Structural and Functional Studies of Sugar ABC Transporters in Thermophiles

A Thesis

Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

by

Monika Chandravanshi
(136106030)

Under supervision of

Dr. Shankar Prasad Kanaujia

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

Dedicated to my Family and Friends

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

Department of Biosciences and Bioengineering

Guwahati-781039

STATEMENT

I do hereby declare that the research findings of this thesis is the result of research work carried out by me in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati, India, under the supervision of **Dr. Shankar Prasad Kanaujia**.

As per the general norms of reporting research findings, due acknowledgments have been made, wherever the research findings of other researchers have been cited in this thesis.

Date: 02/03/2021 **Monika Chandravanshi**

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

Department of Biosciences and Bioengineering

Guwahati-781039

CERTIFICATE

It is certified that the work described in this thesis entitled "**Structural and Functional Studies of Sugar ABC Transporters in Thermophiles**" by **Ms. Monika Chandravanshi** for the award of degree of **Doctor of Philosophy** is an authentic record of the results obtained from the research work carried out under my supervision in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, India, and this work has not been submitted elsewhere for the award of any other degree.

CERTIFIED

Monika

Monika Chandravanshi
(Candidate)
Roll No: 136106030

Shankar Prasad Kanaujia

Dr. Shankar Prasad Kanaujia
(Thesis Supervisor)

Date: 2/3/2021

ACKNOWLEDGEMENTS

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

Foremost, I would like to thank my Ph.D. advisor, **Dr. Shankar Prasad Kanaujia**, for his guidance, advice, and tremendous support throughout my work. He was actively involved in acquiring and providing all the required facilities to carry out the work. More importantly, he engaged with me in scientific discussions to plan out the experiments and share his knowledge. His hard work and scientific curiosity motivated me to work harder. I am grateful for his time and effort in making sure that my research was carried out with good quality. I will be ever grateful for having trained me as a young researcher.

I'm also very grateful to my doctoral committee members: **Prof. Sanjukta Patra, Dr. B. Anand** and **Dr. Priyadarshi Satpati** for their valuable suggestions and healthy discussions over the past six years. I would also like to thank **Dr. Anil Mukund Limaye** for his valuable time for evaluating my second Annual Progress Seminar. I acknowledge the past and present Heads of the Department of Biosciences and Bioengineering, **Prof. V. Venkata Dasu**, **Prof. Kannan Pakshirajan, and Prof. Latha Rangan** for providing all the necessary departmental facilities during their tenure.

I acknowledge the Central instrument facility (CIF) for providing the XRD facility for data collection and ITC for thermodynamic analysis, which were instrumental to this work.

I also acknowledge the Department of Biosciences and Bioengineering (BSBE), IIT Guwahati, and the institute as a whole for providing me an opportunity and all the necessary facilities to carry out my research work. I am also thankful to the office staff member of BSBE and CIF for their technical support.

I would like to acknowledge the funding agencies, the Department of Biotechnology (DBT), Government of India for the funding, and MHRD and Department of Biotechnology (DBT), Government of India, for providing the fellowship during my Ph.D. tenure.

I sincerely thank all the members of the Structural and Computational Biology Laboratory (SCBL) for being very supportive and creating a healthy environment. My thesis would be incomplete without help from **Reshama Smanta**. I thank her for helping me in protein purification, mutagenesis and other biophysical experiments. Her sincerity and cooperation, even during the global Coronavirus pandemic led to the completion of my thesis.

I extend my gratitude to my senior fellow **Dr. Prerana Gogoi,** for helping me with difficulties in experimental work and also for taking up other lab responsibilities. Her expertise in research have helped in solving some issues with my work and I am indebted to her for sharing her experience. I thank my friend and fellow lab member, **Suraj Kumar Mandal,** for sharing his knowledge on ABC transporter and helping me to analyze the *in silico* result. I truly appreciate your constant help over the past six years. I also thank my junior, **Anghsu Dutta** for helping in the XRD facility. I also thank **Pratik Das Gupta** and **Anjaney Sharma** who generously helped me complete the *in silico* analysis. I would like to say a warm thank you to my beloved juniors **Sayan Saha** and **Prerana Mordina** for their help. Special praises to **Shreya LB** for engaging with me in discussions, which provoked me to think in a new direction.

I will always be grateful to the faculties who taught me the fundamental concepts in my undergraduate and postgraduate studies including **Prof. B.N. Tiwary, Prof. Renu Bhat, Dr. Harit Jha, Dr. Pradeep Verma, Dr. D.K. Parihar, Dr. Alka Ekka, Dr. Sudheer Pamidimarri** and **Mrs. Tithi Saha Yadav**. They played an important role in establishing my career as a researcher.

Good friends motivate and help us navigate towards our goal. I am very fortunate to have great friends and peers and would love to express gratitude to them. **Mrs. Pooja Mishra, Dr. Jayant Dewangan** and **Dr. Kedar Sharma** have supported me throughout my career. **Madhuri Patel**, **Komal Choukate**, **Dr. Debika Dutta**, **Dr. Gaurav Pandey**, **Dr. Reshmi Das** and **Amrita Khwairakpam** were always there when I needed them. Lastly, many thanks to my hostel friends **Shruti, Ruchika, Trishana** and **Peeyushi** for sharing all the happy and sad moments in the IITG campus.

Finally, I express my endless gratitude to my parents for their unconditional love, encouragement, patience and endurance. Their faith in me and sacrifices have made this

journey possible. Special thanks to my younger brother and sister for taking care our family, while I pursued my Ph.D and for boosting my morale.

Monika Chandravanshi

ABSTRACT

The uptake of nutrients, including carbohydrates, metal ions, amino acids, and peptides, are required for many biological processes. Bacterial cells scavenge these essential nutrients from microenvironments for their survival. It utilizes a myriad of mechanisms to acquire these essential nutrients from the extracellular environment. One of the mechanism involves the use of several transport proteins, such as ATP-binding cassette (ABC) transporters to import and export the substrates. ABC transporters, conserved across all organisms, are powered by the energy from ATP to move substrates across cellular membranes. Topologically, ABC transporters consist of two transmembrane domains (TMDs) for transport pore and two nucleotide-binding domains (NBDs) for ATP hydrolysis to generate the energy during substrate translocation across the plasma membrane. In addition, the prokaryotic ABC-importer system possesses a functional and structural unit known as substrate (or solute)-binding protein (SBP) for nutrient acquisition inside the cell. In the periplasm, SBPs bind to the specific ligand with high affinity and mediate their transport into the cytoplasm via the cognate inner membrane component, i.e., TMD. As ABC importers perform multiple functions required for cell physiology, it transports various kinds of nutrients inside the cells via SBPs. On the ground of substrate types being transported, SBPs are named as sugar-, amino acid-, sulfate-, ferric iron-, metal ion-, siderophore- and oligopeptide-binding proteins. During substrate binding, SBPs changes its conformation to accommodate the ligand in the active site via an induced-fit mechanism known as the “Venus Fly-trap” mechanism. A range of structural as well as other biophysical methods, have been suggested to elucidate the mechanism of SBP-ligand interactions. However, due to the low sequence identity of SBPs, understanding the structural basis of substrate recognition by SBPs has remained very challenging. In addition to SBP, the structural diversity of ligands makes it more complex to understand the mechanistic basis of ligand recognition via SBPs. One of the essential nutrient molecules for bacterial cell survival is carbohydrate molecules, which exhibit structural diversity as a result of microbial population environmental diversity. Till date, several SBPs bound to carbohydrate have been characterized, however little is known about the selection criteria for carbohydrate diversity. Thus, this study proposes the structural and functional elucidation of the SBP subunit of ABC transporter for mechanistic elucidation.

ABSTRACT

In this study, we have performed data mining for ORFs encoding sugar ABC transporters from the *Thermus thermophilus* HB8 genome. This enabled the collection of 11 ABC transporters for carbohydrates, indicating the importance of ABC transporters in sequestering carbohydrates.

To identify the cognate ligand for each ABC transport system, *in silico* functional characterization of SBP of ABC transport system followed by further verification via metabolic pathway profiling has been executed. With the help of an integrated approach including sequence, structure, genomic, and metabolism out of 11 ABC transporters, six were characterized for their cognate sugars and identified to be involved in the uptake of diverse carbohydrates molecules essential for bacterium survival in the extreme environment. The remaining five were characterized as Ugp and purine ABC transporters. Moreover, from the metabolic pathway profiling, five pathways were found to be functionally associated with these transporters. In addition, SBPs with their probable ligands were characterized for maintaining the multiple specificities of these transporters. Also, numerous improvements regarding gene annotations also have been proposed, which includes the identification of three misannotated sugar ABC transport systems. These misannotated ABC transporters were identified as candidate transporters for phospholipid precursors and purine.

After identification of the total carbohydrate ABC transport system in *T. thermophilus* HB8, three of the ORFs TTHA0356, TTHB082, and TTHA0379 were further characterized using the crystallographic approach. To understand the mechanistic basis of ligand selection of SBPs for diverse carbohydrates, these three selected SBPs were subjected to structural and thermodynamic analysis. In this study, the crystal structure of protein TTHA0356 was solved in complex with trehalose (α-1,1), sucrose (α-1,2), maltose (α-1,4), palatinose (α-1,6) and glucose with resolution in a range 1.6 to 2.0 Å. From the structural analysis, protein TTHA0356 has been identified to be specific for various α-glycosides and thus named as an α-glycosides-binding protein (αGlyBP). Though αGlyBP possesses a multispecificity, it

maintains its stereoselectivity for both glycosidic linkage and an epimeric hydroxyl group. Surprisingly, both thermodynamic and structure data confirm that glucose is competitive sugar for α-glycosides and thus exhibits the paradoxical behavior, where it displaces the higher affinity ligand from the active site. Moreover, a combination of mutagenesis and structural data involving comparative assessment of open and closed conformations of αGlyBP reveals the hinge region is the first interaction site for ligand binding, while encapsulation of ligand inside the active site is achieved through N-terminal domain (NTD) movement. The C-terminal domain (CTD) of αGlyBP was identified to be rigid and postulated for maintaining the interaction with the transmembrane domain (TMD) during substrate translocation.

Next, we did the structure-based functional characterization of protein TTHB082. Both structural and thermodynamics methods delineate that the protein TTHB082 is a β-glycosides-binding protein (βGlyBP, ORF: TTHB082) and akin to αGlyBP it maintains the stereoselectivity for the β-glycosides. Thermodynamic analysis reveals that βGlyBP is multi-specific and binds to different types of β-glycosides varying in glycosidic linkages (β-1,2; β-1,3; β-1,4 and β-1,6). Structurally, βGlyBP shows four consecutive subdomain(s) (N1, N2, C1 and C2) organizations. Surprisingly, unlike the αGlyBP, for ligand encapsulation βGlyBP adopts the modified "Venus Fly-trap" mechanism, where the conformational dynamics of N1 and C1 subdomains drives the ligand-binding rather than NTD and CTD. In addition, comparative active site profiling of αGlyBP and βGlyBP reveals the stereo selection mechanism for carbohydrates in the different SBPs, where α-and β-glycosides occupies a similar position but binds in an opposite orientation. Altogether, the structure of αGlyBP and βGlyBP demonstrate that information for ligand selection and binding mechanism is already pre-encoded in the SBPs and thus varied with protein and carbohydrates.

Among the 11 carbohydrate ABC transport system, protein TTHA0379 was found to be misannotated as a sugar-binding protein and identified as protein UgpB for the transport phospholipid precursors such as glycerophosphocholine (GPC). However, the structural analysis demonstrated that the protein TTHA0379 is a dinucleotide-binding protein rather

than sugar and/or GPC-binding protein. Structurally the protein TTHA0379 is endogenously bound to a novel ligand, which is a derivative of uridylyl-3'-5'-phospho-guanosine (U3G). Thus, owing to endogenous binding of dinucleotide in the active site, protein TTHA0379 is characterized as a dinucleotide-binding protein and re-annotated as a U3G-binding protein (U3GBP). This study reports the first structure of SBP in complex with dinucleotide, and thus U3GBP is classified as a first member of subcluster D-I SBPs in the SBP classification system. In addition to the uptake mechanism of dinucleotide via the ABC transport system, this study also suggests the downstream function of transported dinucleotide U3G as a secondary messenger for several cytosolic proteins involved in tRNA synthesis and/or modification.

In summary, the finding of this study reveals that *T. thermophilus* HB8 possesses two SBPs (ORF IDs: TTHA0356 and TTHB082) of ABC transporters, which selectively transports the α-and β-glycosides, respectively. Depth characterization of protein αGlyBP (ORF ID: TTHA0356) suggests that it maintains the stereoselectivity for linkages and does not allow the passage of β-glycosides. Similarly, protein βGlyBP (ORF ID: TTHB082) maintained the specificity only for β-glycosides and occluded the transport of α-glycosides. In contrast, the characterization of third SBP (ORF ID: TTHA0379) reveals that protein U3GBP (ORF ID: TTHA0379) is a dinucleotide-binding protein rather than sugar-/GPC-binding protein. Altogether, this study provides a basic understanding of the selective nature of ABC transporter over the versatility of carbohydrates.

CONTENTS

CONTENTS

CONTENTS

CONTENTS

CONTENTS

CONTENTS

CONTENTS

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

LIST OF PDBs

α-glycoside-binding protein (αGlyBP)

S.No.	Title	PDB ID
1	Crystal structure of ABC transporter α-glycoside-binding protein in complex with trehalose	6J9W
2	Crystal structure of ABC transporter α-glycoside-binding protein in complex with maltose	6J9Y
3	Crystal structure of ABC transporter α-glycoside-binding protein in complex with palatinose	6JAD
4	Crystal structure of ABC transporter α-glycoside-binding protein in complex with sucrose	6JAG
5	Crystal structure of ABC transporter α-glycoside-binding protein in complex with glucose	6JAH
6	Crystal structure of ABC transporter α-glycoside-binding mutant protein R356A in ligand free form	6JAL
7	Crystal structure of ABC transporter α-glycoside-binding mutant protein R356A in complex with trehalose	6JAM
8	Crystal structure of ABC transporter α-glycoside-binding mutant protein R356A in complex with maltose	6JAN
9	Crystal structure of ABC transporter α-glycoside-binding mutant protein R356A in complex with palatinose	6JAO
10	Crystal structure of ABC transporter α-glycoside-binding mutant protein R356A in complex with sucrose	6JAP
11	Crystal structure of ABC transporter α-glycoside-binding mutant protein R356A in complex with glucose	6JAQ
12	Crystal structure of ABC transporter α-glycoside-binding mutant protein R49A in complex with maltose	6JAR
13	Crystal structure of ABC transporter α-glycoside-binding mutant protein D118A in complex with maltose	6JAI

LIST OF PDBs

14	Crystal structure of ABC transporter α-glycoside-binding mutant protein W287F in complex with trehalose	6JAZ
15	Crystal structure of ABC transporter α-glycoside-binding mutant protein W287A in complex with trehalose	6JB0
16	Crystal structure of ABC transporter α-glycoside-binding mutant protein W287A in complex with maltose	6JB4
17	Crystal structure of ABC transporter α-glycoside-binding mutant protein W287A in complex with palatinose	6JBA
18	Crystal structure of ABC transporter α-glycoside-binding mutant protein W287A in complex with sucrose	6JBB
19	Crystal structure of ABC transporter α-glycoside-binding mutant protein W287A in complex with glucose	6JBE

β-glycoside-binding protein (βGlyBP)

20	Crystal structure of β-glycosides-binding protein of ABC transporter in open state, Form I	7C63
21	Crystal structure of β-glycosides-binding protein of ABC transporter in open state, Form II	7C64
22	Crystal structure of β-glycosides-binding protein of ABC transporter in closed state with cellobiose	7C66
23	Crystal structure of β-glycosides-binding protein of ABC transporter in closed state with cellotriose	7C67
24	Crystal structure of β-glycosides-binding protein of ABC transporter in closed state with cellotetraose	7C68
25	Crystal structure of β-glycosides-binding protein of ABC transporter in closed state with sophorose	7C69
26	Crystal structure of mutant W177X β-glycosides-binding protein of ABC transporter in open state	7C6F

LIST OF PDBs

27	Crystal structure of mutant W177X β-glycosides-binding protein of ABC transporter in open-liganded state with gentiobiose	7C6G
28	Crystal structure of mutant W177X β-glycosides-binding protein of ABC transporter in open-liganded state with laminaribiose	7C6H
29	Crystal structure of mutant W177X β-glycosides-binding protein of ABC transporter in open-liganded state with sophorose	7C6I
30	Crystal structure of mutant W177X β-glycosides-binding protein of ABC transporter in closed state with cellobiose	7C6J
31	Crystal structure of mutant W177X β-glycosides-binding protein of ABC transporter in closed state with cellotriose, Form I	7C6K
32	Crystal structure of mutant W177X β-glycosides-binding protein of ABC transporter in closed state with cellotriose, Form II	7C6L
33	Crystal structure of mutant W177X β-glycosides-binding protein of ABC transporter in closed state with cellotetraose, Form I	7C6M
34	Crystal structure of mutant W177X β-glycosides-binding protein of ABC transporter in closed state with cellotetraose, Form II	7C6N
35	Crystal structure of mutant W177X β-glycosides-binding protein of ABC transporter in closed state with cellopentaose	7C6R
36	Crystal structure of mutant W177X β-glycosides-binding protein of ABC transporter in closed state with laminaritriose, Form I	7C6T
37	Crystal structure of mutant W177X β-glycosides-binding protein of ABC transporter in closed state with laminaritriose, Form II	7C6V
38	Crystal structure of mutant W177X β-glycosides-binding protein of ABC transporter in closed state with laminaritetraose	7C6W
39	Crystal structure of mutant W41A β-glycosides-binding protein of ABC transporter in open state, Form I	7C6X
40	Crystal structure of mutant W41A β-glycosides-binding protein of ABC transporter in open state, Form II	7C6Y
41	Crystal structure of mutant W67A β-glycosides-binding protein of ABC transporter in open state	7C6Z

42	Crystal structure of mutant W67A β-glycosides-binding protein of ABC transporter in open-liganded state with gentiobiose	7C70
43	Crystal structure of mutant E117A β-glycosides-binding protein of ABC transporter in open state	7C71

U3G-binding protein (U3GBP)

44	Crystal structure of a dinucleotide-binding protein of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine (Form I)	7C0F
45	Crystal structure of a dinucleotide-binding protein of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine (Form II)	7C0K
46	Crystal structure of a dinucleotide-binding protein of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine (Form III)	7C0L
47	Crystal structure of a dinucleotide-binding protein (Y56F) of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine	7C0O
48	Crystal structure of a dinucleotide-binding protein (F79A) of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine (Form I)	7C0R
49	Crystal structure of a dinucleotide-binding protein (F79A) of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine (Form II)	7C0S
50	Crystal structure of a dinucleotide-binding protein (N81A) of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine	7C0T
51	Crystal structure of a dinucleotide-binding protein (S127A) of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine	7C0U
52	Crystal structure of a dinucleotide-binding protein (Y224A) of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine (Form I)	7C0V

LIST OF PDBs

53	Crystal structure of a dinucleotide-binding protein (Y224A) of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine (Form II)	7C0W
54	Crystal structure of a dinucleotide-binding protein (T240A) of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine	7C0X
55	Crystal structure of a dinucleotide-binding protein (Y246A) of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine (Form I)	7C0Y
56	Crystal structure of a dinucleotide-binding protein (Y246A) of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine (Form II)	7C0Z
57	Crystal structure of a dinucleotide-binding protein (Q274A) of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine (Form I)	7C14
58	Crystal structure of a dinucleotide-binding protein (Q274A) of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine (Form II)	7C15
59	Crystal structure of a dinucleotide-binding protein (Y224A/Y246A) of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine (Form I)	7C16
60	Crystal structure of a dinucleotide-binding protein (Y224A/Y246A) of ABC transporter endogenously bound to uridylyl-3'-5'-phospho-guanosine (Form II)	7C19
61	Crystal structure of a dinucleotide-binding protein (F79A/Y224A/Y246A and deletion of residues 50-75) of ABC transporter (unbound form)	7C1B

ABBREVIATIONS AND NOTATIONS

ABBREVIATIONS

ABC	ATP-binding cassette
BGL	β-glucosidase
BXY	Arabinose
CBM	Carbohydrate-binding module
CD	Circular dichroism
c-di-AMP	Cyclic diadenylate
c-di-GMP	Cyclic diguanylate
CEL2	Cellobiose
CEL3	Cellotriose
CEL4	Cellotetraose
CEL5	Cellopentaose
CTD	C-terminal domain
CUT	Carbohydrate uptake transporter
CUT1	Carbohydrate uptake transporter-1
CV	Column-volume
DP	Degree of polymerization
G1P	Glucose-1-phosphate
G3P	Glycerol-3-phosphate
GAL	Galactose
GCU	Glucoronic acid
GEN2	Gentiobiose

ABBREVIATIONS AND NOTATIONS

GH	Glycosyl hydrolase
GLC	Glucose
GPC	Glycerophosphocholine
GT	Glycosyltransferase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LAM2	Laminaribiose
LAM3	Laminaritriose
LAM4	Laminaritetraose
LAMn	Laminarin
LAT	Lactose
LB	Luria-Bertani
MAL	Maltose
MAN	Mannose
MFS	Major facilitator superfamily
MLB	Melibiose
MMS	Matrix micro-seeding
MSA	Multiple sequence alignment
N.D.	Not detected
NBD	Nucleotide-binding domain
NCBI	National Center for Biotechnology Information
NTD	N-terminal domain
ORF	Open reading frame

ABBREVIATIONS AND NOTATIONS

PAL	Palatinose
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PGM	Phosphoglucomutase
PMSF	Phenylmethylsulfonyl fluoride
RAF	Raffinose
RMSD	Root mean square deviation
SBP	Substrate (or solute)-binding protein
SOP2	Sophorose
SUC	Sucrose
t6A	N6-threonylcarbamoyladenosine
TAG	Tagatose
ThuA	Trehalose utilization protein
TMD	Transmembrane domain
TRE	Trehalose
U3G	Uridylyl-3'-5'-phospho-guanosine
U3GBP	U3G-binding protein
Ugp	Uptake glycerol phosphate
WT	Wild-type
αGlyBP	α-glycoside-binding protein
βGlyBP	β-glycosides-binding protein
β-ME	β-mercaptoethanol

ABBREVIATIONS AND NOTATIONS

NOTATIONS

~	Approximately
%	Percent
Å	Angstrom
a.u.	Arbitrary unit
bp	Base pair
°C	Degree Celsius
g	Relative Centrifugal Force (RCF)
Ka	Association constant
kb	kilobase
kcal	Kilocalories
Kd	Dissociation constant
kDa	kilodalton
L	Litre
M	Molar
mdeg	Change in ellipticity
mM	Millimolar
ml	Millilitre
n	Binding stoichiometry
s	Second
T	Temperature
Tm	Melting temperature

ABBREVIATIONS AND NOTATIONS

ΔG	Change in free energy
ΔH	Change in enthalpy
ΔS	Change in entropy
λ	Wavelength
µg	Microgram
µl	Microlitre
µM	Micromolar

CHAPTER 1 – INTRODUCTION

1.1 INTRODUCTION

Carbohydrates (or sugars) are widely distributed in nature in almost all kingdoms of life i.e., eukarya (animalia, fungi, crustaceans and plants), bacteria and archaea (Elbein et al., 2003). It is involved in numerous biological functions such as stress-protection, energy supply and signal transduction (Iturriaga et al., 2009). Furthermore, in thermophiles, sugars play a crucial role as an osmoprotectant during severe salt stress and non-inhabitant conditions (Alarico et al., 2007). Moreover, sugars have also been implicated in the regulation of gene expression and in host-pathogen interactions (Iturriaga et al., 2009). For all this, uptake of sugars inside the cell from the extracellular environment is required, which is attained by proteins known as transporters embedded mostly in membranes. The membrane transporters for sugars have been classified primarily into two superfamilies known as (1) ATP-Binding Cassette (ABC) superfamily and (2) Major Facilitator Superfamily (MFS) (Higgins, 1992; Fath and Kolter, 1993; Dean and Allikmets, 1995; Kuan et al., 1995) with Transporter Classification Identifier (TCID) 2.A.1 and 3.A.1, respectively (Saier, 2000a, b, c). ABC transporters are multicomponent primary active-transporter, which transport the solutes by utilizing the ATP-driven energy source (Wilkens, 2015). On the other hand, MFS transporters use chemiosmotic ion-gradient-driven energy for solute transport (Xie, 2008). Despite the recent progress in molecular and biochemical analyses of these transporters, structural information is essential to understand the mechanism of substrate transport across the membranes.

Over the last few decades, crystal structures of different members of ABC and MFS transporter superfamilies have been reported in the literature. Among these, the family of sugar transporters such as maltose ABC transporter and lactose permease (LacY) is the most extensively analyzed (Abramson et al., 2003; Chen et al., 2003; Lu et al., 2005). Crystal structures of these proteins are used as a prototype to understand the mechanistic insights of transport cycle of respective superfamilies. In this study, we disclose the structural and mechanistic studies of sugar transporters belonging to ABC transporter superfamilies.

1.1.1 Bacterial transporters

Transporters are a vital component of all forms of life. It is used by cells to carry out selective export and/or import of essential nutrients such as amino acids, carbohydrates, lipids, inorganic ions, the end products of metabolism and deleterious substances across the membranes. They are also required to maintain homeostasis between inside and outside of cells (Mitchell, 1967). Based on functional aspects, transporters have broadly been classified as porins, channels, porters (primary and secondary active transporters), group translocators and transmembrane electron carriers (Davidson et al., 2008). These transporters have also been classified into ~60 protein superfamilies such as ABC, MFS and voltage-gated ion channel (VIC). These superfamilies have further been sub-classified into ~750 protein families (Saier et al., 2016). Out of which, ~33 families are known to be involved in the uptake of carbohydrates (Walmsley et al., 1998; Saier, 2000a). Sugars being hydrophilic molecules cannot diffuse through the bacterial outer and inner membrane (OM and IM); thus, specific and selective transporters are required for their import inside the cell. However, uptake of sugars inside the cell through the bacterial cell membrane occurs in concatenates of several steps (Wang et al., 1997). A general schematic picture of the import mechanism of carbohydrates has been depicted in Figure 1.1. Initially, carbohydrate (or sugar) molecules such as malto-oligosaccharides, sucrose, raffinose and glucose diffuse through porins (e.g LamB, ScrY, RafY and OprB, respectively), which are located in the OM of prokaryotes (Saier, 2000a). Porins exclusively comprise of β-strands forming β-barrel structures for the passage of carbohydrate molecules. Once these molecules cross the OM and enter the periplasmic space, it is sequestered by proteins termed as **p**eriplasmic-**b**inding **p**roteins (PBPs, also known as **s**olute- or **s**ubstrate-**b**inding **p**roteins, SBPs) of ABC transporter (Wang et al., 1997). Subsequently, carbohydrate molecules are taken up inside the cell through IM by a primary active (e.g MalFGK) and secondary active (e.g ScrA, RafB and Gtr) transporters (Nikaido and Saier, 1992). Once carbohydrate molecules reached the cytoplasm, carbohydrate specific enzymes such as sucrose hydrolase (ScrB), fructose kinase (ScrK) and α-galactosidase (RafA) facilitate their dissimilation for the subsequent metabolism (Figure 1.1).

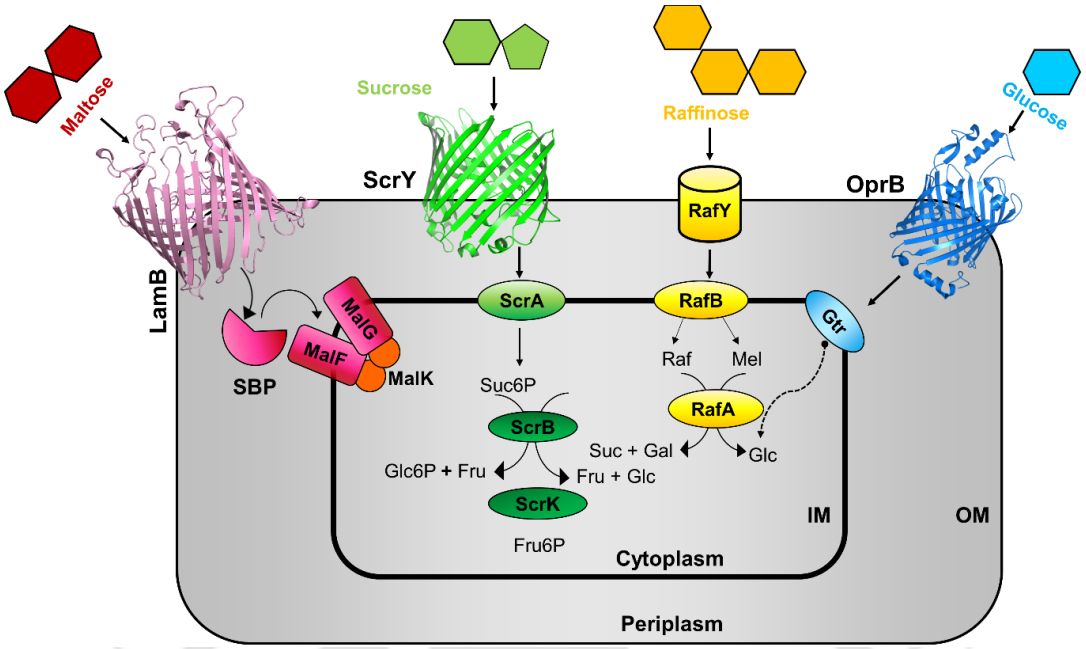


Figure 1.1. Bacterial transport pathways for sugars. The outer membrane (OM) contains porins for sugars such as malto-oligosaccharides (LamB, magenta), sucrose (ScrY, green), raffinose (RafY, yellow) and glucose (OprB, blue). In periplasm, substrate-binding proteins (SBPs) capture sugars. Subsequently, malto-oligosaccharides, sucrose, raffinose and glucose are transported to the cytoplasm by MalFGK (magenta), ScrA (green), RafB (yellow) and Gtr (blue) transporters, respectively. In the cytoplasm, these sugars are metabolized for downstream processes by proteins such as ScrB (green), ScrK (green) and RafA (yellow). Abbreviation: Fru, fructose; Gal, galactose; Glc, glucose; Mel, melibiose; Suc, sucrose.

1.1.2 Transport of carbohydrate (or sugar) across the outer membrane of bacterial cells

Molecules to be transported inside the cytoplasm of a bacterial cell firstly need to pass through the outer membrane (OM) and depending on the nature of the molecule, different pathways are utilized. Hydrophobic molecules mostly diffuse through the OM, whereas hydrophilic molecules usually pass through porins found in the OM of Gram-negative and some mycolic acid-containing bacteria such as *Mycobacteria* (Faller et al., 2004; Shultis et al., 2006). Porins are mostly trimeric in nature with 18 (or 16) β-strands organized into β-barrel closed structures (Schirmer, 1998). It contains loops projected towards the active site cavity forming a constriction contributing to the opening and closing of the channels (Davidson et al., 2008). This constricted zone

defines the selectivity and specificity of solutes that can be transported through these channels (Koebnik et al., 2000). Molecules having a molecular mass < 650 Da can passively diffuse across the porins, whereas, others are transported through specific porins (Nikaido, 1994). Sugar polymers with two (e.g. maltose and sucrose) or more units (e.g. maltodextrin) are transported by selective porins such as maltoporin (or LamB porin) (Szmelcman and Hofnung, 1975). So far, many porins specific for sugars have been identified and classified with TCID 1.B (Saier, 2000a). However, as of now, only a few sugar porins families have been biophysically analyzed (Table 1.1).

Table 1.1. Families of porins for carbohydrate uptake.

Family	Protein name	*TCID	Substrate	Organism	PDB ID
Major intrinsic protein (MIP) family	Glycerol facilitator	1.A.8.1.1	Glycerol, polyols	*Escherichia coli*	1FX8
Sugar porin (SP) family	Maltoporin, LamB (MalL)	1.B.3.1	Maltose, maltoheptaose	*E.coli*	1MPM
	Oligosaccharide porin, ScrY	1.B.3.2	Sucrose	*Staphylococcus typhimurium*	1A0T
	Porin with specificity for β-glucosides, BglH	1.B.3.3	Arbutin, salicin, gentibiose	*E.coli*	-
Raffinose porin (RafY) family	Raffinose porin, RafY	1.B.15.1	Raffinose	*E.coli*	-
Glucose-selective OprB porin family	Glucose-inducible glucose transporting porin, OprB	1.B.19.1	Glycerol, mannitol, fructose, maltose, pentose	*Pseudomonas aeruginosa*	-
	Quorum sensing acyl homoserin lactone porin, OprB	1.B.19.2	Acyl homoserine lactone	*Burkholderia pseudomallei*	-

	Carbohydrate selective porin, OprB	1.B.19.3	Monosachharides	*Pseudomonas putida*	4GF4
Cyclodextrin porin (CDP) family	Cyclodextrin /linear malto-oligosachharide porin, CymA	1.B.26	Cyclodextrin / linear maltodextrin	*Klebsiella oxytoca*	-

* Information of TCID was retrieved from Transporter Classification Database (TCDB, Saier et al., 2016).

1.1.3 Transport of carbohydrate (or sugar) across the inner membrane of bacterial cells

Carbohydrate molecules, once in the periplasm, cannot cross the IM due to its hydrophilic nature, thus specific transporters are required to channelize them into the cytoplasm. The IM sugar transporters are broadly classified into (1) ABC (2) MFS and (3) phosphotransferase system (PTS) superfamilies (Table 1.2). ABC transporters catalyze the transport of carbohydrate molecules against the concentration gradient using the ATP-hydrolyzed energy (Wilkens, 2015). On the other hand, MFS transporters use the chemiosmotic-ion-gradient energy to catalyze the facilitated diffusion of sugars (Xie, 2008). In addition, the group translocators such as phosphoenolpyruvate: sugar phosphotransferase (PTS) transporters, unique to prokaryotes, catalyze the uptake of those sugars, which are further phosphorylated in the cytoplasm (Saier, 2000a,b) (Figure 1.2). ABC superfamily of proteins, the largest of all, is further subdivided into **c**arbohydrate **u**ptake **t**ransporters 1 and 2 (CUT1 and CUT2) families. Similarly, the second largest sugar transport superfamily MFS is classified into uniporters, antiporters and symporters. The third superfamily of sugar transporters is Phosphoryl transfer-driven group translocators (PTS), which is found exclusively in bacteria and is specific only for hexoses (Table 1.2). After sugar transportation, PTS subsequently phosphorylates the transported sugar molecules. In contrast, Sugars transported through ABC and MFS proteins remain unaltered.

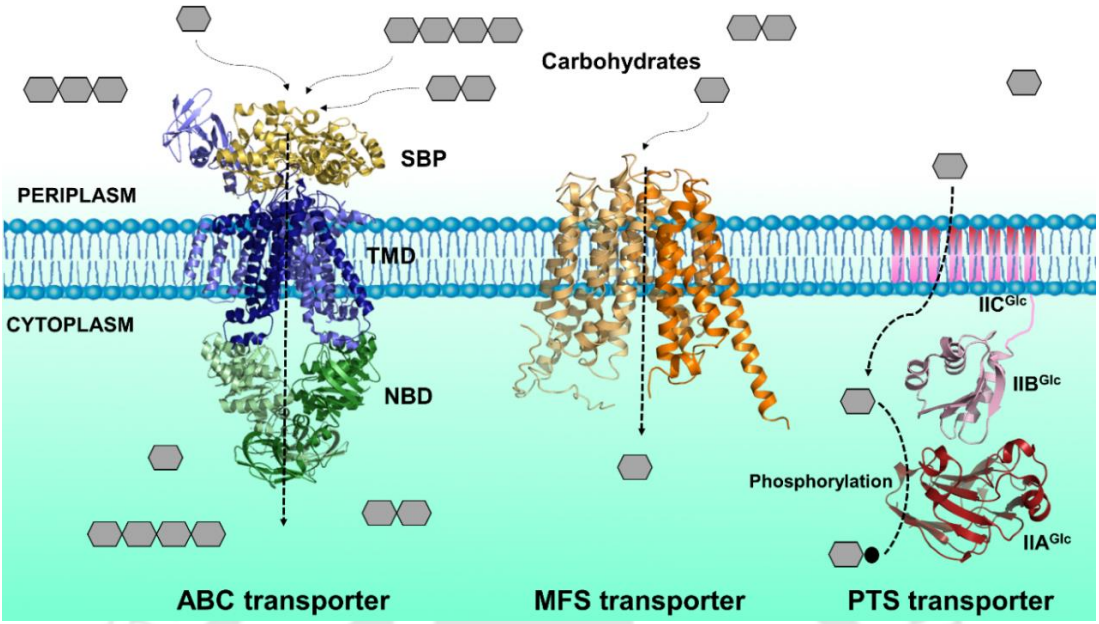


Figure 1.2. Inner membrane transporter for carbohydrate uptake. Three major superfamilies namely ABC transporter (e.g. maltose ABC transporter, PDB ID: 4KHZ), MFS transporter (e.g. GlpT, PDB ID: 1PW4) and PTS transporter (e.g Glucose IICB-IIA, PDB ID: 1O2F) for carbohydrate uptake are shown in blue, orange and magenta color, respectively.

Table 1.2. Inner membrane transporter families for carbohydrate uptake.

Transporter family	*TCID	Example	Sugar
2.A Major Facilitator Superfamily (MFS)			
Sugar Porter (SP) family	2.A.1.1	Xylose:H^+ symporter	Xylose
Organophosphate:Pi antiporter (OPA) family	2.A.1.4	Glycerol-3-phosphate:Pi antiporter	*sn*-glycerol-3-phosphate
Oligosaccharide:H^+ symporter (OHS) family	2.A.1.5	Lactose:H^+ symporter	Lactose
Fucose:H^+ symporter (FHS) family	2.A.1.7	Fucose:H^+ symporter	Fucose
Polyol permease (PP) family	2.A.1.18	D-Arabinitol:H^+ symporter	Arabinitol
Sugar efflux transporter (SET) family	2.A.1.20	Sugar-efflux transporter A	Multisugar
Glycoside-pentoside-hexuronide (GPH):cation symporter family	2.A.2	Melibiose permease	Galactoside

Gluconate:H^+ symporter (GntP) family	2.A.8	D-Gluconate:H^+ symporter	D-Gluconate, L-idonate
The Glycerol Uptake (GUP) Family	2.A.50	Glycerol: H^+ symporter	Glycerol
3.A.1 ATP-binding cassette (ABC) superfamily			
Carbohydrate uptake transporter-1 (CUT1) family	3.A.1.1	MalEFGK transporter	Maltose
Carbohydrate uptake transporter-2 (CUT2) family	3.A.1.2	RbsABC transporter	Ribose
4.A Phosphoryl transfer-driven group translocators			
PTS glucose-glucoside (Glc) family	4.A.1	Glucose porter (PtsG)	Glucose
PTS fructose-mannitol (Fru) family	4.A.2	Fructose, mannose and mannitol specific porters	Fructose, mannose, mannitol
PTS lactose-N,N9 diacetyl chitobiose-b-glucoside (Lac) family	4.A.3	Lactose PTS group translocator	β-glucosides
PTS glucitol (Gut) family	4.A.4	Glucitol porter	Glucitol, 2-C-methyl-D-erythritol
PTS galactitol (Gat) family	4.A.5	Galactitol porter	Hexitol, galactitol
PTS Mannose-fructose-sorbose (Man) family	4.A.6	Mannose porter	Phosphorylate glucose, mannose, fructose

* Information of TCID was retrieved from Transporter Classification Database (TCDB, Saier et al., 2016).

1.1.4 ATP-Binding Cassette (ABC) transporters

ABC transporter proteins are evolutionarily conserved among almost all domains of life, including humans (Dassa and Bouige, 2001; Fuellen et al., 2005; Davidson et al., 2008). On the basis of substrate-transport direction, ABC transporters are broadly classified as ABC exporters (outside cell) and importers (inside cell). ABC exporters are found in all three domains of life i.e. archaea, bacteria and eukarya, whereas importers are present only in prokaryotes and more recently have been identified in plants (Shitan et al., 2003; Terasaka et al., 2005). ABC transporters have been evolved to perform a wide array of functions such as import of nutrients, excretion of toxins,

antigen presentation, cell signaling, cell-wall peptide recycling, quorum sensing and pathogenesis and many other biological functions (Higgins, 1992; Dassa and Bouige, 2001; Jones and George, 2004; Lewis et al., 2012; Paul et al., 2013). Moreover, it has also been reported to play a central role in many physiological processes and in various fields of biomedical research, including multidrug resistance, cancers and genetic disorders. Structural, molecular and biochemical studies of ABC transporters have assisted in inferring their catalytic mechanism.

1.1.4.1 ABC importer

In prokaryotes, uptake of various nutrients inside the cell is majorly mediated by ABC importers. These include a wide variety of carbohydrates, metal ions, siderophore, amino acids, polyamines, oligopeptides and vitamins. To facilitate their uptake, ABC importers employ a periplasmic SBPs, which are specific to these nutrient modules' and subsequently deliver it to the translocator. Based on the substrate being transported, ABC importers are divided into ~24 families, out of which, two families are classified as carbohydrate uptake transporters-1 and-2 (CUT1 and CUT2) (Table 1.3) (Saier, 2000a). In addition to nutrient diversity, the structural characterization of ABC importer reveals that it shows significantly different topological folds and thus further classified as the Type I and II importer (Lewinson and Livnat-Levanon, 2017). In contrast to Type I and Type II importer, structurally divergent ABC importer i.e Energy-Coupling Factor (ECF) transporters, are classified as Type III importers (Figure 1.3). Interestingly, all these three types of ABC importers are found exclusively in prokaryotes. Both Types I and II ABC importers require an additional periplasmic protein i.e. SBPs for substrate acquisition, while Type III importers lack the SBPs instead it use S-component, which is a transmembrane protein (Quiocho and Ledvina, 1996; Rodionov et al., 2009; Berntsson et al., 2010). Among these, three different types, CUT families are proposed to belong to the Type I importers.

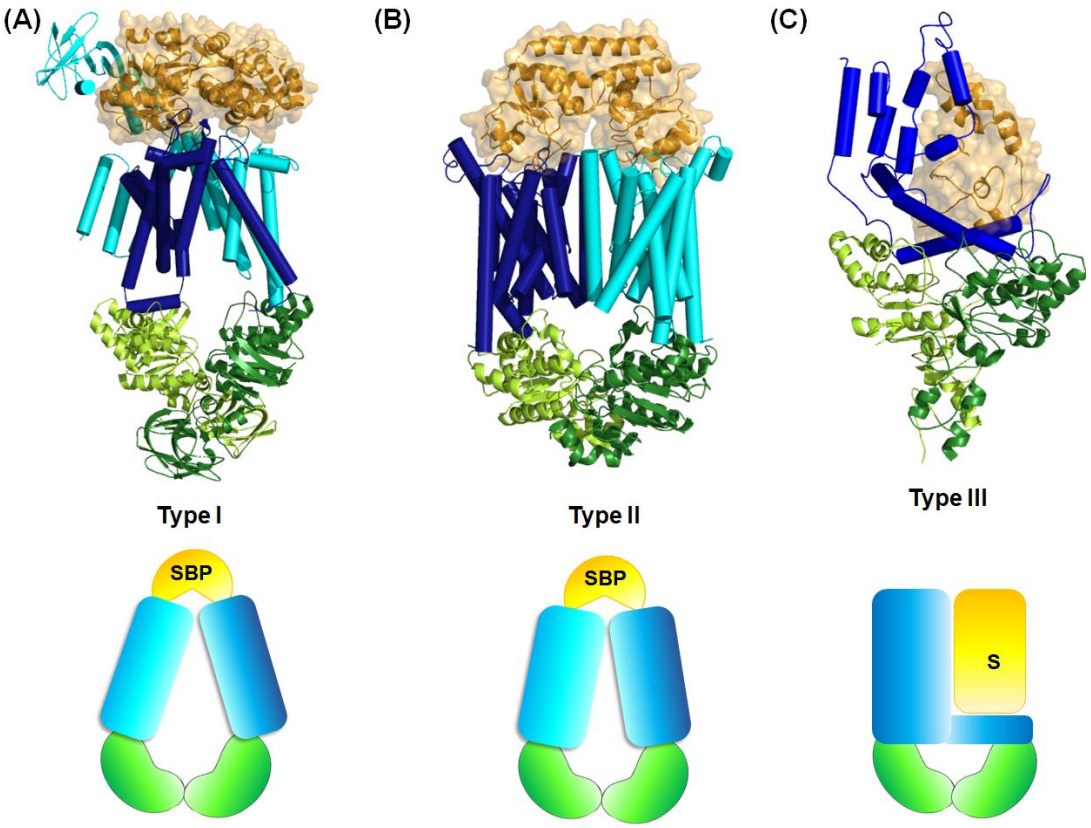


Figure 1.3. Schematic representation of ABC importer types. Ribbon representation of (A) Type I (e.g. maltose ABC transporter, PDB ID: 4KHZ) (B) Type II (e.g. vitamin B12 ABC transporter, PDB ID: 1L7V) and (C) Type III (e.g. ECF transporter, PDB ID: 6FNP) ABC transporter.

Table 1.3. Families of ABC importers for carbohydrate uptake.

Transporter family	*TCID	Example	Substrate
Carbohydrate uptake transporter-1 (CUT1) family	3.A.1.1	MalEFGK of *E. coli*	Maltose
Carbohydrate uptake transporter-2 (CUT2) family	3.A.1.2	RbsABC of *E. coli*	Ribose

* Information of TCID was retrieved from Transporter Classification Database (TCDB, Saier et al., 2016).

1.1.4.2 Carbohydrate uptake transporters family

Bacteria use a variety of sugars as a source of carbon and energy. Genomic sequence analysis of several bacterial genomes reveals that they possess a remarkable arrangement of metabolic pathways, including uptake and processing systems for different types of sugars. Uptake of sugar inside the cell is mediated by CUT1 and CUT2 families having TCID 3.A.1.1 and 3.A.1.2, respectively (Table 1.3). CUT1 and CUT2 families differ on the basis of the substrates being transported. The CUT1 family members facilitate the transport of di- and oligosaccharides, whereas, the CUT2 family members help in transporting monosaccharides only.

Carbohydrate uptake transporters-1 (CUT1)

Protein members of CUT1 family transport a variety of di- and oligosaccharides such as malto-oligosaccharides, cellobiose, glycerol-phosphate and polyols (Table 1.4). The members of the CUT1 family show sequence homology with the ABC transporters involved in oligo and/or dipeptides uptake. Proteins of CUT1 family adopt a typical architecture of ABC importer synthesized by four gene products (an extracellular-binding protein**,** two TMDs spanning the membrane and ATPase subunits). Maltose/maltodextrin transporter from enterobacteria is a well-characterized protein member of the CUT1 family (Oldham and Chen, 2011).

Table 1.4. List of sugar ABC importers from the CUT1 family.

Transporter	TCID	Example	Substrate	PDB ID
Malto-oligosaccharide porter	3.A.1.1.1	MalEFGK of *E. coli*	Maltose	1ANF
Glycerol-phosphate porter	3.A.1.1.3	UgpABCE of *E. coli*	*sn*-glycerol-3- phosphate	4AQ4
Maltose /trehalose porter	3.A.1.1.7	MalEFGK of *T. litoralis*	Trehalose	1EU8
Malto-oligosaccharide porter	3.A.1.1.16	MalEFGK of *P. furiosus*	Maltotriose	1ELJ

Cellobiose / Cellotriose / Cellopentose	3.A.1.1.23	CebEFGMsiK of *S. reticuli*	Cellobiose, cellotriose, cellopentose	2O7I
ABC-type fucose uptake porter	3.A.1.1.47	FucABCD of *S. pneumoniae*	Fucose	2W7Y
The lacto-N-biose I transporter	3.A.1.1.48	The LNB/GNB uptake transporter of *B. longum*	Gal β-1,3-GlcNAc / galacto-N-biose (GNB; Gal β-1,3-GalNAc)	2Z8D

* Information of TCID was retrieved from Transporter Classification Database (TCDB, Saier et al., 2016).

Carbohydrate uptake transporters-2 (CUT2)

The members of the CUT2 family are specific for monosaccharides (Table 1.5) (Schneider, 2001). Its domain organization is similar to that of the CUT1 family containing an SBP, a hydrophobic TMD and an ATPase subunit (Schneider, 2001; Koning et al., 2002). The CUT2 family transporters are found both in bacteria and archaea; however, archaeal proteins have not been well-characterized. The CUT2 family protein, which helps in the import of glucose/galactose and D-allose from *S. solfataricus* are biochemically and biophysically characterized and contains domain organization similar to that of the CUT1 family members (Koning et al., 2002). In addition, the ribose and galactoside transporters from *E. coli* and *Salmonella typhimurium* are also well-characterized members of the CUT2 family (Schneider, 2001).

Table 1.5. List of sugar ABC importers from the CUT2 family.

Transporter	***TCID**	**Example**	**Substrate**	**PDB ID**
Ribose porter	3.A.1.2.1	RbsABC of *E. coli*	Ribose	1DRJ

L-arabinose porter	3.A.1.2.2	AraFGH of *E. coli*	L-arabinose, D-fucose, D-galactose	1ABE
Glucose/ galactose porter	3.A.1.2.3	MglABC of *E. coli*	Glucose/galactose	2FVY
D-allose porter	3.A.1.2.6	AlsABC of *E. coli*	D-allose	1GUB

* Information of TCID was retrieved from Transporter Classification Database (TCDB, Saier et al., 2016).

1.1.5 ABC transporter architecture

The structural organization of ABC transporter and its transport mechanism is common in all three kingdoms of life. Structurally, both ABC exporters as well as importers contain two nucleotide-binding domains (NBDs) and two transmembrane domains (TMDs). The NBDs hydrolyze ATPs to provide free energy and TMDs provide a translocation passage for substrates transport. The NBDs are highly conserved across organisms, whereas TMDs show significantly low sequence conservation (Saurin et al., 1999). In addition to these, ABC importers entail another component termed as substrate (solute)-binding proteins (SBPs). SBPs are an accessory domain required to capture substrate and deliver it to TMDs. In Gram-negative bacteria, SBPs are free-floated in the periplasm, while in Gram-positive bacteria and archaea, SBPs are lipid anchored to the plasma membrane. An overview of ABC transporter architecture and SBP subunit organization has been provided in Figure 1.4.

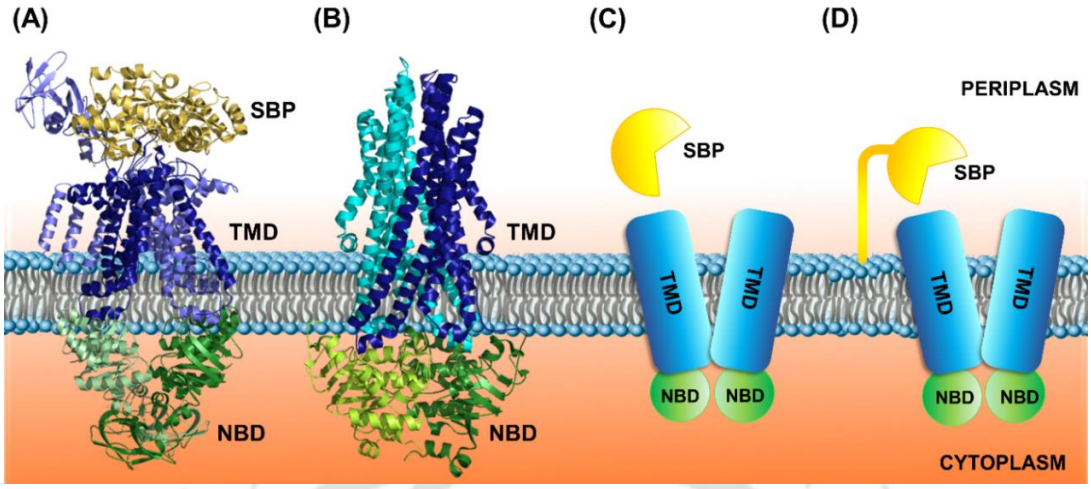


Figure 1.4. The overall topology of ABC importer and exporter. (A and B) Domain architecture of ABC importer e.g. maltose ABC transporter (PDB ID: 4KHZ) and exporter e.g. Sav1866 (PDB ID: 2HYD), respectively. The ABC transporter subunits TMDs, NBSs and SBP are represented as blue, green and yellow, respectively. Schematic representation of SBP organization in (C) Gram-negative bacteria and (D) Gram-positive bacteria and archaea (right). Respective free-floating and lipid anchored SBPs are shown in yellow.

1.1.5.1 Transmembrane domains (TMDs)

TMDs are embedded in the membrane with the core structure containing 10-12 transmembrane (TM) helices, which form a central translocation pathway. The number of TM helices varies in ABC exporters and importers. The central channel of exporters comprises of 12 TM helices, while in importers, this number varies between 10 and 20 (Hollenstein et al., 2007). Depending upon ABC transporters Types (I, II and III), the TMDs folds vary. Also, the TMD fold of Type III importer is related to ABC exporters (Rees et al., 2009). Type I ABC transporter contains six transmembrane (TM) helices per TMD subunit (Parcej et al., 2007). For example, TMD subunit of moldydate ABC transporter (ModB, PDB ID: 3D31) and maltose ABC transporter (MalF and MalG, PDB ID: 4KHZ) contains six to eight TM helices and organized into homo-and hetero-dimer association (Oldham et al., 2007; Gerber et al., 2008). Similarly, Type II transporters such as vitamin B12 ABC transporter (PDB ID: 1L7V) comprises a higher number of TM helices with 10 to 12 α-helices per TMD subunit (Locher et al., 2002). In contrast, in Type III transporters (e.g. ECF

transporter, Sav1866 and P- glycoprotein ABC transporter) TM helices from six helices, transverse the plasma membrane and extended into the cytoplasm (Figure 1.5) (Dawson and Locher, 2006).

TMDs are generally found in two confirmations (1) open towards the periplasmic site (outward open) and (2) open towards the cytoplasmic site (inward open) (Wilkens, 2015). TMDs lack conservation at the primary structure level, thus allowing the transport of various substrates. Inner TM helices provide binding sites for substrate or may form a hydrophobic cavity (Korkhov et al., 2012; Oldham et al., 2013). For example, maltose ABC importer from *E. coli* (Type I fold) contains a site for three sugar rings. During translocation, the reducing end of sugar interacts with the TM helices residues (Oldham et al., 2013). Like Type I importers, ABC exporters also contain substrate-binding sites, but with more number of interaction sites. For instance, P-glycoprotein, a multidrug exporter, contains multiple drug-binding sites leading to the efflux of various drugs (Loo et al., 2003; Aller et al., 2009).

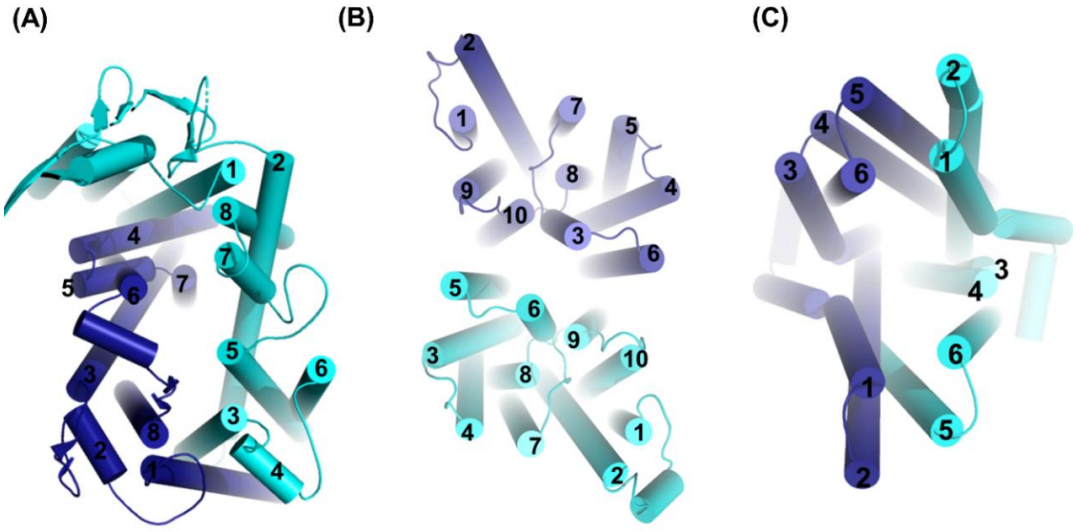


Figure 1.5. Schematic representation of the transmembrane domain (TMD) fold. A canonical transmembrane domain (TMD) fold of (D) Type I, (E) Type II and (F) Type III ABC transporters. Distribution of TM helices per TMD subunit in a different class of ABC transporter are numbered and color-coded according to TMD subunit.

1.1.5.2 Nucleotide-binding domains (NBDs)

NBDs are catalytic domains of ABC transporters and reside on the cytoplasmic phase of the transporter. Crystallographic structures of these domains reveal that bacterial and eukaryotic NBDs have a conserved fold similar to RecA domain. The comparative analysis of the crystal structures of ATP-bound NDBs and free forms (without ATP) suggests that the ATP binding to NBDs is required for its dimerization (Hung et al., 1998; Chen et al., 2003). Two ATP molecules are sandwiched between the dimer to form a 'sandwich dimer' (Kerr, 2002). In this dimeric arrangement, it contains several conserved motifs that are involved in the ATP binding (Schneider and Hunke, 1998; Davidson and Chen, 2004). Two conserved motifs of NBDs known as 'Walker-A or P-loop' and 'LSGGQ' motifs help in binding ATP molecules (Figure 1.6). In addition, the 'Walker B' motif is required for the catalytic activity i.e. hydrolysis of ATP by NBDs. A glutamate residue of 'Walker B' motif plays a role of nucleophile. Another motif termed as 'D loop' mediates the interaction between TMDs and NBDs (Zaitseva et al., 2005; Ambudkar et al., 2006).

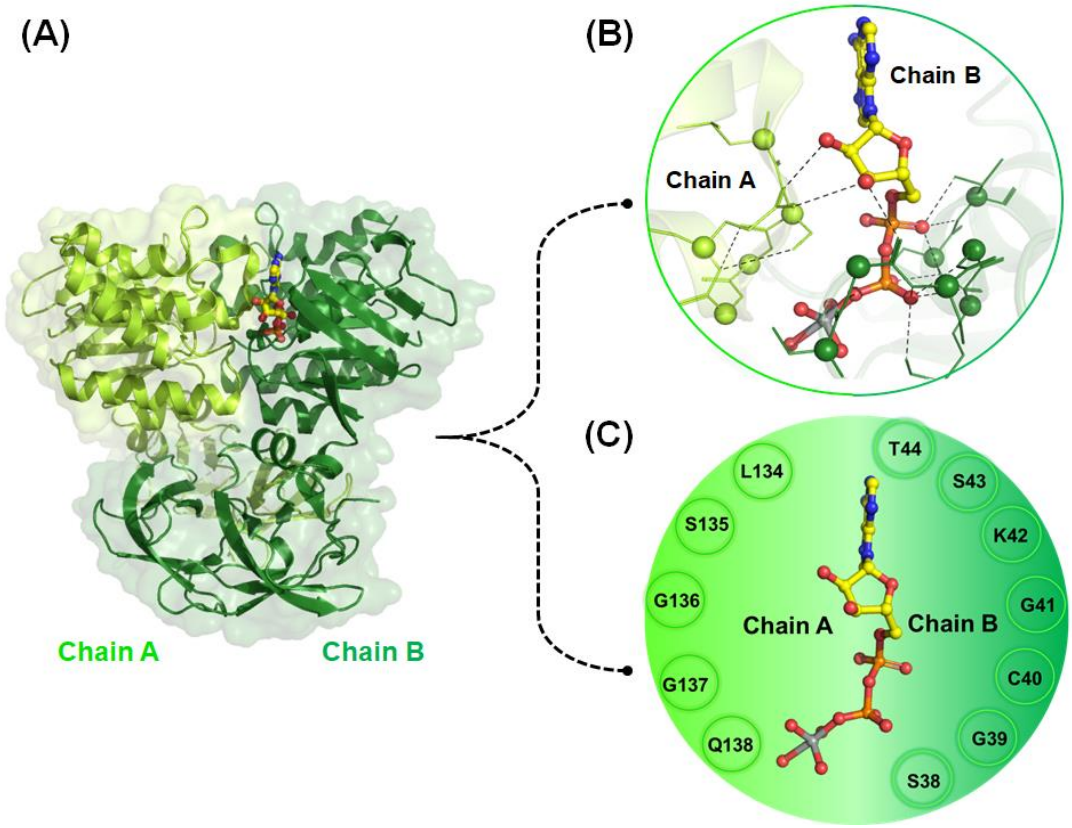


Figure 1.6. Domain architecture of the nucleotide-binding domain (NBD). (A) Dimeric conformation of NBD subunit of maltose ABC transporter (PDB ID: 3PUV). Two protomers (chain A and B) are shown in light and dark green. ATP molecule bound at the interface of the dimer is shown in yellow ball-and stick-model. (B and C) A schematic representation of the ABC signature motif i.e. LSGGQ (light green) and Walker A (dark green) motifs interacting with ATP molecule. Hydrogen bonding between motif residues and ATP molecule are shown with dotted lines. Respective residue number of ABC signature motif is represented in the Figure 1.6C.

1.1.5.3 Substrate (solute)-binding proteins (SBPs)

Substrate (solute)-binding proteins (SBPs) are a class of proteins that is associated with membrane protein complexes for transport of substrates. SBPs are soluble proteins found in the periplasm of the Gram-negative bacteria, sometimes called as periplasmic-binding proteins (PBPs) (Berger and Heppel, 1974). Originally, SBPs were discovered as a component of prokaryotic ABC transporters (Berger, 1973; Berger and Heppel, 1974; Tam and Saier, 1993; Quiocho and Ledvina, 1996; Wilkinson, 2002). Later on, it has been found that besides ABC transporters, it is also associated with prokaryotic tripartite ATP-independent periplasmic (TRAP) transporters (Gonin et al., 2007; Mulligan et al., 2009). In order to import different types of nutrients, ABC importers uses different types of SBPs. Depending upon the substrate, different SBPs are named accordingly like sugar-binding protein, metal-binding proteins and amino acid-binding proteins. One of the well-characterized SBPs is a maltose-binding protein of maltose ABC transporter (MalE-MalFGK2) from *E. coli* (Orelle et al., 2008; Oldham and Chen, 2011). Topologically, all SBPs contains two α/β domains, in which β-sheets are flanked by α-helices. The two domains, N-and C-terminal domains (NTD and CTD) are linked by a 'hinge region', which is a site for ligand binding (Figure 1.7A). In a ligand unbound state (open conformation), these two domains are very flexible, allows the free rotation of NTD and CTD around the hinge region. Upon ligand binding, hinge region stabilizes the two domains (NTD and CTD) and attains the closed state of SBPs (Tang et al., 2007). During open to closed state transition, both NTD and CTD move around the hinge region through a well-known ''Venus Fly-trap'' mechanism (Mao et al., 1982) (Figure 1.7B).

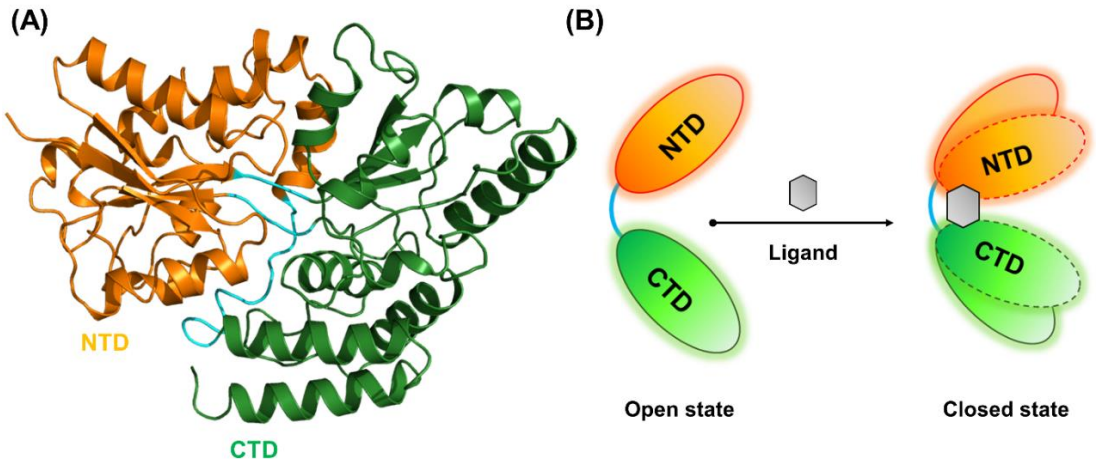


Figure 1.7. Structural topology of SBPs. (A) The tertiary structure of SBP (e.g. α-glycoside-binding protein, PDB ID: 6J9W) is shown as a ribbon model. The NTD and CTD are shown in orange and green, respectively. Hinge region connecting the NTD and CTD are represented in cyan. (B) Schematic model representation of the ‘‘Venus Fly-trap’’ mechanism. The conformational changes of the NTD (orange) and CTD (green) upon ligand (grey) binding are depicted with dotted lines.

1.1.6 Substrate (solute)-binding proteins (SBPs) classification

Initially, based on sheet topology SBPs are categorized into three classes: I, II and III (Fukami-Kobayashi et al., 1999; Lee et al., 1999). Two classes (I and II) differ by their β-sheet arrangements and linker regions. In class I, the β-sheet is organized as β_2-β_1-β_3-β_4-β_5 arrangement, whereas in class II, it adopts β_2-β_1-β_3-β_n-β_4 topology (n is crossover β-strand between NTD and CTD) (Fukami-Kobayashi *et al*., 1999). The hinge region of class I is generally made up of three connecting strands, while in class II, only two connecting strands are required. The class III SBPs are different from classes I and II in that a single α-helix connects the two domains of SBPs and functions as a hinge region (Karpowich et al., 2003).

Later on, many SBPs have been biochemically and biophysically characterized. Even though the amino acid sequence identity between these SBPs is low, their overall three-dimensional structural fold is highly conserved (Berntsson et al., 2010). In 2016, Scheepers et al., collected the 504 unique entries for SBP structures deposited in Protein Data Bank (PDB), which binds to a different type of ligand. Using these

structures, all SBPs have been classified into seven different clusters (A-G) (Scheepers et al., 2016). Based on their structural features and ligand specificity, clusters A, B, D, E and F were further subdivided into subclusters. These clusters possess a unique topology, where the hinge region is identified as describing feature for each cluster (Figure 1.8).

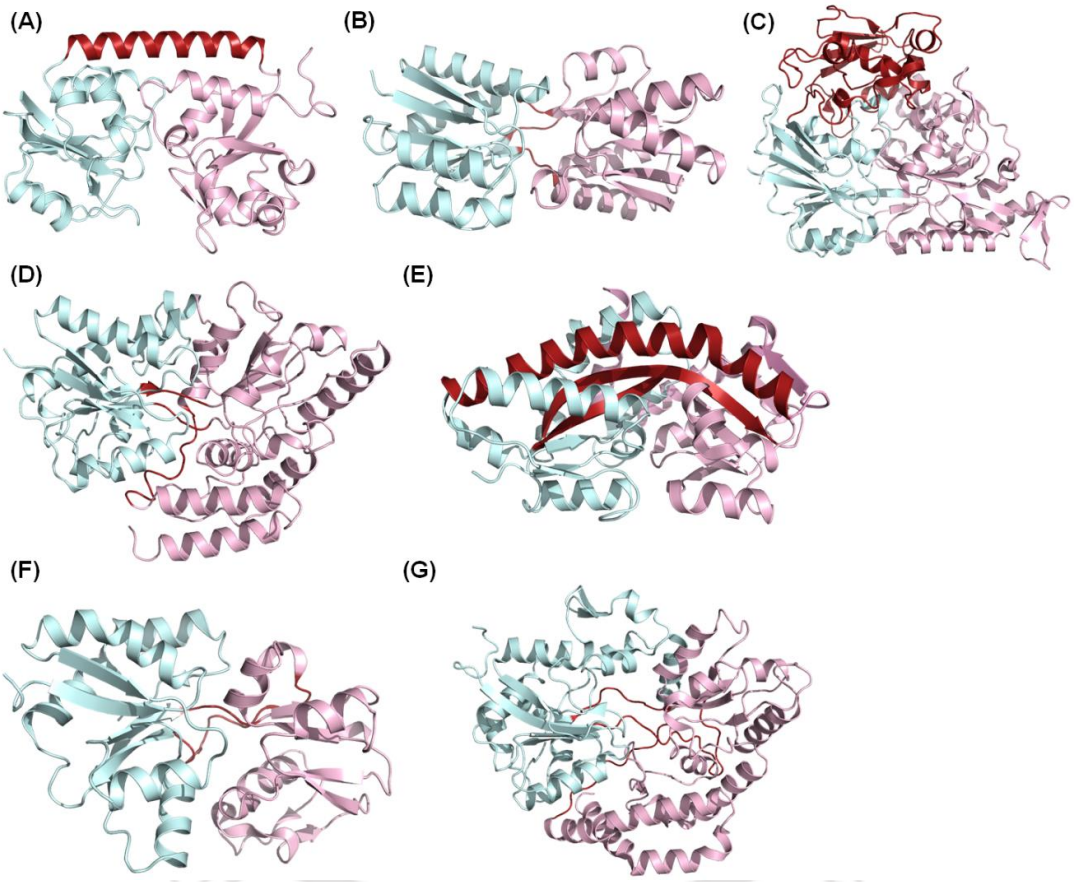


Figure 1.8. Structure based classification of SBPs. Based on overall topology and structural features of hinge regions (highlighted in red) SBPs are classified as (A) cluster A (e.g. BtuF, PDB ID: 1N2Z), (B) cluster B (e.g. RBP, PDB ID: 1DRJ), (C) cluster C (e.g. OppA, PDB ID: 3DRF), (D) cluster D (e.g. ModA, PDB ID: 1AMF), (E) cluster E (e.g. UehA, PDB ID: 3FXB), (F) cluster F (e.g. MetQ, PDB ID: 4GOT) and (F) cluster G (e.g. AlgQ1, PDB ID: 1Y3P). For each SBP, NTD and CTD are represented in cyan and pink, respectively.

1.1.7 Carbohydrate uptake mechanism via ABC importer

Although Type I and Type II ABC importers differ in their topology, they transport substrates using an identical mechanism, known as 'alternate access mechanism' (Rice et al., 2014). In this mechanism, ABC importers undergoes conformational transition from inward to outward to facilitate the import of molecules. One of well-characterized sugar ABC transporter viz. maltose ABC transporter (MalFGK) follows the alternate access mechanism to import of maltose molecules (Oldham et al., 2007; Oldham et al., 2013). In this mechanism, firstly SBP (MalE) interacts with maltose molecule and changes it conformation from open to closed state. Subsequently closed MalE interacts with cognate TMDs (MalFG) and lead to the movement of TMDs. This movement results in inward conformation of the transporter, where NBDs (MalK) are remain in the in a nucleotide-free semi-open state (Figure 1.9, state 2). In this state, binding of two ATP molecules to the NBDs induces the transition of transporter from inward conformation to the outward-facing conformation (Figure 1.9, state 3). Subsequently, it opens the SBPs and lead to the release of maltose molecule into a translocation passage where it interacts with TMDs (Figure 1.9, state 3). Upon ATP hydrolysis, maltose ABC transporter release maltose molecule into the cytosol and returns to its resting state i.e. inward conformation (Figure 1.9, state 4).

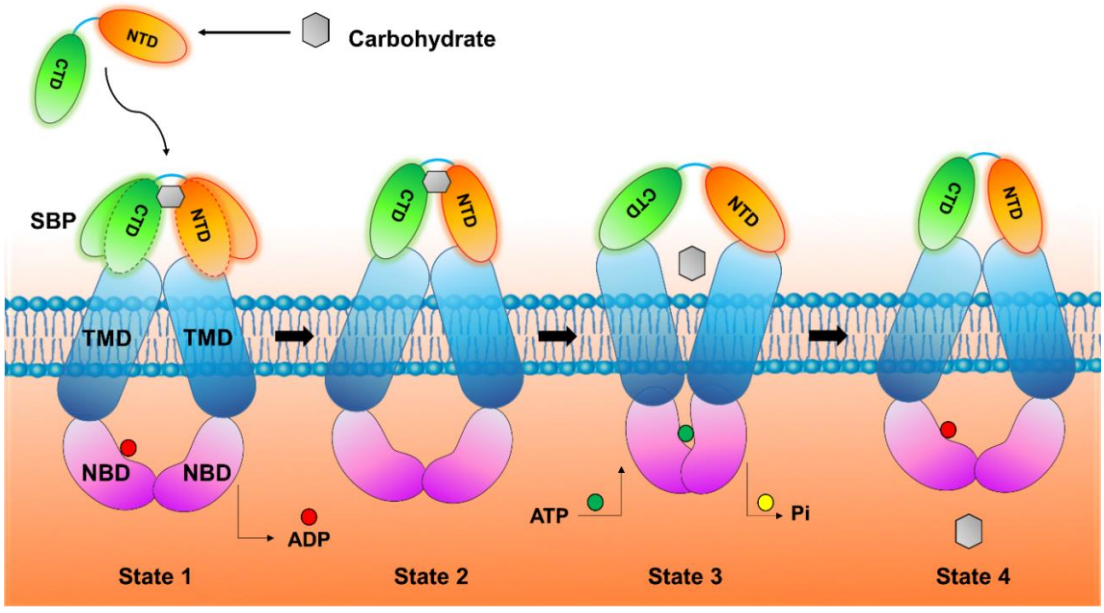


Figure 1.9. Alternate access mechanism of ABC importers. Conformational transition of ABC transporter subunits SBP (green and orange), TMDs (blue) and NBDs (magenta) during sugar molecule (grey) transportation is represented across the four different states 1 to 4. (Mechanism steps are redrawn from Rice et al., 2014).

1.2 IMPORTANCE OF THE STUDY

Uptake of carbohydrates by sugar ABC transporters are major feed lines in pathogenic and nonpathogenic bacteria to fulfill the requirement of carbon and energy source. Many biochemically and biophysical characterization of SBPs of sugar ABC transporters have reported a clear contribution to virulence. In pathogenic bacteria, these roles are associated with colonization in the host environment and escape from the host defense mechanism. As these proteins are specific to only prokaryotic ABC transporter and absent eukaryotes including human, they are proven to be useful targets for antimicrobial therapeutics. However, an application of these SBPs for designing of therapeutic drugs, requires a clear understanding of their transport mechanism. Although, SBPs involved in the carbohydrate uptake are extensively studied, however, due to the vast chemical diversity of carbohydrates, understating of the transport mechanism is remains challenging. Knowledge of the three-dimensional structures of SBP is one way to understand their molecular mechanism of action. However, despite the low sequence identity, SBPs are reported to have a conserve topology and thus (sub)cluster B and D SBPs are specified for the carbohydrate uptake. In contrast to SBPs, carbohydrates exhibit diversity for their length, glycosidic linkage, anomeric configuration, and epimeric state. Hence to understand that how structurally conserves SBPs select the specific carbohydrate molecule and mediates the uptake of diverse carbohydrates, we investigated the carbohydrate specific SBPs from *Thermus thermophilus* HB8. The detailed structural, molecular and transport kinetic information will contribute to understanding the mechanism of SBPs of ABC importers. Also, the findings of the proposed study will further establish the base for designing potent drugs against the SBPs from the target pathogens.

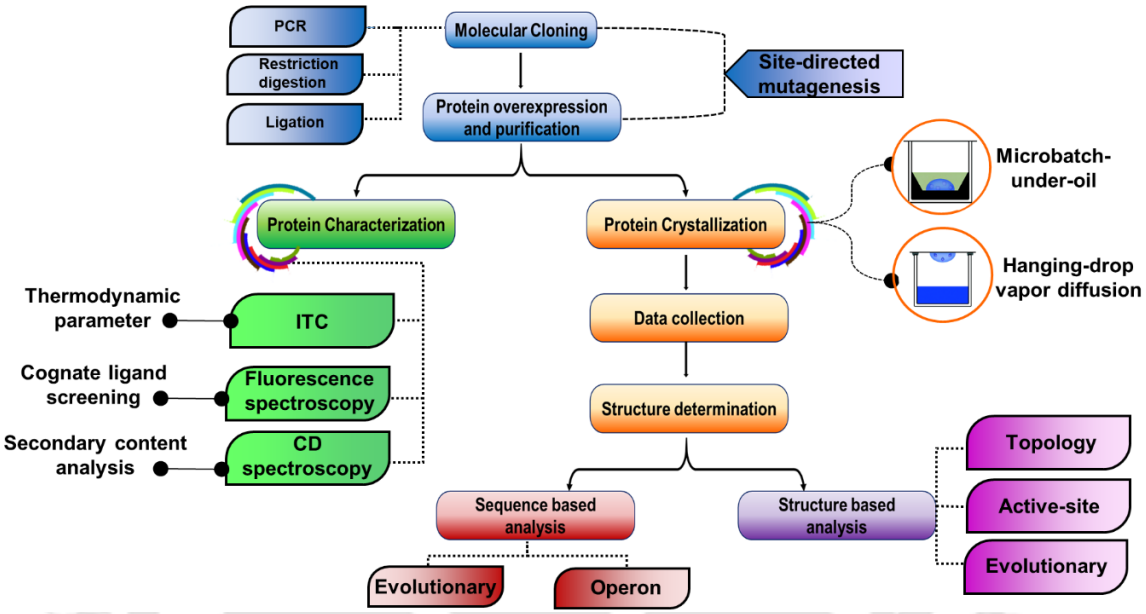
1.3 OBJECTIVES

To get insight the selective transport mechanism of ABC transporter for diverse carbohydrates, the following objectives were phrased-

1. *In silico* characterization of all 11 genes (*TTHV089, TTHA0356, TTHA0979, TTHA1652, TTHA0379, TTHB082, TTHA0688, TTHA0379, TTHA1877, TTHA1936* and *TTHV034*) encoding for substrate-binding proteins from *Thermus thermophilus* HB8.
2. Molecular cloning, over expression and protein purification to its homogeneity of selected three proteins TTHA0356, TTHB082 and TTHA0379.
3. Crystallization of selected three proteins to elucidate their three-dimensional crystal structures.
4. Co-crystallization of the purified proteins with their cognate substrate (sugars) to study their binding and kinetics.
5. Biochemical and biophysical studies to understand the binding mechanism of SBPs.

General outline of the chapter 2

This chapter describes the designing and methodology of various experiments performed for the structural and functional characterization of the selected SBPs of ABC transport systems. General workflow representing the major experimental outline is described below with different colour code.



CHAPTER 2- MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Reagents

For molecular cloning, restriction enzymes, ligase and alkaline phosphatase enzymes were procured from New England Biolabs (NEB) and/or Thermo Fisher Scientific. In addition, different kits used for different purposes, such as plasmid isolation, PCR clean-up and gel extraction were purchased from Himedia and QIAGEN. For mutagenic experiments, Q5 site-specific mutagenesis kit was used procured from NEB. The Ni^{2+}-NTA affinity resin and pierce centrifuge column used for protein purification were purchased from QIAGEN and Thermo Fisher Scientific, respectively. Protein concentrators (Vivaspin turbo 15) were obtained from Sartorius. All crystallization buffers, plates and other crystallization-related chemicals and tools were purchased from Hampton Research (USA) and Molecular Dimensions (UK).

2.1.2 Carbohydrates

Disaccharide β-glycosides (D-(+)-cellobiose and D-(+)-sophorose) and α-glycosides (D-(+)-trehalose and D-(+)-maltose) and monosaccharides (D-(+)-glucose and D-(+)-mannose) were purchased from Sigma-Aldrich (St. Louis, Missouri, USA). Other higher β-glycosides like D-(+)-cellotriose, D-(+)-cellotetraose, D-(+)-cellopentaose, D-(+)-laminaribiose, D-(+)-laminaritriose and D-(+)-laminaritetraose were procured from Megazyme (Wicklow, Ireland). Carbohydrates like D-(+)-gentiobiose, D-(+)-lactose, D-(+)-sucrose, D-(+)-melibiose, D-(+)-raffinose, D-(+)-galactose and D-(+)-tagatose were obtained from the Himedia (Mumbai, India).

2.2 METHODS

2.2.1 Designing of recombinant constructs for protein overexpression

Molecular cloning is a molecular biology technique that is used to generate new recombinant DNA molecules, which possess a novel combination of the nucleotide sequence. In this study, the molecular cloning method was used to create the recombinant construct, in which target gene is inserted into the appropriate vector for propagation. Both vector and gene were mixed together in the presence of ligase enzyme and introduced

inside the host organism, which leads to replication of designed construct and produces the multiple copies of recombinant clones (Figure 2.1). Details of each step of molecular cloning performed in this study are described below.

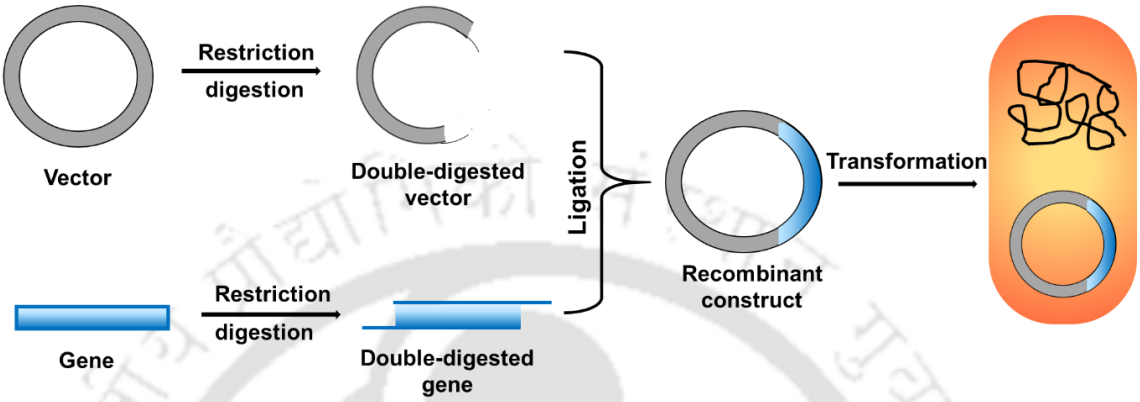


Figure 2.1. Schematic representation for molecular cloning.

Preparation of cloning and expression cells

For cloning and expression, *Escherichia coli* strains were used as competent cells which were prepared by using Calcium chloride ($CaCl_2$) based chemical method. In this method, competency was obtained by suspending the bacterial cells, which were in the log phase to a $CaCl_2$ solution for 30 min at 4°C. Expression cells used in this study are derived strains which contain a chromosomal copy of the T7 RNA polymerase to express the protein under the control of T7 promoter. Overexpression of protein was achieved by inducing the expression cells with lactose analog Isopropyl β- d-1-thiogalactopyranoside (IPTG), which binds to the repressor.

In addition to cloning, for expression optimization and soluble protein production following *E. coli* strains were used in this study.

- DH5-α competent cells have high transformation efficiency and thus used to increase the cloned plasmid copy number.
- BL21(DE3) competent cells were used to express the gene under the control of a T7 promoter.

- Rosetta (DE3) competent cells are BL21 derivatives and used to express the genes that contain rare codons. These strains provide the tRNAs for following rare codons: AGG, AGA, AUA, CUA, CCC and GGA.
- BL21-CodonPlus (DE3)-RIL competent cells are also BL21 derived cells to provide the extra copies of tRNAs for rare codons (AGA/AGG, AUA, and CUA).

Gene amplification

For the target gene amplification, gene-specific customized oligonucleotides primers were designed based on the following parameters melting temperature (Tm), GC content, restriction digestion sites. Each designed primer was analyzed using the web tool OligoAnalyzer to check all parameters. Clone plasmids carrying a full length of target gene was procured from the Biological Research Center, NITE (NBRC), Japan. However, this clone plasmid lacks 6x-His-tag essential for protein purification via Ni-NTA metal-affinity chromatography. Hence, subcloning of target genes was performed by amplifying the gene using the cloned plasmid as a template during polymerase chain reaction (PCR). The PCR reaction was performed in the following steps: 1) initial denaturation (95°C, 5-10 min), 2) final denaturation (95°C, 30-40 sec), 3) annealing (Tm of genes, 45 sec), 4) initial extension (72°C, 50 sec) and 5) final extension (72°C, 10 min). Amplified genes do not contain the periplasmic signal sequence and contain 6X-His was tagged at the C-terminus of the protein to facilitate the purification process by immobilized Ni-NTA metal-affinity chromatography.

Double digestion of gene and vector

In order to carry out the directional cloning, amplified genes were double digested with respective restriction enzymes before inserting into pET series vectors. Restriction enzymes such as *Nde*I and *Bam*HI were chosen, which generates the 5′ and 3′ sticky ends of the genes and vectors. Before digestion, genes were mixed with both restriction enzymes together and incubated at 37°C for 4 hours. At the same time, vectors were mixed with similar restriction enzymes together and incubated at 37°C for 3 hours. After third hour, vector reaction mixture was mixed with alkaline phosphatase to remove the 5′-phosphate group and incubated an additional 1 hour at 37°C. After successful digestion, genes and

vectors were purified using the PCR purification kit (QIAGEN) and gel extraction kit (QIAGEN), respectively.

Ligation and clone confirmation

To finally obtain the recombinant plasmid, a ligation reaction was performed in which the target gene was inserted into a compatibly digested pET series vectors (pET22b and pET28a). This ligation reaction was catalyzed by the T4 DNA ligase enzyme, which forms the covalent bond between two DNA fragments. To performs the ligand experiment, double digested vector and gene having sticky ends were mixed together in a different ratio (1:1, 1:3, 1:4 and 1:5) and incubated at 22°C for 4 hours in the presence of T4 DNA ligase. After incubation, each ligation mixtures were then transformed into *E. coli* DH5α competent cells to propagate the ligated recombinant clone. Positives colonies were screened based on antibiotic selection by growing the transformed cells into antibiotic supplemented Luria Bertani (LB) agar plates. For further confirmation, grown colonies were further inoculated into Luria Bertani (LB) broth supplemented with antibiotics and subjected to plasmid isolation. After plasmid isolation, expected cloned plasmids were incubated with respective restriction enzymes and incubated at 37°C for two hours to confirm the presence of the target gene. After clone confirmation via double digestion of cloned plasmid, further confirmation was affirmed by DNA sequencing.

Site-directed mutagenesis

To understand the ligand-binding mechanism and find the active-site residues crucial for ligand binding site-directed mutagenesis were exploited using the Q5 site-specific mutagenesis kit (NEB, USA). Target site for the site-directed mutagenesis was selected based on preliminary *in silico* analysis, homology to known structure and the strongest binding with an endogenously-bound ligand. As site-directed mutagenesis is an *in vitro* method, custom-designed oligonucleotide primers mentioned in the further chapters are used to introduce the desire mutation in the wild type (WT) cloned plasmid. Experiment was performed in the following three steps: in the first step designed primers, WT clone plasmid as template and High-Fidelity DNA polymerase from the master mix formulation of the kit was used together to amplify the vector. In the second step, amplified products

were incubated with unique enzymes, namely kinase, a ligase and *Dpn*I to facilitate the circularization of the amplified products followed by removal of the template wild type (WT) DNA. Finally, the last step involves the transformation of the circularized PCR product into the is a high-efficient *E. coli* DH5α competent cells. Confirmation of desired mutations in the target gene was confirmed by plasmid DNA sequencing and structural solution of mutant proteins.

2.2.2 Protein overexpression, solubilization and purification

Overexpression of SBPs

SBPs cloned in pET series vectors were overexpressed by transforming the recombinant constructs into the above-mentioned *E. coli* expression host cells, namely BL21 (DE3), Rosetta (DE3) and BL21-CodonPlus (DE3)-RIL. Expressions of each SBP was optimized at 37℃ for different time intervals (1 to 6 hrs). To get the overexpression, IPTG was used as an inducer, however as a higher concentration of IPTG can lead to cell death or unproductive protein formation, the concentration of IPTG (0.1, 0.5 and 1.0 mM) was also optimized. Once the expression host was optimized, optimum temperature, IPTG concentration and expression time were optimized, which was subsequently utilized for the soluble protein production.

Protein solubilization

Once SBPs were overexpressed, the cells were harvested by centrifugation and resuspended in lysis buffer for homogenization. The homogenized suspension was subsequently subjected to the sonicator for cell disruption by using 33% amplitude with 2 s on and 10 s off time. Obtained lysate was heat-treated at 70℃ for 10 min to remove thermolabile host proteins. A clear lysate was obtained by pelleting the cell debris via centrifugation, which yielded the pellet and supernatant fractions. Solubilization of protein was analyzed by examining the obtained pellet and supernatant fractions in 12% SDS-PAGE. Once solubilized protein identified in the supernatant fraction in SDS-PAGE, it was further subjected to protein purification by using chromatography techniques.

Affinity chromatography for protein purification

Immobilized metal affinity chromatography (IMAC) method was employed to purify the desired recombinant SBPs containing 6xHis-tag at C-terminus. IMAC is based on the precise geometry of coordination between a transition metal ion (Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}) and polyhistidine residues. In this study, Ni^{2+} transition metal immobilized on a matrix was used to coordinate with the imidazole ring of 6xHis-tag of recombinant proteins. During purification, binding of the 6xHis-tag proteins was performed in the pierce centrifuge columns packed with the resins with an incubation period of 2 hrs. Following the binding, the column was washed with imidazole containing (10-20 mM) wash buffers to remove the non-specific proteins weakly bound to the column. In addition to imidazole, other additives such as salt (sodium chloride) and reducing agent (β-mercaptoethanol) were also included in the wash to reduce the non-specific binding of host proteins. In order to elute the protein of interest, coordination bond between the Ni^{2+} transition metal ion and histidine residues was disrupted by using higher concertation of imidazole (250-400 mM), which competitively binds to the Ni^{2+} transition metal ion and elutes the tagged proteins. Pure eluted factions analyzed in 12% SDS-PAGE were pooled together and then subjected to gradient dialysis for imidazole removal. After dialysis, proteins were concentrated up to the desired concertation required for biochemical and biophysical experiments.

2.2.3 Protein characterization

Isothermal titration calorimetry

To measure the thermodynamic parameters of protein-ligand interactions, isothermal titration calorimetry (ITC) was used. ITC is a biophysical method which determines the basic chemical details of binding such as stoichiometry (n), change in enthalpy (ΔH), change in entropy (ΔS) and equilibrium association constant (Ka). In this study, incremental titration-based ITC experiments were performed where a precise volume of ligand was titrated into the protein solution at discrete time intervals. This incremental titration can lead to heat changes either in exothermic (heat released) or endothermic way (heat absorbed). Heat changes appear as peak, which is subjected to calculate the area under the peak via integration. After integration, data are normalized and fitted to various

binding models (one binding site and two binding sites, sequential binding site) to calculate affinity, enthalpy and stoichiometry for the interaction. For each experiment control experiment was performed by titrating ligand into a buffer solution to calculate the heat of dilution. This obtained heat of dilution was subtracted from the reactions to get the true heat changes associate with protein-ligand interaction and then plotted against the protein to ligand molar ratio.

Fluorescence spectroscopy

As aromatic amino acids play a crucial role in the binding of carbohydrates or sugars, the binding pocket of SBPs are rich in aromatic amino acids. Hence to assess the sugar-binding with target SBPs, a tryptophan-based intrinsic protein fluorescence approach was used. Sample was excited with a fixed wavelength at 285 nm, which led to the emission of longer wavelength signals (310-450 nm), called fluorescence. Changes in the fluorescence signal strongly depend on the protein conformation and ligand binding or dissociation, which affects the local tryptophan environments. Thus in this study, changes in the fluorescence of SBPs were monitored in the presence and absence of different sugars. Addition of sugar to the protein solution decreases the fluorescence signals (quenching), anticipating that alteration in protein conformations changes the local environment of tryptophan.

Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy is a widely used biophysical technique for protein characterization and thermal stability analysis. It provides the information about the different secondary structural content of proteins such as α-helix, parallel and anti-parallel β-sheet, turn, and random coil in a range between 260 and 190 nm wavelength. In this study, the thermodynamics of protein unfolding was monitored as a function of temperature to calculate the melting temperature (Tm) of proteins. As target SBPs are from thermophilic organism i.e., *T. thermophilus* HB8, the thermal stability of protein was monitored in up to higher temperatures ranging from 20 to 120 ºC. Unfolding of SBPs were investigated by changes in the ellipticity at a wavelength range from 260 and 190 nm and analyzed for the changes in secondary structural content i.e., α-helical protein (at 220 nm) and β-sheet (at 195 nm) as a function of temperature.

Mass spectrometric analysis

To determine the exact molecular weight of proteins, matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) was utilized. Before measurement, our protein of interest was homogenously purified and mixed with the matrix in a ratio (protein: matrix) of 1:2, 1:3 and 1:5. Sinapinic acid is a matrix of choice for analyzing proteins above 10 kDa. The matrix was prepared to a concentration of 10 mg ml^{-1} dissolved in a solvent containing Acetonitritle and 0.1% Trifluoroacetic acid in a ratio of 30:70, respectively. Sinapinic acid was dissolved in the above-mentioned solvent by sonicating the solution for 30 mins and followed by centrifuging at 12000xg for 10 mins, to get rid of excessive matrix crystals. The prepared matrix was mixed with protein samples of 1, 5 and 10 mg ml^{-1} in the ratios specified above. 2 µL spots of the protein-matrix mix were made on the ground steel MALDI plate grooves and allowed to air dry. Post drying, the drops were ionized in Bruker autoflex speed and the data was plotted using the in-built Flex control analysis software.

2.2.4 Crystallization of SBPs

In this study, X-ray crystallographic techniques was used to determine the SBP structures, which require diffraction quality protein crystals. In order to find the good quality protein crystals, the target protein solution was screened into a variety of crystallization to get initial crystal hits. However, identification of crystallization conditions for a new protein usually depends on the trial-and-error experiments with various optimizations. Thermodynamically, the crystallization process has four zones, namely undersaturated, saturated, metastable, labile, and precipitation, where crystal formation occurs at a supersaturated zone (Figure 2.2). To reach the supersaturated state, two major approaches was used (1) vapor diffusion (2) microbatch-under-oil. For a vapor diffusion technique, hanging-drop method was used where a drop containing a mixture of protein solution and precipitant were placed on coverslip and then allowed to equilibrate against the reservoir solution in an inverted position. In contrast, microbatch is an immersed based technique where a mixture of protein and precipitant is submerged under the oil to slow down the evaporation. In the present study, a mixture of paraffin and silicone oil having different density were used as a sealing agent to control the diffusion rate. In order to get the

successful protein crystal, protein solubility was screened at two different temperatures (4 and 20ºC) in both the hanging-drop and microbatch-under-oil method. Once the initial hits were obtained, further optimization was done by separating the two major events of crystallization i.e., nucleation and crystal growth via microseeding method. In this method, submicroscopic crystals were prepared and used as a seed to introduced into equilibrated new drops.

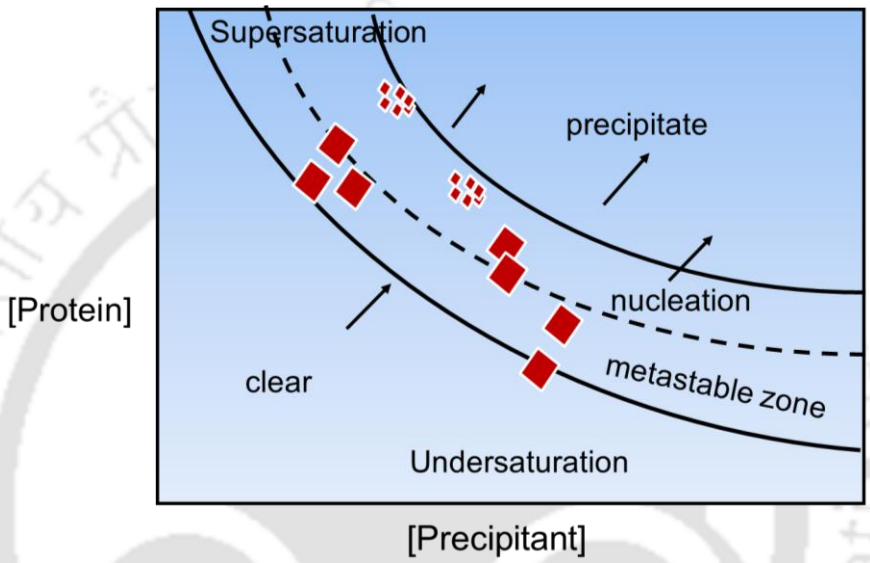


Figure 2.2. Schematic representation for the phase diagram of protein crystallization. (Image is redrawn from Stewart et al., 2011).

2.2.5 Data collection and processing

Data collection

Once the single and suitable crystals were obtained, they were exposed to a monochromatic X-ray beam to get the diffraction pattern. Since crystal contains the periodic arrangement of atoms, X-ray scattering occurs either in constructively or destructively way. Bragg`s law defines the relation between constructive interference and geometrical crystal parameters. According to this law, X-ray scattering depends on the incident beam wavelength (λ), the glancing angle (2θ) and positive integer (n).

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

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

Data processing

In data processing, raw diffraction data is converted into Bragg reflections with measured intensities. It occurs in the following four steps: calibration, integration, merging and scaling. For each step, algorithms are derived and available in software packages. In this study, processing of the raw diffraction images was carried out using the program iMosflm for initial processing (Battye et al., 2011), Aimless for scaling and merging (Evans and Murshudov, 2013). iMosflm program is embedded as part of the CCP4 suite and performs the indexing to provide the index for each reflection and an initial statistic on unit cell dimensions in a crystallographic symmetry. To determine the possible point and space groups, data set were subjected for scaling and merging using program Aimless. Conversion of intensities into structure factors was done by using the program CTRUNCATE distributed in the CCP4 package (Winn et al., 2011).

2.2.6 Structure determination

Obtained intensities from X-ray diffraction experiment provide the information about structure amplitudes, but unable to give the information of phase angles. This phase angles are crucial for calculation of the electron density (Taylor GL., 2010). There are several methods that provides the initial phases, which includes: 1) molecular replacement (MR), (2) multiple-wavelength anomalous dispersion (MAD) and (ii) Molecular isomorphous replacement (MIR). In this study, to obtain the phase information for the structural solution of SBPs, molecular replacement was used (Evans and McCoy, 2008). In MR, a known molecular structure of homologs protein was used to solve the structure of new

target SBP using the program Phaser (McCoy et al., 2007), which utilizes multivariate statistics and maximum likelihood probability theory. In order to get the initial phase by MR, it requires a significant sequence similarity (>25%) between know homologous and unknown proteins. The phases information obtained from the model structure was further used to calculate phases from the target protein intensities.

2.2.7 Model building and structure refinement

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

In this study, refinement of build model was done using the program Refmac5 (Vagin et al., 2004), which utilizes Bayesian statistics and maximum likelihood probability. Refinement of model improves the agreement between calculated structure factors from the model and the observed structure factor from the experiment. After each refinement cycle, agreement between calculated structure factor (Fcalc) was compared with the measured structure factor(Fobs) and can be represented as:

$$R = (\Sigma(Fobs - Fcalc)/(\Sigma Fobs)$$

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

2.2.8 Cross-validation

Model quality was assessed by statistical cross-validation by using R-factor (Brünger, 1992). For cross-validation, before refinement diffraction data are divided into two sets where 5-10% of the diffraction data are removed from the data set as a small

complementary test set. While the remaining 90-95% of the data are used as a large working set and used in the refinement process. Remaining 10% test data set is used to assess the accuracy of model prediction by calculating the R-free value:

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

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

2.2.9 Structure validation

Reliability of three-dimension structure of SBPs were evaluated by structure validation process. In general, validation has three aspects: 1) validity of the experiment measurements 2) consistency of the atomic model and 3) consistency in the known physical and chemical properties of the atomic model. For SBP structure validation following aspect has been considered: (1) geometrical parameter and (2) root mean square deviation (RMSD) of the refined structure. A number of online tools are available for structure validation, out of which PROCHECK and MolProbity was used for the SBP structures.

PROCHECK

After every refinement cycle, the stereochemical quality of the model structure was evaluated by PROCHECK (Laskowski et al., 1993). This program checks the residue-by-residue geometry of the structure. It provides the following information's: Ramachandran plot, Gly & Pro Ramachandran plots, Chi1-Chi2 plots, main-chain parameters, side-chain parameters, residue properties, main-chain bond length distributions, main-chain bond angle distributions, RMS distances from planarity and distorted geometry plots. Among these outputs, the Ramachandran plot was used to used estimates the geometry of the SBPs.

MolProbity

MolProbity is also a web service for structure validation (Davis et al., 2007). It evaluates the three-dimensional structures of the proteins at both the global and local levels. For a given crystal structure, it optimized both polar and nonpolar hydrogen, all-atom contact analysis, covalent-geometry and torsion-angles. In addition, it also improves the local

chemical environment due to the resolution of structures, which includes Ramachandran outliers, flipped branched protein side chains and incorrect sugar puckers. In this study, this program was used to detects the Ramachandran outliers and torsion-angles outliers.

2.2.10 Sequence-and structure-based analysis

Once the structure of target SBP was solved, further functional characterization was performed by using sequence-and structure-based analysis approach as described below.

2.2.10.1 Retrieval of sequence-based information

The nucleotide sequences of each ORFs encoding for SBPs were retrieved from KEGG database (Kanehisa and Goto, 2000) and Gene database of National Centre for Biotechnology (NCBI) (Benson et al., 2000). In addition, operonic information of the ABC transport system for the genetic arrangement of each subunit was also obtained from the Gene database of NCBI (https://www.ncbi.nlm.nih.gov/gene). Similarly, respective protein and homolog sequences analyzed in this study were collected from the protein database i.e., UniProtKB (UniProt Consortium, 2019). To get the information of a total number of ORFs encoding for ABC transporters in *T. thermophilus* HB8 following database has been used for data collection:

- TransportDB 2.0 (Elbourne et al., 2017)
- Structural-Biological Whole-Cell Project (http://www.thermus.org) databases.
- UniProtKB database (The UniProt Consortium, 2019).

2.2.10.2 Sequence-based analysis

Homolog identification and sequence alignment

The homologous proteins of each target SBP were identified by performing search analysis against the non-redundant (nr), UniProt and Protein Data Bank (PDB) database via programs BLAST (Basic Local Alignment Search Tool, Altschul et al., 1990). During protein BLAST, following parameters were used: max target sequence, 100; expected threshold, 10; word size, 6; matrix, BLOSUM62 and gap costs, Existence:11 Extension:1.

After BLAST search, depending upon sequence identity with query coverage sequences were shortlisted and subjected to multiple sequence alignments (MSA) using program Clustal Omega (Sievers and Higgins, 2014). MSA was performed with default parameter, where Clustal Omega uses a combination of seeded guide trees and hidden Markov model (HMM) iterations to generate alignments between the multiple sequences. For the clear representation, output result of MSA was further rendered using the online tool ESPript (Easy Sequencing in PostScript, Gouet et al. 2003).

Phylogenetic tree

A phylogenetic tree is used to find the evolutionary relationships among homolog SBP sequences and to deduces their common ancestors. In this thesis, phylogenetic trees were built from protein sequences which were retrieved from the UniProtKB database. To generate the phylogenetic tree of SBPs, Molecular Evolution and Genetic Analysis (MEGA) software with version 7.0 was used (Kumar at al., 2016). Phylogenetic tree was built in following three steps: In first step, acquire homologous protein sequences were aligned with default parameter (gap opening: 10 and gap extension penalties: 0.2) by Clustal W embedded in the MEGA software. In the second step, a tree was built from the aligned sequences using the neighbor joining (NJ) method. Significance of tree was estimated by bootstrapping method, which defines the confidence level of similarity between the clusters. In the final step, generated was further rendered to convey the clear clustering and relevant information about the evolutionary relationship.

2.2.10.3 Structure-based analysis

Structure prediction

In homology modeling, the theoretical tertiary structure of target protein was constructed from its amino acid sequence and by using a crystal structure of a homologous protein as a template. In order to predict the three-dimension model structure of target SBP online programs RaptorX (Kallberg et al., 2012), SWISS-MODEL (Biasini et al., 2014) and I-TASSER (Yang et al., 2015a) was used, where all server identified the closest homolog as the best template and predicted the model structures. Obtained model structures were

further refined for energy minimization using the program ModRefiner (Xu and Zhang, 2011). Refined and predicted structures were compared together based on their geometrical and stereochemical parameters in the Ramachandran Plot. Based on the obtained best geometrical and stereochemical parameters, the final model was selected and used for further structural analysis.

Structural comparison with homologous proteins

To procure the structural homologs from the Protein Data Bank (PDB) (Berman et al., 2000), a web server DALI (Holm and Rosenström, 2010) was used. Calculation of root mean square deviation (RMSD) between two structures was done using the programs 3dSS (3-dimensional Structural Superposition, Sumathi et al., 2006) and PyMOL (The PyMOL Molecular Graphics System, Schrodinger, LLC). To identify the signature motifs and crucial structural elements, structure-based sequence alignment was performed using program PROMALS3D (PROfile Multiple Alignment with predicted Local Structures and 3D constraints), which aligns multiple three-dimensional structures using a scoring function of BLOSUM62 and performs profile-profile alignments with secondary structures using the hidden Markov model (HMM). Akin to sequences alignment result, the output result of PROMALS3D was also further decorated using the online tool ESPript (Easy Sequencing in PostScript, Gouet et al. 2003).

Structure-based classification of SBPs

A wealth of structural information of SBPs are available in Protein Data Bank (PDB) (Berman et al., 2000). However, despite sequence similarity, SBPs are known to share structural similarity and thus classified into six different clusters. To identify the cluster in which target SBP will belong, structural based classification of SBP was performed. Owing to structural homology with cluster D SBPs, a total of 30 SBPs are shortlisted from different subcluster, which are classified in the previous report (Scheepers et al., 2016). SBPs are shortlisted based on different bacterial genus name. After data retrieval, resulting 30 PDBs along with target SBP were pairwise superimposed with default parameter by using the program PDBeFold (Krissinel and Henrick K, 2019). Resulting RMSD values were further used to generate the distance matrix. Subsequently, this distance matrix with

a suitable input format was submitted into the program dendroUPGMA (GarciaVallvé et al., 1999) to convert the distance matrix into the Newick file format. Obtained Newick file was used to generate the final structural distance tree using online tool iTOL (Letunic and Bork, 2016). Generated tree grouped the target SBP into their respective subcluster, which were cross-verified for distinguishing structural features via manual superimposition in PyMOL (The PyMOL Molecular Graphics System, Schrodinger, LLC).

Surface analysis

In order to identify the nature of active according to cognate ligand surface electrostatic potential charge distribution of the SBPs were calculated using the program Adaptive Poisson-Boltzmann Solver (APBS) embedded in PyMOL (Baker et al., 2001). For the clear representation calculated electrostatic potential charge distribution of each SBPs were color-coded as blue for positive charge with a scale range +1 to +10 kcal mol^{-1}. Similarly, negative charge is also color-coded as red with a scale range of -1 to -10 kcal mol^{-1}. In addition to charge distribution, the surface analysis allowed us to calculate the active site area(s) and volume(s) using the program CASTp with a default probe radius of 1.4 Å (Binkowski et al., 2003).

Domain moment analysis of SBPs

To elucidate conformational changes in the N-terminal and C-terminal domain (NTD and CTD) of SBPs upon ligand binding online program DynDom was used (Hayward and Berendsen, 1998; Taylor et al., 2014). DynDom program determines the rotational angle of domain movement as well as the translation movement of the domain. In addition, it also provides the hinge axes and hinge bending residues information between two conformations of SBPs (open and closed).

Structure and interaction visualization

To visualize all three-dimensional structures of proteins, to locate active sites by comparing known and modeled structures as well as to generate the molecular graphic figures the program PyMOL (The PyMOL Molecular Graphics System, Schrodinger, LLC) was used. For ligand modeling, in place of endogenously-bound ligand program Coot was used. In

addition to identify the inter-molecular interaction between cognate ligand, docked ligand as well modeled ligand and SBPs are analyzed via program Coot and PoseView (Stierand et al., 2006).

Molecular docking

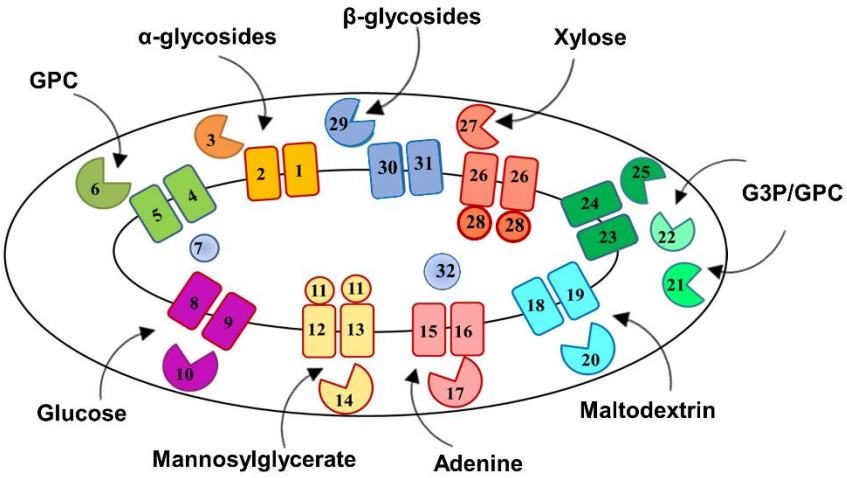
In this study, to predict the binding mode of ligand, ligand selectivity binding affinity of a ligand with SBP via *in silico* methods, molecular docking was performed using the program Autodock version 4.0 (Morris et al., 2009). During the binding experiment, flexible molecular docking was performed, where ligands were kept as flexible with all possible torsional angle, while target protein molecules were remains in rigid conformation. For the molecular docking experiment, the three-dimensional atomic coordinates of proteins and sugar molecules were downloaded from the Protein Data Bank (PDB) (Berman et al., 2000) and PubChem database (Kim et al., 2015), respectively. Before docking, hydrogen atoms were added to protein and ligand molecules as well as partial charges were assigned to both protein and ligand molecules using the Kollman charges and Gasteiger charges, respectively. Each molecular docking experiment was performed with a default set of parameters (i.e., 150 populations, 27000 maximum generations, 0.02 mutation rate, etc.) using the Genetic Algorithm (GA) with 2000 runs to explores the all possible ligand conformation. The docked ligand conformations were clustered with an RMSD of 2.0 Å and subsequently analyzed for the conformation having the lowest binding energy to consider as the potential ligand.

This chapter has been published as:

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

ABSTRACT

Organisms use a variety of carbohydrates and metabolic pathways in order to capitalize in their specific environments. Depending upon their habitat, organism employs different types of transporters to maintain the cellular nutritional balance via central metabolism. A major contributor in this process in bacteria is a carbohydrate ABC transporter. The focus of this study is to get an insight into the carbohydrate transport and metabolism of a hot-spring-dwelling bacterium *Thermus thermophilus* HB8. We applied high-throughput data-mining approaches for identification and characterization of carbohydrate ABC transporters in *T. thermophilus* HB8. This enabled the identification of 11 putative carbohydrate ABC transport systems. To identify the cognate ligands for these transporters, functional annotation was performed. However, scarcity of homologous-protein's function hinder the process of functional annotation. Thus, to overcome this limitation, we integrated the functional annotation of carbohydrate transporters with their metabolic analysis. Our results demonstrate that out of 11 putative carbohydrate ABC transporters, six are involved in the sugar (four for monosaccharides and polysaccharides-degraded products and two for osmotic regulation), four in phospholipid precursor (namely UgpABCE) and the remaining one in purine uptake. Further, analysis suggests the existence of sharing mechanism of transmembrane domains (TMDs) and/or nucleotide-binding domains (NBDs) amongst the 11 carbohydrate ABC transporters.



1. TTHA0354 2. TTHA0355 3. TTHA0356	8. TTHA0685 9. TTHA0686 10. TTHA0688	15. TTHA1301 16. TTHA1302 17. TTHA1304	22. TTHA1936 23. TTHV032 24. TTHV033 25. TTHV034	28. TTHV087 29. TTHB082 30. TTHB083 31. TTHB085
4. TTHA0377 5. TTHA0378 6. TTHA0379	11. TTHA0975 12. TTHA0976 13. TTHA0977 14. TTHA0979	18. TTHA1650 19. TTHA1651 20. TTHA1652	26. TTHV088 26. TTHV088 27. TTHV089	32. TTHA1099
7. TTHA0579		21. TTHA1877		

3.1 INTRODUCTION

Carbohydrates (or sugars) are widespread in nature and are utilized by all domains of life (Toukach and Egorova, 2015). It has multiple roles such as osmoprotection, intracellular signaling and source of carbon and energy in different organisms (Iturriaga and Suarez, 2009). In order to perform these various metabolic activities, cells uptake sugars from extracellular environment with the aid of three major classes of transporters viz. (1) primary active transporters, (2) secondary transporters and (3) group translocators, exclusive in bacteria (Saier, 2000a). The primary active transporters include the largest superfamily of transporters known as the ATP-binding cassette (ABC) transporters (Transporter Classification ID or TCID: 3.A.1), which utilizes the energy produced by ATP hydrolysis as the driving force for nutrient transport (Saurin et al., 1999; Saier, 2000b). ABC transporters are classified into exporters and importers, depending upon the direction of the substrate transport. Architecturally, ABC exporters comprise of a transmembrane domain (TMD) for substrate translocation and a nucleotide-binding domain (NBD) for the ATP hydrolysis process, whereas ABC importers contain an additional substrate (or solute)-binding protein (SBP) for substrate recognition. Among these subunits, SBPs

confer substrate affinity, specificity and directionality of ABC transporters (Maqbool et al., 2015). The primary function of SBPs is to sequester substrates from an extracellular environment and subsequently deliver them to TMDs for their transport (Davidson et al., 2008). ABC transporters employ 45 families for the uptake of different nutrients, of which, carbohydrate uptake transporters 1 and 2 (referred to as CUT1 and CUT2) are known to be specific to carbohydrate transport (Saier, 2000a). Members of the CUT1 family are specific to oligosaccharides, polyols, glycerol and glycerol-phosphate transport while those of the CUT2 family transport only monosaccharides (Schneider, 2001).

Depending upon the varied habitats (e.g. soil, water, animal digestive tracts, etc.), microbes possess a diverse range of carbohydrate ABC transporters and metabolic pathways (Rodionov et al., 2013; Bräsen et al., 2014). *Thermus thermophilus* HB8, a thermophilic gram-negative bacterium dwelling in hot springs, utilizes a limited number of carbohydrates as carbon and energy source. Unlike several other bacteria, it harbors only major facilitator superfamily (MFS, a class of secondary transporters) and ABC transporter superfamily (a class of primary active transporters) for carbohydrate uptake (Elbourne et al., 2017). The bacterium *T. thermophilus* HB8 consists of 127 open reading frames (ORFs) encoding ABC transporter subunits. Out of which, a total of 35 ORFs have been annotated to encode 11 carbohydrate ABC transport systems including 11 SBP, 19 TMD and 5 NBD subunits. Quite interestingly, 9 out of 11 carbohydrate ABC transport systems lack their corresponding NBD subunits (Elbourne et al., 2017). In many gram-positive bacteria, the operons encoding ABC transporters of the CUT1 family are known to be devoid of NBD subunits. In addition, the genes encoding NBD components of carbohydrate ABC transporters are observed to be located at a distant position on the genome (Ferreira and Sa-Nogueira, 2010; Marion et al., 2011; Wuttge et al., 2012). Earlier studies have suggested a sharing mechanism of the NBD subunits among several carbohydrate ABC transporters to compensate for their low number (Eitinger et al., 2011).

Despite having identified the subunits of carbohydrate ABC transporters in *T. thermophilus* HB8, their function remains underexplored. The usual method of sequence-similarity-based characterization of ABC transporters to assign their cognate ligands is often

inadequate. Thus, in addition to the sequence-based analysis, integrating the information of genetic contents, their organization and regulation as well as metabolomic analysis are known to significantly enhance the preciseness of gene annotation (Rodionov et al., 2013). Detailed information regarding carbohydrate transport and metabolism has been made available from many model organisms such as *Thermotoga maritima* (Conners et al., 2005). However, projection of these information for the study of carbohydrate transport in distantly related bacteria is difficult due to different habitats. Thus, it becomes essential to perform a functional annotation of carbohydrate transporters considering the bacterium-specific carbohydrate metabolism.

In this study, we have attempted to assign the cognate substrate to each carbohydrate ABC transporter as well as correlate their function with the carbohydrate metabolic network in *T. thermophilus* HB8. To fulfill this purpose, a total of 35 ORFs encoding the 11 putative carbohydrate ABC transport systems and 106 ORFs encoding proteins involved in carbohydrate metabolism have been investigated. Since the UgpABCE (uptake glycerol phosphate) transporters are the members of the CUT1 family and are also the closest homolog of the sugar ABC transporters, they often get misannotated as sugar ABC transporters. Thus, we have also performed functional annotation of the UgpABCE transporters. In summary, we present the functional characterization of six sugar, four Ugp and one purine ABC transporter(s) utilizing the approaches of sequence-, structure-, genetic- and metabolic-based analysis.

3.2 MATERIALS AND METHODS

3.2.1 Data collection

The ORFs coding for sugar ABC transporters were collected from TransportDB 2.0 (Elbourne et al., 2017) and Structural-Biological Whole Cell Project (http://www.thermus.org) databases and published literature (Ohtani et al., 2012). The annotations and protein sequences of the identified ORFs were retrieved from UniProtKB database (The UniProt Consortium, 2019). The genetic arrangement of each ORF and their functionally-associated genes were obtained from the Gene database of National Center for

Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/gene). To functionally annotate each carbohydrate ABC transport system of *T. thermophilus* HB8, a total of 11 ORFs encoding SBP subunits were considered.

3.2.2 Sequence analysis

The homologous proteins for each SBP were identified using the web tool BLAST (Altschul et al., 1990). The phylogenetic evolutionary tree of SBPs along with their homologs were performed using neighbor joining (NJ) method with 1000 bootstrap replicates embedded in the program MEGA7 (Kumar et al., 2016). To obtain the conserved regions for further validation of the function of SBPs, multiple sequence alignment (MSA) was performed using the program Clustal Omega (Sievers and Higgins, 2014). The MSAs were further rendered using the web tool ESPript 3.0 (Gouet et al., 2003) for the clarity of the sequence alignments. The metabolomic profiling of carbohydrate molecules was generated using an integrated network of their transport and metabolism by retrieving genes from KEGG database (Kanehisa and Goto, 2000). Further, to identify the NBD and TMD subunit(s) of each carbohydrate ABC transporter, protein-protein interaction analysis was performed using the database STRING v10.5 (Szklarczyk et al., 2017) with a default set of parameters.

3.2.3 Structure analysis

Theoretical tertiary structure of each SBP was predicted using the programs RaptorX (Kallberg et al., 2012), SWISS-MODEL (Biasini et al., 2014) and I-TASSER (Yang et al., 2015a). The predicted models were further refined by energy minimization using the program ModRefiner (Xu and Zhang, 2011). Subsequently, the predicted and refined models were validated by calculating the geometrical and stereo-chemical parameters using the Ramachandran Plot available in the program PDBsum (De Beer et al., 2014). The models having the best geometrical and stereo-chemical parameters were chosen for further analysis. The structural homologs of each SBP were searched using the web server Dali (Holm and Rosenstrom, 2010). The surface area of active-site pocket of each SBP was computed using the program CASTp 3.0 (Tian et al., 2018) with a default probe radius of 1.4 Å. To estimate the binding affinities and selectivity of sugar molecules towards SBPs,

the program AutoDock version 4.0 was used (Morris et al., 2008). For this, the three-dimensional atomic coordinates of sugar molecules were retrieved from the PubChem database (Kim et al., 2015). Each molecular docking experiment was performed using Genetic Algorithm (GA) search method for 2000 runs and with default set of parameters (i.e. 150 populations, 27000 maximum generations, 0.02 mutation rate, etc.) and the docked ligand conformations were clustered with a root mean square deviation (rmsd) of 2.0 Å. The interaction analysis of the docked sugar molecules to SBPs was performed using the programs Coot (Emsley et al., 2010) and PoseView (Stierand et al., 2006). The program PyMOL (The PyMOL Molecular Graphics System, Schrodinger, LLC) was used to visualize protein structures, to locate active sites by comparing different known and modeled structures and to generate figures.

3.3 RESULTS AND DISCUSSION

3.3.1 Repertoire of carbohydrate uptake ABC transporters in *Thermus thermophilus* HB8

To identify the list and diversity of ABC transporters for carbohydrate uptake in *T. thermophilus* HB8, we collected all the ORFs annotated as “sugar ABC transporters" in TransportDB 2.0 (Elbourne et al., 2017) and Structural-Biological Whole Cell Project (http://www.thermus.org) databases. The thermophilic bacterium *T. thermophilus* HB8 is a polyploid organism harboring a chromosome and three plasmids viz. pTT27, pTT8 and pVV8 (Ohtani et al., 2012). As ORFs from pVV8 is not present in databases, details of each ORFs was collected from published literature (Ohtani et al., 2012). Altogether, a total of 35 ORFs (28 - sugar and 7 - G3P or Ugp) were identified to encode ABC transporter subunits. Out of which, 25 are located on the chromosome while the remaining 10 ORFs are from the plasmids pTT27 and pVV8 (Table A.1). As ABC transporters are comprised of three components viz. SBP, TMD and NBD, all 35 ORFs were assigned to their respective components (Table A.1); this resulted in 11 (8 - sugar and 3- G3P or Ugp) carbohydrate ABC transport systems. The exact function and type of sugars transported by these 11 carbohydrate ABC transporters are unknown, hence we performed their functional characterization by assigning the cognate sugar to each transporter. Since, SBPs are

specific to their substrates and responsible for maintaining the selectivity of ABC transporters, we focused mainly on these proteins for characterization. In this study, we propose the cognate substrates for all the 11 carbohydrate ABC transporters in the successive sections.

3.3.2 D-xylose ABC transporter

Bacterial cells utilize xylose as a sole carbon source by importing it inside the cell via ABC, MFS and group translocator transporters and metabolizing it through pentose phosphate pathway (Han et al., 2012; Sun et al., 2015). In *T. thermophilus* HB8, the sequence analysis of the plasmid pVV8 revealed the presence of ORFs (TTHV087-TTHV089) encoding D-xylose ABC transporters (Ohtani et al., 2012). In this study, we investigated the mechanism of uptake and metabolism of D-xylose.

The protein TTHV089 is a xylose-binding protein

To understand the xylose uptake mechanism in *T. thermophilus* HB8, functional characterization of the protein TTHV089, which is annotated as "D-xylose ABC transporter periplasmic substrate-binding protein", was performed. A search for homologous proteins against the UniProtKB database reveals that TTHV089 has a significant similarity with xylose-, multisugar- and sugar alcohol-binding proteins with a sequence identity/query coverage of 40/92%, 32/90% and 32/57%, respectively. Phylogenetic analysis of TTHV089 with its homologous proteins exhibits its proximity to xylose-binding proteins (XBPs) (Figure 3.1A). To further substantiate the relationship of TTHV089 with XBPs, its three-dimensional tertiary structure was predicted using various programs. All the programs utilized XBP from *E. coli* (*Ec*XBP, PDB ID: 3M9W, open conformation) as the best template, hinting the protein TTHV089 to be a xylose-binding protein. Further, the structural homology search using the predicted model of TTHV089 shows that apart from xylose-bound proteins, galactose- and glucose-bound proteins also share topological similarity with an average rmsd of ~3.4 Å (Table A.2). To further affirm the similarity and dissimilarity of TTHV089 with xylose-, galactose- and glucose-bound proteins, a closed form of TTHV089 was predicted using the template *Ec*XBP (PDB ID: 3MA0, xylose-bound). Interestingly, unlike the open conformation of TTHV089, its close

conformation shows higher topological similarity with xylose- (PDB ID: 3MA0), galactose- (PDB ID: 3URM) and glucose-bound (PDB ID: 4YS6) proteins with an average rmsd of ~0.7 Å (Figure A.1A). Thus, to identify the characteristic features of xylose-, galactose- and glucose-bound proteins in TTHV089, an MSA of these proteins was performed. The results indicate that all these proteins contain a similar active site (Figure A.1B, A.1C and A.1D), irrespective of their different cognate ligands. Further, the position and orientation of the active-site residues of the protein TTHV089 are significantly similar to that of xylose-, galactose- and glucose-bound proteins (Figure 3.1B), hinting that the protein TTHV089 might bind galactose and glucose, in addition to xylose. However, *Ec*XBP has been reported to be specific to xylose and is hypothesized to avert the binding of hexose sugars due to steric clash created by the residue Leu14 in the active site (Sooriyaarachchi et al., 2010). A comparison of the active site of TTHV089 with that of xylose-, galactose- and glucose-bound proteins shows that the residue Leu27 in glucose-binding protein (PDB ID: 4YS6) retains a similar position to that of Leu14 in *Ec*XBP (Figure 3.1C). As the presence of Leu27 does not affect the binding of glucose (hexose sugar) to the protein, it suggests that *Ec*XBP should also bind to hexose sugars in addition to xylose. Thus, to corroborate this, we performed molecular docking experiments taking *Ec*XBP as the receptor and different sugars (xylose, galactose and glucose) as the ligands. The results suggest that all the three sugars i.e. xylose, galactose and glucose can bind in the active site of the protein *Ec*XBP (Figure 3.1D, 3.1E and 3.1F). Interestingly, the protein-ligand coordination of xylose and glucose with *Ec*XBP are almost identical (Table A.3), wherein the CH_2OH group of glucose interacts with the residue Arg91 (Figure 3.1F and Table A.3).

Similarly, to verify the binding of xylose to the protein TTHV089, we performed molecular docking experiment with xylose as the ligand. The result confirms the binding of xylose to the protein TTHV089 in an identical manner as in *Ec*XBP. The residues Glu12, Arg14, Asp88, Arg89, Asp132, Asn134, Asn193, Asp219 and Lys239 are involved in holding the xylose in the active site of the protein (Figure 3.1G and Table A.3). Interestingly, in contrast to *Ec*XBP and glucose-binding proteins, the position of the residues Leu14 and Leu27, respectively, is occupied by a bulkier residue Glu12 in the protein TTHV089. An

active-site comparison of TTHV089 with that of galactose- and glucose-bound proteins reveals that the distance between the residue Glu12 and the CH_2OH group of hexose sugars (galactose and glucose) are 1.5 Å and 1.2 Å, respectively, suggesting a steric clash (Figure 3.1H and 3.1I). Furthermore, molecular docking performed with hexose sugars shows that both galactose and glucose flip their orientation, most likely due to the presence of the residue Glu12 (Figure 3.1H and 3.1I). Thus, it can be concluded that unlike *Ec*XBP, the protein TTHV089 can bind only xylose.

Our results suggest that *Ec*XBP can bind both xylose (pentose) and glucose (hexose), whereas the protein TTHV089 binds only xylose. However, the binding of both types of sugars (pentose and hexose) to *Ec*XBP raises the question as to why xylose and glucose occupy the same spatial position and interact with same set of residues in the active site of the protein? One of the answers can be inferred from MFS class of xylose transporter viz. XylE. The XylE transporter is xylose-specific, however, it can bind xylose and glucose in an identical position and orientation. Moreover, glucose is known to be an efficient inhibitor for XylE suggesting its competitive binding to the protein (Sun et al., 2012). Thus, it can be proposed that ABC, like MFS transporters, follow a similar regulatory mechanism for the transport of pentose and hexose sugars. Hence, to understand the requirement and metabolism of xylose in *E. coli* and *T. thermophilus* HB8, we identified the existence of an operon for xylose transport and its metabolism via pentose phosphate pathway. Although both the bacteria have a similar set of genes for xylose transport and metabolism, *E. coli* has an additional gene *malS* encoding a periplasmic α-amylase, known to hydrolyze long-chain maltodextrins producing hexose (e.g. glucose) sugars (Figure 3.1J) (Freundlieb and Boos, 1986). In *E. coli*, presence of gene *malS* adjacent to the xylose ABC transporter indicates a glucose-dependent regulatory mechanism for xylose transport like XylE. On the other hand, the absence of the gene *malS* in *T. thermophilus* HB8 indicates the inability of the protein TTHV089 to uptake hexose sugars; this corroborates the docking results. In summary, in *T. thermophilus* HB8, the uptake of xylose is mediated by TTHV087-TTHV089 ABC transporter and is further metabolized by the pentose phosphate pathway (Figure A.2).

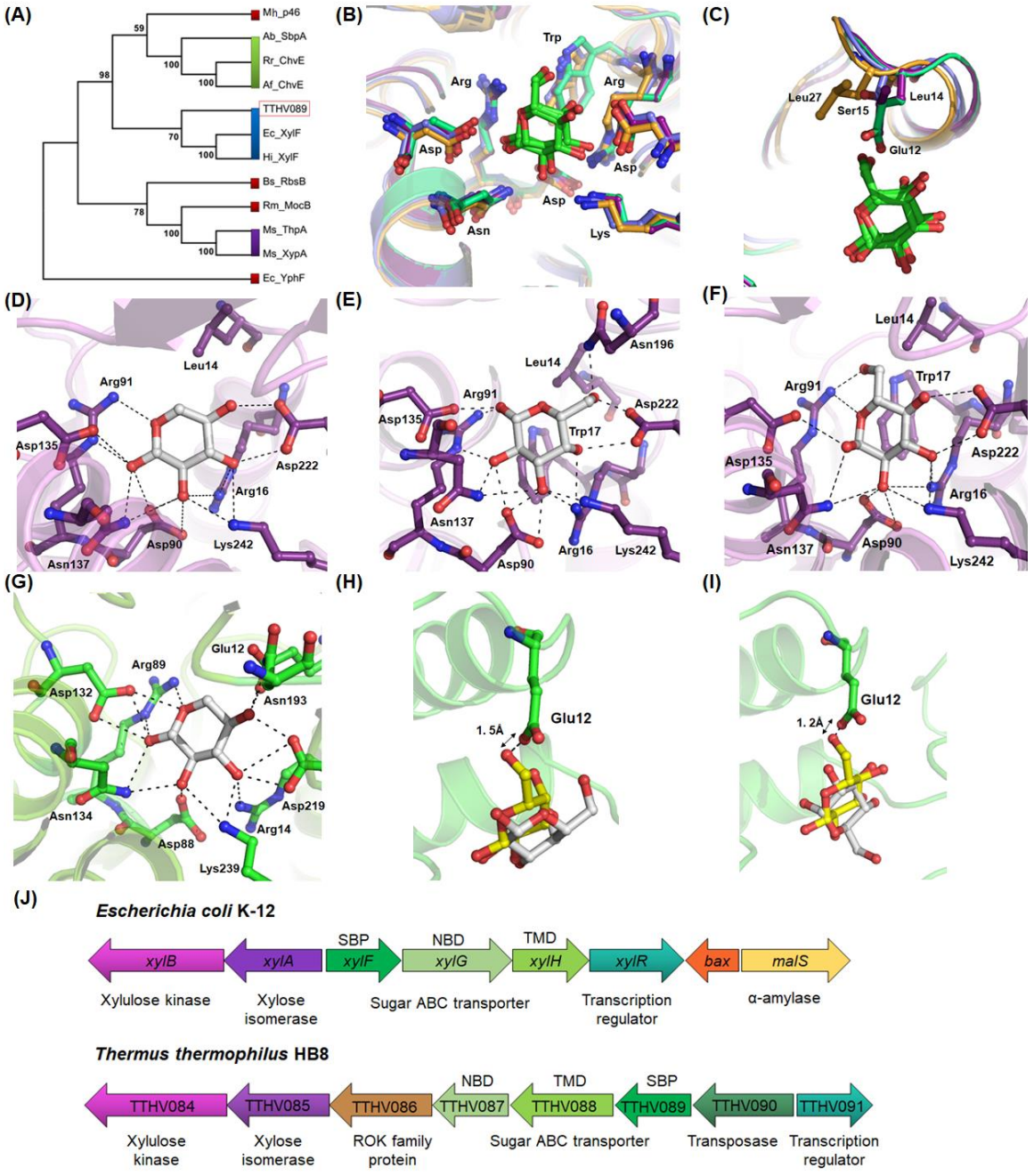


Figure 3.1. Representation of phylogenetic, active site and genetic organization of protein TTHV089. (A) Phylogenetic tree of TTHV089 along with its homologous proteins. The protein sequences used to build the phylogenetic tree are TTHV089, *T. thermophilus* HB8 (G9MBE1); Mh_p46, *Mycoplasma hyopneumoniae* (P0C0J7); Ab_SbpA, *Azospirillum brasilense* (P54083); Rr_ChvE, *Rhizobium radiobacter* (P54082); Af_ChvE, *Agrobacterium fabrum* (P25548); Ec_XylF, *E. coli* (P37387); Hi_XylF, *Haemophilus influenzae* (P45047); Bs_RbsB, *Bacillus subtilis* (P36949); Rm_MocB, *Rhizobium meliloti* (P49308); Ms_ThpA, *Mycobacterium smegmatis* (A0QYB5); Ms_XypA, *M. smegmatis* (A0QYB3) and Ec_YphF, *E. coli* (P77269). The UniProt ID for

each protein are provided in parenthesis. The analysis suggests the grouping of these proteins into three distinct clusters shown in blue (xylose-), green (multisugar-) and violet (sugar alcohol-binding proteins). Proteins, which do not belong to any cluster are marked in red. The protein TTHV089 in the cluster of xylose-binding proteins is shown in red box. (B) Structural superimposition of the active site of TTHV089 (green), *Ec*XBP (PDB ID: 3MA0, violet), galactose- (PDB ID: 3URM, blue) and glucose-binding (PDB ID: 4YS6, orange) proteins. The amino acid residues involved in the sugar binding are shown in ball-and-stick model and rest of the protein is displayed in cartoon. (C) Active site structural difference among TTHV089, xylose-, galactose- and glucose-bound proteins with respect to the position of Glu12, Leu14, Ser15 and Leu27 shown in ball-and stick-model. (D, E, F) Schematic representation of binding of *Ec*XBP with xylose, galactose and glucose, respectively. Residues involved in interaction are shown in violet ball-and-stick model and rest of the protein in cartoon. Hydrogen bonds established between the residues and OH group of sugar are depicted with dotted lines. (G, H, I) Xylose, galactose and glucose binding to the protein TTHV089. The amino acid residues involved hydrogen bonding (dotted line) with the sugar molecule are labeled and shown in green ball-and-stick model. The residue Glu12 (green) in TTHV089 causes the flipping of galactose and glucose (grey) in the active site with respect to galactose and glucose (yellow) of galactose- and glucose-binding proteins, respectively. (J) Genetic organization of genes associated with xylose transport and metabolism in *E. coli* K-12 and *T. thermophilus* HB8. Each gene is represented by an arrow (indicating its direction of transcription), ORF number and respective encoded proteins. The respective genes in both the bacteria are shown as ABC transporter components (green), xylose isomerase (violet), xylulose kinase (magenta), transcription regulator (blue) and α-amylase (yellow).

3.3.3 Trehalose/Maltose ABC transporter

Trehalose, a disaccharide, is used for multiple cellular functions such as carbon and energy source, osmoprotectant and in cell-wall synthesis (Iturriaga et al., 2009). The *Thermus* spp. utilize trehalose as an osmoprotectant for their growth in a high-salt environment (Alarico et al., 2005). Uptake of trehalose in (hyper)thermophilic bacteria is mediated by trehalose/maltose ABC transporter (Silva et al., 2005). However, in *T. thermophilus* HB8, none out of 11 carbohydrate ABC transporters has been annotated as trehalose/maltose ABC transporter. Here, we identified the genes encoding trehalose/maltose ABC transporter in *T. thermophilus* HB8.

The protein TTHA0356 is a trehalose-binding protein

In *T. thermophilus* HB27, uptake of trehalose has been reported to be mediated by trehalose/maltose-binding protein (ORF ID: TTC1627), an SBP subunit of ABC

transporter (ORF IDs: TTC1627-TTC1629) (Silva et al., 2005), which shares a 100% sequence identity and query coverage with the protein TTHA0356 of *T. thermophilus* HB8, suggesting it to be a trehalose/maltose-binding protein. However, a search for homologous proteins of TTHA0356 against the UniProtKB database produces di-, tri- and tetra-saccharide-binding proteins with the sequence identities in the range of 23-39% (query coverage: 57-89%). Further, a phylogenetic analysis of these proteins groups into three different clusters viz. di-, tri- and tetra-saccharide-binding proteins; where the protein TTHA0356 clusters with the disaccharide-binding protein (Figure 3.2A). To further affirm its cognate substrate, tertiary structure of the protein TTHA0356 was predicted; expectedly, all the programs utilized trehalose/maltose-binding protein (TMBP) from *Thermococcus litoralis* (PDB ID: 1EU8) as the default template. Also, it suggests the preference of disaccharides over tri- and tetra-saccharides to the protein TTHA0356. Furthermore, a comparison of the protein sequences of TTHA0356 and TMBP shows that, except Thr44 and Tyr121, all the residues Glu17, Thr46, Arg49, Asp70, Asp123, Glu239, Gly294, Trp295 and Arg364 coordinating the trehalose molecule in the active site of TMBP, are conserved in TTHA0356 as well (Figure 3.2B). Superimposition of the active site of TMBP and TTHA0356 reveals that, except the residue Glu17, all other residues holding trehalose occupy the same position and orientation (Figure 3.2C). Altogether, these results indicate that the protein TTHA0356 is a disaccharide-binding protein and can bind trehalose.

Although the protein TTHA0356 seems to bind trehalose, its closest homolog TMBPs (ORF IDs: TTC1627 and OCC03562 (PDB ID: 1EU8) from *T. thermophilus* HB27 and *T. litoralis*, respectively) show higher binding affinity towards glucose and maltose also (Silva et al., 2005; Diez et al., 2001; Herman et al., 2006; Herman et al., 2007; Povarova et al., 2010; Fonin et al., 2014). Further, it is suggested that two loops (L1 and L2) and three helices (H1, H2 and H3) of sugar SBPs determine the substrate types which can be accommodated in the active site of the protein (Cuneo et al., 2009a). Thus, to investigate whether the protein TTHA0356 has a similar property, we compared the active-site pocket of TTHA0356 along with its structural homologs identified via Dali search (Table A.2). An analysis of four subsites (A-D) of TTHA0356 active site indicates the occupancy of subsite C by the residue Glu174 from helix H3; and thus the availability of only two

subsites (A-B) hints the binding of either mono- (e.g. glucose) or di-saccharides (e.g. trehalose and maltose) (Figure 3.2D). To further substantiate the hypothesis, molecular docking experiments of the protein TTHA0356 with mono- (glucose) and di-saccharide (trehalose and maltose) sugars were performed. The glucose molecule can bind the protein and occupy subsite either A or B with binding affinity of -5.66 and -6.09 kcal mol^{-1}, respectively (Figure A.3A and Table A.3), whereas trehalose and maltose both occupy subsites A-B in the active site of the protein (Table A.3). Interestingly, maltose shows higher binding affinity (-10.22 kcal mol^{-1}) compared to trehalose (-8.47 kcal mol^{-1}) coordinated by a large number of hydrogen bonds (Figure A.3B and Table A.3). An earlier study using molecular dynamics simulation showed the flipping of two glucose units (Glc1 and Glc2) of trehalose into a "maltose-like" configuration of the TMBP (Herman et al., 2007). To investigate a similar process in TTHA0356, we compared the two glucose units of trehalose and maltose docked in subsites A and B. Indeed, the two glucose units prefer a "maltose-like" orientation rather than "trehalose-like" signifying the preference of maltose over trehalose to the protein TTHA0356 (Figure A.3C).

As mentioned, in thermophiles, trehalose is predominantly taken up inside the cell through trehalose/maltose ABC transporter (Silva et al., 2005). However, specific reasons for sugars like glucose, trehalose and maltose being imported by the same ABC transporter are not well known. Thus, to unriddle this enigma, we looked into the genetic mapping of trehalose/maltose ABC transporters and genes involved in trehalose synthesis. Trehalose biosynthesis pathway involves two enzymes namely trehalose-phosphate synthase (TPS) and trehalose-6-phosphate phosphatase (TPP), encoded by genes *otsA* and *otsB*, respectively (Giaever et al., 1988). These two enzymes synthesize trehalose from glucose-6-phosphate, which is generated during glycolysis or by interconversion of glucose-1-phosphate by the phosphoglucomutase (PGM) enzyme (Li et al., 2010). On the other hand, trehalose can also be synthesized from maltose by the enzyme trehalose synthase (TreS), encoded by the gene *treS* (Koh et al., 1998). Out of these, two genes *otsB* and *treS* have already been identified in *T. thermophilus* HB8 (Alarico et al., 2005); presence of which suggests the synthesis of trehalose by both the pathways (Figure A.4). An earlier study reported that in trehalose-deficient medium, *T. thermophilus* RQ-1 cells maintain the

intracellular trehalose level by importing it from extracellular environment or by synthesizing it from glucose and maltose via *otsA*/*otsB* and *treS* genes, respectively (Silva et al., 2003). This observation suggests that transport of glucose and maltose is essential to maintain the intracellular levels of trehalose. Furthermore, in *T. thermophilus* HB8, the gene *pgm* is located upstream to the trehalose/maltose ABC transporter genes (Figure 3.2E). In *T. litoralis* the enzyme trehalose glycosyl transferring synthase (TreT), which catalyzes the breakdown of trehalose, is located upstream to the trehalose/maltose ABC transporter genes (Figure 3.2E) suggesting that the bacterium uses trehalose as a carbon source and not as an osmolyte like *T. thermophilus* HB8, which lacks the enzyme TreT (Qu et al., 2004). Altogether, these observations demonstrate the reason for the transport of glucose and maltose along with the trehalose by trehalose/maltose ABC transporter.

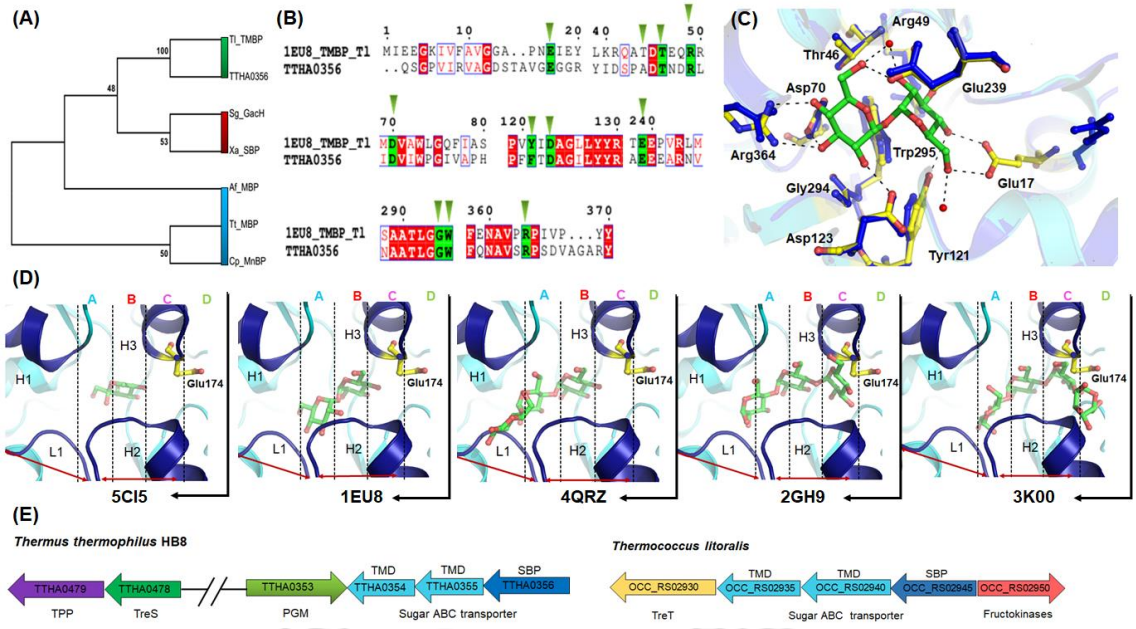


Figure 3.2. Schematic representation of phylogenetic, active site and genetic organization of protein TTHA0356. (A) Phylogenetic analysis of di- (green), tri- (blue) and tetra-saccharide-binding (red) proteins along with TTHA0356. The protein sequences used for the analysis are TTHA0356, *T. thermophilus* HB8 (Q5SLD7); Tl_TMBP, *Thermococcus litoralis* (Q7LYW7); Sg_GacH, *Streptomyces glaucescens* (B0B0V1); Xa_SBP, *Xanthomonas axonopodis* pv. *Citri* (Q8PK66); Af_MBP, *Agrobacterium fabrum* (A9CGI0); Tt_MBP, *T. thermophilus* HB27 (Q72I44) and Cp_MnBP, *Caldanaerobius polysaccharolyticus* (L0E2M2). The UniProt ID of each protein is provided in parenthesis. (B) Pairwise sequence alignment of protein TTHA0356 and TMBP. To get a clear view, only the partial segment of the alignment has been shown here. The conserved residues and

trehalose-interacting residues across the proteins are enclosed in a red and green boxes, respectively. (C) Overlay of active-site pocket of protein TTHA0356 and TMBP. The trehalose molecule and its interacting residues of TMBP are shown as ball-and-stick model in green and yellow, respectively. The position and orientation of the conserved residues of TTHA0356 are depicted as ball-and-stick model in blue. Hydrogen bonds are represented by black-dashed lines and water molecules are shown as red spheres. (D) Schematic representation of four subsites of TTHA0356. The loop (L1) and three helices (H1, H2 and H3) anchoring the subsites are shown in blue; rest of the protein are shown in cyan. Each subsite (A-D) are partitioned by a dotted vertical line. Occupancy of mono-, di-, tri- and tetra-saccharide in four subsites is shown by overlaying TTHA0356 with structural homologs such as D-tagatose- (PDB ID: 5CI5), trehalose- (PDB ID: 1EU8), maltotriose- (PDB ID: 4QRZ and 2GH9) and maltotetraose-binding proteins (PDB ID: 3K00). (E) Genetic organization of genes involved in trehalose transport and metabolism in *T. thermophilus* HB8 and *T. litoralis*. Each gene is represented by an arrow indicating its direction of transcription, ORF number and respective encoded proteins. The ORFs are represented in different colors such as ABC transporter, blue; phosphoglucomutase (PGM), green; trehalose synthase (TreS), dark green; trehalose-6-phosphate phosphatase (TPP), purple; trehalose glycosyltransferring synthase (TreT), yellow and fructokinase, red.

3.3.4 Mannosylglycerate ABC transporter

The protein TTHA0979 is a mannosylglycerate-binding protein

T. thermophilus strains adapt halotolerancy by accumulating compatible solutes (or osmoprotectants) in response to osmotic stress (Alarico et al., 2005); two major osmoprotectants are trehalose and mannosylglycerate (α-D-mannopyranosyl-(1,2)-O-D-glycerate) (Gonçalves et al., 2010). Much like trehalose, mannosylglycerate (MG) is synthesized by two different pathways (Alarico et al., 2007). In one of the pathways, mannosyl-3-phosphoglycerate synthase (MpgS) uses GDP-α-D-mannose and D-glycerate-3-phosphate to form mannosyl-3-phosphoglycerate, which is subsequently hydrolyzed to MG by the enzyme mannosyl-3-phosphoglycerate phosphatase (MpgP). In an alternative pathway, MG can directly be synthesized using GDP-α-mannose and D-glycerate by the enzyme mannosylglycerate synthase (MgS) (Gonçalves et al., 2010). In *T. thermophilus,* trehalose is a principal osmoprotectant and MG is utilized as an alternative during stress conditions (Alarico et al., 2013). Thus, unlike the trehalose, MG molecules are hydrolyzed into mannose and glycerate by the mannosylglycerate hydrolase (MgH) (Miyazaki et al., 2015).

In *T. thermophilus* HB27, mutagenic studies of MG-synthesizing genes showed the restoration of halotolerancy in the presence of externally-supplied MG (Alarico et al., 2013). In this bacterium, the uptake of MG is attained by an ABC-type transport system (ORF IDs: TTC0611–TTC0615), which is found to be functionally associated with MgH (ORF ID: TTC0614) (Figure 3.3A). This functional association of the enzyme MgH with ABC transporter has also been reported in *Meiothermus silvanus* (ORF ID: Mesil_0668) and *Truepera radiovictrix* (ORF ID: Trad_0950) (Figure 3.3B, 3.3C). Similarly, in *T. thermophilus* HB8, the enzyme MgH (ORF ID: TTHA0978) is co-localized with the ABC transporter (ORF IDs: TTHA0975-TTHA0979) (Figure 3.3D). Till date, cognate ligand for this ABC transporter is unknown as the SBP subunit (ORF ID: TTHA0979) being uncharacterized. Thus, to identify the cognate ligand for TTHA0979, a homology search against protein database was performed which shows its similarity with carbohydrate ABC transporter substrate-binding proteins. However, lack of structures of homologous proteins limits the further characterization of the protein TTHA0979. Hence, based on the co-occurrence of genes encoding MgH and ABC transporter, it can be speculated that the protein TTHA0979 can bind MG.

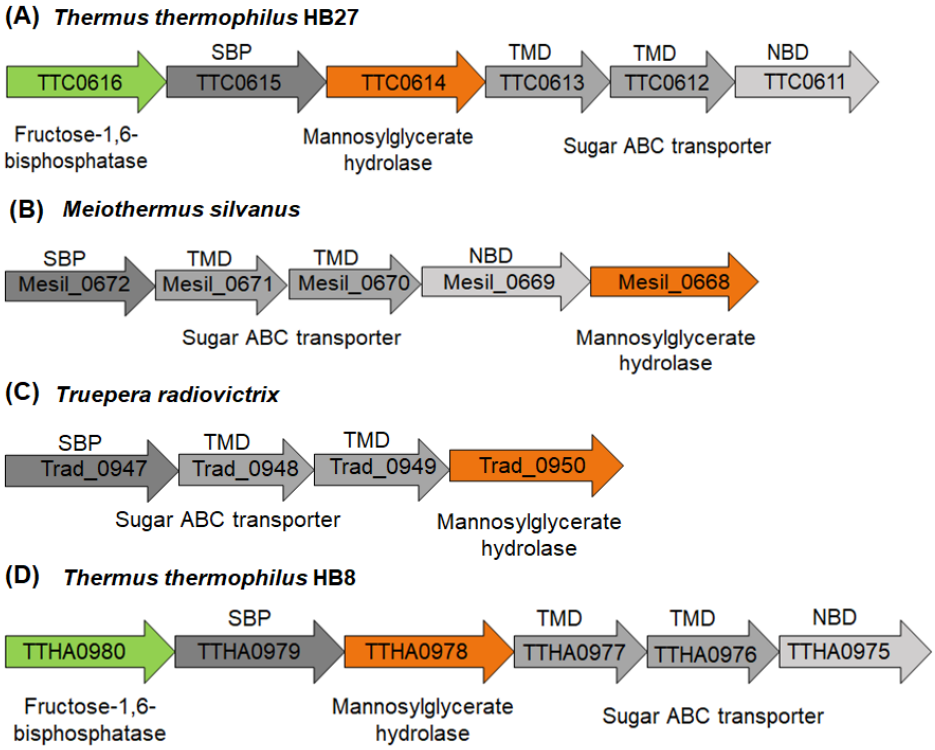


Figure 3.3. Representative operons comprising genes for MG transport and metabolism. A genetic cluster of mannosylglycerate hydrolase (MgH) in conjunction with ABC transporters in (A) *T. thermophilus* HB27, (B) *Meiothermus silvanus*, (C) *Truepera radiovictrix* and (D) *T. thermophilus* HB8. Each gene is represented by an arrow indicating the direction of its transcription. The ORF number and the respective encoded-protein name are also provided. For clarity, the arrows are colored according to their respective function, i.e. ABC transporter, grey; MgH, orange and fructose-1,6-bisphosphatase, green.

3.3.5 Cyclo/Maltodextrin ABC transporter

In *T. thermophilus* HB27, uptake of maltotriose is mediated by a maltose ABC transporter (ORF IDs: TTC1286-TTC1288) (Cuneo et al., 2009a). Genome sequence analysis of *T. thermophilus* HB8 reveals the presence of a similar maltose ABC transporter (ORF IDs: TTHA1650-TTHA1652). In fact, the SBP components (ORF IDs: TTC1288 and TTHA1652) of both the transporters are identical at the primary structure level (sequence identity: 99%, query coverage: 100%). Historically, maltose ABC transporters are known to import maltose and $\alpha_{(1\rightarrow4)}$-linked glucose polymers (up to seven glucose units) such as maltodextrin (Boos and Shuman, 1998). Thus, in this study, we investigated the binding of glucose polymers such as maltodextrin (or cyclodextrin) to the protein TTHA1652.

The protein TTHA1652 can bind cyclo/maltodextrin molecules

In UniProtKB database, the protein TTHA1652 is annotated as "maltose ABC transporter, periplasmic maltose-binding protein" which shares the highest sequence similarity with maltose/maltodextrin-binding protein (ORF ID: TTC1288, PDB ID: 2GH9) from *T. thermophilus* HB27. To infer the probable ligands in addition to maltose, tertiary structure of TTHA1652 was predicted using different programs. Expectedly, all programs used maltose/maltodextrin-binding protein (PDB ID: 2GH9, maltotriose bound and close conformation) as the default template, hinting maltotriose as the cognate substrate for the protein TTHA1652. Topologically, the protein TTHA1652 is homologous to maltose/maltodextrin-binding protein from *T. thermophilus* HB27 (PDB ID: 2GH9, maltotriose bound) and *T. maritima* (PDB IDs: 2GHA and 2FNC, maltotriose bound) and maltodextrin-binding protein from *Thermoactinomyces vulgaris* (*Tvu*CMBP, PDB ID: 2ZYM, cyclodextrin bound) (Table A.2). The protein *Tvu*CMBP binds glucose polymers such as maltodextrin e.g. maltotetraose (G4: 4 glucose units, PDB ID: 2ZYO) and

cyclodextrin (CD) e.g. α-CD (G6: 6 glucose units, PDB ID: 2ZYM), β-CD (G7: 7 glucose units, PDB ID: 2ZYN) and γ-CD (G8: 8 glucose units, PDB ID: 2ZYK) (Tonozuka et al., 2007; Matsumoto et al., 2009). Thus, the protein TTHA1652, which shares a similar topology to that of *Tvu*CMBP, can be speculated to be capable of binding glucose polymers such as cyclo- and malto-dextrin molecules. Furthermore, the residues Leu59, Asp83, Glu129, Tyr175, Asn247, Trp250 and Trp360 of *Tvu*CMBP, which holds the CD molecule in the active site of the protein, are conserved in the protein TTHA1652 as well (Figure 3.4A); suggesting cyclo/maltodextrin as its probable substrate.

Furthermore, to evaluate the binding of cyclo/maltodextrin to the protein TTHA1652, molecular docking experiments with maltose, maltodextrins and cyclodextrin were performed. The results suggest the binding of maltose in subsites either A-B or B-C (Figure A.5A, A.5B) while the smallest maltodextrin i.e. maltotriose and maltotetraose occupy subsites A-C and A-D, respectively (Figure A.5C, A.5D). Interestingly, among all the docked substrates, maltotriose and maltotetraose seems to be the preferred cognate substrates for the protein TTHA1652 (Table A.3). The larger docked molecules are stabilized mostly by stacking forces and few water molecules in the active site of the protein (Figure 3.4B). Thus, it can be suggested that the protein TTHA1652 binds and mediates the transport of maltodextrin molecules preferably up to 4-5 pyranose units.

As *T. thermophilus* HB8 contains cyclo/maltodextrin-metabolizing enzyme maltodextrin glucosidase (MalZ, ORF ID: TTHA1647), which catalyzes the hydrolysis of both maltodextrin and cyclodextrin (only γ-CD and not α- and β-CDs) molecules into glucose and maltose (Boos and Shuman, 1998) and also the homologous protein *Tvu*CMBP transports cyclo/maltodextrin molecules, the transport of γ-CD in addition to maltodextrin can be extrapolated for the protein TTHA1652 as well. To affirm this, we performed molecular docking of TTHA1652 and *Tvu*CMBP (as control) with γ-CD as the ligand. The results demonstrate that like *Tvu*CMBP, the protein TTHA1652 can accommodate γ-CD in its active site with the help of the conserved residues Phe39, Tyr157, Trp235 and Trp345, which stack the G1c2, Glc3 and Glc4 units of γ-CD (Figure 3.4C and Table A.3).

These results were further substantiated by investigating the genetic machineries associated with cyclo/maltodextrin transport and metabolism in *T. thermophilus* HB8, HB27 and *T. maritima* (Figure 3.4D). The presence of functionally-associated cyclo/maltodextrin-metabolizing enzymes upstream to the ABC transport system suggests that in *T. thermophilus* HB8, the import of the cyclo/maltodextrin is performed by the TTHA1650-TTHA1652 ABC transporter for its further metabolism via the enzyme MalZ (ORF ID: TTHA1647). Subsequently, the metabolized products of cyclo/maltodextrins such as glucose and maltose enter the glycolysis pathway for downstream processing (Figure A.6). In *T. thermophilus* HB8, the presence of the enzyme neopullulanase (PDB ID: 2Z1K), which hydrolyses glycosidic linkages (α-1,4 and α-1,6) of pullulan and cyclodextrins (Cheong et al., 2002; Lee et al., 2002), further corroborates our hypothesis.

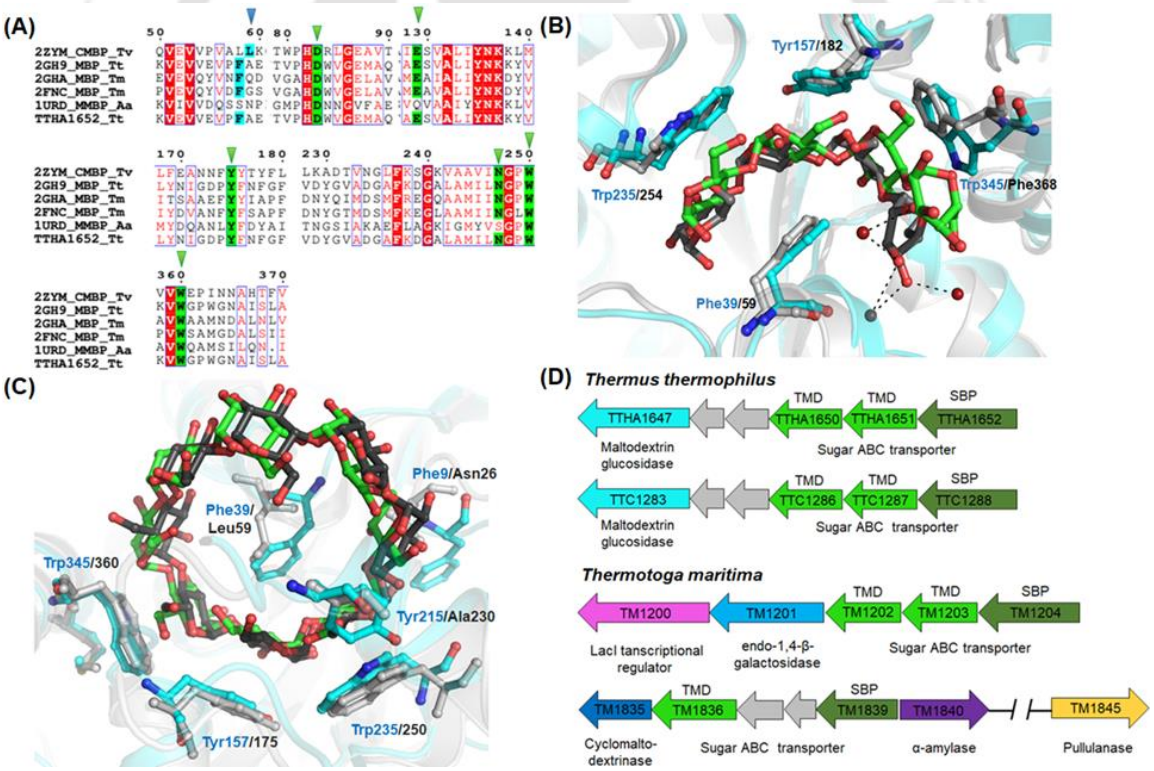


Figure 3.4. Active site and genetic organization of TTHA1652. (A) Multiple sequence alignment of TTHA1652_Tt (*T. thermophilus* HB8) along with 2ZYM_CMBP_Tv (*T. vulgaris*), 2GH9_MBP_Tt (*T. thermophilus* HB27), 2GHA_MBP_Tm (*T. maritima*), 2FNC_MBP_Tm (*T. maritima*) and 1URD_MMBP_Aa (*Alicyclobacillus acidocaldarius*). For the clarity of the figure, only partial alignment is shown. The amino acid residues involved in interaction with G1c2, Glc3 and Glc4 of CD are shaded in green and

highlighted with green downward arrowheads. The hydrophobic residues interacting with the central cavity of CD and maltotetraose are highlighted in cyan and blue downward arrowheads. (B) Superimposition of the active-site pocket of maltotetraose-binding protein (PDB ID: 3K00, grey) and the protein TTHA1652 (cyan). The maltotetraose bound to the crystal structure (PDB ID: 3K00) and that obtained from the molecular docking calculation is presented as ball-and-stick model in black and green, respectively. The aromatic residues of TTHA1652 essential for the binding of maltotetraose are shown in cyan. (C) The comparison of γ-CD binding in *Tvu*CMBP (grey) and TTHA1652 (cyan). The positioning of γ-CD from the crystal structure (PDB ID: 2ZYK) and the molecular docking experiment is shown in black and green, respectively. The aromatic residues Phe9, Phe39, Tyr157, Tyr215, Trp235 and Trp345 of TTHA1652 holding the G1c2, Glc3 and Glc4 units of γ-CD are depicted in cyan. (D) Schematic representation of ORFs for maltodextrin transport and metabolism systems in *T. thermophilus* HB8 (upper) and HB27 (lower) and *T. maritima* (PDB IDs: 2GHA (upper) and 2FNC (lower)). Each gene is represented by an arrow indicating the direction of transcription along with its ORF number and respective encoded protein names. ORFs encoding a functionally similar protein across the organisms are depicted in similar color such as ABC transporter, green; maltodextrin glucosidase, cyan; transcription regulator, magenta; endo-1,4-β-galactosidase, blue; cyclo-maltodextrinase, navy blue; α-amylase, purple and pullulanase, yellow.

3.3.6 β-glucoside transporter

The protein TTHB082 seems to prefer laminaribiose and cellobiose disaccharides

In UniProtKB database, the protein TTHB082 is annotated as "sugar ABC transporter, periplasmic sugar-binding protein", which shares the highest sequence similarity (sequence identity: 27%, query coverage: 94%) with a sugar-binding protein from *Xanthomonas citri* (PDB ID: 3UOR) for which the cognate sugar is unknown. The other closest homologs of the protein TTHB082 are GacH receptor (PDB ID: 3K01, sequence identity: 28% and query coverage: 81%) and D-tagatose-binding protein (PDB ID: 5CI5, sequence identity: 25% and query coverage: 77%). A pairwise sequence alignment of TTHB082 with GacH receptor and D-tagatose-binding protein, shows almost no conservation for sugar-binding residues. Thus, to identify the cognate sugar for the protein TTHB082, we predicted its tertiary structure using three different programs; all of which used sugar-binding protein (PDB ID: 3UOR, unbound and open conformation) as the default template. Thus, for further characterization based on the active site, closed conformation of the protein TTHB082 was predicted using the D-tagatose-binding protein (PDB ID: 5CI5, tagatose bound and closed conformation) as a template. Interestingly, the closed conformation of

the protein TTHB082 exhibits an aromatic and polar amino acids-rich active site (Figure A.7A). This observation is in agreement with other known carbohydrate-binding proteins such as lectins and sugar transporters, which possess a large number of aromatic residues in their active site providing a strong CH…π interactions to the substrate molecule (Spiwok, 2017).

To further substantiate the hypothesis, we compared the active-site pocket area and volume of the protein TTHB082 with its structural homologs. Although result suggests that the active site of TTHB082 has adequate volume and surface area for accommodating tri- and tetra-saccharides (Figure 3.5A), a closer investigation of subsites A-D indicates the possible occupancy of only mono- and di-saccharides in the subsites A-B due to the hindrance created by the helix H3 and the residue Trp177 in subsites C-D (Figure 3.5B). Also, the presence of two aromatic residues Trp41 and Trp256 in subsites A-B can stack the mono- and di-saccharides in the active site (Figure 3.5B). Thus, to identify a probable list of mono- and di-saccharides transported by the protein TTHB082, we looked into the nearby genetic arrangement. The investigation unveils the presence of sugar-metabolizing gene (ORF ID: TTHB087) downstream to the ABC transporter (ORF IDs: TTHB082-TTHB086) (Figure 3.5C). The ORF TTHB087 encodes β-glycosidase, which catalyzes the hydrolysis of disaccharide β-glucosides in the decreasing order of laminaribiose, sophorose and cellobiose into glucose (Dion et al., 1999; Teze et al., 2013). The co-occurrence of genes for the enzyme β-glycosidase and ABC transporter having the same preference for disaccharides suggest their relationship. To further corroborate the hypothesis, molecular docking with various sugars such as tagatose, glucose, laminaribiose, sophorose and cellobiose was performed. The results assert the binding of β-glucosides to the protein TTHB082 with a preference to laminaribiose and cellobiose by making the CH…π interactions (Figure A.7B-A.7E and Table A.3). In summary, *T. thermophilus* HB8 transports β-glucosides by using ABC transporter (ORF IDs: TTHB082-TTHB086), which are subsequently metabolized by the β-glycosidase (ORF ID: TTHB087) for downstream processing (Figure A.8).

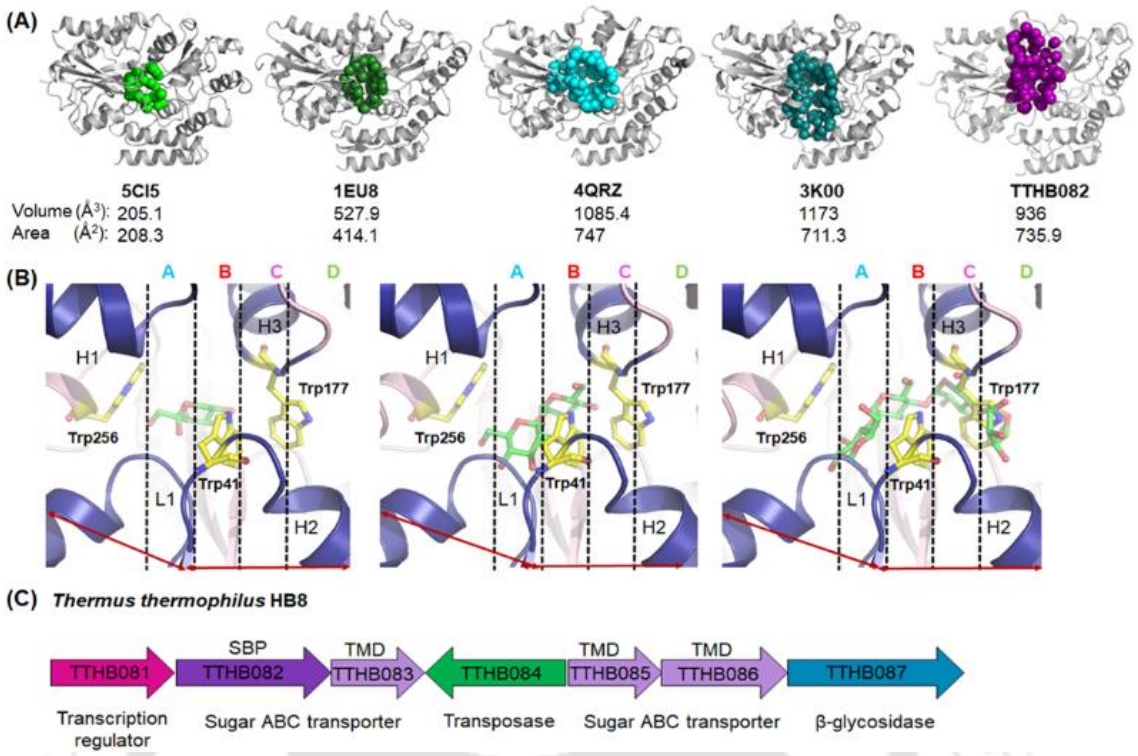


Figure 3.5. Active site and genetic organization of protein TTHB082. (A) Comparison of area and volume of the active-site pocket of the protein TTHB082 with mono- (PDB ID: 5CI5), di- (PDB ID: 1EU8), tri- (PDB ID: 4QRZ) and tetra-saccharide-binding (PDB ID: 2K00) proteins. The active site of each protein is shown as sphere in different colors and rest of the protein in grey. Details of the area and volume of active site of each protein is mentioned at the bottom of the figure. (B) Pictorial representation of four subsites (A-D) of the protein TTHB082. The secondary structural elements (L1 and H1-H3) regulating the subsites are shown in blue while rest of the protein in pink. Four subsites (A-D) are separated by dotted vertical lines and the subsite modulating residues are shown as ball-and-stick model in yellow. The occupancy for mono-, di- and tetra-saccharides in four subsites is shown by superimposing the protein TTHB082 and its structural homologs. D-tagatose- (PDB ID: 5CI5), trehalose- (PDB ID: 1EU8) and maltotetraose-binding proteins (PDB ID: 3K00) shown at left, center and right sides, respectively. (C) A schematic representation of gene cluster for β-glucoside uptake and metabolism. Each gene is represented by an arrow indicating its direction of transcription. The ORF number and its respective encoded protein name is provided along the arrows. The ORFs encoding transcription regulator, ABC transporter and β-glycosidase are colored in magenta, purple and blue, respectively.

3.3.7 Glucose ABC transporter

The protein TTHA0688 is a glucose-binding protein

In UniProtKB database, the protein TTHA0688 is annotated as "sugar ABC transporter, substrate-binding protein" which is almost identical (sequence identity: 99%, query coverage: 98%) to a glucose-binding protein (GBP, ORF ID: TTC0328, PDB ID: 2B3B) from *T. thermophilus* HB27 (Cuneo et al., 2006). Though the cognate ligand for TTHA0688 has not been reported, its highest homology with GBP suggests glucose to be a probable ligand. To further reaffirm this, tertiary structure of TTHA0688 was predicted using different programs, which used GBP (PDB ID: 2B3B) as the best template. Akin to GBP, TTHA0688 shares topological similarity with disaccharide-binding protein (TMBP, PDB ID: 1EU8) rather than monosaccharide-binding protein (PDB ID: 1GLG) (Figure A.9A) (Cuneo et al., 2006). However, similar to GBP, the presence of the residue His347 from loop L1 in the active site of TTHA0688 negates their possible binding with disaccharides and suggests the binding of only monosaccharides (Figure A.9B). To further corroborate the preference for monosaccharide by the protein TTHA0688, its genetic machinery associated with monosaccharide sugar transport and metabolism was analyzed. It has been reported that the genetic organization of the glucose ABC transporter in *Pseudomonas aeruginosa* (ORF IDs: PA3188-PA3190) and *T. thermophilus* HB27 (ORF IDs: TTC0326-TTC0328) are identical (Cuneo et al., 2006). Analysis of the genetic operons in *T. thermophilus* HB8 (ORF IDs: TTHA0685-TTHA0688) also shows a similar genetic organization of the glucose ABC transporter. However, unlike *P. aeruginosa*, both the thermophilic organisms *T. thermophilus* HB27 and HB8 lack genes for the glucose metabolism (Figure A.9C). Instead, the glucose ABC transporter genes are followed by unrelated genes encoding TolQ and TolR proteins involved in transport of group A colicins (Figure A.9C). In summary, although the bacterium *T. thermophilus* HB8 possesses glucose ABC transport system (ORF IDs: TTHA0685-TTHA0688), it lacks the genes for its metabolism or it may be located at distant on the genome.

3.3.8 UgpABCE transporter

For the *de novo* biosynthesis of phospholipids, uptake of two essential metabolites viz. *sn*-glycerol-3-phosphate (G3P) and glycerophosphocholine (GPC) is accomplished mostly through UgpABCE transporters (Wuttge et al., 2012). Structurally, UgpABCE transporters are known to be similar to sugar ABC transporters and belong to the CUT1 family (Wuttge et al., 2012; Jiang et al., 2014). Owing to its high sequence and structural similarities with that of sugar ABC transporters, UgpABCE transporters, particularly its SBP component (referred to as UgpB), are often misannotated as sugar-binding proteins. In our previous work, we reported such a case where the UgpB protein (*Tt*UgpB, ORF ID: TTHA0379) from *T. thermophilus* HB8 was misannotated as a sugar-binding protein (Chandravanshi et al., 2016). In UniProtKB database, other two ORFs TTHA1936 and TTHV034 from *T. thermophilus* HB8 are annotated as UgpB proteins. Thus, here we report the possible reason(s) of the existence of multiple UgpB proteins in *T. thermophilus* HB8.

The proteins TTHA0379, TTHA1877, TTHA1936 and TTHV034 are functionally closer to UgpB rather than sugar-binding proteins

In UniProtKB database, ORFs TTHA0379 and TTHA1877 are annotated as "sugar ABC transporter, periplasmic sugar-binding protein" while TTHA1936 and TTHV034 as "glycerol-3-phosphate ABC transporter, periplasmic glycerol-3-phosphate-binding protein" and "glycerol-3-phosphate ABC transporter substrate-binding protein", respectively. However, a homology search of TTHA0379, TTHA1877, TTHA1936 and TTHV034 against UniProtKB database reveals their decent similarity (in the range of 22-29%) with both the UgpB and sugar-binding proteins. Thus, to assign their exact function, an evolutionary analysis of these proteins and their homologs were performed. The result reveals their association with UgpB proteins rather than sugar-binding proteins and suggests a misannotation of the protein TTHA1877 (Figure 3.6A). To further substantiate the phylogenetic result, a multiple sequence alignment of these proteins with the UgpB proteins from *T. thermophilus* (*Tt*UgpB), *Mycobacterium tuberculosis* (*Mt*UgpB) and *E. coli* (*Ec*UgpB) was performed. The alignment confirms the conservation of the residues Tyr42, Glu66, Ser121, Trp169, Trp172, Gly284, Tyr323 and Arg374 (numbering according to *Ec*UgpB) reported to coordinate G3P or GPC in the active site of UgpB

proteins (Figure 3.6B). Notably, the presence of either one or two tryptophan residues determines the binding of GPC or G3P, respectively (Jiang et al., 2014; Chandravanshi et al., 2016; Adhikari et al., 2017). As all these proteins TTHA0379, TTHA1877, TTHA1936 and TTHV034 possess only one tryptophan residue (Figure 3.6B), it can be concluded that they all might bind GPC rather than G3P.

To locate the GPC-binding site of the proteins TTHA1877, TTHA1936 and TTHV034, their tertiary structures were predicted using three different programs; all of which used the G3P-and GPC-binding proteins from *E. coli* (*Ec*UgpB, PDB ID: 4AQ4) and *M. tuberculosis* (*Mt*UgpB, PDB ID: 4MFI), respectievely, as the default templates. Comparison of theoretical models with crystal structures of G3P- (*Ec*UgpB) and GPC-binding (*Mt*UgpB)proteins establishes that all three proteins TTHA1877, TTHA1936 and TTHV034 can bind GPC rather than G3P (Figure 3.6C). Since the GPC-bound structure of UgpB is unavailable in the literature, molecular docking was executed to obtain the substrate-binding mode to the UgpB protein; the protein TTHA0379, which is already annotated as GPC-binding protein, was taken as control. Results reveal that GPC binds to the active site of protein TTHA1877, TTHA1936 and TTHV034, wherein, the phosphate group of the GPC is held by the residues Tyr302, Tyr303 and Gly284, respectively (Table A.3). Similarly, the trimethylammonium moiety of GPC is held mostly through cation…π interactions by aromatic residues, which form a hydrophobic cage in the active site of these proteins (Figure A.10A-A.10C). Breathtakingly, the molecular docking experiment of GPC with the protein TTHA0379 suggests the binding of GPC in a secondary site, far from the putative active site. The possible reason for the non-binding of the GPC in the active site seems to be due to the hindrance created by the residue Gln274 of the protein TTHA0379 (Figure A.10D). This enforces to hypothesize that GPC may not be a physiological ligand for the protein TTHA0379. Thus, to understand the transport and metabolism of GPC, we looked into the genetic arrangement in the vicinity of all the four ORFs TTHA0379, TTHA1877, TTHA1936 and TTHV034. Astoundingly, the investigation reveals that unlike the *E. coli* UgpABCE transporter, *T. thermophilus* HB8 UgpABCE transporters lack genes for GPC metabolism in their locality. Moreover, the flanking genes in all four operons are different indicating distinct ligand for each transporter (Figure 3.6D). In

summary, all the four proteins are members of UgpB, however, their cognate ligands remain unknown.

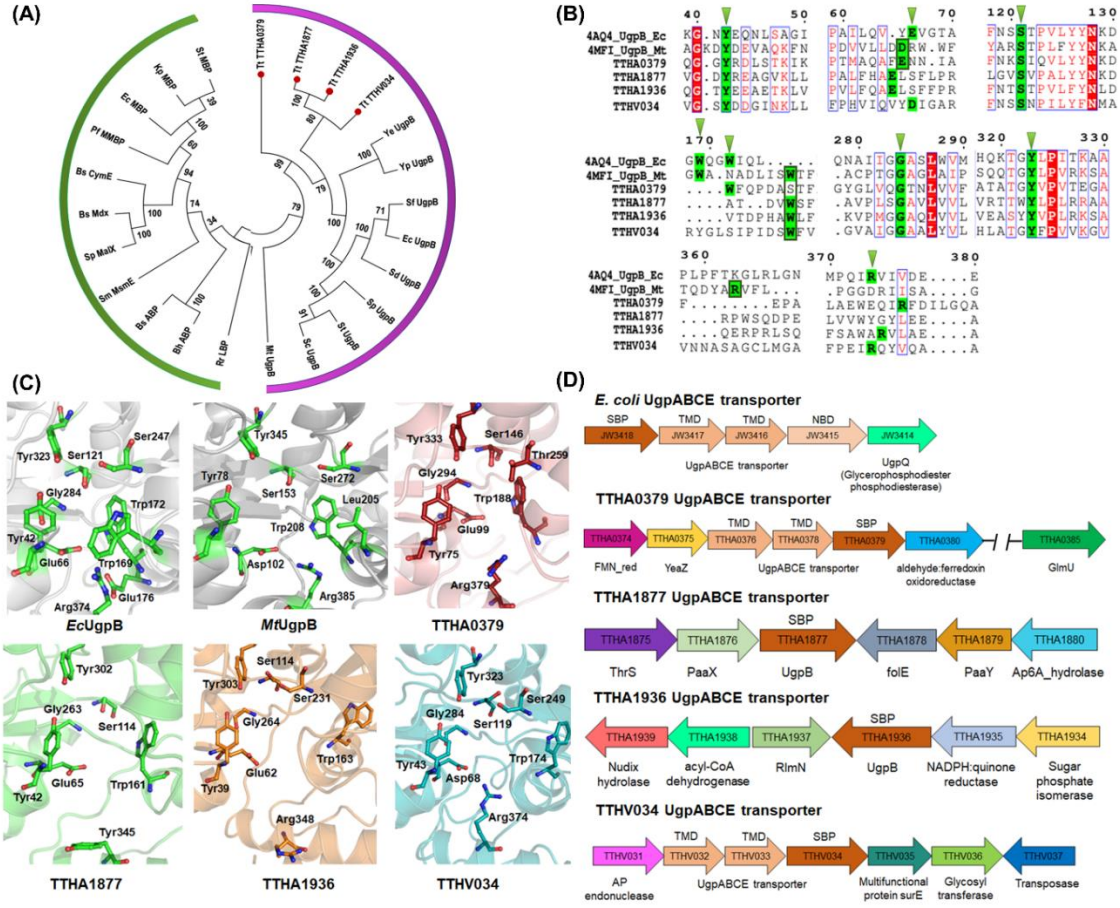


Figure 3.6. Comparative assessment of four UgpB proteins with *Ec*UgpB and *Mt*UgpB. (A) Phylogenetic tree showing the evolutionary relationship of TTHA0379, TTHA1877, TTHA1936 and TTHV034 along with UgpBs and sugar-binding proteins. The phylogenetic tree is constructed using the protein sequences of TTHA0379, TTHA1877, TTHA1936, TTHV034, UgpBs from *Salmonella choleraesuis* (Q57IS0), *S. typhimurium* (Q7CPK0), *S. paratyphi* (Q5PJK8), *Shigella dysenteriae* serotype 1 (Q32AT3), *E. coli* (P0AG80), *S. flexneri* serotype 5b (Q0SZL9), *Yersinia pestis* bv. Antiqua (Q1CNC9), *Y. enterocolitica* (A1JID7) and *M. tuberculosis* (A5U6I5) and sugar-binding proteins from *Rhizobium radiobacter* (P29822), *Bacillus halodurans* (Q9KEE7), *B. subtilis* (P94528, O06989 and O07009), *Streptococcus mutans* serotype c (Q00749), *S. pneumoniae* serotype 4 (P59213), *Pyrococcus furiosus* (P58300), *E. coli* O157:H7 (P0AEY0), *Klebsiella aerogenes* (P18815) and *S. typhimurium* (P19576). The UniProt ID of each protein is provided in parenthesis. In the phylogenetic tree, two separate clusters of UgpBs and sugar-binding proteins are shown in green and violet, respectively. The proteins TTHA0379, TTHA1877, TTHA1936 and TTHV034 are marked with red dots. (B) Multiple sequence alignment of TTHA1877, TTHA1936 and TTHV034 along with reference proteins *Ec*UgpB (P0AG80), *Mt*UgpB (A5U6I5) and TTHA0379 (Q5SLB4); the codes provided in

parenthesis are UniProt IDs. For the clarity of the figure, only a partial alignment has been shown here. The essential conserved residues of *Ec*UgpB proteins are shaded in green and those similar to *Mt*UgpB are enclosed in green box. (C) Comparison of the active-site pockets of TTHA1877, TTHA1936 and TTHV034 with *Ec*UgpB, *Mt*UgpB and TTHA0379. The active-site residues known to be involved in holding GPC are shown as ball-and-stick model while rest of the protein as cartoon. (D) A diagrammatic representation of the genetic cluster of *Ec*UgpB, TTHA0379, TTHA1877, TTHA1936 and TTHV034 UgpABCE transporters. Each gene is represented by an arrow indicating its direction of transcription. The ORF numbers and its respective encoded protein names are furnished along each arrow. The ORFs encoding ABC transporters are shown in brown while ORFs for different proteins are filled with different color.

3.3.9 Purine ABC transporter

The protein TTHA1301 is a purine-binding protein and belongs to the CUT2 family

In UniProtKB database, the protein TTHA1301 is annotated as "uncharacterized protein", whereas it's corresponding TMD components (ORF IDs: TTHA1302 and TTHA1303) are annotated as "sugar ABC transporter, permease protein". In contrast, a homology search of the protein TTHA1301 unveils its highest similarity (sequence identity: 38% and query coverage: 96%) with a purine-binding protein from *Brucella abortus* bound to adenine (PDB ID: 3S99). A sequence comparison of the protein TTHA1301 and the purine-binding protein confirms the conservation of adenine-coordinating residues (Figure 3.7A).

To further affirm the binding of purine to the protein TTHA1301, its tertiary structure was predicted using three different programs; all of which used purine-binding protein from *B. abortus* (PDB ID: 3S99) as the default template. The theoretical model comprises of a typical N- and C-terminal domains linked by hinge region (Figure 3.7B). Further, the structural homologs of the protein TTHA1301 are found to be both purine- and sugar-binding proteins (Table A.2). In addition, even though TTHA1301, purine- and sugar-binding proteins show dissimilarity at the primary structure level, they share a similar structural fold (Figure 3.7C). This characteristic has earlier been observed in the case of the purine nucleoside receptor A (PnrA, PDB ID: 2FQY), which belongs to the CUT2 family and shows resemblance to ribose- and glucose-binding proteins (Deka et al., 2006; Webb and Hosie, 2006). Interestingly, according to structural classification of SBPs, the purine-binding protein from *B. abortus* (PDB ID: 3S99) has been grouped in the cluster B,

a class of sugar-binding proteins (Scheepers et al., 2006). Thus, to unriddle the exact ligand for the protein TTHA1301, we identified the major structural determinants of cluster B subgroups (B-I to B-V). Our results bring out that the purine-binding proteins including TTHA1301 constitute an additional α-helix crossing the interconnecting regions of the protein, which is absent in the sugar-binding proteins (Figure 3.7D). Thus, the protein TTHA1301 can be inferred to be a purine-binding protein. Furthermore, to evaluate the binding of purine to the protein TTHA1301, molecular docking experiments were performed with nitrogenous bases (i.e. purine and pyrimidine). Interestingly, results are in accordance with above hypothesis and suggest the binding of purine (i.e. adenine and guanine) bases to the protein TTHA1301 (Table A.3).

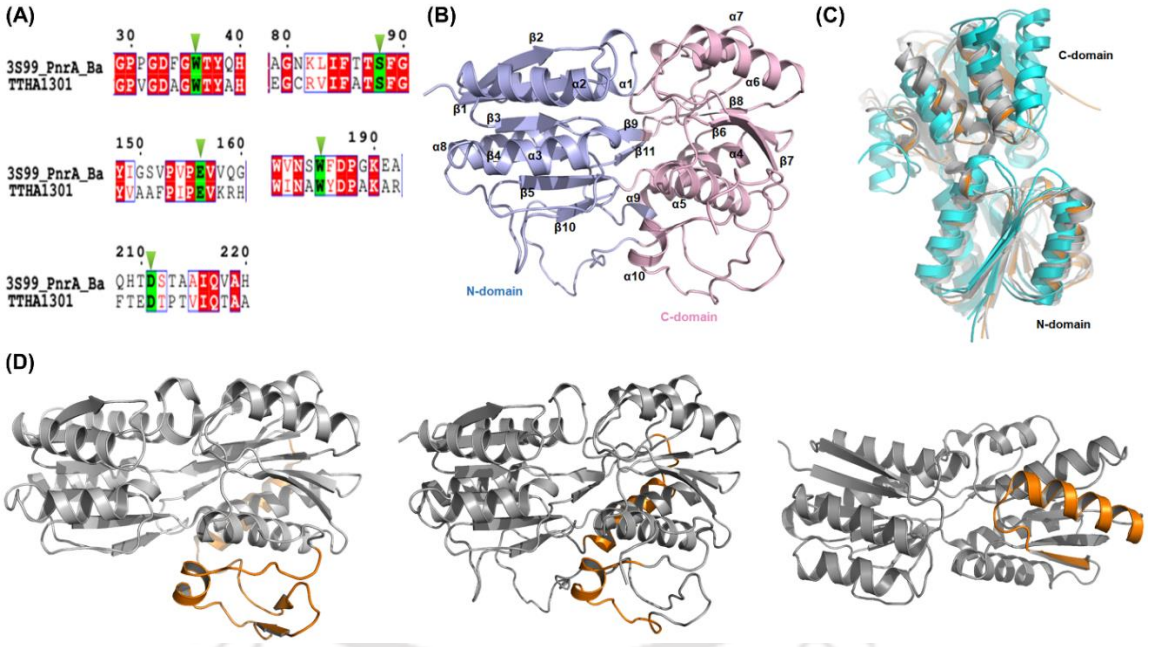


Figure 3.7. Sequence- and structure-based inference of purine binding with TTHA1301. (A) The pairwise sequence alignment of TTHA1301 and purine-binding protein (PDB ID: 3S99) demonstrating the conservation of adenine-binding residues shown in green blocks and downward arrowheads. (B) A theoretical model of the protein TTHA1301 comprising of N- (blue) and C-terminal (pink) domains. (C) Structural superimposition of the TTHA1301 model (orange), purine- (PDB IDs: 3S99 and 2FQY, grey) and sugar-binding proteins (PDB ID: 4RY0 and 2H3H, cyan). (D) Comparison of structural feature of purine-binding protein (PDB ID: 3S99, left, TTHA1301 (model, center) and sugar-binding protein (PDB ID: 4RY0, right). Distinctive secondary structural elements of purine-, TTHA1301, and sugar-binding proteins are highlighted in orange.

3.3.10 Sharing of nucleotide-binding domains (NBDs) and transmembrane domains (TMDs) among carbohydrate ABC transporters

It is well known that an ABC importer comprises of three components viz. substrate (or solute)-binding protein (SBP), transmembrane domain (TMD) and nucleotide-binding domain (NBD). Thus, it becomes imperative to understand whether there remains always a one-to-one mapping of these components in all the ABC import systems. To study the hypothesis, we searched the TransportDB 2.0 database and found a total of 127 ORFs encoding the components of ABC importers in *T. thermophilus* HB8. Out of these, a total of 35 ORFs are annotated as carbohydrate ABC transport systems (Table A.1).

In this study, we have already identified a total of six carbohydrate ABC transporters for sugar uptake comprising of 6 SBPs, 13 TMDs and 2 NBDs. Notably, all, except xylose (ORF IDs: TTHV087-TTHV089) and mannosylglycerate (ORF IDs: TTHA0975-TTHA0979) ABC transport systems, lack their respective NBD component(s). It has already been reported that many gram-positive bacteria usually contain operon of carbohydrate ABC transporters lacking NBD component(s), however, can utilize NBD(s) encoded by ORFs located elsewhere on the genome (Ferreira and Sa-Nogueira, 2010; Marion et al., 2011; Wuttge et al., 2012). Thus, to identify a potential NBD component for each carbohydrate ABC transporter, we used protein-protein network analysis approach. Our results indicate that three ORFs TTHA0579, TTHA0975 and TTHA1630, coding for NBDs, can interact with multiple sugar ABC transport systems (Figure 3.8A-3.8D and Table A.4). Although the protein TTHA0975 is the NBD component for the mannosylglycerate ABC transporter, our results reveal its utilization by other sugar ABC transporters as well (Figure 3.8A-3.8D).

In *T. thermophilus* HB8, out of 11 carbohydrate ABC transport systems, 10 are identified for carbohydrates including six for sugars and four for phospholipid precursors such as G3P/GPC. Interestingly, for the first time, our analysis reveals the absence of TMDs in the UgpABCE transporters. Moreover, out of four UgpABCE transporters, the SBP TTHA0379 lacks only its NBD while TTHA1877 and TTHA1936 lack both their TMD and NBD constituents. A search for NBDs and TMDs for the proteins TTHA0379,

TTHA1877 and TTHA1936 using protein-protein interaction analysis unveils the NBDs and TMDs corresponding to the sugar ABC transporters (Figure 3.8E-3.8G and Table A.4). These observations designate that the NBD and TMD components of the sugar ABC transporters can be utilized by the UgpABCE transporters. This seems to be a possible rationale for the UgpB proteins being evolutionary related to the sugar-binding proteins of ABC transporters.

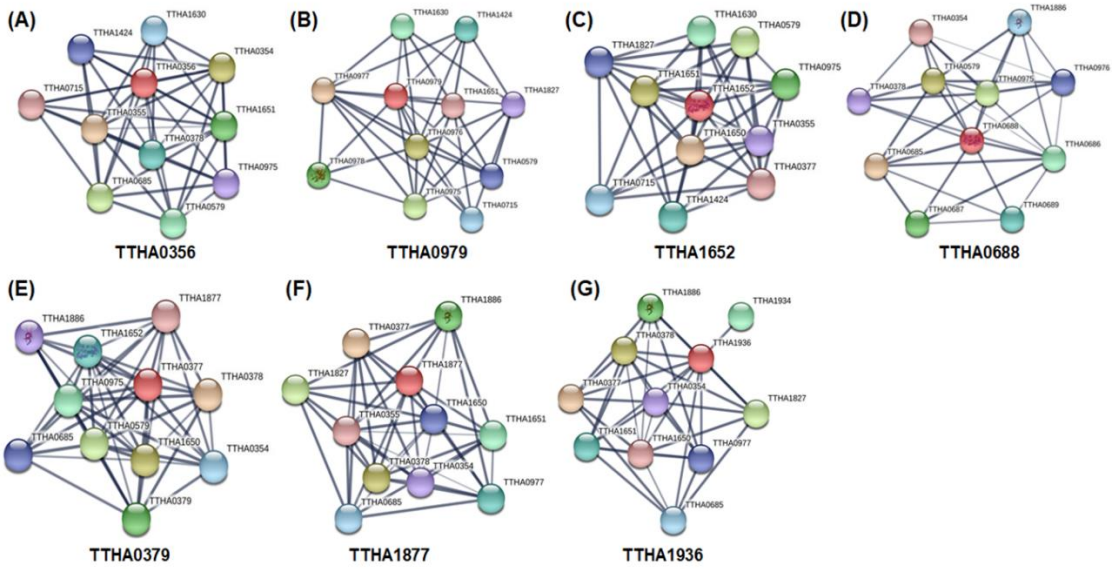


Figure 3.8. Interactome map representing the protein-protein interactions. Network showing the protein-protein interaction among SBP, TMD and NBD subunits of (A) trehalose/maltose, (B) mannosylglycerate (C) cyclo/maltodextrin, (D) glucose ABC transporters and (E, F, G) UgpABCE transporters. Each node represents a subunit of ABC transporter and the connections between nodes represent a predicted interaction. The result demonstrates the interaction of NBD with TMD in all the carbohydrate ABC transporters excluding TTHA0379, for which its TMD (ORF ID: TTHA0377) has been used to get the interaction with its NBD. The ORF number is provided along with each node. The details of each ORF is provided in Table A.4.

3.3.11 Carbohydrate uptake and metabolism network in *T. thermophilus* HB8

To further substantiate the relationship between transport and metabolism of carbohydrates, we collected all the respective genes from the KEGG database (Kanehisa and Goto, 2000). In *T. thermophilus* HB8, 10 out of 11 putative carbohydrate ABC transport systems (6 - sugar and 4 - G3P or Ugp, encoded by 31 genes) and 106 metabolic

enzymes (97 - sugar and 9 – phospholipid, encoded by 106 genes) are found to be involved in carbohydrate (sugar and phospholipid precursor) transport and metabolism, respectively (Figure 3.9, Table A.1 and A.5). In *T. thermophilus* HB8, a metabolic profiling of phospholipids reveals the presence of glycerophosphoryl diester phosphodiesterase (ORF ID: TTHB141), which degrades the glycerophosphodiester such as GPC imported by UgpABCE transporter during the phospholipid *de novo* synthesis (Wuttge et al., 2012; Jiang et al., 2014; Chandravanshi et al., 2016). The existence of genes encoding the enzymes neopullulanase, maltodextrin glucosidase and β-glucosidases affirm the metabolism of polysaccharides such as starch and cellulose. In addition to polysaccharides, metabolism of monosaccharides such as glucose and xylose is identified to be facilitated by glycolysis and pentose phosphate pathways, respectively. Moreover, the metabolism of two key osmoprotectants (trehalose and MG) of *T. thermophilus* HB8 is mediated by starch and sucrose as well as fructose and mannose metabolism pathways, respectively. Remarkably, most of the sugar metabolism networks seem to be interconnected. For instance, the maltodextrin metabolism pathway is linked to the trehalose metabolism pathway. Eventually, all the carbohydrate metabolism pathways are interlinked to the glycolysis pathway which produces energy (Figure 3.9). The bacterium *T. thermophilus* HB8 being a halotolerant thermophile utilizes a carbohydrate-dependent protective mechanism harboring both uptake and biosynthetic systems using two major compatible solutes trehalose and mannosylglycerate as osmolyte. In summary, this study provides a functional relationship between carbohydrate transport and its metabolism.

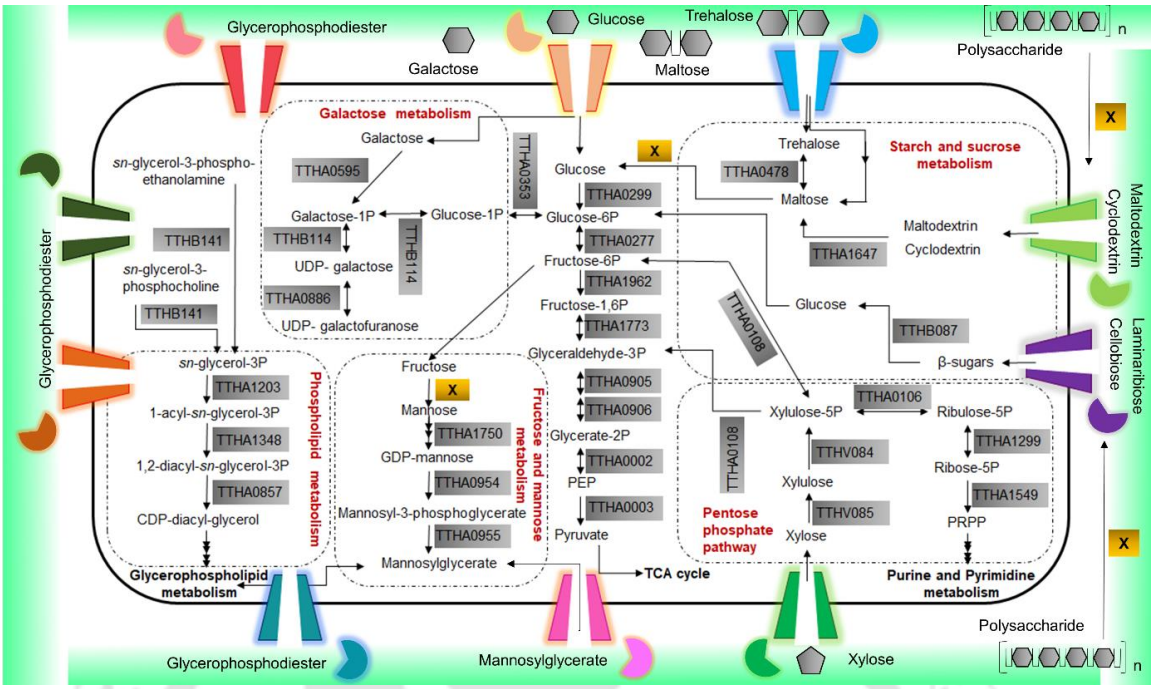


Figure 3.9. Schematic representation of the integrated network of carbohydrate ABC transporters and metabolic pathways. In total, 10 carbohydrate ABC transporters and 5 metabolic pathways enclosed by dotted lines along with the glycolysis pathway in the center. Proteins involved in the transportation of different carbohydrates located in the plasma membrane are represented in different colors while those involved in the carbohydrate metabolism are highlighted in grey. In addition, the unidentified genes of a pathway have been marked with the symbol **X**. For clarity of the figure, only a few proteins from Table A.5 have been shown here. Further details of each protein associated with carbohydrate metabolism are provided in Table A.5.

3.4 CONCLUSION

This chapter demonstrates an integrated approach for the functional annotation of carbohydrate ABC transporters, where cognate ligand for each transporter is verified by their metabolic profiling. Through this study, a total of 11 putative carbohydrate ABC transporters from *T. thermophilus* HB8 are identified and their cognate ligand are assigned. Out of these, six are characterized as sugar, four as phospholipid (viz. UgpABCE) and the remaining one as purine ABC importer. This study also supports the relationship between microbial habitats and the carbohydrate diversity. Functional annotation of carbohydrate ABC transporters and its metabolic profiling suggest the transport and synthesis of two major osmolytes viz. trehalose and mannosylglycerate. In addition, the bacterium *T.*

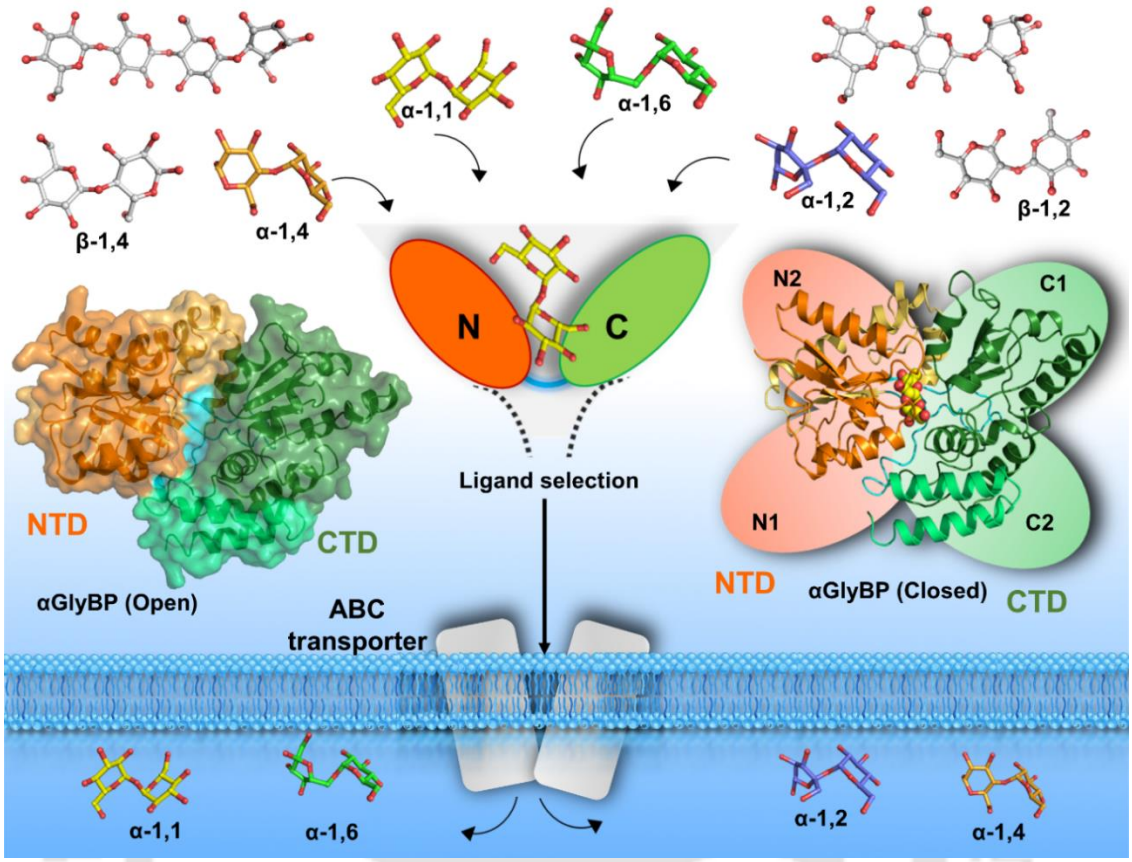
thermophilus HB8 seems to transport starch- and cellulose-degraded products as carbon and energy source. Apart from cognate ligand identification, this study also suggests the hypothesis of the sharing phenomenon of ABC transporter subunits. Further, a thorough analysis of ABC transporter subunits reveals that the NBDs are not coexistent with the carbohydrate ABC transporter operons and thus are shared among multiple carbohydrate ABC transporters. Interestingly, UgpABCE transporters lack both the NDB as well as TMD components and hence seems to borrow from other sugar ABC transport systems. In summary, the predicted carbohydrate metabolism network along with the carbohydrate ABC transport systems suggest that most of the carbohydrate ABC transporters are interlinked by carbohydrate metabolism which, in turn, may help the bacterium to survive even in the scarcity of some carbohydrates.

This chapter has been published as:

Chandravanshi M, Gogoi P and Kanaujia SP (2020). Structural and thermodynamic correlation illuminates the selective transport mechanism of disaccharide α-glycosides through ABC transporter. *FEBS J.*, 287:1576-1597.

ABSTRACT

Carbohydrate (or sugar) molecules are extremely diverse regarding their length, linkage and epimeric state. Acquisition of these essential molecules inside the cell is achieved by the functional and structural unit of ATP-binding cassette (ABC) transport system known as substrate (or solute)-binding protein (SBP) which assists in the selective passage of solutes. Though several SBPs bound to carbohydrates have been characterized, little is known about the selection criteria for carbohydrate diversity. This study reports crystal structures of an α-glycoside-binding protein (αGlyBP, ORF ID: TTHA0356 from *Thermus thermophilus* HB8) in complex with disaccharide α-glycosides namely trehalose (α-1,1), sucrose (α-1,2), maltose (α-1,4), palatinose (α-1,6) and glucose within a resolution range of 1.6 to 2.0 Å. Despite transporting multiple types of sugars, αGlyBP maintains its stereoselectivity for both glycosidic linkage as well as an epimeric hydroxyl group. Out of the two subsites identified in the active-site pocket, subsite B which accommodates the glucose and glycosyl unit of disaccharide α-glycosides is highly conserved. In addition, structural data confirms the paradoxical behavior of glucose where it replaces the high-affinity ligand(s) (disaccharide α-glycosides) from the active site of the protein. Measurement of the binding affinity of αGlyBP for disaccharide α-glycoside evinces that mutation of active-site residues alters the ligand preference. In addition, comparative assessment of open and closed conformations of αGlyBP along with mutagenic and thermodynamic studies identifies the hinge region as the first interaction site for the ligands. On the other hand, encapsulation of ligand inside the active site is achieved through the N-terminal domain (NTD) movement, whereas the C-terminal domain (CTD) of αGlyBP is identified to be rigid and postulated to be responsible for maintaining the interaction with the transmembrane domain (TMD) during substrate translocation.



4.1 INTRODUCTION

Carbohydrates are crucial biological macromolecules which serve as a major source of carbon and energy in a variety of physiological functions. To accomplish these functions, carbohydrates are available in diverse forms. Nature provides diversity to carbohydrates by creating a different combination of the basic monosaccharide units (e.g. glucose, GLC and fructose, FRU) to form higher oligosaccharides and the glycosidic linkages which joins two carbohydrate molecules (e.g. α-1,4 and α-1,6 in starch; β-1,4 in cellulose) (Hölemann and Seeberger, 2004). Altogether, these results in a higher structural diversity of carbohydrates which makes them extremely complex in the context of their length, conformation, monosaccharide-ring constituent and α/β anomeric configuration (Raich et al., 2016). Due to this diversity and complexity, microorganisms adopt a diverse set of transporters for their uptake during cellular metabolism. Uptake of carbohydrates in a cell is executed by three classes of transporters viz. (1) primary active transporter, (2)

secondary transporter and (3) group translocator (Saier, 2000a). Out of which, ATP-binding cassette (ABC) transporter belonging to the primary active transporter is the largest superfamily for carbohydrate transport (Saier, 2000b). ABC transporters utilize ATP as an energy source to facilitate the transport of solute molecules across the membrane and are classified as ABC exporters and importers. Although, architecturally both ABC exporters and importers contain common subunits namely the transmembrane domain (TMD) and the nucleotide-binding domain (NBD), ABC importers possess an additional domain referred to as substrate-binding proteins (SBPs) (Rees et al., 2009; Wilkens, 2015). Functionally, SBPs capture the substrates and bring it towards the TMDs for subsequent translocation and thus renders specificity and directionality to importers (Maqbool et al., 2015). Interestingly, ABC exporters are distributed in all the domains of life while ABC importers are present only in prokaryotes and recently reported in plants too (Kang et al., 2011; Kretzschmar et al., 2011).

Based on overall topology and ligand specificity, SBPs are classified into seven different clusters (viz. A-F) and further subdivided into various sub-clusters. Out of which, SBPs involved in transporting carbohydrates belong to the clusters B and D (particularly, sub-cluster D-I) (Scheepers et al., 2016). SBPs belonging to these clusters comprise of two α/β domains viz. N- and C-terminal domain (NTD and CTD, respectively) connected by a loop which acts as a hinge as well as the carbohydrate-binding site (Berntsson et al., 2010). In an unliganded state, both the domains (NTD and CTD) remain separated (i.e. open conformation) and can freely rotate via the hinge region, whereas upon ligand binding they move asymmetrically closer to each other (i.e. closed conformation) and encapsulate the ligand between them (Pandey et al., 2016). This conformational change of SBP upon ligand binding is proposed as a ‘‘Venus Fly-trap’’ mechanism (Mao et al., 1982). Structurally, SBPs belonging to the sub-cluster D-I possess a common structural fold and can bind a diverse range of carbohydrates varying in size and chemical constituents. Earlier studies have reported the role of structural adaptation at the active-site pocket of SBPs to maintain the selectivity and specificity based on carbohydrate length (Cuneo et al., 2009a). This structural adaptation is governed by five secondary structural elements viz. two loops (L1 and L2) and three helices (H1, H2 and H3) that

modulate the occupancy of carbohydrate at the four subsites (A, B, C and D) of the active site (Cuneo et al., 2009a). Although, many SBPs have been structurally characterized for carbohydrate binding, the detailed enumeration of selective ligand binding mechanism associated with the vast carbohydrate diversity and complexity remains unfulfilled.

The requirement of carbohydrate transporters varies with microbial habitat as it plays a pivotal role in the accumulation of diverse extracellular sugars inside the cells. *Thermus thermophilus* strains are halotolerant and reside in marine hot spring (Alarico et al., 2005). In order to survive the extreme environment, *T. thermophilus* utilizes a variety of carbohydrates including polysaccharides (e.g. starch and glycogen), oligosaccharides (e.g. α- and β-glucosides and galactosides) and monosaccharides (e.g. glucose, galactose and xylose) for its growth (Henne et al., 2004). In addition to these carbohydrates, the bacterium also accumulates compatible solutes such as trehalose and/or mannosylglycerate to maintain the halotolerancy (Santos and Da Costa, 2002). Trehalose is a disaccharide α-glycoside in which two glycosyl units (Glc1 and Glc2) are linked via an α-(1,1) glycosidic bond. For survival, *T. thermophilus* maintains the intracellular level of compatible solutes either by synthesis or by importing it through energy-dependent transport system such as ABC transporter (Silva et al., 2003). The ABC transport system of *Thermus sp.* for trehalose is identified to be multispecific in nature and is able to transport a wide range of substrates namely trehalose (TRE), maltose (MAL), sucrose (SUC), palatinose (PAL) and glucose (GLC) (Silva et al., 2005). Although, the trehalose transport system is identified as a multisubstrate transporter, the molecular basis of this multispecificity with preferential enumeration remains unclear. Currently, it is unknown whether this multispecific transport system possesses inherent selectivity or it transport substrates without any selection criteria. Henceforth, we hypothesize that the transport system exhibits specificity towards α-glycosides possessing various α-glycosidic linkages such as α-1,1 (e.g. trehalose, Glc1-(1,1)-Glc2), α-1,2 (e.g. sucrose, Glc1-(1,2)-Fru1), α-1,4 (e.g. maltose, Glc1-(1,4)-Glc2) and α-1,6 (e.g. palatinose, Glc1-(1,6)-Fru1). Indeed, till date many ABC transport systems have been structurally described that transports α-(1,1)-, α-(1,4)- and α-(1,6)-glycosides from different microorganisms (Diez et al., 2001; Cuneo et al., 2009a; Ejby et al., 2016). However, the mechanism underlying

the transport of all types of α-glycosides through a single transport system still remains elusive. Furthermore, it remains unknown whether the transport system that recognizes the multiple disaccharide α-glycosides can also bind to other complex oligosaccharides that are decorated with α-glycosidic linkages. Also, whether the transport system that exist for multiple α-glycosides exhibits sufficient selectivity to discriminate from sugars that are composed of β-glycosidic linkages. The full evaluation of transport system is crucial to solve the enigma of a selective transport mechanism, where the full understanding of the physiological basis for the transport of multiple sugars through a single transport system is also essential.

Previously, we reported the genetic cluster involved in the transport and metabolism of α-glycosides involved in maintaining the intracellular trehalose level of *T. thermophilus* HB8 as well as discovered the physiological basis of transporting maltose and glucose through trehalose ABC transport system (Chandravanshi et al., 2019). In addition, we also discovered that the transport system possesses a selective mechanism based on the carbohydrate length and exhibit a significant preference for the disaccharides over higher oligosaccharide (Chandravanshi et al., 2019). Since, the closest homolog trehalose/maltose-binding protein (TMBP, ORF: TTC1627) from *T. thermophilus* HB27 has been biochemically characterized for the multisubstrate transport, however the molecular mechanism of multiple substrate translocation and stereoselectivity for α-glycosides remains elusive (Silva et al., 2005). In this study, we have determined the three-dimensional crystal structures of an SBP (ORF: TTHA0356 from *Thermus thermophilus* HB8) of an ABC transporter (ORFs: TTHA0354-TTHA0356) in open as well as closed conformations in complex with disaccharide α-glycosides (e.g. trehalose (α-1,1), sucrose (α-1,2), maltose (α-1,4), palatinose (α-1,6)) as well as with monosaccharide (e.g. glucose) and thus named it as "α-glycoside-binding protein or αGlyBP". Information obtained from the structural and thermodynamic study of αGlyBP enabled us to postulate a selective mechanism for αGlyBP and also understand its ability to distinguish between α- and β-glycosides. In addition, structural characterization along with mutagenic studies provide a new prospect for the ligand-binding mechanism at atomic level which is in contrary to the well-established ''Venus Fly-trap'' mechanism.

4.2 MATERIALS AND METHODS

4.2.1 Cloning and site-directed mutagenesis

Full-length (1-1290bp) gene *TTHA0356* encoding αGlyBP cloned in pET11a vector (plasmid ID: PC010356-41) under the restriction sites *Nde*I and *Bam*HI, was purchased from Biological Research Center, NITE (NBRC), Japan. Sub-cloning of the truncated form of the gene (82-1290bp) was constructed without a signal sequence as predicted by SignalP 4.1 server (Petersen et al., 2011). For sub-cloning, the gene was amplified by polymerase chain reaction (PCR) using the cloned pET11a vector as a template and re-cloned into pET28a(+) vector (Novagen) using the same set of restriction enzymes i.e. *Nde*I and *Bam*HI (New England Biolabs). The oligonucleotide sequences used for amplification are mentioned in Table 4.1. In the resulting recombinant construct, the coding sequence starts at the residue Gln28 with an in-frame ATG start codon and hexa-histidine was tagged at the C-terminus of the protein to facilitate the purification process by immobilized Ni-NTA metal-affinity chromatography. This recombinant wild-type (WT) construct αGlyBP_*WT* was further used as a template to generate the following mutant constructs: *αGlyBP_R49A, αGlyBP_D70A, αGlyBP_D118A, αGlyBP_W287F, αGlyBP_W287A* and *αGlyBP_R356A* using oligonucleotide sequences mentioned in Table 4.1 and Q5 Site-Directed Mutagenesis Kit (New England Biolabs). All the resulting recombinant construct were confirmed by double digestion using the same set of restriction enzymes by incubating the vectors at 37 °C for two hours followed by plasmid DNA sequencing (Figure 4.1).

Table 4.1. List of oligonucleotide sequences used for αGlyBP_WT and αGlyBP_mutant recombinant constructs. In αGlyBP_WT construct, restriction enzyme recognition site and codon encoding for hexa-His tag are underlined and shown in bold, respectively. Similarly, in mutant constructs, codon encoding for substituted amino acid is shown in bold and underlined.

Primer	Oligonucleotide sequence (5′-3′)
αGlyBP_WT_F	ATATCATATGATGCAGTCCGGCCCCGTGATCCGCG
α*GlyBP_WT_R*	ATATGGATCCCTA**ATGATGATGATGATGATG**GCG CAGGATGCGG

αGlyBP_R49A_F	CACCAACGAC**GCC**CTCGCCCTCTACC
αGlyBP_R49A_R	TCGGCGGGGGAGTCAATG
αGlyBP_D70A_F	TACATGATT**GCC**GTCATCTGGCCG
αGlyBP_D70A_R	GACGTCCACATCGGGGCT
αGlyBP_D118A_F	TTCTTCACC**GCC**GCCGGCATC
αGlyBP_D118A_R	GGGCAGGGAGGTGAGCTT
αGlyBP_W287F_F	CTGGGCGGC**TTT**CAGCTCATGGTCTCCG
αGlyBP_W287_R	GGTGGCGGCGTTGGGGGC
αGlyBP_W287A_F	CCTGGGCGGC**GCG**CAGCTCATGGTCTCCG
αGlyBP_W287A_R	GTGGCGGCGTTGGGGGCG
αGlyBP_R356A_F	CGCCGTCTCC**GCT**CCCTCCGACG
αGlyBP_R356A_R	TTCTGGAAGACGGGGAGG

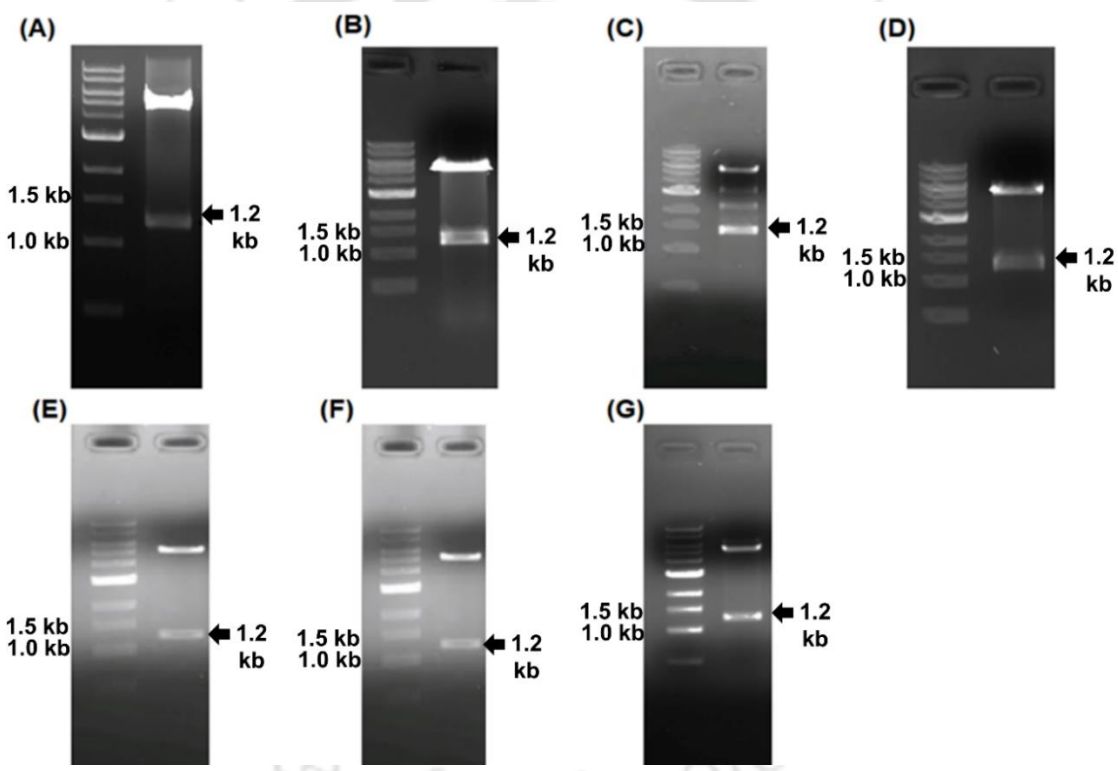


Figure 4.1. Cloning and site-directed mutagenesis of *TTHA0356*. (A) Clone confirmation of *TTHA0356_WT* by double digestion of the vector pET28a carrying the gene *TTHA0356* by *Nde*I and *Bam*HI restriction enzymes (lane 1: DNA ladder, lane 2: insert pop out after double digestion of the plasmid pET28a). Clone confirmation gel of all the TTHA0356 mutant constructs (B) *αGlyBP_R49A* (C) *αGlyBP_D70A* (D) *αGlyBP_D118A* (E) *αGlyBP_W287F* (F) *αGlyBP_W287A* and (G) *αGlyBP_R356A.* Mutagenesis confirmation of all the *TTHA0356* mutant constructs were done by plasmid DNA sequencing.

4.2.2 Over expression and protein purification of wild-type and mutant proteins

BL21 (DE3) *Escherichia coli* and Rosetta (DE3) *E. coli* competent cells (Novagen) were transformed with WT (*αGlyBP_WT*) and mutant recombinant constructs (*αGlyBP_R49A*, *αGlyBP_D70A*, *αGlyBP_D118A*, *αGlyBP_W287F*, *αGlyBP_W287A* and *αGlyBP_R356A*), respectively. Subsequently, all transformants were cultured followed by subculturing in 2.4 L of Luria-Bertani (LB) broth (supplemented with 50 µg/ml kanamycin) and grown at 37ºC to an optical density (OD_{600}) of ~0.6-0.8. Subsequently, the culture was induced with 1 mM isopropyl-1-thio-β-D-thiogalactopyranoside (IPTG) followed by overnight incubation at 25ºC to obtain soluble protein production. The cells were then harvested by centrifugation at 3000 rpm for 15 min and resuspended in lysis buffer (20 mM Tris-HCl pH 7.5, 20 mM imidazole, 150 mM NaCl, 10% glycerol, 1 mM phenylmethylsulfonyl fluoride (PMSF) and 3 mM β-mercaptoethanol (β-ME)). After resuspension, a clear lysate was prepared by disrupting the cells by sonication, followed by heating the cell lysate at 70°C for 10 min and centrifugation at 12000 rpm for 40 min at 4°C. The obtained supernatant was loaded on a pre-equilibrated Ni^{2+}-affinity resin (Qiagen) packed in pierce centrifuge column (Thermo Fisher Scientific) for desired protein purification. After 2 h incubation, the column was gradiently washed with wash buffer A (20 mM Tris-HCl pH 7.5, 10 mM imidazole, 150 mM NaCl, 10% glycerol, 1 mM PMSF and 3 mM β-ME) and wash buffer B (20 mM Tris-HCl pH 7.5, 20 mM imidazole and 150 mM NaCl). The proteins were eluted from the column with a higher concentration (250 mM) of imidazole in wash buffer B. Pure eluted protein fractions analyzed by SDS-PAGE (10%) were pooled together and step-wise dialysis was performed against 20 mM Tris-HCl pH 7.5 and 150 mM NaCl to get rid of imidazole. After dialysis, proteins (αGlyBP_WT, αGlyBP_R49A, αGlyBP_D70A, αGlyBP_D118A, αGlyBP_W287F, αGlyBP_W287A and αGlyBP_R356A) were concentrated using Vivaspin turbo 15 (10 kDa cutoff; Sartorius) up to a concentration of ~16, ~41, ~40, ~35, ~21, ~20 and ~28 mg ml^{-1}, respectively (Figure 4.2). The protein concentration was measured by UV_{280} (ε_{280} = 101300 M^{-1} cm^{-1}) method.

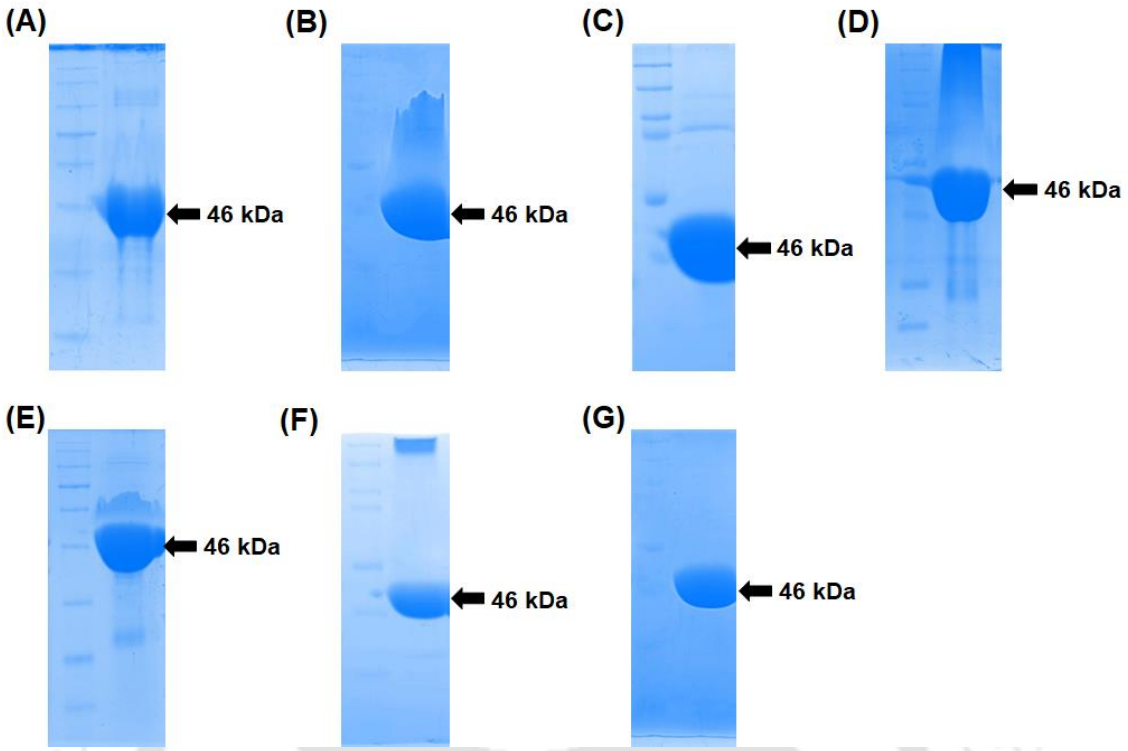


Figure 4.2. Purification of αGlyBP_WT and αGlyBP_mutant protein. SDS-PAGE analysis of purified and concentrated proteins (A) αGlyBP_WT, (B) αGlyBP_R49A, (C) αGlyBP_D70A, (D) αGlyBP_D118A, (E) αGlyBP_W287F, (F) αGlyBP_W287A and (G) αGlyBP_R356A. (lane 1: protein ladder, lane 2: concentrated protein).

4.2.3 Crystallization of wild-type (ligand bound) and mutant (ligand bound and free forms) proteins

Initial crystal screening was done at a protein:buffer ratio of 1:1 using Crystal Screen and PEG/Ion kits from Hampton Research in microbatch-under-oil technique at 4ºC and 20ºC. Crystals of both αGlyBP_WT and αGlyBP_mutant proteins were obtained in microbatch-under-oil at 4 ºC. Protein crystals of αGlyBP_WT were obtained in a buffer containing (0.04 M Citric Acid, 0.06 M Bis-Tris Propane/pH 6.4, 20% PEG 3350) at 4ºC within 3 weeks (Figure 4.3). To obtain the crystals of αGlyBP_WT protein in complex with maltose and glucose, the protein was first pre-incubated with a 100 times higher molar excess of maltose and glucose prepared in dialyzed buffer containing 20 mM Tris-HCl pH 7.5 and 150 mM NaCl for overnight at 4ºC. After pre-incubation, the protein was crystallized in a buffer solution containing (0.03-0.05 M Citric Acid, 0.05-0.07 M Bis-

Tris Propane) pH 3-7, 16% and 20% PEG 3350. After pH optimization, suitable crystals were obtained in the same precipitant at pH 5 and pH 6.4 (Figure 4.3). To obtain sucrose- and palatinose-bound forms of αGlyBP_WT, the protein was initially treated with glucose by pre-incubating with 1% glucose for overnight. Subsequently, the protein was dialyzed with buffer containing 20 mM Tris-HCl pH 7.5 and 150 mM NaCl and finally concentrated up to ~20 mg ml^{-1}. Prior to crystallization, the glucose-treated protein was incubated with 100 times higher molar excess of sucrose and palatinose prepared in a dialyzed buffer for overnight at 4 ºC and crystallized in the same buffer condition as mentioned above. Like other complexes, the best crystals of sucrose and palatinose complexes were also obtained in buffer containing (0.05-0.04 M Citric Acid, 0.05-0.06 M Bis-Tris Propane) pH 5 and pH 6.4, 16% and 20% PEG 3350 (Figure 4.3). In addition to the αGlyBP_WT complexes, mutant proteins in ligand-free and bound to disaccharide α-glycosides (such as trehalose, maltose, palatinose and sucrose) as well as glucose were also obtained in a similar crystallization buffer as described above. All the mutant protein crystals were obtained by using microbatch-under-oil technique at 4°C within 2-3 weeks (Figure 4.3). All the sugars used in this study were purchased from Sigma-Aldrich.

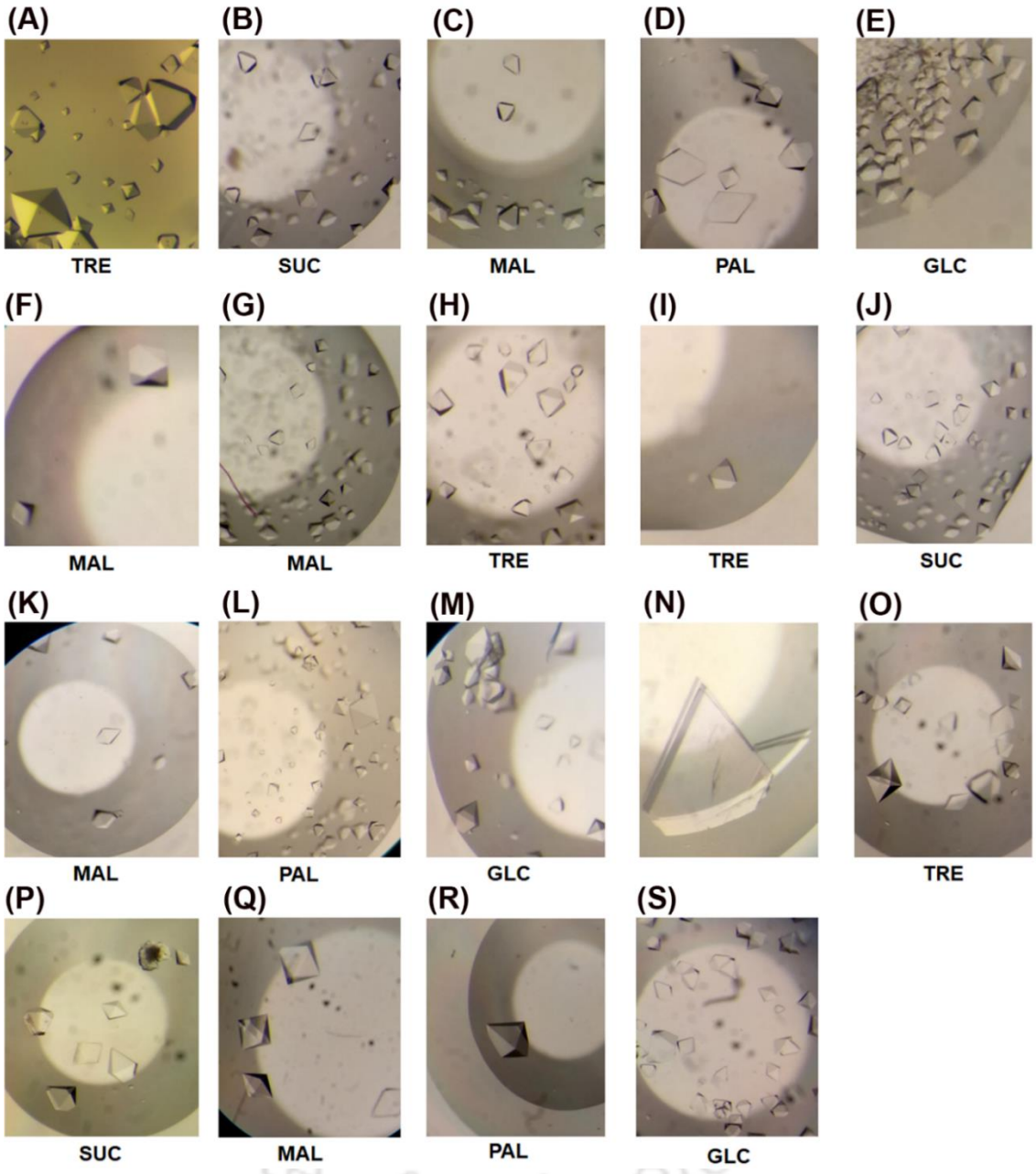


Figure 4.3. Crystallization of αGlyBP_WT and αGlyBP_mutant proteins. (A-E) Protein crystals of αGlyBP_WT complexed with different α-glycosides and glucose. Protein crystals of (F) αGlyBP_R49A, (G) αGlyBP_D118A and (H) αGlyBP_W287F. (I-M) Protein crystals of αGlyBP_W287A in complex with α-glycosides and glucose. (N) Protein crystals of αGlyBP_R356A in ligand free form. (O-S) Protein crystals of αGlyBP_R356A complexed with different α-glycosides and glucose.

4.2.4 Data collection, processing, structure solution, model building and refinement

X-ray intensity diffraction data for unliganded and ligand-bound forms of αGlyBP were collected at -173°C using the home source Rigaku MicroMax-007 HF diffractometer (operated at 40 kV and 30 mA) and R-Axis IV++ image-plate detector with rotating anode X-ray tube at wavelength 1.5418 Å available at Central Instruments Facility (CIF) of Indian Institute of Technology Guwahati, India. Prior to data collection, each single crystal was cryoprotected in mother liquor supplemented with 10-20% glycerol. The detector was kept at a distance range of 100-150 mm and oscillated with 1° followed by an exposure time of 300 s. The data sets of open and closed conformations of crystals were indexed, integrated and scaled using programs iMosflm and Aimless of CCP4 (Battye et al., 2011; Winn, 2011; Evans and Murshudov, 2013). The intensities were converted to structure factors using the program CTRUNCATE embedded in the package CCP4. Data collection statistics of all the crystals are provided in the supplementary Table 4.2-4.8. The three-dimensional structure of αGlyBP_WT protein was determined by the molecular replacement method using the program Phaser (McCoy et al., 2007). The crystal structure of TMBP from *Thermococcus litoralis* (PDB ID: 1EU8) having a sequence identity (query coverage) of 37% (97%) was used as the search model. The structures of αGlyBP_WT protein with different ligand-bound and αGlyBP_mutant protein in unliganded as well as liganded forms were also determined by the molecular replacement method with the program Phaser using three-dimensional atomic coordinates of αGlyBP_WT as the search model. For the calculation of R_{free}, 5% of the total reflections were kept aside from the data set (Brünger, 1992). The first cycle of structure refinement was done using the program Refmac5 (Vagin, 2004) available in CCP4 in which *Fo-Fc* electron density maps revealed the endogenously bound disaccharide α-glycoside (trehalose) in the active site. However, firstly the protein molecule was built using the program *Coot* (Emsley et al., 2010) and refinement using the program Refmac5 (Vagin, 2004) was performed. Subsequently, water molecules and other molecules (such as PEG, glycerol, etc.) from crystallization conditions were modeled and refined. Subsequently, the α-glycoside molecule was modeled and iterative cycles of refinement were carried out. The overall quality of all refined models was

checked and validated by programs PROCHECK (Laskowski et al., 1993) and MolProbity (Chen et al., 2010). The three-dimensional atomic coordinates and the structure factors of all the structures have been deposited in the RCSB Protein Data Bank (Berman et al., 2000) with accession codes 6J9W, 6J9Y, 6JAD, 6JAG, 6JAH, 6JAI, 6JAL, 6JAM, 6JAN, 6JAO, 6JAP, 6JAQ, 6JAR, 6JAZ, 6JB0, 6JB4, 6JBA, 6JBB and 6JBE for αGlyBP_WT•TRE, αGlyBP_WT•MAL, αGlyBP_WT•PAL, αGlyBP_WT•SUC, αGlyBP_WT•GLC, αGlyBP_D118A•MAL, αGlyBP_R356A, αGlyBP_R356A•TRE, αGlyBP_R356A•MAL, αGlyBP_R356A•PAL, αGlyBP_R356A•SUC, αGlyBP_R356A•GLC, αGlyBP_R49A•MAL, αGlyBP_W287F•TRE, αGlyBP_W287A•TRE, αGlyBP_W287A•MAL, αGlyBP_W287A•PAL, αGlyBP_W287A•SUC and αGlyBP_W287A•GLC, respectively. All the molecular graphics figures of protein structures were prepared using PyMOL (Molecular Graphics System, Version 1.7.2.2 Schrödinger, LLC).

Table 4.2. Data collection and refinement statistics of αGlyBP_WT protein (bound to trehalose, maltose and palatinose). The values in parenthesis are for the last resolution shell.

	αGlyBP_WT•TRE	**αGlyBP_WT•MAL**	**αGlyBP_WT•PAL**
Wavelength (Å)	1.5418	1.5418	1.5418
Temperature (K)	100	100	100
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Unit-cell parameters (Å, °)	$a=b=84.91$, $c=145.80$, $\alpha=\beta=\gamma=90$	$a=b=85.24$, $c=145.68$, $\alpha=\beta=\gamma=90$	$a=b=84.77$, $c=145.83$, $\alpha=\beta=\gamma=90$
Resolution (Å)	48.60-1.80 (1.84-1.80)	46.44-1.63 (1.66-1.63)	59.94-1.90 (1.94-1.90)
No. of observed reflections	611481 (35139)	611875 (26380)	279983 (18724)
No. of unique reflections	50048 (2860)	67774 (3302)	42666 (2695)
Mn(I) CC(1/2)	0.999 (0.954)	0.998 (0.970)	0.997 (0.915)
Completeness (%)	99.8 (98.9)	100 (99.9)	99.9 (99.6)
V_M ($Å^3$ Da^{-1})	2.90	2.92	2.89
Solvent content (%)	57.63	57.93	57.50

Mosaicity (°)	0.461	0.422	0.639
MeanI/σ(I)	18.8 (5.8)	18.5 (5.5)	13.2 (3.6)
R_{merge}† (%)	0.088 (0.435)	0.067 (0.242)	0.083 (0.449)
R_{pim} (%)	0.037 (0.185)	0.033 (0.135)	0.050 (0.268)
R_{meas} (%)	0.095 (0.473)	0.075 (0.278)	0.097 (0.524)
Multiplicity	12.2 (12.3)	9.0 (8.0)	6.6 (6.9)
R_{work}/R_{free} (%)	13.25/16.41	13.69/16.27	13.94/18.16
Protein model			
No. of subunits in ASU	1	1	1
Protein atoms	3242	3274	3250
Water molecules	533	575	517
Trehalose	1	-	-
Maltose	-	1	-
Palatinose	-	-	1
Others	2	3	2
Deviation from ideal geometry			
Bond length (Å)	0.020	0.022	0.018
Bond angles (°)	1.848	2.057	1.804
Average *B*-factor (Å²)			
Protein atoms	16.73	16.19	21.60
Water molecules	30.60	31.05	34.18
Trehalose	11.56	-	-
Maltose	-	11.88	-
Palatinose	-	-	15.75
Ramachandran plot			
Favored (%)	97.73	98.47	98.49
Allowed (%)	1.77	1.27	1.26
Remaining (%)	0.51	0.25	0.25
PDB id	6J9W	6J9Y	6JAD

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

TRE: Trehalose; MAL: Maltose; PAL: Palatinose

Table 4.3. Data collection and refinement statistics of αGlyBP_WT (bound to sucrose and glucose) and αGlyBP_D118A mutant protein (bound to maltose). The values in parenthesis are for the last resolution shell.

	αGlyBP_WT•SUC	**αGlyBP_WT•GLC**	**αGlyBP_D118A•MAL**
Wavelength (Å)	1.5418	1.5418	1.5418

Temperature (K)	100	100	100
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Unit-cell parameters (Å, °)	$a=b$=84.93, c=145.95, $\alpha=\beta=\gamma$=90	$a=b$= 84.23, c=145.90, $\alpha=\beta=\gamma$=90	$a=b$= 84.72, c=146.02, $\alpha=\beta=\gamma$=90
Resolution (Å)	60.05-1.85 (1.89-1.85)	48.63-1.63 (1.66-1.63)	59.91-2.10 (2.16-2.10)
No. of observed reflections	823750 (47317)	1181510 (51925)	493303 (38585)
No. of unique reflections	46411 (2826)	66385 (3266)	31860 (2553)
Mn(I) CC(1/2)	0.999 (0.944)	0.999 (0.929)	0.998 (0.957)
Completeness (%)	100 (100)	100 (100)	100 (100)
V_M ($Å^3$ Da^{-1})	2.91	2.86	2.89
Solvent content (%)	57.70	56.98	57.51
Mosaicity (°)	0.573	0.521	0.730
Mean I/σ(I)	24.3 (5.8)	22.3 (4.3)	17.5 (6.0)
R_{merge}† (%)	0.107 (0.573)	0.082 (0.647)	0.126 (0.481)
R_{pim} (%)	0.036 (0.203)	0.028 (0.239)	0.046 (0.180)
R_{meas} (%)	0.113 (0.608)	0.087 (0.690)	0.134 (0.514)
Multiplicity	17.7 (16.7)	17.8 (15.9)	15.5 (15.1)
R_{work}/R_{free} (%)	12.93/16.68	13.48/16.05	14.17/18.99
Protein model			
No. of subunits in ASU	1	1	1
Protein atoms	3255	3258	3246
Water molecules	570	580	448
Sucrose	1	-	-
Glucose	-	1	-
Maltose	-	-	1
Calcium ion	-	-	1
Others	2	7	3
Deviation from ideal geometry			
Bond length (Å)	0.019	0.021	0.017
Bond angles (°)	1.801	1.990	1.705
Average *B*-factor ($Å^2$)			
Protein atoms	14.93	17.28	20.22
Water molecules	28.80	32.41	31.35
Sucrose	8.12	-	-
Glucose	-	10.12	-

Maltose	-	-	21.46
Calcium ion	-	-	18.07
Ramachandran plot			
Favored (%)	97.73	97.73	97.78
Allowed (%)	1.77	1.77	1.51
Remaining (%)	0.51	0.51	0.50
PDB id	6JAG	6JAH	6JAI

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

Table 4.4. Data collection and refinement statistics of αGlyBP_R356A mutant protein in open and closed conformation (bound to trehalose and maltose). The values in parenthesis are for the last resolution shell.

	αGlyBP_R356A	**αGlyBP_R356A•TRE**	**αGlyBP_R356A•MAL**
Wavelength (Å)	1.5418	1.5418	1.5418
Temperature (K)	100	100	100
Space group	$P2_1$	$P4_12_12$	$P4_12_12$
Unit-cell parameters (Å, °)	a= 53.04, b= 85.13, c=55.34, α=γ=90, β=92.99	a=b=85.07, c=146.38, α=β=γ=90	a=b=84.41, c=145.55, α=β=γ=90
Resolution (Å)	44.97-1.56 (1.59-1.56)	48.79-1.63 (1.66-1.63)	59.69-1.77 (1.81-1.77)
No. of observed reflections	318860 (14055)	1241773 (53313)	771334 (40657)
No. of unique reflections	69683 (3351)	67875 (3277)	49679 (2550)
Mn(I) CC(1/2)	0.999 (0.994)	0.999 (0.953)	0.999 (0.967)
Completeness (%)	99.8 (96.4)	100 (100)	96.0 (87.8)
V_M ($Å^3$ Da^{-1})	2.76	2.92	2.86
Solvent content (%)	55.39	57.96	57.06
Mosaicity (°)	0.357	0.542	0.559
Mean I/σ(I)	30.1 (15.0)	22.1 (4.6)	22.1 (6.2)
R_{merge}† (%)	0.031 (0.061)	0.078 (0.532)	0.083 (0.389)
R_{pim} (%)	0.024 (0.048)	0.026 (0.194)	0.029 (0.137)
R_{meas} (%)	0.040 (0.077)	0.082 (0.567)	0.089 (0.413)

Multiplicity	4.6 (4.2)	18.3 (16.3)	15.5 (15.9)
R_{work}/R_{free} (%)	12.71/15.05	13.81/16.59	13.25/16.35
Protein model			
No. of subunits in ASU	1	1	1
Protein atoms	3396	3279	3251
Water molecules	666	587	528
Trehalose	-	1	-
Maltose	-	-	1
Others	7	6	7
Deviation from ideal geometry			
Bond length (Å)	0.022	0.021	0.019
Bond angles (°)	2.155	1.955	1.856
Average *B*-factor (Å^2)			
Protein atoms	10.39	18.23	16.50
Water molecules	25.21	32.83	30.50
Trehalose	-	12.59	-
Maltose	-	-	10.52
Ramachandran plot			
Favored (%)	98.75	97.73	97.98
Allowed (%)	1.00	1.77	1.77
Remaining (%)	0.25	0.51	0.25
PDB id	6JAL	6JAM	6JAN

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

TRE: Trehalose; MAL: Maltose

Table 4.5. Data collection and refinement statistics of αGlyBP_R356A mutant protein in closed conformation (bound to palatinose, sucrose and glucose). The values in parenthesis are for the last resolution shell.

	αGlyBP_R356A•PAL	**αGlyBP_R356A•SUC**	**αGlyBP_R356A•GLC**
Wavelength (Å)	1.5418	1.5418	1.5418
Temperature (K)	100	100	100
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Unit-cell parameters (Å, °)	$a=b$=85.42, c=146.15, $\alpha=\beta=\gamma$=90	$a=b$= 85.41, c=146.29, $\alpha=\beta=\gamma$=90	$a=b$=84.62, c=146.23, $\alpha=\beta=\gamma$=90

Resolution (Å)	60.40-1.77 (1.81-1.77)	60.39-1.77 (1.81-1.77)	59.84-1.95 (2.00-1.95)
No. of observed reflections	940791 (48057)	911320 (46820)	580422 (40079)
No. of unique reflections	53234 (2987)	53038 (2950)	39565 (2725)
Mn(I) CC(1/2)	0.999 (0.982)	0.999 (0.986)	0.999 (0.952)
Completeness (%)	100 (100)	99.7 (98.6)	100 (100)
V_M ($Å^3$ Da^{-1})	2.94	2.95	2.89
Solvent content (%)	58.24	58.27	57.47
Mosaicity (°)	0.481	0.421	0.388
Mean I/σ(I)	24.0 (6.1)	31.2 (8.8)	22.2 (6.0)
R_{merge}† (%)	0.080 (0.302)	0.063 (0.247)	0.108 (0.483)
R_{pim} (%)	0.027 (0.111)	0.022 (0.090)	0.041 (0.185)
R_{meas} (%)	0.085 (0.322)	0.067 (0.263)	0.116 (0.517)
Multiplicity	17.7 (16.1)	17.2 (15.9)	14.7 (14.7)
R_{work}/R_{free} (%)	15.49/19.94	15.04/19.53	13.14/16.71
Protein model			
No. of subunits in ASU	1	1	1
Protein atoms	3273	3274	3247
Water molecules	557	556	531
Palatinose	1	-	-
Sucrose	-	1	-
Glucose	-	-	1
Others	4	3	5
Deviation from ideal geometry			
Bond length (Å)	0.018	0.018	0.019
Bond angles (°)	1.827	1.799	1.783
Average *B*-factor ($Å^2$)			
Protein atoms	19.91	17.81	15.49
Water molecules	34.33	32.16	28.71
Palatinose	12.68	-	-
Sucrose	-	10.13	-
Glucose	-	-	8.79
Ramachandran plot			
Favored (%)	98.47	97.96	97.97
Allowed (%)	1.27	1.53	1.52
Remaining (%)	0.25	0.51	0.51
PDB id	6JAO	6JAP	6JAQ

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

Table 4.6. Data collection and refinement statistics of αGlyBP_R49A (bound to maltose), αGlyBP_W287F and αGlyBP_W287A mutant proteins (bound to trehalose). The values in parenthesis are for the last resolution shell.

	αGlyBP_R49A•MAL	**αGlyBP_W287F•TRE**	**αGlyBP_W287A•TRE**
Wavelength (Å)	1.5418	1.5418	1.5418
Temperature (K)	100	100	100
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Unit-cell parameters (Å, °)	$a=b=84.98$, $c=145.46$, $\alpha=\beta=\gamma=90$	$a=b=85.27$, $c=145.65$, $\alpha=\beta=\gamma=90$	$a=b=84.76$, $c=146.21$, $\alpha=\beta=\gamma=90$
Resolution (Å)	48.49-1.63 (1.66-1.63)	48.55-1.85 (1.89-1.85)	48.74-1.63 (1.66-1.63)
No. of observed reflections	1006044 (44383)	690786 (41102)	804773 (35559)
No. of unique reflections	61325 (2721)	46653 (2839)	65184 (2980)
Mn(I) CC(1/2)	0.999 (0.954)	0.999 (0.949)	0.999 (0.942)
Completeness (%)	91.3 (83.4)	100 (99.9)	96.9 (91.5)
V_M ($Å^3$ Da^{-1})	2.90	2.92	2.90
Solvent content (%)	57.61	57.95	57.61
Mosaicity (°)	0.558	0.370	0.332
Mean I/σ(I)	24.3 (5.1)	21.6 (5.8)	24.4 (5.7)
R_{merge}† (%)	0.064 (0.391)	0.092 (0.511)	0.061 (0.381)
R_{pim} (%)	0.022 (0.139)	0.035 (0.196)	0.025 (0.160)
R_{meas} (%)	0.068 (0.416)	0.098 (0.547)	0.066 (0.414)
Multiplicity	16.4 (16.3)	14.8 (14.5)	12.3 (11.9)
R_{work}/R_{free} (%)	13.99/16.62	13.26/16.83	13.42/16.02
Protein model			

No. of subunits in ASU	1	1	1
Protein atoms	3262	3232	3249
Water molecules	601	539	620
Maltose	1	-	-
Trehalose	-	1	1
Others	3	2	4
Deviation from ideal geometry			
Bond length (Å)	0.020	0.019	0.020
Bond angles (°)	1.932	1.835	1.947
Average *B*-factor (Å^2)			
Protein atoms	17.85	17.06	14.81
Water molecules	33.00	30.34	30.05
Maltose	17.93	-	-
Trehalose	-	18.36	9.46
Ramachandran plot			
Favored (%)	97.96	97.97	97.73
Allowed (%)	1.53	1.52	1.77
Remaining (%)	0.51	0.51	0.51
PDB id	6JAR	6JAZ	6JB0

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

Table 4.7. Data collection and refinement statistics of αGlyBP_W287A mutant protein (bound to maltose, palatinose and sucrose). The values in parenthesis are for the last resolution shell.

	αGlyBP_W287A•MAL	**αGlyBP_W287A•PAL**	**αGlyBP_W287A•SUC**
Wavelength (Å)	1.5418	1.5418	1.5418
Temperature (K)	100	100	100
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$

Unit-cell parameters (Å, °)	$a=b=85.13$, $c=145.54$, $\alpha=\beta=\gamma=90$	$a=b=84.10$, $c=146.02$, $\alpha=\beta=\gamma=90$	$a=b=84.48$, $c=145.93$, $\alpha=\beta=\gamma=90$
Resolution (Å)	48.51-1.63 (1.66-1.63)	59.47-2.00 (2.05-2.00)	59.74-1.95 (2.00-1.95)
No. of observed reflections	628881 (27856)	308917 (22315)	311670 (20698)
No. of unique reflections	67595 (3297)	36237 (2647)	38891 (2614)
Mn(I) CC(1/2)	0.999 (0.891)	0.990 (0.917)	0.997 (0.869)
Completeness (%)	100 (100)	100 (100)	99.0 (96.9)
V_M ($Å^3$ Da^{-1})	2.91	2.85	2.87
Solvent content (%)	57.78	56.88	57.24
Mosaicity (°)	0.400	0.60	0.590
Mean I/σ(I)	17.1 (3.9)	10.4 (3.4)	13.4 (3.2)
R_{merge}† (%)	0.074 (0.457)	0.140 (0.574)	0.118 (0.557)
R_{pim} (%)	0.037 (0.246)	0.073 (0.308)	0.061 (0.285)
R_{meas} (%)	0.082 (0.520)	0.158 (0.653)	0.133 (0.629)
Multiplicity	9.3 (8.4)	8.5 (8.4)	8.0 (7.9)
R_{work}/R_{free} (%)	13.68/16.17	15.06/18.63	14.27/17.69
Protein model			
No. of subunits in ASU	1	1	1
Protein atoms	3246	3254	3241
Water molecules	593	448	473
Maltose	1	-	-
Palatinose	-	1	-
Sucrose	-	-	1
Others	5	4	5
Deviation from ideal geometry			
Bond length (Å)	0.020	0.017	0.017
Bond angles (°)	1.876	1.787	1.796
Average *B*-factor ($Å^2$)			
Protein atoms	16.13	22.37	18.07
Water molecules	30.57	33.84	30.21
Maltose	11.31	-	-

Palatinose	-	20.84	-
Sucrose	-	-	10.82
Ramachandran plot			
Favored (%)	97.73	97.72	97.48
Allowed (%)	1.77	1.77	2.02
Remaining (%)	0.51	0.51	0.50
PDB id	6JB4	6JBA	6JBB

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

Table 4.8. Data collection and refinement statistics of αGlyBP_W287A mutant protein (bound to glucose). The values in parenthesis are for the last resolution shell.

	αGlyBP_W287A•GLC
Wavelength (Å)	1.5418
Temperature (K)	100
Space group	$P4_12_12$
Unit-cell parameters (Å, °)	$a=b=84.99$, $c=145.83$, $\alpha=\beta=\gamma=90$
Resolution (Å)	55.56-1.75 (1.78-1.75)
No. of observed reflections	645449 (32906)
No. of unique reflections	52985 (2590)
Mn(I) CC(1/2)	0.999 (0.918)
Completeness (%)	97.0 (88.3)
V_M ($Å^3$ Da^{-1})	2.91
Solvent content (%)	57.72
Mosaicity (°)	0.448
Mean I/σ(I)	17.4 (3.9)
R_{merge}† (%)	0.092 (0.606)
R_{pim} (%)	0.037 (0.238)
R_{meas} (%)	0.099 (0.652)
Multiplicity	12.2 (12.7)
R_{work}/R_{free} (%)	13.89/17.57
Protein model	
No. of subunits in ASU	1
Protein atoms	3246
Water molecules	573
Glucose	1
Others	6
Deviation from ideal geometry	
Bond length (Å)	0.018

Bond angles (°)	1.770
Average *B*-factor (Å^2)	
Protein atoms	18.64
Water molecules	32.31
Glucose	10.92
Ramachandran plot	
Favored (%)	97.73
Allowed (%)	1.77
Remaining (%)	0.51
PDB id	6JBE

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

4.2.5 Isothermal titration calorimetry

All the isothermal titration calorimetry (ITC) experiments were carried out at 25°C using MicroCal iTC200 (Malvern, UK). For each titration, protein and ligand solutions were prepared and diluted in the dialyzed buffer containing 20 mM Tris-HCl pH 7.5 and 150 mM NaCl. Owing to the capability of αGlyBP to bind trehalose endogenously, thermodynamic parameter of αGlyBP_WT for ligand selection were determined by the displacement titration method. Prior to titration, endogenously bound trehalose was replaced by glucose via dialysis buffer containing 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 1% glucose followed by second and third dialysis in buffer containing 20 mM Tris-HCl pH 7.5 and 150 mM NaCl. After that, glucose-treated αGlyBP_WT was titrated with different disaccharide sugars (mentioned in Table 4.9) to confirm the specificity of αGlyBP_WT for α-glycosides. Once the specificity for α-glycosides has been confirmed, measurement of binding kinetics and thermodynamic parameters of α-glycosides were carried out with the ligand free mutant proteins such as αGlyBP_D70A, αGlyBP_W287A and αGlyBP_R356A. All three mutant proteins were titrated with trehalose, maltose, glucose, sucrose and palatinose where protein and ligand solution with desired concentrations (mentioned in Table 4.9) were loaded into the sample cell and syringe, respectively. In each titration experiment, after pre-injection (0.4 µl), 1.5 µl of ligand solution was injected 24 times into the sample cell at an interval of 120 s and stirred at 250 rpm for homogenous mixing. Apart from titrating sugar, binding energetics of metal ions with αGlyBP_D118A mutant protein was also estimated using the same parameters.

Prior to titration, the protein was treated with EDTA to remove the endogenously bound calcium (Ca^{2+}) ion. For EDTA treatment, purified protein was dialyzed against buffer containing 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 10 mM EDTA followed by second and third dialysis against a buffer containing 20 mM Tris-HCl pH 7.5 and 150 mM NaCl and concentrated up to ~21 mg ml^{-1}. After that, EDTA-treated mutant protein was titrated with $CaCl_2$, $MgCl_2$, $ZnCl_2$ and $MnSO_4$ solution. In all the titration experiment, concentrations of active proteins mentioned in Table 4.9 were determined by curve fitting by adjusting their concentrations to a value resulting in stoichiometric ratio of ligand to protein at half saturation. To get the truly integrated data of heat generation associated with the complex formation, each raw data was subtracted for the heat of dilution by titrating the sugar and metal solution into the dialyzed buffer. Integrated calorimetric data were analyzed by using the software Origin (version 7.0) and fitted into one- or two- site binding models followed by non-linear sigmoidal curve fitting. This fitting curve yields the binding constant (K_a), binding stoichiometry (n) and binding enthalpy (ΔH). In addition, other thermodynamic parameters including a change in Gibbs free energy (ΔG) and entropy ($T\Delta S$) were calculated using the equation $\Delta G = \Delta H - T\Delta S$.

4.2.6 Bioinformatics analysis

Structural homologs of the protein αGlyBP were identified by using the web server Dali (Holm and Rosenstrom, 2010). The three-dimensional atomic coordinates of each homolog protein was retrieved from the Protein Data Bank (PDB) and subsequently utilized for structure-based sequence alignment using the program PROMALS3D (Pei and Grishin, 2014). The clarity of sequence alignment was further improved by the web tool ESPript3.0 (Gouet et al., 2003). In addition to structural homologs, sequence-based homologs were identified by performing BLASTP (Altschul et al., 1990) search against the RefSeq database (O'Leary et al., 2016). Retrieved homologs were further sorted based on sequence identity (>30%) and query coverage (>90%). The resulting 93 sequences were aligned together using the program ClustalW embedded in MEGA7 and subjected to phylogenetic tree construction using the neighbor joining (NJ) method with 1000 bootstrap replicates in the program MEGA7 (Kumar et al., 2016). Further, the organization of genes for ABC transport system along with flanking genes was analyzed

by using genomic viewer in the Gene database of National Center for Biotechnology Information (NCBI) search database. The significance of the functional association of ABC transport system with flanking genes was probed by retrieving the information of each gene from the UniProtKB database (The UniProt Consortium, 2019).

4.3 RESULTS

4.3.1 The overall structure and the active site of αGlyBP

The three-dimensional crystal structure of αGlyBP was determined at 1.8 Å resolution in the space group $P4_12_12$. The protein αGlyBP has a bilobate structure and is composed of N- and C-terminal domains (NTD and CTD, respectively) linked by a tripartite hinge region (Figure 4.4A). The NTD is made up of two small subdomains (N1, residues 3-116 and N2, residues 287-351) constituting of 11 α-helices and five antiparallel β-strands. Similarly, the CTD comprises of two subdomains (C1, residues 121-268 and C2, residues 365-405) containing ten α-helices and three antiparallel β-strands. Each subdomain is connected via three loops L1 (residue 117-120), L2 (residue 269-286) and L3 (residue 352-364), which serve as hinge 1, hinge 2 and hinge 3, respectively. Despite the parallel arrangement of NTD and CTD, subdomains are linked diagonally, where hinge 1, hinge 2 and hinge 3 connect the N1-C1, C1-N2 and N2-C2 subdomains, respectively (Figure 4.4B). Structurally, αGlyBP shares the highest structural homology with member of sub-cluster D-I SBPs such as maltose-binding protein MalE3 (PDB ID: 6DTQ, RMSD: 1.3 Å, Z-score: 57.4), trehalose/maltose-binding protein (TMBP, PDB ID: 1EU8, RMSD: 1.3 Å, Z-score: 57.1), extracellular solute-binding protein family 1 (PDB ID: 5CI5, RMSD: 2.1 Å, Z-score: 46.3), acarbose/maltose-binding protein GacH (PDB ID: 3K00, RMSD: 1.9 Å, Z-score: 46.1), ABC-type sugar transporter (PDB ID: 4QRZ, RMSD: 2.0 Å, Z-score: 45.5) and sugar ABC transporter (PDB ID: 5IAI, RMSD: 2.4 Å, Z-score: 45).

Interestingly, a trehalose molecule was found to be endogenously bound at the active site of αGlyBP near the tripartite hinges occupying the cleft formed by the subdomains (N1-N2-C1-C2) (Figure 4.4C). Considering the highest homology with multi-substrate TMBP

which is a SBP of ABC transport system (ORF: TTC1627 from *T. thermophilus* HB27 with sequence identity and query coverage of 100%), the protein αGlyBP was also examined for multi-substrate specificity. For this, attempts were made to get rid of the endogenously-bound trehalose from the protein by using denaturants such as 8M urea and 6M guanidine hydrochloride. However, denaturation of the protein was unsuccessful most likely due to its high thermal stability and the presence of a stabilizing molecule i.e. trehalose in its active site. Thus, to achieve the structure of αGlyBP in complex with other sugar molecules, co-crystallization was performed with a diverse range of sugars e.g. disaccharide (e.g. maltose), hexoses (e.g. glucose, its epimers and derivatives) and pentoses (e.g. D- and L-arabinose) anticipating the replacement of trehalose from the active site of the protein molecule. Among all the sugars used, a high concentration of maltose (35.3 mM) and glucose (35.3 mM) were able to replace the bound trehalose from the active site. Since homologous TMBP can transport sucrose and palatinose other than trehalose and maltose, αGlyBP was also explored for sucrose and palatinose binding. To obtain the structure of αGlyBP in complex with sucrose and palatinose, endogenously-bound trehalose in the active site was initially replaced by glucose followed by the replacement of glucose with a higher concentration of sucrose and palatinose.

All the complex structures of αGlyBP bound to trehalose, maltose, glucose, sucrose and palatinose are well superimposable upon each other with an RMSD of 0.1 Å. A common feature of all the disaccharide sugars interacting with αGlyBP is the presence of an α-linked glycosyl unit (hereafter referred to as disaccharide α-glycosides). One of the unit, Glc1 of all the disaccharide α-glycosides as well as glucose binds at the subsite B in a similar manner while the other unit, Glc2 (in trehalose and maltose) or Fru1 (in sucrose and palatinose) binds differentially at subsite A (Figure 4.4C-4.4G). In case of αGlyBP_WT•GLC, glucose occupies the subsite B while the subsite A is filled with water molecules (Figure 4.4H). In all the complex structures, two aromatic residues Trp248 (NTD) and Trp287 (CTD) provide the hydrophobic stacking interaction to the pyranose ring of Glc1 unit of all disaccharide α-glycosides and glucose. Along with hydrophobic stacking interaction, the hydroxyl group of Glc1 unit and glucose also forms identical hydrogen bonding with Asp70 (NTD), Asp118 (hinge 1), Gly286 (hinge 2), Arg356

(hinge 3), Glu230, Trp248 (CTD) and a water molecule (Figure 4.4C-4.4G, Table B.1). In contrast to Glc1 unit, the hydrogen-bonding pattern of Glc2 unit (in trehalose and maltose) or Fru1 unit (in sucrose and palatinose) is not conserved among themselves. Even though the interacting residues Asp11, Arg49, Arg323 (NTD) and Asp118 (hinge 1) remain same, the interacting hydroxyl group of the Glc2 or Fru1 are different (Table B.1).

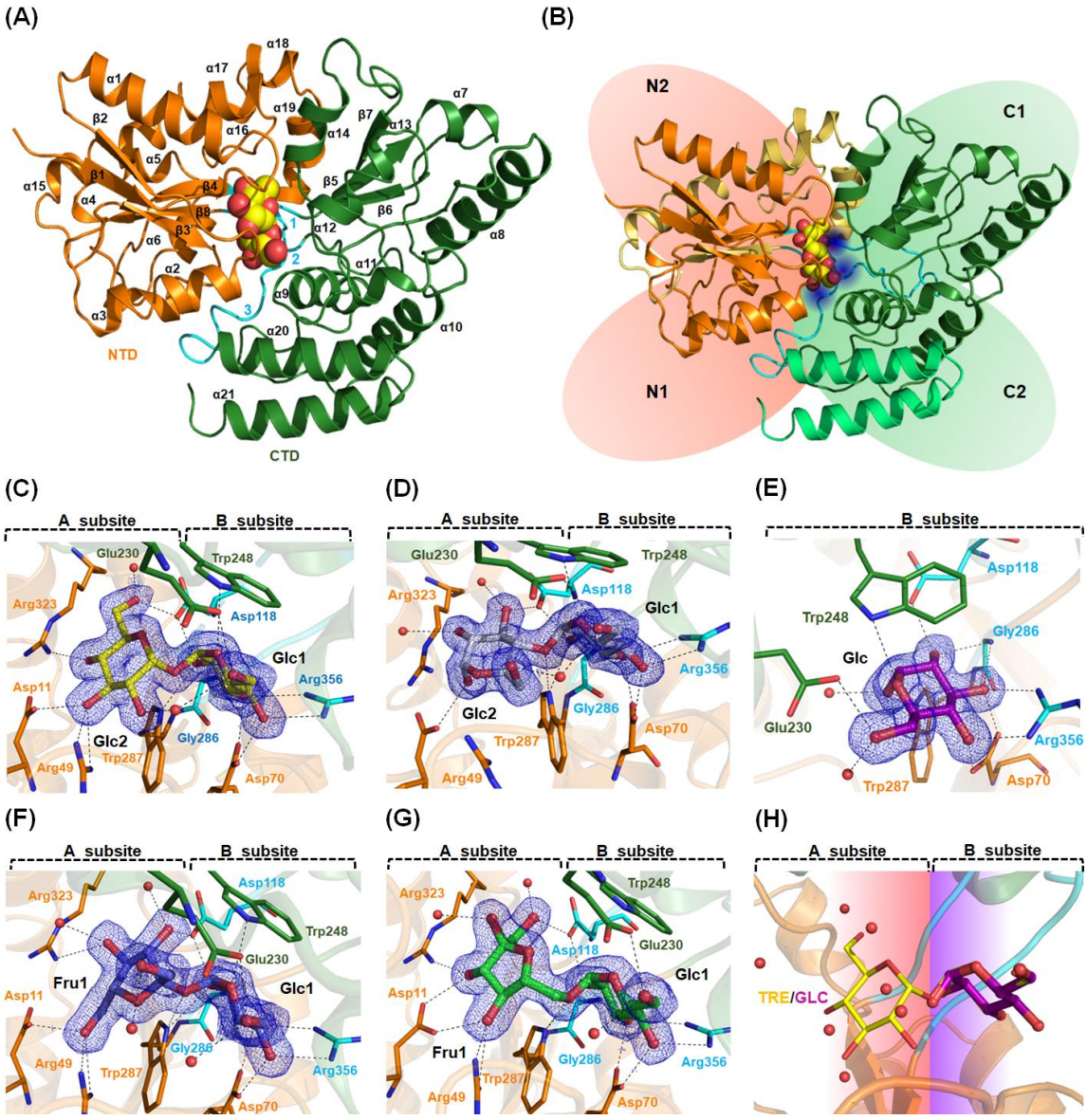


Figure 4.4. Three-dimensional structure and the active site of αGlyBP. (A) The overall topology of αGlyBP with endogenously-bound trehalose (spheres in yellow).

Three regions (NTD, CTD and hinge) of αGlyBP are represented in orange, green and cyan, respectively. (B) Subdomains N1, N2, C1 and C2 are demarcated with orange and green background. (C-G) αGlyBP bound to trehalose, maltose, glucose, sucrose and palatinose, respectively. All the NTD, CTD and hinge region residues interacting with ligand molecules via hydrogen bonding (dotted lines) are shown as color-coded line model and labeled. The *2Fo−Fc* electron density maps of bound sugars contoured at 1.6σ are represented as a blue mesh. (H) Comparison of the two subsites A and B of αGlyBP_WT•GLC and αGlyBP_WT•TRE. Water molecules occupying the subsite A are represented in red spheres while the bound glucose (subsite B) and trehalose (subsites A and B) are shown in violet and yellow, respectively.

4.3.2. αGlyBP exhibits stereo- and glycosidic-linkage selectivity

Apart from glucose, none of the other monosaccharides such as glucose epimers, glucose derivatives and pentose sugars could replace the endogenously-bound trehalose. Thus, to obtain a more in-depth insight at the atomic level, glucose was manually replaced by each sugar at the active site using the program *Coot* (Emsley et al., 2010). In case of galactose and mannose, the pyranose ring forms stacking interaction with two tryptophan residues (Trp248 and Trp287) and all the hydroxyl groups form hydrogen bond with the active-site residues similar to glucose. The only observable structural difference is identified at the epimeric hydroxyl group at C4 and C2 positions which led to the loss of interaction with hinge residues Arg356 (hinge 3) and Asp118 (hinge 1), respectively (Figure 4.5A and 4.5B). This demonstrates the stereoselective nature of the protein for ligand binding, where the interaction with the hinge region is crucial for binding. Similarly, an analysis for glucose derivatives shows a higher number of hydrogen bonding for glucose-1-phosphate and glucuronic acid indicating a stronger binding than glucose. However, the requirement of a positive charge to balance the negative charge of substituted groups (phosphate and carboxylate group) is not fulfilled and thus forbid the binding of glucose derivatives (Figure 4.5C and 4.5D). Unlike hexose sugars, pentose sugar cannot interact with Gly286 (hinge 2) and Arg356 (hinge 3) residues and thus forms lesser hydrogen bonds resulting in the loss of its binding to the protein (Figure 4.5E).

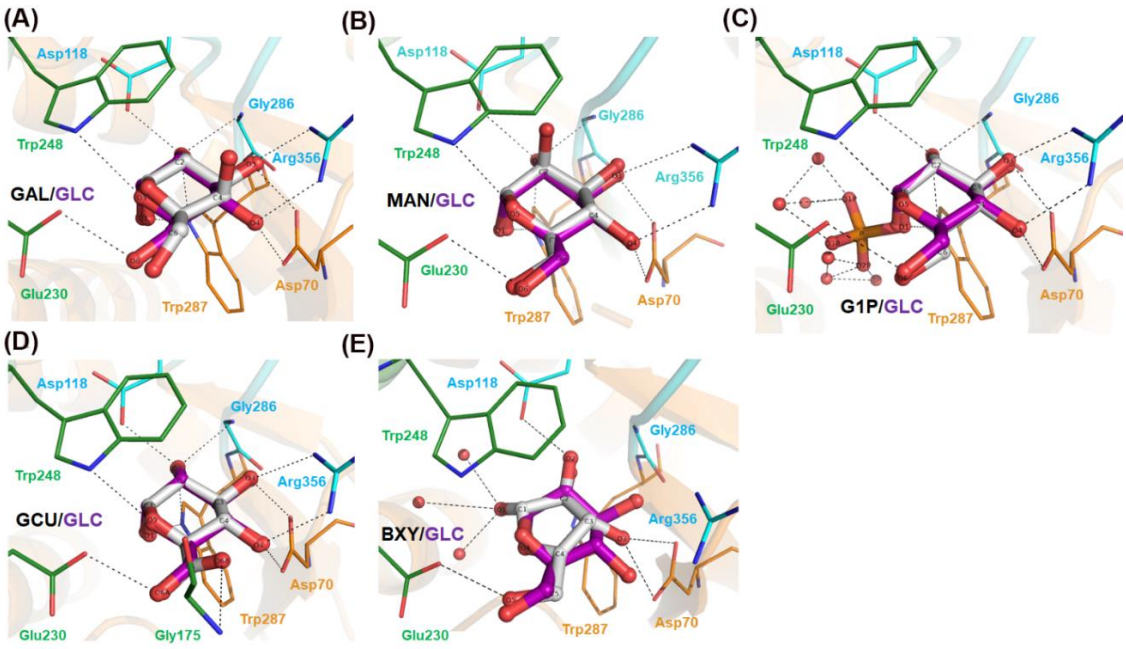


Figure 4.5. Stereo-adaptation of αGlyBP for glucose binding. Superimposition of glucose (GLC) bound to the active site of αGlyBP_WT•GLC with modeled (A) galactose (GAL), (B) mannose (MAN), (C) glucose-1-phosphate (G1P), (D) glucoronic acid (GCU) and (E) arabinose (BXY). The glucose bound to crystal structure and the ligands modeled are represented as ball-and-stick model in purple and grey, respectively. The coordinating residues from NTD, CTD and hinge involved in hydrogen bonding (dotted lines) with ligand molecule are shown in orange, cyan and green lines, respectively.

In addition to maintaining the stereoselectivity, αGlyBP is hypothesized for maintaining the selectivity for glycosidic linkage and thus possesses the specificity for disaccharide α-glycosides. To further affirm the specificity of αGlyBP for α-glycosides, the thermodynamic parameters for both disaccharide α- and β-glycosides binding were measured using ITC experiments. Prior to titration, the endogenously-bound trehalose was replaced with glucose from the active site using dialysis. From the crystallographic experiment, it was obtained that glucose could be more readily replaced by disaccharide α-glycosides. Hence glucose-bound αGlyBP was further used for thermodynamic parameter measurements for the binding of disaccharide α- and β-glycosides during ITC experiments. Binding isotherms showed the replacement of glucose by α-glycosides such as trehalose (α-1,1), sucrose (α-1,2), maltose (α-1,4) and palatinose (α-1,6) with dissociation constant (K_d) in the range of ~7.4-48.3 μM (Table 4.9). This thermodynamic phenomenon corresponds well with the crystallographic data, where

replacement of water and glucose from the subsites A and B by α-glycosides is an endothermic process and is driven by unfavorable enthalpy change and positive entropy change due to large desolvation energy (Figure 4.6A-4.6D). In contrast to α-glycosides, β-glycosides such as sophorose (β-1,2), gentiobiose (β-1,3) and cellobiose (β-1,4) did not exhibit any heat change illustrating the high specificity of αGlyBP towards α-glycosides (Table 4.9 and Figure 4.6E-4.6G).

Table 4.9. Thermodynamic parameters of ligand binding to αGlyBP_WT, αGlyBP_R356A, αGlyBP_D70A, αGlyBP_W287A and αGlyBP_D118A mutant proteins. The values provided in parenthesis are the concentration of protein and ligand used during ITC experiments.

Protein (μM)	**Ligand (mM)**	**Stoichiometry ratio (n)**	**Association (K_a, M^{-1}) / Dissociation (K_d, μM) constant**	**ΔH**	**$T\Delta S$**	**ΔG**
				(kcal mol^{-1})		
αGlyBP_WT (GLC bound, 220)*	TRE (15)	1.36	$1.34x10^5 \pm 1.03x10^5$ / 7.46	6.02	12.99	-6.97
	SUC (15)	0.77	$3.17x10^4 \pm 3.40x10^3$ / 31.5	2.94	9.08	-6.14
	MAL (15)	1.17	$1.17x10^5 \pm 2.37x10^3$ / 8.54	7.69	14.60	-6.91
	PAL (15)	0.90	$2.07x10^4 \pm 782$ / 48.3	6.21	12.09	-5.88
	SOP (15)	N.D.	N.D.	N.D.	N.D.	N.D.
	GEN (15)	N.D.	N.D.	N.D.	N.D.	N.D.
	CEL (15)	N.D.	N.D.	N.D.	N.D.	N.D.
αGlyBP_R356A (200)**	TRE (3.6)	1.28	$1.68x10^5 \pm 4.80x10^4$ / 5.95	-1.02	6.10	-7.12
	SUC (6)	0.96	$6.24x10^4 \pm 9.23x10^3$/ 16.02	-4.26	2.27	-6.53
	MAL (3.6)	1.10	$6.21x10^5 \pm 4.16x10^5$/ 1.61	0.43	8.34	-7.91
	PAL (6)	0.93	$7.20x10^4 \pm 1.51x10^4$/ 13.8	-0.95	5.66	-6.61
	GLC (6.0)	1.0[a]	$8.99x10^3 \pm 2.99x10^3$/ 111.23	-3.06	2.32	-5.38
αGlyBP_D70A (200)**	TRE (20)	1.0[a]	636±34.4 / 1572.32	-5.49	-1.66	-3.83
	SUC (20)	1.0[a]	278±19.7 /3597.12	-3.78	-0.44	-3.34
	MAL (20)	1.21	$1.01x10^4 \pm 534$ / 99	-4.23	1.22	-5.45
	PAL (20)	1.0[a]	$1.08x10^3 \pm 42.0$ / 925.92	-6.07	-1.93	-4.14
	GLC (20)	1.0[a]	166±14.5 /6024.09	-6.74	-3.72	-3.02
αGlyBP_W287	TRE (2.2)	0.97	$1.72x10^7 \pm 2.37x10^7$/ 0.058	-5.70	4.17	-9.87
	SUC (2.2)	0.91	$9.64x10^5 \pm 2.62x10^5$/ 1.03	-9.56	-1.39	-8.17

A (150)**	MAL (4.4)	0.84[b]	$9.59x10^5±7.78x10^5$/ 1.04[b]	-1.83[b]	6.31[b]	8.14
	PAL (2.2)	0.99	$2.42x10^6±7.79x10^5$/ 0.41	-3.41	5.27	-8.68
	GLC (2.2)	1.29	$5.05x10^4±5.85x10^3$/ 19.80	-7.94	-1.53	-6.41
αGlyBP _D118A (100)	Ca^{2+} (1)	0.90	$1.99x10^6±4.27x10^5$/ 0.50	-12.7	-4.11	-8.59
	Mg^{2+} (1)	0.90	$3.48x10^5±8.42x10^4$/ 2.87	-2.13	5.42	-7.55
	Zn^{2+} (2)[c]	1.03	$2.92x10^5±6.70x10^4$/ 3.42	-57.52	-50.06	-7.46
	Mn^{2+} (10)	N.D.	N.D.	N.D.	N.D.	N.D.

GLC: Glucose; TRE: Trehalose; SUC: Sucrose; MAL: Maltose; PAL: Palatinose; SOP: Sophorose; GEN: Gentiobiose; CEL: Cellobiose; N.D.: Not Detected.

* Due to endogenous binding of ligand, thermodynamic parameter of glucose binding with WT proteins is unknown and thus restricted our attempts to fit the calorimetric graph in competitive binding model. All the titration curve for glucose replacement were fitted using one-site binding model.

**Since mutant proteins are ligand free, direct titration was performed to measure the thermodynamic parameter.

[a] Binding parameter was calculated using one binding site fitting model with fixed stoichiometry.

[b] Value for only first site having stoichiometry close to one has been provided for binding parameters as it was calculated using two-site binding model.

[c] Estimated ligand concentration based on estimated protein concentration.

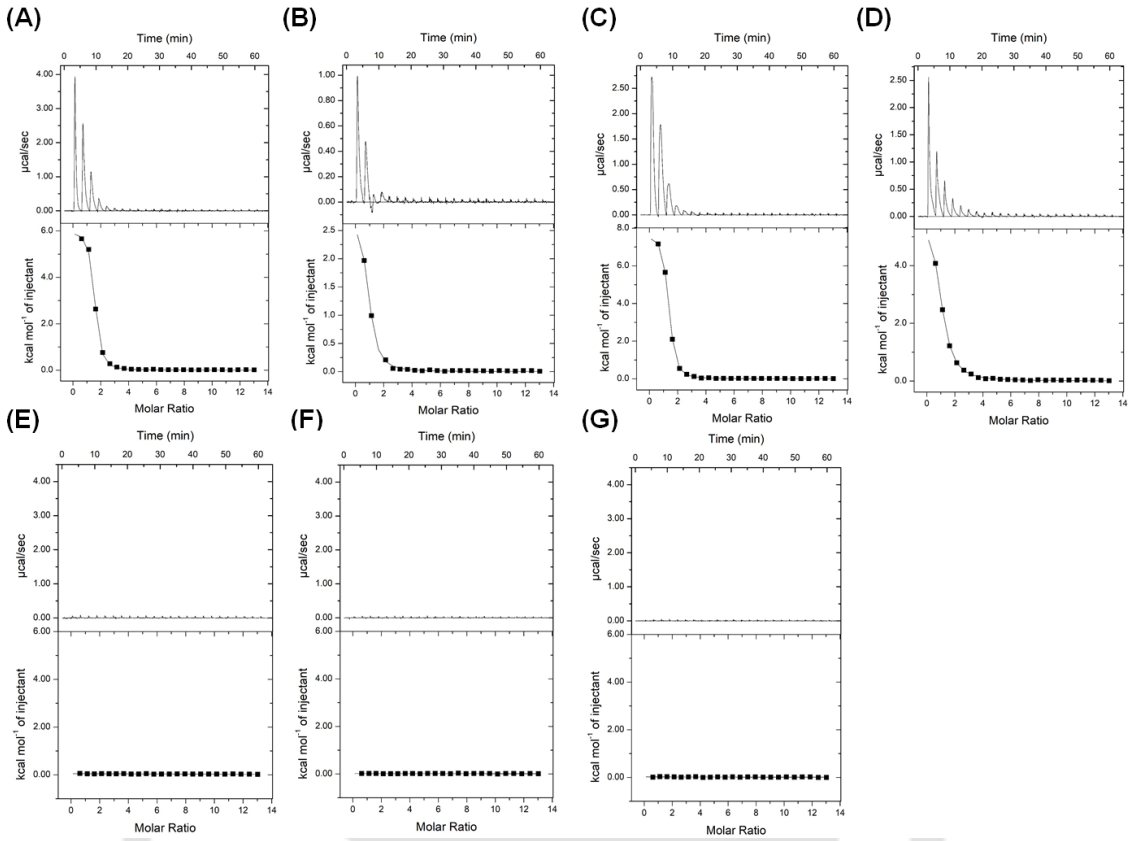


Figure 4.6. Binding energetics for replacement experiment in ITC. Replacement of glucose by (A-D) disaccharide α-glycosides e.g. trehalose, sucrose, maltose, palatinose and (E-G) β-glycosides e.g. sophorose, gentiobiose and cellobiose as measured in ITC. To improve the signals for replacement events, all the experiments were repeated with higher concentrations of protein (220 µM) and ligand (15 mM). In each thermogram, upper panel represents the heat trace released (black line) upon ligand titration, whereas lower panel represents the normalized integrated peak areas (black box) of each injection with the one-site binding model fit (black line). Thermodynamic parameters obtained from fitting model for individual titration experiments are mentioned in the Table 4.9.

4.3.3. N-terminal domain of αGlyBP dictates the open and closed conformations

Even though glucose is a monosaccharide and binds at the active site with lesser number of hydrogen bonds as compared to trehalose, it still retains the ability to replace and occupy the Glc1 position of trehalose at the subsite B. To assess the cause, a comparison of αGlyBP_WT•GLC was drawn with disaccharide α-glycosides (i.e. trehalose, sucrose, maltose and palatinose)-bound structures. Interestingly, at subsite B the Glc1 unit of all the disaccharide α-glycosides is absolutely superimposable on glucose and forms similar number of hydrogen bonds (Figure 4.7A and Table B.1). Furthermore, analysis of the

active site of the structural homolog Atu4361 protein bound to four different type of ligands viz. glycerol (PDB ID: 4QSE), sucrose (PDB ID: 4QSD), maltose (PDB ID: 4QSC) and maltotriose (PDB ID: 4QRZ) reveals a similar mode of binding. Despite the vast diversity of carbohydrates, both αGlyBP and Atu4361 proteins are specific to α-glycosides and conserves a similar binding mode. Hence, to further infer the cause of conserved binding mode for diverse ligands in both the proteins, glycerol-bound Atu4361 protein was compared with sugar-bound structures. Since, glycerol is the smallest ligand among the three, it can occupy any position in the active site. However, akin to glucose in αGlyBP_WT•GLC complex, all three hydroxyl group and carbon backbone of glycerol is well overlaid upon the O2, O3 and O4 oxygen atoms of α-glycosides (sucrose, maltose and maltotriose) (Figure 4.7B). In-depth investigation of the active site of Atu4361 reveals that the three hydroxyl groups of glycerol strongly interact with the residues Asp89, Asn142, Asp298 and Arg367 (Numbering according to Atu4361 protein). Notably, out of four interacting residues, three (Asn142, Asp298, and Arg367) are from the hinge region and one (Asp89) from the NTD, whereas no residue from the CTD is involved in the interaction (Figure 4.7C). Based on this analysis, the hinge region can be considered as the first interacting site for the ligand and Asp89 from NTD as the first residue responsible for the domain closure while no involvement of CTD residues suggest that CTD does not participate in domain closure upon ligand binding. Furthermore, analysis for conservation of NTD and hinge residues in αGlyBP as well as its structural homologs suggests that the subsite B (i.e. Glc1 binding site) is the initial ligand binding site as it possess all the four residues i.e. domain closure (Asp70) and hinge (Asp118, hinge1; Gly286, hinge2 and Arg356, hinge3) residues and thus conserves the similar binding mode for Glc1 unit in αGlyBP and for glycerol in Atu4361 protein (Figure 4.7C and 4.7D).

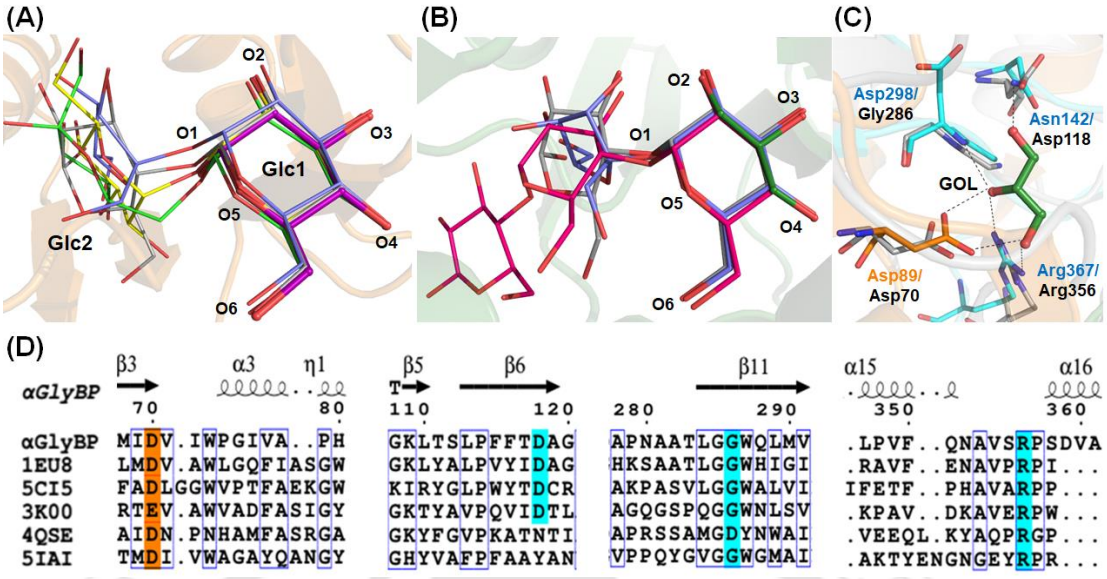


Figure 4.7. Conservation of structural determinant in the subsite B for ligand binding. (A) Overlay of αGlyBP bound to trehalose (yellow), sucrose (blue), maltose (grey), palatinose (green) and glucose (violet). (B) Overlay of Atu4361 protein bound to glycerol (green), sucrose (blue), maltose (grey) and maltotriose (magenta). (C) Active-site comparison of αGlyBP and Atu4361 protein (glycerol bound). Three hinge residues (Asn142, Asp298, and Arg367) and NTD residue (Asp89) interacting with glycerol (ball-and-stick model in green) via hydrogen bonding (dotted lines) are represented in cyan and orange, respectively. The highly conserved hinge residues (Asp118, Gly286 and Arg356) and NTD residue (Asp70) of αGlyBP occupying a similar position are shown with grey line. (D) Structure-based sequence alignment of αGlyBP with trehalose/maltose-binding protein (PDB ID: 1EU8; UniProt ID: Q7LYW7), extracellular solute-binding protein family 1 (PDB ID: 5CI5; UniProt ID: A8F7X5), acarbose/maltose-binding protein GacH (PDB ID: 3K00; UniProt ID: B0B0V1), ABC-type sugar transporter (PDB ID: 4QSE; UniProt ID: A9CGI0) and sugar ABC transporter (PDB ID: 5IAI; UniProt ID: B9JM84) using the program PROMALS3D (Pei and Grishin, 2014) followed by further rendering using online web tool ESPript 3.0 (Gouet et al., 2003). The accession codes for the PDB and UniProt IDs are provided in the parenthesis. For the figure clarity, only a partial alignment has been shown here. Conservation of domain closure (Asp70) and hinge residues (Asp118, Gly286, and Arg356) in all homologous proteins are highlighted in orange and cyan, respectively.

To further affirm the importance of domain closure residue (Asp70) and hinge residues (Asp118, Gly286 and Arg356) in ligand binding, all of them were mutated to alanine except Gly286 as its backbone atom (N) participates in ligand binding. All the mutant proteins were subjected to crystallization in similar condition; however, the crystals of only αGlyBP_R356A mutant protein could be obtained, albeit in a different space group

$P2_1$. Interestingly, the overall structure of αGlyBP_R356A mutant protein reveals the open conformation of αGlyBP in an unliganded state. Topologically, αGlyBP_R356A mutant protein (unbound structure) is identical to that of a αGlyBP_WT•TRE (sugar-bound structure) with an RMSD of 4.2 Å. Overlaying the NTD and CTD of the open and closed conformations using the web server DynDom (Hayward and Berendsen, 1998) showed an RMSD of 1.23 Å and 0.68 Å, respectively, illustrating a larger conformational change for NTD than CTD. Furthermore, the distance between C_α atom of Lys31 (helix α1) and Ala59 (helix α2) in both the open and closed conformations remain constant (~31 Å) indicating that NTD undergoes a rigid translation movement (1.3 Å) (Figure B.1). Calculation of rotational angle of NTD movement by DynDom server shows that change in the torsion angle of hinges allows a rotation of 41.9°. A similar rigid movement of NTD is also identified in the structural homologs such as Atu4361 protein (unbound, PDB ID: 4RJZ and sugar-bound, PDB ID: 4QRZ) and acarbose/maltose-binding protein GacH (unbound, PDB ID: 3K01 and sugar-bound, PDB ID: 3K02), where the distance (~30 Å) between helices α1 and α2 in their respective NTD remains unaltered while the degree of rotation angles (29.5º, 33.4º and 41.9º) increases as the ligand size (GacH receptor-pentasaccharide, Atu4361 protein-trisaccharide and αGlyBP-disaccharide, respectively) decreases (Figure 4.8A-4.8C).

Interestingly, analysis for domain closure mechanism based on DynDom server prediction suggests that upon ligand binding NTD shows a larger conformation change than the CTD which is in accordance with the previously reported asymmetric domain movement mechanism (Pandey et al., 2016). However, based on above mentioned analysis, we propose an extension for asymmetric domain movement mechanism, where only NTD participates in domain closure while CTD plays no role. Moreover, despite the available information for various ligand binding, the molecular details of the mechanism are still underexplored. Hence to delineate the atomic details of domain closure, the molecular organization of active-site residues were compared in both the open and closed conformations. In the unliganded form, the hinge 2 residue Gly286 interacts with residue Asp70 from the NTD via Val71 as well as with residue Asp118 and Arg356 from hinge 1 and hinge 3 region, respectively. Upon ligand binding, the

Glc1 unit interacts with the hinge residues which pulls the domain closure residue Asp70 via Val71 subsequently compelling the NTD to move towards CTD (Figure 4.8D). In addition, further characterization of αGlyBP_D70A mutant protein was done through measurement of the energetic contribution of Asp70 in binding to α-glycosides using ITC experiments considering the active site to be free from endogenously bound ligand due to mutation. Strikingly, a D70A mutation led to a ~210-fold reduction in the binding affinity of trehalose as compared to that of αGlyB_WT protein. This observation suggests that although the Glc1 unit of all α-glycosides binds to the hinge region at subsite B with a negative enthalpy change, mutation of domain closure residue Asp70 to alanine inhibits the NTD to undergo an open-to-closed transition (Figure B.2 and Table 4.9).

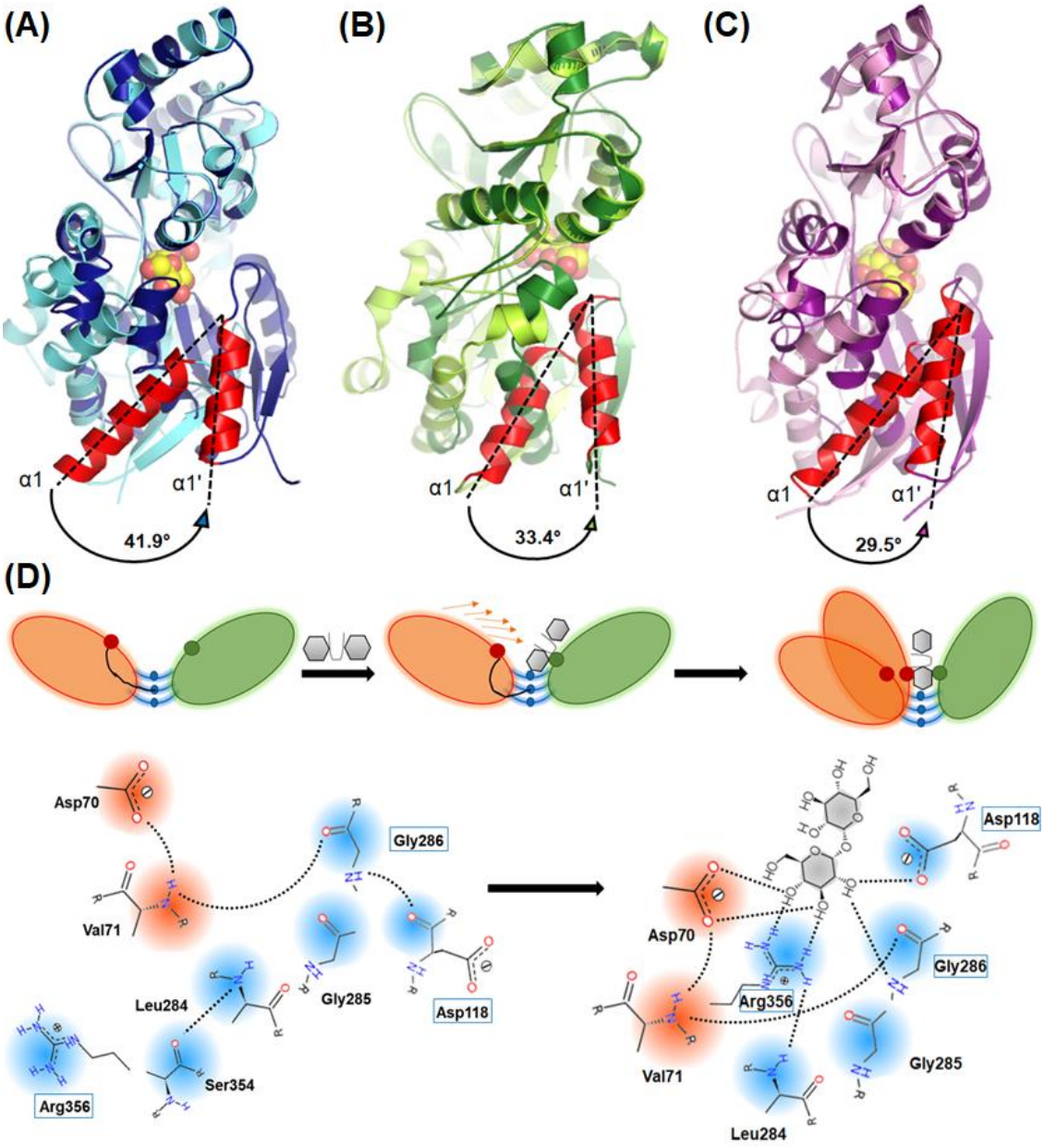


Figure 4.8. Domain movement upon ligand binding. (A-C) Superimposition of unliganded (αGlyBP_R356A-cyan, PDB ID: 6JAL; Atu4361-lime, PDB ID: 4RJZ and GacH-magenta, PDB ID: 3K01) with ligand-bound structures (αGlyBP_WT•TRE-blue, PDB ID: 6J9W; Atu4361-green, PDB ID: 4QRZ and GacH-purple, PDB ID: 3K02). The bound ligand molecule is shown as yellow sphere. Open and closed structures are superimposed at the CTD residues. Rigid movement of NTD is depicted by highlighting the helix α1 (open conformation) and α1' (closed conformation) in red along with their respective angle of movement. (D) A schematic and atomic details of change in the interaction in the presence (left panel) and absence of ligand (right panel) during domain movement.

4.3.4. Mutation of active-site residues alter ligand specificity

Since mutation of hinge 3 residue Arg356 into alanine led to the open conformation of αGlyBP, the energetic contribution of αGlyBP_R356A mutant protein for the disaccharide α-glycosides binding was also determined. Similar to αGlyBP_D70A mutant protein, mutation of Arg356 into alanine leads to a decrease in the binding affinity for trehalose. However, it did not completely lose the binding activity as it showed interaction with all α-glycosides in the ITC experiments. The binding affinity (K_d) of αGlyBP_R356A for disaccharide α-glycosides are found to be in the range of ~1-16 μM, which is in agreement with previous reports for other SBPs (Berntsson et al., 2010). Thermodynamically, similar to αGlyBP_D70A mutant protein, αGlyBP_R356A mutant protein also exhibited a higher affinity for maltose (K_d: 1.61 μM) than trehalose (K_d: 5.95 μM). The binding isotherm for all disaccharide α-glycosides and glucose with αGlyBP_R356A mutant protein are driven by a negative enthalpy change while a positive enthalpy change is observed for maltose (Figure 4.9A, Figure B.3 and Table 4.9). According to the structural data, the three-dimensional structure of all the αGlyBP_R356A mutant protein complexes (αGlyBP_R356A•TRE, αGlyBP_R356A•MAL, αGlyBP_R356A•SUC, αGlyBP_R356A•PAL and αGlyBP_R356A•GLC) are similar to αGlyBP_WT complex structures with an RMSD value of ≤ 0.1 Å for C_α atoms. No differences, except for the loss of hydrogen bonding of O3 and O4 atoms with Arg356 residue, in hydrogen-bonding pattern of the WT and mutant complexes were found. However, in-depth investigation of αGlyBP_R356A mutant protein complexes reveals that in sucrose, O3 and O4 atoms of Glc1 forms two water-mediated interactions with Leu287 (hinge residue) and Glu174 (CTD residue) and thus compensate for the loss of interaction with $N^{\varepsilon 1}$ and $N^{\varepsilon 2}$ atoms of Arg356 due to mutation. Similar to sucrose, the binding of other disaccharide α-glycosides and glucose also restores the hydrogen bonding with the O4 atom, (Figure 4.9B-4.9F). This observation for restoration of hydrogen bonding correspond well with thermodynamic data, where water-mediated interaction plays the role of hinge 3 residue and thus favors the binding of disaccharide α-glycosides and glucose with αGlyBP_R356A mutant protein.

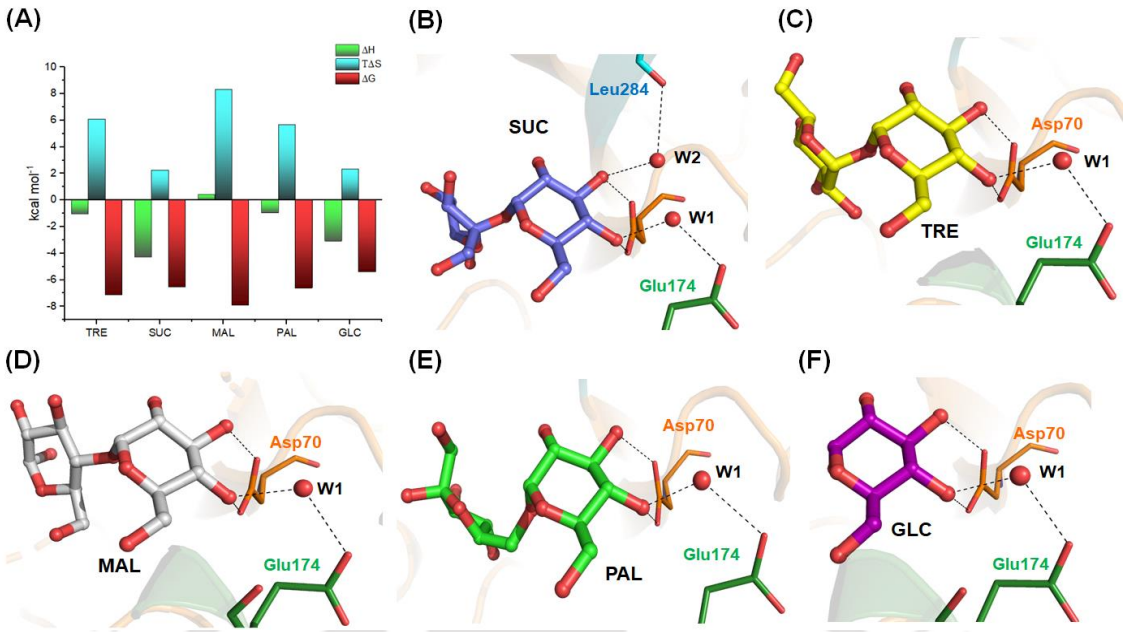


Figure 4.9. Binding of disaccharide α-glycosides with αGlyBP_R356A mutant protein. (A) Thermodynamic profile of ligand binding with αGlyBP_R356A mutant protein, where the change in enthalpy (Δ*H*, green) and entropy (*T*Δ*S*, blue) and the Gibbs free energy of binding (Δ*G*, red) for each ligand are represented as histogram in kcal mol^{-1}. The histograms show that the binding of disaccharide α-glycosides with αGlyBP_R356A mutant is more entropically favourable excluding sucrose which shows the enthalpically favourable binding profile. (B-F) Active site of αGlyBP_R356A mutant protein bound to sucrose (blue), trehalose (yellow), maltose (grey), palatinose (green) and glucose (violet), respectively. In all the complexes, the position of Arg356 is occupied by water (W1, red sphere) molecule, which forms water-mediated (W1) interaction (dotted lines) with Glu174 represented in green line model. In the sucrose complex structure, another water (W2) forms water-mediated interaction (dotted lines) with hinge residue Leu284 shown in cyan line model.

4.3.5. Calcium ion (Ca^{2+}) imparts the role of hinge 1 residue in conferring stability to maltose at the active site

In addition to the hinge 3 residue Arg356, we also determined the functional significance of hinge 1 residue Asp118 for the disaccharide α-glycosides binding. In αGlyBP_WT, out of two Glc units of trehalose, Glc1 interacts with all the three hinge residues Asp118, Gly286 and Arg356 while Glc2 interacts with Asp118 only. Interestingly, upon mutation of Asp118 into alanine, maltose is found to be endogenously bound at the active-site pocket of the mutant protein along with the presence of extra electron density for a metal ion (Figure 4.10A). Initially, several putative divalent cations such as Mg^{2+}, Mn^{2+}, Ca^{2+},

were placed during the structure refinement. However, based on the *B* factor, the Ca^{2+} has been identified as the possible metal ion. To further affirm the presence of Ca^{2+} in αGlyBP_D118A mutant protein, we quantified the binding affinity of Ca^{2+} as well as other divalent metal ions such as Mg^{2+}, Zn^{2+} and Mn^{2+} with EDTA-treated αGlyBP_D118A mutant protein using ITC experiments. Among the tested metal ions, highest affinity was observed for Ca^{2+} as compared to other metal ions Mg^{2+}, Zn^{2+} and Mn^{2+}. All the binding isotherms are exothermic and Ca^{2+} showed typical sigmoidal exothermic isotherm with a dissociation constant (K_d) of 0.50 µM for αGlyBP_D118A mutant protein (Table 4.9). Binding of Ca^{2+} to αGlyBP_D118A mutant protein is driven by favorable enthalpy change and unfavorable entropic contribution where the binding is majorly enthalpically driven. However, the binding of Mg^{2+} is associated with positive entropy changes (Figure 4.10B). Much like Ca^{2+}, the binding of Zn^{2+} is enthalpically favorable, whereas Mn^{2+} does not show any binding activity (Figure B.4).

To further affirm the significance of Ca^{2+} in αGlyBP_D118A mutant protein, examination of the active site of αGlyBP_D118A mutant protein has been performed which reveals that Ca^{2+} is compensating for the absence of Asp118 by stabilizing the maltose molecule in the active site. Although, the functional importance of Ca^{2+} for carbohydrate stabilization has been implicated in many carbohydrate-binding module (CBM) proteins, the binding of a Ca^{2+} to an ABC transporter for rendering stability to the carbohydrate molecule has not been reported, till date. In the active site of αGlyBP_WT, O3-C3-C2-O2 of Glc2 of maltose form a torsion angle of 55.59º where O3 makes an interaction with Asp118. However, upon mutation, Ca^{2+} occupies the position, which is known to form a bidentate interaction with carbohydrates. Hence, to make the bidentate interaction with O2 and O3 of Glc2 of maltose in αGlyBP_D118A mutant protein, Ca^{2+} alters the O3-C3-C2-O2 torsion angle of Glc2 from 55.59º (αGlyBP_WT) to 68.75º (αGlyBP_D118A) and thus, in turn, stabilizes the maltose binding (Figure 4.10C-4.10D). Characteristically, Ca^{2+} forms octahedral coordination with pentagonal bipyramid geometry within 2.4 Å radius. However, this feature is absent in αGlyBP_D118A which forms only tri-coordination bond with O2 and O3 of Glc2 of maltose and a water molecule. Therefore, to further interpret the preferred

coordination of Ca^{2+}, interactions were calculated within 4 Å radius, which revealed a penta-coordination, accomplished through interactions with atoms N (Pro249), N (Tyr250), O2 (Glc2), O3 (Glc2) and O (water) with a square pyramidal geometry (Figure 4.10E).

In addition to Ca^{2+} binding, D118A mutation also altered the ligand priority and maltose was preferred over trehalose. To know the cause for maltose preference, αGlyBP_D118A•MAL the complex was compared with αGlyBP_WT•TRE, αGlyBP_WT•SUC and αGlyBP_WT•PAL complexes. As the two crucial features i.e. a bidentate interaction and torsion angle (O3-C3-C2-O2) of +60° essential for Ca^{2+} binding are inadequate in trehalose, sucrose and palatinose, maltose serves as a better ligand for αGlyBP_D118A mutant protein (Figure B.5).

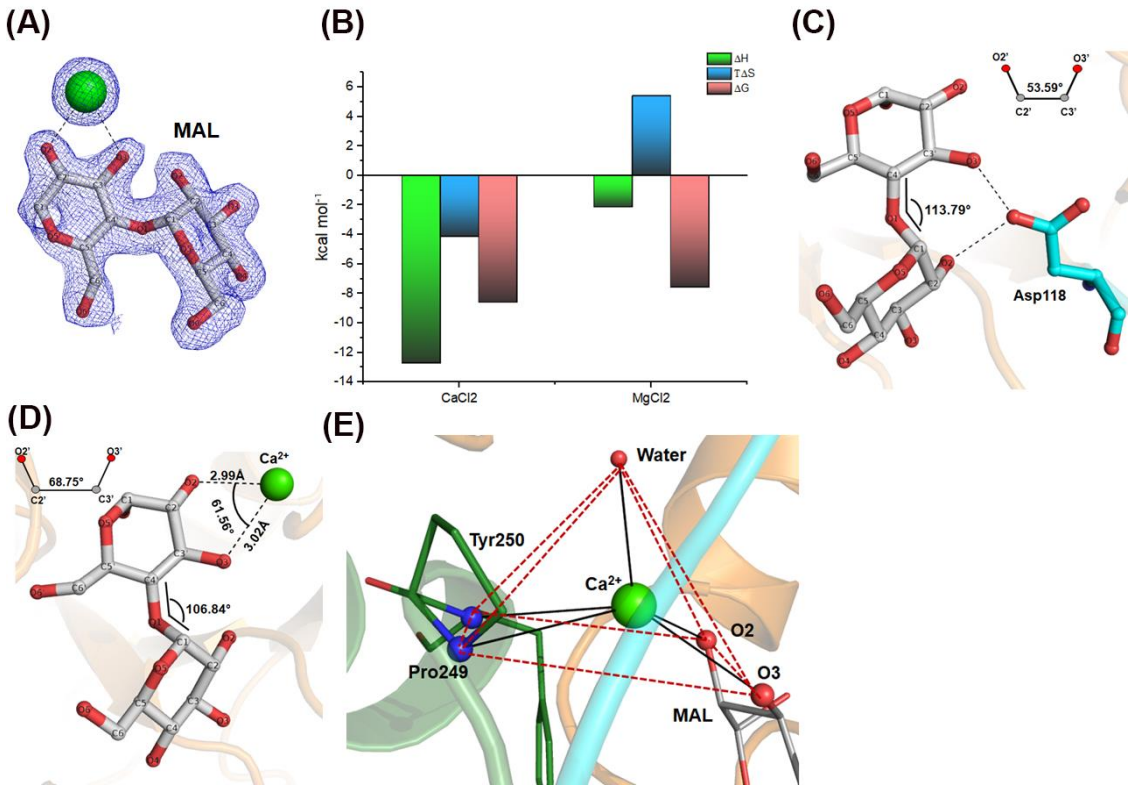


Figure 4.10. The active site of αGlyBP_D118A•MAL complex. (A) *2Fo − Fc* electron density map of endogenously-bound maltose (grey) and metal ion (green sphere) contoured at 1.6σ are represented with blue mesh. (B) Thermodynamic profile of metal binding to αGlyBP_D118A mutant protein, where the change in enthalpy (Δ*H*, green),

entropy ($T\Delta S$, blue) and Gibbs free energy of binding (ΔG, red) for each ligand are represented as histograms in kcal mol^{-1}. The histograms show that the binding of Ca^{2+} with αGlyBP_D118A•MAL mutant is more enthalpically favourable, whereas Mg^{2+} shows the entropically favourable binding profile. (C, D) Comparative assessment of αGlyBP_WT•MAL and αGlyBP_D118A•MAL complexes. The atom O3 of Glc2 forms hydrogen bond (dotted lines) with Asp118 (ball-and-stick model in cyan), whereas in αGlyBP_D118A mutant protein, O3 interacts with Ca^{2+} (green sphere). Changes in the angle of glycosidic linkage and torsion angle of O3-C3-C2-O2 is mentioned in the figure. (E) Penta-coordination of Ca^{2+} (green sphere) with maltose (O2 and O3, red), active-site residue (N, Pro249 and Tyr250, green) and water (O, red sphere) atoms is represented with square pyramidal geometry.

4.3.6. Trehalose and maltose are equally preferred by αGlyBP

To identify the functional significance of other active-site residues apart from domain closure (Asp70) and hinge residues (Asp118 and Arg356), point mutations were also performed for two other crucial residues i.e. subsite A residue (Arg49) and α-glycosidic linkage holding residue (Trp287). Interestingly, equivalent to D70A, D118A and R356A variant, mutation of subsite A residue (Arg49) into alanine altered the ligand specificity of αGlyBP from trehalose to maltose. While the analysis of active site of αGlyBP_R49A suggests that similar to the αGlyBP_WT and αGlyBP_D118A mutant, αGlyBP_R49A mutant protein was also endogenously bound to maltose without any alteration in the hydrogen bonding with the active-site residues (Table B.1). In-depth investigation of the active site reveals that in the open conformation, helices α1 and α2 from N1 subdomain hold each other through electrostatic interaction between Arg49 and Asp11 as well as via polar interaction between Gly10 and Ser42 (Figure B.6A). Except for maltose, this electrostatic interaction with Arg49 is altered upon binding of trehalose, sucrose, palatinose and glucose (Figure B.6B-B.6G). Thus, Arg49 is crucial for the binding of other α-glycosides which is not essential for maltose. Overall, trehalose is the preferred ligand of αGlyBP, however, both thermodynamic and mutagenesis suggests that maltose is equivalently preferred by αGlyBP.

4.3.7. CH…π interaction is crucial for disaccharide α-glycosides binding

In αGlyBP_WT, Glc1 of trehalose is held between two CH…π (or hydrophobic stacking) interaction, where Trp248 forms a parallel-displaced π-stacking and Trp287 establishes a

T-shaped π-stacking with pyranose ring of Glc1. In addition to π-stacking, Trp287 also establishes polar interaction with α-glycosidic linkage (Figure 4.11A). Presuming that the Trp287 might be responsible for maintaining the linkage selectivity between α- and β-glycosidic bonds, it was mutated to phenylalanine. Surprisingly, abolishment of polar interaction of Trp287 with glycosidic linkage did not alter the endogenous binding of trehalose with αGlyBP_W287F mutant protein and suggests that Trp287 does not govern the linkage selection. Further, deep inspection of active site of αGlyBP_W287F•TRE exhibited that Phe287 restored the T-shaped π-stacking of Trp287 with pyranose ring of Glc1 and preserve the glycosidic angle in αGlyBP_W287F (111.98º) as in αGlyBP_WT (116.13º) (Figure 4.11A, 4.11B). To further corroborate the importance of T-shaped π-stacking with Glc1 unit, Trp287 was mutated into alanine and subjected to co-crystallization and thermodynamic studies with α-glycosides. Binding energetics confirms the removal of the endogenously-bound trehalose owing to the loss of π-stacking and reveals a change in the enthalpy upon disaccharide α-glycosides binding. Strikingly, among all the titrated ligands, maltose showed the dual binding mode with αGlyBP_W287A mutant protein (Figure B.7). Hence, to get further insight into the dual binding mode, co-crystallization of αGlyBP_W287A mutant protein with disaccharide α-glycosides and glucose was performed. Crystal structures of all αGlyBP_W287A mutant complexes are comparable to the αGlyBP_WT complex structures with an RMSD value of ≤ 0.1 Å for C_α atoms, indicating negligible deviation from the WT. No alteration, except for the loss of interaction with the mutant residue, is observed in the hydrogen-bonding interaction between the sugars and the active-site residues (Table B.1).

Based on the position of Trp287 in between NTD and hinge region, it can be speculated that being a bulkier residue, Trp287 anchors the torsion for NTD movement (Figure 4.11C). Mutation of Trp287 into alanine led to the loss of torsion and increased the flexibility of NTD movement. This observation corresponded well with the thermodynamic characteristics for a dual binding mode of maltose in the ITC experiments. The negative enthalpy change arises mainly due to the loss of hydrophobic stacking with Trp287 and formation of hydrogen bonding of Glc1 hydroxyl group with the hinge region. On the other hand, the second phase of binding associated with a

positive enthalpy indicates a delay in the NTD closure which arises due to an establishment of hydrophobic interaction between Val15 (NTD) and Glc2 (Figure 4.11D and Figure B.7). Strikingly, unlike the other mutations, abolishment of π-stacking in αGlyBP_W287A mutant protein did not alter the ligand priority and showed the preference for trehalose with a dissociation constant (K_d) of 58 nM over other ligands (Table 4.9).

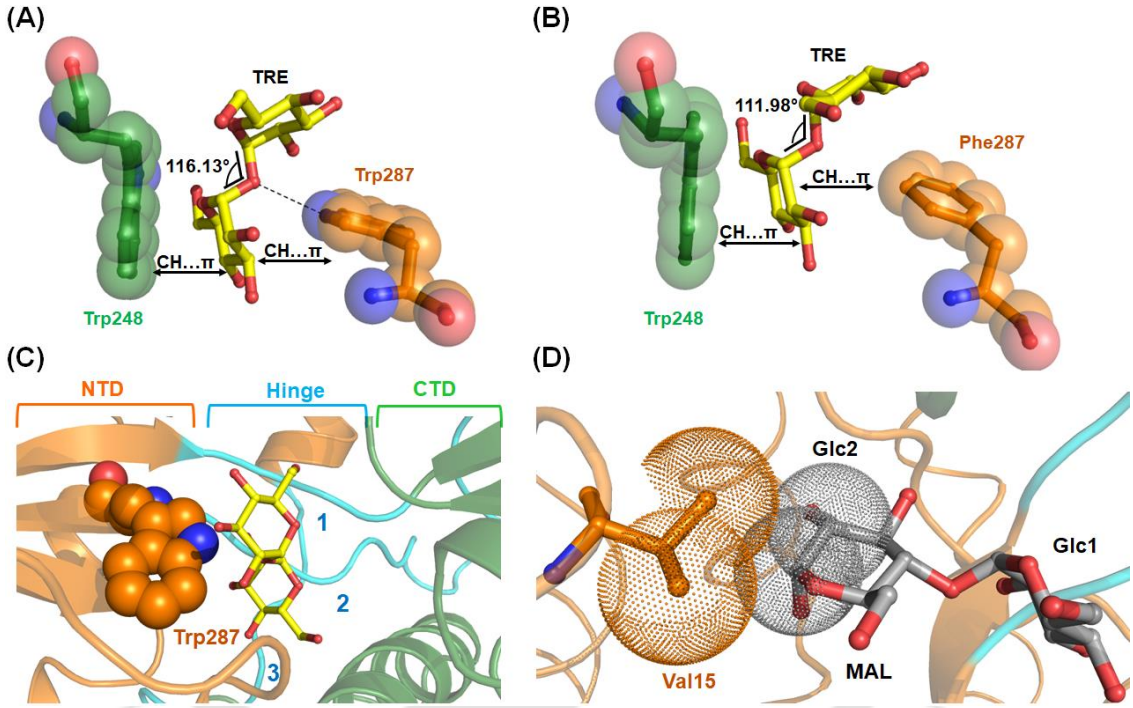


Figure 4.11. Hydrophobic and polar interaction with bound disaccharide α-glycosides. (A) Two CH…π interaction in αGlyBP_WT•TRE complex formed with Trp248 (parallel-displaced) and Trp287 (T-shaped) are marked with a double-headed arrow while the polar interaction between the glycosidic linkage and Trp287 is represented with a dotted line. (B) CH…π interaction of αGlyBP_W287F•TRE complex structure with Trp248 (parallel-displaced) and Phe287 (T-shaped). (C) The bulkier residue Trp287 present between NTD and hinge region is shown as a sphere. (D) Hydrophobic interaction within 4 Å radius formed between Val15 and maltose is shown with orange and grey dotted sphere, respectively.

4.3.8. α-glycosides uptake and metabolism systems are functionally associated

The protein αGlyBP transports five different types of sugar and does not allow the binding of any other sugar as it maintains the stereo- and linkage-selectivity for α-(1,1), α-(1,2), α-(1,4) and α-(1,6) disaccharide α-glycosides only. To identify such potential selective mechanism in other homologous proteins, evolutionary relationship of αGlyBP was examined. As TMBP (PDB ID: 1EU8) from *T. litoralis* also transports trehalose, its homologous proteins were also included for the evolutionary study. Homologous proteins of αGlyBP and TMBP having a significant identity (sequence identity: >30%, query coverage: > 90%) were retrieved from RefSeq database (O'Leary et al., 2016). Evolutionarily, all homologous proteins are distributed among ten different genera and belong to three distinct phyla including Actinobacteria, Firmicutes and Dionococcus (Figure 4.12A). Major homologous proteins of αGlyBP and TMBP are from Actinobacteria and Firmicutes, respectively, whereas the common homologous proteins belong to phyla Dionococcus. Next, we identified the genes involved in transport and metabolism of α-glycosides and compared among ten different genera to see the functional signification of α-glycoside transport. Out of the ten genera, only five genera were examined due to the unavailability of a genetic map in the Gene database and nine ORFs were selected as a representative to cover all the phyla. Despite belonging to different phyla, genes for ABC transport system are conserved while the flanking genes vary. In both Actinobacteria and Firmicutes, the functional unit carries the genes for α-glycosides uptake and metabolism, however in contrast to α-glycosides uptake system diversity can be observed in genes involved in metabolism. In Actinobacteria, the co-occurrence of genes for trehalose utilization and α-glucosidases with ABC transport system led us to postulate that the metabolic requirement of bacteria also governs the selective transport of α-glucosides through ABC transport system. Interestingly, in Firmicutes, operon of *Thermotoga* carries the flanking gene for β-glycosidase instead of α-glucosidase suggesting a loss of linkage selectivity or change in the substrate specificity during the course of evolution. In contrast to both phyla, Dionococcus operon possesses only those genes which encode the ABC transport system while lack the genes responsible for metabolism in the flanking regions (Figure 4.12B).

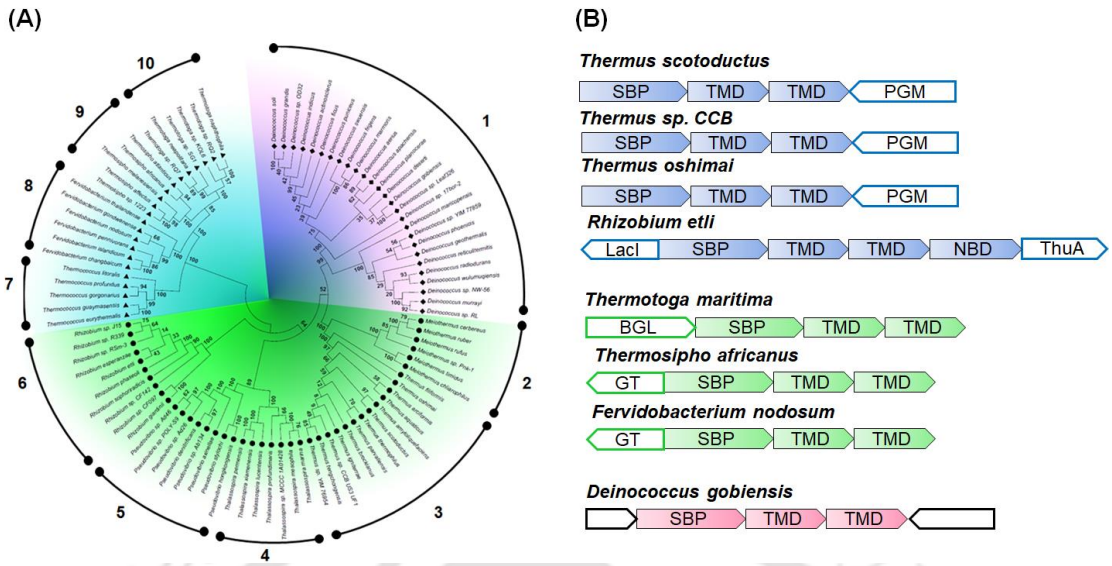


Figure 4.12. Evolutionary distribution for α-glycosides transport. (A) All homologous sequences of αGlyBP and TMBP were used to construct the phylogenetic tree using the neighbor joining (NJ) method and clustered into three phyla viz. Actinobacteria (blue), Firmicutes (green) and Dionococcus (magenta). The numbers on the internal nodes of the phylogenetic tree represent the bootstrap value (1000 replicates) which denotes the significance of evolutionary relation across the three phyla while the numbers on the periphery of the phylogenetic tree symbolize ten different genera. (B) Operonic arrangement showing a functional association between α-glycosides ABC transport and metabolizing systems. Subunits of ABC transporter are labeled and colored coded according to their respective phyla. Abbreviation: SBP, substrate-binding protein; TMD; transmembrane domain; NBD, nucleotide-binding domain; PGM, phosphoglucomutase; ThuA, trehalose utilization protein; BGL, β-glucosidase; GT, glycosyltransferase.

4.4 DISCUSSION

Carbohydrates show a wide diversity in context of their stereochemistry and can be assembled in many combinations to form a wide range of isomers. Unlike the carbohydrate-metabolizing enzymes, which selectively hydrolyze the glycosidic linkage, selective transport of carbohydrates based on its stereochemistry remains unexplored. In *T. thermophilus* HB27, a trehalose ABC transport system was identified for the transport of a multiple α-glycosides (Silva et al., 2005). Previously, we have reported a selection mechanism for disaccharide α-glycosides over higher oligosaccharide α-glycosides and provided a physiological basis for the transport of multiple α-glycosides through a single

transporter (Chandravanshi et al., 2019). However, the basis for binding affinity towards α-glycosides, linkage selection and preference of α- over β-glycosides during transport remains unknown. This study utilizes the structural and thermodynamic data to underpin the mechanism of ABC transporter for selective binding of carbohydrates.

One of the critical aspects identified through structural analysis of αGlyBP for carbohydrate transport is that it transports multiple disaccharide α-glycosides comprising of different kind of α-glycosidic linkages such as α-(1,1), α-(1,2), α-(1,4) and α-(1,6) and has the ability to differentiate between α- and β-configurations of sugars. Analysis of structures of both α- and β-glycosides reveals that change in the planner geometry of glycosidic linkages leads to stereochemical restrains and thus hinders the binding. Furthermore, structural and thermodynamic analyses of wild and mutant proteins together demonstrate that out of the two subsites (A and B), subsite B acts as an 'initial binding site' and stabilizes the ligand at the active site as it maintains the hydrophobic stacking and polar interaction for all ligands comprising of α-linked glucose unit as a structural component. The fact that the residues Asp70, Trp287 and Arg356 which initiates the domain closure, anchor the torsion of NTD movement and facilitates sugar binding, resides at the subsite B indicating it to be accountable for the generation of the closed conformation. According to previous studies, "Venus Fly-trap" mechanism involves an asymmetric domain movement in which movement of NTD is larger than the CTD (Pandey et al., 2016). However, in-depth investigation of the "Venus Fly-trap" mechanism at an atomic level in the αGlyBP and homologous protein Atu4361 in complex with glycerol reveals the involvement of ligand interaction only with NTD and hinge region residues while residues from CTD do not play any role in domain closure (Figure 4.13). Instead, the role of CTD was proposed for substrate translocation in FusA protein belonging to cluster G of SBPs (Culurgioni et al., 2017). SBPs from the cluster G possess the metal ion binding site at CTD and thus has been suggested to be involved in salt bridge formation at the SBP: TMD interface (Culurgioni et al., 2017). Based on these observations, it can be postulated that for substrate translocation, the CTD maintains the interaction with TMD and remains static while the NTD undergoes a conformational change to facilitate the release of the substrate into the translocation passageway.

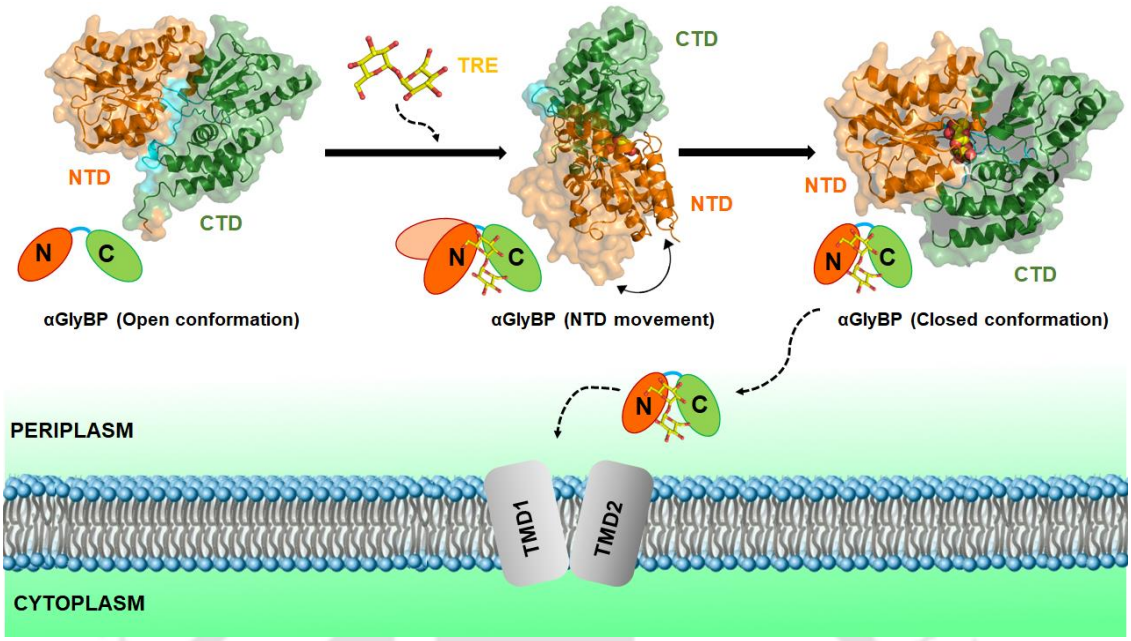


Figure 4.13. Proposed model for α-glycoside binding mechanism of αGlyBP. In this model, the ligand-binding mechanism has been proposed based on the structural studies of SBP, where free-floating SBP i.e. αGlyBP (open conformation, ribbon model) selectively recognizes the cognate ligand (TRE, yellow) in the periplasm. The binding of ligand leads to movement of NTD (orange) for ligand encapsulation and generates the closed conformation of SBP (αGlyBP). The closed form of SBP might further interact with TMDs (grey) and translocate the substrate.

In addition to the selective transport mechanism, another remarkable feature identified for the αGlyBP ABC transport system is the replacement behavior of glucose. The dissociation constant (K_d) of glucose (111 µM) is ~18-fold lower than that of trehalose (5.95 µM), which follows the required criteria of ligand replacement titration (Velazquez-Campoy and Freire, 2006). Till date the replacement of ligand from the active site has been reported by higher-affinity ligands (Krainer and Keller, 2015). For the first time, this study reports the structural data for the reverse mechanism where a lower-affinity ligand (glucose) replaces a higher-affinity ligand (trehalose) from the active site. This leads towards the identification of a remarkable feature of αGlyBP ABC transport system where glucose retains the capability to replace all the other α-glycosides from the active site. However, the reason behind the ability of glucose to replace trehalose still requires further understanding. The phenomena of α-glycosides replacement by glucose at the active site can be correlated to product-based feedback

inhibition mechanism of glycoside hydrolase (GH) enzyme. In this, glucose binds to the active site and inhibits the enzymatic activity by competing with the substrate for the active site (Yang et al., 2015b). Based on this phenomenon, it can be hypothesized that apart from being a cognate ligand itself, glucose might also be involved in regulating the transport of α-glycosides.

In many SBPs viz. glucose/galactose-binding protein (PDB ID: GCG) and fructooligosaccharide-binding protein (PDB ID: 5G5Z), the role of Ca^{2+} is hypothesized to be involved in signaling and substrate translocation (Flocco and Mowbray, 1994; Culurgioni et al., 2017). In contrary, Ca^{2+} of αGlyBP_D118A•MAL complex participates in the binding and stabilization of sugar at the active site. In many CBM proteins (e.g. CBM36 and CBM60), a similar feature has been reported where Ca^{2+} is directly involved in the interaction with carbohydrates. However, unlike αGlyBP_D118A•MAL complex, in CBMs Ca^{2+} mediates the recognition and subsequent binding of carbohydrate at the active site (Abou-Hachem et al., 2002; Jamal-Talabani et al., 2004; Montanier et al., 2010). Despite the different functional role of Ca^{2+} in CBM60 and αGlyBP_D118A•MAL complex, O2 and O3 atoms of sugar form the similar coordination with the Ca^{2+} with a torsion angle (O2-C2-C3-O3) of +60º. This critical geometric feature is reported to be crucial in CBMs as carbohydrates (e.g., mannoside) having a torsion angle of -60º is unable to bind to CBM60 which indicates that ligand selection is also Ca^{2+}-dependent (Montanier et al., 2010; Zhu et al., 2010). Moreover, Ca^{2+} is known to form octa-coordination with the pentagonal bipyramid geometry in a radius of 2.4 Å (Senguen and Grabarek, 2012). However, in αGlyBP_D118A•MAL, Ca^{2+} forms only three coordination in the 2.4 Å radius while with an extended radius (4 Å), it forms the penta-coordination with the square pyramidal geometry. This geometric observation within the 4 Å radius correspond well with Ca^{2+} voltage-gated channel in which Ca^{2+} forms the octa-coordination in 4 Å radius (Tang et al., 2014). Although, Ca^{2+} is the preferred metal ion in CBMs and αGlyBP_D118A•MAL, binding affinity towards Mg^{2+} is also exhibited. Due to smaller ionic radii, a bidentate interaction similar to Ca^{2+} is not possible in case of Mg^{2+} and hence would lead to a change in the O2-C2-C3-O3 torsion angle, which is apparently implausible.

Evolutionary distributions of homolog ABC transport systems for α-glycosides have been identified in three different phyla Actinobacteria, Firmicutes and Dionococcus. To understand the functional significance of transporting multiple α-glycosides through a single transport system, the distribution of functionally associated genes across the genera was compared. In Actinobacteria, the ABC transporter of genera *Thermus* and *Meiothermus* is functionally associated with α-glucosidase while in *Rhizobium etli*, the trehalose-utilizing protein co-exists with the ABC transport system. The ABC transport system identified in *Rhizobium etli* is encoded by *thuEFGK*, which is similar to ABC transport system of *Shinorhizobium meliloti*, is primarily involved in the uptake of trehalose, maltose and sucrose (Willis and Walker, 1999; Jensen et al., 2002). In Firmicutes, the operon of *Thermococcus* carries the genes for trehalose metabolism, whereas in *Thermotoga*, β-glycosidase co-occurs with ABC transporter suggesting the transport of β-glycosides rather than α-glycosides. Based on the highest homology with α-glycoside ABC transporter and co-occurrence with β-glycosidase, one can deduce that like aglEFGAK transport system (Jensen et al., 2002), ABC transport system of *Thermotoga* might transport both α- and β-glycosides. Notably, in Dionococcus, such functional association is not observed and hence mechanistic behavior of α-glycosides transport remains elusive. In general, the fate of sugar transport is linked with functionally-associated proteins (Light et al., 2017). Henceforth, the transport system in Actinobacteria is exclusively responsible for the transport of α-glycosides while Firmicutes transport system shows broad range substrate specificities without differentiating the linkage between the monosaccharide moieties.

4.5. CONCLUSION

This study employs structural and thermodynamic data to provide the basis for the selective transport mechanism of disaccharide α-glycosides through ABC transporter. αGlyBP selectively transports α-(1,1), α-(1,2), α-(1,4) and α-(1,6) disaccharide α-glycosides over β-glycosides. In addition to α-glycosides, glucose is also stereoselectively transported. In contrast to binding kinetics, glucose replaces the endogenously-bound trehalose suggesting that αGlyBP can be potentially used as solubility-tag in the place of maltose-binding protein (MBP) that can be further regenerated by using glucose.

Moreover, protein can be changed from trehalose binding to maltose-binding protein through strategized mutation of the active-site residues which, in turn, can be further exploited for biological applications.

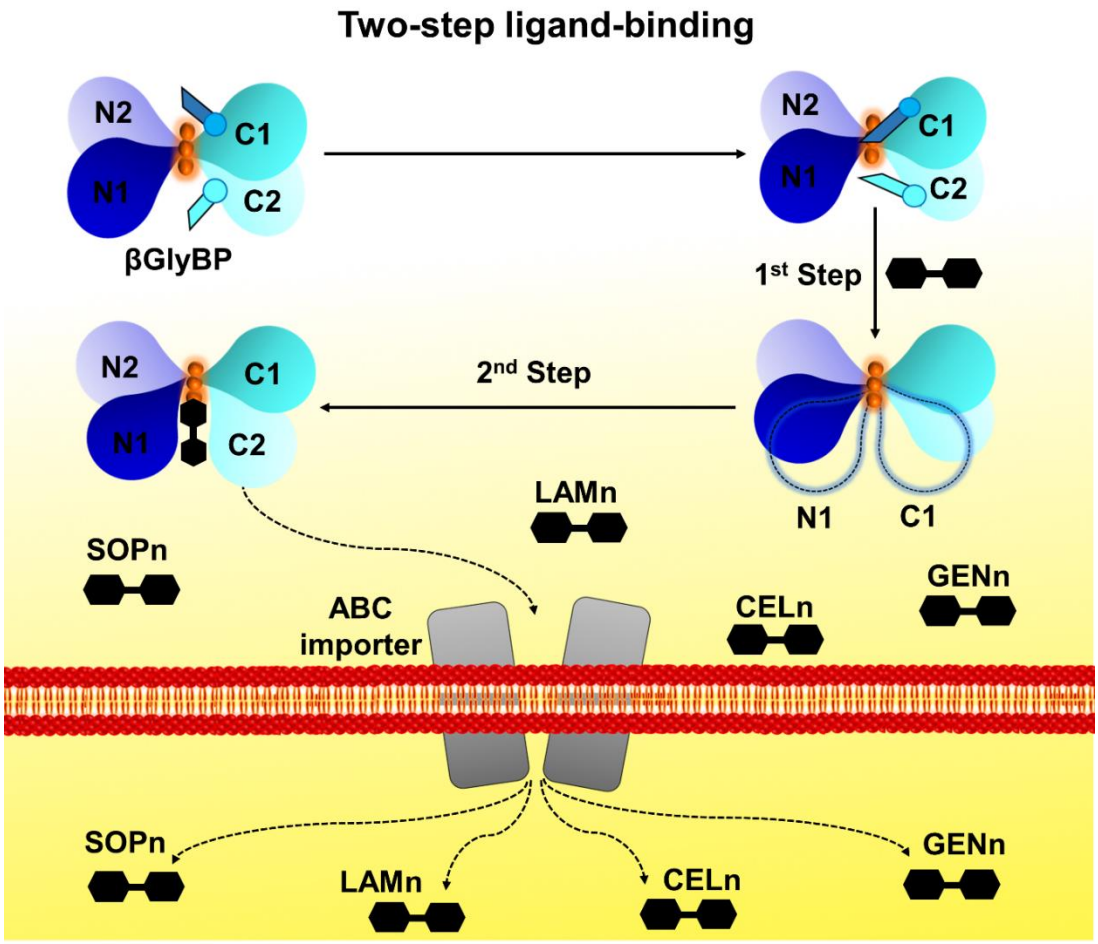
CHAPTER 5 – STRUCTURE OF βGlyBP

This chapter has been published as:

Chandravanshi M, Samanta R and Kanaujia SP (2020). Conformational trapping of a β-glucosides-binding protein unveils the selective two-step ligand-binding mechanism of ABC importers. *J. Mol. Biol.*, 432:5711-5734.

ABSTRACT

Substrate-binding proteins (SBPs), a key component of ATP-binding cassette (ABC) importers, selectively capture ligand(s) and ensure their translocation via its cognate ABC import system. SBPs bind their cognate ligand(s) via an induced-fit mechanism known as the "Venus Fly-trap" mechanism; however, this mechanism lacks the confirmatory evidence(s) in its support. In this study, we used structural and thermodynamic approaches to delineate the atomic details and ligand selection mechanism of an SBP, β-glycosides-binding protein (βGlyBP). The protein βGlyBP is multi-specific and binds to different types of β-glycosides varying in their glycosidic linkages viz. β-1,2; β-1,3; β-1,4 and β-1,6 with a degree of polymerization of 2-5 glycosyl units. The overall structure of βGlyBP possesses four subdomains (N1, N2, C1 and C2) similar to other known SBPs. The unliganded protein βGlyBP remains in an open state, which closes upon binding to sophorose (SOP2), laminari-oligosaccharides (LAMn), cello-oligosaccharides (CELn), and gentiobiose (GEN2) of varying lengths. This study reports, for the first time, four different structural states (open-unliganded, partial-open-unliganded, open-liganded and closed-liganded) of the protein βGlyBP revealing its conformational dynamics. Further, results suggest that the conformational dynamics of N1 and C1 subdomains drive the ligand binding unlike that of the whole N- and C-terminal domains (NTD and CTD) as known in the "Venus Fly-trap" mechanism establishing a two-step ligand-binding mechanism. Additionally, profiling of stereo-selection mechanism for carbohydrates in the subcluster D-I SBPs reveals that α- and β-glycosides occupy a similar position but in an opposite orientation. In summary, results from this study demonstrate that the details of conformational dynamics and ligand selection is pre-encoded in the SBPs that interplay during ligand transportation.



5.1 INTRODUCTION

ATP-binding cassette (ABC) transporters are the largest superfamily of proteins that facilitate the translocation of a diverse array of substrates across the plasma membrane using ATP as the energy source (Higgins, 1992). ABC transporters are classified into importers and exporters depending upon the direction (inside or outside of the cell, respectively) of the substrates being transported (Wilkens, 2015; Szollosi et al., 2018). Although ABC exporters are ubiquitously distributed in all domains of life, ABC importers are reported only in prokaryotes and plants till date (Davidson et al., 2008; Lefevre and Boutry, 2018). Both the ABC importers and exporters share a typical architecture of transmembrane domains (TMDs) and nucleotide-binding domains (NBDs) which function as substrate translocator and energy generator from ATP hydrolysis, respectively. Unlike ABC exporters, ABC importers require an additional component known as substrate (or

solute)-binding protein (SBP) (van der Heide and Poolman, 2002; Marinelli et al., 2011; Scheepers et al., 2016). SBPs capture their cognate ligands from the periplasmic or extracellular environment and deliver them to the TMDs for the subsequent translocation into the cell (Davidson et al., 1992). Irrespective of the types of the ligands, SBPs possess a conserved structural fold having two globular α/β domains with a central β-sheet flanked by α-helices (Berntsson et al., 2010). These two domains are linked by a flexible hinge region which allows the free rotation of these domains for substrate capturing via the ‘‘Venus Fly-trap’’ mechanism (Mao et al., 1982). Based on their topology, particularly of the hinge region, SBPs have been classified into seven different clusters, A-G (Berntsson et al., 2010; Scheepers et al., 2016).

Structural studies on SBPs have demonstrated that the hinge region facilitates the transition from an open to closed state upon ligand binding and controls the equilibration between these two states (Quiocho and Ledvina, 1996; Shilton et al., 1996). As the structural feature of the hinge region varies across the SBP clusters, the degree of domain movement from an open to closed state also differs (Begg et al., 2015; Chandravanshi et al., 2020). Nevertheless, this degree of domain movement is independent of the size and type of ligands bound to the protein (Magnusson et al., 2004; Trakhanov et al., 2005; Pandey et al., 2016; Chandravanshi et al., 2020). Two basic models associated with the domain movement describing the ligand-binding mechanism have been proposed: (1) conformational selection and (2) induced-fit mechanism. In the conformational selection mechanism, ligands bind to a preformed closed-unliganded state, whereas, in the induced-fit mechanism, ligand binding triggers the domain movement to bring the closed conformational changes (de Boer et al., 2019a). Between the two mechanisms, most SBPs follow the latter i.e. the induced-fit mechanism as it enables the translocator (TMDs) to differentiate between the unliganded and liganded states (Doeven et al., 2008).

In the induced-fit mechanism, an essential intermediate (open-liganded) state is known to couple the two end (open and closed) states; it is a thermodynamically unfavorable and transient state (de Boer et al., 2019a). Owing to its transient nature, capturing its molecular details becomes difficult. Moreover, to obtain its mechanistic insights, details of ligand

binding as well as conformational dynamics of the intermediate state is inevitable. Although an array of structural and biophysical data detailing the mechanistic insights into the induced-fit mechanism of SBPs have been reported (Skrynnikov et al., 2000; Trakhanov et al., 2005; Silva et al., 2011; de Boer et al., 2019a, b), these relate to the initial (open) and/or the final (closed) states only. Moreover, the mechanisms for the ligand recognition and selection by SBPs are not well delineated. Although the selectivity of ABC importers that is shown to be governed by the conformational state(s) and ligand-release kinetics of SBPs is well accepted (de Boer et al., 2019b), a precise relationship between selective ligand binding and conformational dynamics of SBPs has not been well established till date.

Carbohydrate-specific SBPs are pertinent to understand this relationship due to the complexity of the substrate (i.e. carbohydrate) having varying length, anomeric configuration, glycosidic linkage and epimeric state (Hölemann and Seeberger, 2004; Raich et al., 2016). Consequently, carbohydrate-specific SPBs attain different topologies and thus are classified into four distinct subclusters B-I, C-IV, D-I and cluster G specific to monosaccharides, linear oligosaccharides, linear, circular & branched oligosaccharides and polysaccharides, respectively (Fukamizo et al., 2019). Surprisingly, despite having different topologies, subclusters C-IV and D-I SBPs are designated to facilitate the uptake of linear β-glycosides (Cuneo et al., 2009b; Abe et al., 2018). Depending upon the glycosidic linkages, β-glycosides are categories as linear glucan such as β-1,2-glucan, β-1,3-glucan (laminarin), β-1,4-glucan (cellulose), β-1,3/1,4-glucan (lichenan), and branched glucan such as β-1,3/1,4-glucan (calocyban) and β-1,3/1,6-glucan (lentinan) (Synytsya and Novak, 2014). These polysaccharides are further catabolized into shorter gluco-oligosaccharides such as sophoro- (SOPn; β-1,2), laminari- (LAMn; β-1,3), cello- (CELn; β-1,4) and gentio-oligosaccharides (GENn; β-1,6), where n represents the number of glycosyl (Glc) unit or a degree of polymerization (DP). These gluco-oligosaccharides are the preferred substrates for β-glucosidases (Chuenchor et al., 2011). Although the understanding of the various gluco-oligosaccharides metabolism by β-glucosidases are well documented, their uptake through an ABC import system remains elusive. Nevertheless, it can be speculated that multi-specificity of the ABC import system would

be essential to fulfill the demand of a broad range of substrates to β-glucosidases. In the previous *in silico* study, we suggested that a single ABC import system (ORF IDs: TTHB082-TTHB086) of a thermophilic gram-negative bacterium *Thermus thermophilus* HB8 which is enough to uptake β-glycosides unlike α-glycosides for which multiple import systems (ORF IDs: TTHA0354-TTHA0356 and TTHA1650-TTHA1652) are required (Chandravanshi et al., 2019). However, this premise required further confirmatory evidences.

Thus, in this study, we report the crystal structures of the SBP (ORF ID: TTHB082) to accomplish the insights into the ligand binding and selection mechanism of β-glycosides. Moreover, this study provides the first-ever structural data of an SBP bound to a variety of β-glycosides such as sophoro- (SOPn; β-1,2), laminari- (LAMn; β-1,3), cello- (CELn; β-1,4) and gentio-oligosaccharides (GENn; β-1,6). Furthermore, the structural data corroborated the thermodynamic data establishing the SBP (ORF ID: TTHB082) as a β-glycosides-binding protein (βGlyBP). Using a combination of structural and thermodynamic data of the wild type and mutants, here we propose the ligand binding and selection mechanism of the protein βGlyBP.

5.2 MATERIALS AND METHODS

5.2.1 Carbohydrates

Disaccharide β-glycosides (D-(+)-cellobiose and D-(+)-sophorose) and α-glycosides (D-(+)-trehalose and D-(+)-maltose) and monosaccharides (D-(+)-glucose and D-(+)-mannose) were purchased from Sigma-Aldrich (St. Louis, Missouri, USA). Other higher β-glycosides like D-(+)-cellotriose, D-(+)-cellotetraose, D-(+)-cellopentaose, D-(+)-laminaribiose, D-(+)-laminaritriose and D-(+)-laminaritetraose were procured from Megazyme (Wicklow, Ireland). Carbohydrates like D-(+)-gentiobiose, D-(+)-lactose, D-(+)-sucrose, D-(+)-melibiose, D-(+)-raffinose, D-(+)-galactose and D-(+)-tagatose were obtained from the Himedia (Mumbai, India).

5.2.2 Construction of wild type and mutant expression plasmids

The cloned pET11a vector (plasmid ID: PC014082-41) carrying full length (1305 bp) *TTHB082* gene coding for the protein βGlyBP was acquired from Biological Research Center, NITE (NBRC), Japan. However, to facilitate the purification of recombinant protein using metal affinity chromatography, the *TTHB082* gene was sub-cloned into pET22b(+) vector (Novagen). For sub-cloning, truncated (1251 bp, lacking the signal sequence 1-54 bp) form of the *TTHB082* gene was amplified using the cloned pET11a vector as a template and the oligonucleotide sequences mentioned in the Table 5.1. The 6xHis-tag was introduced at the C-terminus of the protein βGlyBP. Subsequently, amplified fragment was double digested by using *Nde*I and *Bam*HI and inserted into a pET22b(+) vector double digested by using the same set of restriction enzymes. The clones were confirmed by double digestion using the same set of restriction enzymes by incubating the vectors at 37 °C for two hours (Figure 5.1A). The resulting recombinant construct of wild type was further utilized as a template to generate the mutants of the protein βGlyBP by using oligonucleotide sequences mentioned in the Table 5.1 and Q5 Site-Directed Mutagenesis Kit (New England Biolabs, MA, USA). Subsequently, all these mutant constructs were analyzed in 0.8% agarose gel and subjected to plasmid DNA sequencing for mutagenesis confirmation (Figure 5.1B-5.1F).

Table 5.1. List of oligonucleotide sequences used to construct the wild type (WT) and mutant protein expression plasmids.

Plasmid name	Primer	Oligonucleotide sequence (5′-3′)
βGlyBP_WT	Forward (*Nde*I)	ATAT**CATATG**ATGCAGAAGACCCTCGAGGTCTGGATCATGCC
	Reverse (*Bam*HI)	ATAT**GGATCC**TCA**ATGGTGATGATGATGGTG**CCTCAGGGCCTGGTTG
βGlyBP_W41A	Forward (Trp41Ala)	CGTCCTGGAC***gc***GGGCGTGGCC
	Reverse (Trp41Ala)	GTGACCTTCACCTCCACG
βGlyBP_W67A	Forward (Trp67Ala)	CGGCACCACC***gc***GGTGGGGGCG
	Reverse (Trp67Ala)	AGCTGGGTGAGGTCCGGC
βGlyBP_E117A	Forward (Glu117Ala)	TGGTTCTCGG***c***GCTTCGGGCCTTCTACTAC

	Reverse (Glu117Ala)	GGGCACGGCGGTGGCCTG
*βGlyBP_W177X	Forward (Trp177Ala)	AAAGAACTCC***gc***GGACGTGCTCCACAACGCCGC
	Reverse (Trp177Ala)	CCCGGGGTGCAGAGGGGG
βGlyBP_W256A	Forward (Trp256Ala)	CAGCGGCCCC***gc***GATGATCCAGC
	Reverse (Trp256Ala)	GCGAAGACGGCGCACTTC

* These primers were designed for the alanine substitution; however, multiple random mutations were observed during the site-directed mutagenesis experiment.

- Restriction sites are highlighted in bold and uppercase.
- The 6xHis-tag is highlighted in bold underlined and uppercase.

The altered codon for the alanine substitution is indicated with bold underlined and lowercase.

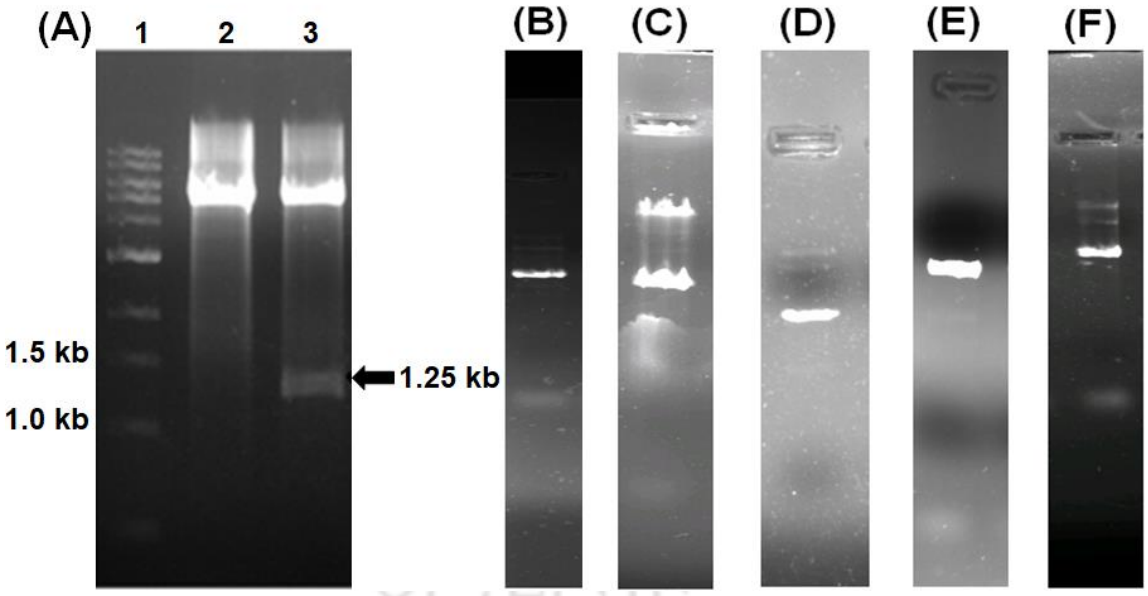


Figure 5.1. Cloning and site-directed mutagenesis of *TTHB082*. (A) Clone confirmation of *TTHB082_WT* by double digestion of the vector pET22b carrying the gene *TTHB082* by *Nde*I and *Bam*HI restriction enzymes (lane 1: DNA ladder, lane 2: negative clone, lane 3: insert pop out after double digestion of the plasmid pET22b). Isolated plasmids of all the *TTHB082* mutant constructs (B) *TTHB082_W41A* (C) *TTHB082_W67A* (D) *TTHB082_E117A* (E) *TTHB082_W177X* and (F) *TTHB082_W256A*. Mutagenesis confirmation of all the *TTHB082* mutant constructs were done by plasmid DNA sequencing.

5.2.3 Overexpression and protein purification of recombinant proteins

The *Escherichia coli* BL21 (DE3) competent cells (Novagen) were transformed by using all the generated recombinant constructs of wild type (*TTHB082_WT*) and mutants (*TTHB082_W41, TTHB082_W67A, TTHB082_E117A, TTHB082_W177X* and *TTHB082_W256A*) for the overexpression. The transformed competent cells were grown at 37°C in Luria-Bertani (LB) medium supplemented with 100 mg ml^{-1} ampicillin until the culture attained the optical density (OD) of 0.6-0.8 at 600 nm. Subsequently, the cells were induced by using 1 mM of isopropyl-1-thio-β-D-thiogalactopyranoside (IPTG) and were allowed to overexpress the recombinant proteins for 12 hrs at 25°C. After which, cells were harvested by centrifugation at 3836*g* for 10 min and resuspended in the lysis buffer containing 20 mM Tris-HCl pH 7.5, 20 mM imidazole, 150 mM NaCl, 10% glycerol, 3 mM β-mercaptoethanol (β-ME) and 1 mM phenylmethylsulfonyl fluoride (PMSF). Subsequently, the resuspended cells were lysed by sonication for 2s on and 10s off at 33% amplitude. As an initial purification step, the crude lysate was heated at 70°C for 10 min followed by centrifugation at 15644*g* for 40 min at 4°C. The recombinant proteins being from a thermophilic bacterium remained in the supernatant while the host proteins precipitated. The clear supernatant was applied to a pre-equilibrated pierce centrifuge column (Thermo Fisher Scientific) packed with Ni^{2+}-affinity resin (Qiagen, Hilden, Germany) and then incubated for 2 hrs for the protein binding. The protein purification was performed in gradient way using wash buffer A (20 mM Tris-HCl pH 7.5, 10 mM imidazole, 150 mM NaCl, 10% glycerol, 1 mM PMSF and 3 mM β-ME) and B (20 mM Tris-HCl pH 7.5, 20 mM imidazole and 150 mM NaCl). Firstly, the affinity column was washed with wash buffers A and then with wash buffer B. The bound recombinant proteins were eluted from the column with 250 mM of imidazole in wash buffer B. Purity of the eluted fractions were analyzed in SDS-PAGE and the elutions were subjected to the gradient dialysis for imidazole removal against 20 mM Tris-HCl pH 7.5 and 150 mM NaCl. After that, the proteins βGlyBP_WT, βGlyBP_W41A, βGlyBP_W67A, βGlyBP_E117A, βGlyBP_W177X and βGlyBP_W256A were concentrated with the Vivaspin turbo 15 (10 kDa cutoff; Sartorius, Göttingen, Germany) up to ~23, ~42, ~51, ~41, ~32 and ~39 mg ml^{-1}, respectively (Figure 5.2). Protein concentration was estimated by measuring the

absorbance at 280 nm followed by calculation using the theoretical extinction coefficient (ε_{280} = 97525 M^{-1} cm^{-1}).

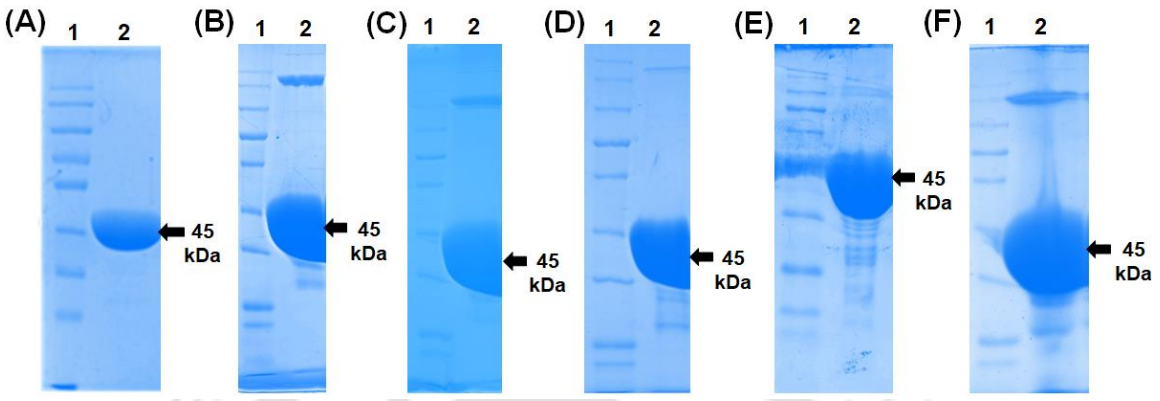


Figure 5.2. Purification of βGlyBP_WT and mutant protein. SDS-PAGE analysis of purified and concentrated (A) βGlyBP_WT protein (B) βGlyBP_W41 mutant protein (C) βGlyBP_W67A mutant protein (D) βGlyBP_E117A mutant protein (E) βGlyBP_W177X mutant protein and (F) βGlyBP_W256A mutant protein. (lane 1: protein ladder, lane 2: concentrated protein).

5.2.4 Fluorescence spectroscopy

To identify the physiological ligand(s) for the protein βGlyBP, different types of carbohydrate(s) were screened using fluorescence spectroscopy. The protein βGlyBP possesses four tryptophan residues at its active site and thus can be utilized to measure the changes in intrinsic fluorescence emission. The protein and carbohydrate(s) both with final concentration of 1 µM and 1 mM, respectively, was prepared in 20 mM Tris-HCl pH 7.5 and 150 mM NaCl. Before measurement, samples were incubated at 4°C for 2 hrs and subjected to the steady-state fluorescence measurements. The measurements were performed in a 10 mm path-length quartz cuvette for both protein and protein-carbohydrate complex at room temperature (25°C) in a FluoroMax-4 fluorimeter. Both protein and protein-carbohydrate complex were excited at a fixed excitation wavelength of 285 nm with a slit width of 5 nm and monitored for emission spectra from 300 to 500 nm. For each measurement, three averaging scans were recorded and subtracted with the blank for the actual fluorescence. Ligand binding was measured by estimating the quenching values of protein fluorescence attributed by the energy transfer.

5.2.5 Crystallization of wild type and mutant βGlyBP

To obtain the protein crystals, the purified wild type protein (βGlyBP_WT) at 16 mg ml^{-1} concentration was subjected to the initial crystal screening by using Crystal Screen and PEG/Ion kits from Hampton Research (CA, USA). The initial crystal screening was performed with a protein:buffer ratio of 1:1 using microbatch-under-oil method at 4 and 20ºC. Initial crystal hits for the apo βGlyBP_WT were obtained in a solution containing 0.2 M ammonium sulphate and 30% (w/v) PEG 8000. Subsequently, these crystals were optimized by using hanging-drop vapor-diffusion method. Although diffractable crystals of the apo βGlyBP_WT were obtained by mixing equal volumes of protein and crystallization condition (0.2 M ammonium sulphate and 30% (w/v) PEG 8000) at 20ºC; crystals grew after almost nine months of incubation (Figure 5.3A). In parallel, co-crystals of βGlyBP_WT with different probable ligands (as mentioned in Table C.1) were being attempted. For that, the protein βGlyBP_WT (1.33 mM) was incubated with ligands (150 mM) prepared in a buffer containing 20 mM Tris-HCl pH 7.5 and 150 mM NaCl for overnight at 4ºC.

Co-crystals of βGlyBP_WT incubated with ligands were obtained in a similar condition as that of without ligands, however, they also grew in almost nine months. Thus, the condition was optimized by varying the PEG 8000 concentration (30-70%) and fortunately several co-crystals (incubated with non-cognate ligands e.g. melibiose (MLB), raffinose (RAF), tagatose (TAG) and fucose (FUC) appeared within two months (Figure 5.3B). Furthermore, the microseed matrix screening (MMS, D'Arcy et al., 2007) method was applied to those crystals which required more than two months to grow to expedite the crystal growth. The MMS method was performed by mixing the protein with a concentrated seed stock in 3:1 ratio and by setting up the crystallization and co-crystallization with cognate ligand such as sophorose (SOP2), cellobiose (CEL2), cellotriose (CEL3) and cellotetraose (CEL4) in microbatch-under-oil and hanging-drop vapor-diffusion method at 20ºC; this yielded suitable crystals within two months (Figure 5.3).

Interestingly, crystals of mutants βGlyBP_W41A, βGlyBP_W67A, βGlyBP_E117A and βGlyBP_W177X appeared within 1-3 months in the same crystallization condition except

the concentration of the mutant proteins as described for the βGlyBP_WT (Figure 5.3). The co-crystals of βGlyBP_W177X mutant with LAM2, CEL2, SOP2 and GEN2 were obtained within one month by varying the PEG 8000 concentration (range 30-60%) while with CEL3, CEL4, CEL5, LAM3 and LAM4 were achieved by employing MMS method within 1-2 months. Similarly, co-crystals of βGlyBP_W67A with GEN2 were accomplished by varying the PEG 8000 concentration (range 30-60%). Co-crystals of βGlyBP_W41A mutant with SOP2, GEN2, CEL3 and CEL4 were grown within 1-3 months using MMS method (Figure 5.3).

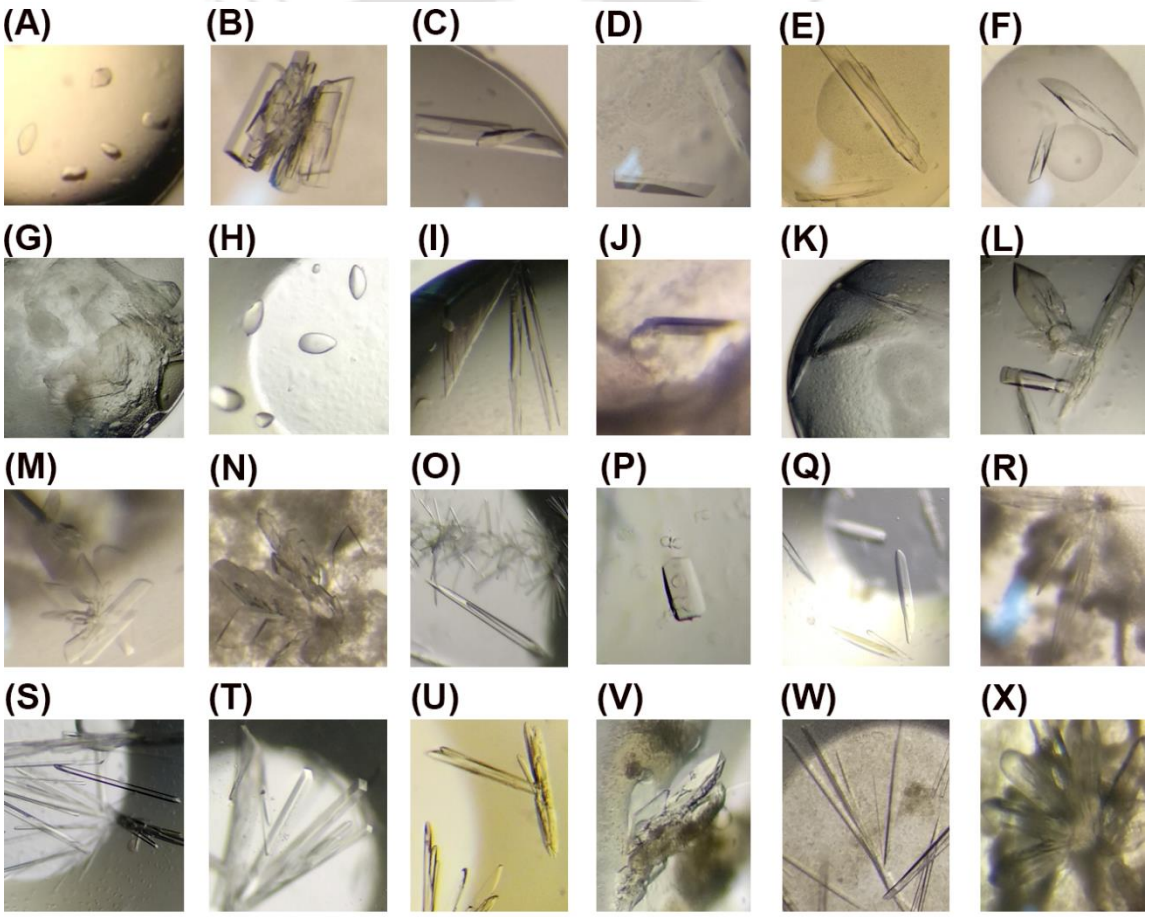


Figure 5.3. Crystallization of protein βGlyBP. (A) Protein crystals of βGlyBP_WT in microbatch. Co-crystal of βGlyBP_WT with (B) fucose (C) sophorose, (D) cellobiose, (E) cellotriose and (F) cellotetraose, (G) Protein crystals of βGlyBP_W41A, (H) Protein crystals of βGlyBP_W41A in presence of sophorose, (I) Protein crystals of βGlyBP_W67A, (J) Protein crystals of βGlyBP_W67A with gentiobiose, (K) Protein crystals of βGlyBP_E117A, (L) Protein crystals of βGlyBP_W177X. Protein crystals of βGlyBP_W177XA with (M) sophorose, (N) laminaribiose, (O and P) laminaritriose, (Q)

laminaritetraose, (R) cellobiose, (S and T) cellotriose, (U and V) cellotetraose, (W) cellopentaose and (X) gentiobiose.

5.2.6 Data collection, processing and structure determination

As all the crystals of the wild type and mutant protein βGlyBP (apo and holo) were grown in high concertation of PEG 8000, hence no further cryo-protectant was used during flash cooling. X-ray diffraction data for all the crystals were collected at -173°C using Cu_*Kα* radiation (wavelength 1.5418 Å) generated by Rigaku MicroMax-007 HF diffractometer (operated at 40 kV and 30 mA) and R-Axis IV++ image-plate detector available at Central Instruments Facility (CIF) of Indian Institute of Technology Guwahati (IITG), India. The data for all the crystals were collected with 1° oscillation and 300s exposure time. All the X-ray data sets were processed and scaled using the programs iMosflm (Battye et al., 2011) and Aimless (Evans and Murshudov, 2013) embedded in the package CCP4 (Winn, 2011). The X-ray intensities were converted to structure factors using program module ctruncate of CCP4. The structure solution of βGlyBP_WT was obtained by molecular replacement method using the program Phaser (McCoy et al., 2007) using the crystal structure of sugar-binding protein MalE from *Xanthomonas citri* (PDB ID: 3UOR) as the search model, which shares a sequence identity (query coverage) of 27(97)% with the protein βGlyBP. The Matthews coefficient (Matthews, 1968) indicated the presence of one molecule in the asymmetric unit. The structure solution of holo βGlyBP_WT and apo & holo forms of mutant proteins βGlyBP_W41A, βGlyBP_W67A, βGlyBP_E117A and βGlyBP_W177X were obtained by the direct molecular replacement method using the program Phaser and the three-dimensional atomic coordinates of the apo βGlyBP_WT as the search model.

A total of 5% of the reflections were kept aside from the data set for the calculation of R_{free} (Brunger, 1992). All the structural refinement of the model was carried out by using program Refmac5 (Vagin, 2004) with a default set of parameters. All the model building was carried out manually using the program Coot (Emsley et al., 2010). Although, in case of all the complexed structures, a clear *Fo*−*Fc* electron density maps could be observed for the bound ligand(s) after the first cycle of the refinement, the protein atoms were firstly fitted into the electron density contoured at 3.0σ and 1.0σ for the *Fo-Fc* and *2Fo-Fc* maps,

respectively. Subsequently, electron density observed for water molecules and other molecules present in the protein buffer and the crystallization conditions were modeled. After each cycle of the model building, structure refinement was performed using the same set of parameters. Finally, the ligands bound to the protein molecule were fitted in the electron density using the difference map contoured at 3.0σ. The geometric parameters of all the refined structures were checked and validated by the programs Procheck (Laskowski et al., 1993) and MolProbity (Chen et al., 2010). All the refinement and validation statistics of the refined models are provided in the Table 5.2-5.7. The atomic coordinates and structure factors of all the refined models have been deposited in the RCSB Protein Data Bank (Berman et al., 2000). All the molecular graphic figures presented in this study were generated using the program PyMOL (Molecular Graphics System, Version 2.1.1 Schrödinger, LLC).

Table 5.2. X-ray crystallographic data collection and refinement statistics for the protein βGlyBP_WT (apo and complexed with CEL2 and CEL3). The values in parenthesis represent the statistics for the last resolution shell.

	βGlyBP_WT_FormI	**βGlyBP_WT_FormII**	**βGlyBP_WT•CEL2**	**βGlyBP_WT•CEL3**
Wavelength (Å)	1.5418	1.5418	1.5418	1.5418
Temperature (K)	100	100	100	100
Space group	$P2_12_12_1$	$P2_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å, °)	a=58.52, b=62.93, c=104.02, $\alpha=\beta=\gamma$=90	a=57.46, b=100.71, c=66.26, $\alpha=\gamma$=90, β=104.19	a=64.24, b=110.82, c=110.91, $\alpha=\beta=\gamma$=90	a=62.67, b=109.88, c=110.89, $\alpha=\beta=\gamma$=90
Resolution (Å)	51.00-1.63 (1.66-1.63)	50.35-1.63 (1.66-1.63)	64.24-2.30 (2.38-2.30)	62.67-2.00 (2.05-2.00)
No. of observed reflections	440139 (19653)	415172 (18698)	252397 (22922)	368548 (25412)
No. of unique reflections	48845 (2397)	91079 (4434)	35932 (3469)	52166 (3744)
Mn(I) CC(1/2)	0.999 (0.961)	0.998 (0.963)	0.984 (0.911)	0.996 (0.882)

Completeness (%)	100.0 (99.8)	99.9 (98.0)	99.9 (99.9)	99.4 (98.3)
V_M ($Å^3$ Da^{-1})	2.09	2.03	2.15	2.08
Solvent content (%)	41.16	39.36	42.90	40.96
Mosaicity (°)	0.40	0.70	0.70	0.80
Mean I/σ(I)	24.7 (4.9)	19.3 (6.5)	8.2 (3.4)	9.6 (3.1)
R_{merge}† (%)	4.7 (28.8)	4.5 (16.8)	17.8 (47.3)	10.1 (48.8)
R_{pim} (%)	2.4 (15.8)	3.6 (13.2)	10.2 (28.7)	5.8 (29.0)
R_{meas} (%)	5.3 (33.0)	5.8 (21.4)	20.4 (55.6)	11.7 (56.9)
Multiplicity	9.0 (8.2)	4.6 (4.2)	7.0 (6.6)	7.1 (6.8)
R_{work}/R_{free} (%)	16.40/20.80	13.26/16.58	19.22/24.84	16.98/23.06
Protein model				
No. of subunits in ASU	1	2	2	2
Protein atoms	3237	6436	6407	6421
Water molecules	464	934	423	419
Carbohydrates (number)	-	-	CEL2 (2)	CEL3 (2)
Other molecules	7	23	10	14
Deviation from ideal geometry				
Bond length (Å)	0.017	0.021	0.013	0.013
Bond angles (°)	2.163	2.380	1.748	1.766
Average *B*-factor ($Å^2$)				
Protein atoms	16.51	10.99	22.27	23.13
Water molecules	32.24	25.06	27.45	34.33
Carbohydrate	-	-	CEL2 (17.15)	CEL3 (26.65)
Ramachandran plot (%)				
Favored	98.02	97.28	97.72	97.57
Allowed	1.98	2.72	2.28	2.43
Remaining	0.00	0.00	0.00	0.00
PDB ID	7C63	7C64	7C66	7C67

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

Abbreviations: CEL2, cellobiose; CEL3, cellotriose.

Table 5.3. X-ray crystallographic data collection and refinement statistics of the protein βGlyBP_WT (complexed with CEL4 and SOP2) and mutant βGlyBP_W177X (apo and complexed with GEN2). The values in parenthesis represent the statistics for the last resolution shell.

	βGlyBP_WT•CEL4	**βGlyBP_WT•SOP2**	**βGlyBP_W177X**	**βGlyBP_W177X•GEN2**
Wavelength (Å)	1.5418	1.5418	1.5418	1.5418
Temperature (K)	100	100	100	100
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_1$
Unit-cell parameters (Å, °)	a=62.63, b=109.74, c=111.57, $\alpha=\beta=\gamma$=90	a=62.79, b=108.93, c=109.99, $\alpha=\beta=\gamma$=90	a=58.21, b=63.10, c=103.79, $\alpha=\beta=\gamma$=90	a=62.34, b=91.54, c=69.39, $\alpha=\gamma$=90, β=112.51
Resolution (Å)	55.78-2.05 (2.11-2.05)	62.79-2.35 (2.43-2.35)	53.92-1.70 (1.73-1.70)	48.75-1.90 (1.94-1.90)
No. of observed reflections	352325 (27050)	279970 (25078)	278893 (12758)	252521 (14977)
No. of unique reflections	48888 (3760)	32179 (3110)	41913 (2061)	53435 (3258)
Mn(I) CC(1/2)	0.993 (0.920)	0.984 (0.901)	0.995 (0.922)	0.995 (0.891)
Completeness (%)	99.7 (99.4)	99.9 (99.7)	98.1 (94.0)	94.5 (89.3)
V_M (Å^3 Da^{-1})	2.09	2.05	2.08	1.99
Solvent content (%)	41.21	40.08	40.87	38.38
Mosaicity (°)	0.90	0.60	0.60	0.80
Mean I/σ(I)	9.6 (3.2)	9.6 (3.8)	14.5 (4.2)	10.6 (2.8)
R_{merge}† (%)	12.3 (46.4)	19.0 (47.9)	8.1 (34.2)	7.6 (38.7)
R_{pim} (%)	7.2 (27.8)	9.8 (26.1)	4.8 (20.9)	6.0 (29.4)
R_{meas} (%)	14.3 (54.3)	21.4 (54.7)	9.4 (40.2)	9.7 (48.9)
Multiplicity	7.2 (7.2)	8.7 (8.1)	6.7 (6.2)	4.7 (4.6)
R_{work}/R_{free} (%)	18.18/22.41	19.96/26.23	17.38/21.58	17.35/22.73
Protein model				
No. of subunits in ASU	2	2	1	2

Protein atoms	6373	6388	3211	6439
Water molecules	463	264	401	470
Carbohydrate (number)	CEL4 (2)	SOP2 (2)	-	GEN2 (2)
Other molecules	15	6	10	10
Deviation from ideal geometry				
Bond length (Å)	0.013	0.011	0.014	0.014
Bond angles (°)	1.828	1.764	1.909	1.920
Average *B*-factor (Å^2)				
Protein atoms	19.89	19.10	19.30	21.69
Water molecules	30.78	26.04	31.44	32.03
Carbohydrate	CEL4 (21.42)	SOP2 (17.98)	-	GEN2 (35.70)
Ramachandran plot (%)				
Favored	97.09	97.23	97.56	96.96
Allowed	2.91	2.65	2.44	3.04
Remaining	0.00	0.12	0.00	0.00
PDB ID	7C68	7C69	7C6F	7C6G

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

Abbreviations: CEL4, cellotetraose; SOP2, sophorose; GEN2, gentiobiose.

Table 5.4. X-ray crystallographic data collection and refinement statistics of the mutant βGlyBP_W177X (complexed with LAM2, SOP2, CEL2 and CEL3). The values in parenthesis represent the statistics for the last resolution shell.

	βGlyBP_W177X•LAM2	**βGlyBP_W177X•SOP2**	**βGlyBP_W177X•CEL2**	**βGlyBP_W177X•CEL3_Form I**
Wavelength (Å)	1.5418	1.5418	1.5418	1.5418
Temperature (K)	100	100	100	100
Space group	$P2_1$	$P2_1$	$P2_12_12_1$	$P2_12_12_1$

Unit-cell parameters (Å, °)	*a*=62.51, *b*=92.25, *c*=69.78, *α*=*γ*=90, *β*=112.42	*a*=62.64, *b*=92.06, *c*=70.01, *α*=*γ*=90, *β*=112.85	*a*=60.90, *b*=100.25, *c*=134.05, *α*=*β*=*γ*=90	*a*=58.58, *b*=62.96, *c*=109.89, *α*=*β*=*γ*=90
Resolution (Å)	54.62-1.85 (1.89-1.85)	54.90-1.70 (1.73-1.70)	55.72-2.10 (2.16-2.10)	58.58-1.77 (1.81-1.77)
No. of observed reflections	271996 (16488)	377894 (17665)	364377 (26595)	355851 (18113)
No. of unique reflections	62461 (3890)	80275 (4085)	48058 (3677)	39933 (2178)
Mn(I) CC(1/2)	0.995 (0.814)	0.993 (0.883)	0.994 (0.928)	0.999 (0.910)
Completeness (%)	100.0 (100.0)	99.8 (96.5)	98.7 (93.9)	99.4 (96.9)
V_M ($Å^3$ Da^{-1})	2.03	2.03	2.33	2.21
Solvent content (%)	39.40	39.41	44.92	40.39
Mosaicity (°)	0.50	0.60	0.60	0.25
Mean I/σ(I)	10.5 (2.6)	11.1 (2.9)	12.8 (4.0)	21.2 (4.0)
R_{merge}† (%)	8.3 (50.3)	8.3 (39.1)	10.8 (45.4)	6.1 (46.5)
R_{pim} (%)	6.9 (41.0)	6.4 (30.2)	5.9 (24.5)	3.0 (24.7)
R_{meas} (%)	10.9 (65.2)	10.5 (49.6)	12.3 (51.7)	6.8 (52.9)
Multiplicity	4.4 (4.2)	4.7 (4.3)	7.6 (7.2)	8.9 (8.3)
R_{work}/R_{free} (%)	16.82/21.67	16.57/20.73	17.84/23.66	16.58/21.74
Protein model				
No. of subunits in ASU	2	2	2	1
Protein atoms	6466	6450	6383	3182
Water molecules	630	817	502	378
Carbohydrate (number)	LAM2 (2)	SOP2 (2)	CEL2 (2)	CEL3 (1)
Other molecules	11	10	6	6
Deviation from ideal geometry				
Bond length (Å)	0.013	0.014	0.014	0.013

Bond angles (°)	1.849	1.984	1.866	1.949
Average *B*-factor (Å^2)				
Protein atoms	18.19	17.47	18.86	18.82
Water molecules	30.25	30.29	29.60	32.20
Carbohydrate	LAM2 (19.04)	SOP2 (23.52)	CEL2 (18.00)	CEL3 (19.18)
Ramachandran plot (%)				
Favored	97.33	97.94	97.32	98.04
Allowed	2.67	2.06	2.55	1.96
Remaining	0.00	0.00	0.12	0.00
PDB ID	7C6H	7C6I	7C6J	7C6K

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

Abbreviations: LAM2, laminaribiose; SOP2, sophorose; CEL2, cellobiose; CEL3, cellotriose.

Table 5.5. X-ray crystallographic data collection and refinement statistics of the mutant βGlyBP_W177X (complexed with CEL3, CEL4 and CEL5). The values in parenthesis represent the statistics for the last resolution shell.

	βGlyBP_W177X•CEL3_FormII	**βGlyBP_W177X•CEL4_FormI**	**βGlyBP_W177X•CEL4_FormII**	**βGlyBP_W177X•CEL5**
Wavelength (Å)	1.5418	1.5418	1.5418	1.5418
Temperature (K)	100	100	100	100
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å, °)	a=61.00, b=100.34, c=133.23, $\alpha=\beta=\gamma$=90	a=58.32, b=62.93, c=110.26, $\alpha=\beta=\gamma$=90	a=60.85, b=100.17, c=133.06, $\alpha=\beta=\gamma$=90	a=58.23, b=63.23, c=110.08, $\alpha=\beta=\gamma$=90
Resolution (Å)	66.61-2.40 (2.49-2.40)	58.32-1.90 (1.94-1.90)	66.53-2.05 (2.11-2.05)	58.23-2.10 (2.16-2.10)
No. of observed reflections	217551 (20686)	279959 (17278)	450232 (31718)	192624 (15215)
No. of unique reflections	32723 (3337)	32723 (2063)	50982 (3827)	22403 (1863)

Mn(I) CC(1/2)	0.994 (0.905)	0.998 (0.920)	0.998 (0.924)	0.992 (0.912)
Completeness (%)	99.9 (99.5)	99.9 (99.6)	98.5 (96.6)	92.0 (94.9)
V_M ($Å^3$ Da^{-1})	2.22	2.21	2.21	2.21
Solvent content (%)	44.72	44.30	44.42	44.39
Mosaicity (°)	0.60	0.50	0.40	0.70
Mean I/σ(I)	11.4 (3.6)	15.8 (3.5)	17.1 (4.4)	10.3 (3.6)
R_{merge}† (%)	12.1 (47.7)	7.7 (46.9)	8.8 (44.1)	12.7 (44.6)
R_{pim} (%)	7.4 (29.6)	4.0 (24.3)	4.5 (23.2)	6.7 (23.2)
R_{meas} (%)	14.2 (56.3)	8.7 (53.0)	9.8 (49.9)	14.4 (50.4)
Multiplicity	6.6 (6.2)	8.6 (8.4)	8.8 (8.3)	8.6 (8.2)
R_{work}/R_{free} (%)	19.11/26.33	17.85/22.90	16.12/21.98	23.01/29.16
Protein model				
No. of subunits in ASU	2	1	2	1
Protein atoms	6361	3173	6371	3174
Water molecules	327	240	648	94
Carbohydrate (number)	CEL3 (2)	CEL4 (1)	CEL4 (2)	CEL5 (1)
Other molecules	8	4	8	3
Deviation from ideal geometry				
Bond length (Å)	0.013	0.014	0.014	0.011
Bond angles (°)	1.846	1.928	1.829	1.803
Average *B*-factor ($Å^2$)				
Protein atoms	22.05	23.32	20.52	19.93
Water molecules	26.31	34.17	31.30	29.15
Carbohydrate	CEL3 (25.34)	CEL4 (31.20)	CEL4 (25.33)	CEL5 (31.15)
Ramachandran plot (%)				
Favored	96.74	98.54	98.05	95.85
Allowed	3.26	1.46	1.95	4.15
Remaining	0.00	0.00	0.00	0.00

PDB ID	7C6L	7C6M	7C6N	7C6R

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

Abbreviations: CEL3, cellotriose; CEL4, cellotetraose; CEL5, cellopentaose.

Table 5.6. X-ray crystallographic data collection and refinement statistics of the mutants βGlyBP_W177X (complexed with LAM3 and LAM4) and βGlyBP_W41A (apo). The values in parenthesis represent the statistics for the last resolution shell.

	βGlyBP_W177X•LAM3_FormI	**βGlyBP_W177X•LAM3_FormII**	**βGlyBP_W177X•LAM4**	**βGlyBP_W41A_FormI**
Wavelength (Å)	1.5418	1.5418	1.5418	1.5418
Temperature (K)	100	100	100	100
Space group	$P2_12_12_1$	$P3_121$	$P2_12_12_1$	$P1$
Unit-cell parameters (Å, °)	a=57.93, b=63.55, c=110.23, $\alpha=\beta=\gamma$=90	$a=b$=185.41, c=74.18, $\alpha=\beta$=90, γ=120	a=58.11, b=62.60, c=109.2, $\alpha=\beta=\gamma$=90	a=80.99, b=116.40, c=132.14, α=83.39, β=88.73, γ=89.95
Resolution (Å)	57.93-2.30 (2.38-2.30)	80.28-2.80 (2.92-2.80)	58.11-1.88 (1.92-1.88)	82.19-2.65 (2.70-2.65)
No. of observed reflections	165548 (15036)	354628 (40473)	240945 (14248)	463080 (20071)
No. of unique reflections	18713 (1771)	36336 (4404)	32582 (1978)	130991 (6096)
Mn(I) CC(1/2)	0.996 (0.961)	0.999 (0.933)	0.998 (0.920)	0.793 (0.806)
Completeness (%)	99.8 (99.3)	100.0 (100.0)	98.5 (95.6)	94.1 (88.3)
V_M ($Å^3$ Da^{-1})	2.21	2.68	2.16	2.25
Solvent content (%)	44.46	54.07	43.16	45.34
Mosaicity (°)	0.60	0.44	0.75	0.80

Mean I/σ(I)	16.1 (5.6)	13.8 (4.5)	15.2 (3.5)	3.5 (2.0)
R_{merge}† (%)	9.9 (33.8)	9.8 (50.1)	7.7 (43.3)	23.6 (37.2)
R_{pim} (%)	4.9 (16.7)	4.8 (25.8)	4.2 (24.1)	23.6 (37.2)
R_{meas} (%)	11.1 (37.8)	10.9 (56.5)	8.8 (49.6)	33.4 (52.7)
Multiplicity	8.8 (8.5)	9.8 (9.2)	7.4 (7.2)	3.5 (3.3)
R_{work}/R_{free} (%)	20.26/26.49	18.66/23.56	18.71/24.23	24.30/26.80
Protein model				
No. of subunits in ASU	1	3	1	12
Protein atoms	3162	9469	3183	37819
Water molecules	139	52	319	484
Carbohydrate (number)	LAM3 (1)	LAM3 (3)	LAM4 (1)	-
Other molecules	6	13	5	22
Deviation from ideal geometry				
Bond length (Å)	0.014	0.013	0.013	0.012
Bond angles (°)	1.931	1.841	1.880	1.622
Average *B*-factor (Å^2)				
Protein atoms	17.96	32.80	16.98	7.70
Water molecules	27.50	36.99	33.74	4.60
Carbohydrate	LAM3 (26.91)	LAM3 (45.92)	LAM4 (29.94)	-
Ramachandran plot (%)				
Favored	97.07	95.20	96.80	96.80
Allowed	2.93	4.80	3.20	3.20
Remaining	0.00	0.00	0.00	0.00
PDB ID	7C6T	7C6V	7C6W	7C6X

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

Abbreviations: LAM3, laminaritriose; LAM4, laminaritetraose.

Table 5.7. X-ray crystallographic data collection and refinement statistics of the mutants βGlyBP_W41A (apo), βGlyBP_W67A (apo and complexed form with GEN2) and βGlyBP_E117A (apo). The values in parenthesis represent the statistics for the last resolution shell.

	βGlyBP_W41A_FormII	**βGlyBP_W67A**	**βGlyBP_W67A•GEN2**	**βGlyBP_E117A**
Wavelength (Å)	1.5418	1.5418	1.5418	1.5418
Temperature (K)	100	100	100	100
Space group	$P2_12_12_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters (Å, °)	a=58.35, b=63.25, c=103.95, $\alpha=\beta=\gamma$=90	a=55.26, b=97.79, c=66.61, $\alpha=\gamma$=90, β=103.42	a=55.24, b=98.96, c=66.56, $\alpha=\gamma$=90, β=103.25	a=57.40, b=100.96, c=66.52, $\alpha=\gamma$=90, β=104.60
Resolution (Å)	54.03-1.70 (1.73-1.70)	48.90-1.63 (1.66-1.63)	49.48-1.63 (1.66-1.63)	55.55-2.10 (2.16-2.10)
No. of observed reflections	373282 (17715)	349654 (16083)	618337 (27904)	173623 (13800)
No. of unique reflections	41739 (2100)	78922 (3823)	85916 (4142)	42869 (3480)
Mn(I) CC(1/2)	0.997 (0.939)	0.996 (0.838)	0.995 (0.918)	0.981 (0.812)
Completeness (%)	97.2 (94.3)	92.4 (90.8)	99.0 (97.0)	100.0 (100.0)
V_M ($Å^3$ Da^{-1})	2.09	1.91	1.93	2.03
Solvent content (%)	41.25	35.62	36.36	39.58
Mosaicity (°)	0.60	1.07	0.60	0.80
Mean I/σ(I)	21.3 (5.7)	9.4 (2.6)	15.9 (4.0)	5.9 (2.6)
R_{merge}† (%)	7.1 (32.9)	7.4 (40.1)	7.6 (38.9)	14.7 (39.7)
R_{pim} (%)	3.6 (17.4)	6.0 (32.0)	4.6 (24.7)	13.4 (36.2)
R_{meas} (%)	7.9 (37.3)	9.6 (51.4)	8.8 (46.3)	20.0 (54.0)
Multiplicity	8.9 (8.4)	4.4 (4.2)	7.2 (6.7)	4.1 (4.0)
R_{work}/R_{free} (%)	17.23/21.08	15.75/19.59	13.65/17.81	17.90/24.39
Protein model				
No. of subunits in ASU	1	2	2	2

Protein atoms	3241	6392	6495	6362
Water molecules	396	782	832	611
Carbohydrate (number)	-	-	GEN2 (2)	-
Other molecules	7	17	17	12
Deviation from ideal geometry				
Bond length (Å)	0.017	0.015	0.017	0.011
Bond angles (°)	2.128	1.963	2.147	1.646
Average *B*-factor (Å^2)				
Protein atoms	16.28	13.37	12.74	14.99
Water molecules	29.09	26.29	26.17	22.50
Carbohydrate	-	-	GEN2 (30.79)	-
Ramachandran plot (%)				
Favored	98.27	97.79	98.00	96.47
Allowed	1.73	2.21	2.00	3.53
Remaining	0.00	0.00	0.00	0.00
PDB ID	7C6Y	7C6Z	7C70	7C71

† $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl, $\sum hkl$ is the sum overall reflections and $\sum i$ is the sum over i measurements of reflection hkl.
Abbreviation: GEN2, gentiobiose.

5.2.7 Measurement of ligand binding affinity using isothermal titration calorimetry

The binding affinity of different β-glycosides with the protein βGlyBP was measured by isothermal titration calorimetry (MicroCal ITC_{200}; GE healthcare) method. The protein and ligand solutions were prepared in the dialyzed buffer containing 20 mM Tris-HCl pH 7.5 and 150 mM NaCl. The protein βGlyBP_WT (100 µM) was filled in the sample cells which was titrated with different ligands (2.5-5 mM) mentioned in the Table C.1. During titration, the sample cell was stirred at 250/450 rpm and ligands filled in the syringe were injected 24 times each of 1.5 µl (excluding 0.4 µl pre-injection) of injection volume at an interval of 120/150s. Binding kinetics of βGlyBP_WT was measured at two different temperatures 25 and 70°C. A similar protocol was followed for measuring the binding kinetics of

different β-glycosides with mutant proteins (βGlyBP_W41A, βGlyBP_W67A, βGlyBP_W177X and βGlyBP_W256A) at 25°C. Along with each titration, a control experiment was also performed for measuring the heat of dilution, by titrating β-glycosides into the dialyzed buffer. The heat of dilution obtained for each control experiment was subtracted from each protein-ligand titration to estimate the actual ligand binding affinity. The reaction heat was integrated using the program Origin (version 7.0) provided with the MicroCal software. The thermodynamic parameters such as association constant (K_a), enthalpy (ΔH) and entropy (ΔS) changes and stoichiometric ratio (number of binding sites, n) were estimated by integrating the reaction heat using a non-linear regression and one-site binding model. The thermodynamic parameter Gibbs free energy change (ΔG) was calculated using the thermodynamic equation $\Delta G = \Delta H - T\Delta S$, *where* T is the reaction temperature. The estimated values for these thermodynamic parameters of each reaction are provided in the Table 5.8 and Table C.1.

5.2.8 Thermal denaturation studies using circular dichroism

The Far-UV circular dichroism (CD) spectra of the protein β-GlyBP_WT was recorded between 190 to 260 nm wavelength and 1 mm optical path length quartz cuvette using the Jasco J-1500 spectrometer (Jasco, Germany). The protein of βGlyBP_WT prepared in a buffer containing 20 mM Tris-HCl pH 7.5 and 150 mM NaCl was diluted into the water with a final concentration of 5 μM. For the spectra measurement, the parameters like response, sensitivity and scan speed were fixed as 2s, 100 milli degrees and 100 nm min^{-1}, respectively. Each spectra measurement includes a three averaging scans and subtraction from blank containing the same buffer as the reference. Analysis of the spectra was carried out by using Spectra Manager software provided by JASCO. Using the optimized spectral profile of βGlyBP_WT, the thermal denaturation profile was analyzed. To monitor the thermal unfolding curves, the protein sample was subjected for the measurement of the CD signals at varying temperatures ranging from 20-120°C with heating rate of 2°C min^{-1}. The analysis of the secondary structural change was carried out between 190 to 260 nm wavelength. The final denaturation profile of the protein βGlyBP_WT was plotted as a function of temperature and wavelength using the software Origin (version 9.6).

5.2.9 Architecture of the genetic operon for β-glycosides metabolism

Genes functionally associated to the protein βGlyBP was identified using the genomic context in Gene database of National Center for Biotechnology Information (NCBI). Each flanking gene of the protein βGlyBP was manually located by analyzing the genetic context for the upstream and downstream regions. These genes were also ascertained whether they belonged to the family of the ABC transporter or glycoside hydrolases (GH). The details of the functional association of GH genes with the β-glycosides metabolism were retrieved from the CAZy database (Lombard et al., 2014). A similar approach was employed to construct the β-glycosides operon of other SBPs from the subclusters C-IV and D-I. The structural homologs from the subcluster D-I SBPs (β-glycosides specific) were searched using the web server Dali (Holm and Rosenstrom, 2010) while that of the subcluster C-IV were compiled from a previous study (Fukamizo et al., 2019). However, those SBPs for which genomic context is unavailable in the Gene database were excluded from the study. The details of each gene and its subcellular location was retrieved from the UniProtKB database (The UniProt Consortium, 2019).

5.3 RESULTS

5.3.1 Fluorescence and thermodynamic data suggest conformational rearrangement of the protein βGlyBP upon β-glycosides binding

In a previous *in silico* study, we proposed the uptake of β-glycosides through an ABC import system encoded by ORF IDs: TTHB082-TTHB086 (Chandravanshi et al., 2019). To corroborate the hypothesis experimentally, a set of carbohydrates were screened to estimate the binding to the SBP subunit (ORF ID: TTHB082) of the ABC importer using the fluorescence-based approach. For this, various carbohydrates including α-glycosides (α-1,1; α-1,2; α-1,4 and α-1,6), β-glycosides (β-1,2; β-1,3; β-1,4 and β-1,6) having different glycosidic linkages and monosaccharides (glucose and its epimers) as detailed in Table C.1 were used to estimate the change in the fluorescence behavior of the protein TTHB082. Results depict that the emission maximum (λ_{max}) of the protein TTHB082 occur at 351 nm in the absence of ligand(s) while in the presence of ligand, the fluorescence intensity decreases and blue-shifts with the λ_{max} in between 345 to 340 nm (Figure 5.4A) indicating

a change in the local environment of tryptophan residues from hydrophilic to hydrophobic (Munishkina and Fink, 2007). Notably, the change in intensity and wavelength can be observed only in the presence of β-glycosides such as SOP2 (β-1,2), CEL2 (β-1,4) and GEN2 (β-1,6) with a blue shift of 6, 11 and 6 nm, respectively (Figure 5.4B). This indicates that the protein TTHB082 undergoes conformational changes in the presence of β-glycosides suggesting reorientation of the tryptophan residues, most likely, providing stacking interaction to the pyranose ring of the β-glycosides (Vázquez-Ibar et al., 2003).

To further probe the binding and specificity of the protein TTHB082 towards β-glycosides, isothermal titration calorimetry (ITC) experiments were performed with a similar set of carbohydrates. Akin to the fluorescence data, thermodynamic analysis also exhibited the binding of β-glycosides to the protein TTHB082 (Figure 5.4C-5.4E and Table C.1). On the other hand, no binding could be observed for the α-glycosides and monosaccharides (Table C.1). Thus, the protein TTHB082 was renamed as a β-glycosides-binding protein (βGlyBP). Among the various β-glycosides, only short-chain β-glycosides, SOP2, CEL2 and GEN2 show a significant binding to the protein βGlyBP with a dissociation constant (K_d) of 5.37, 0.15 and 178.00 μM, respectively (Table C.1), whereas the binding for polysaccharide β-glycoside laminarin (LAMn) and β-galactoside lactose (LAT) could not be detected (Table C.1). In summary, the fluorescence and thermodynamic data suggest a preferential binding of β-glycosides to the protein βGlyBP.

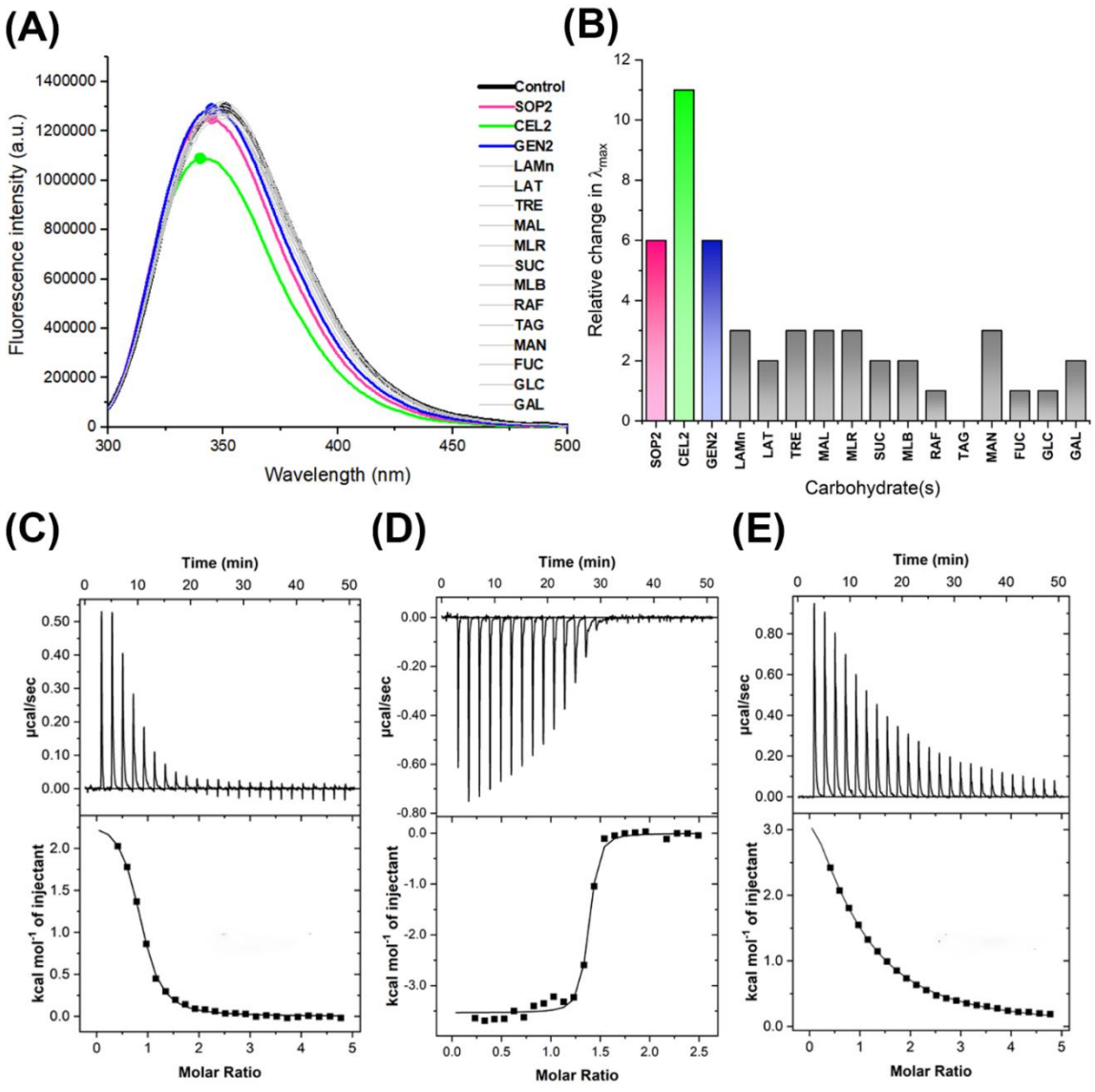


Figure 5.4. The specificity of the protein βGlyBP for β-glycosides. (A) Fluorescence emission spectra of the protein βGlyBP in the absence (black) and presence of β-glycosides (SOP2, CEL2, GEN2, LAMn and LAT), α-glycosides (TRE, MAL, MLR, SUC, MLB and RAF) and monosaccharides (TAG, MAN, FUC, GLC and GAL); details are provided in Table C.1. The change in fluorescence spectra of the protein βGlyBP due to the quenching effect of SOP2, CEL2 and GEN2 has been represented in magenta, green and blue, respectively, while that for other ligands in grey. (B) Relative changes in λ_{max} induced by SOP2 (magenta), CEL2 (green), GEN2 (blue) and other ligands (grey) are represented as histogram. (C-E) Plot showing the respective binding of SOP2, CEL2 and GEN2 to the protein βGlyBP obtained through isothermal titration calorimetry (ITC) data. The upper panels show the raw heat change upon ligand titration while the integrated heat pulse subtracted with their heat of dilution is shown in the lower panels.

5.3.2 The overall structure of the protein βGlyBP

For in-depth investigation of the structural determinants that govern the selection of β-glycosides over other carbohydrates, the protein βGlyBP was crystallized in both the absence and presence of all 14 carbohydrates mentioned in Table C.1. The apo βGlyBP crystallized in two space groups $P2_12_12_1$ (βGlyBP_WT_FormI) and $P2_1$ (βGlyBP_WT_FormII); both the crystals diffracted at 1.63 Å resolution. The asymmetric units of βGlyBP_WT_FormI and _FormII crystals contain one and two molecules, respectively (Figure C.1A-C.1D). The crystals of the protein βGlyBP_WT_FormII was obtained by incubating it with α-glycosides (melibiose, MLB and raffinose, RAF) and monosaccharides (fucose, FUC and tagatose, TAG) and high polyethylene glycol (PEG) concentration (60-70%). However, these carbohydrates could not be observed in the electron density map of the protein βGlyBP_WT_FormII indicating their binding incapability corroborating the fluorescence and ITC data.

The overall structure of the protein βGlyBP in both the forms is similar and typical to that of subcluster D-I SBPs. The structure possesses two α/β globular (N-and C-terminal) domains connected by three loops, which serve as hinge regions (H1, H2 and H3). The N-terminal domain (NTD) is composed of two N1 (3-115) and N2 (298-362) subdomains. Similarly, the C-terminal domain (CTD) is composed of two C1 (120-282) and C2 (372-416) subdomains (Figure 5.5A). Topologically, NTD comprises of five β-strands sandwiched between ten α-helices while CTD possesses three β-strands surrounded by eleven α-helices and two β-strands (Figure 5.5A).

Further, a comparison of the two forms βGlyBP_WT_FormI and _FormII delineates that the two molecules of βGlyBP_WT_FormII are related by a non-crystallographic 2-fold self-rotation juxta-positioning the NTD and CTD of protomers A and B, respectively (Figure C.1C and C1D). Furthermore, the electrostatic surface charge distributions of NTD (+ve) and CTD (-ve) compensate each other in this juxtaposed arrangement (Figure C.1E). In addition, a comparison of *B*-factor of the main-chain atoms of the two forms βGlyBP_WT_FormI and _FormII suggests the latter to be more stable than the former (Figure C.2A and C.2B). Overall, the structures of two forms (βGlyBP_WT_FormI and

_FormII) are similar with a root mean square deviation (rmsd) of 0.35 Å, however, structural differences are notable at the active site loops L1 (173-175) and L2 (264-272) (Figure C.2C).

In βGlyBP_WT_FormII, two PEG molecules present at the active site hold the two protein molecules together (Figure 5.5B and 5.5C). These PEG molecules form hydrogen bonds with the amino acid residues of the two subsites (I and II) present at the active site interface of the two protein molecules. At the subsite I, two PEG molecules coordinate with the residues Arg260 and Thr332 (protomer A) and Asp40 (protomer B) while at the subsite II, the residues Thr50 (protomer A) and Trp67 and Glu117 (protomer B) form hydrogen bonds with the PEG molecules (Figure 5.5C). The binding of the PEG molecules at the active site of the βGlyBP_WT_FormII alters the orientation of two active site aromatic residues (Trp177 and Trp256) from an outward to inward orientation leading to the transition from an open to partial-open state of the protein βGlyBP (Figure 5.5D and 5.5E).

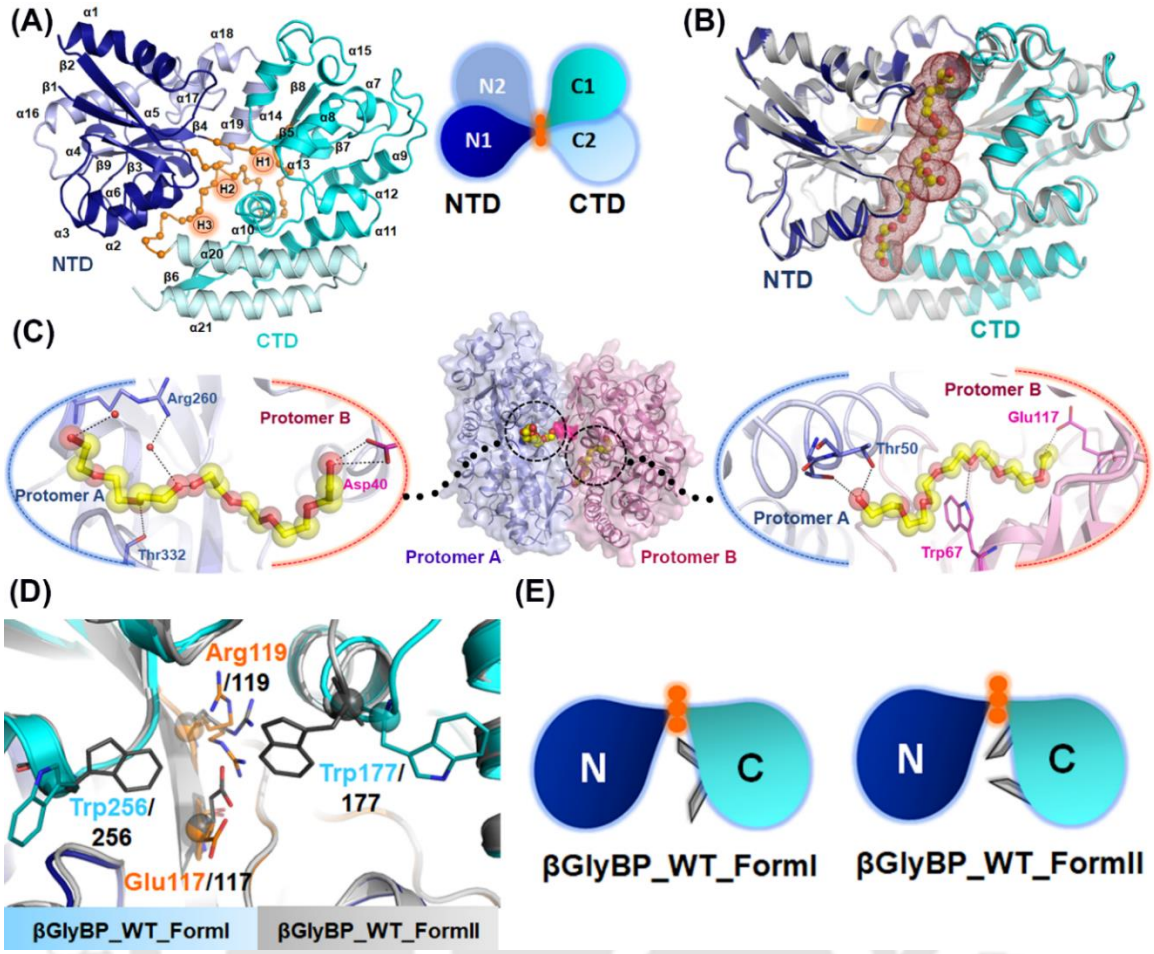


Figure 5.5. Overall structure of apo βGlyBP. (A) (Left) ribbon model of apo βGlyBP with NTD and CTD connected by three hinge regions (H1, H2 and H3 in orange) which are shown in blue and cyan, respectively. (Right) schematic representation of the four N1, N2, C1 and C2 subdomains organization of the protein βGlyBP. (B) Structural superimposition of the protein βGlyBP_WT_FormI (blue & cyan) and _FormII (grey) structures. The PEG molecules bound at the active site of the protein βGlyBP_WT_FormII is shown as yellow dotted-sphere. (C) Self-assembly of the protein βGlyBP via PEG molecules in the βGlyBP_WT_FormII structure. (Center) PEG molecules (yellow sphere) bound at the two subsites (I and II, encircled with dotted line) anchor two protein molecules (protomer A, blue and protomer B, pink) of βGlyBP_WT_FormII. (Left and right), active site residues holding the PEG molecules at subsite I and II, respectively, are shown as line. (D) Active site comparison of βGlyBP_WT_FormI (cyan) and _FormII (grey) structures displaying the conformational changes in the hinge region (orange) and CTD (cyan) residues. (E) Schematic models of two forms βGlyBP_WT_FormI (left) and βGlyBP_WT_FormII (right) of the apo βGlyBP_WT having a different orientation of residues Trp177 and Trp256 (grey lines) as outward and inward, respectively.

5.3.3 The protein βGlyBP exhibits a broad-range β-glycosides specificity under physiological conditions

Although the fluorescence and thermodynamic data demonstrated that the protein βGlyBP selectively binds to short-chain disaccharide β-glycosides, its structural homolog SOPn-binding protein (SO-BP, PDB ID: 5YSB) is reported to bind even higher β-glycosides (Abe et al., 2018). To examine this for the protein βGlyBP, ITC experiments with higher β-glycosides were performed which exhibited an apparent binding only for CEL3 and CEL4 (Figure C.3A and C.3B). Notably, these ITC experiments were performed at room temperature (25°C). However, considering the apparent binding of CEL3 and CEL4, ITC experiments were repeated at physiological growth temperature (70°C) of the bacterium *T. thermophilus* HB8. Expectedly, all the higher β-glycosides (e.g. CELn and LAMn) demonstrated a stronger binding affinity compared to that of short-chain disaccharide β-glycosides such as SOP2, LAM2, CEL2, and GEN2 (Figure 5.6 and Table 5.8). Among these β-glycosides, cello-oligosaccharides (CELn) exhibited the highest binding affinity with the dissociation constant (K_d) of 0.24, 0.13 and 0.08 μM for CEL3, CEL4 and CEL5, respectively (Table 5.8). This observation suggests a positive correlation between the degree of polymerization (DP) of β-glycosides and their binding affinity at physiological temperature. Interestingly, the ITC experiments of short-chain disaccharide β-glycosides performed at two different temperatures (25 and 70°C) did not affect their binding affinities, that corroborates a previous report (Jelesarov and Bosshard, 1999) (Table 5.8 and C.1). The binding isotherms are enthalpically favorable which increases with the temperature increment (Figure 5.6 and Table 5.8).

From this observation, it can be inferred that the increase in temperature modulate either secondary structural elements or the conformational dynamics of the protein βGlyBP. To probe the former effect, circular dichroism (CD) experiment of the protein βGlyBP was performed at different temperatures ranging from 20 to 120°C. The result exhibits that the protein βGlyBP is stable up to 100°C and starts melting afterwards (Figure C.3C). Precisely, overlay of a temperature profile of the protein βGlyBP at 20 and 70°C illustrates that the secondary structures of the protein βGlyBP remains intact even at higher temperature (70°C) (Figure C.3D). This indicates that the latter effect i.e. conformational

dynamics of the protein, instead of the changes in secondary structural elements, favors the binding of the higher β-glycosides to the protein βGlyBP.

Table 5.8. Thermodynamic parameters of β-glycosides binding to the protein βGlyBP at a physiological temperature (70°C) of the bacterium *T. thermophilus* HB8.

Protein (μM)	**Ligand (linkage)**	**Ligand conc. (mM)**	**Stoichiometry ratio (*n*)**	**Association (K_a, M) /Dissociation (K_d, μM) constants**	**ΔH**	**$T\Delta S$**	**ΔG**
					(kcal mol^{-1})		
βGlyBP_WT (333)	SOP2 (β-1,2)	6.66	0.9	1.85x10^5/ 5.40	-7.19	1.07	-8.26
	LAM2 (β-1,3)	3.33	0.7	7.08x10^5/ 1.41	-15.63	-6.44	-9.19
	LAM3 (β-1,3)		0.8	6.25x10^6/ 0.16	-14.17	-3.49	-10.68
	LAM4 (β-1,3)	6.66	1.1	6.44x10^4/ 15.50	-8.66	-1.11	-7.55
	CEL2 (β-1,4)	3.33	0.9	7.35x10^6/ 0.13	-20.20	-9.50	-10.70
	CEL3 (β-1,4)		1.0	4.14x10^6/ 0.24	-13.40	-3.02	-10.38
	CEL4 (β-1,4)		1.3	7.69x10^6/ 0.13	-8.71	2.09	-10.80
	CEL5 (β-1,4)		0.8	1.27x10^7/ 0.08	-16.15	-5.00	-11.15
	GEN2 (β-1,6)	6.66	0.7	5.34x10^3/ 187.00	-6.05	-0.20	-5.85

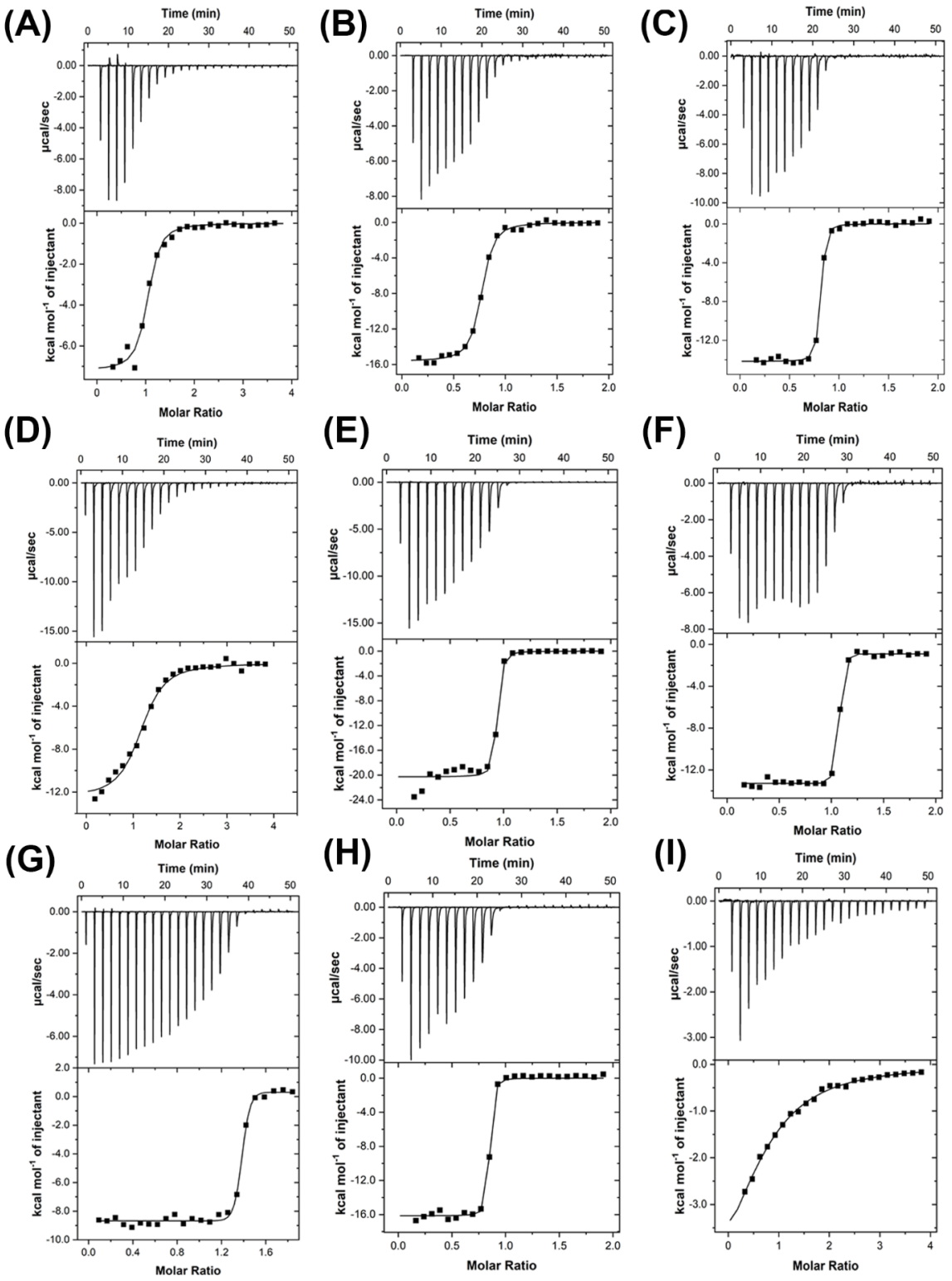


Figure 5.6. Isothermal titration calorimetry of the protein βGlyBP at a physiological temperature (70°C) of the bacterium *T. thermophilus* HB8. The plot displaying the titration of (A) SOP2, (B) LAM2, (C) LAM3, (D) LAM4, (E) CEL2, (F) CEL3, (G) CEL4,

(H) CEL5, and (I) GEN2 with the protein βGlyBP. Upper and lower panels represent raw heat changes and the integrated heat pulse subtracted with the heat of dilution, respectively. In all the binding isotherm, integrated data is plotted against the protein:ligand molar ratio and fitted using a one-site binding model.

5.3.4 Structural basis for the ligand size selection of the protein βGlyBP

To obtain the mechanistic insights into the selection and subsequent transport of various β-glycosides by the protein βGlyBP, it was crystallized with a total of nine β-glycosides identified using the fluorescence and ITC experiments. However, out of these nine cognate ligands, co-crystals could be achieved only with SOP2, CEL2, CEL3 and CEL4. The overall structural topology of the holo βGlyBP (i.e. βGlyBP_WT•SOP2, βGlyBP_WT•CEL2, βGlyBP_WT•CEL3 and βGlyBP_WT•CEL4) remains similar to that of the apo protein (βGlyBP_WT). The bound β-glycosides in the active site of the protein are observed to be sandwiched between NTD and CTD. In all these complex structures, the glycosyl (Glc) units of the β-glycosides are referred to as Glc1, Glc2, …, Glcn (n represents the n^{th} glycosyl unit) where Glc1 is a nonreducing end. Structural analysis of all the complex structures reveal that the β-glycosides bind at the active site of the protein in such a way that its nonreducing end (Glc1) acts as a protein contact point and is buried while other Glc units are exposed to the solvent (Figure C.4). The Glc1 unit coordinates with the residues Thr65 and Thr66 (NTD), Glu117 and Gly297 (hinge) and His181 and Glu376 (CTD) (Figure 5.7 and C.4A). Along with polar interactions, the Glc1 unit is surrounded by three aromatic residues Trp41 and Try67 (NTD) and Trp177 (CTD) forming a hydrophobic cage for nonreducing end (Figure 5.7, C.4B and C.4C). The remaining Glc units interact with the protein βGlyBP mostly through water-mediated hydrogen bonds (Figure 5.7). The presence of a water-mediated network in the vicinity of the active site corresponds well with the thermodynamic data, where an increase in the hydrogen bonding increases the binding affinity of the β-glycosides with a favorable enthalpy and unfavorable entropy change (Table 5.8).

Nevertheless, a higher binding affinity (K_d: 0.13 μM) for CEL2 (β-1,4) compared to that (K_d: 5.4 μM) of SOP2 (β-1,2) remains striking, as it was anticipated to have similar hydrogen bonding network due to their disaccharide nature (Table 5.8 and Figure C.5A).

Comparative structure analysis of both the complexes (βGlyBP_WT•CEL2 and βGlyBP_WT•SOP2) indicates that Glc1 units of CEL2 and SOP2 share a common interaction, thus it can be speculated that the Glc2 unit contributes to this differential binding affinity (Figure C.5B). Crystal structures reveal that the Glc2 unit of CEL2 forms a higher number of water-mediated interactions with the protein βGlyBP than that of the SOP2, that contribute to a higher binding affinity for CEL2 with favorable enthalpy change (Figure C.5C and C.5D).

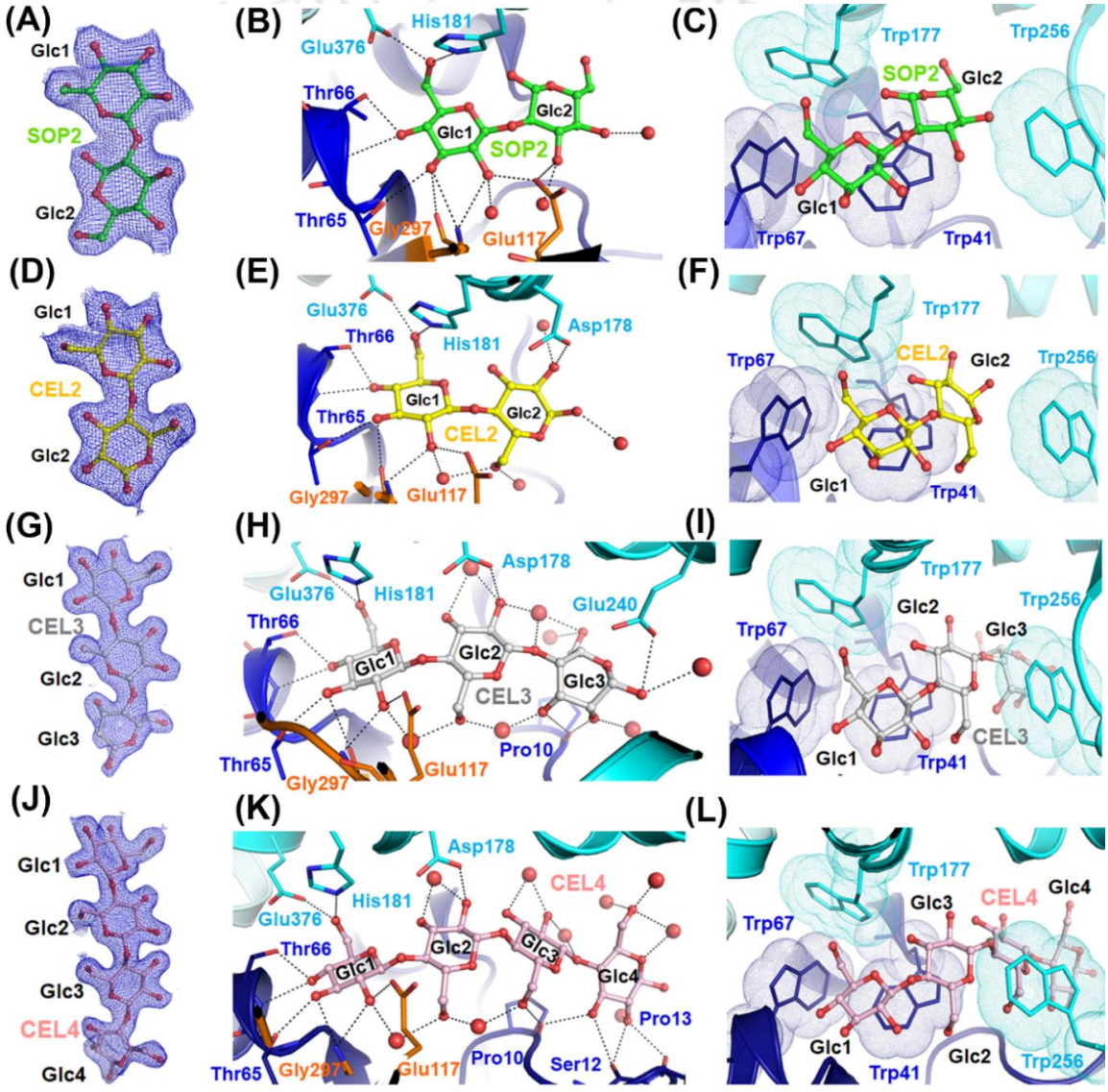


Figure 5.7. Complex structures of the protein βGlyBP with β-glycosides. The protein βGlyBP bound to β-glycosides namely (A-C) SOP2 (green), (D-F) CEL2 (yellow), (G-I)

CEL3 (grey) and (J-L) CEL4 (pink). Each Glc unit of the bound β-glycosides has been labeled as Glcn (n: n^{th} Glc number). (A, D, G, J) The *2Fo-Fc* difference electron density maps of the bound β-glycosides contoured at 1.0σ are displayed in blue mesh. (B, E, H, K) Details of polar interactions between the active site residues of the protein βGlyBP and the β-glycosides. All the interacting residues from NTD (blue), hinge region (orange) and CTD (cyan) are shown as line while hydrogen bonds are represented as dotted lines. (C, F, I, L) The aromatic residues from NTD (blue) and CTD (cyan) involved in the stacking interaction with the β-glycosides are also shown as line while the hydrophobic cage formed for the nonreducing end (Glc1) is represented as dots.

5.3.5 Conserved glycosyl unit of carbohydrate renders initial ligand binding in the subcluster D-I SBPs

The structural and thermodynamic data demonstrate that the protein βGlyBP is a multi-specific and likely involved in the transport of various β-glycosides having different glycosidic linkages such as β-1,2; β-1,3; β-1,4 and β-1,6. In contrast, its homolog protein SO-BP is specific to β-glycosides with β-1,2 linkage. To investigate the mechanistic difference(s) between these two proteins, the complex structures of the protein βGlyBP (βGlyBP_WT•SOP2, βGlyBP_WT•CEL2, βGlyBP_WT•CEL3 and βGlyBP_WT•CEL4) and SO-BP (SO-BP•SOP3, PDB ID: 5YSD; SO-BP•SOP4, PDB ID: 5YSE and SO-BP•SOP5, PDB ID: 5YSF) were compared. Result unveils that in βGlyBP complex structures, only the nonreducing end (Glc1) of all the bound β-glycosides is spatially conserved while in the SO-BP complex structures, Glc2 and Glc3 units are also required to be conserved (Figure 5.8A and 5.8B). Further an in-depth analysis of the active site of these two proteins reveals that in the protein βGlyBP, Glc1 unit of all bound β-glycosides coordinates with the residues from all three active site regions i.e. NTD, hinge and CTD establishing the interaction between NTD and CTD. In contrast, in the protein SO-BP, Glc1 unit of SOP3-5 coordinates with only the NTD residues and thus demands a minimum of three Glc (1-3) units to connect the NTD and CTD as well as to attain the closed conformation of the protein (Figure 5.8A and 5.8B). This observation provides a possible reason as to why the protein SO-BP conserves the spatial position for three Glc (1-3) units and shows no binding preference to the disaccharides (Abe et al., 2018).

Further, to find out whether the other carbohydrate-specific SBPs from the subcluster D-I also possess this feature of spatial conservation, active sites of the subcluster D-I SBPs mentioned in the Table C.2 were investigated. The result establishes that irrespective to the type of carbohydrate bound, the subcluster D-I SBPs conserve the spatial position for at least one Glc unit. However, akin to the proteins βGlyBP and SO-BP, the interactions between the conserved Glc unit and active site regions (NTD, hinge and CTD) vary across the subcluster D-I SBPs (Figure 5.8C). Collectively, these observations suggest that the spatial conservation of the Glc unit is crucial for the establishment of an initial carbohydrate binding in the subcluster D-I SBPs.

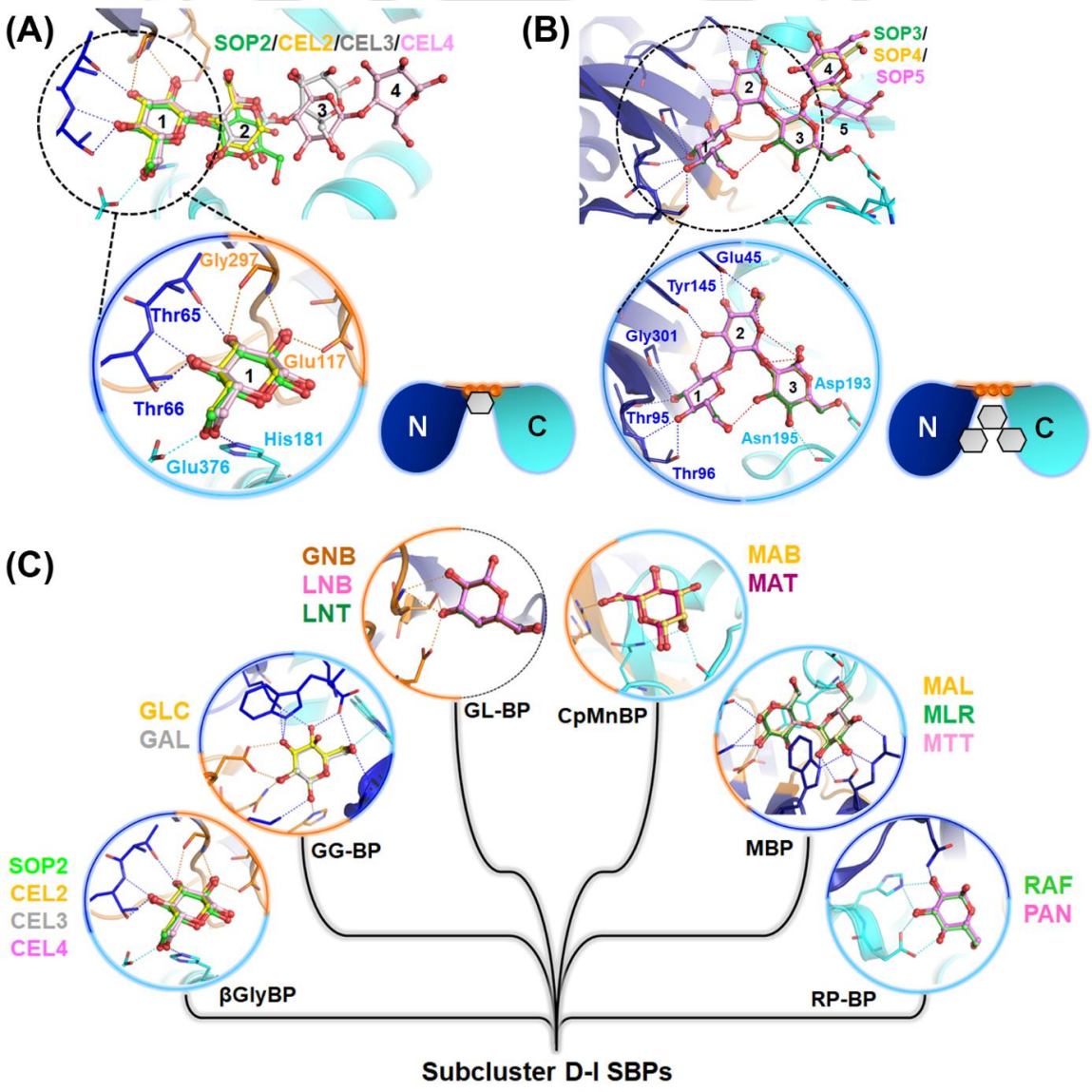


Figure 5.8. Conservation of the glycosyl unit of carbohydrates. (A) Overlay of the protein βGlyBP complex structures bound to SOP2 (green), CEL2 (yellow), CEL3 (grey) and CEL4 (pink). (B) Overlay of the protein SO-BP complex structures bound to SOP3 (green), SOP4 (yellow) and SOP5 (magenta). Each Glc unit (ball-and-stick model) is numbered (1,2,….), considering the nonreducing end as the Glc1 unit. Conserved Glc units are encircled and represented in a close-up view. The amino acid residues hydrogen bonded (dotted-line) with the conserved Glc units are shown in blue (NTD), orange (hinge) and cyan (CTD). (C) Diversity of the interacting regions across the subcluster D-I SBPs. Each circle represents the spatial conservation of Glc unit in the subcluster D-I SBPs for which details are provided in the Table C.2. For the figure clarity, only the conserved Glc unit(s) in each group has(ve) been shown, where active site regions (NTD, hinge and CTD) are represented in blue, orange and cyan, respectively.

5.3.6 Structural determinants distinguishing between the α- and β-glycosides

Strikingly, a structural homology search of the protein βGlyBP using the web server Dali (Holm and Rosenstrom, 2010) yielded subcluster D-I SBPs bound to α-glycosides with high Z-score rather than to β-glycosides (Table C.3). Notably, among these homologs, the maltose-binding protein MalE3 (α-glycoside bound, PDB ID: 6DTQ) is annotated as a "putative β-glucosides ABC transporter, substrate-binding protein" in the UniProtKB database (The UniProt Consortium, 2019), anticipating the shared structural homology between α- and β-glycosides-binding proteins. Since these structural data provide only about their multi-specificity and not about their ligand selection, we sought to perform an in-depth investigation to elucidate the mechanism(s) of glycosidic linkage selection and substrate preferences by the subcluster D-I SBPs. For this, structures of subcluster D-I SBPs α-glycoside-binding protein or αGlyBP (PDB ID: 6J9Y, Chandravanshi et al., 2020) and βGlyBP (PDB ID: 7C66, this study) from the bacterium *T. thermophilus* HB8 were compared as both the proteins strictly maintain their ligand preferences. A structural superimposition of αGlyBP_WT•MAL (α-glycoside bound, α-1,4; PDB ID: 6J9Y) and βGlyBP_WT•CEL2 (β-glycoside bound, β-1,4; PDB ID: 7C66) demonstrates that two structures are almost identical with a rmsd of 0.5 Å (Figure 5.9A). Interestingly, the binding mode of the nonreducing end (Glc1) is also reminiscent of αGlyBP. A comparison of the active site residues of the proteins αGlyBP and βGlyBP coordinating with the respective Glc1 unit of MAL and CEL2 exhibits that the active site regions (NTD, hinge and CTD) are conserved in both the proteins (Figure 5.9B-5.9D). Moreover, the conserved active site residues of the protein αGlyBP/βGlyBP Asp118/Glu117 and Gly286/Gly297 (hinge) and

Asp70/Thr66 (NTD) anchor the Glc1 unit of α- and β-glycosides (Figure 5.9E). These observations designate that the nonreducing end (Glc1) facilitates the initial ligand binding, however, may not govern the selectivity between the α- and β-glycosides.

Thus, to find out the structural factor(s) determining the stereo-chemical selectivity for α- and β-glycosides, the Glc1 unit of α-glycosides (MAL) was manually modeled on to that of the β-glycoside (CEL2) bound to the protein βGlyBP. The result reveals that owing to different glycosidic linkages, Glc2 unit of the α-glycosides alter its position with respect to the Glc2 unit of the β-glycoside and thus causes a steric clash with the residue Trp41 of the protein βGlyBP (Figure 5.9F). Thus, it can be proposed that the residue Trp41 of βGlyBP maintains the stereo-chemical selection and distinguishes between its cognate and non-cognate ligands.

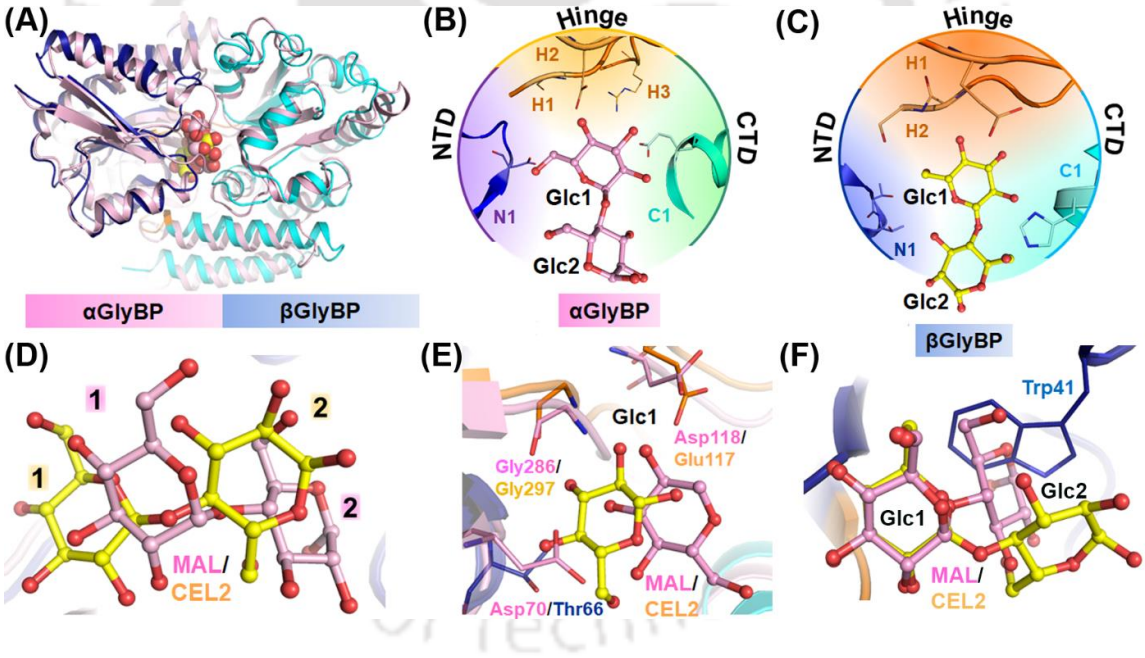


Figure 5.9. Structural determinant(s) for ligand selection. (A) Structural superimposition of the proteins αGlyBP (pink, PDB ID: 6J9Y) and βGlyBP (blue and cyan, PDB ID: 7C66) bound to MAL (pink) and CEL2 (yellow), respectively. The bound ligands are shown as spheres while proteins are as ribbon model. (B and C) Active site comparison of the proteins αGlyBP and βGlyBP, respectively, where the conserved NTD, hinge region and CTD are shown in blue, orange and cyan, respectively. (D) Superimposition of MAL (pink) and CEL2 (yellow) occupying a similar spatial position at the active site with a slight shift in their configuration due to different glycosidic linkages. (E) The comparison of the

active site residues of the proteins αGlyBP and βGlyBP from NTD (blue) and hinge (orange) that interact with the nonreducing end (Glc1) of MAL (pink) and CEL2 (yellow). (F) A close-up view of the active site of the protein βGlyBP with the modeled MAL (pink) in place of CEL2 (yellow). It depicts the steric clash between the Glc2 unit of MAL and the residue Trp41 (blue line).

To further substantiate this proposed ligand selection mechanism, structures of multi-saccharides-binding proteins such as MalE (PDB ID: 6DTU), GacH (PDB ID: 3K00), CpMnBP1 (PDB ID: 4R9G), BlAXBP (PDB ID: 3ZKK) and MBP (PDB ID: 4MBP) from the subcluster D-I SBPs were investigated. Superimposition of the protein βGlyBP with these proteins reveals that irrespective of the ligands type, all carbohydrates occupy a similar spatial position (Figure 5.10A). However, the orientation of β-glycoside bound to the protein βGlyBP is found to be in an opposite direction to that of α-glycoside bound to the protein MalE (Figure 5.10B and 5.10C). Thus, unlike the α-glycosides which occupy subsites A, B, C and D in the active site of MalE, β-glycoside fills four subsites -A, -B, A and B in the active site of protein βGlyBP (Figure 5.10D). To find out the cause of the occupancy of α- and β-glycosides in an opposite orientation, active sites of the two proteins βGlyBP (CEL4 bound, PDB ID: 7C68) and MalE (Maltotetraose or MTT bound, PDB ID: 6DTU) were investigated. Result reveals that the residues from the four secondary structural elements loop1 (L1), helix1 (H1), helix2 (H2) and helix3 (H3) drive the orientation of ligands (Figure 5.10E and 5.10F). A similar feature could also be observed in other subcluster D-I SBPs (Figure C.6).

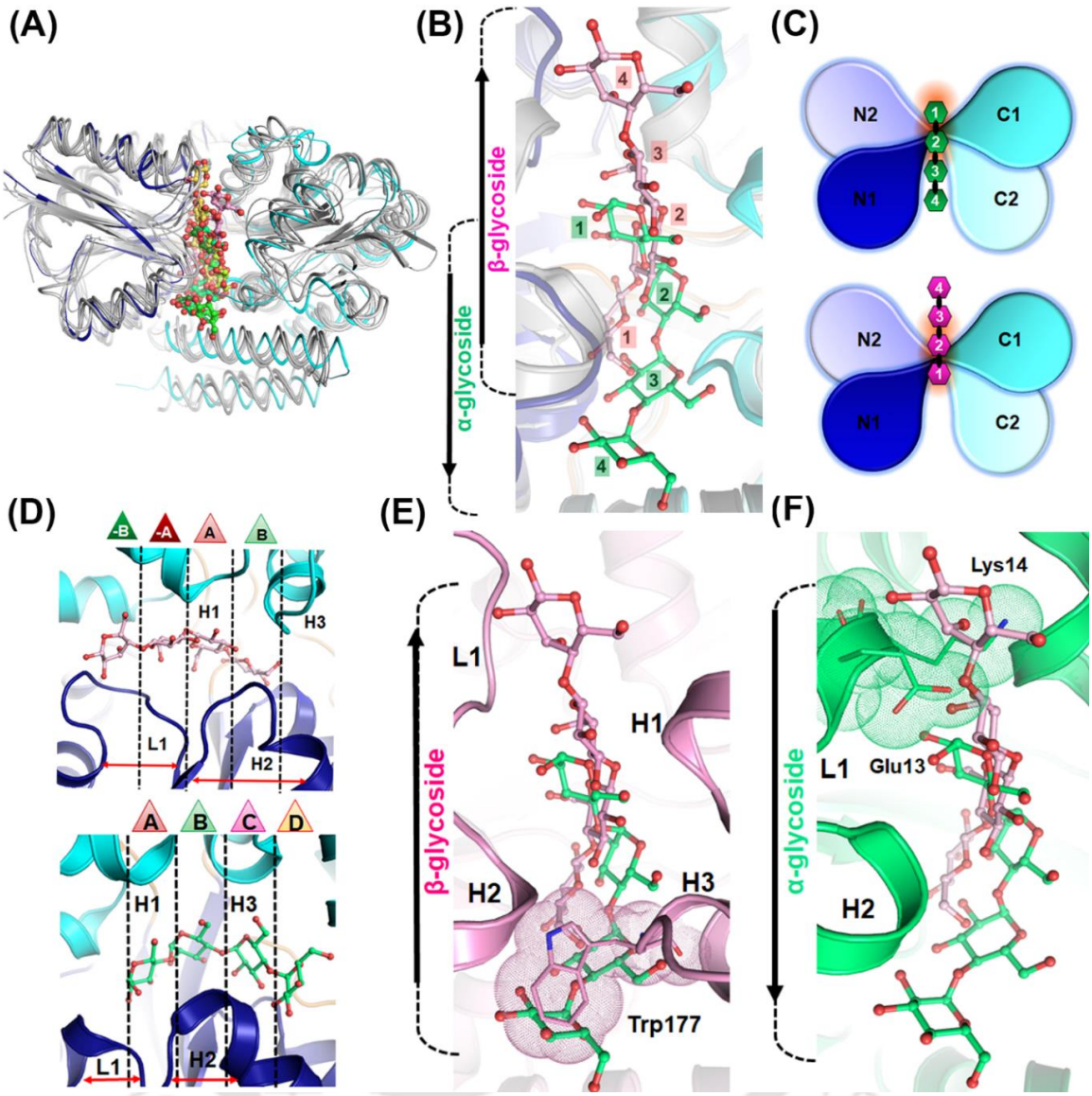


Figure 5.10. Structural determinant(s) guiding the orientations of carbohydrates. (A) Structural superposition of the protein βGlyBP (blue and cyan, PDB ID: 7C68) and the subcluster D-I SBPs MalE (PDB ID: 6DTU), GacH (PDB ID: 3K00), CpMnBP1 (PDB ID: 4R9G), BlAXBP (PDB ID: 3ZKK) and MBP (PDB ID: 4MBP) in grey shows their conserved topology and the ligand binding sites. (B) Comparison of the orientation of the bound β-glycoside (CEL4, pink) to the protein βGlyBP and α-glycoside (MTT, green) to the protein MalE. Each Glc unit is numbered sequentially (1,2,....), considering the nonreducing end as the 1st Glc unit. (C) A schematic representation of α- and β-glycoside binding in an upward (top) and downward direction (bottom), respectively. Four subdomains of the proteins are labeled as N1, N2, C1 and C2 and the Glc units are numbered sequentially (1 to 4). (D) The four subsites (-A, -B, A and B) of βGlyBP (top) bound to a β-glycoside (CEL4, pink) and that (A, B, C and D) of αGlyBP (bottom) bound to an α-glycoside (MTT, green) are partitioned by a dotted vertical line. (E and F) Structural

determinants loop (L1) and helices (H1-H3) guiding the orientation of β-glycoside (CEL4, pink) and α-glycoside (MTT, green), respectively, are shown here. The amino acid residues from the structural elements (L1 and H1-H3) and carbohydrates are represented as dots and ball-and-stick model, respectively. The arrow denotes the directionality of the orientation of the bound ligand.

5.3.7 Two-step ligand-binding mechanism of the protein βGlyBP

Based on the structural insights, we postulate that the ligand binding is primarily driven by an initial interaction between the conserved nonreducing end (Glc1) and the active site hydrophobic cage of the protein βGlyBP. However, the question whether the hydrophobic cage is pre- or post-formed after the rearrangement of three aromatic residues Trp41, Trp67 and Trp177 upon the nonreducing end (Glc1) interaction, remains unanswered. To delineate the role of these aromatic residues Trp41, Trp67 and Trp177 in hydrophobic cage formation and that of the residue Trp256 present in the vicinity were mutated using site-directed mutagenesis approach. Although the residues Trp41, Trp67 and Trp256 could be correctly mutated to alanine, that of the residue Trp177 introduced random mutations as K174R, N175T, S176P, ΔW177, D178R and V179T (hereafter, the mutant is referred to as βGlyBP_W177X). However, to our delight, crystals of the mutant βGlyBP_W177X incubated with nine cognate β-glycosides mentioned in the Table 5.8, appeared within a month compared to that (nine months) of the wild type protein. Thermodynamic analysis of the mutant βGlyBP_W177X using the ITC experiments revealed that in comparison to the protein βGlyBP_WT, the mutant βGlyBP_W177X shows a reduced binding affinity (~24-206 fold) for all cognate β-glycosides and no binding for CEL2 and CEL5 (Table C.1). This observation suggests the role of the residue Trp177 in hydrophobic cage formation for the β-glycosides binding.

Further, the structural data of the mutant βGlyBP_W177X reveals that the binding of disaccharide β-glycosides (SOP2, LAM2 and GEN2) fails to transit the conformation from an open to closed state and thus is trapped in an intermediate transition state (referred to as open-liganded state). In this state, ligands are hydrogen bonded with the active site residues Thr65 and Thr66 (NTD) and Glu117 and Gly296 (hinge region) only (Figure 5.11A-5.11C). This suggests that the initial interaction of β-glycosides with the protein βGlyBP

is independent of the amino acid residues from CTD. This structural feature supports the thermodynamic data, where a decrease in the binding affinity of the mutant βGlyBP_W177X for disaccharide β-glycosides compared to the wild type protein (βGlyBP_WT) are associated with the loss of the domain closure as well as interaction with CTD. Notably, the intermediate transition (open-liganded) state of βGlyBP_W177X is structurally similar to that of βGlyBP_WT_FormII, where the residues Trp256 and Trp177 are in an inward orientation (Figure 5.11A-5.11C). This designates that the structure βGlyBP_WT_FormII is a pre-formed state (referred to as partial-open-unliganded state) of the intermediate transition (open-liganded) sate worthy to bind β-glycosides. Thermodynamically, the mutant βGlyBP_W177X displays a higher binding affinity for the disaccharides β-glycoside with β-1,2 linkage (SOP2) compared to those having β-1,3 (LAM2) and β-1,6 (GEN2) linkages. Structural data of βGlyBP_W177X•SOP2 suggest that the higher binding affinity for SOP2 is due to a large number of water-mediated interactions with the mutant βGlyBP_W177X (Figure 5.11A).

This four different trapped conformational (open-unliganded, partial-open-unliganded, open-liganded and closed-liganded) states of the protein βGlyBP earmarked us to conclude that the ligand binding to the protein βGlyBP is a multi-step process, at least in two steps. In the first step, the nonreducing end (Glc1) of the ligand binds at the active site NTD and hinge regions and remains in the open-liganded state. In the second step, this initial interaction initiates the hydrophobic cage formation and lock the Glc1 unit and thus transits the protein to the closed-liganded state by forming additional interactions with the CTD (Figure C.7).

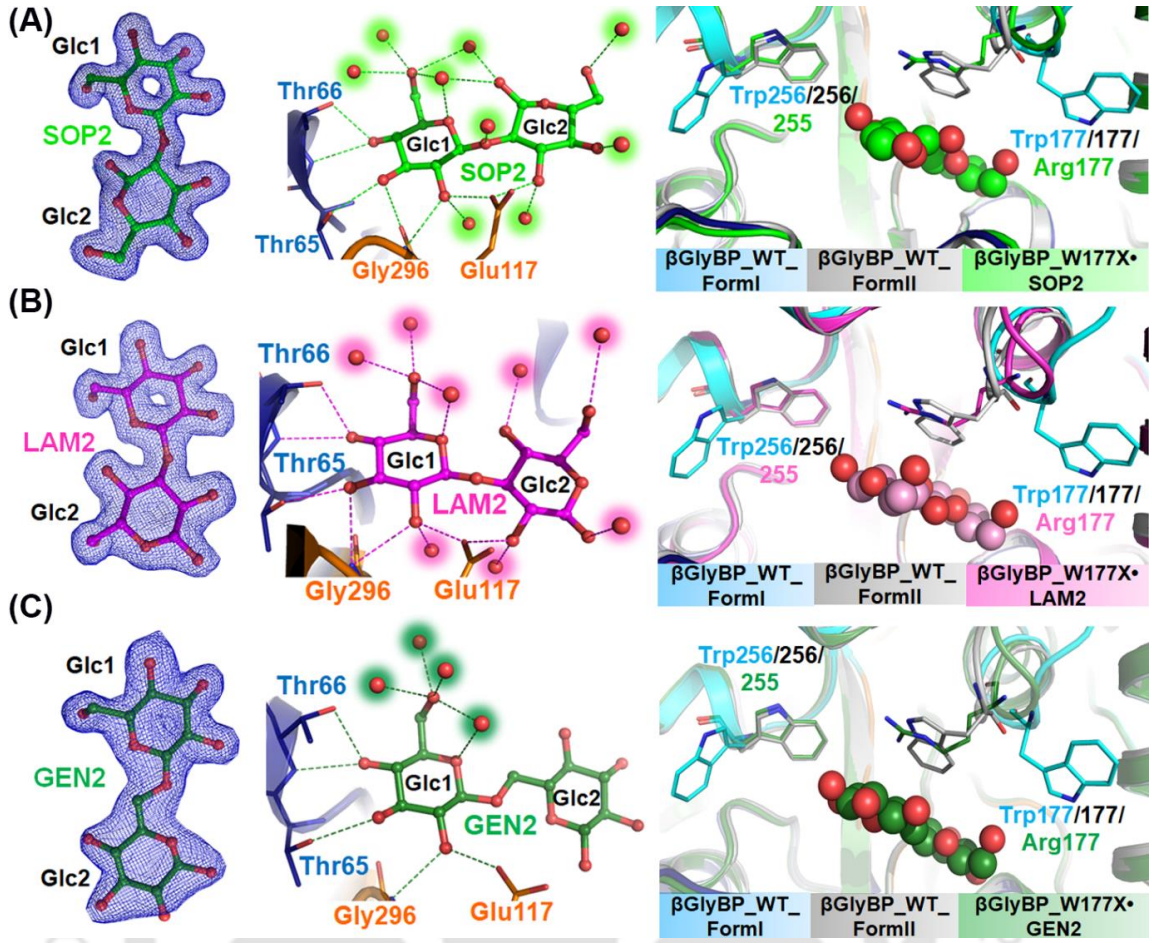


Figure 5.11. The complex structures of the mutant βGlyBP_W177X mimicking an intermediate transition state. (A-C) The mutant βGlyBP_W177X complex structures bound to SOP2, LAM2 and GEN2, respectively. (Left), the *2Fo − Fc* difference electron density map for SOP2 (light green), LAM2 (magenta) and GEN2 (green) contoured at 1.0σ are shown in blue mesh. (Center), the active site residues interacting with SOP2, LAM2 and GEN2 via hydrogen bonds (dotted line) from NTD and hinge regions are colored blue and orange, respectively. The water molecules are represented as red spheres. (Right), overlay of the unliganded structures of βGlyBP_WT_FormI (cyan, PDB ID: 7C63), βGlyBP_WT_FormII (grey, PDB ID: 7C64) with open-liganded structures of (A) βGlyBP_W177X•SOP2 (light green, PDB ID: 7C6I), (B) βGlyBP_W177X•LAM2 (magenta, PDB ID: 7C6H) and (C) βGlyBP_W177X•GEN2 (green, PDB ID: 7C6G) complexes. The inward orientation of the residues Arg177 (random mutation) and Trp255 of the mutant βGlyBP_W177X complex structure is similar to that of βGlyBP_WT_FormII. The amino acid residues and the ligands are represented as ball-and-stick and sphere models, respectively.

Akin to W177X, effect of W41A mutation was also investigated, which reveals that the mutation of Trp41 to alanine (βGlyBP_W41A) reduces the binding affinity of β-glycosides by almost ~436-1034 fold compared to that of the protein βGlyBP_WT (Table C.1). Strikingly, the thermodynamic data of the mutant βGlyBP_W41A with all the nine cognate β-glycosides except LAM4 exhibits an endothermic reaction (Figure C.8). This reduced binding affinity and an endothermic reaction of the mutant βGlyBP_W41A indicate that unlike the mutant βGlyBP_W177X, it might not attain the intermediate transition state upon the β-glycosides binding. To substantiate this, the βGlyBP_W41A was crystallized in the absence and presence of all the nine cognate β-glycosides. The mutant protein βGlyBP_W41A crystallized in two forms (I and II) and are in an open-unliganded state. This observation indicates that unlike the mutant βGlyBP_W177X, binding of the β-glycosides to the mutant βGlyBP_W41A_FormII is transient and thus, the mutant fails to attain the intermediate transition (open-liganded) state. However, the conformation of the residue Trp177 of the mutant βGlyBP_W41A_FormII is observed to be similar to that of the closed state of the protein βGlyBP_WT mimicking a closure dynamic of the CTD due to the co-crystallization effect. Nonetheless, the required shift (4.1 Å) of the residue Trp177 from CTD is not enough to form the hydrophobic cage (Figure 5.12A-5.12C).

Similarly, the thermodynamic data of another mutant βGlyBP_W67A show a reduced binding affinity for all nine cognate β-glycosides, which are similar to that of the mutant βGlyBP_W177X (Figure C.9 and Table C.1). The structural data of the mutant βGlyBP_W67A bound to GEN2 (βGlyBP_W67A•GEN2) reveal that the ligand is sandwiched between the aromatic residues Trp41 and Trp67 (Figure 5.12D). Although it can bind to the ligand, it fails to attain the closed state and thus locked in the intermediate transition (open-liganded) state. Similarly, the thermodynamic data of the mutant βGlyBP_W256A also show its reduced binding affinity for β-glycosides (Figure C.10 and Table C.1). Unfortunately, crystals of the mutant βGlyBP_W256A could not be achieved which restricted to substantiate its thermodynamic data.

Altogether, the mutagenesis experiments of the protein βGlyBP suggests that out of three aromatic residues of hydrophobic cage, residue Trp41 play a major role compared to the

residues Trp177 and Trp67. Interestingly the residue Trp256 present in the vicinity is also found to be equally crucial for β-glycosides as those residues involved in hydrophobic cage formation. Moreover, the relative binding kinetics of these four mutants W256A > W67A > W177X > W41A corroborates the proposed two-step ligand-binding mechanism, where the first step is triggered by the initial hydrophobic cage formation with the residue Trp41 for nonreducing end (Glc1).

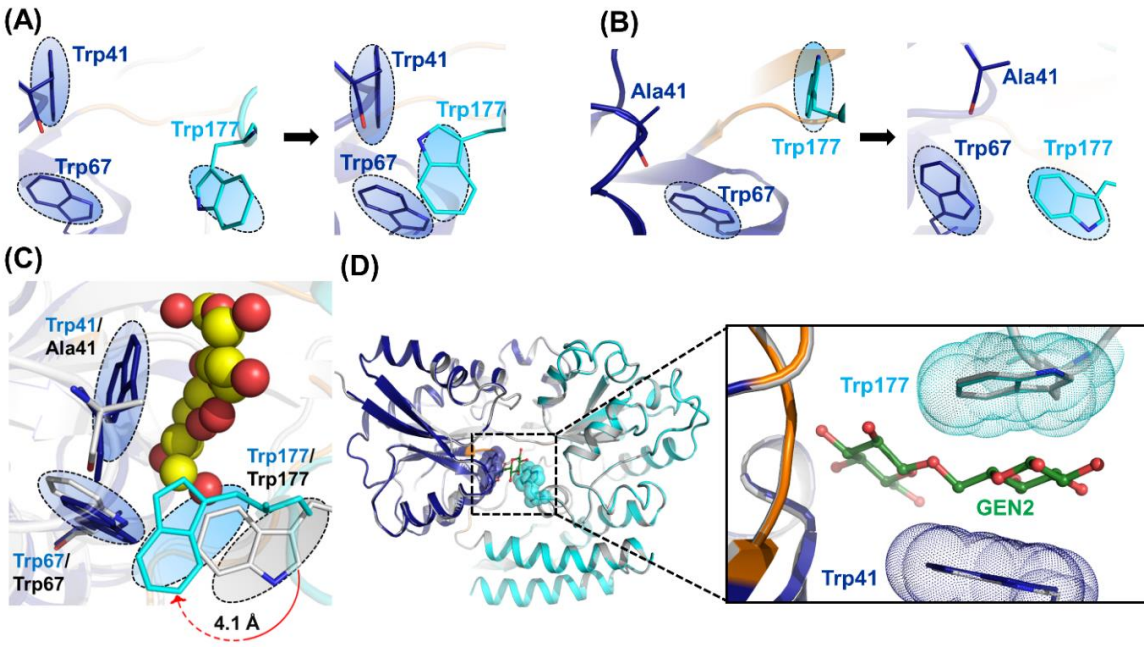


Figure 5.12. Active site comparison of the wild type protein βGlyBP_WT with the mutants βGlyBP_W41A and βGlyBP_W67A. Orientation and position of three aromatic residues Trp41, Trp67 and Trp177 of (A) βGlyBP_WT_FormI in open-unliganded (left) and βGlyBP_WT•CEL2 closed-liganded (right) states, (B) βGlyBP_W41A_FormI in open (left) and βGlyBP_W41A_FormII in partial-open (right) conformations. (C) Superimposition of the active sites of βGlyBP_WT•CEL2 (blue and cyan) and βGlyBP_W41A_FormII (grey). The bound CEL2 is represented as yellow spheres and the distance required to cover by the residue Trp177 to form the aromatic hydrophobic cage is represented by a curved red arrow. (D) Structural superimposition (left) and close-up view (right) of βGlyBP_WT_FormI (blue and cyan, PDB ID: 7C63) and βGlyBP_W67A•GEN2 (grey, PDB ID: 7C70) structures in open conformation. The aromatic residues Trp41 and Trp177, which stack GEN2 (green) are shown as dots. In all the figures, the residues from NTD and CTD are shown in blue and cyan, respectively.

5.3.8 The C2 subdomain holds the N1 and C1 subdomains

The structural data of the mutant βGlyBP_W177X complexed with CEL2 contradicts the thermodynamic data which do not show its binding. Moreover, unlike the open-liganded structures of mutant βGlyBP_W177X with disaccharide β-glycosides (SOP2, LAM2 and GEN2), the mutant βGlyBP_W177X•CEL2 complex structure attains the closed-liganded state. To get insight into this differential behavior, thermodynamic and structural data of CELn (CEL3-5) binding to the mutant βGlyBP_W177X were compared. According to the thermodynamic data, the binding of CELn with the mutant βGlyBP_W177X is an endothermic reaction as a result of a positive entropy change (Figure C.11 and C.12A); this feature has previously been reported for the CELn binding with the carbohydrate binding module (CBM) (Georgelis et al., 2012). Investigation of the CELn-bound structures of βGlyBP_W177X shows that the protein transits from an open to closed state and forms an increased number of water-mediated hydrogen bonds with an increment in the Glc units (Figure C.12B-C.12E). Notably, akin to CEL2, although the binding of CEL5 to βGlyBP_W177X leads to transition in closed state, the thermodynamic data shows negligible heat change (Figure C.11). This thermodynamic loss in the binding affinity of CEL2 and CEL5 can be explained by enthalpy-entropy compensation property of a protein complex, where the enthalpy and entropy change equate to each other resulting in a net zero heat change.

To further investigate the cause for the complete enthalpy-entropy compensation, apo and holo structures of the mutant βGlyBP_W177X were analyzed. In the apo βGlyBP_W177X structure, the C2 subdomain strongly holds the N1 and C1 subdomains through water-mediated hydrogen bonds with the residues Thr66 and Leu179, respectively. Thus, it controls the conformational flexibility of the N1 and C1 subdomains. Interestingly, in the holo βGlyBP_W177X structure, the ligand binding (CEL3, CEL4, LAM3 and LAM4 except for CEL2 and CEL5), followed by the domain closure maintains these inter-domain interactions and conserves water (Cw) molecules (Figure 5.13 and Table C.4). Moreover, the complex structures of the mutant proteins (βGlyBP_W177X•SOP2, βGlyBP_W177X•LAM2 and βGlyBP_W177X•GEN2), which impairs the domain closure in case of bound SOP2, GEN2 and LAM2, also retains these inter-domain interactions

(Figure C.13 and Table C.4). Based on these observations, it can be speculated that the water-mediated conformational flexibility of the subdomains and CEL2/CEL5 binding might contribute to the complete enthalpy-entropy compensation considering that the ITC conditions are far different from the crystallization condition (Figure 5.13 and Table C.4).

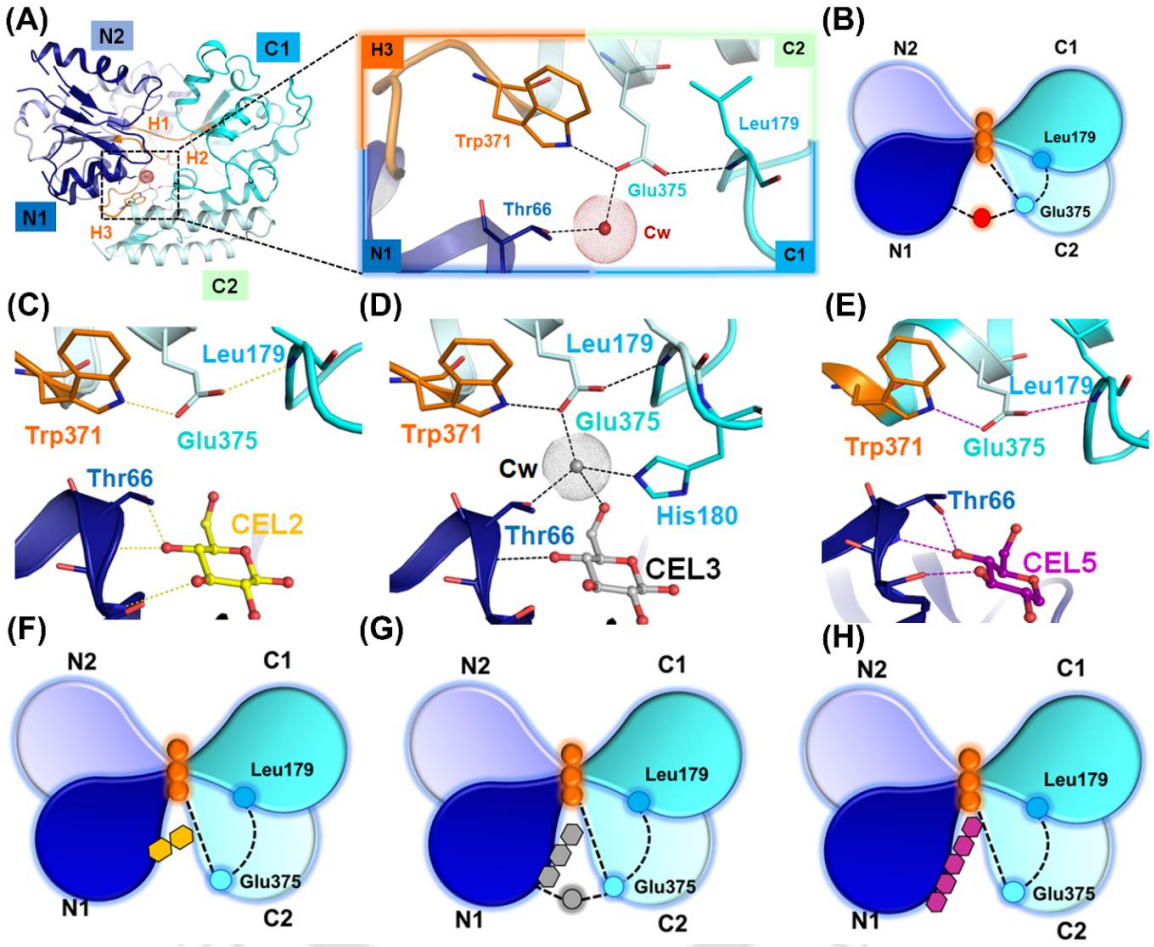


Figure 5.13. Inter-domain interactions maintaining the conformations of the N1, C1 and C2 subdomains. (A) The overall structure (left) and a close-up view (right) of the apo βGlyBP_W177X depicting the polar interactions (dotted lines), which hold the hinge region and the N1, C1 and C2 subdomains. A conserved water (Cw) molecule is represented by a dotted sphere. (B) A schematic representation of the apo βGlyBP_W177X structure representing the inter-domain interactions (dotted line). The amino acid residues, hinge region and water molecule are depicted in blue (cyan), orange and red spheres, respectively. (C-E) (top) Molecular details of the interactions among the N1 (blue), C1 (light cyan), C2 (cyan) subdomains in βGlyBP_W177X•CEL2, βGlyBP_W177X•CEL3 and βGlyBP_W177X•CEL5 complexes, respectively, where hydrogen bonds and the conserved water (Cw) molecules are shown as dotted lines and sphere, respectively; (F-H)

(bottom) schematic representations of the βGlyBP_W177X•CEL2, βGlyBP_W177X•CEL3 and βGlyBP_W177X•CEL5 complex structures, respectively, depicting the loss of inter-domain interaction between the N1 and C2 subdomains in the CEL2 (yellow) and CEL5 (purple) bound structures, where domains and residues are labeled and colored coded as N1 (blue), N2 (light blue), C1 (light cyan) and C2 (cyan).

5.3.9 Second-step dynamics correlate with the differential thermodynamic behavior

Analysis of thermodynamics of LAMn binding to the mutant βGlyBP_W177X shows that in contrast to CELn, the binding profile of LAMn are exothermic, where LAM2 and LAM4 are enthalpically favorable while binding of LAM3 is majorly driven by a positive entropy change (Figure C.11B-C.11D, Figure C.14A and Table C.1). Interestingly, the structural data of βGlyBP_W177X•LAMn complexes also support the thermodynamic data. Unlike LAM2, binding of LAM3 and LAM4 brings the domain closure leading to the closed state with a significant increase in the number of hydrogen bonds with an increment in the Glc units (Figure C.14B-C.14D). Using these structural and thermodynamic data of βGlyBP_W177X•CELn and βGlyBP_W177X•LAMn complexes, it can be speculated that in the first step of the ligand-binding mechanism, both the ligands CELn and LAMn preserve their binding mode through the nonreducing end (Glc1). However, in the second step, the conformational dynamics of domains lead to the differential thermodynamic behavior for CELn (endothermic) and LAMn (exothermic) binding. Moreover, the mutant βGlyBP_W177X complexed with CEL3 and CEL4 crystallizes in two forms (βGlyBP_W177X•CEL3_FormI and _FormII; βGlyBP_W177X•CEL4_FormI and _FormII) both belonging the space group $P2_12_12_1$. In the first form (βGlyBP_W177X•CEL3_FormI and βGlyBP_W177X•CEL4_FormI), it is similar to βGlyBP_WT_FormI with one molecule in the asymmetric unit while in the second form (βGlyBP_W177X•CEL3_FormII and βGlyBP_W177X•CEL4_FormII), two molecules are present in the asymmetric unit with a different crystal packing. Two molecules, in the second form, are arranged in such a way that NTD of one protomer is juxtaposed to that of NTD of the other protomer (Figure C.15A and C.15B). Interestingly, the protein βGlyBP_W177X complexed with LAM3 is crystallized in two form (βGlyBP_W177X•LAM3_FormI and _FormII) with different space group. The first βGlyBP_W177X•LAM3_FormI is crystallized in space group $P2_12_12_1$, while second βGlyBP_W177X•LAM3_FormII is crystallized in a trigonal space group $P3_121$ containing

three molecules in the asymmetric unit, which are associated by a two-fold non-crystallographic symmetry (Figure C.15C).

5.3.10 Conformational dynamics of the N1 and C1 subdomains anchor domain closure

A wealth of open and closed structures of SBPs available in the literature demonstrate that the domain closure occurs through a conformational change of the NTD. Interestingly, an analysis of the conformational dynamics of the protein βGlyBP using the web server DymDom (Lee et al., 2003) reveals that both the N1 and C1 subdomains undergo conformational changes upon the cognate ligand binding. The N1 subdomain shows a greater degree of conformational change (a rotation of 17.8° and a rigid-body translation of -1.0 Å) than the C1 subdomain (a rotation of 11.28° and a rigid-body translation of -0.2 Å). A similar analysis of other subcluster D-I SBPs unveils that irrespective of the ligand types, subcluster D-I SBPs viz. SO-BP (apo/SOP5-bound, PDB ID: 5YSB/5YSD), αGlyBP (apo/TRE-bound, PDB ID: 6JAL/6JAM) GG-BP (apo/GLC-bound, PDB ID: 3OO7/3OO6), GacH (apo/MTT-bound, PDB ID: 3K01/3K00) and MalE (apo/maltopentaose-bound, PDB ID: 2XD2/2XD3) demonstrate only a single domain (NTD) movement with a varying degree of rotation, 25.4, 41.9, 35.9, 29.5 and 26.5°, respectively (Figure 5.14A-5.14F). Based on these observations, it can be proposed that the information for the domain movement is encoded in the protein architecture and is independent of the type of the bound ligand (carbohydrate). Thus, to decode this information, both the apo and holo forms of the protein βGlyBP was investigated. Results show that in the apo form, the residues Glu117 and Arg119 (hinge 1) interacts with the N1 and C1 subdomains through hydrogen bond network. Upon initial interaction of β-glycosides with the residues forming the hydrophobic cage and hinge region, the residues Glu117 (hinge 1) and Arg119 (hinge 2) pull the N1 and C1 subdomains closer for the ligand encapsulation (Figure 5.14G).

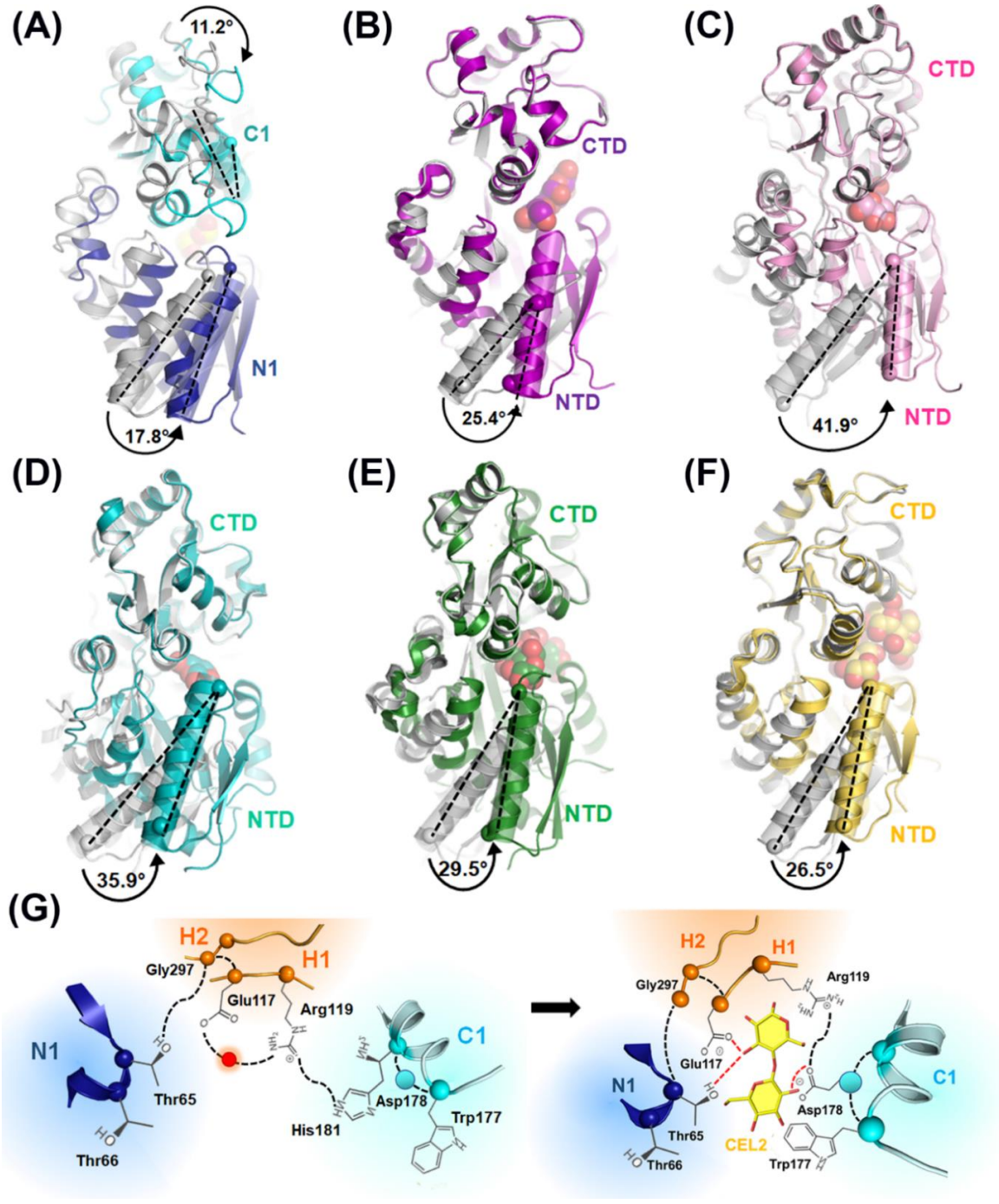


Figure 5.14. Conformation dynamics of domains upon the ligand binding. Superimposition of the open structures (grey) of (A) βGlyBP (PDB ID: 7C63), (B) SO-BP (PDB ID: 5YSB), (C) αGlyBP (PDB ID: 6JAL), (D) GG-BP (PDB ID: 3OO7), (E) GacH (PDB ID: 3K01) and (F) MalE (PDB ID: 2XD2) with the closed structures of βGlyBP (blue and cyan, PDB ID: 7C66), SO-BP (violet, PDB ID: 5YSD), αGlyBP (pink, PDB ID: 6JAM), GG-BP (cyan, PDB ID: 3OO6), GacH (green, PDB ID: 3K00) and MalE (yellow, PDB ID: 2XD3), respectively. The bound ligands are color-coded according to their

respective protein and are represented as spheres. The movement of the domain NTD as well as the N1 and C1 subdomains are depicted by a conformational change of a helix represented as cylinder with a rotation angle. (G) Atomic details of the open-to-closed state transition of the protein βGlyBP. A systematic change in interactions (dotted line) during open (left) to closed (right) state transition has been represented. Secondary structural elements of the N1 subdomain (blue), hinge region (orange) and C1 subdomain (cyan) and the active site residues are represented in 3D and 2D models, respectively. Water molecule is depicted as a red circle. In the closed state (right), new interactions formed between ligand (CEL2, yellow) and the amino acid residues are represented with red dotted lines. For the clarity of the figure, only one ligand CEL2 has been shown here.

5.3.11 The transport and metabolism of β-glycosides inside the cell

The ability of the protein βGlyBP to bind a variety of β-glycosides as compared to its homolog SO-BP inspired to understand its physiological basis. For this, an analysis of the genomic content of the bacterium *T. thermophilus* HB8 was investigated which reveals that genes (ORF IDs: TTHB082-TTHB086) encoding the β-glycosides ABC transporter is organized as a cluster followed by the gene (ORF ID: TTHB087) coding for the enzyme β-glucosidase from the glycosyl hydrolase family 1 (GH1). Since, the proteins βGlyBP and SO-BP belong to the subcluster D-I of SBPs, β-glycosides-specific subcluster C-IV SBPs were also considered for the genetic organization analysis to find out the reason of the β-glycosides transport through two different subclusters. For these genetic architectures of β-glycosides-specific subcluster D-I SBPs e.g. βGlyBP (ORF ID: TTHB082) and SO-BP (ORF ID: Lin1841) were compared with the subcluster C-IV SBPs from *Xanthomonas axonopodis* pv. citri (*Xc*_SBP), *Pseudarthrobacter chlorophenolicus* (*Pc*_SBP), CEL2-binding protein from *Thermotoga maritima* (*Tm*_CBP) and chitin oligasaccharide-binding protein from *Vibrio cholerae serotype O1* (*Vc*_CBP).

A search for the genes involved in the β-glycosides uptake and metabolism yields different genetic architectures for the subcluster C-IV and D-I SBPs: (1) the β-glycosides ABC transporter operon with the β-glucosidase from GH1, (2) the β-glycosides ABC transporter operon with the β-glucosidase from glycosyl hydrolase family 3 (GH3), (3) the β-glycosides ABC transporter operon with the endoglucanase from glycosyl hydrolase family 9 (GH9) and (4) the β-glycosides ABC transporter operon alone. Interestingly, SBPs from the subcluster D-I belong to the 1st and 4th operonic architectures while those of the

subcluster C-IV belong to the 2nd and 3rd types (Figure 5.15A, 5.15B and C.16). The co-occurrence of the glycosyl hydrolase (GH) enzymes with the subcluster C-IV and D-I SBPs provides an answer as to why the uptake of the β-glycosides is executed by both the subclusters C-IV and D-I SBPs. Furthermore, the enzymes GH1 and GH3 β-glucosidases are intracellularly located and thus can mediate the metabolism of β-glycosides imported through functionally-associated ABC transporter (Figure 5.15C and 5.15D).

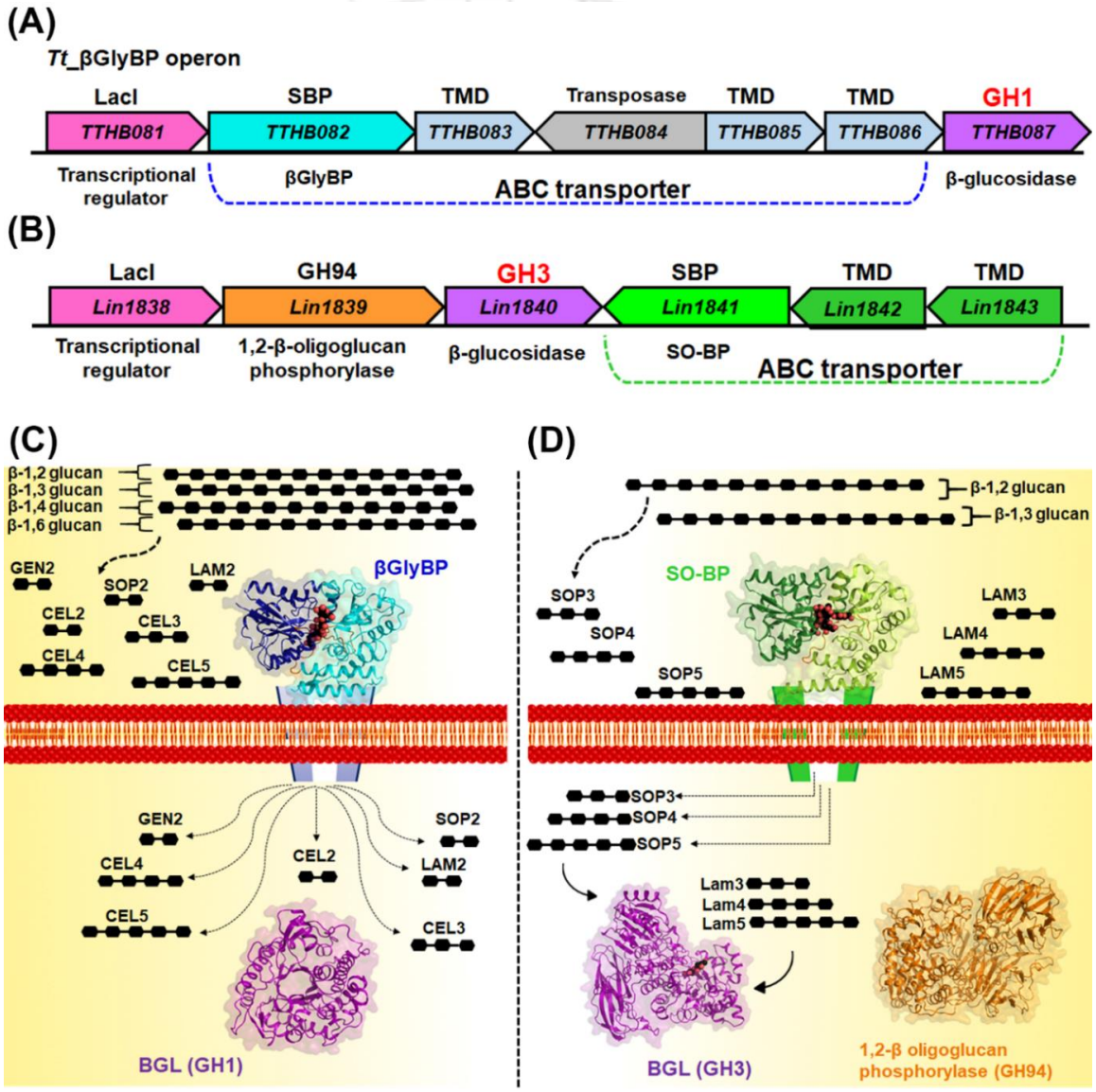


Figure 5.15. Transport and metabolism of β-glycosides. The genetic organization of the β-glycosides ABC transporter along with metabolizing enzymes in (A) *T. thermophilus* HB8 (*Tt*_ βGlyBP) and (B *Listeria innocua* (*Li*_SO-BP. Each gene is represented by an

ORF number with an arrow indicating its transcriptional direction. An integrative proposed model for the β-glycosides transport and metabolism pathway in (C) *T. thermophilus* HB8 and (D *L. innocua*. This includes β-glycoside poly- and oligo-saccharides (black), βGlyBP (blue and cyan) and SO-BP (green and light green) ABC transport systems, β-glucosidase (BGL, violet) from GH3 and GH1 family and 1,2-β-oligoglucan phosphorylase (orange) from GH94 family.

5.4 DISCUSSION

Substrate (solute)-binding proteins (SBPs) specific to ABC importers capture their substrates via a well-known "Venus Fly-trap" mechanism by undergoing a large conformational transition from an open to closed state (Mao et al., 1982). Based on a wealth of structural information, SBPs are known to share a common architecture (Berntsson et al., 2010). Despite structural conservation, SBPs are known to adopt different ligand-binding mechanisms (Begg et al., 2015; Pandey et al., 2016; Chandravanshi et al., 2020). However, a basic question pertaining to SBPs is how do they distinguish among different (dis)similar ligands and maintain their stereo-selectivity between α- and β-glycosides, even though having a conserved active site architecture. This prompted us to investigate the relationship between a ligand selection mechanism and conformational dynamics of SBPs triggered upon ligand binding. This study presents the structural information of a βGlyBP (ORF ID: TTHB082), an SBP subunit of ABC transport system (ORF IDs: TTHB082-TTHB086) from the bacterium *T. thermophilus* HB8. Topologically, the protein βGlyBP possesses four non-sequential subdomain organization N1-C1-N2-C2 as previously reported for the proteins Smon0123 (PDB ID: 5XS8) and αGlyBP (PDB ID: 6J9W) from the subcluster D-I SBPs (Oiki et al., 2017; Chandravanshi et al., 2020).

This study provides the first structure of the protein βGlyBP complexed with various β-glycosides viz. sophorose (SOP2, β-1,2), laminaribiose (LAM2, β-1,3), laminaritriose (LAM3, β-1,3), laminaritetraose (LAM4, β-1,3), cellobiose (CEL2, β-1,4), cellotriose (CEL3, β-1,4), cellotetraose (CEL4, β-1,4), cellopentaose (CEL5, β-1,4) and gentiobiose (GEN2, β-1,6) having different types of glycosidic linkages. These structures reveal that the binding of β-glycosides majorly occurs through the nonreducing end (Glc1) by forming hydrophobic stacking interactions using three aromatic residues Trp41, Trp67 and Trp177, as reflected in the fluorescence quenching experiments. The quenching phenomenon

suggest a strong dehydration of the indole ring of tryptophan residues upon β-glycosides binding as previously suggested for the protein cellulase Cel7A upon carbohydrate binding (Røjel et al., 2019). The thermodynamic data demonstrate the binding of higher β-glycoside(s) at physiological temperature (70°C) of the bacterium *T. thermophilus* HB8. A higher temperature causes conformational changes in protein leading to increased binding affinity by releasing bulk water molecules and exposure of hydrophobic residues (Chen et al., 2003). This fact supports that at a higher temperature, the active site aromatic residues change their conformation to accommodate higher β-glycoside(s) without perturbing the secondary structure of the protein βGlyBP as demonstrated using the circular dichroism (CD) data. This observation is visible in the βGlyBP_WT_FormII structure, where two active site aromatic residues Trp177 and Trp256 reorient their conformation due to the dehydrating effect of the polyethylene glycol (PEG) molecules present at the active site of the protein βGlyBP.

This study not only identifies a range of ligands which can bind the protein βGlyBP, but also uncovers the ligand selection mechanism for the subcluster D-I SBPs. Moreover, the study provides the structural basis for the exclusion of non-cognate ligands from binding to the protein. Interestingly, due to a conserved topological similarity of the subcluster D-I SBPs, they exhibit a conserved pattern in the interacting regions i.e. NTD, hinge and CTD for both type of carbohydrates i.e. α- and β-glycosides. Furthermore, since these SBPs initiate the ligand binding through the nonreducing end (Glc1), they can initially bind both cognate and non-cognate ligands at the active site without any distinction. These findings corroborate an earlier unexplained mechanism of the non-cognate ligand binding to SBPs (de Boer et al., 2019b). However, binding of the non-cognate ligands fail to trigger the conformational changes in the protein as previously reported for the arginine and lysine residues of the subdomains SBD1 and SBD2 of the protein GlnQ (Gouridis et al., 2015). This fact supports our proposed mechanism of ligand stereo-selection, where closed conformation of the protein βGlyBP causes a stearic clash between the residue Trp41 and the Glc2 unit of the MAL (a non-cognate α-glycoside) modeled in the place of CEL2 (a cognate β-glycoside). This inference suggests that ligand binding is a two-step process,

where in the first step, ligand binding is initiated in a nonselective manner while the selection between the cognate and non-cognate ligands is assured in second step.

Akin to the protein βGlyBP, results show that the subcluster D-I SBPs also follow this ligand selection mechanism. Differences in their active site is correlated with the ligand size, which is anchored by five secondary structural elements, two loops (L1 and L2) and three helices (H1, H2 and H3) at four subsites (A, B, C and D) (Cuneo et al., 2009a). In this study, these four secondary structural elements (L1, H1, H2 and H3) direct the ligand orientation in the active site of the protein. For instance, the loop L1 in the protein MalE directs the MTT (α-glycoside) binding in a downward direction, whereas, the helix H3 in the protein βGlyBP directs the CEL4 (β-glycoside) binding in an upward direction. Thus, the orientation of the ligand in the active site provides a plausible way of distinguishing between cognate and non-cognate ligands in the subcluster D-I SBPs.

For an induced-fit mechanism, knowledge of initial, intermediate(s) and final state of transition process is crucial to understand the conformational dynamics of SBPs. The intermediate transition (open-liganded) state is reported to have a short lifetime (~200 ms for the protein FeuA; de Boer et al., 2019a) and thus, is difficult to accomplish its atomic details. This study presents a total of 24 high-resolution crystal structures of the protein βGlyBP in apo and holo forms representing its four states and mimicking its dynamics of the induced-fit mechanism. These include the initial open-unliganded, partial-open-unliganded, intermediate transition (open-liganded) and closed-liganded states. An in-depth analysis of these four states of the protein βGlyBP affirms its two-step induced-fit mechanism for the ligand binding; as reported previously for the proteins TeaA and LAO (Silva et al., 2005; Fabrizio et al., 2011). In the first step, the ligand (β-glycoside) interacts with the active site residues Trp41 and Trp67 from the NTD and the protein remains in an open-liganded state. Subsequently, in the second step, N1 and C1 subdomains along with another aromatic residue Trp177 of the protein modulate their conformation to form the hydrophobic cage, which accelerates the domain closure to accomplish the final closed-liganded state. This two-step ligand-binding mechanism corroborates the thermodynamic profile of the maltose binding to the protein αGlyBP, where a dual binding mode leads to

both exothermic (first step) and endothermic (second step) reactions (Chandravanshi et al., 2020). Combined structural information suggest that binding of β-glycosides to the protein βGlyBP is an induced-fit mechanism, where the second step primarily follows the "Venus Fly-trap" mechanism. According to this mechanism, the movement of the NTD and CTD occurs in a synchronized manner during ligand binding (Mao et al., 1982). Later, another mechanism namely "asymmetric domain movement" was proposed where the movement of the NTD is shown to be larger than that of the CTD (Pandey et al., 2016).

In this study, however, the atomic details of the protein βGlyBP reveals that "Venus Fly-trap" mechanism is confined only to the N1 and C1 subdomains rather than the whole NTD and CTD (Figure 5.16). Moreover, an interaction analysis between domains allowed us to postulate that the C2 subdomain holds the N1 and C1 subdomains and regulates its plasticity. A structural rigidity of the C2 subdomain corroborates the one domain movement mechanism, where CTD remain static, while movement of NTD brings the conformational dynamics (Chandravanshi et al., 2020). This rigidity domain (subdomain) is crucial to maintain the interaction of SBP to TMD during the ligand translocation (Culurgioni et al., 2017). In summary, to some extent, the flexibility (or rigidity) of one domain (subdomain) of an SBP is required to form an energetically-favorable SBP•TMD complex, so that the other domain (subdomain) can confine the flexibility to one subunit of TMD for the substrate translocation (Figure 5.16).

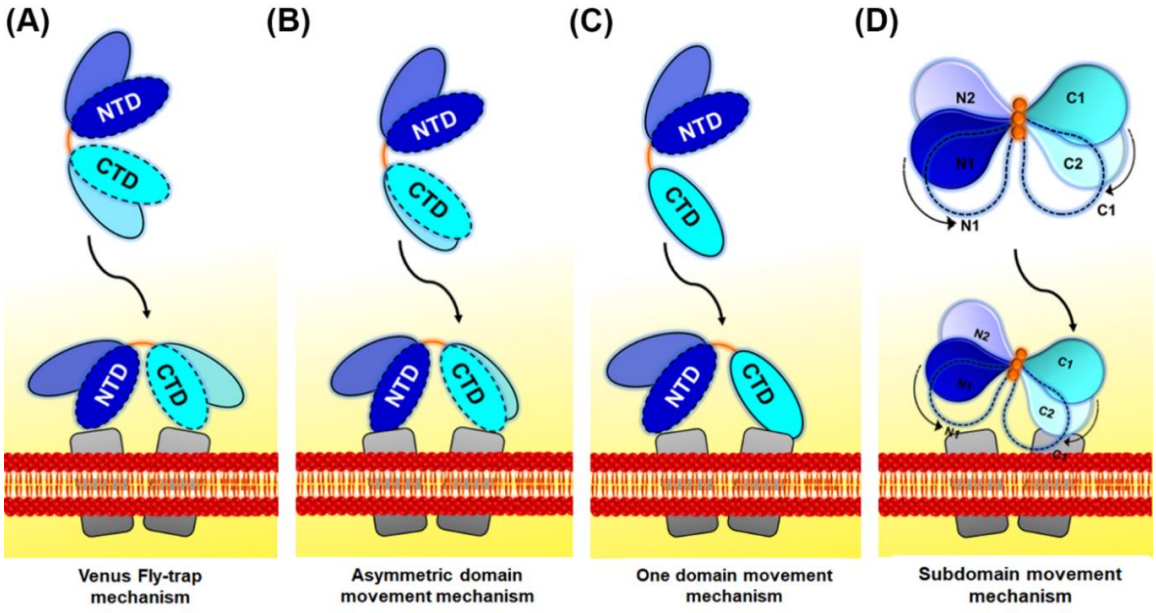


Figure 5.16. Schematic representation of the proposed model for the ligand-binding mechanism. Schematic representation of different proposed modified version of the induced-fit mechanism for the ligand binding. (A) The original "Venus Fly-trap" mechanism (Mao et al., 1982), (B) the "asymmetric domain movement mechanism" (Pandey et al., 2016), (C) the "one domain movement mechanism" (Chandravanshi et al., 2020) and (D) the "subdomain movement mechanism" (this study). The NTD and CTD of SBPs are shown in blue and cyan, respectively, while the TMDs are depicted in grey. The conformational dynamics of the domains (subdomains) are depicted with dotted lines.

The genetic content analysis asserts the functional significance of the protein βGlyBP as a multi-β-glycosides ABC transporter. Moreover, a similar analysis of other members of the subcluster C-IV and D-I SBPs reveals that, despite different type of ligands, the operons encoding β-glycosides ABC transporters always possess genes for carbohydrate-active enzymes (CAZymes), which facilitates the β-glycosides metabolism. These results also indicate that the uptake and metabolism of β-glycosides function in tandem and thus, regulate the ligand selectivity and multi-specificity of the ABC transporter. Operons containing the genes for the protein βGlyBP (ORF ID: TTHB082) and SO-BP (ORF ID: Lin1841) are flanked by the genes encoding GH1 and GH3 β-glucosidases (members of CAZymes), respectively. The enzyme GH1 β-glucosidase (ORF ID: TTHB087) catalyzes the degradation of β-glycosides having preferentially the linkage β-1,3 followed by β-1,2, β-1,4 and β-1,6. Apart from having diverse linkage specificities, the enzyme GH1 β-

glucosidase is specific to cello-oligosaccharides or CELn (up to pentamers), whereas, it shows no enzymatic activity on polysaccharides such as laminarin (Dion et al., 1999). This corroborates the multi-specific nature of the protein βGlyBP. Thus, it can be speculated that the multi-specificity of the downstream metabolic enzyme can lead to the multi-specificity of the associated β-glycosides ABC transporter (ORF IDs: TTHB082-TTHB086). In contrast to the enzyme GH1 β-glucosidase, GH3 β-glucosidase (ORF ID: Lin1840) has been reported to be involved in the β-1,2-glucan dissimilation. Also, the enzyme GH3 β-glucosidase has been demonstrated to have the highest activity towards SOP2 among various sophoro-oligosaccharides (SOPn) and to some extent LAM2 but not for CEL2 and GEN2. Thus, GH3 β-glucosidase was proposed for the SOPn metabolism (Nakajima et al., 2016). Altogether, these observations suggest that the specificity of the enzyme GH3 β-glucosidase for SOPn reflects the similar specificity of SO-BP for SOPn. However, although enzyme GH3 β-glucosidase can hydrolyze LAMn, the protein SO-BP is not shown to bind them, which necessitates further studies to understand the transport of multiple β-glycosides by the protein SO-BP.

5.5 CONCLUSION

This study primarily demonstrates the ligand binding and selection mechanisms of an SBP subunit (βGlyBP) of ABC transport system. Our results suggest that the ligand binding to the protein βGlyBP is a two-step process, where in the first step, the initial interaction between the protein βGlyBP and β-glycosides is established, subsequently, in the second step, the conformational dynamics of subdomain(s) leads to the closure of the active site. Further, results affirm that the protein βGlyBP is multi-specific in nature and binds to different types of β-glycosides via the induced-fit mechanism. The binding of nine cognate ligands, considered in this study, occurs through the conserved nonreducing end (Glc1) as the initial contact unit to the protein suggesting a non-selective nature of the first step of the two-step process. In contrast, the second step determines the selectivity of the ligand in which the conformational dynamics of subdomain(s) lead to a steric clash with non-cognate ligands. Furthermore, results also highlights that the subcluster D-I SBPs, bind α- and β-glycosides in an opposite orientation with the overlapping non-reducing end (Glc1) and

thus, is capable of maintaining the stereo-selection between cognate and non-cognate ligands. In summary, we expect that this study provides the structural and thermodynamic bases for the multi-specificity and stereo-selectivity of SBPs, which would have significant implications in understanding the machinery of ABC transport system(s) and pave way for future drug discovery.

CHAPTER 6 – STRUCTURE OF U3GBP

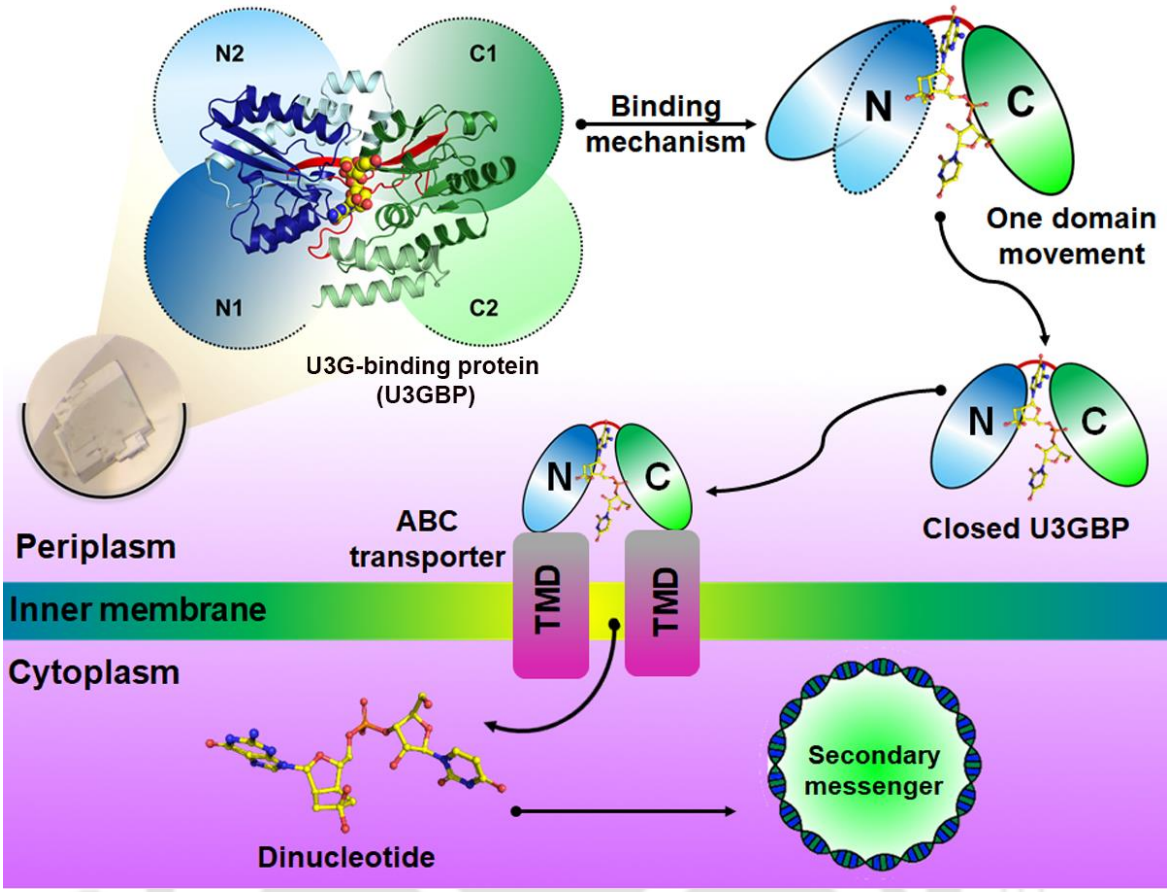
This chapter has been published as:

Chandravanshi M, Gogoi P and Kanaujia SP (2016). Computational characterization of TTHA0379: A potential glycerophosphocholine binding protein of Ugp ATP-binding cassette transporter. *Gene*, 592(2):260-268.

Chandravanshi M, Samanta R and Kanaujia SP (2021). Structural and thermodynamic insights into the novel dinucleotide-binding protein of ABC transporter unveils its moonlighting function. *FEBS J.*, Accepted. DOI: 10.1111/febs.15774.

ABSTRACT

Substrate (or solute)-binding proteins (SBPs) selectively bind the target ligands and deliver them to the ATP-binding cassette (ABC) transport system for their translocation. Irrespective of the different types of ligands, SBPs are structurally conserved. A wealth of structural details of SBPs bound to different types of ligands and the physiological basis of their import are available; however, the same for nucleotides are still deficient. In this study, we elucidated the structural details of an SBP endogenously bound to a novel ligand, a derivative of uridylyl-3'-5'-phospho-guanosine (U3G); thus, we named it as a U3G-binding protein (U3GBP). To the best of our knowledge, this is the first report of U3G (and a dinucleotide) uptake by ABC transporter and thus U3GBP is classified as a first member of subcluster D-I SBPs. Thermodynamic data also suggest that U3GBP can bind phospholipid precursor *sn*-glycerophosphochoine (GPC) at site other than the active site. Moreover, combination of mutagenic and structural information reveals that the protein U3GBP follows the well-known 'Venus fly-trap' mechanism for dinucleotide uptake. We also propose that the molecule U3G might serve as a secondary messenger for several cytosolic proteins involved in the tRNA synthesis and/or modification.



6.1 INTRODUCTION

Acquisition of diverse nutrients from the extra- to the intra-cellular environment of bacterial cells is facilitated by substrate (or solute)-binding proteins (SBPs) (Wilkinson et al., 2003). In gram-negative bacteria, SBPs are located in the periplasmic space, whereas, in gram-positive bacteria, they are anchored to the plasma membrane (van der Heide and Poolman, 2002). SBPs are essential constituents of ATP-binding cassette (ABC) transport systems involved in the uptake of substrates inside the cell (Davidson et al., 2008). SBPs selectively bind to cognate ligand(s) (Miller et al., 1980) and subsequently deliver to them to transmembrane domains (TMDs), which translocate the ligand(s) inside the cell by utilizing the energy of ATP hydrolysis by nucleotide-binding domains (NBDs) (Wilkens, 2015). Depending upon habitats and cellular requirements, bacterial cells possess multiple types of SBPs to facilitate the

uptake of nutrients such as metal ions, sugars, vitamins, peptides, amino acids, etc. (Higgins, 1992; Davidson et al., 2008). Irrespective of ligand types and low sequence similarity, SBPs adapt a similar structural topology. Overall, SBPs preserve a structural fold comprised of α/β N- and C-terminal domains (NTD and CTD), which are connected by a hinge region (Scheepers et al., 2016). SBPs are classified into seven different clusters (A-G) based on their hinge regions and substrate specificities (Berntsson et al., 2010; Scheepers et al., 2016). Upon ligand binding, both NTD and CTD move around the hinge region and transit from an open to closed state through a well-known ‘‘Venus Fly-trap’’ mechanism” (Mao et al., 1982).

SBPs being an essential component of ABC importers, which are present exclusively in prokaryotes and some plants, are considered as drug targets against pathogenic bacteria (Kalscheuer et al., 2010; Lefevre and Boutry, 2018). In pathogens, SBPs mediate the uptake of various crucial nutrients required to maintain their virulence (Culurgioni et al., 2017). For example, syphilis, bejel and yaws disease-causing spirochaete bacterium *Treponema pallidum* lacks the *de novo* purine biosynthesis pathway. Instead, it possesses an ABC transporter which facilitates the uptake of required nucleosi(ti)des and nitrogenous bases (Deka et al., 2006). These nucleotides not only play a role in the nucleic acid biosynthesis, but also serve as signaling molecules (He et al., 2020).

Akin to nucleotides, phospholipids are also biosynthesized through a *de novo* pathway via cytosolic enzyme glycerophosphodiesterase (Ohshima et al., 2008). This enzyme utilizes *sn*-glycerol-3-phosphate (G3P) and *sn*-glycerophosphocholine (GPC) as precursors, which are imported inside the cell by a UgpABCE ABC transporter. The SBP subunit (UgpB) of UgpABCE ABC transporter shares structural similarity with the trehalose/maltose-binding protein (TMBP, PDB ID: 1EU8) and thus has been grouped into the subcluster D-I SBPs (Scheepers et al., 2016). This clustering is in accordance with a previous report, where both sugar and Ugp (uptake glycerol phosphate) ABC transporters are grouped under the CUT1 (carbohydrate uptake transporter-1) family of bacterial ABC transporters (Tommassen et al., 1991; Schneider, 2001; Wuttge et al., 2012). Thus, owing to the structural similarity between the SBP units of sugar and Ugp ABC transporters, some UgpB proteins have been annotated as sugar-binding proteins (Chandravanshi et al., 2019). In our previous study, one such protein

(ORF ID: TTHA0379) from a thermophilic gram-negative bacterium *Thermus thermophilus* HB8 was proposed to be a GPC-binding UgpB protein based on its sequence and structural features (Chandravanshi et al., 2016). In this study, we elucidated the structural details of the protein TTHA0379. The structural analysis suggests that the protein TTHA0379 is a uridylyl-3'-5'-phospho-guanosine (U3G)-binding protein (U3GBP) rather than sugar- and/or GPC-binding proteins. Unlike the other nutrients, structural details and mechanism of nucleotide uptake through ABC transporters are lacking. This study provides mechanistic basis for "Venus Fly-trap" mechanism of U3GBP for U3G (or dinucleotide) binding and correlation between U3G ABC transporter and tRNA synthesis and/or modification system.

6.2 MATERIALS AND METHODS

6.2.1 Preliminary *in silico* analysis of protein TTHA0379

All the amino acid sequences used in this study were extracted from UniProtKB (The UniProt consortium, 2019). The homologous sequences of protein TTHA0379 were searched in the protein database using the tool BLAST (Altschul et al., 1990). To study the evolutionary relationship among protein TTHA0379 and its homologous (UgpB and sugar binding) proteins, phylogenetic analysis was performed using the program MEGA6 (Tamura et al., 2013). It is to be noted that the homologous proteins were selected having different genus, high query coverage (65 to 95%), varying sequence identities (15 to 30%) and reviewed (annotated) status. To identify conserved regions among homologous proteins of TTHA0379, multiple sequence alignment (MSA) was performed using the program Clustal Omega (Sievers and Higgins, 2014). The visualization of MSA was further polished using the tool ESPript (Gouet et al., 2003). The three-dimensional structure of protein TTHA0379 was predicted using the programs I-TASSER (Yang et al., 2015a), RaptorX (Kallberg et al., 2012) and SWISS-MODEL (Biasini et al., 2014). The theoretical models of protein TTHA0379 obtained were further refined by minimizing energy using the program ModRefiner (Xu and Zhang, 2011). These refined models were subsequently validated for their stereo-chemical quality using the program MolProbity (Chen et al., 2010). The final model to be used for further analysis was chosen based on the geometrical parameters. The

structural homologs of protein TTHA0379 were searched using the web server DALI (Holm and Rosenstrom, 2010). To further identify conservation both in primary and tertiary structures of homologous proteins, a structure-based sequence alignment and structural superimposition was performed using the programs PROMALS3D and 3dSS, respectively (Sumathi et al., 2006; Pei and Grishin, 2014). For further functional characterization of protein TTHA0379, ligand binding site was identified and analyzed for its potent substrates. The volume(s) and surface area of ligand binding pocket(s) were calculated using the program CASTp (Binkowski et al., 2003) with a default probe radius of 1.4 Å. All the structural figures were generated using the program PyMOL (The PyMOL Molecular Graphics System, Schrodinger, LLC).

6.2.2 Molecular cloning for U3GBP wild type and its mutants

The cloned pET11a vector (plasmid ID: PC010379-41) containing the open reading frame (ORF) TTHA0379 from *Thermus thermophilus* HB8 coding for full length (1251 bp) U3GBP was procured from Biological Research Center, NITE (NBRC), Japan. Subsequently, a truncated version of the protein (residues 20-417) lacking the N-terminus signal peptide (residues 1-19) was subcloned into the pET22b vector. For subcloning, the truncated gene (1194 bp) was amplified by polymerase chain reaction (PCR) using the full-length recombinant construct as a template and appropriate forward/reverse primers, details of which are provided in the Table 6.1. The amplified gene carrying a 6xHis-tag at the C-terminus was double digested using the enzymes *Nde*I and *Bam*HI before inserting into the pET22b vector. The resultant recombinant construct was confirmed by the double digestion using the same set of enzymes followed by the plasmid DNA sequencing (Figure 6.1A). The recombinant mutant constructs *U3GBP_Y56F*, *U3GBP_F79A*, *U3GBP_N81A*, *U3GBP_S127A*, *U3GBP_Y224A*, *U3GBP_T240A*, *U3GBP_Y246A*, *U3GBP_Q274A*, *U3GBP_Y224A/Y246A*, *U3GBP_F79A/Y224A/Y246A/Δ74-75* and *U3GBP_ F79A/Y224A/Y246A/Δ50-75* were generated by PCR using the specific primer oligonucleotide sequences and templates mentioned in Table 6.1 using Q5 site-directed mutagenesis kit. Subsequently, all these mutant constructs were analyzed in 0.8% agarose gel and subjected to plasmid DNA sequencing for mutagenesis confirmation (Figure 6.1B-6.1L).

Table 6.1. Details of the clones, templates and primer oligonucleotide sequences. Restriction enzyme cutting site (underlined), 6xHis-tag (bold) and point mutations (bold and underlined) are mentioned in the primers.

Clone (primer)	Template	Oligonucleotide sequence (5′-3′)
U3GBP_WT (Forward/ Reverse)	Cloned pET11a plasmid	ATATCATATGATGAAGCCGGAG GATGTGATCAAGGAGC/ ATATGGATCCCTA**ATGATGATG ATGATGATG**GCGGGTCCTGCTG GAG
U3GBP_Y56F (Forward/ Reverse)	U3GBP_WT (recombinant construct)	CAGGGGGGGT**T**CCGGGACCTTTC CAC/ GGGCACGGGGCGGACGCA
U3GBP_F79A (Forward/ Reverse)		GGCCCAGGCC**GC**TGAGAACAAC ATCG/ ATGGTGGGCACCTTCCCG
U3GBP_N81A (Forward/ Reverse)		GGCCTTTGAG**GC**CAACATCGCCC TCTACCTCG/ TGGGCCATGGTGGGCACC
U3GBP_S127A (Forward/ Reverse)		CTTCAACAAG**GC**CATCCAGGTCC TTTACTACAACAAGG/ GGCACCCCGTAGACCACG
U3GBP_Y224A (Forward/ Reverse)		CACCTCGGGC**GC**CATCAACCAG AACCTGGG/ ATGGGCTTGGCCCACCCC
U3GBP_T240A (Forward/ Reverse)		CAGCGTGGAC**G**CCTCCGCGGG/ AAGGCGTAGGGCCCCGAG
U3GBP_Y246A (Forward/ Reverse)		GGGCTACACC**GC**CTACCTCCGGG C/ GCGGAGGTGTCCACGCTG
U3GBP_Q274A (Forward/ Reverse)		CGGCCTCGTC**GC**GGGCACCAAC CTGG/ TAGCCCGGCTGGCCCTTG
U3GBP_Y224A/Y246A (Forward/ Reverse)	U3GBP_Y224A (recombinant construct)	GGGCTACACC**GC**CTACCTCCGGG C/ GCGGAGGTGTCCACGCTG
**U3GBP_F79A/Y224A/Y246A/Δ74-75* (Forward/ Reverse)	U3GBP_Y224A/Y246A (recombinant construct)	GGCCCAGGCC**GC**TGAGAACAAC ATCG/ ATGGTGGGCACCTTCCCG
**U3GBP_F79A/Y224A/Y246A/Δ50-75* (Forward/ Reverse)	U3GBP_Y224A/Y246A (recombinant construct)	GGCCCAGGCC**GC**TGAGAACAAC ATCG/ ATGGTGGGCACCTTCCCG

**U3GBP_F79A* primers were used to generate the triple mutant constructs by using double mutant cloned plasmid as a template, however, during the site-directed mutagenesis experiment random deletions were observed.

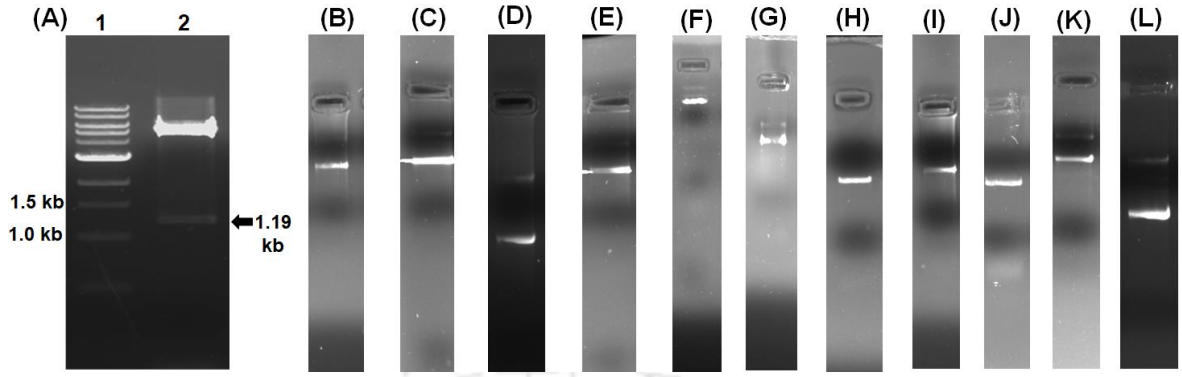


Figure 6.1. Cloning and site-directed mutagenesis of *TTHA0379*. (A) Clone confirmation of recombinant construct of TTHA0379 by double digestion of the vector pET22b carrying the gene *TTHA0379* by *Nde*I and *Bam*HI restriction enzymes (lane 1: DNA ladder, lane 2: insert pop out after double digestion of the plasmid pET22b). Isolated plasmids of all the *TTHA0379* mutant constructs (B) *U3GBP_Y56F* (C) *U3GBP_F79A* (D) *U3GBP_N81A* (E) *U3GBP_S127A*, (F) *U3GBP_Y224A*, (G) *U3GBP_T240A*, (H) *U3GBP_Y246A*, (I) *U3GBP_Q274A*, (J) *U3GBP_Y22A*/Y246A, (K) *U3GBP_ F79A/Y224A/Y246A/*Δ*74-75* and (L) *U3GBP_F79A/Y224A/Y246A/*Δ*50-75*. Clone confirmation of all the *TTHA0379* mutant constructs were done by plasmid DNA sequencing.

6.2.3 Recombinant wild type and mutant protein overexpression and purification

The overexpression of the recombinant protein was optimized in Rosetta (DE3) expression systems at 37°C at different isopropyl-1-thio-β-d-galactopyranoside (IPTG) concentrations (0.1, 0.5 and 1 mM). For protein purification, the wild type (*U3GBP_WT*) and mutant (*U3GBP_Y56F*, *U3GBP_F79A*, *U3GBP_N81A*, *U3GBP_S127A*, *U3GBP_Y224A*, *U3GBP_T240A*, *U3GBP_Y246A*, *U3GBP_Q274A*, *U3GBP_Y224A/Y246A*, *U3GBP_F79A/Y224A/Y246A/*Δ*74-75* and *U3GBP_F79A/Y224A/Y246A/*Δ*50-75)* clones were overexpressed in Rosetta (DE3) cells grown at 37°C in 2-5 L of Luria-Bertani (LB) broth supplemented with 100 μg ml^{-1} ampicillin to an optical density (A_{600}) of 0.6. Subsequently, the grown secondary cultures were induced with IPTG at a final optimized concentration of 1 mM and then cells were further grown at 18°C for an additional 16 hrs. After that, cells were harvested by centrifugation at 3836*g* for 10 min at 4°C. For soluble protein production, the harvested cells were lysed by sonication (2s on and 10s off with 33% amplitude) in a lysis buffer containing 20 mM Tris-HCl pH 7.5, 10 mM imidazole, 150 mM NaCl, 10% glycerol, 3 mM β-mercaptoethanol (β-ME) and 1 mM phenylmethylsulfonyl fluoride (PMSF). The crude lysate of WT and all mutant constructs except

U3GBP_F79A/Y224A/Y246A/Δ74-75 and *U3GBP_F79A/Y224A/Y246A/Δ50-75* were further cleared by heating at 70°C for 10 min followed by centrifugation at 15644*g* for 45 min at 4°C. For the recombinant protein purification, the obtained clear lysate was applied to Ni-NTA column prepared by packing the Ni^{2+}-affinity resin (Qiagen, Hilden, Germany) in a pierce centrifuge column (Thermo Fisher Scientific, Waltham, Massachusetts, United States). Subsequently, the loaded column was washed with 10 column-volume (CV) of wash buffer A (20 mM Tris-HCl pH 7.5, 10 mM imidazole, 150 mM NaCl, 10% glycerol, 3 mM β-ME and 1 mM PMSF) and 10 CV of wash buffer B (20 mM Tris-HCl pH 7.5, 20 mM imidazole and 150 mM NaCl). The recombinant protein was eluted from the column with wash buffer B containing high concentration of imidazole (250 mM). Pure eluted fractions were pooled together and subjected for the imidazole removal through a gradient dialysis method against buffer 20 mM Tris-HCl pH 7.5, 125-0 mM imidazole and 150 mM NaCl. Subsequently, the eluted proteins U3GBP_WT, U3GBP_Y56F, U3GBP_F79A, U3GBP_N81A, U3GBP_S127A, U3GBP_Y224A, U3GBP_T240A, U3GBP_Y246A, U3GBP_Q274A, U3GBP_Y224A/Y246A, U3GBP_F79A/Y224A/Y246A/Δ74-75 and U3GBP_F79A/Y224A/Y246A/Δ50-75 were concentrated up to ~23, ~66, ~81, ~57, ~37, ~39, ~25, ~50, ~22, ~106, ~29 and ~11 mg ml^{-1}, respectively, measured using UV_{280} method (ε_{280} = 50435 M^{-1} cm^{-1}) (Figure 6.2). The final protein buffer contained 20 mM Tris-HCl pH 7.5 and 150 mM NaCl.

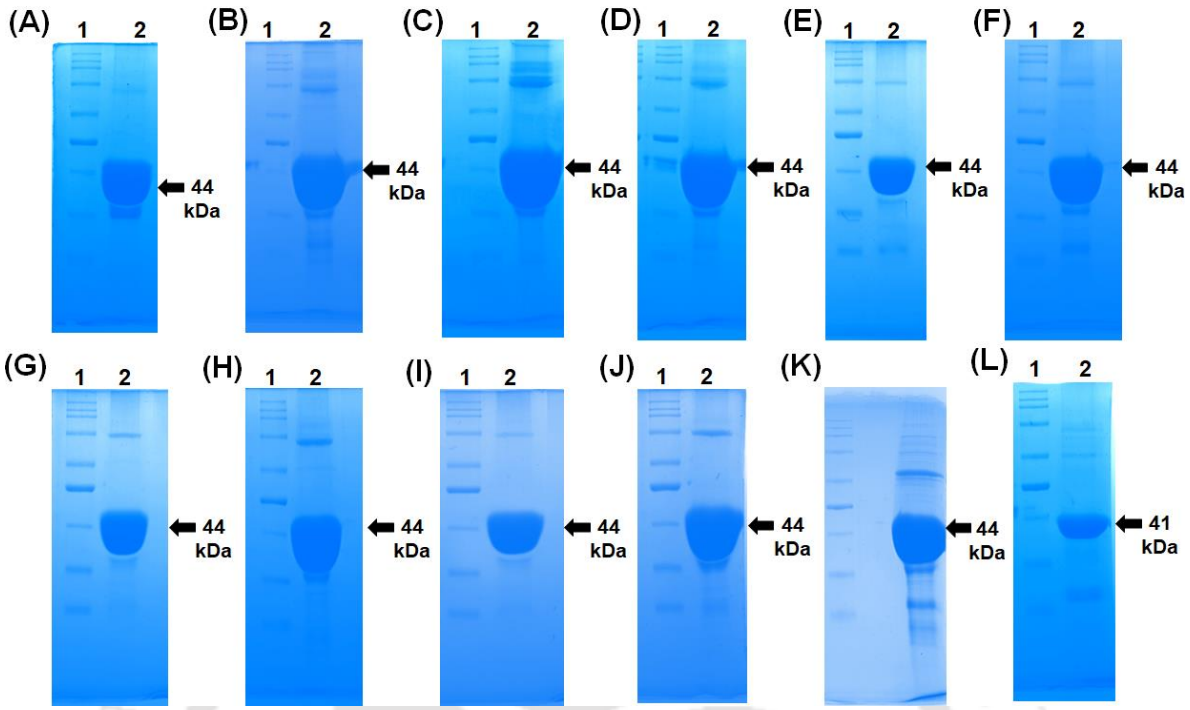


Figure 6.2. Purification of U3GBP_WT and mutant protein. SDS-PAGE analysis of purified and concentrated proteins (A) U3GBP_WT, (B) U3GBP_Y56F (C) U3GBP_F79A (D) U3GBP_N81A (E) U3GBP_S127A, (F) U3GBP_Y224A, (G) U3GBP_T240A, (H) U3GBP_Y246A, (I) U3GBP_Q274A, (J) U3GBP_Y22A/Y246A, (K) U3GBP_ F79A/Y224A/Y246A/Δ74-75 and (L) U3GBP_F79A/Y224A/Y246A/Δ50-75. (lane 1: protein ladder, lane 2: concentrated protein).

6.2.4 Isothermal titration calorimetry

For thermodynamic parameter measurement of ligand binding, isothermal titration calorimetry (ITC) experiments were performed using MicroCal ITC200 (GE healthcare, City, Country). Before each experiment, the protein was dialyzed in a buffer containing 50 mM Tris-HCl pH 7.5 and 150 mM NaCl and ligands were prepared in the same to minimize the heat of dilution. The prepared protein and ligand samples were placed into cells and syringe with a final concentration of 135 µM and 5 mM, respectively. For each titration, 0.4 µl of pre-injection followed by 1.3 µl of 24 injections were titrated into the sample at every 180 s. Titrations were performed at 25°C with stirring speed of 450 rpm. Heat changes associated with dilution of ligands in the buffer was also considered and subtracted from the heat of reactions to get the true heat associated with protein-ligand binding. Analysis of the subtracted data was performed using the program Origin (version 7). All the integrated data were fitted

using a one-site binding model and the thermodynamic parameters such as enthalpy change (ΔH), stoichiometry ratio (n) and association constant (*Ka*) were determined. Other thermodynamic parameters such as change in Gibbs free energy (ΔG) and entropy change (ΔS) were calculated using the equation $\Delta G = \Delta H - T\Delta S$.

6.2.5 Crystallization, data collection and structure determination

The purified wild type (U3GBP_WT) protein was screened for initial crystal hits using crystallization conditions available in commercial kits from Hampton Research utilizing microbatch-under-oil technique at 4 and 20°C temperatures. Initial hits of the crystals were obtained within 2-3 weeks at 4°C in a condition having 0.2 M ammonium sulfate, 0.1 M sodium cacodylate trihydrate pH 6.5 and 30% (w/v) polyethylene glycol 8000. However, morphology of these obtained crystals was a thin plate grown as a part of multi-stacked crystals. Hence, further optimization of crystals was performed by using hanging-drop vapor-diffusion method at 4°C temperature. The optimization of protein crystals was performed by changing the salt (ammonium sulfate to ammonium phosphate) and varying precipitant and glycerol at 1-10% concentrations. This improved the thickness of the crystals and a diffractable crystal could be obtained in a condition having the original condition along with glycerol (7-8%) as an additive (Figure 6.3). With further optimization of the condition, single and bigger crystals could be obtained using matrix micro-seeding (MMS) method in a condition containing 0.2 M ammonium phosphate, 0.1 M sodium cacodylate trihydrate pH 6.5 and 30% (w/v) polyethylene glycol 8000. In a similar manner, crystals of mutants U3GBP_Y56F, U3GBP_F79A, U3GBP_N81A, U3GBP_S127A, U3GBP_Y224A, U3GBP_T240A, U3GBP_Y246A, U3GBP_Q274A, U3GBP_Y224A/Y246A was obtained using MMS method employing both microbatch-under-oil and hanging-drop vapor-diffusion techniques in an optimized condition (0.2 M ammonium phosphate, 0.1 M sodium cacodylate trihydrate pH 6.5, 30% (w/v) polyethylene glycol 8000 and ± 8% glycerol) at 4°C (Figure 6.3). However, crystals of the triple mutants U3GBP_F79A/Y224A/Y246A/Δ50-75) were achieved in a condition containing 0.2 M ammonium sulfate, 0.1 M sodium cacodylate trihydrate pH 6.5 and 30% (w/v) polyethylene glycol 8000 using microbatch-under-oil method.

X-ray intensity data of all the wild type and mutant protein crystals were collected at Central Instruments Facility (CIF), Indian Institute of Technology Guwahati, India using Cu *Kα* radiation (wavelength 1.5418 Å) produced by Rigaku MicroMax-007 HF diffractometer (operated at 40 kV and 30 mA), on R-Axis IV++ image-plate detector equipped with rotating anode X-ray tube. At the time of data collection, all crystals were cryoprotected using 10% glycerol and flash-cooled under stream of liquid nitrogen. All these data of U3GBP were processed, indexed, integrated and scaled using the programs iMosflm (Battye et al., 2011) and Aimless (Evans and Murshudov, 2013) embedded in the CCP4 package (Winn, 2011). Summary for X-ray intensity data collection and processing statistics of all the crystals are provided in Table 6.2-6.6.

Initial phases of the protein U3GBP_WT was determined employing the molecular replacement method using the crystal structure of *Mt*UgpB (PDB ID: 4MFI), which has a sequence identity (query coverage) of 26 (77)% as a search model using the program Phaser (McCoy et al., 2007). The structure solution of the mutant proteins was also obtained by molecular replacement method using program Phaser and the structure of U3GBP_WT as a search model. In order to calculate the R_{free}, 5% of the total reflections were kept aside as a test data set (Brunger, 1992). The atomic model building and iterative cycles of structural parameters refinement were carried out using the programs Coot (Emsley et al., 2010) and Refmac5 (Vagin, 2004), respectively. The first cycle of structure refinement itself showed a clear electron density for the endogenously-bound U3G in the active site of the protein. However, firstly the protein atoms and water molecules were fitted in the electron density map. Finally, the endogenously-bound molecule U3G was fitted. Structural quality of final refined models was validated using programs PROCHECK (Laskowski et al., 1993) and MolProbity (Chen et al., 2010). The details of the structure refinement and validation of the final structure models are provided in the Table 6.2-6.6. The three-dimensional atomic coordinates of all the wild type and mutant structures of U3GBP have been deposited in the RCSB Protein Data Bank (Berman et al., 2000).

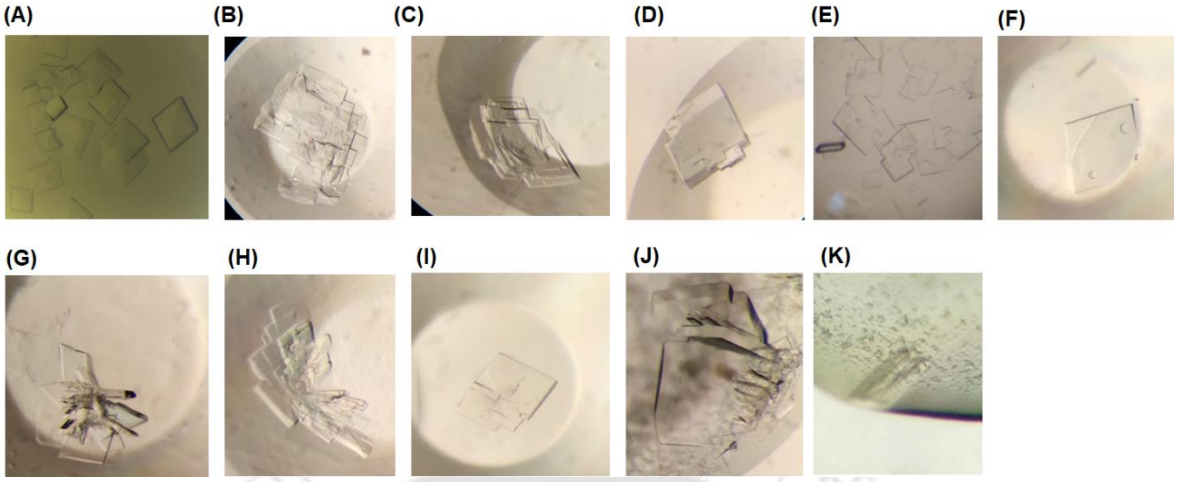


Figure 6.3. Crystallization of protein U3GBP. Protein crystals of U3GBP and its mutant in microbatch and hanging drop (A) U3GBP_WT, (B) U3GBP_Y56F (C) U3GBP_F79A (D) U3GBP_N81A (E) U3GBP_S127A, (F) U3GBP_Y224A, (G) U3GBP_T240A, (H) U3GBP_Y246A, (I) U3GBP_Q274A, (J) U3GBP_Y22A/Y246A and (K) U3GBP_F79A/Y224A/Y246A/Δ50-75.

Table 6.2. Data collection and refinement statistics for U3GBP_WT and U3GBP_Y56F protein. The values in parenthesis represents the statics for the last resolution shell.

	U3GBP_WT-Form I	**U3GBP_WT-Form II**	**U3GBP_WT-Form III**	**U3GBP_Y56F**
Wavelength (Å)	1.5418	1.5418	1.5418	1.5418
Temperature (K)	100	100	100	100
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters (Å, °)	a=58.16, b=120.88, c=66.06, $\alpha=\beta=\gamma$=90	a=66.17, b=57.52, c=121.13, $\alpha=\gamma$=90, β=94.99	a=65.91, b=58.12, c=102.4, $\alpha=\gamma$=90, β=93.04	a=66.60, b=58.45, c=123.09, $\alpha=\gamma$=90, β=95.10
Resolution (Å)	66.06-2.15 (2.21-2.15)	46.57-1.80 (1.83-1.80)	56.73-1.85 (1.89-1.85)	61.15-2.00 (2.05-2.00)
No. of observed reflections	134555 (10324)	361385 (18701)	154455 (8450)	265289 (18095)
No. of unique reflections	48571 (3803)	84314 (4443)	55937 (3383)	61494 (4284)
Mn(I) CC(1/2)	0.987 (0.697)	0.998 (0.792)	0.978 (0.768)	0.980 (0.846)

Completeness (%)	98.2 (95.1)	99.9 (99.6)	83.8 (82.3)	97.2 (92.6)
V_M ($Å^3$ Da^{-1})	2.63	2.60	2.22	2.70
Solvent content (%)	53.22	52.70	44.53	54.48
Mosaicity (°)	0.6	0.35	0.6	0.60
MeanI/σ(I)	6.9 (2.5)	12.3 (2.6)	9.3 (2.9)	8.0 (3.3)
R_{merge}† (%)	10.6 (42.9)	6.5 (46.0)	10.0 (39.4)	11.8 (41.6)
R_{pim} (%)	10.4 (41.0)	5.6 (39.4)	9.9 (38.5)	13.2 (46.4)
R_{meas} (%)	14.8 (59.4)	8.6 (60.8)	14.1 (55.1)	8.6 (60.8)
Multiplicity	2.8 (2.7)	4.3 (4.2)	2.8 (2.5)	4.3 (4.2)
R_{work}/R_{free} (%)	15.82/20.41	14.60/17.62	17.80/22.27	17.92/22.67
Protein model				
No. of subunits in ASU	2	2	2	2
Protein atoms	6114	6168	6093	6157
Water molecules	415	637	520	656
U3G	2	2	2	2
Others	33	47	13	13
Deviation from ideal geometry				
Bond length (Å)	0.015	0.018	0.014	0.016
Bond angles (°)	1.927	2.210	1.882	1.956
Average *B*-factor (Å²)				
Protein atoms	17.70	15.55	11.41	13.65
Water molecules	29.52	31.32	26.75	28.54
U3G	19.45	17.23	16.75	20.32
Ramachandran plot (%)				
Favored	96.80	97.44	97.44	97.19
Allowed	3.20	2.56	2.56	2.81
Remaining	0.00	0.00	0.00	0.00
PDB ID	7C0F	7C0K	7C0L	7C0O

†$R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl, $\sum hkl$ is the sum overall reflections and $\sum i$ is the sum over i measurements of reflection hkl.
U3G: Uridylyl-3'-5'-phospho-guanosine

Table 6.3. Data collection and refinement statistics of U3GBP mutant proteins (F79A, N81A and S127A). The values in parenthesis represents the statics for the last resolution shell.

	U3GBP_F79A -Form I	**U3GBP_F79A-Form II**	**U3GBP_N81A**	**U3GBP_S127A**
Wavelength (Å)	1.5418	1.5418	1.5418	1.5418
Temperature (K)	100	100	100	100
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters (Å, °)	a=58.39, b=121.58, c=66.19, α= β=γ=90	a=66.53, b=58.36, c=123.82, α=γ=90, β=94.81	a=66.61, b=58.34, c=123.48, α=γ=90, β=95.04	a=66.99, b=57.28, c=121.59, α=γ=90, β=94.93
Resolution (Å)	60.79-1.77 (1.80-1.77)	61.67-1.90 (1.94-1.90)	61.50-1.77 (1.80-1.77)	46.91-1.80 (1.83-1.80)
No. of observed reflections	389869 (17149)	316199 (17672)	557411 (25421)	288210 (14898)
No. of unique reflections	85038 (4067)	69777 (4141)	87549 (4100)	82790 (4301)
Mn(I) CC(1/2)	0.991 (0.706)	0.994 (0.890)	0.996 (0.959)	0.985 (0.728)
Completeness (%)	95.4 (86.7)	93.1 (85.8)	95.3 (88.3)	97.1 (92.6)
V_M ($Å^3$ Da^{-1})	2.66	2.71	2.71	2.63
Solvent content (%)	53.76	54.65	54.55	53.26
Mosaicity (°)	0.6	0.6	0.6	0.5
Mean I/σ(I)	9.3 (2.4)	13.1 (4.0)	15.1 (5.0)	7.0 (2.5)
R_{merge}† (%)	15.2 (49.2)	7.3 (36.4)	7.8 (22.0)	10.0 (44.1)
R_{pim} (%)	7.6 (25)	5.9 (30.5)	4.9 (14.1)	8.9 (38.1)
R_{meas} (%)	17.0 (55.3)	9.5 (47.7)	9.3 (26.2)	13.4 (58.5)
Multiplicity	2.3 (2.2)	4.5 (4.3)	6.4 (6.2)	3.5 (3.5)
R_{work}/R_{free} (%)	14.30/18.23	20.48/23.98	18.06/21.86	22.59/27.43
Protein model				
No. of subunits in ASU	2	2	2	2
Protein atoms	6120	6144	6172	6196
Water molecules	700	724	845	711
U3G	2	2	2	2
Others	18	8	13	13
Deviation from ideal geometry				

Bond length (Å)	0.018	0.015	0.016	0.014
Bond angles (°)	2.286	1.997	2.037	1.899
Average *B*-factor (Å^2)				
Protein atoms	15.00	16.84	12.98	11.19
Water molecules	28.54	33.47	29.48	25.84
U3G	20.05	22.24	14.36	14.93
Ramachandran plot (%)				
Favored	97.70	97.31	97.30	97.16
Allowed	2.30	2.69	2.70	2.84
Remaining	0.00	0.00	0.00	0.00
PDB ID	7C0R	7C0S	7C0T	7C0U

†R_{merge} = $\sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl, $\sum hkl$ is the sum overall reflections and $\sum i$ is the sum over i measurements of reflection hkl.
U3G: Uridylyl-3'-5'-phospho-guanosine

Table 6.4. Data collection and refinement statistics of U3GBP mutant proteins (Y224A, T240A and Y246A). The values in parenthesis represents the statics for the last resolution shell.

	U3GBP_Y224A-Form I	**U3GBP_Y224A-Form II**	**U3GBP_T240A**	**U3GBP_Y246A-Form I**
Wavelength (Å)	1.5418	1.5418	1.5418	1.5418
Temperature (K)	100	100	100	100
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters (Å, °)	a=66.33, b=57.78, c=121.84, α=γ=90, β=94.99	a=65.96, b=58.06, c=103.08, α=γ=90, β=93.25	a=66.64, b=58.08, c=122.02, α=γ=90, β=95.34	a=65.08, b=58.52, c=102.64, α=γ=90, β=92.75
Resolution (Å)	60.69-1.90 (1.94-1.90)	58.06-2.10 (2.16-2.10)	60.75-1.77 (1.81-1.77)	56.13-1.70 (1.73-1.70)
No. of observed reflections	201108 (12057)	183894 (14618)	355385 (16007)	372306 (18135)
No. of unique reflections	68792 (4085)	45231 (3648)	83929 (3881)	83006 (4230)
Mn(I) CC(1/2)	0.964 (0.654)	0.983 (0.797)	0.974 (0.764)	0.991 (0.888)

Completeness (%)	94.6 (87.6)	99.0 (97.6)	93.1 (81.7)	97.6 (94.6)
V_M ($Å^3$ Da^{-1})	2.63	2.23	2.66	2.21
Solvent content (%)	53.29	44.87	53.79	44.35
Mosaicity (°)	0.6	0.6	0.6	0.6
Mean I/σ(I)	5.9 (2.3)	6.9 (3.2)	6.6 (2.6)	8.7 (3.7)
R_{merge}† (%)	13.1 (42.8)	13.6 (41.9)	14.6 (41.7)	11.0 (33.2)
R_{pim} (%)	12.9 (42.3)	12.1 (35.5)	12.0 (34.2)	5.5 (17.9)
R_{meas} (%)	18.4 (60.2)	18.3 (55.1)	19.0 (54.3)	12.3 (37.9)
Multiplicity	2.9 (3.0)	4.1 (4.0)	4.1 (4.2)	4.5 (4.3)
R_{work}/R_{free} (%)	19.45/24.56	20.06/25.00	18.56/22.64	17.68/21.44
Protein model				
No. of subunits in ASU	2	2	2	2
Protein atoms	6108	6095	6147	6182
Water molecules	665	411	756	729
U3G	2	2	2	2
Others	12	6	15	10
Deviation from ideal geometry				
Bond length (Å)	0.013	0.012	0.016	0.015
Bond angles (°)	1.795	1.781	2.025	1.971
Average *B*-factor (Å²)				
Protein atoms	13.33	14.35	14.33	12.39
Water molecules	29.43	25.83	28.50	29.26
U3G	18.29	18.75	16.49	16.61
Ramachandran plot (%)				
Favored	97.57	96.42	97.16	97.54
Allowed	2.43	3.58	2.84	2.46
Remaining	0.00	0.00	0.00	0.00
PDB ID	7C0V	7C0W	7C0X	7C0Y

†$R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl, $\sum hkl$ is the sum overall reflections and $\sum i$ is the sum over i measurements of reflection hkl.
U3G: Uridylyl-3'-5'-phospho-guanosine

Table 6.5. Data collection and refinement statistics of U3GBP mutant proteins (Y246A, Q274A and Y224A/Y246A). The values in parenthesis represents the statics for the last resolution shell.

	U3GBP_Y246 A-Form II	U3GBP_Q274 A-Form I	U3GBP_Q274 A-Form II	U3GBP_Y224A/ Y246A-Form I
Wavelength (Å)	1.5418	1.5418	1.5418	1.5418
Temperature (K)	100	100	100	100
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters (Å, °)	a=58.80, b=104.39, c=65.74, α=β=γ=90	a=66.20, b=58.29, c=121.90, α=γ=90, β=94.78	a=57.73, b=102.56, c=66.18, α=γ=90, β=90.09	a=58.52, b=121.69, c=65.40, α=γ=90, β=90.07
Resolution (Å)	65.74-2.20 (2.27-2.20)	65.97-2.40 (2.49-2.40)	57.73-1.80 (1.84-1.80)	60.84-1.90 (1.94-1.90)
No. of observed reflections	172451 (13680)	151772 (15635)	264684 (14954)	294910 (18576)
No. of unique reflections	39359 (3320)	36616 (3809)	69911 (3973)	71958 (4654)
Mn(I) CC(1/2)	0.978 (0.778)	0.951 (0.714)	0.987 (0.728)	0.969 (0.692)
Completeness (%)	97.9 (95.9)	100 (100)	98.3 (95.4)	100 (100)
V_M ($Å^3$ Da^{-1})	2.28	2.65	2.22	2.63
Solvent content (%)	46.16	53.65	44.55	53.35
Mosaicity (°)	0.6	0.6	0.6	0.45
Mean I/σ(I)	7.1 (2.9)	6.4 (2.7)	7.2 (2.3)	6.8 (2.3)
R_{merge}† (%)	15.0 (44.8)	19.4 (49.3)	10.6 (47.7)	13.3 (50.2)
R_{pim} (%)	12.5 (38.4)	17.3 (43.6)	9.5 (38.8)	11.8 (42.7)
R_{meas} (%)	19.6 (54.2)	26.1 (66.1)	14.2 (61.7)	17.8 (66.2)
Multiplicity	4.4 (4.1)	4.1 (4.1)	3.8 (3.8)	4.1 (4.0)
R_{work}/R_{free} (%)	18.78/23.78	22.74/29.37	16.13/20.94	16.29/20.44
Protein model				
No. of subunits in ASU	2	2	2	2
Protein atoms	6068	6096	6116	6073
Water molecules	110	317	521	535
U3G	2	2	2	2
Others	1	4	14	20
Deviation from ideal geometry				
Bond length (Å)	0.013	0.010	0.017	0.018
Bond angles (°)	2.030	1.720	2.157	2.229

Average *B*-factor (Å^2)				
Protein atoms	13.95	7.46	15.25	13.63
Water molecules	20.16	18.69	29.38	25.44
U3G	24.69	23.51	19.02	21.26
Ramachandran plot (%)				
Favored	94.85	95.80	96.14	97.07
Allowed	5.15	4.20	3.86	2.93
Remaining	0.00	0.00	0.00	0.00
PDB ID	7C0Z	7C14	7C15	7C16

†$R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl, $\sum hkl$ is the sum overall reflections and $\sum i$ is the sum over i measurements of reflection hkl.
U3G: Uridylyl-3'-5'-phospho-guanosine

Table 6.6. Data collection and refinement statistics of U3GBP mutant proteins (Y224A/Y246A and F79A/Y224A/Y246A/Δ50-75). The values in parenthesis represents the statics for the last resolution shell.

	U3GBP_Y224A/Y246A-Form II	**U3GBP_F79A/Y224A/Y246A/Δ50-75**
Wavelength (Å)	1.5418	1.5418
Temperature (K)	100	100
Space group	$P2_1$	$P2_12_12_1$
Unit-cell parameters (Å, °)	a=66.31, b=58.84, c=103.58, α=γ=90, β=93.10	a=52.93, b=55.47, c=124.40, α= β=γ=90
Resolution (Å)	58.84-1.77 (1.81-1.77)	62.20-2.30 (2.38-2.30)
No. of observed reflections	341086 (16991)	181309 (17280)
No. of unique reflections	77250 (4209)	16950 (1607)
Mn(I) CC(1/2)	0.980 (0.911)	0.997 (0.928)
Completeness (%)	99.7 (96.0)	100 (100)
V_M (Å^3 Da^{-1})	2.28	2.20
Solvent content (%)	46.16	44.17
Mosaicity (°)	0.6	0.6
Mean I/σ(I)	11.0 (3.6)	11.5 (4.1)
R_{merge}† (%)	9.7 (24.3)	14.5 (53.3)
R_{pim} (%)	8.0 (20.1)	6.6 (24.6)
R_{meas} (%)	12.7 (31.7)	16.0 (58.8)
Multiplicity	4.4 (4.0)	10.7 (10.8)
R_{work}/R_{free} (%)	18.13/22.17	19.49/27.13

Protein model		
No. of subunits in ASU	2	1
Protein atoms	6125	2817
Water molecules	742	120
U3G	2	-
Others	9	2
Deviation from ideal geometry		
Bond length (Å)	0.014	0.013
Bond angles (°)	1.892	1.898
Average *B*-factor (Å^2)		
Protein atoms	12.34	20.88
Water molecules	32.06	28.04
U3G	21.7	-
Ramachandran plot (%)		
Favored	97.29	95.59
Allowed	2.71	4.41
Remaining	0.00	0.00
PDB ID	7C19	7C1B

†$R_{merge} = \sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|/ \sum_{hkl}\sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection *hkl*, $\sum hkl$ is the sum overall reflections and $\sum i$ is the sum over *i* measurements of reflection *hkl*.
U3G: Uridylyl-3'-5'-phospho-guanosine

6.2.6 Circular dichroism measurement

Temperature-dependent conformational changes in the secondary structure of the wild type and mutant proteins of U3GBP were recorded by measuring the Far-UV CD spectra between the wavelength 190 to 260 nm using circular dichroism (CD) spectrometer JASCO J-1500 (JASCO, Pfungstadt, Germany). Each protein sample prepared in the buffer 20 mM Tris-HCl pH 7.5 and 150 mM NaCl were diluted in water to a final concentration of 5 µM and placed in the 0.1 cm optical path-length quartz cuvette for spectra measurement. CD spectra of the blank sample (only buffer) was subtracted from that of the protein sample. Each measurement was performed with a default set of parameters such as scan speed 100 nm min^{-1}, response 2s and sensitivity 100 millidegree. Each CD run includes the three-repeat scans followed by averaging to generate the final spectra. The final spectra with these set of parameters was further used to probe the thermal denaturation profile of the wild type and mutant proteins of U3GBP over the temperature range of 20 to 120°C with a heating rate of 2°C min^{-1} and wavelength range of 190 to 260 nm. The denaturation analysis of the spectra was performed using the Spectra Manager software available with the JASCO J-1500. To calculate the

melting temperature (Tm) of wild type and mutant proteins of U3GBP, change in ellipticity (mdeg) was monitored over different temperature and recorded as melting temperature (Tm) at which a sharp decline was observed. For final representation, the thermal denaturation profile was smoothened and plotted in a two-dimensional graph as a function of temperature with respect to different wavelengths using the program Origin (version 9.6).

6.2.7 Mass spectrometry analysis

To estimate the molecular weight of U3GBP variants, matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) method was employed. The homogenously-purified protein samples were mixed with the matrix (sinapinic acid) in a (protein:matrix) ratio of 1:2, 1:3 and 1:5. The sinapinic acid matrix was prepared to a final concentration of 10 mg ml^{-1} dissolved in a solvent containing 100% acetonitritle and 0.1% trifluoroacetic acid in a ratio of 30:70, respectively. Subsequently, the sinapinic acid was dissolved in the solvents by sonicating it for 30 min followed by centrifugation at 12000*g* for 10 min to get rid of the excessive matrix particles. The final prepared sinapinic acid matrix was mixed with the protein samples of 1, 5 and 10 mg ml^{-1} in the ratios specified above. Subsequently, a 2 µL spot of the protein:matrix mix were settled in the grooves of the steel MALDI plate and was allowed to dry. Post drying, the drops were ionized in a Bruker autoflex speed and all data were analyzed using the in-built Flex control analysis software. For final representation, graphs were plotted using the program Origin (version 9.6).

6.2.8 Bioinformatics analysis of U3GBP crystal structure

Members of the (sub)cluster D SBPs and other structural homologs were identified using the web server Dali (Holm and Rosenström, 2010). For evolutionary relation with these proteins and to identify the (sub)cluster of U3GBP, the structure- and sequence-based phylogenetic trees were generated by retrieving the structural homologs from (sub)cluster D SBPs reported in a previous report (Scheepers et al., 2016). Five SBPs from each (sub)cluster D SBPs were selected based on different bacterial genera. The resulting 30 SBPs along with the protein U3GBP structures were aligned pairwise using the web server PDBeFold (Krissinel and Henrick K, 2019), which generated a distance-based matrix of their root mean square deviation (rmsd) and sequence identity values. Subsequently, the rmsd and sequence identity

matrices were converted into a tab-delimited format to provide the suitable input to the program dendroUPGMA (Garcia-Vallvé and Puigbo, 2009), which converts a distance matrix into the Newick file format. Finally, the structure- and sequence-based distance trees were generated from the Newick file using a web tool iTOL (Letunic and Bork, 2016). The electrostatic surface charge distribution of protein was calculated using the plugin APBS embedded in the visualization program PyMOL (Molecular Graphics System, Version 1.7.2.2 Schrödinger, LLC). In addition to structural homolog, other homologous proteins of U3GBP based on sequence similarity was identified by using the web tool BLAST against the UniProtKB database (The UniProt Consortium, 2019). All the identified homologs were further subjected to a comparative genetic operon analysis. To visualize the operonic arrangement of each homolog protein, functionally-associated flanking genes were extracted from the Gene database, located at the National Center for Biotechnology Information (NCBI). Further, protein sequence of TTHA0375, extracted from the UniProtKB database, was used as a query to identify its functional homologs using the web tool BLAST (Altschul et al., 1990). All the multiple sequence alignments were performed using the program Clustal Omega (Sievers and Higgins, 2014). The output of the multiple sequence alignment was further furnished using the web tool ESPript (Gouet et al., 2003). The theoretical tertiary structure model of the protein TTHA0375 was predicted using the program SWISS-MODEL (Biasini et al., 2014). The theoretical structural models were further validated based on their geometrical and stereo-chemical parameters in the Ramachandran plot using the web tool PDBsum (De Beer et al., 2014). Further, to identify the functionally-associated proteins of TTHA0375, a protein-protein interactome network was generated using the database STRING v11 with a default set of parameters (Szklarczyk et al., 2019). The generated protein interaction networks were further rendered using the program Gephi (Bastian et al., 2009). All the molecular graphic figures of this study were prepared using the molecular visualization tool PyMOL.

6.3 RESULTS

6.3.1 Protein TTHA0379 is misannotated as sugar-binding protein

The protein TTHA0379 is annotated as "sugar ABC transporter, periplasmic sugar-binding protein" in the UniProtKB database (The UniProt Consortium, 2019). A search for homologous of protein TTHA0379 using the tool PSI-BLAST against UniProt database, reveals that protein UgpB and sugar-binding proteins show significant sequence identity in the range of 18 to 26% to the protein TTHA0379. Further, phylogenetic analysis of protein UgpB and sugar-binding protein along with protein TTHA0379 clearly demonstrates the grouping of protein TTHA0379 with protein UgpB instead of sugar-binding protein (Figure 6.4A). To gain further insight into similarity and dissimilarity of protein TTHA0379 with both UgpB and sugar-binding proteins, structural details of protein TTHA0379 was analyzed. The three-dimensional structure of protein TTHA0379 was modeled using three different programs (I-TASSER, RaptorX and SWISS-MODEL). Notably, all these programs used the tertiary structure of protein UgpB from *Escherichia coli* (*Ec*UgpB, PDB ID: 4AQ4) and *Mycobacterium tuberculosis* (*Mt*UgpB, PDB ID: 4MFI) as the default template. The root mean square deviation (RMSD) between the final theoretical model of protein TTHA0379 and the backbone atoms of the templates (4AQ4 and 4MFI) is 3.0 Å and 1.0 Å, respectively, indicating that tertiary structure of protein TTHA0379 is more similar to *Mt*UgpB than to that of *Ec*UgpB. Surprisingly, structural comparison of modeled protein TTHA0379 with protein UgpB and sugar-binding proteins shows a significant resemblance among each other (Figure 6.4B). Similar to protein UgpB, protein TTHA0379 shows conservation with sugar-binding proteins only in the β-sheet of the N-domain and the hinge region (Figure 6.4B).

To further shed light on the similarity of protein TTHA0379 with protein UgpB, the ligand binding sites of both (UgpB and sugar-binding) proteins along with protein TTHA0379 were superimposed. Earlier studies have shown that a tryptophan residue (Trp172) present in the active site of *Ec*UgpB (PDB ID: 4AQ4) is replaced by a tyrosine residue in sugar-binding proteins (Wuttge et al., 2012). Akin to *Ec*UgpB, active site of protein TTHA0379 also contains a tryptophan (Trp188) residue instead of a tyrosine residue as in the case of sugar-binding proteins. Furthermore, two residues glutamate and arginine (Glu66 and Arg374 in

*Ec*UgpB, Asp102 and Arg385 in *Mt*UgpB) are the signature residues of protein UgpBs. Moreover, protein TTHA0379 also comprises of the signature residues Glu99 and Arg379 in the active site, similar to *Ec*UgpB (Figure 6.4C). Thus, active site superimposition reveals that the ligand binding sites of proteins UgpB and TTHA0379 are different from that of sugar-binding proteins, wherein the signature residues (tryptophan and glutamate) partially occupy the active site. Furthermore, earlier studies of sugar-binding proteins suggested that they contain ten conserved regions, which stabilizes the sugar molecule in the active site (Medrano et al., 2014). Out of which, two regions (namely, IV: polar and/or acidic and VII: tryptophan rich) show highest conservation among all sugar-binding proteins (Medrano et al., 2014). A comparison of protein TTHA0379 with sugar-binding proteins unveils that two regions (IV and VII) of sugar-binding proteins are completely absent in protein UgpB including protein TTHA0379 (Figure 6.4D).

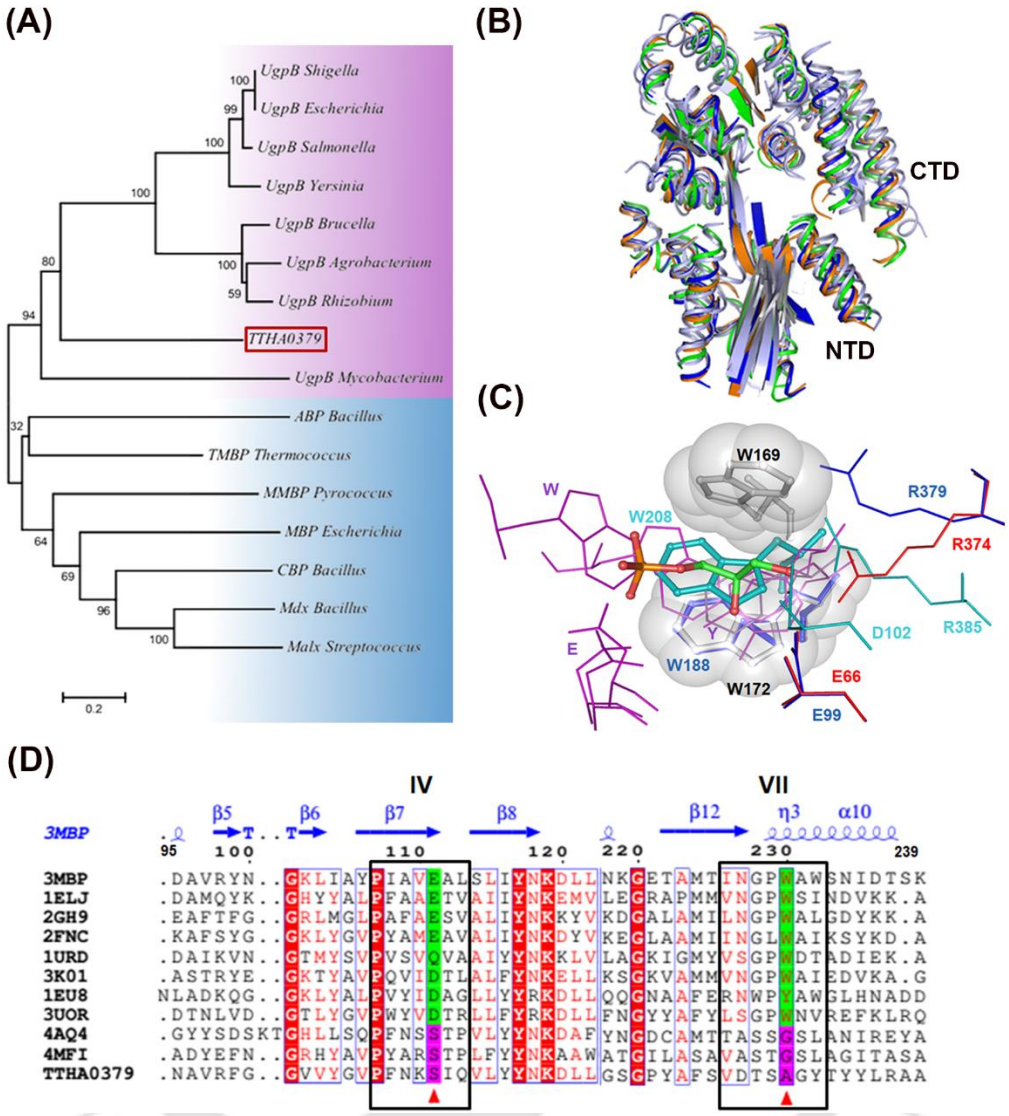


Figure 6.4. **Comparison between proteins TTHA0379, UgpB and sugar-binding protein.** (A) Phylogenetic tree showing the evolutionary relationship among TTHA0379, UgpB and sugar binding proteins. The phylogenetic tree was constructed for a set of 15 protein sequences of UgpB (Q83PU5: *Shigella flexneri*, P0AG80: *Escherichia coli*, Q7CPK0: *Salmonella typhimurium*, Q1CNC9: *Yersinia pestis*, Q8UB32: *Agrobacterium fabrum*, Q926H8: *Rhizobium meliloti* and Q8YCA7: *Brucella melitensis*, A5U6I5: *Mycobacterium tuberculosis*) and sugar binding proteins (P94528: *Bacillus subtilis*, Q7LYW7: *Thermococcus litoralis*, P58300: *Pyrococcus furiosus*, P0AEY0: *Escherichia coli*, O07009: *Bacillus subtilis*, O06989: *Bacillus subtilis* and P59213: *Streptococcus pneumoniae*) from various organisms along with the protein TTHA0379 (Q5SLB4: *Thermus thermophilus*). The two major clusters UgpB and sugar-binding proteins of the phylogenetic tree are highlighted in purple and blue, respectively. The protein TTHA0379 is enclosed in a red box. Significance of each cluster was evaluated by 1000 bootstrap replicates. (B) Structural superimposition of modeled protein TTHA0379 (blue), *Ec*UgpB (green), *Mt*UgpB (orange) and sugar-binding proteins (grey) such as trehalose/maltose binding proteins (PDB ID: 1EU8), sugar binding protein (PDB ID:

3UOR) and ABC-type sugar transporter (PDB ID: 4QRZ). The N- and C-terminal domains of the proteins are represented by 'NTD' and 'CTD', respectively. The highest conservation among superimposed structures is found in the β-sheet of N-domain and hinge region. (C) Structural overlay of ligand binding pockets of TTHA0379 (blue), *Ec*UgpB (red), *Mt*UgpB (cyan) and sugar-binding proteins (violet). The G3P molecule bound to *Ec*UgpB is depicted as stick in green. The two tryptophan residues of *Ec*UgpB responsible for creating space for G3P binding are shown as spheres, while those of *Mt*UgpB and TTHA0379, which contain only one tryptophan residue, are shown as stick in blue and cyan, respectively. (D) Structure-based sequence alignment of TTHA0379, UgpB and sugar-binding proteins. Two unique conserved regions (acidic, IV and tryptophan rich, VII) of sugar-binding proteins are enclosed in black box. The conserved residues of regions IV and VII are enclosed in green box, absence of which in UgpB proteins is highlighted in magenta box.

6.3.2 Resemblance between protein TTHA0379 and GPC-binding UgpB protein

To further substantiate the conservation of ligand binding residues in protein TTHA0379 similar to that of protein UgpB, analysis of the amino acid sequences of these proteins were performed. We first identified the amino acid residues in TTHA0379 required for binding G3P considering *Ec*UgpB as the reference protein. The MSA analysis of these proteins exhibits that the residues Tyr42 & Tyr323 and Glu66 & Arg374, required for anchoring the phosphate and hydroxyl groups of G3P in *Ec*UgpB, are also present in protein TTHA0379 (Figure 6.5A). However, out of two tryptophan residues (Trp169 and Trp172) present in the vicinity of *Ec*UgpB active site, only one is conserved in TTHA0379. As reported earlier, in *Mt*UgpB protein also, all the G3P binding residues are conserved and lack only one tryptophan residue and thus it binds to GPC instead of G3P (Figure 6.5A) (Jiang et al., 2014). Presumably, it can be speculated that protein TTHA0379 might bind to GPC, much like *Mt*UgpB.

Further a comparison of active site reveals that, though the position and orientation of active site residues in all three proteins *Ec*UgpB, *Mt*UgpB and protein TTHA0379 are similar; the absence of a tryptophan in *Mt*UgpB and protein TTHA0379 provides more space to accommodate a larger ligand such as GPC. To further corroborate this hypothesis, we calculated the volume and active site pocket area of these proteins expecting that G3P-binding protein will have smaller volume and area than that of GPC-binding protein. As anticipated, the active site surface area (volume) of G3P-and GPC-binding proteins is 466 $Å^2$ (593 $Å^3$) and 1317 $Å^2$ (2940 $Å^3$), respectively (Figure 6.5B-6.5D). A similar analysis of

protein TTHA0379 displays that its active site surface area and volume are 1326 $Å^2$ and 2629 $Å^3$, respectively (Figure 6.5B-6.5D). This observation affirms that protein TTHA0379 is a most likely a GPC-binding UgpB protein.

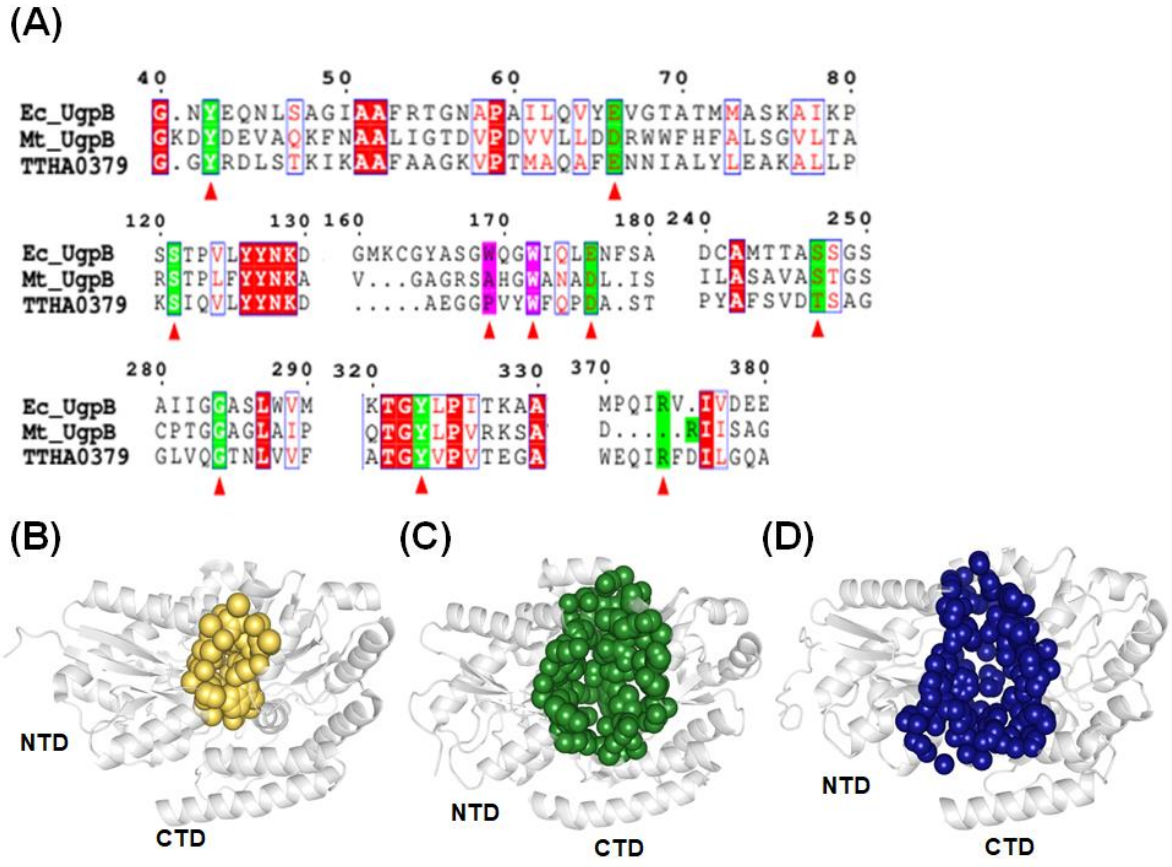


Figure 6.5. Active site pocket analysis of proteins TTHA0379 and UgpB. (A) Multiple sequence alignment of TTHA0379 (UniProt ID: Q5SLB4) along with *Ec*UgpB (UniProt ID: P0AG80) and *Mt*UgpB (UniProt ID: A5U6I5). Only partial MSA having preferred conserved region has been shown in the figure for clarity. The amino acid residues conserved across these three proteins are highlighted in red blocks. The residues essential for ligand binding are highlighted in green blocks and marked with red triangles at the bottom of the alignment. (B-D) Active site volume and area. A pictorial representation of the active site volume and area of *Ec*UgpB (G3P binding protein)**,** *Mt*UgpB (GPC binding protein) and protein TTHA0379, respectively. The active sites of *Ec*UgpB, *Mt*UgpB and TTHA0379 are filled with spheres in yellow, green and blue, respectively.

6.3.3 Thermodynamically, the protein TTHA0379 displays specificity towards GPC

According to *in silico* study, the protein TTHA0379 was proposed to bind a phospholipid precursor GPC based on its sequence similarity with that of the GPC-binding UgpB protein from *Mycobacterium tuberculosis* (*Mt*UgpB, ORF ID: Rv2833c). To further validate the proposal experimentally and to identify its cognate ligand(s), the purified protein TTHA0379 was screened for GPC, G3P and sugar binding through isothermal titration calorimetry (ITC). The result confirms the binding of GPC to the protein TTHA0379, however, no binding to the G3P and sugar molecules (Figure 6.6). Binding of GPC was optimized at different concentration of GPC ranging from 10 to 74 times higher than protein concentration, however best fit was observed at 37 times higher concentration of GPC. The dissociation constant (K_d) of GPC for the protein TTHA0379 is 202 µM (Table 6.7), which is in a comparable range of the reported values for the protein UgpB from *Escherichia coli* (*Ec*UgpB) and *M. tuberculosis* (*Mt*UgpB) (Wuttge et al., 2012; Jiang et al., 2014). The binding kinetics of GPC to the protein TTHA0379 is majorly driven by a favorable enthalpy change (Table 6.7).

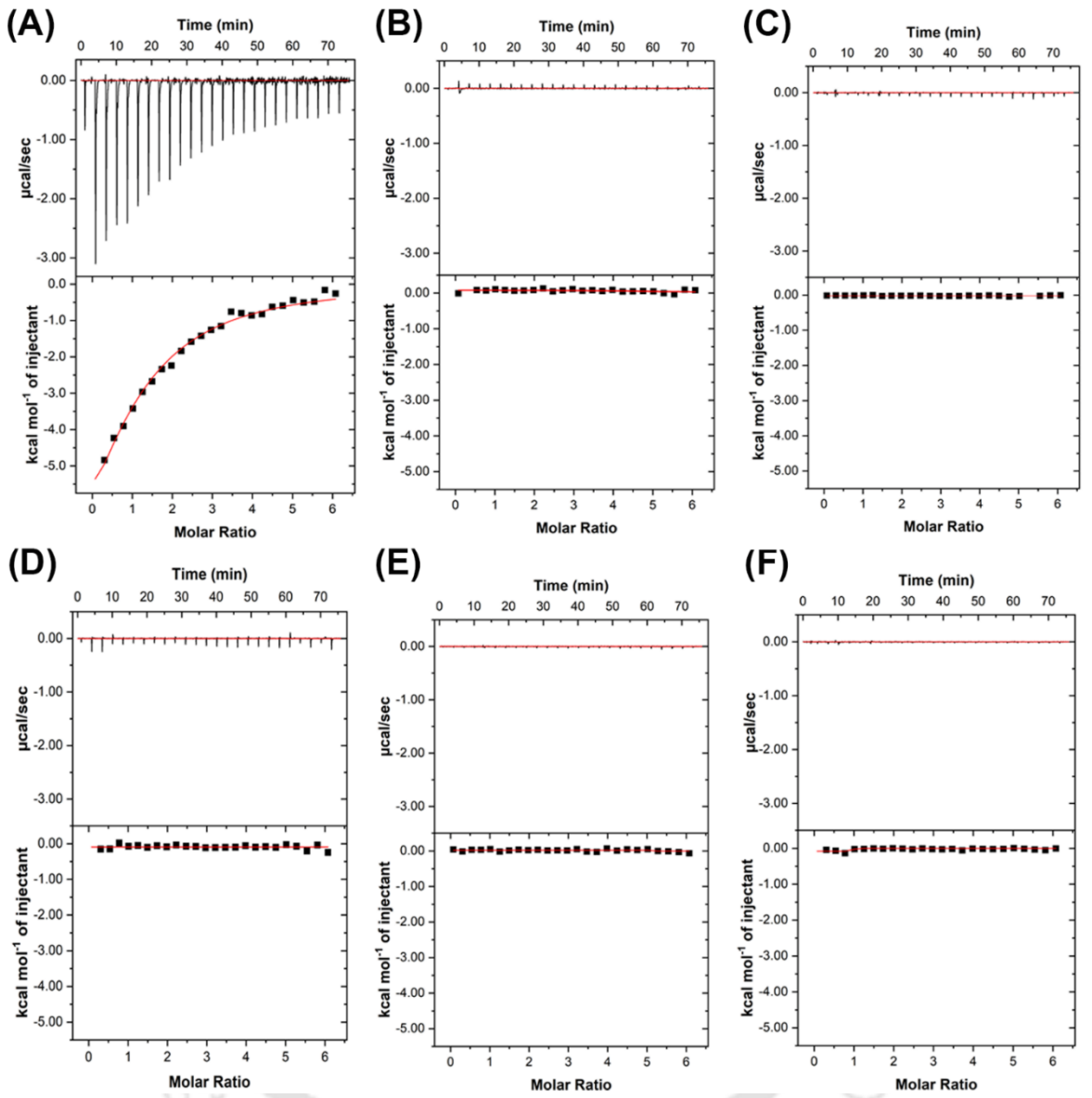


Figure 6.6. Representative isothermal titration calorimetry data of protein TTHA0379 ligand binding. Binding kinetics of the protein TTHA0379 titrated with (A) *sn*-glycerophosphocholine (GPC), (B) *sn*-glycerol-3-phosphate (G3P), (C) Maltose (MAL), (D) Maltotriose (MLR), (E) Trehalose (TRE) and (F) Glucose (GLC). The upper panel displays the heat changes (black lines) upon ligand titrations while the lower panel shows the integrated heat of reactions. All integrated heat pulses are represented with square dots and the one-site binding model fitting is represented with red solid lines. All data are normalized and subtracted with the heat of dilution.

Table 6.7. Thermodynamic parameters of ligand binding to the protein TTHA0379 estimated through isothermal titration calorimetry (ITC).

Protein (135 µM)	Ligand (5 mM)	Stoichiometry ratio (n)	Association (*Ka*, M) / Dissociation (K_d, mM) constant	Δ*H*	*T*Δ*S*	Δ*G*
				(kcal mol^{-1})		
TTHA0379	GPC*	1[a]	$4.93x10^3$/ 0.202	-13.65	-8.61	-5.04
	G3P	N.D.	N.D.	N.D.	N.D.	N.D.
	MAL	N.D.	N.D.	N.D.	N.D.	N.D.
	MLR	N.D.	N.D.	N.D.	N.D.	N.D.
	TRE	N.D.	N.D.	N.D.	N.D.	N.D.
	GLC	N.D.	N.D.	N.D.	N.D.	N.D.

[a]Fixed binding stoichiometry.
Abbreviation: GPC, *sn*-glycerophosphocholine; G3P, *sn*-glycerol-3-phosphate; MAL, Maltose; MLR, Maltotriose; TRE, Trehalose; GLC, Glucose; N.D., Not detected.
*Reproducibility of GPC binding is observed at different conc. of GPC as well as with different batches of purified protein.

6.3.4 The protein TTHA0379 is endogenously bound to a dinucleotide

To further affirm the binding of GPC to the protein TTHA0379, it was crystallized in the presence and absence of GPC. In both cases, crystal structure was determined to a high resolution (of 2.15 and 1.8 Å, respectively) in the space group $P2_1$ containing two protomers in the asymmetric unit. The crystal structure of the protein TTHA0379 adopts the α/β topology and akin to other SBPs, contains two N- and C-terminal domains (NTD and CTD) linked by a hinge region comprised of hinge 1 (H1), hinge 2 (H2) and hinge 3 (H3) (Figure 6.7A). The NTD is composed of two subdomains (N1, residues 1–121 and N2, residues 276–345) and possesses a four β-stranded sheet flanked by nine α-helices. Similarly, the CTD is made up of two subdomains (C1, residues 135–258 and C2, residues 356–395) constituted of a four β-stranded sheet sandwiched by ten α-helices (Figure 6.7B). To our surprise, a strong electron density for a molecule similar to a dinucleotide, uridylyl-3'-5'-phospho-guanosine (U3G) could be observed at the active site of the protein TTHA0379 in both crystal conditions i.e. in the presence and absence of GPC (Figure 6.7C). Thus, the molecule dinucleotide uridylyl-3'-5'-phospho-guanosine (U3G) was fitted into the electron density after building all the protein and water molecules (Figure 6.7C). This endogenous binding of molecules U3G with protein was observed in a consistent way in every batch of purified protein. Thus, owing to 100% occupancy of of U3G in the active site of protein TTHA0379, we referred the molecule U3G as a cognate ligand and protein TTHA0379 to as a U3G-binding protein (U3GBP).

Interestingly, the protein U3GBP packs in three different ways in crystals (referred to as a U3GBP_WT-Form I, -Form II and -Form III), in the space group $P2_1$ (Table 6.2). In U3GBP-Form I, two protomers of the asymmetric unit are arranged in antiparallel direction to the a- and c-axes (Figure D.1). In contrast, in U3GBP_WT-Form II and -Form III, two protomers are arranged in the antiparallel manner facing in an opposite direction (Figure D.1). However, cell parameters of these two forms U3GBP_WT-Form II (a=66.17, b=57.52, c=121.13, $\alpha=\gamma$=90, β=94.99) and U3GBP_WT-Form III (a=65.91, b=58.12, c=102.40, $\alpha=\gamma$=90, β=93.04) are different (Table 6.2). In addition, inter-molecular contacts between the two protomers are different in three crystal forms. However, no significant structural differences could be observed among these forms (Figure D.2). Thus, the structure of only form (i.e. U3GBP_WT-Form I) was considered for further analysis.

A search for the structural homologs of U3GBP using the web server Dali (Holm and Rosenström, 2010) reveals that it shares the highest structural similarity with the members of the (sub)cluster D SBPs including UgpBs, *Mt*UgpB (apo and GPC-bound, PDB IDs: 4MFI/6R1B, RMSD: 2.7/2.2 Å, Z-score: 42/43.5) and *Ec*UgpB (G3P-bound, PDB ID: 4AQ4, RMSD: 2.7 Å, Z-score: 40.5) and sugar-binding proteins, trehalose/maltose-binding protein (TMBP, PDB ID: 1EU8, RMSD: 2.6 Å, Z-score: 41.1), maltose-binding protein MalE3 (MBP, PDB ID: 6DTQ, RMSD: 2.6 Å, Z-score: 41) and acarbose/maltose-binding protein GacH (PDB ID: 3K00, RMSD: 2.5 Å, Z-score: 40.7). A structural comparison of U3GBP with these structural homologs reveals that the protein U3GBP contains an additional N-terminal helix (α1) (Figure 6.7D). This helix is absent in its closest homologs *Mt*UgpB (PDB ID: 4MFI) and *Ec*UgpB (PDB ID: 4QA4) (Figure 6.7E and 6.7F). Further, a structural comparison of U3GBP with other members of the (sub)cluster D SBPs shows that the N1 subdomain is conserved across all the members of the (sub)clusters D, D-I, -II, -III-a and -III-b SBPs, however, they lack the N-terminal helix (α1). Interestingly, the substrate-binding protein (PDB ID: 3C9H) from the subcluster D-IIIa contains an additional N-terminal β-strand (β1) (Figure 6.7G). All other members of the (sub)clusters D, D-I, -II, -III-a and -III-b SBPs lack both these additional N-terminal helix (α1) and strand (β1) (Figure 6.7H and Table D.1).

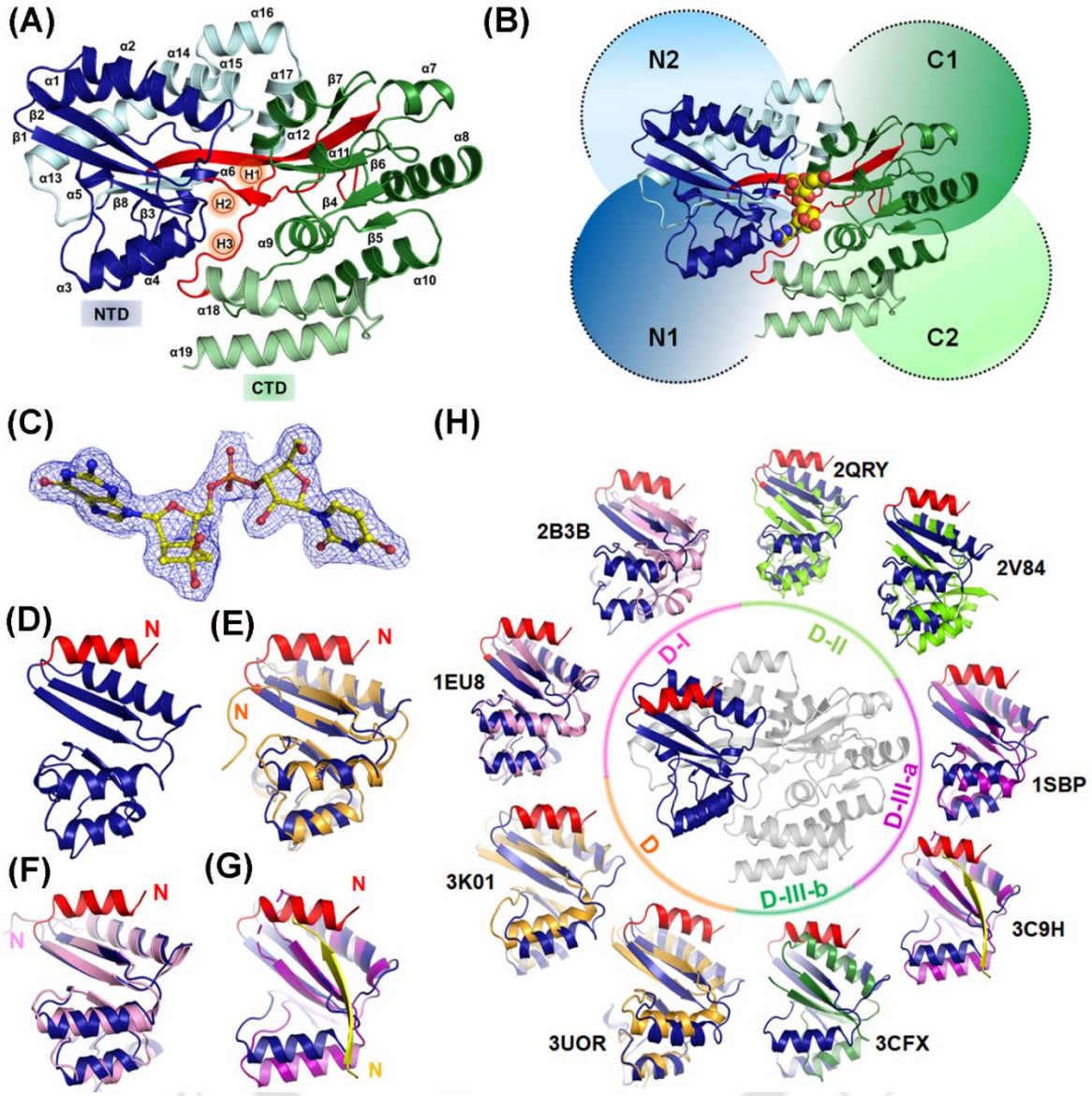


Figure 6.7. Structural details of the protein U3GBP. (A) The overall structure of protein U3GBP with NTD, CTD and hinge region shown in blue, green and red, respectively. (B) The organization of the N1, N2, C1 and C2 subdomains of U3GBP has been displayed in blue, light blue, green and light green backgrounds, respectively. The endogenously-bound molecule U3G has been represented as yellow spheres. (C) The *2Fo−Fc* electron density map contoured at 1.0σ for the molecule U3G is delineated as a blue mesh. (D) The secondary structural content of the N1 subdomain (blue) with an additional N-terminal helix (α1, red) and its comparison with that of (E) *Mt*UgpB (orange, PDB ID: 4MFI), (F) *Ec*UgpB (pink, PDB ID: 4AQ4) and (G) substrate-binding protein (violet, PDB ID: 3C9H). (H) Superimposition of the N1 subdomain (blue) of U3GBP with that of other SBPs from the (sub)cluster D (orange), D-I (pink), D-II (light green), D-III-a (violet) and D-III-b (green). The additional N-terminal helix (α1) of U3GBP is shown in red while the (sub)clusters are

labeled and represented in different colors. Each member from (sub)clusters D, D-I, -II, -III-a and -III-b SBP is denoted by their respective PDB ids (see Table D.1 for details).

6.3.5 Architecture of U3GBP binding pocket

A huge electron density observed at the active site of the protein was fitted with a dinucleotide uridylyl-3'-5'-phospho-guanosine (U3G) containing two ribose sugars, a phosphate group and two nitrogenous bases (uracil and guanine) (Figure 6.8A). Among the two ribose sugars, one is attached to uracil base as the nucleoside uridine while second sugar seems to be modified, which could not be identified. Thus, we could fit a derivative of U3G as an unknown molecule (Figure 6.8A). A search against small molecule databases such as Protein Data Bank (Berman et al., 2000), ZINC (Irwin and Shoichet, 2005), *E. coli* Metabolome Database ECMDB 2.0 (Guo et al., 2012) and PubChem (Kim et al., 2019) could not yield any molecule like this.

As identified ligand is unknown natural ligand and possess a high tendency to endogenously bind with the protein U3GBP, it restricted our attempt to determine the U3G binding kinetics using other biophysical methods. Hence, based on the crystal structures, binding features of U3G with the protein U3GBP was estimated. One molecule of U3G is bound to the active site of each protomer of the asymmetric unit. Akin to other substrates of SBPs (de Boer et al., 2019b), U3G is also bound at the hinge region sandwiched between the NTD and CTD. The nitrogenous bases of the molecule U3G is coordinated by the residues Asn81, Asn226 and Glu357 through hydrogen bonds (Figure 6.8B and 6.8C). Similarly, the phosphate group of U3G is anchored by the residues Tyr56, Ser127, Thr240, Gly275 and Tyr314 and the ribose sugar is stabilized by water-mediated interactions with the protein (Figure 6.8B). The residue Arg360 plays a pivotal role in stabilizing the guanine base by forming planar cation–π stacking interaction (Figure 6.8D). Interestingly, out of four aromatic residues in the vicinity of the active site, residues Tyr224 and Tyr246 sandwich the uracil base by forming π–π stacking interactions while residue Phe79 forms π–π stacking interaction with the guanine base (Figure 6.8E). The fourth aromatic residue Phe361 forms CH–π stacking interaction in a perpendicular orientation to the guanine base (Figure 6.8F).

To further probe the significance of these residues in holding the molecule U3G at the active site of the protein, the residue Tyr56 was mutated into the residue phenylalanine while other residues Phe79, Asn81, Ser127, Tyr224, Thr240 and Tyr246 were mutated into the residue alanine. In addition, the residue Gln274, which was suggested in our previous study (Chandravanshi et al., 2016) to play a significant role in the binding of the molecule GPC to the protein, was also mutated into the residue alanine. To assess the effect of these mutations on the binding of the molecule U3G, these mutants were crystallized and their structures were determined to high resolution (Table 6.2-6.6). Akin to U3GBP_WT structures, U3GBP mutants also exhibited the variation for unit cell parameters and thus were categorized into different forms. To our surprise, none of these mutations could hinder the binding of the molecule U3G to the protein U3GBP. In each mutant, the molecule U3G was observed to be endogenously bound at the active site of the protein U3GBP with all the conserved interactions, except those which lost due to the mutation of the residue (Table D.2-D.6).

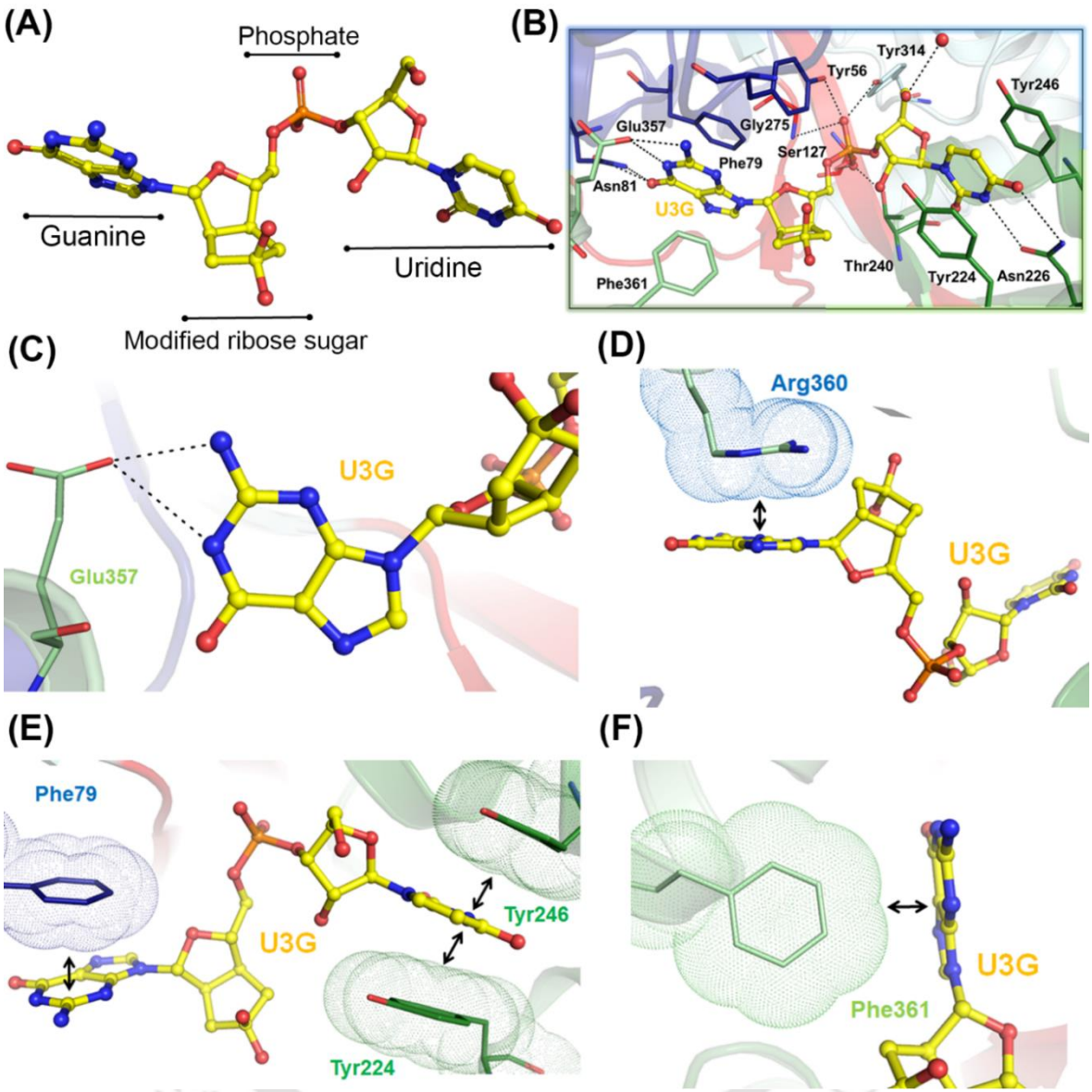


Figure 6.8. Active site of U3GBP. (A) The fitted U3G molecule at the active site of the protein U3GBP is shown as ball-and-stick model. (B) The residues interacting with the molecule U3G (yellow ball-and-stick) from the N1, N2, C1 and C2 subdomains are depicted as lines in blue, light blue, green and light green, respectively. The polar interactions are represented as black-dashed lines. (C) The residue Glu357 (green line) interacts with 1-NH and 2-NH_2 atoms of the guanine base of U3G, where hydrogen bonds are shown as black-dashed lines. (D) The cation–π stacking interaction between the residue Arg360 guanidinium group (blue dots) and the guanine base of U3G is marked with a double-headed black arrow. (E) The π–π stacking interaction between aromatic residues (blue and green dots) and U3G is demarcated with a double-headed black arrow. (F) The CH–π stacking interaction between the residue Phe361 (light green dots) and the guanine base of U3G is represented by a double-headed black arrow.

6.3.6 Endogenously-bound U3G adapts a conformation as that of c-di-GMP/AMP

To explore the conformations adapted by the endogenously-bound U3G and that of other dinucleotides bound to their cognate proteins were investigated. Since no structural data of SBPs bound to dinucleotides are available in the literature, the comparison was made with cyclic dinucleotides such as cyclic diguanylate (c-di-GMP) and diadenylate (c-di-AMP). As per reports, c-di-GMP/AMP adapts four different conformations depending upon their receptor proteins (He et al., 2020). These include U-type, in which both nitrogenous bases are oriented in a parallel way (c-di-GMP bound, PDB ID: 2RDE; c-di-AMP bound, PDB ID: 4XTT), V-type, where both nitrogenous bases point in the outward direction in such way that it maintains approximately 45° angle between the nitrogenous bases (c-di-GMP bound, PDB ID: 4F3H; c-di-AMP bound, PDB ID: 4QLN), E-type, an extended and flat conformation where both nitrogenous bases arranged in an opposite directions (c-di-GMP bound, PDB ID: 3GFX; c-di-AMP bound, PDB ID: 5UXF) and O-type, an orthogonal conformation in which nitrogenous bases are orthogonally arranged to the each other with central cyclic conformation (c-di-AMP bound, PDB ID: 5KS7) (Figure 6.9A-6.9D). Comparison of U3G conformation with these four conformations reveals that U3G conformation resembles with V-shape, indicating that it adapts V-type conformation like c-di-GMP/AMP. In the V-type, the uracil base and modified ribose sugar point in the same direction. Interestingly, within this V-type conformation of U3G, a water molecule (W296) forms water-mediated interactions between two ribose sugars of U3G and forms the cyclic organization. Due to this cyclic organization, adaption to O-type conformation of molecule U3G in the active site U3GBP is also identified (Figure 6.9E and 6.9F). Interestingly, even in the structures of U3GBP mutants (U3GBP_Y56F, U3GBP_F79A, U3GBP_N81A, U3GBP_S127A, U3GBP_Y224A, U3GBP_T240A, U3GBP_Y246A and U3GBP_Q274A), the molecule U3G maintains the V-type conformation, where it also adapts the O-type with conserved water molecules (Figure D.3).

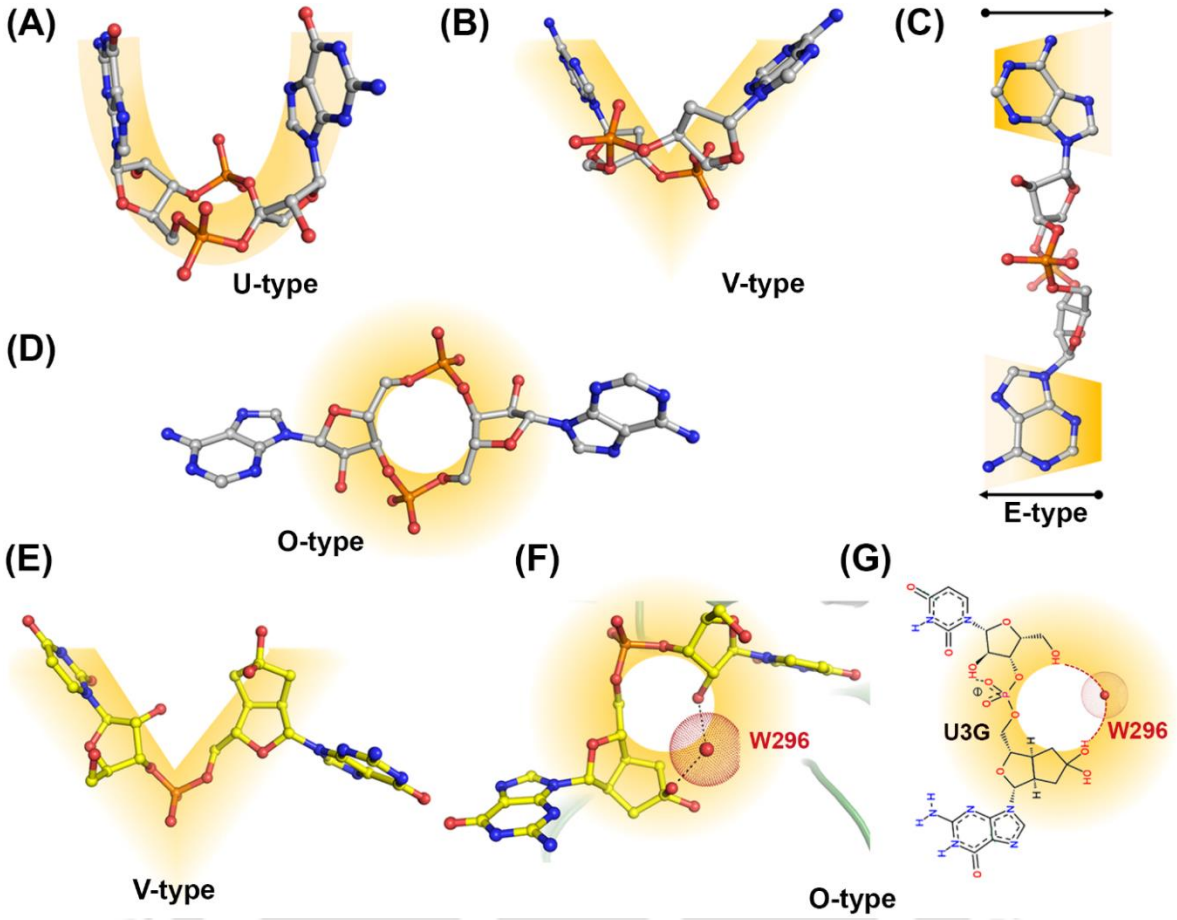


Figure 6.9. Different conformations adapted by c-di-GMP/AMP and U3G. Four different conformations adapted by c-di-GMP/AMP are (A) U-type (PDB ID: 2RDE), (B) V-type (PDB ID: 4QLN), (C) E-type (PDB ID: 5UXF) and (D) O-type (PDB ID: 5KS7) (He et al., 2020). Similarly, two conformations of U3G are (E) V-type and (F) O-type (this study). The different shapes are depicted with yellow background where the molecules c-di-GMP/AMP and U3G are represented as ball-and-stick model in grey and yellow, respectively. (G) For the clarity of the figure, a two-dimensional O-type conformation of U3G has been displayed, where the water molecule (W296, red sphere) forms hydrogen bonds (red-dotted lines) with the ribose sugars of U3G.

6.3.7 U3GBP is a new member of the subcluster D-I SBPs

To classify protein U3GBP, a comparative structural analysis with other SBPs was performed which reveals that protein U3GBP possesses novel structural features and a cognate ligand not reported till date. Despite a low sequence identity with its homologs, U3GBP shares the highest structural similarity with (sub)cluster D SBPs. A structure-based evolutionary analysis of U3GBP and its homologs unveils that the protein U3GBP belongs to the subcluster D-I

containing protein UgpB (*Ec*UgpB, PDB ID: 4AQ4) and sugar-binding protein (TMBP, PDB ID: 1EU8) (Figure 6.10A). However, interestingly, a sequence-based evolutionary analysis including these proteins clusters the protein U3GBP along with a molybdate-binding protein (ModA, PDB ID: 2H5Y) (Figure 6.10B and Table D.7).

Further, an investigation of the ligand binding pocket of U3GBP and closest homolog identified by the web server Dali reveals that irrespective of the ligand types, all the proteins viz. UgpB (*Mt*UgpB and *Ec*UgpB), sugar-binding protein (maltose-binding protein MalE or MBP) and U3GBP possess similar active site architecture and conserve the four subsites (A, B, C and D) (Figure 6.11A-6.11D). Moreover, an electrostatic surface charge distribution of these proteins demonstrates that the proteins *Mt*UgpB (GPC-bound, PDB ID: 6R1B), *Ec*UgpB (G3P bound, PDB ID: 4AQ4) and MBP (maltose-bound, PDB ID: 6DTQ) possess a negatively-charged active site regardless of their negatively-charged substrates (Figure 6.11E-6.11G). On other hand, the protein U3GBP comprises a highly positively charged surface favoring the binding of the negatively-charged U3G (Figure 6.11H).

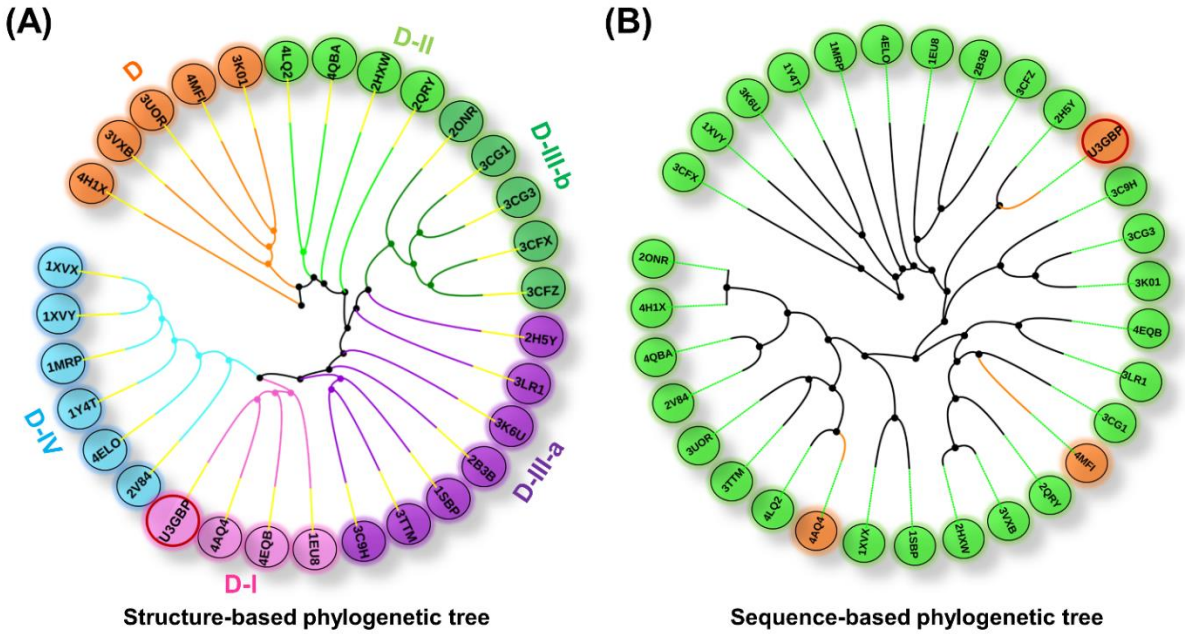


Figure 6.10. Evolutionary relationship of U3GBP with the (sub)cluster D SBPs. A total of 30 SBPs from different (sub)clusters D belonging to different bacterial genera have been chosen for classification based on structure and sequence similarity. (A) Structure-based phylogeny tree depicting the clustering of U3GBP with the subcluster D-I SBPs. (B) Phylogeny tree depicting an evolutionary relatedness of U3GBP and UgpBs (orange) with

other (sub)cluster D (green) SBPs using sequence similarities. Each SBP is denoted by their respective PDB ids (see Table D.7 for details).

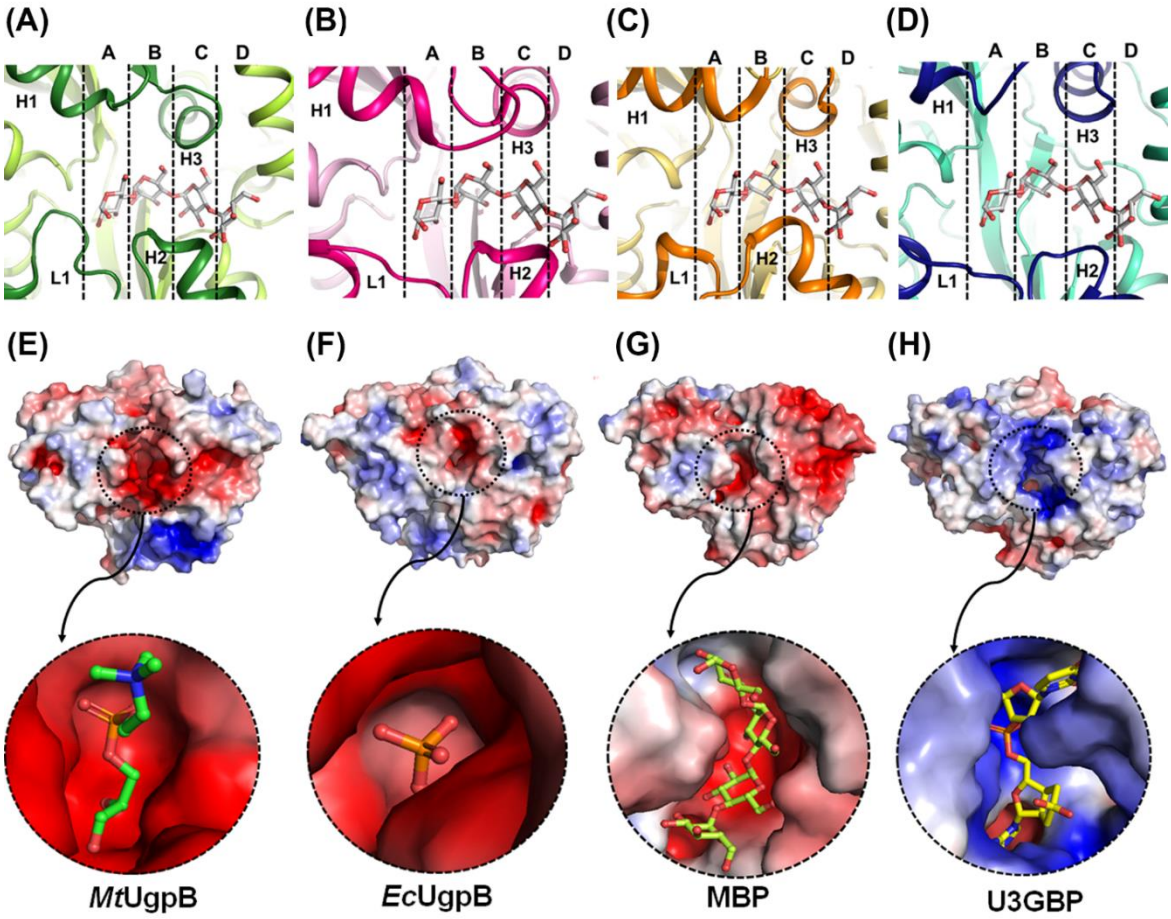


Figure 6.11. Active site architecture of UgpBs, MBP and U3GBP. (A-D) The subsite comparison of *Mt*UgpB (green, PDB ID: 4MFI), *Ec*UgpB (pink, PDB ID: 4AQ4), MBP (orange, PDB ID: 6DTQ) and U3GBP (blue, PDB ID: 7C0F). The secondary structural elements helices (H1, H2 and H3) and loop (L1) are shown in dark green, dark pink, dark orange and dark blue, respectively. The four subsites (A, B, C and D) of these proteins are separated by vertical dotted lines. The availability of these four subsites in *Mt*UgpB, *Ec*UgpB and U3GBP is depicted by overlaying the sugar-binding protein MalE (MBP, PDB ID: 6DTQ). (E-H) Electrostatic surface charge distribution of *Mt*UgpB (PDB ID: 6R1B), *Ec*UgpB, MBP and U3GBP. The calculated electrostatic potential of each protein is color-coded from blue (positive) to red (negative) with a scale of -1 to +1 kcal mol^{-1}, respectively. The ligands bound at the active site (zoomed in) of the protein are shown as ball-and stick-model.

6.3.8 Binding of U3G at the N-terminal domain brings the domain closure

Although the proteins *Mt*UgpB, *Ec*UgpB and U3GBP show differences for the cognate ligand, however, the ligand binding site and in particular the phosphate group of GPC, G3P and U3G, respectively, occupy the same spatial position (Figure D.4A). Also, the residues Tyr56, Ser127, Thr240 and Tyr314 of U3GBP, which coordinate the phosphate group of U3G, are conserved in the UgpBs (Figure D.4A). This suggests that regardless of different ligands, the binding mode of these proteins to the cognate ligand is preserved. Hence, to further probe the role of these conserved residues, mutants Y56F, S127A and T240A of U3GBP were generated. Surprisingly, akin to U3GBP_WT, the mutants U3GBP_Y56F, U3GBP_S127A and U3GBP_T240A are also observed to be endogenously bound the molecule U3G in the active site (Figure D.4B-D.4D). This indicates that the abolishment of hydrogen bonds with phosphate group is not enough to hinder the binding of the molecule U3G. Possibly, the π-π stacking interaction created by the residues Tyr224 and Tyr246 from CTD is dominant to the higher affinity for the molecule U3G. Thus, it can be hypothesized that the phosphate group initiates the binding at the N-terminal while the aromatic residues at C-terminal stabilize the ligand at the active site and originate the domains closure.

To further establish the role of CTD in the domain closure, both the aromatic residues Tyr224 and Tyr246 were mutated to alanine. Using circular dichroism (CD) spectral analysis, the stability of these mutants as a function of melting temperature (Tm) was compared with that of the wild type U3GBP_WT. The results exhibit that the melting temperatures (Tm) of U3GBP_WT and single point mutants (U3GBP_Y224A and U3GBP_Y246A) are similar (~120ºC) (Figure 6.12A-6.12C). However, a double mutant (U3GBP_Y224A/Y246A) shows a modest decrease in the melting temperature (Tm) of 110ºC (Figure 6.12D). Thus, it can be anticipated that the double mutant (U3GBP_Y224A/Y246A) of the protein might have hindered the binding of U3G. To affirm this further, the double mutant protein (U3GBP_Y224A/Y246A) was crystallized and its tertiary structure was elucidated. To our surprise, the structural data demonstrates that the molecule U3G remains bound at the active site of the double mutant protein (U3GBP_Y224A/Y246A). Thus, it can be speculated that the CTD might not alone direct the domain closure of the protein upon ligand binding.

Thus, to find out whether NTD, instead of CTD alone, contributes to the domain closure, a triple mutant (U3GBP_F79A/Y224A/Y246A) was generated. However, due to random errors, residues (74-75 and 50-75) were also deleted and thus two triple mutant constructs (U3GBP_F79A/Y224A/Y246A/Δ74-75 and U3GBP_F79A/Y224A/Y246A/Δ50-75) were generated. Notably, the residues Thr74 and Met75 hold the two subdomains (N1 and N2) of the protein U3GBP_WT (Figure D.5). Thus, due to the deletion of these two residues, the triple mutant U3GBP_F79A/Y224A/Y246A/Δ74-75 exhibits a reduced thermal stability with a melting temperature (Tm) of 90°C (Figure 6.12E). This reduced thermal stability hints that the triple mutant protein might be in apo state due to perturbation in π-π stacking interaction for U3G binding as well as inhibition of NTD domain closure due to loss of the inter-subdomain interaction of N1 and N2. However, lack of the structural details of the triple mutant U3GBP_F79A/Y224A/Y246A/Δ74-75 hinders the insights into the NTD closure mechanism.

In our previous study, we showed that for the protein αGlyBP, an active site residue Asp70 of the N1 subdomain initiates the NTD closure (Chandravanshi et al., 2020). A structural comparison of the proteins αGlyBP and U3GBP reveals that they conserve the N1 subdomain including the N-terminal residues Asp70 and Phe79, respectively, suggesting the role of NTD in the domain closure (Figure D.6). The triple mutant U3GBP_F79A/Y224A/Y246A/Δ50-75, which lacks the residues 50-75 forming the helix α3 of the N1 subdomain, is involved in holding the N2 subdomain of the wild type protein. The deletion of the residues 50-75 was further confirmed by a matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) analysis with a calculated mass of 41.5 kDa compared to that U3GBP_WT (44.0 kDa) (Figure D.7). Moreover, a thermal stability analysis of the triple mutant U3GBP_ F79A/Y224A/Y246A/Δ50-75 exhibits a significant decrease in its melting temperature (Tm, 80°C) compared to that of U3GBP_WT (120°C) and other mutants (Figure 6.12F and 6.12G). Further, to understand the lower thermal stability of the triple mutant U3GBP_F79A/Y224A/Y246A/Δ50-75, the protein was crystallized and its structure was determined. To our delight, the molecule U3G is found to be absent from the active site of the triple mutant U3GBP_F79A/Y224A/Y246A/Δ50-75. As expected, a structural comparison of the both wild type and triple mutant proteins reveals that due to loss of the inter-subdomain

interaction of N1 and N2 subdomains as well as loss of polar interaction between N1 subdomain and molecule U3G, triple mutant is unable to binds the molecule U3G (Figure D.5B-D.5D).

To further investigate the role of NTD in the domain closure of the protein, the web server DynDom (Taylor et al., 2014) was used to evaluate the conformational changes of protein U3GBP (unbound, PDB ID: 7C1B and U3G-bound, PDB ID: 7C0F). The result reveals that the bending of three hinge regions leads to a rigid-body rotation of 21.9° of the NTD while the CTD remains static (Figure 6.13A). This type of NTD-mediated domain closure can also be observed in the proteins *Mt*UgpB (unbound, PDB ID: 4MFI and GPC-bound, PDB ID: 6R1B) and maltose-binding protein MalE (MBP; unbound, PDB ID: 6DTR and maltose-bound, PDB ID: 6DTQ) where the degree of rigid-body rotation of the NTD is 21.8º and 35.8º, respectively (Figure 6.13B and 6.13C).

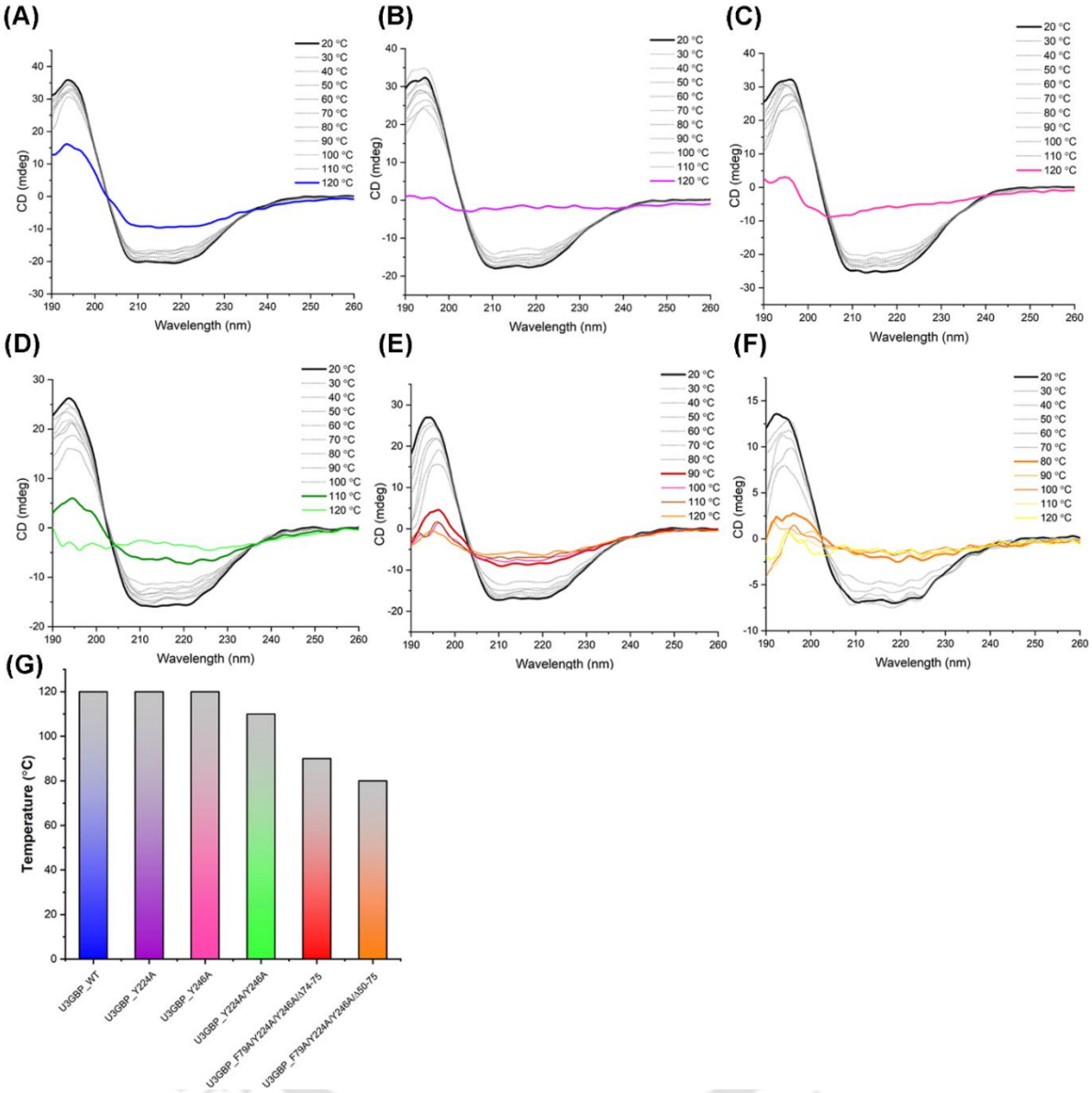


Figure 6.12. Thermal melting plots of the wild type and mutants U3GBP. Circular dichroism (CD) spectra of (A) U3GBP_WT (blue), (B) U3GBP_Y224A (violet), (C) U3GBP_Y246A (pink), (D) U3GBP_Y224A/Y246A (green), (E) U3GBP_F79A/Y224A/Y246A/Δ74-75 (red) and (F) U3GBP_ F79A/Y224A/Y246A/Δ50-75 (yellow). The CD spectra was measured in wavelength range of 260 to 170 nm. The spectra obtained at different temperatures ranging from 20 to 120°C is overlaid all together to identify the melting temperature (Tm) at which the changes in secondary structure content was observed. The initial spectra at 20°C is depicted with black line while the subsequent spectrum is shown as dotted grey lines. A deviation in the spectrum are demarcated with colored lines. (G) A histogram plot of the melting temperature (Tm) of the wild type and mutants U3GBP.

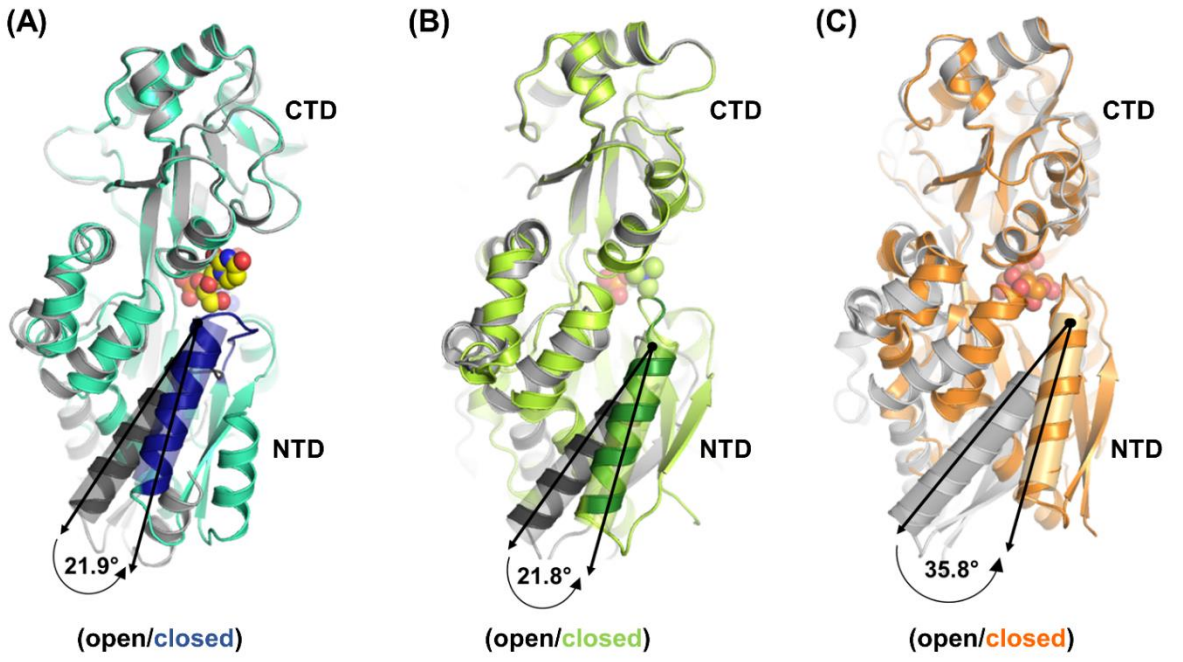


Figure 6.13. Rigid-body movement of N-terminal domain (NTD) upon ligand binding. Superimposition of open structures (grey) of U3GBP (PDB ID: 7C1B), *Mt*UgpB (PDB ID: 4MFI) and MBP (PDB ID: 6DTR) with their respective closed structures of (A) U3GBP (blue, PDB ID: 7C0F), (B) *Mt*UgpB (green, PDB ID: 6R1B) and (C) MBP (orange, PDB ID: 6DTQ). The bound ligands are shown as spheres. The helix, which depicts the rigid-body movement of the NTD, is represented as a cylindrical model.

6.3.9 Functionally-associated genes are involved in tRNA synthesis and/or modification

To understand the rationale for the uptake of U3G through an ABC transporter in *T. thermophilus* HB8, an analysis of the genetic organization of the U3G ABC transporter (ORF IDs: TTHA0376-TTHA0379) was performed. A homology search of the protein U3GBP against the UniProtKB database yields three closest homologs (ORF IDs: Mesil_2057, Mrose_01072 and Ocepr_1138) having a sequence identity greater than 50%. However, the details of the genetic organization of only two proteins (ORF IDs: Mesil_2057 and Ocepr_1138) are available in the Gene database (NCBI) and thus were used for further analysis. An analysis of the genetic operon of U3G ABC transport system reveals that it is functionally associated with a tRNA N^6-threonylcarbamoyladenosine biosynthesis protein TsaB (ORF ID: TTHA0375) (Figure 6.14A). A close strain of *T. thermophilus* HB8, *T.*

thermophilus HB27 also possesses a U3G ABC transport system (ORF IDs: TTC0009-TTC0011) with a similar genetic organization (Figure 6.14B). On the other hand, in *Meiothermus silvanus*, the homolog ABC transport system (ORF IDs: Mesil_2055-Mesil_2057) is preceded by a histidyl-tRNA aminoacylating protein (HisRS, ORF ID: Mesil_2054) (Figure 6.14C). Similarly, in *Oceanithermus profundus*, the homolog ABC transporter (ORF IDs: Ocepr_1136-Ocepr_1138) is preceded by tmRNA binding protein (SmpB, ORF ID: Ocepr_1135) (Figure 6.14D). As U3GBP and its homologs share a similar genetic organization, it can be proposed that the functionally-associated genes of the U3G ABC transport system are involved in tRNA synthesis and/or modification.

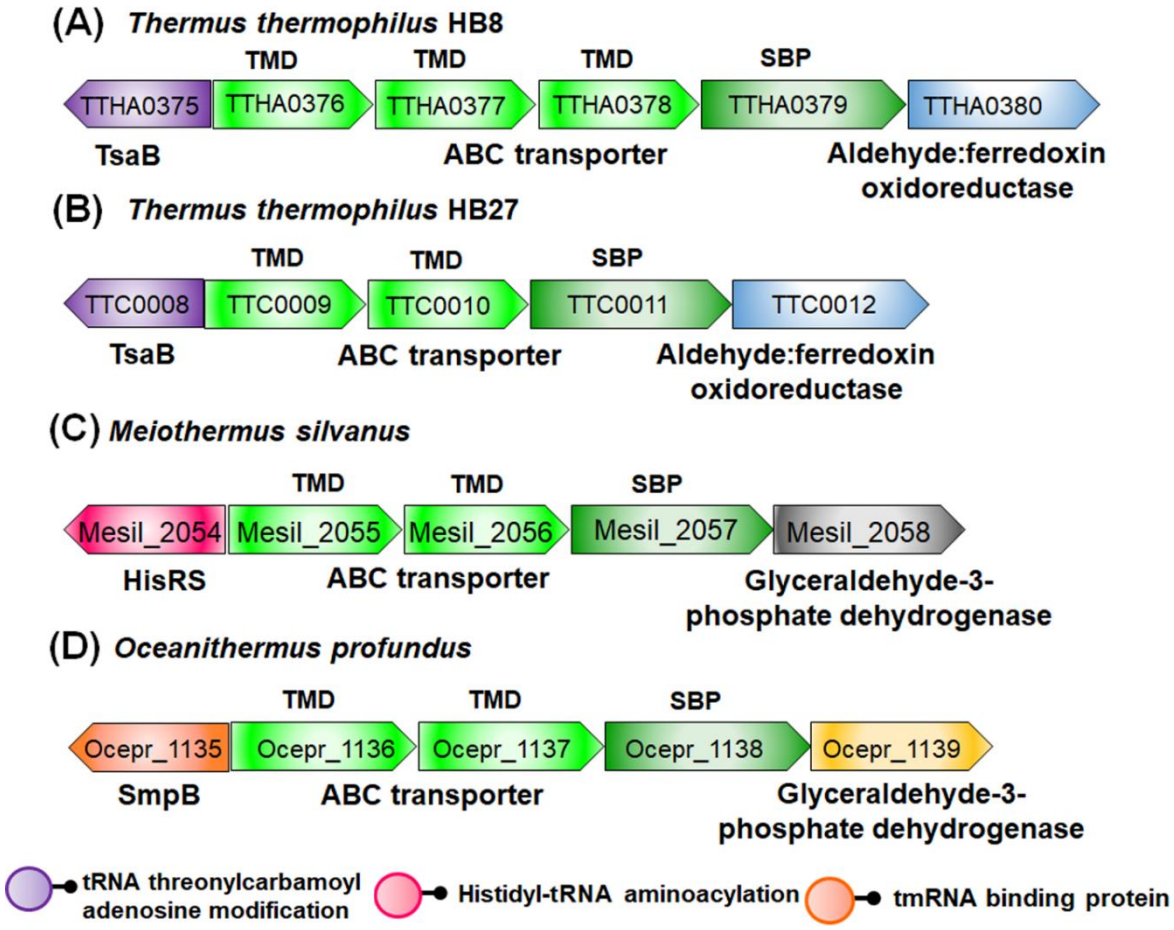


Figure 6.14. The operonic arrangement of the U3G ABC transport system. The ORFs involved in the U3G transport are labeled and represented as an arrow depicting their direction of transcription. The ORFs encoding the components of the ABC transporter depicted in green while the functionally-associated tRNA synthesis and/or modification systems (TsaB, HisRS and SmpB) are in violet, pink and orange, respectively.

6.3.10 The molecule U3G might serve as a signaling molecule for N^6-threonylcarbamoyladenosine biosynthetic pathway

Owing to the conformational similarity of U3G with secondary messenger c-di-GMP/AMP, it is tempting to speculate that U3G also might serve as a signaling molecule for its functionally-associated proteins. To strengthen this, the protein TTHA0375 was further characterized using bioinformatics approaches. The gene *TTHA0375* is located upstream to the U3G ABC transporter (ORF IDs: TTHA0376-TTHA0379) and has been annotated as a "putative glycoprotein endopeptidase" in the UniProtKB database. On the other hand, in the Gene database (NCBI), it is annotated as "tRNA (adenosine(37)-N^6)-threonylcarbamoyl transferase complex dimerization subunit type 1 TsaB". A homology search using the web tool BLAST (Altschul et al., 1990) against the protein data bank (PDB) indicates that the protein TTHA0375 shares the highest similarity with the protein YeaZ (or TsaB) from *Vibrio parahaemolyticus* (*Vp*YeaZ) having a sequence identity (query coverage) of 31(96)%. To further identify the sequence features of the protein YeaZ in TTHA0375, a multiple sequence alignment was performed. The result exhibits that out of the three signature motifs D-X-G, G-X-G and G-P-G-X_2-T-G-X-R, crucial for the nucleotide binding to YeaZ, the protein TTHA0375 contains two motifs D-X-G and G-P-G-X_2-T-G-X-R (where X denotes any amino acid) (Figure 6.15A). This suggests that the protein TTHA0375 might also function as YeaZ (or TsaB). It's to be noted that, considering the *in silico* analysis of protein TTHA0375, we cannot preclude the possibility of other physiological function of protein TTHA0375 based on its crystal structure. However, owing to the conservation for major functional features of YeaZ, protein TTHA0375 seems to be YeaZ (or TsaB)

Physiologically, the protein YeaZ functions as dimer, where the hydrophobic residues Leu40, Leu47, Ile74, Ile78, Leu82 (numbering as per *Vp*YeaZ) play a crucial role in dimeric formation (Aydin et al., 2011). Interestingly, these residues are conserved in the protein TTHA0375 as well, hinting it to form a biological dimer (Figure 6.15B). To locate the spatial positions of these conserved hydrophobic residues, a tertiary structure of the protein TTHA0375 was predicted employing the homology modelling method using the program SWISS-MODEL which used the protein YeaZ (PDB ID: 1OKJ) as the template. The result shows that the predicted model contains two domains (I and II) with βββαβαβα structural fold

(Figure 6.15C). Further, the conserved hydrophobic residues are located at the dimeric interface (Figure 6.15D). To further affirm the function of the protein TTHA0375, a protein-protein interaction network analysis of YeaZ (TsaB), YrdC (TsaC), YgjD (TsaD) and YjeE (TsaE), which are known to be involved the universal tRNA modification system, was performed. The result revels that the proteins involved in the tRNA synthesis and/or modification such as tRNA-specific 2-thiouridylase (ORF ID: TTHA0258), UPF0102 (ORF ID: TTHA0372), YrdC-like domain-containing protein (ORF ID: TTHA0793) PolyA polymerase (ORF ID: TTHA0831), YgjD (or TsaD, ORF ID: TTHA1252) and tRNA(Ile)-lysidine synthase (ORF ID: TTHA1542) are functionally linked to the protein TTHA0375 (Figure 6.16 and Table D.8). Notably, the protein TTHA0375 inter-links the tRNA synthesis and/or modification and U3G ABC transport system (ORF IDs: TTHA0376-TTHA0379) hinting that functional association between them (Figure 6.16 and Table D.8).

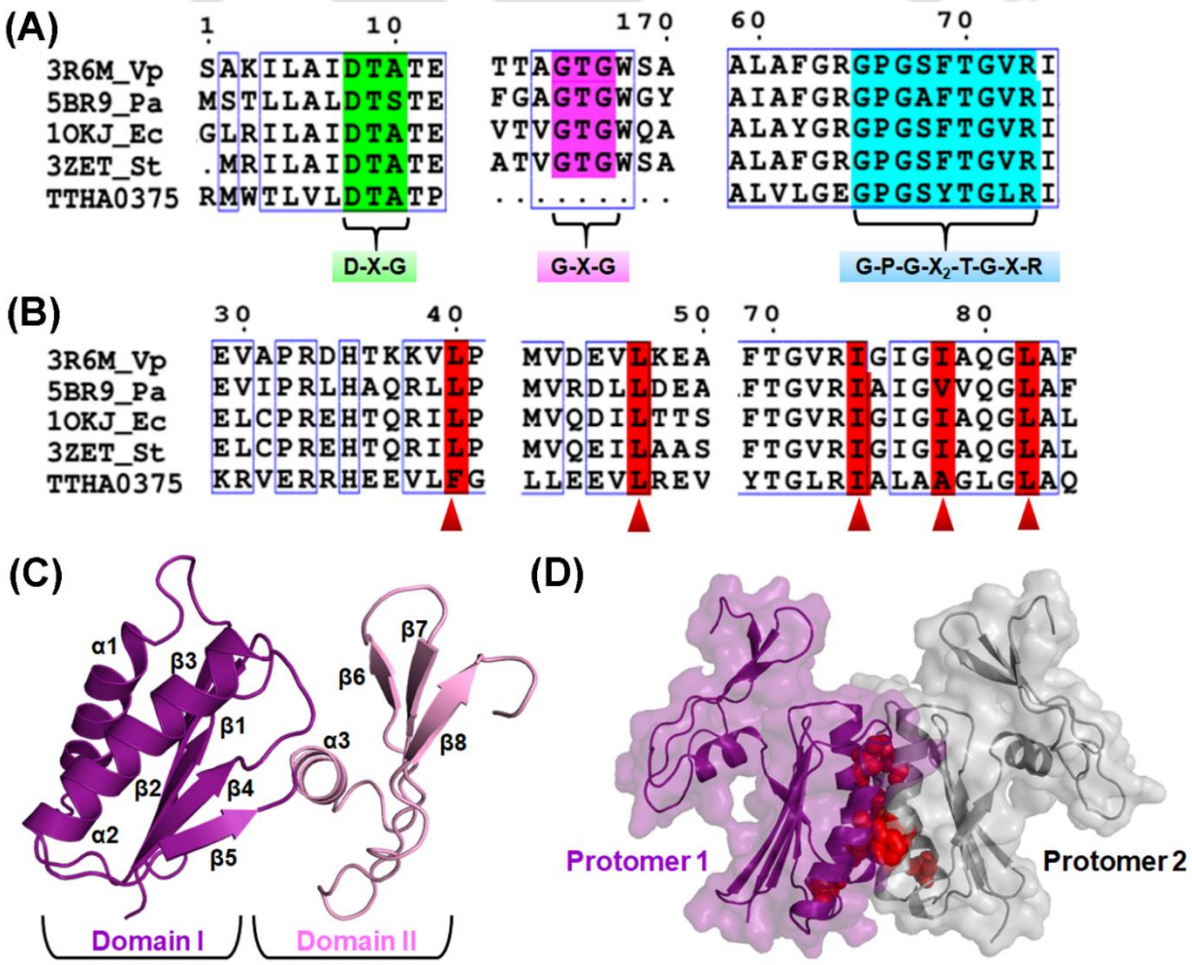


Figure 6.15. Multiple sequence alignment and structure prediction of the protein TTHA0375. (A and B) Multiple sequence alignments of the protein TTHA0375 along with the protein YeaZ from *V. parahaemolyticus* (PDB ID: 3R6M), *Pseudomonas aeruginosa* (PDB ID: 5BR9), *E. coli* (PDB ID: 1OKJ) and *Salmonella typhimurium* (PDB ID: 3ZET). The motifs (D-X-G, G-X-G and G-P-G-X_2-T-G-X-R, where X denotes any amino acid) involved in the nucleotide binding highlighted in green, pink and cyan, respectively. The hydrophobic residues present at the dimeric interface are highlighted in red and marked with upward arrowheads. (C) The predicted model of the protein TTHA0375. (D) Spatial position of the hydrophobic residues located at the dimeric interface are represented as red spheres.

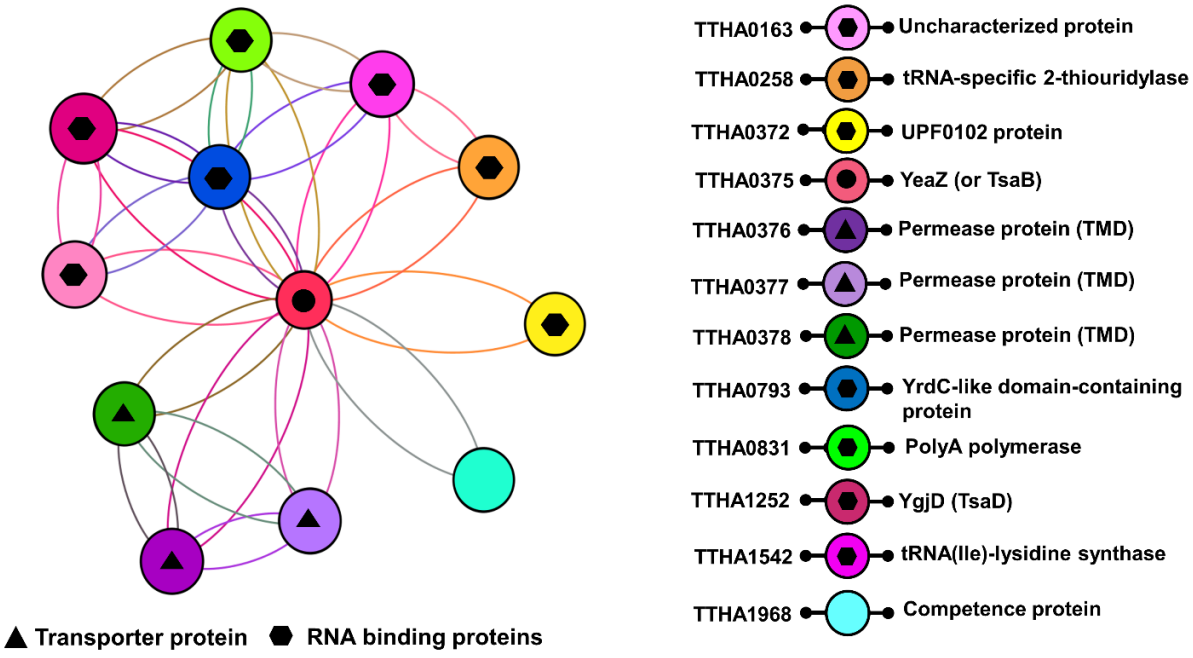


Figure 6.16. Schematic representation of protein interactome map. A protein-protein interaction network of the protein TTHA0375 along with the components of the U3G ABC transport and tRNA synthesis and/or modification systems. Proteins are represented as nodes (circles) and their interactions as solid lines (see Table D.8 for further details).

6.4 DISCUSSION

In this chapter, we performed the structural characterization of the protein TTHA0379 from a thermophilic bacterium *T. thermophilus* HB8. Based on our *in silico* study, the protein TTHA0379 was proposed to function as a substrate (or solute)-binding protein of ABC transporter involved in the uptake of a phospholipid precursor GPC essential for the *de novo*

biosynthesis of phospholipids (Chandravanshi et al., 2016). Thermodynamic data, confirmed the binding of GPC to the protein TTHA0379 while no binding for the G3P and sugars. In contrast, the crystal structure determination of the protein TTHA0379 revealed that it binds to a derivative of uridylyl-3'-5'-phospho-guanosine (U3G) and thus, it was renamed as a U3G-binding protein (U3GBP). Endogenous binding of U3G in the active site confirms that U3GBP is a dinucleotide-binding protein where U3G is a cognate ligand rather than GPC. However, thermodynamic data suggested the binding of GPC to the protein U3GBP, which is endogenously bound to the molecule U3G. Thus, it can be hypothesized that the molecule GPC might bind to a site other than the active site of the protein U3GBP. An attempt to crystalize the protein U3GBP with GPC was unfortunately unsuccessful. It is noteworthy that in our previous *in silico* study, we reported that the molecule GPC binds to the protein U3GBP at a site far from the active site (Chandravanshi et al., 2019). However, further experimental studies are required to validate these results. In summary, it is tempting to propose that GPC is not a physiological ligand for the protein U3GBP.

The overall structure of U3GBP possesses a similar topological fold and architecture as that of other (sub)cluster D SBPs such as UgpBs and sugar-binding proteins (Berntsson et al., 2010; Wuttge et al., 2012; Scheepers et al., 2016; Fenn et al., 2019). However, U3GBP possesses an additional N-terminal helix (α1), which is absent in the (sub)cluster D SBPs. This suggests that additional N-terminal helix (α1) might contribute to the plasticity of the N1 subdomain for domain closure or might involve in the establishment of SBP and TMD interaction. Moreover, analysis of the active site of the protein U3GBP reveals that U3G follows the binding feature of c-di-GMP/AMP nucleotides and thus binds to U3GBP using four different kinds of interactions (hydrogen bonds, π–π, cation–π and CH–π) reported for c-di-AMP/GMP (He et al., 2020). In the binding of c-di-GMP, residues aspartate and glutamate form hydrogen bonds with the 1-NH and 2-NH_2 atoms of the guanine base (Chou and Galperin, 2016). In a similar manner, residue Glu357 forms hydrogen bonds with the 1-NH and 2-NH_2 atoms of the guanine base of the molecule U3G. In addition, the guanidinium group of an arginine residue from the conserved motif R-X_2-D of c-di-GMP-binding proteins, forms a planar cation-π stacking interaction with the guanine base (Chou and Galperin, 2016). In the protein U3GBP also, the residue Arg360 forms a similar cation-π stacking interaction with the

guanine base of the molecule U3G. Furthermore, the π-π stacking interaction has been reported to be the most crucial factors stabilizing the c-di-GMP/AMP in the active site of the proteins (He et al., 2020). In the protein U3GBP also, three aromatic residues Phe79, Tyr224 and Tyr226 stack with the nitrogenous bases (uracil and guanine) of U3G. In addition, significance of four interactions was also assessed via mutagenesis which reveals that single interaction abolishment is not enough to prevent the endogenous binding of U3G. Thus, it can be suggested that all interactions work together for dinucleotide binding, where protein U3GBP maintains a similar interaction with the ligand U3G as in other dinucleotide-binding proteins.

Despite the presence of novel ligand, similar topology of proteins U3GBP with that of UgpBs and sugar-binding proteins leads to attainment of four similar active site subsites (A, B, C and D) as reported in maltotriose-binding proteins (Cuneo et al., 2009a). Moreover, though the active sites of (sub)cluster D SBPs are similar, the orientations of the bound ligands have been reported to be different (Pandey et al., 2016). Our structural data revealed that the orientation of the endogenously-bound U3G is like those of secondary messengers c-di-GMP/AMP which exhibits four major conformations viz. U-, V-, E- and O-types (He et al., 2020). Out of these four conformations, U3G bound to the protein U3GBP adapted a V-type conformation where the uracil base and modified ribose sugar maintains the V-shape. Moreover, through a water-mediated interaction, it can form an O-type conformation as well. An adaptation to V-type conformation of U3G was observed due to maintenance of the stacking interaction between uracil base and aromatic residues (Tyr224 and Tyr246) while O-type conformation interlinks both the ribose sugars and might play a role for ligand stabilization. The flexibility in the conformation of the molecule U3G suggests that akin to c-di-GMP/AMP, its adaptive conformations of U3G can help it to bind to different downstream proteins.

An investigation of binding mechanism of SBP for dinucleotide reveals that protein U3GBP is a new member of the subcluster D-I SBPs and shares structural homology with UgpBs and sugar-binding proteins. Thus, it can be suggested that the differences at the active sites might determine the different ligand binding as observed in distinct charge distribution for the active sites. A previous study has also reported that UgpB and sugar-binding proteins differ in the

composition of the active site residues and hence, sugar molecules could not bind to the protein UgpB (Jiang et al., 2014). Surprising, *Mt*UgpB and *Ec*UgpB bind to different ligands, however contains similar set of active site residues for ligand binding (Wuttge et al., 2012; Jiang et al., 2014). In the protein *Ec*UgpB, residues Tyr42, Ser121, Ser247 and Tyr323 interact with phosphate group of G3P (Wuttge et al., 2012). Likewise, in the *Mt*UgpB similar set of residues Tyr78, Ser153, Ser272 and Tyr345 interact with the phosphate moiety of GPC (Fenn et al., 2019). To our surprise, these residues are found to preserve their spatial positions in all three protein *Mt*UgpB, *Ec*UgpB and U3GBP, where U3GBP residues Tyr56, Ser127, Thr240 and Tyr314 interact with the phosphate group of U3G molecule. As a result, it can be suggested that these proteins follow a similar binding mode, where phosphate group initiates the ligand binding. Subsequent to binding mode investigation, an analysis of U3GBP domain closure via "Venus Fly-trap" mechanism reveals that upon ligand binding NTD undergoes a rigid translation movement while CTD remains static, which is in accordance with previous reports for other SBPs (Pandey et al., 2016; Chandravanshi et al., 2020). In summary, it can be concluded that during open to closed transition, phosphate group of U3G initiates the ligand binding while movement of NTD leads to the encapsulation of ligands in the active site.

The structural data of this study suggested that the ABC transporter (ORF IDs: TTHA0376-TTHA0379) is involved in the uptake of dinucleotides. However, observation of a U3G derivative in the active site of the protein U3GBP remains enigmatic. Further, the biogenesis and modification of the molecule U3G in the bacterial periplasm enhances the complexity of the transporter. Since the molecule U3G co-purified with the recombinant protein U3GBP overexpressed in the cytosol of *E. coli* host cells, it indicates that the molecule U3G is crucial for cytosolic proteins. This was further corroborated by the analysis of the genetic organization and the protein interactome map which predicted the interaction of U3GBP to a cytosolic protein YeaZ (ORF ID: TTHA0375). In *E. coli*, the four cytosolic proteins YeaZ (TsaB), YrdC (TsaC), YgjD (TsaD) and YjeE (TsaE) are known to be involved in the biosynthesis of N^6-threonylcarbamoyladenosine (t^6A) (Grosjean et al., 2010; El Yacoubi et al., 2012). The biosynthesis of the molecule t^6A occurs in two steps, where L-threonylcarbamoyladenylate (TCA) is synthesized followed by the transfer of its L-threonylcarbamoyl moiety at position 37 of ANN-decoding tRNAs by the enzymes YeaZ,

YgjD and YjeE (Zhang et al., 2015). The protein-protein interaction network analysis confirms the presence of N^6-threonylcarbamoyladenosine (t^6A) biosynthesis machinery in *T. thermophilus* HB8, where proteins YeaZ (TsaB, ORF ID: TTHA0375) and YgjD (TsaD, ORF ID: TTHA1252) functionally interacts with each other. Functional association of U3G ABC transport system with proteins involved in either tRNA synthesis and/or modification and similarity of U3G with secondary messengers c-di-GMP/AMP suggest that the molecule U3G can also serve as a secondary messenger for these cytosolic proteins. However, we cannot also rule out other possible unexpected function of molecule U3G associated with its novel structural features as well as with protein TTHA0357, which requires further functional confirmation at the structural level. In summary, this study orchestrates the U3G ABC transport system with tRNA synthesis and/or modification machinery for cellular metabolism.

6.5 CONCLUSION

Despite a considerable amount of information available on the nucleotide *de novo* biosynthesis, its physiological basis for the uptake and subsequent downstream processing are poorly described. In this chapter, the structural data establish that the protein U3GBP, an SBP subunit of the ABC transport system (ORF IDs: TTHA0376-TTHA0379), is involved in the uptake of a dinucleotide uridylyl-3'-5'-phospho-guanosine (U3G). Furthermore, the study provides supporting data which suggest that the imported molecule U3G is functionally associated with tRNA synthesis and/or modification system. In summary, this study enhances the understanding of dinucleotide uptake, which is preferentially depended on the requirement of downstream process.

SUMMARY

This study reports the computational and structural studies of putative sugar-binding proteins from *Thermus thermophilus* HB8 and suggests the atomic basis of ligand selection mechanism of ABC transport system.

- An initial *in silico* analysis indicates that *T. thermophilus* HB8 possess 11 ABC transport system for carbohydrate uptake. Out of 11 ABC transport system, six were characterized for the uptake of different types of sugar, while remaining four for phospholipid precursors and purine uptake.

- One of the noticeable feature identified for 11 ABC transport system is the absence NBD. An operonic and protein-protein interaction analysis of ABC transporter subunit reveals that the NBDs are not coexist with sugar ABC transporter operons and thus shared among multiple sugar ABC transporters.

- A thorough structural analysis confirms the *in silico* results and suggest that protein TTHA0356 binds to multiple disaccharides irrespective of their linkage [(α-(1,1)), (α-(1,2)), (α-(1,4)) and (α-(1,6))] and able to differentiate from the β-configuration sugars. Based on the multispecificity of protein TTA0356 for various α-glycosides, its named as α-glycosides-binding protein (or αGlyBP).

- Structurally, αGlyBP active site comprises two subsites (A and B), where subsite B initiates the binding and stabilizes the ligand in the active site as it conserves the hydrophobic stacking and polar interaction for all ligand comprising α-linked glucose unit as a structural feature.

- Mutagenesis of subsite B residues unveil the atomic details of ligand finding mechanism and suggests that information for domain closure is already encoded in the subsite B, where hinge residue viz. Arg356 facilitate the sugar binding, residue Asp70 initiates the domain closure and residue Trp287 control domain movement of NTD.

- Comparative analysis of open and closed structures of αGlyBP discloses the ligand binding mechanism where αGlyBP follows the one domain movement mechanism rather than 'Venus Fly-trap' mechanism for ligand binding.

- Akin to protein TTHA0356, structural analysis of protein TTHB082 reveals the multispecificity towards various sugars having different type of glycosidic linkages such as Cellobiose (CEL2, β-1,4), Cellotriose (CEL3, β-1,4), Cellotetraose (CEL2, β-1,4), Cellopentaose (CEL2, β-1,4), Laminaribiose (LAM2, β-1,3), Laminaritriose (LAM3, β-1,3), Laminaritetraose (LAM4, β-1,3), Sophorose (SOP2, β-1,2) and Gentiobiose (GEN2, β-1,6). Hence, we annotated the protein TTHB082 as β-glycosides-binding protein (βGlyBP).

- Structural and mutagenic data discloses the atomic detail for ligand binding mechanism, where binding of all β-glycosides occurs though nonreducing end (Glc1) via hydrophobic stacking interaction with three aromatic residues (Trp41, Trp67 and Trp177) of hydrophobic cage.

- Comparative analysis of αGlyBP and βGlyBP reveals that both protein shares similar topology, where four secondary structural elements namely two loops (L1 and L2) and two helices (H1 and H3) account for the ligand size selection as well as directs the ligand orientation. This opposite orientation of ligand is identified as a one way to differentiate between cognate and non-cognate ligand.

- For ligand binding βGlyBP follows the two-step process, where in the first step ligand initially establish the interaction with protein, while in the second step conformational dynamics of subdomains brings the domain closure for ligand encapsulation in the active site.

- The elucidation of three-dimensional structure of protein TTHA0379 reveals that it is structurally very similar to the both sugar-binding proteins αGlyBP and βGlyBP.

However, thermodynamically its identified to be involved in the uptake of a phospholipid precursor GPC during *de novo* biosynthesis of phospholipids.

- Surprisingly, the outcome of the crystal structure of the protein TTHA0379 revealed that although it shares topology sugar-and/or GPC-binding proteins, but it binds to a derivative of uridylyl-3'-5'-phospho-guanosine (U3G) rather than sugar or GPC molecules and thus we have renamed the protein TTHA0379 as a U3G-binding protein (U3GBP).

- An in-depth structural analysis of the U3GBP also provides an understanding of the different structural aspects between sugar-, GPC- and U3G-binding proteins and reveals that U3GBP binds to the dinucleotide via well-known 'Venus Fly-trap' mechanism.

- In summary, the outcome of this study hints towards the fact that though SBPs share the structural topology as identified among three proteins αGlyBP (ORF ID: TTHA0356), βGlyBP (ORF ID: TTHB082) and U3GBP (ORF ID: TTHA0379), however, differences in the active site account for the different types of ligand binding.

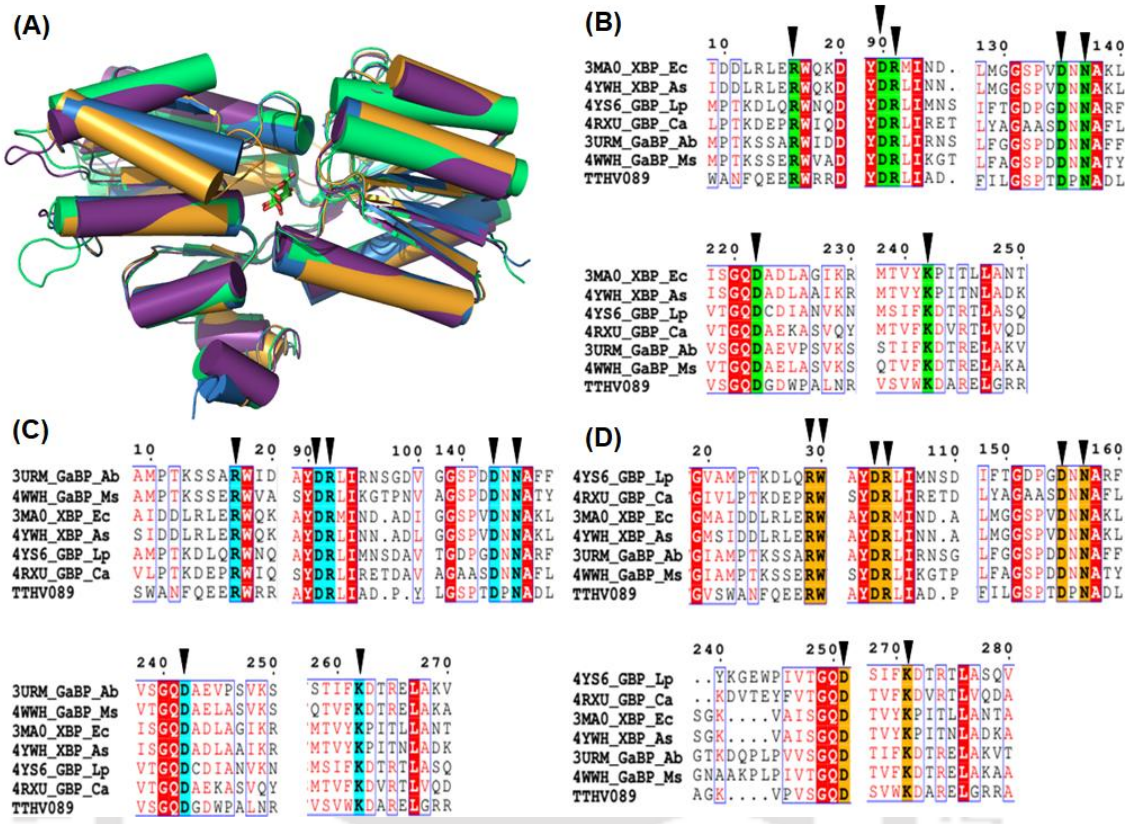


Figure A.1. Comparison of TTHV089 with xylose- (PDB ID: 3MA0), galactose- (PDB ID: 3URM) and glucose-bound (PDB ID: 4YS6) proteins. (A) Structural superimposition of TTHV089 (green) with xylose- (violet), galactose- (blue) and glucose-bound (orange) proteins. Despite the lower sequence identity, structural topology of all the proteins are identical (average rmsd: ~0.7 Å). (B, C, D) MSA of TTHV089 with homologous proteins listed in Table S2. Conservation of xylose-, galactose- and glucose-binding residues has been shown in green, cyan and orange, respectively. Abbreviations: GaBP, galactose-binding protein; GBP, glucose-binding protein and XBP, xylose-binding protein.

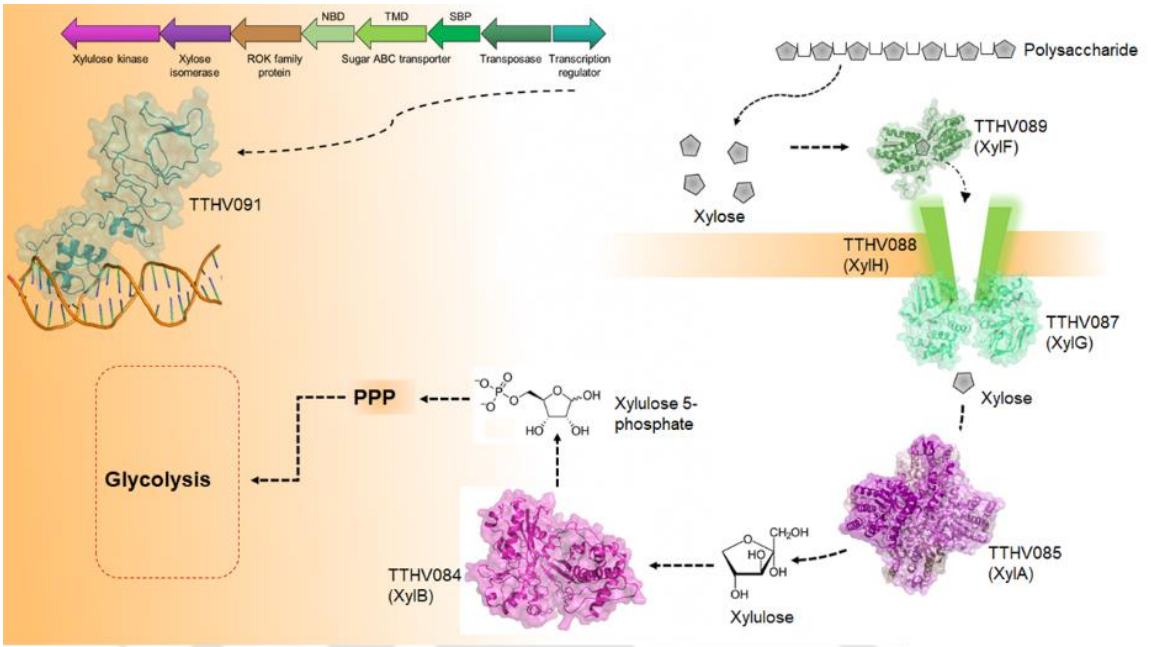


Figure A.2. Schematic representation of proposed xylose import and metabolism. The transcription regulator (ORF ID: TTHV091, blue) regulates the transcription of the operon TTHV084-TTHV089. During metabolism, polysaccharide-degraded xylose is captured by the SBP XylF (ORF ID: TTHV089, green) and subsequently transported via XylG-XylH ABC transporter (ORF IDs: TTHV087-TTHV089, green). The imported xylose is then converted into xylulose by cytosolic xylose isomerase XylA (ORF ID: TTHV085, violet). The enzyme xylulose kinase XylB (ORF ID: TTHV084, magenta) then catalyzes the phosphorylation of xylulose producing xylulose-5-phosphate, which enter the pentose phosphate pathway (PPP) for further metabolism. In figure, the crystal structure of TTHV085 (PDB ID: 1BXB) is shown, whereas for other proteins (ORF IDs: TTHV084, TTHV087, TTHV089 and TTHV091), their theoretical models are shown.

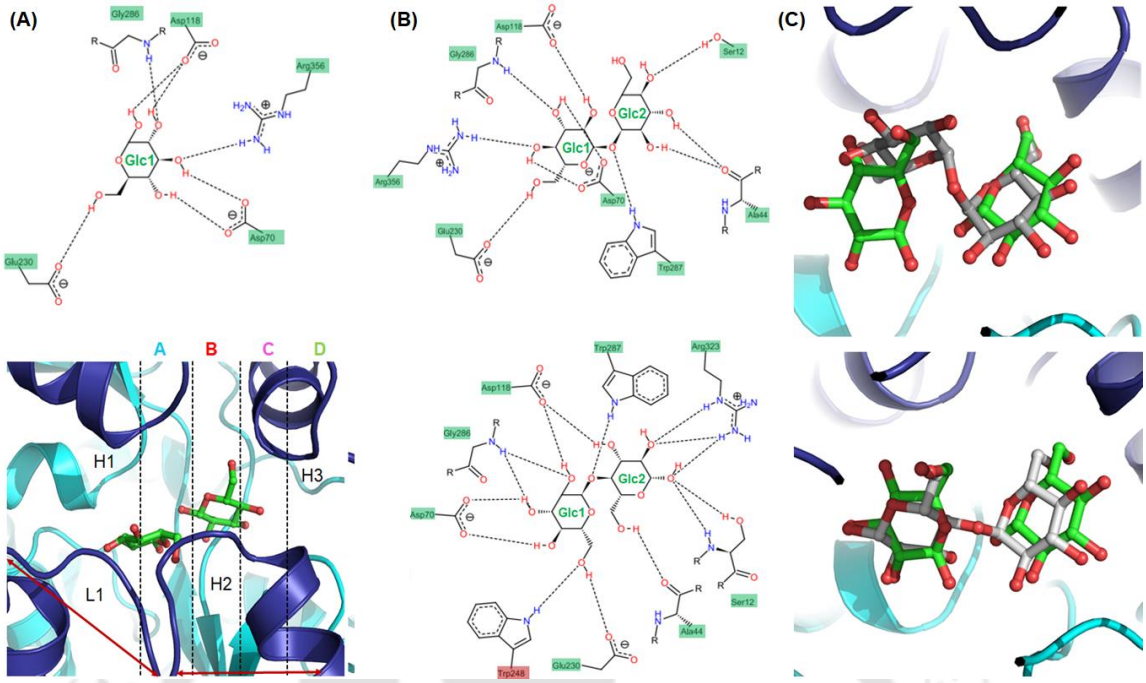


Figure A.3. Active-site pocket of the protein TTHA0356. (A) Schematic representation of the glucose-protein interaction (upper panel). The residues involved in the interaction with the glucose molecule has been labelled. Binding of glucose in the subsite (A-B) of TTHA0356 (lower panel). All the four subsites (A-D) are separated by broken lines. The secondary-structural elements such as loop (L1) and three helices (H1, H2 and H3), which modulate the four subsites, are shown in blue while rest of the protein in cyan. (B) Interaction of trehalose (upper panel) and maltose (lower panel) with active-site residues of TTHA0356. The residues interacting with all three sugars (glucose, trehalose and maltose) are highlighted in green while those specific to a particular sugar are marked in red. The hydrogen bonds established between the OH group of the sugar and the protein residues are depicted by dotted lines. (C) Superimposition of glucose (green) with trehalose (grey, upper panel) and maltose (white, lower panel), respectively.

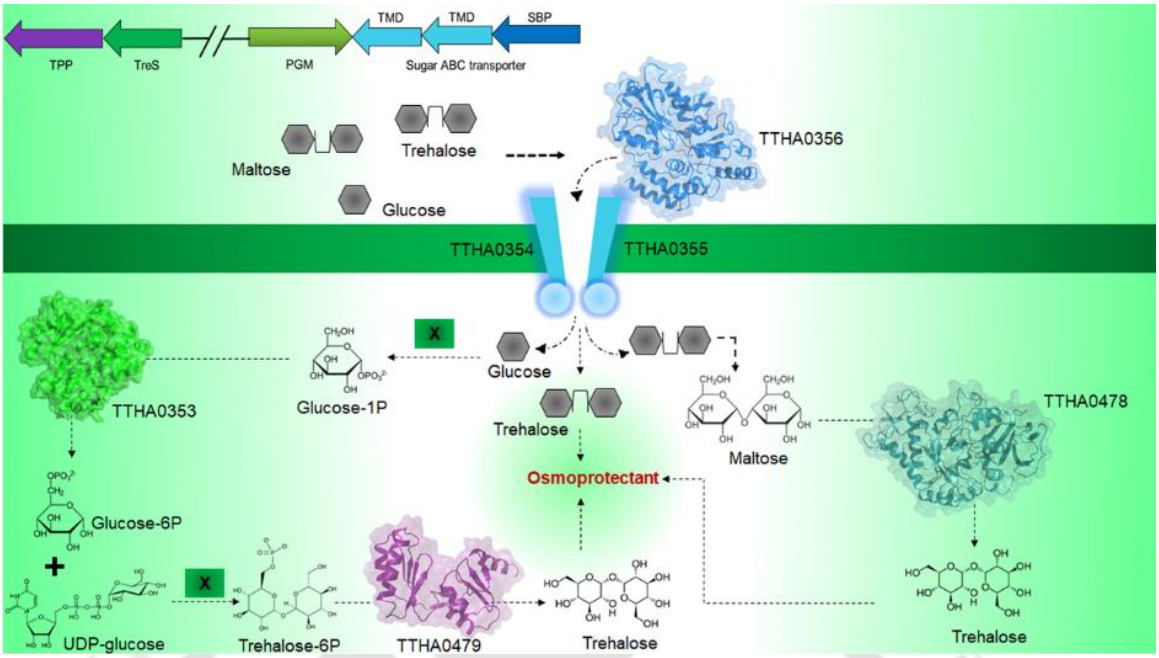


Figure A.4. Mechanism of osmoregulation in *T. thermophilus* HB8. In high-osmotic environments, bacteria accumulate trehalose via biosynthesis or importing it from the extracellular environment. Trehalose/Maltose ABC transporter (ORF IDs: TTHA0354-TTHA0356, cyan) is a major uptake system for the trehalose. However, during scarcity of trehalose in external medium, bacteria maintains the trehalose level via the *de novo* biosynthesis. To maintain the consistent requirement of trehalose, it synthesizes trehalose by two routes. In one route, imported or intracellularly available maltose is converted into trehalose by trehalose synthase TreS (ORF ID: TTHA0478, blue). Other alternative route involves transport of glucose via Trehalose/Maltose ABC transporter. In this pathway, glucose is converted into glucose-1-phosphate which is further converted to glucose-6-phosphate by phosphoglucomutase PGM (ORF ID: TTHA0353, green). Subsequently, glucose-6-phosphate is condensed with UDP-glucose and produces trehalose-6-phosphate. Produced intermediate is dephosphorylated by trehalose-phosphate phosphatase TPP (ORF ID: TTHA0479, purple) yielding a trehalose. For the protein TTHA0353, its crystal structure (PDB ID: 2Z0F) is shown, whereas for other proteins (ORF IDs: TTHA0478, TTHA0479 and TTHA0356), their theoretical models are shown here.

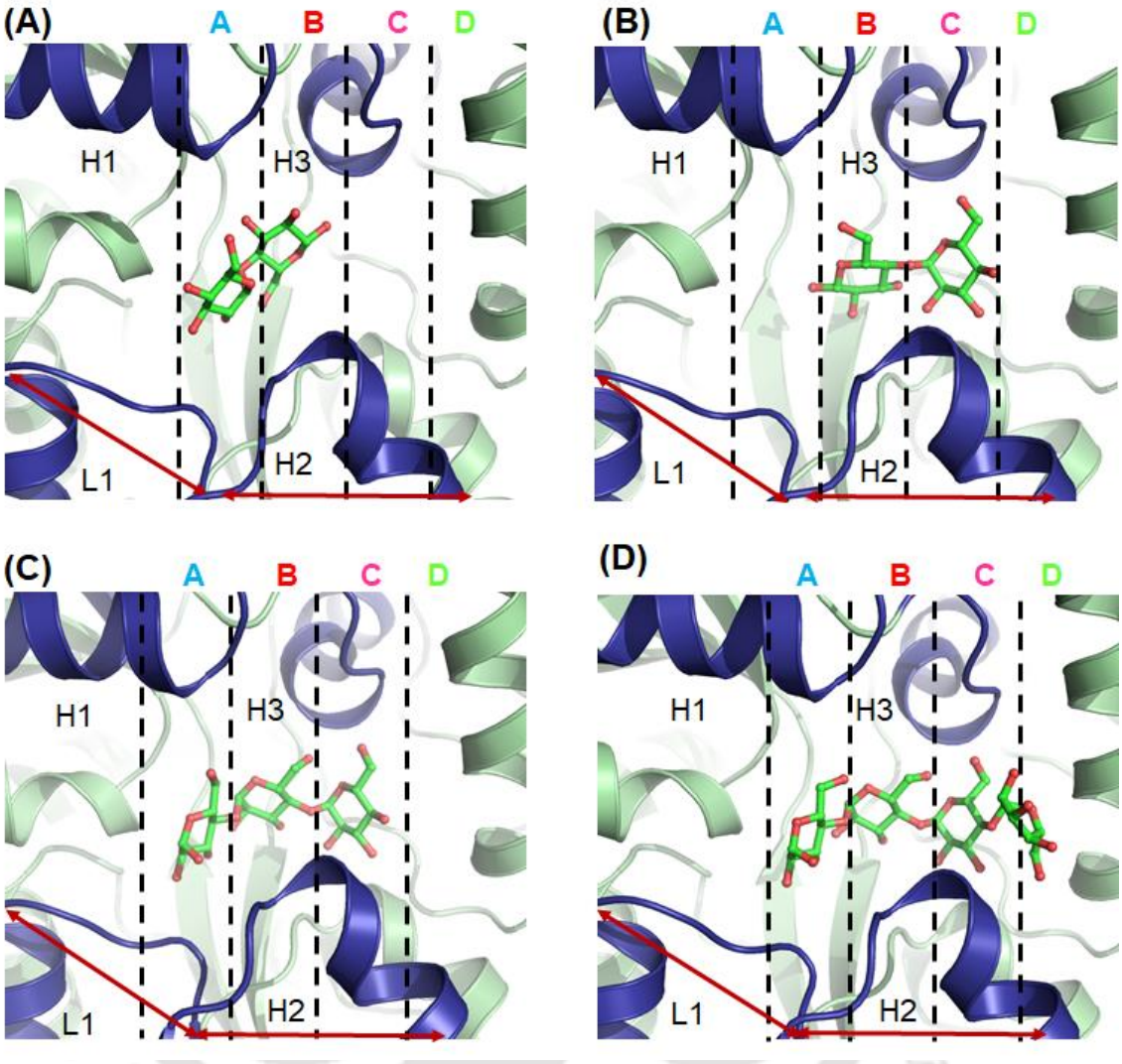


Figure A.5. Active-site pocket of the protein TTHA1652. The four subsites-modulating loop (L1) and helices (H1, H2 and H3) are shown in blue while rest of the protein in green. Each subsite (A-D) is separated by vertical broken lines. The accommodation for the glucose polymer (di-, tri- and tetra-saccharides) in the four subsites is analyzed by molecular docking experiment of the protein TTHA1652 with the ligands (A, B) maltose, (C) maltotriose and (D) maltotetraose, shown as ball-and-stick model in green. The results illustrate that the active-site-modulating secondary structural elements, H3 in particular, are far from the cavity allowing the binding of glucose polymers.

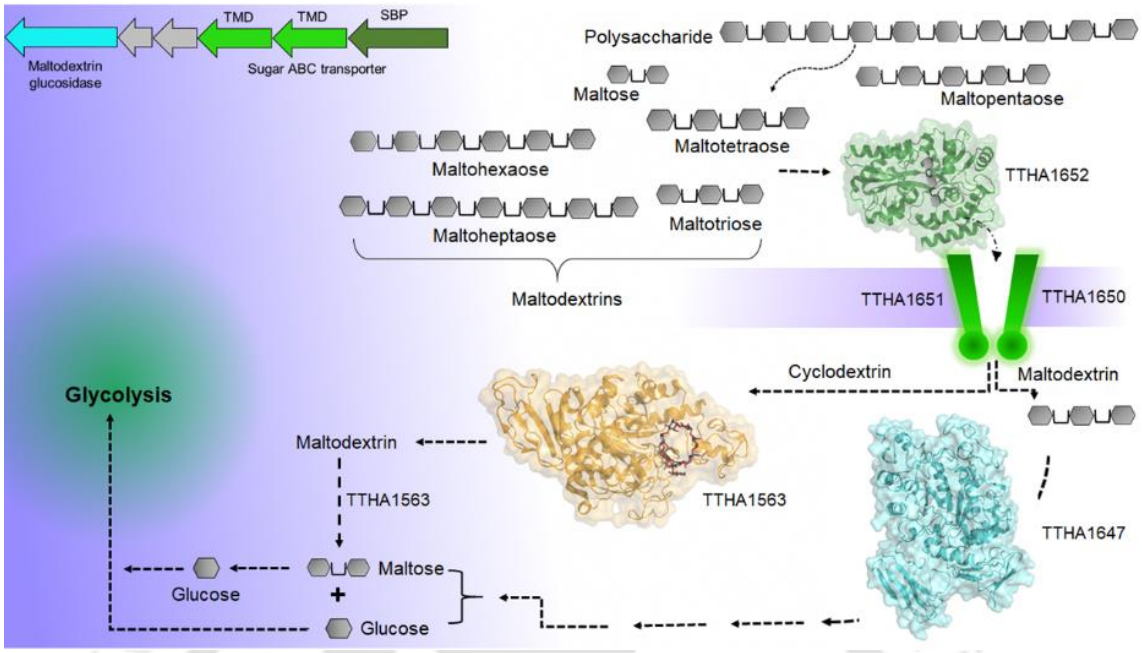


Figure A.6. Schematic representation of the uptake and metabolic systems for cyclo/maltodextrin. During the metabolism, glucose polymers (grey), liberated as a starch-degraded product, are recognized by the protein TTHA1652 (green) for its transportation via the ABC transporter (ORF IDs: TTHA1650-TTHA1652, green). The imported glucose polymers (i.e. cyclo/maltodextrins) are further metabolized by the maltodextrin glucosidase (ORF ID: TTHA1647, blue) and neopullulanase (ORF ID: TTHA1563, orange) into simple sugar molecules which enter the glycolysis pathway for further processing. In the figure, the crystal structure of TTHA1563 (PDB ID: 2Z1K) is shown while the theoretical models for the protein TTHA1652 and TTHA1647 are provided.

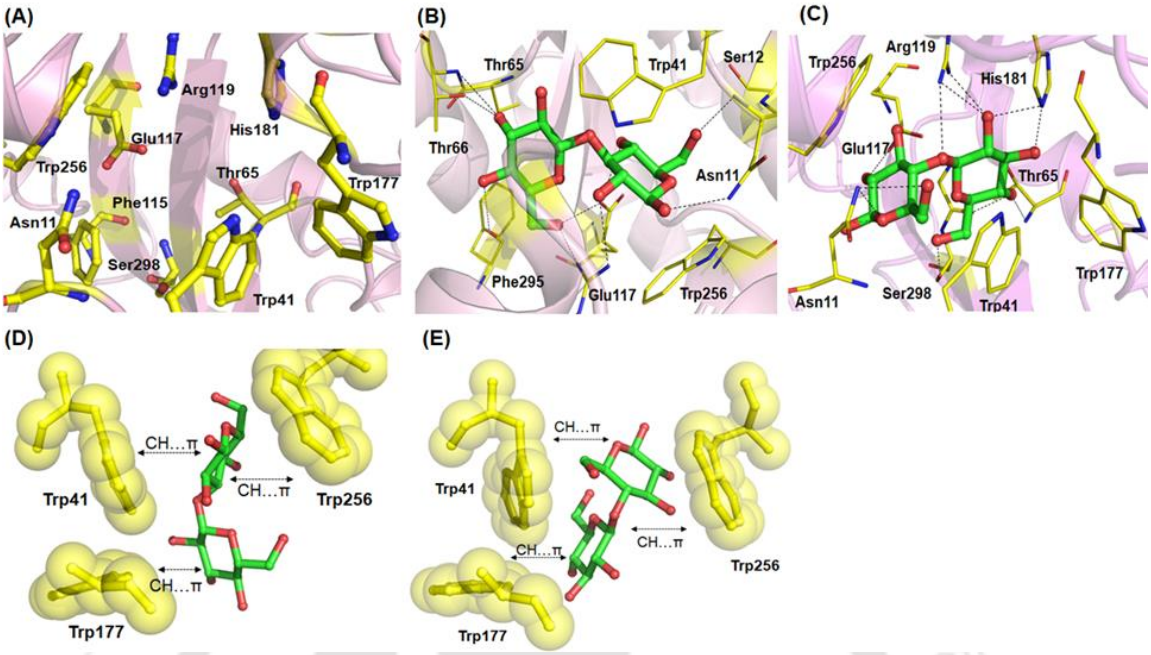


Figure A.7. Active-site pocket of the protein TTHB082. (A) Aromatic and polar residues-rich active site of the protein TTHB082 is shown as ball-and-stick model in yellow, respectively. (B, C) Schematic representation of binding of the protein TTHB082 with laminaribiose and cellobiose, respectively. The amino acid residues and the bound sugars are shown in yellow lines and green ball-and-stick model, respectively. Hydrogen bonds are depicted by black-dotted lines. (D, E) Tryptophan residues Trp41, Trp177 and Trp256 (yellow spheres) making the CH…π interaction with the glucose unit of the laminaribiose and cellobiose (green), respectively, are also shown.

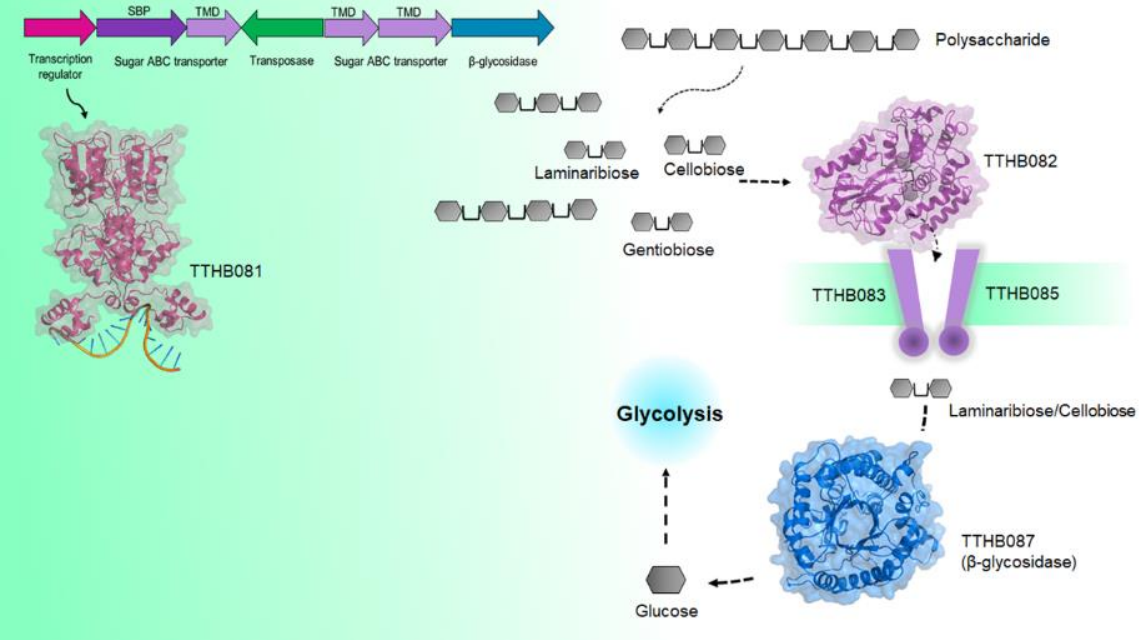


Figure A.8. Schematic illustration of the mechanisms of uptake and metabolism of β-glucosides. The transcription regulator (ORF ID: TTHB081, red) regulates the transcription of ABC transporter (ORF IDs: TTHB082-TTHB085, purple) and β-glycosidase (ORF ID: TTHB087, blue). During the metabolism, β-glucosides (grey), a degraded product of cellulose (grey), are recognized by the protein TTHB082 for its transportation. The imported β-glucosides is subsequently hydrolyzed to glucose molecules by the enzyme β-glycosidase. For the protein TTHB087, its crystal structure (PDB ID: 4BCE) is shown, whereas for other proteins (ORF IDs: TTHB081 and TTHB082), their theoretical models are shown here.

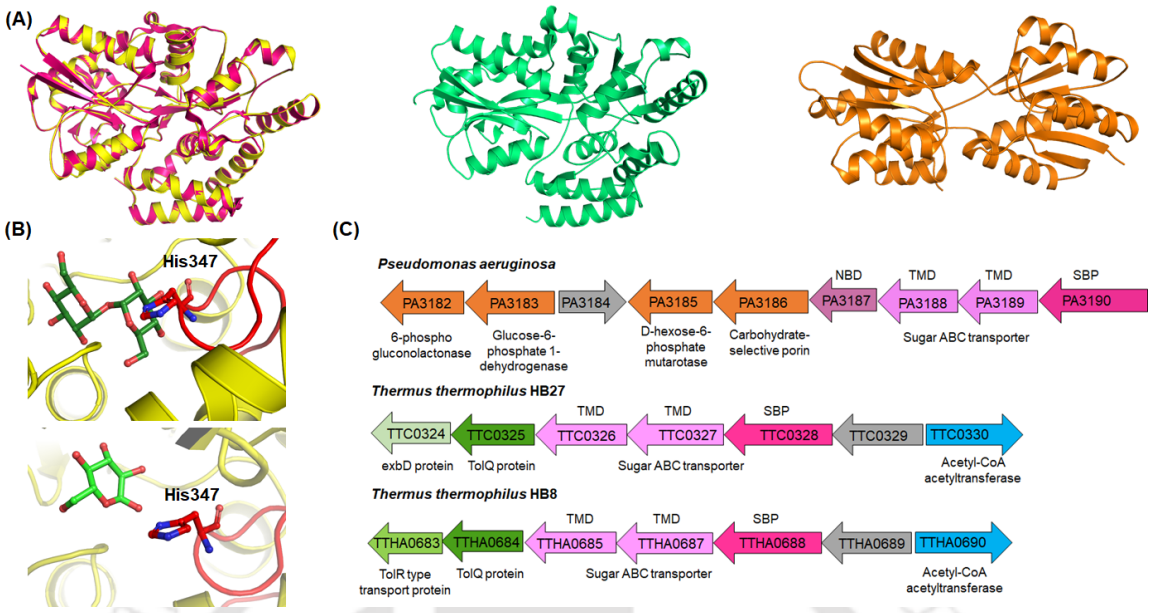


Figure A.9. Schematic representation of active site of TTHA0688 and genetic cluster for glucose uptake and metabolism systems. (A) Structural superimposition of TTHA0688 (yellow) with GBP (PDB ID: 2B3B, magenta, left) and its topological comparison with disaccharide binding protein TMBP (PDB ID: 1EU8, green, center) and monosaccharide binding protein (PDB ID: 1GLG, orange, right). (B) Active-site comparison of protein TTHA0688 with TMBP. The loop L1 and the residue His347, which determine the accommodation of monosaccharide (glucose, green) over disaccharide (trehalose, green), is highlighted in red while the rest of the protein in yellow. (C) Genetic cluster for glucose uptake and metabolism from *P. aeruginosa*, *T. thermophilus* HB27 and *T. thermophilus* HB8. Each gene is represented by an arrow indicating its direction of transcription. ORF number with its respective encoded putative protein name is mentioned inside and below the arrow, respectively. ORFs encoding various proteins are depicted by different colors e.g. glucose uptake via ABC transporter (magenta), glucose-metabolizing genes (orange), other functionally-associated protein (green and blue) and uncharacterized protein (grey).

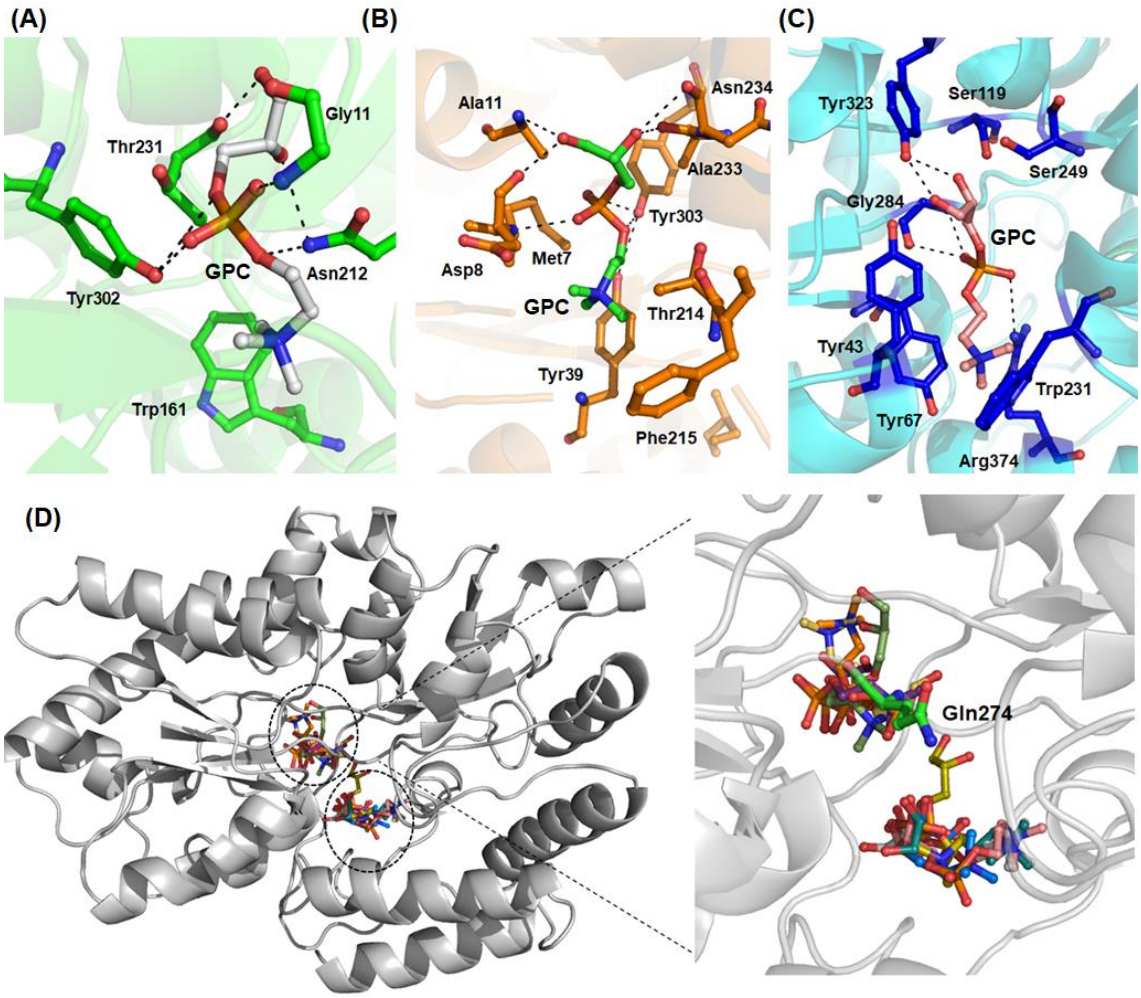


Figure A.10. Schematic representation of GPC binding to UgpB proteins. (A) TTHA1877, (B) TTHA1936 and (C) TTHV034. Residues involved in interaction are represented as ball-and-stick model and hydrogen bonds are depicted by black-dotted lines. (D) In TTHA0379, the GPC binding site of different clusters obtained through molecular docking experiment are encircled and zoomed (right panel) showing the hindrance of the GPC binding by the residue Gln274 (green) in the active site of the protein.

APPENDIX A- SUPPLEMENTARY DATA TO CHAPTER 3

Table A.1. List of proteins of the carbohydrate ABC transporter family from *Thermus thermophilus* HB8 and their proposed ligands.

S. No.	[+]Protein name (ORF number)	Gene location	Cellular location (subunit)	Previous annotated ligand	Present annotated ligand (this study)	
					1st preference	2nd preference
1.	ABC transporter-like protein (TTHV087)	Mini plasmid (TTHV)	Cytosol (NBD)	Sugar	Xylose	-
2.	Permease protein, ABC-type xylose transporter (TTHV088)		Membrane (TMD)	Sugar		
3.	D-xylose ABC transporter periplasmic substrate-binding protein (TTHV089)		Periplasm (SBP)	Sugar		
4.	ABC transporter, permease protein, MalFG family (TTHA0354)	Chromosome (TTHA)	Membrane (TMD)	Sugar	Trehalose	Maltose, Glucose
5.	ABC transporter, permease protein, MalFG family (TTHA0355)		Membrane (TMD)	Sugar		
6.	ABC transporter, periplasmic substrate-binding protein (TTHA0356)		Periplasm (SBP)	Sugar		
7.	Sugar ABC transporter, ATP-binding protein (TTHA0975)		Cytosol (NBD)	Sugar	Mannosylglycerate	-
8.	Sugar ABC transporter, permease protein(TTHA0976)		Membrane (TMD)	Sugar		
9.	Sugar ABC transporter, permease protein (TTHA0977)		Membrane (TMD)	Sugar		
10.	Uncharacterized protein (TTHA0979)		Periplasm (SBP)	Sugar		

11.	Maltose ABC transporter, permease protein (TTHA1650)		Membrane (TMD)	Sugar (maltose)	Maltotriose	Maltodextrin, Cyclodextrin
12.	Maltose ABC transporter, permease protein (TTHA1651)		Membrane (TMD)	Sugar		
13.	Maltose ABC transporter, periplasmic maltose-binding protein(TTHA1652)		Periplasm (SBP)	Sugar (maltose)		
14.	Sugar ABC transporter, periplasmic sugar-binding protein (TTHB082)	Mega plasmid (TTHB)	Periplasm (SBP)	Sugar	Laminaribose	Cellobiose
15.	Sugar ABC transporter, permease protein (TTHB083)		Membrane (TMD)	Sugar		
16.	Sugar ABC transporter, permease protein (TTHB085)		Membrane (TMD)	Sugar		
17.	Sugar ABC transporter, permease protein (TTHB086)		Membrane (TMD)	Sugar		
18.	Sugar ABC transporter, permease protein (TTHA0685)	Chromosome (TTHA)	Membrane (TMD)	Sugar	Glucose	Galactose
19.	Sugar ABC transporter, permease protein [C-terminal] (TTHA0686)		Membrane C-terminal (TMD)	Sugar		
20.	Sugar ABC transporter, permease protein [N-terminal] (TTHA0687)		Membrane N-terminal (TMD)	Sugar		
21.	Sugar ABC transporter, substrate-binding protein (TTHA0688)		Periplasm (SBP)	Sugar		
22.	*Sugar ABC transporter, permease protein (TTHA0377)		Membrane (TMD)	Sugar	Glycerophosphocholine (GPC)	-

23.	*Sugar ABC transporter, permease protein (TTHA0378)		Membrane (TMD)	Sugar		
24.	*Sugar ABC transporter, periplasmic sugar-binding protein (TTHA0379)		Periplasm (SBP)	Sugar		
25.	Sugar ABC transporter, periplasmic sugar-binding protein(TTHA1877)		Periplasm (SBP)	Sugar	Glycerophosphocholine (GPC)	-
26.	Glycerol-3-phosphate ABC transporter, periplasmic glycerol-3-phosphate-binding protein (TTHA1936)		Periplasm (SBP)	Glycerol-3-phosphate (G3P)	Glycerophosphocholine (GPC)	-
27.	Glycerol-3-phosphate ABC transporter permease (TTHV032)	Mini plasmid (TTHV)	Membrane (TMD)	Glycerol-3-phosphate (G3P)	Glycerophosphocholine (GPC)	-
28.	*sn*-glycerol-3-phosphate transport system permease UgpE (TTHV033)		Membrane (TMD)	Glycerol-3-phosphate (G3P)		
29.	Glycerol-3-phosphate ABC transporter substrate-binding protein (TTHV034)		Periplasm (SBP)	Glycerol-3-phosphate (G3P)		
30.	Uncharacterized protein (TTHA1301)	Chromosome (TTHA)	Periplasm (SBP)	Sugar	Adenine	Guanine
31.	Sugar ABC transporter, permease protein (TTHA1302)		Membrane (TMD)	Sugar		
32.	Sugar ABC transporter, permease protein (TTHA1303)		Membrane (TMD)	Sugar		
33.	Sugar ABC transporter, ATP-binding protein (TTHA1304)		Cytosol (NBD)	Sugar		

34.	Sugar ABC transporter, ATP-binding protein (TTHA0579)		Cytosol (NBD)	Sugar	-	-
35.	Sugar ABC transporter, ATP-binding protein (TTHA1099)		Cytosol (NBD)	Sugar	-	-

[+] Protein names are according to UniProtKB database.
* These ORFs encode Ugp ABC transporter as reported in the reference [68].

Table A.2. List of structural homologs of SBP subunits of sugar and purine ABC transporters from *Thermus thermophilus* HB8.

ORF number	PDB ID	UniProt ID	[+]Protein name	Bound ligand	*Z-score	RMSD (Å)	LALI	NRES	ID (%)
TTHV089	3M9W	P37387	D-xylose-binding periplasmic protein	-	48.7	0.1	303	305	41
	3M9X	P37387	D-xylose-binding periplasmic protein	D-xylose	47.6	0.3	303	305	41
	3MA0	P37387	D-xylose-binding periplasmic protein	D-xylose	40.9	3.6	301	306	41
	4YWH	A6VLM7	D-xylose ABC transporter, periplasmic substrate-binding protein	D-xylose	40.3	3.5	302	307	40
	3URM	P25548	Multiple sugar-binding periplasmic receptor ChvE	D-galactose	34.5	3.3	302	329	32
	3UUG	P25548	Multiple sugar-binding periplasmic receptor ChvE	D-glucuronic acid	34.3	3.4	302	329	32
	4WWH	A0QT42	ABC transporter	D-galactose	34.1	3.5	303	329	31
	4YS6	A9KQP6	Putative solute-binding component of ABC transporter	D-glucose	32.2	3.8	300	324	30
	4RXU	A9WDY0	Periplasmic sugar-binding protein	D-glucose	31.6	3.7	300	340	32
	1BA2	P02925	Ribose import binding protein RbsB	-	31.1	2.5	268	271	22
	1URP	P02925	Ribose import binding protein RbsB	-	30.7	2.2	268	271	22

	4KQ9	D3F6W3	Ribose ABC transporter, substrate binding protein	-	30.5	2.3	279	298	24
	3C6Q	Q9WXW9	Sugar ABC transporter, periplasmic sugar-binding protein	D-xylose	30.1	2.3	290	305	22
	2FN9	Q9X053	Ribose ABC transporter, periplasmic ribose-binding protein	-	30.1	2.2	270	280	24
	1GUD	P39265	D-allose-binding periplasmic protein	-	29.7	2.5	270	288	22
	4RWE	Q7CFZ8	Sugar-binding transport protein	-	29.3	2.3	277	287	25
	5HSG	A6T990	Putative ABC transporter, nucleotide binding/ATPase protein	-	29.3	2.1	277	285	20
	1GUB	P39265	D-allose-binding periplasmic protein	-	29.1	2.6	270	288	22
	2QW1	P0AEE5	D-galactose-binding periplasmic protein	3-O-methyl D-glucose	28.6	3.3	276	305	26
	2FW0	P0AEE5	D-galactose-binding periplasmic protein	-	28.5	3.2	275	305	26
TTHA0356	1EU8	Q7LYW7	Trehalose/maltose-binding protein MalE	Trehalose	69	0.3	394	407	38
	5CI5	A8F7X5	Extracellular solute-binding protein family 1	D-tagatose	46.7	2.2	380	393	24
	4QRZ	A9CGI0	ABC transporter, substrate binding protein (Sugar)	Maltotriose	45.8	2.0	370	383	20
	4QSE	A9CGI0	ABC transporter, substrate binding protein (Sugar)	Glycerol	45.6	2.0	370	382	20
	4QSC	A9CGI0	ABC transporter, substrate binding protein (Sugar)	Maltose	45.5	2.0	371	383	20
	3K00	B0B0V1	Acarbose/maltose binding protein GacH	Maltotetraose	45	2.0	372	388	26
	5IAI	B9JM84	Sugar ABC transporter	D-ribitol	45	2.3	383	399	20

	4QSD	A9CGI0	ABC transporter, substrate binding protein (Sugar)	Sucrose	44.9	2.0	369	382	21
	3JZJ	B0B0V1	Acarbose/maltose binding protein GacH	Acarbose	44.9	2.0	373	388	25
	3K02	B0B0V1	Acarbose/maltose binding protein GacH	**TXT	44.9	2.0	374	388	25
	4RYA	B9JRF8	ABC transporter substrate binding protein (Sorbitol)	D-mannitol	42.5	2.6	383	417	19
	5YSE	Q92AS8	Lin1841 protein	Sophorotetraose	41.6	2.3	368	387	20
	5YSD	Q92AS8	Lin1841 protein	Sophorotriose	41.5	2.4	369	389	20
	5YSF	Q92AS8	Lin1841 protein	Sophoropentaose	41.4	2.4	369	389	20
	5F7V	Q8YAE9	Lmo0181 protein	Cycloalternan	40.8	2.1	363	388	18
	4RJZ	A9CGI0	ABC transporter, substrate binding protein (Sugar)	-	40.4	3.7	367	382	20
	6FFL	Q6MNM0	Maltose/maltodextrin transport permease homologue	Maltotriose	40.4	2.5	368	383	18
	3K01	B0B0V1	Acarbose/maltose binding protein GacH	-	39.9	3.8	373	391	26
	2GH9	Q72I44	Maltose/maltodextrin-binding protein	Maltotriose	39.3	2.3	362	378	19
	2GHA	Q9X0T1	Maltose ABC transporter, periplasmic maltose-binding protein	Maltotriose	39.3	2.3	362	374	20
TTHA1652	2GH9	Q72I44	Maltose/maltodextrin-binding protein	Maltotriose	67.3	0.1	378	378	99
	2GHA	Q9X0T1	Maltose ABC transporter, periplasmic maltose-binding protein	Maltotriose	49.8	1.5	366	374	35
	2FNC	Q9S5Y1	Maltose ABC transporter, periplasmic maltose-binding protein	Maltotriose	49.6	1.7	367	379	36

	2ZYM	Q9AJF5	Maltodextrin-binding protein	α-cyclodextrin	47.4	1.9	369	381	34
	1URD	Q9RHZ6	Maltodextrin-binding protein	Maltotriose	47.4	1.9	365	370	30
	2ZYN	Q9AJF5	Maltodextrin-binding protein	β-cyclodextrin	47.3	1.9	369	383	34
	1URG	Q9RHZ6	Maltodextrin-binding protein	Maltose	47.2	1.9	366	373	30
	2ZYK	Q9AJF5	Maltodextrin-binding protein	γ-cyclodextrin	47.0	1.9	369	381	34
	1URS	Q9RHZ6	Maltodextrin-binding protein	Maltotriose	46.9	1.9	362	365	30
	6FFL	Q6MNM0	Maltose/maltodextrin transport permease homologue	Maltotriose	46.4	1.8	368	383	26
	3C4M	P0AEX9	$^{++}$Fusion protein of maltose-binding periplasmic protein and Parathyroid hormone/parathyroid hormone-related peptide receptor	Maltose	45.1	2.0	366	469	30
	2VGQ	P0AEX9	$^{++}$Sugar ABC transporter substrate-binding protein, Mitochondrial antiviral-signaling protein	Maltotetraose	45.0	2.0	366	461	30
	5AZ8	P0AEX9	$^{++}$Maltose-binding periplasmic protein, mitochondrial import receptor subunit TOM20 homolog	Maltose	44.9	2.0	366	433	31
	5Z0R	P0AEX9	$^{++}$Extracellular solute-binding protein family 1,viral genome protein	Maltose	44.8	2.0	366	445	30
	5AZ7	P0AEX9	$^{++}$Maltose-binding periplasmic protein, mitochondrial import receptor subunit TOM20 homolog	Maltose	44.8	2.1	366	434	31

	5Z0V	P0AEX9	++Extracellular solute-binding protein family 1,viral genome protein	Maltose	44.7	2.1	366	444	30
	5H7N	P0AEX9	NLRP12-PYD with MBP tag	Maltotetraose	44.7	2.0	366	457	30
	4EGC	P0AEX9	++Maltose-binding periplasmic protein, Homeobox protein SIX1 chimera	Maltose	44.7	2.0	366	539	30
	4XR8	P0AEX9	++Maltose-binding periplasmic protein, ubiquitin ligase E6AP	Maltose	44.7	2.0	366	381	29
TTHB082	5CI5	A8F7X5	Extracellular solute-binding protein family 1	D-tagatose	65.6	0.6	387	393	21
	1EU8	Q7LYW7	Trehalose/maltose binding protein MalE	Trehalose	45.5	2.2	387	407	21
	4QRZ	A9CGI0	ABC transporter, substrate binding protein (Sugar)	Maltotriose	45.0	2.0	369	383	20
	4QSC	A9CGI0	ABC transporter, substrate binding protein (Sugar)	Maltose	44.8	2.0	369	383	20
	3K00	B0B0V1	Acarbose/maltose binding protein GacH	Maltotetraose	44.8	2.0	371	388	25
	4QSE	A9CGI0	ABC transporter, substrate binding protein (Sugar)	Glycerol	44.7	2.0	368	382	20
	3JZJ	B0B0V1	Acarbose/maltose binding protein GacH	Acarbose	44.7	2.0	371	388	25
	3K02	B0B0V1	Acarbose/maltose binding protein GacH	**TXT	44.6	2.0	371	388	25
	4QSD	A9CGI0	ABC transporter, substrate binding protein (Sugar)	Sucrose	44.3	2.1	368	382	20
	5YSE	Q92AS8	Lin1841 protein	Sophorotetraose	44.2	2.1	371	387	24
	5YSD	Q92AS8	Lin1841 protein	Sophorotriose	43.9	2.2	371	387	24
	5YSF	Q92AS8	Lin1841 protein	Sophoropentaose	43.9	2.2	371	388	24
	5F7V	Q8YAE9	Lmo0181 protein	Cycloalternan	42.2	2.1	368	388	18

	5IAI	B9JM84	Sugar ABC transporter		41.7	2.5	379	399	21
	2GH9	Q72I44	Maltose/maltodextrin-binding protein	Maltotriose	40.2	2.2	365	378	21
	2FNC	Q9S5Y1	Maltose ABC transporter, periplasmic maltose-binding protein	Maltotriose	40.2	2.2	362	379	22
	4RJZ	A9CGI0	ABC transporter, substrate binding protein (Sugar)	-	40.0	3.9	367	382	20
	4RYA	B9JRF8	ABC transporter substrate binding protein (Sorbitol)	D-mannitol	39.9	2.9	384	417	18
	5YSB	Q92AS8	Lin1841 protein	-	39.8	3.9	370	386	24
	2GHA	Q9X0T1	Maltose ABC transporter, periplasmic maltose-binding protein	Maltotriose	39.6	2.2	362	374	23
TTHA1301	3S99	Q2YKI6	Purine-binding protein BAB2_0673	Adenine	57.6	0.2	328	330	41
	2FQY	P29724	Membrane lipoprotein TmpC	Adenosine	34.1	1.8	294	316	18
	2FQW	P29724	Membrane lipoprotein TmpC	Inosine	33.4	2.0	294	316	18
	2FQX	P29724	Membrane lipoprotein TmpC	Guanosine	33.4	2.0	294	316	18
	4PEV	Q9Y8P1	Membrane lipoprotein family protein	Adenosine	31.5	2.4	298	370	19
	4IIL	Q56328	ABC transporter riboflavin-binding protein RfuA	Riboflavin	29.7	2.3	289	314	18
	2HQB	Q9K8W3	Transcriptional activator of comK protein	-	28.2	3.1	278	283	22
	4P98	D3F9N7	Basic membrane lipoprotein	-	27.6	3.7	284	300	21
	5IBQ	Q2JZQ5	Probable ribose ABC transporter, substrate-binding protein	α-D-apiose	18.9	3.0	254	287	14
	4RY0	Q2JZQ5	Probable ribose ABC transporter, substrate-binding protein	D-ribose	18.9	3.0	254	287	14
	1SXG	P46828	Catabolite control protein A	-	18.9	3.2	252	273	10

	2H3H	Q9WXW9	Sugar ABC transporter, periplasmic sugar-binding protein	D-glucose	18.8	3.2	250	313	16
	1SXI	P46828	Catabolite control protein A	-	18.7	3.2	252	273	10
	3L6U	B1YL86	Periplasmic binding protein/LacI transcriptional regulator	-	18.6	3.2	246	274	14
	3O74	Q88PQ6	Catabolite repressor-activator, DNA-binding transcriptional dual regulator	-	18.6	3.1	245	271	11
	5DTE	A6VKG5	Monosaccharide-transporting ATPase	D-allose	18.6	3.1	249	286	11
	2QVC	Q9WXW9	Sugar ABC transporter, periplasmic sugar-binding protein	D-glucose	18.6	3.3	249	302	16
	3C6Q	Q9WXW9	Sugar ABC transporter, periplasmic sugar-binding protein	D-xylose	18.5	3.3	248	305	15
	4WUT	B9K0B2	ABC transporter substrate binding protein (Ribose)	D-fucose	18.5	3.4	250	290	12
	2RJO	B2TEP5	Periplasmic binding protein/LacI transcriptional regulator	D-glucose	18.5	3.1	259	322	10

[+] Protein names are according to UniProtKB database.

[++] Fusion protein tagged with maltose binding protein; protein names are according to the PDB database.

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

** TXT: ligand ID of 4,6-dideoxy-4-{[(1S,2R,3R,4S,5S)-2,3,4-trihydroxy-5-(hydroxymethyl)cyclohexyl]amino}-alpha-D-allopyranosyl-(1->4)-alpha-D-glucopyranosyl-(1->4)-alpha-D-glucopyranosyl-(1->4)-alpha-D-glucopyranose.

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

Table A.3. Details of the molecular docking results of various sugars with SBPs.

ORF number	Ligand	*EFBE (kcal mol^{-1})	*NHBOND	Ligand subunit	Interactions	
					Ligand atom	Protein residue atom
JW3538 (*Ec*XBP, PDB ID: 3MA0)	Xylose	-8.24	13	-	O1	Asp90 $O^{\delta2}$, Arg91 N^{ε}, Asp135 $O^{\delta2}$, Asn137 $N^{\delta2}$
					O2	Asp90 $O^{\delta2}$, Asn137 $N^{\delta2}$, Lys242 N^{ζ}
					O3	Arg16 $N^{\eta1}$, Asp222 $O^{\delta1}$, Lys242 N^{ζ}
					O4	Asn196 $N^{\delta2}$, Asp222 $O^{\delta2}$
					O5	Arg91 $N^{\eta2}$
	Galactose	-7.06	13	-	O1	Arg91 $N^{\eta2}$, Asp135 $O^{\delta2}$, Trp169 $N^{\varepsilon1}$
					O2	Asp90 $O^{\delta2}$, Arg91 N^{ε}, Asn137 $N^{\delta2}$
					O3	Arg16 $N^{\eta1}$, Asp90 $O^{\delta1}$ $O^{\delta2}$, Lys242 N^{ζ}
					O4	Arg16 $N^{\eta1}$, Asp222 $O^{\delta1}$
					O5	-
					O6	Asn196 $N^{\delta2}$, Asp222 $O^{\delta1}$ $O^{\delta2}$
	Glucose	-6.57	13	-	O1	Asp135 $O^{\delta2}$, Asn137 $N^{\delta2}$
					O2	Asp90 $O^{\delta2}$, Asn137 $N^{\delta2}$, Lys242 N^{ζ}
					O3	Arg16 $N^{\eta1}$, Asp222 $O^{\delta1}$, Lys242 N^{ζ}
					O4	Asn196 $N^{\delta2}$, Asp222 $O^{\delta2}$
					O5	Arg91 $N^{\eta2}$
					O6	Arg91 $N^{\eta2}$, Trp169 $N^{\varepsilon1}$
TTHV089	Xylose	-8.3	16	-	O1	Arg89 N^{ε}, Asp132 $O^{\delta1}$ & $O^{\delta2}$, Asn134 $N^{\delta2}$
					O2	Asp88 $O^{\delta2}$, Asn134 $N^{\delta2}$, Lys239 N^{ζ}
					O3	Arg14 $N^{\eta1}$, Asp219 $O^{\delta1}$ & $O^{\delta2}$, Lys239 N^{ζ}
					O4	Glu12 $O^{\varepsilon1}$, Asn193 $N^{\delta2}$, Asp219 $O^{\delta2}$
					O5	Arg89 $N^{\eta2}$, Asp132 $O^{\delta2}$
	Galactose	-6.57	14	-	O1	Arg89 $N^{\eta2}$, Asp132 $O^{\delta2}$, Trp166 $N^{\varepsilon1}$
					O2	Arg89 N^{ε}, Asp132 $O^{\delta2}$, Asn134 $N^{\delta2}$
					O3	Asp88 $O^{\delta2}$, Asn134 $N^{\delta2}$, Lys239 N^{ζ}
					O4	Arg14 $N^{\eta1}$, Asp219 $O^{\delta2}$, Lys239 N^{ζ}
					O5	-
					O6	Asn193 $N^{\delta2}$, Asp219 $O^{\delta2}$
	Glucose	-7.09	13	-	O1	Asp88 $O^{\delta2}$, Arg89 N^{ε}
					O2	Glu12 $O^{\varepsilon1}$, Arg89 $N^{\eta2}$

					O3	Glu12 $O^{\varepsilon1}$, Asn193 $N^{\delta2}$, Asp219 $O^{\delta2}$
					O4	Gln218 $O^{\varepsilon1}$, Asp219 $O^{\delta1}$, Lys239 N^{ζ}
					O5	Asn134 $N^{\delta2}$
					O6	Gln218 $O^{\varepsilon1}$ & $N^{\varepsilon2}$
TTHA0356	Glucose (subsite A)	-5.66	10	-	O1	Asp118 $O^{\delta2}$
					O2	Arg323 N^{ε}
					O3	Ser12 N & O^{γ}, Arg323 $N^{\eta2}$
					O4	Ser12 N & O^{γ}
					O5	Trp287 $N^{\varepsilon1}$
					O6	Ala44 O
	Glucose (subsite B)	-6.09	10	-	O1	Asp118 $O^{\delta2}$, Trp248 $N^{\varepsilon1}$
					O2	Asp118 $O^{\delta2}$, Gly286 N
					O3	Asp70 $O^{\delta1}$
					O4	Asp70 $O^{\delta2}$, Arg356 $N^{\eta2}$
					O5	Trp248 $N^{\varepsilon1}$
					O6	Glu230 $O^{\varepsilon2}$
	Trehalose (subsites A-B)	-8.47	14	Glc1	O1	-
					O2	Asp118 $O^{\delta2}$, Trp287 $N^{\varepsilon1}$
					O3	Asp70 $O^{\delta1}$, Gly286 N
					O4	Asp70 $O^{\delta2}$, Arg356 $N^{\eta2}$
					O5	Trp248 $N^{\varepsilon1}$
					O6	Glu230 $O^{\varepsilon2}$
				Glc2	O1	-
					O2	Ala44 O
					O3	Ala44 O
					O4	Ser12 N & O^{γ}
					O5	-
					O6	Ser12 O^{γ}
	Maltose (subsites A-B)	-10.22	20	Glc1	O1	Trp287 $N^{\varepsilon1}$
					O2	Asp118 $O^{\delta2}$, Gly286 N
					O3	Asp70 $O^{\delta1}$, Gly286 N
					O4	Asp70 $O^{\delta2}$
					O5	Trp248 $N^{\varepsilon1}$
					O6	Glu230 $O^{\varepsilon2}$, Trp248 $N^{\varepsilon1}$
				Glc2	O1	Trp287 $N^{\varepsilon1}$
					O2	Asp118 $O^{\delta2}$
					O3	Arg323 N^{ε} & $N^{\eta2}$
					O4	Ser12 N & O^{γ}, Arg323 $N^{\eta2}$
					O5	-
					O6	Ala44 O
TTHA1652	Maltose (subsites A-B)	-6.85	11	Glc1	O1	-
					O2	Gln269 $N^{\varepsilon2}$, Arg305 $N^{\eta1}$
					O3	His8 $N^{\delta1}$, Arg305 $N^{\eta2}$

					O4	-
					O5	-
					O6	-
				Glc2	O1	-
					O2	Asp155 $O^{\delta 2}$
					O3	-
					O4	-
					O5	-
					O6	Glu113 $O^{\varepsilon 2}$, Gln269 $N^{\varepsilon 2}$
	Maltose (subsites B-C)	-7.56	11	Glc1	O1	Glu113 $O^{\varepsilon 2}$
					O2	Glu113 $O^{\varepsilon 2}$, Gln269 $N^{\varepsilon 2}$
					O3	-
					O4	-
					O5	-
					O6	-
				Glc2	O1	-
					O2	Asp65 $O^{\delta 1}$
					O3	Asp65 $O^{\delta 2}$, Trp66 $N^{\varepsilon 1}$
					O4	-
					O5	-
					O6	Asp155 $O^{\delta 1}$, Tyr157 N
	Maltotriose (subsites A-C)	-10.23	16	Glc1	O1	-
					O2	His8 $N^{\delta 1}$, Arg305 $N^{\eta 2}$
					O3	Arg305 $N^{\eta 1}$
					O4	-
					O5	-
					O6	-
				Glc2	O1	-
					O2	Glu113 $O^{\varepsilon 2}$
					O3	-
					O4	-
					O5	-
					O6	Asp115 $O^{\delta 1}$
				Glc3	O1	-
					O2	Asp65 $O^{\delta 1}$
					O3	Asp65 $O^{\delta 2}$, Trp66 $N^{\varepsilon 1}$
					O4	-
					O5	-
					O6	Asp155 $O^{\delta 1}$, Tyr157 N
	Maltotetraose (subsites A-D)	-10.00	20	Glc1	O1	-
					O2	His8 $N^{\delta 1}$, Arg305 $N^{\eta 2}$
					O3	-
					O4	-
					O5	-
					O6	-
				Glc2	O1	-
					O2	Glu113 $O^{\varepsilon 2}$
					O3	-

					O4	-
					O5	-
					O6	Asp155 $O^{\delta1}$ & $O^{\delta2}$
				Glc3	O1	-
					O2	Asp65 $O^{\delta1}$ & $O^{\delta2}$
					O3	-
					O4	-
					O5	-
					O6	Asp155 $O^{\delta1}$, Tyr157 N
				Glc4	O1	-
					O2	-
					O3	Gly346 N
					O4	-
					O5	-
					O6	-
	γ-cyclodextrin (subsites A-D)	-8.32	11	Glc1	O1	-
					O2	His8 $N^{\delta1}$
					O3	-
					O4	-
					O5	Trp235 $N^{\varepsilon1}$
					O6	-
				Glc2	O1	-
					O2	Glu113 $O^{\varepsilon2}$
					O3	-
					O4	-
					O5	-
					O6	-
				Glc3	O1	-
					O2	Asp65 $O^{\delta1}$
					O3	Asp65 $O^{\delta2}$, Trp66 $N^{\varepsilon1}$
					O4	-
					O5	-
					O6	Asp155 $O^{\delta1}$, Tyr157 N
				Glc4	O1	-
					O2	-
					O3	-
					O4	-
					O5	-
					O6	-
				Glc5	O1	-
					O2	Ala40 O
					O3	Ala40 O
					O4	-
					O5	-
				Glc6	O1	-
					O2	-
					O3	-
					O4	-

					O5	-
					O6	-
				Glc7	O1	-
					O2	-
					O3	-
					O4	-
					O5	-
					O6	-
				Glc8	O1	-
					O2	-
					O3	His8 O
					O4	-
					O5	-
					O6	-
TTHB082	Tagatose (subsite B)	-5.81	11	-	O1	His181 $N^{\varepsilon 2}$
					O2	Glu117 $O^{\varepsilon 2}$, Arg119 $N^{\eta 2}$
					O3	Glu117 $O^{\varepsilon 2}$, Gly297 N & O
					O4	Thr65 N & $O^{\gamma 1}$, Gly297 O
					O5	Phe295 O
					O6	Arg119 $N^{\eta 1}$
	Glucose (subsite B)	-5.21	11	-	O1	Thr66 $O^{\gamma 1}$
					O2	Thr65 N & $O^{\gamma 1}$, Gly297 N & O
					O3	Gly297 N & O, Glu117 $O^{\varepsilon 2}$
					O4	Glu117 $O^{\varepsilon 2}$
					O5	His181 $N^{\varepsilon 2}$
					O6	Arg119 $N^{\eta 2}$
	Laminaribose (subsites A-B)	-7.11	12	Glc1	O1	-
					O2	-
					O3	Thr65 N & $O^{\gamma 1}$, Thr66 N
					O4	Phe295 O
					O5	-
					O6	Glu117 $O^{\varepsilon 2}$, Arg119 $N^{\eta 1}$
				Glc2	O1	Asn11 $N^{\delta 2}$
					O2	Arg119 $N^{\eta 2}$
					O3	-
					O4	Glu117 $O^{\varepsilon 2}$
					O5	Asn11 $N^{\delta 2}$
					O6	Asn11 N
	**Sophorose	No binding				
	Cellobiose (subsites A-B)	-6.32	14	Glc1	O1	Arg119 $N^{\eta 2}$
					O2	Arg119 $N^{\eta 1}$ & $N^{\eta 2}$, His181 $N^{\varepsilon 2}$
					O3	His181 $N^{\varepsilon 2}$
					O4	Thr65 N, Gly297 O
					O5	-
					O6	Ser298 O^{γ}
				Glc2	O1	Ser12 N
					O2	Glu117 $O^{\varepsilon 1}$

					O3	Glu117 $O^{\varepsilon 2}$
					O4	-
					O5	Asn11 N & $N^{\delta 2}$
					O6	Asn11 $N^{\delta 2}$
TTHA1877	GPC	-4.84	7	-	Glycerol group	Gln212 $N^{\varepsilon 2}$, Thr231 O & $O^{\gamma 1}$
					PO_4 group	Gly11 N, Gln212 $N^{\varepsilon 2}$, Tyr302 O^{η}
					$(CH_3)_3$ group	Trp161
TTHA1936	GPC	-4.82	7	-	Glycerol group	Asp8 O, Ala11 N, Ala233 O, Asn234 O
					PO_4 group	Tyr303 O^{η}, Asp8 N
					$(CH_3)_3$ group	Tyr39
TTHV034	GPC	-4.95	7	-	Glycerol group	Tyr43 O^{η}, Ser249 O^{γ}, Tyr323 O^{η}
					PO_4 group	Gly284 O & N
					$(CH_3)_3$ group	Tyr67, Trp231
TTHA0379	GPC	No binding				
TTHA1301	Adenine	-5.21	4	-	N1	Glu136 $O^{\varepsilon 1}$
					N3	Phe69 N
					N6	Glu136 $O^{\varepsilon 2}$
					N7	Asp191 $O^{\delta 1}$
	Guanine	-6.01	4	-	N2	Asp191 $O^{\delta 2}$
					N3	Asp191 $O^{\delta 1}$
					O6	Phe69 N
					N7	Glu136 $O^{\varepsilon 2}$
	Cytosine	No binding				
	Thymine	No binding				

* EFBE: estimated free energy of binding, NHBOND: number of hydrogen bonds.
** Molecular docking experiment was carried out using the Fischer projection (linear form) of sophorose due to the unavailability of the Haworth projection (cyclic form) in the PubChem database.

Table A.4. List of interacting proteins with sugar ABC transporters in *Thermus thermophilus* HB8. In the database STRING v10, the protein-protein interaction data are available only for ORFs located on the chromosome.

ORF number	*ORF number of the interacting protein	+Protein name	Subunit
TTHA0356	TTHA0355	ABC transporter, permease protein, MalFG family	TMD
	TTHA0354	ABC transporter, permease protein, MalFG family	TMD
	TTHA0685	Sugar ABC transporter, permease protein	TMD
	TTHA1651	Maltose ABC transporter, permease protein	TMD
	TTHA0579	Sugar ABC transporter, ATP-binding protein	NBD
	TTHA0378	Sugar ABC transporter, permease protein	TMD
	TTHA1630	Iron ABC transporter, ATP-binding protein	NBD
	TTHA1424	ABC transporter ATP-binding protein (CycB)	NBD
	TTHA0975	Sugar ABC transporter, ATP-binding protein	NBD
	TTHA0715	ABC transporter ATP-binding protein	NBD
TTHA0979	TTHA0977	Sugar ABC transporter, permease protein	TMD
	TTHA0976	Sugar ABC transporter, permease protein	TMD
	TTHA0975	Sugar ABC transporter, ATP-binding protein	NBD
	TTHA0978	Uncharacterized protein	-
	TTHA1630	Iron ABC transporter, ATP-binding protein	NBD
	TTHA1424	ABC transporter ATP-binding protein (CycB)	NBD
	TTHA0715	ABC transporter ATP-binding protein	NBD
	TTHA0579	Sugar ABC transporter, ATP-binding protein	NBD
	TTHA1827	Sugar ABC transporter permease protein	TMD
	TTHA1651	Maltose ABC transporter, permease protein	TMD
TTHA1652	TTHA1650	Maltose ABC transporter, permease protein	TMD
	TTHA1651	Maltose ABC transporter, permease protein	TMD
	TTHA0579	Sugar ABC transporter, ATP-binding protein	NBD
	TTHA0975	Sugar ABC transporter, ATP-binding protein	NBD
	TTHA1630	Iron ABC transporter, ATP-binding protein	NBD
	TTHA1424	ABC transporter, ATP-binding protein (CycB)	NBD
	TTHA0715	ABC transporter ATP-binding protein	NBD
	TTHA1827	Sugar ABC transporter permease protein	TMD
	TTHA0355	ABC transporter, permease protein, MalFG family	TMD
	TTHA0377	Sugar ABC transporter, permease protein	TMD
TTHA0688	TTHA0685	Sugar ABC transporter, permease protein	TMD
	TTHA0579	Sugar ABC transporter, ATP-binding protein	NBD
	TTHA0975	Sugar ABC transporter, ATP-binding protein	NBD
	TTHA0687	Sugar ABC transporter, permease protein [N-terminal]	TMD
	TTHA0686	Sugar ABC transporter, permease protein [C-terminal]	TMD
	TTHA0689	Putative xylose repressor [C-terminal]	-
	TTHA1886	ABC transporter, permease protein	TMD
	TTHA0976	Sugar ABC transporter, permease protein	TMD

	TTHA0378	Sugar ABC transporter, permease protein	TMD
	TTHA0354	ABC transporter, permease protein, MalFG family	TMD
**TTHA0379	TTHA0378	Sugar ABC transporter, permease protein	TMD
	TTHA1650	Maltose ABC transporter, permease protein	TMD
	TTHA0579	Sugar ABC transporter, ATP-binding protein	NBD
	TTHA0379	Sugar ABC transporter, periplasmic sugar-binding protein	SBP
	TTHA0975	Sugar ABC transporter, ATP-binding protein	NBD
	TTHA1652	Maltose ABC transporter, periplasmic maltose-binding protein	SBP
	TTHA0354	ABC transporter, permease protein, MalFG family	TMD
	TTHA0685	Sugar ABC transporter, permease protein	TMD
	TTHA1886	ABC transporter, permease protein	TMD
	TTHA1877	Sugar ABC transporter, periplasmic sugar-binding protein	SBP
TTHA1877	TTHA0377	Sugar ABC transporter, permease protein	TMD
	TTHA0378	Sugar ABC transporter, permease protein	TMD
	TTHA1827	Sugar ABC transporter permease protein	TMD
	TTHA1886	ABC transporter, permease protein	TMD
	TTHA1651	Maltose ABC transporter, permease protein	TMD
	TTHA0977	Sugar ABC transporter, permease protein	TMD
	TTHA0685	Sugar ABC transporter, permease protein	TMD
	TTHA1650	Maltose ABC transporter, permease protein	TMD
	TTHA0354	ABC transporter, permease protein, MalFG family	TMD
	TTHA0355	ABC transporter, permease protein, MalFG family	TMD
TTHA1936	TTHA0377	Sugar ABC transporter, permease protein	TMD
	TTHA0378	Sugar ABC transporter, permease protein	TMD
	TTHA1827	Sugar ABC transporter permease protein	TMD
	TTHA1886	ABC transporter, permease protein	TMD
	TTHA1934	Uncharacterized protein	-
	TTHA1651	Maltose ABC transporter, permease protein	TMD
	TTHA0685	Sugar ABC transporter, permease protein	TMD
	TTHA0977	Sugar ABC transporter, permease protein	TMD
	TTHA0354	ABC transporter, permease protein, MalFG family	TMD
	TTHA1650	Maltose ABC transporter, permease protein	TMD

[+] Protein names are according to UniProtKB database.

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

** In this case, the interaction network was generated using the TMD (ORF ID: TTHA0377) rather than the SBP (ORF ID: TTHA0379) subunit.

Table A.5. Genes associated with carbohydrate metabolism in *Thermus thermophilus* HB8.

Pathway	ORF number	[+]Protein name	E.C. number	KO identifier	PDB ID	*Template
Glycolysis	TTHA0299	Glucokinase	2.7.1.2	K00845	3VOV	-
	TTHA0277	Glucose-6-phosphate isomerase	5.3.1.9	K01810	1ZZG	-
	TTHA1962	6-phosphofructokinase	2.7.1.11	K00850	-	-
	TTHA1446	Fructose 1,6-bisphosphatase II	3.1.3.11	K02446	-	-
	TTHA1773	Fructose-1,6-bisphosphate aldolase	4.1.2.13	K01624	-	-
	TTHA0980	Fructose-1,6-bisphosphate aldolase/phosphatase	3.1.3.11, 4.1.2.13	K01622	-	-
	TTHA0947	Triosephosphate isomerase	5.3.1.1	K01803	1YYA	-
	TTHA0905	Glyceraldehyde 3-phosphate dehydrogenase	1.2.1.12	K00134	-	-
	TTHA0906	Phosphoglycerate kinase	2.7.2.3	K00927	1V6S	-
	TTHB012	Phosphoglycerate mutase family protein	5.4.2.12	K15634	-	-
	TTHA0050	Phosphoglycerate mutase	5.4.2.12	K15634	-	-
	TTHA0368	Phosphoglycerate mutase	5.4.2.12	K15634	-	-
	TTHB049	Alpha-ribazole-5'-phosphate phosphatase	5.4.2.12	K15634	1V37	-
	TTHA0116	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	5.4.2.12	K15635	-	-
	TTHA0002	Phosphopyruvate hydratase	4.2.1.11	K01689	-	-
	TTHA0003	Pyruvate kinase	2.7.1.40	K00873	-	-
	TTHA0185	Pyruvate dehydrogenase subunit E1	1.2.4.1	K00163	-	-
	TTHA0939	Pyruvate dehydrogenase E1-subunit alpha	1.2.4.1	K00161	-	-
	TTHA0938	Pyruvate dehydrogenase E1 component subunit beta	1.2.4.1	K00162	-	-
	TTHA0184	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex	2.3.1.12	K00627	2EQ8	-
	TTHA0232	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex	2.3.1.12	K00627	2EQ9	-
	TTHA0287	2-oxoglutarate dehydrogenase E3 component (dihydrolipoamide dehydrogenase)	1.8.1.4	K00382	2EQ7	-
	TTHA0233	Pyruvate dehydrogenase complex, dihydrolipoamide	1.8.1.4	K00382	2EQ6	-

		dehydrogenase E3 component				
	TTHA1955	2-oxoacid--ferredoxin oxidoreductase subunit alpha	1.2.7.11, 1.2.7.3	K00174	-	-
	TTHA1956	2-oxoglutarate ferredoxin oxidoreductase subunit beta	1.2.7.11, 1.2.7.3	K00175	-	-
	TTHA1113	L-lactate dehydrogenase	1.1.1.27	K00016	2V6M	-
	TTHA0865	NAD-dependent aldehyde dehydrogenase	1.2.1.3	K00128	-	-
	TTHA1248	Acetyl-coenzyme A synthetase	6.2.1.1	K01895	-	-
	TTHA1249	Acetyl-coenzyme A synthetase	6.2.1.1	K01895	-	-
	TTHA1250	Acetyl-coenzyme A synthetase	6.2.1.1	K01895	-	-
	TTHA1285	Acetyl-coenzyme A synthetase	6.2.1.1	K01895	-	-
	TTHA0353	Phosphoglucomutase	5.4.2.2	K01835	2Z0F	-
	TTHA0278	Phosphoenolpyruvate carboxykinase	4.1.1.49	K01610	1J3B	-
Pentose phosphate pathway	TTHA0277	Glucose-6-phosphate isomerase	5.3.1.9	K01810	1ZZG	-
	TTHA0106	Ribulose-phosphate 3-epimerase	5.1.3.1	K01783	-	-
	TTHA0108	Transketolase	2.2.1.1	K00615	2E6K	-
	TTHA1066	Translaldolase	2.2.1.2	K00616	1WX0	-
	TTHA1299	Ribose-5-phosphate isomerase A	5.3.1.6	K01807	-	-
	TTHA1186	2-deoxyribose-5-phosphate aldolase	4.1.2.4	K01619	1J2W	-
	TTHA0431	Sugar kinase	2.7.1.15	K00852	-	-
	TTHA0324	Phosphopentomutase	5.4.2.7	K01839	-	-
	TTHA0353	Phosphoglucomutase	5.4.2.2	K01835	2Z0F	-
	TTHA1549	Ribose-phosphate pyrophosphokinase	2.7.6.1	K00948	-	-
	TTHA1637	Ribose-phosphate pyrophosphokinase	2.7.6.1	K00948	5T3O	-
	TTHB072	4-hydroxy-2-oxoglutarate aldolase/2-deydro-3-deoxyphosphogluconate aldolase	4.1.3.42, 4.1.2.14	K01625	2YW3	-
	TTHB079	2-dehydro-3-deoxygluconokinase	2.7.1.45	K00874	1V19	-
	TTHA0152	Tungsten-containing aldehyde:ferredoxin oxidoreductase	1.2.7.5	K03738	-	-
	TTHA0380	Aldehyde:ferredoxin oxidoreductase	1.2.7.5	K03738	-	-

	TTHA0500	Glycerate dehydrogenase/hydroxypyruvate reductase	2.7.1.165	K11529	-	-
	TTHA1773	Fructose-1,6-bisphosphate aldolase	4.1.2.13	K01624	-	-
	TTHA0980	Fructose-1,6-bisphosphate aldolase/phosphatase	3.1.3.11, 4.1.2.13	K01622	-	-
	TTHA1446	Fructose-1,6-bisphosphatase	3.1.3.11	K02446	-	-
	TTHA1962	6-phosphofructokinase	2.7.1.11	K00850	-	-
	TTHV084	Xylulose kinase	2.7.1.17	K00854	-	5VM1
	TTHV085	Xylose isomerase	5.3.1.5	K01805	1BXB	-
Fructose and mannose pathway	TTHA0992	Fructokinase	2.7.1.4	K00847	-	-
	TTHA1345	Mannose-6-phosphate isomerase	5.3.1.8	K01809	-	-
	TTHA1750	Mannose-1-phosphate guanylyl transferase (GDP)/mannose-6-phosphate isomerase	2.7.7.13	K00971	2CU2	-
	TTHA1811	Fuculose-1-phosphate aldolase	4.1.2.17	K01628	2FK5	-
	TTHB077	Short chain dehydrogenase/reductase family oxidoreductase	1.1.1.-	K18335	-	-
	TTHA0954	Mannosyl-3-phosphoglycerate synthase	2.4.1.217	K05947	-	-
	TTHA0955	Mannosyl-3-phosphoglycerate phosphatase	3.1.3.70	K07026	-	-
	TTHA1962	6-phosphofructokinase	2.7.1.11	K00850	-	-
	TTHA1446	Fructose-1,6-bisphosphatase	3.1.3.11	K02446	-	-
	TTHA1773	Fructose-1,6-bisphosphate aldolase	4.1.2.13	K01624	-	-
	TTHA0980	Fructose-1,6-bisphosphate aldolase/phosphatase	3.1.3.11, 4.1.2.13	K01622	-	-
	TTHA0947	Triosephosphate isomerase	5.3.1.1	K01803	1YYA	-
Galactose metabolism pathway	TTHA0595	Galactokinase	2.7.1.6	K00849	-	-
	TTHB114	Galactose-1-phosphate uridylyltransferase	2.7.7.12	K00965	-	-
	TTHA0591	UDP-glucose 4-epimerase	5.1.3.2	K01784	2P5U	-
	TTHA0353	Phosphoglucomutase	5.4.2.2	K01835	2Z0F	-
	TTHA0299	Glucokinase	2.7.1.2	K00845	3VOV	-
	TTHB032	Putative β-galactosidase	3.2.1.23	K01190	-	-
	TTHA0886	UDP-galactopyranose mutase	5.4.99.9	K01854	-	-
	TTHB115	α-galactosidase	3.2.1.22	K07407	-	-
	TTHA1962	6-phosphofructokinase	2.7.1.11	K00850	-	-
	TTHA1647	Maltodextrin glucosidase	3.2.1.20	K01187	-	5BN7
	TTHB033	α-glucosidase	3.2.1.20	K01187	-	-
	TTHA0481	Oligo-1,6-glucosidase	3.2.1.20	K01187	-	-

Starch and sucrose metabolism pathway	TTHA1647	Maltodextrin glucosidase	3.2.1.20	K01187	-	5BN7
	TTHB033	α-glucosidase	3.2.1.20	K01187	-	-
	TTHA0481	Oligo-1,6-glucosidase	3.2.1.20	K01187	-	-
	TTHB087	β-glucosidase	3.2.1.21	K05350	4BCE	-
	TTHA0022	Glucose-1-phosphate adenylyltransferase	2.7.7.27	K00975	-	-
	TTHA0018	Glycogen synthase	2.4.1.21	K00703	-	-
	TTHA1902	1,4-alpha-glucan-branching protein	2.4.1.18	K16149	1UFA	-
	TTHA1172	Alpha-glucan phosphorylase	2.4.1.1	K00688	-	-
	TTHA0478	Trehalose synthase	3.2.1.1, 5 .4.99.16	K05343	-	5H2T
	TTHA1563	Neopullulanase	3.2.1.135 , 3.2.1.13 3, 3.2.1.54	K01208	2Z1K	-
	TTHA1261	4-alpha-glucanotransferase	2.4.1.25	K00705	-	-
	TTHA0479	Trehalose-6-phosphate phosphatase	3.1.3.12	K01087	-	1U02
	TTHA0353	Phosphoglucomutase	5.4.2.2	K01835	2Z0F	-
	TTHA0299	Glucokinase	2.7.1.2	K00845	3VOV	-
	TTHA0277	Glucose-6-phosphate isomerase	5.3.1.9	K01810	1ZZG	-
	TTHA0992	Fructokinase	2.7.1.4	K00847	-	-
Inositol metabolism pathway	TTHA0077	Inositol monophophatase family protein	3.1.3.25	K01092	-	-
	TTHA0947	Triosephosphate isomerase	5.3.1.1	K01803	1YYA	-
Glycerophospholipid metabolism pathway	TTHA1740	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase	1.1.1.94	K00057	-	-
	TTHB143	Putative glycerol-3-phohsphate dehydrogenase	1.1.5.3	K00111	-	-
	TTHA1746	Glycerol-3-phosphate acyltransferase	2.3.1.15	K03621	-	-
	TTHA1203	Glycerol-3-phosphate acyltransferase	2.3.1.15	K08591	-	-
	TTHA0927	Hypothetical protein	2.3.1.15	K08591	-	-
	TTHA1348	1-acyl-sn-glycerol-3-phosphate acyltransferase	2.3.1.51	K00655	-	-
	TTHA1044	Diacylglycerol kinase	2.7.1.107	K00901	-	-
	TTHB141	Putative glycerophosphoryldiester phosphodiesterase	3.1.4.46	K01126	-	-
	TTHA0857	Phosphatidatecytidylyltransferase	2.7.7.41	K00981	-	-

[+] Protein names are according to UniProtKB database.
* Used as template for model building in this study.

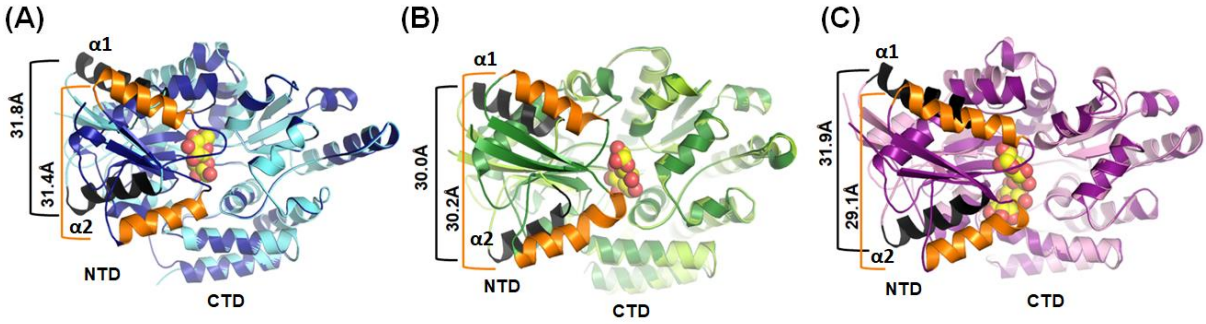


Figure B.1. **Rigid movement of NTD.** Conformational change upon sugar binding in (A) αGlyBP (open-cyan and close-blue) (B) Atu4361 protein (open-lime and close-green) and (C) Acarbose/maltose binding protein GacH (open-magenta and close-purple). Each ligand-bound structure (closed conformation) is superimposed with the unliganded structure (open conformation) at the CTD. Rigid movement of NTD is represented by comparing the distance between the helices α1 and α2 in open (black) and closed (orange) conformation. The bound sugar in all complex structures is shown as a yellow sphere.

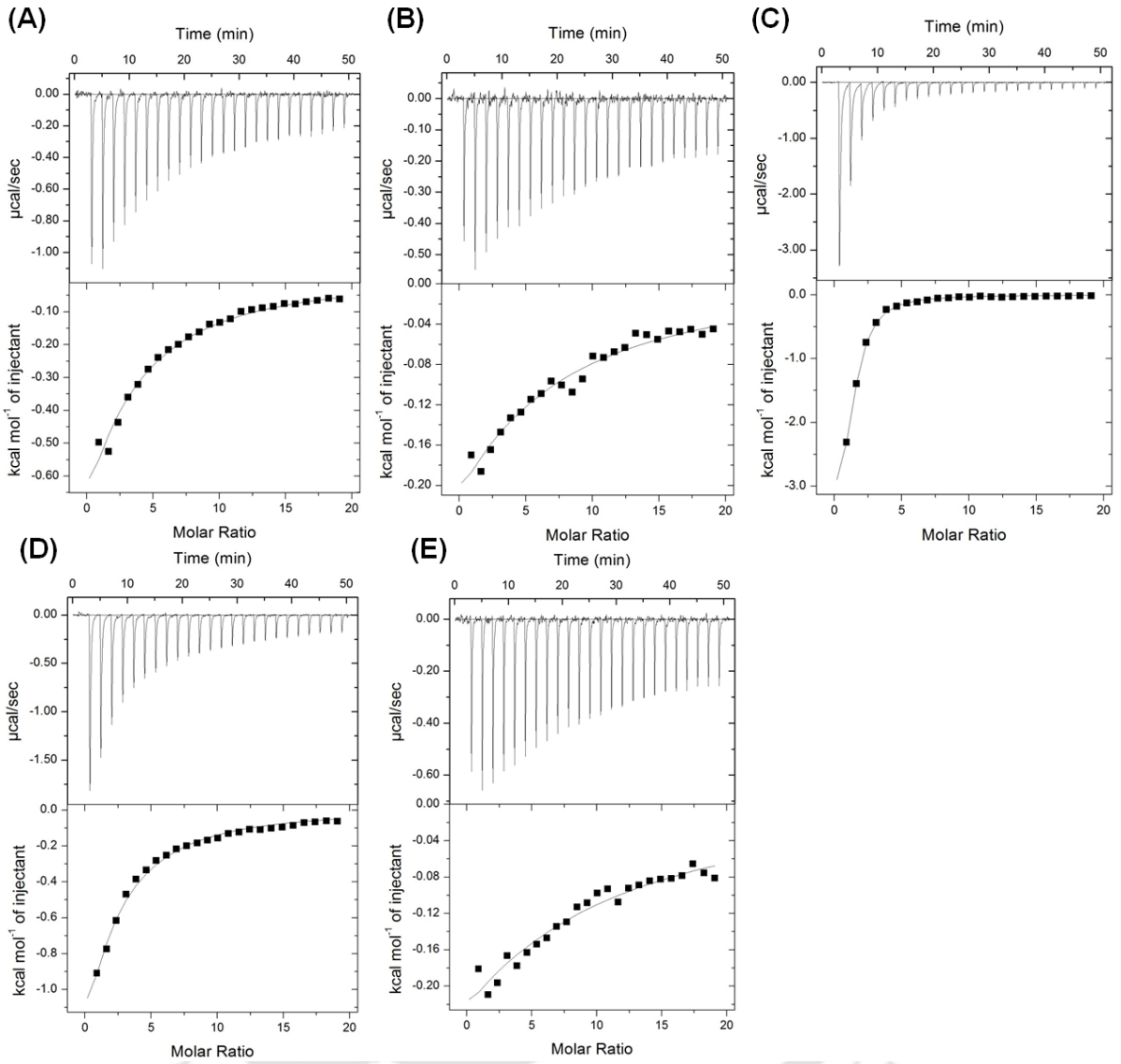


Figure B.2. Binding energetics of disaccharide α-glycosides and glucose with αGlyBP_D70A mutant protein determined by ITC. αGlyBP_D70A mutant protein was titrated with (A) trehalose (B) sucrose (C) maltose (D) palatinose and (E) glucose. In ITC thermogram, upper panel shows raw data (black line) for ligand titration, whereas lower panel shows the integrated and subtracted peak areas (red box) of each titration with the one-site binding model fit (black line). Titrations are replicated at different concentrations for binding isotherm optimization. Thermodynamic parameters obtained from fitting model for individual titration experiments are mentioned in the Table 4.2.

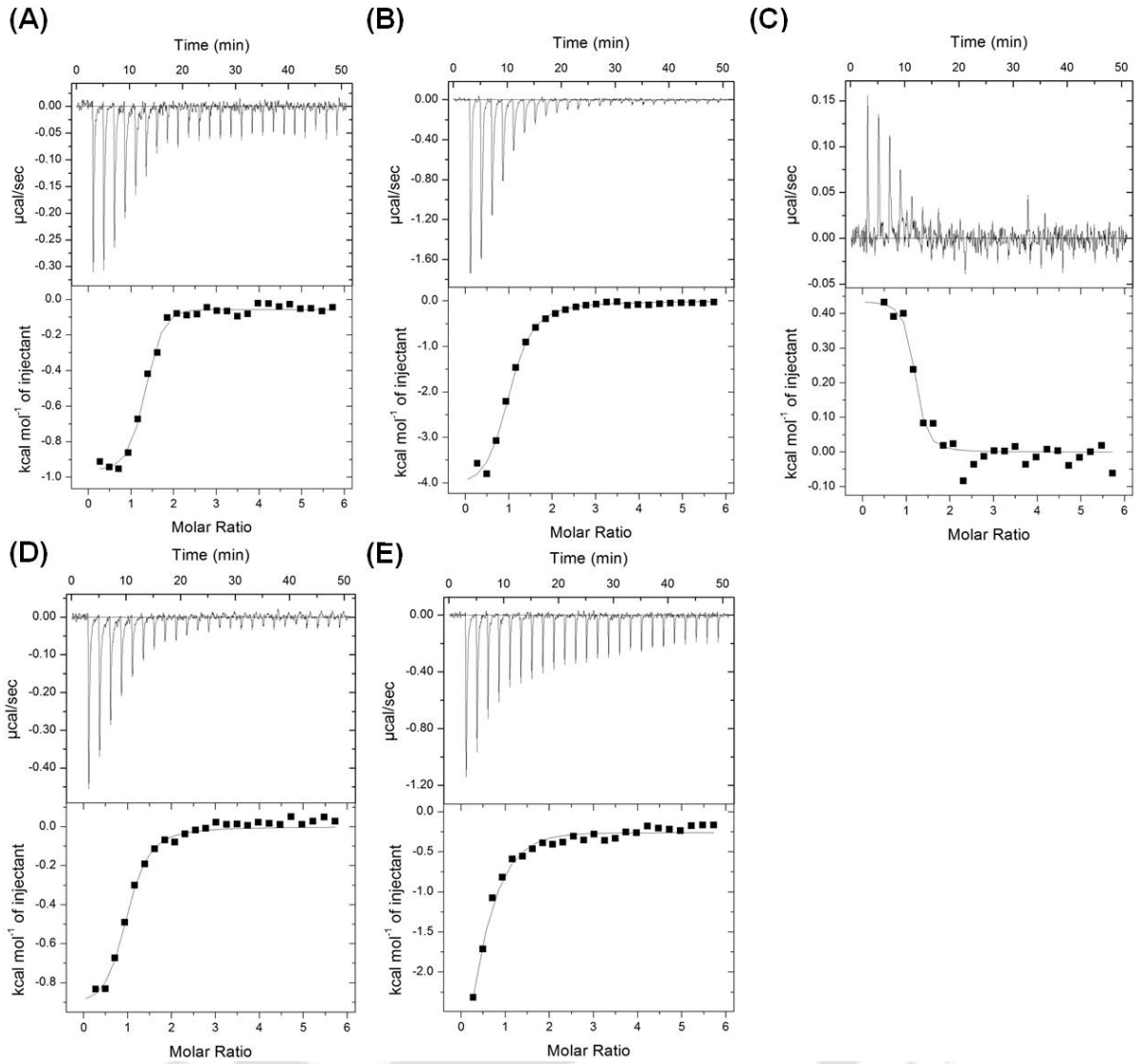


Figure B.3. Binding energetics of disaccharide α-glycosides and glucose with αGlyBP_R356A mutant protein determined by ITC. αGlyBP_R356A mutant protein was titrated with (A) trehalose (B) sucrose (C) maltose (D) palatinose and (E) glucose. In ITC thermogram, upper panel shows raw data (black line) for ligand titration, whereas lower panel shows the integrated and subtracted peak areas (red box) of each titration with the one-site binding model fit (black line). Titrations are replicated at different concentrations for binding isotherm optimization. Thermodynamic parameters obtained from fitting model for individual titration experiments are mentioned in the Table 4.2. The calorimetric profile for maltose binding is different than the other α-glycosides, however binding isotherm is reproducible.

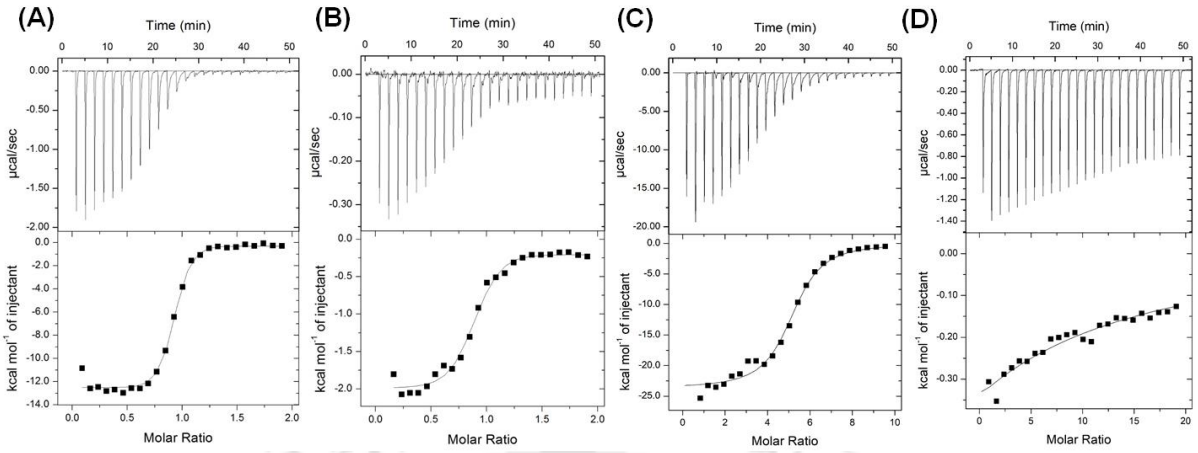


Figure B.4. Binding energetics of metal ions with αGlyBP_D118A mutant protein determined by ITC. αGlyBP_D118A mutant protein was titrated with (A) $CaCl_2$ (B) $MgCl_2$ (C) $ZnCl_2$ and (D) $MnSO_4$. In ITC thermogram, upper panel shows raw data (black line) for metal titration, whereas lower panel shows the integrated and subtracted peak areas (red box) of each titration with the one-site binding model fit (black line). Titrations are replicated at different concentrations for binding isotherm optimization. Thermodynamic parameters obtained from fitting model for individual titration experiments are mentioned in the Table 4.2.

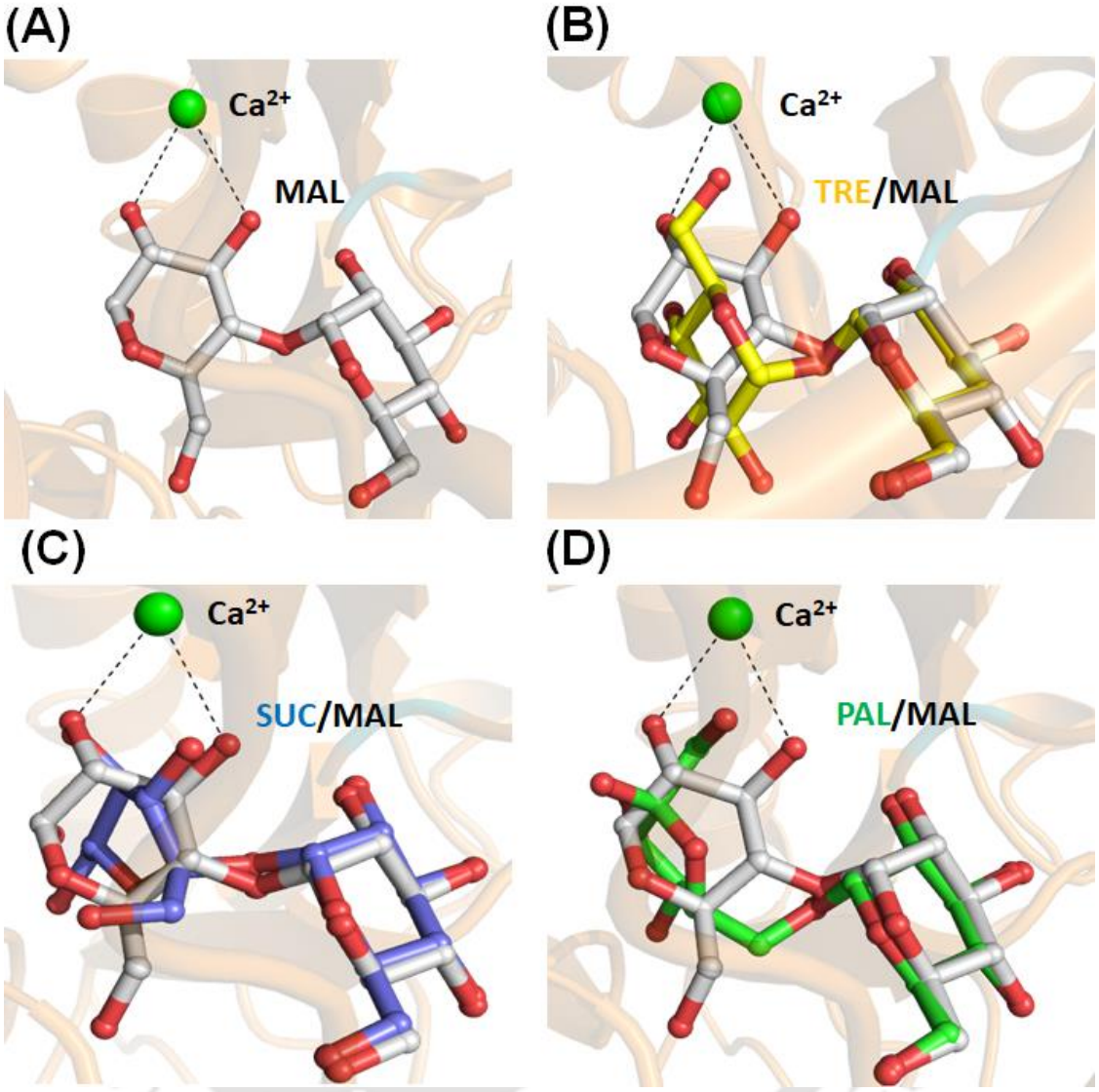


Figure B.5. Ligand selection by αGlyBP_D118A mutant protein. (A) Bidentate interaction of the Ca^{2+} with maltose is represented in dashed lines. Bound Ca^{2+} and maltose are shown as a green sphere and grey ball-and-stick model, respectively. Overlaying of αGlyBP_D118A•MAL complex with (B) αGlyBP_WT•TRE (C) αGlyBP_WT•SUC and (D) αGlyBP_WT•PAL complex demonstrates their inability to form Ca^{2+} bidentate interaction in the presence of trehalose (yellow), sucrose (blue) and palatinose (green).

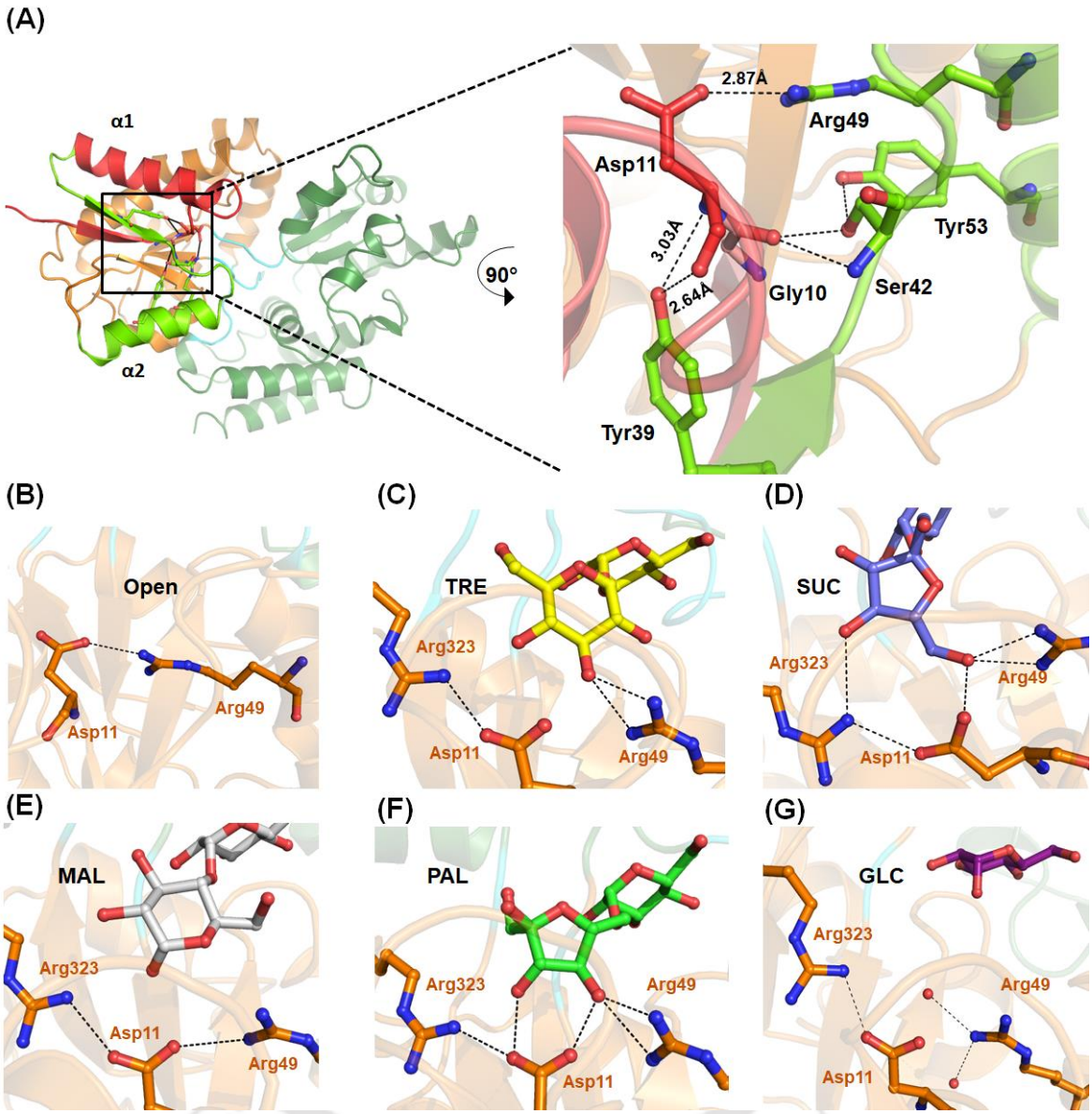


Figure B.6. Role of Arg49 in disaccharide α-glycoside and glucose binding. (A) The overall topology of the open conformation of αGlyBP where α1 (red) and α2 (green) helices of N1 subdomain hold each other at intersecting points (left panel). Close-up view of N1 subdomain depicting the electrostatic and polar interaction involved in stabilization of α1 and α2 helices are shown as dashed lines (right panel). Electrostatic interactions of Arg49 with Asp11 in (B) open conformation (C) trehalose (yellow) (D) sucrose (blue) (E) maltose (grey) (F) palatinose (green) bound structures and with water in (G) glucose (violet) bound structure is represented as dashed lines. All interacting residues and bound ligands are labeled and shown as ball-and-stick model.

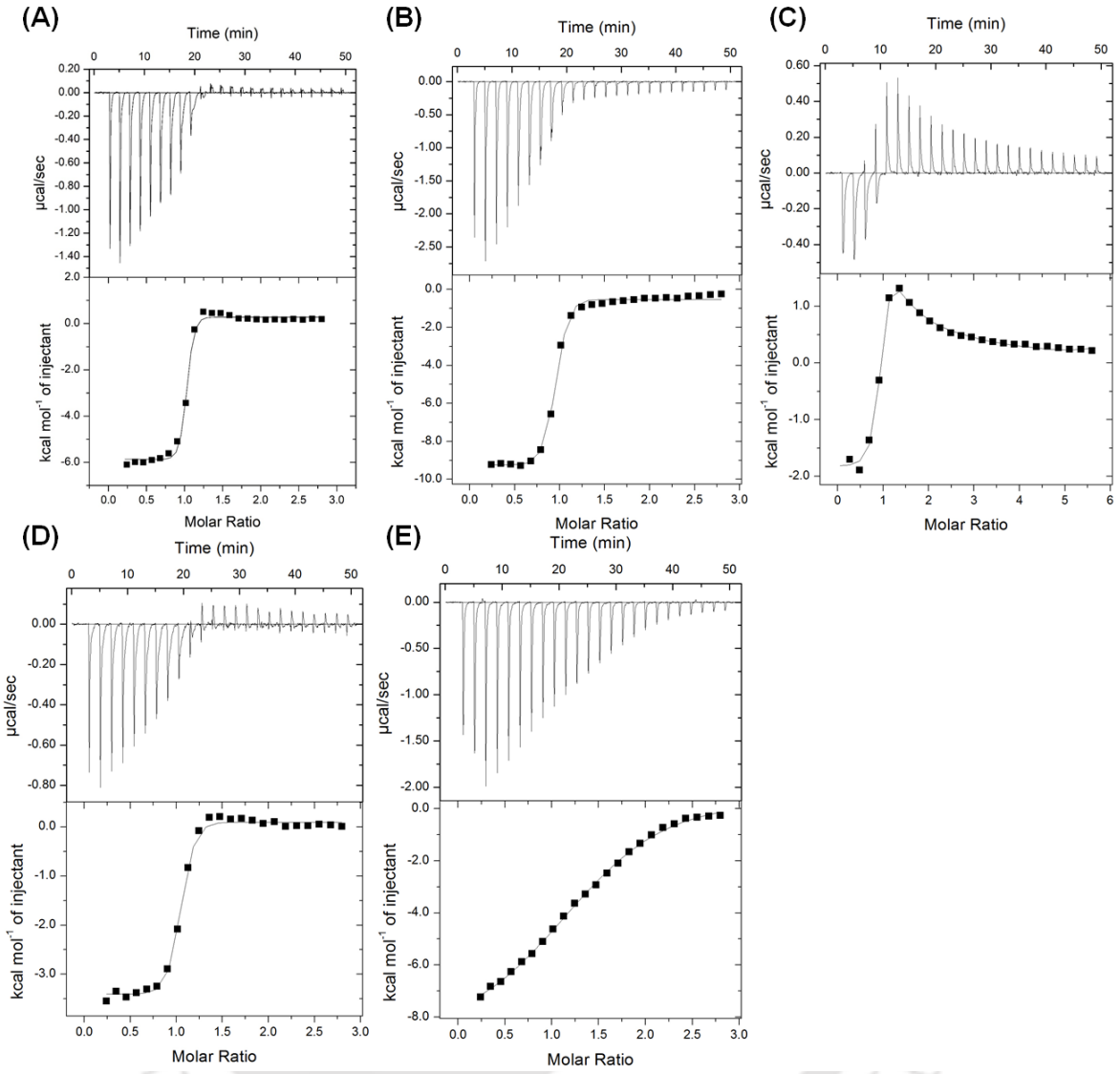


Figure B.7. Binding energetics of disaccharide α-glycosides and glucose with αGlyBP_W287A mutant protein determined by ITC. αGlyBP_W287A mutant protein was titrated with (A) trehalose (B) sucrose (C) maltose (D) palatinose and (E) glucose. In ITC thermogram, upper panel shows raw data (black line) for ligand titration, whereas lower panel shows the integrated and subtracted peak areas (red box) of each titration with the one-site binding model fit (black line). Titrations are replicated at different concentrations for binding isotherm optimization. Thermodynamic parameters obtained from fitting model for individual titration experiments are mentioned in the Table 4.2. In contrast to crystallography data where maltose has one-binding site, thermodynamically it exhibited the two different binding mode with αGlyBP_W287A protein and thus resisted our attempt to fit the calorimetric profile in one-site binding model. Binding isotherm is reproducible and was fitted using two-sitebinding model.

Table B.1. Interactions of αGlyBP_WT, αGlyBP_W287A, αGlyBP _R356A and αGlyBP _R49A mutant proteins with disaccharide α-glycosides.

Ligand		**αGlyBP_WT**	**αGlyBP_W287A**	**αGlyBP _R356A**	**αGlyBP _R49A**
	Atom	Protein / Water (distance in Å)	Protein / Water (distance in Å)	Protein / Water (distance in Å)	Protein / Water (distance in Å)
Trehalose (Glc1)	O2	Asp118 $O^{\delta2}$ (2.7), Gly286 N (3.0), Trp287 $N^{\varepsilon1}$ (3.0)	Asp118 $O^{\delta2}$ (2.7), Gly286 N (3.1), EDO (2.8)	Asp118 $O^{\delta2}$ (2.7), Gly286 N (3.2), Trp287 $N^{\varepsilon1}$ (3.0)	-
	O3	Gly286 N (3.2), Asp70 $O^{\delta1}$ (2.6), Arg356 $N^{\eta2}$ (3.0)	Gly286 N (3.1), Asp70 $O^{\delta1}$ (2.6), Arg356 $N^{\eta2}$ (3.0)	Gly286 N (3.0), Asp70 $O^{\delta1}$ (2.6)	-
	O4	Asp70 $O^{\delta2}$ (2.7), Arg356 $N^{\eta1}$ (2.8)	Asp70 $O^{\delta2}$ (2.7), Arg356 $N^{\eta1}$ (2.8)	Asp70 $O^{\delta2}$ (2.), HOH (2.8)	-
	O5	Trp248 $N^{\varepsilon1}$ (3.1)	Trp248 $N^{\varepsilon1}$ (3.1)	Trp248 $N^{\varepsilon1}$ (3.1)	-
	O6	Glu230 $O^{\varepsilon2}$ (2.7), HOH (2.8)	Glu230 $O^{\varepsilon2}$ (2.7), HOH (2.8)	Glu230 $O^{\varepsilon2}$ (2.8), HOH (2.8)	-
Trehalose (Glc2)	O2	HOH (2.8)	HOH (2.8), HOH (2.8)	HOH (2.9)	-
	O3	Arg49 $N^{\eta1}$ (3.0), Arg49 $N^{\eta2}$ (2.7), Asp11 $O^{\delta1}$ (2.8)	Arg49 $N^{\eta1}$ (3.0), Arg49 $N^{\eta2}$ (2.9), Asp11 $O^{\delta1}$ (2.8)	Arg49 $N^{\eta1}$ (3.0), Arg49 $N^{\eta2}$ (2.8), Asp11 $O^{\delta1}$ (2.8)	-
	O4	Arg323 $N^{\eta1}$ (3.1)	Arg323 $N^{\eta1}$ (3.1)	Arg323 $N^{\eta1}$ (3.1)	-
	O6	Asp118 $O^{\delta2}$ (2.6), HOH (2.5)	Asp118 $O^{\delta2}$ (2.8), HOH (2.7)	Asp118 $O^{\delta2}$ (2.7), HOH (2.6)	-
Maltose (Glc1)	O1	Trp287 $N^{\varepsilon1}$ (3.1)	-	Trp287 $N^{\varepsilon1}$ (3.0)	Trp287 $N^{\varepsilon1}$ (3.1)
	O2	Gly286N (3.0), Trp287 $N^{\varepsilon1}$ (3.1), Asp118 $O^{\delta2}$ (2.7)	Gly286N (2.9), HOH (2.9), Asp118 $O^{\delta2}$ (2.6)	Gly286N (3.0), Trp287 $N^{\varepsilon1}$ (3.0), Asp118 $O^{\delta2}$ (2.7)	Gly286N (3.0), Trp287 $N^{\varepsilon1}$ (3.1), Asp118 $O^{\delta2}$ (2.6)
	O3	Gly286 N (3.2), Asp70$O^{\delta1}$ (2.6), Arg356 $N^{\eta2}$ (3.0)	Gly286 N (3.1), Asp70$O^{\delta1}$ (2.6), Arg356 $N^{\eta2}$ (3.0)	Gly286 N (3.1), Asp70$O^{\delta1}$ (2.7)	Asp70$O^{\delta1}$ (2.6), Arg356 $N^{\eta2}$ (3.0)
	O4	Asp70 $O^{\delta2}$ (2.6), Arg356 $N^{\eta1}$ (2.8)	Asp70 $O^{\delta2}$ (2.7), Arg356 $N^{\eta1}$ (2.8)	Asp70 $O^{\delta2}$ (2.6), Arg356 $N^{\eta1}$ (2.7)	Asp70 $O^{\delta2}$ (2.6), Arg356 $N^{\eta1}$ (2.8)
	O5	Trp287 $N^{\varepsilon1}$ (3.1)	Trp287 $N^{\varepsilon1}$ (3.0)	Trp287 $N^{\varepsilon1}$ (3.1)	Trp287 $N^{\varepsilon1}$ (3.1)
	O6	Glu230 $O^{\varepsilon2}$ (2.7), HOH (2.7)	Glu230 $O^{\varepsilon2}$ (2.8), HOH (2.7)	Glu230 $O^{\varepsilon2}$ (2.7), HOH (2.7)	Glu230 $O^{\varepsilon2}$ (2.7), HOH (2.7)
Maltose (Glc2)	O1	Asp11 $O^{\delta1}$ (2.9)	Asp11 $O^{\delta1}$ (2.8), HOH (2.8)	Asp11 $O^{\delta1}$ (2.9)	Asp11 $O^{\delta1}$ (3.1)

	O2	HOH (2.9)	HOH (2.8), Arg323 $N^{\eta 1}$ (3.1)	Arg323 $N^{\eta 1}$ (3.1), HOH (3.0)	HOH (2.8)
	O3	HOH (3.2), Asp118 $O^{\delta 2}$ (2.7)	HOH (3.0), Asp118 $O^{\delta 2}$ (2.7)	HOH (3.0), Asp118 $O^{\delta 2}$ (2.7)	HOH (3.2), Asp118 $O^{\delta 2}$ (2.6)
	O6	HOH (2.6)	HOH (3.0), HOH (2.7), Ala44 O (2.9)	HOH (3.1), HOH (2.7), Ala44 O (3.2)	HOH (2.8), Ala44 O (2.9)
Palatinose (Glc1)	O1	Trp287 $N^{\varepsilon 1}$ (3.0)		Trp287 $N^{\varepsilon 1}$ (3.0)	-
	O2	Gly286 N (3.0), Trp287 $N^{\varepsilon 1}$ (3.1), Asp118 $O^{\delta 2}$ (2.7)	Gly286 N (3.1), HOH (3.0), Asp118 $O^{\delta 2}$ (2.7)	Trp287 $N^{\varepsilon 1}$ (3.0), Asp118 $O^{\delta 2}$ (2.7)	-
	O3	Gly286 N (3.1), Asp70 $O^{\delta 1}$ (2.6), Arg356 $N^{\eta 2}$ (3.0)	Gly286 N (3.1), Asp70 $O^{\delta 1}$ (2.5), Arg356 $N^{\eta 2}$ (3.0)	Gly286 N (3.0), Asp70 $O^{\delta 1}$ (2.7)	-
	O4	Asp70 $O^{\delta 2}$ (2.6), Arg356 $N^{\eta 1}$ (2.8)	Asp70 $O^{\delta 2}$ (2.7), Arg356 $N^{\eta 1}$ (2.7)	Asp70 $O^{\delta 2}$ (2.6), HOH (2.5)	-
	O5	Trp248 $N^{\varepsilon 1}$ (3.1)	Trp248 $N^{\varepsilon 1}$ (3.1)	Trp248 $N^{\varepsilon 1}$ (3.2)	-
	O6	Glu230 $O^{\varepsilon 2}$ (2.8), HOH (2.7)	Glu230 $O^{\varepsilon 2}$ (2.8), HOH (2.8)	Glu230 $O^{\varepsilon 2}$ (2.7), HOH (2.7)	-
Palatinose (Fru1)	O3	Asp 11$O^{\delta 1}$ (2.6), Arg49 $N^{\eta 1}$ (3.1), Arg49 $N^{\eta 2}$ (2.7)	Arg49 $N^{\eta 1}$ (3.0), Arg49 $N^{\eta 2}$ (2.8)	Asp 11$O^{\delta 1}$ (2.6), Arg49 $N^{\eta 1}$ (3.2), Arg49 $N^{\eta 2}$ (2.8)	-
	O4	Asp11 $O^{\delta 2}$ (2.7), Arg323 $N^{\eta 1}$ (2.8)	Arg323 $N^{\eta 1}$ (2.9)	Asp11 $O^{\delta 2}$ (2.8), Arg323 $N^{\eta 1}$ (2.8)	-
	O5	HOH (2.8)	HOH (2.7)	HOH (2.9)	-
	O6	HOH (2.7), Asp118 $O^{\delta 2}$ (2.6)	HOH (2.6), Asp118 $O^{\delta 2}$ (2.8)	HOH (2.7), Asp118 $O^{\delta 2}$ (2.7)	-
Sucrose (Glc1)	O1	Trp 287 $N^{\varepsilon 1}$ (3.1)	-	Trp 287 $N^{\varepsilon 1}$ (3.1)	-
	O2	Gly286 N (2.9), Trp 287 $N^{\varepsilon 1}$ (3.1), Asp118 $O^{\delta 2}$ (2.6)	Gly286 N (2.9), Trp 287 $N^{\varepsilon 1}$ (3.0), Asp118 $O^{\delta 2}$ (2.7)	Gly286 N (3.0), Trp 287 $N^{\varepsilon 1}$ (3.0), Asp118 $O^{\delta 2}$ (2.6)	-
	O3	Asp 70 $O^{\delta 1}$ (2.6), Arg356 $N^{\eta 2}$ (2.9)	Asp 70 $O^{\delta 1}$ (2.6), Arg356 $N^{\eta 2}$ (2.9), Gly286 N (3.2)	Asp 70 $O^{\delta 1}$ (2.6), HOH (3.1), Gly286 N (3.2)	-
	O4	Asp 70 $O^{\delta 2}$ (2.7), Arg356 $N^{\eta 1}$ (2.8)	Asp 70 $O^{\delta 2}$ (2.7), Arg356 $N^{\eta 1}$ (2.9)	Asp 70 $O^{\delta 2}$ (2.6), HOH (2.7)	-
	O5	Trp248 $N^{\varepsilon 1}$ (3.1)	Trp248 $N^{\varepsilon 1}$ (3.1)	Trp248 $N^{\varepsilon 1}$ (3.1)	-
	O6	Glu230 $O^{\varepsilon 2}$ (2.7), HOH (2.8)	Glu230 $O^{\varepsilon 2}$ (2.7), HOH (2.8)	Glu230 $O^{\varepsilon 2}$ (2.7), HOH (2.7)	-
Sucrose (Fru1)	O1	HOH (2.9)	HOH (2.8)	HOH (2.8)	-
	O3	HOH (3.1), Asp118 $O^{\delta 2}$ (2.7)	HOH (3.0), Asp118 $O^{\delta 2}$ (2.7)	HOH (3.1), Asp118 $O^{\delta 2}$ (2.6)	-
	O4	HOH (2.7)	HOH (2.7), Arg323 $N^{\eta 1}$ (3.2)	HOH (2.8), Arg323 $N^{\eta 1}$ (3.2)	-

	O6	Arg49 $N^{\eta1}$ (2.9), Arg49 $N^{\eta2}$ (2.9), Asp11 $O^{\delta1}$ (2.7)	Arg49 $N^{\eta1}$ (2.8), Arg49 $N^{\eta2}$ (2.9), Asp11 $O^{\delta1}$ (2.8)	Arg49 $N^{\eta1}$ (2.8), Arg49 $N^{\eta2}$ (2.8), Asp11 $O^{\delta1}$ (2.7)	-
Glucose	O1	HOH (2.6)	HOH (2.7), EDO (3.1)	HOH (2.8), Trp287 $N^{\varepsilon1}$ (3.0)	-
	O2	Gly286 N (3.0), Trp287 $N^{\varepsilon1}$ (3.1), Asp118 $O^{\delta2}$ (2.7)	Gly286 N (3.0), EDO (3.1), Asp118 $O^{\delta2}$ (2.7)	Gly286 N (3.0), Trp287 $N^{\varepsilon1}$ (3.0), Asp118 $O^{\delta2}$ (2.7)	-
	O3	Gly286 N (3.2), Asp70 $O^{\delta1}$ (2.6), Arg356 $N^{\eta2}$ (3.0)	Gly286 N (3.1), Asp70 $O^{\delta1}$ (2.6), Arg356 $N^{\eta2}$ (3.0)	Gly286 N (3.1), Asp70 $O^{\delta1}$ (2.7)	-
	O4	Asp70 $O^{\delta2}$ (2.7), Arg356 $N^{\eta1}$ (2.8)	Asp70 $O^{\delta2}$ (2.6), Arg356 $N^{\eta1}$ (2.9)	Asp70 $O^{\delta2}$ (2.7), HOH (2.7)	-
	O5	Trp248 $N^{\varepsilon1}$ (3.2)	Trp248 $N^{\varepsilon1}$ (3.2)	Trp248 $N^{\varepsilon1}$ (3.2)	-
	O6	Glu230 $O^{\varepsilon2}$ (2.7), HOH (2.7)	Glu230 $O^{\varepsilon2}$ (2.7), HOH (2.7)	Glu230 $O^{\varepsilon2}$ (2.7), HOH (2.8)	-

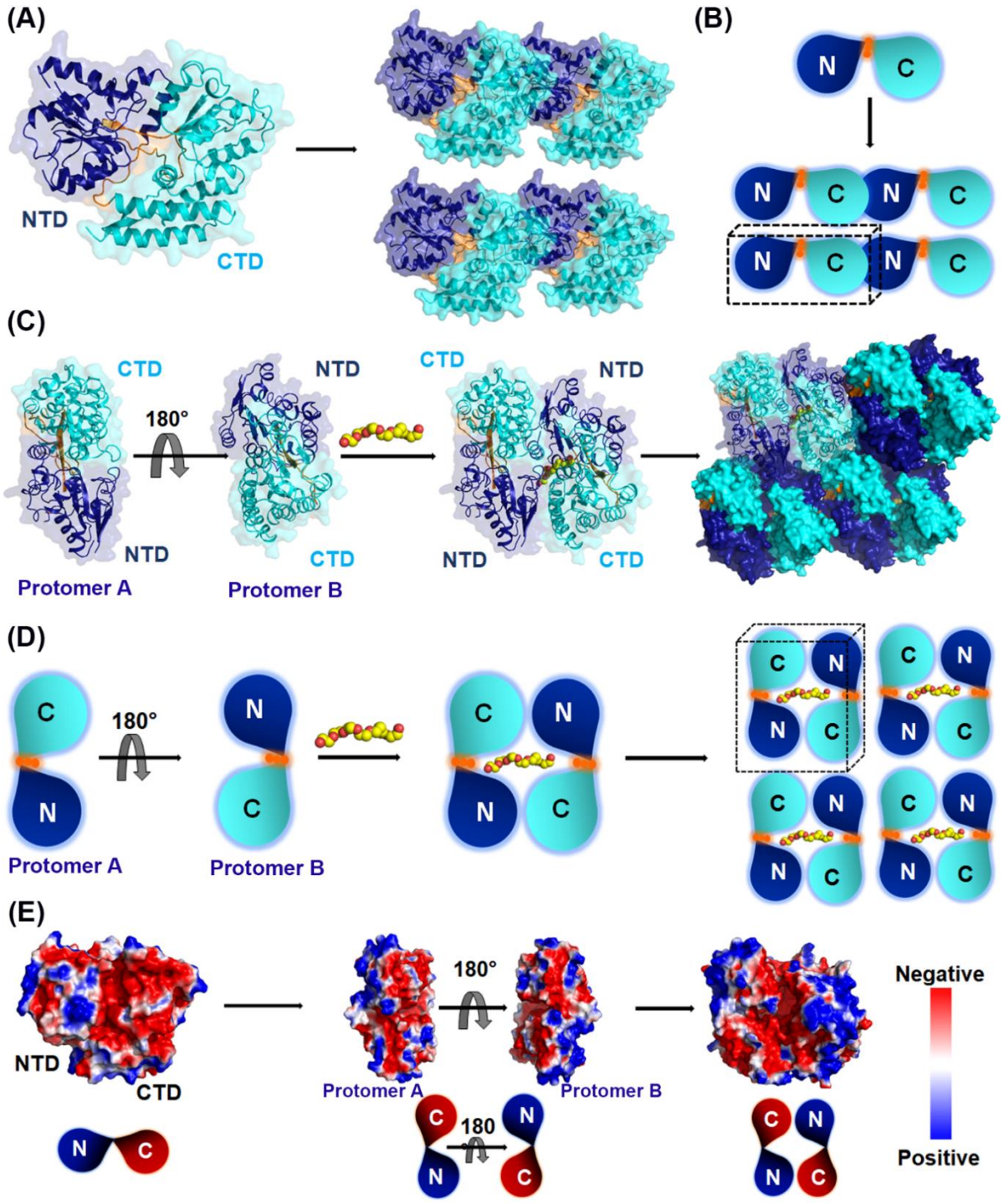


Figure C.1. Crystal packing of the protein βGlyBP in the space groups $P2_12_12_1$ and $P2_1$. (A and B) Surface and schematic representations of the protein βGlyBP showing the molecular arrangement in the space group $P2_12_12_1$. Two domains NTD and CTD of the protein βGlyBP are shown in blue and cyan, respectively. In space group $P2_12_12_1$, the protein molecules are arranged by a non-crystallographic translational symmetry with one

molecule in the asymmetric unit (enclosed with black-dotted lines). (C and D) Surface and schematic representations of the protein βGlyBP depicting the molecular arrangement in the space group $P2_1$. In this space group, two molecules of βGlyBP are arranged by a non-crystallographic rotational symmetry in such a way that the NTD (blue, protomer A) and CTD (cyan, protomer B) are juxtaposed. Both the molecules of the protein βGlyBP are packed together via PEG molecules (yellow sphere). The asymmetric unit having two molecules of the protein βGlyBP is enclosed by black-dotted lines. (E) Surface (top) and schematic (bottom) representations of the protein βGlyBP exhibiting its electrostatic surface charge distribution in the space group $P2_1$. The NTD and CTD are shown in blue and red according to their respective positive and negative potentials.

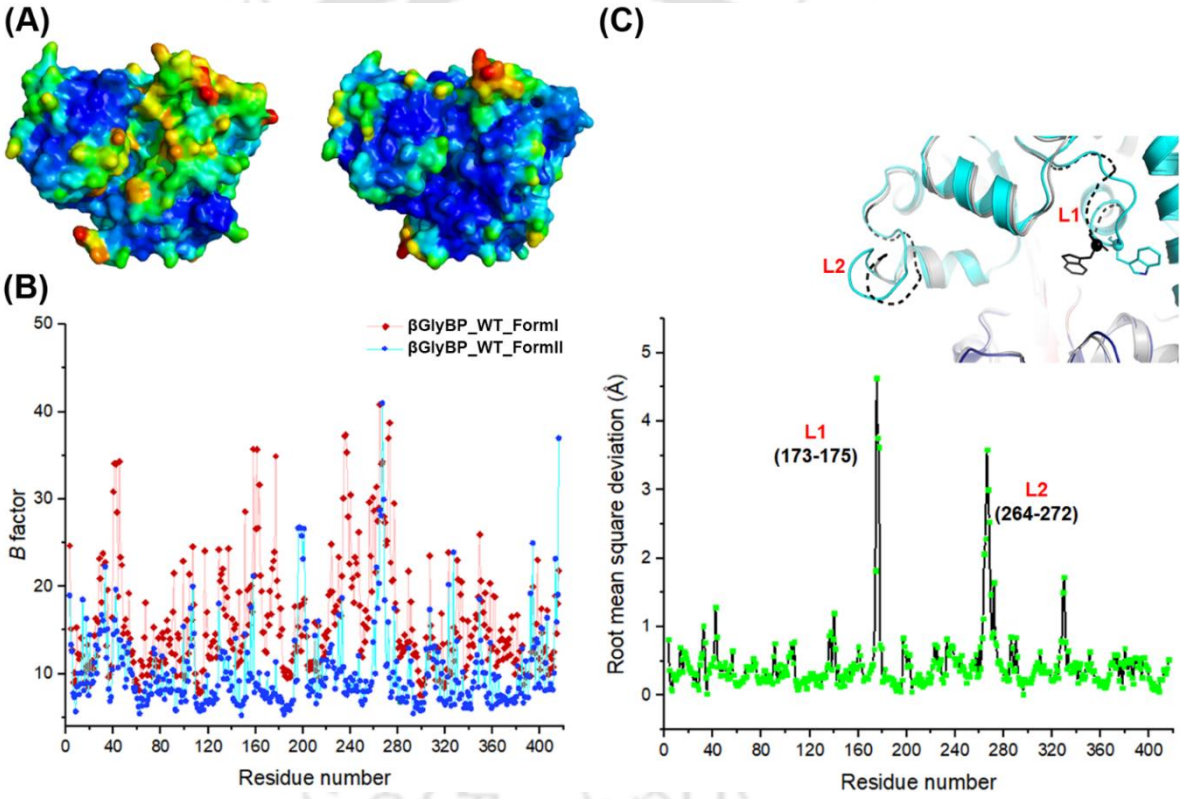


Figure C.2. Structural comparison of βGlyBP_WT_FormI and _FormII. (A) Surface representation of the average *B*-factor of each residue of the two forms βGlyBP_WT_FormI (left) and _FormII (right). (B) The plot showing the average *B*-factor distribution of each residue of the βGlyBP_WT_FormI (red) and _FormII (blue) structures. The X- and Y-axis represents the residue number and average *B*-factor, respectively. (C) The root mean square deviation (rmsd) plot showing the differences in the βGlyBP_WT_FormI and _FormII structures at the C_α atom level. The X- and Y-axis represents the amino acid numbers and rmsd value, respectively.

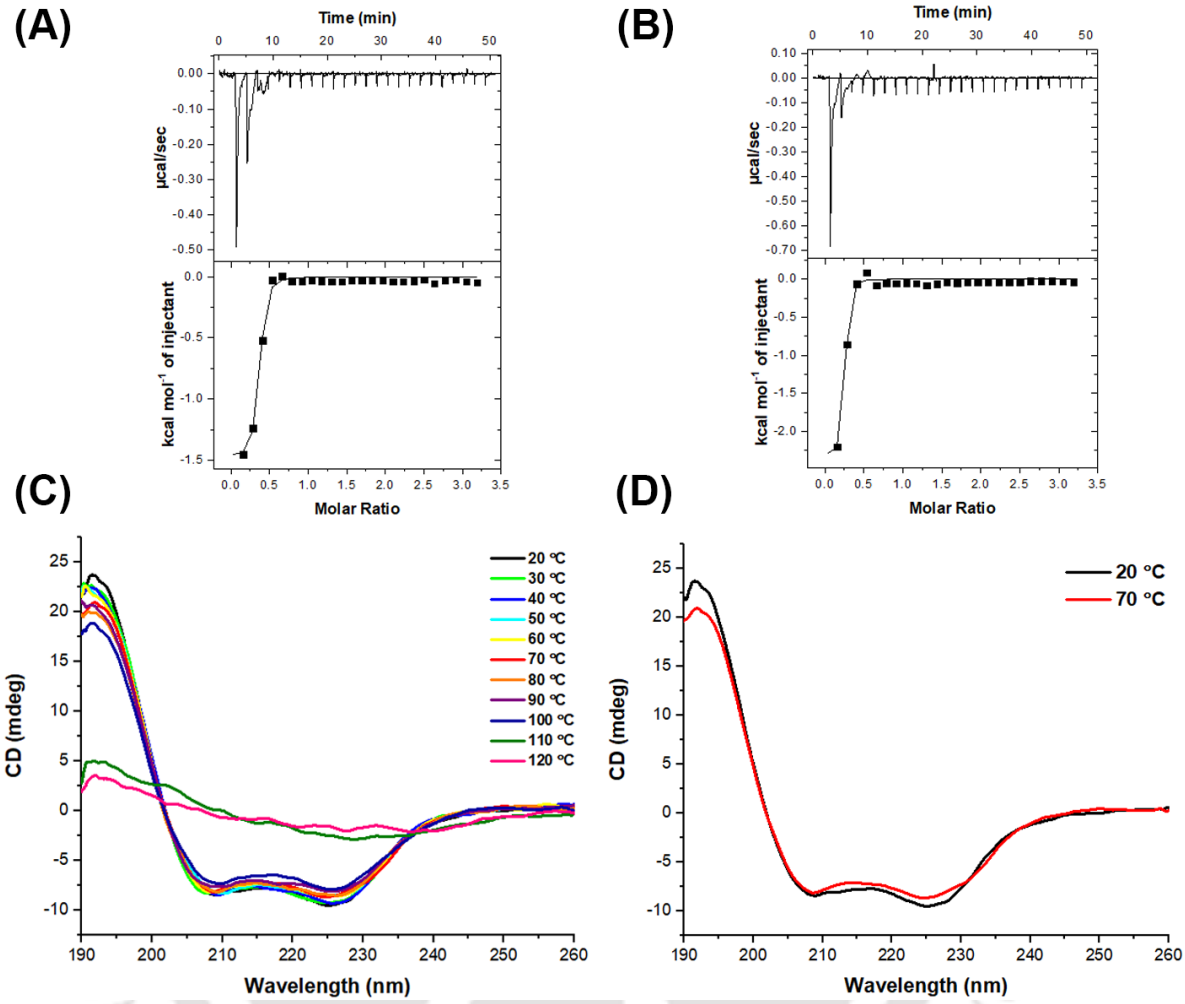


Figure C.3. Thermodynamic data of CEL3 and CEL4 binding to the protein βGlyBP. An initial exothermic heat pulse upon binding of (A) CEL3 and (B) CEL4 to the protein βGlyBP at 25°C. The upper and lower panels represent the raw heat pulse upon ligand titration and the integrated data, respectively. (C) Circular dichroism (CD) spectra of the protein βGlyBP at different temperatures (20-120°C) depicting its effect on the secondary structural contents of the protein. (D) CD spectra of the protein βGlyBP at 20°C (black) and 70°C (red) indicating no change in the secondary structural contents of the protein.

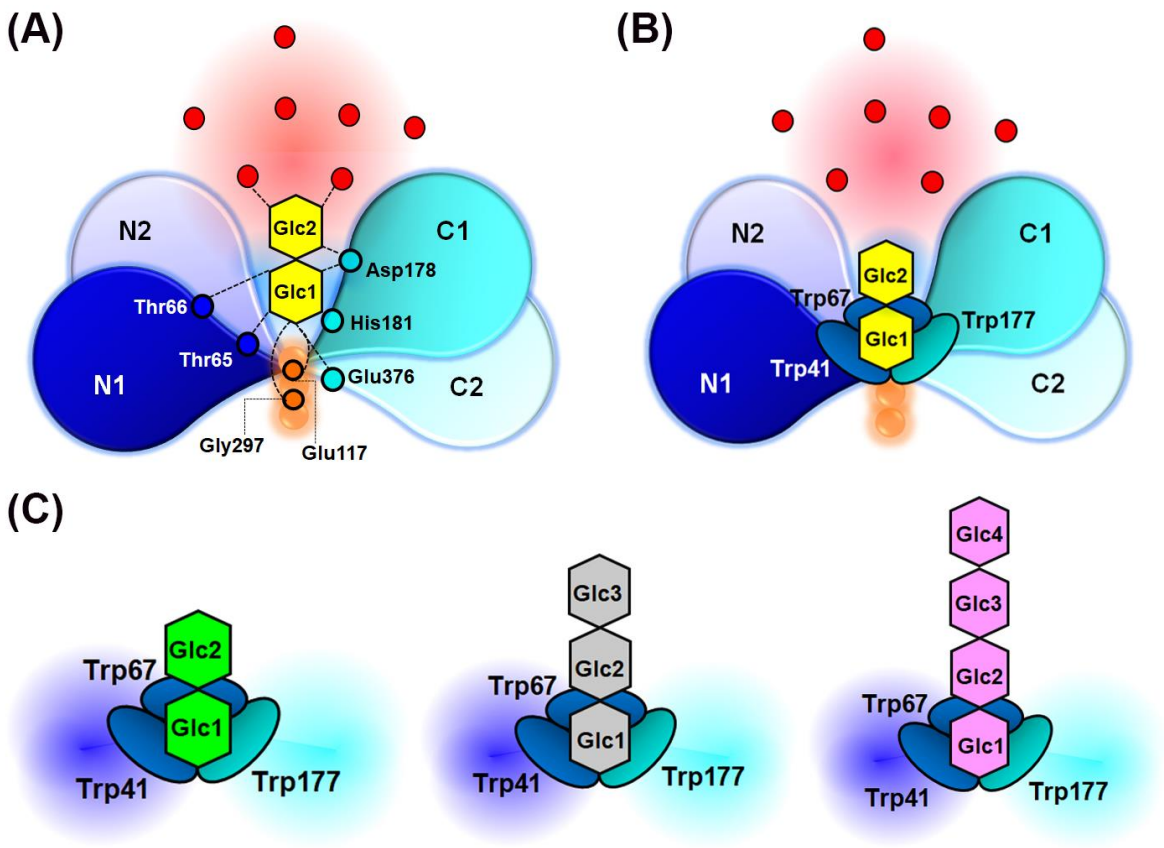


Figure C.4. Schematic representation of the active site of the protein βGlyBP. (A) Active site residues involving polar interactions with β-glycoside (CEL2, yellow) from the NTD, hinge region and CTD are represented in blue, orange and cyan circles, respectively, while the hydrogen bonds are depicted by dotted lines. (B) Aromatic residues forming the hydrophobic cage and stacking interactions with β-glycosides are labeled and depicted with an oval circle. The buried region for nonreducing end (Glc1) is situated far from the solvent exposed region (demarcated in red background). Water molecules are shown as red filled circle. (C) The binding mode of SOP2 (green), CEL3 (grey) and CEL4 (pink) with a hydrophobic cage (demarcated in blue and cyan background) shows the binding of the nonreducing end (Glc1) to the buried region.

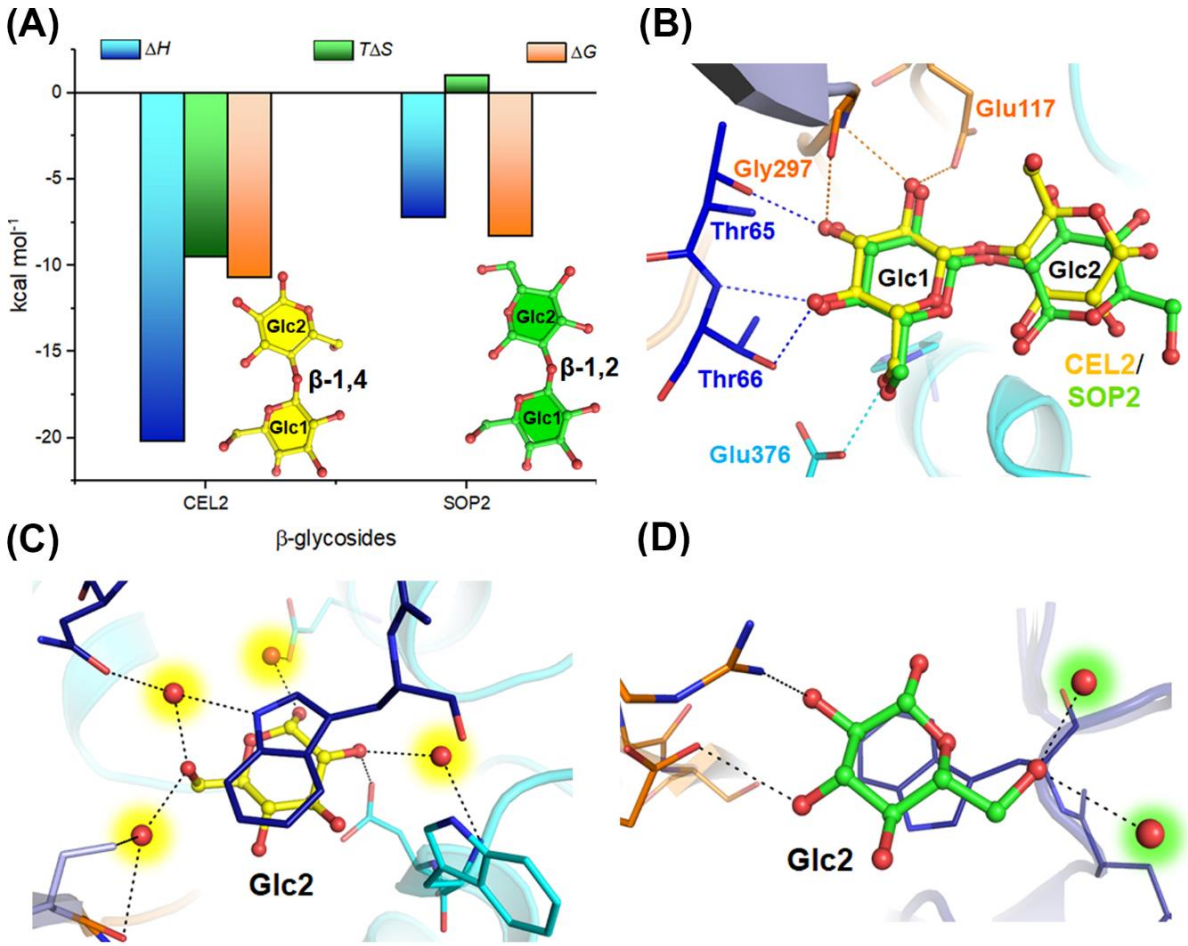


Figure C.5. Water-mediated interactions influencing the ligand-binding affinity. (A) Histograms showing the thermodynamic change in enthalpy (ΔH, blue), entropy (ΔS, green) and free energy (ΔG, orange) for CEL2 and SOP2 binding to the protein βGlyBP. (B) Structural superimposition of protein βGlyBP active site bound to CEL2 (yellow) and SOP2 (green). Hydrogen bonds between the protein and ligand are shown as a dotted line. Water-mediated interactions with the Glc2 unit of (C) CEL2 and (D) SOP2. All participating water molecules are shown as red spheres. Active site residues from NTD, hinge region and CTD forming hydrogen bond(s) are displayed in blue, orange and cyan, respectively.

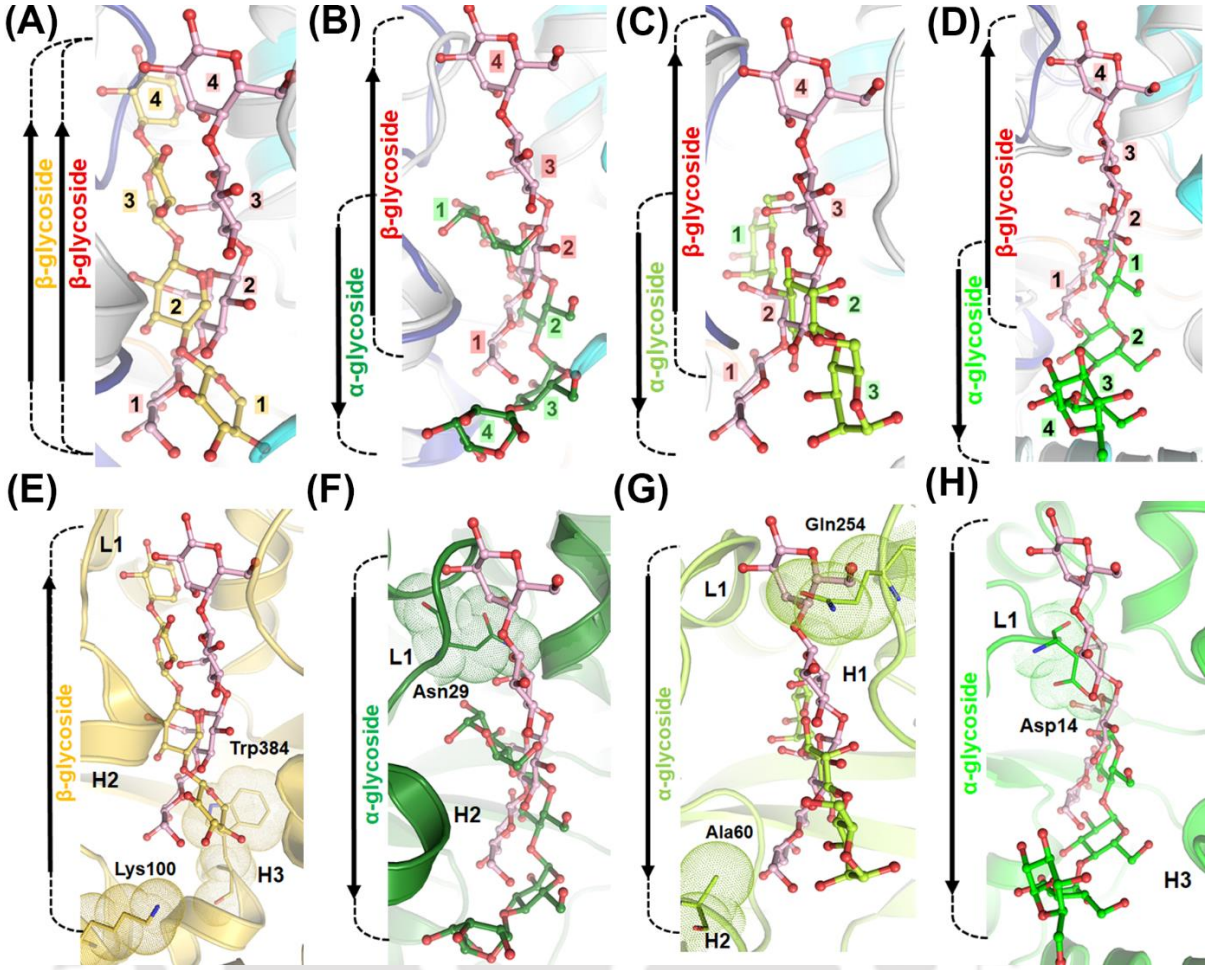


Figure C.6. Schematic representation of the orientation of carbohydrates in subcluster D-I SBPs. Structural superposition of the protein βGlyBP_WT•CEL4 (PDB ID: 7C68) with subcluster D-I SBPs (A) BlAXBP (yellow, PDB ID: 3ZKK), (B) GacH (green, PDB ID: 3K00), (C) CpMnBP1 (limon, PDB id: 4R9G) and (D) MBP (light green, PDB id: 4MBP). The orientation of CEL4 (pink, β-glycoside) at the binding pocket was compared with that of bound (A) xylotetraose (yellow, β-glycoside), (B) maltotetraose (green, α-glycoside), (C) mannotriose (limon, α-glycoside) and (D) maltotetraose (light green, α-glycoside). Each Glc unit is numbered starting from the nonreducing end (Glc1). Active site residues from structural element(s) loop (L1) and helices (H1-H3) that direct the orientation of (E) xylotetraose (yellow, β-glycoside), (F) maltotetraose (green, α-glycoside), (G) mannotriose (limon, α-glycoside) and (H) maltotetraose (light green, α-glycoside) are shown as line and dots. In all the figures, the arrow denotes the orientation of bound ligand. Carbohydrates are shown as ball-and-stick model. Abbreviations: BlAXBP, xylo-oligosaccharide-binding protein; GacH, acarbose/maltose-binding protein; CpMnBP1, mannobiose-binding protein; MBP, maltose-binding protein.

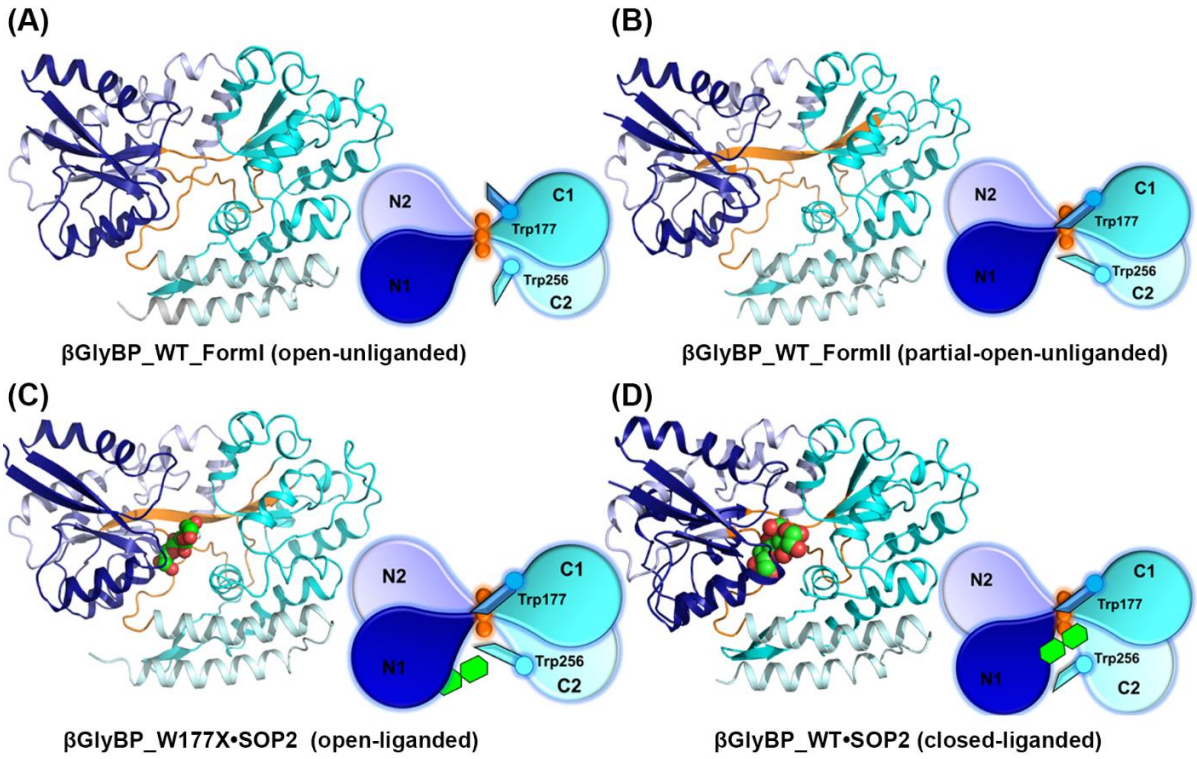


Figure C.7. Structural and schematic representatives of the entrapped configurations of the protein βGlyBP. Four different conformational states representing the postulated two-step ligand-binding mechanism includes open-unliganded state of (A) βGlyBP_WT_FormI, partial-open-unliganded state of (B) βGlyBP_WT_FormII, intermediate transition or open-liganded state of (C) βGlyBP_W177X•SOP2 and closed-liganded state of (D) βGlyBP_WT•SOP2. The first step of the induced-fit mechanism includes the first three conformations while the second step involves the final closed state.

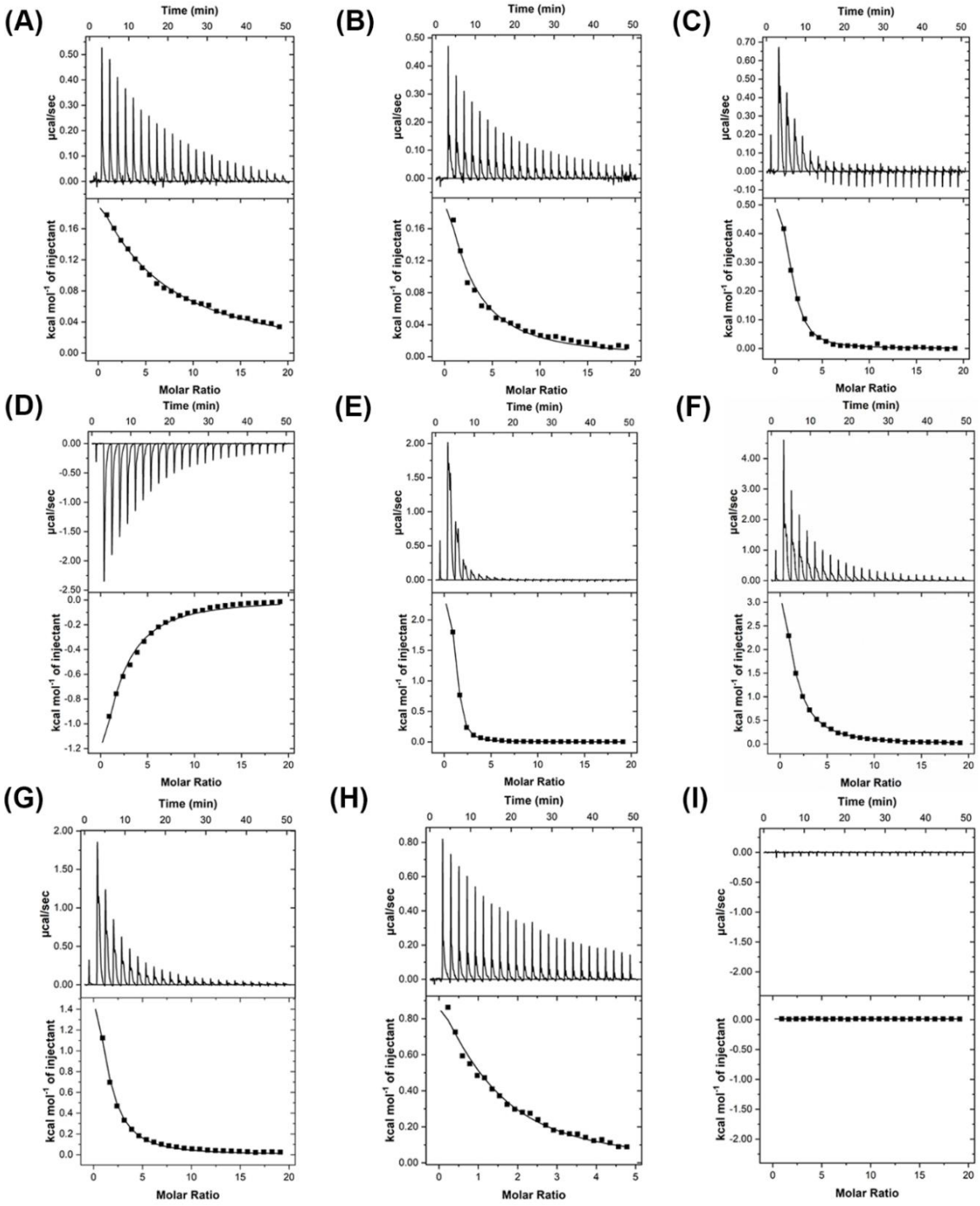


Figure C.8. Thermodynamic data of the mutant βGlyBP_W41A. Binding isotherm of the mutant βGlyBP_W41A titrated with (A) SOP2, (B) LAM2, (C) LAM3, (D) LAM4, (E) CEL2, (F) CEL3, (G) CEL4, (H) CEL5 and (I) GEN2. The upper and lower panels represent the raw heat changes and the integrated heat pulse subtracted with the heat of dilution, respectively. In all the binding isotherms, the integrated data are plotted against the protein:ligand molar ratio and are fitted using a one-site binding model.

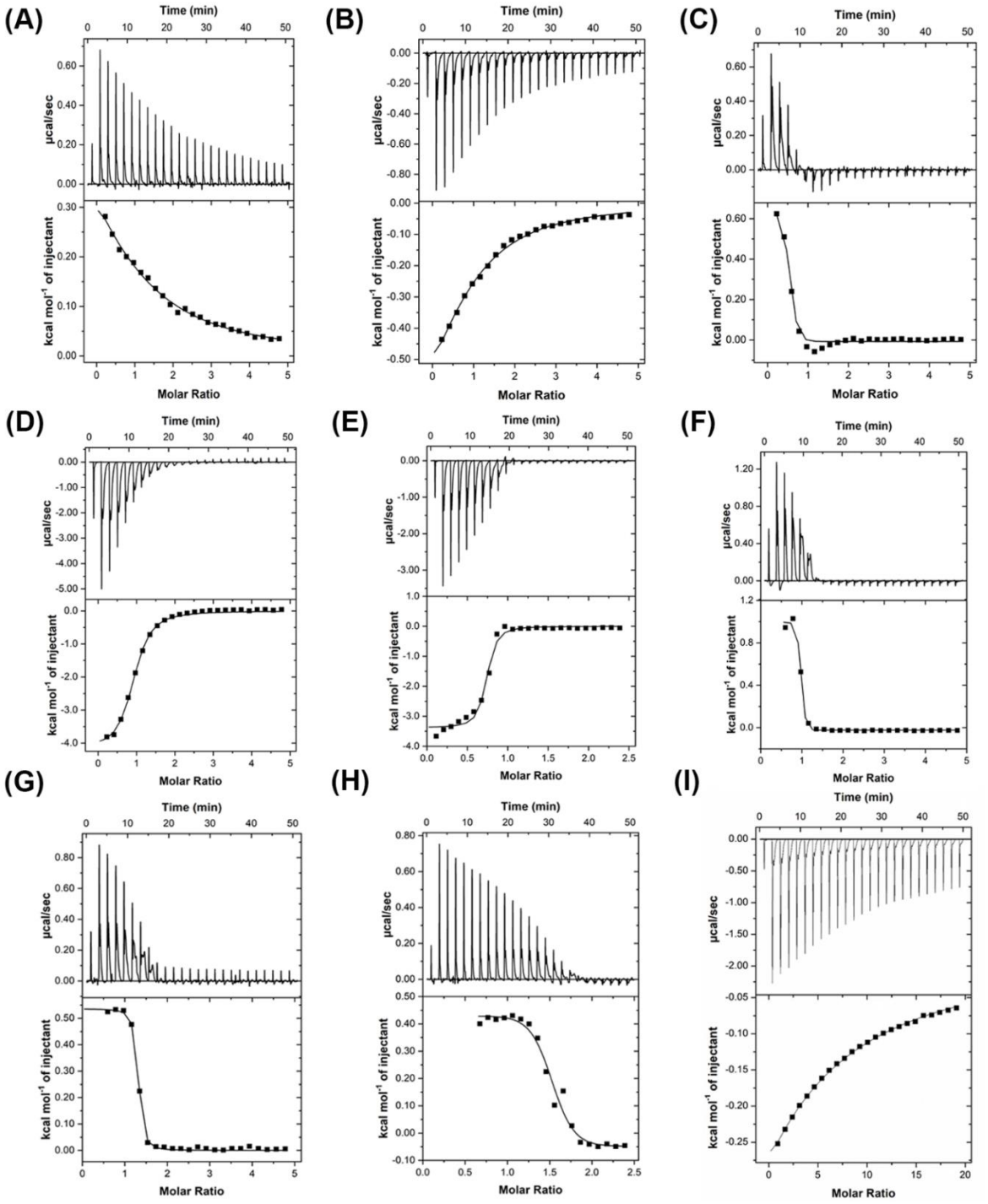


Figure C.9. Thermodynamic data of the mutant βGlyBP_W67A. Binding isotherm of the mutant βGlyBP_W67A titrated with (A) SOP2, (B) LAM2, (C) LAM3, (D) LAM4, (E) CEL2, (F) CEL3, (G) CEL4, (H) CEL5 and (I) GEN2. The upper and lower panels represent the raw heat changes and the integrated heat pulse subtracted with the heat of dilution, respectively. In all the binding isotherms, the integrated data are plotted against the protein:ligand molar ratio and are fitted using a one-site binding model.

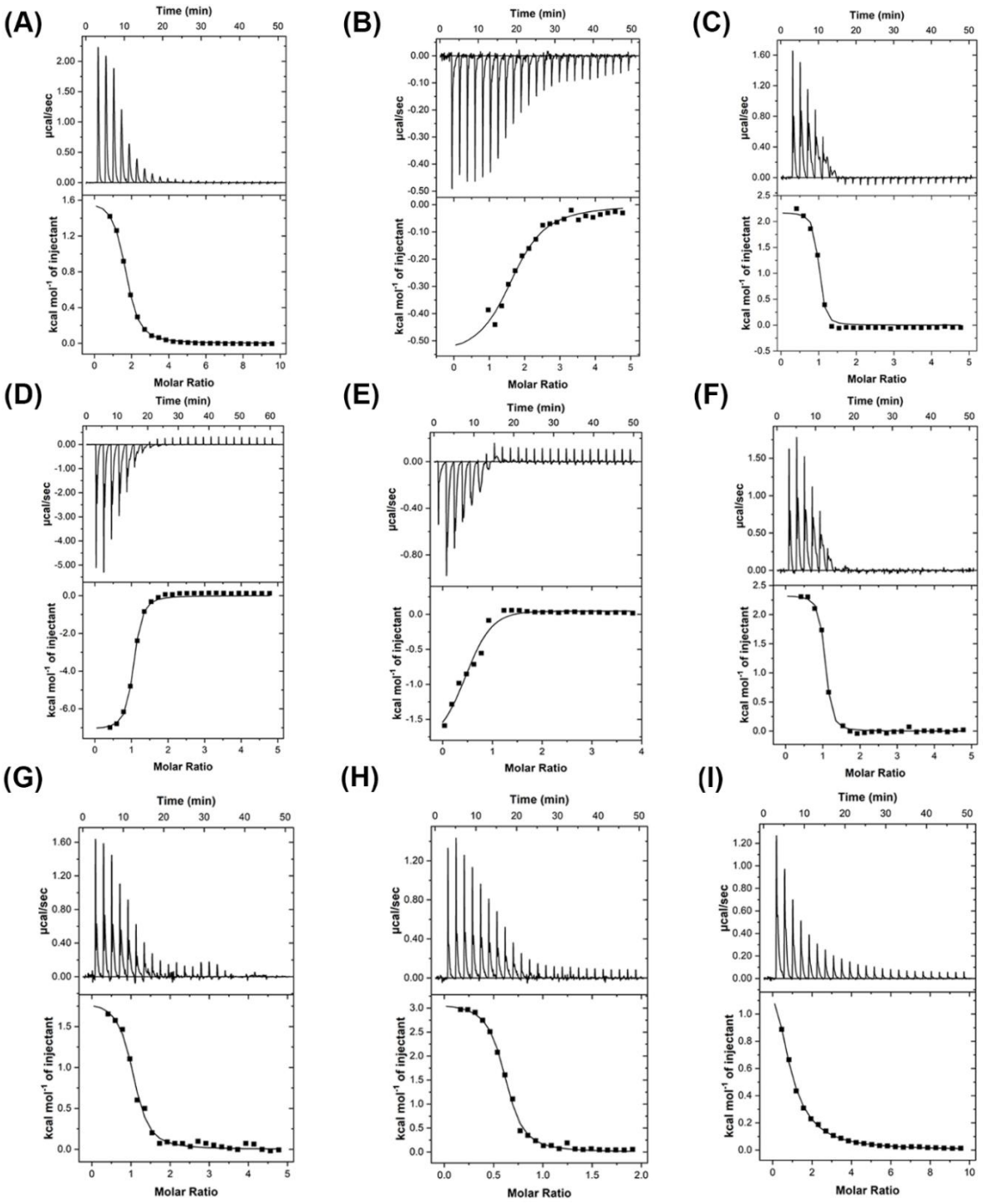


Figure C.10. Thermodynamic data of the mutant βGlyBP_W256A. Binding isotherm of the mutant βGlyBP_W256A titrated with (A) SOP2, (B) LAM2, (C) LAM3, (D) LAM4, (E) CEL2, (F) CEL3, (G) CEL4, (H) CEL5 and (I) GEN2. The upper and lower panels represent the raw heat changes and the integrated heat pulse subtracted with the heat of dilution, respectively. In all the binding isotherms, the integrated data are plotted against the protein:ligand molar ratio and are fitted using a one-site binding model.

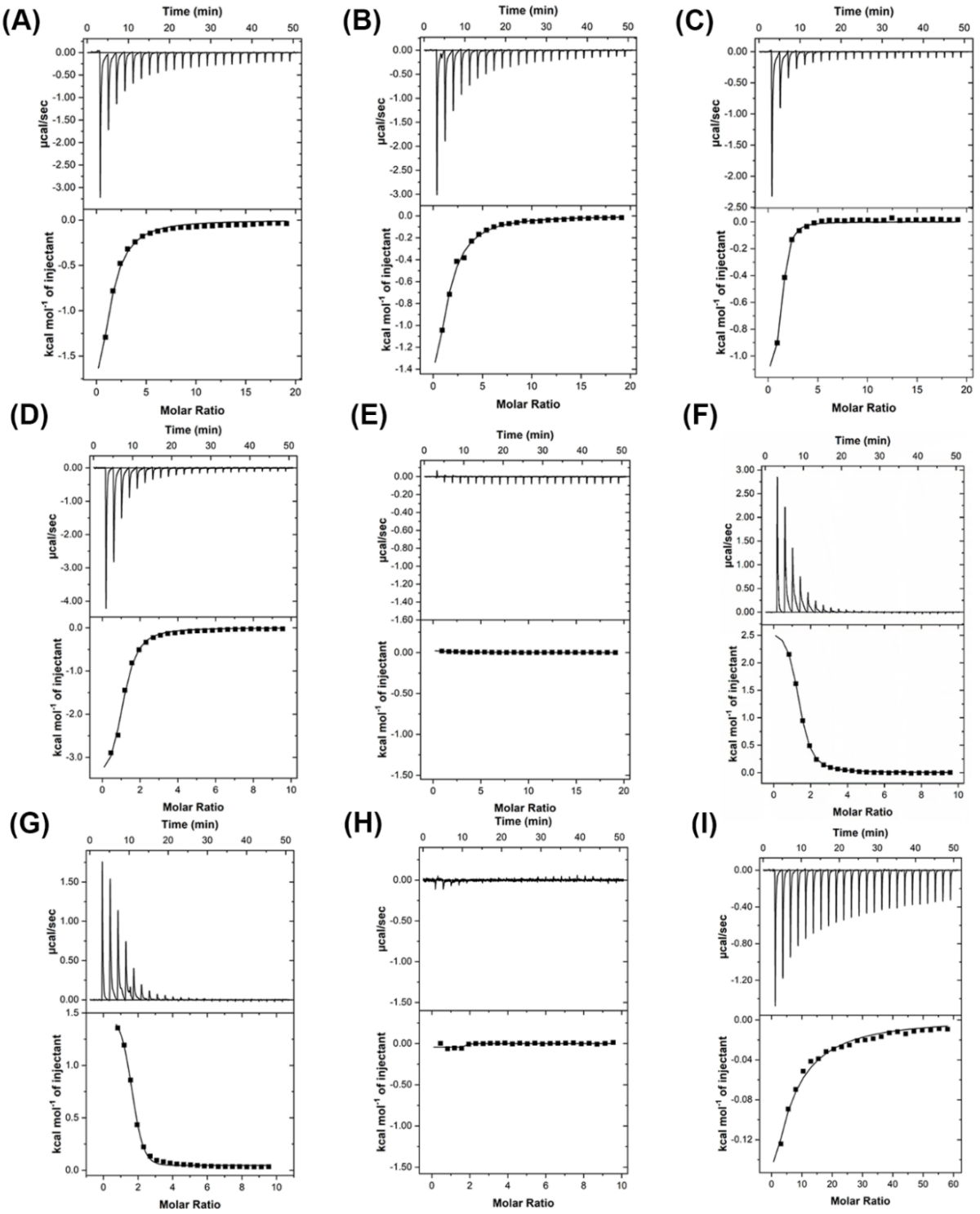


Figure C.11. Thermodynamic data of the mutant βGlyBP_W177X. Binding isotherm of the mutant βGlyBP_W177X titrated with (A) SOP2, (B) LAM2, (C) LAM3, (D) LAM4, (E) CEL2, (F) CEL3, (G) CEL4, (H) CEL5 and (I) GEN2. The upper and lower panels represent the raw heat changes and the integrated heat pulse subtracted with the heat of

dilution, respectively. In all the binding isotherms, the integrated data are plotted against the protein:ligand molar ratio and are fitted using a one-site binding model.

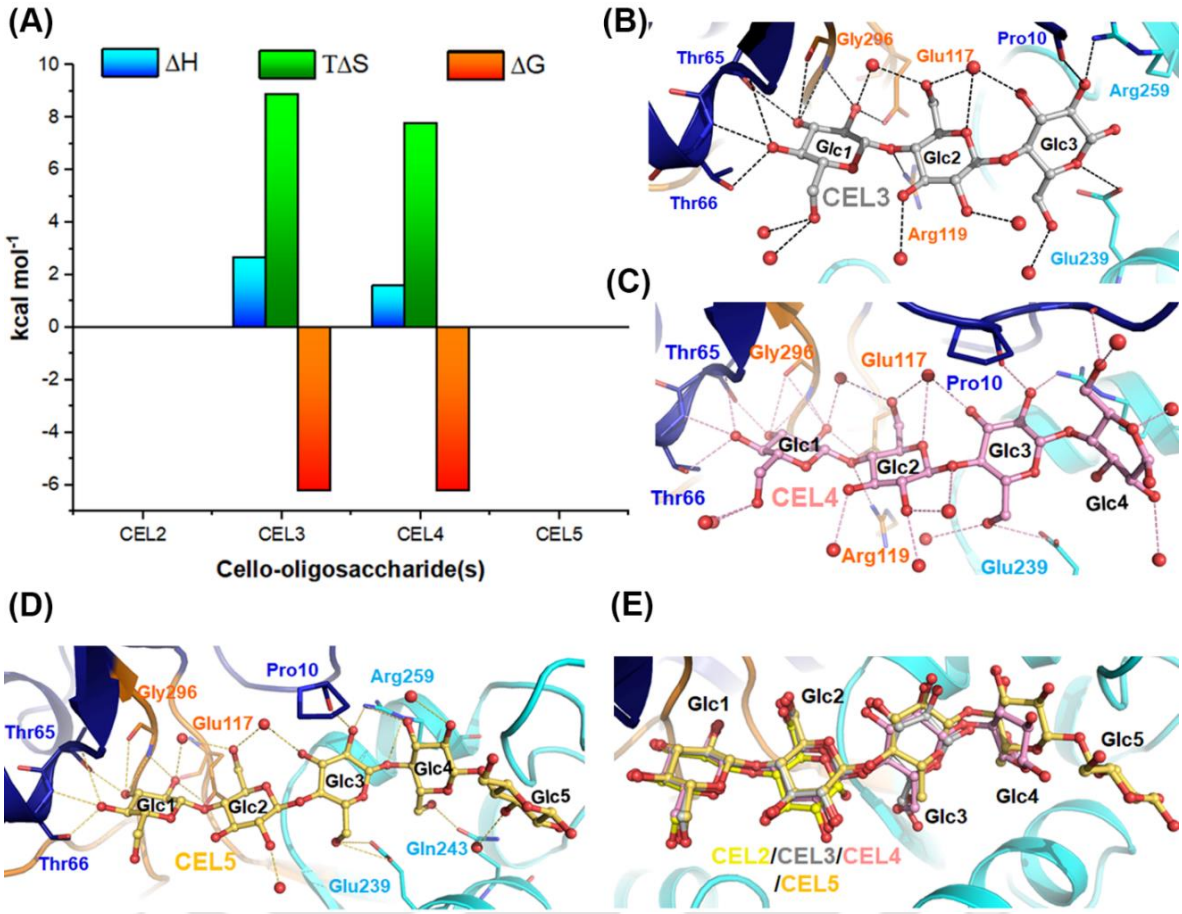


Figure C.12. Thermodynamic parameters and structural details of cello-oligosaccharide (CELn) binding to the mutant βGlyBP_W177X. (A) Thermodynamic change in enthalpy (ΔH, blue), entropy (ΔS, green) and free energy (ΔG, orange) for the binding of CEL2, CEL3, CEL4 and CEL5 to the mutant βGlyBP_W177X are represented as histograms. Hydrogen bonding network in (B) βGlyBP_W177X•CEL3, (C) βGlyBP_W177X•CEL4 and (D) βGlyBP_W177X•CEL5 complex structures. In each complex structure, the Glc unit of CEL3 (grey), CEL4 (pink) and CEL5 (light yellow) are labeled as Glcn (n: n^{th} glycosyl unit). The established hydrogen bonds between the active site residues and the ligand are depicted as black dotted lines while the water molecules are represented as red spheres. (E) Structural alignment of the active sites of the mutant βGlyBP_W177X bound to CEL2 (yellow), CEL3 (grey), CEL4 (pink) and CEL5 (light yellow). In all diagrams, the active site residues from NTD, hinge region and CTD participating in the polar interaction are shown in blue, orange and cyan, respectively.

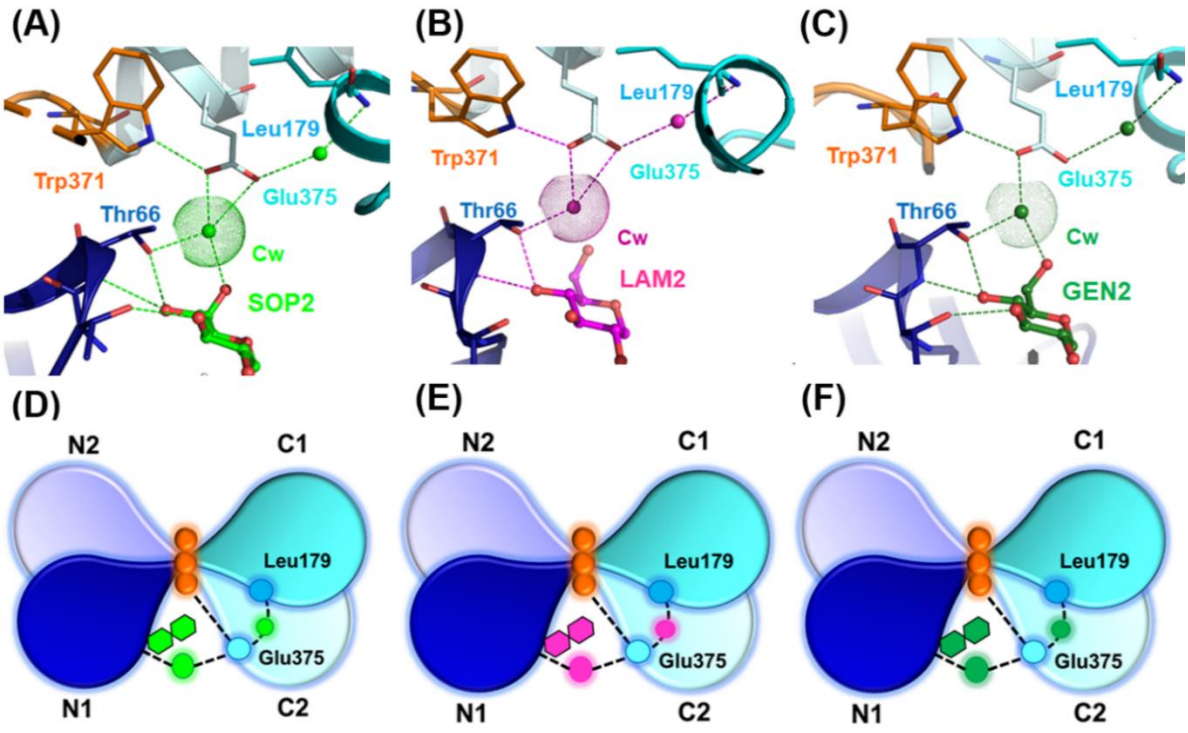


Figure C.13. Inter-domain interaction(s) holding the conformations of the N1, C1 and C2 subdomains. Molecular details of the inter-domain interaction(s) between the N1 (blue), C1 (light cyan) and C2 (cyan) subdomains in (A) βGlyBP_W177X•SOP2, (B) βGlyBP_W177X•LAM2 and (C) βGlyBP_W177X•GEN2 complex structures. Hydrogen bonds and conserved water (Cw) are shown as dotted lines and dotted spheres, respectively. For the figure clarity, model structures of (D) βGlyBP_W177X•SOP2, (E) βGlyBP_W177X•LAM2 and (F) βGlyBP_W177X•GEN2 have been shown, which represent the water-mediated interaction(s) between the N1, C1 and C2 subdomains. The water molecules and the active site residues are depicted as circles. In all the figures, the N1, N2, C1 and C2 subdomains and the amino acid residues from them are shown blue, light blue, light cyan and cyan, respectively.

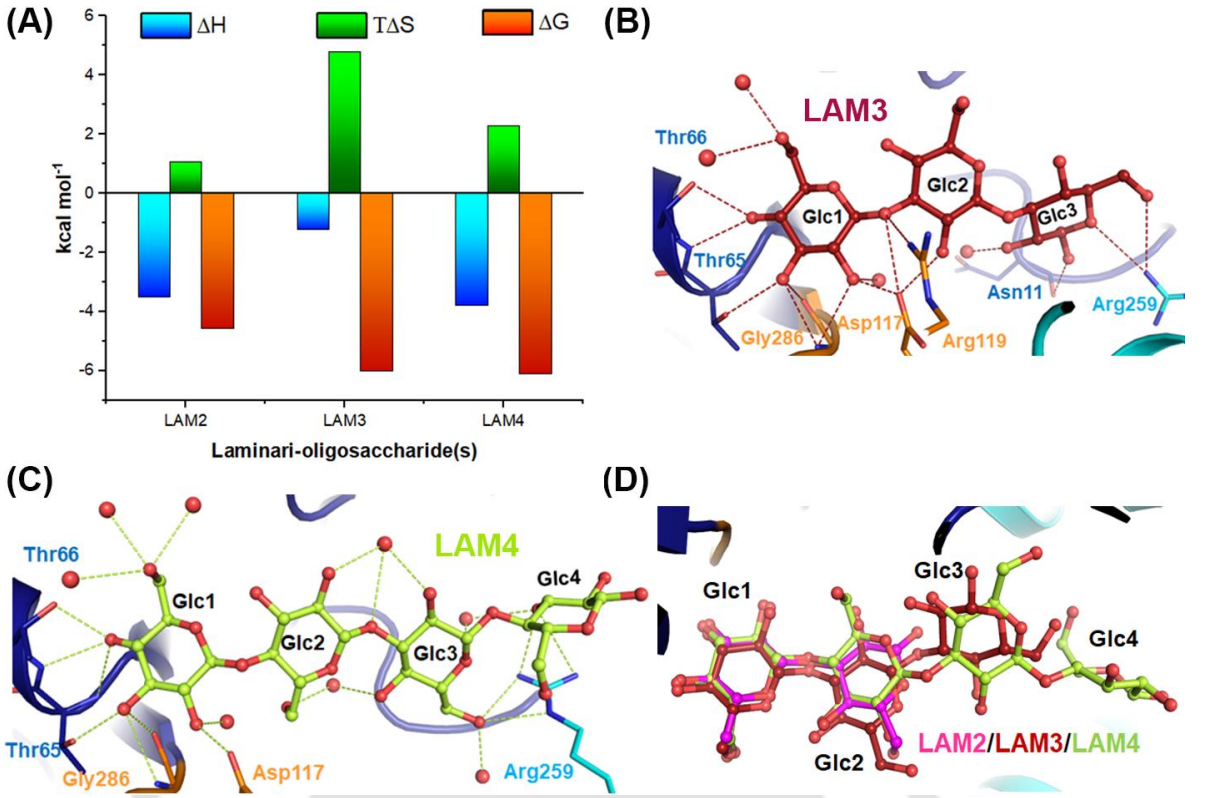


Figure C.14. Thermodynamic parameters and structural details of laminari-oligosaccharide (LAMn) binding to the mutant βGlyBP_W177X. (A) Thermodynamic change in enthalpy (ΔH, blue), entropy (ΔS, green) and free energy (ΔG, orange) for the binding of LAM2, LAM3 and LAM4 to the mutant βGlyBP_W177X are shown as histograms. Hydrogen-bonding network in (B) βGlyBP_W177X•LAM3 and (C) βGlyBP_W177X•LAM4 complex structures. In each complex structure, the Glc unit of LAM3 (red) and LAM4 (light green) are labeled as Glcn (n: n^{th} glycosyl unit). The established hydrogen bonds between the active site residues and the ligand are depicted as dotted lines while the water molecules are represented as red spheres. (D) Structural alignment of the active sites of mutant βGlyBP_W177X bound to LAM2 (magenta), LAM3 (red) and LAM4 (light green). In all the diagrams, the active site residues from NTD, hinge region and CTD participating in the polar interaction(s) are displayed in blue, orange and cyan, respectively.

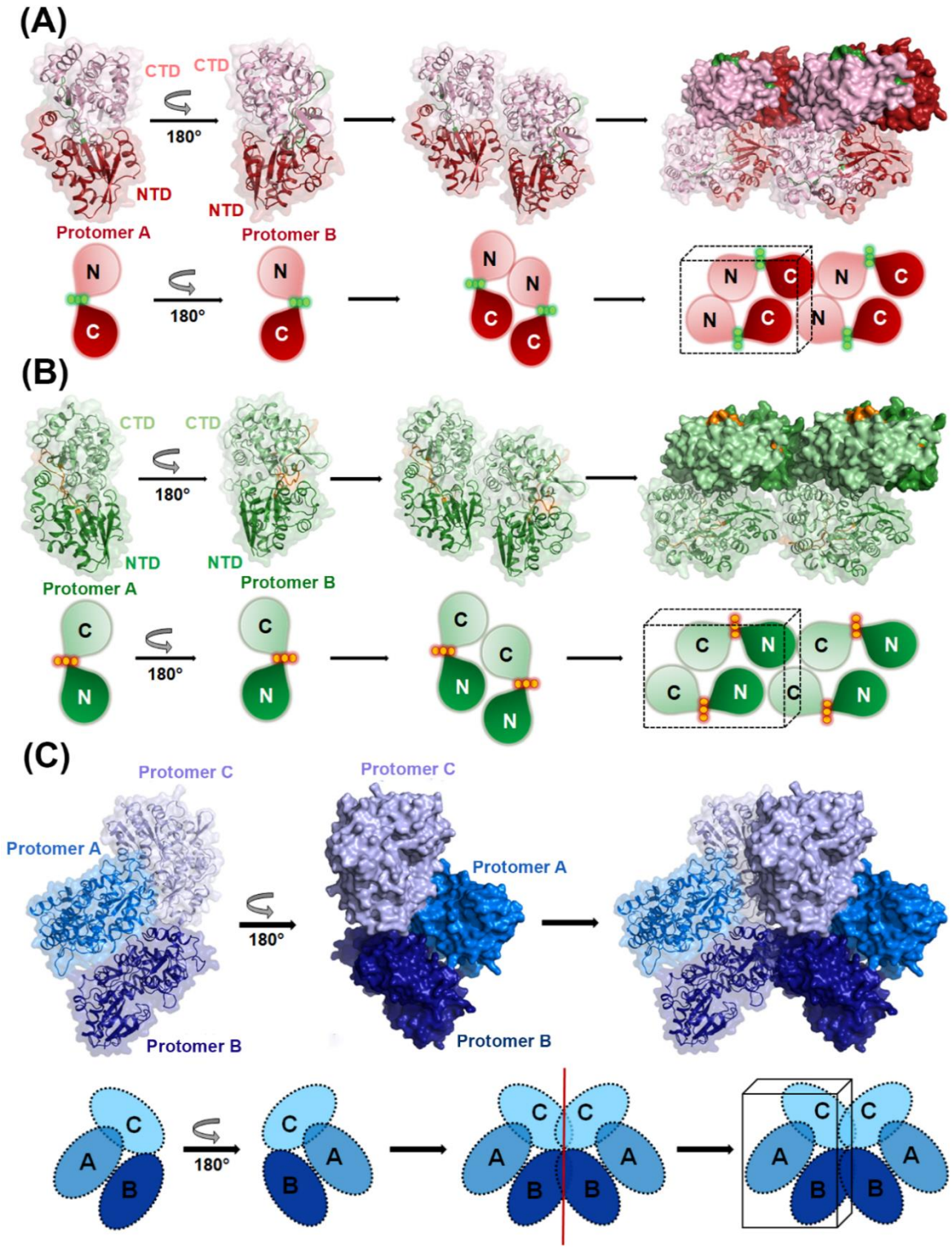


Figure C.15. Crystal packing of the mutant βGlyBP_W177X complexed with CEL3, CEL4 and LAM3. Surface (top) and schematic (bottom) representations of the molecular arrangements of (A) βGlyBP_W177X•CEL3 and (B) βGlyBP_W177X•CEL4 in space group $P2_12_12_1$. In βGlyBP_W177X•CEL3 structure, two domains NTD and CTD are

depicted in red and light red, respectively, while in βGlyBP_W177X•CEL4 structures, they are represented in green and light green, respectively. In space group $P2_12_12_1$, both the protein molecules are arranged by a non-crystallographic rotational symmetry in such a way that the NTD (protomer A) and NTD (protomer B) are juxtaposed. The asymmetric unit, enclosed with black dashed lines, contains two molecules of βGlyBP_W177X. The domains NTD and CTD of a complex structure are labelled as N and C, respectively. (C) Surface (top) and schematic (bottom) representations of the molecular arrangement of βGlyBP_W177X•LAM3 containing three molecules in the asymmetric unit of the space group $P3_121$. Three molecules are represented as ovals in blue (protomer A), dark blue (protomer B) and light blue (protomer C). In 3D space, these three molecules are arranged with two-fold symmetry (red line) in such a way that asymmetric unit (black box) contains three molecules of βGlyBP_W177X•LAM3.

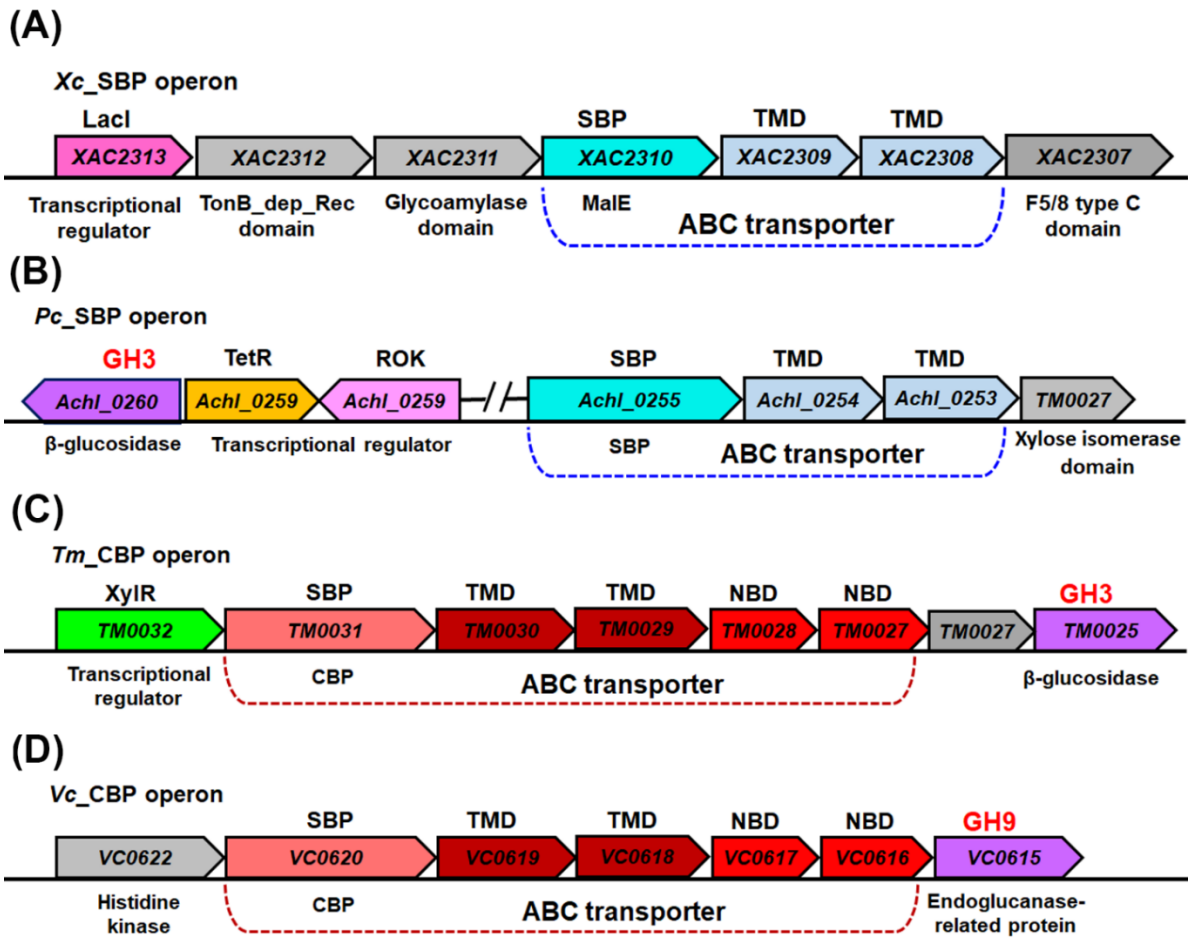


Figure C.16. Genetic cluster of the subclusters C-IV and D-I SBPs of β-glycosides ABC transporter. Genetic organization of the β-glycosides ABC transporter from the subclusters C-IV and D-I and their functionally-associated metabolizing enzymes in (A) *X. axonopodis* pv. *citri* (Xc_SBP), (B) *P. chlorophenolicus* (*Pc*_SBP), (C) *T. maritima* (*Tm*_CBP) and (D) *V. cholerae serotype O1* (*Vc*_CBP). ORFs encoding the β-glycosides ABC transporter from the subclusters C-IV and D-I are represented in red and blue, respectively, while the glycosyl hydrolase (GH) enzymes are depicted in violet. In all the operons, each gene is represented with an arrow indicating its transcriptional direction.

Table C.1. Thermodynamic parameters of β-glycosides binding with wild type (WT) and mutant βGlyBP. All these isothermal titration calorimetry (ITC) experiments were performed at 25°C.

Protein (μM)	Ligand name (Linkage)	Ligand conc. (mM)	Stoichiometry ratio (n)	Association (*Ka*, M) / Dissociation (K_d, μM) constant	Δ*H*	*T*Δ*S*	Δ*G*
					(kcal mol^{-1})		
βGlyBP_WT (200)	*SOP2 (β-1,2)	2.5	0.8	1.86x10^5/ 5.37	2.37	9.58	-7.21
	*CEL2 (β-1,4)	2.0[a]	1.3	6.82x10^6/ 0.15	-3.53	5.78	-9.31
	GEN2 (β-1,6)	5.0	0.8	5.61x10^3/ 178.00	6.68	11.8	-5.12
	LAMn (β-1,3)	1.0[b]	N.D.	N.D.	N.D.	N.D.	N.D.
	LAT (β-1,4)	5.0	N.D.	N.D.	N.D.	N.D.	N.D.
	TRE (α-1,1)	5.0	N.D.	N.D.	N.D.	N.D.	N.D.
	SUC (α-1,2)	5.0	N.D.	N.D.	N.D.	N.D.	N.D.
	MAL (α-1,4)	5.0	N.D.	N.D.	N.D.	N.D.	N.D.
	MLB (α-1,6)	5.0	N.D.	N.D.	N.D.	N.D.	N.D.
	RAF (α-1,6)	5.0	N.D.	N.D.	N.D.	N.D.	N.D.
	GLC	5.0	N.D.	N.D.	N.D.	N.D.	N.D.
	GAL	5.0	N.D.	N.D.	N.D.	N.D.	N.D.
	MAN	5.0	N.D.	N.D.	N.D.	N.D.	N.D.
	TAG	5.0	N.D.	N.D.	N.D.	N.D.	N.D.
βGlyBP_W177X (316)	**SOP2 (β-1,2)	30.0	1[c]	2.99x10^3/ 334.00	-3.62	1.11	-4.73
	LAM2 (β-1,3)	31.6	0.9	2.22x10^3/ 450.00	-3.49	1.07	-4.56
	LAM3 (β-1,3)	31.6	1.1	2.54x10^4/ 39.30	-1.21	4.79	-6.00
	LAM4 (β-1,3)	15.8	1.0	1.93x10^4/ 51.80	-3.79	2.30	-6.10
	CEL2 (β-1,4)	31.6	N.D.	N.D.	N.D.	N.D.	N.D.
	CEL3 (β-1,4)	15.8	1.3	3.48x10^4/ 28.70	2.69	8.89	-6.20
	CEL4 (β-1,4)	15.8	1.5	3.71x10^4/ 26.90	1.59	7.80	-6.21
	CEL5 (β-1,4)	15.8	N.D.	N.D.	N.D.	N.D.	N.D.
	**GEN2 (β-1,6)	100.0	1[c]	2.28x10^2/ 4385.00	-2.31	0.90	-3.21
βGlyBP_W41A (380)	SOP2 (β-1,2)	38.0	1[c]	179/ 5586.00	2.96	6.04	-3.08
	LAM2 (β-1,3)	38.0	1[c]	572/ 1748.00	1.06	4.82	3.76

	LAM3 (β-1,3)	38.0	1.5	$5.54x10^3$/ 180.00	0.65	5.75	-5.10
	LAM4 (β-1,3)	38.0	1[c]	775/ 1290.00	-5.23	-1.29	-3.94
	CEL2 (β-1,4)	38.0	1.0	$1.76x10^4$/ 56.80	2.65	8.43	-5.78
	CEL3 (β-1,4)	38.0	0.8	$1.72x10^3$/ 581.00	8.96	13.35	-4.39
	CEL4 (β-1,4)	38.0	1[c]	$1.94x10^3$/ 515.00	3.46	7.95	-4.49
	CEL5 (β-1,4)	9.5	1[c]	$1.82x10^3$/ 549.00	2.08	6.52	-4.44
	GEN2 (β-1,6)	38.0	N.D.	N.D.	N.D.	N.D.	N.D.
βGlyBP _W67A (800)	SOP2 (β-1,2)	20.0	0.9	771/ 1279.00	0.79	4.73	-3.94
	LAM2 (β-1,3)	20.0	0.8	$1.34x10^3$/ 746.00	-1.01	3.24	-4.25
	LAM3 (β-1,3)	20.0	0.5	$1.57x10^5$/ 6.36	0.62	7.71	-7.09
	LAM4 (β-1,3)	20.0	0.8	$2.12x10^4$/ 47.10	-4.21	1.16	-5.37
	CEL2 (β-1,4)	10.0	0.7	$2.55x10^5$/ 3.92	-3.38	3.99	-7.37
	CEL3 (β-1,4)	20.0	0.8	$9.39x10^6$/ 0.11	0.99	10.48	-9.49
	CEL4 (β-1,4)	20.0	1.2	$4.99x10^5$/ 0.20	0.53	8.31	-7.77
	CEL5 (β-1,4)	10.0	1.4	$3.19x10^5$/ 3.13	0.42	7.92	-7.50
	GEN2 (β-1,6)	80.0	1[c]	61.3/ 16313.20	-5.66	-3.21	-2.45
βGlyBP _W256A (487)	SOP2 (β-1,2)	24.0	1.6	$2.54x10^4$/ 39.30	1.61	7.62	-6.01
	LAM2 (β-1,3)	12.0	1.7	$1.38x10^4$/ 72.40	-0.56	5.09	-5.65
	LAM3 (β-1,3)	12.0	0.9	$3.60x10^5$/ 2.77	2.17	9.74	-7.57
	LAM4 (β-1,3)	12.0	0.9	$1.61x10^5$/ 6.21	-7.11	-0.01	-7.10
	***CEL2 (β-1,4)	16.0	0.5	$1.86x10^4$/ 53.70	-1.60	4.23	-5.83
	CEL3 (β-1,4)	12.0	0.9	$3.06x10^5$/ 3.26	2.33	9.80	-7.47
	CEL4 (β-1,4)	12.0	1.0	$5.02x10^4$/ 19.00	1.82	8.22	-6.40
	CEL5 (β-1,4)	5.0	0.6	$1.41x10^5$/ 7.00	3.03	10.04	-7.01
	GEN2 (β-1,6)	24.0	0.7	$3.06x10^3$/ 320.00	2.17	6.94	-4.77

* The protein βGlyBP _WT concentration used for these experiments was 100 µM during titration.

** The protein βGlyBP _W177X concentration used for these experiments was 300 µM during titration.

*** The protein βGlyBP _W177X concentration used for this experiment was 800 µM during titration.

[a] The ligand concentration was estimated during curve fitting according to the molar ratio.

[b] The polysaccharide laminarin (LAMn) was used at 1% (w/v) during titration.

[c] The binding stoichiometry ratio was fixed as n=1 during one-site-binding fitting model. Abbreviations: SOP2, sophorose (β-1,2); CEL2, cellobiose (β-1,4); GEN2, gentiobiose (β-1,6); LAMn, laminarin (β-1,3); LAT, lactose (β-1,4); TRE, trehalose (α-1,1); SUC, sucrose (α-1,2); MAL, maltose (α-1,4); MLB, melibiose (α-1,6); RAF, raffinose (α-1,2 and α-1,6); GLC, glucose; GAL, galactose; MAN, mannose; TAG, tagatose; N.D., not detected.

Table C.2. List of subcluster D-I SBPs specific to multiple carbohydrates.

Sub cluster	Protein	UniProt ID	Organism	Ligand (PDB ID)
D-I	Glucose/galactose-binding protein (GG-BP)	H7BRJ8	*Pseudomonas putida*	Galactose or GAL (5DVJ) and glucose or GLC (5DVI)
	Galacto-N-biose-/lacto-N-biose I-binding protein (GL-BP)	A8W790	*Bifidobacterium longum*	Lacto-N-biose or LNB (2Z8D), lacto-N-tetraose or LNT (2Z8F) and galacto-N-biose or GNB (2Z8E)
	CpMnBP1	L0E2M2	*Caldanaerobius polysaccharolyticus*	Mannobiose or MAB (4R9F) and mannotriose or MAT (4R9G)
	Maltose-binding protein or MalE (MBP)	P0AEX9	*Escherichia coli*	Maltose or MAL (1ANF), maltotriose or MLR (3MBP) and maltotetraose or MTT (4MBP)
	Raffinose/panose-binding protein (RP-BP)	A0A1A9TAB8	*Bifidobacterium animalis*	Raffinose or RAF (4ZZA) and panose or PAN (4ZZE)

Table C.3. List of structural homologs of the protein βGlyBP from the subcluster D-I SBPs identified by the web server Dali (Holm and Rosenström, 2010).

PDB ID	Z-score	RMSD (Å)	Ligand	Protein
5CI5	44.9	2.4	Tagatose	Extracellular solute-binding protein family 1
6DTQ	44	2.3	α-glycoside	Maltose-binding protein MalE3 or MBP
3K00	42.9	2.3	α-glycoside	Acarbose/maltose-binding protein GacH
5YSD	42.4	2.2	β-glycoside	SOPn-binding protein (SO-BP) or lin1841 protein
1EU8	42.4	2.5	α-glycoside	Trehalose/maltose-binding protein or TMBP
4QSC	42.4	2.4	α-glycoside	Abc-type sugar transporter

Table C.4. List of inter-domain interaction(s) holding the N1, C1 and C2 subdomains of the βGlyBP_W177X mutant.

Protein state	Ligand	Conserved water (Cw) number	List of interaction(s)	
			C1-C2* and C1-Cw-C2**	N1-Cw-C2†
Apo (open)	-	W128	Leu179 N-(2.90)-Glu375 $O^{\delta 2}$	Thr66 $O^{\gamma 1}$-(2.82)-W128-(2.68)-Glu375 $O^{\delta 1}$
Holo (closed)	CEL2	-	Leu179 N-(3.07)-Glu375 $O^{\delta 2}$	-
	CEL3	W15	Leu179 N-(2.90)-Glu375 $O^{\delta 2}$	Thr66 $O^{\gamma 1}$-(2.76)-W15-(2.65)-Glu375 $O^{\delta 1}$
	CEL4	W6	Leu179 N-(2.82)-Glu375 $O^{\delta 2}$	Thr66 $O^{\gamma 1}$-(2.86)-W6-(2.80)-Glu375 $O^{\delta 1}$
	CEL5	-	Leu179 N-(3.11)-Glu375 $O^{\delta 2}$	-
	LAM2	W101	Leu179 N-(2.85)-W266-(2.83)-Glu375 $O^{\delta 2}$	Thr66 $O^{\gamma 1}$-(3.08)-W101-(2.52)-Glu375 $O^{\delta 1}$
	LAM3	W31	Leu179 N-(2.73)-Glu375 $O^{\delta 2}$	Thr66 $O^{\gamma 1}$-(2.80)-W31 (2.66)-Glu375 $O^{\delta 1}$
	LAM4	W112	Leu179 N-(2.71)-Glu375 $O^{\delta 2}$	Thr66 $O^{\gamma 1}$-(2.83)-W112-(2.70)-Glu375 $O^{\delta 1}$
	SOP2	W187	Leu179 N-(2.86)-W134-(2.92)-Glu375 $O^{\delta 2}$	Thr66 $O^{\gamma 1}$-(2.66)-W187-(2.62)-Glu375 $O^{\delta 1}$
	GEN2	W71	Leu179 N-(3.06)-W183-(3.00)-Glu375 $O^{\delta 2}$	Thr66 $O^{\gamma 1}$-(2.63)-W71-(2.49)-Glu375 $O^{\delta 1}$

***** C1-(Hydrogen bond distance in Å)-C2.
** C1-(Hydrogen bond distance in Å)-Cw-(Hydrogen bond distance in Å)-C2.
† N1-(Hydrogen bond distance in Å)-Cw-(Hydrogen bond distance in Å)-C2.

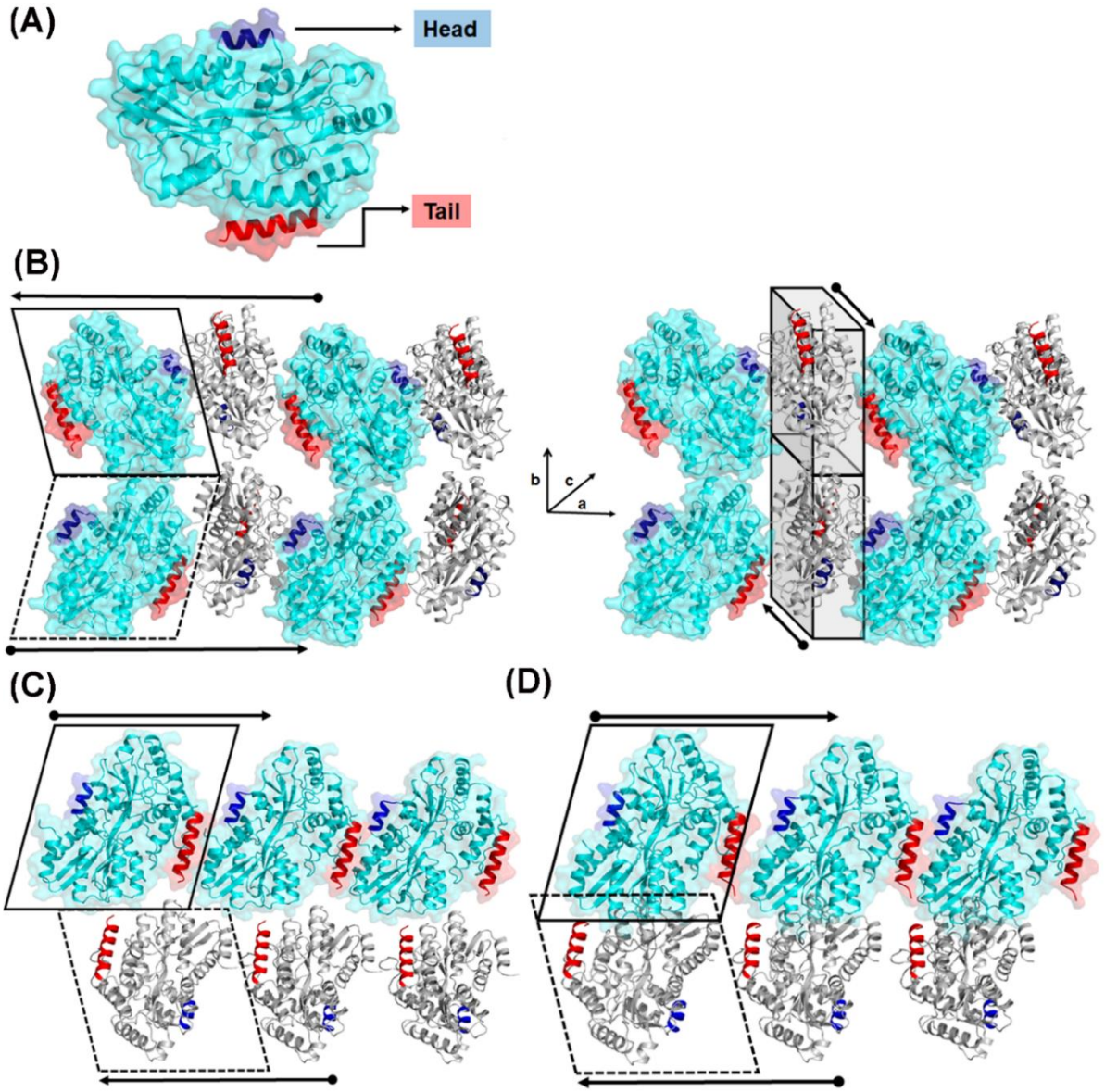


Figure D.1. Different crystal packing of the protein U3GBP. (A) One protomer of U3GBP is represented in cyan. To get the clear view for antiparallel arrangement of protomers, the head and tail of the protomer is demarcated with blue and red, respectively. (B) U3GBP_WT-Form I crystal packing. (Left) arrangement of 1st protomer (cyan) in a direction to the a-axis; (right) arrangement of 2nd protomer (grey) in a direction to the c-axis. (C and D) Crystal packing of the protein U3GBP_WT-Form II and -Form III, respectively. In all three different crystal packings of U3GBP_WT, two protomers are represented in cyan and grey and packing direction of protomers is represented with an arrow. Each protomer of an asymmetric unit are enclosed in black box.

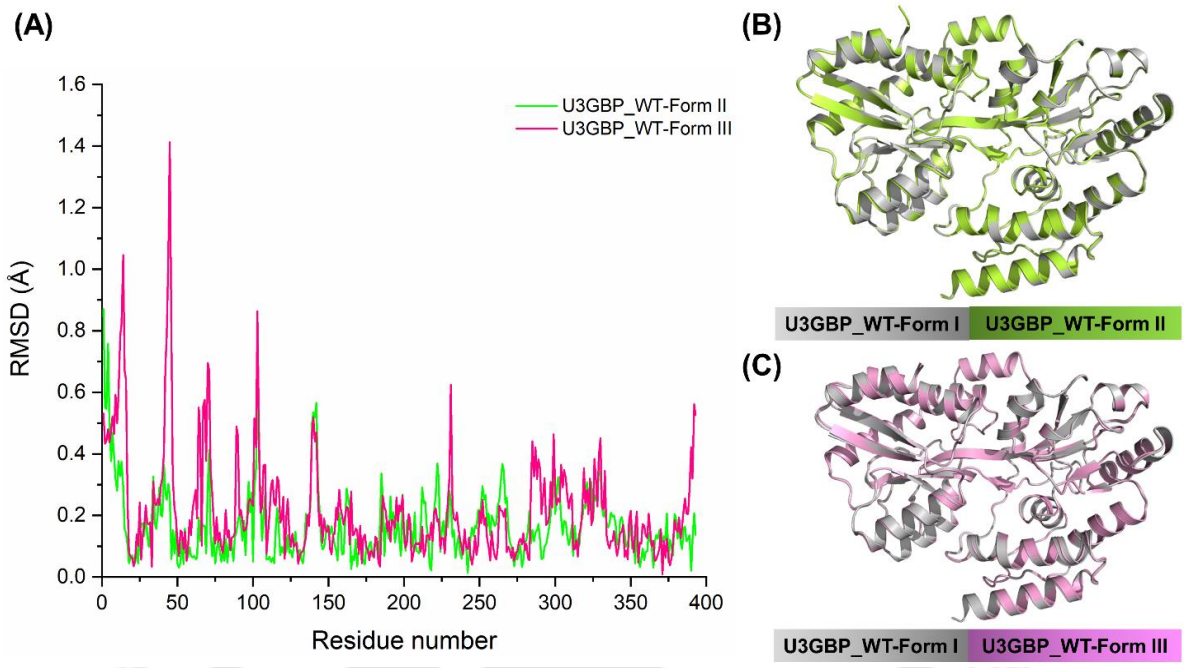


Figure D.2. Structural comparison between different forms of the protein U3GBP_WT. (A) A root mean square deviation (rmsd) plot depicting the comparative structural (dis)similarities between the residues (Cα atom) of the U3GBP_WT-Form II (green) and -Form III (pink) with respect to the U3GBP_WT-Form I. (B and C) Structural superimposition of the U3GBP_WT-Form I (grey) with U3GBP_WT-Form II (green) and U3GBP_WT-Form III (pink), respectively.

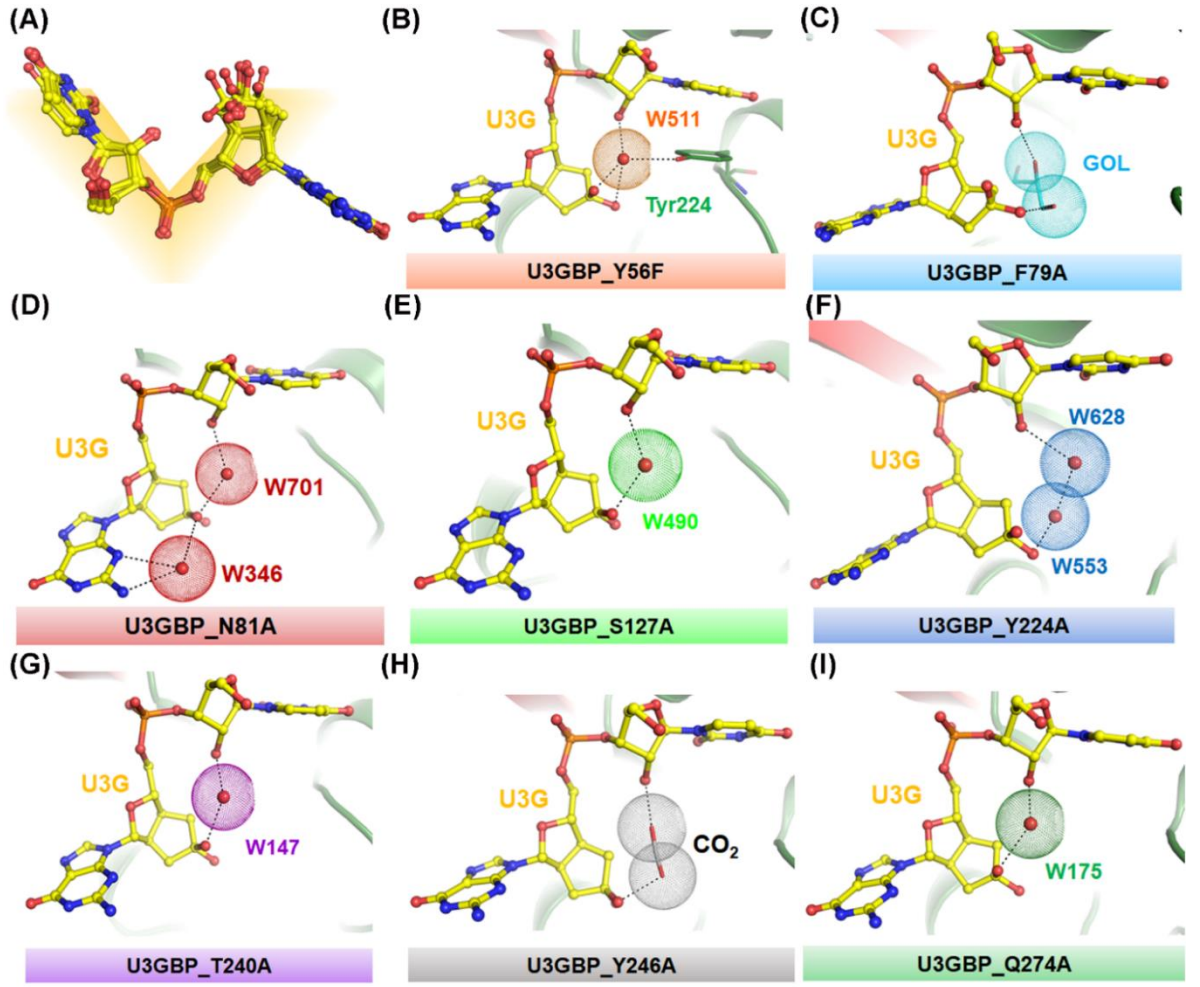


Figure D.3. The V- and O-type conformations of U3G endogenously-bound to the mutant proteins of U3GBP. (A) Conservation of the V-type conformation of the endogenously-bound U3G (yellow ball-and-stick model) of the proteins U3GBP_WT, U3GBP_Y56F, U3GBP_F79A, U3GBP_N81A, U3GBP_S127A, U3GBP_Y224A, U3GBP_T240A, U3GBP_Y246A and U3GBP_Q274A is depicted here. (B-I) The conserved O-type conformation of the endogenously-bound U3G of the proteins U3GBP_Y56F, U3GBP_F79A, U3GBP_N81A, U3GBP_S127A, U3GBP_Y224A, U3GBP_T240A, U3GBP_Y246A and U3GBP_Q274A, respectively. The conserved water molecule (red sphere) and other molecules (glycerol and CO_2) are represented as dotted-spheres. Hydrogen bonds are represented as black-dashed lines.

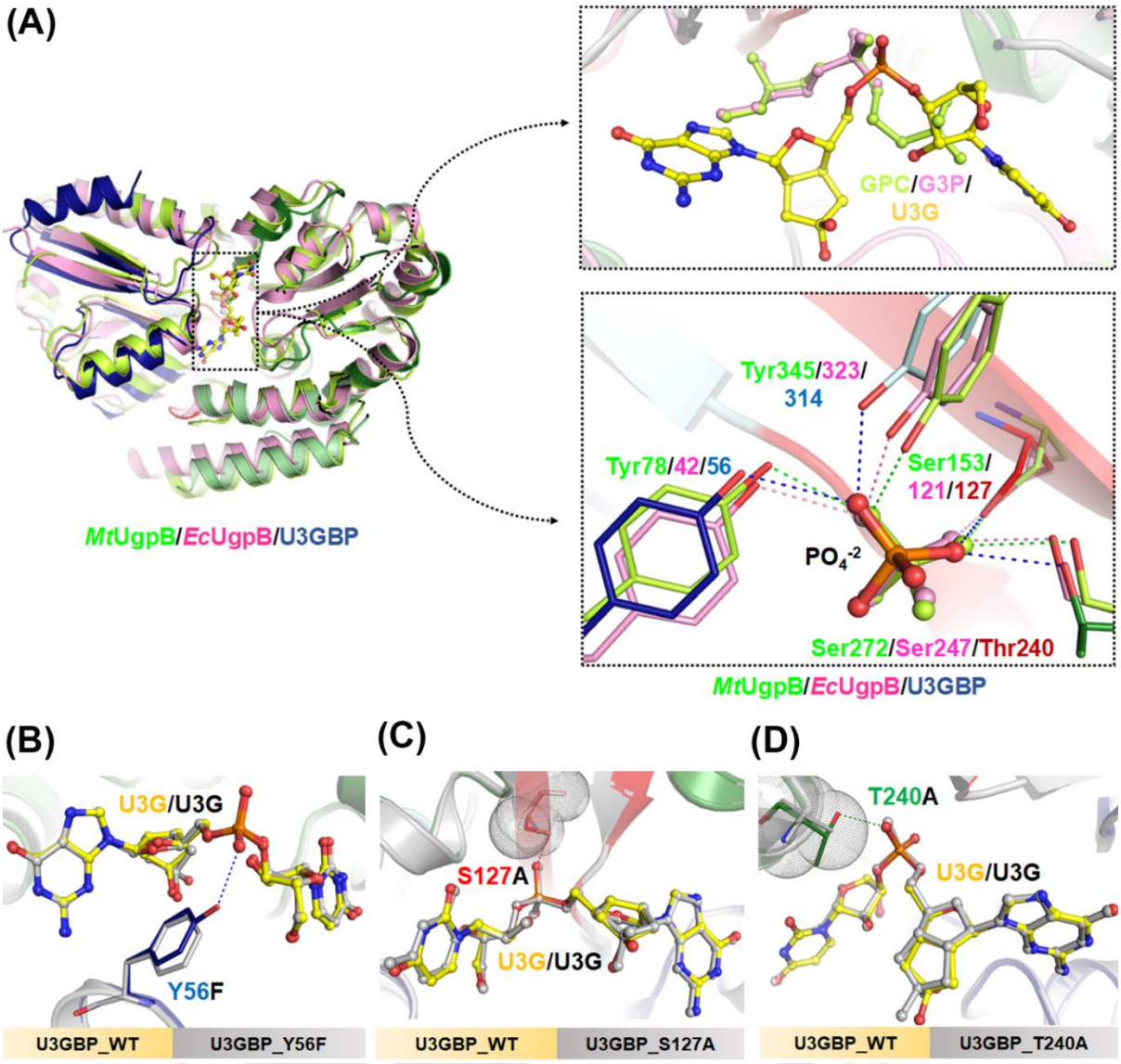


Figure D.4. Active site comparison of UgpBs and U3GBP. (A) Left, structural superimposition of *Mt*UgpB (green, PDB ID: 6R1B), *Ec*UgpB (pink, PDB ID: 4AQ4) and U3GBP (blue, PDB ID: 7C0F). Right, a close-up view of the superimposed structures showing the spatial conservation of GPC (green), G3P (pink) and U3G (yellow) as well as their phosphate group (PO_4^{-2}, ball-and-stick model) of GPC, G3P and U3G in green, pink and orange, respectively. The conserved residues of *Mt*UgpB, *Ec*UgpB and U3GBP, which forms the hydrogen bonding with phosphate group (PO_4^{-2}), are displayed in green, pink and blue, respectively. (B-D) The active site overlay of U3GBP_WT with U3GBP_Y56F (grey), U3GBP_S127A (grey) and U3GBP_T240A (grey), respectively. The endogenously-bound U3G to U3GBP_WT is shown as ball-and-stick model in yellow while that in mutant proteins are in grey. The mutated residues Tyr56, Ser127 and Thr240 are shown in blue, red and green, respectively

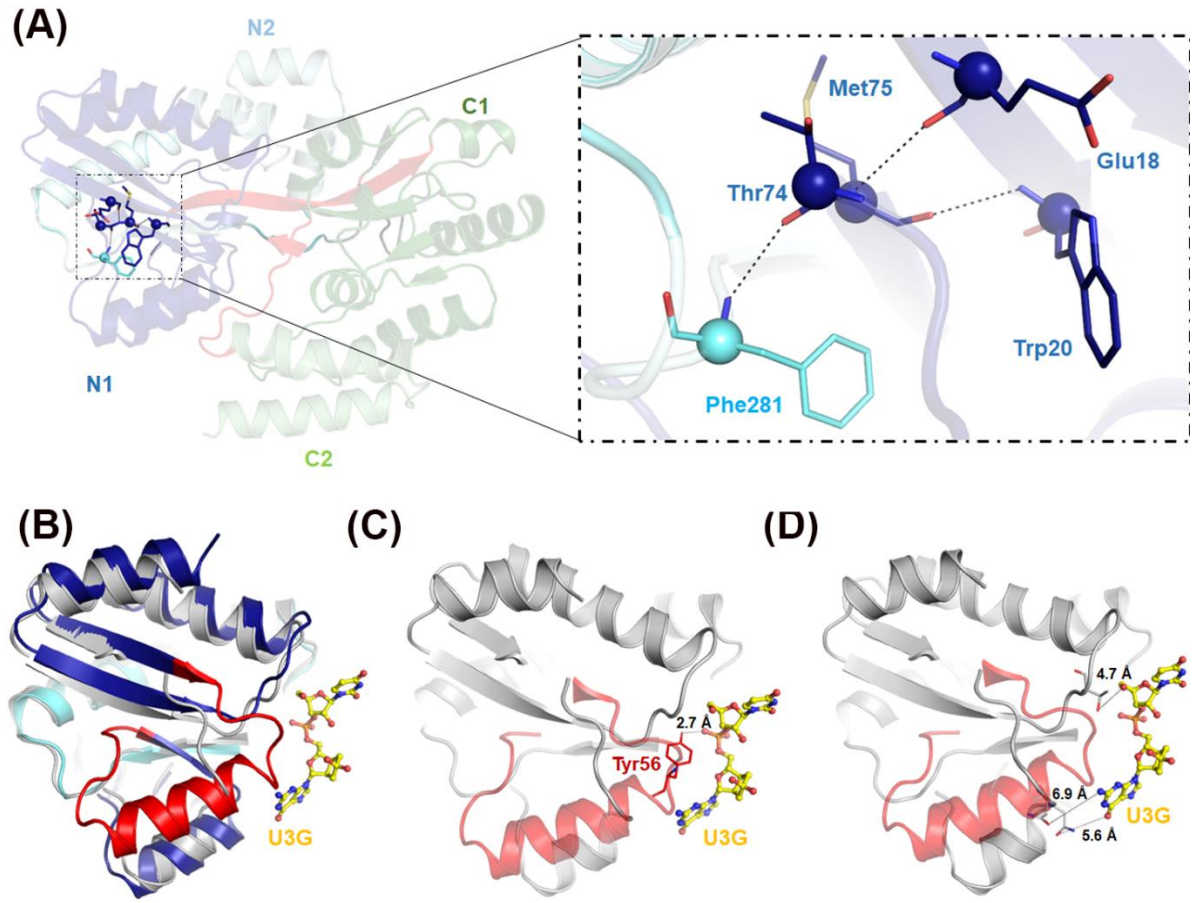


Figure D.5. Interaction between the N1 and N2 subdomains. (A) Left, the overall topology of the wild type protein U3GBP, where four subdomains (N1, N2, C1 and C2) are shown in cartoon while inter-subdomain interacting residues as lines. Right, a close-up view of inter-subdomain interaction of N1 and N2 where the residues from the them involved in hydrogen bonding (black-dotted lines) are shown in blue and cyan, respectively. (B) Superimposition of N1 subdomain of proteins U3GBP_WT (blue) and U3GBP_F79A/Y224A/Y246A/Δ50-75 (grey). Deleted region from wild type protein is highlighted in red. (C) and (D) Superimposition of deleted region (red) and N1 subdomain of U3GBP_F79A/Y224A/Y246A/Δ50-75 (grey). (C) Polar interaction between deleted region and U3G molecule (yellow) is depicted in dotted black line. (D) loss of U3G binding with mutant protein is represented by loss in the establishment of hydrogen bonds (black dotted lines) due to the deleted region which lead to the larger distance between N1 subdomain (grey) of mutant of U3G molecule (yellow).

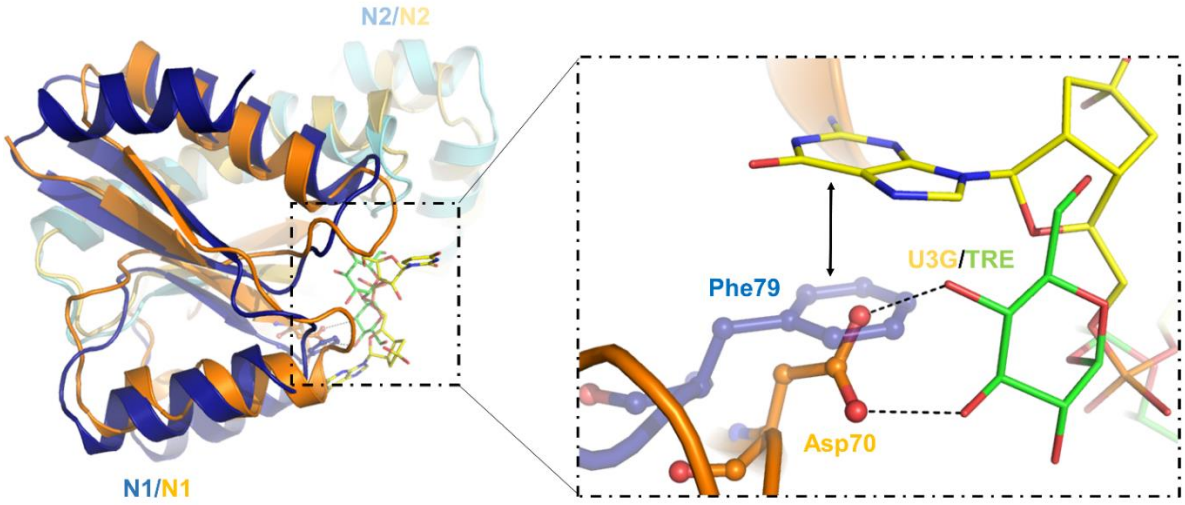


Figure D.6. (Dis)similarity of the N1 subdomain of the wild type proteins αGlyBP and U3GBP. Left, overlay of the N1 subdomain of the wild type proteins αGlyBP (orange, PDB ID: 6J9W) and U3GBP (blue, PDB ID: 7C0F) are shown as ribbon model. Right, a zoomed-in view of the active site showing a conserved position of the residues Asp70 (orange) and Phe79 (blue) both as ball-and-stick model. The ligands trehalose (TRE) and U3G bound to these two proteins are depicted in yellow and green, respectively. Hydrogen bonds are shown as black-dotted lines, whereas π-π stacking interaction as double-sided arrow.

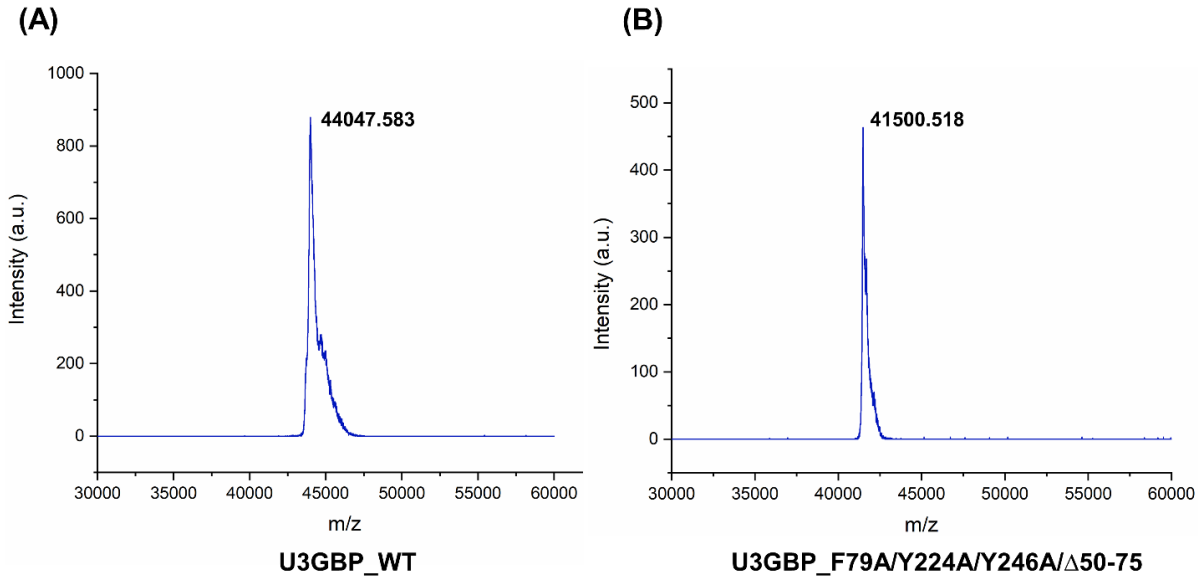


Figure D.7. MALDI-TOF mass spectrometry analysis. MALDI-TOF mass spectra of proteins (A and B) U3GBP_WT and U3GBP_F79A/Y224A/Y246A/Δ50-75, respectively.

APPENDIX D- SUPPLEMENTARY DATA TO CHAPTER 6

Table D.1. List of the sub(cluster) D SBPs considered for N-terminal helix (α1) analysis.

(sub)cluster	PDB ID*	UniProt ID	Protein name	Organism name	Ligand bound
D-I	1EU8	Q7LYW7	MalE	*Thermococcus litoralis*	Maltose, Trehalose
D-I	2B3B	Q72KX2	GBP	*Thermus thermophilus* HB27	Glucose
D-II	2QRY	P31550	ThiB	*Escherichia coli*	Thiamine
D-II	2V84	O83661	PotD	*Treponema pallidum*	Spermidine, Putrescine
D-III-a	1SBP	P02906	SBP	*Salmonella typhimurium*	Sulfate
D-III-a	3C9H	Q7CWZ6	SBP	*Agrobacterium fabrum*	Iron
D-III-b	3CFX	Q8TTZ5	ModA	*Methanosarcina acetivorans*	Tungstate
D	3UOR	Q8PK66	MalE	*Xanthomonas axonopodis*	Maltose
D	3K01	B0B0V1	GacH	*Streptomyces glaucescens*	Acarbose, Maltose

*List of PDBs was sorted from a previous report (Scheepers et al., 2016) based on superimposition with the protein U3GBP and visual inspection of the N1 subdomain.

Table D.2. Polar and hydrophobic interactions between protein U3GBP (WT-Form I, WT-Form II, WT-Form III and Y56F) and endogenously-bound U3G. The distance cut-off for hydrogen bonds and stacking interactions were kept at 3.5 and 4.0 Å, respectively.

Ligand (U3G)			Protein (chain A) / water atoms			
Group	**Interaction type**	**Atom**	**U3GBP_WT-Form I**	**U3GBP_WT-Form II**	**U3GBP_WT-Form III**	**U3GBP_Y56F**
Nitrogenous base (uracil)	Hydrogen bond	N1	Asn226 $O^{\delta1}$	Asn226 $O^{\delta1}$	Asn226 $O^{\delta1}$	Asn226 $O^{\delta1}$
		O4	HOH222	HOH316	HOH168	HOH191
		O5	Asn226 $N^{\delta2}$	Asn226 $N^{\delta2}$, HOH266	Asn226 $N^{\delta2}$, HOH409	Asn226 $N^{\delta2}$, HOH423
	π-π stacking	-	Tyr224, Tyr246	Tyr224, Tyr246	Tyr224, Tyr246	Tyr224, Tyr246
Ribose sugar	Hydrogen bond	O1	HOH296, HOH297	HOH521, HOH597	HOH395	HOH424, HOH511
		O3	HOH291, HOH298	HOH87	HOH57, HOH509	HOH131
Phosphate group		O11	Ser127 O^{γ}, Thr240 $O^{\gamma1}$	Ser127 O^{γ}, Thr240 $O^{\gamma1}$	Ser127 O^{γ}, Thr240 $O^{\gamma1}$	Ser127 O^{γ}, Thr240 $O^{\gamma1}$

		O12	Tyr56 O^{η}, Tyr314 O^{η}	Tyr56 O^{η}, Tyr314 O^{η}	Tyr56 O^{η}, Tyr314 O^{η}	**Gly275 N, Tyr314 O^{η}**
Modified ribose sugar		O9	HOH179, HOH296	HOH386, HOH521	HOH395	HOH511
		O10	-	$CO_2$11, HOH463	-	EDO6, HOH188, HOH511
Nitrogenous base (guanine)	Hydrogen bond	N3	Arg360 N^{ε}	Arg360 N^{ε}, HOH542	Arg360 N^{ε}	Arg360 N^{ε}
		N4	Glu357 $O^{\varepsilon 2}$	Glu357 $O^{\varepsilon 2}$	Glu357 $O^{\varepsilon 2}$	Glu357 $O^{\varepsilon 2}$
		N5	HOH179	HOH386	-	-
		N6	Glu357 $O^{\varepsilon 2}$, HOH179	Glu357 $O^{\varepsilon 2}$, HOH386	Glu357 $O^{\varepsilon 1}$ and $O^{\varepsilon 2}$	Glu357 $O^{\varepsilon 2}$
		O8	Asn81 $O^{\delta 1}$ and $N^{\delta 2}$	Asn81 $O^{\delta 1}$ and $N^{\delta 2}$, HOH542	Asn81 $O^{\delta 1}$ and $N^{\delta 2}$	Asn81 $N^{\delta 2}$
	π-π and CH-π stacking	-	Phe79, Phe361	Phe79, Phe361	Phe79, Phe361	Phe79, Phe361

CO_2: carbon dioxide, EDO: ethylene glycol.
Note: The interactions highlighted in bold shows changes in the wild type U3GBP_WT and mutant proteins.

Table D.3. Polar and hydrophobic interactions between protein U3GBP (F79A, N81A and S127A) and endogenously-bound U3G. The distance cut-off for hydrogen bonds and stacking interactions were kept at 3.5 and 4.0 Å, respectively.

Ligand (U3G)			**Protein (chain A) / water atoms**			
Group	**Interaction type**	**Atom**	**U3GBP_F79A-Form I**	**U3GBP_F79A-Form II**	**U3GBP_N81A**	**U3GBP_S127A**
Nitrogenous base (uracil)	Hydrogen bond	N1	Asn226 $O^{\delta 1}$	Asn226 $O^{\delta 1}$	Asn226 $O^{\delta 1}$	Asn226 $O^{\delta 1}$
		O4	HOH289	HOH221	HOH332	HOH173
		O5	Asn226 $N^{\delta 2}$, HOH380	Asn226 $N^{\delta 2}$	Asn226 $N^{\delta 2}$, HOH511	Asn226 $N^{\delta 2}$, HOH333
	π-π stacking	-	Tyr224, Tyr246	Tyr224, Tyr246	Tyr224, Tyr246	Tyr224, Tyr246
Ribose sugar	Hydrogen bond	O1	GOL17	-	PEG9, HOH701	HOH490, HOH492
		O3	HOH378	HOH252	HOH284	HOH52, HOH630
Phosphate group		O11	Ser127 O^{γ}, Thr240 $O^{\gamma 1}$	Ser127 O^{γ}, Thr240 $O^{\gamma 1}$	Ser127 O^{γ}, Thr240 $O^{\gamma 1}$	**Thr240 $O^{\gamma 1}$**

		O12	Tyr56 O^{η}, Tyr314 O^{η}	Tyr56 O^{η}, Tyr314 O^{η}	Tyr56 O^{η}, Tyr314 O^{η}	Tyr56 O^{η}, Tyr314 O^{η}, Gly275 N
Modified ribose sugar		O7	HOH444	HOH30	-	-
		O9	HOH449	Tyr224 O^{η}	HOH346, HOH701	HOH490
		O10	GOL17	-	HOH702	HOH491, HOH494, HOH634
Nitrogenous base (guanine)	Hydrogen bond	N2	HOH444	HOH30	-	-
		N3	Arg360 N^{ε}, HOH601	Arg360 N^{ε}, HOH515	Arg360 N^{ε}, HOH278	Arg360 N^{ε}, HOH56
		N4	Glu357 $O^{\varepsilon 2}$	-	Glu357 $O^{\varepsilon 2}$	Glu357 $O^{\varepsilon 2}$
		N5	-	HOH249	HOH346	-
		N6	Glu357 $O^{\varepsilon 2}$	Glu357 $O^{\varepsilon 2}$	Arg57 N, Glu357 $O^{\varepsilon 2}$, HOH346	Glu357 $O^{\varepsilon 2}$
		O8	Asn81 $N^{\delta 2}$	Asn81 $N^{\delta 2}$, HOH515	-	Asn81 $N^{\delta 2}$, HOH56
	π-π and CH-π stacking	-	**Phe361**	**Phe361**	Phe79, Phe361	Phe79, Phe361

GOL: glycerol, PEG: polyethylene glycol.
Note: The interactions highlighted in bold shows changes in the wild type U3GBP_WT and mutant proteins.

Table D.4. Polar and hydrophobic interactions between protein U3GBP (Y224A, T240A and Y246A) and endogenously-bound U3G. The distance cut-off for hydrogen bonds and stacking interactions were kept at 3.5 and 4.0 Å, respectively.

Ligand (U3G)			**Protein (chain A) / water atoms**			
Group	**Interaction type**	**Atom**	**U3GBP_Y224A-Form I**	**U3GBP_Y224A-Form II**	**U3GBP_T240A**	**U3GBP_Y246A-Form I**
Nitrogenous base (uracil)	Hydrogen bond	N1	Asn226 $O^{\delta 1}$	Asn226 $O^{\delta 1}$	Asn226 $O^{\delta 1}$	HOH619
		O4	-	-	-	HOH233, HOH619
		O5	Asn226 $N^{\delta 2}$, HOH144	Asn226 $N^{\delta 2}$	Asn226 $N^{\delta 2}$, HOH359	Asn226 $N^{\delta 2}$
	π-π stacking	-	**Tyr246**	**Tyr246**	Tyr224, Tyr246	**Tyr224**
Ribose sugar	Hydrogen bond	O1	HOH628	-	HOH147, HOH722	$CO_2$1

		O3	HOH81	HOH69	HOH177, HOH401	HOH357
Phosphate group		O6	-	-	-	-
		O11	Ser127 O^{γ}, Thr240 $O^{\gamma 1}$	Ser127 O^{γ}, Thr240 $O^{\gamma 1}$	**Ser127 O^{γ}**	Ser127 O^{γ}, Thr240 $O^{\gamma 1}$
		O12	Tyr56 O^{η}, Tyr314 OH	Tyr56 O^{η}, Tyr314 O^{η}	Tyr56 O^{η}, Tyr314 O^{η}	Tyr56 O^{η}, Tyr314 O^{η}
Modified ribose sugar (guanine)		O7	-	-	-	-
		O9	HOH553	-	HOH147, HOH621	$CO_2$2
		O10	HOH553	$CO_2$2, HOH185, HOH393	HOH213	Asp173 $O^{\delta 2}$, $CO_2$1, $CO_2$2, HOH99
		N3	Arg360 N^{ε}, HOH69	Arg360 N^{ε}	Arg360 N^{ε}, HOH628	Arg360 N^{ε}, HOH504
		N4	Glu357 $O^{\varepsilon 2}$	Glu357 $O^{\varepsilon 2}$	Glu357 $O^{\varepsilon 2}$	Glu357 $O^{\varepsilon 2}$
		N5	-	-	HOH621	-
		N6	Arg57 N, Glu357 $O^{\varepsilon 2}$	Glu357 $O^{\varepsilon 2}$	Arg57 N, Glu357 $O^{\varepsilon 2}$, HOH621	Glu357 $O^{\varepsilon 2}$
		O8	Asn81 $N^{\delta 2}$	Asn81 $N^{\delta 2}$	Asn81 $N^{\delta 2}$	Asn81 $N^{\delta 2}$, HOH504
	π-π and CH-π stacking	-	Phe79, Phe361	Phe79, Phe361	Phe79, Phe361	Phe79, Phe361

CO_2: carbon dioxide.
Note: The interactions highlighted in bold shows changes in the wild type U3GBP_WT and mutant proteins.

Table D.5. Polar and hydrophobic interactions between protein U3GBP (Y246A, Q274A and Y224A/Y246A) and endogenously-bound U3G. The distance cut-off for hydrogen bonds and stacking interactions were kept at 3.5 and 4.0 Å, respectively.

Ligand (U3G)			**Protein (chain A) / water atoms**			
Group	**Interaction type**	**Atom**	**U3GBP_Y246A-Form II**	**U3GBP_Q274A-Form I**	**U3GBP_Q274A-Form II**	**U3GBP_Y224A/Y246A-Form I**
Nitrogenous base (uracil)	Hydrogen bond	N1	Asn226 $O^{\delta 1}$	Asn226 $O^{\delta 1}$	Asn226 $O^{\delta 1}$	Asn226 $O^{\delta 1}$
		O4	HOH5	-	HOH176	-
		O5	Asn226 $O^{\delta 1}$ and $N^{\delta 2}$	Asn226 $N^{\delta 2}$	Asn226 $N^{\delta 2}$, HOH122	Asn226 $N^{\delta 2}$, Gln227 $N^{\varepsilon 2}$, HOH270

	π-π stacking	-	**Tyr224**	Tyr224, Tyr246	Tyr224, Tyr246	-
Ribose sugar	Hydrogen bond	O1	-	-	HOH175	-
		O3	-	HOH123	HOH41, HOH398	HOH523
		O11	Ser127 O^{γ}, Thr240 $O^{\gamma 1}$	Ser127 O^{γ}, Thr240 $O^{\gamma 1}$	Ser127 O^{γ}, Thr240 $O^{\gamma 1}$	Ser127 O^{γ}, Thr240 $O^{\gamma 1}$
		O12	Tyr56 O^{η}, Tyr314 O^{η}	Tyr56 O^{η}, Gly275 N, Tyr314 O^{η}	Tyr56 O^{η}, Tyr314 O^{η}	Tyr56 O^{η} , Tyr314 O^{η}
		O9	HOH77	-	HOH175	
		O10	-	HOH248	$CO_3$9	$CO_2$4, EDO11, HOH368, HOH442
Nitrogenous base (guanine)	Hydrogen bond	N2	-	Arg360 $N^{\eta 1}$	-	-
		N3	Arg360 N^{ε}	Arg360 N^{ε} , HOH216	Arg360 N^{ε} , HOH402	Arg360 N^{ε}, HOH421
		N4	-	Glu357 $O^{\varepsilon 2}$	Glu357 $O^{\varepsilon 2}$	Glu357 $O^{\varepsilon 2}$
		N5	-	-	-	HOH241
		N6	Glu357 $O^{\varepsilon 2}$	Arg57 N, Glu357 $O^{\varepsilon 2}$	Arg57 N, Glu357 $O^{\varepsilon 2}$	Arg57 N, Glu357 $O^{\varepsilon 2}$, HOH241
		O8	Asn81 $N^{\delta 2}$	Asn81 $O^{\delta 1}$ and $N^{\delta 2}$	Asn81 $N^{\delta 2}$	Asn81 $N^{\delta 2}$
	π-π and CH-π stacking	-	Phe79, Phe361	Phe79, Phe361	Phe79, Phe361	Phe79, Phe361

CO_3: carbon trioxide, CO_2: carbon dioxide, EDO: ethylene glycol.
Note: The interactions highlighted in bold shows changes in the wild type U3GBP_WT and mutant proteins.

Table D.6. Polar and hydrophobic interactions between protein U3GBP (Y224A/Y246A) and endogenously-bound U3G. The distance cut-off for hydrogen bonds and stacking interactions were kept at 3.5 and 4.0 Å, respectively.

Ligand (U3G)			**Protein (chain A) / water atoms**
Group	**Interaction type**	**Atom**	**U3GBP_Y224A/Y246A-Form II**
Nitrogenous base (uracil)	Hydrogen bond	N1	Asn226 $O^{\delta 1}$
		O4	HOH213
		O5	Asn226 $N^{\delta 2}$, HOH352

	π-π stacking	-	-
Ribose sugar	Hydrogen bond	O1	$CO_2$3
		O3	HOH510
Phosphate group		O6	-
		O11	Ser127 O^{γ}, Thr240 $O^{\gamma 1}$
		O12	Tyr56 O^{η} , Tyr314 O^{η}
Modified ribose sugar		O7	-
		O9	HOH364
		O10	$CO_2$3, PEG9, HOH670
Nitrogenous base (guanine)	Hydrogen bond	N2	-
		N3	Arg360 N^{ε}, HOH393
		N4	Glu357 $O^{\varepsilon 2}$
		N5	HOH364
		N6	Glu357 $O^{\varepsilon 2}$, HOH364
		O8	Asn81 $N^{\delta 2}$
	π-π and CH-π stacking	-	Phe79, Phe361

CO_2: carbon dioxide, PEG: polyethylene glycol.
Note: The interactions highlighted in bold shows changes in the wild type U3GBP_WT and mutant proteins.

Table D.7. Details of the SBPs considered in the structure- and sequence-based evolutionary analyses (Scheepers et al., 2016).

PDB ID*	**UniProt ID**	**Protein name**	**Organism name**	**Ligand bound**	**Existing (sub)cluster****	**Proposed (sub)cluster (this study)†**
7C0F	Q5SLB4	U3GBP	*T. thermophilus* HB8	U3G	-	**D-I**
4EQB	D6ZS00	PotD	*Streptococcus pneumoniae*	Spermidine, Putrescine	D-I	D-I
3K6U	Q8TTZ5	ModA	*Methanosarcina acetivorans*	Molybdate	D-I	**D-III-a**
2B3B	Q72KX2	GBP	*T. thermophilus* HB27	Glucose	D-I	**D-III-a**
4AQ4	P0AG80	UgpB	*E. coli*	Diester	D-I	D-I
1EU8	Q7LYW7	MalE	*T. litoralis*	Maltose, Trehalose	D-I	D-I
2QRY	P31550	ThiB	*E. coli*	Thiamine	D-II	D-II
2V84	O83661	PotD	*T. pallidum*	Spermidine, Putrescine	D-II	**D-IV**

3TTM	Q9I6J1	SpuD	*Pseudomonas aeruginosa*	Putrescine	D-II	**D-III-a**
4LQ2	P06614	CysB	*Salmonella enterica*	O-Acetylserine	D-II	D-II
4QBA	N.A.	CcpE	*Staphylococcus aureus*	Citrate	D-II	D-II
4H1X	P0C2M5	PstS2	*S. pneumoniae*	Phosphate	D-III-a	**D**
3LR1	Q749P2	TupA	*Geobacter sulfurreducens*	Tungstate	D-III-a	D-III-a
3C9H	Q7CWZ6	SBP	*A. fabrum*	Iron	D-III-a	D-III-a
1SBP	P02906	SBP	*S. typhimurium*	Sulfate	D-III-a	D-III-a
2H5Y	Q8PHA1	ModA	*Xanthomonas axonopodis*	Molybdate	D-III-a	D-III-a
3CG1	Q8U4K5	WtpA	*Pyrococcus furiosus*	Molybdate, Tungstate	D-III-b	D-III-b
3CG3	O57890	WtpA	*P. horikoshii*	Molybdate, Tungstate	D-III-b	D-III-b
3CFX	Q8TTZ5	ModA	*M. acetivorans*	Tungstate	D-III-b	D-III-b
3CFZ	Q58586	WtpA	*Methanocaldococcus jannaschii*	Molybdate, Tungstate	D-III-b	D-III-b
2ONR	O30142	WtpA / ModA	*Archaeoglobus fulgidus*	Molybdate, Tungstate	D-III-b	D-III-b
4ELO	Q5SHV2	FbpA	*T. thermophilus* HB8	Iron	D-IV	D-IV
1MRP	P35755	FbpA	*Haemophilus influenzae*	Iron	D-IV	D-IV
1XVY	P21408	FbpA	*Serratia marcescens*	Iron	D-IV	D-IV
1XVX	A1JLH5	YfuA	*Yersinia enterocolitica*	Iron	D-IV	D-IV
1Y4T	Q0PBW4	CfbpA	*Campylobacter jejuni*	Iron	D-IV	D-IV
3VXB	Q76BU9	BxlD	*Streptomyces thermoviolaceus*	Sugar	D	D
4MFI	A5U6I5	UgpB	*Mycobacterium tuberculosis* H37Ra	Glycerophosphocholine	D	D
3UOR	Q8PK66	MalE	*X. axonopodis*	Maltose	D	D
3K01	B0B0V1	GacH	*S. glaucescens*	Acarbose, Maltose	D	D
2HXW	Q0PBL7	Peb3	*C. jejuni*	Citrate	D	**D-II**

*List of PDBs was sorted from a previous report (Scheepers et al., 2016).
N.A.: Not available.
** Different (sub)clusters of SBPs as per a previous classification (Scheepers et al., 2016).

[†](Sub)cluster for same set of SBPs proposed in this study. The newly proposed (sub)clusters are highlighted in bold.

Table D.8. Details of the proteins considered in the protein-protein interaction analysis. The score provided by the program STRING v 11 is also provided.

ORF ID	UniProt ID	Protein name*	Score
TTHA0163	Q5SLY0	Uncharacterized protein	0.738
TTHA0258	Q5SLN5	tRNA-specific 2-thiouridylase MnmA	0.412
TTHA0372	Q5SLC1	UPF0102 protein	0.416
TTHA0375**	Q5SLB8	Putative glycoprotein endopeptidase	***
TTHA0376	Q5SLB7	Sugar ABC transporter, permease protein	0.629
TTHA0377	Q5SLB6	Sugar ABC transporter, permease protein	0.574
TTHA0378	Q5SLB5	Sugar ABC transporter, permease protein	0.477
TTHA0793	Q5SK59	YrdC-like domain-containing protein	0.516
TTHA0831	Q5SK21	PolyA polymerase family protein	0.46
TTHA1252	Q5SIW2	tRNA N^6-adenosine threonylcarbamoyltransferase; TsaD	0.99
TTHA1542	Q5SI38	tRNA(Ile)-lysidine synthase; tilS	0.66
TTHA1968	Q5SGW2	Competence protein ComEC	0.489

*Protein name has been retrieved from the UniProtKB database.
**The ORF ID TTHA0375 was characterized as a protein YeaZ, in this study.
***The ORF ID TTHA0375 was used as a search query in the STRING database.

BIBLIOGRAPHY

- Abe K, Sunagawa N, Terada T, Takahashi Y, Arakawa T, Igarashi K, Samejima M, Nakai H, Taguchi H, Nakajima M and Fushinobu S (2018). Structural and thermodynamic insights into β-1, 2-glucooligosaccharide capture by a solute-binding protein in *Listeria innocua*. *J. Biol. Chem*., 293(23):8812-8828.
- Abou-Hachem M, Karlsson EN, Simpson PJ, Linse S, Sellers P, Williamson MP, Jamieson SJ, Gilbert HJ, Bolam DN and Holst O (2002). Calcium binding and thermostability of carbohydrate binding module CBM4-2 of Xyn10A from *Rhodothermus marinus*. *Biochemistry*, 41(18):5720-5729.
- Abramson J, Smirnova I, Kasho V, Verner G, Kaback HR and Iwata S (2003). Structure and mechanism of the lactose permease of *Escherichia coli*. *Science*, 301(5633):610-615.
- Adhikari R, Singh D, Chandravanshi M, Dutta A and Kanaujia SP (2017). UgpB, a periplasmic component of the UgpABCE ATP-binding cassette transporter, predominantly follows the Sec translocation pathway. *Meta Gene*, 13:129-139.
- Alarico S, Empadinhas N and da Costa MS (2013). A new bacterial hydrolase specific for the compatible solutes α-D-mannopyranosyl-(1→2)-D-glycerate and α-D-glucopyranosyl-(1→2)-D-glycerate. *Enzyme Microb. Technol*., 52:77-83.
- Alarico S, Empadinhas N, Mingote A, Simoes C, Santos MS and da Costa MS (2007). Mannosylglycerate is essential for osmotic adjustment in *Thermus thermophilus* strains HB27 and RQ-1. *Extremophiles*, 11:833-840.
- Alarico S, Empadinhas N, Simoes C, Silva Z, Henne A, Mingote A, Santos H and da Costa MS (2005). Distribution of genes for synthesis of trehalose and mannosylglycerate in *Thermus* spp. and direct correlation of these genes with halotolerance. *Appl. Environ. Microbiol*., 71: 2460-2466.
- Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, Zhuo R, Harrell PM, Trinh YT, Zhang Q, Urbatsch IL and Chang G (2009). Structure of P glycoprotein reveals a molecular basis for poly-specific drug binding. *Science*, 323:1718-1722.
- Altschul SF, Gish W, Miller W, Myers EW and Lipmanl DJ (1990). Basic Local Alignment Search Tool. *J. Mol. Biol*., 215:403-210.
- Ambudkar SV, Kim IW, Xia D and Sauna ZE (2006). The A-loop, a novel conserved aromatic acid subdomain upstream of the Walker A motif in ABC transporters, is critical for ATP binding. *FEBS Lett.*, 580:1049-1055.
- Aydin I, Saijo-Hamano Y, Namba K, Thomas C and Roujeinikova A (2011). Structural analysis of the essential resuscitation promoting factor YeaZ suggests a mechanism of nucleotide regulation through dimer reorganization. *PLoS One*, 6(8).
- Balamurugan B, Roshan MN, Shaahul Hameed B, Sumathi K, Senthilkumar R, Udayakumar A, Venkatesh Babu KH, Kalaivani M, Sowmiya G, Sivasankari P and Saravanan S (2007). PSAP: protein structure analysis package. *J. Appl. Crystallogr*., 40(4):773-777.
- Bastian M, Heymann S and Jacomy M (2009). Gephi: an open source software for exploring and manipulating networks. In Third international AAAI conference on weblogs and social media.
- Battye TGG, Kontogiannis L, Johnson O, Powell HR and Leslie AG (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.*, 67:271-281.

BIBLIOGRAPHY

- Begg SL, Eijkelkamp BA, Luo Z, Couñago RM, Morey JR, Maher MJ, Ong CL, McEwan AG, Kobe B, O'Mara ML and Paton JC (2015). Dysregulation of transition metal ion homeostasis is the molecular basis for cadmium toxicity in *Streptococcus pneumoniae*. *Nat. Commun*., 6:6418.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA and Wheeler DL (2000). GenBank. *Nucleic Acids Res.*, 28(1):15-18.
- Berger EA (1973). Different mechanisms of energy coupling for the active transport of proline and glutamine in *Escherichia coli. Proc. Natl. Acad. Sci*. USA, 70:1514-1518.
- Berger EA and Heppel LA (1974). Different mechanisms of energy coupling for the shock-sensitive and shock-resistant amino acid permeases of *Escherichia coli*. *J. Biol. Chem.*, 249:7747-7755.
- Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, Weissig H and Westbrook J (2000). The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol*., 7:957-959.
- Berntsson RP, Smits SH, Schmitt L, Slotboom DJ and Poolman B (2010). A structural classification of substrate-binding proteins. *FEBS Lett.*, 584(12):2606-2617.
- Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L and Schwede T (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*., 42:252-258.
- Binkowski TA, Naghibzadeh S and Liang J (2003). CASTp: computed atlas of surface topography of proteins. *Nucleic Acids Res*., 31:352-355.
- Boos W and Shuman H (1998). Maltose/maltodextrin system of *Escherichia coli*: transport, metabolism, and regulation. *Microbiol. Mol. Biol. Rev*., 62:204-229.
- Bräsen C, Esser D, Rauch B and Siebers B (2014). Carbohydrate metabolism in Archaea: current insights into unusual enzymes and pathways and their regulation. *Microbiol. Mol. Biol. Rev.*, 78(1):89-175.
- Brunger AT (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355:472-475.
- Chandravanshi M, Gogoi P and Kanaujia SP (2016). Computational characterization of TTHA0379: A potential glycerophosphocholine binding protein of Ugp ATP-binding cassette transporter. *Gene*, 592:260-268.
- Chandravanshi M, Sharma A, Dasgupta P, Mandal SK and Kanaujia SP (2019). Identification and characterization of ABC transporters for carbohydrate uptake in *Thermus thermophilus* HB8. *Gene*, 696:135-148.
- Chandravanshi M, Gogoi P and Kanaujia SP (2020). Structural and thermodynamic correlation illuminates the selective transport mechanism of disaccharide α-glycosides through ABC transporter. *FEBS J*., 287(8):1576-1597.
- Chen J, Lu G, Lin J, Davidson AL and Quiocho FA (2003). A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Mol. Cell*., 12:651-661.
- Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS and Richardson DC (2010). MolProbity: all-atom

structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr*., 66:12-21.

- Chen WY, Huang HM, Lin CC, Lin FY and Chan YC (2003). Effect of temperature on hydrophobic interaction between proteins and hydrophobic adsorbents: studies by isothermal titration calorimetry and the van't Hoff equation. *Langmuir*, 19(22):9395-9403.
- Cheong KA, Kim TJ, Yoon JW, Park CS, Lee TS, Kim YB, Park KH and Kim JW (2002). Catalytic activities of intracellular dimeric neopullulanase on cyclodextrin, acarbose and maltose. *Biotechnol. Appl. Biochem*., 35:27-34.
- Chou SH and Galperin MY (2016). Diversity of cyclic di-GMP-binding proteins and mechanisms. *J. Bacteriol*., 198(1):32-46.
- Chuenchor W, Pengthaisong S, Robinson RC, Yuvaniyama J, Svasti J and Cairns JR (2011). The structural basis of oligosaccharide binding by rice BGlu1 beta-glucosidase. *J. Struct. Biol.*,173(1):169-179.
- Conners SB, Montero CI, Comfort DA, Shockley KR, Johnson MR, Chhabra SR and Kelly RM (2005). An expression-driven approach to the prediction of carbohydrate transport and utilization regulons in the hyperthermophilic bacterium *Thermotoga maritima*. *J. Bacteriol*., 187:7267-7282.
- Culurgioni S, Harris G, Singh AK, King SJ and Walsh MA (2017). Structural basis for regulation and specificity of fructooligosaccharide import in *Streptococcus pneumoniae*. *Structure*, 25(1):79-93.
- Cuneo MJ, Changela A, Beese LS and Hellinga HW (2009a). Structural adaptations that modulate monosaccharide, disaccharide, and trisaccharide specificities in periplasmic maltose-binding proteins. *J. Mol. Biol.*, 389(1):157-166.
- Cuneo MJ, Beese LS and Hellinga HW (2009b). Structural analysis of semi-specific oligosaccharide recognition by a cellulose-binding protein of *Thermotoga maritima* reveals adaptations for functional diversification of the oligopeptide periplasmic binding protein fold. *J. Biol. Chem.*, 284(48):33217-33223.
- Cuneo MJ, Changela A, Warren JJ, Beese LS and Hellinga HW (2006). The crystal structure of a thermophilic glucose binding protein reveals adaptations that interconvert mono and di-saccharide binding sites. *J. Mol. Biol.*, 362:259-270.
- D'Arcy A, Villard F and Marsh M (2007). An automated microseed matrix-screening method for protein crystallization. *Acta Crystallogr. D Biol. Crystallogr.*, 63(4):550-554.
- Dassa E and Bouige P (2001). The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. *Res. Microbiol*., 152:211-229.
- Davidson AL and Chen J (2004). ATP-binding cassette transporters in bacteria. *Annu. Rev. Biochem.*, 73:241-268.
- Davidson AL, Dassa E, Orelle C and Chen J (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. *Micro. Mol. Biol. Rev*., 72:317-364.
- Davidson AL, Shuman HA and Nikaido H (1992). Mechanism of maltose transport in *Escherichia coli*: transmembrane signaling by periplasmic binding proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 89:2360-2364.

BIBLIOGRAPHY

- Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall III WB, Snoeyink J, Richardson JS and Richardson DC (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, 35(suppl_2):W375-W383.
- Dawson RJP and Locher KP **(**2006) Structure of a bacterial multidrug ABC transporter. *Nature*, 443:180-185
- De Beer TA, Berka K, Thornton JM and Laskowski RA (2014). PDBsum additions. *Nucleic Acids Res.*, 42(D1):D292-D296.
- de Boer M, Gouridis G, Muthahari YA and Cordes T (2019a). Single-molecule observation of ligand binding and conformational changes in FeuA. *Biophys. J.*, 117(9):1642-1654.
- de Boer M, Gouridis G, Vietrov R, Begg SL, Schuurman-Wolters GK, Husada F, Eleftheriadis N, Poolman B, McDevitt CA and Cordes T (2019b). Conformational and dynamic plasticity in substrate-binding proteins underlies selective transport in ABC importers. *Elife*, 8: e44652.
- Dean M and Allikmets R (1995). Evolution of ATP-binding cassette transporter genes. *Curr. Opin. Genet. Dev.*, 5:779-785.
- Deka RK, Brautigam CA, Yang XF, Blevins JS, Machius M, Tomchick DR and Norgard MV (2006). The PnrA (Tp0319; TmpC) lipoprotein represents a new family of bacterial purine nucleoside receptor encoded within an ATP-binding cassette (ABC)-like operon in *Treponema pallidum*. *J. Biol. Chem.*, 281:8072-8081.
- 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. *J. Mol. Biol.*, 305:905-915.
- Dion M, Fourage L, Hallet JN and Colas B (1999). Cloning and expression of a β-glycosidase gene from *Thermus thermophilus*. Sequence and biochemical characterization of the encoded enzyme. *Glycoconj. J.*, 16(1):27-37.
- Doeven MK, Van den Bogaart G, Krasnikov V and Poolman B (2008). Probing receptor-translocator interactions in the oligopeptide ABC transporter by fluorescence correlation spectroscopy. *Biophys. J.*, 94(10):3956-3965.
- Eitinger T, Rodionov DA, Grote M and Schneider E (2011). Canonical and ECF-type ATP-binding cassette importers in prokaryotes: diversity in modular organization and cellular functions. *FEMS Microbiol Rev.*, 35:3-67.
- Ejby M, Fredslund F, Andersen JM, Žagar AV, Henriksen JR, Andresen TL, Svensson B, Slotboom DJ and Hachem MA (2016). An ATP-binding cassette transporter mediates the uptake of α-(1, 6)-linked dietary oligosaccharides in Bifidobacterium and correlates with competitive growth on these substrates. *J. Biol. Chem.*, 291(38): 20220-20231.
- El Yacoubi B, Bailly M and de Crécy-Lagard V (2012). Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu. Rev. Genet.*, 46:69-95.
- Elbein AD, Pan YT, Pastuszak I and Carroll D (2003). New insights on trehalose: a multifunctional molecule. *Glycobiology*, 13:17-27.

BIBLIOGRAPHY

- Elbourne LD, Tetu SG, Hassan KA and Paulsen IT (2017). TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res.*, 45(D1):D320-D324.
- Emsley P, Lohkamp B, Scott WG and Cowtan K (2010). Features and development of Coot. *Acta. Crystallogr. D Biol. Crystallogr.*, 66:486-501.
- Evans P and McCoy A (2008). An introduction to molecular replacement. *Acta Crystallogr D Biol Crystallogr*., 64(Pt 1):1-10.
- Evans PR and Murshudov GN (2013). How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.*, 69:1204-1214.
- Faller M, Niederweis M and Schulz GE (2004). The structure of a mycobacterial outer-membrane channel. *Science*, 303(5661):1189-1192.
- Fath MJ and Kolter R (1993). ABC transporters: bacterial exporters. *Microbiol. Rev.*, 57:995-1017.
- Fenn JS, Nepravishta R, Guy CS, Harrison J, Angulo J, Cameron AD and Fullam E (2019). Structural basis of glycerophosphodiester recognition by the *Mycobacterium tuberculosis* substrate-binding protein UgpB. *ACS Chem. Biol*.,14(9):1879-1887.
- Ferreira MJ and Sa-Nogueira I (2010). A multitask ATPase serving different ABC-type sugar importers in *Bacillus subtilis*. *J. Bacteriol*., 192:5312-5318.
- Flocco MM and Mowbray SL (1994). The 1.9 A X-ray structure of a closed unliganded form of the periplasmic glucose/galactose receptor from *Salmonella typhimurium*. *J. Biol. Chem.*, 269(12):8931-8936.
- Fonin AV, Povarova OI, Staiano M, D'Auria S, Turoverov KK and Kuznetsova IM (2014). The trehalose/maltose-binding protein as the sensitive element of a glucose biosensor. *Opt. Mater*., 36(10):1676-1679.
- Freundlieb S and Boos W (1986). Alpha-amylase of *Escherichia coli*, mapping and cloning of the structural gene, *malS*, and identification of its product as a periplasmic protein. *J. Biol. Chem*., 261(6), 2946-2953.
- Fuellen G, Spitzer M, Cullen P and Lorkowski S (2005). Correspondence of function and phylogeny of ABC proteins based on an automated analysis of 20 model protein data sets. *Proteins*, 61:888-899.
- Fukami-Kobayashi K, Tateno Y and Nishikawa K (1999) Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *J. Mol. Biol.*, 286:279-290.
- Garcia-Vallvé, S and Puigbo, PERE (2009). DendroUPGMA: a dendrogram construction utility. Universitat Rovira i Virgili.
- Garmory HS and Titball RW (2004). ATP-Binding Cassette Transporters Are Targets for the Development of Antibacterial Vaccines and Therapies. *Infect. Immun.*, 72(12):6757-6763.
- Georgelis N, Yennawar NH and Cosgrove DJ (2012). Structural basis for entropy-driven cellulose binding by a type-A cellulose-binding module (CBM) and bacterial expansin. *Proc. Natl. Acad. Sci. U.S.A*.,109(37):14830-14835.
- Gerber S, Comellas-Bigler M, Goetz BA and Locher KP (2008). Structural basis of trans-inhibition in a molybdate/tungstate ABC transporter. *Science*, 321(5886):246-250.

- Giaever HM, Styrvold OB, Kaasen IN and Strøm AR (1998). Biochemical and genetic characterization of osmoregulatorytrehalose synthesis in *Escherichia coli*. *J. Bacteriol*., 170(6): 2841-2849.
- Gonçalves S, Borges N, Esteves AM, Victor BL, Soares CM, Santos H and Matias PM (2010). Structural analysis of *Thermus thermophilus* HB27 mannosyl-3-phosphoglycerate synthase provides evidence for a second catalytic metal ion and new insight into the retaining mechanism of glycosyltransferases. *J. Biol. Chem.*, 285(23):17857-17868.
- Gonin S, Arnoux P, Pierru B, Lavergne J, Alonso B, Sabaty M and Pignol D (2007). Crystal structures of an Extracytoplasmic Solute Receptor from a TRAP transporter in its open and closed forms reveal a helix-swapped dimer requiring a cation for alpha-keto acid binding. *BMC Struct. Biol.*, 7:11.
- Gouet P, Robert X and Courcelle E (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res*., 31:3320-3323.
- Gouridis G, Schuurman-Wolters GK, Ploetz E, Husada F, Vietrov R, de Boer M, Cordes T and Poolman B (2015). Conformational dynamics in substrate-binding domains influences transport in the ABC importer GlnPQ. *Nat. Struct. Mol. Biol.*, 22:57-64.
- Grosjean H, de Crecy-Lagard V and Marck C (2010). Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Lett.*, 584:252-264.
- Guo AC, Jewison T, Wilson M, Liu Y, Knox C, Djoumbou Y, Lo P, Mandal R, Krishnamurthy R and Wishart DS (2012). ECMDB: the *E. coli* Metabolome Database. *Nucleic Acids Res.*, 41(D1):D625-630.
- Han Y, Agarwal V, Dodd D, Kim J, Bae B, Mackie RI, Nair SK and Cann IK (2012). Biochemical and structural insights into xylan utilization by the thermophilic bacterium *Caldanaerobius polysaccharolyticus*. *J. Biol. Chem*., 287(42):34946-34960.
- Hayward S and Berendsen HJ (1998) Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins*, 30: 144-154.
- He J, Yin W, Galperin MY and Chou SH (2020). Cyclic di-AMP, a second messenger of primary importance: tertiary structures and binding mechanisms. *Nucleic Acids Res.*, 48(6):2807-2829.
- Henne A, Brüggemann H, Raasch C, Wiezer A, Hartsch T, Liesegang H, Johann A, Lienard T, Gohl O, Martinez-Arias R and Jacobi C (2004). The genome sequence of the extreme thermophile *Thermus thermophilus*. *Nat. Biotechnol*., 22(5):547.
- Herman P, Barvik I, Staiano M, Vitale A, Vecer J and D'Auria S (2007). Temperature modulates binding specificity and affinity of the d-trehalose/d-maltose-binding protein from the hyperthermophilic archaeon *Thermococcus litoralis*. *Biochim. Biophys. Acta, Proteins Proteomics*, 1774(5):540-544.
- Herman P, Staiano M, Marabotti A, Varriale A, Scire A, Tanfani F, Vecer J, Rossi M and D'Auria S (2006). D-Trehalose/D-maltose-binding protein from the hyperthermophilic archaeon *Thermococcus litoralis*: The binding of trehalose and

maltose results in different protein conformational states. *Proteins: Struct. Funct. Bioinf.*, 63(4):754-767.

- Higgins CF (1992). ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol*., 8:67-113.
- Hölemann A and Seeberger PH (2004). Carbohydrate diversity: synthesis of glycoconjugates and complex carbohydrates. *Curr. Opin. Biotechnol.*, 15:615-622.
- Hollenstein K, Dawson RJ and Locher KP (2007) Structure and mechanism of ABC transporter proteins. *Curr Opin Struct Biol*.,17(4):412-418.
- Holm L and Rosenstrom P (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res*., 38:545-549.
- Hung LW, Wang IX, Nikaido K, Liu PQ, Ames GF and Kim SH (1998). Crystal structure of the ATP-binding subunit of an ABC transporter. *Nature*, 396:703-707.
- Hurtubise Y, Shareck F, Kluepfel D and Morosoli R (1995). A cellulase/xylanase-negative mutant of *Streptomyces lividans* 1326 defective in cellobiose and xylobiose uptake is mutated in a gene encoding a protein homologous to ATP-binding proteins. *Mol. Microbiol*., 17:367-377.
- Hussain AS, Shanthi V, Sheik SS, Jeyakanthan J, Selvarani P and Sekar K (2002). PDB Goodies–a web-based GUI to manipulate the Protein Data Bank file. *Acta Crystallogr D Biol Crystallogr*., 58(8):1385-6.
- Irwin JJ and Shoichet BK (2005). ZINC− a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model*, 45(1):177-182.
- Iturriaga G, Suarez R and Nova-Franco B (2009). Trehalose metabolism: from osmoprotection to signaling. *Int. J. Mol. Sci.*, 10:3793-3810.
- Jamal-Talabani S, Boraston AB, Turkenburg JP, Tarbouriech N, Ducros VM and Davies GJ (2004). *Ab initio* structure determination and functional characterization of CBM36: a new family of calcium-dependent carbohydrate binding modules. *Structure*, 12(7):1177-1187.
- Jelesarov I and Bosshard HR (1999). Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J. Mol. Biol.*, 12(1):3-18.
- Jensen JB, Peters NK and Bhuvaneswari TV (2002). Redundancy in periplasmic binding protein-dependent transport systems for trehalose, sucrose, and maltose in *Sinorhizobium meliloti*. *J. Bacteriol.*, 184(11):2978-2986.
- Jiang D, Zhang Q, Zheng Q, Zhou H, Jin J, Zhou W, Bartlam M and Rao Z (2014). Structural analysis of *Mycobacterium tuberculosis* ATP-binding cassette transporter subunit UgpB reveals specificity for glycerophosphocholine. *FEBS J.*, 281:331-341.
- Jones PM and George AM (2004). The ABC transporter structure and mechanism: perspectives on recent research. *Cell. Mol. Life Sci*., 61:682-689.
- Kallberg M, Wang H, Wang S, Peng J, Wang Z, Lu H and Xu J (2012). Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc*., 7:1511-1522.
- Kalscheuer R, Weinrick B, Veeraraghavan U, Besra GS and Jacobs WR (2010). Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA*, 107(50): 21761-21766.

BIBLIOGRAPHY

- Kanehisa M and Goto S (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, 28:27-30.
- Kang J, Park J, Choi H, Burla B, Kretzschmar T, Lee Y and Martinoia E (2011). Plant ABC transporters. The Arabidopsis book/American Society of Plant Biologists, 9.
- Karpowich NK, Huang HH, Smith PC and Hunt JF (2003). Crystal structures of the BtuF periplasmic-binding protein for vitamin B12 suggest a functionally important reduction in protein mobility upon ligand binding. *J.Biol. Chem.*, 278:8429-8434
- Kerr ID (2002). Structure and association of ATP-binding cassette transporter nucleotide-binding domains. *Bioch. Biophs. Acta.*, 1561:47-64.
- Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B and Zaslavsky L (2019). PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.*, 47(D1):D1102-1109.
- Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA and Wang J (2015). PubChem substance and compound databases. *Nucleic Acids Res.*, 44(D1):D1202-D1213.
- Koebnik R, Locher KP and Van Gelder P (2000). Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol. Microbiol.*, 37:239-253.
- Koning SM, Albers SV, Konings WN and Driessen AJM (2002). Sugar transport in (hyper) thermophilic archaea. *Res. Microbiol.*, 153:61-67.
- Korkhov VM, Mireku SA and Locher KP (2012). Structure of AMP-PNP bound vitamin B12 transporter BtuCD-F. *Nature*, 490:367-372.
- Krainer G and Keller S (2015). Single-experiment replacement assay for quantifying high-affinity binding by isothermal titration calorimetry. *Methods*, 76:116-123.
- Kretzschmar T, Burla B, Lee Y, Martinoia E and Nagy R (2011). Functions of ABC transporters in plants. *Essays in biochemistry*, 50:145-160.
- Krissinel E and Henrick K (2019). Protein structure comparison service PDBeFold at European Bioinformatics Institute.
- Kuan G, Dassa E, Saurin W, Hofnung M and Saier MH Jr (1995). Phylogenetic analyses of the ATP-binding constituents of bacterial extracytoplasmic receptor-dependent ABC-type nutrient uptake permeases. *Res. Microbiol.*, 146:271-278.
- Kumar S, Stecher G and Tamura K (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.*, 33(7):1870-1874.
- Laskowski RA, MacArthur MW, Moss DS and Thornton JM (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, 26:283-291.
- Lee HS, Kim MS, Cho HS, Kim JI, Kim TJ, Choi JH, Park C, Lee HS, Oh BH and Park KH (2002). Cyclomaltodextrinase, neopullulanase, and maltogenic amylase are nearly indistinguishable from each other. *J. Biol. Chem.*, 277(24):21891-21897.
- Lee RA, Razaz M and Hayward S (2003). The DynDom database of protein domain motions. *Bioinformatics*, 19:10.
- Lee YH, Deka RK, Norgard MV, Radolf JD and Hasemann CA (1999) Treponema pallidum TroA is a periplasmic zinc-binding protein with a helical backbone. *Nat Struct Biol.*, 6(7):628-633.

BIBLIOGRAPHY

- Lefevre F and Boutry M (2018). Towards identification of the substrates of ATP-binding cassette transporters. Plant physiol., 178(1):18-39.
- Letunic I and Bork P (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*., 44(W1):W242-245.
- Lewinson O and Livnat-Levanon N (2017). Mechanism of action of ABC importers: conservation, divergence, and physiological adaptations. *J. Mol. Biol*., 429(5):606-619.
- Lewis VG, Ween MP and McDevitt CA (2012). The role of ATP-binding cassette transporters in bacterial pathogenicity. *Protoplasma*, 249:919-942.
- Li H, Ji X, Zhou Z, Wang Y and Zhang X (2010). *Thermus thermophilus* proteins that are differentially expressed in response to growth temperature and their implication in thermos adaptation. *J. Proteome Res.*, 9(2):855-864.
- Light SH, Cahoon LA, Halavaty AS, Freitag NE and Anderson WF (2017). Structure to function of an α-glucan metabolic pathway that promotes *Listeria monocytogenes* pathogenesis. *Nature Microbiol*., 2(2):16202.
- Locher KP, Lee AT and Rees DC (2002) The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science*, 296:1091-1098.
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM and Henrissat B (2014) The Carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.*, 42:D490-D495.
- Loo TW, Bartlett MC and Clarke DM (2003). Simultaneous binding of two different drugs in the binding pocket of the human multidrug resistance P-glycoprotein. *J. Biol. Chem.*, 278:39706-39710.
- Lu G, Westbrooks JM, Davidson AL and Chen J (2005). ATP hydrolysis is required to reset the ATP-binding cassette dimer into the resting-state conformation. *Proc. Natl. Acad. Sci. USA*, 102:17969-17974.
- Magnusson U, Salopek-Sondi B, Luck LA and Mowbray SL (2004). X-ray structures of the leucine-binding protein illustrate conformational changes and the basis of ligand specificity. *J. Biol. Chem.*, 279:8747-8752.
- Mao B, Pear M, McCammon J and Quiocho F (1982). Hinge-bending in L-arabinose-binding protein. The "Venus's-flytrap" model. *J. Biol. Chem*., 257:1131-1133.
- Maqbool A, Horler RS, Muller A, Wilkinson AJ, Wilson KS and Thomas GH (2015). The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity. *Biochem. Soc. Trans.*, 43:1011-1017.
- Marinelli F, Kuhlmann SI, Grell E, Kunte HJ, Ziegler C and Faraldo-Gómez JD (2011). Evidence for an allosteric mechanism of substrate release from membrane-transporter accessory binding proteins. *Proc. Natl. Acad. Sci. U.S.A*., 108(49):E1285-E1292.
- Marion C, Aten AE, Woodiga SA and King SJ (2011). Identification of an ATPase, MsmK, which energizes multiple carbohydrate ABC transporters in *Streptococcus pneumonia*. *Infect. Immun*., 79:4193- 4200.
- Matsumoto N, Yamada M, Kurakata Y, Yoshida H, Kamitori S, Nishikawa A and Tonozuka T (2009). Crystal structures of open and closed forms of cyclo/maltodextrin-binding protein. *FEBS J.*, 276(11):3008-3019.

BIBLIOGRAPHY

- Matthews BW (1968). Solvent content of protein crystals. *J. Mol. Biol.* 33:491-497.
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007). Phaser crystallographic software. *J. Appl. Crystallogr.*, 40:658-674.
- Medrano FJ, Souza CSD, Romero A and Balan A (2014). Structure determination of a sugar-binding protein from the phytopathogenic bacterium *Xanthomonas citri*. *Acta Crystallogr F Struct Biol Commun.*, 70: 564-571.
- Miller D, Olson JS and Quiocho FA (1980). The mechanism of sugar binding to the periplasmic receptor for galactose chemotaxis and transport in *Escherichia coli*. *J. Biol. Chem.*, 255:2465-2471.
- Mitchell P (1967). Translocations through natural membranes. *Adv. Enzymol*., 29:33-87.
- Miyazaki T, Ichikawa M, Iino H, Nishikawa A and Tonozuka T (2015). Crystal structure and substrate-binding mode of GH63 mannosylglycerate hydrolase from *Thermus thermophilus* HB8. *J. Struct. Biol.*, 190(1):21-30.
- Montanier C, Flint JE, Bolam DN, Xie H, Liu Z, Rogowski A, Weiner D, Ratnaparkhe S, Nurizzo D, Roberts SM and Turkenburg JP (2010). Circular permutation provides an evolutionary link between two families of calcium-dependent carbohydrate binding modules. *J. Biol. Chem.*, 285(41):31742-31754.
- Morris GM, Huey R and Olson AJ (2008) Using AutoDock for ligand–receptor docking. *Curr. Protoc. Bioinformatics*, 24:8.14.1–8.14.40.
- Mulligan C, Geertsma E, Severi E, Kelly D, Poolman B and Thomas G (2009). The substrate-binding protein imposes directionality on an electrochemical sodium gradient-driven TRAP transporter. *Proc. Natl. Acad.Sci.* USA, 106:1778-1783.
- Munishkina LA and Fink AL (2007). Fluorescence as a method to reveal structures and membrane-interactions of amyloidogenic proteins. *Biochim. Biophys. Acta*, 1768(8):1862-1885.
- Nakajima M, Yoshida R, Miyanaga A, Abe K, Takahashi Y, Sugimoto N, Toyoizumi H, Nakai H, Kitaoka M and Taguchi H (2016). Functional and structural analysis of a β-glucosidase involved in β-1, 2-glucan metabolism in *Listeria innocua*. *PLoS One*,11(2).
- Nikaido H (1994). Prevention of drug access to bacterial targets: permeability barriers and active efflux. *Science.*, 264(5157):382-388.
- Nikaido H and Saier MH Jr (1992). Transport proteins in bacteria: common themes in their design. *Science*, 258:936-942.
- Ohshima N, Yamashita S, Takahashi N, Kuroishi C, Shiro Y and Takio K (2008). *Escherichia coli* cytosolic glycerophosphodiester phosphodiesterase (UgpQ) requires Mg^{2+}, Co^{2+}, or Mn^{2+} for its enzyme activity. *J. Bacteriol.*, 190:1219-1223.
- Ohtani N, Tomita M and Itaya M (2012). The third plasmid pVV8 from *Thermus thermophilus* HB8: isolation, characterization, and sequence determination Extrem. *Life Extrem. Cond.*, 16: 237-244.
- Oiki S, Kamochi R, Mikami B, Murata K and Hashimoto W (2017). Alternative substrate-bound conformation of bacterial solute-binding protein involved in the import of mammalian host glycosaminoglycans. *Sci. Rep.*, 7(1):1-3.
- Oldham ML and Chen J (2011). Crystal structure of the maltose transporter in a pre-translocation intermediate state. *Science*, 332:1202-1205.

BIBLIOGRAPHY

- Oldham ML, 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(45):18132-18137.
- Oldham ML, Khare D, Quiocho FA, Davidson AL and Chen J (2007) Crystal structure of a catalytic intermediate of the maltose transporter. *Nature*, 450(7169):515-521.
- O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD and Pruitt KD (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, 44(D1): D733-745.
- Orelle C, Ayvaz T, Everly RM, Klug CS and Davidson AL (2008). Both maltose-binding protein and ATP are required for nucleotide-binding domain closure in the intact maltose ABC transporter. *Proc. Natl. Acad. Sci. USA*, 105:12837-12842.
- Pandey S, Modak A, Phale PS and Bhaumik P (2016). High resolution structures of periplasmic glucose binding protein of *Pseudomonas putida* CSV86 reveal structural basis of its substrate specificity. *J. Biol. Chem.*, 291(15):7844-7857.
- Parcej D and Tampé R (2007). Caught in the act: an ABC transporter on the move. *Structure*, 15(9):1028-1030.
- Paul S, Diekema D and Moye-Rowley WS (2013). Contributions of *Aspergillus fumigatus* ATP-binding cassette transporter proteins to drug resistance and virulence. *Eukaryot. Cell*, 12:1619- 1628.
- Pei J and Grishin NV (2014). PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods Mol. Biol*., 1079:263-271
- Petersen TN, Brunak S, von Heijne G and Nielsen H (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, 8(10):785.
- Potrykus K and Cashel M (2008). (p)ppGpp: still magical? Annu. Rev. Microbiol., 62:35-51.
- Povarova OI, Stepanenko OV, Sulatskaya AI, Kuznetsova IM, Turoverov KK, Staiano M, Vitale A and D'Auria S (2010). High stability of trehalose/maltose binding protein from *Thermococcus litoralis* makes it a good candidate as a sensitive element in biosensor systems for sugar control. *J. Spectrosc*., 24(3-4):349-53.
- Qu Q, Lee SJ and Boos W (2004). TreT, a novel trehaloseglycosyltransferring synthase of the hyperthermophilic archaeon *Thermococcus litoralis*. *J. Biol. Chem.*, 279 (46):47890-47897.
- Quiocho FA and Ledvina P (1996). Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol. Microbiol*., 20:17-25.

BIBLIOGRAPHY

- Raich L, Nin-Hill A, Ardèvol A and Rovira C (2016). Enzymatic Cleavage of Glycosidic Bonds: Strategies on How to Set Up and Control a QM/MM Metadynamics Simulation. In Methods in enzymology, Academic Press, 577:159-183.
- Rees DC, Johnson E and Lewinson O (2009). ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.* 10, 218-227.
- Rice AJ, Park A and Pinkett HW (2014). Diversity in ABC transporters: Type I, II and III importers. *Crit. Rev. Biochem. Mol. Biol.*, 49(5):426-437.
- Rodionov DA, Hebbeln P, Eudes A, ter Beek J, Rodionova IA, Erkens GB, Slotboom DJ, Gelfand MS, Osterman AL, Hanson AD and Eitinger T (2009). A novel class of modular transporters for vitamins in prokaryotes. *J. Bacteriol*., 191:42-51.
- Rodionov DA, Rodionova IA, Li X, Ravcheev DA, Tarasova Y, Portnoy VA, Zengler K and Osterman AL (2013). Transcriptional regulation of the carbohydrate utilization network in *Thermotogamaritima*. *Front Microbiol.*, 4.
- Røjel N, Kari J, Sørensen TH, Badino SF, Morth JP, Schaller K, Cavaleiro AM, Borch K and Westh P (2019). Substrate binding in the processive cellulase Cel7A: Transition state of complexation and roles of conserved tryptophan residues. *J. Biol. Chem.*, RA119.
- Saier MH Jr (2000a). Families of transmembrane sugar transport proteins**.** *Mol. Microbiol.*, 35(4):699-710.
- Saier MH (2000b). A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol. Mol. Biol. Rev.*, 64:354-411.
- Saier MH Jr (2000c). Families of proteins forming transmembrane channels. *J. Membr. Biol.*, 175:165-180.
- Saier MH, Reddy VS, Tsu BV, Ahmed MS, Li C and Moreno-Hagelsieb G (2016). The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res*., 44:D372-D379.
- Santos H and Da Costa MS (2002). Compatible solutes of organisms that live in hot saline environments. *Environ. Microbiol*., 4(9):501-509.
- Saurin W, Hofnung M and Dassa E (1999). Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *J. Mol. Evol.*, 48:22-41.
- Scheepers GH, Lycklama a Nijeholt JA and Poolman B (2016). An updated structural classification of substrate-binding proteins. *FEBS Lett.*, 590(23):4393-401.
- Schirmer T (1998). General and specific porins from bacterial outer membranes. *J. Struct. Biol.*, 121(2):101-109.
- Schlösser A (2000). MsiK-dependent trehalose uptake in *Streptomyces reticuli*. *FEMS Microbiol. Lett*., 184:187-192.
- Schlösser A, Kampers T and Schrempf H (1997). The *Streptomyces* ATP binding component MsiK assists in cellobiose and maltose transport. *J. Bacteriol.*, 179:2092-2095.
- Schneider E (2001). ABC transporters catalyzing carbohydrate uptake. *Res. Microbiol.*, 152:303-310.

BIBLIOGRAPHY

- Schneider E and Hunke S (1998). ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP hydrolyzing subunits/domains. *FEMS Microbiol. Rev.*, 22:1-20.
- Senguen FT and Grabarek Z (2012). X-ray structures of magnesium and manganese complexes with the N-terminal domain of calmodulin: insights into the mechanism and specificity of metal ion binding to an EF-hand. *Biochemistry*, 51(31):6182-6194.
- Shaw Stewart PD, Kolek SA, Briggs RA, Chayen NE and Baldock PF (2011). Random microseeding: a theoretical and practical exploration of seed stability and seeding techniques for successful protein crystallization. *Cryst. Growth Des.*, 11(8):3432-3441.
- Shilton BH, Flocco MM, Nilsson M and Mowbray SL (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.
- Shitan N, Bazin I, Dan K, Obata K, Kigawa K, Ueda K, Sato F, Forestier C and Yazaki K (2003). Involvement of CjMDR1, a plant multidrug resistance-type ATP-binding cassette protein, in alkaloid transportin *Coptis japonica*. *Proc. Natl. Acad. Sci. USA*, 100:751-756.
- Shultis DD, Purdy MD, Banchs CN and Wiener MC (2006). Outer Membrane Active Transport: Structure of the BtuB: TonB Complex. *Science*, 312(5778):1396-1399.
- Sievers F and Higgins DG (2014). Clustal omega, accurate alignment of very large numbers of sequences. *Methods Mol. Biol.*, 1079:105-116.
- Silva DA, Bowman GR, Sosa-Peinado A and Huang X (2011). A role for both conformational selection and induced fit in ligand binding by the LAO protein. *PLoS Comput. Biol.*, 7(5).
- Silva Z, Alarico S, Nobre A, Horlacher R, Marugg J, Boos W, Mingote AI and da Costa MS (2003). Osmotic adaptation of *Thermus thermophilus* RQ-1: lesson from a mutant deficient in synthesis of trehalose. *J. Bacteriol*., 185(20):5943-52.
- Silva Z, Sampaio MM, Henne A, Böhm A, Gutzat R, Boos W, da Costa MS and Santos H (2005). The high-affinity maltose/trehalose ABC transporter in the extremely thermophilic bacterium *Thermus thermophilus* HB27 also recognizes sucrose and palatinose. *J. Bacteriol.*, 187(4):1210-1218.
- Skrynnikov NR, Goto NK, Yang D, Choy WY, Tolman JR, Mueller GA and Kay LE (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 beta-cyclodextrin. *J. Mol. Biol.*, 295:1265-1273.
- Spiwok V (2017). CH/π Interactions in Carbohydrate Recognition. *Molecules*, 22(7):1038.
- Stierand K, Maaß P and Rarey M (2006). Molecular Complexes at a Glance: Automated Generation of two-dimensional Complex Diagrams. *Bioinformatics*, 22:1710-1716.
- Sumathi K, Ananthalakshmi P, Roshan MM and Sekar K (2006). 3dSS: 3D structural superposition. *Nucleic Acids Res*., 34(suppl_2):W128-W132.
- Sun L, Zeng X, Yan C, Sun X, Gong X, Rao Y and Yan N (2012). Crystal structure of a bacterial homologue of glucose transporters GLUT1-4. *Nature*, 490(7420):361.

BIBLIOGRAPHY

- Sun Z, Chen Y, Yang C, Yang S, Gu Y and Jiang W (2015). A novel three-component system-based regulatory model for d-xylose sensing and transport in *Clostridium beijerinckii*. *Mol. Microbiol*., 95(4):576-89.
- Synytsya A and Novak M (2014). Structural analysis of glucans. *Ann. Transl. Med.*, 2(2).
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP and Kuhn M (2014). STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*., 43(D1):D447-D452.
- Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P and Jensen LJ (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.*, 47(D1):D607-D613.
- Szmelcman S and Hofnung (1975). Maltose transport in *Escherichia coli* K12: involvement of the bacteriophage receptor. *J. Bacteriol*., 124:112-118.
- Szollosi D, Rose-Sperling D, Hellmich UA and Stockner T (2018). Comparison of mechanistic transport cycle models of ABC exporters. *Biochim. Biophys. Acta Biomembr.*,1860:818-832.
- Tam R and Saier MH (1993). Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol. Rev*., 57:320-346.
- Tamura K, Stecher G, Peterson D, Filipski A and Kumar S (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol*., 30:2725-2729.
- Tang C, Schwieters CD and Clore GM (2007). Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature*, 449:1078-1082.
- Tang L, El-Din TM, Payandeh J, Martinez GQ, Heard TM, Scheuer T, Zheng N and Catterall WA (2014). Structural basis for Ca^{2+} selectivity of a voltage-gated calcium channel. *Nature*, 505(7481):56.
- Taylor D, Cawley G and Hayward S (2014). Quantitative method for the assignment of hinge and shear mechanism in protein domain movements. Bioinformatics, 30(22):3189-3196.
- Taylor GL (2010). Introduction to phasing. *Acta Crystallogr D Biol Crystallogr*., 66(Pt 4):325-38.
- Terasaka K, Blakeslee JJ, Titapiwatanakun B, Peer WA, Bandyopadhyay A, Makam SN, Lee OR, Richards EL, Murphy AS, Sato F and Yazaki K (2005). PGP4, an ATP binding cassette P-glycoprotein, catalyzes auxin transport in *Arabidopsis thaliana* roots. *Plant Cell*, 17:2922-2939.
- Teze D, Hendrickx J, Czjzek M, Ropartz D, Sanejouand YH, Tran V, Tellier C and Dion M (2013). Semi-rational approach for converting a GH1 β-glycosidase into a β-transglycosidase. *Protein Eng. Des. Sel.*, 27(1):13-19.
- The UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, 47: D506-515.

BIBLIOGRAPHY

- Tommassen J, Eiglmeier K, Cole ST, Overduin P, Larson TJ and Boos W (1991). Characterization of two genes, *glpQ* and *ugpQ*, encoding glycerophosphoryl diester phosphodiesterases of *Escherichia coli*. *Mol. Gen. Genet.*, 226:321-327.
- Tonozuka T, Sogawa A, Yamada M, Matsumoto N, Yoshida H, Kamitori S, Ichikawa K, Mizuno M, Nishikawa A and Sakano Y (2007). Structural basis for cyclodextrin recognition by *Thermoactinomyces vulgaris*cyclo/maltodextrin-binding protein. *FEBS J.*, 274(8):2109-20.
- Trakhanov S, Vyas NK, Luecke H, Kristensen DM, Ma J and Quiocho FA (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.
- Tsusaki K, Nishimoto T, Nakada T, Kubota M, Chaen H, Fukuda S, Sugimoto T and Kurimoto M (1997). Cloning and sequencing of trehalose synthase gene from *Thermus aquaticus* ATCC33923. *Biochim. Biophys. Acta, Gen. Subj*., 1334(1):28-32.
- Vagin AA (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D Biol. Crystallogr*., 60:2184-2195.
- van der Heide T and Poolman B (2002). ABC transporters: one, two or four extracytoplasmic substrate-binding sites? *EMBO Rep.*, 3:938-943.
- Vázquez-Ibar JL, Guan L, Svrakic M and Kaback HR (2003). Exploiting luminescence spectroscopy to elucidate the interaction between sugar and a tryptophan residue in the lactose permease of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, 100(22):12706-12711.
- Velazquez-Campoy A and Freire E (2006). Isothermal titration calorimetry to determine association constants for high-affinity ligands. *Nat. Protocols*, 1(1):186.
- Walmsley AR , Barrett MP, Bringaud F and Gould GW (1998). Sugar transporters from bacteria, parasites and mammals: structure-activity relationships. *Trends Biochem. Sci*., 23(12):476-481.
- Wang YF, Dutzler R, Rizkallah PJ, Rosenbusch JP and Schirmer T (1997). Channel specificity: structural basis for sugar discrimination and differential flux rates in maltoporin. *J. Mol. Biol.*, 272(1):56-63.
- Wilkens S (2015). Structure and mechanism of ABC transporters. *F1000prime Rep.*, 7.
- Wilkinson AJ (2002). ABC Proteins: From Bacteria to Man. (Holland, B., Kuchler, K., Cole, S.P. and Higgins, C., Eds.), Elsevier Science and Technology Books, London.
- Wilkinson T, Verschueren KHG, Wilkinson T and Verschueren KHG (2003). Crystal structures of periplasmic solute binding proteins in ABC-transport complexes illuminate their function. ABC proteins: from bacteria to man. 187:208.
- Willis LB and Walker GC (1999). A novel *Sinorhizobium meliloti* operon encodes an α-glucosidase and a periplasmic-binding-protein-dependent transport system for α-glycosides. *J. Bacteriol.*, 181(14):4176-4184.
- Winn MD (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, 67:235-242.

BIBLIOGRAPHY

- Wuttge S, Bommer M, Jager F, Martins BM, Jacob S, Licht A, Scheffel F, Dobbek H and Schneider E (2012). Determinants of substrate specificity and biochemical properties of the sn-glycerol-3-phosphate ATP binding cassette transporter (UgpB–AEC2) of *Escherichia coli*. *Mol. Microbiol*., 86:908-920.
- Xie H (2008). Activity assay of membrane transport proteins. *Acta Biochim. Biophys*., 40:269-277.
- Xu D and Zhang Y (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys. J*., 101:2525-2534.
- Yang J, Yan R, Roy A, Xu D, Poisson J and Zhang Y (2015a). The I-TASSER suite: protein structure and function prediction. *Nat. Methods.*, 12:7-8.
- Yang Y, Zhang X, Yin Q, Fang W, Fang Z, Wang X, Zhang X and Xiao Y (2015b). A mechanism of glucose tolerance and stimulation of GH1 β-glucosidases. *Sci. Rep*., 5:17296.
- Zaitseva J, Jenewein S, Jumpertz T, Holland IB and Schmitt L (2005). H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB. *EMBO J*., 24:1901-1910.
- Zhang W, Collinet B, Perrochia L, Durand D and Van Tilbeurgh H (2015). The ATP-mediated formation of the YgjD–YeaZ–YjeE complex is required for the biosynthesis of tRNA t6A in *Escherichia coli*. *Nucleic Acids Res.*, 43(3):1804-1817.
- Zhu Y, Suits MD, Thompson AJ, Chavan S, Dinev Z, Dumon C, Smith N, Moremen KW, Xiang Y, Siriwardena A and Williams SJ (2010). Mechanistic insights into a Ca^{2+}-dependent family of α-mannosidases in a human gut symbiont. *Nat. Chem. Biol.*, 6(2):125.

LIST OF PUBLICATIONS

Publications from thesis

1. **Chandravanshi M,** Samanta R and Kanaujia SP (2021). Structural and thermodynamic insights into the novel dinucleotide-binding protein of ABC transporter unveils its moonlighting function. *FEBS J.*, Accepted. DOI: 10.1111/febs.15774.

2. **Chandravanshi M,** Samanta R and Kanaujia SP (2020). Conformational trapping of a β-glucosides-binding protein unveils the selective two-step ligand-binding mechanism of ABC importers. *J. Mol. Biol*., 432:5711-5734.

3. **Chandravanshi M**, Gogoi P and Kanaujia SP (2020). Structural and thermodynamic correlation illuminates the selective transport mechanism of disaccharide α-glycosides through ABC transporter. *FEBS J*., 287(8):1576-1597.

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

5. **Chandravanshi M**, Gogoi P and Kanaujia SP (2016). Computational characterization of TTHA0379: A potential glycerophosphocholine binding protein of Ugp ATP-binding cassette transporter. *Gene*, 592(2):260-268.

Publications from other collaborative work

1. **Chandravanshi M**, Tripathi SK and Kanaujia SP (2021). Mechanistic insights into the ligand binding of the substrate-binding proteins. Under submission.

2. Dutta A, **Chandravanshi M** and Kanaujia SP (2021). Conserved features of MlaD domain aid the trafficking of hydrophobic molecules. Under submission.

3. Sharma A, Gogoi P, **Chandravanshi M** and Kanaujia SP (2021). Water-mediated structural rearrangement establishes active conformation of caspases for apoptosis and inflammation. *J. Biomol. Struct. Dyn.*, Accepted.

4. Sinha AK, Dutta A, **Chandravanshi M** and Kanaujia SP (2019). An insight into bacterial phospholipase C classification and their translocation through Tat and Sec pathways: A data mining study. *Meta Gene*, 20:100547.

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

LIST OF PUBLICATIONS

6. Adhikari R, Singh D, **Chandravanshi M**, Dutta A and Kanaujia SP (2017). UgpB, a periplasmic component of the UgpABCE ATP-binding cassette transporter, predominantly follows the sec translocation pathway. *Meta Gene*, 13:129-139.

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

8. Gogoi P, **Chandravanshi M**, Mandal SK, Srivastava A and Kanaujia SP (2015). Heterogeneous behavior of metalloproteins towards metal ion binding and selectivity: insights from molecular dynamics studies. *J. Biomol. Struct. Dyn.*, 34(7):1470-1485.