

**A detailed structural study of microbial
membrane-intrinsic prenyltransferase and
acyltransferase belonging to two dynamic
superfamilies**

Thesis Submitted to AcSIR
For the Award of the Degree of

DOCTOR OF PHILOSOPHY
In
BIOLOGICAL SCIENCES



By
DEBJYOTI BORAL
Registration Number: 10BB14A26042

Under the guidance of

Dr. V. Koteswara Rao
(Supervisor)

Dr. Sureshkumar Ramasamy
(Co-Supervisor)

Biochemical Sciences Division
CSIR-National Chemical Laboratory
Pune-411008, India

JUNE 2020

Certificate

This is to certify that the work incorporated in this Ph.D. thesis entitled **A detailed structural study of microbial membrane-intrinsic prenyltransferase and acyltransferase belonging to two dynamic superfamilies** submitted by Mr. **Deb jyoti Boral** to Academy of Scientific and Innovative Research (AcSIR) in fulfillment of the requirements for the award of the Degree of **Doctor of Philosophy**, embodies original research work under my/our supervision/guidance. We further certify that this work has not been submitted to any other University or Institution in part or full for the award of any degree or diploma. Research material obtained from other sources has been duly acknowledged in the thesis. Any text, illustration, table etc., used in the thesis from other sources, have been duly cited and acknowledged.

It is also certified that this work done by the student under my supervision, is plagiarism free.



Deb jyoti Boral
(Research Student)



Dr. V. Koteswara Rao
(Research Supervisor)



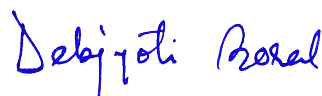
Dr. Sureshkumar Ramasamy
(Research Co-Supervisor)

Place: Pune

Date: 29/06/2020

DECLARATION BY THE CANDIDATE

I hereby declare that the thesis entitled " **A detailed structural study of microbial membrane-intrinsic prenyltransferase and acyltransferase belonging to two dynamic superfamilies** ", submitted for the Degree of **Doctor of Philosophy in Biological Sciences** to Academy of Scientific and Innovative Research (AcSIR) is the record of work carried out by me at Biochemical Sciences Division, CSIR-National Chemical Laboratory, Pune-411008, India under the supervision of **Dr. V. Koteswara Rao (Research guide)** and **Dr. Sureshkumar Ramasamy (Research co-guide)**. The work is original and has not been submitted in part or full by me for any other degree or diploma to any other University. I further declared that the material obtained from other sources has been duly acknowledged in the thesis.



Debjyoti Boral

Biochemical Sciences Division,
CSIR- National Chemical Laboratory,
Pune – 411008, Maharashtra, India

Place: Pune

Date: [29/06/2020](#)

DEDICATED TO.....

Dearest Mom & Dad

Acknowledgement

This thesis marks the completion of the research work that I have undertaken during my tenure as a doctorate student in CSIR-NCL, Pune, India. It's been a long journey full of scientific and philosophical hurdles and a tale of overcoming the same, which has enriched me with knowledge that would hopefully bear fruit in my future. There are numerous individuals who have in many ways contributed to my thesis work and I would like to attempt, to do justice to them in the next few phrases. However, there is no doubt that this list would be incomplete.

Firstly, I would like to express my sincere gratitude to my research guide, Dr. V. Koteswara Rao for his constant support and encouragement during my research work. He has patiently stood by me during tough times and given me absolute freedom to carry out my work and I couldn't be more grateful.

I would like to extend my utmost and heartfelt gratitude to my thesis advisor and co-guide Dr. Sureshkumar Ramasamy for the continuous support of my Ph.D. study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in every aspect of my research and writing of this thesis. He has offered me immense encouragement and helped me solve major scientific roadblocks during my research work. He has been very considerate and kind towards all of my shortcomings and been a constant source of scientific insights, innovative ideas and constructive criticism. I could not have imagined having a better advisor and mentor for my foray into the world of scientific research and consider myself privileged to have worked as a part of his research group.

I would like to thank my previous supervisors, Dr. C. G. Suresh and Dr. Archana V. Pundle for their invaluable contributions to different aspects of my research including moral guidance, abundant support and necessary scientific inputs.

I am grateful to the members of my doctoral advisory committee, Dr. Shubhangi Umbarkar, Dr. Mahesh S. Dharne, and Dr. Mahesh J. Kulkarni for their periodic assessment of my research work, valuable comments, different perspectives, astute suggestions and constant encouragement.

I am eternally grateful to Dr. Abhrajyoti Ghosh for welcoming me to work in his lab at Bose Institute, Kolkata, where an important part of my thesis work was conceived and executed under his supervision. He provided me with the archaeal cultures from which my archaeal lipid extract related experiments were performed. The members of Dr. Ghosh's Lab were extremely warm and helpful to me during my tenure there and without their help this work would have been incomplete. Special thanks to Mousam, Koustav, Shayantan, Arghya and Sayandeep Gupta for helping me with the day-to-day lab work and scientific inputs.

I am sincerely grateful to Dr. Gayathri Pananghat and Dr. Saikrishnan Kayarat of IISER, Pune for allowing me to use their lab facilities as and when required and specially Dr. Gayathri for being available for scientific comments on my research work.

I am immensely thankful to Dr. Subashchandraboese Chinnathambi who supported my work by giving extensive access to his lab facilities. I would also like to extend my sincere thanks to the entire Neurobiology group specially Shweta, Nalini, Abhishek, Tushar, Hari and Rashmi for treating me like an extended lab member and being available all the time for my work.

I am thankful to Dr. Radha Chauhan and Dr. Janesh Kumar for allowing me to use the SEC MALLS and Mosquito facility in NCCS, especially Sangeeta, Ashwini, Ananth and Dr. Pravin Dewangan for helping out during the experiments. I am grateful to Dr. Ravindra Makde, Dr. Ashwani Kumar and Dr. Biplab Ghosh from PXBL21 beamline, Indus-II, RRCAT, DAE, Indore, for allowing to use the synchrotron facilities for XRD data collection of crystals.

My sincere thanks to the H.O.D, Biochemical Sciences, Dr. Ashok P. Giri and Dr. Rakesh Joshi for allowing me to use lab facilities and a special mention should go to Shounak, Varun and Meenakshi for their cooperation and help during my research work.

I feel indebted to Dr. Sushama Gaikwad and my guide Dr. V. Koteswara Rao for helping me with my HPLC experiments, and Sanskruti for her inputs and guidance in CD experiments and data analysis.

I would take this opportunity to thank my seniors and present lab members, Dr. Manas Sule, Dr. Priyabrata Panigrahi, Dr. Ruby Singh, Dr. Deepak Chand, Dr. Manu M. S., Dr. Ameya Bendre, Dr. Yashpal Yadav, Dr. Deepanjan Ghosh, Dr. Vijay Rajput, Shiva Shankar S., Shridhar Chougule, Tejashri Hingmire, Vishwambar Navale and Amol Sawant for being a part of my journey at different stages of my Ph.D. tenure. They were a constant source of positive energy, measured concern and unadulterated fun in the lab. Special mention must go to Shiva Shankar for introducing me to research lab work and contributing to the ground work related to my Ph.D. project and to Deepanjan, Vijay, Yashpal, Deepak Chand, Manu M.S. and Ameya Bendre for their encouragement, comments and support.

This list wouldn't be complete without extending my sincerest thanks to my Ph.D. batch mates Nirbhik, Parag, Kushal, Amarnath, Dr. Amruta, Monika, Sneha, Rajeshwari and Prachi for their help and support throughout my tenure. They were almost as big, a part of my research work, as I was. A special mention to Prajna for her inputs, help and support in my research work. Special credits to Rajeshwari for helping me with the basics of mass spectrometry and also helping me design my experiments. A small note of mention to Prajna and Monika for cheering me up during difficult times and helping me focus back on the job at hand and Priya for being a close friend through trying times. Heartfelt thanks to Parag, Nirbhik and Kushal for being prominently supportive throughout my research tenure. Due credits should go to Sneha, Zenia, Debopriya and Dr. Kiran Kulkarni for the use of X-ray lab facilities for my research work. A ton of thanks to Veena, Santosh, Bhakti, Ashwini, Gopal, Sinduri, Baba, Mrinmayee, Dr. Rahul Salunkhe, Dr. Reema Banarjee, Dr. Anand Sukeerthi, Dr. Avinash Pandreka and all others from Dr. Subashchandraboese Chinnathambi lab, Dr. Dhanasekaran Shanmugam lab, Dr. Santosh Jha lab, Dr. Asmita Prabhune lab, Dr. Mahesh Kulkarni lab, Dr. Kiran Kulkarni lab for being there through thick and thin, at different time points of my tenure at CSIR-NCL. My sincere regards to my trainees Venkat Vaithyanathan and

Sugavaneshwaran K who worked hard with no complaints and other lab trainees for their help in my research. My deepest gratitude to Suraj Patil for contributing notably to my thesis work. Special mention to my MBL lab members, Dr. Prabhakar, Dr. Atul, Dr. Krithika and Dr. Dipesh. My sincerest thanks to all my friends and batch mates from my Masters, especially Dr. Amruta, Manoj and Shubhant for just being there as invisible pillars of support. I am also thankful to all my friends from my Bachelors, specially Nilanjan, Ankita, Arunima, Jayjit and Aniket for their support. I am thankful to my friends in Pune, in and out of CSIR-NCL, hostel mates, SAO personnel, AcSIR office personnel and other nameless and faceless sources of help in CSIR-NCL. A glowing tribute to my teachers in Bachelors and Masters, especially Dr. Anup Kumar Guin, Dr. H.S. Aparna and Dr. Geetha Nagaraj who have inspired me in this journey.

I gratefully acknowledge UGC, New Delhi, India for fellowship. I also thank Director, CSIR-National Chemical Laboratory (CSIR-NCL, Pune) and Chair, Biochemical Sciences for giving me the opportunity to work in this great institution, their support and for providing state of art research infrastructure.

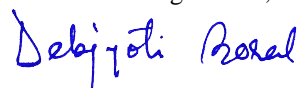
I owe a ton of thanks to my family, specially my cousins Dr. Sreya and Ishita, and all my close friends.

Finally, and most importantly, I would like to thank those who have been the source of my inspiration, my parents.

To my life-coach, my late father, I owe it all to you, forever in life.

To my eternal cheerleader, my mother, I could not have been more blessed to have a selfless parent like you, who has supported me all through my journey till date and has been my constant pillar of support.

With love and gratitude,



DEBJYOTI BORAL

Table of Contents

Serial Number	Title	Page Number
	Table of Contents	I
	List of Figures	IV
	List of Tables	VII
	Abbreviations	VIII
	Abstract	1
	Thesis Overview	3
Chapter I	Introduction	6
1.1	Proteins: Their synthesis, structures and functions	8
1.2	Protein classification based on their localization: Membrane proteins and their subtypes	8
1.3	Introduction to Enzymes	10
1.4	Enzyme classification	11
1.5	Transferases and their subtypes	13
1.6	Prenyltransferases and the UbiA superfamily	15
1.7	Digeranylgeranylglyceryl phosphate synthase and the Archaeal membrane biosynthesis pathway	19
1.8	Archaeal membrane lipids and their uniqueness	22
1.9	Structural Insights into the UbiA prenyltransferase Superfamily	23
1.10	Acyltransferases and the LPLAT superfamily	26
1.11	Acyltransferase of the HtrB/MsbB family and the Raetz pathway	29
1.12	Lipid A and the bacterial lipopolysaccharides layer	32
1.13	Structural Insights into the LPLAT acyltransferase Superfamily	34
1.14	Role of UbiA prenyltransferases in diseases	39
1.15	Role of acyltransferases in diseases	40
1.16	Statement of problem	41
Chapter II	Expression and functional characterization of Digeranylgeranylglyceryl phosphate synthase (DGGGPS)	44
2.1	Introduction	45
2.2	Materials and Methods	46
2.3	Results	59

2.3.1	Isolation of a novel DGGGPS gene from <i>Thermococcus kodakarensis</i>	59
2.3.2	Expression and Purification of Recombinant Tk-DGGGPS and Tk-DGGGPS-Bril	60
2.3.3	Checking the stability of the Protein Detergent Complex	62
2.3.4	Estimating the oligomeric profile of Tk-DGGGPS by SEC MALLS and Blue Native PAGE	63
2.3.5	Gene Cloning of Tk-GGGPS	64
2.3.6	Expression, Purification and confirmation of Tk-GGGPS	65
2.3.7	Enzymatic synthesis of GGGP by Tk-GGGPS	66
2.3.8	Enzymatic activity of Tk-DGGGPS	67
2.3.9	ITC analysis of Tk-DGGGPS with GGPP	70
2.3.10	Circular dichroism study of Tk-DGGGPS	71
2.3.11	Fluorescence spectroscopy of Tk-DGGGPS	72
2.3.12	Surface structure morphology analysis of Tk-DGGGPS by ANS binding	74
2.3.13	Archaeal lipid isolation, archaeosome preparation and visualization (TEM)	76
2.3.14	Interaction of archaeal lipids with Tk-DGGGPS	76
2.3.15	Archaeosome stabilization by Tk-DGGGPS	77
2.4	Discussion	78
Chapter III	Structural insights and computational analysis of Di-geranylgeranyl glyceryl phosphate synthase (Tk-DGGGPS)	81
3.1	Introduction	82
3.2	Materials and Methods	83
3.3	Results	91
3.3.1	Initial crystallization trials of Tk-DGGGPS and Tk-DGGGPS-Bril variants	91
3.3.2	Crystallization trials of Tk-DGGGPS with advanced membrane protein crystallization screens and additives	92
3.3.3	Crystal optimization of Tk-DGGGPS by HILIDE and LCP methods	93
3.3.4	Tk-DGGGPS mutant purification and crystallization	94
3.3.5	Screening detergents for the optimal PDC (Protein Detergent complex)	96
3.3.6	Tk-DGGGPS purification in new detergents with low micellar size	97
3.3.7	Sequence analysis of DGGGPS and other UbiA superfamily members	99
3.3.8	Phylogenetic analysis of DGGGPS and other UbiA superfamily members	103
3.3.9	Structure prediction and Homology modelling of Tk-DGGGPS	105
3.3.10	Molecular docking studies to understand the dynamics at the ligand binding pocket	110

3.3.11	Studying the oligomeric interface	116
3.3.12	Studying the behavior of the Tk-DGGGPS in a membrane environment	119
3.3.13	Study of disease-causing mutations in human homologs of DGGGPS	126
3.4	Discussion	130
Chapter IV	Structural Insights into a Membrane Intrinsic Acyltransferase from <i>Chlorobium tepidum</i>	135
4.1	Introduction	136
4.2	Materials and Methods	137
4.3	Results	140
4.3.1	Isolation of a Novel Acyltransferase Gene of <i>Chlorobium tepidum</i>	140
4.3.2	Expression and Purification of Recombinant <i>Chlorobium tepidum</i> Acyltransferase	141
4.3.3	Sequence analysis of Ct-acyltransferase and other LPLAT superfamily members	143
4.3.4	Phylogenetic Analysis of the LPLAT acyltransferases	149
4.3.5	Homology Modeling: Molecular Models of Ct-Acyltransferase	150
4.3.6	Salient structural characteristics of Ct-acyltransferase	151
4.3.7	Membrane interaction and Oligomeric assembly of Ct-acyltransferase	155
4.4	Discussion	159
Chapter V	Summary and Inferences	160
	Bibliography	171
	List of Publications	179

List of Figures

Figure Number	Title	Page Number
Chapter I		
Fig 1.1	Types of membrane protein (Schematic representation)	9
Fig 1.2	General enzyme mechanism of UbiA prenyltransferases	16
Fig 1.3	Dendrogram of UbiA superfamily members	17
Fig 1.4	Reactions catalyzed by UbiA prenyltransferases	19
Fig 1.5	Reaction catalyzed by DGGGP synthase	20
Fig 1.6	Diagrammatic representation of the Archaeal membrane lipid biosynthesis pathway	21
Fig 1.7	Structural differences between archaeal ether lipids and bacterial & eukaryotic ester lipids	22
Fig 1.8	Core structural components of archaeal membrane lipids	23
Fig 1.9	Crystal structure of the archaeal homolog Ap-UbiA (PDB: 4OD5)	24
Fig 1.10	Crystal structure of the archaeal homolog Af-UbiA (PDB: 4TQ3)	25
Fig 1.11	Dendrogram of the LPLAT acyltransferase superfamily	28
Fig 1.12	Diagrammatic representation of the Lipid A biosynthetic 'Raetz' pathway	30
Fig 1.13	Detailed diagram of the components of the bacterial LPS layer	33
Fig 1.14	Crystal structure of squash GPAT	35
Fig 1.15	Crystal structure of mycobacterial PatA	36
Fig 1.16	Crystal structure of <i>A. baumannii</i> LpxM	37
Fig 1.17	Crystal structure of <i>T. maritima</i> AGPAT	38
Fig 1.18	UbiA homologs associated with human diseases	40
Fig 1.19	Acyltransferases associated with human diseases	41
Chapter II		
Fig 2.1	Map of vectors	49
Fig 2.2	Flow diagram for the steps involved in lipid extraction from archaea	57
Fig 2.3	Cloning of TK1957 in pET28a and pET28a-BRIL plasmids	60
Fig 2.4	Purification of Tk-DGGGPS	61
Fig 2.5	Purification of Tk-DGGGPS-BRIL	62
Fig 2.6	Ultracentrifugation Dispersity Sedimentation assay	63
Fig 2.7	Oligomeric profiling of Tk-DGGGPS	64
Fig 2.8	Cloning of TK1026 in pET28a plasmid	65
Fig 2.9	Purification of Tk-GGGPS	66
Fig 2.10	The two step reaction to generate the archaeol intermediate Di-o-geranyl geranyl glyceryl phosphate	67
Fig 2.11	Verification of the enzyme reactions by TLC	68
Fig 2.12	HPLC profiles of the reaction mixtures	69
Fig 2.13	MS analysis of the reaction mixtures	69

Fig 2.14	Isothermal titration calorimetry measurement of Tk-DGGGPS and GGPP binding interactions	70
Fig 2.15	Far-UV CD spectra analysis of Tk-DGGGPS	71
Fig 2.16	Global structure analysis and estimation of T _m of Tk-DGGGPS from Far-UV CD analysis	72
Fig 2.17	Intrinsic fluorescence spectra of Tk-DGGGPS	73
Fig 2.18	ANS binding dynamics of Tk-DGGGPS	75
Fig 2.19	TEM visualization of the archaeosomes generated from the Saci_lipids	76
Fig 2.20	SDS-PAGE of pull down analysis of Tk-DGGGPS with archaeal total lipid extracts	77
Fig 2.21	Steady-state fluorescence anisotropy measurements of DPH incubated with Tk-DGGGPS	78
Chapter 3		
Fig 3.1	Diagrammatic representation of the HILIDE technique of crystallization to increase the probability of formation of ordered Type I crystals	84
Fig 3.2	Sequence electrogram of site-directed mutagenesis	86
Fig 3.3	Protocol overview for detergent screening of membrane proteins	87
Fig 3.4	Crystals obtained using the PACT suite	92
Fig 3.5	Crystals obtained using membrane protein specialized screens	93
Fig 3.6	Crystals obtained using the HILIDE method	94
Fig 3.7	A comparison of the yields of the mutants and wild type Tk-DGGGPS	95
Fig 3.8	The entire set of detergents used for the high throughput detergent screening with the best hits highlighted	96
Fig 3.9	Tk-DGGGPS purified in CYMAL-6 detergent	97
Fig 3.10	Multiple sequence alignment of the representative sequences of the UbiA superfamily, homologous to DGGGPS	98
Fig 3.11	Multiple sequence alignment of DGGGPS sequences from different classes of archaea	100
Fig 3.12	Consurf sequence conservation diagram of Tk-DGGGPS	101
Fig 3.13	The sequence conservations of DGGGPS arranged according to archaeal taxonomic classes	102
Fig 3.14	The HHpred sequence conservation diagram of the sodium glucose transporter from <i>Vibrio parahaemolyticus</i> (PDB: 2XQ2)	102
Fig 3.15	Phylogenetic tree of all the UbiA homologs of Tk-DGGGPS	103
Fig 3.16	Phylogenetic tree of DGGGPS from different archaeal taxonomic classes	104
Fig 3.17	Overall structure of Tk-DGGGPS	106
Fig 3.18	Particulars of the Tk-DGGGPS homology model	107
Fig 3.19	Overview of the Tk-DGGGPS structure	108
Fig 3.20	Electrostatic surface potential of Tk-DGGGPS structure	110
Fig 3.21	GGPP docked in the predicted active site	111
Fig 3.22	GGPP docked in the predicted allosteric site of Tk-DGGGPS	112
Fig 3.23	GGGP docked in the predicted active site of Tk-DGGGPS	114
Fig 3.24	GGGP docked in the predicted active site of GGPP bound Tk-DGGGPS	115
Fig 3.25	DGGGP docked in the predicted active site of Tk-DGGGPS	116
Fig 3.26	Detailed molecular interactions in the oligomeric assembly of Tk-DGGGPS	117

Fig 3.27	Lipid Bilayer with Tk-DGGGPS embedded	118
Fig 3.28	Snapshots of the Tk-DGGGPS structure before and after the simulation	120
Fig 3.29	A detailed analysis of the 50 ns simulation of Tk-DGGGPS in an ether lipid bilayer	122
Fig 3.30	Superpose analysis of the 50 ns membrane simulation of Tk-DGGGPS	123
Fig 3.31	Mutation hotspots in Tk-DGGGPS and their connection to its human UbiA homologs implicated in diseases	127
Fig 3.32	Mutation hotspot residues of Tk-DGGGPS from STRUM webserver	129
Chapter 4		
Fig 4.1	Cloning of CT0211 in pET28a and pET28a-BRIL plasmids	141
Fig 4.2	Purification of Ct-acyltransferase	142
Fig 4.3	SEC profile of Ct-acyltransferase	143
Fig 4.4	Multiple sequence alignment of the Ct-acyltransferase homologs belonging to the LPLAT superfamily	145
Fig 4.5	Multiple sequence alignment of the HtrB/MsbB family of acyltransferases	146
Fig 4.6	Match correlation matrix of Ct-acyltransferase	147
Fig 4.7	Phylogenetic tree of LPLAT acyltransferases	148
Fig 4.8	Overall structure of Ct-acyltransferase	150
Fig 4.9	Particulars of the Ct-acyltransferase homology model	151
Fig 4.10	Overview of the Ct-acyltransferase structure	152
Fig 4.11	The Ct-acyltransferase enzyme active site pocket	153
Fig 4.12	A schematic diagram of the structural components of Ct-acyltransferase	154
Fig 4.13	Molecular details of the Ct-acyltransferase active site pocket	155
Fig 4.14	The membrane interacting structural components of Ct-acyltransferases	156
Fig 4.15	A comparative view of the spatial location of the oligomeric interface of Ct-acyltransferase	156
Fig 4.16	The transmembrane domain of Ct-acyltransferase	157
Fig 4.17	Molecular interactions in the oligomeric interface of Ct-acyltransferase	158

List of Tables

Table Number	Title	Page Number
Table 1.1	Membrane protein distribution among different classes of enzymes	13
Table 2.1	Details of the primers used in the study	48
Table 2.2	PCR reaction mixture for amplification of the TK1957 gene	48
Table 3.1	A list of new hydrogen bonds formed after the 50ns simulation involving conserved residues of Tk-DGGGPS	125
Table 4.1	Details of the primers used in the study	138

Abbreviations

ABBREVIATION	FULL FORM
μL	Microliter
AA	Amino acid
AH	Amphipathic helix
AU	Absorbance unit
BLAST	Basic local alignment search tool
BP	Base pair
CD	Circular dichroism
Ct-Acyltransferase	HtrB/MsbB acyltransferase from <i>Chlorobium tepidum</i>
CV	Column volume
DAB	3,3'-diaminobenzidine
DDM	N-Dodecyl-β-D-Maltopyranoside
DGGGP	(S)-2,3-Di-O-geranylgeranyl glyceryl phosphate
DGGGPS	(S)-2,3-Di-O-geranylgeranyl glyceryl phosphate synthase
DM	N-Decyl-β-D-Maltopyranoside
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
GGGP	Sn-3-O-(Geranylgeranyl) glycerol 1-phosphate
GGPP	Geranylgeranyl pyrophosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria-Bertani media
LDAO	N-Dodecyl-N,N-Dimethylamine-N-Oxide
NM	Nanometer
OD	Optical density
PBS	Phosphate buffer saline
PBST	Phosphate buffer saline Tween20
PCR	Polymerase chain reaction
PDB	Protein data bank
PIC	Protease inhibitor cocktail
PMSF	Phenylmethylsulfonyl fluoride
RMSD	Root mean square deviation
RPM	Revolutions per minute
SDS	Sodium dodecyl sulfate
Tk-DGGGPS	(S)-2,3-Di-O-geranylgeranyl glyceryl phosphate synthase from <i>Thermococcus kodakarensis</i>
Tk-GGGPS	Sn-3-O-(Geranylgeranyl) glycerol 1-phosphate synthase from <i>Thermococcus kodakarensis</i>
T_m	Melting temperature
TM	Transmembrane
β-ME	β-mercaptoethanol
β-OG	Octyl-β-D-glucopyranoside

ABSTRACT

Among the transferases, which are perhaps the most diverse yet ubiquitously present class of enzymes in the living world, the prenyltransferases and acyltransferases are perhaps the most dynamic in terms of the substrates they modify enzymatically. By transferring carbon chain backbones to small core biomolecules like glycerol and disaccharides, these two transferases play a major role in shaping the external boundaries of both prokaryotic and eukaryotic life forms. Their involvement can be observed in crucial steps of major biochemical processes like respiration, photosynthesis and metabolism of various biologically important molecules. As a result, both prenyltransferases and acyltransferases have been extensively studied but the primary focus has been on the ones found in the cytosolic fraction of the cell. Not much is known about the prenyl- and acyltransferases which are present in the membrane fraction of the cell. Two prominent superfamilies, namely the UbiA superfamily of prenyltransferases and the LPLAT superfamily of acyltransferases containing enzymes located in the membrane fraction of cells are therefore the present focus of many research studies. Both these superfamilies include a diverse range of enzymes, in terms of the biochemical pathways they are a part of. Among the UbiA prenyltransferases, the enzyme (*S*)-2,3-Di-*O*-geranylgeranylgeranyl phosphate synthase (DGGGPS), involved in the prenylation of archaeal ether-linked membrane lipid precursors, plays a notable role in the structural uniqueness of the membrane lipids between archaea and other prokaryotes and eukaryotes. One of the unique characteristic features of the domain archaea, are the lipids that form the hydrophobic core of their cell membrane. These membrane lipids are characterized by unique isoprenoid biochemistry and the building blocks are two core lipid structures, *sn*-2,3-diphytanyl glycerol diether (archaeol) and *sn*-2,3-dibiphytanyl diglycerol tetraether (caldarchaeol). Both archaeol and caldarchaeol have phytanyl chains (C₂₀) in a bilayer structure, two and four respectively, connected to the glycerol moiety by an ether bond which is the primary uniqueness between these archaeal and other prokaryotic and eukaryotic membrane lipids. Also DGGGPS has prominent homologs in humans and other higher mammals, which are implicated in several human diseases. The HtrB/MsbB acyltransferases on the other hand are responsible for catalyzing the final step of Lipid A biosynthesis in Gram negative bacteria. Lipid A is an integral part of the bacterial liposaccharides (LPS) which anchors the LPS layer to outer

membrane of the bacteria. The Lipid A moiety is responsible for the endotoxin activity shown by the LPS and induces innate immune response activation in the human body by TLR4 activation, primarily. A lack of the secondary acylations by HtrB/MsbB acyltransferases results in increased permeability in the OM and higher susceptibility towards cationic antimicrobial peptides (CAMPs) and other bactericidal compounds. There are human homologs of HtrB/MsbB acyltransferases which are involved prominently in some common human diseases. These two enzymes have major significance, as potential drug targets against human diseases and towards the control of pathogenic bacteria, apart from having novel therapeutic applications in drug delivery and separation processes, among others.

This thesis includes a structural and functional characterization of the Tk-DGGGPS and Ct-acyltransferase, along with thorough computational analysis to study the conserved core structures in the prenyltransferases and acyltransferases, respectively and how that translates into their functional distinctiveness. From the above mentioned studies, we also hypothesize new moonlight functionalities of DGGGPS and a probable feedback control mechanism for efficient energy resource conservation in prokaryotes.

THESIS OVERVIEW

Chapter 1: Introduction

A general introduction about proteins, enzymes and their classification have been covered in this chapter, followed by an extensive review of literature about prenyl- and acyltransferases and the present status of research on both. The chapter then delves into further detail about the enzymes in focus, the Di-geranylgeranylgeranyl phosphate synthase (DGGGPS) and the HtrB/MsbB acyltransferase, highlighting their structural and functional features and the biochemical pathways they are involved in. It also includes a comparative analysis of the structural properties of the other homologs of the membrane intrinsic prenyltransferase and acyltransferase superfamilies, respectively. The probable roles of the human homologs in these superfamilies in human physiology have also been emphasized.

Chapter 2: Expression and functional characterization of Digeranylgeranylgeranyl phosphate synthase (DGGGPS)

This chapter deals with the cloning, expression and purification of Tk-DGGGPS and its functional and biophysical characterization. Heterologous expression of archaeal DGGGPS was achieved in *E. coli* with and without a helical apocytochrome tag. The substrate specificity of the enzyme was verified along with the stereo specificity of the first enzyme (Tk-GGGPS) of membrane lipid biosynthesis pathway. The thermodynamic parameters of the interaction between the prenyl chain donating ligand (GGPP) and the enzyme were analyzed with identification of different binding sites for ligand binding resulting in a hypothesis about the mode of action of the enzyme. The lipid interaction dynamic of the Tk-DGGGPS, studied by various biophysical methods, highlighted the prominent affinity of the enzyme to interact with a variety of lipids, especially with lipids of archaeal origin. Structure transition studies were carried out on the enzyme as well, which gave novel insights into the thermal stability of the enzyme as well its effective catalysis at higher temperatures. The presence of Tk-DGGGPS in archaeal membrane was also found to increase the stability of the membrane indicated by an increase in anisotropy measurements after treatment with DPH.

Chapter 3: Structural insights and computational analysis of Di-geranylgeranyl glyceryl phosphate synthase (Tk-DGGGPS)

In this chapter a detailed structural characterization of Tk-DGGGPS was attempted using X ray crystallography, employing crystallization techniques developed especially for membrane protein crystallization. Tk-DGGGPS mutants were generated using site-directed mutagenesis to address the entropic limitations of a few residues on the protein surface which could have an impact on the conformational stability of the protein three dimensional structure. We obtained positive results showing higher protein yield and faster crystal nucleation in these mutants. Due to the detergent component of the PDC (Protein Detergent complex) being a crucial factor in the crystallization of a membrane protein, a high throughput detergent screening was carried out to identify detergents with a micelle better suited for the co-purification of Tk-DGGGPS. Sequence conservations in the enzyme were analyzed and important residues were identified which were involved in ligand interactions and further verified by a membrane simulation. A moonlight functionality of the enzyme was hypothesized on basis of the simulation results and secondary structure based conservation analysis. Also mutational hotspots in Tk-DGGGPS were reported, some of which coincided with the known mutations in the human UbiA homologs of Tk-DGGGPS, thereby lending pathological significance to these newly identified hotspots.

Chapter 4: Structural Insights into a Membrane Intrinsic Acyltransferase from *Chlorobium tepidum*

The Lipid A species are synthesized through a highly conserved multistep biosynthetic pathway whose final step is catalyzed by acyltransferases of the HtrB/MsbB family. The contents of this chapter represent a detailed study on a putative dual functional HtrB/MsbB acyltransferase from *Chlorobium tepidum* (Ct-acyltransferase). In this study, the enzyme was cloned and purified to homogeneity via heterologous expression in a bacterial system. Further, the conserved patches of amino acid sequences were analyzed towards detailed understanding of their involvement in structural stability of the enzyme as well as the probable functional implications. The prominent dual-storied hydrophobic pocket hosting the active site was studied, which was unique in the Ct-

acyltransferase from other HtrB/MsbB acyltransferases. The predicted oligomeric assembly and membrane association interfaces were further studied and proposed.

Chapter 5: Summary and inferences

The principle findings of the research work and the conclusions drawn from the same are discussed in this chapter. The future prospects of the work and the putative therapeutic applications are highlighted as well.

Chapter I – Introduction

The ancestry of living organisms had been one of the earliest areas of interest among the scientifically inclined, since biology came into existence as an area of learning. In 1962, Roger Stanier and C. B. van Niel, in their widely regarded research article, first brought about the concept of a broad classification of all living organisms into prokaryotes and eukaryotes, on basis of cellular organization, wherein they defined prokaryotes to be devoid of a cell nucleus [1]. The system of five Kingdoms thereby came to be accepted as the standard mode of phylogenetic classification which included the prokaryotes (bacteria) and the eukaryotes (plants, animals, fungi, & protists). However, in 1977, Carl Woese and George E Fox brought about a major change in the ‘Tree of Life’ by discovering a new group of prokaryotes named Archaea, on basis of phylogenetic taxonomy of the 16S ribosomal RNA. The word *archaea* was derived from the ancient Greek word “ἀρχαῖα”, meaning "ancient things". Later, Woese proposed that life should be divided into three domains: Archaea, Bacteria, and Eukarya on basis of phylogenetic relationships instead of the then prevalent classification based on morphological similarities [2].

The domains Archaea and Bacteria were composed of only unicellular organisms while Eukarya primarily consisted of multicellular organisms. Bacteria were found to be the more dynamic among the prokaryotes in terms of habitats while Archaea were notably found to thrive in extreme habitats such as hot springs and salt lakes, hence popularly termed as “Life’s extremists”. Apart from their vast difference in genetic makeup, bacteria and archaea were also very distinctive in the composition of their external envelope. The highlights of the archaeal external envelope were the composition of their cell membrane, where the membrane lipids varied from those in the other domains of life in terms of their unique chemical nature and stereochemical structure [3-7]. In bacteria (mainly the Gram negative subtype), the uniqueness of the cell envelope lied in the second lipid membrane or the outer lipid membrane, which contains lipopolysaccharides (LPS layer) in their outer leaflet and phospholipids in the inner leaflet [8]. It is of notable mention, that the above described unique features of archaea and bacteria keenly influence the lipid biochemistry of the respective organisms and are maintained by defined pathways. These pathways are shepherded by specialized proteins called enzymes, which belong to different superfamilies and have characteristic mechanisms by which they catalyze their substrates.

Our work is centered on enzymes belonging to two such superfamilies, one belonging to the prenyltransferases and the other belonging to the acyltransferases, where we try and understand

the structure function relationships of these enzymes from different organisms which would give us interesting insights into the respective diverse superfamilies.

1.1 Proteins: Their synthesis, structures and functions

Proteins (derived from the word *proteios*) are the most abundant and versatile among biological macromolecules. All of them are polypeptides constructed from a ubiquitous set of standard 20 amino acid residues, linked by a substituted amide linkage called peptide bonds [9]. The protein diversity information stored in the genetic material in the DNA, is translated into different combinations and sequences of these 20 different amino acids [10, 11].

The *primary structure* of the enzyme is nothing but a characteristic linear arrangement of amino acids. This linear polypeptide chain folds into regular recurring *secondary structural elements* namely α -helices, β -sheets, loops and turns, whose formation is influenced by the intramolecular interactions between the amino group and the carbonyl group of different amino acid residues [12]. Further folding of the *secondary structure* forms a three-dimensional structure known as the *tertiary structure* of the protein, which is generally the functionally active form of the same. Certain enzymes are made up of two or more polypeptide chains/subunits and assembles in the form of a *quaternary structure* [13, 14]. Proteins also interact among each other and with other biomolecules like lipids, nucleic acids and carbohydrates to form macromolecular assemblies known as *quinary* structures which are known to modulate the protein folding and stability [15].

Proteins are known to perform a diverse range of functions ranging from acting as catalysts to biochemical reactions in the cell, being involved in transport of other proteins and small molecules within and between cells, acting as storage molecules for the cell, acting as structural components to provide structural integrity to the cell to being involved in signal transduction and amplification via protein networks [16].

1.2 Protein classification based on their localization: Membrane proteins and their subtypes

In a living cell, proteins can be localized either in the cell membrane or in the cytoplasm based on the overall nature of the amino acid residues exposed on the protein surface. If the protein surface is composed of mainly hydrophilic residues, the protein is soluble in water and hence are known as soluble proteins. Soluble proteins can be of two types, fibrous proteins (known to play a relatively inert structural role in cellular functions) and globular proteins (known to play major

functional roles like catalysis or regulatory roles) [17]. However, if the protein surface contains primarily hydrophobic residues, then it is unable to exist, surrounded by water, in the cytoplasm and is therefore found to be present in the cell membrane where their hydrophobic regions are masked by the lipid bilayer. These proteins are known as membrane proteins.

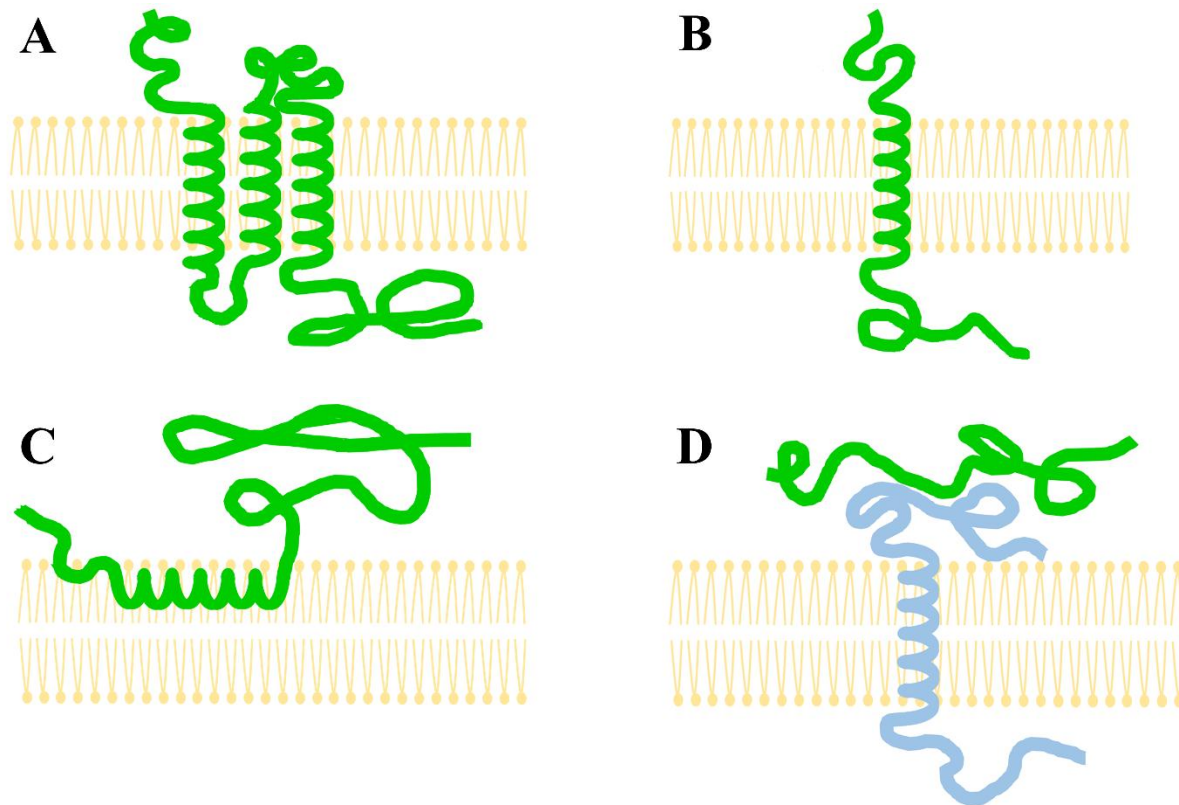


Fig 1.1: Types of membrane protein (Schematic representation) (A) Integral Polytopic, (B) Integral Bitopic, (C) Integral Monotopic and (D) Peripheral.

Membrane proteins can be both structurally important proteins and functionally important enzymes. Among notable functions of membrane proteins are signal transduction between the cell's internal and external environments, transport of molecules and ions across membranes and their involvement in immune response mechanisms. Membrane proteins are implicated in a plethora of human diseases and is thereby targeted by most present medicinal drugs. There are two major types of membrane proteins, namely the integral membrane proteins and peripheral membrane proteins (Fig 1.1A, D). The integral membrane proteins have transmembrane domains spanning across the lipid bilayer whereas the peripheral membrane proteins are temporarily attached to the lipid bilayer or to the integral membrane proteins through an amalgam of

electrostatic, hydrophobic and other non-covalent interactions. The integral membrane proteins can also be classified on the basis of their relationship to the lipid bilayer, into three following subtypes. A) The integral polytopic proteins which span the entire length of the membrane more than once; B) The bitopic proteins which span the lipid bilayer only once; C) The integral monotopic proteins which are attached to one side of the membrane and do not span the entire lipid bilayer (Fig 1.1A, B, C) [18].

1.3 Introduction to Enzymes

All the biochemical reactions which take place in a living organism are catalyzed by specialized proteins, known as enzymes. The term *enzyme* was first coined by Frederick Wilhelm Kühne in 1877, derived from the Greek word *énsymo* (*én* meaning ‘in’; *symo* meaning ‘yeast’). Enzymes are highly selective bio-catalysts which controls the rate and specificity of diverse metabolic reactions in the living cell. Most enzymes are proteinaceous in nature with only a few exceptions like catalytic RNA. Most enzymes require non-protein cofactors (either coenzymes, prosthetic groups or metal-ion activators) for catalyzing biochemical reactions. Enzymes act by lowering the activation energy of biochemical reactions thereby maintaining them at the specific rates necessary to sustain life [19, 20]. Every enzyme has a specific optimum pH and optimum temperature where it functions most efficiently and any change in these factors adversely affect the rate of catalysis due to a resulting change in the enzyme stability. Enzymes generally bind to a variety of substrates with different binding specificity and affinity and this is controlled by certain key residues in the active site of the protein which are often important for catalysis, hence called catalytic residues. The active site generally consists about ten to twelve amino acids [21] and histidine, cysteine and aspartic acid are most frequently involved in catalytic activities among all amino acids [22]. Enzymes are known to have four different types of specificities as follows [23].

- *Absolute specificity* – enzyme specific to only one substrate molecule (Examples: *Lactase* which acts only on lactose; *Carbonic anhydrase* which acts only on carbonic acid)
- *Group specificity* – enzyme specific to substrates containing a specific functional group (Example: *Pepsin* which is an endopeptidase responsible for hydrolysis of the central peptide bond involving amino group of aromatic amino acids; *Amino peptidase* which is an exopeptidase responsible for hydrolysis of amino terminal peripheral peptide bonds)

- *Linkage specificity* – enzyme specific to substrates having particular types of chemical bond (Example: *Fructose 1,6-Bisphosphate Aldolase* which acts only on fructose 1,6-bisphosphate; *Lipase* which hydrolyzes ester bonds in triglycerides)
- *Stereo specificity* – enzyme specific to only one stereoisomer of a substrate (Example: α -glycosidase & β -glycosidase/amylase which acts only on α -glycosidic & β -glycosidic bonds respectively; D-lactate dehydrogenase which acts on D-lactic acid only)

In 1833, Anselme Payen and Jean-François Persoz discovered the first enzyme, diastase/amylase which ushered in the fermentation technology. James Batcheller Sumner provided a breakthrough in enzyme studies in 1926 when he isolated and crystallized the urease enzyme. He then followed it up by isolating and crystalizing another enzyme, catalase, in 1937. He along with John Howard Northrop and Wendell Meredith Stanley definitively proved the hypothesis that enzymes are pure proteins and shared the Nobel Prize in Chemistry with them in 1946. In 1960, Christian Boehmer Anfinsen Jr. hypothesized that ‘*at least for a small globular protein in its standard physiological environment, the native structure is determined only by the protein's amino acid sequence*’ on basis of his protein folding experiments on Ribonuclease A. The X-ray crystallographic structure of lysozyme solved in 1965 marked the beginning of the molecular elucidation of enzyme biochemistry and reaction mechanisms [24].

1.4 Enzyme classification

All known enzymes are categorized into six main enzyme classes, namely:

1. Oxidoreductases – They catalyze the transfer of electrons from a reductant (electron donor) to an oxidant (electron acceptor) molecule and mostly use nicotinamide adenine dinucleotide phosphate (NADP) or nicotinamide adenine dinucleotide (NAD) as cofactors. Systemic names are in a form of ‘*donor: acceptor oxidoreductase*’; while in most cases ‘*donor dehydrogenase*’ is much more common. Oxidoreductases can be either oxidases or dehydrogenases and they play significant roles in both aerobic and anaerobic metabolisms.
2. Transferases – They catalyze the transfer of specific functional groups (prenyl, acetyl, phosphate groups etc.) from one donor molecule to an acceptor molecule. Systematic names assigned to transferases are ‘*donor: acceptor grouptransferase*’, while ‘*acceptor*

grouptransferase or *donor grouptransferase* is more common. Hydrogen is however not considered as a functional group in case of transferases.

3. Hydrolases – They are a class of hydrolytic enzymes which generally catalyzes the breakage of a chemical bond like C-O, C-N, C-C and others, utilizing water, in order to divide a large molecule into two smaller ones. Hydrolases are pivotal for the conversion of large molecules into fragments for synthesis, for excreting waste materials, and for providing carbon sources for the production of energy, during which many biopolymers are converted to monomers. Systemic names assigned are *'substrate hydrolases'*, while they are more commonly referred to as *'substrate-ase'*.
4. Lyases – They catalyze the cleavage of various chemical bonds like C-O, C-N, C-C and others by means of an 'elimination' reaction which often results in the formation of a new cyclic structure or a new double bond. A reverse reaction known as a "Michael addition" might also possibly happen under the catalysis of lyase. While only one substrate is required for the reaction in the forward direction, but two substrates are essential for the reverse reaction. Systemic names assigned are *'substrate group-lyases'*. Lyases generally have narrow substrate specificity.
5. Isomerases – They catalyze reactions involving geometrical or structural changes within a molecule, i.e. stereoisomers and structural isomers. Racemases, epimerases and cis-trans isomerases are involved in formation of stereoisomers, whereas intramolecular lyases, oxidoreductases and transferases catalyze the formation of structural isomers.
6. Ligases – They catalyze reactions joining two molecules through the formation of new chemical bonds (such as C-O, C-S, C-N), by hydrolysis of energy rich diphosphate bonds of ATP/GTP. T4 DNA and RNA ligases are the most commonly known among this class of enzymes. Systemic name assigned to them are *'XY ligases'*, while synthetase is the common name as they are involved in the synthesis of new compounds.
7. Translocases – They catalyze the movement of molecules and ions across cell membranes and are even involved in the separation of ions and molecules within membranes. Common names for this class of enzymes are transporters.

International Union of Biochemistry and Molecular Biology created a system for enzyme nomenclature where enzymes could be assigned unique code numbers based on the chemical

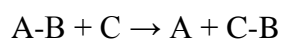
reactions they catalyzed. The code number known as Enzyme Commission (EC) number have the letters EC followed by four numbers punctuated by dots. The first number represents the mechanism of the reaction (the major enzyme classes), the second number represents the subclass, the third stands for the sub-subclass and the last one is the enzyme serial number in its sub-subclass. For example, EC 1.1.1.1 is the nomenclature of alcohol dehydrogenases which performs redox reactions, hence belong to the class of *oxidoreductases*. The present study involves the *transferase* class of enzymes (EC 2.x.x.x) [25].

Table 1.1 Membrane protein distribution among different classes of enzymes

Enzyme classes	Membrane protein distribution	Examples
Oxidoreductases	Abundant	Glutathione-Insulin transhydrogenase (EC 1.8. 4.2) Succinate: Quinone oxidoreductases (EC 1.3.5.1)
Transferases	Abundant	Digeranylgeranylglyceryl phosphate synthase (EC 2.5.1.42) Lipid A biosynthesis myristoyltransferase (EC 2.3.1.243)
Hydrolases	Abundant	Microsomal epoxide hydrolase (EC 3.3.2.9) Mammalian serine amidase (EC 3.5.1.99)
Lyases	Sparse	Sphingosine-1-phosphate lyase 1 (EC 4.1.2.27)
Isomerases	Very sparse (only domains present)	Peptidylprolyl isomerase or PPLase domain in SurA (Outer Membrane Protein Chaperone)
Ligases	Sparse	E3 Ubiquitin ligase (EC 6.3.2.19)
Translocases	Moderately abundant	ABC-type oligosaccharide transporter (EC 7.5.2.2) TIM17-23 complex

1.5 Transferases and their subtypes

Transferases are involved in the transfer of functional groups (acyl, glycosyl, amino, prenyl, phosphoryl groups etc.) from a donor molecule to an acceptor molecule. Most transferases are very specific in identifying the biochemical groups and this contributes to their substrate selectivity. This selectivity is a defining feature of the transferases and is regulated by the residues present in the active site of the enzyme as well as by the conformation of the active site pocket of these enzymes. The basic mechanism of reaction catalyzed by the transferases are:



Here ‘A’ is the donor, ‘C’ is the acceptor and ‘B’ is the group transferred in the reaction. In many cases the ‘donor’ is also a coenzyme.

Transferases are involved in many different biochemical pathways distributed across various living organisms, and are essential for most of the important biological processes. There are more than

450 different unique enzymes of transferases which has been identified and classified till date [26]. Transferases are categorized into ten subtypes, primarily on basis of the nature of functional group transferred by the enzymes.

- Single carbon group transferases (EC 2.1); Example: Amidotransferases and Methyltransferases
- Aldehyde and ketone group transferases (EC 2.2); Example: Transketolases and Transaldolases
- Acyl group transferases (EC 2.3); Example: Lauroyl transferases and Peptidyl transferase (the ribozyme)
- Glycosyl, pentosyl and hexosyl group transferases (EC 2.4); Example: Lactose synthases and Glycogen synthases
- Alkyl and aryl group transferases (EC 2.5); Example: Prenyltransferases and Cysteine synthases
- Nitrogenous group transferases (EC 2.6); Example: Transaminases and oximinotransferases
- Phosphorus group transferases (EC 2.7); Example: Cyclin-dependent kinases (CDKs)
- Sulfur group transferases (EC 2.8); Example: Alcohol sulfotransferases
- Selenium group transferases (EC 2.9); Example: Selenocysteine synthases
- Metal group transferases (EC 2.10); Example: Molybdopterin molybdotransferases

Perhaps the most well-known among transferases are the alpha 1-3-N-acetylgalactosaminyltransferases and alpha 1-3-galactosyltransferases (A and B glycosyltransferases, respectively) in the human ABO blood group system[27]. ‘A glycosyltransferases’ are involved in the biogenesis of the A-antigen by addition of N-acetylgalactosamine to the H-antigen and ‘B glycosyltransferases’ are involved in the addition of a galactose molecule to H-antigen synthesizing the B-antigen. Both these transferases are highly similar in respect to their amino acid sequences only differing at four positions and only this variation is enough to alter their substrate specificity.

1.6 Prenyltransferases and the UbiA superfamily

Prenyltransferases (EC 2.5.1.1) transfer prenyl groups from donor to acceptor molecules. The donor molecules are mostly linear polyprenyl pyrophosphates (lipids containing one or more than one isoprene units)[28]. The most common prenyl donors used by these enzymes are IPP/DMAPP (Isopentenyl pyrophosphate/ Dimethylallyl pyrophosphate; C₅), GPP (Geranyl pyrophosphate; C₁₀), FPP (Farnesyl pyrophosphate; C₁₅) and GGPP (Geranylgeranyl pyrophosphate; C₂₀). The acceptor molecules catalyzed by these enzymes are far more diverse in their structure ranging from linear compounds to cyclic compounds.

The mechanism involved is a condensation reaction between the first and last carbon atoms of the acceptor and donor polyprenyl diphosphates, respectively. In case of IPP as a donor, the C₄ atom condenses with the C₁ of an allylic acceptor pyrophosphate. In 1968, Cornforth and Popjak hypothesized that these enzymes could trigger prenyl transfer/hydrolysis by removal of the negative charge from the pyrophosphate moiety of the allylic substrate. However further studies suggested that initial ionization could generate an allylic cation which then condenses with isopentenyl pyrophosphate or it could also take place in assistance of the π electrons in the double bond of isopentenyl pyrophosphate, implying the bonding event between the two substrates to be closely related with ionization. The elimination of a proton from C₂ completes the condensation reaction to form a double bond between C₂ and C₃. Finally, the pyrophosphate group previously released from the allylic substrate would assist in elimination of the proton. The common name used for this class of enzymes are prenyl diphosphate synthases. Prenyltransferases also cause hydrolysis of a higher carbon prenyl diphosphate molecule by catalyzing the C₁-O bond and the active site involved for both the prenyl transfer and hydrolysis reactions are the same [29]. The concentration of inorganic pyrophosphate, which is the by-product of the enzyme reaction regulates the rate of the hydrolysis. The predicted general enzyme mechanism of the UbiA superfamily of enzymes is shown in Fig 1.2.

Prenyltransferases are known to be involved in terpenoid metabolism in archaea, bacteria, fungi, animals and plants. Terpenoids are known precursors for sterol/steroid production in animals which act as precursors to vitamins A, D, E and K and provide structural stability to cell membranes, while in plants, they have diverse functional roles, including acting as growth

hormones, photosynthetic pigments, electron carriers and structural membrane components (phytosterols). Most of the available knowledge about prenyltransferases and their roles in cellular processes are limited to a certain category, i.e. the prenyltransferases located in the cytosolic fraction of the cell. However, there is not much detailed study on the prenyltransferases which are located in the membrane fraction of the cell. These prenyltransferases are categorized into a superfamily of enzymes known as the UbiA superfamily.

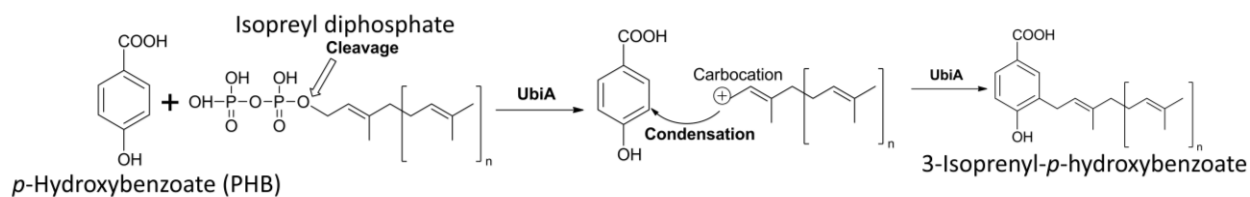


Fig 1.2 General enzyme mechanism of UbiA prenyltransferases

The UbiA superfamily of intramembrane prenyltransferases catalyzes vital biosynthetic steps in the production of ubiquinones, menaquinones, plastoquinones, hemes, chlorophylls, vitamin E, and structural lipid components of membranes. These lipophilic compounds perform a varied range of functions, like acting as electron and proton carriers for cellular respiration and photosynthesis, as antioxidants to reduce cell damage, and as structural components of archaeal membranes [30]. These integral membrane proteins have been generally composed of 7–9 transmembrane segments. The superfamily contains the following enzymes (Fig 1.3):

1. 4-hydroxybenzoate polyprenyltransferases (bacteria, archaea and eukaryotic mitochondria),
2. DHNA (1,4-dihydroxy-2-naphthoic acid) prenyltransferases (bacteria, archaea and human mitochondria),
3. Protohaem IX farnesyltransferase/ haem O synthase (eukaryotes and aerobic microbes),
4. Chlorophyll synthase (plants and photosynthetic bacteria),
5. Digeranylgeranylgeranyl phosphate synthase (archaea),
6. Homogentisate phytyl- and prenyltransferases (plants and cyanobacteria)
7. Decaprenyl-phosphate 5-phosphoribosyl transferase synthase (Corynebacteriaceae).

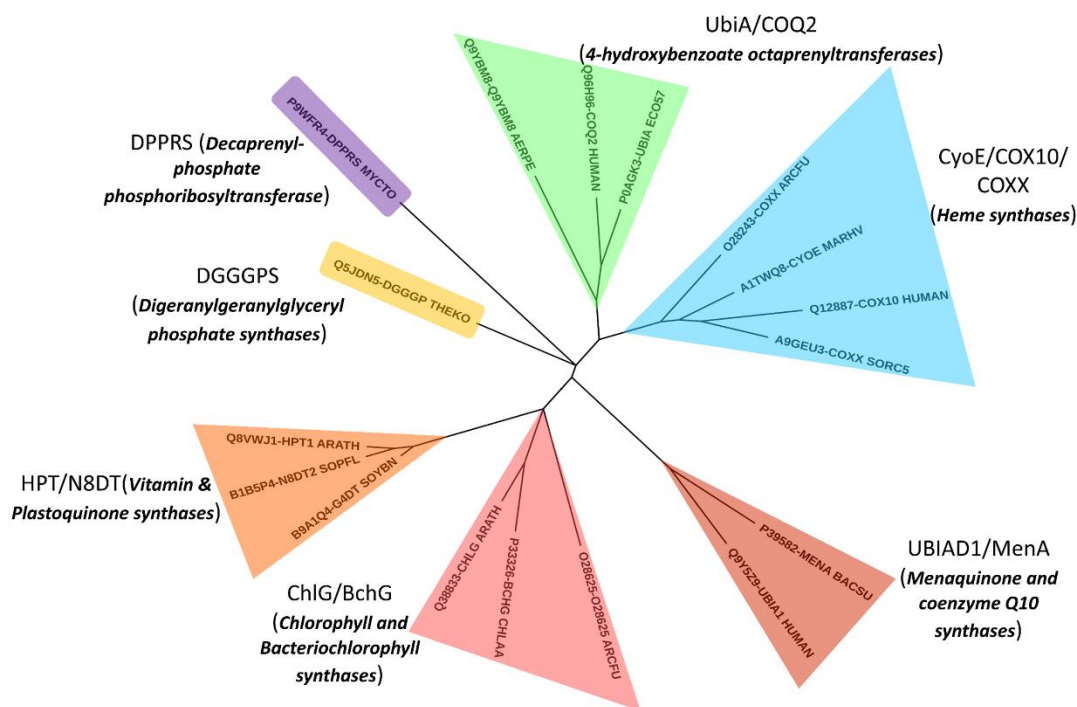
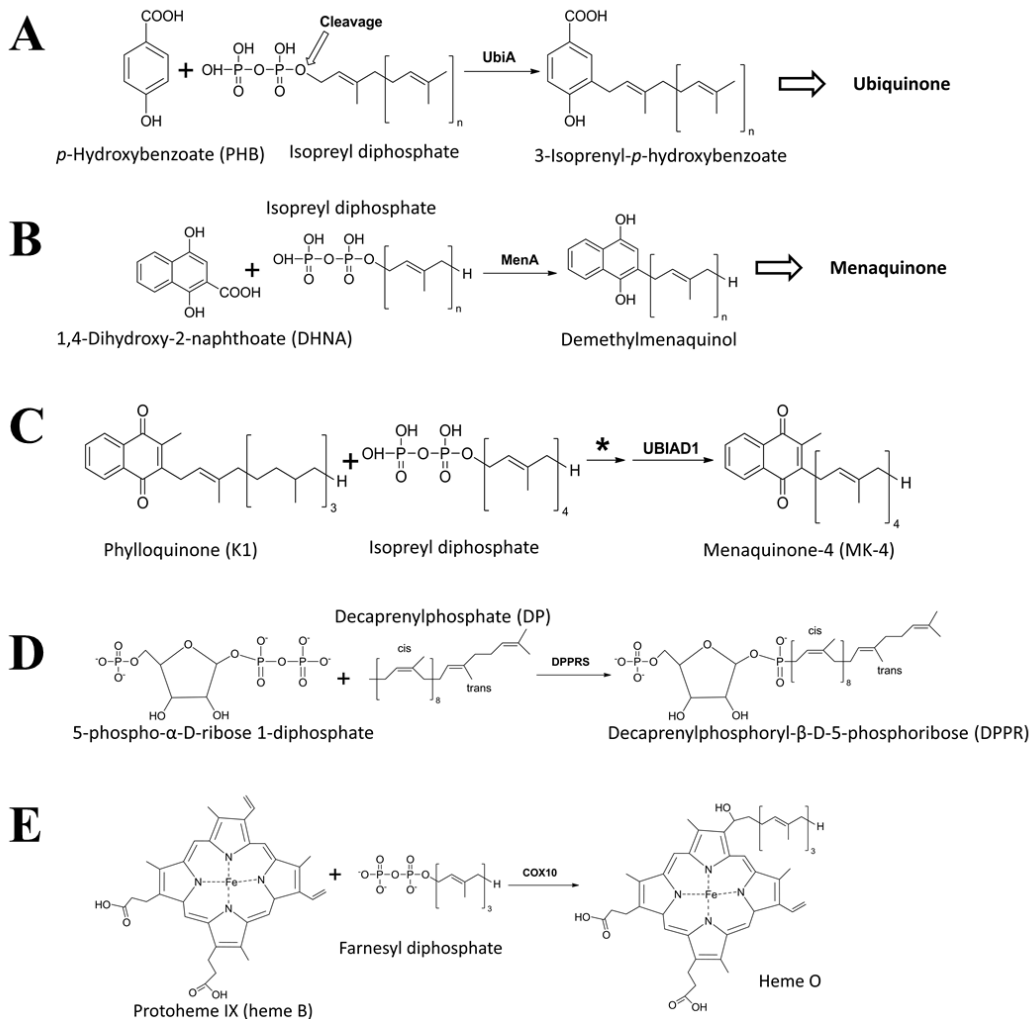


Fig 1.3 Dendrogram of UbiA superfamily members Generated using the MEGA suite, this tree clearly shows the clustering among the UbiA prenyltransferases, which is mainly on the basis of enzymes having similar functions.

The UbiA superfamily of enzymes are involved in numerous biochemical pathways (Fig 1.4), distributed over a wide variety of species and thereby differ significantly in their amino acid sequences and homologies. This make them functionally diverse and enable them to act on chemically different and/or structurally related substrates which are the hallmarks of a *superfamily* of enzymes [31]. Only recently two crystal structures of enzymes were solved in this superfamily, both are from the domain Archaea [32, 33]. Structural elucidation of these enzymes might be the key to understanding the ability of these enzymes to catalyze such diverse substrates and take part in so many different biochemical pathways while still preserving the same core structure. Also, the obvious differences in the chemical nature among these cytosolic and membrane prenyltransferases, warrants a detailed study of these membrane prenyltransferases to have a greater depth in understanding the functionality and importance of prenyltransferases in life processes.

The members of the UbiA Superfamily are enzymes that catalyze transfer of isoprenoid moieties to acceptor molecules which are majorly aromatic ring containing compounds to generate a wide spectrum of lipophilic end products which are essential for different metabolic processes in living

organisms. However very interestingly, one member of this superfamily, DGGGPS have been reported to use, as prenyl acceptor, a linear prenyl chain containing compound, namely (S)-3-O-geranylgeranylgeranyl glyceryl phosphate (GGGP). While the members of the superfamily are quite flexible in its choice of substrates, this specific member bears a certain uniqueness from the structural perspective. Though it remains unclear how these intra- membrane enzymes select their prenyl donor and acceptor substrates, a detailed study of the interacting residues in the active pocket of the enzymes from this family, more so from DGGGPS, will give us a notable insight into the same and with it give us the opportunity to bioengineer the enzyme and possibly generate novel lipids which might have commercial or biomedical applications.



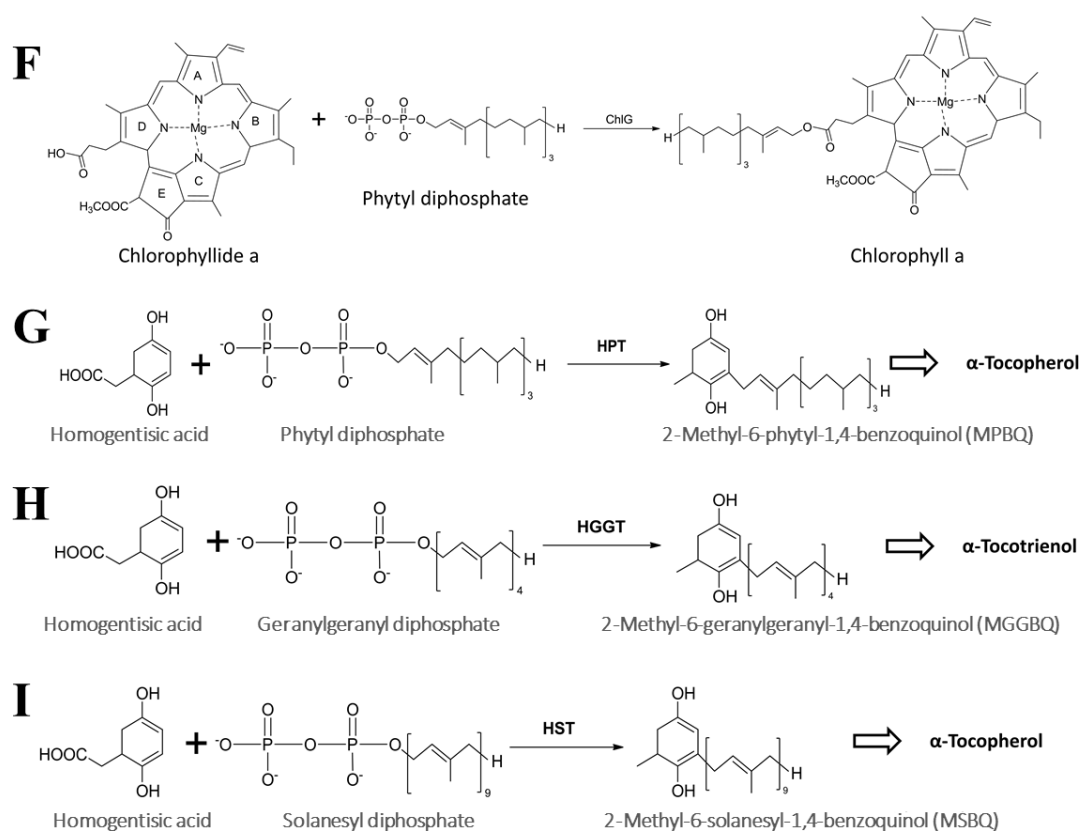


Fig 1.4 Reactions catalyzed by UbiA prenyltransferases (A) UbiA 4-hydroxybenzoate polyprenyltransferase, (B) DHNA (1,4-dihydroxy-2-naphthoic acid) prenyltransferase, (C) UbiA prenyltransferase domain-containing (UBIAD1) protein, (D) DPPR (decaprenyl-phosphate 5-phosphoribosyltransferase) synthase, (E) Protohaem IX farnesyltransferase, (F) Chlorophyll synthase, (G) (H) & (I) Homogentisate phytyl and prenyltransferases.

1.7 Digeranylgeranylgeranyl phosphate synthase and the Archaeal membrane biosynthesis pathway

(S)-2,3-Di-O-geranylgeranylgeranyl phosphate synthase (DGGGPS) catalyzes the transfer of a geranylgeranyl group from geranylgeranyl diphosphate (GGPP) to (S)-3-O-geranylgeranylgeranyl phosphate (GGGP), to form digeranylgeranylgeranyl phosphate (DGGGP) (Fig 1.5). Based on sequence alignment and phylogenetic analysis, DGGGPS appeared to be a member of a large family of proteins, known as the UbiA prenyltransferase superfamily [34, 35]. This family of protein is known to consist of prenyltransferases that are involved in the biosynthesis of diverse compounds like respiratory quinones, hemes, chlorophylls, vitamin E, and shikonin by catalyzing the fusion of a phytyl or isoprenyl chain to an aromatic ring. The prenyl acceptor in DGGGP

synthase is a linear compound, which is a notable exception to the aromatic ring structures used by other prenyltransferases in the superfamily.

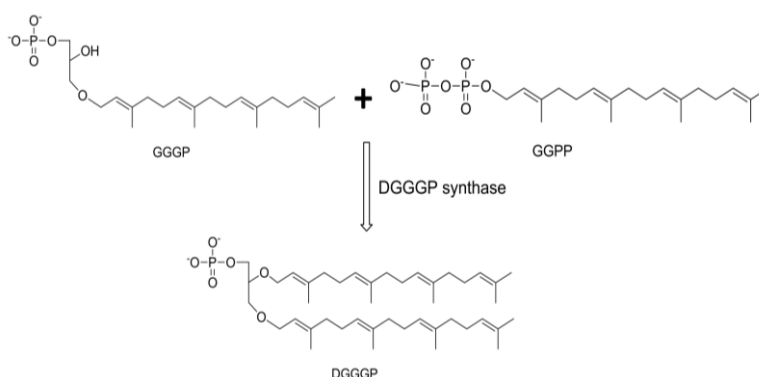


Fig 1.5 Reaction catalyzed by DGGGP synthase GGPP is the prenyl chain donating ligand while GGPP is the acceptor. It is also the first reaction of the archaeal membrane lipid biosynthesis pathway to take place in the membrane itself.

The initiation of the archaeal isoprenoid biosynthesis occurs via two major building blocks, namely isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). In Archaea, these building blocks are obtained via the ‘alternative’ Mevalonate pathway, characteristic of this domain of life. A similar version of this pathway, known as the ‘classical’ Mevalonate pathway are observed in most eukaryotes (and Sulfolobales) and differs in the last three steps from the Archaeal Mevalonate pathway [36]. However, most bacteria including *Escherichia coli*, obtain IPP via the MEP/DOXP pathway [37, 38]. To synthesize archaeol, IPP undergoes isomerization to DMAPP by IPP isomerase [39]. IPP molecules consequently condense with DMAPP, and undergoes chain elongation to initially form geranyl diphosphate (GPP; C₁₀), which then generates geranylgeranyl diphosphate (GGPP; C₂₀), by addition of IPP units via condensation reactions, catalyzed by the enzyme, GGPP synthase. The glycerol back bone, as mentioned above, is provided by sn-glycerol-1-phosphate (G1P), which is produced by the reduction of dihydroxy acetone phosphate (DHAP), in presence of NADH/NADPH, with the enzyme involved being the Zn²⁺ dependent Glycerol-1-phosphate dehydrogenases (G1PDH). Following this occurs the first distinct step of ether lipid biosynthesis, where GGPP is attached to G1P via an ether bond to form geranylgeranylglyceryl phosphate (GGGP), catalyzed by GGGP synthases (GGGPs). A second GGPP is attached to GGGP via the second ether bond to give digeranylgeranylglyceryl phosphate (DGGGP), catalyzed by DGGGP synthases (DGGGPs) [40]. Next, utilizing CTP as a substrate,

cytidine-diphosphate (CDP) archaeol synthases replace the phosphate group of the unsaturated DGGGP with CDP, generating unsaturated CDP-DGGGP. A CMP molecules is then removed from the CDP headgroup, only to be replaced by a polar headgroup (like Serine, Ethanolamine) by CDP-alcohol phosphatidyl transferases. The next reaction that follows, is the saturation/hydrogenation of the side chains, thought to be mediated by Geranylgeranyl reductases in a stereospecific manner in presence of flavin adenine dinucleotide (FAD). The formation of the tetraether lipids (GDGTs) is thought to involve a head to head coupling between the two archaeol lipids, followed by internal cyclization to form cyclopentane and cyclohexane moieties in certain archaea [41]. However, the reaction specifics in this last step is yet to be determined completely (Fig 1.6).

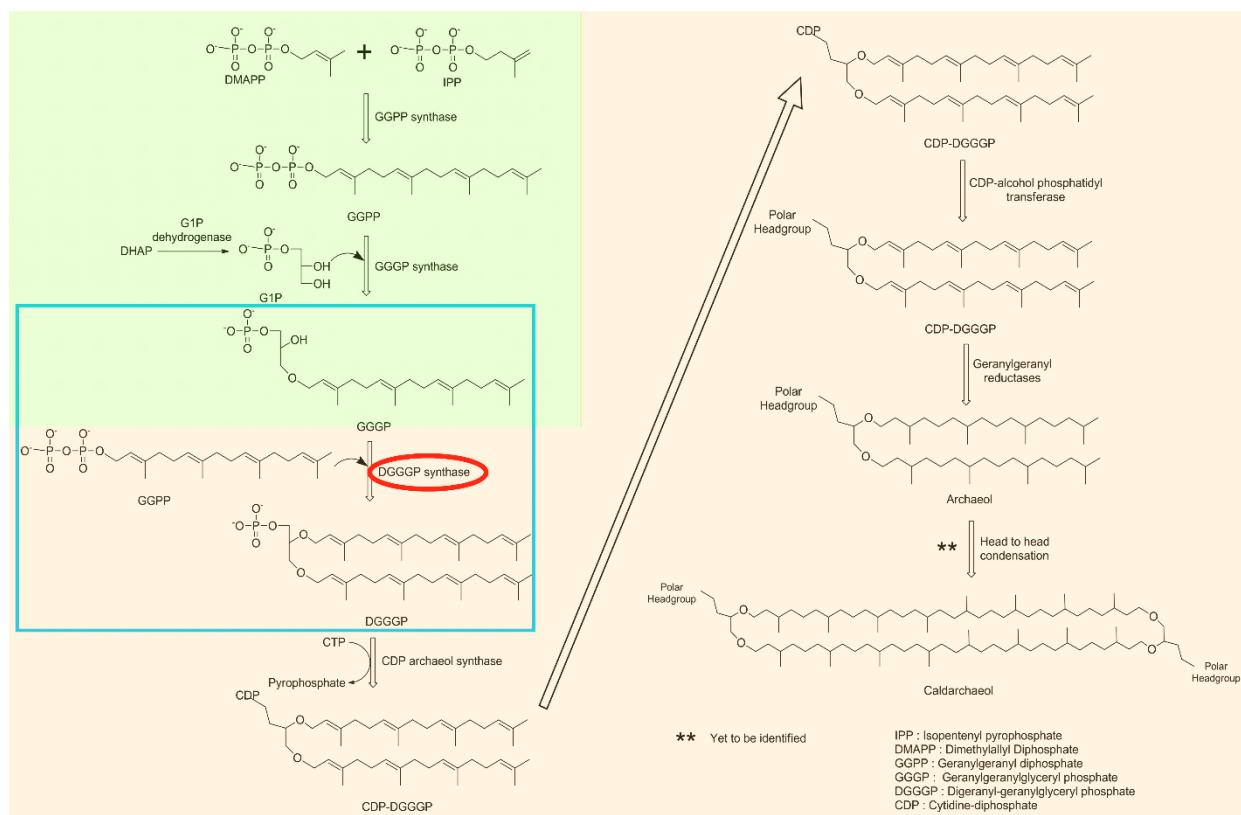


Fig 1.6 Diagrammatic representation of the Archaeal membrane lipid biosynthesis pathway The first step shown is a common step for many other biosynthetic pathways while the others are unique to this pathway. The enzyme in focus is highlighted in a blue box. While the first two steps take place in the cytosol (in light green backdrop), the rest take place in the plasma membrane (in light orange backdrop).

1.8 Archaeal membrane lipids and their uniqueness

The membrane lipids of Archaea have unique structural features, which sets it apart from the other two domains of life, Eubacteria and Eukarya. Although they essentially are glycerolipid analogues from bacteria or eukaryotes, they have the following distinct structural characters (Fig 1.7) [42].

- 1) The hydrocarbon chains of archaeal lipids are branched isoprenoids typically derived from geranylgeranyl diphosphate (GGPP), while the bacterial/eukaryotic lipids are made up of linear acyl groups.
- 2) In Archaea, the isoprenoid chains are linked with the glycerol moiety with ether bonds, while the hydrocarbon moieties are linked by ester bonds in their bacterial and eukaryotic counterparts.
- 3) The glycerol moiety of archaeal lipids is derived from sn-glycerol-1-phosphate (G-1-P), which is an enantiomer of sn-glycerol-3-phosphate, the precursor for bacterial/eukaryotic glycerolipids.

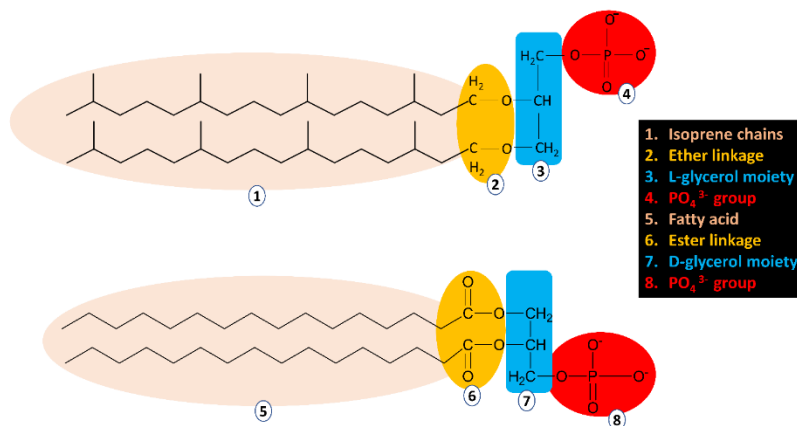


Fig 1.7 Structural differences between archaeal ether lipids and bacterial & eukaryotic ester lipids The differences lie not only in the kind of bonds which links the glycerol moiety and the carbon chains, as is apparent from the names of the lipids, but also in the carbon chains which are saturated in ether lipids unlike the ester lipids, and the enantiomeric properties of respective lipids

The archaeal and bacterial membrane lipids have opposite chiralities at their glycerol moieties i.e. the glycerol moieties of the archaeal and bacterial glycerolipids are sn-2,3-di-O-alkylated and sn-1,2-di-O-acylated, respectively [41]. Even though the core membrane lipid structures show occasional variation among archaeal species, the most commonly found basic structural forms are the glycerol diether lipids and the glycerol tetra ether lipids (Fig 1.8). The glycerol di-ether core lipid structure, 2,3-di-O-phytanyl-sn-glycerol, also called archaeol, includes two fully saturated C20 isoprenoid chains. The tetra-ether core lipid, sn 2,3-di- O-dibiphytanyl-sn-diglycerol, (also

known as glycerol dibiphytanyl glycerol tetraether (GDGT) and caldarchaeol), appears to be constructed by the head to head coupling of two archaeol molecules, linked covalently at the reduced hydrocarbon chain ends to give a bipolar, macro-cyclic tetra-ether structure containing two C40 isoprenoid chains enabling the formation of monolayer membranes. GDGTs can also contain 0–8 cyclopentane moieties. These isoprenoid ether lipids, which constitute the lipid component of archaeal membranes, are a distinguishing evolutionary feature of the archaea.

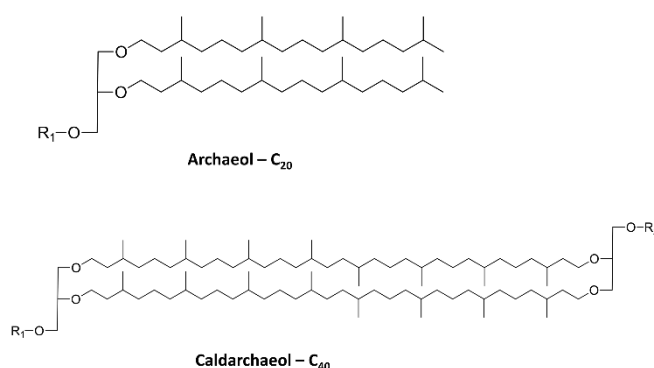


Fig 1.8 Core structural components of archaeal membrane lipids Archaeol are diether lipids (C₂₀) and Caldarchaeol are tetraether lipids (C₄₀) formed by head-to-head condensation of two archaeol molecules. Archaeol is the basic unit of archaeal membrane lipids.

1.9 Structural Insights into the UbiA prenyltransferase Superfamily

Recently, crystal structures of two paralogs of DGGGPS have been determined [32, 33] from hyperthermophilic archaeal species, *Aeropyrum pernix* (Ap) and *Archaeoglobus fulgidus* (Af) respectively. The Af structure complexed with GPP and Mg²⁺ (PDB id: 4tq3) was determined at 2.41 Å resolution, and has five water molecules in the active site which indicates there might be an involvement of the same in conjunction with two Mg²⁺ ions, in the prenyl transfer reactions. The Af apo structure (4tq5) had a resolution of 3.07 Å. The Ap apo structure (4od4) was determined at a resolution of 3.3 Å and its PHB and geranyl thiolodiphosphate (GSPP) bound form (4od5) was solved at a resolution of 3.6 Å. Together, these structures have contributed significantly, towards the understanding of the enzymatic mechanism used by the UbiA superfamily.

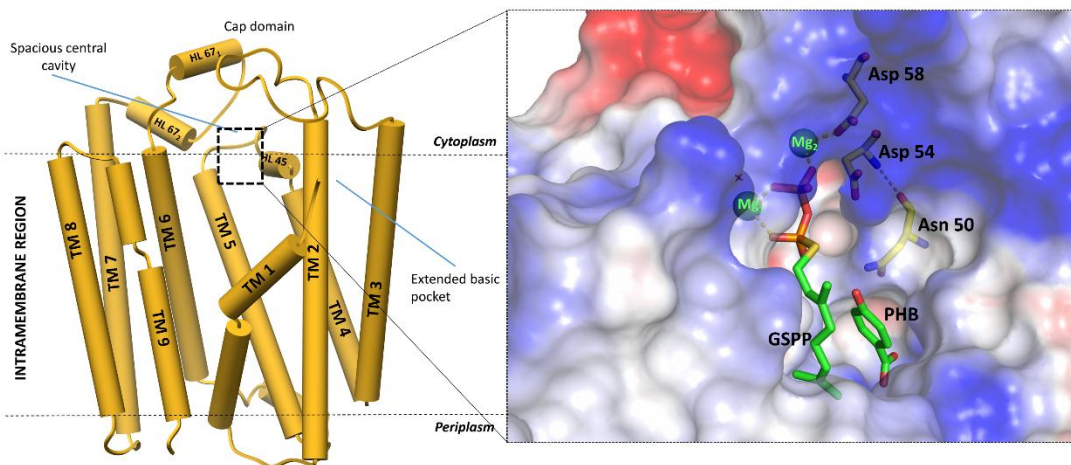


Fig 1.9 Crystal structure of the archaeal homolog Ap-UbiA (PDB: 4OD5) A stick model representation of the enzyme shows the nine transmembrane helices of the enzyme and a surface view of the substrate binding pocket is included with detailed inter-molecular interactions.

The crystal structure of the Ap-UbiA homolog [33] shows the presence of nine transmembrane domains (TMs), arranged in a clockwise fashion forming a reverse U-shaped architecture with a spacious central cavity (Figure 1.9). The central cavity is surrounded by TM1, TM2, and TM9 at the front and TM5 and TM6 at the backside when viewed from the cytoplasmic side. The C terminal ends of TM2, TM4, and TM6 are found to be kinked, generating a short extra-membrane helix followed by a C terminal loop. These helix/loop regions are termed HL2-3, HL4-5, and HL6-7, respectively and form the extra-membrane cap domain over the central cavity and contain most of the conserved residues, including two Asp rich motifs that are essential for the activity of various enzymes in the superfamily [43-46]. The non-cleavable XPP analog, geranyl thiolodiphosphate (GSPP), essentially a competitive inhibitor, binds in the central cavity via interactions between extensive networks of conserved residues. Both Asp-rich motifs form coordinate bonds with Mg²⁺ ions and engage the diphosphate group of GSPP. One Mg²⁺ ion may bridge the O1 atom of the pyrophosphate to Asp 54 and Asp 58, and the second Mg²⁺ ion may bridge the O₂ atom to Asp 182 and Asp 186 through a water molecule. (Figure 13; right) Asp 54 and Asp 182 seem to have an additional role of positioning the side chains of Arg 63 and Arg 67, respectively. These two arginine residues, together with Tyr 115 and Lys 119, may in turn interact with the other oxygen of the pyrophosphate group. A hydrophobic residue at the bottom and a small basic pocket, are two other components and primary determinants of the central cavity, which were proposed to

bind the isoprenyl chain and an aromatic substrate, respectively [33]. Location of this possible binding pocket is aptly supported by the finding, that residues in TM5 determine the specificity for aromatic substrates [47]. The central cavity has a lateral opening to the lipid bilayer (Figure 13; left), which creates a passage to the active site that might facilitate the binding of long-chain substrates and release of products in membranes. This capacious binding chamber might be the secret behind the chain-length promiscuity of UbiA, which accepts a variety of XPP lengths to generate UQ6–10 in different species.

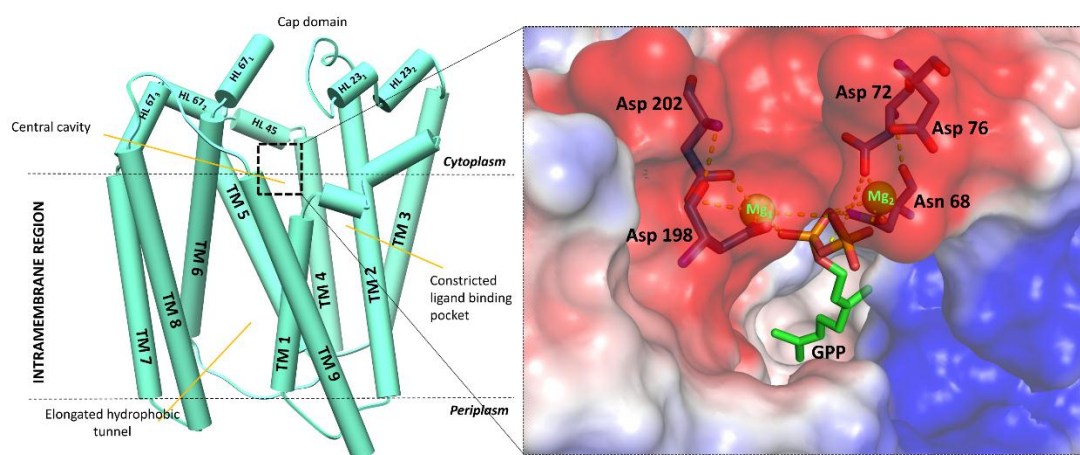


Fig 1.10 Crystal structure of the archaeal homolog Af-UbiA (PDB: 4TQ3) The transmembrane helices of the enzyme are shown in a pipe cartoon representation with the substrate binding pocket in a surface charge distribution view.

The overall architecture of the Af-UbiA homolog [32] is similar to that of Ap-UbiA. However, the two substrate-binding pockets are quite distinctly different. Unlike the Ap structure, the Af structure, has one small substrate binding pocket, restricted by the protein backbone (Figure 1.10) (not capable of accommodating longer XPPs) and a significantly longer tunnel leading from the active site, which might be the putative binding site for longer XPPs. In this scenario, the probable explanation would be that, the prenyl-acceptor substrate of Af might bind to the longer tunnel, while the small pocket might bind the prenyl donors like GPP or DMAPP in this case. The long tunnel in Af is due to the larger spacing between TM9 and TM8 (Figure 14; left). Also, TM9 and TM1 are much closer together in Af, hence the direct opening of the Af central cavity is blocked to lipids. In this structure, the predicted HL4-5 is missing, though not because of lack of residue

conservation, but due to different conformations adopted by the extramembrane cap region of this enzyme.

The Ap and Af homologs have distinct substrates and functions, as is apparent from their structural differences and these structural variations indicate that the difference in positions of TM1, TM8 and TM9 and the presence/absence of the helix loop regions HL2-3, HL4-5 and HL6-7 might be the factor which determines the substrate specificity of the enzymes belonging to this superfamily. The varied spacing between the TMs and the changes in critical residues in the active site (either charges on the specific residue or even the type of the residue), together might generate considerable variations in the two substrate-binding pockets, thereby allowing different prenyltransferases to use remarkably distinctive substrates. However, XPP is the prenyl donor substrate used most commonly by the members of the superfamily, sans DPPR synthase, therefore the residues that recognize the diphosphate groups of XPP are likely to be conserved.

1.10 Acyltransferases and the LPLAT superfamily

Long chain hydrophobic acyl residues play vital roles in many important biological structures. The processes like forming an integral part of the inner hydrophobic layers of biological membranes, functioning as intracellular storage compounds and modifying protein properties/function as membrane anchors. Activation of the acyl chains are generally caused by esterification of the carboxyl groups with the thiol group of coenzyme A (CoA) or of the acyl carrier protein (ACP), generating acyl-thioesters. Acyltransferases (EC 2.3.1.1) or transacylases utilize these acyl-thioesters and polymerize them into lipophilic storage compounds or catalyze their transfer to different substrates to synthesize many biologically relevant molecules [48]. Acyltransferases can be roughly divided into two groups with respect to their catalytic mechanism. The majority of them (GPAT, LPAAT, Lpx acyltransferases, Polyketide-associated acyltransferases, and Chloramphenicol acetyltransferase) are characterized by a conserved Hxxx(x)D/E-like motif in their active site with a catalytically active histidine residue and an acidic aspartate or glutamate and involves the formation of a noncovalent transition state, while the others like lipases, RtxC acyltransferases, PhaG transacylases, or PHA synthases employ a catalytic triad and form covalent intermediates during catalysis. The acyltransferases involving a crucial

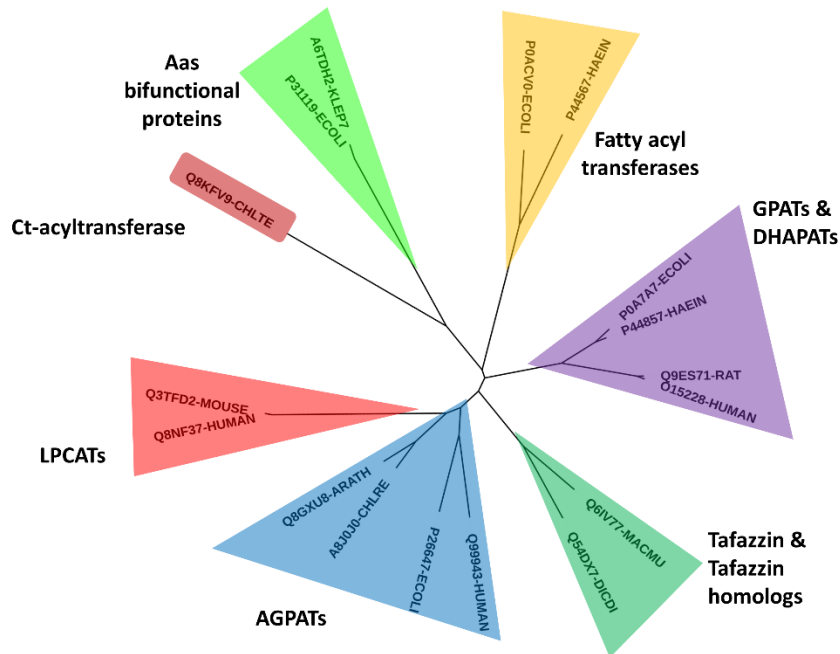
catalytic histidine residue at their active site are generally considered as “real acyltransferases” since the others resemble hydrolases in their catalytic mechanism [49].

Acyltransferases are primarily membrane associated or membrane intrinsic in nature which might be due to their involvement in generating components of the cell membrane or cell envelope. However, the acyltransferases involved in manufacture of the storage lipids, antibiotic resistance factors and pathogenic virulence factors are also seen to be membrane intrinsic which indicates that in contrast to prenyltransferases, these class of enzymes, even though with completely different structures and/or catalytic mechanisms, primarily love the hydrophobic lipid environment for their functional mechanisms. The only notable exceptions are the Early Lpx acyltransferases (LpxA and LpxD) for synthesis and modification of Lipid A, which are soluble proteins in nature. These enzymes are also known to be functionally dependent oligomers which indicates that often the structural assembly of these homo oligomeric enzymes plays a key part in their catalytic mechanisms.

The general catalytic mechanism involves the catalytically active histidine in the heptapeptide motif HX₄DG. To enable the nucleophilic attack on the acyl donor, the histidine in the above mentioned motif, acts as a base and initiates the deprotonation of the hydroxyl group of the substrate. This deprotonation mechanism varies subtly over the different acyltransferases. In some acyltransferases like GPATs, the negative residue aspartate in the motif forms a charge relay system with histidine, while in others like the AtfA acyltransferases involved in storage lipid synthesis and Chloramphenicol acyltransferases, the aspartate residue isn't of major significance. This may be due to the difference in conformation of the active site of the above enzymes and the spatial arrangement of the amino acid residues. The generated oxyanion then initiates a nucleophilic attack to form the oxoester bond, while a proton transfer occurs from histidine to the CoA-S residue, resulting in the release of a CoA-SH molecule and subsequent regeneration of the catalytic histidine residues [50].

Among the entire class of acyltransferases, the Lpx acyltransferases involved in biogenesis of Lipid A have both soluble (LpxA and LpxD) and membrane bound members (LpxL, LpxM and LpxP). The membrane bound members composes the HtrB/MsbB family. The soluble counterparts have been studied in detail while only very recently a crystal structure of one of the membrane

counterparts (LpxM; 5KN7,5KNK) have been solved. The membrane bound members, also known as the “Late Lpx acyltransferases”, belong to a diverse superfamily of proteins, the lysophospholipid acyltransferase (LPLAT) superfamily. They catalyze the final steps of the Raetz pathway, by performing secondary O-acylations at the attached R-3-hydroxy fattyacyl chains, at positions 2' and 3' of KDO₂-lipid IVA to produce lipid A. This superfamily also contains the glycerol-3-phosphate acyltransferases (GPATs) and acylglycerolphosphate acyltransferases (AGPATs), the dihydroxyacetone-phosphate acyltransferases (DHAP-ATs) and the membrane-bound O-acyltransferases (MBOATs). The soluble GPATs and membrane intrinsic AGPATs are involved in triglyceride biosynthesis in bacteria while the DHAP-ATs along with GPATs and AGPATs are involved in the pathway of glycerophospholipid metabolism in mammals (Fig 1.11).



25

Fig 1.11 Dendrogram of the LPLAT acyltransferase superfamily Generated using the MEGA suite, this tree shows the distinct clustering among the LPLAT acyltransferases, where each cluster has acyltransferases which are involved in unique biochemical pathways.

The MBOATs are involved in storage lipid biosynthesis, protein/peptide acylation, and membrane phospholipid remodeling in mammalian species. The MBOATs are also unique in their conserved motifs among the other LPLAT members. They do have the conserved catalytic His residue, but lack the aspartate/glutamate residue which indicates that no charge relay system occurs here in the

deprotonation mechanism [51]. The motif containing the histidine residue looks like WHGxxxGYxxxF and the presence of Tyr and Phe residues in them indicate a slightly different reaction mechanism of these proteins. Four crystal structures have been solved from this superfamily till date, a stromal GPAT from *Cucurbita moschata* which is only distantly related to bacterial LPLATs [52], a membrane associated PatA from *Mycobacterium smegmatis* [53], an AGPAT from *Thermotoga maritima* [54] and a LpxM from the pathogenic bacterium *A. baumannii* [55]. LpxM from *A. baumannii* was found to produce holo-ACP even in the absence of Lipid A precursors, suggesting that this enzyme has an acyl-protein thioesterase activity as well which is unique from any other acyltransferase known. This could be a characteristic property of the LPLAT superfamily members and hints at a dynamic structural role of the residues in the active site cleft of these enzymes yet unknown in other acyltransferases.

While the basic mechanism of the reaction catalyzed by LpxM from *A. baumannii* was the charge transfer mechanism involving the catalytic dyad HX₄D/E, as mentioned before, the proposed model of its activity was a unique ordered binding and donor substrate concentration dependent “reset” mechanism. Detailed structural studies of these enzymes might be the key to understanding the prominent variability of the acyltransferases in terms of their involvement in different pathways, their variety of core structures and their divergent evolution.

1.11 Acyltransferase of the HtrB/MsbB family and the Raetz pathway

Lipid A biosynthesis in Gram negative bacteria, initiated on the inner leaflet of the internal membrane results in the production of a bi-phosphorylated, tetra acyl chain containing lipid A structure. The lipid A core is composed of a fatty acylated β -glucosamine-(1 \rightarrow 6)-glucosamine-1-phosphate base. Among the final modifications made to lipid A before its transport to the outer membrane, are the secondary acyl chain additions to the 2' and 3' positions of the glucosamine backbone of lipid A, by the acyltransferases belonging to the HtrB/MsbB family. Known members of the HtrB/MsbB family are LpxL and LpxM, which are nothing but orthologs of each other, and mediates the addition of laurate (C12:0 fatty acid) and a myristate (C14:0 fatty acid) respectively [56-58]. There is also a third LpxL ortholog in *E. coli* called LpxP involved in the transfer of the palmitoleate (C16:1 unsaturated fatty acid) to Lipid A depending on the growth conditions

encountered by the bacteria. The final steps of the Raetz pathway are therefore dependent on the HtrB/MsbB family of acyltransferases which results in the formation of fully functional Lipid A.

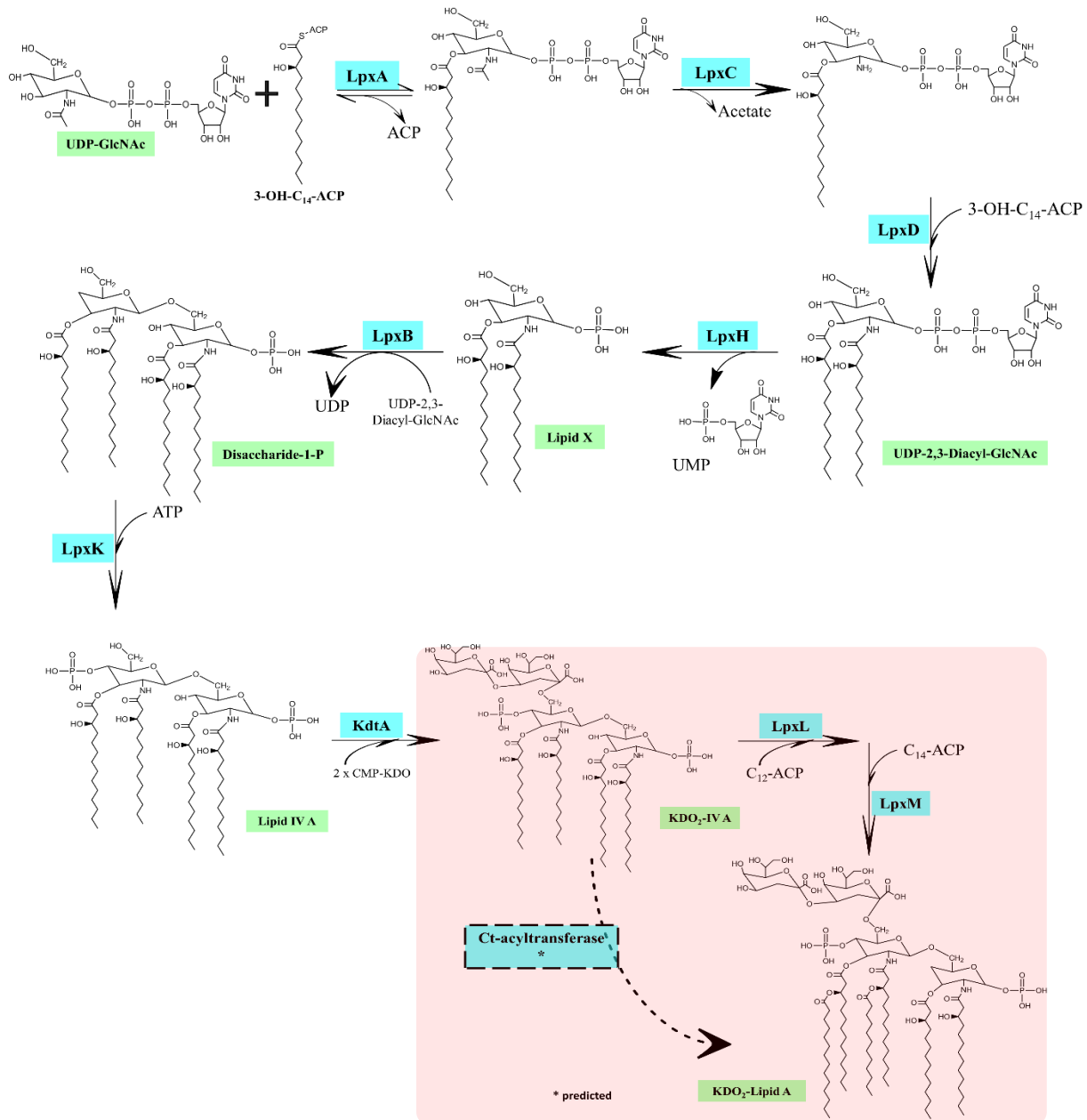


Fig 1.12 Diagrammatic representation of the Lipid A biosynthetic ‘Raetz’ pathway The pathway takes place in the inner leaflet of the internal cell membrane and only the mature Lipid A are transported to the outer envelope. The enzyme in focus, involved in the secondary acylation of the immature Lipid A, is highlighted by a light red backdrop and comprises the final steps of the pathway. These secondary acylations are often very important in the virulence and survival of a Gram negative bacterial strain and is a key component of the entire pathway unlike what the name suggests.

Lipid A biosynthesis in Gram negative bacteria like *E. coli* proceeds via nine enzyme catalyzed steps, commonly referred to as the Raetz pathway (Fig 1.12). All of the enzymes belonging to the pathway are constitutively expressed. The starting substrates for Lipid A biosynthesis are UDP-N-acetylglucosamine (UDP-GlcNAc) and β -hydroxymyristoyl-ACP. UDP-GlcNAc is also a known substrate in peptidoglycan synthesis while β -hydroxymyristoyl-ACP can act as a precursor during phospholipid metabolism. The first three steps of the lipid A pathway occur in the cytoplasm. In the first step of Lipid A biosynthesis, the cytosolic enzyme LpxA adds acyl chains to UDP-GlcNAc with β -hydroxymyristoyl-ACP acting as the donor in a reversible reaction. The product catalyzed by LpxA, UDP-3-O-(β -hydroxymyristoyl)-NAc, is then deacetylated by the cytosolic enzyme LpxC in an essentially irreversible reaction, making this the first committed step to this pathway. Next, the cytosolic enzyme LpxD, catalyzes the incorporation of a second hydroxymyristate moiety onto the deacylated product of LpxC, UDP-3-O-(β -hydroxymyristoyl)-D-glucosamine. Both LpxD reaction products however act as competitive or non-competitive inhibitors against the substrates to inhibit the LpxD reaction. After the first three steps of the pathway, which takes place in the cytosol, the next two steps catalyzed by peripheral membrane proteins. The first of them, LpxH generates lipid X by cleaving the UDP moiety off the LpxD product, UDP-2,3-bis(β -hydroxymyristoyl)-D-glucosamine, leaving a single phosphate group for further reactions. In the next step, the second peripheral membrane protein, LpxB catalyzes the combination of lipid X and UDP-2,3-bis(β -hydroxymyristoyl)-D-glucosamine, to form lipid A disaccharide. The next four steps are catalyzed by integral membrane enzymes and takes place in the membranes. In the first step, LpxK generates lipid IVA by phosphorylating the lipid A disaccharide. In the following step, WaaA (previously known as KdtA) transfers two KDO (3-deoxy- α -D-manno-octulosonic acid) sugars to lipid IVA in a sequential manner to produce KDO₂-lipid IVA in an irreversible reaction. Finally, in the last two steps, the “late acyltransferases” LpxL and LpxM, add lauroyl and myristoyl chains to the KDO₂-lipid IVA to form the final KDO₂-Lipid A which is then joined to core oligosaccharide and flipped to the outer leaflet of the inner membrane from the inner leaflet, by an ABC transporter protein, MsbA. Then, the O-antigen is added to the KDO₂-Lipid A by several enzymes to form LPS, and transported to the outer leaflet of the outer membrane [59].

1.12 Lipid A and the bacterial lipopolysaccharides layer

Lipopolysaccharide (LPS) is located in the outer layer of the outer membrane of Gram-negative bacteria and is exposed on the cell surface in case of strains lacking capsules. Intact bacterial LPS, also known lipoglycans have molecular masses of 10-20 kDa and three structural components (Fig 1.13):

- (i) A hydrophobic lipid moiety, Lipid A forming the innermost domain of the LPS
- (ii) A core polysaccharide chain composed of an inner and outer core joined by a covalent bond
- (iii) A repeating oligosaccharide side chain called the hydrophilic O-antigen which forms the outermost domain of the LPS

Lipid A is the phospholipid component which anchors the LPS layer to the outer leaflet of the outer membrane of Gram negative bacteria. Lipid A composes of two β (1 \rightarrow 6) linked glucosamine units, with fatty acyl chains attached at the 2' and 3' positions of the glucosamine residues, and one phosphate group attached to each of the carbohydrates. The Lipid A moiety has been identified to be absolutely critical for the endotoxin activity shown by the LPS. It is also reportedly critical for immune response activation in the human body.

The inner core of the core polysaccharide typically contains 1-4 KDO molecules attached to the disaccharide core. The inner core also contains ketoheptose monosaccharides and phosphorylated glycan residues. These phosphate-containing groups of the LPS increases the overall negative charge of the cell membrane and helps in stabilization of its structure. It is to be noted that KDO requirement is not absolute for viability of the organism. The outer core contains regular hexose molecules like glucose, galactose, and N-acetylglucosamine but has far higher structural diversity than the inner core [60].

The O-antigen is a repeating glycan polymer typically comprising of three to six sugars. The O-antigen has the highest degree of structural variability among all the components and is different among bacterial species. Different serogroups have been assigned to *Escherichia coli*, *Salmonella enterica*, and *Vibrio cholera* on basis of distinctive O-antigen structures. Based on the presence of the O-chains, the LPS are considered smooth or rough. Smooth LPS contains full length O-chains whereas rough LPS sport reduced or absent O-chains. Hence rough LPS is generally more

hydrophobic and thereby bacteria with rough LPS have higher penetrability to hydrophobic antibiotics in their cell membranes. O antigen being the exposed part of LPS in bacteria, is therefore targeted by host antibodies.

LPS majorly contributes to the structural integrity in Gram negative bacteria and protects the bacterial plasma membrane from various chemical attacks like bile salts and lipophilic antibiotics. LPS are heat stable endotoxins and are known to cause septic shock (septicemia) in humans and can induce strong immune responses in mammalian cells. LPS is also known to influence different facets of bacterial ecology, like surface adhesion and bacteriophage susceptibility. The CD14/TLR4/MD2 receptor complex is known to be involved in LPS recognition and promotes the secretion of proinflammatory cytokines (TNF α and ILs) as a response. However, for NF- κ B activation, the polysaccharide component of LPS along with Lipid A is necessary. Due to its connection to septicemia, LPS has been studied in detail, as possible targets for antibodies and also to find inhibitors to LPS biosynthesis [61].

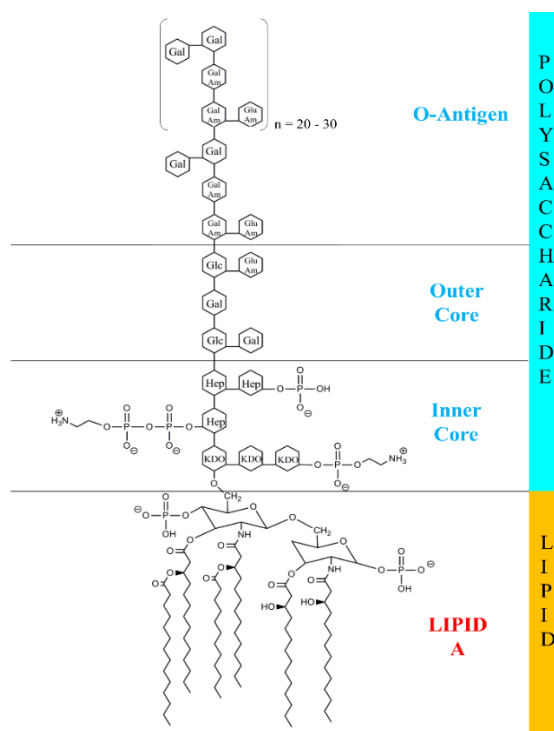


Fig 1.13 Detailed diagram of the components of the bacterial LPS layer The LPS layer composed of a lipid component (Lipid A) and a saccharide component (inner core, outer core and O-antigen) are shown. While the inner core has KDO sugars, the outer core and O-antigen has similar compositions of hexoses and their N-acetylated variants only differing in the type of hexoses (glucose for the outer core and galactose for the O-antigen).

1.13 Structural Insights into the LPLAT acyltransferase Superfamily

Four crystal structures have been solved in the LPLAT acyltransferase superfamily, GPATs from *Cucurbita moschata* (1K30 and 1IUQ), PatA from *Mycobacterium smegmatis* (5F2T), LpxM from *Acinetobacter baumannii* (5KN7 and 5KNK) and AGPAT/LPAAT (PlsC) from *Thermotoga maritima* (5KYM). The Cm-GPATs were resolved at 1.9 Å and 1.55 Å respectively and both were apo structures without any ligand co-crystalized. The Ms-PatA structure was solved at 2.06 Å and interestingly was co-crystalized with a Palmitic acid. While the Ab-LpxM apo structure was solved at 1.99 Å, the catalytic residue mutant apo structure (5KNK; E127A) was solved at 1.9 Å. The core structures of PatA and LpxM have the most prominent similarities among all the four structures. The presence of a n-dodecyl- β -D-maltoside (DDM) molecule in a large hydrophobic pocket of the 5KN7 structure, which incidentally is also present in the PatA structure with a Palmitic acid co-crystalized, confirms that this is the binding pocket for the acyl chain donor/acceptor molecule. The Tm-PlsC structure was solved at 2.8 Å resolution and was crystallized along with 1-Heptadecanoyl-2-tridecanoyl-3-glycerol-phosphonyl choline, nonane and DDM molecules. The structures of the GPAT and PatA also show some conservation around the predicted binding pockets indicating that the core structure involving the important active site residues are somewhat conserved. Further comparisons between these structures may lead to discovery of structural elements that are broadly conserved among mammalian LPLAT members and could be more relevant for certain subfamilies.

GPATs catalyze the transfer of acyl chains from acyl-ACPs (mainly palmitoyl-ACP or octadecenoyl-ACP in plants) to the sn-1 position of glycerol-3-phosphate (G3P). The GPAT structure is composed of two domains, an α helical domain and an α/β domain connected by a short loop. The helical domain has a four helix bundle while the α/β domain core has an open twisted α/β architecture with parallel and antiparallel β strands. Molecular surface calculations show that the structure has three tunnels and three pockets where the tunnels act as connectors between the pockets. Each tunnel had necessary dimensions for containing the acyl chain of palmitoyl-CoA. The predicted G3P binding pocket is rich in positive residues and is connected via a hydrophobic tunnel, can be predicted as the acyl chain donor binding site, to the surface opening of the enzyme. This highlights the underlying mechanism of action of the GPATs where the acyl chain transfer

occurs between acyl-CoAs and G3P at the junction of the pocket and tunnel as mentioned above [52].

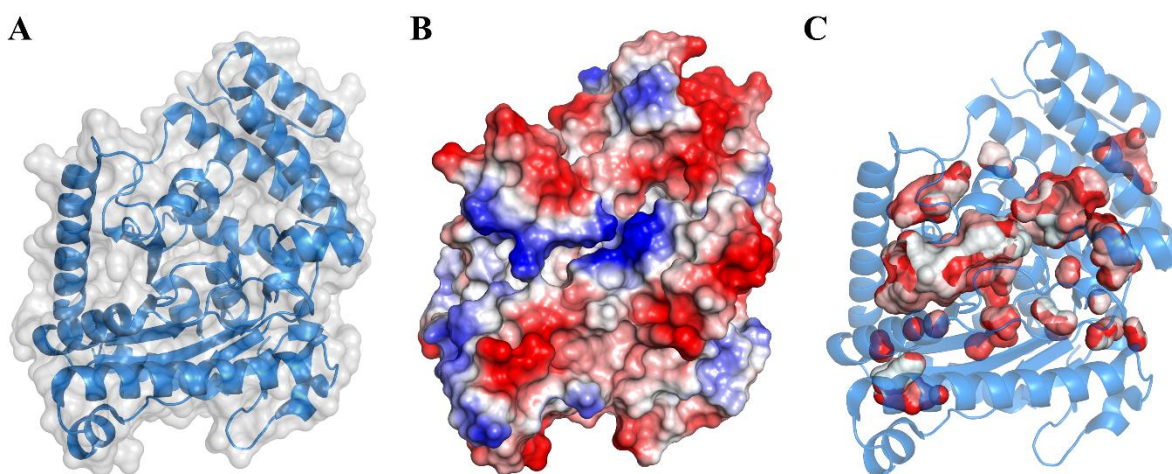


Fig 1.14 Crystal structure of squash GPAT Three different views of the crystal structure (PDB: 1k30) showing secondary structural elements and the binding cleft of the enzyme. (A) Cartoon & surface representation depicting the core domains, (B) Electrostatic surface representation showing polar residue distribution over the protein surface, (C) Culled hydrophobic surface representation of the binding cleft bisecting the enzyme showing the prominence of non-polar residues in the pocket.

An essential step for the biosynthesis of PIMs is the transfer of a palmitoyl moiety to the 2-position of myo-inositol in PIM1/PIM2 at the 6-position of the Manp ring. This step is catalyzed by PatA. The PatA central core has a topology composed of six β strands with four parallel and two antiparallel β sheets, surrounded by α helices. The active site is located in a prominent groove running parallel to the protein surface which also contains an extended, narrow and predominantly hydrophobic tunnel which is perpendicular to the central core. The groove is flanked by four α helices, 2 β strands and connecting loops while the tunnel comprises of the central β sheet, two α helices and connecting loops. PatA has notable hydrophobic patches interspersed with positively charged clusters adjacent to the major groove. The acyl chain of the donor in the hydrophobic tunnel has the carboxylate group facing the groove and the acyl tail extending into the hydrophobic core. The entrance of this hydrophobic tunnel is rich in charged residues consistent with the need for interaction with the buried carboxylate group. The predicted acceptor binding site is the end side of the main groove, close to the hydrophobic tunnel. The underlying mechanism is therefore initiated by a conserved histidine residue (H126) in congregation with a conserved glutamate

(E200) located at the entrance of the tunnel which carries out the nucleophilic attack on palmitoyl-CoA resulting in a covalent bond formation between the palmitate moiety and the mannose ring of PIM1/PIM2 [53].

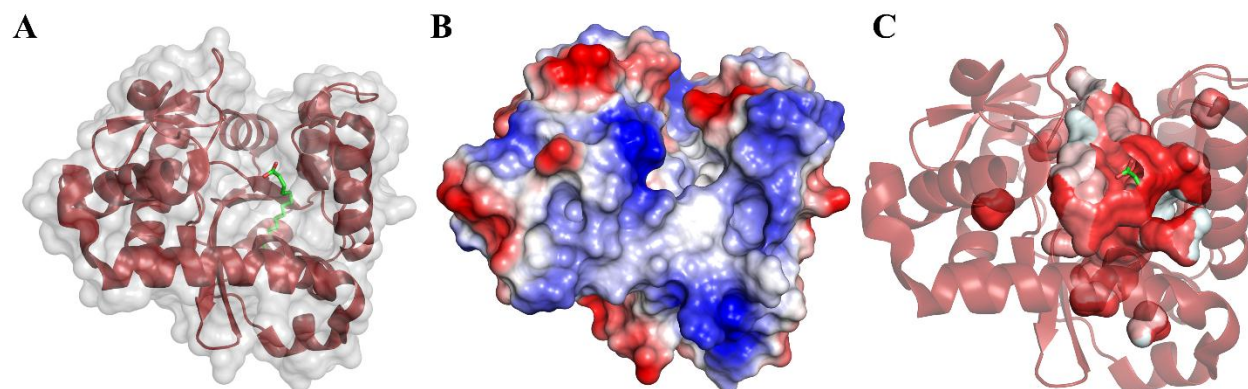


Fig 1.15 Crystal structure of mycobacterial PatA Three different views of the crystal structure (PDB: 5f2t) showing secondary structural elements and the binding cleft of the enzyme. (A) Cartoon & surface representation depicting the core domains, (B) Electrostatic surface representation showing polar residue distribution over the protein surface, (C) Culled hydrophobic surface representation of the binding cleft, at the heart of the enzyme, containing a co-crystallized palmitate, showing predominantly non-polar residues in the pocket.

Ab-LpxM has a large seven-stranded β -sheet composed of parallel and antiparallel β sheets at its core bordered by α helices. It has a very extended hydrophobic pocket, with a bound n-dodecyl- β -D-maltoside (DDM) molecule. The opening of the pocket is considerably solvent exposed and broad, which makes it a probable binding site of the bulky acyl chain acceptor (Lipid A precursor). The predicted binding site of the donor acyl-ACP is the deep hydrophobic channels towards the core of the enzyme, which are continuations of the same hydrophobic cavity. There are more of such hydrophobic channels which may probably act as hydrocarbon “rulers” for determining the acyl chain lengths of acyl-ACP or lipid A precursor. The active site involving the catalytic dyad HX₄D/E is thereby located in this hydrophobic cleft. A conserved arginine residue (R159) located at the starting point of these channels was predicted to be involved in the binding and positioning of the di-glucosamine head group of lipid IVA for the acyl transfer. Ab-LpxM was found to produce holo-ACP in the absence of Lipid A precursor, hinting towards a previously unreported acyl-protein thioesterase activity. It was also proposed to have more than one binding sites for ACP which contributed to a feedback mechanism to control the levels of holo-ACP and Lipid A precursors in the cell. An electropositive patch located in the solvent exposed region of the active

site represents a putative ACP-binding site which could be responsible for the substrate inhibition where a binding of ACP doesn't allow further binding of the Lipid A precursor, hence contributing to the feedback mechanism mentioned above. The overall enzyme mechanism involves a nucleophilic attack initiated by the conserved histidine residue (H122) and a conserved glutamate (E127) located near the opening of the extended hydrophobic channel where the lauryl/myristyl chain transfer takes place at position 3' on the R-3-hydroxylaurate in what is known as a secondary acylation of the KDO₂Lipid A to produce mature Lipid A [55].

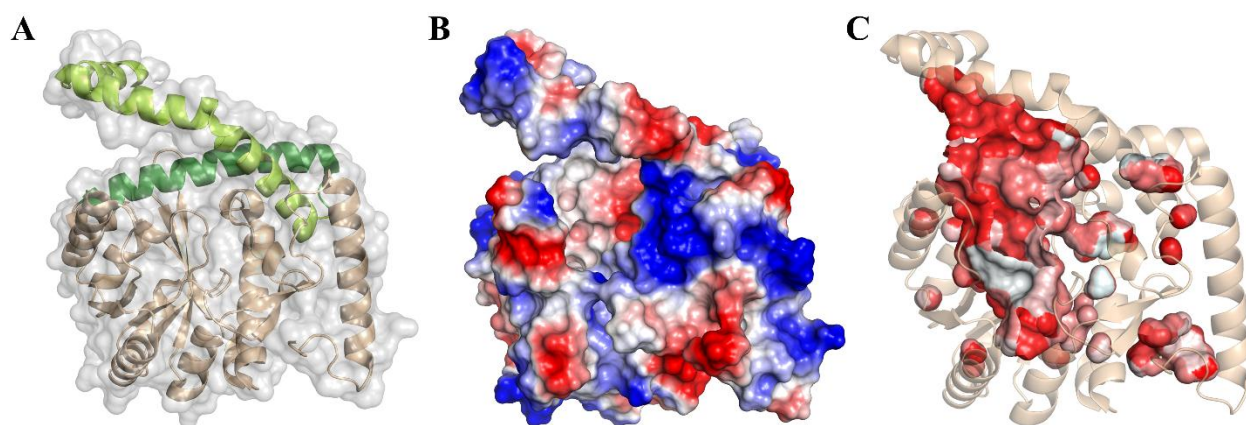


Fig 1.16 Crystal structure of *A. baumannii* LpxM Three different views of the crystal structure (PDB: 5kn7) showing secondary structural elements and the binding cleft of the enzyme. (A) Cartoon & surface representation showing the transmembrane helices in different shades of green, (B) Electrostatic surface representation showing polar residue distribution over the protein surface, (C) Culled hydrophobic surface representation showing the extended binding cleft of the enzyme showing the prominence of non-polar residues in the pocket.

Tm-PlsC comprises of two domains, a unique two-helix motif in the N terminal, and an α/β domain in the C terminal. Like the previous GPAT, PatA and LpxM structures, this α/β domain had seven β strands arranged into a sheet at its core which was surrounded by α and 3_{10} helices and housed the AGPAT conserved motifs including the catalytic dyad HX₄D. Three of these conserved motifs compose the active site region. The other motif, known as the N terminal two-helix motif is the new feature in this crystal structure and is predicted to be responsible for positioning the active site for catalysis. It is not present in the GPAT and PatA structures while in LpxM a structurally different helix motif composing of a kinked helix and a normal helix could be predicted to do a similar function. The first helix of the two-helix motif is mainly hydrophobic while the second helix has exposed positively charged residues and has a close association with the α/β domain.

This two-helix motif is responsible for the membrane association of the Tm-PlsC. The HX4D catalytic motif is positioned in a prominent deep cleft which almost dissects the Tm-PlsC structure. The amino acid residues on one side of the binding cleft had high B-factor indicating their tendency to fluctuate their positions, which could lead to one possible conclusion, the flexible nature of the binding cleft, wherein the cleft opens while the substrate binding, closes during catalysis and again opens to release the product. A hydrophobic tunnel is located as an extension of the binding cleft and is ideally positioned to bind the acyl chain of the donor and interestingly lies at the conjunction of the α/β domain and the two-helix motif making this an interesting structural feature which needs to be studied in more detail. An overview of the enzyme mechanism is a nucleophilic attack on the thioester bond of the acyl CoA/ACP by a proton abstracted from the 2- hydroxyl of LPA by the N ϵ 2 nitrogen of the conserved histidine residue (His84) aided by a conserved aspartate (Asp89) leading to the incorporation of acyl chain species into the 2-position of bacterial phospholipids [54].

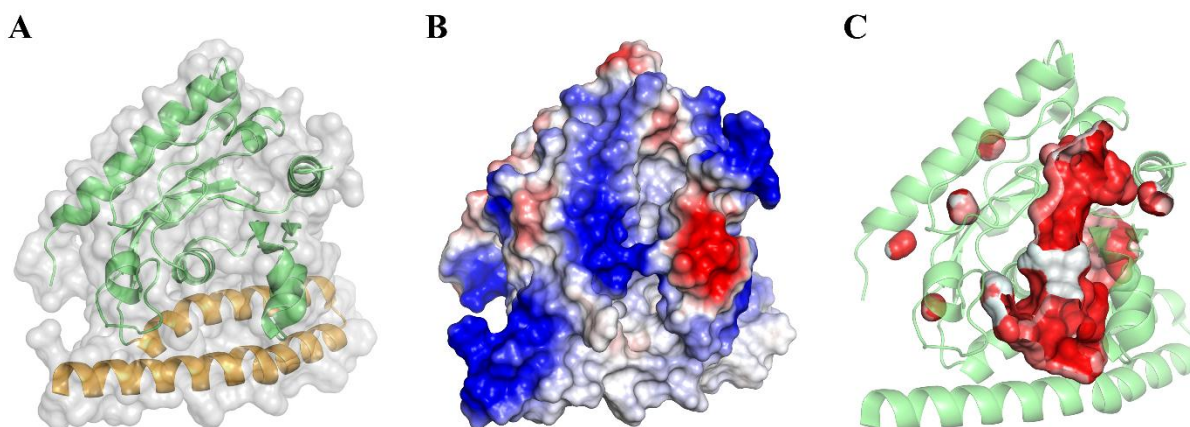


Fig 1.17 Crystal structure of *T. maritima* AGPAT Three different views of the crystal structure (PDB: 5kym) showing secondary structural elements and the binding cleft of the enzyme. (A) Cartoon & surface representation showing the transmembrane helices in yellow orange, (B) Electrostatic surface representation showing polar residue distribution over the protein surface, (C) Culled hydrophobic surface representation of the binding cleft, almost bisecting the enzyme, showing the prominence of non-polar residues in the pocket and a small patch of polar residues.

Overall the α/β domain with parallel and antiparallel β strands surrounded by α helices form the core of these enzymes and are essentially conserved over all the structures. The membrane bound members had a two-helix domain which helped in their membrane association and in the proper active site orientation. The binding clefts were mostly parallel to the enzyme surface and almost

bisected the enzymes through the middle. The core α/β domain contained the conserved motifs including the catalytic HX₄D/E motif and the binding pockets for the acyl chain donor and acceptor molecules which are located almost adjacently to the active site in most cases.

1.14 Role of UbiA prenyltransferases in diseases

There are three notably significant members of the UbiA prenyltransferases in humans, namely UBIAD1, COQ2 and COX10, all of which are known to have major disease implications (Fig 1.18). The eukaryotic homologs of prokaryotic UbiA, COQ2 mediates the second step in the final reaction sequence of coenzyme Q (CoQ) biosynthesis, which is the condensation of the polyisoprenoid side chain with PHB, generating the first membrane-bound Q intermediate. In humans, mutations in COQ2 have been linked to multiple system atrophy, hypercholesterolemia, steroid resistant nephrotic syndrome [62]. The eukaryotic homologs of prokaryotic MenA, UBIAD1 mediates the formation of menaquinone-4 (MK-4) and coenzyme Q10. UBIAD1 is also known to promote endothelial cell survival, thereby sustaining vascular homeostasis [63], prevent oxidative damage in cardiovascular tissues [64], and maintain mitochondrial function [65]. Dysfunctional UBIAD1, however has been linked to cardiovascular degeneration [63, 64], Parkinson's disease [65], Schnyder corneal dystrophy [66, 67], and urologic cancers [68]. COX10 converts protoheme IX and farnesyl diphosphate to heme O. Mutations in this gene have been linked to several cases of cardiomyopathy (Antonicka et al.,2003), mitochondrial cytopathies involving a variety of multisystem disorders targeting the muscle and nervous systems [69] and Cytochrome C oxidase deficiency [70] featuring a generalized weakness of skeletal muscles (myotonia), abnormalities of the heart and kidneys, and/or abnormally high levels of lactic acid in the blood [71].

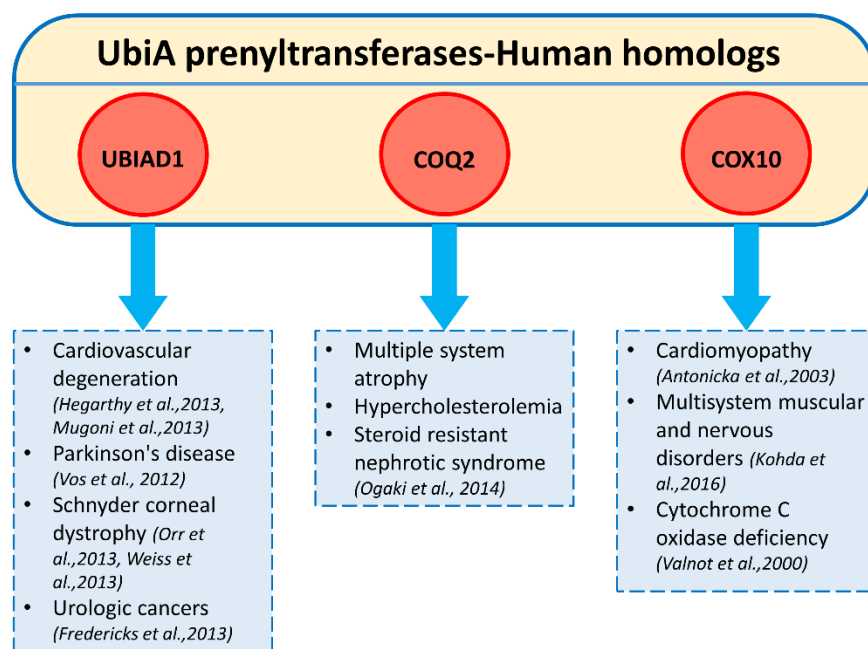


Fig 1.18 UbiA homologs associated with human diseases Human UbiA homologs UBIAD1 COQ2 and COX10 are reported to have prominent roles in various cardiovascular and neurological diseases and even in in some selective cancerous diseases.

1.15 Role of acyltransferases in diseases

A few acyltransferases are widely implicated in human diseases. Monoacylglycerol acyltransferases MGAT1 and MGAT2 have been pathologically connected to obesity related diseases. MGAT1 is considered to be a probable target for reducing hepatic steatosis and obesity [72], while MGAT2 has been identified as an attractive target for treating obesity, Type 2 diabetes mellitus, and nonalcoholic steatohepatitis [73]. Membrane-bound O-acyltransferases (MBOAT4/GOAT) are drug targets for obesity treatment. Several MBOAT members are drug targets for major human diseases, including atherosclerosis, Alzheimer disease, and viral infections [51]. Acyl-coenzyme A: cholesterol acyltransferase 1 (ACAT1) is known to be involved in the Acetoacetyl-CoA thiolase deficiency, and are potential drug targets for Alzheimer's disease [74] and atherosclerosis [75]. Diacylglycerol O-acyltransferase 1 (DGAT1) is implicated in Congenital diarrhoeal disorder, protein-losing enteropathy type. DGAT1 is also a drug target for obesity and type II diabetes [76]. Roles of GPAT/AGPAT in obesity-related diseases were also revealed by identification of the disease causing genes and analysis of genetically manipulated

animals. GPATs and AGPATs are known to be involved in lipodystrophy and hypertrophic cardiomyopathy [77]. Tafazzin is implicated in Barth syndrome (Fig 1.19) [78].

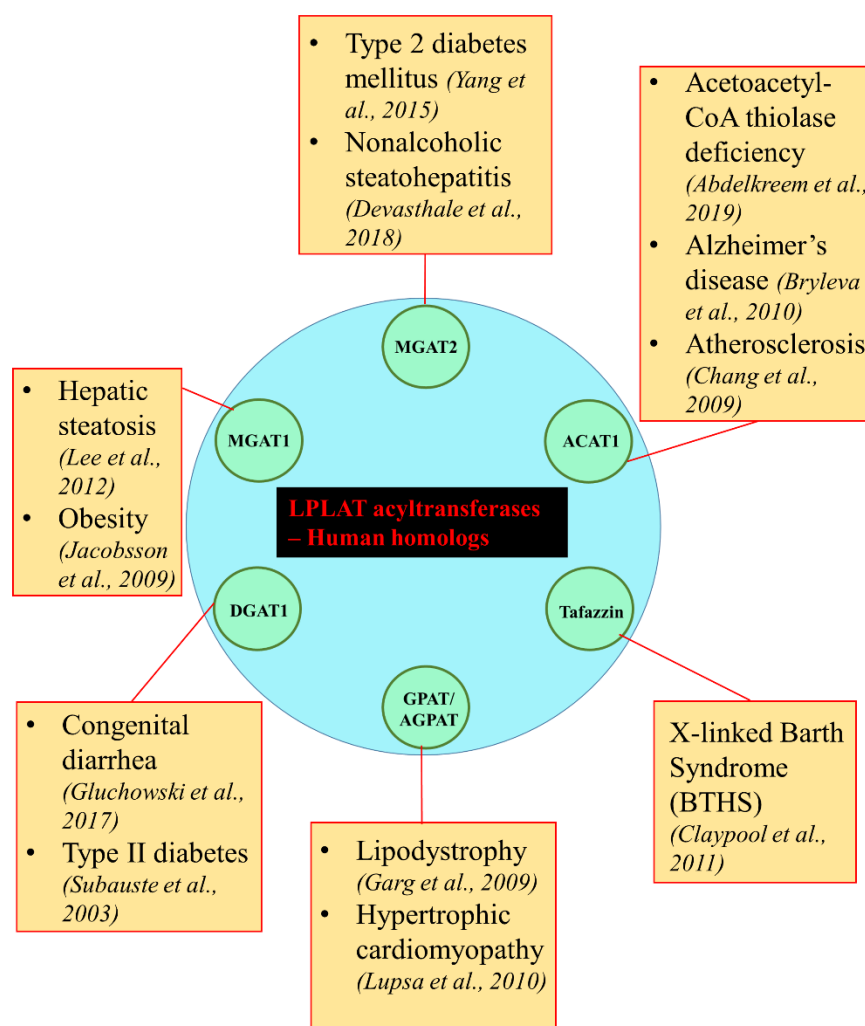


Fig 1.19 Acyltransferases associated with human diseases Human LPLAT acyltransferases are reported to be involved in obesity related diseases, diabetes and neurodegenerative disorders

1.16 Statement of problem

For life to sustain, the unit of life, the cell, functions in harmony among a myriad of interconnected biochemical pathways. For these pathways to co-exist energetically and maintain their appropriate order, they need to be regulated systematically. Be it by preserving and harnessing the genetic code (the Central Dogma of life) stored inside the cell or by maintaining the structural integrity of the cell or by generating the energy reserves for the cell or by providing energy for cell survival or

even by assembling molecules necessary for cell survival, enzymes maintain the entire dynamics of the cell environment. Among these enzymes, transferases and their ten subclasses are the only class of enzymes to be involved in all known aspects of cell survival. The preliminary function of a cell in regards to its survival is to control the nature of its boundaries. The most important component of these boundaries are the differentially permeable cell membranes which vary in fluidity in different organisms, often influenced by their compositions. Other components include the cell wall and an outer membrane in prokaryotes, which also contribute to the integrity of the outer membrane of these organisms. The common structural component of both the ‘inner’ cell membrane and the outer membrane are lipids predominantly containing long carbon chains. The two varieties of these long carbon chains are the ‘acyl chains’ which are saturated and have a COOH group at one end while the other one are the ‘prenyl chains’ which are unsaturated and contain a pyrophosphate moiety at one end. The prenyltransferases are involved in the transfer of the prenyl chains in archaeal inner membrane lipids while the acyltransferases do the same for acyl chains in bacteria apart from contributing to the formation of bacterial storage lipids, making them major players in the lipid dynamics inside the cell.

In this work, we attempted to look at the prenyltransferases and acyltransferases in detail from a structural perspective and tried to understand the structure function relationship in these enzymes. Since these enzymes perform essentially the same function with a small but very notable difference (difference in their substrate specificity with regards to the type of carbon chain they use), we wanted to study their structural differences and fathom how these changes are translated into their functional difference. For this purpose, we have selected a DGGGPS (prenyltransferase) from *Thermococcus kodakarensis* (Tk-DGGGPS) and a HtrB/MsbB family acyltransferase from *Chlorobium tepidium* (Ct-acyltransferase) for the study.

The DGGGPS are unique among the prenyltransferases since they catalyze the transfer of prenyl chains to linear substrates in contrast to others who have aromatic ring containing compounds as their prenyl acceptors. The key to this distinctive nature should lie in its binding pocket, which in this case is a buried central cavity with distinct openings for key steps of the enzyme catalysis. The members of the UbiA superfamily of prenyltransferases are responsible for the catalysis of a very diverse substrate pool since they are involved in many different biochemical pathways over a variety of species and yet they show prominent conservational features. The human homologs in

the superfamily are implicated in many diseases like Parkinson's, Schnyder corneal dystrophy and urologic cancers making structural knowhow about these enzymes vitally important in clinical drug designing. The HtrB/MsbB family acyltransferases are different from the others as they are membrane bound while most acyltransferases are soluble proteins. The prominent feature that is responsible for the above is a two-helix motif which not only tethers the enzyme to the membrane but also properly orients the active site for proper catalysis. The members of the LPLAT superfamily are involved in mainly storage lipid and membrane lipid biosynthesis and are notably quite diverse in their amino acid sequence as well as overall structures. However, they do have some similarities in their core structures which are only obvious when studied in great detail. The human homologs in the superfamily are implicated in many diseases like Alzheimer's, Type 2 diabetes mellitus and atherosclerosis making structural studies important in potential drug target screening.

Our study optimized the expression of in bacterial system for structural studies, which demand an active and relatively large amount of highly pure protein. Initially, heterologous expression of Tk-DGGGPS and Ct-acyltransferase, cloned in pET28a vector were attempted in different strains of *E. coli* such as BL21 (DE3) and C41 (DE3), of which the latter met with greater success. Further biochemical, biophysical and structural analyses were done with certain objectives.

Our aim was to study the core conserved regions and the architecture of the catalytic pocket/active site of the Tk-DGGGPS and Ct-acyltransferase to understand the key to their functional differences and their diversity substrate pool. We also wanted to find out the phylogenetic relationships between the members of individual superfamilies and attempt to figure out the functional significance of the evolutionarily conserved patches and how they are implicated in the disease causing homologs.

Chapter II – Expression and
functional characterization of
Digeranylgeranylglyceryl
phosphate synthase (DGGGPS)

2.1 Introduction

(S)-2,3-Di-O-geranylgeranyl glyceryl phosphate synthase (DGGGPS) is an intrinsic membrane protein reported to be found in most classes of archaea. It is known to be primarily involved in archaeal ether lipid formation but is also involved in the transfer of prenyl groups to hydrophobic ring structures like quinones and hemes [40]. DGGGPS is a member of a large family of proteins, known as the UbiA prenyltransferase superfamily (Pfam: PF01040) which belongs to the clan known as the Terpenoid synthases superfamily (Pfam: CL0613) [34-36]. This family is known to consist of prenyltransferases that are involved in the biosynthesis of diverse compounds like respiratory quinones, hemes, chlorophylls, vitamin E, and shikonin by catalyzing the fusion of a phytyl or isoprenyl chain to an aromatic ring. The prenyl acceptor in DGGGPS synthase is a linear compound, which is a notable exception to the aromatic ring structures used by other prenyltransferases in the superfamily. The atypical substrate specificity of DGGGPS may be the key towards unravelling the dynamics behind the biogenesis of the unique archaeal membrane lipids.

DGGGPS (EC 2.5.1.42) catalyzes the second pathway-specific step in the biosynthesis of the core membrane di-ether lipids in archaea. It is reported to catalyze the transfer of a geranylgeranyl group (a 20-carbon lipophilic isoprene unit) from geranylgeranyl diphosphate (GGPP) to (S)-3-O-geranylgeranyl glyceryl phosphate (GGGP), to form digeranylgeranyl glyceryl phosphate (DGGGP) [40]. Divalent metal cations like Mg^{2+} and Mn^{2+} seem to be a requirement for optimal activity of this enzyme. DGGGPS basically constructs the diether lipid skeleton in an unsaturated precursor form, from which diether lipids (archaeol; C20) and tetraether lipids (caldarchaeol; C40) in archaea are generated. This membrane bound enzyme has been reported to be quite rigid and highly substrate specific and substrates like 4-Hydroxybenzoate, 1,4-dihydroxy 2-naphthoate, homogentisate and alpha-glycerophosphate which are well known prenyl acceptors in other prenyltransferases of the enzyme family, cannot act as prenyl-acceptor substrates for the enzyme. In the only previous study, a gene encoding a homologue of ubiquinone-biosynthetic prenyltransferase was identified from the whole genome sequence of *Sulfolobus solfataricus* as the candidate for the DGGGPS gene. The gene was then cloned and expressed in *Escherichia coli*, and shown to specifically catalyze the formation of DGGGP from GGGP and GGPP. The catalytic

mechanism of DGGGPS is not well elucidated, however Asp rich domains are predicted to be an integral part of the active site which is involved in the ligand binding and catalysis [40].

Only very limited number of studies have been carried out and reported on this enzyme, since it is not only difficult to purify it from the membrane fraction, but equally challenging to keep it stable and biologically active form after purification. Most membrane-bound proteins in general are α -helix rich structures, which are quite dynamic and flexible in their cellular environment and prominently hydrophobic in nature. Thereby, in order to extract it from the membrane fraction, we have to mimic it with a similar micelle environment as the membranes, wherein comes the need of non-ionic detergents which are mild enough to extract the protein in an active form by maintaining the native protein conformation. Other ionic detergents are harsh enough to destabilize and permanently denature α -helix rich intrinsic membrane proteins.

In the present chapter, we have reported for the first time, a comprehensive account of the expression and purification of biologically active DGGGPS from *Thermococcus kodakarensis* (Tk-DGGGPS). We have also presented primary data regarding the interaction of this membrane protein with archaeal lipids in the form of archaeosomes which are nothing but liposomes generated using archaeal lipid extracts and also studied the role of Tk-DGGGPS in membrane stabilization. We have also investigated the thermodynamic parameters of the enzyme substrate reaction involving Tk-DGGGPS which gives an insight into the spontaneity of the reaction. We already know that during higher stress condition, the composition of the archaeal membrane lipids change into higher carbon chain containing lipids [79]. The involvement of DGGGPS in stabilization of membrane thereby indicates a far more prominent role of this enzyme in the archaeal stress response mechanism.

2.2 Materials and Methods

2.2.1 Materials

LB media, 2xYT and LB agar used for all the bacterial culture were purchased from Hi-media. Kanamycin used for selection of transformants was obtained from Sigma-Aldrich, USA. Chemicals used for the purification of proteins such as Trizma, Sodium chloride, Imidazole, β -mercaptoethanol, glycerol, DTT, Nickel sulphate, Bromophenol-blue (BPB), Acrylamide, N,N'-

methylene bisacrylamide, Sodium dodecyl sulfate (SDS), Acetic acid, Methanol, TEMED (N,N,N',N'- Tetramethylethylenediamine), Ammonium persulfate (APS), were purchased from SigmaAldrich, USA. Molecular weight marker for SDS-PAGE was purchased from Bio-Rad Laboratories, USA. Ni-NTA used for affinity purification were purchased from Qiagen, Germany. Size exclusion columns were obtained from GE, USA. Protein samples were concentrated using Amicon® ultra centrifugal filters procured from Merck-Millipore, USA. All detergents were procured from Anatrace USA.

2.2.2 Cloning of Tk-DGGGPS and Tk-DGGGPS Bril variant

The TK1957 gene was amplified by means of a Polymerase Chain Reaction by using forward and reverse primers specific to the 5- and 3-ends, respectively. The genome of *Thermococcus kodakarensis*, as a template, and AccuPrime™ Pfx DNA Polymerase were used in the reaction. The PCR conditions were a single cycle of pre-denaturation at 95°C (5 minutes) followed by 28 cycles of denaturation at 95°C (30 seconds), primer annealing at 58°C (30 seconds) and extension at 68°C (1 minute), a final extension at 68°C (10 minutes) and then cooled to 4°C. The PCR amplified products were analyzed on 1% agarose gel. The restriction sites in the primers, the NdeI and SalI sites are highlighted in bold, and ligation was carried out with the NdeI–SalI-digested pET-28a vector. The same gene was also cloned into a modified pET-28a plasmid, with a fusion of apocytochrome b562 (BRIL). Detailed information about the primers used are tabulated in Table 2.1 and the composition of the PCR mix is mentioned in Table 2.2. The ligated vectors were then transformed in competent DH5α cells and the colonies obtained were screened against any possible false negatives by colony PCR and restriction digestion simultaneously. Once successfully cloned in the desired pET28a and pET28a-BRIL vectors, the clones were stored at –80 °C. Also the clones were examined by DNA sequencing to check for proper gene sequence insertion and absence of point mutations.

Table 2.1 Details of the primers used in the study

Primers	Primer sequence
Tk-DGGGPS	
Forward primer (Tk-DGGGPS-NdeFor)	5' ACT AGT CAT ATG ATG GAA TTC AAG GCA TTC ATC 3'
Reverse primer (Tk-DGGGPS-SalRev)	5' AGC ACT GTC GAC TCA CCT CAC GAG CGA AGC TAT 3'
Tk-DGGGPS-BRIL	
Forward primer (Tk-DGGGPS-BRIL-BamFor)	5' ACT AGT GGA TCC ATG GAA TTC AAG GCA TTC ATC 3'
Reverse primer (Tk-DGGGPS-BRIL-SalRev)	5' AGC ACT GTC GAC TCA CCT CAC GAG CGA AGC TAT 3'

Table 2.2 PCR reaction mixture for amplification of the TK1957 gene

Serial No.	PCR components	PCR mix
1	Accuprime <i>Pfx</i> Supermix (1x)	45 μ l
2	Forward primer	1 μ l
3	Reverse primer	1 μ l
4	Template	0.5 μ l
5	ddH ₂ O	2.5 μ l

2.2.3 Expression and SEC of Tk-DGGGPS

To carry out heterologous over-expression of Tk-DGGGPS, the expression host *E. coli* C41 (DE3) was transformed with Tk-DGGGPS/pET28a and Tk-DGGGPS/pET28a-BRIL constructs by calcium chloride treatment. Selection of transformed colonies was performed on LB agar plate containing Kanamycin (50 μ g/ml). 10ml of the overnight culture was used to inoculate 1000 ml of fresh 2xYT broth supplemented with kanamycin (50 μ g/ml) at 37°C. When the absorbance at 600nm reached around ~ 0.6, the culture was induced with 1 mM IPTG and grown for 12 hours at the same temperature. Bulk batches of 10/20 liters were harvested by centrifugation at 4500 rpm for 10 minutes at 4°C. The cell pellets obtained after centrifugation were subjected to complete lysis by homogenization in Lysis Buffer A1, pH 8.0 containing 1mM PMSF and PIC. The cell lysate was spun in an ultracentrifuge at 125,000 \times g for 50 min at 4 °C. The pellet obtained was then manually homogenized in Solubilizing Buffer A2, pH 8.0 and excess detergent DDM was added to the 1% (w/v) solution and solubilized in a mixer, overnight. The overnight homogenate kept for detergent extraction of the membrane protein was then subjected to a second round of ultracentrifugation at 150,000 \times g for 40 min at 4°C and the supernatant was collected. The supernatant fraction was then purified by Ni NTA affinity chromatography using a pre-equilibrated Ni NTA bead column. The column was initially washed with Wash Buffer A3, pH 8.0 followed

by Elution Buffer A4, pH 8.0 to obtain the 6XHis-tagged Tk-DGGGPS. An SDS-PAGE was carried out to confirm the monomeric molecular mass of the eluted protein. Size exclusion chromatography was performed thereafter in ÄKTA Pure (GE) using the pooled elution fractions in the Superdex™ S200 10/300 GL Increase column (GE), using Purifying Buffer A5, pH 8.0 to further purify the protein to homogeneity and the peaks obtained were analyzed in SDS-PAGE.

Lysis Buffer A1 (pH 8.0)

- 50 mM Tris
- 300 mM NaCl
- 5% (v/v) glycerol

Solubilizing Buffer A2 (pH 8.0)

- 50 mM Tris
- 100 mM NaCl
- 5% (v/v) glycerol

Wash Buffer A3 (pH 8.0)

- 50 mM Tris
- 100 mM NaCl
- 30 mM Imidazole
- 1 mM DTT
- 0.03% (w/v) DDM
- 5% (v/v) glycerol

Elution Buffer A4 (pH 8.0)

- 50 mM Tris
- 100 mM NaCl
- 300 mM Imidazole
- 1 mM DTT
- 0.03% (w/v) DDM
- 5% (v/v) glycerol

Purifying Buffer A5 (pH 8.0)

- 50 mM Tris
- 100 mM NaCl
- 1 mM DTT
- 0.03% (w/v) DDM
- 5% (v/v) glycerol

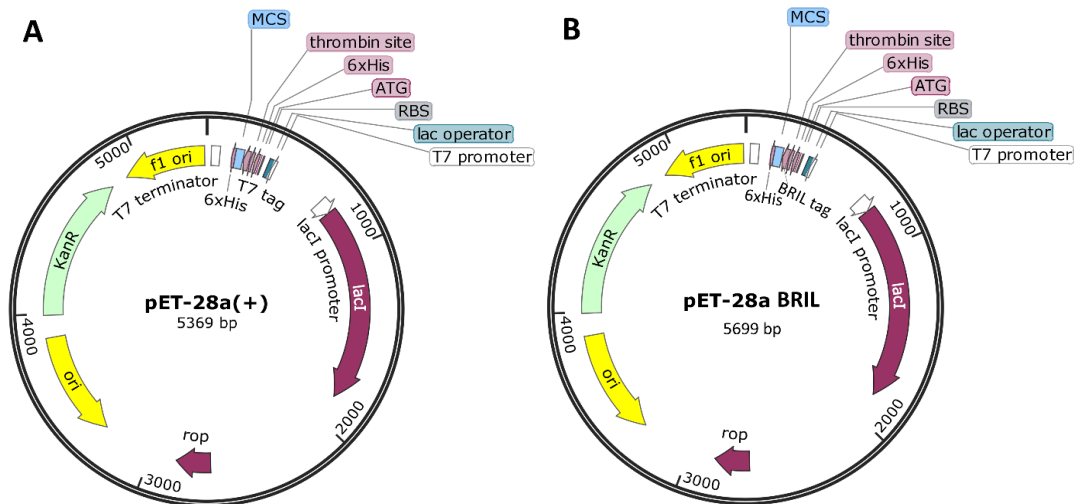


Fig 2.1 Map of vectors (A) pET 28a, (B) pET28a-BRIL

2.2.4 Western Blot analysis

Western blot was performed to confirm the expression of the protein by electro-blotting

on to a PVDF membrane. PVDF membrane was blocked with 5% bovine skim milk in 1X PBS for two hours. The PVDF membrane was then incubated with primary anti-penta-his (Qiagen) mouse antibody at 1:3000 dilution for two hours. To remove any unbound antibody, three PBST washes were performed and the membrane was further incubated with secondary anti-mouse IgG (Fc-region specific) HRP conjugated antibody at 1:2000 dilutions for one hour followed by three PBST washes. For visualizing the bands, the membrane was incubated with substrate (DAB).

2.2.4 Ultracentrifugation Dispersity Sedimentation Assay

An ultracentrifugation dispersity sedimentation (UDS) assay was also carried out to check if the membrane protein was stable in the PDC (protein-detergent complex). The UDS assay is principled on the hypothesis that once protein aggregates form, they become orders of magnitude heavier than the normal dispersed state of the protein particles and hence sedimentation at high 'g' forces can be used to remove them. All ultracentrifugation steps were performed on a Beckman Coulter Optima™ XPN 100 K centrifuge at 4 °C and detergent (DDM) was used at 3X their CMC, 0.03% (w/v). In the first step of the UDS assay, initial step of ultracentrifugation (100,000×g; 50 min) was carried out using a 45-Ti (Beckman Coulter) rotor. After the ultracentrifugation step was carried out, a 1-mL control sample of the protein is removed from the cell pellet re-suspension (using TSG 100) after overnight incubation with DDM, mixed with SDS-PAGE loading buffer, and stored. In the second step of UDS assay, a second round of ultracentrifugation (100,000×g; 40 min) was carried out using a 50-Ti (Beckman Coulter) rotor, to ensure the removal of any aggregates that might have accumulated during the purification. A sample was collected after storing the supernatant overnight and analyzed in SDS-PAGE using a standard dual color Bio-Rad protein marker.

2.2.5 Blue Native PAGE

BN PAGE was performed essentially as described previously [80] with a few modifications. Non-gradient, uniform 8% acrylamide gels were used for the analysis with 3.5% stacking. The protein samples were obtained from the SEC purified protein fractions and Ponceau S stain was used for visualization during the run. Protein samples were first centrifuged at 15000 rpm for 15 minutes to remove possible aggregates. They were further mixed with 5% glycerol and 0.01% Ponceau S

from the Ponceau S/glycerol stock solution to prepare the samples. The buffer of the protein samples was ensured to be of low ionic strength, lacking imidazole altogether and containing 25mM Tris and 50mM NaCl. Two comparative molecular weight markers were used for reference and these samples were run with Coomassie Blue G250 dye. One was a mixture of native proteins from the Amersham™ HMW Calibration Kit for Native Electrophoresis and another was a mixture of monomeric and dimeric BSA, both dissolved in Phosphate buffer, pH 7.0. The electrophoresis was performed at 6°C with the power supply set at 100V at a constant current of 11 mA for 6 hours. After the protein samples entered the resolving gel, cathode buffer B was replaced by cathode buffer B/10 to facilitate better visualization and quicker running of samples. The BN gel was then fixed in Thermo stain for permanent staining of the protein bands.

Cathode Buffer B (pH 7.0)

- 50 mM Tricine
- 7.5 mM Imidazole
- 0.02% Coomassie Blue G-250

Cathode Buffer B/10 (pH 7.0)

- 50 mM Tricine
- 7.5 mM Imidazole
- 0.002% Coomassie Blue G-250

Anode Buffer (pH 7.0)

- 25 mM Imidazole

2.2.6 SEC-MALLS

The SEC-MALLS/RI/UV system comprised of a Superose™ 6 Increase 5/150 GL GE column for SEC, a MALLS detector (DAWN HELEOS, $k = 658$ nm, Wyatt Technologies, USA), and an RI detector (Optilab T-rEX, $k = 658$ nm, Wyatt Technologies, USA). Data acquisition and processing were performed using ASTRA 6.1.7.17 software (Wyatt Technologies) attached to the MALLS system (Wyatt Technologies, USA). The SEC conditions used were the following, sample concentration of 0.1 mg/ml, injection volume of 100 μ L and the flow rate maintained was 0.3 ml/min. The column, PDA, RI detector cells and the MALLS detector were maintained at ambient temperature. The samples were centrifuged at 15000 rpm for 15 minutes to remove any possible aggregates before injection. First a triplicate of standard runs of BSA in the SEC buffer (2mg/ml) were carried out for system calibration and normalization to determine the experimental dn/dc value. M_w and M_n value of the sample protein were calculated from the MALLS data using $dn/dc = 0.185$ ml/g, derived from the BSA runs. The slope value was obtained from double-logarithmic plots of the molar mass versus root-mean-square (rms) radius or conformation plots. SEC-elution profiles were detected by MALLS, IR, and UV at 280 nm to determine the SEC-elution profiles and molar mass/rms relationships of Tk-DGGGPS samples.

2.2.7 Cloning of Tk-GGGPS

The TK1026 gene was amplified by means of a Polymerase Chain Reaction by using primers specific to the 5- and 3-ends: Tk-GGGPS-NdeFor (5' ACT GCG CAT ATG ATG CTC AAG CTT GGA AAA GTC 3') and Tk-GGGPS-SalRev (5' AGC GCG GTC GAC CTA TCC CTT AAC GCC CCT GTT 3'), respectively, the *Thermococcus kodakarensis* genome, as a template, and AccuPrime™ Pfx DNA Polymerase in the PCR reaction. The amplified gene product was then digested by restriction enzyme digestion and inserted into digested pET28a vectors, which were then ligated. The ligated vectors were then transformed in competent DH5α cells and the colonies obtained were screened against any possible false negatives by colony PCR and restriction digestion simultaneously. The clones were examined by DNA sequencing to check for proper gene sequence insertion and absence of point mutations and stored at – 80°C.

2.2.8 Expression of Tk-GGGPS

E. coli BL21 Star™ (DE3) expression cells were transformed with Tk-GGGPS/pET28a construct by calcium chloride treatment. Selection of transformed colonies was performed on LB agar plate containing Kanamycin (50µg/ml). 10ml of the overnight culture was used to inoculate 1000 ml of fresh LB broth supplemented with kanamycin (50µg/ml) at 37°C. When the absorbance at 600nm reached around ~ 1.0, the culture was induced with 0.1 mM IPTG and grown overnight at the same temperature. Bulk batches of 2-4 liters were harvested by centrifugation at 4500 rpm for 10 minutes at 4°C. The cell pellets obtained were re-suspended in Lysis Buffer A1, pH 8.0 containing 1mM PMSF and sonicated twice for 5 minutes each, with 5 second on/off pulse and 45% amplitude. After the cell disruption, the lysate was centrifuged for 20 minutes at 14000 rpm. The supernatant was collected and subjected to Ni-NTA affinity chromatography. Wash Buffer A6, pH 8.0 was used to initially remove the non-specific binding to the Ni beads, followed by an Elution Buffer A7, pH 8.0 to elute the 6XHis-tagged Tk-GGGPS. An SDS-PAGE was carried out to confirm the monomeric molecular mass of the eluted protein. Size exclusion chromatography was performed thereafter in ÄKTA Pure (GE) using the pooled elution fractions in the Superdex™ S200 10/300 GL Increase column (GE) column, using Purifying Buffer A8, pH 8.0 to further purify the protein to homogeneity and the peaks obtained were analyzed in SDS-PAGE. A western blot analysis was

further carried out using His tag-specific primary antibodies (from Qiagen) and HRP-conjugated anti-mouse secondary antibodies (from Sigma) to confirm the purified protein.

Wash Buffer A6 (pH 8.0)

- 50 mM Tris
- 100 mM NaCl
- 30 mM Imidazole
- 1 mM DTT
- 5% (v/v) glycerol

Elution Buffer A7 (pH 8.0)

- 50 mM Tris
- 100 mM NaCl
- 300 mM Imidazole
- 1 mM DTT
- 5% (v/v) glycerol

Purifying Buffer A8 (pH 8.0)

- 50 mM Tris
- 100 mM NaCl
- 1 mM DTT
- 5% (v/v) glycerol

2.2.9 Enzyme assay of Tk-GGGPS and Tk-DGGGPS

Enzyme assays performed here were two pronged. First the Tk-GGGPS was used to generate a substrate for the Tk-DGGGPS and then the Tk-DGGGPS assay was carried out in the same reaction mixture. For the first reaction the enzyme activity of Tk-GGGPS was verified. The reaction mixture contained, in a final volume of 500 μ l, 50 μ mol of GGPP, 350 μ mol of Glycerol-1-phosphate, 1500 μ mol of $MgCl_2$, 7500 μ mol of sodium phosphate buffer, pH 6.0 and 1 μ g of freshly purified Tk-GGGPS protein. This mixture was incubated at 55 °C for 1 hour, and the reaction was stopped by adding 200 μ l of cold, saturated NaCl solution. The mixture was shaken with 600 μ l of 1-butanol saturated with H_2O for 30 minutes. Then the upper layer was removed and concentrated for further analysis. A blank reaction was also performed wherein the Tk-GGGPS wasn't added to the reaction mix.

For the second reaction, a final volume of 500 μ l had the following compositions, 100 μ mol of GGPP, 700 μ mol of G-1-P, 3000 μ mol of $MgCl_2$ and 10000 μ mol of sodium phosphate buffer, pH 5.8. 2 μ g of freshly purified Tk-GGGPS protein was added and the reaction was incubated at 55°C for 1 hour. Then 5 μ g of freshly purified Tk-DGGGPS protein and 100 μ mol of GGPP was added and the reaction was incubated at 55°C for another hour. The reaction was stopped by adding 200 μ l of cold, saturated NaCl solution and the mixture was shaken with 600 μ l of 1-butanol saturated with H_2O for 30 minutes following which the upper layer was removed and concentrated for further analysis. A blank reaction was performed wherein the Tk-DGGGPS wasn't added to the reaction mix.

The reaction samples were analyzed by TLC, reverse phase HPLC and MS. Concentrated samples were loaded on a TLC Silica gel 60 F₂₅₄ aluminium plate. A TLC chamber saturated with the

solvent system 65:25:4 (v/v/v) :: Chloroform:Methanol:Water was used for the TLC runs. The TLC plate was developed in a saturated iodine chamber and the spots obtained were visualized. Concentrated samples dried using a SpeedVac chamber and re-solubilized in HPLC grade methanol were also loaded onto C18 reverse phase columns (Guard column: Symmetry C18 Sentry Guard Cartridge, 300Å, 3.5 µm, 2.1 mm X 10 mm; Main column: Waters Symmetry300™ C₁₈, 300Å, 5 µm, 4.6 mm X 250 mm) connected to a high performance liquid chromatography (HPLC) system. They were eluted with Eluant A (25 mM NH₄HCO₃) isocratically for the first 5 min and then with a linear gradient from 100% Eluant A to 100% Eluant B (acetonitrile) through 30 min, and finally with Eluant B for 7 min, at a flow rate of 0.5 ml/min, under the method described previously (Zhang and Poulter) with minor modifications. Elution of the compounds was detected by UV absorption at 210 nm. Same samples were also loaded onto a Agilent 6550 UHD Accurate Mass QTOF MS in negative ion mode and the prominent m/z peaks were recorded.

2.2.10 Isothermal Calorimetry

Micro calorimetric measurements of ligand binding to Tk-DGGGPS was performed on the MicroCal PEAQ-ITC (Malvern). The parameters of the interaction between prenyl chain donor GGPP and Tk-DGGGPS was recorded at the reaction temperature of 55°C. Tk-DGGGPS (25 µM) was titrated with GGPP (750 µM) in Purifying Buffer A5 + 10mM MgCl₂ buffer, pH 8.0 with a stirring speed of 500 r.p.m. The titration experiment comprised of a total of 25 injections of 1.5 µl for 3 seconds at an interval of 180 seconds between each injection. An initial injection of 0.4 µl was given for stabilization of reaction cell. The data was analyzed and interpreted using PEAQ-ITC software using the sequential binding sites model. The final baseline corrected data and heat plot were plotted using Sigma plot 10.2. The values of ΔG , ΔH and $T\Delta S$ were used to calculate the binding constant for the interaction.

2.2.11 Circular Dichroism Spectroscopy

The far-UV Circular Dichroism spectra was recorded on a Jasco J-815-150S spectropolarimeter (Jasco, Tokyo, Japan) with a PTC 343 Peltier unit in a quartz cuvette of path length 0.1cm, after incubation at temperatures ranging from 25-95°C. The spectra was collected using a data pitch of

1 nm, data integration time of 4 s, bandwidth of 2 nm and a scan speed of 20 nm/min. Far-UV CD spectra was recorded as a mean of three scans for ellipticity in millidegrees. The purified enzyme of concentration 2 μ M was used for the experiment. All spectra were corrected for buffer contributions and observed ellipticity values were converted to mean residue ellipticity (MRE) in deg cm² dmol⁻¹ defined as the following equation.

$$MRE = \frac{M}{r} \times \frac{\theta}{10c} \times d$$

Where M is the molecular weight of the protein, θ is ellipticity in millidegree, d is the path length in cm, c is the protein concentration in mg/ml and r is the number of peptide bonds in the protein (number of residues). The relative content of secondary structure elements was calculated by using CD pro software [81]. Low NRMSD values were observed for analysis with CONTINLL and SELCON.

To analyze the thermally induced unfolding transition of the protein between two states, folded and unfolded, we used the a two-state N \leftrightarrow U unfolding mechanism. The denaturation data were converted into plots of fraction unfolded (f_u) defined by the following equation and T_m of the protein was calculated from the f_u plot.

$$f_u = \frac{\exp \left[\frac{\Delta H_g \left(\frac{T}{T_g} - 1 \right) + \Delta C_p [T_g - T + T \ln (T/T_g)]}{RT} \right]}{1 + \exp \left[\frac{\Delta H_g \left(\frac{T}{T_g} - 1 \right) + \Delta C_p [T_g - T + T \ln (T/T_g)]}{RT} \right]}$$

where T_g referred to the midpoint of the thermal transition where ΔG is equal to zero, ΔH_g referred to the change in enthalpy at T_g and ΔC_p meant change in heat capacity in the unfolding reaction in presence of a denaturant (in this case thermal denaturation).

2.2.12 Fluorescence Spectroscopy

Tyrosine and Tryptophan fluorescence spectra of Tk-DGGGPS (2.5 μM) in Purifying Buffer A5, pH 8.0 were recorded from 300 to 410 nm with preheated protein samples ranging from 30°C-80°C at intervals of 5°C using a fluorimeter (Hitachi F-700, Hitachi High-Technologies Corporation, Japan). The excitation wavelength was set to 295 nm. Data were collected at a 0.5 nm wavelength resolution. The effect of hydrophobic environment on the intrinsic fluorescence was evaluated following similar measurements after incubation with archaeal total lipids and the ligand GGPP at three temperatures 50°C, 55°C and 60°C. All the data points are the mean of triplicate experiments.

2.2.13 ANS binding assay

Fluorescence of ANS was measured using a Hitachi F-700 fluorimeter (Hitachi High-Technologies Corporation, Japan). Fluorescence spectra of ANS bound to Tk-DGGGPS were taken in two separate experiments. In one a standard concentration of Tk-DGGGPS (5 μM) was treated with increasing concentration of ANS (from 0-200 μM) and the changes in ANS fluorescence intensity were recorded. Finally, the number of ANS binding sites and the dissociation constant were calculated according to the following formulae.

$$\Delta F(L) = n\Delta F_{max} \frac{[L]}{[L] + K_D}$$

In the other one, a urea melt was carried out with different concentrations of urea in the range of 0-8M in presence of ANS (10 μM) and Tk-DGGGPS (5 μM). An excitation wavelength of 380 nm was used, and fluorescence emission spectra were recorded in the range of 400–600 nm for both the experiments with a data pitch of 0.5 nm. Both excitation and emission slits were set at 5 nm. The scan rate used for this assay was 240 nm/min. The reported spectra were the mean of three scans.

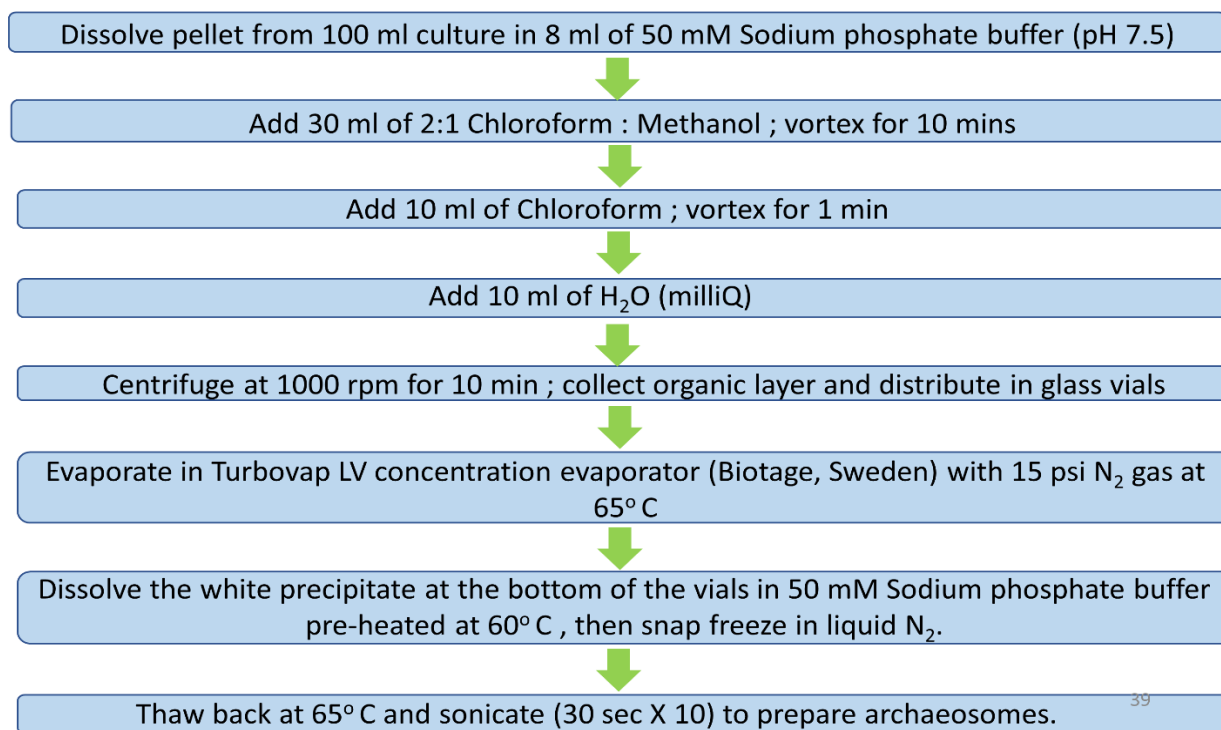


Fig 2.2 Flow diagram for the steps involved in lipid extraction from archaea

2.2.14 Archaeal total lipid isolation and generating archaeosomes from archaeal lipids

Archaeal lipids were isolated from two different species, a thermoacidophilic strain *Sulfolobus acidocaldarius* and a haloarchaeal strain, *Haloferax volcanii*. *S. acidocaldarius* cells were grown aerobically in Brock medium at a pH of 3 and 76°C [34], while the *H. volcanii* strain H53 cells were grown in 18% MGM medium supplemented with uracil and tryptophan at 37 °C for 48 h [38]. The archaeal membranes were isolated following the Bligh-Dyer method with minor modifications [39] (Fig 2.2). These archaeal membranes were used to prepare archaeosomes to test the interaction of Tk-DGGGPS with membrane lipids. Archaeosomes were prepared from isolated membranes by sonicating the solution in a Soniprep 150 sonicator for 5 min (10 cycles × 30 s). Stewart assay [40] was used to quantify lipid concentration which was found to be 0.3 mg/ml for *S. acidocaldarius* (Saci) lipids and 0.32 mg/ml for *H. volcanii* (Hvo) lipids respectively.

2.2.15 TEM Sample Preparation

The archaeosome samples were prepared for visualization in Transmission Electron Microscopy in carbon grids. The concentrated samples were prepared in 1/10 and 1/100 dilutions on carbon coated copper grids (400 mesh Ted Pella Inc.). The grids were dipped in the archaeosome solution of appropriate dilutions for 90 seconds and then transferred to 3% glutaraldehyde solution for a minute followed by two washes with filtered MilliQ water for 45 seconds each. The samples were negatively stained for 5 minutes using 2% uranyl acetate and then kept for drying overnight placed on a clean filter paper and properly covered. The dried grids were analyzed using Tecnai G2 20 S-Twin transmission electron microscope at 200kV.

2.2.16 Pull down assay with archaeal lipids

10 μ M Tk-DGGGPS was mixed with different concentrations of archaeosome solution to check the interaction between the protein and archaeosome. Other than the concentration of the original archaeosome solution, 10 and 100-fold dilutions were also used for the assay. The three reaction mixtures containing archaeosomes and the enzyme were diluted in 50mM Sodium phosphate buffer pH 7.5, to make the total reaction volume to 20 μ l. After incubation for 30 min at room temperature and at 55 °C, the mixtures were centrifuged at 12000 rpm for 20 min. The supernatant was collected in separate microfuge tubes, and pellets were re-suspended in 20 μ l buffer having the same composition as mentioned previously. Supernatant and pellet for each of the archaeosome dilutions were run in parallel on a 12% SDS-PAGE. For the control experiment, a similar assay was performed with Tk-DGGGPS without archaeosomes where Tk-DGGGPS and archaeosome were centrifuged separately at 12000 rpm for 20 min and run on a 12% SDS-PAGE to check the presence of aggregates and presence of protein contamination in Tk-DGGGPS and archaeosome respectively.

2.2.17 DPH anisotropy

A 10 mM solution of DPH (1,6-Diphenyl-1,3,5-hexatriene) (Sigma) was prepared in DMSO. It was then added to a 100-fold diluted sample of archaeosome with the final DPH concentration

being 3 μM and incubated at 16 $^{\circ}\text{C}$ for 10 min. Anisotropy of DPH was then measured in a JASCO fluorimeter (Jasco, Japan) at 50 $^{\circ}\text{C}$, 55 $^{\circ}\text{C}$ and 60 $^{\circ}\text{C}$. Anisotropy of DPH was also measured at identical temperatures with archaeosome mixed with 10 μM of Tk-DGGGPS and TSG100+DDM buffer where the latter was performed as a control experiment. Anisotropy percentage was calculated by considering anisotropy at 25 $^{\circ}\text{C}$ as 100% and plotted against temperature for each of the three conditions. Steady-state fluorescence anisotropy (r) was calculated according to the following formulae.

$$r = \frac{I_{VV} - I_{VH}G}{I_{VV} + 2I_{VH}G}$$

where the I referred to the emission intensity, the subscripts v (vertical) and h (horizontal) referred to the position of polarizers in the excitation beam and the emission beam respectively, for every measurement. The instrumental correction factor, G , is equal to the ratio of I_{HV} to I_{HH} .

2.3 Results

2.3.1 Isolation of a novel DGGGPS gene from *Thermococcus kodakarensis*

The TK1957 gene corresponding to Di-geranylgeranylgeranyl glyceryl phosphate synthase (DGGGPS) was amplified using PCR from the genomic DNA of *Thermococcus kodakarensis* (strain ATCC BAA-918 / JCM 12380 / KOD1). The amplified product (Fig 2.3 A) was then cloned between the restriction sites of pET28a and pET28a-BRIL respectively. The restriction sites used for pET 28a were NdeI and SalI while for pET28a-BRIL the BamHI and SalI sites were used. The positive clones (Fig 2.3 B, C) were selected by colony PCR and plasmids were isolated. Restriction digestion of respective plasmids indicated the presence of a 0.9 kb insert cloned in the plasmid. The positive clones were maintained separately and plasmids were isolated and sent for sequenced. Sequencing results confirmed that Tk-DGGGPS was cloned and devoid of mutation in both the pET28a and pET28a-BRIL plasmids.

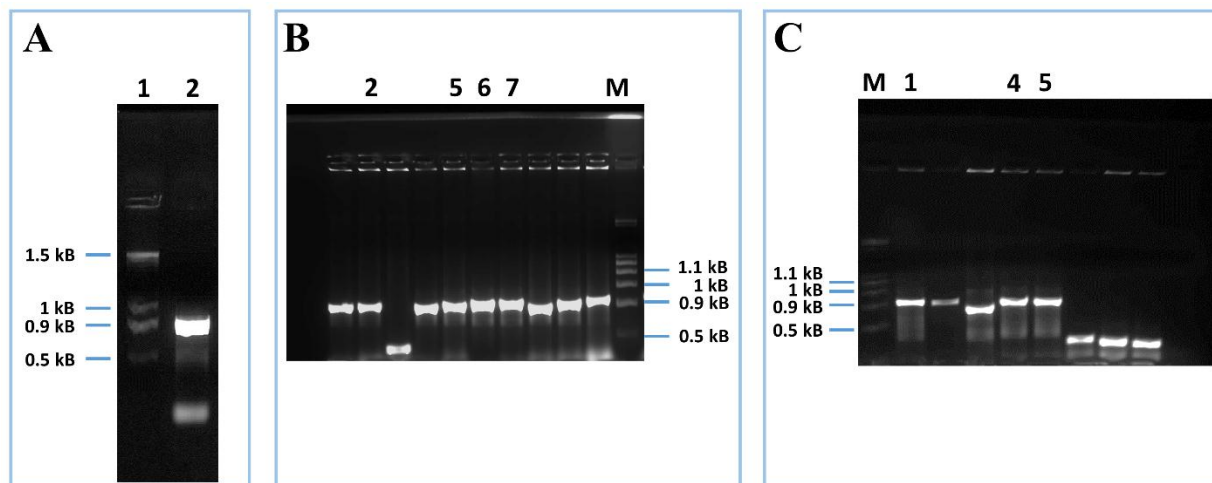


Fig 2.3 Cloning of TK1957 in pET28a and pET28a-BRIL plasmids (A) PCR amplified product of DGGGPS in 2% agarose gel. Lane 1: NEB 2Kb plus DNA ladder; Lane 2: Amplified gene product, (B) Colony PCR for confirmation of Tk-DGGGPS clones. Lane 2,5,6 & 7 are the positive clones; Lane M is the DNA ladder, (C) Colony PCR for confirmation of Tk-DGGGPS-BRIL clones. Lane 1,4 & 5 are the positive clones; Lane M is the DNA ladder.

2.3.2 Expression and Purification of Recombinant Tk-DGGGPS and Tk-DGGGPS-Bril

The recombinant Tk-DGGGPS protein with an N terminal 6XHis tag and its BRIL fusion construct was expressed in the *E. coli* C41 (DE3) cells. The stable membrane protein was then over-expressed by inducing with 1 mM IPTG at an optimized temperature of 37°C. Different detergents were used on a trial basis to determine which one extracted the protein most efficiently from the pellet fraction and kept it most stable. Notable success was obtained using DDM (n-Dodecyl- β -D Maltopyranoside). The protein was extracted from the cell membrane fraction of the cell pellet homogenate using 1% DDM after a single round of ultracentrifugation and subjected to another round of ultracentrifugation wherein the membrane protein was obtained in solution in the form of protein detergent complex (PDC). A two-step purification protocol, using Ni-NTA followed by size exclusion chromatography (SEC) was employed for Tk-DGGGPS purification. Ni-NTA purification was carried out in buffers containing different detergents at 2-3X concentrations of their CMCs (critical micelle concentrations) like 0.1% DM, 0.03% DDM, 0.7% β -OG and 0.2% LDAO, but the maximum extraction efficiency and stability of the protein was obtained using DDM. SDS-PAGE analysis of Ni-NTA affinity chromatography purified fractions on a 12% acrylamide gel indicated relatively pure protein bands at ~27 kDa (Fig 2.4 & 2.5 A).

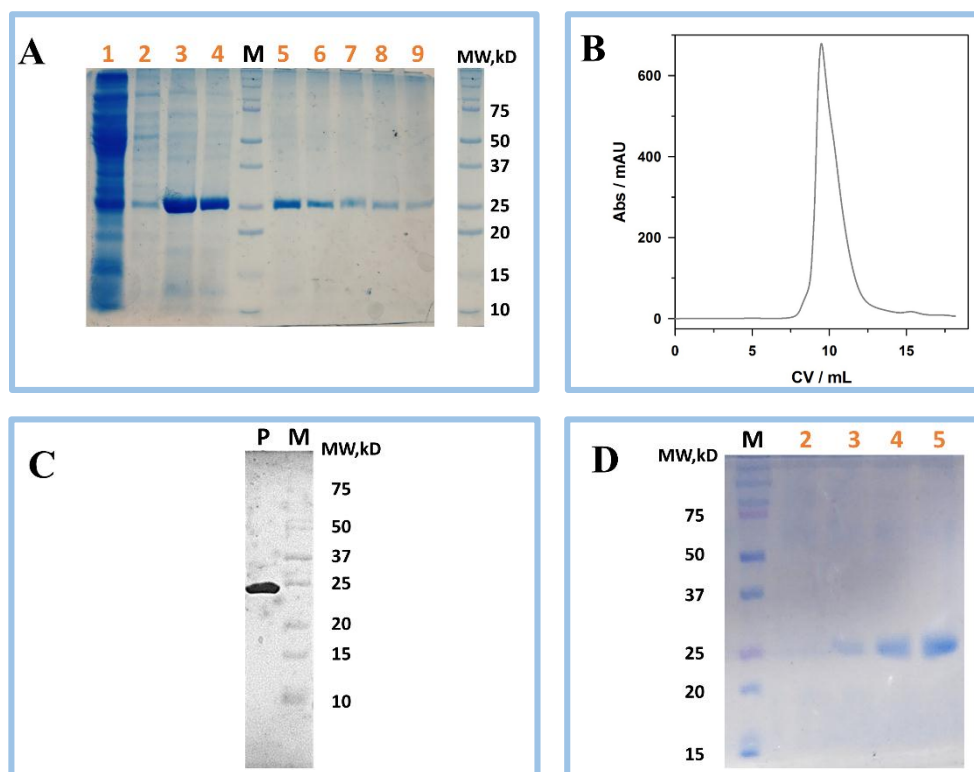


Fig 2.4 Purification of Tk-DGGGPS (A) 12% SDS-PAGE analysis showing expression of Tk-DGGGPS protein. Lane 1: Unbound fraction; Lane 2: Wash fraction; Lane M: Dual color Bio-Rad protein marker; Lane 3–9 Elution fractions, (B) SEC profile of Tk-DGGGPS in a Superdex 200 Increase columns showing a prominent, clean peak suggesting lack of aggregates, (C) Western blot analysis of purified Tk-DGGGPS, (D) SDS-PAGE of the SEC purified protein fractions showing noticeably less contaminant co-purified proteins.

Size exclusion chromatography (SEC) of the pooled positive Ni-NTA fractions showed a distinct peak with no brother peaks in its elution profile. However, a comparison with a standard run doesn't give us the correct estimate of the molecular weight of the protein due to the presence of excess DDM micelles which emphasizes the need for another technique to estimate its molecular weight more accurately (Fig 2.4 & 2.5 B, D). The western blot analyses (Fig 2.4 & 2.5 C) using His tag specific primary antibodies (PentaHis, Qiagen) confirmed that, the isolated protein was the desired Di-geranylgeranyl glyceryl phosphate synthase from *Thermococcus kodakarensis* (Tk-DGGGPS).

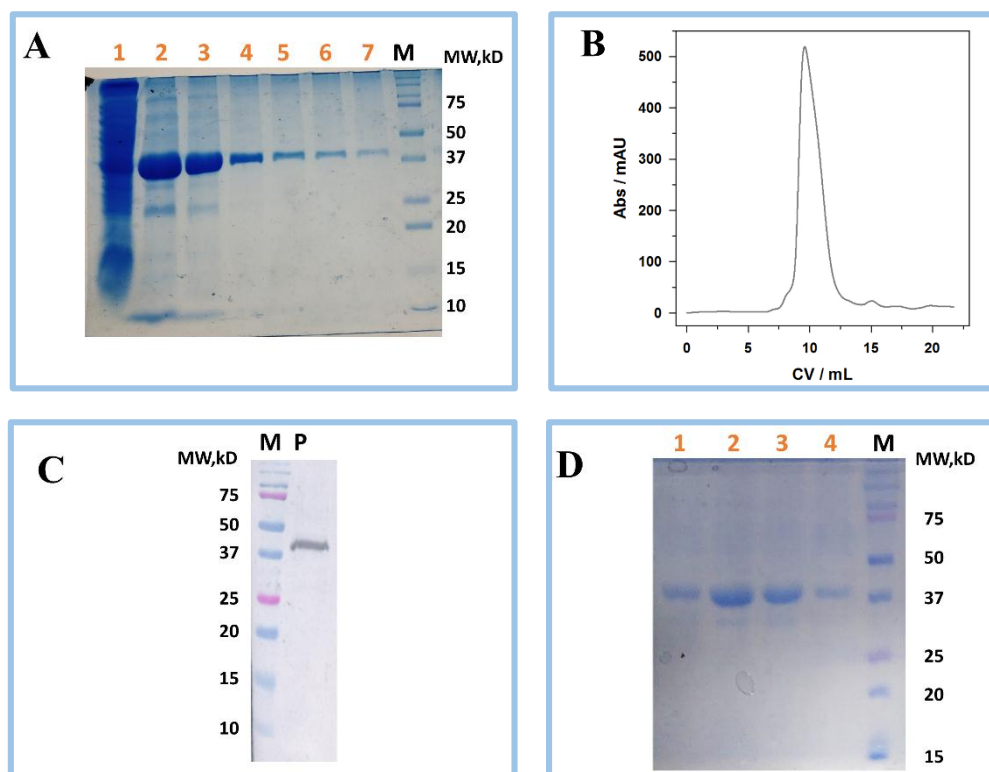


Fig 2.5 Purification of Tk-DGGGPS-BRIL (A) 12% SDS-PAGE analysis showing expression of Tk-DGGGPS-BRIL protein. Lane 1: Unbound fraction; Lane 2–7: Elution fractions; Lane M: Dual color protein marker, (B) Chromatogram of SEC purified Tk-DGGGPS-BRIL, (C) Western blot analysis of pure Tk-DGGGPS-BRIL protein, (D) SDS-PAGE of the SEC purified protein fractions.

2.3.3 Checking the stability of the Protein Detergent Complex

An ultracentrifugation dispersity sedimentation (UDS) assay was carried out to check if Tk-DGGGPS was stable in the PDC (protein-detergent complex) with DDM. The UDS assay is principled on the hypothesis that once protein aggregates form, they become orders of magnitude heavier than the normal dispersed state of the protein particles and hence sedimentation at high ‘g’ forces can be used to remove them. The samples obtained from each of the two steps of the assay were run on SDS-PAGE and the intensity of the bands showed some differences which can be expected (Fig 2.6). However, the inference from the experiment was that the PDC was stable initially but not over a long period of time. This can be justified when we take into consideration that the detergent micelles only mimic the membrane environment and thereby cannot provide

stability to the protein over a long period of time. However, this indicated that a search for the detergents which can keep this protein stable for a longer duration needed to be carried out.

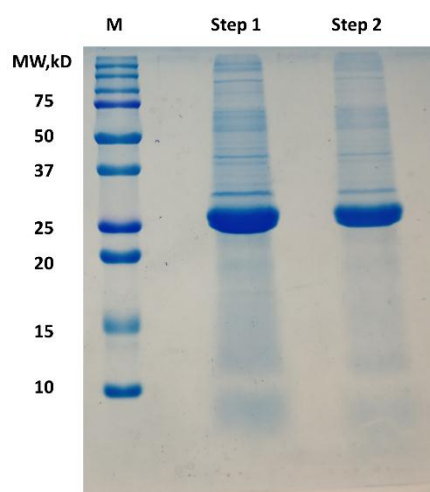


Fig 2.6 Ultracentrifugation Dispersity Sedimentation assay SDS-PAGE analysis of the samples from each step of the UDS assay indicating that the PDC is stable for the duration of the assay but has the tendency towards aggregate formation over time.

2.3.4 Estimating the oligomeric profile of Tk-DGGGPS by SEC MALLS and Blue Native PAGE

The Tk-DGGGPS PDC gives a peak which corresponds to a molecular weight of ~ 300 kDa in its SEC profile which includes the DDM micelle size as well. So to understand the actual oligomeric nature and the native molecular mass of the protein, a Blue Native PAGE (BN-PAGE) analysis of the SEC purified Tk-DGGGPS samples was carried out. The BN-PAGE profile indicates a rather heterogeneous oligomeric state profile of the protein with the presence of 2-3 distinct oligomeric states in the purified Tk-DGGGPS fractions. However, the two most prominent bands were seen to be flanking the band corresponding to the 140 kDa protein standard (Lactate dehydrogenase, bovine) (Fig 2.7 A). The higher molecular weight bands can be considered as concentration dependent aggregates of the protein.

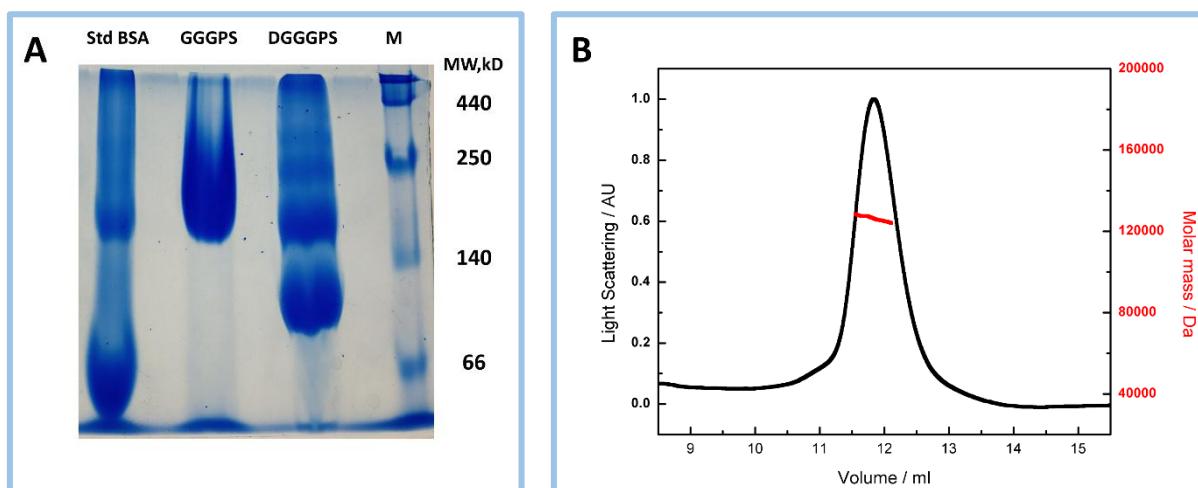


Fig 2.7 Oligomeric profiling of Tk-DGGGPS (A) 8% BN-PAGE analysis showing presence of more than one oligomeric states among which the 120kD band is the dominant state and others are soluble aggregations of the protein. Lane 1: Mixture of monomeric and dimeric BSA (1mg/ml); Lane 2: Purified Tk-GGGPS; Lane 3: Purified Tk-DGGGPS; Lane 4: Mixture of native proteins (AmershamTM make), (B) SEC-MALLS profile of Tk-DGGGPS displaying the molecular mass corresponding to the protein. The calculations were performed after proper curve fitting with refractive index and UV absorbance readings.

To study the oligomeric nature of Tk-DGGGPS more accurately, the protein was subjected to a Size Exclusion Chromatography-Multiple Angle Laser Light Scattering (SEC-MALLS) analysis. The SEC coupled with "on-line" static laser light scattering (LS), refractive index (RI) and UV detection provides a universally accepted approach for determining the molar mass and the oligomeric state as well in case of membrane protein PDCs in non-ionic detergents. The molecular mass of the purified Tk-DGGGPS was estimated to be ~120 kDa (with a polydispersity index of 1 and error percentage of 0.5%) which correlated with the results of BN-PAGE analysis (Fig 2.7 B). The monomeric weight of the protein being ~27 kDa, thereby proves the protein to exist in a homo tetrameric state in solution.

2.3.5 Gene Cloning of Tk-GGGPS

The TK1026 gene corresponding to Geranylgeranylgeranyl phosphate synthase (GGGPS) was amplified using PCR (Fig 2.8 A) from the genomic DNA of *Thermococcus kodakarensis* (strain ATCC BAA-918 / JCM 12380 / KOD1). The amplified product was then cloned between the NdeI

and Sall restriction sites of pET28a with N terminal His-Tag. The positive clones were selected by colony PCR (Fig 2.8 B) and plasmids were isolated. Restriction digestion of respective plasmids indicated the presence of a 759 bp insert cloned in the plasmid. The positive clones were maintained separately and plasmids were isolated from the same for confirmation using sequencing. Sequencing results confirmed that Tk-GGGPS was cloned without any mutation into the pET28a plasmid.

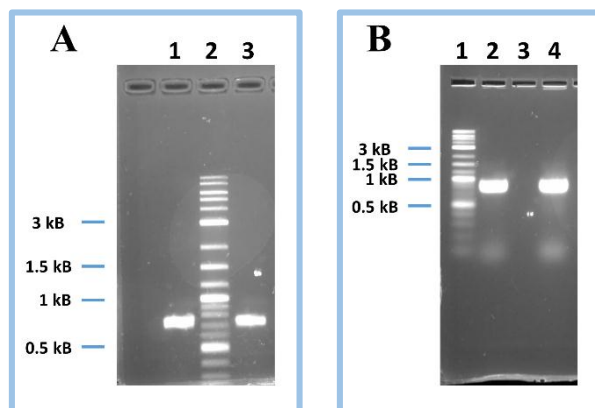


Fig 2.8 Cloning of TK1026 in pET28a plasmid (A) PCR amplified product of GGGPS in 2% agarose gel. Lane 1 & 3: Amplified gene product; Lane 2: DNA ladder, (B) Colony PCR for confirmation of Tk-GGGPS clones. Lane 2 & 4 are the positive clones; Lane M is the DNA ladder; Lane 2: PCR blank as negative control.

2.3.6 Expression, Purification and confirmation of Tk-GGGPS

The recombinant Tk-GGGPS protein with an N terminal 6XHis tag was expressed in the *E. coli* BL21 StarTM (DE3) cells. The stable and soluble protein was then over-expressed by inducing with 0.5 mM IPTG at an optimized temperature of 37°C. A two-step purification protocol, using Ni-NTA followed by size exclusion chromatography (SEC) was employed for Tk-GGGPS purification. After cell lysis by sonication the supernatant obtained was used for further purification of the protein by Ni-NTA chromatography. An SDS-PAGE analysis of the Ni-NTA purified fractions were carried out and distinct bands at ~30 kDa were observed (Fig 2.9 A). Size exclusion chromatography (SEC) of the pooled positive Ni-NTA fractions showed that Tk-GGGPS exists as a probable pentamer in solution with molecular mass ~130 kDa (Fig 2.9 C) further confirmed by BN-PAGE (Fig 2.7 A). Western blot analysis using His tag specific primary antibodies (PentaHis, Qiagen) confirmed that, the isolated protein was the desired

Geranylgeranylgeranyl phosphate synthase from *Thermococcus kodakarensis* (Tk-GGGPS) (Fig 2.9 B).

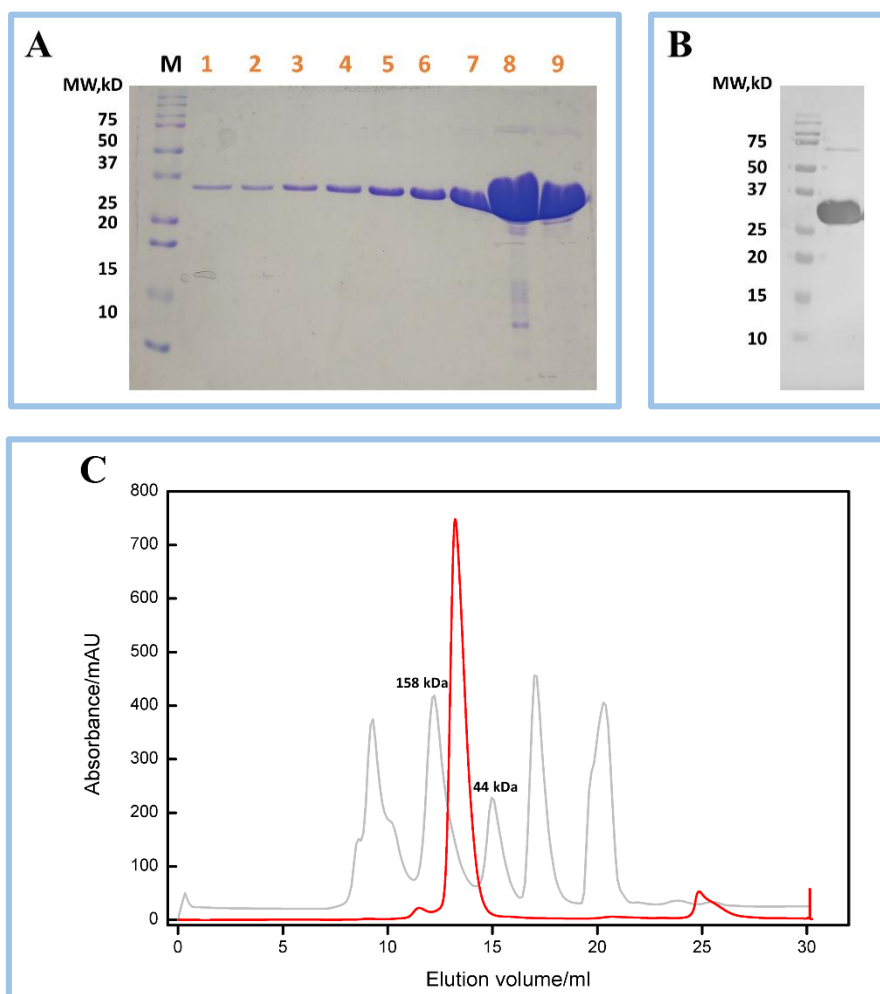


Fig 2.9 Purification of Tk-GGGPS (A) 12% SDS-PAGE analysis showing expression of Tk-GGGPS protein. Lane 1–9: Elution fractions; Lane M: Dual color protein marker, (B) Western blot analysis of pure Tk-DGGGPS-BRIL protein, (C) SEC profile of Tk-GGGPS; Peaks in gray are standard molecular weight proteins and the protein under evaluation is in red; The red peak flanked by grey colored peaks corresponding to Aldolase (158 kDa) and Ovalbumin (44 kDa) suggests that the molecular mass of the protein lies around 130 kDa further supported by BN-PAGE analysis in Fig 2.7.

2.3.7 Enzymatic synthesis of GGGP by Tk-GGGPS

Tk-DGGGPS is reported to catalyze the transfer of the geranylgeranyl group from geranylgeranyl pyrophosphate (GGPP) to the C2 hydroxyl group of (S)-3-O-geranylgeranylgeranyl phosphate (GGGP) in the second ether bond formation step in the biosynthesis of archaeal membrane lipids

(Fig 2.10). However, the prenyl group acceptor, GGGP is not commercially available. Therefore, to synthesize GGGP which is the substrate for the enzyme catalyzed reaction we cloned and purified the prenyltransferase Tk-GGGPS.

In the reaction mixture containing Glycerol-1-phosphate (G-1-P) and GGPP, addition of purified Tk-GGGPS resulted in formation of a new spot, when detected by TLC analysis of the product extracted with 1-butanol (Fig 2.11 A, B). The reaction did not proceed as expected however, when Glycerol-3-phosphate (G -3-P) was used instead of G-1-P, indicating towards the substrate specificity of the enzyme involved. A blank reaction was also performed wherein no enzyme was added to verify that the new spot obtained was not a derivative of the parent compound. A reverse phase HPLC analysis of the butanol extract of the reaction mixture also revealed a new peak, different from the peak given by GGPP (Fig 2.12 A, B). The new product was confirmed to be GGGP by negative mode ESI-MS, wherein an ion peak with an m/z value of 443 [82], corresponding to the $[M-H]^-$ of GGGP was identified (Fig 2.13 A, B).

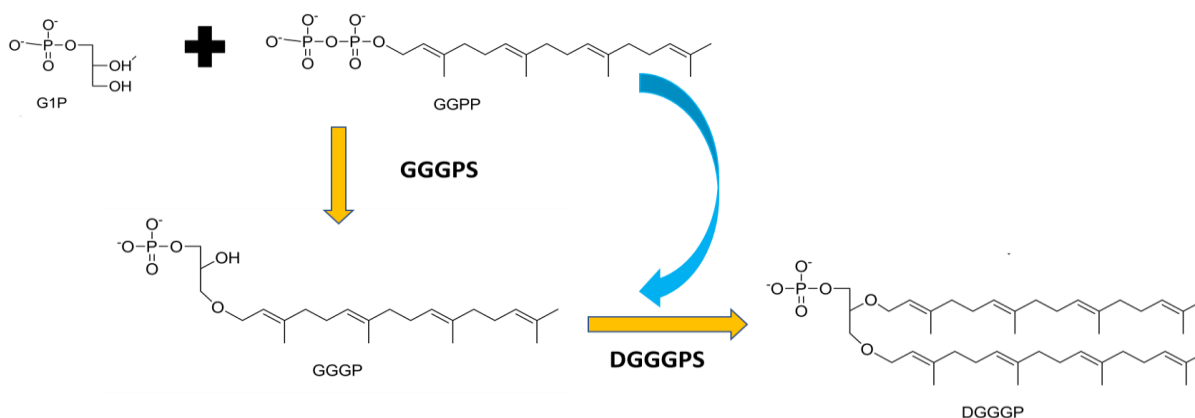


Fig 2.10 The two step reaction to generate the archaeol intermediate Di-o-geranyl geranyl glyceryl phosphate
The first step adds a glycerol backbone to GGPP (a key intermediate of terpenoid biosynthesis) to produce GGGP which is then converted to DGGGP, basically by a condensation of two GGGP molecules in the next step involving an addition of a C₂₀ chain to the glycerol moiety of the other.

2.3.8 Enzymatic activity of Tk-DGGGPS

In the reaction mixture with same components as above, purified enzyme Tk-GGGPS was added, incubated and following that purified enzyme Tk-DGGGPS was added. Since GGPP and GGGP was already present in the reaction mixture, the substrates of Tk-DGGGPS were available now. A

TLC analysis of the reaction product extracted with 1-butanol showed the generation of a new spot which indicated the formation of a new product (Fig 2.11 C). If Tk-GGGPS was not added to the reaction mixture, nothing apart from a spot for GGPP was observed and the new spot disappeared. A HPLC analysis using a C18 reverse-phase column, showed a smaller peak for GGGP and a new peak with a higher retention time than GGGP, formed after incubation with Tk-DGGGPS (Fig 2.12 C). The smaller peak of GGGP indicated its consumption in the reaction when incubated with Tk-DGGGPS. Negative mode ESI-MS of the new product was found to give an ion peak having its m/z value as 716 [40], which corresponds to $[M-H]^-$ of DGGGP (Fig 2.13 C). Collectively all the above data confirmed that the Tk-DGGGPS was enzymatically active. On substitution of GGPP by other prenyl donors like FPP for this reaction, formation of new products was not observed indicating that the reaction catalyzed by Tk-DGGGPS is very substrate specific. The enzyme activity of Tk-DGGGPS was found to be absent when 5mM Mg^{2+} was replaced with the same concentration of EDTA, which is a known metal-chelator indicating the importance of Mg^{2+} for the enzyme catalysis. The optimum pH for the enzyme catalyzed reaction was determined to be 6.0 by comparing the intensity of the spots obtained by TLC analysis (Fig 2.11).

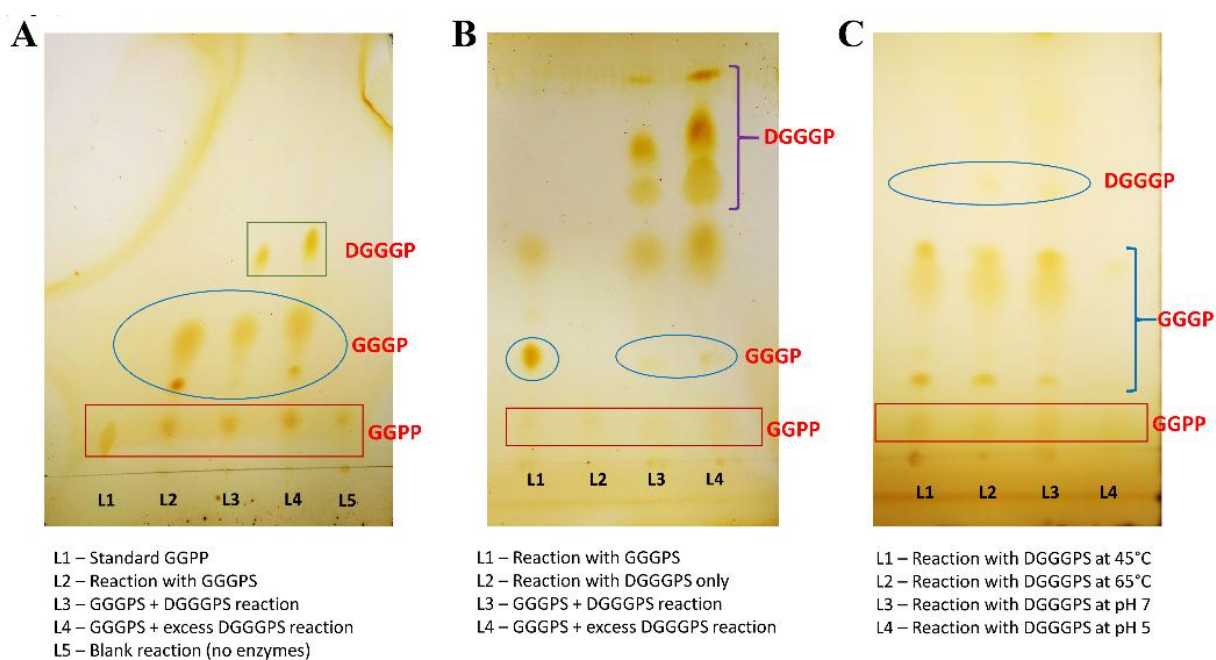


Fig 2.11 Verification of the enzyme reactions by TLC DGGGPS reaction gives low product yield at acidic pH conditions but no such variations at different temperatures. The plates were developed by exposure to iodine vapors.

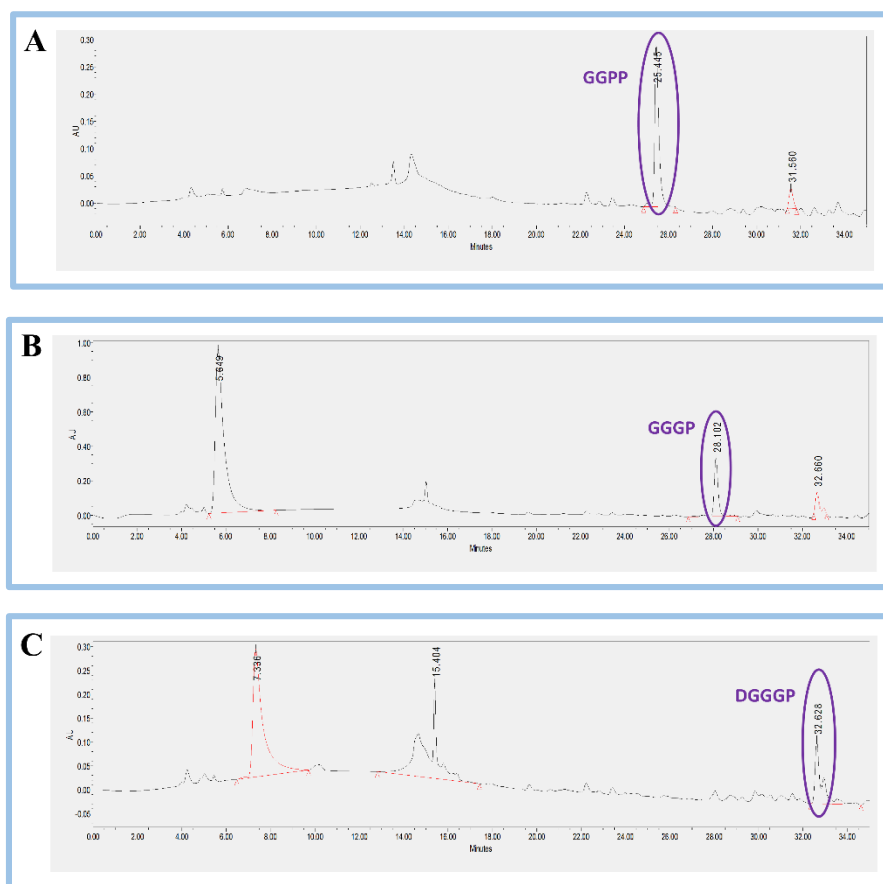


Fig 2.12 HPLC profiles of the reaction mixtures (A) Blank reaction run (B) GGGPS reaction run (C) DGGGPS reaction run. The differentiation is based on the separate retention times of GGPP (25.445 minutes), GGGP (28.133 minutes) and DGGGP (32.628 minutes) which confirm that the peaks indeed belong to three different compounds, hence confirming the success of the enzyme reaction.

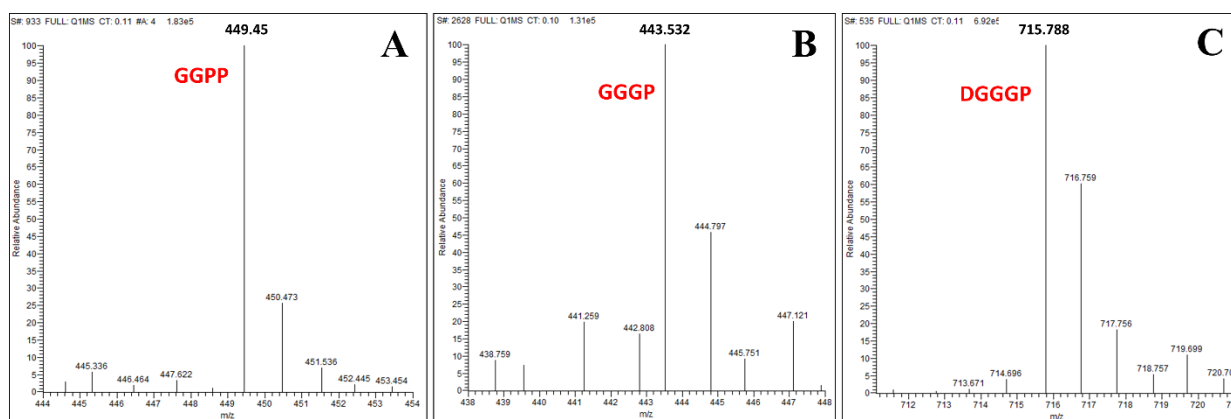


Fig 2.13 MS analysis of the reaction mixtures The experimental m/z ratios of the respective moieties (GGPP: 449, GGGP:443, DGGGP:716) confirmed their presence in the reaction mixture.

2.3.9 ITC analysis of Tk-DGGGPS with GGPP

To analyse the thermodynamic parameters of the reaction where GGPP binds to Tk-DGGGPS, ITC was performed. Our molecular docking results between GGPP and Tk-DGGGPS suggested the GGPP binding to Tk-DGGGPS at more than one site, predictably two (for more detail please refer Chapter III). So, the “Sequential binding sites” model was used to analyse the ITC results with the number of binding sites as two. The binding isotherm generated, showing the heat change over the course of titration, indicated a gradual decrease in the initial part of the titration, while it found to stabilize towards the end point of titration in a 25 injections program (Fig 2.14 A). The negative value of free energy changes (ΔG) of the reaction indicated that GGPP binding to Tk-DGGGPS was spontaneous and favourable. The K_D determined from the thermodynamic parameters for the GGPP binding at individual sites were confirmed to be 866 nM (K_{D1}) and 273 μ M (K_{D2}) respectively indicating nearly a 300-fold difference between the binding propensities at the two sites (Fig 2.14 B). These results, also supported by our molecular docking results (refer to Chapter III), suggested that the initial binding between GGPP and Tk-DGGGPS was weak and probably transient in nature and influenced the second GGPP binding with higher K_{D1} in an allosteric fashion. Towards the end of the titration, the system tended to move towards a state of equilibrium, characterized by constant association and dissociation of GGPP and Tk-DGGGPS.

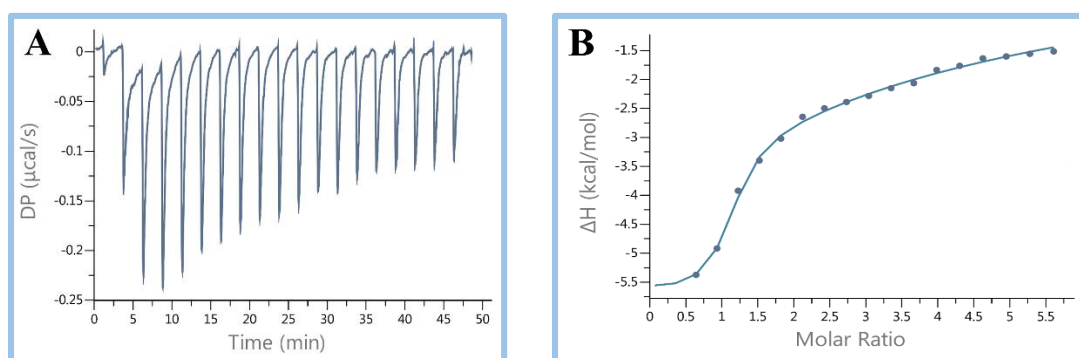


Fig 2.14 Isothermal titration calorimetry measurement of Tk-DGGGPS and GGPP binding interactions 25 μM of purified Tk-DGGGPS and 750 μM of GGPP were used for the study and titrations performed at 55°C. (A) The heat change occurring in the reaction is shown in micro calorie per second per injection, (B) The change in enthalpy (kcal/mol) as a function of the ligand–protein molar ratio is shown, derived from a sequential binding site model fit at $n=2$ binding sites; The K_D of binding at the sites suggest that one site has weaker interaction ($K_{D2} = 273 \mu\text{M}$) while the other site has a strong affinity ($K_{D1} = 866 \text{ nM}$) of ligand binding to the enzyme.

2.3.10 Circular dichroism study of Tk-DGGGPS

The far UV Circular dichroism (CD) spectrum of Tk-DGGGPS had positive ellipticity maxima at 200 nm and negative minima at 208 nm and 222 nm (Fig 2.15 A). The spectrum was analyzed with CD Pro to estimate the secondary structure contents as 70% α -helix, 8% loops and 22% random coils at 25°C, suggesting the Tk-DGGGPS to be very rich in α -helices and β -sheets were found to be totally absent. The ordered CD profile also indicates that the recombinant protein was in a stable and properly folded form. Thermal stability of the protein provides information about the non-covalent and covalent forces involved in stabilizing the enzyme structure. Secondary structure monitoring during thermal denaturation CD spectroscopy revealed almost unperturbed CD profile up to 70°C after which the protein starts unfolding and starts precipitating at ~85°C (Fig 2.15 B, C). The protein might have started precipitating due to partial unfolding as evident from the graph which is also in-line with the comparison of the CD spectra at different temperatures.

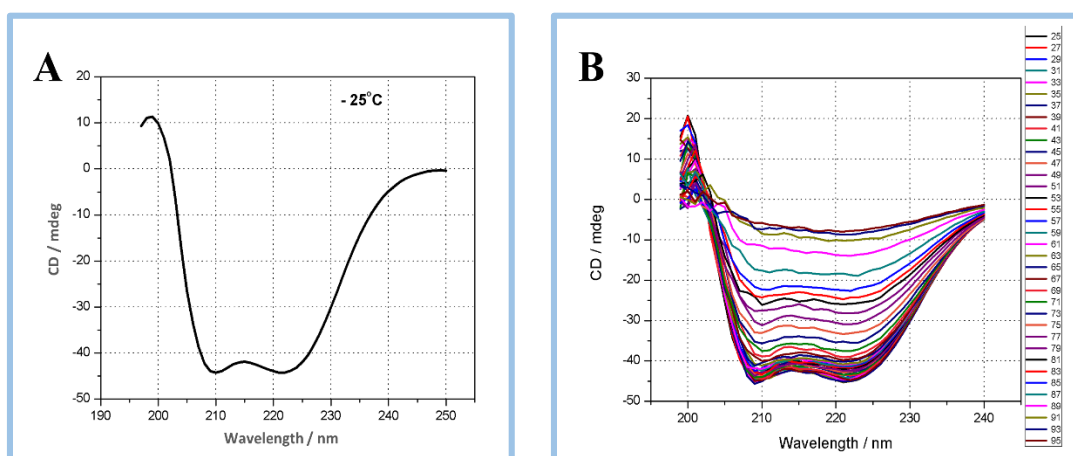


Fig 2.15 Far-UV CD spectra analysis of Tk-DGGGPS (A) A scan at 25°C shows prominent grooves at 208nm and 222nm, characteristic of α -helix rich proteins, (B) A temperature melt of the protein shows that the overall secondary structure has no notable change till around 70-75°C after which it starts to denature, probably due to visible precipitation in the protein.

We used the two-state model to fit the temperature induced unfolding transition, monitored by Far-UV CD. The midpoint of transition or melting temperature (T_m) of the protein appeared to be ~75°C. The decrease of the α -helix percentage in Tk-DGGGPS secondary structure was evident, as the mean residual ellipticity at both 208 nm and 222 nm gradually increased with increasing

temperature (more prominently at temperatures above 75°C), which is an indication of decreasing α -helical content.

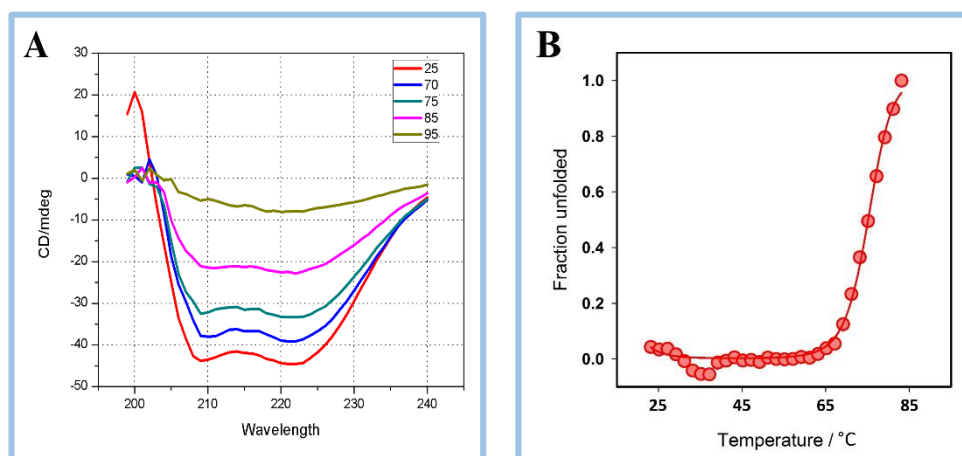


Fig 2.16 Global structure analysis and estimation of T_m of Tk-DGGGPS from Far-UV CD analysis (A) The thermal denaturation plot shows the changes in secondary structure as a function of wavelength for five different temperatures color coded, ranging from 25°C to 95°C, (B) The fraction of protein in the unfolded form (f_U), determined using the given equation, is plotted as a function of increasing temperatures to generate a thermally-induced denaturation curve.

2.3.11 Fluorescence spectroscopy of Tk-DGGGPS

Intrinsic fluorescence spectra of Tk-DGGGPS was recorded using fluorescence spectroscopy, where the protein showed λ_{max} at 332 nm indicating that the Tryptophan and Tyrosine residues are partially buried in the membrane. The tertiary structure of the protein was evaluated under native conditions and at increasing temperatures using intrinsic Trp and Tyr fluorescence. The protein was excited at 280nm to check for the excitation emissions of Tyr and Trp. The Trp and Tyr fluorescence intensity of Tk-DGGGPS was found to decrease gradually upon heating at higher temperatures (30, 35, 40, 45, 50, 55, 60, 65, 70, 75 and 80 °C) indicating a perturbation of the microenvironment of the four Trp and nine Tyr residues which can be a result of the loosening of the tertiary structure of the protein (Fig 2.17 A).

At the emission maxima, we observed that there is a gradual decrease in the fluorescent intensities with increase in temperature, with immediate effect. This is in contrast to the CD experimental results where the change in CD signal is apparent only after 70°C. This indicates that even though there is no considerable change in the secondary structure of the protein, the tertiary structure

loosening is taking place via a probable formation of a molten globule ensemble at the intermediate temperatures. Molten globules are intermediate states where the secondary structure remains intact while the tertiary structure is partially loosened. It is interesting to note however that the temperature ranges at which the enzyme is most enzymatically active is 50-60°C which are included in these intermediate temperatures mentioned above.

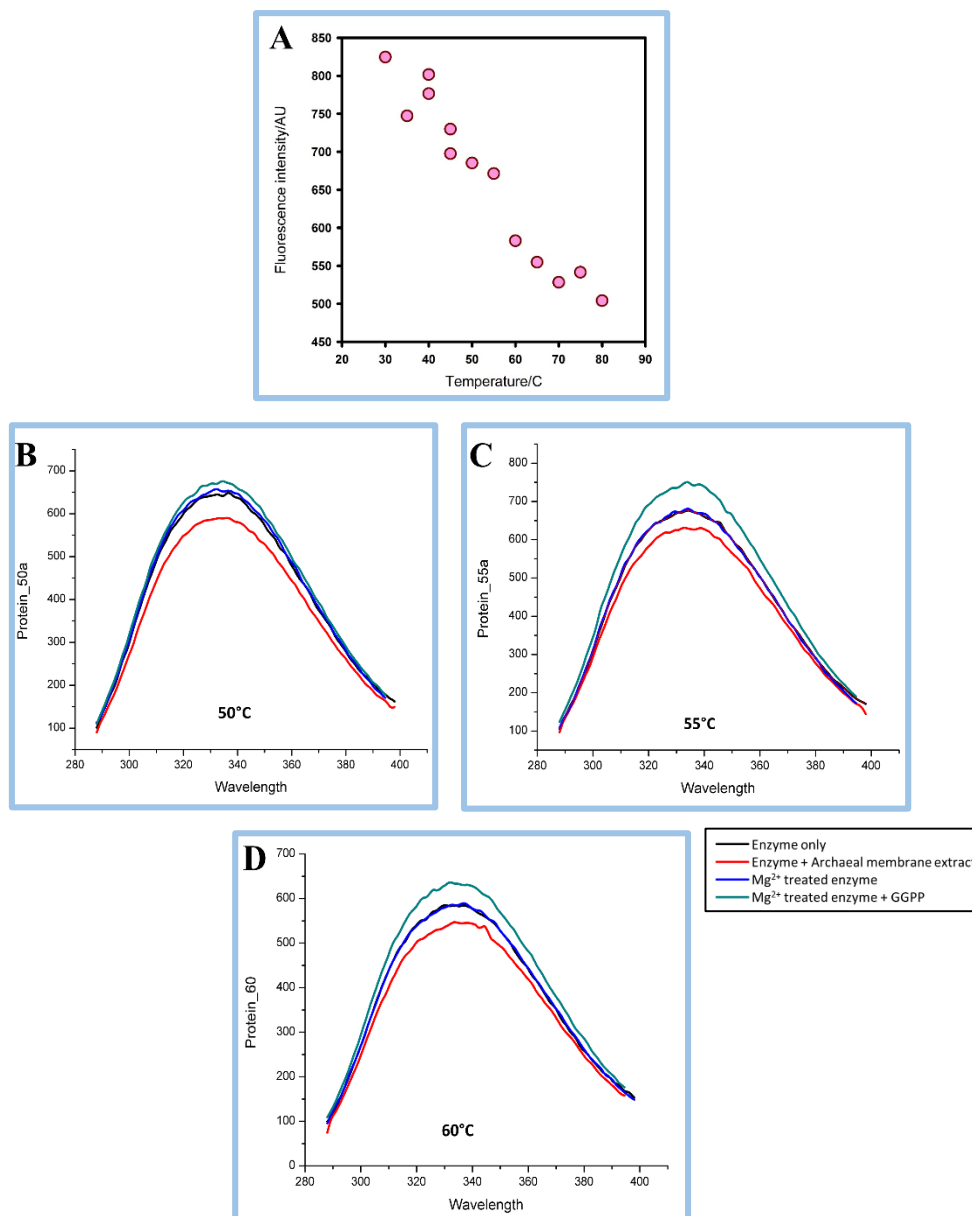


Fig 2.17 Intrinsic fluorescence spectra of Tk-DGGGPS (A) Temperature induced changes in the intrinsic Tyr & Trp fluorescence spectra of Tk-DGGGPS, (B) (C) & (D) Comparative fluorescence spectra analysis of the enzyme at three different temperatures 50°C, 55°C and 60°C (where the enzyme is the most active), in four reaction conditions.

To study in further detail about tertiary structure changes at this temperature range (50-60°C), intrinsic fluorescence spectra of the following reactions, only enzyme, enzyme incubated with archaeal membrane extract, Mg²⁺ treated enzyme and Mg²⁺ treated enzyme incubated with GGPP were recorded. Interestingly, the fluorescence spectra of the native enzyme and Mg²⁺ treated enzyme were identical indicating no distinct structural changes. However, there was an upward shift in the fluorescence spectra in the Mg²⁺ treated enzyme + GGPP reaction while there was visible quenching of fluorescence in the enzyme + archaeal membrane extract incubation. From these results we could infer that the binding of the ligand (GGPP) to the enzyme brought about a distinct change in the microenvironment of the Trp and Tyr residues which indicated their influence on the active site conformational change and suggested their involvement in the overall active site chemistry (Fig 2.17 B, C & D). The quenching effect in the archaeal membrane extract treated enzyme might be due to binding of the lipids to the residues around the surface aromatic amino acids or the hydrophobic patches of the cytoplasmic cap.

This points us towards a temperature dependent conformational change in the active site structure of the Tk-DGGGPS where with increase in temperature there are distinct changes in the structural dynamics of the active site residues leading to a probable temperature induced activation of the enzymes to catalyze its substrates more effectively.

2.3.12 Surface structure morphology analysis of Tk-DGGGPS by ANS binding

To understand whether the changes in tertiary and quaternary structures of Tk-DGGGPS under pre-heating conditions brought about any variations in the hydrophobic binding pockets on the surface of Tk-DGGGPS, surface hydrophobicity was measured using a hydrophobic probe, ANS (8-anilino-1-naphthalenesulfonic acid). ANS is an extensively utilized fluorescent probe for the characterization of protein binding sites and buried hydrophobic sites of proteins. This hydrophobic probe exhibits very weak fluorescence in an aqueous environment, however when it binds to the hydrophobic surface of any protein, its fluorescence is enhanced. Two experiments were carried out with ANS. In the first one, the protein was incubated with increasing concentrations of ANS. This study indicated that the protein has a well-structured hydrophobic cavity where ANS binds with dissociation constant of 19.2 μM. An identical non-interacting

binding site count of 1.5 indicated that there may be possibly two prominent hydrophobic grooves, one located centrally and another laterally (Fig 2.18 A).

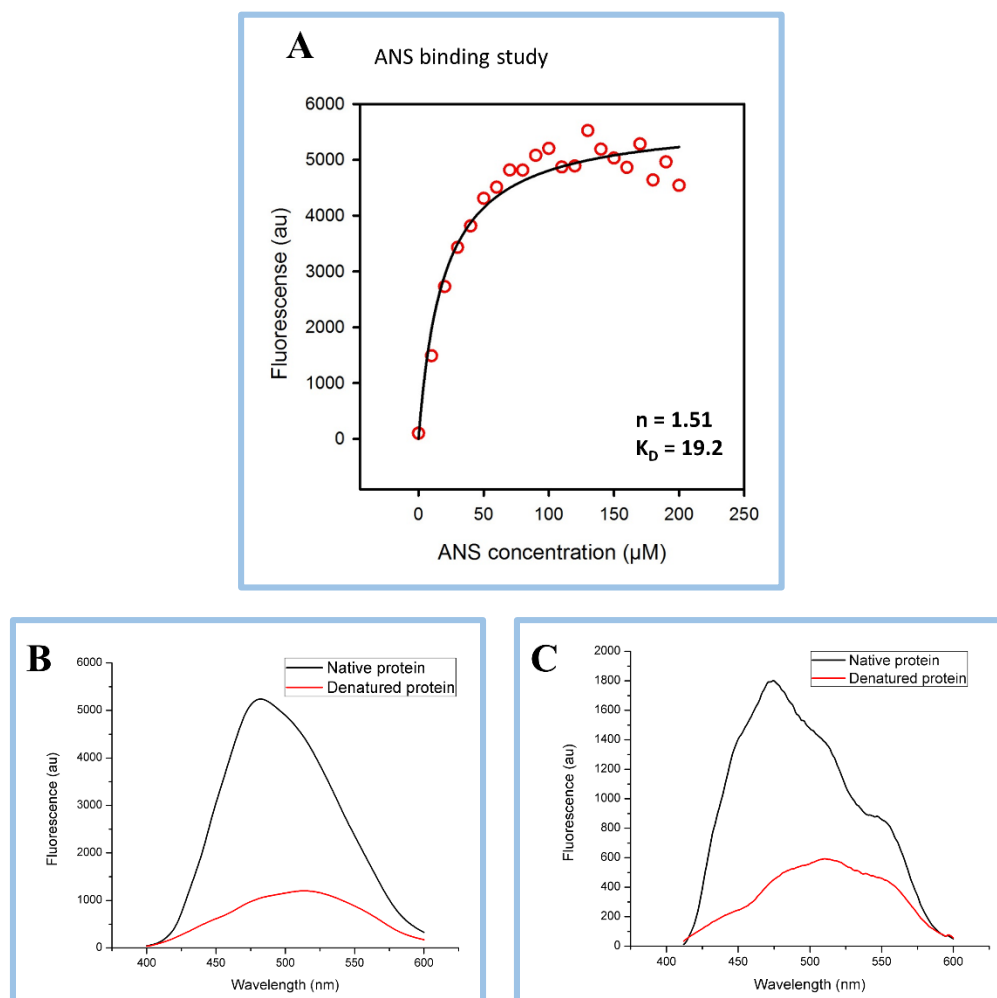


Fig 2.18 ANS binding dynamics of Tk-DGGGPS (A) In presence of its increasing concentrations, ANS binds to the hydrophobic grooves of the protein with a high K_D of $19.2 \mu\text{M}$ and calculations on basic of the curve fitting suggested the binding site count to be 1.5. The system only started to attain equilibrium around a high concentration of ANS ($\sim 200 \mu\text{M}$), suggesting the presence of bulky binding sites where the fluorescent hydrophobic probe could bind, (B) The urea induced denaturation severely hampered the ANS binding to the protein, apparent from the clear shift of the emission maxima to 525 nm from the experimentally determined maxima of 480 nm , confirming the loss of hydrophobic pockets lead to the inability of ANS to bind to the protein, (C) The buffer corrected plot supporting the data from the previous plot.

In the second study, commonly known as a Urea melt, the fluorescence intensity of ANS bound to Tk-DGGGPS treated with increasing concentrations of urea ($0\text{-}8\text{M}$) was found to reduce markedly, suggesting a decrease in surface hydrophobicity of the protein with increasing concentrations of

urea and thereby indicating a denaturation of the overall protein structure. Denatured proteins lose the architecture of their hydrophobic cavities, hence ANS was found not to bind and showed free dye emission maxima at ~525 nm (Fig 2.18 B, C). The study also suggests that these hydrophobic binding pockets are highly specific for substrate binding since the structure of ANS is similar to the substrates of the other UbiA superfamily members.

2.3.13 Archaeal lipid isolation, archaeosome preparation and visualization (TEM)

Total archaeal lipid extracts were isolated from cell pellets of 100 ml cultures each of *Sulfolobus acidocaldarius* DSM639 and *Haloferax volcanii* strain H53 using the solvent system of 2:1 Chloroform/ Methanol and their concentration was determined using the Stewart assay for lipid quantification as 0.3 mg/ml for the Saci_lipids and 0.32 for the Hvo_lipids. The lipid extracts were subjected to ultra-sonication for 10 cycles of 30 second on and 15 second off each to form archaeosomes. The structure of the archaeosomes formed using the Saci_lipid extracts (Fig 2.19) and Hvo_lipid extracts visualized using TEM indicates the archaeosomes to be distinctly ordered.

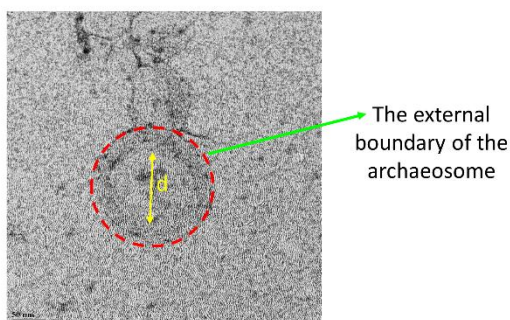


Fig 2.19 TEM visualization of the archaeosomes generated from the Saci_lipids The image shows the external boundary of the archaeosome and the internal diameter of the archaeosome is indicated by 'd'.

2.3.14 Interaction of archaeal lipids with Tk-DGGGPS

Pull-down assays were performed using the archaeal lipid extracts in which Tk-DGGGPS was incubated with the Saci_lipid extracts at different dilutions of the lipids (1:1, 1:10, 1:100). The analyses were carried out at room temperature and at 55°C. Positive results obtained for the first two dilutions showed that there was increased interaction between the Tk-DGGGPS and the archaeal lipid extracts at higher temperatures compared to no detectable interactions at room temperature (Fig 2.20 A, B). Among which the dilution of 1:10 showed comparatively better

results. This experiment indicates the increase in affinity of probable binding sites to the archaeal lipids in a temperature dependent manner. The justification for the same could be changes in surface morphology at specific temperatures which renders the enzyme active due to changes in the inter-residue interactions at such temperatures resulting in possible disruptions of local bonds which in turn facilitates the interaction of the lipids with the enzyme Tk-DGGGPS.

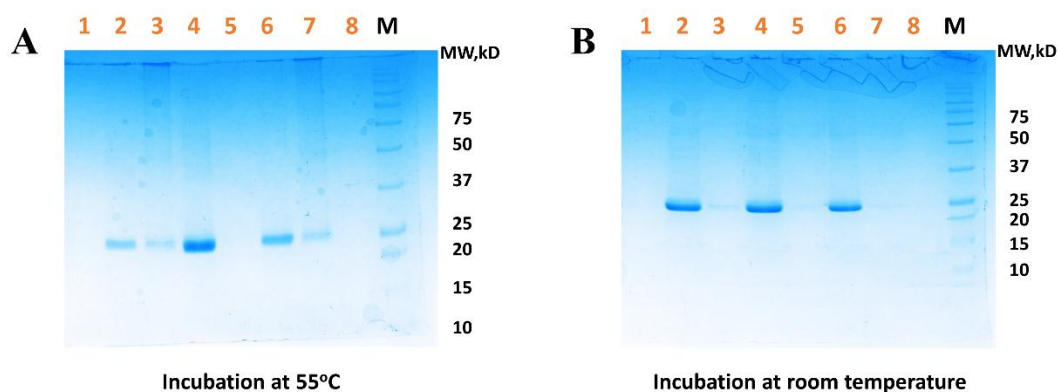


Fig 2.20 SDS-PAGE of pull down analysis of Tk-DGGGPS with archaeal total lipid extracts A 12% SDS-PAGE with sample details as following: Lane 1: Pellet of centrifuged lipid solution; Lane 2: Supernatant of protein incubated with lipid; Lane 3: Pellet of protein incubated with lipid; Lane 4: Supernatant of centrifuged protein solution; Lane 5: Pellet of centrifuged protein solution; Lane 6: Supernatant of protein incubated with lipid diluted in 1:10 ratio; Lane 7: Pellet of protein incubated with lipid diluted in 1:10 ratio; Lane 8: Blank reaction; Lane 9: Pre-stained protein marker. (A) After incubation at 55°C there was visible interaction between the enzyme and the Saci_lipids, (B) Room temperatures incubation showed a lack of interaction between the enzyme and lipids.

2.3.15 Archaeosome stabilization by Tk-DGGGPS

Archaeosomes generated from the isolated lipid extracts from *S. acidocaldarius* and *H. volcanii* strains were treated with DPH, and the steady-state fluorescence anisotropy of DPH then measured the variation of membrane fluidity in the absence and presence of Tk-DGGGPS (Figure). A control experiment was also run with the buffer containing DDM in which the protein was maintained. There was a distinct increase in the anisotropy measurements reflecting the decrease in fluidization of archaeosomes (Fig 2.21) due to the presence of Tk-DGGGPS. The measurements showed an increase in membrane compactness of ~40% for Saci-archaeosome and a rise of ~60% for Hvo-archaeosome. Comparative results were obtained in the presence of BSA, which was used as a

negative control, and showed no changes in the membrane fluidization of archaeosomes. These results indicated an important and hitherto unknown role of Tk-DGGGPS in influencing membrane fluidity in archaea.

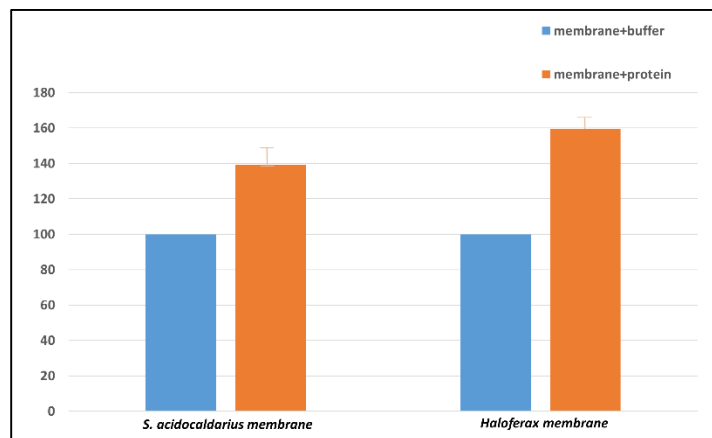


Fig 2.21 Steady-state fluorescence anisotropy measurements of DPH incubated with Tk-DGGGPS A prominent increase in the anisotropy readings of the reaction mixtures in presence of the enzyme and the respective lipids showed a percent increase in archaeosome stability of the respective lipids. A reaction blank was taken using the buffer (containing the co-purifying detergent DDM) in which the enzyme was solubilized, to negate any influence DDM could have on the stability of the archaeosome. The error bar represents the spread in the data from two independent experiments.

2.4 Discussion

Di-geranyl geranyl glyceryl phosphate synthase (DGGGPS) catalyzes the second committed step towards the biosynthesis of archaeal membrane lipids. It is responsible for a key step towards the formation of the bi-prenylated /lipidated precursors of archaeol. The importance of this enzyme lies in the fact that it is also the first enzyme of the pathway which is located in the membrane of the archaeal cell. Therefore, it shifts the lipid generation machinery from the cytosol to the membrane where the rest of the pathway is carried out. What also makes it an important study among other prenyltransferases is the fact that DGGGPS deals with strongly hydrophobic, long carbon prenyl chain containing, linear substrates as the precursors to its final catalyzed product DGGGP. No other prenyltransferase is known to deal with such an array of substrates which makes its reaction dynamics unique from other prenyltransferases. Its location in the membrane makes it important to look at the membrane assembly and the probable interactions with the lipid

component of the membrane, which could have an important role to play in the biochemistry of the enzyme. The presence of the enzyme in the archaeal cell membrane meant that it could have an impact on the membrane stability and its functioning could be influenced according to changes in external environments.

The protein was purified in two forms, one having a N terminal 6XHis tag and the other one having two tags at the N-termini, the 6XHis tag and a BRIL (cytochrome b562RIL) tag. While the initial BRIL construct was made from a crystallographic point of view since BRIL tags are known to successfully assist crystallization of GPCRs [83], we observed a positive influence on both the protein yield and stability in solution. However, since the location of the tags in the protein structure were essentially on its surface and not buried, the influence of the BRIL tag on the functionality of the protein were not studied. Since the theoretical isoelectric point of Tk-DGGGPS was found to be 9.7, buffer of pH 8.0 was used for purification of the protein. Being a membrane protein, there was another important aspect to the Tk-DGGGPS purification which was the co-purifying detergent with which the protein forms a complex. The detergent micelle mimics the lipid environment of the membrane to the protein which is therefore encapsulated in the micelle while it exists in the buffer solution of suitable pH. However, even though the protein was found to be initially stable in the PDC, the dynamics of the micellar environment resulted in the time dependent aggregation of Tk-DGGGPS. This makes it important to choose the detergent component with care to ensure proper structural and functional studies of membrane proteins [84].

Oligomeric profiling of proteins and more importantly enzymes, are often suggestive of how their assembly influences the protein biochemistry. Here the purified Tk-DGGGPS showed the presence of a variety of oligomers in its BN PAGE profile which could include the natural oligomeric form of the enzyme as well as intermediate states and aggregates. Determination of the dominant oligomeric form of the protein was thereby carried out using multiple angle light scattering (MALS). This is a very effective technique since it correlates the differential refractive index and light scattering to distinguish between the size of the particles present in solution, in this case the protein molecules and the detergent molecules and in the process gives us definitive values for the molar mass of the protein in solution. Since there is only one reported chain in Tk-DGGGPS protein, the results suggested the formation of a homo tetramer. It would be interesting to study in future, how the molecular mechanism of catalysis is influenced by this oligomeric assembly.

The Tk-DGGGPS and Tk-GGGPS, both prenyltransferases of the archaeal membrane lipid biosynthesis pathway, but residing in different fractions of the cell, were found to be very substrate specific. The experimental data indicated that these enzymes were influenced by temperature for their efficiency of catalysis. These two aspects of these proteins could be subject to further protein engineering to synthesize novel lipids which could have important medical implications. The basis of this temperature-based efficiency could be a change in the local hydrogen bonds between residues at the active site, due possible change in orientation or conformation of the amino acids brought about by changes in temperature. The ITC results suggest the allosteric activation of the Tk-DGGGPS enzyme. This holds important significance in the membrane biology of archaea, where only an excess amount of free GGPP would result in further archaeol precursors to be synthesized. This could be energetically favorable for the archaeal survival in harsh conditions. Also, the archaeosome stabilization studies showed that presence of Tk-DGGGPS in the membrane had a positive influence on its structural compactness. The fluidity of the archaeosome was found to decrease prominently in presence of Tk-DGGGPS indicating a two pronged role of this enzyme in membrane stabilization, i.e., by influencing the structural make-up of the membrane lipids and stabilization by its membrane assembly. All of these indicated that Tk-DGGGPS could be an important clog in the survival of archaea and also could be a primary tool for engineering medically important lipids.

Chapter III – Structural insights
and computational analysis of
Di-geranylgeranylglyceryl
phosphate synthase (Tk-
DGGGPS)

3.1. Introduction

An initial assessment of the archaeal membrane lipid biosynthesis pathway indicated it to be a fairly heterogeneous pathway in terms of the enzymes involved. The initial steps of the pathway have since then, been robustly defined and established [41]. The final steps of the pathway however remain to be fully elucidated. Notably, in different species of archaea, there are alternative enzymes designated to more than one of the steps involved [34]. The first five steps however are well defined in terms of their order and catalyzing enzyme and conserved over all known species of archaea. The first dedicated step of the pathway is the second step which occurs in the cytosol. It is followed by the first of the rest of the steps which take place in the membranes, catalyzed by DGGGPS. This step is important since it marks the transition of the membrane building machinery from cytosol to the membrane and a protein structure of atomic resolution can give detailed insight into the mechanism of action of this enzyme and how it varies from the soluble prenyltransferases. DGGGPS also catalyzes prenyl transfer between highly hydrophobic linear prenyl donors (GGPP) and acceptors (GGGP) which makes it very distinct from the other UbiA superfamily members. The chain lengths of the above molecules, each consisting of 20 carbon atoms, would preferably require elongated voluminous cavities, for the binding of each of the molecules. So, DGGGPS could have a prolonged tunnel shaped opening in its active site along the characteristic central cavity (like the Af-UbiA homolog) [32] and also an extended basic pocket (more along the line of the Ap-UbiA homolog) [33] to assist in the interaction with the pyrophosphate group. So essentially it could encompass two distinct structural features of the two homologs structurally elucidated till date and hence may be more prominent from an evolutionary point of view, indicating that DGGGPS might have been one of the earlier evolved prenyltransferases. This is a very adventurous claim and needs to be backed up with more structural and functional data. Structural studies may also reveal in much more explicit detail, how these prenyltransferases can accommodate such a wide variety of substrates in its enzymatic reactions. Other aspects of studies in this field would be to engineer the residues in the active site, in this case the central cavity and the basic pocket, in order to synthesize novel lipids of biological importance.

In this chapter, we attempt to delve into more details on the structural aspect of this enzyme, Tk-DGGGPS and thereby gain crucial insight into its active site morphology and try to fathom, why

the core structure of this family of enzymes are conserved over such a wide range of organisms in spite of being involved in key steps of many different biosynthetic pathways. We also try to relate it with residue conservation studies to determine important residues which are key towards their structural uniqueness and how their location in the protein structure influences its function. We even predict a moonlighting function of the enzyme in a membrane environment where it also functions as a transporter in absence of catalyzable substrates. We also aim to investigate, the mutational hotspots in this versatile superfamily of enzymes which happen to be among the conserved residues found throughout this superfamily and evaluate their impact on the biochemistry of these enzymes.

3.2. Materials and Methods

3.2.1 Materials

Commercial screens procured from Hampton research, USA and Qiagen, Germany were used for initial crystallization screening. Sodium cacodylate, lithium sulfate, PEG 4000, Ethylene glycol, glycerol, 2-methyl-2,4-pentanediol, propan-2-ol *et cetera*, used in crystallization trials were obtained from Sigma-Aldrich, USA. Two-well sitting-drop plates were obtained from Hampton research, USA. 24 well plates and coverslips were obtained from Corning® (Sigma- Aldrich, USA) and Blue Star, India respectively. Other specialized chemicals and instruments used in the experiments are mentioned in the appropriate places.

3.2.2 Crystallization of Tk-DGGGPS and Tk-DGGGPS-BRIL

Purified Tk-DGGGPS and Tk-DGGGPS-BRIL were purified to homogeneity and concentrated up to 10 mg/ml using Amicon ultra 50 kDa cutoff centrifugal filters (Millipore, USA) at 4 °C at 4500 rpm. To find the crystallization conditions, the proteins were initially screened against several commercially available crystallization screens including, Index (Hampton Research), PEGRx 1 (Hampton Research), MembFac (Hampton Research), Nextal PACT (Qiagen), Nextal Classics Suite (Qiagen) and JCSG plus (Molecular Dimensions). Screenings were performed using the Mosquito Robot (TTP Labtech) by vapor diffusion method in 96 well sitting drop Hampton MRC-SD2 plates with different protein to screen ratios of 1:1, 1:2, 1.5:1 and 2:1 in drop volumes of 300µl, 350µl and 400µl and incubated at 20°C in a RuMed vibration free incubator.

Further the Tk-DGGGPS and its Bril variant was screened against commercially available crystallization screens optimized for membrane proteins namely, MEMGOLD and MEMGOLD2 suites (Molecular Dimensions) in 96 well MRC2 sitting drop plates Hampton MRC-SD2 plates, by vapor diffusion method, with protein to screen ratios of 1:1, 1.5:1 and 2:1 in 400 μ l drop volumes using the Mosquito robot (TTP Labtech) and incubated at 20°C. Further optimization was carried out using additive screens, Silver Bullets (Hampton Research).

Two specialized crystallization methods primarily for membrane proteins, HILIDE and LCP were attempted with the proteins. Crystals were checked under a light microscope.

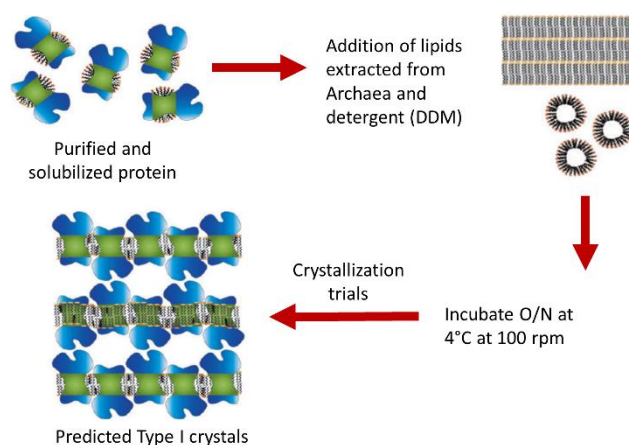


Fig 3.1 Diagrammatic representation of the HILIDE technique of crystallization to increase the probability of formation of ordered Type I crystals.

3.2.3 HILIDE (High Lipid High Detergent)

For crystallization attempts of Tk-DGGGPS in excess lipid and excess detergent, the archaeal total lipid extracts obtained in the previous chapter, were distributed in different vials and vacuum dried to remove the phosphate buffer. The lipids were then dissolved in chloroform to maintain two vials in the same concentration as before (0.3 mg/ml) and another two vials in half the concentration (0.15 mg/ml). They were spun at 14000 rpm for 10 minutes to remove any precipitation and transferred to GC vials, where the solvent was evaporated under a stream of nitrogen gas. The concentrated Tk-DGGGPS-DDM purified protein solution (10mg/ml) was added slowly along the sides of the GC vials and then excess detergent DDM was added so that the final detergent concentration in solution was maintained at 1%. It was incubated at 4°C in a shaker with low rpm (~50) overnight. The overnight mixture was then mixed thoroughly with pipette and checked for

precipitation of which none was observed. Crystal setup was done in MRC-SD2 plates using Mosquito robot with the positive precipitation conditions from the previous experiment with manual variations in the compositions as well as the standard commercial screen.

3.2.4 LCP (Lipidic Cubic Phase)

For *in-meso* crystallization attempts of Tk-DGGGPS, the mesophase was prepared by combining monoolein as the hosting lipid with the concentrated Tk-DGGGPS-DDM purified protein solution (5mg/ml) in coupled Hamilton Gastight® Syringes in a volume ratio of 3/2, as described previously [85, 86]. The monoolein was first melted at ~50°C to facilitate loading it, as an oil, into the microsyringe. The experiment was carried out at an ambient temperature above 25°C, and the syringe was also warmed to ~25°C ahead of filling with molten lipid. The protein solution was equilibrated to $\geq 25^\circ\text{C}$. The mesophase achieves a characteristic smooth viscous texture and is optically clear and isotropic (nonbirefringent) when formed properly. The protein-laden mesophase once formed is transferred to a 10 μl microsyringe mounted in the Mosquito LCP, as described previously [86] for convenient delivery of the mesophase into crystallization wells. The repeating dispenser delivers a bolus of mesophase of volume varying from 25-200 nl. Next, LCP crystallization was set up on 24 well glass sandwich plates (Molecular Dimensions) using 1:2 and 1:5 ratios of protein-laden mesophase and precipitant solution, from the commercially available membrane protein screens (half diluted with filtered MilliQ water) using automated Mosquito LCP. As quickly as possible, glass cover slide for the 24 well plate was placed over the wells and sealed to ensure that the diluted bolus didn't get dried. The plates were then incubated at 20°C.

3.2.5 Site-directed mutagenesis of Tk-DGGGPS

The Site Directed mutagenesis primers were designed in accordance to the Quick-change protocol of mutant generation. The primers were complementary having the mutations and amplified the whole plasmid by extending in opposite directions. Different mutants of wild-type Tk-DGGGPS, K¹⁸⁵A, K¹⁸⁸A and K^{185,188}A were prepared using Tk-DGGGPS/pET28a plasmid extracted from transformed DH5alpha cells grown overnight in 10ml LB media in Kanamycin. PCR was performed at 95 °C for 3 min and 18 cycles of (94/ 30 seconds, 55 °C for 30 seconds, 72 °C for 6

min 30 seconds) and final ramping of 72 C for 10 min, using Takara Ex-Taq DNA polymerase. PCR mixture was then treated with DpnI enzyme to degrade the methylated parental DNA and incubated at 37 °C for 3 hrs. 5 µl of PCR mixture is further is transformed in *E. coli* DH5α cells. The mutant plasmids were screened by plating the transformed DH5α cells on Kanamycin LB agar plates. Plasmids were purified and sequenced for confirmation of the desired mutation. List of primers used for the desired mutations are listed below.

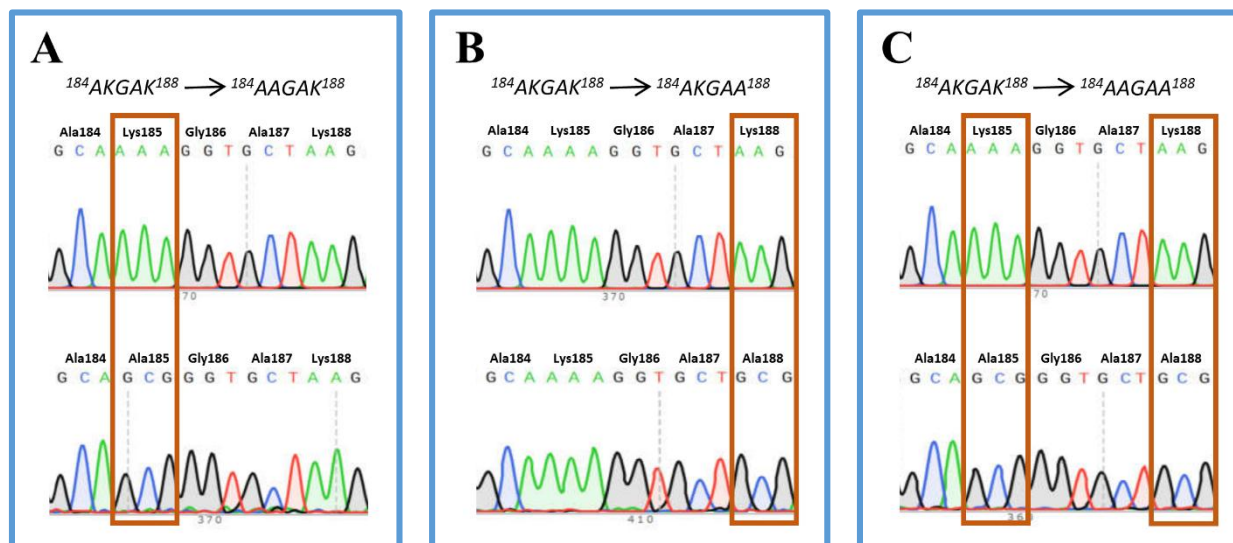


Fig 3.2 Sequence electrogram of site-directed mutagenesis The upper panel is the wild type sequences and the lower panel is the mutant sequence. The brown box highlights the mutation, (A) Mutation K¹⁸⁵A, (B) Mutation K¹⁸⁸A, (C) Mutation K^{185,188}A.

3.2.6 Detergent Screening by Differential Filtration Assay

A detergent quick screening 96 well kit, DSOL-MK 1KT (Qiagen) was used to check the protein stability in different detergents in a high throughput manner based on the principle of differential filtration assay. 200µl of the membrane fraction containing the protein of interest was mixed with 610µl of the Solubilizing Buffer A2+25mM imidazole+1mM DTT and 90µl of the available detergents to obtain final detergent concentration of 1%. The mixture was incubated in a shaker at 1200rpm for an hour at 4°C in order to suspend the membrane protein in solution. It was then spun for an hour at 3000rpm and 4°C. 700µl of the supernatant was transferred to a 96 well filter plate (1.2µm), spun at 3000rpm for 30 minutes at 4°C and the flow-through collected. 200µl of pre-equilibrated Ni-NTA resin slurry was added to the flow-through, shaken at 4°C for an hour at

3000rpm and transferred to a 1.2µm filter plate. The incubated mixture was centrifuged at 3000rpm for 20 minutes at 4°C. This was followed by washing with 500µl of Solubilizing Buffer A2+25mM imidazole + detergent of 2XCMC (twice) and centrifugation at 3000rpm for 15 minutes at 4°C. Next the membrane proteins were eluted out with 200µl of Solubilizing Buffer A2+250mM imidazole+1mM DTT+detergent of 2XCMC and centrifuged at 3000rpm for 15 minutes at 4°C, after which the flow-through was collected. Fluorescence was measured at 280nm to estimate the total protein in the flow-through and then passed through 100kDa MWCO filter plates (Pall Corporation) by centrifuging at 3000rpm for 30 minutes, wherein retentates were considered for analysis. The absorbance at 280nm and 340nm were recorded and its ratio Abs_{340}/Abs_{280} were calculated to final interpretation. A diagrammatic representation of the protocol is given below.

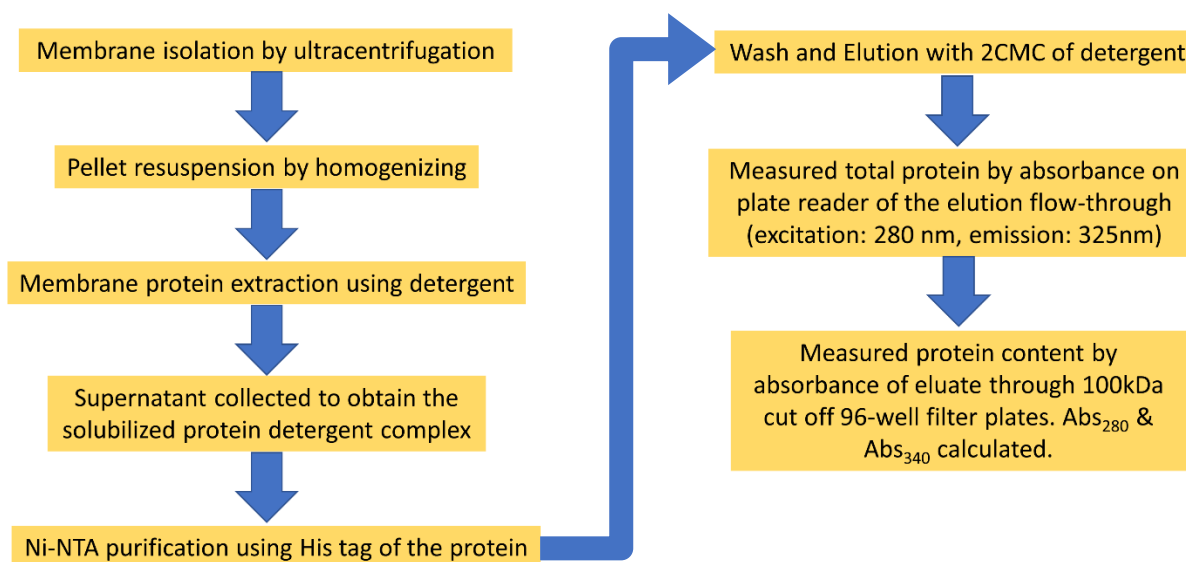


Fig 3.3 Protocol overview for detergent screening of membrane proteins

3.2.7 Protein purification (Mutant Tk-DGGGPS and Tk-DGGGPS in new detergents)

Same protocol was followed in mutant protein purification as in protein purification in Chapter 2 but in smaller batch volumes (5 liter batches). For Tk-DGGGPS purification with new detergents, the detergent extraction was done in 1% (w/v) DDM. Then Ni-NTA purification was done in a buffer containing 0.1% (w/v) CYMAL-6. Initially multiple washes were given with Wash Buffer A9, pH 8.0 to remove the excess DDM and then elution was done using Elution Buffer A10, pH 8.0. This was followed by SEC with Purifying Buffer A11, pH 8.0 to obtain pure protein.

Wash Buffer A9 (pH 8.0)	Elution Buffer A10 (pH 8.0)	Purifying Buffer A11 (pH 8.0)
<ul style="list-style-type: none">• 50 mM Tris• 100 mM NaCl• 30 mM Imidazole• 1 mM DTT• 0.1% (w/v) CYMAL-6• 5% (v/v) glycerol	<ul style="list-style-type: none">• 50 mM Tris• 100 mM NaCl• 300 mM Imidazole• 1 mM DTT• 0.1% (w/v) CYMAL-6• 5% (v/v) glycerol	<ul style="list-style-type: none">• 50 mM Tris• 100 mM NaCl• 1 mM DTT• 0.1% (w/v) CYMAL-6• 5% (v/v) glycerol

3.2.8 Sequence and Phylogenetic Analysis of Tk-DGGGPS

Protein sequence of the TK1957 gene of sequence length 277 amino acids was obtained from the UniProtKB/Swiss-Prot database (<http://www.uniprot.org/>). The sequence was checked against the entire existing UniProt database of bacterial, archaeal, plant and eukaryotic proteins using BLAST, Delta-BLAST and PSI-BLAST for identifying putative homologs of the same, which was further verified by entries in the Pfam database and the MGDB server. DGGGPS sequences from all archaeal species were also annotated separately to identify homologs in all classes of archaea. The homologs were aligned in the software Molecular Evolutionary Genetics Analysis X 10.1 (MEGA) [87], using the algorithm, Multiple Sequence Comparison by Log- Expectation (MUSCLE). The best substitution model to generate the phylogenetic trees were estimated and the model with the lowest BIC scores (Bayesian Information Criterion) was chosen to generate the Maximum Likelihood (ML) tree for the alignments. Initial tree(s) for the heuristic search was obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The base substitution models, LG+G+I+F was used for the archaeal DGGGPS tree and the JTT+G+I+F model for the UbiA superfamily tree. LG and JTT are the amino acid exchange rate matrices, +G stands for the discrete Gamma distribution parameter, +I stands for the number of evolutionarily invariant sites and +F stands for empirical base frequencies from the alignment.

3.2.9 Molecular Modeling of Tk-DGGGPS

To identify templates for homology model building, a similarity search using the Basic Local Alignment Search Tool (BLAST) algorithm [88] was carried out against only Protein Data Bank (PDB) [13] entries to select the high-resolution crystal structures of homologous proteins to the

Tk-DGGGPS. The sequence identity cut-off was set to $\geq 30\%$ (E value cut-off=1). The list of pdb entries obtained were verified using the HHPred server results [89]. Homology modeling was then carried out using the MODELLER 9.16 [90] by using four structures having the highest homology to the target protein, as templates, namely, 4od4, 4od5 (UbiA homolog, 4-hydroxybenzoate octaprenyltransferase from *Aeropyrum pernix* K1) [33] and 4tq3, 4tq4 (UbiA homolog, Bacteriochlorophyll synthase, 33kDa subunit from *Archaeoglobus fulgidus*) [32]. The initial models were first refined in the GalaxyWeb server and then refined using impref minimization of Protein Preparation Wizard of Maestro, Schrödinger [91], followed by minimization and finally loop refining in Prime, Schrödinger [92]. The models obtained after validation and refining were evaluated for the stereochemical quality of the protein backbone and side chains using PROCHECK [93]. Environment of the atoms in the protein model was checked by ERRAT server [94]. Verify3D plot showed the compatibility of the 3-D structure with respect to the protein sequence [95]. These energy minimized final Tk-DGGGPS models were further used for the molecular docking studies with its known ligands.

3.2.10 Molecular Docking of Tk-DGGGPS with its ligands

The ligand molecules considered in the present study were downloaded from Pubchem database in the sdf format. The protein was prepared using protein preparation utility of Maestro, before proceeding with the docking studies. Hydrogens were added subsequently to carry out restrained minimization of the protein model. The minimization was done using impref utility of Maestro in which the heavy atoms were restrained such that the strains generated upon protonation could be relieved. The root mean square deviation (RMSD) of the atomic displacement for terminating the minimization was set as 0.3 \AA . Similarly, the ligands were refined with the help of LigPrep 2.5 to define their charged state and enumerate their stereoisomers. The processed receptors and ligands were further used for the docking studies using Glide 5.8 [96]. A Sitemap analysis [97] was performed on the prepared receptor molecule, to identify the probable binding sites for our ligands of interest, since no prior information was available regarding the same. Next, grids were generated by selecting the best hits among the Sitemap points obtained above. Flexible ligand docking was carried out using the standard precision option. A total of 6 poses with the respective

ligand and different sites were generated and scored on the basis of their docking score, glide score and E-model values. The hydrogen bond interactions between the protein and ligands were visualized using PyMOL. The oligomeric assembly of Tk-DGGGPS was predicted using webservers GALAXYGEMINI and GALAXYHOMOMER [98, 99].

3.2.11 Molecular Dynamic Simulations of Tk-DGGGPS in ether lipid bilayer

First a membrane bilayer was constructed using the Membrane/Bilayer Builder module of the online CHARMM-GUI server [100]. To build the protein/membrane system, the Tk-DGGGPS homology model and ether lipids (DHPCE) were used and the detailed parameters like upper and lower leaflet ratios and experimental surface area were obtained with minor modifications from previous studies (Shinoda et al 2004). First lipids and water thickness (default -17.5Å) were chosen and the protein was inserted into the membrane by the Replacement method. Calcium ions were used for system neutralization and additional concentration and system preparation was performed. After the system is prepared the Tk-DGGGPS inserted in a lipid bilayer membrane was visualized and inspected manually in PyMol for any unusual conformations of both the protein and the lipids. Then the input files were generated for dynamic simulations to be run in GROMACS [101]. These input files were then subjected to multiple rounds of energy minimization (10 ns) to stabilize the protein-lipid system, followed by the final 50ns simulation. The trajectories were visualized using Visual Molecular Dynamics program (VMD) [102]. The RMSD of the protein embedded in the lipid bilayer was monitored for the entire duration of the simulation. The molecular interactions between the protein and lipids were recorded and the interesting presence of water molecules in the central cavity was studied. The Rg (Radius of gyration) was also monitored for the protein to check its stability in the lipid bilayer as well as its overall compactness. RMSF graph of the amino acid residues over the course of the simulation were plotted to find out the dynamic residue patches which were further studied.

3.2.12 Mutational analysis of Tk-DGGGPS homologs

The reported disease mutations of UbiA homologs namely UBIAD1, COQ2 and COX10 were obtained from the Human Gene Mutation Database [103]. Pairwise alignment was done using

WATER algorithm from online server. An attempt was made to study the structural stability changes in Tk-DGGGPS upon single-point mutation at all the amino acid residues using a web server called STRUM [104]. It uses the local structural environment (influenced by main chain conformation, solvent accessibility and hydrogen bonding class) of the wild-type and mutant residues to calculate a stability score, analogous to the free energy difference between the wild-type and mutant protein, and terms it as the pseudo $\Delta\Delta G$. These stability score predictions would be useful for designing site-directed mutagenesis experiments or for predicting whether a mutation would impact protein structure and have a role in disease. A negative score indicates that the mutation is destabilizing whereas a positive score indicates that the mutation is stabilizing.

3.3 Results

3.3.1 Initial crystallization trials of Tk-DGGGPS and Tk-DGGGPS-Bril variants

Purified Tk-DGGGPS protein obtained after size exclusion chromatography, was concentrated to ~10 mg/ml using 50kDa MWCO filters (Amicon ultra 50kDa cutoff centrifugal filters; Millipore, USA) at 4 °C at 4500 rpm for both the wild type and Brill variants respectively. The proteins were screened against several commercially available crystallization screens including, Index (Hampton Research), PEGRx 1 (Hampton Research), MembFac (Hampton Research), Nextal PACT (Qiagen), Nextal Classics Suite (Qiagen) and JCSG plus (Molecular Dimensions) by vapour diffusion method in 96 well MRC-SD2 (MRC2 Sitting Drop plates, Hampton) using different protein to screen ratios of 1:1, 1:2, 1.5:1 and 2:1 in different sitting drop volumes of 300 μ l, 350 μ l and 400 μ l and incubated at 20°C. Initial crystal hits were obtained in the wild type variant in the PACT, MembFac and PEGRx 1 suites while quite surprisingly the Brill variant failed to show prominent crystallization tendencies apart from occasionally forming branched micro-crystals. Tk-DGGGPS wild type crystals were seen to form in two screens of the PACT suite and their compositions were 200mM MgCl₂, 100mM MES (pH 6.0), 20% w/v PEG 6000 and 200mM MgCl₂, 100mM Tris (pH 8.0), 50% w/v PEG 6000 respectively (Fig 3.4 A, B). Tk-DGGGPS crystals were also obtained in the 200mM MgCl₂.6H₂O, 100mM HEPES (pH 7.5), 15% v/v 2-Propanol screen (MembFac suite) and in the 100mM Tris (pH 8.0), 30% w/v PEG monomethyl ether 2000 screen (PEGRx 1 suite). However, the crystal sizes were quite small and hence further

refinement were necessary. But crystal reproduction using manually prepared screens of the same composition was not successful, hence crystallization suites optimized for membrane proteins were used for the next rounds of crystallization attempts.

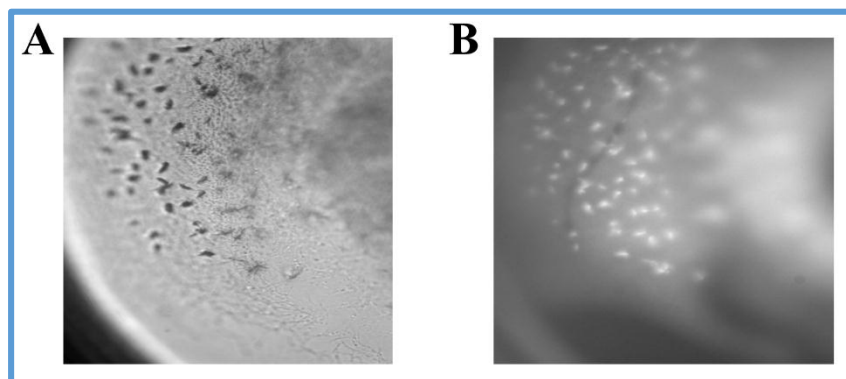


Fig 3.4 Crystals obtained using the PACT suite Crystal images from the MES buffer screen grown at 20°C (A) Bright field images (B) UV images

3.3.2 Crystallization trials of Tk-DGGGPS with advanced membrane protein crystallization screens and additives

Purified Tk-DGGGPS concentrated using the above mentioned method were further subjected to crystallization trials using MEMGOLD and MEMGOLD2 suites (Molecular Dimensions) in 96 well MRC-SD2 plates using different protein to screen ratios of 1:1, 1.5:1 and 2:1 in 400 μ l sitting drops and incubated at 20°C. Positive hits were obtained in the 0.2M MgCl₂.6H₂O, 0.1M KCl, 0.025M Sodium citrate (pH 4.0), 33% v/v PEG 400 and 0.002M ZnCl₂.7H₂O, 0.08M HEPES (pH 7.0), 25% v/v Jeffamine ED 2003 screens of the MEMGOLD suite (Fig 3.5 C) while three screens of MEMGOLD2 also gave crystal hits with compositions of 0.15M CaCl₂; 0.1M Glycine (pH 9.0); 35% v/v PEG 400 , 0.5M MgCl₂; 0.02M LiCl; 0.02M Glycine (pH 10.0); 33% w/v PEG 1000 and 0.1M Sodium cacodylate (pH 6.5); 24% w/v PEG 1500 respectively. Out of the above the best crystals visually were seen in the 0.5M MgCl₂; 0.02M LiCl; 0.02M Glycine (pH 10.0); 33% w/v PEG 1000 and 0.1M Sodium cacodylate (pH 6.5); 24% PEG 1500 screens and they were reproducible manually as well (Fig 3.5 A). However, these crystals were too small to be picked and thereby further optimization was carried out using additive screens, namely Silver Bullets (Hampton Research). Crystals obtained after using Silver Bullets were found to improve in terms of their size and morphology and X-ray diffraction was carried out for the same at the BL21px

beamline, RRCAT Indore (Fig 3.6 B, D). However, they diffracted poorly and only a few diffraction spots were obtained which was not enough to index and for further processing. The diffraction quality remains to be further optimized and micro focusing beam is necessary for these membrane protein crystals due to their proneness towards radiation damage even at cryogenic temperatures.

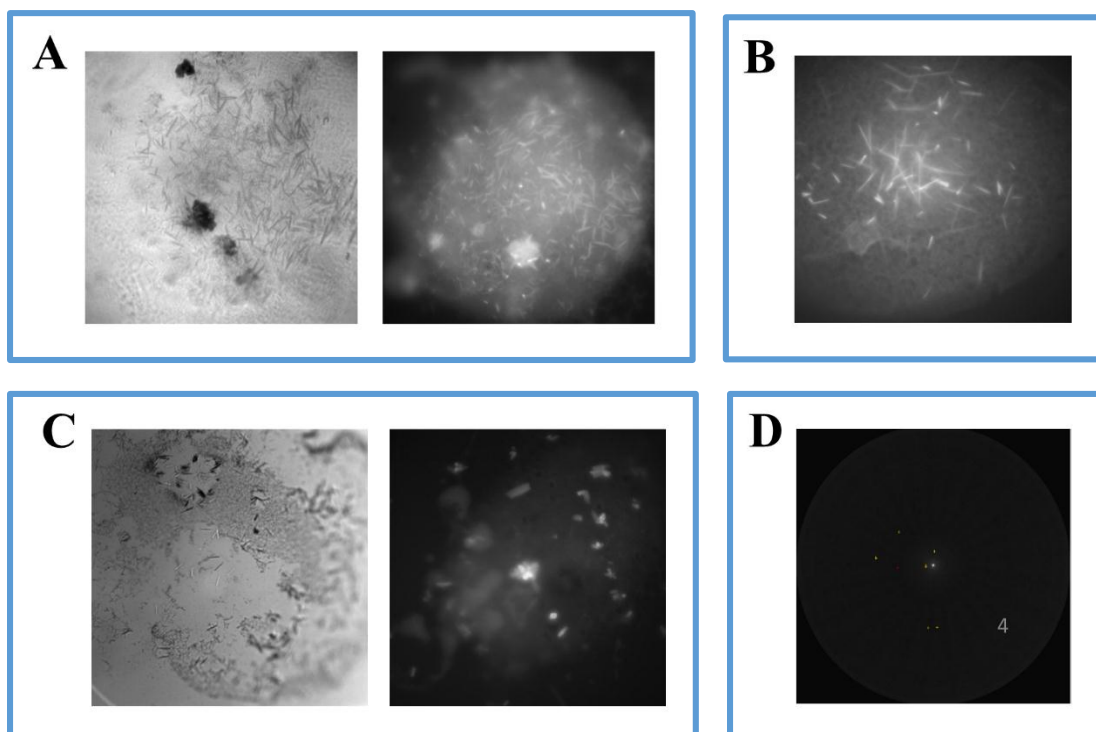


Fig 3.5 Crystals obtained using membrane protein specialized screens (A) Bright field and UV images of crystals from the screen with the best hit in the MEMGOLD 2 suite, (B) UV images of crystals of 'A' optimized using SB additive screen, (C) Bright field and UV images of crystals from the citrate buffer screen in the MEMGOLD suite, (D) Diffraction spots obtained after X-ray diffraction of the crystals picked from 'B'.

3.3.3 Crystal optimization of Tk-DGGGPS by HILIDE and LCP methods

To obtain better quality crystals large enough for diffraction analysis, attempts were made to perform crystallization using two alternative methods. In the first method, crystallization was carried out in high concentrations of lipid and detergent (HILIDE) which centers on a re-lipidation step with relatively high concentrations of lipid(s) and detergent(s) prior to the crystallization process. Two different concentrations of the archaeal total lipid extract from *S. acidocaldarius* and *H. volcanii* (Saci_lipids and Hvo_lipids) were used for the experiment. In the

second method, crystallization was carried out using a membrane mimetic medium known as the lipidic cubic phase (LCP) where crystallization takes place in lipidic mesophases. The lipid used for the experiment was monoolein, 9.9 MAG (monounsaturated monoacylglycerol lipids; the acyl chain contains 9 carbon atoms between the ester and *cis*-olefinic bonds and 9 carbon atoms between the *cis*-olefinic bond and the end of the chain). The reproducible crystal conditions from the previous experiment was used in both cases and the HILIDE method produced crystals which were bigger in size than the ones produced earlier (Fig 3.6 A, B). However, no crystals were observed in the LCP plates, which indicates that further optimization is necessary for the same. The crystals obtained using the HILIDE method remain to be diffracted.

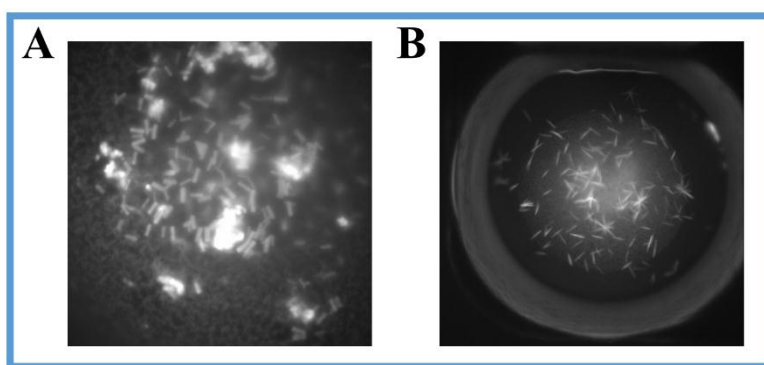


Fig 3.6 Crystals obtained using the HILIDE method The sodium cacodylate buffer screen from MEMGOLD2 suite was used as the base screen to which the archaeal total lipid extracts were added (A) UV images crystals using the Saci_lipids, (B) UV images crystals using the Hvo_lipids.

3.3.4 Tk-DGGGPS mutant purification and crystallization

In every protein there are a few entropically dynamic residues especially on the surface, which may or may not be essential for the structural and functional integrity of the protein. However due to presence of such residues, crystal formation is often hampered due to interference in the crystal lattices and lack of stable crystal contacts. In order to screen such residues in our protein based on available crystal data and thermodynamic calculations, the Surface Entropy Reduction prediction (SERp- <https://services.mbi.ucla.edu/SER/>) server [105] was used. This tool attempts to help in identification of sites that are most suitable for mutations designed to enhance crystallizability of the protein by a Surface Entropy Reduction approach, wherein it replaces flexible, solvent-exposed residues with residues having low conformational entropy. As per the prediction of the SERp server, K185 and K188 were found to be the most potent sites for mutations (SERp score: 3.87) in

the cluster $^{184}\text{AKGAK}^{188}$ which would predictably not hamper the structural stability of the protein, yet make its crystallizability more thermodynamically viable. Generally, cluster scores of 3.0 and above are good candidates for mutation studies. Site Directed Mutagenesis was hence performed on Tk-DGGGPS to generate single mutants K^{185}A and K^{188}A and a double mutant $\text{K}^{185,188}\text{A}$.

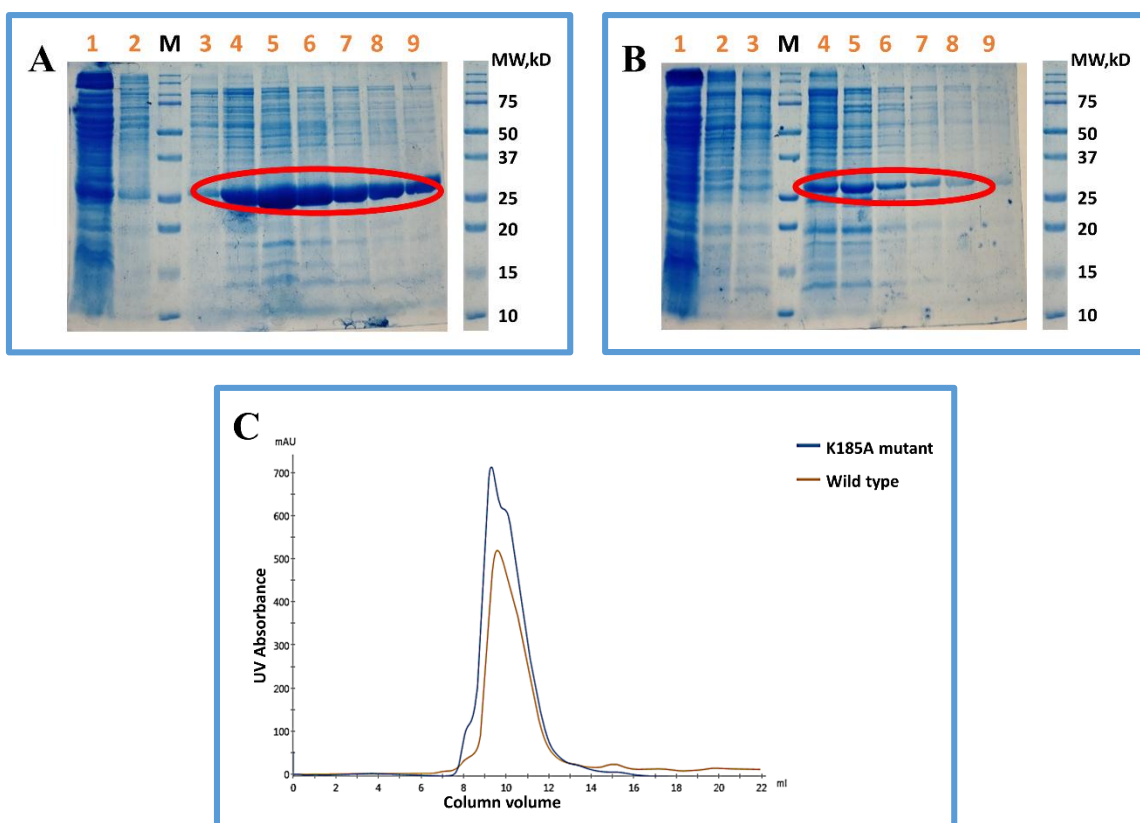


Fig 3.7 A comparison of the yields of the mutants and wild type Tk-DGGGPS (A) SDS-PAGE profile of the K^{185}A mutant expression. Lane 1: Unbound fraction; Lane 2: Wash fractions; Lane 3-9: Elution fractions; Lane M: Dual protein marker (B) SDS-PAGE profile of the K^{188}A mutant expression. Lane 1: Unbound fraction; Lane 2: Wash fractions; Lane 3-9: Elution fractions; Lane M: Dual protein marker (C) A comparison of the yield of the wild type Tk-DGGGPS and the K^{185}A mutant by its SEC profile showing the distinctly high yield which becomes even more relevant since the wild type yield was from a 20-liter batch culture while the mutant was only from a 5-liter batch culture.

Attempts at purification of the Tk-DGGGPS mutants resulted in an interesting outcome. The K^{185}A mutant was found to have the highest expression among all the mutants (Fig 3.7 A, B) and it was seen that it gave almost four fold increased yield than the Tk-DGGGPS wild type and the Tk-DGGGPS Brill variant which indicates that the mutation had in some prominent way influenced the overall stability of the enzyme (Fig 3.7 C). The mutant protein was also found to show

enzymatic activity which indicates that the mutation doesn't hamper the functional motifs in the enzyme. The Size exclusion chromatography profile of the K¹⁸⁵A mutant was found to be similar to the wild-type protein. The K¹⁸⁸A mutant also showed a prominent yield which was slightly more in comparison to the wild type Tk-DGGGPS while the double mutant K^{185,188}A showed very less yield. The K¹⁸⁸A mutant however showed a proneness towards precipitation and was thereby considered unsuitable for further studies while the low yield of the double mutant defeated the purpose of the mutational studies. Thereby among the three mutants the K¹⁸⁵A mutant was found to be most suitable for further crystallization studies.

Anapoe-58	Fos-Choline-16	Tetradecyl-β-D-Maltopyranoside	Anapoe-80	Tridecyl-β-D-Maltopyranoside	Dodecyl-β-D-Thiomaltop	Fos-Choline-15	Anapoe-C12E9	Fos-Choline-14	Anapoe-C12E8	e Glycol Monododecyl Ether	Anapoe-C13E8
Anzergent 3-14	Anapoe-20	n-Tetradecyl-N,N-Dimethylami	n-Dodecyl-α-D-	n-Dodecyl-β-D-	CYMAL-7	Anapoe-35	Anapoe-X-114	n-Decyl-β-D-Thiomaltopyranoside	n-Undecyl-β-D-Thiomaltopyra	Anapoe-C12E10	Anapoe-X-100
Sucrose Monododecanoate	Cyclofos-7	n-Dodecyl-N,N-Dimethylami	Fos-Choline-13	CYMAL-6	n-Undecyl-α-D-	n-Undecyl-β-D-	Pentaethylen e Glycol Monodecyl	Anapoe-C10E6	n-Dodecyl-N,N-Dimethylglycin e	DHPC	Fos-Choline-12
Fos-Choline-11	Anapoe-C10E9	n-Decyl-α-D-Maltopyranoside	n-Decyl-β-D-Maltopyran	n-Nonyl-β-D-Thiomaltopyranoside	Anzergent 3-12	Cyclofos-6	Big CHAP, Deoxy	CYMAL-5	Cyclofos-5	Fos-Mea-10	Anapoe-X-405
Nonyl-β-D-Glucopyranoside	Fos-Choline-Unsat-11-10	Deoxycholic Acid, Sodium Salt	Big CHAP	e Glycol Monoethyl Ether (C8E4)	Nonyl-β-D-Maltopyranoside	n-Octyl-β-D-Thiomaltopyranoside	Fos-Choline-10	CYMAL-4	Glycol Monoethyl Ether (C8E6)	Cholic Acid, Sodium Salt	Sodium Dodecanoyl Sarcosine
Cyclofos-4	n-Decyl-N,N-Dimethylglycine	CHAPS	CHAPSO	n-Octyl-β-D-Glucopyranoside	Fos-Mea-8	Anameg-7	Heptyl-β-D-Thioglucopyranoside	CYGLU-3	n-Octyl-β-D-Maltopyranoside	Fos-Choline-ISO-11	Fos-Choline-ISO-9
Anapoe-X-305	Anzergent 3-10	4-Heptyl-β-D-Maltopyranoside	Fos-Choline-9	Cyclofos-3	CYMAL-3	Heptyl-β-D-Glucopyranoside	Pentyl Maltopyranoside	MEGA-8	Fos-Choline-8	CYMAL-2	Hexyl-β-D-Glucopyranoside
Cyclofos-2	n-Hexyl-β-D-Maltopyran	Anzergent 3-8	CYMAL-1	PMAL-C8							

Fig 3.8 The entire set of detergents used for the high throughput detergent screening with the best hits highlighted.

3.3.5 Screening detergents for the optimal PDC (Protein Detergent complex)

Even though DDM is one of the most successful detergents to be used for membrane protein purification and crystallization till date based on the number of membrane protein structures available in PDB, the higher micelle size of the detergent remains the major bottleneck for proteins of lower molecular masses, wherein a compact arrangement of the protein molecules might be hampered due to the bigger micellar size. So to check, if other detergents were better suited for both the protein stability and crystallizability of Tk-DGGGPS, high throughput detergent screening was performed based on the basic principle of differential filtration wherein an observed

variation in the amount of protein in filtrates or retentates after parallel passage of protein in buffer containing different detergents through different MWCO filters provides information on stability (aggregation) and size of the PDC.

All the detergents in the 96 well suite (DSOL-MK 1 KT; Qiagen) used for the detergent screening had molecular weights above 100kD, hence only the retentate absorbance readings at 280nm and 340nm were considered. The Abs_{340}/Abs_{280} ratio was calculated for all the absorbance readings and the ones found to remain constant over the period Day 1 – Day 4 would be considered for further studies. Since absorbance at 340 nm is known to be predominantly due to protein aggregates, so the lower ratios were considered as positive hits. Among the positive hits, the detergents CYMAL-5 and CYMAL-6 were found suitable for further purification based on their properties (Fig 3.8). CYMAL-5 and CYMAL-6 are nothing but maltoside detergents with cyclohexyl aliphatic tail having distinctly low micellar size compared to DDM which might help in proper packing and orientation of the protein in the PDC.

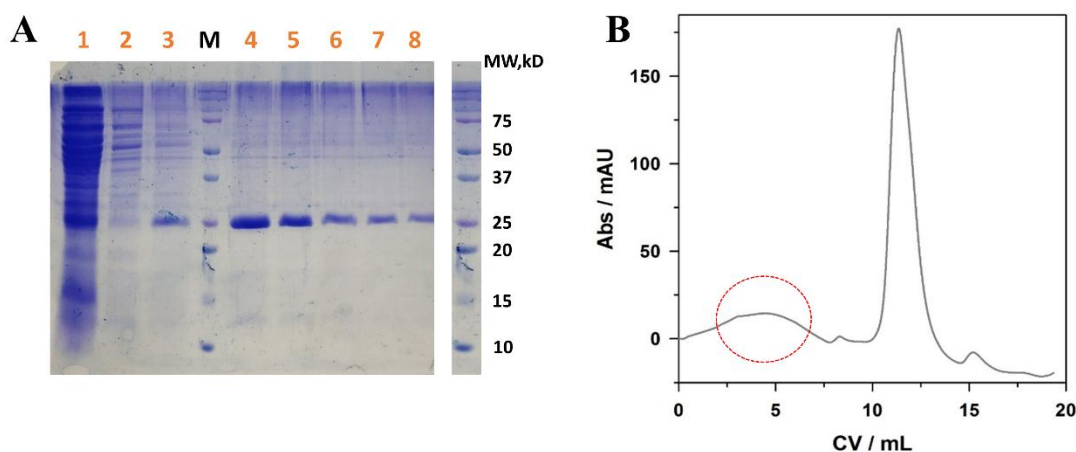


Fig 3.9 Tk-DGGGPS purified in CYMAL-6 detergent (A) SDS-PAGE profile of the CYMAL-6 purification of wild type Tk-DGGGPS. Lane 1: Unbound fraction; Lane 2: Wash fraction; Lane 3-8: Elution fractions; Lane M: Dual protein marker, (B) SEC profile of Tk-DGGGPS in CYMAL-6 showing a small peak of aggregation which could be the result of improper detergent exchange between DDM and CYMAL-6 and a sharp peak for the purified Tk-DGGGPS-CYMAL-6 PDC.

3.3.6 Tk-DGGGPS purification in new detergents with low micellar size

CYMAL-5 has the smallest micelle among CYMAL-5 and CYMAL-6 at 23 kD, however due to its relatively high cmc of ~ 0.12 , CYMAL-6 with a lower cmc of ~ 0.028 (0.56 mM) and a small

micelle size of 32 kDa was considered to be the first detergent of choice for further purification studies.

The detergent exchange with 0.1% w/v CYMAL-6 was carried out after detergent extraction with 1% w/v DDM and a Size exclusion chromatography (SEC) was carried out to check the behavior and stability of the Tk-DGGGPS embedded in the detergent micelle (Fig 3.9 A, B). The SEC profile of Tk-DGGGPS/CYMAL-6 showed a prominent peak which eluted at a similar position when compared to the Tk-DGGGPS/DDM profile, however it also shows presence of different oligomeric states/aggregation which could be due to the possible formation of higher molecular weight functional oligomers due to the lower micelle size of CYMAL-6. Or it could also indicate towards an improper detergent exchange between DDM and CYMAL-6 due to the higher micellar size of DDM. Hence further optimization is necessary to determine if a different detergent would be ideal for this protein to be stable in solution.

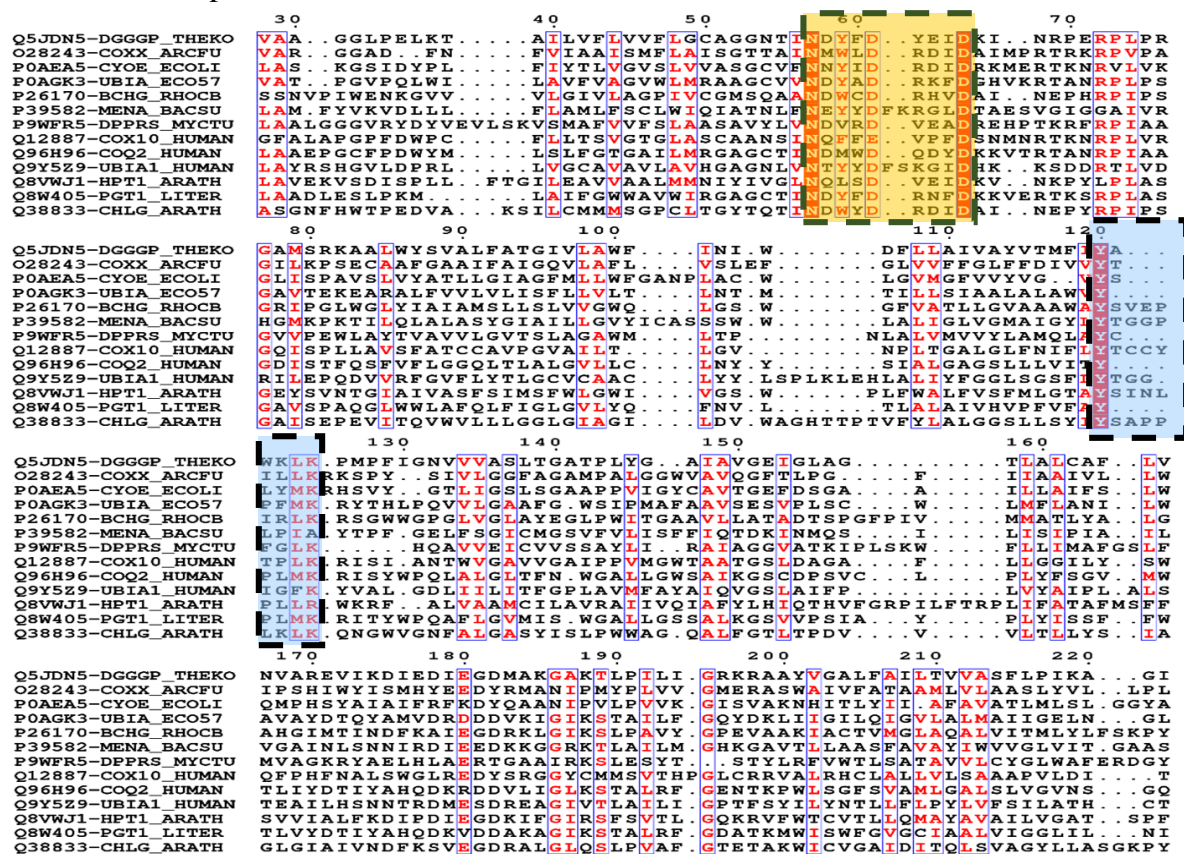


Fig 3.10 Multiple sequence alignment of the representative sequences of the UbiA superfamily, homologous to DGGGPS. The notably conserved *NxxxDxxxxD* residue cluster and the *YxxxxxL/MK/R* cluster are highlighted in orange and blue respectively.

3.3.7 Sequence analysis of DGGGPS and other UbiA superfamily members

In most family of enzymes, the core structure is generally found to be distinctly conserved, which is only possible due to the presence of identical or similar amino acid residues in the specific functional domains of the protein. So to understand the structure-function dynamics, it is important to first compare the sequences of the members of the protein family to identify those conserved residues which can be a part of the enzyme binding site. Also, due to the uniqueness of DGGGPS among all the other members of the UbiA prenyltransferase superfamily in terms of its substrate sensitivity, it was interesting to study its amino acid sequence. Differentially conserved residues obtained by sequence alignment between these prenyltransferases could possibly give an insight into their roles in the respective enzyme mechanisms.

The sequences obtained from a Domain Enhanced Lookup Time Accelerated BLAST (DELTA-BLAST) and Position-Specific Iterated BLAST (PSI-BLAST) against Tk-DGGGPS (Uniprot ID: Q5JDN5), further annotated from Pfam database [106] and the Microbial Genome Data Base web server (<http://mbgd.genome.ad.jp/>), gave us an interesting array of putative homologs from a vast variety of organisms ranging from archaea to humans. In plants it included homogentisate geranylgeranyltransferases, homogentisate phytyltransferases, glycinol 4-dimethylallyl transferases, naringenin 8-dimethylallyltransferases and chlorophyll synthases, a varied cluster of enzymes, all functionally distinct but evolutionarily linked. In bacteria bacteriochlorophyll synthases (BChG), protoheme IX farnesyltransferases (COXX), 1,4-dihydroxy-2-naphthoate octaprenyltransferases (MenA) and 4-hydroxybenzoate octaprenyltransferases (UbiA), 4-hydroxybenzoate solanesyltransferase and homogentisate phytyltransferase are the various homologs that are functionally different but have distinct sequence similarities. In archaea, apart from other Di-O-geranylgeranylglyceryl phosphate synthases (DGGGPS), protoheme IX farnesyltransferases (COXX) and lycopene elongase were indicated as possible homologs. In humans the mitochondrial isoform (Isoform 3) of 4-hydroxybenzoate polyprenyltransferase (COQ2), mitochondrial protoheme IX farnesyltransferase (COX10) and UbiA prenyltransferase domain-containing protein (UBIAD1) were identified as homologs.

Multiple sequence alignment of the representative homologous sequences of the UbiA superfamily highlights a few residues which were entirely conserved and some residues which were highly

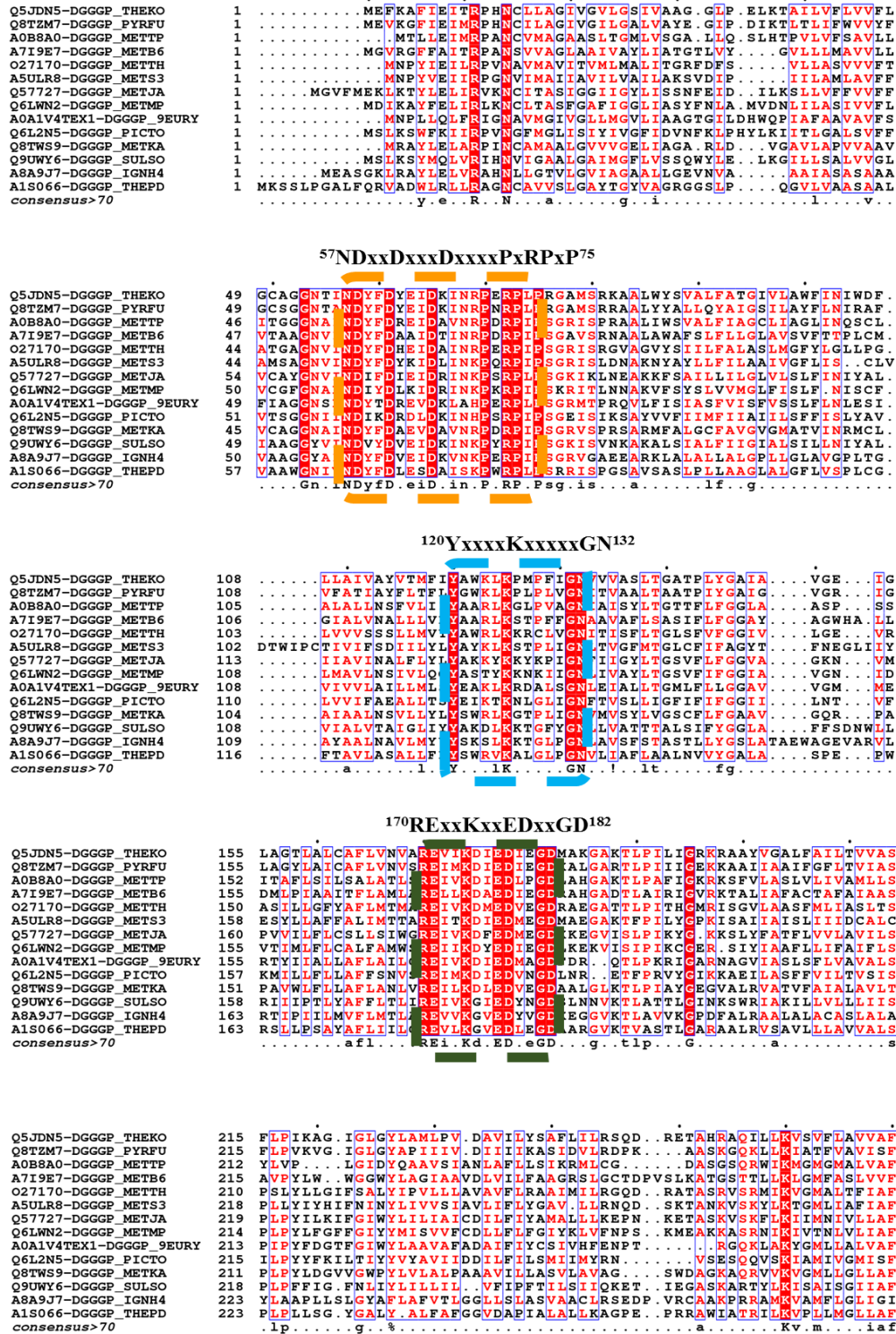


Fig 3.11 Multiple sequence alignment of DGGGPS sequences from different classes of archaea. The conserved patches are highlighted in dotted boxes. The *RExxKxxEDxxGD* motif highlighted in a green dotted box is unique from the motifs obtained in all UbiA members and shows more prominent conservation among DGGGPS.

conserved. These residues were seen to form distinct clusters out of which two were most prominent (Fig 3.10). One was the already reported $NxxxDxxxxxD$ motif which is a predicted functional motif in prenyltransferases, and the other was the $YxxxxxxL/MK/R$ motif which contained conserved residues located in the central cavity of the protein, where the prenyl transfer reaction is predicted to occur. A multiple sequence alignment of representative sequences of DGGGPS from each taxonomic class of Archaea (Fig 3.11), on the basis of protein homology with the Tk-DGGGPS, further annotated on basis of the order they belong to showed conserved motifs similar to the above mentioned ones $^{57}NDxxDxxxDxxxxPxRPxP^{75}$ and $^{120}YxxxxKxxxxGN^{132}$, respectively. In addition, another distinct cluster $^{170}RExxKxxEDxxGD^{182}$ was also visualized. The significance of these clusters were checked using the CONSURF analysis [107] which predicted all of them to be highly conserved structural and functional residues (Fig 3.12).

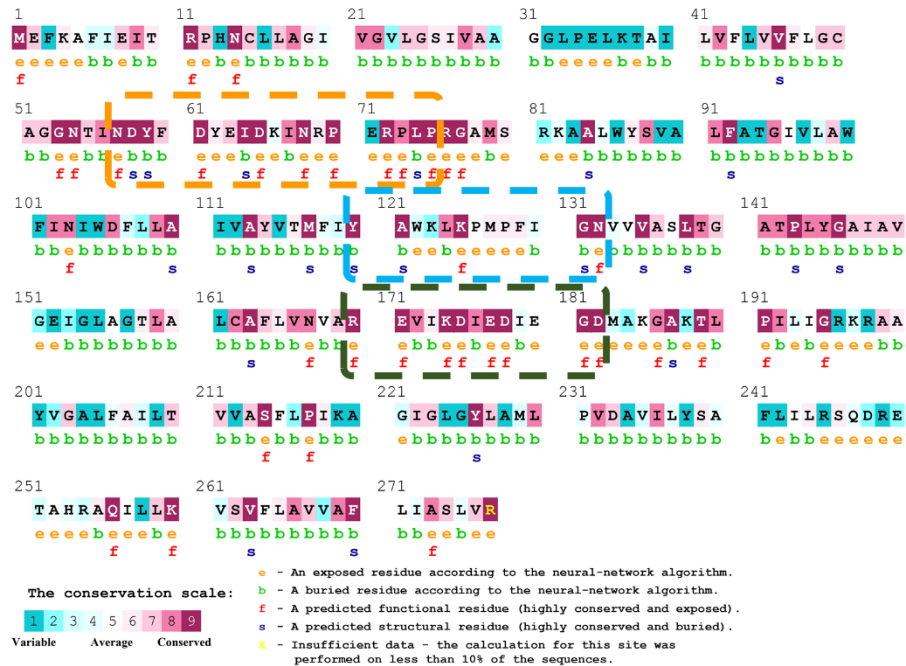


Fig 3.12 Consurf sequence conservation diagram of Tk-DGGGPS. The $NDxxDxxxDxxxxPxRPxP$, the $YxxxxKxxxxGN$ and the $RExxKxxEDxxGD$ residue clusters highlighted in orange, blue and green dotted boxes respectively, as in Fig 3.11, contain residues with high conservation scores and happen to be predicted functionally and structurally relevant residues as well.

The Tk-DGGGPS was further analyzed using the HHPRED webserver (<https://toolkit.tuebingen.mpg.de/tools/hhpred>) which employs sequence database searching on basis of pairwise comparison of profile hidden Markov models (HMMs) and verifies it with the

predicted secondary structure similarity, available from protein crystal structure data in PDB to check if it has any remote homologs whose sequence identity is even below 20%. The analysis results showed a homology with a sodium/glucose cotransporter (PDB: 2XQ2) from *Vibrio parahaemolyticus* [108] with a probability of 71.43% and a negative E-value (Fig 3.14). A normal protein BLAST of the Tk-DGGGPS sequence against the *Sulfolobus acidocaldarius* putative proteome also gives a similar hit against a major facilitator superfamily (MFS) transporter (the most commonly found transporter protein in Archaea) with a sequence identity of 31.91 % and a query coverage of 99%, further verified in the entire Sulfolobus genus, indicating that DGGGPS might also be involved in a secondary ion/sugar regulating pathway in certain classes of Archaea.

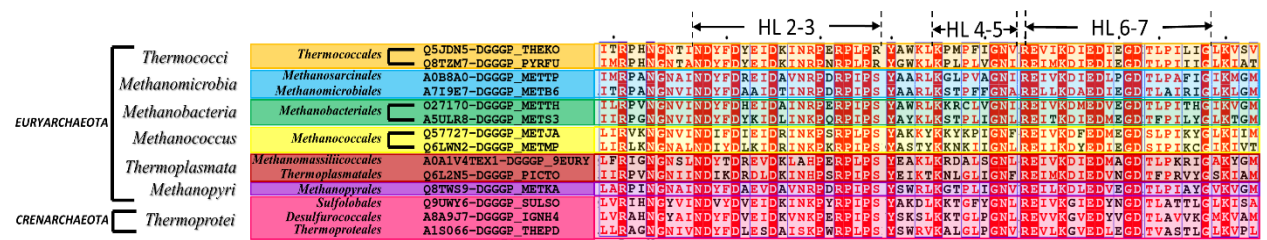


Fig 3.13 The sequence conservations of DGGGPS arranged according to archaeal taxonomic classes. The conserved residue clusters, mentioned in Fig 3.11, are located in HL regions between the TM helices of the protein.



Fig 3.14 The HHpred sequence conservation diagram of the sodium glucose transporter from *Vibrio parahaemolyticus* (PDB: 2XQ2). The notably conserved regions are shown here. The color code is based on physicochemical properties, wherein positively charged residues are presented in red, negative charged in blue and aliphatic residues in green. The *ss* refers to secondary structure prediction by PSIPRED and using DSSP from known structures.

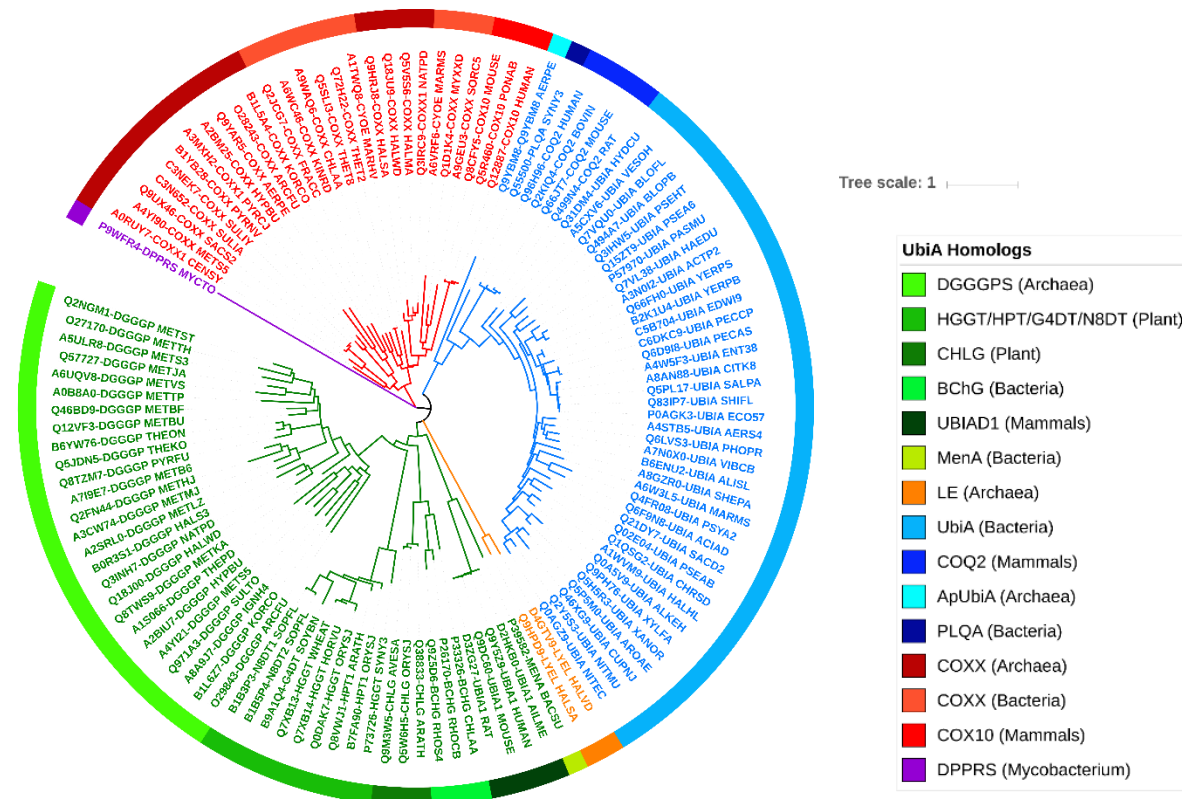


Fig 3.15 Phylogenetic tree of all the UbiA homologs of Tk-DGGGPS Generated using the MEGA suite, this tree shows the distinct grouping of DGGGPS with all plant UbiA prenyltransferases and respiratory quinone synthases in bacteria and humans. The color code is based on the organism of origin and the involvement in different pathways.

3.3.8 Phylogenetic analysis of DGGGPS and other UbiA superfamily members

A phylogenetic tree obtained using the above MSA of the representative homologous sequences of the UbiA superfamily, divided the total number of sequences into five distinct clusters (Fig 3.15). The tree showed a common ancestry of DGGGPS, UbiA superfamily representatives in plants (notably the Chlorophyll synthases and the homologs involved in the secondary metabolite phytoalexins, vitamin E, sophoraflavanone G synthesis) and mammalian UBIAD1, which is quite unexpected and noteworthy considering the fact that they belong to three different species very distant in the general evolutionary tree. This gave us insights into the vast functional diversity of the members of the UbiA superfamily. Out of the five distinct evolutionary groups, there are three major groups, A, B & C. Group A includes the archaeal DGGGPS, the entire plant homologous protein faction (CHLG, HGGT, HPT, N8DT), the bacterial BChG homologs, the bacterial MenA & its mammalian homologs UBIAD1. Group B includes all the bacterial variants of

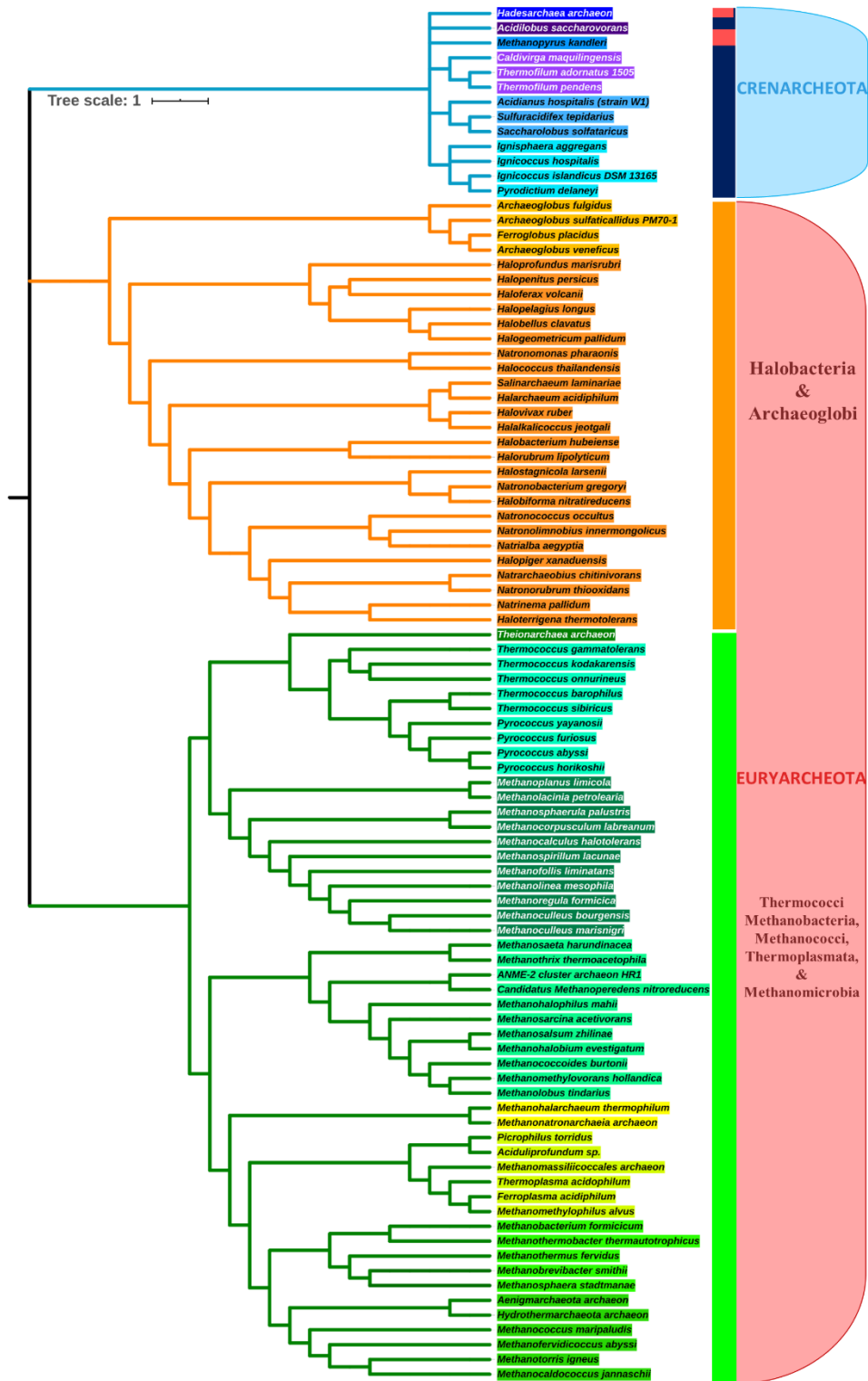


Fig 3.16 Phylogenetic tree of DGGGPS from different archaeal taxonomic classes Generated using the MEGA suite this tree shows that DGGGPS are present in almost all known classes of archaea with prominent segregation of the two phyla, Crenarchaeotes and Euryarchaeotes.

4-hydroxybenzoate octaprenyltransferases (UbiA) and the mammalian variants of 4-hydroxybenzoate polyprenyltransferases (COQ2), Group C includes the bacterial archaeal and mammalian variants of protoheme IX farnesyltransferases (CyoE, CtaB & COX10). The two minor groups are composed of the mycobacterial DPPR synthases and the lycopene elongases, respectively.

Another phylogenetic tree was constructed using the MSA of archaeal DGGGPS sequences from different taxonomic classes. It showed the putative DGGGPS sequences to be clustered according to the main archaeal orders and form three distinct clades (Fig 3.16). The blue clade was composed of predominantly Crenarcheotes with the only exception of *Hadesarchaea archeon* and *Methanopyrus kandleri*. The green and orange clade were composed of Euryarchaeotes entirely indicating that the DGGGPS in the two major kingdoms of archaea evolved distinctly and therefore might have distinctive features of their own like the one mentioned previously where DGGGPS in the *Sulfolobus* genus was found to have high sequence similarity with MFS transporters. Among the Euryarchaeotes, the orange clade composed of DGGGPS from classes Halobacteria and Archaeoglobi where the former were found to have the most abundant representatives and were more prominently clustered indicating higher homology between them. The green clade featured members of different classes of Euryarchaeotes and were clustered according to their orders out of which there were two prominent divisions, one composed of orders Thermococcales and Methanomicrobiales and the other composed of orders Methanobacteriales, Methanococcales, Thermoplasmatales, Methanonatronarchaeales and Methanosarcinales. The prominent clusters indicated that DGGGPS were highly conserved in respective classes of Archaea and thereby could be an important member in the Archaeal proteome.

3.3.9 Structure prediction and Homology modelling of Tk-DGGGPS

Due to many unsuccessful attempts on our effort towards X-ray crystallographic structure determination of Tk-DGGGP we proceeded with the homology modelling. Currently there are no structures available for archaeal DGGGPS in PDB. Thereby to study the distinct structural features of the enzyme, Tk-DGGGPS was modelled in MODELLER ver.9.17 [90] using a UbiA homolog apo structure (4OD4) from *A. pernix* and a UbiA homolog apo structure (4TQ5) from *A. fuldigus*

as templates. The model was further refined in Galaxy web server and verified in PROCHECK and ERRAT servers [93, 94]. The Ramachandran plot of the final model showed 93.8% of the amino acid residues to be in the most favored regions, 5.8% residues in additional allowed regions and 0.4% residues in disallowed regions. The overall quality factor as determined by ERRAT for the homology model was 100% indicating that the model was properly refined and there were no unusually close non-bonded contacts (Fig 3.18 A, C). The RMSD of the Tk-DGGGPS structure with the templates were 4.46 and 1.97 respectively.

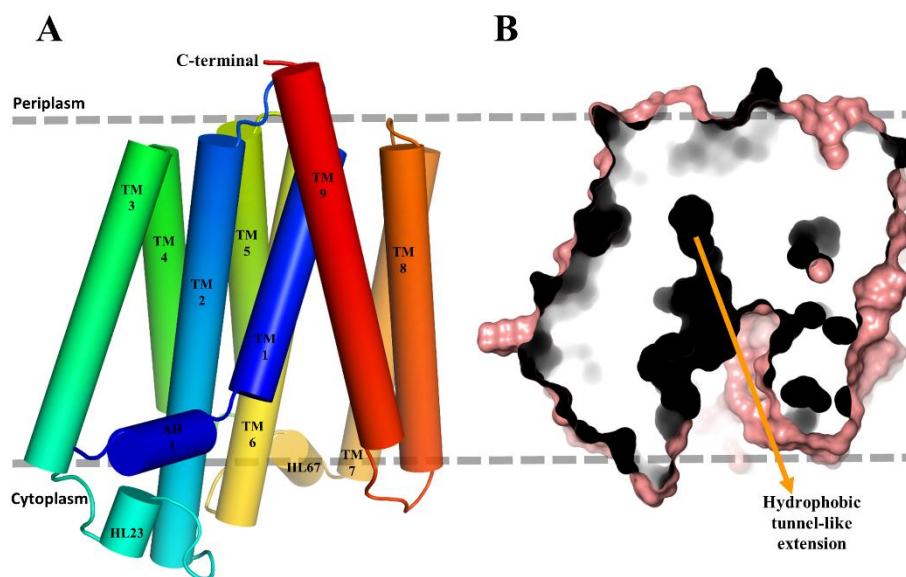


Fig 3.17 Overall structure of Tk-DGGGPS (A) Cartoon pipes & planks representation depicting the nine transmembrane helices of the enzyme (TM1-TM9) and an amphipathic helix AH1 connected to TM1 by a proline kink; Helix-loop regions connecting the TMs HL23 and HL67 are also shown which lie towards the cytoplasmic opening whereas only loops are present between the TMs at the peripheral end, (B) A vertical cross section of the molecular surface representation of the protein structure showing the prominent extended central cavity with a broad opening at the cytoplasmic end; This central cavity hosts most of the prominent conserved residues and is highly functionally relevant.

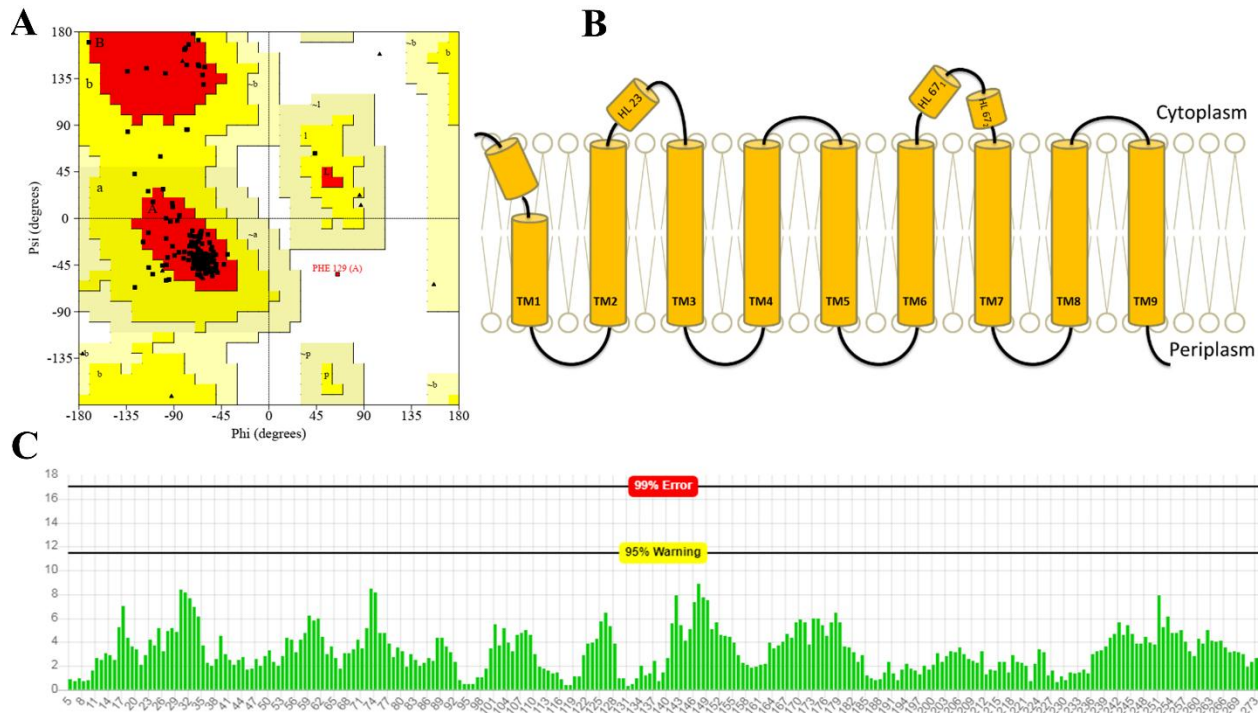


Fig 3.18 Particulars of the Tk-DGGGPS homology model (A) Ramachandran plot (B) Schematic diagram of the structural components and their arrangement in the lipid bilayer (C) The ERRAT quality score plot showing the homology model to be energetically compliant without major steric clashes among residues.

The N-terminal residue of the enzyme was predicted to be buried inside the membrane and formed a component of a kinked transmembrane helix (TM helix), which was also the first of nine transmembrane helices of Tk-DGGGPS. The kinked TM helix was composed of a predicted short amphipathic helix AH1 (Phe3-Thr10) and TM1 (His13-Ala29) with a kink at Arg11 and Pro 12 among which Arg11 was found to be highly conserved in all archaeal DGGGPS. TM1 was connected to TM2 (Leu36-Phe60) via a short periplasmic loop. The TM2 and TM3 (Arg81-Trp100) was connected by a stretch known as the helix-loop (HL) region (Fig 3.17 A, 3.18 B). The HL regions are a characteristic of this superfamily of enzymes. TM2 and HL23 were found to contain residues which are a part of the $^{57}\text{NDXXDXXXDXXXXPXRXP}^{75}$ conserved domain of DGGGPS, as mentioned before. Due to the presence of three conserved proline residues namely Pro 70,73 and 75, HL23 might be a region with important roles in the binding of the prenyl donor and acceptor to the enzyme during its catalytic activity. Another residue present in the HL23 region, Asp65, was predicted to be a catalytic residue (Ap-UbiA; 4OD4) and also was found to be entirely conserved in all DGGGPS further highlighting the importance of these HL regions. The

TM3 and TM4 (Ile104-Lys125) were again connected by a very short periplasmic loop. TM4 was found to have three conserved residues in Lys125, Gly131 and Asn132 among which Lys125 is a predicted functional residue as well. TM5 (Ile130-Val150) was found to be parallel to TM4 and were connected by a short cytoplasmic loop. TM5 and TM6 (Ile153-Glu177) is separated by a three residue loop located in the periplasmic side of the membrane. TM6 was found to have conserved residues belonging to the domain $^{170}\text{REXXKXXEDXXGD}^{182}$, namely Arg170, Glu171, Lys174 and Glu177 all of which were predicted functionally relevant residues. The TM6 and TM7 (Lys197-Lys219) was connected by a cytoplasmic HL67 as well which included conserved residues Gly181 and Asp182 among which Asp182 is a predicted catalytic residue (AfUbiA; 4TQ5). TM7 was further connected to TM8 (Leu224-Leu244) by a short periplasmic loop while TM8 was linked to TM9 (Thr251-Val276) by a cytoplasmic loop. The C terminus of the protein extends into the periplasm till Arg277 (Fig 3.17 A). Most of the TM domains were found to be nearly parallel to each other with a slight angle towards the center, thereby enclosing a central cavity where the predicted active site pocket was located. TM1, TM2, TM4, TM5 and TM6 appear to be the helices which forms the boundary of the hydrophobic core of the enzyme along with the HL regions and thereby have the most conserved residues. The other TMs make up the external morphology of the Tk-DGGGPS (Fig 3.19).

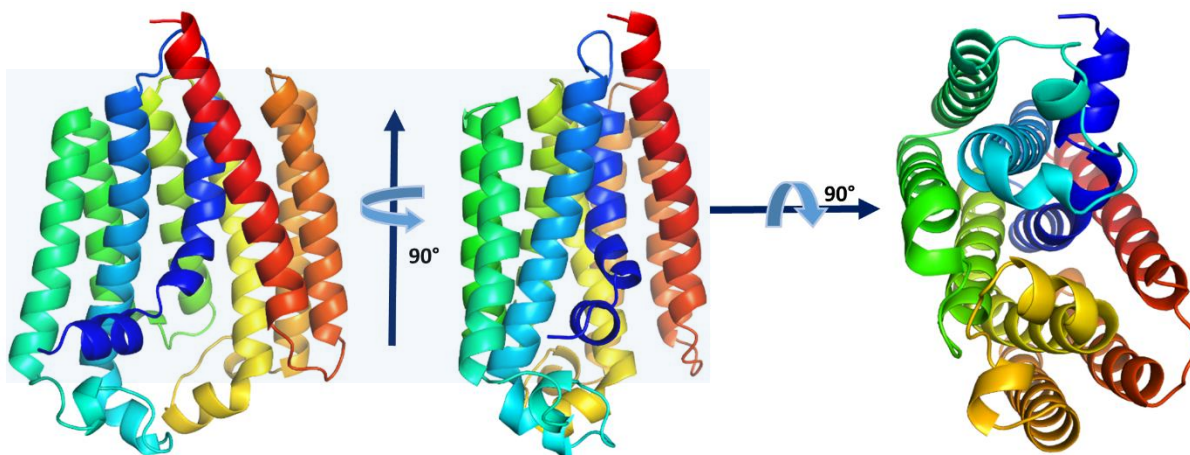
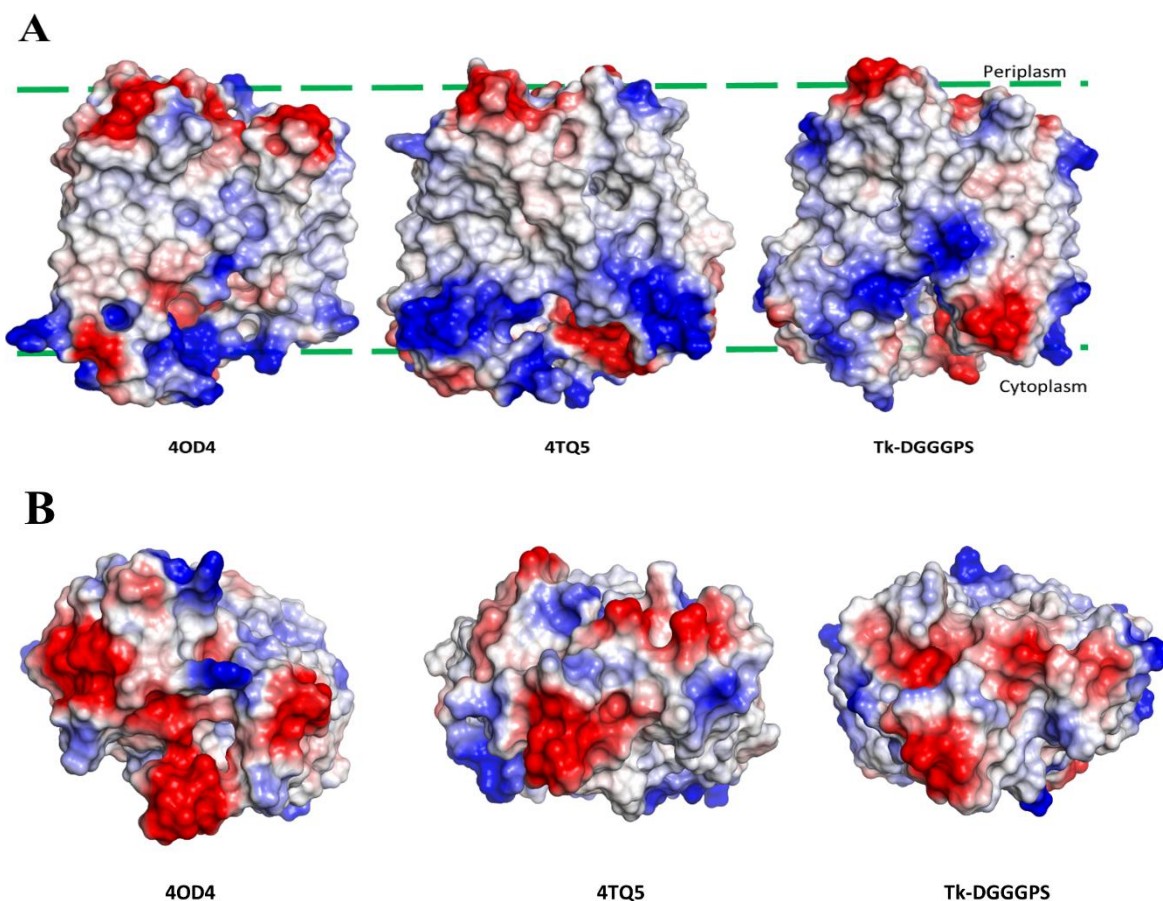


Fig 3.19 Overview of the Tk-DGGGPS structure Cartoon representation of the structure, color-ramped from the N to C terminus (blue to red) and viewed as it is embedded in the membrane, at a rotation of 90° in the same plane and from the top of the membrane.

Electrostatic surface potential was mapped for Tk-DGGGPS and distinct charged patches were found to be buried near the membrane region which was notably less than that observed in case of Ap-UbiA [33] and Af-UbiA [32] (Fig 3.20). However, in all the three structures negative electrostatic patches were more prominent in their periplasmic side in comparison. The charge distribution on the frontal face of Tk-DGGGPS, Ap-UbiA and Af-UbiA were quite distinct with prominent positive patches in respective structures (Fig 3.20 A). However, the periplasmic cap region of the protein showed similar charge distribution in the three structures (Fig 3.20 B). The cytoplasmic cap had distinct differences in charge distribution, where a positive electrostatic patch was observed in case of Tk-DGGGPS and both ApUbiA and AfUbiA were observed to have negative patches (Fig 3.20 C). The Tk-DGGGPS appeared to have more compact molecular architecture compared to its Ap & Af homologs with a cavity running through its very center. The central cavity (Fig 3.17 B) which hosted the putative active site had a prominent hydrophobic tunnel-like extension at its base (towards the periplasmic face of the protein) which could be the probable binding site for a long prenyl chain.



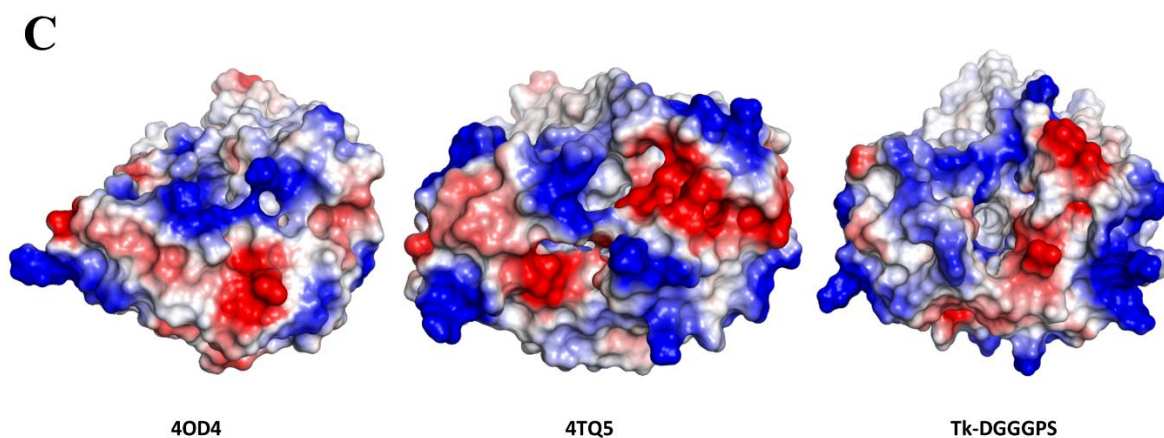


Fig 3.20 Electrostatic surface potential of Tk-DGGGPS structure The surface charge diagram depicting negative residues in red and positive residues in blue, is shown in comparison with Ap-UbiA and Af-UbiA. (A) Frontal view, (B) Periplasmic view, (C) Cytoplasmic view.

3.3.10 Molecular docking studies to understand the dynamics at the ligand binding pocket

In order to determine the binding pocket of the prenyl donor and acceptor ligands of DGGGPS, a molecular docking analysis was carried out with Geranylgeranyl pyrophosphate (GGPP; prenyl donor) and Sn-3-O-Geranylgeranyl glycerol 1-phosphate (GGGP; prenyl acceptor) respectively. Also another docking experiment was carried out with the product catalyzed by the enzyme, (S)-2,3-di-O-geranylgeranyl glyceryl phosphate (DGGGP). Docking experiments carried out with both the reacting ligands were unsuccessful which indicated that the binding of these ligands might be mutually exclusive events. A possible course of the binding events might be the initial binding of the prenyl donor GGPP to the enzyme Tk-DGGGPS, causing a change in the conformation of the binding pocket which allows the binding of the prenyl acceptor GGGP and the geranyl chain transfer occurs almost instantly, accompanied with release of a diphosphate moiety, after which the first ligand ceases to exist and the product DGGGP is formed.

In the docking experiment two binding sites were predicted for the prenyl donor GGPP, however, the docking scores of the sites varied distinctly. One site (on TM7 near the HL67) had a docking score of -5.879 while the other (in the predicted active site pocket) had a higher score of -9.572. This score might indicate that the first site has weak/transient binding while the next site is a thermodynamically stable binding site for GGPP. This, along with the fact that the interacting residues were distinctly different in the two sites, led us to predict a hypothesis that the first site of

binding might be the key to the allosteric change in the active site (binding pocket) which then further allows the binding of the prenyl donor and acceptor molecule to the binding pocket located in the central cavity where the actual carbon chain transfer occurs and results in formation of the final product DGGGP.

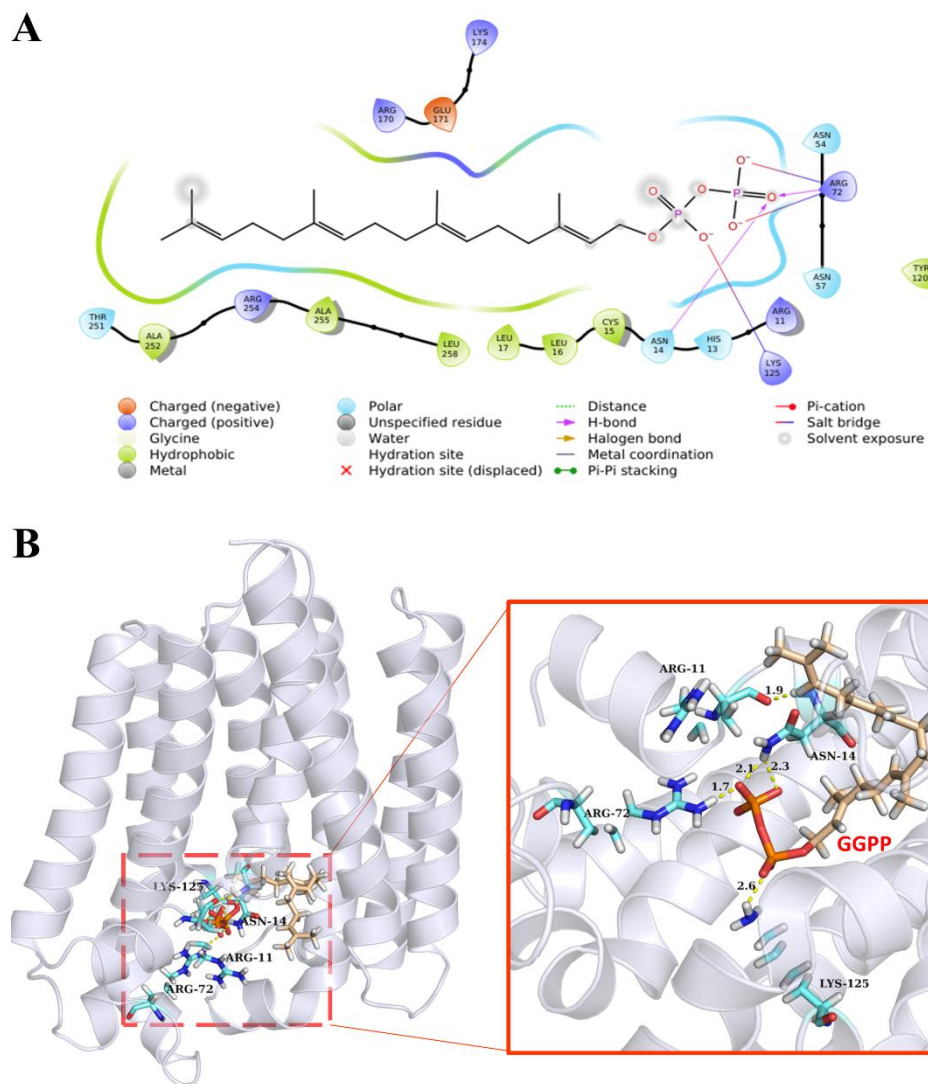


Fig 3.21 GGPP docked in the predicted active site This site of ligand binding has the higher docking score among the two predicted binding sites and the ligand forms hydrogen bonds with three prominent residues N14, R72 and K125. (A) Ligand interaction diagram for GGPP binding to Tk-DGGGPS showing the interacting residues (B) The site of binding shown in greater detail highlighting the molecular interactions between enzyme and ligand.

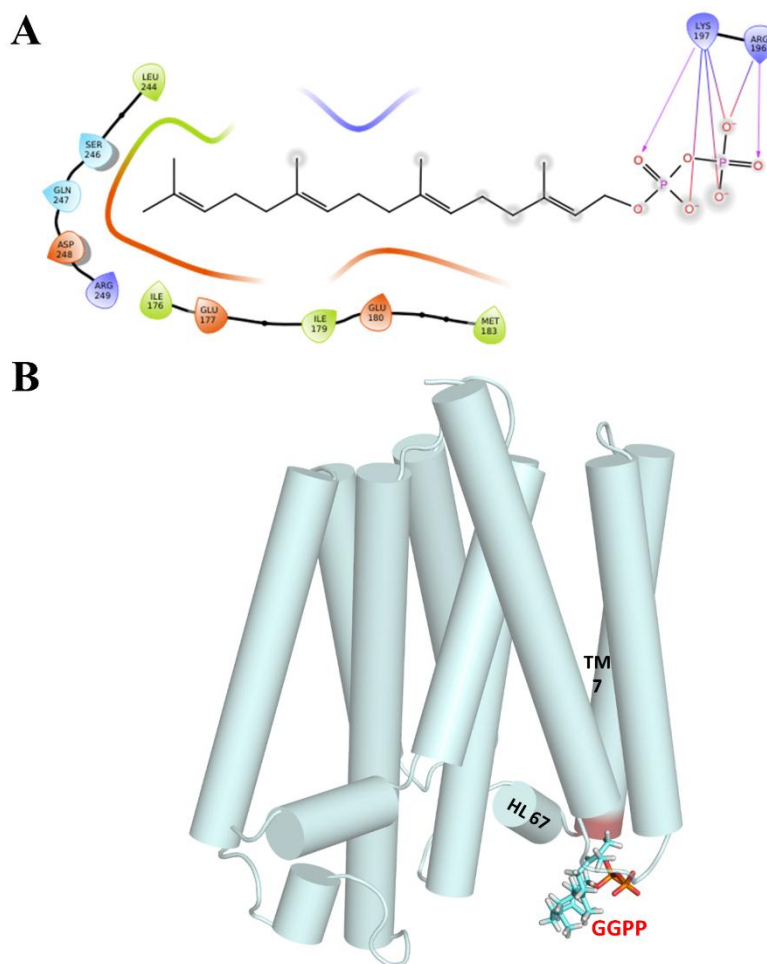


Fig 3.22 GGPP docked in the predicted allosteric site of Tk-DGGGPS This site of ligand binding is predicted to be a transient binding site for GGPP and it forms hydrogen bonds with two prominent residues R196 and K197. (A) Ligand interaction diagram for GGPP binding to Tk-DGGGPS showing the interacting residues, (B) The transient site of binding of the ligand near the HL67 region on the cytoplasmic surface of the enzyme is shown.

This is also supported and validated by ITC experimental data (please refer Chapter 2) where the thermodynamic binding parameters could only be satisfied with the assumption that there were two binding sites with distinctly different dissociation constants (K_{D1} and K_{D2}) for the prenyl donor GGPP, along the lines of the docking scores obtained here. However, the binding and prenyl exchange at the active site pocket might be a spontaneous event as a result of which we were unable to capture it in the docked structure. In the predicted active site pocket, the GGPP interacting residues Arg72 and Lys125 were located in and around the helix loop regions further supporting the fact that the HL regions and its immediate surrounding residues include important

residues in the binding pocket of the enzyme (Fig 3.21 A). Lys125 could be involved in the nucleophilic attack on the phosphate moiety which leads to the formation of the donor prenyl carbocation. The other interacting residue Asn14, located near the amphipathic helix AH1, seem to be involved in the binding and stabilization of the donor ligand along with Arg72 by interacting with the negatively charged phosphate moieties of GGPP (Fig 3.21 B). The site of the transient binding of GGPP showed ligand interactions at Arg196 and Lys197 involving the ligand pyrophosphate moieties (Fig 3.22 A). These residues were again located on and in the immediate vicinity of the HL67 respectively (Fig 3.22 B). This indicated a very evident role of the cytoplasmic helix loop regions on the donor ligand binding at the active site pocket and it was also found to house the putative allosteric binding site of the enzyme.

The prenyl acceptor (GGGP) binding to the pocket showed a higher affinity, indicated by a high negative docking score of -9.699, while the final product DGGGP had a slightly lower docking score of -8.631 but the site of binding of both of these were the same which can be justifiable by the proposed chain of events in the active site pocket of the enzyme. However, the interacting residues were mutually exclusive. The ligand interacting residues Arg11, His13 and Asn14 were located near the amphipathic helix AH1 (Fig 3.23 A, B). The docking of GGGP was also attempted in the GGPP docked Tk-DGGGPS (at the transient binding site near HL67). While the site of docking remained same, interestingly the docking score of -10.143 was more negative than the previous one, indicating a higher propensity of binding of GGGP to the predicted active site pocket, with GGPP bound near HL67. The GGGP interacting residues also changed as the ligand seemed to bind further inside the active site pocket (Fig 3.24 A). The interacting residues Asp175, Asp178 and Lys185 were located in and around the HL67 region while other interacting residues like Asn57 and Lys125 were also located near other HL regions again highlighting the importance of the HL regions. Asp175 and Asp178 were found to interact with the phosphate moiety which could involve proton donation to the moiety, while Lys125, another phosphate binding residue, could be involved in metal dependent (Mg^{2+} dependent) stabilization of the electrophilic phosphate group (Fig 3.24 B). The docking experiment with DGGGP showed the interacting residues to be Arg11 and Thr251 (Fig 3.25 A). Arg11 was located near the amphipathic helix AH1, while Thr251 was located on the TM9 near the cytoplasmic loop connecting TM8 and TM9 (Fig 3.25 B).

Docking of DGGGP to the GGPP bound Tk-DGGGPS (near HL67) did not yield any significant changes in the docking scores and the interacting residues remained the same.

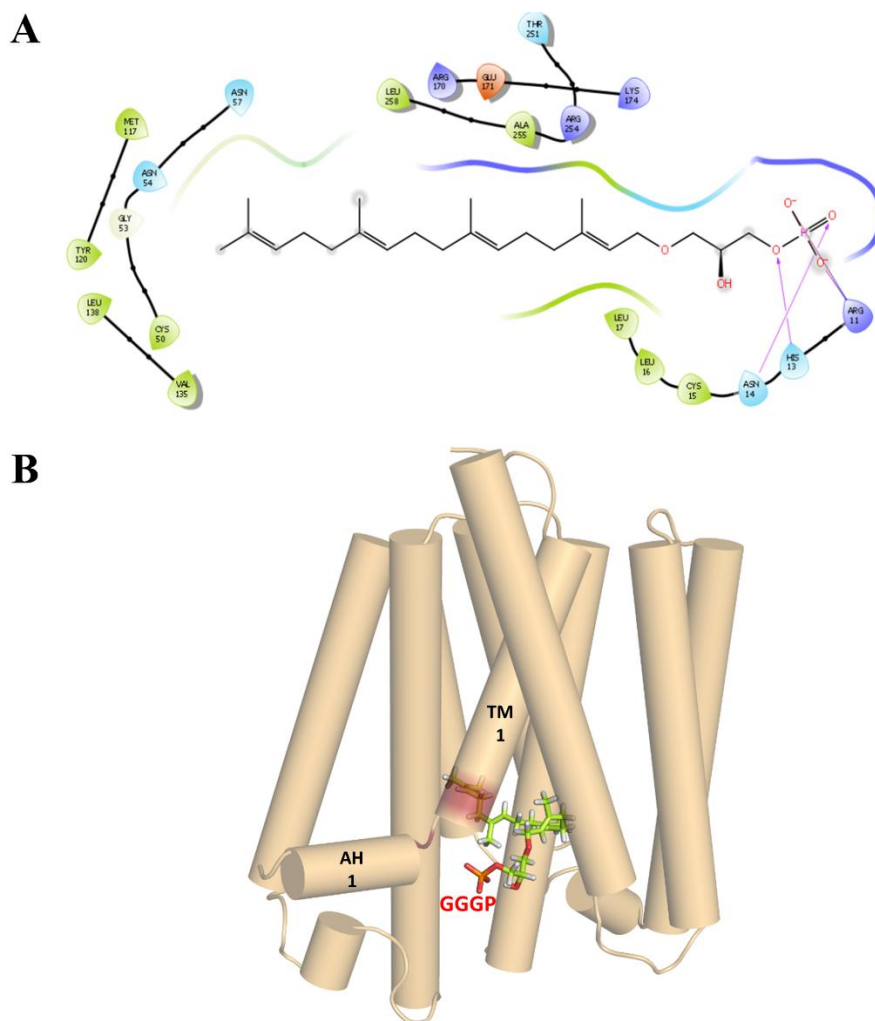


Fig 3.23 GGGP docked in the predicted active site of Tk-DGGGPS GGGP binds to Tk-DGGGPS at the opening of enzyme active site, in absence of GGPP bound at the predicted allosteric site and forms hydrogen bonds with three prominent residues R11, H13 and N14. (A) Ligand interaction diagram for GGGP binding to Tk-DGGGPS showing the interacting residues, (B) The site of binding of the ligand at the opening of the central cavity of the enzyme is shown.

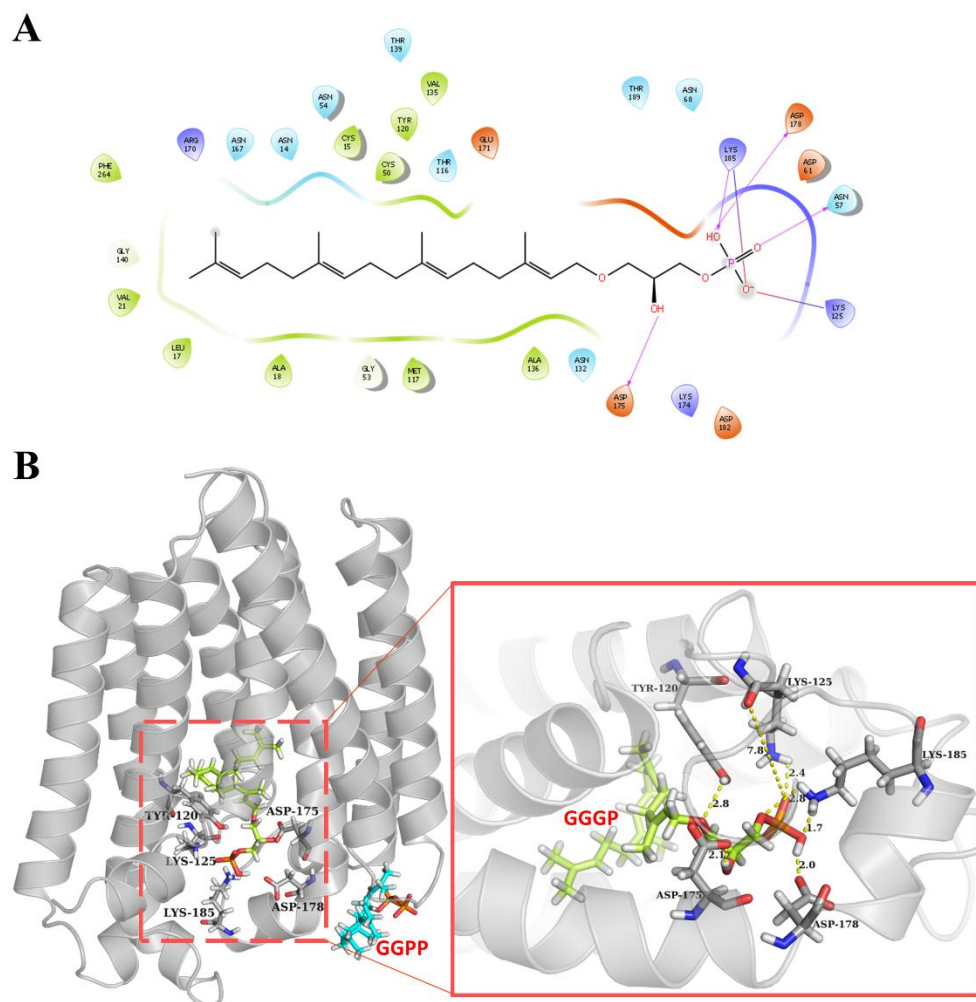


Fig 3.24 GGGP docked in the predicted active site of GGPP bound Tk-DGGGPS GGGP binds to Tk-DGGGPS at a site further inside the central cavity, with GGPP bound to the predicted allosteric site and forms hydrogen bonds with five prominent residues Y120, K125, D175, D178 and K185. This deeper binding site compared to Fig 3.23 might add to the thermodynamic favorability of the prenylation reaction and lower the activation energy barrier as well. (A) Ligand interaction diagram for GGGP binding to Tk-DGGGPS showing the interacting residues, (B) The site of binding of the ligand at the central cavity of the enzyme highlighting the molecular interactions between enzyme and ligand.

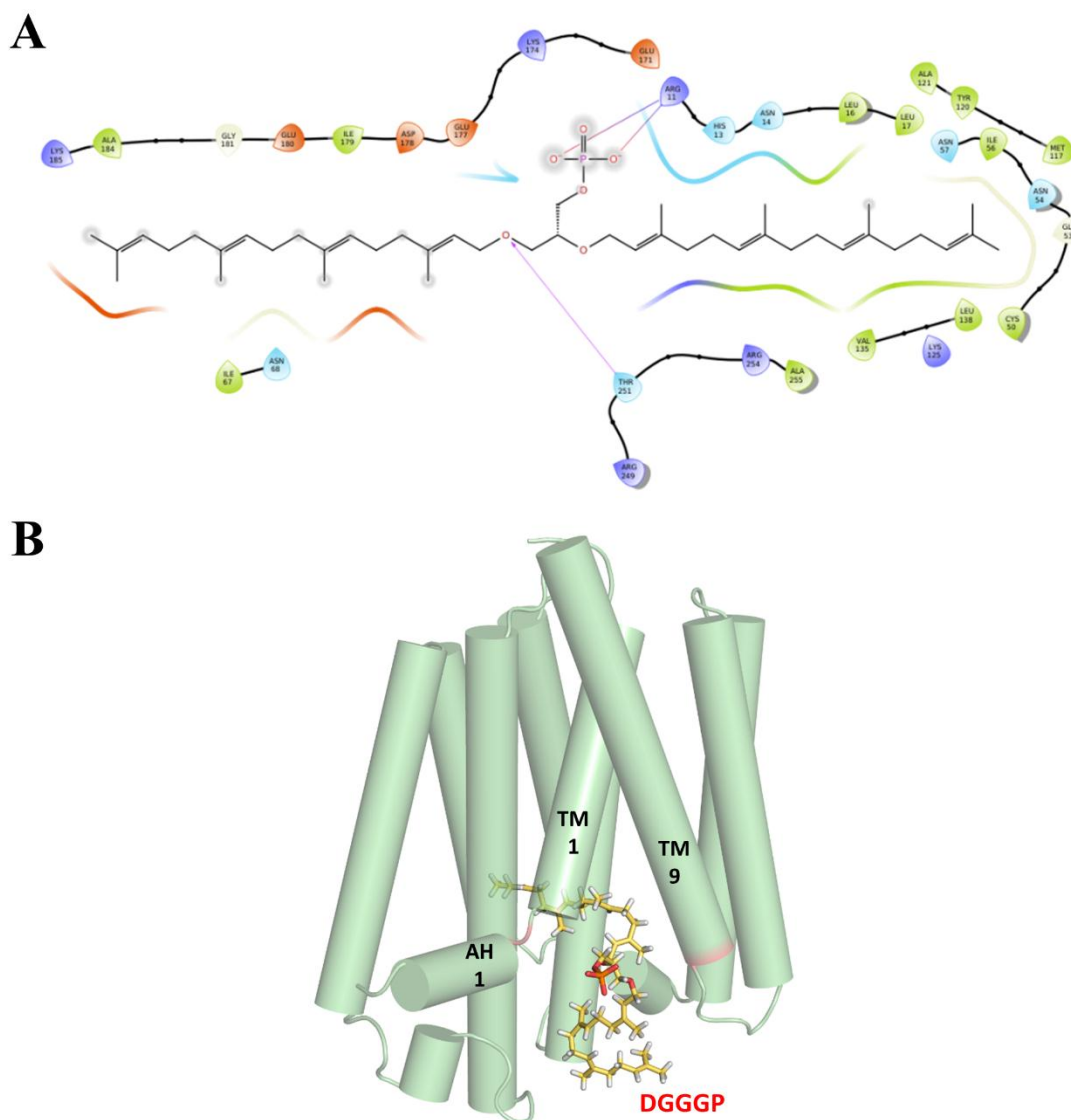


Fig 3.25 DGGGP docked in the predicted active site of Tk-DGGGPS DGGGP binds to Tk-DGGGPS at the predicted active site and forms hydrogen bonds with two prominent residues R11 and T251. (A) Ligand interaction diagram for DGGGP binding to Tk-DGGGPS showing the interacting residues, (B) The site of binding of the ligand at the central cavity of the enzyme highlighting the molecular interactions between enzyme and ligand.

3.3.11 Studying the oligomeric interface

All of the above experiments were performed with the monomer. However, since the enzyme was proven to form a tetrameric assembly in solution from our MALLS experiment, it is important to

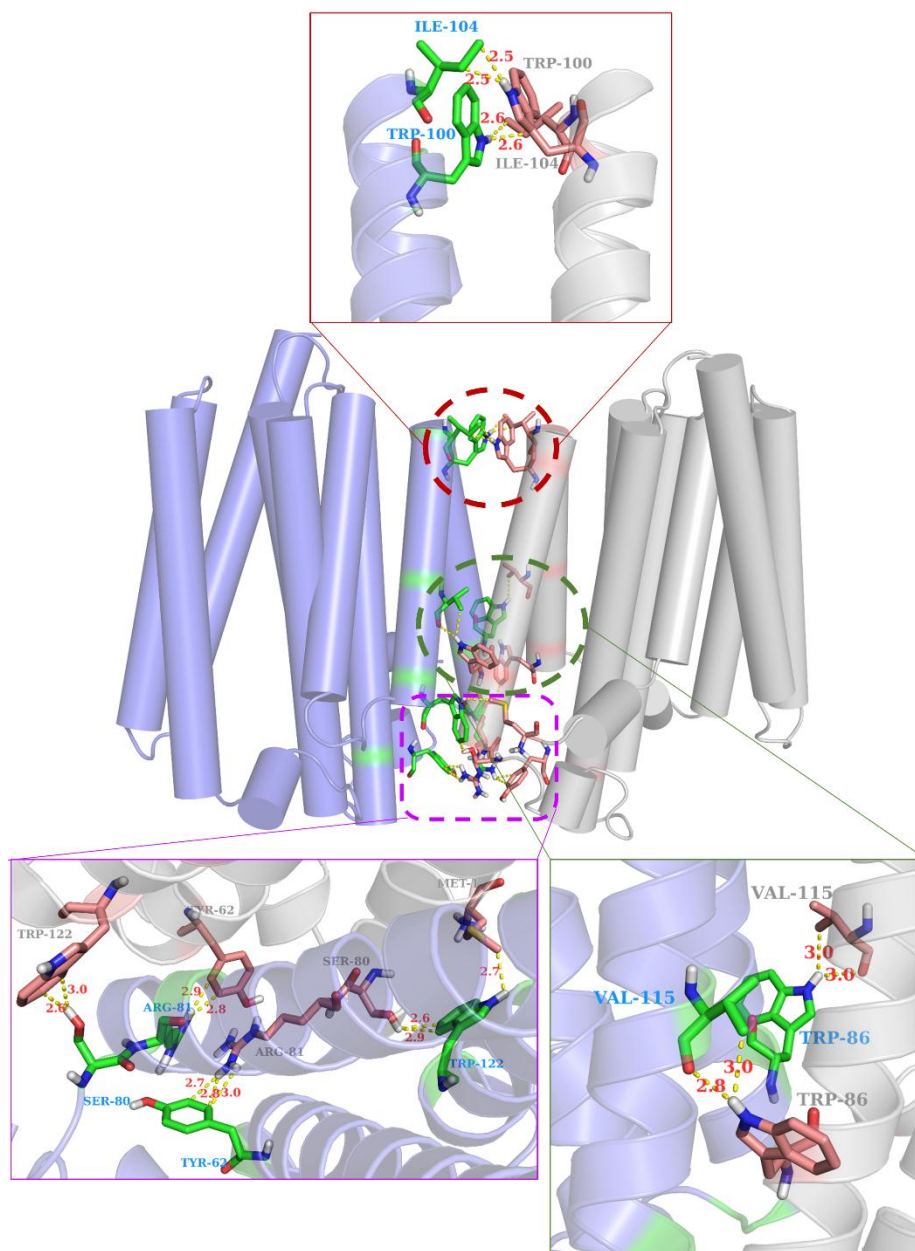


Fig 3.26 Detailed molecular interactions in the oligomeric assembly of Tk-DGGGPS The dimeric assembly is shown here with the subunits color coded in blue and grey respectively. The same residues in pairs from each subunit are involved in the dimeric interface and appears to compose an isologous interface. Among all those residues Y62 and S80 located in the HL23 were found to be evolutionarily conserved and hence could be prominently involved in the formation of a tetrameric interface as well.

study the oligomeric interfaces that help in this assembly and how it impacts the function of the protein as a whole. A prediction study of the dimeric interface of this protein showed us that the

interface is composed with similar amino acids and forms what is known as an isologous interface. The monomers adhere to each other at their sides by face to face manner and the TM3 and TM4 was found to contain most of the residues involved with a distinct role for the loop region of the HL23 in the assembly (Fig 3.26). Historically, loops have been identified as major factors for oligomeric assemblies in well-known enzyme oligomers. According to CONSURF analysis, the only evolutionarily conserved residues among the implicated residues were located in the loop region of HL23 (Tyr62 and Ser80) while the others were deemed variable in their conservation scores. Therefore, this dimeric assembly seems only to be an intermediate state before the enzyme forms its experimentally proven biological assembly, which is a tetramer. This tetramer would also be a putative isologous oligomer with a pseudo-tetrahedral geometry. The residues involved in the assembly are mostly hydrophobic and buried, which is often the case in stable oligomeric interface associations [109, 110]. The oligomeric interfaces for the dimeric and tetrameric association are generally predicted to be different, with two different domains of bonding involved characteristically. A detailed study on the oligomeric nature of the protein might give us insight in to the underlying reaction mechanism as well as the membrane orientation of the enzyme assembly.

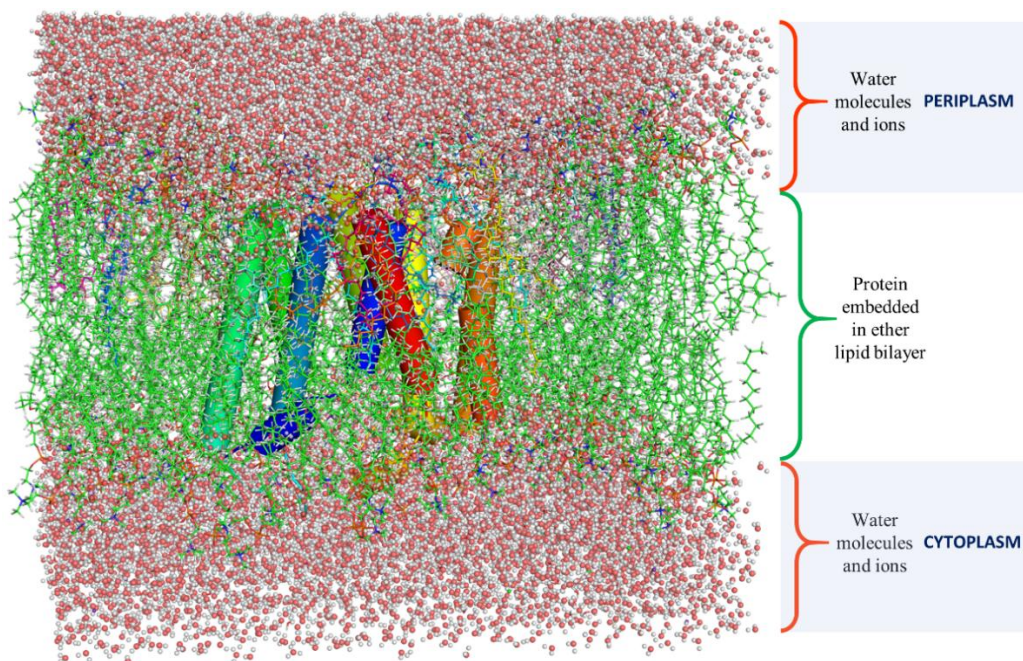


Fig 3.27 Lipid Bilayer with Tk-DGGGPS embedded A cartoon representation of Tk-DGGGPS embedded in an ether DHPC lipid bilayer membrane in a simulated intra-cellular environment.

3.3.12 Studying the behavior of the Tk-DGGGPS in a membrane environment

To understand the native and optimized conformation of a membrane protein under the bilayer environment, which is a complex two-dimensional liquid crystalline system and where mapping the details of protein-membrane interactions using experimental techniques has proved challenging, Molecular dynamic simulation has proved to be a valuable tool to decipher these details and to overcome the above mentioned technical shortcomings. A 50 ns simulation was hence carried out in a membrane environment mimicking the archaeal membrane. The lipids used to build the membrane were ether linked DHPC molecules which is a distinctive character of the archaeal membrane lipids (Fig 3.27). After multiple cycles of minimization and equilibration the system was deemed ready for a prolonged simulation. The results of the simulation were quite revealing of the behavior of the Tk-DGGGPS in a membrane environment. We have previously mentioned that Tk-DGGGPS might have a dual functional role, that of an ion transporter. That presumption gains some validation from these results as we can see that the central cavity transforms into a water funnel during the course of the simulation.

The entire central cavity is packed by water molecules in the final PDB file obtained at the end of the simulation (Fig 3.28 A). All the nine TM helices were found to harbor conserved residues which were interacting with these water molecules out of whom core residues of the central cavity, found to interact with water molecules were located on TM1, TM2, TM4, TM5, TM6 and TM8 helices. This could be implicated in the water-dependent transport of ions or could be involved in the hydration of the important active site polar and charged residues like arginine (Arg11, Arg170 & Arg254), asparagine (Asn14, Asn54, Asn57 & Asn132) and histidine (HSD13 & HSD253) primarily. There are previous reports stating that charged and polar residues in active site pockets of membrane proteins located inside the lipid bilayer require hydration by water funnels. The HL regions were found to have major interactions with the water molecules surrounding the lipid bilayer indicating that their conformational stability could be intricately related to these interactions. The important active site residues located in the ⁵⁷NDxxDxxxDxxxxPxRPxP⁷⁵ and the ¹⁷⁰RExxKxxEDxxGD¹⁸² clusters were found to have access to the water molecules located in the central cavity during simulation. Interestingly, a number of surface amino acid residues were also found to have interaction with the membrane lipids, mainly via their side chains. Among these residues, some were found to belong to the ¹²⁰YxxxxKxxxxGN¹³² cluster and others in HL

regions (HL23, HL45 & HL89). Notably one such residue, Lys197 was also a predicted docking residue for the ligand GGPP in what we predicted as its allosteric binding site of the enzyme. This suggests that lipid interaction could have a major role in the biochemistry of Tk-DGGGPS. A prominent metal interaction was identified with the C-terminal residues Ser274, Leu275 and Arg277 which suggests its role in stabilization of the membrane protein.

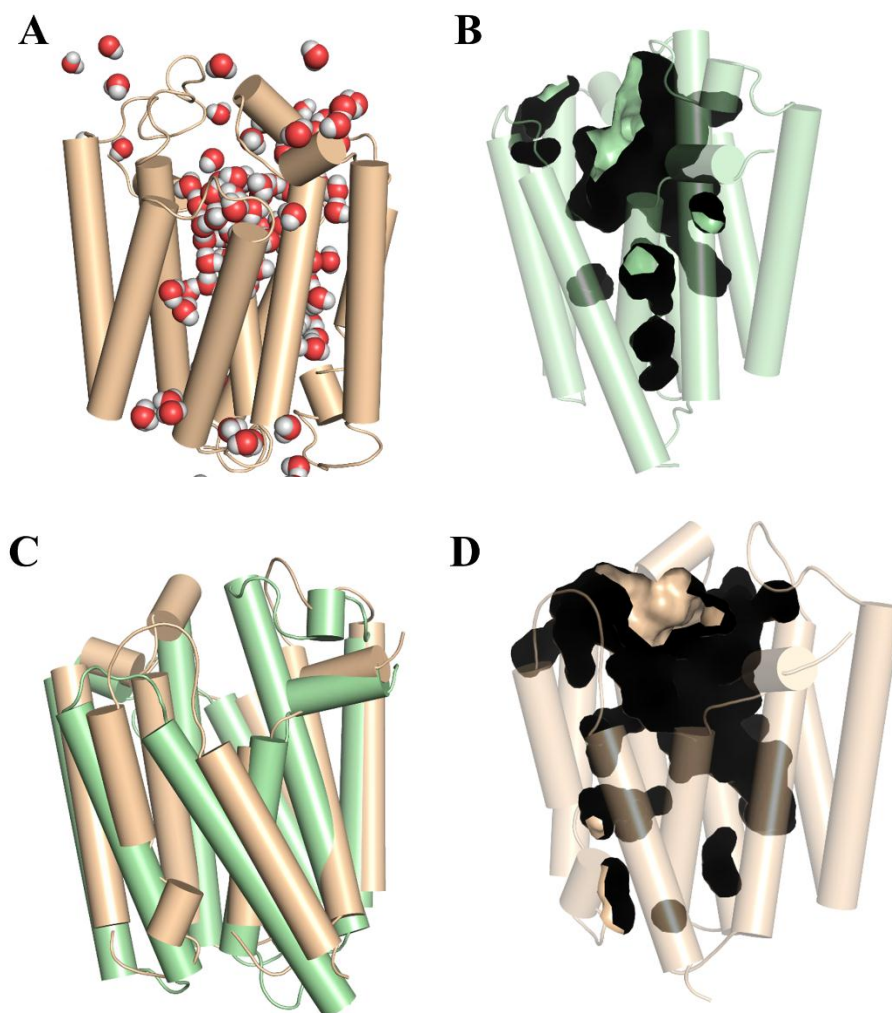


Fig 3.28 Snapshots of the Tk-DGGGPS structure before and after the simulation (A) A water pocket is found to exist in the central cavity of the protein at the end of the 50ns simulation hinting at a unknown role of these water molecules in absence of the ligands of the enzyme; the role could be that of an ion transporter as predicted before (Section 3.3.7), (B) The narrow and constricted central cavity of Tk-DGGGPS before the start of the 50ns simulation, (C) An alignment between the pre and post simulated protein structure (pre-simulated color in green) showing major changes in the spatial arrangement of the secondary structural elements of the Tk-DGGGPS homology model, (D) The spacious central cavity at the end of the simulation indicating major molecular changes in inter-residue interaction and the water-residue interaction throughout the simulation.

The system stabilizes over the last ten nanoseconds of the simulation as indicated by the RMSD graph of the simulation and the R_g is also notably stable throughout the simulation indicating that the protein is structurally stable over the course of the simulation (Fig 3.29 A, C). However, in spite of the stable R_g value throughout the simulation, when the final optimized structure of the Tk-DGGGPS was aligned with the pre-simulation structure we could observe some notable and significant differences (Fig 3.28 C). Though the overall RMSD between the two structures was $\sim 1.5\text{\AA}$, the average number of molecular contacts in the system were found to remain nearly equivalent (Fig 3.29 D). However, an analysis using the PIC server (Protein Interactions Calculator) [111] gave a detailed insight into the prominent changes in the intra-chain hydrogen bonding of the pre-simulation and post-simulation Tk-DGGGPS structure. Almost a 25% decrease in the overall hydrogen bonding between the main-chain carbon chains were observed after the membrane simulation. The new hydrogen bonds formed as a result of the simulation could play important roles in stabilizing the intramolecular interactions in the central cavity of the enzyme, Tk-DGGGPS. The hydrogen bonds between the main chain of the residues generally influence the mobility of the structural elements of the protein while the hydrogen bonding between the main chain and side chain of residues influences the stability of the structural elements like α helices and contribute prominently to the changes in the secondary structure components during the simulation [112]. Only the hydrogen bonding dynamic of the conserved residues were tabulated in Table 3.1 A, B.

By studying the simulated structure in more detail, we could identify certain patches of the protein which were relatively much more mobile in comparison to the others. The areas showing highest RMSD differences were located in either loops or the TM helices that is the part of the exterior protein surface. Most prominent among them was the TM7 which can therefore be predicted to be a probable interface for an association involving the external surface residues, namely the oligomeric interface. The TM 8 was found to form a kinked helix with a kink angle of 8.28° after the simulation, while a new helix showed up in the HL67 region where a loop was present in the original structure and the HL23 region lost its helix component (Fig 3.28 C). It was notable that both the HL23 and HL67 regions had adjacent polar residues interacting with the ether lipids of the bilayer (Lys82, Lys197 & Arg198 respectively) and therefore a notable influence of the membrane lipids on the overall conformation of Tk-DGGGPS could be inferred. A detailed look

at the short kinked region revealed that it had a fairly conserved valine residue and a proline residue which was surprisingly non-conserved. This proline kink, on its either side, had two residues Leu230 and Asp232 which were prominently interacting with water molecules during the simulation and thereby could be pivotal to the stability of the protein assembly in the membrane environment [113] and could have a major role in the overall functional efficiency of these enzymes.

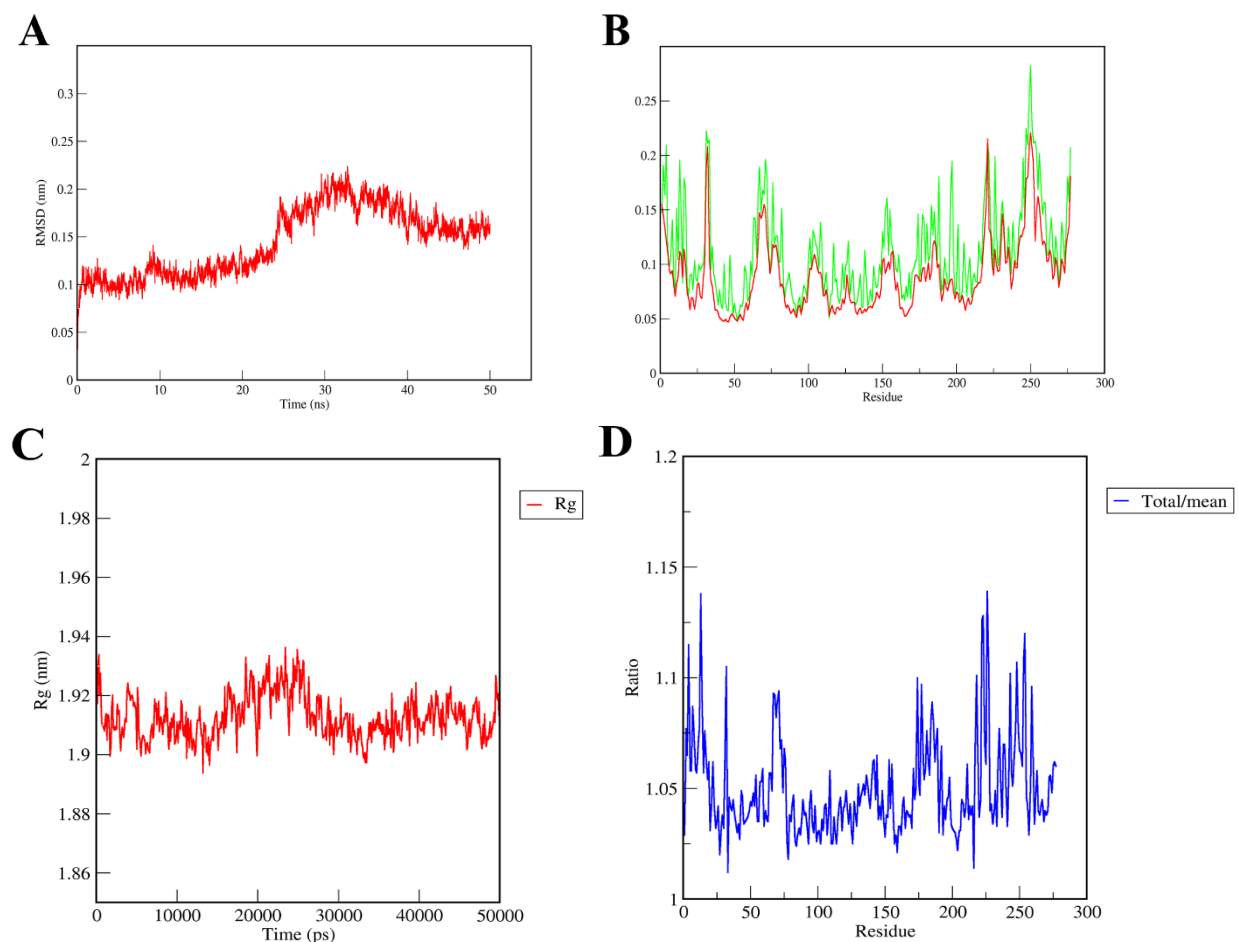


Fig 3.29 A detailed analysis of the 50 ns simulation of Tk-DGGGPS in an ether lipid bilayer (A) The backbone RMSD plot of the Tk-DGGGPS structure over the entire duration of 50 ns reflecting the stabilization of the entire enzyme-membrane system over the last 10 ns, (B) The RMSF plot of the amino acid residues of Tk-DGGGPS showing a comparison between the RMSF of the backbone carbon chain and the side chains of the residues highlighting four regions of notable fluctuations and high B factor, (C) The radius of gyration (R_g) plot showing the protein structure was stable throughout the simulation, (D) The ratio of total molecular interaction and the mean interactions between the residues is plotted and highlights the system stability as the average number of molecular contacts remains same.

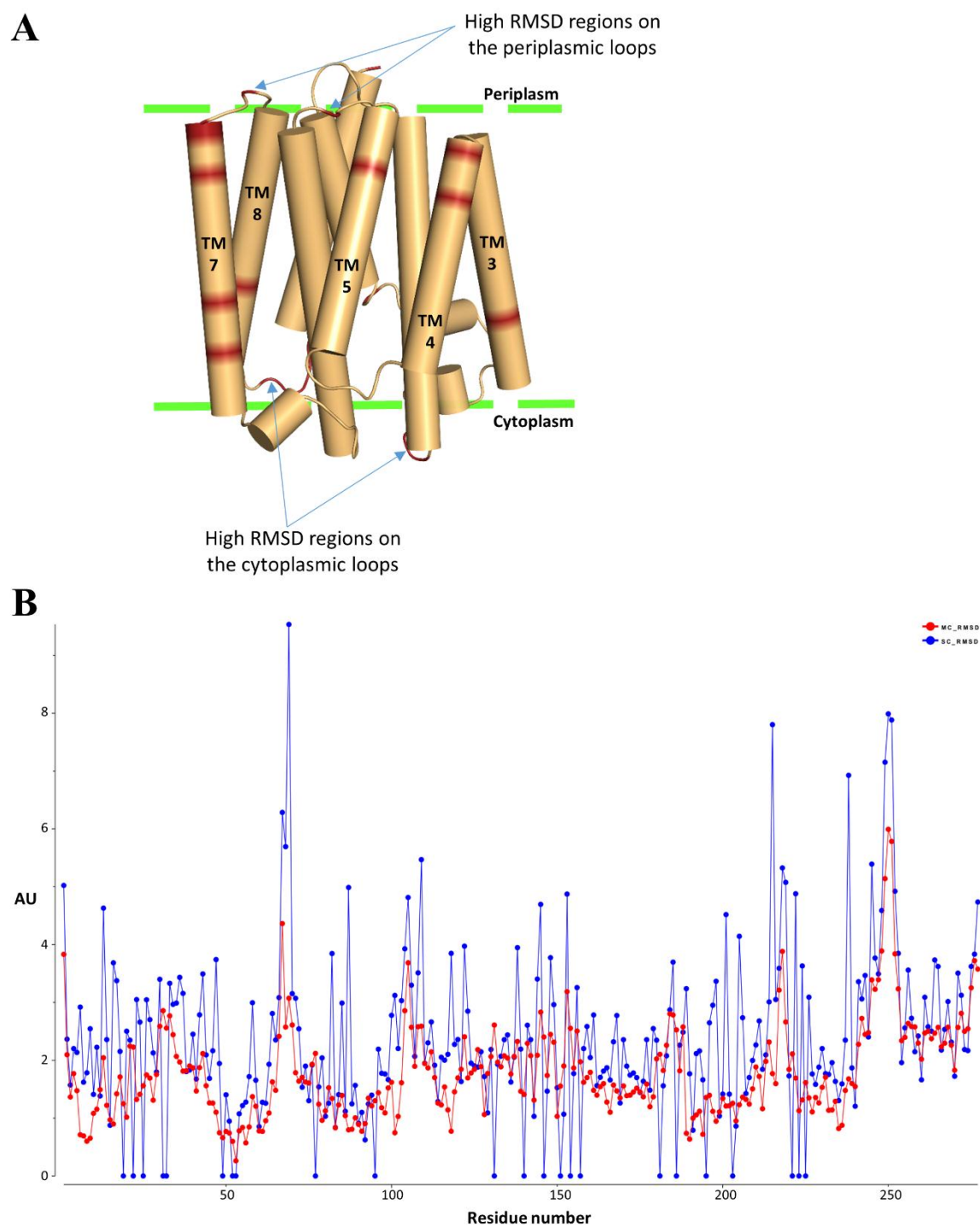


Fig 3.30 Superpose analysis of the 50 ns membrane simulation of Tk-DGGGPS (A) The location of the residues with high RMS deviations in the secondary structural elements are highlighted, among which the notable ones are located in the loops present in the periplasmic and cytoplasmic exposed surfaces of the protein and the TM3, TM4, TM5, TM7 and TM8 helices, (B) The main chain RMSD and side chain RMSD of the amino acid residues of Tk-DGGGPS color coded in red and blue respectively are shown.

Overall the central cavity seemed to have relaxed further after the simulation with an increase in the cavity size (Fig 3.28 B, D). The RMSF graph of the residues over the course of the simulation further confirmed that there existed a few prominent patches which were thermally more unstable, dynamic and had high B factor values (Fig 3.29 B). An analysis of these residues showed that they were located either around the central cavity or on the external surface of the enzyme. In the central cavity, the patches are predominant near its cytoplasmic opening and at the end of the two prominent hydrophobic channels, one of which is an extension of the central cavity and has an outlet on the protein surface at the periplasmic end while the other extends perpendicularly from the central cavity and is close ended. The tunnel extension of the central cavity is close-ended in the original homology model but becomes open ended here indicating a possible role in the pore formation necessary for acting as a transporter (Fig 3.28 D). The tunnel perpendicular to the central cavity only shows up after the simulation and therefore must have significant roles in the protein activity, either enzymatic or transporter or both. All of this indicate a probable involvement of these dynamic residue patches in the functionality of the Tk-DGGGPS. The per residue RMSD values, before and after the dynamic simulation were analyzed using Superpose and it indicated two prominent patches where the side chain RMSD were notably high (Fig 3.30). Both of these patches were found to be present at the cytoplasmic face located on the HL23 region and between the TM8 and TM9 (HL89). While the HL23 residues were predicted active site residues and conserved, the residues in HL89 were found to be non-conserved. This indicated that HL89 probably had roles in interacting with the archaeal membrane lipids responsible in maintaining the membrane stability. A high number of residues on TM7 were also found to have prominently high RMS deviations indicating that they could have an influence on the oligomeric make-up of the Tk-DGGGPS. Two other isoleucine residues were found at the periplasmic face of the Tk-DGGGPS with distinctly high side chain RMSD values located on loops between TM5&6 and TM7&8 of which one was a conserved residue. These residues might be involved in influencing the hydration dynamics of this membrane protein near the archaeal membrane or even act as interacting partners to peripheral membrane proteins. Notable residue fluctuations were also observed on the TM helices which formed the exterior surface of the protein (TM3, TM4, TM8 and TM9) out of which TM3 and TM4 have been predicted to form a part of the dimeric interface of the enzyme assembly. This is also justified by the high side chain RMSF of these residues indicating that the absence of

their interacting counterparts in another monomer maybe a prime cause for this fluctuation. Membrane proteins are often known to perform more than one functions in the membrane in what is popularly known as a moonlight functionality of the enzyme [114]. This makes the presence of the protein in the membrane, thermodynamically more favorable. Generally, the primary responsibility of these enzymes are enzymatic catalysis while secondary roles involve signal transduction, ion channel functionalities and transcriptional regulation among many others. Tk-DGGGPS is therefore a putative moonlighting protein which acts a prenyltransferase primarily and as transporter when it is not involved in enzyme catalysis.

Table 3.1 A list of new hydrogen bonds formed after the 50ns simulation involving conserved residues of Tk-DGGGPS (A) Hydrogen bonding between the main chains of donor and acceptor amino acid residues, (B) Hydrogen bonding between the main chain of donor residues and the side chain of acceptor residues.

A

Intra-chain hydrogen bonding			
Main chain DONOR		Main chain ACCEPTOR	
Position	Residue	Position	Residue
1	Met	78	Ala
21	Val	17	Leu
43	Phe	22	Gly
60	Phe	58	Asp
61	Asp	58	Asp
76	Arg	72	Arg
79	Met	77	Gly
133	Val	131	Gly
134	Val	131	Gly
135	Val	131	Gly
141	Ala	138	Leu
149	Ala	146	Gly
150	Val	146	Gly
189	Thr	187	Ala
193	Leu	191	Pro
215	Phe	213	Ala
263	Val	260	Lys
271	Leu	269	Ala

B

Intra-chain hydrogen bonding			
Main chain DONOR		Side chain ACCEPTOR	
Position	Residue	Position	Residue
15	Cys	10	Thr
15	Cys	11	Arg
50	Cys	14	Asn
54	Asn	11	Arg
54	Asn	14	Asn
55	Thr	09	Ile
68	Asn	64	Ile
69	Arg	75	Pro
69	Arg	75	Pro
69	Arg	76	Arg
76	Arg	70	Pro
88	Ser	52	Gly
145	Tyr	106	Asp
256	Gln	252	Ala
260	Lys	256	Gln
277	Arg	33	Leu
277	Arg	34	Pro

3.3.13 Study of disease-causing mutations in human homologs of DGGGPS

A detailed study was performed about the DGGGPS homologs in humans, namely UBIAD1, COQ2 and COX10 and the reported mutations of these proteins obtained from the HGMD database. In UBIAD1, most reported SCD mutation were observed to be clustered around the central core with as many as 12 mutations in the helix loop regions which form an integral part of the extra-membrane cytoplasmic cap domain of these enzymes (Fig 3.31). In COQ2, the mutations were far more widely distributed to all over the protein including the external surface of the protein, while not many COX10 mutations were majorly implicated in diseases

A pairwise alignment carried out between the UBIAD1, COQ2 and COX10 protein sequences and Tk-DGGGPS respectively showed that many of these mutations were also conserved residues. Only the conserved mutations were considered further. In UBIAD1, the mutations D118G, R119G and L121V/F were located on HL23, mutations G186R and L188H on HL45, and mutations

D236E, D240N and I245N on HL67. A hot-spot for SCD mutations, N102S was located on TM2 along with G98S and D112G/N, while Y174C was located on TM4 [32, 33]. Most of these mutations are reportedly involved in differential perturbations of cholesterol metabolism. Mutations G98S, N102S, D112G/N, R119G, D236E and D240N may be involved in either substrate binding or stabilization of the carbocation intermediate (R119G and D236E predicted to affect the binding of the XPP substrate) [115]. The others L121V/F, Y174C, G186R and I245N are predicted to be involved in the structural stability changes in the protein also known as TERE1 [116, 117].

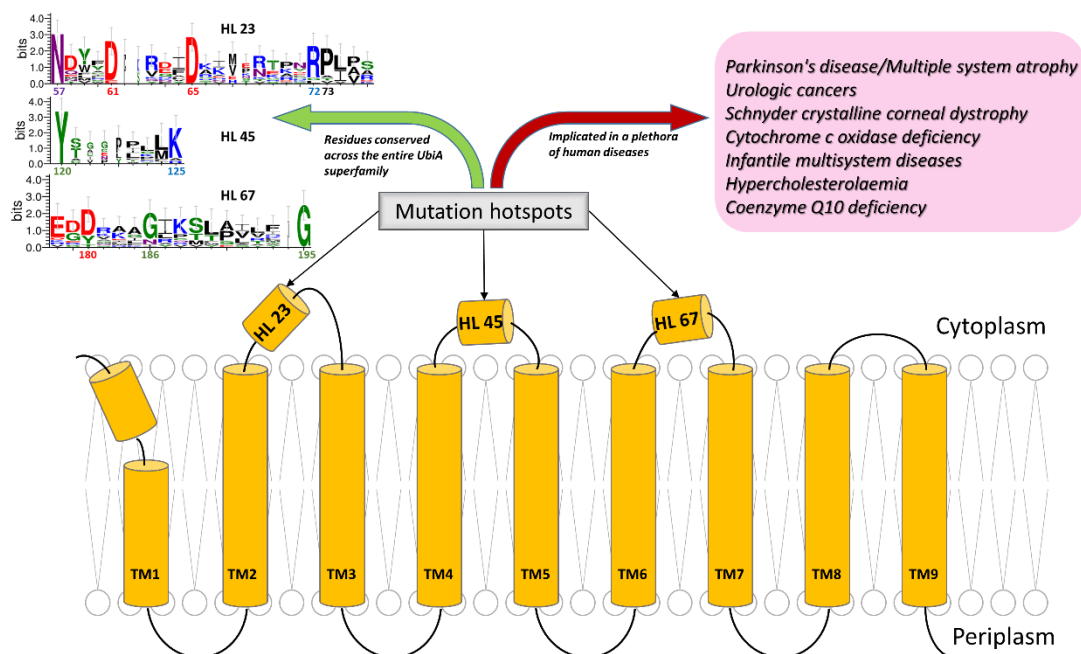
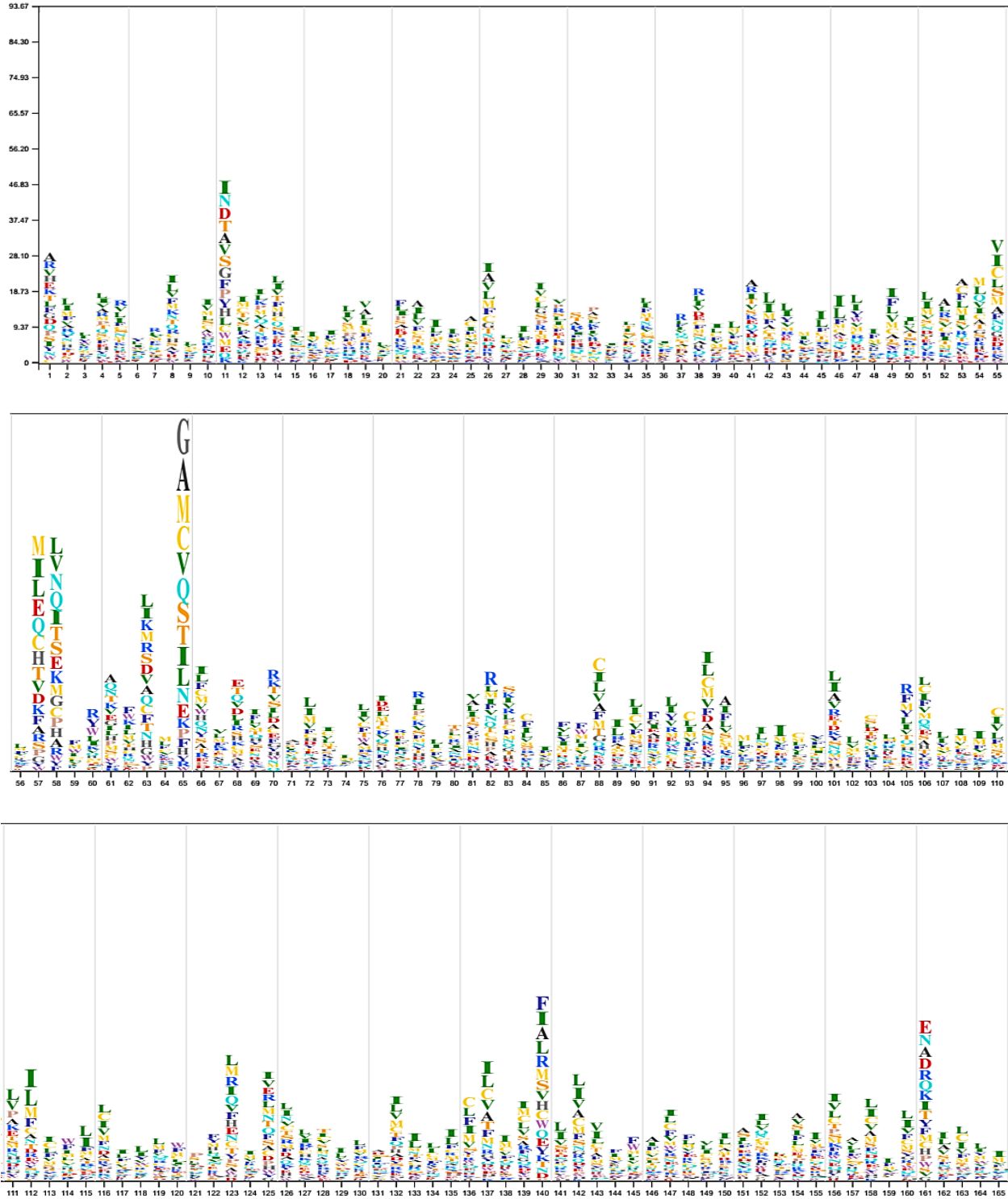


Fig 3.31 Mutation hotspots in Tk-DGGGPS and their connection to its human UbiA homologs implicated in diseases Most of the reported disease causing mutations are found in the HL regions of the UbiA human homologs and in the Tk-DGGGPS, they are the predicted mutational hotspots as well. The WEBLOGO shows the high conservation scores in these HL regions.

In COQ2, the only conserved mutation to be located in HL23 was R197H. The mutation M128V was located in the AH1 while mutations I147T, L162F and L286F were located in TM1, TM2 and TM6 respectively. Deletion mutants at codon positions 233 and 398 were located in TM3 and TM9 respectively. N386I, V393A and L402F mutants were both located on TM9 as well. The only mutation found in the periplasmic loop region between TM1 and TM2 was P157S. Among them, R197H and N386I may have functional implications in the enzyme catalysis with R197H possibly

affecting the diphosphate binding of the XPP substrate [118, 119]. The mutations L286F and V393A (associated with multiple system atrophy) may be involved in structural destabilization of the COQ2 enzyme [120, 121].



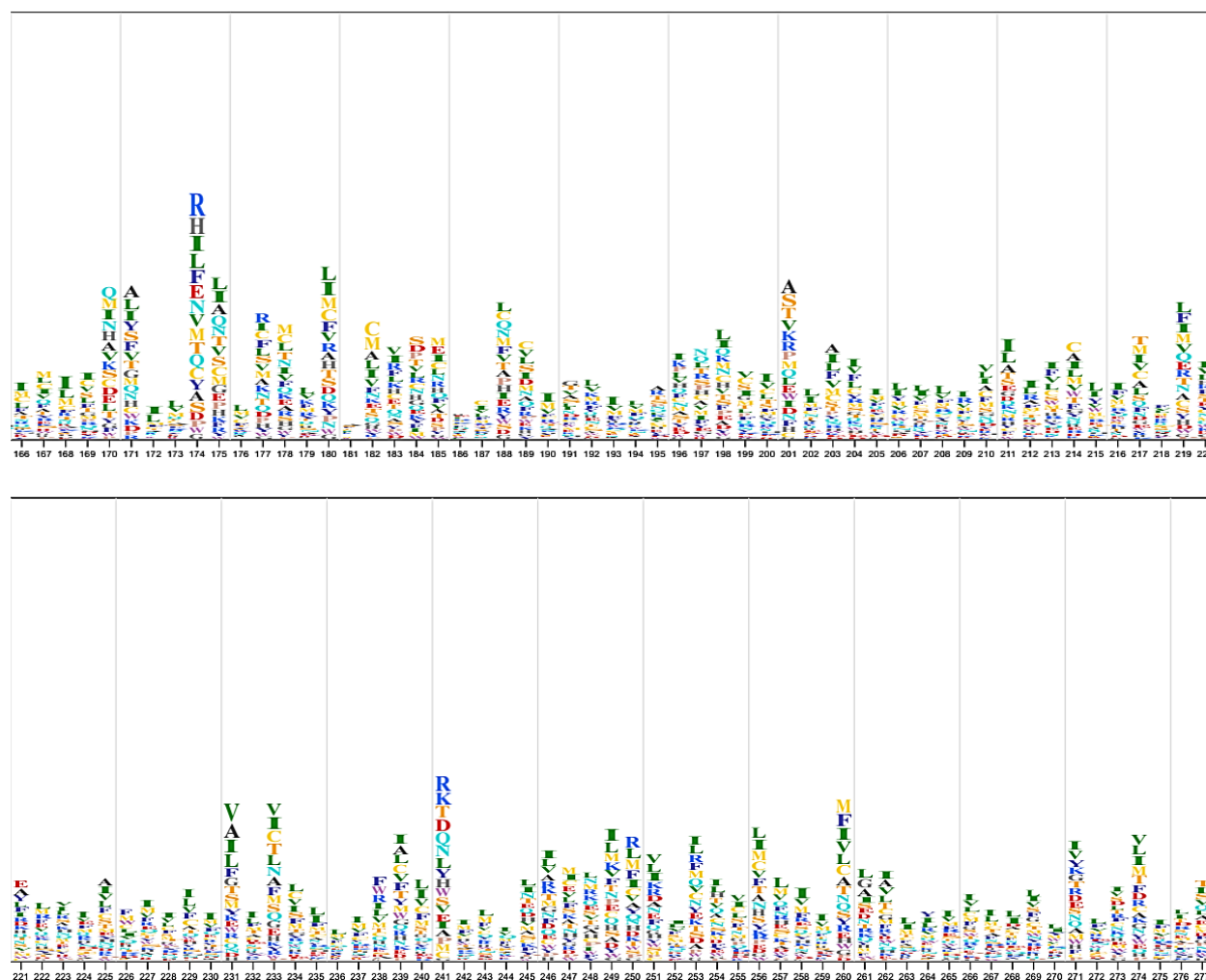


Fig 3.32 Mutation hotspot residues of Tk-DGGGPS from STRUM webserver Probability scores for the propensity of the residues to be mutational hotspots according to calculations from the STRUM server highlights eight residues of which five are predicted structurally (D58) and functionally relevant residues (R11, N57, D65 & K174).

In COX10, the conserved mutations P225L and R228H were found to be in HL23, while mutations G288R and D336V/G were located in HL45 and HL67. The conserved mutation N204K was located in TM2. All the mutations apart from G288R were predicted to have some involvement in interaction between the enzyme and its respective XPP substrates, while G288R may be causing a structural anomaly of the enzyme [69].

To study which residues in Tk-DGGGPS were most prone towards mutation and verify if there were any similarities to its human homologs, an attempt was made to study the structural stability changes in Tk-DGGGPS by single-point mutation at all the amino acid residues using a web server

called STRUM [104]. The results highlighted a few mutational hotspots in our protein, namely residues Arg11, Asn57, Asp58, Glu63, Asp65, Gly140, Leu161 and Lys174 with the highest score being for Asp 65 (Fig 3.32). It is interesting to note that among the above 8 hotspots, the predicted functional conserved residues Arg11, Asn57, Asp65 and Lys174 and the predicted structural conserved residue Asp58 had the highest probabilities for mutation and among them Asn57 and Asp65 were found to align with known SCD mutations in human UBIAD1, N102S and D112G/N [122]. The other residues could thereby new mutation hotspots which could be useful in further disease related studies.

3.4 Discussion

Even though archaeal proteins have shown high tendencies towards crystallization, like other proteins, they require a favorable surrounding environment to reach the energy minimized state where they can maintain a relatively stable conformation for proper packing in the crystal unit cell. This is more difficult in case of membrane proteins wherein due to their intrinsic hydrophobic nature, a very critical mixture of solvents and ions, along with enough presence of lipidic components are required to thermodynamically favor the hydrophobic interfaces to form an ordered crystal structure. Often due to this very critical dynamic, many membrane protein crystal structures have regions which are lacking in electron density [123]. This is due of the fact that, those regions of the protein were unable to assume a stable conformation during crystallization and the major player for this is the surrounding environment of the protein in solution.

The crystallization trials of Tk-DGGGPS confirmed that only the crystal screens optimized for membrane proteins had the optimal dynamic mix of components necessary for providing positive results for purely alpha helical proteins such as our enzyme. However, the crystals obtained were not necessarily diffract-able. To obtain diffraction quality crystals further additives were required which could include a variety of molecules ranging from small molecules like multivalent salts to detergents, polyalcohols and non-volatile organic compounds. The observations made during crystallization of the Tk-DGGGPS were, the positive influence of magnesium salts in crystal formation along with a need for low-high molecular weight PEGs for crystallization. This could be justified by the fact that prenyltransferases are known to be Mg²⁺ dependent for their catalytic

activities. While the influence of Mg^{2+} ions is known to be functionally significant, our results suggests that it might also be involved in structural stability of the protein, specially the active site pocket. Often due to the presence of entropic shields on the surface of the protein due to amino acid residues with flexible polar side chains in the protein surface, it becomes difficult for the protein to incorporate into a crystalline lattice [124]. For a crystal lattice to form, the side chains of the amino acids present at crystal contact interfaces along with the solvent molecules present contribute to a loss in entropy which results in the immobilization of the protein in a thermodynamically stable crystal structure. This makes it highly necessary to mutate these residues to obtain entropically restricted overall structure of the protein [125]. Often one or two such mutations can make the difference between a good quality and a poor quality crystal or even successfully assist in the crystallization of an otherwise structurally dynamic protein. In Tk-DGGGPS such residues were mainly observed on the external solvent accessible cytoplasmic surface and were three clusters of lysine residues. The elongated polar side chain of lysine has the highest mean side-chain conformational entropy ($T\Delta S_{conf}$) score of -1.89 which tells us that it has a major impact on the stability of the protein [126]. A mutation to a more non-polar sidechain containing amino acid like alanine for instance, thereby makes the protein more stable conformationally and this was experimentally proven when there was a notable increase in protein yield in the K185A Tk-DGGGPS mutant. However sometimes the conformational flexibility is required for the protein to maintain its structural and functional integrity and therefore the K188A and the K185,188A mutants were not expressed in high amount. This makes it highly important that mutant generation be done with extreme care and after careful consideration.

Moreover, the crystallization of membrane proteins not only involves the protein but also the co-purifying detergent in the form of a PDC. Therefore, the detergent component of the PDC also plays a major role in the crystal formation. Mostly mild detergents like alkyl maltopyranosides and alkyl glucopyranosides have been successfully used for membrane protein crystallization. In general, it's been observed that shorter chain detergents help in diffraction at better resolutions [127]. Hence detergent screening is necessary for identifying the ideal detergent/s for a particular protein. However, there exists no set pattern regarding the detergents which would be favorable to certain types of membrane protein and a high throughput method for detergent screening should be ideally performed for obtaining initial hits or for further optimization. We found that DDM

remains the most efficient detergent for extracting the membrane protein from the membrane fraction obtained after ultracentrifugation. While the yield is high with DDM extracted proteins, they face a problem of detergent micelle accumulation during concentration of the protein for crystallization due to the big size of the DDM micelles. Also our results indicated that efficient detergent exchange is also a roadblock that needs to be addressed while looking towards new detergents to co-purify membrane proteins of interest. Recently, an important improvement in the field of membrane protein crystallization has been the introduction of lipidic mesophases in the form of LCP crystallization which has had considerable success in GPCR crystallization. The LCP method along with membrane protein crystallization in excess lipids and detergents (HILIDE) have hence become prospective methods for obtaining quality crystals of membrane proteins.

Multiple sequence analysis between a protein and its homologs or between the members of an enzyme superfamily can give us insights into the important amino acid residues which could be structurally or functionally important for the protein. Generally, these residues are found to be conserved throughout the enzymes or at least replaced by similar residues. Often these residues are found to be clustered and form known functional motifs of the protein. Often there are very low sequence identity between proteins which have evolved due to convergent evolution but perform similar functions in different species. Such remotely homologous proteins have conserved structural core features which are maintained by similar set of amino acid residues, but not necessarily in the same order, for being identified by sequence alignment. These can be identified by pairwise comparison of profile hidden Markov models (HMMs) and verification with predicted secondary structure similarity, available from protein crystal structure data in PDB. While trying to understand the phylogenetic connection between protein, it has been observed that more the homology between proteins, more are their propensity to be grouped in the same clade. Often even remotely homologous proteins can occupy a leaf in the same clade as their counterparts. However, the phylogenetic tree is quite relative to the entire population of proteins considered for the analysis and it is highly important to consider the significance of the tree and verify it with supporting facts before trying to draw any conclusions based on a phylogenetic analysis. Structural analysis of a protein can give insights into a variety of aspects including the oligomeric assembly, the biological significance and the functional mechanism of the protein and its related homologs.

Knowhow about the molecular structures of prenyltransferases of the UbiA superfamily are fairly limited with only two known crystal structures (Ap-UbiA and Af-UbiA). Even though these structures have been studied in detail, and molecular details of the reaction mechanism have been partially elucidated, more insights are necessary for understanding the different variables in structural features of the prenyltransferases. The substrate promiscuity and the notable conservation of the core structures of archaeal homologs while having different catalytic responsibilities shifted our focus to a prenyltransferase unique to archaea (DGGGPS) and we through our studies have results indicating that this protein might be an ancestral homolog among UbiA prenyltransferases and have structural features pertaining to both the known archaeal homologs (Ap-UbiA and Af-UbiA). The Tk-DGGGPS showed a remote homology with a MFS transporter by sequence analysis, which was further supported by our findings in the MD simulation of the protein in a membrane environment. This indicated towards a previously unknown moonlighting function of DGGGPS, that of a transporter. Moonlighting proteins are an interesting area to study for development of novel therapeutics, identification of new cellular processes, finding the connecting partners among different biochemical pathways, elucidation of novel protein mechanisms and obtaining new insights into the evolution of protein structures. The Tk-DGGGPS overall structure showed conservations in the TM helices and the HL regions emphasizing their importance in the structure function relationship of the protein. The TM helices which make up the protein surface exterior along with the cytoplasmic HL regions connecting these helices are prominently involved in the assembly of the protein oligomer in the archaeal membrane. Many of these conserved residues were also found to be mutational hotspots. It was found that in the Tk-DGGGPS human homologs, most of the known diseases-causing mutations were conserved residues which indicated at their importance in the biological functioning of these homologous proteins. It is also important to note that these TM helices also form the boundary of the hydrophobic tunnels inside the protein central cavity which could be implicated in determining the carbon chain length of the prenyl groups involved in the catalysis by acting as carbon chain rulers. An engineering of these residues could be the breakthrough which leads us to the biogenesis of novel biologically important lipids. The proposed allosteric activation of Tk-DGGGPS could be an energy saving maneuver in an archaeal cell, where only in presence of excess GGPP, the protein is activated for generating the di-geranylated archaeol precursors. Otherwise the protein

remains involved as a MFS transporter in regulating the archaeal membrane environment. This could be a very efficient feedback mechanism amongst others, which helps archaea to survive in harsh environmental conditions.

Chapter IV – Structural insights
into a membrane intrinsic
acyltransferase from *Chlorobium*
tepidum

4.1 Introduction

The uniqueness of the outer membranes of the Gram-negative bacteria lies in its asymmetric composition, which varies from all the other forms of life. Lipopolysaccharide (LPS) is a major component of this outer leaflet [128] and the hydrophobic anchor of LPS to the outer membrane is known as lipid A [129]. These LPSs shield the bacterium from harsh external environments and offer protection against the influx of harmful antimicrobial compounds by making it less permeable to such compounds. This functional barrier of sorts is thought to have two different aspects in its mode of operation, the inner lipid bilayer, composed of glycerophospholipids, instrumental in the exclusion of polar molecules, and the non-polar molecules effectively kept out by the closely packed polysaccharide domains of LPS. However, LPS is also a potent activator of the innate immune response and is recognized even at picomolar levels by the Toll-like receptor 4 (TLR4) [130].

Lipid A is produced in Gram-negative bacteria via a nine-step conserved pathway known as the Raetz pathway [60]. In this pathway, UDP-d-GlcNAc (Uridine diphosphate N-acetyl-d-glucosamine) is first converted to the UDP-diacetylglucosamine precursor of Lipid A, by the N- and O- linkage of the acyl groups donated by the R-3-hydroxymyristoyl acyl carrier protein (ACP). The secondary acylation of the Lipid A molecules is quite important since it has been observed that mutants lacking the genes responsible for this secondary acylation, namely, LpxL and LpxM, have been found to have increased permeability in the outer membrane and more sensitivity to cationic antimicrobial peptides (CAMPs). Particularly, LpxM has been found to have a direct impact on the virulence of the respective organisms, namely, the nosocomial pathogen, *Acinetobacter baumannii* [131], pathogenic *Escherichia coli* strains [132, 133], pathogenic strains of *Shigella flexneri* [134, 135], and many more.

Acyltransferases are a key family of enzymes, which catalyzes the transfer of acyl chains to different acceptor molecules. They are involved in the biosynthesis of triacylglycerols and 3-hydroxyalkanoic acids, the most important storage lipids found in eukaryotes and prokaryotes, respectively. Till date, very limited structural and biochemical studies have been carried out in this family of membrane intrinsic enzymes and despite the presence of structural data, the enzymatic mechanisms of these enzymes are still not elucidated completely. In this study, we have isolated

an acyltransferase from *Chlorobium tepidum* (strain ATCC 49652), belonging to the HtrB/ MsbB family. We have selected thermophile *C. tepidum*, as the expression of the membrane proteins from thermophile has relatively high success rate. We have purified this enzyme to homogeneity, biochemically characterized and attempted to predict its structure via homology modeling using known structures from the same family.

4.2 Materials and Methods

4.2.1 Isolation of Bacterial Acyltransferase Genes

The CT0211 gene was amplified by means of a Polymerase Chain Reaction by using primers specific to the 5- and 3-ends, respectively. The genome of *C. tepidum*, as a template, and AccuPrime™ Pfx DNA Polymerase were used in the reaction. The PCR conditions were a single cycle of pre-denaturation at 95°C (5 minutes) followed by 28 cycles of denaturation at 95°C (30 seconds), primer annealing at 58°C (30 seconds) and extension at 68°C (1 minute), a final extension at 68°C (10 minutes) and then cooled to 4°C. The PCR amplified products were analyzed on 1% agarose gel. The restriction sites in the primers, the NdeI and NotI sites, are highlighted in bold and ligation was carried out with the NdeI–NotI-digested pET-28a vector. The same gene was also cloned into a modified pET-28a plasmid, with a fusion of apocytochrome b562 (BRIL). Detailed information about the primers used are tabulated in Table 2.1 and the composition of the PCR mix is mentioned in Chapter 2 (section 2.2.2). The ligated vectors were then transformed in competent DH5α cells and the colonies obtained were screened against any possible false negatives by colony PCR and restriction digestion simultaneously. Once successfully cloned in the desired pET28a and pET28a-BRIL vectors, the clones were stored at – 80 °C. The samples were also sent for sequencing to check the integrity of the sequence.

4.2.2 Expression and Purification of the Recombinant Enzymes

Escherichia coli BL21(DE3) cells and C41 cells were transformed with each of the above-mentioned plasmids, to do a comparative study of the yield of our desired protein from either cell. They were cultivated in 6X1 L batches of 2XYT broth supplemented with kanamycin (50 mg/L)

at 37 °C. The cultures were induced with 1.0 mM Isopropyl β -d-1- thiogalactopyranoside (IPTG), when the A600 of the culture reached 0.4 at 37 °C. The cultures were then grown overnight and the cells were harvested and disrupted by homogenization in Lysis Buffer C1, pH 8.0 containing 1mM PMSF and PIC. The homogenate was spun in an ultracentrifuge at 125,000 \times g for 50 min at 4 °C, and the supernatant was removed. The pellet was re-dissolved in Solubilizing Buffer C2, pH 8.0 into a homogenous solution, using a hand homogenizer and excess detergent Decyl β -d-maltopyranoside (DM) was added to the 1% (w/v) solution and incubated in a mixer overnight. The homogenate was then subjected to a second round of ultracentrifugation at 150,000 \times g for 40 min at 4 °C and the supernatant was collected. The supernatant fraction was further purified by affinity chromatography using a Ni NTA column pre-equilibrated with Wash Buffer C3, pH 8.5. The column was washed with the wash buffer and then an Elution Buffer C4, pH 8.5 was used to elute the 6XHis tagged protein. An SDS-PAGE was carried out to check the purity of the eluted protein. Size exclusion chromatography was performed thereafter using these pooled fractions in the Superdex 200 column, using Purifying Buffer C5, pH 8.5 to further purify the protein to homogeneity and the peaks obtained were analyzed in SDS-PAGE. A western blot analysis was further carried out using His tag-specific primary antibodies (from Qiagen) and HRP-conjugated anti-mouse secondary antibodies (from Sigma) to confirm the isolated protein.

Table 4.1 Details of the primers used in the study

Primers	Primer sequence
Ct-Acyltransferase	
Forward primer (Ct-Acyltransferase-NdeFor)	5' ATT AAT CAT ATG ATG TGG CCC CTC TTC GCC CTG CTC 3'
Reverse primer (Ct-Acyltransferase-NotRev)	5' TAT AAT TGC GGC CGC TTA CGA GGG CGA GGT AGG GGC CGA AGG 3'
Ct-Acyltransferase-BRIL	
Forward primer (Ct-Acyltransferase-BRIL-BamFor)	5' ATT AAT GGA TCC ATG TGG CCC CTC TTC GCC CTG CTC 3'
Reverse primer (Ct-Acyltransferase-BRIL-NotRev)	5' TAT AAT TGC GGC CGC CGA GGG CGA GGT AGG GGC CGA AGG 3'

4.2.3 Ultracentrifugation Dispersity Sedimentation Assay

An ultracentrifugation dispersity sedimentation (UDS) assay was also carried out to check if the membrane protein was stable in the PDC (protein-detergent complex). The UDS assay is principled on the hypothesis that once protein aggregates form, they become orders of magnitude heavier than the normal dispersed state of the protein particles and hence sedimentation at high 'g'

forces can be used to remove them. All ultracentrifugation steps were performed on a Beckman Coulter Optima™ XPN 100 K centrifuge at 4 °C and detergent (DM) was used at 3× their CMC of 0.1% (w/v). In the first step of the UDS assay, initial step of ultracentrifugation (100,000×g; 50 min) was carried out using a 45-Ti (Beckman Coulter) rotor. After the ultracentrifugation step was carried out, a 1-mL control sample of the protein is removed from the cell pellet re-suspension (using TSG 100) after overnight incubation with DM, mixed with SDS-PAGE loading buffer, and stored. In the second step of UDS assay, a second round of ultracentrifugation (100,000×g; 40 min) was carried out using a 50-Ti (Beckman Coulter) rotor, to ensure the removal of any aggregates that might have accumulated during the purification. A sample was collected after storing the supernatant overnight and analyzed in SDS-PAGE using a standard dual color Bio-Rad protein marker.

Lysis Buffer C1 (pH 8.0)

- 50 mM Tris
- 300 mM NaCl
- 10% (v/v) glycerol

Solubilizing Buffer C2 (pH 8.0)

- 50 mM Tris
- 100 mM NaCl
- 10% (v/v) glycerol

Wash Buffer C3 (pH 8.5)

- 50 mM Tris
- 100 mM NaCl
- 25 mM Imidazole
- 1 mM DTT
- 0.1% (w/v) DM
- 10% (v/v) glycerol

Elution Buffer C4 (pH 8.5)

- 50 mM Tris
- 100 mM NaCl
- 250 mM Imidazole
- 1 mM DTT
- 0.1% (w/v) DM
- 10% (v/v) glycerol

Purifying Buffer A5 (pH 8.5)

- 50 mM Tris
- 100 mM NaCl
- 1 mM DTT
- 0.1% (w/v) DM
- 10% (v/v) glycerol

4.2.4 Sequence and Phylogenetic Analysis

Protein sequence of CT0211 of sequence length 278 amino acids was obtained from the UniProtKB/Swiss-Prot database (<http://www.uniprot.org/>). The identified protein sequences were first aligned in the software Molecular Evolutionary Genetics Analysis 5.1 (MEGA) [136], using the algorithm named Multiple Sequence Comparison by Log- Expectation (MUSCLE). Initial tree(s) for the heuristic search was obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value.

4.2.5 Molecular Modeling of *C. tepidum* Acyltransferase

A search using Basic Local Alignment Search Tool (BLAST) algorithm was carried out against the Protein Data Bank (PDB) to select the high-resolution crystal structures of homologous proteins to the *C. tepidum* acyltransferase gene. The sequence identity cut-off was set to $\geq 30\%$ (E value cut-off=1). Homology modeling was carried out using the MODELLER [137] by taking structures homologous to the target protein as templates, namely, 5F2T, 5F2Z, 5F31, 5F34, 5KN7, and 5KNK in order to study their structural architecture, possible binding pockets, and affinity towards its substrates. The models obtained were evaluated for the stereochemical quality of the protein backbone and side chains using PROCHECK [93]. Environment of the atoms in the protein model was checked by ERRAT server [94]. Verify3D plot showed the compatibility of the 3-D structure with respect to the protein sequence [95]. Errors in the model structures were also checked with ProSA server [138]. After model validation, initial models were refined using impref minimization of Protein Preparation Wizard of Maestro [91], Schrödinger, followed by minimization and finally loop refining in Prime, Schrödinger [92].

4.3 Results

4.3.1 Isolation of a Novel Acyltransferase Gene of *Chlorobium tepidum*

In the entire genome sequence of *C. tepidum*, no gene was found corresponding to the already identified LpxL and LpxM genes found in *E. coli*, which are known to encode the enzymes that transfer laurate and myristate residues from respective ACPs to KDO bound IVA molecules. Therefore, the CT0211 gene transcripts can be predicted to transfer both the above-mentioned residues to generate the Lipid IVA molecules in this Gram-negative bacterium. The CT0211 gene has a significant and almost similar identity of 25% with the LpxL and LpxM genes of *E. coli*. The result of an ortholog analysis using the Microbial Genome Data Base web server (<http://mbgd.genome.ad.jp/>) indicated that CT0211 is indeed an ortholog of *E. coli* LpxL and LpxM, suggesting that the enzyme is likely the acyltransferase involved in cell wall biosynthesis in *C. tepidum*. The ORF, encoding a protein of 278 amino acids, was amplified by PCR from the genome of *C. tepidum* (Fig 4.1 A), and the amplified DNA fragment was digested with restriction enzymes and inserted into pET28a and pET28a-BRIL respectively. The restriction sites used for

pET 28a were NdeI and NotI while for pET28a-BRIL the BamHI and NotI sites were used. The positive clones (Fig 4.1 B, C) were selected by colony PCR and plasmids were isolated. Restriction digestion of respective plasmids indicated the presence of a 933 bp insert cloned in the plasmid. The positive clones were maintained separately and plasmids were isolated and sent for sequenced. Sequencing results confirmed that Ct-acyltransferase was cloned and devoid of mutation in both the pET28a and pET28a-BRIL plasmids.

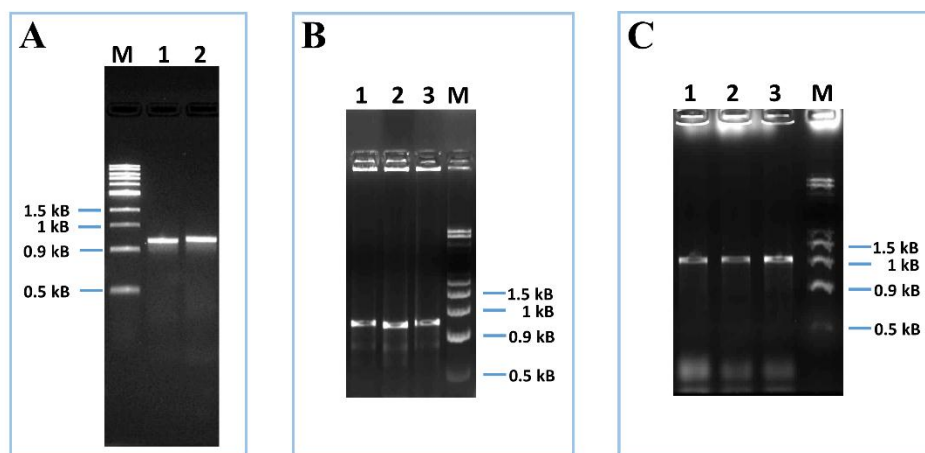


Fig 4.1 Cloning of CT0211 in pET28a and pET28a-BRIL plasmids (A) PCR amplified product of Ct-acyltransferase in 2% agarose gel. Lane M: NEB 2Kb plus DNA ladder; Lane 1&2: Amplified gene product, (B) Colony PCR for confirmation of Ct-acyltransferase clones. Lane 1,2 & 3 are the positive clones; Lane M is the DNA ladder, (C) Colony PCR for confirmation of Ct-acyltransferase-BRIL clones. Lane 1,2 & 3 are the positive clones; Lane M is the DNA ladder.

4.3.2 Expression and Purification of Recombinant *Chlorobium tepidum* Acyltransferase

The recombinant acyltransferase protein with an N terminal 6XHis tag and its BRIL fusion construct was expressed in the E. coli BL21 Star (DE3) cells. Different detergents were used on a trial basis to determine which one extracted the protein most efficiently from the pellet fraction and kept it most stable. Notable success was obtained using DDM (n-Dodecyl- β -d Maltopyranoside). The protein was extracted from the cell membrane fraction of the cell pellet homogenate using 1% DDM after a single round of ultracentrifugation and subjected to another round of ultracentrifugation wherein the membrane protein was obtained in solution in the form of protein detergent complex (PDC). A two-step purification protocol, using Ni-NTA followed by size exclusion chromatography (SEC) was employed for Ct-acyltransferase purification. Ni-NTA

purification was carried out in buffers containing different detergents at 2-3X concentrations of their CMCs (critical micelle concentrations) like 0.1% DM, 0.03% DDM, 0.7% β -OG and 0.2% LDAO, but the best results were obtained using DM. SDS-PAGE analysis of Ni-NTA affinity chromatography purified fractions on a 12% acrylamide gel indicated relatively pure protein bands at ~37 kDa. The BRIL fusion protein showed bands at ~50 kDa, at a ~12 kDa difference with the wild type, which is about the exact known molecular weight of the BRIL tag. (Fig 4.2 A, B)

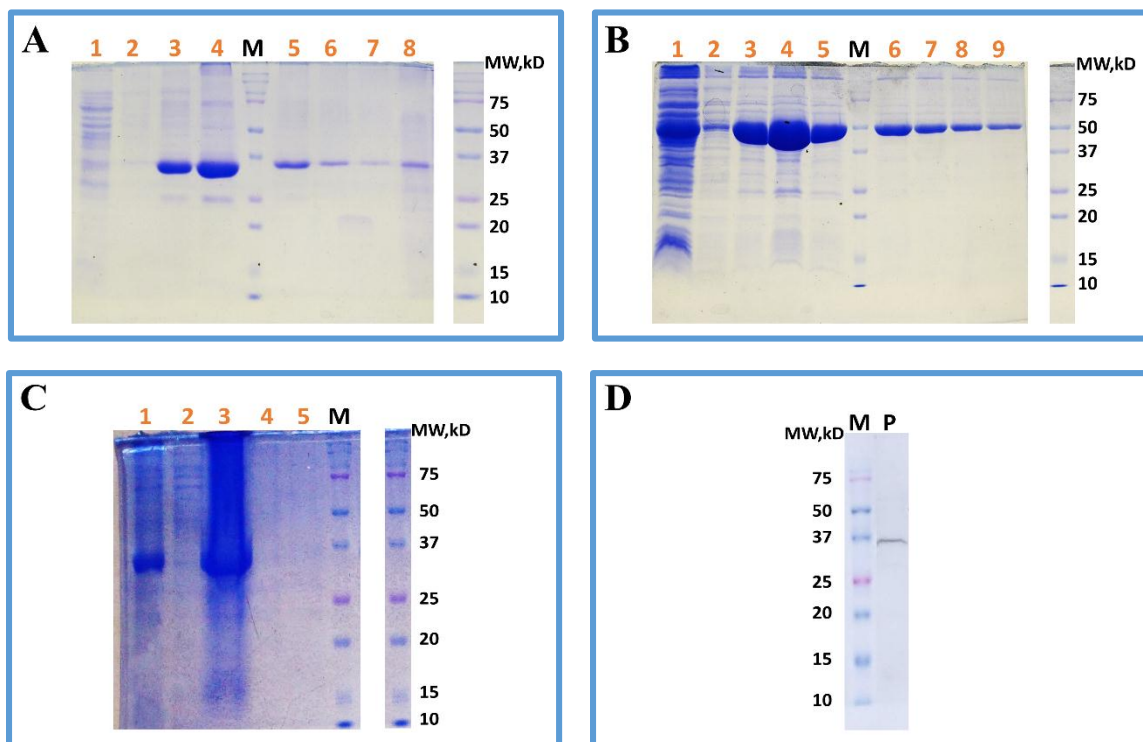


Fig 4.2 Purification of Ct-acyltransferase (A) SDS-PAGE analysis showing expression of Ct-acyltransferase. Lane 1: Unbound fraction; Lane 2: Wash fraction; Lane 3–8 Elution fractions; Lane M: Dual Protein marker, (B) SDS-PAGE analysis showing expression of Ct-acyltransferase-BRIL. Lane 1: Unbound fraction; Lane 2: Wash fraction; Lane 3–8 Elution fractions; Lane M: Dual Protein marker, (C) SDS-PAGE of the samples from each step of the UDS assay indicating that the PDC is stable for the duration of the assay. Lane 1: Sample from Step 1; Lane 3: Sample from Step 2; Lane 2,4&5: Blank; Lane M: Dual Protein marker, (D) Western blot analysis of purified Ct-acyltransferase.

Size exclusion chromatography (SEC) of the pooled positive Ni-NTA fractions showed a prominent peak with one broader peak in its elution profile indicating a heterogeneous oligomeric nature of the protein (Fig 4.3). However, a comparison with a standard run doesn't give us the correct estimate of the molecular weight of the protein due to the presence of excess DDM micelles. The western blot analysis using His-tag specific primary antibodies (PentaHis, Qiagen)

confirmed that, the isolated protein was the desired acyltransferase from *C. tepidum* (Ct-acyltransferase) (Fig 4.2 D). A comparative study of the protein yield obtained in either case indicated that the BRIL tag increases the yield in case of this protein. However, when stored for a few days at 4 °C, the BRIL fusion protein was relatively stable without any visible precipitation. This probably meant that the addition of the BRIL segment might also have stabilized the protein better than its wild type. The UDS assay samples were analyzed in SDS-PAGE. The band intensity was similar between the control and DM-solubilized fractions which indicated the stability of the protein in 0.1% (w/v) DM (Fig 4.2 C).

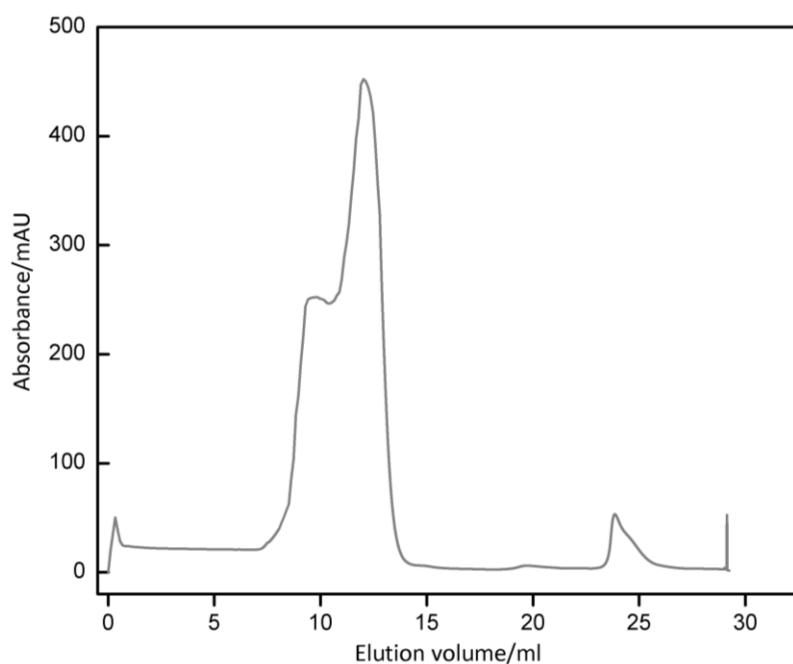


Fig 4.3 SEC profile of Ct-acyltransferase Chromatogram of SEC purified Ct-acyltransferase showing a broader peak to the main peak, indicative of possible soluble aggregate formation.

4.3.3 Sequence analysis of Ct-acyltransferase and other LPLAT superfamily members

There are certain enzyme families which contain both soluble and membrane intrinsic proteins. Even though the overall spatial alignment of charge varies in such proteins, they maintain a conserved core structure. This conservation is only possible due to the presence of identical or similar amino acid residues which make up the specific functional domains of the protein. So, a comparison of the amino acid sequences between the members of the family might help to identify the conserved residues which can be the key to understanding the structure function dynamics of

these family of enzymes. The presence of only one acyltransferase of the HtrB/MsbB family in *C. tepidum* instead of two, as seen in the case of many Gram negative bacteria, indicated that it could be an enzyme that capable of carrying out both the secondary acylations of KDO₂-lipid IVA and hence it was interesting to study its amino acid sequence in more detail. Differentially conserved residues obtained by sequence alignment between these acyltransferases could possibly give insights into how these conservations have impacts on the enzyme mechanism and highlight their importance to the overall structure of these enzymes as well.

The sequences obtained from a Domain Enhanced Lookup Time Accelerated BLAST (DELTA-BLAST) and Position-Specific Iterated BLAST (PSI-BLAST) against Ct-acyltransferase (Uniprot ID: Q8KFBV9), further annotated from Pfam database and the Microbial Genome Data Base web server (<http://mbgd.genome.ad.jp/>), gave us an interesting array of putative homologs from a vast variety of organisms ranging from bacteria to humans. In plants it included the lysophospholipid acyltransferases (LPEATs), the N-acylphosphatidylethanolamine synthase (NAPE synthases), the 1-acyl-sn-glycerol-3-phosphate acyltransferases (AGPATs) and the glycerol-3-phosphate acyltransferases (GPATs) which are functionally distinct but evolutionarily linked. In bacteria the glycerol-3-phosphate acyltransferases (GPATs), the 1-acyl-sn-glycerol-3-phosphate acyltransferases (AGPATs), the fatty-acyl transferases (HtrB/MsbB family), the Vi polysaccharide export protein (VexE) and the Bifunctional protein Aas (having the lysophospholipid acyltransferase and the Acyl-ACP synthetase domains) are the various homologs that show unique substrate specificities and functions, but have distinct sequence similarities. In slime molds a putative lysophosphatidylcholine acyltransferase homolog was identified. In insects, lysophosphatidylcholine acyltransferases (LPCATs) and Tafazzin homologs (involved in cardiolipin biosynthesis) were indicated as possible homologs. In humans and other mammals, the Dihydroxyacetone phosphate acyltransferase (DHAP-ATs), the lysophosphatidylcholine acyltransferases (LPCATs), the 1-acyl-sn-glycerol-3-phosphate acyltransferases (AGPATs) and Tafazzin (involved in cardiolipin metabolism) were identified as homologs.

Multiple sequence alignment of the representative sequences of all the homologs of the LPLAT superfamily showed that there was very high sequence variation between the proteins which is the stark uniqueness of this family. However, some conserved residues were identified which were grouped into clusters (Fig 4.4).

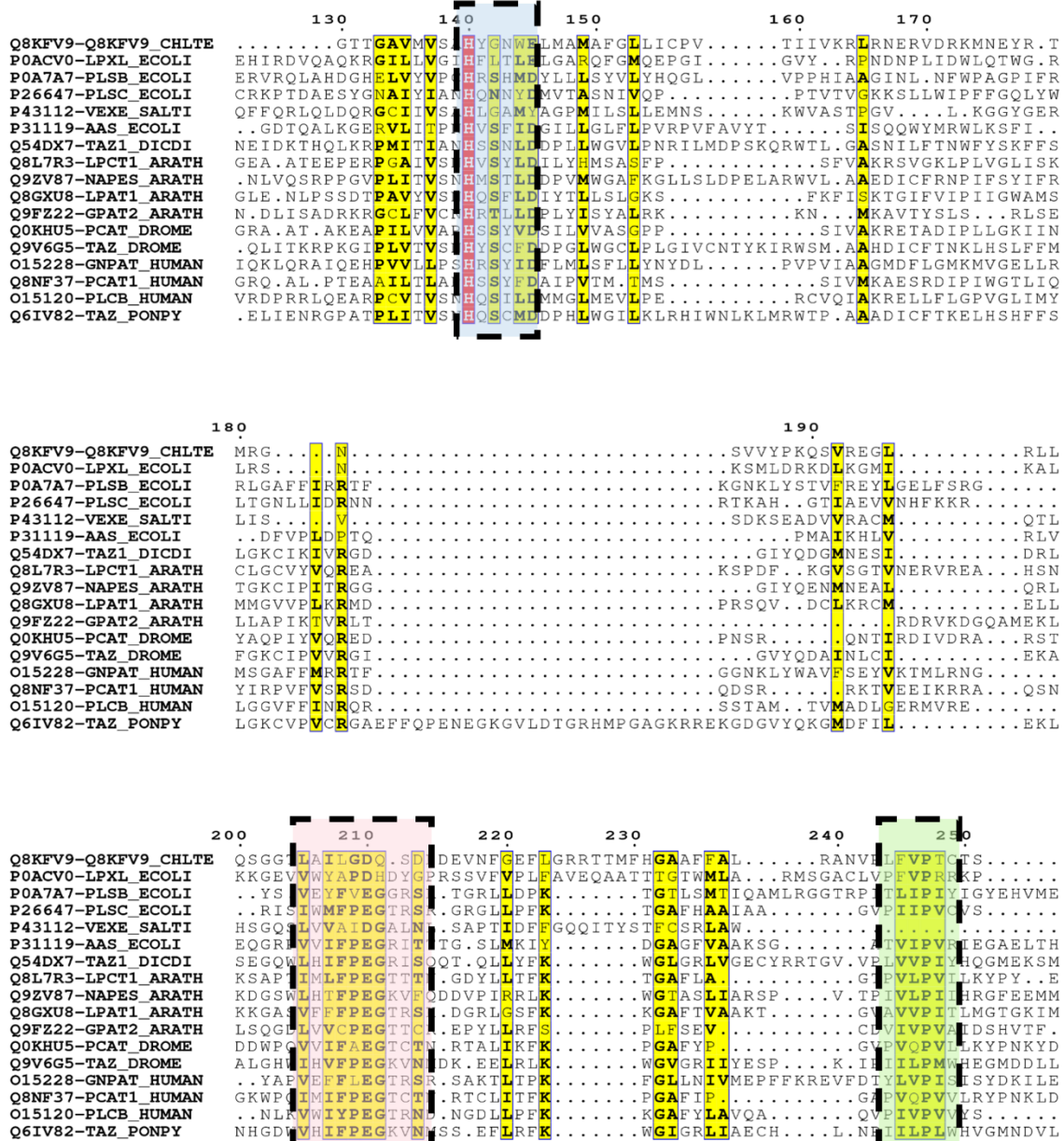


Fig 4.4 Multiple sequence alignment of the Ct-acyltransferase homologs belonging to the LPLAT superfamily

In an otherwise highly diverse sequence pool, the dotted box highlighted in blue depicts the HxxxxD/E conserved motif present in all ‘real’ acyltransferases and the dotted boxes in red and green represent other fairly conserved residue patches which could be possible conserved motifs as well. The box highlighted in red is centered around a highly conserved Asp/Glu residue while the one in green is centered around a moderately conserved Pro residue.

The first cluster included the entirely conserved histidine and the highly conserved aspartate/glutamate residue arranged in the order, HxxxxD/E, which was a known conserved motif for acyltransferases of this family.

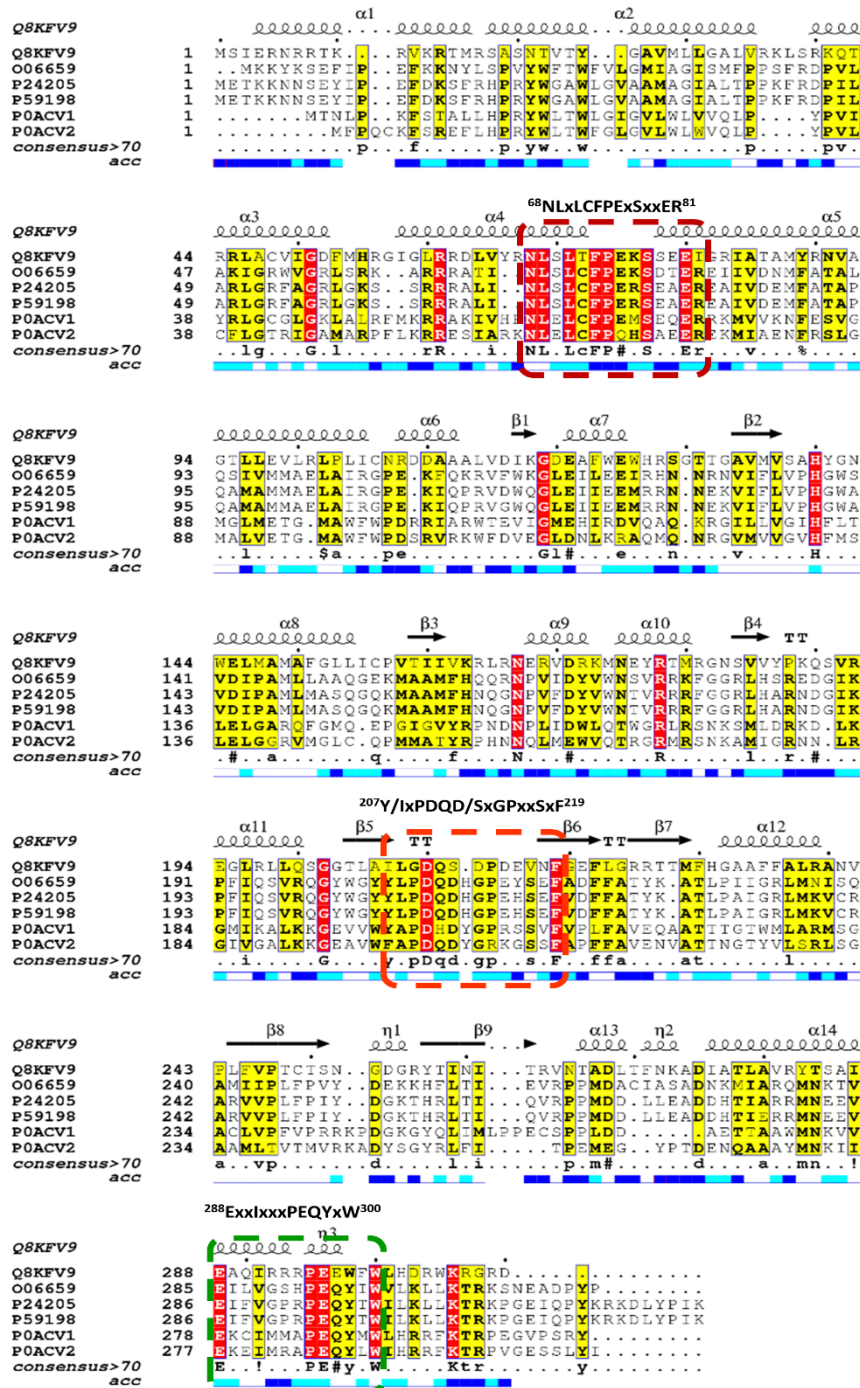


Fig 4.5 Multiple sequence alignment of the HtrB/MsbB family of acyltransferases. The overall residue conservation was prominently more than the entire LPLAT superfamily and three prominent residue clusters obtained are depicted here in maroon, orange and green dotted boxes, respectively.

The second cluster had a highly conserved aspartate/glutamate residue with other notably conserved residues around it. We could also identify a third cluster where a moderately conserved

proline residue was located. These clusters are all located in the core regions of the protein which indicated that even though there was prominent variation in overall sequences of these acyltransferases, the central α/β domain hosted most of the similar and conserved sequences. However, the sequence variation should be comparatively lower among similar types of acyltransferases. To verify the above statement, multiple sequence alignment of the representative sequences of the HtrB/MsbB family of acyltransferases were carried out. It was observed that the amino acid residue conservation was prominently more as expected. The conserved residues were also seen to be grouped in clusters which could be putative functional motifs in these enzymes. Apart from the well-known conserved motif, $^{140}\mathbf{H}_{xxxx}\mathbf{D/E}^{145}$, three other clusters were identified, Cluster 1 ($^{68}\mathbf{NLxLCFPExSxxER}^{81}$), Cluster 2 ($^{207}\mathbf{Y/IxPDQD/SxGPxxSxF}^{219}$) and Cluster 3 ($^{288}\mathbf{ExxIxxxPEQYxW}^{300}$). The significance of these clusters were checked using the CONSURF analysis which predicted the residues in bold to be highly conserved structural and functional residues (Fig 4.5).

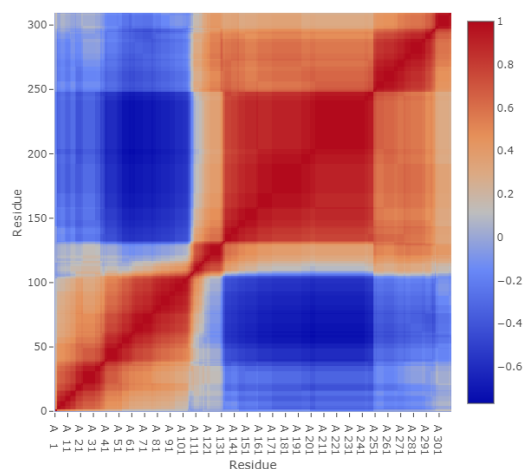


Fig 4.6 Match correlation matrix of Ct-acyltransferase This was generated by the DALI server and shows that there are four prominent subdomains in the Ct-acyltransferase which in addition to the fact that Ct-acyltransferase holds structural similarity to four known crystal structures belonging to functionally different LPLAT acyltransferases make a strong case for this protein being an ancestral protein among the HtrB/MsbB family of acyltransferases.

The Ct-acyltransferase was further analyzed using the HHPRED webserver (<https://toolkit.tuebingen.mpg.de/tools/hhpred>) [89] which employs sequence database searching on basis of pairwise comparison of profile hidden Markov models (HMMs) and verifies it with the predicted secondary structure similarity, available from protein crystal structure data in PDB to

check if it has any remote homologs whose sequence identity is even below 20%. The analysis results showed prominent homologies with four known LPLAT acyltransferase structures (Ab-LpxM-5KNK, Ms-PatA-5F2T, Tm-AGPAT-5KYM and Cm-GPAT-1IUQ) with probabilities above 99% and E-value in higher exponential of negatives. This was further confirmed using a DALI server [139] search where the match correlation matrix had anti-correlated blocks indicating the presence of four prominent subdomains in the protein (Fig 4.6). Therefore, it could be concluded that in spite of having high sequence variability the LPLATs have core conserved secondary structures which are maintained by different residues. These changes might have taken place in course of divergent evolution that these acyltransferases might have gone through while adapting to new roles in different biochemical pathways.

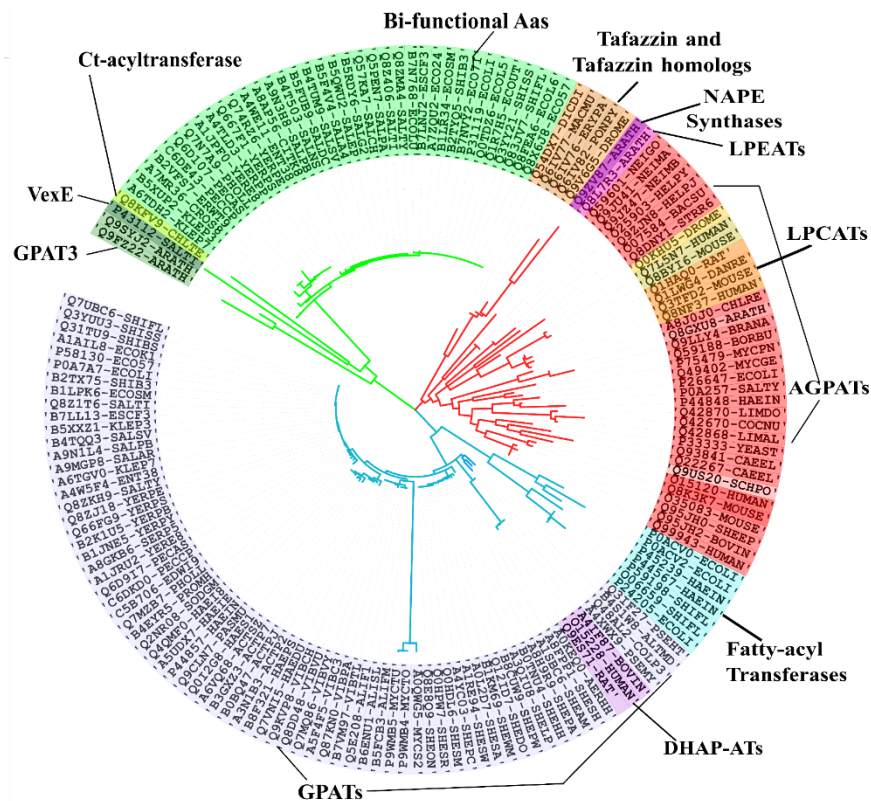


Fig 4.7 Phylogenetic tree of LPLAT acyltransferases A Maximum Likelihood tree generated by the MEGA suite using sequences of LPLAT acyltransferases showing prominent clustering of GPATs, AGPATs, fatty-acyl transferases and Bi-functional Aas proteins. Notably, the Ct-acyltransferase clusters closely with the bi-functional proteins further supporting the claims of being a multifunctional acyltransferase.

4.3.4 Phylogenetic Analysis of the LPLAT acyltransferases

To study the evolutionary relationships between Ct-acyltransferase and other members of the LPLAT acyltransferase superfamily, a phylogenetic tree was constructed using the multiple sequence alignment of the representative members. The members were comprised of just the acyltransferase domain or combination of the glycosyl transferase-2 domain and the acyltransferase domain. The sequences got clustered in three major clades containing 6 sub clusters. The major clade comprised of predominantly Glycerol-3-phosphate acyltransferases (GPATs), with the mammalian AcylCoA:Dihydroxyacetone phosphate acyltransferases (DHAP-ATs) and the Lipid A biosynthesizing lauroyl, palmitoleoyl, and myristoyl transferases consisting of the lesser populated members in what appears to be the least diverse of the clades. The second clade was found to consist of 1-acylglycerol-3-phosphate O-acyltransferases (AGPATs) most prominently with Lysophosphatidylcholine acyltransferases (LPCATs), Lysophospholipid acyltransferases (LPEATs), N-acylphosphatidylethanolamine (NAPE) synthases, Tafazzin and Tafazzin homologs populating the rest of the clade making it the most diverse among the three. The bi-functional protein Aas including the domains 2-acylglycerophosphoethanolamine acyltransferase and Acyl-ACP synthetase, the Vi polysaccharide export protein (VexE), glycerol-3-phosphate acyltransferase 3 (GPAT3), and the Ct-acyltransferase was found to be grouped in the other clade where the Aas forms the major component (Fig 4.7). Interestingly, the Ct-acyltransferase clusters distinctly from its orthologs and paralogs, i.e., the Lipid A fatty-acyl transferases, but at the same time it is found to group together with the only other bi-functional proteins (having two functional domains) of this superfamily of enzymes. The Ct-acyltransferase can hence be predicted to have a putative dual functional nature, since it is the only acyltransferase involved in the transfer of laurate/myristate groups in the biosynthesis of Lipid A in *C. tepidum*.

It is important to note that *C. tepidum* is a non-pathogenic strain, and Lipid A is highly implicated as an endotoxin in pathogenic Gram negative bacteria. Hence there is high probability that the Ct-acyltransferase could be involved in other functional capacities its respective organism. This is also partially supported by the above mentioned HHPRED results, which suggested, with probability scores of more than 99%, that there were structural similarities with three acyltransferases with different functions than the HtrB/MsbB family enzymes. The distinct point

of divergence of the Ct-acyltransferase from the ancestral tree also makes it a candidate for one of the probable ancestral genes among acyltransferases.

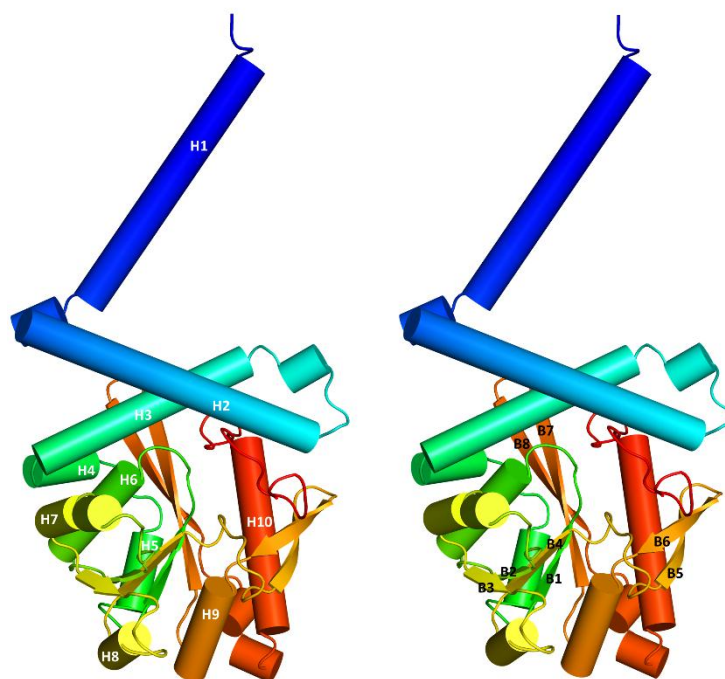


Fig 4.8 Overall structure of Ct-acyltransferase Cartoon representations of the Ct-acyltransferase homology model are shown, with the helices depicted as pipes. Among the total count of ten α -helices and six β -strands, six α -helices and four β -strands compose the core α/β domain of the protein.

4.3.5 Homology Modeling: Molecular Models of Ct-Acyltransferase

The molecular structures of acyltransferase belonging to the HtrB/MsbB family from *C. tepidum* were modeled in order to study how it incorporated the above mentioned domains in its overall three dimensional structure. Further, the relevance of the core α/β domain which is found to be conserved among acyltransferases along with the pore lining helix residues (Ala87–Leu102) were also to be studied in greater detail. Detailed analysis of the structure could reveal the key behind the intriguing absence of conserved residues, even near the predicted ligand binding site throughout the family and their role in ligand binding. Also a prediction of the putative enzymatic mechanism of this protein was attempted from the location of conserved residues using the CONSURF analysis. The initial few residues could not be modeled, due to the absence of the corresponding regions in the template sequences. The model was refined in the GalaxyWeb server and verified in PROCHECK and ERRAT servers. The Ramachandran plot of the final model

showed 83.0% of the amino acid residues to be in the most favored regions, 13.7% residues in additional allowed regions, 2.5% in generously allowed regions and 0.7% residues in disallowed regions. The RMSD of the C-alpha atoms of the models and the templates was ~ 2.7 Å. A z score of -7.15 indicated that the model had a z score similar to experimentally determined PDB structures (-7.75 for one of the templates 5KN7). The overall quality factor as determined by ERRAT for the homology model was 97.1% indicating that the model was well refined and there were very few unusually close non-bonded contacts (Fig 4.9).

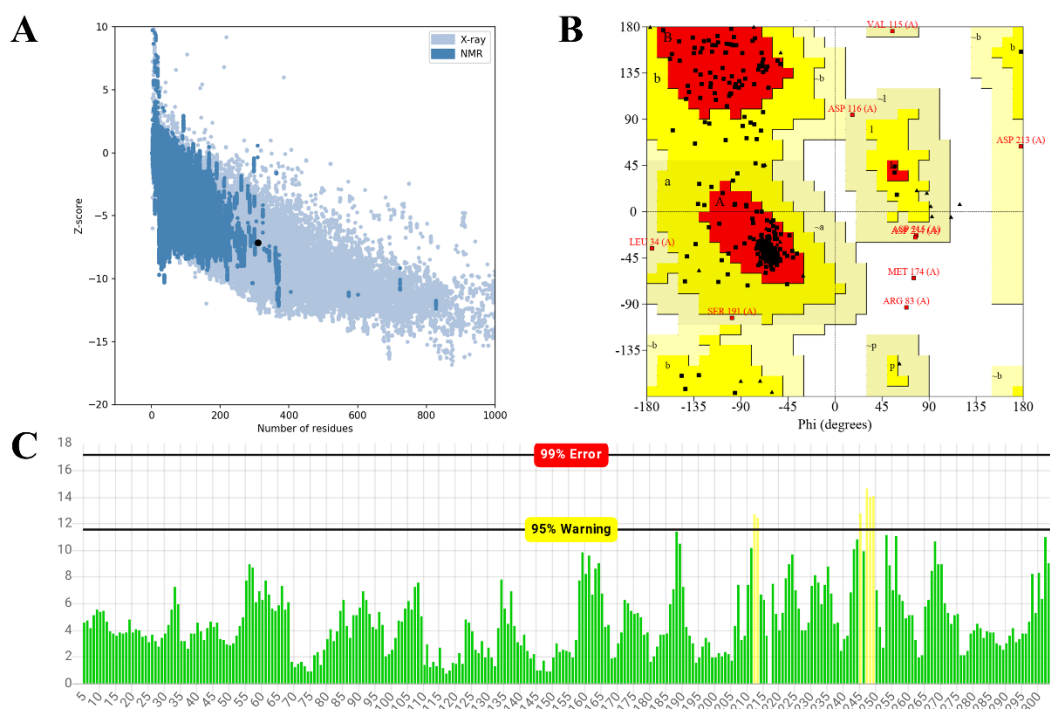


Fig 4.9 Particulars of the Ct-acyltransferase homology model (A) Z score from PROSA analysis, (B) Ramachandran plot, (C) The ERRAT quality score plot showing the homology model to be energetically compliant without major steric clashes among residues.

4.3.6 Salient structural characteristics of Ct-acyltransferase

The model depicts the Ct-acyltransferase to have a globular structure, the core of which is comprised of a pair of parallel and antiparallel β strands along with four parallel β strands and six α helices, composing the α/β domain (Fig 4.8, 4.12). These strands are composed of residues that are fairly conserved. The helices make up almost the entire external surface of the enzyme structure (Fig 4.10) with most of it being the solvent exposed areas of the protein as well. The helices also

contain most of the acidic and basic amino acid residues, including the one involved in the catalytic dyad, Glu145. Notably it was observed that the conserved acidic and basic residues were strategically placed in the overall protein structure. It was observed that three of the conserved basic amino acid residues and two of the conserved acidic residues were located in the binding cleft of the protein (Fig 4.13) which indicates that they might play an important role in the acyl-transfer reaction catalyzed by these enzymes. The presence of polar conserved residues all over the protein solvent accessible surface indicated that they notably influence the overall protein stability and functionality of the enzyme.

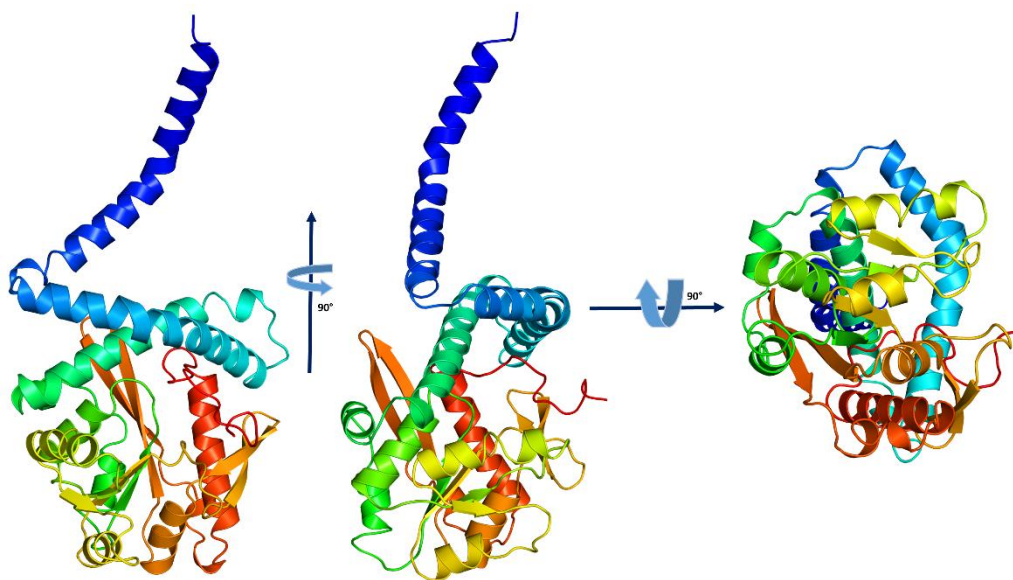


Fig 4.10 Overview of the Ct-acyltransferase structure Cartoon representation of the structure, color-ramped from the N to C terminus (blue to red) and showed from a frontal view, at a rotation of 90° in the same plane and from the bottom of the protein.

In the **HxxxxD/E** motif, His140 was present in a loop at the core of the predicted site of catalysis of the enzyme. In particular, Asp210, along with the catalytic dyad component composing the characteristic **HxxxxD/E** motif, appeared to have important roles in the charge relay system initiated by the catalytic core. The possible mechanism would be histidine extracting the proton from the hydroxyl group, thereby activating the oxygen for a nucleophilic attack on the acyl thioester carbon. Interestingly, a trio of positively charged residues Lys163, Arg178, and Arg181 were visualized and seemed to be strategically placed in the binding pocket of Ct-acyltransferase among which Lys163 and Arg178 were prominently conserved (Fig 4.11, 4.13, 4.16). Their

positioning was very close to the putative catalytic core residues among which Arg178 was located almost adjacent to the catalytic histidine and could be possibly involved in the charge transfer mechanism, while Lys163 is a notable difference from its closest PDB relative (Ab-LpxM; 5KN7) [55]. It formed a bridge across the binding cleft and showed some similarity in this aspect to the PatA (Ms-PatA; 5F2T) structure [53] and might be involved in the opening and closing of the binding pocket as well during catalysis. Among other conserved residues, the positively charged Arg61 could have possible roles in the substrate binding of the phosphate group of the glucosamine moiety of the Lipid IVA precursors. Another important conserved residue was K306, which forms a positively charged patch at the opening of a tunnel on the protein surface (Fig 4.13). There was also a branched cavity near the opening of the binding pocket and was connected by the above mentioned tunnel to the protein surface. The walls of this cavity were composed of hydrophilic residues and could be another acyl-ACP binding site in the enzyme. This site could in turn be responsible for the acyl-protein thioesterase activity, proposed in Ab-LpxM. Conserved divalent metal binding sites were also determined based on CONSURF and PSIPRED analysis. However, they were not located in or near the binding pocket, indicating that metal binding could be necessary for the structural stability of the enzyme and might not have a prominent role in catalysis.

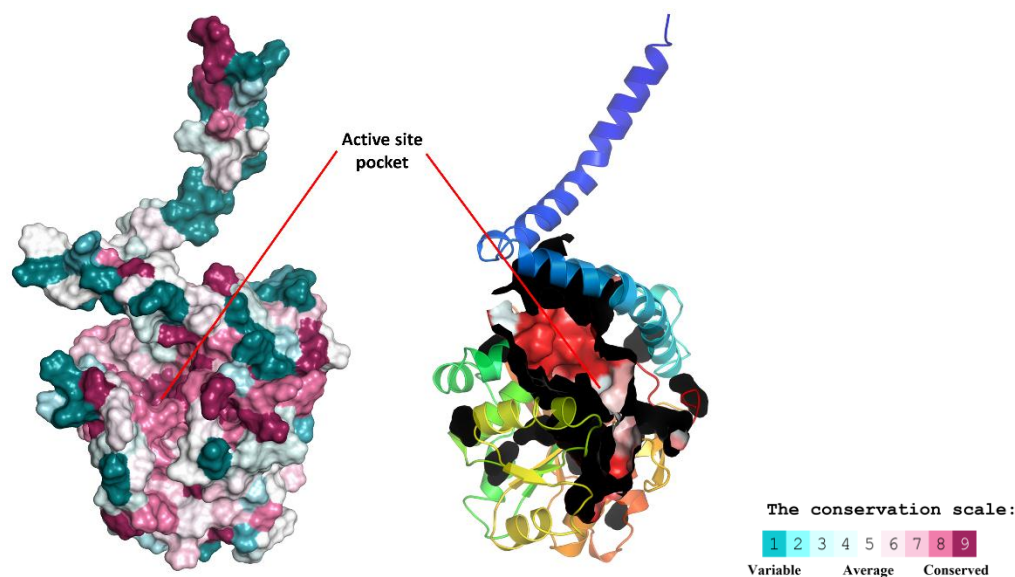


Fig 4.11 The Ct-acyltransferase enzyme active site pocket A surface conservation analysis and a hydrophobic surface analysis showing the architecture of the active site cavity of Ct-acyltransferase. The active site is seen to contain a high number of conserved residues which include both hydrophobic and hydrophilic residues.

PSIPRED analysis also revealed that the enzyme was partially embedded into the membrane by its N terminal helices and the pore lining helix. The N terminal was towards the cytoplasmic side of the membrane while the C terminal was extracellular in nature (Fig 4.14). This is agreeable with the location of the enzymatic reaction as well since Lipid A is generated on the outer membrane. The predicted active site of the Ct-acyltransferase has been already reported in a related fatty-acyl transferase (5KN7). The notable differences in the active sites of the Ct-acyltransferase and Ab-LpxM enzymes only seem to support the fact that there is only one fatty-acyl transferase in *Chlorobium tepidum*, in contrast to other Gram-negative strains. Unlike the active site of the *A. baumannii* ortholog, this one has a dual-storied architecture. The active site is composed of a deep hydrophobic pocket that is the probable binding site of the acyl chain and it is distinctly divided into two halves by a belt of hydrophilic residues (Fig. 4.13). This hints at a probable mechanism that is different from the recently predicted mechanism of the Lipid A secondary acyltransferase LpxM from *A. baumannii*. Here the attachment of the two acyl chains might take place simultaneously in each sub-division of the active site cleft, instead of being separate processes [140].

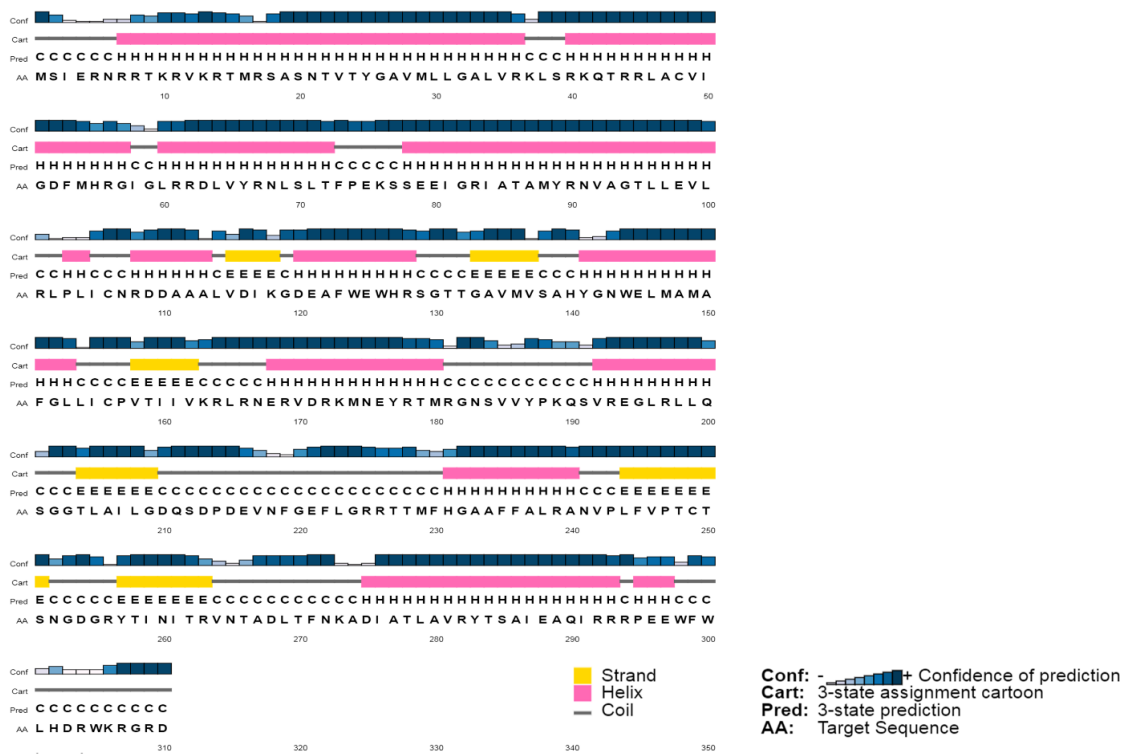


Fig 4.12 A schematic diagram of the structural components of Ct-acyltransferase

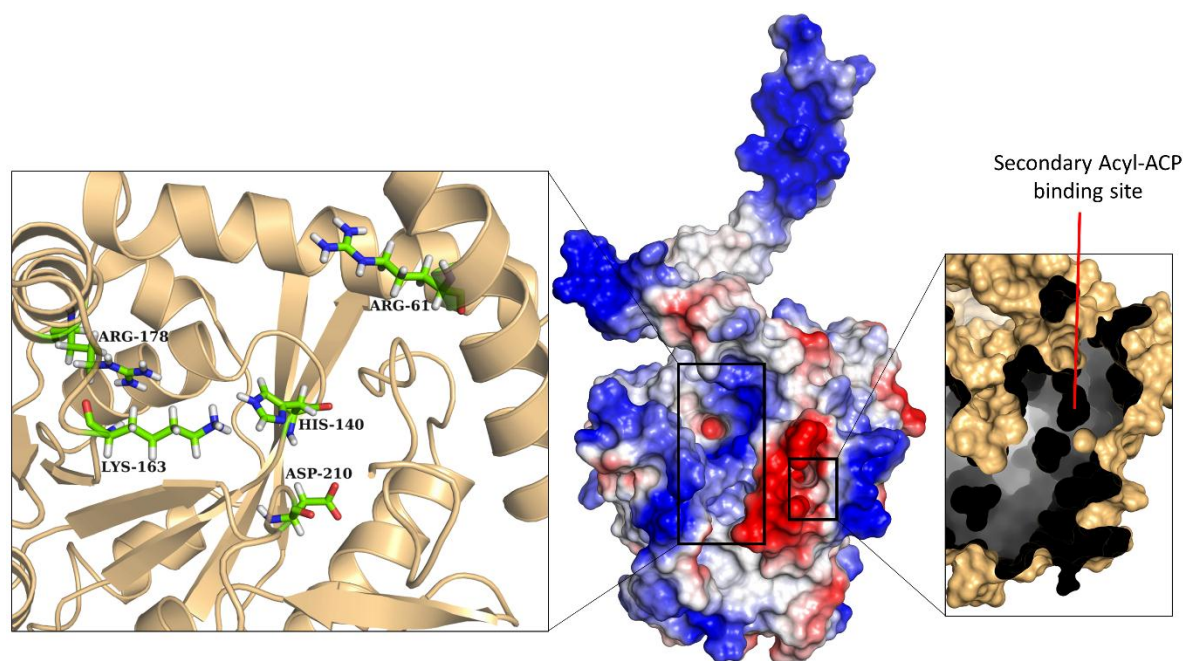


Fig 4.13 Molecular details of the Ct-acyltransferase active site pocket The electrostatic surface potential of the protein depicting negative residues in red and positive residues in blue showing that the active site pocket contains predominantly positively charged residues and the secondary ACP binding site has mostly negatively charged residues. The important residues of the active site are highlighted as well.

4.3.7 Membrane interaction and Oligomeric assembly of Ct-acyltransferase

Ct-acyltransferase was found to interact with the membrane through only a small patch of amino acid residues, distributed over the exposed regions of the helices H1, H2 and H3 (Fig 4.15). Three prominent groups of residues made up the oligomeric interface of which two groups were found to be prominently evolutionarily conserved and one included sparsely conserved residues (Fig 4.16). This indicated that the interaction with the bacterial membrane was an inherent property of this enzyme and suggested that Ct-acyltransferase was an important component of the Gram negative bacterial outer membrane. The predicted oligomeric assembly was observed to be an isologous association since the residues which contributes to the binding in one monomer were also similarly implicated in the other monomer. The oligomeric interface was found to be quite extensive and contained 28 residues which had an equal distribution of conserved and non-conserved residues (Fig 4.16). This could probably mean that not all of the residues implicated are essential for oligomerization.

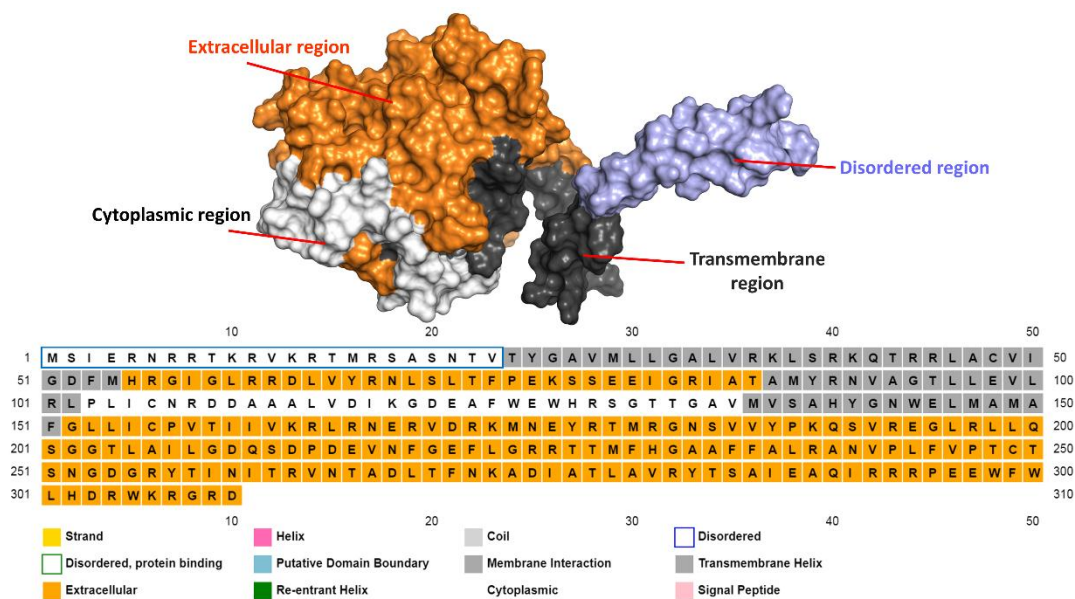


Fig 4.14 The membrane interacting structural components of Ct-acyltransferases Three transmembrane patches located in helices H1, H2 & H3 at almost the junction of the α/β core domain and the protruding helical arms of the Ct-acyltransferase monomer are shown. The extracellular regions and the transmembrane residue patches are color coded in orange and grey.

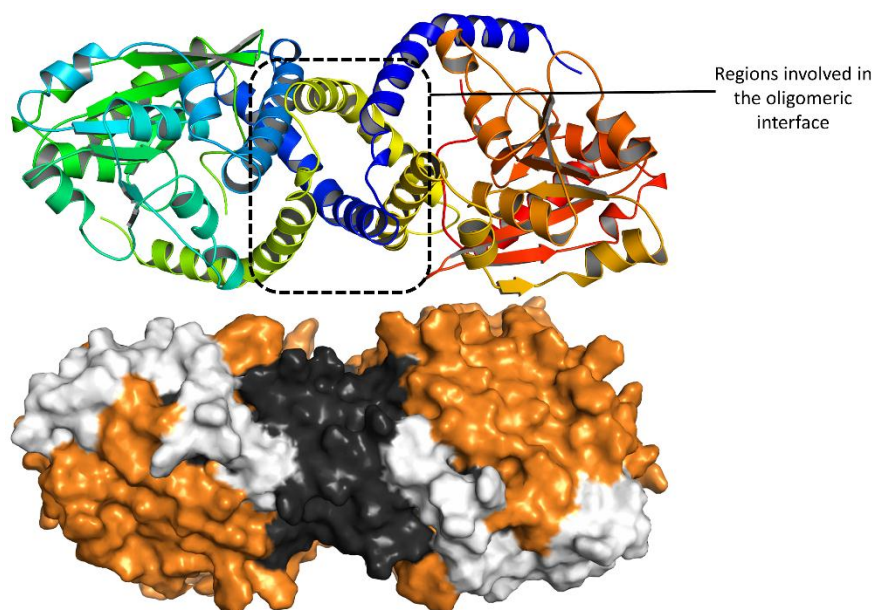


Fig 4.15 A comparative view of the spatial location of the oligomeric interface of Ct-acyltransferase The dimeric structure is shown in cartoon and surface representation and color coded as in Fig 4.14. The oligomeric interface seems to coincide partially with the membrane interacting residue patches. It is shown in greater detail in Fig 4.17.

It was also not clear how this assembly influenced the orientation of the enzyme in the membrane. The residues involved in the oligomeric assembly were evenly distributed over the predicted transmembrane regions and the extracellular region of the protein. However, they were prominently absent in the cytoplasmic part of the protein indicating that the oligomeric interface was located outside the bacterial cell. The Metsite predictions [141] coupled with the CONSURF conservation results indicated that divalent metal binding had an important influence on the oligomeric assembly and membrane interaction of Ct-acyltransferase (Fig 4.16). Conserved buried and exposed residues present in transmembrane regions of the protein (Thr95 and Ser138) and in the oligomeric interface (Ser251) were found to have high propensity towards metal binding based on Metsite prediction scores. Generally, hydrophobic residues are primarily found in the interfaces of stable oligomeric assemblies [109, 110]. However, in this case, the residues involved in the oligomeric interface were found to be mostly polar and charged with only one hydrophobic residue involved (Fig 4.17). Therefore, this association might be weak and transient in the absence of divalent metal ions and the metal ion interaction might be essential for the overall function of the Ct-acyltransferase enzyme.

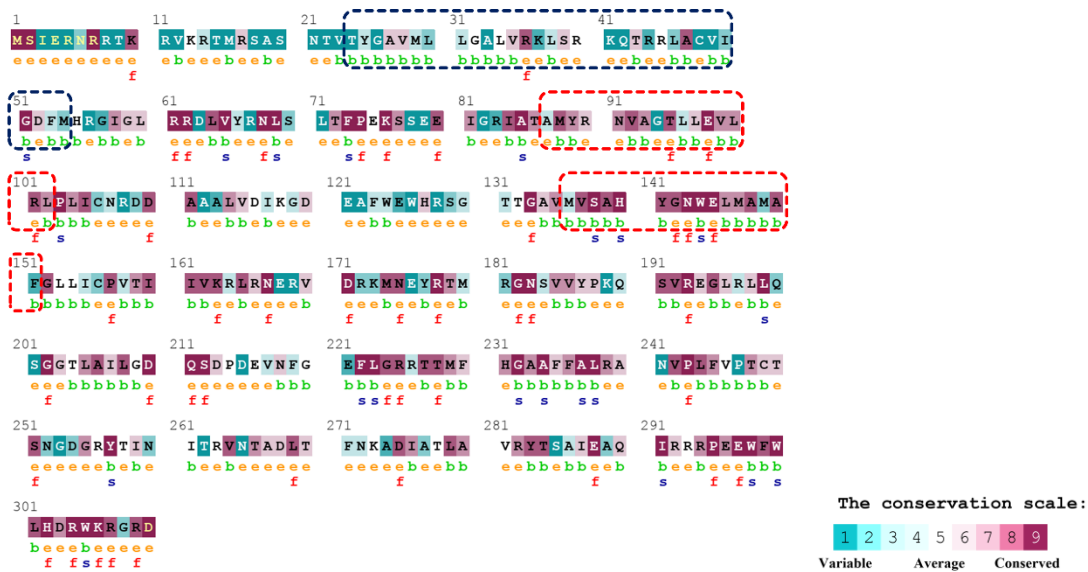


Fig 4.16 The transmembrane domain of Ct-acyltransferase The per residue evolutionary conservation of Ct-acyltransferase obtained by Consurf analysis with the dotted boxes indicating the residue patches involved in interactions of the protein with the membrane. The red dotted boxes show the highly conserved patches and contains residues which are predicted metal binding sites of the protein, while the blue dotted box shows the sparsely conserved region.

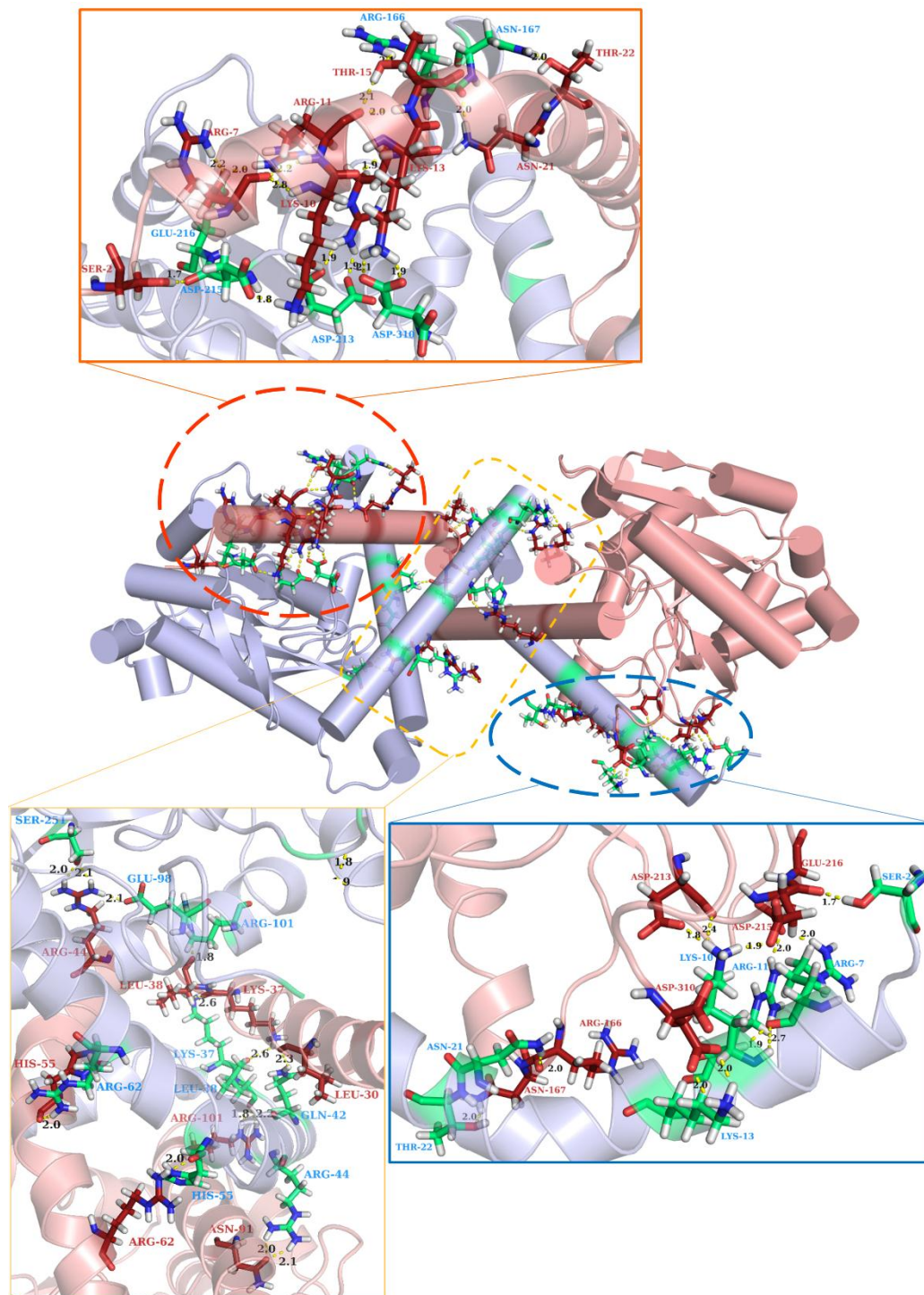


Fig 4.17 Molecular interactions in the oligomeric interface of Ct-acyltransferase The three interfaces of the oligomeric assembly are shown in greater detail. Even though most of these residues are polar unlike what is generally reported in oligomeric interfaces, S251 is one conserved residue which is involved in metal interactions as well as in the oligomeric assembly. This could point toward a possible role of metal binding in this oligomeric assembly.

4.4 Discussion

Acyltransferases are known to be major enzymes in a number of important biological processes in different organisms. Such functional diversity in one single superfamily of enzymes emphasizes the fact that, acyltransferases are essentially enzymes which warrant detailed studies, not only because of the sheer number of unique pathways they are involved in but also for its highly potent pharmacological importance. Among them the LPLAT superfamily of acyltransferases are the most prominent and functionally diverse. Structurally elucidation and analysis may give us insight into the MBOAT acyltransferases that are implicated in multiple human diseases like obesity, diabetes and Alzheimer's. Many acyltransferases are already being harvested as drug targets for tackling human disease related problems [73], but this has considerable scope in microbiological applications as well.

Our study of the Ct-acyltransferase can give a robust insight into the active site morphology of the LPLAT superfamily of enzymes since it is structurally unique from known structures of the HtrB/MsbB family of acyltransferases. The putative dual substrate specificity of this enzyme holds the key to the unique variability of the biochemical reactions catalyzed by these enzymes. The fact that the Ct-acyltransferase structure encompasses two structural features unique to known LPLAT acyltransferases (PatA and LpxM), only confirms the its case as an enzyme whose active site is diverse and thereby a perfect candidate for studying the nuances of the acyltransferase structure that makes it so functionally diverse. The two storied structure of the active site probably gives the necessary flexibility to this specific enzyme to add both laurate and myristate residues to the KDO₂-lipid IVA. Detailed structural studies might assist in the search for finding potent inhibitors of lipid A maturation in known Gram-negative pathogens. An analytical study of the same could reveal potent allosteric inhibitors to targeted enzymes relevant to human diseases as well. Overall, our study opens new frontiers in the study of the omnipresent acyltransferase superfamily and provides a broad working model for high throughput screening of inhibitors for these enzymes.

Chapter V – Summary and Inferences

Archaeal prenyltransferases (Di-geranylgeranyl glyceryl phosphate synthase; DGGGPS) and bacterial acyltransferases of the HtrB/MsbB family (Ct-acyltransferase), involved in determining the lipidic composition of the outer envelopes among prokaryotic organisms have been explored in this thesis work. These enzymes are not only important to the survival of these prokaryotes, thereby having considerable evolutionary significance, but they have close structural homologs in the eukaryotes as well. Notably, the humans orthologs have important roles to play in many diseases which emphasizes the clinical and pharmaceutical relevance of these enzymes as well. The main objective of our study was to understand the comparative structure-function relation of these enzymes with special emphasis on their core structural architecture which could possibly enhance their prospective applications. The study involved biochemical, biophysical and computational evaluation of these proteins.

DGGGPS are enzymes that catalyze the transfer of prenyl C₂₀ chains to the first archaeol precursor via formation of an ether bond which is a unique characteristic of the archaeal membrane lipids. The gene coding for archaeal DGGGPS was obtained from a sulfur-reducing, hyperthermophilic archaeal species, *Thermococcus kodakarensis*, using gene-specific primers, cloned in a bacterial system and purified by heterologous expression in a bacterial strain engineered for optimal expression of membrane proteins, C41(DE3) cells. As expected, the protein was obtained in the membrane fraction while harvesting and was further purified by Ni-NTA chromatography coupled with size exclusion chromatography. The purified protein was qualitatively estimated by western blotting and the enzyme activity was established by TLC, HPLC and mass spectrometry.

The energetics of the enzymatic reaction was assessed using ITC which highlighted the spontaneity of binding of specific prenyl chain containing ligands to the enzyme. It also hinted towards a possible allosteric activation of the enzyme by its prenyl donating ligand. The experiment indicated two binding sites for the prenyl donor, one with weak/transient binding (K_D in millimolar range) and another with strong binding (K_D in nanomolar range). Being an enzyme from a thermophilic organism, it could be expected that Tk-DGGGPS would be stable at high temperatures (verified by circular dichroism study) and show activity at elevated temperatures, however we found interesting evidence that a molten globule dynamic might have a positive influence on the catalytic process. The enzyme assembly in the membrane fraction, was verified to be a tetrameric assembly by MALLS and was found to increase the membrane compactness by its presence in an archaeal

liposome. The interaction of the protein with its surrounding components studied using hydrophobic probes (ANS) and archaeal lipid extracts revealed well-structured hydrophobic grooves, predictably involved in ligand binding interactions or membrane lipid component interactions.

Investigation of these structural features of the Tk-DGGGPS enzyme by crystallization revealed that appropriate lipid environment was important for the protein to form an ordered state required for crystal formation. Due to this reason only crystal screens optimized for membrane protein crystallization like MEMGOLD and MEMGOLD2 gave reproducible protein crystals whose quality needed further fine-tuning for better diffraction results. Also improvement in crystal size was observed as well as the time required for nucleation was found to be reduced in presence of archaeal total lipid extracts, however the diffraction quality remains to be determined. Mutants generated to curtail the entropy of the surface residues of the Tk-DGGGPS enzyme were found to influence the yield of the enzyme during heterologous expression, however only one among three mutations, the K185A mutant, showed promise for further studies. Another interesting conundrum we faced was determining the most favorable detergent to keep the enzyme stable in solution. Whereas DDM was the best option to extract the enzyme from the membrane fraction of the cell, it wasn't necessarily the ideal option for long time storage. While other detergents like CYMAL5 and CYMAL6 were estimated suitable for maintaining the protein in solution by high throughput detergent screening methods, they were unsuccessful in giving desired result, when purification of larger batches were attempted. This could probably have been due to improper detergent exchange between the extraction detergent DDM and the purifying CYMAL detergents.

A detailed amino acid sequence analysis of DGGGPS and its UbiA homologs highlighted differentially conserved residues arranged in prominent clusters. An already reported functional motif in prenyltransferases, the **NxxxDxxxxxD** motif, was found to be among these clusters, which suggested that the other cluster, **YxxxxxxL/MK/R** could be a possible conserved functional motif as well. In addition to the above two clusters, another distinct conserved amino acid cluster **RExxKxxEDxxGD** was observed in case of DGGGPS from different archaeal species. All of the residues in these clusters were also found to have high evolutionary conservation scores which further reinforced the importance of these clusters to the prenyltransferases of the UbiA superfamily. These sequence alignments were further studied by phylogenetic analysis which

highlighted the diversity of species in which these different UbiA homologs were present and the different biochemical pathways in which these enzymes were involved in prenyl chain transfers. To study the three dimensional spatial arrangement of these amino acid conservations and to fathom how integral they were to the stability and functionality of the Tk-DGGGPS enzyme, a homology model was generated. The three dimensional structure of the enzyme was determined to have an inverted V-shaped architecture dominated by membrane spanning α helices (transmembrane domains; TM) and a combination of helix-connecting loops and short helices on the cytoplasmic side with only loops exposed to the periplasm. These helix-loop (HL) regions were found to be the major contributors to the enzyme mechanism undertaken by the Tk-DGGGPS. They were also identified as mutational hotspots and found to house most of the already reported disease inducing mutations in the human UbiA homologs.

The central cavity composed of the space enclosed by the TM helices was determined to be the ligand binding site as well as the active site of the enzyme by molecular docking, where the prenyl C₂₀ chain-transfer took place. In absence of the ligands this cavity was also found to act as a water channel suggesting possible transporter functionality of the Tk-DGGGPS enzyme. This is commonly seen in integral membrane proteins where apart from their primary responsibility, they also perform other moonlighting roles like transporting ions, among others. The site for allosteric activation of the enzyme was also identified to be located in the HL regions. The GGPP binding at this site was found to facilitate binding of the ligands further inside the central cavity (than where it was found to bind without GGPP at the putative allosteric site) and showed better docking scores towards the binding of both ligands at the putative active site of the enzyme. The MD simulation of Tk-DGGGPS in membrane environments showed that the central cavity had few residues which required the presence of hydration channels for their conformational stability, supported by the fact that the central cavity had a notable increase in its volume after the simulation. Certain residue patches showed prominent fluctuation after simulation and most of them were either located on the external surface of the protein or near the putative active site in the central cavity. In the central cavity these patches were primarily located in the HL regions towards the cytoplasmic opening and near the hydrophobic channels suggesting important roles towards transporter and enzymatic activities of the protein. On the other hand, the patches located

in the external surfaces were found in the exposed TM helices and periplasmic loops which could be responsible for stability in the membrane environment and other bio-molecular interactions.

The notable points that could be inferred from the observations recorded in this study included the prominent role of DGGGPS in both the structural rigidity of the archaeal membrane as well as the biosynthesis of its lipid components. DGGGPS catalyzes the first effective step towards membrane lipid biosynthesis by shifting the lipid synthesis machinery to the membrane. The oligomeric form of the purified, heterologously expressed Tk-DGGGPS protein indicated the inherent propensity of the enzyme monomers to associate by intermolecular interactions. The N-termini BRIL tag was found to have a positive influence on the stability and expression of the protein since it forms an additional helical interface on the cytoplasmic membrane surface resulting in probable stabilizing molecular interactions. Mutants generated on basis of surface entropy based calculations also showed notable changes in yield and stability of this integral membrane protein. Experimental results suggested that proper oligomerization is dependent on the detergent concentration, glycerol concentration in purifying buffer and the variety of detergent used. The most dominant oligomeric form obtained in solution was that of a homo tetramer. Tk-DGGGPS was found to function most efficiently at slightly acidic environments, have high specificity in their substrate preferences and was notably thermostable making it a prime candidate for protein engineering to synthesize lipids of choice in industry. The preferred temperature range of catalysis was 50°C -60°C, where the enzyme showed optimal product yield. CD spectrum revealed a high α -helical composition (70%) with 8% loops and 22% random coils. From CD and fluorescence analysis of Tk-DGGGPS, it could be said that the enzyme catalyzes its substrates most efficiently in a molten globule ensemble. The binding of the prenyl chain donor, GGPP to the enzyme along with changes in temperature were observed to cause notable conformational changes in the active site. The enzyme was found to have a high affinity of interaction (low K_{D1} ; 866nM) with GGPP at one site and another low affinity interaction (high K_{D2} ; 273 mM) as well. We inferred in congregation with fluorescence analysis and molecular docking results that the later could be a site for allosteric regulation of the enzyme by its dedicated ligand. The overall three dimensional morphology including the hydrophobic grooves were also influenced by increase in temperatures and at temperatures above 70°C there was a prominent decrease in the surface hydrophobicity of the protein wherein its interaction with its hydrophobic surroundings were hampered. Crystallization

studies established the paramount importance of the surrounding hydrophobic environment for the conformational stability of the Tk-DGGGPS. The lipid content, be it detergents or archaeal total lipids, was found to increase the effectiveness of crystal nucleation. The size of the co-purifying detergent micelle was found to be an important factor in crystallization as well. DDM was found to be ideal for the extraction of this intrinsic membrane protein from the cell membrane fraction but it could keep the protein from precipitating for only a limited period of time. The way forward could be an efficient detergent transfer to low micelle sized detergents which would prevent excessive detergent accumulation when the protein is concentrated for carrying out crystallization studies. The overall sequence conservation in DGGGPS was quite prominent with three notable clusters. These clusters had high evolutionarily conserved amino acid residues and were found to border the cytoplasmic opening of the central cavity and were primarily present in the HL regions. These residues were found to be implicated in ligand interactions and inter-residue interactions between homomers, therefore highlighting their importance in the overall biology of the enzyme. In the phylogenetic tree DGGGPS was found to group separately from the naming members of the UbiA superfamily, which is the bacterial UbiA and notably shared the same clade with mammalian homologs (UBIAD1) and plant UbiA homologs, which led us to believe that there could be key differences between the already reported structures (Ap-UbiA) and DGGGPS. Homology modelling of Tk-DGGGPS provided further structural insight into the overall morphology of the enzyme. In an α helix dominated three dimensional structure, the TM helices constituting the active site were found to be more prominently conserved than the TM helices on the external surface of the protein. The presence of multiple conserved proline residues around the active site indicated at the robustness of the architecture of the active site. The HL regions were primarily responsible for ligand interactions and binding of GGPP near the HL67 resulted in better docking scores of the prenyl acceptor GGPP and hence it can be concluded that it facilitates the prenyl chain transfer more efficiently. This site was deduced to be the allosteric binding site of GGPP to the enzyme. The HL regions were found to have roles in oligomeric assembly as well with conserved residues involved in the isologous dimeric association of the monomers located in HL23. The dimeric assembly seems to be the preliminary step towards the final tetramer formation. A water channel was found throughout the entire length of the Tk-DGGGPS central cavity after subjecting it to a 50 nanosecond simulation in presence of theoretically constructed ether lipid bilayer. This could

be prominent proof towards a water-dependent transporter role taken up by Tk-DGGGPS in absence of lipid substrates and/or could also be a mechanism by which some residues in and around the active site could maintain stabilizing interactions between their dynamic buried polar side chains and water molecules. Other notable observations were changes in some secondary structure elements of the protein in two notable regions of the three dimensional structure. One area of impact were the HL regions towards the cytoplasmic end of the enzyme near the active site, where there were changes in the loop and helix content indicating the dynamic nature of the region. This dynamic nature could be a key to efficient catalysis by the enzyme. The other area of impact was the surface exposed TM8 which formed a kinked helix after simulation where the kink region was found to interact with the lipid molecules of the theoretical bilayer. This, coupled with the other residues with high RMSD located towards both the cytoplasmic and periplasmic ends of this membrane intrinsic enzyme could be a key factor in the membrane assembly of the enzyme. The simulated structure showed a prominent change in the size and structure of the central cavity with formation of new hydrophobic channels. This could be a key finding towards determining the orientation of the hydrophobic substrates of the enzyme and could also be the key towards the co-existence of water molecules in the active site and the prenyl chain containing substrates. The HL regions were also found to be mutational hotspots with already known disease-related mutations in the human homologs. Few new residues were identified which could be useful in identification of other disease-related mutations.

The Ct-acyltransferase of the HtrB/MsbB family are involved in the secondary O-acylation of KDO bound Lipid IVA precursors for generation of matured Lipid A molecules. The transfer of acyl chains to the attached R-3-hydroxy fattyacyl chains at C2 and C3 positions of the KDO₂-lipid IVA precursor are catalyzed by this enzyme. Due to the presence of only one acyltransferase gene of the HtrB/MsbB family in *C. tepidium*, the CT0211 gene, this acyltransferase enzyme could be responsible for carrying out both the secondary acylations of the KDO₂-lipid IVA. The gene coding for this bacterial acyltransferase was obtained from the anaerobic, thermophilic, gram negative, green sulfur bacterial species, *Chlorobium tepidium*, using gene-specific primers, cloned in a bacterial system and purified by heterologous expression in a bacterial strain engineered for optimal heterologous expression of recombinant proteins, BL21 Star(DE3) cells. The protein was expressed and localized in the membrane fraction of the expression cells. A complete lysis of the

cells by homogenization was required for efficient extraction of the protein while harvesting where DDM was used as the detergent for overnight extraction. Further purification was done in a lower carbon chain length containing maltoside, DM and further purification was carried out by Ni-NTA chromatography coupled with size exclusion chromatography. The native protein and the BRIL tagged variant had prominent differences in the protein yield as well stability in solution. It appeared that the BRIL tag reduced the overall dynamic nature of the protein which was reflected in the increased yield and stability. However, we failed to obtain stable PDC at high concentrations required for crystallization which led us to check the general stability of the PDC which seemed fine. This could therefore be the case of a concentration dependent precipitation which could be due to the influence of different variables. Due to the roadblocks in obtaining an experimental three dimensional structure of the protein we proceeded with homology modelling to study the salient features of this acyltransferase. A HHPRED analysis revealed that the Ct-acyltransferase had very high structural homologies with other LPLAT acyltransferases which had completely different substrate preferences and were involved in totally unrelated biochemical pathways in a vast variety of organisms ranging from bacteria, plants and mammals. A DALI analysis further emphasized the presence of four prominent subdomains which comprised of the entire central α/β domain. This α/β domain seems to be the core conserved region in all the acyltransferases of the LPLAT superfamily which had quite insignificant amino acid sequence conservations. Four notable clusters of conserved amino acid residues were observed in the multiple sequence analysis of the HtrB/MsbB family acyltransferases. Apart from the already reported **HxxxxD/E** motif, which is omnipresent in all acyltransferases, the clusters **NLxLCFPExSxxER**, **Y/IxPDQD/SxGPxxSxF** and **ExxIxxxPEQYxW** were identified, which had prominent evolutionarily conserved residues as well. Further studies on the sequence alignments by phylogenetic analysis highlighted the diversity of species in which these acyltransferases are present and revealed the separate clustering of Ct-acyltransferases from the HtrB/MsbB acyltransferases. The enzyme was found to cluster with two functional domain containing proteins, Aas, indicating that it too could have a dual function which found support from the findings from its predicted three dimensional structure. The homology model generated housed a prominent α/β domain composed of parallel and antiparallel β strands having fairly conserved residues. The α helices contained key polar and charged residues and was found to constitute the exposed surface

area of the enzyme. The catalytic dyad residues of the HxxxxD/E motif (His140 and Glu145) were located in these α helices and the connecting loops at the center of the α/β domain which is also the predicted catalytic site of the enzyme. Two conserved positively charged residues (Lys163 and Arg178) were strategically placed in the binding pocket near the α helices where Arg178 was located next to the catalytic His140 and Lys163 formed a bridge across the binding cleft. Conserved residues were found at the substrate binding sites and conserved putative metal binding sites were found near the transmembrane regions of the enzyme. The active site was found to possess a dual storied architecture and consisted of a deep hydrophobic pocket distinctly divided into two halves by a belt of hydrophilic residues. The Ct-acyltransferase was predicted to be partially embedded into the membrane by its N terminal helices and the pore lining helix with the N terminal towards the cytoplasmic side of the membrane and C terminal towards the extracellular side. The enzyme was predicted to exist as a dimer in the membrane with the helices H1, H2 and H3 being the ones involved primarily. The residues which were predicted to be involved in the dimeric interface were found to be evolutionarily conserved as well out of which Ser251 was a residue which was a putative divalent metal binding site. Other predicted metal binding sites like the positively charged Thr95 and Ser138 residues were found in the transmembrane regions. The oligomeric interface was found to be quite extensive and contained 28 residues which had an equal number of conserved and non-conserved residues. The residues involved in the oligomeric assembly were absent in the cytoplasmic part of the protein, indicating that the oligomeric interface was located outside the bacterial cell, however it wasn't clear how it influenced the membrane orientation of the enzyme.

From this study the following inferences about the Ct-acyltransferase and its superfamily could be drawn. Ct-acyltransferase was the only member of the HtrB/MsbB family of acyltransferases in *Chlorobium tepidium* which carry out the secondary acylations of the immature Lipid A moieties. It was probably due to the absence of the regular LpxL and LpxM homologs in the organism, that this photoautotrophic Gram negative bacterium lacked virulence. However, the Ct-acyltransferase appeared to have other functions than the one designated function of its enzyme family as well, emphasized by its clustering with acyltransferases with two functional domains. It was predicted to have an acyl-protein thioesterase functionality where it recycles the acyl ACP to holo ACP in absence of an acyl chain acceptor. This was further supported by the finding that there

were two putative sites for the binding of acyl ACP, each prominent in their location, of which one could be the regulatory binding site for the substrate controlled feedback mechanism. The fact that it shared a unique structural feature of mycobacterial PatA acyltransferase and its unique point of divergence in the phylogenetic tree of the LPLAT superfamily make a strong case for it being one of the ancestral genes of the superfamily. However, the high variability in the sequences indicate that the members of this superfamily might have undergone divergent evolution. In spite of the high sequence variability a few notable conserved clusters which were identified seemed to have important roles in the active site and the membrane interface of the enzyme. The dimeric assembly and its location in the membrane couldn't be elucidated thoroughly but it appeared that it has prominent dependence on divalent metal interactions for stability. The dimer interface containing an unlikely number of polar residues and its location outside the cell could mean that the oligomeric assembly could be weak and transient, only stable during the availability of substrates and catalysis. The dual-storied morphology of the active site probably gives the necessary flexibility to the enzyme for addition of both laurate and myristate residues to the KDO₂-lipid IVA. The Lys163 residue was strategically placed quite similar to the mycobacterial PatA and could be involved in the opening and closing of the active site cleft during catalysis. The conserved α/β domain located at the core of the Ct-acyltransferase globular structure appeared to be most evolutionarily conserved region in the protein. Analytical study of such conserved pockets could help in finding potent allosteric inhibitors to targeted acyltransferases relevant to human diseases and could even help in design of the pathogenic Lipid A inhibitors.

Overall this thesis work highlights the unique structure-function dynamic of membrane intrinsic transferase enzymes, which shape the lipid environment of the prokaryotic cell exterior using cloning, expression and enzyme characterization by various biochemical, biophysical and *in silico* techniques. This study, with two enzymes in focus, Tk-DGGGPS from archaea and Ct-acyltransferase from bacteria, attempts to gain a holistic understanding of how diversely the membrane lipid biochemistry can be influenced by such enzymes in different organisms. These enzymes apart from having important regulatory roles in their respective organisms, have probable commercial/industrial applications as well. Both the enzymes could be harvested as possible drug targets for the diseases in which their human homologs are implicated, due to the conserved structural features that these respective enzymes possess, in spite of sequence variations. DGGGPS

could be an important cog in wheel of archaeal liposome manufacture both *in vivo* and *in vitro* by protein bioengineering and gene transfer into relevant bacterial strains. Liposomes hold promise as systems of drug delivery due to their size, amphipathic nature and biocompatibility. Archaeal lipids are naturally resistant to oxidation and action of esterases and hence can be used to manufacture liposomes those are stable over a broad range of temperatures and pH. These archaeal liposomes or archaeosomes are thereby superior to the naturally isolated ones, due to their chemical and thermal inertness. It is this ultra-high stability factor, which increases the potential usefulness of archaeosomes in biotechnology, the most prominent among which are vessels for drug delivery and the construction of novel membranes for separation processes. These liposome of archaeal lipid origin are even resistant to autoclaving, thereby can be microbial and viral contaminant free as well, which would have been impossible in the case of normal, polyester-derived liposomes because they would lose structural integrity. DGGGPS is the enzyme central to the biosynthesis of lipid components of these archaeosomes and if it can be obtained in a purified and active state and maintained in a stable form, there would be no requirement to obtain these polar lipids from archaeal extracts, thereby making the archaeosome production process much more economically feasible for the biomedical industry. The secondary Lipid A acyltransferases could help us in developing potent inhibitors against Lipid A maturation in other Gram-negative pathogens. It could also help in development of inhibitors towards the secondary acyl ACP binding, thereby preventing the thioesterase activity which maintains the metabolic balance in the bacteria. For this purpose, more detailed understanding of the structural make up of these acyltransferases are required and thereby atomic structure elucidation is the next roadblock which needs to be investigated. For both proteins, their contribution to drug development will be more exhaustive with their high resolution three dimensional structures and hence further research in this direction is imperative.

BIBLIOGRAPHY

1. Stanier, R.Y. and C. Van Niel, *The concept of a bacterium*. Archiv für Mikrobiologie, 1962. **42**(1): p. 17-35.
2. Woese, C.R. and G.E. Fox, *Phylogenetic structure of the prokaryotic domain: the primary kingdoms*. Proceedings of the National Academy of Sciences, 1977. **74**(11): p. 5088-5090.
3. Langworthy, T.A., *Lipids of archaeobacteria*. The bacteria, 1985. **8**: p. 459-497.
4. Kates, M. *Archaeobacterial lipids: structure, biosynthesis and function*. in *Biochemical Society Symposium*. 1992.
5. Kates, M., *Membrane lipids of extreme halophiles: biosynthesis, function and evolutionary significance: Biology of halophilic bacteria. II*. Experientia, 1993. **49**(12): p. 1027-1036.
6. Kamekura, M. and M. Kates, *Structural diversity of membrane lipids in members of Halobacteriaceae*. Bioscience, biotechnology, and biochemistry, 1999. **63**(6): p. 969-972.
7. Koga, Y., et al., *Ether polar lipids of methanogenic bacteria: structures, comparative aspects, and biosyntheses*. Microbiology and Molecular Biology Reviews, 1993. **57**(1): p. 164-182.
8. Sheu, C.W. and E. Freese, *Lipopolysaccharide layer protection of gram-negative bacteria against inhibition by long-chain fatty acids*. Journal of Bacteriology, 1973. **115**(3): p. 869-875.
9. Creighton, T.E., *Proteins: structures and molecular properties*. 1993: Macmillan.
10. Branden, C. and J. Tooze, *Prediction, engineering, and design of protein structures*. Introduction to Protein Structure, Garland Publishing Inc., New York, 1991: p. 247.
11. Anfinsen, C.B., *Principles that govern the folding of protein chains*. Science, 1973. **181**(4096): p. 223-230.
12. Kendrew, J.C., *The three-dimensional structure of a protein molecule*. Scientific American, 1961. **205**(6): p. 96-111.
13. Berman, H.M., et al., *The Protein Data Bank, 1999–*, in *International Tables for Crystallography Volume F: Crystallography of biological macromolecules*. 2006, Springer. p. 675-684.
14. Berman, H.M., *The past and future of structure databases*. Current opinion in biotechnology, 1999. **10**(1): p. 76-80.
15. McConkey, E.H., *Molecular evolution, intracellular organization, and the quinary structure of proteins*. Proceedings of the National Academy of Sciences, 1982. **79**(10): p. 3236-3240.
16. Zuckerkandl, E., *The appearance of new structures and functions in proteins during evolution*. Journal of molecular evolution, 1975. **7**(1): p. 1-57.
17. Astbury, W., *Relation between 'Fibrous' and 'Globular' Proteins*. Nature, 1937. **140**(3553): p. 968-969.
18. von Heijne, G., *Membrane-protein topology*. Nature reviews Molecular cell biology, 2006. **7**(12): p. 909-918.
19. Kraut, J., *How do enzymes work?* Science, 1988. **242**(4878): p. 533-540.
20. Fersht, A., *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*. 1999: Macmillan.

21. Gutteridge, A. and J. Thornton, *Conformational changes observed in enzyme crystal structures upon substrate binding*. Journal of molecular biology, 2005. **346**(1): p. 21-28.
22. Holliday, G.L., J.B. Mitchell, and J.M. Thornton, *Understanding the functional roles of amino acid residues in enzyme catalysis*. Journal of molecular biology, 2009. **390**(3): p. 560-577.
23. Bennett, T.P. and E. Frieden, *Modern topics in biochemistry: structure and function of biological molecules*. 1969: Macmillan.
24. Blake, C., et al., *Structure of hen egg-white lysozyme: a three-dimensional Fourier synthesis at 2 Å resolution*. Nature, 1965. **206**(4986): p. 757-761.
25. McDonald, A.G. and K.F. Tipton, *Fifty-five years of enzyme classification: advances and difficulties*. The FEBS journal, 2014. **281**(2): p. 583-592.
26. McDonald, A.G., S. Boyce, and K.F. Tipton, *Enzyme classification and nomenclature*. eLS, 2001: p. 1-11.
27. Alfaro, J.A., et al., *ABO (H) blood group A and B glycosyltransferases recognize substrate via specific conformational changes*. Journal of Biological Chemistry, 2008. **283**(15): p. 10097-10108.
28. Liang, P.H., T.P. Ko, and A.H.J. Wang, *Structure, mechanism and function of prenyltransferases*. European Journal of Biochemistry, 2002. **269**(14): p. 3339-3354.
29. Poulter, C.D. and H.C. Rilling, *Prenyltransferase: the mechanism of the reaction*. Biochemistry, 1976. **15**(5): p. 1079-1083.
30. Li, W., *Bringing Bioactive Compounds into Membranes: The UbiA Superfamily of Intramembrane Aromatic Prenyltransferases*. Trends Biochem Sci, 2016. **41**(4): p. 356-70.
31. Todd, A.E., C.A. Orengo, and J.M. Thornton, *Evolution of function in protein superfamilies, from a structural perspective*. Journal of molecular biology, 2001. **307**(4): p. 1113-1143.
32. Huang, H., et al., *Structure of a membrane-embedded prenyltransferase homologous to UBIADI*. PLoS biology, 2014. **12**(7): p. e1001911.
33. Cheng, W. and W. Li, *Structural insights into ubiquinone biosynthesis in membranes*. Science, 2014. **343**(6173): p. 878-81.
34. Boucher, Y., M. Kamekura, and W.F. Doolittle, *Origins and evolution of isoprenoid lipid biosynthesis in archaea*. Mol Microbiol, 2004. **52**(2): p. 515-27.
35. Lombard, J., P. López-García, and D. Moreira, *The early evolution of lipid membranes and the three domains of life*. Nature Reviews Microbiology, 2012. **10**(7): p. 507.
36. Lombard, J. and D. Moreira, *Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life*. Molecular biology and evolution, 2011. **28**(1): p. 87-99.
37. Barkovich, R. and J.C. Liao, *Metabolic engineering of isoprenoids*. Metab Eng, 2001. **3**(1): p. 27-39.
38. Khosla, C. and J.D. Keasling, *Metabolic engineering for drug discovery and development*. Nature Reviews Drug Discovery, 2003. **2**(12): p. 1019.
39. Koga, Y. and H. Morii, *Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations*. Microbiol. Mol. Biol. Rev., 2007. **71**(1): p. 97-120.
40. Hemmi, H., et al., *(S)-2, 3-Di-O-geranylgeranylgeranyl glyceryl phosphate synthase from the thermoacidophilic archaeon Sulfolobus solfataricus: molecular cloning and*

- characterization of a membrane-intrinsic prenyltransferase involved in the biosynthesis of archaeal ether-linked membrane lipids.* Journal of Biological Chemistry, 2004. **279**(48): p. 50197-50203.
41. Villanueva, L., J.S.S. Damsté, and S. Schouten, *A re-evaluation of the archaeal membrane lipid biosynthetic pathway.* Nature Reviews Microbiology, 2014. **12**(6): p. 438.
 42. Balleza, D., et al., *Ether-versus ester-linked phospholipid bilayers containing either linear or branched apolar chains.* Biophysical journal, 2014. **107**(6): p. 1364-1374.
 43. Ohara, K., et al., *Functional characterization of LePGT1, a membrane-bound prenyltransferase involved in the geranylation of p-hydroxybenzoic acid.* Biochemical Journal, 2009. **421**(2): p. 231-241.
 44. Ohara, K., K. Mito, and K. Yazaki, *Homogeneous purification and characterization of LePGT1—a membrane-bound aromatic substrate prenyltransferase involved in secondary metabolism of Lithospermum erythrorhizon.* The FEBS journal, 2013. **280**(11): p. 2572-2580.
 45. Huang, H., et al., *Identification of amino acids and domains required for catalytic activity of DPPR synthase, a cell wall biosynthetic enzyme of Mycobacterium tuberculosis.* Microbiology, 2008. **154**(3): p. 736-743.
 46. Stec, E. and S.-M. Li, *Mutagenesis and biochemical studies on AuaA confirmed the importance of the two conserved aspartate-rich motifs and suggested difference in the amino acids for substrate binding in membrane-bound prenyltransferases.* Archives of microbiology, 2012. **194**(7): p. 589-595.
 47. Sasaki, K., et al., *Molecular characterization of a membrane-bound prenyltransferase specific for isoflavone from Sophora flavescens.* Journal of Biological Chemistry, 2011. **286**(27): p. 24125-24134.
 48. Röttig, A. and A. Steinbüchel, *Acyltransferases in bacteria.* Microbiol. Mol. Biol. Rev., 2013. **77**(2): p. 277-321.
 49. Heath, R.J. and C.O. Rock, *A conserved histidine is essential for glycerolipid acyltransferase catalysis.* Journal of bacteriology, 1998. **180**(6): p. 1425-1430.
 50. Shindou, H., T. Harayama, and D. Hishikawa, *Lysophospholipid Acyltransferases*, in *Bioactive Lipid Mediators*. 2015, Springer. p. 3-21.
 51. Chang, C.C., J. Sun, and T.-Y. Chang, *Membrane-bound O-acyltransferases (MBOATs).* Frontiers in Biology, 2011. **6**(3): p. 177.
 52. Turnbull, A.P., et al., *Analysis of the structure, substrate specificity, and mechanism of squash glycerol-3-phosphate (1)-acyltransferase.* Structure, 2001. **9**(5): p. 347-353.
 53. Albesa-Jové, D., et al., *Structural basis for selective recognition of acyl chains by the membrane-associated acyltransferase PatA.* Nature communications, 2016. **7**(1): p. 1-12.
 54. Robertson, R.M., et al., *A two-helix motif positions the lysophosphatidic acid acyltransferase active site for catalysis within the membrane bilayer.* Nature structural & molecular biology, 2017. **24**(8): p. 666.
 55. Dovala, D., et al., *Structure-guided enzymology of the lipid A acyltransferase LpxM reveals a dual activity mechanism.* Proceedings of the National Academy of Sciences, 2016. **113**(41): p. E6064-E6071.
 56. Clementz, T., Z. Zhou, and C.R. Raetz, *Function of the Escherichia coli msbB gene, a multicopy suppressor of htrB knockouts, in the acylation of lipid A acylation by MsbB*

- follows laurate incorporation by HtrB*. Journal of Biological Chemistry, 1997. **272**(16): p. 10353-10360.
57. Brozek, K.A. and C. Raetz, *Biosynthesis of lipid A in Escherichia coli. Acyl carrier protein-dependent incorporation of laurate and myristate*. Journal of Biological Chemistry, 1990. **265**(26): p. 15410-15417.
 58. Clementz, T., J.J. Bednarski, and C.R. Raetz, *Function of the htrB high temperature requirement gene of Escherichia coli in the acylation of lipid a HtrB catalyzed incorporation of laurate*. Journal of Biological Chemistry, 1996. **271**(20): p. 12095-12102.
 59. Raetz, C.R., et al., *Lipid A modification systems in gram-negative bacteria*. Annu. Rev. Biochem., 2007. **76**: p. 295-329.
 60. Whitfield, C. and M.S. Trent, *Biosynthesis and export of bacterial lipopolysaccharides*. Annu Rev Biochem, 2014. **83**: p. 99-128.
 61. Pelletier, M.R., et al., *Unique structural modifications are present in the lipopolysaccharide from colistin-resistant strains of Acinetobacter baumannii*. Antimicrobial agents and chemotherapy, 2013. **57**(10): p. 4831-4840.
 62. Ogaki, K., et al., *Analysis of COQ2 gene in multiple system atrophy*. Molecular neurodegeneration, 2014. **9**(1): p. 44.
 63. Hegarty, J.M., H. Yang, and N.C. Chi, *UBIAD1-mediated vitamin K2 synthesis is required for vascular endothelial cell survival and development*. Development, 2013. **140**(8): p. 1713-1719.
 64. Mugoni, V., et al., *Ubiad1 is an antioxidant enzyme that regulates eNOS activity by CoQ10 synthesis*. Cell, 2013. **152**(3): p. 504-518.
 65. Vos, M., et al., *Vitamin K2 is a mitochondrial electron carrier that rescues pink1 deficiency*. Science, 2012. **336**(6086): p. 1306-1310.
 66. Orr, A., et al., *Mutations in the UBIAD1 gene, encoding a potential prenyltransferase, are causal for Schnyder crystalline corneal dystrophy*. PLoS One, 2007. **2**(8).
 67. Weiss, J.S., et al., *Mutations in the UBIAD1 gene on chromosome short arm 1, region 36, cause Schnyder crystalline corneal dystrophy*. Investigative ophthalmology & visual science, 2007. **48**(11): p. 5007-5012.
 68. Fredericks, W.J., et al., *The bladder tumor suppressor protein TERE1 (UBIAD1) modulates cell cholesterol: implications for tumor progression*. DNA and cell biology, 2011. **30**(11): p. 851-864.
 69. Kohda, M., et al., *A comprehensive genomic analysis reveals the genetic landscape of mitochondrial respiratory chain complex deficiencies*. PLoS genetics, 2016. **12**(1).
 70. Valnot, I., et al., *A mutation in the human heme A: farnesyltransferase gene (COX10) causes cytochrome c oxidase deficiency*. Human Molecular Genetics, 2000. **9**(8): p. 1245-1249.
 71. Antonicka, H., et al., *Mutations in COX10 result in a defect in mitochondrial heme A biosynthesis and account for multiple, early-onset clinical phenotypes associated with isolated COX deficiency*. Hum Mol Genet, 2003. **12**(20): p. 2693-702.
 72. Lee, Y.J. and J.-w. Kim, *Monoacylglycerol O-acyltransferase 1 (MGAT1) localizes to the ER and lipid droplets promoting triacylglycerol synthesis*. BMB reports, 2017. **50**(7): p. 367.

73. Devasthale, P. and D. Cheng, *Monoacylglycerol Acyltransferase 2 (MGAT2) Inhibitors for the Treatment of Metabolic Diseases and Nonalcoholic Steatohepatitis (NASH) Miniperspective*. Journal of medicinal chemistry, 2018. **61**(22): p. 9879-9888.
74. Chang, T.Y., et al., *Neuronal cholesterol esterification by ACAT1 in Alzheimer's disease*. IUBMB life, 2010. **62**(4): p. 261-267.
75. Chang, T.-Y., et al., *Acyl-coenzyme A: cholesterol acyltransferases*. American Journal of Physiology-Endocrinology and Metabolism, 2009. **297**(1): p. E1-E9.
76. Birch, A.M., L.K. Buckett, and A.V. Turnbull, *DGAT1 inhibitors as anti-obesity and anti-diabetic agents*. Current opinion in drug discovery & development, 2010. **13**(4): p. 489-496.
77. Sukumaran, S., et al., *Functional characterization of the human 1-acylglycerol-3-phosphate-O-acyltransferase isoform 10/glycerol-3-phosphate acyltransferase isoform 3*. Journal of molecular endocrinology, 2009. **42**(6): p. 469-478.
78. Claypool, S.M., et al., *The cardiolipin transacylase, tafazzin, associates with two distinct respiratory components providing insight into Barth syndrome*. Molecular biology of the cell, 2008. **19**(12): p. 5143-5155.
79. Matsumi, R., et al., *Isoprenoid biosynthesis in Archaea—biochemical and evolutionary implications*. Research in microbiology, 2011. **162**(1): p. 39-52.
80. Wittig, I., H.-P. Braun, and H. Schägger, *Blue native PAGE*. Nature protocols, 2006. **1**(1): p. 418.
81. Sreerama, N. and R.W. Woody, *Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set*. Analytical biochemistry, 2000. **287**(2): p. 252-260.
82. Chhonker, Y.S., et al., *Simultaneous Quantitation of Isoprenoid Pyrophosphates in Plasma and Cancer Cells Using LC-MS/MS*. Molecules, 2018. **23**(12): p. 3275.
83. Sun, B., et al., *Crystal structure of the adenosine A2A receptor bound to an antagonist reveals a potential allosteric pocket*. Proceedings of the National Academy of Sciences, 2017. **114**(8): p. 2066-2071.
84. Wolfe, A.J., et al., *Detergent desorption of membrane proteins exhibits two kinetic phases*. The journal of physical chemistry letters, 2018. **9**(8): p. 1913-1919.
85. Cheng, A., et al., *A simple mechanical mixer for small viscous lipid-containing samples*. Chemistry and Physics of Lipids, 1998. **95**(1): p. 11-21.
86. Caffrey, M. and V. Cherezov, *Crystallizing membrane proteins using lipidic mesophases*. Nature protocols, 2009. **4**(5): p. 706.
87. Kumar, S., et al., *MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms*. Mol Biol Evol, 2018. **35**(6): p. 1547-1549.
88. Mount, D.W., *Using the Basic Local Alignment Search Tool (BLAST)*. CSH Protoc, 2007. **2007**: p. pdb top17.
89. Söding, J., A. Biegert, and A.N. Lupas, *The HHpred interactive server for protein homology detection and structure prediction*. Nucleic acids research, 2005. **33**(suppl_2): p. W244-W248.
90. Webb, B. and A. Sali, *Protein structure modeling with MODELLER*, in *Protein Structure Prediction*. 2014, Springer. p. 1-15.

91. Sastry, G.M., et al., *Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments*. Journal of computer-aided molecular design, 2013. **27**(3): p. 221-234.
92. Jacobson, M.P., et al., *A hierarchical approach to all-atom protein loop prediction*. Proteins: Structure, Function, and Bioinformatics, 2004. **55**(2): p. 351-367.
93. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of protein structures*. Journal of applied crystallography, 1993. **26**(2): p. 283-291.
94. Colovos, C. and T.O. Yeates, *Verification of protein structures: patterns of nonbonded atomic interactions*. Protein science, 1993. **2**(9): p. 1511-1519.
95. Eisenberg, D., R. Lüthy, and J.U. Bowie, [20] *VERIFY3D: assessment of protein models with three-dimensional profiles*, in *Methods in enzymology*. 1997, Elsevier. p. 396-404.
96. Friesner, R.A., et al., *Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy*. Journal of medicinal chemistry, 2004. **47**(7): p. 1739-1749.
97. Halgren, T.A., *Identifying and characterizing binding sites and assessing druggability*. Journal of chemical information and modeling, 2009. **49**(2): p. 377-389.
98. Baek, M., et al., *GalaxyHomomer: a web server for protein homo-oligomer structure prediction from a monomer sequence or structure*. Nucleic acids research, 2017. **45**(W1): p. W320-W324.
99. Lee, H., et al., *GalaxyGemini: a web server for protein homo-oligomer structure prediction based on similarity*. Bioinformatics, 2013. **29**(8): p. 1078-1080.
100. Jo, S., et al., *CHARMM-GUI: a web-based graphical user interface for CHARMM*. Journal of computational chemistry, 2008. **29**(11): p. 1859-1865.
101. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free*. Journal of computational chemistry, 2005. **26**(16): p. 1701-1718.
102. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics*. Journal of molecular graphics, 1996. **14**(1): p. 33-38.
103. Stenson, P.D., et al., *The human gene mutation database: 2008 update*. Genome medicine, 2009. **1**(1): p. 13.
104. Quan, L., Q. Lv, and Y. Zhang, *STRUM: structure-based prediction of protein stability changes upon single-point mutation*. Bioinformatics, 2016. **32**(19): p. 2936-2946.
105. Goldschmidt, L., et al., *Toward rational protein crystallization: A Web server for the design of crystallizable protein variants*. Protein Science, 2007. **16**(8): p. 1569-1576.
106. Finn, R.D., et al., *The Pfam protein families database: towards a more sustainable future*. Nucleic Acids Res, 2016. **44**(D1): p. D279-85.
107. Celniker, G., et al., *ConSurf: using evolutionary data to raise testable hypotheses about protein function*. Israel Journal of Chemistry, 2013. **53**(3-4): p. 199-206.
108. Watanabe, A., et al., *The mechanism of sodium and substrate release from the binding pocket of vSGLT*. Nature, 2010. **468**(7326): p. 988-991.
109. Jones, S., A. Marin, and J. M. Thornton, *Protein domain interfaces: characterization and comparison with oligomeric protein interfaces*. Protein Engineering, 2000. **13**(2): p. 77-82.
110. Duarte, J.M., et al., *An analysis of oligomerization interfaces in transmembrane proteins*. BMC structural biology, 2013. **13**(1): p. 21.
111. Tina, K., R. Bhadra, and N. Srinivasan, *PIC: protein interactions calculator*. Nucleic acids research, 2007. **35**(suppl_2): p. W473-W476.

112. Vijayakumar, M., H. Qian, and H.X. Zhou, *Hydrogen bonds between short polar side chains and peptide backbone: Prevalence in proteins and effects on helix-forming propensities*. *Proteins: Structure, Function, and Bioinformatics*, 1999. **34**(4): p. 497-507.
113. Schmidt, T., A.J. Situ, and T.S. Ulmer, *Structural and thermodynamic basis of proline-induced transmembrane complex stabilization*. *Scientific reports*, 2016. **6**(1): p. 1-7.
114. Jeffery, C.J., *Moonlighting proteins: old proteins learning new tricks*. *TRENDS in Genetics*, 2003. **19**(8): p. 415-417.
115. Fredericks, W.J., et al., *The bladder tumor suppressor protein TERE1 (UBIAD1) modulates cell cholesterol: implications for tumor progression*. *DNA Cell Biol*, 2011. **30**(11): p. 851-64.
116. Fredericks, W.J., et al., *The TERE1 protein interacts with mitochondrial TBL2: Regulation of trans-membrane potential, ROS/RNS and SXR target genes*. *Journal of cellular biochemistry*, 2013. **114**(9): p. 2170-2187.
117. McGarvey, T.W., T.B. Nguyen, and S.B. Malkowicz, *An interaction between apolipoprotein E and TERE1 with a possible association with bladder tumor formation*. *Journal of cellular biochemistry*, 2005. **95**(2): p. 419-428.
118. Diomedi-Camassei, F., et al., *COQ2 nephropathy: a newly described inherited mitochondriopathy with primary renal involvement*. *J Am Soc Nephrol*, 2007. **18**(10): p. 2773-80.
119. Chen, Y.P., et al., *Mutation scanning of the COQ2 gene in ethnic Chinese patients with multiple-system atrophy*. *Neurobiology of aging*, 2015. **36**(2): p. 1222. e7-1222. e11.
120. Sadowski, C.E., et al., *A single-gene cause in 29.5% of cases of steroid-resistant nephrotic syndrome*. *Journal of the American Society of Nephrology*, 2015. **26**(6): p. 1279-1289.
121. Collaboration, M.-S.A.R., *Mutations in COQ2 in familial and sporadic multiple-system atrophy*. *New England Journal of Medicine*, 2013. **369**(3): p. 233-244.
122. Boral, D., V. Rao, and S. Ramasamy, *Archeal Di-O-geranylgeranyl glyceryl phosphate synthase of a UbiA superfamily member provides insight into the multiple human diseases*. *Protein and peptide letters*, 2019.
123. Vinothkumar, K.R. and R. Henderson, *Structures of membrane proteins*. *Quarterly reviews of biophysics*, 2010. **43**(1): p. 65-158.
124. Derewenda, Z.S. and A. Godzik, *The "Sticky Patch" Model of Crystallization and Modification of Proteins for Enhanced Crystallizability*, in *Protein Crystallography*. 2017, Springer. p. 77-115.
125. Cooper, D.R., et al., *Protein crystallization by surface entropy reduction: optimization of the SER strategy*. *Acta Crystallographica Section D: Biological Crystallography*, 2007. **63**(5): p. 636-645.
126. Trbovic, N., et al., *Protein side-chain dynamics and residual conformational entropy*. *Journal of the American Chemical Society*, 2009. **131**(2): p. 615-622.
127. Bill, R.M., et al., *Overcoming barriers to membrane protein structure determination*. *Nature biotechnology*, 2011. **29**(4): p. 335-340.
128. Ruiz, N., D. Kahne, and T.J. Silhavy, *Advances in understanding bacterial outer-membrane biogenesis*. *Nat Rev Microbiol*, 2006. **4**(1): p. 57-66.
129. Zhou, Z., et al., *Function of Escherichia coli MsbA, an essential ABC family transporter, in lipid A and phospholipid biosynthesis*. *Journal of Biological Chemistry*, 1998. **273**(20): p. 12466-12475.

130. Zughailer, S.M., et al., *Differential induction of the toll-like receptor 4-MyD88-dependent and-independent signaling pathways by endotoxins*. Infection and immunity, 2005. **73**(5): p. 2940-2950.
131. Boll, J.M., et al., *Reinforcing lipid A acylation on the cell surface of Acinetobacter baumannii promotes cationic antimicrobial peptide resistance and desiccation survival*. MBio, 2015. **6**(3): p. e00478-15.
132. Somerville, J.E., L. Cassiano, and R.P. Darveau, *Escherichia coli msbB gene as a virulence factor and a therapeutic target*. Infection and immunity, 1999. **67**(12): p. 6583-6590.
133. Xu, H., et al., *Role of the lpxM lipid A biosynthesis pathway gene in pathogenicity of avian pathogenic Escherichia coli strain E058 in a chicken infection model*. Veterinary microbiology, 2013. **166**(3-4): p. 516-526.
134. d’Hauteville, H., et al., *Two msbB genes encoding maximal acylation of lipid A are required for invasive Shigella flexneri to mediate inflammatory rupture and destruction of the intestinal epithelium*. The Journal of Immunology, 2002. **168**(10): p. 5240-5251.
135. Goldman, S.R., Y. Tu, and M.B. Goldberg, *Differential regulation by magnesium of the two MsbB paralogs of Shigella flexneri*. Journal of bacteriology, 2008. **190**(10): p. 3526-3537.
136. Tamura, K., et al., *MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods*. Molecular biology and evolution, 2011. **28**(10): p. 2731-2739.
137. Eswar, N., et al., *Comparative protein structure modeling using Modeller*. Current protocols in bioinformatics, 2006. **15**(1): p. 5.6. 1-5.6. 30.
138. Wiederstein, M. and M.J. Sippl, *ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins*. Nucleic acids research, 2007. **35**(suppl_2): p. W407-W410.
139. Holm, L. and P.i. Rosenström, *Dali server: conservation mapping in 3D*. Nucleic acids research, 2010. **38**(suppl_2): p. W545-W549.
140. Boral, D., K.R. Vankudoth, and S. Ramasamy, *Structural Insight into a Membrane Intrinsic Acyltransferase from Chlorobium tepidum*. Current microbiology, 2019. **76**(11): p. 1290-1297.
141. Sodhi, J.S., et al., *Predicting metal-binding site residues in low-resolution structural models*. Journal of molecular biology, 2004. **342**(1): p. 307-320.

List of publications

- **Boral, Debjyoti**, Koteswara Rao Vankudoth, and Sureshkumar Ramasamy. "Structural Insight into a Membrane Intrinsic Acyltransferase from *Chlorobium tepidum*." *Current microbiology* 76.11 (2019): 1290-1297.
- **Boral, Debjyoti**, and Sureshkumar Ramasamy. "Structural characterization of a UbiA superfamily member of archaeal origin." *ACTA CRYSTALLOGRAPHICA A-FOUNDATION AND ADVANCES*. Vol. 73. 2 ABBEY SQ, CHESTER, CH1 2HU, ENGLAND: INT UNION CRYSTALLOGRAPHY, 2017.
- **Boral, D.**, V. K. Rao, and S. Ramasamy. "Archeal Di-O-geranylgeranyl glyceryl phosphate synthase of a UbiA superfamily member provides insight into the multiple human diseases." *Protein and peptide letters* (2019).
- Ghosh, D., **Boral, D.**, Vankudoth, K. R., & Ramasamy, S.* (2019). Analysis of haloarchaeal twin-arginine translocase pathway reveals the diversity of the machineries. *Heliyon*, 5(5), e01587.
- Sonawane, Shweta Kishor, Abhishek Ankur Balmik, **Debjyoti Boral**, Sureshkumar Ramasamy, and Subashchandrabose Chinnathambi. "Baicalein suppresses Repeat Tau fibrillization by sequestering oligomers." *Archives of Biochemistry and Biophysics* 675 (2019): 108119.
- Shweta Kishor Sonawane, Hariharakrishnan Chidambaram, **Debjyoti Boral**, Nalini Vijay Gorantla, Abhishek Ankur Balmik, Abha Dangi, Sureshkumar Ramasamy, Udaya Kiran Marelli and Subashchandrabose Chinnathambi. "EGCG impedes human Tau aggregation and interacts with Tau" *Nature Scientific Reports* 2020 (Accepted)