Functional and in-silico characterization of Mce1R of *Mycobacterium tuberculosis*

A thesis

submitted in partial fulfillment of the requirements

for the award of the degree of

DOCTOR OF PHILOSOPHY

in

BIOTECHNOLOGY

by

Dipanwita Maity

(Roll No. 701636)

under the guidance of

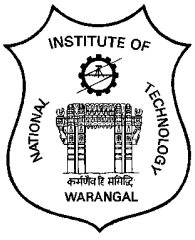
Dr. Amitava Bandhu

Associate Professor, Department of Biotechnology, NIT Warangal

and

Dr. P. Sreenivasa Rao

Professor, Department of Biotechnology, NIT Warangal



DEPARTMENT OF BIOTECHNOLOGY

NATIONAL INSTITUTE OF TECHNOLOGY WARANGAL

September, 2023

DEDICATED TO

MY PARENTS

NATIONAL INSTITUTE OF TECHNOLOGY
WARANGAL – 506004.
DEPARTMENT OF BIOTECHNOLOGY

THESIS APPROVAL FOR Ph.D.

The thesis entitled **"Functional and in-silico characterization of Mce1R of *Mycobacterium tuberculosis*"** by **Mrs. Dipanwita Maity, Roll No. 701636** is approved for the degree of **Doctor of Philosophy.**

Examiners

Supervisors

Dr. Amitava Bandhu

Associate Professor
Department of Biotechnology
National Institute of Technology Warangal

Dr. Parcha Sreenivasa Rao

Professor
Department of Biotechnology
National Institute of Technology Warangal

Chairman

Dr. Prakash Saudagar
Associate Professor & Head
Department of Biotechnology
National Institute of Technology Warangal

Date:

NATIONAL INSTITUTE OF TECHNOLOGY
WARANGAL – 506004.
DEPARTMENT OF BIOTECHNOLOGY

CERTIFICATE

This is to certify that the thesis entitled **"Functional and in-silico characterization of Mce1R of *Mycobacterium tuberculosis*"** is being submitted by **Mrs. Dipanwita Maity, Roll No. 701636** in the partial fulfilment for the award of the degree of **Doctor of Philosophy** (**Ph.D.**) in the **Department of Biotechnology, National Institute of Technology, Warangal**, is a record of bonafide work carried out by her under our guidance and supervision. The results embodied in this thesis have not been submitted to any other universities or institutes for the award of any degree or diploma.

Dr. Amitava Bandhu
(Supervisor)
Associate Professor
Department of Biotechnology
National Institute of Technology Warangal

Dr. Parcha Sreenivasa Rao
(Co-supervisor)
Professor
Department of Biotechnology
National Institute of Technology Warangal

NATIONAL INSTITUTE OF TECHNOLOGY
WARANGAL – 506004.
DEPARTMENT OF BIOTECHNOLOGY

DECLARATION

This is to certify that the work presented in the thesis entitled **"Functional and in-silico characterization of Mce1R of *Mycobacterium tuberculosis*"** is a bonafide work done by me under the supervision of **Dr. Amitava Bandhu** and **Prof. Parcha Sreenivasa Rao** and has not been submitted elsewhere for the award of any degree.

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all the principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea / data / fact / source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Date:

Place: Warangal

Dipanwita Maity

Roll No: 701636

ACKNOWLEDGEMENTS

Firstly, I would like to express my deepest gratitude to my supervisor, **Dr. Amitava Bandhu**, Associate Professor and Co-supervisor **Prof. Parcha Sreenivasa Rao**, Department of Biotechnology for the continuous support and inspiration throughout my research. It is because of their immense motivation and scientific freedom that encouraged me to continue work with new ideas. Under their esteemed guidance, I have learnt a lot, overcame many difficulties and have been inspired for which I would always be indebted.

I was supported by the fellowship from the Science and Engineering Research Board (SERB), under the Department of Science and Technology (DST), Govt. of India initially during my tenure.

I sincerely thank the Director, National Institute of Technology Warangal and other authorities who gave me this opportunity to carry out my research work at this prestigious institute.

I wish to extend my heartfelt thanks to past and present Heads of the Department of Biotechnology, for providing necessary laboratory facilities and helping me in this journey.

I would like to thank my doctoral scrutiny committee, **Dr. Vidyasagar Shilapuram**, Associate Professor, Department of Chemical Engineering, **Dr. Asim bikas Das**, Associate Professor, Department of Biotechnology and **Dr. V. Kohila**, Associate Professor, Department of Biotechnology, National institute of Technology Warangal for their valuable comments and encouragement in all stages of my research path till date.

I would like to extend my thanks to all the faculty members in the Department of Biotechnology who have directly or indirectly helped me during my work. I also acknowledge all the supporting technical staffs of the Department of Biotechnology for providing research environment and necessary facilities when required.

I would like to thank my lab mates **Dr. Katreddy Rajasekhara Reddy, Mr. Dheeraj Singh, Mr. E. Rohith Raj** and all the research scholars in the Department of Biotechnology for supporting me during my work.

Most of all, I express my deepest gratitude and indebtedness to my parents and parents in laws and my beloved husband and my sons. Whatever I am today is all because of

their immeasurable sacrifices, endless love and support. Even in the most difficult situations unflinching courage and inspiration have led me to continue my research work without which this would not have been possible. I also owe a very much appreciation to my supportive sister and brother.

I assuredly behold a ultimate thanks to "**GOD**" for warmth and blessings showered upon me. It aided in keeping up my strength and patience to work hard even a tough moments of my journey till here and always.

Dipanwita Maity

Abstract

Mycobacterium tuberculosis mostly causes latent tuberculosis infection in humans by entering in a state of "Dormancy" where it silently resides within the host cellular system without establishing active disease symptoms. During dormancy, the pahogen utilizes host derived fatty acids and cholesterol as sole sources of carbon and energy to promote survival and pathogenicity. *M. tuberculosis* imports fatty acids from extracellular environment through Mce1 transporter which is under transcriptional control of Mce1R, a VanR-type regulator. Mce1R deletion mutants of *M. tuberculosis* are unable to cause persistent infection which makes Mce1R a novel drug target for anti-tuberculosis drug discovery approaches. In this work functional and in-silico characterization of Mce1R has been performed to some extent. Mce1R gene has been cloned, expressed and purified to homogeneity. Purified Mce1R could specifically bind to the *mce1* promoter DNA (operator) with moderate affinity ($K_d = 0.35 \pm 0.02$ μM). Initially, the monomeric unit structure of Mce1R has been modeled using Phyre2 server and validated by computational and experimental methods. Since VanR-type regulators form dimers, the dimeric model of Mce1R was modeled using the Galaxy Homomer server and validated again. The structure is found to carry an N-terminal unstructured arm with distinct N- and C-terminal domains like that of VanR-type regulators. Coarse grain molecular dynamics simulation suggests that the N-terminal domain including the N-terminal arm structure is more flexible while the C-terminal domain is comparatively rigid. Structure-guided sequence alignment among the structural orthologs of Mce1R revealed that the N-terminal domain of Mce1R is rich in many highly conserved residues while the C-terminal domain residues are mostly substituted by similar types of residues which suggests that structural dynamics of Mce1R is preserved among the structural orthologs. A ligand-binding cavity has been identified at the C-terminal domain of Mce1R and through binding site matching approach employed by the ProBis server; fatty acids were selected as possible ligands for Mce1R. Molecular docking followed by analysis of Mce1R-fatty acids interactions reveled that several cavity residues are mediating hydrophobic interaction with the fatty acid ligands. All atom molecular dynamics simulation of the docked complexes using GROMACS suggests that ligand binding stabilizes the structure of Mce1R. Interestingly, Mce1R is found to preferably form stable complexes with long chain fatty acids and undergo distinct structural changes after binding.

Organization of the thesis

Chapter 1: This chapter contains introduction and review of literatures. It represents the infection cycle of *Mycobacterium tuberculosis* and current understandings on mycobacterial pathogenicity, genetic factors important for virulence and dormancy. It also summarizes current understanding on genetic organization, structure, function and regulation of *mce* operons and their roles in pathogenicity.

Chapter 2: This chapter underlines gaps in the present knowledge and the objectives framed for the present study.

Chapter 3: This chapter contains detailed descriptions of the materials used in this study and the methods used to complete the objectives. It includes detailed description of molecular cloning, protein expression and purification methods, in-vitro DNA binding experiments, etc., and in-silico characterization of protein structure, function and dynamics.

Chapter 4: This chapter includes a detailed description of the results obtained in this study and in-depth discussion. This chapter is divided into four sections. The section I consists of results of molecular cloning, expression and purification followed by DNA binding activity analysis of Mce1R. Section II contains the in-silico results for structure determination, validation, structural dynamics, sequence analysis and identification of cavity of Mce1R. In the section III ligand identification, molecular docking and analysis of the interaction between ligand and Mce1R has been discussed. Section IV consists of all atom molecular dynamics simulation analysis of the docked complexes, ligand induced structural changes in Mce1R and analysis of stabilities of the complexes have been included with detailed discussion.

Chapter 5: This chapter represents conclusion and scope for future research.

List of abbreviations and symbols

Symbols	Description
Å	Angstrom
°C	Degree Celsius
fs	femtosecond
g	Gram
g/mol	Gram per mole
hr	Hour
K	Kelvin
kBp	kilo Base pair
kCal/mol	kilo Calorie per mole
kDa	kilo Dalton
kJ/mol	kilo Joule per mole
λmax	Lambda max
l	litre
M	Molar
µg	microgram
µg/ml	microgram per millilitre
µl	microlitre
µM	micromolar
mBp	mega Base pair
mg	milligram
mg/ml	milligram per milliliter
ml	millilitre
mm	millimeter
mM	millimolar
mV	millivolt
min	minutes
ng	nanogram
nm	nanometer
nM	nanomolar
ns	nanosecond
ps	picosecond

%	percentage
rpm	rotations per minute
sec	seconds
v/v	volume by volume
w/v	weight by volume
Abbreviations	**Descriptions**
APS	Ammonium per sulphate
ATP	Adenosine tri Phosphate
BSA	Bovine Serum Albumin
CTD	C–terminal domain
EDTA	Ethylene di amine tetra acetic acid
EtBr	Ethidium Bromide
FCD	FadR C-terminal domain
IgG–AP	Immunoglobulin G Alkaline Phosphatase
IPTG	Isopropyl β D Thio Galactopyranoside
MMGBSA	Molecular Mechanics Generalized Born Surface Area
NCBI	National Centre of Biotechnology Information
Ni^{2+}–NTA	Nickel ion –nitrilotriacetic acid
NMR	Nuclear magnetic resonance
NTD	N–terminal domain
OADC	Oleic acid Albumin Dextrose Catalase
ORF	Open Reading Frame
PAGE	Poly Acrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Poly Ethylene Glycol
PSI–BLAST	Position Specific Iterated Basic Local Alignment Search Tool
PVDF	Polyvinyl di Fluoride
R_{gavg}	Average radius of gyration
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SDS	Sodium dodecyl sulphate

TBE	Tris borate EDTA
TBS	Tris buffer saline
TBST	Tris buffer saline tween 20
TEMED	Tetra–methyl Ethylene di–amine
TGS	Tris glycine SDS
UV	Ultra Violet
VMD	Visual Molecular Dynamics
wHTH	winged helix turn telix

List of figures

List of tables

Contents

CHAPTER 1

INTRODUCTION

AND

REVIEW OF LITERATURE

1 Introduction

1.1 *Mycobacterium tuberculosis*, a dreadful pathogen

The pathogen *Mycobacterium tuberculosis*, initially identified by Dr. Robert Koch in 1882 [23], continues to challenge public health world–wide by causing the communicable disease– Tuberculosis. According to the recent Global Tuberculosis Report, published by the World Health Organization, Geneva, Switzerland in 2022, Tuberculosis caused highest deaths world–wide apart from COVID–19 induced pandemic [1]. The report also describes that nearly one–quarter of the total world population is infected from this pathogen which depicts the severity of the disease. Tuberculosis infections in patients are of two types– active and latent [2]. Almost 10% individuals develop active form of Tuberculosis upon exposure to the pathogen which is characterized by the continued replication of the pathogen within the host system followed by clinical manifestation of the disease symptoms. The remaining 90% individuals develop latent tuberculosis infection where the pathogen undergoes a dormant stage upon infecting the host and maintains an asymptomatic form of infection characterized by absence of clinical disease symptoms [2]. This state of infection is stable for years to decades until the immunity of the individual wanes; i.e. due to ill health, poor lifestyle, HIV co–infection or use of immunosuppressive drugs [3]. Unlike active Tuberculosis infection, latent form of infection is difficult to treat as the dormant pathogen shows tolerance to many anti–microbial drugs, thus creating an obstacle to eradicate the disease. Moreover, the latent Tuberculosis infection initiates the generation of multi–drug resistant (MDR) strains of *M. tuberculosis* which are resistant against frontline drugs– isoniazid and rifampicin [4]. In the year 2019, a new drug regimen approved by the FDA (Food and Drug Administration), BPaL which is consisted of three novel drugs bedaquiline, pretomanid, and linezolid, has been introduced for the treatment of MDR strains of *M. tuberculosis* for a duration of 6–9 months which is a significant improvement over the standard 20 months treatment regimen [5]. Additionally, two other drug resistant strains– extensively drug resistant (XDR) [6] and totally drug resistant (TDR) [7] variants of *M. tuberculosis* also generated against which no effective treatment regime is available presently [8]. Clearly, generation of different drug resistant variants of this pathogen has made the scenario more difficult to combat against this disease. Therefore a deeper understanding on *M. tuberculosis* infection cycle, pathogenic determinants and identification of new drug targets could lead towards novel drug discovery approaches which would be helpful to eradicate the disease.

1.2 The infection cycle of *M. tuberculosis*: an overview

M. tuberculosis disseminates from an infected person, carried through micro droplets in a form of aerosol, generated during coughing or sneezing. These micro droplets may enter into the body of a healthy individual through the respiratory tract. To cause a successful infection, it is found that only 2–3 bacilli are sufficient in a micro droplet underlining the capacity of the pathogen to evade host immunity with such a low infective dose [9]. After reaching at the lower respiratory tract the bacilli are internalized by the process of phagocytosis by the immune cells– neutrophil, macrophage and dendritic cells [10]. Additionally, *M. tuberculosis* also invades cells from non–myelocytic origins such as epithelial and M cells (microfold cells) of the lungs to initiate infection [10, 11]. Infected dendritic cells will elicit a local inflammatory response which includes migration to the draining lymph nodes to prime the T cells which then return to the site of infection. This continues until an effective immune response develops which then recruits T cells, B cells and activated macrophages surrounding the infected cells to eventually develop a structure which is called "granuloma"; the pathological hallmark of tuberculosis infection [12–14]. The environment of granuloma is very different than those of healthy tissues which renders the pathogen to stop replication thereby the immunological control over bacterial burden is established; resulting in the state of infection termed latent infection [15]. A latently infected person develops no symptoms of active tuberculosis disease although elicits an adaptive immune response without the sign of culturable bacilli. This state leads to two possible outcomes which include clearance of the pathogen by host immunity or subclinical infection [15]. The state of subclinical infection is quite stable if the host immunity could contain the pathogen inside the granuloma; if not, the infection leads to develop active tuberculosis diseases of various types– from cavitary lung disease to focal infection [15, 16]. The cavitary lung disease is most infectious among other types and can disseminate viable bacilli in microdroplets generated during coughing or sneezing by the patients which can infect a healthy person. The infection cycle of *M. tuberculosis* is shown in the Fig. 1.1. Dissemination of the pathogens to any organ can also occur through the lymphatics and lymph nodes in extrapulmonary tuberculosis infection [17]. Studies reveal that the M cells contribute to dissemination of pathogens in such type of infection to lymph nodes where the lymphatic endothelial cells are the primary targets [18, 19]. The other target sites for the extrapulmonary tuberculosis infection are the adipose tissue and bone marrow where the pathogen could persist for a long time [20, 21]. It is therefore quite clear that long term persistence of *M. tuberculosis* within the host system (different target cells) represents a major hurdle to eradicate the disease.

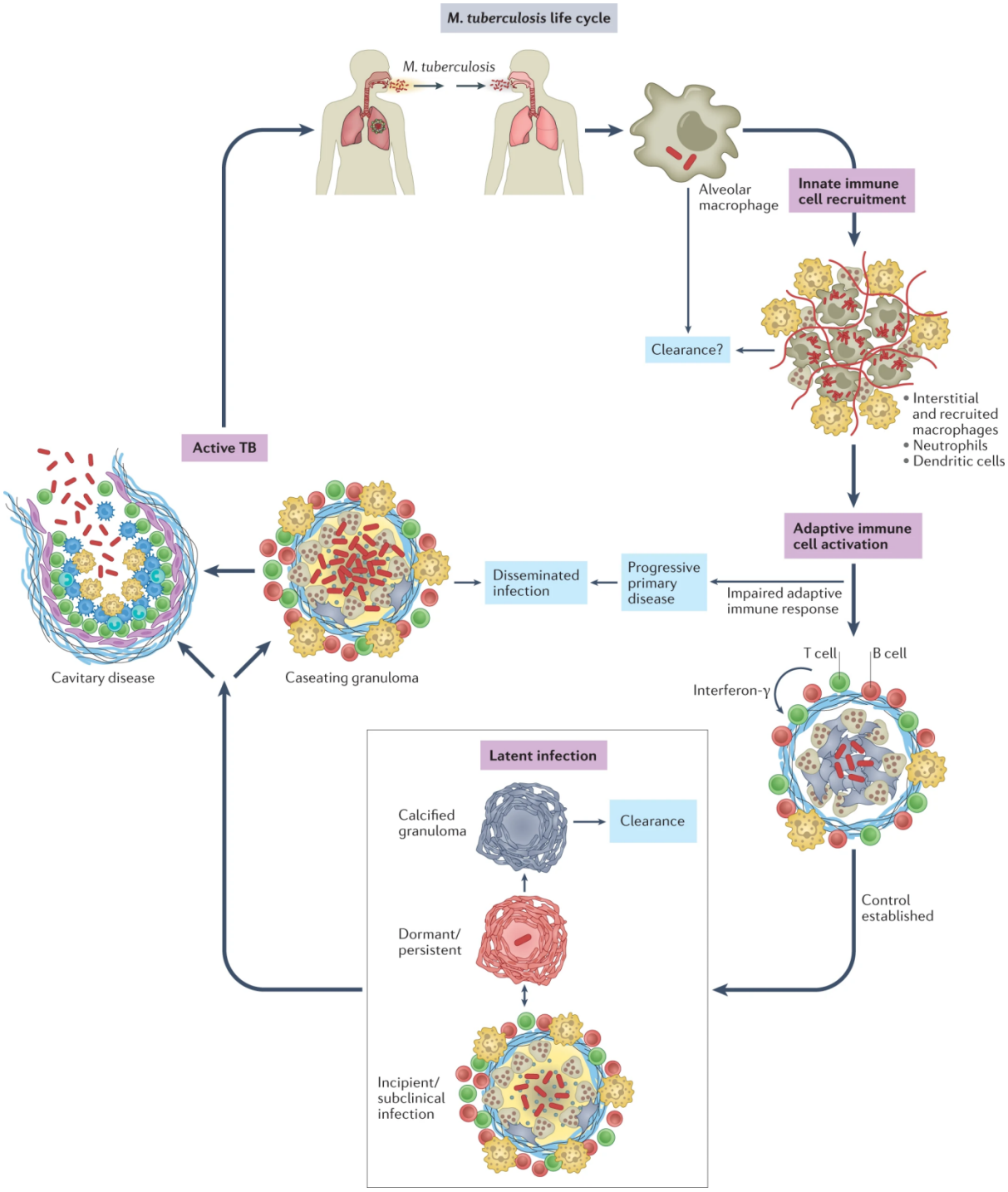


Fig. 1.1 **The infection cycle of *Mycobacterium tuberculosis*.** The possible stages in the pulmonary tuberculosis infection are shown here. The picture is taken from Chandra et al. 2022 [22].

1.3 Dormancy in *M. tuberculosis* infection

As described earlier; although dormancy is one of the two possible outcomes of tuberculosis infection to a patient, in ~90% cases it is the general outcome [2]. From the pathogen's

survival perspective, choosing dormancy is somewhat beneficial for the pathogen as during latent infection, *M. tuberculosis* can persist within the host without being eliminated by the immune system [15]. In this state of infection, the pathogen remains metabolically less active and does not replicate under unfavorable metabolic and immunological conditions faced in the host. *M. tuberculosis* adopts various strategies for survival including active functioning of efflux pumps, formation of biofilms, altered metabolic strategies, etc. Eventually, this serves as the generation of persisters and/or phenotypically distinct traits of the pathogen; such as acquiring resistance to antimicrobial drugs [24]. Therefore, to explore the molecular mechanisms of initiation and sustenance of dormancy, various in–vivo and in–vitro models have been used to mimic several environmental parameters which are encountered during dormancy [24–26]. Studies using various models revealed a large number of genes were upregulated during dormancy. These genes belong to several classes such as– dormancy regulon genes which also includes two component systems, a number of alternate sigma factors, genes maintaining redox homeostasis, synthesis of cell wall–specific complex lipids, transport and metabolism of host derived lipids including fatty acids and cholesterol [27–31]. Interestingly, the genome of *M. tuberculosis* is found to carry 250 genes, encoding enzymes required for synthesis and degradation of lipids; a highest number of genes dedicated for lipid metabolism compared to other prokaryotes [32, 33]. Moreover, lipids constitute up to ~60% of cellular dry weight of *M. tuberculosis* [34] which suggests that lipid metabolism plays critical roles in cellular physiology and pathogenicity [35, 36].

1.4 Cell envelope lipids of *M. tuberculosis*

Majority of the cellular lipids are present at the cell envelop. The cell envelop of *M. tuberculosis* is unusually rich in lipids which accounts for 40% of cell envelop weight [37]. The structure of cell envelop of *M. tuberculosis* is complex compared to conventional gram positive and gram negative bacteria. The cytosol is enclosed by plasma membrane composed of a lipid bilayer. A thick layer of peptidoglycan is present outside the plasma membrane which is covalently attached with phosphatidylinositol mannoside (PIM) residues. The peptidoglycan layer is surrounded by a carbohydrate polymer, arabinogalactan, covalently attached by lipoarabinomannan (LAM). An outer layer, composed of mycolic acid and other extractable lipids like phthiocerol dimycocerosates (PDIM), sulfolipid–1(SL–1), trehalose monomycolate (TMM), trehalose dimycolate (TDM), phenolic glycolipid (PGL) etc. surrounds the entire cellular structure [38, 39]. The cell envelop structure is shown in Fig.

1.2. Presence of various lipids at the cell envelop of *M. tuberculosis* makes the cell wall more impervious which also contribute to inherent drug tolerance [41].

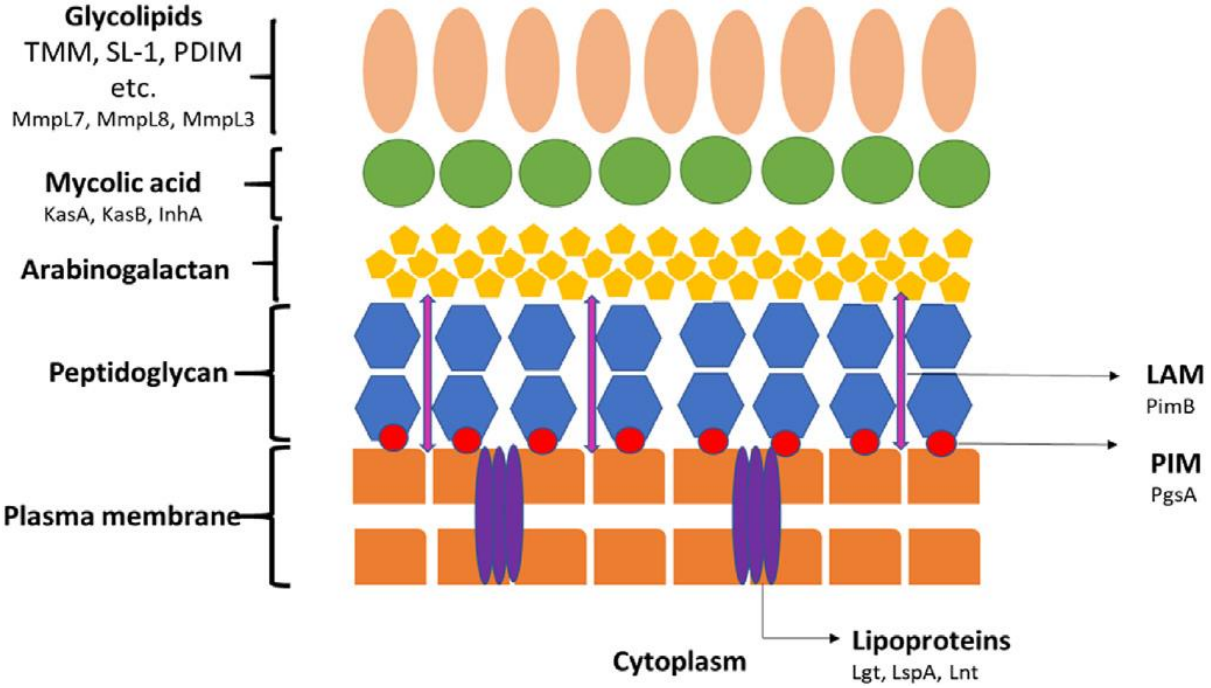


Fig. 1.2 The cell envelope of *M. tuberculosis*. The proteins involved in the transport of PDIM (Mmpl7) [34], synthesis of SL–1 (Mmpl8) [42], transport of TMM (Mmpl3) [43], biosynthesis of mycolic acid (KasA, KasB and InhA) [44], assembly of LAM (PimB) [46] and assembly of PIM (PgsA) [45] are shown below their respective lipids. Lgt, LspA and Lnt are responsible for anchoring the lipoproteins to the lipid bilayer [47]. The figure is taken from Moopanar et al. 2020 [40].

1.4.1 Cell envelop lipids contribute to pathogenicity of *M tuberculosis*

Components of the cell envelops of many bacteria are found to detach from the outermost part; many of which are known to interfere with the host functions, such as LPS (lipopolysaccharide) of gram negative bacteria [48]. Similarly, *M. tuberculosis* is also reported to shed various lipids from its outermost layer of the cell envelop through specific secretion pathway and fusion of membrane vesicles with host cell membrane which modulate host immune response to evade the defense mechanisms [49–52]. Here, roles of some of the cell envelop lipids in pathogenicity of *M. tuberculosis* are discussed.

The lipid mycolic acid is the most abundant component of the outermost layer of *M. tuberculosis* cell envelop, present as free entity or conjugated with arabinogalactan layer [53, 54]. Mycolic acid decreases the production of IL–8 by inhibiting TLR2 mediated signaling in

alveolar epithelial cells [55]. It also activates DAP12 associated triggering receptor on myeloid cell 2 to increase MCP–1 production and recruiting macrophages to promote active disease progression [56, 57]. By regulating abundance and structure of mycolic acid in response to the extracellular environment, *M. tuberculosis* can modulate host immune response [58]. The PGL present in the outer layer also similarly induces secretion of MCP–1 which interacts with the host chemokine receptor 2 (CCR2) and recruits macrophages permissive to the pathogen [62]. The lipids– TMM and TDM are shed by *M. tuberculosis* also influence host immune response by binding with the pattern recognition receptors (PRR) of macrophages and dendritic cells to trigger SHP–1 and FcγRIIB signaling pathways to inhibit phagosome arrest [59–61]. Another cell envelop lipid PDIM masks the pathogen–associated molecular patterns (PAMPS) present on the cell surface unique molecules of *M. tuberculosis*, thereby helping the pathogen evading TLR–mediated recognition by the innate immune cells [62, 63]. PIM, LAM and manLAM (additional mannose residue attached with LAM) functions similarly. They interact through their mannose residues with several C–type lectin receptors of macrophages and dendritic cells to initiate phagocytosis [64, 65], facilitating uptake of the pathogen and providing a niche for initial replication. They also inhibit phagosome maturation and acidification to protect the pathogen from getting degraded [66, 67].

1.5 Lipids are sources of nutrients and units of biosynthesis in *M. tuberculosis*

Use of lipids as nutrient source for *M. tuberculosis* was first discovered by Segal and Bloch with the observation that *M. tuberculosis*, cultured from mouse lung, could perform respiration *ex–vivo* in presence of lipids but not in presence of carbohydrates [68]. Afterwards, several studies have established that *M. tuberculosis* can utilize host derived lipids (fatty acids and cholesterol) as carbon and energy sources during infection [32, 69]. Lipid utilization is particularly important as it has been observed that fatty acid and cholesterol metabolizing genes were specifically upregulated during tuberculosis infection [71–73]. Consistent to this; the extracellular environment of *M. tuberculosis* is found to be rich in fatty acids and cholesterol. Macrophages infected with *M. tuberculosis* accumulate lipid bodies composed of triacyl glycerol (TAG) and cholesteryl ester therefore appear as foamy macrophages. The pathogen secrets various lipolytic enzymes to degrade those lipid inclusions of macrophages and imports simple lipids (fatty acids and cholesterol) into the cytosol which are later metabolized during persistence [74, 75]. Metabolism of fatty acids and cholesterol is used to fuel central carbon metabolism and the end products are mediating

pathogenicity, survival and drug tolerance [70]. The metabolic product of cholesterol degradation, propionyl–CoA [76] is directed towards three different metabolic pathways– (a) assimilation in methylcitrate cycle [77, 78], (b) assimilation in methylmalonyl cycle [76] and (c) biosynthetic incorporation in the synthesis of methyl branched and polyketide lipids [79]. Propionyl–CoA is converted to methylmalonyl–CoA which acts as precursor for synthesis of PDIM, polyacyl trehalose and SL–1 cell envelop lipids [79] which are important for pathogenicity. The fatty acids in the animal cells are majorly of even chain lengths [80] and therefore their degradation by β–oxidation yield pool of acetyl–CoA which is used to fuel carbon metabolism and biosynthetic processes. The polyketide synthases utilizes the fatty acyl–AMP intermediates with the addition of malonyl–CoA and/or methyl malonyl–CoA units to synthesize poleketide lipids [79]. The fatty acid synthase–II complex (FAS–II) of *M. tuberculosis* synthesizes full length acyl chains of mycolic acid using elongated fatty acyl units [81]. Fatty acids also can be directly incorporated as phospholipids into the membrane of *M. tuberculosis* and provides structural integrity and/or used to produce triacylglycerol (TAG) as lipid bodies at the cytosol [82] which could be used as sources of carbon and energy during persistent infection.

1.6 Import of lipids (fatty acids and cholesterol) through the cell wall

Although *M. tuberculosis* harbors the fatty acid synthase–I (FAS–I) enzyme and can synthesize fatty acids of $C_{16–22}$ chain lengths [83], it prefers to scavenge host–derived fatty acids and cholesterol since it is energetically less expensive than *de–novo* fatty acid biosynthetic pathway [79]. Lipids imported from outside is utilized to synthesize intracellular lipid bodies and/or cell envelop lipids. For importing lipids, *M. tuberculosis* employs four different Mce transporters (Mce1–4) [84]. Studies reveal that Mce1 transporter imports fatty acids while Mce4 transporter is involved in only cholesterol import [85–87]. The substrates transported through other two Mce transporters, Mce2 and Mce3 have not yet been identified but they are also thought to be lipids. Given the importance of the Mce transporters in lipid import, they have been studied to some extent which is described below briefly.

1.6.1 The Mce transporters

M. tuberculosis contains four *mce* operons (*mce1–4*) in its genome [32]. The core genetic organizations of these operons are highly similar. Each core operon begins with the permease genes *yrbEA* and *yrbEB* followed by six respective *mce* genes (*mceA–F*) (Fig. 1.3) [84]. These transporters are powered by the cytoplasmic ATPase MceG, coded by *rv0655*, not

localized with the *mce* operons [84, 88]. The YrbE permeases bear the conserved EExDA sequence, analogous to the EEA sequence of the ABC transporters [89], at the final cytoplasmic loop region. The ATPase MceG also displays sequence similarity with those of ATPases of ABC transporters and contains Walker A and Walker B motifs important for ATP binding. MceG also contains the signature sequence motif LSGGQ like other ATPases of ABC transporters [84]. MceA–F proteins are analogous to the SBP (substrate binding protein) proteins of the ABC transporters. These similarities suggest that Mce transporters belong to the family of ABC transporters. However, there are specific features distinct to Mce transporters.

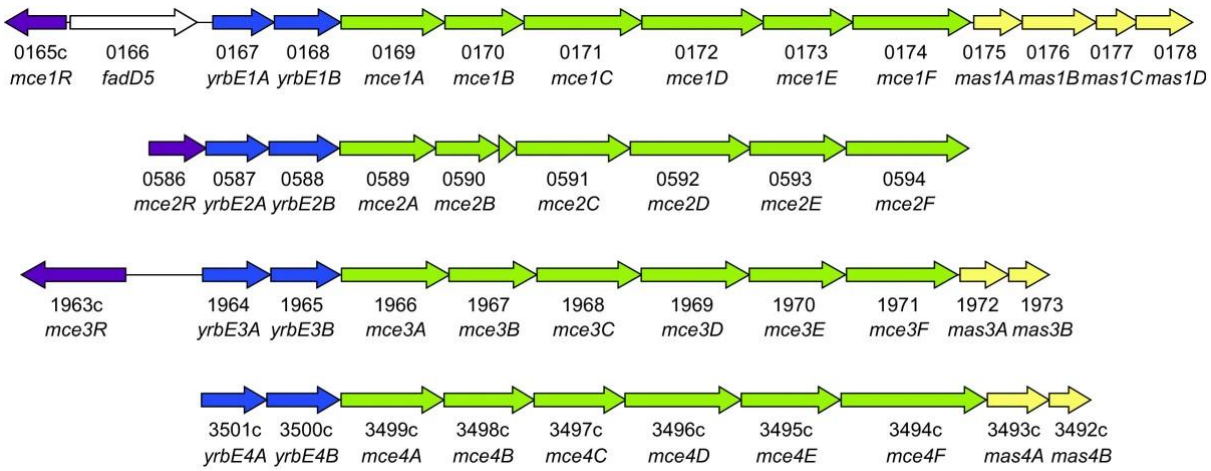


Fig. 1.3 The genetic organization of the *mce* operons of *M. tuberculosis*. The permease genes are shown in blue colors and the core *mce* genes are displayed in green colors. Other accessory genes are depicted in yellow colors. The three transcriptional regulator genes *mce1R*, *mce2R* and *mce3R* for the respective *mce* operons are shown in purple colors. The figure is taken from Casali et al. 2007 [84].

The Mce transporters employ six SBPs (MceA–F) [84] while the ABC transporters have one dedicated SBPs. Each Mce protein (MceA–F) contains a Cholesterol_Uptake_Portar_1 (CUP_1) domain composed of α helices and a variable domain in addition to the Mce domain common to the ABC transporters [90]. Moreover, Mce Associated Membrane (Mam) proteins and Orphaned Mce Associated Membrane (Omam) proteins also contribute to the functions played by these Mce transporters. Consistent to these; OmamA is reported to promote the import of palmitic and oleic acids through the Mce1 transporter and cholesterol import through the Mce4 transporter [86, 91]. OmamA also stabilizes the components of Mce1 and Mce4 transporters. Very recently, another protein

LucA is also reported to play similar role like that of OmamA [86]. These observations suggest that the Mce transporters might function as large multi–protein transmembrane complexes. This is supported by the recent evidences which showed that the Mce1 and Mce4 transporters of *M. smegmatis*, a non–pathogenic relative of *M. tuberculosis*, functioned as large multi–protein transmembrane complexes [92, 93].

Although the much of the functions of the Mce proteins are known today but the detailed mechanism of lipid import through these transporters remains to be determined. However, apart from importing lipids, Mce proteins are also involved in entry of the pathogen in to the host system and modulating host immune response. For example; Mce1A and Mce3C proteins promote uptake of *M. tuberculosis* by the host cells [94–96]. In *M. bovis* BCG, the Mce2E and Mce3E proteins inhibit ERK1/2 signaling pathway to reduce the expression of TNF and cytokine IL–6 [97, 98]. Deletion mutants of *mce1–4* operons result in attenuated infection in a murine model which indicates that Mce proteins are also important for pathogenicity [99–101]. However, there are some discrepancies regarding the results for the deletion mutant of *mce1* operon in mice model. A few studies reported that deletion of *mce1* operon caused hypervirulent symptom in the BALB/c mice when infected through tail vain [102], or intraperitoneally [99]. Another study showed *mce1* operon deletion mutant caused increased bacterial burden in C57BL/6 mice lungs when infected through aerosol route [103]. On the other side; when BALB/c mice was infected with *mce1* operon deletion mutant of *M. tuberculosis* intratracheally, virulence attenuated phenotype was observed [99]. The reason for the discrepancies is not known with certainty; however it is possible that different route of infection in mice with different genetic background might have caused such discrepancies.

1.6.1.1 Regulation of expression of *mce* operons in *M. tuberculosis*

In the bacilli, isolated from the lungs of rabbit at 24 weeks post infection, expressions from *mce1*, *mce3* and *mce4* operons were detected while the *mce2* operon was repressed [105]. In separate studies using RAW murine macrophage and bone marrow derived murine macrophage infection models, the *mce1* operon genes are found to be downregulated till 48 hr post infection [72, 104]. In the spleens of the guinea pigs; only the expression of *mce4* operon is observed at 16 weeks post infection [105]. All these reports suggest that expressions of the *mce* operons are regulated differentially under different host cellular environments and tissue–specific manner. *M. tuberculosis* employs specific transcriptional regulators to regulate the expression of the *mce* operons. The expressions of *mce1–3* operons

are regulated by the transcriptional regulators Mce1R, Mce2R and Mce3R respectively, associated with the respective operons (Fig. 1.3) [104, 106, 107]. The *mce4* operon is regulated by KstR1 [108, 109], located far away from the operon.

These transcriptional regulators have been studied to some extent. The transcriptional regulators, KstR1 and Mce3R have been classified as TetR–type regulators [107, 108] while Mce1R and Mce2R belong to VanR and FadR–types of regulators respectively [110], depending on their predicted secondary structure profiles. In–vitro DNA binding studies using purified Mce2R demonstrated presence of two similar DNA binding sites within the –35 promoter element of *mce2* operon [106] which suggests that Mce2R binding might hinder promoter recognition by RNA polymerase; thereby repressing the expression. Two sequence motifs were also reported to be present in the *mce3* promoter region which are specifically bound by the transcriptional regulator Mce3R [111]. However, the binding sites are located –214 to –182 and from –142 to –111 relative to the transcription start site. Therefore it appears that Mce3R can mediate long range transcriptional control over *mce3* operon. KstR1 regulates the expression of a large number of genes including *mce4* operon by binding with specific binding sites located at the corresponding promoters of those genes [108]. It is a master regulator which controls the expression of several cholesterol metabolism genes. The regulator of the *mce1* operon, Mce1R also has been described as a global negative regulator, reported to control the expression of various genes involved in pH balance, cell wall synthesis, intracellular iron balance, virulence, antibiotic resistance, cell cycle and cell division including its own expression by binding with their respective promoters [112]. Thus far no other systemic characterizations of these regulators have been performed yet. Given the importance of these regulators in controlling the expression of *mce* operons, characterizing these regulators may provide valuable information for anti–tuberculosis drug discovery.

CHAPTER 2

OBJECTIVES

2 Objectives

2.1 Origin of the proposal

Among the regulators of the *mce* operons, Mce1R is particularly important since it was found that *mce1R* deletion mutants of *M. tuberculosis* are unable to cause persistent infection in the host [117] and it acts as a global regulator [112]. Mce1R is a VanR–type regulator [110]. A common function for TetR, VanR or FadR types of regulators is that their N–terminal domains are used for binding to cognate DNA sequences and the C–terminal domains are involved in binding to specific ligands which in turn regulate their DNA binding activities [113, 114]. Consistent to this, the DNA binding function of KstR1 is found to be inhibited by binding to its specific ligand 3-hydroxy-cholest-5-ene-26-oyl-CoA (3OCH–CoA) [115], a catabolic intermediate of cholesterol degradation pathway which is also under the transcriptional control of KstR1 [107]. Similarly, the DNA binding activity of Mce2R has been shown to be inhibited by binding to its specific ligands– long chain fatty acyl CoAs, specifically palmitoyl CoA and arachidonoyl CoA [116]. However, the natural ligand(s) for Mce1R have not yet been identified. Mce1R also regulates its own expression in addition to being a global regulator [112]. To control the expression of several genes, sufficient intracellular concentration of Mce1R should be maintained which indicates that Mce1R–mediated repression of *mce1R* promoter may not be very strict. Therefore, it is important to determine the binding affinity of Mce1R for its promoter DNA. Apart from being classified as a VanR–type regulator based on predicted secondary structure profile [110], no other structural characterization of Mce1R has been performed yet. Being an important global regulator for the persistence and virulence of *M. tuberculosis*, it holds the potentiality to act as target for novel anti–tuberculosis drug discovery approaches. Therefore it is important to characterize this regulator. Under this perspective, the objectives are set as mentioned below:-

2.2 Objectives of the present work

A) Cloning, expression, purification of Mce1R and analyzing its DNA binding activity

1. Extraction of genomic DNA from *M. tuberculosis* H37Ra
2. PCR amplification of *mce1R* gene and cloning in the expression vector pET28a
3. PCR amplification and cloning of *mce1* promoter (operator) DNA
4. Expression and purification of C–terminal His–tagged Mce1R by Ni^{2+} affinity chromatography

5. Estimating operator DNA binding affinity of Mce1R by gel–shift assay

B) Structural modeling, validation, dynamics and sequence analysis of Mce1R

1. Modeling of monomeric structure of Mce1R by Phyre2 server and its validation
2. Coarse grain molecular dynamics simulation of Mce1R by CABS–Flex server
3. Modeling of the dimeric structure of Mce1R by GalaxyHomomer server and its validation
4. Identification of conserved amino acid residues of Mce1R by multiple sequence alignment
5. Detection of ligand binding cavity in the modeled structure

C) Identification of the specific ligand(s) for Mce1R

1. Identification of putative ligand(s) for Mce1R by cavity similarity search method using the ProBis server
2. Molecular docking analysis using those ligand(s) by AutoDockVina
3. Analyzing the interactions between the ligand(s) and the cavity residues of Mce1R by the LigPlot software

D) Analyzing the stability and dynamic properties of the docked Mce1R–ligand complexes

1. All atom molecular dynamics simulation of Mce1R–ligand complexes using GROMACS
2. Analyzing RMSD, RMSF and Radius of gyration of Mce1R in those complexes
3. Analyzing ligand–induced dynamic changes in the secondary structure of Mce1R
4. Analyzing ligand stabilities and dynamics of hydrogen bonding interactions with the cavity of Mce1R
5. Determination of binding free energy for those ligand(s) following MMGBSA approach

CHAPTER 3

MATERIALS

AND

METHODS

3 Materials and methods

3.1 Materials

3.1.1 Non-radioactive chemicals

Agar powder, OADC growth supplement, Kanamycin, dehydrated media LB and 7H9 were purchased from Himedia. Tris, Glycine, Na_2HPO_4, NaH_2PO_4, Glycerol, Sodium hydroxide, Potassium acetate, Sodium acetate, Sodium chloride, o-phosphoric acid, β-mercaptoethanol, Isopropanol, Ethanol, Methanol, Butanol, EDTA (Ethylene di-amine tetra acetic acid) and PVDF membrane were purchased from MERCK (Germany). Agarose, acrylamide, bis-acrylamide, APS (Ammonium per–sulfate), TEMED (Tetra–methyl Ethylene di–amine), SDS (Sodium Dodecyl Sulfate), Proteinase K, Lysozyme, Carbinic anhydrase, EtBr (Ethidium Bromide), IPTG and coomassie brilliant blue (G and R 250), Poly dI-dC were procured from SIGMA. Plasmid extraction kit, gel extraction kit, PCR purification kit, 1 kBp DNA ladder, protein ladder, restriction enzymes, Phusion polymerase, Taq Polymerase and T4 DNA ligase were purchased from Thermo Scientific. PCR primers were supplied by Bioserve. Ni^{2+}–NTA resin was obtained from Qiagen. Sarkosyl, mouse anti–His antibody (primary antibody) were purchased form MERCK and the secondary antibody goat anti–mouse IgG–AP (alkaline phosphatase) was purchased from SIGMA. The substrate NBT/BCIP was purchased from Amresco. Developer and fixer powders were purchased from Prime Pvt. Ltd., India. X–Ray films were purchased from Fuji.

3.1.2 Radioactive chemical

[γ-^{32}P]–ATP was purchased from BRIT (Board of Radiation and Isotope Technology), Hyderabad.

3.1.3 Bacterial strains

E. coli DH5α was used for molecular cloning and *E. coli* BL21(DE3) was used for protein expression. *M. tuberculosis* H37Ra strain (Catalog no 300) was received from MTCC (Microbial Type Culture Collection), Chandigarh, India.

3.1.4 Plasticwares and Glasswares

Plasticwares like microtips, microcentrifuge tubes, autoclavable petriplates, magnetic stir bars, 15 ml and 50 ml centrifuge tubes were purchased from Tarsons Products Pvt. Ltd., India. Glasswares like 15 ml test tubes, 250 ml beakers, 100 ml, 250 ml and 500 ml conical

flasks, used for microbial culture preparation and protein induction, were purchased from Borosil International, India.

3.1.5 Filter paper

Blotting paper (3 mm) was purchased from Whatmann Limited (England).

3.2 Methods

3.2.1 Growth of bacteria

E. coli bacteria were grown in LB broth at 37ºC with or without Kanamycin (50µg/ml) with constant shaking in incubator at 180 rpm. *M. tuberculosis* H37Ra were cultured in 7H9 media with 1% OADC and 0.5% glycerol at 37ºC with constant shaking in incubator at 180 rpm.

3.2.2 Molecular biological methods

3.2.2.1 Isolation of plasmid DNA by alkali lysis method

Plasmid Extraction (Alkali lysis, mini prep method)

1. Centrifuge the 3ml of saturated culture and pellet down the cells in 1.5ml centrifuge tube.
2. Add 1ml cold STE buffer, vortex to re-suspend completely.
3. Centrifuge and pellet down the cells.
4. Add 100µl of Alkaline lysis I solution with 3µl RNAse A(20mg/ml) and vortex to complete resuspension.
5. Add 200µl of alkaline lysis II solution and invert 3-4 times gently.
6. Add 150µl of alkaline lysis III solution and invert 3-4 times gently.
7. Store on ice for 5 minutes and spin down at 12000 rpm for 15 minute at 4°C
8. Collect the sup and add equal volume of Phenol:chloroform:iso-amyl alcohol (25:24:1).
9. Mix 3-4 times inverting the tube and spin down at 5000rpm for 5 min. at room temperature.
10. Take out the aqueous layer and add double volume of absolute ethanol (cold).
11. Spin at 12000rpm for 15minute at 4°C and discard the supernatant.
12. Add 1ml of 70% ethanol (cold) and spin at 12000rpm for 15 minute at 4°C.
13. Discard the supernatant and air-dry the DNA pellet in the centrifuge tube.
14. Add 15µl of 1X TE buffer and dissolve with gentle mixing.

3.2.2.2 Digestion of DNA by restriction enzymes

Digestion of DNA with restriction endonucleases was carried out according to a standard method [118] in the buffer supplied by the manufacturer. A typical reaction mixture contained 2 – 3 µg of DNA and 1 X digestion buffer in 20 – 30 µl reaction volume. Required amount of enzyme was added to it keeping in mind that glycerol concentration in the reaction does not exceed 2.5%. The reaction mixture was incubated at the appropriate temperature, recommended for activity of the enzyme, for 3–5 h. Placing the digestion mix at 65ºC for 15 min inactivated the enzyme. Gel loading dye was added to it and loaded into the gel slot.

3.2.2.3 Agarose gel electrophoresis

Preparation of agarose gel was performed by a standard method [118]. Briefly, after weighing required amount of agarose was put into 0.5X TBE buffer. It was melted in boiling water bath followed by cooling it down to about 50°C. Ethidium bromide (EtBr) solution was added to it to a final concentration of 0.5 µg/ml followed by pouring it onto a gel-tray with a slot former placed in position. The solidified gel was placed in the electrophoresis tank, submerged in 0.5X TBE buffer. The electrophoresis was carried out at 5 – 10 V/cm after loading the samples till optimal resolution was achieved.

3.2.2.4 Elution of DNA from agarose gel

DNA purification from the agarose gel was achieved using an agarose gel extraction kit from the Thermo Fisher Scientific India Pvt. Ltd. The manufacturer supplied all the required buffers along with the kit. The procedure is as follows: gel slice containing the DNA was weighed and mixed with 1 volume of binding buffer (100 mg of gel slice was considered as 100µl). It was then incubated at 55°C with occasional vortexing till the gel slice dissolved completely. This solution was passed through a GenJET spin column by centrifugation in a microfuge for 1 min at full-speed. The flow-through was discarded and the column was washed with 0.7 ml of wash buffer (working solution prepared by mixing the supplied one with ten volumes of dehydrated ethanol). All the flow-through was discarded. The column was spun again for one additional min to ensure complete removal of ethanol from the column. The DNA was eluted with 50 µl of elution buffer (10 mM Tris-HCl, pH 8.0).

3.2.2.5 Ligation of DNA fragments

DNA ligation was essentially carried out by a standard procedure [118]. Briefly, digested vector DNA (~ 100 ng) was mixed with digested insert DNA at a molar ratio of 1:3 and 1:5 in a 10 µl reaction volume for cohesive and blunt end ligation reactions, respectively. To the reaction mixture 1 µl of 10 x ligation buffer and 1 µl of T4 DNA ligase were added and mixed well. The ligation reaction was carried out at either 16ºC for 16 hr or 22ºC for 1 hr for cohesive and blunt end ligations.

3.2.2.6 Labeling of 5’ end of linear DNA fragments by γ-^{32}P ATP

It was done by the method as demonstrated by Sambrook et al. (2014) [118]. Briefly, T4 polynucleotide kinase was used to label 5’ end of DNA with γ-^{32}P ATP. Reaction mixture containing 2 pmol of 5’-termini of DNA was labeled in 20 µl reaction volumes in the

presence of 2 µl of 10x reaction buffer B, 40 pmol of γ-^{32}P ATP, 4 µl of 24% (w/v) PEG 6000 solution and 10 units of T4 polynucleotide kinase. The reaction was carried out at 37ºC for 30 min. The unincorporated ATP was removed by using the same agarose gel extraction kit (mentioned before) with some modification. At first the reaction volume was made to 100 µl by adding 80 µl of sterile water. Then 100 µl binding buffer was added and the next steps were followed as mentioned above section 3.2.2.4.

3.2.2.7 Transformation of *E. coli* following $CaCl_2$ method

It was essentially carried out by a standard procedure [118]. Briefly, a single colony of *E. coli* cells was grown overnight in LB to saturation. This culture was used to inoculate a fresh culture (10 ml) and grown to OD_{590} ~ 0.5 - 0.6, followed by its centrifugation at 4000 rpm for 10 min to pellet down the cells. The cells were re-suspended gently in 10 ml of chilled 100 mM $CaCl_2$, kept on ice for 30 min and centrifuged at 4000 rpm for 10 min at 4°C. The cell pellet was re-suspended in 5 ml of chilled 100 mM $CaCl_2$, and centrifuged again at 4000 rpm for 10 min at 4°C. The cell pellet was re-suspended in 0.5 ml chilled 100 mM $CaCl_2$ and kept on ice for 16 hr.

For transformation, an aliquot of 100 µl of this cell was mixed with 10 µl of ligation mixture (or ~50 ng of plasmid DNA), incubated on ice for 40 min, and kept for 90 seconds at 42°C. Seven volumes of LB broth was added to it and kept at 37°C for 1 hr. The transformation mixture was plated onto LA plates containing proper antibiotic for selection of transformants and incubated overnight at 37°C for obtaining visible colonies.

3.2.2.8 Genomic DNA extraction from *M. tuberculosis* H37Ra

The genomic DNA from *M. tuberculosis* H37Ra has been extracted as per the protocol mentioned here. *M. tuberculosis* H37Ra culture of volume 25 ml grown in Middlebrook 7H9 broth supplemented with 1% Middlebrook OADC growth supplement was centrifuged at 10,000 rpm for 10 min at 4ºC in two separate centrifuge tubes. The pellets were washed twice with 2 ml of 1X TE buffer (10 mM Tris–Cl, 1 mM EDTA, pH 8.0) after re–suspension and centrifugation at 10000 rpm for 10 min at 4ºC. The obtained pellets were re–suspended in 2 ml 1X TE buffer each and combined in to a single tube to obtain total 4 ml cell suspension after vortexing. To the cell suspension, 150 µl of lysozyme (100 mg/ml) and 6 µl RNaseA (10 mg/ml) were added and the cells were kept for incubation at 37ºC at 150 rpm (gentle mixing) overnight. To the cell lysate, 550 µl of 10X Proteinase K buffer (100 mM Tris–Cl, 50 mM EDTA, 500 mM NaCl, pH 8.0), 40 µl of Proteinase K (20 mg/ml) and 700 µl 10%

SDS were added and incubated at 37ºC at 150 rpm for 2.30 hours. After this, 2 ml of 2.5 M NaCl and 1100 µl of 10% CTAB (pre–warmed to 65ºC) were added and the lysate was incubated for 15 min at 65ºC. To the lysate, equal volume of phenol:chloroform:isoamyl alcohol mixture (25:24:1) was added and incubated for 20 min at room temperature with intermittent gentle mixing. Centrifugation was then carried out at 6000 rpm for 7 min at room temperature and the top aqueous layer was carefully transferred to a fresh tube. To the aqueous solution, equal volume of chloroform:isoamyl alcohol mixture (24:1) was added and mixed gently by inverting the tube. The mixture was then centrifuged at 6000 rpm for 7 min at room temperature and the top aqueous layer was again carefully transferred to another fresh tube. Ice–cold isopropanol was added to the collected solution at final concentration of 70% and incubated at –20ºC for 2 hours. The genomic DNA was precipitated by centrifugation at 12000 rpm for 15 min at 4ºC. The pellet was washed with 1 ml of ice–cold 70% ethanol and the DNA was re–precipitated by centrifugation at 12000 rpm for 15 min at 4ºC. The pellet was air–dried for overnight at room temperature and dissolved in 400 µl 1X TE buffer. The extracted genomic DNA was checked on 0.7% agarose gel to confirm integrity and its concentration was estimated using UV spectrophotometer at 260 nm wavelength.

3.2.2.9 PCR amplification of *mce1R* and the operator DNA

The primers; MRA_0173cForword (5'- TTTCCATGGTGAACGCACCTCTATCGGC, *Nco*I site underlined) and MRA_0173cReverse (5'- TTTCTCGAGGCCAGGGCCTCCGTC, *Xho*I site underlined) were designed to amplify the mce1R ORF by PCR, using the *M. tuberculosis* H37Ra genomic DNA as template, without the stop codon. The operator DNA also has been amplified from the *M. tuberculosis* H37Ra genomic DNA by PCR using the primers Mce1ROpForward (5'-AAAGAATTCGCAGAACTGGGTCAACCAG, *EcoR*I site underlined) and Mce1ROpReverse (5'- AAAAAGCTTGACTCGACGAACTCGGTG, *Hind*III site underlined). The PCR reaction composition and conditions are described briefly. Thirty five cycles of amplification using Phusion Polymerase (1 unit) were performed in a 50 µl reaction volume containing 1X Phusion HF buffer with 1.5 mM $MgCl_2$, 0.2 mM dNTP mix, 30 ng *M. tuberculosis* H37Ra genomic DNA and 200 nM of both primers. Before the beginning of amplification cycle, an initial denaturation step at 98°C for 30s was performed for both *mce1R* and operator DNA amplification. The actual amplification cycle for *mce1R* consisted of denaturation at 98°C for 10s, annealing at 63°C for 10s, extension at 72°C for 30s. To amplify the operator DNA, annealing was done at 61°C for 10s and extension at

72°C for 15s. After the 35 amplification cycles, extra extension at 72°C for were done for 5 min. The PCR fragments were purified using PCR purification kit (Thermo) as per the supplier's protocol. The amplified *mce1R* ORF was digested with *Nco*I and *Xho*I and ligated in pET28a vector at the identical sites to obtain plasmid pAB1014. The amplified operator DNA fragment was also separately cloned in pET28a vector at *Eco*RI and *Hind*III sites similarly. The resultant plasmid was named as pAB1013. The DNA inserts in both of these plasmids were confirmed by DNA sequencing (data not shown).

3.2.2.10 Protein expression and purification technique

From a saturated primary culture of *Escherichia coli* BL21(DE3) cells carrying the plasmid pAB1014, ~1.5 ml has been transferred to a fresh 150 ml of LB medium supplemented with 50 µg/ml kanamycin and incubated at 37°C with constant shaking at 180 rpm till the OD_{600} was 0.5 – 0.6. After that temperature was changed to 20°C and cells were kept with constant shaking for 20 min. Protein was induced with addition of 200 µM IPTG at 20°C with constant shaking for 5 hr. After that cells were harvested by centrifugation at 4°C for 5 min at 7000 rpm. The cell pellet was washed with 0.9% NaCl and stored at –20°C till further use.

The cell pellet was kept on ice for 10 min and re–suspended in 8 ml lysis buffer (25 mM TAPS, pH 9.0, 400 mM NaCl, 5% glycerol). After re–suspension, 0.01% Triton X–100 and 100 µg/ml PMSF were added to the cell suspension and sonicated at 20% amplitude for 5 sec on and 10 sec off for 2 min on ice using Sonics Vibra Cell VCX–500 ultrasonicator. The crude lysate was centrifuged at 12,000 rpm for 10 min at 4°C and the pellet was collected. The pellet was rinsed and re–suspended twice with 5 ml lysis buffer containing 0.01% Triton X–100 and again centrifuged at 12,000 rpm for 10 min at 4°C. The final pellet was re–suspended in lysis buffer containing 0.5% sarcosyl and 0.01% Triton X–100 and centrifuged at 12,000 rpm for 10 min. The supernatant was collected for the affinity purification using Ni–NTA column. The supernatant was added to the Ni–NTA column (1 ml bead volume), pre–equilibrated in the same lysis buffer with 0.5% sarcosyl, and incubated at 4°C for 30 min with gentle mixing for binding of His tagged Mce1R to Nickel column. After that the flowthrough were allowed to pass out of the column. The column was washed with 5 ml of wash buffer I (25 mM TAPS, pH 9.0, 300 mM NaCl, 5% glycerol, 3 mM imidazole) and 3 ml of wash buffer II (25 mM TAPS, pH 9.0, 300 mM NaCl, 5% glycerol, 9 mM imidazole). The bound His–tagged Mce1R was eluted from the column using 5 ml of elution buffer (25 mM TAPS, pH 9.0, 300 mM NaCl, 5% glycerol, 0.1% sarcosyl, 300 mM imidazole). The eluted protein was concentrated to 1 ml volume using centrifugal concentrator (10 KDa cut off) and

EDTA (Ethylene Diamine Tetra Acetic Acid) was added at 40 mM final concentration to chelate the leached Ni^{2+} ion. The protein solution was dialyzed against dialysis buffer (50 mM Na–Phosphate, pH 8.0, 300 mM NaCl, 5% glycerol, 0.05% sarcosyl) at 4°C for overnight and its concentration was measured by Bradford's method using BSA as standard [119].

3.2.2.11 Estimation of protein concentration

Protein was estimated by a standard method [119]. The Bradford dye 5X concentrate at 4˚C was diluted five folds with water before use. A standard curve was prepared with a series of protein samples containing different amounts of BSA. Basically, 1 ml of diluted Bradford reagents were mixed with approximately 10 µl of protein sample and incubated for 5 min. The blue color developed was measured in a spectrophotometer at OD_{595}. The amount of protein in non-BSA sample was estimated from the standard curve.

3.2.2.12 Polyacrylamide gel electrophoresis (PAGE)

Two types of PAGE were performed. These are SDS-PAGE and native PAGE. The standard protocols [118] are mentioned below.

3.2.2.12.1 Tris-glycine SDS-PAGE

The resolving gel part of SDS-PAGE was prepared by mixing required amount of resolving buffer, acrylamide solution as mentioned below. APS solution and the TEMED were added, mixed, and the resultant solution was poured between the glass plates separated by spacers and polymerized below butanol. After allowing sufficient time for the resolving gel to polymerize, the stacking gel mix was prepared with appropriate amount of stacking buffer, acrylamide solution and water and polymerized above the resolving gel with the insertion of a proper slot former. The protein samples were mixed with protein gel-loading dye at final 1X concentration and kept in a boiling water bath for 2 min. After that the samples were loaded in the gel. Gel electrophoresis was carried out in the presence of 1X TGS (Table 3.1) initially at 80 V and then at 100 V using running buffer 1X TGS till the optimal resolution was achieved.

SDS-10%PAGE:

Resolving gel (7 ml):

Water – 2.66ml

30% acrylamide – 2.38ml

Resolving buffer – 1.82ml

10% SDS – 70µl

10% APS – 70µl

TEMED – 5.6µl

Stacking gel (2 ml):

Water – 1.33ml

30% acrylamide – 400µl

Stacking buffer – 250µl

10% SDS – 20µl

10% APS – 20µl

TEMED – 2µl

3.2.2.12.2 Native PAGE

Native PAGE was used mainly to detect DNA-protein interaction by gel shift assay. Usually a 3.2% native PAGE was used for this purpose. The composition of the gel is given below:

For 10 ml gel:

30% Acrylamide solution:	1.06 ml
2X TBE:	2.5 ml
Double distilled water:	6.33 ml
10% APS:	0.1 ml
TEMED:	10 µl
Total:	10.0 ml

3.2.2.13 Staining of polyacrylamide gels with Coomassie brilliant blue

After completion of electrophoresis, the polyacrylamide gel was washed with double distilled water and stained using the staining solution (Table 3.1) at room temperature. After the gel turned blue, it was incubated in the destaining solution till the background was clear and the protein bands were visible. The gel was documented using Gel Doc.

3.2.2.14 Drying of polyacrylamide gels

After the run was complete, the polyacrylamide gel was removed and transferred onto a Whatman 3 mm filter paper. The gel was wrapped with a thin plastic saran wrap. It was then dried using a gel drier (Allied Scientific, India) at 80˚C for 2-3 hr.

3.2.2.15 Autoradiography and developing the X-Ray films

The dried radioactive gel was exposed to X-ray film (Fuji) using intensifying screen. After a suitable time of exposure the film was developed by PRIMER X-ray developer and fixed with the PRIMER fixer solution. The developed films were washed thoroughly and left for air-drying in hanging condition carefully.

3.2.2.16 Gel–shift assay and determination of K_d for Mce1R and operator DNA binding

The gel–shift assay was performed according to the standard protocol [120, 121] with modifications. Briefly, purified Mce1R at varying concentrations were added to several 20µl reaction mixtures containing 1 nM of ^{32}P–labeled operator DNA in buffer A (50 mM Na–Phosphate, pH 8.0, 50 mM NaCl, 5% glycerol, 1 µg/ml BSA) and incubated at room temperature for 30 min. The reaction mixtures were mixed with 4 µl of 6X DNA loading dye without SDS and loaded on a native 3.2% PAGE (prepared in 0.5X TBE). The gel was run in 0.5X TBE at 4°C for ~2.5 hr at constant 80 Volt. After that, the gel was transferred to 3 MM Whatman filter paper and dried in a gel dryer. The dried gel was exposed to X–Ray film for ~20 hr and developed. The resulting image was analyzed by ImageJ 1.52a software (https://imagej.nih.gov/ij/) to quantify the individual band intensities of ^{32}P–labeled operator DNAs. From this fractional DNA binding values were calculated using the following equation

$$y = \frac{i_{bound}}{i_{free}+i_{bound}} \quad (1)$$

where y, i_{bound} and i_{free} indicate fractional DNA binding, band intensities of protein–DNA complex and free DNA respectively. The values of y were plotted against respective Mce1R molar concentrations. The resulting plot was fitted using the following Hill equation in OriginPro 8 software (OriginLab Corporation, MA)

$$y = V_{max}\frac{x^n}{k^n+x^n} \quad (2)$$

where y, V_{max}, x, k and n indicate fractional DNA binding, maximum DNA binding, Mce1R molar concentration, equilibrium dissociation constant (k_d) and Hill coefficient respectively.

The competition gel–shift assay has been performed according to the standard protocol [120] with modifications. Briefly, gel–shift assay was performed using radiolabeled operator DNA premixed with either 10 fold or 25 fold molar excess unlabeled operator DNA and incubated for 30 min. As nonspecific competitor, 100 ng of poly dI–dC [122] was added to the radiolabeled operator DNA and gel–shift assay was performed as mentioned above. All the reaction mixtures were resolved on a non–denaturing 3.2% PAGE, and subjected to autoradiogram.

Gel–shift assay using heat–inactivated Mce1R also has been performed in addition to the competition gel–shift assay to further confirm binding specificity as described earlier [112, 123] with modification. Briefly, Mce1R was heated for 10 min at 95°C in a PCR machine and allowed to cool to room temperature. After that the radiolabeled operator DNA has been mixed. As positive control, active Mce1R also used in a separate reaction and gel–shift assay has been performed as mentioned above.

3.2.2.17 Western blotting

To check the purity of His–tagged Mce1R, SDS–10%PAGE was used and for western blotting, the gel was blotted on a PVDF membrane activated before blotting by submerging in methanol for a few seconds. The detailed protocol is mentioned below:-

1. After the transfer, the membrane was treated with 3% BSA for 1 h at room temperature.
2. To remove unbound BSA, the membrane was washed single time with TBST (10 min) and two times with TBS (10 min) buffers sequentially.
3. After washing, the membrane was incubated with anti-his mouse primary antibody (1: 5000 dilution) for 2hr at room temperature followed by two wash each of TBST and TBS for 10 min.
4. Then, the membrane was incubated with alkaline phosphatase-tagged goat anti-mouse secondary antibody IgG1-AP (1: 10000 dilution) for 1 hr at room temperature followed by its washing with TBS-T 10 min and two times with TBS.
6. Finally, the membrane was treated with NBT/BCIP in dark till the brown colored bands appeared. The membrane then was washed in sterile pure water and kept in it.

3.2.2.18 Intrinsic tryptophan fluorescence quenching analysis

Quenching analysis of intrinsic tryptophan fluorescence of Mce1R by acrylamide was performed according to standard method [127] with modification. Briefly, increasing concentrations (0–0.9 M) of acrylamide were added to final fixed concentrations (5 µM) of

Mce1R solution of final 200 µl volume at room temperature in buffer A without BSA. The acrylamide stock solution was prepared in the same buffer A to avoid dilution of buffer components during preparation of the protein–acrylamide mixtures. The mixtures were incubated for 2 hr at room temperature and subjected to tryptophan fluorescence measurement in a fluorescence spectrophotometer (Horiba Fluorolog 3-21, USA). The steady–state fluorescence spectra were recorded between 310–400 nm by exciting the sample at 295 nm with 5 nm band–pass on both sides. The quartz cuvette had optical pathlengths 1 and 0.4 cm on emission and excitation sides respectively. The recorded spectra were corrected by deducting corresponding buffer spectra and adjusting inner filter effect as described [128]. As Mce1R contains more than one tryptophan residues, the quenching data were analyzed according to Lehrer relation [129] assuming equal quenching constants for all fluorophores (tryptophan residues) to determine fractional accessibility of the tryptophan residues. The Lehrer relation is mentioned below:

$$\frac{F_0}{\Delta F} = \frac{1}{f_a K[Q]} + \frac{1}{f_a} \quad (3)$$

Where, F_0 is fluorescence intensity without quenching, $\Delta F=F_0–F$ is the decrease in fluorescence intensity after addition of quencher at concentration [Q], f_a is the maximum fraction of all accessible fluorophores, K is the quenching constant. The plot $F_0/\Delta F$ versus 1/[Q] yields a straight line which can be extrapolated at 1/[Q]=0 to determine the value of $1/f_a$ from which f_a can be determined. The linear fitting has been performed using the OriginPro 8 software (OriginLab Corporation, MA).

3.2.2.19 Gel–filtration chromatography

Gel–filtration chromatography of Mce1R was performed in an HP 1200 (Agilent Technologies India Pvt. Ltd.) equipment at room temperature in buffer A without BSA following the procedure described by Ganguly et al., 2007 [120]. As standards, BSA (~66 kDa), carbonic anhydrase (~25 kDa) and lysozyme (~14 kDa) were run.

3.2.3 Bioinformatic methods

3.2.3.1 Generation of three–dimensional structure

Protein sequence of Mce1R (Rv0165c) (GenBank accession no: CCP42891.1) was retrieved from Protein database available in NCBI (https://www.ncbi.nlm.nih.gov/protein). Because of the lack of suitable homologous protein of sequence identity and query coverage with experimentally determined structure in the PDB database (https://www.rcsb.org/), the

sequence of Mce1R was used to generate the three–dimensional structure of Mce1R using the Phyre2 fold recognition server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) [124]. The server was set to run in "intensive" mode as this mode generates complete structure of the protein that could be downloaded as PDB file which helped in further analysis.

3.2.3.2 Refinement and validation of the predicted three–dimensional structure

The generated structure of Mce1R was further refined using the Swiss PDB Viewer tool (V4.1.0) [125] to lower the internal energy and to increase the stability of the structure. The energy minimization was performed using 1200 steps of steepest descent of the Swiss PDB Viewer. The energy minimized structure was then saved as PDB file and validated using the SAVES (V6.0) (https://saves.mbi.ucla.edu/) and the ProSA servers [126].

3.2.3.3 Coarse grain molecular dynamics simulation of Mce1R by CABS–Flex server

CABS–flex 2.0 server [130] was used in the simulation process, an updated version of CABS–flex 1.0 server which provides almost native structural dynamics of proteins in a 10 ns simulation time, using all atom explicit water in the four force fields – amber, grooms, OPLS and CHARMM [131]. During the simulation, samplings of protein conformations were done following Monte Carlo method [132]. The server returns 10 models of different conformations in 1 ns intervals up to total 10 ns simulation time. To perform 100 ns simulation, the server was run total 10 times using the last conformation of Mce1R after 10 ns simulation time as input for the next 10 ns simulation. All sampled conformations (100 conformations for 100 ns simulation time) were superimposed using PyMOL v2.4.0. The server calculates the residue–wise fluctuations as RMSF (Root Mean Square Fluctuation) values and plots them after global superimposing of the sampled structures. The RMSF values after each 10 ns simulation were collected and 100 ns RMSF values were generated taking the highest residue–wise RMSF values among 10 such simulations of 10 ns times (Appendix 6.1).

3.2.3.4 Prediction and validation of the dimeric structure of Mce1R

The structure of Mce1R was submitted to the GalaxyHomomer server (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=HOMOMER) [133] which generated 5 different dimeric structures of Mce1R. The structure with highest TM score has been selected and energy minimization has been performed using the Swiss PDB Viewer tool (V4.1.0)

[125] and the resulting structure was saved in pdb format. Validation of the structure has been performed using the SAVES (V6.0) (https://saves.mbi.ucla.edu/) and ProSA [126] servers respectively as described previously in section 3.2.8. To analyze non–covalent interactions between the two subunits in the Mce1R dimer, the PIC webserver (http://pic.mbu.iisc.ernet.in/job.html) [134] was used.

3.2.3.5 Alignment of homologous structures

To identify the homologous structures to Mce1R, DaliLite (V5.0) server [135] was run using the Mce1R model structure as query. To select the strong matches, the cut off for the Z score was set at 18.3 following the function $\frac{n}{10} - 4$ [136], where n is the number of amino acid residues of the query structure. As the numbers of structures were low, the cut off score was further reduced to 17.3 and the structures were selected for structure guided sequence alignment using the PROMALS3D server with default settings [137].

3.2.3.6 Multiple sequence alignment

The alignment was performed using the MEGA–X version 10.1.8. [138]. PSI–BLAST was performed up to 5 iterations using the Mce1R sequence as query sequence. Those sequences showing more than 80% query coverage and more than 40% similarity were selected as orthologs. From each genus maximum 4 sequences were taken and aligned using the MUSCLE algorithm.

3.2.3.7 Cavity identification

To detect the cavity in the C–terminal domain, the Mce1R structure was submitted to the Cavityplus web server [139]. The server returns the detected cavity with the residue names constituting it.

3.2.3.8 Search for cavities similar to that of Mce1R

The validated structure of Mce1R, obtained as mentioned in section 3.2.8, was used as input to search for cavities similar to that of Mce1R using the ProBis server (http://probis.cmm.ki.si/) [140] (Job id: 12032171021175). The cavity of Mce1R has been specified using the residue numbers of the residues forming the cavity of Mce1R during the search using ProBis. In addition to identification of protein structures carrying cavities similar to that of Mce1R, the server also identifies possible ligands; those may bind with

Mce1R specifically or non–specifically. The ProBis server arranged the results in a descending order depending on the corresponding confidence scores.

3.2.3.9 Molecular docking analysis

Initially, protein–ligand complex (for example; the protein–ligand complex; PDB id 2NNJ) was downloaded from the PDB and protein structure was extracted. The extracted protein structure was used as the receptor molecule and the respective ligand was again re–docked at the target cavity site of the protein using the AutoDock Vina software [141] to validate the docking protocol employed in this work. The docking protocol mentioned below had successfully re–docked the ligand to its specific cavity of the protein, suggesting the validation of the docking protocol (data not shown).

For performing the docking on the dimeric Mce1R structure, at first the ideal structure data files for the ligands to be tested (myristic, palmitic, oleic and stearic acids) were downloaded from the PDB website as sdf files. The sdf files were then converted to pdb format using the PyMOL v 2.5.2 software. Using the AutoDock Tools, the polar hydrogen atoms and Kolman charges were added to the Mce1R dimeric structure (pdb file). The resulting structure was saved as pdbqt format to be used in the docking analysis using the AutoDock Vina software [141]. Gasteiger charges and torsion roots were added to the ligands automatically using the AutoDock tools and saved as pdbqt formats. For docking analysis, at first, the cavity in the chain A of Mce1R was chosen as the target site. The grid parameters for the same were as follows: X_center = 37.339, Y_center = 45.356 and Z_center = 48.89. The best docked poses of the ligands were chosen and the docked complexes were converted to respective pdb files using the PyMOL v 2.5.2 software. The docked complexes were again converted to pdbqt files using the AutoDock Tools following the procedures mentioned above and again used as the receptors for docking with respective ligands targeting the cavity site present in the chain B of Mce1R. The grid parameters for the same were as follows: X_center = 10.884, Y_center = 32.751, Z_center = 36.107. The exhaustiveness for performing docking at both the cavity sites was kept at 40. The grid box dimensions (X_size, Y_size, Z_size) were kept as (30, 30, 30). Again the best docked poses of the ligands were chosen and the final docked complexes were converted to pdb files using the PyMOL v2.5.2 software. The resulting docked complexes thus contained two ligands docked within the two cavities of Mce1R. To analyze various non–bonded interactions between the ligands and receptor, the LigPlot^{+} v2.2 software [142] was used.

3.2.3.10 All atom molecular dynamics simulation of Mce1R and Mce1R–fatty acid complexes using GROMACS

The GROMACS 2021.2 software package was used to perform molecular dynamics simulation of Mce1R and Mce1R–fatty acids complexes to analyze conformational flexibilities and stabilities of the protein and protein–ligand complexes. From the docked complexes, the ligands were extracted and hydrogen atoms were added using the PyMOL v2.5.2 software. The parameter and force field files of the ligands were then generated using the Acpype server [143] using the in–built AMBER forcefield. For performing the molecular dynamics simulation of Mce1R and Mce1R–fatty acids complexes, the AMBER ff99SB–ILDN forcefield [144] was used. The protein and protein–ligand complexes were kept at the center of a dodecahedral simulation box with 1Å buffering distance from the edge of the box surrounding the protein. The systems were solvated using the TIP3P simple point charge explicit water molecules and neutralized with 150 mM of Na^+ and Cl^- ions. Energy minimization of the systems were performed without any position restraining on water molecules using steepest descent algorithm until the force threshold value reached ≤1000 kJ/mol/nm for the systems. Equilibration of the systems was performed at first at constant volume (NVT) at 300K using the modified Berendsen thermostat for 1 ns. After that the systems were equilibrated at constant pressure of 1 bar and constant temperature (NPT) for 1 ns using the Parrinello–Rahman barostat. Two coupling groups, protein–ligand(s) and water and ions, with coupling times 0.1 ps and 2 ps (for NVT and NPT respectively) were set. Lincs constraint algorithm was employed to fix the bond lengths. The long range electrostatics was treated using Particle Mesh Ewald method (PME). The V–rescale method was used to control the temperature of the system. For non–bonded interactions, the cut–off value were kept at 1 nm. The final MD runs were performed for 200 ns without any restraint with 2 fs timestep. During the simulation, the coordinates of the intermediate structures were saved every 10 ps.

After the simulation, the trajectories of the systems were corrected using the *trjconv* tool available in GROMACS and the periodic boundary conditions were removed. The corrected trajectories were used to analyze various dynamic properties of the systems, such as– root mean square deviation (RMSD), root mean square fluctuation (RMSF), average radius of gyration (R_{gavg}) and secondary structure analysis using in–built scripts of GROMACS. The hydrogen bond analysis between the protein and ligand was performed using the VMD software [145].

3.2.3.11 Determination of binding free energy for Mce1R–fatty acid ligands

The binding free energies of all Mce1R–fatty acid complexes have been computed following the MMGBSA (Molecular Mechanics Generalized Born Surface Area) approach using the gmx_MMPBSA tool [146] that uses the MMPBSA.py python script [147] of AmberTools21. From the simulation trajectories, snapshots were taken at 1.25 ns interval from 150 ns to 200 ns timepoints. Therefore total 41 frames were taken and the average energy values were computed form those frames.

CHAPTER 4

RESULTS

AND

DISCUSSION

Section 1

4.1 Cloning, expression, purification of Mce1R and analyzing its DNA binding activity

4.1.1 Objective A.1- Extraction of genomic DNA from *M. tuberculosis* H37Ra

The genomic DNA from *M. tuberculosis* H37Ra has been extracted following the protocol mentioned in the section 3.2.2.8. After dissolving the pellet in 400 µl of 1XTE buffer, the absorbance (A_{260}) was recorded using a spectrophotometer at 260 nm wavelength taking 5 µl DNA in 100 µl final volume of 1XTE buffer after deducting the absorbance value of 1XTE buffer at the same wavelength. A suitable UV–transparent quartz cuvette was used for this purpose. The absorbance values were recorded three times and the average was determined from them. The absorbance values are shown in the below Table 4.1.

Table 4.1 Absorbance values of the genomic DNA sample at 260 nm wavelength

Sl. No.	A_{260}
1.	0.15
2.	0.18
3.	0.13
Avarage	0.153

The average A_{260} value 0.153 corresponds to 7.65 ng/µl of DNA concentration which is 20 fold diluted (5 µl of DNA from the original stock to final volume 100 µl). Therefore, the stock concentration of the isolated genomic DNA is ~153 ng/µl.

The DNA concentration obtained from the spectrophotometer measurement is adequate for the subsequent steps to meet the next objectives, however, the integrity of the extracted DNA is also a major requirement for the PCR process. Therefore, the integrity of the genomic DNA was checked by loading 2 µl sample on a 0.7% agarose gel alongside the DNA marker. Fig. 4.1.1 shows the presence of single intense band in the lane 1 migrated less compared to that of the 10 kBp band of the DNA marker. The size of the genomic DNA of *M. tuberculosis* H37Ra is ~4.4 mBp [148]. The band position of the genomic DNA with respect to the DNA markers in the Fig. 4.1.1 is also very similar with the published results for genomic DNA of high molecular size [149]. Moreover presence of the single band suggests that integrity of the genomic DNA is maintained and the DNA is not sheared during the extraction process.

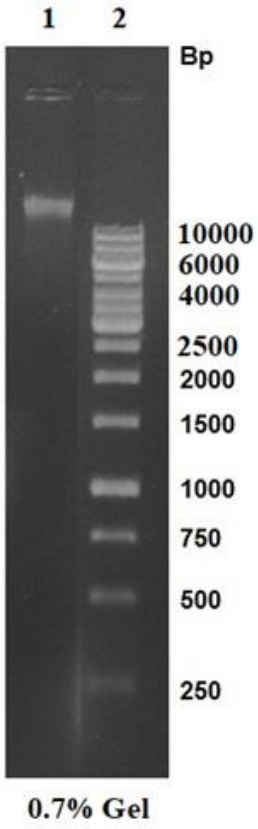


Fig. 4.1.1 Extracted genomic DNA of *M. tuberculosis* H37Ra. Lane 1: Extracted genomic DNA is loaded; Lane 2: 1 kBp DNA marker. A single band appeared in the Lane 1. The base pairs of the DNA marker bands are shown next to the respective bands. The DNA was resolved on 0.7% agarose gel.

4.1.2 Objective A.2- PCR amplification of *mce1R* gene and cloning in the expression vector pET28a

Objective A.3- PCR amplification and cloning of *mce1* promoter (operator) DNA

The entire DNA region of *M. tuberculosis* H37Rv, from *mce1R* to the last gene of *mce1* operon (*rv0165c* to *rv0178*; nucleotide number 194144 to 209672), is identical to the corresponding DNA region of *M. tuberculosis* H37Ra (*MRA_0173* to *MRA_0186*; nucleotide number 195504 to 211032) [148]. Previously, Casali et al. had shown that the start codon GTG, for *mce1R* translation was located at nucleotide number 194815 in *M. tuberculosis* H37Rv [104] that corresponds to the nucleotide number 196175 in *M. tuberculosis* H37Ra. Accordingly the primers were designed to amplify the region from nucleotide number 196175 to 195507 (669 bp) by PCR, using the *M. tuberculosis* H37Ra genomic DNA as template, without the stop codon (Fig. 4.1.2a). The operator DNA also has been amplified from the region 196427 to 196627 (201 bp) in *M. tuberculosis* H37Ra genomic DNA (corresponding region 195067 to 195267 in *M. tuberculosis* H37Rv genomic DNA, containing the putative operator site [112]) by PCR using the appropriate primers (Fig. 4.1.2b). The PCR fragments were purified using PCR purification kit (Thermo) as per the supplier's protocol. The DNA fragment carrying the *mce1R* ORF was cloned in pET28a vector at the *Nco*I and *Xho*I sites to transcriptionally fuse C–terminal His–tag with Mce1R. The recombinant plasmid carrying the *mce1R* ORF was named as pAB1014. The operator DNA fragment of 201 bp was also separately cloned in pET28a vector at *EcoR*I and *Hind*III sites. The resultant plasmid was named as pAB1013. Presence of the correct DNA inserts were confirmed by digestion of the plasmids with respective restriction enzymes (Fig. 4.1.2c) and DNA sequencing (data not shown).

4.1.3 Objective A4- Expression and purification of C–terminal His–tagged Mce1R by Ni^{2+} affinity chromatography

To confirm optimum production of Mce1R under the laboratory condition, whole cell lysate sample from the IPTG induced culture was resolved on a 13.5% SDS–PAGE and stained by coomassie brilliant blue. As negative control, crude lysate sample from the culture without IPTG was also resolved alongside the induced lysate. Fig. 4.1.3a shows the presence of an over–induced protein band in the induced lysate of more than 21 kDa molecular weight in lane 2. The theoretical molecular weight of His–tagged Mce1R is ~25.23 kDa according to the ProtParam analysis tool [150]. To confirm whether the observed over–induced band specific to C–terminal His–tagged Mce1R, western blotting was performed

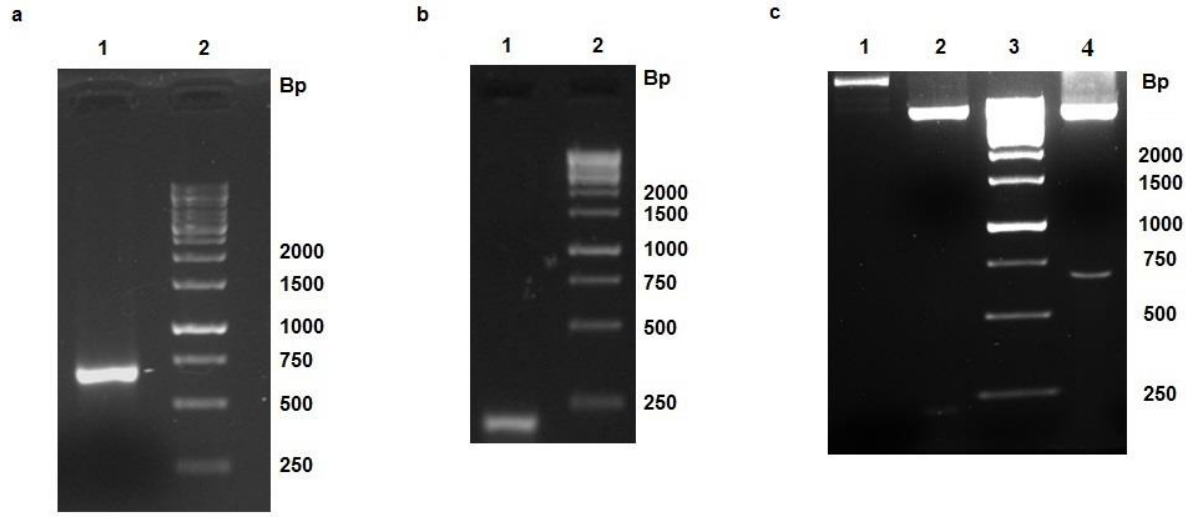


Fig. 4.1.2 PCR amplification and confirmation of the cloning of *mce1R* gene and promoter DNA (operator). (a) Lane 1: PCR amplified *mce1R* ORF; Lane 2: 1 kBp DNA marker. (b) Lane 1: PCR amplified operator DNA; Lane 2: 1 kBp DNA marker. (c) Lane 1: Undigested pET28a; Lane 2: *EcoR*I and *Hind*III digested plasmid pAB1013; Lane 3: 1 kBp DNA marker; Lane 4: *Nco*I and *Xho*I digested plasmid pAB1014. Bp indicates the sizes of the DNA fragments in the 1 kBp DNA marker.

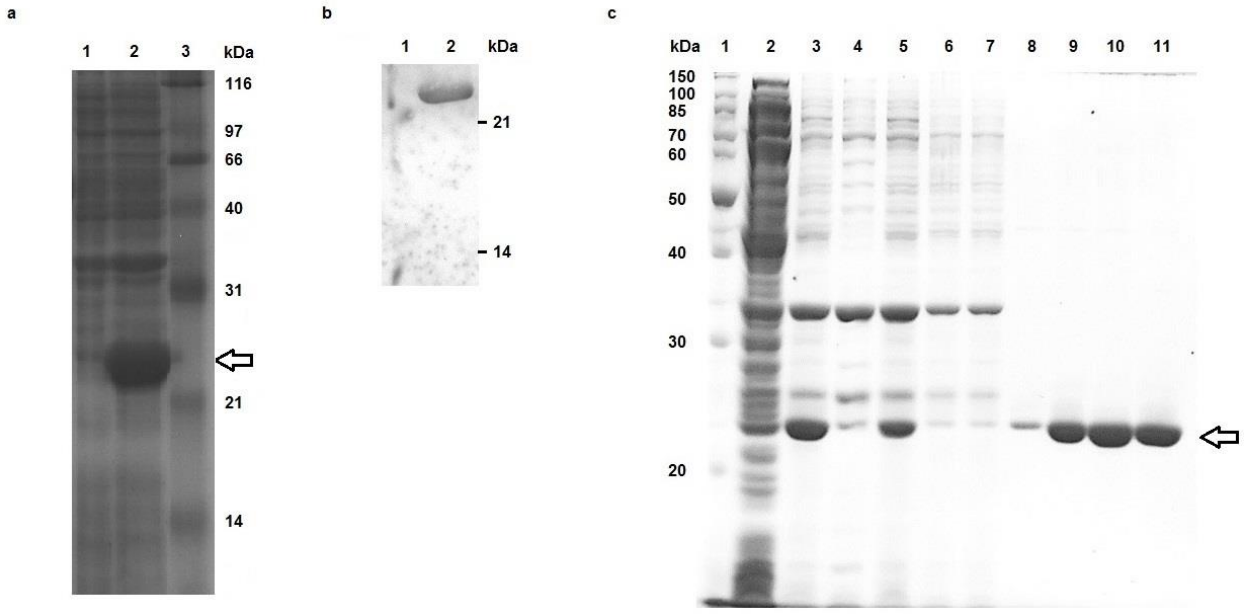


Fig. 4.1.3 Expression and purification of C–terminal His–tagged Mce1R. (a) Lanes 1 and 2: Protein extracts from *E. coli* BL21(DE3) cells without and with IPTG respectively; Lane 3: Protein molecular weight marker. (b) Western blot of protein extracts from *E. coli* BL21(DE3) cells without and with IPTG in lanes 1 and 2 respectively. (c) Lane 1: Protein molecular weight marker; Lanes 2 and 3: Supernatant and pellet fractions after sonication in lysis buffer without sarkosyl respectively; Lanes 4 and 5: Pellet and supernatant fractions after dissolving the inclusion bodies in lysis buffer with sarkosyl; Lane 6: Flow-through from the Ni^{2+}–NTA column; Lanes 7 and 8: Wash fractions from the column; Lanes 9, 10 and 11: Elution fractions from the column. Sizes of protein bands (kDa) in the protein molecular weight markers are indicated adjacent to the corresponding bands. The arrows indicate Mce1R–specific bands.

after SDS–PAGE using mouse anti–His primary antibody. As shown in Fig. 4.1.3b, the same over–induced band was detected in induced lysate (lane 2) but no band appeared in the un–induced lysate (lane 1). Taken together the data indicate that under the experimental conditions, the C–terminal His–tagged Mce1R is strongly expressed.

The whole cell lysate from the over–induced culture cell pellet was centrifuged to separate pellet and supernatant fractions. SDS–PAGE analysis revealed that a major fraction of expressed Mce1R is present in the pellet fraction as inclusion bodies while very less amount is present in soluble fraction (Fig. 4.1.3c; lanes 2 and 3). The pellet was re–dissolved in lysis buffer containing sarkosyl and re–clarified by centrifugation. The resultant supernatant was allowed to incubate with Ni^{2+}–NTA resin and the flow–through from the column (Fig. 4.1.3c; lane 6) was found to carry very less amount of Mce1R, indicating that most of the Mce1R interacted and bound with the Ni^{2+}–NTA resin. Lanes 7 and 8 show the profiles of two successive washes of the column. It was observed that a small fraction of bound Mce1R was lost during washing steps along with other contaminating proteins. Lanes 9, 10 and 11 show the elution fractions from the resin which indicates that Mce1R is purified to homogeneity without any contamination. The protein was dialyzed to remove excess sarkosyl and its concentration was determined by Bradford's method [119]. This highly purified Mce1R was used in all other subsequent studies.

Adding a His–tag to proteins not only makes the purification process quicker and easier, but also many times helps in expression and folding of the produced proteins [151]. To minimize the addition of extra amino acids, Mce1R was purified as C–terminal His–tagged form which carries only two additional amino acids (Leu and Glu) with the six histidine residues (His–tag). As shown in the Fig. 4.1.3a, the recombinant Mce1R was strongly expressed under the conditions used in the laboratory. Majority of the expressed Mce1R formed inclusion bodies inside the cell, consistent to the observation that nearly 70% of recombinant *M. tuberculosis* proteins form inclusion bodies in the *E. coli* host [152]. As changing the induction conditions such as induction temperature, IPTG concentration and induction time did not decrease the amount of Mce1R inclusion bodies (data not shown), protein purification has been performed from these inclusion bodies since these contain intact and pure recombinant proteins [153] thereby greatly reducing the possible contaminations that might occur from other cellular proteins.

4.1.4 Objective A5- Estimating operator DNA binding affinity of Mce1R by gel–shift assay

To measure the activity of purified Mce1R, *in–vitro* DNA binding experiment (gel–shift assay) was performed using ^{32}P–labeled operator DNA and Mce1R at room temperature. After incubating the DNA–protein mixtures for 30 min, the reaction mixtures were resolved on a non–denaturing 3.2% PAGE. The autoradiogram shown in the Fig. 4.1.4a clearly indicates the formation of Mce1R–operator DNA complexes as shifted DNA bands in presence of 0.4–2.8 µM concentrations of Mce1R. Operator DNA binding increases with concomitant increase in Mce1R concentration. As negative control, a reaction mixture was prepared similarly without the addition of Mce1R where no shifted band was observed indicating that formation of protein–DNA complexes is attributed to Mce1R–specific DNA–binding function. The concentrations of ^{32}P–labeled operator DNA in all the reactions were maintained at 1 nM. The intensities of the bands were measured using ImageJ 1.52a software after subtracting the background intensity and fractional DNA binding is determined using the equation (1) as mentioned in the "Materials and Methods" section 3.2.2.16. The gel–shift assay experiment was performed thrice using identical Mce1R and operator DNA concentrations and one representative autoradiogram is shown in the Fig. 4.1.4a. The average fractional DNA binding values obtained from three independent experiments were plotted against corresponding Mce1R concentrations and curve fitting was performed using the Hill equation (equation (2) in "Materials and Methods" section 3.2.2.16) (Fig. 4.1.4c). Error bars are indicative of standard deviation. The R–square value was ~ 0.99 for the fitting curve indicating a good fit. The K_d value obtained from the curve was 0.35±0.02 µM and the Hill coefficient (n) was determined to be 1.04±0.13 which is very close to 1. This suggests that Mce1R possibly binds to the operator DNA without cooperativity.

To check whether observed Mce1R-operator DNA interaction is specific, competition gel–shift assay was performed. Fig. 4.1.4b shows the result of the experiment. The intensity of the protein–DNA complex has been decreased to nearly 66% (analyzed using ImageJ 1.52a software after subtracting the background intensity) in presence of 10 fold molar excess unlabeled operator DNA which further decreases to ~26.7% compared to the DNA–protein complex band intensity without any competition when 25 fold molar excess unlabeled operator DNA was added. Unlike specific competitor, Mce1R–operator DNA complex did not dissociate much in presence of non–specific competitor poly dI–dC.

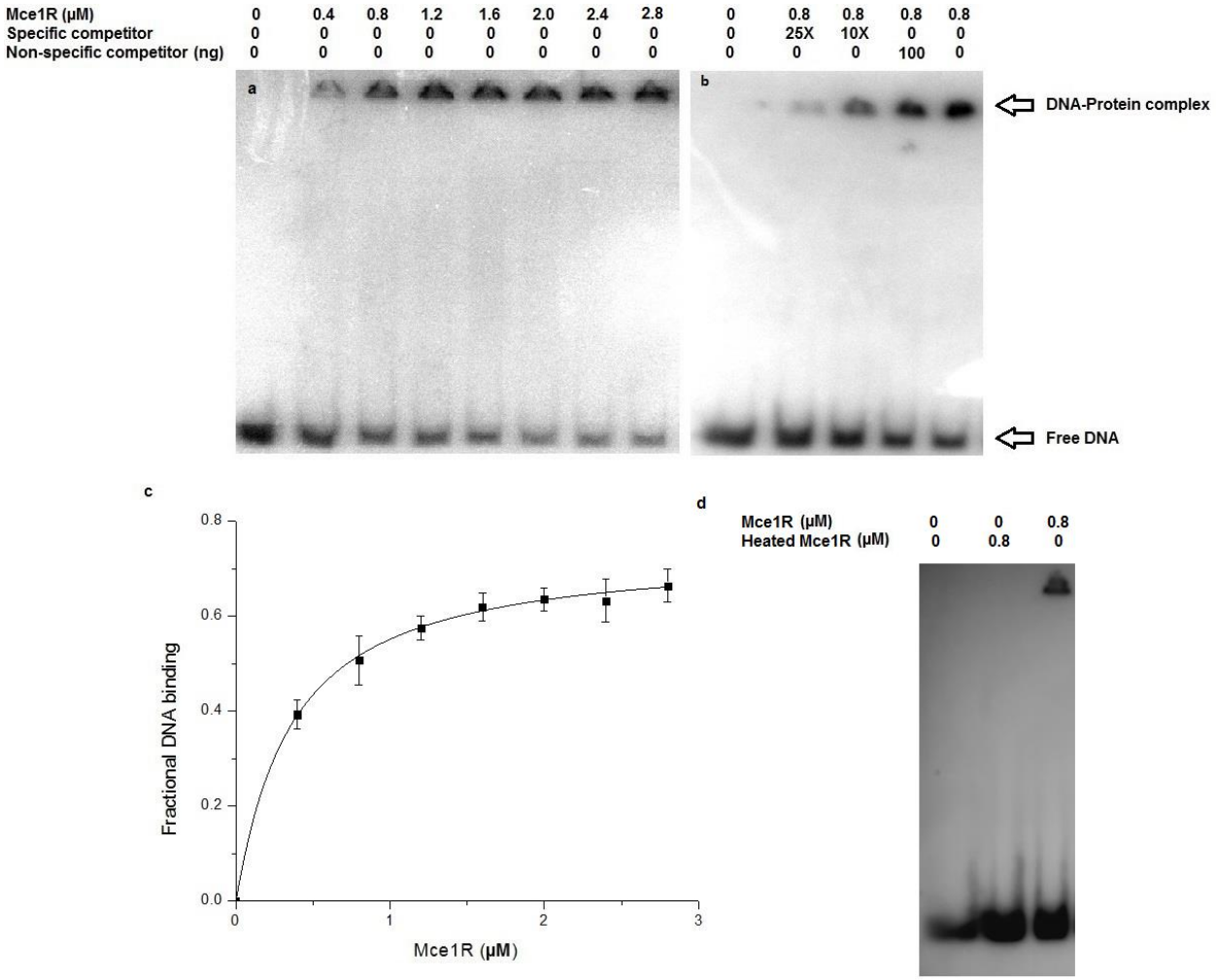


Fig. 4.1.4 Gel-shift assay of Mce1R–operator DNA interaction. (a) Equilibrium binding of Mce1R to the operator DNA. Molar concentrations of Mce1R (0–2.8 μM) used in each lanes are shown in the figure. No specific or non–specific competitor DNA was used in this experiment. (b) Binding specificity of Mce1R to its operator DNA. Mce1R concentration (0.8 μM) was kept constant. 10X and 25X indicate 10 and 25 fold molar excess of unlabeled operator DNA (specific competitor) in the respective reactions. 100 ng of Poly dI–dC used as non–specific competitor [122]. (c) The plot of fractional DNA binding vs Mce1R concentrations obtained from the equilibrium binding of Mce1R and operator DNA. (d) Gel–shift assay using heat–inactivated Mce1R. Molar concentrations of heat–inactivated and active Mce1R are shown at the top of the picture. All these experiments were performed thrice and the representative pictures are shown here. Error bars indicate standard deviations of three independent experiments.

Further, to confirm the observed shifted band caused due to the activity of Mce1R, gel–shift assay has been performed using heat–inactivated Mce1R as a negative control. The result from the autoradiogram (Fig. 4.1.4d) showed that heat–inactivated Mce1R failed to bind the operator DNA which strongly supports that observed shift of the operator DNA bands in the above experiments caused due to the activity of Mce1R only. Therefore, the results of the competition assay and the control experiment confirm that Mce1R specifically binds with its cognate operator DNA.

The K_d value for Mce1R–operator DNA interaction is 0.35±0.02 µM. Compared to other FadR–type transcriptional regulators having K_d values as low as 1 nM [154] to ~6 µM (approximation made by us from the published EMSA gel picture Fig. 3b; not determined by the authors) [155], Mce1R tends to possess moderate affinity to its operator DNA. Detailed analysis on DNA binding activity of extensively characterized FCD family of regulators, FadR of *E. coli*, has revealed a DNA binding mode distinct than most of the regulators bearing wHTH motifs with highly similar topologies of the N–terminal DNA binding domains. *E. coli* FadR uses specific amino acid residues of the N–terminal region of the wHTH motif rather than using the recognition helix (helix α3) to specifically bind cognate DNA [156]. The amino acid residues of FadR critical for making specific DNA contacts are Arg35, Arg45, Arg49 and His65. Hydrogen bonding interactions with bases of DNA, mediated by these residues are important for binding specificity and affinity. Interestingly, amino acid residues corresponding to that of the above mentioned amino acids of FadR, for a number of well characterized FCD family of regulators (Tm0439 of *T. maritima*; McbR of *E. coli*; Cgl2915 of *C. glutamicum*) with wHTH motifs of highly similar topologies, are also reported to play similar roles in making specific interactions to DNA [157–159], suggesting that positions and types of amino acid residues are important for making stable DNA–protein complexes. Notably, significant sequence similarity has been observed among Tm0439, McbR and Mce1R particularly at the wHTH motifs (can be observed in Fig. 4.2.4), implying that Mce1R might interact with DNA in a way similar to that of Tm0439, McbR, Cgl2915 or FadR because of highly similar topology of the wHTH motif. However, sequence comparison between the wHTH motifs of FadR and that of Mce1R revealed that the amino acid residues responsible for making specific contacts with DNA in FadR (Arg35, Arg45, Arg49 and His65) were not conserved in Mce1R (except Arg49 position) and replaced by Asp46, Val56, Arg60 and Pro76 respectively [104] which might be the reason for the reduced affinity of Mce1R compared to FadR as residues like valine and proline do not form hydrogen bonding

interactions [160]. It is possible that Val56 may stabilize the wHTH fold through hydrophobic interaction with the hydrophobic residues in the 'turn' region of the wHTH motif [161] instead of participating directly in DNA–binding. Pro76 may interact through weak ring–stacking interaction with DNA bases [162]. Consistent to this, in–vivo overexpression of Mce1R caused only 2–3 fold repression of the promoter activity of mce1R [112] thus supporting our observation that interaction between Mce1R and the operator DNA is not very strong. A different study on several regulons associated with divergently transcribed FadR–type regulators also supports the fact that autoregulation of FadR–type regulators are weaker than the regulation of the corresponding regulons [160]. This is in agreement with the essentiality to maintain an intracellular concentration of Mce1R for stringent regulation of *mce1* operon consistent with the observation that *mce1* operon remains repressed up to 4 weeks post–infection [117]. Although the recognition sequence of Mce1R (operator DNA) remains to be identified, presence of different amino acid residues in the wHTH motif of Mce1R suggests that the recognition sequence for Mce1R might be distinct than that of FadR.

Section II

4.2 Structural modeling, validation, dynamics and sequence analysis of Mce1R

4.2.1 Objective B1- Modeling of monomeric structure of Mce1R by Phyre2 server and its validation

A protein's function is invariably dependent on its proper three–dimensional structure. Alterations in the primary structures of proteins; even by a single amino acid residues many times cause significant changes in structures, functions and stabilities of proteins [169, 170]. Despite being categorized as VanR–type regulator based on secondary structure profile [110], Mce1R displays little similarity at the primary structure level compared to many VanR–type regulators with known three–dimensional structures excluding the wHTH motif. To confirm whether high level of dissimilarities at the primary structure level has made any significant deviation in the three–dimensional structure of Mce1R from the canonical VanR–type structures or not, it was sought to model the structure of Mce1R computationally. The protein sequence of Mce1R was retrieved from the Protein database of NCBI and searched against PDB database for identifying orthologs by PSI–BLAST analysis to use as templates to construct structure through homology modeling. As there was lack of suitable orthologs of significant sequence similarity and query coverage of known structures, Mce1R three-dimensional structure was generated using the Phyre2 fold recognition server [124]. The result showed that more than 96% of total residues have been modeled with 100% confidence score. The generated structure was subjected to energy minimization using 1200 steps of steepest descent available in the Swiss PDB Viewer to remove any high energy configuration and to increase stability of the modeled structure. The computed internal energy of the model was –2873.882 kJ/mol and –13120.676 kJ/mol on before and after energy minimization respectively.

The energy minimized structure was then validated using the SAVES and ProSA [126] servers. The SAVES server simultaneously runs a number of programs out of which Verify 3D [163], ERRAT [164] and PROCHECK [165] programs are most significant to evaluate the structure. The Ramachandran plot (Fig. 4.2.1a) generated by the PROCHECK program showed that ~84.3% residues were in most favored regions, 13.6% residues were in additional allowed regions and 1% residues were in generously allowed regions. Only 1% residues (2 residues) were in Ramachandran outliers (disallowed region).

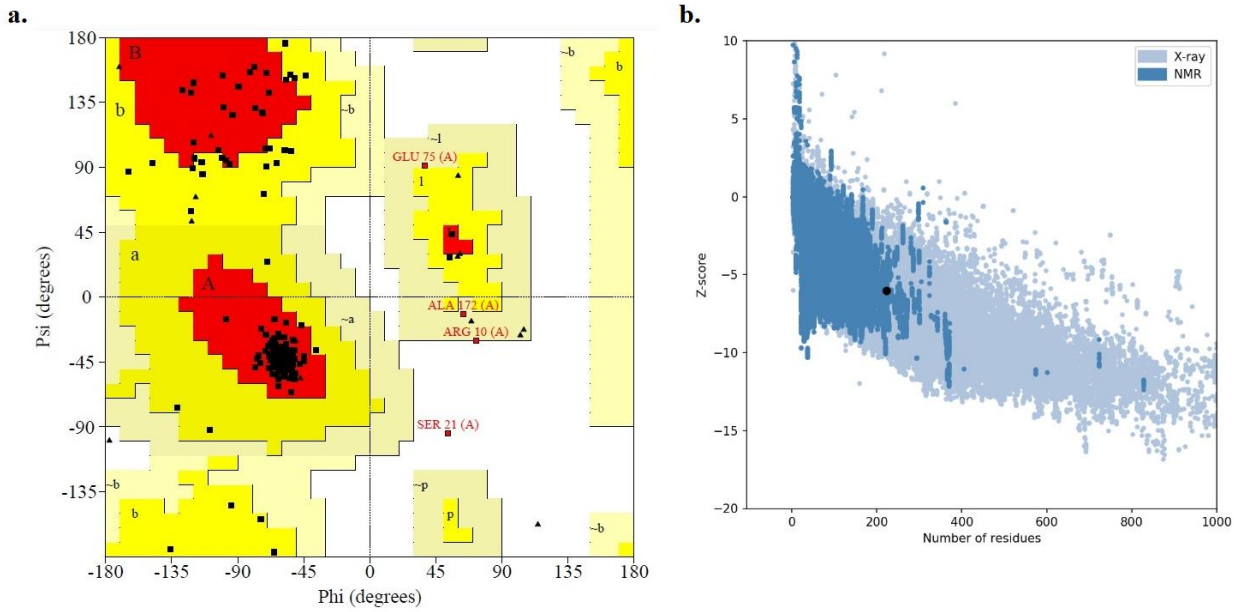


Fig. 4.2.1 Computational validation of Mce1R monomeric structure. (a) Ramachandran plot analysis by the PROCCHECK server. The dark–filled triangles (▲) indicate glycine residues. (b) Z score plot determined by the ProSA server. The black dot is showing the location of the modeled structure.

The overall G factor computed by PROCHECK was –0.05, well within the acceptable range <–0.5 [166], indicative of structure of good sterio–chemical quality. The Ramachandran plot for Mce1R structure displays that only 2 residues (Ser21 and one glycine residue) are placed at the disallowed region. Notably, Ser21 is located at the N–terminal region of the first helix (α1) just after the structurally disordered 'arm' region; and therefore possibly making it deviated from suitable favorable region. As glycine imparts flexibility to the peptide backbone (because of the presence of smallest side chain; hydrogen atom), it can attain a wide range of torsion angles (φ and ψ) and therefore sometimes may be located in the disallowed region in the Ramachandran plot. The remaining residues are present mostly in favored or allowed regions suggesting that the conformation of the modeled structure is energetically stable. The ERRAT computed quality factor for the structure was 82.71 which suggested that 82.71% of protein residues error value falls below the 95% rejection limit. Compared to high resolution protein structures where the quality factor lies in the region 95–100, the Mce1R modeled structure seems to be of low resolution with quality factor below 90. The compatibility of the structure was further checked using Verify 3D which showed that 73.54% of residues have average 3D–1D score ≥ 0.2 which is more than the cut off score 65%. The Z score computed by the ProSA server was –6.03 which also showed that the structure lied within the normal distribution of native protein structures of similar sizes (Fig. 4.2.1b). Overall, these data have computationally validated the modeled Mce1R structure. The final model (Fig. 4.2.2a) displays that α helices are major structural components. The structure also indicated presence of disordered region at the N–terminal end, composed of first ~19 amino acid residues, devoid of any secondary structure elements. A small glycine-rich C–terminal end was also designated as disordered region. The final model appears to possess two domains– an N–terminal domain consists of three α helices and two β sheets, connected by a small loop (the Wing) and the C–terminal domain composed of all remaining six α helices, comprising a bundle of antiparallel helices, similar to canonical VanR–type proteins [110, 114].

The modeled structure also showed that all three tryptophan residues (Trp95, Trp161 and Trp214) were fully exposed at the surface (Fig. 4.2.2b). To check whether the purified active Mce1R carried the tryptophan residues at the surface, acrylamide mediated quenching of intrinsic fluorescence of the tryptophan residues has been performed following a standard method [127]. The major advantage of using acrylamide as quencher is that it is an electrically neutral and polar quencher of high quenching efficiency (γ = 1) [167]

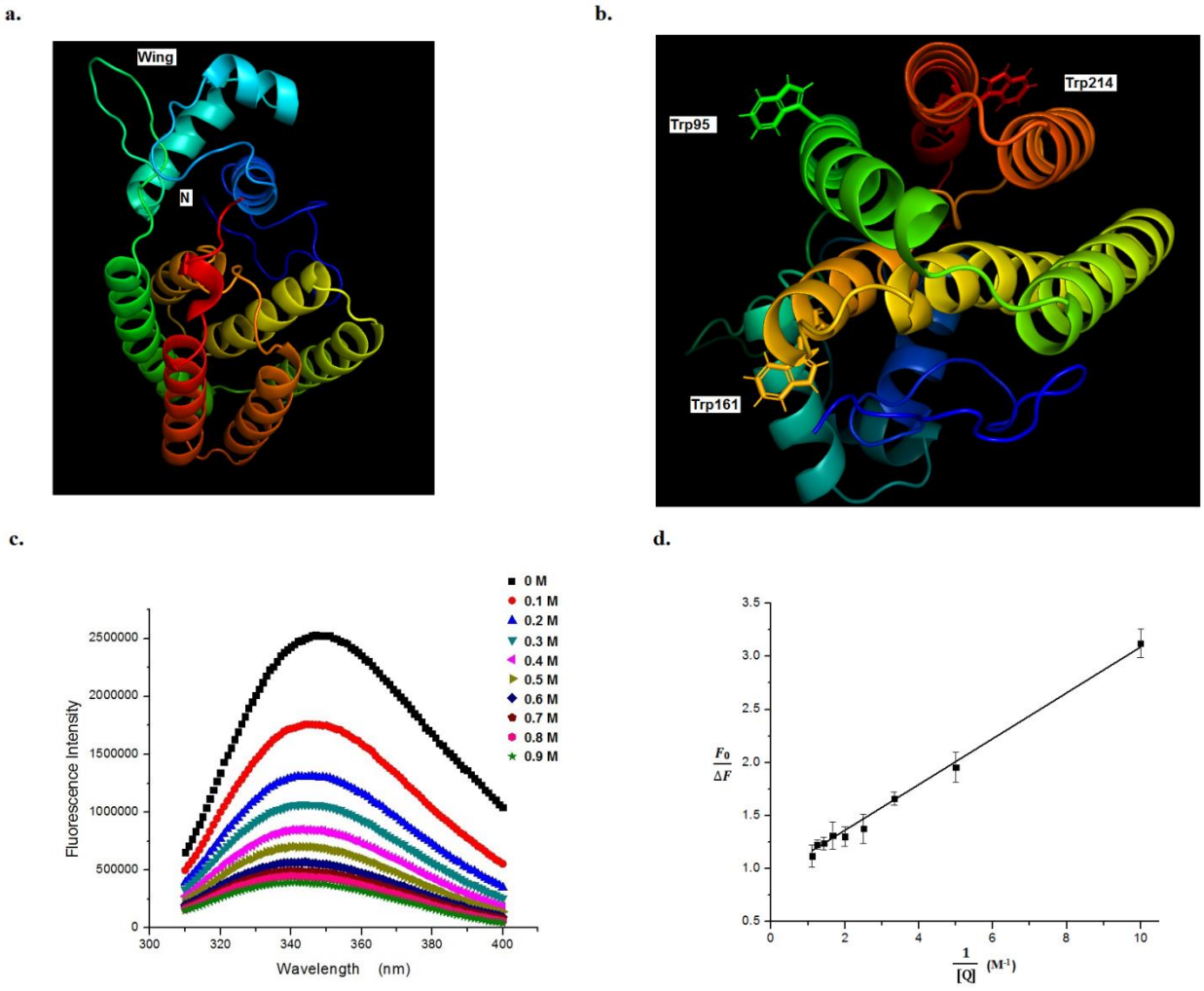


Fig. 4.2.2 Surface exposure of the tryptophan residues of Mce1R. (a) The energy minimized final model of Mce1R. N and Wing indicate the N-terminal end and the wing generated by the β-sheets respectively. (b) The positions of the tryptophan residues are shown in the model. (c) Tryptophan fluorescence quenching spectra of Mce1R in presence of (0-9 M) acrylamide. At the top right corner, the concentrations of acrylamide and colors of corresponding data are displayed. (d) The Lehrer plot to determine fractional accessibility of the tryptophan residues.

which can easily interact with exposed tryptophan residues, specifically those present in an electrically charged environment, without being influenced by any electrostatic forces exerted by neighboring positively or negatively charged amino acid sidechains. The Fig. 4.2.2c displays the tryptophan fluorescence emission spectra in presence of various acrylamide concentrations (0–0.9 M). Notably, the maximum fluorescence emission for Mce1R without acrylamide was observed at 347 nm (λmax) which suggested that tryptophan residues might be exposed and accessible to the solvent molecules since the λmax value for the exposed tryptophan residues generally lies around ~350 nm [168]. To confirm surface exposure of the tryptophan residues, the Lehrer plot [129] analysis (Fig. 4.2.2d) following the equation (3), was performed as described in the "Materials and Methods" section 3.2.2.18. The average data obtained from three independent experiments were fitted with straight line using the OriginPro 8 software (OriginLab Corporation, MA) and the R–square value obtained for the linear fit was more than 0.99 which indicated a good fit. The error bars are indicative of standard deviation. The value of the intercept of the straight line on vertical axis was 0.93±0.04 from which the fa (fractional accessibility) was determined to be 1.07±0.04, which indicated that all three tryptophan residues were exposed at the surface. The good linearity (R-square value > 0.99) of the plot also suggests that the three tryptophan residues of Mce1R are present in identical microenvironment and all are involved in dynamic quenching process with equal quenching constant K = 4.65±0.01 M^{-1}. Taken together, the computationally predicted Mce1R model is validated by both computational and experimental methods.

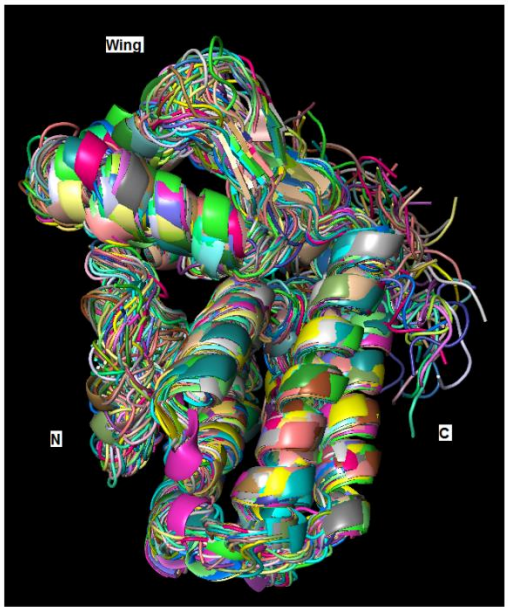
Tryptophan, being the most hydrophobic amino acid, bears highest propensity, compared to other hydrophobic amino acid residues, to be located at the hydrophobic core of the protein structures (buried from the solvent) rather than at the surfaces (exposed to the solvent) [171]. Tryptophan fluorescence is highly sensitive to its specific micro–environments (buried or exposed) and therefore analysis on tryptophan fluorescence spectra of proteins serves an excellent experimental means to confirm stability and correct folding of proteins [127, 129, 167,168, 172]. Mce1R structure shows that all the three tryptophan residues are exposed at the surface and it has been confirmed by acrylamide mediated quenching analysis of intrinsic fluorescence of those tryptophan residues. Such good correlation between the results of computational and experimental methods suggests that in reality, Mce1R assumes the modeled structure. However, surface exposure of tryptophan residues is not very rare and neither always indicates structural destabilization [173]. Interestingly, presence of hydrophobic residues adjacent to the exposed tryptophan residues many times stabilizes the

exposed conformation of tryptophan residues by forming a hydrophobic cluster [173–175]. The residues adjacent to Trp95 and Trp161 are Ile93, Phe94 and Leu96 and Leu159, Ala160, Phe162, Leu163 and Leu164 respectively, might be stabilizing those exposed tryptophan residues. However, the condition for Trp214 is different than that of Trp95 and Trp161 as it is located at the C–terminal tail region which is composed of mostly glycine and charged residues and therefore exposed to solvent.

4.2.2 Objective B2- Coarse grain molecular dynamics simulation of Mce1R by CABS–Flex server

Protein molecules are not entirely rigid and for their function and stability, structural dynamics plays an important role [176, 177]. Although the monomeric structure of Mce1R has been modeled and validated by computational and experimental methods, it is necessary to assess the structural dynamics of Mce1R to correlate the modeled structure with different specific aspects of its functions. To investigate structural dynamics of Mce1R at molecular level, the CABS Flex 2.0 server has been used [130]. The figure (Fig. 4.2.3a) shows the structural dynamics of 100 ns simulation time with superimposed picture of all sampled models with different conformations during the trajectory of the simulation. The residue–wise fluctuation plot during the simulation is shown in Fig. 4.2.3b. The plot displays several regions of Mce1R structure with different degree of flexibilities. The specific regions with significant fluctuations are marked with the corresponding residues in the figure (Fig. 4.2.3b). The N–terminal first 1–22 residues, those comprise the disordered region, showed fluctuations within the range ~1.8–3Å. The residues 33–40 and 50–60 have RMSF values within ~1–2.7Å and ~1.4–3.9Å respectively which overlap within the HTH DNA binding motif of Mce1R. Notably, these residues show a bit more flexibility probably due to playing direct roles in binding to DNA, suggesting possible conformational rearrangements upon DNA binding. The residues, 73–81 fluctuating within the range ~1.4–4Å, comprise the "wing" region of the wHTH motif. The wing is typically composed of two anti–parallel β–sheets connected by a short loop, a distinctive structural element of the wHTH motif which plays an important role in DNA recognition and specificity [156]. Considering the fluctuations of the disordered region, HTH and wing region, the N–terminal region of Mce1R seems to have more flexibility than the rest of the structures. The superimposed conformations of the N–terminal region are represented in Fig. 4.2.3a.

a.



b.

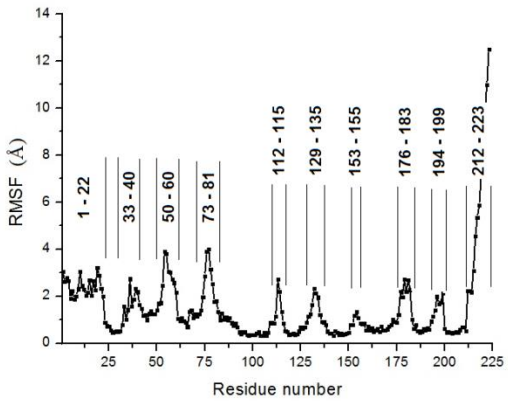


Fig. 4.2.3 Course grain molecular dynamics simulation of Mce1R. (a) All conformations of Mce1R during the total simulation time are superimposed to show structural dynamics. 'N', 'C' and 'Wing' indicate N– and C– terminals and wing structure of Mce1R. (b) Residue–wise fluctuation plot of Mce1R. RMSF values are plotted along the y axis and the residue positions are plotted along x axis. The regions of Mce1R showing higher fluctuations are marked over the plot by the corresponding residue numbers.

The other flexible residues are 112–115, 129–135, 153–155, 176–183 and 194–199 located at the C–terminal region, having RMSF values within ~1–2.5Å which suggest less flexibility compared to the N–terminal region. These regions mostly map the linkers between two helices. But the C–terminal α helices have shown high structural stability (RMSF < 1Å) during the whole simulation process. The C–terminal end residues (212–223) show flexibility within ~2.2–12.5Å as this region is rich in glycine and polar residues therefore showing highest fluctuation.

The server CABS Flex 2.0 uses the principles of coarse–grain molecular dynamics simulation in order to simplify the simulation process but at the same time preserves main chain hydrogen bond attractions and specifically side chain contact potentials to accommodate mutual orientations of amino acid side chains using knowledge–based force field derived from statistical analysis of known protein structures [130–132]. The server has been used in several studies including analysis of dynamics and flexibilities of proteins (reviewed in [178]). Since the wHTH motif of Mce1R is directly involved in DNA recognition, sufficient conformational flexibility of this motif is a prerequisite for efficient recognition and binding to DNA. Conformational flexibilities of DNA–binding HTH motifs have been demonstrated by a number of research groups previously [179–182]. Consistent to this, molecular dynamics simulation study suggested that the N–terminal domain of Mce1R including the 'arm' was structurally dynamic. Notably, the residues– Asp46, Val56, Arg60 and Pro76 of wHTH motif of Mce1R, exhibit significant fluctuations which further supports their possible roles in DNA binding. Several studies have demonstrated structural stabilization of the flexible HTH motifs after binding to DNA [183–185] to increase the stability of the bound complex which suggests that the N–terminal domain of Mce1R may attain more stable conformation after binding to DNA compared to the unbound state. In contrast, the C–terminal domain was found to be structurally more rigid compared to the N–terminal domain. Interestingly, the two exposed tryptophan residues– Trp95 and Trp161 of Mce1R, were also observed to be located within the rigid regions of the C–terminal domain which suggests that exposure of tryptophan residues did not affect the stability of the structure of Mce1R and therefore supports the result of quenching analysis of tryptophan fluorescence spectra mentioned before.

Next it was hypothesized that structures similar to that of Mce1R might exhibit structural dynamics similar to that of Mce1R. To identify structural homologs of Mce1R, the DaliLite server (V5.0) [135] was run using the structure of Mce1R as input to search against PDB

database. The server arranges the structures in descending order according to their respective Z scores. A high Z score indicates high similarity between the query and subject structures. The structures up to the cut off Z score 17.3 were selected (see Materials and Methods; section 3.2.3.5) and compared with that of Mce1R using the PROMALS3D server [137]. The Table 4.2 displays the DaliLite server scores for the selected structures. Among the selected structures, the structure for the PDB id 6az6 was not considered for the PROMALS3D server analysis as 6az6 was structurally related to FadR subfamily. The PROMALS3D server did not consider the structure of the PDB code 3c7j as the RMSD score was reasonably high (6.1Å) during aligning the structures. The final alignment is shown in the figure (Fig. 4.2.4). The secondary structural elements are numbered (α1–9; β1 and β2). The flexible residues of Mce1R and the corresponding aligned residues of other structures were marked by the boxes. At the N–terminal ends of all the proteins there were extra amino acids (see Table 4.2.) those did not resolute in the crystal structures possibly because of high flexibility. Notably at the N–terminal domain of Mce1R (up to the second β–sheet; β2) majority of the flexible residues of Mce1R aligned with corresponding residues of other structures which are either identical amino acids or bearing sidechains of similar chemical properties. Since position and chemical properties of amino acid residues play important roles in structural dynamics in proteins [186], the N–terminal domains of other homologous structures most likely to exhibit structural dynamics similar to that of Mce1R. However, at the C–terminal domains (from α4 to the end residue), the putative flexible residues showed little conservancy but the six α helices (α4–9) of all structures exhibited significant presence of amino acids bearing sidechains of similar chemical properties which suggest that the helices at the C–terminal regions of the homologous structures might be structurally more stable compared to their N–terminal residues. Taken together, the structural dynamics of the homologous structures seems to be similar with that of Mce1R.

Table 4.2 Alignment scores of homologous structures from the DaliLite V5.0 server

Sl. No.	PDB id	Chain	Z score	RMSD (Å)	% similarity	No of amino acids before the first helix (α1)
1.	3fms	A	22.7	2.1	18	8
2.	4p9f	A	20.2	2.8	19	11
3.	3ihu	B	19.6	2.5	22	16
4.	3c7j	B	19.3	6.1	16	Not aligned
5.	6ep3	B	18.1	2.6	5	7
6.	6az6	A	18.0	2.6	17	Not aligned
7.	2hs5	A	17.8	3.0	12	28

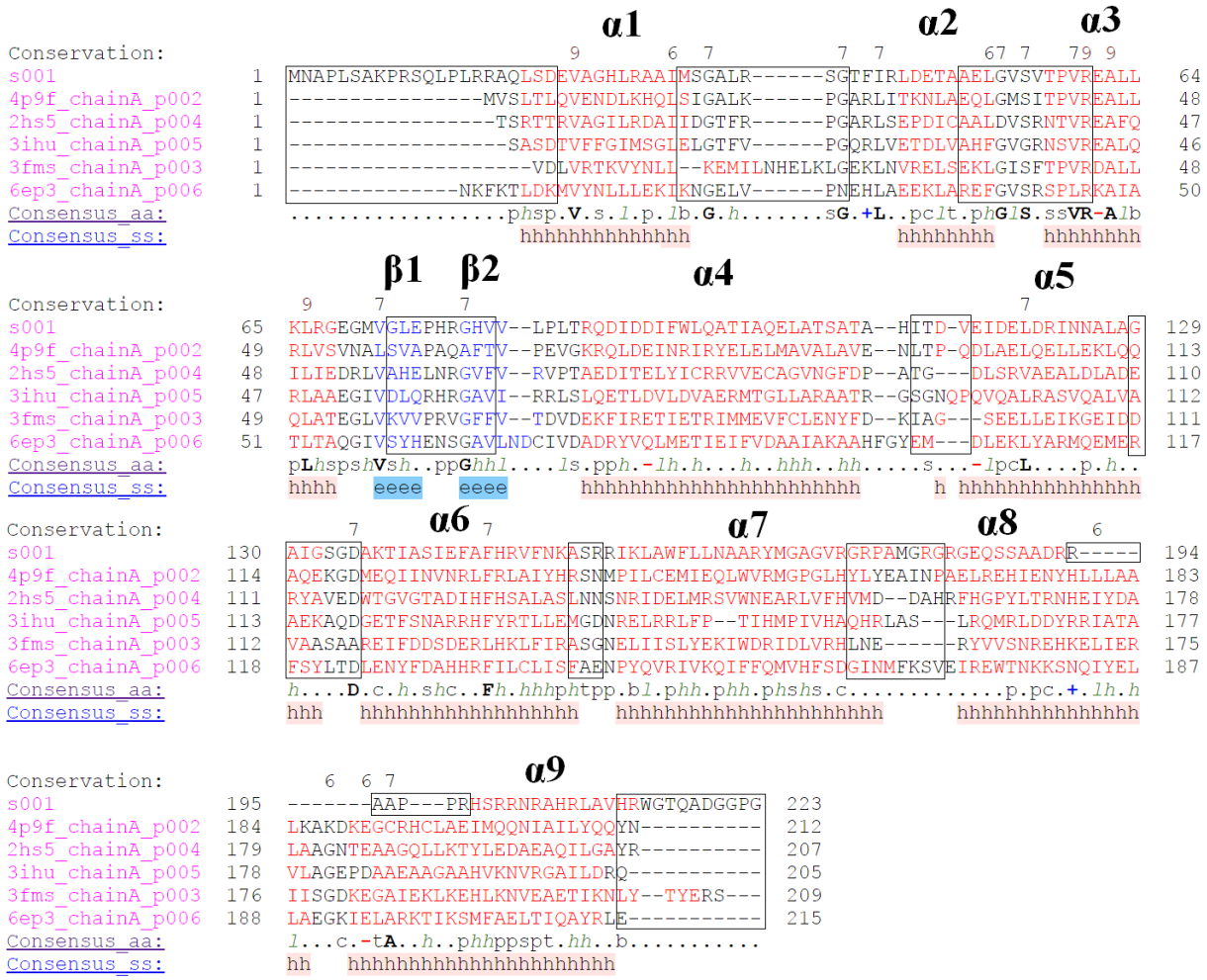


Fig. 4.2.4 PROMALS3D server analysis of structures homologous to Mce1R. S001 indicates the Mce1R structure. All other structures are represented by their respective PDB ids followed by chain ids. The digits over the aligned sequences are indicative of sequence conservancy. Higher digits indicate high conservancy of the particular residue. Below the aligned sequences the bold letters are indicating the conserved residues. Consensus amino acid residues with small, bulky, charged, aliphatic, positively charged, hydrophobic, polar, tiny, negatively charged and aromatic side chains are indicated by the letter codes “s”, “b”, “c”, “l”, “+”, “h”, “p”, “t”, “-” and “@” respectively. α1–9 indicate the α helices numbered sequentially from the N-terminal ends. β1&2 indicate the β sheets. The regions with possible similar structural flexibilities are boxed. The region between α8 and α9 of Mce1R did not align well with the other structures and therefore only this Mce1R region is boxed.

4.2.3 Objective B3- Modeling of the dimeric structure of Mce1R by GalaxyHomomer server and its validation

The proteins from the GntR superfamily have been predicted to form dimers through their C–terminal domains [114]. All crystal structures of different proteins from the GntR superfamily, deposited till date in the PDB database (https://www.rcsb.org/), also display their dimeric nature. Mce1R belongs to the VanR group of FCD family of GntR superfamily of regulators [110] and therefore bears the possibility to exist as dimeric form. To generate the dimeric structure of Mce1R, the modeled monomeric Mce1R structure from section 4.2.1, was submitted to the GalaxyHomomer server [133]. The server detects most suitable five templates for the prediction of oligomeric state and modeling the oligomer for a submitted protein structure following sequence and structure based scoring functions [187, 188]. Initially the server performs a sequence and secondary structure based similarity search following the HHsearch algorithm [189] and assigns a similarity score (S score) against each template. Five most suitable templates are then selected according to the ranking based on the S scores among those templates whose monomeric structures are most similar (TM score > 0.5) to the submitted structure (structural similarity is measured as the TM score, determined using the TM–align program [188]). The server also simultaneously determines the oligomeric state for the given structure by estimating the oligomeric state ratios for the selected templates as described [133]. As the five selected templates are structurally dimeric, the oligomeric state for Mce1R has been selected to be dimer. The server has generated five dimeric models in pdb format for Mce1R (Table 4.3). Interestingly, two dimeric models for Mce1R (Model 1 and Model 3) have been generated by the server using the same protein structure (McbR; PDB Id: 4P9F) with different interface areas and TM scores. Notably, the two monomers of McbR are not structurally identical [158]; because of which differences in the interface area and TM score have been reported by the server. The model with highest TM score among others (Model 1) has been selected for subsequent analysis since the submitted Mce1R structure is most similar to that particular template subunit (PDB Id: 4P9F).

The Model 1 has been submitted to the Swiss PDB Viewer tool (V4.1.0) [125] for energy minimization following 1500 steps of steepest descent algorithm using the GROMOS96 43B1 forcefield. The energy for the structure before and after energy minimization was –20306 KJ/mol and –29198.043 KJ/mol respectively.

Table 4.3 Oligomeric models of Mce1R predicted by the GalaxyHomomer server

Sl No.	Oligomer template	No. of subunits	Interface area $(Å)^2$	Sequence similarity	TM Score
Model 1	4P9F	2–mer	1688.5	17.9	0.8246
Model 2	3SXY	2–mer	1460.3	14.8	0.7434
Model 3	4P9F	2–mer	1199.4	17.9	0.7777
Model 4	3FMS	2–mer	1095.8	15.2	0.7685
Model 5	4P96	2–mer	2210.0	12.1	0.6407

The energy minimized dimeric structure, shown in Fig. 4.2.5a, was submitted to the SAVES (V6.0) (https://saves.mbi.ucla.edu/) server for validation. The SAVES server simultaneously runs a number of structure validation programs out of which the results obtained from ERRAT [164], Verify 3D [163] and PROCHECK [165] are important to assess the quality of the structure. ERRAT first determines the error functions for a 9–residue sliding window along the submitted protein structure by statistically analyzing the non–bonded interactions between different atom types in comparison with experimental structures of high resolution. The error functions are then plotted against different positions of the 9–residue sliding window. The ERRAT score for the structure has been determined to be 81.6471, suggesting that ≈ 81.64% of the total amino acid residues of Mce1R have calculated error values lower than 95% rejection limit. Generally, for high resolution structures (< 2Å) the ERRAT scores used to be more than 95 and for low resolution structures (between 2.5 – 3Å) the ERRAT scores found to be around 91 [164]. Therefore, it appears that the modeled structure is of low resolution. The other program, Verify 3D evaluates the compatibility of a given three–dimensional structure with its amino acid sequence (3D–1D score) by first estimating the respective propensities of each residue to be present in specific classes of structural environment in the given model, calculated based on the statistics for that of the residues of the experimentally determined structures of high resolution, present in the PDB (https://www.rcsb.org/). The program then scores the given structure by adding the individual propensities of each residue [163]. The lower and upper limits of such scores are –1 (poor) and +1 (good) where at least 80% of the residues in a given structure need to have 3D–1D scores ≥ 0.2 in order to be considered as a structure of good compatibility. Analysis of Mce1R structure by Verify 3D showed that 90.81% of the residues have 3D–1D scores ≥ 0.2, suggesting that the generated model is reasonably compatible with its 1D profile. Moreover, the average scores for all the amino acid residues of the both chains, determined by Verify 3D (shown in Fig. 4.2.5b), shown to have positive values, suggesting the proposition made above. The Ramachandran plot (Fig. 4.2.5c) generated by the program PROCHECK [165] showed that 85.3% of the residues were in most favored regions, 12.6% residues were is additional allowed regions, 1.6% residues were in generously allowed regions and only 2 residues (Ser34 of the A and B chains) were in disallowed regions. These two residues belong to the flexible regions of the protein structure (Fig. 4.2.3b) which probably allowed them to attain such combinations of torsion angles. The overall G factor, computed by PROCHECK,

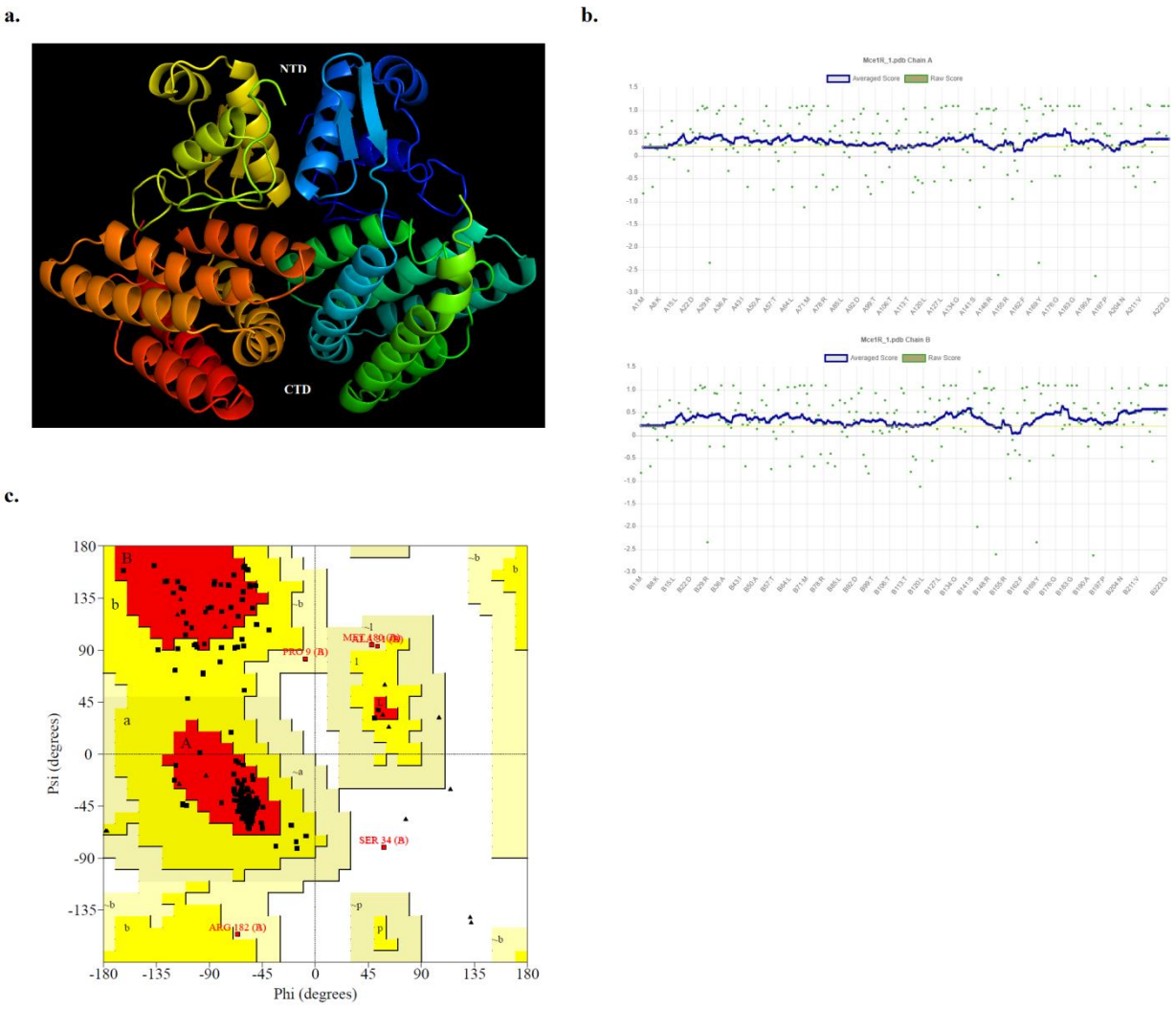


Fig. 4.2.5 Dimeric structure of Mce1R. (a) The validated dimeric structure of Mce1R showing distinct two domain structure for each monomer. The N– and C–terminal domains are marked as "NTD" and "CTD" respectively. (b) The residue–specific scores predicted by the Verify 3D program for each subunit. (c) The Ramachandran plot for the dimeric Mce1R predicted by the PROCHECK program, showing the steriological positions of each residues. The residue Ser34 of both subunits present in disallowed region (white field) hence marked. The dark–filled triangles (▲) indicate glycine residues. The most favorable, generously allowed and additionally allowed regions are shown in red, yellow and light yellow colors respectively.

considering both the torsion angles and covalent geometry, for the structure of Mce1R was found to be –0.02 which is within the lower limit –0.05, suggesting conformational stability of the generated dimeric structure [166].

The programs, ERRAT, Verify 3D and PROCHECK, available in the SAVES server, evaluate a given protein structure based on the criteria attributed to the properties of different side chains of amino acid residues such as; non–bonded interactions among the amino acid side chains, specific classes of structural environments and steric interactions among the side chains of the amino acid residues in a three–dimensional space. But these programs do not consider the contribution of different forces which stabilize the native fold of a protein structure in a solution. The native fold of a protein structure is best described by the spatial arrangements of Cα atoms of the amino acid residues constituting the peptide backbone, i.e. the Cα trace [190]. Therefore, a predicted protein structure with altered conformation (non–native folds; altered Cα trace) may appear to these programs as natively folded structure since it may satisfy all or most of the criteria of these programs. Consistent to this, the predicted random coil model of Tanford, Flory and Ramachandran suggested the occupancy of each amino acid residues in all sterically allowed regions in the Ramachandran plot with almost equal probabilities [191–193]. Therefore, to address this issue, the ProSA server [126] was used which checks a given protein structure for possible errors (local regions of higher energies) considering the potentials of the mean forces acting on the Cα trace. Initially, from the PDB database of experimentally solved globular protein structures (both X–ray and NMR structures), the stabilizing forces were extracted for individual protein structures in the form of Cα potentials of mean force using Boltzmann's principle [194] from which the energies of the structures are computed based on a distance–based pair potential and potential of the residues exposed to solvent [195, 196]. For each of these structures, a 'z' score is then calculated [190] from these computed energies and plotted against the number of residues of each structures. The Mce1R structure was submitted to the ProSA server and the 'z' score was estimated to be –5.92 which was found to lie within the range of the 'z' scores determined for the experimental native protein structures of similar sizes (Fig. 4.2.6a; the • spot), suggesting that the submitted structure consisted of native structural folds. The server also generated a plot of residue energies for Mce1R structure (Fig. 4.2.6b), showing the mean energies for a 40 residue sliding window positions along the entire sequence of Mce1R (the dark green line). The mean energy plot suggests that the modeled structure does not contain any unfolded or high energy region(s) as the mean energies for almost all residue

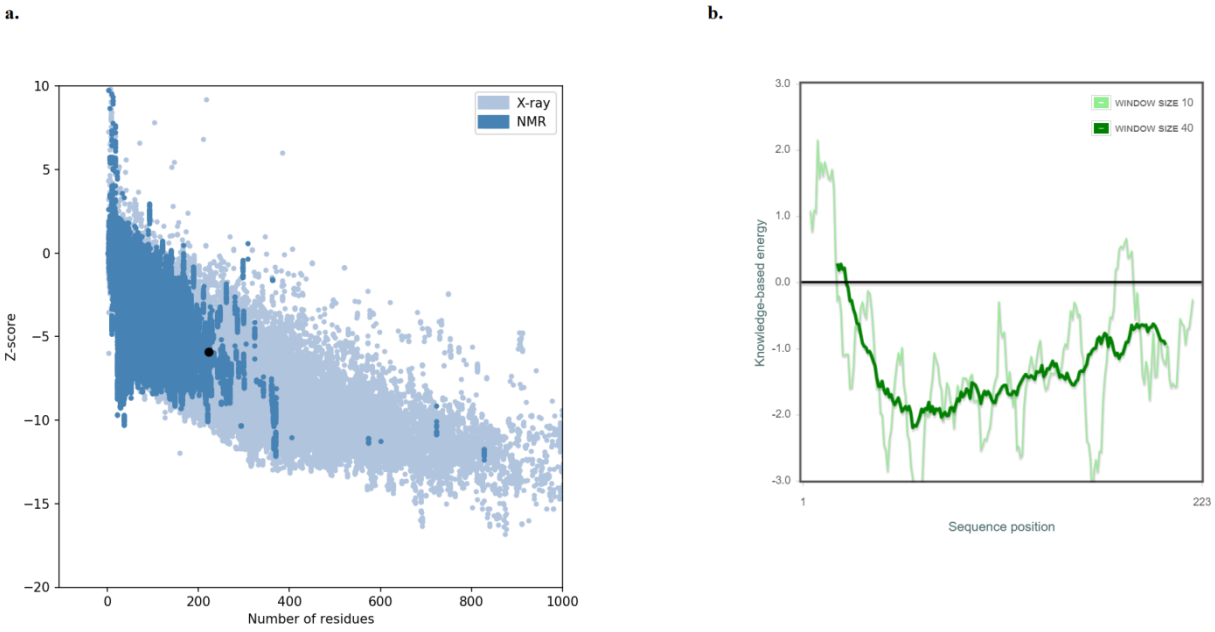


Fig. 4.2.6 Mce1R dimeric structure validation by ProSA. (a) The "z" score predicted by the ProSA server for the dimeric Mce1R model. The dark–filled circle (•) in the plot showing the position of Mce1R dimer relative to the experimentally determined (NMR and X–Ray methods) structures of similar sizes. (b) The residue–specific knowledge–based mean energies determined by the ProSA server for Mce1R in a 40–residue window are plotted in dark green line. The negative values indicate energetically stable native fold of Mce1R.

positions bear reasonable negative values [126]. Overall these results have computationally validated the generated dimeric model of Mce1R.

A protein dimer is stabilized by several types of inter–subunit interactions. The dimeric protein McbR (a VanR–type protein; PDB id 4P9F) of *E. coli*, the structural ortholog of Mce1R, has been shown to be stabilized by various types of non–covalent interactions contributed by two monomers [158]. Therefore, to assess the stability of the generated Mce1R dimer, the structure was submitted to the Protein Interactions Calculator (PIC) webserver (http://pic.mbu.iisc.ernet.in/job.html) [134] which analyzed various interactions between two subunits. Among various types of non–covalent interactions, the hydrophobic interaction is considered to be the major stabilizing interaction to form a dimer or protein–protein interaction in general [197–200]. The Table 4.4 displays the hydrophobic interactions between various residues of the two subunits of Mce1R. It was observed that majority of the residues of the C–terminal domains from both subunits (Table 4.4; Sl. No. 6–19) participated in hydrophobic interactions while a relatively less number of residues from the N–terminal domains (Table 4.4; Sl. No. 1–5) contributed to the same. Additionally, various hydrophilic interactions (ionic and hydrogen bond interactions) also have been reported by the server, involving various residues from the both subunits (Figs. 4.2.7a, b and c), thereby providing additional stability to the dimeric structure of Mce1R. Analysis of the structure of McbR by the PIC server also revealed such types of stabilizing interactions between the subunits [158].

Computational methods mentioned above suggest that Mce1R forms stable dimer. To validate it further, gel–filtration chromatography of purified Mce1R was performed following the protocol mentioned in the section 3.2.2.19. The elution times of BSA, carbonic anhydrase and lysozyme were ~13.709 min, ~15.714 min and ~16.304 min respectively (Fig. 4.2.8). The molecular weight of monomeric His–tagged Mce1R is ~25.3 kDa. Therefore, for dimeric species of Mce1R, the peak was expected to appear well below that of BSA. Interestingly, the chromatogram of Mce1R showed the appearance of a single strong peak at ~13.618 min which closely matches with that of BSA (~13.709 min). The most possible reason for this discrepancy could be the specific molecular shape of dimeric Mce1R. For the protein molecules of not completely globular or symmetrical shapes, the elution times are reported to be aberrantly higher compared to symmetrical globular protein molecules of similar masses [239]. The distinct two–domain structure with a flexible N–terminal arm of Mce1R dimer also suggests that the structure is not of completely globular shape which might have caused such elution profile. Considering all, Mce1R is found to exist as dimer in solution.

Table 4.4 Hydrophobic interactions among the residues from chain A and chain B of Mce1R

Sl. No.	Position	Residue	Chain	Position	Residue	Chain
1.	1	Met	A	74	Leu	B
2.	1	Met	A	76	Pro	B
3.	64	Leu	A	64	Leu	B
4.	74	Leu	A	1	Met	B
5.	76	Pro	A	1	Met	B
6.	95	Trp	A	159	Leu	B
7.	96	Leu	A	100	Ile	B
8.	96	Leu	A	159	Leu	B
9.	96	Leu	A	162	Phe	B
10.	100	Ile	A	100	Ile	B
11.	100	Ile	A	96	Leu	B
12.	159	Leu	A	95	Trp	B
13.	159	Leu	A	96	Leu	B
14.	161	Trp	A	161	Trp	B
15.	161	Trp	A	162	Phe	B
16.	162	Phe	A	161	Trp	B
17.	162	Phe	A	162	Phe	B
18.	162	Phe	A	96	Leu	B
19.	169	Tyr	A	161	Trp	B

a.

Protein-Protein Main Chain-Side Chain Hydrogen Bonds

Rasmol Jmol [help]

[View the original hbond output]

Mce1R_Dimer1.pdb

DONOR				ACCEPTOR				PARAMETERS				
POS	CHAIN	RES	ATOM	POS	CHAIN	RES	ATOM	MO	Dd-a	Dh-a	A(d-H-N)	A(a-O=C)
17	A	ARG	NH2	64	B	LEU	O	1	3.02	2.09	146.01	86.13
17	A	ARG	NH2	64	B	LEU	O	2	3.02	3.62	47.12	86.13
88	A	GLN	NE2	155	B	ARG	O	1	3.28	2.70	115.05	132.11
88	A	GLN	NE2	155	B	ARG	O	2	3.28	3.24	83.04	132.11
161	A	TRP	NE1	68	B	GLY	O	-	2.82	2.00	147.10	159.61
17	B	ARG	NE	64	A	LEU	O	-	3.44	2.73	127.20	113.76
17	B	ARG	NH2	64	A	LEU	O	1	3.01	2.11	143.38	86.29
17	B	ARG	NH2	64	A	LEU	O	2	3.01	3.71	40.99	86.29
157	B	ILE	N	92	A	ASP	OD2	-	3.36	2.93	106.23	154.05
158	B	LYS	N	92	A	ASP	OD2	-	3.33	2.34	175.52	150.97
161	B	TRP	NE1	68	A	GLY	O	-	2.80	1.97	149.42	153.56

Dd-a = Distance Between Donor and Acceptor
Dh-a = Distance Between Hydrogen and Acceptor
A(d-H-N) = Angle Between Donor-H-N
A(a-O=C) = Angle Between Acceptor-O=C
MO = Multiple Occupancy
Note that angles that are undefined are written as 999.99

b.

Protein-Protein Side Chain-Side Chain Hydrogen Bonds

Rasmol Jmol [help]

[View the original hbond output]

Mce1R_Dimer1.pdb

DONOR				ACCEPTOR				PARAMETERS				
POS	CHAIN	RES	ATOM	POS	CHAIN	RES	ATOM	MO	Dd-a	Dh-a	A(d-H-N)	A(a-O=C)
95	A	TRP	NE1	103	B	GLU	OE1	-	2.94	2.25	131.79	999.99
156	A	ARG	NE	92	B	ASP	OD1	-	2.78	1.76	168.18	999.99
95	B	TRP	NE1	103	A	GLU	OE1	-	2.93	2.23	131.42	999.99
156	B	ARG	NE	92	A	ASP	OD1	-	2.93	1.96	155.70	999.99

Dd-a = Distance Between Donor and Acceptor
Dh-a = Distance Between Hydrogen and Acceptor
A(d-H-N) = Angle Between Donor-H-N
A(a-O=C) = Angle Between Acceptor-O=C
MO = Multiple Occupancy
Note that angles that are undefined are written as 999.99

c.

Protein-Protein Ionic Interactions

Rasmol Jmol [help]

Mce1R_Dimer1.pdb

Ionic Interactions within 6 Angstroms

Position	Residue	Chain	Position	Residue	Chain
61	GLU	A	67	ARG	B
67	ARG	A	61	GLU	B
89	ASP	A	158	LYS	B
91	ASP	A	156	ARG	B
92	ASP	A	156	ARG	B
92	ASP	A	158	LYS	B
156	ARG	A	91	ASP	B
156	ARG	A	92	ASP	B
158	LYS	A	89	ASP	B
158	LYS	A	92	ASP	B

Fig. 4.2.7 (a), (b) and (c) Various ionic and hydrogen bond interactions among the residues of two subunits of Mce1R dimer identified by the PIC server.

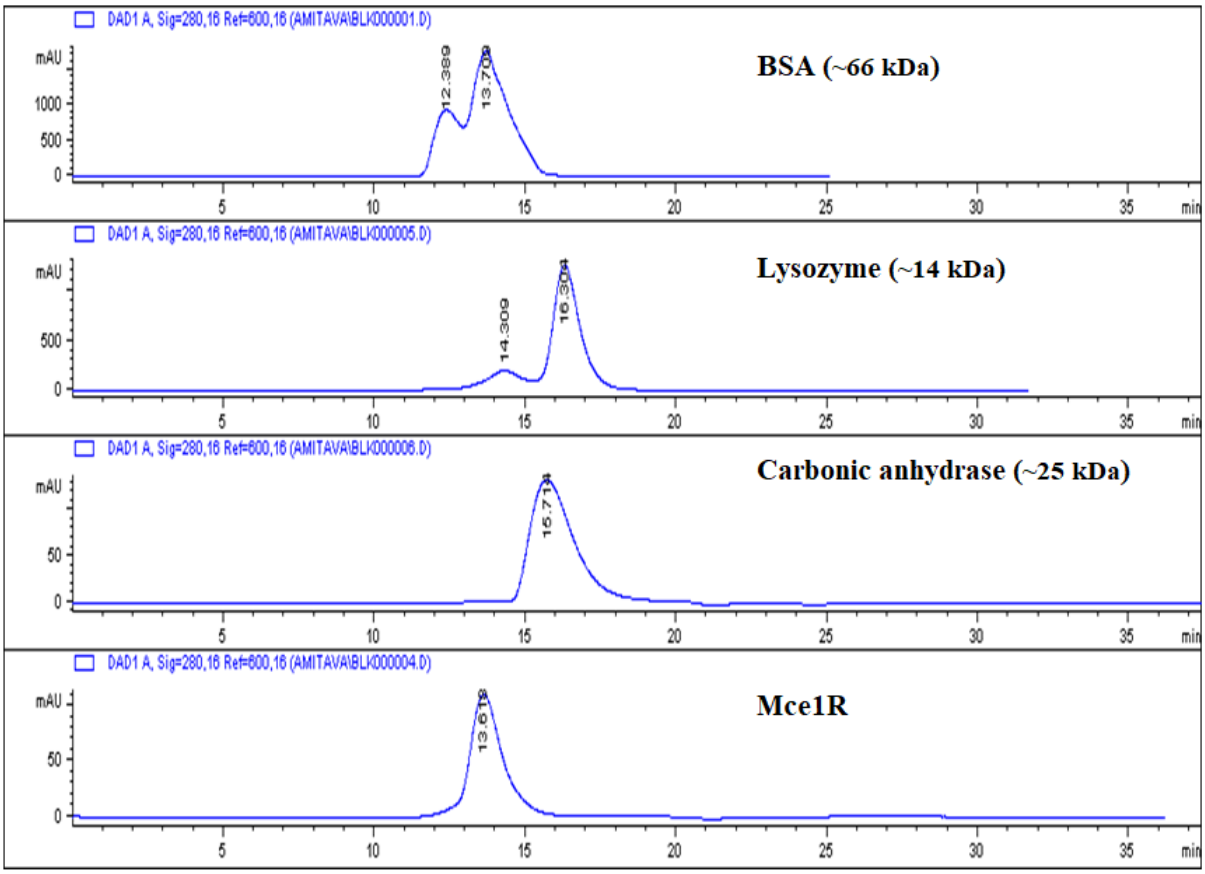


Fig. 4.2.8 Gel–filtration chromatography of Mce1R. Elution profiles of molecular weight markers; (BSA, carbonic anhydrase and lysozyme) are shown. Molecular masses of each marker are also indicated. Elution profile of Mce1R is shown in the lowest panel.

4.2.4 Objective B4- Identification of conserved amino acid residues of Mce1R by multiple sequence alignment

Amino acid residues critical for maintaining structure and function of proteins tend to be conserved among orthologs. To identify orthologs of Mce1R, PSI–BLAST was run on MEGA–X platform and those selected sequences were aligned using the MUSCLE algorithm. The alignment is shown in the figure 4.2.9a. The fully conserved residues are marked with (*) sign. The alignment shows that a significant number of residues of the N–terminal domain (up to β2) of Mce1R are highly conserved across different organisms which suggest that structures of the N–terminal domains and DNA binding activities of the orthologs might be highly similar to that of Mce1R. Notably, at the N–terminal disordered region of Mce1R, some of the residues (Pro, Gln and Arg) are also found to be highly conserved or replaced by amino acid of similar properties (e.g. Lys) (Fig. 4.2.9a; residues marked by dark filled circles). Such N–terminal disordered regions of proteins, carrying positively charged residues, were reported to play important roles in DNA binding [201–207]. Being positively charged, these regions of Mce1R may make contacts with the DNA during protein–DNA complex formation. The entire wHTH motif of Mce1R carries highly conserved residues. In contrast, the C–terminal domain (from α4 to the end residue) of the Mce1R exhibits little sequence conservancy. This observation is in agreement with the reported diversity of the C–terminal regions of VanR family of proteins, involved in binding of varieties of small inducer molecules [110, 114]. Despite sequence diversity, at many positions in the C–terminal domain of Mce1R, specifically hydrophobic residues are substituted by other residues with highly similar sidechains (i.e. hydrophobic to hydrophobic substitutions; colored by the yellow color in Fig. 4.2.9a) in the orthologs suggesting that preserving hydrophobic environment in those specific positions are important; probably for maintaining proper structure of the C–terminal domains. However, A few residues of the C–terminal domain are found to be highly conserved among the various orthologs, suggesting that they might be involved in performing specific functions.

4.2.5 Objective B5- Detection of ligand binding cavity in the modeled structure

Generally, the proteins with FCD domains carry cavities which act as the inducer/ligand binding sites [157, 158, 208]. To test whether in Mce1R such type of cavity was present, the Mce1R structure was submitted in the CavityPlus web server [139]. The server detected a

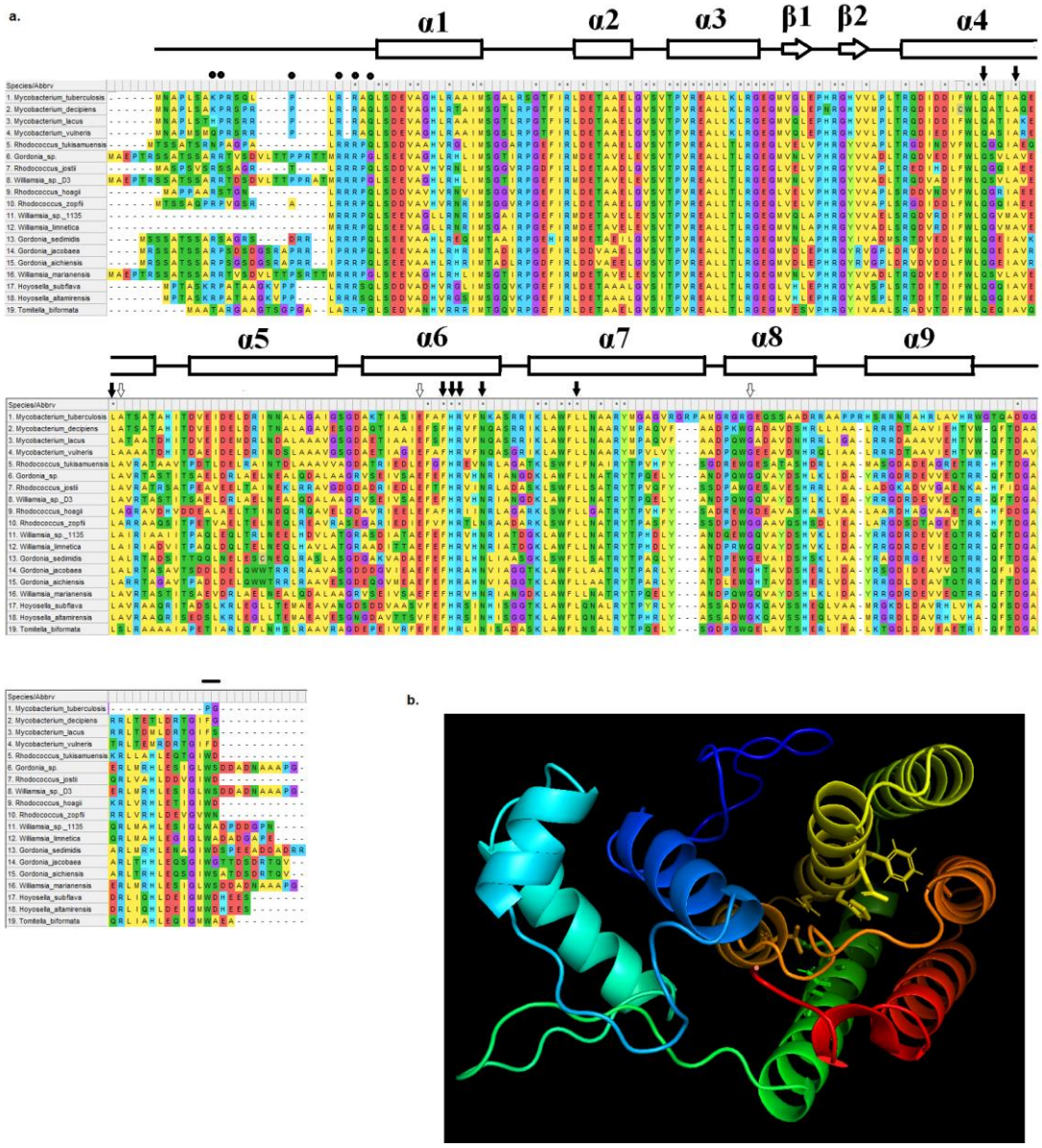


Fig. 4.2.9 The multiple sequence alignment of Mce1R and its orthologs. (a) The fully conserved residues are marked by the * signs. The amino acid sidechains with highly similar properties are marked by identical colors. The respective organism names are shown on the left sides of the alignments. The secondary structure elements of Mce1R are shown above the alignment. At the C–terminal domain of Mce1R, residues conserved in the cavity are marked by the down arrows. The dark arrows are indicating fully conserved residues detected by the software. The light arrows are representing highly conserved residues with very less dissimilarities. The dark filled circles are indicating conserved residues at the N–terminal. (b) The down arrow marked conserved residue sidechains (except Arg148) in the cavity are shown.

cavity at the C–terminal domain and the residues constituting the cavity were– **Gln97**, Ala98, Ile100, **Ala101**, Gln102, Glu103, **Leu104**, **Ala105**, Thr106, ala108, **Glu143**, **Phe146**, **His147**, **Arg148**, Phe150, **Asn151**, **Leu163**, Gly181, Arg184, **Gly185**, Glu186, Ser188, Ser189, His200, Arg203, Asn204, Arg205, His207 and Arg208 (highly conserved residues are in **bold** front and marked by down arrows in the alignment; Fig. 4.2.9a). The predicted max pKd score is 9.04 and Drug score is 109; as computed by the server for the cavity which suggests that the cavity may bind a specific ligand with high affinity and/or suitable drug compound may be designed targeting this cavity. The sidechains of all those highly conserved residues (except Arg148) at the C–terminal domain are shown in the figure (Fig. 4.2.9b). Except Arg148, the sidechains of all conserved residues are oriented towards the interior of the cavity suggesting that some of these residues (if not all) might be playing critical roles in recognizing and binding of specific ligand.

Interestingly, some of those above–mentioned highly conserved residues forming the cavity in Mce1R, found to align with some of those residues of McbR and FadR of *E. coli*, critical for interacting with ligand. Fig. 4.2.4 shows the alignment between Mce1R and other structural orthologs including McbR (PDB code 4p9f) where the alignment of McbR–specific ligand–binding residues [158] with that of Mce1R can be observed. Similarly, the alignment of FadR–specific ligand–binding residues [209] with that of Mce1R can be observed in the previous work, reported by Casali et al. [104]. The residues Arg89 of McbR (in the alignment Arg81; Fig. 4.2.4) and Leu101 of FadR align with Gln97 of Mce1R; Glu93 of McbR (in the alignment Glu85) and Arg105 of FadR align with Ala101 of Mce1R; Asn135 of McbR (in the alignment Asn127) and Asp145 of FadR align with Glu143 of Mce1R; Arg139 of McbR (in the alignment Arg131) align with His147 of Mce1R and Leu165 of FadR align with Leu163 of Mce1R. To further validate this result, structures of FadR (PDB id 1hw1) and McbR (PDB id 4p9f) have been analyzed similarly using the CavityPlus server which has also identified the same cavities, as determined experimentally in their C–terminal domains [158, 209]. The predicted max pKd value and Drug score for FadR was 10.84 and 248 and for that of McbR, were 10.47 and 1557 respectively. Notably, the pKd values of the cavities of FadR and McbR are comparable to that of Mce1R. Taken together, strong correlations among some of the ligand–binding residues of McbR and FadR with some of the conserved cavity–forming residues of Mce1R and comparable pKd values strongly suggest the possible roles for those conserved residues of Mce1R in ligand–binding. Interestingly, the Drug score for McbR was found to be significantly higher than that of Mce1R and FadR. A possible

explanation for this would be that the structure of McbR (PDB id 4p9f) used for cavity analysis, has been described as ligand–bound conformation [158] whereas the FadR structure (PDB id 1hw1) was not. We also have noticed that for ligand–bound structure of FadR (PDB id 1h9g), the Drug score was 773 while the max pKd value was 9.71. Therefore, it seems likely that the modeled structure of Mce1R is not a ligand–bound conformation. However, roles played by the other conserved residues (not marked by arrows) of the C–terminal domain are not clear. As many of them are hydrophobic in nature, we speculate that they might be important to preserve the overall structure of the C–terminal domain since hydrophobic interaction plays an important role in protein folding and stability.

Section III

4.3 Identification of the specific ligand(s) for Mce1R

4.3.1 Objective C1- Identification of putative ligand(s) for Mce1R by cavity similarity search method using the ProBis server

Generally, protein cavities, binding to ligands of similar sizes and physicochemical properties, share similarity in geometric shapes and physicochemical environments of their cavities although many times they do not share common sequence patterns [210–212]. Therefore, it is possible to identify the cavities of different proteins; those are structurally and functionally similar (binding to similar type of ligands) to a given cavity by searching against a non–redundant database of protein structures (and protein–ligand complexes) using binding site (cavity) matching algorithms [213]. Simultaneously, it is also possible to identify the potential ligands; those may bind to the given protein cavity, previously bound to the similar cavities of different protein structures. To identify the protein cavities similar to that of Mce1R and potential ligand(s) which may bind to Mce1R, the ProBis server [140] was used. The server detected the protein structures carrying the cavities which were structurally and functionally similar to that of Mce1R by comparing against a non–redundant database of protein structures. The level of such similarity between two protein cavities was computed by the server in terms of the confidence score (Z score) considering the geometric similarities of the cavities and physicochemical similarities among the corresponding amino acid residues constituting the cavities [214]. Existing ligands bound to the cavities of such identified protein structures, were then transposed to the cavity of Mce1R by the ProBis server to assess the possibilities of occurrence of specific or non–specific interactions between the ligands and the amino acid residues forming the cavity of Mce1R. Depending on the cavity confidence scores and nature of the possible interactions between such ligands and the cavity of Mce1R, the server has listed the possible ligands (both specific and non–specific) following the descending order of confidence scores. ProBis has also identified four different binding sites (binding sites 2–5) in Mce1R in addition to the primary binding site (binding site 1) which has been designated as the cavity in the section 4.2.5. However, the minor binding sites (sites 2–5) displayed only non–specific interactions with very less number of the ligands and therefore these were not considered for further studies. The result showed that Mce1R cavity has been compared with total 4448 cavities of many different protein structures (Supplementary file 1; sheet 1 [238]) and a large no of ligands might be

specifically or non–specifically interacting with Mce1R (Supplementary file 1; sheet 2 [238]). The ligands predicted to specifically interact within the cavity are displayed in the Table 4.5. Interestingly, all the putative ligands are found to be various types of lipids suggesting that the cavity of Mce1R is primarily lipid–specific in nature. Further, to reduce the possible false–positive ligands for Mce1R, the localizations and functions of those lipids and their corresponding protein receptors have been explored. It is found that except the ligands; PLM [215, 216], MYR [217], SVR [218, 219] and SEF [220], the remaining ligands are localized at the biological membranes where they perform specific cellular functions as complexes with their respective receptors [221–226]. Being a transcriptional regulator, Mce1R functions in the cytosol; not at the membrane neither it is localized at the membrane. Therefore it is not possible for the remaining molecules to act as the specific ligands for Mce1R. Moreover, the ligands; SVR and SEF are synthetic molecules [218, 220]; hence are not present naturally in the cytosol of *M. tuberculosis*. Therefore, SVR and SEF cannot be the natural ligands for Mce1R.

Multiple lines of evidence reported that *M. tuberculosis* utilized host–derived fatty acids and cholesterol as sole source of carbon and energy during infection [85, 227–229]. The granuloma (necrotic lesion at the lungs of the patients with pulmonary tuberculosis; where the pathogen resides in) environment is rich in fatty acids [230]. The sputum of the tuberculosis patients also reported to contain cholesterol, palmitic, stearic and oleic acids as the lipid components [231, 232] surrounding the pathogen. Import of palmitic and oleic acids by *M. tuberculosis* was greatly reduced when the Mce1 transporter was not present (*Δmce1* mutant of *M. tuberculosis* strain) [86]. The *mce1* operon was also observed to be transcriptionally de–repressed in the media supplemented with palmitic [233] or oleic acids [104]. Considering the prediction made by the ProBis server and several scientific reports mentioned above, it seems possible that the fatty acids (palmitic, myristic, stearic and oleic acids) might act as specific ligand(s) for Mce1R. Although stearic and oleic acids were not predicted to be potential ligands by the ProBis server, it was sought to test their compatibilities as ligands for Mce1R along with palmitic and myristic acids by molecular docking approach.

Table 4.5 Ligands binding specifically within the cavity of Mce1R predicted by ProBis

Sl. No.	Ligand name	PDB ID of ligand	PDB ID of the protein bound with the ligand	Confidence score
1.	Protoporphyrin IX containing Fe	HEM	3ZSN	1.71
2.	(1s,8e)-1-{[(2s)-1-hydroxy-3-{[(1s)-1-hydroxypentadecyl]oxy}propan-2-yl]oxy}heptadec-8-en-1-ol	2WA	4OGQ	1.66
3.	Palmitic acid	PLM	2NNJ	1.56
4.	N-[(1s,2s)-2-hydroxy-1-({[(2r,3r,4s,5s,6r)-3, 4,5-trihydroxy-6-(hydroxymethyl)tetrahydro-2h-pyran-2-yl]oxy}methyl)octadecyl] octadecanamide	GM2	1GZP	1.47
5.	Myristic acid	MYR	2K4I	1.39
6.	8,8'-[carbonylbis[imino-3,1-phenylenecarbonylimino(4-methyl-3,1-phenylene)carbonylimino]]bis-1,3,5-naphthalenetrisulfonic acid	SVR	2NYR	1.29
7.	Ethyl (r)-{10-[(hept-6-yn-1-ylcarbamoyl) oxy]decyl}phosphonofluoridate	SEF	4JLL	1.18
8.	Rhodopin glucoside	RG1	2FKW	1.13
9.	Chlorophyll a	CLA	4KT0	1.13
10.	Peridinin	PID	2C9E	1.12

4.3.2 Objective C2- Molecular docking analysis using those ligand(s) by AutoDockVina

Molecular docking was performed using the AutoDock Vina [141] software to test the compatibilities of the fatty acids (myristic, palmitic, oleic and stearic acids) as ligands targeting the cavity sites of the dimeric Mce1R molecule. The best docking scores corresponding to the ligand poses assigned by the docking software are shown in the Table 4.6. Again AutoDock4 software was used to perform docking with the fatty acid ligands on Mce1R structure to revalidate the results obtained from AutoDock Vina. The mean binding energies computed by AutoDock4 closely match with the docking scores assigned by AutoDock Vina which indicates reliability of the docking scores computed by AutoDock Vina for Mce1R–fatty acid ligands. The data form the Table 4.6 indicated that while the docking scores for the other ligands were highly similar, myristic acid acted as slightly more preferred ligand than the other ligands. However, all the ligands were found to dock within the cavities of the dimeric Mce1R (Fig. 4.3.1a–d).

Table 4.6 Docking scores of the fatty acid ligands

Sl No.	Ligands	Docking score* (kCal/mol)	Mean binding energy† (kCal/mol)
1	Myristic acid	–6.1 (Chain A & B)	–6.08 (Chain A & B)
2	Palmitic acid	–5.8 (Chain A)	–5.82 (Chain A)
		–5.9 (Chain B)	–5.75 (Chain B)
3	Oleic acid	–5.8 (Chain A)	–5.84 (Chain A)
		–5.7 (Chain B)	–5.55 (Chain B)
4	Stearic acid	–5.8 (Chain A)	–5.65 (Chain A)
		–5.7 (Chain B)	–5.44 (Chain B)

* Docking scores obtained from AutoDock Vina

† Mean binding energy values obtained from first ranked clusters computed by AutoDock4

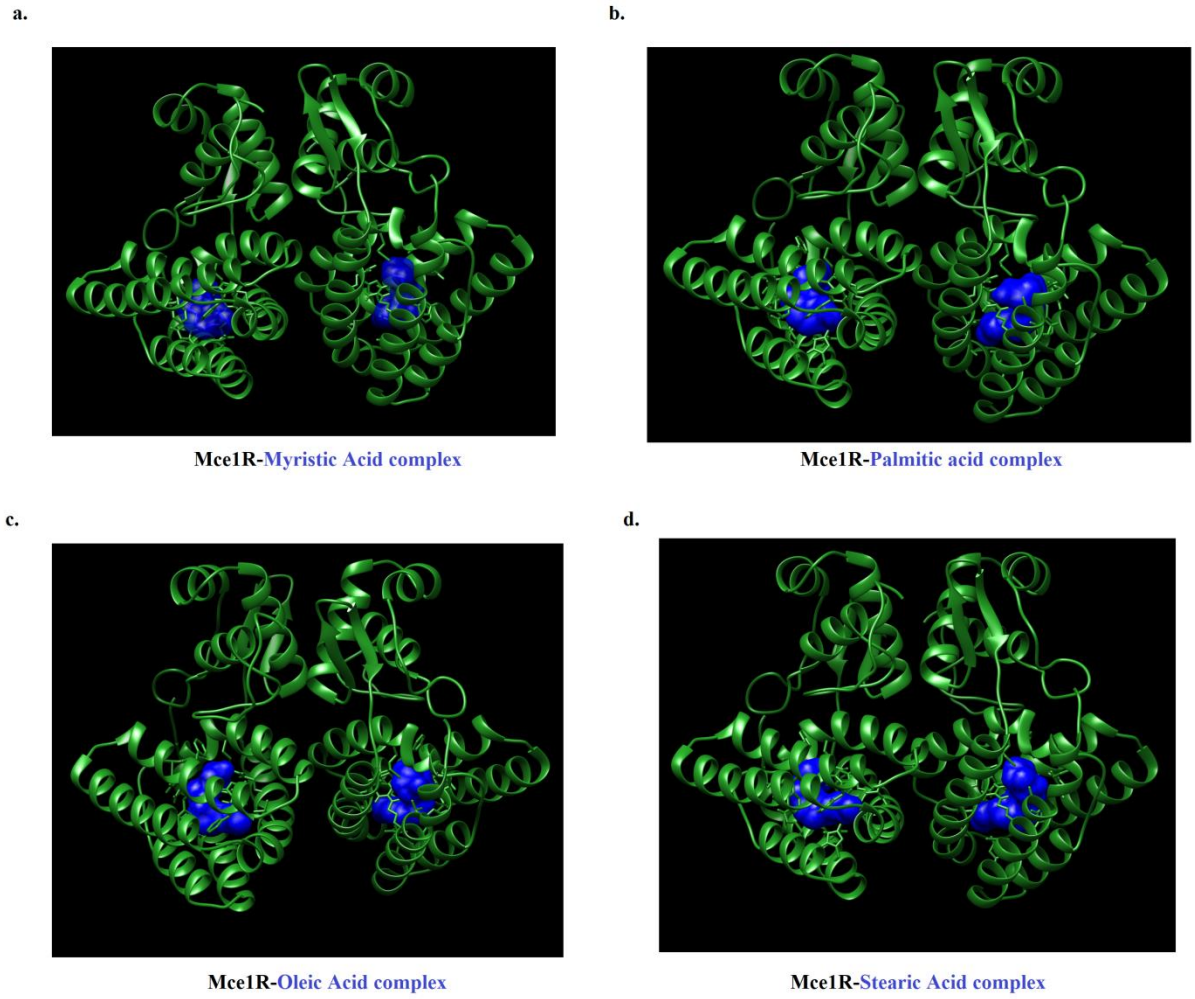


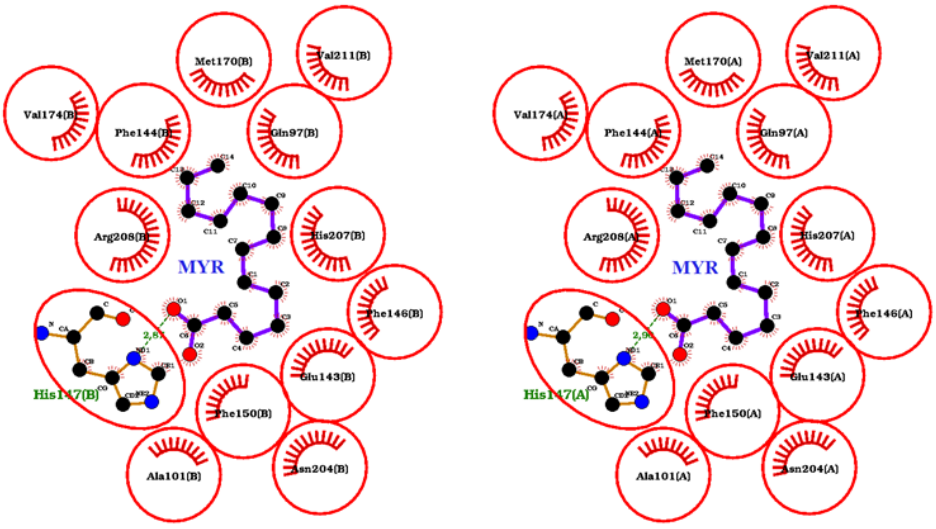
Fig. 4.3.1 Dimeric Mce1R complexed with fatty acids. (a) Myristic (b) Palmitic, (c) Oleic and (d) Stearic acids are docked within the same cavities of Mce1R. The docked ligand structures within the cavities of the CTDs of Mce1R and the ligand named are colored in blue. The ligands are represented as molecular surfaces for better understanding. Only the best docked poses (top score complexes) of the fatty acid ligands are shown here.

4.3.3 Objective C3- Analyzing the interactions between the ligand(s) and the cavity residues of Mce1R by the LigPlot software

The docked Mce1R–fatty acid complexes were further analyzed using the LigPlot^{+} v2.2 software [142]. The figures Fig. 4.3.2a and b; Fig. 4.3.3a and b show the non–bonded interactions between the cavity residues and ligands. For all the ligands, the major stabilizing forces were found to be hydrophobic interactions, mediated by several number of cavity residues. For a particular ligand, the common interacting cavity residues in both the subunits of Mce1R structure are encircled. A few particular residues were found to make interactions with the fatty acids, which are– Gln97, Ala101, Leu104, Ala105, Glu143, **Phe144**, Phe146, His147, Phe150, **Ala167**, **Met170**, **Val174**, **Arg177**, Asn204, His207, Arg208 and **Ser211**. The bold front residues are not cavity–forming but mediating interactions with the ligands while the remaining residues are cavity–forming among which the red colored residues already identified to be highly conserved among various VanR–type proteins in section 4.2.5. Interestingly, the conserved residues Gln97, Ala101, Glu143 and His147 of Mce1R have been shown to align with Arg81, Glu85, Asn127 and Arg131 of McbR respectively (Fig. 4.2.4), which had been implicated to interact with ligand [158], suggesting that the interactions observed between these conserved residues of Mce1R with the fatty acids, most possibly do occur in reality. Notably, almost all the interacting residues are involved in mediating hydrophobic interactions with the ligands, suggesting that the cavity environment is primarily hydrophobic in nature. However, a few residues– Gln97, Glu143, His147, Arg177 His207 and Arg208 could also form hydrogen bonds with the carboxyl terminals of the fatty acids besides making hydrophobic interactions (Fig. 4.3.2a and b; Fig. 4.3.3a and b). Such types of hydrogen bond interactions contribute to the overall stability of Mce1R–fatty acids complexes and also to the orientations of the fatty acid ligands during forming the complexes. Although being predominantly hydrophobic in nature, the cavity could specifically recognize the fatty acids (myristic, palmitic, oleic and stearic acids) as the ligands but not cholesterol; because cholesterol could not dock within the cavity (data not shown) although it was found to be present as one of the lipid components of the environment surrounding *M. tuberculosis* in the host system [231, 232].

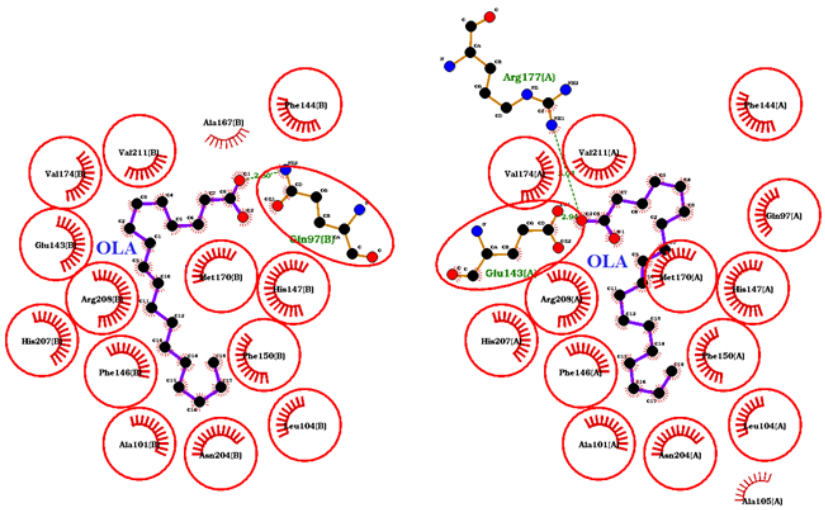
Notably, the docking scores for the Mce1R–fatty acids interactions (–5.7 to –6.1 kCal/mol; Table 4.6) appear to be less. To see whether naturally occurring fatty acid–protein complex also represents similar docking score or not, it was sought to choose the fatty acid–protein complex whose ligand binding site was similar to that of Mce1R and re–dock

a.



Mce1R-Myristic acid complex

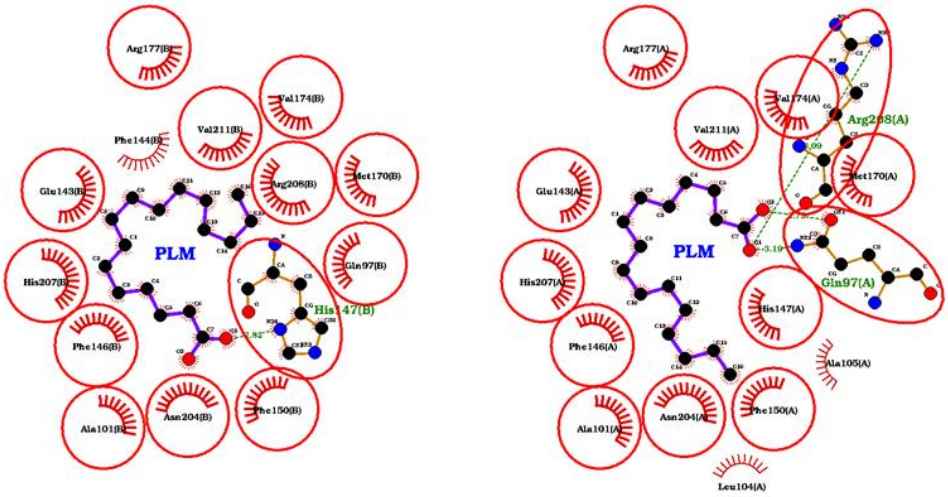
b.



Mce1R-Oleic acid complex

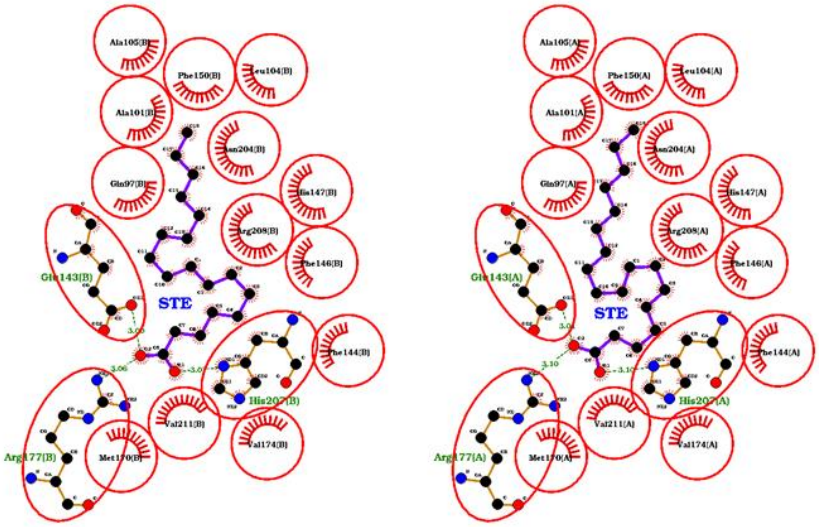
Fig. 4.3.2 Analysis of Mce1R–fatty acids (myristic and oleic acids) interactions. (a and b) Analyses of the docked complexes were performed by the LigPlot^{+} v2.2 [142] software. "MYR" and "OLA" indicate the docked myristic and oleic acids ligands respectively. Hydrogen bond interactions between ligands and residues are shown in green dotted line. Other residues are mediating hydrophobic interactions with the ligands. Interacting residues common to both subunits of Mce1R are encircled.

a.



Mce1R-Palmitic acid complex

b.



Mce1R-Stearic acid complex

Fig. 4.3.3 Analysis of Mce1R–fatty acids (palmitic and stearic acids) interactions. (a and b) Analyses of the docked complexes were performed by the LigPlot^{+} v2.2 [142] software. "PLM" and "STE" indicate the docked palmitic and stearic acids ligands respectively. Hydrogen bond interactions between ligands and residues are shown in green dotted line. Other residues are mediating hydrophobic interactions with the ligands. Interacting residues common to both subunits of Mce1R are encircled.

the fatty acid to its specific protein cavity following the docking protocol mentioned in the "Materials and methods" section 3.2.3.9 and then to compare the docking score with that of Mce1R–fatty acid complex. From Table 4.5, it was observed that palmitic acid was a natural ligand for cytochrome P450 2C8 (PDB id 2NNJ) whose ligand binding site (cavity) was similar to that of Mce1R. Almost similar docking score (–6.5 kCal/mol) was obtained for cytochrome P450 2C8 with its natural ligand palmitic acid [215] with nearly identical to the original bound pose, (data not shown). The reason for such less docking scores for Mce1R–fatty acid ligands could be the altered conformations of the cavities, as the cavities of Mce1R had been earlier described to be in a conformation not suitable for optimized ligand binding, i.e. not in a ligand–bound conformation (section 4.2.5). Although the fatty acid ligands could dock within the cavity of Mce1R and the ligand poses have been found to be stabilized by many non–covalent interactions mediated by several cavity residues, the stability and other dynamic properties of the docked complexes were explored following molecular dynamics simulation approach (see in the next section).

Section IV

4.4 Analyzing the stability and dynamic properties of the docked Mce1R–ligand complexes

4.4.1 Objective D1- All atom molecular dynamics simulation of Mce1R–ligand complexes using GROMACS

Objective D2- Analyzing RMSD, RMSF and Radius of gyration of Mce1R in those complexes

Objective D3- Analyzing ligand–induced dynamic changes in the secondary structure of Mce1R

All atom molecular dynamics simulations of Mce1R with and without ligands were performed for 200 ns duration. To detect structural perturbations occurring at the backbone of the protein molecule, the RMSD values were calculated for the protein backbone atoms throughout the simulation and plotted against simulation times (ns). The RMSD plots for each of the systems are shown in the Fig. 4.4.1. It is observed that without any ligands, Mce1R has displayed highest RMSD values compared to those with ligands, throughout the simulation, suggesting that ligand binding stabilizes the structure of Mce1R more than its unbound state. All the systems got stabilized from ~100 ns onward. The average RMSD values for Mce1R, Mce1R–Myristic acid, Mce1R–Oleic acid, Mce1R–Palmitic acid and Mce1R–Stearic acid complexes, considering 100–200 ns time points; after they got stabilized, determined to be 0.902, 0.634, 0.737, 0.547 and 0.732 nm respectively which indicated that binding of oleic and stearic acids to Mce1R caused relatively more structural perturbations compared to other two ligands. The ligand palmitic acid stabilized Mce1R most while the other ligand myristic acid provided intermediate stability to Mce1R. Although all protein–ligand complexes seem to be stable after ~100 ns, their RMSD plots have shown distinct features. Mce1R–Myristic acid complex displayed a sudden change around 100 ns time point and then fairly maintained a stable RMSD values with less fluctuation. Mce1R–Palmitic acid complex showed a bit changes in the RMSD values at ~40, ~60, ~80 and ~120 ns time points and then became stabilized. Although Mce1R–Oleic and Mce1R–Stearic acids displayed highly similar RMSD values from ~80 ns onward, at the time points ~40 to ~70 ns; Mce1R–Stearic acid complex showed bit more structural fluctuations than the Mce1R–Oleic

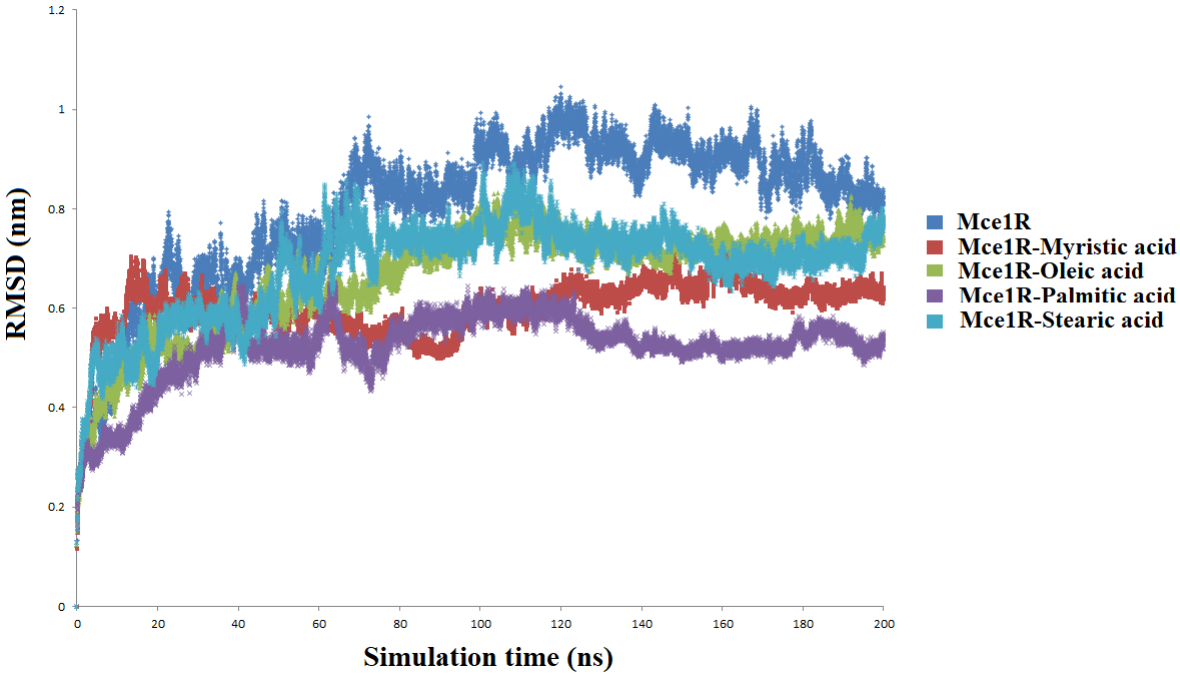


Fig. 4.4.1 RMSD plots of Mce1R and its fatty acids docked complexes. The RMSD values calculated in nm were plotted against simulation time (200 ns).

acid complex. Such distinctive changes in the RMSD values indicate conformational changes in the backbone of Mce1R mediated by binding with different fatty acid ligands. Although the fatty acid ligands are structurally and functionally similar, interestingly they could induce distinct changes to the structure of Mce1R. However, visual inspection of the trajectories of all the complexes by VMD [145] has revealed that the complexes were stable throughout the simulation.

To determine residue–specific average fluctuations in the protein structure during the simulation, the RMSF values of both the subunits (chains A and B) were plotted against each residue positions for Mce1R and Mce1R–fatty acid complexes (Fig. 4.4.2a&b). The plots showed that the unstructured N–terminal arms of protein and protein–ligand complexes, composed of first ~20 amino acid residues of both subunits had highest fluctuations (RMSF values ~0.2–1.8 nm) compared to the remaining regions of the protein. Interestingly, the N–terminal arm of the chain B of Mce1R showed more increased fluctuations at the amino acid positions ~5 to 19 (RMSF values ~0.46–1.24 nm) compared to chain A and other ligand–bound chains of Mce1R. The Mce1R structure also displayed distinctive enhanced flexibilities at other amino acid positions, ~55 to 83 (RMSF values ~0.24–0.47 nm) and ~100 to 121 (RMSF values ~0.17–0.51 nm) of the chain A compared to the ligand–bound Mce1R. In the chain B, amino acid positions ~176 to 184 (RMSF values ~0.27–0.51 nm) also displayed similar enhanced flexibilities compared to the ligand–bound Mce1R. Notably, the structural dynamic profiles of both subunits of dimeric Mce1R were not similar. Such regions of enhanced flexibilities made the total RMSD values higher for Mce1R than other Mce1R–fatty acid complexes. Interestingly, a few cavity residues of the chain A of Mce1R; Ile100, Ala101, Gln102, Glu103, Leu104, Ala105, Thr106 and Ala108 were found to lie within the amino acid positions 100–121 whereas, Gly181 and Arg184 of chain B were found to lie within 176–184 region, thus bearing a bit higher RMSF values. It therefore appears that structural dynamics of the cavities of both subunits are not similar. In contrast, binding of myristic, palmitic and oleic acids to Mce1R caused significant decrease in the overall fluctuation profiles of both the subunits, including the regions where those specific cavity residues (mentioned above) were located; suggesting that binding to these ligands stabilized the structure of Mce1R. Moreover, less RMSF values for those cavity residues suggest that during the simulation the ligands have mediated stable non–covalent interactions with those residues. On the other hand, binding of stearic acid to Mce1R mostly caused more fluctuations compared to other protein–ligand complexes, throughout the structure

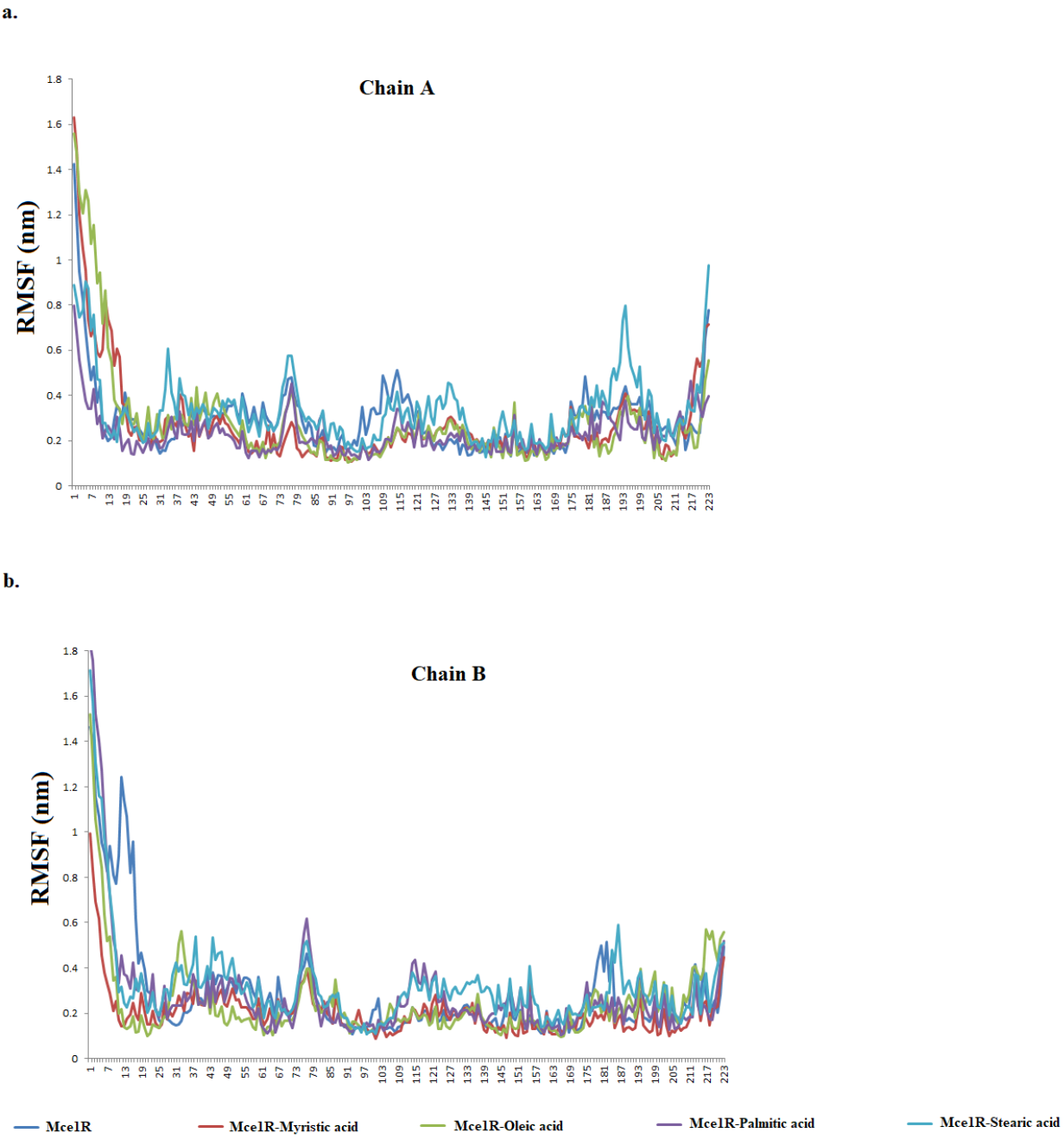


Fig. 4.4.2 RMSF plots of Mce1R and its fatty acid complexes. RMSF values of chain A (a) and chain B (b). The colors for Mce1R and its complexes are shown in the lower panel. The RMSF values are available in Maity et al. 2022; Supplementary file 12 [238].

of Mce1R which accounts for fluctuations at the RMSD values observed at the time points ~40 to ~70 ns during the simulation.

The RMSF plots of Mce1R and its ligand–bound forms point towards distinct structural changes in Mce1R, those might occur during the simulation. To detect such changes, the dynamics of secondary structures of Mce1R and its ligand–bound complexes have been analyzed using the *do_dssp* tool available in GROMACS. The result is shown in the figures (Fig. 4.4.3a–e). For all these figures, the amino acid residue numbers were plotted along the y axis and the simulation time along the x axis. In these plots the residue numbers of chain A were plotted from 1–223 and after that the residue numbers of chain B were plotted as consecutive numbers by the program. It was observed that the secondary structural elements of the N–terminal domains (α1–3; β1&2) of both subunits of all protein structures remained mostly stable throughout the simulation while in that of the C–terminal domains a few distinct changes have occurred due to binding of the ligands. The α4 helices of both subunits in the protein–ligand complexes became more stable compared to that of Mce1R during the simulation. The cavity residues of Mce1R– Ile100, Ala101, Gln102, Glu103, Leu104, Ala105, Thr106 and Ala108 were present within this α4 helices which were shown to have less RMSF values (Fig. 4.4.2). Except Mce1R–myristic acid complex (Fig. 4.4.3b), most of the residues at the C–terminals of α7 helices of chain A of the remaining systems have shown transitions from helix to 'bend' or 'turn' conformations during the simulation. The residues– Ala167, Met170, Val174 and Arg177; detected to mediate non–covalent interactions with the ligands (Fig. 4.3.2), found to be located within the C–terminals of the α7 helices, those were undergoing stable transitions from helix to 'bend' or 'turn' conformations, similar to that of Mce1R without any ligand, suggesting that these interactions are also important for binding to the ligands. Although they are not cavity–forming, they might help to stabilize the protein–ligand complexes. At the B chains also most of the residues of the α7 helices for all the protein–ligand complexes have undergone such stable transitions. Interestingly, the cavity– residues– Arg184, Gly185, Glu186, Ser188 and Ser189, located at the α8 helices of the chain A of Mce1R–oleic and palmitic acid complexes (Figs. 4.4.2c & d) went through similar types of transitions, suggesting that these residues also mediate non–covalent interactions with the ligands during the simulation although these residues remained undetected initially during docking analysis. Involvement of the α8 helices of the chain A in binding to oleic and palmitic acids seems to be specific to these two ligands as such transitions were not

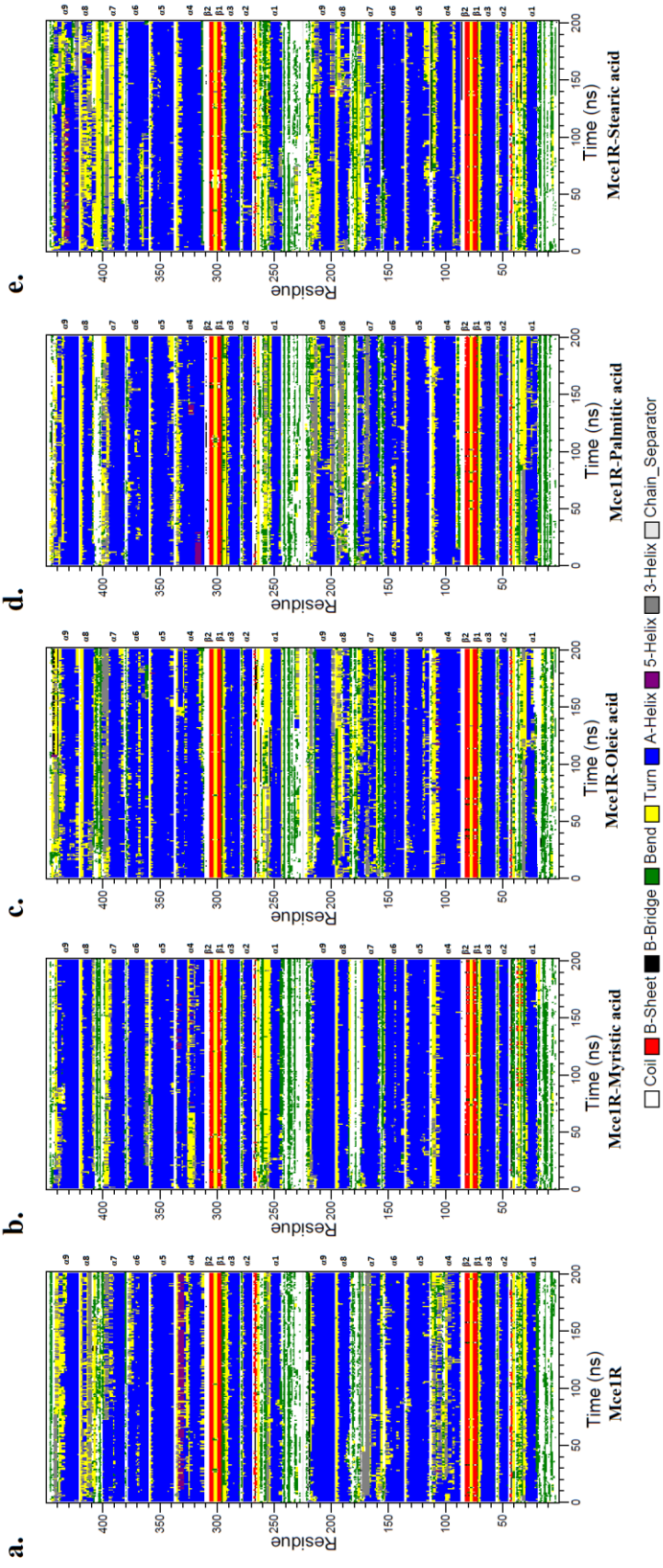


Fig. 4.4.3 Dynamic changes in the secondary structure profile of Mce1R induced upon ligand binding. (a–e) Myristic, oleic, palmitic and stearic acid induced changes in the secondary structure of Mce1R are shown. The secondary structural elements of Mce1R (α1–9; β1&2) of both subunits are shown in the side panel. The specific colors for different structures are shown in lower panel.

observed for myristic and stearic acid complexes (Figs. 4.4.3b & e). On the other hand the α8 helix of the chain B of Mce1R–stearic acid complex was observed to undergo such conformational transitions thus might be involved in binding stearic acid specifically (Fig. 4.4.3e). Except specific ligand–induced conformational changes, the other secondary structural elements of Mce1R were found to be mostly stable which also suggest that the systems (Mce1R–ligand complexes) were stable throughout the simulation.

To assess the effects of such ligand–induced changes in the secondary structure of Mce1R on its overall structure, average radius of gyration (R_{gavg}) of all the systems were estimated. Radius of gyration (R_{gavg}) represents the compactness of a protein structure during simulation by measuring the average distance between the center of mass of the protein molecule and its terminal over the course of simulation. Lower R_{gavg} value indicates more compact whereas higher value indicates expanded or unfolded structure. The R_{gavg} values for Mce1R, Mce1R–myrictic acid, Mce1R–oleic acid, Mce1R–palmitic acid and Mce1R–stearic acid complexes were found to be 2.35, 2.35, 2.37, 2.29 and 2.40 nm respectively. It therefore appears that there is negligible effect on the overall structure of Mce1R due to binding of myristic acid. Binding of palmitic acid caused more compactness while stearic acid caused expansion in the structure of Mce1R upon binding. Binding of oleic acid also expanded the structure of Mce1R a bit. Taken together, it seems that except myristic acid, all the other ligands caused significant perturbations in the structure of Mce1R upon binding.

4.4.2 Objective D4- Analyzing ligand stabilities and dynamics of hydrogen bonding interactions with the cavity of Mce1R

In order to form a stable complex with Mce1R, the ligands should bind to it in an optimum pose where its structural deviations would be minimized. To detect structural deviations in the ligands bound with Mce1R, RMSD plots for the ligands for each cavity sites were obtained from the simulation trajectory files (Figs. 4.4.4a and 4.4.5a). At the cavity of chain A of Mce1R, the RMSD values of all the ligands became stable from ~90 ns time points and maintained almost similar RMSD values, whereas stearic acid maintained relatively higher RMSD values throughout the simulation (Fig. 4.4.4a). At the cavity site of chain B of Mce1R, the ligand oleic acid has undergone some conformational switching at ~100 ns time point and then became stabilized. All the other ligands were mostly stable during the simulation (Fig. 4.4.5a).

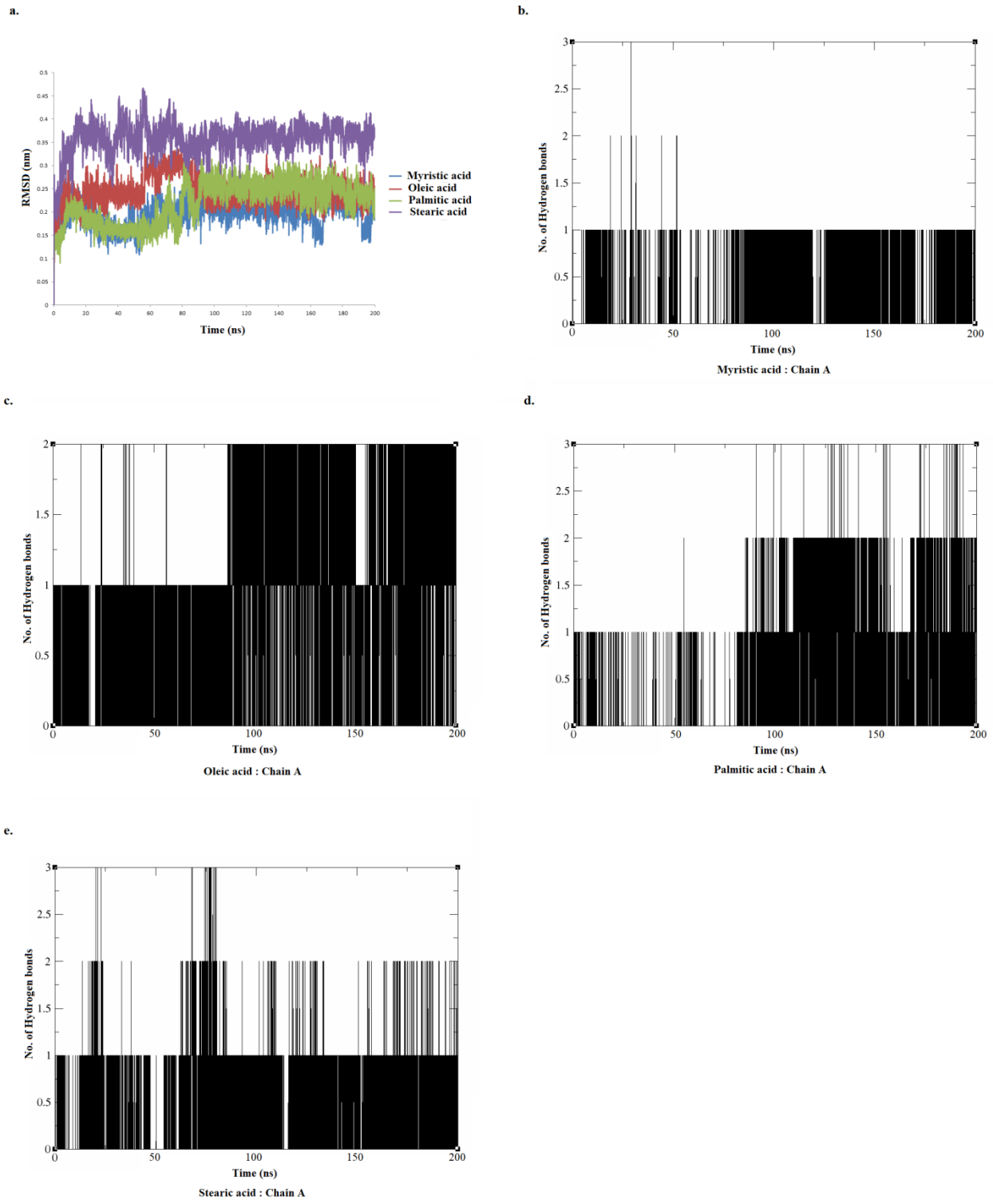


Fig. 4.4.4 Mce1R–ligand interactions at the subunit A. (a) RMSD analysis of the ligands in the complexes. (b–e) Dynamics of hydrogen bond formation between ligands and cavity residues of subunit A of Mce1R.

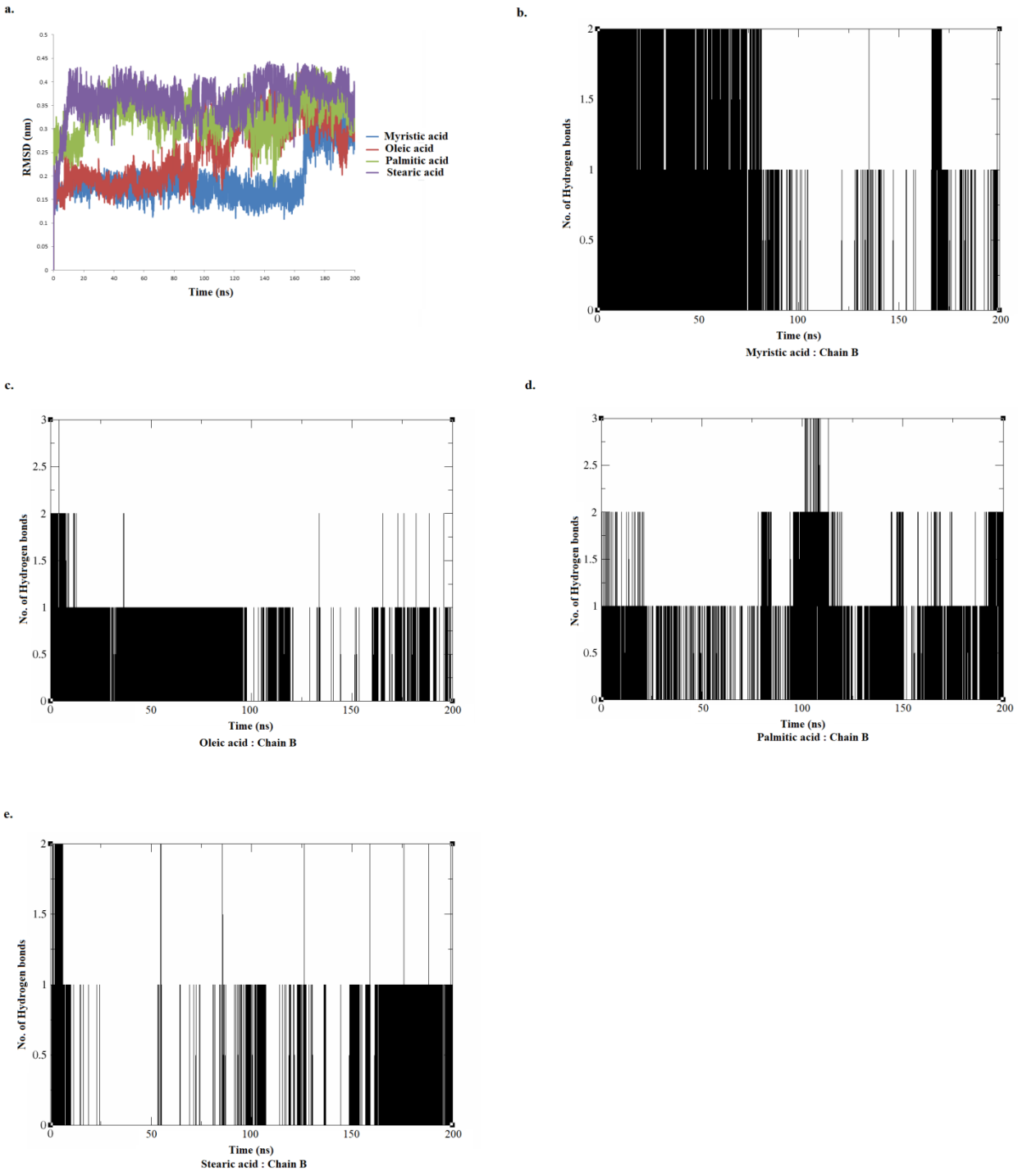


Fig. 4.4.5 Mce1R–ligand interactions at the subunit B. (a) RMSD analysis of the ligands in the complexes. (b–e) Dynamics of hydrogen bond formation between ligands and cavity residues of subunit B of Mce1R.

A ligand stably bound to a receptor molecule, participates in several non–covalent interactions with the amino acid side chains of the ligand–binding site (cavity) of the receptor. Hydrogen bond interaction is such type of non–covalent interaction considered to be important for both stability of the formed complex and proper orientation of the bound ligand within the cavity. The hydrogen bond interactions throughout the simulation have been analyzed using the VMD [145] software and the results are shown in the figures 4.4.4 & 4.4.5 b–e. At the cavity of chain A of Mce1R, oleic and palmitic acids formed single hydrogen bonds up to ~100 ns and maintained double hydrogen bonds thereafter till 200 ns (Figs. 4.4.4c & d). As shown in the figure 4.4.4a, at ~100 ns time point, both oleic and palmitic acids had undergone conformational switching (fluctuation in RMSD values) which possibly resulted in formation of double hydrogen bonds with the protein molecule and thereby more stabilizing the complexes. Stearic acid mostly formed single hydrogen bond with transient double hydrogen bonds throughout the simulation (Fig. 4.4.4e) while, myristic acid predominantly formed single hydrogen bond (Fig. 4.4.4b). At the cavity of the B chain of Mce1R, oleic acid predominantly formed single hydrogen bond and palmitic acid formed mostly single and intermittent double to triple hydrogen bonds during the simulation (Figs. 4.4.5c & d). Stearic acid formed mainly single hydrogen bond (Fig. 4.4.5e) while myristic acid maintained double hydrogen bonds initially up to ~80 ns and thereafter single hydrogen bond (Fig. 4.4.5b).

4.4.3 Objective D5- Determination of binding free energy for those ligand(s) following MMGBSA approach

All the fatty acid ligands have shown to form distinct pattern of hydrogen bond interactions and hydrophobic interactions with the residues within and near the cavities of Mce1R, revealed by the molecular docking analysis (Fig. 4.3.2). The docking scores have been assigned for such types of protein–ligand interactions in the form of binding energies (kCal/mol; Table 4.6). However, form molecular dynamics simulation studies; it was observed that a few residues from the α8 helix might be involved in interacting with the ligands, whose contributions were not considered by the docking software earlier. Moreover, the docking softwares generally use the scoring functions those bear many approximations to calculate various components of the binding free energy in order to speed up the docking processes [234]. Compared to many docking softwares and other methods (such as free

energy perturbation or thermodynamic integration methods [235]) to calculate ligand binding free energies, ensemble–averaged MMGBSA method has gained wide popularity because of its forcefield based method which provides fair accuracy at reasonable computation cost [236]. Therefore, to accurately determine the binding free energies for all the fatty acid ligands with the both subunits of Mce1R, ensemble–averaged MMGBSA method was employed using the gmx_MMPBSA tool [146]. Total 41 snapshots were taken at 1.25 ns interval from the complex trajectories from 150 ns to 200 ns time points as all the complexes were mostly stable (Fig. 4.4.1). The binding energies computed by MMGBSA method are displayed in the Table 4.7. The data displayed that binding free energies for the ligand myristic acid to both the subunits of Mce1R are lowest among that of the other ligands, suggesting that stability of the Mce1R–myristic acid complex is lowest compared to other complexes. For both the subunits of Mce1R, the ligand palmitic acid has showed similar binding free energies. The ligand stearic acid also showed similarity in the binding free energies to the both subunits of Mce1R. Oleic acid showed highest binding free energy for the cavity of the A subunit than the B subunit. The reason for this large difference in the binding free energies for oleic acid could be either the presence of double bond within its structure which has restricted some of the conformational flexibilities of the ligand or the difference in cavity dynamics between the subunits of Mce1R, or both. Interestingly, although docking scores for myristic acid were highest (Table 4.6), MMGBSA method–based binding free energies for this ligand are found to be lowest. It is observed that Van der Wall energies for all the ligands have contributed more to the favorable binding free energies compared to other energy components (Table 4.7), suggesting that hydrophobic interactions played most important roles to stabilize the complexes during the simulation [237]. Notably, the stabilities of the complexes found to be increased if the chain lengths of the ligands are increased (at least more than that of myristic acid) which indicates that Mce1R forms more stable complexes preferably with long–chain fatty acids.

Notably myristic acid was considered initially as a possible natural ligand for Mce1R since it was identified by the ProBis server, although it is not naturally present as one of the lipid components surrounding *M. tuberculosis* [230–232]. On the other hand, oleic and stearic acids were not been identified by the ProBis server as potential ligands; probably because of the particular conformation of the cavity structure of Mce1R or the transposition process for the ligands used by the server or both. Therefore, inclusion of myristic, oleic and stearic acids

Table 4.7 Binding free energies of fatty acid ligands to Mce1R by MMGBSA method

Ligand	Cavity of subunit	ΔE_{VDW} kCal/mol	ΔE_{ELE} kCal/mol	ΔG_{GB} kCal/mol	ΔG_{SASA} kCal/mol	ΔG_{bind} kCal/mol
Myristic acid	A	–20.61±17.71	–13.78±12.15	17.51±15.12	–3.18±2.69	–20.07±17.06
	B	–33.43±6.09	–4.05±6.92	15.76±6.32	–4.99±0.83	–26.71±4.73
Oleic acid	A	–53.85±12.55	–19.83±5.33	24.13±5.67	–7.64±1.74	–57.21±13.38
	B	–39.06±20.79	–6.29±5.48	18.79±10.70	–5.59±2.98	–32.15±17.21
Palmitic acid	A	–33.58±16.75	–28.06±18.92	33.65±18.57	–5.18±2.57	–33.19±17.13
	B	–31.11±16.88	–19.94±16.84	25.61±16.95	–4.93±2.64	–30.38±16.79
Stearic acid	A	–42.24±7.47	–24.24±5.88	29.36±5.77	–6.54±1.10	–43.66±7.61
	B	–47.86±2.26	–18.32±5.14	28.68±4.44	–7.11±0.24	–44.61±2.48

in this study helped us to perform a comparative analysis for the probable fatty acid ligands of various chain–lengths. The dimeric structure of Mce1R was needed because during molecular dynamics simulation it was essential to use the biologically active molecule in order to mimic various interactions those might occur between the protein and ligand(s). Indeed the gel-filtration chromatography also suggests that Mce1R exists as dimer in solution thus experimentally validating the modeled dimeric structure. Importantly, it was also observed that the dimeric Mce1R structure was stable during the molecular dynamics simulation. Simulation studies of the docked complexes showed that while oleic, palmitic and stearic acids caused distinct perturbations in the overall structure of Mce1R, the effects of binding of myristic acid were negligible to cause any significant change in the structure of Mce1R. This indicates that to cause significant changes in the structure of Mce1R upon ligand–binding (which is a critical characteristic of VanR or GntR–type proteins [113, 114]), chain length is an important factor. Additionally, the ensemble–averaged binding free energy analysis of the docked complexes by the MMGBSA method also indicates that Mce1R–myristic acid complex was least stable compared to the other complexes with ligands of longer chain lengths. Taken together, all these data suggest that fatty acid ligands with chain lengths higher than that of myristic acid are most preferred natural ligands for Mce1R.

CHAPTER 5

CONCLUSIONS

AND

FUTURE PROSPECTS

5 Conclusions

5.1 Section I

The genomic DNA from *M. tuberculosis* H37Ra has been successfully isolated with high integrity and purity which is the primary requirement for the subsequent PCR amplification and molecular cloning. The operator DNA (*mce1* promoter) and *mce1R* ORF were successfully cloned in suitable vector without any mutation(s). Mce1R has been adequately expressed under the laboratory condition and purified to homogeneity. The purified Mce1R is shown to specifically bind with the operator DNA with moderate affinity compared to other VanR–type regulators which is necessary to maintain intracellular concentration of Mce1R sufficient to regulate the expression of several genes.

5.2 Section II

The structure of Mce1R has been modeled and validated. Mce1R is found to possess VanR–type structure with an unstructured arm at the N–terminal of ~19 amino acid residues. The Mce1R structure is composed of N– and C–terminal domains. An winged HTH motif is present at the N–terminal domain while the C–terminal domain contains a bundle of six α–helices. Coarse grain molecular dynamics simulation studies suggest that the N–terminal domain is more structurally dynamic than the C–terminal domain and such mode of structural dynamics is most likely similar in other structural homologs. Despite significant sequence conservancy at the N–terminal domain of Mce1R, the residues important for DNA–binding are not completely conserved which might be the reason for its moderate DNA–binding affinity, supported by in–vitro experiment using purified Mce1R. Additionally, the N–terminal flexible arm of Mce1R harbors a few highly conserved positively charged residues, those might be important for binding with DNA and the C–terminal domain contains a cavity where a few residues are highly conserved and most possibly involved in binding to specific ligand(s).

5.3 Sections III and IV

From the cavity similarity search by the ProBis server, it was observed that the cavity of Mce1R was primarily responsive to lipids, among which the fatty acids (palmitic, myristic, stearic and oleic acids) could be the potential ligands. The molecular docking analysis performed between Mce1R and the fatty acid ligands suggested that the formation of Mce1R–fatty acid complexes were possible. Interestingly, in addition to the conserved cavity

residues, several other residues located near the cavity mediated mostly hydrophobic interactions with the fatty acid ligands. All atom molecular dynamics simulation of the docked complexes showed that while oleic, palmitic and stearic acids caused distinct perturbations in the overall structure of Mce1R, the effects of binding of myristic acid were negligible to cause any significant change in the structure of Mce1R. Additionally, the ensemble–averaged binding free energy analysis of the docked complexes by the MMGBSA method also indicated that Mce1R–myristic acid complex was least stable compared to the other complexes with ligands of longer chain lengths. Taken together, all these data suggest that fatty acid ligands with chain lengths higher than that of myristic acid are most preferred natural ligands for Mce1R.

5.4 Future prospects

To understand Mce1R–mediated regulation of *mce1* operon in more detail it is important to characterize the *mce1* promoter (determining the +1, –10 and –35 regions). Also the specific binding site of Mce1R on the *mce1* promoter DNA needs to be identified following experimental methods (e.g. DNA footprinting). The roles of the conserved amino acid residues in maintaining structure and function of Mce1R is necessary to explore by genetic and biochemical experiments. The fatty acid induced changes in the structure and function of Mce1R needs to be validated by experimental methods.

APPENDICES

6 Appendices

6.1 Media, different buffer and solutions compositions

Media and solutions	Composition
Middlebrook 7H9 broth	7H9 broth base 0.47%, Glycerol 0.4%, OADC 10%
Middlebrook Agar	Middlebrook 7H9 broth with 1.5% Agar
LB Broth	0.5% NaCl, 0.5% Yeast extract, 1% Tryptone
LB agar	1.5% agar in LB broth
TE Buffer	10 mM Tris-Cl buffer pH 8.0, 1 mM EDTA pH 8.0
2XTBE Buffer 500 ml	10.8g Tris base, 5.5g Boric acid, 4 ml of 0.5 M EDTA pH 8.0
Acrylamide – bis – acrylamide solution	29% (w/v) solution of acrylamide and 1% (w/v) solution of N. N'-Bis-acrylamide
Resolving buffer	1.5 M Tris-Cl, 0.4% (w/v) SDS, pH 8.8
Stacking buffer	0.5 M Tris-Cl, 0.4% (w/v) SDS, pH 6.8
6X DNA gel loading Dye	0.25% Bromophenol Blue, 30% Glycerol pH adjusted to 8.0 with NaOH
4X Protein Gel loading Dye	200 mM Tris-Cl pH 6.8, 8% SDS, 0.4% Bromophenol Blue, 40% Glycerol
Staining Solution	0.25% (w/v) Commassie Brilliant Blue R 250, 50% Methanol, 10% Acetic acid
Destaining Solution	20% Methanol, 10% Acetic acid
Bradford Reagent (5X)	Per 200 ml: 100 mg Coomasiae Brilliant Blue G-250 in 50 ml 95% ethanol, 100 ml 85% phosphoric acid.
EtBr solution	1% Ethidium bromide
Transfer Buffer	39 mM glycine, 48 mM Tris base, 0.037% SDS, 20% methanol.
TBS	10 mM Tris-Cl, 150mM NaCl, pH 7.5
TBS-T	TBS with 0.1% (v/v) Tween 20.

Developer (Primer)	As per manufacturer's instruction
Fixer (Primer)	As per manufacturer's instruction
TGS	25 mM Tris-Cl, 250 mM Glycine (pH 8.3), 0.1% SDS.
10X Proteinase K buffer	100 mM Tris–Cl, 50 mM EDTA, 500 mM NaCl, pH 8.0
Cell lysis buffer	25 mM TAPS, pH 9.0, 400 mM NaCl, 5% glycerol
Wash buffer I	25 mM TAPS, pH 9.0, 300 mM NaCl, 5% glycerol, 3 mM imidazole
Wash buffer II	25 mM TAPS, pH 9.0, 300 mM NaCl, 5% glycerol, 9 mM imidazole
Elution buffer	25 mM TAPS, pH 9.0, 300 mM NaCl, 5% glycerol, 0.1% sarcosyl, 300 mM imidazole
Dialysis buffer	50 mM Na–Phosphate, pH 8.0, 300 mM NaCl, 5% glycerol, 0.05% sarcosyl
Solution I	50 mM D-Glucose, 25 mM Tris-Cl, pH 8.0, 10 mM EDTA
Solution II	0.2 M NaOH, 1% SDS
Solution III 100 ml	5 M KOAc 60 ml, Glacial acetic acid 11.5 ml, Water 28.5 ml
STE Buffer 10 ml	10 mM Tris-Cl pH 8.0, 0.1 M NaCl, 1 mM EDTA

6.2 RMSF values from the CABS-Flex 2.0 server

Residues	**10 ns**	**20 ns**	**30 ns**	**40 ns**	**50 ns**	**60 ns**	**70 ns**	**80 ns**	**90 ns**	**100 ns**	**Max value**
1	2.307	1.015	2.765	0.614	3.056	2.431	2.307	1.906	1.282	1.273	3.056
2	1.852	0.89	2.62	2.261	1.585	1.357	1.873	1.525	1.48	1.349	2.62
3	2.045	1.655	2.782	2.464	1.444	1.438	1.943	1.167	1.617	1.12	2.782
4	1.325	2.007	2.263	1.845	1.137	1.472	1.472	1.064	2.667	1.152	2.667
5	1.912	1.521	1.525	1.511	1.228	0.982	1.191	1.311	1.194	0.981	1.912
6	1.812	1.813	2.233	1.395	0.966	0.861	0.973	0.862	0.991	0.82	2.233

7	1.228	1.828	1.875	1.159	0.684	0.619	0.894	0.829	1.046	0.543	1.875
8	1.952	2.011	1.991	1.052	0.789	0.722	0.839	0.797	0.905	0.765	2.011
9	1.825	2.318	1.751	1.235	1.116	1.25	1.041	1.239	1.036	1.194	2.318
10	3.059	2.395	1.477	1.008	1.323	1.23	0.871	1.467	0.704	1.849	3.059
11	2.449	2.179	1.556	1.449	1.494	1.745	1.145	1.817	0.942	1.025	2.449
12	2.231	1.842	1.208	1.224	0.952	1.479	2.314	1.34	0.936	1.367	2.314
13	2.042	1.24	1.284	1.387	0.967	1.192	0.964	1.184	0.651	0.694	2.042
14	2.174	1.653	1.75	1.682	0.811	1.151	0.885	1.536	0.881	1.315	2.174
15	2.696	1.576	1.243	1.457	0.741	1.238	0.853	1.303	0.811	0.938	2.696
16	1.878	1.334	1.333	1.215	0.491	2.032	0.987	1.968	1.017	1.211	2.032
17	1.368	0.936	1.218	1.824	0.333	2.675	1.229	2.218	1.138	1.095	2.675
18	1.305	0.896	1.861	1.473	0.484	1.612	1.777	2.264	1.145	1.104	2.264
19	1.27	1.103	3.231	1.294	0.796	1.934	2.727	1.966	1.321	1.269	3.231
20	0.673	2.062	2.89	1.124	0.621	0.715	1.52	1.058	0.791	0.803	2.89
21	1.313	0.988	2.341	1.414	0.591	0.787	0.821	0.787	0.639	0.688	2.341
22	0.862	1.657	2.019	1.135	0.911	0.701	0.457	0.472	0.566	0.528	2.019
23	0.575	0.503	0.893	0.791	0.795	0.493	0.418	0.64	0.68	0.632	0.893
24	0.338	0.449	0.741	0.488	0.675	0.297	0.326	0.546	0.505	0.391	0.741
25	0.456	0.442	0.72	0.319	0.392	0.185	0.272	0.427	0.502	0.359	0.72
26	0.477	0.282	0.493	0.33	0.276	0.318	0.249	0.255	0.564	0.437	0.564
27	0.285	0.234	0.467	0.354	0.323	0.196	0.247	0.337	0.473	0.397	0.473
28	0.423	0.484	0.436	0.338	0.397	0.137	0.192	0.357	0.471	0.397	0.484
29	0.333	0.147	0.533	0.343	0.277	0.123	0.157	0.197	0.398	0.326	0.533
30	0.259	0.243	0.474	0.381	0.288	0.131	0.216	0.502	0.481	0.477	0.502
31	0.359	0.26	0.523	0.385	0.324	0.194	0.203	0.502	0.481	0.528	0.528
32	0.37	0.275	0.787	0.344	0.411	0.175	0.181	0.433	0.489	0.509	0.787
33	0.367	0.279	1.571	0.262	0.527	0.353	0.363	0.312	0.606	0.597	1.571
34	0.697	0.596	0.831	0.66	0.592	0.591	0.584	0.798	0.913	1.012	1.012
35	1.029	0.975	1.426	1.006	0.738	0.73	0.931	0.888	1.094	1.263	1.426
36	1.237	1.042	2.342	2.745	0.862	0.75	1.048	0.672	1.117	1.198	2.745
37	1.029	0.874	1.573	1.509	0.583	0.491	1.028	0.514	0.96	1.015	1.573
38	1.162	1.125	1.764	1.881	0.899	0.703	1.331	0.752	1.156	1.292	1.881

39	1.126	1.323	2.162	2.322	1.618	0.973	1.184	0.644	1.101	1.584	2.322
40	0.967	1.289	1.896	2.099	2.184	0.975	1.025	0.627	0.879	1.426	2.184
41	0.761	1.064	1.583	1.658	1.059	0.787	0.658	0.62	0.644	1.042	1.658
42	0.786	0.814	1.435	1.485	0.721	0.797	0.538	0.603	0.601	1.094	1.485
43	0.83	0.918	1.187	1.119	0.535	0.572	0.493	0.455	0.327	0.763	1.187
44	0.989	0.531	1.231	0.924	0.832	0.691	0.526	0.656	0.395	0.972	1.231
45	0.825	0.381	0.792	0.665	1	0.952	0.635	0.798	0.578	0.817	1
46	0.965	0.305	1.051	0.805	1.264	1.235	0.705	0.746	0.65	1.078	1.264
47	0.86	0.284	0.956	0.925	1.059	1.362	0.744	0.76	0.811	1.365	1.365
48	0.65	0.165	0.66	0.746	0.883	1.25	0.667	0.846	0.803	1.246	1.25
49	0.626	0.167	0.709	0.803	1.063	1.246	0.728	0.695	0.7	1.158	1.246
50	0.822	0.227	0.696	1.025	1.338	1.403	0.907	0.763	0.845	1.406	1.406
51	0.645	0.204	0.767	1.046	1.567	1.614	0.909	0.739	0.926	1.694	1.694
52	0.645	0.197	0.709	0.98	1.677	1.674	0.966	0.812	0.948	1.727	1.727
53	0.833	0.381	1.032	1.51	2.126	2.47	1.227	0.91	1.202	1.945	2.47
54	1.014	0.574	1.416	2.094	3.437	3.899	1.74	1.401	1.682	2.546	3.899
55	0.895	0.685	2.341	3.235	3.573	3.799	2.084	1.618	1.84	2.856	3.799
56	0.663	0.4	2.505	3.01	3.044	3.017	2.095	1.738	2.669	2.741	3.044
57	0.574	0.39	3.014	2.281	2.39	2.401	1.738	1.367	2.272	1.913	3.014
58	0.536	0.373	1.1	1.744	2.625	2.716	1.644	1.097	1.289	1.442	2.716
59	0.226	0.165	0.62	0.708	1.449	2.578	1.291	0.796	0.851	0.685	2.578
60	0.341	0.255	0.84	0.345	1.573	2.158	1.657	0.948	0.838	0.84	2.158
61	0.331	0.292	0.672	0.337	0.724	0.985	1.057	0.792	0.679	0.745	1.057
62	0.371	0.361	0.526	0.342	0.531	0.911	0.967	0.518	0.605	0.566	0.967
63	0.333	0.313	0.583	0.295	0.565	0.778	1.091	0.689	0.512	0.37	1.091
64	0.354	0.551	0.548	0.309	0.653	0.783	0.957	0.727	0.526	0.564	0.957
65	0.266	0.314	0.443	0.33	0.334	0.895	0.751	0.579	0.342	0.505	0.895
66	0.217	0.461	0.46	0.284	0.408	0.61	0.712	0.683	0.316	0.508	0.712
67	0.225	0.618	0.523	0.522	0.596	0.473	0.91	1.337	0.299	0.477	1.337
68	0.452	0.431	0.445	1.077	0.328	0.706	0.885	1.425	0.352	0.852	1.425
69	0.359	0.369	0.506	0.67	0.381	0.596	0.786	1.077	0.699	0.88	1.077
70	0.593	0.596	0.853	1.123	0.51	0.698	0.899	1.253	0.868	1.099	1.253

71	0.508	0.919	1.156	0.864	0.599	0.321	0.536	0.887	0.417	0.637	1.156
72	0.267	1.244	1.252	0.753	0.792	0.555	0.52	0.775	0.548	0.713	1.252
73	0.463	1.153	1.409	1.084	0.652	0.748	0.628	0.934	0.559	0.925	1.409
74	0.681	1.457	1.422	1.979	0.745	0.881	0.701	0.731	0.575	0.752	1.979
75	1.316	1.88	1.88	2.872	0.926	1.169	0.927	0.807	0.638	1.098	2.872
76	2.457	2.291	2.705	3.896	1.29	1.563	1.223	1.119	0.927	1.268	3.896
77	3.077	2.597	2.831	4.007	1.638	1.816	2.016	1.036	1.203	1.733	4.007
78	2.534	1.586	3.155	2.238	1.563	1.604	2.388	0.936	1.364	1.734	3.155
79	1.477	2.532	2.181	2.368	1.115	1.091	1.063	0.669	0.718	1.129	2.532
80	0.654	1.517	1.786	1.539	0.725	0.614	0.59	0.417	0.454	0.93	1.786
81	0.403	1.523	1.812	1.363	0.439	0.393	0.343	0.422	0.462	0.557	1.812
82	0.384	1.355	1.154	1.005	0.474	0.624	0.387	0.505	0.449	0.845	1.355
83	0.393	0.977	0.902	0.851	0.361	0.492	0.431	0.512	0.536	0.723	0.977
84	0.565	0.972	1.286	0.84	0.935	1.118	0.845	0.613	1.03	1.155	1.286
85	0.657	0.711	0.89	0.684	0.603	0.682	1.01	0.605	0.998	0.936	1.01
86	0.773	0.661	0.874	0.64	0.601	0.841	0.866	0.702	0.99	1.107	1.107
87	0.552	0.601	0.69	0.508	0.739	0.85	0.795	0.742	1	0.814	1
88	0.634	0.565	0.446	0.585	0.701	0.767	0.776	0.653	1.09	0.987	1.09
89	0.519	0.515	0.463	0.495	0.384	0.58	0.557	0.553	0.873	0.851	0.873
90	0.346	0.392	0.332	0.303	0.397	0.544	0.475	0.519	0.715	0.765	0.765
91	0.367	0.582	0.395	0.54	0.661	0.785	0.498	0.522	0.762	0.863	0.863
92	0.493	0.567	0.372	0.371	0.491	0.508	0.333	0.551	0.598	0.691	0.691
93	0.342	0.399	0.159	0.189	0.153	0.364	0.263	0.479	0.442	0.321	0.479
94	0.294	0.317	0.141	0.168	0.13	0.352	0.147	0.386	0.305	0.321	0.386
95	0.478	0.381	0.161	0.18	0.143	0.277	0.153	0.426	0.369	0.353	0.478
96	0.282	0.347	0.325	0.188	0.199	0.295	0.319	0.49	0.392	0.126	0.49
97	0.328	0.3	0.283	0.162	0.176	0.355	0.13	0.126	0.271	0.108	0.355
98	0.225	0.218	0.186	0.328	0.128	0.257	0.336	0.108	0.269	0.109	0.336
99	0.23	0.362	0.274	0.371	0.18	0.244	0.284	0.129	0.26	0.128	0.371
100	0.221	0.206	0.252	0.355	0.283	0.254	0.276	0.108	0.347	0.311	0.355
101	0.229	0.318	0.084	0.355	0.119	0.243	0.17	0.106	0.265	0.11	0.355
102	0.398	0.43	0.17	0.28	0.153	0.317	0.18	0.134	0.262	0.129	0.43

103	0.213	0.268	0.33	0.51	0.396	0.361	0.369	0.311	0.269	0.169	0.51
104	0.15	0.324	0.121	0.332	0.16	0.249	0.174	0.24	0.266	0.27	0.332
105	0.147	0.186	0.102	0.284	0.187	0.303	0.179	0.232	0.261	0.21	0.303
106	0.181	0.275	0.318	0.459	0.452	0.433	0.458	0.42	0.285	0.288	0.459
107	0.327	0.328	0.301	0.341	0.326	0.315	0.25	0.268	0.315	0.306	0.341
108	0.45	0.201	0.097	0.254	0.336	0.399	0.329	0.407	0.346	0.311	0.45
109	0.449	0.486	0.316	0.509	0.636	0.66	0.553	0.706	0.884	0.537	0.884
110	0.591	0.429	0.604	0.731	0.665	0.574	0.872	0.503	0.648	0.727	0.872
111	0.572	0.529	0.775	0.75	0.642	0.579	0.742	0.542	0.699	0.859	0.859
112	0.625	0.333	0.969	0.829	0.695	0.681	0.878	0.602	1.159	1.016	1.159
113	0.678	0.458	1.808	0.997	0.844	0.772	1.286	0.868	2.717	1.392	2.717
114	0.566	0.426	2.198	0.73	0.599	0.55	0.981	0.597	1.752	1.764	2.198
115	0.307	0.388	0.958	0.458	0.494	0.391	0.705	0.524	1.338	0.899	1.338
116	0.159	0.346	0.239	0.339	0.363	0.274	0.309	0.332	0.871	0.864	0.871
117	0.141	0.318	0.206	0.343	0.378	0.242	0.321	0.376	0.519	0.377	0.519
118	0.287	0.306	0.175	0.504	0.497	0.437	0.301	0.478	0.507	0.391	0.507
119	0.173	0.298	0.147	0.287	0.371	0.21	0.254	0.322	0.34	0.13	0.371
120	0.118	0.391	0.143	0.26	0.385	0.19	0.259	0.319	0.278	0.113	0.391
121	0.37	0.346	0.127	0.303	0.435	0.204	0.27	0.344	0.396	0.131	0.435
122	0.178	0.286	0.138	0.261	0.378	0.321	0.258	0.328	0.335	0.135	0.378
123	0.108	0.292	0.155	0.247	0.384	0.338	0.246	0.318	0.324	0.12	0.384
124	0.11	0.306	0.175	0.35	0.366	0.512	0.437	0.444	0.448	0.428	0.512
125	0.322	0.363	0.238	0.681	0.496	0.385	0.305	0.402	0.372	0.121	0.681
126	0.35	0.296	0.372	0.602	0.535	0.619	0.45	0.465	0.389	0.115	0.619
127	0.159	0.423	0.431	0.435	0.443	0.699	0.47	0.419	0.462	0.098	0.699
128	0.507	0.514	0.565	0.759	0.61	0.879	0.685	0.671	0.514	0.157	0.879
129	0.646	0.59	0.494	0.909	0.787	1.159	0.803	0.682	0.753	0.247	1.159
130	0.657	0.456	0.554	0.696	0.712	1.241	0.776	0.537	0.437	0.156	1.241
131	0.653	0.712	0.536	0.782	0.945	1.649	0.887	0.562	0.449	0.163	1.649
132	0.964	1.09	0.727	1.667	1.282	2.324	1.159	0.704	0.678	0.443	2.324
133	1.154	0.964	0.796	1.613	1.272	2.145	1.147	0.696	0.756	0.315	2.145
134	1.181	0.977	1.028	1.704	1.267	1.957	1.113	0.85	0.83	0.571	1.957

135	0.864	0.716	0.82	0.926	0.978	1.228	0.76	0.756	0.775	0.598	1.228
136	0.522	0.606	0.596	0.658	0.804	0.925	0.496	0.535	0.554	0.361	0.925
137	0.558	0.588	0.58	0.555	0.691	0.903	0.407	0.472	0.573	0.452	0.903
138	0.475	0.38	0.413	0.462	0.556	0.793	0.431	0.515	0.338	0.123	0.793
139	0.199	0.411	0.282	0.472	0.438	0.512	0.438	0.366	0.274	0.112	0.512
140	0.22	0.308	0.359	0.401	0.42	0.436	0.361	0.357	0.404	0.101	0.436
141	0.152	0.424	0.36	0.414	0.207	0.373	0.368	0.429	0.244	0.224	0.429
142	0.098	0.257	0.316	0.315	0.325	0.163	0.26	0.178	0.21	0.101	0.325
143	0.148	0.41	0.518	0.158	0.221	0.103	0.219	0.167	0.348	0.329	0.518
144	0.096	0.2	0.398	0.146	0.117	0.108	0.126	0.149	0.381	0.451	0.451
145	0.093	0.369	0.376	0.166	0.128	0.093	0.112	0.144	0.248	0.252	0.376
146	0.088	0.423	0.303	0.164	0.26	0.087	0.115	0.13	0.247	0.163	0.423
147	0.146	0.367	0.377	0.291	0.334	0.09	0.108	0.121	0.227	0.101	0.377
148	0.114	0.393	0.343	0.237	0.159	0.104	0.104	0.129	0.232	0.254	0.393
149	0.096	0.373	0.212	0.254	0.432	0.12	0.11	0.126	0.203	0.258	0.432
150	0.439	0.291	0.115	0.293	0.327	0.24	0.189	0.468	0.457	0.346	0.468
151	0.392	0.378	0.245	0.291	0.454	0.168	0.24	0.35	0.637	0.796	0.796
152	0.219	0.324	0.135	0.454	0.515	0.31	0.425	0.534	0.782	0.796	0.796
153	0.26	0.344	0.219	0.43	0.477	0.524	0.393	0.686	1.18	1.012	1.18
154	0.214	0.421	0.436	0.249	0.551	0.626	0.402	0.639	1.346	1.057	1.346
155	0.593	0.737	0.661	0.657	0.695	0.816	0.532	0.862	1.085	0.773	1.085
156	0.464	0.673	0.649	0.718	0.766	0.75	0.592	0.626	0.868	0.82	0.868
157	0.48	0.787	0.817	0.484	0.724	0.647	0.545	0.529	0.777	0.636	0.817
158	0.305	0.861	0.651	0.744	0.759	0.566	0.398	0.554	0.737	0.765	0.861
159	0.226	0.573	0.354	0.46	0.553	0.569	0.298	0.23	0.399	0.442	0.573
160	0.257	0.393	0.453	0.301	0.724	0.511	0.227	0.235	0.446	0.294	0.724
161	0.293	0.381	0.336	0.3	0.547	0.459	0.336	0.222	0.489	0.526	0.547
162	0.246	0.47	0.351	0.378	0.656	0.33	0.563	0.215	0.607	0.431	0.656
163	0.182	0.185	0.289	0.319	0.515	0.26	0.341	0.206	0.455	0.177	0.515
164	0.119	0.176	0.269	0.457	0.125	0.423	0.564	0.138	0.277	0.24	0.564
165	0.158	0.42	0.461	0.48	0.133	0.44	0.631	0.341	0.405	0.263	0.631
166	0.489	0.361	0.324	0.505	0.116	0.304	0.489	0.161	0.448	0.24	0.505

167	0.316	0.51	0.305	0.719	0.125	0.103	0.428	0.324	0.181	0.092	0.719
168	0.572	0.448	0.393	0.552	0.138	0.386	0.302	0.199	0.372	0.343	0.572
169	0.48	0.431	0.548	0.468	0.257	0.27	0.418	0.452	0.201	0.385	0.548
170	0.328	0.411	0.609	0.214	0.162	0.103	0.346	0.144	0.289	0.308	0.609
171	0.738	0.5	0.585	0.372	0.161	0.342	0.598	0.276	0.343	0.423	0.738
172	0.769	0.689	0.784	0.67	0.253	0.47	0.582	0.183	0.573	0.602	0.784
173	0.932	0.859	1.013	0.512	0.691	0.529	0.718	0.357	0.824	0.779	1.013
174	0.937	0.708	0.738	0.485	0.617	0.527	0.507	0.571	0.539	0.521	0.937
175	0.901	0.671	0.787	0.488	0.448	0.486	0.661	0.636	0.401	0.461	0.901
176	1.203	1.102	1.024	0.679	1.066	1.025	0.702	0.801	0.569	0.459	1.203
177	1.487	1.845	0.982	0.7	2.245	2.223	0.576	0.534	0.635	0.437	2.245
178	0.825	1.977	0.688	0.561	1.641	1.403	0.745	0.501	0.444	0.559	1.977
179	0.827	1.087	0.599	0.607	2.485	2.717	0.653	0.342	0.204	0.69	2.717
180	0.696	0.968	0.767	0.506	2.187	1.406	0.434	0.182	0.201	0.181	2.187
181	0.665	0.883	0.666	0.59	2.701	1.757	0.811	0.284	0.287	0.54	2.701
182	0.628	0.626	0.476	0.425	2.283	1.697	0.629	0.317	0.303	0.508	2.283
183	0.541	0.684	0.494	0.441	1.031	0.718	0.704	0.443	0.445	0.579	1.031
184	0.632	0.506	0.383	0.48	0.468	0.374	0.503	0.174	0.27	0.237	0.632
185	0.778	0.453	0.233	0.335	0.433	0.371	0.383	0.144	0.203	0.094	0.778
186	0.533	0.399	0.195	0.468	0.472	0.425	0.398	0.232	0.372	0.317	0.533
187	0.47	0.407	0.193	0.301	0.319	0.139	0.312	0.146	0.358	0.097	0.47
188	0.512	0.387	0.139	0.273	0.328	0.135	0.367	0.143	0.351	0.106	0.512
189	0.612	0.426	0.126	0.278	0.304	0.132	0.319	0.134	0.28	0.094	0.612
190	0.518	0.565	0.204	0.518	0.316	0.14	0.324	0.142	0.556	0.099	0.565
191	0.483	0.632	0.366	0.496	0.312	0.146	0.327	0.133	0.644	0.106	0.644
192	0.456	0.604	0.201	0.356	0.287	0.143	0.319	0.124	0.406	0.109	0.604
193	0.637	0.942	0.233	0.471	0.304	0.15	0.34	0.162	0.496	0.113	0.942
194	0.67	1.16	0.653	0.712	0.394	0.369	0.522	0.392	0.8	0.29	1.16
195	0.597	1.194	1.383	0.752	0.48	0.372	0.509	0.399	1.42	0.753	1.42
196	0.642	1.279	1.985	1.057	0.714	0.587	0.733	0.653	1.693	1.767	1.985
197	0.505	1.347	1.686	1.18	0.696	0.545	0.771	0.799	1.447	1.174	1.686
198	0.593	1.251	1.732	1.129	0.75	0.535	0.844	0.748	1.656	0.516	1.732

199	0.574	0.694	0.683	0.799	0.423	0.267	0.659	0.342	2.136	0.172	2.136
200	0.286	0.327	0.309	0.507	0.368	0.199	0.456	0.186	0.613	0.168	0.613
201	0.349	0.248	0.281	0.477	0.245	0.214	0.349	0.179	0.279	0.126	0.477
202	0.505	0.406	0.293	0.47	0.45	0.466	0.487	0.411	0.448	0.165	0.505
203	0.28	0.167	0.171	0.458	0.17	0.126	0.344	0.176	0.306	0.1	0.458
204	0.322	0.284	0.139	0.426	0.164	0.177	0.338	0.18	0.329	0.1	0.426
205	0.385	0.142	0.147	0.452	0.152	0.155	0.306	0.182	0.293	0.111	0.452
206	0.483	0.146	0.196	0.228	0.146	0.125	0.305	0.174	0.281	0.107	0.483
207	0.455	0.142	0.155	0.166	0.127	0.105	0.286	0.162	0.266	0.084	0.455
208	0.579	0.293	0.141	0.183	0.119	0.105	0.289	0.168	0.274	0.084	0.579
209	0.684	0.373	0.184	0.189	0.129	0.12	0.297	0.175	0.173	0.118	0.684
210	0.683	0.396	0.176	0.19	0.131	0.135	0.305	0.362	0.402	0.43	0.683
211	0.497	0.426	0.242	0.184	0.122	0.399	0.322	0.531	0.428	0.319	0.531
212	0.614	0.505	0.25	0.201	0.123	0.122	0.363	2.236	0.836	0.573	2.236
213	0.99	0.576	0.633	0.646	0.397	0.43	0.689	1.42	2.227	0.867	2.227
214	1.222	0.628	0.777	0.6	0.901	0.949	1.697	1.624	2.185	0.81	2.185
215	1.561	0.949	1.192	0.855	1.438	1.465	3.101	2.108	2.381	1.105	3.101
216	1.423	0.961	1.175	0.726	2.059	1.769	4.561	4.215	3.884	1.212	4.561
217	1.391	0.846	0.808	0.488	2.058	1.755	4.601	5.35	3.938	1.479	5.35
218	1.535	0.8	1.232	0.62	3.284	3.509	4.9	5.879	4.279	1.672	5.879
219	2.272	1.343	1.282	0.92	4.018	4.554	6.673	7.146	4.081	2.001	7.146
220	3.365	2.968	1.958	0.985	5.218	5.952	8.265	8.282	4.977	2.666	8.282
221	4.556	4.601	3.082	2.981	7.096	8.12	9.443	9.858	5.897	3.303	9.858
222	6.141	6.194	4.377	4.159	8.791	9.885	10.824	10.996	7.202	3.645	10.996
223	7.4	7.543	5.298	4.654	10.353	11.705	12.485	12.496	8.945	5.236	12.496

REFERENCES

References:

1. Global tuberculosis report 2022. Geneva: World Health Organization; 2022. License: CC BY-NC-SA 3.0 IGO.
2. Lee SH (2016) Tuberculosis infection and latent tuberculosis. Tuberc Respir Dis (Seoul) 79:201-206.
3. HonerzuBentrop K, Russell DG (2001) Mycobacterial persistence: adaptation to a changing environment. Trends Microbiol 9:597–605.
4. Prasad R, Gupta N, Banka A (2018) Multidrug–resistant tuberculosis / rifampicin–resistant tuberculosis: principles of management. Lung India 35:78–81.
5. Haley CA, Macias P, Jasuja S, Jones BA, Rowlinson MC, Jaimon R, Onderko P, Darnell E, Gomez ME, Peloquin C, Ashkin D, Goswami ND (2021) Novel 6–month treatment for drug–resistant tuberculosis, United States. Emerg Infect Dis 27:332–334.
6. Masjedi MR, Farnia P, Sorooch S, Pooramiri MV, Mansoori SD, Zarifi AZ, Akbarvelayati A, Hoffner S (2006) Extensively drug-resistant tuberculosis: 2 years of surveillance in Iran. Clinical Infect Dis 43:841–847.
7. Velayati AA, Masjedi MR, Farnia P, Tabarsi P, Ghanavi J, ZiaZarifi AH, Hoffner SE (2009) Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in Iran. Chest 136:420–425.
8. Velayati AA, Farnia P, Masjedi MR (2013) The totally drug resistant tuberculosis (TDR–TB). Int J Clin Exp Med 6:307–309.
9. Donald PR, Diacon AH, Lange C, Demers A–M, von Groote–Bidlingmaier F, Nardell E (2018) Droplets, dust and guinea pigs: an historical review of tuberculosis transmission research, 1878–1940. Int J Tuberc Lung Dis 22:972–982.
10. Eum SY, Kong JH, Hong MS, Lee YJ, Kim JH, Hwang SH, Cho SN, Via LE, Barry CE 3rd (2010) Neutrophils are the predominant infected phagocytic cells in the airways of patients with active pulmonary TB. Chest 137:122–128.
11. Lerner TR, Borel S, Gutierrez MG (2015) The innate immune response in human tuberculosis. Cell Microbiol 17:1277–1285.
12. Norris BA, Ernst JD (2018) Mononuclear cell dynamics in *M. tuberculosis* infection provide opportunities for therapeutic intervention. PLoS Pathog 14:e1007154.
13. Cohen SB, Gern BH, Delahaye JL, Adams KN, Plumlee CR, Winkler JK, Sherman DR, Gerner MY, Urdahl KB (2018) Alveolar macrophages provide an early *Mycobacterium tuberculosis* niche and initiate dissemination. Cell Host Microbe 24:439–446.e4.
14. Rubin EJ (2009) The granuloma in tuberculosis–friend or foe? N Engl J Med 360:2471–2473.
15. Boom WH, Schaible UE, Achkar JM (2021) The knowns and unknowns of latent *Mycobacterium tuberculosis* infection. J Clin Invest 131(3):e136222.
16. Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC, Ginsberg A, Swaminathan S, Spigelman M, Getahun H, Menzies D, Raviglione M (2016) Tuberculosis. Nat Rev Dis Primers 2:16076.
17. Behr MA,Waters WR (2014) Is tuberculosis a lymphatic disease with a pulmonary portal? Lancet Infect Dis 14:250–255.

18. Nair VR, Franco LH, Zacharia VM, Khan HS, Stamm CE, You W, Marciano DK, Yagita H, Levine B, Shiloh MU (2016) Microfold cells actively translocate *Mycobacterium tuberculosis* to initiate infection. Cell Rep 16:1253–1258.
19. Lerner TR, de Souza Carvalho-Wodarz C, Repnik U, Russell MR, Borel S, Diedrich CR, Rohde M, Wainwright H, Collinson LM, Wilkinson RJ, Griffiths G, Gutierrez MG (2016) Lymphatic endothelial cells are a replicative niche for *Mycobacterium tuberculosis*. J Clin Invest 126:1093–1108.
20. Beigier-Bompadre M, Montagna GN, Kühl AA, Lozza L, Weiner J 3rd, Kupz A, Vogelzang A, Mollenkopf HJ, Löwe D, Bandermann S, Dorhoi A, Brinkmann V, Matuschewski K, Kaufmann SHE (2017) *Mycobacterium tuberculosis* infection modulates adipose tissue biology. PLoS Pathog 13(10):e1006676.
21. Mayito J, Andia I, Belay M, Jolliffe DA, Kateete DP, Reece ST, Martineau AR (2019) Anatomic and cellular niches for *Mycobacterium tuberculosis* in latent tuberculosis infection. J Infect Dis 219:685–694.
22. Chandra P, Grigsby SJ, Philips JA (2022) Immune evasion and provocation by *Mycobacterium tuberculosis*. Nat Rev Microbiol 20:750–766.
23. Cambau E, Drancourt M (2014) Steps towards the discovery of *Mycobacterium tuberculosis* by Robert Koch, 1882. Clin Microbiol Infect 20:196–201.
24. Bhaskar A, Dwivedi VP, Nandicoori VK (2019) "Eliminating mycobacterial persistence: Novel targets for anti–TB therapy," in Pathogenicity and drug resistance of human pathogens. Eds. S Hameed, Z Fatima, et al. (Singapore: Springer Singapore), 57–79.
25. Alnimr AM (2015) Dormancy models for *Mycobacterium tuberculosis*: A minireview. Braz J Microbiol 46:641–647.
26. Singh R, Dwivedi SP, Gaharwar US, Meena R, Rajamani P, Prasad T (2020) Recent updates on drug resistance in *Mycobacterium tuberculosis*. J Appl Microbiol 128:1547–1567.
27. Iona E, Pardini M, Mustazzolu A, Piccaro G, Nisini R, Fattorini L, Giannoni F (2016) *Mycobacterium tuberculosis* gene expression at different stages of hypoxia-induced dormancy and upon resuscitation. J Microbiol 54:565–72.
28. Du P, Sohaskey CD, Shi L (2016) Transcriptional and physiological changes during *Mycobacterium tuberculosis* reactivation from non-replicating persistence. Front Microbiol 7:1346.
29. Talaat AM, Lyons R, Howard ST, Johnston SA (2004) The temporal expression profile of *Mycobacterium tuberculosis* infection in mice. Proc Natl Acad Sci U S A 101:4602–4607.
30. Hudock TA, Foreman TW, Bandyopadhyay N, Gautam US, Veatch AV, LoBato DN, Gentry KM, Golden NA, Cavigli A, Mueller M, Hwang SA, Hunter RL, Alvarez X, Lackner AA, Bader JS, Mehra S, Kaushal D (2017) Hypoxia sensing and persistence genes are expressed during the intragranulomatous survival of *Mycobacterium tuberculosis*. Am J Respir Cell Mol Biol 56:637–647.
31. Fenhalls G, Stevens L, Moses L, Bezuidenhout J, Betts JC, Helden Pv Pv, Lukey PT, Duncan K (2002) In situ detection of *Mycobacterium tuberculosis* transcripts in human lung granulomas reveals differential gene expression in necrotic lesions. Infect Immun 70:6330–6338.
32. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K,

Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393:537–544.

33. Crellin PK, Luo C-Y, Morita YS (2013) Metabolism of plasma membrane lipids in mycobacteria and corynebacteria. Lipid Metabolism 6:119–148.
34. Cox JS, Chen B, McNeil M, Jacobs WR Jr (1999) Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. Nature 402:79–83.
35. Reed MB, Domenech P, Manca C, Su H, Barczak AK, Kreiswirth BN, Kaplan G, Barry CE 3rd (2004) A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. Nature 431:84–87.
36. Rao V, Fujiwara N, Porcelli SA, Glickman MS (2005) *Mycobacterium tuberculosis* controls host innate immune activation through cyclopropane modification of a glycolipid effector molecule. J Exp Med 201:535–543.
37. Brennan PJ (2003) Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. Tuberculosis (Edinb) 83:91–97.
38. Jackson M (2014) The mycobacterial cell envelope–lipids. Cold Spring Harb Perspect Med 4(10):a021105.
39. Alderwick LJ, Harrison J, Lloyd GS, Birch HL (2015) The mycobacterial cell Wall–peptidoglycan and arabinogalactan. Cold Spring Harb Perspect Med 5(8):a021113.
40. Moopanar K, Mvubu NE (2020) Lineage-specific differences in lipid metabolism and its impact on clinical strains of *Mycobacterium tuberculosis*. Microb Pathog 146:104250.
41. Ercan B, Low W-Y, Liu X, Chng S-S (2019) Characterization of interactions and phospholipid transfer between substrate binding proteins of the OmpC-Mla system. Biochemistry 58:114–119.
42. Converse SE, Mougous JD, Leavell MD, Leary JA, Bertozzi CR, Cox JS (2003) MmpL8 is required for sulfolipid-1 biosynthesis and *Mycobacterium tuberculosis* virulence. Proc Natl Acad Sci U S A 100:6121–6126.
43. Grzegorzewicz AE, Pham H, Gundi VA, Scherman MS, North EJ, Hess T, Jones V, Gruppo V, Born SE, Korduláková J, Chavadi SS, Morisseau C, Lenaerts AJ, Lee RE, McNeil MR, Jackson M (2012) Inhibition of mycolic acid transport across the *Mycobacterium tuberculosis* plasma membrane. Nat Chem Biol 8:334–341.
44. Slayden RA, Barry CE 3rd (2002) The role of KasA and KasB in the biosynthesis of meromycolic acids and isoniazid resistance in *Mycobacterium tuberculosis*. Tuberculosis (Edinb) 82:149–160.
45. Jackson M, Crick DC, Brennan PJ (2000) Phosphatidylinositol is an essential phospholipid of mycobacteria. J Biol Chem 275:30092–30099.
46. Schaeffer ML, Khoo KH, Besra GS, Chatterjee D, Brennan PJ, Belisle JT, Inamine JM (1999) The *pimB* gene of *Mycobacterium tuberculosis* encodes a mannosyltransferase involved in lipoarabinomannan biosynthesis. J Biol Chem 274:31625–31631.
47. Rezwan M, Grau T, Tschumi A, Sander P (2007) Lipoprotein synthesis in mycobacteria. Microbiology (Reading) 153:652–658.

48. Zamyatina A, Heine H (2020) Lipopolysaccharide recognition in the crossroads of TLR4 and Caspase–4/11 mediated inflammatory pathways. Front Immunol 11:585146.
49. Layre E (2020) Trafficking of *Mycobacterium tuberculosis* envelope components and release within extracellular vesicles: host-pathogen interactions beyond the wall. Front Immunol 11:1230.
50. Augenstreich J, Briken V (2020) Host cell targets of released lipid and secreted protein effectors of *Mycobacterium tuberculosis*. Front Cell Infect Microbiol 10:595029.
51. Prados–Rosales R, Carreño LJ, Batista–Gonzalez A, Baena A, Venkataswamy MM, Xu J, Yu X, Wallstrom G, Magee DM, LaBaer J, Achkar JM, Jacobs WR Jr, Chan J, Porcelli SA, Casadevall A (2014) Mycobacterial membrane vesicles administered systemically in mice induce a protective immune response to surface compartments of *Mycobacterium tuberculosis*. mBio 5:e01921–14.
52. Siegrist MS, Bertozzi CR (2014) Mycobacterial lipid logic. Cell Host Microbe 15:1–2.
53. Bansal-Mutalik R, Nikaido H (2014) Mycobacterial outer membrane is a lipid bilayer and the inner membrane is sually rich in diacyl phosphatidylinositol dimannosides. Proc Natl Acad Sci U S A 111:4958–4963.
54. Ojha AK, Baughn AD, Sambandan D, Hsu T, Trivelli X, Guerardel Y, Alahari A, Kremer L, Jacobs WR Jr, Hatfull GF (2008) Growth of *Mycobacterium tuberculosis* biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. Mol Microbiol 69:164–174.
55. Sequeira PC, Senaratne RH, Riley LW (2014) Inhibition of toll-like receptor 2 (TLR-2)-mediated response in human alveolar epithelial cells by mycolic acids and *Mycobacterium tuberculosis* mce1 operon mutant. Pathog Dis 70:132–140.
56. Iizasa E, Chuma Y, Uematsu T, Kubota M, Kawaguchi H, Umemura M, Toyonaga K, Kiyohara H, Yano I, Colonna M, Sugita M, Matsuzaki G, Yamasaki S, Yoshida H, Hara H (2021) TREM2 is a receptor for non-glycosylated mycolic acids of mycobacteria that limits anti-mycobacterial macrophage activation. Nat Commun 12:2299.
57. Flores-Villanueva PO, Ruiz-Morales JA, Song CH, Flores LM, Jo EK, Montano M, Barnes PF, Selman M, Granados J (2005) A functional promoter polymorphism in monocyte chemoattractant protein–1 is associated with increased susceptibility to pulmonary tuberculosis. J Exp Med 202:1649–1658.
58. Sharma NK, Rathor N, Sinha R, Gupta S, Tyagi G, Garima K, Pathak R, Singh P, Jain A, Bose M, Varma–Basil M (2019) Expression of mycolic acid in response to stress and association with differential clinical manifestations of tuberculosis. Int J Mycobacteriol 8:237–243.
59. Patin EC, Geffken AC, Willcocks S, Leschczyk C, Haas A, Nimmerjahn F, Lang R, Ward TH, Schaible UE (2017) Trehalose dimycolate interferes with FcgammaR–mediated phagosome maturation through Mincle, SHP-1 and FcgammaRIIB signalling. PLoS ONE 12: e0174973.
60. Bowker N, Salie M, Schurz H, van Helden PD, Kinnear CJ, Hoal EG, Möller M (2016) Polymorphisms in the pattern recognition receptor mincle gene (CLEC4E) and association with tuberculosis. Lung 194:763–767.
61. Huber A, Kallerup RS, Korsholm KS, Franzyk H, Lepenies B, Christensen D, Foged C, Lang R (2016) Trehalose diester glycolipids are superior to the monoesters in binding to Mincle, activation of macrophages in vitro and adjuvant activity in vivo. Innate Immun 22:405–418.

62. Cambier CJ, Takaki KK, Larson RP, Hernandez RE, Tobin DM, Urdahl KB, Cosma CL, Ramakrishnan L (2014) Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids. Nature 505:218–222.
63. Day TA, Mittler JE, Nixon MR, Thompson C, Miner MD, Hickey MJ, Liao RP, Pang JM, Shayakhmetov DM, Sherman DR (2014) *Mycobacterium tuberculosis* strains lacking surface lipid phthiocerol dimycocerosate are susceptible to killing by an early innate host response. Infect Immun 82:5214–5222.
64. Zheng RB, Jegouzo SAF, Joe M, Bai Y, Tran H-A, Shen K, Saupe J, Xia L, Ahmed MF, Liu Y-H, Patil PS, Tripathi A, Hung S-C, Taylor MF, Lowary TL, Drickamer K (2017) Insights into interactions of mycobacteria with the host innate immune system from a novel array of synthetic mycobacterial glycans. ACS Chem Biol 12:2990–3002.
65. Li K, Underhill DM (2020) C-type lectin receptors in phagocytosis. Curr Top Microbiol Immunol 429: 1–18.
66. Torrelles JB, Knaup R, Kolareth A, Slepushkina T, Kaufman TM, Kang P, Hill PJ, Brennan PJ, Chatterjee D, Belisle JT, Musser JM, Schlesinger LS (2008) Identification of *Mycobacterium tuberculosis* clinical isolates with altered phagocytosis by human macrophages due to a truncated lipoarabinomannan. J Biol Chem 283:31417–31428.
67. Birch HL, Alderwick LJ, Appelmelk BJ, Maaskant J, Bhatt A, Singh A, Nigou J, Eggeling L, Geurtsen J, Besra GS (2010) A truncated lipoglycan from mycobacteria with altered immunological properties. Proc Natl Acad Sci U S A 107:2634–2639.
68. Bloch H, Segal W (1956) Biochemical differentiation of *Mycobacterium tuberculosis* grown in vivo and in vitro. J Bacteriol 72:132–141.
69. Russell DG, Cardona PJ, Kim MJ, Allain S, Altare F (2009) Foamy macrophages and the progression of the human tuberculosis granuloma. Nat Immunol 10:943–948.
70. Wilburn KM, Fieweger RA, VanderVen BC (2018) Cholesterol and fatty acids grease the wheels of *Mycobacterium tuberculosis* pathogenesis. Pathog Dis 76(2):fty021.
71. Fontán P, Aris V, Ghanny S, Soteropoulos P, Smith I (2008) Global transcriptional profile of *Mycobacterium tuberculosis* during THP-1 human macrophage infection. Infect Immun 76:717–725.
72. Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, Dolganov G, Efron B, Butcher PD, Nathan C, Schoolnik GK (2003) Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: insights into the phagosomal environment. J Exp Med 198:693–704.
73. Rohde KH, Abramovitch RB, Russell DG (2007) *Mycobacterium tuberculosis* invasion of macrophages: linking bacterial gene expression to environmental cues. Cell Host Microbe 2:352–364.
74. Neyrolles O (2014) Mycobacteria and the greasy macrophage: getting fat and frustrated. Infect Immun 82:472–475.
75. Peyron P, Vaubourgeix J, Poquet Y, Levillain F, Botanch C, Bardou F, Daffé M, Emile J-F, Marchou B, Cardona P-J, de Chastellier C, Altare F (2008) Foamy macrophages from tuberculous patients' granulomas constitute a nutrient–rich reservoir for *M. tuberculosis* persistence. PLoS Pathog 4(11): e1000204.

76. Savvi S, Warner DF, Kana BD, McKinney JD, Mizrahi V, Dawes SS (2008) Functional characterization of a vitamin B12-dependent methylmalonyl pathway in *Mycobacterium tuberculosis*: implications for propionate metabolism during growth on fatty acids. J Bacteriol 190:3886–3895.
77. Muñoz-Elías EJ, Upton AM, Cherian J, McKinney JD (2006) Role of the methylcitrate cycle in *Mycobacterium tuberculosis* metabolism, intracellular growth, and virulence. Mol Microbiol 60:1109–1122.
78. Eoh H, Rhee KY (2014) Methylcitrate cycle defines the bactericidal essentiality of isocitrate lyase for survival of *Mycobacterium tuberculosis* on fatty acids. Proc Natl Acad Sci USA 111:4976–4981.
79. Quadri LE (2014) Biosynthesis of mycobacterial lipids by polyketide synthases and beyond. Crit Rev Biochem Moll 49:179–211.
80. (2004) A developmental role for fatty acids in eukaryotes. PLoS Biol 2(9): e293.
81. Portevin D, De Sousa–D'Auria C, Houssin C, Grimaldi C, Chami M, Daffé M, Guilhot C (2004) A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms. Proc Natl Acad Sci U S A 101:314–319.
82. Daniel J, Maamar H, Deb C, Sirakova TD, Kolattukudy PE (2011) *Mycobacterium tuberculosis* uses host triacylglycerol to accumulate lipid droplets and acquires a dormancy–like phenotype in lipid–loaded macrophages. PLoS Pathog 7(6):e1002093.
83. Takayama K, Wang C, Besra GS (2005) Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*. Clin Microbiol Rev 18:81–101.
84. Casali N, Riley LW (2007) A phylogenomic analysis of the Actinomycetales *mce* operons. BMC Genomics 8:60.
85. Pandey AK, Sassetti CM (2008) Mycobacterial persistence requires the utilization of host cholesterol. Proc Natl Acad Sci U S A 105:4376–4380.
86. Nazarova EV, Montague CR, La T, Wilburn KM, Sukumar N, Lee W, Caldwell S, Russell DG, VanderVen BC (2017) Rv3723/LucA coordinates fatty acid and cholesterol uptake in *Mycobacterium tuberculosis*. eLife 6:e26969.
87. Mohn WW, van der Geize R, Stewart GR, Okamoto S, Liu J, Dijkhuizen L, Eltis LD (2008) The actinobacterial mce4 locus encodes a steroid transporter. J Biol Chem 283:35368–35374.
88. Joshi SM, Pandey AK, Capite N, Fortune SM, Rubin EJ, Sassetti CM (2006) Characterization of mycobacterial virulence genes through genetic interaction mapping. Proc Natl Acad Sci U S A 103:11760–11765.
89. Saurin W, Köster W, Dassa E (1994) Bacterial binding protein-dependent permeases: characterization of distinctive signatures for functionally related integral cytoplasmic membrane proteins. Mol Microbiol 12:993–1004.
90. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD (2019) The Pfam protein families database in 2019. Nucleic Acids Res 47:D427–D432.
91. Perkowski EF, Miller BK, McCann JR, Sullivan JT, Malik S, Allen IC, Godfrey V, Hayden JD, Braunstein M (2016) An orphaned Mce-associated membrane protein of *Mycobacterium tuberculosis* is a virulence factor that stabilizes Mce transporters. Mol Microbiol 100:90–107.

92. Rank L, Herring LE, Braunstein M (2021) Evidence for the mycobacterial Mce4 transporter being a multiprotein complex. J Bacteriol 203:e00685-20.
93. Chen J, Fruhauf A, Fan C, Ponce J, Ueberheide B, Bhabha G, Ekiert DC (2023) Structure of an endogenous mycobacterial MCE lipid transporter. Nature https://doi.org/10.1038/s41586-023-06366-0.
94. Arruda S, Bomfim G, Knights R, Huima-Byron T, Riley LW (1993) Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. Science 261:1454–1457.
95. Chitale S, Ehrt S, Kawamura I, Fujimura T, Shimono N, Anand N, Lu S, Cohen-Gould L, Riley LW (2001) Recombinant *Mycobacterium tuberculosis* protein associated with mammalian cell entry. Cell Microbiol 3:247–254.
96. Zhang Y, Li J, Li B, Wang J, Liu CH (2018) *Mycobacterium tuberculosis* Mce3C promotes mycobacteria entry into macrophages through activation of β2 integrin-mediated signalling pathway. Cell Microbiol 20(2):10.1111/cmi.12800.
97. Qiang L, Wang J, Zhang Y, Ge P, Chai Q, Li B, Shi Y, Zhang L, Gao GF, Liu CH (2019) *Mycobacterium tuberculosis* Mce2E suppresses the macrophage innate immune response and promotes epithelial cell proliferation. Cell Mol Immunol 16:380–391.
98. Li J, Chai Q-Y, Zhang Y, Li B-X, Wang J, Qiu X-B, Liu CH (2015) *Mycobacterium tuberculosis* Mce3E Suppresses Host Innate Immune Responses by Targeting ERK1/2 Signaling. J Immunol 194:3756–3767.
99. Gioffré A, Infante E, Aguilar D, Santangelo MD la P, Klepp L, Amadio A, Meikle V, Etchechoury I, Romano MI, Cataldi A, Hernández RP, Bigi F (2005) Mutation in *mce* operons attenuates *Mycobacterium tuberculosis* virulence. Microbes Infect 7:325–334.
100. Sassetti CM, Rubin EJ (2003) Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci 100:12989–12994.
101. Marjanovic O, Miyata T, Goodridge A, Kendall LV, Riley LW (2010) Mce2 operon mutant strain of *Mycobacterium tuberculosis* is attenuated in C57BL/6 mice. Tuberc Edinb Scotl 90:50–56.
102. Shimono N, Morici L, Casali N, Cantrell S, Sidders B, Ehrt S, Riley LW (2003) Hypervirulent mutant of *Mycobacterium tuberculosis* resulting from disruption of the *mce1* operon. Proc Natl Acad Sci 100:15918–15923.
103. Lima P, Sidders B, Morici L, Reader R, Senaratne R, Casali N, Riley LW (2007) Enhanced mortality despite control of lung infection in mice aerogenically infected with a *Mycobacterium tuberculosis mce1* operon mutant. Microbes Infect 9:1285–1290.
104. Casali N, White AM, Riley LW (2006) Regulation of the *Mycobacterium tuberculosis mce1* operon. J Bacteriol 188:441–449.
105. Kumar A, Bose M, Brahmachari V (2003) Analysis of expression profile of mammalian cell entry (*mce*) operons of *Mycobacterium tuberculosis*. Infect Immun 71:6083–6087.
106. Santangelo Mde L, Blanco F, Campos E, Soria M, Bianco MV, Klepp L, Alito A, Zabal O, Cataldi A, Bigi F (2009) Mce2R from *Mycobacterium tuberculosis* represses the expression of the *mce2* operon. Tuberculosis (Edinb), 89, 22–28

107. Santangelo MAP, Goldstein J, Alito A, Gioffre A, Caimi K, Zabal O, Zumarraga MN, Romano MI, Cataldi AA, Bigi F (2002) Negative transcriptional regulation of the *mce3* operon in *Mycobacterium tuberculosis*. Microbiology (Reading) 148:2997–3006.
108. Kendall SL, Withers M, Soffair CN, Moreland NJ, Gurcha S, Sidders B, Frita R, Ten Bokum A, Besra GS, Lott JS, Stoker NG (2007) A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. Mol Microbiol 65:684–699.
109. Wipperman MF, Sampson NS, Thomas ST (2014) Pathogen roid rage: cholesterol utilization by *Mycobacterium tuberculosis*. Crit Rev Biochem Mol Biol 49:269–293.
110. Vindal V, Ranjan S, Ranjan A (2007) In silico analysis and characterization of GntR family of regulators from *Mycobacterium tuberculosis*. Tuberculosis 87:242–247.
111. Santangelo MP, Klepp L, Nuñez–García J, Blanco FC, Soria M, García–Pelayo MDC, Bianco MV, Cataldi AA, Golby P, Jackson M, Gordon SV, Bigi F (2009) Mce3R, a TetR-type transcriptional repressor, controls the expression of a regulon involved in lipid metabolism in *Mycobacterium tuberculosis*. Microbiology (Reading) 155:2245–2255.
112. Zeng J, Cui T, He ZG (2012) A genome–wide regulator–DNA interaction network in the human pathogen *Mycobacterium tuberculosis* H37Rv. J Proteome Res 11:4682–4692.
113. Cuthbertson L, Nodwell JR (2013) The TetR family of regulators. Microbiol Mol Biol Rev 77:440–475.
114. Rigali S, Derouaux A, Giannotta F, Dusart J (2002) Subdivision of the helix–turn–helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. J Biol Chem 277:12507–12515.
115. Ho NAT, Dawes SS, Crowe AM, Casabon I, Gao C, Kendall SL, Baker EN, Eltis LD, Lott JS (2016) The structure of the transcriptional repressor KstR in complex with CoA thioester cholesterol metabolites sheds light on the regulation of cholesterol catabolism in *Mycobacterium tuberculosis*. J Biol Chem 291:7256–7266.
116. Yousuf S, Angara RK, Roy A, Gupta SK, Misra R, Ranjan A (2018) Mce2R/Rv0586 of *Mycobacterium tuberculosis* is the functional homologue of $FadR^{E.\ Coli}$. Microbiology 164:1133–1145.
117. Uchida Y, Casali N, White A, Morici L, Kendall LV, Riley LW (2007) Accelerated immunopathological response of mice infected with *Mycobacterium tuberculosis* disrupted in the *mce1* operon negative transcriptional regulator. Cell Microbiol 9:1275–1283.
118. Sambrook J, Russell DW (2014) Molecular cloning: a laboratory manual, 4^{th} edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
119. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. Anal Biochem 72:248–254.
120. Ganguly T, Bandhu A, Chattoraj P, Chanda PK, Das M, Mandal NC, Sau S (2007) Repressor of temperate mycobacteriophage L1 harbors a stable C–terminal domain and binds to different asymmetric operator DNAs with variable affinity. Virol J 4:64.
121. Bandhu A, Ganguly T, Jana B, Mondal R, Sau S (2010) Regions and residues of an asymmetric operator DNA interacting with the monomeric repressor of temperate mycobacteriophage L1. Biochemistry 49:4235–4243.

122. Larouche K, Bergeron M–J, Leclere S, Guerin SL (1996) Optimization of competitor poly (dI–dC)•poly (dI–dC) levels is advised in DNA–protein interaction studies involving enriched nuclear proteins. Biotechniques 20:439–444.
123. Chen X, Farmer G, Zhu H, Prywes R, Prives C (1993) Cooperative DNA binding of p53 with TFIID (TBP): a possible mechanism for transcriptional activation. Genes Dev 7:1837–1849.
124. Kelly LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858.
125. Guex N, Peitsch MC (1997) SWISS–MODEL and the Swiss–Pdb Viewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723.
126. Wiederstein M, Sippl MJ (2007) ProSA–web: Interactive web service for the recognition of errors in three–dimensional structures of proteins. Nucleic Acids Res 35:W407–W410.
127. Zargarian L, Tilly VL, Jamin N, Chaffotte A, Gabrielsen OS, Toma F, Alpert B (1999) Myb–DNA recognition: role of tryptophan residues and structural changes of the minimal DNA binding domain of c–Myb. Biochemistry 38:1921–1929.
128. Larsson T, Wedborg M, Turner D (2007) Correction of inner–filter effect in fluorescence excitation–emission matrix spectrometry using Raman scatter. Anal Chim Acta 583:357–363.
129. Lehrer SS (1971) Solute perturbation of protein fluorescence. The quenching of the tryptophyl fluorescence of model compounds and of lysozyme by iodide ion. Biochemistry 10:3254–3263.
130. Kuriata A, Gierut AM, Oleniecki T, Ciemny MP, Kolinski A, Kurcinski M, Kmiecik S (2018) CABS–flex 2.0: a web server for fast simulations of flexibility of protein structures. Nucleic Acids Res 46:W338–W343.
131. Jamroz M, Orozco M, Kolinski A, Kmiecik S (2013) Consistent view of protein fluctuations from all–atom molecular dynamics and coarse–grained dynamics with knowledge–based force–field. J Chem Theory Comput 9:119–125.
132. Koliński A (2004) Protein modeling and structure prediction with a reduced representation. Acta Biochim Pol 51:349–371.
133. Baek M, Park T, Heo L, Park C, Seok C (2017) GalaxyHomomer: a web server for protein homo–oligomer structure prediction from a monomer sequence or structure. Nucleic Acids Res 45(W1):W320–W324.
134. Tina KG, Bhadra B, Srinivasan N (2007) PIC: Protein Interactions Calculator. Nucleic Acids Res 35:W473–W476.
135. Holm L (2019) Benchmarking fold detection by DaliLite v.5. Bioinformatics 35:5326–5327.
136. Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite V.3. Bioinformatics 24:2780–2781.
137. Pei J, Kim BH, Grishim NV (2008) PROMALS3D: a tool for multiple sequence and structure alignment. Nucleic Acids Res 36:2295–2300.
138. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA–X: Molecular evolutionary genetic analysis across computing platforms. Mol Biol Evol 35:1547–1549.

139. Xu Y, Wang S, Hu Q, Gao S, Ma X, Zhang W, Shen Y, Chen F, Lai L, Pei J (2018) CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. Nucleic Acids Res 46:W374–W379.
140. Konc J, Miller BT, Stular T, Lesnik S, Woodcock HL, Brooks BR, Janezic D (2015) ProBiS–CHARMMing: Web Interface for prediction and optimization of ligands in protein binding sites. J Chem Inf Model 55:2308–2314.
141. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J Comp Chem 31:455–461.
142. Laskowski RA, Swindells MB (2011) LigPlot^{+}: multiple ligand–protein interaction diagrams for drug discovery. J Chem Inf Model 51:2778–2786.
143. Sousa da Silva AW, Vranken WF (2012) ACPYPE –AnteChamber PYthon Parser interfacE. BMC Res Notes 5:367.
144. Lindorff–Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE (2010) Improved side–chain torsion potentials for the Amber ff99SB protein force field. Proteins 78:1950–1958.
145. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 33–8:27–28.
146. Valdés–Tresanco MS, Valdés–Tresanco ME, Valiente PA, Moreno E (2021) gmx_MMPBSA: A new tool to perform end–state free energy calculations with GROMACS. J Chem Theory Comput 17:6281–6291.
147. Miller BR 3rd, McGee TD Jr, Swails JM, Homeyer N, Gohlke H, Roitberg AE (2012) MMPBSA.py: An efficient program for end–state free energy calculations. J Chem Theory Comput 8:3314–3321.
148. Zheng H, Lu L, Wang B, Pu S, Zhang X, Zhu G, Shi W, Zhang L, Wang H, Wang S, Zhao G, Zhang Y (2008) Genetic Basis of Virulence Attenuation Revealed by Comparative Genomic Analysis of *Mycobacterium tuberculosis* Strain H37Ra versus H37Rv. PLoS ONE 3(6): e2375.
149. Gong R, Li S (2014) Extraction of human genomic DNA from whole blood using a magnetic microsphere method. Int J Nanomedicine 9:3781–3789.
150. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press, pp. 571–607.
151. Gräslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S. et al. (2008) Protein production and purification. Nat Methods 5:135–146.
152. Goldstone RM, Moreland NJ, Bashiri G, Baker EN, Shaun Lott J (2008) A new Gateway vector and expression protocol for fast and efficient recombinant protein expression in *Mycobacterium smegmatis*. Protein Expr Purif 57:81–87.
153. Yamaguchi H, Miyazaki M (2014) Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. Biomolecules 4:235–251.
154. DiRusso CC, Heimert TL, Metzger AK (1992) Characterization of FadR, a global transcriptional regulator of fatty acid metabolism in *Escherichia coli*: Interaction with the *fadB* promoter by long chain fatty acyl Coenzyme–As. J Biol Chem 267:8685–8691.

155. Rao M, Liu H, Yang M, Zhao C, He ZG (2012) A copper–responsive global repressor regulates expression of diverse membrane-associated transporters and bacterial drug resistance in mycobacteria. J Biol Chem 287:39721–39731.
156. Xu Y, Heath RJ, Li Z, Rock CO, White SW (2001) The FadR•DNA complex: Transcriptional control of fatty acid metabolism in *Escherichia coli*. J Biol Chem 276:17373–17379.
157. Zheng M, Cooper DR, Grossoehme NE, Yu M, Hung LW, Cieslik M, Derewenda U, Lesley SC, Wilson IA, Giedroc DP, Derewenda ZS (2009) Structure of *Thermotoga maritima* TM0439: implications for the mechanism of bacterial GntR transcriptional regulators with Zn^{2+}–binding FCD domains. Acta Cryst D 65:356–365.
158. Lord DM, Baran AU, Soo VWC, Wood TK, Peti W, Page R (2014) McbR/YncC: Implications for the mechanism of ligand and DNA binding by a bacterial GntR transcriptional regulator involved in biofilm formation. Biochemistry 53:7223–7231.
159. Gao YG, Suzuki H, Itou H, Zhou Y, Tanaka Y, Wachi M, Watanabe N, Tanaka I, Yao M (2008) Structural and functional characterization of the LldR from *Corynebacterium glutamicum*: a transcriptional repressor involved in L–lactate and sugar utilization. Nucleic Acids Res 36:7110–7123.
160. Suvorova IA, Korostelev YD, Gelfand MS (2015) GntR family of bacterial transcription factors and their DNA binding motifs: structure, positioning and co–evolution. PLoS ONE. 10, e0132618.
161. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM (2005) The many faces of the helix–turn–helix domain: Transcription regulation and beyond. FEMS Microbiol Rev 29:231–262.
162. Luscombe NM, Laskowski RA, Thornton JM (2001) Amino acid–base interactions: a three–dimensional analysis of protein–DNA interactions at an atomic level. Nucleic Acids Res 29:2860–2874.
163. Bowie JU, Luthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. Science 253:164–170.
164. Colovos, C. and Yeates, T. O. (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein. Sci. 2, 151–1519.
165. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK – a program to check the stereochemical quality of protein structures. J App Cryst 26:283–291.
166. Roy S, Maheshwari N, Chauhan R, Sen NK, Sharma A (2011) Structure prediction and functional characterization of secondary metabolite proteins of Ocimum. Bioinformation 6:315–319.
167. Eftink MR, Ghiron CA (1981) Fluorescence quenching studies with proteins. Anal Biochem 114:199–227.
168. Vivian JT, Callis PR (2001) Mechanisms of tryptophan fluorescence shifts in proteins. Biophys J 80:2093–2109.
169. Bromberg Y, Rost B (2009) Correlating protein function and stability through the analysis of single amino acid substitutions. BMC Bioinformatics 10:S8.
170. Teng S, Wang L, Srivastava AK, Schwartz CE, Alexov E (2011) Structural assessment of the effects of amino acid substitutions on protein stability and protein–protein interaction. Int J Comput Biol Drug Des 3:334–349.

171. Miller S, Janin J, Lesk A, Chothia C (1987) Interior and surface of monomeric proteins, J Mol Biol 196:641–656.
172. Jana B, Bandhu A, Mondal R, Biswas A, Sau K, Sau S (2012) Domain structure and denaturation of a dimeric Mip–like peptidyl–prolyl cis–trans isomerase from *Escherichia coli*. Biochemistry 51:1223–1237.
173. Tisi LC, Evans PA (1995) Conserved structural features on protein surfaces: Small exterior hydrophobic clusters. J Mol Biol 249:251–258.
174. Frigerio F, Margarit I, Nogarotto R, de Filippis V, Grandi G (1996) Cumulative stabilizing effects of hydrophobic interactions on the surface of the neutral protease from *Bacillus subtilis*. Protein Eng 9:439–445.
175. Kannan N, Vishveshwara S (2000) Aromatic clusters: A determinant of thermal stability of thermophilic proteins. Protein Eng 13:753–761.
176. Karplus M, Petsko GA (1990) Molecular dynamics simulations in biology. Nature 347:631–639.
177. Childers MC, Daggett V (2017) Insights from molecular dynamics simulations for computational protein design. Mol Syst Des Eng 2:9–33.
178. Kmiecik S, Kouza M, Badaczewska–Dawid AE, Kloczkowski A, Kolinski A (2018) Modeling of protein structural flexibility and large–scale dynamics: Coarse–Grained simulations and elastic network models. Int J Mol Sci 19:3496.
179. Teichmann M, Dumay–Odelot H, Fribourg S (2012) Structural and functional aspects of winged–helix domains at the core of transcription complexes. Transcription 3:2–7.
180. Fernández–Tornero C, Böttcher B, Rashid UJ, Steuerwald U, Flörchinger B, Devos DP, Lindner D, Muller CW (2010) Conformational flexibility of RNA polymerase III during transcriptional elongation. EMBO J 29:3762–3772.
181. Clubb RT, Mizuuchi M, Huth JR, Omichinski JG, Savilahti H, Mizuuchi K, Clore GM, Gronenborn AM (1996) The wing of the enhancer–binding domain of Mu phage transposase is flexible and is essential for efficient transposition. Proc Natl Acad Sci USA 93:1146–1150.
182. Milano T, Gulzar A, Narzi D, Guidoni L, Pascarella S (2017) Molecular dynamics simulation unveils the conformational flexibility of the interdomain linker in the bacterial transcriptional regulator GabR from *Bacillus subtilis* bound to pyridoxal 5'–phosphate. PLoS ONE 12:e0189270.
183. Komeiji Y, Uebayasi M (1999) Change in fonformation by DNA–peptide association: molecular dynamics of the Hin–recombinase–hixl complex. Biophys J 77:123–138.
184. Ghosh S, Bagchi A (2019) Structural study to analyze the DNA–binding properties of DsrC protein from the *dsr* operon of sulfur–oxidizing bacterium *Allochromatium vinosum*. J Mol Model 25:74.
185. Kachhap S, Singh N (2015) Role of DNA conformation & energetic insights in Msx–1–DNA recognition as revealed by molecular dynamics studies on specific and nonspecific complexes. J Biomol Struct Dyn 33:2069–2082.
186. Boehr DD, D'Amico RN, O'Rourke KF (2018) Engineered control of enzyme structural dynamics and function. Protein Sci 27:825–838.
187. Ko J, Park H, Seok C (2012) GalaxyTBM: template–based modeling by building a reliable core and refining unreliable local regions. BMC Bioinformatics 13:198.

188. Zhang Y, Skolnick J (2005) TM–align: a protein structure alignment algorithm based on the TM–score. Nucleic Acids Res 33:2302–2309.
189. Soding J (2005) Protein homology detection by HMM–HMM comparison. Bioinformatics 21:951–960.
190. Sippl MJ (1993) Recognition of errors in three–dimensional structures of proteins. Proteins 17:355–362.
191. Ramachandran GN, Ramachandran C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7:95–99.
192. Flory PJ (1969) Statistical mechanics of chain molecules. Wiley & Sons. New York, USA, pp. 30–31.
193. Tanford C (1968) Protein denaturation. Adv Protein Chem 23:121–282.
194. Sippl MJ (1990) Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge–based prediction of local structures in globular proteins. J Mol Biol 213:859–883.
195. Sippl MJ (1995) Knowledge-based potentials for proteins. Curr Opin Struct Biol 5:229–235.
196. Sippl MJ (1993) Boltzmann's principle, knowledge–based mean fields and protein folding. An approach to the computational determination of protein structures. J Comput Aided Mol Des 7:473–501.
197. Jones S, Thornton JM (1996) Principles of protein–protein interactions. Proc Natl Acad Sci USA 93:13–20.
198. Keskin O, Gursoy A, Ma B, Nussinov R (2008) Principles of protein–protein interactions: What are the preferred ways for proteins to interact? Chem Rev 108:1225–1244.
199. Tsai CJ, Lin SL, Wolfson HJ, Nussinov R (1997) Studies of protein–protein interfaces: A statistical analysis of the hydrophobic effect. Protein Sci 6:53–64.
200. Tsai CJ, Nussinov R (1997) Hydrophobic folding units at protein–protein interfaces: Implications to protein folding and to protein–protein association. Protein Sci 6:1426–1437.
201. Clarke ND, Beamer KJ, Goldberg HF, Berkower C, Pabo CO (1991) The DNA binding arm of lambda repressor: critical contacts form a flexible region. Science 254:267–270.
202. Das AK, Kumar VA, Sevalkar RR, Bansal R, Sarkar D (2013) Unique N–terminal arm of *Mycobacterium tuberculosis* PhoP protein plays an unusual role in its regulatory function. J Biol Chem 288:29182–29192.
203. Palena CM, Tron AE, Bertoncini CW, Gonzalez DH, Chan RL (2001) Positively charged residues at the N–terminal arm of the homeodomain are required for efficient DNA binding by homeodomain–leucine zipper proteins. J Mol Biol 308:39–47.
204. Miyazono K, Zhi Y, Takamura Y, Nagata K, Saigo K, Kojima T, Tanokura M (2010) Cooperative DNA–binding and sequence–recognition mechanism of aristaless and clawless. EMBO J 29:1613–1623.
205. Domsic JF, Chen H-S, Lu F, Marmorstein R, Lieberman PM (2013) Molecular basis for oligomeric–DNA binding and episome maintenance by KSHV LANA. PLoS Pathog 9(10):e1003672.
206. Tioni MF, Viola IL, Chan RL, Gonzalez DH (2005) Site–directed mutagenesis and footprinting analysis of the interaction of the sunflower KNOX protein HAKN1 with DNA. FEBS J 272:190–202.

207. Hart BR, Mishra PK, Lintner RE, Hinerman JM, Herr AB, Blumenthal RM (2011) Recognition of DNA by the helix–turn–helix global regulatory protein Lrp is modulated by the amino terminus. J Bacteriol 193:3794–3803.
208. Pinheiro J, Lisboa J, Pombinho R, Carvalho F, Carreaux A, Brito C, Pontinen A, Korkeala H, dos Santos NMS, Morais–Cabral JH, Sousa S, Cabanes D (2018) MouR controls the expression of the *Listeria monocytogenes* Agr system and mediates virulence Nucleic Acids Res 46:9338–9352.
209. Van Aalten DMF, Dirusso CC, Knudsen J (2001) The structural basis of acyl Coenzyme A–dependent regulation of the transcription factor FadR. EMBO J 20:2041–2050.
210. Moodie SL, Mitchell JB, Thornton JM (1996) Protein recognition of adenylate: an example of a fuzzy recognition template. J Mol Biol 263:486–500.
211. Denessiouk KA, Rantanen VV, Johnson MS (2001) Adenine recognition: a motif present in ATP–, CoA–, NAD–, NADP–, and FAD–dependent proteins. Proteins 44:282–291.
212. Shulman–Peleg A, Nussinov R, Wolfson HJ (2004) Recognition of functional sites in protein structures. J Mol Biol 339:607–633.
213. Naderi M, Lemoine JM, Govindaraj RG, Kana OZ, Feinstein WP, Brylinski M (2019) Binding site matching in rational drug design: algorithms and applications. Brief Bioinform 20:2167–2184.
214. Konc J, Janezic D (2010) ProBis algorithm for detection of structurally similar protein binding sites by local structural alignment. Bioinformatics 26:1160–1168.
215. Schoch GA, Yano JK, Sansen S, Dansette PM, Stout CD, Johnson EF (2008) Determinants of cytochrome P450 2C8 substrate binding: structures of complexes with montelukast, troglitazone, felodipine, and 9–cis–retinoic acid. J Biol Chem 283:17227–17237.
216. Daily EB, Aquilante CL (2009) Cytochrome P450 2C8 pharmacogenetics: a review of clinical studies. Pharmacogenomics 10:1489–1510.
217. Saad JS, Ablan SD, Ghanam RH, Kim A, Andrews K, Nagashima K, Soheilian F, Freed EO, Summers MF (2008) Structure of the myristylated human immunodeficiency virus type 2 matrix protein and the role of phosphatidylinositol– (4,5) –bisphosphate in membrane targeting. J Mol Biol 382:434–447.
218. Schuetz A, Min J, Antoshenko T, Wang CL, Allali–Hassani A, Dong A, Loppnau P, Vedadi M, Bochkarev A, Sternglanz R, Plotnikov AN (2007) Structural basis of inhibition of the human NAD^+–dependent deacetylase SIRT5 by suramin. Structure 15:377–389.
219. Du Y, Hu H, Hua C, Du K, Wei T (2018) Tissue distribution, subcellular localization and enzymatic activity analysis of human SIRT5 isoforms. Biochem Biophys Res Commun 503:763–769.
220. Rajagopalan S, Wang C, Yu K, Kuzin AP, Richter F, Lew S, Miklos AE, Matthews ML, Seetharaman J, Su M, Hunt JF, Cravatt BF, Baker D (2014) Design of activated serine–containing catalytic triads with atomic–level accuracy. Nat Chem Biol 10:386–391.
221. Li S, Tietz DR, Rutaganira FU, Kells PM, Anzai Y, Kato F, Pochapsky TC, Sherman DH, Podust LM (2012) Substrate recognition by the multifunctional Cytochrome P450 Mycg in mycinamicin hydroxylation and epoxidation reactions. J Biol Chem 287:37880–37890.
222. Hassan SS, Cramer WA (2014) Internal lipid architecture of the hetero–oligomeric cytochrome b6f complex. Structure 22:1008–1015.

223. Gadola SD, Zaccai NR, Harlos K, Shepherd D, Castro-Palomino JC, Ritter G, Schmidt RR, Jones EY, Cerundolo V (2002) Structure of human CD1b with bound ligands at 2.3 Å, a maze for alkyl chains. Nat Immunol 3:721–726.
224. Cherezov V, Clogston J, Papiz MZ, Caffrey M (2006) Room to move: Crystallizing membrane proteins in swollen lipidic mesophases. J Mol Biol 357:1605–1618.
225. Mazor Y, Nataf D, Toporik H, Nelson N (2014) Crystal structures of virus–like photosystem I complexes from the mesophilic cyanobacterium *Synechocystis* PCC 6803. Elife 3:e01496.
226. Schulte T, Sharples FP, Hiller RG, Hofmann E (2009) X–Ray structure of the high–salt form of the Peridinin–Chlorophyll a–protein from the Dinoflagellate *Amphidinium Carterae*: Modulation of the spectral properties of pigments by the protein environment. Biochemistry 48:4466–4475.
227. Bishai W (2000) Lipid lunch for persistent pathogen. Nature 406:683–685.
228. McKinney JD, Honer zu Bentrup K, Muñoz-Elias EJ, Miczak A, Chen B, Chan WT, Swenson D, Sacchettini JC, Jacobs Jr WR, Russell DG (2000) Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. Nature 406:735–738.
229. Marrero J, Rhee KY, Schnappinger D, Pethe K, Ehrt S (2010) Gluconeogenic carbon flow of tricarboxylic acid cycle intermediates is critical for *Mycobacterium tuberculosis* to establish and maintain infection. Proc Natl Acad Sci USA 107:9819–9824.
230. Kim MJ, Wainwright HC, Locketz M, Bekker LG, Walther GB, Dittrich C, Visser A, Wang W, Hsu FF, Wiehart U, Tsenova L, Kaplan G, Russell DG (2010) Caseation of human tuberculosis granulomas correlates with elevated host lipid metabolism. EMBO Mol Med 2:258–274.
231. Cha D, Cheng D, Liu M, Zeng Z, Hu X, Guan W (2009) Analysis of fatty acids in sputum from patients with pulmonary tuberculosis using gas chromatography–mass spectrometry preceded by solid–phase microextraction and post–derivatization on the fiber. J Chromatogr A 1216:1450–1457.
232. Mourão MPB, Denekamp I, Kuijper S, Kolk AHJ, Janssen HG (2016) Hyphenated and comprehensive liquid chromatography × gas chromatography–mass spectrometry for the identification of *Mycobacterium tuberculosis*. J Chromatogr A 1439:152–160.
233. Forrellad MA, McNeil M, de la Paz Santangelo M, Blanco FC, Garcia E, Klepp LI, Huff J, Niederweis M, Jackson M, Bigi F (2014) Role of the Mce1 transporter in the lipid homeostasis of *Mycobacterium tuberculosis*. Tuberculosis 94:170–177.
234. Zhang X, Perez–Sanchez H, Lightstone FC (2017) A comprehensive docking and MM/GBSA rescoring study of ligand recognition upon binding antithrombin. Curr Top Med Chem 17:1631–1639.
235. Beveridge DL, DiCapua FM (1989) Free energy via molecular simulation: applications to chemical and biomolecular systems. Annu Rev Biophys Biophys Chem 18:431–492.
236. Zhang X, Péréz–Sánchez H, Lightstone F (2015) Molecular dynamics simulations of ligand recognition upon binding antithrombin: A MM/GBSA approach. In: Ortuño F, Rojas I (ed) Bioinformatics and biomedical engineering. Vol. 9044. Springer International Publishing, Switzerland, pp 584–593.
237. Verma S, Grover S, Tyagi C, Goyal S, Jamal S, Singh A, Grover A (2016) Hydrophobic interactions are a key to MDM2 inhibition by polyphenols as revealed by molecular dynamics simulations and MM/PBSA free energy calculations. PLoS ONE 11(2):e0149014.

238. Maity D, Singh D, Bandhu A (2023) Mce1R of *Mycobacterium tuberculosis* prefers long–chain fatty acids as specific ligands: a computational study. Mol Divers 6:2523–2543.
239. Ó'Fágáin C, Cummins PM, O'Connor BF (2017) Gel–filtration chromatography. Methods Mol Biol 1485:15–25.

PUBLICATIONS

Research publications from this present work

1. Maity D, Katreddy RR, Bandhu A (2021) Molecular cloning, purification and characterization of Mce1R of *Mycobacterium tuberculosis*. Mol Biotechnol 63:200-220.
2. Maity D, Singh D, Bandhu A (2022) Mce1R of *Mycobacterium tuberculosis* prefers long-chain fatty acids as specific ligands: a computational study. Mol Divers. https://doi.org/10.1007/s11030-022-10566-7.