Structural and Functional Studies of Metal Ion ABC Transporters in Prokaryotes

A Thesis

Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

by

Suraj Kumar Mandal
(146106004)

Under supervision of

Shankar Prasad Kanaujia
Professor

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

December 2021

Dedicated to

My beloved

Grandparents and Parents

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING

DECLARATION

I do hereby declare that the research findings of this thesis entitled "**Structural and Functional Studies of Metal Ion ABC Transporters in Prokaryotes**" is the result of research work carried out by me in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati, India, under the supervision of **Prof. Shankar Prasad Kanaujia**.

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

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

Date: 02-12-2021

Place: IIT Guwahati

Suraj Kumar Mandal

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING

CERTIFICATE

This is to certify that the work described in the thesis entitled "**Structural and Functional Studies of Metal Ion ABC Transporters in Prokaryotes**" is an authentic record of the results obtained from the research work carried out by **Suraj Kumar Mandal** in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Assam, India, under my supervision and this work in part or as a whole has not been submitted elsewhere for the award of any other degree.

Shankar A. Kanaujia

Date: 02-12-2021 **Prof. Shankar Prasad Kanaujia**

Place: IIT Guwahati (Thesis Supervisor)

ACKNOWLEDGEMENTS

The Ph.D. work presented in this thesis involved the contribution of many people. During my Ph.D. tenure, I have worked with a number of people whose support and guidance made this possible. It is a pleasure to express my gratitude to all of them.

With deepest respect and esteem regards, I owe a profound gratitude to my Ph.D. advisor, **Prof. Shankar Prasad Kanaujia**, for providing me an opportunity to place my first step towards accomplishing a remarkable thing in the field of structural biology. Without his persistent guidance, tremendous support and encouragement, this thesis would not have seen the light of the day. His unwavering enthusiasm for science and active involvement in acquiring and providing all the required facilities made this work possible for me. His words and actions taught me "to aim high and to stay grounded" which I would surely carry forward in my career. Thank you Sir.

I would like to express my sincere thanks to the doctoral committee members: **Prof. Kannan Pakshirajan, Prof. Sachin Kumar** and **Prof. Biplab Mandal** for their valuable suggestions and healthy discussions. I acknowledge the past and present Heads of the Department of Biosciences and Bioengineering, **Prof. V. Venkata Dasu**, **Prof. Kannan Pakshirajan and Prof. Latha Rangan** for providing all the necessary departmental facilities during their tenure.

I acknowledge the infrastructural support and favorable academic environment from **Department of Biosciences and Bioengineering** at IIT Guwahati. I also owe my gratitude to the **Central Instruments Facility (CIF)**, IIT Guwahati for providing me access to the sophisticated instruments such as X-ray diffractometer and isothermal titration calorimetry including others, which were instrumental to this work.

I sincerely acknowledge the financial aid from **Ministry of Human Resource and Development (MHRD)**, Government of India for providing me fellowship and **Department of Biotechnology (DBT)**, Government of India for providing research grant to our laboratory.

Acknowledgements

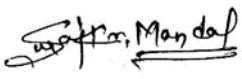
I would take this opportunity to thank the past and present members of the Structural and Computational Biology Laboratory (SCBL) for being supportive and creating a healthy environment. This work would not have been possible without the intellectual contributions from my senior fellows **Dr. Prerana Gogoi** and **Dr. Monika Chandravanshi** and therefore, I extend my deepest gratitude to them. Their continuous support and encouragement helped in with my work and I am indebted to them for sharing their experience. I also thank my fellow juniors, **Anghsu, Sayan, Pratik, Prerana, Kalyan, Sakshi and Ankita** for their cooperation and timely help. I also thank **Anjaney, Rahi** and **Satya** for their generous help to complete the *in silico* analysis.

I am fortunate to have great friends and peers who supported and encouraged me to navigate through my highs and lows. I extend warm and generous thanks to my friends **Dr. Anshul Assaiya, Jinti Hazarika, Dr. Faheem, Dr. Abshar, Dr. Tasleem, Dr. Bhuvan, Dr. Bapi, Dilip, Madan** and **Janardhan** who were always there when I needed them.

Above ground, I am indebted to my parents, **Mr. Santosh Kumar Mandal** and **Mrs. Mamata Mandal** and sisters **Radhika Mandal** and **Menaka Mandal** whose value to me only grow with age. This dissertation would not have been possible without their unconditional love, encouragement, patience and endurance. Their faith in me and sacrifices have made this journey possible. Special thanks to my brother-in-law's **Amit Mandal** and **Ajay Mandal** for taking care of our family, while I pursued my Ph.D. and for boosting my morale.

At last, I would like to thank everybody who was involved as well as expressing my apology if I have missed anyone.

Last but not the least, I would like to devote my wholehearted gratitude to the Almighty God for his blessings to achieve the goals.



Suraj Kumar Mandal

December, 2021

TABLE OF CONTENTS

Section	Content	Page No.
	Abstract	xi
	Abbreviations and notations	xvii
	List of figures	xix
	List of tables	xxiii
	Chapter 1 - Introduction	1
1.	Introduction	3
1.1.	Importance of metal ions	3
1.1.1	Nutritional immunity	4
1.1.2	Metal ion transporters in prokaryotes	5
1.2	ATP-binding cassette (ABC) transporters	9
1.2.1	ABC importers	11
1.2.1.1	Transmembrane domain (TMD)	14
1.2.1.2	Nucleotide-binding domain (NBD)	16
1.2.1.3	Substrate (solute)-binding domain (SBP)	18
1.2.1.4	Classification of substrate (solute)-binding proteins (SBPs)	20
1.2.2	Mechanism of substrate translocation by ABC transporter	24
1.3	Significance of the study	25
1.4	Objectives	27
	Chapter 2 – Materials and methods	29
	General outline of chapter 2	31
2.1.	Materials	31
2.1.1.	Reagents	31
2.1.2.	Metal ions and their complexes	32
2.2.	Methods	32
2.2.1.	Sequence- and structure-based analyses	32
2.2.1.1.	Identification of ORFs and retrieval of sequences	32
2.2.1.2.	Sequence-based analysis	33

2.2.1.2.1.	Identification of homologous proteins and sequence alignment	33
2.2.1.2.2.	Phylogenetic tree	33
2.2.1.3.	Structure-based analysis	34
2.2.1.3.1.	Structure prediction	34
2.2.1.3.2.	Structural comparison with homologous proteins	34
2.2.1.3.3.	Structure-based classification of SBPs	35
2.2.1.3.4.	Surface analysis	35
2.2.2.	Molecular docking	36
2.2.3.	Recombinant construction for protein overexpression	37
2.2.3.1.	Competent cell preparation for gene cloning and protein overexpression	37
2.2.3.2.	Gene amplification	38
2.2.3.3.	Double digestion of the gene of interest and vector	39
2.2.3.4.	Ligation of the gene of interest and clone confirmation	39
2.2.3.5.	Site-directed mutagenesis	40
2.2.4.	Protein overexpression, solubilization and purification	40
2.2.4.1.	Overexpression of the selected SBPs	40
2.2.4.2.	Solubility of protein	41
2.2.4.3.	Affinity chromatography for protein purification	41
2.2.5.	Protein characterization	42
2.2.5.1.	Mass spectrometric analysis	42
2.2.5.2.	Isothermal titration calorimetry	42
2.2.5.3.	Fluorescence spectroscopy	43
2.2.5.4.	Circular dichroism spectroscopy	44
2.2.5.5.	Difference absorption spectroscopy	44
2.2.5.6.	Energy Dispersive X-ray analysis	44
2.2.6.	Crystallization of the selected SBPs	45
2.2.7.	Data collection and processing	46
2.2.7.1.	Data collection	46
2.2.7.2.	Data processing	47

2.2.7.3.	Structure determination	47
2.2.7.4.	Model building and structure refinement	48
2.2.7.5.	Cross-validation	49
2.2.8.	Structure validation	49
2.2.8.1	PROCHECK	49
2.2.8.2.	MolProbity	50
2.2.9.	Domain movement analysis of SBPs	50
2.2.10.	Structure and protein-ligand interaction visualization	50
	Chapter 3 – Metal uptake ABC transporters in *T. thermophilus* HB8	51
	Abstract	53
	Graphical abstract	54
3.1.	Introduction	54
3.2.	Materials and methods	56
3.2.1.	Sequence analysis	56
3.2.2.	Structure analysis	56
3.2.3.	Molecular docking studies	57
3.3.	Results	58
3.3.1.	Metal uptake ABC transporters in *T. thermophilus* HB8	58
3.3.2.	Metal uptake SBPs cluster into functionally distinct clades	59
3.3.3.	Thermophilic metal uptake SBPs are co-localized with their cognate TMDs and NBDs, except TTHA0723	62
3.3.4.	Metal uptake SBPs from *T. thermophilus* HB8 belong to clusters A and D	64
3.3.5.	Active-site pocket of metal-binding proteins are fine-tuned to interact with their cognate substrates	66
3.3.6.	TTHB220 and TTHA0723 might potentially function as heme-binding proteins	68
3.3.7.	TTHA0746 is potential cyanocobalamin- rather than an iron-binding protein	70
3.3.8.	TTHB177 is potentially Mg/Fe^{2+}-citrate-binding protein	73

Table of contents

3.3.9.	TTHA0069 and TTHA0717 are potential WO_4^{2-}- and MoO_4^{2-}-binding proteins, respectively	75
3.3.10.	The protein TTHA0596 is a potential Zn^{2+}-binding protein	77
3.4.	Discussion	80
	Chapter 4 – Metal uptake ABC transporters in *M. tuberculosis* H37Rv	85
	Abstract	87
	Graphical abstract	88
4.1.	Introduction	88
4.2.	Materials and methods	90
4.2.1.	Data collection and sequence analysis	90
4.2.2.	Structure analysis	91
4.2.3.	Molecular docking	92
4.3.	Results	93
4.3.1.	Repertoire of the metal uptake ABC transporters and their genetic organization in *M. tuberculosis* H37Rv and *E. coli* K-12	93
4.3.2.	The probable metal-specific SBPs and their cognate NBDs cluster into functionally distinct clades	96
4.3.3.	The proteins Rv0265c and Rv3044 belong to a novel subcluster A-II-b of SBPs	100
4.3.4.	Active-site pocket of the metal uptake SBPs are quintessential to accommodate their cognate substrates	101
4.3.5.	The proteins Rv0265c, Rv2895c and Rv3044 code for iron-uptake ABC transporters	103
4.3.6.	The protein Rv1857 is a plausible molybdate-binding protein	106
4.3.7.	The protein Rv2059 is a potential Zn^{2+}-binding protein	108
4.3.8.	The proteins Rv1859, Rv2326c and Rv3041c conserve all the intrinsic features of a typical nucleotide-binding domain (NBD) of ABC importers	110

4.3.9.	The proteins Rv0908 and Rv1030 are P-type ATPases associated with metal ion transport	112
4.4.	Discussion	114
	Chapter 5 – Structural and functional characterization of the protein MctA	119
	Abstract	121
	Graphical abstract	122
5.1.	Introduction	122
5.2.	Materials and methods	125
5.2.1.	Construction of wild-type and mutant expression plasmids	125
5.2.2.	Protein over expression and purification	127
5.2.3.	Mass spectrometric analysis	130
5.2.4.	Crystallization of the wild-type and mutant MctA proteins	130
5.2.5.	Data collection, processing and structure determination	133
5.2.6.	Energy Dispersive X-ray analysis	141
5.2.7.	Circular dichroism for the wild type and mutant proteins of MctA	142
5.2.8.	Isothermal titration calorimetry	142
5.2.9.	Bioinformatic analysis	143
5.3.	Results	143
5.3.1.	Overall structure of the protein MctA	143
5.3.2.	The protein MctA is poly-specific towards divalent metal ions	147
5.3.3.	Physiological ligand of the protein MctA	149
5.3.4.	The protein MctA is a novel divalent metal ion-complexed citrate transporter belonging to ABC transporter superfamily	152
5.3.5.	The active-site pockets of the protein MctA and its homolog Atu2467 are made for similar ligands	155
5.3.6.	The residues Thr55 and Arg246 are crucial for maintaining the overall fold of the protein MctA	157

5.3.7.	Mutation of active-site residues possessing aliphatic hydroxyl group alters ligand specificity	159
5.3.8.	The residue Asp28 is important for metal binding, whereas mutation in Tyr221 confers an apo state of the protein MctA	160
5.3.9.	The protein MctA follows a gating mechanism of ligand entry	164
5.3.10.	The protein MctA follows an "asymmetric domain movement" mechanism of ligand binding	166
5.4.	Discussion	168
	Chapter 6 – Structural and functional characterization of the protein MhuP	171
	Abstract	173
	Graphical abstract	174
6.1.	Introduction	174
6.2.	Materials and methods	176
6.2.1.	Construction of wild-type expression plasmids	176
6.2.2.	Protein overexpression and purification	178
6.2.3.	Mass spectrometric analysis	180
6.2.4.	Crystallization and co-crystallization of the protein MhuP	180
6.2.5.	Data collection, processing and structure determination	181
6.2.6.	Circular dichroism analysis	184
6.2.7.	Isothermal titration calorimetry	185
6.2.8.	Difference absorption spectroscopy	185
6.2.9.	Fluorescence spectroscopy	186
6.2.10.	Molecular docking and other bioinformatic analyses	186
6.3.	Results	187
6.3.1.	The protein MhuP is stable at both compositional and conformational levels	187
6.3.2.	Overall structure of the protein MhuP	190
6.3.3.	The protein MhuP is evolutionarily closer to Fe^{3+}-siderophore-binding protein	192

6.3.4.	Active site of the protein MhuP is favourable for hydroxamate-type siderophores	194
6.3.5.	Difference absorption spectroscopy suggest the protein MhuP can bind hemin	196
6.3.6.	The protein MhuP shows specificity towards heme molecules	198
6.3.7.	Docking studies suggest a strong affinity of the protein MhuP towards the heme	200
6.3.8.	Protein-protein docking suggests a transient t of heme from the protein MhuP to DppA	203
6.4.	Discussion	205
	Conclusions and Summary	209
	Appendix-A - Supplementary data to chapter 3	215
	Appendix-B - Supplementary data to chapter 4	249
	Appendix-C - Supplementary data to chapter 5	279
	Appendix-D - Supplementary data to chapter 6	289
	Bibliography	295
	List of publications, conferences and workshops	319

ABSTRACT

ABSTRACT

Enormous proteins function in numerous cellular processes by utilizing different metal ions to accomplish their designated structural and/or functional roles. The alkali earth metals such as magnesium & calcium and d-block or transition-group metal ions such as manganese, iron, cobalt, nickel, zinc and to some extent, molybdenum and tungsten are principally involved in playing crucial roles in biological processes. Metal ions are essential for a plethora of cellular or physiological processes, including DNA replication, protein biosynthesis, cell wall biosynthesis, cellular respiration, signal transduction, etc. However, alteration (either increase or decrease) in the cellular concentration of metal ions turns out to be toxic for microorganisms and therefore, metal ion homeostasis becomes an important cellular phenomenon. In prokaryotes, several high-affinity transport machinery govern the metal ion homeostasis and help in scavenging these essential metal ions from different environmental niches. It exploits a myriad of mechanisms to sequester these essential metal ions from the extracellular environment. One of the mechanisms involves the use of sophisticated transport machinery popularly known as ATP-binding cassette (ABC) transporters to import and export the substrates. ABC transporters are conserved across all domains of life and are fueled by the energy from ATP hydrolysis to facilitate the translocation of substrate(s) across cellular membranes. Architecturally, an ABC transport system consists of two transmembrane domains (TMDs) which form the transport pore and two nucleotide-binding domains (NBDs) where ATP hydrolysis occurs to generate the energy during substrate translocation across the plasma membrane. In addition, the prokaryotic ABC-importers possess an extracellular domain known as substrate (or solute)-binding protein (SBP) for nutrient acquisition inside the cell. In the periplasm, SBPs recognize and bind their specific or cognate ligand with high affinity and facilitate their transport inside the cytoplasm via the cognate inner membrane component, i.e., TMD. Due to the involvement of ABC importers in multiple functions necessary for cell physiology, it transports several kinds of (micro)nutrients into the cytoplasm via SBPs. Based on the type of substrates being transported, SBPs are named as metal ion-, siderophore-, sugar-, amino acid-, vitamin-, di- & oligopeptide and sugar-binding proteins. During substrate binding, conformational changes occur in the SBPs to accommodate the

substrate molecule in its active site following an induced-fit mechanism known as "Venus Fly-trap". Many structural and biophysical methods have been proposed in order to elucidate the mechanism of ligand binding by SBPs. However, due to the low level of sequence similarity among SBPs, understanding the mechanism of substrate recognition and binding by SBPs has remained very enigmatic. In addition to the low sequence identity, the diversity of ligands (substrates) makes it even more complex to determine the mechanism of ligand recognition by SBPs.

In the present study, we have performed a data mining approach to identify the probable ORFs, which encodes for metal-specific ABC transporters from a thermophilic and pathogenic bacteria *Thermus thermophilus* HB8 and *Mycobacterium tuberculosis* HB8, respectively. To identify the cognate metal ion for each ABC transporter, an integrated *in silico* approach has been executed to characterize the identified ORFs probably involved in metal uptake. The results enable the identification of 22 ORFs, resulting in eight ABC transport systems in *T. thermophilus*, which are potentially involved in the uptake of metal ions. This study suggests that three out of eight metal-specific ABC import systems are involved in iron transport. Among the remaining five, two are particular to divalent metal ion transport, for instance, Mg^{2+} and Zn^{2+}/Mn^{2+}; another two are for tetrahedral oxyanions such as MoO_4^{2-} and WO_4^{2-} and the remaining one imports cyanocobalamin. Similarly, in *M. tuberculosis*, 19 ORFs resulting in seven ABC transport systems and two P-type ATPases were identified, which are potentially involved in the uptake of different metal ions. The results from this study suggest that three out of the seven metal uptake ABC transporters are specific to iron complexes. Among the remaining four, one is specific to the d-block metal ion Zn^{2+}, one each for the metal oxyanion MoO_4^{2-} and cobalamin (vitamin B12) and the remaining one is an energy-coupling factor (ECF)-type ABC importer specific to Co^{2+} ion. Moreover, the results also demonstrate the existence of a subunit sharing mechanism in both the organisms where the TMDs and NBDs are shared among different ABC transporters hinting at the import of multiple substrates via a single ABC transport system. Altogether, this study combinedly reflects an overview of the repertoire of metal-specific ABC transport systems in *T. thermophilus* HB8 and *M.*

tuberculosis H37Rv, providing potential therapeutic targets for the future and for other biotechnological applications.

After identification of the complete metal uptake ABC transport systems, one ORF each from *T. thermophilus* HB8 and *M. tuberculosis* H37Rv viz., TTHB177 (referred as MctA) and Rv0265c (referred as MhuP), respectively, were further characterized using the experimental approach. The experimental studies report the crystal structures of an Mg^{2+}-citrate-binding protein MctA (ORF id: TTHB177) from the Gram-negative *Thermus thermophilus* HB8 in apo as well as holo forms at a resolution range of 1.63 to 2.50 Å. The results suggested that the physiological metal ion for the protein MctA is Mg^{2+}. However, the protein MctA is able to bind several other divalent metal ions as well. The results from this study unveil a novel sub-classification of cluster D SBPs capable of binding divalent metal ion complexed citrate molecules. The open and closed conformations of the wild-type and mutant proteins of MctA exhibit a gating mechanism of ligand entry followed by an asymmetric domain movement mechanism of ligand-binding from the N-terminal domain (NTD).

Additionally, this study also reports the crystal structures of the protein MhuP (ORF id: Rv0265c) at a resolution of 1.8 and 2.01 Å in two different space groups $P2_1$ and $C2$, respectively. The absorption and fluorescence spectrometric results exhibit a selective as well as specific binding of heme with the protein MhuP. Moreover, the docking experiments suggested the heme-binding potential of the protein MhuP as well as the plausible heme transfer mechanism between the proteins MhuP and DppA. Thus, this study confirms the binding of heme with the protein MhuP and its transient transport via the DppABCD transporter in *M. tuberculosis*.

In summary, the findings from this study provide the first ever-preliminary glimpse of the repertoire of metal uptake ABC transporters in *T. thermophilus* HB8 and *M. tuberculosis* H37Rv. The structural and functional approach to characterize the protein MctA from *T. thermophilus* HB8 reveals the existence of a novel Mg^{2+}-citrate-binding protein in Gram-negative bacteria and reports their substrate-binding mechanism. In contrast, the structural

and functional characterization of the protein MhuP from *M. tuberculosis* H37Rv exhibited the specificity of the protein towards heme. Moreover, the results also confirmed the earlier hypothesis of heme transfer between the proteins MhuP and DppA.

ABBREVIATIONS AND NOTATIONS

ABBREVIATIONS

ABC	ATP-binding cassette
CD	Circular dichroism
CTD	C-terminal domain
CV	Column-volume
DAS	Difference absorption spectroscopy
DP	Degree of polymerization
EDX	Energy dispersive X-ray
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LB	Luria-Bertani
MFS	Major facilitator superfamily
MSA	Multiple sequence alignment
N.D.	Not detected
NBD	Nucleotide-binding domain
NCBI	National Center for Biotechnology Information
NTD	N-terminal domain
ORF	Open reading frame
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonyl fluoride
rmsd	Root mean square deviation
SBP	Substrate (or solute)-binding protein
TMD	Transmembrane domain
WT	Wild-type
β-ME	β-mercaptoethanol

NOTATIONS

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

LIST OF FIGURES

Figure 1.1.	Nutritional immunity	5
Figure 1.2.	Metal ion transporters in microorganisms	7
Figure 1.3.	Overall architecture of ABC transporters	10
Figure 1.4.	Different types of ABC importers	14
Figure 1.5.	Schematic representation of the transmembrane domain (TMD) fold	16
Figure 1.6.	Multiple sequence alignment and domain architecture of the nucleotide-binding domain (NBD)	17
Figure 1.7.	Structural topology of a SBP	20
Figure 1.8.	Structure-based classification of the SBPs	23
Figure 1.9.	Schematic representation of substrate translocation via ABC importer	24
Figure 2.1.	A general work plan of the methodology used to accomplish the objectives	31
Figure 2.2.	Schematic representation for molecular cloning	37
Figure 2.3.	Schematic representation for the phase diagram of protein crystallization	46
Figure 3.1.	Phylogenetic analysis of the metal uptake SBPs along with their homologous proteins	60
Figure 3.2.	Genetic organization and protein-protein network analysis of the metal uptake SBPs	63
Figure 3.3.	The Overall three-dimensional models of the proteins and structure-based evolutionary analysis	65
Figure 3.4.	Electrostatic potential charge distribution and active-site area and volume of the metal uptake SBPs	67
Figure 3.5.	Active-site residue analysis and molecular docking of the proteins TTHB220 and TTHA0723	69
Figure 3.6.	Active-site residue analysis and molecular docking of the protein TTHA0746	72

Figure 3.7.	Comparing the spatial orientation of the acitve-site residues of the protein TTHA1628 and TTHB177 and molecular docking of the protein TTHB177	74
Figure 3.8.	Active-site residue analysis and molecular docking of the proteins TTHA0069 and TTHA0717	76
Figure 3.9.	Active-site residue analysis and their spatial orientation at the acitve-site in the protein TTHA0596	79
Figure 3.10.	Schematic representation of metal ABC transporters available in *T. thermophilus* HB8.	84
Figure 4.1.	Metal-specific ABC transport systems of *M. tuberculosis* H37Rv and *E. coli* K-12	95
Figure 4.2.	Evolutionary relationship and protein-protein interatome analysis for probable metal uptake ABC transporters in *M. tuberculosis* H37Rv	98
Figure 4.3.	Structure-based distance tree of cluster A and D SBPs	101
Figure 4.4.	Electrostatic potential charge distribution and active-site area and volume of the metal uptake SBPs	103
Figure 4.5.	Active-site residue analysis and molecular docking of the proteins Rv0265c and Rv3044 with their probable ligands	105
Figure 4.6.	Active-site residue analysis and molecular docking of the protein Rv1857 with MoO_4^{2-} and WO_4^{2-}	107
Figure 4.7.	Comparing the spatial orientation of the acitve-site residues of the protein Rv2059 with its homologous proteins	109
Figure 4.8.	Multiple sequence alignment of the proteins Rv1859, Rv2326c and Rv3041c with their homologous proteins	111
Figure 4.9.	Primary and tertiary structure analysis for the proteins Rv0908 and Rv1030	113
Figure 4.10.	Schematic representation of the repertoire of probable metal uptake ABC transporters available in *M. tuberculosis* H37Rv	117
Figure 5.1.	Cloning and site-directed mutagenesis of the gene *mctA*	127

List of Figures

Figure 5.2.	Purification of MctA_WT and mutant proteins	129
Figure 5.3.	Crystallization of the protein MctA	132
Figure 5.4.	Overall three-dimensional structure of the protein MctA	146
Figure 5.5.	Isothermal titration calorimetry of MctA_WT protein	148
Figure 5.6.	Coordination chemistry for Mg^{2+} and Mn^{2+} ion followed by MctA_WT, MctA_S26A mutant and Atu2467 protein	151
Figure 5.7.	Subclassification of the subcluster D-IV	154
Figure 5.8.	Active-site analysis of the protein MctA with its closest homolog Atu2467	156
Figure 5.9.	Circular dichroism spectra analysis and location of the residues Thr55 and Arg246 in the protein MctA	158
Figure 5.10.	Mutagenic study of the active-site residues possessing aliphatic hydroxyl group	161
Figure 5.11.	Mutagenic study of the active-site residues Asp28 and Tyr221	163
Figure 5.12.	Gating mechanism of ligand entry in the protein MctA	165
Figure 5.13.	Asymmetric domain movement confers ligand binding in the protein MctA	167
Figure 6.1.	Cloning of the gene *mhuP*	177
Figure 6.2.	Over expression, solubility and purification of the protein MhuP	179
Figure 6.3.	Crystallization of the protein MhuP	181
Figure 6.4.	Compositional and conformational stability of the protein MhuP	189
Figure 6.5.	Overall three-dimensional structure and differences in $P2_1$ and $C2$ space group structures of the protein MhuP	191
Figure 6.6.	Phylogenetic analysis of the protein MhuP with its homologous proteins along with known heme- and di- and oligo-peptide binding proteins	193
Figure 6.7.	Comparision of the active site of the protein MhuP with Fe^{3+}-siderophore and heme-binding proteins	195

List of Figures

Figure 6.8.	Difference absorption spectroscopy of the protein MhuP	197
Figure 6.9.	Fluorescence study of the protein MhuP with hemin and ferrichrome	199
Figure 6.10.	Molecular docking studies of the protein MhuP	200
Figure 6.11.	Protein-protein docking of the proteins MhuP (heme bound) and *Mt*DppA	204
Figure 6.12.	Schematic representation of heme acquisition by *M. tuberculosis*	208

LIST OF TABLES

Table 1.1.	Classification of active transporters citing some major examples	8
Table 1.2.	Metal ion uptake ABC importers involved in bacterial virulence	11
Table 1.3.	An overview of the SBP classification based on their ligand specificity	22
Table 5.1.	List of oligonucleotide sequences used for *mctA* and *mctA_mutant* recombinant constructs	125
Table 5.2.	Data collection and refinement statistics of MctA_WT protein (bound to Mg^{2+}-citrate in $P2_1$ and $P2_12_12_1$ space groups and citrate only)	134
Table 5.3.	Data collection and refinement statistics of MctA_S26A protein (bound to Mg^{2+}-citrate and Mn^{2+}-citrate) and MctA_D28A	136
Table 5.4.	Data collection and refinement statistics of MctA_S79A, MctA_S164A and MctA_T199A protein	138
Table 5.5.	Data collection and refinement statistics of MctA_Y221A and MctA_Y221F protein	139
Table 5.6.	Thermodynamic parameters of various divalent metal ions and $FeCl_3$ binding to the protein MctA_WT	149
Table 6.1.	X-ray crystallographic data collection and refinement statistics for the protein MhuP	183
Table 6.2.	Dissociation constants (K_d), Hill coefficients (n) and regression coefficient for the fluorescence spectrum for the protein MhuP with different ligands	200

CHAPTER 1

Introduction

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

1. INTRODUCTION

1.1. Importance of metal ions

More than one-third of the proteins necessitate metal ions as cofactors which renders them structural stability and ability to accomplish various catalytic activities. Apart from being involved in various biological processes such as DNA replication, protein biosynthesis, cell wall biosynthesis, cellular respiration, etc. (Kehres and Maguire, 2003; Counago et al., 2012; Andreini et al., 2013), metal ions also contribute significantly towards microbial pathogenesis (Kehl-Fie and Skaar, 2010). Among different metal ions, the *d*-block elements, preferably known as transition metal ions, are also utilized by bacteria for a variety of cellular demands (Waldron and Robinson, 2009; Kehl-Fie and Skaar, 2010). Nearly 50% of the prokaryotic including archaeal and eukaryotic proteins containing copper (Cu) and non-heme iron (Fe) are either oxidoreductases or other electron transfer proteins (Andreini et al., 2008). The heme-iron complex is also an important cofactor for several biosynthetic and metabolic processes, including respiration. In spite of Magnesium (Mg) being the most prevalent non-redox metal ion present in enzymes, zinc (Zn) is the most prevalent non-redox metal ion among transition row metal ions (Andreini et al., 2008). Both Mg and Zn are important structural and catalytic cofactors for various proteins (Andreini et al., 2008). Mg plays a crucial role in many physiological functions and is involved in several energy-requiring metabolic processes, membrane integrity, protein biosynthesis, nervous tissue conduction, neuromuscular excitability, hormone secretion, intermediary metabolism, just to name a few (Laires et al., 2004). Nearly 5-6% of the prokaryotic proteome contains zinc-binding proteins (Andreini et al., 2006). Recently, it has been proposed that the *Streptococcus pneumoniae* genome consists of more than 200 (~9% of proteome) Zn/Cu-containing proteins (Sun et al., 2011). Certainly, zinc and copper are necessary for the proper activity of many Streptococcal enzymes (Lagartera et al., 2005; Hammerschmidt, 2009). In contrast, the presence of strictly manganese (Mn)-dependent enzymes is only a handful in both prokaryotes and eukaryotes. This is probably due to the readily interchangeable characteristic of Mn with other divalent cations in metalloenzymes (Andreini et al., 2008). In prokaryotes, Mn ions are not only involved as a cofactor for free-

radical detoxifying enzymes but also involves in central carbon metabolism (Kehres and Maguire, 2003).

Although all living cells require metal ions for various catabolic and anabolic processes to survive, the catalytic activity of metal ions also leads to their toxicity depending on their concentrations inside the cell. The level of these metal ions inside the cells must be controlled to its survival. Besides, the mechanisms exploited by host systems to limit the freely available metal ions to the microorganisms residing inside them also serve as a countermeasure against the invading pathogens.

1.1.1. Nutritional immunity

The human body is a reservoir rich in essential (micro)nutrients for several microorganisms residing inside. To prevent the infection or virulence caused by the residing pathogens, humans or other host systems restrict access to these essential micronutrients, in particular, metal ions and this process of metal ion restriction to pathogens are termed as 'nutritional immunity'. Although nutritional immunity was originally coined to demonstrate the host-mediated restriction of iron ions, the term is also applied to the withholding mechanisms of other essential metal ions as well (Figure 1.1) (Weinberg et al., 1975, 2009; Hood and Skaar, 2012). In mammalian cells, transition metal ions are present, either in a protein-bound form or sequestered within the cell, and thus, limiting their accessibility to pathogens (Diaz-Ochoa et al., 2014). For example, during *S. aureus* infection, iron, the most abundant transition metal ion in the human body, is maintained intracellularly or sometimes complexed with the host's high-affinity binding proteins (Weinberg, 2009). This sequestration results in an extracellular concentration of iron which is inadequate to support the growth of *S. aureus* (Braun et al., 1998).For its proliferation, *S. aureus* must overcome this iron sequestration for a successful infection to host tissues. Accordingly, *S. aureus* and other *Staphylococcus* species have evolved ways to acquire, process and de-toxicate iron and host proteins and molecules which contain iron. Evidence from the literature suggests that the principle of nutritional immunity has also augmented to other *d*-block divalent metal ions, including manganese and zinc (Corbin et al., 2008). However, an excess concentration of metal ions can be toxic to the microorganisms as well as metal ion

deprivation can also inhibit microbial growth. Thus, efficient management and regulation of metal ion homeostasis is a critical aspect of prokaryotic chemical biology (Agranoff and Krishna, 1998; Moore and Helmann, 2005; McDevitt et al., 2011). Over time, bacteria have developed sophisticated transport systems to deal with these limitations (Figure 1.1) (Wakeman and Skaar, 2012; Porcheron et al., 2013).

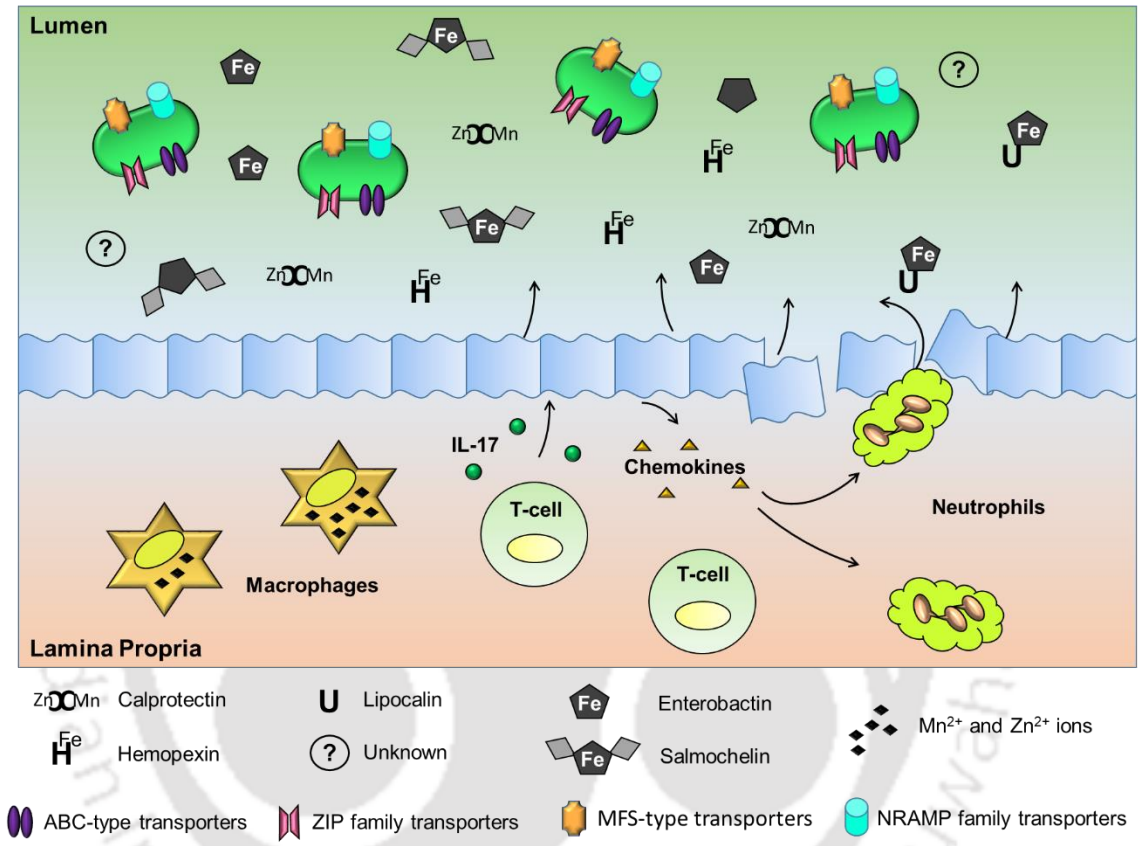


Figure 1.1. Nutritional immunity: A battle for metals in the intestinal mucosa between the host and pathogens. The figure is redrawn from Diaz-Ochoa et al., 2014.

1.1.2. Metal ion transporters in prokaryotes

As discussed, metal ions are crucial components of important biomolecules and their homeostasis is central to various cellular processes. The key elements controlling the distribution of metal ions in several cellular compartments are transmembrane transporters. However, the differences in the chemical properties of metal ions such as transition row,

alkali, alkali earth and other main group metal ions, require transporters having different structural-functional characteristics for their translocation. The plethora of structural information and functional studies have exhibited the distinctive features of these metal transporters (Ferguson and Deisenhofer, 2004; Ma et al., 2009; Wilkens, 2015; Lewinson and Livnat-Levanon, 2017). In addition, different membrane components of Gram-negative and positive bacteria lead to the types of transporters they possess. Both Gram-negative and positive bacteria consist of an outer membrane (OM) and an inner cytoplasmic membrane (IM) through which the metal ion passes before entering the cytosolic region. Additionally, Gram-negative bacteria also possess a periplasmic space in between the OM and IM.

The membrane transport system is broadly divided into two classes: 1) Passive transporter and 2) Active transporters. Passive transporters follow a concentration gradient-based transport of ions and do not require energy as input (Figure 1.2). Passive transporters are further subdivided as diffusion, where spontaneous translocation of a membrane-permeable substance occurs across the membrane and facilitates transport where the movement of a membrane-impermeable substance occurs across the membrane via transporters such as gated ion channels. The trimeric β-barrel proteins referred to as 'porins' embedded in the OM is an example of passive diffusion, which allows non-specific diffusion of metal ions across OM. On the other hand, active transporters require energy to transport substances across the cell membrane against the concentration gradient, which is further sub-divided as primary and secondary active transporters based on their energy requirement. Primary active transporters contain an ATPase responsible for ATP hydrolysis to generate the energy required for transport (Figure 1.2). In contrast, in secondary active transporter, the potential difference created by pumping ions out of the cell is employed rather than exploiting direct coupling of ATP hydrolysis (Figure 1.2).

However, to meet the threshold metal ion concentration, the cytosol is concentrated with various metal ions effectively (Ma et al., 2009). Apart from maintaining the threshold concentration of these essential metal ions, microorganisms possess several types of these transporters to regulate the concentration of these metal ions, referred to as metal ion

homeostasis (Porcheron et al., 2013). These transporters are very eminent and efficient in their functions. Some of the examples of the active transport systems involved in metal ion uptake and its regulation are provided in Table 1.1. One of the eminent mechanisms developed by microorganisms to sequester metal ions effectively is a primary active transporter called ATP binding cassette (ABC) transporters.

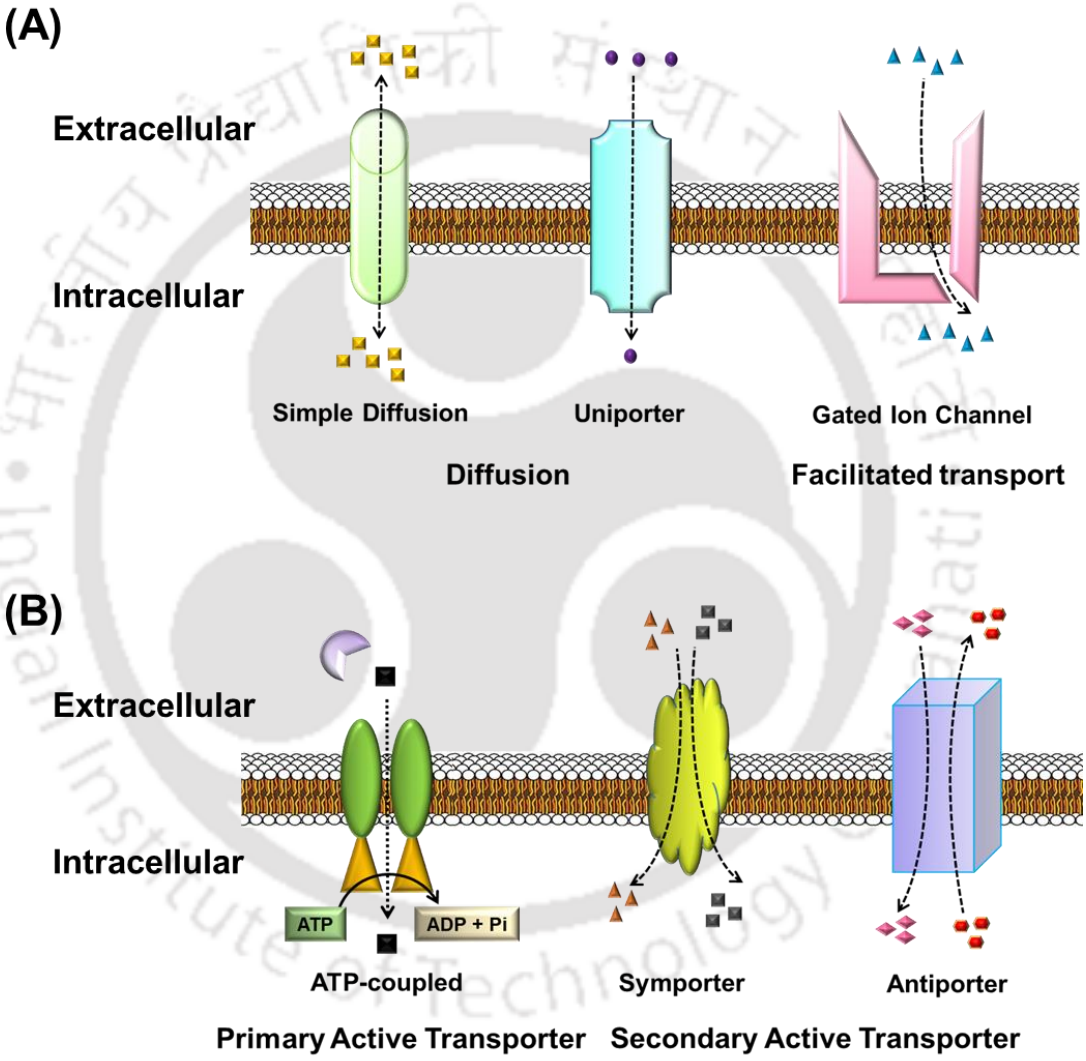


Figure 1.2. Metal ion transporters in microorganisms. (A) Passive transporters and (B) Active transporters.

Table 1.1. Classification of active transporters citing some major examples.

Transporter Class	Transporter Family	Gene Name	Substrate	Energetics
Secondary transporter	Proton dependent Mn^{2+} transporter	*mntH*	Mn^{2+}, Fe^{2+}, Zn^{2+}/H^{+} symporter	PMF*
	Metal ion transporters	*corA*	?[#], Mg^{2+}, Co^{2+}, Fe^{2+}	?
	Cation diffusion facilitator	-	?, Cd^{2+}, Zn^{2+}, Co^{2+}	
	Ni^{2+}, Co^{2+} transporter	*nicT*	Ni^{2+}, Co^{2+}	PMF
	Arsenical resistance-3	*arsC*	AsO^{3-}	?, PMF,
	Ca^{2+}:cation antiporter	*chaA*	Ca^{2+}	
	Monovalent cation:proton antiporter-1	*yjcE*	?, Na^{+}/H^{+} antiporter	
	Monovalent cation:proton antiporter-2	*kefB*	$?/K^{+}$, Na^{+}/H^{+} antiporter	
Primary active transporter	P-type ATPase	*ctpA*	?, heavy metal, $Cu^{+/2+}$	ATP hydrolysis
		ctpB	?, heavy metal, $Cu^{+/2+}$	
		ctpC	?, heavy metal, Cd^{2+}/Fe^{2+}	
		ctpD	?, heavy metal, Cd^{2+}	
		ctpE	?	
		ctpF	?, Ca^{2+}, Mg^{2+}	
		ctpG	?, heavy metal, Cd^{2+}	
		ctpH	?	
		ctpI	?	
		ctpJ	?, heavy metal, Cd^{2+}	
		ctpV	?, heavy metal, Cu^{2+}	
		kdpB	K^{+} uptake	
	ATP-binding cassette	*fecB*	$Fe^{2+/3+}$	

		znuA	Zn^{2+}	
		sitA	Mn^{2+}/Fe^{2+}	
		psaA	Mn^{2+}	
		mntC	Mn^{2+}	
		troA	Zn^{2+}	
		btuF	Cobalamin	
		bhuT, *hmuT*, *rhuT*	Heme	
Unclassified transporter	Mg^{2+} transporter	*mgtE*	?, Mg^{2+}, Co^{2+} uptake	?
	MerTP mercuric ion permease	*merT*	?, Hg^{2+} uptake	

?, unknown; * PMF, Proton motive force.

1.2. ATP-binding cassette (ABC) transporters

The expeditiously growing family of **A**TP-**b**inding **c**assette (ABC) transporters consists of an extremely diversified group of membrane transport proteins that translocate solutes (substrates) across cellular membranes. This translocation of substrates is fueled by ATP hydrolysis. Members of this superfamily are not only involved in the uptake of nutrients but are also actively involved in several other cellular processes such as protein secretion, cell-wall peptide recycling, quorum sensing, signal transduction, antigen presentation, drug or antibiotic resistance, bacterial pathogenesis and sporulation (Higgins, 1992; Lewis et al., 2012; Paul et al., 2013). The involvement of ABC transporters is reported to be central in many physiological processes and in several areas of biomedical research which includes, multidrug resistance (MDR), bacterial pathogenesis, a certain type of cancers and genetic disorders. ABC transporters involved in various functions have been identified and found to be evolutionarily conserved in all domains of life, including humans and thus, can be considered as an early proteinaceous machinery for nutrients to cross membranes against a concentration gradient (Schneider and Hunke, 1997; Davidson et al., 2008). Based on the directionality of substrate transport, ABC transporters are broadly classified as exporters

and importers. ABC exporters are the efflux systems that confer translocation of substrates like toxins, drug and peptide antibiotic molecules from the cytoplasmic to the exoplasmic leaflet of the cell membrane (Wilkens, 2015). In contrast, ABC importers, as the name suggests, import nutrients essential for the survival of the microorganisms from the extracellular milieu to the cytosolic region. ABC exporters are ubiquitously present in all domains of life (archaea, bacteria and eukarya), whereas ABC importers are exclusively present in prokaryotes and, more recently, have also been reported in some plants (Shitan et al., 2003; Terasaka et al., 2005). A canonical architecture of an ABC transporter is organized of four functional units or domains viz., two nucleotide-binding domains (NBD; NBD1 and NBD2) and two transmembrane domains (TMD; TMD1 and TMD2) (Figure 1.3A and B). Eukaryotes have a majority of ABC transport system constituted by a single polypeptide chain having all the four functional domains with few members assembled from 'half' transporters having either identical (homodimers) or different (heterodimers) halves (Procko et al., 2009; Wilkens, 2015). In contrast, bacteria possess these four domains either in a combination of individual or pairwise identical subunits or a combination of fused TMDs and/or NBDs. Besides the four functional units, bacterial importers require an additional subunit responsible for substrate (solute) capturing and delivering it to the TMDs known as substrate-binding proteins or domains (SBPs) (Figure 1.3B).

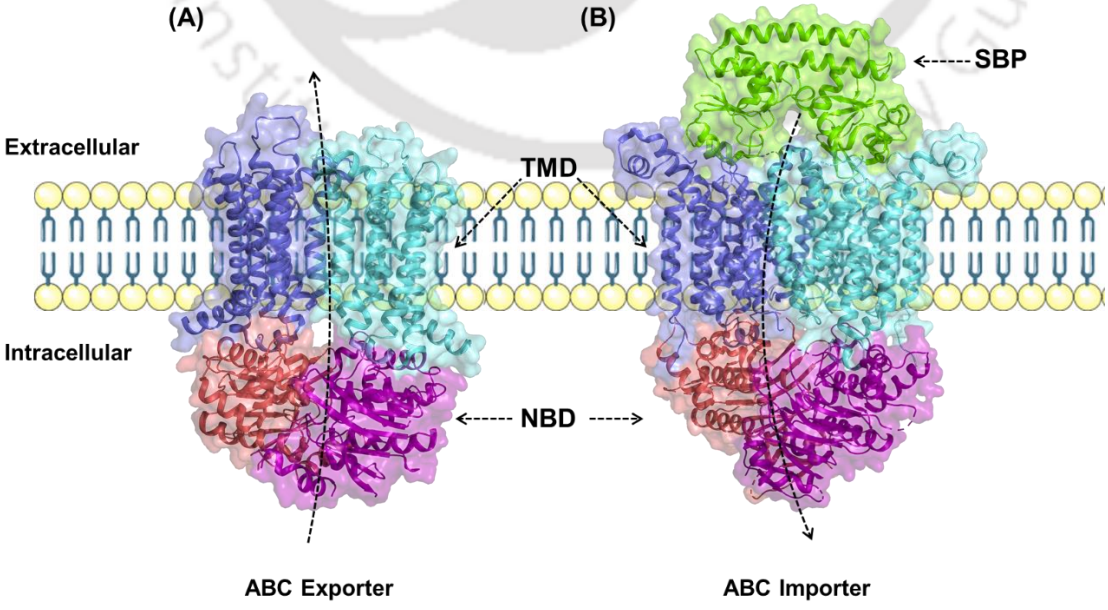


Figure 1.3. The overall architecture of ABC transporters. Three-dimensional structure showing the domain organization in (A) ABC exporter (PDB id: 6XJH) and (B) ABC importer (PDB id: 5B58). The transmembrane and nucleotide-binding domains (TMD in blue and cyan and NBD in salmon and purple) are shown for both the transporters along with the substrate-binding protein SBP (green) of ABC importer.

1.2.1. ABC importers

ABC importers mediate the uptake of various nutrients from the extracellular milieu to the inside of the cell in prokaryotes. These nutrients or solute molecules include a wide range of substrates such as metal ions, siderophores, amino acid precursors, di- and oligo-peptides, carbohydrates, vitamins. For the ease of substrate uptake, ABC importers exploit a substrate-binding protein (SBP), which are specific to their cognate substrate or solute molecule and subsequently deliver it to their cognate TMD for further translocation. The available structures exhibit divergence in the overall architecture of ABC importers, which is not quite understood. ABC importers associated with metal ion uptake system, such as iron, manganese and zinc, have been shown to play crucial roles in bacterial virulence in disease models (Table 1.2). These metal ions are essential for bacterial growth, and knockout studies of the ABC transporters responsible for metal import have revealed deleterious effects on the virulence of pathogenic bacteria (Boyer et al., 2002; Garmory and Titball, 2004).

Table 1.2. Metal ion uptake ABC importers involved in bacterial virulence.

ABC importer or protein	**Organism**	**Transported metal ion (s)**	**Details of virulence**	**Reference(s)**
ZnuA	*Haemophilus ducreyi*	Zinc	*znuA* mutant exhibits decrease in virulence in rabbit model	Lewis et al., 1999

EfaCBA	*Enterococcus faecalis*	Manganese	*efaCBA* mutant exhibits attenuated virulence in mouse models	Singh et al., 1998
PsaABC	*Streptococcus pneumoniae*	Manganese	*psaA* mutant exhibits attenuated virulence in mice model	Berry and Paton, 1996
SitABCD	*Salmonella enterica*	Iron, manganese	*sitABCD* is necessary for full virulence in mice model	Boyer et al., 2002
MtsABC	*Streptococcus pyogenes*	Iron, manganese	*mtsABC* mutant exhibits attenuated virulence in mice	Janulczyk et al., 2003
SloABC	*Streptococcus mutans*	Iron, manganese	*sloC* and *sloA* mutants exhibits attenuated virulence in rat models	Kitten et al., 2002; Paik et al., 2003
YbtPQ	*Yersinia pestis*	Iron	*ybtP* mutants were avirulent in mice model	Fetherston et al., 1999
YfeABCD	*Yersinia pestis*	Iron	*yfeAB* mutant exhibits reduced virulence in mice model	Bearden and Perry, 1999
FeoABC	*Salmonella enterica*	Iron	*feoB* mutant exhibits attenuated virulence in mice	Boyer et al., 2002
SufABCD	*Erwinia chrysanthemi*	Iron	*sufA* and *sufC* mutants exhibits reduced ability to cause maceration of chicory leaves whereas *sufC* is required for systemic invasion in *Saintpaulia ionantha*	Nachin et al., 2001, 2003

Pit1 and Pit2	*Streptococcus pneumoniae*	Iron	*pit1/pit2* mutant attenuated for virulence in mouse models	Brown et al., 2001

The structural and biochemical evidences have permitted the classification of ABC importers into three different group's namely Type I, II and III ABC importers which are exclusively present in prokaryotes (Rice et al., 2014). Both Type I and II importers possess the additional SBP component for substrate uptake; however, there are several other characteristics that distinguishes Type I importers from type II (Figure 1.4A and B). Type I importers possess fewer transmembrane helices and show negligible levels of ATPase activity in the absence of SBPs. Whereas Type II importers possess a greater number of transmembrane helices and the core helices undergo distinct conformational changes to translocate substrates (Korkhov et al., 2012; Rice et al., 2014). The presence of a substrate molecule weakens the association of SBP with Type II importers. In contrast to Type I and II, Type III importers are structurally much divergent, which includes the energy coupling factor (ECF) ABC transporters (Figure 1.4C). Type III importers lack the SBP component but possess an 'S'-component responsible for substrate binding and is a transmembrane protein (Figure 1.4C and D). Interestingly, metal ion importers are present in all three types of importers. Several studies have reported the structure of metal ion importers from Type I, II and III ABC importers. For example, the ModABC importer from *Archaeoglobus fulgidus* involved in molybdate transport is a Type I importer, whereas the HmuTUV and BtuFCD from *Yersinia pestis* and *Escherichia coli*, involved in heme and cobalamin transport, respectively, are Type II ABC importers (Hollenstein et al., 2007; Mattle et al., 2010; Woo et al., 2012). In addition, the structure of a nickel/cobalt ECF-type ABC transporter from *Thermoanaerobacter tengcogensis*, NikM2, has also been reported (Yu et al., 2014).

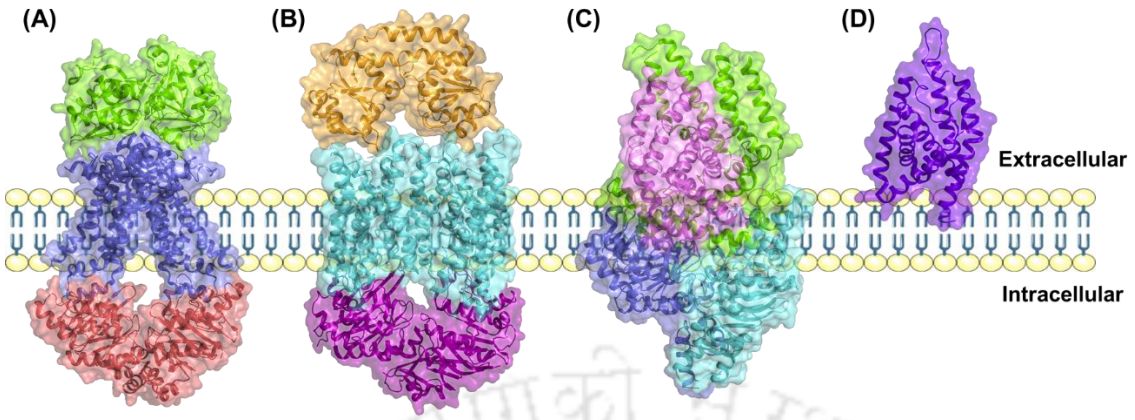


Figure 1.4. Different types of ABC importers. A complete (A) Type I ABC importer (PDB id: 2ONK), (B) Type II ABC importer (PDB id: 3NU1 and 4G1U) and (C) Type III ABC importer (ECF-type) (PDB id: 4HUQ). (D) S-component of an ECF-type ABC transporter (PDB id: 4M5B). The TMDs, NBDs and SBPs of Type I and II as well as different components of Type III importers, are shown in different colors.

1.2.1.1. Transmembrane domain (TMD)

The TMDs are membrane-embedded structures comprising of 10-12 transmembrane helices in their core forming a central translocation pathway. The number of transmembrane helices comprising the TMD varies in ABC exporters and importers. The central channel in ABC exporters possesses ~12 transmembrane helices, while in ABC importers, the number of transmembrane helices varies from 10 to 20 (Hollenstein et al., 2007). The overall fold in TMDs varies in different types of ABC transporters (Type I, II and III) (Figure 1.5A-C). Additionally, the overall TMD fold of Type III ABC importers is related to ABC exporters (Rees et al., 2009). The Type I ABC importers possess six transmembrane helices, whereas Type II ABC importers consist of 10-12 transmembrane helices per TMD subunit (Figure 1.5A and B) (Locher et al., 2002; Parcej et al., 2007). For instance, TMD subunit of Type I importer, i.e. maltose (MalF and MalG, PDB id: 4KHZ) and molybdate (ModB, PDB id: 3D31) ABC transporter possess 6-8 transmembrane helices and organized into a homodimeric and heterodimeric association, respectively (Oldham et al., 2007; Gerber et al., 2008). In contrast, TMD subunit of a Type II importer

viz., cyanocobalamin ABC transporter (PDB id: 1L7V) consists of a higher number (10-12) transmembrane helices (Locher et al., 2002). On the contrary, Type III transporters such as ECF-type, Sav1866 and P-glycoprotein ABC transporters contain six transmembrane helices which transverse the plasma membrane and extend into the cytoplasmic region (Figure 1.5C) (Dawson and Locher, 2006).

In general, TMDs are found to be packed in two different confirmations (1) open towards the periplasmic site where the SBP is located (outward open) and (2) open towards the cytoplasmic site where the NBD is located (inward open) (Figure 1.5D and E) (Wilkens, 2015). Although TMDs are highly conserved in their tertiary structures, they lack conservation at their primary structure level. The lack of conservation in their primary structure is expected owing to their diverse nature of transport substrates. The transport substrate has to interact with the residues of the transmembrane helices that form the transmembrane pore at one point or another for further translocation. In Type I importers (e.g., the maltose ABC transporter from *E. coli*), crystallographic as well as mutagenesis studies have revealed the specific residues interacting with the substrate (Oldham and Chen, 2011; Chen, 2013). However, for Type II importers (e.g., vitamin B12 ABC transporter from *E. coli*), the specific substrate-binding site in the translocation path is unknown till date, but the presence of a mid-membrane hydrophobic cavity is known, which can be blocked from either side of the lipid bilayer (Korkhov et al., 2012). On the contrary, in ABC exporters such as a multidrug transporter P-glycoprotein, multiple overlapping drug-binding sites have been identified, suggesting their poly-specific nature towards their transport substrates (Gutmann et al., 2010; Li et al., 2014).

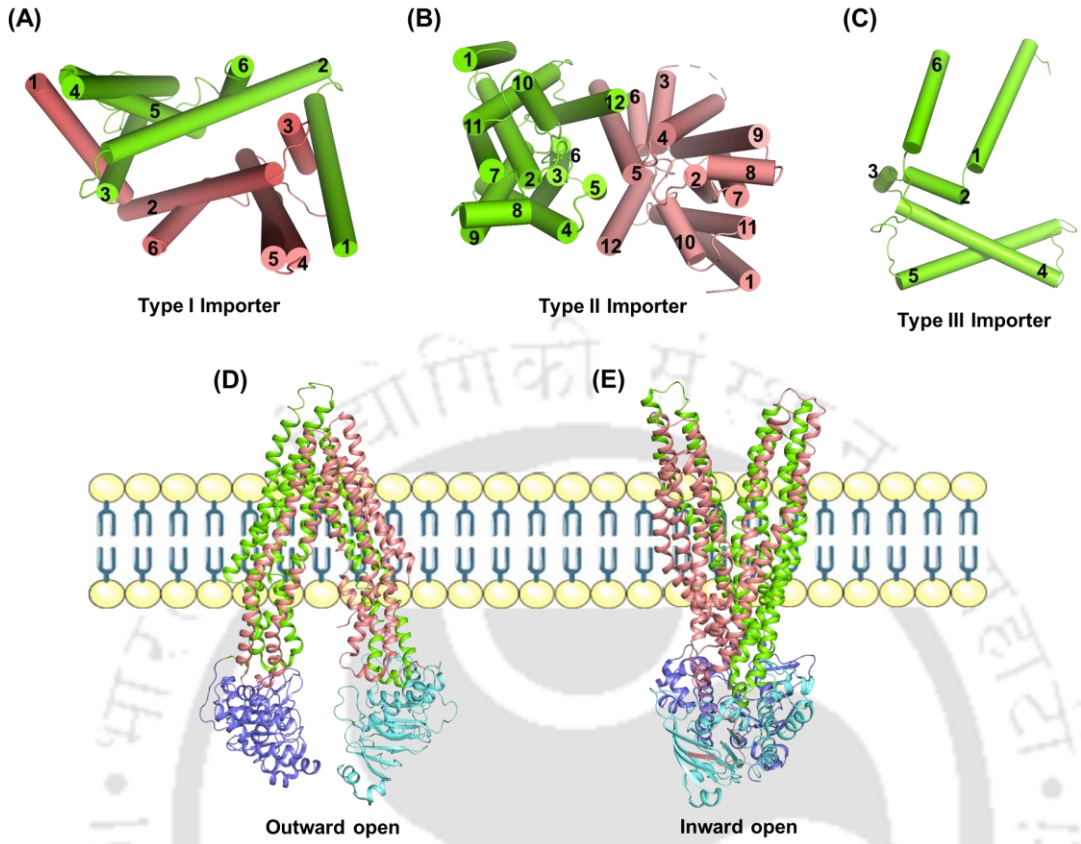


Figure 1.5. Schematic representation of the transmembrane domain (TMD) fold. A canonical transmembrane domain (TMD) fold of (A) Type I, (B) Type II and (C) Type III ABC transporters. The distribution of TM helices per TMD subunit in a different class of ABC transporter is numbered and colored according to the TMD subunit. (D) Outward- and (E) Inward-facing conformation of TMD subunit (PDB id: 2HYD and 4M2T). The TMD1 and TMD2 subunits are colored in blue and cyan with their associated NBDs in green and salmon, respectively.

1.2.1.2. Nucleotide-binding domain (NBD)

NBDs are the catalytic domains of an ABC transport system that reside in the cytoplasmic region of the transporter. Owing to its high level of sequence similarity and presence of various conserved motifs, the NBDs are considered as the hallmark of ABC transporter superfamily. The sequence identity of the NBDs within and between prokaryotic and

eukaryotic ABC transporters ranges from 30-50%, suggesting an overall similar three-dimensional fold and a conserved energy coupling mechanism. Several structures of NBDs are elucidated to date such as, the histidine permease (Hung et al., 1998), which unveils the existence of a RecA-like domain with a Rossman fold and Walker A motif along with a helical domain possessing the 'LSGGQ' signature sequence exclusive to NBDs (Figure 1.6A and B). Studies reporting the crystal structures of ATP-bound NBDs exhibit the engagement of two NBDs as a symmetric dimer possessing two molecules of ATPs sandwiched within the dimeric interface. In this 'sandwich dimer', one ATP molecule is bound to NBD1 via P-loop residues of NBD1 and signature sequence residues of NBD2 and vice versa for the second ATP (Figure 1.6A). Additionally, the 'Walker B' motif is essential for the catalytic activity, i.e. ATP hydrolysis by NBDs (Figure 1.6B). A glutamate residue in the 'Walker B' motif acts as a nucleophile to accomplish the catalysis. Another motif termed as the 'aspartate or D-loop' facilitates the interaction between the TMDs and NBDs (Zaitseva et al., 2005; Ambudkar et al., 2006). A comparison between the nucleotide bound and free NBDs suggested that the dimerization of NBDs is nucleotide-dependent, which drives significant conformational changes to their cognate TMDs (Chen et al., 2003).

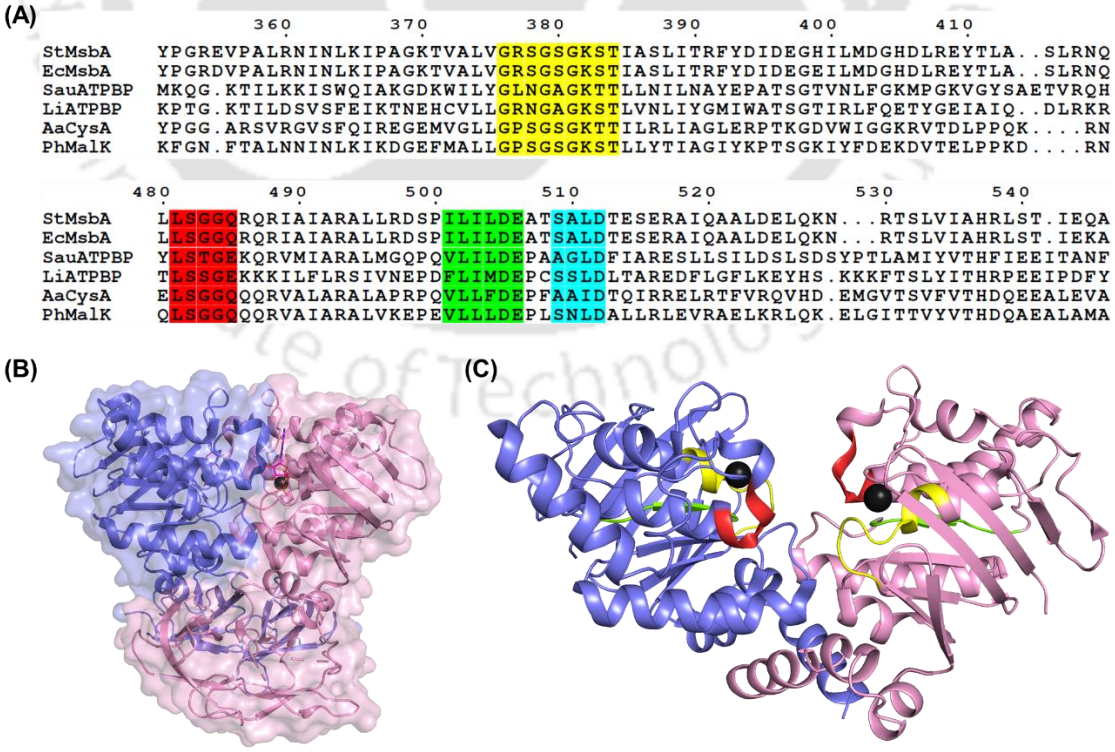


Figure 1.6. Multiple sequence alignment and domain architecture of the nucleotide-binding domain (NBD). (A) Multiple sequence alignment of known NBDs showing the conservation of important motifs and signature sequences. The proteins used to build the MSA are StMsbA (*Salmonella typhimurium*, P63359), EcMsbA (*E. coli*, P60752), SauATPBP (*Staphylococcus aureus*, A0A0H2WZ98), LiATPBP (*Leptospira interrogans*, Q72QN4), AaCysA (*Alicyclobacillus acidocaldarius*, Q9RHZ7) and PhMalK (*Pyrococcus horikoshii*, O57942); the organism name and UniProt id of each protein is provided in the parenthesis. The Walker A, Walker B and LSGGQ motifs are shown in yellow, green and red, respectively, along with the 'D loop' in cyan. (B) Dimeric conformation of an NBD subunit in a closed state (maltose ABC transporter, PDB id: 3PUV). The bound ATP molecule at the interface of the dimeric structure is represented in the magenta stick model. (C) Dimeric conformation of an NBD subunit in an open state (molybdate ABC transporter, PDB id: 2ONK). The two protomers of the protein are shown in blue and pink, respectively, with the bound Mg^{2+} ion in the black sphere. The location of the ABC signature motifs such as the Walker A, Walker B and LSGGQ motifs are shown in yellow, green and red, respectively.

1.2.1.3. Substrate (solute)-binding domain (SBP)

In addition to the TMDs and NBDs, ABC importers possess a domain called substrate-binding protein or domain (SBP). The role of SBP is to recognize and bind the substrate molecule and deliver it to the cognate TMD. SBPs are soluble proteins located in the periplasmic space of the Gram-negative bacteria and, therefore, sometimes referred to as periplasmic-binding proteins (PBPs). However, in Gram-positive bacteria, SBPs are anchored to the cytoplasmic membrane via a lipid moiety located near its cognate TMD (Tanaka et al., 2018). Although SBPs were originally discovered as a component of bacterial ABC transport system (Berger, 1973; Tam and Saier, 1993; Wilkinson, 2002), it has now been discovered to be associated with prokaryotic tripartite ATP-independent periplasmic (TRAP) transporters as well (Gonin et al., 2007; Mulligan et al., 2009). Since ABC importers influx in a wide range of substrates, they utilize different types of SBPs to accomplish the substrate translocation. Generally, SBPs are located within the same operon as the TMD and NBD subunits of the ABC transporter. However, the plethora of microbial

genome sequencing has revealed numerous SBPs which lack an apparent cognate ABC transport system and therefore, these SBPs are referred to as 'orphan SBPs'. Recent studies reported the association of these orphan SBPs with pre-existing ABC transport complexes (Lewis et al., 2012). This suggests that the substrate transport capability of these ABC transporters is substantially larger and is not governed solely by the SBP located within the operon (Chen et al. 2010). SBPs are responsible for the selection and uptake of a broad array of substrates from the extracellular milieu; different SBPs are named accordingly like metal-binding proteins, sugar-binding proteins, peptide-binding proteins, etc. Since SBPs are responsible for maintaining the specificity and selectivity of the substrate(s) as well as directionality of the transport, they have been shown to be a crucial component of prokaryotic ABC transport systems for substrate transport. Several knockout and mutagenesis studies have revealed the crucial role of SBPs, in particular, SBPs involved in metal ion uptake. For example, mutant studies on PsaA, an Mn^{2+}-binding protein of PsaABC transporter from *Streptococcus pneumoniae* exhibited attenuated virulence in mice models (Berry and Paton, 1996). Similarly, mutational studies on ZnuA, a Zn^{2+}-binding protein of ZnuABC transporter from *Haemophilus ducreyi*, showed a decrease in virulence in rabbits (Lewis et al., 1999).

Despite having low sequence similarity, SBPs share a common architecture. SBPs are comprised of two α/β domains where the β-sheets are flanked by α-helices, numbers of which vary for different types of SBPs. These two domains are termed the N- and C-terminal domains (NTD and CTD). The NTD and CTD are linked by a 'hinge region' forming the active-site cavity for ligand or substrate binding (Figure 1.7). In a ligand-free state (i.e., open confirmation), the NTD and CTD show enough flexibility allowing the free rotation of these two domains around the hinge region (Figure 1.7A). However, upon ligand binding, the NTD and CTD come closer stabilized by the hinge region to attain a close confirmation of the SBPs (Figure 1.7B) (Tang et al., 2007). In a transition state from open to closed, both the domains move around the flexible hinge region following a well-known 'Venus Fly-trap' mechanism (Figure 1.7C) (Mao et al., 1982).

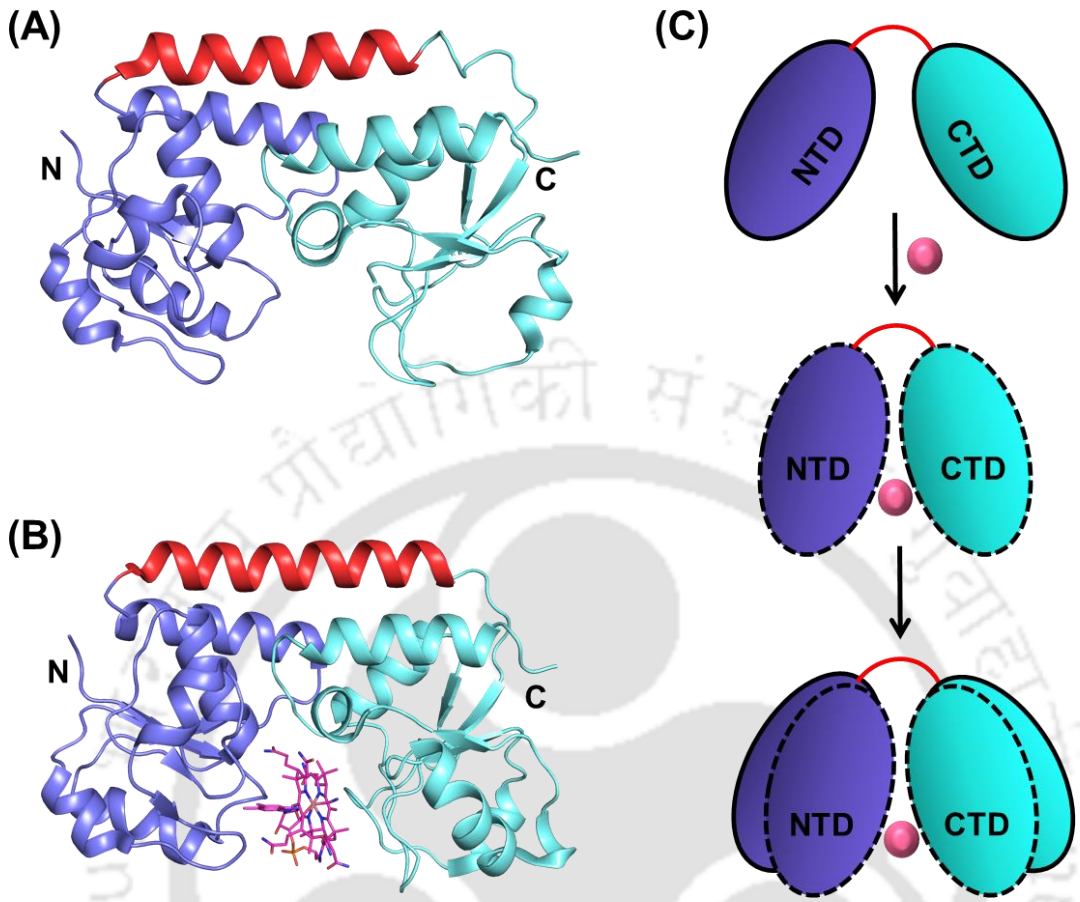


Figure 1.7. Structural topology of a SBP. The tertiary structure of (A) ligand free (vitamin B12-binding protein, PDB id: 1N4D) and (B) ligand-bound (vitamin B12-binding protein, PDB id: 1N4A) SBP. The bound ligand is shown in the magenta stick model. The NTD and CTD are shown in blue and cyan, respectively, with the hinge region connecting them in red. (C) Schematic representation of the classical ‘‘Venus Fly-trap’’ mechanism of ligand binding. The conformational changes of the NTD (blue) and CTD (cyan) upon ligand (pink sphere) binding are depicted with dotted lines.

1.2.1.4. Classification of substrate (solute)-binding proteins (SBPs)

Initially, on the basis of the topology of the SBPs, they were classified into three different classes, viz I, II and III (Fukami-Kobayashi et al., 1999; Lee et al., 1999). Out of the three classes, two classes (I and II) differ by their β-sheet arrangements and linker region. In

class I and II, the β-sheet organization follows β_2-β_1-β_3-β_4-β_5 and β_2-β_1-β_3-β_n-β_4 topology (n is crossover β-strand between NTD and CTD), respectively (Fukami-Kobayashi et al., 1999). In general, the hinge region of class I SBPs is made up of three connecting strands, whereas class II SBPs require only two connecting strands as the hinge region. On the contrary, class III SBPs possess a single α-helix function as a hinge region that links the two domains of the protein (Karpowich et al., 2003).

Later on, many SBPs have been biochemically and biophysically characterized. Even though the amino acid sequence identity between these SBPs is low, their overall three-dimensional structural fold is highly conserved (Berntsson et al., 2010). With the technological advancements in the field of structural biology, a plethora of structures is available from different classes and types of SBPs. This ultimately led to a new classification of SBPs based on their overall architecture and the type of ligands they bind proposed by Berntsson and colleagues in 2010 followed by an updated classification of SBPs (Berntsson et al., 2010; Scheepers et al., 2016). In total, 504 unique entries for SBP structures that bind different types of ligands were collected from the Protein Data Bank (PDB) for their classification. Using these structures, all the SBPs have been categorized into seven different clusters, viz cluster A-G (Table 1.3 and Figure 1.8) (Scheepers et al., 2016). SBPs involved in the uptake of metal ions are clustered in clusters A and D whereas; those involved in sugar uptake are clustered in clusters B, C and D (Table 1.3). Interestingly, amino acid binding SBPs are spread throughout the SBP clusters (Table 1.3). Depending on the subtle variations in the structural features and ligand specificity, clusters A, B, D, E and F were further sub-categorized into subclusters. The different clusters of SBPs are topologically unique, where the hinge region connecting the NTD and CTD is identified as the defining feature for each cluster (Figure 1.8).

Table 1.3. An overview of the SBP classification based on their ligand specificity.

SBP classification by Fukami-Kobayashi et al., 1999	SBP classification by Scheepers et al., 2016		Types of ligands
Class III	Cluster A	Cluster A-I	Metal ions
		Cluster A-II	Siderophores, hemes, vitamin B12
Class I	Cluster B	Cluster B-I	Sugar, sugar alcohols, autoinducer 2
		Cluster B-II	Amino acids
		Cluster B-III	Aromatic acids
		Cluster B-IV	Natriuretic compounds
		Cluster B-V	Glutamate
Class II	Cluster C		Di- and oligo-peptides, cellobiose, opine, nickel
Class II	Cluster D	Cluster D-I	Sugar, sugar alcohols, spermidine/putrescine
		Cluster D-II	Thiamine, spermidine/putrescine
		Cluster D-III-a	Phosphate, molybdate, tungstate
		Cluster D-III-b	molybdate, tungstate
		Cluster D-IV	Iron ions
Class II	Cluster E	Cluster E-I	Organic acids, amino alcohols, dipeptides, glycerol-3-phosphate
		Cluster E-II	Amino acids
Class II	Cluster F	Cluster F-I	Thiamine, pyrimidines, sulfonates, bicarbonate
		Cluster F-II	Methionine

		Cluster F-III	Compatible solutes
		Cluster F-IV	Amino acids
N.D*	Cluster G		Alginate, unknown

*N.D: Not determined

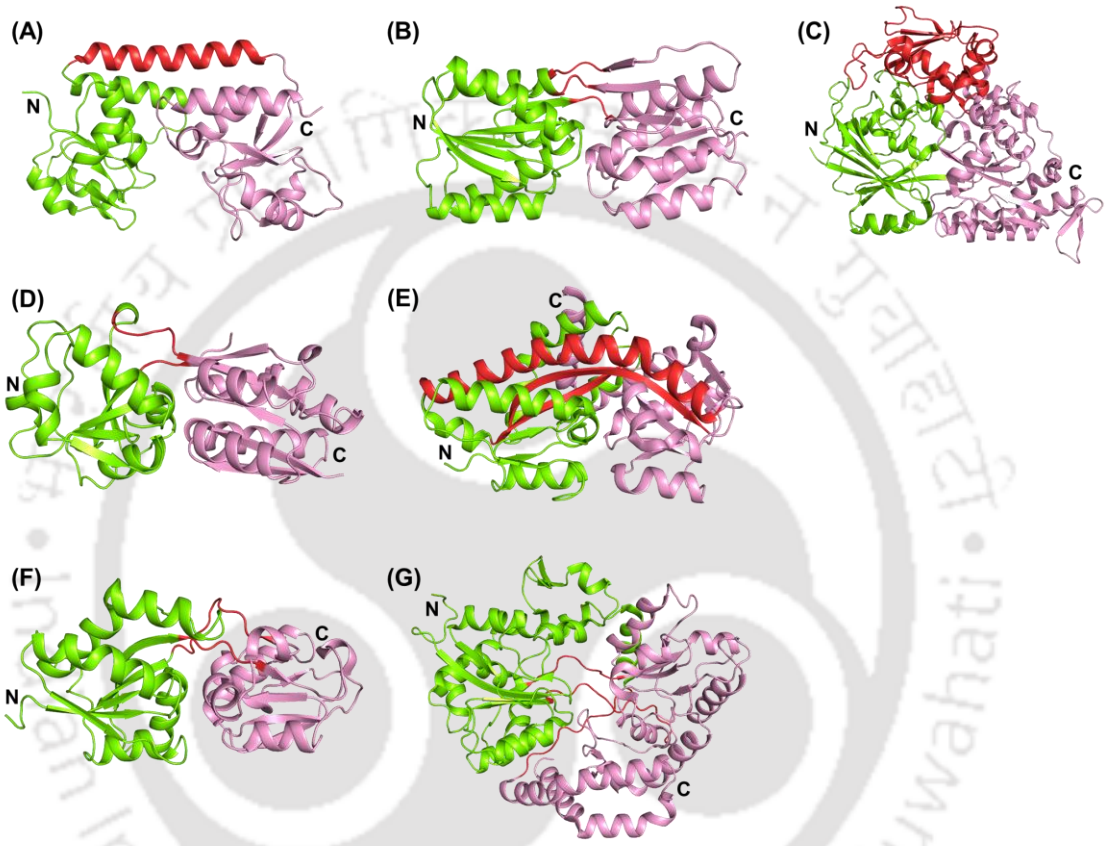


Figure 1.8. Structure-based classification of the SBPs. Based on the overall topology and differences in the structural features of hinge regions of the SBPs, they are classified as (A) cluster A (PDB ID: 1N2Z), (B) cluster B (PDB id: 1DRJ), (C) cluster C (PDB id: 3DRF), (D) cluster D (PDB id: 1AMF), (E) cluster E (PDB id: 3FXB), (F) cluster F (PDB id: 4GOT) and (F) cluster G (PDB id: 1Y3P). For each SBP, the NTD and CTD are represented in green and pink, respectively, with the hinge region connecting them in red.

1.2.2. Mechanism of substrate translocation by ABC transporter

Despite having topological differences, Type I and Type II ABC importers follow an identical mechanism of substrate transport, known as the "alternate access mechanism" (Figure 1.9) (Rice et al., 2014). In this mechanism, the ABC importers undergo significant conformational transitions from inward to outward directions to facilitate the influx of molecules or substrates. At an initial stage, the SBP interacts with the substrate molecule and changes its conformation from an open to a closed state where the substrate binding occurs (Figure 1.9A). Subsequently, the substrate-bound closed SBP interacts with its cognate TMDs, which ultimately leads to conformation changes and movement in TMDs (Figure 1.9B). This movement results in an inward conformation of the transporter, where the NBD subunits remain in a semi-open state devoid of any nucleotide, here ATP. At this state, two ATP molecules bind to the NBD subunits at their interface, which induces the transition of the transporter from an inward conformation to an outward-facing conformation (Figure 1.9B). Subsequently, it opens the SBP which leads to the release of the substrate molecule following a translocation passage where the SBP interacts with the TMD subunits (Figure 1.9C). Upon ATP hydrolysis, the ABC transporter releases the substrate molecule into the cytoplasmic region and reorients in an inward conformation that is the initial resting state of the transporter (Figure 1.9C and D).

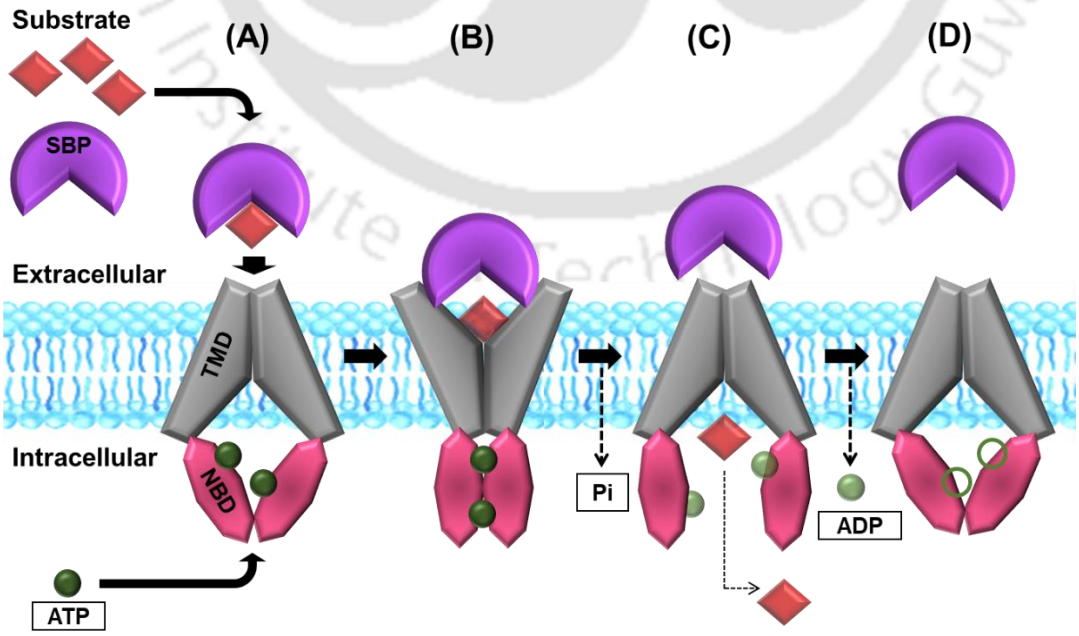


Figure 1.9. Schematic representation of substrate translocation via ABC importer. (A) The binding of substrate-loaded SBPs to the open inward TMD induces the conformational changes in the TMDs. (B) Changes in the TMDs conformation stimulate the ATPase activity of NBD by stabilizing the catalytic transition state for ATP hydrolysis. (C) The TMD opens inwards, facilitating the translocation of substrate into the cytoplasm, ATP is converted to ADP (D) The transporter returns to resting state following ATP hydrolysis, substrate is transported, and SBP is released. SBP: magenta, substrate: red, TMD: green, NBD: dark blue, ATP: brown and ADP: orange.

1.3. Significance of the study

Uptake of metal ions either in their free or complex forms by ABC transporters is major feed lines in pathogenic and non-pathogenic bacteria to accomplish various physiological processes. Several studies have reported the contribution of the SBPs of ABC transport systems specific to metal ions in virulence. In pathogenic bacteria, these roles are associated with colonization in the host environment and escape from the host defense mechanism. As eukaryotes, including human beings, lack SBPs and an exclusive presence of SBPs in prokaryotes proves them as useful targets for antimicrobial therapeutics. However, designing therapeutic drug molecules or inhibitors against SBPs requires a clear understanding of their substrate binding and transport mechanisms. Although, SBPs involved in the metal ion/complex uptake are extensively studied, due to the vast chemical properties of metal ions such as transition row and alkali earth metal ions, understating their transport mechanism via ABC transporters remains challenging. Knowledge of the three-dimensional structures of SBPs in their ligand-free (apo) and bound (holo) states is one way to understand their molecular mechanism of action in detail. Despite having a low level of sequence identity, SBPs are reported to share a conserve topology. The cluster A and D SBPs are specified for metal uptake where clusters A and D are sub-clustered as A-I and A-II and D-I to D-IV. Subclusters A-I and A-II are involved in the uptake of divalent metal ions and metal ion complexes, respectively, whereas members of cluster D-III and D-IV are reported to uptake metal oxyanions and iron ($Fe^{2+/3+}$) ions, respectively. Owing to the distribution of metal ion uptake SBPs in different clusters and varying properties,

including the coordination chemistry of different metal ions, the selection and binding mode of several metal ions with SBPs is not clearly understood. Hence, it is important to understand how topologically conserved SBPs recognize and select the specific metal ion and mediates their uptake in free or complex forms. In this study, we have investigated the metal-specific SBPs from a Gram-negative thermophilic bacterium *Thermus thermophilus* HB8 and a pathogenic bacterium *Mycobacterium tuberculosis* HB8. Detailed information of the structural, molecular and transport kinetics will confer an easy understanding of substrate binding mechanism of SBPs in ABC importers. Also, the findings from the proposed study will furnish the base for designing potent inhibitor molecules against the selected SBPs from the target pathogens.

1.4. OBJECTIVES

To understand the selective transport mechanism of ABC transporter for diverse metal ions, the following objectives were phrased:

1. *In silico* characterization of the repertoire of probable metal uptake ABC transporters from *Thermus thermophilus* HB8 and *Mycobacterium tuberculosis* H37Rv.

2. Molecular cloning, over expression and protein purification to its homogeneity of two selected proteins, MctA (ORF id: TTHB177) from *T. thermophilus* HB8 and MhuP (ORF id: Rv0265c) from *M. tuberculosis* H37Rv.

3. Crystallization of the selected two proteins to elucidate their three-dimensional crystal structures.

4. Co-crystallization of the purified proteins with their cognate substrate(s) (metal ions/complexes) to study their binding and kinetics.

5. Biochemical and biophysical studies to understand the metal ion/complex-binding mechanism of SBPs.

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

CHAPTER 2

Materials and methods

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

General outline of chapter 2

This chapter explains the bioinformatic approach used to study the repertoire of metal uptake ABC transporters in *Thermus thermophilus* HB8 and *Mycobacterium tuberculosis* H37Rv. This chapter also describes the designing and methodology of different experiments for the structural and functional characterization of the selected SBPs of ABC transport systems from both the microorganisms. A general work plan representing the major experimental outline is demonstrated in Figure 2.1.

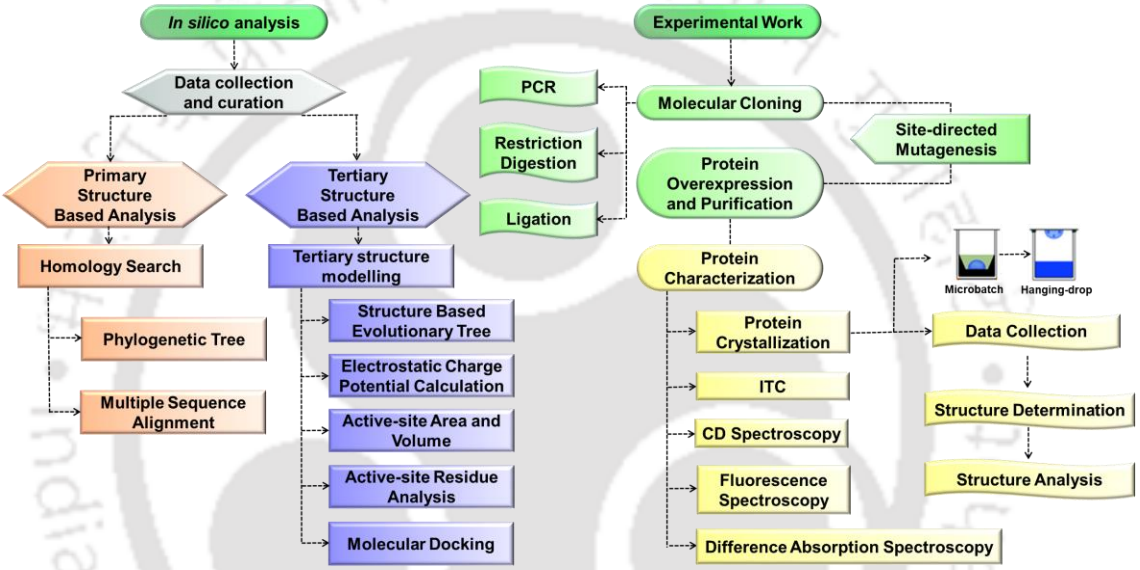


Figure 2.1. A general work plan of the methodology used to accomplish the objectives.

2.1. MATERIALS

2.1.1. Reagents

The enzymes used for molecular cloning, such as restriction enzymes, shrimp alkaline phosphatase, T4 DNA ligase, were purchased from New England Biolabs (NEB) and/or Thermo Fisher Scientific. Additionally, the plasmid isolation, PCR clean-up and gel extraction kits used for recombinant cloning were procured from QIAGEN (Germany), HiMedia (India) and GCC Biotech (India). For site-directed mutagenesis, the Q5 site-specific mutagenesis kit used was purchased from NEB. The Ni-NTA affinity resin and pierce centrifuge columns used for purifying wild type and mutant proteins were procured

from QIAGEN (Germany) and Thermo Fisher Scientific (USA), respectively. The Amicon Ultra-15 and Vivaspin Turbo 15 protein concentrators were purchased from Merck (USA) and Sartorius (Germany), respectively. The various crystallization buffers, microbatch and hanging-drop plates and other protein crystallization-related chemicals and tools were procured from Hampton Research (USA) and Molecular Dimensions (UK).

2.1.2. Metal ions and their complexes

Various divalent and trivalent metal ions used in this study, viz. $MgCl_2$, $CaCl_2$, $MnSO_4$, $MgCl_2$, $CoCl_2$, $CdCl_2$, $NiCl_2$, $ZnCl_2$, $Fe(II)SO_4$ and $Fe(III)Cl_2$ along with Hemin, ferrichrome and cyanocobalamin (vitamin B12) were purchased from the Sigma Aldrich (USA).

2.2. METHODS

2.2.1. Sequence- and structure-based analyses

Once the structure of selected SBPs was solved, further functional characterization was performed by sequence- and structure-based analysis methods as described below.

2.2.1.1. Identification of ORFs and retrieval of sequences

The nucleotide sequences of each ORFs encoding for the selected SBPs were retrieved from KEGG (Kanehisa and Goto, 2000) and Gene database of National Centre for Biotechnology (NCBI) (Benson et al., 2000). In addition, the operonic information of the ABC transporters to identify the genetic arrangement of each subunit was also obtained from the Gene database of NCBI. Similarly, the respective proteins and their homologous sequences analyzed in this study were retrieved from the UniProtKB database (UniProt Consortium, 2019). To obtain the information of a total number of ORFs encoding for probable metal uptake ABC transport systems in *T. thermophilus* HB8 and *M. tuberculosis* H37Rv, the following databases has been used for data collection:

- Structural-Biological Whole-Cell Project (http://www.thermus.org)
- TB Database (Reddy et al., 2009)
- MycoBrowser (Kapopoulou et al., 2011)

- TubercuList (Lew et al., 2011)
- TransportDB 2.0 (Elbourne et al., 2017)
- UniProtKB database (The UniProt Consortium, 2019).

2.2.1.2. Sequence-based analysis

2.2.1.2.1. Identification of homologous proteins and sequence alignment

The homologous proteins for each targeted SBPs were identified by performing a search analysis against the non-redundant (nr), UniProt/SwissProt and Protein Data Bank (PDB) databases using the program BLAST (Basic Local Alignment Search Tool, Altschul et al., 1990). For protein BLAST, following parameters were used: max target sequence, 100; expected threshold, 10; word size, 6; matrix, BLOSUM62 and gap costs, Existence: 11 Extension: 1. After BLAST search, depending on the sequence identity and query coverage, homologous sequences were listed out and subjected to multiple sequence alignments (MSA) using the program Clustal Omega (Sievers and Higgins, 2014). Clustal Omega uses a combination of seeded guide trees and hidden Markov model (HMM) iterations to build alignments between the multiple sequences. The MSAs were performed with a default set of parameters. For clarity, the outputs of the MSA results were further rendered using the web tool ESPript 3.0 (Easy Sequencing in PostScript, Gouet et al., 2003).

2.2.1.2.2. Phylogenetic tree

A phylogenetic tree is used to establish an evolutionary relationship among homologous protein sequences and to deduce their common ancestors. In the present work, phylogenetic trees were generated using the protein sequences of the selected SBPs and their homologs. To generate the phylogenetic trees of SBPs used in this study, the software MEGA version 7.0 was used (Molecular Evolution and Genetic Analysis, Kumar et al., 2016). The phylogenetic trees were built in MEGA 7.0 in the following three steps: (1) the retrieved homologous protein sequences were aligned with the default set of parameters (gap opening: 10 and gap extension penalties: 0.2) using the program ClustalW embedded in the software MEGA 7.0. (2) A tree was generated from the aligned sequences using the neighbor-joining (NJ) method. The significance of the tree was evaluated using the

bootstrapping method, which defines the confidence level of similarity between the clusters formed. (3) The generated phylogenetic tree was further rendered for clear clustering and to obtain the relevant information about the evolutionary relationship.

2.2.1.3. Structure-based analysis

2.2.1.3.1. Structure prediction

For theoretical three-dimensional structure prediction, homology modeling was used. In homology modeling, the theoretical three-dimensional model of the targeted protein was constructed using its primary structure (amino acid sequence) by using a crystal structure of the homologous protein as a template. To predict the tertiary model structure of the selected SBPs, online three-dimensional structure prediction programs such as Phyre 2.0 (Kelley et al., 2015), SWISS-MODEL (Biasini et al., 2014), RaptorX (Kallberg et al., 2012) and I-TASSER (Yang et al., 2015) was used. All the online programs identified the closest homolog of the selected SBPs as the best template and predicted their model structures. The modeled structures were further refined for energy minimization using the programs UCSF Chimera (Petterson et al., 2004), GalaxyRefine (Ko et al., 2012) and ModRefiner (Xu and Zhang, 2011). The initially predicted and the refined model structures were compared together based on their geometrical and stereochemical parameters. Theoretical models of the SBPs having the best geometrical and stereochemical parameters were finalized after validation for further structural analysis.

2.2.1.3.2. Structural comparison with homologous proteins

To obtain the structural homologs of the selected SBPs from the Protein Data Bank (PDB) (Berman et al., 2000), the web-server DALI (Holm and Rosenström, 2010) was used. The root mean square deviation (RMSD) calculations among the selected and homologous structures were performed using the program 3dSS (3-dimensional Structural Superposition, Sumathi et al., 2006) and PyMOL (The PyMOL Molecular Graphics System, Schrodinger, LLC).

2.2.1.3.3. Structure-based classification of SBPs

In the PDB, a wealth of structural information of SBPs is available. Interestingly, despite low sequence similarity, SBPs share high structural similarity among themselves. Therefore, based on the subtle differences in their architecture and the type of ligands they bind, SBPs are classified into seven different clusters. To determine the cluster in which the selected SBPs belong, a structure-based classification of SBPs was performed. Although the metal uptake SBPs are categorized in clusters A and D, SBPs from other clusters were also selected for the analysis for metal uptake SBPs of *T. thermophilus* HB8. However, in the case of *M. tuberculosis* H37Rv, owing to the structural homology of the metal uptake SBPs with cluster A and D SBPs, a total of 45 SBPs are shortlisted from different subclusters. The root-mean-square deviation (rmsd) among all the selected SBPs from different clusters were computed using a homebuilt python script and the program PyMOL (PyMOL Molecular Graphics System, Schrodinger, LLC). The obtained RMSD distance matrix for the shortlisted as well as the targeted SBPs was used to generate a structure-based distance tree to study the structural relationship among all the SBPs. The RMSD matrix with a suitable input format was submitted in the program dendroUPGMA (GarciaVallvé et al., 1999) to convert the RMSD matrix into a Newick (.nwk) file format. The obtained .nwk file was further subjected to MEGA 7.0 to generate the final structure-based evolutionary tree (Kumar et al., 2016). The resultant evolutionary tree grouped the targeted SBPs into their respective subclusters, which were further cross-verified to distinguish the structural features of SBPs by superposing the structures in PyMOL (The PyMOL Molecular Graphics System, Schrodinger, LLC).

2.2.1.3.4. Surface analysis

To determine the electrostatic nature of the active-site pocket of the selected SBPs and their homologs, electrostatic potential charge distribution was calculated using the program Adaptive Poisson-Boltzmann Solver (APBS) embedded in PyMOL (Baker et al., 2001). For clarity, the calculated electrostatic potential charge distribution of each protein was colored in blue for positively charged and red for the negatively-charged surface on a scale of -10 to +10 kcal mol^{-1}. Additionally, the active site area and volume of the selected

proteins, as well as their homologs, were calculated using the program CASTp with a default probe radius of 1.4 Å (Binkowski et al., 2003).

2.2.2. Molecular docking

The binding affinities of different metal ions and their complexes to SBPs were estimated by molecular docking analysis performed using a freely available program AutoDock v4.2 (Morris et al., 2009). For that, the three-dimensional atomic coordinates of the metal ions/complexes were retrieved from their protein-bound structures available at the RCSB Protein Data Bank (Berman et al., 2000). The rotameric orientations of active-site residues were adjusted using the program Coot (Emsley and Cowtan, 2004; Emsley et al., 2010) in order to achieve the closed-like confirmation of the SBPs to accommodate the ligands. Before each docking experiment, wherever applicable, the hydrogen atoms were added to the protein and ligand, using the module available in the program AutoDock. Subsequently, the partial charges were assigned to the protein atoms using the Gasteiger charge algorithm (Gasteiger and Marsili, 1980). Similarly, default as well as calculated (as applicable) set of partial charges for the metal ions, cobalt (Co), iron (Fe), molybdenum (Mo), sulfur (S) and tungsten (W) available in the AutoDock library were assigned. To perform the blind docking of the ligands with the protein, a grid size of 126x126x126 with a grid spacing of 0.375 Å was generated by taking the center of mass of the targeted protein as the grid center. The rigid molecular docking was performed by keeping the protein atoms as rigid and the ligand atoms as flexible about the rotatable bonds. For each molecular docking experiment, a total of 2000 runs of the Lamarckian genetic algorithm (LGA) was set up. The docked ligand conformations were clustered with an rmsd cut-off of 2.0 Å. The atomic interactions between the protein and the docked ligand(s) were identified using the program Coot. The molecular interactions of docked ligands and other structural figures were prepared using the program PyMOL.

For protein-protein docking, the freely available online docking tool ClusPro was used (Kozakov et al., 2017). In ClusPro, the one protein molecule was submitted as a receptor, whereas the ligand-bound protein was submitted as a ligand molecule, keeping a default set of parameters provided by the program. Multiple protein-protein complexes provided

as results were analyzed to obtain the plausible protein complex. The balanced results, which provides 10 clusters, were finally considered for analysis. The atomic interactions between the protein and the docked ligand(s) were identified using the program Coot (Emsley et al., 2010). The molecular interactions of docked ligands and other structural figures were prepared using the program PyMOL.

2.2.3. Recombinant construction for protein overexpression

Molecular cloning is a molecular biology technique that is used to generate new recombinant DNA molecules, which possess a novel combination of the nucleotide sequence. In this study, the molecular cloning method was used to create the recombinant construct, in which the target gene is inserted into the appropriate vector for propagation. Both vector and gene were mixed together in the presence of ligase enzyme and introduced inside the host organism, which leads to replication of the designed construct and produces multiple copies of recombinant clones (Figure 2.2). Details of each step of molecular cloning performed in this study are described below.

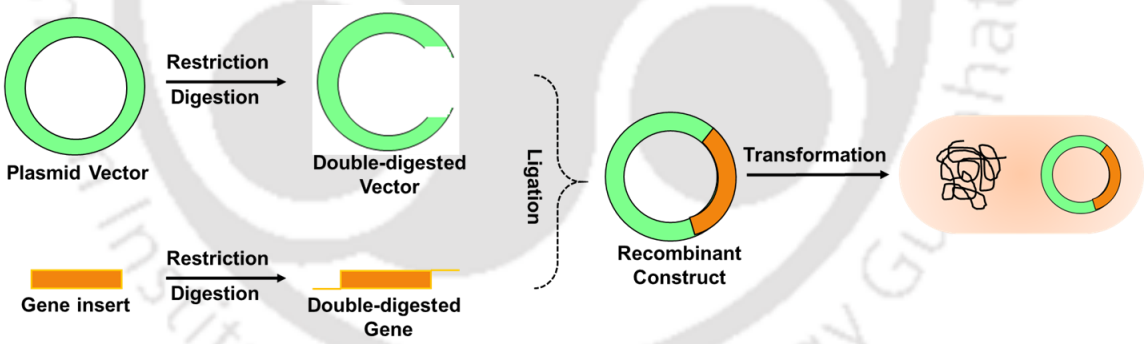


Figure 2.2. Schematic representation for molecular cloning.

2.2.3.1. Competent cell preparation for gene cloning and protein overexpression

The competent cells for different strains of *Escherichia coli* used for molecular cloning and protein overexpression were prepared by using the calcium chloride ($CaCl_2$) based chemical method. The competency of the bacterial cells was achieved by resuspending

them in their log phase in 100 mM $CaCl_2$ solution followed by ice incubation at 4°C for 30 min. Finally, the competent cells were resuspended in an autoclaved mixture of 100 mM $CaCl_2$ and 15% glycerol and stored at -80°C after flash freezing in liquid nitrogen. The expression cells used in this study are derived strains of *E. coli* possessing a chromosomal copy of the T7 RNA polymerase for expression of the protein under the control of T7 promoter. Overexpression of the targeted proteins was achieved by inducing the expression cells with an analog of lactose, i.e. Isopropyl β- d-1-thiogalactopyranoside (IPTG), which binds to the repressor.

For the purpose of molecular cloning and protein overexpression and solubility optimization, following *E. coli* strains were used in this study.

- DH5-α competent cells: have high transformation efficiency responsible for increasing the copy number of the cloned plasmid.
- BL21 (DE3) competent cells: used to overexpress the gene under the control of a T7 promoter using an inducer (IPTG).

2.2.3.2. Gene amplification

For amplifying the gene of interest, gene-specific customized primers were designed considering parameters such as GC content, melting temperature (Tm), restriction digestion sites, etc. The primers were analyzed using the web tool OligoAnalyzer of IDT Technologies to check and verify all the parameters. For the gene *mctA* (from *T. thermophilus* HB8), a cloned plasmid carrying the full length of the target gene was procured from the Biological Research Center, NITE (NBRC), Japan. However, this cloned plasmid lacked the 6x-His-tag necessary for Ni-NTA affinity chromatography-based protein purification. Hence, sub-cloning of the gene *mctA* was performed by amplifying the gene using the cloned plasmid as a template during polymerase chain reaction (PCR). For amplifying the gene *mhuP* (from *M. tuberculosis* H37Rv), the genomic DNA of *M. tuberculosis* H37Rv procured from ATCC was used as a template in PCR. The PCR reaction was performed in the following steps: 1) initial denaturation (95°C, 5-10 min), 2) final denaturation (95°C, 30-40 sec), 3) annealing (Tm of genes, 35-45 sec), 4) initial extension (72°C, 40-50 sec) and 5) final extension (72°C, 10 min). The amplified product

lacks the periplasmic signal sequence and contains a 6x-His-tag at the C-terminus of the protein to facilitate the purification using immobilized Ni-NTA-based affinity chromatography.

2.2.3.3. Double digestion of the gene of interest and vector

For directional cloning, double digestion of the amplified genes was performed using the respective restriction enzymes before inserting them into the pET22b plasmid vector. Restriction enzymes such as NdeI, XhoI and BamHI were chosen, which generates the 5′ and 3′ sticky ends in both the genes and the vector. Prior to restriction digestion, the genes were added with the required restriction enzymes and allowed to incubate at 37°C for 3 hours. Simultaneously, the pET22b plasmid vector was added with the same restriction enzymes and incubated at 37°C for 3 hours. After the third hour, alkaline phosphatase was added to the vector reaction mixture and incubated for an additional 1 hour at 37°C to remove the 5′-phosphate group. After incubation, the double digested genes and vector were purified using the PCR purification kit (QIAGEN) and gel extraction kit (QIAGEN), respectively.

2.2.3.4. Ligation of the gene of interest and clone confirmation

At the final step of molecular cloning, a ligation reaction was performed to insert the gene of interest into a compatibly digested pET22b vector. The ligation reaction was catalyzed by the T4 DNA ligase enzyme, which forms a covalent bond between the two DNA fragments. For ligation, the double digested vector and gene (insert) having sticky ends were mixed together in different ratios (1:1, 1:2, 1:3, 1:4 and 1:5) and allowed to incubate at 22°C for 4 hours in the presence of the T4 DNA ligase. After incubation, each ligation mixture was then transformed into *E. coli* DH5α competent cells to propagate the ligated recombinant clone. The positive colonies were screened by growing the transformed cells in antibiotic supplemented Luria Bertani (LB) agar plates. For further confirmation, the fully-grown colonies were inoculated into Luria Bertani (LB) broth supplemented with selective antibiotics and subjected to plasmid isolation. After plasmid isolation, the expected cloned plasmids were incubated with respective restriction enzymes followed by incubation at 37°C for 3 hours to confirm the presence of the gene of interest. After

confirming the clones via double digestion, the clones were subjected to DNA sequencing for further confirmation.

2.2.3.5. Site-directed mutagenesis

In order to obtain the apo state of the protein MhuP, to find the active-site residues important for ligand binding and to understand the ligand-binding mechanism, site-directed mutagenesis was performed using the Q5 site-specific mutagenesis kit (NEB, USA). The target site for site-directed mutagenesis was selected based on i) the preliminary *in silico* analysis, ii) homology to known structure and iii) the strongest binding of the protein MhuP with an endogenously bound ligand. For *in vitro* site-directed mutagenesis, customized oligonucleotide primers were designed (mentioned in Chapter 5) to introduce the desired mutation in the wild-type (WT) cloned plasmid. Site-directed mutagenesis was performed in the following three steps:

i) The vector was amplified using a High-Fidelity (HF) DNA polymerase from the master mix formulation added to the WT cloned plasmid used as a template.

ii) The amplified products were incubated with unique enzymes viz. kinase, ligase and *Dpn*I to assist the circularization of the amplified products, followed by removal of the template WT DNA.

iii) Transformation of the high-efficient *E. coli* DH5α competent cells with the circularized PCR product.

Finally, the desired point mutations in the target gene were confirmed by plasmid DNA sequencing and structural solution of mutant proteins.

2.2.4. Protein overexpression, solubilization and purification

2.2.4.1. Overexpression of the selected SBPs

The selected SBPs cloned in pET22b vector were overexpressed by transforming the *E. coli* BL21 (DE3) expression host cells with the recombinant constructs. Protein overexpression for MctA and MhuP was optimized at 37°C for different time intervals (1 to 7 hrs). For overexpression, IPTG was used as an inducer. Since a higher concentration of IPTG leads to cell death or unproductive protein formation, the IPTG concentration (0.1, 0.5 and 1.0 mM) was also optimized. Once the expression host, optimum temperature,

IPTG concentration and expression time were optimized, the overexpressed protein was checked for solubility.

2.2.4.2. Solubility of protein

The overexpressed protein of interests was harvested by centrifugation followed by resuspension in lysis buffer for homogenization. Subsequently, the homogenized suspension was sonicated for cell disruption at 26% amplitude for 15 min keeping the sonication cycle to 3 s on and 5 s off. For the thermophilic protein MctA, the obtained cell lysate was heat-treated in a boiling water bath at 70°C for 13 min to remove the thermolabile host proteins. However, no such heat treatment was given to the protein MhuP. The cell lysate was centrifuged at a maximum speed of 18894g for 20 min to segregate the pellet and supernatant fractions. The solubility of the protein was analyzed by running the obtained pellet and supernatant fractions in 10% SDS-PAGE. Once the solubilized protein was identified in the supernatant fraction in SDS-PAGE analysis; it was further subjected to protein purification by using chromatography techniques.

2.2.4.3. Affinity chromatography for protein purification

The immobilized metal affinity chromatography (IMAC) method was employed to purify the protein of interest. Since IMAC follows the precise coordination geometry between a transition metal ion (Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}) and polyhistidine residues, a 6X-His was tagged at the C-terminus of the recombinant proteins, which coordinate with the immobilized Ni^{2+} ions on a matrix. Prior to purification, the pierce centrifuge columns were packed with Ni-NTA resins followed by equilibrating the resins using desired buffers. During purification, the packed and equilibrated Ni-NTA resins were allowed to incubate for ~3 hrs with the supernatant fraction of the protein of interest. After binding, the columns were washed with wash buffers containing varying concentrations of imidazole (10-20 mM) to remove the non-specific proteins weakly bound to the column. In addition to imidazole, other additives such as salt (NaCl), a reducing agent (β-mercaptoethanol) and a protease inhibitor phenylmethanesulphonyl fluoride (PMSF) were also added in the wash buffer to reduce the non-specific binding of host proteins. To elute the protein of interest, a higher concentration of imidazole (250-400 mM) was used in the elution buffer to disrupt

the coordination bond between the Ni^{2+} ions (present in the resin) and histidine residues (tagged in the recombinant protein), which competitively binds to the Ni^{2+} transition metal ion and elutes the tagged proteins. The purity of the eluted fractions was analyzed in 10% SDS-PAGE. The purified elutes were pooled together, followed by gradient dialysis for complete removal of imidazole. To remove the endogenously bound metal ions in the protein MctA, dialysis was performed against 10 mM of EDTA along with the requisite buffer composition. The dialyzed proteins were then concentrated using protein concentrators up to the desired concertation required for biophysical and biochemical experiments.

2.2.5. Protein characterization

2.2.5.1. Mass spectrometric analysis

The exact molecular weight of proteins was determined using matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF). The selected SBPs were homogenously purified and mixed with the matrix in a (protein:matrix) ratio of 1:2, 1:3 and 1:5. Since, the theoretical molecular weight of the selected SBPs is more than 10 kDa, sinapinic acid was used for analyzing the proteins. The matrix was prepared in acetonitrile and 0.1% trifluoroacetic acid in a ratio of 30:70, respectively, to a concentration of 10 mg ml^{-1}. Sinapinic acid was dissolved in the mentioned solvent by sonicating the solution for 30 mins followed by centrifugation at 12000xg for 10 mins to remove the excess matrix crystals. Further, the prepared matrix was mixed with protein samples in varying concentrations of 1, 5 and 10 mg ml^{-1} in the ratios specified above. 2 µL spots of the protein-matrix mixture were made on the ground steel MALDI plate grooves and air-dried. Post drying, the drops were ionized in a Bruker autoflex speed and the data were plotted using an in-built Flex control analysis software.

2.2.5.2. Isothermal titration calorimetry

The thermodynamic parameters of protein-ligand interactions were measured using isothermal titration calorimetry (ITC). ITC is a biophysical method that provides the thermodynamic parameters such as stoichiometry (n), change in enthalpy (ΔH), change in entropy (ΔS) and equilibrium association constant (Ka) for protein-ligand binding. In this

study, an incremental titration using ITC was performed where a fixed volume of ligand was titrated into the protein solution at discrete time intervals. Incremental titration leads to the change in heat due to exothermic (heat released) or endothermic reactions (heat absorbed) taking place between the protein and ligand. The change in heat produces a peak utilized to calculate the area under the peak via the peak integration method. After integration, the data was normalized and fitted to one binding site model to calculate the stoichiometric ratio and change in affinity and enthalpy for the interaction. For each ITC experiment, a control experiment was performed by titrating the ligand into a buffer solution to determine the heat of dilution. The obtained heat of dilution from the control experiment was subtracted from the reaction to get the actual heat of dilution associated with protein-ligand interaction. The subtracted heat change was further used to plot the graph against the protein to ligand molar ratio. The Gibb's free energy for each binding experiment was calculated using the expression:

$$\Delta G = \Delta H - T\Delta S$$

Where ΔH is the change in enthalpy, T is the temperature of the adiabatic system and ΔS is the change in the entropy for the reaction.

2.2.5.3. Fluorescence spectroscopy

As aromatic amino acids play an important role in iron-complexes such as hemin, the binding pocket of the selected SBPs is rich in aromatic amino acids. Hence, to assess the binding of iron complexes with SBPs, a tryptophan-based intrinsic protein fluorescence was performed. The samples were excited at a fixed wavelength at 285 nm, which led to the emission of wavelength signals at 310-450 nm, called fluorescence. The changes in the fluorescence signal highly depend upon the protein conformation and association or dissociation of ligand molecules, which ultimately affects the local tryptophan(s) environment. Thus, in this study, the changes in the tryptophan environment via fluorescence of SBPs were monitored in the presence and absence of several metal ions and their complexes. The addition of metal ions and their complexes to the protein solution decreases the fluorescence signals (quenching), indicating the change in the local environment of tryptophan residues present in the protein.

2.2.5.4. Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy is a biophysical technique widely used for protein characterization and thermal stability analysis. CD provides the secondary structural information of proteins such as α-helix, parallel and anti-parallel β-sheet, turns and random coil in a wavelength range of 260 to 190 nm. In this study, the secondary structural content, as well as the unfolding thermodynamics of the selected proteins, were monitored. The changes in the secondary structural content of the proteins were analyzed for α-helices and β-sheets at the wavelengths ~220 nm and ~195 nm as a function of temperature, respectively. The thermodynamics of protein unfolding as a function of temperature results in melting temperature (Tm) calculation of the targeted SBPs. The thermal stability of the selected SBPs was monitored in a temperature range of 20 to 120ºC.

2.2.5.5. Difference absorption spectroscopy

To check the binding of porphyrin ring containing molecules with the selected protein, Difference absorption spectroscopy (DAS) was performed. For DAS, fresh hemin solutions were prepared first by solubilizing hemin in 100% dimethyl sulfoxide (DMSO) and then making working concentrations of hemin using the dialysate buffer of the respective protein. An equimolar amount of hemin was added to the selected protein at a concentration of 25 µM and incubated at room temperature for 10 min. For DAS, heme-binding was monitored using an infinite M200PRO plate reader (TECAN, Switzerland). An appropriate blank experiment using only hemin solution prepared in the final dialysate buffer was also performed. The actual heme-binding was obtained by subtracting the free hemin spectra from the protein incubated hemin spectra.

2.2.5.6. Energy Dispersive X-ray analysis

The Energy Dispersive X-ray (EDX) microanalysis of the selected protein was performed using 80T EDX detector (Oxford Instruments, Tokyo, Japan) equipped in TEM-Jeol 2100F transmission electron microscopy available at Central Institute Facility (CIF) of Indian Institute of Technology Guwahati, India. The EDX detector is devised with a superatmospheric thin window (SATW) and a 50 mm^2 silicon drift detector (SDD), placed with a takeoff angle of 30°, having a resolution of 129 eV at the Mn K-L3 line (in IUPAC

notation; Siegbahn notation Mn $K_{\alpha 1}$) under the conditions used in this study. The EDX detector is monitored by INCA (version 5.05, Oxford Instruments). For EDX spectral processing, INCA uses a filtered least squares (FLS) approach for background removal, fitting, and deconvolution of the peaks, which also includes a pulse pile-up correction. Quantitative data are obtained from the spectra after matrix correction, which follows the XPP exponential model procedure. All of the details of EDX spectral processing software can be found in the INCA Energy Operator Manual. Because of the large amounts of light elements in the samples, we did not seek to obtain full quantitative data but instead focused on the S/In ratio. Using INCA, this was made by indicating that the sample was coated with a 50 nm carbon layer, selecting all the elements other than sulfur and indium as "deconvolution elements" (which means that their presence is taken into account for the calculation but that they are not displayed in the final result), and using the "Quant Standardization (Extended Set)" provided with the software. Ratios were calculated with the sulfur K-series peaks and the indium L-series peaks. S/Na ratios were obtained using the same procedure.

2.2.6. Crystallization of the selected SBPs

In this study, the method of X-ray crystallography was used to determine the three-dimensional structures of the selected SBPs. X-ray crystallography requires diffractable quality protein crystals. In order to obtain diffractable protein crystals, the target protein solution was screened with a variety of crystallization screens (buffers) to get an initial crystal hit, including various optimizations. Thermodynamically, the protein crystallization progression has four zones viz., undersaturated, metastable, labile and precipitation, where crystal formation occurs at a supersaturated stage in a metastable zone (Figure 2.3). Supersaturated state of protein crystallization can be reached using two major approaches (1) vapor diffusion (2) microbatch-under-oil. For the vapor diffusion technique, hanging-drop method was utilized. In the hanging-drop vapor-diffusion method, a drop containing a mixture of the protein solution and precipitant was placed on a coverslip and subjected to equilibration against the reservoir solution in an inverted position. In contrast, microbatch under oil is a method where a mixture of the protein and precipitant is submerged under the oil to slow down the evaporation. In the present study, paraffin and

silicon oils having different densities were used in a combination to control the diffusion rate. In order to obtain a protein crystal, the solubility of the protein was screened at two different temperatures viz., 4 and 20ºC in both hanging-drop and microbatch-under-oil methods. Once the initial crystal hits were obtained, further optimization was performed by separating the two major events of crystallization, i.e. nucleation and crystal growth, via several crystallization optimization methods including microseeding. In microseeding, submicroscopic crystals were prepared and used as a seed to introduce equilibrated new drops.

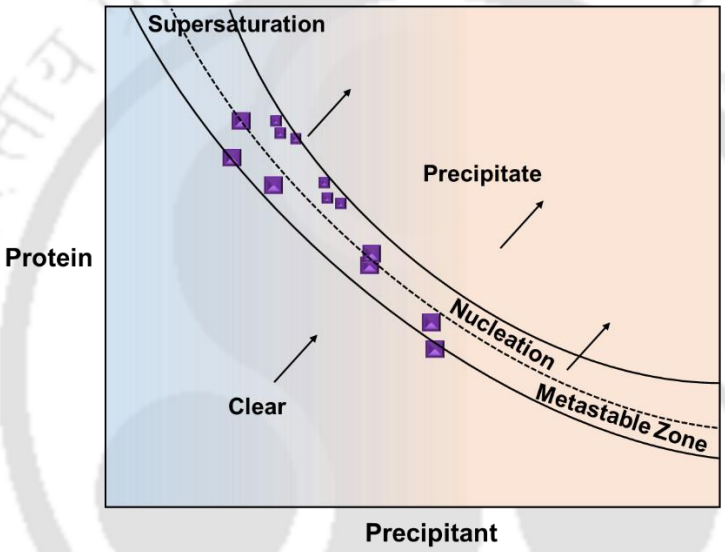


Figure 2.3. Schematic representation for the phase diagram of protein crystallization. (Image is redrawn from Stewart et al., 2011).

2.2.7. Data collection and processing

2.2.7.1. Data collection

Once a suitable protein crystal was obtained for the selected proteins, they were exposed to a monochromatic beam of X-ray to get the diffraction patterns. Since crystal contains a periodic arrangement of atoms, therefore, molecules, X-ray scattering occurs either in a constructive or destructive way. Bragg`s law defines the relationship between constructive

interference and the geometry of crystal parameters. According to Bragg`s law, the X-ray scattering depends on the incident beam wavelength (λ), the glancing angle (2θ) and a positive integer (n).

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

In this study, all the X-ray intensity data were collected at a fixed wavelength (λ) of the incident X-ray beam (1.5418 Å), which is recorded on an image plate (IP) detector. For data collection with all possible diffraction peaks, the angle of oscillation for the diffracted crystals was kept as 1.0° (1.0 degree per image) during data collection. Depending on the crystal's space group, the desired number of images or frames were collected for the complete data set. Since the data was collected at a fixed wavelength (λ), other parameters such as the crystal-to-detector distance (d) in a range of 80 - 200 mm and exposure time (300 - 600 s) were considered to improve the data collection process with reasonable statistics.

2.2.7.2. Data processing

In X-ray data processing, the raw diffraction data is converted into Bragg reflections with measured intensities. It follows a four-step process: calibration, integration, merging and scaling. For each step, different algorithms are derived, which are available in software packages. In the present study, data processing for the raw diffraction images was performed using the program iMosflm for initial processing (Battye et al., 2011), Aimless for scaling and merging (Evans and Murshudov, 2013). The program iMosflm is embedded as part of the CCP4 suite. The program iMosflm performs indexing to provide an index for each reflection and initial statistics on unit cell dimensions in a crystallographic symmetry. To calculate the possible point and space groups, data sets were subjected for scaling and merging using the program Aimless. Conversion of the obtained intensities into structure factors was performed by using the module ctruncate distributed in the CCP4 package (Winn et al., 2011).

2.2.7.3. Structure determination

Obtained intensities from the X-ray diffraction methods provide the information about structure amplitudes only but unable to give the phase angle information. Information of

phase angles is essential for the calculation of the electron density (Taylor GL., 2010). Several methods are known to obtain the initial phase information, which includes: 1) molecular replacement (MR), (2) single- and multiple-wavelength anomalous dispersion (SAD and MAD) and (3) Molecular isomorphous replacement (MIR). To obtain the phase information for the structural solution of the selected SBPs, the molecular replacement was used in this study (Evans and McCoy, 2008). In MR, a known molecular structure of the homologous protein is used to solve the structure of a new target protein using the program Phaser (McCoy et al., 2007). Phaser utilizes multivariate statistics and maximum likelihood probability theory. In order to obtain the initial phase by MR, it requires a significant sequence similarity (>25%) between the known homologous protein and the targeted unknown protein. The phase information obtained from the model structure was further utilized to calculate the phases from the target protein intensities.

2.2.7.4. Model building and structure refinement

Once the initial phase angles were obtained from the target protein reflections, an electron density map was calculated. The electron density map helps to build a model structure by tracing and modeling the secondary structural elements (α-helices and β-sheets) using the program Coot (Crystallographic Object-Oriented Toolkit, Emsley et al., 2010). During model building, chemical and physical parameters such as bond lengths, bond angles, torsion angles and atomic overlaps, etc. of the model was analyzed using the programs such as PDB Goodies (Hussain et al., 2002) and PSAP (Protein Structure Analysis Package, Balamurugan et al., 2007).

In the present study, model refinement was done using the program Refmac5 (Vagin et al., 2004), which utilizes Bayesian statistics and maximum likelihood probability. Model refinement improves the agreement between the calculated structure factors from the model and observed structure factors from the experiment. After each refinement cycle, an agreement between the calculated structure factors (F_{calc}) was compared with that of the measured structure factors (F_{obs}) and can be represented as:

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

After refinement, the improvement in the model was validated by inspecting the R-factor values and electron-density maps ($2F_O$-F_C and F_O-F_C) contoured at 1.0σ and 3.0σ, respectively.

2.2.7.5. Cross-validation

The quality of the model was evaluated by statistical cross-validation using R-factor (Brunger, 1992). For cross-validation, the before refinement diffraction data were divided into two sets. 5-10% of the diffraction data were removed from the complete data set as a small complimentary test set, while the remaining 90-95% of the data were used as a large working set during the refinement process. The remaining 10% test data set was used to evaluate the accuracy of the predicted model by calculating the R-free value:

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

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

2.2.8. Structure validation

The reliability of the three-dimensional structures of SBPs was evaluated by the structure validation process. In general, a structure validation process has three aspects: 1) validity of the experimental measurements, 2) consistency in the atomic model and 3) consistency of the known chemical and physical properties of the atomic model. For structure validation of the selected proteins, the following aspects haave been considered in this study: (1) geometrical parameters and (2) root mean square deviation (RMSD) of the refined structures. For structure validation, a number of online web servers and tools are availabl,e out of which PROCHECK (Laskowski et al., 2012) and MolProbity (Chen et al., 2010) were used for the selected SBP in the present study.

2.2.8.1 PROCHECK

After every refinement cycle, the stereochemical quality of the model structure was assessed using PROCHECK (Laskowski et al., 1993). This program checks the geometry of the model structure residue-by-residue. PROCHECK provides the following information: Ramachandran plot, Gly & Pro Ramachandran and Chi1-Chi2 plots, main-

chain and side-chain parameters, residue properties, main-chain bond length and angle distributions and RMS distances from planarity and distorted geometry plots. Among these outputs, the Ramachandran plot was used to estimate the geometry of the selected SBPs.

2.2.8.2. MolProbity

MolProbity is a web server used for structure validation (Davis et al., 2007). This program evaluates the three-dimensional structures of the proteins both at global and local levels. For a given crystal structure, MolProbity optimizes both polar and nonpolar hydrogens, all-atom contact analysis, covalent-geometry and torsion-angles of the protein. Additionally, it also enhances the local chemical environment due to the resolution of structures, which includes Ramachandran outliers and flipped branched protein side chains. In the present study, the program MolProbity was used to detect the Ramachandran and torsion-angles outliers.

2.2.9. Domain movement analysis of SBPs

To elucidate the conformational changes in the N- and C-terminal domains (NTD and CTD) of the selected SBPs upon ligand binding, the domain movement of the protein was analyzed using the online program DynDom (Hayward and Berendsen, 1998; Taylor et al., 2014). DynDom server calculates the rotational angle as well as the translational movement of the domains in a protein. Additionally, it also provides information on the residues involved in hinge axes and hinge bending between two conformations of the SBPs (open and closed).

2.2.10. Structure and protein-ligand interaction visualization

In this study, the program PyMOL (The PyMOL Molecular Graphics System, Schrodinger, LLC) was used to visualize all the three-dimensional structures of proteins, to locate the active site of the already known and modeled structures by comparing them. In addition, the program was also used to generate all the molecular graphic figures used in this study. To identify the intermolecular interaction(s) between the cognate as well as docked ligand and the SBPs are analyzed using the program Coot and PoseView (Stierand et al., 2006; Emsley et al., 2010).

CHAPTER 3

Metal uptake ABC transporters in *T. thermophilus* HB8

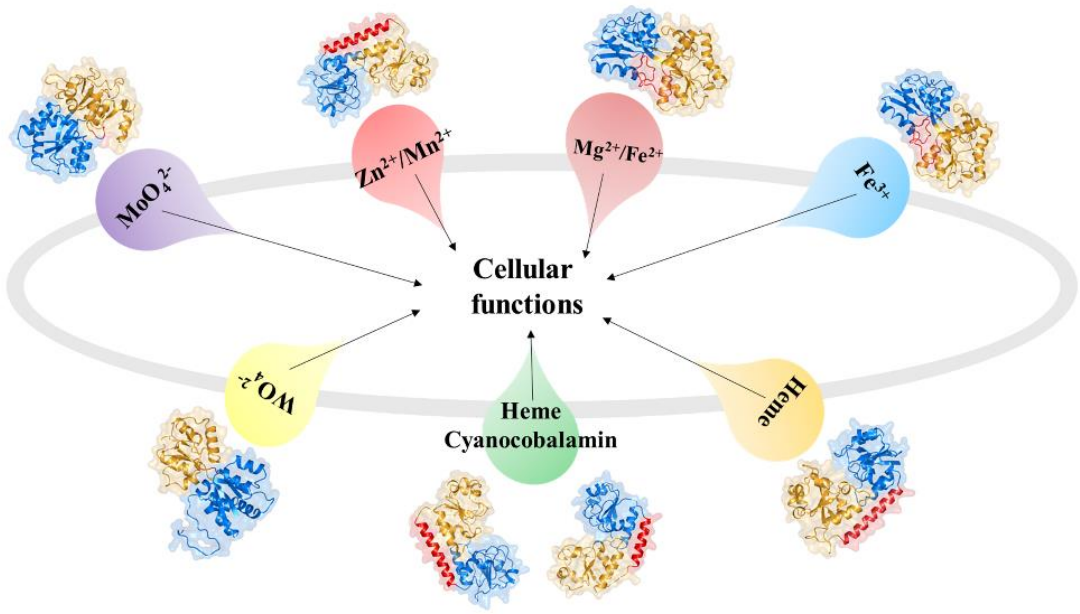
This chapter has been published as:

1. **Mandal SK**, Adhikari R, Sharma A, Chandravanshi M, Gogoi P and Kanaujia SP (2019). Designating ligand specificities to metal uptake ABC transporters in *Thermus thermophilus* HB8. ***Metallomics***, 11, 597-612.
2. **Mandal SK**, Chandravanshi M, Gogoi P and Kanaujia SP (2017). *In silico* characterization of TTHA0596: A potential Zn^{2+} binding protein of ATP-binding cassette transporter. ***Gene Reports***, 6, 132-141.

ABSTRACT

Micronutrients such as metal ions are indispensable for the growth and survival of microorganisms in assorted environmental niches. However, change in the cellular concentration of metal ions is pernicious for an organism; thus, metal ion homeostasis is crucial for their survival and growth. An eminent mechanism for maintaining metal ion homeostasis in microorganisms is ATP-binding cassette (ABC) transporters, which transport metal ions in their ionic/complex forms across the cell membrane. For the uptake, metals are sequestered by substrate-binding proteins (SBPs) and transferred to transmembrane domains (TMDs) for their transport. In this work, a high-throughput data mining analysis has been performed to identify open reading frames (ORFs) encoding metal-specific ABC transporters in a thermophilic bacterium *Thermus thermophilus* HB8. In total, 22 ORFs resulting in eight ABC transport systems were identified, which are potentially involved in the uptake of metal ions. This study suggests that three out of eight metal-specific ABC import systems are specific to iron ions. Among the remaining five, two are particular to divalent metal ions such as Mg^{2+} and Zn^{2+}/Mn^{2+}; another two are for tetrahedral oxyanions such as MoO_4^{2-} and WO_4^{2-} and the remaining one imports cyanocobalamin (vitamin B12). Besides these, the results of this study demonstrate the existence of a mechanism where TMD and NBD components are shared among different ABC transport systems hinting that multiple substrates can be imported via a single transporter. This study, thus, provides the first-ever preliminary glimpse into the entire repertoire of metal uptake ABC transporters in a thermophilic organism.

Graphical abstract



3.1. INTRODUCTION

A large number of proteins function in various cellular processes by utilizing different metal ions, especially the transition metal ions, for accomplishing their assigned structural and/or functional roles (Holm et al., 1996; Waldron and Robinson, 2009). The d-block or transition-group metal ions such as manganese, iron, cobalt, nickel, copper, zinc and to some extent, vanadium, molybdenum and tungsten, have been reported to play crucial roles in biological processes (Zhen et al., 2009). Iron, for instance, has the ability to fine-tune the regulation of many cellular, biosynthetic and metabolic processes owing to its unstable redox chemistry, making it a cofactor obligatory for each organism (Crichton and Boelaert, 2001). Divalent metal ions also regulate and inhibit many enzymatic reactions such as protein kinase activity (Knape et al., 2017). Similarly, molybdenum, in its oxyanionic form (i.e., molybdate, MoO_4^{2-}), is utilized by *Pseudomonas aeruginosa* for nitrate reduction to produce energy (Pedrick et al., 2014). Succinctly, metal ions are pivotal for a plethora of cellular processes, including DNA replication, protein biosynthesis, cell wall biosynthesis, cellular respiration, signal transduction, etc. (Waldron and Robinson, 2009; Kehres and Maguire, 2003; Hasse and Rink, 2007; Kehl-Fie and Skaar 2010; Counago et al., 2012). Although highly essential, alteration (either increase or decrease) in metal ion

concentration turns out to be toxic for microorganisms and therefore, its homeostasis becomes a crucial cellular phenomenon (Klein and Lewinson, 2011). In microorganisms, metal ion homeostasis is governed by high-affinity transport systems. Among many others, ATP-binding cassette (ABC) transporters are one of the most efficient transport systems involved in the transport of metal ions (Klein and Lewinson, 2011; Wilkens, 2015).

ABC transporters are functionally involved in the unidirectional translocation of substrates across cellular membranes and are broadly classified into exporters and importers (Wilkens, 2015). The efflux of a substrate(s) from the cytoplasm is carried out by ABC exporters, which is ubiquitous to all domains of life (Wilkens, 2015). On the other hand, ABC importers extract substrate(s) from the extracellular milieu into the cytoplasmic space, which is notably present only in bacteria, archaea and lately identified in plants (Shitan et al., 2003; Terasaka et al., 2005). Architecturally, although both ABC exporters and importers possess a transmembrane domain (TMD) and a nucleotide-binding domain (NBD), ABC importers contain an additional substrate-binding protein (SBP) (Wilkens, 2015; Lewinson and Livnat-Levanon, 2017). SBPs are responsible for maintaining the specificity and selectivity of the substrate(s) as well as the directionality of the transport (Webb et al., 2008). Even though SBPs possess a low sequence similarity, they show high structural resemblance (Berntsson et al., 2010). Based on its architecture and the types of substrate(s) being transported, SBPs have been classified into seven different clusters viz. A-G (Scheepers et al., 2016). The SBPs of metal uptake ABC transporters have been grouped into clusters A and D, further subclustered into A-I to A-II and D-I to D-IV, respectively (Berntsson et al., 2010). Subcluster A-I binds particularly divalent metal ions, whereas members of A-II are responsible for sequestering metal ion complexes like siderophores and cyanocobalamin. Cluster D members show a broad substrate specificity, including carbohydrates (D-I), polyamines (D-II), tetrahedral oxyanions (D-III) and iron ions (D-IV).

Bacterial species living in different environmental niches have a varied number of ABC transporters encoded by their genome to adapt to different conditions (Harland et al., 2005). *Thermus thermophilus* is one such extensively studied gram-negative thermophilic bacterium dwelling in marine hot springs, which possesses at least 42

ABC transporters (Henne et al., 2004). However, the total number of metal uptake ABC transporters in *T. thermophilus* is unknown, as most of the ORFs encoding these systems are either unidentified or annotated as 'hypothetical proteins'. Thus, in this study, the complete genomic framework of *T. thermophilus* HB8 has been analyzed to identify and characterize the metal ABC importers based on the unique features of SBPs. Identification of the metal ABC importers in *T. thermophilus* also provides an insight into the repertoire of metal uptake ABC transporters in a thermophilic organism.

3.2. MATERIALS AND METHODS

3.2.1. Sequence analysis

The list of open reading frames (ORFs) encoding metal-specific components of ABC transporters from *T. thermophilus* HB8 were identified and obtained from the TransportDB 2.0 and Structural-Biological Whole-Cell Project (http://www.thermus.org) databases (Elbourne et al., 2017). The amino acid sequences and annotations of these ORFs were extracted from the UniProtKB database (The UniProt Consortium, 2018). The genetic cluster of each identified ORF and the functionally-associated genes were extracted from the Gene database available at the National Center for Biotechnology Information (NCBI). The protein-protein interaction analyses of each protein were performed using the database STRING v10 (Szklarczyk et al., 2017). The sequence homology search was performed using the web tool BLAST (Altschul et al., 1990). To study the evolutionary relationships among SBPs, a phylogenetic tree was generated using the neighbor-joining (NJ) method with bootstrapping of 1000 replicates using the program MEGA7 (Kumar et al., 2016). To identify the conserved regions in the protein sequences, multiple sequence alignment (MSA) was executed using the program Clustal Omega (Sievers and Higgins, 2014). Subsequently, the MSAs were further rendered using the web tool ESPript3.0 for visual clarity (Gouet et al., 2003).

3.2.2. Structure analysis

The three-dimensional structures of all the SBPs considered in this study were predicted using the programs RaptorX (Kallberg et al., 2012), SWISS-MODEL (Biasini et al., 2014), I-TASSER (Yang et al., 2015) and Phyre2 (Kelley et al., 2015). Subsequently, the predicted models were refined by an energy-minimization method using the

programs UCSF Chimera and ModRefiner (Pettersen et al., 2004; Xu and Zhang, 2011). The stereo-chemical properties of all the refined models were validated using the web-server RAMPAGE (Lovell et al., 2003); models with the fittest geometrical parameters were chosen for further analysis. The structural homologs of each SBP were searched using the web-server Dali (Holm and Rosenstrom, 2010). The root mean square deviation (RMSD) among all the selected 106 SBPs were calculated using a home-built python script and the program PyMOL (PyMOL Molecular Graphics System, Schrodinger, LLC). The RMSD values obtained were subsequently used to generate a structure-based distance tree for the study of structural relationships. The three-dimensional structures were compared using the web-server 3dSS to examine the spatial orientation of the residues at the active-site pocket (Sumathi et al., 2006). The area and volume of the active-site pocket of each SBP were computed using the program CASTp with a default probe radius of 1.4 Å (Binkowski et al., 2003). The electrostatic potential charge distribution of each SBP was calculated using the module PDB2PQR embedded in the tool APBS plugged in the program PyMOL (Jurrus et al., 2018).

3.2.3. Molecular docking studies

To estimate the binding affinities of different metal ions/complexes to SBPs, molecular docking experiments were performed using a freely available program AutoDock version 4.0 (Morris et al., 2009). For this, the three-dimensional atomic coordinates of each metal ion/complex were extracted from their protein-bound states available in the Protein Data Bank (PDB) (Berman et al., 2000). In the cases of the closed conformation of proteins, the orientation of the active-site residues was adjusted using the program Coot to accommodate the ligands (Emsley and Cowtan, 2004). Hydrogens were added using the module available in AutoDock. The partial charges for the protein atoms were assigned using the Gasteiger charge algorithm (Gasteiger and Marsili, 1978; Gasteiger and Marsili, 1980). The partial charge for the metal ions viz. magnesium (Mg^{2+}) and iron (Fe^{2+}) was kept as +0.8 (Chen et al., 2007) while that of molybdenum (Mo) and tungsten (W) was assigned +1, as the oxygen atoms in molybdate (MoO_4^{2-}) and tungstate (WO_4^{2-}) were assigned a partial charge of -0.25 each. Similarly, the partial charge of cobalt (Co^{2+}) was assigned to +0.857 as predicted by the Gasteiger charge algorithm. Thus, a modified version of the AD4_parameters available in the literature

(http://bioinf.modares.ac.ir/Courses/Docking/AD4_parameters.dat) was used. The grid size was fixed to 126x126x126 points, keeping a grid spacing of 0.375 Å, taking the center of mass of the protein as the grid center. In each molecular docking experiment, the protein atoms were kept rigid while the ligand molecules were flexible about the rotatable bonds. For each molecular docking experiment, a total of 2000 runs of the Lamarckian genetic algorithm (GA) was simulated. The docked conformations of ligands in each molecular docking experiment were clustered with an RMSD cut-off of 2.0 Å. The list of interactions between the protein and ligand atoms was visualized and identified using the program Coot (Emsley and Cowtan, 2004). All the structural figures, including the molecular interactions of docked poses, were prepared using the program PyMOL.

3.3. RESULTS

3.3.1. Metal uptake ABC transporters in *T. thermophilus* HB8

The repertoire of metal uptake ABC transporters in the thermophilic bacterium *T. thermophilus* HB8 was collected by searching the ORFs having the phrase "metal uptake ABC transporter" from TransportDB (Elbourne et al., 2017) and Structural-Biological Whole-Cell Project (http://www.thermus.org) databases. Unlike the mesophilic bacterium *Escherichia coli* K-12, which possesses seven metal-specific ABC transporters encoded by *btuFCD*, *fecBCDE*, *fepBDGC*, *fhuDBC*, *znuACB*, *nikABCDE* and *modABC* (Moussatova et al., 2008), *T. thermophilus* HB8 contains 22 ORFs encoding SBPs, TMDs and NBDs resulting in eight metal-specific ABC transport systems (Table A1). As SBPs are unique to ABC importers and determine their selectivity and specificity (Biemans-Oldehinkel et al., 2006), they were chosen as the target protein for the functional characterization of the ABC transport systems. Out of the eight SBPs, one (ORFs: TTHA1628) has already been functionally characterized as Fe^{3+}-carbonate-binding (FbpA) proteins, respectively (Wang et al., 2014). Thus, in this study, the remaining seven SBPs (ORFs: TTHB220, TTHA0723, TTHA0746, TTHB177, TTHA0069, TTHA0717 and TTHA0596) have been characterized for their functional annotation.

3.3.2. Metal uptake SBPs cluster into functionally distinct clades

In the UniProtKB database, three ORFs TTHA0723, TTHA0746 and TTHB177 have been annotated as "iron (III) ABC transporter, solute-binding protein", "iron ABC transporter, substrate-binding protein" and "iron ABC transporter, periplasmic iron-binding protein", respectively. The availability of multiple isoforms of Cu^{+} and other metal ion-ATPases has been recognized earlier in microorganisms which tend to possess a compact genome size (Arguello et al., 2011). Thus, to confirm the presence of three iron-binding SBPs and to accomplish their function in *T. thermophilus* HB8, a homology search analysis was performed. Unlike the annotation provided in the UniProtKB database, TTHA0723 and TTHA0746 show the highest similarity to cyanocobalamin-binding (BtuF) proteins (sequence identity: 24-32%, query coverage: 70-95%). On the other hand, TTHB177 shows significant similarity to Mg^{2+}-citrate-binding protein (PDB ID: 3C9H, sequence identity: 37%, query coverage: 89%). Since, in *T. thermophilus* HB8, another ORF TTHA1628 has already been established as a ferric (Fe^{3+})-binding protein (FbpA, PDB ID: 3WAE) (Wang et al., 2014) and TTHB177 protein is also annotated for iron transport, a comparative study of TTHB177 with TTHA1628 was performed for the identification of their physiological ligand. Furthermore, the ORFs TTHA0069 and TTHA0717 are annotated as "ABC transporter, periplasmic binding protein" and "molybdenum ABC transporter molybdate-binding protein", respectively, in the UniProtKB database. A homology search analysis reveals that both the proteins TTHA0069 and TTHA0717 show similarity to WO_4^{2-}- and/or MoO_4^{2-}-binding proteins (TupA/WtpA/ModA) (sequence identity: 24-40%, query coverage: 79-97%). On the other hand, TTHB220 is annotated as "hemin ABC transporter, periplasmic hemin-binding protein" in the UniProtKB database and exhibits the highest similarity with heme-binding proteins (sequence identity: 77-92%, query coverage: 22-49%). The protein TTHA0596 is annotated as an ABC transporter solute-binding protein in the UniProtKB database. However, the type of solute molecule it can bind and transport is not yet revealed. The homology search result reveals the highest homology of TTHA0596 with metal-binding proteins belonging to cluster A SBPs.

To mitigate the uncertainty regarding the functional annotation of the metal uptake SBPs, evolutionary analysis was performed along with their homologous proteins. The result reveals that the protein TTHB220 and TTHA0746 are evolutionarily closer to

heme- and cyanocobalamin-binding proteins, respectively (Figure 3.1). Interestingly, the protein TTHA0723 is grouped in between heme- and cyanocobalamin-binding proteins showing its probable binding with both type of molecules. In agreement with Wang et al. (2014), the protein TTHA1628 is grouped with other ferric-binding proteins, whereas TTHB177 is evolutionarily closer to *Af*Atu, which is an Mg^{2+}-citrate-binding protein and not an iron-binding protein (Figure 3.1). The phylogenetic tree exhibits that the proteins TTHA0069 and TTHA0717 are evolutionarily closer to WO_4^{2-}-binding TupA and MoO_4^{2-}-binding ModA proteins, respectively. The results also exhibit that the protein TTHA0596 is evolutionarily closer to divalent metal ion-binding proteins (Figure 3.1).

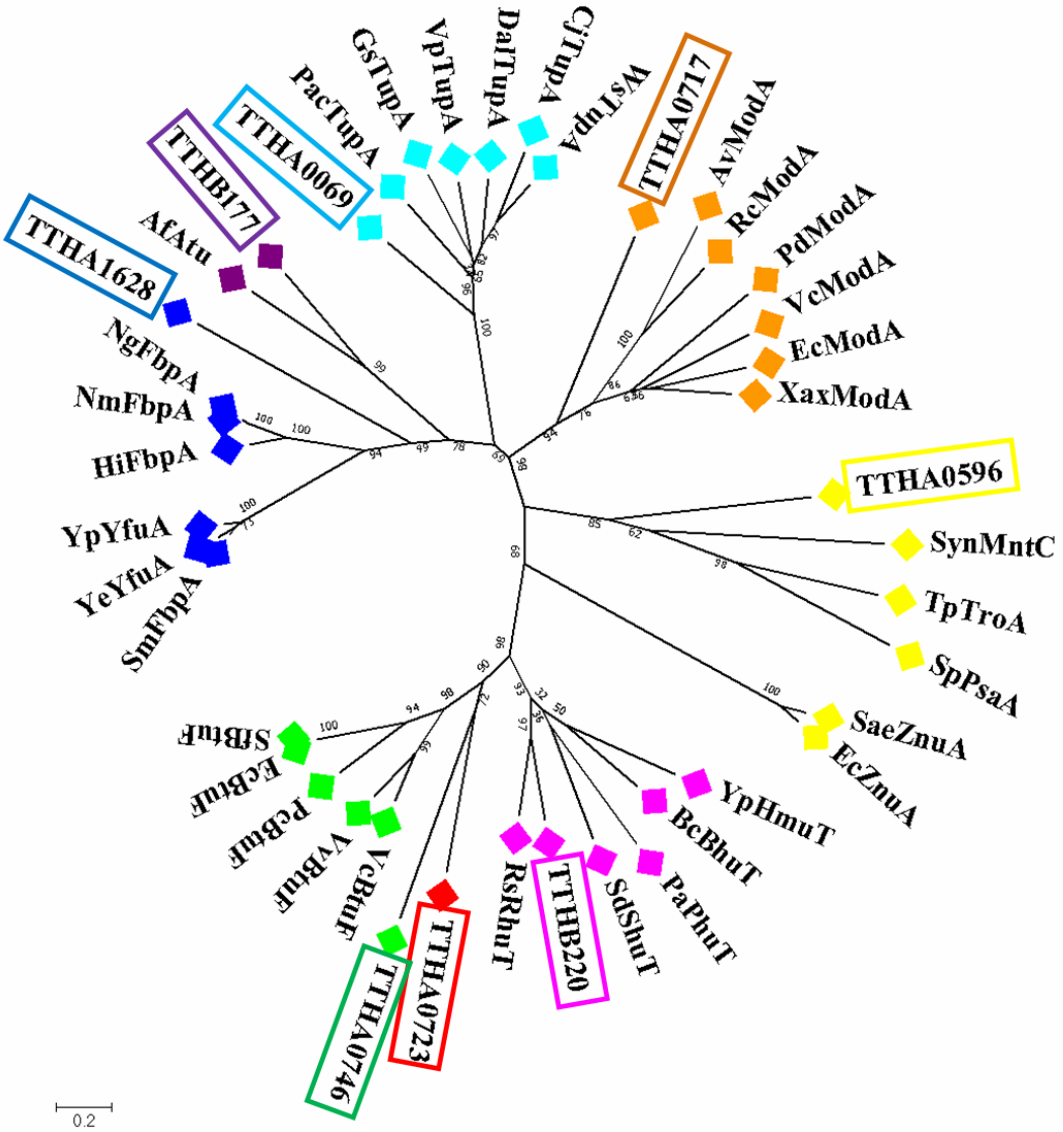


Figure 3.1. Phylogenetic analysis of the metal uptake SBPs along with their homologous proteins. The proteins used to build the phylogenetic tree are from *Roseiflexus spp.* (*Rs*RhuT, A5UZ69), *Shigella dysentriae* (*Sd*ShuT, O70018), *Burkholderia cenocepacia* (*Bc*BhuT, B4EKB3), *Yersinia pestis* (*Yp*HmuT, Q56991), *P. aeruginosa* (*Pa*PhuT, O68879), *Vibrio cholera* (*Vc*BtuF, Q9KPI6), *V. vulnificus* (*Vv*BtuF, Q7MNT2), *Pectobacterium carotovorum* (*Pc*BtuF, C6DC28), *Escherichia coli* (*Ec*BtuF, P37028), *S. flexneri* (*Sf*BtuF, Q83MD7), *Serratia marcescens* (*Sm*FbpA, P21408), *Y. enterolitica* (*Ye*YfuA, A1JLH5), *Y. pestis* (*Yp*YfuA, A0A0E1NNQ6), *Haemophilus influenzae* (*Hi*FbpA, P35755), *Neisseria meningitidis* (*Nm*FbpA, P0A0Y4), *N. gonorrhoeae* (*Ng*FbpA, P17259), *Agrobacterium fabrum* (*Af*Atu, Q7CWZ6), *Peptoclostridium acidaminophilum* (*Pac*TupA, Q93KD6), *Geobacter sulfurreducens* (*Gs*TupA, Q749P2), *V. parahaemolyticus* (*Vp*TupA,Q87PK2), *Desulfovibrio alaskensis* (*Dal*TupA, Q316W1), *Campylobacter jejuni* (*Cj*TupA, Q0P885), *Wolinella succinogenes* (*Ws*TupA, Q7M8V9), *A. vinelandii* (*Av*ModA, Q7SIH2), *Rhodobacter capsulatus* (*Rc*ModA, Q08383), *V. cholerae* (*Vc*ModA, Q9KLL7), *P. difficile* (*Pd*ModA, Q18A64), *E. coli* (*Ec*ModA, P37329), *Xanthomonas axonopodis* (*Xax*ModA, Q8PHA1), *Salmonella enterica* (*Sae*ZnuA, Q8ZNV8), *E. coli* (*Ec*ZnuA, P39172), *Treponema pallidum* (*Tp*TroA, P96116), *SynechocystiS spp.* (*Syn*MntC, Q79EF9), *Streptococcus pneumoniae* (*Sp*PsaA, P0A4G2); the UniProt ID of each protein is mentioned in the parenthesis. The result shows TTHB220 (pink box) clades with heme- (pink diamond) and TTHA0746 (green box) with cyanocobalamin-binding proteins (green diamond), whereas TTHA0723 (red box) clusters in between both of them. TTHB177 (violet box) clusters with the Mg^{2+}-citrate-binding protein (violet diamond), whereas the protein TTHA1628 (blue box) is grouped along with ferric-binding proteins (blue diamond). Phylogenetic tree showing the clustering of the protein TTHA0069 (cyan box) with WO_4^{2-}-binding proteins (cyan diamond), whereas that of TTHA0717 (orange box) with MoO_4^{2-}-binding proteins (orange diamond). The results exhibit the clading of the protein TTHA0596 (yellow box) with divalent metal ion binding proteins (yellow diamond).

3.3.3. Thermophilic metal uptake SBPs are co-localized with their cognate TMDs and NBDs, except TTHA0723

As genes encoding ABC components are mostly found to be located in an operonic form on the genome, we investigated the genetic organization of all the metal uptake ABC transporters in *T. thermophilus* HB8 and made a comparison with a mesophilic organism *E. coli* to distinguish the unique feature of metal uptake ABC transport system of a thermophilic organism. The results reveal that *T. thermophilus* HB8 possesses two heme uptake ABC transport systems (Figure 3.2A), whereas no heme uptake system exists in *E. coli* K-12. Apart from heme uptake systems, *T. thermophilus* HB8 also has a transporter (ORFs: TTHA1628-TTHA1630) exclusive for iron transport. *E. coli* K-12, on the other hand, lacks ABC transporters for heme uptake while possesses three transporters viz. FecBCDE, FhuDBC and FepBDGC for the uptake of iron (Figure 3.2B). *E. coli* K-12 have two divalent metal uptake ABC transporters viz. NikABCDE and ZnuABC for Ni^{2+} and Zn^{2+} uptake, respectively, where NikABCDE is comprised of two TMD and NBD subunits. Likewise, *T. thermophilus* HB8 also possesses transport systems (ORFs: TTHA0596-TTHA0598 and TTHB175-TTHB177) for Zn^{2+}/Mn^{2+} and Fe^{2+}/Mg^{2+}-citrate uptake, respectively, while no gene encoding an ABC transporter for Ni^{2+} uptake is available (Figure 3.2C). Similarly, both the organisms owns an ABC transport system for the uptake of cyanocobalamin (Co^{2+} bound to tetrapyrrole ring) (Figure 3.2D). For uptake of tetrahedral oxyanions, *E. coli* K-12 possesses a single uptake system (ModABC), whereas *T. thermophilus* HB8 has two distinct transporters (ORFs: TTHA0069-TTHA0071 and TTHA0715-TTHA0717) for WO_4^{2-} and MoO_4^{2-} uptake, respectively (Figure 3.2E).

In *E. coli* K-12, the genes encoding the cyanocobalamin uptake ABC transport system lacks synteny and two of the iron uptake ABC systems (FecBCDE and FepBDGC) possess two TMDs. On the other hand, in *T. thermophilus* HB8, five of the SBPs (ORFs: TTHB220, TTHA0746, TTHB177, TTHA0069 and TTHA0717) are arranged in tandem with their single cognate TMDs and NBDs (Figure 3.2A-E). However, TTHA0723 lacks its cognate TMD and NBD components and is flanked by two unrelated enzymes, namely histidinol dehydrogenase (ORF: TTHA0722) and serine protease (ORF: TTHA0724). Thus, to identify the cognate TMD and NBD subunits of TTHA0723, an interactome map analysis was performed. The result predicts its relationship with the TMD and NBD components of another putative iron ABC

transporter (ORFs: TTHA0747-TTHA0748) (Figure 3.2F and Table A2). However, ORFs: TTHA0747-TTHA0748 have already been identified as the cognate TMD and NBD components of the SBP (ORF: TTHA0746); this is supported by its interactome map analysis as well (Figure 3.2F and Table A2). This suggests that the protein TTHA0723 shares the TMD and NBD components of another metal uptake ABC transporter (ORFs: TTHA0747-TTHA0748).

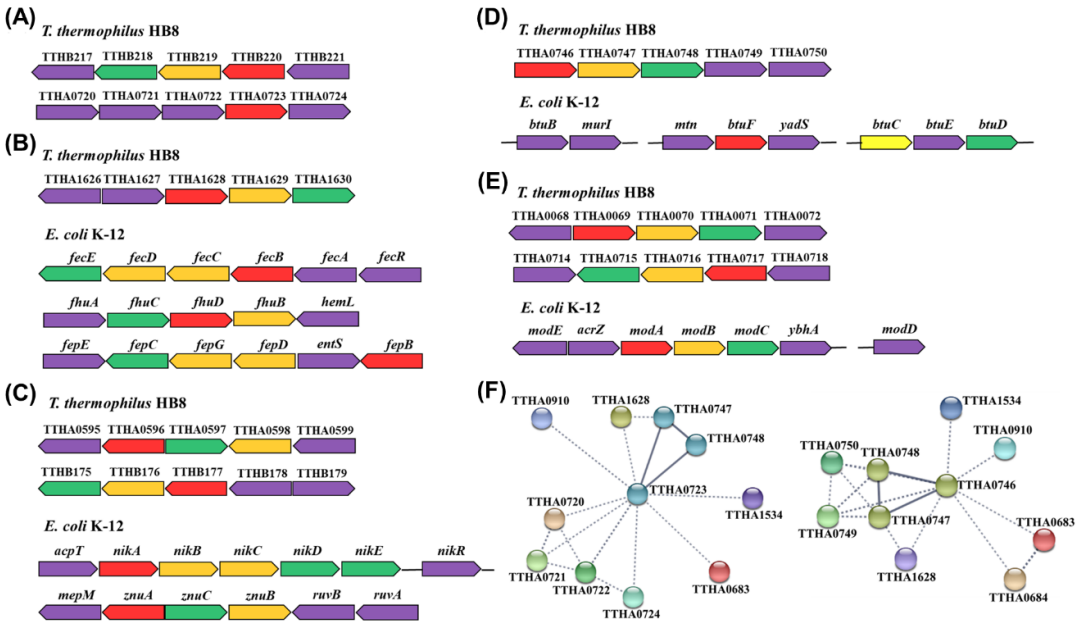


Figure 3.2. Genetic organization of (A) TTHB220 and TTHA0723 from *T. thermophilus* HB8, (B) TTHA1628 from *T. thermophilus* HB8 and *fecBCDE*, *fhuDBC* and *fepBDGC* from *E. coli* K-12, (C) TTHA0596 and TTHB177 from *T. thermophilus* HB8 and *nikABCDE* and *znuACB* from *E. coli* K-12, (D) TTHA0746 from *T. thermophilus* HB8 and *btuFCD* from *E. coli* K-12, (E) TTHA0069 and TTHA0717 from *T. thermophilus* HB8 and *modABC* from *E. coli* K-12. The SBPs are shown in red arrows and their cognate TMD (yellow arrow) and NBD (green arrow) are labeled with their respective gene names. The other functionally associated genes are represented in the purple arrow. (F) Protein-protein interaction network of the protein TTHA0723 and TTHA0746 with their interacting partners. The interactions of metal uptake SBPs with their cognate TMD and NBD subunits are highlighted in bold lines. The ORF number is provided along with each node.

3.3.4. Metal uptake SBPs from *T. thermophilus* HB8 belong to clusters A and D

Primary structure analysis provided a preliminary idea of the functions of each of the metal uptake SBPs. To gain the knowledge about the cognate function of all the seven SBPs at the tertiary structure level, their three-dimensional structures were modeled. The tertiary model for each SBP viz. TTHB220, TTHA0723, TTHA0746, TTHB177, TTHA0069, TTHA0717 and TTHA0596 were modeled using homologous proteins as template mentioned in Table A3. The modeled structures display an overall topology similar to classical SBPs of ABC transporters comprising of two globular α/β domains viz. N- and C-terminal domains connected by a hinge region. A difference could, however, be perceived between the heme- & cyanocobalamin-binding proteins and divalent metal ion- & oxyanion-binding proteins. In TTHB220, TTHA723, TTHA0746 and TTHA0596 (heme-, cyanocobalamin and divalent metal ion-binding proteins), the N- and C-terminal domains are linked by a long and rigid α-helix similar to cluster A SBPs (Figure 3.3A). On the other hand, in TTHB177, TTHA0069 and TTHA0717 (divalent metal ion & oxyanion-binding proteins), the N- and C-terminal domains are connected by a flexible loop similar to cluster D SBPs (Figure 3.3B).

To further determine the subcluster of the seven metal uptake SBPs, a structure-based distance tree containing their structural homologs and SBPs belonging to different clusters (Table A4 and A5) was built. The analysis shows that the proteins TTHB220, TTHA0723 and TTHA0746 are distinctly placed in subcluster A-II (Figure 3.3C), indicating their probable substrates to be a metal-complex/siderophore. The results also exhibit that the protein TTHA0596 clades with cluster A-I SBPs suggesting its probable ligand to be a transition row divalent metal ion (Figure 3.3C). A search for the structural homologs of the protein TTHB177 reveals Mg^{2+}-citrate- and iron-binding proteins, including TTHA1628, as its closest members (Table A4). However, the structural distance tree demonstrates that the protein TTHA1628 and TTHB177 belong to subcluster D-IV (specific to Fe^{3+} ions) and subcluster D-III (specific to metal ion-complexes), respectively (Figure 3.3C and Table A5). Akin to TTHB177, the proteins TTHA0069 and TTHA0717 also belong to the subcluster D-III of SBPs (Figure 3.3C and Table A5).

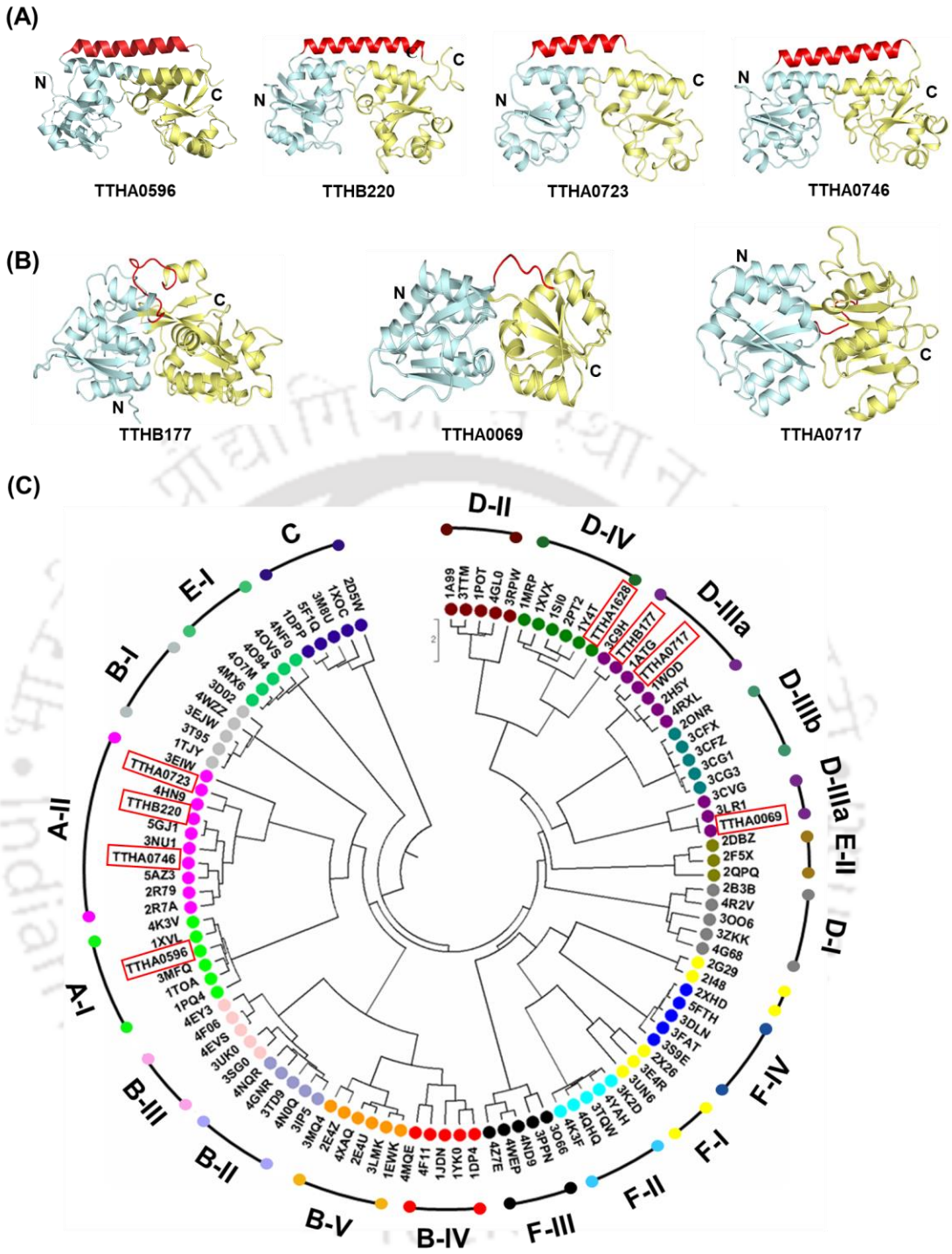


Figure 3.3. The Overall three-dimensional models of the proteins (A) TTHA0596, TTHB220, TTHA0723, TTHA0746 and (B) TTHB177, TTHA0069 and TTHA0717. The N- and C-terminal lobes are shown in cyan and yellow, respectively, with the α-helix/loop linking region in red. (C) Structural distance tree of SBPs. The structures used for the analysis are denoted by their PDB IDs. Each cluster of SBPs is separated

by a diamond symbol in different colors. The tree suggests the grouping of the targeted proteins (highlighted in red box) into distinct clusters.

3.3.5. Active-site pocket of metal-binding proteins are fine-tuned to interact with their cognate substrates

To further substantiate the results, the active-site pockets of the modeled structures of these proteins were analyzed. The electrostatic potential charge distribution of the active-site pockets of TTHB220 and TTHA0723 are observed to be positively charged, much like known heme-binding proteins required for accommodating negatively-charged heme molecules (Figure 3.4A). Similarly, the electrostatic potential charge distribution of TTHA0746 demonstrates the presence of both positively- and negatively-charged residues in its active site, favorable for the binding of a cyanocobalamin molecule (Figure 3.4A). Further, the active-site area and volume of TTHB220 (365 $Å^2$ and 461 $Å^3$) and TTHA0723 (435 $Å^2$ and 812 $Å^3$) are comparable to that of heme-binding *Rs*RhuT (474 $Å^2$ and 934 $Å^3$) and *Bc*BhuT (577 $Å^2$ and 1155 $Å^3$) proteins, respectively (Figure 3.4B). Interestingly, the volume of the protein TTHB220 is lesser (by two folds), suggesting the possible binding of only one heme molecule in its active site, unlike *Rs*RhuT, which is capable of binding two heme molecules simultaneously. The active-site area and volume of TTHA0746 protein (821 $Å^2$ and 1294 $Å^3$) is comparable to that of *Sd*ShuT (963 $Å^2$ and 1363 $Å^3$). However, a twofold increase in the volume of cyanocobalamin-binding BtuF protein from *E. coli* K-12 (*Ec*BtuF) (area: 980 $Å^2$, volume: 2275 $Å^3$) has been observed, which is possible as the tertiary structure of TTHA0746 protein is modeled using a heme-binding protein (*Sd*ShuT) as the template (Figure 3.4B). Furthermore, an analysis of the electrostatic potential charge distribution of TTHA1628 and TTHB177 proteins reveals that the protein TTHA1628 contains a positively-charged active site favorable for Fe^{3+} ion coordinated by two negatively-charged bicarbonates, whereas that of TTHB177 is composed of a mix of negatively- and positively-charged residues required for the binding of citrate and a divalent metal ion (Figure 3.4A). The electrostatic potential charge distribution of TTHA0069 and TTHA0717 confirms the existence of a positively-charged active site required for the binding of the negatively-charged WO_4^{2-} and MoO_4^{2-} ions (Figure 3.4A).

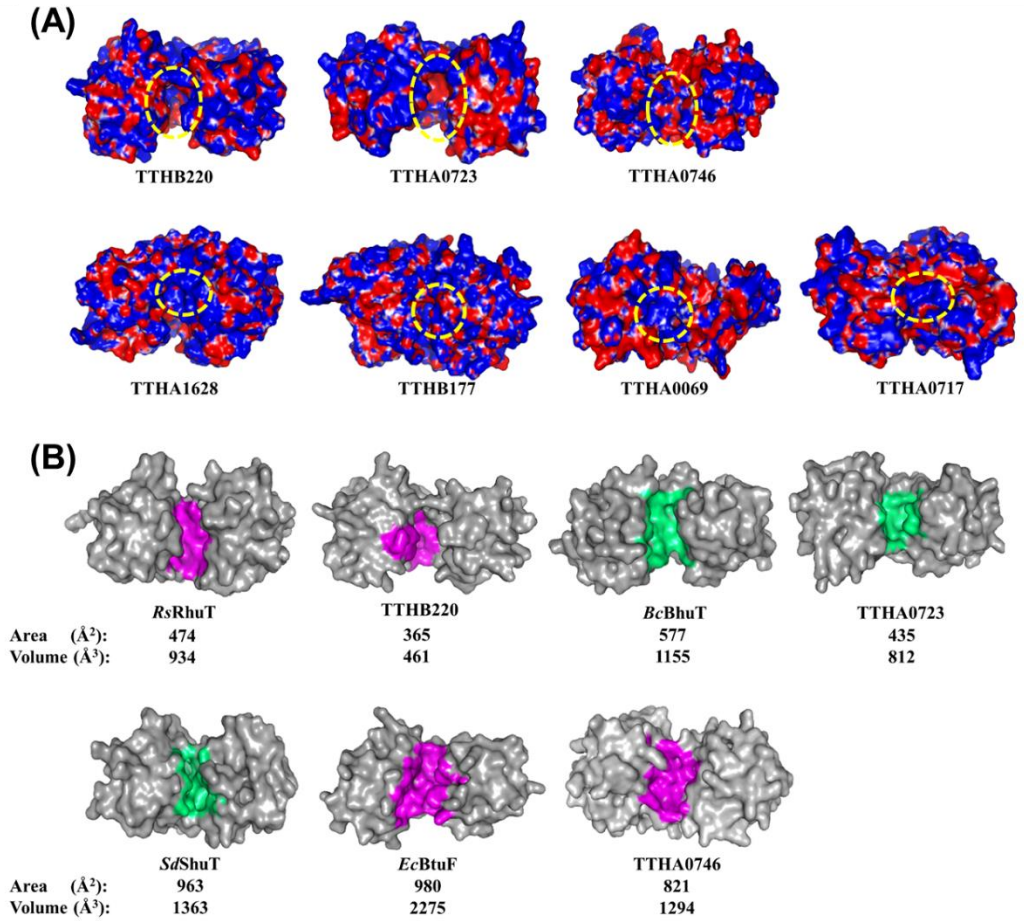


Figure 3.4. (A) Electrostatic potential charge distribution of the metal uptake SBPs. The positively- and negatively-charged surfaces are shown in blue (+1 kT/e) and red (−1 kT/e), respectively. The analysis suggests that the proteins TTHB220, TTHA0723, TTHA1628, TTHA0069 and TTHA0717 possess a positively-charged active site, whereas the active site of TTHA0746 and TTHB177 are composed of both positively- and negatively-charged residues. (B) Comparison of active-site area and volume of TTHB220, TTHA0723 and TTHA0746 with their closest homologs. The analysis shows that the active site of proteins TTHB220 and TTHA0723 are comparable to that of *Rs*RhuT and *Bc*BhuT, respectively. However, the comparison of active-site area and volume of heme- (*Sd*ShuT, PDB ID: 2R7A) and cyanocobalamin-binding (*Ec*BtuF, PDB ID: 1N2Z) proteins along with TTHA0746 suggests a comparable active-site area of TTHA0746 to *Sd*ShuT, but a twofold difference in the volume of the protein TTHA0746 and *Ec*BtuF.

3.3.6. TTHB220 and TTHA0723 might potentially function as heme-binding proteins

To reaffirm the binding of heme and cyanocobalamin molecules to the proteins TTHB220 and TTHA0723, respectively, the amino acid residues required for substrate coordination were identified by performing multiple sequence alignment (MSA). Earlier studies have shown that heme recognition and binding by the SBPs involved different modes of coordination which varies from protein to protein (Ho et al., 2007; Mattle et al., 2010; Naoe et al., 2017). The analysis reveals that the protein TTHB220 possesses the residues Tyr60, Arg63, Tyr161 and Arg163 known to coordinate the heme molecule in the active site. Interestingly, the protein TTHA0723 comprises residues Pro7, Phe28, Tyr44, Lys46, Tyr170, Asp220 and His224 required to accommodate both the heme and cyanocobalamin molecules in its active site (Figure 3.5A). This characteristic of the protein TTHA0723 raises the speculation of its binding to both heme and cyanocobalamin molecules. To clarify this, the position and orientation of the active-site residues conserved in TTHB220 and TTHA0723 were compared with those of known heme- and cyanocobalamin-binding proteins. The analysis reveals that although heme-binding residues in both the proteins TTHB220 and TTHA0723 occupy a similar conformation to that of *Rs*Rhut and *Bc*Bhut (Figure 3.5B and 3.5C), the cyanocobalamin-binding residues, except Pro7 and Asp220, are flipped outside compared to that of a cyanocobalamin-binding protein from *E. coli* K-12 (*Ec*BtuF, PDB ID: 1N2Z) (Figure 3.5D). As the bacterium *T. thermophilus* inhabits the iron-deficient hot springs (Fujino et al., 2016), the requirement of multi-substrate-binding proteins such as TTHA0723 might be essential.

Furthermore, to affirm the cognate substrates for TTHB220 and TTHA0723, molecular docking experiments with heme and cyanocobalamin molecules were performed. The results reveal that a heme molecule can be accommodated in the active site of the protein TTHB220 with an estimated binding free energy of -9.83 kcal mol^{-1} coordinated by the residues Tyr161 and Arg163 from the C-terminal lobe (Figure 3.5E and Table A6). On the other hand, heme molecules are observed to bind to the active site of the protein TTHA0723 in two different conformations with binding energies of -8.19 and -8.02 kcal mol^{-1}, respectively (Table A6). This corroborates the results obtained through active-site analysis, which suggests the binding of two heme molecules to the protein TTHA0723. The two docked heme molecules in the active site of TTHA0723 are

coordinated by residues Tyr44, Thr45, Lys191 and Lys195 (Figure 3.5F and Table A6). However, cyanocobalamin could not be docked to the protein TTHA0723 due to its near-to-close conformation. In summary, both the proteins TTHB220 and TTHA0723 seem to bind heme molecules. The overall analysis suggests that the protein TTHA0723 might bind cyanocobalamin, in addition to heme.

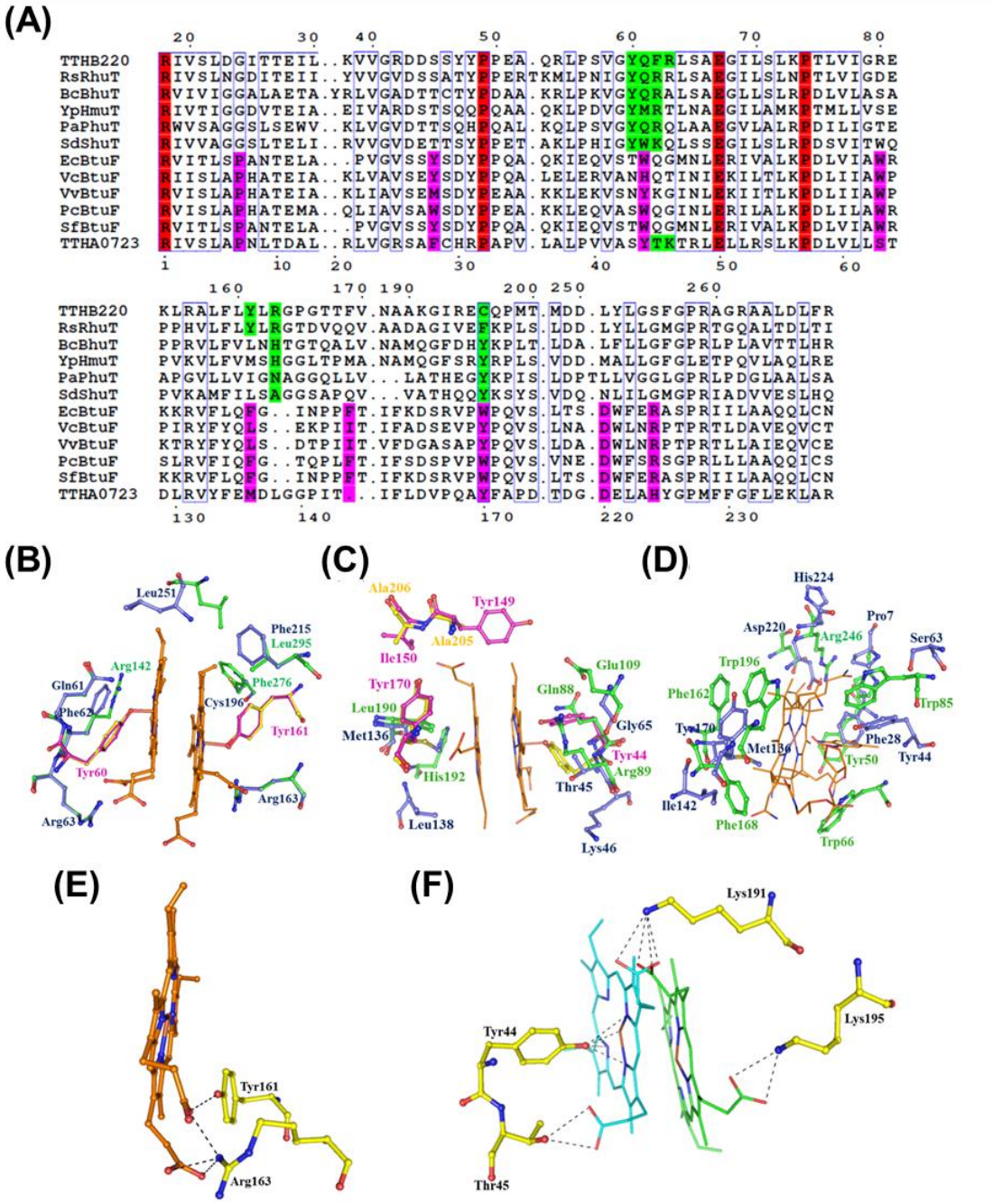


Figure 3.5. (A) Multiple sequence alignment (MSA) of the proteins TTHB220 and TTHA0723 along with other known heme- and cyanocobalamin-binding proteins. The protein sequences used to build the MSA are from *Roseiflexus spp.* (*Rs*RhuT, A5UZ69), *B. cenocepacia* (*Bc*BhuT, B4EKB3), *Y. pestis* (*Yp*HmuT, Q56991), *P. aeruginosa* (*Pa*PhuT, O68879), *S. dysentriae* (*Sd*ShuT, O70018), *E. coli* (*Ec*BtuF, P37028), *V. cholera* (*Vc*BtuF, Q9KPI6), *V. vulnificus* (*Vv*BtuF, Q7MNT2), *P. carotovorum* (*Pc*BtuF, C6DC28) and *S. flexneri* (*Sf*BtuF, Q83MD7); the UniProt ID of each protein is mentioned in the parenthesis. For the clarity of the figure, only the partial MSA has been shown. The heme- and cyanocobalamin-coordinating residues are shaded in green and pink, respectively. (B) The active-site pocket comparison of the protein TTHB220 with *Rs*RhuT (PDB ID: 5GJ3). All the active-site residues are represented as a ball-and-stick model. The active-site residues of the proteins *Rs*RhuT and TTHB220 are shown in green and blue, respectively. Similarly, the residues involved in the heme (orange line) binding of the proteins *Rs*RhuT and TTHB220 are shown in yellow and pink, respectively. (C) The active-site pocket comparison of the protein TTHA0723 with *Bc*BhuT (PDB ID: 5Y8A). The active-site residues of the proteins *Bc*BhuT and TTHA0723 are presented in green and blue, respectively. The heme-interacting residues in *Bc*BhuT and TTHA0723 are shown in yellow and pink, respectively. (D) The active-site pocket comparison of the protein TTHA0723 with *Ec*BtuF (PDB ID: 1N2Z). The cyanocobalamin (orange line)-anchoring residues of *Ec*BtuF and TTHA0723 are shown in green and blue, respectively. (E, F) Molecular docking analysis of TTHB220 and TTHA0723 with heme molecule. The active-site residues holding one heme (green) and two heme (green and cyan) molecules in TTHB220 and TTHA0723 proteins, respectively, are represented in yellow.

3.3.7. TTHA0746 is potential cyanocobalamin- rather than an iron-binding protein

Evolutionary and preliminary tertiary structure analyses hint that the protein TTHA0746 is probably cyanocobalamin- rather than heme-binding protein. To identify the determinants distinguishing between heme- and cyanocobalamin-binding proteins, an MSA analysis of TTHA0746 and its homologs were performed. The result reveals the absence of a positively-charged arginine or lysine residue crucial for maintaining the deprotonated state of heme-binding tyrosine residue (Tyr60) in the protein

TTHA0746 (Figure 3.6A). However, residues Pro22, Tyr43, Tyr60, Ile153, Tyr158, Phe186, Asp238 and Arg243 required to coordinate cyanocobalamin molecule in the active site are mostly conserved in TTHA0746 as well (Figure 3.6A). Although these cyanocobalamin-coordinating residues are not universally conserved across organisms, its similarity with that of BtuF proteins hints the protein TTHA0746 to be a probable cyanocobalamin-binding protein. Furthermore, the structural superimposition of TTHA0746 and *Ec*BtuF demonstrates that the active-site residues Pro22, Tyr43, Tyr60, Ser78, Ile153, Tyr158, Asp238 and Arg243 occupy a similar position implying a plausible occupancy of cyanocobalamin in its active-site pocket (Figure 3.6B).

To reaffirm the preference of TTHA0746 towards cyanocobalamin over heme, molecular docking experiments were performed. The molecular docking results show that the docked cyanocobalamin is coordinated by the residues Ser78, Thr99, Glu100, Asp238 and Arg243 with a binding affinity of -4.16 kcal mol^{-1} (Figure 3.6C and Table A6). In addition, molecular docking of TTHA0746 with one and two heme molecules separately was performed, which shows that the heme molecule(s) bind at the C-terminal of the protein non-specifically, though with a higher binding affinity of -8.39 kcal mol^{-1} (Figure 3.6D and Table A6). In summary, the results hint a plausible misannotation of the protein TTHA0746 as an iron-binding protein and is most likely a cyanocobalamin-binding protein.

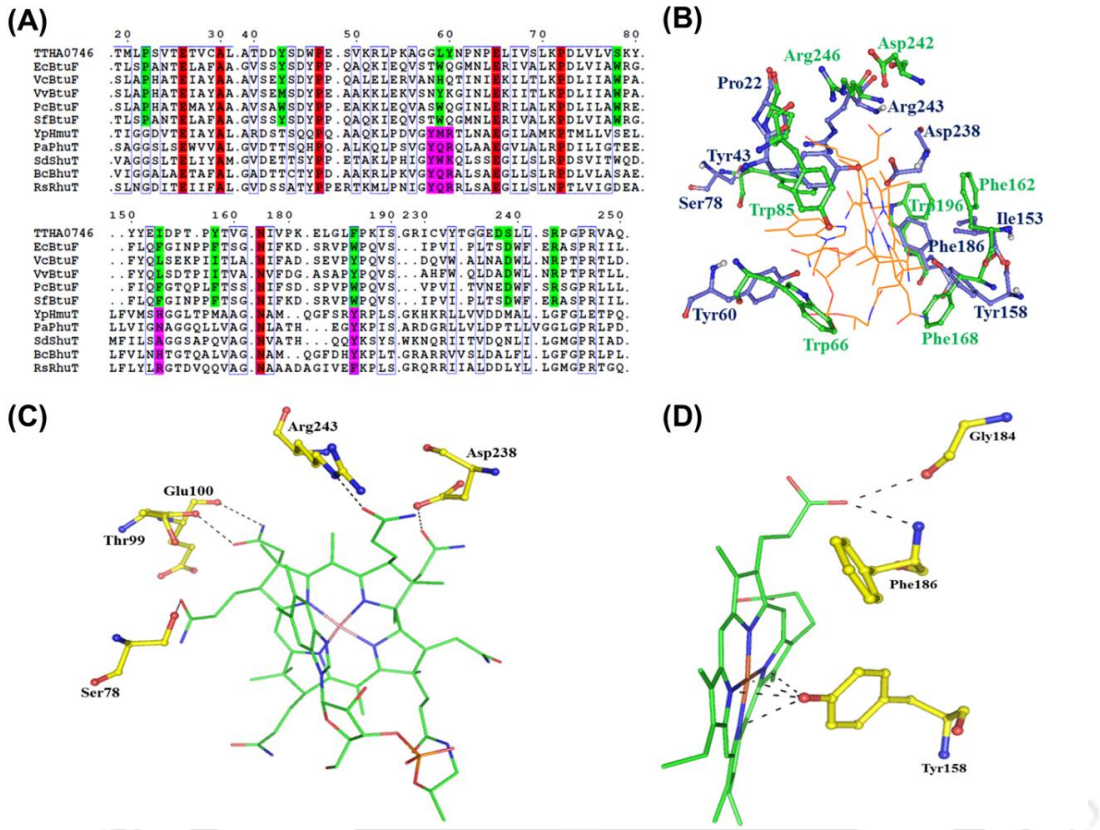


Figure 3.6. (A) Multiple sequence alignment (MSA) of TTHA0746, heme- and cyanocobalamin-binding proteins. The protein sequences used to build the MSA are from *E. coli* (*Ec*BtuF, P37028), *V. cholera* (*Vc*BtuF, Q9KPI6), *V. vulnificus* (*Vv*BtuF, Q7MNT2), *P. carotovorum* (*Pc*BtuF, C6DC28), *S. flexneri* (*Sf*BtuF, Q83MD7), *Y. pestis* (*Yp*HmuT, Q56991), *P. aeruginosa* (*Pa*PhuT, O68879), *S. dysentriae* (*Sd*ShuT, O70018), *B. cenocepacia* (*Bc*BhuT, B4EKB3) and *Roseiflexus spp.* (*Rs*RhuT, A5UZ69); the UniProt ID of each protein is mentioned in the parenthesis. For the clarity of the figure, only the partial MSA has been shown. The amino acid residues involved in the substrate binding in BtuF proteins are shaded in green, whereas those involved in heme binding are in pink. (B) Superimposition of the active-site pockets of the proteins TTHA0746 and *Ec*BtuF (PDB ID: 1N2Z). The residues involved in the binding of cyanocobalamin molecule (orange line) in *Ec*BtuF and TTHA0746 are shown as a ball-and-stick model in green and blue, respectively. (C, D) Molecular docking results of TTHA0746 with cyanocobalamin and heme molecules, respectively,are shown in green. The amino acid residues coordinating the ligand are represented as a ball-and-stick model in yellow.

3.3.8. TTHB177 is potentially Mg/Fe^{2+}-citrate-binding protein

To affirm that TTHB177 is a probable Mg^{2+}-citrate-binding protein, the conservation of the Mg^{2+}-citrate-binding residues was investigated by performing an MSA analysis. The result confirms the Mg^{2+}-citrate-coordinating residues Ser26, Asp28, Thr55, Ser79, Tyr221 and Arg246 to be conserved in TTHB177, however residues Tyr176, Gly182 and Ser216 present in *Af*Atu are substituted by Trp159, Ala165 and Thr199, respectively, in TTHB177 (Figure 3.7A and 3.7B). On the other hand, comparison of TTHB177 with ferric (Fe^{3+})-binding proteins along with TTHA1628 shows that two consecutive tyrosine residues (Tyr219 and Tyr220) and occasionally a third tyrosine residue (Tyr162) in TTHA1628 essential for the binding of Fe^{3+} ion, are absent in TTHB177 negating its possibility to be a ferric-binding protein (Figure 3.7A and 3.7C).

To further validate the binding of Mg^{2+}-citrate to the protein TTHB177, a molecular docking experiment was performed with both Mg^{2+} and Fe^{2+} ions along with the citrate molecule. Interestingly, the results suggest that both Mg^{2+}-citrate and Fe^{2+}-citrate molecules can bind to the protein TTHB177 with a binding affinity of -10.02 and -8.94 kcal mol^{-1}, respectively (Table A6). Both the ions in the docked conformation are penta-coordinated by the residues Ser26, Asp28 and the citrate molecule (Figure 3.7D, 3.7E and Table A6). In conclusion, based on these results, the protein TTHB177 seems to be a potential divalent metal ion-citrate-binding protein.

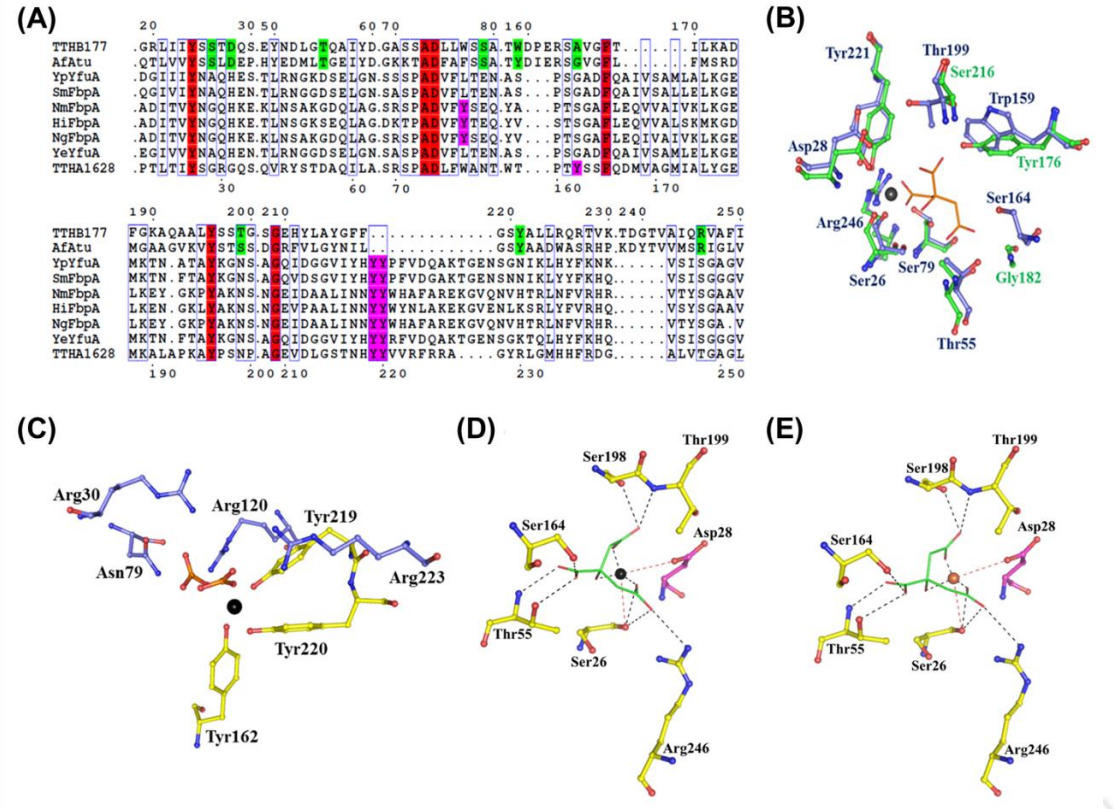


Figure 3.7. (A) Multiple sequence alignment (MSA) of the proteins TTHA1628, TTHB177 and their homologs. The protein sequences used for the MSA are from *A. fabrum* (*Af*Atu, Q7CWZ6), *Y. pestis* (*Yp*YfuA, A0A0E1NNQ6), *S. marcescens* (*Sm*FbpA, P21408), *N. meningitidis* (*Nm*FbpA, P0A0Y4), *H. influenzae* (*Hi*FbpA, P35755), *N. gonorrhoeae* (*Ng*FbpA, P17259) and *Y. enterolitica* (*Ye*YfuA, A1JLH5); the UniProt ID of each protein is mentioned in the parenthesis. For the clarity of the figure, only the partial MSA has been shown. The Mg^{2+}-citrate-coordinating residues are shaded in green. Three tyrosine residues (at positions 162, 219 and 220) required for the binding of Fe^{3+} in TTHA1628 are shaded in pink. (B, C) The active-site analysis of TTHB177 superimposed with *Af*Atu (PDB ID: 3C9H) and TTHA1628 (PDB ID: 3WAE), respectively. The residues holding the Mg^{2+}-citrate molecule (black sphere and orange line) in *Af*Atu are shown as a ball-and-stick model in green (blue for TTHB177). The residues binding the two bicarbonates (orange) are represented in blue, whereas the tyrosine residues important for coordinating the Fe^{3+} ion (black sphere) are shown in yellow. (D, E) Molecular docking of the protein TTHB177 with Mg^{2+}-citrate and Fe^{2+}-citrate complexes. The residues coordinating the citrate (green) molecule are shown in yellow. The Mg^{2+} and Fe^{2+} ions are shown as black and brown spheres, respectively. The residue Asp28 interacting with the metal ion is shown in pink.

3.3.9. TTHA0069 and TTHA0717 are potential WO_4^{2-}- and MoO_4^{2-}-binding proteins, respectively

Initial analyses of the primary and tertiary structure reveal that the proteins TTHA0069 and TTHA0717 are probable WO_4^{2-}- and MoO_4^{2-}-binding proteins. In WO_4^{2-}-binding proteins, a motif termed as 'TTTS' motif along with residues His59, Arg118, Thr124 and Asp170 (numbering according to TupA protein from *Desulfovibrio alaskensis*, *Dal*TupA) are known to anchor the WO_4^{2-} ion in the active site (Otrelo-Cardoso et al., 2017). Thus, the conservation of these residues required for the binding of WO_4^{2-} and MoO_4^{2-} oxyanions in the proteins TTHA0069 and TTHA0717 were investigated using an MSA analysis. The result demonstrates the conservation of the 'TTTS' motif and its supporting residues in the protein TTHA0069 as well (Figure 3.8A). Further, structural comparison of TTHA0069 with TupA protein from *Geobacter sulfurreducens* (*Gs*TupA, PDB ID: 3LR1, WO_4^{2-}-bound and closed conformation) and *Dal*TupA (PDB ID: 5MY5, open conformation) exhibits that the motif 'TTTS' and the residues His55, Arg115, Thr121 and Asp165 occupy a similar spatial position. Interestingly, in the open state of the protein (i.e. *Dal*TupA), the conformation of the supporting residues is flipped outside, indicating their closure upon the substrate (WO_4^{2-}) binding (Figure 3.8B). On the other hand, in MoO_4^{2-}-binding proteins, the substrate-anchoring residues are generally variable in nature due to their tendency to bind both WO_4^{2-} and MoO_4^{2-}. For instance, the residue Tyr117 and Tyr118 in TTHA0717 and ModA from *Azotobacter vinelandii* (*Av*ModA, PDB ID: 1ATG), respectively, is replaced by Ala125 in ModA from *E. coli* K-12 (*Ec*ModA, PDB ID: 1AMF) (Hu et al., 1997; Lawson et al., 1998). In *Ec*ModA, another residue Tyr170 from the C-lobe, which interacts with the MoO_4^{2-} ion (Hu et al., 1997), is partially conserved in the protein TTHA0717. Also, residues Thr9 and Asn10 in *Av*ModA are substituted by Ser7 and Asp8, respectively, in TTHA0717. Similarly, residues Val147 and Val152 in *Av*ModA and *Ec*ModA, respectively, are replaced by Ile186 in TTHA0717 (Figure 3.8A). Moreover, the active-site comparison of TTHA0717 with *Av*ModA (PDB ID: 1ATG, WO_4^{2-}-bound) and *Ec*ModA (PDB ID: 1AMF, MoO_4^{2-}-bound) displays the residues Ser7, Asp8, Ser36 and Tyr117 of TTHA0717 are in close proximity to the oxyanions (Figure 3.8C). However, the residue Ile186, hypothesized to coordinate the ligand in the active site, is located far away (~12-14 Å) from the active site hinting at its inability to bind the ligand (Figure 3.8C).

To corroborate these observations, molecular docking experiments of both the proteins TTHA0069 and TTHA0717 with WO_4^{2-} and MoO_4^{2-} molecules were performed. As anticipated, the molecule WO_4^{2-} is docked in the active site of TTHA0069 with a binding affinity of -2.92 kcal mol^{-1} and is coordinated by the residues Thr37, His55, Arg115, Gly120 and Thr121 from the 'TTTS' motif present in the active site (Figure 3.8D and Table A6). Similarly, both the molecules WO_4^{2-} and MoO_4^{2-} can dock to the active site of the protein TTHA0717 with a binding affinity of -2.73 and -2.89 kcal mol^{-1}, respectively, coordinated by residues Ser7, Asp8, Ser36 and Ala164 (Figure 3.8E, 3.8F and Table A6). Thus, it can be suggested that the protein TTHA0069 might bind WO_4^{2-} while the protein TTHA0717 might interact with both the molecules MoO_4^{2-} and WO_4^{2-}, preferably the former one.

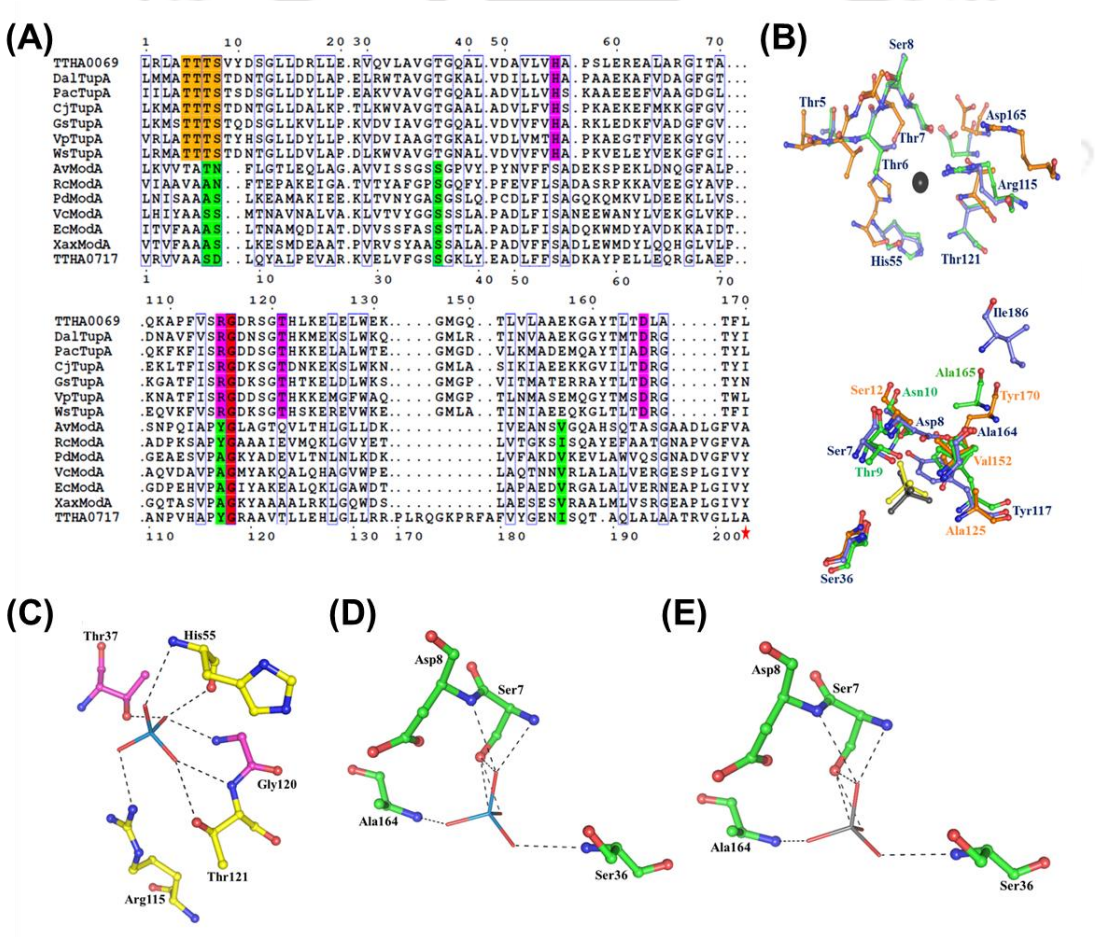


Figure 3.8. (A) Multiple sequence alignment (MSA) of TTHA0069 and TTHA0717, along with their homologs. The protein sequences used to build the MSA are from *D. alaskensis* (*Dal*TupA, Q316W1), *P. acidaminophilum* (*Pac*TupA, Q93KD6), *C. jejuni*

(*Cj*TupA, Q0P885), *G. sulfurreducens* (*Gs*TupA, Q749P2), *V. parahaemolyticus* (*Vp*TupA, Q87PK2), *W. succinogenes* (*Ws*TupA, Q7M8V9), *A. vinelandii* (*Av*ModA, Q7SIH2), *R. capsulatus* (*Rc*ModA, Q08383), *P. difficile* (*Pd*ModA, Q18A64), *V. cholerae* (*Vc*ModA, Q9KLL7), *E. coli* (*Ec*ModA, P37329) and *X. axonopodis* (*Xax*ModA, Q8PHA1); the UniProt ID of each protein is mentioned in the parenthesis. For the clarity of the figure, only the partial MSA has been shown. The 'TTTS' motif in TupA proteins are shaded in orange and the amino acid residues involved in binding of the WO_4^{2-} and MoO_4^{2-} molecules are shaded in pink and green, respectively. The residue Tyr170 in *Ec*ModA, essential for holding the MoO_4^{2-} molecule, is marked with a red star. (B) The active-site pocket analysis of TTHA0069 superimposed with *Gs*TupA (PDB ID: 3LR1) and *Dal*TupA (PDB ID: 5MY5). The amino acid residues involved in the ligand (black sphere) binding are shown as a ball-and-stick model in blue (TTHA0069), green (*Gs*TupA) and orange (*Dal*TupA), respectively. (C) The active-site pocket analysis of TTHA0717 superimposed with *Av*ModA (PDB ID: 1ATG) and *Ec*ModA (PDB ID: 1AMF) proteins. The amino acid residues involved in the binding of MoO_4^{2-} are shown as a ball-and-stick model in blue (TTHA0717), green (*Av*ModA) and orange (*Ec*ModA), respectively. (D-F) Molecular docking analysis of TTHA0069 with WO_4^{2-} (cyan lines) and TTHA0717 with WO_4^{2-} (cyan lines) and MoO_4^{2-} (grey lines). The ligand-interacting residues are shown in yellow and green in TTHA0069 and TTHA0717, respectively. The residues shown in pink are also involved in holding the WO_4^{2-} ion in TupA proteins.

3.3.10. The protein TTHA0596 is a potential Zn^{2+}-binding protein

To reaffirm the active site similarity of TTHA0596 protein with Zn^{2+} and Mn^{2+} binding SBPs, a multiple sequence alignment (MSA) of the homologous proteins from cluster A-I family along with TTHA0596 protein was performed to identify the amino acid residues involved in metal binding. Interestingly, both Zn^{2+} and Mn^{2+} contradict each other in their preferences for coordination chemistry. Theoretically, Zn^{2+} prefers tetrahedral (n = 4) coordination geometry, whereas, Mn^{2+} can be accommodated in both tetrahedral and octahedral (n = 6) coordination geometry (Harding, 2006). Earlier studies have shown that Zn^{2+} binding SBPs requires a glutamate or aspartate residue along with three histidine residues for metal binding (Zn^{2+}), whereas, Mn^{2+} binding SBPs requires only two histidines, a glutamate and an aspartate residue in the binding

cavity, to interact with the Mn^{2+} ion (Li and Jogl, 2007; Counago et al., 2014). However, a typical variability has been reported to exist at the active site pocket of Zn^{2+}-binding SBPs from different organisms. For example, the Zn^{2+}-binding pocket of ZnuA from *E. coli* (*Ec*ZnuA) is composed of Glu59, His60, His143 and His207 (Li and Jogl, 2007). A similar binding pocket is required for ZnuA from *S. Enterica* (*Sae*ZnuA) (Ilari et al., 2011). However, Zn^{2+}-binding TroA protein form *T. pallidum* (*Tp*TroA) utilizes an aspartate (Asp279) and three histidines (His68, His133 and His199) residues to anchor Zn^{2+} ion (Lee et al., 1999; Lee et al., 2002). Additionally, TroA from *S. Suis* (*Ss*TroA) requires an Asp289 along with three histidine residues (His76, His139 and His205) to interact with Zn^{2+} ion (Zheng etal., 2011). Replacement of the glutamate residue by an aspartate residue renders a Zn^{2+}-binding SBP with a certain affinity towards Mn^{2+} (Desrosiers et al., 2007). Our MSA result shows that TTHA0596 protein also possesses three histidine residues (His52, His123 and His185) which are found to be conserved along with other Zn^{2+} binding SBPs (Figure 3.9A). TTHA0596 also lacks the preceding glutamate residue (Glu59and Glu78) present in *Ec*ZnuA and *Sae*ZnuA, respectively, and owns an Asp259 similar to two of the Zn^{2+} binding SBPs *Tp*TroA (Asp279) and *Ss*TroA (Asp289) (Figure 3.9A). In contrast, $Fe^{2+/3+}$ generally forms com-plexes with siderophore molecules and does not interact directly with amino acid ligands (Miethke and Marahiel, 2007); henceforth, no amino acid conservation has been observed in the MSA (Figure 3.9A).To further corroborate the above-inferred conclusion, the orientation of the active-site residues conserved in the protein TTHA0596 was compared with those of Mn^{2+} and Zn^{2+} ion-binding residues. The structural comparison of the protein TTHA0596 with three Mn^{2+}-binding SBPs, i.e. *Sau*MntC, *Stp*SitA and *Sp*PsaA reveal that Mn^{2+} ion is coordinated by two histidines, a glutamate and an aspartate residue (Figure 3.9B). Though the position and orientation of two histidine residues and the aspartate residue were found to be conserved in TTHA0596 protein model, Glu205 of *Sp*PsaA was replaced by His185 in TTHA0596 protein, which further suggests the preference for another metal ion instead of Mn^{2+} ion (Figure 3.9B). In contrast, structural superimposition of TTHA0596 protein with three Zn^{2+}-binding SBPs such as *Tp*TroA, *Ss*TroA and *Syn*ZnuA exhibits that TTHA0596 occupies a similar orientation for three histidine residues in the active site (Figure 3.9C). Surprisingly, *Tp*TroA and TTHA0596 possess an aspartate (Asp279 and Asp259, respectively) residue in the active site similar to Mn^{2+}-binding SBPs. It has earlier been reported that, though the natural ligand for *Tp*TroA is Zn^{2+} ion, it can also interact with

Mn^{2+} ion (Desrosiers et al., 2007). Thus, it can be postulated that the protein TTHA0596 possesses a similar characteristic and may interact with Mn^{2+} ion in addition to Zn^{2+} ion.

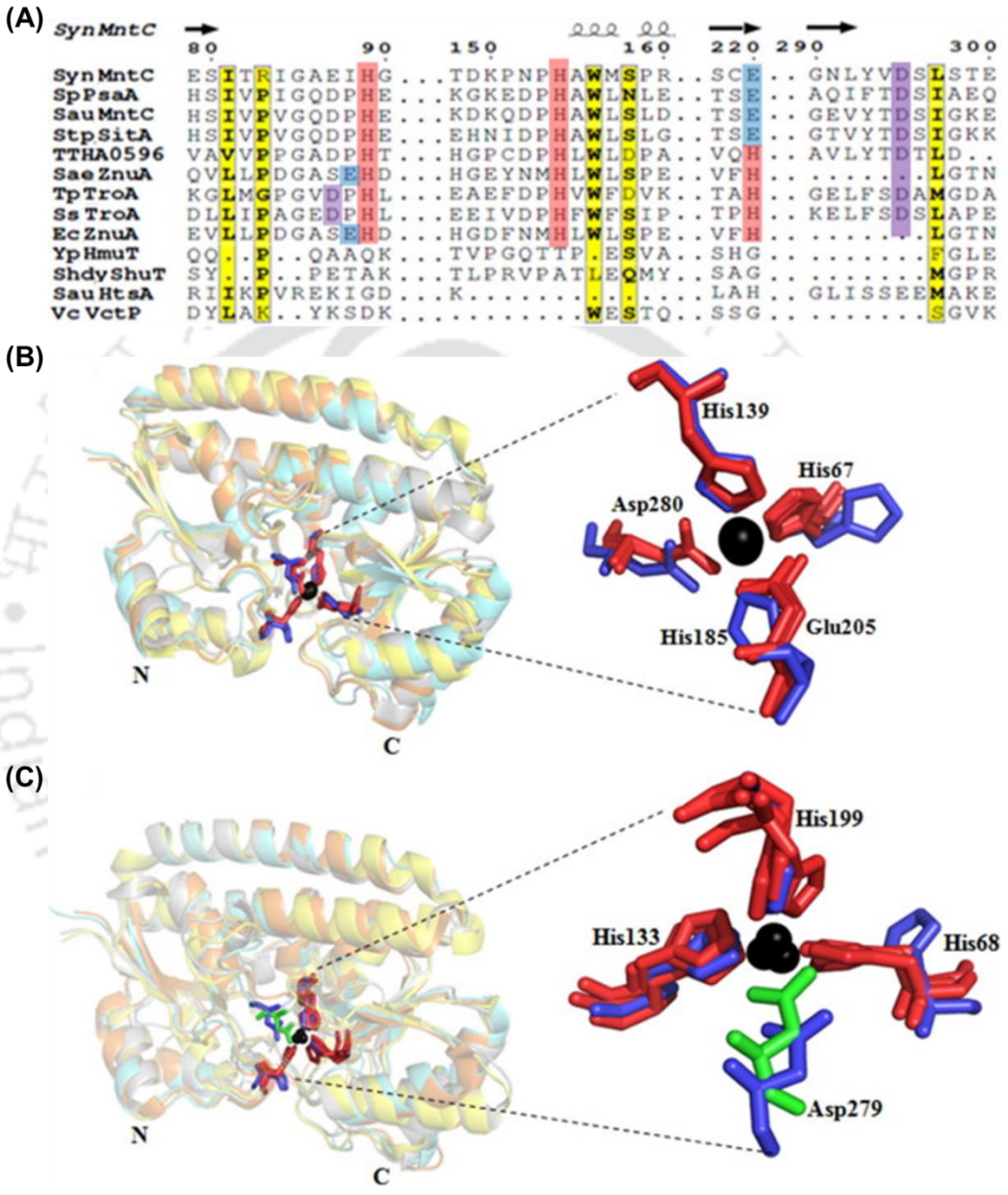


Figure 3.9. (A) Multiple sequence alignment of Zn^{2+}, Mn^{2+} and $Fe^{2+/3+}$-binding SBPs along with the protein TTHA0596. The protein sequences chosen for analysis include Mn^{2+}-binding SBPs *Syn*MntC, *Sp*PsaA, *Sau*MntC and *Stp*SitA, Zn^{2+}-binding SBPs *Sae*ZnuA, *Tp*TroA, *Ss*TroA and *Ec*ZnuA and $Fe^{2+/3+}$-binding SBPs *Yp*HmuT, *Shdy*ShuT, *Sau*HtsA and *Vc*VctP. For the clarity of the figure, only the partial MSA has

been shown. Conservation of three histidine residues in Zn^{2+}-binding SBPs is shaded in red. The substitution of the third histidine by glutamate residue in Mn^{2+}-binding SBPs and the glutamate residues preceding the first histidine in *Sae*ZnuA and *Ec*ZnuA are in blue. In addition, an aspartate residue conserved in both Zn^{2+} and Mn^{2+}-binding SBPs is shaded in purple. This analysis shows that the active site of the protein TTHA0596 is more similar to Zn^{2+}-binding SBPs. (B) Structural superimposition of the protein TTHA0596 (modeled, cyan) with *Sau*MntC (PDB id: 4K3V in orange), *Stp*SitA (PDB id: 4OXR, yellow) and *Sp*PsaA (PDB id: 3ZTT, grey). The active-site residues of TTHA0596 protein model are in blue and those of *Sau*MntC, *Stp*SitA and *Sp*PsaA are represented in the red stick model. The zoomed view of the active-site pocket of *Sau*MntC, *Stp*SitA and *Sp*PsaA (red) is compared with model protein TTHA0596 (blue). The black sphere represents the Mn^{2+} ion. (C) Structural superimposition of the protein TTHA0596 (modeled, cyan), *Tp*TroA (PDB id: 1TOA, grey), *Ss*TroA (PDB id: 3MFQ, yellow) and *Syn*ZnuA (PDB id: 1PQ4 in grey). The active-site residues of TTHA0596 protein model are in blue and those of *Tp*TroA, *Ss*TroA and *Syn*ZnuA are in represented in the red stick model. The zoomed view of the active-site pocket of *Tp*TroA, *Ss*TroA and *Syn*ZnuA is compared with the model TTHA0596 protein. The black sphere represents the Zn^{2+} ion. The N- and C-terminal domains are labeled for clarity. The residues His67, His139 and Asp280 of *Sp*PsaA are found conserved in TTHA0596 also, however, Glu205 (*Sp*PsaA) of Mn^{2+}-binding SBPs is replaced by His185 in the protein TTHA0596. In addition, the three histidine residues (His68, His133 and His199 in *Tp*TroA) of Zn^{2+} binding SBPs are found conserved in TTHA0596 also. In addition, Asp279 (green) of *Tp*TroA is also conserved in the protein TTHA0596 (Asp259 shown in blue).

3.4. DISCUSSION

Thermophilic microorganisms reside in diverse habitats such as volcanic areas, geothermal springs, hot water regions, geothermal power plants and marine hydrothermal vents (Ranawat and Rawat, 2018). Such harsh environments rich in metal and metalloid ions necessitate the availability of specific metal uptake ABC transport systems to maintain intracellular metal ion homeostasis. However, there is a lack of information regarding the repertory of metal uptake ABC transport systems in

thermophilic organisms. In this study, a total of 22 ORFs encoding eight different metal uptake ABC transport systems were identified in the bacterium *T. thermophilus* HB8. However, out of these eight metal uptake ABC transport systems, only one has been characterized, leaving the remaining six transport systems as uncharacterized. Thus, this study involved an integrated *in silico* approach for the functional annotation of metal-binding periplasmic proteins of ABC transporters in *T. thermophilus* HB8.

Being an essential component of ABC transport system, SBPs provide ligand specificity and thus designates function to the entire transport system. Although one of the SBP viz. TTHA1628 in *T. thermophilus* HB8 has been assigned with their function as Fe^{3+}-binding protein, the cognate ligands for the remaining seven SBPs are unassigned. From this study, a total of three iron-binding proteins (TTHA1628, TTHB220 and TTHA0723) and one protein each for cyanocobalamin (TTHA0746), Mg^{2+}-citrate complex (TTHB177), WO_4^{2-} (TTHA0069), MoO_4^{2-} (TTHA0717) and Zn^{2+} (TTHA0596) binding have been identified in *T. thermophilus* HB8. This repertoire of the metal uptake ABC transport system in *T. thermophilus* HB8 is comparable to that of a mesophile (*E. coli* K-12), albeit with certain noteworthy differences in the number of transporter for a particular metal ion. Owing to the difference in their habitats and different response patterns, the availability of the metal uptake ABC transporters widely varies in diverse cell types (Hwang et al., 2016).

In *T. thermophilus* HB8, a high number of ABC transport systems for iron, both in ionic (Fe^{2+}) and complex (heme) forms, uptake could be perceived. In *E. coli* K-12, exogenous heme is not utilized and therefore, no heme uptake or receptors have been identified (Letoffe et al., 2006). However, the requirement for molecular iron is achieved via Fe^{3+}-citrate, Fe^{3+}-enterobactin and Fe^{3+}-hydroxamate uptake ABC transporters and other iron receptors (Maussatova et al., 2008; Letoffe et al., 2006). In contrast to *E. coli* K-12, thermophilic aerobes require heme for the formation of heme-copper oxidases, a terminal enzyme required for catalyzing the reduction of molecular oxygen into water in the electron transport chain (ETC) (Giuffre et al., 1999; Werner et al., 2010). Being an aerobic thermophile, *T. thermophilus* HB8 might fulfill the heme requirement for the synthesis of heme-copper oxidases by employing two heme uptake ABC transport systems (ORFs: TTHB218-TTHB220 and TTHA0723-TTHA0747-TTHA0748). Additionally, another ABC transporter (ORFs: TTHA1628-TTHA1630),

which transports Fe^{3+}-carbonate, maintains the threshold molecular iron concentration in *T. thermophilus* HB8 (Wang et al., 2014).

In contrast to heme, the cellular demand for cyanocobalamin is achieved by the *de novo* synthesis of cyanocobalamin as well as cyanocobalamin uptake ABC transport system (Henne et al., 2004; Romine et al., 2017). Cyanocobalamin is an essential cofactor for many key enzymes catalyzing a diverse range of transmethylation as well as other rearrangement reactions and therefore is required in supplemental amounts (Martens et al., 2002). Similar to *E. coli* K-12, where the cyanocobalamin requirement is achieved by the salvage pathway and BtuCDF ABC transporter, *T. thermophilus* HB8 also utilizes its *de novo* synthesis pathway and cyanocobalamin uptake ABC transporter (ORFs: TTHA0746-TTHA0748) depending on the requirement (Borths et al., 2002; Fang et al., 2017). One of the proteins, TTHA0723, exhibits the ability to bind heme as well as cyanocobalamin. Although no thermophilic protein has been reported to exhibit binding affinity towards two different metal ions, this phenomenon has been observed in some mesophilic soil bacteria e.g., *Salmonella enterica* and *Streptomyces reticuli* (Lucana et al., 2014; Lucana et al., 2016). It might be speculated that TTHA0723, which initially served as a cyanocobalamin-binding protein, later evolved to function as a heme-binding protein as well. The *de novo* synthesis of cyanocobalamin reduces the requirement of two cyanocobalamin uptake ABC transporters and instead employs it in heme uptake from an iron-deficient environment (Henne et al., 2004) to maintain the threshold concentration of intracellular iron.

Out of eight SBPs, seven are present with their cognate TMD and NBD subunits in *T. thermophilus* HB8. This synteny is, however, not maintained in *E. coli* K-12. The protein TTHA0723, on the other hand, completely lacks its respective NBD and TMD components. It might be possible that the protein TTHA0723 employs the TMD (ORF id: TTHA0747) and NBD (ORF: TTHA0748) of the protein TTHA0746 for translocation of the heme molecule. This notion of subunit sharing has been identified for ECF-type ABC transporters, where the NBD is shared by many TMDs to form a holo-transporter (Eitinger et al., 2011). Additionally, the TMDs and NBDs of metal uptake ABC transporters are inert in nature and thus, the substrate specificity is conferred by their corresponding SBPs (Ter Beek et al., 2014).

Apart from ABC transporters involved in the uptake of iron and cyanocobalamin, *T. thermophilus* HB8 harbors ABC transporters for other transition row metal ions as well. For instance, the protein TTHA0596 is involved in Zn^{2+}/Mn^{2+} binding, whereas TTHB177 is identified as a probable divalent metal ion-citrate-binding protein that may translocate either Fe^{2+} or Mg^{2+} by using citrate as a chelator. Interestingly, Mn^{2+} has an ionic radius of 0.80 Å (in aqueous solution) and lies between Fe^{2+} (0.76 Å) and Mg^{2+} (0.99 Å). The closeness of the ionic radii of Mn^{2+} with Fe^{2+} and its ability to interchange with Mg^{2+} suggests Mn^{2+} as a probable candidate to be transported by the protein TTHB177 by forming an Mn^{2+}-citrate complex (Jakubovics and Jenikson, 2001).

Instead of being transported in their ionic or complex form similar to Zn^{2+}, Mn^{2+} or Fe^{2+} ion, molybdenum (Mo) and tungsten (W) are transported in their oxyanionic form (i.e., MoO_4^{2-} and WO_4^{2-}, respectively). Unlike *E. coli* K-12, which possesses a common transporter (ModABC) for the uptake of Mo and W both, *T. thermophilus* HB8 harbors two exclusive ABC transport systems (ORFs: TTHA0069-TTHB0071 and TTHA0715-TTHA0717) for both the oxyanions. This might be due to the fact that microbes thriving only in extreme environments are obligatory tungsten-dependent to accomplish their biological processes (Kletzin and Adams, 1996; Andreesen and Makdessi, 2008). Thus, this necessitates the requirement of separate ABC transport systems for the uptake of Mo and W to maintain their optimum intracellular level.

It can, thus, be concluded that *T. thermophilus* HB8 possesses eight metal uptake ABC transporters that supports the thermophilic organism to thrive in their extreme natural environment. A similar integrative methodology can also be implicated to functionally annotate the metal ion/complex/oxyanion ABC transporters of pathogens, which lack valid functional annotation.

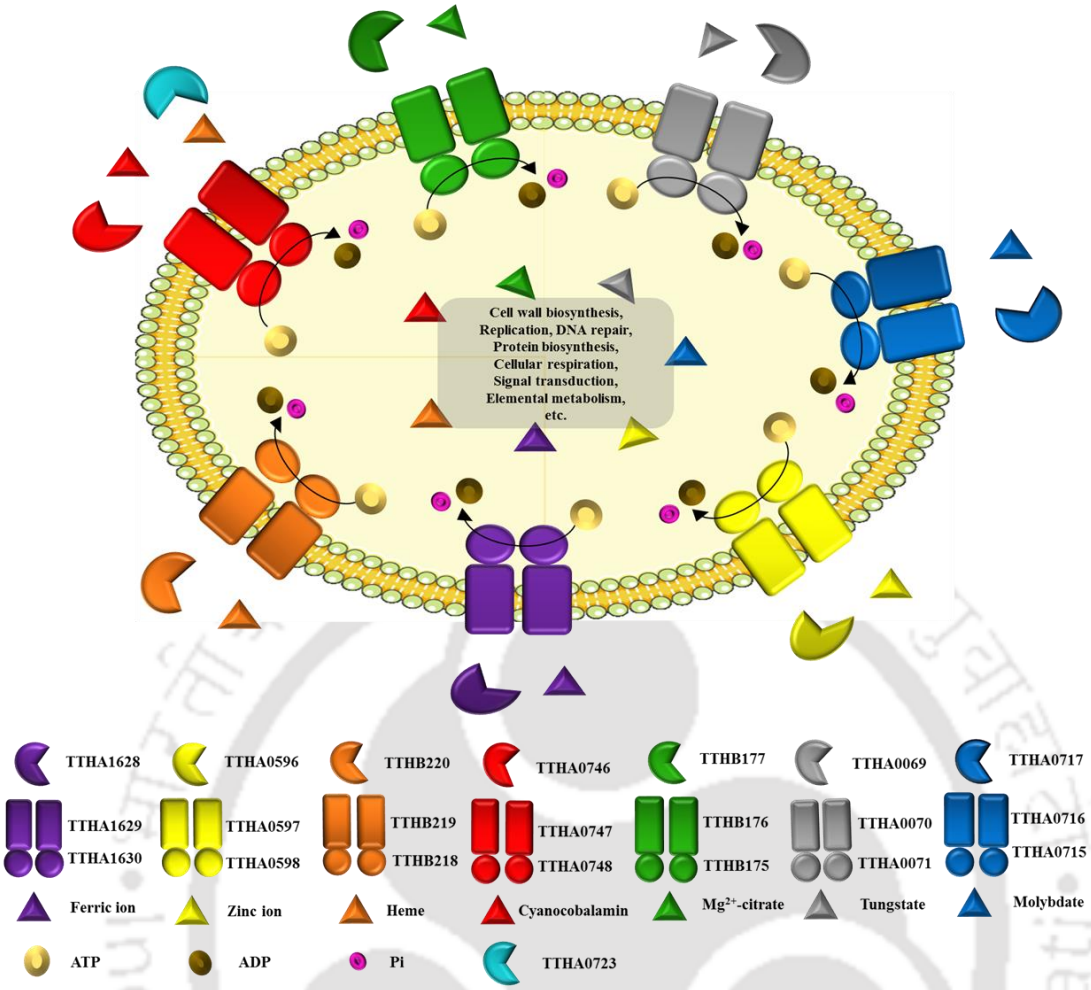


Figure 3.10. Schematic representation of metal ABC transporters available in *T. thermophilus* HB8.

CHAPTER 4

Metal uptake ABC transporters in *M. tuberculosis* H37Rv

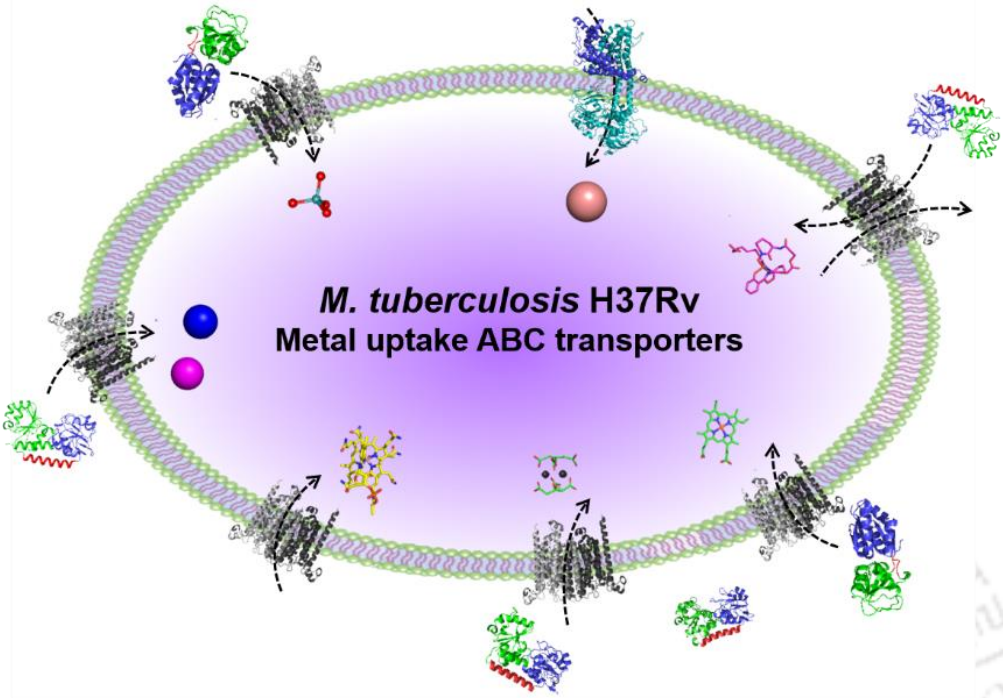
This chapter has been published as:

1. **Mandal SK**, Nayak SG, and Kanaujia SP (2021). Identification and characterization of metal uptake ABC transporters in *Mycobacterium tuberculosis* unveil their ligand specificity. ***International Journal of Biological macromolecules***, 185, 324-337.

ABSTRACT

Mycobacterium tuberculosis, one of the major threats to mankind, requires micronutrients such as metal ions for their survival and pathogenicity inside the host system. Intracellular pathogens such as *M. tuberculosis* have co-evolved to combat the nutritional immunity developed by the host. *M. tuberculosis* has developed eminent mechanisms to sequester essential metal ions from the host system. One such prominent mechanism to scavenge metal ions to thrive in the host cell involves ATP-binding cassette (ABC) transporters, which transport metal ions (in free and/or complex forms) across the cell membrane. To influx, metal ions are selected by substrate-binding proteins (SBPs) and then transferred to the transmembrane domains (TMDs) for further transport into the cytosol. This study employs a high-throughput data mining analysis to identify open reading frames (ORFs) encoding metal uptake ABC transporters in *M. tuberculosis* H37Rv. In total, 19 ORFs resulting in seven ABC transport systems and two P-type ATPases were identified, which are potentially involved in the uptake of different metal ions. This study suggests that three out of the seven metal uptake ABC transporters are specific to iron complexes. Among the remaining four, one is specific to the transition row metal ion Zn^{2+}, one each for MoO_4^{2-} and cobalamin and the remaining one is an energy-coupling factor (ECF)-type ABC transporter importing Co^{2+} ion. The results also suggest the existence of a subunit sharing mechanism in *M. tuberculosis* where the TMDs and NBDs are shared among different ABC transport systems indicating the import of multiple substrates via a single ABC transporter. Thus, this study reflects an overview of the repertoire of metal-specific ABC transport systems in *M. tuberculosis* H37Rv, providing potential therapeutic targets for the future.

Graphical abstract



4.1. INTRODUCTION

Metal ions are one of the most crucial micronutrients required for the survival and growth of all domains of life, including humans (Waldron et al., 2009; Zoroddu et al., 2019). They are utilized by cells in various biological processes such as cell wall biosynthesis, cellular respiration, DNA replication, oxidative stress responses, protein biosynthesis, signal transduction, etc. (Kehres and Maguire, 2003; Hasse and Rink, 2007; Kehl-Fie and Skaar, 2010). Due to the requirement of metal ions by all organisms, there is always an ongoing battle between the host and pathogen(s), known as nutritional immunity (Diaz-Ochoa et al., 2014). For parasitic pathogens, this becomes more challenging as they need to depend on the availability of these metal ions inside the host cell (Hood and Skaar, 2012). Host cells are known to produce various cofactors to chelate available metal ions in the cytosol (Diaz-Ochoa et al., 2014; Hood and Skaar, 2012). Survival in such an environment is a challenging task for parasitic pathogens. However, in due course of time, these pathogens have evolved various proteins, in particular transporters and cofactors which can sequester the required metal ions inside the host cell for their survival and growth (Diaz-Ochoa et al., 2014; Hood and Skaar,

2012). Among the different types of transporters, ATP-binding cassette (ABC) transporters are widely present and utilized by almost all organisms (Diaz-Ochoa et al., 2014; Wilkens, 2015). It is estimated that, on average, ~1-3% of archaeal and bacterial genome codes for ABC transporters (Wilkens, 2015; Tomii and Kanehisa, 1998). These transporters are architecturally and functionally divided into exporters and importers (Wilkens, 2015). ABC exporters are ubiquitously present in all domains of life and export out substrates such as antibiotics, drugs and lipids (Wilkens, 2015). On the contrary, ABC importers are involved in nutrient influx and reported to be present only in prokaryotes and also identified in some plants as well (Wilkens, 2015; Shitan et al., 2003).

In general, ABC transporters are structurally comprised of transmembrane (TMDs) and nucleotide-binding (NBDs) domains (Rees et al., 2009). However, ABC importers require another subunit known as substrate (solute)-binding proteins (SBPs). SBPs bind to their cognate substrates specifically and deliver them to the TMD subunits for their translocation inside the cell. The energy required for the substrate translocation is provided by the NBD subunits where ATP hydrolysis takes place (Lewinson and Livnat-Levanon, 2017). Due to their uniqueness of being present only in prokaryotes, ABC importers are considered to be potential drug targets (Counago et al., 2012). Thus, characterizing the ABC importers, in particular SBPs, becomes essential to understand the mechanism(s) of substrate translocation in order to design drug-like molecules against pathogens as well as to employ SBPs as drug delivery systems.

In this study, we targeted a parasitic pathogen, *Mycobacterium tuberculosis*, the causative agent of tuberculosis, due to which ~10 million cases and ~1.5 million deaths occur per year globally (World Health Organization, 2020). *M. tuberculosis* utilizes various micronutrients such as alcohols, amino acids, carbohydrates, fatty acids, metal ions and vitamins for its growth, survival and pathogenicity (Braibant et al., 2000; Niederweis, 2008). A large number of *Mycobacterium* proteins require d-block transition row metal ions such as manganese (Mn), iron (Fe), cobalt (Co), nickel (Ni), copper (Cu), zinc (Zn), molybdenum (Mo) and tungsten (W) to attain their assigned structural and/or functional role(s) (Niederweis, 2008; Rodriguez and Neyrolles, 2014). Although metal ions are essential for cellular processes, alteration in their concentration is lethal for microorganisms (Klein and Lewinson, 2011). Microorganisms, in particular

M. tuberculosis, govern their metal ion homeostasis via various high-affinity transport systems (Neyrolles et al., 2015). Knock-out studies of ABC importers in *M. tuberculosis* also suggest that the bacterium is unable to grow inside the host system without specific nutrients (Rodriguez and Smith, 2006; Mitra et al., 2019). The vitality of ABC importers in *M. tuberculosis* offers a new paradigm in identifying potential drug targets to develop novel and effective inhibitor molecules. Additionally, the SBPs can also serve as a successful vaccine candidate(s) upon the production of antibodies against the pathogenic *M. tuberculosis*.

The complete genome sequence of *M. tuberculosis* H37Rv is available (Cole et al., 1998; Cabibbe et al., 2018). However, to enhance the understanding of its biological processes, an integrative approach to systematically investigate the available data is required. Thus, in this study, an *in silico* characterization of metal uptake ABC transporters was attempted. Since a large number of ORFs in *M. tuberculosis* H37Rv are annotated as 'unknown or hypothetical proteins', we analyzed the complete genomic ORFs to identify and characterize the probable metal uptake ABC transporters. The results suggest that the bacterium contains various metal-specific ABC transporters for different metal ions either in free or complex forms.

4.2. METHODOLOGY

4.2.1. Data collection and sequence analysis

The complete list of open reading frames (ORFs) encoding metal-specific components of ABC transporters present in *M. tuberculosis* H37Rv were identified and extracted from the TB, Mycobrowser and TubercuList databases (Reddy et al., 2009; Kapopoulou et al., 2011; Lew et al., 2011). Initially, keywords such as 'ABC transporter', 'metal uptake ABC transporter', 'substrate-binding protein' and 'metal-binding protein' were used to search the probable ORFs in the databases. Further, the genome locations were looked for in the 'MAIN SEARCHES' panel of TubercuList database for all listed ORFs in *M. tuberculosis* H37Rv and checked for their annotations in both TubercuList and UniProtKB databases (The Uniprot Consortium, 2019). The ORFs annotated as 'hypothetical protein', 'putative protein' and 'uncharacterized protein' were marked and their amino acid sequences were retrieved from the UniProtKB database to perform a homology search using the web tool BLAST to identify their probable function

(Altschul et al., 1990). A total of 3993 ORFs were checked and verified to identify the probable metal uptake ABC transporters in *M. tuberculosis* H37Rv. The protein sequences of these identified ORFs were retrieved from the UniProtKB database. The functionally associated genes present in an operon of each identified ORF were located from the Gene database of the National Center for Biotechnology Information (NCBI). The evolutionary relationships among different metal-specific components of ABC transporters were studied by constructing phylogenetic trees employing the neighbor-joining (NJ) method using the program MEGA7 (Kumar et al., 2016). The phylogenetic trees were further validated by performing bootstrapping for 1000 replicates in the same program. The protein-protein interaction networks for each protein were generated using the STRING v11.0 database (Szklarczyk et al., 2019). Protein-protein interactome maps were generated with a default set of parameters keeping the top 10 interactors as the results. Subsequently, the interactome networks were further rendered using the tool Cytoscape 3.8.1 for visual clarity (Shannon et al., 2003). To identify the conserved residues and motifs in the protein sequences, multiple sequence alignment (MSA) was performed using the program Clustal Omega (Sievers and Higgins, 2014). The MSAs were further rendered using the web tool ESPript 3.0 (Gouet et al., 2003).

4.2.2. Structure analysis

The three-dimensional structures of all the selected SBPs, NBDs and TMDs, wherever applicable, were modeled using the programs RaptorX, SWISS-MODEL, I-TASSER, and Phyre2 (Kallberg et al., 2012; Biasini et al., 2014; Yang et al., 2015; Kelley et al., 2015). Subsequently, the programs UCSF Chimera and GalaxyRefine were used for model refinement and energy minimization (Pettersen et al., 2004; Ko et al., 2012). The stereo-chemical attributes of all the refined three-dimensional models were validated using the web-server PDBsum (Laskowski et al., 2001); models with the fittest geometrical parameters were chosen for the analysis. The structure-based homology search for each model was performed using the web-server Dali (Holm et al., 2010). The root-mean-square deviation (rmsd) among all the selected SBPs belonging to clusters A and D were computed using a homebuilt python script and the program PyMOL (PyMOL Molecular Graphics System, Schrodinger, LLC). The obtained rmsd distance matrix was used to generate a structure-based distance tree to study the structural relationship among all the SBPs using the programs DendroUPGMA and MEGA7 (Kumar et al., 2016; Garcia-Vallve et al., 1999). A comparison of the three-

dimensional structures to examine the spatial orientation of the active-site residues was performed using the web-server 3dSS (Sumathi et al., 2006). The active-site pocket area and volume of all the validated models were calculated using the program CASTp using a default probe radius of 1.4 Å (Binkowski et al., 2003). The electrostatic potential charge distribution of all the validated models was computed using the PDB2PQR module embedded in the APBS tool plugged in the program PyMOL.

4.2.3. Molecular docking

The binding affinities of different metal ions and their complexes to SBPs were estimated by molecular docking analysis performed using a freely available program AutoDock v4.2 (Morris et al., 2009). For that, the three-dimensional atomic coordinates of the metal ions/complexes were retrieved from their protein-bound structures available at the RCSB Protein Data Bank (Berman et al., 2000). The rotameric orientations of active-site residues were adjusted using the program Coot in order to achieve the closed-like confirmation of the SBPs to accommodate the ligands (Emsley et al., 2010). Before each docking experiment, wherever applicable, the hydrogen atoms were added to the protein and ligand, using the module available in the program AutoDock. Subsequently, the partial charges were assigned to the protein atoms using the Gasteiger charge algorithm (Gasteiger and Marsili, 1980). Similarly, a default set of partial charges for the metal ions, cobalt (Co), iron (Fe), molybdenum (Mo), sulfur (S) and tungsten (W) available in the AutoDock library were assigned. To perform the blind docking of the ligands to the protein, a grid size of 126x126x126 with a grid spacing of 0.375 Å was generated by taking the center of mass of the targeted protein as the grid center. The rigid molecular docking was performed by keeping the protein atoms as rigid and the ligand atoms as flexible about the rotatable bonds. For each molecular docking experiment, a total of 2000 runs of the Lamarckian genetic algorithm (LGA) was set up. The docked ligand conformations were clustered with an rmsd cut-off of 2.0 Å. The atomic interactions between the protein and the docked ligand(s) were identified using the program Coot. The molecular interactions of docked ligands and other structural figures were prepared using the program PyMOL.

4.3. RESULTS

4.3.1. Repertoire of the metal uptake ABC transporters and their genetic organization in *M. tuberculosis* H37Rv and *E. coli* K-12

A total of 19 ORFs encoding SBPs, TMDs and NBDs constituting seven probable metal-specific ABC transporters could be identified from the TB, MycoBrowser and TubercuList databases (Table B1). As the SBPs are known to selectively and specifically bind their cognate ligands, they were utilized to functionally characterize their cognate ABC transport system(s). Among the seven probable metal-specific ABC transporters, two (ORF ids: Rv1348-Rv1349-Rv2895c and Rv1819c) have already been functionally characterized as 'mycobacterial Fe^{3+}-siderophore' and 'cobalamin ABC importers', respectively (Rodriguez and Smith, 2006; Arnold et al., 2020; Rempel et al., 2020). Thus, in this study, the remaining five ABC transporters (ORF ids: Rv0265c, Rv1857-Rv1858-Rv1859, Rv2059-Rv2060, Rv2325c-Rv2326c and Rv3044-Rv3041c) were characterized for their cognate substrate(s) and functionally annotated. In the UniProtKB database, the ORF ids Rv0265c, Rv1857, Rv2059 and Rv3044 have been annotated as 'probable periplasmic iron-transport lipoprotein', 'molybdate-binding protein ModA', 'uncharacterized protein' and 'probable Fe(III)-dicitrate-binding periplasmic lipoprotein FecB', respectively. Thus, to verify the presence of two additional iron-binding proteins (Rv0265c and Rv3044), besides IrtAB (Rv1348-Rv1349) Fe-mycobactin ABC transport system, and to annotate the protein Rv2059, their homologous proteins were searched in the databases. The results corroborate the similarity (query coverage: 46–81% and sequence identity: 17-31%) of the proteins Rv0265c and Rv3044 with iron, iron-siderophore and ferric-dicitrate-binding proteins. Similarly, the protein Rv1857 shows similarity (query coverage: 60–99% and sequence identity: 26-32%) with molybdate-binding proteins. Interestingly, the homology search result suggests that the protein Rv2059 has a low sequence similarity (query coverage: 37-52% and sequence identity: 19-27%) with transition metal ion-binding proteins. From the homology search results, it can be anticipated that akin to *E. coli* K-12, *M. tuberculosis* H37Rv also possesses three possible iron uptake ABC transport systems, out of which, the IrtAB (Rv1348-Rv1349-Rv2895c) system has been characterized to transport ferric carboxymycobactin (Figure 4.1A). Likewise, the number of divalent metal ions, molybdenum and cyanocobalamin uptake ABC transport systems are also similar in both organisms (Figure 4.1B-D). However, the differences are observed in

the genetic arrangement of these genes. In *E. coli*, all the seven metal uptake ABC transport systems, except for cobalamin, are located in tandem along with their SBPs, TMDs and NBDs subunits (Figure 4.1A-D). In contrast, in *M. tuberculosis*, all the probable metal uptake ABC transport systems, except for molybdenum, do not follow a tandem domain organization pattern (Figure 4.1A-D).

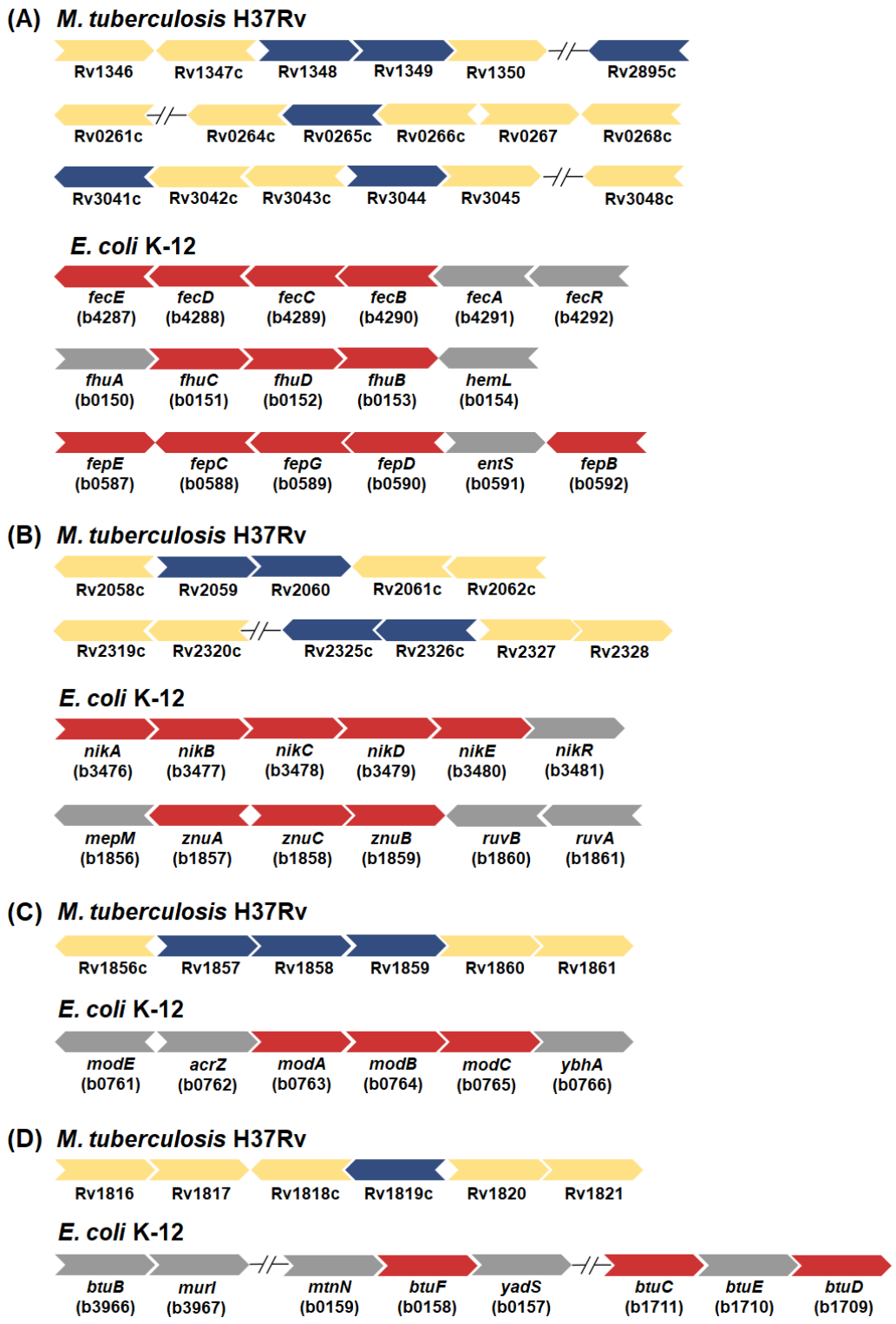


Figure 4.1. Metal-specific ABC transport systems of *M. tuberculosis* H37Rv and *E. coli* K-12. A comparison of the genetic organization of probable (A) iron, (B) transition row divalent metal ion, (C) molybdenum and (D) cyanocobalamin uptake ABC transporters in *M. tuberculosis* H37Rv and *E. coli* K-12. The ORF ids of the SBP-TMD-

NBD subunits of ABC transport systems and other functionally associated genes in *E. coli* K-12 are given in parenthesis. The genes encoding the SBP-TMD-NBD subunits of the ABC transport system in *M. tuberculosis* and *E. coli* are shown in blue and red, whereas other functionally associated genes are in wheat and grey arrows, respectively.

4.3.2. The probable metal-specific SBPs and their cognate NBDs cluster into functionally distinct clades

In *M. tuberculosis*, since the operonic arrangement of the subunits (SBPs, TMDs and NBDs) of the metal-specific ABC transport systems are not in tandem, their cognate subunits were identified by assessing their evolutionary relationships. To find the evolutionary relationships of the targeted SBPs along with their homologous proteins, a sequence-based phylogenetic tree was generated. The result suggests that the proteins Rv0265c and Rv3044, as expected, are clustered along with free or complex iron-binding proteins (Figure 4.2A). Similarly, the protein Rv1857 is cladded with molybdate-binding proteins (Figure 4.2A). Notably, the protein Rv2059, which showed homology with divalent metal ion-binding proteins, lies closer to iron-binding proteins rather than other metal ions (e.g., zinc, magnesium, etc.)-binding proteins (Figure 4.2A).

A similar analysis for ATP-binding domains (NBDs) of these proteins was also performed. Initial data mining suggests the presence of five probable nucleotide-binding domains (NBDs, ORF ids: Rv0908, Rv1030, Rv1859, Rv2326c and Rv3041c) associated with metal uptake ABC transporters in *M. tuberculosis* H37Rv. A homology search of the proteins Rv0908 and Rv1030 reveals sodium, potassium, calcium and copper transporting ATPases as their homologous proteins (query coverage: >70% and sequence identity: >30%). The protein Rv1859 shows homology (query coverage: >60% and sequence identity: >35%) with the oxyanion import ATP-binding proteins, whereas the protein Rv2326c shows homology (query coverage: 27-87% and sequence identity: 24-42%) with ECF-type ABC transporters ATP-binding protein EcfA1/2. Similarly, the protein Rv3041c shows homology (query coverage: 67-88% and sequence identity: 25-40%) with different iron-complex transport system ATP-binding proteins. Furthermore, a sequence-based phylogenetic tree was performed to establish an evolutionary relationship of these homologous proteins. The results reveal that the

proteins Rv3041c and Rv1859 are clustered with the NBDs of iron-complex and oxyanion-binding proteins, respectively (Figure 4.2B). Similarly, the protein Rv2326c, which belongs to the ECF-type of ABC importers, is found to be closer to the EcfA2 proteins (Figure 4.2B). Another two proteins, Rv0908 and Rv1030, which plausibly belong to the P-type ATPases, are expectedly clustered with P-type ATPases (Figure 4.2B). The analysis was validated by including NBDs of ABC exporters (e.g., MbsA) as well, which forms a completely distinct clade showing ancestral differences among NBDs of ABC importers and exporters (Figure 4.2B).

To further affirm the association of probable metal-uptake SBPs with their cognate NBDs and to identify their respective TMD subunits, a protein-protein interaction network analysis was performed. The analysis demonstrates that the protein Rv3044 (SBP) shows interaction with its cognate NBD subunit (Rv3041c) (Figure 4.2C). Similarly, the protein Rv1857 (SBP) interacts with Rv1858 (TMD) and Rv1859 (NBD) subunits of a probable molybdate ABC transport system (Figure 4.2C). Interestingly, the proteins Rv1857-Rv1858-Rv1859 interact with other molybdopterin transferase enzymes implying their role in molybdenum transport (Table B2). Interestingly, the protein Rv0265c (SBP), which lacks both TMD and NBD subunits in its operon, interacts with Rv0264c (an uncharacterized protein), Rv0266c (OplA, a probable 5-oxoprolinase), Rv0453 (PPE11, a protein playing a role in the persistence of mycobacteria in host cells), Rv3041c (NBD of Rv3044) and Rv1857 (SBP) (Figure 4.2C and Table B2). The protein Rv2059 (SBP) shows interaction with Rv2060 (a conserved integral membrane protein) along with proteins belonging to the transport family (OppB, MntH and Zur) and ribosomal proteins (L28, L33 and S14) (Figure 4.2D and Table B2). The protein Rv2326c (a probable NBD of an ECF-type ABC importer) shows interaction with the protein Rv2325c (a probable integral membrane protein) and various conserved hypothetical and transcription regulatory proteins (Figure 4.2E and Table B2).

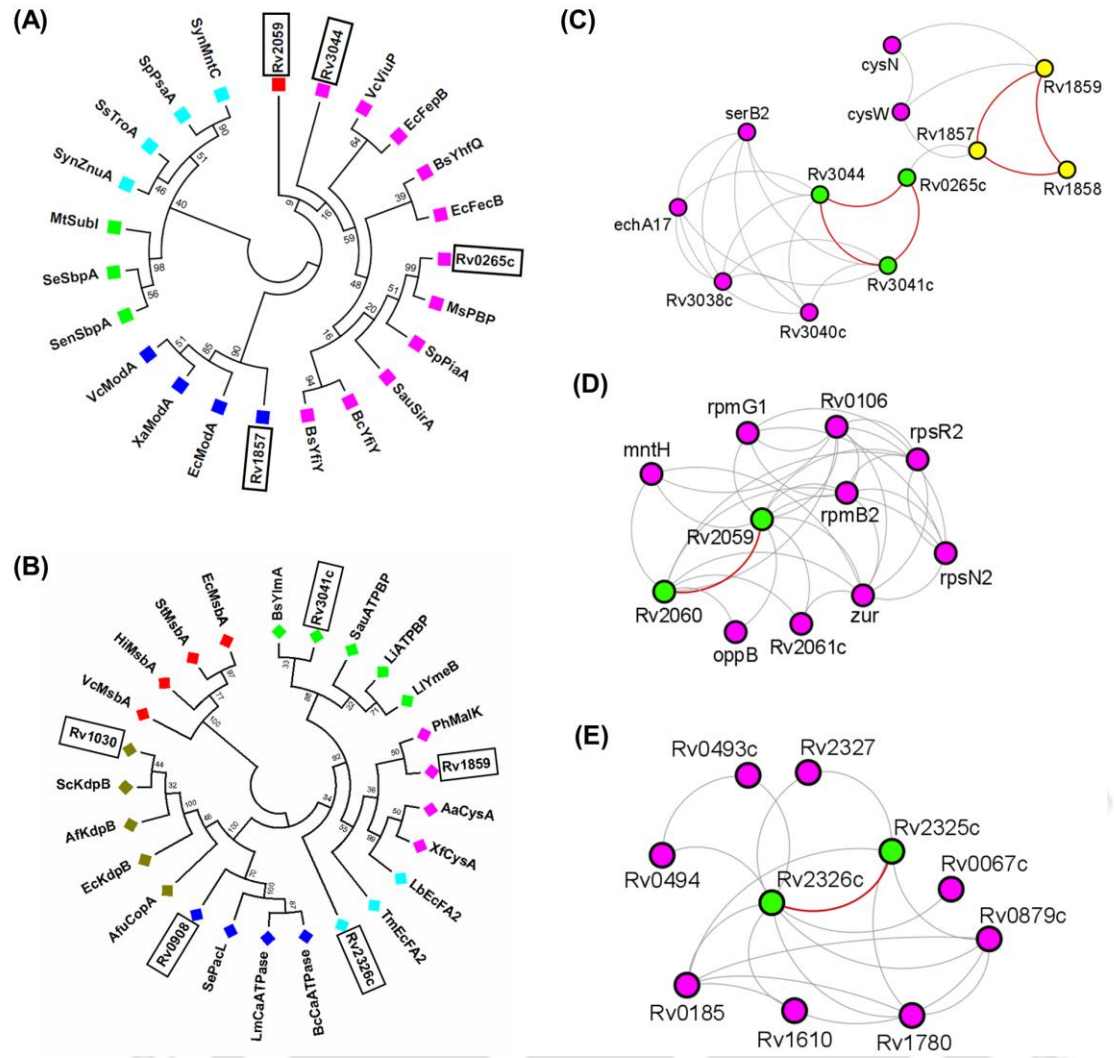


Figure 4.2. Evolutionary relationship and protein-protein interactome analysis for probable metal uptake ABC transporters in *M. tuberculosis* H37Rv. (A) Phylogenetic analysis of probable SBPs from *M. tuberculosis* H37Rv with their homologous proteins. The proteins used to build the phylogenetic tree are Rv2059 (*M. tuberculosis*, O07257), Rv3044 (*M. tuberculosis*, O53291), *Vc*ViuP (*Vibrio cholera*, A0A2R8EE08), *Bs*YhfQ (*Bacillus subtilis*, C0SP94), *Ec*FepB (*E. coli*, P0AEL6), *Ec*FecB (*E. coli*, P15028), Rv0265c (*M. tuberculosis*, L7N6B2), *Ms*PBP (*M. smegmatis*, A0QPL3), *Sp*PiaA (*Streptococcus pneumoniae*, A0A0H2UPT5), *Sau*SirA (*Staphylococcus aureus*, A0A0H3JJC6), *Bc*YfiY (*B. cereus*, UPI00027FCE27), *Bs*YfiY (*Bacillus subtilis*, O31567), Rv1857 (*M. tuberculosis*, P9WGU3), *Ec*ModA (*E. coli*, P37329), *Xa*ModA (*Xanthomonas axonopodis*, Q8PHA1), *Vc*ModA (*Vibrio cholera*, Q9KLL7), *Sen*SbpA (*Salmonella enterica*, P02906), *Se*SbpA (*Synechococcus elongatus*, P27366), *Mt*SubI (*M. tuberculosis*, P71744), *Syn*ZnuA (*Synechocystis spp.*, P73085), *Ss*TroA (*S.*

suis, A4VY63), *Sp*PsaA (*Streptococcus pneumoniae*, P0A4G2) and *Syn*MntC (*Synechocystis spp.*, Q79EF9); the organism name and UniProt id of each protein is provided in the parenthesis. The protein Rv2059, free/complex iron-, molybdate-, sulfate- and Zn^{2+}/Mn^{2+}-binding proteins are represented as red, magenta, blue, green and cyan squares, respectively. (B) Phylogenetic analysis of the probable NBDs and ATPases of *M. tuberculosis* H37Rv with their homologous proteins. The proteins used to build the phylogenetic tree are *Bs*YlmA (*Bacillus subtilis*, O31723), Rv3041c (*M. tuberculosis*, I6YF11), *Sau*ATPBP (*Staphylococcus aureus*, A0A0H2WZ98), *Li*ATPBP (*Leptospira interrogans*, Q72QN4), *Ll*YmeB (*Lactococcus lactis*, Q02151), *Ph*MalK (*Pyrococcus horikoshii*, O57942), Rv1859 (*M. tuberculosis*, P9WQL3), *Aa*CysA (*Alicyclobacillus acidocaldarius*, Q9RHZ7), *Xf*CysA (*Xylella fastidiosa*, Q9PDN2), *Lb*EcfA2 (*Lactobacillus brevis*, Q03PY6), *Tm*EcfA2 (*Thermotoga maritima*, Q9WY65), *Bc*CaATPase (*Bacillus cereus*, Q73E41), *Lm*CaATPase (*Listeria monocytogenes*, Q8Y8Q5), *Se*PacL (*Synechococcus elongatus*, P37278), Rv0908 (*M. tuberculosis*, P9WPT1), *Afu*CopA (*Archaeoglobus fulgidus*, O29777), *Ec*KdpB (*E. coli*, P03960), *Af*KdpB (*Agrobacterium fabrum*, Q8U9D9), *Sc*KdpB (*Streptomyces coelicolor*, Q9X8Z9), Rv1030 (*M. tuberculosis*, P9WPU3), *Vc*MsbA (*V. cholerae*, 9KQW9), *Hi*MsbA (*Haemophilus influenza*, P44407), *St*MsbA (*Salmonella typhimurium*, P63359) and *Ec*MsbA (*E. coli*, P60752); the organism name and UniProt id of each protein is provided in the parenthesis. The free/compelx iron and oxyanion transporting NBDs are denoted as green and magenta diamonds, respectively. Similarly, the ECF-type ABC transporters (cyan) and P-type APases (blue and olive) are also represented as diamonds. The NBDs of ABC exporters (MsbA) are marked as red diamonds. (C-E) The protein-protein interaction network of the proteins Rv0265c (green), Rv1857 (yellow), Rv2059 (green), Rv2326c (green) and Rv3044 (green). The interacting proteins belonging to the same operon are also shown in the matching color. The nodes of the other interacting proteins (ORF id or gene name) are represented in magenta. The protein-protein interactions of the targeted proteins of the same operon are denoted as solid red lines, while those of other proteins as solid grey lines. The other details of each interacting protein are provided in Table B2.

4.3.3. The proteins Rv0265c and Rv3044 belong to a novel subcluster A-II-b of SBPs

Primary structure analysis of the probable SBPs Rv0265c, Rv1857, Rv2059 and Rv3044 suggested their probable metal ions/complexes. To further verify the results, their three-dimensional structures were modeled (Figure B1A-B1D); the other details are provided in Table B3. Since the crystal structure of the protein Rv0265c is available (PDB id: 4PM4) in the literature, it was directly used for the analysis (Figure B1A). Further, since the three-dimensional tertiary structure of the closest homolog of the protein Rv3044, a ferric dicitrate-binding protein from *E. coli* (*Ec*FecB), is unavailable in the literature, its structure was also modeled (Figure B1E and Table B3). The overall topology of the modeled structures of these proteins is similar to that of the canonical SBPs of ABC transporters possessing the N- and C-terminal domains connected by a hinge region (Figure B1). The SBPs of ABC transporters are structurally classified into seven clusters, viz. A-G (Scheepers et al., 2016). The defining features of each of these clusters are the hinge region of the SBPs. The structure/models of the proteins Rv0265c, Rv2059, Rv3044 and *Ec*FecB display topological similarity with that of the cluster A SBPs (Figure B1). On the other hand, the structural model of the protein Rv1857 exhibits similarity with the cluster D SBPs (Figure B1D). Based on their substrate specificity, the clusters A and D SBPs are further subclassified into two (A-I and A-II) and four (D-I, D-II, D-III and D-IV) subclusters, respectively (Scheepers et al., 2016).

Thus, to identify the probable ligands for the proteins Rv0265c, Rv1857, Rv2059 and Rv3044, including *Ec*FecB, a structure-based evolutionary tree with the structural homologs of these proteins (a total of 45) was performed; the details of the homologs considered are provided in Table B4 and B5. The result reveals that the proteins Rv0265c and Rv3044, including its closest homolog *Ec*FecB (ORF id: b4290) groups with the subcluster A-II SBPs (Figure 4.3). Further, an in-depth analysis of the phylogenetic tree demonstrates that the subcluster A-II can be subclassified into two clades. The SBPs clustered in the first clade bind molecules having porphyrin or corrin-like rings such as heme and cyanocobalamin, while that cladding in the second group bind iron-complexes such as iron citrate, staphyloferrin A and schizokinen molecules (Figure 4.3 and Figure B1). Thus, in this study, we renamed these groups as subcluster A-II-a and A-II-b, respectively. That suggests that the proteins Rv0265c, Rv3044 and *Ec*FecB (b4290) would possibly bind iron-complexes and thus belong to the subcluster

A-II-b SBPs (Figure 4.3). Interestingly, the results reveal that the proteins Rv2059 and Rv1857 grouped with the subcluster A-I and D-III-a SBPs, indicating their probable substrates to be a divalent metal ion and oxyanions, respectively (Figure 4.3).

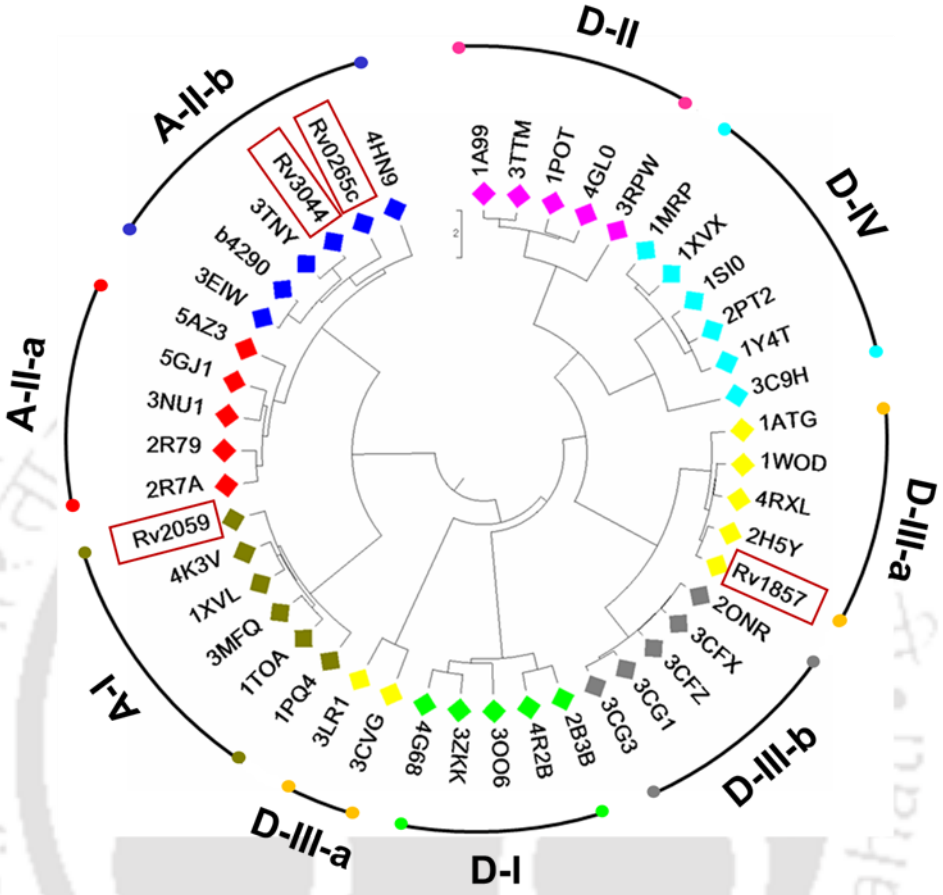


Figure 4.3. Structure-based distance tree of cluster A and D SBPs. The structures used for the analysis are denoted by their PDB ids, except for the modeled structures of the targeted proteins (red rectangles) and FecB protein from *E. coli* (*Ec*FecB), which is denoted by its ORF id (b4290). Each subcluster of the clusters A and D SBPs are represented in different colored diamonds.

4.3.4. Active-site pocket of the metal uptake SBPs are quintessential to accommodate their cognate substrates

The initial sequence and structure-based evolutionary analysis suggested the potential cognate substrate molecules for the targeted SBPs. To reaffirm the probable substrates which can be accommodated, the electrostatic potential charge distribution, area and

volume of the active site of the target proteins were computed. The electrostatic potential charge distribution result suggests that the protein Rv0265c possesses a negative charge at its active site, similar to its closest homolog, an iron-siderophore-binding protein PiaA, from S. *pneumoniae* (*Sp*PiaA, PDB id: 4HMQ) (Figure 4.4A). On the other hand, the heme-binding protein from *Burkholderia cenocepacia* (*Bc*BhuT, PDB id: 5Y89) contains a combination of both positively- and negatively-charged active site (Figure 4.4A). The protein Rv3044, a probable iron-complex-binding protein, is comprised of a combination of both positively and negatively-charged active sites; however, its closest homolog *Ec*FecB shows an absolute positively-charged active-site pocket indicating that simple iron complexes might possess varying electrostatic charges in their active site (Figure 4.4B). The protein Rv1857 contains a positively charged active site suggesting the binding of a negatively-charged molecule such as molybdate (MoO_4^{2-}) and tungstate (WO_4^{2-}); similar to its closest homolog, a MoO_4^{2-}-binding protein from *E. coli* (*Ec*ModA, PDB id: 1AMF) (Figure 4.4C). Similarly, the protein Rv2059 shows the presence of a negatively-charged active site favorable for the binding of positively-charged divalent metal ions such as Mn^{2+}/Zn^{2+} ions; akin to *Syn*ZnuA, a Zn^{2+}-binding and *Syn*MntC, an Mn^{2+}-binding protein from *Synechocystis spp.* (PDB ids: 1PQ4 and 1XVL, respectively) (Figure 4.4D). An analysis of the active-site area and volume shows that the protein Rv0265c (289 $Å^2$ and 156 $Å^3$) is comparable to that of the *Bc*BhuT protein (376 $Å^2$ and 205 $Å^3$), in contrast to the protein *Sp*PiaA (489 $Å^2$ and 380 $Å^3$) (Figure 4.4E). Similarly, the active-site area and volume of the protein Rv3044 (344 $Å^2$ and 500 $Å^3$) are comparable to its closest homolog *Ec*FecB (287 $Å^2$ and 458 $Å^3$) (Figure 4.4F). Interestingly, the active-site area and volume of the protein Rv1857 (26 $Å^2$ and 7 $Å^3$) are identical to that of *Ec*ModA (24 $Å^2$ and 7 $Å^3$) (Figure 4.4G).

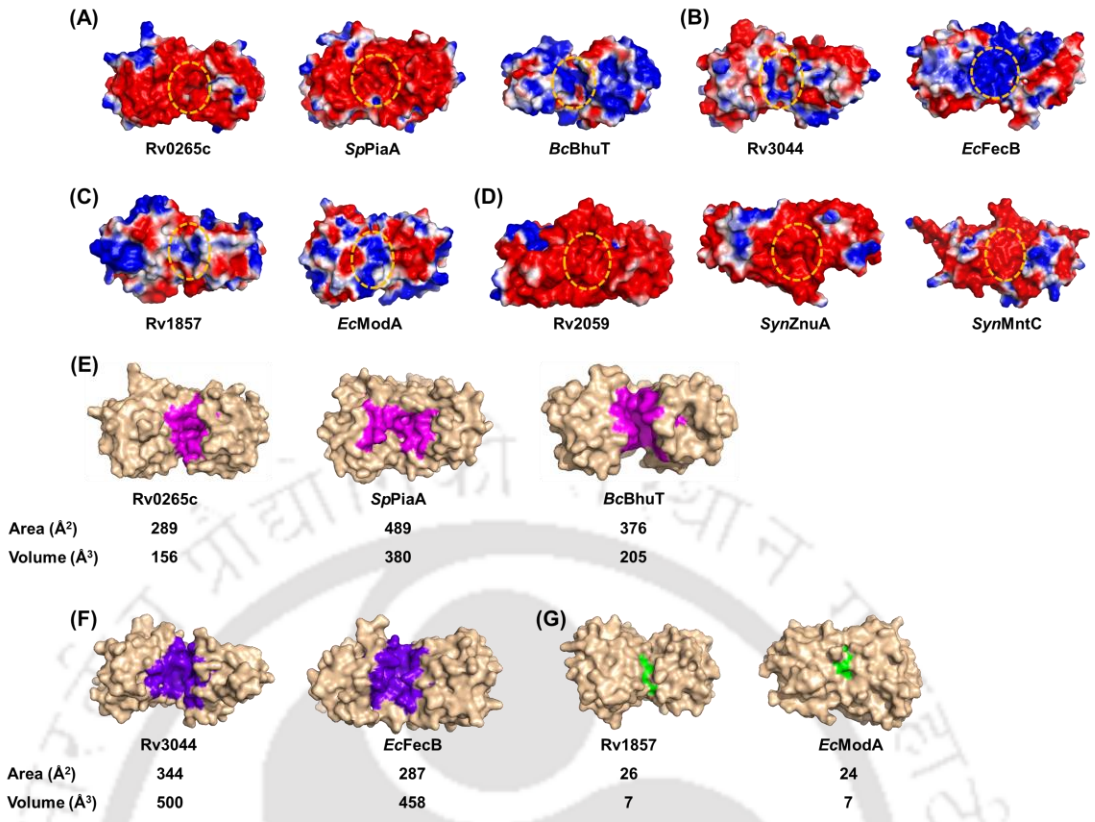


Figure 4.4. Electrostatic potential charge distribution and active-site area and volume of the metal uptake SBPs. Electrostatic potential charge calculation for the protein (A) Rv0265c compared with *Sp*PiaA (PDB id: 4HMQ) and *Bc*BhuT (PDB id: 5Y89), (B) Rv3044 and *Ec*FecB (modeled structure), (C) Rv1857 and *Ec*ModA (PDB id: 1AMF) and (D) Rv2059 compared with *Syn*ZnuA (PDB id: 1PQ4) and *Syn*MntC (PDB id: 1XVL). The positively- and negatively-charged surfaces are shown in blue (+1 kT/e) and red (-1 kT/e), respectively. Comparison of active-site area and volume of the protein (E) Rv0265c with its closest homolog *Sp*PiaA and *Bc*BhuT, (F) Rv3044 with its closest homolog *Ec*FecB and (G) Rv1857 with its closest homolog *Ec*ModA. The active-site region of Rv0265c, Rv3044 and Rv1857, along with their homologous proteins, are represented in magenta, blue and green, respectively.

4.3.5. The proteins Rv0265c, Rv2895c and Rv3044 code for iron-uptake ABC transporters

The ABC transport system IrtAB and Rv2895c (Rv1348-Rv1349-Rv2895c) has already been characterized experimentally to uptake iron inside bacterial cells (Rodriguez and

Smith, 2006; Arnold et al., 2020). In addition to this, our preliminary sequence analysis suggests that the proteins Rv0265c and Rv3044 also encode probable free and/or complex iron-binding proteins. To further confirm that the protein Rv0265c is an iron-siderophore-binding protein, the conservation of the iron-siderophore-binding residues was identified by performing a pairwise sequence alignment with its closest homolog *Sp*PiaA and heme-binding proteins. The result reveals that the iron-siderophore-coordinating residues Phe59, Phe131 and Arg198 are found to be conserved in the protein Rv0265c (Figure 4.5A). In contrast, no conservation of the active-site residues could be found between Rv0265c and heme-binding proteins (Figure B2A). Furthermore, a structural superimposition of the proteins Rv0265c (PDB id: 4PM4) and *Sp*PiaA (PDB id: 4HMQ) exhibits that the residues Phe59, Phe131 and Arg198 occupy a similar position to that of the ferrichrome-binding residues Asn83, Trp158 and Asn231 in *Sp*PiaA (Figure 4.5B). Thus, to further validate the binding of ferrichrome and heme molecules to the protein Rv0265c, molecular docking calculations were performed. These results corroborate the binding of ferrichrome and heme molecules to the protein Rv0265c with a binding energy of -8.18 and -11.23 kcal mol^{-1}, respectively (Figure 4.5C, 4.5D and Table B6). The docked ferrichrome molecule is held by the residues Trp58 and Gln182, whereas the residues Tyr39, Trp58, Arg198 and Arg221 coordinate the heme molecule in the active site of the protein Rv0265c (Figure 4.5C, 4.5D and Table B6). A similar analysis was performed for the protein Rv3044. A multiple sequence alignment (MSA) of the protein Rv3044 and known iron/heme-binding proteins show no conservation of the active-site residues (Figure B2A). However, the presence of an Arg218-Val219-Tyr220 sequence in Rv3044 akin to heme-binding proteins (Tyr87-Gln88-Arg89 in *Bc*BhuT protein) shows its involvement in Fe^{3+} ion binding (Figure B2A). The presence of a positively-charged arginine residue (Arg218) is crucial in maintaining the deprotonated state of an iron-binding tyrosine residue (Tyr220) in the protein Rv3044. Since the structural information of the closest homolog of the protein Rv3044, i.e., *Ec*FecB (a ferric-dicitrate-binding protein from *E. coli*) is unavailable, the tertiary structures of both the proteins Rv3044 and *Ec*FecB were modeled (Figure B1). A structural comparison of the proteins Rv3044 and *Ec*FecB exhibits that both the proteins possess a completely different active site (Figure B2B). This was further corroborated by a pairwise sequence alignment of both the proteins Rv3044 and *Ec*FecB (Figure B2C). To substantiate and estimate the binding of Fe^{3+}-citrate and -dicitrate molecules to the proteins Rv3044 and

*Ec*FecB, molecular docking simulations were performed. The results suggest that *Ec*FecB can bind Fe^{3+}-dicitrate molecules with a binding energy of -5.96 kcal mol^{-1}. The residues Arg66, Arg108, Tyr216, Ser247 and Arg252 interact with the Fe^{3+}-dicitrate molecule in the active-site pocket (Figure 4.5E and Table B6). Similarly, the protein Rv3044 shows a binding energy of -4.84 and -6.72 kcal mol^{-1}, with Fe^{3+}-citrate and Fe^{3+}-dicitrate, respectively (Figure 4.5F, 4.5G and Table B6). Interestingly, as expected, the active-site residues Gly162, Gln211, Arg218, Tyr220, Asn224, Phe225, Gln314, Thr315 and Glu317 interact with the Fe^{3+}-citrate and Fe^{3+}-dicitrate molecules (Figure 4.5F, 4.5G and Table B6). In summary, all the three proteins Rv0265c, Rv2895c and Rv3044 seem to be involved in the uptake of iron ions in different complex forms.

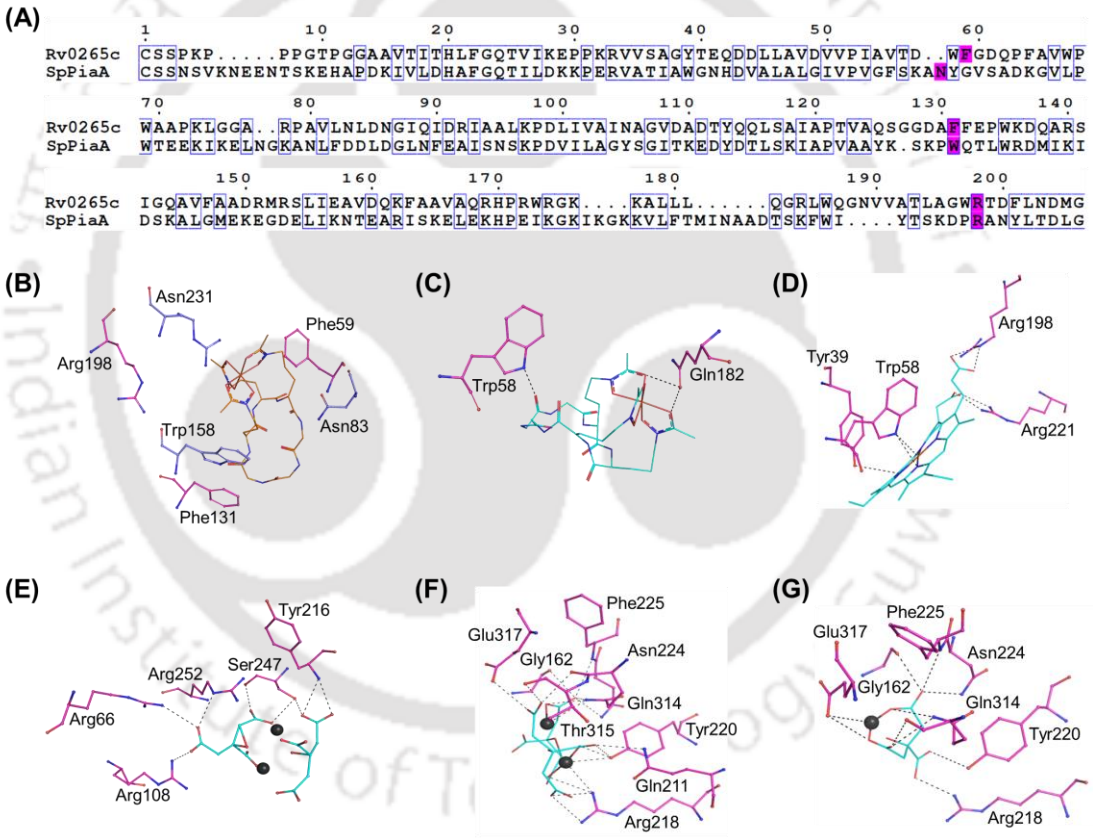


Figure 4.5. Active-site residue analysis and molecular docking of the proteins Rv0265c and Rv3044 with their probable ligands. (A) Pairwise sequence alignment of the protein Rv0265c with its closest homolog PiaA from *S. pneumoniae* (*Sp*PiaA, A0A0H2UPT5); the protein name and UniProt id is provided in the parenthesis). The ligand-binding residues are highlighted in magenta. (B) The active-site pocket comparison of the protein Rv0265c with *Sp*PiaA (PDB id: 4HMQ). All the active-site residues are

represented as a ball-and-stick model. The active-site residues involved in ferrichrome (orange line) binding in *Sp*PiaA and the corresponding residues in Rv0265c protein are shown in blue and magenta, respectively. (C and D) Molecular docking of the protein Rv0265c with ferrichrome and heme molecules. The residues coordinating the ferrichrome and heme (both in cyan line) are shown in magenta ball-and-stick. (E) Molecular docking of the protein *Ec*FecB with Fe^{3+}–dicitrate molecule. The molecular docking of the protein Rv3044 with (F) Fe^{3+}–dicitrate and (G) Fe^{3+}–citrate molecules. The residues coordinating the Fe^{3+}–citrate and Fe^{3+}–dicitrate (cyan line) molecules are shown in magenta ball-and-stick. The Fe^{3+} ion is shown as a black sphere.

4.3.6. The protein Rv1857 is a plausible molybdate-binding protein

Preliminary sequence analysis of the protein Rv1857 revealed its similarity with MoO_4^{2-} and/or SO_4^{2-}-binding proteins. To identify the residues involved in the MoO_4^{2-} and/or SO_4^{2-} binding, an MSA of the protein Rv1857 and its homologous MoO_4^{2-} and SO_4^{2-}-binding proteins was performed. The result reveals that the MoO_4^{2-}-binding residues Ser26, Ser54, Val157, and Tyr175 are conserved in the protein Rv1857 also (Figure 4.6A). However, the conserved residue Ala149 (*Ec*ModA) is substituted by Cys133 in the protein Rv1857 (Figure 4.6A). This can be supported by the active-site comparison of the protein Rv1857 and *Ec*ModA (Figure 4.6B). In contrast, most of the active-site residues, except Ser54 and Tyr175, of the SO_4^{2-}-binding proteins are not conserved in the protein Rv1857 (Figure 4.6A). To further corroborate the binding of MoO_4^{2-}, and not SO_4^{2-}, molecular docking calculations of the protein Rv1857 with MoO_4^{2-}, tungstate (WO_4^{2-}) and SO_4^{2-} molecules were performed. The docking results demonstrate that MoO_4^{2-} and WO_4^{2-} ions bind to the protein with a binding energy of -3.01 and -2.13 kcal mol^{-1}, respectively (Figure 4.6C, 4.6D and Table B6). As anticipated, the MoO_4^{2-} ion is coordinated by the residues Ser26, Ser54, Cys133, Ser156, Val157 and Tyr175 in the active site of the protein (Figure 4.6C and Table B6). Akin to MoO_4^{2-}, the WO_4^{2-} ion is anchored by all the active-site residues, except Ser26 and Tyr175, in the active site of the protein (Figure 4.6D and Table B6). Interestingly, the protein Rv1857 does not show any binding to the SO_4^{2-} ion in the vicinity of the active site (Table B6).

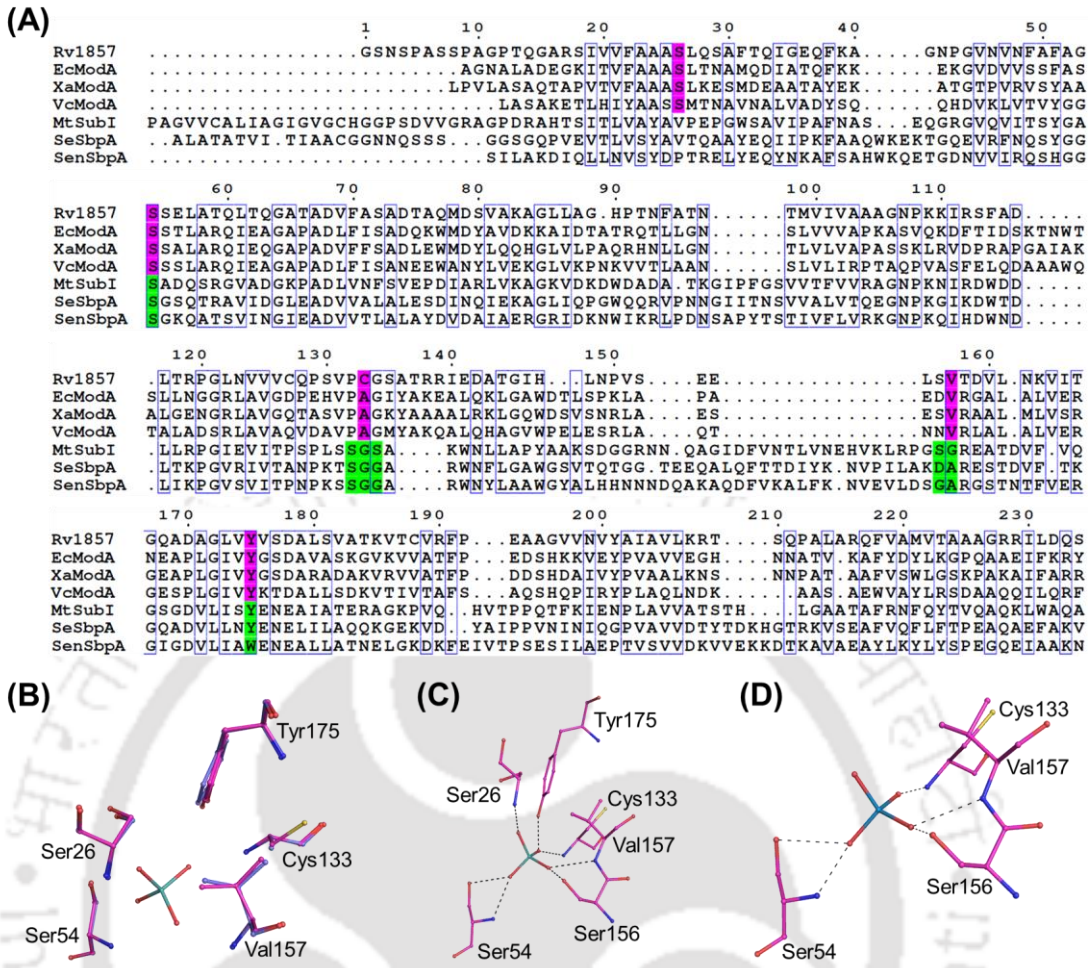


Figure 4.6. Active-site residue analysis and molecular docking of the protein Rv1857 with MoO_4^{2-} and WO_4^{2-}. (A) Multiple sequence alignment of Rv1857 protein with its closest homologs. The proteins used in MSA are *Ec*ModA (*E. coli*, P37329), *Xa*ModA (*Xanthomonas axonopodis*, Q8PHA1), *Vc*ModA (*Vibrio cholera*, 9KLL7), *Mt*SubI (*M. tuberculosis*, P71744), *Se*SbpA (*Synechococcus elongatus*, P27366) and *Sen*SbpA (*Salmonella enterica*, P02906); the organism name and UniProt id of each protein are provided in the parenthesis. The amino acid residues which interact with MoO_4^{2-} in the protein ModA are shaded in magenta, whereas those involved in SO_4^{2-} binding in the proteins SubI/SbpA are shaded in green. (B) The active-site pocket comparison of the protein Rv1857 with *Ec*ModA (PDB id: 1AMF). All the active-site residues are represented as a ball-and-stick model. The active-site residues involved in MoO_4^{2-} (cyan ball-and-stick) binding in *Ec*ModA and the corresponding residues in the protein Rv1857 are shown in blue and magenta, respectively. The residue Ala149 (blue) in the protein *Ec*ModA is replaced by Cys133 in the protein Rv1857. (C and D) Molecular

docking of the protein Rv1857 with MoO_4^{2-} and WO_4^{2-} ions. The residues coordinating the MoO_4^{2-} and WO_4^{2-} (cyan ball-and-stick) are shown in the magenta ball-and-stick model.

4.3.7. The protein Rv2059 is a potential Zn^{2+}-binding protein

The protein Rv2059 (length: 511 amino acid residues), which is annotated as an "uncharacterized protein" in the UniProtKB database, shows a distant evolutionary relationship with cluster A-I metal (e.g., manganese, zinc, etc.) uptake SBPs (length: 250-350 amino acid residues) (Figure 4.2A). The operonic, as well as protein-protein interaction network analysis, suggested its probable TMD subunit as Rv2060; however, no cognate NBD subunit was observed (Figure 4.2D). It is known that the cluster A-I SBPs conserve three histidine residues to chelate zinc ions (Wei et al., 2007), while two histidines and either two glutamates or one glutamate and one aspartate residue for coordinating the manganese ions (Counago et al., 2014). To confirm the presence of these residues in the protein Rv2059, MSA of the target protein with its homologous cluster A-I divalent metal SBPs was performed. The MSA results demonstrate that the protein Rv2059 also conserves two histidines (His137 and His197) and two glutamate residues (Glu261 and Glu341), suggesting the binding of the manganese ions (Figure 4.7A). Interestingly, the protein Rv2059 lacks the histidine-rich flexible loop, a signature loop of the Zn^{2+}-binding cluster A-I SBPs (Figure 4.7A) (Mandal et al., 2017). Further, structural comparison of the protein Rv2059 and two of its closest homologs ZnuA and MntC from *Synechocystis* sp. (PDB ids: 1PQ4 and 1XVL) displays spatial conservation of the two histidine residues His137 and His197 (His83 and His179 in *Syn*ZnuA) (Figure 4.7B and 4.7C). However, a third histidine residue (His243 in *Syn*ZnuA) is substituted by the residue Glu261 (in Rv2059). Similarly, structural comparison of the protein Rv2059 and *Syn*MntC exhibits the conservation of the residues His137, His197 and Glu261. On the other hand, Asp295 (in *Syn*MntC) is replaced by the residue Glu341 (in Rv2059) (Figure 4.7C). In the modeled structure, which mimics an open conformation of the protein Rv2059, the loop L15 consisting of the residue Glu341 is ~7.0 Å away from the loop L17 having the residue Asp295 in *Syn*MntC (a closed conformation) (Figure 4.7D). This suggests that either the loop L15 of the protein Rv2059 might shift towards the metal ion (Zn^{2+}/Mn^{2+}) in its close confirmation or a water molecule might play an important role in Zn^{2+}/Mn^{2+} binding in

the protein Rv2059. Altogether, the results suggest that the protein Rv2059 is probably an unusual Zn^{2+}-binding protein which can also bind Mn^{2+} ion.

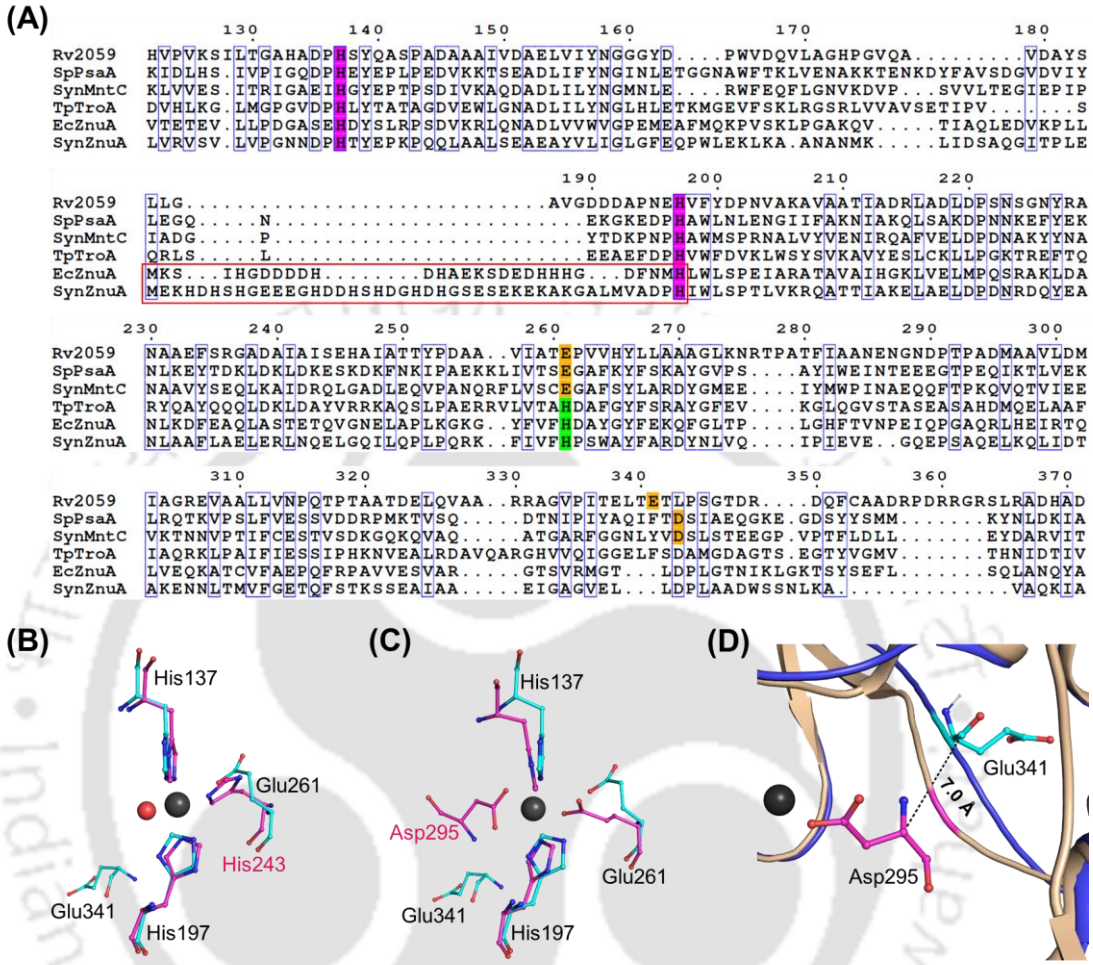


Figure 4.7. Comparing the spatial orientation of the active-site residues of the protein Rv2059 with its homologous proteins. (A) Multiple sequence alignment of the protein Rv2059 with its closest homologs. The proteins used in MSA are *Sp*PsaA (*S. pnuemoniae*, P0A4G2), *Syn*MntC (*Synechocystis spp.*, Q79EF9), *Tp*TroA (*Treponema pallidum*, P96116), *Ec*ZnuA (*E. coli*, P39172) and *Syn*ZnuA (*Synechocystis spp.*, P73085); the organism name and UniProt id of each protein is provided in the parenthesis. The residues binding to both the metal Zn^{2+} and Mn^{2+} ions are highlighted in magenta, while those specific to Mn^{2+} and Zn^{2+} ions are shown in orange and green shades, respectively. The histidine-rich loop in *Ec*ZnuA and *Syn*ZnuA is shown in the red box. The active-site pocket comparison of the protein Rv2059 with (B) *Syn*ZnuA (PDB id: 1PQ4) and (C) *Syn*MntC (PDB id: 1XVL). All the active-site residues are represented as a ball-and-stick model. The active-site residues involved in Zn^{2+} and

Mn^{2+} ion (black sphere) binding in *Syn*ZnuA and *Syn*MntC, respectively, are shown in magenta, whereas the corresponding residues in Rv2059 are in the cyan ball-and-stick model including Glu341. The water molecule which interacts with the Zn^{2+} ion in *Syn*ZnuA is represented as a red sphere. (D) The residue Glu341 (cyan ball-and-stick) present in the loop L15 of Rv2059 (blue) is located ~7.0 Å away from the Asp295 (magenta ball-and-stick) present in loop L17 of *Syn*MntC (wheat).

4.3.8. The proteins Rv1859, Rv2326c and Rv3041c conserve all the intrinsic features of a typical nucleotide-binding domain (NBD) of ABC importers

A preliminary database search for metal uptake ABC transporters in *M. tuberculosis* H37Rv revealed that it contains a total of five probable NBDs, viz. Rv0908, Rv1030, Rv1859, Rv2326c and Rv3041c. Two of them, Rv0908 and Rv1030, possibly belong to the P-type ATPases instead of ABC transporter NBDs. The proteins Rv1859 and Rv3041c show homology with the ATP-binding proteins of oxyanion and different iron-complex transport systems, respectively. The protein Rv2326c seems to be associated with the protein Rv2325c (query coverage: 27-87%, sequence identity: 24-42%), an 'energy coupling factor ATP-binding protein EcfA1/2' and 'energy coupling factor transporter transmembrane protein EcfT', respectively, potentially involved in cobalt ion transport. Moreover, a structural homology search of the predicted models of the proteins Rv2325c and Rv2326c exhibits similarities with the energy coupling factor transmembrane and ATP-binding proteins involved in cobalt transport (Table B4). Since the NBDs of the ABC transporter superfamily show a high sequence and structural similarities, they follow a conserved energy coupling mechanism (Wilkens, 2015). At the primary structure level, NBDs contains highly conserved motifs, viz. two Walker (A and B), a signature motif and glutamine (Q) and histidine (H) loops along with weakly conserved loops such as tyrosine (Y), proline (P) and aspartate (D) (Procko et al., 2006). Additionally, a structurally diverse region (SDR) of ~30-40 amino acid residue long is known to be unique and control the interaction of NBDs to TMDs (Oswald et al., 2006). To affirm the presence of these signature features in the proteins Rv1859, Rv2326c and Rv3041c, an MSA with known NBDs of both ABC transporters was performed. The result confirms the conservation of the Walker A & B and signature motifs along with Q- and H-loops in these proteins as well (Figure 4.8). Interestingly, the Y-, P- and D-loops are also conserved (Figure 4.8). Moreover, the SDR is also

present in these proteins (Figure 4.8). In summary, the proteins Rv1859, Rv2326c and Rv3041c are possibly the NBDs of the metal uptake ABC transporters in *M. tuberculosis* H37Rv.

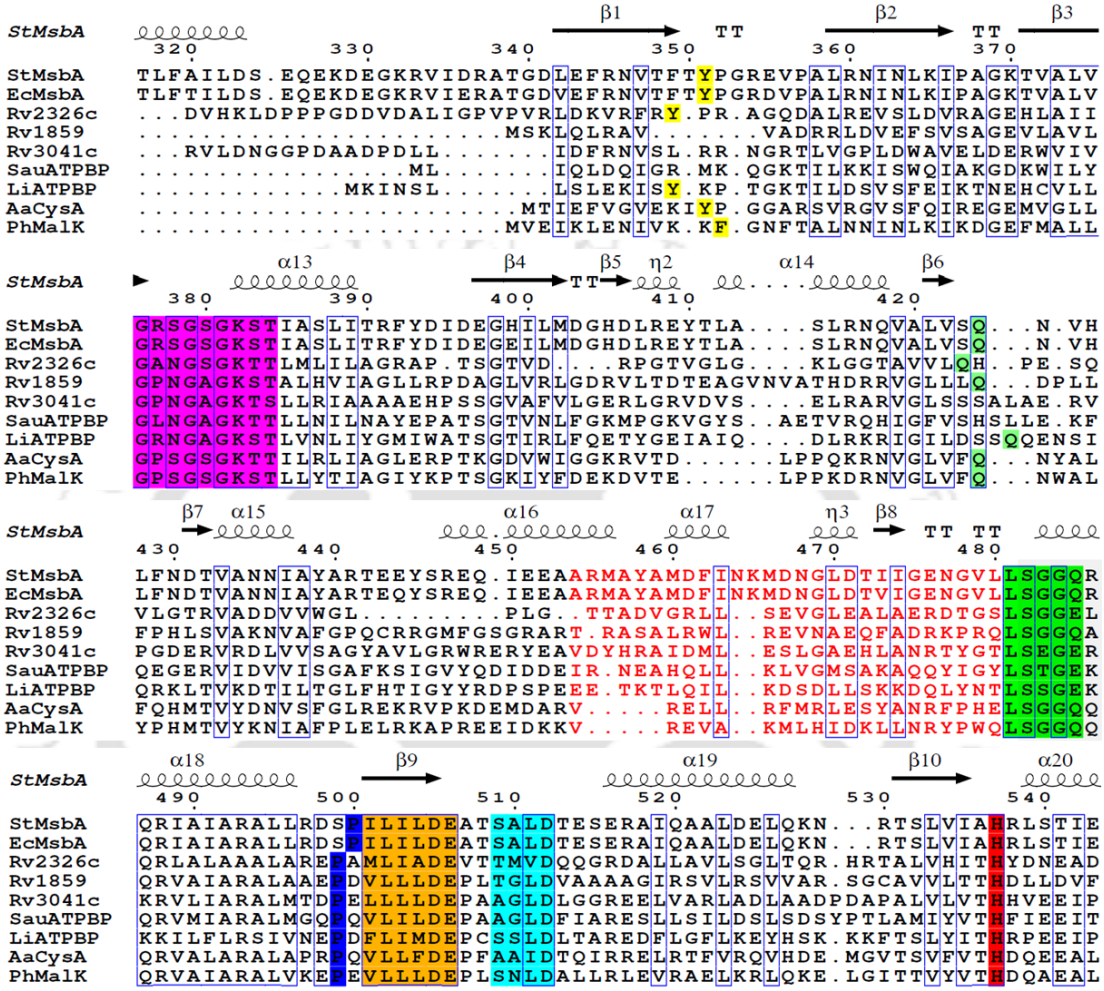


Figure 4.8. Multiple sequence alignment of the proteins Rv1859, Rv2326c and Rv3041c with their homologous proteins. The proteins used in the alignment are *St*MsbA (*Salmonella typhimurium*, P63359), *Ec*MsbA (*E. coli*, P60752), *Sau*ATPBP (*Staphylococcus aureus*, A0A0H2WZ98) *Li*ATPBP (*Leptospira interrogans*, Q72QN4) *Aa*CysA (*Alicyclobacillus acidocaldarius*, Q9RHZ7) and *Ph*MalK (*Pyrococcus horikoshii*, O57942); the organism name and UniProt id of each protein are provided in the parenthesis. The Walker A and B motifs are highlighted in magenta and orange, whereas the signature motif in green, respectively. The Y-, Q-, P-, D- and H-loops are shown in yellow, light green, blue, cyan and red shades, respectively. The amino acid residues highlighted in red texts represent the structurally diverse region of

NBDs. The secondary structure shown on top of the alignment is of the protein MsbA from *Salmonella typhimurium* (PDB id: 6BL6).

4.3.9. The proteins Rv0908 and Rv1030 are P-type ATPases associated with metal ion transport

Two ORF ids Rv0908 and Rv1030, were also targeted for the analysis due to their annotation as 'ATP-binding subunits' in the protein databases. Thus, to verify if these proteins belong to the ABC transporters, they were included in the analysis. Homology search suggests their association with sodium, potassium, calcium and copper transporting ATPases. Further, an evolutionary analysis of these proteins along with their homologs demonstrates their grouping with P-type ATPases (Figure 4.2B). P-type ATPases are a large family of integral membrane proteins transporting cations across the cell membrane against a concentration gradient by coupling ATP hydrolysis (Kuhlbrandt, 2004). P-type ATPases are comprised of four principal domains, namely the phosphorylation (P), the nucleotide-binding (N), the actuator (A) and the membrane (M) domain having maximum conservation in P- and N-domain (Bublitz et al., 2011). Thus, an MSA of these proteins with homologs was performed. The result exhibits that the protein Rv0908 and Rv1030 possess the P-type ATPases signature motif DKTGTLT (Figure 4.9A). In addition, the TGE loop, important for ion translocation (Kuhlbrandt, 2004), in the A-domain is also conserved in the proteins Rv0908 and Rv1030 (Figure 4.9A). To identify the structural features of the proteins Rv0908 and Rv1030, their three-dimensional structures were predicted (Table B3). A structural homology search of these proteins reveals their structural similarity with "calcium-transporting ATPase" and "potassium-transporting ATPase", respectively (Table B4). Further, the conserved motifs (DKTGTLT and TGE) occupy the same spatial positions as in the protein KdpB from *E. coli* (*Ec*KdpB, PDB id: 5MRW) (Figure 4.9B). An electrostatic potential charge distribution analysis reveals that the proteins Rv0908 and Rv1030 are comparable to their closest homologs, the *Ec*KdpB and Ca^{2+}-ATPase from *Oryctolagus cuniculus* (*Oc*Ca^{2+}-ATPase, PDB id: 3FPS), respectively (Figure 4.9C-F). In summary, these observations suggest that the proteins Rv0908 and Rv1030 belong to the P-type ATPases rather than to the ABC transport systems.

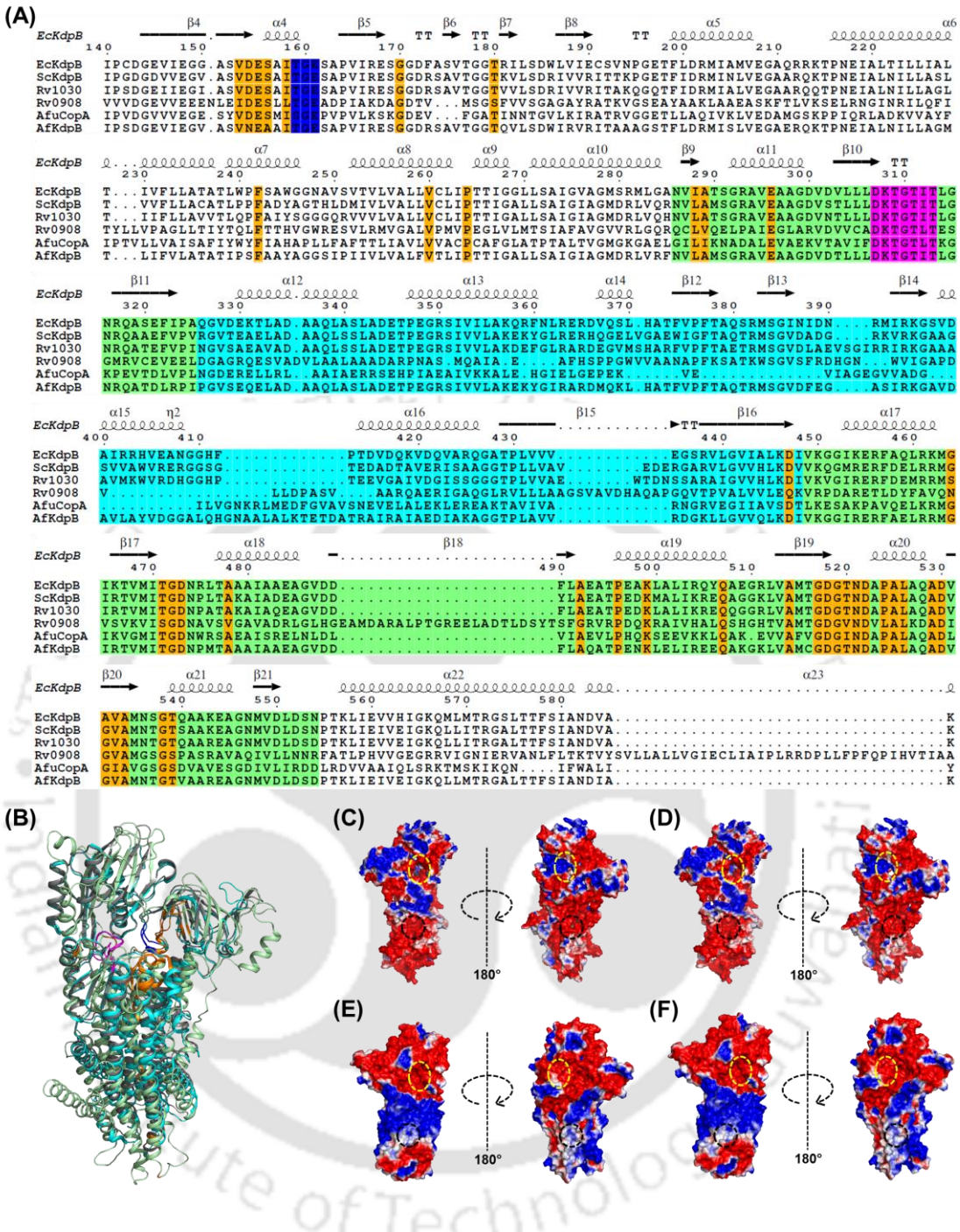


Figure 4.9. Primary and tertiary structure analysis for the proteins Rv0908 and Rv1030. (A) Multiple sequence alignment of the proteins Rv0908 and Rv1030 with their homologous proteins. The proteins used in the alignment are *Ec*KdpB (*E. coli*, P03960), *Sc*KdpB (*Streptomyces coelicolor*, Q9X8Z9), *Af*KdpB (*Agrobacterium fabrum*, Q8U9D9) and *Afu*CopA (*Archaeoglobus fulgidus*, O29777); the organism name and UniProt id of each protein are provided in the parenthesis. The P-domain is shown in green background, whereas the N-domain is highlighted in the cyan background. The conserved ATP-binding domain residues are highlighted in magenta

along with the aspartate residue (marked in red asterisk), which is the phosphorylation site of P-type ATPases. (B) Structural superimposition of Rv0908 (pale green) and Rv1030 (grey) with *Ec*KdpB (cyan) protein (PDB id: 5MRW). The motifs DKTGTLT and TGE, as well as other conserved residues, are represented in magenta, blue and orange, respectively, both at the sequence and structural levels. Electrostatic potential charge distribution of (C) *Ec*KdpB, (D) Rv0908, (E) *Oc*Ca^{2+}-ATPase (PDB id: 3FPS) and (F) Rv1030. The yellow dotted circles represent the phosphorylation site, while the black dotted circle shows the part of the cytoplasmic cation access pathway.

4.4. DISCUSSION

Mycobacterium tuberculosis, an etiological agent of tuberculosis (TB), has an adaptive nature to remain and survive within the host system for a long period of time (Nguyen, 2016). Inside host cells, it requires various micronutrients, including metal ions, to survive (Chai et al., 2018). The ABC importers play an important role in this endeavor of the bacterium (Soni et al., 2020). In this study, a total of 19 ORFs encoding nine metal uptake transporters were identified, out of which seven belong to the ABC importers. Among these seven ABC import systems, two have been characterized experimentally (Rempel et al., 2020; Farhana et al., 2008; Santhagopalan and Rodriguez, 2012). In this study, the remaining five transport systems were characterized using *in silico* approaches.

One of the essential metal ions required by *M. tuberculosis* for its replication is iron. However, the host system sequesters these metal ions more efficiently than the bacterium (Zhang et al., 2020). Thus, the pathogen requires multiple systems to uptake the iron ions, either in free or complex form. In the bloodstream during the initial infection, the bacterium chelates iron from transferrin, lactoferrin, ferritin and heme (Chao et al., 2018). In addition, hemolysins are known to lyse erythrocytes releasing haemoglobin into the bloodstream, making it a promising source of heme molecule for the bacterium (Chao et al., 2018).

In *M. tuberculosis* H37Rv, three ABC import systems, Rv0265c, Rv1348-Rv1349-Rv2895c and Rv3044-Rv3041c seem to play key roles in sequestering iron ions either

in free or in complex forms. The protein Rv0265c has been reported previously to be associated with the DppABCD dipeptide ABC import system encoded by the ORFs Rv3666c-Rv3663c for heme acquisition (Mitra et al., 2019). The results of this study indicate that the protein Rv0265c can also be associated with the ABC import system encoded by the ORFs Rv3044-Rv3041c and may aid in the uptake of Fe^{3+}-siderophore. These results suggest that the protein Rv0265c is an orphan SBP possibly involved in the uptake of both heme and Fe^{3+}-complexes to fulfill the iron requirement of the bacterium *M. tuberculosis*. In conclusion, based on these results, the protein Rv0265c seems to be a unique SBP, which is capable of binding both heme and other ferric-siderophores transiently. Structure-based evolutionary analysis for the iron-binding proteins also suggested a further classification of subcluster A-II SPBs as subcluster A-II-a and A-II-b based on the type of ligands they bind and transport. Subcluster A-II-a is involved in the binding of metal ions trapped in porphyrin or corrin-like ring structures such as heme or cyanocobalamin, which are highly complex in nature, while subcluster A-II-b SBPs bind relatively simpler iron-complexes such as iron-citrate, staphyloferrin A or schizokinen. However, the paucity of the three-dimensional structures of the subcluster A-II-b SBPs in the RCSB protein data bank impeded further validation of this classification.

Apart from Fe^{3+}-siderophores and heme, cobalamin (vitamin B12) also plays a crucial role in *M. tuberculosis* for its growth and survival (Gopinath et al., 2013). The pathogen *M. tuberculosis* synthesizes cobalamin *de novo* in the presence of cobalt as well as scavenge cobalamin from the host cells (Rempel et al., 2020). Cobalamin uptake in *M. tuberculosis* is facilitated by the protein Rv1819c, an ABC type transport system (Rempel et al., 2020). In this study, the results suggest that the cobalt uptake can be fulfilled by an ECF-type ABC importer encoded by ORFs Rv2325c-Rv2326c. In addition to iron and cobalamin, the bacterium *M. tuberculosis* requires other transition row metal ions, e.g. zinc (Riccardi et al., 2008). Several Zn^{2+}-transporting proteins such as CorA (or MgtE) uptake systems (Rv1293c, Rv0362 or Rv1232c), P-type ATPase CtpF (Rv1997) and zinc phosphate transporter PitA (Rv0545c) are reported in the literature (Riccardi et al., 2008). In this study, an ABC transport system encoded by the proteins Rv2059-Rv2060 seems to be involved in free Zn^{2+} ion uptake. The protein Rv2060 is an ortholog of *E. coli* ZnuB, a TMD subunit of ZnuABC transporter. The

results in this study also suggest that the protein Rv2059 is possibly an extracellular Zn^{2+}-binding protein, having a certain binding affinity for Mn^{2+} ion also.

The human pathogen *M. tuberculosis* also requires oxyanions such as molybdate for its survival and pathogenesis (Forrellad et al., 2013). In this study, results affirm that MoO_4^{2-} can be imported by an ABC transporter encoded by the ORFs Rv1857-Rv1858-Rv1859. Interestingly, The MoO_4^{2-} ABC transporter in *M. tuberculosis* is highly similar to that of *E. coli*. In MoO_4^{2-}-binding proteins, the amino acid residues anchoring the substrate are generally variable due to their additional binding to both MoO_4^{2-} and WO_4^{2-} ions (Mandal et al., 2019). Interestingly, in contrast to other microorganisms where the majority of the metal uptake ABC subunits (SBPs, TMDs and NBDs) are located in tandem at the genetic level, most of the ABC transporter subunits, except Rv1857-Rv1858-Rv1859, in *M. tuberculosis* are found to be spread on the genome. The sharing mechanisms of TMDs and NBDs of different SBPs have been reported for metal and sugar uptake ABC transporters previously (Mandal et al., 2019; Chandravanshi et al., 2019). The results from this study and the available literature hint that the protein Rv0265c shares the TMD and NBD subunits of either Rv3041c or DppA proteins or both for substrate transport (Mitra et al., 2019). Similarly, the proteins Rv3044-Rv3041c utilizes the TMD subunit of some unknown ABC transporter for Fe^{3+}-dicitrate translocation into the cytosol. Likewise, the proteins Rv2059-Rv2060 seem to lack their canonical NBD subunit and possibly share it from an unknown ABC transport system. The notion of subunit sharing has earlier been reported for ABC transporters. For example, ECF-type ABC transporters are known to share their NBD subunit with many TMDs to form a hollow transporter for different substrates (Eitinger et al., 2011).

Thus, this study suggests that the bacterium *M. tuberculosis* H37Rv is comprised of seven metal uptake ABC transporters (Figure 4.10). Out of which, three (Rv0265c, Rv2895c and Rv3044) are likely to be involved in the uptake of iron-complex (siderophores), one (Rv1857) for MoO_4^{2-} (and WO_4^{2-}) and one (Rv2059) for Zn^{2+}/Mn^{2+} uptake (Figure 4.10). The other ABC transport system Rv2325c-Rv2326c, an ECF-type or Type-III ABC transporter, is identified to be a probable cobalt transporter. Two proteins Rv0908 and Rv1030, which are annotated as ATP-binding proteins in the protein databases, are identified as probable P-type ATPases (Figure 4.10). This study

presents an instance of identifying and annotating other hypothetical proteins from pathogens using an integrative *in silico* approach. The results of which might help in identifying potential drug targets for human pathogens such as *M. tuberculosis*.

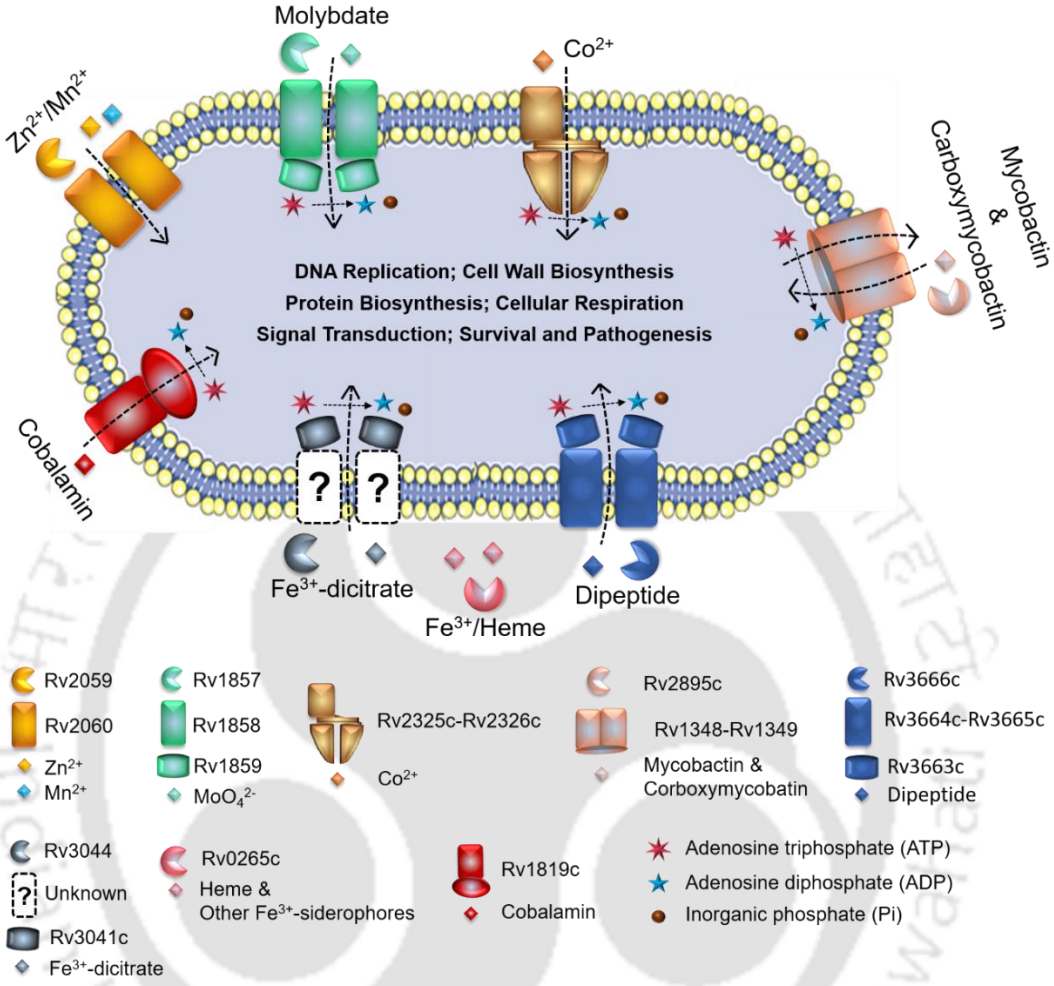


Figure 4.10. Schematic representation of the repertoire of probable metal uptake ABC transporters available in *M. tuberculosis* H37Rv.

CHAPTER 5

Structural and functional characterization of the protein MctA

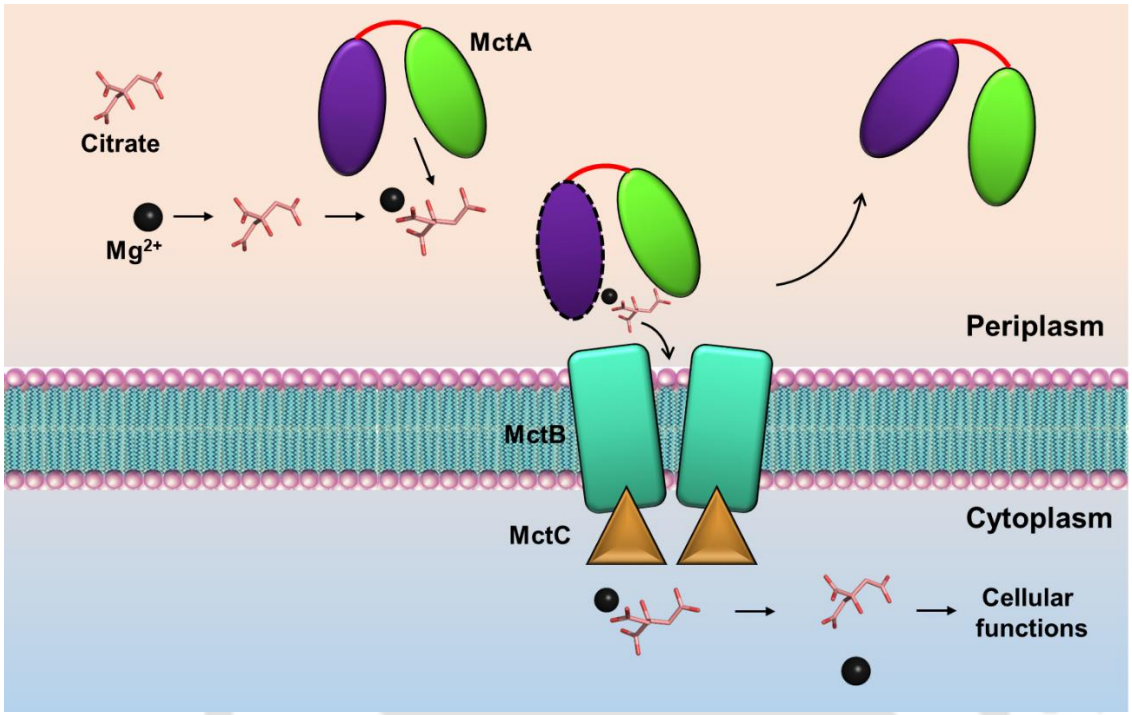
The chapter has been published as:

1. **Mandal SK** and Kanaujia SP (2021). Structural and thermodynamic insights into a novel Mg^{2+}-citrate-binding protein of ABC transporter. ***Acta Crystallographica Section D: Structural Biology***, D77, https://doi.org/10.1107/S2059798321010457.

ABSTRACT

More than one-third of the proteins require metal ions to accomplish their functions, making the metal ions to be obligatory for the growth and survival of microorganisms in varying environmental niches. In prokaryotes, besides their involvement in various cellular and physiological processes, metal ions also stimulate the uptake of citrate molecules. Citrate is a source of carbon and energy and is reported to be transported by secondary transporters. In Gram-positive bacteria, citrate molecules are transported in complex with divalent metal ions, while in Gram-negative bacteria, they are translocated by Na^+/citrate symporters (CitS). In this study, we report the presence of a novel divalent metal ion-complexed citrate uptake system that belongs to the primary active ABC transporter superfamily. Members of this superfamily possess an eminent mechanism of substrate transport by hydrolyzing ATP molecules which fuel the transport system. For the uptake, the metal ion-complexed citrate molecules are sequestered by substrate-binding proteins (SBPs) and transferred to transmembrane domains (TMDs) for their transport. This study reports the crystal structures of an Mg^{2+}-citrate-binding protein MctA (ORF id: TTHB177) from a Gram-negative thermophilic bacteria *Thermus thermophilus* HB8 in both apo and holo forms at a resolution range of 1.63 to 2.50 Å. Despite being capable of binding various divalent metal ions, the protein MctA follows the coordination geometry to bind its physiological metal ion, Mg^{2+}. The results also suggest a novel sub-classification of cluster D SBPs, which can bind and transport divalent metal ion-complexed citrate molecules. Comparative assessment of the open and closed conformations of the wild type and mutant proteins suggests a gating mechanism of ligand entry following an "asymmetric domain movement" of the N-terminal domain (NTD) for ligand binding.

Graphical abstract



5.1. INTRODUCTION

Nearly one-third of the protein-mediated processes necessitate metal ions to accomplish the fundamental roles in biological systems (Holm et al., 1996; Waldron and Robinson, 2009). Metal ions not only serve as essential cofactors in various cellular and physiological processes but also govern the structures and functions of various biological molecules. The divalent metal ions such as magnesium (Mg), calcium (Ca), manganese (Mn), zinc (Zn), iron (Fe), nickel (Ni) and copper (Cu) have been reported to play pivotal roles in numerous cellular processes (Zhen et al., 2009). For example, Fe is capable of fine-tuning the regulation of many biosynthetic, metabolic and therefore, cellular processes due to its unstable redox chemistry obligating it as a cofactor for organisms (Crichton and Boelaert, 2001). Similarly, Mg is crucial for many energy-requiring metabolic processes, protein synthesis, for maintaining membrane integrity, neuromuscular excitability and nervous tissue conduction, muscle contraction, hormone secretion, etc. (Laires et al., 2004). Additionally, many divalent metal ions regulate and inhibit several enzymatic reactions like protein kinase activity (Knape et al., 2017). Succinctly, divalent metal ions are central for a plethora of life

processes, including DNA replication, protein and cell-wall biosynthesis, cellular respiration, signal transduction, etc. (Waldron and Robinson, 2009; Kehres and Maguire, 2003; Hasse and Rink, 2007; Kehl-Fie and Skaar 2010; Counago et al., 2012).

In prokaryotes, one of the important roles of divalent metal ions is to stimulate the uptake of citrate molecules. Being ubiquitously present, citrate is the primary metabolite used as a source of carbon and energy under both aerobic and anaerobic conditions. Uptake of citrate in Gram-positive bacteria occurs via the secondary transporters present in their cytoplasmic membrane and are electrochemically driven. Earlier studies have stated that the presence of citrate (secondary) transporters are only predicted, are not reported in Gram-negative bacteria (Lensbouer et al., 2008). However, recent studies have reported the presence of CitS Na^+-dependent citrate symporters in Gram-negative *Salmonella enterica* and *Klebsiella pneumoniae* (Wohlert et al., 2015; Kim et al., 2017). Interestingly, the presence of divalent metal ion stimulated secondary citrate transporters are believed to be absent in Gram-negative bacteria. In contrast, the presence of primary active transporters involved in ferric citrate (FecBCD) transport in Gram-negative bacterium *E. coli* is well documented (Braun and Herrmann, 2007; Banerjee et al., 2016). Moreover, the presence of FecBCD homologs have also been identified in other microorganisms as well, which basically belong to the superfamily of ATP-binding cassette (ABC) transporters.

ABC transporters are one of the largest superfamilies of proteins involved in the efficient transport of a diverse array of metal ions and their complexes across plasma membranes (Higgins, 1992). The transport of metal ions or any other substrate molecules by ABC transporters is fueled by ATP hydrolysis. ABC transporters are broadly classified as ABC exporters and importers based on the directionality of substrate transport (Wilkens, 2015; Szollosi et al., 2018). ABC exporters efflux out toxins, drug molecules, or other substrates to the extracellular region and are ubiquitous to all domains of life. In contrast, ABC importers influx the essential micronutrients inside the cytoplasm and are reported to be present exclusively in prokaryotes and in some plants (Davidson et al., 2008; Lefevre and Boutry, 2018). Architecturally, both ABC exporters and importers are canonically comprised of transmembrane and nucleotide-binding domains (TMDs and NBDs). TMDs are the substrate translocators

providing passage to substrates, whereas NBDs are the energy generators where ATP hydrolysis occurs to fuel the transport. In addition to the TMDs and NBDs, ABC importers require an additional protein termed as substrate (or solute)-binding proteins (SBPs) (van der Heide and Poolman, 2002; Marinelli et al., 2011; Scheepers et al., 2016). SBPs are soluble proteins and are located in the periplasm of Gram-negative bacteria and, therefore, sometimes referred to as periplasmic-binding proteins (PBPs). On the other hand, the SBPs are anchored to the cytoplasmic membrane via a lipid moiety located near its cognate TMD subunit in Gram-positive bacteria and archaea (Tanaka et al., 2018). The SBPs recognize and capture their physiological ligands from the extracellular environment and deliver it to their respective TMD subunits for further translocation of the substrates inside the cell (Davidson et al., 1992). Since SBPs are involved in maintaining the selectivity and specificity of the substrate(s) and directionality of the transport, they have been reported to be pivotal for prokaryotic ABC transport systems for substrate transport. Although SBPs have low sequence similarity among themselves, they share a common architecture and are comprised of two α/β globular domains where the β-sheets are flanked by α-helices. These two domains are called the N- and C-terminal domains (NTD and CTD), which are linked by a flexible hinge region forming the ligand or substrate-binding pocket. The hinge region facilitates rotational movement to these domains for substrate binding via the popular ‘‘Venus Fly-trap’’ mechanism (Mao et al., 1982). SBPs have been categorized into seven different clusters, viz clusters A-G, based on their topology and type of substrates they bind and transport (Berntsson et al., 2010; Scheepers et al., 2016). SBPs involved in metal ion uptake either in their ionic or complex forms are classified in clusters A and D, which are further subclassified into subcluster A-I and A-II and D-I to D-IV. The subclusters A-I and A-II are involved in the uptake of divalent metal ions and metal ion complexes, respectively, whereas the subclusters D-III and D-IV are reported to be involved in metal oxyanions and iron transport (Berntsson et al., 2010; Scheepers et al., 2016).

The requirement of metal ion uptake transporters varies with microbial niches as it plays a crucial role in the accumulation and utilization of diverse metal ions for several metabolic processes inside the cells. *Thermus thermophilus* is one such profoundly studied Gram-negative thermophilic bacterium that resides in marine hot springs that

can grow at higher temperatures of ~50 to 82°C (Alarico et al., 2005; Ohtani et al., 2012). Previously, in our *in silico* analysis, we have reported the presence of a Mg^{2+}/Fe^{3+} complexed citrate transport system in *Thermus thermophilus* HB8 (Mandal et al., 2019). In this study, we confirm the presence of a novel divalent metal ion stimulated citrate binding protein (ORF id: TTHB177; hereafter referred to as **M**etal ion **c**itrate **t**ransporter MctA), which belongs to the superfamily of ABC transporters. The structural and functional characterization of the protein MctA suggests that it is a novel protein present only in Gram-negative bacteria and is capable of binding and transporting Mg^{2+}-complexed citrate molecule.

5.2. MATERIALS AND METHODS

5.2.1. Construction of wild-type and mutant expression plasmids

The full-length *mctA* gene (1077 bp) construct cloned in pET11a vector coding for a divalent metal ion complexed citrate binding protein was procured from Biological Research Center, NITE (NBRC), Japan. For recombinant protein using immobilized metal affinity chromatography, further sub-cloning of the gene *mctA* in pET22b(+) vector (Novagen, India) was performed. For sub-cloning, a truncated form of the gene *mctA* (1041 bp, including 18 bp for 6xHis-tag and lacking the signal sequence of 1–54 bp) was first amplified using the cloned pET11a vector as a template. The C-terminus of the protein MctA was tagged with a 6xHis-tag and the oligonucleotide sequences are mentioned in Table 5.1. The amplified product was double digested using *Nde*I and *Bam*HI restriction enzymes and inserted into a pET22b(+) vector excised by the same restriction enzymes. The clones were confirmed by double digestion using the restriction enzymes by incubating the vectors at 37°C for two hours (Figure 5.1A). The resulting wild-type construct was further used as a template to generate various mutants of the protein MctA using Q5 Site-Directed Mutagenesis Kit (New England Biolabs, USA) and the oligonucleotide sequences mentioned in Table 5.1. The mutant clones were confirmed by double digestion using the restriction enzymes by incubating the vectors at 37°C for two hours (Figure 5.1B-J) followed by DNA sequencing.

Table 5.1. List of oligonucleotide sequences used for *mctA_WT and mutants* recombinant constructs. In *mctA_WT* construct, restriction enzyme recognition site and codon encoding for 6-His-tag are underlined and shown in bold, respectively. Similarly,

in mutant constructs, codon encoding for a substituted amino acid is shown in bold and underlined.

Primer	Oligonucleotide sequence (5′-3′)
mctA_WT_F	ATATCATATGATGCAGCAGTCCCGGCCTGCCTCG GAC
mctA_WT_R	ATATGGATCCTTA**ATGATGATGATGATGATG**CT GGCCCCGGACGGC
mctA_S26A_F	CATCTACTCC**GCC**ACGGATCAATC
mctA_S26A_R	ATGAGGCGCCCTTCTTTC
mctA_D28A_F	TCCTCCACG**GCT**CAATCCAGC
mctA_D28A_R	GTAGATGATGAGGCGCCC
mctA_T55A_F	TGATCTGGGT**GCC**CAAGCCAT
mctA_T55A_R	TTGTATTCAATCTGTATAAAGGGATAAAGTTTG
mctA_S79A_F	TCTCTGGTCC**GCG**GCCATGGAGCTCCAGG
mctA_S79A_R	AGGTCTGCGCTGGAGGCT
mctA_S164A_F	CCCCGAGCGG**GCC**GCCGTGGGC
mctA_S164A_R	TCCCAGGTGGCCACCCGC
mctA_T199A_F	CTACTCCTCC**GCC**GGGGCCGC
mctA_T199A_R	AGGGCGGCCTGGGCTTTG
mctA_Y221A_F	CTTCGGCTCC**GCG**GCCCTCCTCCGG
mctA_Y221A_R	AAGCCGTAGGCCAGGTAG
mctA_Y221F_F	TTCGGCTCC**TTC**GCCCTCCTC
mctA_Y221F_R	GAAGCCGTAGGCCAGGTAG
mctA_R246A_F	GGCCATCCAG**GCG**GTGGCCTTCATCAACAAG
mctA_R246A_R	ACGGTGCCGTCGGTGAGG

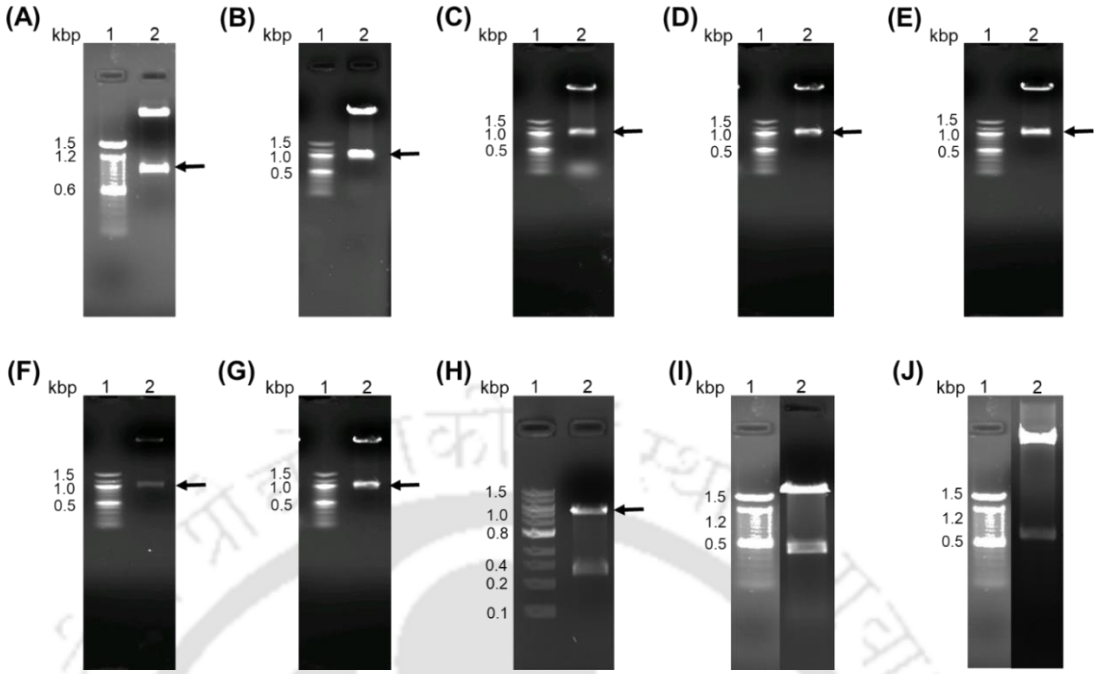


Figure 5.1. Cloning and site-directed mutagenesis of the gene *mctA*. (A) Clone confirmation of *mctA_WT* by double digestion of the vector pET22b carrying the gene *mctA* by *Nde*I and *Bam*HI restriction enzymes (lane 1: DNA ladder, lane 2: insert pop out after double digestion of the plasmid pET22b). Clone confirmation gel of all the *mctA* mutant constructs, (B) *mctA_S26A*, (C) *mctA_D28A*, (D) *mctA_T55A*, (E) *mctA_S79A*, (F) *mctA_S164A*, (G) *mctA_T199A*, (H) *mctA_Y221A*, (I) *mctA_Y221F* and (J) *mctA_R246A*. Mutagenesis confirmation of all the *mctA* mutant constructs was done by plasmid DNA sequencing.

5.2.2. Protein over expression and purification

The *Escherichia coli* BL21 (DE3) competent cells (Novagen) were transformed with the recombinant constructs of the wild-type (MctA_WT) and mutants (MctA_S26A, MctA_D28A, MctA_T55A, MctA_S79A, MctA_S164A, MctA_T199A, MctA_Y221A, MctA_Y221F and MctA_R246A) for the over expression. The transformed cells were grown at 37°C in 100 mg ml^{-1} ampicillin supplemented Luria-Bertani (LB) medium till an optical density (OD) of 0.6–0.8 at 600 nm. Subsequently, the cells were induced with 0.5 mM of Isopropyl ß-D-1-thiogalactopyranoside (IPTG) and further incubated at 37°C for 3.5 h. Following this, the overexpressed cells were harvested by centrifugation at 3836g for 15 min and resuspended in the lysis buffer A

containing 20 mM Tris–HCl (pH 7.5), 5 mM imidazole, 150 mM NaCl, 5% glycerol, 3 mM β-mercaptoethanol (β-ME) and 1 mM phenylmethylsulfonyl fluoride (PMSF). The resuspended cells were incubated on ice for 1 h and then disrupted using a sonicator. As an initial purification step, the obtained cell lysate was incubated in a water bath at 70°C for 10 min followed by centrifugation at 18,514g for 1 h at 4°C to remove the insoluble debris. The supernatant fraction was applied to a pierce centrifuge column (Thermo Fisher Scientific, USA) packed with Ni-nitrilotriacetic acid (Ni-NTA) affinity resin (Qiagen, Hilden, Germany) pre-equilibrated with buffer A. After 3 h of incubation at 4°C, the column was gradiently washed with 10 column volumes each of buffer B containing 20 mM Tris–HCl (pH 7.5), 10 mM imidazole, 150 mM NaCl, 5% glycerol, 1 mM PMSF and 3 mM β-ME and buffer C having 20 mM Tris–HCl (pH 7.5), 20 mM imidazole, 150 mM NaCl, 5% glycerol, 1 mM PMSF and 3 mM β-ME. The recombinant protein(s) bound to the Ni-NTA resin was then eluted with buffer D containing 20 mM Tris–HCl (pH 7.5), 250 mM imidazole, 150 mM NaCl, 5% glycerol and 1 mM PMSF. The eluted fractions were pooled together and subjected to gradient dialysis to remove imidazole against the final protein buffer E containing 20 mM Tris-HCl (pH 7.5), 150 mM NaCl and 5% glycerol. The dialyzed proteins MctA_WT, MctA_S26A, MctA_D28A, MctA_T55A, MctA_S79A, MctA_S164A, MctA_T199A, MctA_Y221A, MctA_Y221F and MctA_R246A were concentrated in a Vivaspin turbo 15 centricon (10 kDa cutoff; Sartorius, Göttingen, Germany) up to a concentration of ~14, ~26, ~18, ~23, ~10, ~15, ~13, ~18, ~19 and ~10 mg ml^{-1}, respectively (Figure 5.2A-J). For metal-free protein, a similar approach of protein purification was followed for MctA_WT and mutants except that at the time of dialysis, the dialysis buffer was supplemented with 10 mM of EDTA in two consecutive steps. However, the last two steps of dialysis were performed in buffer E in the absence of EDTA. The EDTA-dialyzed proteins MctA_WT, MctA_S26A, MctA_D28A, MctA_T55A, MctA_S79A, MctA_S164A, MctA_T199A, MctA_Y221A, MctA_Y221F and MctA_R246A were concentrated in a Vivaspin turbo 15 centricon (10 kDa cutoff; Sartorius, Germany) up to a concentration of ~6, ~7, ~8, ~6, ~5, ~6, ~5, ~5, ~7 and ~5 mg ml^{-1}, respectively. SDS-PAGE analysis was performed after each step of protein purification to check the purity of the protein (Figure 5.2K-T). The protein concentration was calculated using the theoretical extinction coefficient (ε_{280}= 48820 M^{-1} cm^{-1}) after measuring the absorbance of protein(s) at 280 nm.

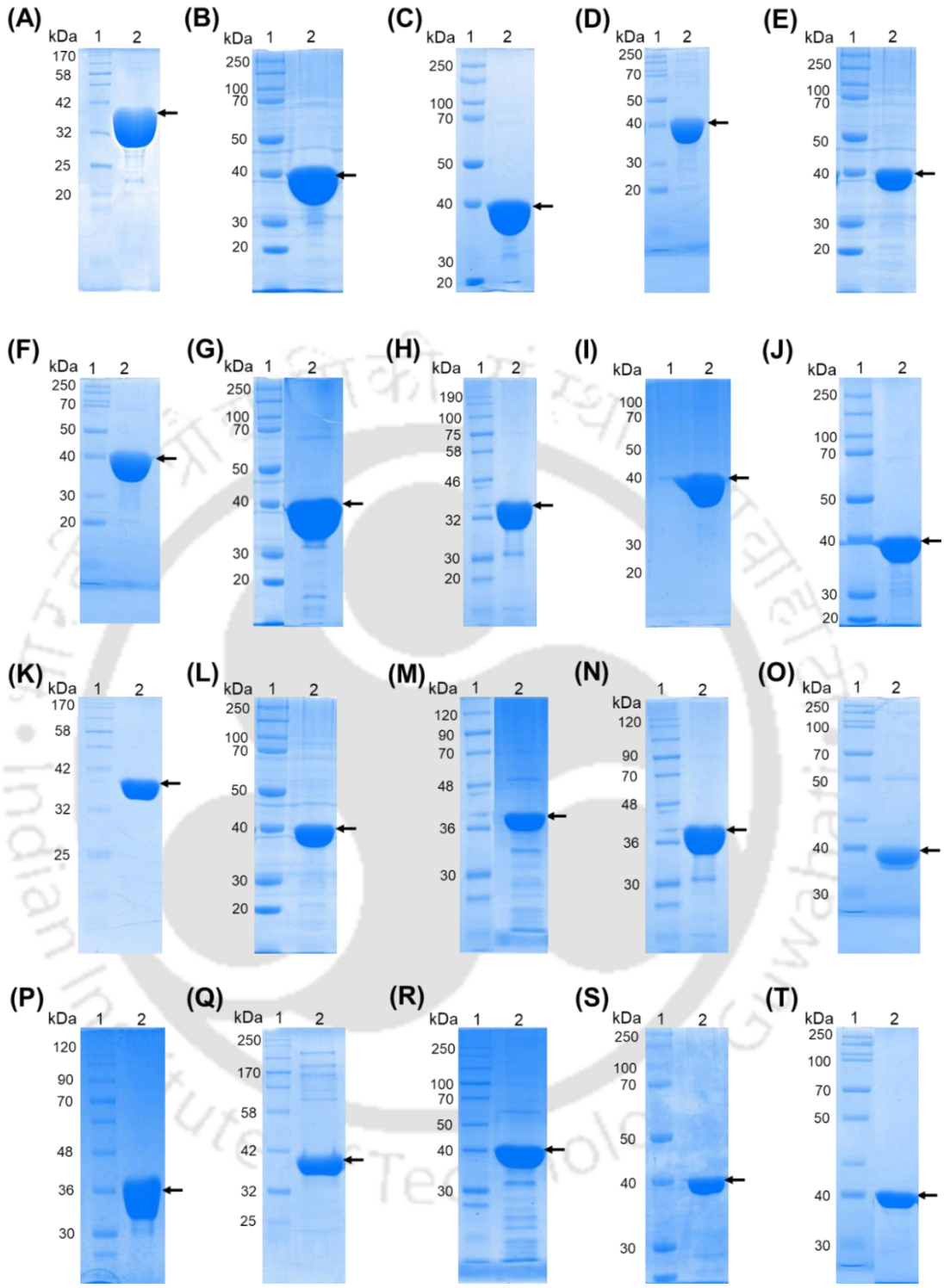


Figure 5.2. Purification of MctA_WT and mutant proteins. SDS-PAGE analysis of the purified and concentrated proteins (A) MctA_WT, (B) MctA_S26A, (C) MctA_D28A, (D) MctA_T55A, (E) MctA_S79A, (F) MctA_S164A, (G) MctA_T199A, (H) MctA_Y221A, (I) MctA_Y221F and (J) MctA_R246A (lane 1: protein ladder, lane 2: concentrated protein). SDS-PAGE analysis of the purified and

concentrated proteins of (K) MctA_WT, (L) MctA_S26A, (M) MctA_D28A, (N) MctA_T55A, (O) MctA_S79A, (P) MctA_S164A, (Q) MctA_T199A, (R) MctA_Y221A, (S) MctA_Y221F and (T) MctA_R246A (lane 1: protein ladder, lane 2: concentrated protein) after dialysis with 10 mM of EDTA to remove the endogenously bound metal ion.

5.2.3. Mass spectrometric analysis

The exact molecular weight of the protein MctA _WT was determined using matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF). For MALDI_TOF analysis, the homogenously purified MctA _WT protein was mixed with the matrix in a (protein:matrix) ratio of 1:2, 1:3 and 1:5. Since, the theoretical molecular weight of the protein MctA is more than 10 kDa, sinapinic acid was prepared in acetonitrile and 0.1% trifluoroacetic acid in a ratio of 30:70, respectively, to a concentration of 10 mg ml^{-1}. Sinapinic acid was dissolved in the mentioned solvent by sonicating the solution for 30 mins followed by centrifugation at 12000g for 10 mins, to remove the excess crystals formed by the matrix. Further, the prepared matrix was mixed with protein samples in varying concentrations of 1, 5 and 10 mg ml^{-1} in the ratios specified above. 2 µL spots of the MctA _WT protein-matrix mixture was made on the ground steel MALDI plate grooves and air-dried. Post drying, the drops were ionized in a Bruker autoflex speed and the data were plotted using an in-built Flex control analysis software.

5.2.4. Crystallization of the wild-type and mutant MctA proteins

Preliminary crystallization screening of the purified MctA _WT protein (~8 mg ml^{-1}) was performed using crystal screen and PEG/Ion kits (Hampton Research, CA, USA) by mixing 2 µl each of protein and buffer (1:1 ratio) using microbatch-under-oil technique at 4 and 20°C. Initial crystal hits for the protein MctA _WT were obtained in a buffer containing 0.2 M ammonium fluoride and 20% PEG 3350 at 20°C. The obtained crystal hits were reproduced in microbatch-under-oil as well as optimized using hanging-drop vapor-diffusion method (Figure 5.3A). The reproduced crystal at 20°C in microbatch obtained after ~four weeks was diffracted in a home source X-ray diffractometer and X-ray intensity data was collected. With a different batch of purified

protein (14 mg ml^{-1}), another crystal was obtained in microbatch-under-oil technique at 20°C in a buffer having 0.1 M sodium citrate tribasic dihydrate pH 5.5 and 16% PEG 8000 (Figure 5.3B). The obtained crystal was also diffracted and the X-ray data was collected. Moreover, the EDTA-treated MctA _WT protein was also screened for crystallization using the crystal screen and PEG/Ion (Hampton Research, USA) and JCSG Plus and Wizard Cryo (Molecular Dimensions, UK) screens. A crystal appeared after ~four weeks of incubation in crystallization buffer having 0.2 M ammonium sulfate, 0.1 M MES monohydrate pH 6.5, 30% PEG monoethyl ether 5000 at 20°C, which was diffracted and the X-ray data for the same was collected.

In parallel, co-crystallization experiments for the EDTA-treated MctA _WT protein with various divalent metal ions were performed using multiple approaches. In the first method, the protein was incubated with varying concentrations of different divalent metal ions (10 to 100X) prepared in buffer E for 3-4 h prior to setting up crystal drops. In the second approach, the prepared ligands were directly added at the time of setting up crystal drop at varying concentrations. Initial co-crystallization experiments were performed in a hanging-drop vapor-diffusion method using the original crystallization buffer in which the crystal of MctA _WT was obtained. Upon failure, crystal screen and PEG/Ion kits (Hampton Research, USA) were used for buffer screening to pre-incubated divalent metal ion(s) and TTHB177_WT protein. Additionally, JCSG Plus and Wizard Cryo screens (Molecular Dimensions, UK) were also screened for co-crystallization experiments with several divalent metal ions. However, none of the co-crystallized drops produced any crystal.

For crystallization of the mutant proteins of MctA, initially, the crystallization experiments were performed in hanging-drop vapor-diffusion method using the original crystallization buffer in which crystal of MctA_WT was obtained. Since no crystal hits were obtained, crystal screen, PEG/Ion (Hampton Research, USA), JCSG Plus and Wizard Cryo screens (Molecular Dimensions, UK) were exploited for screening using microbatch-under-oil technique at 4 and 20°C. Crystals appeared after six to twelve weeks in different conditions for the MctA mutant proteins list of which is provided in Table C1 (Figure 5.3C-J). Unfortunately, no crystal hits were obtained for MctA _T55A and MctA _R246A mutants in any crystallization trials. Simultaneously, co-

crystallization experiments with EDTA-treated protein batches of the MctA mutants were performed in their originally obtained crystals in hanging-drop vapor diffusion as well as in microbatch-under-oil methods. Upon failure, all the available crystallization screens in our lab were screened for co-crystallization of the MctA mutant proteins with divalent metal ions. Although the co-crystallization experiments were unsuccessful, only MctA _S26A mutant protein was able to form a crystal with $MnSO_4$ salts in a buffer having 0.2 M ammonium fluoride and 20% PEG 3350 at 20°C, which was diffracted and the X-ray data were collected (Figure 5.3D).

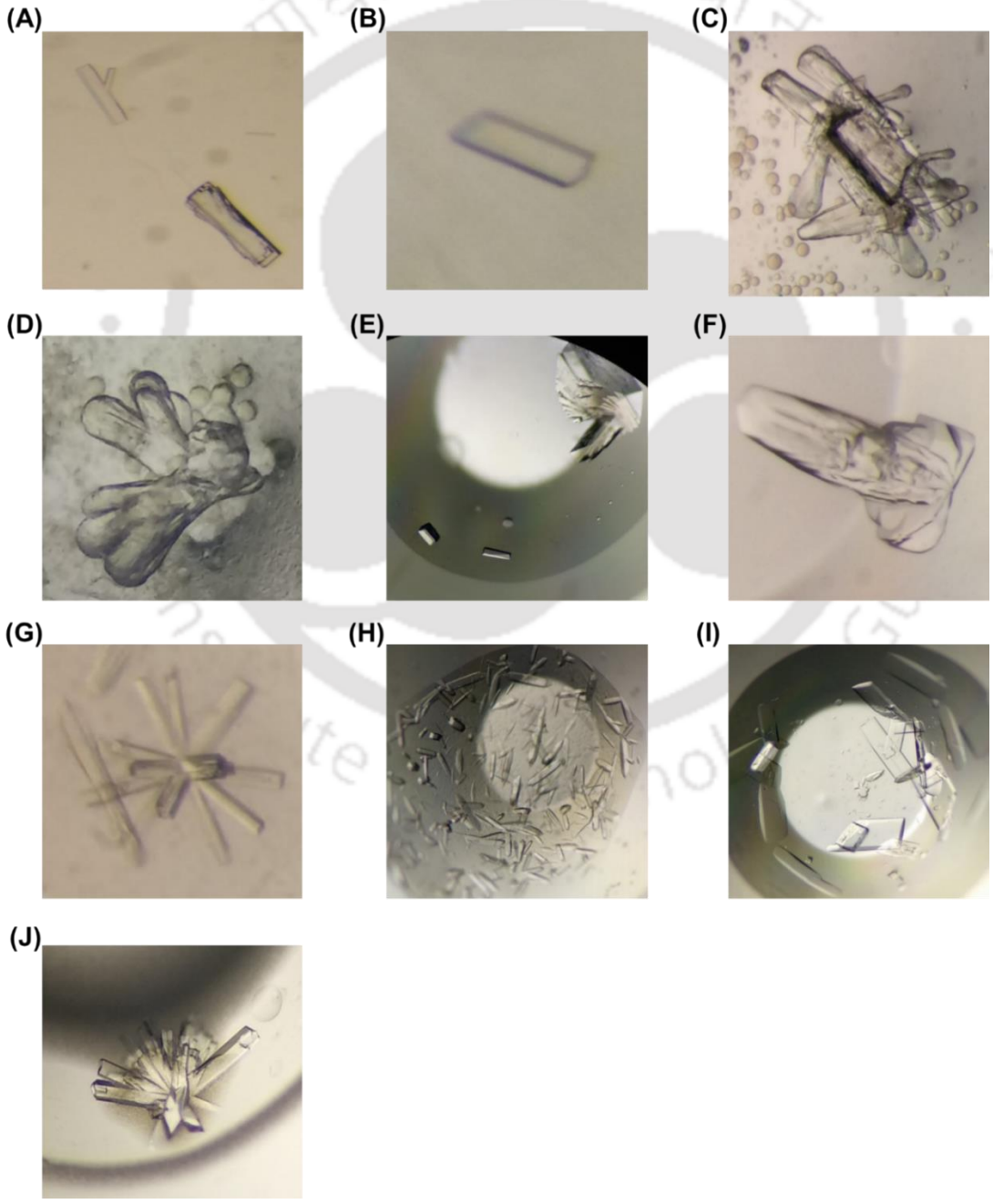


Figure 5.3. Crystallization of the protein MctA**.** (A, B) Protein crystals of MctA_WT in microbatch. (C) Protein crystals of MctA _S26A. (D) Protein co-crystals of MctA _S26A with $MnSO_4$. Protein crystals of (E) MctA _D28A (F) MctA _S79A, (G) MctA _S164A, (H) MctA _T199A, (I) MctA _Y221A and (J) MctA _Y221F.

5.2.5. Data collection, processing and structure determination

X-ray intensity diffraction data for the MctA _WT and mutant protein crystals were collected at -173°C using the home source Rigaku MicroMax-007 HF diffractometer (operated at 40 kV and 30 mA) and R-Axis IV++ image-plate detector available at the Central Instrument Facility (CIF) of Indian Institute of Technology Guwahati, India. The X-ray data sets for all the crystals were collected with 1° oscillation and 300 s exposure time while keeping the detector at a distance range of 150-200 mm. Processing and scaling of all the X-ray data sets were performed using the programs iMosflm (Battye et al., 2011) and Aimless (Evans and Murshudov, 2013) and the intensities were converted to structure factors using the module ctruncate embedded in the CCP4 package (Evans, 2011). The structure solution of MctA _WT protein was determined by molecular replacement method using the program Phaser (McCoy et al., 2007). The crystal structure of an Mg^{2+}-citrate binding protein from *Agrobacterium fabrum* (PDB id: 3C9H), sharing a sequence identity (query coverage) of 37% (89%) was used as a search model. The structure solution of MctA _WT, MctA _S26A, MctA _SD28A, MctA _S79A, MctA _S164A, MctA _T199A, MctA _Y221A and MctA _Y221F proteins were also determined by molecular replacement with the program Phaser using the three-dimensional atomic coordinates of MctA_WT as a search model. A total of 5% of the reflections were kept aside from the data set for R_{free} calculations (Kleywegt and Jones, 1997). For all the models, the program Coot (Emsley et al., 2010) was used for model building, followed by structural refinement of the models using the program *REFMAC5* (Murshudov et al., 2011) with default parameters. Although a clear *Fo−Fc* map was observed for the bound ligands after the first cycle of refinement for all the complexed structures, firstly, the protein atoms were fitted into an electron density countered at 3.0σ and 1.0σ for the *Fo−Fc* and *2Fo−Fc* maps, respectively. Subsequently, the electron densities observed for water and other molecules present in the protein and crystallization buffer were modeled. Structure refinement was

performed with each cycle of the model building using the same set of parameters. The geometrical parameters for all the refined structures were checked and validated by the program PROCHECK (Laskowski et al., 1993) and MolProbity (Williams et al., 2018). The refinement and validation statistics for all the refined models are provided in Table 5.2 - 5.5. The three-dimensional atomic factors and structure factors of all the refined structures are deposited in RCSB Protein Data Bank. The molecular graphic figures used in this study are prepared using the program PyMol (Molecular Graphics System, Version 2.1.1 Schrödinger, LLC).

Table 5.2. Data collection and refinement statistics of MctA_WT (bound to Mg^{2+}-citrate in $P2_1$ and $P2_12_12_1$ space groups and citrate only). The values in parenthesis are for the last resolution shell.

Parameters	**MctA_WT•Mg•CIT _FormI**	**MctA_WT•Mg•CIT_ FormII**	**MctA_WT•CIT**
Wavelength (Å)	1.5418	1.5418	1.5418
Temperature (K)	100	100	100
Space group	$P2_1$	$P2_12_12_1$	$P2_1$
Unit-cell parameters (Å, °)	a=45.53, b=144.99, c=79.95, α=γ=90, β=100.05	a=44.65, b=100.87, c=143.66, α= β=γ=90	a=45.55, b=145.11, c=79.24, α =γ=90, β=100.44
Resolution (Å)	52.93-1.77 (1.80-1.77)	71.83-2.10 (2.16-2.10)	77.93-2.10 (2.16-2.10)
No. of observed reflections	486645 (21491)	225737 (17002)	213306 (15755)
No. of unique reflections	92348 (4278)	38780 (3095)	58898 (4570)
Mn(I) CC(1/2)	0.999 (0.879)	0.994 (0.881)	0.994 (0.833)

Completeness (%)	95.8 (88.9)	99.8 (99.2)	100.0 (99.7)
V_M ($Å^3$ Da^{-1})	2.23	2.15	2.29
Solvent content (%)	44.98	42.91	46.40
Mosaicity (°)	0.30	0.40	0.60
Mean I/σ(I)	15.2 (3.3)	10.1 (3.1)	9.2 (2.8)
R_{merge}† (%)	6.0 (37.2)	11.0 (47.5)	8.8 (36.6)
R_{pim} (%)	4.2 (27.1)	7.3 (32.4)	8.1 (32.6)
R_{meas} (%)	7.4 (46.3)	13.2 (57.7)	12.0 (49.2)
Multiplicity	5.3 (5.0)	5.8 (5.5)	3.6 (3.4)
R_{work}/R_{free} (%)	13.69/17.15	16.52/21.63	17.93/21.29
Protein model			
No. of subunits in ASU	3	2	3
Protein atoms	8095	5325	7970
Water molecules	716	318	480
Citrate	3	2	3
Metal ions (Mg^{2+})	3	2	-
Other molecules	37	8	16
Deviation from ideal geometry			
Bond length (Å)	0.014	0.014	0.013
Bond angles (°)	1.811	1.822	1.680
Average *B*-factor ($Å^2$)			
Protein atoms	13.76	18.54	23.55
Water molecules	32.17	34.01	34.20
Citrate	16.21	21.97	32.5
Metal ion (Mg^{2+})	18.22	27.58	-

Ramachandran plot (%)			
Favored	97.39	96.85	96.75
Allowed	2.61	3.15	3.25
PDB id	**7F6E**	**7F6F**	**7F6K**

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

Table 5.3. Data collection and refinement statistics of MctA_S26A (bound to Mg^{2+}-citrate and Mn^{2+}-citrate) and MctA_D28A mutant. The values in parenthesis are for the last resolution shell.

Parameters	**MctA_S26A•Mg•CIT**	**MctA_S26A•Mn•CIT**	**MctA_D28A•CIT**
Wavelength (Å)	1.5418	1.5418	1.5418
Temperature (K)	100	100	100
Space group	$P2_1$	$P2_12_12_1$	$P2_1$
Unit-cell parameters (Å, °)	a=44.51, b=142.92, c=79.27, α =γ=90, β=100.51	a=44.73, b=100.09, c=143.30, α= β=γ=90	a=45.31, b=146.26, c=79.09, α =γ=90, β=99.89
Resolution (Å)	77.94-2.30 (2.38-2.30)	71.65-2.50 (2.60-2.50)	73.13-2.15 (2.21-2.15)
No. of observed reflections	194948 (19012)	168215 (17862)	290017 (21756)
No. of unique reflections	43237 (4260)	23063 (2526)	52451 (4081)
Mn(I) CC(1/2)	0.974 (0.770)	0.971 (0.862)	0.997 (0.851)
Completeness (%)	100.0 (100)	99.9 (99.7)	95.5 (91.4)
V_M ($Å^3$ Da^{-1})	2.21	2.14	2.30

Solvent content (%)	44.34	42.68	46.59
Mosaicity (°)	0.60	0.60	0.45
Mean I/σ(I)	7.7 (2.9)	6.4 (3.0)	12.7 (3.1)
R_{merge}† (%)	15.4 (44.3)	21.8 (52.4)	8.4 (55.9)
R_{pim} (%)	12.9 (37.4)	12.0 (32.6)	5.8 (38.5)
R_{meas} (%)	20.2 (58.4)	25.4 (61.5)	10.2 (68.2)
Multiplicity	4.5 (4.5)	7.3 (7.1)	5.5 (5.3)
R_{work}/R_{free} (%)	19.45/24.23	19.40/26.52	17.41/21.46
Protein model			
No. of subunits in ASU	3	2	3
Protein atoms	7960	5316	7964
Water molecules	307	182	451
Citrate (number)	3	2	3
Metal ions	Mg^{2+} (3)	Mn^{2+} (2)	Mg^{2+} (1)
Other molecules	9	11	7
Deviation from ideal geometry			
Bond length (Å)	0.011	0.012	0.013
Bond angles (°)	1.663	1.782	1.698
Average *B*-factor (Å^2)			
Protein atoms	18.03	11.39	23.70
Water molecules	26.18	18.38	37.25
Citrate	22.71	27.83	45.10
Metal ion	Mg^{2+} (28.89)	Mn^{2+} (57.51)	Mg^{2+} (41.16)
Ramachandran plot (%)			
Favored	95.90	95.94	96.99
Allowed	4.10	4.06	3.01
PDB id	**7F6N**	**7F6O**	**7F6P**

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

Table 5.4. Data collection and refinement statistics of MctA_S79A, MctA_S164A and MctA_T199A mutant. The values in parenthesis are for the last resolution shell.

Parameters	MctA_S79A	MctA_S164A	MctA_T199A
Wavelength (Å)	1.5418	1.5418	1.5418
Temperature (K)	100	100	100
Space group	$P3_121$	$I222$	$I222$
Unit-cell parameters (Å, °)	$a=b=72.17$, $c=114.30$, $\alpha=\beta=90$, $\gamma=120$	$a=44.53$, $b=108.7$, $c=163.99$, $\alpha=\beta=\gamma=90$	$a=44.29$, $b=108.85$, $c=163.57$, $\alpha=\beta=\gamma=90$
Resolution (Å)	42.18-1.63 (1.66-1.63)	81.99-2.09 (2.15-2.09)	48.75-1.80 (1.84-1.80)
No. of observed reflections	456622 (20414)	241478 (17371)	358806 (20436)
No. of unique reflections	43745 (2113)	24055 (1838)	37221 (2179)
Mn(I) CC(1/2)	0.999 (0.979)	0.999 (0.929)	0.999 (0.941)
Completeness (%)	100.0 (99.5)	100 (100)	100 (100)
V_M ($Å^3$ Da^{-1})	2.28	2.63	2.61
Solvent content (%)	46.10	53.21	52.87
Mosaicity (°)	0.45	0.40	0.464
Mean I/σ(I)	24.7 (5.1)	19.5 (5.4)	19.0 (3.8)
R_{merge}† (%)	5.2 (21.9)	9.4 (42.3)	6.6 (45.8)
R_{pim} (%)	2.4 (10.7)	4.5 (21.1)	3.2 (23.0)
R_{meas} (%)	5.8 (24.5)	10.4 (47.4)	7.4 (51.3)
Multiplicity	10.4 (9.7)	10.0 (9.5)	9.6 (9.4)

R_{work}/R_{free} (%)	17.55/21.32	15.59/21.41	15.13/18.16
Protein model			
No. of subunits in ASU	1	1	1
Protein atoms	2726	2665	2701
Water molecules	355	219	236
Other molecules	14	12	9
Deviation from ideal geometry			
Bond length (Å)	0.018	0.016	0.017
Bond angles (°)	2.218	1.948	2.145
Average *B*-factor (Å^2)			
Protein atoms	15.31	12.54	17.77
Water molecules	34.43	35.76	38.35
Ramachandran plot (%)			
Favored	97.21	98.20	98.49
Allowed	2.79	1.80	1.51
PDB id	**7F6Q**	**7F6R**	**7F6S**

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

Table 5.5. Data collection and refinement statistics of MctA_Y221A and MctA_Y221F mutant. The values in parenthesis are for the last resolution shell.

Parameters	MctA_Y221A	MctA_Y221F
Wavelength (Å)	1.5418	1.5418
Temperature (K)	100	100
Space group	*I*222	*I*222
Unit-cell parameters (Å, °)	a=44.42, b=108.61, c=163.17, $\alpha=\beta=\gamma$=90	a=44.70, b=108.37, c=163.41, $\alpha=\beta=\gamma$=90

Resolution (Å)	54.30-1.75 (1.78-1.75)	54.18-1.90 (1.94-1.90)
No. of observed reflections	324728 (16706)	180548 (11373)
No. of unique reflections	40385 (2190)	31862 (2042)
Mn(I) CC(1/2)	0.997 (0.907)	0.999 (0.867)
Completeness (%)	100.0 (100)	100 (100)
V_M ($Å^3$ Da^{-1})	2.61	2.64
Solvent content (%)	52.87	53.35
Mosaicity (°)	0.60	0.394
Mean I/σ(I)	13.2 (2.5)	18.0 (3.4)
R_{merge}† (%)	8.4 (46.8)	6.4 (45.6)
R_{pim} (%)	4.5 (27.0)	4.5 (31.9)
R_{meas} (%)	9.6 (54.1)	7.8 (55.8)
Multiplicity	8.0 (7.6)	5.7 (5.6)
R_{work}/R_{free} (%)	20.39/23.62	17.26/20.73
Protein model		
No. of subunits in ASU	1	1
Protein atoms	2694	2683
Water molecules	281	249
Other molecules	12	10
Deviation from ideal geometry		
Bond length (Å)	0.020	0.016
Bond angles (°)	2.134	2.051
Average *B*-factor ($Å^2$)		
Protein atoms	15.57	11.94
Water molecules	37.16	33.54
Ramachandran plot (%)		
Favored	96.69	97.58
Allowed	3.31	2.42
PDB id	**7F6U**	**7F6T**

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

5.2.6. Energy Dispersive X-ray analysis

The Energy Dispersive X-ray (EDX) microanalysis of the protein MctA_WT and MctA_S26A was performed using 80T EDX detector (Oxford Instruments, Tokyo, Japan) equipped in TEM-Jeol 2100F transmission electron microscopy available at Central Institute Facility (CIF) of Indian Institute of Technology Guwahati, India. The EDX detector is devised with a super atmospheric thin window (SATW) and a 50 mm^2 silicon drift detector (SDD), placed with a takeoff angle of 30°, having a resolution of 129 eV at the Mn K-L3 line (in IUPAC notation; Siegbahn notation Mn $K_{\alpha 1}$) under the conditions used in this study. The EDX detector is monitored by INCA (version 5.05, Oxford Instruments) EDX spectra acquisition and processing. For EDX spectra, protein samples were first carbon-coated using a carbon film 300 mesh procured from Electron Microscopy Sciences (Hatfield, Pennsylvania, USA). The analysis was performed under vacuum at an acceleration voltage of 20 kV, a 10 mm working distance, a 40000x magnification and a beam size of 100 nm.

Under these conditions, probe currents of 0.75−1.5 nA were measured by pointing the beam in a Faraday cup. This allowed spectral acquisition rates of 10000−15000 counts per second (cps) with a dead time of ∼10%. The total acquisition time was automatically set by INCA and usually did not exceed 40 s with total numbers of counts of ∼310000−320000. For EDX spectral processing, INCA uses a filtered least squares (FLS) approach for background removal, fitting, and deconvolution of the peaks, which also includes a pulse pile-up correction. Quantitative data are obtained from the spectra after matrix correction, which follows the XPP exponential model procedure. All of the details of EDX spectral processing software can be found in the INCA Energy Operator Manual. Because of the large amounts of light elements in the samples, we did not seek to obtain full quantitative data but instead focused on the S/In ratio. Using INCA, this was made by indicating that the sample was coated with a 50 nm carbon layer, selecting all the elements other than sulfur and indium as “deconvolution elements” (which means that their presence is taken into account for the calculation but that they are not

displayed in the final result), and using the "Quant Standardization (Extended Set)" provided with the software. Ratios were calculated with the sulfur K-series peaks and the indium L-series peaks. S/Na ratios were obtained using the same procedure.

5.2.7. Circular dichroism for the wild type and mutant proteins of MctA

The far-UV circular dichroism (CD) spectra of the protein MctA_WT and mutants were recorded between 190-260 nm wavelengths with 1 mm optical path length quartz cuvette using the JASCO J-1500 spectrometer (Jasco, Germany). MctA_WT and mutant proteins were prepared with their respective dialysis buffer E and diluted into milliQ water to a final concentration of 2.5 µM. CD parameters like a response, sensitivity and scan speed were fixed as 2 s, 100 mdeg and 100 nm min^{-1}, respectively, for spectra measurement. Each measurement includes three averaging scans to which the blank containing the same buffer as reference was subtracted. The CD spectra obtained were analyzed using the Spectra Manager software provided by JASCO. The secondary structural changes were scanned between 190 to 260 nm wavelengths. The final spectral profile of MctA_WT and mutant proteins were plotted as a function of ellipticity in mdeg and wavelength using the Origin software (v 9.6).

5.2.8. Isothermal titration calorimetry

The binding affinity of different divalent and trivalent metal ions and sodium citrate with EDTA-treated batches of MctA_WT and mutant proteins were measured using isothermal titration calorimetry (MicroCal ITC200; GE Healthcare, USA) method. The EDTA-treated MctA_WT protein and the ligands used for ITC experiments were prepared in the dialyzed buffer E. 40 µM of MctA_WT protein was filled in the sample cell and titrated against 4 mM of different ligands mentioned in Table 5.6. Each injection consists of 1.5 µl of ligand injected 24 times (excluding 0.4 µl pre-injection) with a duration of 120 s and the sample cell was stirred at 250 rpm. An exactly same protocol was followed for EDTA-treated MctA mutant proteins, MctA_S26A, MctA_SD28A, MctA_T55A, MctA_S79A, MctA_S164A, MctA_T199A, MctA_Y221A, MctA_Y221F and MctA_R246A. The thermodynamic parameters of MctA_WT and mutant proteins were measured at 25°C. A control experiment to measure the heat of dilution was also performed for each ITC experiment by titrating different ligands into the dialyzed buffer. The obtained heat of dilution in each control

experiment was subtracted from their respective protein-ligand titration to dictate the real ligand binding affinity. The integration of the reaction heat was performed using the program Origin (v 7.0) provided with the MicroCal software. The stoichiometric ratio (the number of binding sites, n), association constant (Ka) and the change in enthalpy (ΔH) and entropy (ΔS) were determined by reaction heat integration using the non-linear regression and one-site binding model. The Gibbs free energy change (ΔG) was calculated using the thermodynamic equation:

$$\Delta G = \Delta H - T\Delta S$$

Where T is the reaction temperature. The estimated values for these thermodynamic parameters for each ITC reaction are provided in Table 5.6 and C4.

5.2.9. Bioinformatic analysis

The conserved residues between the proteins MctA and its closest homolog Atu2467 werer identified by performing a pairwise sequence alignment using the program Clustal Omega (Sievers and Higgins, 2014). The alignment was further rendered using the web tool ESPript 3.0 (Gouet et al., 2003). Homologous proteins for MctA were identified by performing BLASTP (Altschul et al., 1990) search against the non-redundant, UniProtKb/SwissProt and PDB databases (O'Leary et al., 2016). The genus name representing the homologous proteins of MctA from different species were listed out to check for their Gram's characteristic. Structural homologs of the protein MctA were identified by using the web server Dali (Holm and Rosenstrom, 2010). The three-dimensional atomic coordinates of each homolog protein were retrieved from the Protein Data Bank (PDB). The root-mean-square deviation (rmsd) among all the selected SBPs belonging to cluster D were computed using a homebuilt python script and the program PyMOL (PyMOL Molecular Graphics System, Schrodinger, LLC). The obtained rmsd distance matrix was used to generate a structure-based distance tree to study the structural relationship among all the selected cluster D SBPs using the programs DendroUPGMA and MEGA7 (Garcia-Vallve et al., 1999; Kumar et al., 2016).

5.3. RESULTS

5.3.1. Overall structure of the protein MctA

The three-dimensional crystal structure of the protein MctA was determined in two space groups $P2_1$ (MctA_WT_FormI) and $P2_12_12_1$ (MctA_WT_FormII), at a resolution of 1.77 and 2.10 Å, respectively. Each asymmetric unit of MctA_WT_FormI and _FormII contains three and two molecules, respectively. The five molecules from both the crystal forms are very similar, indicating an average rmsd of ~0.2 Å for all Cα atoms. Both at the primary and tertiary structural levels, MctA shares the highest homology with the protein Atu2467 from *Agrobacterium fabrum* structure of which is solved in $P2_12_12_1$ space group at a resolution of 1.9 Å. The structure of Atu2467 was deposited in the PDB with id 3C9H. A comparison of the arrangement of the molecules in the asymmetric unit in the proteins Atu2467, MctA_WT_FormI and FormII suggest significant differences in their packing (Figure 5.4A-C). Both the proteins Atu2467 and MctA_WT_FormII solved in $P2_12_12_1$ space group have two protomers in the asymmetric unit. Chains A and B of Atu2467 are facing each other at the active site pocket (Figure 5.4A). In contrast, the two chains are arranged in opposite directions in the protein MctA_WT_FormII (Figure 5.4B). The protein MctA_WT_FormI having three monomeric units in the asymmetric unit, is arranged in a completely different manner (Figure 5.4C).

The overall three-dimensional structure of the protein MctA is comprised of an N- and C-terminal domain (NTD and CTD) linked by two loop regions forming the hinge (Figure 5.4D). Such architecture of SBPs has been proposed for members of cluster D SBPs (Berntsson et al., 2010). Furthermore, in the structure, an extra density other than that of amino acid residues was also observed at the active-site pocket (Figure 5.4D). The extra density was identified as a citrate molecule that was endogenously bound at the active-site pocket. The presence of a citrate molecule along with the protein was also suggested in our mass spectrometric analysis using MALDI-TOF (Figure C1A). The theoretical molecular weight (MW) of the mature protein is ~38750 Da but the obtained MW from MALDI-TOF analysis was 38968 Da. The approximate difference of ~200 Da is the molecular weight of the citrate molecule (189.1 g mol^{-1}). In addition to citrate, a density for a metal ion was also observed close to citrate (Figure 5.4D). Since citrate is a metal chelator; several divalent metal ions were kept at that density

and checked for their *B* factor to which Mg^{2+} ion shows higher stability. To eliminate the doubt that the bound metal ion is Mg^{2+} and no other metal ions at the active-site pocket, we performed Energy Dispersive X-ray (EDX) analysis. The EDX results exhibit that the protein solution is comprised of Mg^{2+} (0.16% of the total weight) in addition to carbon, nitrogen and oxygen (Figure C1B and C1C). Moreover, the template protein used to solve the structure of MctA, i.e. Atu2467 from *Agrobacterium fabrum* was also bound to Mg^{2+}-citrate molecule (PDB id: 3C9H), which corroborates our results. Since the structure of MctA was solved with an endogenously-bound ligand; attempts were made to obtain an apo structure of the protein as well. Dialyzing the purified protein against 10 mM of EDTA allowed us to obtain the structure of a metal ion-free but citrate-bound protein at a resolution of 2.10 Å in $P2_1$ space group (Figure 5.4E). The overall conformation of the metal ion-free citrate bound MctA is comparable to that of the Mg^{2+}-citrate bound MctA protein with an rmsd of 0.12 Å (Figure 5.4F). All the three structures of MctA_WT protein solved in this study show a similar secondary structural conformation comprising 18 α-helices and 11 β-sheets connected by several loop regions (Figure 5.4G).

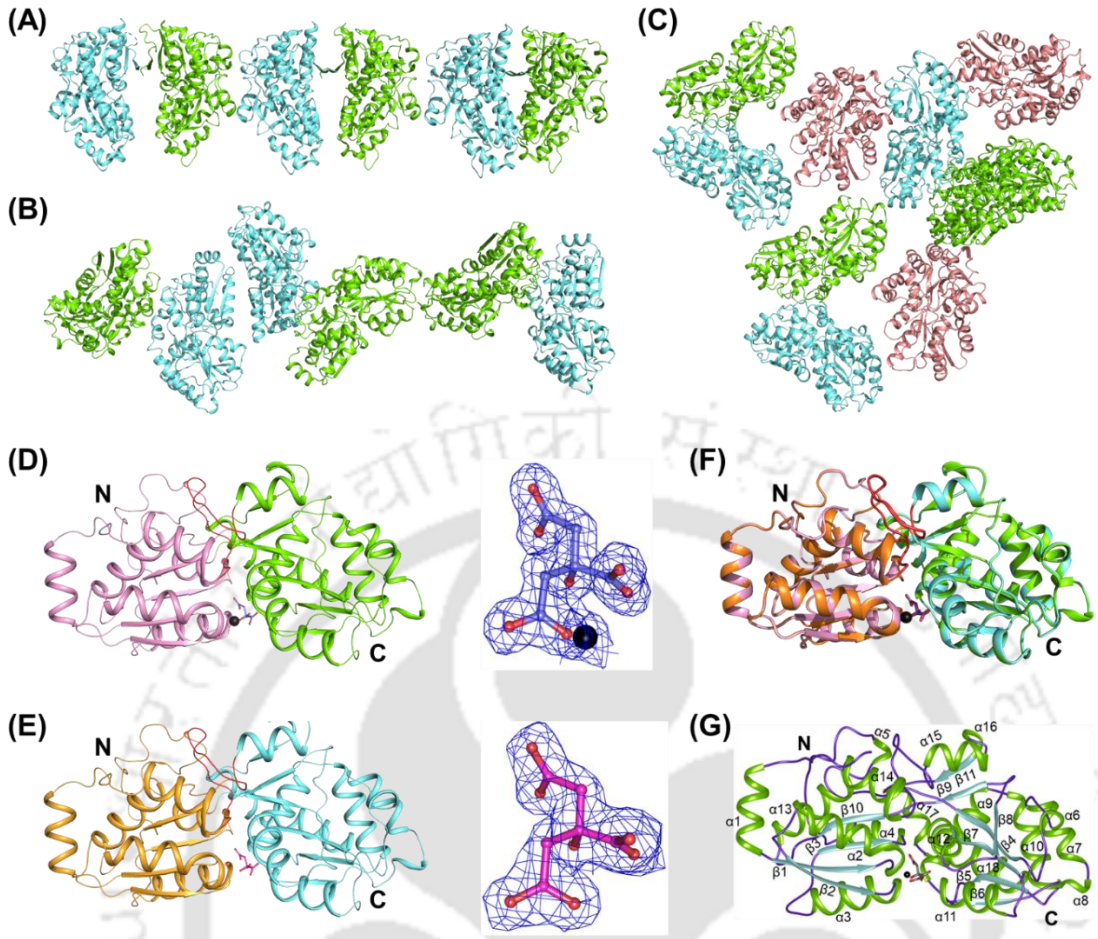


Figure 5.4. Overall three-dimensional structure of the protein MctA. Arrangement of protein molecules in an asymmetric unit in different space groups for the proteins (A) Atu2467 (space group: $P2_12_12_1$), (B) MctA (space group: $P2_12_12_1$) and (C) MctA (space group: $P2_1$). Different chains of the monomers in an asymmetric unit is represented in a different color. Chains A and B are shown in cyan and green, respectively, whereas chain C in the protein MctA ($P2_1$) is shown in magenta. Overall three-dimensional structure of the protein MctA in (D) metal ion-complexed citrate and (E) metal ion free but citrate bound state. The bound citrate is shown in blue (D) and magenta (E) ball-and-stick, whereas the metal ion (Mg^{2+}) is in the black sphere. The $2F_o$-F_c map of the ligand molecules contoured at 1.0 σ is shown adjacent to the protein structures. (F) Superposed image of the metal-bound and free MctA protein indicating no overall conformational change in them. The N- and C-terminal domains of the metal-bound state of MctA protein is shown in pink and green, respectively, whereas metal-free state is in orange and cyan. (G) The α-helix and β-sheet content of the protein MctA. The α-

helices are represented in green, whereas the β-sheets are in cyan with the connecting loops in blue. The bound Mg^{2+}-citrate molecule is shown in black sphere and grey stick.

5.3.2. The protein MctA is poly-specific towards divalent metal ions

Since the protein MctA was endogenously bound to Mg^{2+} ion, the bound Mg^{2+} was chelated out using EDTA chelation. Additionally, in the UniProtKB database, the protein TTHB177 is annotated as an "iron ABC transporter, periplasmic iron-binding protein". Therefore, to determine the binding affinity of the protein MctA towards Mg^{2+} and to examine whether the protein can bind other divalent metal ions, the binding of citrate-bound MctA protein with different metal ions including iron in both ferrous and ferric forms were performed using isothermal titration calorimetry (ITC). ITC experiments of the protein MctA with alkali earth, transition row as well as other divalent metal ions viz. $CaCl_2$, $MgCl_2$, $MnCl_2$, $FeSO_4$, $FeCl_3$, $NiCl_2$, $CoCl_2$, $ZnCl_2$ and $CdCl_2$ were performed (Figure 5.5A-I and Table 5.6). For all the ITC reactions, 40 µM of protein solubilized in 20 mM Tris pH 7.5, 150 mM NaCl and 5% glycerol buffer was used and titrated against freshly prepared 4 mM of different metal ions dissolved in the same buffer. The ITC results exhibit that apart from Mg^{2+} (dissociation constant (K_d): 5.88 µM), the protein MctA also shows the strongest affinity towards Ca^{2+}, Mn^{2+}, Cd^{2+}, Ni^{2+} and Co^{2+} ions with a K_d value of 0.27, 0.42, 5.88, 1.02 and 1.41 µM, respectively (Figure 5.5B-F and Table 5.6). Also, the protein MctA shows slihgtly lower binding affinity towards Zn^{2+} with a K_d of 10.00 µM, (Figure 5.5G and Table 5.6). Interestingly, in contrast to its UniProtKB annotation, the protein MctA shows almost negligible binding towards ferrous or ferric salts (Figure 5.5H and 5.5I). All the ITC reactions were exothermic in nature and are enthalpically favorable in accordance with the previous reports for other metal-binding cluster D SBPs (Sit et al., 2015).

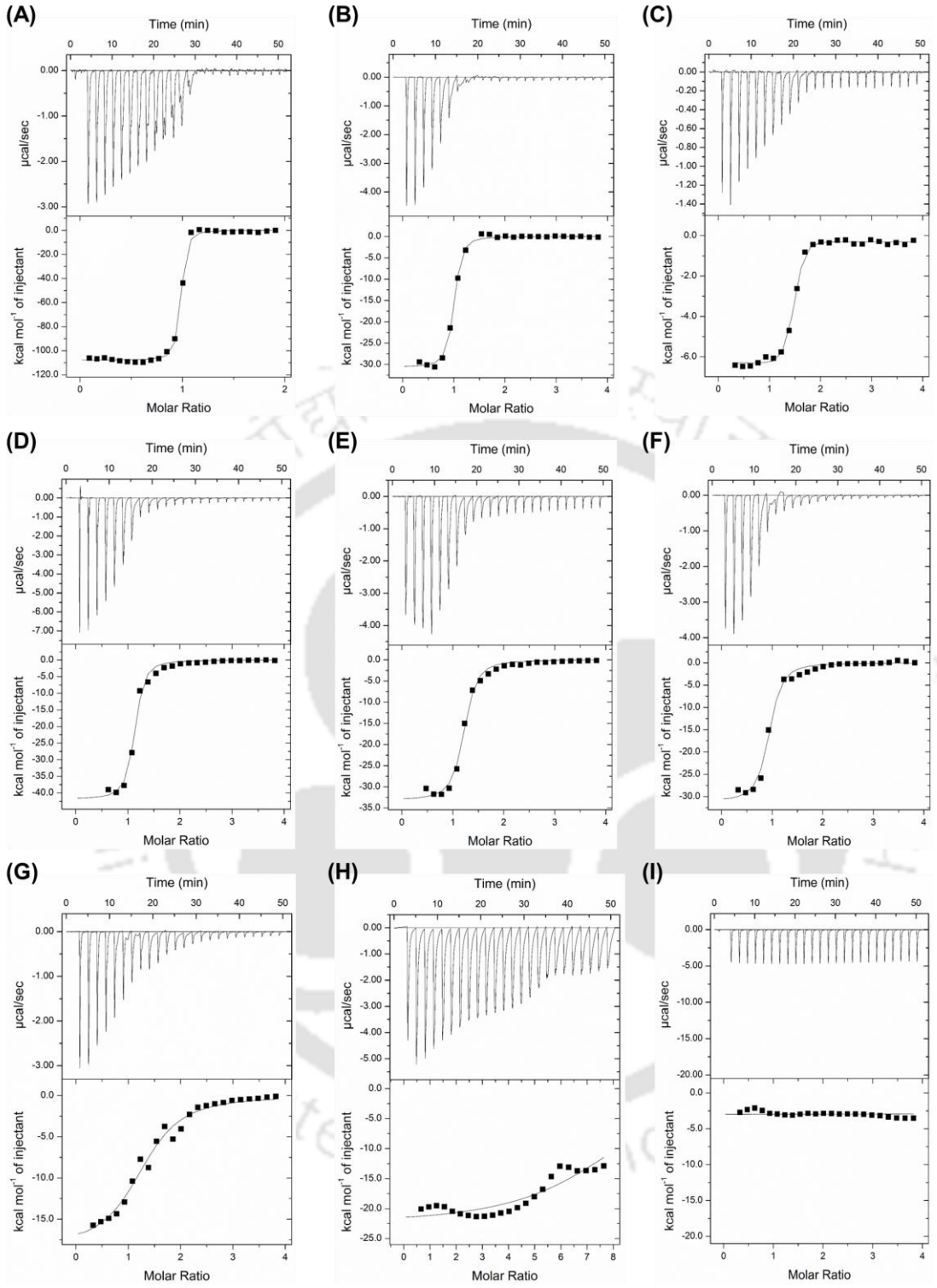


Figure 5.5. Isothermal titration calorimetry of MctA_WT protein with (A) $CaCl_2$, (B) $MnCl_2$, (C) $MgCl_2$, (D) $CdCl_2$, (E) $NiCl_2$, (F) $CoCl_2$, (G) $ZnCl_2$, (H) $FeSO_4$ and (I) $FeCl_3$.

Table 5.6. Thermodynamic parameters of various divalent metal ions and $FeCl_3$ binding to the protein MctA_WT.

Ligand	n	K_a x 10^6 (M^{-1})	K_d (µM)	ΔH	TΔS	ΔG
				(kcal mol^{-1})		
$MgCl_2$	1.43 ± 0.01	0.17 ± 0.04	5.88	-6.45	1.78	-8.24
$CaCl_2$	0.95 ± 0.00	3.72 ± 0.82	0.27	-107.90	-97.44	-10.45
$MnSO_4$	0.94 ± 0.00	2.38 ± 0.49	0.42	-30.62	-22.05	-8.57
$CdCl_2$	1.07 ± 0.01	0.17 ± 0.04	5.88	-41.86	-33.40	-8.46
$NiCl_2$	1.16 ± 0.01	0.98 ± 0.17	1.02	-33.17	-25.03	-8.14
$CoCl_2$	0.89 ± 0.01	0.71 ± 0.16	1.41	-31.16	-23.24	-7.92
$ZnCl_2$	1.30 ± 0.04	0.10 ± 0.02	10.00	-18.47	-11.62	-6.85
$FeSO_4$	Data are beyond calorimetric determination.					
$FeCl_3$	No binding.					

5.3.3. Physiological ligand of the protein MctA

The thermodynamic studies using isothermal titration calorimetry of the protein MctA with different divalent metal ions indicated its poly-specific nature of metal binding. Therefore, in order to identify the physiological metal ion for the protein MctA, the coordination chemistry of Mg^{2+} and Mn^{2+} was analyzed. In general, divalent metal ions follow three-coordination geometries to interact with metalloproteins viz., (i) tetrahedral (n = 4), (ii) trigonal bipyramidal (n = 5) and (iii) octahedral (n = 6) (Schmidt and Husted, 2019) (Figure 5.6A and 5.6B). Among the three, metal ions having the tightest first coordination sphere closest to ideal octahedral geometry is the most stable (Zheng et al., 2017; Schmidt and Husted, 2019). The coordinating environment in terms of coordination chemistry is an important factor in deciding the physiological ligands for metal binding SBPs (Vigouroux et al., 2019). Hence, the metal ion coordination for Mg^{2+} ion bound to the protein MctA and Atu2467 was analyzed. The analysis exhibits that Mg^{2+} ion forms an octahedral coordination geometry making three interactions from the citrate molecules and one each from Ser26 (Ser41 in Atu2467), Asp28 (Asp43

in Atu2467) and a water molecule (W1) (Figure 5.6C and 5.6D). Moreover, the structure of MctA_S26A mutant protein was solved in Mg^{2+}-citrate as well as Mn^{2+}-citrate bound (co-crystallized) states. Upon analyzing the metal ion coordination in MctA_S26A protein, we observed that similar to the wild type protein and Atu2467, Mg^{2+} ion is octahedrally coordinated, making a much stable metal-protein complex in MctA_S26A mutant (Figure 5.6E). However, the Mn^{2+} ion is electrostatically held by the citrate molecule and the residues Ser26 and Asp28. The interaction of the water molecule with the metal ion is absent in this case, making a penta-coordinated state following trigonal bipyramidal geometry of metal ion coordination (Figure 5.6F). Furthermore, it is noteworthy that rank-wise, Mg^{2+} appears fourth following Ca^{2+}>Mn^{2+}>Cd^{2+} in terms of binding affinity *in vitro* (Table 5.6). This is probably to facilitate its release to the cognate transmembrane domains. Such lower affinity of the physiological metal ions have also been reported earlier (Counago et al., 2014; Vigouroux et al., 2019).

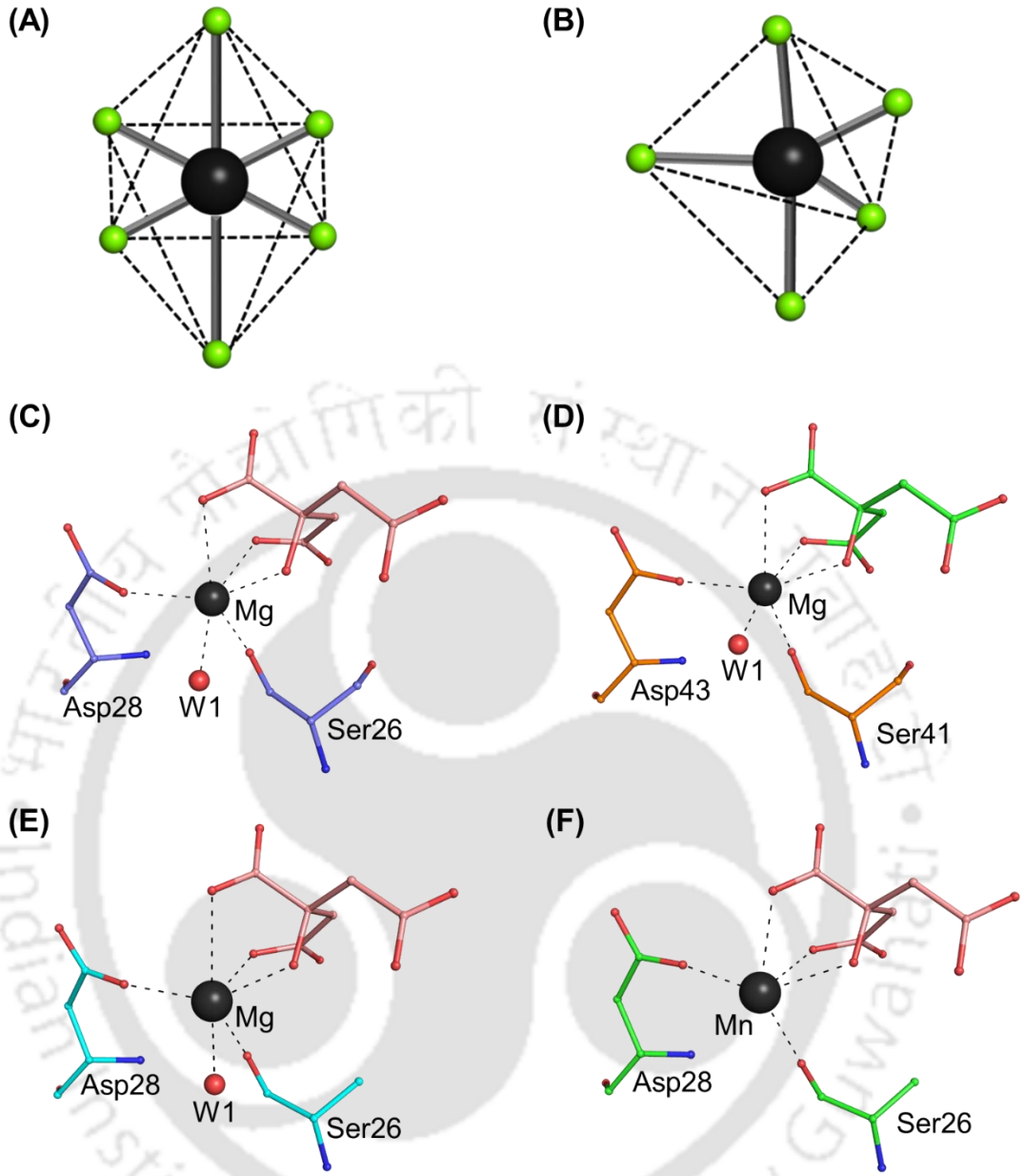


Figure 5.6. Coordination chemistry for Mg^{2+} and Mn^{2+} ion followed by MctA_WT, MctA_S26A mutant and Atu2467 protein. The two coordination geometries and coordination numbers observed in divalent metal ions (black sphere) containing metalloproteins are (A) octahedral (n = 6) and (B) trigonal bipyramidal (n = 5). Interaction of Mg^{2+} ion by citrate molecule, Ser26, Asp28 and a water molecule (W1) in the proteins (C) MctA_WT, (D) Atu2467 and (E) MctA_S26A mutant. (F) Interaction of Mn^{2+} ion by citrate molecule, Ser26 and Asp28 in the protein MctA_S26A. The bound citrate molecules in MctA_WT and mutant protein is represented in pink, whereas Atu2467 is shown in the green ball-and-stick model.

5.3.4. The protein MctA is a novel divalent metal ion-complexed citrate transporter belonging to ABC transporter superfamily

Interestingly, the presence of citrate uptake secondary transporters is reported only in Gram-positive bacteria and predicted in Gram-negative bacteria (Lensbouer et al., 2008). However, studies have reported the presence of CitS and CitC citrate uptake transporters in *Klebsiella pneumoniae* and *Salmonella enterica* which is symported with Na^+ ions (Wohlert et al., 2015; Kim et al., 2017). The presence of CitS is only predicted in *Thermus thermophilus* HB8 and several other Gram-negative bacteria. However, a homology search analysis of the protein MctA and its closest homolog Atu2467 against non-redundant, SwissProt and Protein Data Bank databases suggest that their homologous proteins to be present only in Gram-negative bacteria (Table C2). This suggests that a novel metal complexed (Mg^{2+} in this study) citrate transporter belonging to the ABC transporter superfamily is present exclusively in Gram-negative bacteria, which can efficiently uptake divalent metal complexed citrate molecules into the cytosol.

The overall three-dimensional architecture of the protein MctA is akin to other members of cluster D SBPs. Cluster D SBPs are further subclassified into four different subclusters viz., D-I to D-IV, wherein the subcluster D-I binds sugar, sugar alcohols and some amino acid precursors and D-II binds thiamine and amino acid precursors like spermidine/putrescine. The subcluster D-III is further subdivided into D-III-a and D-III-b, which binds oxyanions like molybdate, tungstate and phosphate. Members of the subcluster D-IV are reported to be involved in iron binding, whereas a separate cluster D group is also present in which SBPs not belonging to other subclusters are kept (Berntsson et al., 2010; Scheepers et a., 2016). Thus, to identify the sub-clustering for the protein MctA, a structure-based evolutionary tree was performed along with members of cluster D SBPs; details of the proteins considered are provided in Table C3. The result reveals that the protein MctA and its closest homolog, Atu2467, group with the subcluster D-IV SBPs. Further, an in-depth analysis of the evolutionary tree demonstrates that the subcluster D-IV can be subclassified into two clades. The SBPs grouped in the first clade bind iron either in ferric or ferrous form, while that cladding in the second group binds divalent metal ion complexed citrate molecules (Figure 5.7A and Table C3). Further investigations revealed that the length of an extended β-sheet

(β4 in MctA) varies in members of both groups. For instance, in iron-binding proteins FbpA and FutA1 (PDB id: 3WAE and 2PT2, respectively), this long extended β-sheet is comprising 12 and 13 amino acid residues, respectively, whereas, in MctA (this study) and Atu2467 (PDB id: 3C9H), this extended β-sheet is comparatively short with eight amino acid residues (Figure 5.7B). Thus, in this study, we renamed these groups as subcluster D-IV-a and D-IV-b, respectively. That suggests that the Mg^{2+}-citrate bound proteins MctA and Atu2467 belong to the subcluster D-IV-b SBPs.

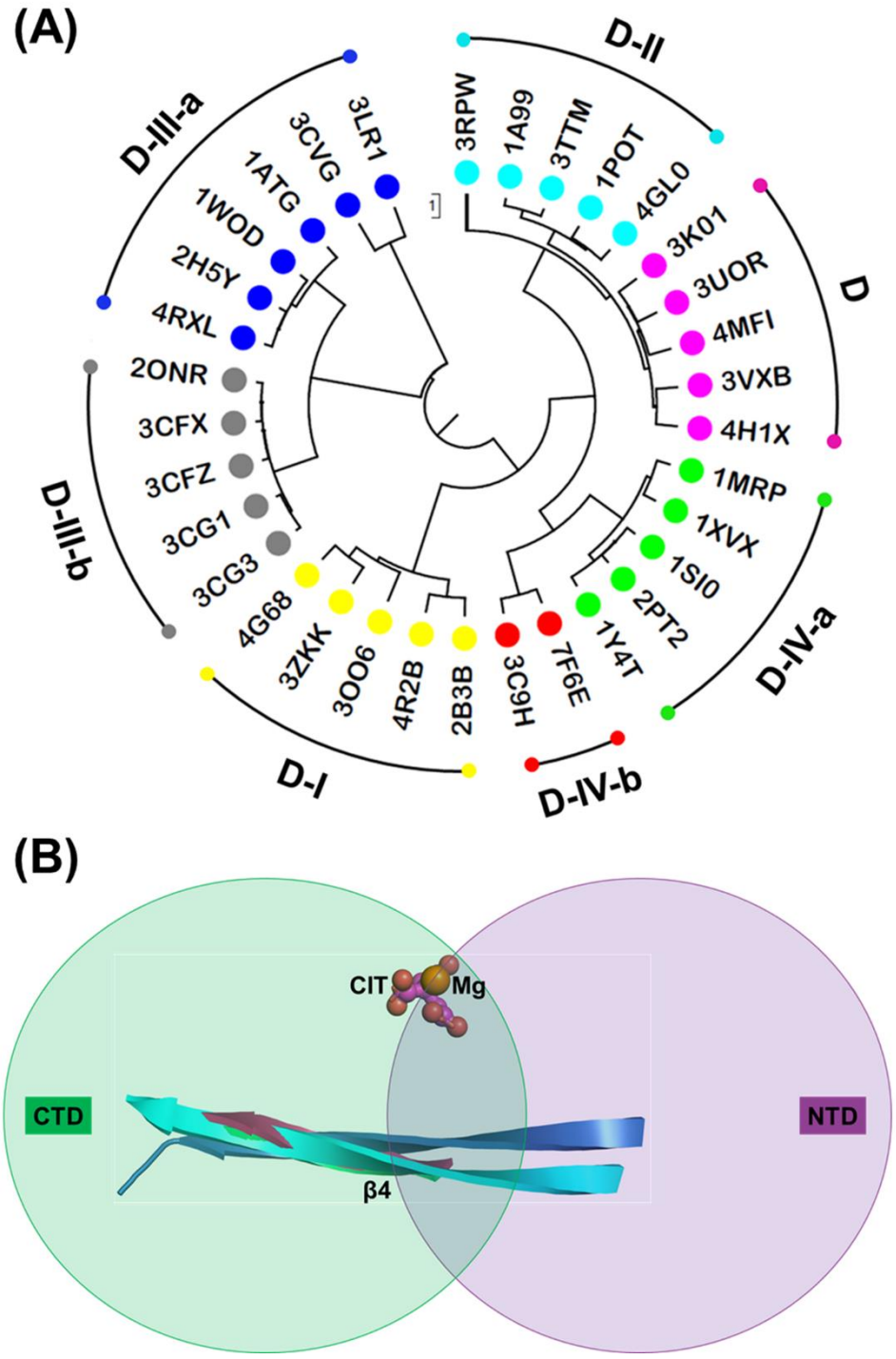


Figure 5.7. Subclassification of the subcluster D-IV. (A) Structure-based phylogeny tree of cluster D SBPs. The structures used for the analysis are denoted by their PDB ids. Each subcluster of the cluster D SBPs is represented in different colored dots. (B) Comparison of iron-binding subcluster D-IV-a SBPs (FbpA and FutA1) and Mg^{2+}-citrate binding D-IV-b SBPs (MctA and Atu2467). The long extended β4-strand of the

protein MctA (PDB id: 7F6E, waterpink), Atu2467 (PDB id: 3C9H, green), FbpA (PDB id: 3WAE, blue) and FutA1 (PDB id: 2PT2, cyan) are compared.

5.3.5. The active-site pockets of the protein MctA and its homolog Atu2467 are made for similar ligands

Since the protein MctA shares the highest homology with Atu2467 from *Agrobacterium fabrum* with a sequence identity and query coverage of 37% and 89%, respectively, a comparison between both the proteins has been drawn. The overall Cα rmsd between both the proteins is 1.24 Å with similar overall conformation (Figure 5.8A). The crystal structure of the protein Atu2467 is bound to Mg^{2+}-citrate, similar to the protein MctA. Therefore, the active-site residues interacting with the Mg^{2+}-citrate molecule in the protein Atu2467 were analyzed and compared with that of the protein MctA. Similar to MctA, the endogenously-bound Mg^{2+} ion in Atu2467 protein is also hexa-coordinated by the residues Ser41, Asp43, a water molecule and the citrate molecule itself following an octahedral coordination chemistry (Figure 5.8B and 5.8C). The citrate molecule is interacted by the residues Ser41, Thr70, Ser94, Tyr176, Gly182, Thr215, Tyr238 and Arg263, similar to the protein MctA (Figure 5.8B and 5.8C). To confirm whether the position of the active-site residues is also conserved at the sequence level also, a pairwise sequence alignment between both the proteins was performed. The alignment results exhibit that akin to the structural conservation, the ligand-interacting residues are conserved at sequence level as well (Figure 5.8D). Moreover, the alignment results also suggest the conservation of the secondary structures in both the proteins (Figure 5.8D). All these analyses collectively suggest that both the proteins MctA and Atu2467 have the potential to bind and transport similar metal ion complexed citrate molecules.

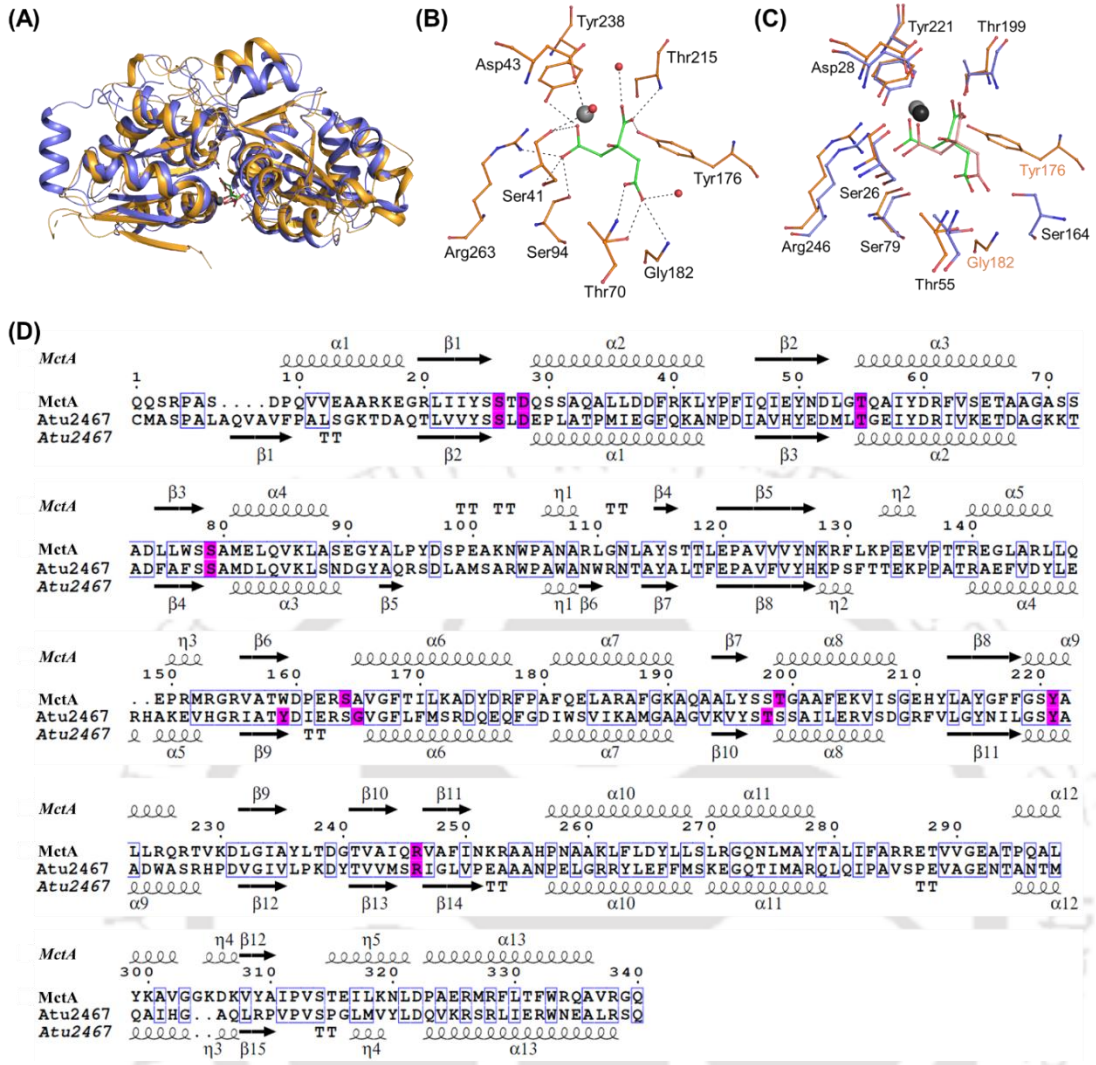


Figure 5.8. Active-site analysis of the protein MctA with its closest homolog Atu2467. (A) Superposition of MctA with Atu2467 (PDB id: 3C9H). The protein MctA is shown in blue with the bound Mg^{2+}-citrate in the black sphere and pink ball-and-stick, whereas Atu2467 is shown in orange with the bound Mg^{2+}-citrate in the grey sphere and green ball-and-stick model. (B) Active-site of the protein Atu2467 showing the interacting amino acid residues with Mg^{2+}-citrate. The active-site residues are shown in orange with the interacting citrate molecules in green ball-and-stick. The bound Mg^{2+} ion is in the grey sphere. (C) Superposition of the active-site residues of the protein MctA with Atu2467. The active-site residues of MctA and Atu2467 are shown in blue and orange with the interacting citrate molecules in pink and green ball-and-stick, respectively. The bound Mg^{2+} ion is in black and grey spheres in MctA and Atu2467 proteins. (D) Secondary structure as well as sequence alignment of the proteins MctA with Atu2467.

The residues interacting with the ligand molecule in both the proteins are highlighted in magenta. The secondary structural conservation is shown for the proteins MctA and Atu2467 in the top and bottom of the alignment, respectively.

5.3.6. The residues Thr55 and Arg246 are crucial for maintaining the overall fold of the protein MctA

Since multiple efforts to purify citrate-free MctA protein were unsuccessful, we performed site-directed mutagenesis of the active-site residues to obtain its apo structure. Hence, to determine the importance of the residues Ser26, Asp28, Thr55, Ser79, Ser164, Thr199, Tyr221 and Arg246 in Mg^{2+}-citrate binding as well as to obtain an apo structure of the protein MctA, we performed point mutation of these residues. With the mutant proteins, crystallization experiments were performed initially in the original buffer in which crystals of MctA_WT protein were obtained. Upon failure, screening of a series of crystallization buffer screening using microbatch-under-oil method leads to crystal formation for the mutant proteins. The crystallization as well as data collection and refinement statistics for the mutant proteins are provided in Table 5.2-5.5 and C1. Notably, no protein crystals were obtained for MctA_T55A and MctA_R246A mutant proteins after several efforts. Since folding and stability of the protein is a critical factor for proteins to form crystals, we performed circular dichroism (CD) for all mutants, including the wild-type protein. The CD spectra analysis suggested that akin to the protein MctA_WT, other mutants (except MctA_T55A and MctA_R246A) also conserve the overall folding and secondary structural elements of the protein (Figure 5.9A). Two of the mutants MctA_T55A and MctA_R246A, show distorted secondary structural content compared to the wild type and other mutant proteins (Figure 5.9A). Interestingly, we were also not able to obtain any protein crystals for these two mutants as well. The residue Thr55 originates in α3 helix, whereas Arg246 is part of the flexible hinge region connecting the NTD and CTD (Figure 5.9B). Energetics study using ITC for these two mutants corroborated the CD and crystallization results, suggesting no binding of MctA_T55A and MctA_R246A mutant proteins with any of the divalent metal ions (Table C4). Altogether, it can be concluded that mutation in Thr55 and Arg246 misshape the overall folding of the protein MctA, indicating these residues to be important to maintain the folding of the protein.

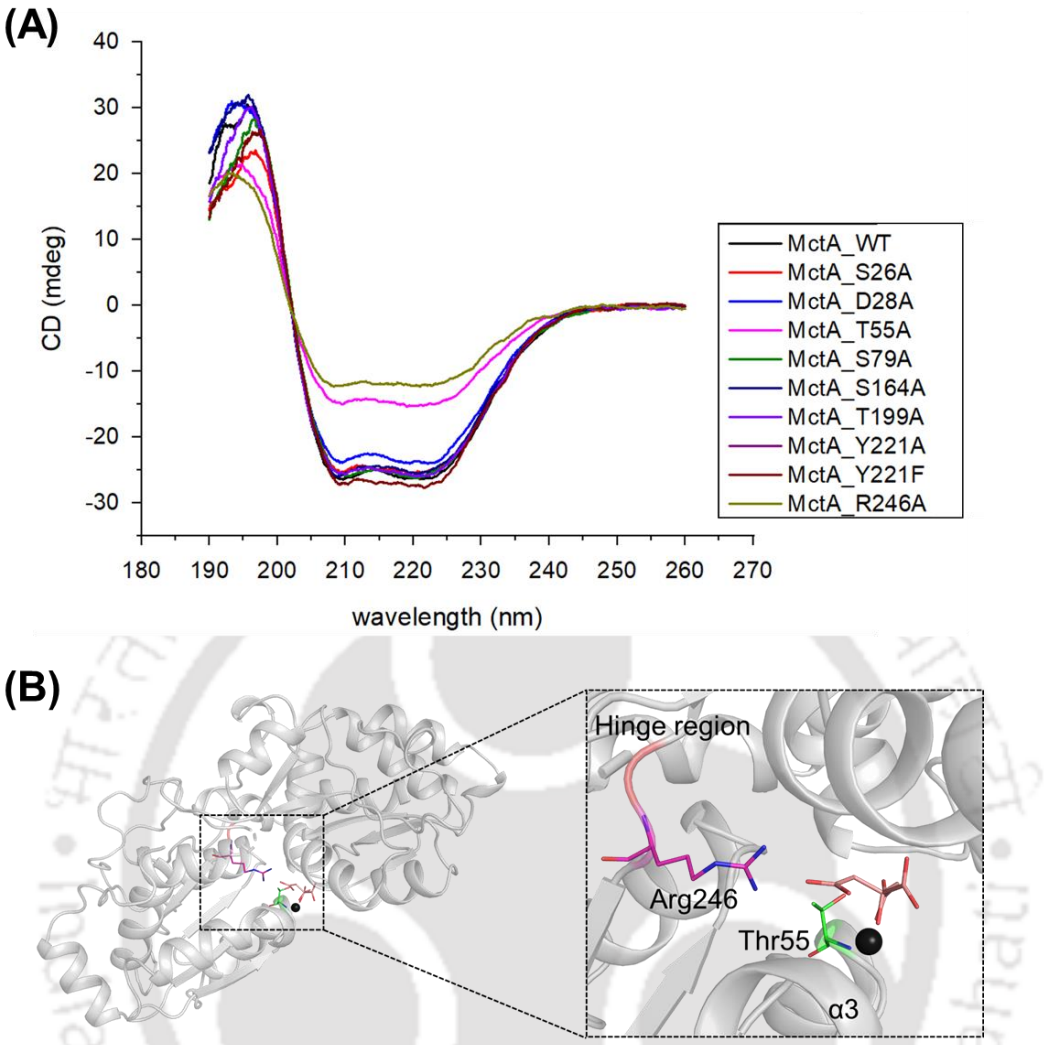


Figure 5.9. Circular dichroism spectra analysis and location of the residues Thr55 and Arg246 in the protein MctA. (A) CD spectra analysis for the protein MctA_WT and its mutants at a protein concentration of 2.5 µM. The spectra obtained for wild-type and mutant proteins of MctA are represented in different colors and are labeled. (B) Location of the residues Thr55 and Arg246 in the active site of MctA protein. The α3 helix and hinge region, which contains the Thr55 (green stick) and Arg246 (magenta stick) residue, is shown in grey and red cartoon and are labeled. The bound ligand to the wild-type protein is represented in the black sphere (Mg^{2+}) and pink stick (citrate).

5.3.7. Mutation of active-site residues possessing aliphatic hydroxyl group alters ligand specificity

Since no protein crystals were obtained for MctA_T55A mutant protein, we proceeded with the crystallization of other active-site residues possessing aliphatic hydroxyl groups. Hence, to determine the importance of the residues Ser26, Ser79, Ser164 and Thr199 in Mg^{2+}-citrate binding as well as to obtain an apo structure of the protein MctA, we performed point mutation of these residues. Mutation of the residue Ser26 into alanine does not affect the binding of Mg^{2+}-citrate with the protein MctA. Even after mutation, the protein MctA_S26A was endogenously bound to Mg^{2+}-citrate and the interaction of the active-site residues with the ligand molecule upon mutation at Ser26 was completely intact (Figure 5.10A). Along with the crystallization of the mutant proteins, co-crystallization experiments with different divalent metal ions were also performed. Although all the co-crystallization experiments were unsuccessful in yielding protein crystals, we were able to obtain a crystal for MctA_S26A mutant co-crystallized with $MnSO_4$. After solving the X-ray intensity data, a clear density for a metal ion was observed, which was suggested to be Mn^{2+} ion, upon analyzing the B-factor for different metal ions at that intensity. The Mn^{2+} ion was electrostatically bound to the protein via Ala26 (after mutation) and Asp28 along with citrate molecule forming a penta-coordination (Figure 5.10B). Interestingly, the interaction of Arg246 with the citrate molecule is lost in this case (Figure 5.10B). To confirm whether the bound metal ion in co-crystallization experiment was Mn^{2+} and to check if MctA_S26A mutant is capable of binding any other metal ions, ITC was performed. Interestingly, the binding isotherms suggest that MctA_S26A mutant enhances the binding affinity of the protein towards Mn^{2+} with a K_d of 0.11 µM (Figure 5.10C and Table C4). Moreover, the protein MctA_S26A does not exhibit binding with any other metal ions, including Mg^{2+} (Table C4).

The mutation of Ser79 into alanine confers a completely apo state of the protein MctA. Analysis of the active-site pocket suggests that the spatial orientation of the active-site residues is intact in apo form also, except for Tyr221 and Arg246 (Figure 5.10D). The ITC of MctA_S79A with metal ions suggests that in the apo-state, MctA is capable of binding divalent metal ions and Ser79 does not play any role in binding or stabilizing metal ion in the active site (Figure 5.10E, 5.10F, C2A-D and Table C4). Akin to Ser79,

mutation of Ser164 to alanine also confers apo state of the protein MctA although in a different space group (Table 5.4). Similar to Ser79 mutation, MctA_S164A also conserves the spatial orientation of the active-site residues except for Tyr221 and Arg246 (Figure 5.10G). Additionally, the mutant MctA_S164A is also capable of binding divalent metal ions at the active site of the protein with slightly lower affinity (Figure 5.10H, 5.10I, C2E-H and Table C4).

Similar to the residues Ser79 and Ser164 mutants, mutation of Thr199 to alanine also resulted in an apo state of the protein MctA with conserved spatial orientation of the active-site residues except for Tyr221 and Arg246 (Figure C2I). However, in contrast to serine mutants, MctA_T199A does not exhibit binding with any of the divalent metal ions suggesting their role in indirect stability to the metal ion (Table C4). Additionally, none of the mutant proteins shows binding with sodium citrate molecule even at 100X concentration (4 mM), suggesting their importance in citrate binding at the active-site pocket (Table C4).

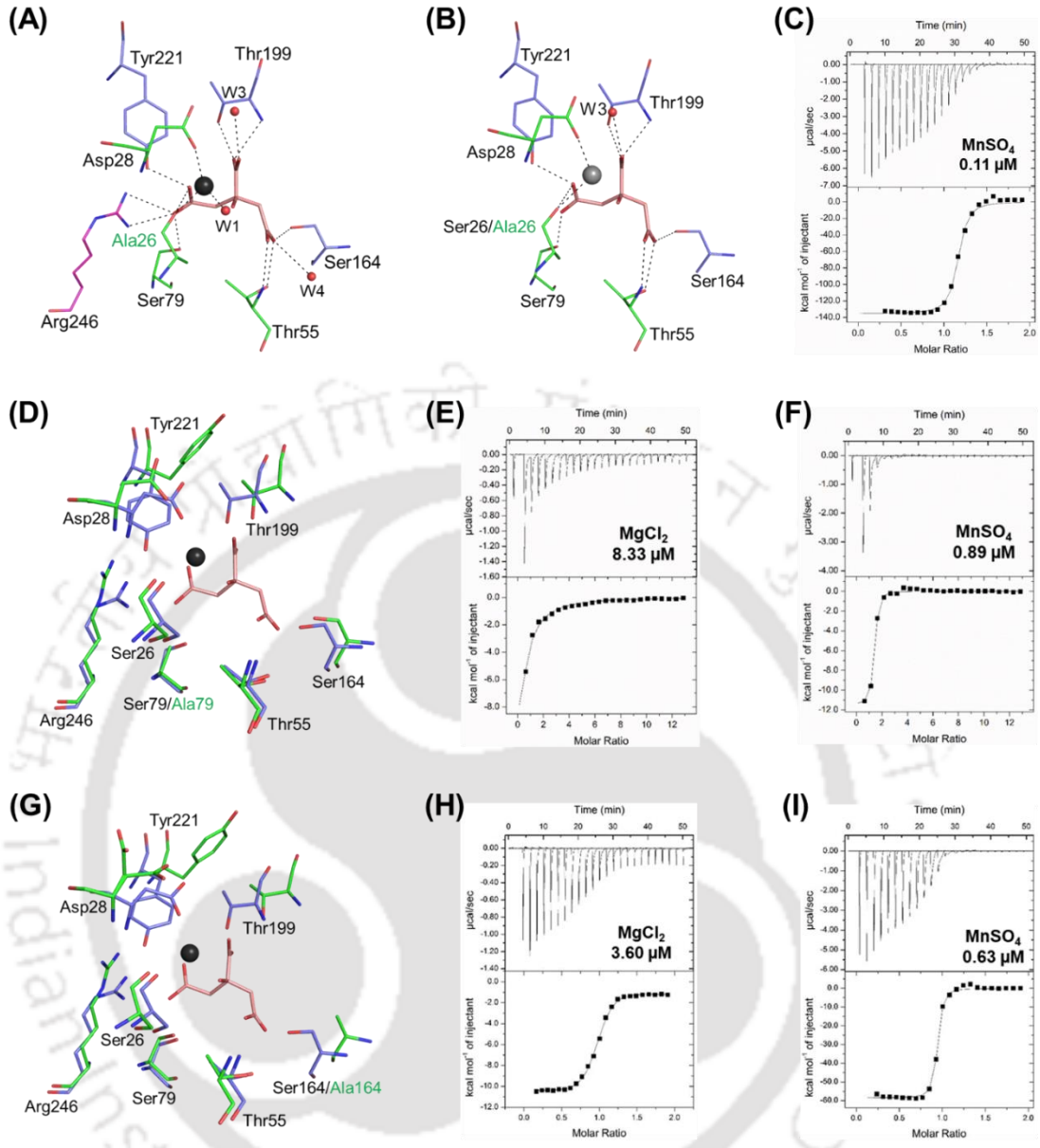


Figure 5.10. Mutagenic study of the active-site residues possessing aliphatic hydroxyl group. Interaction of the active-site residues of MctA_S26A mutant with (A) Mg^{2+}-citrate and (B) Mn^{2+}-citrate. The interacting residues from the N- and C-terminal domains are shown in green and blue, respectively, whereas the hinge region is in the magenta stick model. The Mg^{2+} and Mn^{2+} ions are in black and grey spheres, respectively, with the bound citrate in the pink stick. (C) ITC of the protein MctA_S26A with $MnSO_4$. (D) Superposition of the active-site residues of the ligand-bound MctA_WT protein and ligand unbound MctA_S79A mutant. ITC of the protein MctA_S79A with (E) $MgCl_2$ and (F) $MnSO_4$. (G) Superposition of the active-site residues of the ligand-bound MctA_WT protein and ligand unbound MctA_S164A

mutant. ITC of the protein MctA_S164A with (H) $MgCl_2$.The active-site residues of the proteins MctA_WT and mutants are shown in blue and green sticks, respectively, with the bound ligand in the wild type protein in the black sphere (Mg^{2+}) and pink stick (citrate).

5.3.8. The residue Asp28 is important for metal binding, whereas mutation in Tyr221 confers an apo state of the protein MctA

Mutation of serine residues except for Ser26 present in the vicinity of the active site confers an apo state of the protein. In addition to the citrate, Ser26 also interacts with the metal ion Mg^{2+}. Akin to Ser26, Asp28 also establishes interaction with the metal ion in the active-site pocket. Mutagenic studies of Asp28 to alanine results in a metal-free but citrate-bound structure of the protein MctA, suggesting that Asp28 is crucial for metal binding at the active site (Figure 5.11A). The active-site analysis of the MctA_D28A mutant structure suggests that all the interactions between the active-site residues and citrate molecule is conserved with almost identical bond lengths (Figure 5.11A). Comparision of the active-site pocket of MctA_D28A mutant with the wild-type protein suggested an intact spatial orientation of the active-site residues having similar interactions with the citrate molecule exhibiting an rmsd of 0.20 Å for Cα atoms (Figure 5.11B). Since the Mg^{2+} ion is absent at the active site of the mutant protein MctA_D28A, but citrate was endogenously bound, we performed ITC to check if the Asp28 mutation is capable of binding any other divalent metal ions. The results exhibited no binding for any divalent metal ion with the protein MctA_D28A suggesting its pivotal role in metal ion binding and stability (Table C4).

In contrast to Asp28 mutation, mutating Tyr221 residue to alanine and phenylalanine confers an apo state of the protein MctA in the space group *I*222 similar to Ser164 and Thr199 mutants. Akin to the residues Ser79, S164A and Thr199 mutants, MctA_Y221A and MctA_Y221F also conserve the spatial orientation of the active-site residues except for the residue Arg246 (Figure 5.11C and 5.11D). Moreover, in contrast to other mutants, where the apo state of the protein was obtained, the orientation of Phe221 (mutated from Tyr221) faces the active site of the protein towards the ligand. However, the absence of citrate molecule in MctA_Y221F suggests that the residue

Tyr221 is important for citrate-binding at the active-site pocket (Figure 5.11D). Additionally, the mutant MctA_Y221A and MctA_Y221F do not exhibit binding with any of the divalent metal ions as well as citrate molecule at the active-site pocket of the protein (Table C4).

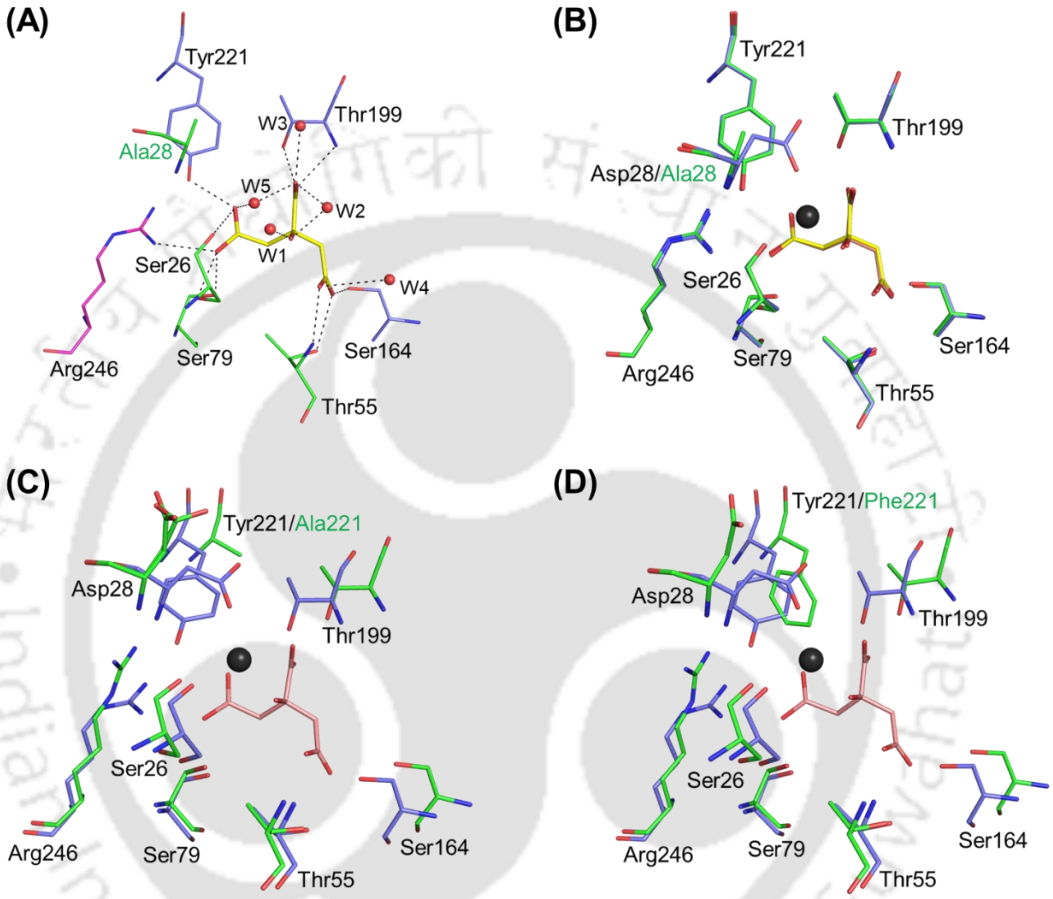


Figure 5.11. Mutagenic study of the active-site residues Asp28 and Tyr221. (A) Interaction of the active-site residue of MctA_D28A mutant with citrate molecule. The bound citrate molecule in MctA_WT and MctA_D28A is represented in pink and yellow sticks, respectively. Superposition of the active-site residues of the Mg^{2+}-citrate bound MctA_WT protein with (B) citrate bound MctA_D28A, (C) ligand unbound MctA_Y221A and (D) ligand unbound MctA_Y221F mutants. The active-site residues of the proteins MctA_WT and mutants are shown in blue and green sticks, respectively, with the bound ligand in the wild type protein in the black sphere (Mg^{2+}) and pink stick (citrate).

5.3.9. The protein MctA follows a gating mechanism of ligand entry

To affirm the importance of the residue Tyr221 in citrate binding and to gain insights into the mechanism of ligand entry, the mutant structures in their apo states were compared with the ligand-bound MctA_WT protein. An in-depth analysis of the active site of the protein MctA and its comparison with the apo states of the protein shows that the residue Tyr221 is present at the entry point near the active site (Figure 5.12A). Upon comparing the apo and the holo states of the protein MctA, significant rotameric shifts were observed in the residue Tyr221 at the active-site pocket (Figure 5.12B-E). The residue Tyr221 exhibits a considerable degree of rotation (~105°) towards the Mg^{2+}-citrate molecule upon its entry to the active-site pocket. This suggests that the residue Tyr221 acts like a 'gate' at the ligand entry site controlling the entry of the ligand molecule in the active site (Figure 5.12B-E). A comparison of the ligand-bound closed state of the protein MctA with ligand-free MctA_Y221F mutant structure revealed a rotation of Tyr221 to Phe221 by 18.7° (Figure 5.12B). However, in a ligand-unbound open state, this rotameric movement of Tyr221 increases to 99.3° and 102.4° in the case of MctA_S79A and MctA_T199A mutants, making a transition state of ligand entry (Figure 5.12C and 5.12D). However, in a completely open form, the rotameric shift of Tyr221 is 105.7° showing the maxima of conformational change in Tyr221 for entry of Mg^{2+}-citrate molecule at the active site (Figure 5.12E). This indicates that the protein MctA follows a 'gating mechanism' of ligand entry where the residue Tyr221 acts as a 'gate' which swaps in different rotameric confirmations allowing entry of the ligand.

Moreover, the residue Tyr221 also interacts with the residue Asp28 from the N-terminal domain. In a ligand-unbound open state, the residue Asp28 flips in an outward direction losing interaction with the Mg^{2+}, whereas, in a ligand-bound closed state, Asp28 interacts with the metal ion Mg^{2+} (Figure 5.12F and 5.12G). Subsequent changes in the movement of the helix α11, which contains the residue Asp28 towards the active-site pocket was observed (Figure 5.12F and 5.12G). Notably, the residues Asp28 and Tyr221 are present in the N- and C-terminal domains, respectively, of the MctA protein. The interaction of the residues Tyr221 (C-terminal) with Asp28 (N-terminal) upon ligand binding suggests a locking mechanism of securing the bound state of the ligand. Altogether, it can be suggested that the residue Tyr221 acts as a gate for ligand entry

and exit, wherein upon ligand entry, the gate closes itself and locks the active site by interacting with the residue Asp28.

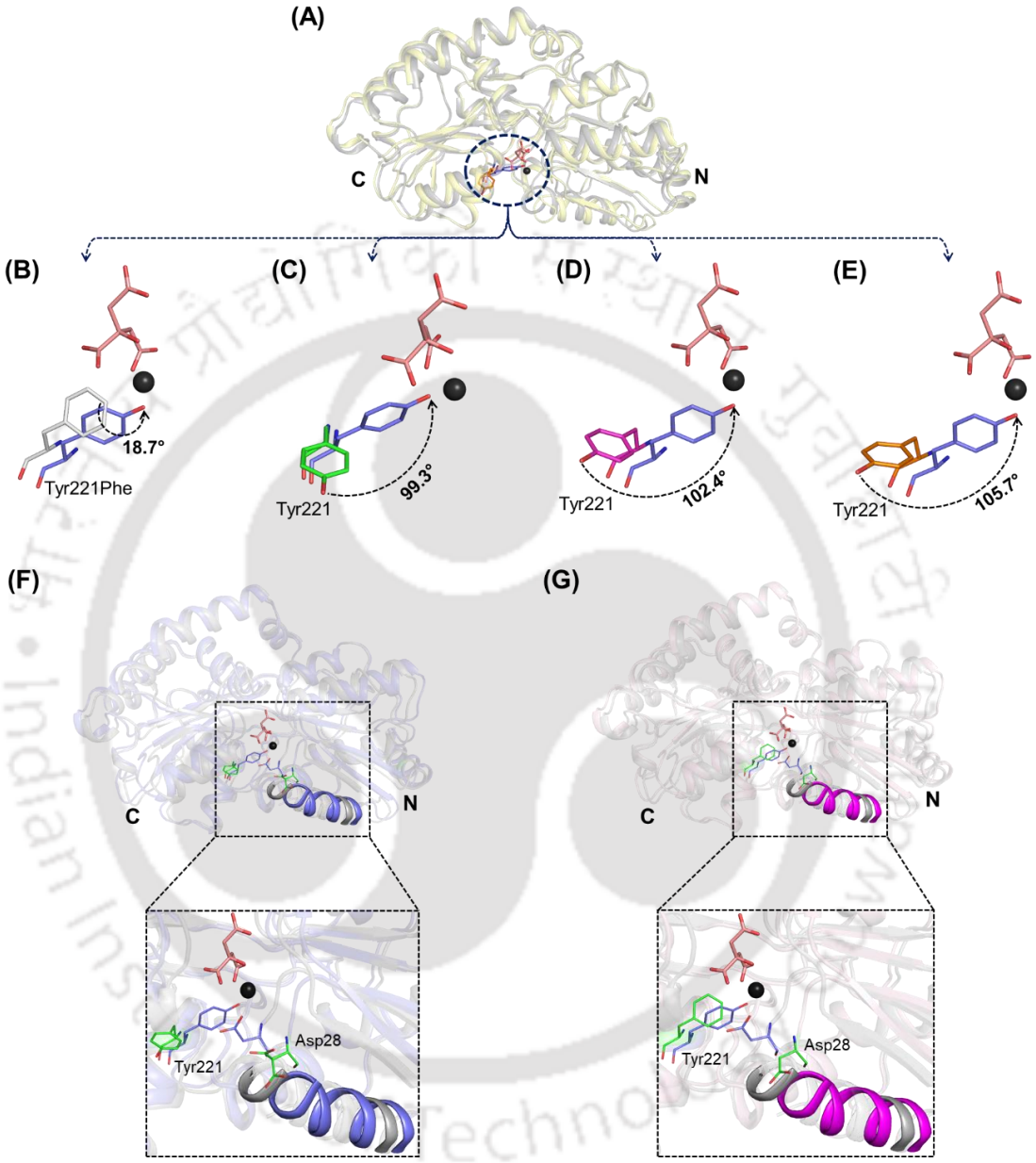


Figure 5.12. Gating mechanism of ligand entry in the protein MctA. (A) Location of the residue Tyr221 cited as a gate at the entry site of the ligand Mg^{2+}-citrate. Opening and closing movement of the gate residue Tyr221 showing different angles of movement in (B) MctA_Y221F, (C) MctA_S79A, (D) MctA_T199A and (E) MctA_S164A mutants compared with the protein MctA_WT. Tyr221 of MctA_WT structure is shown in the blue stick. (F, G) Comparison of the open (MctA_S164A and

MctA_Y221F) and closed (MctA_WT) structures of wild-type and mutant MctA showing the loss of interaction between Tyr221 of C-terminal and Asp28 of N-terminal. The interaction between the Tyr221 and Asp28 in MctA_WT is represented in blue stick, whereas the outward orientation of Tyr221 and Asp28 in ligand-unbound structures is represented in green sticks. The bound Mg^{2+}-citrate molecule is represented the black sphere and pink stick model.

5.3.10. The protein MctA follows an "asymmetric domain movement" mechanism of ligand binding

The overall three-dimensional structure of MctA_S164A mutant protein reveals a completely open conformation in an unliganded state. Topologically, MctA_S164A mutant protein (unbound structure) is identical to that of the protein MctA_WT (Mg^{2+}-citrate-bound structure) with an rmsd of 1.58 Å. Overlaying the NTD and CTD of the open and closed conformations of the protein MctA using the web server DynDom (Hayward and Berendsen, 1998) showed an rmsd of 0.67 Å for 176 amino acid residues comprising the NTD, whereas 0.39 Å for 156 amino acid residues comprising the CTD, illustrating a moderately larger conformational change for NTD than CTD. Furthermore, the distance between Cα atom of Lys41 (helix α3) and Ala68 (helix α4) in both the open and closed conformations remain constant (~34.8 Å), indicating that the NTD undergoes a rigid-body translation movement (Figure 5.13A). Calculation of the rotational angle of the NTD movement by the DynDom web server shows a torsion angle change of ~14° on hinges (Figure 5.13B). Although, compared to SBPs binding larger molecules like carbohydrates, the rotational movement of 14° is comparatively small; however, considering the fact that the protein MctA binds a small Mg^{2+}-citrate molecule, this rotation is enough for ligand binding (Chandravanshi et al., 2020). Interestingly, analysis for the domain closure mechanism, the analysis suggests that upon ligand binding, NTD shows a moderately larger conformational change than the CTD, which is in accordance with the previously reported "asymmetric domain movement" mechanism of ligand binding to the SBPs (Figure 5.13C-E) (Pandey et al., 2016; Chandravanshi et al., 2020). Notably, the "asymmetric domain movement" of ligand binding has been reported earlier for sugar-binding proteins only (Pandey et al., 2016; Chandravanshi et al., 2020). Therefore, we believe that this is the first study

reporting insights of the "asymmetric domain movement" for a metal ion (Mg^{2+}) complexed citrate binding protein belonging to the subcluster D-IV-b or broadly in cluster D (Figure 5.13E).

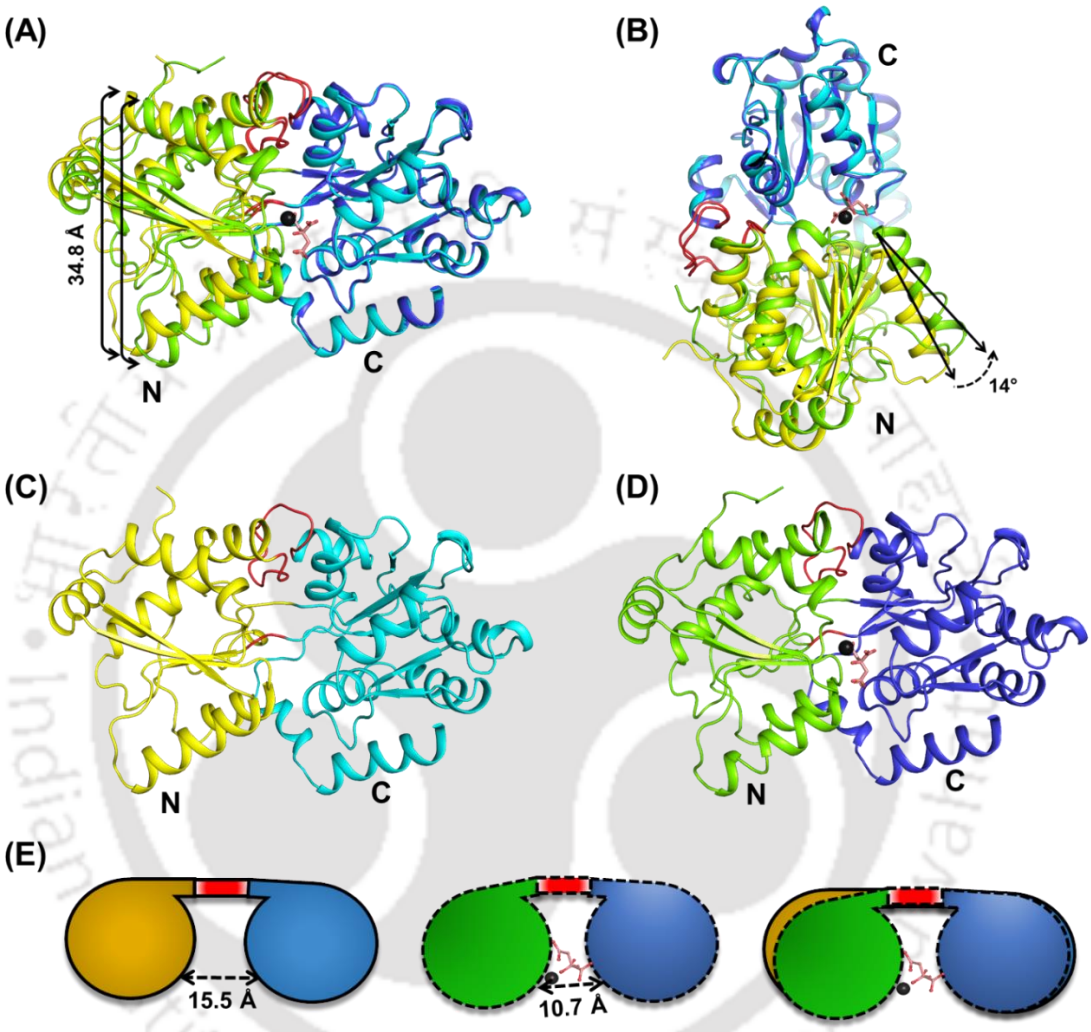


Figure 5.13. Asymmetric domain movement confers ligand binding in the protein MctA. Superposition of the ligand-bound (MctA_WT) and unbound (MctA_S164A) MctA protein suggesting (A) no overall change between them and (B) an angular shift of ~14° of the N-terminal domain towards the active site. The overall three-dimensional structure of the (C) open (MctA_S164A, unbound) and (D) closed (MctA_WT, ligand-bound) state of the protein. The N- and C-terminal domains in ligand-bound state of the protein are shown in green and blue, respectively, whereas in ligand-free state are shown in yellow and cyan, respectively. (E) Schematic representation of the asymmetric domain movement of the N-terminal domain for ligand binding in the

protein MctA. The conformational changes of the NTD (green) and CTD (blue) upon ligand binding are depicted with dotted lines. The bound Mg^{2+}-citrate molecule is represented as the black sphere and pink stick model.

5.4. DISCUSSION

Thermophilic microorganisms are capable of surviving in diverse habitats such as geothermal springs, volcanic areas, hot water regions, marine hydrothermal vents and geothermal power plants (Ranawat and Rawat, 2018). Such punitive environments rich in metal, metalloid and metal ion complexes necessitate the availability of specific metal uptake transporters in thermophilic microorganisms (Mandal et al., 2019). Among the various functions of divalent metal ions in cellular and physiological processes, one of the important functions is to stimulate the uptake of citrate molecules in prokaryotes (Krom et al., 2000). Although divalent metal ions such as Mg^{2+}, Mn^{2+}, Ca^{2+} complexed citrate transporters are reported in Gram-positive bacteria, which belongs to the secondary transport system, Na^{+}-dependent citrate transporters (CitS symporters) are reported in Gram-negative bacteria as well (Saeir, 2000; Krom et al., 2000, Kim et al., 2017). To the best of our knowledge, divalent metal ion complexed citrate transporter is not yet reported in Gram-negative bacteria. In this study, the structural and functional characterization of a novel Mg^{2+}-citrate-binding protein MctA is performed, which is the SBP component of a ABC transporter TTHB175-TTHB176-TTHB177 (referred to as MctCBA) from a thermophilic bacteria *T. thermophilus* HB8.

Sequence-based homology search suggested that protein homologs of MctA are present only in Gram-negative bacteria. Citrate is a primary metabolite and a source of carbon and energy in microorganisms. The threshold concentration of metal ions and citrate molecules in Gram-positive bacteria are maintained by multiple secondary transporters such as CitMHS; however, these transporters are only predicted in Gram-negative bacteria (Lensbouer et al., 2008). To date, the presence of only CitS, which is a Na^{+}/citrate symporter, is reported in *K. pneumoniae* and *S. enterica* for citrate transport (Wohlert et al., 2015; Kim et al., 2017). Therefore, in order to maintain the threshold concentration of metal ions and citrate, Gram-negative bacteria possess an ABC transport system which probably is also involved in metal ion homeostasis. The

structural architecture of the protein MctA suggests that it belongs to the family of cluster D SBPs possessing an N- and C-terminal domain liked by two hinge region (Berntsson et al., 2010). Moreover, the structure-based evolutionary tree suggested a novel classification of subcluster D-IV as D-IV-a and D-IV-b. The protein MctA, along with its homologous protein Atu2467 (from *A. fabrum*, PDB id: 3C9H), belongs to subcluster D-IV-b SBPs. However, the paucity of structural information on metal ion complexed citrate binding ABC transporter proteins from other microorganisms impedes further confirmation of this new sub-clustering. Topologically, the difference between subcluster D-IV-a and D-IV-b lies in the length of an extended β-sheet, which is longer in the case of iron-binding D-IV-a SBPs whereas shorter in Mg^{2+}-citrate binding D-IV-b SBPs.

Earlier studies have reported that although citrate uptake is stimulated by a divalent metal ion, one particular citrate uptake secondary transporter can uptake only one type of metal ion complexed citrate molecule. For instance, CitM can transport citrate in complex with Mg^{2+}, Ni^{2+}, Mn^{2+}, Co^{2+}, and Zn^{2+} and not with Ca^{2+}, Ba^{2+}, and Sr^{2+}. Similarly, CitH can transport citrate in complex with Ca^{2+}, Ba^{2+}, and Sr^{2+} and not with Mg^{2+}, Ni^{2+}, Mn^{2+}, Co^{2+}, and Zn^{2+} (Krom et al., 2000). Our *in vitro* studies exhibit that the protein MctA is capable of binding several divalent metal ions such as Ca^{2+}, Mn^{2+}, Co^{2+}, Ni^{2+}, Cd^{2+} and Zn^{2+} in addition to Mg^{2+}. Although citrate can chelate iron with higher affinity and the protein MctA is also annotated as an “iron ABC transporter, periplasmic iron-binding protein”, it does not show an affinity for iron in both ferrous and ferric forms. This suggests that the binding of the metal ion is not only decided by the citrate molecule in the extracellular environment, but the protein MctA also plays a significant role is deciding the type of metal ion to be transported. Moreover, mutations in serine residues present in the active site enhance the metal binding affinity for some of the divalent metal ions. Since *T. thermophilus* is a thermophilic organism capable of surviving in higher temperatures, engineering the protein MctA might enhance its potentiality to be used as a drug delivery system or for synthesizing metal nanoparticles.

Comparison of the active-site pocket of the apo and holo MctA proteins suggested that the protein follows a “gating mechanism” of ligand entry. The presence of the residue Tyr221 at the ligand entry point of the protein and its movement at different angles

(~18° to 105°) like a gate allows the entry of the Mg^{2+}-citrate for binding. The gating mechanism of ligand entry is earlier proposed for SitA protein from *Staphylococcus pseudintermedius* which is an Mn^{2+}/Zn^{2+} binding protein where a histidine residue (His64) acts as a gate for ligand entry (Abate et al., 2014). Moreover, the apo and holo-structures of the protein MctA suggested that the protein binds the ligand Mg^{2+}-citrate following an "asymmetric domain movement" mechanism of ligand binding. The NTD moves towards the active site by ~14° for ligand binding. For binding of small molecules like Mg^{2+}-citrate, a movement of the NTD by ~14° is enough for ligand binding. Although the residues from the CTD also interacts with the ligand, CTD does not show much movement for ligand binding. This could be due to the presence of the residue Tyr221 as a gate of the protein in the C-terminal domain, which is involved in allowing the ligand molecule to enter the active site of the protein. Earlier, the "asymmetric domain movement" mechanism of ligand binding has been reported for *pp*GBP and αGlyBP (from *Pseudomonas putida* and *T. thermophilus*), which binds carbohydrate molecules (Pandey et al., 2016; Chandravanshi et al., 2020). In both *pp*GBP and αGlyBP, only the NTD participates in domain closure. This is the first study reporting an "asymmetric domain movement" mechanism of ligand binding for a divalent metal ion complexed (Mg^{2+})-citrate molecule. Altogether, this study employs structural and thermodynamic analysis to provide the basis for the ligand (Mg^{2+}-citrate)-binding through an ABC transporter. Moreover, the poly-specific nature of the protein MctA and an enhanced binding affinity for divalent metal ions upon mutation can be further exploited for biological applications.

CHAPTER 6

Structural and functional characterization of the protein MhuP

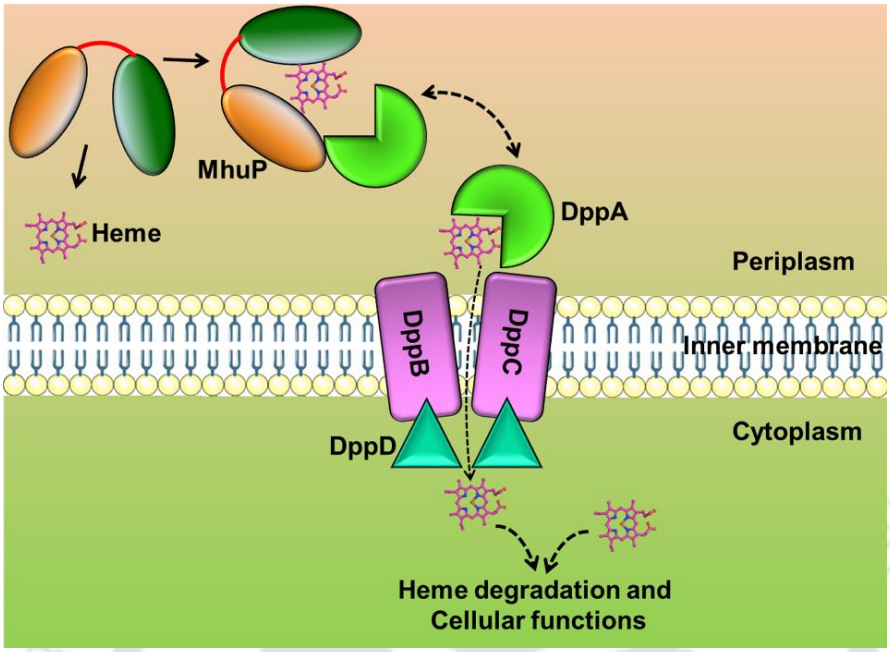
The manuscript for this chapter is under revision as:

1. **Mandal SK** and Kanaujia SP. Role of an orphan substrate-binding protein MhuP in transient heme transfer in *Mycobacterium tuberculosis*. ***Under Revision***.

ABSTRACT

The redox property of iron makes it an essential cofactor for numerous enzymes involved in various metabolic processes. In vertebrates, iron is bound to either heme or other circulatory proteins and therefore, its accessibility is restricted for bacterial pathogens residing inside the host. Due to the importance of iron in all organisms, there is always an ongoing battle between the host system and pathogens, known as nutritional immunity. To capture the bound iron from the human hosts, intracellular pathogens like *Mycobacterium tuberculosis* secrete siderophore molecules which are ultimately sequestered by eminent transport machineries such as ABC transporters. More than 70% of the available iron in the human body is strongly bound to heme or hemoglobin, making it a major source of iron for *M. tuberculosis*. Apart from secreting iron-sequestering siderophore molecules, *M. tuberculosis* has evolved themselves to uptake and utilize iron-bound heme molecules from the human host. Earlier reports have suggested the presence of a heme-binding protein MhuP (ORF id: Rv0265c) in *M. tuberculosis* which transiently transfers the bound iron to a DppA protein for further heme transport by utilizing its cognate transport machinery (DppBCD). In the present study, we report the crystal structure of the protein MhuP in two different space groups $P2_1$ and $C2$, at a resolution of 1.8 and 2.01 Å, respectively. The binding experiments of heme with the protein MhuP in solution suggest selective as well as specific binding of the protein towards heme. Moreover, docking studies exhibited the binding of the protein MhuP with heme as well as with DppA protein for heme transfer. Thus, the results confirm the binding of heme with the protein MhuP and its transient transport via the DppABCD transport system in *M. tuberculosis*.

Graphical abstract



6.1. INTRODUCTION

Iron is one of the most abundant elements on the earth and is required by all domains of life to accomplish various life processes. Iron has an incomplete d-orbital, which allows it to attain diverse oxidation states and redox chemistry (Cotton, 2000). Iron exists in two oxidation states, ferrous (Fe^{2+}) and ferric (Fe^{3+}), and is involved in various oxido-reduction reactions (Andrews et al., 2013). Fe^{3+} ionic form is dominantly present in oxygenated environments having very low solubility, which appears as a problem for aerobic microorganisms (Andrews et al., 2013). In contrast, in anaerobic or microaerobic environments with low pH, Fe^{2+} ionic form is the most abundant (Andrews et al., 2003). However, due to this unique redox chemistry, iron is essential as a cofactor for numerous enzymes and plays pivotal roles in electron transport, oxygen metabolism and oxidative phosphorylation, tricarboxylic acid (TCA) cycle, peroxide detoxification, lipid metabolism, DNA biosynthesis and xenobiotic metabolism (Hentze et al., 2004; Krewulak and Vogel, 2008; Cornelis et al., 2011; Zhang, 2014; Ma et al., 2015). Succinctly, iron is intimately involved in numerous crucial cellular and biological processes. However, pathogenic bacteria face problems inside host systems to scavenge iron due to the unavailability of a free ion. Iron in the

human body is bound to heme molecules or by circulatory proteins such as lactoferrin and transferrin (Finkelstein et al., 1983; Cornelissen and Sparling, 1994). It is estimated that ~70% of the available iron in the human body is tightly associated with heme and is complexed with hemoglobin making it the only major source of iron in the human host (Huang and Wilks, 2017; Mitra et al., 2019). Since iron is required virtually by all pathogenic bacteria, vertebrates, including human beings, restrict the access of iron for these pathogens, which is a potent defense mechanism against infections. This restriction of iron for the bacterial pathogens imposed by the host system is termed as 'nutritional immunity' (Hennigar and McClung, 2016). Reports have suggested the presence of a high concentration of iron-scavenging proteins and iron restricting factors at the necrotic centers of granulomas of TB patients, which likely establishes an iron-deficient environment in *M. tuberculosis* (Kurthkoti et al., 2017). In addition to iron restriction at the primary infection sites, studies have also suggested that iron starvation for a longer period of time triggers the transition of *M. tuberculosis* to a persistent state leading to chronic TB (Kurthkoti et al., 2017).

However, to overcome or counter the nutritional immunity imposed by the host, intracellular pathogens such as *Mycobacterium tuberculosis* have sophisticated mechanisms to scavenge iron ions inside the host. *M. tuberculosis* secretes extremely high iron-binding siderophores such as (carboxy)mycobactins which are sequestered by the well-characterized IrtAB ABC transport system (Snow, 1970; Rodriguez, 2006; Rodriguez and Smith, 2006; Arnold et al., 2020). Since a major portion of iron in the host system is heme-bound, *M. tuberculosis* has evolved to sequester and utilized heme and hemoglobin as iron source as well (Jones and Niederweis, 2011; Tullius et al., 2011). Several studies have reported the importance of these iron sequestering mechanisms for *M. tuberculosis* virulence. For instance, the IrtAB transport system is necessary for efficient uptake of iron from ferric carboxymycobactin as well as replication of the pathogen in macrophages and mice (Rodriguez and Smith, 2006). An *M. tuberculosis* mutant, which lacks the iron storage protein ferritin, is incapable of establishing a chronic infection in mice and the main iron homeostasis regulator in *M. tuberculosis* IdeR is essential for virulence (Pandey and Rodriguez, 2012; Pandey and Rodriguez, 2014).

Highlighting the fact that more than 70% of iron ions are tightly bound to heme and complexed with hemoglobin, heme is the principal source of iron in human hosts (Huang and Wilks, 2017). In bacteria with two cell membranes (diderm), the heme or hemoproteins present in the host system are scavenged by the outer membrane receptors (OMRs) to transport heme molecules into the periplasmic space. Further, the heme is translocated across the inner membrane by heme-specific ABC importers (Huang and Wilks, 2017). Recently, Niederweis and co-workers have reported two membrane-anchored heme-binding proteins PPE36 and PPE62, present on the cell surface of *M. tuberculosis* required for heme utilization (Mitra et al., 2017). Earlier, mycobacterial membrane proteins MmpL3 and MmpL11 were proposed as inner membrane heme importers (Tullius et al., 2011). Later on, MmpL3 and MmpL11 are characterized as RND-type efflux pumps involved in trehalose monomycolate, lipids, and/or other lipid-like molecule export for mycobacterial cell wall maintenance (Varela et al., 2012; La Rosa et al., 2012; Pacheco et al., 2013). Thus, the role of MmpL3 and MmpL11 in heme transport is unclear in *M. tuberculosis* (Mitra et al., 2017).

Niederweis and co-workers have reported the presence of a heme-binding protein (ORF id: Rv0265c; hereafter referred to as ***M****ycobacterium* **h**eme **u**ptake **p**rotein MhuP) in *M. tuberculosis* H37Rv, which binds and transports heme transiently to a peptide-binding protein DppA which belongs to the DppABCD transporter. In this study, we have performed the structural and functional characterization of the protein MhuP, which confirms the binding of heme with the protein. The study also exhibits the plausible binding of a heme-bound MhuP protein with DppA protein proposing a binding mechanism between the two proteins.

6.2. MATERIALS AND METHODS

6.2.1. Construction of wild-type expression plasmids

The recombinant construct of the gene *mhuP* (993 bp) was obtained by molecular cloning of the gene in pET22b(+) plasmid vector. For cloning, a truncated form of the gene *mhuP* (894 bp, including 18 bp for 6xHis-tag and lacking the region of first 57 bp forming a lipid anchored region in the protein MhuP) was amplified using the *Mycobacterium tuberculosis* H37Rv genome procured from ATCC as a template. The

C-terminal end of the protein MhuP was tagged with a 6xHis-tag for obtaining a protein using immobilized metal-affinity chromatography. For gene amplification, suitable forward (5'- GCGCATATGATGGCGGTAACTATCACCCACCTG -3' with *Nde*I (underlined) restriction site) and the reverse (5'-GTTGCTCGAGTCA**ATGATGATGATGATGATG**TGCGCCCAAGATCTGGCT -3' with the *Xho*I (underlined) restriction site and 6xHis-tag (bold)) primers were used in polymerase chain reaction (PCR) (Figure 6.1A). The amplified gene was double digested with *Nde*I and *Xho*I restriction enzymes before inserting into the pET22b(+) vector. The ligation of the inserted gene was performed using T4 DNA ligase. Subsequently, *E. coli* DH5α competent cells were transformed with the recombinant plasmid containing the gene *mhuP* for plasmid amplification. The resultant recombinant construct was further confirmed by double digestion (Figure 6.1B) followed by DNA sequencing.

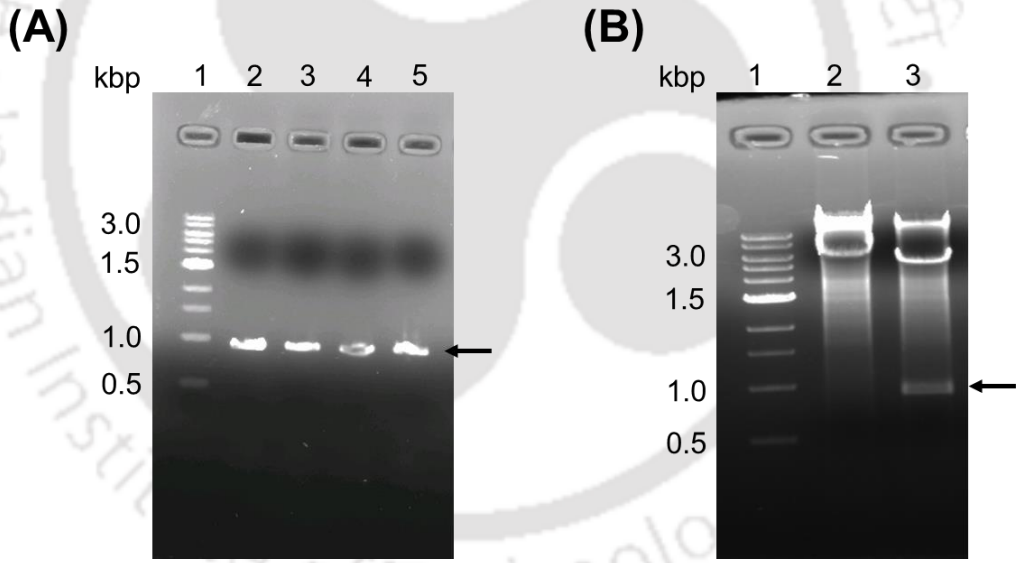


Figure 6.1. Cloning of the gene *mhuP*. (A) Amplification of the gene MhuP at 67.1°C using static PCR (lane 1: DNA ladder, lane 2-5: amplified gene). (B) Clone confirmation of the gene *mhuP* by double digestion of the vector pET22b(+) carrying the gene *mhuP* by *Nde*I and *Xho*I restriction enzymes (lane 1: DNA ladder, lane 2: negative control and lane 3: insert pop out after double digestion of the plasmid pET22b(+)).

6.2.2. Protein overexpression and purification

The *Escherichia coli* BL21 (DE3) competent cells (Novagen, India) were transformed with the recombinant constructs of the gene *mhuP* for protein overexpression. The transformed cells were grown at 37°C in 100 mg ml^{-1} ampicillin supplemented Luria-Bertani (LB) medium till an optical density (OD) of 0.6–0.8 at 600 nm. Subsequently, the cells were induced with 0.5 mM of Isopropyl ß-D-1-thiogalactopyranoside (IPTG) and further incubated at 37°C for 4 hrs to get the maximum overexpressed protein (Figure 6.2A). However, at this condition, the protein was found to be insoluble and therefore, further optimizations were carried out (Figure 6.2B). Finally, an IPTG induction at a concentration of 0.5 mM along with 3% alcohol followed by incubation for an additional 18 hrs at 18°C worked successfully for MhuP protein solubility (Figure 6.2C). Following this, the overexpressed cells were harvested by centrifugation at 3836g for 15 min and resuspended in the lysis buffer A containing 20 mM Tris–HCl (pH 7.5), 5 mM imidazole, 150 mM NaCl, 5% glycerol, 3 mM β-mercaptoethanol (β-ME) and 1 mM phenylmethylsulfonyl fluoride (PMSF). The resuspended cells were incubated on ice for 1 h and then disrupted using a sonicator. The sonicated lysate was then subjected to centrifugation at 18514g for 1 h at 4°C to remove the insoluble debris. The supernatant fraction was applied to a pierce centrifuge column (Thermo Fisher Scientific, USA) packed with Ni-nitrilotriacetic acid (Ni-NTA) affinity resin (Qiagen, Hilden, Germany) pre-equilibrated with buffer A. After 3 hrs of incubation at 4°C, the column was gradiently washed with 10 column volumes each of buffer B (20 mM Tris–HCl (pH 7.5), 150 mM NaCl, 5% glycerol, 1 mM PMSF and 3 mM β-ME) followed by 10 column volumes of buffer C (20 mM Tris–HCl (pH 7.5), 10 mM imidazole, 150 mM NaCl, 5% glycerol, 1 mM PMSF and 3 mM β-ME) followed by a final wash using 10 column volumes of buffer D (20 mM Tris–HCl (pH 7.5), 20 mM imidazole, 150 mM NaCl, 5% glycerol, 1 mM PMSF and 3 mM β-ME). The recombinant protein(s) bound to the Ni-NTA resin was then eluted with buffer E (20 mM Tris–HCl (pH 7.5), 250 mM imidazole, 150 mM NaCl, 5% glycerol and 1 mM PMSF) (Figure 6.2D). The eluted fractions were pooled together and subjected to gradient dialysis to remove imidazole against the final protein buffer F (20 mM Tris-HCl (pH 7.5), 150 mM NaCl and 5% glycerol). The dialyzed protein MhuP was concentrated in a Vivaspin turbo 15 centricon (10 kDa cutoff; Sartorius, Göttingen, Germany) up to a concentration of ~27 mg ml^{-1} (Figure 6.2E). SDS-PAGE analysis was performed after each step of protein

purification to check the purity of the protein (Figure 6.2E). The protein concentration was calculated using the theoretical extinction coefficient (ε_{280}= 48470 M^{-1} cm^{-1}) after measuring the absorbance of protein(s) at 280 nm.

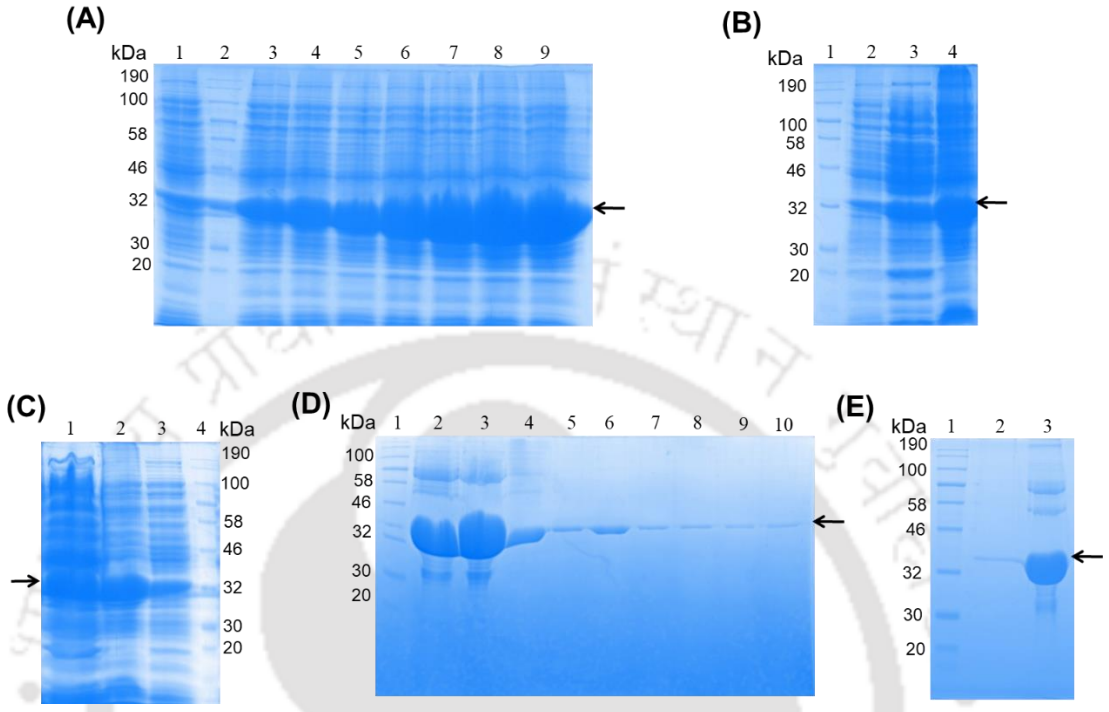


Figure 6.2. Overexpression, solubility and purification of the protein MhuP. (A) Overexpression of the protein MhuP in BL21 (DE3) at 0.5 mM IPTG at a different time interval (lane 1: uninduced or 0 hrs, lane 2: protein marker, lane 3: 1 hr, lane 4: 2 hr, lane 5: 3 hr, lane 6: 4hr, lane 7: 5 hr, lane 8: 6 hr and lane 9: 7 hr) after induction. (B) SDS-PAGE analysis for protein solubility at 37°C for 4 hrs (lane 1: protein marker, lane 2: uninduced sample, lane 3: supernatant and lane 4: pellet fraction). (C) SDS-PAGE analysis for the optimized protein solubility after incubation at 18°C for 18 hrs post IPTG induction (lane 1: supernatant fraction, lane 2: pellet fraction, lane 3: uninduced and lane 4: protein marker). (D) SDS-PAGE analysis for the eluted fractions of the protein using Ni-NTA metal-affinity chromatography (lane1: protein marker, lanes 2-10: eluted fractions of the protein). (E) SDS-PAGE analysis for the concentrated protein (lane 1: protein marker, lane 2: collector fraction and lane 3: concentrated protein).

6.2.3. Mass spectrometric analysis

The exact molecular weight of the protein MhuP was determined using matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF). For MALDI_TOF analysis, the homogeneously purified MhuP protein was mixed with the matrix in a (protein:matrix) ratio of 1:2, 1:3 and 1:5. Since the theoretical molecular weight of the protein MhuP is more than 10 kDa, Sinapinic acid, which was prepared in acetonitrile and 0.1% trifluoroacetic acid in a ratio of 30:70, respectively, to a concentration of 10 mg ml^{-1} was used as a matrix. Sinapinic acid was dissolved in the mentioned solvent by sonicating the solution for 30 mins followed by centrifugation at 12000g for 10 mins, to remove the excess crystals formed by the matrix. Further, the prepared matrix was mixed with protein samples in varying concentrations of 1, 5 and 10 mg ml^{-1} in the ratios specified above. 2 µL spots of the MhuP protein-matrix mixture was made on the ground steel MALDI plate grooves and air-dried. Post drying, the drops were ionized in a Bruker autoflex speed and the data weres plotted using an in-built Flex control analysis software.

6.2.4. Crystallization and co-crystallization of the protein MhuP

Preliminary crystallization screening of the purified MhuP protein (~27.0 mg ml^{-1}) was performed using crystal screen and PEG/Ion kits (Hampton Research, USA) by mixing 2 µl each of protein and buffer (1:1 ratio) using microbatch-under-oil technique at 4 and 20°C. Initial crystal hits for the protein MhuP were obtained in a buffer consisting of 0.1 M citrate pH 5.0 and 20% PEG 6000 at 20°C. The obtained crystal hits were reproduced in microbatch-under-oil as well as optimized using hanging-drop vapor-diffusion method (Figure 6.3A). The reproduced crystal at 20°C in microbatch obtained after ~three weeks of incubation was diffracted in a home source X-ray diffractometer and X-ray intensity data was collected.

In parallel, co-crystallization experiments for the protein MhuP with hemin were performed using multiple approaches. For co-crystallization, the hemin was solubilised separately in 100% dimethyl sulfoxide (DMSO) as well as in 10 mM NaOH. Further, the solubilized hemin was freshly prepared in buffer F (20 mM Tris-HCl (pH 7.5), 150 mM NaCl and 5% glycerol) prior to setting up crystal drops. In the first method, the protein was incubated in the original crystallization buffer with varying concentrations

of hemin (5X to 50X). In the second approach, the prepared ligands were directly added at the time of setting up crystal drop at varying concentrations. In another approach, the previously obtained crystals were soaked overnight with varying concentrations of hemin. Although we obtained multiple crystals in co-crystallization experiments, to our dismay, none of them were bound to hemin. Upon failure, crystal screen and PEG/Ion kits from Hampton Research were used for buffer screening to pre-incubated hemin and MhuP protein. Additionally, JCSG Plus and Wizard Cryo screens (Molecular Dimensions, UK) were also screened for co-crystallization experiments with pre-incubated hemin. However, none of the co-crystallized drops forms any crystal. Interestingly, in one of the soaking experiment where the original crystal was obtained in a buffer having 0.2 M lithium sulfate monohydrate, 0.1 M sodium acetate trihydrate pH 4.6 and 30% w/v PEG 8000 to co-crystallize hemin (Figure 6.3B), we were able to collect an X-ray intensity data with a different space group *C*2.

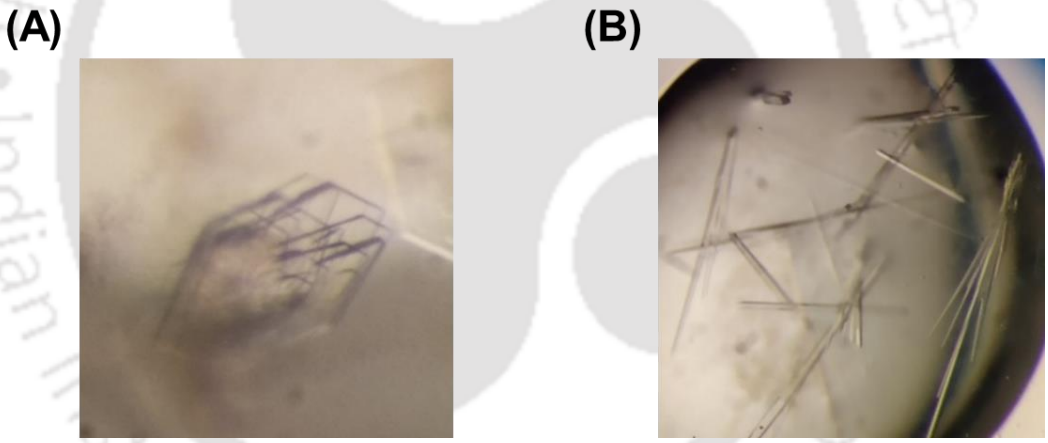


Figure 6.3. Crystallization of the protein MhuP. (A) Protein crystals of MhuP in microbatch and (B) Protein crystals of MhuP after soaking with hemin overnight.

6.2.5. Data collection, processing and structure determination

X-ray intensity diffraction data for MhuP protein crystals were collected at -173°C using the home source Rigaku MicroMax-007 HF diffractometer (operated at 40 kV and 30 mA) and R-Axis IV++ image-plate detector available at the Central Instrument Facility (CIF) of Indian Institute of Technology Guwahati, India. The X-ray data sets for all the crystals were collected with 1° oscillation and 300 s exposure time while

keeping the detector at a distance range of 150-200 mm. Processing and scaling of all the X-ray data sets were performed using the programs iMosflm (Battye et al., 2011) and Aimless (Evans and Murshudov, 2013) and the intensities were converted to structure factors using the module ctruncate embedded in the CCP4 package (Evans, 2011). The structure solution of the protein MhuP was determined by the molecular replacement method using the program Phaser (McCoy et al., 2007). The crystal structure of MhuP has already been deposited under the program TB Structural Genomics Consortium (TBSGC) at a resolution of 2.2 Å with PDB id 4PM4. Hence, to solve the structure for the X-ray intensity data obtained from this study, the already submitted structure was used as a template. A total of 5% reflections were kept aside from the data set for R_{free} calculations (Kleywegt and Jones, 1997). For all the models, the program Coot (Emsley et al., 2010) was used for model building, followed by structural refinement of the models using the program *REFMAC*5 with default parameters (Murshudov et al., 2011). Although a clear *Fo−Fc* map was observed for the bound ligands after the first cycle of refinement for all the complexed structures, firstly, the protein atoms were fitted into an electron density countered at 3.0σ and 1.0σ for the *Fo−Fc* and *2Fo−Fc* maps, respectively. Subsequently, the electron densities observed for water and other molecules present in the protein and crystallization buffer were modeled. Structure refinement was performed with each cycle of a model building using the same set of parameters. The geometrical parameters for all the refined structures were checked and validated by the program PROCHECK and MolProbity (Laskowski et al., 1993; Williams et al., 2018). The refinement and validation statistics for all the refined models are provided in Table 6.1. The three-dimensional atomic factors and structure factors of all the refined structures are deposited in the RCSB Protein Data Bank. The molecular graphic figures used in this study are prepared using the program PyMol (Molecular Graphics System, Version 2.1.1 Schrödinger, LLC).

Table 6.1. X-ray crystallographic data collection and refinement statistics for the protein MhuP. The values in parenthesis represent the statistics for the last resolution shell.

Parameters	MhuP (Form I)	MhuP (Form II)
Wavelength (Å)	1.5418	1.5418
Temperature (K)	100	100
Space group	$P2_1$	$C2$
Unit-cell parameters (Å, °)	a=93.34, b=39.41, c=94.67, α =γ=90, β=108.34	a=202.38, b=39.67, c=92.97, α=γ=90, β=113.38
Resolution (Å)	55.03-1.80 (1.84-1.80)	53.19-2.01 (2.06-2.01)
No. of observed reflections	291219 (16183)	178682 (10728)
No. of unique reflections	61369 (3564)	45599 (3243)
Mn(I) CC(1/2)	0.998 (0.819)	0.989 (0.896)
Completeness (%)	100.0 (99.8)	99.0 (96.1)
V_M (Å^3 Da^{-1})	2.61	2.71
Solvent content (%)	52.94	54.59
Mosaicity (°)	0.30	0.60
Mean I/σ(I)	14.1 (3.2)	8.7 (3.6)
R_{merge}† (%)	6.5 (40.8)	9.4 (26.3)
R_{pim} (%)	5.2 (33.2)	8.3 (22.7)
R_{meas} (%)	8.3 (52.8)	12.6 (34.8)
Multiplicity	4.7 (4.5)	3.9 (3.3)
R_{work}/R_{free} (%)	15.49/18.86	17.14/21.52
No. of subunits in ASU	2	2
Protein atoms	4530	4514
Water molecules	696	621
Other molecules	9	8
Deviation from ideal geometry		

Bond length (Å)	0.014	0.014
Bond angles (°)	1.809	1.785
Average *B*-factor (Å^2)		
Protein atoms	15.14	13.16
Water molecules	31.91	29.03
Ramachandran plot (%)		
Favoured	96.64	96.15
Allowed	3.36	3.68
Remaining	0.00	0.18
PDB id	**7FHM**	**7FHP**

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

6.2.6. Circular dichroism analysis

The Far-UV circular dichroism (CD) spectra of the protein MhuP were recorded between 190 to 270 nm wavelength and 1 mm optical path length quartz cuvette using the Jasco J-1500 spectrometer (Jasco, Germany). The protein of MhuP prepared in a buffer F (20 mM Tris-HCl pH 7.5, 150 mM NaCl and 5% glycerol) was diluted into milliQ grade water with a final concentration of 2.5 µM. For measuring the spectra, the parameters like response, sensitivity and scan speed were fixed as 2 s, 100 milli degrees and 100 nm min^{-1}, respectively. Each spectra measurement includes an average of three scans and subtraction from blank containing the same buffer as the reference. Analysis of the spectra was performed using Spectra Manager software provided by JASCO. Using the optimized spectral profile obtained for the protein MhuP, the thermal denaturation profile was analyzed. To monitor the thermal unfolding curves, the protein sample of MhuP was subjected to the measurement of the CD signals at a varying temperature range of 20-120°C with a heating rate of 2°C min^{-1}. The changes in the secondary structure of the protein MhuP were analyzed between 190 to 270 nm wavelengths. The final denaturation profile of the protein MhuP was plotted as a function of temperature and wavelength using the software Origin (version 9.6).

6.2.7. Isothermal titration calorimetry

The binding affinity of probable ligands for the protein MhuP was measured using isothermal titration calorimetry (MicroCal ITC200; GE healthcare) method. The protein MhuP and the ligands used for ITC experiments were prepared in the dialyzed buffer F. 40 µM of the protein MhuP was filled in the sample cell and titrated against 4 mM of different ligands such as hemin, ferrichrome, $MgCl_2$, $MnCl_2$, $CoCl_2$ and $NiCl_2$. For ITC, two different batches of hemin were prepared by solubilizing it in 100% DMSO and another batch in 10 mM NaOH. The solubilized hemin was then prepared to a required concentration using buffer F. For ferreted ferrichrome, equimolar concentrations of $FeCl_3$ (freshly prepared) and ferrichrome were added in buffer F whereas, other ligands were directly prepared in buffer F. In ITC experiments, each injection consists of 1.5 µl of ligand injected 24 times (excluding 0.4 µl pre-injection) with a duration of 120 s and the sample cell was stirred at 250 rpm. The thermodynamic parameters of the protein MhuP were measured at 25°C. A control experiment to measure the heat of dilution was also performed for each ITC experiment by titrating different ligands into the dialyzed buffer F. The obtained heat of dilution in each control experiment was subtracted from their respective protein-ligand titration to dictate the real ligand binding affinity. The integration of the reaction heat was performed using the program Origin (v 7.0) provided with the MicroCal software.

6.2.8. Difference absorption spectroscopy

To check the binding of heme with the protein MhuP, difference absorption spectroscopy (DAS) was performed. For DAS, fresh hemin solutions were prepared first by solubilizing hemin in 100% DMSO and then making working concentrations of hemin using buffer F. An equimolar amount of hemin was added to the protein MhuP at a concentration of 25 µM and incubated at room temperature for 10 min. For DAS, heme-binding was monitored using an infinite M200PRO plate reader (TECAN). An appropriate blank experiment using only hemin solution prepared in buffer F was also performed. The actual heme-binding was obtained by subtracting the free hemin spectra from the protein incubated hemin spectra.

6.2.9. Fluorescence spectroscopy

To identify the physiological ligand(s) for the protein MhuP, probable ligands based on the literature and an earlier *in silico* analysis were screened using fluorescence spectroscopy. The protein MhuP possesses 22 aromatic amino acid residues, out of which eight are tryptophan residues. Out of these eight tryptophans, four are present at its active site and thus, can be utilized to measure the changes in intrinsic fluorescence emission. For fluorescence studies, 5 µM of protein was used and incubated with increasing concentrations of the ligands viz., hemin, ferrichrome and glycyl glutamate. For all the three ligands, a concentration range of 25 to 250 µM was used with an increment of 5X. The protein and ligand molecules were prepared in buffer F except for hemin, which was earlier solubilized with 100% DMSO. The measurements were performed in a 10 mm path-length quartz cuvette for both protein and protein-ligand complex at room temperature (25°C) in a FluoroMax-4 spectrofluorometer (Horiba Scientific, Japan). Both the protein and protein-ligand complexes were excited at a fixed excitation wavelength of 285 nm with a slit width of 5 nm and monitored for emission spectra from 300 to 500 nm. For each spectrum, three averaging scans were recorded and subtracted with their respective blank to obtain the actual fluorescence. Ligand binding or the change in the aromatic environment or the protein was measured by estimating the quenching values of protein fluorescence attributed to the energy transfer.

6.2.10. Molecular docking and other bioinformatic analyses

The conserved residues between the proteins MhuP and its closest homolog PiaA from *Streptococcus pneumoniae* (*Sp*PiaA) were identified by performing a pairwise sequence alignment using the program Clustal Omega (Sievers and Higgins, 2014). The alignment was further rendered using the web tool ESPript 3.0 (Gouet et al., 2003). Structural homologs of the protein MhuP were identified by using the web-server DALI (Holm and Rosenstrom, 2010). The three-dimensional atomic coordinates of the homologous proteins were retrieved from the Protein Data Bank (PDB). The binding affinities of the heme molecule with the protein MhuP were estimated by molecular docking analysis performed using a freely available program AutoDock v4.2 (Morris et al., 2009). For that, the three-dimensional atomic coordinates of heme were retrieved from its protein-bound structures available at the RCSB Protein Data Bank (Berman et

al., 2000). Before each docking experiment, wherever applicable, the hydrogen atoms were added to the protein and ligand, using the module available in the program AutoDock. Subsequently, the partial charges were assigned to the protein atoms using the Gasteiger charge algorithm (Gasteiger and Marsili, 1980). Similarly, the partial charge for iron (Fe), was assigned as +1. To perform the blind docking of the ligands to the protein, a grid size of 126x126x126 with a grid spacing of 0.375 Å was generated by taking the center of mass of the targeted protein as the grid center. The rigid molecular docking was performed by keeping the protein atoms as rigid and the ligand atoms as flexible about the rotatable bonds. For each molecular docking experiment, a total of 2000 runs of the Lamarckian genetic algorithm (LGA) was set up. The docked ligand conformations were clustered with an rmsd cut-off of 2.0 Å.

For protein-protein docking, the freely available online docking tool ClusPro was used (Kozakov et al., 2017). In ClusPro, the protein DppA was submitted as a receptor molecule, whereas the heme-bound MhuP protein was submitted as ligand molecule, keeping a default set of parameters provided by the program. Although ClusPro gives multiples models of protein-protein complex as results, the balanced results provide 10 clusters. The results in the 2nd cluster having 75 similar confirmations was utilized for analysis. The atomic interactions between the protein and the docked ligand(s) were identified using the program Coot (Emsley et al., 2010). The molecular interactions of docked ligands and other structural figures were prepared using the program PyMOL.

6.3. RESULTS

6.3.1. The protein MhuP is stable at both compositional and conformational levels

For a protein to be crystallized, the protein should possess two major flavors of stability, (i) compositional stability and (ii) conformational stability (Deller et al., 2016). Compositional stability can be achieved at the time of protein purification by maintaining the same species within the crystallization experiment. This chemical homogeneity of the protein can be monitored by SDS-PAGE analysis and mass spectrometry (Deller et al., 2016). In section 6.2.2., the SDS-PAGE analysis of the concentrated protein was analyzed for its homogeneity (Figure 6.2E). With the homogenously purified protein, mass spectrometry was performed using MALDI-TOF

to determine the molecular weight of the protein in order to verify its compositional stability. The theoretical molecular weight of the protein MhuP was calculated to be 32321.89 Da. The mass spectrometric analysis using MALDI-TOF suggested the molecular weight of the protein is ~32322 Da, which suggested that the protein MhuP is compositionally stable (Figure 6.4A).

Further, to determine the conformational stability, secondary structural components, as well as the thermal stability of the protein was analysed using circular dichroism. A normal CD spectra analysis of the protein MhuP suggested that the protein is comprised of ~27% α-helices and 20% β-sheets (Figure 6.4B). To probe the temperature effect, thermal ramping of the protein was performed at different temperatures ranging from 10 to 110°C. The results exhibit that the protein MhuP is stable up to 50°C and starts denaturing afterward (Figure 6.4C). Being a substrate-binding protein from a mesophilic bacterium *M. tuberculosis*, thermal stability till 50°C suggests significant conformational stability of the protein MhuP.

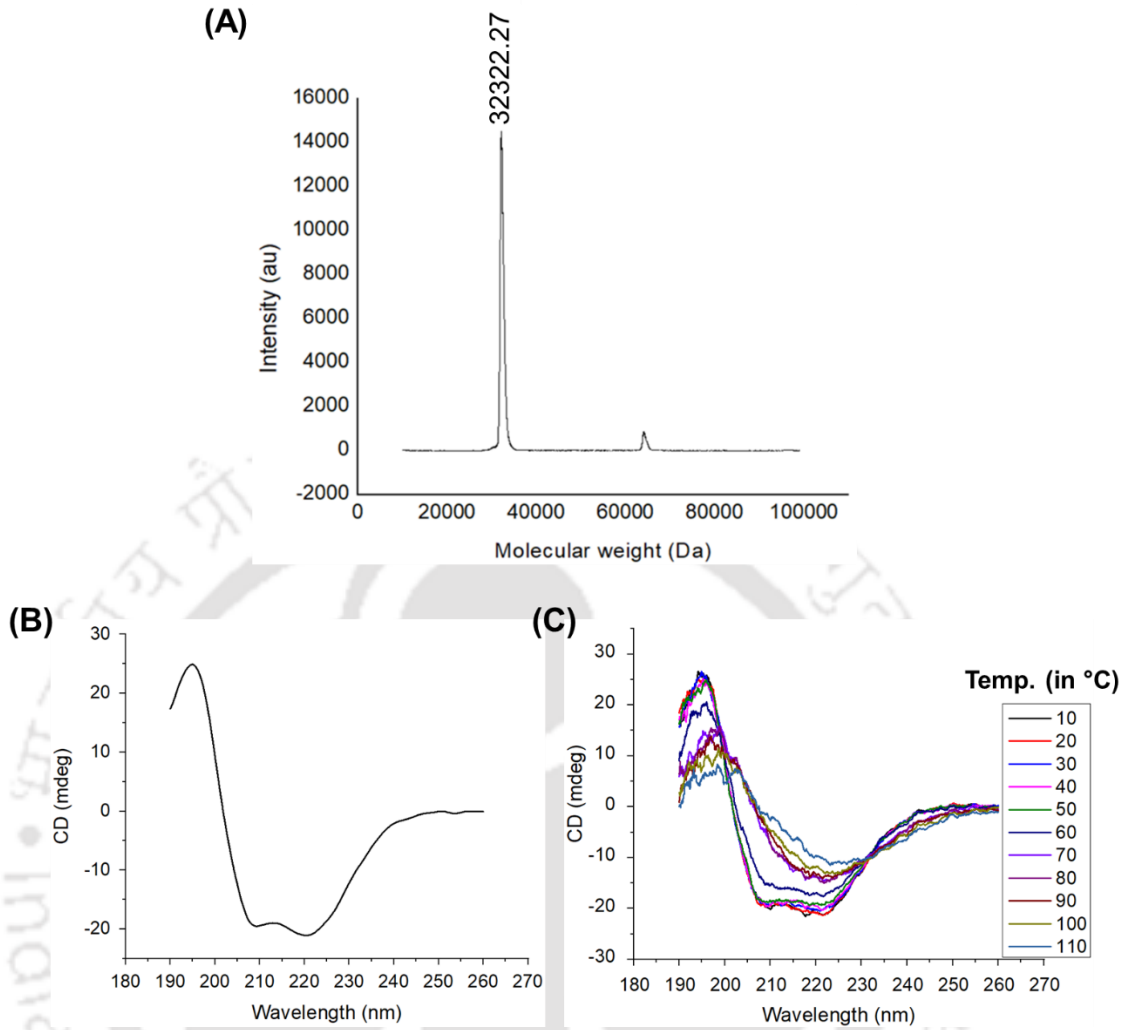


Figure 6.4. Compositional and conformational stability of the protein MhuP. (A) Mass spectra of the protein MhuP using MALDI-TOF analysis suggesting the molecular weight of the protein is 32322.27 Da which is equivalent to the theoretically calculated molecular weight. (B) A normal CD spectrum for the protein MhuP providing its secondary structural content. (C) Thermal melting plots for the protein MhuP. The CD spectra were measured for a wavelength range of 260-170 nm. The CD spectra obtained at temperatures ranging from 10 to 110°C with an interval of 10°C is overlaid together to identify the T_m at which changes in the secondary structure of the protein occurred. The initial spectra at 10°C are represented in black line followed by spectra for other temperatures in different colors.

6.3.2. Overall structure of the protein MhuP

Since, at the time of cloning, the initial amino acid residues from 1-39 were truncated, the mature protein construct used for crystallization consisted of 291 residues (amino acid residues 40–330). Hereafter, the truncated form of the protein MhuP is used for amino acid residue numbering. The crystal structure of the protein was solved in $P2_1$ space group at a resolution of 1.8 Å. Akin to the members of cluster A SBPs, the overall three-dimensional structure of the protein MhuP consists of two globular domains, N- and C-terminal domains (NTD and CTD) linked by an α-helical hinge region comprising of Ala143-Arg163 (Figure 6.5A). Architecturally, the protein MhuP is comprised of six β-strands flanked by six α-helices in the N-terminal domain, whereas five β-strands flanked by nine α-helices in the C-terminal domain (Figure 6.5B). Both the NTD and CTD are connected by a long helical hinge α7. The structure of the protein MhuP was obtained in an unliganded open state (Figure 6.5A and 6.5B). Since the structure of the protein was solved in its apo form, several co-crystallization trials were performed using probable ligands for the protein MhuP. In one of the experiments, 10 mM of hemin was soaked with one of the obtained crystals of the protein MhuP for overnight, for which X-ray intensity data was collected. We solved the crystal structure for the protein MhuP from the collected X-ray data in *C*2 space group; however, the density for hemin was not observed in the density maps. Altogether, the crystal structure of the protein MhuP was solved in two different space groups, i.e. $P2_1$ and *C*2. The overall three-dimensional structure of both the structures looked similar with an rmsd of 0.16 Å (Figure 6.5C). For both the space groups, the asymmetric unit consists of two protomers. To identify the difference in the structure of the protein MhuP solved in different space groups, their molecular arrangement in the unit cell was analyzed. In the case of $P2_1$ space group, in a unit cell two molecules of the protein from different asymmetric units i.e., chain A from two different asymmetric units were arranged with cell dimensions of a = 93.34, b = 39.41 and c = 94.67. One chain of the protein from one asymmetric unit is rotated by 180° with respect to each other chain from another asymmetric unit (Figure 6.5D). In contrast to $P2_1$, the crystal structure of the protein MhuP in *C*2 space group is comprised of eight molecules from four asymmetric units in a unit cell and are rotated 180° with respect to each other (Figure 6.5E). In spite of differences in the unit cell arrangement of the molecules, the arrangement of molecules in an asymmetric unit looked identical in both the space groups. An in-depth analysis

of the molecules arranged in the asymmetric unit suggested that the Chain B of the protein is rotated by 180° with respect to chain A in the asymmetric unit. However, in the protein structure solved in $P2_1$ space group, the CTD of chain B is more inclined towards NTD of chain A whereas, in $C2$, the CTD of chain B is comparatively less inclined towards NTD of chain A (Figure 6.5F and 6.5G).

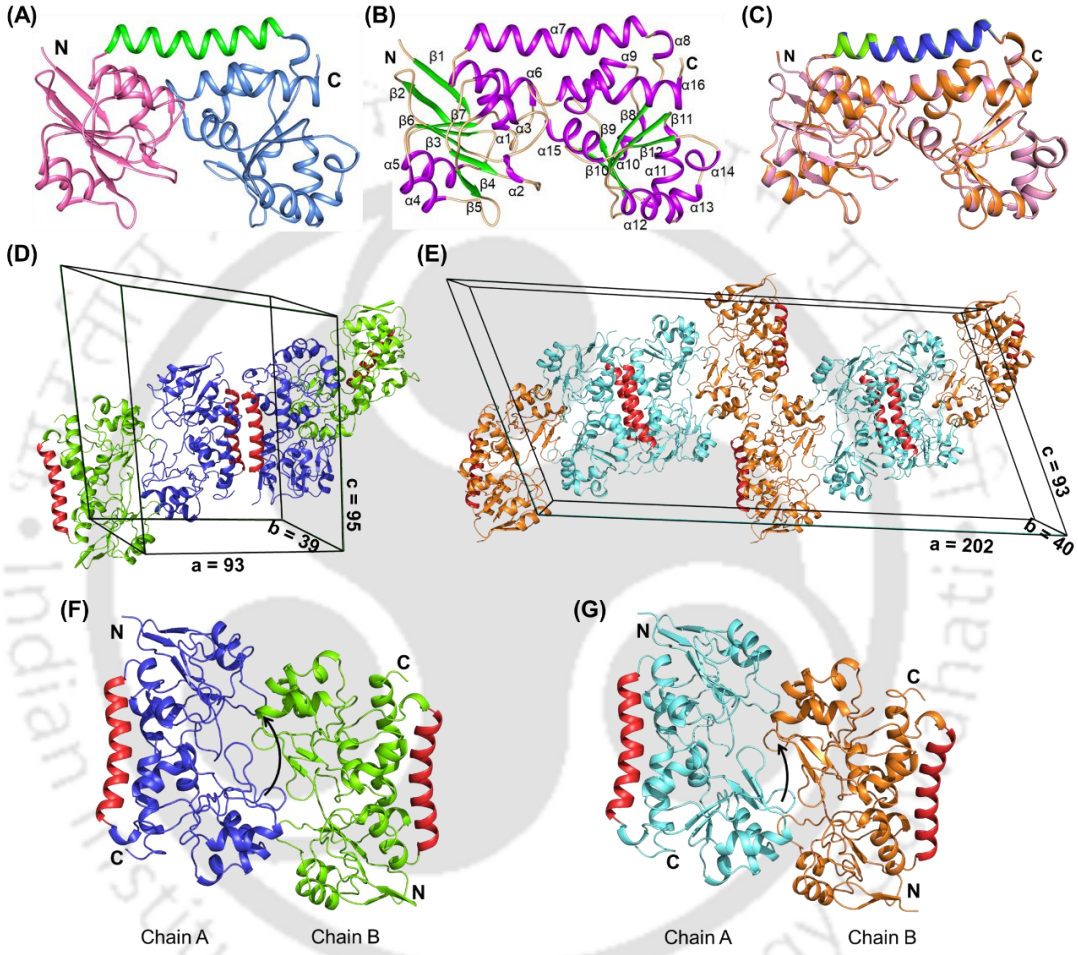


Figure 6.5. Overall three-dimensional structure and differences in $P2_1$ and $C2$ space group structures of the protein MhuP. (A) The overall three-dimensional structure of the protein MhuP. The N- and C-terminal domains are represented in pink and light blue, with the connecting long α-helix in green. (B) The α-helix and β-sheet content of the protein MhuP. The α-helices are represented in violet, while the β-sheets are in green with the connecting loops in wheat. (C) Superposed image of the structure of the protein MhuP solved in $P2_1$ and $C2$ space groups indicating no overall conformational change in them. The structure of MhuP solved in $P2_1$ and $C2$ space groups are shown

in pink and orange with the connecting α-helix in green and blue, respectively. (D, E) Crystal packing of the protein MhuP solved in $P2_1$ and $C2$ space groups. One unit cell with one layer of molecules is shown. In $P2_1$, one molecule each from two asymmetric units is packed in a unit cell and is rotated 180° with respect to each other. In $C2$, eight molecules from four asymmetric units are packed in a unit cell and are rotated 180° with respect to each other. Chains A and B of an asymmetric unit of the protein are shown in blue & green and cyan & orange in $P2_1$ and $C2$ space groups, respectively, with the α-helix hinge region in red. The approximate unit cell dimension (a, b and c) in Å for both the space groups are labeled for clarity. (F, G) The arrangement of two molecules (chain A and B) of the protein MhuP in an asymmetric unit in $P2_1$ and $C2$ space groups. Chains A and B of an asymmetric unit of the protein are shown in blue & green and cyan & orange in $P2_1$ and $C2$ space groups, respectively, with the α-helix hinge region in red.

6.3.3. The protein MhuP is evolutionarily closer to Fe^{3+}-siderophore-binding protein

Since the structure of the protein MhuP was solved in apo state and the co-crystallization trials for the protein were unsuccessful, attempts were made to identify the cognate ligand of the protein MhuP. Using the solved structure of the protein, structurally homologous proteins for MhuP were identified using DALI search (Holm and Rosenstrom, 2010) (Table D1). The results suggest Fe^{3+}-siderophore binding proteins from different microorganisms as the closest homologs for the protein MhuP (Table D1). However, in the literature, the protein MhuP was reported to be a heme-binding protein (Mitra et al., 2017). Moreover, Niederweis and co-workers have also hypothesized that the protein MhuP binds and delivers heme to a dipeptide-binding protein DppA from *M. tuberculosis* (PDB id: 6E4D). Therefore, an evolutionary relationship between the proteins MhuP with Fe^{3+}-siderophore-binding proteins, known heme- and peptide-binding proteins were established. The results exhibited that the protein MhuP is closer to Fe^{3+}-siderophore-binding proteins followed by heme-binding proteins (Figure 6.6). The results also suggested that the protein MhuP is distantly related to the peptide-binding proteins (Figure 6.6).

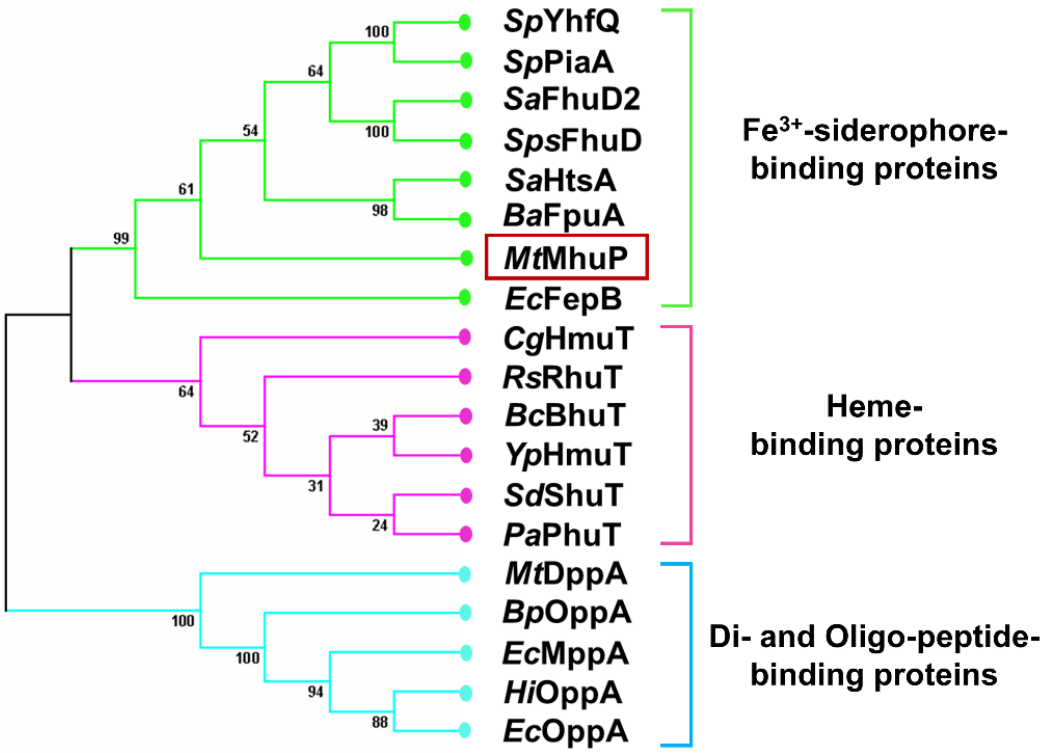


Figure 6.6. Phylogenetic analysis of the protein MhuP with its homologous proteins along with known heme- and di- and oligo-peptide binding proteins. The proteins used to build the phylogenetic tree are *Sp*YhfQ (*Streptococcus pneumoniae*, A0A062WLD9), *Sp*PiaA (*S. pneumoniae*, A0A0H2UPT5), *Sa*FhuD2 (*Staphylococcus aureus*, Q7BGA5), *Sps*FhuD (*St. pseudintermedius*, A0A2P5JFB2), *Sa*HtsA (*St. aureus*, A0A0H3JM89), *Ba*FpuA (*Bacillus anthracis*, Q81L65), MhuP (*M. tuberculosis*, L7N6B2), *Ec*FepB (*Escherichia coli*, P0AEL6), *Cg*HmuT (*Corynebacterium glutamicum*, Q8NTB8), *Rs*RhuT (*Roseiflexus spp.*, A5UZ69), *Bc*BhuT (*Burkholderia cenocepacia*, B4EKB3), *Yp*HmuT (*Yersinia pestis*, A0A0E1NWS3), *Sd*ShuT (*Shigella dysentriae*, O70018), *Pa*PhuT (*Pseudomonas aeruginosa*, O68879), *Mt*DppA (*M. tuberculosis*, I6X811), *Bp*OppA (Burkholderia pseudomallei, Q63ID0), *Ec*MppA (*E. coli*, P77348), *Hi*OppA (*Haemophilus influenzae*, Q4QLH0) and *Ec*OppA (*E. coli*, P23843); the organism name and UniProt id of each protein is provided in the parenthesis.

6.3.4. Active site of the protein MhuP is favourable for hydroxamate-type siderophores

To affirm the similarity of the protein MhuP with Fe^{3+}-siderophore-binding proteins, the electrostatic potential charge distribution at the active site of the protein was compared with Fe^{3+}-siderophore- and heme-binding proteins. From the structural homology search, majorly two types of Fe^{3+}-siderophore-binding proteins viz. carboxylate- and hydroxamate-type Fe^{3+}-siderophores were found to be similar to the protein MhuP (Table D1). The example of carboxylate-type siderophore is staphyloferrin A and hydroxamate-type siderophore in ferrioxamine A and ferrichrome. Therefore, the electrostatic potential charge distribution of both types of siderophore-binding proteins, a heme-binding protein, along with MhuP was calculated. The results exhibit that hydroxamate-type siderophores require a completely positively-charged active-site pocket such as HtsA protein from *Staphylococcus aureus* (*Sa*HtsA, PDB id: 3LHS), which bind staphyloferrin A (Figure 6.7A). Akin to hydroxamate-type siderophores, heme-binding protein, BhuT from *Burkholderia cenocepacia* (*Bc*BhuT, PDB id: 5Y89) possesses a positively-charged active-site pocket (Figure 6.7B). In contrast, carboxylate type siderophore-binding proteins such as YhfQ and PiaA from *Streptococcus pneumoniae* bound to ferrioxamine E and ferrichrome, respectively, (*Sp*YhfQ, PDB id: 5JJ5 and *Sp*PiaA, PDB id: 4HMQ) possess a completely negatively-charged active-site pocket (Figure 6.7C and 6.7D). Akin to carboxylate-type siderophore-binding proteins, MhuP also possesses a negatively charged active site, which favors the binding of carboxylate-type siderophores (Figure 6.7E). To further confirm this, a pairwise sequence alignment of the protein MhuP with its closest homolog *Sp*PiaA was performed to identify the conservation of active-site residues. The results reveal that the ferrichrome-interacting residues Asn83, Trp158 and Arg231 in *Sp*PiaA are found to be partially conserved in the protein MhuP (Trp52, Phe125 and Arg192) (Figure 6.7F). Moreover, the residues forming the active-site pocket in *Sp*PiaA are also observed to be partially conserved except for Phe255, which is replaced by a Gly216 in MhuP (Figure 6.7F). Additionally, the two glutamate residues Glu119 and Glu262, which were reported to interact with the cognate TMD of *Sp*PiaA protein are also replaced by Asp85 and Asp221 in the protein MhuP (Figure 6.7G). The spatial conformation of the active-site residues interacting with ferrichrome in *Sp*PiaA is also found to be partially similar to the protein MhuP (Figure 6.7G). Altogether, the analysis

suggests that the active site of the protein MhuP is more similar to other Fe^{3+}-siderophore-binding proteins.

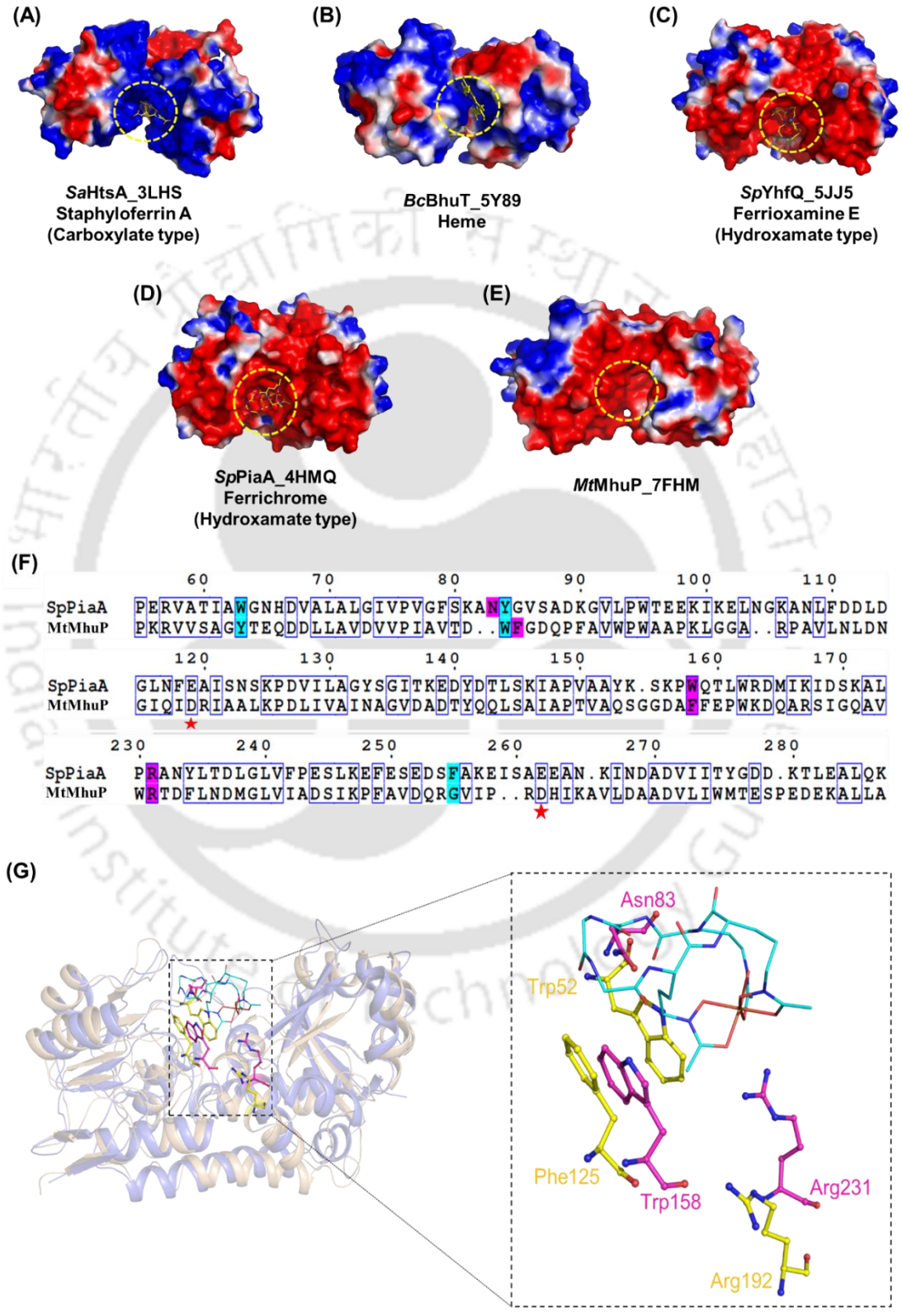


Figure 6.7. Comparison of the active site of the protein MhuP with Fe^{3+}-siderophore and heme-binding proteins. Electrostatic potential charge distribution on the protein (A) HtsA from *Staphylococcus aureus*, (B) BhuT from *Burkholderia cenocepacia*, (C) YhfQ from *Streptococcus pneumoniae*, (D) PiaA from *S. pneumoniae* and (E) MhuP. The active-site pocket in all the proteins is encircled in the yellow dash. The positively- and negatively-charged surfaces are shown in blue (+1 kT/e) and red (-1 kT/e), respectively. The PDB ids and the bound ligands for all the reference proteins are labeled for clarity. (F) Pairwise sequence alignment of the protein MhuP with a ferrichrome-binding protein *Sp*PiaA from *S. pneumoniae*. The ferrichrome binding residues of *Sp*PiaA are conserved in MhuP protein as well and are highlighted in magenta, whereas the residues involved in forming the active-site vicinity in *Sp*PiaA are highlighted in cyan. The red star marks the position of two glutamate residues (Glu119 and Glu262) involved in interaction with the cognate TMD of *Sp*PiaA. (G) Superposition of the protein MhuP over *Sp*PiaA (PDB id: 4HMQ) showing the active-site residues involved in ferrichrome-binding in *Sp*PiaA. The proteins MhuP and *Sp*PiaA are shown in wheat and blue with the active-site residues involved in ferrichrome-binding (in *Sp*PiaA) in yellow and magenta ball-and-stick, respectively, and are labeled in their respective color. The bound ferrichrome molecule to the protein *Sp*PiaA is shown as cyan lines.

6.3.5. Difference absorption spectroscopy suggest the protein MhuP can bind hemin

Earlier analysis and reports from literature suggested that the protein MhuP is capable of binding carboxylate type siderophore or heme molecule (Mitra et al., 2017; Mitra et al., 2019). To clear the doubt whether the protein MhuP can bind carboxylate-type siderophore and/or heme molecule or not, binding of the protein with both the ligands was performed using ITC (Figure D1A and D1B). However, the ITC results suggest that the protein MhuP is unable to bind either of the ligands (Figure D1C-F). Moreover, binding of the protein MhuP with divalent metal ions such as $MgCl_2$, $MnCl_2$, $CoCl_2$ and $NiCl_2$ has also been performed. Interestingly, the protein MhuP shows significant binding isotherms for Co^{2+} and Ni^{2+} ions; however, it doesn't show any binding for Mg^{2+} and Mn^{2+} (Figure D1G and D1H). The binding of the protein MhuP to Co^{2+} and

Ni^{2+} ions is probably due to the presence of 6x His-tag in the protein, which has a high affinity for these metal ions and is therefore used in immobilized metal affinity chromatography.

Although ITC data do not exhibit the binding of heme by the protein MhuP, Niederweis and co-workers have reported the binding of heme to the protein MhuP using surface plasmon resonance (SPR) (Mitra et al., 2017). Therefore, we further examined the heme-binding property of the protein MhuP using difference absorption spectroscopy (DAS). Heme is a porphyrin ring-containing molecule (Figure D1A), which produces a Soret peak. Soret peak or band is an intense peak in the blue wavelength region of the visible spectrum, which arises principally due to an electron dipole movement that allows π-π* transitions; most commonly observed in porphyrin ring containing molecules (Egawa et al., 2001). Subtracting the free hemin spectra from the hemin incubated protein spectra resulted in a characteristic Soret peak at ~416 nm along with a broad Q band peak at ~550 nm, which indicates the binding of heme by the protein MhuP in solution (Figure 6.8A). The DAS analysis for the negative control, i.e. hemin free MhuP protein does not yield any peak at the blue wavelength region (~400-430 nm) of the visible spectrum (Figure 6.8B). These results demonstrate that the protein MhuP is a heme-binding protein.

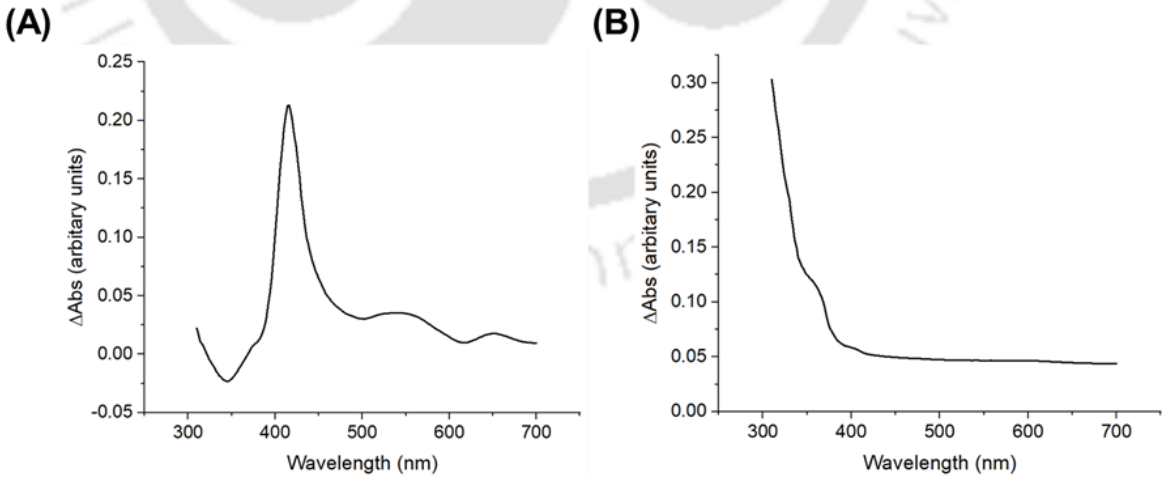


Figure 6.8. Difference absorption spectroscopy (DAS) of the protein MhuP. (A) DAS of hemin incubated MhuP and (B) hemin-free MhuP protein as a negative control. The

free hemin spectra were subtracted from the hemin-incubated MhuP protein spectra at protein concentrations of 25 µM.

6.3.6. The protein MhuP shows specificity towards heme molecules

To reaffirm the binding of the heme molecule to the protein MhuP and to determine whether ferrichrome can bind the protein or not, further confirmation was performed using a fluorescence-based approach. To perform fluoroscence spectroscopic analysis, varying concentrations of hemin (dissolved in 100% DMSO and diluted in buffer F) and ferrichrome in its iron-free and bound forms were prepared and added to the protein MhuP. The results exhibit that the emission maxima (λ_{max}) of the protein MhuP occurs at ~336 nm in the absence of the ligand(s) (Figure 6.9A-C). However, the fluorescence intensity decreases gradually in the presence of the hemin molecule with increasing concentrations suggesting that hemin is capable of quenching the protein MhuP (Figure 6.9A). Additionally, a blue shift of ~10 nm in the λ_{max} in between 336 to 327 nm indicative of a change in the local environment of the tryptophan residues from hydrophilic to hydrophobic (Munishkina and Fink, 2007). In contrast, both ferreted as well as non-ferreted ferrichorme molecules were unable to quench the protein indicating the protein MhuP cannot bind ferrichrome molecules or carboxylate type siderophores as a whole (Figure 6.9B and 6.9C). Moreover, the molar intensity of the fluorescence emission spectrum was also calculated, which suggested a decrease in the fluorescence intensity with increasing concentrations of hemin (Figure 6.9D). The regression coefficient for the plotted graph is very significant, with R-square (COD) of ~0.99 (Table 6.2). With this, the (dissociation constant) K_d value and the stoichiometric ratio (n) of the reaction between the protein MhuP and hemin were also determined. The binding affinity (K_d) for the protein MhuP towards the heme molecule is calculated to be ~389 µM with n value of ~0.7 (Table 6.2). The binding affinity of the protein MhuP with heme has been reported to be ~300 µM using SPR (Mitra et al., 2017). The molar intensity of the fluorescence emission spectrum for the protein MhuP with ferreted and non-ferreted ferrichrome was not determined as the regression coefficient of the plotted graph was very insignificant (Figure 6.9E, 6.9F and Table 6.2). Moreover, since the protein MhuP was reported to bind heme and then deliver it to a peptide-binding protein DppA, binding of the protein MhuP to a dipeptide glycyl glutamate was

also checked to know if the protein is also capable of binding peptide molecules using fluorescence spectrometry. However, the dipeptide molecule, i.e. glycyl glutamate is unable to quench the protein MhuP suggesting no changes in the local environment of the tryptophan residues of the protein (Figure D2).

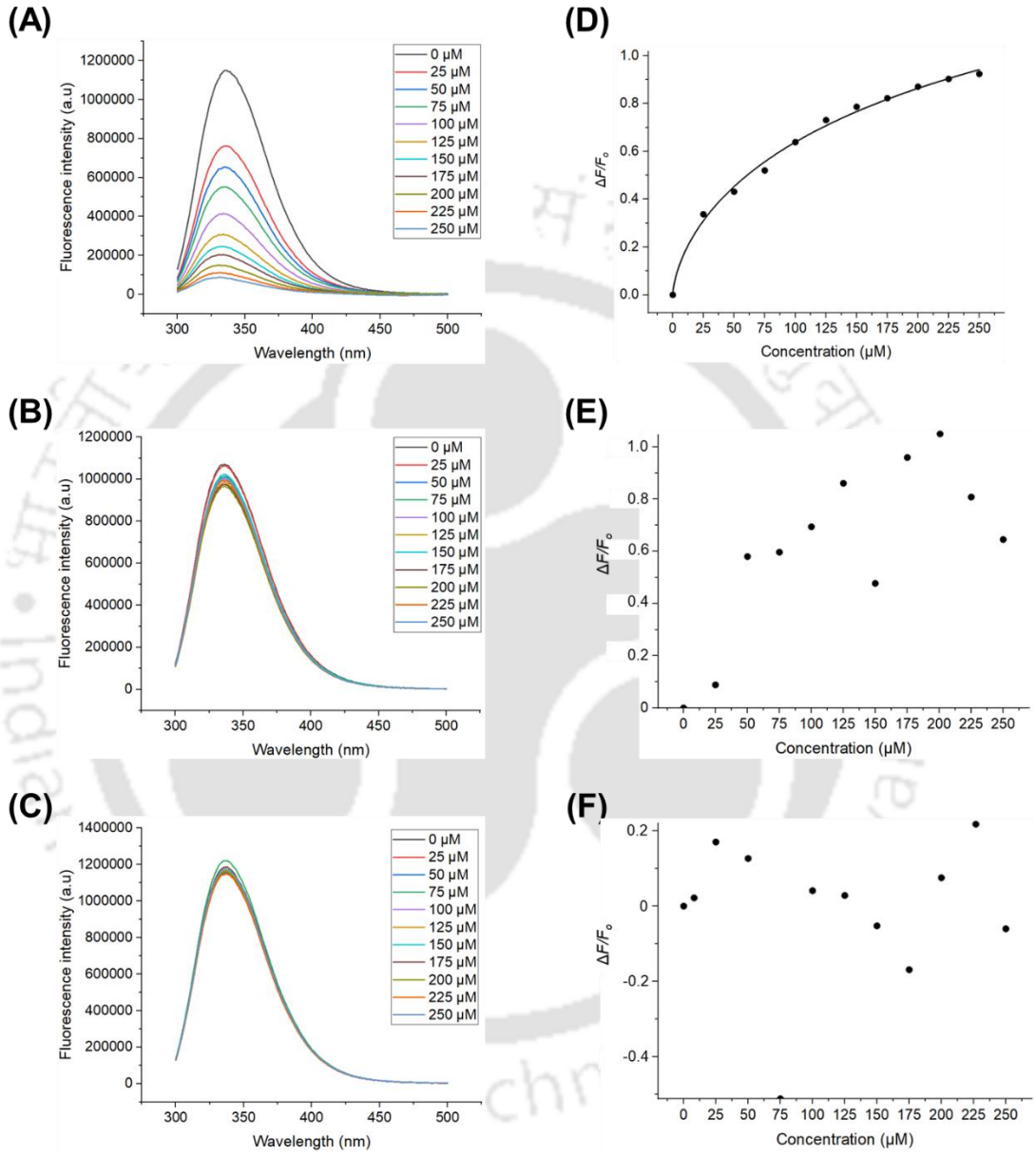


Figure 6.9. Fluorescence study of the protein MhuP with hemin and ferrichrome. (A) Fluorescence emission spectra of the protein MhuP in the absence (black) and presence of (A) hemin, (B) ferreted ferrichrome and (C) non-ferreted ferrichrome. The change in fluorescence spectra of the protein MhuP due to the quenching effect of hemin and ferreted and non-ferreted ferrichrome has been represented in different colors and is labeled for the varying concentrations of ligands. A saturation isotherm generated from

the fluorescence quenching data by plotting the change in fluorescence intensity at λ_{max} as a function of added ligand for (D) hemin, (E) ferreted ferrichrome and (F) non-ferreted ferrichrome. Saturation upon increasing concentrations of the ligand was achieved by hemin only.

Table 6.2. Dissociation constants (K_d), Hill coefficients (n) and regression coefficient for the fluorescence spectrum for the protein MhuP with different ligands.

Ligand	K_d (in µM)	n	R-square (COD)
Hemin	389.62	0.7	0.99538
Ferreted ferrichrome	-	-	0
Non-ferreted ferrichrome	-	-	0.00322

6.3.7. Docking studies suggest a strong affinity of the protein MhuP towards the heme

Since the protein MhuP exhibits significant binding towards the heme molecule in solution, we performed molecular docking calculations to identify the amino acid residues at the active-site pocket interacting with heme. The docking results corroborate the binding of the protein with heme with a binding energy of -10.38 kcal mol^{-1} (Figure 6.10A). The docked heme molecule is held by the residues Trp52 from the NTD and Gln176, Arg192 and Arg215 from the CTD (Figure 6.10A and Table D1). For known heme-binding proteins, the importance of a tyrosine residue has been shown at the active-site pocket along with a charged residue near the iron-binding tyrosine residue (Naoe et al., 2017). The tyrosine residue interacts with the heme iron, whereas the charged amino acid residue is crucial to maintain the deprotonated state of the iron-binding tyrosine residue (Naoe et al., 2017). For instance, in *Bc*BhuT and *Rs*RhuT (PDB id: 5Y89 and 5GJ3, respectively), the heme iron as well as the heme molecule is coordinated by Tyr87 and Tyr140, respectively. The residues Tyr87 and Tyr140 are maintained in their deprotonated state by Arg89 and Arg143 in *Bc*BhuT and *Rs*RhuT, respectively, upon heme iron binding (Figure 6.10B and 6.10C). Interestingly, a tyrosine residue, Tyr33 is present in the active-site vicinity of the protein MhuP (Figure 6.10D). Since the structure of the protein MhuP is solved in an open unliganded state;

the residue Tyr33 is located ~6.0 Å away from the heme iron in the docked conformation (Figure 6.10D). In order to check if Tyr33 is capable of interacting with the heme iron, we changed its rotameric conformation bringing it a bit closer to the heme iron. Following this, a molecular docking experiment was again performed for the protein MhuP having a changed Tyr33 rotamer with heme molecule. As anticipated, the docking results suggested the interaction of Tyr33 with the heme iron along with the interactions from Trp52, Gln176, Arg192 and Arg215 residues (Figure 6.10E). Surprisingly, the binding affinity of the protein with heme molecule also increased to -12.37 kcal mol^{-1} compared to the earlier docking with untranslated Tyr33 residue of the protein MhuP. Further analysis of the active-site region of the protein MhuP suggested that the residues Glu35 and Gln36 are located near to the residue Tyr33, which can probably interact with it to maintain its deprotonated stated upon heme iron binding in ligand-bound closed conformation.

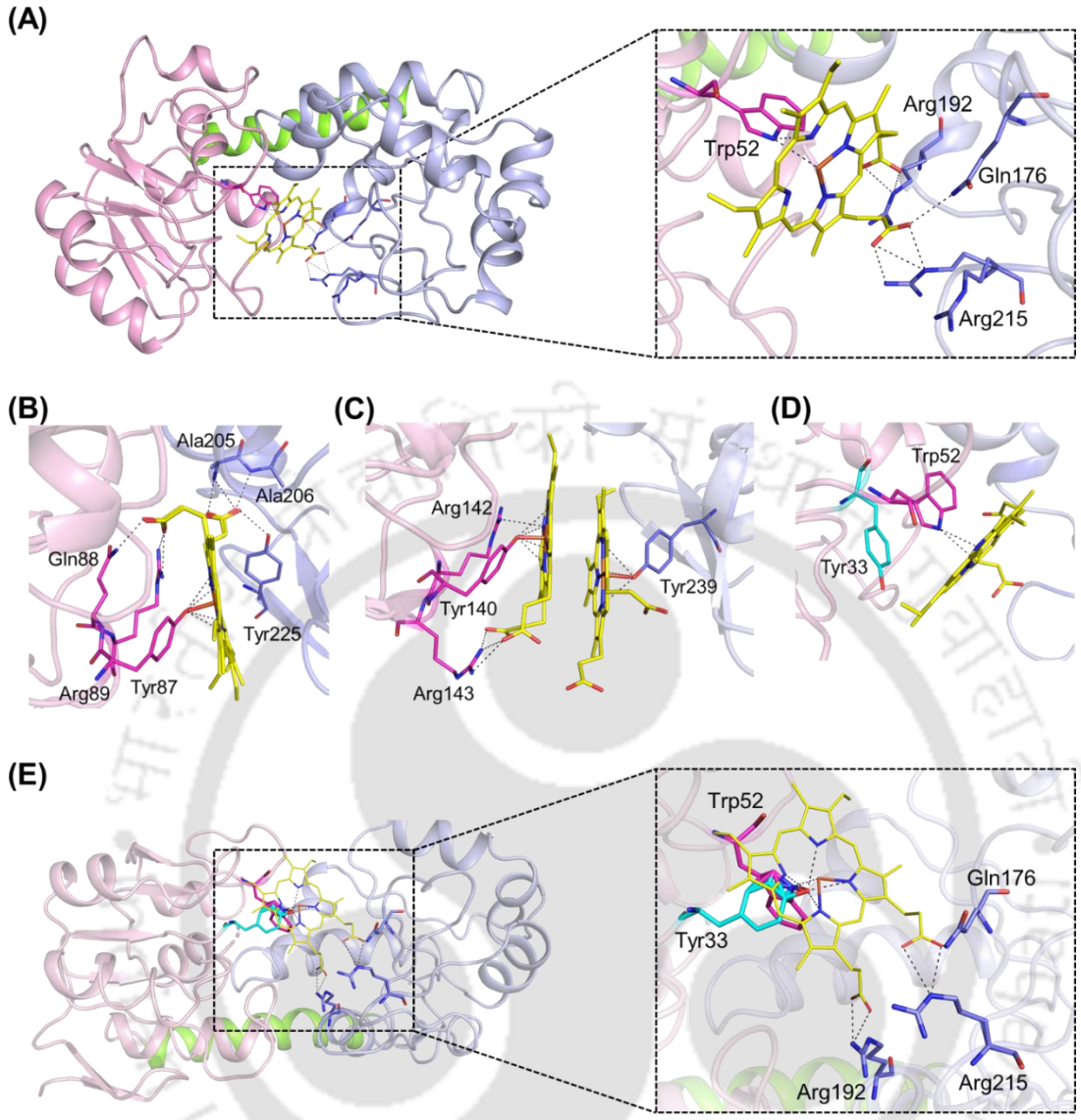


Figure 6.10. Molecular docking studies of the protein MhuP. (A) Molecular docking of the protein MhuP with heme molecule. The active-site residues coordinating with the heme molecule in (B) *Bc*BhuT (PDB id: 5Y89) and (C) *Rs*RhuT (PDB id: 5GJ3) showing the interaction of a tyrosine residue with heme. (D) Location of Tyr33 (untranslated) at the active-site pocket of the protein MhuP docked with heme. (E) Molecular docking of the Tyr33 translated MhuP protein with heme molecule. The N- and C-terminal domains of the protein are represented in light pink and blue, with the helical hinge region connecting them in green. The amino acid residues coordinating the docked heme (yellow stick) in the active-site pocket are shown in the magenta (from NTD) and blue (from CTD) stick model. The translated Tyr33 (from NTD) coordinating with the heme molecule is in the cyan stick model.

6.3.8. Protein-protein docking suggests a transient t of heme from the protein MhuP to DppA

From the earlier analysis, a heme docked conformation of the protein MhuP is achieved from molecular docking. Niederweis and co-workers have earlier proposed that the protein MhuP binds heme and delivers it to the DppA protein, which further translocates it to the DppBCD transporter for heme influx into the cytosol (Mitra et al., 2019). The protein DppA is a dipeptide-binding protein from *M. tuberculosis* belonging to cluster C SBPs (Figure 6.11A). Niederweis and co-workers have shown the binding of heme with DppA using computational docking where the binding was predicted by normal mode analysis (Mitra et al., 2019). The DppA-heme docking results suggest that the heme molecule was held by the residues His131, Ser134, Arg179, Leu477 and Glu481 of the protein DppA (Figure 6.11A). For the transient delivery of heme from the protein MhuP to DppA, the heme-bound MhuP protein have to interact with DppA protein for substrate delivery. To verify this, we performed a protein-protein docking of the heme-bound MhuP and DppA (PDB id: 6E4D) proteins. The results suggest that the heme-bound face of the protein MhuP interacts with the DppA protein at its heme-binding location (Figure 6.11B). As per our docking results, the interaction between these two proteins is mediated by a loop region in DppA (Ala143-Arg151) and Tyr33 of MhuP where Arg151 from DppA interacts with Tyr33 of MhuP protein (Figure 6.11B-C). This suggests that the MhuP bound heme molecule is plausibly delivered to the loop region of DppA (Ala143-Arg151) from where it slides into the heme-binding site of the protein DppA. The heme-bound DppA protein further delivers the heme molecule to the DppBCD transport system for further translocation of the heme into the cytoplasmic region of *M. tuberculosis*.

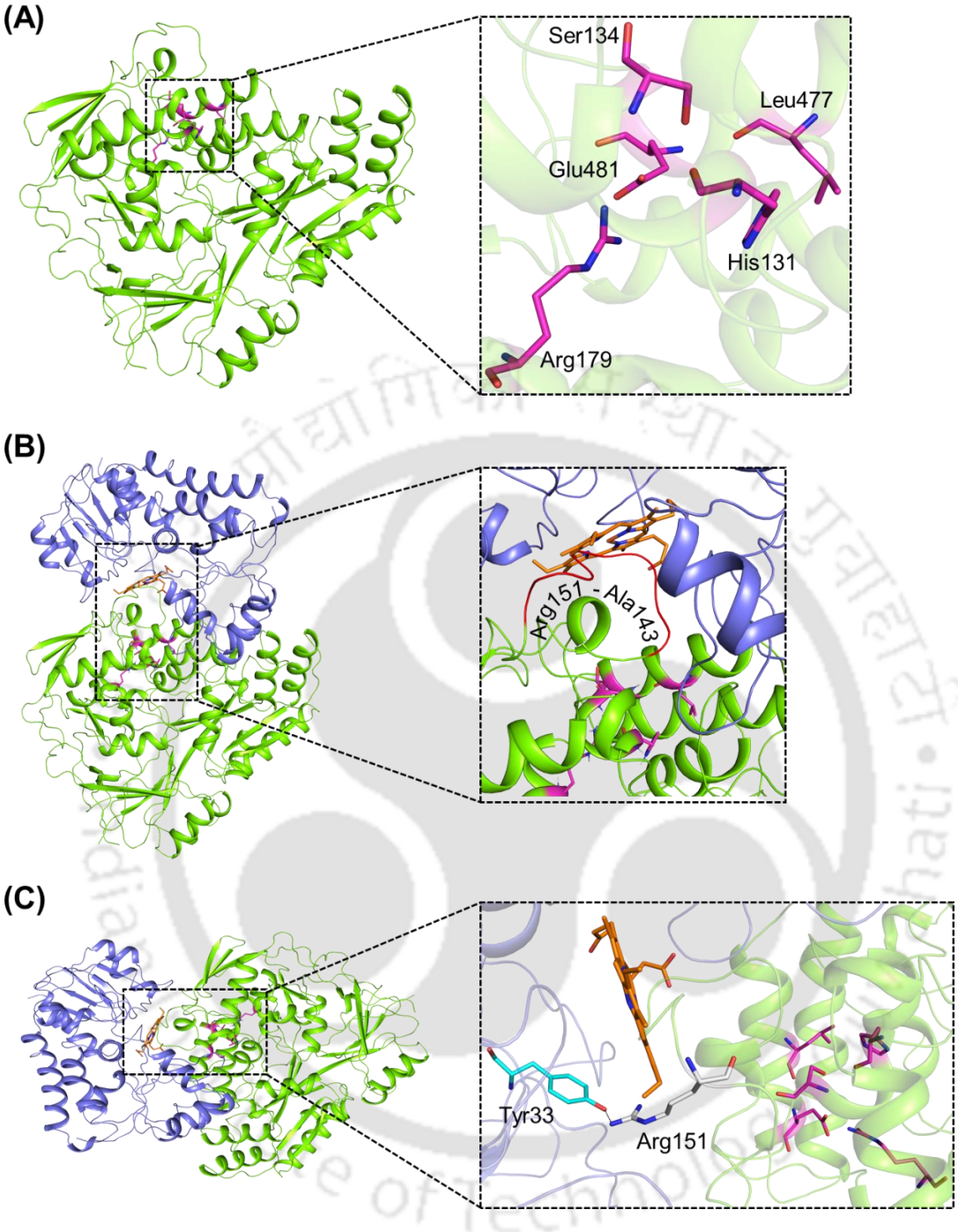


Figure 6.11. Protein-protein docking of the proteins MhuP (heme bound) and *Mt*DppA. (A) Overall three-dimensional structure of *Mt*DppA (PDB id: 6E4D). The amino acid residues reported to interact with the heme molecule are shown in the magenta stick model. (B) Docked conformation (vertical) of the protein MhuP with *Mt*DppA. The loop region (Ala143-Arg151) of *Mt*DppA, which is supposed to interact with MhuP and receives the heme molecule, is represented in red. (C) Docked confirmation (horizontal, 90°) showing the interaction between the proteins MhuP and *Mt*DppA. The

amino acid residues, Tyr33 of MhuP and Arg151 of *Mt*DppA interacting with each other are represented as cyan and grey stick model, respectively. The heme-bound MhuP protein is shown in blue, whereas *Mt*DppA is shown in green with the heme-coordinating residues of *Mt*DppA in magenta. The docked heme molecule with the protein MhuP is represented as the orange stick.

6.4. DISCUSSION

In humans, more than 70% of the iron is bound to heme, making it the most prevalent source of iron in humans (Runyen-Janecky and Role, 2013). Since iron is an essential cofactor for several enzymes which play pivotal roles in numerous cellular and physiological processes, it is required by all domains of life. Hence, to restrict the iron ions to be utilized by pathogens residing inside the host body, they have developed an innate immune system termed as 'nutritional immunity' against such pathogens. However, intracellular pathogens such as *M. tuberculosis* which solely depends on the host system for its iron supplements, have developed eminent mechanisms to sequester iron from the host environment. For instance, *M. tuberculosis* possesses the IrtAB import-export system, which is involved in the transport of (carboxy)mycobatins (Rodriguez and Smith, 2006; Arnold et al., 2020). Another important mechanism of iron transport exploited by *M. tuberculosis* is to scavenge the iron-bound heme molecules inside the host system using an ABC transport system. In this study, we have characterized a heme-binding protein MhuP (ORF id: Rv0265c) from *M. tuberculosis* both structurally and functionally. An operonic analysis for the gene *mhuP* suggested that the protein lacks its cognate TMD and NBD subunits indicating that MhuP is an orphan SBP. The protein MhuP was earlier reviewed as an orphan SBP with unknown function (Sutcliffe and Harrington, 2004). These orphan SBPs are not just remnants of non-functional ABC transport systems but are functional SBPs, which exploit the TMD and NBD subunits of some other ABC transporters encoded elsewhere on the genome (Thomas, 2010). *Pseudomonas aeruginosa* has more than 25 orphan SBPs majority of which are uncharacterized. The murein tripeptide-binding protein MppA from *E. coli* utilizes the transport system of OppBCDF for murein transport (park et al., 1998). Moreover, *E. coli* also possesses other orphan OppA orthologues of unknown function dispersed over the genome (Thomas, 2010).

The overall three-dimensional structure of the protein MhuP is similar to the members of cluster A SBPs having an NTD and CTD linked by a long helical hinge region. Akin to other known heme-binding proteins, the structure of MhuP is suggested to be a member of subcluster A-II SBPs. In an aqueous solution hemin forms dimer even at lower concentrations (10-20 µM) (Schmitt et al., 1993). However, the area and volume of the protein MhuP were observed to accommodate only one heme molecule. This might be the reason for obtaining a heme-bound state of the protein MhuP. Additionally, the solubility of ferreted molecules is an issue at higher concentrations which also probably hinders the co-crystallization experiments of the protein MhuP with hemin.

An *in silico* analysis, as well as bioinformatic results from this study suggested that the protein MhuP is similar to Fe^{3+}-siderophore-binding proteins. The protein MhuP is suggested to be a carboxylate type siderophore binding protein at both sequence and structural level using bioinformatic studies. However, the experimental characterization exhibits that the protein MhuP is a heme-binding protein in this study. The presence of an intense Soret band at 416 nm, including the Q band at ~550 nm in the hemin-incubated MhuP protein, strongly suggests the binding of heme to the protein. However, the intensity of the Soret peak (obtained at 25 µM concentrations) is less compared to the Soret peak obtained for hemin incubated DppA protein (obtained at 10 µM concentrations) by Niederweis and co-workers (Mitra et al., 2019). Additionally, the dissociation constant (K_d) of the protein MhuP towards hemin is ~389 µM compared to the 1.5 µM for DppA protein. Succinctly, the experimental evidence suggest that the binding affinity of MhuP towards hemin is less compared to DppA. A comparatively lower affinity suggests that the protein MhuP has to release the bound heme molecule for its transport via DppABCD transport system. The heme released by MhuP is further received by DppA protein with higher affinity, which is further translocated to the TMD subunits DppBC. This corroborates the hypothesis proposed by Niederweis and co-workers (Mitra et al., 2019) that the protein MhuP binds heme and then delivers it to DppA for further transport of the substrate.

The docking experiments performed for the protein MhuP with heme strongly corroborate our experimental analysis. Additionally, docking of the protein MhuP after

changing the rotameric conformation of Tyr33 with heme enhances the binding affinity of the protein towards heme. This indicates that Tyr33 plays an important role in heme binding in the active-site pocket of the protein MhuP. The role of a tyrosine residue in heme-binding has earlier been reported for many known heme uptake proteins of cluster A-II SBPs (Naoe et al., 2017). In ShuT protein from *Shigella dysentriae* (*Sd*ShuT), a tyrosine residue penta-coordinates the heme molecule and functions as an axial iron ligand (Eakanunkul et al., 2005). The residue Tyr33 in the protein MhuP also penta-coordinates the heme molecule exhibiting an axial binding with the iron atom in our docking analysis. Moreover, the protein-protein docking for heme-bound MhuP and DppA corroborate the already established hypothesis about their interaction (Mitra et al., 2019). Therefore, based on the available literature and the outcomes from this study, we propose that heme transport in *M. tuberculosis* occurs in a transient manner. Initially, the heme molecule is first introduced into the periplasmic space through the heme-binding membrane proteins PPE36/PE22 and PPE62 (Mitra et al., 2017). In the periplasmic space, the heme molecule is recognized and sequestered by the protein MhuP (Figure 6.12). The heme-bound MhuP protein then interacts with the protein DppA and delivers the ligand, followed by delivering the heme molecule to the cognate TMD subunits of the DppABCD transporter from where the heme substrate is ultimately influxes into the cytoplasm (Figure 6.12). In conclusion, the iron-acquisition mechanisms of *M. tuberculosis* are a complex system compared to other prokaryotes since *M. tuberculosis* is capable of capturing iron from different sources. The exclusive presence of SBPs in prokaryotes such as MhuP protein from *M. tuberculosis* and its importance in iron acquisition in the pathogen makes it a potential therapeutic target.

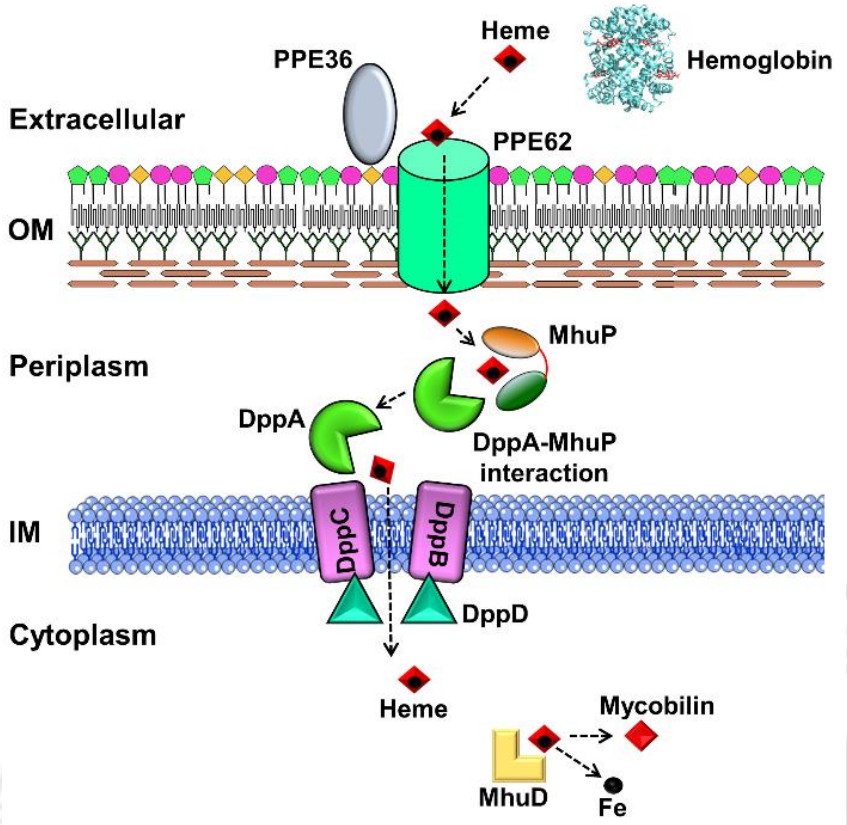


Figure 6.12. Schematic representation of heme acquisition by *M. tuberculosis*. Initially, the cell surface proteins PPE36 and PPE62 bind heme (black sphere inside a red diamond) and are anchored in the outer membrane (OM) (Mitra et al., 2017). After crossing the OM, heme is recognized and sequestered by the periplasmic protein MhuP. The heme-bound MhuP interacts with the dipeptide-binding DppA lipoprotein, which plausibly delivers heme to its cognate the DppBCD transport system for uptake across the inner membrane (IM). The influxed heme in the cytoplasm is further degraded by the oxygenase MhuD to mycobilin resulting in the release of the iron ion (black sphere). This study completes the story of heme transfer from the protein MhuP to DppA for successful translocation of heme inside the cytoplasm. The figure is redrawn from Mitra et al., 2017 and Mitra et al., 2019 using the results obtained from this study.

CONCLUSIONS AND SUMMARY

CONCLUSIONS AND SUMMARY

The work presented in this thesis can be broadly grouped into two parts. First is the computational characterization of the complete metal uptake ABC transporters in a thermophilic *Thermus thermophilus* HB8 and a pathogenic *Mycobacterium tuberculosis* H37Rv. The second part deals with the structural and functional characterization of one SBP of the metal-specific ABC transport system from each of the microorganisms using various experimental approaches. The bioinformatic analysis of the repertoire of metal uptake ABC transport systems using a data mining approach results in the identification of 22 and 19 ORFs which results in eight and seven metal uptake ABC transport systems in *T. thermophilus* HB8 and *M. tuberculosis* H37Rv, respectively. In *T. thermophilus*, these metal uptake ABC transporters are further characterized using various *in silico* approaches to further identify their probable cognate ligands. The results obtained from the analysis suggest that out of the eight metal-specific ABC import systems, three are involved in the import of iron ions either in their ionic or complex form. Among the remaining five, two of the ABC import systems are involved in divalent metal ions such as Mg^{2+} and Zn^{2+}/Mn^{2+} transport, whereas two are specific for MoO_4^{2-} and WO_4^{2-} import. The remaining ABC transporter is probably involved in the translocation of cyanocobalamin (vitamin B12) into the cytoplasm. Similarly, in *M. tuberculosis* H37Rv, the 19 ORFs identified resulted in seven probable metal uptake ABC transport systems and two P-type ATPases potentially involved in the uptake of metal ions. Further characterization of these transport systems in *M. tuberculosis* leads to the identification of a total of three ABC transporters responsible for iron import. Out of the remaining ABC transport systems identified, one is involved in the transport of d-block metal ion Zn^{2+} and one each for MoO_4^{2-} and cobalamin uptake. The remaining one system is identified as an energy-coupling factor (ECF)-type ABC transporter, probably importing Co^{2+} ion. In addition to the identification of the complete repertoire of probable metal uptake ABC transporters in both the organisms, their distribution on the genome and their interaction with the cognate TMD and NBD subunits suggests the presence of a subunit sharing mechanism where the TMD and NBD subunits can be shared among different metal uptake ABC transport systems. The existence of a subunit sharing mechanism indicates the plausible import of multiple substrates (in particular metal ions) via a single ABC transport system in both organisms. To conclude, this part of the thesis represents the

first-ever preliminary glimpse of the complete metal uptake ABC transport systems suggesting their possible cognate ligands in *T. thermophilus* and *M. tuberculosis*. This will further help in identifying potential therapeutic targets in tuberculosis-causing pathogen *M. tuberculosis*, as well as also onset an idea of different metal transporting proteins for other biotechnological applications in the future.

In the second part of the thesis, two of the metal uptake ABC transport protein, one each from *T. thermophilus* HB8 and *M. tuberculosis* H37Rv are characterized structurally and functionally using various experimental approaches. In *T. thermophilus* HB8, the protein MctA (ORF id: TTHB177) is characterized experimentally, which is suggested to be a divalent metal ion complexed citrate transporter in our *in silico* analysis. In this study, the crystal structure of the protein MctA from the Gram-negative *Thermus thermophilus* HB8 is solved at a resolution range of 1.63 to 2.50 Å in apo as well as holo forms. The study revealed that although the physiological metal ion for the protein MctA is Mg^{2+}, the protein could also bind several other divalent metal ions with high affinity. Moreover, the results also uncover a novel subclassification of cluster D SBPs that are capable of binding divalent metal ion complexed citrate molecules. However, the paucity of structures in the protein data bank impedes further characterization of these proteins in a more detailed manner. A comparison between the open (ligand unbound) and closed (ligand-bound) structure of the wild type and mutant proteins of MctA demonstrates a gating mechanism of ligand entry. The results also reveal that upon ligand entry, the protein MctA follows an asymmetric domain movement mechanism to bind the ligand, i.e. Mg^{2+}-citrate.

The second part of the thesis also reports the crystal structures of an SBP MhuP (ORF id: Rv0265c) from *M. tuberculosis* H37Rv solved in two different space groups $P2_1$ and $C2$, at a resolution of 1.8 and 2.01 Å, respectively. The spectrometric studies conducted on the protein MhuP indicate a selective and specific heme-binding property of the protein in solution. The results also suggest that the protein MhuP is incapable of binding other Fe^{3+}-siderophores, in particular carboxylate-type siderophore. Furthermore, the molecular docking results indicate the potential binding of the protein MhuP with heme, as well as the possible mechanism of heme transfer from the protein MhuP to DppA. To conclude, this study confirms the binding of heme with the protein

MhuP as well as also confirms the earlier hypothesis of transient transport of heme via the DppABCD transporter in *M. tuberculosis*.

Altogether, the outcomes of this study reflect an overview of the repertoire of metal uptake ABC transport systems in *T. thermophilus* HB8 and *M. tuberculosis* H37Rv, suggesting their probable ligands. The structural and functional characterization of the protein MctA from *T. thermophilus* HB8 reveals the presence of a novel Mg^{2+}-citrate-binding protein in Gram-negative bacteria and its mechanistic details of substrate entry and binding. Moreover, the structural and functional characterization of the protein MhuP from *M. tuberculosis* H37Rv exhibited its specificity towards heme molecules and also suggests a possible mechanism of heme transfer between the protein MhuP and DppA via DppABCD transport system.

APPENDIX A

Supplementary data to chapter 3

APPENDIX-A

Table A1. Details of the open reading frames (ORFs) identified as metal uptake ABC transporter proteins from the transporter class 3.A.1 extracted from the TransportDB 2.0 and Structural-Biological Whole-Cell Project (http://www.thermus.org) databases.

S. No.	ORF number	UniProt ID	+Protein name	*(SP) Protein length	Cellular location	**ABC subunit
1.	TTHA0596	Q5SKP7	ABC transporter solute-binding protein	(1-21) 281	Periplasm	SBP
2.	TTHA0597	Q5SKP6	Uncharacterized protein	(-) 553	-	-
3.	TTHA0598	Q5SKP5	Putative transport integral membrane protein (chloride channel)	(-) 481	Membrane	TMD
4.	TTHA1628	Q5SHV2	Iron ABC transporter, periplasmic iron-binding protein	(-) 330	Periplasm	SBP
5.	TTHA1629	Q5SHV1	Iron ABC transporter, permease protein	(-) 516	Membrane	TMD
6.	TTHA1630	Q5SHV0	Iron ABC transporter, ATP-binding protein	(-) 350	Cytosol	NBD
7.	TTHB218	Q53VW1	ABC transporter, ATP-binding protein	(-) 258	Cytosol	NBD

8.	TTHB219	Q53VW0	Hemin ABC transporter, permease protein	(-) 341	Membrane	TMD
9.	TTHB220	Q53VV9	Hemin ABC transporter, periplasmic hemin-binding protein	(1-36) 301	Periplasm	SBP
10.	TTHA0723	Q5SKC0	Iron(III) ABC transporter, solute-binding protein	(1-18) 265	Periplasm	SBP
11.	TTHA0746	Q5SKA2	Iron ABC transporter, substrate-binding protein	(1-22) 284	Periplasm	SBP
12.	TTHA0747	Q5SKA1	Iron ABC transporter, permease protein	(-) 342	Membrane	TMD
13.	TTHA0748	Q5SKA0	Iron ABC transporter, ATP-binding protein	(-) 250	Cytosol	NBD
14.	TTHB175	Q53VZ4	ABC transporter, ATP-binding protein	(-) 368	Cytosol	NBD
15.	TTHB176	Q53VZ3	Putative iron ABC transporter, permease protein	(-) 565	Membrane	TMD
16.	TTHB177	Q53VZ2	Iron ABC transporter, periplasmic iron-binding protein	(1-18) 358	Periplasm	SBP

17.	TTHA0069	Q5SM74	ABC transporter, periplasmic binding protein	(1-46) 293	Periplasm	SBP
18.	TTHA0070	Q5SM73	ABC transporter permease protein	(-) 211	Membrane	TMD
19.	TTHA0071	Q5SM72	ABC-transporter, ATP-binding subunit	(-) 224	Cytosol	NBD
20.	TTHA0715	Q5SKC8	ABC transporter ATP-binding protein	(-) 339	Cytosol	NBD
21.	TTHA0716	Q5SKC7	Molybdenum ABC transporter, permease protein	(-) 224	Membrane	TMD
22.	TTHA0717	Q5SKC6	Molybdenum ABC transporter molybdate-binding protein	(1-19) 289	Periplasm	SBP

[+] Protein names are according to the UniProtKB database.

* SP: Signal peptide.

** SBP: Solute (substrate)-binding protein; TMD: Transmembrane domain; NBD: Nucleotide-binding domain.

Table A2. Protein annotation and UniProt ids of the ORFs interacting with the respective metal binding ABC transporter protein obtained in the interactome map analysis.

ORF number	***ORF number of the interacting protein**	**UniProt ID of the interacting protein**	**[+]Protein name**
TTHA0723	TTHA0747	Q5SKA1	Iron ABC transporter permease

	TTHA0748	Q5SKA0	Iron ABC transporter ATP-binding protein
	TTHA0722	Q5SKC1	Histidinol dehydrogenase
	TTHA1534	Q5SI46	Ribonuclease II family protein
	TTHA0910	Q5SJU8	Ribonuclease R
	TTHA0724	Q5SKB9	Serine protease
	TTHA0721	Q5SKC2	Hypothetical protein
	TTHA0720	Q5SKC3	Hypothetical protein
	TTHA1628	Q5SHV2	Iron ABC transporter substrate-binding protein
	TTHA0683	Q5SKG0	TolR-type transport protein
TTHA0746	TTHA0748	Q5SKA0	Iron ABC transporter ATP-binding protein
	TTHA0747	Q5SKA1	Iron ABC transporter permease
	TTHA0749	Q5SK99	Hypothetical protein
	TTHA1534	Q5SI46	Ribonuclease II family protein
	TTHA0910	Q5SJU8	Ribonuclease R
	TTHA0750	Q5SK98	3-oxoacyl-ACP reductase
	TTHA1628	Q5SHV2	Iron ABC transporter substrate-binding protein
	TTHA0683	Q5SKG0	TolR-type transport protein
	TTHA0684	Q5SKF9	TolQ-type transport protein
TTHA0069 and TTHA0717	TTHA0716	Q5SKC7	Molybdenum ABC transporter, permease protein
	TTHA0070	Q5SM73	ABC transporter permease protein
	TTHA0071	Q5SM72	ABC-transporter, ATP-binding subunit
	TTHA0715	Q5SKC8	ABC transporter ATP-binding protein
	TTHA1159	Q5SJ55	Amino acid ABC transporter, ATP-binding protein

	TTHA1630	Q5SHV0	Iron ABC transporter, ATP-binding protein
	TTHA1424	Q5SIF0	ABC transporter ATP-binding protein
	TTHA0975	Q5SJN3	Sugar ABC transporter, ATP-binding protein
	TTHA0579	Q5SKR4	Sugar ABC transporter, ATP-binding protein
	TTHA0380	Q5SLB3	Aldehyde: ferredoxin oxidoreductase

* Only the top 10 interacting proteins sorted by the best confidence score in the decreasing order are provided.

+ Protein names are according to the UniProtKB database.

Table A3. Details of the template protein used for the model building of each SBP.

ORF number	Template	PDB ID of template	Microorganism	Protein name	Bound Ligand
TTHB220	RhuT	5GJ3	*Roseiflexus sp.*	*Rs*RhuT	Heme
TTHA0723	BhuT	5Y8A	*Burkholderia cenocepacia*	*Bc*BhuT	Heme
TTHA0746	ShuT	2R7A	*Shigella dysentriae*	*Sd*ShuT	Heme
TTHB177	Atu	3C9H	*Agrobacterium fabrum*	*Af*Atu	Mg^{2+}-citrate
TTHA0069	TupA	3LR1	*Geobacter sulfurreducens*	*Gs*TupA	Tungsten (W^{6+}) ion
TTHA0717	ModA	1ATG	*Azotobacter vinelandii*	*Av*MoDA	Tungstate (WoO_4^{2-})

Table A4. Top 20 structural homologs (ranked by the Z-score) of the selected proteins obtained using the web-server Dali (Holm and Rosenstrom, 2010).

ORF no.	PDB ID	UniProt ID	+Protein name	Bound ligand	**Z-score	RMSD (Å)	LALI	NRES	SI (%)
TTHB220	5GJ3	A5UZ69	Periplasmic binding protein	Heme	42.0	0.5	260	270	49
	5Y8B			-	38.9	1.1	258	263	49
	5GIZ	B4EKB3	Putative hemin transport system, substrate-binding protein	-	33.0	1.8	260	271	37
	3MD9	Q56991	Hemin-binding periplasmic protein HmuT	-	32.9	1.5	252	253	34
	5Y8A	B4EKB3	Putative hemin transport system, substrate-binding protein	Heme	32.9	1.8	260	268	37
	3NU1	A0A0E1NWS3	Hemin-binding	Heme	32.8	1.7	253	254	34

			periplasmic protein						
	5Y89	B4EKB3	Putative hemin transport system, substrate-binding protein	Heme	32.3	2.3	260	267	37
	2R7A	O70018	Uncharacterized protein	Heme	30.1	2.0	252	255	36
	2RG7			-	29.6	1.8	250	255	36
	5AZ3	Q8NTB8	ABC-type transporter, periplasmic component	Heme	28.5	2.3	254	309	35
	2R79	O68879	Periplasmic binding protein	Heme	28.2	2.0	253	279	30
	5KHL	A0A0H3AE70	Hemin ABC transporter, periplasmic hemin-binding protein HutB	-	28.0	2.1	251	257	30

	4M7O	A0A0H2VHB1	Iron binding protein	-	26.0	3.0	251	258	21
	1N2Z	P37028	Vitamin B12-binding protein	Co-cyanocobalamin	23.2	3.1	242	245	26
	1N4D			-	23.0	2.9	241	244	27
	2Q8Q	Q7A652	High-affinity heme uptake system protein IsdE	Heme	22.8	3.1	241	258	21
	5JOQ	Q7AP55	Lmo2184 protein	-	22.7	3.5	246	258	22
	4HN9	C4Z2B5	Iron complex transport system substrate-binding protein	-	22.2	2.5	245	323	20
	4N01	D1BNS2	Periplasmic binding protein	-	21.7	2.9	247	296	17
	3PSH	P44206	Uncharacterized protein HI_1472	Molybdate	22.7	3.3	261	326	24

TTHA0723	5Y8A	B4EKB3	Putative hemin transport system, substrate-binding protein	Heme	34.2	1.1	240	271	24
	5GIZ			-	33.7	1.2	240	271	23
	5Y89			Heme	32.3	1.8	241	267	23
	3MD9	Q56991	Hemin-binding periplasmic protein HmuT	-	30.5	1.7	239	253	22
	3NU1	A0A0E1NWS3	Hemin-binding periplasmic protein	Heme	30.2	2.1	240	254	22
	5Y8B	A5UZ69	Periplasmic binding protein	-	29.2	1.8	241	263	22
	5GJ3			Heme	29.1	1.8	241	270	23
	2R7A	O70018	Uncharacterized protein	Heme	26.9	2.0	233	255	21
	2RG7			-	26.4	2.1	233	255	21
	2R79	O68879	Periplasmic binding protein	Heme	24.3	2.3	233	279	21
	5KHL	A0A0H3AE70	Hemin ABC transporter,	-	24.2	2.4	236	257	20

			periplasmic hemin-binding protein HutB						
	1N4A	P37028	Vitamin B12-binding protein	Co-cyanocobalamin	23.9	2.6	234	244	25
	5AZ3	Q8NTB8	ABC-type transporter, periplasmic component	Heme	23.2	2.4	236	309	17
	4FI3	P06609	Vitamin B12 import system permease protein BtuC	-	23.1	2.7	233	245	24
	4M7O	A0A0H2VHB1	Iron binding protein	-	23.0	2.9	235	258	25
	5M29	P37028	Vitamin B12-binding protein	Cobinamide	22.7	2.5	220	229	26
	2Q8Q	Q7A652	High-affinity heme	Heme	20.1	3.0	226	258	19

			uptake system protein IsdE						
	5JOQ	Q7AP55	Lmo2184 protein	-	19.9	3.2	231	258	19
	4MLZ	C7R301	Periplasmic binding protein	-	19.7	2.7	236	299	19
	5AD1	Q0P8Q4	Enterochelin uptake periplasmic binding protein	Fe^{3+}	19.5	2.6	231	290	16
TTHA0746	2R7A	O70018	Uncharacterized protein	Heme	29.7	1.8	244	255	25
	2RG7			-	29.6	1.8	244	255	25
	4HN9	C4Z2B5	Iron complex transport system substrate-binding protein	-	29.1	2.1	258	323	21
	5Y8B	A5UZ69	Periplasmic binding protein	-	27.6	1.9	244	263	27
	4M7O	A0A0H2VHB1	Iron-binding protein	-	27.6	2.1	245	258	27

	5GJ3	A5UZ69	Periplasmic binding protein	Heme	27.2	1.9	244	270	27
	5Y8A	B4EKB3	Putative hemin transport system, substrate-binding protein	Heme	27.1	1.9	243	268	23
	2R79	O68879	Periplasmic binding protein	Heme	26.8	2.1	244	279	25
	5GIZ	B4EKB3	Putative hemin transport system, substrate-binding protein	-	26.7	2.1	246	271	22
	3NU1	A0A0E1NWS3	Hemin-binding periplasmic protein	Heme	26.3	2.5	240	254	26
	1N2Z	P37028	Vitamin B12-binding protein	Co-cyanocobalamin	26.2	2.4	238	245	26
	5Y89	B4EKB3	Putative hemin transport system, substrate-	Heme	26.1	2.7	243	267	23

			binding protein						
	3MD9	Q56991	Hemin-binding periplasmic protein HmuT	-	26.1	2.2	240	253	26
	5KHL	A0A0H3AE70	Hemin ABC transporter, periplasmic hemin-binding protein HutB	-	25.9	1.9	243	257	19
	3PSH	P44206	Uncharacterized protein HI_1472	Molybdate	25.3	2.8	258	326	20
	3PSA			Tungstate	25.2	2.8	258	326	20
	5AZ3	Q8NTB8	ABC-type transporter, periplasmic component	Heme	25.0	2.7	256	309	19
	4OVK	D1BR49	Periplasmic binding protein	-	25.0	3.3	259	329	22

	4N01	D1BNS2	Periplasmic binding protein	-	25.0	2.3	253	296	21
	5M29	P37028	Vitamin B12-binding protein	Cobinamide	24.9	2.3	223	229	27
TTHB177	3C9H	Q7CWZ6	ABC transporter, substrate binding protein (Iron)	Mg^{2+} citrate	53.1	0.7	336	339	35
	4R74	A3N294	ABC-type Fe3+ transport system, periplasmic component	Fructose-6-phosphate	31.0	2.6	303	319	20
	4R73			Glucose-6-phosphate, Mannose-6-phosphate	30.8	2.6	303	321	18
	4R75			Sedoheptulose-7-	30.7	2.4	300	318	18

				phosphate					
	4ELR	Q5SHV2	Iron ABC transporter, periplasmic iron-binding protein	Fe^{3+} carbonate	29.5	3.3	298	310	21
	1D9Y	Q50964	FbpA protein	Fe^{3+} phosphate	28.2	2.8	293	309	16
	1SI0	Q9Z4N6	Iron ABC transporter substrate-binding protein	Fe^{3+} carbonate	28.2	3.2	294	316	16
	2PT2	P72827	Iron uptake protein A1	Fe^{2+} sulfate	27.9	3.1	288	316	19
	3E13	Q0PBW4	Putative iron-uptake ABC transport system, periplasmic iron-binding protein	Fe^{3+}	27.8	3.1	290	317	18
	1XVX	A1JLH5	Iron(III)-binding	Fe^{3+} carbonate	27.7	3.3	297	311	17

			periplasmic protein						
	1MRP	P35755	Iron-utilization periplasmic protein	Fe^{3+} phosphate	27.5	3.0	292	309	14
	2OWT	Q7VXW9	Putative iron binding protein	Fe^{3+} carbonate	27.5	3.1	290	315	17
	4R6Y	P96062	Putative 2-aminoethyl phosphonate-binding periplasmic protein	*Acetate	27.4	3.1	298	317	18
	1Q35	Q9Z4N6	Iron ABC transporter substrate-binding protein	*Formic acid	26.9	3.3	292	317	16
	2VP1	Q55835	Iron uptake protein A2	Fe^{3+}	26.9	3.2	286	313	19
	1XVY	P21408	Fe(3+)-binding periplasmic protein	*Citric acid	26.6	3.6	291	307	16
	2QRY	P31550	Thiamine-binding periplasmic protein	Thiamin phosphate	26.3	2.8	291	318	16

	3TTM	Q9I6J1	Putrescine-binding periplasmic protein SpuD	Putrescine	25.5	2.9	292	341	15
	1XC1	P17259	Major ferric iron-binding protein	Oxo-zirconium(IV) cluster	25.4	3.6	293	309	16
	1O7T	Q50964	FbpA protein	Metal nanoclusters	25.3	3.7	292	309	16
TTHA0069	3LR1	Q749P2	Tungstate ABC transporter, periplasmic tungstate-binding protein	Tungsten ion	38.1	0.2	224	232	39
	5MY5	Q316W1	ABC transporter periplasmic substrate-binding protein	-	25.5	3.0	229	250	38
	3KN3	Q7M8V9	Putative periplasmic protein	*Citrate	24.9	3.3	225	238	35

	3CVG	J3KFC8	Uncharacterized protein	-	24.8	2.3	225	270	25
	3MUQ	Q87PK2	Uncharacterized protein	-	21.9	5.3	151	233	32
	1AMF	P37329	Molybdate-binding protein ModA	Molybdate	18.7	2.9	207	231	12
	1ATG	Q7SIH2	Binding protein	Tungstate	18.5	3.0	205	231	14
	1SBP	P02906	Sulfate-binding protein	Sulfate ion	17.4	2.9	216	309	11
	4Q8R	A0A0H2YSI2	Phosphate ABC transporter, phosphate-binding protein	Phosphate ion	17.4	2.8	205	241	18
	5Y2W	P73862	Rubisco operon transcriptional regulator	*2-phosphoglycolic acid	17.3	3.0	198	208	15
	3CFZ	Q58586	Molybdate/tungstate-binding protein WtpA	Tungstate	17.2	3.2	210	289	11

	5UM2	Q8PNN7	ABC transporter sulfate binding protein	Sulfate ion	17.2	2.9	216	316	11
	2ONR	O30142	Molybdate /tungstate-binding protein WtpA	Molybdate	17.1	3.3	213	314	12
	5TED	Q8Y9N7	Lmo0488 protein	Shikimate	17.1	2.7	200	210	11
	4GD5	A0A0H2YSI2	Phosphate ABC transporter, phosphate-binding protein	Phosphate ion	17.1	2.8	204	243	18
	3FJ7	Q0PBL7	Major antigenic peptide PEB3	L-phospholactate	17.1	3.3	207	231	14
	4JB7	Q9KTQ6	Accessory colonization factor AcfC	D-malate	16.7	3.3	206	233	9
	2FYI	Q47083	HTH-type transcriptional regulator cbl	-	16.7	2.6	195	223	13

	3GLB	P07774	HTH-type transcriptional regulator CatM	*Hexa-2,4-diene dioic acid	16.3	2.8	198	215	16
	2QL3	Q0SFM8	Probable transcriptional regulator, LysR family protein	-	16.3	3.1	194	205	14
TTHA0717	1ATG	Q7SIH2	Binding protein	Tungstate	33.2	1.2	228	231	29
	3GZG	Q8PHA1	Molybdate-binding protein ModA	Molybdate	29.1	1.9	221	232	29
	1WOD	P37329	Molybdate-binding protein ModA	Tungstate	28.8	1.8	224	231	26
	4RXL	Q9KLL7	Molybdenum ABC transporter, periplasmic molybdenum-binding protein	Tungstate	28.7	1.9	223	229	23

	3CFX	Q8TTZ5	Uncharacterized solute-binding protein MA_0280	Tungstate	26.3	2.1	238	293	18
	3CIJ	O30142	Molybdate/tungstate-binding protein WtpA	Tungstate	26.0	2.1	239	291	18
	3CFZ	Q58586	Molybdate/tungstate-binding protein WtpA	Tungstate	25.4	2.2	235	289	20
	3CG1	Q8U4K5	Molybdate/tungstate-binding protein WtpA	Tungstate	25.3	2.1	238	293	16
	3CG3	O57890	Molybdate/tungstate-binding protein WtpA	Tungstate	24.6	2.3	241	315	17
	4KD5	Q18A64	ABC-type transport system, molybdenum-specific extracellul	-	24.0	3.1	224	229	21

			ar solute-binding protein						
	1SBP	P02906	Sulfate-binding protein	Sulfate ion	22.5	2.7	228	309	14
	5UM2	Q8PNN7	ABC transporter sulfate binding protein	Sulfate ion	22.3	2.6	228	316	15
	3FJM	Q0PBL7	Major antigenic peptide PEB3	Phosphate ion	19.5	3.1	218	231	14
	4R73	A3N294	ABC-type Fe^{3+} transport system, periplasmic component	Glucose-6-phosphate, Mannose-6-phosphate	19.4	3.3	225	319	14
	3CVG	J3KFC8	Uncharacterized protein	-	19.2	3.6	228	270	11
	3LR1	Q749P2	Tungstate ABC transporter, periplasmic tungstate-	Tungsten ion	18.8	2.8	210	232	14

			binding protein						
	3C9H	Q7CWZ6	ABC transporter, substrate binding protein (Iron)	Mg^{2+} citrate	18.7	2.9	226	339	12
	2QRY	P31550	Thiamine-binding periplasmic protein	Thiamin phosphate	18.5	3.4	224	309	13
	4JB7	Q9KTQ6	Accessory colonization factor AcfC	D-malate	18.4	3.2	218	233	14
	4Q8R	A0A0H2YSI2	Phosphate ABC transporter, phosphate-binding protein	Phosphate ion	18.0	3.0	209	241	8

[+] Protein names are according to the UniProtKB database.

* These molecules bound at the active site of the protein were present in the crystallization buffer.

** Z-score = [(S(A, B) – m(N))/ (σ(N)], where S(A, B) = similarity score between protein A and B; m(N) = mean score of average length N; σ = average standard deviation.

Abbreviations: RMSD, root mean square deviation; LALI, length of the alignment; NRES, number of residues; SI: sequence identity.

Table A5. Details of the proteins used for generating the structure-based distance tree for SBP classification.

Cluster	PDB ID	UniProt ID	Protein name	Organism	Proposed ligand(s)
A-I	1PQ4	P73085	ZntC	*Synechocystis sp.*	Zinc
A-I	1TOA	P96116	TroA	*Treponema pallidum*	Zinc
A-I	1XVL	Q79EF9	MntC	*Synechocystis sp.*	Manganese
A-I	3MFQ	A4VY63	TroA	*Streptococcus suis*	Zinc
A-I	4K3V	D9RF12	MntC	*Staphylococcus aureus*	Manganese
A-II	2R7A	O70018	ShuT	*Shigella dysenteriae*	Heme
A-II	2R79	O68879	PhuT	*Pseudomonas aeruginosa*	Heme
A-II	3EIW	A0A0H3JM89	HtsA	*Staphylococcus aureus*	Staphyloferrin A
A-II	3NU1	A0A0E1NWS3	HmuT	*Yersinia pestis*	Heme
A-II	4HN9	C4Z2B5	N.D.	*Eubacterium eligens*	Iron complex
A-II	5AZ3	Q8NTB8	HmuT	*Corynebacterium glutamicum*	Heme
A-II	5Y8A	B4EKB3	BhuT	*Burkholderia cenocepacia*	Heme
B	4F11	O75899	GABBR2	*Homo sapiens*	GABA
B	4MQE	Q9UBS5	GABBR1	*Homo sapiens*	GABA
B-I	1TJY	Q8ZKQ1	IsrB	*Salmonella typhimurium*	Autoinducer2

B-I	3D02	A6THR5	N.D.	*Klebsiella pneumoniae*	Sugar
B-I	3T95	Q74PW2	IsrB	*Yersinia pestis*	Autoinducer 2
B-I	4WZZ	A9KIX1	N.D.	*Clostridium phytofermentans*	Sugar
B-II	3IP5	Q7CX36	Atu2422	*Agrobacterium fabrum*	Amino acids (GABA)
B-II	3TD9	Q9X0L9	LivK	*Thermotoga maritima*	Branched-chain amino acids
B-II	4GNR	J9PBT7	N.D.	*Streptococcus pneumoniae*	Branched-chain amino acids
B-II	4N0Q	Q8YEE8	N.D.	*Brucella melitensis*	Leucine/ Isoleucine/ Valine
B-II	4NQR	Q3MFZ5	N.D.	*Anabaena variabilis*	Amino acid, amide
B-III	3SG0	Q2IR47	N.D.	*Rhodopseudomonas palustris*	Aromatic compounds
B-III	3UK0	Q139W5	N.D.	*Rhodopseudomonas palustris*	Aromatic compounds
B-III	4EVS	Q6NB45	N.D.	*Rhodopseudomonas palustris*	Aromatic compounds
B-III	4EY3	Q13AR6	N.D.	*Rhodopseudomonas palustris*	N.D.
B-III	4F06	Q2IXT5	N.D.	*Rhodopseudomonas palustris*	N.D.
B-IV	1DP4	P18910	Npr1	*Rattus norvegicus*	Natiuretic peptides
B-IV	1JDN	P17342	NPR3	*Homo sapiens*	Natiuretic peptides

B-IV	1YK0	P01160	NPPA	*Homo sapiens*	Natiuretic peptides
B-V	1EWK	P23385	Grm1	*Rattus norvegicus*	Glutamate
B-V	2E4U	P31422	Grm3	*Rattus norvegicus*	Glutamate
B-V	2E4Z	P35400	Grm7	*Rattus norvegicus*	Glutamate
B-V	3LMK	P41594	GRM5	*Homo sapiens*	Glutamate
B-V	3MQ4	Q14831	GRM7	*Homo sapiens*	Glutamate
B-V	4XAQ	Q14416	GRM2	*Homo sapiens*	Glutamate
C	1DPP	P23847	DppA	*Escherichia coli*	Dipeptide
C	1XOC	P42061	AppA	*Bacillus subtilis*	Oligopeptide
C	2D5W	Q5SHU6	N.D.	*Thermus thermophilus*	Peptide
C	3M8U	B8F653	HbpA	*Haemophilus parasuis*	Heme
C	5F1Q	A0A0H2YM47	N.D.	*Yersinia pestis*	Dipeptide
D-I	2B3B	Q72KX2	GGBP	*Thermus thermophilus*	Glucose
D-I	3OO6	Q27GR2	AcbH	*Actinoplanes sp.*	Galactose
D-I	3ZKK	A0A0J9X1B4	N.D.	*Bifidobacterium animalis*	Sugar
D-I	4G68	J9PBT4	N.D.	*Caldanaerobius*	N.D.
D-I	4R2B	A6X5L9	N.D.	*Ochrobactrum anthropi*	Sugar
D-II	1A99	P31133	PotF	*Escherichia coli*	Putrescine
D-II	1POT	P0AFK9	PotD	*Escherichia coli*	Spermidine, Putrescine
D-II	3RPW	Q6N0W2	N.D.	*Rhodopseudomonas palustris*	N.D.
D-II	3TTM	Q9I6J1	SpuD	*Pseudomonas aeruginosa*	Putrescine

D-II	4GL0	Q8Y8T4	N.D.	*Listeria monocytogenes*	Spermidine, Putrescine
D-III(A)	1ATG	Q7SIH2	ModA	*Azotobacter vinelandii*	Tungstate (VI)
D-III(A)	1WOD	P37329	ModA	*Escherichia coli*	Molybdate
D-III(A)	2H5Y	Q8PHA1	ModA	*Xanthomonas axonopodis*	Molybdate
D-III(A)	3C9H	Q7CWZ6	Atu	*Agrobacterium fabrum*	Iron
D-III(A)	3CVG	J3KFC8	N.D.	*Coccidioides immitis*	Metal
D-III(A)	3LR1	Q749P2	TupA	*Geobacter sulfurreducens*	Tungstate
D-III(A)	4RXL	Q9KLL7	N.D.	*Vibrio cholerae*	Molybdate
D-III(B)	2ONR	O30142	WtpA/ ModA	*Archaeoglobus fulgidus*	Molybdate/ Tungstate
D-III(B)	3CFX	Q8TTZ5	ModA	*Methanosarcina acetivorans*	Tungstate
D-III(B)	3CFZ	Q58586	WtpA	*Methanocaldococcus jannaschii*	Molybdate/ Tungstate
D-III(B)	3CG1	Q8U4K5	WtpA	*Pyrococcus furiosus*	Molybdate/ Tungstate
D-III(B)	3CG3	O57890	WtpA	*Pyrococcus horikoshii*	Molybdate/ Tungstate
D-IV	1MRP	P35755	FbpA	*Haemophilus influenzae*	Iron
D-IV	1SI0	Q9Z4N6	FbpA	*Mannheimia haemolytica*	Iron
D-IV	1XVX	A1JLH5	YfuA	*Yersinia enterocolitica*	Iron

D-IV	1Y4T	Q0PBW4	CfbpA	*Campylobacter jejuni*	Iron
D-IV	2PT2	P72827	FutA1	*Synechocystis sp.*	Iron
E-I	4MX6	Q8ECK4	DctP	*Shewanella oneidensis*	Succinate
E-I	4NF0	Q9HVH5	N.D.	*Pseudomonas aeruginosa*	L-maltate
E-I	4O7M	A3QCW5	N.D.	*Shewanella loihica*	L-malate
E-I	4O94	Q2IUT5	N.D.	*Rhodopseudomonas palustris*	Succinate
E-I	4OVS	D1AZL7	N.D.	*Sulfurospirillum deleyianum*	Succinate
E-II	2DVZ	Q7W0A0	BugE	*Bordetella pertussis*	Glutamate
E-II	2F5X	Q7VTS1	BugD	*Bordetella pertussis*	Aspartate
E-II	2QPQ	Q7W019	Bug27	*Bordetella pertussis*	N.D.
F-I	2G29	P73452	NrtA	*Synechocystis sp.*	Nitrate
F-I	2I48	Q55460	CmpA	*Synechocystis sp.*	Bicarbonate
F-I	2X26	P75853	SsuA	*Escherichia coli*	Aliphatic sulfonate
F-I	3E4R	Q8PHQ1	SsuA	*Xanthomonas axonopodis*	Alkanesulfonate
F-I	3UN6	Q2G1I5	N.D.	*Staphylococcus aureus*	N.D.
F-II	3K2D	Q8DFC1	LlpA	*Vibrio vulnificus*	Methionine
F-II	3TQW	Q83F42	N.D.	*Coxiella burnetii*	Methionine
F-II	4K3F	Q9HT68	N.D.	*Pseudomonas aeruginosa*	N.D.
F-II	4QHQ	B4EA88	N.D.	*Burkholderia cenocepacia*	Methionine

F-II	4YAH	P28635	MetQ	*Escherichia coli*	Methionine
F-III	3O66	A0A0H3K0Z1	OpuCC	*Staphylococcus aureus*	Glycine Betaine/ Carnitine/ Choline
F-III	3PPN	O32243	OpuCC	*Bacillus subtilis*	Glycine Betaine/ Carnitine/ Choline
F-III	4ND9	A9CKK7	N.D.	*Agrobacterium fabrum*	Proline/ Glycine/ Betaine
F-III	4WEP	P33362	YehZ	*Escherichia coli*	Glycine/ Betaine
F-III	4Z7E	Q8Y775	N.D.	*Listeria monocytogenes*	N.D.
F-IV	2XHD	P42262	GRIA2	*Homo sapiens*	Glutamate
F-IV	3DLN	P19492	Gria3	*Rattus norvegicus*	Glutamate
F-IV	3FAT	P19493	Gria4	*Rattus norvegicus*	Glutamate
F-IV	3S9E	P42264	Grik3	*Rattus norvegicus*	Glutamate
F-IV	5FTH	P19491	Gria2	*Rattus norvegicus*	Glutamate

N.D.: Not defined.

Table A6. Details of the molecular docking results of various metal ions and/or their complexes with SBPs.

ORF number	**Ligand**	**EFBE (kcal mol^{-1})**	**NHBOND**	**Interactions**	
				Ligand atom	**Protein residue atom**
TTHB220	Heme	-9.83	4	O1D	Tyr161 O, Arg163 N$^{\eta 2}$

				O1A	Arg163 $N^{\eta 2}$
				O2A	Arg163 $N^{\eta 2}$
TTHA0723	Heme (conformation-A)	-8.19	4	O1A	Lys191 N^{ζ}
				O2A	Lys191 N^{ζ}
				O1D	Lys195 N^{ζ}
				O2D	Lys195 N^{ζ}
	Heme (conformation-B)	-8.02	8	Fe	Tyr44 O
				NA	Tyr44 O
				NB	Tyr44 O
				ND	Tyr44 O
				O1A	Thr45 $O^{\gamma 1}$
				O2A	Thr45 $O^{\gamma 1}$
				O1D	Lys191 N^{ζ}
				O2D	Lys191 N^{ζ}
TTHA0746	Cyanocobalamin	-4.16	5	O28	Asp238 $O^{\delta 2}$
				O34	Arg243 N^{ε}
				O39	Thr99 O
				O44	Ser78 O^{γ}
				N40	Glu100 O
	Heme	-8.39	6	O2A	Gly184 O, Phe186 N
				NB	Tyr158 O
				NC	Tyr158 O
				ND	Tyr158 O
				Fe	Tyr158 O
TTHB177	Mg^{2+}-Citrate	-8.94	21	O1	Mg^{2+}
				O2	Mg^{2+}, Ser198 O^{γ}, Thr199 N
				O3	Ser26 O, Arg246 $N^{\eta 1}$
				O4	Mg^{2+}, Ser26 O

				O5	Thr55 $O^{\gamma 1}$, Ser164 O^{γ}
				O6	Thr55 N
				O7	Mg^{2+}
	Fe^{2+}-Citrate	-10.02	20	O1	Fe^{2+}
				O2	Fe^{2+}, Ser198 O^{γ}, Thr199 N
				O3	Fe^{2+}, Ser26 O
				O4	Ser26 O, Arg246 $N^{\eta 1}$
				O5	Thr55 O^{γ}, Ser164 O^{γ}
				O6	Thr55 N
				O7	Fe^{2+}
TTHA0069	WO_4^{2-}	-2.92	7	O1	His55 N
				O2	Thr37 $O^{\gamma 1}$, His55 O, Gly120 N
				O3	Thr121 N & $O^{\gamma 1}$
				O4	Arg115 $N^{\eta 2}$
TTHA0717	WO_4^{2-}	-2.73	13	O1	Ser7 N & O^{γ}, Asp8 N
				O2	Ala164 N
				O3	Ser7 O^{γ}
				O4	Ser36 N
				W	Ser7 O^{γ}
	MoO_4^{2-}	-2.89	13	O1	Ser7 N & O^{γ}, Asp8 N
				O2	Ser7 O^{γ}
				O3	Ala164 N

				O4	Ser36 N
				Mo	Ser7 O^{γ},

Abbreviations: EFBE, estimated free energy of binding; NHBOND, number of hydrogen bonds.

APPENDIX B

Supplementary data to chapter 4

APPENDIX B

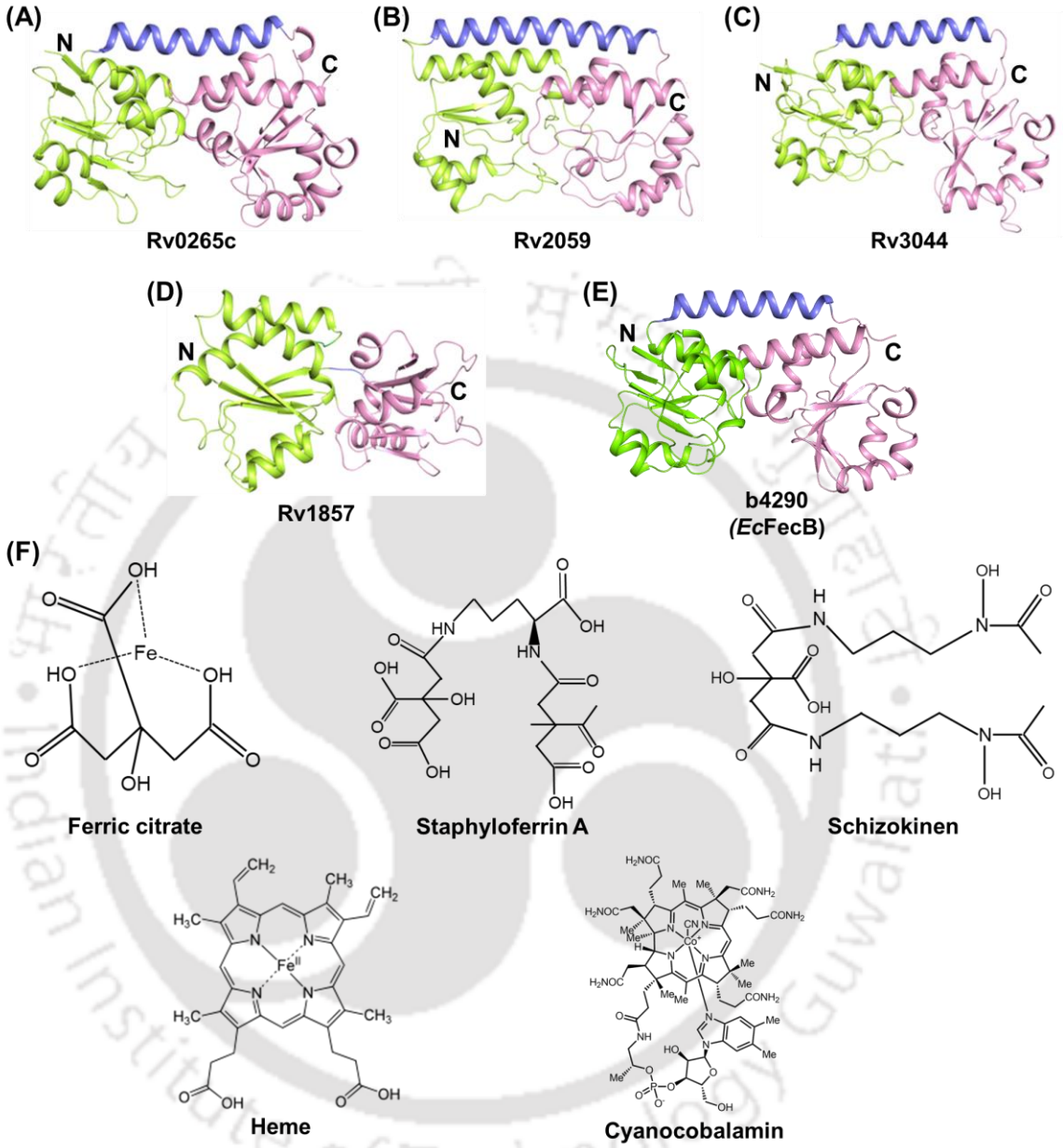


Figure B1. The three-dimensional crystal structure of (A) Rv0265c (PDB id: 4PM4) and theoretical models of (B) Rv2059, (C) Rv3044, (D) Rv1857 and (E) b4290 (*Ec*FecB) proteins. The N- and C-terminal domains are displayed in green and pink, respectively. The hinge region connecting these two domains are highlighted in blue. (F) The schematic two-dimensional chemical structures of iron complexes such as ferric citrate, staphyloferrin A and schizokinen and heme & cyanocobalamin (porphyrin or corrin-ring containing) molecules.

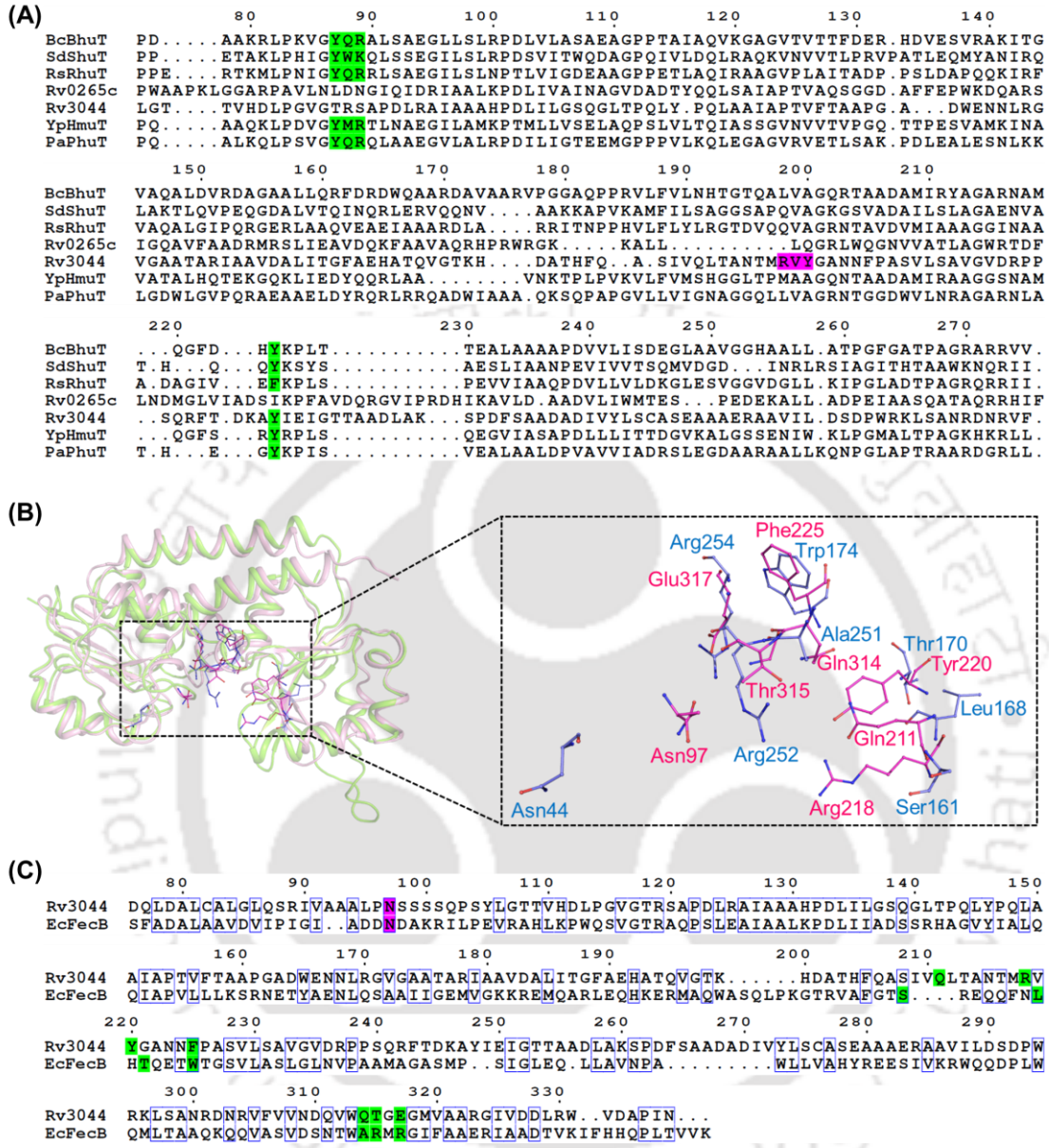


Figure B2. (A) Multiple sequence alignment of the proteins Rv0265c and Rv3044 with heme-binding proteins. The proteins used for the MSA are heme-binding proteins from *Burkholderia cenocepacia* (*Bc*BhuT, B4EKB3), *Shigella dysenteriae* (*Sd*ShuT, O70018), *Roseiflexus spp*. (*Rs*RhuT, A5UZ69), *Yersinia pestis* (*Yp*HmuT, Q56991) and *Pseudomonas aeruginosa* (*Pa*PhuT, O68879); the protein name and UniProt id of each protein are provided in the parenthesis. The heme-binding residues are highlighted in green whereas, the presence of Arg218-Val219-Tyr220 involved in Fe^{3+} ion-binding of Rv3044 is represented in magenta. (B) Superposition of the protein Rv3044 and its

closest homolog *Ec*FecB from *E. coli* showing the differences in their active site residues. The protein *Ec*FecB and Rv3044 are shown in pink and green along with their active site residues in the blue and magenta ball-and-stick models, respectively. (C) Pairwise sequence alignment of the protein Rv3044 with its closest homolog *Ec*FecB. The conserved and unconserved active-site residues of the proteins Rv3044 and *Ec*FecB are highlighted in magenta and green, respectively.

Table B1. Details of the open reading frames (ORFs) identified as metal uptake ABC transporter proteins extracted from the TB, MycoBrowser and TubercuList databases.

S. No.	ORF number	UniProt id	+Protein name	*(SP) Protein length	Cellular location	*ABC subunit
1.	Rv1348	P9WQJ9	Mycobactin import ATP-binding/ permease protein IrtA	(-) 859	Membrane	Complete ABC transporter
2.	Rv1349	P9WQJ7	Mycobactin import ATP-binding /permease protein IrtB	(-) 579	Membrane	Incomplete ABC transporter
3.	Rv2895c	P9WL31	Uncharacterized protein	(-) 283	Periplasm	SBP
4.	Rv1819c	P9WQI9	Hydrophilic compounds import ATP-binding/ permease protein BacA	(-) 639	Membrane	Complete ABC transporter

5.	Rv0265c	L7N6B2	Probable periplasmic iron-transport lipoprotein	(39) 330	Periplasm	SBP
6.	Rv1857	P9WGU3	Molybdate-binding protein ModA	(21) 261	Periplasm	SBP
7.	Rv1858	P9WG13	Molybdenum transport system permease protein ModB	(-) 264	Membrane	TMD
8.	Rv1859	P9WQL3	Molybdenum import ATP-binding protein ModC	(-) 369	Cytosol	NBD
9.	Rv2059	O07257	Uncharacterized protein	(-) 511	Membrane	SBP
10.	Rv2060	O86339	Possible conserved integral membrane protein	(-) 133	Membrane	TMD
11.	Rv2325c	P9WPI7	Uncharacterized protein Rv2325c	(-) 282	Membrane	ECF-type ABC transporter
12.	Rv2326c	P9WQI7	Uncharacterized ABC transporter ATP-binding protein	(-) 697	Membrane	ECF-type ABC transporter
13.	Rv3044	O53291	Probable FEIII-dicitrate-binding periplasmic	(28) 359	Periplasm	SBP

			lipoprotein FecB			
14.	Rv3041c	I6YF11	Probable conserved ATP-binding protein ABC transporter	(-) 287	Membrane	NBD
15.	Rv0908	P9WPT1	Calcium-transporting ATPase CtpE	(-) 797	Membrane + cytoplasm	P-type ATPase transporter
16.	Rv1028c	P9WGL3	Sensor protein KdpD	(-) 860	Membrane	P-type ATPase transporter
17.	Rv1029	P9WKF3	Potassium-transporting ATPase potassium-binding subunit	(-) 571	Periplasm	
18.	Rv1030	P9WPU3	Potassium-transporting ATPase ATP-binding subunit	(-) 709	Cytosol	
19.	Rv1031	P9WKF1	Potassium-transporting ATPase KdpC subunit	(-) 189	Cytosol	

[+] Protein names are according to the UniProtKB database.

* SP: Signal peptide; SBP: Solute (substrate)-binding protein; TMD: Transmembrane domain; NBD: Nucleotide-binding domain.

Table B2. Protein annotation and UniProt id(s) of the ORFs interacting with the respective metal-binding ABC transporter proteins from *M. tuberculosis* H37Rv obtained in the interactome map analysis.

Target ORF id(s)	Details of interacting partner			
	Gene name	*ORF id	UniProt id	+Protein name
Rv0265c, Rv1857 and Rv3044	*oplA*	Rv0266c	P95223	Probable 5-oxoprolinase OplA (5-oxo-L-prolinase) (Pyroglutamase) (5-OPASE)
	PPE11	Rv0453	P9WI39	PPE family protein PPE11
	Rv0264c	Rv0264c	P95221	AHS1 domain-containing protein
	Rv2073c	Rv2073c	P9WGR3	Uncharacterized oxidoreductase Rv2073c
	Rv3041c	Rv3041c	I6YF11	Probable conserved ATP-binding protein ABC transporter
	mobA	Rv2453c	P9WJQ9	Probable molybdenum cofactor guanylyltransferase
	moaA1	Rv3109	P9WJS3	GTP 3',8-cyclase 1
	moeA1	Rv0994	P9WJQ7	Molybdopterin molybdenumtransferase 1
Rv2059	Rv2060	Rv2060	O86339	Possible conserved integral membrane protein
	Rv2061c	Rv2061c	O86340	Conserved protein
	mntH	Rv0924c	P9WIZ5	Divalent metal cation transporter MntH
	zur	Rv2359	P9WN85	Zinc uptake regulation protein
	oppB	Rv1283c	P9WFZ7	Putative peptide transport permease protein Rv1283c
	rpmB2	Rv2058c	P9WHA9	50S ribosomal protein L28-2
	rpmG1	Rv2057c	P9WH97	50S ribosomal protein L33 1

	rpsN2	Rv2056c	P9WH59	30S ribosomal protein S14
	rpsR2	Rv2055c	P9WH47	30S ribosomal protein S18 2
	Rv0106	Rv0106	P9WPI5	Uncharacterized protein Rv0106
Rv2326c	Rv2325c	Rv2325c	P9WPI7	Uncharacterized protein Rv2325c
	Rv2327	Rv2327	P71885	Conserved protein
	Rv0067c	Rv0067c	O53612	Possible transcriptional regulatory protein (Possibly TetR-family)
	Rv0879c	Rv0879c	P9WKR1	Uncharacterized protein Rv0879c
	Rv1780	Rv1780	O53931	Conserved protein
	Rv1610	Rv1610	O06128	Possible conserved membrane protein
	Rv0185	Rv0185	A0A0Y1B7U6	DUF2786 domain-containing protein
	Rv0494	Rv0494	P9WMG7	Uncharacterized HTH-type transcriptional regulator Rv0494
	Rv0493c	Rv0493c	P9WKU7	Uncharacterized protein Rv0493c

*Only the top 8-10 interacting proteins sorted by the best confidence score in the decreasing order are provided.

[+]Protein names are according to the UniProtKB database.

Table B3. Details of the proteins used as templates for three-dimensional model building for the targeted proteins.

Target ORF id	Template details			
	PDB id	**Organism name**	**Protein name**	**Bound ligand**
Rv1857	2H5Y	*Xanthomonas citri*	ModA	Molybdate
Rv2059	5KZJ	*Paracoccus denitrificans*	AztC	Zinc
Rv3044	3TNY	*Bacillus cereus*	YfiY	Fe^{3+}-schizokinen
b4290 (*Ec*FecB)	3EIW	*Staphylococcus aureus*	HtsA	-
Rv2325c	5JSZ	*Lactobacillus delbrueckii*	Folate ECF transporter	-
Rv2326c	5KC0	*Thermotoga maritima*	RibU	Riboflavin
Rv0908	3FPS	*Oryctolagus cuniculus*	Ca^{2+}-ATPase	Cyclopiazonic and ADP
Rv1030	5MRW	*Escherichia coli*	KdpFABC	Potassium

Table B4. Details of the top 20 structural homologs, if available, keeping one representative for each UniProt id (ranked by the Z-score) of the probable SBPs, ECF-type ABC and P-type ATPase transporter proteins obtained using the web-server Dali (Holm and Rosenstrom, 2010).

Target ORF id	Details of the structural homolog								
	PDB id	**UniProt id**	**+Protein name**	**Bound ligand**	***Z-score**	**RMSD (Å)**	**LALI**	**NRES**	**Id (%)**
Rv0265c	4MDY	A0QPL3	Periplasmic binding protein	-	42.9	0.8	282	285	65

	5JJ5	A0A062WLD9	ABC transporter substrate-binding protein	Ferrioxamine E and Fe^{3+}	27.9	3.3	285	304	22
	4H59	Q8DPY6	ABC transporter substrate-binding protein-iron transport	Bis-tris propane	27.6	3.3	285	306	22
	4HMQ	A0A0H2UPT5	Iron-compound ABC transporter, iron compound-binding protein	Ferrichrome	27.4	3.3	285	303	22
	3TNY	UPI00027FCE27	ABC transporter periplasmic binding domain	Fe^{3+} schizokinen	25.0	2.6	265	280	21
	3EIW	A0A0H3JM89	SA1979 protein	-	24.3	2.5	269	292	20
	3LHS	A0A0H3K9U6	Ferrichrome ABC transporter lipoprotein	Staphyloferrin A	24.2	2.5	269	291	20

	3MWF	A0A0H3JJC6	Iron-regulated ABC transporter siderophore-binding protein SirA	Staphyloferrin B	24.1	2.9	269	292	22
	4B8Y	Q2FVW9	Fe/B12 periplasmic-binding domain-containing protein	Ferrichrome	23.3	2.7	263	277	14
	6ALL	Q81L65	Petrobactin-binding protein FpuA	Tridecaoxahentetracontane-1,41-diol	22.8	2.8	267	289	18
	4FNA	Q7BGA5	ABC transporter substrate-binding protein	-	22.7	3.0	264	278	15
	3TLK	P0AEL6	Ferrienterobactin-binding periplasmic protein	Ferrienterobactin	22.7	3.1	270	296	19
	3BE6	Q8XBR1	Ferrichrome-	-	22.7	2.9	265	294	18

			binding protein						
	5FLY	A0A2P5JFB2	ABC transporter substrate-binding protein	-	22.6	2.9	262	277	16
	3R5S	Q9RCF6	Fe2+-enterobactin ABC transporter substrate-binding protein	-	22.0	3.1	266	301	19
	3HXP	P37580	Iron(3+)-hydroxamate-binding protein FhuD	-	21.9	2.8	257	272	16
Rv1857	3GZG	Q8PHA1	Molybdate-binding protein ModA	Molybdate	38.5	0.6	220	232	33
	1AMF	P37329	Molybdate-binding protein ModA	Molybdate	35.8	1.0	220	231	30
	4RXL	Q9KLL7	Molybdenum ABC transporter, periplasmi	Tungstate	33.7	1.2	218	229	24

			c molybdenum-binding protein						
	6NIO	UPI000FE182A6	Molybdate ABC transporter substrate-binding protein	Formic acid	29.3	2.7	220	231	29
	1ATG	Q7SIH2	Binding protein	Tungstate	29.2	1.8	215	231	20
	4KD5	Q18A64	ABC-type transport system, molybdenum-specific extracellular solute-binding protein	Formic acid	25.3	3.3	217	229	25
	3CFZ	Q58586	Molybdate/tungstate-binding protein WtpA	Tungstate	24.6	1.9	214	289	24
	2ONR	O30142	Molybdate/tungstate-binding protein WtpA	Molybdate	24.2	1.9	216	314	24

	1SBP	P02906	Sulfate-binding protein	Sulphate	23.0	2.2	217	309	17
	4R74	A3N294	ABC-type Fe^{3+} transport system, periplasmic component	Fructose-6-phosphate	20.0	2.7	214	319	13
	3LR1	Q749P2	Tungstate ABC transporter, periplasmic tungstate-binding protein	Tungstate	19.8	2.8	205	232	15
	3FIR	Q0PBL7	Major antigenic peptide PEB3	Citrate	19.7	3.1	207	231	16
	4Q8R	A0A0H2YSI2	Phosphate ABC transporter, phosphate-binding protein	Phosphate	19.7	2.7	205	241	14
Rv2059	5KZJ	A1B2F3	Periplasmic solute	Zinc	36.7	0.8	254	272	21

			binding protein						
	5I4K	Q8Y653	Manganese-binding lipoprotein MntA	Manganese	31.6	1.5	253	280	20
	3HH8	P0A4G4	Metal ABC transporter substrate-binding lipoprotein	Ferric ion	31.4	1.6	254	280	16
	1XVL	Q79EF9	Mn^{2+} transporter MntC	Manganese	31.4	1.4	253	271	21
	3ZTT	P0A4G2	Manganese ABC transporter substrate-binding lipoprotein	Manganese	31.1	1.6	254	278	17
	5UYB	Q56952	Periplasmic chelated iron-binding protein YfeA	Ferric ion	30.9	1.5	253	274	17
Rv3044	3TNY	UPI00027FCE27	ABC transport system substrate-binding protein	Fe^{3+}-schizokinen	44.1	0.5	274	280	23

	3MWF	A0A0H3JJC6	Iron-regulated ABC transporter siderophore-binding protein SirA	staphyloferrin B	33.9	1.8	272	292	23
	3LI2	A0A0H3K9U6	Ferrichrome ABC transporter lipoprotein	Staphyloferrin A	28.7	2.1	265	289	20
	3EIW	A0A0H3JM89	SA1979 protein	Chlorine	27.7	2.2	265	292	20
	5FLY	A0A2P5JFB2	ABC transporter substrate-binding protein	-	25.6	2.6	262	279	19
	4B8Y	Q2FVW9	Fe/B12 periplasmic-binding domain-containing protein	Ferrichrome	25.5	2.4	262	277	18
	3BE6	Q8XBR1	Ferrichrome-binding protein	-	25.2	2.7	264	294	21
	3GFV	P94421	Petrobactin-binding protein YclQ	-	25.2	2.8	268	286	16

	6ALL	Q81L65	Petrobactin-binding protein fpua	Tridecaoxahentetracontane-1,41-diol	25.0	2.6	266	289	18
	3HXP	P37580	Iron(3+)-hydroxamate-binding protein FhuD	-	24.8	2.4	258	272	17
	5OD5	Q0P8Q4	Enterochelin uptake periplasmic binding protein	Fe^{3+}-Azotochelin	24.6	2.8	264	288	21
	2XUZ	P40409	Iron-uptake system-binding protein	Ferri-Enterobactin	24.5	2.9	266	281	15
	5A5D	Q0P8Q4	Enterochelin uptake periplasmic binding protein	Fe(III)-5-LICAM	24.5	2.8	264	289	21
	4PM4	L7N6B2	Probable periplasmic iron-transport lipoprotein	-	24.4	2.6	264	290	18

Rv2325c	4HUQ	Q03PY7	Energy-coupling factor transporter transmembrane protein EcfT	-	23.3	1.5	242	245	21
	5JSZ	A0A061BSU4	Energy-coupling factor transporter transmembrane protein EcfT	-	22.4	2.1	251	259	16
	5D3M	Q1GBI8	Energy-coupling factor transporter transmembrane protein EcfT	-	21.8	2.0	249	259	16
	5X3X	D5AUZ7	Cobalt transport protein CbiQ	-	14.0	3.3	227	242	17
	6IZ2	B5SYS5	Major ampullate spidroin 1A	-	6.0	3.3	91	145	7

	6UTG	P28061	Proteasome subunit beta	-	6.0	9.9	111	218	12
	6BFI	A0A3B6UES5	Vinculin	-	5.9	9.5	122	806	7
	6IRR	Q06507	Cyclic AMP-dependent transcription factor ATF-4	-	5.9	3.3	82	133	6
	6XMJ	P60900	Proteasome subunit alpha type-6	-	5.9	10.1	109	207	10
	6IUH	Q9Z272	ARF GTPase-activating protein GIT1	-	5.9	4.3	86	123	10
	5MSJ	P97371	Proteasome activator complex subunit 1	-	5.8	9.2	98	204	6
	4FIX	O53585	Galactofuranosyltransferase GlfT2	Manganese	5.8	11.3	85	629	9
	6KCR	P0AF10	Mannitol operon repressor	-	5.8	5.0	95	169	6
	6FUY	P18206	Vinculin	Calcium	5.7	5.6	114	965	9

	3C8G	Q83Q96	Fumarase E	-	5.7	3.9	87	166	9
Rv2326c	6S7P	P21439	Phosphatidylcholine translocator ABCB4	Magnesium, ATP	31.6	4.5	487	1128	20
	6MSM	P13569	Cystic fibrosis transmembrane conductance regulator	Magnesium, ATP	28.8	15.6	496	1181	17
	4HUQ	Q03PY6	Energy-coupling factor transporter ATP-binding protein EcfA2	-	27.1	1.0	222	284	33
	6TMF	Q980K5	RNase L inhibitor	Magnesium, Phosphoamino phosphonic acid adenylate ester	27.0	13.4	409	582	20
	4MKI	Q8R7Y5	Energy-coupling factor	-	26.3	1.0	222	285	31

			transporter ATP-binding protein EcfA2						
	6C0V	P08183	ATP-dependent translocase ABCB1	Magnesium, ATP	26.2	4.4	504	1154	20
	4RFS	Q1GBI9	Energy-coupling factor transporter ATP-binding protein EcfA2	-	25.6	1.0	222	284	33
	5W81	Q1LX78	Cystic fibrosis transmembrane conductance regulator	Magnesium, ATP	25.1	9.9	480	1173	20
	6C3P	Q09428	ATP-binding cassette sub-family C member 8	Magnesium, ADP, ATP	24.9	8.6	504	1359	18
	4HLU	Q9WY65	Energy-coupling factor	ADP	24.2	1.5	220	265	34

			transporter ATP-binding protein EcfA2						
	5X40	O68106	Cobalt import ATP-binding protein CbiO	Magnesium, Phosphomethylphosphonic acid adenylate ester	23.2	2.0	216	280	34
	5LL6	Q03195	Translation initiation factor RLI1	Magnesium, Phosphoaminophosphonic acid adenylate ester	23.0	14.6	407	278	20
	6O1V	P13569	Cystic fibrosis transmembrane conductance regulator	Magnesium, ATP	22.9	17.9	499	1179	17
	3TUJ	P30750	Methionine import ATP-	-	22.8	1.7	220	344	27

			binding protein MetN						
Rv0908	4UU0	P04191	Sarcoplasmic/endoplasmic reticulum calcium ATPase 1	Octanoic acid, Tertiary-butyl alcohol, Potassium, Magnesium	44.3	1.6	767	994	23
	5MPM	P11607	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	Cyclopiazonic Acid, Magnesium, Potassium	41.7	3.9	759	992	23
	3NAM	B6CAM1	Calcium-transporting ATPase	Thapsigargin, Magnesium, Sodium	40.6	1.6	765	994	23
Rv1030	5MRW	P03960	Potassium-transporting ATPase ATP-binding subunit	Potassium	51.1	0.3	668	674	64
	3AR5	P04191	Sarcoplasmic/endop	Phosphatidylet	18.2	7.4	452	994	20

			lasmic reticulum calcium ATPase 1	hanola mine, Soidum , Octanoi c acid					
	3J09	O29777	Probable copper-exporting P-type ATPase	-	17.9	7.5	346	723	21
	3NAL	B6CA M1	Calcium-transportin g ATPase	Potassi um	17.5	6.3	376	994	21
	4BEV	Q5ZW R1	Copper-exporting P-type ATPase	Trifluor omagne sate, Magnes sium	16.8	7.2	362	663	21

[+] Protein names are according to the UniProtKB database.

* Z-score = [(S(A, B) – m(N))/ (σ(N)], where S(A, B) = similarity score between protein A and B; m(N) = mean score of average length N; σ = average standard deviation.

Abbreviations: rmsd: root-mean-square deviation; LALI: length of the alignment; NRES: number of residues; Id: sequence identity.

Table B5. Details of the proteins used for generating the structure-based distance tree for SBP classification.

Cluster	PDB id	UniProt id	Protein name	Organism name	Known ligand(s)
A-I	1PQ4	P73085	ZntC	*Synechocystis sp.*	Zinc
	1TOA	P96116	TroA	*Treponema pallidum*	Zinc
	1XVL	Q79EF9	MntC	*Synechocystis sp.*	Manganese
	3MFQ	A4VY63	TroA	*Streptococcus suis*	Zinc
	4K3V	D9RF12	MntC	*Staphylococcus aureus*	Manganese
A-II-a	2R7A	O70018	ShuT	*Shigella dysenteriae*	Heme
	2R79	O68879	PhuT	*Pseudomonas aeruginosa*	Heme
	3NU1	A0A0E1NWS3	HmuT	*Yersinia pestis*	Heme
	5AZ3	Q8NTB8	HmuT	*Corynebacterium glutamicum*	Heme
	5GJ1	B4EKB3	BhuT	*Burkholderia cenocepacia*	Heme
A-II-b	3EIW	A0A0H3JM89	HtsA	*Staphylococcus aureus*	Staphyloferrin A
	3TNY	UPI00027FCE27	YfiY	*Bacillus cereus*	Iron (III) schizokinen
	4HN9	C4Z2B5	-	*Eubacterium eligens*	Iron complex
D-I	2B3B	Q72KX2	GGBP	*Thermus thermophilus*	Glucose
	3OO6	Q27GR2	AcbH	*Actinoplanes sp.*	Galactose
	3ZKK	A0A0J9X1B4	-	*Bifidobacterium animalis*	Sugar
	4G68	J9PBT4	-	*Caldanaerobius*	-
	4R2B	A6X5L9	-	*Ochrobactrum anthropi*	Sugar
D-II	1A99	P31133	PotF	*Escherichia coli*	Putrescine

	1POT	P0AFK9	PotD	*Escherichia coli*	Spermidine, Putrescine
	3RPW	Q6N0W2	-	*Rhodopseudomonas palustris*	-
	3TTM	Q9I6J1	SpuD	*Pseudomonas aeruginosa*	Putrescine
	4GL0	Q8Y8T4	-	*Listeria monocytogenes*	Spermidine, Putrescine
D-III-a	1ATG	Q7SIH2	ModA	*Azotobacter vinelandii*	Tungstate (VI)
	1WOD	P37329	ModA	*Escherichia coli*	Molybdate
	2H5Y	Q8PHA1	ModA	*Xanthomonas axonopodis*	Molybdate
	3CVG	J3KFC8	-	*Coccidioides immitis*	Metal
	3LR1	Q749P2	TupA	*Geobacter sulfurreducens*	Tungstate
	4RXL	Q9KLL7	-	*Vibrio cholerae*	Molybdate
D-III-b	2ONR	O30142	WtpA/ModA	*Archaeoglobus fulgidus*	Molybdate/ Tungstate
	3CFX	Q8TTZ5	ModA	*Methanosarcina acetivorans*	Tungstate
	3CFZ	Q58586	WtpA	*Methanocaldococcus jannaschii*	Molybdate/ Tungstate
	3CG1	Q8U4K5	WtpA	*Pyrococcus furiosus*	Molybdate/ Tungstate
	3CG3	O57890	WtpA	*Pyrococcus horikoshii*	Molybdate/ Tungstate
D-IV	1MRP	P35755	FbpA	*Haemophilus influenzae*	Iron
	1SI0	Q9Z4N6	FbpA	*Mannheimia haemolytica*	Iron
	1XVX	A1JLH5	YfuA	*Yersinia enterocolitica*	Iron

	1Y4T	Q0PBW4	CfbpA	*Campylobacter jejuni*	Iron
	2PT2	P72827	FutA1	*Synechocystis sp.*	Iron
	3C9H	Q7CWZ6	Atu	*Agrobacterium fabrum*	Mg^{2+}-citrate

Table B6. Details of the molecular docking results of various metal ions and/or their complexes with targeted SBPs.

Target ORF id	**Ligand**	**EFBE (kcal mol^{-1})**	**NHBOND**	**Interactions**		
				Ligand atom		**Protein atom(s)**
Rv0265c	Ferrichrome	-8.18	3	O2		Trp58 $N^{\varepsilon 1}$
				O6		Gln182 $O^{\varepsilon 1}$
				O8		Gln182 $O^{\varepsilon 1}$
	Heme	-11.23	9	Fe		Trp58 $N^{\varepsilon 1}$
				NB		Tyr39 O^{η}
				NC		Trp58 $N^{\varepsilon 1}$
				O1D		Arg198 N^{ε}
				O2D		Arg198 $N^{\eta\varepsilon}$, Arg198 $N^{\eta 1}$
				O1A		Arg221 $N^{\eta 2}$
				O2A		Arg198 NH1, Arg221 $N^{\eta 2}$
b4290 (*Ec*FecB)	Fe^{3+}-dicitrate	-5.96	8	FLC742	OG1	Tyr216 N, Ser247 O^{γ}
					OG2	Tyr216 N
				FLC743	OA1	Arg66 $N^{\eta 2}$, Arg252 N^{ε}
					OA2	Arg108 $N^{\eta 1}$
					OG1	Ser247 O

					OG2	Ser247 O^{γ}
Rv3044	Fe^{3+}-citrate	-4.84	13	Fe		Gly162 N, Glu317 $O^{\varepsilon 2}$
				O1		Glu317 $O^{\varepsilon 2}$
				O2		Gln314 $N^{\varepsilon 2}$, Gln314 O
				O3		Gly162 N
				O4		Gly162 O, Gln314 $N^{\varepsilon 2}$, Asn224 $N^{\delta 2}$, Phe225 N
				O5		Arg218 $N^{\eta 1}$, Tyr220 O^{η}
				O6		Tyr220 O^{η}
	Fe^{3+}-dicitrate	-6.72	17	Fe744		Arg218 $N^{\eta 1}$, Tyr220 O^{η},
				Fe745		Gln314 $N^{\varepsilon 2}$
				FLC742	OA1	Thr315 O, Gln314 O
					OA2	Glu317 $O^{\varepsilon 2}$
					OB1	Gln211 $N^{\varepsilon 2}$, Tyr220 O^{η}
					OB2	Tyr220 O^{η}
					OG1	Arg218 $N^{\eta 1}$ and $N^{\eta 2}$
				FLC743	OA1	Gln314 $^{\varepsilon 2}$

					OA2	Gly162 O, Asn224 $N^{\delta 2}$, Phe225 N, Gln314 $^{\varepsilon 2}$
					OG2	Arg218 $N^{\eta 1}$
Rv1857	Molybdate	-3.01	7	O1		Ser156 O^{γ}, Val157 N
				O2		Ser54 N, Ser54 O^{γ}
				O3		Cys133 N, Tyr175 O^{η}
				O4		Ser26 N
	Tungstate	-2.13	6	O1		Ser156 O^{γ}, Val157 N
				O3		Cys133 N
				O4		Ser54 N, Ser54 O^{γ}
	Sulphate	-				

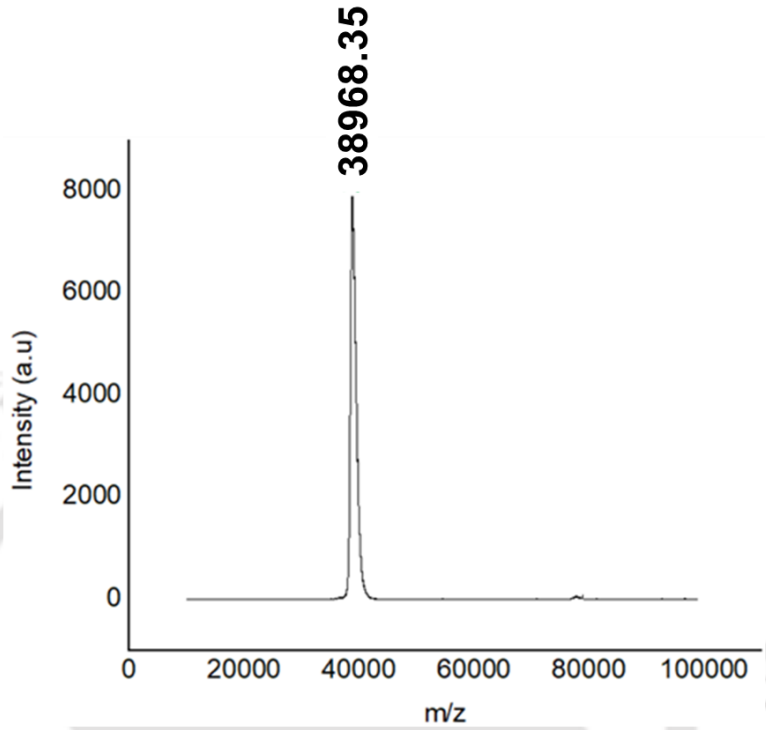
Abbreviations: EFBE, estimated free energy of binding; NHBOND, number of hydrogen bonds.

APPENDIX C

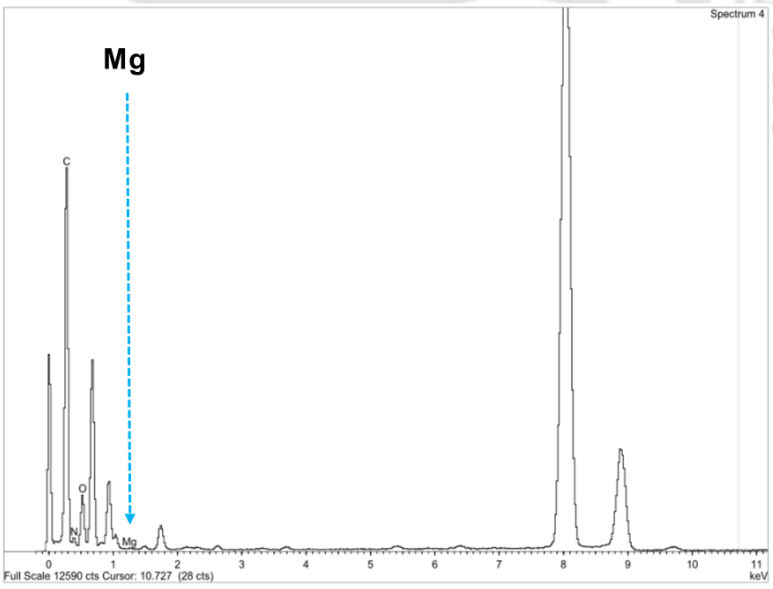
Supplementary data to chapter 5

APPENDIX C

(A)



(B)



(C)

Element	Weight (%)	Atomic (%)
Carbon	85.83	88.97
Nitrogen	0.47	0.42
Oxygen	13.54	10.54
Magnesium	0.16	0.08

Figure C1. (A) Mass spectrometric analysis of the protein MctA_WT using MALDI-TOF. The obtained molecular weight of the protein in Daltons (unit) is labeled. (B) Energy dispersive X-ray **(**EDX) graph for the protein MctA_WT. A downward cyan arrow indicates the presence of Mg^{2+}. (C) The total elemental content of the protein solution is provided.

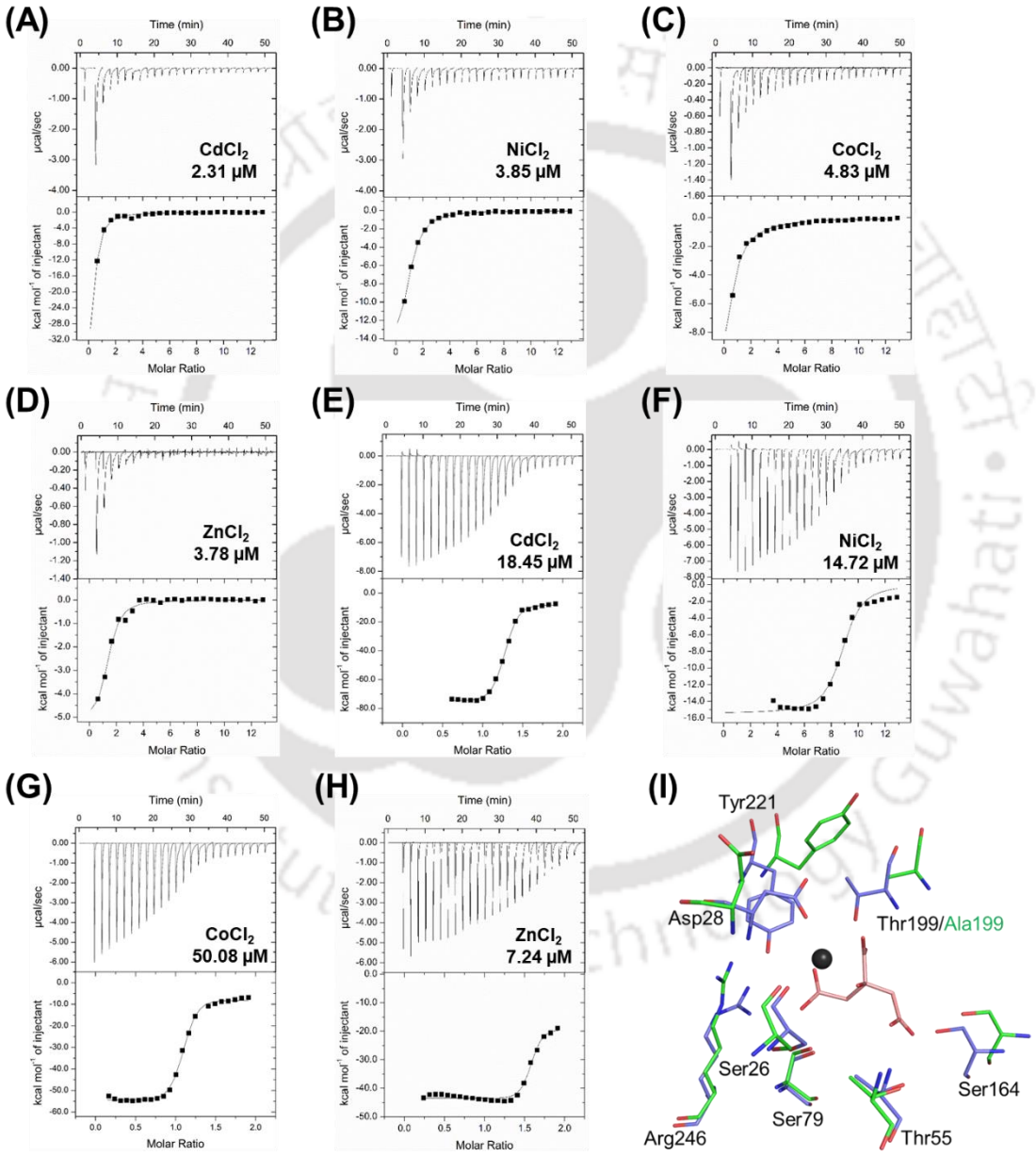


Figure C2. Isothermal titration calorimetry of MctA_S79A mutant protein with (A) $CdCl_2$, (B) $NiCl_2$, (C) $CoCl_2$ and (D) $ZnCl_2$. ITC of the protein MctA_S164A with (E) $CdCl_2$, (F) $NiCl_2$, (G) $CoCl_2$ and (H) $ZnCl_2$. (I) Superposition of the active-site residues

of the ligand-bound MctA_WT protein and ligand unbound MctA_T199A mutant. The active-site residues of the proteins MctA_WT and mutants are shown in blue and green sticks, respectively, with the bound ligand in the wild type protein in the black sphere (Mg^{2+}) and pink stick (citrate).

Table C1. List of wild type and mutant proteins of MctA along with Atu2467 protein with their crystallization buffers.

Sl. No.	Protein	Space group	Ligand	Crystallization buffer
1	MctA_WT	$P2_1$	Mg^{2+}-citrate bound	0.1 M Sodium citrate tribasic dihydrate pH 5.5, 16% PEG 8000
2	MctA_WT	$P2_12_12_1$	,,	0.2 M Ammonium fluoride, 20 % PEG 3350
3	MctA_WT	$P2_1$	Citrate bound	0.2 M Ammonium sulfate, 0.1 M MES Monohydrate, pH 6.5, 30% w/v PEG monoethyl ether 5000
4	MctA_S26A	$P2_1$	Mg^{2+}-citrate bound	0.2 M Ammonium fluoride, 20 % PEG 3350
5	MctA_S26A	$P2_12_12_1$	Mn^{2+}-citrate bound	0.2 M Ammonium fluoride, 20 % PEG 3350
6	MctA_D28A	$P2_1$	Mg^{2+}-citrate bound	0.2 M Ammonium phosphate dibasic, 20 % PEG 3350
7	MctA_S79A	$P3_121$	Unbound	0.2 M Ammonium sulfate, 0.1 M MES monohydrate pH 6.5, 30% w/v PEG monomethyl ether 5000
8	MctA_S164A	$I222$	Unbound	0.01 M $Co(II)Cl_2$ hexahydrate, 0.1 M MES monohydrate pH 6.5, 1.8 M Ammonium sulfate

9	MctA_T199A	*I*222	Unbound	0.2 M Ammonium tartarate dibasic pH 7.0, 20% PEG 3350
10	MctA_Y221A	*I*222	Unbound	0.2 M Ammonium sulfate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 30 % w/v PEG 8000
11	MctA_Y221F	*I*222	Unbound	0.2 M Ammonium phosphate dibasic, 20% PEG 3350
12	Atu2467 (PDB id: 3C9H)	$P2_12_12_1$	Mg^{2+}-citrate bound	0.1M Sodium citrate, 0.2M Ammonium acetate, 30% PEG 4000, 10mM ATP, pH 5.6,

Table C2. List of bacteria having proteins homologous to the protein MctA with their Gram's property. The table suggests that homologous proteins of MctA are present only in Gram-negative bacteria.

Microorganism (Genus name)	Gram's property	Query coverage (%)	Sequence identity (%)
Homology search against non-redundant proteins			
Thermus spp.	Gram-negative	96-100	89-98
Deinococcus spp.	Gram-negative	93	55-58
Proteobacteria spp.	Gram-negative	98	49
Deltaproteobacteria spp.	Gram-negative	99	48
Massilia spp.	Gram-negative	92-98	47-49
Azospirullum spp.	Gram-negative	99	47
Rhizobiales spp.	Gram-negative	96	47
Duganella spp.	Gram-negative	92-98	47
Pseudoduganella spp.	Gram-negative	92	47
Janthinobacteirum spp.	Gram-negative	92-96	45-48
Rugamonas	Gram-negative	96	45
Homology search against protein data bank			
Agrobacterium spp.	Gram-negative	89	37
Actinobacillus spp.	Gram-negative	71	24

Table C3. Details of the proteins belonging to cluster D used for generating the structure-based distance tree for SBP classification.

Cluster	PDB id	UniProt id	Protein name	Organism name	Known ligand(s)
D	3K01	B0B0V1	GacH	*Streptomyces glaucescens*	Acarbose
	3UOR	Q8PK66	MalE	*Xanthomonas citri*	Maltose
	4MFI	A5U6I5	UgpB	*Mycobacterium tuberculosis*	Glycerophosphocholine
	3VXB	Q76BU9	BxlE	*Streptomyces thermoviolaceus*	Xylooligosaccharide
	4H1X	P0C2M5	PstS2	*Streptococcus pneumoniae*	Phosphate
D-I	2B3B	Q72KX2	GGBP	*Thermus thermophilus*	Glucose
	3OO6	Q27GR2	AcbH	*Actinoplanes sp.*	Galactose
	3ZKK	A0A0J9X1B4	-	*Bifidobacterium animalis*	Sugar
	4G68	J9PBT4	-	*Caldanaerobius*	-
	4R2B	A6X5L9	-	*Ochrobactrum anthropi*	Sugar
D-II	1A99	P31133	PotF	*Escherichia coli*	Putrescine
	1POT	P0AFK9	PotD	*Escherichia coli*	Spermidine, Putrescine
	3RPW	Q6N0W2	-	*Rhodopseudomonas palustris*	-
	3TTM	Q9I6J1	SpuD	*Pseudomonas aeruginosa*	Putrescine
	4GL0	Q8Y8T4	-	*Listeria monocytogenes*	Spermidine, Putrescine
D-III-a	1ATG	Q7SIH2	ModA	*Azotobacter vinelandii*	Tungstate (VI)
	1WOD	P37329	ModA	*Escherichia coli*	Molybdate

	2H5Y	Q8PHA1	ModA	*Xanthomonas axonopodis*	Molybdate
	3CVG	J3KFC8	-	*Coccidioides immitis*	Metal
	3LR1	Q749P2	TupA	*Geobacter sulfurreducens*	Tungstate
	4RXL	Q9KLL7	-	*Vibrio cholerae*	Molybdate
D-III-b	2ONR	O30142	WtpA/ ModA	*Archaeoglobus fulgidus*	Molybdate/ Tungstate
	3CFX	Q8TTZ5	ModA	*Methanosarcina acetivorans*	Tungstate
	3CFZ	Q58586	WtpA	*Methanocaldococcus jannaschii*	Molybdate/ Tungstate
	3CG1	Q8U4K5	WtpA	*Pyrococcus furiosus*	Molybdate/ Tungstate
	3CG3	O57890	WtpA	*Pyrococcus horikoshii*	Molybdate/ Tungstate
D-IV-a	1MRP	P35755	FbpA	*Haemophilus influenzae*	Iron
	1SI0	Q9Z4N6	FbpA	*Mannheimia haemolytica*	Iron
	1XVX	A1JLH5	YfuA	*Yersinia enterocolitica*	Iron
	1Y4T	Q0PBW4	CfbpA	*Campylobacter jejuni*	Iron
	2PT2	P72827	FutA1	*Synechocystis sp.*	Iron
D-IV-b	3C9H	Q7CWZ6	Atu	*Agrobacterium fabrum*	Mg^{2+}-citrate

Table C4. The binding affinity of various divalent metal ions and sodium citrate to the various mutant protein of MctA.

Mutant proteins	$MgCl_2$	$MnSO_4$	$CdCl_2$	$NiCl_2$	$CoCl_2$	$ZnCl_2$	Fe_2SO_4	Na-citrate
	K_d (µM)							
MctA_S26A	-	0.11	-	-	-	-	-	-
MctA_D28A	-	-	-	-	-	-	-	-
MctA_T55A	-	-	-	-	-	-	-	-
MctA_S79A	8.33	0.89	2.31	3.85	4.83	3.78	-	-
MctA_S164A	3.60	0.63	18.45	14.72	52.08	7.24	-	-
MctA_T199A	-	-	-	-	-	-	-	-
MctA_Y221A	-	-	-	-	-	-	-	-
MctA_Y221F	-	-	-	-	-	-	-	-
MctA_R246A	-	-	-	-	-	-	-	-

- No binding

APPENDIX D

Supplementary data to chapter 6

APPENDIX D

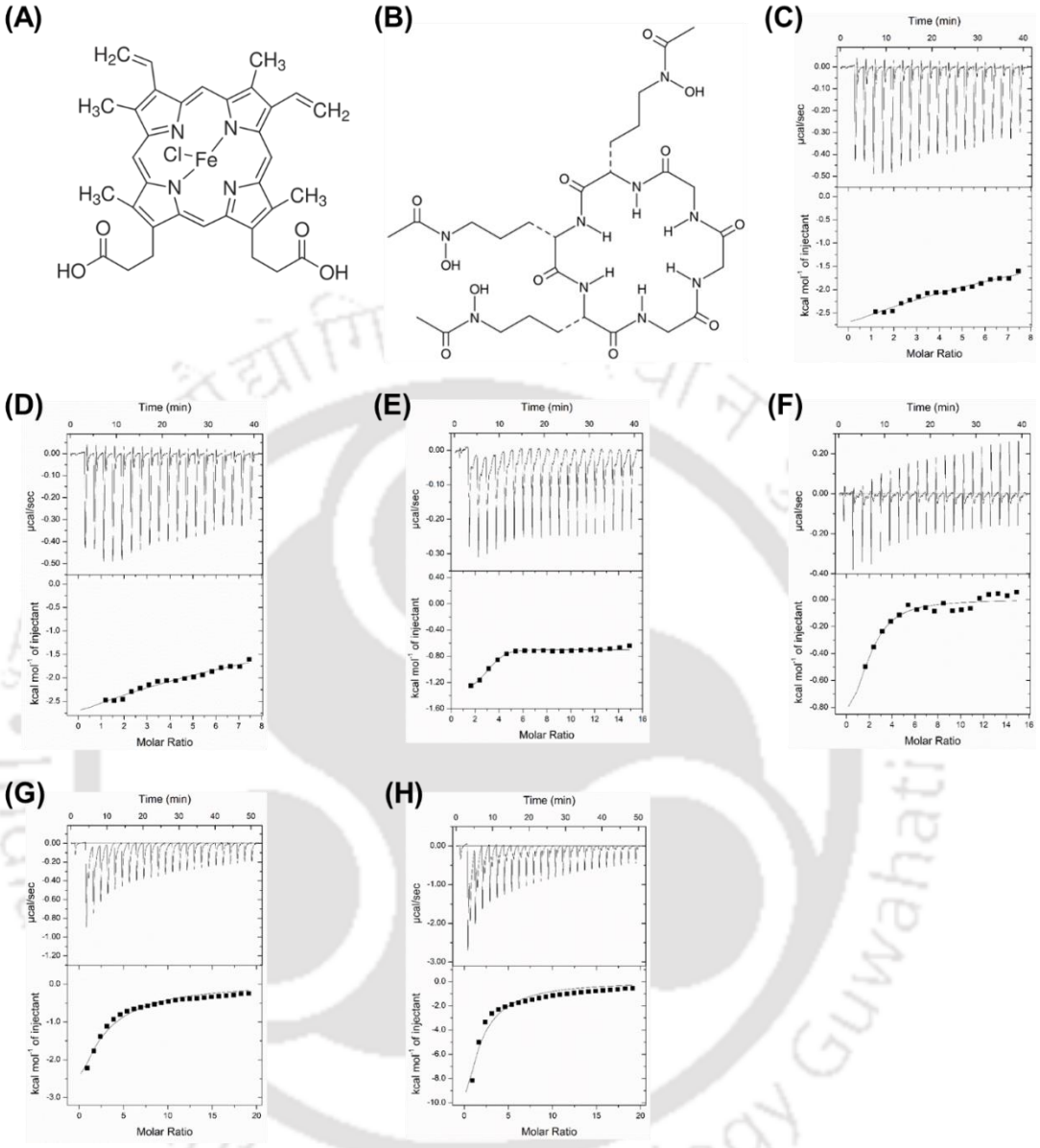


Figure D1. Binding isotherm for the protein MhuP with different ligands using ITC. The schematic two-dimensional chemical structures of (A) heme and (B) ferrichrome. ITC for the protein MhuP with heme prepared in buffer F (20 mM Tris-HCl (pH 7.5), 150 mM NaCl and 5% glycerol) prior to solubilization in (C) 100% DMSO and (D) 10 mM NaOH. ITC of the protein MhuP with (E) ferreted ferrichrome, (F) non-ferreted ferrichrome, (G) $CoCl_2$ and (H) $NiCl_2$. The protein MhuP shows no binding for heme and ferrichrome, whereas it shows binding with Co^{2+} and Ni^{2+} ions; however, the

thermodynamic parameters obtained for the binding were beyond ITC limits for both the metal ions.

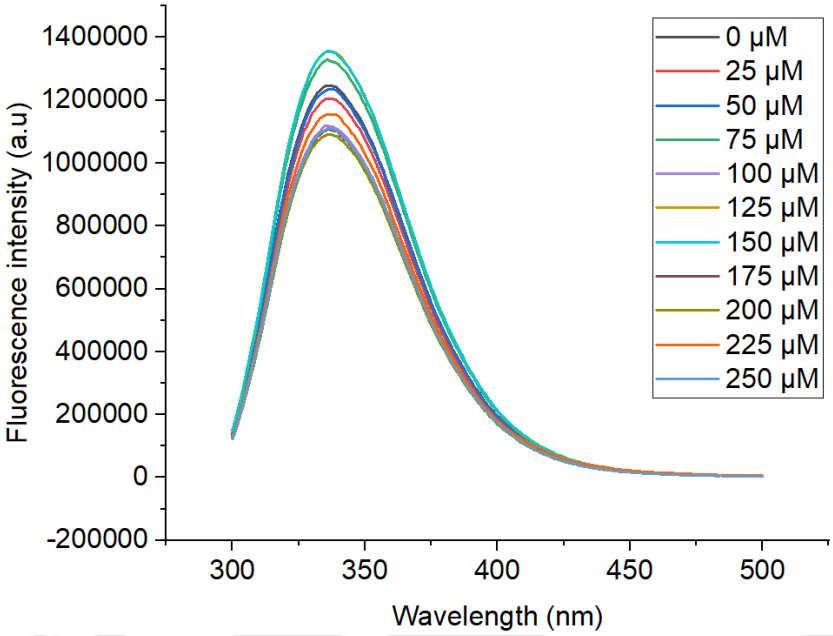


Figure D2. Fluorescence emission spectra of the protein MhuP with a dipeptide glycyl glutamate. The change in fluorescence spectra of the protein MhuP has been represented in different colors and is labeled for the varying concentrations of ligands.

Table D1. Details of the top 20 structural homologs, if available, keeping one representative for each UniProt id (ranked by the Z-score) for the protein MhuP obtained using the web-server Dali (Holm and Rosenstrom, 2010).

Details of the structural homologs of the protein MhuP								
PDB id	**UniProt id**	**[+]Protein name**	**Bound ligand**	***Z-score**	**rmsd (Å)**	**LALI**	**NRES**	**Id (%)**
4MDY	A0QPL3	Periplasmic binding protein	-	42.9	0.8	282	285	65
5JJ5	A0A062WLD9	ABC transporter substrate-binding protein	Ferrioxamine E and Fe^{3+}	27.9	3.3	285	304	22
4H59	Q8DPY6	ABC transporter substrate-	Bis-tris propane	27.6	3.3	285	306	22

		binding protein-iron transport						
4HMQ	A0A0H2UPT5	Iron-compound ABC transporter, iron compound-binding protein	Ferrichrome	27.4	3.3	285	303	22
3TNY	UPI00027FCE27	ABC transporter periplasmic binding domain	Fe^{3+} schizokinen	25.0	2.6	265	280	21
3EIW	A0A0H3JM89	SA1979 protein	-	24.3	2.5	269	292	20
3LHS	A0A0H3K9U6	Ferrichrome ABC transporter lipoprotein	Staphyloferrin A	24.2	2.5	269	291	20
3MWF	A0A0H3JJC6	Iron-regulated ABC transporter siderophore-binding protein SirA	Staphyloferrin B	24.1	2.9	269	292	22
4B8Y	Q2FVW9	Fe/B12 periplasmic-binding domain-containing protein	Ferrichrome	23.3	2.7	263	277	14
6ALL	Q81L65	Petrobactin-binding protein FpuA	Tridecaoxahentetracontane-1,41-diol	22.8	2.8	267	289	18
4FNA	Q7BGA5	ABC transporter substrate-binding protein	-	22.7	3.0	264	278	15
3TLK	P0AEL6	Ferrienterobactin-binding periplasmic protein	Ferrienterobactin	22.7	3.1	270	296	19

3BE6	Q8XBR1	Ferrichrome-binding protein	-	22.7	2.9	265	294	18
5FLY	A0A2P5JFB2	ABC transporter substrate-binding protein	-	22.6	2.9	262	277	16
3R5S	Q9RCF6	Fe^{2+}-enterobactin ABC transporter substrate-binding protein	-	22.0	3.1	266	301	19
3HXP	P37580	Iron(III)-hydroxamate-binding protein FhuD	-	21.9	2.8	257	272	16

[+] Protein names are according to the UniProtKB database.

* Z-score = [(S(A, B) – m(N))/ (σ(N)], where S(A, B) = similarity score between protein A and B; m(N) = mean score of average length N; σ = average standard deviation.

Abbreviations: rmsd: root-mean-square deviation; LALI: length of the alignment; NRES: number of residues; Id: sequence identity.

BIBLIOGRAPHY

1. Abate F, Malito E, Cozzi R, Lo Surdo P, Maione D and Bottomley MJ, 2014. Apo, Zn^{2+}-bound and Mn^{2+}-bound structures reveal ligand-binding properties of SitA from the pathogen *Staphylococcus pseudintermedius*. *Biosci. Rep.*, **34**. e00154.
2. Agranoff DD and Krishna S, 1998. Metal ion homeostasis and intracellular parasitism. *Mol. Microbiol*., **28**, 403-412.
3. Alarico S, Empadinhas N, Simoes C, Silva Z, Henne A, Mingote A, Santos H and da Costa MS, 2005. Distribution of genes for synthesis of trehalose and mannosylglycerate in *Thermus* spp. and direct correlation of these genes with halotolerance. *Appl. Environ. Microbiol*., **71**, 2460-2466.
4. Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ, 1990. Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403-410.
5. Ambudkar SV, Kim IW, Xia D and Sauna ZE, 2006. The A-loop, a novel conserved aromatic acid subdomain upstream of the Walker A motif in ABC transporters, is critical for ATP binding. *FEBS Lett.*, **580**, 1049-1055.
6. Andreini C, Banci L, Bertini I and Rosato A, 2006. Zinc through the three domains of life. *J. Proteome Res.*, **5**, 3173-3178.
7. Andreini C, Bertini I, Cavallaro G, Holliday GL and Thornton JM, 2008. Metal ions in biological catalysis: from enzyme databases to general principles. *J. Biol. Inorg. Chem.*, **13**, 1205-1218.
8. Andreini C, Cavallaro G, Lorenzini S, and Rosato A, 2013. MetalPDB: a database of metal sites in biological macromolecular structures. *Nucleic acids Res*., **41**, D312-D319.
9. Andrews S, Norton I, Salunkhe AS, Goodluck H, Aly WS, Mourad-Agha H and Cornelis P, 2013. Control of iron metabolism in bacteria. *Metallomics and the Cell*, Springer, Dordrecht, 203-239.
10. Arnold FM, Weber MS, Gonda I, Gallenito MJ, Adenau S, Egloff P, Zimmermann I, Hutter CA, Hürlimann LM, Peters EE and Piel J, 2020. The ABC exporter IrtAB imports and reduces mycobacterial siderophores. *Nature*, **580**, 413-417.
11. Balamurugan B, Roshan MN, Shaahul Hameed B, Sumathi K, Senthilkumar R, Udayakumar A, Venkatesh Babu KH, Kalaivani M, Sowmiya G, Sivasankari P and Saravanan S, 2007. PSAP: protein structure analysis package. *J. Appl. Crystallogr*., **40**, 773-777.

12. Banerjee S, Paul S, Nguyen LT, Chu BC and Vogel HJ, 2016. FecB, a periplasmic ferric-citrate transporter from *E. coli*, can bind different forms of ferric-citrate as well as a wide variety of metal-free and metal-loaded tricarboxylic acids. *Metallomics*, **8**, 125-133.
13. Battye TGG, Kontogiannis L, Johnson O, Powell HR and Leslie AG, 2011. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D: Biol. Crystallogr.*, **67**, 271-281.
14. Bearden SW and Perry RD, 1999. The Yfe system of *Yersinia pestis* transports iron and manganese and is required for full virulence of plague. *Mol. Microbiol.*, **32**, 403-414.
15. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA and Wheeler DL, 2000. GenBank. *Nucleic Acids Res.*, **28**, 15-18.
16. Berger EA, 1973. Different mechanisms of energy coupling for the active transport of proline and glutamine in *Escherichia coli. Proc. Natl. Acad. Sci*. USA, **70**, 1514-1518.
17. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE, 2000. The protein data bank. *Nucleic Acids Res.*, **28**, 235-242.
18. Berntsson RP, Smits SH, Schmitt L, Slotboom DJ and Poolman B, 2010. A structural classification of substrate-binding proteins. *FEBS Lett.*, **584**, 2606-2617.
19. Berry AM and Paton JC, 1996. Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of *Streptococcus pneumoniae*. *Infect. Immun.*, **64**, 5255-5262.
20. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L and Schwede T, 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*, **42**, 252-258.
21. Biemans-Oldehinkel E, Doeven MK and Poolman B, 2006. ABC transporter architecture and regulatory roles of accessory domains. *FEBS Lett.*, **580**, 1023-1035.
22. Binkowski TA, Naghibzadeh S and Liang J, 2003. CASTp: computed atlas of surface topography of proteins. *Nucleic Acids Res.*, **31**, 3352-3355.

23. Borths EL, Locher KP, Lee AT and Rees DC, 2002. The structure of *Escherichia coli* BtuF and binding to its cognate ATP binding cassette transporter. *Proc. Natl. Acad. Sci.*, **99**, 16642-16647.
24. Boyer E, Bergevin I, Malo D, Gros P and Cellier MFM, 2002. Acquisition of Mn (II) in addition to Fe (II) is required for full virulence of *Salmonella enterica* serovar Typhimurium. *Infect. Immun.*, **70**, 6032-6042.
25. Braibant M, Gilot P and Content J, 2000. The ATP binding cassette (ABC) transport systems of *Mycobacterium tuberculosis*. *FEMS Microbiol. Rev.*, **24**, 449-467.
26. Braun V and Herrmann C, 2007. Docking of the periplasmic FecB binding protein to the FecCD transmembrane proteins in the ferric citrate transport system of *Escherichia coli*. *J. Bacteriol.*, **189**, 6913-6918.
27. Braun V, Hantke K and Koester W, 1998. Bacterial iron transport: mechanisms, genetics, and regulation. *Metal ions in biological systems*, **35**, 67-146.
28. Brown JS, Gilliland SM and Holden DW, 2001. A *Streptococcus pneumoniae* pathogenicity island encoding an ABC transporter involved in iron uptake and virulence. *Mol. Microbiol.*, **40**, 572-585.
29. Brunger AT, 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472-475.
30. Bublitz M, Morth JP and Nissen P, 2011. P-type ATPases at a glance. *J. Cell Sci.*, **124**, 2515-2519.
31. Cabibbe AM, Walker TM, Niemann S. and Cirillo DM, 2018. Whole genome sequencing of *Mycobacterium tuberculosis*. *Eur. Respir. J.*, **52**.
32. Chai Q, Zhang Y and Liu CH, 2018. *Mycobacterium tuberculosis*: an adaptable pathogen associated with multiple human diseases. *Front. Cell. Infect. Microbiol.*, **8**, 158.
33. Chandravanshi M, Gogoi P and Kanaujia SP, 2020. Structural and thermodynamic correlation illuminates the selective transport mechanism of disaccharide α-glycosides through ABC transporter. *FEBS J.*, **287**, 1576-1597.
34. Chandravanshi M. Sharma A, Dasgupta P, Mandal SK and Kanaujia SP, 2019. Identification and characterization of ABC transporters for carbohydrate uptake in *Thermus thermophilus* HB8. *Gene*, **696**, 135-148.
35. Chao A, Sieminski PJ, Owens CP and Goulding CW, 2018. Iron acquisition in *Mycobacterium tuberculosis*. *Chem. Rev.*, **119**, 1193-1220.

36. Chen C, Malek AA, Wargo MJ, Hogan DA and Beattie GA, 2010. The ATP-binding cassette transporter Cbc (choline/betaine/carnitine) recruits multiple substrate-binding proteins with strong specificity for distinct quaternary ammonium compounds. *Mol. Microbiol.*, **75**, 29-45.
37. Chen D, Menche G, Power TD, Sower L, Peterson JW and Schein CH, 2007. Accounting for ligand-bound metal ions in docking small molecules on adenylyl cyclase toxins. *Proteins: Struct. Funct. Bioinf.*, **67**, 593-605.
38. Chen J, 2013. Molecular mechanism of the *Escherichia coli* maltose transporter. *Curr. Opin. Struct. Biol.*, **23**, 492-498.
39. Chen J, Lu G, Lin J, Davidson AL and Quiocho FA, 2003. A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Mol. Cell.*, **12**, 651-661.
40. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS and Richardson DC, 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12-21.
41. Cole S, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry C, and Tekaia F, 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, **393**, 537-544.
42. Corbin BD, Seeley EH, Raab A, Feldmann J, Miller MR, Torres VJ, Anderson KL, Dattilo BM, Dunman PM, Gerads R and Caprioli RM, 2008. Metal chelation and inhibition of bacterial growth in tissue abscesses. *Science*, **319**, 962-965.
43. Cornelis P and Dingemans J, 2013. *Pseudomonas aeruginosa* adapts its iron uptake strategies in function of the type of infections. *Front. Cell. Infect. Microbiol.*, **3**, 75.
44. Cornelissen CN and Sparling PF, 1994. Iron piracy: acquisition of transferrin-bound iron by bacterial pathogens. *Mol. Microbiol.*, **14**, 843-850.
45. Cotton FA, 2000. A millennial overview of transition metal chemistry. *J. Chem. Soc., Dalton Trans.*, **13**, 1961-1968.
46. Counago RM, McDevitt CA, Ween MP and Kobe B, 2012. Prokaryotic substrate-binding proteins as targets for antimicrobial therapies. *Curr. Drug Targets*, **13**, 1400-1410.

47. Counago RM, Ween MP, Begg SL, Bajaj M, Zuegg J, O'mara ML, Cooper MA, McEwan AG, Paton JC, Kobe B and McDevitt CA, 2014. Imperfect coordination chemistry facilitates metal ion release in the Psa permease. *Nat. Chem. Biol.*, **10**, 35-41.
48. Crichton RR and Boelaert JR, 2001. Inorganic biochemistry of iron metabolism: from molecular mechanisms to clinical consequences. John Wiley & Sons, Hoboken, New Jersey.
49. Davidson AL, Dassa E, Orelle C and Chen J, 2008. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev.*, **72**, 317-364.
50. Davidson AL, Shuman HA and Nikaido H, 1992. Mechanism of maltose transport in *Escherichia coli*: transmembrane signalling by periplasmic binding proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 2360-2364.
51. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall III WB, Snoeyink J, Richardson JS and Richardson DC, 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375-W383.
52. Dawson RJP and Locher KP, 2006. Structure of a bacterial multidrug ABC transporter. *Nature*, **443**, 180-185.
53. Deller MC, Kong L and Rupp B, 2016. Protein stability: a crystallographer's perspective. *Acta Crystallogr. F: Struct. Biol. Commun.*, **72**, 72-95.
54. Desrosiers DC, Sun YC, Zaidi AA, Eggers CH, Cox DL and Radolf JD, 2007. The general transition metal (Tro) and Zn^{2+} (Znu) transporters in *Treponema pallidum*: analysis of metal specificities and expression profiles. *Mol. Microbiol.*, **65**, 137-152.
55. Diaz-Ochoa V, Jellbauer S, Klaus S and Raffatellu M, 2014. Transition metal ions at the crossroads of mucosal immunity and microbial pathogenesis. *Front. Cell. Infect. Microbiol.*, **4**, 2.
56. Eakanunkul S, Lukat-Rodgers GS, Sumithran S, Ghosh A, Rodgers KR, Dawson JH and Wilks, A, 2005. Characterization of the periplasmic heme-binding protein shut from the heme uptake system of *Shigella dysenteriae*. *Biochemistry*, **44**, 13179-13191.
57. Egawa T, Proshlyakov DA, Miki H, Makino R, Ogura T, Kitagawa T and Ishimura Y, 2001. Effects of a thiolate axial ligand on the $\pi \rightarrow \pi^*$ electronic states

of oxoferryl porphyrins: A study of the optical and resonance Raman spectra of compounds I and II of chloroperoxidase. *J. Biol. Inorg. Chem.*, **6**, 46-54.

58. Eitinger T, Rodionov DA, Grote M and Schneider E, 2011. Canonical and ECF-type ATP-binding cassette importers in prokaryotes: diversity in modular organization and cellular functions. *FEMS Microbiol. Rev.*, **35**, 3-67.
59. Elbourne JD, Tetu SG, Hassan KA and Paulsen IT, 2017. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res.*, **45**, D320-D324.
60. Emsley P and Cowtan K, 2004. Coot: model-building tools for molecular graphics. *Acta. Crystallogr. D Biol. Crystallogr*., **60**, 2126-2132.
61. Emsley P, Lohkamp B, Scott WG and Cowtan K, 2010. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr*., **66**, 486-501.
62. Evans PR and Murshudov GN, 2013. How good are my data and what is the resolution?. *Acta Crystallogr. D: Biol. Crystallogr.*, **69**, 1204-1214.
63. Evans PR, 2011. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D: Biol. Crystallogr.*, **67**, 282-292.
64. Fang H, Kang J and Zhang D, 2017. Microbial production of vitamin B12: a review and future perspectives. *Microb. Cell Fact.*, **16**, 15.
65. Farhana A, Kumar S, Rathore SS, Ghosh PC, Ehtesham NZ, Tyagi AK and Hasnain SE, 2008. Mechanistic insights into a novel exporter-importer system of *Mycobacterium tuberculosis* unravel its role in trafficking of iron. *PLoS One*, **3**, 2087.
66. Ferguson AD and Deisenhofer J, 2004. Metal import through microbial membranes. *Cell*, **116**, 15-24.
67. Fetherston JD, Bertolino VJ and Perry RD, 1999. YbtP and YbtQ: two ABC transporters required for iron uptake in *Yersinia pestis*. *Mol. Microbiol.*, **32**, 289-299.
68. Finkelstein RA, Sciortino CV and McIntosh MA, 1983. Role of iron in microbe-host interactions. *Rev. Infect. Dis.*, **5**, S759-S777.
69. Fischer K, Llamas A, Tejada-Jimenez M, Schrader N, Kuper J, Ataya FS, Galvan A, Mendel RR, Fernandez E and Schwarz G, 2006. Function and structure of the molybdenum cofactor carrier protein from *Chlamydomonas reinhardtii*. *J. Biol. Chem.*, **281**, 30186–30194.

70. Forrellad MA, Klepp LI, Gioffré A, Sabio y Garcia J, Morbidoni HR, Santangelo MDLP, Cataldi AA and Bigi F, 2013. Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence*, **4**, 3-66.
71. Fujino Y, Nagayoshi Y, Iwase M, Yokoyama T, Ohshima T and Doi K, 2016. Silica-Induced Protein (Sip) in thermophilic bacterium *Thermus thermophilus* responds to low iron availability. *Appl. Environ. Microbiol.*, **82**, 3198-3207.
72. Fukami-Kobayashi K, Tateno Y and Nishikawa K, 1999. Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *J. Mol. Biol.*, **286**, 279-290.
73. Garcia-Vallvé S, Palau J and Romeu A, 1999. Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in *Escherichia coli* and *Bacillus subtilis*. *Mol. Biol. Evol.*, **16**, 1125-1134.
74. Garmory HS and Titball RW, 2004. ATP-binding cassette transporters are targets for the development of antibacterial vaccines and therapies. *Infect. Immun.*, **72**, 6757-6763.
75. Gasteiger J and Marsili M, 1978. A new model for calculating atomic charges in molecules. *Tetrahedron Lett.*, **19**, 3181-3184.
76. Gasteiger J and Marsili M, 1980. Iterative partial equalization of orbital electronegativity-a rapid access to atomic charges. *Tetrahedron*, **36**, 3219–3228.
77. Gerber S, Comellas-Bigler M, Goetz BA and Locher KP, 2008. Structural basis of trans-inhibition in a molybdate/tungstate ABC transporter. *Science*, **321**, 246-250.
78. Giuffre A, Stubauer G, Sarti P, Brunori M, Zumft WG, Buse G and Soulimane T, 1999. The heme-copper oxidases of *Thermus thermophilus* catalyze the reduction of nitric oxide: Evolutionary implications. *Proc. Natl. Acad. Sci.*, **96**, 14718-14723.
79. Gonin S, Arnoux P, Pierru B, Lavergne J, Alonso B, Sabaty M and Pignol D, 2007. Crystal structures of an Extracytoplasmic Solute Receptor from a TRAP transporter in its open and closed forms reveal a helix-swapped dimer requiring a cation for alpha-keto acid binding. *BMC Struct. Biol.*, **7**, 11.
80. Gopinath K, Venclovas C, Ioerger TR, Sacchettini JC, McKinney, JD, Mizrahi V and Warner DF, 2013. A vitamin B12 transporter in *Mycobacterium tuberculosis*. *Open Biol.*, **3**, 120175.

81. Gouet P, Robert X and Courcelle E, 2003. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic acids Res.*, **31**, 3320-3323.
82. Gutmann DA, Ward A, Urbatsch IL, Chang G and van Veen HW, 2010. Understanding polyspecificity of multidrug ABC transporters: closing in on the gaps in ABCB1. *Trends Biochem. Sci.*, **35**, 36-42.
83. Haase H and Rink L, 2007. Signal transduction in monocytes: the role of zinc ions. *Biometals*, **20**, 579-585.
84. Hammerschmidt S, 2009. Surface-exposed adherence molecules of *Streptococcus pneumoniae*. *Host-Pathogen Interactions: Methods and Protocols*, 29-45.
85. Harding MM, 2006. Small revisions to predicted distances around metal sites in proteins. *Acta Crystallogr. D: Biol. Crystallogr.*, **62**, 678-682.
86. Harland DN, Garmory HS, Brown KA and Titball RW, 2005. An association between ATP binding cassette systems, genome sizes and lifestyles of bacteria. *Res. Microbiol.*, **156**, 434–442.
87. Hayward S and Berendsen HJ, 1998. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins*, **30**, 144-154.
88. Henne A, Brüggemann H, Raasch C, Wiezer A, Hartsch T, Liesegang H, Johann A, Lienard T, Gohl O, Martinez-Arias R and Jacobi C, 2004. The genome sequence of the extreme thermophile *Thermus thermophilus*. *Nature Biotechnol.*, **22**, 547-553.
89. Hennigar SR and McClung JP, 2016. Nutritional immunity: starving pathogens of trace minerals. *Am. J. Lifestyle Med.*, **10**, 170-173.
90. Hentze MW, Muckenthaler MU and Andrews NC, 2004. Balancing acts: molecular control of mammalian iron metabolism. *Cell*, **117**, 285-297.
91. Higgins CF, 1992. ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol.*, **8**, 67-113.
92. Ho WW, Li H, Eakanunkul S, Tong Y, Wilks A, Guo M and Poulos TL, 2007. Holo-and apo-bound structures of bacterial periplasmic heme-binding proteins. *J. Biol. Chem.*, **282**, 35796-35802.
93. Hollenstein K, Frei D C and Locher KP, 2007. Structure of an ABC transporter in complex with its binding protein. *Nature*, **446**, 213-216.

94. Holm L and Rosenstrom P, 2010. Dali server: conservation mapping in 3D. *Nucleic acids Res.*, **38**, 545-549.
95. Holm RH, Kennepohl P and Solomon EI, 1996. Structural and functional aspects of metal sites in biology. *Chem. Rev.*, **96**, 2239-2314.
96. Hood MI and Skaar EP, 2012. Nutritional immunity: transition metals at the pathogen–host interface. *Nat. Rev. Microbiol.*, **10**, 525-537.
97. Hu Y, Rech S, Gunsalus RP and Rees DC, 1997. Crystal structure of the molybdate binding protein ModA. *Nat. Struct. Mol. Biol*., **4**, 703-707.
98. Huang W and Wilks A, 2017. Extracellular heme uptake and the challenge of bacterial cell membranes. *Annu. Rev. Biochem.*, **86**, 799-823.
99. Hung LW, Wang IX, Nikaido K, Liu PQ, Ames GF and Kim SH, 1998. Crystal structure of the ATP-binding subunit of an ABC transporter. *Nature*, **396**, 703-707.
100. Hussain AS, Shanthi V, Sheik SS, Jeyakanthan J, Selvarani P and Sekar K, 2002. PDB Goodies–a web-based GUI to manipulate the Protein Data Bank file. *Acta Crystallogr. D Biol. Crystallogr*., **58**, 1385-6.
101. Ilari A, Alaleona F, Petrarca P, Battistoni A and Chiancone E, 2011. The X-ray structure of the zinc transporter ZnuA from *Salmonella enterica* discloses a unique triad of zinc-coordinating histidines. *J. Mol. Biol.*, **409**, 630-641.
102. Janulczyk R, Ricci S and Björck L, 2003. MtsABC is important for manganese and iron transport, oxidative stress resistance, and virulence of *Streptococcus pyogenes*. *Infect. Immun.*, **71**, 2656-2664.
103. Jokubovics NS and Jenikson HF, 2001. Out of the iron age: new insights into the critical role of manganese homeostasis in bacteria. *Microbiology*, **147**, 1709-1718.
104. Jones CM and Niederweis M, 2011. *Mycobacterium tuberculosis* can utilize heme as an iron source. *J. Bacteriol.*, **193**, 1767-1770.
105. Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, Brookes DH, Wilson L, Chen J, Liles K, Chun M, Li P, Gohara DW, Dolinsky T, Konecny R, Koes DR, Nielsen JE, Head-Gordon T, Geng W, Krasny R, Wei GW, Holst MJ, McCammon JA, Baker NA, 2018. Improvements to the APBS biomolecular solvation software suite. *Protein Science*, **27**, 112-128.

106. Kallberg M, Wang H, Wang S, Peng J, Wang Z, Lu H and Xu J, 2012. Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.*, **7**, 1511-1522.
107. Kanehisa M and Goto S, 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27-30.
108. Kapopoulou A. Jones LM and Cole ST, 2011. The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis*, **91**, 8-13.
109. Karpowich NK, Huang HH, Smith PC and Hunt JF, 2003. Crystal structures of the BtuF periplasmic-binding protein for vitamin B12 suggest a functionally important reduction in protein mobility upon ligand binding. *J.Biol. Chem.*, **278**, 8429-8434.
110. Kehl-Fie TE and Skaar EP, 2010. Nutritional immunity beyond iron: a role for manganese and zinc. *Curr. Opin. Chem. Biol.*, **14**, 218-224.
111. Kehres DG and Maguire ME, 2003. Emerging themes in manganese transport, biochemistry and pathogenesis in bacteria. *FEMS Microbiol. Rev.*, **27**, 263-290.
112. Kelley LA, Mezulis S, Yates CM, Wass MN and Sternberg MJ, 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845-858.
113. Kim JW, Kim S, Kim S, Lee H, Lee JO and Jin MS, 2017. Structural insights into the elevator-like mechanism of the sodium/citrate symporter CitS. *Scientific Rep.*, **7**, 1-10.
114. Kitten T, Munro CL, Michalek SM and Macrina FL, 2000. Genetic characterization of a *Streptococcus mutans* LraI family operon and role in virulence. *Infect. Immun.*, **68**, 4441-4451.
115. Klein JS and Lewinson O, 2011. Bacterial ATP-driven transporters of transition metals: physiological roles, mechanisms of action, and roles in bacterial virulence. *Metallomics*, **3**, 1098-1108.
116. Knape MJ, Ballez M, Burghardt NC, Zimmermann B, Bertinetti D, Kornev AP and Herberg FW, 2017. Divalent metal ions control activity and inhibition of protein kinases. *Metallomics*, **9**, 1576-1584.
117. Ko J, Park H, Heo L and Seok C, 2012. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res.*, **40**, 294-297.

118. Korkhov VM, Mireku SA and Locher KP, 2012. Structure of AMP-PNP-bound vitamin B 12 transporter BtuCD–F. *Nature*, **490**, 367-372.
119. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D and Vajda S, 2017. The ClusPro web server for protein–protein docking. *Nat. Protoc.*, **12**, 255.
120. Krewulak KD and Vogel HJ, 2008. Structural biology of bacterial iron uptake. *Biochim. Biophys. Acta Biomembr.*, **1778**, 1781-1804.
121. Krom BP, Warner JB, Konings WN and Lolkema JS, 2000. Complementary metal ion specificity of the metal-citrate transporters CitM and CitH of *Bacillus subtilis*. *J. Bacteriol.*, **182**, 6374-6381.
122. Kuhlbrandt W, 2004. Biology, structure and mechanism of P-type ATPases. *Nat. Rev. Mol. Cell Biol.*, **5**, 282-295.
123. Kumar S, Stecher G and Tamura K, 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.*, **33**, 1870-1874.
124. Kurthkoti K, Amin H, Marakalala MJ, Ghanny S, Subbian S, Sakatos A, Livny J, Fortune SM, Berney M and Rodriguez GM, 2017. The capacity of *Mycobacterium tuberculosis* to survive iron starvation might enable it to persist in iron-deprived microenvironments of human granulomas. *MBio*, **8**, e01092-17.
125. La Rosa V, Poce G, Canseco JO, Buroni S, Pasca MR, Biava M, Raju RM, Porretta GC, Alfonso S, Battilocchio C and Javid B, 2012. MmpL3 is the cellular target of the antitubercular pyrrole derivative BM212. *Antimicrob. Agents Chemother.*, **56**, 324-331.
126. Lagartera L, Gonzalez A, Hermoso JA, Saiz JL, Garcia P, Garcia JL and Menendez M, 2005. Pneumococcal phosphorylcholine esterase, Pce, contains a metal binuclear center that is essential for substrate binding and catalysis. *Protein Sci.*, **14**, 3013-3024.
127. Laires MJ, Monteiro CP and Bicho M, 2004. Role of cellular magnesium in health and human disease. *Front. Biosci.*, **9**, 76.
128. Laskowski RA, 2001. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res.*, **29**, 221-222.
129. Laskowski RA, MacArthur MW, Moss DS and Thornton JM, 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283-291.

130. Lawson DM, Williams CE, Mitchenall LA and Pau RN, 1998. Ligand size is a major determinant of specificity in periplasmic oxyanion-binding proteins: the 1.2 Å resolution crystal structure of *Azotobacter vinelandii* ModA. *Structure*, **6**, 1529-1539.
131. Lee YH, Deka RK, Norgard MV, Radolf JD and Hasemann CA, 1999. *Treponema pallidum* TroA is a periplasmic zinc-binding protein with a helical backbone. *Nat. Struct. Biol.*, **6**, 628-633.
132. Lee YH, Dorwart MR, Hazlett KR, Deka RK, Norgard MV, Radolf JD and Hasemann CA, 2002. The crystal structure of Zn (II)-free *Treponema pallidum* TroA, a periplasmic metal-binding protein, reveals a closed conformation. *J. Bacteriol.*, **184**, 2300-2304.
133. Lefevre F and Boutry M, 2018. Towards identification of the substrates of ATP-binding cassette transporters. *Plant physiol*., **178**, 18-39.
134. Lensbouer JJ, Patel A, Sirianni JP and Doyle RP, 2008. Functional characterization and metal ion specificity of the metal-citrate complex transporter from *Streptomyces coelicolor*. *J. Bacteriol.*, **190**, 5616-5623.
135. Lew JM, Kapopoulou A, Jones LM and Cole ST, 2011. TubercuList—10 years after. *Tuberculosis*, **91**, 1-7.
136. Lewinson O and Livnat-Levanon N, 2017. Mechanism of action of ABC importers: conservation, divergence, and physiological adaptations. *J. Mol. Biol.*, **429**, 606-619.
137. Lewis DA, Klesney-Tait J, Lumbley SR, Ward CK, Latimer JL, Ison CA and Hansen EJ, 1999. Identification of the znuA-encoded periplasmic zinc transport protein of *Haemophilus ducreyi*. *Infect. Immun.*, **67**, 5060-5068.
138. Lewis VG, Ween MP and McDevitt CA, 2012. The role of ATP-binding cassette transporters in bacterial pathogenicity. *Protoplasma*, **249**, 919-942.
139. Li H and Jogl G, 2007. Crystal structure of the zinc-binding transport protein ZnuA from *Escherichia coli* reveals an unexpected variation in metal coordination. *J. Mol. Biol.*, **368**, 1358-1366.
140. Li W, Sharma M and Kaur P, 2014. The DrrAB efflux system of *Streptomyces peucetius* is a multidrug transporter of broad substrate specificity. *J. Biol. Chem.*, **289**, 12633-12646.

141. Locher KP, Lee AT and Rees DC, 2002. The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science*, **296**, 1091-1098.
142. Lovell SC, Davis IW, Arendall WB, de Bakker PI, Word JM, Prisant MG, Richardson JS and Richardson DC, 2003. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Struct. Funct. Bioinf.*, **50**, 437-450.
143. Lucana DOO, Fedosov SN, Wedderhoff I, Che EN and Torda AE, 2014. The extracellular heme-binding protein HbpS from the soil bacterium *Streptomyces reticuli* is an aquo-cobalamin binder. *J. Biol. Chem*., 289, 34214-34228.
144. Lucana DOO, Hickey N, Hensel M, Klare JP, Geremia S, Tiufiakova T and Torda AE, 2016. The crystal structure of the C-terminal domain of the *Salmonella enterica* PduO protein: an old fold with a new heme-binding mode. *Front. Microbiol*., **7**, 1010.
145. Ma Z, Jacobsen FE and Giedroc DP, 2009. Coordination chemistry of bacterial metal transport and sensing. *Chem. Rev.*, **109**, 4644-4681.
146. Mandal SK, Adhikari R, Sharma A, Chandravanshi M, Gogoi P and Kanaujia SP, 2019. Designating ligand specificities to metal uptake ABC transporters in *Thermus thermophilus* HB8. *Metallomics*, **11**, 597-612.
147. Mandal SK, Chandravanshi M, Gogoi P and Kanaujia SP, 2017. *In silico* characterization of TTHA0596: A potential Zn^{2+} binding protein of ATP-binding cassette transporter. *Gene Rep.*, **6**, 132-141.
148. Mao B, Pear M, McCammon J and Quiocho F, 1982. Hinge-bending in L-arabinose-binding protein. The "Venus's-flytrap" model. *J. Biol. Chem*., **257**, 1131-1133.
149. Marinelli F, Kuhlmann SI, Grell E, Kunte HJ, Ziegler C and Faraldo-Gómez JD, 2011. Evidence for an allosteric mechanism of substrate release from membrane-transporter accessory binding proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, E1285-E1292.
150. Martens JH, Barg H, Warren MJ and Jahn D, 2002. Microbial production of vitamin B_{12}. *Appl. Microbiol. Biotechnol.*, **58**, 275–285.
151. Mattle D, Zeltina A, Woo JS, Goetz BA and Locher KP, 2010. Two stacked heme molecules in the binding pocket of the periplasmic heme-binding protein HmuT from *Yersinia pestis*. *J. Mol. Boil*., **404**, 220-231.

152. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ, 2007. Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658-674.
153. McDevitt CA, Ogunniyi AD, Valkov E, Lawrence MC, Kobe B, McEwan AG and Paton JC, 2011. A molecular mechanism for bacterial susceptibility to zinc. *PLoS Pathog.*, **7**, e1002357.
154. Miethke M and Marahiel MA, 2007. Siderophore-based iron acquisition and pathogen control. *Microbiol. Mol. Biol. Rev.*, **71**, 413-451.
155. Mitra A, Ko YH, Cingolani G and Niederweis M, 2019. Heme and hemoglobin utilization by *Mycobacterium tuberculosis*. *Nat. Commun.*, **10**, 1-14.
156. Moore CM and Helmann JD, 2005. Metal ion homeostasis in *Bacillus subtilis*. *Curr. Opin. Microbiol.*, **8**, 188-195.
157. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS and Olson AJ, 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.*, **30**, 2785-2791.
158. Mota CS, Valette O, Gonza´lez PJ, Brondino CD, Moura JJG, Moura I, Dolla A and Rivas MG, 2011. Effects of molybdate and tungstate on expression levels and biochemical characteristics of formate dehydrogenases produced by *Desulfovibrio alaskensis* NCIMB 13491. *J. Bacteriol.*, **193**, 2917-2923.
159. Mulligan C, Geertsma E, Severi E, Kelly D, Poolman B and Thomas G, 2009. The substrate-binding protein imposes directionality on an electrochemical sodium gradient-driven TRAP transporter. *Proc. Natl. Acad.Sci.* USA, **106**, 1778-1783.
160. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F and Vagin AA, 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D: Biol. Crystallogr.*, **67**, 355-367.
161. Nachin L, El Hassouni M, Loiseau L, Expert D and Barras F, 2001. SoxR-dependent response to oxidative stress and virulence of *Erwinia chrysanthemi*: the key role of SufC, an orphan ABC ATPase. *Mol. Microbiol.*, **39**, 960-972.
162. Naoe Y, Nakamura N, Rahman M, Tosha T, Nagatoishi S, Tsumoto K, Shiro Y and Sugimoto H, 2017. Structural basis for binding and transfer of heme in bacterial heme-acquisition systems. *Proteins: Struct. Funct. Bioinf.*, **85**, 2217-2230.

163. Neyrolles O, Wolschendorf F, Mitra A and Niederweis M, 2015. Mycobacteria, metals, and the macrophage. *Immunol. Rev.*, **264**, 249-263.
164. Nguyen L, 2016. Antibiotic resistance mechanisms in *M. tuberculosis*: an update. *Arch. Toxicol.*, **90**, 1585-1604.
165. Niederweis M, 2008. Nutrient acquisition by mycobacteria. *Microbiology*, **154**, 679-692.
166. Ohtani N, Tomita M and Itaya M, 2012. The third plasmid pVV8 from *Thermus thermophilus* HB8: isolation, characterization, and sequence determination. *Extremophiles*, **16**, 237-244.
167. Oldham ML and Chen J, 2011. Crystal structure of the maltose transporter in a pre-translocation intermediate state. *Science,* **332**, 1202-1205.
168. Oldham ML, Khare D, Quiocho FA, Davidson AL and Chen J, 2007. Crystal structure of a catalytic intermediate of the maltose transporter. *Nature*, **450**, 515-521.
169. Oswald C, Holland IB and Schmitt L, 2006. The motor domains of ABC-transporters. *Naunyn-Schmiedeberg's Arch. Pharmacol.*, **372**, 385-399.
170. Otrelo-Cardoso AR, Nair RR, Correia MA, Cordeiro RSC, Panjkovich A, Svergun DI, Santos-Silva T and Rivas MG, 2017. Highly selective tungstate transporter protein TupA from *Desulfovibrio alaskensis* G20. *Sci. Rep.*, **7**, 5798.
171. Pacheco SA, Hsu FF, Powers KM and Purdy GE, 2013. MmpL11 protein transports mycolic acid-containing lipids to the mycobacterial cell wall and contributes to biofilm formation in *Mycobacterium smegmatis*. *J. Biol. Chem.*, **288**, 24213-24222.\
172. Paik S, Brown A, Munro CL, Cornelissen CN and Kitten T, 2003. The sloABCR operon of *Streptococcus mutans* encodes an Mn and Fe transport system required for endocarditis virulence and its Mn-dependent repressor. *J. Bacteriol.*, **185**, 5967-5975.
173. Pandey R and Rodriguez GM, 2012. A ferritin mutant of *Mycobacterium tuberculosis* is highly susceptible to killing by antibiotics and is unable to establish a chronic infection in mice. *Infect. Immun.*, **80**, 3650-3659.
174. Pandey R and Rodriguez GM, 2014. IdeR is required for iron homeostasis and virulence in *Mycobacterium tuberculosis*. *Mol. Microbiol.*, **91**, 98-109.

175. Pandey S, Modak A, Phale PS and Bhaumik P, 2016. High resolution structures of periplasmic glucose binding protein of *Pseudomonas putida* CSV86 reveal structural basis of its substrate specificity. *J. Biol. Chem.*, **291**, 7844-7857.
176. Parcej D and Tampé R, 2007. Caught in the act: an ABC transporter on the move. *Structure*, **15**, 1028-1030.
177. Park JT, Raychaudhuri D, Li H, Normark S and Mengin-Lecreulx D, 1998. MppA, a periplasmic binding protein essential for import of the bacterial cell wall peptidel-alanyl-γ-d-glutamyl-meso-diaminopimelate. *J. Bacteriol.*, **180**, 1215-1223.
178. Paul S, Diekema D and Moye-Rowley WS, 2013. Contributions of *Aspergillus fumigatus* ATP-binding cassette transporter proteins to drug resistance and virulence. *Eukaryot. Cell*, **12**, 1619-1628.
179. Pederick VG, Eijkelkamp BA, Ween MP, Begg SL, Paton JC and McDevitt CA, 2014. Acquisition and role of molybdate in *Pseudomonas aeruginosa*. *Appl. Environ. Microbiol.*, **80**, 6843-6852.
180. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE, 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605-1612.
181. Porcheron G, Garénaux A, Proulx J, Sabri M and Dozois CM, 2013. Iron, copper, zinc, and manganese transport and regulation in pathogenic Enterobacteria: correlations between strains, site of infection and the relative importance of the different metal transport systems for virulence. *Front. Cell. Infect. Microbiol.*, **3**, 90.
182. Procko E, Ferrin-O'Connell I, Ng SL and Gaudet R, 2006. Distinct structural and functional properties of the ATPase sites in an asymmetric ABC transporter. *Mol. cell*, **24**, 51-62.
183. Ranawat P and Rawat S, 2018. Metal-tolerant thermophiles: metals as electron donors and acceptors, toxicity, tolerance and industrial applications. *Environ. Sci. Pollut. Res.*, **25**, 4105-4133.
184. Reddy TBK, Riley R, Wymore F, Montgomery P, DeCaprio D, Engels R, Gellesch M, Hubble J, Jen D, Jin H, Koehrsen M, Larson L, Mao M, Nitzberg M, Sisk P, Stolte C, Weiner B, White J, Zachariah ZK, Sherlock G, Galagan JE, Ball CA and Schoolnik GK, 2009. TB database: an integrated platform for tuberculosis research. *Nucleic Acids Res.*, **37**, 499-508.

185. Rees DC, Johnson E and Lewinson O, 2009. ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.*, **10**, 218-227.
186. Rempel S, Gati C, Nijland M, Thangaratnarajah C, Karyolaimos A, de Gier JW, Guskov A and Slotboom DJ, 2020. A mycobacterial ABC transporter mediates the uptake of hydrophilic compounds. *Nature*, **580**, 409-412.
187. Riccardi G, Milano A, Pasca MR and Nies DH, 2008. Genomic analysis of zinc homeostasis in Mycobacterium tuberculosis. *FEMS Microbiol. Lett.*, **287**, 1-7.
188. Rice AJ, Park A and Pinkett HW, 2014. Diversity in ABC transporters: type I, II and III importers. *Crit. Rev. Biochem. Mol. Biol.*, **49**, 426-437.
189. Rodriguez GM and Neyrolles O, 2014. Metallobiology of tuberculosis. *Microbiol. Spectr.*, **2**, MGM2-0012-2013.
190. Rodriguez GM and Smith I, 2006. Identification of an ABC transporter required for iron acquisition and virulence in *Mycobacterium tuberculosis*. *J. Bacteriol.*, **188**, 424-430.
191. Rodriguez GM, 2006. Control of iron metabolism in *Mycobacterium tuberculosis*. *Trends Microbiol.*, **14**, 320-327.
192. Romine MF, Rodionov DA, Maezato Y, Anderson LN, Nandhikonda P, Rodionova IA, Carre A, Li X, Xu C, Clauss TRW, Kim Y, Metz TO and Wright AT, 2017. Elucidation of roles for vitamin B12 in regulation of folate, ubiquinone, and methionine metabolism. *Proc. Natl. Acad. Sci.*, **114**, 1205-1214.
193. Runyen-Janecky LJ, 2013. Role and regulation of heme iron acquisition in gram-negative pathogens. *Front. Cell. Infect. Microbiol.*, **3**, 55.
194. Saier MH, 2000. A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol. Mol. Biol. Rev.*, **64**, 354-411.
195. Santhanagopalan SM and Rodriguez GM, 2012. Examining the role of Rv2895c (ViuB) in iron acquisition in *Mycobacterium tuberculosis*. *Tuberculosis*, **92**, 60-62.
196. Scheepers GH, Lycklama a Nijeholt JA and Poolman B, 2016. An updated structural classification of substrate-binding proteins. *FEBS Lett.*, **590**, 4393-4401.
197. Schmidt SB and Husted S, 2019. The biochemical properties of manganese in plants. *Plants*, **8**, 381.

198. Schmitt TH, Frezzatti WA and Schreier S, 1993. Hemin-induced lipid membrane disorder and increased permeability: a molecular model for the mechanism of cell lysis. *Arch. Biochem. Biophys*, **307**, 96-103.
199. Schneider E and Hunke S, 1998. ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains. *FEMS Microbiol. Rev.*, **22**, 1-20.
200. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T, 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498-2504.
201. Shitan N, Bazin I, Dan K, Obata K, Kigawa K, Ueda K, Sato F, Forestier C and Yazaki K, 2003. Involvement of CjMDR1, a plant multidrug-resistance-type ATP-binding cassette protein, in alkaloid transport in Coptis japonica. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 751-756.
202. Sievers F and Higgins DG, 2014. Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol. Biol.*, **1079**, 105-116.
203. Singh KV, Coque TM, Weinstock GM and Murray BE, 1998. *In vivo* testing of an *Enterococcus faecalis* efaA mutant and use of efaA homologs for species identification. *FEMS Immunol. Med. Microbiol.*, **21**, 323-331.
204. Sit B, Crowley SM, Bhullar K, Lai CCL, Tang C, Hooda Y, Calmettes C, Khambati H, Ma C, Brumell JH and Schryvers AB, 2015. Active transport of phosphorylated carbohydrates promotes intestinal colonization and transmission of a bacterial pathogen. *PLoS Pathog.*, **11**, e1005107.
205. Snow GA, 1970. Mycobactins: iron-chelating growth factors from mycobacteria. *Bacteriol. Rev.*, **34**, 99.
206. Soni DK, Dubey SK and Bhatnagar R, 2020. ATP-binding cassette (ABC) import systems of *Mycobacterium tuberculosis*: target for drug and vaccine development. *Emerg. Microbes & Infect.*, **9**, 207-220.
207. Stierand K, Maaß P and Rarey M, 2006. Molecular Complexes at a Glance: Automated Generation of two-dimensional Complex Diagrams. *Bioinformatics*, **22**, 1710-1716.
208. Sumathi K, Ananthalakshmi P, Roshan MM and Sekar K, 2006. 3dSS: 3D structural superposition. *Nucleic Acids Res.*, **34**, 128-132.

209. Sun X, Xiao C, Ge R, Yin X, Li H, Li N, Yang X, Zhu Y, He X and He QY, 2011. Putative copper-and zinc-binding motifs in *Streptococcus pneumoniae* identified by immobilized metal affinity chromatography and mass spectrometry. *Proteomics*, **11**, 3288-3298.
210. Sutcliffe IC and Harrington DJ, 2004. Lipoproteins of *Mycobacterium tuberculosis*: an abundant and functionally diverse class of cell envelope components. *FEMS Microbiol. Rev.*, **28**, 645-659.
211. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P and Jensen LJ, 2019. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.*, **47**, 607-613.
212. Szollosi D, Rose-Sperling D, Hellmich UA and Stockner T, 2018. Comparison of mechanistic transport cycle models of ABC exporters. *Biochim. Biophys. Acta Biomembr.*, **1860**, 818-832.
213. Tam R and Saier MH, 1993. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol. Rev.*, **57**, 320-346.
214. Tanaka KJ, Song S, Mason K and Pinkett HW, 2018. Selective substrate uptake: the role of ATP-binding cassette (ABC) importers in pathogenesis. *Biochim. Biophys. Acta Biomembr.*, **1860**, 868-877.
215. Tang C, Schwieters CD and Clore GM, 2007. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature*, **449**, 1078-1082.
216. Taylor GL, 2010. Introduction to phasing. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 325-38.
217. Ter Beek J, Guskov A and Slotboom DJ, 2014. Structural diversity of ABC transporters. *J. gen. physiol.*, **143**, 419-435.
218. Terasaka K, Blakeslee JJ, Titapiwatanakun B, Peer WA, Bandyopadhyay A, Makam SN, Lee OR, Richards EL, Murphy AS, Sato F and Yazaki K, 2005. PGP4, an ATP binding cassette P-glycoprotein, catalyzes auxin transport in *Arabidopsis thaliana* roots. *Plant Cell*, **17**, 2922-2939.
219. The UniProt Consortium, 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506-515.

220. Thomas GH, 2010. Homes for the orphans: utilization of multiple substrate-binding proteins by ABC transporters. *Mol. Microbiol.*, **75**, 6-9.
221. Tomii K and Kanehisa M, 1998. A comparative analysis of ABC transporters in complete microbial genomes. *Genome Res.*, **8**, 1048-1059.
222. Tullius MV, Harmston CA, Owens CP, Chim N, Morse RP, McMath LM, Iniguez A, Kimmey JM, Sawaya MR, Whitelegge JP and Horwitz MA, 2011. Discovery and characterization of a unique mycobacterial heme acquisition system. *Proc. Natl. Acad. Sci*. USA, **108**, 5051-5056.
223. Vagin AA, 2004. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D Biol. Crystallogr*., **60**, 2184-2195.
224. van der Heide T and Poolman B, 2002. ABC transporters: one, two or four extracytoplasmic substrate-binding sites? *EMBO Rep.*, **3**, 938-943.
225. Varela C, Rittmann D, Singh A, Krumbach K, Bhatt K, Eggeling L, Besra GS and Bhatt A, 2012. MmpL genes are associated with mycolic acid metabolism in mycobacteria and corynebacteria. *Chem. Biol.*, **19**, 498-506.
226. Vigonsky E, Fish I, Livnat-Levanon N, Ovcharenko E, Ben-Talb N and Lewinson O, 2015. Metal binding spectrum and model structure of the *Bacillus anthracis* virulence determinant MntA. *Metallomics*, **7**, 1407-1419.
227. Vigouroux A, Aumont-Nicaise M, Boussac A, Marty L, Lo Bello L, Legrand P, Brillet K, Schalk IJ and Moréra S, 2020. A unique ferrous iron binding mode is associated with large conformational changes for the transport protein FpvC of *Pseudomonas aeruginosa*. *FEBS J.*, **287**, 295-309.
228. Wakeman CA and Skaar EP, 2012. Metalloregulation of Gram-positive pathogen physiology. *Curr. Opin. Microbiol*., **15**, 169-174.
229. Waldron KJ and Robinson NJ, 2009. How do bacterial cells ensure that metalloproteins get the correct metal? *Nat. Rev. Microbiol*., **7**, 25-35.
230. Waldron KJ, Rutherford JC, Ford D and Robinson NJ, 2009. Metalloproteins and metal sensing. *Nature*, **460**, 823-830.
231. Wang S, Ogata M, Horita S, Ohtsuka J, Nagata K and Tanokura M, 2014. A novel mode of ferric ion coordination by the periplasmic ferric ion-binding subunit FbpA of an ABC-type iron transporter from *Thermus thermophilus* HB8. *Acta Crystallogr. D Biol. Crystallogr*., **70**, 196-202.

232. Webb AJ, Homer KA and Hosie AH, 2008. Two closely related ABC transporters in *Streptococcus* mutans are involved in disaccharide and/or oligosaccharide uptake. *J. Bacteriol.*, **190**, 168-178.
233. Wei B, Randich AM, Bhattacharyya-Pakrasi M, Pakrasi HB and Smith TJ, 2007. Possible regulatory role for the histidine-rich loop in the zinc transport protein, ZnuA. *Biochemistry*, **46**, 8734-8743.
234. Weinberg ED, 1975. Nutritional immunity: host's attempt to withhold iron from microbial invaders. *JAMA*, **231**, pp.39-41.
235. Weinberg ED, 2009. Iron availability and infection. *Biochim. Biophys. Acta Gen. Sub.*, **1790**, 600-605.
236. Werner C, Richter OH and Ludwig B, 2010. A novel heme *a* insertion factor gene cotranscribes with the *Thermus thermophilus* cytochrome *ba*3 oxidase locus. *J. Bacterial.*, **192**, 4712-4719.
237. Wilkens S, 2015. Structure and mechanism of ABC transporters. *F1000 Prime Rep.*, **7**.
238. Wilkinson AJ, 2002. ABC Proteins: From Bacteria to Man. (Holland B, Kuchler K, Cole SP and Higgins C, Eds.), Elsevier Science and Technology Books, London.
239. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB and Jain S, 2018. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.*, **27**, 293-315.
240. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A and McNicholas SJ, 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D: Biol. Crystallogr.*, **67**, 235-242.
241. Wohlert D, Grotzinger MJ, Kuhlbrandt W and Yildiz O, 2015. Mechanism of Na^+-dependent citrate transport from the structure of an asymmetrical CitS dimer. *Elife*, **4**, e09375.
242. Woo JS, Zeltina A, Goetz BA and Locher KP, 2012. X-ray structure of the *Yersinia pestis* heme transporter HmuUV. *Nat. Struct. Mol. Biol.*, **19**, 1310.
243. World Health Organization, 2020. Global tuberculosis report 2020. Geneva. ISBN 978-92-4-001313-1.

244. Xu D and Zhang Y, 2011. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys. J.*, **2011**, 101, 2525-2534.
245. Yang J, Yan R, Roy A, Xu D, Poisson J and Zhang Y, 2015. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods*, **12**, 7-8.
246. Yu Y, Zhou M, Kirsch F, Xu C, Zhang L, Wang Y, Jiang Z, Wang N, Li J, Eitinger T and Yang M, 2014. Planar substrate-binding site dictates the specificity of ECF-type nickel/cobalt transporters. *Cell Res.*, **24**, 267-277.
247. Zaitseva J, Jenewein S, Jumpertz T, Holland IB and Schmitt L, 2005. H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB. *EMBO J.*, **24**, 1901-1910.
248. Zhang C, 2014. Essential functions of iron-requiring proteins in DNA replication, repair and cell cycle control. *Protein Cell*, **5**, 750-760.
249. Zhang L, Hendrickson RC, Meikle V, Lefkowitz EJ, Ioerger TR and Niederweis M, 2020. Comprehensive analysis of iron utilization by *Mycobacterium tuberculosis*. *PLoS Pathog.*, **16**, e1008337.
250. Zhen M, Jacobsen FE and Giedroc DP, 2009. Metal transporters and metal sensors: how coordination chemistry controls bacterial metal homeostasis. *Chem. Rev.*, **109**, 4644-4681.
251. Zheng B, Zhang Q, Gao J, Han H, Li M, Zhang J, Qi J, Yan J and Gao GF, 2011. Insight into the interaction of metal ions with TroA from *Streptococcus suis*. *PLoS One*, **6**, e19510.
252. Zheng H, Cooper DR, Porebski PJ, Shabalin IG, Handing KB and Minor W, 2017. CheckMyMetal: a macromolecular metal-binding validation tool. *Acta Crystallogr. D: Struct. Biol.*, **73**, 223-233.
253. Zoroddu MA, Aaseth J, Crisponi G, Medici S, Peana M and Nurchi VM, 2019. The essential metals for humans: a brief overview. *J. Inorg. Biochem.*, **195**, 120-129.

LIST OF PUBLICATIONS, CONFERENCES AND WORKSHOPS

PUBLICATIONS FROM THESIS

1. **Mandal SK** and Kanaujia SP. Role of an orphan substrate-binding protein MhuP in transient heme transfer in *Mycobacterium tuberculosis*. ***Under Revision***.

2. **Mandal SK** and Kanaujia SP (2021). Structural and thermodynamic insights into a novel Mg^{2+}-citrate-binding protein of ABC transporter. ***Acta Crystallographica Section D: Structural Biology***, D77, https://doi.org/10.1107/S2059798321010457.

3. **Mandal SK**, Nayak SG, and Kanaujia SP (2021). Identification and characterization of metal uptake ABC transporters in *Mycobacterium tuberculosis* unveil their ligand specificity. ***International Journal of Biological macromolecules***, 185, 324-337.

4. **Mandal SK**, Adhikari R, Sharma A, Chandravanshi M, Gogoi P and Kanaujia SP (2019). Designating ligand specificities to metal uptake ABC transporters in *Thermus thermophilus* HB8. ***Metallomics***, 11, 597-612.

5. **Mandal SK**, Chandravanshi M, Gogoi P and Kanaujia SP (2017). *In silico* characterization of TTHA0596: A potential Zn^{2+} binding protein of ATP-binding cassette transporter. ***Gene Reports***, 6, 132-141.

PUBLICATIONS FROM OTHER COLLABORATIVE WORK

1. **Mandal SK** and Kanaujia SP. Identification and characterization of complete repertoire of metal ion uptake ABC transporters in an archaea. ***Under preparation.***

2. Saha S, **Mandal SK** and Kanaujia SP. Characterization of putative archaeal 5-methylcytosine RNA methyltransferases: an *in silico* approach. ***Under preparation***.

3. Kumar H, **Mandal SK**, Gogoi P and Kanaujia SP. Structural and functional role of invariant water molecules in matrix metalloproteinases: a data-mining approach. ***Journal of Biomolecular Structure and Dynamics***, 1-12.

4. Chandravanshi M, Sharma A, Dasgupta P, **Mandal SK** and Kanaujia SP (2019). Identification and characterization of ABC transporters for carbohydrate uptake in *Thermus thermophilus* HB8. ***Gene***, 696, 135-148.

5. Gogoi P, Chandravanshi M, **Mandal SK**, Srivastava A and Kanaujia SP (2015). Heterogeneous behavior of metalloproteins towards metal ion binding and selectivity: insights from molecular dynamics studies. ***Journal of Biomolecular Structure and Dynamics***, 34, 1470-1485.

LIST OF CONFERENCES AND WORKSHOPS

1. Suraj Kumar Mandal and **Shankar Prasad Kanaujia**. Deciphering the mechanism of transient heme transfer by an orphan substrate-binding protein MhuP in *Mycobacterium tuberculosis*. National Seminar on Crystallography (NSC-48). November 25-27, 2021, IIT Roorkee, Uttarakhand, India. **[Best Oral Presentation]**.
2. Suraj Kumar Mandal and **Shankar Prasad Kanaujia**. Structural and functional characterization of an Mg^{2+}-citrate-binding protein from a thermophilic bacteria. PDB 50th Anniversary Symposium in Asia. November 24, 2021, PDBj and Institute for Protein Research, Osaka University, Japan. **[Poster Presentation]**.
3. **Suraj Kumar Mandal** and Shankar Prasad Kanaujia. A transient mechanism of heme transfer by an orphan substrate-binding protein MhuP in Mycobacterium tuberculosis. Bringing Molecular Structure to Life: 50 Years of the PDB. October 20-22, 2021, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. **[Poster Presentation]**.
4. **Suraj Kumar Mandal** and Shankar Prasad Kanaujia. Understanding the structural and functional aspects of a novel Mg^{2+}-citrate-binding protein. "25th Congress and General Assembly of the International union of Crystallography (IUCr)". August 14-22, 2021, Prague, Czech Republic. **[Poster Presentation]**.

5. **Suraj Kumar Mandal** and Shankar Prasad Kanaujia. Metal binding spectrum of a putative metal-specific substrate binding protein from *Thermus thermophilus* HB8. National Seminar on Crystallography (NSC-47). June 19-22, 2019, BARC Mumbai, India. **[Poster Presentation]**.

6. **Suraj Kumar Mandal** and Shankar Prasad Kanaujia. Functional annotation of metal sequestering ABC transporters in *Thermus thermophilus* HB8. Emerging Trends in Disease Model Systems. March 25-27, 2019, NCCS Pune, India**. [Poster Presentation]**.

7. **Suraj Kumar Mandal** and Shankar Prasad Kanaujia. Structural and functional characterization of a putative metal-specific substrate binding protein from *Thermus thermophilus* HB8. National Seminar on Crystallography (NSC-46). June 27-29, 2018, NIMHANS Bangalore, India. **[Best Oral Presentation Award]**.

8. **Suraj Kumar Mandal** and Shankar Prasad Kanaujia. *In silico* characterization of a potential Zn^{2+} ABC transporter. "24th Congress and General Assembly of the International union of Crystallography (IUCr)" held at Hyderabad International Convention Centre. August 21-27, 2017, Hyderabad 500081, India. **[Poster presentation].**

9. **Suraj Kumar Mandal**. "Latest Methods in Structural Biology" course organized by "Global Initiative for Academic Networks", November 15-25, 2016 in JNU, New Delhi, India. **[Workshop]**.

10. **Suraj Kumar Mandal**. National Workshop on "Principles of Drug Designing" at Bioinformatics Infrastructure Facility (BIF), DMBB, March 5-6, 2016, Tezpur University, Tezpur, Assam, India. **[Resource Person].**

11. Prerana Gogoi, Monika Chandravanshi, **Suraj Kumar Mandal**, Ritiritao Brahma, Kanchan Chauhan and Shankar Prasad Kanaujia. Structural studies of Proteins involved in Crucial Biological Processes using X-Ray Crystallography. Research Conclave'16 organized by Students' Academic Board. March 2016, IIT Guwahati, Assam, India. **[Poster Presentation]**.

12. **Suraj Kumar Mandal**. National Conference on "Recent Developments in Molecular Biology, Structural Biology and Drug Discovery (RDMBSBDD)" organized by Dept. of Biosciences and Bioengineering, December 6-7, 2015, IIT Guwahati, Assam, India.

13. **Suraj Kumar Mandal**. Symposium cum Workshop on "Advances in Computational Biology and Computer Aided Drug Design" organized by Bioinformatics Infrastructure Facility (BIF), Dept. of Biosciences and Bioengineering. June 24-26, 2015, IIT Guwahati, Assam, India.