

Probing Structure, Function and Dynamics in Bacterial Primary and Secondary Transporter-Associated Binding Proteins

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Shantanu Shukla
December 2020

Copyright © 2020 by Shantanu Shukla
All rights reserved.

Dedication

To my family and my teachers for making me who I am today

Acknowledgements

I would like to sincerely thank many people without whom this work and my Ph.D. would not have been made possible. First and foremost, I want to extend my most sincere gratitude to my advisor Dr. Dean Myles who went above and beyond to make sure I had everything I needed to finish my research at the Spallation Neutron Source (SNS), Oak Ridge National Laboratory (ORNL). I have been fortunate to have such a wonderful mentor whose guidance and support I have cherished at every moment of this academic journey.

I would also like to extend my sincere gratitude to my committee members, Prof. Barry Bruce, Dr. Hugh O'Neill, Dr. Jerry Parks, and Dr. Matthew Cuneo, for lending their time and expertise and for their valuable suggestions to shape my Ph. D. research. Thank you all for being so helpful, kind, and supportive. I particularly remember the encouragement I have received from Prof. Bruce during my prelim examination. I would especially like to mention that Dr. O'Neill helped me a lot with getting neutron beamtime for my work, and for which I shall always remain very grateful. I will also remember the motivation I received from Dr. Parks during my prelims and thesis committee meetings. I appreciate Dr. Cuneo's suggestions and critical insights into my research.

It goes without saying that none of this would have been possible without the Genome Science and Technology (GST) program at the University of Tennessee at Knoxville (UTK) accepting me into this wonderful interdisciplinary program. I want to thank Dr. Albrecht von Arnim for being an incredibly supportive Program Director who always had time for me when I knocked on his door. Additionally, I also want to thank Teresa Yeatts for helping me navigate through all the administrative complexities at UTK.

Next, I would like to thank my collaborators, Dr. Christopher Stanley, Dr. Jose Borreguero Calvo, Dr. Laura Stingaciu, and Dr. Eugene Mamontov at the SNS, ORNL, who mentored me and helped me in designing, executing, and explaining complex neutron scattering experiments. I would also like to thank Dr. Pratul Agarwal for helping me with computational simulations and analysis.

I would like to extend a special note of thanks to Dr. Kevin Wiess and Qiu Zhang for making everything easily available for research at the CSMB facility at SNS, ORNL. Additionally, I would also like to thank my colleagues from the GST program and at SNS for their help and encouragement. I want to especially thank Dr. Khushboo Bafna for being incredibly supportive throughout my Ph.D.

Last but certainly not the least I want to thank my family, especially my parents, who despite all the struggles never stopped believing in me. I want to also thank my grandfather for being the source of inspiration to me. It was the strong trust of my family in me and my abilities that always kept me going.

Abstract

Substrate binding proteins (SBPs) are ubiquitous in all life forms and have evolved to perform diverse physiological functions, such as in membrane transport, gene regulation, neurotransmission, and quorum sensing. It is quite astounding to observe such functional diversity among the SBPs even when they are restricted by their fold space. Therefore, the SBPs are an excellent set of proteins that can reveal how proteins evolution novel function in a structurally conserved/constrained fold. This study attempts to understand the phenomenon of affinity and specificity evolution in SBPs by combining a set of biochemical, biophysical, and structural studies on the SBPs involved in translocation of substrates across the membrane using ATP-binding cassette (ABC) transporters and tripartite ATP-independent periplasmic (TRAP) transporters in gram-negative bacteria *Thermotoga maritima* and *Pseudomonas aeruginosa*, respectively. Additionally, experimental, and computational methods were used in conjunction to highlight the variation in the dynamics of these SBPs. The results from this study highlight an intricate role of dynamics in complementing the structural alterations that are required for high-affinity ligand binding. Moreover, first ever neutron structure of a SBP was determined during my study to delineate the extensive network of water in the binding cavities of the SBPs that help stabilize larger substrates by forming water-mediated hydrogen bond interactions with the bound substrates. Furthermore, structures of two SBPs from *T. maritima* were determined in both substrate-free (apo) and substrate-bound (holo) forms and subsequently used for computational molecular dynamics simulation to determine the variation in dynamics due to substrate-binding. The novel TRAP SBP identified in *P. aeruginosa* was identified as a promiscuous binder of several tricarboxylic acid cycle (TCA) cycle intermediates. A total of six SBP structures were determined using X-ray crystallography and one SBP structure was determined using neutron crystallography. Finally, experimental neutron scattering was used to experimentally characterize the picosecond to nanosecond dynamics in SBPs and highlighted differences in the translational, rotational, and internal dynamical signatures of two SBP isoforms. Overall, the findings of this study can be broadly applied in biotechnology and biosensor development by artificially engineer affinity or specificity for a particular ligand.

Table of Contents

Chapter 1 Introduction.....	1
1.1 Protein-substrate interaction: affinity vs. specificity.....	2
1.2 Primary transport through the ABC systems.....	5
1.2.1 The maltose importer system	11
1.3 TRAP systems: a secondary transport system for nutrient uptake.....	13
1.4 Alternate access model of substrate translocation through ABC and TRAP transporters	17
1.5 Substrate binding proteins: role in substrate selectivity and translocation	20
1.5.1 Structure, function, and classification of SBPs	21
1.5.2 Dynamics of SBPs.....	26
1.6 Proposed study and its significance.....	28
1.6.1 Specific aims and hypothesis	29
1.6.2 Significance of the study.....	30
List of References	31
Chapter 2 Differential substrate recognition by maltose binding proteins influenced by structure and dynamics	45
2.1 Abstract.....	47
2.2 Introduction	47
2.3 Materials and Methods.....	50
2.4 Results.....	55
2.4.1 Overall structure of tmMBPs.....	55
2.4.2 Structural comparison of tmMBPs with E. coli MBP (ecMBP) and Thermococcus litoralis MBP (tlMBP)	56

2.4.3	Conformational dynamics in tmMBPs	67
2.4.4	Interaction energy of substrates with tmMBPs and ecMBP	84
2.5	Discussion	89
List of References		93
 Chapter 3 Accessing the substrate-induced dynamical changes		
in maltose-binding protein isoforms using neutron		
scattering..... 101		
3.1	Abstract.....	102
3.2	Introduction	102
3.3	Materials and Methods.....	106
3.4	Results.....	109
3.4.1	Solution-state structure of tmMBPs	109
3.4.2	Picosecond – nanosecond dynamics of tmMBP isoforms using NSE	111
3.4.3	Determination of water-mediated tmMBP-substrate interaction dynamics in using QENS	115
3.4.4	Dynamical substates in tmMBP as revealed by computational MD	119
3.5	Discussion	122
List of References		126
 Chapter 4 Determining water-mediated substrate stabilization		
in maltose-binding proteins using neutron diffraction		
methods..... 133		
4.1	Abstract.....	134
4.2	Introduction	134

4.3	Materials and methods.....	136
4.4	Results.....	137
4.4.1	Neutron structure of tmMBP-MTT	137
4.4.2	Determination of protonation states of physiologically relevant residues.....	142
4.5	Discussion	143
List of References		149
 Chapter 5 A promiscuous TRAP transporter associated binding protein helps in translocating essential metabolic intermediates in Pseudomonas aeruginosa.....		
		155
5.1	Abstract.....	156
5.2	Introduction	156
5.3	Material and Methods	162
5.4	Results.....	166
5.4.1	Dissecting the TRAP transporter operons in P. aeruginosa	166
5.4.2	paTRAP-SBP has a broad substrate range.....	167
5.4.3	Structure of paTRAP-SBP reveal its mode of substrate binding..	178
5.4.4	Comparative structural and functional analysis of three paTRAP- SBPs from P. aeruginosa	185
5.5	Discussion	188
List of References		191
 Chapter 6 Conclusion and Future Direction.....		
		197
6.1	Conclusion	198
6.2	Future Directions.....	201
List of References		203

Vita	205
-------------	------------

List of Tables

Table 1.1: Conserved motifs in the NBD of the ABC transporter.....	8
Table 2.1: Starting X-ray coordinates for MD simulations.	53
Table 2.2: X-ray data collection and refinement statistics for tmMBPs.....	57
Table 2.3: Radius of gyration (R_g) values of tmMBPs.	73
Table 2.4: Residues of the N-terminal domain, Hinge and C-terminal domain in MBPs.	76
Table 2.5: Averaged protein-substrate interaction energy.	85
Table 2.6: List of residues making favorable contact obtained from interaction energy analysis.	87
Table 4.1: Data collection and refinement statistics for tmMBP2-MTT Neutron/X-ray joint refinement.	139
Table 5.1: Comparative analysis of protein stability in presence of different TCA cycle substrates.	169
Table 5.2: Small-angle X-ray scattering data analysis results for paTRAP-SBP.	177
Table 5.3: Data collection and refinement statistics.....	181

List of Figures

Figure 1.1: Schematics of the ABC importer and exporter systems.	6
Figure 1.2: Structure and domain organization in <i>E. coli</i> MalEFGK2 maltose ABC transporter system.	9
Figure 1.3: Schematics of the TRAP and TRAP-like transporter systems.	14
Figure 1.4: Mechanism of substrate translocation in ABC and TRAP systems.	18
Figure 1.5: Structural and functional diversity in SBPs.	23
Figure 2.1: Structural details of the three tmMBP isoforms in complex with maltotetraose and maltose.	59
Figure 2.2: Differential binding of tetra- and di-saccharides in MBPs.	61
Figure 2.3: Structure based sequence alignment of thermophilic and mesophilic MBPs.	63
Figure 2.4: Differential binding of trisaccharide maltotriose.	64
Figure 2.5: Radius of gyration (R_g) computed from MD simulations.	70
Figure 2.6: Root mean square deviations (RMSD) computed from the MD simulations.	71
Figure 2.7: Radius of gyration (R_g) computed from MD simulations.	72
Figure 2.8: SAXS of apo and substrate-bound tmMBPs.	74
Figure 2.9: Variation of hinge angle for apo MBPs.	75
Figure 2.10: Comparison of tmMBP1 and tmMBP3 conformational flexibility based on RMSF ₁₀	78
Figure 2.11: Comparison of crystallographic β -factors with computational all atom root mean square fluctuations (RMSF ₁₀) of tmMBPs.	79
Figure 2.12: Comparison of crystal B-factors with computational all atom root mean square fluctuations (RMSF ₁₀) of ecMBP.	80
Figure 2.13: All atom RMSF ₁₀ plot of tmMBP and ecMBP systems.	81
Figure 2.14: Errors associated with computed RMSF ₁₀	82
Figure 2.15: Conformational flexibility of the trehalose-bound tMBP.	83
Figure 2.16: Protein-substrate interaction energy.	86
Figure 3.1: Various scattering techniques and their time-scale and length-scale windows.	104

Figure 3.2: Small-angle neutron scattering (SANS) results of hydrogenated tmMBP2 and tmMBP3 in D ₂ O.	110
Figure 3.3: Structure factor analysis of tmMBPs.....	112
Figure 3.4: Neutron spin echo (NSE) spectral data on tmMBP isoforms.....	114
Figure 3.5: Quasi-elastic neutron scattering data on tmMBP isoforms.....	117
Figure 3.6: Root mean square fluctuations (RMSF ₁₀) of tmMBP isoforms.	120
Figure 3.7: Quasi-anharmonic top mode of tmMBP2 and tmMBP3.....	121
Figure 4.1: Neutron structure of tmMBP2-MTT complex.....	141
Figure 4.2: Neutron no-fill 2FoFc Maps of tmMBP2-MTT complex.	144
Figure 4.3: H/D exchangeable sites on tmMBP2-MTT complex.....	145
Figure 5.1: Topologies of TRAP, TAXI-TRAP, TTT transporters.....	158
Figure 5.2: C4-dicarboxylate TRAP transporter in <i>Pseudomonas aeruginosa</i>	160
Figure 5.3: Thermal denaturation data on paTRAP-SBP non-cognate ligand.	170
Figure 5.4: Thermal denaturation data from DSC for paTRAP-SBP with cognate substrates.....	172
Figure 5.5: Small-angle X-ray scattering data of paTRAP-SBP in presence and absence of ligand.....	175
Figure 5.6: Overall structure of paTRAP-SBP.....	179
Figure 5.7: Substrate binding by paTRAP-SBP.	183
Figure 5.8: Comparison of three TRAP-SBPs from <i>P. aeruginosa</i>	186

List of Abbreviations

ΔH	Enthalpy change
ΔS	Entropy change
ABC	Adenosine triphosphate-binding cassette transporter
ADP	Adenosine diphosphate
AMBER	Assisted Model Building and Energy Refinement
AMP	Adenosine monophosphate
APBS	Adaptive Poisson-Boltzman Solver
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
DLS	Dynamic Light Scattering
DSC	Differential Scanning Calorimetry
ecMBP	<i>E. coli</i> Maltose Binding Protein
E_{el}	electrostatic energy
E_{tot}	total interaction energy
E_{vdw}	van der Waal's energy
GLU	glucose
H/D	Hydrogen/Deuterium
H ⁺	Proton
H-bond	Hydrogen Bond
HIFR	High Flux Isotope Reactor
IM	Inner membrane
MAL	Maltose
MBP	Maltose Binding Protein
MD	Molecular Dynamics
MFS	Major facilitator superfamily
mM	Millimolar
ms	millisecond
MS	Mass Spectroscopy

MTR	Maltotriose
MTT	Maltotetrose
NBD	Nucleotide-binding domain
Neu5Ac	N-Acetylneuraminic acid/sialic acid
NMR	Nuclear Magnetic Resonance
NPC	Neutron Protein Crystallography
ns	nanosecond
NSE	Neutron Spin Echo
OM	Outer membrane
ORNL	Oak Ridge National Laboratory
pa	<i>Pseudomonas aeruginosa</i>
PBP	Periplasmic Binding Protein
PDB	Protein Data Bank
ps	picosecond
QAA	Quasi Anharmonic Analysis
QENS	Quasi-Elastic Neutron Scattering
R _g	Radius of Gyration
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuations
SANS	Small Angle Neutron Scattering
SAXS	Small Angle X-ray Scattering
SBP	Substrate Binding Protein
SNS	Spallation Neutron Source
TAXI	TRAP associated extracytoplasmic immunogenic
tlMBP	<i>Thermococcus litoralis</i> Maltose Binding Protein
T _m	Melting Temperature
TM	Transmembrane
TMD	Transmembrane domain
tmMBP	<i>Thermotoga maritima</i> Maltose Binding Protein
TRAP	Tripartite Adenosine Triphosphate-Independent Periplasmic Transporter
TTT	Tripartite Tricarboxylate Transporter

VPI-2	Vibrio pathogenicity island-2
VWD	van der Waal
XRD	X-ray Diffraction
μs	microsecond

Chapter 1 Introduction

1.1 Protein-substrate interaction: affinity vs. specificity

The history of life on Earth has been an incredible journey of evolution and diversification. From animals to plants to microbes, evolution touches all life forms and molds these organisms' behavior to adapt to their respective environments. Just as the diversity among the various life-forms inhabiting this planet has shaped their microcosm role, evolutionary forces at the molecular levels have also diversified cellular behavior. This is demonstrated quite dramatically in the case of cancer tumor heterogeneity, where normal cells undergo a series of molecular evolutionary events in just a few generations to become tumorigenic.

⁽¹⁾ Since proteins form the structural and functional units of the cellular machinery, they are under immense selection pressure to perform a multitude of varying cellular functions.

^(2,3) Understanding protein behavior becomes even more challenging when we contrast the paradox of having several millions of protein sequences in the sequence databases but only around 150,000 experimentally determined structures in the Protein Data Bank (PDB; as on 11/20/2020), which further falls under the ambit of only 1487 existing protein folds listed on the Structure Classification of Proteins (SCOP) database. The limited number of possible folds adopted in nature thus, significantly constrains how a protein molecule evolves to perform new functions. Often, protein function evolves to support the growth and development of the organism when the surrounding abiotic conditions change over the course of time.⁽⁴⁾ Most of the protein functions are a result of the staggering molecular interactions that they make with other macromolecules in the cell over the course of time. Modern bioinformatics and network biology methods use the molecular interaction information to discern the function of a hypothetical protein.^(5,6) Therefore, these molecular interactions form the basis of all life processes.

Macromolecular interactions can be in the form of intra- and intercellular information relay, molecular assemblies, activation and inhibition of a pathway or binding and release of substrates from a protein.⁽⁷⁻⁹⁾ An important factor that influences interaction between two molecules is their respective concentrations in the surrounding environment that controls their relative affinities.⁽⁹⁾ Proteins for example, show substrate binding affinities that range between femtomolar to millimolar range. However, the cellular environment is very heterogeneous with concentrations of specific metabolites often varying in the nanomolar to micromolar range. This heterogeneous behavior causes a spike in non-specific

interactions of the proteins with their surrounding substrates. Maintenance of specificity can become even more challenging when there are very subtle variations in structure of the interacting molecules. Protein promiscuity, on the other hand, can result in phenotypic plasticity in cells without causing any modification to their genomes. Therefore, promiscuity in proteins can significantly affect survival of cells in changing nutrient environments, allowing the cell to metabolize a broad spectrum of substrates. Given all these factors, what remains unclear is to why particular proteins show specificity for only a single substrate, whereas others bind to a range of substrate with near equal affinities? One way to quantify the relationship between affinity (ΔG) vs. specificity ($\Delta\Delta G$) is by discerning the kinetics of protein-substrate(ligand) interaction.⁽¹⁰⁾ For instance, the affinity for a protein P for ligand L can be described by the equilibrium constant of the protein-ligand complex PL, such as:

$$K_{eq} = \frac{[PL]}{[P_{free}][L_{free}]} \quad 1$$

Where, K_{eq} is the equilibrium constant of the total protein-ligand complex $[PL]$ formation divided by the product of the concentrations of free protein $[P_{free}]$ and free cognate ligand $[L_{free}]$. Similarly, K'_{eq} and $[L'_{free}]$ represent equilibrium constant of complex formation between a protein and a non-cognate ligand, and concentration of free non-cognate ligand, respectively. Moreover, the free energy of protein-ligand complex formation for the cognate ligand can be given by equation 2, such as:

$$\Delta G_{PL} = -RT \ln K_{eq} \quad 2$$

And, that for a non-cognate ligand can be given by equation 3:

$$\Delta G_{PL'} = -RT \ln K'_{eq} \quad 3$$

Where, R is the universal gas constant, and T is the temperature in degree Kelvin.

Now, specificity $\Delta\Delta G_{LL'}$ can be calculated as a difference in the affinities for cognate vs. non-cognate ligands and is given as:

$$\Delta\Delta G_{LL'} = \Delta G_{PL} - \Delta G_{PL'} \quad 4$$

Furthermore, the values of $\Delta\Delta G$ and ΔG are shaped by the vitality of the biological process and whether the said process demands high substrate specificity or not.

More generally, there are several determining factors that can influence specificity and affinity in proteins for a number of substrate molecules. These can include shape complementarity, non-bonded interactions (h-bonds, electrostatics, van der Waals, etc.), steric clashes, organization of the binding/active site, solvent accessibility and water-mediated interactions, conformational sampling in protein, and allostery. In large protein superfamilies, like the adenosine triphosphate [ATP]-binding cassette (ABC) transporters associated, and more recently tripartite ATP-independent periplasmic (TRAP) transporters associated, substrate binding protein (SBP) , it is observed that the homologous proteins show conservation in the overall structural fold, yet they differ considerably in their substrate affinity profiles. Similar results have also been observed in other protein superfamilies as well, for example in ribonucleases and kinases. These findings have strongly suggested that a mere similarity in structure is not a sufficient determinant of specificity in proteins, and coupled factors, such as dynamics of the protein, might play a more influential role in regulating their affinity profiles. Additionally, there are design elements in proteins that are evolutionarily selected and can steer them away from unfavorable interactions. These design elements may not alter the binding of a substrate to the protein, but rather modulate the binding affinity allosterically by making or breaking specific interactions around the binding site. Therefore, studying systems like the SBPs from ABC and TRAP transporter complex can help us understand how multifaceted regulation, of structure and dynamics, occur in proteins that exhibit a selective specificity profiles and transports very distinct chemical entities across the cellular membrane.

1.2 Primary transport through the ABC systems

Translocation of substrates across the cell membrane is often facilitated by proteins that are classified as transporters. ABC transporters constitute a superfamily of ATP-powered membrane bound proteins that facilitate the selective active transport of a myriad of chemical entities across the membrane barrier. The ABC transporters are also involved in some non-transporter functions, such as in transcription regulation, mRNA translation, DNA damage repair and chromosome biosynthesis,⁽¹¹⁾ and biogenesis of iron-sulfur clusters. ABC proteins are ubiquitous in all kingdoms of life, with some even existing in the genomes of large viruses, and form one of the largest and most diverse class of paralogous proteins that catalyze a wide array of cellular reactions.⁽¹²⁾ ABC transporters cater to the diverse metabolic appetite of the cell by transporting metal ions, sugars, vitamins, amino acids, peptides, opines, xenobiotics, drugs, lipids, and siderophores.^(12,13) These transporters also work as exporters of efflux pumps and help in flushing out toxins, drugs, and lipids across the cellular membrane. ABC importers are found largely in bacteria, whereas the ABC exporters occur in both prokaryotes and eukaryotes. Evolutionarily, it has been suggested that the transporter proteins evolved from intergenic peptides through gene duplication events, and subsequently ABC importers evolved from ABC exporter proteins and that both groups later diverged to form separate functional classes.⁽¹⁴⁾ Given the fact that the ABC transporters are so important for cellular physiology, a significant part of the organismal genome codes for the different transporter systems. For example, in the bacterium *Escherichia coli*, around five percent of the genome essentially codes for over 80 different ABC transporter proteins.⁽¹⁵⁾ Since these transporters are driven by ATP hydrolysis, they are very specific towards their cognate ligands that they transport. This specificity helps the cells to conserve energy and to translocate only those nutrients that are required for important cellular processes. To understand how these systems achieve specificity in ligand binding and translocation, we must first appreciate their structural organization.

The core structure of the ABC transporter protein is formed by four domains, two transmembrane domains (TMDs) and two ABCs or nucleotide-binding domains (NBDs) (Figure 1.1). Depending upon the type of ABC transporter, this core structure may exist

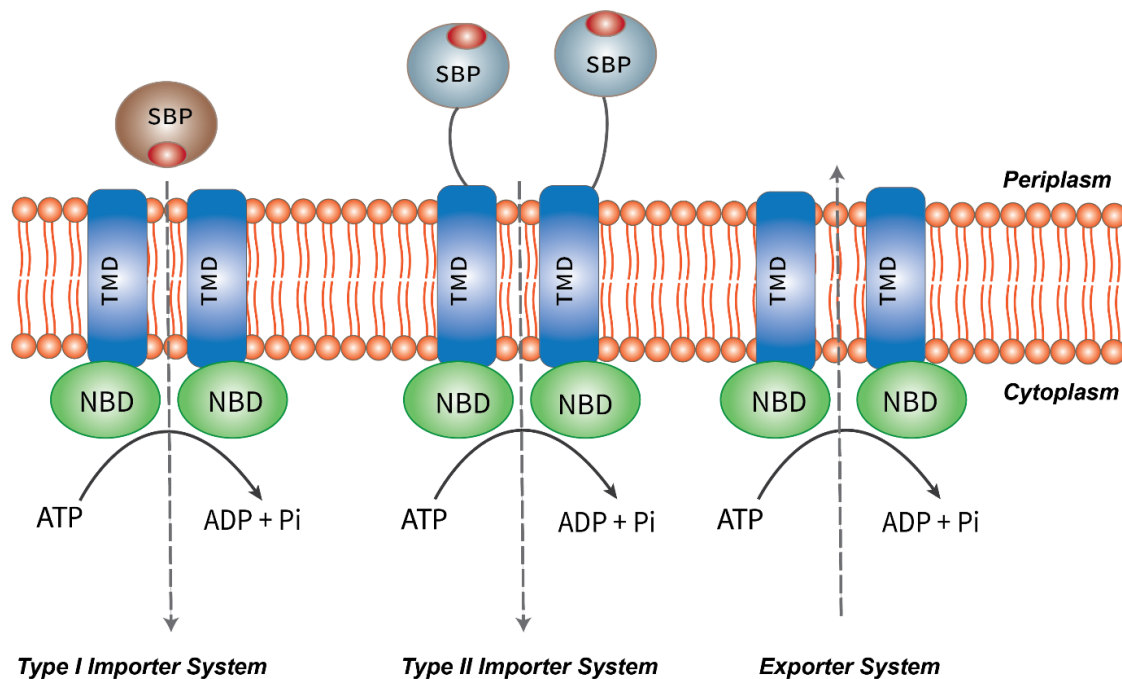


Figure 1.1: Schematics of the ABC importer and exporter systems. ABC importer type I and type II systems and the ABC exporter system are shown here. The transmembrane domain (TMD), which spans the entire length of the inner membrane are represented in blue. The nucleotide binding domains (NBD) are inside the cytoplasm and each NBD can bind one ATP molecule. Both type I and type II importers have a substrate binding protein (SBP), which is tethered to the TMD in the latter. ABC exporters also have TMDs and NBDs, but they lack SBPs and instead the TMDs function as direct binding sites for the molecules being exported out of the cytoplasm. The dotted arrow shows the direction of substrate translocation.

as a single polypeptide or as separate peptide subunits. The transmembrane domains span the inner membrane bilayer and are evolutionarily non-conserved. On the other hand, the NBDs are present at the cytoplasmic end of the transporter and contain highly conserved motifs and are among the most conserved proteins on Earth, with sequence conservation rivaling to that of tRNAs and rRNAs.⁽¹⁶⁾ The sequence variability in the TMDs enable them to bind and translocate different chemical entities. Among different ABC transporter proteins the TMD units can contain anywhere from 6 to 20 helical subunits, and the extent of the tightness of winding of the two TMDs during the translocation process can further influence substrate specificity.⁽¹⁷⁾ The ATP hydrolysis in the NBDs brings about a conformational change in the TMDs, allowing the translocation of bound substrate through the membrane barrier. There are some very important conserved sites in the NBD that regulate the function of the whole transporter, these are listed in Table 1.1. Other than containing these structurally and evolutionarily conserved motifs, the NBDs in some transporter complexes also contain a degenerated ATPase site, which further consist of large number of sequence variations.⁽¹⁸⁾ The TMDs and the NBDs are connected by a short cytoplasmic helix that runs parallel to the membrane plane. These cytoplasmic helices, termed as the coupling helix (also called helical loops), allow bidirectional conformational changes to occur between the NBDs and the TMDs (Figure 1.2).

In ABC importers each TMD has one coupling helix, whereas there are two coupling helices per TMD in ABC exporter systems.⁽¹⁷⁾ More recently, the presence of a reverse adenylate kinase (rAK)-like phosphoryl transfer reaction activity of the NBDs has been described, which enables a molecule each of ATP and adenosine monophosphate (AMP) from two molecules of adenosine diphosphate (ADP) during ATP scarcity.⁽¹⁹⁾ The rAK activity of the NBDs has since been reported for ABC exporter and lipid A flippase MsbA,⁽²⁰⁾ ABC exporters LmrA and TmrAB,⁽²⁰⁾ cystic fibrosis membrane conductance regulator (CFTR),⁽²¹⁾ and DNA damage repair protein Rad50.⁽²²⁾ Moreover, in some ABC importers there is an additional domain, termed as a regulatory domain (RD), attached to the NBDs that can bind substrate or regulatory protein halting ATP hydrolysis and in turn the entire translocation cycle.⁽²³⁾ Thus, the process of substrate translocation through the ABC transporter system is very complex, with a well concerted series of interactions that

Table 1.1: Conserved motifs in the NBD of the ABC transporter. Individual motifs and their functions are elaborated to highlight their role in ATP binding, positioning, and hydrolysis.

<i>Motif</i>	Function
<i>P loop (or Walker-A)</i>	Binds α - and β -phosphates of the ATP molecule
<i>A loop</i>	Provides aromatics side chain that stabilizes the purine ring of adenine
<i>Walker-B motif</i>	Provides catalytic glutamate
<i>Signature motif (LSGGQ)</i>	Orients ATP during hydrolysis
<i>Switch histidine</i>	Stabilizes the transition-state geometry
<i>Q-loop</i>	Provides contact to the TMD
<i>Dimerization or D loop</i>	Couples hydrolysis to transport

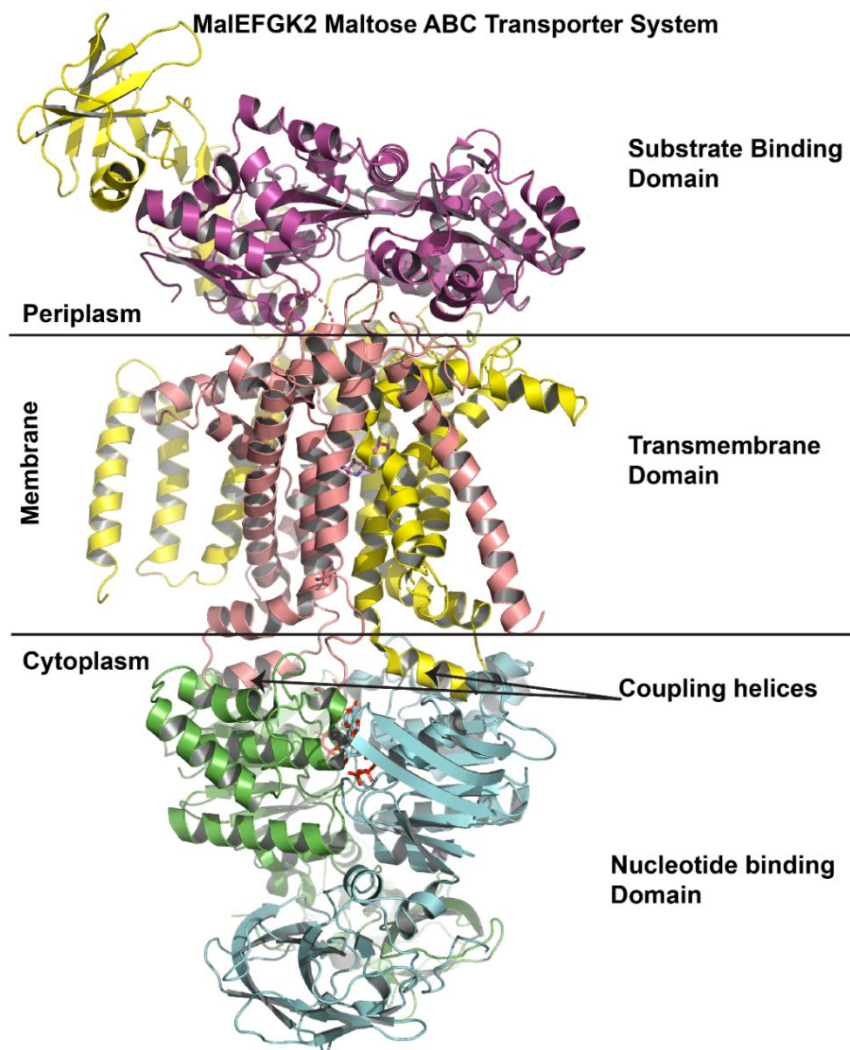


Figure 1.2: Structure and domain organization in *E. coli* MalEFGK2 maltose ABC transporter system. MalEFGK2 is a type I ABC importer with two transmembrane domains (TMDs) MalF (shown in yellow) and MalG (shown in salmon), two nucleotide binding domains (NBDs) shown in green and cyan. The periplasmic substrate binding protein (SBP) is shown in pink. The TMDs and the NBDs are connected by two coupling helices each from MalF and MalG that help relay the allosteric signal from periplasm to cytoplasm and vice versa. (PDB ID: 2R6G)

couple the individual domains and their correlated motions, which result in an allosteric coupling of ATP hydrolysis and substrate translocation.

ABC importers are employed by both gram-positive and gram-negative bacterial genomes to facilitate the transport and acquisition of substrates essential for their growth, such as metal ions, oligopeptides, oligosaccharides, amino acids, sugars, nucleosides, hemes, siderophores, and vitamins. ^(13,24) There are two different classes of ABC importer systems. The first and the larger class comprises of the ABC importer systems associated with the periplasmic substrate-binding proteins, which facilitate faster substrate uptake. In type-I systems, such as the molybdate/tungstate transporter (ModB₂C₂A) from *Archeoglobus fulgidus*, maltose transporter (MalFGK₂) from *E. coli*, and methionine transporter (MetNI), the substrate translocation pathway is formed by the dimerized core of TMDs, each having five TM helices. ⁽²⁵⁾ On the other hand, the type II importers, like the *E. coli* vitamin B₁₂ transporter (Btu(CD)₂), are homodimers of TMD subunit C and the NBD subunit D and consists of a much larger TMD than the type I systems. A recent structure of the Rv1819c transporter, a type II system, from *Mycobacterium tuberculosis* has shown that this transporter contains a very large substrate translocation permease domain with a cavity volume of around 7700 Å³. This allows the transporter to translocate structurally different substrates, such as the antimicrobial peptide bleomycin and cobalamin, and thus aids in bacterial pathogenesis. ⁽²⁶⁾ Given the novel fold architecture of type II importer systems and these recent findings, the type II importer systems are ideally suited for the uptake of much larger substrates.

Proteins associated with the second ABC importer class are not associated with a SBP and have only been structurally and mechanistically delineated in the past decade. This second class of ABC importers, also known as the energy-coupling factor (ECF) transporter systems, were initially identified in the gram-positive bacteria *Lactobacillus casei*, wherein they facilitated the transport of biotin, folate, and thiamine. ⁽²⁷⁻²⁹⁾ The ECF transporters are likely to have evolved to transport substrates that were required in very trace amounts by cellular enzymes to function, such as cofactors and ions. ⁽³⁰⁾ Structurally, the ECF transporters consist of a high-affinity substrate-binding transmembrane subunit (EcfS), a transmembrane coupling domain (EcfT), and a pair of NBDs (EcfAA'). ⁽³¹⁾

Mechanistically the ECF transporters differ significantly from the type-I ABC importers, although they also use the hydrolysis of two ATP molecules for substrate translocation. Substrate translocation in ECF transporters is explained as a “release and catch” mechanism. The transport process is initiated by the binding of two ATP molecules to the EcfAA' heterodimeric subunits, which is followed by a large scale domain motion in the coupling helix of the EcfT domain and a subsequent release of the EcfS domain for substrate capture. ⁽³²⁾ Once the substrate is captured, the EcfS protein interacts with the coupling helix of the EcfT, followed by the ATP hydrolysis event that topples the EcfS domain and facilitates substrate translocation in the cytoplasm. ⁽³²⁾

1.2.1 The maltose importer system

An extensive accumulation of biochemical, biophysical, and structural studies over the past two decades have made the *E. coli* maltose transporter system (MalEFGK₂) one of the best characterized ABC systems to date. This transporter system was initially studied as a model system to comprehend the mechanism of positively regulated regulons and their interaction with the λ phage. ⁽¹²⁾ It was later identified that the λ phage interacts with the outer membrane protein LamB or maltoporins, which help move maltose and malodextrins across the bacterial outer membrane into the periplasm, thus explaining the susceptibility of maltose transport to λ phage infections. ^(33,34) Subsequently, efforts were made to identify individual components of the maltose transporter system, to identify and annotate the involved genes, and to clone, express and eventually determine their structures. ⁽¹²⁾

The *E. coli* maltose regulon encodes proteins that are tasked with the uptake of $\alpha(1-4)$ -linked glucose oligosaccharide. ⁽³⁵⁾ There are three components of the maltose ABC importer system: (i) the membrane-bound TMDs MalF and MalG, (ii) two cytoplasmic NDBs MalK homodimers, and (iii) a soluble periplasmic SBP MalE, the quintessential third component (Figure 1.2). MalE is a bi-domain protein connected by a flexible hinge that identifies and binds to maltose and malodextrins at the bi-domain interface. Subsequently, the SBP docks on to the membrane embedded TMDs, triggering a series of conformational changes that regulates ATP hydrolysis and results in the translocation of

substrates across the membrane. The unidirectional movement of the substrate occurs in an alternate access mode,⁽¹⁸⁾ with both the docking of substrate-bound MalE to the TMDs and the binding of 2 ATP molecules to the MalK subunit required for substrate translocation to occur. The stoichiometry of MalEFGK₂ subunits was initially determined by immunoprecipitation of the detergent stabilized *E. coli* membrane by Davidson et al.⁽³⁶⁾ and later the crystal structure revealed that there are single copies of TMDs MalF and MalG whereas the NBD MalK existed as homodimer.⁽³⁷⁾

The periplasmic SBP MalE was the first component of the maltose ABC transporter system to be characterized and crystallized.⁽³⁸⁻⁴⁰⁾ Initial knockout studies on the MalE subunit established it as an essential component required for *in vivo* maltose uptake, as in its absence, even at high substrate concentrations, transporter function was seized.⁽³⁴⁾ MalE binds to maltose and malto-oligomers with varying affinity, which typically ranges in the low micromolar concentrations.^(41,42) Binding of these cognate substrates induces a signature hinge-bending motion bringing the two domains of the SBP closer and trapping the substrate in the middle (closed conformation). Binding of the MalE protein to the TMDs initiates a reverse hinge motion, separating the two domains and releasing the substrate in the process (open conformation). MalE also binds several non-cognate ligands with very high affinity, yet these ligands fail to induce the characteristic hinge bending motion from open to closed conformation and results in the failure to translocate these ligands.⁽⁴³⁻⁴⁵⁾ Nonetheless, some ABC transporters directly interact with the substrates through an additional substrate binding site on their TMDs. For example, the MalFGK₂ from *E. coli* and Art(QM)₂ from *Thermoanaerobacter tengcongensis* contain such sites on their TMDs, suggesting that these transporters may have a basal rate of substrate translocation in absence of the SBPs.^(46,47) It has also been proposed that substrate binding pockets may exist in other ABC importer systems where the TMDs transition from outward- to inward-facing states.⁽⁴⁸⁻⁵⁰⁾

Another important binding event in the maltose transporters is the binding and subsequent hydrolysis of 2 ATP molecules by the homodimeric MalK subunits in the cytoplasm. The power stroke event that enables the substrate translocation across the membrane in maltose transporter is strongly dependent on the presence of MalE and substrate.^(36,51)

The alternate access mechanism that follows binding of the loaded SBP to the TMDs and ATP binding to the homodimeric NBDs allows the substrate to move down the TMD translocation pore in a concerted fashion that involves several intermediate and transient forms of the maltose transporter. These shall be discussed in detail in section 1.4. of this chapter.

1.3 TRAP systems: a secondary transport system for nutrient uptake

Unlike the primary transporters, secondary transporters do not rely upon ATP hydrolysis for substrate translocation and instead couple the substrate movement across the membrane with a proton motive force (PMF) or an ion-electrochemical gradient. In 1997, Forward et al. discovered a novel genes encoding an ATP-independent secondary transporter system in *Rhodobacter capsulatus* that consisted of a binding protein and two unequal membrane proteins, which the authors named TRAP (tripartite ATP-independent periplasmic) transporters.⁽⁵²⁾ The unique feature of this system was that it was the only secondary transporter system known then to consist of a SBP. As the name suggests, the TRAP transporters consist of three functional units: a periplasmic SBP, a smaller TMD consisting of four TM helices, and a larger TMD consisting of 12 TM helices (Figure 1.3). With only a handful of transporter protein characterized to date, and with limited structural information available, the TRAP-mediated transport of nutrients is shrouded in mystery. However, recent biochemical, biophysical, and structural characterization of TRAP systems, predominantly the SBPs, have shed some light on their cognate substrate profiles and a possible mode of substrate translocation.⁽⁵³⁾

Results from the past decade suggest that bacteria require multiple types of transporters to translocate carboxylates across its inner membrane into the cytoplasm (54). These are predominantly transporters from the ABC superfamily and the major facilitator superfamily (MFS). The ABC-transporters are primary transporters that use ATP hydrolysis to drive uptake. Conversely, MFS energizes the movement of substrate across the membrane by associating the translocation with a pre-generated ion gradient.

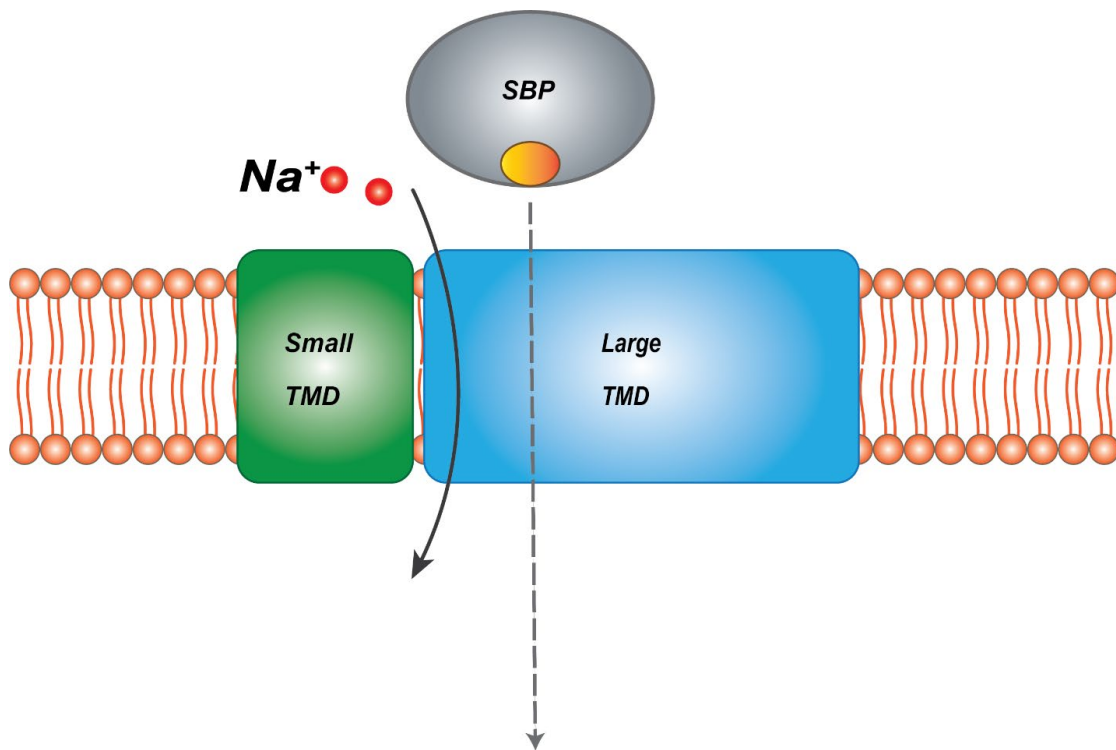


Figure 1.3: Schematics of the TRAP and TRAP-like transporter systems. The TRAP transporter system consists of three domains, two of which are transmembrane domains (TMDs) and one is a periplasmic substrate binding protein (SBP). The smaller TMD (shown in green) is made up of four transmembrane helices (TMHs) and the larger TMD (shown in blue) comprises of 12 TMHs. Substrate translocation is tied to the Na^+ gradient in TRAPs, where two Na^+ ions are required per molecule of substrate for the transport to occur. In TRAP transporters and tripartite tricarboxylate transporters (TTTs) the small TMD and large TMD work as separate proteins, while in TRAP associated extra-cytoplasmic immunogenic (TAXI) proteins the two TMDs are fused by an addition of an extra helix.

These include several subclasses of secondary transporter that differ in their mechanism of uptake of specific carboxylates, such as sialic acid uptake. Examples include NanT transporter that depends on H^+ gradient, SBP associated TRAP transporter and their variants, and the sodium solute symporter.^(54,55) Additionally, two different types of ABC-transporters from *Haemophilus ducreyi*⁽⁵⁶⁾ and *Corynebacterium glutamicum*,⁽⁵⁷⁾ SatABCD and SatABC, are also known to translocate sialic acid.^(54,55) While most bacteria use more than one transporter to acquire sialates, in some cases the TRAP transporter is the sole translocator of sialic acid. This makes TRAP transporter systems essential for the growth and development of pathogenic bacteria and can thus be tested for potential druggability.

The TRAP transporters are members of the SBP-dependent secondary transporters that couple substrate translocation across the membrane with a pre-established Na^+ electrochemical gradient.⁽⁵⁸⁾ Unlike the ubiquitous ABC transporters, the TRAP transporters are present only in archaea and bacteria and can co-translocate C_4 - and C_6 -dicarboxylates and ions across the membrane in either direction based on the concentration gradient.^(55,59) TRAP transporters lie at the intersection of primary and secondary transporter systems. Like ABC transporters, the TRAP transporter has a high-affinity SBP that presents the substrate in the periplasm. However, membrane co-translocation is not dependent on the energy derived from ATP-hydrolysis, but is instead dependent on electrochemical Na^+ gradient, like the secondary symporters.

The energy cost vs. efficiency ratio is higher for TRAP transporters than for the ABC transporters, making TRAP systems well suited for organisms inhabiting substrate deprived conditions.⁽⁶⁰⁾ Recently, these systems have been implicated in the enhanced survival of several pathogenic microbes in human hosts through high-affinity sialic acid uptake, particularly in *V. cholerae*.⁽⁶¹⁻⁶³⁾ In particular, TRAP system SiaPQM (VC1777-1779), the most in-depth studied TRAP system yet, is the sole transporter of N-Acetylneuraminic acid (Neu5Ac) or sialic acid, a C_4 - dicarboxylate, in *Vibrio cholerae* and is encoded within Vibrio pathogenicity island 2 (VPI-2).⁽⁶³⁾

The SiaPQM transporter system comprises of a two-domain SBP (SiaP) and two unequal-sized TM proteins (SiaQ and SiaM with 4 and 12 TM helices, respectively) in a 1:1:1 ratio.

⁽⁶¹⁾ SiaP binds Neu5Ac with very high affinity and specificity, forming a salt bridge between the carboxylate moiety of the substrate and a conserved arginine residue in its binding pocket.⁽⁶⁰⁾ Furthermore, SiaP falls under cluster E of the new SBP classification system and shows structural divergence from the ABC-transporter associated SBP.^(64,65) Thus, while TRAP associated SBPs resemble their distant homologs from the ABC transporter system, they have evolved to carry-out different functions and display significant divergence in their local structures, even though the overall fold is conserved. For example, like the ABC-transporter associated SBPs, TRAP SBPs have undergone subtle changes in their binding sites, which allows them to bind myriad substrates with differing affinity profiles. For instance, both *Haemophilus influenzae* and *V. cholerae* SiaP binds Neu5Ac with a K_d of about 110 nM.⁽⁵⁸⁾ However, changes in the sugar acid substrates drastically affects the affinity of the SBP by over a hundred-fold in some cases.⁽⁵⁸⁾

While structural information on TRAP transporters is well elucidated for the SBP, no atomistic structure is available for the TM proteins to date. The larger twelve TM SiaM subunit is predicted to be homologous to other bacterial secondary transporters⁽⁵⁵⁾, like the anion/ Na^+ cotransporter (NadC; 40% sequence similarity), but no defined function has been established for the smaller 4 TM SiaQ subunit (Figure 1.4). However, it has been proposed that SiaQ might provide additional surface for SiaP binding^(55,58), and that SiaQ might also help in proper folding of SiaM in the membrane, as in its absence transportation is stalled.^(55,61) Experimental and topological predictions using bioinformatics tools have shown that both N- and C-terminal regions are in the cytoplasm of SiaQ (N_{IN} C_{IN}) and in the periplasm for SiaM (N_{OUT} C_{OUT}), which is similar to that observed in other secondary transporters.^(55,66) While SiaM has two additional TMH insertions in the N- and C-terminal regions of the protein, the core 10 TMH architecture is conserved like other secondary transporters.^(67,68)

Although there is limited structural data on the TM protein components, there is significant biochemical information on the SiaPQM from *V. cholerae* (vcSiaPQM). Proteoliposome reconstituted vcSiaPQM showed dependence only on the Na^+ gradient and membrane potential but not on pH, unlike other secondary transporters.⁽⁶¹⁾ It has been proposed that binding of the holo SiaP to the TM SiaQ and SiaM causes conformational changes in

the SBP, which is passed down to the TM transporter proteins, allowing the release of Neu5Ac to the outward open conformation of the TM transporter.^(55,60) The TRAP symporter then conjoins the movement of two Na⁺ ions with Neu5Ac and transports it across the membrane by an alternate access mechanism.⁽⁶¹⁾ However, currently there is no structural evidence to support this model of substrate translocation by TRAP transporters. Based on the available literature on TRAPs, an alternate access model of substrate translocation is presented in the next section.

1.4 Alternate access model of substrate translocation through ABC and TRAP transporters

Members of both the primary and the secondary transporter superfamilies translocate substrate across both sides of the membrane using this alternating access mechanism.^(69,70) As stated above, the maltose ABC transporter system MalEFGK₂ is one of the best studied type I importer systems. Broadly, ABC importers alternate between large-scale conformational changes that allow them two transitions between an inward-open state and an outward-open state exposing the substrate binding site on the open side. This pattern of sequential switching between the two conformational states that allows substrate to transit through the translocation pore and across the membrane, is termed as an alternate-access model. The same mechanism has also been proposed for the translocation of substrates across the TRAP transporter system,⁽⁷¹⁾ though there is currently no structural data to confirm this mode of substrate translocation.

Extensive biochemical and structural studies on the MalEFGK₂ has suggested that there are four states of the type I transporter proteins. These include (i) substrate-binding to the SBP MalE, (ii) docking of the substrate-bound MalE onto the MalFG, (iii) ATP-hydrolysis and substrate delivery to MalFGK₂, and (iv) substrate translocation across membrane and transporter reset (Figure 1.4). Initially, the resting state of the MalFGK₂ transporter was found be devoid of the SBP and there was a clear separation between the MalK homodimers, which did not bind an ATP molecule each.⁽⁷²⁾ In the resting state, the substrate binding site on the MalF subunit of the TMDs is exposed towards the cytoplasm.

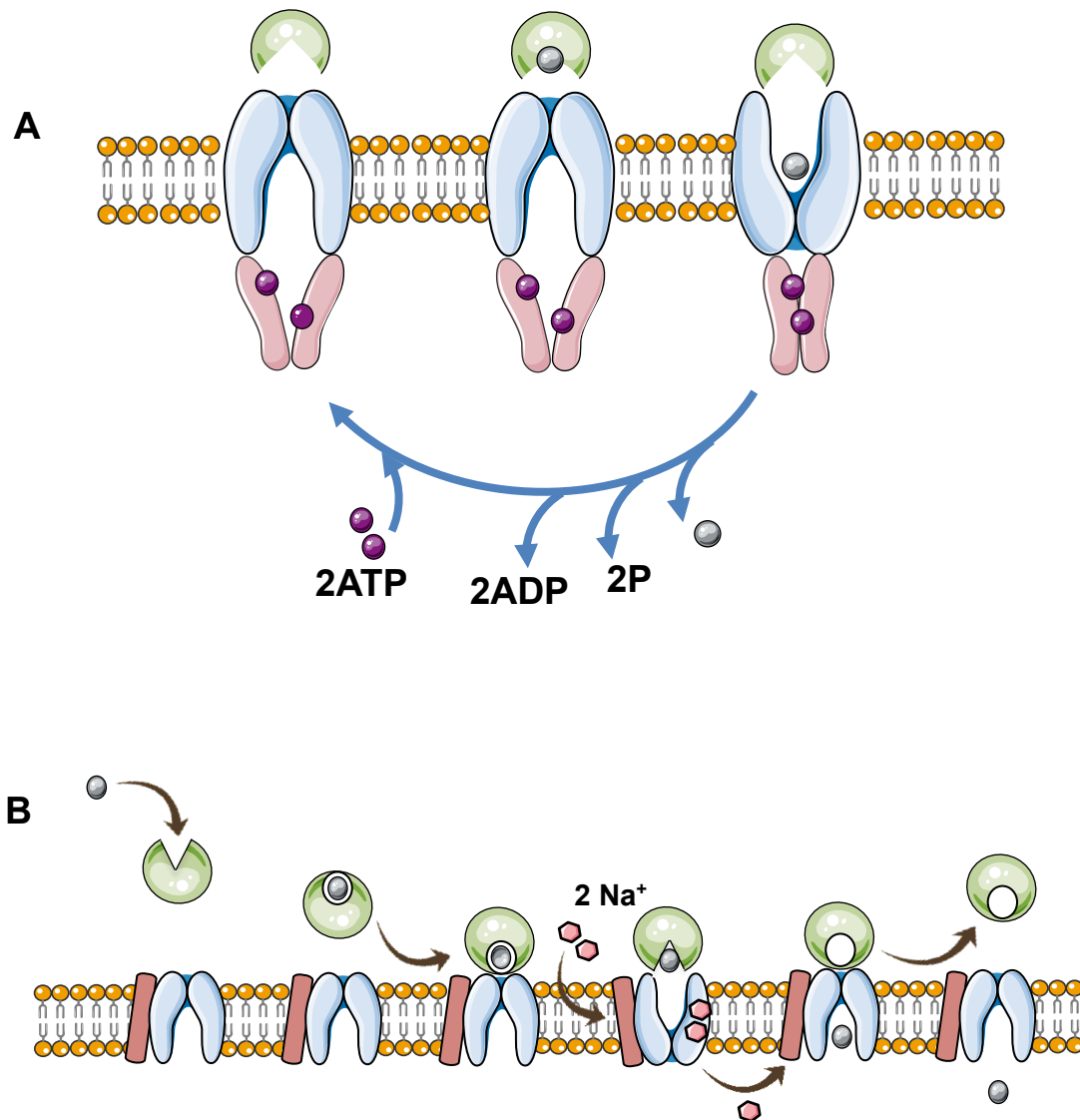


Figure 1.4: Mechanism of substrate translocation in ABC and TRAP systems. Both ABC and TRAP transporters translocate substrates across the biological membrane using an alternate-access mechanism. (A) ABC type I importers undergo a transition between inward- and outward-facing forms. The SBP is shown in green, TMDs are shown in blue and the NBD and RD in pink. Substrate maltose is shown as grey sphere. (B) TRAP transporter components are shown here, SBP is shown in green, smaller TM subunit is shown in red, and larger TM subunit is shown in blue. Substrate is shown in grey.

Furthermore, the translocation pore formed by the TMDs is closed on the periplasmic side of the transporter, possibly inhibiting any low-affinity substrate interactions.

Subsequently, the substrate-bound MalE protein docks onto the transporter protein and this interaction is stabilized by the P2-loop of the MalF subunit, which interacts with the N-terminal of MalE.⁽⁷³⁾ The P2-loop connects the third and fourth TM helices of MalF subunit and binds both substrate-bound MalE (7 μ M) and apo-MalE (22 μ M) with varying affinities.⁽⁷⁴⁾ This step constitutes the pre-translocation arrangement of the transporter where the transporter is still in the inward facing state. Here, the distance between the two MalK homodimers is reduced, bringing the two NBDs to partially overlap and their key catalytic motifs (listed in Table 1.1.) coming in close proximity of each other. The interactions between catalytic residues on both NBDs further aid in binding of one ATP and one Mg^{2+} ion to each MalK subunit. Binding of the second ATP molecule causes a closed NBD state, thus allowing the transporter to shift from a pre-translocation to translocation state.⁽⁷³⁾

In the translocation state of the transporter, the two NBDs overlap allowing the hydrolysis of the ATP molecules and providing the required power-stroke for substrate translocation across the membrane. In this state, the transporter acquires an outward-open state where the substrate-binding site on the MalF is exposed towards the periplasmic side. This allows the allosteric interactions from the NBDs to release the substrate from MalE into a large translocation pore that can accommodate molecules as large as a maltoheptose.⁽²⁵⁾ Crystallization experiments suggested that a substrate-bound MalE allows the stabilization of the outward facing state of the transporter.^(75,76) Interestingly, it was identified that the maltose molecule forms multiple hydrogen-bonding and stacking interactions only with the MalF subunit and there are no observable interactions with the MalG subunit of the TMD in the outward open state.⁽⁷⁷⁾ Finally, the release of MalE and ADP-Pi from NBD allows the substrate to enter the cytosol and resets the transporter in the resting state.

The absence of such structural information on the TRAP TM proteins means knowledge of how the TRAP and TRAP-like proteins work is primarily based upon structural studies

on the SBP and biochemical analysis of the SiaPQM TRAP system. The translocation process in TRAPs begin with the binding of substrate to the SBP, inducing a conformational change in the SBP that locks the substrate. This conformationally "closed" state is recognized by the SiaQM TM subunits, which at this point are in an inward-facing resting state. The assumption about the resting state of the membrane permease comes from the fact that TRAP TM subunits, especially the larger SiaM subunit, shows similarity with other secondary transporter permeases, like LacY, and GlpT.⁽⁷¹⁾ The subsequent binding of the SBP to the TM subunit allows binding of 2 Na⁺ ions to the TM permease through an unknown mechanism. This causes a further conformational change in the TM subunit SiaQM towards an outward-open state, thus releasing the substrate from the binding protein in the transmembrane channel. It is thought that SiaM acts as the permease domain in TRAPs, binding both the Na⁺ ions and the substrate.⁽⁷¹⁾ Once SiaM receives the substrate from the SBP, it flips back into the inward-facing state and releases the substrate and the sodium ions in the cytosol. This mechanism of substrate translocation in TRAPs is more energy efficient than those in ABC transporters or other secondary transporter systems, and the presence of the SBP allows them to show comparable substrate affinity profiles with that of the ABC transporter systems.⁽⁷⁸⁾

1.5 Substrate binding proteins: role in substrate selectivity and translocation

The survival of the bacterial cells hinges squarely on their ability to selectively uptake only the essential metabolites while keeping the cytotoxic chemicals at bay.⁽⁷⁹⁾ This direct dependence on the metabolic transport system for survival has incentivized the evolution of specificity towards particular substrates in transporter proteins. As stated above, a significantly large chunk (~5%) of *E. coli* K-12 genome is dedicated to code for the transporter proteins, with 281 functionally characterized transporter and binding proteins, and an additional 146 putative transporter and binding proteins.⁽⁸⁰⁾ SBPs were initially discovered when gram-negative bacteria were subjected to plasmolysis, wherein an osmotic shock damages the OM causing release of the contents of the bacterial periplasm and inhibiting nutrient transport by periplasmic permeases.⁽⁸¹⁾ For this reason, the specific

class of proteins released due to the osmotic shock were loosely termed "periplasmic proteins". Today, these proteins are known as either periplasmic binding proteins (PBPs) or SBPs.

In gram-negative bacteria, periplasmic SBPs predominantly operate in conjunction with the ATP-binding cassette (ABC) transporters and the tripartite ATP-independent transporter (TRAP) systems, where they bind a diverse array of metabolites in the cellular milieu. The ABC transporters are primary transporter systems that couple the substrate translocation across the bacterial membrane to a subsequent hydrolysis of two ATP molecules. On the other hand, the TRAP transporters symport the substrate translocation with a sodium ion gradient, like the secondary transporter systems. On the other hand, in gram positive bacteria the substrate-binding domains (SBDs) are tethered to the TMDs and are only associated with type II ABC systems.

Even though the overall architecture of the SBPs is restricted by a bidomain fold, nonetheless, their binding affinities are highly tuned to the environmental niche of the host organism, which results in differing affinity profiles for identical ligands by different homologous SBPs. Where ABC transporters can facilitate the translocation of peptides, sugar oligomers, ions and vitamins, TRAP proteins are uniquely suited for moving sugar acids intermediates of various metabolic cycles, such as the tricarboxylic acid (TCA) cycle and amino acid biosynthetic pathways. These facts further underscore the importance of studying the SBPs in greater detail as they are perfect model systems to comprehensively delineate the evolution of specificity in a complex protein fold.

1.5.1 Structure, function, and classification of SBPs

Structurally, SBPs have a distinctive bi-domain architecture with the N- and C-terminal domains linked by a flexible linker region that serves as a hinge. The substrate binding site resides at the interface of these two domains, and the hinge is responsible for the ligand dependent domain motions that are characteristic of this superfamily. The substrate binding site is located within this flexible hinged interface, which allows the evolution of flexibility and specificity in these systems without affecting the overall fold. All SBPs identified to date have a signature $\alpha\beta$ -fold with the β -strands arranged at the core

surrounded by α -helices. The SBPs can vary extensively in their size (25 – 70 kilodaltons (KDa)) yet they retain the same structural element and overall fold.

The first classification of the SBPs divided them into three sub-classes based on the topological organization of their secondary structural elements.⁽⁸²⁾ This initial classification, based on multiple sequence alignment, also showed for the first time that a single evolutionary event of domain dissociation resulted in evolution of Class II SBPs from Class I SBPs. The substantial growth in the number of SBP structures in the following decade enabled the classification system to be further expanded and provided a more coherent and robust understanding of functional segregation in SBPs. Thus, Berntsson et al. compared structures of 107 SBPs (103 substrate-bound and 4 substrate-free SBP structures) and identified six independent clusters (A to F).⁽⁸³⁾ Later, this classification was expanded by the addition of a seventh cluster, cluster G, when the same structural classification was carried out with 387 more SBPs, along with an initial 114 from Berntsson et al.⁽⁸⁴⁾ SBPs within each cluster share some degree of structural similarity, but display a remarkable diversity in specificity and function.

Cluster A consists of SBPs that were previously classified as class III SBPs and they are mainly associated with the ABC type II importer systems that transport metal ions and chelates. This cluster is further subdivided into clusters A-I and A-II. While the overall structural features of the proteins belonging to subcluster A-I and A-II are similar, especially the single α -helix that serves as the hinge (Figure 1.5. B), they differ significantly in the organization of the loops that line the binding cavity. Functionally, SBPs in subcluster A-I facilitate the binding of metal ions whereas subcluster A-II helps in binding siderophores and heme.^(83,84) Examples belonging to the subclass A-I include the zinc binding protein ZnuA from *E.coli* (PDB ID: 2PRS)⁽⁸⁵⁾ and the manganese-binding protein from *Listeria monocytogenes* (PDB ID: 5HX7). Members of subclass A-II include the iron-binding protein from *Theriotoga maritima* (PDB ID: 2ETV) and the siderophore-binding protein FitE from *E. coli* (PDB ID: 3BE6).⁽⁸⁶⁾

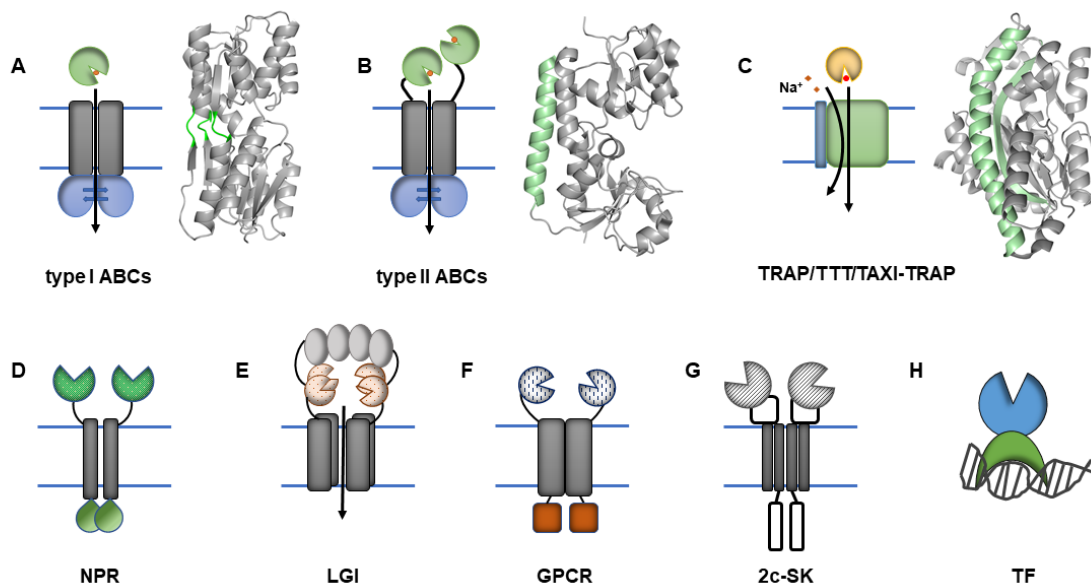


Figure 1.5: Structural and functional diversity in SBPs. (A) Schematics of free SBP (green) associated with the type I ABC importer system and a cartoon representation of the SBP with the hinge region highlighted in green. (B) schematic representation of type II ABC importer system showing the tethered SBP (green) linked to the TMD through lipopolysaccharides, and a cartoon of SBP with single α -helical hinge (green). (C) TRAP transporter and SBP SiaP showing the long domain-spanning helix. Other systems that also are associated with SBPs are (D) the natriuretic peptide receptors (NPR), (E) the ligand-gated ion (LGI) channels, (F) G-protein coupled receptors (GPCR), (G) two-component signal transduction systems, and (H) transcription factors

Cluster B primarily consists of SBPs that have a hinge comprising of three interconnecting β -strands that connect the two α/β -domains. SBPs belonging to cluster B are among a diverse set of proteins and are associated with type-I ABC transporters (Figure 1.5 A), natriuretic peptide receptors (NPR; Figure 1.5 D), G-coupled protein receptors (GPCRs; Figure 1.5 F), two component systems (Figure 1.5 G), and transcription regulators (Figure 1.5 H).^(83,84) Cluster B is further subdivided into five subclusters (B-I through B-V) based on the type of substrates they bind to, and presence or absence of an additional helix near the binding cavity.⁽⁸⁴⁾ The subcluster B-I comprises of members of the *lac*-repressor type transcriptional regulators that modulate the sugar metabolism by binding carbohydrates. Examples of B-I SBPs include the D-xylose receptor (PDB ID: 3M9W)⁽⁸⁷⁾, and the arabinose receptor (PDB ID: 1ABF)⁽⁸⁾. SBPs in subclusters B-II and B-III are associated with type II ABC transporter systems and they bind and aid in the translocation of amino acids and aromatic compounds, respectively.⁽⁸⁴⁾ One representative SBP from B-II subcluster is the glutamate-binding domain of the human ionotropic glutamate receptor (PDB ID: 2WJW).⁽⁸⁸⁾ An example of B-III SBP is the benzoyl-formic acid binding protein from *Rhodopseudomonas palustris* (PDB ID: 3SG0).⁽⁸⁹⁾ Subclusters B-IV and B-V are involved in receptor-mediated signal transduction as they interact with NPRs and GPCRs, respectively. An example of a B-IV SBP includes the natriuretic peptide hormone binding domain of the human NPR (PDB ID: 1JDN)⁽⁹⁰⁾, and the human metabotropic glutamate receptor domain (PDB ID: 3MQ4) is a B-V type SBP.

Cluster C is made up of those SBPs which have an additional domain between the two α/β -domains and are largest among all the SBPs, with sizes ranging from 55-70 KDa.⁽⁸³⁾ The function of the additional domain is not very well understood, but it is considered to provide extra binding surface for larger ligands. Examples of cluster C SBPs include the nickel-binding protein NikA from *E. coli* (PDB ID: 1UIV)⁽⁹¹⁾ and the oligopeptide-binding protein OppA from *Lactococcus lactis* (PDB ID: 3DRF)⁽⁹²⁾. These SBPs interact with ABC transporters.

SBPs forming the cluster D are associated with type II ABC transporter systems and with the *lysR*-type transcriptional regulators (LTTR). Cluster D SBPs bind a number of substrates that include carbohydrates, putrescine, thiamine, tetrahedral oxyanions, and

ferric and ferrous ions. ⁽⁸³⁾ Subcluster D-I contains proteins that bind carbohydrates, like maltose and glucose, and show structural similarity with proteins in clusters C-F. D-I SBPs are also much larger in size than SBPs forming Cluster B, even when both can bind similar substrates. However, the increased size of D-I SBPs allow them to bind larger sugar oligomers. For example, the *E. coli* maltose-binding protein MalE (PDB ID: 1ANF) can bind various maltose oligomers with varying degree of substrate specificity. ⁽⁴²⁾ The second subcluster (D-II) consists of SBPs that belong to the LTTRs and differ considerably in their structural organization from subcluster B-I, which comprises of the *lac*-repressor type transcriptional regulators. Protein CysB from *Salmonella typhimurium* in complex with O-acetylserine (PDB ID: 4LQ2) is an example for D-II type SBPs. Subcluster D-III comprise of the tetrahedral oxanion-binding proteins that can bind phosphate, molybdate, and tungstate. ⁽⁸³⁾ Subsequent reclassification further subdivided D-III into D-IIIa and D-IIIb based on inherent structural feature differences among the proteins. D-IIIa type SBPs include the *E. coli* phosphate binding protein PBP (PDB ID: 1IXH) ⁽⁹³⁾, whereas the molybdate/tungstate binding protein ModA/WtpA from *Pyrococcus furiosus* (PDB ID: 3CG1) ⁽⁹⁴⁾ belongs to subcluster D-IIIb. Finally, subcluster D-IV consists of proteins that can bind ferrous or ferric ions and are collectively called iron-binding proteins (FBPs). One example of a FBP is the *Haemophilus influenzae* Fe(+3)-binding protein (hFBP) (PDB ID: 1MRP). ⁽⁹⁵⁾

SBPs that are integral to the TRAP/TAXI-TRAP and TT transporters exclusively form cluster E (Figure 1.5 C). These SBPs bind several C2-, C3-, C4-carboxylates, such as succinate, fumarate, malate, and sialic acid. These proteins have several unique structural features that are different from other SBPs, especially with the ones that interact with the ABC transporters. These proteins have long β -strands and a long α -helix that spans across both domains. The specific difference between TRAP SBPs and TTT SBPs is that the domain spanning α -helix in the latter is shortened. Therefore, cluster E was subdivided into E-I, which consists of TRAP SBPs, and E-II, which comprises of TTT SBPs. The sialic acid binding protein SiaP from *H. influenzae* is the most extensively studied TRAP SBPs (PDB ID: 3B50) ⁽⁹⁶⁾, whereas the glutamate-binding protein BugE (PDB ID: 2DVZ) and aspartate-binding protein BugD (PDB ID: 2F5X) from *Bordetella pertussis* are examples of TTT associated SBPs. ^(97,98)

The two domains of the SBPs belonging to cluster F are connected by longer hinges, typically 8-10 amino acids, which confers a higher flexibility during conformational switching between open and closed conformations.⁽⁸³⁾ These binding proteins are associated with type I ABC importers, ligand-gated ion channels, and GPCRs⁽⁸⁴⁾ and bind a variety of substrates ranging from trigonal planar anion, like nitrates and bicarbonates, to amino acids, and glycine betaine.⁽⁸³⁾ Moreover, it is this variation in the type of substrate, and not an inherent difference in the structural organization of cluster F SBPs, that is the basis for further subdividing this cluster into four subclusters. Subcluster F-I binds trigonal planar anions and includes proteins such as the nitrate-binding protein NtrA from *Synechocystis sp.* (PDB ID: 2G29)⁽⁹⁹⁾. The smaller F-II subcluster consists of proteins that bind to amino acid methionine and dipeptides. The methionine-binding protein MetQ from *E. coli* (PDB ID: 4YAH)⁽¹⁰⁰⁾ is an example of SBPs that form subcluster F-II. Proteins assigned to subcluster F-III bind to several cellular osmoregulatory solutes, such as the glycine betaine, proline betaine and choline.⁽⁸³⁾ Uniquely, the F-III subcluster binding proteins use cation- π interactions to stabilize their substrates in the binding cavity.⁽⁸³⁾ The glycine betaine binding protein ProX from *E. coli* (PDB ID: 1R9L)⁽¹⁰¹⁾ falls under subcluster F-III. Finally, subcluster F-IV binding proteins interact with charged amino acids such as the arginine-binding protein ArgBP from *T. maritima* (PDB ID: 4PRS)⁽¹⁰²⁾, which neutralizes the charges on the amino acids and binds them by making hydrogen bonds with the side chains of these charged amino acids.⁽⁸³⁾

A new cluster, cluster G, was identified while expanding the classification of SBPs by Scheepers et al.⁽⁸⁴⁾ in 2016. Proteins in cluster G show large structure (~60 KDa) and they also contain a metal binding site. The alginate binding protein AlgQ1 (PDB ID: 1Y3P) and AlgQ2 (PDB ID: 1J1N) from the *Sphingomonas sp.* were assigned to this cluster.^(103,104)

1.5.2 Dynamics of SBPs

Several biophysical, structural, and computational molecular dynamics simulation studies have delineated a complex dynamical profile of the SBPs, which directly influences the rate of substrate translocation across the cellular membrane.⁽¹⁰⁵⁻¹⁰⁷⁾ The conformational switching of the two domains/lobes of the SBP into an open (two lobes are separated) or

closed (two lobes come in close proximity of each other) is induced by substrate binding at the binding site between the two domains. Additionally, there are two predominant motions observed between apo (open) and holo (closed) states: (1) the hinge bending motion where the two domains move towards one another, (2) the twisting motion of one domain relative of the other. ⁽³⁹⁾ However, the extent to which these lobes close or open varies considerably among SBPs. The closing of the two domains in the Vitamin B-12 binding protein BtuF makes an angle of $\sim 35^\circ$, whereas the amino acid binding protein LivJ shows a significantly large hinge bending angle of $\sim 60^\circ$. ^(108,109) This trend was also recently seen in MBP isoforms from *Thermotoga maritima* where isoforms tmMBP1 showed a hinge bending angle of $\sim 50^\circ$ compared to a $\sim 35^\circ$ hinge bending angle for tmMBP3 (see Chapter 2 for more details). ⁽¹¹⁰⁾ These studies highlight that SBPs show remarkably different dynamical behaviors that do not necessarily depend on their structural organization.

Perhaps the most important dynamic event associated with the SBPs is its hinge bending and twisting event, which is a defining feature of each SBP cluster. ^(83,84) Additionally, both experimental and computational studies have shown that the apo-SBPs displays inherent conformational dynamics in the absence of substrate, which might suggest a complex dynamical behavior associated with substrate-recognition and binding. ⁽¹¹¹⁻¹¹⁵⁾ Moreover, it has also been shown that the conformational sampling of these SBPs differs significantly in presence of cognate versus non-cognate ligands. ^(43,45,116) Transport in the case of non-cognate ligands fail because the SBPs either undergo partial hinge bending thereby causing an incomplete domain closure, or the binding of a non-cognate ligand may result in a very tight binding profile. ⁽¹¹⁶⁻¹¹⁸⁾ Thus, the resulting undesirable conformational events in presence of a non-cognate ligand can either prevent the proper docking of the SBP on the transporter protein, or might affect the release of substrate to the transporter due to very tight binding. ⁽¹¹⁸⁾ These findings suggest that the SBP-substrate interaction is a far more complex event than previously thought, which has a profound effect on the ability of a substrate to be transported.

In particular, the event of cognate-substrate binding and translocation have also been studied in MBPs in presence of several malto-oligomers of varying sizes. ^(116,119) Using a

series of experiments that combined single-molecule Förster resonance energy transfer (smFRET) and transport measurements, Boer *et al.* showed that MalE-ligand complexes exhibit at least three distinct conformational states induced by the binding of the cognate ligands.⁽¹¹⁶⁾ Maltopentose, maltohexaose, and maltoheptose bound to MalE in a partially closed states, whereas maltose, maltotriose, and maltotetraose induced a fully closed conformational state of the protein.⁽¹¹⁶⁾ Even with a varying degree of domain closure in MalE, malto-oligomers up to maltoheptose can be successfully translocated across the membrane through the MalFGK₂ maltose transporter system.⁽¹¹⁶⁾ In contrast, peptide binding protein OppA did not show a differential conformation sampling in presence of four cognate substrates, suggesting that substrate-induced sampling is not the same in all SBPs. This data supports the hypothesis that even when SBP structures are captured in either open or closed states, a successful interaction between the SBP and the transporter system doesn't depend on a single SBP conformation. Rather, a larger number of conformationally viable states of SBP-ligand complex can interact with the transporter and induce substrate-translocation.⁽¹²⁰⁻¹²²⁾

1.6 Proposed study and its significance

Understanding the role of dynamics and allostery in controlling the substrate affinity in a protein molecule is important towards its optimal function. While studies have been done towards understanding this structure-function relationship in proteins, there remains significant ambiguity in our understanding of how structurally similar protein homologs behave in strikingly disparate ways when they encounter different substrate molecules in varying concentrations in cellular environment. Given the complex physical properties and dynamical behaviors that are attributed to studying such complex systems, we employed a combination of structural, biophysical, and computational techniques to approach this topic in a multifaceted way. We primarily used structure determination techniques like X-ray diffraction and solution scattering coupled with biophysical techniques such as differential scanning calorimetry (DSC) to extract the atomistic details of the system under investigation and to understand its physical properties *in vitro*. We combined this experimentally derived knowledge with computational molecular dynamics (MD) simulations and detected subtle changes (at the atomistic level) in the protein structure

when it interacted with different substrate molecules with varying specificity profiles. Additionally, we also used more sophisticated techniques, like H/D – exchange neutron diffraction to analyze into the non-covalent bonding patterns and solvent-protein interactions of the studied systems with their cognate ligand molecules. Also, a higher order statistical technique called Quasi-anharmonic analysis (QAA) was employed to predict conformational sub-states corresponding to protein motions that promote substrate-binding. All the proposed neutron diffraction, scattering and spectroscopy experiments were carried out at EQ-SANS, BASIS and NSE beamlines at the spallation neutron source (SNS) and at the IMAGINE beamline of the high-flux isotope reactor (HFIR). Additional biochemical analysis and XRD and XRS were performed at the Center for Structural Molecular Biology at SNS.

1.6.1 Specific aims and hypothesis

Central Hypothesis: Molecular motions coupled with binding site-specific redecoration fine-tunes the mechanism of substrate-recognition and binding in primary and secondary transporter associated binding proteins.

Aim 1: To investigate specificity and substrate affinity in the subgroup B-1 PBPs from *T. maritima*

Hypothesis: Selectivity vs. affinity profile in tmMBPs is attributed to the dynamics of conformational sampling and electrostatic potential changes around the binding site.

Approach: We determined the structures of the apo and ligand-bound states of the three homologs of maltose-binding proteins, tmMBP1, tmMBP2 and tmMBP3, from *T. maritima* using X-ray and neutron crystallography and scattering, and all-atom molecular dynamics (MD) simulation.

Aim 2: To capture the dynamics of tmMBPs using experimental neutron scattering techniques

Hypothesis: The differences in the large-scale dynamical hinge motions in tmMBP isoforms is attributed to the variation in their local dynamical behaviors in presence and absence of cognate ligands.

Approach: We will measure the slow dynamics of domain motion using neutron spin-echo (NSE) spectroscopy and the faster vibrational motions using backscattering methods on BASIS. These experimental measurements will be compared with the MD simulations to selectively identify the differences in the dynamical modes between tmMBP isoforms.

Aim 3: To understand the substrate binding profiles of a novel secondary transporter associated binding protein from *P. aeruginosa*

Hypothesis: The TRAP system aids in the uptake of a large variety of similar sized small molecule intermediates of the tricarboxylic acid cycle

Approach: The SBP from *P. aeruginosa* will be expressed, purified and crystallized in presence of different cognate and non-cognate ligands in addition to isothermal titration calorimetric (ITC) data collection to understand its affinity profiles.

1.6.2 Significance of the study

While it is clear how the changes at the amino acid rearrangement can result in entirely different affinity profiles of homologs, there is some evidence to suggest that conformational dynamics impact the overall function of ligand binding and release. Given the importance of these binding proteins in shaping the development of bacterial systems, gaining a deeper understanding of affinity and specificity tuning in these proteins can open avenues for designer microbes with biomass-selective transport systems for getting a higher product titer in biotechnology applications. More generally, the study can also highlight the basis for engineering specificity in other tightly regulated fold spaces leading to a more efficient substrate binding profile. Moreover, the TRAP transporter system is implicated in virulence development in several pathogenic bacteria and understanding the mode of substrate binding in these proteins can aid in the design and development of therapeutic interventions.

List of References

1. Almendro, V., Marusyk, A., and Polyak, K. (2013) Cellular heterogeneity and molecular evolution in cancer. *Annu Rev Pathol* **8**, 277-302
2. Korotkova, N., Chattopadhyay, S., Tabata, T. A., Beskhlebnaya, V., Vigdorovich, V., Kaiser, B. K., Strong, R. K., Dykhuizen, D. E., Sokurenko, E. V., and Moseley, S. L. (2007) Selection for functional diversity drives accumulation of point mutations in Dr adhesins of Escherichia coli. *Mol Microbiol* **64**, 180-194
3. Todd, A. E., Orengo, C. A., and Thornton, J. M. (2001) Evolution of function in protein superfamilies, from a structural perspective. *J Mol Biol* **307**, 1113-1143
4. Echave, J., Spielman, S. J., and Wilke, C. O. (2016) Causes of evolutionary rate variation among protein sites. *Nat Rev Genet* **17**, 109-121
5. Jacq, B. (2001) Protein function from the perspective of molecular interactions and genetic networks. *Briefings in bioinformatics* **2**, 38-50
6. Karp, P. D. (2000) An ontology for biological function based on molecular interactions. *Bioinformatics* **16**, 269-285
7. Schreiber, G., and Keating, A. E. (2011) Protein binding specificity versus promiscuity. *Curr Opin Struct Biol* **21**, 50-61
8. Quiocho, F. A., Wilson, D. K., and Vyas, N. K. (1989) Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature* **340**, 404-407
9. Eaton, B. E., Gold, L., and Zichi, D. A. (1995) Let's get specific: the relationship between specificity and affinity. *Chem Biol* **2**, 633-638
10. Szwajkajzer, D., and Carey, J. (1997) Molecular and biological constraints on ligand-binding affinity and specificity. *Biopolymers* **44**, 181-198
11. Doolittle, R. F., Johnson, M. S., Husain, I., Van Houten, B., Thomas, D. C., and Sancar, A. (1986) Domainal evolution of a prokaryotic DNA repair protein and its relationship to active-transport proteins. *Nature* **323**, 451-453
12. Davidson, A. L., 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, table of contents
13. Rees, D. C., Johnson, E., and Lewinson, O. (2009) ABC transporters: the power to change. *Nat Rev Mol Cell Biol* **10**, 218-227
14. Saier, M. H., Jr. (2016) Transport protein evolution deduced from analysis of sequence, topology and structure. *Curr Opin Struct Biol* **38**, 9-17

15. Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453-1462
16. Isenbarger, T. A., Carr, C. E., Johnson, S. S., Finney, M., Church, G. M., Gilbert, W., Zuber, M. T., and Ruvkun, G. (2008) The most conserved genome segments for life detection on Earth and other planets. *Origins of Life and Evolution of Biospheres* **38**, 517-533
17. Holland, I. B. (2019) Rise and rise of the ABC transporter families. *Res Microbiol* **170**, 304-320
18. Locher, K. P. (2016) Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat Struct Mol Biol* **23**, 487-493
19. Kaur, H., Abreu, B., Akhmetzyanov, D., Lakatos-Karoly, A., Soares, C. M., Prisner, T., and Glaubitz, C. (2018) Unexplored Nucleotide Binding Modes for the ABC Exporter MsbA. *J Am Chem Soc* **140**, 14112-14125
20. Kaur, H., Lakatos-Karoly, A., Vogel, R., Noll, A., Tampe, R., and Glaubitz, C. (2016) Coupled ATPase-adenylate kinase activity in ABC transporters. *Nat Commun* **7**, 13864
21. Randak, C. O., Dong, Q., Ver Heul, A. R., Elcock, A. H., and Welsh, M. J. (2013) ATP and AMP mutually influence their interaction with the ATP-binding cassette (ABC) adenylylate kinase cystic fibrosis transmembrane conductance regulator (CFTR) at separate binding sites. *Journal of Biological Chemistry* **288**, 27692-27701
22. Bhaskara, V., Dupré, A., Lengsfeld, B., Hopkins, B. B., Chan, A., Lee, J.-H., Zhang, X., Gautier, J., Zakian, V., and Paull, T. T. (2007) Rad50 adenylylate kinase activity regulates DNA tethering by Mre11/Rad50 complexes. *Molecular cell* **25**, 647-661
23. Rice, A. J., Park, A., and Pinkett, H. W. (2014) Diversity in ABC transporters: type I, II and III importers. *Crit Rev Biochem Mol Biol* **49**, 426-437
24. Higgins, C. F. (1992) ABC transporters: from microorganisms to man. *Annu Rev Cell Biol* **8**, 67-113

25. Orelle, C., Oldham, M. L., and Davidson, A. L. (2014) The maltose ABC transporter: where structure meets function. in *Membrane transport mechanism*, Springer. pp 181-205
26. Rempel, S., Gati, C., Nijland, M., Thangaratnarajah, C., Karyolaimos, A., de Gier, J. W., Guskov, A., and Slotboom, D. J. (2020) A mycobacterial ABC transporter mediates the uptake of hydrophilic compounds. *Nature* **580**, 409-412
27. Henderson, G. B., and Huennekens, F. (1974) Transport of folate compounds into *Lactobacillus casei*. *Archives of biochemistry and biophysics* **164**, 722-728
28. Henderson, G. B., Zevely, E. M., and Huennekens, F. (1976) Folate transport in *Lactobacillus casei*: solubilization and general properties of the binding protein. *Biochemical and Biophysical Research Communications* **68**, 712-717
29. Henderson, G. B., Zevely, E. M., Kadner, R. J., and Huennekens, F. (1977) The folate and thiamine transport proteins of *Lactobacillus casei*. *Journal of supramolecular structure* **6**, 239-247
30. Rempel, S., Stanek, W. K., and Slotboom, D. J. (2019) ECF-Type ATP-Binding Cassette Transporters. *Annu Rev Biochem* **88**, 551-576
31. Eitinger, T., Rodionov, D. A., Grote, M., and Schneider, E. (2011) Canonical and ECF-type ATP-binding cassette importers in prokaryotes: diversity in modular organization and cellular functions. *FEMS microbiology reviews* **35**, 3-67
32. Karpowich, N. K., Song, J. M., Cocco, N., and Wang, D. N. (2015) ATP binding drives substrate capture in an ECF transporter by a release-and-catch mechanism. *Nat Struct Mol Biol* **22**, 565-571
33. Luckey, M., and Nikaido, H. (1980) Specificity of diffusion channels produced by lambda phage receptor protein of *Escherichia coli*. *Proc Natl Acad Sci U S A* **77**, 167-171
34. Szmecman, S., and Hofnung, M. (1975) Maltose transport in *Escherichia coli* K-12: involvement of the bacteriophage lambda receptor. *J Bacteriol* **124**, 112-118
35. Boos, W., and Shuman, H. (1998) Maltose/maltodextrin system of *Escherichia coli*: transport, metabolism, and regulation. *Microbiol Mol Biol Rev* **62**, 204-229
36. Davidson, A. L., and Nikaido, H. (1991) Purification and characterization of the membrane-associated components of the maltose transport system from *Escherichia coli*. *J Biol Chem* **266**, 8946-8951

37. Oldham, M. L., Khare, D., Quijcho, F. A., Davidson, A. L., and Chen, J. (2007) Crystal structure of a catalytic intermediate of the maltose transporter. *Nature* **450**, 515-521
38. Duplay, P., Bedouelle, H., Fowler, A., Zabin, I., Saurin, W., and Hofnung, M. (1984) Sequences of the malE gene and of its product, the maltose-binding protein of Escherichia coli K12. *J Biol Chem* **259**, 10606-10613
39. Sharff, A. J., Rodseth, L. E., Spurlino, J. C., and Quijcho, F. A. (1992) Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry* **31**, 10657-10663
40. Spurlino, J. C., Lu, G. Y., and Quijcho, F. A. (1991) The 2.3-A resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. *J Biol Chem* **266**, 5202-5219
41. Ferenci, T., Muir, M., Lee, K.-S., and Maris, D. (1986) Substrate specificity of the Escherichia coli maltodextrin transport system and its component proteins. *Biochimica Et Biophysica Acta (BBA)-Biomembranes* **860**, 44-50
42. Quijcho, F. A., Spurlino, J. C., and Rodseth, L. E. (1997) Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure* **5**, 997-1015
43. Hall, J. A., Thorgeirsson, T. E., Liu, J., Shin, Y. K., and Nikaido, H. (1997) Two modes of ligand binding in maltose-binding protein of Escherichia coli. Electron paramagnetic resonance study of ligand-induced global conformational changes by site-directed spin labeling. *J Biol Chem* **272**, 17610-17614
44. Sharff, A. J., Rodseth, L. E., and Quijcho, F. A. (1993) Refined 1.8-A structure reveals the mode of binding of beta-cyclodextrin to the maltodextrin binding protein. *Biochemistry* **32**, 10553-10559
45. Skrynnikov, N. R., Goto, N. K., Yang, D., Choy, W.-Y., Tolman, J. R., Mueller, G. A., and Kay, L. E. (2000) Orienting domains in proteins using dipolar couplings measured by liquid-state NMR: differences in solution and crystal forms of maltodextrin binding protein loaded with β -cyclodextrin. *Journal of molecular biology* **295**, 1265-1273

46. Oldham, M. L., Chen, S., and Chen, J. (2013) Structural basis for substrate specificity in the Escherichia coli maltose transport system. *Proc Natl Acad Sci U S A* **110**, 18132-18137
47. Yu, J., Ge, J., Heuveling, J., Schneider, E., and Yang, M. (2015) Structural basis for substrate specificity of an amino acid ABC transporter. *Proc Natl Acad Sci U S A* **112**, 5243-5248
48. Woo, J. S., Zeltina, A., Goetz, B. A., and Locher, K. P. (2012) X-ray structure of the Yersinia pestis heme transporter HmuUV. *Nat Struct Mol Biol* **19**, 1310-1315
49. Pinkett, H. W., Lee, A. T., Lum, P., Locher, K. P., and Rees, D. C. (2007) An inward-facing conformation of a putative metal-chelate-type ABC transporter. *Science* **315**, 373-377
50. Locher, K. P., Lee, A. T., and Rees, D. C. (2002) The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* **296**, 1091-1098
51. Davidson, A. L., and Nikaido, H. (1990) Overproduction, solubilization, and reconstitution of the maltose transport system from Escherichia coli. *J Biol Chem* **265**, 4254-4260
52. Forward, J. A., Behrendt, M. C., Wyborn, N. R., Cross, R., and Kelly, D. J. (1997) TRAP transporters: a new family of periplasmic solute transport systems encoded by the dctPQM genes of Rhodobacter capsulatus and by homologs in diverse gram-negative bacteria. *Journal of bacteriology* **179**, 5482-5493
53. Rosa, L. T., Bianconi, M. E., Thomas, G. H., and Kelly, D. J. (2018) Tripartite ATP-independent periplasmic (TRAP) transporters and tripartite tricarboxylate transporters (TTT): from uptake to pathogenicity. *Frontiers in cellular and infection microbiology* **8**, 33
54. Thomas, G. H. (2016) Sialic acid acquisition in bacteria—one substrate, many transporters. *Biochemical Society Transactions* **44**, 760-765
55. Mulligan, C., Fischer, M., and Thomas, G. H. (2010) Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS microbiology reviews* **35**, 68-86

56. Post, D. M., Mungur, R., Gibson, B. W., and Munson, R. S. (2005) Identification of a novel sialic acid transporter in *Haemophilus ducreyi*. *Infection and immunity* **73**, 6727-6735
57. Gruteser, N., Marin, K., Krämer, R., and Thomas, G. H. (2012) Sialic acid utilization by the soil bacterium *Corynebacterium glutamicum*. *FEMS microbiology letters* **336**, 131-138
58. Kelly, D. J., and Thomas, G. H. (2001) The tripartite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea. *FEMS microbiology reviews* **25**, 405-424
59. Poolman, B., and Konings, W. N. (1993) Secondary solute transport in bacteria. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1183**, 5-39
60. Mulligan, C., Geertsma, E. R., Severi, E., Kelly, D. J., Poolman, B., and Thomas, G. H. (2009) The substrate-binding protein imposes directionality on an electrochemical sodium gradient-driven TRAP transporter. *Proceedings of the National Academy of Sciences* **106**, 1778-1783
61. Mulligan, C., Leech, A. P., Kelly, D. J., and Thomas, G. H. (2012) The membrane proteins SiaQ and SiaM form an essential stoichiometric complex in the sialic acid tripartite ATP-independent periplasmic (TRAP) transporter SiaPQM (VC1777–1779) from *Vibrio cholerae*. *Journal of Biological Chemistry* **287**, 3598-3608
62. Almagro-Moreno, S., and Boyd, E. F. (2009) Sialic acid catabolism confers a competitive advantage to pathogenic *Vibrio cholerae* in the mouse intestine. *Infection and immunity* **77**, 3807-3816
63. Chowdhury, N., Norris, J., McAlister, E., Lau, S. K., Thomas, G. H., and Boyd, E. F. (2012) The VC1777–VC1779 proteins are members of a sialic acid-specific subfamily of TRAP transporters (SiaPQM) and constitute the sole route of sialic acid uptake in the human pathogen *Vibrio cholerae*. *Microbiology* **158**, 2158-2167
64. Berntsson, R. P.-A., Smits, S. H., Schmitt, L., Slotboom, D.-J., and Poolman, B. (2010) A structural classification of substrate-binding proteins. *FEBS letters* **584**, 2606-2617
65. Scheepers, G. H., Lycklama a Nijeholt, J. A., and Poolman, B. (2016) An updated structural classification of substrate-binding proteins. *FEBS letters* **590**, 4393-4401

66. Wyborn, N. R., Alderson, J., Andrews, S. C., and Kelly, D. J. (2001) Topological analysis of DctQ, the small integral membrane protein of the C4-dicarboxylate TRAP transporter of *Rhodobacter capsulatus*. *FEMS microbiology letters* **194**, 13-17
67. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y., and Gouaux, E. (2005) Crystal structure of a bacterial homologue of Na⁺/Cl⁻-dependent neurotransmitter transporters. *Nature* **437**, 215
68. Nie, R., Stark, S., Symersky, J., Kaplan, R. S., and Lu, M. (2017) Structure and function of the divalent anion/Na⁺ symporter from *Vibrio cholerae* and a humanized variant. *Nature Communications* **8**
69. Mitchell, P. (1957) A general theory of membrane transport from studies of bacteria. *Nature* **180**, 134-136
70. Jardetzky, O. (1966) Simple allosteric model for membrane pumps. *Nature* **211**, 969-970
71. Mulligan, C., Fischer, M., and Thomas, G. H. (2011) Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS microbiology reviews* **35**, 68-86
72. Khare, D., Oldham, M. L., Orelle, C., Davidson, A. L., and Chen, J. (2009) Alternating access in maltose transporter mediated by rigid-body rotations. *Mol Cell* **33**, 528-536
73. Oldham, M. L., and Chen, J. (2011) Crystal structure of the maltose transporter in a pretranslocation intermediate state. *Science* **332**, 1202-1205
74. Jacso, T., Grote, M., Daus, M. L., Schmieder, P., Keller, S., Schneider, E., and Reif, B. (2009) Periplasmic loop P2 of the MalF subunit of the maltose ATP binding cassette transporter is sufficient to bind the maltose binding protein MalE. *Biochemistry* **48**, 2216-2225
75. Austermuhle, M. I., Hall, J. A., Klug, C. S., and Davidson, A. L. (2004) Maltose-binding protein is open in the catalytic transition state for ATP hydrolysis during maltose transport. *Journal of Biological Chemistry* **279**, 28243-28250
76. Chen, J., Sharma, S., Quiocho, F. A., and Davidson, A. L. (2001) Trapping the transition state of an ATP-binding cassette transporter: evidence for a concerted

- mechanism of maltose transport. *Proceedings of the National Academy of Sciences* **98**, 1525-1530
77. Khare, D., Oldham, M. L., Orelle, C., Davidson, A. L., and Chen, J. (2009) Alternating access in maltose transporter mediated by rigid-body rotations. *Molecular cell* **33**, 528-536
 78. Mulligan, C., Geertsma, E. R., Severi, E., Kelly, D. J., Poolman, B., and Thomas, G. H. (2009) The substrate-binding protein imposes directionality on an electrochemical sodium gradient-driven TRAP transporter. *Proc Natl Acad Sci U S A* **106**, 1778-1783
 79. Ferenci, T. (1996) Adaptation to life at micromolar nutrient levels: the regulation of *Escherichia coli* glucose transport by endoinduction and cAMP. *FEMS microbiology reviews* **18**, 301-317
 80. Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., and Mayhew, G. F. (1997) The complete genome sequence of *Escherichia coli* K-12. *science* **277**, 1453-1462
 81. Ames, G. F. (1986) Bacterial periplasmic transport systems: structure, mechanism, and evolution. *Annu Rev Biochem* **55**, 397-425
 82. Fukami-Kobayashi, K., Tateno, Y., and Nishikawa, K. (1999) Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *Journal of molecular biology* **286**, 279-290
 83. Berntsson, R. P., Smits, S. H., Schmitt, L., Slotboom, D. J., and Poolman, B. (2010) A structural classification of substrate-binding proteins. *FEBS Lett* **584**, 2606-2617
 84. Scheepers, G. H., Lycklama, A. N. J. A., and Poolman, B. (2016) An updated structural classification of substrate-binding proteins. *FEBS Lett* **590**, 4393-4401
 85. Yatsunyk, L. A., Easton, J. A., Kim, L. R., Sugarbaker, S. A., Bennett, B., Breece, R. M., Vorontsov, II, Tierney, D. L., Crowder, M. W., and Rosenzweig, A. C. (2008) Structure and metal binding properties of ZnuA, a periplasmic zinc transporter from *Escherichia coli*. *J Biol Inorg Chem* **13**, 271-288
 86. Shi, R., Proteau, A., Wagner, J., Cui, Q., Purisima, E. O., Matte, A., and Cygler, M. (2009) Trapping open and closed forms of FitE: a group III periplasmic binding protein. *Proteins* **75**, 598-609

87. Sooriyaarachchi, S., Ubhayasekera, W., Park, C., and Mowbray, S. L. (2010) Conformational changes and ligand recognition of Escherichia coli D-xylose binding protein revealed. *J Mol Biol* **402**, 657-668
88. Clayton, A., Siebold, C., Gilbert, R. J., Sutton, G. C., Harlos, K., McIlhinney, R. J., Jones, E. Y., and Aricescu, A. R. (2009) Crystal structure of the GluR2 amino-terminal domain provides insights into the architecture and assembly of ionotropic glutamate receptors. *Journal of molecular biology* **392**, 1125-1132
89. Tan, K., Chang, C., Cuff, M., Osipiuk, J., Landorf, E., Mack, J. C., Zerbs, S., Joachimiak, A., and Collart, F. R. (2013) Structural and functional characterization of solute binding proteins for aromatic compounds derived from lignin: p-coumaric acid and related aromatic acids. *Proteins* **81**, 1709-1726
90. He, X., Chow, D., Martick, M. M., and Garcia, K. C. (2001) Allosteric activation of a spring-loaded natriuretic peptide receptor dimer by hormone. *Science* **293**, 1657-1662
91. Heddle, J., Scott, D. J., Unzai, S., Park, S. Y., and Tame, J. R. (2003) Crystal structures of the liganded and unliganded nickel-binding protein NikA from Escherichia coli. *J Biol Chem* **278**, 50322-50329
92. Berntsson, R. P., Doeven, M. K., Fusetti, F., Duurkens, R. H., Sengupta, D., Marrink, S. J., Thunnissen, A. M., Poolman, B., and Slotboom, D. J. (2009) The structural basis for peptide selection by the transport receptor OppA. *EMBO J* **28**, 1332-1340
93. Wang, Z., Luecke, H., Yao, N., and Quiocho, F. A. (1997) A low energy short hydrogen bond in very high resolution structures of protein receptor-phosphate complexes. *Nature structural biology* **4**, 519-522
94. Hollenstein, K., Comellas-Bigler, M., Bevers, L. E., Feiters, M. C., Meyer-Klaucke, W., Hagedoorn, P. L., and Locher, K. P. (2009) Distorted octahedral coordination of tungstate in a subfamily of specific binding proteins. *J Biol Inorg Chem* **14**, 663-672
95. Bruns, C. M., Nowalk, A. J., Arvai, A. S., McTigue, M. A., Vaughan, K. G., Mietzner, T. A., and McRee, D. E. (1997) Structure of Haemophilus influenzae Fe³⁺-binding protein reveals convergent evolution within a superfamily. *Nature structural biology* **4**, 919-924

96. Johnston, J. W., Coussens, N. P., Allen, S., Houtman, J. C., Turner, K. H., Zaleski, A., Ramaswamy, S., Gibson, B. W., and Apicella, M. A. (2008) Characterization of the N-acetyl-5-neuraminic acid-binding site of the extracytoplasmic solute receptor (SiaP) of nontypeable *Haemophilus influenzae* strain 2019. *J Biol Chem* **283**, 855-865
97. Huvent, I., Belrhali, H., Antoine, R., Bompard, C., Loch, C., Jacob-Dubuisson, F., and Villeret, V. (2006) Structural analysis of *Bordetella pertussis* BugE solute receptor in a bound conformation. *Acta Crystallogr D Biol Crystallogr* **62**, 1375-1381
98. Huvent, I., Belrhali, H., Antoine, R., Bompard, C., Loch, C., Jacob-Dubuisson, F., and Villeret, V. (2006) Crystal structure of *Bordetella pertussis* BugD solute receptor unveils the basis of ligand binding in a new family of periplasmic binding proteins. *J Mol Biol* **356**, 1014-1026
99. Koropatkin, N. M., Pakrasi, H. B., and Smith, T. J. (2006) Atomic structure of a nitrate-binding protein crucial for photosynthetic productivity. *Proc Natl Acad Sci U S A* **103**, 9820-9825
100. Nguyen, P. T., Li, Q. W., Kadaba, N. S., Lai, J. Y., Yang, J. G., and Rees, D. C. (2015) The contribution of methionine to the stability of the *Escherichia coli* MetNIQ ABC transporter-substrate binding protein complex. *Biol Chem* **396**, 1127-1134
101. Schiefner, A., Breed, J., Bösner, L., Kneip, S., Gade, J., Holtmann, G., Diederichs, K., Welte, W., and Bremer, E. (2004) Cation- π interactions as determinants for binding of the compatible solutes glycine betaine and proline betaine by the periplasmic ligand-binding protein ProX from *Escherichia coli*. *Journal of Biological Chemistry* **279**, 5588-5596
102. Ruggiero, A., Dattelbaum, J. D., Staiano, M., Berisio, R., D'Auria, S., and Vitagliano, L. (2014) A loose domain swapping organization confers a remarkable stability to the dimeric structure of the arginine binding protein from *Thermotoga maritima*. *PLoS One* **9**, e96560
103. Momma, K., Mishima, Y., Hashimoto, W., Mikami, B., and Murata, K. (2005) Direct evidence for *Sphingomonas* sp. A1 periplasmic proteins as macromolecule-binding proteins associated with the ABC transporter: molecular insights into alginate transport in the periplasm. *Biochemistry* **44**, 5053-5064

104. Mishima, Y., Momma, K., Hashimoto, W., Mikami, B., and Murata, K. (2003) Crystal structure of AlgQ2, a macromolecule (alginate)-binding protein of *Sphingomonas* sp. A1, complexed with an alginate tetrasaccharide at 1.6-Å resolution. *Journal of Biological Chemistry* **278**, 6552-6559
105. Bucher, D., Grant, B. J., Markwick, P. R., and McCammon, J. A. (2011) Accessing a hidden conformation of the maltose binding protein using accelerated molecular dynamics. *PLoS Comput Biol* **7**, e1002034
106. Shilton, B. H., Flocco, M. M., Nilsson, M., and Mowbray, S. L. (1996) Conformational changes of three periplasmic receptors for bacterial chemotaxis and transport: the maltose-, glucose/galactose- and ribose-binding proteins. *J Mol Biol* **264**, 350-363
107. Quijcho, F. A., and Ledvina, P. S. (1996) Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol Microbiol* **20**, 17-25
108. Karpowich, N. K., Huang, H. H., Smith, P. C., and Hunt, J. F. (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
109. Trakhanov, S., Vyas, N. K., Luecke, H., Kristensen, D. M., Ma, J., and Quijcho, F. A. (2005) Ligand-free and -bound structures of the binding protein (LivJ) of the *Escherichia coli* ABC leucine/isoleucine/valine transport system: trajectory and dynamics of the interdomain rotation and ligand specificity. *Biochemistry* **44**, 6597-6608
110. Shukla, S., Bafna, K., Gullett, C., Myles, D. A. A., Agarwal, P. K., and Cuneo, M. J. (2018) Differential Substrate Recognition by Maltose Binding Proteins Influenced by Structure and Dynamics. *Biochemistry* **57**, 5864-5876
111. Flocco, M., and Mowbray, S. L. (1994) The 1.9 Å x-ray structure of a closed unliganded form of the periplasmic glucose/galactose receptor from *Salmonella typhimurium*. *J Biol Chem* **269**, 8931-8936
112. Oswald, C., Smits, S. H., Hoing, M., Sohn-Bosser, L., Dupont, L., Le Rudulier, D., Schmitt, L., and Bremer, E. (2008) Crystal structures of the choline/acetylcholine

- substrate-binding protein ChoX from *Sinorhizobium meliloti* in the liganded and unliganded-closed states. *J Biol Chem* **283**, 32848-32859
113. Tang, C., Schwieters, C. D., and Clore, G. M. (2007) Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449**, 1078-1082
 114. Feng, Y., Zhang, L., Wu, S., Liu, Z., Gao, X., Zhang, X., Liu, M., Liu, J., Huang, X., and Wang, W. (2016) Conformational Dynamics of apo-GlnBP Revealed by Experimental and Computational Analysis. *Angew Chem Int Ed Engl* **55**, 13990-13994
 115. Feng, Y., Zhang, L., Wu, S., Liu, Z., Gao, X., Zhang, X., Liu, M., Liu, J., Huang, X., and Wang, W. (2020) Corrigendum: Conformational Dynamics of apo-GlnBP Revealed by Experimental and Computational Analysis. *Angew Chem Int Ed Engl* **59**, 2144
 116. de Boer, M., Gouridis, G., Vietrov, R., Begg, S. L., Schuurman-Wolters, G. K., Husada, F., Eleftheriadis, N., Poolman, B., McDevitt, C. A., and Cordes, T. (2019) Conformational and dynamic plasticity in substrate-binding proteins underlies selective transport in ABC importers. *Elife* **8**
 117. Couñago, R. M., Ween, M. P., Begg, S. L., Bajaj, M., Zuegg, J., O'mara, M. L., Cooper, M. A., McEwan, A. G., Paton, J. C., and Kobe, B. (2014) Imperfect coordination chemistry facilitates metal ion release in the Psa permease. *Nature chemical biology* **10**, 35-41
 118. de Boer, M., Cordes, T., and Poolman, B. (2020) Kinetic Modelling of Transport Inhibition by Substrates in ABC Importers. *Journal of Molecular Biology* **432**, 5565-5576
 119. Ferenci, T. (1980) The recognition of maltodextrins by *Escherichia coli*. *Eur J Biochem* **108**, 631-636
 120. Fulyani, F., Schuurman-Wolters, G. K., Slotboom, D. J., and Poolman, B. (2016) Relative Rates of Amino Acid Import via the ABC Transporter GlnPQ Determine the Growth Performance of *Lactococcus lactis*. *J Bacteriol* **198**, 477-485
 121. Wolters, J. C., Berntsson, R. P., Gul, N., Karasawa, A., Thunnissen, A. M., Slotboom, D. J., and Poolman, B. (2010) Ligand binding and crystal structures of the substrate-binding domain of the ABC transporter OpuA. *PLoS One* **5**, e10361

122. Doeven, M. K., Abele, R., Tampe, R., and Poolman, B. (2004) The binding specificity of OppA determines the selectivity of the oligopeptide ATP-binding cassette transporter. *J Biol Chem* **279**, 32301-32307

Chapter 2 Differential substrate recognition by maltose binding proteins influenced by structure and dynamics

This chapter is a slightly revised version of a manuscript by the same title published in the journal, *Biochemistry* published in 2018

Reprinted (adapted) with permission from Shantanu Shukla,^{1,2} Khushboo Bafna,¹ Caeley Gullett,² Dean A. A. Myles,^{1,2} Pratul K. Agarwal,^{3,*} Matthew J. Cuneo^{2,4,*}

¹Graduate School of Genome Science and Technology, The University of Tennessee, Knoxville, Tennessee

²Neutron Sciences Directorate, Oak Ridge National Laboratory, Oak Ridge, Tennessee

³Department of Biochemistry & Cellular and Molecular Biology, The University of Tennessee, Knoxville, Tennessee

⁴Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, Tennessee

*Corresponding authors: pratul@agarwal-lab.org , matt.cuneo@stjude.org

This author contributed the following to the manuscript: (1) X-tal structure determination, (2) molecular dynamics simulations, (3) Small angle Scattering (4) data analysis and organization and (5) Manuscript writing.

2.1 Abstract

The genome of the hyperthermophile *Thermotoga maritima* contains three isoforms of maltose binding protein (MBP) that are high affinity receptors for di-, tri- and tetra-saccharides. Two of these proteins (tmMBP1 and tmMBP2) share significant sequence identity, approximately 90%, while the third (tmMBP3) shares less than 40% identity. MBP from *Escherichia coli* (ecMBP) shares 35% sequence identity with the tmMBPs. This subset of MBP isoforms offer an interesting opportunity to investigate the mechanisms underlying the evolution of substrate specificity and affinity profiles in a genome where redundant MBP genes are present. In this study, the X-ray crystal structures of tmMBP1, tmMBP2 and tmMBP3 are reported in absence and presence of oligosaccharides. tmMBP1 and tmMBP2 show larger binding pockets than tmMBP3 enabling them to bind to larger substrates, while tmMBP1 and tmMBP2 also undergo larger substrate induced hinge bending motions ($\sim 55^\circ$) than tmMBP3 ($\sim 35^\circ$). Small angle X-ray scattering was used to compare protein behavior in solution and computer simulations provided insights into dynamics of these proteins. Comparing quantitative protein-substrate interactions and dynamical properties of tmMBPs with the promiscuous ecMBP and di-saccharide selective *Thermococcus litoralis* MBP provide insights into the features that enable selective binding. Collectively, the results provide insights into how the structure and dynamics of tmMBP homologs enable them to differentiate between a myriad of chemical entities whilst maintaining their common fold.

2.2 Introduction

Bacteria utilize diverse mechanisms to survive in challenging environmental niches, which include borrowing and duplicating genes.⁽¹⁻³⁾ This provides functional robustness to bacterial genomes by allowing the duplicated components to compensate for any loss of function due to changes in the original gene.⁽⁴⁾ Operon redundancy can also result in function segregation through specialization of protein isoforms that share significant sequence similarity.⁽⁵⁻⁷⁾ As selection pressure typically removes redundancy, the duplicated genes frequently diverge and vary their roles in bacterial survival, thus escaping evolutionary scrutiny.^(8,9) Even when there is reduced sequence identity, such homologs

could maintain the same structural fold, allowing the organism to increase functional diversification with minimal structural changes.⁽¹⁰⁻¹²⁾

Periplasmic binding proteins constitute a large protein super-family commonly involved in substrate binding and translocation of metabolites across the bacterial membrane. In such a large super-family, homologous proteins show conservation in their overall structural fold even with low sequence similarities. The members of this super-family bind a myriad of metabolites; their binding affinities are highly tuned to the environmental niche of the host organism, resulting in differing affinity profiles for identical ligands.⁽¹³⁻¹⁶⁾ Adjustment of protein structure, especially around the binding pocket, is known to play a role in recognition of ligand(s) in periplasmic binding proteins;^(17,18) additionally conformational dynamics can also help fine-tune affinity and specificity by altering the thermodynamics and kinetics of ligand binding in these systems.⁽¹⁸⁾ While periplasmic binding proteins are subjected to the constraints of the common structural fold, these proteins manage a wide spectrum of specificity profiles by tailoring the non-bonded interactions in and around their binding sites.^(13,19) Moreover, it has also been shown that some periplasmic binding proteins contain inbuilt allosteric mechanisms, which are evolutionarily selected to modulate the affinity profiles of these proteins to their preferred ligands by controlling their conformational switching between the apo and ligand bound states,⁽²⁰⁻²²⁾ this property can also be engineered in these proteins to alter their ligand affinities.⁽²³⁾ Therefore, periplasmic binding proteins can provide unique insights into how proteins specialize in their designated activities while being subjected to the constraints of a structural fold. They can also aid in understanding the process of transport associated molecular recognition, which is vital for bacterial physiology and survival.

Members of the periplasmic binding protein super-family operate in conjunction with adenosine tri-phosphate binding cassette (ABC) transport systems, where they bind a diverse array of metabolites in the cellular milieu.^(24,25) Periplasmic binding proteins have a two domain architecture with the N- and C-terminal domains linked by a flexible linker region that serves as a hinge.^(15,26) The substrate binding site resides at the interface of these two domains, and the hinge is responsible for the characteristic ligand dependent domain motions of this super-family.^(13,15) It has been proposed that the location of the

substrate binding site within this flexible hinged interface, allows the evolution of flexibility in these systems without affecting the overall fold.⁽²⁷⁾ While it is not clear how the changes at the amino acid level result in entirely different affinity profiles of homologs, there is some evidence to suggest that conformational dynamics impact the overall function of ligand binding and release.^(18,28,29) Based on these findings, it is hypothesized that the modification of the binding site and the changes in the overall dynamics of the system is a potential mechanism to encode functional diversity in this protein fold, beyond what can be attained by simply tuning molecular recognition through structural interactions in the binding pocket.

Maltose binding proteins (MBPs) are ubiquitous periplasmic binding proteins that serve as the primary receptors for alpha-linked glucose based oligosaccharides.⁽¹⁹⁾ While many bacteria have only a single copy of a more promiscuous MBP, some organisms contain more than one MBP isoform.⁽³⁰⁾ The hyperthermophile *Thermotoga maritima* was recently shown to have three isoforms of MBP ⁽³¹⁾ (tmMBP1-3) that are high affinity receptors for di-, tri- and tetra-saccharides respectively. ⁽³²⁾ tmMBP1 (391 residues) and tmMBP2 (393 residues) share significant sequence identity (approximately 90%) but tmMBP3 (411 residues) shares less than 40% identity. The tmMBPs share roughly 35% sequence identity with the *Escherichia coli* MBP (ecMBP). ⁽³²⁾ Thus, this cluster of MBP isoforms offer an interesting opportunity to investigate the mechanisms underlying the evolution of specificity and affinity profiles in a genome where MBP gene redundancy is present.

In this study, the X-ray crystal structures of tmMBP1, tmMBP2 and tmMBP3 in absence and presence of substrates have been solved, to understand the structural basis of specificity encoded in this subset of proteins. To account for the contribution of protein dynamics in substrate recognition, which are not very well captured in the static X-ray crystal structures, solution-based small angle X-ray scattering, and microsecond time-scale (500 nanoseconds) computer simulations and theoretical modeling were also performed on these systems. A strategy of comparing structural and dynamical properties of the three tmMBP isoforms to the promiscuous ecMBP and the di-saccharide selective MBP from *Thermococcus litoralis* (tlMBP) was used to characterize features of promiscuous versus non-promiscuous binding by periplasmic binding proteins. This

comparison was further aided by quantitative estimates of MBP-substrate interactions obtained using theoretical calculations. Further the results indicate that the structure and dynamics of these homologs and their substrate interaction profiles enable them to differentiate between myriad chemical entities whilst maintaining their common fold.

2.3 Materials and Methods

Protein expression and purification

tmMBP proteins were purified using the published protocol with some minor changes.⁽¹⁶⁾ Briefly, the hexahistidine-tagged tmMBP isoforms were cloned in a pET21a vector and expressed in BL21 RIL *E.coli* cells. The apo form of the proteins was produced using Enfor's minimal media to avoid binding of endogenous carbohydrates from the medium. The substrate bound form of the proteins was grown in Terrific Broth. Both media were incubated at 37° C and induced using 1 mM IPTG. The induced cells were harvested and lysed using sonication before an affinity purification step (HiTrap Chelating HP column). The column was washed with affinity buffer A (20 mM imidazole, 300 mM NaCl) and subsequently with affinity buffer B (75 mM Imidazole and 150 mM NaCl) and the washes were collected. Finally, elution buffer (300 mM Imidazole, 150 mM NaCl) was used to elute the bound protein from the affinity column. Eluted samples were subsequently loaded onto a S75 26/60 gel filtration column (GE Healthcare Life Sciences). The protein fractions were examined using SDS page and the fractions containing tmMBP protein were pooled and concentrated. The concentrated proteins were dialyzed in 20 mM Tris (pH 7.5) and 40 mM NaCl overnight at room temperature.

Crystallization and X-ray data collection

Dialyzed tmMBP proteins (both apo and holo forms) were concentrated to 15 mg/ml for crystallization. Crystals were grown by hanging drop vapor diffusion with a drop size of 2 μ L for each of the protein and the mother liquor and equilibrated against 900 μ L of the mother liquor. Maltotetraose (MTT) bound tmMBP1 was crystallized in 25% poly-ethylene glycol (PEG) 4000, 18% isopropanol and 0.1M sodium citrate. MTT bound tmMBP2 was crystallized in 23% PEG 3350, 0.3 M sodium acetate and 0.1 M bis-tris pH 5.5. Crystals of the maltose (MAL) bound tmMBP3 was obtained in 9% PEG 3350 supplemented with 0.4 M sodium thiocyanate. Crystals for apo tmMBP2 grew in 26% PEG 3350 and 0.1 M

sodium diphosphate. The apo form of the tmMBP3 was crystallized in 20% PEG 4000 and 0.2 M sodium sulfate. tmMBP2 apo and tmMBP2-MTT crystals were transferred to 35% (wt./vol.) PEG 3350 for cryoprotection, mounted in a nylon loop, and flash frozen in liquid nitrogen. X-ray datasets for tmMBP1-MTT, tmMBP3-MAL and tmMBP3 apo were collected at room temperature on a Rigaku 007HFmicromax X-ray generator with a Raxis IV++ detector. Scaling and integration of the X-ray data were carried out using HKL3000.⁽³³⁾ The structures of apo tmMBP2 and tmMBP3 were determined to a resolution of 1.90 Å and 2.30 Å, respectively. Structures of tmMBP1 and tmMBP2 bound to maltotetraose were determined at 1.50 Å each. Maltose bound tmMBP3 structure was refined to a resolution of 2.15 Å. Atomic coordinates and structure factors for the following systems have been deposited in the Protein Data Bank (PDB): apo tmMBP2 (accession code 6DTT), apo tmMBP3 (6DTR), tmMBP1-maltotetraose complex (6DTU), tmMBP2-maltotetraose complex (6DTS) and tmMBP3-maltose complex (6DTQ).

Small-angle X-ray scattering

The dialyzed proteins were concentrated to 8 mg/ml. Data were collected at 50° C and in the presence of 1mM ligand for the holo forms of the protein. Circular averaging of the scattering data was used to determine the relative scattering intensity (I) as a function of momentum transfer vector, q ($q = 4\pi\sin\theta/\lambda$). Data were buffer subtracted prior to analysis. The solution scattering data were analyzed using the ATSAS software package;⁽³⁴⁾ the GNOM software package⁽³⁵⁾ was used for all $P(r)$ and I_o analyses. The three-dimensional envelope was generated from the SAXS data using the GASBOR program.⁽³⁶⁾ Each calculation was repeated at least ten times with different random starting points for the simulated annealing algorithm; no predefined shape or symmetry constraints were used. Subsequently, these 10 model structures were averaged using DAMAVER.⁽³⁷⁾ These models were used to generate an average ensemble, which was subsequently used to generate the envelope structure. To superimpose the modeled envelope on to the PDB structure, the SUPCOMB program was used.⁽³⁸⁾ The experimental scattering data was fitted on to the simulated scattering data from the crystal structure using CRY SOL.⁽³⁹⁾

Computational Modeling

Molecular dynamics (MD) simulations were performed to model MBPs in complex with various oligosaccharides in explicit water solvent. Model preparation and simulations were performed using the AMBER v14 suite of programs for biomolecular simulations.⁽⁴⁰⁾ AMBER's *ff14SB(41)* and *GLYCAM 06Epb* (for sugar units)⁽⁴²⁾ force-fields were used for all simulations. MD simulations were performed using NVIDIA graphical processing units (GPUs) and AMBER's *pmemd.cuda* simulation engine using our lab protocols published previously.^(43,44)

A total of 21 separate simulations were performed, based on the crystal structures determined in this study, and previously solved crystal structures, as well as computationally modeled protein-substrate complexes (Table 2.1). The alternate substrates, where coordinates were not available from X-ray structures, were modeled based on other crystal structures available for the protein with other substrates. These alternate substrates were modeled by deleting one or more of the sugar units present in the available crystal structure and the missing hydrogen atoms were added by AMBER's *tleap* program. After processing the coordinates of the protein and substrate, all systems were neutralized by addition of counter-ions and the resulting system were solvated in a rectangular box of SPC/E water, with a 10 Å minimum distance between the protein and the edge of the periodic box. The prepared systems were equilibrated using a protocol described previously.⁽⁴⁵⁾ The equilibrated systems were then used to run 0.5 μs of production MD under constant energy conditions (NVE ensemble). The use of NVE ensemble is preferred as it offers better computational stability and performance,⁽⁴⁶⁾ and as other ensembles (constant temperature or constant pressure) could affect dynamics of the system due to coordinates/velocities scaling. The production simulations were performed at a temperature of 325 K to match the conditions of SAXS experiments. As NVE ensemble was used for production runs, these values correspond to initial temperature at start of simulations. Temperature adjusting thermostat was not used in simulations; over the course of 0.5 μs simulations the temperature fluctuated around 325 K with RMS fluctuations between 2-4 K, which is typical for well equilibrated systems. A total of 10,000 conformational snapshots (stored every 50 ps) collected for each system (apo and substrate bound) was used for analysis.

Table 2.1: Starting X-ray coordinates for MD simulations. PDB codes or other source used for modeling are noted.

	tmMBP1	tmMBP2	tmMBP3	ecMBP	tIMBP
<i>Apo</i>	2GHB	Crystallized in this study	Crystallized in this study	1JW4	
<i>Glucose</i>			modeled ^a	modeled ^b	
<i>Maltose</i>	modeled ^c	modeled ^d	Crystallized in this study	1ANF	
<i>Trehalose</i>					1EU8
<i>Maltotriose</i>	2GHA	2FN8		3MBP	
<i>Maltotetraose</i>	Crystallized in this study	Crystallized in this study		4MBP	

^a To explore alternate binding sites, 2 alternate simulations were done for tMBP3-glucose (referred as tmMBP3-GLU1 and tmMBP3-GLU2)

^b To explore alternate binding sites, 2 alternate simulations were done for ecMBP-glucose (referred as ecMBP-GLU1 and ecMBP-GLU2)

^c To explore alternate binding sites, 2 alternate simulations were done for tMBP1-maltose (referred as tmMBP1-MAL1 and tmMBP1-MAL2)

^d To explore alternate binding sites, 2 alternate simulations were done for tMBP1-maltose (referred as tmMBP2-MAL1 and tmMBP2-MAL2)

RMSF and RMSF₁₀ calculations

All atoms root mean square fluctuations (RMSF) were computed based on the conformational snapshots collected during the MD simulations. To identify global motions on slower time-scales from MD, for each of the 21 systems the fluctuations associated with the first (slowest) 10 quasi-harmonic modes (RMSF₁₀) were also computed and aggregated. It is well known that slowest 10 modes contribute to the majority of fluctuations in proteins (>80%) and the use of RMSF₁₀, instead of all modes (RMSF), removes the faster stochastic motions of the protein, allowing focus on intrinsic dynamics of proteins.⁽⁴⁷⁾ Both these calculations were performed using AMBER's *ptraj* analysis program. All trajectory conformations were first aligned to a common structure, to remove any translation and overall molecular rotation during the simulations.

Radius of gyration (R_g)

AMBER's *ptraj* module was used to calculate the R_g for the protein in its apo and substrate bound forms for each stored conformational snapshot.

Protein-substrate interactions: The energy for the enzyme-substrate interactions ($E_{pro-sub}$) were calculated as a sum of electrostatic (E_{el}) and van der Waals energy (E_{vdw}) between atom pairs, based on an approach developed in our group.^(48,49) All protein and substrate atom pairs were included in the calculations and resulting interaction energies were summed up per residue pair. The energies were calculated for 10,000 snapshots, every 50 ps, sampled during the full 0.5 μ s simulation and were averaged over these 10,000 snapshots.

The energy for the enzyme-substrate interactions ($E_{pro-sub}$) were calculated as a sum of electrostatic and van der Waals energy between atom pairs.

$$E_{pro-sub} = \sum (E_{el} + E_{vdw}) \quad (1)$$

E_{el} is the electrostatic contribution, E_{vdw} is the van der Waals term and the summation runs over all atom pairs for the enzyme and substrate. The E_{el} and E_{vdw} terms were computed as follows

$$E_{el} = \frac{q_i q_j}{\epsilon(r) r_{ij}} \quad \text{and} \quad E_{vdw} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \quad (2)$$

where q_i , q_j are partial charges, and A_{ij} , B_{ij} are Lennard-Jones parameters. These

parameters were obtained from AMBER *ff14SB* force field. A distance-dependent dielectric function was used:

$$\varepsilon(r_{ij}) = A + \frac{B}{1 + k \exp(-\lambda B r_{ij})} \quad (3)$$

$B = \varepsilon_o - A$; $\varepsilon_o = 78.4$ for water; $A = -8.5525$; $\lambda = 0.003627$ and $k = 7.7839$.

For calculations of the errors associated of interaction energy values, the sampled conformations were divided into two parts. The first set considered odd numbered frames stored during MD production runs, while the other set considered the even numbered frames. The difference between averaged values for these two sets is considered indicative of the errors associated with these values.

2.4 Results

2.4.1 Overall structure of tmMBPs

A total of five X-ray crystal structures of the *T. maritima* MBPs (tmMBPs) were determined in this study: tmMBP1 bound to maltotetraose; tmMBP2 in apo form; tmMBP2 bound to maltotetraose; tmMBP3 in apo form; and tmMBP3 bound to maltose. The structure of the apo form of tmMBP1 has previously been deposited to the Protein Data Bank (PDB ID: 2GHB). An effort to co-crystallize tmMBP1 and tmMBP2 with disaccharide (maltose) resulted in structures with the apo form of these two proteins, while an attempt to crystallize tmMBP3 with maltotetraose resulted in protein co-crystallized with disaccharide. Crystallographic data collection and refinement statistics are presented in Table 2.2.

All three tmMBPs possess the signature α/β fold characteristic of MBPs,⁽⁵⁰⁾ and the three isoforms form several common hydrogen bonds (H-bonds) and hydrophobic interactions with their substrates (Figure 2.1). tmMBP1 and tmMBP2 have significant structural similarity (RMSD = 0.6 Å) and an essentially identical interaction network with the maltotetraose (Figure 2.1). These two proteins share high sequence similarity (~90%).

tmMBP3 has several amino acid insertions and is larger than the other two isoforms, yet it has a smaller and a more closed off binding pocket. A significant number of water-mediated interactions between the protein and the ligands were observed for tmMBP1 and tmMBP2, however, the smaller and closed off binding site of tmMBP3 reduces the number of H-bonded water molecules to three. There is also a considerable difference in the extent of hinge bending motion in the three isoforms upon substrate binding: tmMBP1 (52.3°) and tmMBP2 (51.2°) showing a hinge movement of >15° more than tmMBP3 (35.3°) (Figure 2.1).

2.4.2 Structural comparison of tmMBPs with *E. coli* MBP (ecMBP) and *Thermococcus litoralis* MBP (tlMBP)

It has been previously reported that the MBP fold contains multiple binding sub-sites in the ligand binding cavity.^(19,51) These sites, termed S1, S2, S3 and S4, each accommodate a single sugar ring, with the reducing end of the sugar being placed in the first occupied subsite. The interactions in the subsites are modulated through residue insertions and deletions (indels) which tune substrate specificity and affinity profiles (Figures 2.2 – 2.4). Identified indels include two loops, termed L1 and L2 and three helical insertions H1, H2 and H3 (Figure 2.3). To characterize the binding of saccharides of different lengths in the conserved sub-sites, the obtained structures of the tmMBPs were compared to the promiscuous *E. coli* maltose binding protein (ecMBP) and the disaccharide specific *T. litoralis* trehalose/maltose binding protein (tlMBP).

In the tmMBPs, the reducing end of the saccharide binds to subsite S1 (Figure 2.2). However, in ecMBP subsite S1 is sterically hindered by two residues found in the L1 loop (Asp14 and Lys15). These residues extend into the S1 subsite and also form H-bonds with the first sugar ring in subsite S2 (Figure 2.2). In the tmMBPs, subsite S1 has a tryptophan (Trp233 in tmMBP1 and tmMBP2, and Trp230 in ecMBP) residue which forms hydrophobic contacts with the substrate. Further, in the case of tmMBP1 and tmMBP2, the subsite S1 also include a conserved arginine (Arg303) and a serine (Ser12) that interact with both tetra- and tri-saccharide substrates (Figure 2.2 and Figure 2.4). Additionally, helix H1 also contains a conserved tyrosine (Tyr213) in tmMBP1 and tmMBP2, which provides for additional hydrophobic interactions with the first sugar ring.

Table 2.2: X-ray data collection and refinement statistics for tmMBPs. A total of five X-ray crystal structures of the *T. maritima* MBPs (tmMBPs) were determined in this study: tmMBP1 bound to maltotetraose, tmMBP2 in its apo form, tmMBP2 bound to maltotetraose, tmMBP3 in its apo form, and tmMBP3 bound to maltose.

	tmMBP1- MTT	apo tmMBP2	tmMBP2- MTT	apo tmMBP3	tmMBP3- MAL
<i>PDB ID</i>	6DTU	6DTT	6DTS	6DTR	6DTQ
Data collection					
Resolution range (Å)	50.0 – 1.5 (1.55 – 1.50)	30.0 – 1.9 (1.97 – 1.90)	50.0 – 1.5 (1.54 – 1.50)	40.0 – 2.3 (2.38 – 2.30)	40.0 – 2.1 (2.23 – 2.15)
Space group	P 1	P 2 ₁ 2 ₁ 2 ₁	P 1	C 2 2 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Unit cell					
a, b and c (Å)	34.6, 55.4, 93.5	78.9, 91.0, 100.2	35.4, 56.1, 88.3	65.5, 137.4, 145.0	82.3, 124, 174.8
α, β, and γ °	96.8, 93.6, 101.9	90.0, 90.0, 90.0	89.5, 84.6, 89.6	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Unique reflections	90224	55922	98262	27256	93556
Multiplicity*	2.1 (2.1)	3.0 (2.8)	1.9 (1.9)	2.4 (2.3)	3.2 (2.9)
Completeness (%)	83.4 (83.3)	97.1 (98.3)	90.3 (85.4)	92.5 (92.4)	95.2 (88.5)
R-merge	0.077 (0.464)	0.056 (0.472)	0.036 (0.125)	0.129 (0.641)	0.086 (0.497)
Mean I/σ(I)	8.3 (3.0)	16.8 (2.3)	19.7 (7.1)	7.6 (2.0)	11.2 (2.1)
Wilson B-factor	16.99	26.62	18.53	40.09	28.14
Refinement					
Reflections used in refinement	90177	55874	98260	27237	93498
Reflections used for R-free	1999	2000	1969	1363	4624
R-work	0.205 (0.304)	0.193 (0.280)	0.169 (0.203)	0.164 (0.283)	0.174 (0.280)
R-free	0.227 (0.318)	0.234 (0.331)	0.191 (0.225)	0.193 (0.319)	0.205 (0.332)
Number of non-hydrogen atoms in refinement					
macromolecules	5930	5854	5891	3145	12526
ligands	90	0	90	5	96
solvent	524	389	730	88	514
RMS Bond length (Å)	0.007	0.006	0.013	0.004	0.002
RMS Bond angle °	0.9	0.8	1.2	0.7	0.6
Ramachandran allowed (%)	0.8	1.7	0.7	2.8	1.5
Ramachandran outliers (%)	0.0	0.1	0.0	0.0	0.0
Ramachandran favored (%)	99.2	98.2	99.3	97.2	98.5

Table 2.2 continued

	tmMBP1- MTT	apo tmMBP2	tmMBP2- MTT	apo tmMBP3	tmMBP3- MAL
<i>Clashscore</i>	3.5	2.5	3.1	4.4	2.2
<i>Average B-factor</i>	20	28	21	48	31
<i>macromolecules</i>	20	28	21	48	31
<i>ligands</i>	16	N/A	17	44	19
<i>solvent</i>	26	33	29	46	33

^aNumber in parentheses represent values in the highest resolution shell

Figure 2.1: Structural details of the three tmMBP isoforms in complex with maltotetraose and maltose. (A) Comparison of apo (gray) and maltotetraose bound (red) structures of tmMBP1(left) and the structural details of the substrate interaction in the binding pocket (right). (B) Apo and maltotetraose bound structures of tmMBP2. (C) Apo and maltose bound structures of tmMBP3. In the left panels, substrate is shown as green sticks. In the right panels, the ligands are shown as blue sticks and protein residues as green (hydrophobic contact with substrates) and gray (hydrophilic contact with substrates) stick representation. Direct hydrogen bonds between the protein and the ligands are shown as black dashed lines; magenta spheres indicate oxygen from water molecules observed in crystal structures; and water mediated H-bonds to the ligands as red dashed lines and indirect water mediated H-bonds with binding site residues (cyan sticks) as yellow dashed lines. Some residues making two different types of interactions include: hydrophobic and indirect H-bonds (F41 of tmMBP1); hydrophobic and direct H-bonds (W67 of tmMBP1 and W296 of tmMBP3); and direct and indirect H-bonds (E32 and D133 of tmMBP3).

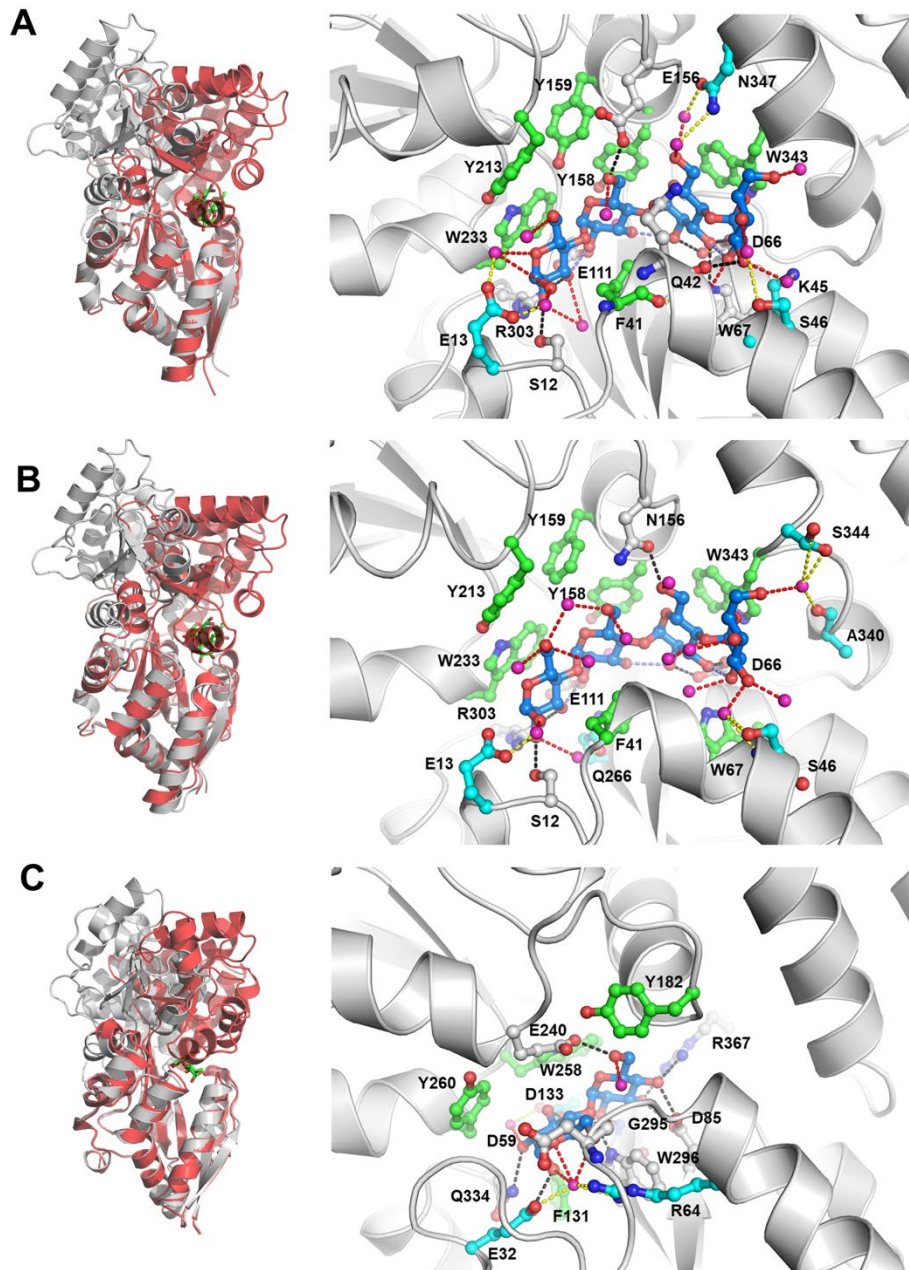


Figure 2.1: Structural details of the three tmMBP isoforms in complex with maltotetraose and maltose.

Figure 2.2: Differential binding of tetra- and di-saccharides in MBPs. (A) Maltotetraose bound to tmMP1, tmMBP2, and ecMBP. (B) Maltose bound to tmMBP3, trehalose bound to tIMBP and maltose bound to ecMBP. The bound substrates are shown in blue sticks and protein residues are shown as green (hydrophobic contact with substrates) and gray (hydrophilic contact with substrates) sticks. Sub-sites in the binding pocket (S1, S2, S3 and S4) are separated by gray vertical dashed lines. Loop 1 (L1), and helices H1, H2 and H3 are also marked for each complex.

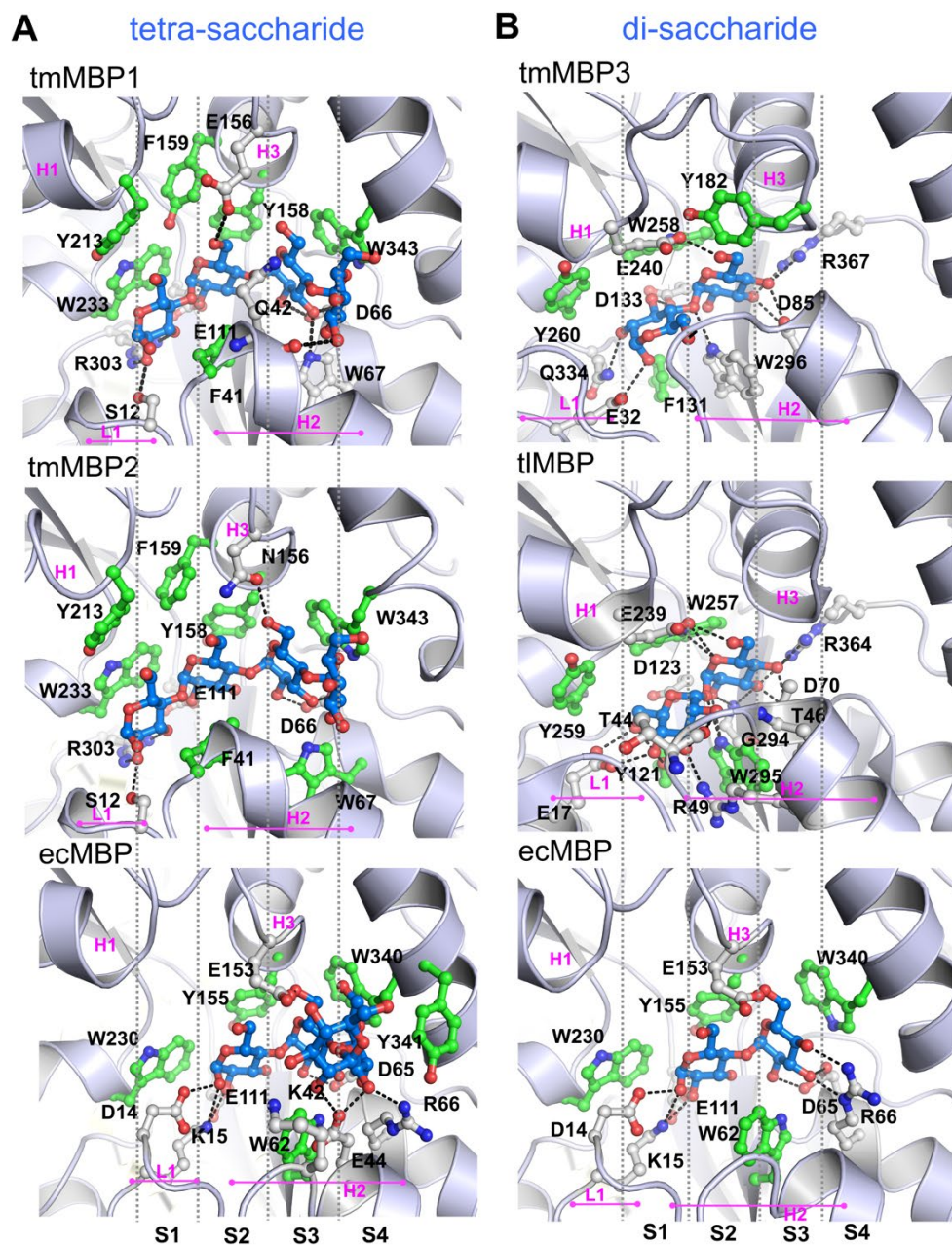


Figure 2.2: Differential binding of tetra- and di-saccharides in MBPs.

			L1			H2		
ecMBP	1	KIEEG	KLVIWINGDK	-GYNGLAEVGGKKFEKDTGIKVTVEHPD	-----	KLEEKFPQVAAT	53	
tmMBP1	3	--MQP	KLTIWCSEK	-QVDILQKLGEFFKAKYGVEVEVQYVN	FQDI--	KS--KFLTAAPE	54	
tmMBP2	5	----	TKLTIWCSEK	-QVDILQKLGEFFKAKYGIPVEVQYVD	FGSI--	KS--KFLTAAPE	54	
tmMBP3	19	---AV	KITMTSGGVG	KELEVLKKQLEM	FHQYPDIEVEIIPM	PDSSTE	75	
tMBP	2	-IEEG	KIVFAVGGAP	NEIEYWKGVIAE	FEKKYPGVTVELKRO	ATDTEQ	60	
ecMBP		GDGP	DIIFWAHDRFGGYAQSGLLAEITP	----DKAFQDKLYPFTW-DAVRYN	KLIIAYPI	108		
tmMBP1		GQG	ADIIVGAHDWVGELAVNGLIEIPN	----FSDLKNFYETAL-NAFSYG	GKLYGIPY	108		
tmMBP2		GQG	ADIIVGAHDWVGELAVNGLIEIPN	----FSDLKNFYDTAL-KAFSYG	GKLYGVPY	108		
tmMBP3		ETDP	DVLMLDVIWPAEFAPF--LEDL--TADKDYFELGEFLPGTV-MSVTVN	GRIVAVPW	130			
tMBP		SSDP	DVFLMDVAWLGGFIASGWLEPLDDYVQKDNVDLSVFFQSVINLADKQG	GKLYALPV	120			
					H3			
ecMBP		AVE	ALSLIYNKDLLP	----NPKTWEEIPALDKELKAK	-----GKSALMFNLQ	EPYF-	155	
tmMBP1		AME	IALIYNKDYP	----EPKTMDELIEIAKQIDEEFGG	---EVRGFITSAA	EFYY-	159	
tmMBP2		AME	AVALIYNKDYP	----SVPKTMDELIEIAKQIDEEYGG	---EVRGFIYDVA	NFYF-	159	
tmMBP3		FTD	AGLLYRKDLLEKYGYDHAPRTWDEL	VEMAKKISQAE	-----GIHGFVWQGA	RYEGL	185	
tMBP		YID	AGLLYRKDLLEKYGYSKPETWQEL	VEMAQKIQSGERETNP	NFWGFWQGGK	QYEG	180	
					H1			
ecMBP		---	TwPLIAADGGYAFKYENGKYDIKDVGVNAGAKAGLTFLVDLIK	NKHMNADTDY--	213			
tmMBP1		---	IAPFIFGYGGYVFKQTEKGLDVNDIGLANEGAIGVKLLKRLVDE	GILDP	SDNY--GT	215		
tmMBP2		---	SAPFILGYGGYVFKETPQGLDVTDIGLANEGAVKGAKLIKRMIDEG	VLT	PGDNY--QT	215		
tmMBP3		VCDF	FLEYLWSFGGDVLDE--SG--KVVIDSPEAVAALQFMVDLIYKH	KVTP	EGVTT--YM	239		
tMBP		VCDF	VEYVYSNGGSLGEFKDGK--VPTLNK	PENVEALQFMVDLIH	KYISPPNTYT--EM	237		
			H1					
ecMBP		--	SIAEAAFNKGETAMTINGPW	AWSNIDT--S--KVNYGVTVL	PTF-KGQPSKPFVGVLS	263		
tmMBP1		MD	----SMFREGQAAMIINGPW	AIKAYKD--A--GIDYGVAPI	PDLEPGVPARPFVGVQG	267		
tmMBP2		MD	----SMFKEGLAAMIINGLW	AIKSYKD--A--GINYGVAPI	PELEPGVPAKPFVGVQG	267		
tmMBP3		EE	-DARRIFQNGEAVFMRNWPY	AWSLVNSDESPIKGKVG	VAPLPMG-PGGRAATLGGWV	297		
tMBP		TE	PVRLMFQQGNAAFERNWPY	AWGLHNADDSPVKGV	VAPLPHF-PGHKSAATLGGWH	296		
ecMBP		AGIN	AASPN--KELAKEFL--ENYLLTDEGLEAVNKDKPLGAVALKS	YEEELAKDPRIA	318			
tmMBP1		FMV	NAKSPN--KLLAIEFL--TSFIAKKETMYRIYLGDPRLP	-SRKDVLELVKDN	PDVV	321		
tmMBP2		FMIN	AKSPN--KVIAMEFL--TNFIARKETMYKIYLADPRLP	-ARKDVLELVKDN	PDVV	321		
tmMBP3		LGINK	FSSPEEKEAAKKLIKFLTSYDQQL--YKAINAGQN	PTKAVYKDPKLKEA	APFMV	355		
tMBP		IGISKY	SDN--KALAWEFVKFVESYSVQK--GFAMNLGWN	PGRVDVYDDPAVVS	SKSPHLK	352		
ecMBP		ATMEN	AQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDE	ALKDAQTRITK	-----	370		
tmMBP1		GFTL	SAANGIPMPNVPQMAAVWAAMNDALNLVVNGKATVEE	ALKNAVERIKAQIQ	---	376		
tmMBP2		AFTQ	SASMGTPMPNVPEMAPVWSAMGDALSIINGQASVED	ALKEAVEKIK	QIEKGS	379		
tmMBP3		ELLG	VFINALPRPRVANYTEVSDVIQRYVHAALTRQTTSED	AIKNI	AKELKFL	---	409	
tMBP		ELRA	VFENAVPRIVPYYPQLSEIIQKYVNSALAGKISPQ	ALDKAQKEA	EELVKQ	--	408	

Figure 2.3: Structure based sequence alignment of thermophilic and mesophilic MBPs. Structural elements that determine the differential substrate binding are highlighted: Loop 1 (L1) in green; helix 1 (H1) in yellow; helix 2 (H2) in red; and helix 3 (H3) in cyan. The analysis was performed using the PROMALS3D server and Clustal Omega.

Figure 2.4: Differential binding of trisaccharide maltotriose. The bound substrates are shown in blue sticks and protein residues are shown as green (hydrophobic contact with substrates) and gray (hydrophilic contact with substrates) sticks. Sub-sites in the binding pocket (S1, S2, S3 and S4) are separated by gray vertical lines.

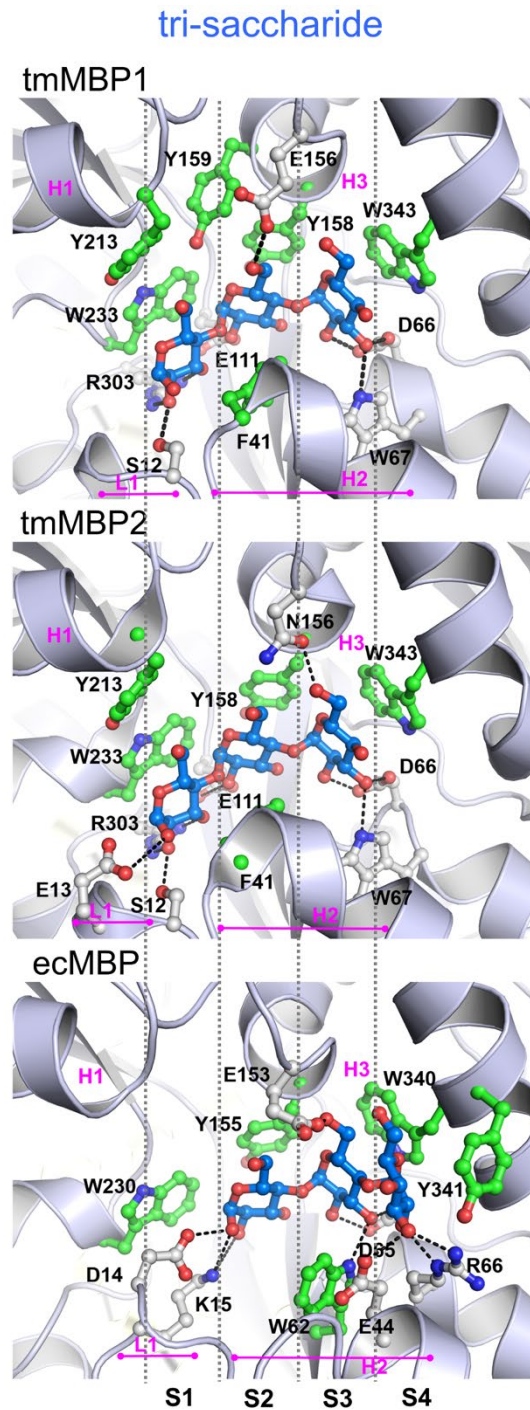


Figure 2.4: Differential binding of trisaccharide maltotriose.

In the di-saccharide specific tmMBP3 and tIMBP, multiple H-bonding interactions are formed with sugar ring occupying the S1 subsite. Other common subsite S1 interactions include the conserved Glu32 in tmMBP3 and Glu17 in tIMBP, and Asp133 in tmMBP3 and Asp123 in tIMBP, which form direct H-bonds with sugar ring bound in the S1 subsite (Figure 2.2). Similarly, a tyrosine that forms hydrophobic interactions is also conserved in both tmMBP3 (Tyr260) and tIMBP (Tyr259). Loop L1 flips in tmMBP3 to allow the opening of subsite S1, in comparison it partially occludes subsite S1 in tIMBP. Furthermore, due to an insertion in helix H1 in tmMBP3 and tIMBP, as compared to the other MBPs, the orientation of pyranose ring is changed in the S1 subsite by almost 90°, where it forms hydrophobic interactions with a conserved tryptophan residue (Trp296 and Trp295 in tmMBP3 and tIMBP, respectively) that protrudes from helix H2 (Figure 2.2).

Subsite S2 binds to the second glucose ring of the saccharides in tmMBPs, as well as in tIMBP, while it binds the first glucose unit of the saccharides bound to ecMBP (Figure 2.2 and Figure 2.4). However, at subsite S2, π - π stacking interactions between aromatic residues (tyrosine, tryptophan or phenylalanine) and the saccharide occurs in all the tmMBPs. Two glutamic acid residues (Glu156 and Glu111) form H-bonds with the second pyranose ring of the saccharide molecule in tmMBP1. However, for trisaccharide bound tmMBP2 the H-bonding occurs only with Glu111; and Asn156 instead forms an interaction with the third glucose ring founds at subsite S3 (Figure 2.4). Similar to tmMBP1 and tmMBP2, ecMBP shows a significant number of H-bonds between substrate in the S2 subsite and water molecules (Figure 2.1). Subsite S2 in tmMBP3 and tIMBP binds the second pyranose ring of the maltose and trehalose, respectively, using extensive H-bonding interactions with the side-chains lining the binding cavity (Figure 2.2), with the interacting amino acids involving charged aspartate and glutamate residues. Moreover, tIMBP has two threonine residues which are on the either ends of the binding pocket and form H-bonds with the substrate at S1 (Thr44) and at S2 (Thr46). Both these residues are present on helix H2 that spans across the entire length of the substrate molecule in tIMBPs (Figure 2.1B).

Subsite S3 binds the third glucose unit of a triose or a larger sugar in tmMBP1 and tmMBP2 and the second sugar ring in ecMBP. For tmMBP1 the overall H-bond (Asp66,

Trp67) and hydrophobic (Trp343) interactions are conserved between maltotriose and maltotetraose for site S3 (Figure 2.2 and Figure 2.4). Nonetheless, for tmMBP2 there is a loss of one H-bond with Trp67 at site S3 when tri-saccharide is substituted with a tetra-saccharide in the binding pocket (Figure 2.2 and Figure 2.4). In ecMBP, the pyranose ring of the di-, tri- and tetra-saccharides forms H-bonds with an aspartate (Asp65) and a glutamate (Glu153; helix H3) and hydrophobic_interactions with tryptophan (Trp340). Additionally, in the disaccharide bound ecMBP it is also observed that there are two additional H-bonds with an arginine residue (Arg66), which is present on the helix adjacent to helix H2 in the N-terminal domain. However, in ecMBP some of the interactions that stabilize the third glucose ring in maltotriose are rewired to support the solvent exposed fourth glucose unit when binding maltotetraose. Both, subsite S3 and S4 are occluded in tmMBP3 and tlMBP due to steric hindrance from helix H2 and H3, which prevents anything larger than a disaccharide from binding.

Subsite S4 is occupied by the third sugar unit of a trisaccharide, or a larger saccharide in ecMBP, and the fourth ring of a tetrasaccharide in tmMBP1 and tmMBP2. Anything larger than a trisaccharide is pushed outside the binding subsite S4 in ecMBP, with only three glucose rings located within the binding pocket. In ecMBP, maltotetraose interactions include an H-bond with Arg66 and Glu44. Additionally, a lysine residue (Lys42), along with an aspartate (Asp65) restricts the movement of the last solvent exposed glucose unit. This reduces the total number of H-bonds associated with third sugar ring at subsite S4 in ecMBP to two (one each with Asp65 and Arg66). In tmMBP1, the fourth glucose ring of maltotetraose is stabilized by a H-bond with backbone of glutamine (Gln42) residue from helix H2. While in tmMBP2, the fourth sugar ring is stabilized by direct a water H-bond, and water mediated H-bonds with residues Ser344, Ser46, and Ala340.

2.4.3 Conformational dynamics in tmMBPs

Computer simulations of apo and holo (substrate bound) forms of tmMBP and ecMBP, and holo form of tlMBP were performed to obtain insights into the conformational dynamics of the MBPs. Note, all simulations were performed at a temperature of 325 K (~52 °C), to match the limits of the solution characterization conditions of the tmMBPs (discussed below). Although the optimal growth temperature of *T. maritima* is 80 °C, we expect the

results from simulations to be qualitatively similar and indicative of results at higher temperatures. In addition to the solved crystal structure, computer modeling allowed investigation of MBPs in complex with smaller oligomeric sugars for the three isoforms that did not crystallize. The motivation behind generating the models with smaller sugars was to understand the role of protein dynamics and protein-substrate interactions in differential binding of substrates, and further to quantify the preference of each substrate for a specific site in the binding pocket. MD simulations of 0.5 μ s duration were performed for each of the four apo systems (tmMBP1, tmMBP2, tmMBP3, and ecMBP) and 17 substrate bound systems (11 simulations for tmMBP, 5 for ecMBP and 1 for tMBP; see Table 2.1 for details of the protein-substrate complexes simulated).

Characterization of the computer simulations indicated changes in the apo and substrate bound protein conformations, as well as provided insights into the substrate stability in the binding pocket. Variation in radius of gyration (R_g) provides an indication of changes in protein conformation. On average the R_g of substrate bound systems is smaller by approximately 1.0-1.5 Å (5-7 %) compared to the apo systems (Figure 2.5). The smaller R_g of substrate bound systems indicate compact conformations, due to increased interactions in the binding pocket located adjacent to the hinge region of the MBPs. The increased interactions in the hinge region are responsible for the observed bending between the N- and C-terminal domains (Figure 2.2). For the smaller maltose and glucose substrates, the MBPs could have multiple binding positions. These were explored computationally, with two starting positions for the saccharide tested: MAL1 simulations started with maltose in S1+S2 subsites and MAL2 simulations started with maltose in the S2+S3 subsites. The results indicate that MAL2 simulations for tmMBP1 and tmMBP2 are unstable (panel B and F in Figure 2.5) with protein conformations opening up (increasing R_g values, observed in the blue curves) and returning to the apo conformation, and as the substrate is ejected completely from the binding pocket in case of tmMBP2. However, MAL1 simulations for tmMBP1 and tmMBP2 are stable, indicating that the substrate is stable in S1+S2 subsites, which is consistent with the observed X-ray structures. Similarly, for tmMBP3, glucose can bind in alternate sites S1 (GLU1 simulation, panel I in Figure 2.5) or S2 (GLU2 simulation panel J in Figure 2.5). The simulation results indicate stable binding in S1 subsite, however, S2 subsite was less preferred as the simulations indicated

relative opening up of the protein. Larger RMSD values were also seen in simulations where substrates were ejected (Figure 2.6).

For the larger maltotriose and maltotetraose substrates, tmMBP1 and tmMBP2 show stable binding as indicated by the smaller R_g values throughout the simulations (panels C, D, G and H Figure 2.5). Similarly, for proteins with smaller binding pockets, tmMBP3 and tlMBP, binding to di-saccharide substrates showed stable simulations (panel K in Figure 2.5 and panel F of Figure 2.7). The promiscuous ecMBP shows interesting behavior (Figure 2.7) in simulations, where the smaller glucose substrate was unstable and ejected out of the binding pocket in both S1 and S2 subsites, however, maltose and maltotriose show stable binding in the pocket. Maltotetraose shows mostly stable binding with some minor fluctuations, possibly due to the fact that the last sugar unit does not have room in the binding pocket.

The computational R_g trends in the tmMBPs were confirmed by small-angle X-ray scattering (SAXS) studies of these proteins in the solution state. The fits of the experimental SAXS data to theoretical scattering data generated from the crystal structures (Figure 2.8) are in agreement (Table 2.3). The experimental R_g decreases by ~ 1.5 Å after substrate binding (Table 2.3), which is comparable to the observed computer simulation results (Figure 2.5 and Figure 2.7). It is also interesting to note that the apo proteins in MD simulations for all MBPs show stable R_g values throughout the period of simulation, consistent with a body of literature demonstrating the MBP proteins do not populate the ligand bound conformation in the absence of ligand.^(28,29,52,53) The hinge angle of the apo proteins shows little deviations over the course of MD simulations (Figure 2.9), ranging between -10° to $+10^\circ$ from the first frame of the simulations and consistent with previous simulations of ecMBP.⁽²⁸⁾

The conformational flexibility (or intrinsic dynamics) of the MBP systems were quantified using quasi-harmonic analysis (QHA) of the protein conformations sampled during the MD simulations. The aggregated root means square fluctuation of top 10 slowest QHA modes ($RMSF_{10}$) provides a good measure of protein's conformational flexibility. Such an analysis

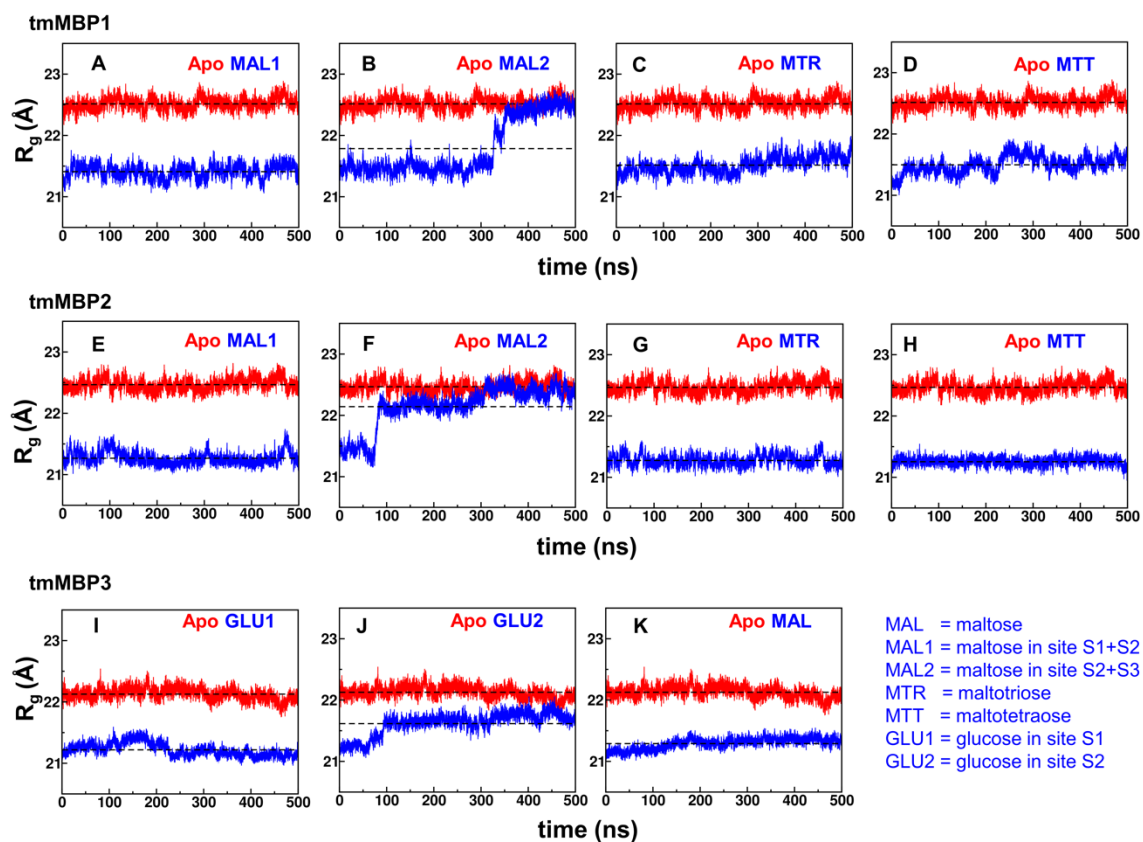


Figure 2.5: Radius of gyration (R_g) computed from MD simulations. Panels A-D show tmMBP1 results for MAL1, MAL2, MTR and MTR; panels E-H show tmMBP2 results for MAL1, MAL2, MTR and MTR; and panels I-K show tmMB1 results for GLU1, GLU2, and MAL. R_g for apo simulations depicted in red and the substrate bound simulations in blue, and average value of R_g computed from all simulation snapshots are shown by horizontal lines. MAL1 and MAL2 indicate simulations with maltose starting in two alternate positions; site S1+S2 for MAL1 and S2+S3 for MAL2. GLU1 simulation started with glucose in S1 binding pocket and GLU2 in the S2 binding pocket.

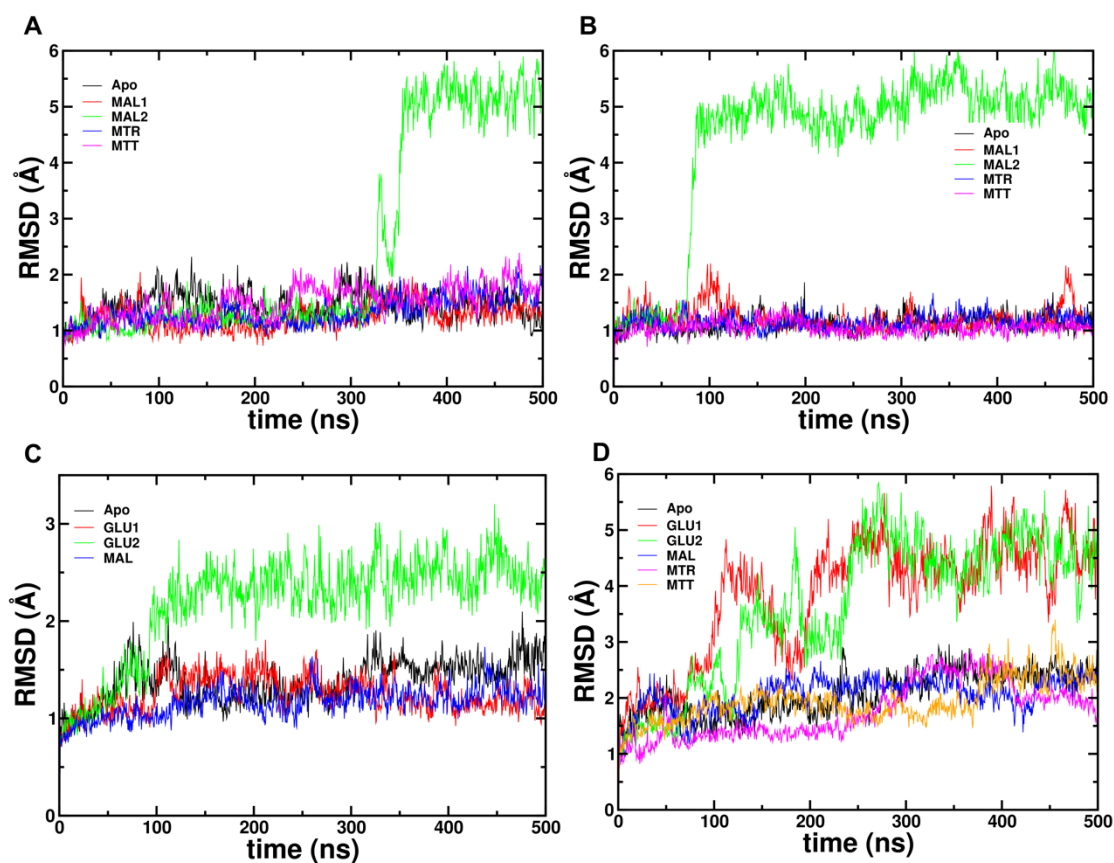


Figure 2.6: Root mean square deviations (RMSD) computed from the MD simulations. (A) tmMBP1, (B) tmMBP2, (C) tmMBP3, (D) ecMBP. See Figure 2.5 of main text for substrate key. First structure of the simulation was used as reference for RMSD calculations. Large deviations indicate change in conformation of the protein due to substrate leaving the pocket and protein opening up.

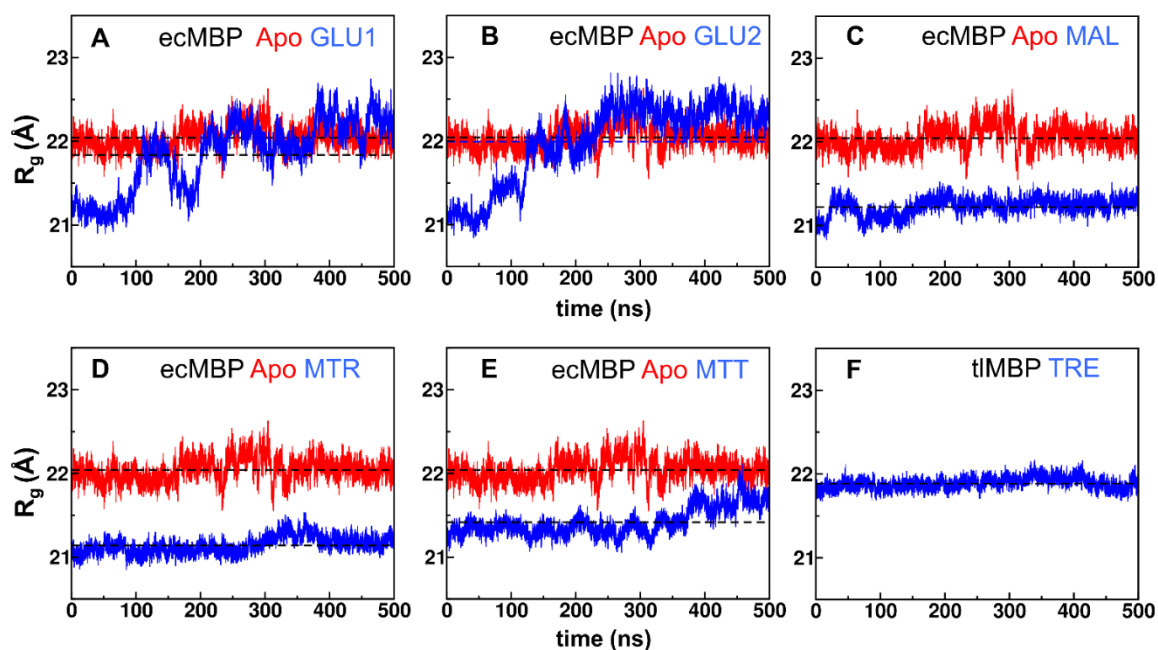


Figure 2.7: Radius of gyration (R_g) computed from MD simulations. R_g for apo simulations depicted in red and the substrate bound simulations in blue, and average value of R_g computed from all simulation snapshots are shown by horizontal lines. Results are shown for ecMBP bound to glucose (GLU, panels A-B), maltose (MAL, panel C), maltotriose (MTR, panel D) and maltotetraose (MTT, panel E); and trehalose (TRE) bound to tIMBP (panel F). GLU1 simulation started with glucose in S1 binding pocket and GLU2 in the S2 binding pocket.

Table 2.3: Radius of gyration (R_g) values of tmMBPs. The experimental R_g determined by SAXS experiments decreases after substrate binding, which is also observed in the computer simulation results.

<i>Protein</i>	SAXS Conc. (mg/ml)	SAXS R_g (Å)	SAXS D_{max} (Å)	χ^2*	Averaged R_g (Å), from computations
<i>apo tmMBP1</i>	2.7	24.1	82.3	1.19	22.5
<i>tmMBP1-MTT</i>	2.6	22.6	78.3	1.06	21.5
<i>apo tmMBP2</i>	2.6	23.5	79.1	1.29	22.4
<i>tmMBP2-MTT</i>	2.4	22.2	69.9	1.21	21.2
<i>apo tmMBP3</i>	3.6	22.4	70.3	1.18	22.1
<i>tmMBP3-MAL</i>	2.5	22.6	72.9	1.40	21.3

* χ^2 generated from the fit of the CRY SOL calculated scattering profile to the experimental data.

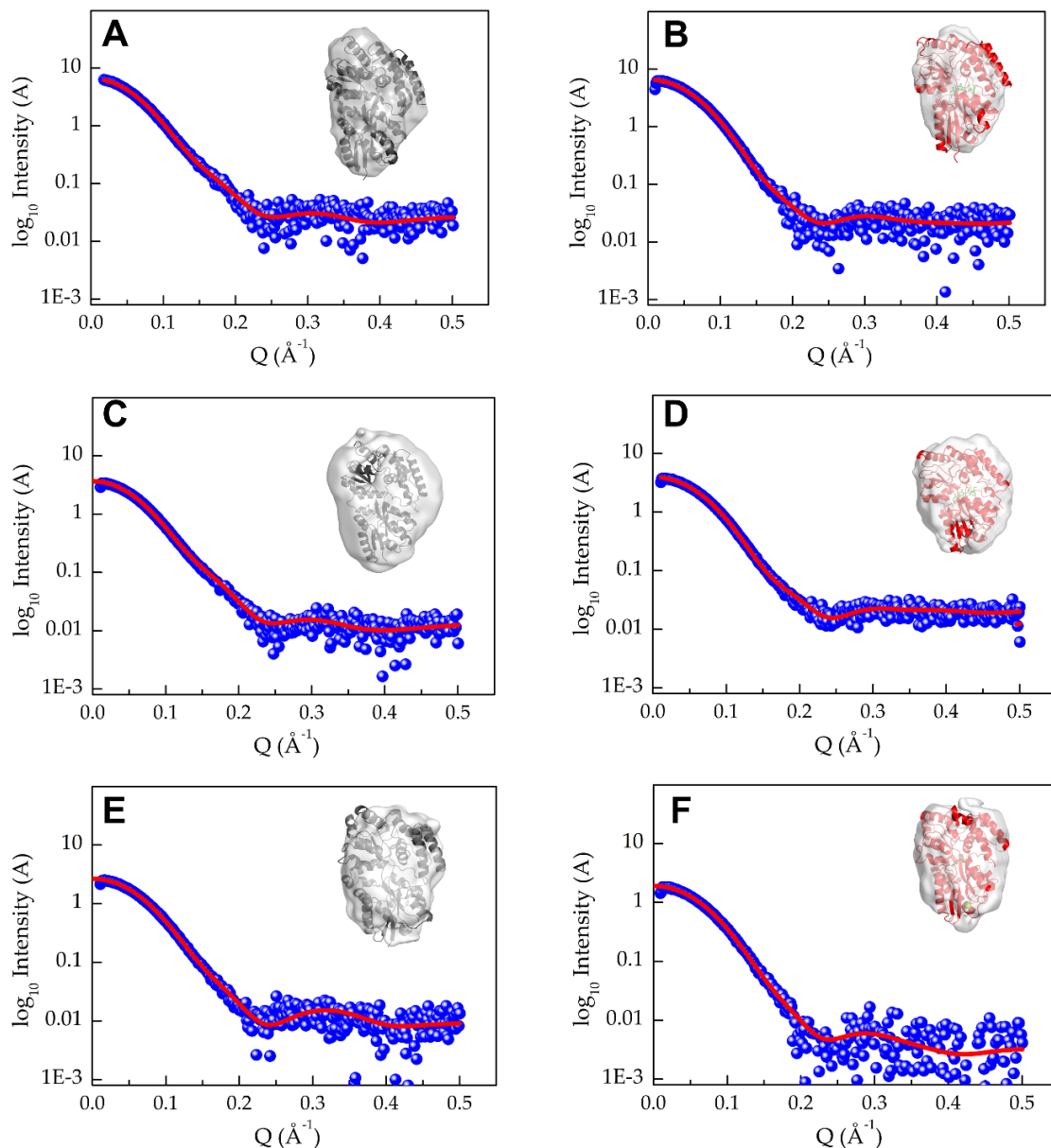


Figure 2.8: SAXS of apo and substrate-bound tmMBPs. Experimental SAXS data of (A) apo tmMBP1, (B) maltotetraose bound tmMBP1, (C) apo tmMBP2, (D) maltotetraose bound tmMBP2, (E) apo tmMBP3, and (F) maltose bound tmMBP3. Solid red lines are the SAXS data calculated from the respective crystal structures. Inset is the ab-initio model generated from the SAXS data superimposed on the crystal structures.

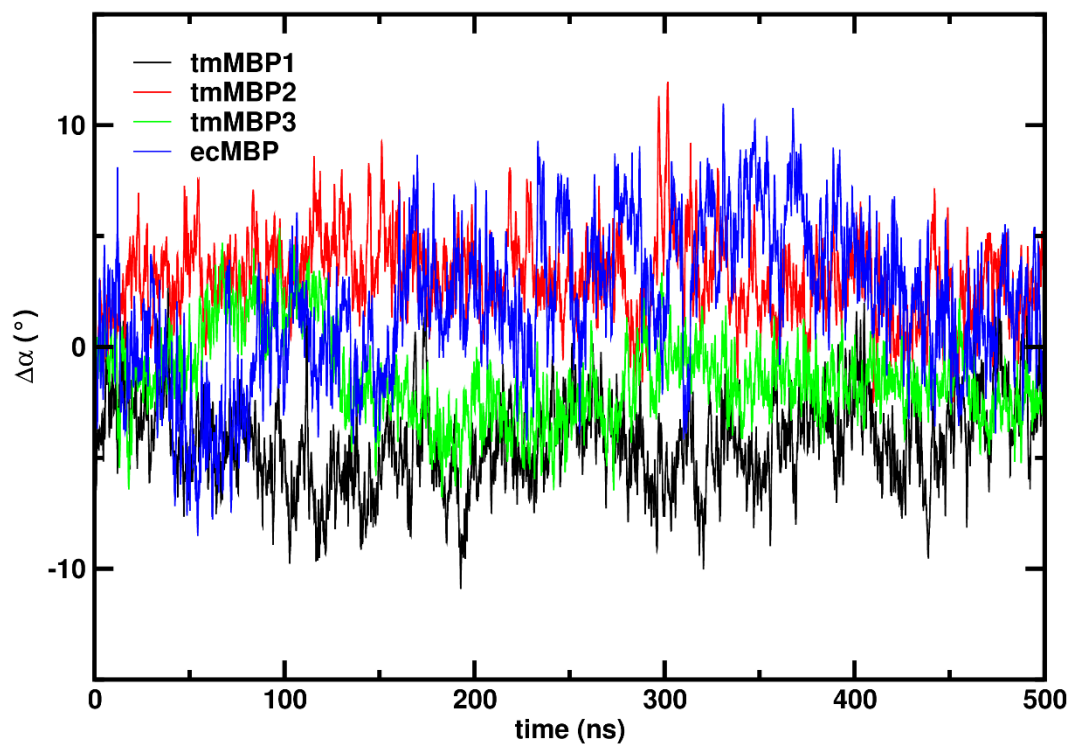


Figure 2.9: Variation of hinge angle for apo MBPs. The hinge angle variation from first frame is depicted over the course of MD for tmMBP1 (black line), tmMBP2 (red), tmMBP3 (green) and ecMBP (blue). The angle was computed between the center of mass for N-terminal domain, hinge region, and C-terminal domain as defined by residues in the Table 2.4.

Table 2.4: Residues of the N-terminal domain, Hinge and C-terminal domain in MBPs. The hinge angle was computed between the center of mass for N-terminal domain, hinge region, and C-terminal domain as defined by residues in this table.

	N-terminal domain	Hinge	C-terminal domain
<i>tmMBP1</i>			
<i>Apo</i>	4-110, 265-315	111, 264	112-263, 316-376
<i>MAL1/MAL2</i>	3-110, 265-315	111, 264	112-263, 316-376
<i>MTR</i>	3-110, 265-315	111, 264	112-263, 316-376
<i>MTT</i>	3-110, 265-315	111, 264	112-263, 316-377
<i>tmMBP2</i>			
<i>Apo</i>	5-110, 265-317	111, 264	112-263, 318-377
<i>MAL1/MAL2</i>	6-110, 265-315	111, 264	112-263, 316-384
<i>MTR</i>	6-110, 265-315	111, 264	112-263, 316-384
<i>MTT</i>	5-110, 265-317	111, 264	112-263, 318-383
<i>tmMBP3</i>			
<i>Apo</i>	19-133, 296-352	134, 295	135-294, 353-411
<i>GLU1/GLU2</i>	19-133, 296-352	134, 295	135-294, 353-409
<i>MAL</i>	19-133, 296-352	134, 295	135-294, 353-409
<i>ecMBP</i>			
<i>All</i>	1-110, 261-313	111, 260	112-259, 314-370

has been successfully used for relating dynamics to function for a number of other proteins.⁽⁵⁴⁻⁵⁶⁾ Both crystallographic B-factors and the RMSF₁₀ showed similar trends of reduction in the overall conformational flexibility of MBPs when they bind to preferred substrates (Figures 2.8 – 2.12). The results indicate that large fluctuations occur in the regions 40-60, 130-150 and 165-195 for apo as well as substrate bound systems for tmMBP1, tmMBP2 and ecMBP. These regions in tmMBP1 and corresponding regions in tmMBP3, are highlighted with ellipses and marked as 1, 2, and 3 in Figure 2.10 respectively. Region 165-195 shows the largest flexibility, interestingly this region corresponds to region 194-218 in tmMBP3 (see Figure 2.3); and as seen in Figure 2.10, part of this region in tmMBP3 (residue 197-211) also shows increased flexibility. This region is located in the C-terminal domain and interacts with the N-terminal domain in the apo form of tmMBP and ecMBP structures. Previous experimental studies have shown altering this loop enhances the substrate binding affinities in ecMBPs,⁽⁵⁷⁾ and plays a role in the energetics of the conformational dynamics in the apo state of the protein.⁽²⁸⁾ It is therefore likely that this region plays a role in substrate selectivity. Additionally, subtle variations in the dynamics were also observed when the same protein was interacting with different substrates. The lower panels of Figure 2.10 provide quantitative comparison of the flexibility of apo and different substrate bound cases (dashed red line shows the average over all residues of apo protein for comparison). For ecMBP, tmMBP1 and tmMBP2, it is observed that with an addition of each pyranose ring in the substrate, the dynamics of only certain regions in the protein changes (see Figures 2.11 – 2.15). These variations in dynamics were not just observed around the binding site, rather the changes occurred at sites distally located from the substrate binding cavity (Figures 2.10 – 2.14), which supports the potential role of these regions in substrate binding or specificity. Unfortunately, the current simulations cannot differentiate whether dynamical variations are caused due to different substrate binding or these motions playing an active role in tuning binding preference.

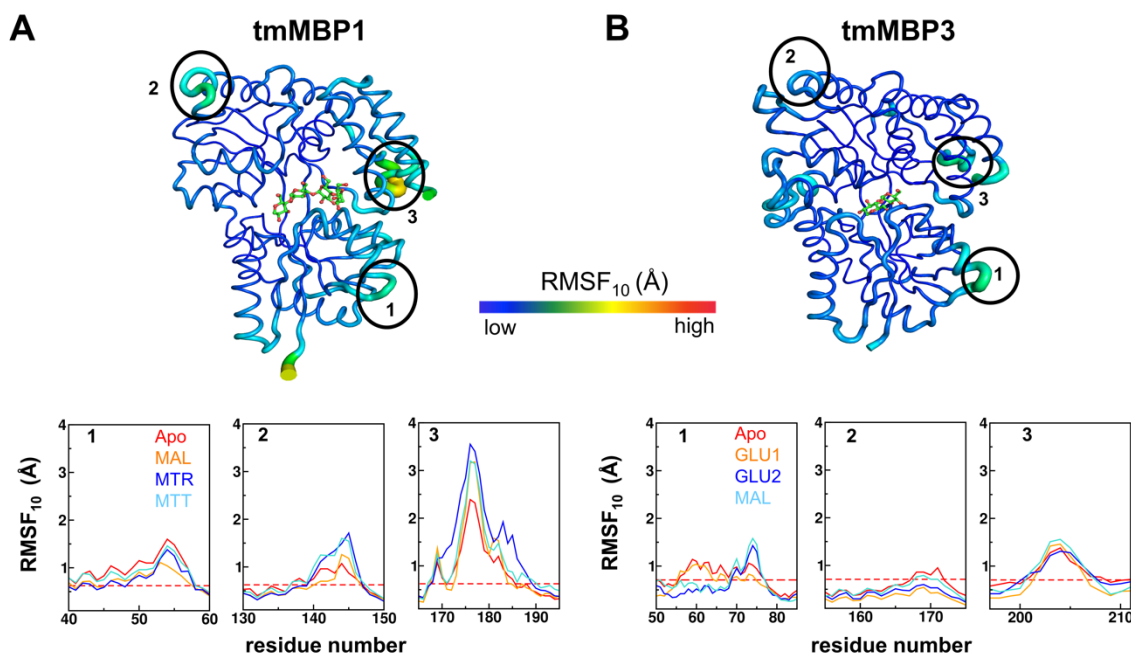


Figure 2.10: Comparison of tmMBP1 and tmMBP3 conformational flexibility based on RMSF10. (A) tmMBP1 and (B) tmMBP3. The tube thickness corresponds to degree of flexibility, with thicker tubes (green/yellow) indicating more flexible regions than the rest of rigid protein (dark blue). Surface exposed regions displaying higher than average conformational flexibility are marked by black ellipses (40-60, 130-150, and 165-195 for tmMBP1 and 50-85, 155-175, and 197-211 for tmMBP3), and the plots below compare the observed values for apo and protein in complex with various substrates. The dashed red line in the plots indicate the average value of RMSF10 observed in the apo proteins for comparison.

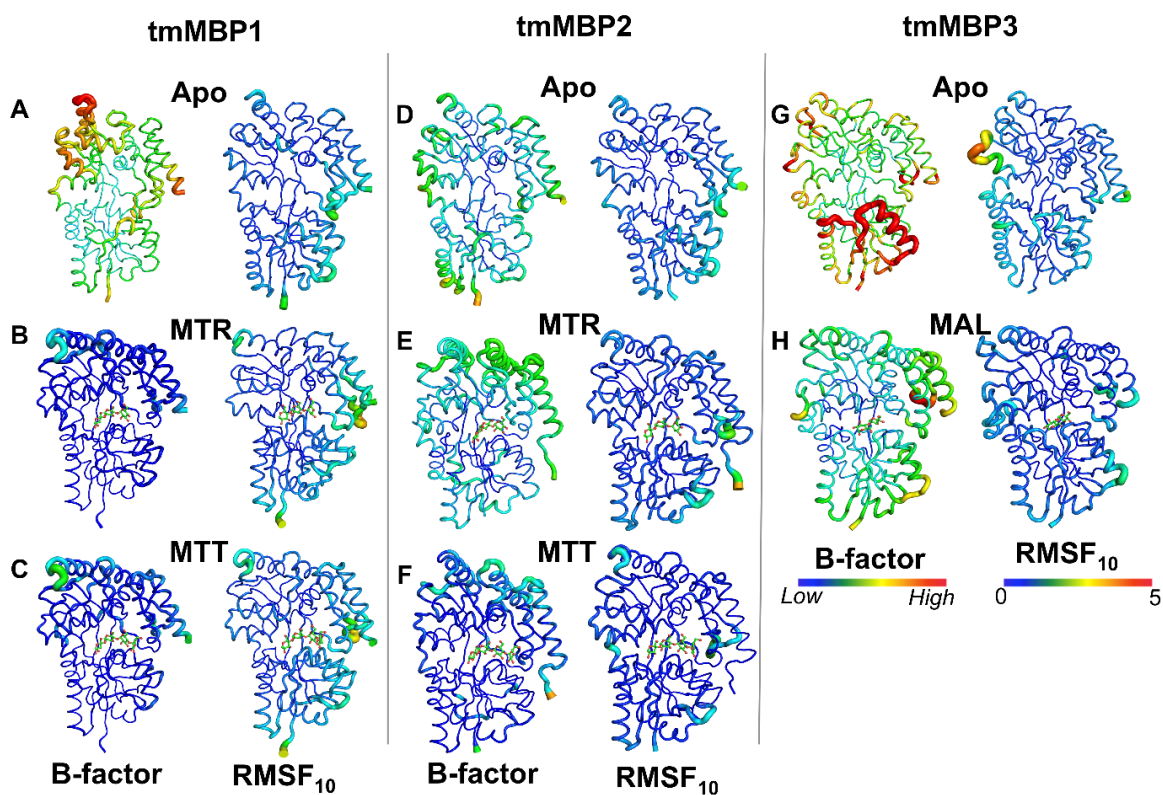


Figure 2.11: Comparison of crystallographic β -factors with computational all atom root mean square fluctuations (RMSF₁₀) of tmMBPs. (A) tmMBP1 Apo, (B) tmMBP1 maltotriose, (C) tmMBP1 maltotetraose, (D) tmMBP2 apo, (E) tmMBP2 maltotriose, (F) tmMBP2 maltotetraose, (G) tmMBP3 apo and (H) tmMBP3 maltose. Proteins are represented with tube width corresponding to crystallographic B-factor on left. On right side tube width corresponds to RMSF₁₀ fluctuations.

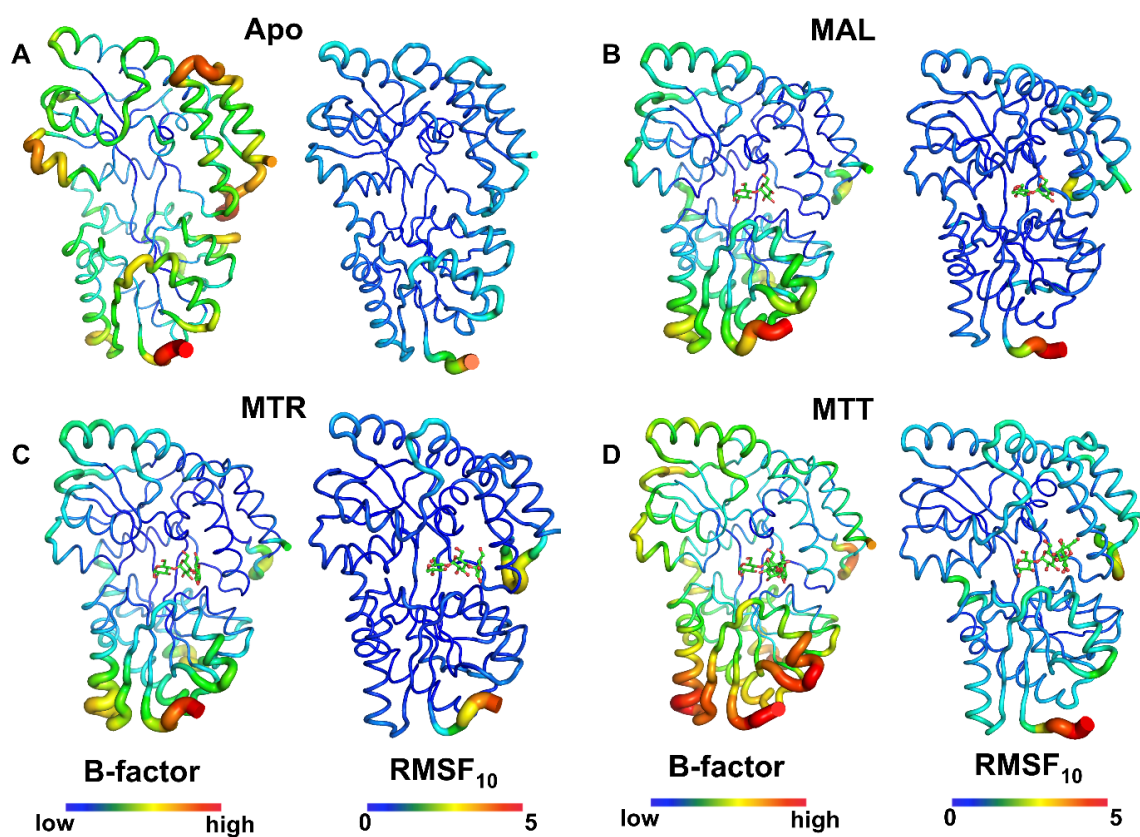


Figure 2.12: Comparison of crystal B-factors with computational all atom root mean square fluctuations ($RMSF_{10}$) of ecMBP. (A) ecMBP apo, (B) ecMBP maltose, (C) ecMBP maltotriose, and (D) ecMBP maltotetraose. Proteins are represented with tube width corresponding to crystallographic B-factor on left. On right side tube width corresponds to $RMSF_{10}$ fluctuations.

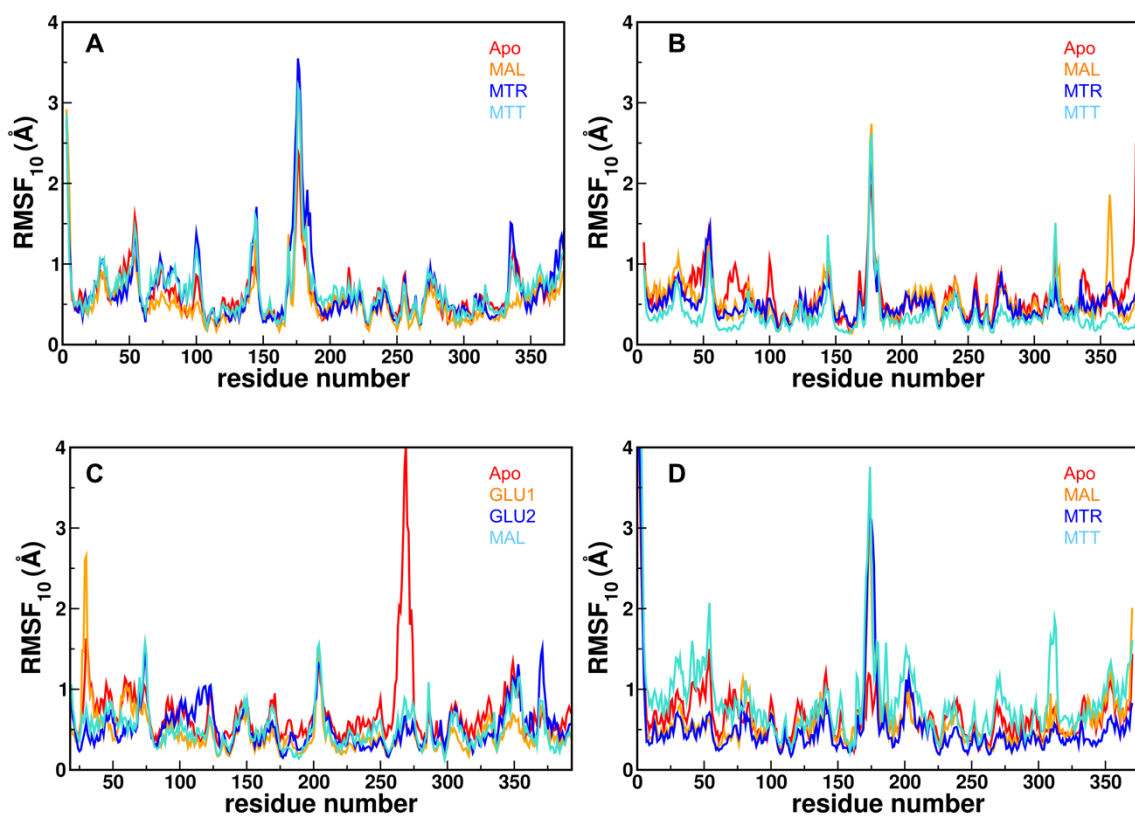


Figure 2.13: All atom RMSF_{10} plot of tmMBP and ecMBP systems. (A) tmMBP1 apo and substrate-bound forms, (B) tmMBP2 apo and substrate-bound forms, (C) tmMBP3 apo and substrate-bound forms, (D) ecMBP apo and substrate-bound forms. All the apo proteins show larger change in RMSF_{10} values than the substrate bound proteins.

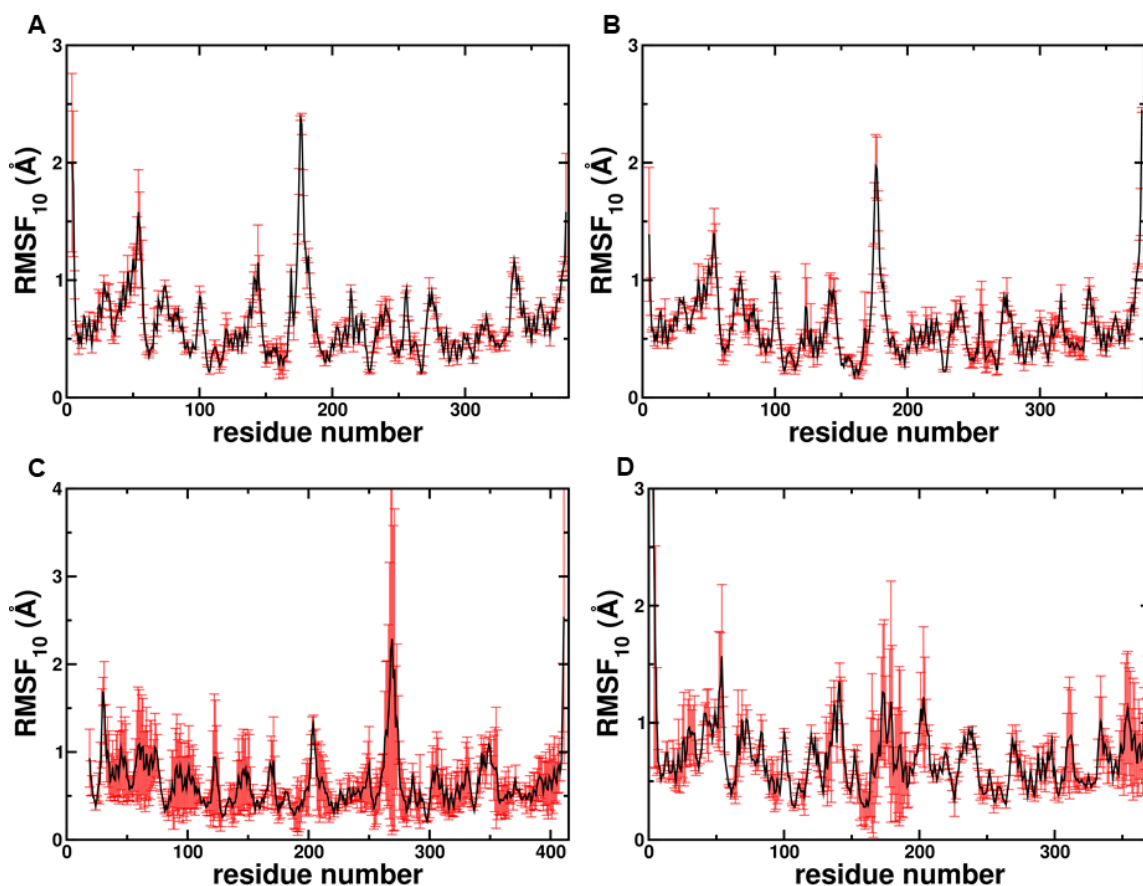


Figure 2.14: Errors associated with computed RMSF_{10} . (A) apo tmMBP1, (B) apo tmMBP2, (C) apo tmMBP3, (D) apo ecMBP. The errors were computed by dividing the MD trajectories into two halves: 0-0.25 μs and 0.25-0.50 μs . The RMSF_{10} values were calculated for each half, the difference in values is considered as error (red bars), while the average of the two values is plotted as black curves. The data shown in the main manuscript and previous supporting figure is based on computation of RMSF_{10} based on the entire 0.5 μs trajectory.

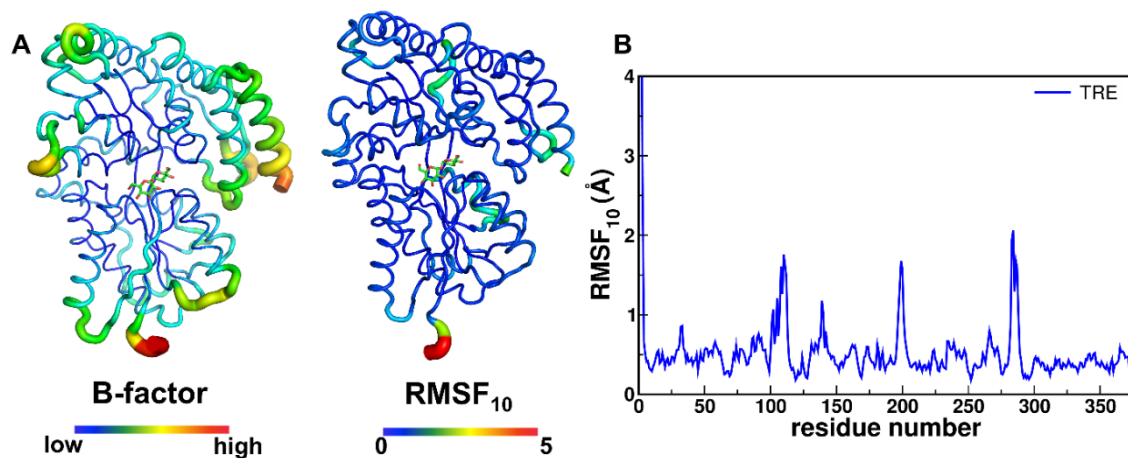


Figure 2.15: Conformational flexibility of the trehalose-bound tIMBP. (A) B-factor and RMSF₁₀ tube structure showing the overall change in the dynamics of the protein after 0.5 μ s of simulations. The thickness of the tubes suggests the degree of fluctuations at that particular site. (B) An all atom RMSF₁₀ plot showing the larger fluctuations in the structure.

2.4.4 Interaction energy of substrates with tmMBPs and ecMBP

The interaction energies for substrate bound systems (Table 2.5), calculated from computer simulations, also supports the differential binding in MBPs. These computational estimates only provide enthalpic contributions, as accurate estimates of entropic contributions are difficult to obtain. Larger negative values in Table 2.5 are considered to be more favorable interactions than smaller negative values. The magnitude of interaction energies, averaged per residue (Table 2.5), were stronger for the known natural binders (which are known to be maltotriose and maltotetraose for tmMBP1/tmMBP2; maltose for tmMBP3; maltose, maltotriose, and maltotetraose for ecMBP; and trehalose for tMBP). In the case of the promiscuous ecMBP, maltotriose has the most favorable interaction energy, followed by maltotetraose and maltose, in a trend identical to the previously experimentally measured binding constants for this system, ⁽¹⁹⁾ which suggests the simulations capture the elements that lead to ligand discrimination accurately. Disaccharide binding to the specific tmMBP3 and tMBP proteins produced significantly stronger interaction energies when compared to the promiscuous ecMBP. tmMBP1 and tmMBP2 show a similar interaction energy in the case of maltotriose, however, of these two proteins tmMBP2 shows significantly stronger interactions with maltotetraose. The substrate-protein complex co-crystallized in this study exhibited more favorable interaction energy profiles when compared to those substrate-protein interactions that failed to co-crystallize. The cases where the ligand showed relative unstable binding during the computer simulation, indicated weaker interaction energies (marked with a, b in Table 2.5), compared to the cases of stable substrate binding. Previous studies of oligosaccharide binding to ecMBP, based on isothermal titration calorimetry (ITC), indicated dissociation constant (K_d) for maltose and maltotriose is tighter when compared to maltotetraose. ⁽¹⁹⁾ The computational results obtained in this study are in agreement with these results as maltotriose shows more favorable binding than maltose and maltotetraose (Table 2.5).

Figure 2.16 provides more details of the interaction energy at protein residue and substrate unit pair level. The protein residues showing most favorable contacts are highlighted. For each protein, there were some common residues that showed favorable interactions across substrates of different lengths (Figure 2.16). These include: Glu13, Phe41, Glu111, Tyr158, Trp233 and Arg303 for tmMBP1 and tmMBP2; Glu32, Asp85,

Table 2.5: Averaged protein-substrate interaction energy. All values are shown in kcal/mol. The values in parenthesis are per sugar unit. GLU1/GLU2 and MAL1/MAL2 indicate two different simulations with substrate alternate sites (see Figure 2.10 and its legend for details). The errors associated are provided; see SI text for details on how the errors were calculated.

	Glucose		Maltose		Maltotriose	Maltotetraose
	GLU1	GLU2	MAL1	MAL2		
<i>tmMBP1</i>			-55.5 ± 0.35 (-27.8)	-42.6 ± 0.18 (-21.3) ^a	-79.4 ± 0.09 (-26.5)	-90.0 ± 0.44 (-22.5) ^c
<i>tmMBP2</i>			-55.7 ± 0.10 (-27.8)	-24.8 ± 0.27 (-12.4) ^b	-79.4 ± 0.13 (-26.5)	-92.0 ± 0.36 (-23.0) ^c
<i>tmMBP3</i>	-40.0 ± 0.18	-33.6 ± 0.09 ^a	-72.6 ± 0.14 (-36.3) ^c			
<i>ecMBP</i>	-26.6 ± 0.15 ^a	-33.9 ± 0.20 ^a	-62.2 ± (-31.1)		-81.1 ± (-27.0)	-78.5 ± (-19.6)
<i>tlMBP</i>			-68.3 ± (-34.2)			

^a substrate changed significantly from the original orientation as a number of protein-substrate contacts are broken

^b substrate was completely ejected out of the binding pocket.

^c indicates the substrate in complex with the protein for X-ray solved in this study.

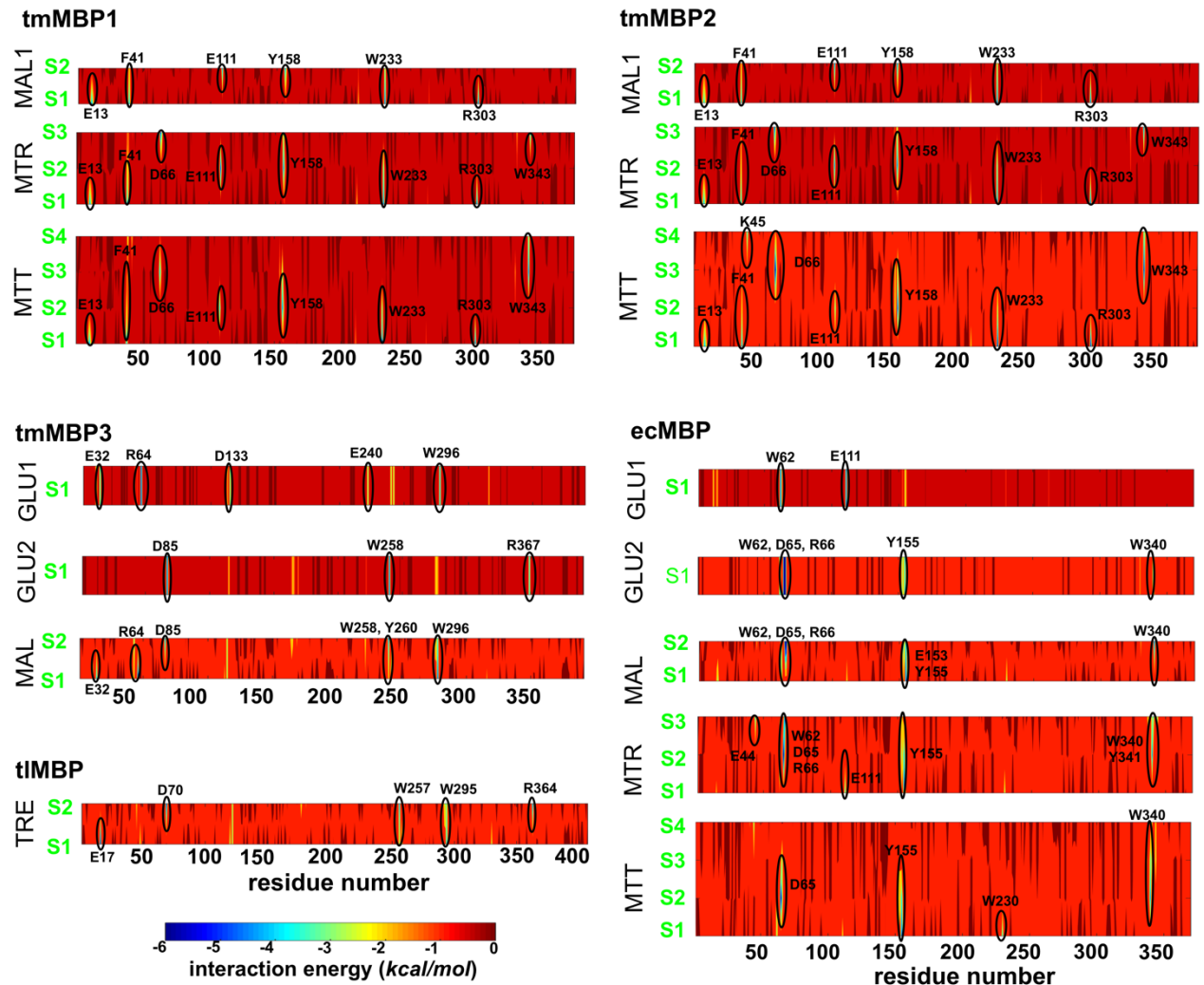


Figure 2.16: Protein-substrate interaction energy. The interaction energy was computed as a sum of electrostatics and van der Waals contributions between each protein residue and substrate unit pair. The yellow-green-blue areas with more favorable contacts are highlighted, and the corresponding protein residue is marked. See Figure 2.5 legend in main manuscript for the substrate key.

Table 2.6: List of residues making favorable contact obtained from interaction energy analysis. An interaction is considered strong if the enzyme residues interact with the substrate with interaction energy < -3 kcal/mol of E_{total} . The enzyme residues moderately interacting with the substrate are also listed.

<i>Protein</i>	Substrate	Strong < -6 kcal/mol	Strong -5 to -3 kcal/mol	Moderately Favorable -2 kcal/mol
<i>tmMBP1</i>	maltose (MAL)		Glu13, Phe41, Glu111, Tyr158, Trp233, Arg303	Lys14, Tyr159
	maltotriose (MTR)		Glu13, Phe41, Asp66, Glu111, Tyr158, Trp233, Arg303, Trp343	Lys14, Trp67
	maltotetraose (MTT)		Glu13, Phe41, Asp66, Glu111, Tyr158, Trp233, Arg303, Trp343	Lys14, Gln42
<i>tmMBP2</i>	maltose (MAL)		Glu13, Phe41, Glu111, Tyr158, Trp233, Arg303	Lys14, Tyr213
	maltotriose (MTR)		Glu13, Phe41, Asp66, Glu111, Tyr158, Trp233, Arg303, Trp 343	Lys14, Trp67
	maltotetraose (MTT)	Asp66 Trp233	Glu13, Phe41, Lys45, Glu111, Tyr158, Arg303, Trp 343	Lys14, Trp67, Asn156

Table 2.6 continued

<i>Protein</i>	Substrate	Strong < -6 kcal/mol	Strong -5 to -3 kcal/mol	Moderately Favorable -2 kcal/mol
<i>tmMBP3</i>	glucose (GLU1)		Glu32, Arg64, Asp133, Glu240, Trp296	Val29, Trp258, Tyr260
	glucose (GLU2)		Asp85, Trp258, Arg367	Asp133
	maltose (MAL)	Arg64	Glu32, Asp85, Trp258, Tyr260, Trp296	Val29, Ser61, Asp133, Glu183
<i>ecMBP</i>	glucose (GLU1)		Trp62, Glu111	Asn12, Tyr155
	glucose (GLU2)	Asp65	Trp62, Arg66, Tyr155, Trp 340	Glu153
	maltose (MAL)	Asp65	Trp62, Arg66, Glu153, Tyr155, Trp340	Lys15, Glu111, Trp230
	maltotriose (MTR)	Asp65	Glu44, Trp62, Arg66, Glu111, Tyr155, Trp340, Tyr341	Glu153, Trp230, Ser337, Arg344
	maltotetraose (MTT)		Asp65, Tyr155, Trp230, Trp340	Glu45, Trp62, Arg66, Ser337, Trp340, Tyr341, Arg344
<i>tlMBP</i>	trehalose (TRE)	Glu17	Asp70, Trp257, Trp295, Arg364	Arg49, Tyr121, Asp123, Tyr259, Gly293, Gly294, Trp331

and Trp296 for tmMBP3; Trp62, Asp65, Arg66, Tyr155 and Trp230 for ecMBP. A full list of residues is available in Table 2.6. Not surprisingly these residues are also the ones which make hydrophilic and hydrophobic contacts in the obtained tmMBP X-ray structures. These interactions do not show up in the case of weak binding substrates.

2.5 Discussion

The periplasmic binding protein super-family is a collection of proteins with a conserved structural fold; however, these proteins show considerable diversity in the size of substrates that they bind. ⁽⁵⁸⁾ One subset of the periplasmic binding protein super-family, oligosaccharide binding MBPs, is of interest with regards to understanding the binding specificity in these proteins, as members of this family share a common protein fold and bind a wide diversity of substrates. ⁽⁵⁹⁾ Intriguingly three different MBP genes are present in *T. maritima*, which encode for three different isoforms with different substrate binding preferences. Two of these isoforms, tmMBP1 and tmMBP2, show 90% sequence similarity while the third, tmMBP3, only shares 40%; however, all three share the same structural fold. In this study, we have used structural and computational studies to characterize the substrate binding preferences of these three isoforms, which provide insights into presence of multiple genes in *T. maritima*, with related function. The obtained X-ray structures of tmMBP1, tmMBP2 and tmMBP3, in substrate free and substrate bound forms, provide detailed information about the interaction of native oligosaccharide substrates in the binding pocket. The binding pockets of tmMBP1 and tmMBP2 are more open, and suitable to accommodate larger substrates, however, in the case of tmMBP3 the binding pocket is sterically hindered by helices H2 and H3. This observation is similar to the binding characteristics of *T. litoralis* trehalose/maltose binding protein reported previously. ⁽⁶⁰⁾

One of the largest structural change observed between the three isoforms, upon substrate binding, is the extent of the hinge bending motion between the N- and C-terminal domains. The binding pocket of periplasmic binding proteins is partially constituted by the residues around the hinge region. In the case of tmMBP1 and tmMBP2 the substrate binding induces the relative orientation of the two domains to change by $\sim 50^\circ$ while in the case of tmMBP3 the change is only $\sim 35^\circ$ (see Figure 2.2). Previous reports have indicated that

the substrate binding affinities are affected by the residues (and mutations) in the hinge region, ⁽¹⁸⁾ which interestingly also manifests through observed differences in the magnitude of the hinge bending motion. ⁽⁵²⁾

The larger binding pockets of tmMBP1 and tmMBP2 also show the presence of water molecules in and around the binding pocket, which have previously been shown to contribute towards ligand specificity in periplasmic binding proteins. ⁽⁶¹⁾ In both tmMBP1 and tmMBP2, the majority of water involving interactions are found near subsite S1. Similarly, in tmMBP3 a number of H-bond and hydrophobic interactions stabilize the first pyranose ring of the bound maltose. It was observed that in all the stable computer simulations (where the substrate stays in the binding cavity for the entire simulation) subsite S1 was occupied consistently. Interestingly, simulations that started with a vacant S1 subsite (substrate present in other subsites) showed unstable substrate binding in other subsites as well (see Figure 2.10). These results suggest that there is possibly a subsite preference for a tighter binding of malto-oligosaccharides in MBPs, which is mediated by bulk water along with the protein-substrate H-bonds. Further, it is possible that binding at the subsite S1 may help in aligning larger oligosaccharides in a suitable orientation within the binding pocket.

For tmMBP1 and tmMBP2, which show similar binding to maltotriose and maltotetraose, there are subtle differences in the interactions with residues present in the binding pocket, and a similar magnitude of conformational changes upon encountering ligand. In the case of tmMBP3, however, the binding of substrate did not induce a similar magnitude of conformational changes in solution or in simulations. The computer simulations indicate that the conformational flexibility of these 3 isoforms significantly reduced after binding to the substrates, which has been previously shown for other sugar binding proteins. ⁽⁶²⁾ The degree of conformational flexibility of the protein plays a functional role in the conformational selection associated with binding of the native substrate. An important observation in the current study is that the conformational changes were not limited to the canonical hinge bending motion, but distally located solvent exposed regions were modulated as well in response to ligand binding (Figure 2.10). In simulations, this behavior varied between the apo/substrate bound forms as well as the three isoforms investigated.

It is interesting to note that previous reports ^(20,21) have suggested the presence of allosteric modulation in the broader periplasmic binding protein super-family. ^(23,57) An increasing number of proteins/enzymes are showing that protein dynamics is an important contributor to allosteric regulation. ⁽⁶³⁻⁶⁵⁾ Further studies are needed to characterize the relation, if any, between dynamics and allosteric modulation in MBPs. Such insights will have important implications towards an understanding of the regulation of periplasmic binding protein mediated transmembrane metabolite transport.

Computer simulations provide unique insights into substrates binding by different tmMBP isoforms. The computed interaction energies (based on electrostatics and van der Waals terms) between the protein residues and substrate provide quantitative information about the preferred interactions in the binding pocket. A number of alternate simulations were performed with the substrate in the different sub-sites within the binding pocket. The native substrates which crystallized showed the most stable interactions confirming that the observed structures are the preferred binding sites. Interestingly, the simulations with ligand positions different from the crystallized structures demonstrated non-stable interactions, supporting the combined requirements of correct ligand binding site, dynamics and conformational energetics.

In the broader context, the present study touches upon the interesting interplay among gene duplication and gene specialization in ABC transporters. Yet the more omnipresent question that still remains is, why *E. coli* uses a promiscuous MBP, while *T. maritima* uses redundant multiple copies of similar proteins? One possible reason for presence of multiple genes is that binding of longer, or shorter saccharides is optimized, subtly, in the particular fold (e.g. tmMBP1/2 versus tmMBP3), which is not as efficiently achieved in a single protein such as in ecMBP. Indeed, *T. maritima* utilizes promiscuous periplasmic carbohydrate binding proteins, yet the oligopeptide binding PBP fold is used rather than MBP. ^(16,30,51,66) From an evolutionary perspective the encoding of promiscuity is tied to the evolutionary drive to minimize genome size and gene redundancy. Yet, *E. coli* and *T. maritima* live in significantly different environmental niches, one that fluxes between high and low metabolite availability and the other which is vastly carbohydrate deficient. This in-turn has led to different carbohydrate ABC transport systems in these two organisms,

with *T. maritima* encoding transport operons for a variety of beta linked carbohydrates that are lacking from *E. coli*.⁽³⁰⁾ It is thus possible that due to the carbohydrate limiting environment of *T. maritima*, the energy expenditure of operon redundancy in carbohydrate acquisition outweighs the metabolic benefit of the ability to scavenge saccharides. Collectively the results presented here, and the future studies could provide vital clues to understanding of how the limitations imposed on members of a protein super-family due to a structural fold can be compensated by gene duplication and neofunctionalization.

List of References

1. Lawrence, J. G. (1999) Gene transfer, speciation, and the evolution of bacterial genomes. *Current opinion in microbiology* **2**, 519-523
2. Zhang, J. (2003) Evolution by gene duplication: an update. *Trends in ecology & evolution* **18**, 292-298
3. Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000) Lateral gene transfer and the nature of bacterial innovation. *nature* **405**, 299
4. Mendonça, A. G., Alves, R. J., and Pereira-Leal, J. B. (2011) Loss of genetic redundancy in reductive genome evolution. *PLoS computational biology* **7**, e1001082
5. Shukla, S., Bafna, K., Sundar, D., and Thorat, S. S. (2014) The bitter barricading of prostaglandin biosynthesis pathway: understanding the molecular mechanism of selective cyclooxygenase-2 inhibition by amarogentin, a secoiridoid glycoside from *Swertia chirayita*. *PLoS one* **9**, e90637
6. Recinos, D. A., Sekedat, M. D., Hernandez, A., Cohen, T. S., Sakhtah, H., Prince, A. S., Price-Whelan, A., and Dietrich, L. E. (2012) Redundant phenazine operons in *Pseudomonas aeruginosa* exhibit environment-dependent expression and differential roles in pathogenicity. *Proceedings of the National Academy of Sciences* **109**, 19420-19425
7. Cole, J. (1996) Nitrate reduction to ammonia by enteric bacteria: redundancy, or a strategy for survival during oxygen starvation? *FEMS microbiology letters* **136**, 1-11
8. Hughes, A. L. (1994) The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. Lond. B* **256**, 119-124
9. Qian, W., and Zhang, J. (2014) Genomic evidence for adaptation by gene duplication. *Genome research* **24**, 1356-1362
10. Russell, R. B., Saqi, M. A., Sayle, R. A., Bates, P. A., and Sternberg, M. J. (1997) Recognition of analogous and homologous protein folds: analysis of sequence and structure conservation1. *Journal of molecular biology* **269**, 423-439
11. Cygler, M., Schrag, J. D., Sussman, J. L., Harel, M., Silman, I., Gentry, M. K., and Doctor, B. P. (1993) Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins. *Protein Science* **2**, 366-382

12. Friedberg, I., and Margalit, H. (2002) Persistently conserved positions in structurally similar, sequence dissimilar proteins: roles in preserving protein fold and function. *Protein Science* **11**, 350-360
13. Quioco, F. A., and Ledvina, P. S. (1996) Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Molecular microbiology* **20**, 17-25
14. Dwyer, M. A., and Hellinga, H. W. (2004) Periplasmic binding proteins: a versatile superfamily for protein engineering. *Current opinion in structural biology* **14**, 495-504
15. Quioco, F. A. (1990) Atomic structures of periplasmic binding proteins and the high-affinity active transport systems in bacteria. *Phil. Trans. R. Soc. Lond. B* **326**, 341-352
16. Ghimire-Rijal, S., Lu, X., Myles, D. A., and Cuneo, M. J. (2014) Duplication of genes in an ATP-binding cassette transport system increases dynamic range while maintaining ligand specificity. *Journal of Biological Chemistry* **289**, 30090-30100
17. Vermersch, P. S., Lemon, D. D., Tesmer, J. J., and Quioco, F. A. (1991) Sugar-binding and crystallographic studies of an arabinose-binding protein mutant (Met108Leu) that exhibits enhanced affinity and altered specificity. *Biochemistry* **30**, 6861-6866
18. Seo, M.-H., Park, J., Kim, E., Hohng, S., and Kim, H.-S. (2014) Protein conformational dynamics dictate the binding affinity for a ligand. *Nature Communications* **5**, 3724
19. Quioco, F. A., Spurlino, J. C., and Rodseth, L. E. (1997) Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure* **5**, 997-1015
20. Kadaba, N. S., Kaiser, J. T., Johnson, E., Lee, A., and Rees, D. C. (2008) The high-affinity E. coli methionine ABC transporter: structure and allosteric regulation. *Science* **321**, 250-253
21. Li, L., Ghimire-Rijal, S., Lucas, S. L., Stanley, C. B., Wright, E., Agarwal, P. K., Myles, D. A., and Cuneo, M. J. (2017) Periplasmic Binding Protein Dimer Has a Second Allosteric Event Tied to Ligand Binding. *Biochemistry* **56**, 5328-5337

22. Marinelli, F., Kuhlmann, S. I., Grell, E., Kunte, H.-J., Ziegler, C., and Faraldo-Gómez, J. D. (2011) Evidence for an allosteric mechanism of substrate release from membrane-transporter accessory binding proteins. *Proceedings of the National Academy of Sciences* **108**, E1285-E1292
23. Rizk, S. S., Paduch, M., Heithaus, J. H., Duguid, E. M., Sandstrom, A., and Kossiakoff, A. A. (2011) Allosteric control of ligand-binding affinity using engineered conformation-specific effector proteins. *Nature Structural and Molecular Biology* **18**, 437
24. Rees, D. C., Johnson, E., and Lewinson, O. (2009) ABC transporters: the power to change. *Nature reviews Molecular cell biology* **10**, 218
25. Locher, K. P. (2016) Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nature Structural and Molecular Biology* **23**, 487
26. Orelle, C., Oldham, M. L., and Davidson, A. L. (2014) The maltose ABC transporter: where structure meets function. in *Membrane transport mechanism*, Springer. pp 181-205
27. Marvin, J. S., and Hellinga, H. W. (2001) Conversion of a maltose receptor into a zinc biosensor by computational design. *Proceedings of the National Academy of Sciences* **98**, 4955-4960
28. Bucher, D., Grant, B. J., Markwick, P. R., and McCammon, J. A. (2011) Accessing a hidden conformation of the maltose binding protein using accelerated molecular dynamics. *PLoS computational biology* **7**, e1002034
29. Tang, C., Schwieters, C. D., and Clore, G. M. (2007) Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449**, 1078
30. Nanavati, D. M., Nguyen, T. N., and Noll, K. M. (2005) Substrate specificities and expression patterns reflect the evolutionary divergence of maltose ABC transporters in *Thermotoga maritima*. *Journal of bacteriology* **187**, 2002-2009
31. Boucher, N., and Noll, K. M. (2011) Ligands of thermophilic ABC transporters encoded in a newly sequenced genomic region of *Thermotoga maritima* MSB8 screened by differential scanning fluorimetry. *Applied and environmental microbiology* **77**, 6395-6399

32. Wassenberg, D., Liebl, W., and Jaenicke, R. (2000) Maltose-binding protein from the hyperthermophilic bacterium *Thermotoga maritima*: stability and binding properties¹. *Journal of molecular biology* **295**, 279-288
33. Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallographica Section D: Biological Crystallography* **62**, 859-866
34. Franke, D., Petoukhov, M., Konarev, P., Panjkovich, A., Tuukkanen, A., Mertens, H., Kikhney, A., Hajizadeh, N., Franklin, J., and Jeffries, C. (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of applied crystallography* **50**, 1212-1225
35. Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of applied crystallography* **25**, 495-503
36. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophysical journal* **80**, 2946-2953
37. Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *Journal of applied crystallography* **36**, 860-864
38. Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high-and low-resolution structural models. *Journal of applied crystallography* **34**, 33-41
39. Svergun, D., Barberato, C., and Koch, M. H. (1995) CRY SOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of applied crystallography* **28**, 768-773
40. Case, D. A., Berryman, J. T., Betz, R. M., Cerutti, D. S., Cheatham III, T. E., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Götz, A. W., Homeyer, N., S., I., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Luchko, T., Luo, R., Madej, B., Merz, K. M., Monard, G., Needham, P., H., N., Nguyen, H. T., Omelyan, I., Onufriev, A., Roe, D. R., Roitberg, A., Salomon-Ferrer, R., Simmerling, C. L., Smith, J., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu,

- X., York, D. M., and Kollman, P. A. (2015) AMBER 15. *University of California, San Francisco*.
41. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation* **11**, 3696-3713
 42. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *Journal of computational chemistry* **29**, 622-655
 43. Ramanathan, A., Agarwal, P. K., Kurnikova, M., and Langmead, C. J. (2010) An online approach for mining collective behaviors from molecular dynamics simulations. *J Comput Biol* **17**, 309-324
 44. Ramanathan, A., Savol, A. J., Agarwal, P. K., and Chennubhotla, C. S. (2012) Event detection and sub-state discovery from biomolecular simulations using higher-order statistics: application to enzyme adenylate kinase. *Proteins* **80**, 2536-2551
 45. Ramanathan, A., and Agarwal, P. K. (2011) Evolutionarily Conserved Linkage between Enzyme Fold, Flexibility, and Catalysis. *PLoS Biology* **9**, e1001193
 46. Beck, D. A., and Daggett, V. (2004) Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods* **34**, 112-120
 47. Ramanathan, A., and Agarwal, P. K. (2009) Computational identification of slow conformational fluctuations in proteins. *The Journal of Physical Chemistry B* **113**, 16669-16680
 48. Agarwal, P. K. (2004) Cis/trans isomerization in HIV-1 capsid protein catalyzed by cyclophilin A: insights from computational and theoretical studies. *Proteins* **56**, 449-463
 49. Gagne, D., Narayanan, C., Nguyen-Thi, N., Roux, L. D., Bernard, D. N., Brunzelle, J. S., Couture, J. F., Agarwal, P. K., and Doucet, N. (2016) Ligand Binding Enhances Millisecond Conformational Exchange in Xylanase B2 from *Streptomyces lividans*. *Biochemistry* **55**, 4184-4196

50. Cuneo, M. J., Changela, A., Warren, J. J., Beese, L. S., and Hellinga, H. W. (2006) The crystal structure of a thermophilic glucose binding protein reveals adaptations that interconvert mono and di-saccharide binding sites. *Journal of molecular biology* **362**, 259-270
51. Cuneo, M. J., Changela, A., Beese, L. S., and Hellinga, H. W. (2009) Structural adaptations that modulate monosaccharide, disaccharide, and trisaccharide specificities in periplasmic maltose-binding proteins. *Journal of molecular biology* **389**, 157-166
52. Millet, O., Hudson, R. P., and Kay, L. E. (2003) The energetic cost of domain reorientation in maltose-binding protein as studied by NMR and fluorescence spectroscopy. *Proc Natl Acad Sci U S A* **100**, 12700-12705
53. Bucher, D., Grant, B. J., and McCammon, J. A. (2011) Induced fit or conformational selection? The role of the semi-closed state in the maltose binding protein. *Biochemistry* **50**, 10530-10539
54. Narayanan, C., Bernard, D. N., Bafna, K., Gagné, D., Chennubhotla, C. S., Doucet, N., and Agarwal, P. K. (2018) Conservation of dynamics associated with biological function in an enzyme superfamily. *Structure* **26**, 426-436. e423
55. Gagné, D., Narayanan, C., Bafna, K., Charest, L.-A., Agarwal, P. K., and Doucet, N. (2017) Sequence-specific backbone resonance assignments and microsecond timescale molecular dynamics simulation of human eosinophil-derived neurotoxin. *Biomolecular NMR assignments* **11**, 143-149
56. Agarwal, P. K., Schultz, C., Kalivretanos, A., Ghosh, B., and Broedel Jr, S. E. (2012) Engineering a Hyper-catalytic Enzyme by Photoactivated Conformation Modulation. *The Journal of Physical Chemistry Letters* **3**, 1142-1146
57. Telmer, P. G., and Shilton, B. H. (2003) Insights into the conformational equilibria of maltose-binding protein by analysis of high affinity mutants. *Journal of Biological Chemistry* **278**, 34555-34567
58. Lawson, D. M., Williams, C. E., Mitchenall, L. A., and Pau, R. N. (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

59. Locher, K. P. (2004) Structure and mechanism of ABC transporters. *Current opinion in structural biology* **14**, 426-431
60. Diez, J., Diederichs, K., Greller, G., Horlacher, R., Boos, W., and Welte, W. (2001) The crystal structure of a liganded trehalose/maltose-binding protein from the hyperthermophilic archaeon *Thermococcus litoralis* at 1.85 Å. *Journal of molecular biology* **305**, 905-915
61. Quijcho, F., Wilson, D. K., and Vyas, N. (1989) Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature* **340**, 404
62. Gagné, D., Narayanan, C., Nguyen-Thi, N., Roux, L. D., Bernard, D. N., Brunzelle, J. S., Couture, J.-F., Agarwal, P. K., and Doucet, N. (2016) Ligand binding enhances millisecond conformational exchange in xylanase B2 from *Streptomyces lividans*. *Biochemistry* **55**, 4184-4196
63. Bouskila, M., Hunter, R. W., Ibrahim, A. F., Delattre, L., Pegg, M., Van Diepen, J. A., Voshol, P. J., Jensen, J., and Sakamoto, K. (2010) Allosteric regulation of glycogen synthase controls glycogen synthesis in muscle. *Cell metabolism* **12**, 456-466
64. Freiburger, L. A., Baettig, O. M., Sprules, T., Berghuis, A. M., Auclair, K., and Mittermaier, A. K. (2011) Competing allosteric mechanisms modulate substrate binding in a dimeric enzyme. *Nature Structural and Molecular Biology* **18**, 288
65. Dagliyan, O., Tarnawski, M., Chu, P.-H., Shirvanyants, D., Schlichting, I., Dokholyan, N. V., and Hahn, K. M. (2016) Engineering extrinsic disorder to control protein activity in living cells. *Science* **354**, 1441-1444
66. Munshi, P., Stanley, C. B., Ghimire-Rijal, S., Lu, X., Myles, D. A., and Cuneo, M. J. (2013) Molecular details of ligand selectivity determinants in a promiscuous beta-glucan periplasmic binding protein. *BMC Struct Biol* **13**, 18

**Chapter 3 Accessing the substrate-induced dynamical
changes in maltose-binding protein
isoforms using neutron scattering**

3.1 Abstract

Functional segregation among protein isoforms depends on the interplay of their overall structures and the dynamics of these structures. *Thermotoga maritima* maltose binding proteins (tmMBPs) isoforms show size-dependent differential binding of maltose and malto-oligomers. Understanding the effects of protein dynamics on this differential ligand binding and recognition behavior requires detailed characterization in native like aqueous conditions. To investigate the structural and dynamical behavior of tmMBPs experimentally, small-angle neutron scattering (SANS), neutron spin echo (NSE) spectroscopy and dynamic light scattering (DLS) were used in conjunction with computational molecular dynamics simulation. SANS provides information on the overall structure of the molecules in solution, while NSE was used to determine the dynamics at various time and length scales. Both tmMBP2 and tmMBP3 have a bidomain architecture linked with a flexible hinge that forms the binding pocket in the cleft between the two domains. We find that tmMBP2 and tmMBP3 showed distinctly different solution dynamics, with the translational and rotational components dominating the dynamics of both systems resulting in a clear differentiation of their diffusion. Additionally, we find an integral component of internal dynamics that differs for the two isoforms. Moreover, significant differences were observed between the ligand-free (apo) and ligand-bound (holo) states of both proteins. This study demonstrates the ability of neutron scattering to determine dynamical differences between structurally related proteins.

3.2 Introduction

The evolution of a new fold space in proteins is estimated to take a billion years. ⁽¹⁾ Therefore, protein structures are finely tuned to achieve an astounding diversity in function within the confinement of functional protein folds. ⁽²⁾ The evolution of function in a constrained fold space is often brought about by redecorating the binding or active sites of the protein molecule. ⁽³⁾ Additionally, functional segregation between closely related proteins and their isoforms can also be achieved by tuning protein dynamics to match the relevant timescales of cellular activities. ⁽⁴⁾ These alterations in structure and dynamics help organisms circumvent the large evolutionary cost of generating a new fold for each new function. ^(5,6) Many proteins adopt modular domain architecture connected by flexible

regions that facilitate large scale dynamics, enabling functional domain-domain interactions. The timescales of these interdomain interactions govern diverse protein functions, including protein-ligand interaction. ⁽⁷⁻⁹⁾ Depending on the inherent protein function, these large-scale domain motions can range in timescales from picoseconds up to milliseconds. ⁽¹⁰⁻¹²⁾ Current methods that methods used to determine structures of macromolecules provide little direct insight into their dynamical behavior. This gap is addressed using a combination of computational, and conventional biochemical and biophysical techniques that are limited in resolution that are beginning to let us fully appreciate the extent of protein motions at biologically relevant timescales and conditions.

Since most proteins are associated with aqueous environments, it remains essential to study them in their native or near-native conditions. Therefore, quantifying the effects of different factors then influence protein structure, dynamics, and overall function is thus critically important. Solution scattering experiments can help decipher different types of dynamical motions associated with a protein molecule, including fluctuations in protein backbone and sidechain residues, domain-domain motions, and information on global rotational, translational, and internal motions. ⁽¹³⁾ However, while the time-averaged data obtained by diffraction or SAXS/SANS experiments give an excellent overview of the structural organization of a protein molecule or ensemble in solution, , specific information on their local and internal time-dependent dynamics is more difficult to obtain (Figure 3.1.).⁽¹⁴⁾ Without this dynamical information, individual protein structures often cannot be comprehensively used for interpretation of the protein functions. ⁽¹⁵⁾ However, recent advances in the coherent neutron scattering technique of neutron spin echo (NSE) spectroscopy offers an excellent tool to study protein dynamics in near native-like conditions. ^(16,17) Since NSE captures protein dynamics in the picosecond to nanosecond timescales, these experiments also provide an excellent benchmark for comparison with data from molecular dynamics (MD) simulations. ^(18,19) Additionally, incoherent neutron scattering techniques, like quasi-elastic neutron scattering, can provide details on the influence of hydration on local dynamics of protein side chains. ^(20,21) The time and length-scales covered by some of the experimental scattering techniques is summarized in Figure 3.1.

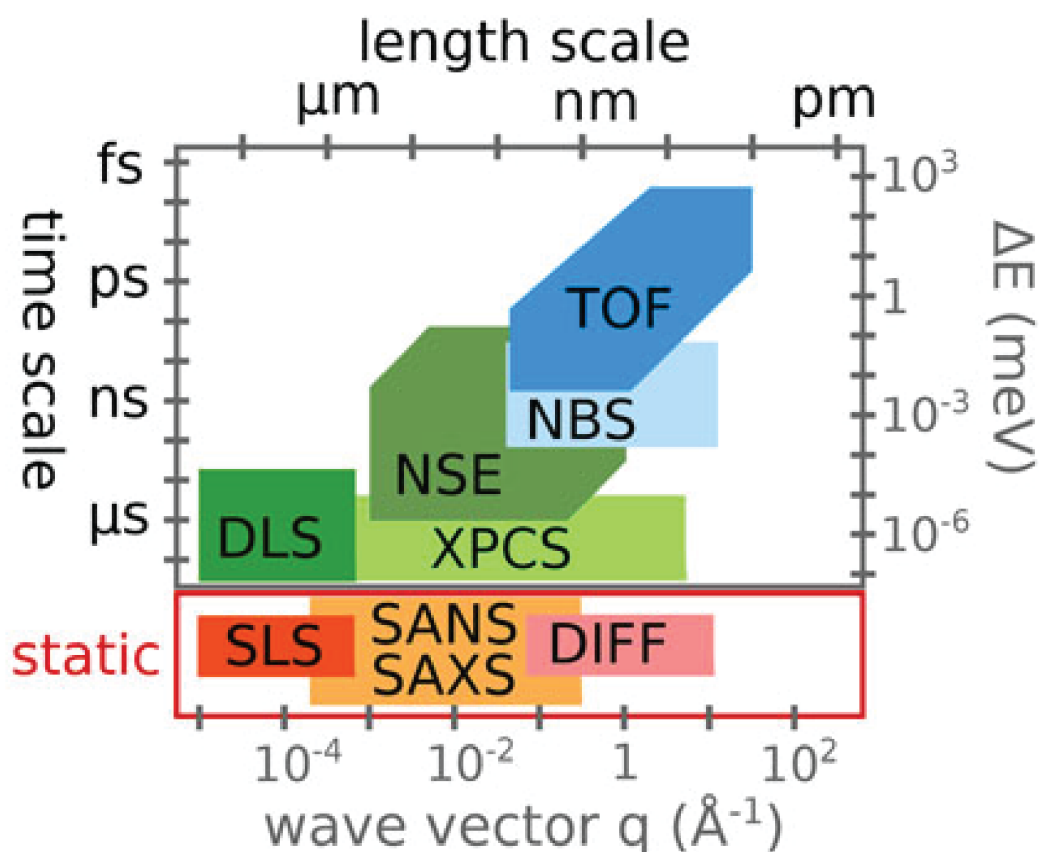


Figure 3.1: Various scattering techniques and their time-scale and length-scale windows. The static techniques include static light scattering (SLS), small-angle X-ray/neutron scattering (SAXS/SANS), and diffraction (DIFF) that provide a time-averaged structure of proteins at different length-scales given by the wave vector q in inverse angstroms. On the contrary, the dynamic scattering techniques, such as dynamic light scattering (DLS), X-ray photon correlation spectroscopy (XPCS), neutron-spin echo (NSE) spectroscopy, neutron backscattering spectroscopy (NBS), and neutron time-of flight (TOF) spectroscopy, can detect changes in the energy (ΔE) with varying sensitivity and can provide information on time-scales ranging from femtoseconds (fs) to microseconds (μ s), and length-scales from picometers (pm) to micrometers (μ m). Figure reference (10).

Maltose-binding proteins (MBPs) represent a subclass of the periplasmic substrate binding proteins that are associated with the maltose ATP-binding cassette (ABC) transporter system.⁽²²⁾ MBPs are very specific towards their substrates (α -linked glucose-based oligomers) with binding affinities ranging in higher nanomolar to low micromolar levels.⁽²³⁾ In mesophilic bacteria, like the *Escherichia coli*, there is a single promiscuous MBP protein that binds with varying specificity to maltose and its oligomers.⁽²³⁾ *Thermotoga maritima*, an extreme-thermophilic bacterium, was shown to contain three copies of the MBP.⁽²⁴⁾ The three isoforms: tmMBP1, tmMBP2 and tmMBP3 have distinct periplasmic binding protein (PBP)-like $\alpha\beta$ fold and a bidomain architecture. tmMBP1 (391 residues) and tmMBP2 (393 residues) share 90% sequence identity, whereas tmMBP3 shares less than 40% sequence identity with both tmMBP1 and tmMBP2.⁽²⁵⁾ It was initially identified that tmMBP1, tmMBP2 and tmMBP3 show differential promiscuous binding of maltose and malto-oligomers with varying affinity.^(24,26) However, our recent work suggested that tmMBP1 and tmMBP2 can bind to substrates that are triose sugar or larger oligomers whereas tmMBP3 has affinity for only maltose.⁽²⁵⁾ Additionally, using MD simulations we predicted that the dynamics of tmMBP1 and tmMBP2 showed distinct similarity but differed significantly with that of the tmMBP3.⁽²⁵⁾ Therefore, there is a need to identify how the underlying role of dynamics may contribute to or determine substrate affinity and specificity profiles in these structurally similar isoforms.

Here, the experimental techniques of dynamic light scattering (DLS), small-angle neutron scattering (SANS), neutron spin echo (NSE), and quasi-elastic neutron scattering (QENS) in conjunction with computational MD simulations were used to address the differences in the dynamical landscapes of isoforms tmMBP2 and tmMBP3. Data on both these systems was collected in absence (apo) and presence (holo) of their cognate substrates. SANS was used to determine the overall shape of the proteins and DLS and NSE were used to determine translation related diffusion and overall dynamics. QENS gave a detailed picture of how the protein motion changes locally at the atomic level. MD simulations provide a model that can be compared to the signal profiles from NSE and QENS to provide details on the predominant motions governing the protein-ligand interaction in these protein isoforms. Together, these techniques have helped differentiate the collective

motions of the structurally related isoforms of tmMBPs and present a potentially new method for detecting protein dynamics experimentally for structurally related proteins.

3.3 Materials and Methods

Protein Expression and Purification:

Both tmMBP2 and tmMBP3 were expressed and purified as per the protocol given in ⁽²⁵⁾ with a few modifications. Briefly, both isoforms were grown in 3L of Enfor's minimal media ⁽²⁷⁾ in a fed-batch bioreactor at 32°C until the O.D. of 20 was reached. Multiple bioreactor runs were combined to produce the cell mass required for gram-level protein production. The cells were then induced using 1M IPTG and grown at 25°C for overnight induction. Cell disruption by sonication was followed by incubation of the lysate at 80°C for 15 minutes to remove contaminating *Escherichia coli* proteins. The lysate was spun down, and the supernatant was subjected to a two-step purification process involving an His-tag affinity purification followed by gel filtration. The collected fractions of purified protein were concentrated to approximately 51 mg/ml and were buffer exchanged with gel filtration buffer (20mM Tris pH 7.5 and 40mM NaCl) containing D₂O. Dried powder samples were separately prepared for the QENS experiment. The details of QENS sample preparation are provided later in section 3.3.5.

Small-angle Neutron Scattering (SANS):

SANS experiments were performed at the EQ-SANS beamline (BL-6), a time-of-flight SANS instrument, at the Spallation Neutron Source (SNS), Oak Ridge National Laboratory (ORNL). ⁽²⁸⁾ A series of different dilutions of the dialyzed tmMBP2 and tmMBP3 were used for SANS experiments (1mg/ml, 5mg/ml, 10mg/ml, 20mg/ml, 50mg/ml). The generated data was background corrected using D₂O dialysis buffer. All the datasets were collected at 323 K to match the NSE measurement. The Q-range explored during the SANS experiment was from $\sim 0.01 \text{ \AA}^{-1}$ to $\sim 0.60 \text{ \AA}^{-1}$. The relative scattering intensities (I) as a function of momentum transfer vector q ($q = 4\pi \sin\theta/\lambda$) was calculated by circular averaging of the scattered data. ATSAS ⁽²⁹⁾ was used for the I_0 the $P(r)$ analysis. Form factor per unit mass $F(q)$ was estimated by scaling the background-corrected scattering data and extrapolating it to the zero concentration for each q -value. The form factor gives

information about the shape of the molecule. The structure factor, $S(q)$, on the other hand describes the protein-protein interaction at higher concentrations, which was calculated by subtracting $F(q)$ from the concentration scaled scattering data using the SASView program (<http://www.sasview.org/>). SANS was performed on the same samples that were later used for NSE experiment to maintain data consistency.

Dynamic Light Scattering

DLS experiments were performed using DyanoPro NanoStar (Wyatt Technology Corporation) at a scattering angle of 90° . 100 μ l samples of dilutions ranging from 1mg/ml to 50mg/ml were used for the experiment. The samples were centrifuged and filtered prior to the experiment to remove large aggregates. CONTIN algorithm ⁽³⁰⁾ in the DYNAMIC software (Wyatt Technology Corporation) was used to analyze the data and to identify polydispersity and molecular mass. All measurements were performed at 323K (50°C).

Neutron Spin Echo (NSE) Spectroscopy

NSE experiments were performed at the NSE spectrometer (Beamline-15), JCNS1 outstation at SNS, ORNL. ⁽³¹⁾ Hellma® Quartz cells of 40 x 30 mm² and a path length of 4mm were used to store ~4ml of sample. Three different incident wavelengths of 8Å, 9Å and 11Å were used to measure the NSE spectra with a dynamical range of $0.1 < t_{\text{max}} < 100$ ns (Fourier times) for q values ranging from 0.03 Å⁻¹ to 0.21 Å⁻¹. Data was also collected on dialysis buffer prepared in D₂O for buffer subtraction. Sample concentration measured for both tmMBP2 and tmMBP3 was ~51 mg/ml. All the experiments were performed in an enclosed sample environment with a constant temperature of 323 K (50 °C). No sample aggregation or degradation was observed after a prolonged data collection time that extended several days. Data reduction were performed using DRSPINE software package of the NSE instrument. ⁽³²⁾

Quasi-elastic neutron scattering (QENS)

The protein was concentrated and extensively dialyzed in a dialysis buffer (10mM Tris, 10mM NaCl, pH 7.5). Dry powder of both apo and holo forms of tmMBP2 and tmMBP3 were prepared by overnight lyophilization of a concentrated protein solution (~60 mg/ml). This resulted in ~1.5 gram of dry protein per protein sample (~3 grams each for tmMBP2

and tmMBP3). Three different samples were prepared from the lyophilized protein, (1) dried protein sample, (2) sample rehydrated with D₂O, and (3) sample rehydrated using H₂O. The dried protein powder was subsequently rehydrated using vapor diffusion against 100% D₂O inside an anaerobic chamber. The protein powder was weighed every 30 minutes until it was fully saturated with D₂O. Similarly, rehydrated samples were also prepared by vapor exchanging the protein powder against H₂O. Finally, all the samples were carefully wrapped in aluminum foil and transferred to the high-pressure sample holders. The sample cells were thoroughly sealed using an indium gasket-seal. The QENS dataset was collected on BASIS spectrometer at SNS, ORNL ⁽³³⁾ for a q-range between 0.3 Å⁻¹ to 1.9 Å⁻¹. The instrument was operating on a bandwidth chopper of 30 Hz and an incident neutron band center at 6.15 Å. The neutron energy transfer was analyzed in the range of -100 µeV to 500 µeV. All the datasets were collected at 323 K.

Molecular Dynamics (MD) Simulation

Details of the system preparation and MD simulation protocol used have been discussed previously in ⁽¹²⁾. Briefly, simulations were performed on tmMBP2 and tmMBP3 in apo and holo states using AMBER version 14 suite in explicit water solvent. ⁽³⁴⁾ SPC/E water model was used to solvate a periodic rectangular box, with 10 Å distance between protein and the edge of the box. System equilibration protocol has been previously described. ⁽³⁵⁾ Final production run of 500 ns was performed under constant-energy conditions (NVE ensemble). All the simulations were performed at 323 K to match the experiments. A total of 10000 conformational snapshots from last 100ns of simulation time were used for final analysis and coherent intermediate scattering function calculations. Analysis of trajectories was performed using CPPTRAJ. ⁽³⁶⁾ The intermediate scattering functions were calculated using the in-house python script that used functions from the MDAnalysis program ⁽³⁷⁾ for coherent intermediate scattering function calculations.

Quasi-anharmonic analysis (QAA)

An all-atom QAA analysis was performed on 50000 frames extracted from 500 nanosecond (ns) trajectories of tmMBP2-apo, tmMBP2-MTT, tmMBP3-apo, and tmMBP3-MAL. Top three QAA vectors were scrutinized for differences in the apo vs. substrate-bound states of both tmMBP isoforms. Large numbers of frames were used to ensure

reproducibility of results and only protein atomic positions were considered for the calculation of modes. The fourth-order statistics implemented in QAA helps differentiation and assignment of independent vectors to each transient substate within a state, thereby capturing even rare dynamical transitions that are otherwise not scored. QAA has been employed to identify critical structural and dynamical features in proteins that are directly tied to their function. ^(38,39)

3.4 Results

3.4.1 Solution-state structure of tmMBPs

SANS experiments were performed on both apo and holo forms of tmMBP isoforms with a protein concentration gradient ranging from 1 – 50 mg/ml in D₂O buffer. Subsequently, data collected on the buffer was subtracted out to account for only corrected protein intensities.

Prior to the structure factor calculation, the SANS intensity data were fitted against the theoretically derived intensities of published crystal structures (described in Chapter 2) of tmMBP2 (PDB: 6DTT and 6DTS) and tmMBP3 (PDB: 6DTR and 6DTQ) in both apo and holo forms, respectively. The solution scattering intensities and the theoretical intensities were in agreement, suggesting a good representation of the solution state of the tmMBP isoforms by their crystalline forms (Figure 3.2.A). Subsequently, the calculated $P(r)$ function of the tmMBP dataset showed variation in the radius of gyration (R_g) in the apo and the substrate-bound forms of both tmMBP isoforms. As expected, the difference was significant larger between tmMBP2-apo vs. tmMBP2-MTT than between the apo vs. substrate-bound forms of tmMBP3 (Figure 3.2.B). These results agree with our previous $P(r)$ estimated from SAXS (presented in Chapter 2) and suggested that tmMBP2 is much more flexible than tmMBP3. ⁽²⁵⁾

A SANS analysis is usually performed prior to collecting NSE datasets to provide information on the form factor $F(Q)$ and concentration-dependent structure factor $S(Q)$, which can be extracted from the scattering intensity $I(Q)$. The form factor provides information on the shape of the molecule, whereas the structure factor gives details of

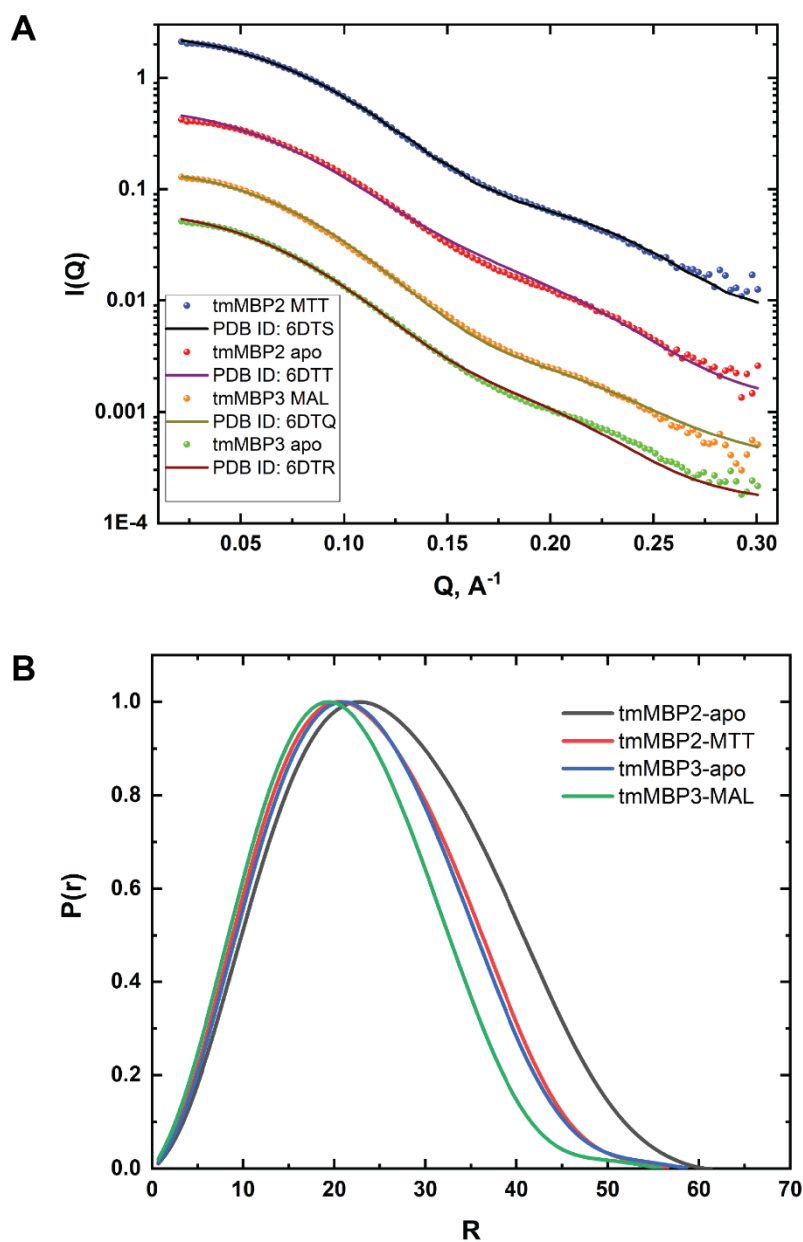


Figure 3.2: Small-angle neutron scattering (SANS) results of hydrogenated tmMBP2 and tmMBP3 in D_2O . (A) An intensity vs. Q plot of tmMBP samples with their structure fits shown as solid lines. (B) A normalized $P(r)$ analysis of tmMBP proteins in the presence and absence of their cognate substrates.

protein-protein interaction. Since, protein-protein interaction can be modulated depending on the vicinity of a nearby protein, $S(Q)$ is dependent on the concentration of the protein in the solution. This information is important since the time-averaged $S(Q)$ from SANS is used as the initial structure factor at time $t=0$, such that averaged $S(Q)$ from SANS becomes $S(Q,0)$ for NSE. $S(Q)$ was determined when the intensities from 1 mg/ml and 50 mg/ml protein datasets were overlapped and fitted with the $F(Q)$ from the lower concentration. This $F(Q)$ value, corresponding to the structure of the protein in the solution, is subtracted to produce the estimation of concentration dependent $S(Q)$. The higher limit of concentration for NSE experiment is chosen based on the $S(Q)$ value, as it can suggest at which concentration does protein-protein interaction start dominating the $I(Q)$.

The calculated structure factors were subsequently fitted with the “hardsphere” and “stickyhardsphere” models in SASView (Figure 3.3).⁽⁴⁰⁾ tmMBP2 apo and substrate-bound $S(Q)$ were fitted using the hardsphere model. This model best represents a low molecular interaction among particles thereby allowing a faster diffusion of the particle even in a very crowded environment.^(41,42) On the other hand, the stickyhardsphere model was fitted onto the structure factor data from tmMBP3-apo and tmMBP-MAL samples. The stickyhardsphere model calculates the hardsphere model with a narrow, attractive, potential well. The extent of “stickiness” of a particle is determined by the strength of its attractive well.^(41,43,44) Close inspection of the structure factor further indicated not just very different protein-protein interaction behaviors of both isoforms, but also showed that there was a large difference in the structure factor of tmMBP2 apo vs. substrate-bound forms, possibly attributed to the more flexible nature of the apo protein. No observable difference in structure factor between the apo and holo states of tmMBP3 were noted. These results were also consistent with our previous findings on the flexibility of both isoforms.⁽²⁵⁾

3.4.2 Picosecond – nanosecond dynamics of tmMBP isoforms using NSE

The NSE experiments were performed on apo and substrate-bound samples of tmMBP2 and tmMBP3 to quantify their picosecond to nanosecond solution-state dynamics in near-native, non-invasive conditions. The dynamic structure factor $S_i(Q,t)/S_i(Q,0)$ was measured out to $t \sim 100$ ns. Data were collected on each sample over a Q -range of $0.03 \text{ \AA}^{-1}$ to $0.21 \text{ \AA}^{-1}$. Each protein sample was concentrated to ~ 50 mg/ml and the data

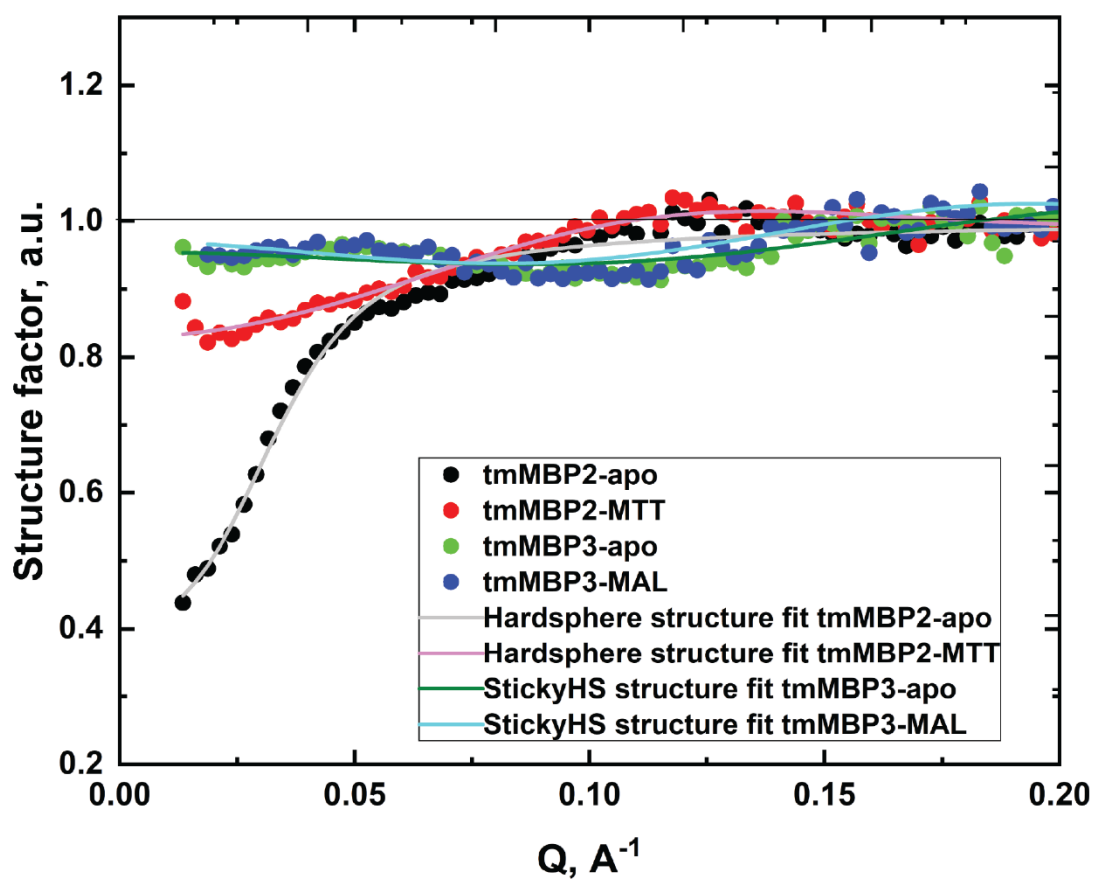


Figure 3.3: Structure factor analysis of tmMBPs. The structure factor calculated for tmMBP2 (apo and MTT-bound forms) and tmMBP3 (apo and MAL-bound forms) is shown here along with the fitted structure factor models of hardsphere and stickyhardsphere (StickyHS), respectively

collection temperature was set at 323 K to keep the results consistent with other experiments. Prior to collecting datasets on each sample, the equipment was equilibrated and tested for polarization and for average signal on the detector. The low-Q region of the spectra contains largely coherent signals whereas in the high-Q regions the signal is mainly constituted by incoherent scattering. In this study, the NSE dataset is collected till the coherent-incoherent transition point, which falls around a Q-value of $0.2 \text{ \AA}^{-1}$.

The dynamic structure factor $S(Q,t)/S(Q,0)$ was measured at various Q-values for the tmMBP isoforms for a total time of $t \sim 100 \text{ ns}$. Figure 3.4.A highlights three Q-values each representing low-Q (0.05), intermediate-Q (0.1), and high-Q regions (0.15). As we move from low-Q towards the high-Q, we transition from diffusive/translational motions towards rotational motion in the intermediate Q-ranges and finally to the high-Q regions dominated by internal dynamics of the proteins. At all these q-values, we can observe faster signal decays while going from low-Q to high-Q region. Additionally, since the Q is the inverse of the distance (in angstroms), it is a measure of the length-scales being observed. The slow dynamics of the low-Q region are contributed by the slower diffusive motions, which can take several nanoseconds to milliseconds, and are observed for the entire molecule. The intermediate-Q range shows a faster decay compared to the low-Q diffusive motions, since this Q-region is dominated by rotational motions of the protein. Depending on the size of the protein being studied, one can selectively screen the dynamics of an individual domain of a protein. Finally, the high-Q region represents a more “zoomed-in” view of the protein structure that reports on the flexibility of a particular secondary structure element in the target protein. Together, the signal decay at various Q-values broadly suggests that the sample is highly dynamic. However, the signal decay is much faster at higher-Q values indicating that secondary structure elements show much faster dynamics. In case of tmMBP isoforms, all the samples showed similar signal decay for individual Q-values suggesting that both apo and holo forms of these proteins exhibit significant dynamical profiles.

The effective diffusion (D_{eff}) measures the decay in the NSE signal as a function of the length scale or Q. ⁽⁴⁵⁾ It encompasses translational and rotational motions, which are directly affected by the interparticle distance, as well as the contributions from the internal

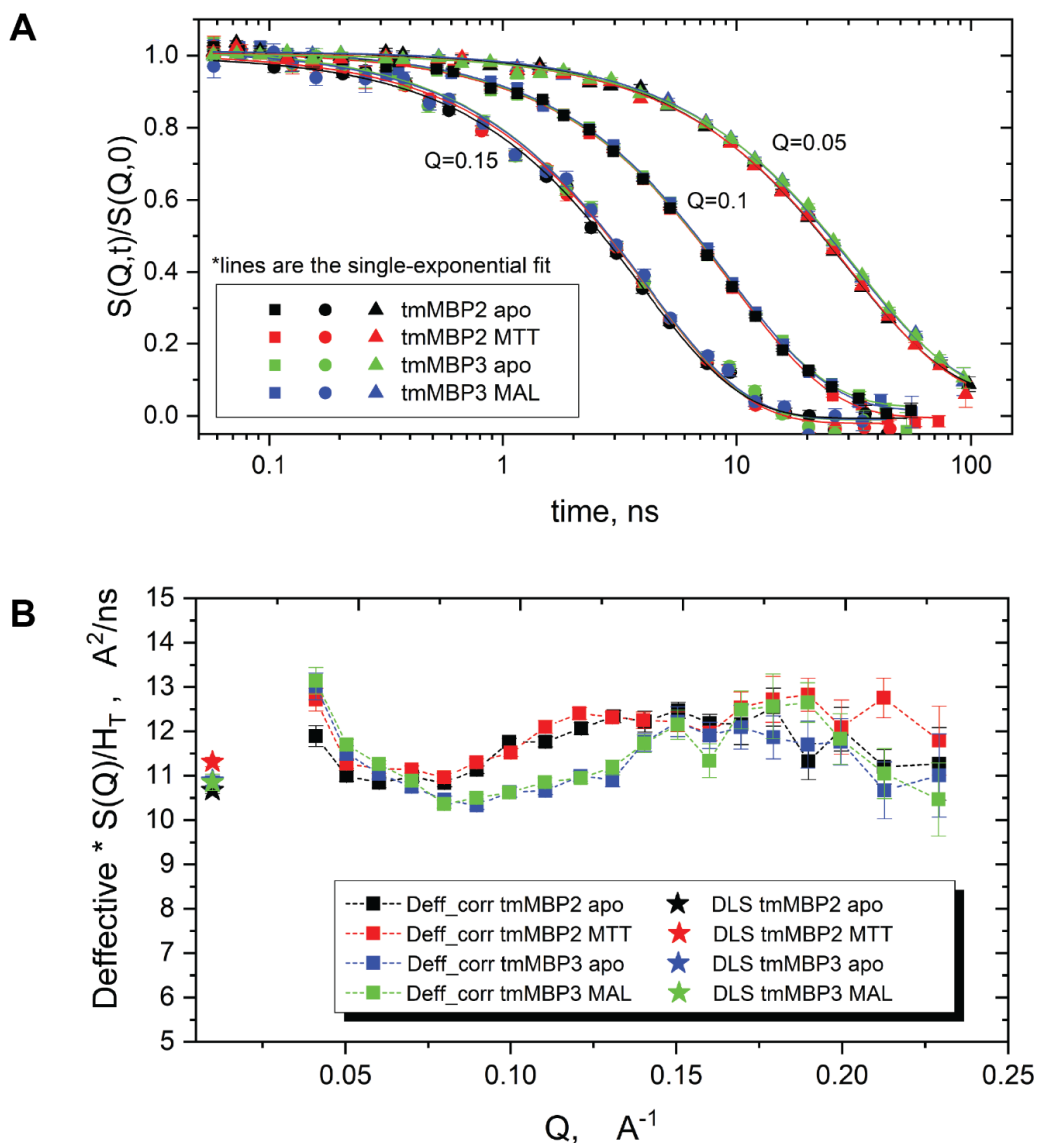


Figure 3.4: Neutron spin echo (NSE) spectral data on tmMBP isoforms. (A) Intermediate scattering function $I(Q,t)$ for low ($Q = 0.05$), intermediate ($Q = 0.1$), and high ($Q = 0.15$) Q -range variation with time in nanosecond (ns). (B) Effective diffusion of tmMBPs as a function of the Q -range, corrected by effective diffusion derived from DLS experiments.

dynamical modes of the protein. Therefore, this is perhaps the most important measure that can help differentiate the contributions of individual motion, either global or local based on the Q , in a protein. It is usually indispensable when comparing two structurally related isoforms that are functionally related. D_{eff} is also influenced by the interparticle interaction, given by $S(Q)$ from SANS, and hydrodynamic interactions (H_T), which is determined from DLS experiments.

For tmMBP2 and tmMBP3, the effective diffusion analysis showed a significant difference between the intermediate and high- Q regions. Around the Q -range of $0.08 \text{ \AA}^{-1}$ to $0.15 \text{ \AA}^{-1}$, which comprises of signals from the rotational and internal dynamics regime, both tmMBP2-apo and tmMBP2-MTT showed much faster diffusion when compared to the apo and maltose-bound tmMBP3. This clearly shows that the dynamical profile of both isoforms is remarkably different from one another. Even though the data higher than $Q=0.15 \text{ \AA}^{-1}$ was noisy, there is strong indication that the apo and the substrate-bound forms of tmMBP2 had divergent internal dynamics. On the contrary, large differences in the internal dynamics of apo vs. substrate-bound forms of tmMBP3 were not observed. This further supports the structure factor analysis results on the two isoforms presented above and suggests that internal motions on these scales cease in tmMBP2-apo after substrate-binding. Additionally, similar trends in the internal dynamics of the two isoforms was also accounted for in our previously published work (see Chapter 2 for details).⁽²⁵⁾ Taken together, these results suggest that despite structural similarity, the isoforms are distinguished by differential dynamical modes that may underly or contribute to functional segregation.

3.4.3 Determination of water-mediated tmMBP-substrate interaction dynamics in using QENS

Quasi-elastic spectra were determined using the backscattering spectrometer BASIS at the SNS.⁽⁴⁶⁾ The dynamic structure factor $S(Q, \Delta E)$ was measured for $t \sim 60$ ps. All the samples were measured at 323 K. The crystal structures of the tmMBP isoforms show the presence of water at the substrate-binding site, which either directly forms H-bond interactions or mediate the same between protein residues and the substrate. Therefore,

it was important to examine whether changes in local dynamics of the protein in the presence and absence of hydration water contributes or facilitates substrate binding. For this experiment, dry and hydrated samples, both in H₂O and D₂O, were prepared and loaded onto a sample holder and sealed with an indium gasket. The dried sample measurements were used to account for the methyl fluctuations, the hydrogenated sample was used to account for the dynamics of water, and the deuterated sample for observing protein dynamics. Figure 3.5.A shows the susceptibility plot that depicts protein dynamics induced by hydration water. The Q-dependence was used to determine the scattering pattern from individual samples. The H₂O sample shows a translation-type Q-dependence, indicating that this represents the dynamical contribution from hydration water. Scattering from the D₂O hydrated sample on the other hand more closely resembled that of the dry sample, albeit with an enhanced Q-dependent signal. Therefore, this confirms that the QENS signal arises predominantly from the hydrogenated protein, and not the hydration D₂O.

Analysis using a conventional Lorentzian fit proved inadequate to observe dynamical transition profile of the tmMBP isoforms in D₂O. Generally, a resolution-convolved superimposition of delta function at zero energy transfer with a model scattering function and a background function is used to fit the QENS data, such that:

$$I(E) = [x\delta(E) + (1 - x)S(E)] \otimes R(E) + (C_1E + C_2) \quad 1$$

A simple forced Lorentzian fit is used most commonly as the model scattering function for fitting QENS data and is given as:

$$S(E) = \frac{1}{\pi} \frac{\Gamma}{E^2 + \Gamma^2} \quad 2$$

a Lorentzian (Eq. 2) is a limiting case of a Cole-Cole (Eq. 3) when the parameter alpha approaches zero. Cole-Cole is thus just a more general distribution, of which a Lorentzian (Γ) is a limiting case (for alpha=0).

$$S(E) = \frac{1}{\pi E_0} \frac{(E/E_0)^{-\alpha} \cos \frac{\pi\alpha}{2}}{1 + 2(E/E_0)^{1-\alpha} \sin \frac{\pi\alpha}{2} + (E/E_0)^{2(1-\alpha)}} \quad 3$$

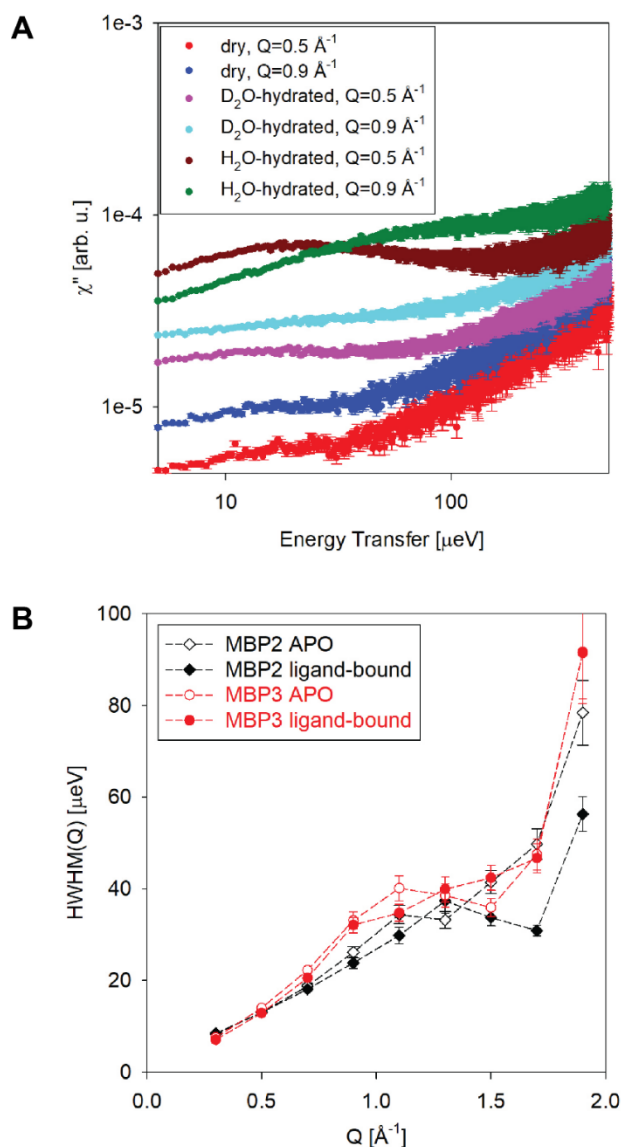


Figure 3.5: Quasi-elastic neutron scattering data on tmMBP isoforms. (A) QENS susceptibility plot showing variation in the signal from the dried, H_2O -hydrated, and D_2O -hydrated tmMBP2 sample at $Q=0.5 \text{ \AA}^{-1}$ and $Q=0.9 \text{ \AA}^{-1}$. (B) A half width half maximum versus Q plot showing difference in sample behavior for tmMBP2 (black) versus tmMBP2 (red) proteins at a Q -range of $0.3 \text{ \AA}^{-1}$ to $1.8 \text{ \AA}^{-1}$.

If the fits with Cole-Cole converged to alpha equal, or nearly equal, zero, that would effectively become a Lorentzian fit, however, the alpha term in analyses of these tmMBP isoforms stayed above zero. The best fit using a single forced Lorentzian function thus proved inadequate to fit the dynamical transitions observed in tmMBP2 and tmMBP3 in apo and substrate-bound states. Guided by the susceptibility plot (Figure 3.5 A), which showed only a single peak (more predominant in the H₂O sample), it was clear that a even a double Lorentzian function would be insufficient in fitting the data.

Therefore, the QENS data was fitted with Cole-Cole distribution function, which is a complex one-component function, as shown in Figure 3.5.B. At $Q=1.7 \text{ \AA}^{-1}$, the characteristic relaxation time for tmMBP2-MTT became 21.3 ps compared to 13.2 ps for apo tmMBP2, whereas at $Q=1.9 \text{ \AA}^{-1}$, the characteristic relaxation time for tmMBP2-MTT was 11.7 ps compared to 8.4 ps for the apo tmMBP2. In contrast, relaxation times for the tmMBP3-MAL sample was calculated to be 13.4 ps at and that of the substrate-bound tmMBP3 to be 14.1 ps at $Q=1.7 \text{ \AA}^{-1}$. The relaxation times for both apo and substrate-bound tmMBP3 were the same (7.2 ps) at $Q=1.9 \text{ \AA}^{-1}$. The half-width half-maximum (HWHM) plot, which gives the signal width, or an inverted relaxation time verses the Q , shows clearly that tmMBP2-MTT is an outlier as it becomes visibly more rigid at $Q=1.7 \text{ \AA}^{-1}$ to $1.9 \text{ \AA}^{-1}$. On the other hand, the tmMBP2-apo, tmMBP3-apo, and tmMBP3-MAL do not show such rigid behavior around the same Q -range. The effect of water on binding dynamics does not seem to be significant for tmMBP3 when compared to tmMBP2., However, the local dynamics in tmMBP2 is significantly affected by ligand-binding in the presence of water in and around the substrate-binding pockets. As our crystal structures show, the substrate-binding site in tmMBP2 is shallow and can accommodate 10 or more water molecules, whereas tmMBP3 has a deep binding pocket that completely encloses the substrate inside with only 3 water molecules to make h-bonds. Additionally, we note that the substrate-bound tmMBP2 showed greater structural stability through 500 ns MD run than substrate-bound tmMBP3. These data, therefore, these dynamics data add to our picture of how large surface area of interaction between substrate and tmMBP2 is stabilized by the water molecules.

3.4.4 Dynamical substates in tmMBP as revealed by computational MD

MD simulations were performed on apo and substrate-bound tmMBP isoforms for a total of 500 ns. The details of the MD simulations have been reported in Chapter 2. MD simulations were performed to compare the results from neutron scattering experiments as well as to understand the functional implications of the neutron scattering data. Additionally, MD trajectories were also used to calculate quasi-anharmonic modes.

Here, I have highlighted the root mean square fluctuations of the top 10 modes (RMSF_{10}) analysis in Figure 3.6. As we compare the two isoforms and their apo and substrate-bound states, it can be stated that the dynamical signatures of tmMBP2 are very different from tmMBP3. This is true for both apo and substrate-bound states of the proteins. In tmMBP2, the apo protein shows a relatively flexible structure when compared to the substrate-bound state. Most of the flexibility in the apo protein is observed in the loop region belonging to the C-terminal domain that interacts with the N-terminal domain (shown as region 1 in Figure 3.7.A). Additionally, N-terminal helices close to the binding site are also seen to be more flexible. However, once the substrate binds to the tmMBP2, there is a sudden freeze in the overall dynamics of the whole structure. Similar observations have also been made for tmMBP2 apo and substrate-bound forms of the protein using the QAA analysis (Figure 3.7.A). Here again, the predominant mode confirmed the fluctuations observed in the RMSF_{10} analysis. However, despite the reduction in the dynamics of the protein after substrate-binding, there were local fluctuations observed in the surface exposed loop regions. These observations support our assertions from neutron spectroscopy where it was observed that the substrate-bound tmMBP2 is very rigid.

On the other hand, tmMBP3 apo and substrate-bound forms show remarkably different dynamics. As predicted by both RMSF_{10} and QAA modes, tmMBP3 seems to become little more flexible after substrate binding. Again, a different C-terminal loop region around the binding site of the protein, which interacts with the N-terminal domain after substrate binding, is seen to be very flexible but does not contribute towards a larger transfer of the local motion throughout the protein. Additionally, the C-terminal loop showing flexibility in tmMBP2 seem to be interacting in the apo form of the tmMBP3 but that interaction gets disrupted in substrate-bound tmMBP3. This is interesting as this loop region is known to allosterically modulate the affinity of MBPs towards their cognate substrates. In case of

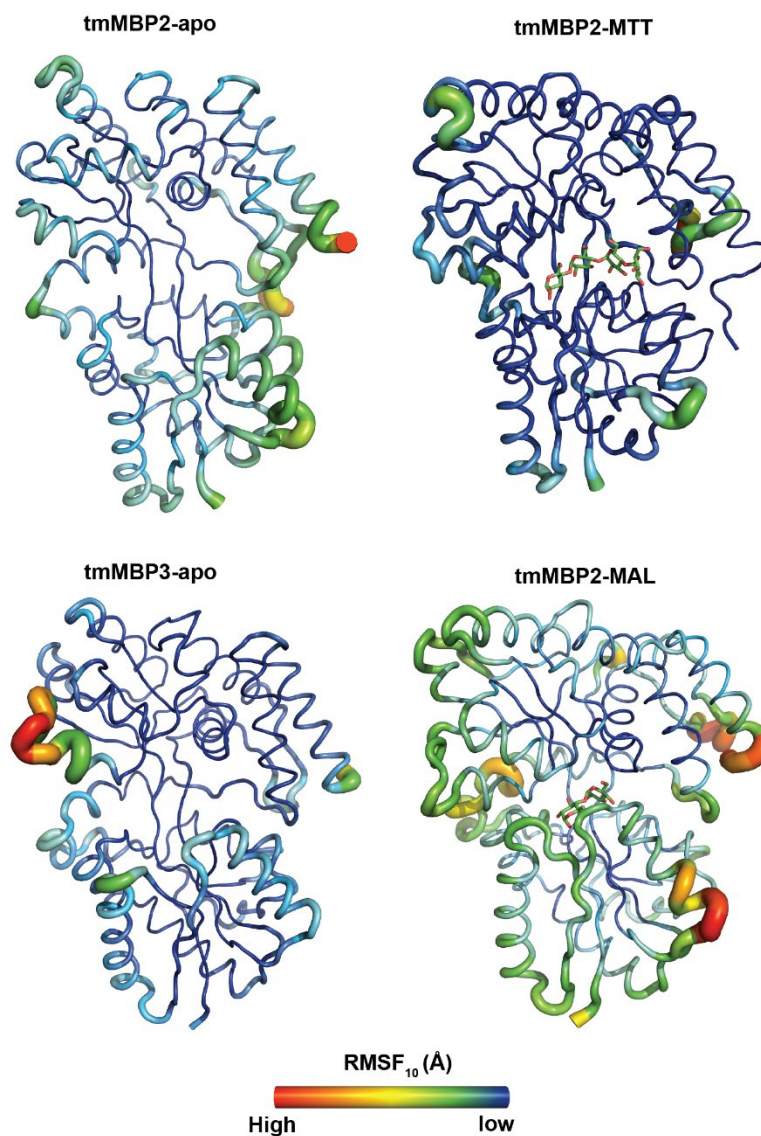


Figure 3.6: Root mean square fluctuations (RMSF₁₀) of tmMBP isoforms. RMSF₁₀ analysis was performed on tmMBP2 (apo and MTT-bound) and tmMBP3 (apo and MAL-bound). The width of the tube represents the sampled area of the residues. Blue regions represent rigid or less flexible regions whereas regions highlighted in red represent highly dynamic profiles.

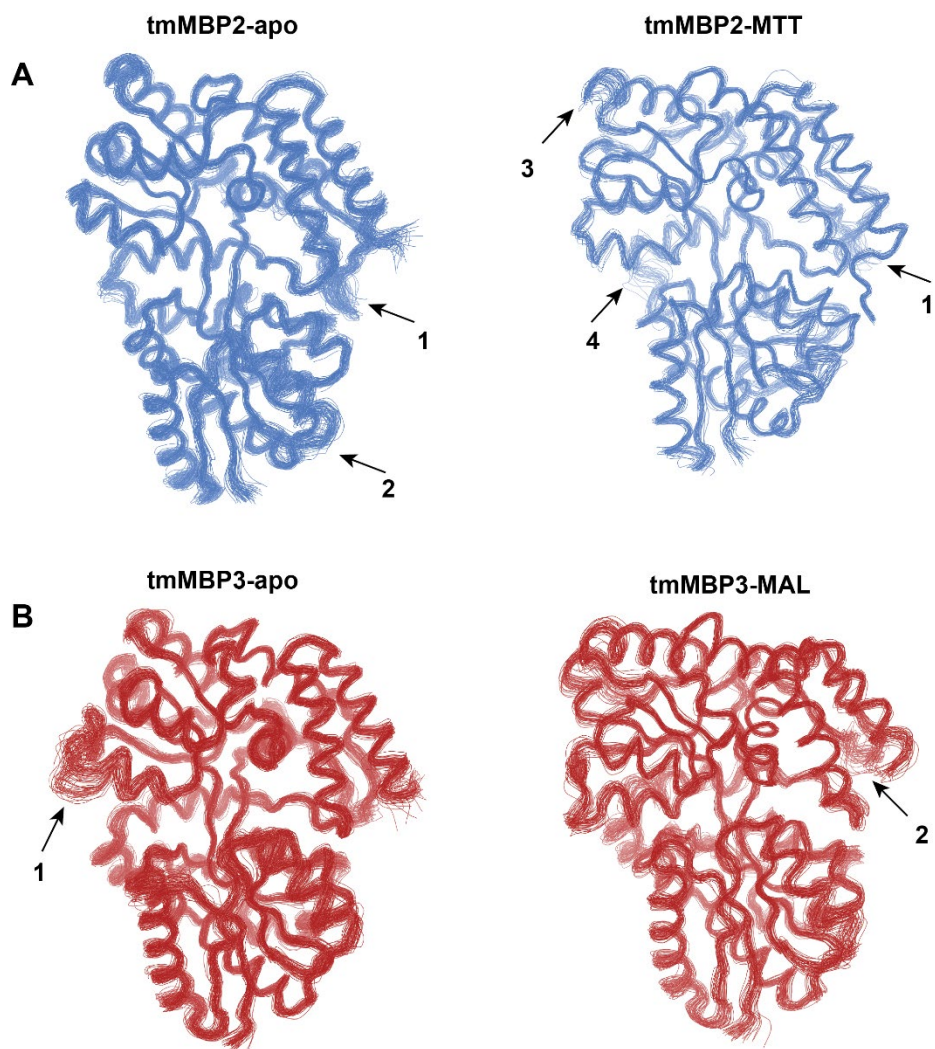


Figure 3.7: Quasi-anharmonic top mode of tmMBP2 and tmMBP3. (A) QAA mode 1 from tmMBP2 apo and MTT-bound systems showing variation in the highly dynamic regions before and after substrate binding. (B) QAA mode 1 from tmMBP3 apo and MAL-bound system highlighting large dynamical variations in both structures before and after substrate binding. Regions of interest are marked on the structures.

tmMBP2, this loop is stabilizing the closed state while in case of tmMBP3, the loop stabilizes the open states. The QAA modes also support this phenomenon as there is differential fluctuation observed in tmMBP2 but not in tmMBP3.

3.5 Discussion

Substrate binding and translocation across the bacterial membrane is predominantly facilitated by the ABC transporter system. These importer systems are associated with high-affinity globular protein that are active in the bacterial periplasm and help bind myriad substrates.⁽⁴⁷⁾ These ABC associated binding protein or simply SBPs show a remarkable structural conservation across the evolutionary spectrum even when they share low sequence similarity.⁽⁴⁸⁾ All members of this family share an archetypal flexible bi-domain architecture, the interface of which forms the binding pocket. SBPs maintain a functional protein fold while binding a diverse array of substrates that range from large peptides and oligosaccharides to ions.⁽⁴⁹⁾ Even when these proteins are structurally similar, they vary considerably in their substrate-binding properties. SBPs have evolved specificity towards a specific set of substrates that they have very high affinities for.⁽⁵⁰⁾ Changes in this affinity profile from one substrate to the other in a SBP has been attributed to fewer alterations in the binding site environment and chemistry in the protein.⁽⁵¹⁾ Given the large variation in substrate-recognition and binding in the SBPs, it is unlikely that local structural changes alone underlie such large functional diversity. Therefore, the SBP superfamily offers a unique set of proteins to understand what additional intrinsic physiochemical features or properties control the evolution of substrate specificity and affinity in structurally similar proteins. We posit that differences in the large-scale dynamical hinge motions in SBP isoforms is attributed to the variation in their local dynamical behaviors in presence and absence of cognate ligands. Here, to investigate this and to better understand the role of dynamics in shaping protein substrate binding profiles, we designed a series of experiments using two isoforms of maltose-binding protein from *T. maritima* (tmMBP2 and tmMBP3) that enabled motions across different time and length-scales to be studied in detail.

Small angle scattering experiments enable the molecular shape and conformational flexibility of proteins to be determined in solution in the presence and absence of cognate

ligand.⁽⁵²⁾ Initial SANS profiles of intensity vs. Q for apo and substrate-bound states of tmMBP2 and tmMBP3 showed good agreement with previous SAXS studies.⁽²⁵⁾ Moreover, extraction of the concentration dependent structure factor $S(Q)$ from the intensities after removing the form factor $F(Q)$ revealed significant differences in the intermolecular interaction profiles for both proteins and behaviors that were best fitted using different models. Specifically, analysis of the tmMBP2 data showed characteristic features of a “hardsphere” model,^(41,42) that indicates less interaction with surrounding proteins, whereas the tmMBP3 data showed properties similar to a “stickyhardsphere” model^(41,43,44) where the protein shows greater interaction with surrounding protein molecule in a crowded solution. Additionally, tmMBP2 showed large differences in the structure factor of its apo vs. substrate-bound forms, suggesting that the apo protein exhibited a more flexible state which is locked down upon binding substrate. The large difference in the structure factor for tmMBP2-apo may correlate with its relatively higher interparticle interaction.

The NSE experiments enable us to study the nature of these interactions in more detail, namely in enabling us to quantify their picosecond to nanosecond solution-state dynamics across differing Q -ranges in both apo and holo forms. The NSE results clearly identified significant differences in global motions in the two isoforms, especially in their rotational-internal regime ($Q = 0.08 \text{ \AA}^{-1} - Q = 0.15 \text{ \AA}^{-1}$). While tmMBP2 showed a higher Q -dependent effective diffusion in the intermediate Q -range, tmMBP3 had a much slower global motion. As it has been suggested that global motions in proteins are coupled to their local dynamics,⁽⁵³⁾ this large variation in D_{eff} could result in not just a large difference in global dynamics of the two isoforms, but also distinctively affect their internal dynamics. These results strongly suggest that the slower translational and rotational dynamics in these proteins can couple their faster stochastic picosecond to nanosecond internal dynamics. Further, the NSE results indicate that both tmMBP isoforms differ in the extent of their interactions with the surrounding environment, and this difference manifests in a significantly large difference in their slower global dynamics. This is consistent with the idea that the surrounding environment of a protein can also have either plasticizer or stabilizer effect and can allow or deny conformational sampling of protein between conformational sub-states.

QENS experiments were performed in conjunction with the NSE to identify the more local dynamical changes at Q-values ranging from $0.3 \text{ \AA}^{-1}$ to $1.8 \text{ \AA}^{-1}$ for faster timescales of $\sim t=60\text{ps}$ in the microenvironment of the two isoforms in presence and absence of a substrate. QENS in association with MD simulation has previously been used to predict local stability and dynamical properties of proteins. As a large number of ordered water molecules ($\sim 30 - 40$ molecules) were at the binding site of tmMBP2 (PDB ID: 6DTS) when compared to only 3 water molecules present in tmMBP3-MAL binding site (PDB ID: 6DTU), it was hypothesized that presence of tightly associated water molecule in tmMBP2 makes the protein more stable in presence of the substrate. This hypothesis was supported in part by our MD simulations which showed tmMBP2-MTT complex is stable based on the RMSF₁₀ values. The analysis of dry, hydrogenated, and deuterated proteins was tested to determine the difference in signal from methyl rotations, water contributions, and internal protein motions, respectively. The final QENS data showed large difference in tmMBP2 and tmMBP3 in their internal dynamics. It especially highlighted a significant difference in tmMBP2-MTT complex and other three samples in the higher Q regions of the data, suggesting that tmMBP2-MTT complex was more rigid than the other three samples. These experiments were done on samples vapor exchanged with D₂O, such that each protein can be considered surrounded by a similar single solvent layer. One interpretation of the observed differences could therefore be that the presence of large numbers of ordered solvent molecules in and around the cavity contributes to the overall stability of the tmMBP2-MTT complex. Conversely, the lack of ordered solvent water around tmMBP2-apo or tmMBP3-apo and substrate-bound proteins may give rise to a comparatively less stable complex. Moreover, the QENS also showed that substrate-binding event segregates the local dynamics in both tmMBP2 and tmMBP3, as in its absence both isoforms show similar local dynamical behavior.

Finally, we note that the QAA modes calculated from the MD trajectories provided quantitatively similar results to that of QENS, where it was observed that internal motions were restricted in tmMBP2-MTT complex. QAA of the tmMBP2-apo molecule showed very flexible structure with the loop region 1 in Figure 3.7.A being highly dynamic. However, binding of the substrate to the binding cavity reduced the fluctuation of this loop region. Studies have shown that truncation of this N-terminal interaction loop region can result in

modulated affinities of MBPs for their cognate substrates. ^(54,55) Since this loop region interacts with the N-terminal domain in tmMBP2-apo, it might have a stabilizing effect on the open-state of the protein. It has been shown that the SBP recognizes the cognate versus non-cognate substrate and binding of the non-cognate substrate may result in an incomplete domain-domain closure. The lack of significant dynamical variations in tmMBP3 on the timescales observed in SANS, NSE, QENS, MD, and QAA suggest that the overall slow global dynamics reduces the internal dynamics, which may also be affected by the insertions of several topological elements in the same fold-space due to larger size of tmMBP3. ⁽⁵⁶⁾

This study, therefore, has demonstrated that neutron scattering can detect and differentiate the distinct dynamical profiles of two structurally and functionally related isoforms. Moreover, as MD simulations are complementary to neutron scattering and often provide a complete picture of atomistic dynamics, computational simulations can help identify the substates that are associated with protein dynamics. However, correlating these experimental dynamical signatures with molecular dynamics simulations in order to better understand the molecular basis of these differences remains a significant challenge. Specifically, there remains significant mis-match in the time and length scales across which it is necessary to extract, calculate and quantitatively compare the information from each technique. Advancement in joint computational and experimental data analysis can lead to better energy resolutions, which can enable study of slower dynamical processes. Efforts are being made at SNS, ORNL to overcome these technical hurdles and to integrate the experimental and computational datasets. More broadly, this study indicates how the techniques of neutron scattering and spectroscopy can be applied to decipher the subtle dynamical and conformational differences in structurally related proteins and enzymes, and to help ascertain the underlying physiochemical factors and properties that influence their overall function.

List of References

1. Wang, M., Jiang, Y. Y., Kim, K. M., Qu, G., Ji, H. F., Mittenthal, J. E., Zhang, H. Y., and Caetano-Anolles, G. (2011) A universal molecular clock of protein folds and its power in tracing the early history of aerobic metabolism and planet oxygenation. *Mol Biol Evol* **28**, 567-582
2. Thornton, J. M., Orengo, C. A., Todd, A. E., and Pearl, F. M. (1999) Protein folds, functions and evolution. *J Mol Biol* **293**, 333-342
3. Cuneo, M. J. (2007) *Exploring the structural diversity and engineering potential of thermophilic periplasmic binding proteins*, Citeseer
4. Narayanan, C., Bernard, D. N., Bafna, K., Gagné, D., Chennubhotla, C. S., Doucet, N., and Agarwal, P. K. (2018) Conservation of dynamics associated with biological function in an enzyme superfamily. *Structure* **26**, 426-436. e423
5. Snell-Rood, E. C., Van Dyken, J. D., Cruickshank, T., Wade, M. J., and Moczek, A. P. (2010) Toward a population genetic framework of developmental evolution: the costs, limits, and consequences of phenotypic plasticity. *Bioessays* **32**, 71-81
6. Caetano-Anolles, G., and Caetano-Anolles, D. (2003) An evolutionarily structured universe of protein architecture. *Genome Res* **13**, 1563-1571
7. Feher, V. A., and Cavanagh, J. (1999) Millisecond-timescale motions contribute to the function of the bacterial response regulator protein Spo0F. *Nature* **400**, 289-293
8. Henzler-Wildman, K., and Kern, D. (2007) Dynamic personalities of proteins. *Nature* **450**, 964-972
9. Turton, D. A., Senn, H. M., Harwood, T., Lapthorn, A. J., Ellis, E. M., and Wynne, K. (2014) Terahertz underdamped vibrational motion governs protein-ligand binding in solution. *Nature communications* **5**, 1-6
10. Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M., and Kern, D. (2007) A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913-916
11. Austin, R. H., Beeson, K., Eisenstein, L., Frauenfelder, H., and Gunsalus, I. (1975) Dynamics of ligand binding to myoglobin. *Biochemistry* **14**, 5355-5373
12. Boehr, D. D., Dyson, H. J., and Wright, P. E. (2006) An NMR perspective on enzyme dynamics. *Chemical reviews* **106**, 3055-3079

13. Grimaldo, M., Roosen-Runge, F., Zhang, F., Schreiber, F., and Seydel, T. (2019) Dynamics of proteins in solution. *Quarterly Reviews of Biophysics* **52**
14. Yang, L.-Q., Sang, P., Tao, Y., Fu, Y.-X., Zhang, K.-Q., Xie, Y.-H., and Liu, S.-Q. (2014) Protein dynamics and motions in relation to their functions: several case studies and the underlying mechanisms. *Journal of Biomolecular Structure and Dynamics* **32**, 372-393
15. Marino-Buslje, C., Monzon, A. M., Zea, D. J., Fornasari, M. S., and Parisi, G. (2019) On the dynamical incompleteness of the Protein Data Bank. *Brief Bioinform* **20**, 356-359
16. Biehl, R., Monkenbusch, M., and Richter, D. (2011) Exploring internal protein dynamics by neutron spin echo spectroscopy. *Soft Matter* **7**, 1299-1307
17. Callaway, D. J., and Bu, Z. (2017) Visualizing the nanoscale: protein internal dynamics and neutron spin echo spectroscopy. *Current opinion in structural biology* **42**, 1-5
18. Smolin, N., Biehl, R., Kneller, G., Richter, D., and Smith, J. C. (2012) Functional domain motions in proteins on the~ 1–100 ns timescale: comparison of neutron spin-echo spectroscopy of phosphoglycerate kinase with molecular-dynamics simulation. *Biophysical journal* **102**, 1108-1117
19. Smith, J. C., Tan, P., Petridis, L., and Hong, L. (2018) Dynamic neutron scattering by biological systems. *Annual review of biophysics* **47**, 335-354
20. Jansson, H., Bergman, R., and Swenson, J. (2006) Protein and solvent dynamics as studied by QENS and dielectric spectroscopy. *Journal of non-crystalline solids* **352**, 4410-4416
21. Swenson, J., Jansson, H., Hedström, J., and Bergman, R. (2007) Properties of hydration water and its role in protein dynamics. *Journal of Physics: Condensed Matter* **19**, 205109
22. Sharff, A. J., Rodseth, L. E., Spurlino, J. C., and Quiocho, F. A. (1992) Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry* **31**, 10657-10663

23. Quijoch, F. A., Spurlino, J. C., and Rodseth, L. E. (1997) Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure* **5**, 997-1015
24. Boucher, N., and Noll, K. M. (2011) Ligands of thermophilic ABC transporters encoded in a newly sequenced genomic region of *Thermotoga maritima* MSB8 screened by differential scanning fluorimetry. *Applied and environmental microbiology* **77**, 6395-6399
25. Shukla, S., Bafna, K., Gullett, C., Myles, D. A. A., Agarwal, P. K., and Cuneo, M. J. (2018) Differential Substrate Recognition by Maltose Binding Proteins Influenced by Structure and Dynamics. *Biochemistry* **57**, 5864-5876
26. Wassenberg, D., Liebl, W., and Jaenicke, R. (2000) Maltose-binding protein from the hyperthermophilic bacterium *Thermotoga maritima*: stability and binding properties. *Journal of molecular biology* **295**, 279-288
27. Törnkvist, M., Larsson, G., and Enfors, S.-O. (1996) Protein release and foaming in *Escherichia coli* cultures grown in minimal medium. *Bioprocess Engineering* **15**, 231-237
28. Do, C., Heller, W. T., Stanley, C., Gallmeier, F. X., Doucet, M., and Smith, G. S. (2014) Understanding inelastically scattered neutrons from water on a time-of-flight small-angle neutron scattering (SANS) instrument. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **737**, 42-46
29. Franke, D., Petoukhov, M., Konarev, P., Panjkovich, A., Tuukkanen, A., Mertens, H., Kikhney, A., Hajizadeh, N., Franklin, J., and Jeffries, C. (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of applied crystallography* **50**, 1212-1225
30. Ju, R. T., Frank, C. W., and Gast, A. P. (1992) CONTIN analysis of colloidal aggregates. *Langmuir* **8**, 2165-2171
31. Ohl, M., Monkenbusch, M., Arend, N., Kozielski, T., Vehres, G., Tiemann, C., Butzek, M., Soltner, H., Giesen, U., and Achten, R. (2012) The spin-echo spectrometer at the Spallation Neutron Source (SNS). *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **696**, 85-99

32. Zolnierczuk, P. A., Holderer, O., and Monkenbusch, M. (2018) DrSPINE: Data reduction for SPIN echo experiments. in *AIP Conference Proceedings*, AIP Publishing LLC
33. Mamontov, E., Zamponi, M., Hammons, S., Keener, W. S., Hagen, M., and Herwig, K. W. (2008) BASIS: A new backscattering spectrometer at the SNS. *Neutron News* **19**, 22-24
34. Salomon-Ferrer, R., Case, D. A., and Walker, R. C. (2013) An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **3**, 198-210
35. Ramanathan, A., and Agarwal, P. K. (2011) Evolutionarily conserved linkage between enzyme fold, flexibility, and catalysis. *PLoS Biol* **9**, e1001193
36. Roe, D. R., and Cheatham III, T. E. (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation* **9**, 3084-3095
37. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., and Beckstein, O. (2011) MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of computational chemistry* **32**, 2319-2327
38. Ramanathan, A., Savol, A. J., Langmead, C. J., Agarwal, P. K., and Chennubhotla, C. S. (2011) Discovering conformational sub-states relevant to protein function. *PLoS One* **6**, e15827
39. Ramanathan, A., Savol, A. J., Agarwal, P. K., and Chennubhotla, C. S. (2012) Event detection and sub-state discovery from biomolecular simulations using higher-order statistics: Application to enzyme adenylate kinase. *Proteins: Structure, Function, and Bioinformatics* **80**, 2536-2551
40. Butler, P., Alina, G., Hernandez, R., Doucet, M., Jackson, A., Kienzle, P., Kline, S., and Zhou, J. (2013) SASView for Small Angle Scattering Analysis. *There is no corresponding record for this reference*
41. Kotlarchyk, M., and Chen, S. H. (1983) Analysis of small angle neutron scattering spectra from polydisperse interacting colloids. *The Journal of chemical physics* **79**, 2461-2469
42. Percus, J. K., and Yevick, G. J. (1958) Analysis of classical statistical mechanics by means of collective coordinates. *Physical Review* **110**, 1

43. Menon, S., Manohar, C., and Rao, K. S. (1991) A new interpretation of the sticky hard sphere model. *The Journal of chemical physics* **95**, 9186-9190
44. Baxter, R. (1968) Percus–Yevick equation for hard spheres with surface adhesion. *The Journal of chemical physics* **49**, 2770-2774
45. Callaway, D. J., Farago, B., and Bu, Z. (2013) Nanoscale protein dynamics: A new frontier for neutron spin echo spectroscopy. *The European Physical Journal E* **36**, 76
46. Mamontov, E., and Herwig, K. W. (2011) A time-of-flight backscattering spectrometer at the Spallation Neutron Source, BASIS. *Review of Scientific Instruments* **82**, 085109
47. Rice, A. J., Park, A., and Pinkett, H. W. (2014) Diversity in ABC transporters: type I, II and III importers. *Crit Rev Biochem Mol Biol* **49**, 426-437
48. Cui, J., and Davidson, A. L. (2011) ABC solute importers in bacteria. *Essays Biochem* **50**, 85-99
49. Locher, K. P. (2016) Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nature structural & molecular biology* **23**, 487
50. Maqbool, A., Horler, R. S., Muller, A., Wilkinson, A. J., Wilson, K. S., and Thomas, G. H. (2015) The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity. *Biochemical Society Transactions* **43**, 1011-1017
51. Quiocho, F. A. (1989) Protein-carbohydrate interactions: basic molecular features. *Pure and Applied Chemistry* **61**, 1293-1306
52. Chen, P.-c., and Hennig, J. (2018) The role of small-angle scattering in structure-based screening applications. *Biophysical reviews* **10**, 1295-1310
53. Hawkins, R. J., and McLeish, T. C. (2006) Coupling of global and local vibrational modes in dynamic allostery of proteins. *Biophysical journal* **91**, 2055-2062
54. Rizk, S. S., Paduch, M., Heithaus, J. H., Duguid, E. M., Sandstrom, A., and Kossiakoff, A. A. (2011) Allosteric control of ligand-binding affinity using engineered conformation-specific effector proteins. *Nature structural & molecular biology* **18**, 437

55. Telmer, P. G., and Shilton, B. H. (2003) Insights into the conformational equilibria of maltose-binding protein by analysis of high affinity mutants. *Journal of Biological Chemistry* **278**, 34555-34567
56. Paciaroni, A., Cinelli, S., and Onori, G. (2002) Effect of the environment on the protein dynamical transition: a neutron scattering study. *Biophysical journal* **83**, 1157-1164

**Chapter 4 Determining water-mediated substrate
stabilization in maltose-binding proteins
using neutron diffraction methods**

4.1 Abstract

Thermotoga maritima maltose binding protein isoform 2 (tmMBP2) binds tri- and tetra-saccharide and helps in their membrane translocation through an associated ABC transporter system. The XRD of the tmMBP2-MTT complex showed a large network of interconnected water molecules stretching from one end of the substrate binding pocket to another. It was also observed that these water molecules were positioned to form multiple h-bonded interactions and to mediate several other indirect H-bonded interactions between the protein and substrate molecule. However, since the XRD is relatively insensitive to H-atoms, and unable to directly determine the positions of hydrogens in crystal structures even when atomic (<1.2 Å) resolution data are available, we wanted to investigate the extent of these H-bonded water networks and understand their broader role in the stability of a protein-substrate interaction using neutron protein crystallography (NPC). NPC is uniquely sensitive to H (and D) in biological structures and was used to correctly identify the precise positions and orientations of hydrogen atoms on the water molecules and how they interacted with the substrate. The results also provided other insights into the local chemistry of the protein molecule by identifying the protonation states of various charged and aromatic residues that carpet the binding site. This first MBP neutron structure unambiguously identified the nature and orientation of the H-bonded water and water-mediated interactions in stabilizing large oligosaccharide substrate inside the binding site of the protein. Additionally, the NPC data was also used to investigate the local dynamics and solvent accessibility within the protein by mapping the pattern and extent of hydrogen/deuterium (H/D) exchange on backbone amides. More broadly, these results reveal show the intricate mechanisms that alter the substrate-specificity and affinity in MBPs through water-mediated hydrogen bonding interaction.

4.2 Introduction

Water plays a fundamental role in all life processes. For most proteins, water provides a versatile environment that promotes folding, local and global dynamics, and protein function.⁽¹⁾ Additionally, the key role of water has been implicated in enzyme catalysis, protein-ligand interactions and allostery.^(2,3) Water can simultaneously act as an h-bond donor or an h-bond acceptor, promoting protein stability by forming multiple h-bonds with

polar side chains of the protein molecule. ⁽⁴⁾ Specifically, in protein-ligand interactions, the presence of a water molecule improves the favorable contributions to free energy instead of a void at the same site. ⁽⁴⁾ This is particularly true for the water molecules present in close vicinity of a substrate molecule, which tend to be more energetically favored than the bulk water. ⁽⁵⁾ These “crystal” waters often exhibit such reduced dynamics that they can be resolved using X-ray/neutron crystallography. In certain examples, such as the peptide binding protein OppA, it has been estimated that the binding site water molecules have B-factors that are even marginally lower than the average protein B-factors. ⁽⁶⁾ This strongly suggests that some water molecules form an indispensable part of a protein-substrate-solvent interaction network and therefore, studying the effects of water-mediated hydrogen bonding interactions on such protein-substrate interactions may help provide insights into their affinity and specificity profiles.

The ABC transporter associated binding proteins are known to bind a diverse range of substrate molecules and help in their translocation across the membrane. These bidomain proteins have a conserved $\alpha\beta$ -fold in which the β -strand core is sandwiched between the surrounding α -helices. ⁽⁷⁾ The class of maltose-binding proteins (MBPs) form a distinctive unit of the larger SBP superfamily. These proteins have evolved to bind maltose and malto-oligomers with varying degrees of substrate specificity. While most mesophilic bacteria contain a single MBP gene, *T. maritima* has three isoforms of this protein, namely tmMBP1, tmMBP2, and tmMBP3 (discussed in detail in Chapter 2). X-ray crystal structures of these tmMBPs revealed a size-selective differential substrate binding profile, with tmMBP1 and tmMBP2 binding larger oligosaccharides (triose and above) and tmMBP3 binding only to maltose. ⁽⁸⁾ One important difference between the large substrate binding isoforms and the small substrate binding isoform was the distribution of water molecules in and around the binding site. While tmMBP2 had a network of 13 water molecules forming h-bonds with the substrate, in tmMBP3 only 3 water molecules that formed h-bonds with substrate.

Thus, to ascertain and distinguish whether “crystal” waters play a specific role towards substrate stabilization in the tmMBP2-MTT complex, neutron crystallography experiments were designed that would allow the explicit orientation of the h-bonded water networks to

be determined at the binding site. The initial crystallization condition used for the XRD data collection was further optimized for growing large ($>>1\text{mm}^3$) neutron-sized crystals. Subsequently, these crystals were checked for their diffraction quality and further rounds of optimization to improve the overall resolution of the neutron data. Finally, neutron diffraction data were collected to better than 2.1 Å resolution on tmMBP2 crystals in complex with MTT. The refined neutron structure explicitly reveals the extent of the H-bonded interactions and water network at the substrate-binding site, and the protonation state of amino residues in the protein. The implications of such water networks on the nature and strength of substrate-protein-solvent interactions is discussed in detail in later sections.

4.3 Materials and methods

Expression, purification, and crystallization of tmMBP2

tmMBP2 was purified as per the protocol mentioned in reference ⁽⁸⁾. Briefly, the Enfors minimal media grown His-tag containing tmMBP2 protein was purified in a two-step chromatographic protein prep consisting of an initial nickel-affinity chromatography followed by a size exclusion purification step. The purified protein was extensively dialyzed for several rounds of buffer exchanged (20 mM Tris (pH 7.5), 40 mM NaCl) at room temperature. The protein was concentrated down to 25 mg/ml and used for crystallization under our previously optimized conditions of 24% - 25% (wt./vol.) PEG 3350, 0.2 M sodium acetate, 0.1 M BisTris (pH 5.5) in H₂O solutions. Prior to setting up the trays, tmMBP2 was mixed with 2 mM MTT. Sitting drop crystallization was performed on tmMBP2-MTT in a sandwich box setup with a nine-well siliconized glass plate (Hampton Research). Each sitting drop was made by mixing 400 µl of protein with equal volume of the mother liquor. The mother liquor was supplemented with 10% (vol./vol.) deuterated glycerol which reduced crystal pathologies, by slowing the rate of crystal growth crystallization. The trays were incubated at 20 °C for 25 – 30 days during which large crystals (3 – 10 mm³) to appear. Subsequently, the crystals were transferred once a week into sandwich boxes with fresh deuterated mother liquor to promote H/D exchange. A total of four D₂O mother liquor exchanges were performed on the crystals. Finally, the crystals were screened under a microscope fitted with a polarizer lens and large single crystals were transferred carefully in 2 – 3 mm quartz capillaries (Hampton Research) with a small amount of mother

liquor and sealed carefully with bee's wax. These capillary mounted crystals were later used to collect neutron diffraction data.

Neutron diffraction data collection and joint refinement

Neutron beamtime was awarded (IPTS: 2475) initially for 20 days and subsequently extended for 9 more days. The neutron diffraction data was collected on the IMAGINE neutron diffractometer at the High Flux Isotope Reactor (HFIR), Oak Ridge National Laboratory (ORNL).^(9,10) A total of 35 quasi-Laue ($\lambda = 2.78 - 4.50$ Å) diffraction frames were collected at room temperature in 10° steps using the standard single-phi goniometer setup at the diffractometer. The data were collected at three crystals settings, in order to fill in the blind region data. The final diffraction data extended to 2.1 Å. An initial peak search, indexing and profile fitted data reduction was completed using the Lauegen program⁽¹¹⁾, the reduced frames were wavelength normalized using LSCALE,⁽¹²⁾ and subsequently merged in SCALA.⁽¹³⁾ Additionally, the same crystal was used to collect an X-ray room temperature dataset on the Rigaku HighFlux HomeLab instrument at the SNS (ORNL), which is equipped with a MicroMax-007 X-ray source and Osmic VariMax optics. Realtime diffraction data were collected on a EIGER R 4M hybrid photon counting detector. The data were integrated using the *CryAlis^{Pro}* program (Rigaku, The Woodlands, Texas). The integrated data were then reduced and scaled in *AIMLESS* program from the CCP4 software package.⁽¹⁴⁾ The X-ray data extended to 1.70 Å and was 100% complete (see Table 4.1. for details). Molecular replacement was used to determine the phases in the Phaser program with the 3D coordinates of tmMBP2-MTT (PDB: 6DTS) as a search model.⁽¹⁵⁾ X-ray and neutron joint refinement was performed using PHENIX Refine^(16,17) and manual model building was carried out in COOT molecular graphics suite.⁽¹⁸⁾ Ramachandran plot and MolProbity were used to determine correct stereochemistry of the modelled structure.⁽¹⁹⁾

4.4 Results

4.4.1 Neutron structure of tmMBP-MTT

The neutron structure was determined at 2.10 Å and was initially refined to an $R_{\text{work}}/R_{\text{free}}$ of 23.24/26.27%. The X-ray crystal structure was determined at 1.70 Å with an $R_{\text{work}}/R_{\text{free}}$ of 14.78/18.73%. A total of 3099 non-hydrogen atoms were then modeled after neutron/X-

ray joint refinement, of which 2893 were protein atoms, 45 ligand atoms, and 161 solvent atoms. A joint neutron and X-ray refinement typically provides a better model at these diffraction resolutions ^(20,21) Although the 2.1 Å resolution neutron data had overall completeness of ~73%, the completeness in the highest (2.21- 2.10) resolution shell was better than 50%. In joint neutron/X-ray structure refinement, the X-ray data serve to refine the core structural model for the non-hydrogen heavy atoms (C, N, O, S) on which the hydrogen positions are determined and refined using the neutron data. Additionally, since neutrons do not ionize the sample, all neutron data collection was performed at physiologically relevant temperature. Standard data collection statistics and details on the refined model are given in Table 4.1.

Overall, the room temperature neutron structure of tmMBP2 (Figure 4.1.A) was very similar (r.m.s.d.= 0.33 on backbone atoms) to the X-ray structure of the same protein published earlier at cryogenic temperatures (PDB ID: 6DTS) ⁽⁸⁾, and clearly showed the signature $\alpha\beta$ -fold with β -strands in both N- and C-terminal domains surrounded by α -helices ⁽²²⁾. Moreover, the neutron structure enabled the addition of 2939 H atoms on the protein, 322 D atoms on solvent water molecules and identified distinct patterns of H/D exchange across the 249 titratable protein residues. Additionally, 2Fo-Fc no-fill neutron maps also showed clearly resolved ligand density for MTT within the tmMBP2 binding site, which was not included in the original structural model (Figure 4.2.B) However, we note that the overall number of solvent molecules modelled for the X-ray cryo-structure were 365 per protein molecule, which is more than double those identified using neutron structure. Despite this and most critically, the neutron structure was able to identify all 13 water molecules in the cavity that were inferred to provide h-bond interactions in the X-ray model (Figure 4.1.C). The determination of the positions and orientations of these D₂O water molecules in the binding cavity can be ascertained with high confidence due to the fact that the current dataset had a resolution of 2.1 Å. ⁽²³⁾ Thus, the neutron structure was able to explicitly show how the network of water molecule that traversed the entire length of the substrate molecule formed precise intermolecular h-bonds with nearby amino-acids residues. Moreover, several water molecules also aided in mediating in-direct h-bond interactions between the protein side chains and the substrate molecule, as had been previously inferred from the cryogenic structure (PDB ID: 6DTS). ⁽⁸⁾

Table 4.1: Data collection and refinement statistics for tmMBP2-MTT Neutron/X-ray joint refinement. Statistics for the highest-resolution shell are shown in parentheses. Both neutron and X-ray data was collected separately and then jointly refined to get the final structure.

	Neutron	X-ray
Source	IMAGINE (HFIR, ORNL)	Rigaku HighFlux HomeLab
Wavelength	2.78 – 4.50 Å	1.54 Å
Resolution range	35.71 - 2.10 (2.21- 2.10)	89.75 - 1.70 (1.76 – 1.70)
Space group	P 1 21 1	
Unit cell	a=35.91 b=56.33 c=90.00 α =90 β =94.25 γ =90	
Total reflections	52954 (4675)	786386 (72561)
Unique reflections	15172 (1628)	39602 (3961)
Multiplicity	3.5 (2.9)	19.9 (18.3)
Completeness (%)	72.36 (51.49)	100.00 (100.00)
Mean I/sigma(I)	7.40 (1.70)	51.66 (9.85)
Wilson B-factor	7.73	18.65
R-merge	0.14 (0.30)	0.12 (0.38)
R-meas	0.16 (0.34)	0.12 (0.40)
R-pim	0.07 (0.16)	0.02 (0.09)
CC1/2	NA	0.99 (0.96)
Neutron/X-ray Joint Refinement		
Reflections used in refinement	15145 (1055)	39564 (3957)
Reflections used for R-free	730 (47)	1932 (178)
R-work (%)	23.24 (32.06)	14.78 (17.26)
R-free (%)	26.27 (38.53)	18.73 (22.88)
Number of non-hydrogen atoms	3099	

Table 4.1 continued

	Neutron	X-ray
<i>Number of hydrogen atoms</i>	2939	
<i>Number of deuterium atoms</i>		
<i>macromolecules</i>	553	
<i>ligands</i>	14	
<i>solvent</i>	322	
<i>Protein residues</i>	379	
<i>RMS (bonds)</i>	0.100	
<i>RMS (angles)</i>	1.48	
<i>Ramachandran favored (%)</i>	99.47	
<i>Ramachandran allowed (%)</i>	0.53	
<i>Ramachandran outliers (%)</i>	0.00	
<i>Rotamer outliers (%)</i>	2.66	
<i>Clashscore</i>	3.41	
<i>Average B-factor</i>	25.26	
<i>macromolecules</i>	24.37	
<i>ligands</i>	17.69	
<i>solvent</i>	43.31	

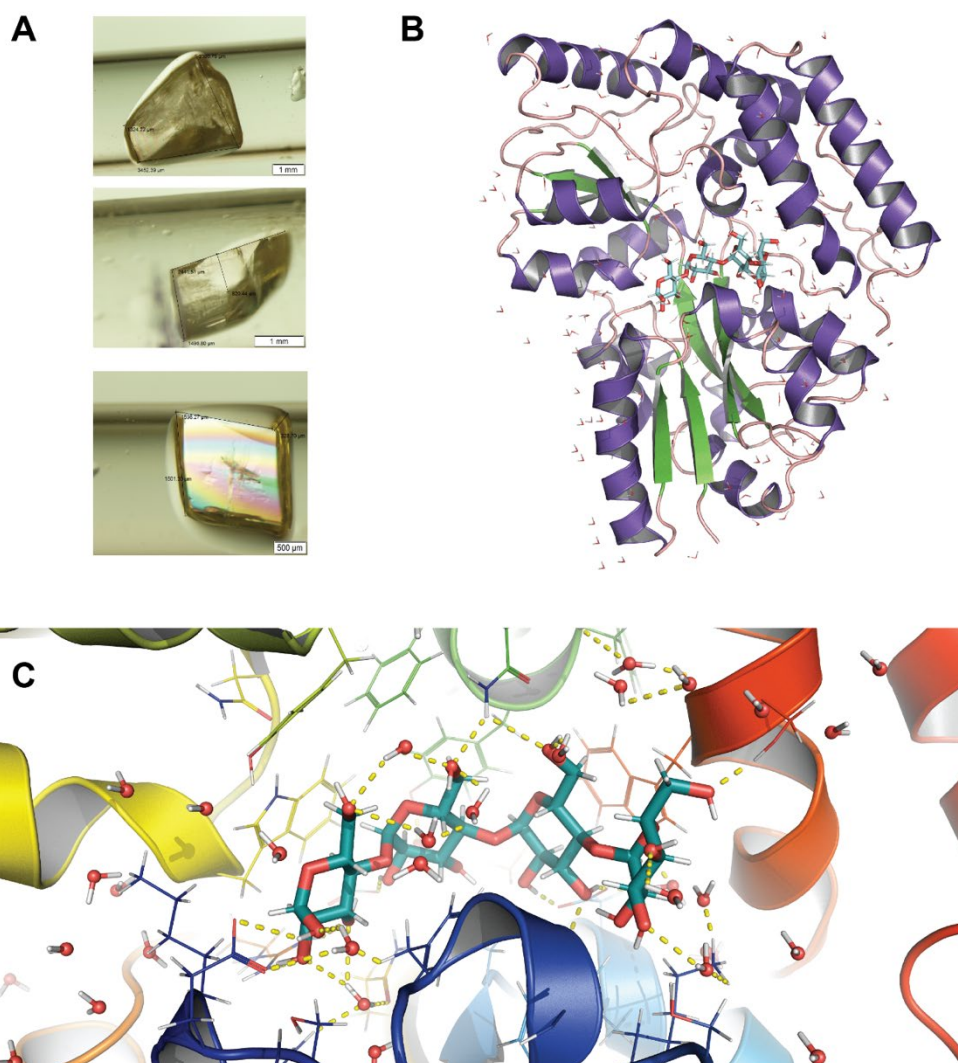


Figure 4.1: Neutron structure of tmMBP2-MTT complex. (A) Large crystal grown for neutron diffraction that range from $\sim 3 - 10 \text{ mm}^3$, the crystals showed some striation marks but diffracted to $\sim 2.0 \text{ \AA}$. (B) A cartoon representation of tmMBP-MTT complex showing the signature $\alpha\beta$ -fold. MTT is shown in cyan. Solvent was modeled as D_2O . (C) A detailed view of the tmMBP2 binding site showing MTT in green surrounded by solvent D_2O shown as red-spheres with protruding deuterium in white. Important non-bonded interaction forming protein residues are shown in the wireframe. The yellow dotted lines represent hydrogen-bond interactions..

Therefore, it is evident from the current neutron structure that this extensive network of water molecules plays a key role in not just binding and stabilization of large substrates, like MTT, in shallow binding pockets of tmMBPs, but also in influencing the favorable interactions between the protein and substrate itself. Furthermore, these additional interactions formed due to the presence of additional non-bonded interaction increases the rigidity as measured by lower B-factors in refinement of the overall structure, allowing the substrate to be safely locked between the two domains. These findings also suggest that there is no significant difference observed in the interaction profiles of tmMBP2 measured under cryo or room temperature conditions.

4.4.2 Determination of protonation states of physiologically relevant residues

Another very important result derived from neutron crystallography is the determination of the explicit protonation states of the important side chain residues at physiologically relevant conditions. Since the neutron structure of tmMBP2-MTT complex was collected at room temperature, it is possible to find the physiologically relevant protonation states of key residues, especially those that line the binding site and interact with the substrate. For example, tmMBP2 has only a single histidine (His-65) which is doubly protonated at pD of ~6.0 (Figure 4.2.A). The protonated N ϵ of His-65 forms a hydrogen bond with the backbone carboxylate of Met-333 with a distance of 1.9 Å. The close vicinity of His-65 from the binding site may also help it in regulating the local electrostatic interaction, as in a doubly protonated histidine carries a net +1 charge. ⁽²⁴⁾ Additionally, 18 tyrosine and 4 tryptophan residues, most of which contribute in forming non-bonded contacts with the substrates were all protonated. Some of these residues also formed hydrogen bonds with the solvent, further stabilizing their conformation during substrate binding. Moreover, the neutron density helped in determine the explicit protonation states of the important residues that stabilize the substrate inside the binding pocket of tmMBP2, such as Tyr-38, His-65, Trp-67, Tyr-158, Tyr-213, Tyr-233, Trp-343. Some of these residues and their associated neutron densities are shown in Figure 4.2. A.

4.4.3. Identifying the H/D exchangeable positions on tmMBP2

The pattern and extent of H/D exchangeable backbone amides sites can be clearly determined using the neutron crystal structure. The extent of H/D exchange backbone amides sites on the protein during a H/D back-exchange experiment can yield important information on the local dynamics of proteins as well as on their hydration properties and solvent accessibility.^(25,26) Therefore, all the exchangeable sites on tmMBP2 were manually inspected and the exchangeable amide sites were shown on the 3D structure (Figure 4.3.A). As expected, most of these exchangeable sites were found on solvent exposed regions of the protein. Significant H/D exchange was also observed in and around the substrate-binding site. This can be attributed to the fact that although the protein is constrained by packing forces within the crystal lattice, the protein remains sufficiently flexible locally to allow D₂O access to the binding cavity and for H/D exchange to occur through the successive rounds of D₂O dialysis. In contrast, the β -strand core of the protein did not undergo significant H/D exchange. (Figure 4.3.B). Amide exchange is frequently used in nuclear magnetic resonance (NMR) and mass spectroscopy (MS) to determine or infer the flexibility of a proteins.⁽²⁷⁻²⁹⁾ Residues showing both side-chain and main-chain amide exchanges included histidine, tyrosine, tryptophan, arginine, asparagine, glutamine, threonine, and serine. Here, residues were designated as deuterated if the refined amide D-occupancy was above 0.8 and as hydrogenated when D-occupancy refined to less than 0.2 (Figure 4.3. A and C). Out of a total 379 residues, 169 underwent significant backbone amide H/D exchanges, 80 showed exchange of both main chain amide and side chain hydrogens, whereas only the side chains hydrogens were exchangeable in 7 residues. The exchangeable amide H/D exchanges were confirmed using the neutron Fo-Fc omit maps (Figure 4.3. B). Taken together, this analysis provides a unique prospective towards understanding the local dynamics and solvent accessibility of the tmMBP2 protein in near-native conditions.

4.5 Discussion

NPC provides a powerful tool for determining hydrogen and water orientations in a protein structure.⁽³⁰⁾ Neutron protein crystallography can also provide unique insights into the explicit protonation states of important catalytic residues as well as in identifying key structural and mechanistic features such as like “low-barrier” h-bonds.^(31,32) In addition,

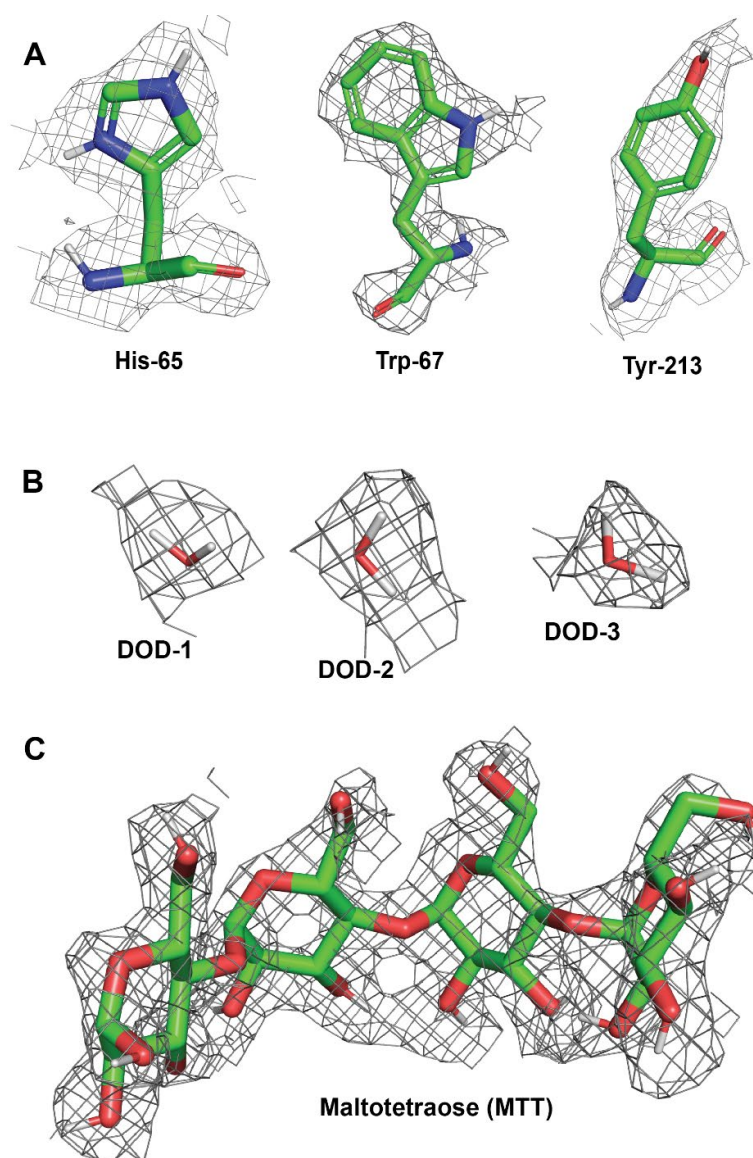


Figure 4.2: Neutron no-fill 2FoFc Maps of tmMBP2-MTT complex. 2Fo-Fc maps were contoured at 1.0 r.m.s.d. in PyMol (A) Neutron density showing various protonation states of some important aromatic residues Histidine 65, Tryptophan 67, and Tyrosine 203. (B) Density around water molecules indicating the positions of deuterium atoms (C) Neutron density around the substrate MTT showing four glucose moieties linked by α 1-4 glycosidic linkage fitted within the 2Fo-Fc map

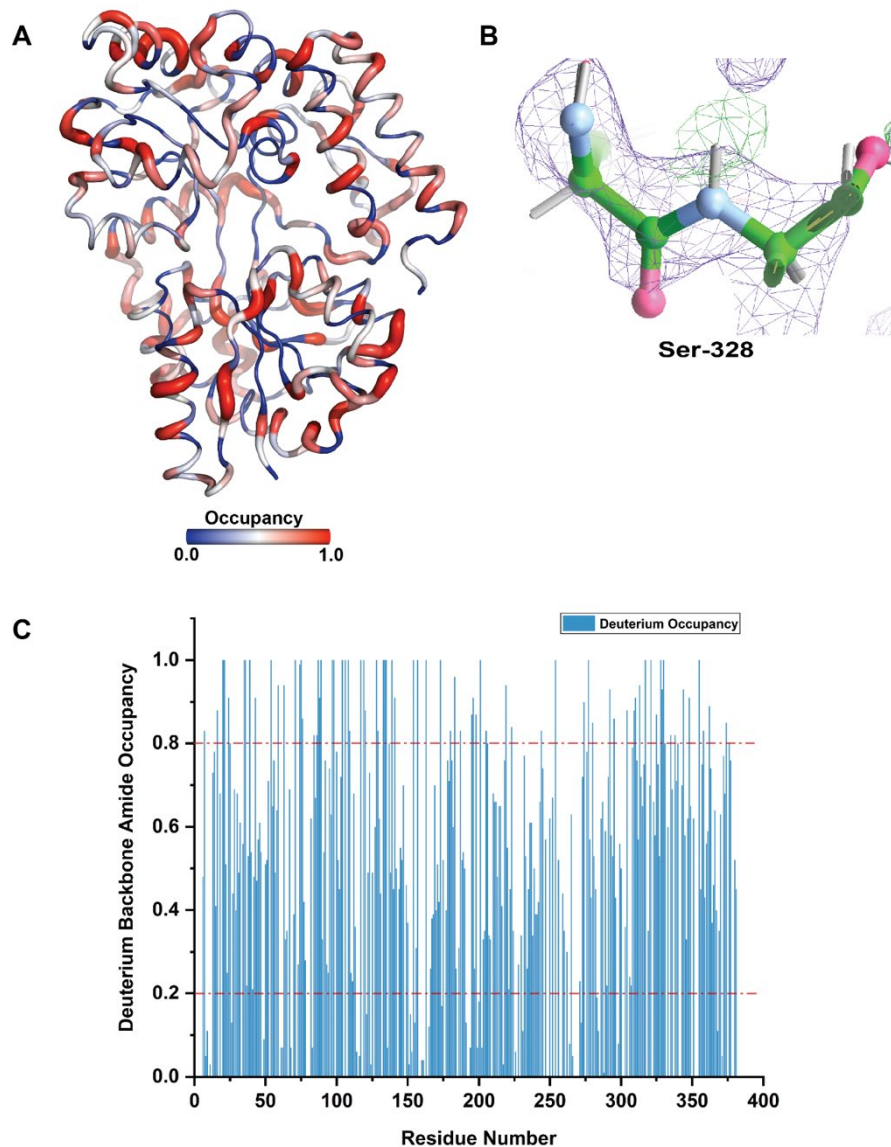


Figure 4.3: H/D exchangeable sites on tmMBP2-MTT complex. (A) A cartoon representation of tmMBP2 structure color coded with amide deuterium occupancy, with fully deuterated as red (occupancy of 1.0) and fully hydrogenated as blue (occupancy of 0.0) (B) An omit map (Fo-Fc map) showing the exchangeable amide site in Serine 328 as a green positive density (C) A percent deuteriation vs. residue map amide H/D exchange pattern in tmMBP2-MTT complex. An occupancy above the cutoff value of 0.8 was considered as fully deuterated site whereas the sites showing occupancy lower than 0.2 were classified as non-exchangeable or hydrogenated site.

NPC can reveal all of the above at physiological temperatures without incurring any observable radiation damage to the sample.⁽³³⁾ However, despite its advantages NPC is flux limited, and thus is not a trivial technique. There are frequently significant challenges in terms of growing very large crystals, often larger than 1 mm³, and sometimes unpredictable (and undesirable) effects from protein perdeuterations on crystallization behavior.⁽³⁴⁾ However, when large crystals can be obtained, the positives of NPC and added information on hydrogen far outweigh its limitations.

The premise of the current study is to investigate the stability of protein-substrate interactions that relate to the presence of water and water-mediated nonbonded contacts. As identified previously using XRD, the entire length of the substrate binding site in tmMBP2 is populated by water molecules that were inferred to form possible h-bonding contacts. Moreover, the binding site was also lined by several charged residues with exchangeable oxygen and nitrogen. Additionally, since the initial X-ray dataset on tmMBP2 was collected at cryogenic temperatures, we were aware that the large water interaction network could possibly have been an artifact of freezing the crystal, which can result in contraction of the crystal unit cell and local reorganization of solvent. Therefore, neutron diffraction experiments were designed to identify the extent of the water and water-mediated h-bond networks and to precisely identify the protonation states of the binding site forming residues at room temperature. NPC has been extensively used to determine the h-bond dynamics originating from the water molecules in previous studies.^(35,36)

The 2.1 Å neutron structure of tmMBP2 is the first neutron crystal structure of any MBP. As MBPs are uniquely structured to bind and transfer substrates to their respective ABC transporter systems,⁽⁷⁾ the neutron data can shed new light upon the specific details of h-bond interactions between the substrate and the protein, which are often mediated by water in their substrate-binding cavities. Indeed, the neutron structure of tmMBP2-MTT complex also showed that there was a large number of water (~30 molecules) that traverse the entire length of the substrate binding pocket out of which 13 form water-mediated interactions with the substrate MTT as was previously identified using XRD. These solvent molecules perform two major functions: (1) they form direct h-bonds with the substrate or

mediate h-bonds between protein and the substrate molecule, and (2) they stabilize the substrate-bound form of the protein by making additional non-bonded interactions with protein residues surrounding the binding cavity (as previously noted, the X-ray structure of tmMBP2 (PDB ID: 6DTS) was determined at cryogenic temperatures, it was also important to identify how the solvent in and around the binding cavity was ordered at more physiologically relevant temperatures). The room temperature neutron structure of tmMBP2 did show that these water molecules were indeed well ordered in and around the substrate binding site in tmMBP2 (Fig 4.1). Given the fact that the binding cavity in tmMBP2 is shallow and broad to accommodate larger oligosaccharide substrates, ⁽⁸⁾ we reason that the presence of an extended network of water molecule helps keep the substrate bound firmly to the protein. It was recently shown that perturbing this water network can significantly alter substrate-binding affinity of the protein. ^(37,38) Furthermore, the enthalpic and entropic contributions resulting in displacement of disorder water in apo-tmMBP2 with a more coordinated substrate-water interaction network in tmMBP2-MTT complex further improve protein stability. ^(39,40) Additionally, we note that the observed direct water-substrate h-bonds may possibly serve to prevent fluctuations of both ends of the tetrasaccharide that partially protrude out of the binding site.

The characterization of the protonation states of important residues in protein can help understand its interactions with substrates and can offer insights into the effects of local chemistry in substrate-binding or catalysis. ^(41,42) The neutron structure of tmMBP2-MTT complex revealed the protonation states of key charged and aromatic residues that participate extensively in forming h-bonds and van der Waals (VDW) interactions with the substrate. Furthermore, some of these residues (His-65, Trp-67, Tyr-213; Figure 4.2. A) also stabilize the closed-state of the protein by forming tight intermolecular h-bond interactions with adjoining residues. It was observed that hydroxyl side chain containing residues (Ser and Thr), amide side chain containing residues (Asn and Gln), charged residues (Lys, Arg, Asp, and Glu), and aromatic side chain containing residues (His, Tyr and Trp) all showed singly protonated site. Most of these residues formed h-bond interactions with nearby residues; however, some of the side chains also formed direct interactions with ordered solvent. The only exception was His-65 that had both N δ and N ϵ

sites H/D exchanged with prominent supporting neutron density suggesting a doubly protonated state. The doubly protonated state of His-65 is relevant given the crystallization condition of pD ~6.0.

Finally, the back exchange of the hydrogenated tmMBP2 in a D₂O environment provided some additional insight into the flexibility of the protein and its solvent interaction profile within the confines of the crystal lattice. As expected, most of the surface exposed and solvent accessible side chain groups had undergone significant H/D exchange. However, analysis of the pattern of H/D exchange on backbone amide positions showed significant differences in the extent of exchange across structure of the protein. Specifically, the beta-sheet core of the protein showed only limited H/D exchange, suggesting that the limited flexibility and packing of the beta sheet structure provided protection against H/D amide exchange. In contrast, some of the helices adjoining the binding pocket showed significant levels of amide H/D exchange. In general, the protein showed flexibility in the surface-exposed loop regions that were shown to be more flexible in apo-tmMBP2 using MD simulation.⁽⁸⁾ This could further suggest that the dynamics in tmMBP2 is influenced predominantly by the solvent-interacting loop regions that then modulate the motions towards the protein core. Together, these findings suggest the possible reason for a very tight substate binding interaction between tmMBP2 and MTT, as is also reported in Chapter 3.

List of References

1. Levy, Y., and Onuchic, J. N. (2006) Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 389-415
2. Barillari, C., Taylor, J., Viner, R., and Essex, J. W. (2007) Classification of water molecules in protein binding sites. *Journal of the American Chemical Society* **129**, 2577-2587
3. Michel, J., Tirado-Rives, J., and Jorgensen, W. L. (2009) Energetics of displacing water molecules from protein binding sites: consequences for ligand optimization. *Journal of the American Chemical Society* **131**, 15403-15411
4. Ladbury, J. E. (1996) Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chemistry & biology* **3**, 973-980
5. Levitt, M., and Park, B. H. (1993) Water: now you see it, now you don't. *Structure* **1**, 223-226
6. Lu, Y., Wang, R., Yang, C. Y., and Wang, S. (2007) Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J Chem Inf Model* **47**, 668-675
7. Duan, X., Hall, J. A., Nikaido, H., and Quiocho, F. A. (2001) Crystal structures of the maltodextrin/maltose-binding protein complexed with reduced oligosaccharides: flexibility of tertiary structure and ligand binding. *Journal of molecular biology* **306**, 1115-1126
8. Shukla, S., Bafna, K., Gullett, C., Myles, D. A. A., Agarwal, P. K., and Cuneo, M. J. (2018) Differential Substrate Recognition by Maltose Binding Proteins Influenced by Structure and Dynamics. *Biochemistry* **57**, 5864-5876
9. Meilleur, F., Munshi, P., Robertson, L., Stoica, A. D., Crow, L., Kovalevsky, A., Koritsanszky, T., Chakoumakos, B. C., Blessing, R., and Myles, D. A. (2013) The IMAGINE instrument: first neutron protein structure and new capabilities for neutron macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* **69**, 2157-2160
10. Meilleur, F., Kovalevsky, A., and Myles, D. A. (2020) IMAGINE: The neutron protein crystallography beamline at the high flux isotope reactor. in *Methods in Enzymology*, Elsevier. pp 69-85

11. Campbell, J. (1995) LAUEGEN, an X-windows-based program for the processing of Laue diffraction data. *Journal of Applied Crystallography* **28**, 228-236
12. Arzt, S., Campbell, J. W., Harding, M. M., Hao, Q., and Helliwell, J. R. (1999) LSCALE—the new normalization, scaling and absorption correction program in the Daresbury Laue software suite. *Journal of applied crystallography* **32**, 554-562
13. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography* **62**, 72-82
14. Collaborative, C. P. (1994) The CCP4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography* **50**, 760
15. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., and Grosse-Kunstleve, R. W. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* **66**, 213-221
16. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D: Biological Crystallography* **68**, 352-367
17. Afonine, P. V., Mustyakimov, M., Grosse-Kunstleve, R. W., Moriarty, N. W., Langan, P., and Adams, P. D. (2010) Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallographica Section D: Biological Crystallography* **66**, 1153-1163
18. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* **60**, 2126-2132
19. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* **66**, 12-21
20. Afonine, P. V., Mustyakimov, M., Grosse-Kunstleve, R. W., Moriarty, N. W., Langan, P., and Adams, P. D. (2010) Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **66**, 1153-1163

21. Logan, D. T. (2020) Interactive model building in neutron macromolecular crystallography. in *Methods in Enzymology*, Elsevier. pp 201-224
22. Rees, D. C., Johnson, E., and Lewinson, O. (2009) ABC transporters: the power to change. *Nat Rev Mol Cell Biol* **10**, 218-227
23. Liebschner, D., Afonine, P. V., Urzhumtsev, A. G., and Adams, P. D. (2020) Implementation of the riding hydrogen model in CCTBX to support the next generation of X-ray and neutron joint refinement in Phenix. in *Methods in Enzymology*, Elsevier. pp 177-199
24. Li, X., Jacobson, M. P., Zhu, K., Zhao, S., and Friesner, R. A. (2007) Assignment of polar states for protein amino acid residues using an interaction cluster decomposition algorithm and its application to high resolution protein structure modeling. *Proteins: Structure, Function, and Bioinformatics* **66**, 824-837
25. Kossiakoff, A. A. (1984) Use of the neutron diffraction—H/D exchange technique to determine the conformational dynamics of trypsin. in *Neutrons in Biology*, Springer. pp 281-304
26. Kossiakoff, A. A. (1985) The application of neutron crystallography to the study of dynamic and hydration properties of proteins. *Annual review of biochemistry* **54**, 1195-1227
27. Skinner, J. J., Lim, W. K., Bedard, S., Black, B. E., and Englander, S. W. (2012) Protein dynamics viewed by hydrogen exchange. *Protein Sci* **21**, 996-1005
28. Wales, T. E., and Engen, J. R. (2006) Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom Rev* **25**, 158-170
29. Rist, W., Rodriguez, F., Jorgensen, T. J., and Mayer, M. P. (2005) Analysis of subsecond protein dynamics by amide hydrogen exchange and mass spectrometry using a quenched-flow setup. *Protein Sci* **14**, 626-632
30. Helliwell, J. R. (2020) Fundamentals of neutron crystallography in structural biology. *Methods Enzymol* **634**, 1-19
31. Tomanicek, S. J., Standaert, R. F., Weiss, K. L., Ostermann, A., Schrader, T. E., Ng, J. D., and Coates, L. (2013) Neutron and X-ray crystal structures of a perdeuterated enzyme inhibitor complex reveal the catalytic proton network of the Toho-1 β -lactamase for the acylation reaction. *Journal of Biological Chemistry* **288**, 4715-4722

32. Kumar, P., Agarwal, P. K., Waddell, M. B., Mittag, T., Serpersu, E. H., and Cuneo, M. J. (2019) Low-Barrier and Canonical Hydrogen Bonds Modulate Activity and Specificity of a Catalytic Triad. *Angewandte Chemie* **131**, 16406-16412
33. Myles, D. A. (2006) Neutron protein crystallography: current status and a brighter future. *Current opinion in structural biology* **16**, 630-637
34. Fisher, S., Blakeley, M., Howard, E. I., Petit-Haertlein, I., Haertlein, M., Mitschler, A., Cousido-Siah, A., Salvay, A. G., Popov, A., and Muller-Dieckmann, C. (2014) Perdeuteration: improved visualization of solvent structure in neutron macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* **70**, 3266-3272
35. Oksanen, E., Chen, J. C.-H., and Fisher, S. Z. (2017) Neutron crystallography for the study of hydrogen bonds in macromolecules. *Molecules* **22**, 596
36. Chen, J. C.-H., Hanson, B. L., Fisher, S. Z., Langan, P., and Kovalevsky, A. Y. (2012) Direct observation of hydrogen atom dynamics and interactions by ultrahigh resolution neutron protein crystallography. *Proceedings of the National Academy of Sciences* **109**, 15301-15306
37. Darby, J. F., Hopkins, A. P., Shimizu, S., Roberts, S. M., Brannigan, J. A., Turkenburg, J. P., Thomas, G. H., Hubbard, R. E., and Fischer, M. (2019) Water networks can determine the affinity of ligand binding to proteins. *Journal of the American Chemical Society* **141**, 15818-15826
38. Oh, B.-H., Ames, G., and Kim, S.-H. (1994) Structural basis for multiple ligand specificity of the periplasmic lysine-, arginine-, ornithine-binding protein. *Journal of Biological Chemistry* **269**, 26323-26330
39. Biela, A., Nasief, N. N., Betz, M., Heine, A., Hangauer, D., and Klebe, G. (2013) Dissecting the hydrophobic effect on the molecular level: the role of water, enthalpy, and entropy in ligand binding to thermolysin. *Angew Chem Int Ed Engl* **52**, 1822-1828
40. Snyder, P. W., Mecinovic, J., Moustakas, D. T., Thomas, S. W., 3rd, Harder, M., Mack, E. T., Lockett, M. R., Heroux, A., Sherman, W., and Whitesides, G. M. (2011) Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc Natl Acad Sci U S A* **108**, 17889-17894

41. Yokoyama, T., Mizuguchi, M., Nabeshima, Y., Kusaka, K., Yamada, T., Hosoya, T., Ohhara, T., Kurihara, K., Tomoyori, K., and Tanaka, I. (2012) Hydrogen-bond network and pH sensitivity in transthyretin: Neutron crystal structure of human transthyretin. *Journal of structural biology* **177**, 283-290
42. Adachi, M., Ohhara, T., Kurihara, K., Tamada, T., Honjo, E., Okazaki, N., Arai, S., Shoyama, Y., Kimura, K., and Matsumura, H. (2009) Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography. *Proceedings of the National Academy of Sciences* **106**, 4641-4646

**Chapter 5 A promiscuous TRAP transporter
associated binding protein helps in
translocating essential metabolic
intermediates in *Pseudomonas aeruginosa***

5.1 Abstract

Tripartite ATP-independent periplasmic (TRAP) transporter is associated with the translocation of C4-dicarboxylates in bacteria and archaea. Additionally, TRAP transporters are implicated in promoting growth in several gram-negative bacterial pathogens inside their host by translocating alternate carbon sources. Opportunistic bacterial pathogen *Pseudomonas aeruginosa*, which is the cause of several human nosocomial infections, contains a TRAP transporter system that aids in the transport of succinate, malate, and other C4-dicarboxylate intermediates of the tricarboxylic acid (TCA) cycle. This study focused on delineating the mechanism of substrate-binding by the periplasmic substrate binding protein (SBP) associated with TRAP transport. Using X-ray crystallography and small-angle X-ray scattering (SAXS), it was shown that the *P. aeruginosa* TRAP-associated SBP (paTRAP-SBP) has the signature bidomain PBP fold, with each domain consisting of central β -strand core surrounded by α -helices. Two β -strands and a long α -helix that spans across both N- and C-terminal domains form the “mixed hinge” in paTRAP-SBP. The substrate was stabilized in the binding pocket with hydrogen bonds, salt-bridge, and van der Waals interactions. Biophysical characterization identified succinate, malate, and fumarate as cognate substrates for this TRAP transporter. Additionally, computational characterization of two more SBPs, one associated with another TRAP system and the other an orphan SBP, was performed. The three identified SBPs had high structural similarity with near overlap of the binding site residues. Therefore, this study has identified two TRAP systems that potentially translocate C4-DA intermediates of TCA cycle in *P. aeruginosa*.

5.2 Introduction

One of the most important areas of research in biomedical science is to comprehend and curtail the growing risk of multiple drug resistance in pathogenic bacteria. However, a failure to develop antibacterial lead compounds at a faster rate has left the scientific community to play catch-up with the ever-evolving pathogens. ⁽¹⁾ One strategy to develop and design better and perhaps more targeted therapeutics against multiple drug resistance is to identify novel targets and mechanisms that target the growth and

development of the pathogens, especially their central metabolic pathways such as the tricarboxylic acid (TCA) cycle. ⁽²⁾

Gram negative opportunistic bacterium *Pseudomonas aeruginosa* in particular has proven effective in developing resistance against multiple antibiotic agents and is a leading cause of nosocomial infections. *P. aeruginosa* is also implicated in several other ailments, including acute pneumonia, bacteremia and wound, urogenital-sepsis, intra-abdominal sepsis, and chronic airway infections with a mortality rate that can exceed 60%. ⁽³⁾ The bacterium has both intrinsic as well as acquired extrinsic factors, which often include targeted mutations, increased expression of drug efflux proteins, and expression of drug metabolizing enzymes, that help it evade antimicrobial agents. ⁽³⁾ Today *P. aeruginosa* features as the one of the “Priority 1 or Critical” organisms listed by the World Health Organization’s for which developing a new therapeutic agent is of highest priority. ⁽⁴⁾ It is therefore urgent to check the proliferation of this pathogen by targeting that counter the metabolism and metabolite acquisition in *P. aeruginosa*.

Bacteria proliferate in challenging environment niches, often surviving under dwindling levels of nutrients that can support their growth. To gain a competitive advantage in nutrient uptake, bacteria and archaea have developed a set of membrane transport proteins that can mediate critical substrate translocation without incurring an energy deficit. This ATP-independent ability to uptake essential nutrients is central to bacterial adaptability, especially pathogenic bacteria, particularly while adapting from native to host environment. TRAP and TRAP-like transporter proteins form an important class of secondary transporters that couple substrate translocation across the membrane with an Na⁺/H⁺ gradient. ⁽⁵⁾ TRAP transporters are essential for the translocation of a myriad substrate catalog, which predominantly includes C4-dicarboxylic acids (C4-DAs) and their derivatives. Some of these C4-DAs, such as succinate, malate, and fumarate, are important intermediates of the TCA cycle and help ATP production. ⁽⁵⁾

The TRAP and TRAP-like transporters can be classified in three main types: tripartite ATP-independent periplasmic (TRAP) transporter, TRAP associated extra-cytoplasmic immunogenic (TAXI) protein, and tripartite tricarboxylate transporter (TTT) (Figure 5.1.).

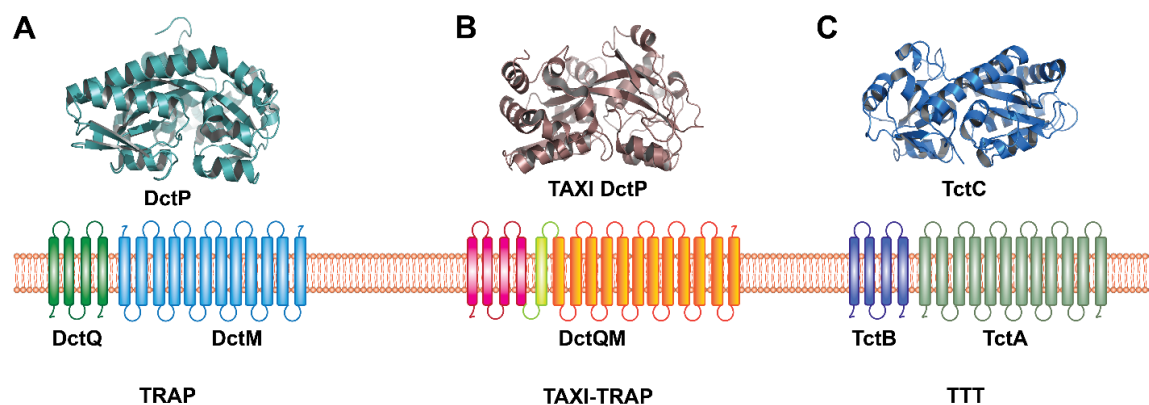


Figure 5.1: Topologies of TRAP, TAXI-TRAP, TTT transporters. (A) TRAP transporters have three components, an extracellular periplasmic substrate-binding protein DctP (PDB ID: 3B50), and two unequal membrane proteins DctQ (4 TM; shown in green) and DctM (12 TM; shown as cyan). TRAP DctPs have a distinctively large domain spanning helix that makes up the hinge with two long β -strands. DctQ has $N_{IN} C_{IN}$ topology, and DctM shows a $N_{OUT} C_{OUT}$ topology of TM organization in the membrane. (B) TAXI-TRAP transporters are unique in having only two components, DctP (PDB ID: 1US5) and a membrane permease DctQM, where DctQ (shown in pink) and DctM (shown in orange) is linked by an additional TM helix (shown in yellow), consisting of 17 TM helix and a $N_{IN} C_{OUT}$ topology. The hinge of the DctP is made up of only two central domain-spanning β -strands. (C) TTT transporters have a periplasmic SBP termed TctC has a helix consisting of two central β -strands, one large β -strand, and a shorter β -strand, along with a domain spanning helix, which is shorter than the TRAP DctP. The membrane proteins TctB (shown in blue) and TctA (shown in olive) consist of 4 and 12 TM helices. Both TctA and TctB have a $N_{IN} C_{IN}$ topology.

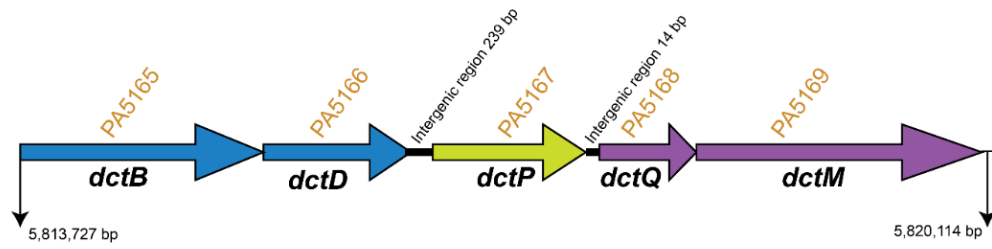
⁽⁶⁾ Structurally these three transporters are similar. The TRAP and TTT systems have three components, comprising one SBP and two unequal TM proteins with 4 and 12 TM helices. In TAXI-TRAPs the individual membrane proteins are connected by a single TM helix (Figure 5.2.B, shown in yellow). There are also subtle differences in the SBP associated with these individual transporter systems. Together these systems offer a path towards importing several essential sugar acids and trace metabolites and allow the bacteria to modify their metabolism and incorporate the imported metabolites as carbon or energy sources. ⁽⁷⁾

P. aeruginosa is known to utilize several TCA-cycle intermediates as carbon or energy source. ^(8,9) Cell growth assays on deletion mutants of *dctPQM* (PA5167 to PA5169) genes (Figure 5.2.A) showed no growth on media supplemented with malate and fumarate, a basal growth on succinate, and normal growth on oxaloacetate. ⁽⁹⁾ This suggested that this TRAP transporter was able to translocate malate and fumarate, and possibly succinate. The study also proposed a possibility of a second TRAP-like transporter system that was essential for the translocation of succinate across the bacterial membrane. Given the dependence of *P. aeruginosa* on TCA cycle intermediates, it is important to characterize these bacterial TRAP transporter systems to gain insight into the molecular mechanism of substrate translocation in this opportunistic pathogen.

This study aimed to identify TRAP transporter associated proteins in *P. aeruginosa* that are involved in binding and translocating C4-DAs using both computational and experimental approaches. Computational operon analysis combined with gene ontology searches and homology model building of important substrate binding proteins (SBP) was performed to identify and characterize the structural and functional organization of TRAP transporter systems in *P. aeruginosa*. Additionally, biophysical, and experimental structure determination techniques were used to characterize the substrate-binding profiles of one of the SBPs associated with a TRAP transporter. Overall, our results support the previously made assessment that indicated the presence of more than one TRAP transporter system that is involved in C4-DA translocation in *P. aeruginosa*.

Figure 5.2: C4-dicarboxylate TRAP transporter in *Pseudomonas aeruginosa*. (A) Operon-II structure of TRAP transporter showing genes coding for a two-component systems *dctB* (PA5165) and *dctD* (PA5166), a *periplasmic binding protein* *dctP* (PA5167), and two unequally sized TM proteins *dctQ* (PA5168) and *dctM* (PA5169). (B) Schematic representation of C4-dicarboxylate transport in *P. aeruginosa*, where availability of C4-dicarboxylates, like succinate, triggers their initial translocation across the outer membrane (OM) through porins. In the periplasm, an increased titer of C4-dicarboxylates results in the activation of two-component systems DctB, which subsequently passes on the signal to DctD upon phosphorylation. DctD in presence of regulatory factor RpoN upregulates the *dctPQM* operon resulting a tripartite system (DctP, DctQ, and DctM) that utilizes sodium ion gradient to transport C4-dicarboxylates inside the cell.

A



B

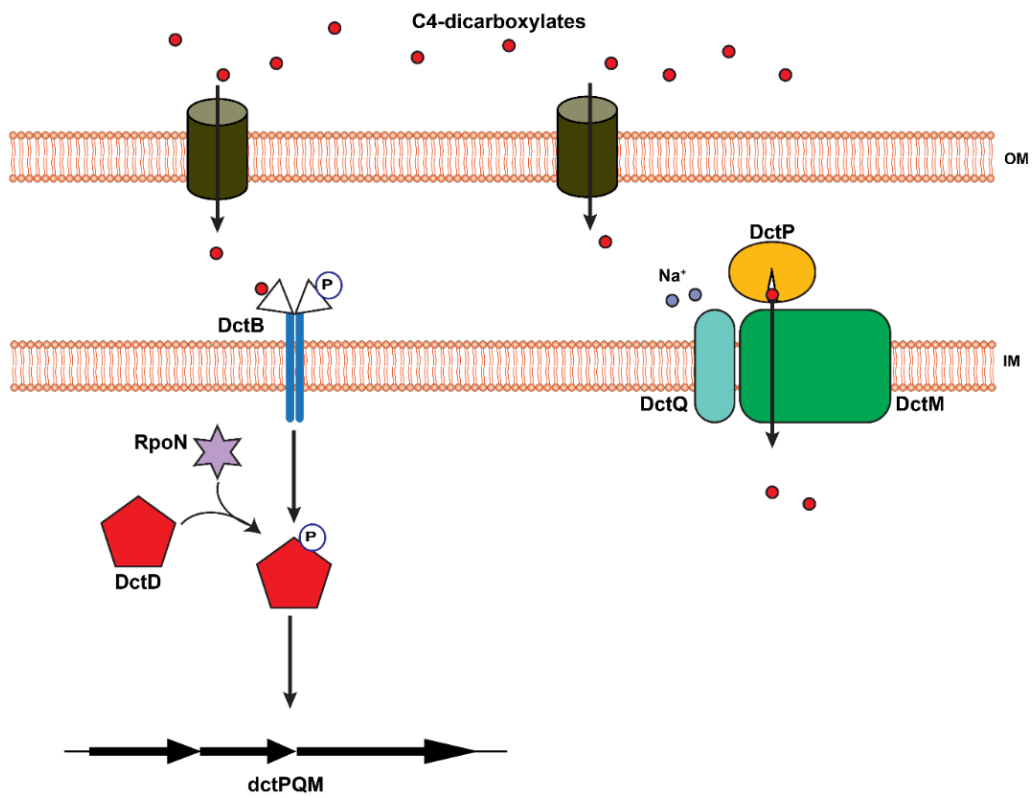


Figure 5.2: C4-dicarboxylate TRAP transporter in *Pseudomonas aeruginosa*.

5.3 Material and Methods

Computational sequence analysis and protein modelling

To identify the TRAP operons in Sequence similarity searches were carried out using the National Center for Biotechnology Information (NCBI) basic local alignment search tool (BLAST) program suite.⁽¹⁰⁾ tBLASTn program was used with an amino acid sequence of PA5167 as query against the RefSeq genome database (refseq_genomes) for organism *P. aeruginosa* PAO1.⁽¹¹⁾ Results were manually scrutinized for the loci, and sequence identity scores. Subsequently, each locus was manually checked for the presence of a TRAP operon. Potential TRAP proteins were identified using sequence similarity searches using the position specific iterated BLAST (PSI-BLAST) algorithm under the protein BLAST search (BLASTp) to identify high sequence conservation. BLASTp results were also compared with a similar result from the HHPred server, which does a profile-based alignment of the multiple sequences rather than position specific iterations used in PSI-BLAST.⁽¹²⁾

A homology model of PA0884 protein was built using MODELLER version 9.25.⁽¹³⁾ Signal sequence of the SBP was identified using SignalP and 23 amino acids showing high signal sequence score were removed out prior to a BLAST search. The TRAP-SBP from *Rhodopseudomonas plausris* (PDB ID: 4O94)⁽¹⁴⁾ was identified as the best template suited for homology modelling using a PSI-BLAST search against the protein data bank (PDB). The template showed a percent identity of 60.53% and a query coverage of 98%. 24 N-terminal and 2 C-terminal residues not observed in the crystal structure were not modeled. The models were assessed based on the DOPE score and molpdf score calculated by MODELLER.⁽¹⁵⁾ The best-scored structure was selected for comparative analysis with other TRAP SBPs from *P. aeruginosa*. Multiple protein sequence alignment was performed using Clustal Omega webserver with default parameters.⁽¹⁶⁾

Expression and purification of paTRAP-SBP

The identified paTRAP-SBP gene (PA5167) was cloned in a pCOLADuet vector with a hexahistidine tag and subsequently used to transform BL21 RIL *E. coli* cells. The transformed cells were grown in Luria-Bertani (LB) broth, supplemented with 0.1% glucose to promote cell growth, for the substrate-bound form of the protein. The apo protein was

grown in the Enfor's minimal media, which was devoid of any sugar acid derivatives and instead used glycerol as a carbon source. The cells were initially grown at 37 °C to an optical density (O.D.) reached 0.7. It was then induced using 1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) and grown overnight at 20 °C. The cells were spun down at 7000 rpm with centrifuge temperature set to 4 °C. The supernatant was discarded, and the cell pellets were stored at -80 °C for subsequent purification steps. A small fraction of the transformed cells was lysed, and the presence of target protein (37 KDa) was confirmed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) against standard protein markers.

paTRAP-SBP purification was carried out using a two-step chromatographic purification process involving an initial Ni-affinity chromatography followed by a gel filtration setup. The entire purification protocol was carried out at 8 °C in a cold room to minimize temperature-dependent protein denaturation. Before purification, the cells stored at -80 °C were thawed on ice and subsequently mixed with ice-cold lysis buffer/Buffer A (20 mM Imidazole, 20 mM MOPS, and 300 mM NaCl pH 7.5). Once the cells homogenously mixed with the buffer, they were lysed using a sonicator on ice with 0.5 sec/0.5 sec on/off pulses for 90 seconds at 50% amplitude. The lysed cell mixture was transferred into 50 ml Oak Ridge centrifugation tubes and 0.1% w/v of 10% polyethyleneimine (PEI) at pH 7.5 was added to the tubes to remove DNA or RNA during the centrifugation process. The lysate was then centrifuged at 18,000 rpm at 4 °C for 30 minutes. The pellets were discarded, and the supernatant was filtered and used for affinity purification on HiTrap Chelating HP column (GE Healthcare Life Sciences) using an Aktä Prime (GE Healthcare Life Sciences). To begin with, the HiTrap column was thoroughly washed with 10 column volumes (CVs) of water to remove any traces of the storage solution followed by five CVs of 100 mM NiSO₄ to charge the column. Subsequent wash with 10 CVs of water removed the unbound nickel ions. Finally, the column was charged using 10 CVs of Buffer A and the filtered supernatant was loaded onto the column using a sample pump with a loading rate of 1 ml/min to compensate with the reduction in binding affinity of the HiTrap columns at colder temperatures. The loaded column was washed with 10 CVs of Buffer A to remove the unbound proteins before finally eluting the target protein using Buffer B (200 mM Imidazole, 20 mM MOPS, 300 mM NaCl pH 7.5). The affinity purified protein was

subsequently dialyzed in the gel filtration buffer (20 mM TRIS, 20 mM MOPS, 150 mM NaCl pH 7.5) overnight at 8 °C.

For the second step of the purification process, Superdex 75 HiLoad 26/60 gel filtration column (GE Healthcare Life Sciences) was equilibrated using 2 CVs of deionized filtered and degassed water followed by 2 CVs of filtered and degassed gel filtration buffer on an Aktä Prime (GE Healthcare Life Sciences) at 8 °C. 10 ml of dialyzed sample was loaded onto the column using a superloop. Finally, fractions were collected, and the presence of target protein (37 KDa) was estimated using SDS-PAGE. Fractions containing purified target protein were pooled together and concentrated using Vivaspin 10,000 MW concentrator (Sartorius Tech.) and dialyzed in dialysis buffer (20 mM TRIS, 20 mM MOPS, 40 mM NaCl pH 7.5) overnight at 8 °C.

Differential Scanning Calorimetry (DSC)

The proteins used for the DSC experiments was expressed in minimal media grown cells, in order to reduce chance binding of C4-DAs from the medium. The purified protein was extensively dialyzed in the phosphate buffer pH 7.5 and concentrated to a final concentration of 2 mg/ml. The protein solution and buffer were vacuum degassed to remove small bubbles, and subsequently 300 µl of protein was loaded onto a Nano DSC (TA Instruments). The buffer was loaded in the reference chamber and the protein solution in the sample chamber using sterile pipettes. The DSCrun program (TA Instruments) was used to control the run on the instrument. The thermal denaturation was carried out for temperature ranging between 10 °C to 100 °C, at a rate of 1 °C/min. The generated data was analyzed in the NanoAnalyze software (TA instruments).

Small-angle X-ray scattering (SAXS)

All SAXS datasets were generated on Rigaku BioSAXS-2000 running on a MicroMax-007 X-ray generator at the Oak Ridge National Laboratory (ORNL). 100 µl of buffer and protein solution, at different percent dilutions, were loaded into the auto-sampler. All the SAXS data were collected at 20 °C. The initial data reduction and buffer subtraction was performed using Rigaku Automatic Data Subtraction Pipeline. The scattering dataset was transformed into one-dimensional scattering profiles as a function of the scattering vector

Q , where $Q = 4\pi\sin\theta/\lambda$, λ is the wavelength of the incident X-ray radiation and 2θ is the detector angle. All the generated SAXS datasets were analyzed in the ATSAS program suite.⁽¹⁷⁾ The $P(r)$ analysis was performed using GNOM⁽¹⁸⁾ and the three-dimensional envelop was generated using GASBORI script.⁽¹⁹⁾ The crystal structure was superimposed on the generated envelop using SUPCOMB.⁽²⁰⁾ Finally, the experimental and simulated scattering data from solution scattering and crystal structure, respectively, were fitted using CRY SOL.⁽²¹⁾

Crystallization of paTRAP-SBP

Purified and extensively dialyzed protein sample in crystallization buffer (20 mM Tris, 20 mM MOPS, and 40 mM NaCl pH 7.5) was used for crystallization experiments. The protein was concentrated to 15 mg/ml and initially used for a sparse-matrix screening for apo, succinate-bound, and malate-bound proteins using Gryphon robotics (Art Robbins International). Among the over 900 different screens tried, the target protein crystallized in over 30 and predominantly in solutions containing polyethylene glycol (PEG). Optimizations were carried out on most promising conditions using vapor diffusion sitting and hanging drop trays. Crystals appeared in both hanging and sitting drops. The method of microseeding was also used to promote crystallization in meta-stable zone. Final crystallization conditions 2 μ l of paTRAP-SBP at 15mg/ml was mixed with 2 μ l of mother liquor and equilibrated against 0.1 M Tris HCl, 0.2 M $MgCl_2$, 28% Polyethylene glycol 4000 MW. The crystal trays were incubated at 10 °C and crystals appeared in and around two weeks from the date of trays being setup.

Diffraction data collection and analysis

Since the crystals grew as radiating needles that broadened on the tips (Figure 5.6. A), a single needle (~200 microns) was removed out carefully using crystal manipulation tools under a microscope. The crystal was subsequently washed in the mother liquor and mounted on a nylon loop. The MiTeGen (Ithica New York) room-temperature MicroRT plastic sleeve mounts were used to prepare the crystal for a room temperature data collection. The data were collected on the Rigaku HighFlux HomeLab instrument at the Oak Ridge National Laboratory equipped with a MicroMax-007 X-ray source and Osmic VariMax optics. Realtime diffraction data were collected on a EIGER R 4M hybrid photon

counting detector. The data were integrated using the *CryAlis^{Pro}* program (Rigaku, The Woodlands, Texas). The integrated data were then reduced and scaled in *AIMLESS* program from the CCP4 software package. ⁽²²⁾ A molecular replacement was performed using the Phaser program in the Phenix suite with PDB 4O94 as a template. ⁽²³⁾ The initial model was built using Phenix automated modelling tool. ⁽²⁴⁾ Refinement of the built model was carried out using Phenix Refine tool. ⁽²⁵⁾ Manual model building was carried out using the Coot molecular builder program. ⁽²⁶⁾ Finally the model geometry and validity were accessed using MOLPROBITY ⁽²⁷⁾ and PDBSum. ⁽²⁸⁾

5.4 Results

5.4.1 Dissecting the TRAP transporter operons in *P. aeruginosa*

Three independent TRAP-associated SBPs were identified in this study using bioinformatic analysis of *P. aeruginosa* PAO1 reference genome. PA5169 sequence was used as a query and three different SBP loci were identified across the genome that had sequence similarity of above 50%. Out of the three, two (PA5167 and PA0884) were found to be associated with operons coding for additional TRAP proteins, the two component systems DctB/DctD and membrane permeases DctQ/DctM. The third SBP (PA4616) was identified as an orphan SBP that was not associated with either a TRAP two-component system or TRAP permeases on the same operon. The structure of the orphan SBP was determined and published in 2015 by Vetting et al. (PDB ID: 4NF0). ⁽¹⁴⁾ The published structure was crystallized with malate bound at the protein binding site. Moreover, an earlier detailed biochemical analysis of one of the two TRAP operons also suggested that the cognate substrates of PA5167 were malate, succinate, and fumarate. ⁽⁹⁾ That study also noted that there was a basal growth rate observed in the deletion mutants of PA5165 to PA5169 in presence of succinate and concluded that there might be another succinate transporter in the *P. aeruginosa*. Here, we used bioinformatics tools to identify another operon that codes for a novel TRAP system (PA0882 to PA0886), which might also be involved in translocation of succinate, malate, and fumarate and making it the possible third transporter of succinate in *P. aeruginosa*.

For the two SBPs associated with other TRAP-proteins, the functional annotation of the genes on both operons have upstream two component system genes *dctB/dctD* followed

by the *dctPQM* transporter protein coding genes. Depending on their position in the *P. aeruginosa* genome here they are classified as operon-I (PA0882 to PA0886) and operon-II (PA5165 to PA5169). Arrangement of genes on Operon-II is depicted in Figure 5.2. A. Interestingly, there are two intergenic regions on this operon that flank both 5'- and 3'-ends of the *dctP* gene, which codes for the SBP. This could suggest that both two-component proteins and membrane permease proteins might be co-expressing. This is known to be true for the *Vibrio cholera* SiaQM proteins that don't remain functionally viable if expressed separately. ^(29,30) Operon-I also has similar intergenic elements inserted on both ends of *dctP* gene. The larger intergenic region of 239 base pairs (bp) on operon-II has a AmrZ binding site, which is a global transcriptional regulator that plays a key role in biofilm development and environmental adaptation of *P. aeruginosa* and related subspecies. ^(31,32) This suggests a possible role for these TRAP systems in promoting proliferation of *P. aeruginosa* during a shift from their natural habitat into the host, potentially aiding in their pathogenesis.

The two-component systems DctB and DctD are involved in the expression regulation of downstream DctPQM TRAP transport. ⁽⁹⁾ In presence of an elevated titer of C4-DA, DctB, a membrane-bound sensor protein, auto-phosphorylates and subsequently phosphorylates DctD. Activated DctD subsequently employs RpoN sigma factor to incorporate RNA polymerase that translates *dctPQM* (Figure 5.2. B).

5.4.2 paTRAP-SBP has a broad substrate range

In a previous study it was observed that the wild-type *P. aeruginosa* PAO1 was able to grow in minimal media supplemented with succinate, malate, and fumarate but did not show significant growth in presence of oxaloacetate and glucose. ⁽⁹⁾ Subsequently, cell-based growth assay on the TRAP deletion mutants indicated that the deletion mutants of *dctPQM* (operon-II) genes were unable to grow on minimal media supplemented with malate and fumarate, whereas some basal growth was observed in presence of succinate. This suggested that gene-products of the operon-II were the sole translocators of malate and fumarate. This DctPQM TRAP system could also translocate succinate, but the deletion mutant studies concluded that there was another succinate transporter in *P. aeruginosa* PAO1.

To access the ability of the SBP (PA5167) to bind different intermediates of the TCA cycle, a DCS experiment was performed on the SBP in the presence of five substrates (succinate, malate, fumarate, citrate, and α -ketoglutarate). The change in the thermal melting temperature (ΔT_m) for the protein-substrate interactions was compared to the ΔT_m of the apo-protein to identify substrates that can cause a significant shift in ΔT_m as the result of a stable protein-substrate interaction. The more specific an interaction, the higher the ΔT_m of the protein in the presence of that particular substrate due to an increased number of non-bonded contacts, especially hydrogen bonds (h-bonds), that stabilize the substrate in the binding pocket of the protein.

The DCS results are summarized in the Table 5.1. The apo SBP showed a multiple transition state with overlaps. A T_m of approximately 60 °C (Figure 5.3. A) was estimated by fitting a two-step scaled model. A total of 4 transition peaks were fitted with this model. Similar transition states were also observed for α -ketoglutarate (Figure 5.3. B) and citrate (Figure 5.3. C), with the T_m of roughly 53 °C. As previously observed in case of antibodies, overlapping peaks can result due to a flexible hinge region and closely interacting regions on the protein. ⁽³³⁾ Additionally, it can also be noted that the overall T_m reduced when α -ketoglutarate and citrate were mixed with the protein. The similar behavior of the SBP protein in the absence and presence of substrates suggests a lack of affinity for these two TCA cycle C4-DAs. Therefore, both α -ketoglutarate and citrate are likely non-cognate ligands that do not get transported by the TRAP transporter.

In contrast, SBP in the presence of fumarate (Figure 5.4.A), malate (Figure 5.4.B), and succinate (Figure 5.4.C) all showed average changes in ΔT_m by over 15 °C when compared to the apo protein. The single denaturation event was also fitted using the two-state scaled model and there were observable differences in the peak shapes of the protein in presence of these substrates. For protein in the presence of succinate, the peak shape is much narrower when compared with the peak shape in the presence of malate and fumarate, suggesting a tighter and more cooperative binding with the protein.

Table 5.1: Comparative analysis of protein stability in presence of different TCA cycle substrates. All the experiments were performed in the temperature range of 10 °C to 100 °C. A constant sample concentration of 2 mg/ml was used for the data collection.

<i>Sample</i>	<i>T_m</i> (°C)	ΔH (kJ/mol)	ΔS (kJ/(mol·K))	Pre- Transition Start Temp (°C)	Pre- Transition End Temp (°C)	Post- Transition Start Temp (°C)	Post- Transition End Temp (°C)
Apo	60.58	892.77	2.67	22.10	30.99	75.25	84.13
α-Ketoglutarate	53.05	897.44	2.75	20.66	29.24	75.02	83.60
Citrate	53.00	817.26	2.50	19.99	29.44	74.38	83.83
Fumarate	76.74	664.20	0.87	65.87	69.59	80.77	84.49
Malate	76.51	743.70	1.24	63.78	68.39	84.55	89.66
Succinate	78.44	751.20	1.23	65.26	68.83	86.24	89.81

Figure 5.3: Thermal denaturation data on paTRAP-SBP non-cognate ligand. An overlapped thermal denaturation peaks of apo protein (A), α -ketoglutarate (B), and citrate (C) fitted using the TwoStateScaled model in NanoAnalyze. All three systems exhibited four intermediate states between the native and denatured states that were fitted using four different models. The sum of all four models (dashed line) is overlaid over the raw data (green solid line)

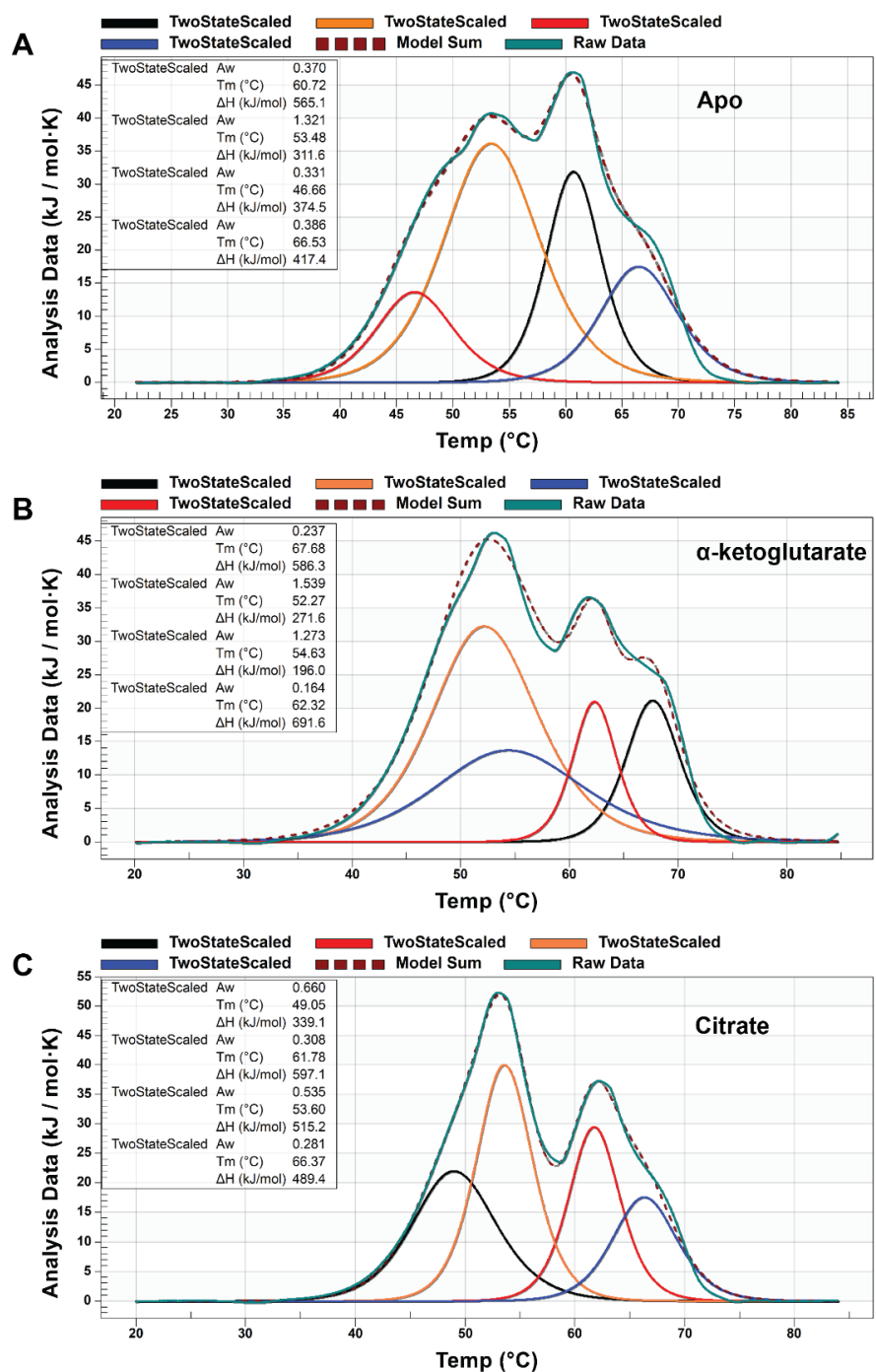


Figure 5.3: Thermal denaturation data on paTRAP-SBP non-cognate ligand.

Figure 5.4: Thermal denaturation data from DSC for paTRAP-SBP with cognate substrates. A single transition peak observed for cognate substrates fumarate (A), malate (B), and succinate (C) showing TwoStateScaled model in NanoAnalyze. The raw data is represented as a solid line and the corresponding model fit in dashed lines.

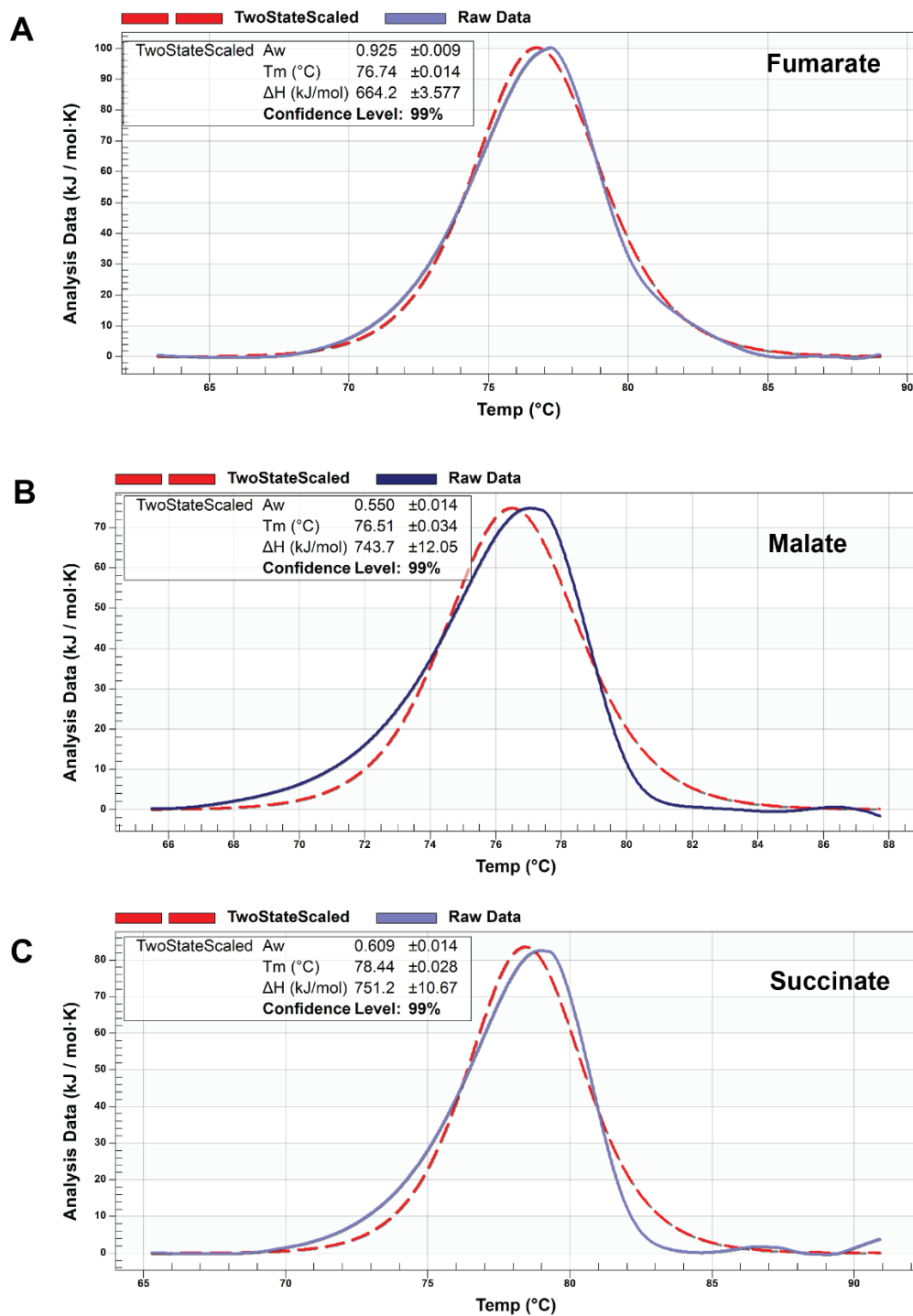


Figure 5:4: Thermal denaturation data from DSC for paTRAP-SBP with cognate substrates.

The biophysical assay to characterize protein-substrate binding using DCS was also supported by structural characterization using small-angle X-ray scattering (SAXS) data. SAXS was performed on the apo protein and on succinate- and malate-bound proteins (Figure 5.5). There were significant differences observed in the radius of gyration (R_g) and D_{max} values of apo vs. substrate-bound proteins (Table 5.2), specifically a difference of ~ 1 Å in D_{max} in apo vs. substrate-bound systems, which is indicative of the hinge-bending in the SBP upon substrate binding. Similar observations have been made in the solution scattering data from other members of this superfamily. ^(34,35) Moreover, given that both succinate and malate showed a higher T_m and a significantly higher ΔH from the DSC analysis, they do stabilize the SBP after binding. Similarly, an overlay fit of the solution scattering data with the three-dimensional diffraction data (PDB: 2CEY for apo sample, paTRA-SBP XRD structure from this study for succinate and malate-bound systems) using CRY SOL showed a good agreement between the experimental and theoretical scattering intensities, suggesting that the three-dimensional crystal structures of the SBP are consistent with its solution state structure. It also confirmed a conformational change post substrate binding. Furthermore, the three-dimensional envelope reconstruction using GASBOR and a subsequent superimposition of the crystal structures using SUPCOMB clearly showed that the envelope compacted for substrate-bound systems (χ^2 of 1.14 and 1.39 for succinate and malate-bound systems, respectively) and was elongated for the apo protein (χ^2 of 2.02). This indicates that binding of succinate and malate to the apo-protein induced a large-scale conformational change in the protein structure mediated by the bending of the hinge, which locked the substrate inside the protein binding cavity.

Thus, both biophysical and structural characterization of the paTRAP-SBP strongly indicate that succinate, malate, and fumarate are cognate substrates for this protein, whereas citrate and α -ketoglutarate are non-cognate ligands that do not bind to the SBP. These results are consistent with the initial growth kinetics experiment by Valentini *et al.* ⁽⁹⁾ where it was shown that the deletion mutants of *dctPQM* (Operon-II) in *P. aeruginosa* were unable to grow on minimal media supplemented with succinate, malate, and fumarate, but the cells grew on oxaloacetate. Taken together, these results indicate that the paTRAP system is selective to succinate, fumarate, and malate but it is not involved in the transport of citrate, α -ketoglutarate, and oxaloacetate

Figure 5.5: Small-angle X-ray scattering data of paTRAP-SBP in presence and absence of ligand. (A) Q vs. intensity plot of the apo paTRAP-SBP showing scattering data as green dots and the corresponding structure fit (PDB: 2CEY) and the envelope structure superimposed with the X-ray crystal structure in the cartoon form. Scattering data of succinate- (B) and malate-bound (C) SBP showing one dimensional solution and crystal structure fits with calculated envelope structures superimposed on the crystal forms of these proteins. paTRAP-SBP structure determined in this study was used to fit the substrate-bound SBP scattering datasets.

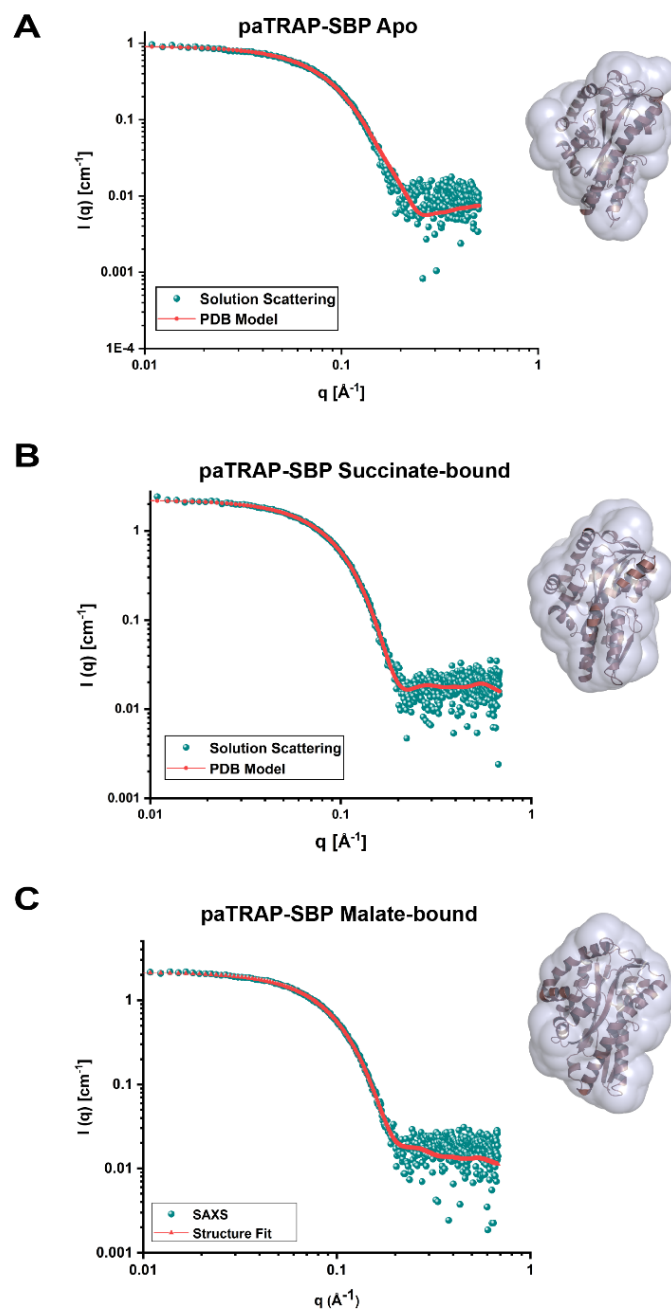


Figure 5.5: Small-angle X-ray scattering data of paTRAP-SBP in presence and absence of ligand.

Table 5.2: Small-angle X-ray scattering data analysis results for paTRAP-SBP. The SAXS data analysis results for paTRAP-SBP APO and in presence of substrates succinate and malate. The analysis was done using ATSAS software package.

<i>Sample</i>	Sample conc. (mg/ml)	R_g (Å)	D_{max} (Å)	χ^2*
<i>paTRAP-SBP APO</i>	6.5	20.1	64.9	2.02
<i>paTRAP-SBP Succinate</i>	7.3	19.9	63.6	1.14
<i>paTRAP-SBP Malate</i>	7.2	19.8	63.1	1.39

* χ^2 generated from the fit of the CRY SOL calculated scattering profile to the experimental data

5.4.3 Structure of paTRAP-SBP reveal its mode of substrate binding

The X-ray crystal structure of the succinate-bound paTRAP-SBP (PA5167) from the operon-II of *P. aeruginosa* PAO1 has been determined. The protein was also crystallized in its apo form, and with malate bound to the protein in initial sparse-matrix trials. However, these crystals were too small or clustered for diffraction and repeated rounds of optimization did not yield diffraction sized single crystals for either the apo or malate-bound protein. For the succinate-bound SBP, crystals grew as radiating needles that flattened-out on the edges. These needle crystals were surgically removed by teasing them apart carefully from the base of the needle cluster. A room temperature data collection resulted in the crystal diffracting to 2.4 Å. While processing the paTRAP-SBP diffraction dataset in CrysAlis^{PRO} (Oxford Diffractions Ltd.), the autoindexing program in CrysAlis^{PRO} flagged an unusually large c edge of the unit cell, which was initially estimated at ~340 Å. However, on a closer examination of the measured data it was observed that every fifth peak was systematically brighter on the largest cell edge. Visual inspection of the raw diffraction images revealed two overlapping crystal lattices, indicating that the crystal was double, the largest edge was shortened to a fifth of its initial value and the data were reintegrated with appropriate twinning filters turned on. This resulted in two independent single crystal datasets with an overall data completeness of 86.10% and 86.67%, respectively. The data sets from the two individual crystals were merged and used for structure refinement.

Molecular replacement was performed using the Phaser program in the Phenix suite with PDB 4O94 as a template, which had a sequence homology of 60% with the target protein. Initial model was calculated using the Phenix Autobuild module and this model was later manually fitted in COOT ⁽³⁶⁾ and refined in Phenix ⁽³⁷⁾ to get the final structure (Figure 5.6.A). Crystallographic data collection and accompanying statistics are given in Table 5.3.

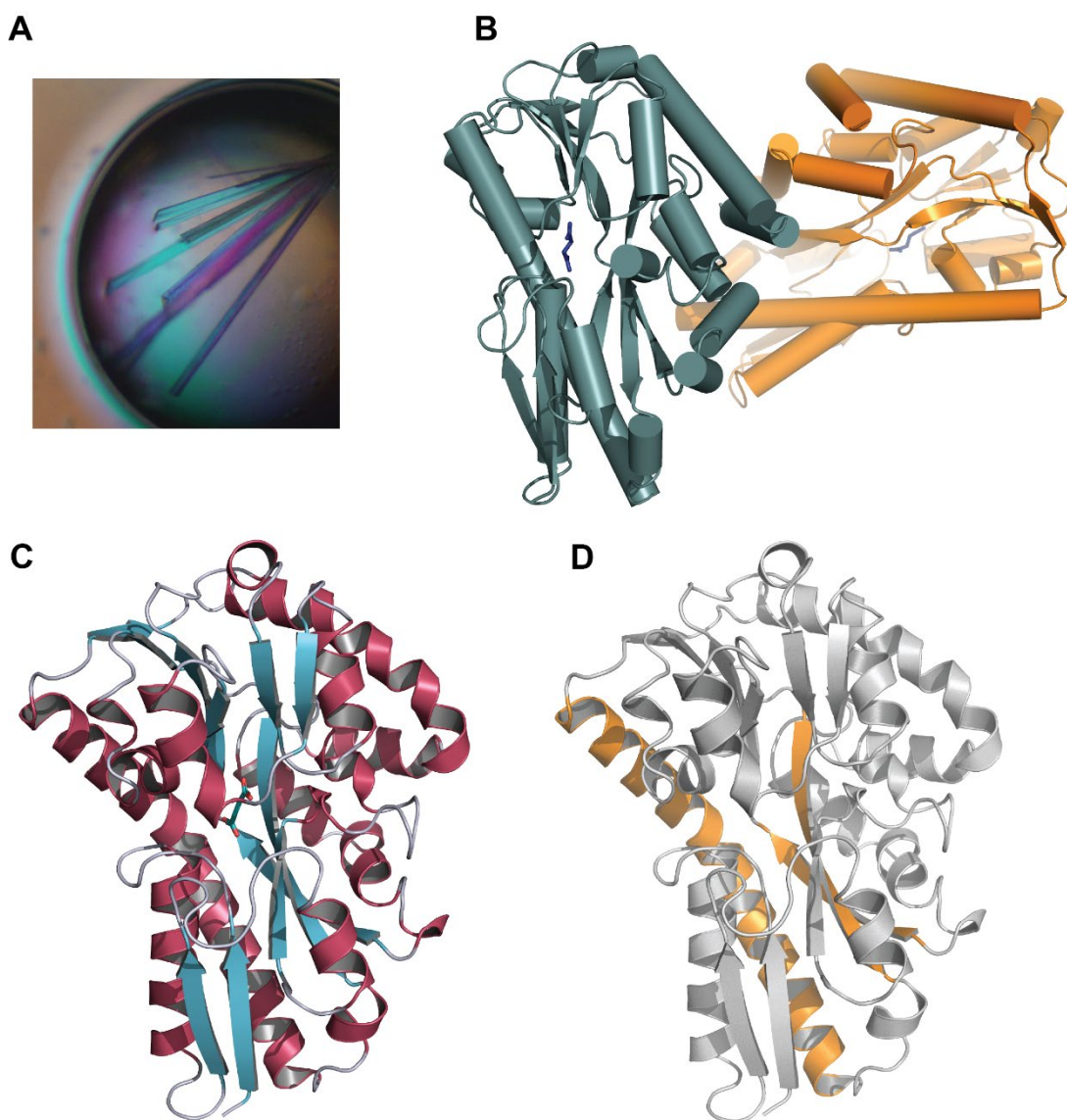


Figure 5.6: Overall structure of paTRAP-SBP. (A) paTRAP-SBP crystals bound with succinate growing as a cluster of needles. (B) Crystallographic arrangement of two asymmetric units in paTRAP-SBP. (C) The modelled structure of paTRAP-SBP showing the $\alpha\beta$ -fold characteristic of PBPs. (D) The hinge region of paTRAP-SBP highlighted in orange consisting of two β -strands and a long α -helix

The structure-model of paTRAP-SBP has two independent chains in the asymmetric unit (ASU), with each monomer containing residues 26 - 328 of protein (Figure 5.6.C). A total of 186 water molecules were also added to the model. There was a clear positive density associated with a single succinate molecule at the binding site of both protein molecules (which was not present in the MR search model). Ramachandran plot estimated 98.17% residues in the favored region and 1.83% in the allowed region. There were no residues in the Ramachandran outlier and rotamer outlier regions of the model. Both copies of the protein molecule showed high structural similarity with a root mean square deviation (RMSD) of 0.126 Å when their backbones were aligned. The final model had a R-factor of 18.22% and a R-free of 23.86%.

The succinate-bound paTRAP-SBP structure (Figure 5.6) shows the classical $\alpha\beta$ -fold architecture exhibited by all members of the PBP superfamily. ^(5,38) The bidomain architecture is similar to other TRAP binding proteins associated with C4-DA translocation, with the β -strands arranged at the center of each domain flanked by α -helices. Moreover, just like other members of the cluster E PBPs ^(39,40), TRAP-associated binding proteins, paTRAP-SBP also has a “mixed hinge” that consists of two β -strands and one α -helix that spans across both domains of the protein (Figure 5.6.B). This hinge allows the SBP to transition between open and closed conformation and aids in trapping the substrate between the cleft formed by the N- and C-terminal domains. Although an apo structure of the paTRAP-SBP has not yet been determined, other studies have noted a change in approximately 30° of hinge bending between apo and substrate-bound forms of TRAP SBPs. ⁽⁴¹⁾ Lack of a very large-scale hinge motion, as observed in some other ABC transporter associated SBPs ⁽³⁴⁾, could be attributed to the fact that the long helix stretching from N- to C-terminal domain restricts the motion of the more flexible β -strands of the hinge region.

The substrate binding occurs at the cleft formed by the two domains, as shown in Figure 5.7.A. The binding sites of TRAP SBPs are decorated with residues that can form non-bonded interactions, such as hydrogen bonds (H-bonds), van der Waals (VDW), salt bridges, and electrostatic interactions. The non-bonded contacts for paTRAP-SBP are

Table 5.3: Data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses

	paTRAP-SBP
Wavelength	1.5406
Resolution range	30.38 - 2.4 (2.49 - 2.4)
Space group	P 1
Unit cell	a = 34.80, b = 61.76, c = 70.65, $\alpha = 79.86$, $\beta = 78.88$, $\gamma = 89.99$
Total reflections	38449
Unique reflections	19895 (2076)
Multiplicity	1.9 (1.95)
Completeness (%)	89.51 (91.45)
Mean I/sigma(I)	5.6 (4.5)
Wilson B-factor	13.74
Rmerge	0.178 (0.292)
Rmeas	0.252 (0.306)
Rpim	0.178 (0.216)
CC1/2	0.843 (0.739)
Reflections used in refinement	19256 (2030)
Reflections used for R-free	1049 (102)
R-work	0.182 (0.215)
R-free	0.239 (0.277)
Number of non-hydrogen atoms	5021
macromolecules	4819

Table 5.3 continued

	paTRAP-SBP
<i>ligands</i>	16
<i>solvent</i>	186
<i>Protein residues</i>	606
<i>RMS (bonds)</i>	0.005
<i>RMS (angles)</i>	0.76
<i>Ramachandran favored (%)</i>	98.17
<i>Ramachandran allowed (%)</i>	1.83
<i>Ramachandran outliers (%)</i>	0.00
<i>Rotamer outliers (%)</i>	0.00
<i>Clashscore</i>	3.50
<i>Average B-factor</i>	21.54
<i>macromolecules</i>	21.72
<i>ligands</i>	5.86
<i>solvent</i>	18.47
<i>Number of TLS groups</i>	1

Figure 5.7: Substrate binding by paTRAP-SBP. (A) Binding site region of paTRAP-SBP highlighted by a black box and details of the interactions stabilizing the substrate succinate inside the binding pocket. (B) A two-dimensional ligplot showing the H-bond and VDW forming residues.

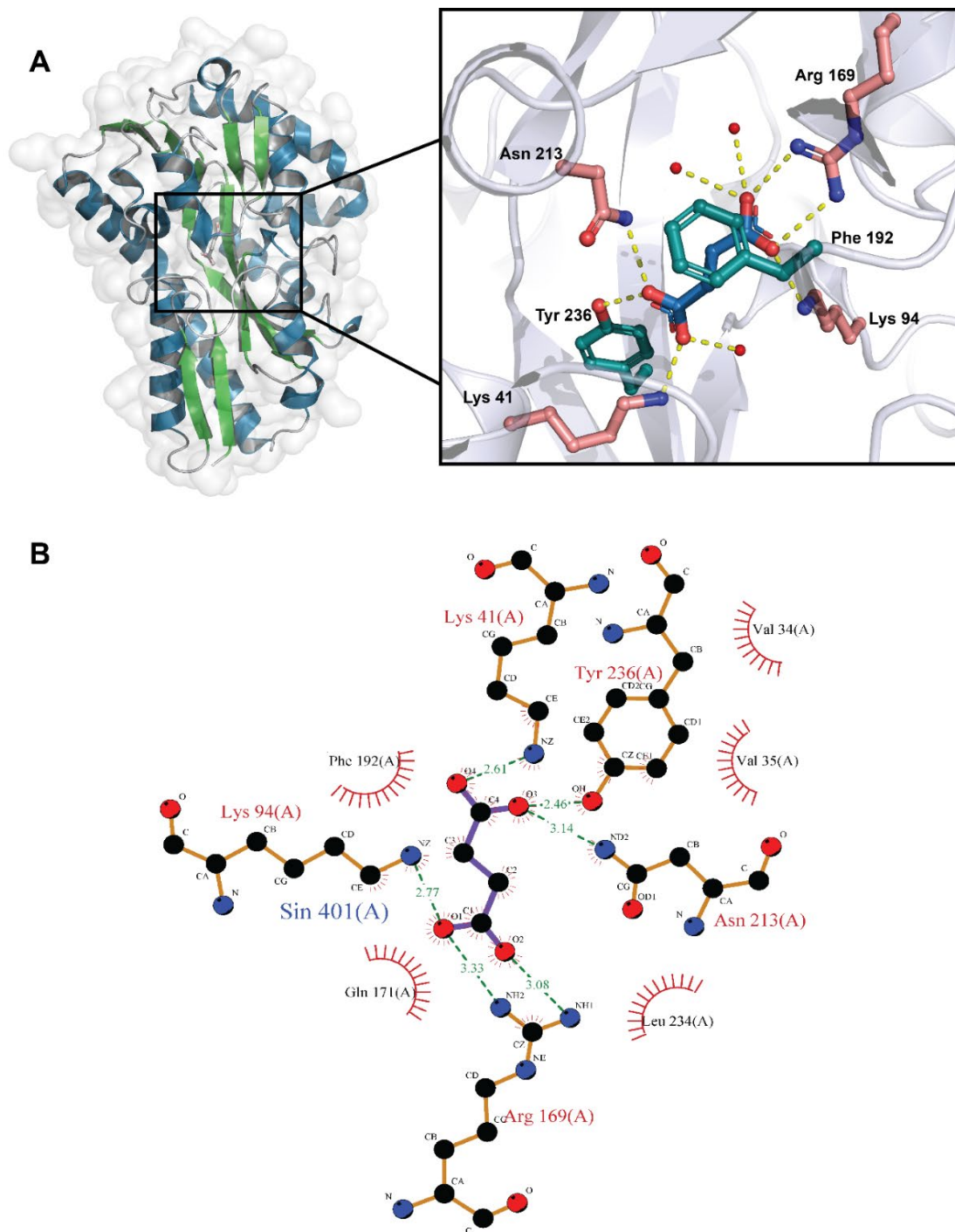


Figure 5.7: Substrate binding by paTRAP-SBP.

shown as an inset in Figure 5.7.A and were validated using MOLPROBITY. The substrate succinate is bound nearly perpendicular to the width of the binding pocket and is stabilized by H-bonds with residues lining the binding site, three coordinated water molecules, and several hydrophobic and aromatic residues forming VDW interactions. H-bonding interactions were made between the carboxylate moieties of succinate and Asn-123, Lys-41, Tyr-236, Lys-94, and Asn-213. Additionally, the conserved Arg-169 makes a salt bridge interaction with the carboxylic acid moiety of the substrate. The signature salt bridge interaction, which is universally present in all C4-DA binding TRAP SBPs and differentiates them from ABC transporter associated SBPs, is formed between a conserved arginine residue and the carboxylate moiety of the substrate. Additionally, residues Phe-192, Gln-171, Leu-234, Val-34, and Val-35 surround the substrate and form VDW interactions. A summary of these interactions is shown in 3D (Figure 5.7.A) and 2D-ligplot (Figure 5.7.B).

To summarize, the XRD structure determined in this study shows all the characteristics of a TRAP-like SBP with signature structural elements of a C4-DA binding DctP protein. The structure also sheds a light on the selective binding of TCA cycle substrates to the SBP in *P. aeruginosa*.

5.4.4 Comparative structural and functional analysis of three paTRAP-SBPs from *P. aeruginosa*

As stated earlier in section 5.4.1, this study identifies three TRAP-like SBPs in the *P. aeruginosa* PAO1 genome. PA5167 and PA0884 are associated with a TRAP *dctPQM* operon, whereas PA4616 was an orphan SBP that did not have an associated TRAP operon. The three SBP amino acid sequences are almost the same size, with PA4616 having a phenylalanine insertion at the 14th position. However, an initial sequence comparison showed remarkable levels of conservation in the binding site of these proteins (Figure 5.8.A), especially in the residues known to stabilize TCA cycle intermediates, such as succinate and malate. This sequence conservation becomes even more interesting when the overall sequence similarity between the three proteins is compared. The orphan SBP (PA4616) shows ~53% sequence conservation with the other two SBPs. On the contrary, PA5167 and PA0884 show around 70% sequence conservation.

Figure 5.8: Comparison of three TRAP-SBPs from *P. aeruginosa*.(A) Multiple sequence alignment of three TRAP-SBPs showing conserved H-bonding residues (red stars) and conserved VDW interactions (green circles), and conserved region highlighted in black. (B) A 3D structure alignment of PA4616 (PDB: 4NF0) shown in green, PA5167 (XRD data from this study) in cyan, and homology model of PA0884 in pink. (C) Alignment of binding site residues from three SBPs also showing substrates succinate (cyan ball and sticks) and malate (green ball and sticks). Residues from three SBPs are shown as sticks. Color combinations are the same as in (B).

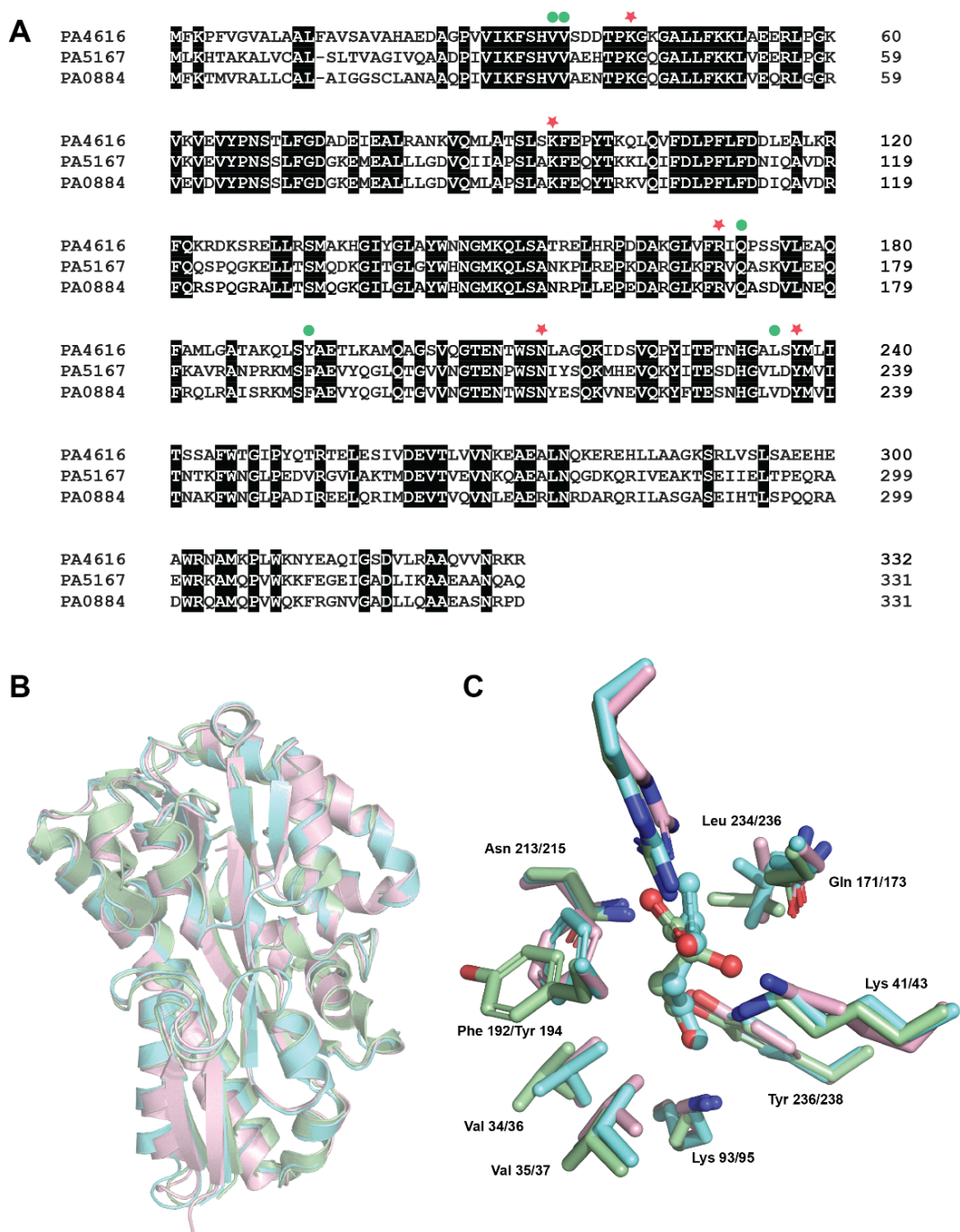


Figure 5.8: Comparison of three TRAP-SBPs from *P. aeruginosa*.

As there was a published structure of PA4616 (PDB: 4NF0) with malate bound to its binding site, it was interesting to examine the differences in its interaction profile against the succinate-bound paTRAP-SBP structure from this study. Additionally, since there was high sequence conservation between PA5167 and PA0884, a homology model of the latter was constructed to compare the possible substrate-binding modes of the third SBP. Structural alignments of the three proteins (Figure 5.8.B) showed a high degree of structural conservation with the r.m.s.d. of alignments between the PA0884 homology model and the published PA4616 crystal structure being just 0.85 Å. Similarly, the r.m.s.d of homology model vs. our paTRAP-SBP crystal structure was estimated to be 0.48 Å. Finally, the two crystal structures of PA4616 and PA5167 showed a r.m.s.d of 0.71 Å.

Moreover, a closer examination of the binding site residues in the three structures revealed that the arrangement of the binding site residues showed strong structural conservation (Figure 5.8.C). For the crystal structures binding similar TCA cycle intermediates, it was observed that similar H-bond and VDW interactions stabilized both substrates inside the binding pockets. Similarly, when the binding site residues of the third SBP was overlayed on the binding site residues of PA4616 and PA5167, it was observed that PA0884 structure also showed a similar architecture of the binding site residues as in the two crystal structures. The results of both sequence and structural comparison therefore strongly suggest that the ligand binding patterns are closely similar in all three SBPs. This may also indicate that the third novel SBP may also bind malate and succinate, consistent with the other two isoforms.

5.5 Discussion

Even though it has been known for the last two decades that *P. aeruginosa* is an opportunistic pathogen that can affect immune-compromised patients, nonetheless, mechanistic insights into its mode of infection have only recently started to be unraveled. *P. aeruginosa* has a direct role in the progression of human diseases like cystic fibrosis and chronic obstructive pulmonary disease (COPD) and recent studies have suggested that it uses succinate (which is released as a signaling molecule by the phagocytic cells of the immune system) as its primary carbon source. ⁽⁷⁾ It is also suggested that unlike other opportunistic pathogens, *P. aeruginosa* alters its metabolism inside the host by

utilizing several immune metabolites as carbon sources. ⁽⁴²⁾ Thus, establishing the mechanism of uptake of these metabolites by the bacteria may prove effective in guiding the design of new therapeutic interventions against this pathogen.

Most bacteria rely on high-affinity substrate uptake systems, like the ABC import system and the TRAP transporter systems, to speed-up the process of metabolite acquisition. There are mechanistic differences between the ABC and TRAP transporter systems but they both employ SBPs for high affinity binding of substrates in the periplasm and to accelerate substrate translocation across the cellular membrane through a membrane embedded transporter protein. Recent reports have specifically identified the TRAP transporter systems as being associated with an increased pathogenicity in some important disease-causing bacteria. ^(43,44) Moreover, opportunistic pathogen *P. aeruginosa* potentially utilizes succinate and other TCA cycle intermediate C4-DAs as carbon and energy sources and help in bacterial proliferation. ^(8,9) The central question addressed by this present study was how high-affinity binding and translocation of different TCA cycle intermediates occur in *P. aeruginosa* through the TRAP transporters.

Bioinformatics analysis of the *P. aeruginosa* PAO1 genome yielded three SBP genes, two of which (PA5167 and PA0884) were associated with additional TRAP transport proteins and one was an orphan SBP (PA4616). Structure of the orphan SBP (PA4616) has been previously characterized using XRD. We characterized the other two SBPs; PA5167 using XRD, and PA0884 by constructing a homology model of the protein. Using computational analysis, we were able to identify that PA5167 and PA0884 SBPs have a potential to translocate succinate and other TCA cycle intermediates. Functionally, PA4616 and PA5167 can bind TCA cycle intermediates malate and succinate, respectively. The homology model suggested that the third SBP (PA0884) also has conserved binding site residues and can possibly bind C4-DAs as well. This suggests about the third transporter system that can translocate succinate. Genes forming TRAP operon-I show a classical TRAP system controlled by DctB/DctD two component systems. Combining the genomic and structural information we can infer that operon-I codes for another TRAP transporter system that can bind and translocate C4-DA intermediates of the TCA cycle, similar to operon-II proteins. High sequence and structure similarity between PA5167 and PA0884

may also suggest a possible gene duplication event. The orphan SBP (PA4616), which shows lower sequence similarity with other two SBPs, may have been acquired through horizontal gene transfer by the bacteria from a closely associated species.

These assertions were also strongly supported by the biophysical DSC and SAXS analysis. It was clearly observed that malate, succinate, and fumarate substantially stabilized the apo form of the SBP, thereby suggesting a tight interaction. Additionally, SAXS analysis also determined a change in the overall R_g and the shape of the molecular envelop post substrate binding. Co-crystallization of PA5167 and PA4616 with succinate and malate, respectively, also strengthened this argument that these SBPs evolved to bind C4-DAs. Our results, combined with the findings of Valentini *et al.* ⁽⁹⁾, therefore strongly suggest that succinate, malate, and fumarate are cognate substrates of the paTRAP transporter system. Other intermediates of TCA cycle, like oxaloacetate, citrate, and α -ketoglutarate, do not get transported through this system.

To conclude, our findings suggest the possibility that more than one TRAP system is capable of translocating TCA cycle intermediates. Tracking substrate translocation through these molecular machines can provide information on metabolic efficiency of these organisms during infections. Additionally, these TRAP transporter systems could become potential targets for drugs that can alleviate *P. aeruginosa* MDR infections in immunocompromised patients suffering from cystic fibrosis and COPD.

List of References

1. Brown, E. D., and Wright, G. D. (2016) Antibacterial drug discovery in the resistance era. *Nature* **529**, 336-343
2. Murima, P., McKinney, J. D., and Pethe, K. (2014) Targeting bacterial central metabolism for drug development. *Chem Biol* **21**, 1423-1432
3. Page, M. G., and Heim, J. (2009) Prospects for the next anti-Pseudomonas drug. *Curr Opin Pharmacol* **9**, 558-565
4. Theuretzbacher, U. (2017) Global antimicrobial resistance in Gram-negative pathogens and clinical need. *Curr Opin Microbiol* **39**, 106-112
5. Mulligan, C., Fischer, M., and Thomas, G. H. (2011) Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS Microbiol Rev* **35**, 68-86
6. Rosa, L. T., Bianconi, M. E., Thomas, G. H., and Kelly, D. J. (2018) Tripartite ATP-Independent Periplasmic (TRAP) Transporters and Tripartite Tricarboxylate Transporters (TTT): From Uptake to Pathogenicity. *Front Cell Infect Microbiol* **8**, 33
7. Riquelme, S. A., Wong Fok Lung, T., and Prince, A. (2020) Pulmonary Pathogens Adapt to Immune Signaling Metabolites in the Airway. *Front Immunol* **11**, 385
8. Liu, P. (1952) Utilization of carbohydrates by Pseudomonas aeruginosa. *J Bacteriol* **64**, 773-781
9. Valentini, M., Storelli, N., and Lapouge, K. (2011) Identification of C4-Dicarboxylate Transport Systems in Pseudomonas aeruginosa PAO1. *Journal of bacteriology* **193**, 4307-4316
10. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S., and Madden, T. L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res* **36**, W5-9
11. O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., and Ako-Adjei, D. (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research* **44**, D733-D745
12. Söding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research* **33**, W244-W248

13. Webb, B., and Sali, A. (2017) Protein structure modeling with MODELLER. in *Functional Genomics*, Springer. pp 39-54
14. Vetting, M. W., Al-Obaidi, N., Zhao, S., San Francisco, B., Kim, J., Wichelecki, D. J., Bouvier, J. T., Solbiati, J. O., Vu, H., Zhang, X., Rodionov, D. A., Love, J. D., Hillerich, B. S., Seidel, R. D., Quinn, R. J., Osterman, A. L., Cronan, J. E., Jacobson, M. P., Gerlt, J. A., and Almo, S. C. (2015) Experimental strategies for functional annotation and metabolism discovery: targeted screening of solute binding proteins and unbiased panning of metabolomes. *Biochemistry* **54**, 909-931
15. Eramian, D., Shen, M. Y., Devos, D., Melo, F., Sali, A., and Marti-Renom, M. A. (2006) A composite score for predicting errors in protein structure models. *Protein Sci* **15**, 1653-1666
16. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., and Söding, J. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **7**, 539
17. Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., Kikhney, A. G., Hajizadeh, N. R., Franklin, J. M., Jeffries, C. M., and Svergun, D. I. (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* **50**, 1212-1225
18. Semenyuk, A., and Svergun, D. (1991) GNOM—a program package for small-angle scattering data processing. *Journal of applied crystallography* **24**, 537-540
19. Lienig, J., and Thulasiraman, K. (1996) GASBOR: a genetic algorithm approach for solving the switchbox routing problem. *Journal of Circuits, Systems, and Computers* **6**, 359-373
20. Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high-and low-resolution structural models. *Journal of applied crystallography* **34**, 33-41
21. Svergun, D., Barberato, C., and Koch, M. H. (1995) CRY SOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of applied crystallography* **28**, 768-773
22. Evans, P. (2014) CCP4 Tools for X-ray Integration and Data Processing. *Nihon Kessho Gakkaishi* **56**, s24-s24

23. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Echols, N., Headd, J. J., Hung, L. W., Jain, S., Kapral, G. J., Grosse Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2011) The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94-106
24. Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., and Sacchettini, J. C. (2008) Automated structure solution with the PHENIX suite. in *Structural Proteomics*, Springer. pp 419-435
25. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., and McCoy, A. J. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology* **75**, 861-877
26. Casanal, A., Lohkamp, B., and Emsley, P. (2020) Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci* **29**, 1069-1078
27. Prisant, M. G., Williams, C. J., Chen, V. B., Richardson, J. S., and Richardson, D. C. (2020) New tools in MolProbity validation: CaBLAM for CryoEM backbone, UnDownser to rethink “waters,” and NGL Viewer to recapture online 3D graphics. *Protein Science* **29**, 315-329
28. Laskowski, R. A., Jablonska, J., Pravda, L., Varekova, R. S., and Thornton, J. M. (2018) PDBsum: Structural summaries of PDB entries. *Protein Sci* **27**, 129-134
29. Mulligan, C., Leech, A. P., Kelly, D. J., and Thomas, G. H. (2012) The membrane proteins SiaQ and SiaM form an essential stoichiometric complex in the sialic acid tripartite ATP-independent periplasmic (TRAP) transporter SiaPQM (VC1777-1779) from *Vibrio cholerae*. *J Biol Chem* **287**, 3598-3608
30. Mulligan, C., Leech, A. P., Kelly, D. J., and Thomas, G. H. (2011) The membrane proteins, SiaQ and SiaM, form an essential stoichiometric complex in the sialic acid TRAP transporter SiaPQM (VC1777-1779) from *Vibrio cholerae*. *Journal of Biological Chemistry*, jbc. M111. 281030

31. Jones, C. J., Newsom, D., Kelly, B., Irie, Y., Jennings, L. K., Xu, B., Limoli, D. H., Harrison, J. J., Parsek, M. R., White, P., and Wozniak, D. J. (2014) ChIP-Seq and RNA-Seq reveal an AmrZ-mediated mechanism for cyclic di-GMP synthesis and biofilm development by *Pseudomonas aeruginosa*. *PLoS Pathog* **10**, e1003984
32. Martínez-Granero, F., Redondo-Nieto, M., Vesga, P., Martín, M., and Rivilla, R. (2014) AmrZ is a global transcriptional regulator implicated in iron uptake and environmental adaption in *P. fluorescens*F113. *BMC genomics* **15**, 237
33. Oi, V. T., Vuong, T., Hardy, R., Reidler, J., Dangl, J., Herzenberg, L., and Stryer, L. (1984) Correlation between segmental flexibility and effector function of antibodies. *Nature* **307**, 136-140
34. Shukla, S., Bafna, K., Gullett, C., Myles, D. A. A., Agarwal, P. K., and Cuneo, M. J. (2018) Differential Substrate Recognition by Maltose Binding Proteins Influenced by Structure and Dynamics. *Biochemistry* **57**, 5864-5876
35. Cuneo, M. J., Changela, A., Miklos, A. E., Beese, L. S., Krueger, J. K., and Hellinga, H. W. (2008) Structural analysis of a periplasmic binding protein in the tripartite ATP-independent transporter family reveals a tetrameric assembly that may have a role in ligand transport. *Journal of Biological Chemistry* **283**, 32812-32820
36. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132
37. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-367
38. Kelly, D. J., and Thomas, G. H. (2001) The tripartite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea. *FEMS Microbiol Rev* **25**, 405-424
39. Berntsson, R. P., Smits, S. H., Schmitt, L., Slotboom, D. J., and Poolman, B. (2010) A structural classification of substrate-binding proteins. *FEBS Lett* **584**, 2606-2617
40. Scheepers, G. H., Lycklama, A. N. J. A., and Poolman, B. (2016) An updated structural classification of substrate-binding proteins. *FEBS Lett* **590**, 4393-4401
41. Marinelli, F., Kuhlmann, S. I., Grell, E., Kunte, H. J., Ziegler, C., and Faraldo-Gomez, J. D. (2011) Evidence for an allosteric mechanism of substrate release

- from membrane-transporter accessory binding proteins. *Proc Natl Acad Sci U S A* **108**, E1285-1292
42. Riquelme, S. A., Liimatta, K., Wong Fok Lung, T., Fields, B., Ahn, D., Chen, D., Lozano, C., Saenz, Y., Uhlemann, A. C., Kahl, B. C., Britto, C. J., DiMango, E., and Prince, A. (2020) *Pseudomonas aeruginosa* Utilizes Host-Derived Itaconate to Redirect Its Metabolism to Promote Biofilm Formation. *Cell Metab* **31**, 1091-1106 e1096
 43. Chowdhury, N., Norris, J., McAlister, E., Lau, S. Y. K., Thomas, G. H., and Boyd, E. F. (2012) The VC1777-VC1779 proteins are members of a sialic acid-specific subfamily of TRAP transporters (SiaPQM) and constitute the sole route of sialic acid uptake in the human pathogen *Vibrio cholerae*. *Microbiology (Reading)* **158**, 2158-2167
 44. Severi, E., Randle, G., Kivlin, P., Whitfield, K., Young, R., Moxon, R., Kelly, D., Hood, D., and Thomas, G. H. (2005) Sialic acid transport in *Haemophilus influenzae* is essential for lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite ATP-independent periplasmic transporter. *Molecular microbiology* **58**, 1173-1185

Chapter 6 Conclusion and Future Direction

6.1 Conclusion

This study attempts to unravel the essential factors in structurally related proteins that shape their specificity and affinity towards cognate vs. non-cognate substrates. In particular, the PBP superfamily provides a unique blend of proteins that strictly adhere to their three-dimensional fold and show tremendous diversity in the type of substrate they bind to. Thus, SBPs provide an ideal system to study substrate-specificity and affinity profiles in proteins constrained by a fold space. The SBPs are predominantly associated with bacterial ABC importer systems, TRAP transporter systems, TAXI-TRAP transporter systems, and TTT systems. ^(1,2) Therefore, understanding the specifics of the structural basis for substrate recognition and binding by the SBPs can be translated into a broader comprehension of the possible ways by which proteins adapt and evolve novel function. This knowledge about protein structure-function evolution may also help to delineate the underlying factors that promote substrate specificity in structurally related proteins.

Chapter 2 highlighted how affinity and specificity were controlled among structurally related isoforms that carried out similar sugar binding functions. ⁽³⁾ *T. maritima* MBP isoforms tmMBP1, tmMBP2, and tmMBP3 share remarkable structural conservation and were shown to bind and translocate maltose and malto-oligosaccharides in association with the MalFGK₂ maltose-ABC transporter system. ^(4,5) These studies observed that the tmMBPs could bind to various malto-oligomers, albeit with varying specificity profiles. To investigate the underlying factors of this affinity-specificity conundrum in tmMBPs, both experimental and computational techniques were used. First, XRD and SAXS were used to detail the ligand-protein interactions and corresponding changes in protein structure post ligand binding. The crystallization trials of the three isoforms gave surprising results, as the trials where tmMBP1 and tmMBP2 were incubated with the disaccharide maltose yielded only crystals of the apo protein. Conversely, both tmMBP1 and tmMBP2 isoforms co-crystallized with maltotriose and maltotetraose. On the other hand, tmMBP3 yielded crystals that were either apo or maltose-bound when it was incubated with maltotriose and maltotetraose, respectively. These crystal structure results clearly suggested that the isoforms exhibit differential specificity for substrates based on their size. A closer examination of the crystal structures also revealed the presence of highly coordinated water network in tmMBP1 and tmMBP2 that serves to help stabilize larger ligands in the

shallow pockets of these isoforms. Comparison of the apo and holo forms of these isoforms also showed large difference in their hinge bending angles by $\sim 20^\circ$, suggesting possible differences in flexibility of the isoforms. This potential modulation in the dynamical behavior of tmMBPs was further studied using computational MD simulation techniques, which underlined a distinctive difference between the global and local dynamics in these proteins. Taken together, these results gave a detailed picture of how substrate specificity evolved in proteins showing high structural conservation. This study laid the groundwork for further investigation of how the structure and dynamical properties of these three isoforms effect specificity and function, using advanced neutron diffraction, scattering, spectroscopic and computational techniques.

Understanding the role of intrinsic dynamics of the tmMBP isoforms in shaping their substrate specificity profile is a very intriguing premise. To corroborate the initial dynamical profiles of tmMBPs calculated using the MD simulation, complementary neutron scattering experiments were designed to scan a similar picosecond to nanosecond timescale window in order to capture and distinguish the important dynamical events in presence and absence of the cognate substrate. These experiments and their results showing the differential dynamical signatures of tmMBP isoforms were presented and discussed in **Chapter 3**. A combination of SANS, NSE, and QENS techniques highlighted not only that these isoforms differed significantly in their intrinsic dynamics, but also that they exhibit considerably varied dynamical responses in presence and absence of their cognate substrates. The largest difference was observed in the intermediate Q-values ($Q = 0.8 \text{ \AA}^{-1} - 0.15 \text{ \AA}^{-1}$), which represents variation in their dynamics propelled by their rotational and internal dynamics. Furthermore, QENS data also bolstered the results from the MD simulation by showing that the apo-tmMBP2 to holo-tmMBP2 underwent a much larger change in signal intensity when compared to that observed for tmMBP3 apo and holo forms, suggesting that the tmMBP2 was much more flexible than tmMBP3. These results for the first time demonstrated the differential dynamical behavior in structurally similar protein isoforms that differed significantly in binding specificity using experimental neutron scattering.

As neutrons are uniquely suited for determining the position of hydrogens and water, ⁽⁶⁾ I wanted to further investigate the role of extended water-mediated H-bonded networks in

stabilizing large oligosaccharide substrates inside the binding pocket of tmMBPs. In **Chapter 4**, this principle was tested by growing large ($\sim 4 - 8 \text{ mm}^3$) crystals of tmMBP2 with cognate substrate MTT and by back-exchanging the exchangeable hydrogen atoms with deuterium. A neutron diffraction data collected on these H/D exchanged crystals resulted in a 2.1 Å dataset that was subsequently used for X-ray/neutron joint refinement. The resulting structure clearly showed the positions of essentially all hydrogen atoms in the protein, and density for deuterium atoms in the solvent water as well as on the exchangeable amide backbone and flexible side chain residues. The tmMBP2-MTT neutron structure also directly revealed how water mediated H-bond stabilize interactions with the substrate molecule, as well as coordinating H-bonds between protein side chains and substrate. The pattern and extent of H/D exchange on backbone amide positions was also used as a measure of protein flexibility and solvent accessibility in presence of a substrate. This was also the first time ever when an ABC transporter associated SBP was studied using neutron diffraction. This study demonstrated that water was indeed used by this MBP to stabilize its large substrate inside the binding pocket, a mechanism that may possibly be used by other large substrate binding PBPs

Chapter 5 also investigates specificity vs. affinity profiles in SBPs but is focused on the TCA cycle intermediate binding SBPs associated with the TRAP transporter system. As for the ABC transporters, the TRAP system also has an associated binding protein. ⁽⁷⁾ In this work, I initially identified The TRAP transporter systems in *P. aeruginosa* by mining the publicly available whole genome data from NCBI genome database. Three SBPs, two of which were associated with two different TRAP systems, were identified. The structure of one these SBPs (PA5167) was determined using XRD and a homology model was generated for PA0884. Additionally, calorimetric and SAXS data analysis provided a comprehensive picture of cognate vs. non-cognate substrates of these proteins. The overlay of these structures shows the three isoforms to be remarkably similar, with near identical binding site residue architecture, even when there was moderate sequence conservation. These results may be extremely informative in comprehending the metabolic efficiency of an opportunistic pathogen such as *P. aeruginosa*, which uses utilizes several intermediates of the TCA cycle as a carbon or energy source. ⁽⁸⁾ Findings from this study can also be applied to further investigate the druggability of the TRAP

transporter in *P. aeruginosa* and towards finding therapeutic interventions that target the C4-dicarboxylate metabolism in these pathogens.

Taken together, these studies follow the common theme. Given that the number of possible protein folds is limited ⁽⁹⁾, my research can help further our understanding on how proteins evolve complex functions in a tightly regulated structural space. The study may find application in guiding design of substrate specificity for binding novel ligands, such as designing uptake of a specific carbon source by the bacteria for bioremediation or for biofuel production. Additionally, this research could also be used towards designing novel drug candidates that target bacterial substrate uptake and metabolism.

6.2 Future Directions

The present research has made some important advances in understanding protein-ligand-solvent properties and interactions in these SBP proteins, but barely scratches the surface of understanding of how the interplay of such factors influence such interactions in these and other transporters system. There is still a large volume of work that needs to be done in this area to expand our knowledge on how protein optimize substrate interactions. Here are some of the areas where this present research can move forward to:

- Calorimetric data is required for the three tmMBP isoforms to understand the subtle variation in binding caused by the size of the interacting substrate. Subsequently, these studies would also be performed on tmMBP mutants that are designed with a significantly altered binding site to identify residues that are critical for substrate binding and specificity. Additionally, highly dynamic surface exposed regions will be targeted to understand the allosteric control of substrate binding. Indeed, there are loop regions in SBPs that are known to modulate the affinity of a protein for a substrate molecule. ⁽¹⁰⁾ These allosteric interactions help stabilize either the apo or the substrate-bound states of SBPs and can be exploited for engineering specificity in these proteins.
- Quantative methods should be developed to correlate the NSE and QENS results with the MD simulation calculations to map out the predominant dynamical modes

exhibited by tmMBP isoforms in the presence and absence of a substrate. Unwrapped MD simulation trajectories will be used to account for translational, rotational, and internal motions of the protein and protein-substrate complex. This analysis will help in precisely determining the regions on the tmMBP isoforms that differ significantly in their dynamical behavior and provide a rigorous framework for future studies.

- Crystallization trials will be carried out on both per-deuterated and hydrogenated tmMBP2 in presence of different cognate substrates, for detailed neutron crystallographic analysis of the binding interactions that are involved. Additionally, this process will be repeated for tmMBP3 isoform for a comparative estimation of the extent of influence of water in substrate stabilization.
- Finally, the TRAP transporters from *P. aeruginosa* should be studied in detail. Calorimetric measures will be performed on an expanded range of substrate to check cognate substrates. The protein will be crystallized in presence of newly identified cognate substrates to identify and delineate the common principles of substrate binding by this class of PBPs. An attempt will be made to determine the structure of the DctPQM protein using crystallography and cryo-electron microscopy (cryoEM) to identify the mode of substrate translocation by an intact TRAP-transporter protein. MD simulations will also be used to determine the intrinsic dynamics of the three *P. aeruginosa* SBP isoforms.

List of References

1. Rees, D. C., Johnson, E., and Lewinson, O. (2009) ABC transporters: the power to change. *Nature reviews Molecular cell biology* **10**, 218-227
2. Mulligan, C., Fischer, M., and Thomas, G. H. (2011) Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS microbiology reviews* **35**, 68-86
3. Shukla, S., Bafna, K., Gullett, C., Myles, D. A. A., Agarwal, P. K., and Cuneo, M. J. (2018) Differential Substrate Recognition by Maltose Binding Proteins Influenced by Structure and Dynamics. *Biochemistry* **57**, 5864-5876
4. Boucher, N., and Noll, K. M. (2011) Ligands of thermophilic ABC transporters encoded in a newly sequenced genomic region of *Thermotoga maritima* MSB8 screened by differential scanning fluorimetry. *Appl Environ Microbiol* **77**, 6395-6399
5. Wassenberg, D., Liebl, W., and Jaenicke, R. (2000) Maltose-binding protein from the hyperthermophilic bacterium *Thermotoga maritima*: stability and binding properties. *J Mol Biol* **295**, 279-288
6. Li, L., Shukla, S., Meilleur, F., Standaert, R. F., Pierce, J., Myles, D. A., and Cuneo, M. J. (2017) Neutron crystallographic studies of T4 lysozyme at cryogenic temperature. *Protein Science* **26**, 2098-2104
7. Rosa, L. T., Bianconi, M. E., Thomas, G. H., and Kelly, D. J. (2018) Tripartite ATP-independent periplasmic (TRAP) transporters and tripartite tricarboxylate transporters (TTT): from uptake to pathogenicity. *Frontiers in cellular and infection microbiology* **8**, 33
8. Valentini, M., Storelli, N., and Lapouge, K. (2011) Identification of C4-Dicarboxylate Transport Systems in *Pseudomonas aeruginosa* PAO1. *Journal of bacteriology* **193**, 4307-4316
9. Kolodny, R., Pereyaslavets, L., Samson, A. O., and Levitt, M. (2013) On the universe of protein folds. *Annual review of biophysics* **42**, 559-582
10. Rizk, S. S., Paduch, M., Heithaus, J. H., Duguid, E. M., Sandstrom, A., and Kossiakoff, A. A. (2011) Allosteric control of ligand-binding affinity using engineered conformation-specific effector proteins. *Nature structural & molecular biology* **18**, 437

Vita

Shantanu Shukla was born on July 6th, 1987 in Gaya, India, to Mr. Kaushal Kishore Shukla and Mrs. Kanchan Mishra. He grew up in a small colonial era industrial town of Jamalpur where he attended junior and middle school at the Notre Dame Academy before moving to Gaya for a high school education in 2003. He enrolled in a Bachelor of Science degree in Biotechnology at the J.J. College of Arts and Science, Bharatidasan University. Upon graduation, Shantanu pursued a Master's degree in Bioinformatics at the Pondicherry University. He completed his master's program in 2011 and worked as a Research Fellow at the Institute of Bioresources & Sustainable Development in eastern Himalayan region of India, and later at the Indian Institute of Technology (IIT) Delhi, in New Delhi India. In 2015, Shantanu joined the Graduate School of Genome Science and Technology at the University of Tennessee Knoxville in the USA in Fall 2015 and earned his Ph.D. in Life Sciences in Fall 2020.