

STRUCTURAL INVESTIGATIONS INTO ASPECTS OF MICROBE-HOST
INTERACTIONS

By

Yi Zheng

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Microbiology and Molecular Genetics

2011

ABSTRACT

STRUCTURAL INVESTIGATIONS INTO ASPECTS OF MICROBE-HOST INTERACTIONS

By

Yi Zheng

Microbes and host organisms continuously interact with each other and have co-evolved delicate systems to cope with these interactions for survival. Investigating this important relationship will lead to better understanding of the pathogenesis mechanism of pathogenic bacteria for better protection as well as better utilization of the symbiotic interaction that will enhance the control of biomass accumulation in agriculture. Structural biology is an indispensable approach that can provide us information at the molecular level to better understand the mechanism of the microbe-host interaction and to develop better treatments. In this thesis, three important proteins, namely YpfP from *Bacillus subtilis*, Sus1 from *Arabidopsis thaliana* and Eha proteins from pathogenic *Escherichia coli* O157: H7 were investigated by biochemical characterization as well as X-ray crystallography to provide the knowledge base for future research.

Structures of AtSus1 in complex with UDP and fructose were determined at resolutions of 2.8 Å and 2.85 Å, respectively. These structures provided insights into sucrose synthesis and cleavage as well as the potential link between the AtSus1 activity and its spatial interactions with cellular targets. They also helped to elucidate the retaining glycosyltransferase mechanism in a broader aspect. In addition, comparison of the retaining glycosyltransferase AtSus1 with an inverting glycosyltransferase revealed an interesting relationship that the substrate specificity

between the two might be related by a simple domain rotation.

Structures of EhaA, B, and D proteins from pathogenic *E. coli* O157: H7 were studied by X-ray crystallography. Diffraction data were collected for EhaA, B, and D C-terminal β domains with the structure of EhaB_c successfully solved to 2.2 Å. Crystals were also obtained for EhaA and EhaB full length proteins but further pursuit of crystals with better diffraction quality are still required. Analysis of the EhaB_c structure revealed an interesting difference with other solved structures of the autotransporter translocation units: EhaB_c has a significantly higher number of bulky aromatic residues inside the barrel and a narrower channel indicated by the lower conductance measured in planar Bilayer Lipid Membranes experiments. However, how these preliminary observations are related to the adhesion function of the Eha proteins is still unknown. Further detailed analysis of structures and activity of Eha proteins is still needed to elucidate this structure-function relationship.

A giant liposome assay was developed to assay the glycosyl transfer activity of BsYpfP. It transferred two glucose moieties to the diacylglycerol with lipid chain varying from C10:0 to C18:1 in this assay. The initial unsuccessful efforts of crystallizing BsYpfP led to an investigation of the fusion module method to facilitate its crystallization. Crystals were obtained with the N-terminal fusion protein of the helix-turn-helix module from TagF protein, but they diffracted poorly. However, this effort offered a clue for the crystallization conditions as well as valuable experience on this method. Further manipulation and characterization are still needed to obtain diffraction quality crystals.

ACKNOWLEDGEMENTS

Among the many people deserving my thanks, I would first and foremost like to express sincere gratitude to my mentor, Professor R. Michael Garavito. His enthusiasm, keen intelligence and breadth of knowledge have always inspired and motivated me. I feel very fortunate to work with him on many challenging but rewarding research projects. I have learned so much from our conversations about protein crystallography and scientific research over the years, and it has been an honor to have worked with such a creative and supportive mentor.

I am deeply grateful to my guidance committee members, Dr. Dennis Arvidson, Dr. Christoph Benning, Dr. Robert Hausinger, and Dr. Rosetta Reusch, for their advice and support throughout my Ph.D. program. I would also like to thank Dr. Carrie Hiser and Fei Li for reviewing my thesis.

Of the many great colleagues that I have had over the years, I would particularly like to thank Dr. Ling Qin, a former member of Dr. Shelagh Ferguson-Miller's lab, for teaching me many practical aspects of the membrane protein crystallography during synchrotron trips. Warm thanks also go to Dr. Eric Moellering and Dr. Changcheng Xu for introducing me into the wonderful world of glycolipids.

I would also like to thank all past and present members of the Garavito's group, for creating such a friendly working environment in the lab and sharing knowledge during various conversations: Amy, Chris, Christine, Dexin, Mike D, Nicole, Rachel, Yanfeng, Young-moon. Their friendship and help make my experience here wonderful. I also owe my thanks to

Professor Shelagh Ferguson-Miller for her encouragement and support and other members of her lab, Denise, Jian, Leann, Martyn, Namjoon, Shujuan, and Xi, for their friendship and good advice as well as the kindred spirits on the structural biology.

The last two years of My Ph.D. work were essentially funded by the Department of Biochemistry and Molecular Biology, where I held a teaching position. I would like to thank my bosses, Dr. Neil Bowlby and Dr. Kathleen Foley, for choosing me for the position and sharing with me the teaching techniques as well as the passion for teaching. I gained more confidence and patience by interacting with students.

I would like to thank staff scientists at LS-CAT and DND-CAT at Advanced Photon Source, Argonne National Laboratory, particularly Dr. Spencer Anderson, Dr. Joseph Brunzelle and Dr. Zdzislaw Wawrzak for their training and help.

I would also like to thank our collaborators, Dr. Yingrui Dai, Dr. Michael Feig, Dr. Alexander Negoda, Dr. Denis Proshlyakov, Dr. Sachin Jadhav, and Dr. Mark Worden for their valuable contributions and stimulating discussions on my research.

Last, but certainly not the least, I would like to thank all my dear friends and family who have supported me in my efforts, showed an interest in my work, or helped take my mind off science for a while. I am especially grateful to my parents for their constant encouragement and unwavering support.

East Lansing, July 2011.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS.....	xii
CHAPTER 1 An introduction to the thesis.....	1
1.1 Background and introduction on microbe-host interactions	1
1.2 Research foci in this thesis.....	2
1.3 References.....	5
CHAPTER 2 Structural and functional studies of glycosyltransferase YpfP from <i>Bacillus subtilis</i>	7
2.1 Background and introduction.....	7
2.1.1 Biological membrane structure and membrane curvature.....	7
2.1.2 Glycolipids in bacteria.....	10
2.1.3 Diacylglycerol-lipoteichoic acid biosynthesis network in <i>B. subtilis</i>	11
2.1.4 Diacylglycerol metabolism and membrane curvature in <i>B. subtilis</i>	14
2.1.5 Glycosyltransferase structural biology	18
2.1.6 Glycosyltransferase YpfP from <i>B. subtilis</i>	22
2.1.7 Crystallization strategies for YpfP	23
2.2 Materials and Methods.....	25
2.2.1 Materials	25
2.2.2 Methods	25
2.3 Results.....	31
2.3.1 Choice of <i>Escherichia. coli</i> strains for the heterologous expression of BsYpfP	31
2.3.2 Choice of detergents to suppress the protein aggregation	34
2.3.3 Production of glycolipids in <i>E. coli</i> as the proof of <i>in vivo</i> activity	38
2.3.4 Determination of the <i>in vitro</i> activity of BsYpfP	44
2.3.5 X-ray crystallography studies of the recombinant BsYpfP	58
2.4 Discussion	61
2.5 Conclusions.....	64
2.6 References	67
CHAPTER 3 Structural studies of adhesins from enterohemorrhagic <i>E. coli</i> (EHEC) O157:H7	74
3.1 Background and introduction.....	74
3.1.1 Pathogenesis of <i>E. coli</i>	74
3.1.2 Secretion systems in Gram negative-bacteria	77
3.1.3 Type V secretion system of Gram-negative bacteria	82
3.1.4 The structure of V _a autotransporters.....	86

3.1.5	The structural implications of autotransporter biogenesis	87
3.1.6	Research aims on Eha proteins	92
3.2	Experimental procedures	93
3.2.1	Materials	93
3.2.2	Expression and purification of Eha proteins	93
3.2.3	Crystallization and cyroprotection of Eha proteins	96
3.2.4	X-ray diffraction data collection and structure determination	97
3.2.5	Planar bilayer lipid membrane (BLM) assay for ion channel activity	97
3.2.6	Multiple sequence alignment of β domain of Eha proteins	98
3.3	Results and Discussion	101
3.3.1	Purification of Eha full length autotransporter proteins and translocation domains	101
3.3.2	Crystallization of EhaA and EhaB full length proteins	102
3.3.3	Crystallization of Eha translocation domains	105
3.3.4	The crystal structure of the EhaB translocation domain (EhaB ₆₇₈₋₉₈₀)	109
3.3.5	The planar BLM studies of the EhaB translocation domain	115
3.4	Conclusions	119
3.5	References	121
CHAPTER 4 The crystal structure determination of sucrose synthase-1 from <i>Arabidopsis thaliana</i>		129
4.1	Introduction	129
4.1.1	Sucrose metabolism and sucrose synthesis in plants	129
4.1.2	Sucrose synthase structure biology	130
4.2	Experimental procedures	131
4.2.1	Materials	132
4.2.2	Methods	132
4.3	Results and Discussion	141
4.3.1	Overall structure of AtSus1	141
4.3.2	Implications of the AtSus1 quaternary structure	152
4.3.3	Sucrose synthase active site	154
4.3.4	Insights into the evolution of the retaining and inverting GT-B glycosyltransferases.	166
4.3.5	Functional implications of the AtSus1 structure	171
4.4	Conclusions	175
4.5	References	177
CHAPTER 5 Conclusion and future directions		185
5.1	Conclusion of the thesis	185
5.2	Future directions	186
5.2.1	Future research directions on BsYpfP	186
5.2.2	Future research directions on Eha proteins	187
5.2.3	Future research directions on AtSus1	189
5.3	References	192

LIST OF TABLES

Table 2.1 <i>Bs</i> YpfP constructs with different tags used in this study.....	32
Table 2.2 Summary of glycolipids generated by YpfP in <i>E. coli</i> revealed by TLC analysis.	39
Table 3.1 Data collection, phasing and refinement statistics for Eha structures.	108
Table 3.2 Aromatic residue distributions in the EhaB, BalP and Hbp β domains.....	113
Table 4.1 Data collection, phasing and refinement statistics for AtSus1 structures.....	137
Table 4.2A Average B-factors for all protein atoms (by chain) in SUS1/Fru/UDP complex	142
Table 4.2B Average B-factors for all protein atoms (by chain) in SUS1/UDP-glc complex	142

LIST OF FIGURES

Figure 2.1 Cartoon representations of membrane structures induced by lipids with different spontaneous membrane curvatures.	9
Figure 2.2 Structures of diacylglycerol and lipoteichoic acid	13
Figure 2.3 Diacylglycerol-lipoteichoic acid pathway enzymes in <i>B. subtilis</i>	13
Figure 2.4 DAG metabolic pathways.....	16
Figure 2.5 Diversity of glycosyltransfer reaction.	19
Figure 2.6 Typical structure folds of GT-A and GT-B glycosyltransferases.	21
Figure 2.7 A representation of giant liposome.....	28
Figure 2.8 Hybrid synthesis of UDP 6-deoxy glucose.	30
Figure 2.9 Representative purifications of <i>Bs</i> YpfP with different tags by Ni-NTA column..	33
Figure 2.10 Gel filtration elution profile of <i>Bs</i> YpfP.....	36
Figure 2.11 <i>In vivo</i> activity of <i>Bs</i> YpfP.	40
Figure 2.12 Multiple sequence alignment of YpfP, MGD synthase and MurG.	42

Figure 2.13 <i>In vivo</i> activity of <i>BsYpfP</i> mutants.....	43
Figure 2.14 Time course studies of <i>BsYpfP</i> -His6 <i>in vitro</i> activity.....	46
Figure 2.15 <i>In vitro</i> glycolipids accumulation curve of <i>BsYpfP</i> -His6.	47
Figure 2.16 Substrate specificity of <i>BsYpfP</i>	49
Figure 2.17 The pH profile of <i>BsYpfP</i> <i>in vitro</i> activity.....	51
Figure 2.18 HPLC chromatogram of hybrid synthesis of UDP-6-deoxy-glucose.....	53
Figure 2.19 Photographs of HTH _{TagF} - <i>BsYpfP</i> -His6 crystals.	60
Figure 3.1 Cartoon representations of bacterial secretion systems.....	79
Figure 3.2 Carton representation of SPATE, TAA, AIDA-I subfamily of AT in <i>E. Coli</i>	85
Figure 3.3 A cartoon representation of a classical type V _a autotransporter structure.....	88
Figure 3.4 Proposed mechanisms of biogenesis of AT.....	89
Figure 3.5 Multiple sequence alignment of β domain of Eha proteins.	100
Figure 3.6 SDS-PAGE analysis of purified Eha proteins.	103
Figure 3.7 Elution profile of ion-exchange chromatography of EhaB _{fl}	104
Figure 3.8 Representative pictures of Eha protein crystals.....	107
Figure 3.9 Structure of EhaB _c	110
Figure 3.10 Cut-away view of NalP _c , EhaB _c and Hbp _c	114

Figure 3.11 Behavior of EhaB_c in planar BLM.....	117
Figure 3.12 EhaB_c is a rectifying voltage-gated channel.	118
Figure 4.1 Sequence alignment of N terminal regulatory domains of selected SUS enzymes... 140	
Figure 4.2 A ribbon drawing of the AtSus1 tetramer.	143
Figure 4.3 A packing diagram of the AtSus1 tetramers in the unit cell.....	144
Figure 4.4 A stereo view of the CTD in subunit H in AtSus1 complexed with UDP and fructose.	145
Figure 4.5 Views of the overall fold of AtSus1 and its subunit interfaces.	149
Figure 4.6 Topology diagram of the GT-B glycosyltransferase domain in AtSus1.	151
Figure 4.7 Stereo views of the active site in AtSus1.	157
Figure 4.8 Schematic diagram of potential hydrogen-bonding interactions of bound UDP, LCN and fructose with the protein, from the superposition of the two complexes.	158
Figure 4.9 A diagram of the S_Ni -like reaction scheme (adapted from Lairson <i>et al.</i> [37]).....	161
Figure 4.10 Stereoview of the conserved interactions in the active site of AtSus1 and the retaining glycosyltransferases glycogen synthase from <i>E. coli</i> (EcGS; PDB 3GUH) [58] and UDP-glucosyltransferase OtsA from <i>Mycobacterium tuberculosis</i> (PDB 3C4V) [63]......	165
Figure 4.11 Structural comparison of inverting and retaining GTs.	169
Figure 4.12 Structural comparison of AtSus1 GT-B _N domain in the retaining and inverting positions..	170
Figure 4.13 Juxtaposition of the CTD and EPBD domains with the AtSus1 active site.	173

LIST OF ABBREVIATIONS

AIDA	Adhesin involved in diffuse adhesion
AS	Ammonium sulfate
ASU	Asymmetric unit
AT	Autotransporter
β -EtSH	β -mercaptoethanol
BLM	Bilayer Lipid Membranes
CHAPS	3-[(3-Cholamidopropyl)-dimethylammonio]-1-propane sulfonate
CU	Chaperone-Usher
CV	Column volume
DAEC	Diffusely adherent <i>E. coli</i>
DAG	Diacylglycerol
DGlcDAG	Diglycosyl-DAG
EAEC	Enteroaggregative <i>E. coli</i>
EHEC	Enterohaemorrhagic <i>E. coli</i>
EIEC	Enteroinvasive <i>E. coli</i>
ENOD40	Early nodulin 40

EPEC	Enteropathogenic <i>E. coli</i>
ETEC	Enterotoxigenic <i>E. coli</i>
G3P	Glycerol-3-phosphate
GroP	Glycerolphosphate
GT	Glycosyltransferase
IM	Inner membrane
LAM	Lipoarabinomannan
LAPAO	3-Dodecylamido-N,N'-dimethylpropyl amine oxide
LCN	Lichenan
LM	Lipomannan
LTA	Lipoteichoic acid
MGlcDAG	Monoglucosyl-DAG
MBP	Maltose binding protein
MFP	Membrane fusion protein
MSA	Multi sequence alignment
NCS	Non-crystallographic symmetry
NHF	1,5-anhydro fructose
OM	Outer membrane
OMP	Outer membrane protein
PA	Phosphatidic acid
PBS	Phosphate buffer saline
PC	Phosphatidylcholine

PG	Phosphatidylglycerol
R.M.S	Root mean square
SeMet	Selenomethionyl
STEC	Shiga toxin-producing <i>E. coli</i>
Stx	Shiga-like toxins
SUS	Sucrose synthase
TAA	Trimeric autotransporter
TLC	Thin layer chromatography
TLS	Translation Libration Screw-motion
TM	Transmembrane domain
TXSS	Type X secretion system, where X = I, II, III, IV, V, or VI
UTI	Urinary tract infection
VT	Verocytotoxin
WTA	Wall teichoic acid
Z3-12	n-Dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate
Z3-14	n-Tetradecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate

CHAPTER 1

AN INTRODUCTION TO THE THESIS

1.1 Background and introduction on microbe-host interactions

Microbes were the first forms of life to appear on earth approximately 3–4 billion years ago [1]. Ever since then, they have formed intimate relationships not only with each other, but also with higher organisms. Such interactions led to the evolution of a variety of complex and highly specialized processes between a microbe and a host organism. These interactions can be beneficial for both, as in many symbiotic relationships, or deleterious to the host cell, in case of the pathogenic microbes. Microbe-host interactions have been a major area of research ever since the identification of bacteria as a primary cause of many human diseases. On the other hand, the beneficial commensal relationships, particularly with regards to human health, have also been investigated to provide a better understanding and control of the microflora under normal conditions. Finally, microbiological research has also focused on symbiotic microbe-host relationships to understand how mutual benefit is achieved by reciprocal adaptation of the metabolisms of the host and microbe.

Despite the amazing diversity of microbes, the fundamental mechanisms of the host-microbe interaction follow several similar strategies. For example, the host-microbe recognition usually involves functional co-adaptation of glycans, glycoproteins and glycolipids on both the cell surfaces of the hosts and microbes, while the colonization and infection process involves a broad range of virulence factors including pili, flagella, adhesins, and extracellular membrane degrading enzymes. Moreover, a variety of proteins are released by microbes to modulate the

host cell's metabolic function and defenses. High resolution structures of important proteins involved in these processes will provide the valuable knowledge base for understanding the mechanism of host-microbe interaction and help create a platform for developing better treatments, in the case of microbial infections, or for exploiting the beneficial microbe-host interactions.

1.2 Research foci in this thesis

The structural biology of proteins involved in host-microbe interactions is the focus of this thesis. The underlying rationale for carrying out this research is that the determination of high resolution structures of key proteins in host-microbe interactions can provide a wealth of new insights into the crucial biochemical mechanisms in these relationships and would very likely provide good targets for modulating important physiological interactions *in vivo*. In this work, I studied the structure and function of two glycosyltransferases (sucrose synthase-1 from *Arabidopsis thaliana* and UDP-glucose:1,2-diacylglycerol-3- β -D-glucosyltransferase YpfP from *Bacillus subtilis*) and of three adhesin proteins from the pathogenic *Escherichia coli* O157: H7. Each of these proteins was expressed, purified, and studied biochemically, and their X-ray crystal structures were determined, when X-ray quality crystals were produced. The efforts to grow crystals of these membrane-active proteins also led to the extension of the fusion module method to facilitate protein crystallization.

Chapter 2 of the thesis focuses on YpfP, a membrane-bound UDP-glucose: 1,2-diacylglycerol-3- β -D- glucosyltransferase from the Gram-positive bacterium *Bacillus subtilis*. The enzyme YpfP is a member of the GT-B family glycosyltransferases and can transfer up to four glucose moieties to diacylglycerol (DAG), an important component of the bacterial cell membrane. Detailed biochemical characterization of YpfP was carried out, which provided new

information on its enzymatic activity *in vitro* and *in vivo*. To accomplish this work, a novel giant liposome activity assay for YpfP and an on-column hybrid synthesis scheme for UDP-6-deoxy glucose were developed. Although extensive crystallization trials produced no crystals of suitable quality for X-ray diffraction analysis, valuable experience was gained during the process of crystallization and provided new developments for crystallization by the fusion module method.

Chapter 3 of the thesis focuses on the structures of Eha (EHEC autotransporter [2]) family of adhesins from pathogenic *Escherichia coli* O157: H7 strain EDL933. EhaA, B, C, and D all belong to the AIDA-I (Adhesin Involved in Diffuse Adherence) family of autotransporters that is involved in biofilm formation and host cell binding. A series of structures were determined during this study and provide new insights on the structure-function relationships in this protein family. In addition, these structures may serve as a platform for developing vaccines targeted to structurally related autotransporters and for novel systems of protein autodisplay.

Chapter 4 switches the focus back to glycosyltransferase structure and functions with a structural study of sucrose synthase, a crucial metabolic enzyme from the plant *Arabidopsis thaliana*. The SUS1 isoform from *A. thaliana* (AtSus1) is a key enzyme in the regulation of the flux and metabolic utilization of the plant primary metabolite sucrose between the source and the sink organs. It is also suggested that AtSus1 plays a role in the host response to bacterial and fungal symbiosis occurring during nitrogen fixation. AtSus1 was crystallized as a complex with different substrates, and two X-ray crystal structures allowed the building of detailed atomic models. The analysis of these structures in comparison with other GT-B family proteins has provided many new insights into the mechanism of the retaining and inverting glycosyl transfer

reactions catalyzed by this family of proteins, as well as the mechanism for its physiological regulation.

In chapter 5 of the thesis, I will revisit some of the important results and conclusions from chapters 2 to chapter 4 and outline future research directions for each project. Overall, the structure determinations for the *BsYpfP* and *Eha* proteins are still not all completed and further experimentation is still needed to get better crystals. At the same time, detailed characterization of their biochemical and biophysical properties, including activity, stability, and homogeneity, still need to be carried out. The results of these studies will undoubtedly facilitate better understanding of the *BsYpfP* and *Eha* proteins' functions and the crystallization process. Although the crystal structure of *AtSus1* has been determined to a reasonable resolution, detailed understanding of how the regulatory domains modulate the enzyme activity is still lacking. Nonetheless, the experimental results presented in this thesis have contributed towards a better understanding of the structure-function relationships in three different systems involved in microbial-host interactions.

REFERENCES

REFERENCES

- 1 Schopf, J. W. (1994) Disparate rates, differing fates: tempo and mode of evolution changed from the Precambrian to the Phanerozoic. *Proc Natl Acad Sci U S A* **91**, 6735-6742
- 2 Wells, T. J., Sherlock, O., Rivas, L., Mahajan, A., Beatson, S. A., Torpdahl, M., Webb, R. I., Allsopp, L. P., Gobius, K. S., Gally, D. L. and Schembri, M. A. (2008) EhaA is a novel autotransporter protein of enterohemorrhagic *Escherichia coli* O157 : H7 that contributes to adhesion and biofilm formation. *Environmental Microbiology* **10**, 589-604

CHAPTER 2

STRUCTURAL AND FUNCTIONAL STUDIES OF GLYCOSYLTRANSFERASE

YpfP FROM *BACILLUS SUBTILIS*

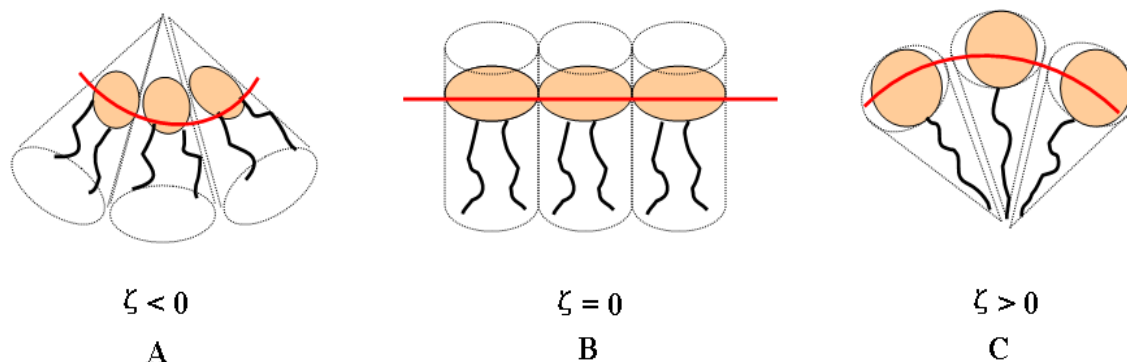
2.1 Background and introduction

2.1.1 Biological membrane structure and membrane curvature

Biological membranes are fluid, continuous, and dynamic bilayer structures that encapsulate cells and organelles. They are formed from amphipathic lipids, which consist of a hydrophobic and a hydrophilic portion. The physical bases of the spontaneous formation of biological membranes are the propensity of the hydrophobic moieties to undergo entropy-driven self-association and the tendency of the hydrophilic moieties to interact with aqueous environments and with each other. It is widely recognized that besides their barrier function biological membranes also play essential roles during cell division, biological reproduction, adaptation to abiotic stresses, and intracellular membrane trafficking. These events can be regulated by altering membrane curvature or composition or by modifying, degrading, and re-synthesizing of membrane lipids during budding, tabulation, fission and fusion processes [1].

Membrane lipids also allow particular proteins within membranes to aggregate, and others to disperse. Moreover, membrane lipids can act as the first and second messengers in signal transduction and molecular recognition processes. Last but not least, some lipids function to define membrane domains, which recruit proteins from the cytosol that subsequently organize secondary signaling or effector complexes. New evidence has suggested that pathological processes, such as viral infection, have been unexpectedly connected to specific membrane lipids and their biophysical properties [2].

Membrane curvature is the geometrical measurement or characterization of the curving nature of membranes. In biological membranes, the curvature is predominately determined by the natural spontaneous curvature (ζ) exhibited by the lipid components of the membrane. The concept of ζ is introduced to define the shape of a lipid monolayer made up of single lipid species [3]. The ζ of a specific lipid species is determined by the length and saturation of its fatty acids [4], which in turn determine the space requirement of the tail in comparison to its head group, as well as environmental conditions such as pH and/or ion concentrations [5]. When the space occupied by the polar headgroups and apolar tails are similar, the molecule is considered to be cylindrical with a spontaneous curvature $\zeta=0$ (Figure 2.1B). Lipids commonly seen in the cell membranes such as phosphatidylcholines (PC) and phosphatidylglycerol (PG), all belong to this group because of their tendency to form a lamellar phase (i.e., bilayer) in an aqueous environment. The lipids with a cone shape, however, are designated with a negative spontaneous curvature ($\zeta < 0$), which is caused by the different space requirements of the small polar headgroups and the large tails. The accumulation of these lipid molecules gives rise to inverted micelles (Figure 2.1A). Lipids with negative spontaneous curvature, represented by diacylglycerol (DAG) and cholesterol, are not as abundant as the bilayer forming lipids in the nature, but still play important roles in the biological membrane system. In contrast, the inverted cone shape seen in lipids with large polar heads and small tails induces a positive spontaneous curvature ($\zeta > 0$) that gives rise to micelles (Figure 2.1C). Lysophospholipids are natural examples of micelle-forming lipids while detergents are also molecules with positive spontaneous curvatures that are designed to form micelles in solution.



For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation

Figure 2.1 Cartoon representations of membrane structures induced by lipids with different spontaneous membrane curvatures. (A): inverted micelle structure formed by lipids with negative spontaneous membrane curvature ($\zeta < 0$); (B) one leaflet of the lipid bilayer formed by lipids with a spontaneous membrane curvature of zero ($\zeta = 0$); (C) micelle structure formed by lipid with a positive spontaneous membrane curvature ($\zeta > 0$).

2.1.2 Glycolipids in bacteria

A glycolipid refers to any compound that contains one or more monosaccharide moieties bound by a glycosidic linkage to a hydrophobic moiety, such as a (DAG), a dolichyl phosphate, a sphingoid, a ceramide (*N*-acylsphingoid), a sterol, or even free fatty acid. DAG-based glycolipids are major membrane constituents in both photosynthetic and non-photosynthetic bacteria [6-8], while dolichyl phosphate-based glycolipids are critical intermediates and lipid carriers for bacterial murein biosynthesis [9, 10]. Glycosphingolipids have been recently identified in several species of the genus *Sphingomonas* [11, 12] as examples of less common bacterial glycolipids. Similarly, steryl glucosides have been found in some bacterial pathogens, such as the gram-negative *Helicobacter pylori* [13-16], which is the causative agent of gastritis and gastric ulcers, and *Borrelia burgdorferi*, which is responsible for Lyme disease [17, 18]. Biosurfactants, such as rhamnolipids [19, 20] secreted by *Pseudomonas aeruginosa* and trehalose lipids [21-23] secreted by *Rhodococcus sp.*, are unusual glycolipids where the sugar moiety is directly linked to an acyl or a hydroxyl hydrocarbon chain without the intermediate of the glycerol backbone.

Partly because of the diverse environments and sugar sources to which bacteria have access, the diversity of the sugar moiety has allowed glycolipids to play important roles in almost every aspect of bacterial life. In brief, glycolipids serve as cell markers, mediate cell proliferation, adhesion, and host colonization, and participate in signaling pathways involving quorum sensing and swarming migration [24-29]. Glycolipids are also required for the maximal efficiency of photosynthesis [6] and/or biofilm formation [25]. Glycolipids sometimes function as membrane anchors in the biosynthesis of the extracellular cell wall superstructure, as in the cases of peptidylglycans, teichoic acids [30-32], lipomannan (LM), and lipoarabinomannan (LAM) glycolipids [33]. Due to the similar spontaneous membrane curvature, glycolipids have also been

demonstrated to be a functional replacement of the phospholipids under phosphate limiting conditions [34].

2.1.3 Diacylglycerol-lipoteichoic acid biosynthesis network in *Bacillus subtilis*

Bacillus subtilis is a well-studied model microorganism belonging to the *Bacilli* class of low G+C Gram-positive bacteria. DAG is the most abundant neutral lipid (more than 90%) in *B. subtilis* [35]; it is also one of the few natural lipids with a quite negative spontaneous membrane curvature ($\zeta < 0$) [36] as its polar head group is very small in comparison to that of other lipids. In terms of cell wall structure, *B. subtilis* has a thick cell wall composed of proteins and peptidoglycans that is typical for the Gram-positive bacteria, in contrast to the outer cell membrane and the relatively thin cell wall in the Gram-negative bacteria. In addition, two unique types of macromolecules, namely lipoteichoic acids (LTA) and wall teichoic acids (WTA), are found only in cell walls of the Gram-positive bacteria from the phylum *Firmicutes*, to which *B. subtilis* belongs.

The generic structure of DAG (left panel of Figure 2.2) consists of two fatty acids tails (shaded in orange) linked to the *sn*-1 and *sn*-2 positions of the glycerol backbone through ester bonds and a free hydroxyl group at the *sn*-3 position. R and R' represent the alkyl chain in the fatty acid and reflect the composition of the parent phospholipids from which DAG was derived. LTA and WTA are both anionic polymers threading vertically through the peptidoglycan layer that surrounds the plasma membrane. However, their biosynthesis proceeds through different pathways. The biosynthesis of both LTA and WTA includes four distinct steps, namely the formation of the glycolipid anchor, the elongation of the anionic polymer backbone onto the

anchor, the decoration of the polymer with alaninyl or glucosyl groups, and the transport of the lipid molecule across the bilayer membrane.

LTA has a simpler chemical structure (Figure 2.2, right panel) that typically consists of the glycolipid glucosyl (β 1-6) glucosyl (β 1-3) DAG (Glc2-DAG) (shaded in pink) embedded in the membrane as the anchor and a polyglycerolphosphate (Gro-P) chain that is linked to it via a diglucose linker. The biosynthesis of LTA in *B. subtilis* (Figure 2.3) starts at the branch point of glycolysis where glucose-6-phosphate is converted to UDP-glucose via two key enzymatic conversions catalyzed by phosphoglucomutase and UDP-glucose pyrophosphorylase. The processive glycosyltransferase YpfP then transfers two glucose moieties from UDP-Glc to DAG in the inner leaflet of the membrane [37-39]. In the next step, a permease-like integral membrane protein, which is unknown in *B. subtilis* but identified as LtaA in *Staphylococcus aureus*, flips the glycolipid anchor across the membrane to the outer leaflet. The ProG backbone is primed by YvgJ and polymerized by one of the three recently characterized ProG polymerases, LtaS, YfnI, or YqgS, from *B. subtilis*. These four proteins are all integral membrane proteins with five transmembrane helices and with a cytosolic C-terminal catalytic domain. However, the ProG backbone is predicted to be formed by the headgroups of the PG. The final D-alanylation of the LTA is catalyzed by proteins from the *dlt* operon.

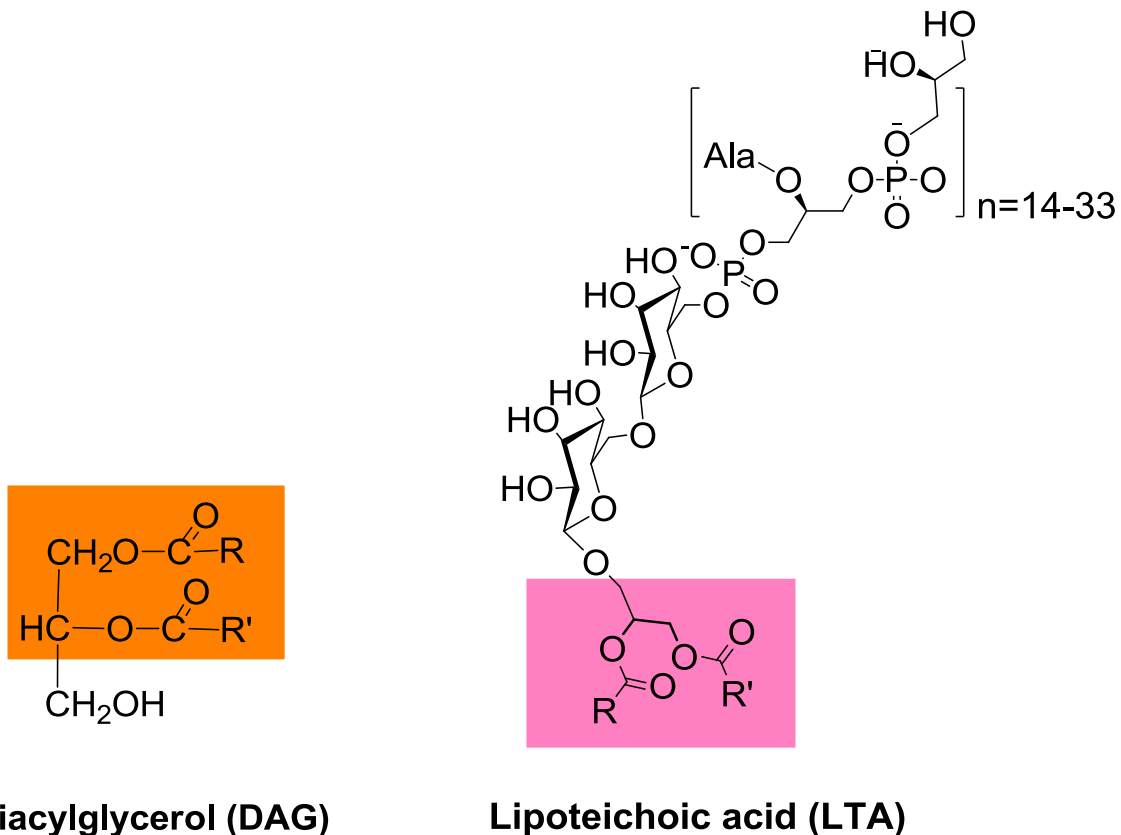


Figure 2.2 Structures of diacylglycerol and lipoteichoic acid

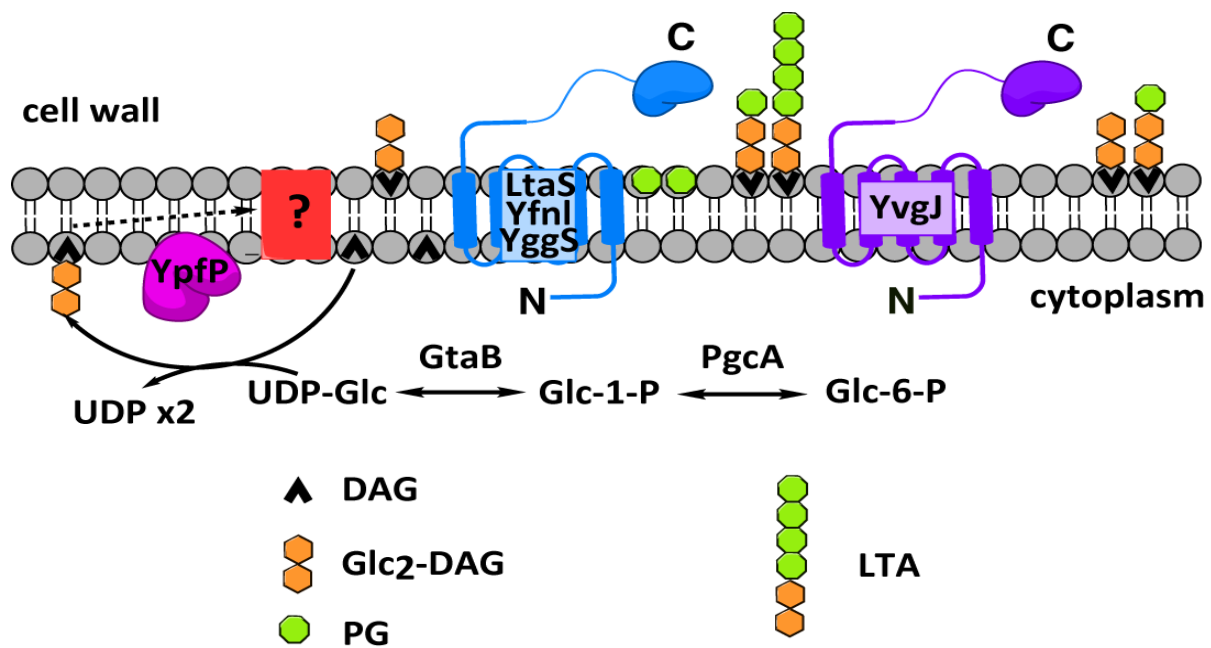


Figure 2.3 Diacylglycerol-lipoteichoic acid pathway enzymes in *B. subtilis*

WTA has a slightly more complex structure where ribitol phosphate (RboP), glycerolphosphate (GroP) or more complex sugar-containing polymers are polymerized on an undecaprenylphosphate precursor within the cytoplasm and transported across the membrane before being covalently linked to the peptidoglycan [40]. In term of the biosynthesis of WTA, the glycolipid anchor is synthesized through sequential transfer of the UMP-activated GlcNAc phosphate and the UDP-activated ManNAc onto an undecaprenylphosphate by TagO and TagA. The CDP-activated GroP is first transferred onto the glycolipid anchor by a primase TagB and then polymerized by a GroP polymerase TagF before an RboP polymerase TagL adds more CDP-activated RboP monomer onto the growing chain. Finally the WTA precursor is transported across the membrane by an ABC-type transporter complex TagG/H before covalently attached onto the peptidoglycan chain.

2.1.4 Diacylglycerol metabolism and membrane curvature in *B. subtilis*

The significant difference in space requirements of the hydroxyl head group and the two acyl tails is the primary reason for the negative curvature caused by DAG. The ratio of DAG to other bilayer forming lipids plays an important role in regulation of cell functions, while the negative curvature in a local region of the membrane due to an enrichment of DAG may have deleterious effects on the cell physiology under stress conditions [41]. On the other hand, the generation of a transitory negative curvature by DAG formation is found to function as a signal for protein recruitment to the membrane [3]. The conversion of DAG into other lipid molecules with less negative membrane curvature will in turn relieve the membrane tension and terminate the signal cascade. The free hydroxyl group, which is subjected to diverse modification such as acylation, phosphorylation, and glycosylation, renders DAG an important regulator of membrane curvature through a complicated lipid metabolism network (Figure 2.4A) [42].

In *B. subtilis*, membrane lipids are synthesized from the common precursor phosphatidic acid (PA), which is in turn converted from fatty acid by lysophosphatidic acid (LPA) acyltransferase. In the case of PG, PA is converted to CDP-diacylglycerol (CDP-DAG) by adding a CDP from CTP by CDP-DAG synthase. Condensation of CDP-DAG with glycerol-3-phosphate (G3P), followed by removal of the phosphate, leads to PG, the only essential complex lipid and the major bilayer-forming lipid in *B. subtilis*. On the other hand, glycolipids are created by dephosphorylation of PA to DAG, which is then glucosylated by YpfP using UDP-glucose to form the monoglycosyl-DAG (MGlcDAG) and diglycosyl-DAG (DGlcDAG). DGlcDAG can be subsequently used in the biosynthesis of LTA with PG as the other substrate. The free DAG molecules released from PG then reenter the network as the substrate for YpfP. By balancing the biosynthesis of the bilayer-forming lipid PG and DGlcDAG and the non-bilayer forming lipid DAG and MGlcDAG, the membrane curvature as well as related functions can be tightly regulated by this network (Figure 2.4B).

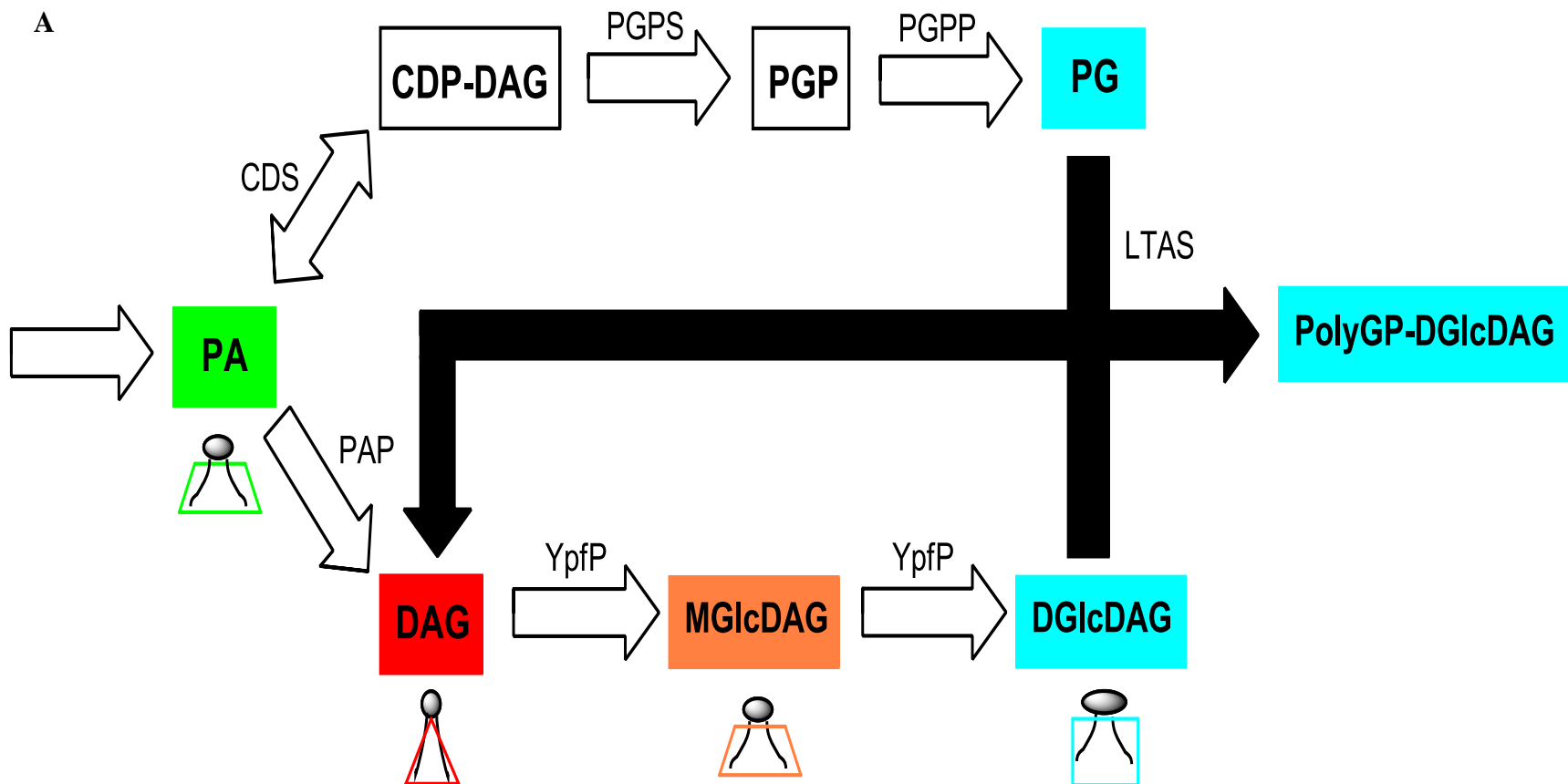
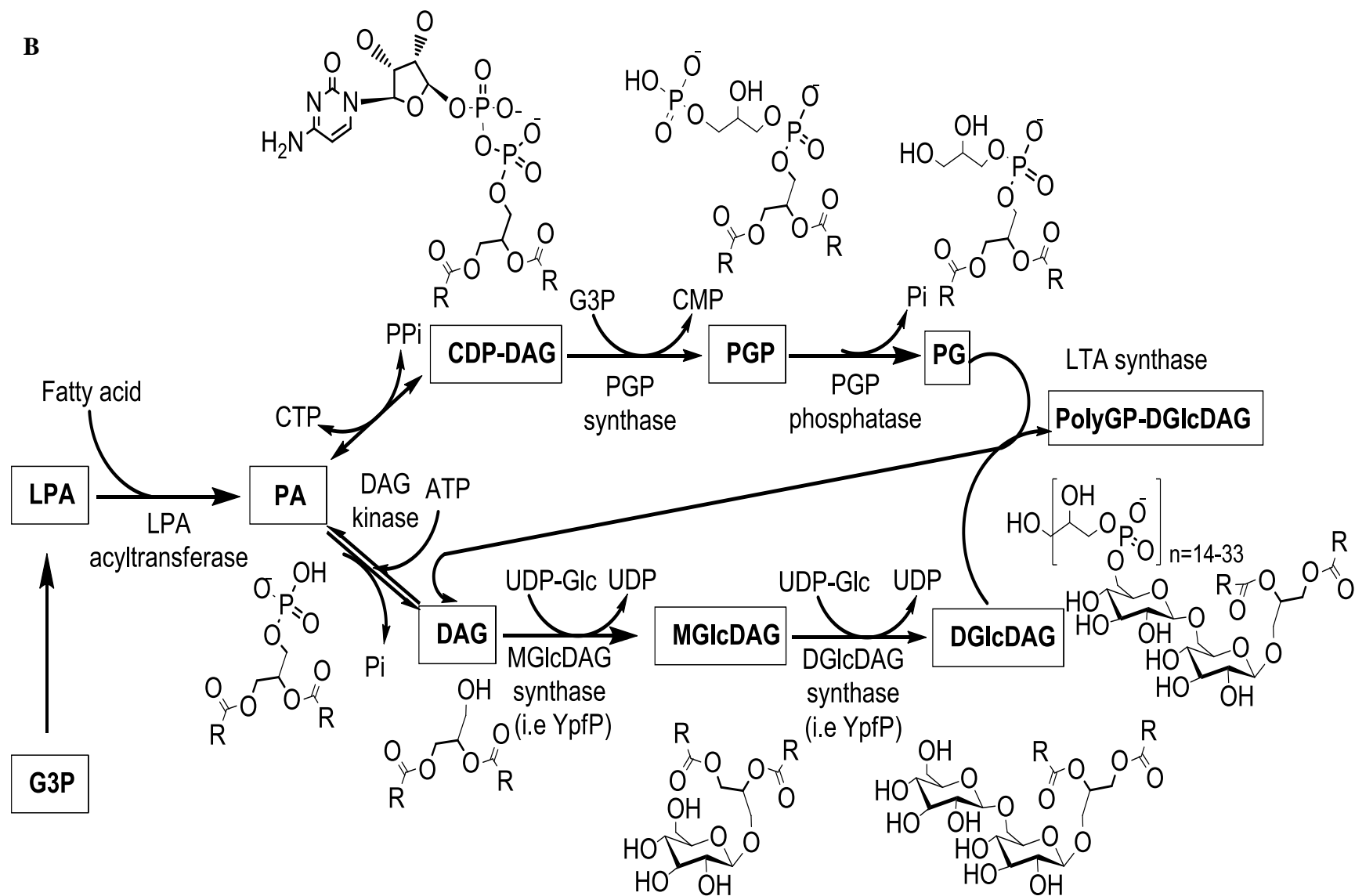


Figure 2.4 DAG metabolic pathways. (A) Interconversion of membrane curvatures in *B. subtilis* during DAG metabolism. (B) DAG metabolism network in *B. subtilis*.

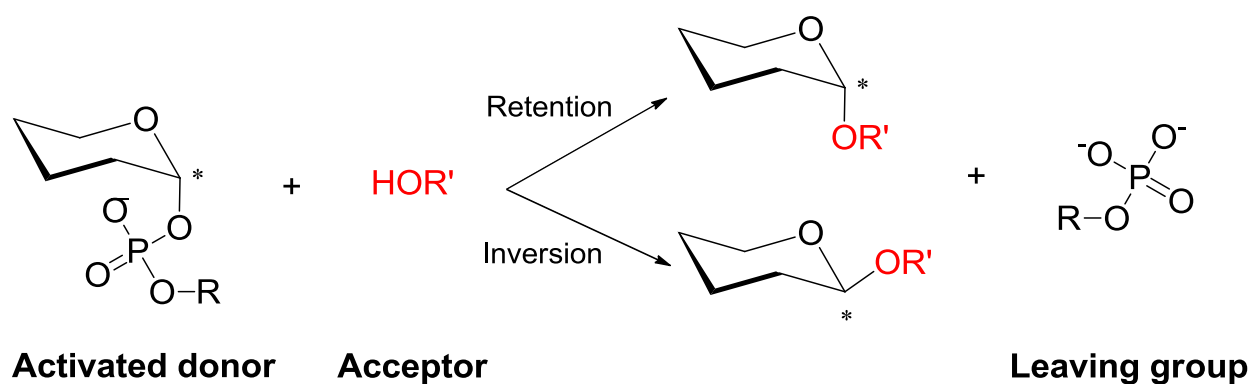
Figure 2.4 continued.

B



2.1.5 Glycosyltransferase structural biology

Glycosyltransferases constitute a large family of enzymes that catalyze the transfer of the sugar moiety from an activated donor to the sugar acceptor. The functional diversity of glycosyltransferases arises from four different aspects in this simple reaction (Figure 2.5): 1) the sugar moiety itself, the leaving group of the sugar donor; 2) the relative configuration of the anomeric carbon in the sugar donor; 3) the relative configuration of the anomeric carbon in the product (i.e., inversion or retention); and 4) the various biopolymers (polysaccharides, proteins, nucleic acids, and lipids) as well as numerous small molecules (saccharides, antibiotics, pigments) that can act as the sugar acceptor. As a result of such a high level of functional diversity, glycosyltransferases have been grouped into more than 90 families based on their highly diverse primary sequences (designated GT_n, where n = 1, 2, ...) (<http://www.CAZy.org>) [43]. However, their tertiary structural folds are conserved. The reported glycosyltransferases structures have revealed only two general structural folds, namely GT-A and GT-B fold, for the nucleotide sugar-dependent enzymes.



R = Nucleotide monophosphate, nucleotide, lipid etc.

R' = Sugar, lipid, protein, antibiotic, nucleic acid etc.

Figure 2.5 Diversity of glycosyltransfer reaction. The diversity of the glycosyltransfer reaction is presented with the generic structure of the substrates. In this reaction, the sugar donor is activated by a phospho-nucleotide attached at the anomeric C1 position where R can be nucleotide, nucleoside monophosphate, or lipid phosphate (R=lipid). The acceptor is shown in red as R'OH, where R' represents a sugar, a lipid, a protein, an antibiotic, a nucleic acid, etc. The attacking of the acceptor in different mechanisms results in two possible conformations in the product: the inversion or retention of the anomeric configuration of the donor.

The GT-A fold (Figure 2.6 (A)) is described as a single Rossmann-like domain consisting of an open twisted β -strains surrounded by α -helices on both sides [44]. The primary GT-A fold can be further divided into two sub domains. The *N*-terminal half, shown in marine blue, has four parallel β -strains while the *C*-terminal half, shown in green, has three anti-parallel β -strains, and the two sub domains abut each other, leading to the formation of a continuous central β -strain. The sugar donor binds to the N sub domain, while the sugar acceptor binds to the pocket created by the C sub domain. In most of the GT-A glycosyltransferases, a conserved divalent metal cation, shown as a purple sphere (i.e., Mg^{2+} or Mn^{2+}), is found to be coordinated by two carboxylates from a signature Asp-X-Asp (DXD) motif to stabilize the leaving diphosphate nucleotide.

Compared to the GT-A enzymes, GT-B proteins (Figure 2.6 (B)) are ion independent and have two distinct and complete Rossmann folds (six parallel β -strains surrounded by three α -helices on each sides), namely N and C domains. The two domains are connected by a double stranded hinge: an unstructured linker from N to C domain and another linker between a tandem two-helix-tail which spans the two domains from C to N. The active site of the GT-B enzymes lies within the large cleft between the two domains. Various loop regions between consecutive β -sheets and α helices clustered inside this cleft determine the substrate specificity. Contrary to GT-A folds, the reported structures of the GT-B members reveal that the sugar donor and the acceptor bind to C domain and N domain, respectively [45, 46].

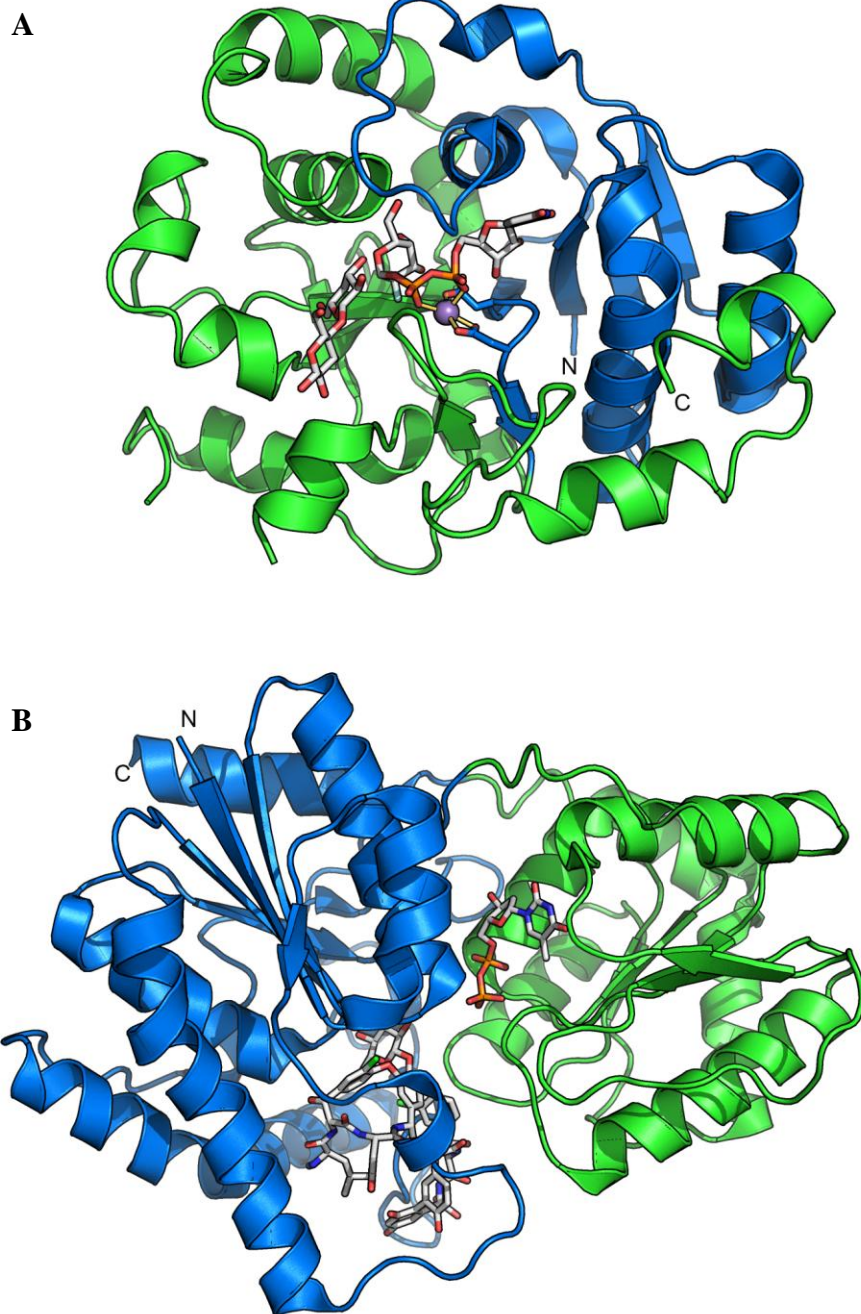


Figure 2.6 Typical structure folds of GT-A and GT-B glycosyltransferases. (A) GT-A fold structure of LgtC, PDB 1GA8. (B) GT-B fold structure of GtfD, PDB 1RRV. Marine blue represents N domain and green represents C domain in both structures. Ligands are shown as gray sticks, metal ion is shown as a purple sphere.

2.1.6 Glycosyltransferase YpfP from *B. subtilis*

YpfP, also known as UgtP, is a glycosyltransferase in *B. subtilis*. The *ypfP* gene was first identified and cloned based on homology to the catalytic domain of the cucumber transmembrane protein MGD synthase from *Cucumis sativus* [37, 47]. The *ypfP* gene is not essential for the bacteria growth, but its deletion null mutant displays defects in swarming motility and biofilm formation. A 90% reduction of LTA contents and the resulting formation of abnormal LTAs, where the ProG was directly linked to DAG, were also observed in the null mutants of *Staphylococcus aureus* [48].

YpfP, a UDP- α -D-glucose: 1, 2-diacylglycerol-3- β -D-glucosyltransferase (EC 2.4.1.157), is classified into family GT28 in the CAZy database, along with MurG (EC 2.4.1.227) and MGD synthase (EC.2.4.1.46). Members of this family all have a GT-B fold structure and catalyze the glycosyltransfer reaction through the inverting mechanism. Due to the hydrophobic nature of the DAG, YpfP is a membrane associated protein potentially interacting with the negatively charged membrane surface via a positively charged helix. Compared to MurG and MGD synthase, YpfP possesses a unique processive character in that it can successively transfer up to four glucosyl moieties to DAG. Among the possible products, the glucosyl (β 1-6) glucosyl (β 1-6) DAG is the major product and the glycolipid anchor for LTAs. The available DAG source in *B. subtilis* is primarily from the dephosphorylation of PA and by-products of LTA biosynthesis (Figure 2.4A). On the other hand, the generation of UDP-glucose starts at the branch point of enzymatic conversions of glucose-6-phosphate to UDP-glucose (Figure 2.3). Therefore YpfP functions as a key enzyme connecting the primary metabolism to the biosynthesis of cell wall macromolecules. YpfP is also shown to be localized to the bacterial cell division site in a nutrient-dependent

manner and inhibits assembly of the tubulin-like cell division protein FtsZ [49], suggesting an important role in nutrient sensing and cell division regulation by YpfP.

2.1.7 Crystallization strategies for YpfP

Despite extensive and exhaustive screening with commercially available crystallization kits, no obvious and reproducible crystallizing conditions were identified for the protein YpfP. Therefore, the fusion module strategy was applied to YpfP with bacterial wall teichoic acid polymerase TagF as the fusion module.

2.1.7.1 Fusion protein as crystallization module

Protein crystallization refers to a process in which protein molecules gradually phase out of the solution and assemble into regular, periodic crystalline structures through stable self interactions. This process is the critical and often the rating –determining step in determination of the atomic structure of proteins by X-ray crystallography. These self interactions are not only determined by the intrinsic properties of the target protein, including irregularity and accessibility of protein interacting surfaces as well as conformational homogeneity, but are also significantly influenced by the specific crystallizing condition that promotes conditional changes of the protein. However, quantification and confident prediction of the protein self interactions still poses a great computational challenge. In fact, some pioneering research [50] has shown that a narrow B_2 value (the second virial coefficient) range indeed exists where protein with a B_2 value in this range has a high possibility of forming well ordered crystals. It is also consistent with the experimental observations that B_2 values for the amorphous protein aggregation due to very strong self interactions and the protein solution of high solubility where the self interactions

are very weak are both outside this range. However, experimental determination of the B_2 value is still time- and material-consuming despite recent progress in this area [50]. Therefore, the prediction of crystallization conditions and systematically designing of crystallization experiments for a new target protein are still far from realistic.

Another approach that may increase the possibility of crystallization is to take advantage of a highly crystallizable fusion module to facilitate the crystallization of the unknown target protein. In this method, the number of screening conditions might be significantly reduced due to the fact that the crystallization condition for the fusion protein can be derived based on the crystallization condition of the fusion module. The primary assumption is that the fusion module plays an important role in initiating and facilitating the self interaction of the entire fusion protein. Several successful examples have demonstrated the possibility of this idea [51-53]. However, large protein fusion modules have only been successfully used in the crystallization of small proteins. It is therefore suggested that the proteins that served as modules could form only weak crystal contacts that tend to be easily disrupted by the addition of larger target proteins and a fusion module with more potent crystal contacts would be more capable to impart its crystallization properties to the target [54].

2.1.7.2 Crystallization of YpfP with the TagF fusion module

TagF was chosen because the recently solved structure of TagF (PDB 3L7I) showed that TagF has a fold similar to GT-B enzymes and has a strong positively charged helix-turn-helix (HTH) *N*-terminal motif, through which the tetramer interface is formed. The small HTH motif is relatively independent from the core structure of TagF such that the structure of the HTH motif is unlikely to change with the removal of the interactions with the GT-B core structure of TagF.

The similarity of the GT-B fold in TagF to the putative GT-B fold of YpfP will probably preserve the structure of both HTH and YpfP in the fusion protein while facilitating the crystallization of YpfP at the same time. Therefore, the HTH motif from TagF was a very tempting fusion module for YpfP. The use of the TagF-HTH as a fusion module with YpfP indeed yielded reproducible crystals.

2.2 Materials and Methods

2.2.1 Materials

Chemicals were purchased from Sigma-Aldrich. LAPAO (Anagrade) was purchased from Anatrace. Lipids were purchased from Avanti Polar Lipids. The cDNA library of *Arabidopsis thaliana* was generously provided by Dr. Eric Moellering at Michigan State University and the pLW01 [55] vector was a gift from Dr. Lucy Waskell at University of Michigan. The *Staphylococcus epidermidis* cell culture was provided by Dr. Chris Waters at Michigan State University.

2.2.2 Methods

2.2.2.1 Cloning and construction of YpfP expression plasmid

The 1146 bp ORF of *ypfP* was cloned from the genomic DNA of *B. subtilis* and sub-cloned into the pLW01 vector to generate the pRMG_ypfP_his6 expression plasmid. The PCR products were purified with the QIAquick PCR purification kit (QIAGEN) and ligated into the pGEM-T-easy vector (Promega). Ligation products were transformed into *E. coli* DH5 α for selection. Plasmids containing the *ypfP* ORF were isolated and digested with restriction enzymes NcoI and SalI (New England Biolabs) and subsequently purified with the QIAquick gel

extraction kit (QIAGEN). The *ypfP* gene fragment was then ligated by T4 DNA ligase (Invitrogen) into the pLW01 vector fragment digested with NcoI and XhoI (a SalI isoschizomer). Ligation products were transformed into *E. coli* DH5 α competent cells. Positive clones were picked, and plasmid DNAs were isolated. The sequence-verified plasmid is designated as pRMG_ypfP_his6.

The genomic DNA of *S. epidermidis* was purified from 2 ml of stationary phase cell culture using the Wizard® Genomic DNA Purification Kit (Promega) supplemented with 1 mg lysostaphin (Sigma, L7386). A 108 bp DNA fragment corresponding to the *tagF* ORF 948-1052 was amplified and subcloned into pRMG_ypfP_his6 by using the NcoI site. Site-directed mutagenesis was carried out to change the second methionine residue introduced by the second NcoI site to an aspartate residue. The sequence-verified plasmid is designated as pRMG_tagF_HLH_ypfP_his6.

2.2.2.2 Expression and purification of the recombinant YpfP protein

The pRMG_ypfP_his6 was transformed into the *E. coli* C41 (DE3) strain. *E. coli* were grown at 37 °C with shaking at 200 rpm until the O.D₆₀₀ reached 1.0. Recombinant protein expression of YpfP was induced by adding 0.1 mM IPTG to the cell culture and the cultures were grown at 23 °C for an additional 16 h. Cells were harvested by centrifugation and stored at -70 °C. The cell pellet was thawed and resuspended in buffer A (30 mM Tris-HCl, 300 mM NaCl, 10 mM β -mercaptoethanol, 0.1 mM EDTA, pH 8.0); the cells were broken by two passes through Emusilflex-C3 at 20,000 p.s.i. of air pressure. Cell debris was discarded after centrifugation at 12,000 \times g for 20 min at 4 °C, and the supernatant with the cell membranes were centrifuged at 170,000 \times g for 1 h at 4 °C. The membrane pellet was resuspended in buffer

A and homogenized before the BCA protein assay. Then detergent LAPAO was added to 0.7 % (w/v) to a protein concentration of 10 mg/mL. The solution was mixed and incubated for 20 min at 4 °C and any insoluble material was discarded after the second round of ultracentrifugation at $170,000 \times g$ for 1 h at 4 °C. The supernatant containing YpfP-detergent complexes was immediately loaded onto a pre-equilibrated gravity flow column containing 20 mL Ni-NTA agarose slurry. The column was washed extensively with buffer B (30 mM Tris-HCl, 300 mM NaCl, 10 mM β -mercaptoethanol, 25 mM imidazole, LAPAO 0.1%, pH 8.0) to remove any non-specific binding proteins until the $A_{280\text{nm}}$ reading was below 0.08. YpfP was then eluted by Buffer C (30 mM Tris-HCl, 100 mM NaCl, 10 mM β -mercaptoethanol, 200 mM imidazole, LAPAO 0.1%, pH 8.0) and concentrated to 0.2 mg/mL by an Amicon ultracentrifugal filter with a molecular weight cutoff of 50 kDa (Millipore). Purified YpfP was distributed into 0.5 mL aliquots, flash frozen by liquid nitrogen, and stored at $-70\text{ }^{\circ}\text{C}$ freezer.

2.2.2.3 Giant liposome in vitro activity assay

The DAG-containing giant liposomes were prepared according to Hoeger et al. [56]. Ultra-low melting temperature agarose was first dissolved in distilled water to 1% (w/v) and heated in a microwave oven for 20 s. Then 60 μL of the agarose solution was dispensed onto a glass cover slide (\varnothing 15 mm), spread evenly, and dried on a heating block at $50\text{ }^{\circ}\text{C}$ until a thin film of agarose formed. The slide was then placed into a plastic vial. Thirty μL of lipid in chloroform solution was deposited onto the cover slide and dried under the air for 30 ~ 40 min. At last, 800 μL of PBS was added into the vial and the giant liposome was formed spontaneously within an h. Figure 2.7 shows a representative giant liposome made by this method.

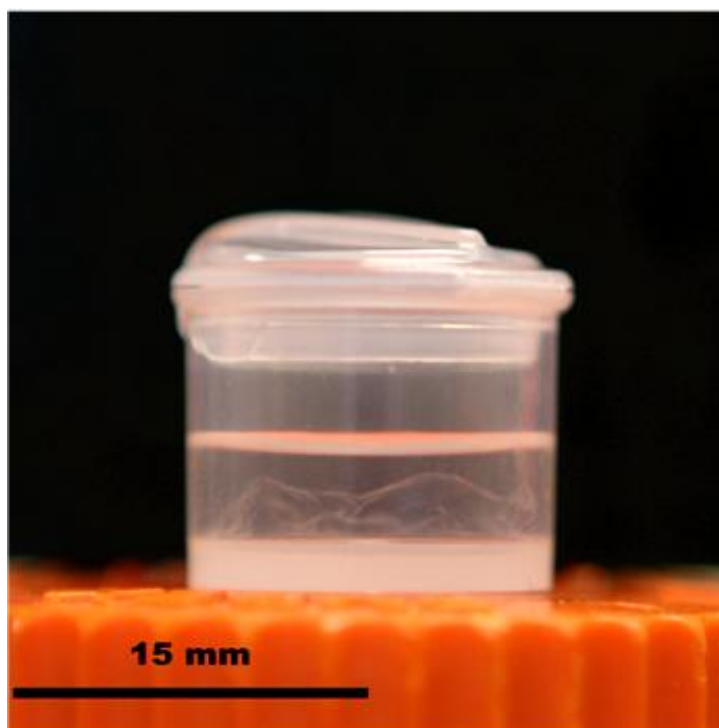


Figure 2.7 A representation of giant liposome.

To assay the *in vitro* glycosyltransferase activity of YpfP, 1mM UDP-glucose and 40 μ L 0.2 mg/mL YpfP were added into the vial with giant liposome to start the reaction. The reaction was stopped for lipid analysis by thin layer chromatography (TLC) after incubation at 23 $^{\circ}$ C for a certain period of time.

2.2.2.4 Hybrid synthesis of UDP 6-deoxy glucose

The reaction scheme for synthesizing UDP-6-deoxy glucose is shown in Figure 2.8. On-column conversion of the UDP-glucose into UDP-4-keto-6-deoxy-glucose by a dehydrotase is carried out in PBS buffer supplied with 2 mM NAD^{+} as the cofactor. The 4-keto group of the product is further chemically reduced with 1 mM NaBH_4 [57] to form UDP 6-deoxy glucose (i.e., UDP-quinovose) and UDP-6-deoxy-galactose (i.e., UDP-fucose), respectively. Products of the hybrid synthesis were analyzed by HPLC with a Carbopac PA1 anion-exchange column (250 X 4.0 mm, Dionex Corp., Sunnyvale, CA) according to the method described by Oka *et al.* [58]. After sample injection, the column was eluted with solvent A (20 mM K_2HPO_4 - KHPO_4 , pH 7.5) at a flow rate of 0.7 mL/min for 5 min, and analyzed isocratically with solvent B (200 mM K_2HPO_4 - KHPO_4 , pH 7.5) at a flow rate of 0.7 ml/min for 30 min. UDP-sugars were detected by UV absorbance at 260 nm.

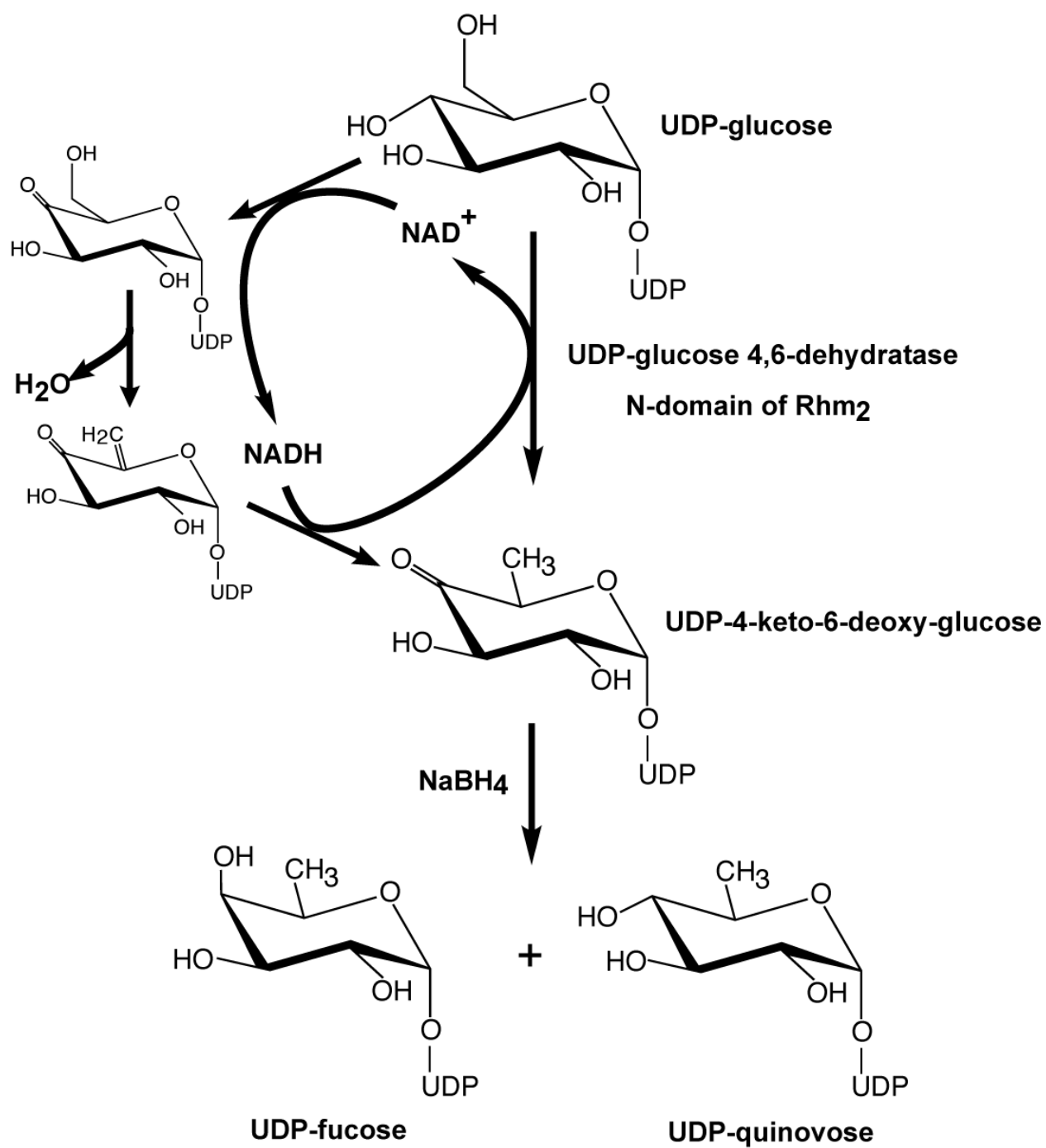


Figure 2.8 Hybrid synthesis of UDP 6-deoxy glucose.

2.3 Results

2.3.1 Choice of *E. coli* strains for the heterologous expression of *BsYpfP*

Previous studies showed half of the recombinant *BsYpfP*-His6 protein forms inclusion bodies when expressed in *E. coli* BL21 (DE3) strain. In order to increase the yield of soluble or membrane associated fraction, the expression temperature, the expression strain, and the addition of the affinity tag were optimized. The properties of *BsYpfP* with different tags used in this study are summarized in Table 2.1 and representative purifications are shown in Figure 2.8. *E. coli* C41 (DE3) strain, which is an uncharacterized mutant of the BL21 strain that is generally used for the expression of membrane proteins, turned out to be the best expression host for the heterologous expression of *BsYpfP*. A low expression temperature (23 °C) is another important parameter to increase the yield by reducing the inclusion body formation inside the cell. Typically, 10 ~ 20 mg protein per liter of culture media can be obtained for various constructs. Among all the fusion proteins, the fusion protein of *BsYpfP* with a maltose binding protein (MBP) at the *N*-terminal was designed to increase the solubility of the fusion protein based on the rationale that a large affinity tag, MBP, is able to mask the hydrophobic region of the *BsYpfP*. Unexpectedly, this fusion protein as well as the fusion protein with a *C*-terminal His6 resulted in higher percentage of protein in the membrane fraction instead of the soluble fraction. The *BsYpfP* protein with a tandem FLAG (*N*-DYKDDDDK-*C*, 1012 Da) and a His6 tag at the *C*-terminus was chosen for the initial crystallization experiments because of its relatively high percentage of protein in the soluble fraction.

Table 2.1 *Bs*YpfP constructs with different tags used in this study

Constructs for heterologous expression	MW kDa	pI	<i>In vivo</i> activity	Solubilizing detergent	Gel filtration	<i>In vitro</i> activity	Crystallization Hits
<i>Bs</i>YpfP-Flag-His6	45.7	7.27	active	Z3-12	aggregates	N/D	n/a
MBP-<i>Bs</i>YpfP-His6	84.9	6.09	active	CHAPS/LAPAO	oligomeric	N/D	n/a
<i>Bs</i>YpfP-His6	44.7	8.36	active	LAPAO	monomeric	active	n/a
HTH_{TagF}-<i>Bs</i>YpfP-His6	49.1	9.06	N/D ¹	Z3-14	N/D	N/D	Yes

1. N/D: Not determined

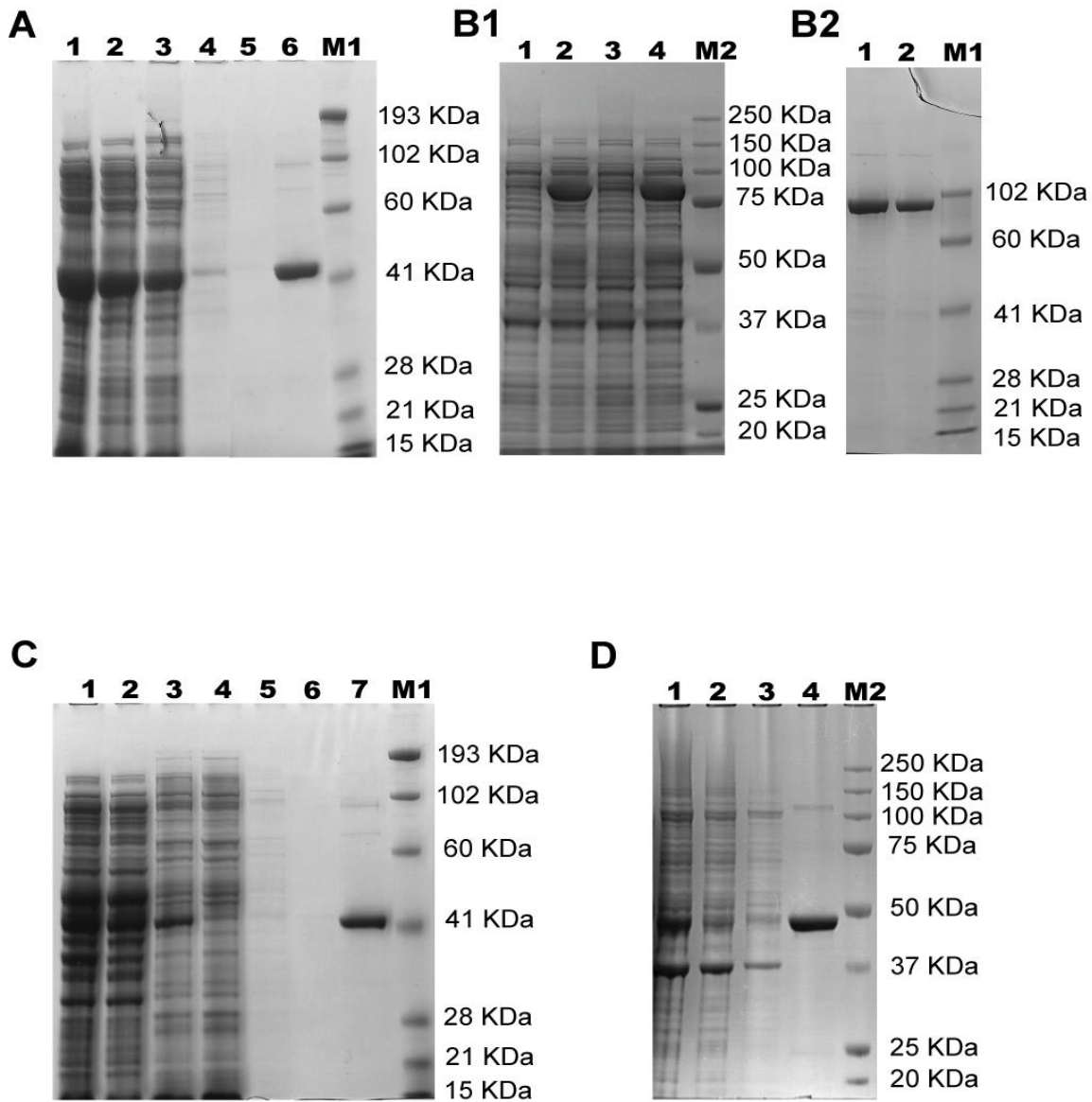


Figure 2.9 Representative purifications of *BsYpfP* with different tags by Ni-NTA column. (A) Purification of *BsYpfP*-Flag-His6. Lane 1: low spin supernatant (LSS); lane 2: high spin supernatant (HSS); lane 3: flow through; lane 4: wash; lane 5: wash 2; lane 6: elution. (B1) Expression of MBP-*BsYpfP*-His6. Lanes 1&3: uninduced cell culture; lanes 2&4 induced cell culture. (B2) Purified MBP-*BsYpfP*-His6 from high spin pellet supernatant (HSPS). Lanes 1&2. (C) Purification of *BsYpfP*-His6. Lane 1: LSS; lane 2: HSPS; lane 3: flow through; lane 4: wash; lane 5: wash; lane 6: wash; lane 7: elution. (D) Purification of HTH_{TagF}-*BsYpfP*-His6. Lane 1: HSPS; lane 2: flow through; lane 3: wash; lane 4: elution. M stands for protein Mw standards.

2.3.2 Choice of detergents to suppress the protein aggregation

It is clear from Figure 2. 9 that *BsYpfP* can be obtained at very high purity by a Ni-NTA column. However, the homogeneity was not optimal initially, and thus a gel filtration assay was used to screen detergents to reduce the aggregates. In order to prevent aggregation, detergents and/or osmolytes were added to the protein solution after Ni-NTA purification. The mixture was then loaded on the gel filtration column to determine the effectiveness of the added compounds. If the protein peak eluted at the void volume (0.3 CV), indicating aggregates, the additive was considered ineffective. On the contrary, if a protein peak eluted at a volume corresponding to its molecular weight at a particular oligomer state, the additive was considered to selectively stabilize that oligomer state and effectively disrupt aggregates. Despite being purified from the soluble fraction, *BsYpfP*-FLAG-His6 protein was prone to aggregate and easily precipitated out of the solution at a concentration of only 2 mg/mL, as suggested by the gel filtration profile shown in Figure 2.10 (A). However, except for some minor improvements indicated by a smaller peak at the void volume (8 ml) in Figure 2.10 (B) compared to Figure 2.10 (A), aggregates still exist after extensive tests.

In another approach, small aliquots of the membrane fraction, instead of the soluble fraction, were resuspended and solubilized with a variety of detergents. The mixture was then ultracentrifuged and the supernatant was collected. A western blot was used to distinguish which detergent can keep the protein in solution. Subsequently, a large scale membrane protein purification was performed by using the detergent chosen from the test. The purified protein was also applied to a gel filtration column equilibrated with a buffer containing the detergent to confirm the result. Two detergents, namely LAPAO and CHAPS, were found to successfully suppress the protein aggregation for MBP-*BsYpfP*-His6 constructs as indicated by a major peak

around 14 mL but very little aggregation peak around 8 mL in Figure 2.10 (C) and (D). LAPAO was found to be particularly useful for the *BsYpfP*-His6 construct. The gel filtration profile of *BsYpfP*-His6 in buffer supplemented with LAPAO (Figure 2.10 (D)) shows no aggregation peak and pretty good homogeneity of the main peak around 15.6 mL.

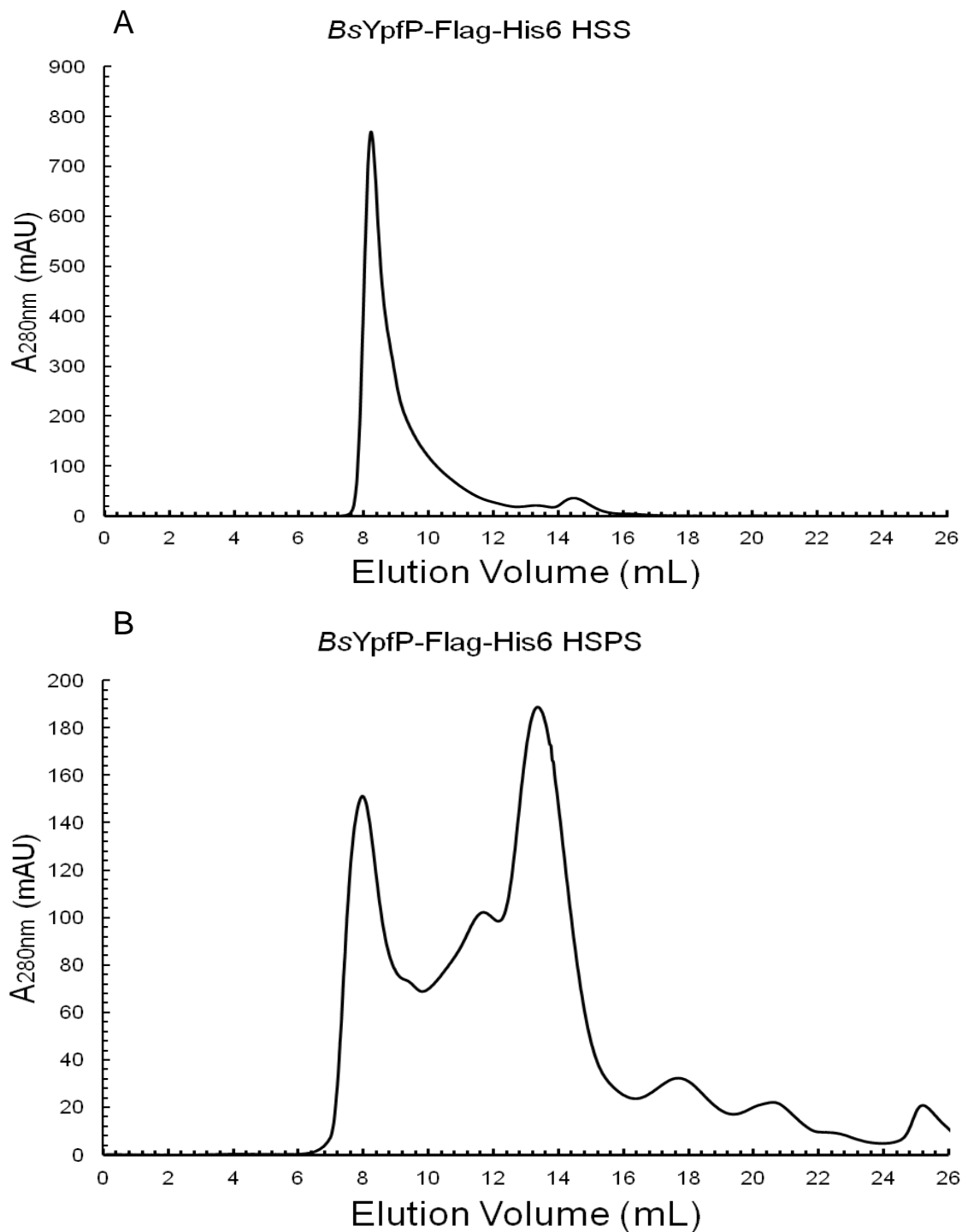
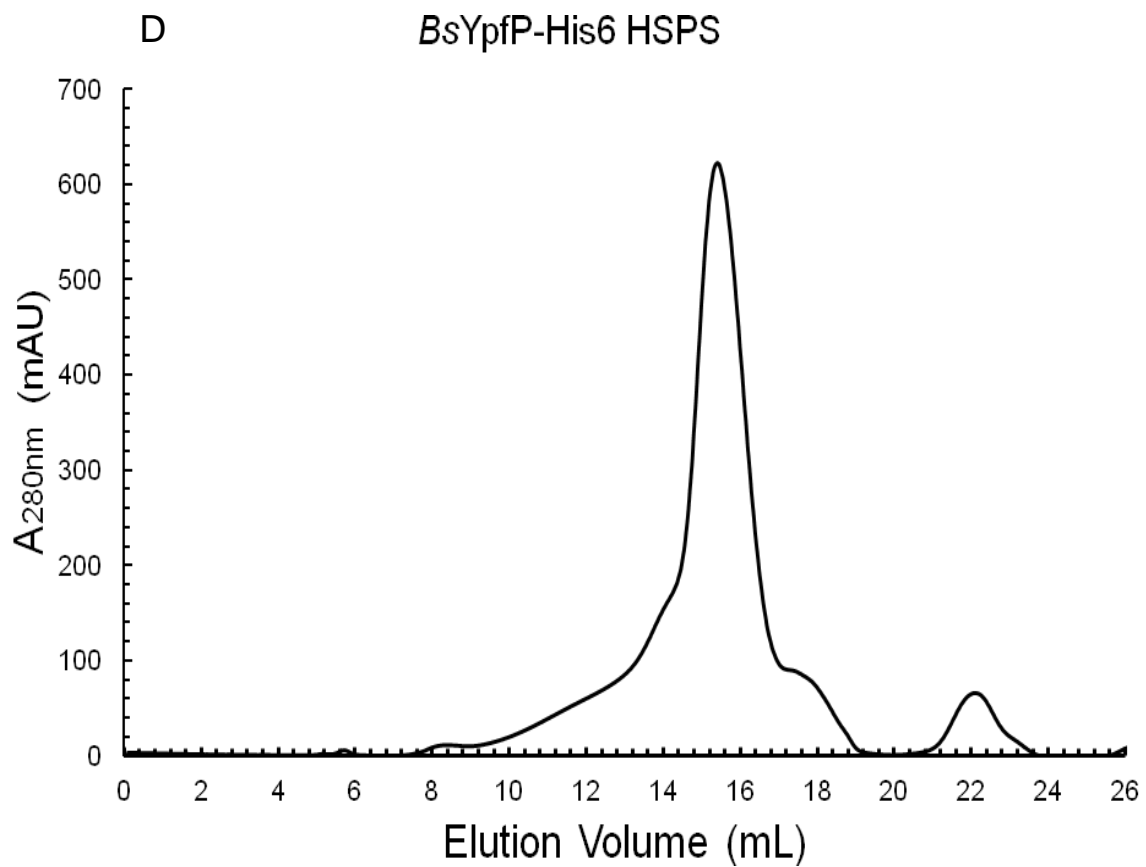
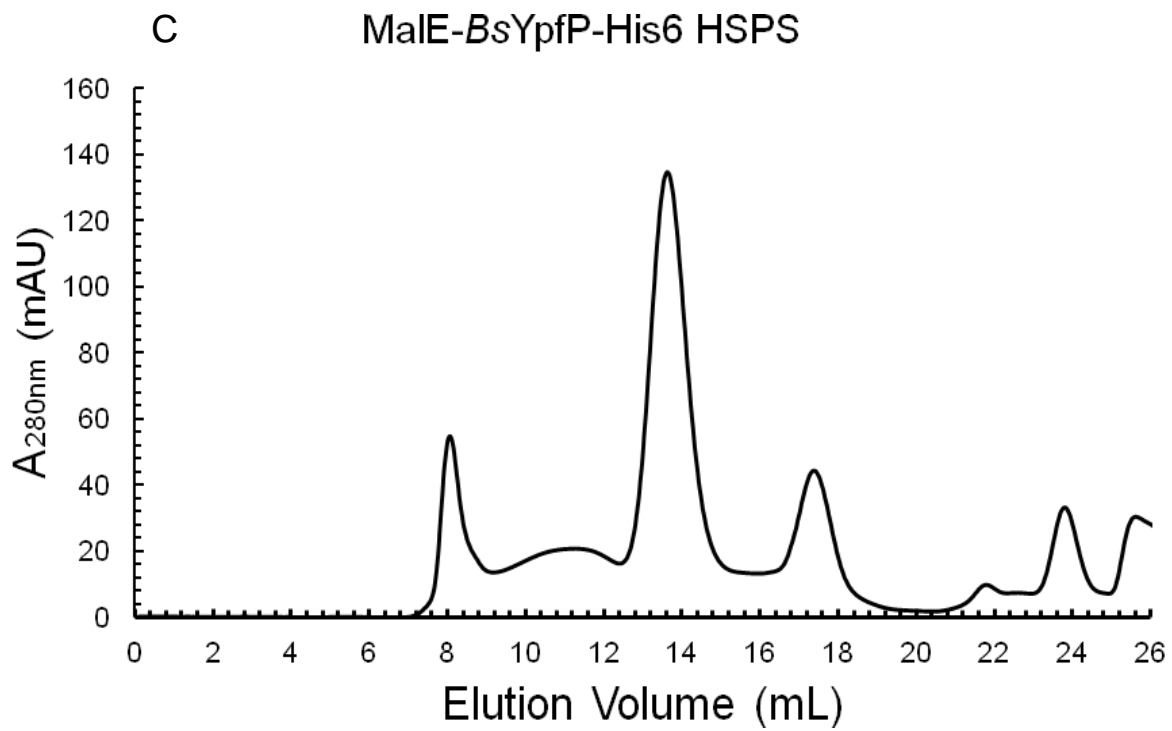


Figure 2.10 Gel filtration elution profile of *BsYpfP*. (A) *BsYpfP*-Flag-His6 purified from HSS fraction run on Superdex[®] 200 column without detergent. (B) *BsYpfP*-Flag-His6 purified from HSPS fraction run on Superdex[®] 200 column in detergent Z3-12. (C) MalE-*BsYpfP*-Flag-His6 purified from HSPS fraction run on Superose[®] 6 column in detergent LAPAO. (D) *BsYpfP*-His6 purified from HSPS fraction run on Superose[®] 6 column in detergent LAPAO.

Figure 2.10 Continued.



2.3.3 Production of glycolipids in *E. coli* as the proof of *in vivo* activity

Since *E. coli* cannot synthesize DAG-based glycolipids, the introduction of a foreign glycosyltransferase gene like *ypfP* may produce glycolipids in the *E. coli* plasma membranes, which can then be detected by TLC. This can be readily used as an *in vivo* activity assay for recombinant YpfP expressed in *E. coli*. Although the endogenous DAG level in *E. coli* is not very high [59], all the *BsYpfP* constructs can render the transformed *E. coli* capable of producing glucosyl-glycosyl DAG. Interestingly, two additional glycolipids bands were observed for the *BsYpfP*-FLAG-His6 construct (Figure 2.11). The detection of glycolipids in the *E. coli* whole cell membranes confirms the *in vivo* activity of the recombinant *BsYpfP*. Based on the analysis of multiple sequence alignment (Figure 2.12), many residues are indicated to have a role in catalysis and/or substrate binding. Three residues of YpfP, specifically N16, H18, and D136, were thus changed to alanine or similar residues by the QuikChangeTM mutagenesis procedure. Results of the whole cell lipid analysis were summarized in Table 2.2 and Figure 2.13.

Table 2.2 Summary of glycolipids generated by YpfP in *E. coli* revealed by TLC analysis.

YpfP Constructs for heterologous expression in <i>E. coli</i>	β MGlcDG	β DGlcDG	β TriGlcDG	Unkown GL I	Unkown GL II
<i>Bs</i>YpfP-Flag-His6	--	++	++++	++	++
MBP-<i>Bs</i>YpfP-His6	--	++	--	--	--
<i>Bs</i>YpfP-His6	--	++	--	--	--
<i>Bs</i>YpfP-Flag-His6 H18A	--	--	--	--	--
<i>Bs</i>YpfP-Flag-His6 H136A	--	--	--	--	--
<i>Bs</i>YpfP-His6 N16D	--	++	--	--	--
<i>Bs</i>YpfP-His6 H18N	--	+	--	--	--
<i>Bs</i>YpfP-His6 H18D	--	--	--	--	--

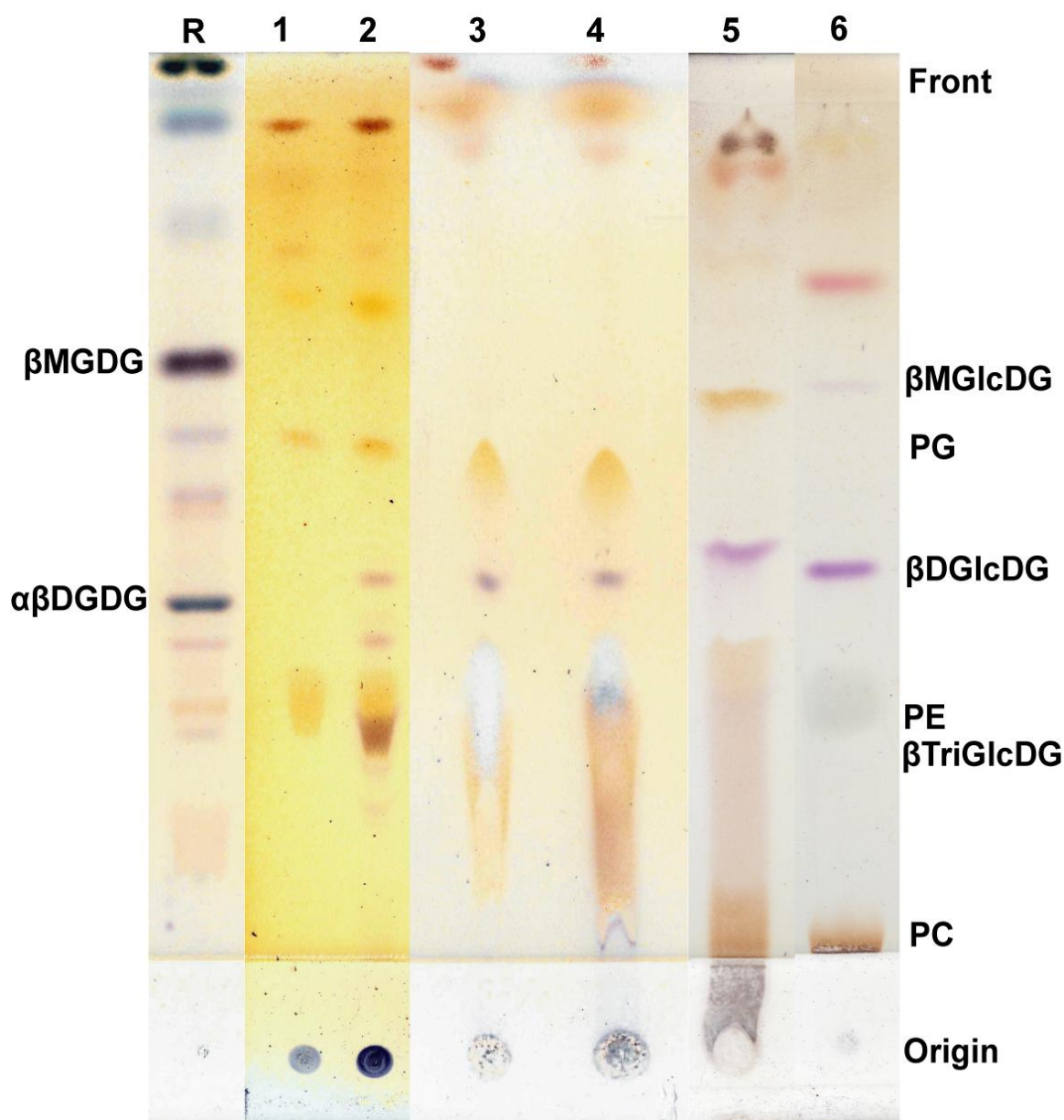
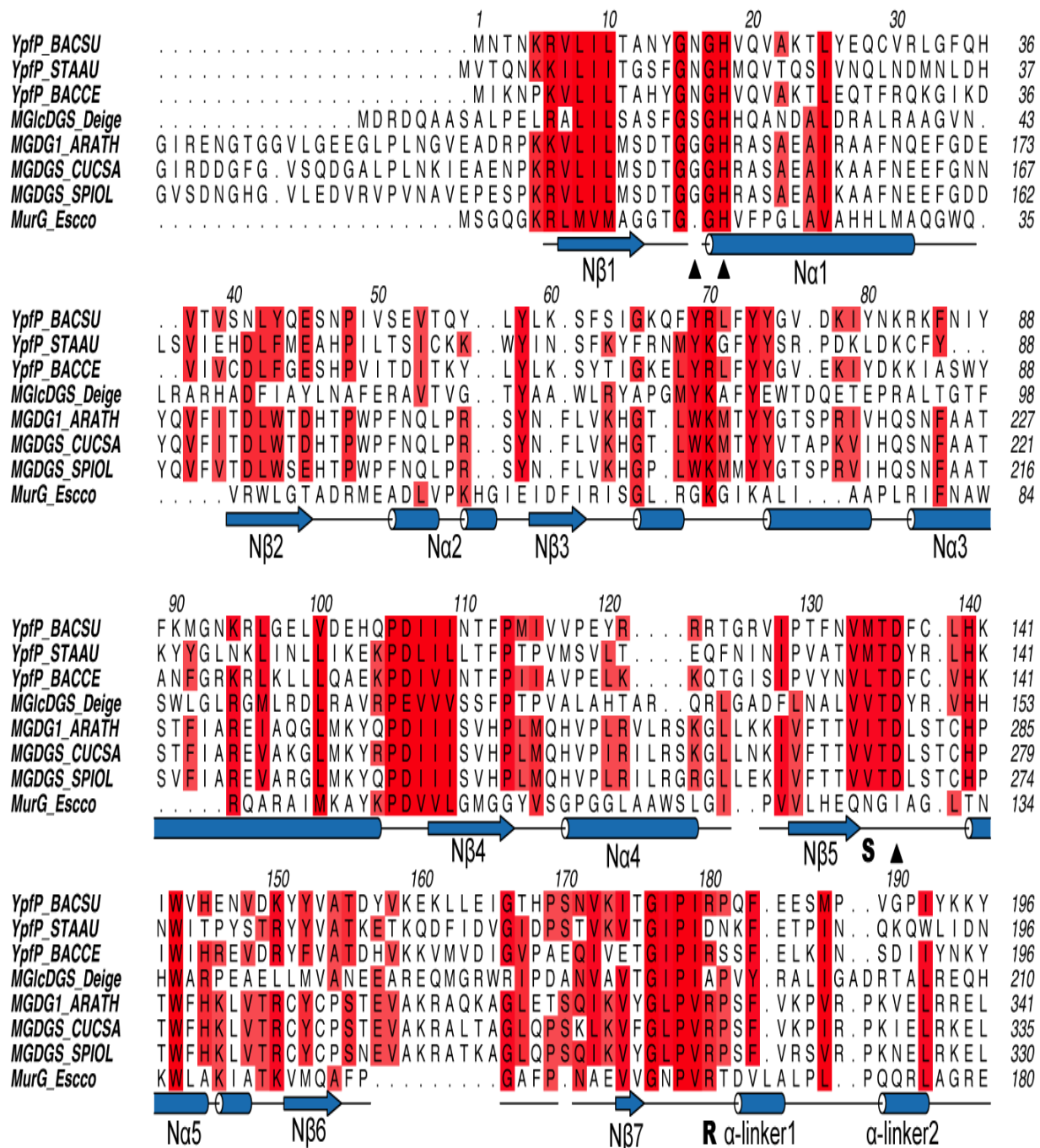


Figure 2.11 *In vivo* activity of *BsYpfP*. TLC of lipid extracts from *E. coli* whole cell membranes stained for glycolipids with α -naphthol reagent. Lane R: plant galactolipids from *A. thaliana* chloroplast membranes as reference; lane 1: whole cell membrane lipids from *E. coli* transformed with empty vector control; lane 2: glycolipids from *E. coli* with overexpression of *BsYpfP*-Flag-His6; lanes 3&4: glycolipids from *E. coli* with overexpression of MBP-*BsYpfP*-Flag-His6; lane 5, glycolipids from *E. coli* with overexpression of *BsYpfP*-His6; lane 6: glycolipids generated from *in vitro* liposome assay with *BsYpfP*-His6.

Figure 2.12



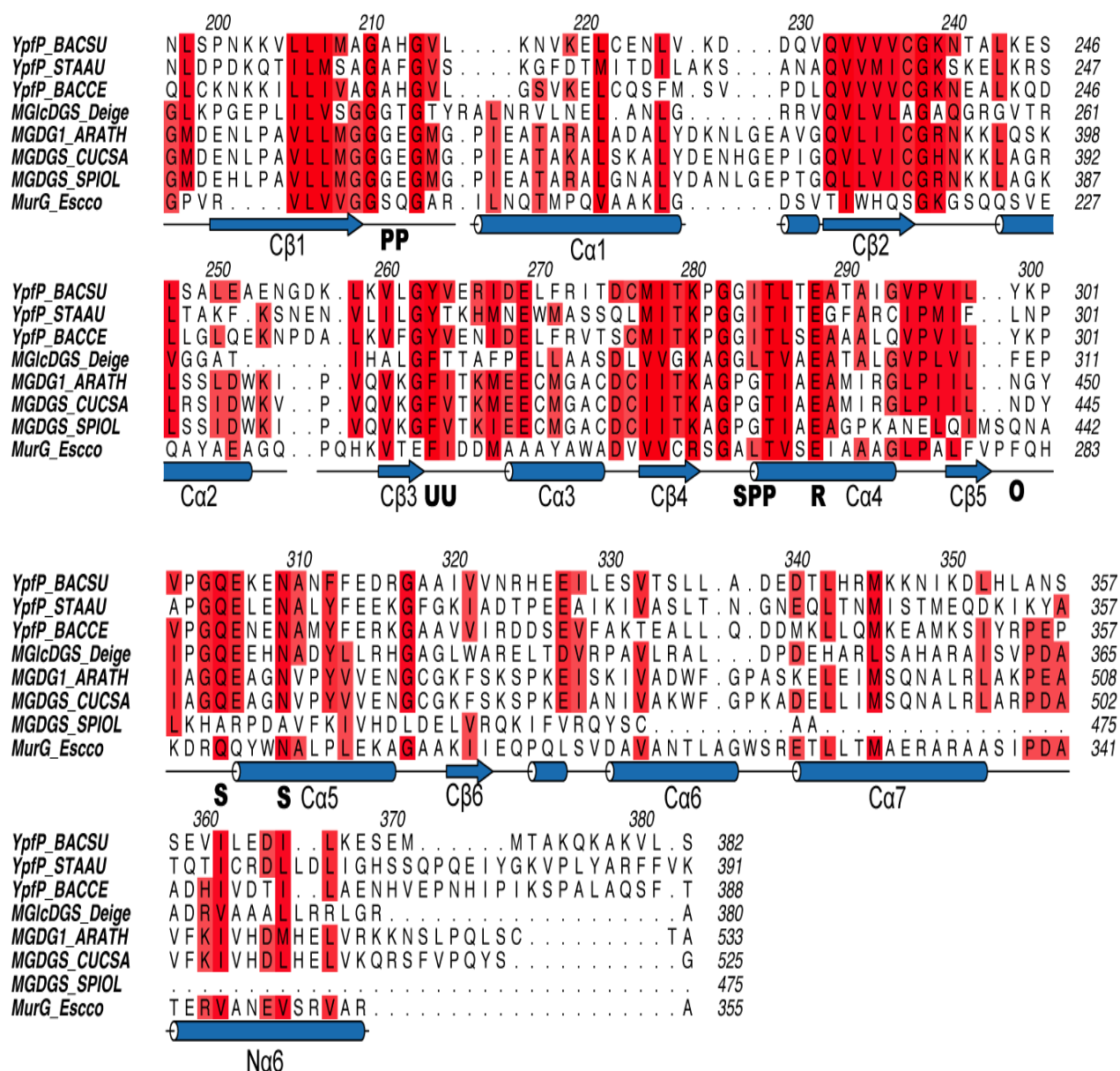


Figure 2.12 Multiple sequence alignment of YpfP, MGD synthase and MurG. Multiple sequence alignment was built through 3D-TCoffee server [60] against the 3D structure of *E. coli* MurG (PDB 1nlm). The numbering scheme and secondary structure profile at the bottom of the alignment refers to *E. coli* MurG sequence. Shown in the alignment are YpfP_BACSU (YpfP from *B. subtilis*), YpfP_STAAU (YpfP from *Staphylococcus aureus*), YpfP_BACCE (YpfP from *Bacillus cereus*), MGlcDGS_DEIGE (YpfP from *Deinococcus geothermalis*), MGDG1_ARATH (MGD synthase-1 from *A. thaliana*), MGDGS_CUCSA (MGD synthase from *Cucumis sativa*), MGDGS_SPIOL (MGD synthase from *Spinacia olearcea*), CLOTE_MGlcDGS (YpfP from *Clostridium tetani*), MurG_ESHEC (MurG from *Escherichia coli*). Identical residues are highlighted in red. Point mutations N16D, H18A and D136A in YpfP_BACSU mentioned in the thesis are labeled with filled triangle. Residues related to substrate binding in MurG are labeled with substrates where P stands for phosphate, R stands for sugars, U stands for uracil, and O stands for N-acetyl group.

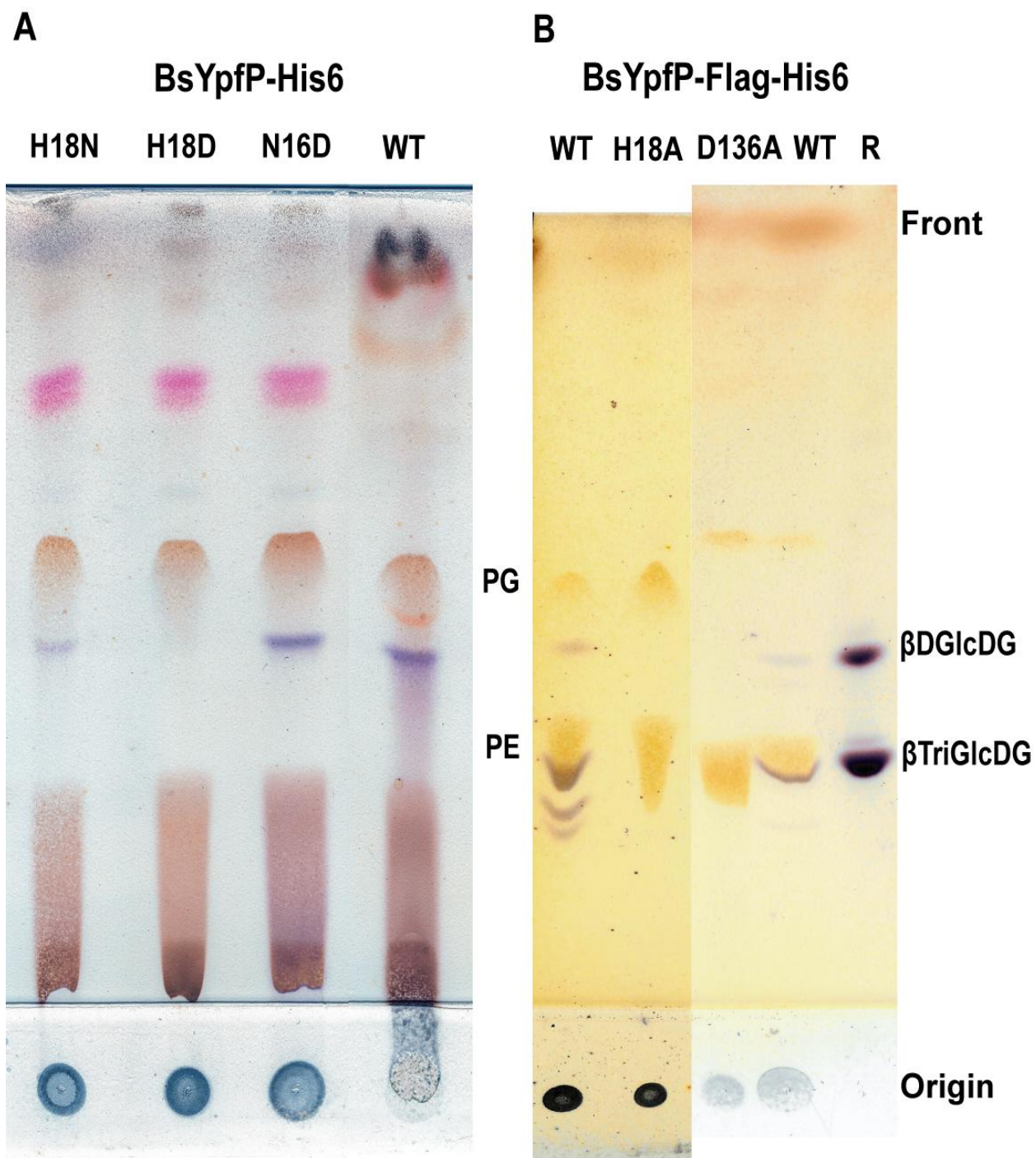


Figure 2.13 *In vivo* activity of *BsYpfP* mutants. TLC of lipid extracts stained for glycolipids with α -naphthol reagent. (A) from left to right: whole cell membrane from *E. coli* with overexpression of *BsYpfP*-His6 carrying mutations H18N, H18D, N16D, and its wild type; (B) from left to right: whole cell membrane from *E. coli* with overexpression of wild type *BsYpfP*-Flag-His6, mutants H18A, D136A, a repeated wild type sample, and purified glycolipids as reference.

2.3.4 Determination of the *in vitro* activity of *BsYpfP*

2.3.4.1 Incorporation of DAG into the giant liposome for *in vitro* activity assay

Giant liposomes made upon the agarose swelling on a glass substrate can reach millimeter in both height and diameter by merging of smaller liposomes. A continuous multilayer membrane can be visualized even by the naked eye. The large surface area of the liposome provides a pseudo-native membrane environment for the purified *BsYpfP*. Small volumes of DAG and matrix phospholipids dissolved in chloroform and the usage of non-radioactive soluble substrate substantially reduced the cost of the activity assay. Smaller size was observed for the DAG-containing liposome than the one with phospholipid only. The negative membrane curvature of DAG may account for the membrane discontinuity and the release of the swelling pressure. Nonetheless, *in vitro* studies show that the recombinant *BsYpfP* can indeed act on DAG and produce glycolipids, as seen by the formation of purple bands on the TLC plate (Figure 2.14); no corresponding bands were observed in the control assays where either the enzyme or the substrate DAG was absent. The TLC results also show a much darker band for the diglucosyldiacylglycerol (DGDG) than the monoglucosyldiacylglycerol (MGDG). A time course has been measured for the production of mono- and di-glucosyl lipids (Figure 2.15). The result also shows that *BsYpfP* predominantly transfers two glucose moieties to the DAG with a high ratio of DGDG/MGDG.

2.3.4.2 Recombinant *BsYpfP* has *in vitro* processive activity

MGDG and DGDG are the major glycolipids formed in the reaction catalyzed by *BsYpfP*-His6. Densitometry analysis revealed the ratio of DGDG:MGDG was more or less constant, at all of the 12 time points recorded. This result is in line with a previous study,

suggesting that a two-step accumulation of mono-glucosylated and di-glucosylated diacylglycerol was not observed [39]. Although the detailed mechanism is unknown, these results suggest that the transfer of the second glucosyl moiety happens much faster than the first, or the dissociation rate of the MGDG from the enzyme is much slower than the final product. Nonetheless these data clearly indicate that the recombinant *BsYpfP* has processive activity *in vitro*.

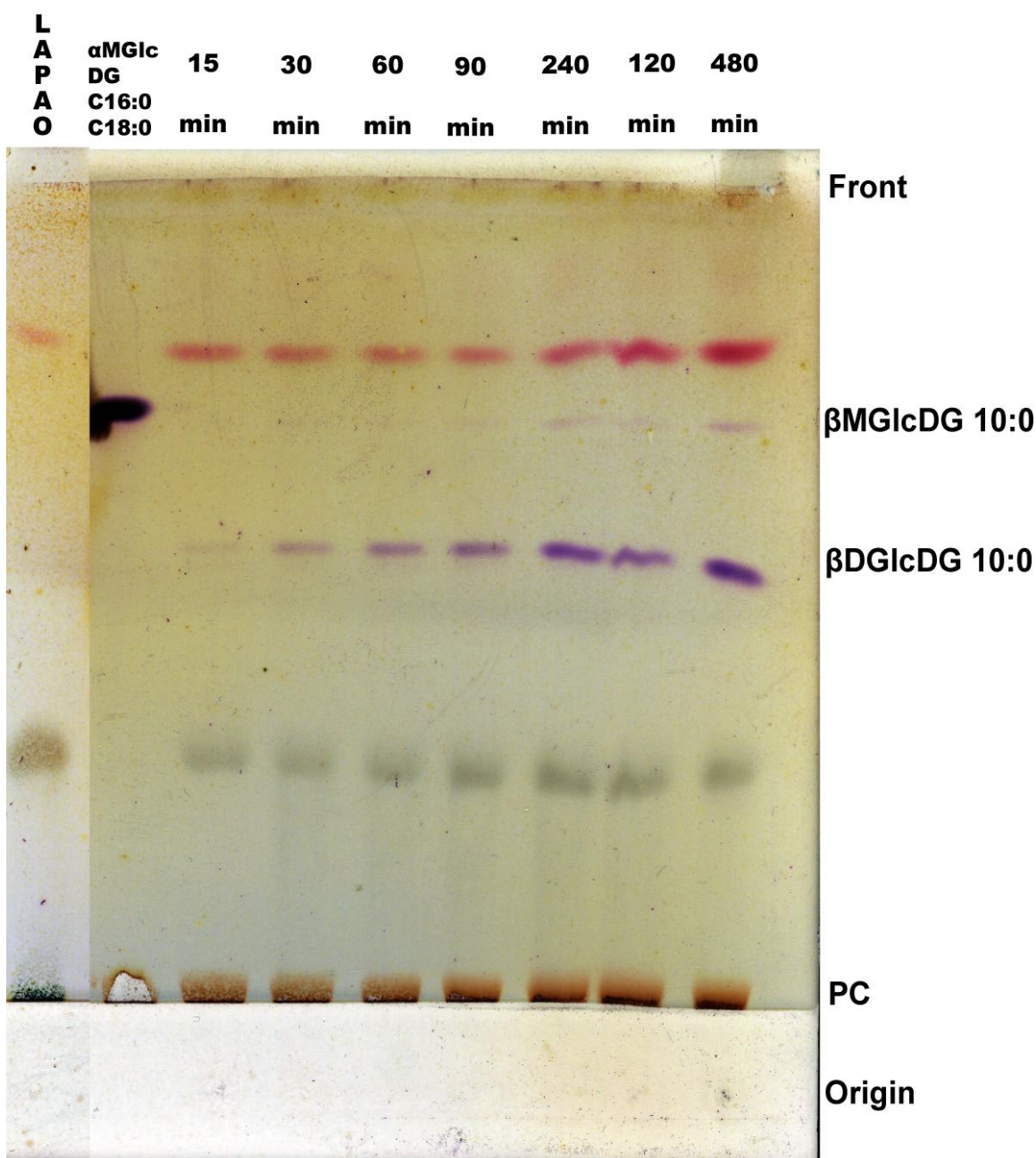


Figure 2.14 Time course studies of *BsYpfP*-His6 *in vitro* activity. Lipids extracted at different time points from the giant liposome assay are analyzed by TLC and visualized with α -naphthol staining. From left to right: the first lane is a negative control where the substrate DAG was absent; the second lane is the commercial mono α -glucosyl DAG as a reference.

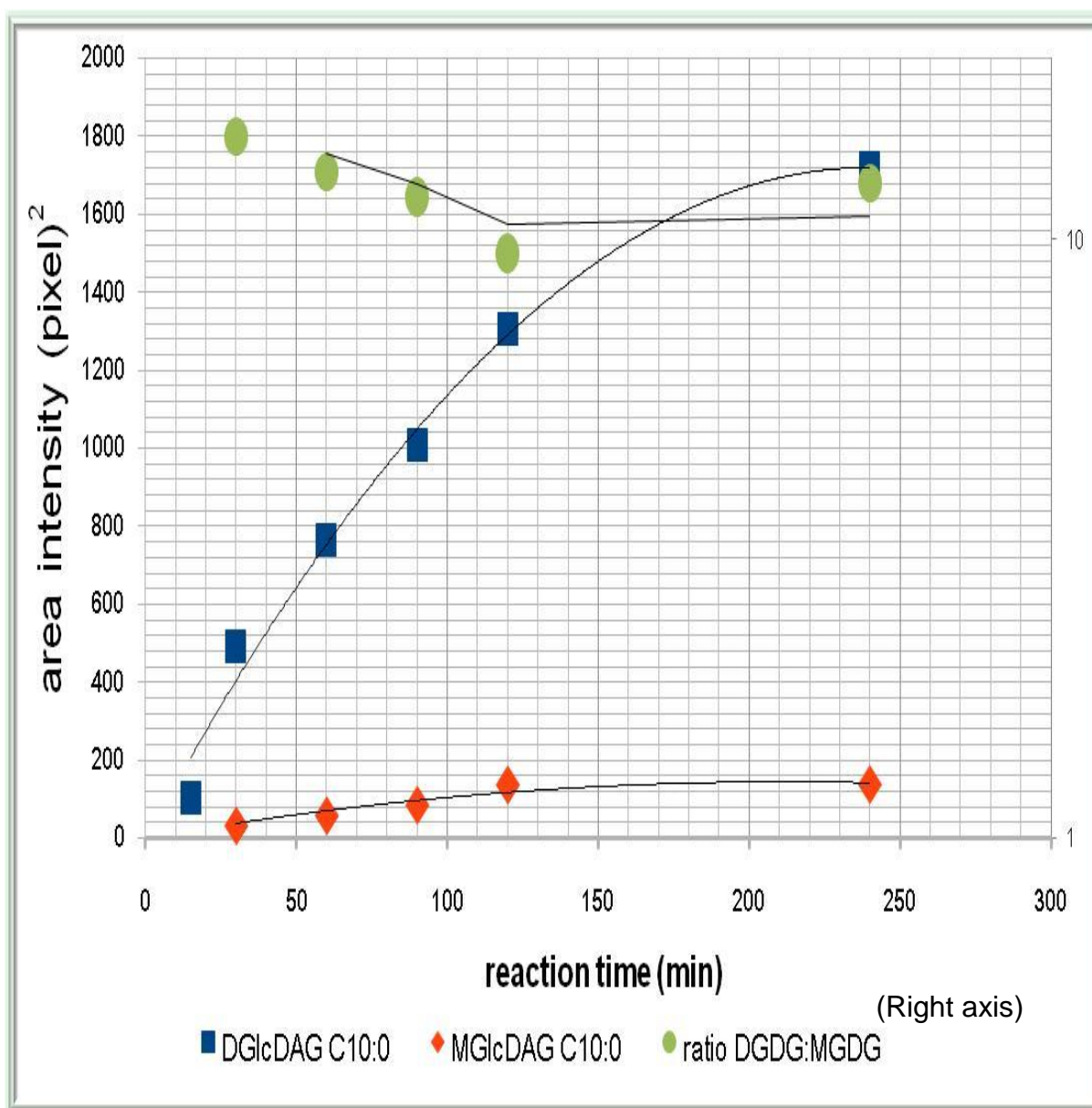


Figure 2.15 *In vitro* glycolipids accumulation curve of *BsYpfP-His6*. Quantification of the glycolipids by intensity count is done by Scion Image software.

2.3.4.3 Recombinant BsYpfP can utilize DAG of various acyl chain lengths

Pure *sn*-1,2 DAG with different acyl chain lengths was tested as the substrate of the recombinant BsYpfP-His6. The results show that YpfP adds the sugar substrate to the DAG regardless of the length of acyl chain (Figure 2.16). This result also indicates that substrate specificity of the DAG is assured mostly through the headgroup region where the five potential hydrogen bonding oxygen atoms are located. TLC analysis showed a strong correlation between a higher mobility of the diglucosyldiacylglycerol and a longer acyl chain length of the DAG. This result is expected due to the decrease in polarity of the glycolipids with longer acyl chain length.

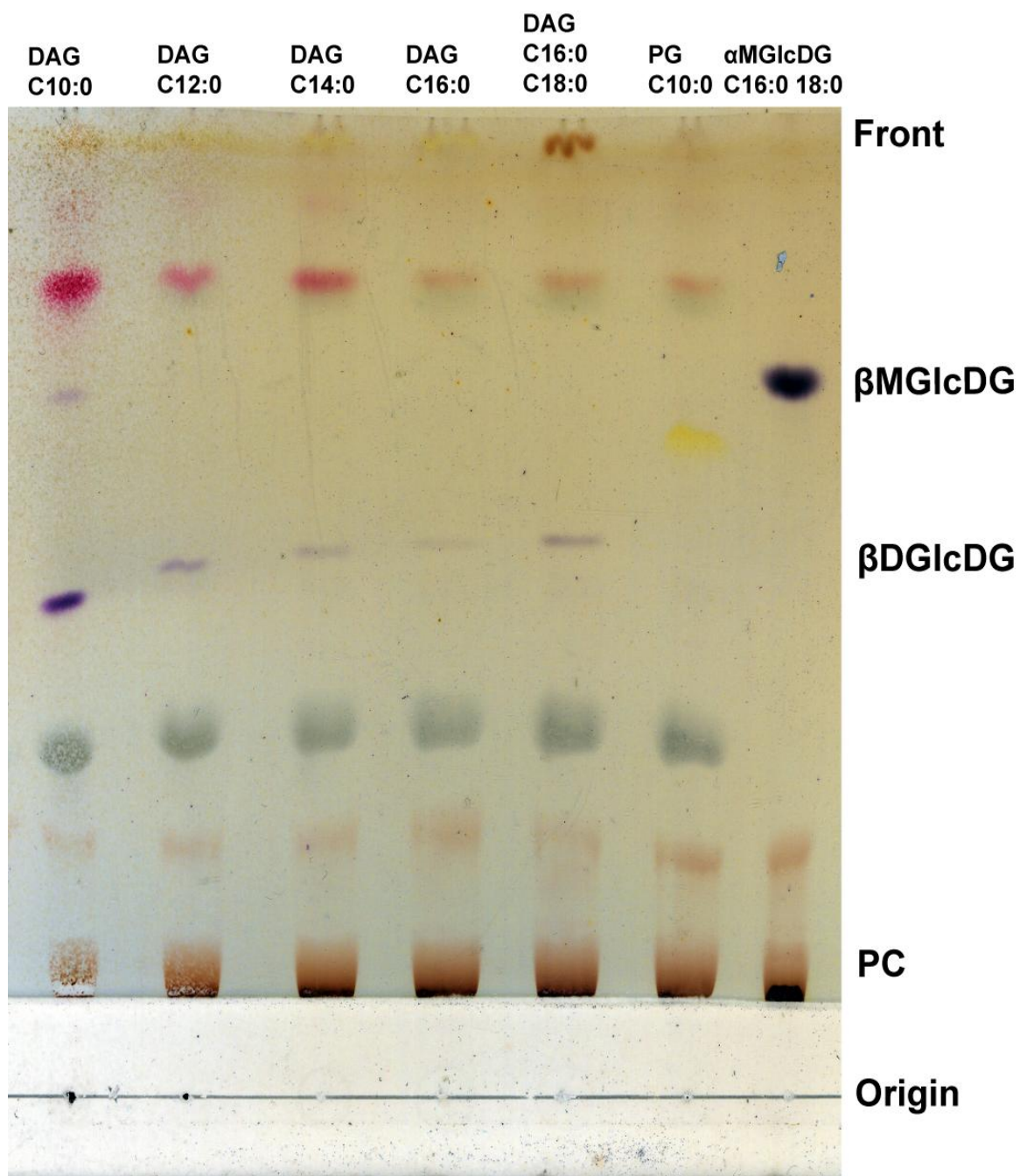


Figure 2.16 Substrate specificity of *BsYpfP*. From left to right, DAGs with increasing acyl chain lengths are tested as *in vitro* substrates. Production of the glycolipids is analyzed by TLC stained with the α -naphthol reagent.

2.3.4.4 Recombinant *Bs*YpfP-His6 has highest activity at pH 7.5

A pH profile of *in vitro* activity was measured for the *Bs*YpfP-His6 construct based on the intensity of alpha-naphthol stain shown on TLC corresponding to the glycolipids formed under various pH conditions. Intensities of the spots were measured by densitometry using ScionImage software. Maximum activity was observed at pH 7.5 for *Bs*YpfP-His6 as shown in Figure 2.17.

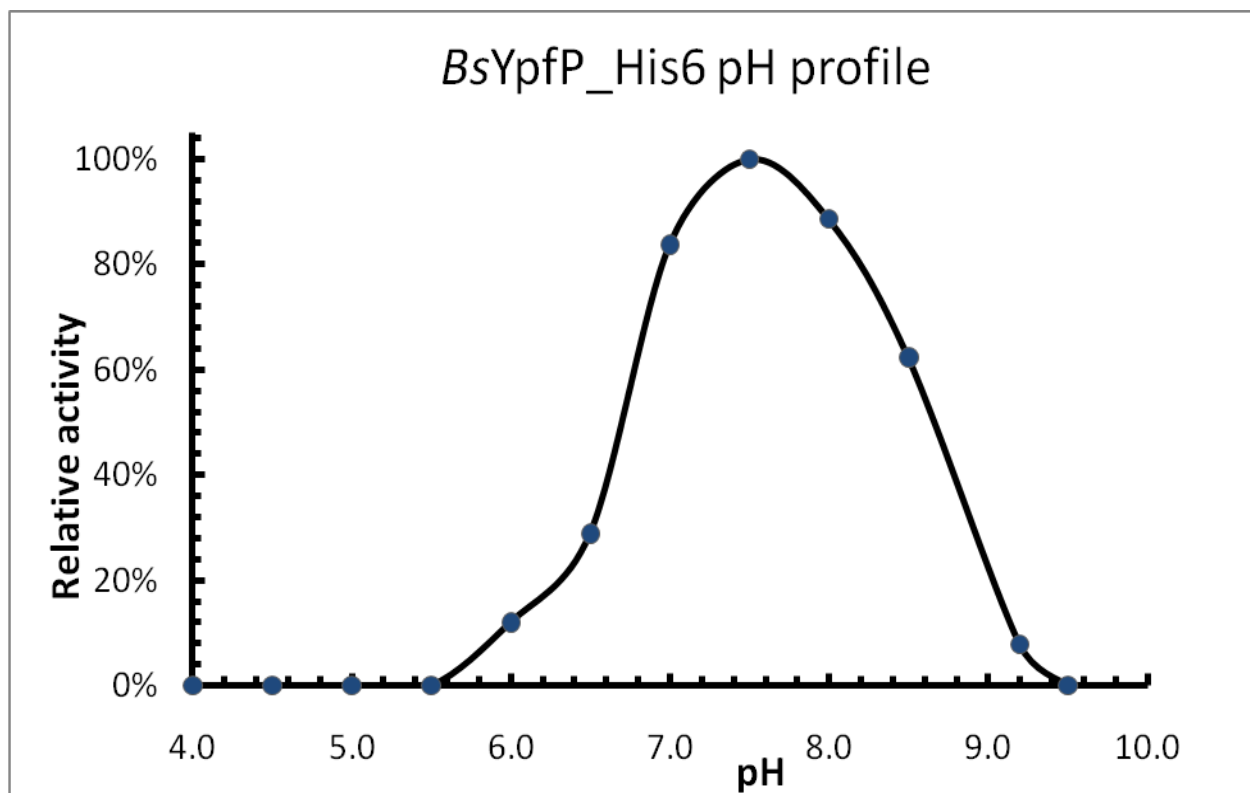


Figure 2.17 The pH profile of *BsYpfP* *in vitro* activity. *In vitro* activity of *BsYpfP*-His6 was measured at different pH conditions based on the production of glycolipids. The relative activity is calculated with highest activity as 100% at pH 7.5.

2.3.4.5 Successful synthesis of the UDP-6-deoxy-glucose

In order to obtain detailed enzymatic parameters of the first transfer reaction catalyzed by YpfP, a scheme was designed for the synthesis of the substrate UDP-6-deoxy-glucose from UDP-glucose (Figure 2.8). Preliminary HPLC results (Figure 2.18 (B)) showed an approximately 95% conversion rate from UDP-glucose to UDP-4-keto-6-deoxy-glucose was achieved in a reaction catalyzed by Rhm2N, a UDP-glucose dehydratase from *A. thaliana*. Two new peaks were found in the HPLC chromatogram with retention times of 15.51 and 16.86 min, respectively, when UDP-4-keto-6-deoxy-glucose was chemically reduced by sodium borohydride (NaBH_4) as shown in Figure 2.18 (C-E). Responding to the higher amount of the reducing agent, the peak areas of the two peaks increased while the UDP-4-keto-6-deoxy-glucose decreased. The two peaks are well separated and putatively assigned to UDP- α -D-fucose and UDP- α -D-quinovose at this time. The ratio of the two peak areas is 2 to 3. However, the separation and purification of UDP- α -D-fucose and UDP- α -D-quinovose on a preparative column was not further pursued because no column was available for the large scale purification.

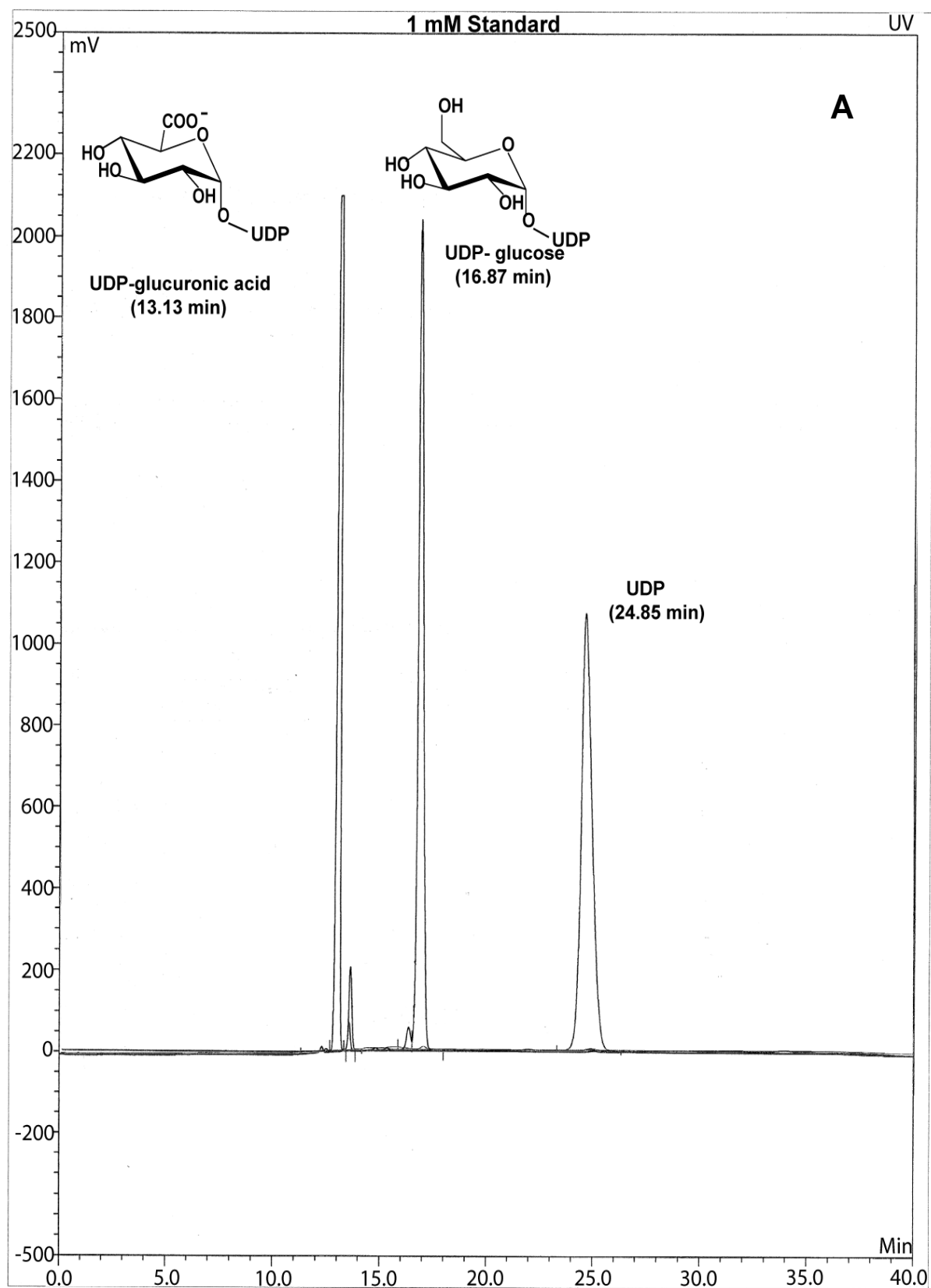


Figure 2.18 HPLC chromatogram of hybrid synthesis of UDP-6-deoxy-glucose. (A) Standards. (B) On column conversion of UDP-glucose to UDP-6-deoxy-glucose by Rhm2N. (C, D, E) Chemical reduction of UDP-6-deoxy-glucose to UDP-fucose/UDP-quinovose mixture by 2 μ L, 10 μ L, and 40 μ L of 100 mM NaBH₄, respectively.

Figure 2.18 continued.

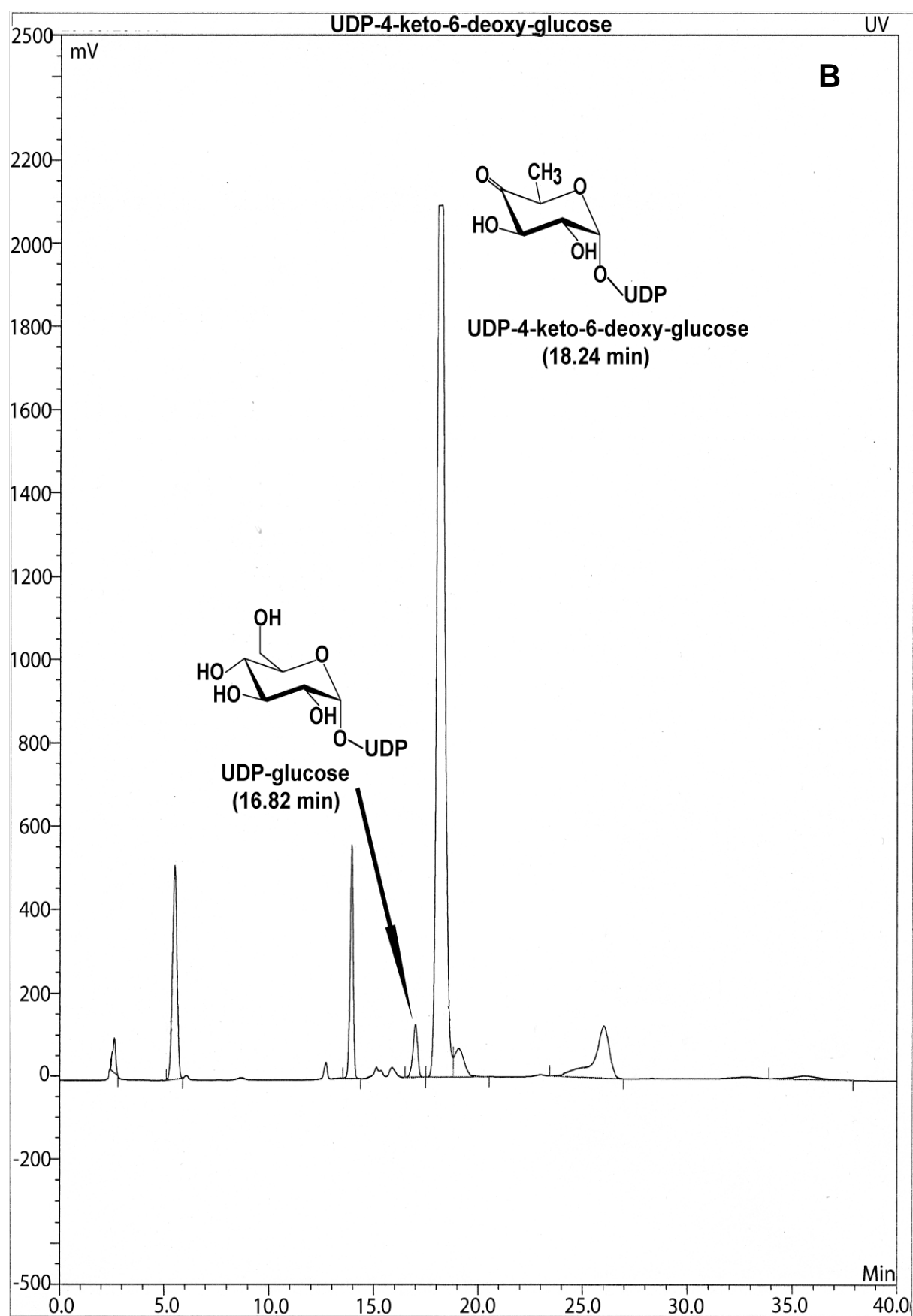


Figure 2.18 continued.

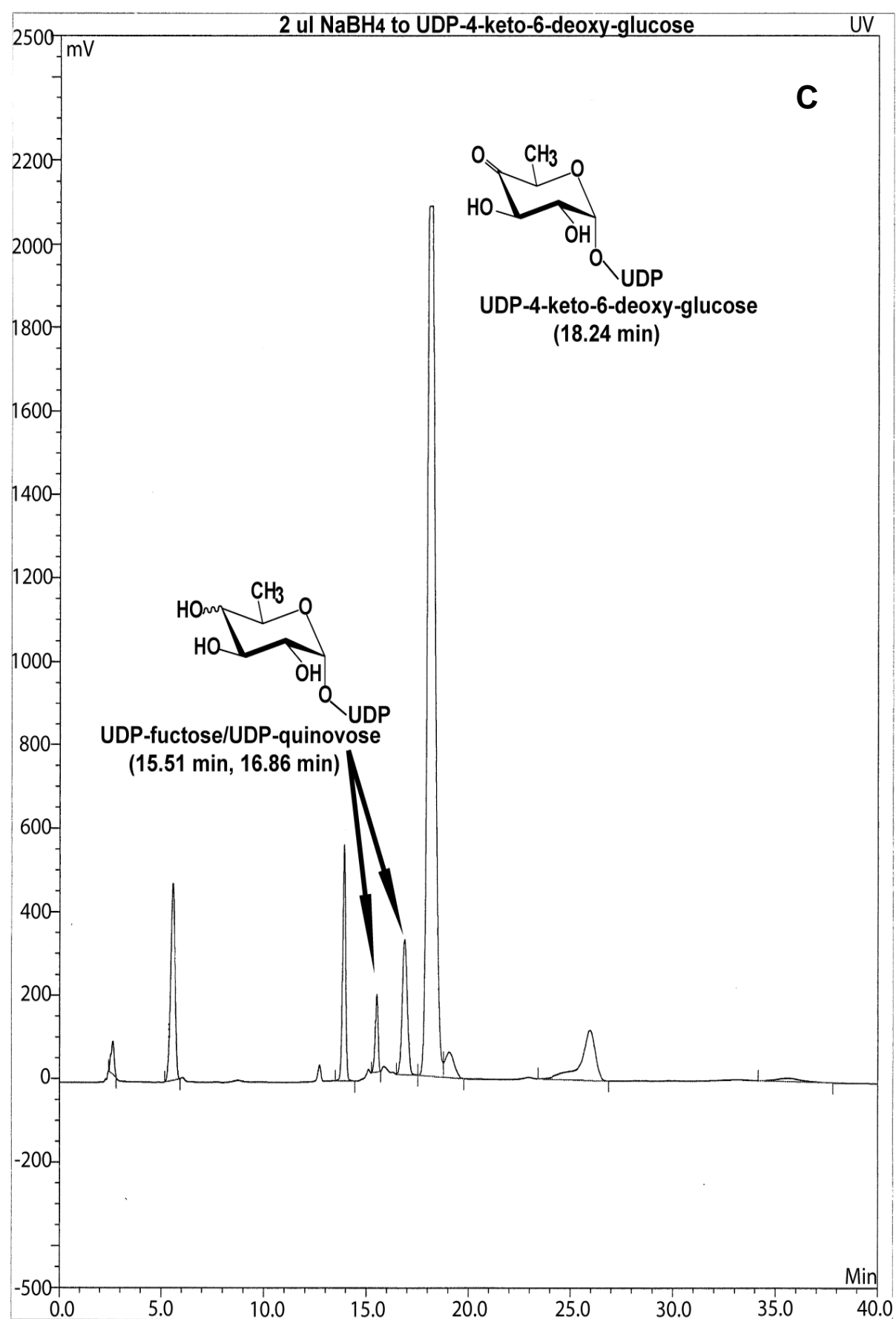


Figure 2.18 continued.

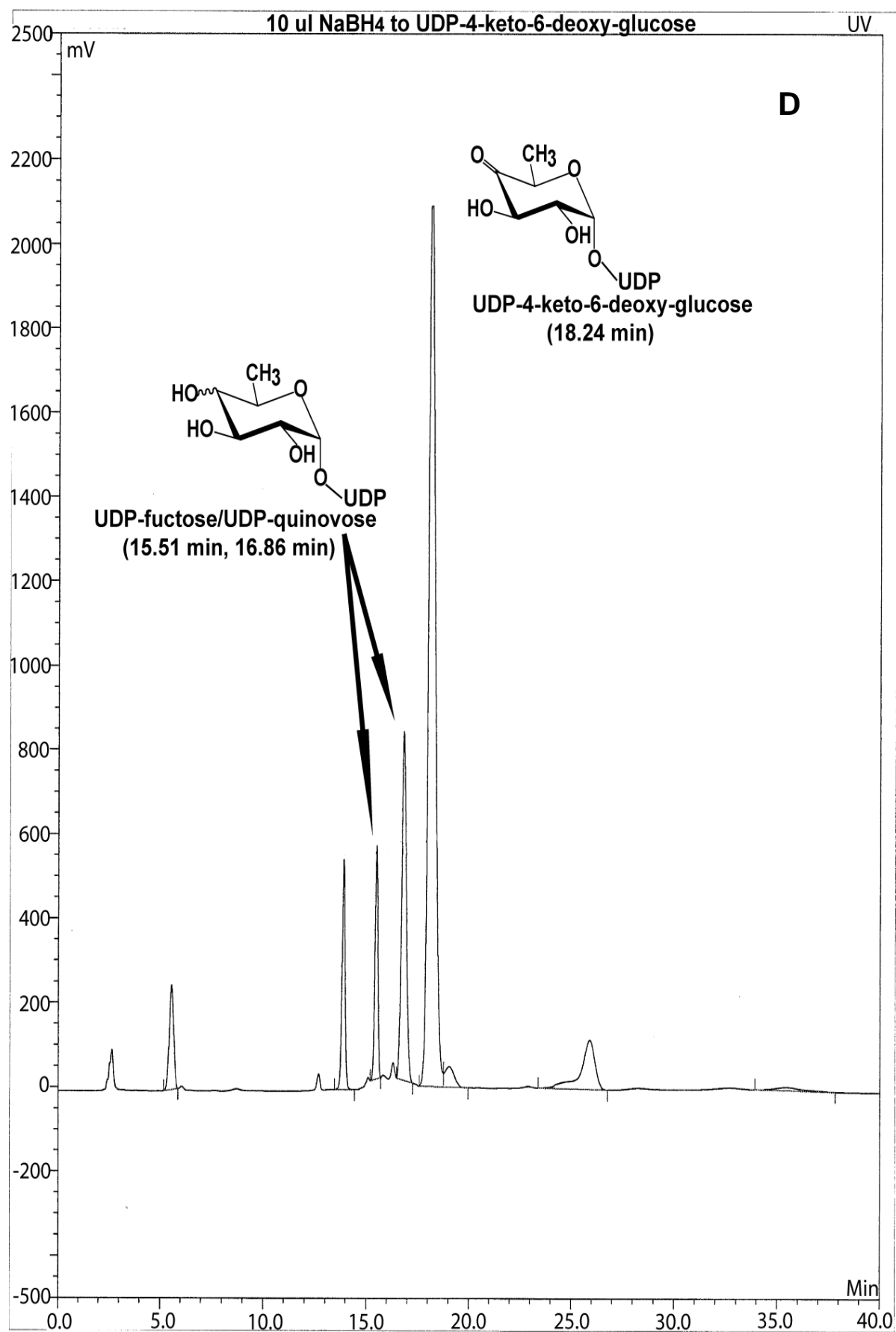
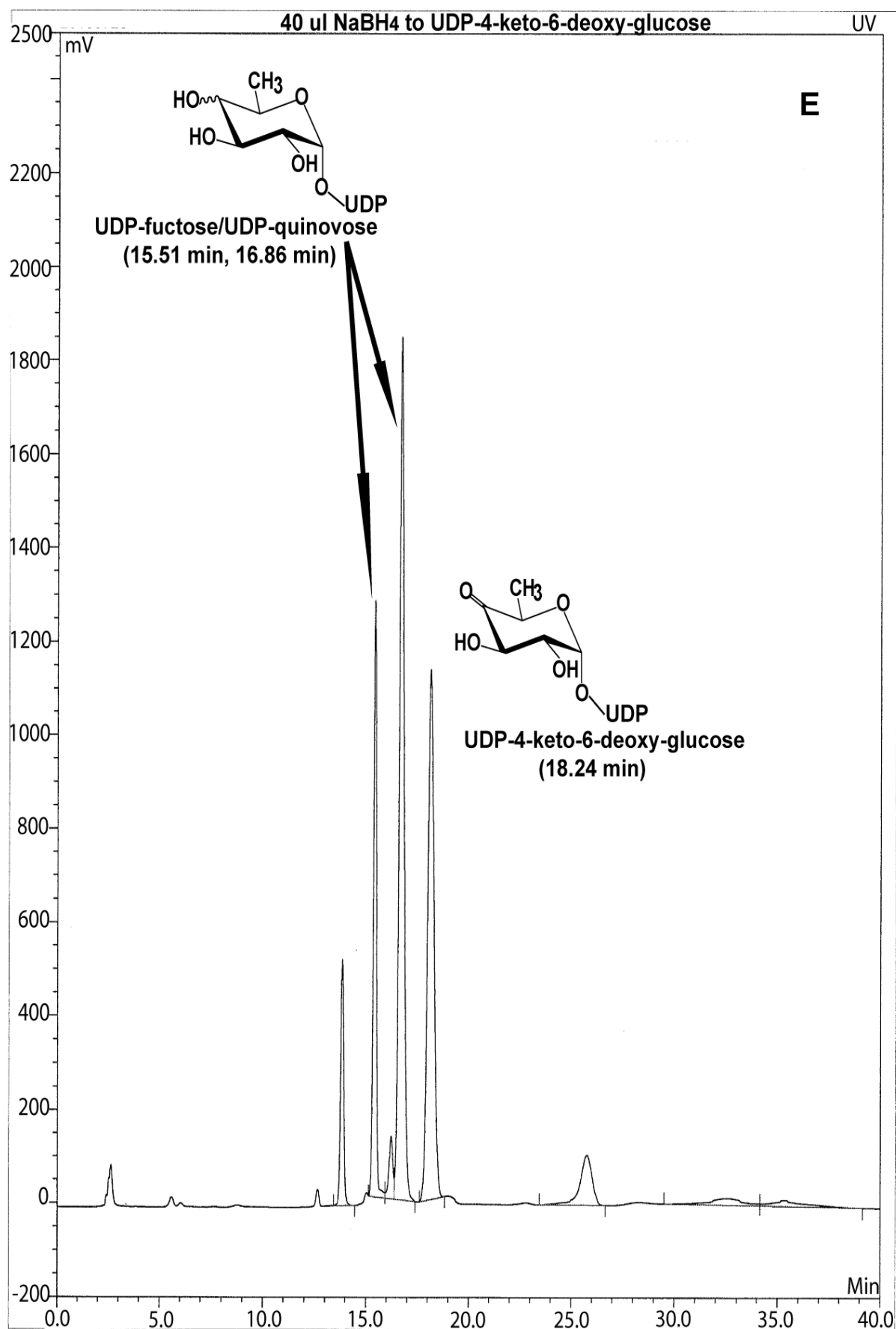


Figure 2.18 continued.



2.3.5 X-ray crystallography studies of the recombinant *BsYpfP*

2.3.5.1 Extensive screen of crystallization conditions for the *BsYpfP* with different tags

BsYpfP-His6, *BsYpfP*-FLAG-His6, MBP-*BsYpfP*-His6 were subjected to various high throughput screening conditions in either sitting drop or microbatch under oil format. In general, equal amounts of protein solution and screen solution were dispensed to the wells by the robot and mixed. In sitting drop format, plates were sealed in the presence of 55 μ L screen solution in the reservoir well. In microbatch format, plates are topped with a layer of Al's oil and vapor was constantly drawn to the surrounding wells with 0.75 M of LiCl solution. Plates were kept in 20 °C for crystals to grow. *BsYpfP* proteins easily precipitated out of solution and formed heavy brown amorphous precipitates in either high or low molecular weight polyethylene glycol (PEG) conditions while *BsYpfP* has relatively high solubility in high concentration salt solutions. In spite of extensive conditions screened and various strategies employed to refine screening conditions, all efforts failed to yield crystal hits worth further pursuing. Occasionally, detergent LAPAO crystals were obtained.

2.3.5.2 Successful crystallization of the recombinant HTH_{TagF}-*BsYpfP*-His6

Crystal growth and quality often requires stable crystal contacts preferably in all three directions. The purpose of the introduction of the HTH structure of *SeTagF* into *BsYpfP* was to reproduce such contacts found in TagF structures to promote the growth of *BsYpfP* crystals. HTH_{TagF}-*BsYpfP*-His6 was purified to homogeneity in the presence of the detergent Z3-14. One crystallization condition from the Hampton crystal screen II was found to produce hut-shape crystals of the fusion protein (Figure 2.19). This condition contained 100 mM sodium citrate, pH 5.6, 500 mM ammonium sulfate, and 1.0 M lithium

sulfate. Further optimization of buffer pH to 5.8 produced large enough crystals for X-ray diffraction tests. Interestingly, *SeTagF* crystallized in a similar condition of 100 mM HEPES, 2 M ammonium sulfate. Moreover, *BsYpfP*-His6 protein remained soluble in 2 M ammonium sulfate solution, while *HTH_{TagF}* -*BsYpfP*-His6 was precipitated by 1.6 M ammonium sulfate.

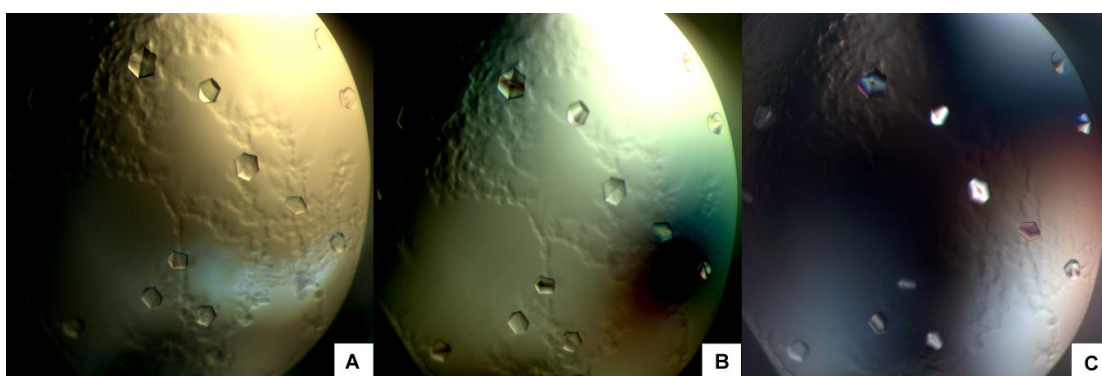


Figure 2.19 Photographs of HTH_{TagF} -*BsYpfP*-His6 crystals. (A) Photograph is taken under light microscopy. (B) and (C) are taken under crossed polarizing light.

2.3.5.3 Ring pattern diffraction images of the HTH_{TagF}-BsYpfP-His6 crystals

The crystals of HTH_{TagF}-BsYpfP-His6 were cryoprotected in 2.2 M lithium sulfate and tested at the synchrotron for X-ray diffraction. Unfortunately, the crystals were very fragile and partially deformed during mounting with cryoloops. Diffraction images of the crystals showed several rings instead of well separated spots. The ring diffraction was at low resolution and was not the typical solution ring observed for ice crystals, suggesting that the crystals were indeed YpfP crystals. This result might reflect the poor packing of the molecules within the crystals and further optimization and screening is needed.

2.4 Discussion

In general, a lipid-modifying membrane protein is usually known to be recalcitrant to crystallization. The more hydrophobic and the smaller the size of headgroup of the substrate lipids, the more difficult it is to get the protein crystals. Part of the reason may be that, in order to access the specific lipid substrate, a hydrophobic patch and/or transmembrane helices are often required for the enzyme. However, the hydrophobicity poses a great challenge for purification and crystallization of the enzyme. Unfortunately, diacylglycerol, a very biochemically important lipid, is one of the lipids with the smallest headgroup and, therefore, brings great challenges for purification and characterization of the enzymes that process it. The hydroxyl group that is very reactive in a series of enzymes, including kinases, glycosyltransferases, and acyltransferases, render DAG one of the most versatile molecules in metabolism and regulation. However, only the structures of diacylglycerol kinases have been determined up to now. Therefore, little is known regarding the structure of this important family of proteins, and more detailed biochemical and biophysical characterization is still needed. In this study we set out to purify and crystallize

one of these important DAG processing enzymes, YpfP, a glycosyltransferase from *B. subtilis*.

I have optimized the choices of the expression strains and membrane solubilizing detergents for recombinant *BsYpfP* during this study to set up a good platform for its further characterization. Also *BsYpfP* is a peripheral membrane protein and the recombinant protein can be expressed as soluble protein. The requirement for the detergent LAPAO for the purification of recombinant *BsYpfP* protein into the monodisperse and homogeneous solution confirmed the idea that a hydrophobic patch is present in the protein that renders it more prone to aggregation without the presence of membrane-like environments.

DAG is insoluble in water and is difficult to handle in an aqueous buffer alone. In addition, mixing DAG with phospholipids still gives rise to a turbid liposome solution upon hydration partly because of the negative spontaneous membrane curvature of DAG. It has been found that addition of a micelle-forming detergent, such as Triton X-100 or CHAPS, can reduce the turbidity at OD_{350nm}. Thus the *in vitro* activity of DAG kinases and MGlcDG synthase were assayed with the hydrophobic substrate DAG present in the form of mixed micelles [61-67]. In both cases, a radioactive soluble substrate (e.g. ATP or UDP-glucose) was used to increase the sensitivity, where the radioactivity in the organic phase was recorded. However, the high concentration of DAG used is very hard to dissolve and made this method inapplicable for YpfP. Therefore a giant liposome assay was developed for *in vitro* activity measurement of recombinant *BsYpfP* protein. This assay is an important improvement on the existing mixed micelle assay by providing the enzyme with a membrane-like environment without using the radioactive substrate. In addition, the relatively simple experimental setup makes mixing the diacylglycerol with the matrix

phospholipids more effective and less time consuming. The processivity of the *BsYpfP* observed by using the giant liposome assay is consistent with previous studies [39] and further confirmed the reliability of this assay system.

Having the capability to assay *BsYpfP* *in vitro* allows for more detailed and controlled mutagenic studies to characterize the active site residues. Multi-sequence alignment and homology modeling analysis (results not shown) suggested that the His18 residue may play a role as the general base expected in the inverting GT-B glycosyltransferases. However, the mutagenesis studies on the His18 mutants, measured with the giant liposome assay, did not indicate any significant change of activity. This result suggests that His18 may not play a critical role in catalysis, which raises the question as to which residue acts as the general base.

The strategy of masking the hydrophobic patch by addition of a protein aid was explored to assist the crystallization of *BsYpfP*. Initially a large protein, MBP, was fused at the *N*-terminus of *BsYpfP*. Quite unexpectedly, this fusion protein was found mostly in membrane fractions, indicating the patch was not sufficiently covered. As an alternative, a structural HTH motif found in TagF, a structurally-determined protein that has a similar GT-B fold with YpfP, was fused also at the *N*-terminus. This motif may function as an interface for higher order oligomeric assembly and also serve as a nucleation site for crystallization. Although the relatively small size of the HTH motif should not cover the hydrophobic patch, the spatial arrangement of the GT-B fold may enhance the relative exposure of the more hydrophilic C domain, allowing it to form the crystal contacts, while the more hydrophobic N domain is sequestered in the middle by the surrounding HTH motif. Regardless, this design indeed led to the successful crystallization of the HTH_{TagF}-*BsYpfP*-

His6. Even more, the similar crystallization conditions of TagF and HTH_{TagF}-BsYpfP-His6 confirmed that the HTH motif played a very important role in crystallization and reduced the need for extensive screening. This clearly confirmed the idea that a good crystallization module should impart its crystallization ability to its target protein, and therefore the crystallization condition of the fusion protein can be derived and optimized from the crystallization condition of the fusion module. Unfortunately, the current crystals were very fragile and diffracted poorly, suggesting that the fusion module alone could not provide enough crystal contacts to create high quality crystals. Thus, optimizing the crystal contacts from the target protein and the crystallization aid is still the challenge in the quest to obtain crystals of high quality for successful structural determinations.

2.5 Conclusions

Two novel approaches were applied in the biochemical and biophysical studies of the membrane peripheral protein BsYpfP. One was to make the DAG containing giant liposome for the successful determination of BsYpfP *in vitro* activity. The simplified lipid hydration and liposome forming process renders this approach very attractive for development into a more general assay for enzymes involved lipid metabolism by providing the membrane-like environments. The other one is to use a fusion protein for the successful crystallization of BsYpfP. The selection criteria of this fusion protein from known structure are: (1) stable crystal contacts to maintain high order of interactions under crystallizing conditions; (2) minimum sterical clashes between the fusion and target proteins; (3) minimum sterical clashes between the target proteins due to the geometrical arrangement of the fusion protein. It is reasonable to suggest that based on these criteria, a targeted search within the structure database will yield many more fusion motifs or proteins suitable for a

specific protein, such as TagF for YpfP used in this study. They may greatly enrich the tool sets available to crystallographers to tackle many more challenging proteins (i.e., weak self interacting proteins) such as membranes associated proteins.

REFERENCES

REFERENCES

- 1 van Meer, G., Voelker, D. R. and Feigenson, G. W. (2008) Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol* **9**, 112-124
- 2 Marsh, M. and Helenius, A. (2006) Virus entry: open sesame. *Cell* **124**, 729-740
- 3 Carrasco, S. and Merida, I. (2007) Diacylglycerol, when simplicity becomes complex. *Trends in Biochemical Sciences* **32**, 27-36
- 4 Szule, J. A., Fuller, N. L. and Rand, R. P. (2002) The effects of acyl chain length and saturation of diacylglycerols and phosphatidylcholines on membrane monolayer curvature. *Biophys J* **83**, 977-984
- 5 Kooijman, E. E., Chupin, V., Fuller, N. L., Kozlov, M. M., de Kruijff, B., Burger, K. N. and Rand, P. R. (2005) Spontaneous curvature of phosphatidic acid and lysophosphatidic acid. *Biochemistry* **44**, 2097-2102
- 6 Holzl, G. and Dormann, P. (2007) Structure and function of glycolipids in plants and bacteria. *Prog Lipid Res* **46**, 225-243
- 7 Pieringer, R. A. (1968) The metabolism of glyceride glycolipids. I. Biosynthesis of monoglucosyl diglyceride and diglucosyl diglyceride by glucosyltransferase pathways in *Streptococcus faecalis*. *J Biol Chem* **243**, 4894-4903
- 8 Block, M. A., Dorne, A. J., Joyard, J. and Douce, R. (1983) Preparation and characterization of membrane fractions enriched in outer and inner envelope membranes from spinach chloroplasts. II. Biochemical characterization. *J Biol Chem* **258**, 13281-13286
- 9 Mancuso, D. J. and Chiu, T. H. (1982) Biosynthesis of glucosyl monophosphoryl undecaprenol and its role in lipoteichoic acid biosynthesis. *J Bacteriol* **152**, 616-625
- 10 Crouvoisier, M., Mengin-Lecreulx, D. and van Heijenoort, J. (1999) UDP-N-acetylglucosamine:N-acetylmuramoyl-(pentapeptide) pyrophosphoryl undecaprenol N-acetylglucosamine transferase from *Escherichia coli*: overproduction, solubilization, and purification. *FEBS Lett* **449**, 289-292

- 11 Kawahara, K., Lindner, B., Isshiki, Y., Jakob, K., Knirel, Y. A. and Zahringer, U. (2001) Structural analysis of a new glycosphingolipid from the lipopolysaccharide-lacking bacterium *Sphingomonas adhaesiva*. *Carbohydr Res* **333**, 87-93
- 12 Takeuchi, M., Hamana, K. and Hiraishi, A. (2001) Proposal of the genus *Sphingomonas* sensu stricto and three new genera, *Sphingobium*, *Novosphingobium* and *Sphingopyxis*, on the basis of phylogenetic and chemotaxonomic analyses. *Int J Syst Evol Microbiol* **51**, 1405-1417
- 13 Haque, M., Hirai, Y., Yokota, K. and Oguma, K. (1995) Steryl glycosides: a characteristic feature of the *Helicobacter* spp.? *J Bacteriol* **177**, 5334-5337
- 14 Lee, H., Kobayashi, M., Wang, P., Nakayama, J., Seeberger, P. H. and Fukuda, M. (2006) Expression cloning of cholesterol α -glucosyltransferase, a unique enzyme that can be inhibited by natural antibiotic gastric mucin O-glycans, from *Helicobacter pylori*. *Biochem Biophys Res Commun* **349**, 1235-1241
- 15 Wunder, C., Churin, Y., Winau, F., Warnecke, D., Vieth, M., Lindner, B., Zahringer, U., Mollenkopf, H. J., Heinz, E. and Meyer, T. F. (2006) Cholesterol glucosylation promotes immune evasion by *Helicobacter pylori*. *Nat Med* **12**, 1030-1038
- 16 Lee, S. J., Lee, B. I. and Suh, S. W. (2011) Crystal structure of the catalytic domain of cholesterol- α -glucosyltransferase from *Helicobacter pylori*. *Proteins*
- 17 Ben-Menachem, G., Kubler-Kielb, J., Coxon, B., Yergey, A. and Schneerson, R. (2003) A newly discovered cholesteryl galactoside from *Borrelia burgdorferi*. *Proc Natl Acad Sci U S A* **100**, 7913-7918
- 18 LaRocca, T. J., Crowley, J. T., Cusack, B. J., Pathak, P., Benach, J., London, E., Garcia-Monco, J. C. and Benach, J. L. (2010) Cholesterol lipids of *Borrelia burgdorferi* form lipid rafts and are required for the bactericidal activity of a complement-independent antibody. *Cell Host Microbe* **8**, 331-342
- 19 Maier, R. M. and Soberon-Chavez, G. (2000) *Pseudomonas aeruginosa* rhamnolipids: biosynthesis and potential applications. *Appl Microbiol Biotechnol* **54**, 625-633
- 20 Abdel-Mawgoud, A. M., Lepine, F. and Deziel, E. (2010) Rhamnolipids: diversity of structures, microbial origins and roles. *Appl Microbiol Biotechnol* **86**, 1323-1336

- 21 Passeri, A., Lang, S., Wagner, F. and Wray, V. (1991) Marine biosurfactants, II. Production and characterization of an anionic trehalose tetraester from the marine bacterium *Arthrobacter* sp. EK 1. *Z Naturforsch C* **46**, 204-209
- 22 Peng, F., Liu, Z., Wang, L. and Shao, Z. (2007) An oil-degrading bacterium: *Rhodococcus erythropolis* strain 3C-9 and its biosurfactants. *J Appl Microbiol* **102**, 1603-1611
- 23 Tokumoto, Y., Nomura, N., Uchiyama, H., Imura, T., Morita, T., Fukuoka, T. and Kitamoto, D. (2009) Structural characterization and surface-active properties of a succinoyl trehalose lipid produced by *Rhodococcus* sp. SD-74. *J Oleo Sci* **58**, 97-102
- 24 Salzberg, L. I. and Helmann, J. D. (2008) Phenotypic and transcriptomic characterization of *Bacillus subtilis* mutants with grossly altered membrane composition. *J Bacteriol* **190**, 7797-7807
- 25 Theilacker, C., Sanchez-Carballo, P., Toma, I., Fabretti, F., Sava, I., Kropec, A., Holst, O. and Huebner, J. (2009) Glycolipids are involved in biofilm accumulation and prolonged bacteraemia in *Enterococcus faecalis*. *Mol Microbiol* **71**, 1055-1069
- 26 Kasahara, K., Sanai, Y., Nakamura, K. and Hashimoto, Y. (2001) [Glycolipid assembly/rafts and cellular functions]. *Tanpakushitsu Kakusan Koso* **46**, 812-820
- 27 Vanier, M. T. (1999) Lipid changes in Niemann-Pick disease type C brain: personal experience and review of the literature. *Neurochem Res* **24**, 481-489
- 28 Veldman, R. J., Klappe, K., Hoekstra, D. and Kok, J. W. (1998) Interferon-gamma-induced differentiation and apoptosis of HT29 cells: dissociation of (glucosyl)ceramide signaling. *Biochem Biophys Res Commun* **247**, 802-808
- 29 Schnaar, R. L. (2004) Glycolipid-mediated cell-cell recognition in inflammation and nerve regeneration. *Arch Biochem Biophys* **426**, 163-172
- 30 D'Elia, M. A., Henderson, J. A., Beveridge, T. J., Heinrichs, D. E. and Brown, E. D. (2009) The N-acetylmannosamine transferase catalyzes the first committed step of teichoic acid assembly in *Bacillus subtilis* and *Staphylococcus aureus*. *J Bacteriol* **191**, 4030-4034

- 31 Webb, A. J., Karatsa-Dodgson, M. and Grundling, A. (2009) Two-enzyme systems for glycolipid and polyglycerolphosphate lipoteichoic acid synthesis in *Listeria monocytogenes*. *Mol Microbiol* **74**, 299-314
- 32 Wormann, M. E., Corrigan, R. M., Simpson, P. J., Matthews, S. J. and Grundling, A. (2011) Enzymatic activities and functional interdependencies of *Bacillus subtilis* lipoteichoic acid synthesis enzymes. *Mol Microbiol* **79**, 566-583
- 33 Morita, Y. S., Patterson, J. H., Billman-Jacobe, H. and McConville, M. J. (2004) Biosynthesis of mycobacterial phosphatidylinositol mannosides. *Biochem J* **378**, 589-597
- 34 Benning, C., Huang, Z. H. and Gage, D. A. (1995) Accumulation of a novel glycolipid and a betaine lipid in cells of *Rhodobacter sphaeroides* grown under phosphate limitation. *Arch Biochem Biophys* **317**, 103-111
- 35 Clejan, S., Krulwich, T. A., Mondrus, K. R. and Seto-Young, D. (1986) Membrane lipid composition of obligately and facultatively alkalophilic strains of *Bacillus* spp. *J Bacteriol* **168**, 334-340
- 36 Shemesh, T., Luini, A., Malhotra, V., Burger, K. N. and Kozlov, M. M. (2003) Prefission constriction of Golgi tubular carriers driven by local lipid metabolism: a theoretical model. *Biophys J* **85**, 3813-3827
- 37 Jorasch, P., Wolter, F. P., Zahringer, U. and Heinz, E. (1998) A UDP glucosyltransferase from *Bacillus subtilis* successively transfers up to four glucose residues to 1,2-diacylglycerol: expression of *ypfP* in *Escherichia coli* and structural analysis of its reaction products. *Mol Microbiol* **29**, 419-430
- 38 Jorasch, P., Warnecke, D. C., Lindner, B., Zahringer, U. and Heinz, E. (2000) Novel processive and nonprocessive glycosyltransferases from *Staphylococcus aureus* and *Arabidopsis thaliana* synthesize glycosylglycerolipids, glycosylphospholipids, glycosylsphingolipids and glycosylsterols. *Eur J Biochem* **267**, 3770-3783
- 39 Kiriukhin, M. Y., Debabov, D. V., Shinabarger, D. L. and Neuhaus, F. C. (2001) Biosynthesis of the glycolipid anchor in lipoteichoic acid of *Staphylococcus aureus* RN4220: role of *YpfP*, the diglucosyldiacylglycerol synthase. *J Bacteriol* **183**, 3506-3514
- 40 Xia, G. and Peschel, A. (2008) Toward the pathway of *S. aureus* WTA biosynthesis. *Chem Biol* **15**, 95-96

- 41 Jerga, A., Lu, Y. J., Schujman, G. E., de Mendoza, D. and Rock, C. O. (2007) Identification of a soluble diacylglycerol kinase required for lipoteichoic acid production in *Bacillus subtilis*. *Journal of Biological Chemistry* **282**, 21738-21745
- 42 Alley, S. H., Ces, O., Templer, R. H. and Barahona, M. (2008) Biophysical regulation of lipid biosynthesis in the plasma membrane. *Biophysical Journal* **94**, 2938-2954
- 43 Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* **37**, D233-D238
- 44 Lairson, L. L., Henrissat, B., Davies, G. J. and Withers, S. G. (2008) Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **77**, 521-555
- 45 Hu, Y. N., Chen, L., Ha, S., Gross, B., Falcone, B., Walker, D., Mokhtarzadeh, M. and Walker, S. (2003) Crystal structure of the MurG : UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 845-849
- 46 Mulichak, A. M., Lu, W., Losey, H. C., Walsh, C. T. and Garavito, R. M. (2004) Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: Interactions with acceptor and nucleotide ligands. *Biochemistry* **43**, 5170-5180
- 47 Price, K. D., Roels, S. and Losick, R. (1997) A *Bacillus subtilis* gene encoding a protein similar to nucleotide sugar transferases influences cell shape and viability. *J Bacteriol* **179**, 4959-4961
- 48 Fedtke, I., Mader, D., Kohler, T., Moll, H., Nicholson, G., Biswas, R., Henseler, K., Gotz, F., Zahringer, U. and Peschel, A. (2007) A *Staphylococcus aureus* ypfP mutant with strongly reduced lipoteichoic acid (LTA) content: LTA governs bacterial surface properties and autolysin activity. *Mol Microbiol* **65**, 1078-1091
- 49 Weart, R. B., Lee, A. H., Chien, A. C., Haeusser, D. P., Hill, N. S. and Levin, P. A. (2007) A metabolic sensor governing cell size in bacteria. *Cell* **130**, 335-347
- 50 Ahamed, T., Ottens, M., van Dedem, G. W. K. and van der Wielen, L. A. M. (2005) Design of self-interaction chromatography as an analytical tool for predicting protein phase behavior. *Journal of Chromatography A* **1089**, 111-124

- 51 Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P. and Kobe, B. (2003) Crystal structures of fusion proteins with large-affinity tags. *Protein Science* **12**, 1313-1322
- 52 Zhan, Y., Song, X. and Zhou, G. W. (2001) Structural analysis of regulatory protein domains using GST-fusion proteins. *Gene* **281**, 1-9
- 53 Donahue, J. P., Patel, H., Anderson, W. F. and Hawiger, J. (1994) Three-dimensional structure of the platelet integrin recognition segment of the fibrinogen gamma chain obtained by carrier protein-driven crystallization. *Proc Natl Acad Sci U S A* **91**, 12178-12182
- 54 Nauli, S., Farr, S., Lee, Y. J., Kim, H. Y., Faham, S. and Bowie, J. U. (2007) Polymer-driven crystallization. *Protein Science* **16**, 2542-2551
- 55 Bridges, A., Gruenke, L., Chang, Y. T., Vakser, I. A., Loew, G. and Waskell, L. (1998) Identification of the binding site on cytochrome P450 2B4 for cytochrome b(5) and cytochrome P450 reductase. *Journal of Biological Chemistry* **273**, 17036-17049
- 56 Horger, K. S., Estes, D. J., Capone, R. and Mayer, M. (2009) Films of agarose enable rapid formation of giant liposomes in solutions of physiologic ionic strength. *J Am Chem Soc* **131**, 1810-1819
- 57 Szu, P. H., Ruszczycky, M. W., Choi, S. H., Yan, F. and Liu, H. W. (2009) Characterization and mechanistic studies of DesII: a radical S-adenosyl-L-methionine enzyme involved in the biosynthesis of TDP-D-desosamine. *J Am Chem Soc* **131**, 14030-14042
- 58 Oka, T., Nemoto, T. and Jigami, Y. (2007) Functional analysis of *Arabidopsis thaliana* RHM2/MUM4, a multidomain protein involved in UDP-D-glucose to UDP-L-rhamnose conversion. *J Biol Chem* **282**, 5389-5403
- 59 Raetz, C. R. and Newman, K. F. (1978) Neutral lipid accumulation in the membranes of *Escherichia coli* mutants lacking diglyceride kinase. *J Biol Chem* **253**, 3882-3887
- 60 O'Sullivan, O., Suhre, K., Abergel, C., Higgins, D. G. and Notredame, C. (2004) 3DCoffee: combining protein sequences and structures within multiple sequence alignments. *J Mol Biol* **340**, 385-395

- 61 Ostroski, M., Tu-Sekine, B. and Raben, D. M. (2005) Analysis of a novel diacylglycerol kinase from *Dictyostelium discoideum*: DGKA. *Biochemistry* **44**, 10199-10207
- 62 Vikstrom, S., Li, L., Karlsson, O. P. and Wieslander, A. (1999) Key role of the diglucosyldiacylglycerol synthase for the nonbilayer-bilayer lipid balance of *Acholeplasma laidlawii* membranes. *Biochemistry* **38**, 5511-5520
- 63 Zhou, C. and Roberts, M. F. (1997) Diacylglycerol partitioning and mixing in detergent micelles: relevance to enzyme kinetics. *Biochim Biophys Acta* **1348**, 273-286
- 64 Li, L., Karlsson, O. P. and Wieslander, A. (1997) Activating amphiphiles cause a conformational change of the 1,2-diacylglycerol 3-glucosyltransferase from *Acholeplasma laidlawii* membranes according to proteolytic digestion. *J Biol Chem* **272**, 29602-29606
- 65 Karlsson, O. P., Dahlqvist, A., Vikstrom, S. and Wieslander, A. (1997) Lipid dependence and basic kinetics of the purified 1,2-diacylglycerol 3-glucosyltransferase from membranes of *Acholeplasma laidlawii*. *J Biol Chem* **272**, 929-936
- 66 Marechal, E., Block, M. A., Joyard, J. and Douce, R. (1994) Kinetic properties of monogalactosyldiacylglycerol synthase from spinach chloroplast envelope membranes. *J Biol Chem* **269**, 5788-5798
- 67 Marechal, E., Block, M. A., Joyard, J. and Douce, R. (1994) Comparison of the kinetic properties of MGDG synthase in mixed micelles and in envelope membranes from spinach chloroplast. *FEBS Lett* **352**, 307-310

CHAPTER 3

STRUCTURAL STUDIES ON ADHESINS FROM ENTEROHEMORRHAGIC

ESCHERICHIA COLI (EHEC) O157:H7

3.1 Background and introduction

3.1.1 Pathogenesis of *E. coli*

3.1.1.1 Pathogenic *E. coli*

The bacterium *Escherichia coli* is one of the most common bacteria found in the normal intestinal microflora of warm-blooded animals. Typically, it colonizes the gastrointestinal tract of human infants within 40 h after birth and coexists with other commensal microbes in its host within the mucous layer of the mammalian colon for decades. *E. coli* is also the workhorse model organism for molecular biology that has been extensively investigated and developed as the host of various cloning and expression vectors. However, several *E. coli* strains are highly pathogenic and can cause severe human disease.

Pathogenic *E. coli* strains have acquired specific virulence factors that allow them to infect host organisms and cause a broad spectrum of diseases. In many cases, novel combinations of virulence factors, arising through horizontal gene transfer, enable pathogenic *E. coli* strains not only to survive the selective pressures in their habitat, but also to become specific “pathotypes” of *E. coli* that are capable of causing disease in healthy individuals. There are three types of *E. coli* infections in humans: urinary tract infections (UTI), neonatal meningitis, and intestinal diseases (gastroenteritis). The diseases caused by a particular strain of *E. coli* depend on the distribution and expression of an array of virulence determinants,

including adhesins, invasins, toxins, and abilities to withstand host defenses. Among intestinal pathogens, there are six categories : enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC) and diffusely adherent *E. coli* (DAEC) [1]. The various pathotypes of *E. coli* can also be grouped into serotypes based on shared O (lipopolysaccharide), H (flagellar) and K (capsular) antigens. Historically, serotyping is important in distinguishing the small number of *E. coli* strains that cause disease and over 700 serotypes of pathogenic *E. coli* strains have been recognized so far.

3.1.1.2 Pathogenic *E. coli* serotype O157:H7

The enterohemorrhagic *E. coli* (EHEC) serotype O157:H7 first came into play as a pathogen in 1982 in a multi-state outbreak of unusual gastrointestinal illness caused by contaminated hamburgers and was subsequently isolated in 1983 [2]. Since then, it has been implicated in many outbreaks of food-borne illnesses, including acute diarrhea, hemorrhagic colitis, and hemolytic uremic syndrome, which may lead to kidney failure [3]. *E. coli* O157:H7 remains a worldwide threat to public health; in the United States, it is estimated that close to 75,000 cases of *E. coli* O157:H7 infections occur every year [4]. The severity of disease, the lack of effective treatment, and the potential for large-scale outbreaks have propelled intensive research on the pathogenesis of *E. coli* O157:H7, especially the mechanism of the infection in a hope that a comprehensive understanding of the infection pathway may lead to new developments of vaccines and treatments.

E. coli O157:H7 is notorious for its ability to cause serious food poisoning in humans and its extremely low dose required for infection [5]. As a result, *E. coli* O157:H7 has been characterized in detail in terms of pathogenesis, clinical diagnosis and detection [3]. The

toxins that are responsible for the illness are believed to be the shiga-like toxins (Stx), which is also known as verocytotoxin. It is named after the Shiga toxin originally produced by *Shigella dysenteriae* and there is evidence suggesting that *E. coli* O157:H7 obtained its pathogenesis from *Shigella* in a horizontal gene spreading event caused by bacteriophage [6]. Stx consists of five identical B subunits that are responsible for binding the holotoxin to the glycolipid globotriaosylceramide on the target cell surface, and a single A subunit that cleaves ribosomal RNA to cease the host cell protein synthesis [7]. In human, Stx is produced by *E. coli* O157:H7 attached to the colon epithelium. Stx then travels through the bloodstream to the kidney, where it damages the renal endothelial cells and causes microvasculature occlusions through a combination of direct toxicity and induction of local cytokine and chemokine production that eventually results in renal inflammation [8]. The damage can lead to hemolytic uremia syndrome, which is characterized by hemolytic anemia, thrombocytopenia, and potentially fatal acute kidney failure. Several other potential adherence factors have also been identified in O157:H7, but the significance of these factors in human disease has not been determined.

3.1.1.3 Targeting secretion systems for vaccine development

Similar to other mucosal pathogens, pathogenic *E. coli* strains use a multi-step scheme to colonize a mucosal site, evade host defenses, multiply, and damage host cells during infection [5]. Secretion systems of *E. coli* play critical roles during these processes as suggested by the significantly higher percentage of the genome encoding secretion related proteins in pathogenic *E. coli* [9] compared to the model *E. coli* K-12 strain. First of all, pathogenic *E. coli* strains need to express specific adherence factors, such as fimbriae or fibrillae, to allow them to colonize sites where commensal *E. coli* cannot. These virulence

proteins and adhesins are usually synthesized inside the bacteria and secreted by different types of secretion systems onto the bacterial cell surface, which then allows their attachment to the host cell surface [10-12]. In addition, the actual toxins and effector proteins that affect the host cell viability are also synthesized by the pathogenic *E. coli* and get secreted into the host cell through some of the same secretion systems [13-15]. The Type IV secretion system has been found to transport plasmids carrying virulence factor genes between *E. coli* strains and other pathogens that may facilitate the evolution of pathogenic *E. coli*. Therefore, targeting virulence proteins, adhesins, and their cognate secretion systems of pathogenic *E. coli* is a promising path for developing new vaccines against the severe disease caused by pathogenic *E. coli*.

3.1.2 Secretion systems in Gram negative-bacteria

Secretion in bacteria refers to the processes by which bacteria use to transport large substrate molecules, such as enzymes or toxins, across the cell membrane and envelope. It is an important aspect of bacterial physiology in the adaptation and survival to their natural environment. Moreover, it is an indispensable mechanism for bacterial pathogens to insert harmful toxins, adhesins, degradative enzymes, and other translocated effectors during the infection of a host organism.

Gram-negative bacteria have evolved a remarkable number of secretion pathways that are typically composed of multiprotein complexes forming nano-machines to facilitate the secretion process. So far, six secretion systems have been identified and characterized in Gram-negative bacterial pathogens [16]. A number of groups are working on these exquisite nano-machineries and published excellent reviews of the details of each individual system [17-24]. Here, I will briefly review the architecture and comparison of these secretion

systems in Gram-negative bacteria (Figure 3.1). Generally, these secretion systems utilize one of the two secretion mechanisms: the one-step mechanism or the two-step mechanism. Type I (T1SS), III (T3SS), IV (T4SS) and VI (T6SS) secretion systems use the one-step mechanism whereby substrates are translocated directly from the bacterial cytoplasm to the extracellular medium or into the target cell. Therefore, they are usually composed of secretion machinery that can span both the inner and outer bacterial membranes. In contrast, the Type II (T2SS) and V (T5SS) secretion systems, as well as pili assembled by the Chaperone-Usher pathway (CU) (not shown in Figure 3.1), contain protein machinery that can only transverse the outer membrane and use a two-step mechanism whereby substrates are first translocated across the bacterial inner membrane into the periplasm and then subsequently targeted to one of the secretion systems for release into the outside.

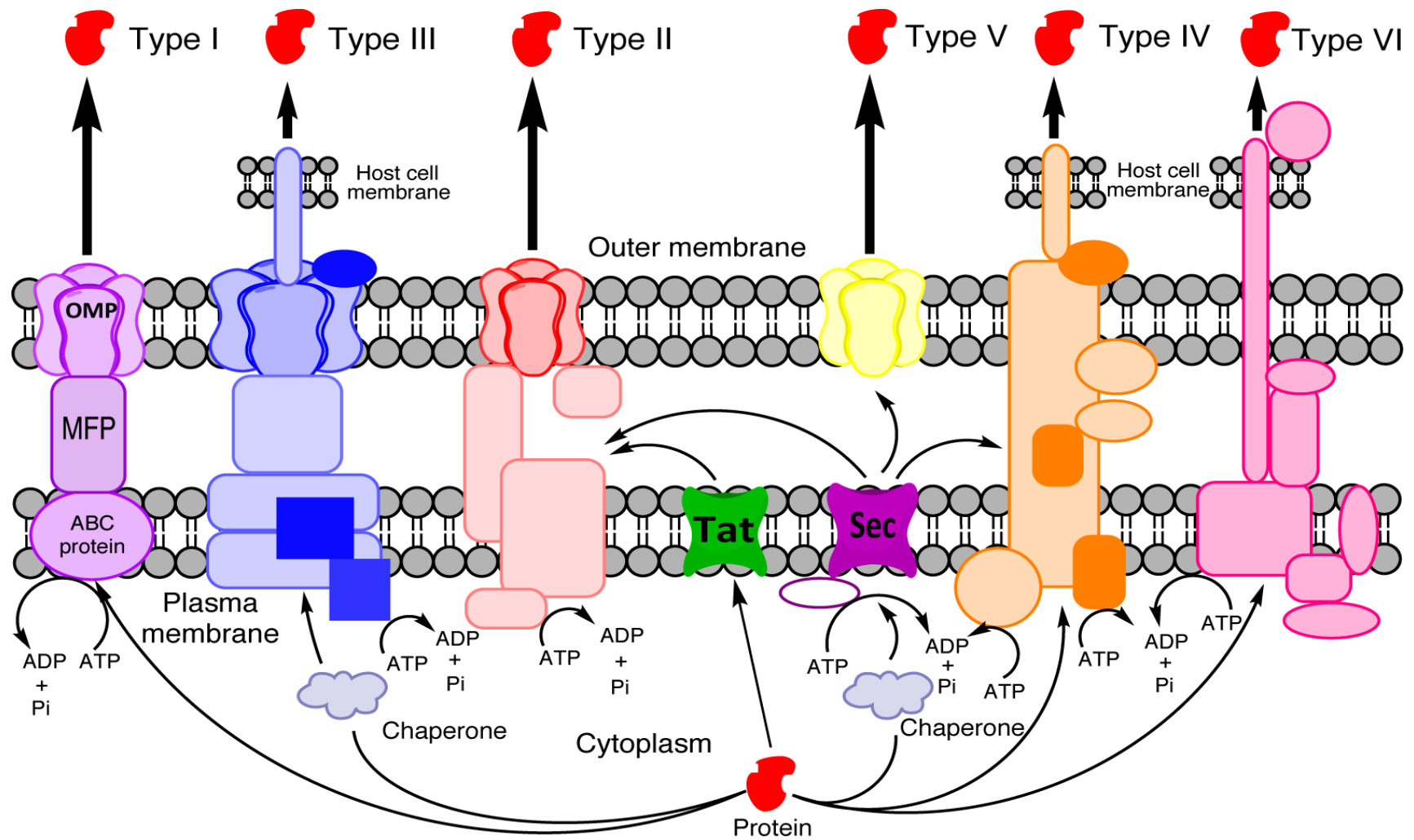


Figure 3.1 Cartoon representations of bacterial secretion systems

T1SS (purple in Figure 3.1) is well conserved among all the three kingdoms; the bacterial system is similar to the ABC transporter found in eukaryotes and composed of only three proteins: the ABC protein, the membrane fusion protein (MFP), and the outer membrane protein (OMP). However despite this simple design, T1SS secretes proteins from 10 kDa to almost 1000 kDa through both inner and outer membrane in a single step without stable periplasmic intermediates. T2SS (red in Figure 3.1) usually consists of 12-16 protein components, but only two of them are effectively located in the OM. Despite the fact that it transverses both the inner and outer membranes, T2SS is a two-step secretion system. Proteins assigned to T2SS need to be translocated by either the Tat translocon, which only transports folded proteins, or the Sec complex, which allows unfolded proteins through the inner membrane into the periplasmic space. Once in the periplasmic space, the effector protein is recognized and transported across the outer membrane by the T2SS.

T3SS (blue in Figure 3.1), also known as the injectisome for its ability of injecting effector protein directly into the host cell in a single step, has been subjected to extensive investigation because the effector proteins injected through T3SS frequently alter the basic cell function of the host cell. The basal structure of T3SS (light blue) is well conserved and composed of at least 20 proteins that transverse the inner and outer bacterial membranes as well as the membrane of host cell. A family of customized cytoplasmic chaperones is required to deliver a specific effector to T3SS and several accessory proteins, including the ATPase, will facilitate the transport of different effector proteins. In the last step, the external needle will inject the effector protein through the host cytoplasmic membrane.

T4SS (orange in Figure 3.1) is widely distributed in both Gram-negative and Gram-positive bacteria and fulfils a wide variety of functions. It not only injects effector proteins

into the host cell, but has also been found to mediate the DNA/plasmid transport among bacteria and between bacteria and host cells. T4SS usually consists of 12 components that span the inner and outer bacterial membranes. Similar to T3SS, T4SS also has an extracellular pilus composed of one major subunit and one minor subunit that is responsible for injecting effector proteins or genetic materials into the target cell. T4SS is also powered by ATP hydrolysis and three protein components on the cytoplasmic side have been found to have ATPase activity.

The newly discovered T6SS [25] (pink in Figure 3.1) also contains a needle-like structure to inject the effector proteins into a host cell, except that this long needle is attached at its base in the inner membrane and extends all the way into the host cell through two layers of membranes. Not much is known about the T6SS, but about 15-20 proteins have been suggested to be in the complex. Several new genetic and structural evidences [26, 27] also suggest that the T6SS resembles an upside-down bacteriophage tail complex.

Finally, T5SS (yellow in Figure 3.1) is the simplest and most predominant among the secretion systems described. It is also very unique in terms that (1) it is composed of a single protein located in the outer membrane and (2) it does not need ATP hydrolysis to provide the power for the translocation. More interestingly, the effector is a part of the transporter and the whole protein is translocated across the inner membrane by the Sec complex.

As mentioned above, all the secretion systems are related, to some extent, to bacterial pathogenesis, toxicity and virulence factor spreading. Intensive investigation has been focused on these secretion systems in order to understand the fundamental mechanism of bacterial-host interaction during the infection process. Since there is demand for new antibiotics for treating bacterial infections, research on the secretion systems will provide

potential targets for discovery and design of novel antibiotics. Structural studies of protein components and targets of the secretion systems offer the possibility to understand the molecular mechanisms in detail and will provide a rational framework for the design of drugs that can block these machineries to attenuate the infections caused by pathogenic bacteria that use these secretion systems as virulence factors. As the simplest system and the most widely used by a broad range of bacteria, T5SS seems to be a very promising target for research into vaccine development.

3.1.3 Type V secretion system of Gram-negative bacteria

The T5SS, also known as the autotransporter (AT) system, is the simplest and most widespread among all the secretion systems in Gram-negative bacteria. More than 700 proteins are identified to be transported to the bacterial cell surface through this secretion system. The canonical (or Type V_a) AT consists of a single protein acting both as the "translocator" at the C-terminal and the passenger domain at the N-terminal. Two additional subgroups of the T5SS have been subsequently identified and characterized, namely the two-partner system (TPS, type V_b) and the oligomeric coiled-coil adhesin system (Oca, type V_c). Depending on the function of the specific effector protein, the secreted proteins will either remain intact or be cleaved from their OM anchors after being transported. Nonetheless, they all have been predicted to protrude about 10 ~ 15 nm from the bacterial cell surface [28]. The passenger domains have been implicated in several different aspects of bacterial virulence, such as proteolytic activity, bacterial motility, adhesion, hemagglutination, serum resistance, cytotoxicity, and biofilm formation [29].

3.1.3.1 Major types of autotransporters in *E. coli*

ATs are found in all *E. coli* strains. A recent bioinformatic survey of twenty-eight selected *E. coli* genome sequences revealed a total of 215 AT-encoding sequences [30].

Partly because there were only a few pairs of TPSs identified in *E. coli*, Wells *et al.* [30] grouped the 215 AT proteins into three broad subfamilies: (1) the trimeric AT adhesins (TAA), (2) serine protease ATs of the *Enterobacteriaceae* (SPATEs), and (3) the AIDA-I type ATs. Some distinct features of each subfamily are recapitulated in the simplified carton below. More detailed descriptions can be found in papers [30, 31]. As is indicated by the name, the SPATE subfamily AT has a distinct peptidase-S6 domain that possesses a serine protease activity at the N-terminal passenger domain. Both SPATE and AIDA-I subfamily ATs have pertactin domains that can be cleaved after the transport is completed. On the contrary, the YadA (*Yersinia* adhesin) TAA subfamily does not have the pertactin domain and the passenger domain remains attached to the C-terminal translocator after being translocated. The YadA [32] and the *Haemophilus influenzae* adhesion protein Hia [33] are two well characterized TAAs that demonstrate some of the general features that also hold for the *E. coli* TAA. As the name indicates, members of this subfamily all behave as a trimer and their major activity relates to the adhesion of the pathogen to the host cell [34]. Thus the passenger domains are not cleaved after the transport is completed. A structural model has also been proposed for TAA based on data from electron microscopy and amino acid sequence analysis [35].

3.1.3.2 AIDA-I type AT and Eha proteins

The focus of this chapter is on the EHEC autotransporter (Eha) proteins, which belong to the AIDA-I subfamily of the Type V_a ATs. Detailed analysis of the structure and biogenesis of AT will be discussed in a later section with the canonical AT as the model. AIDA stands for Adhesin Involvement in Diffuse Adherence [36]. Members of the AIDA-I type ATs are all involved in biofilm formation and host cell binding. EhaA, B, C, and D were identified in the enterohemorrhagic *E. coli* O157:H7 EDL933 strain genome in 2008 and

shown to be prevalent in *E. coli* O157:H7 and a selective lab collection of Shiga toxin-producing *E. coli* (STEC) strains [31]. Besides the standard pertactin (green in Figure 3.2) and AT domain (light blue in Figure 3.2) in all AIDA-I type ATs, EhaA, C, and D have a unique proline-rich repeats (lavender in Figure 3.2) between the pertatin and AT domains. Eha A has been shown to induce cell aggregation and increases adherence of *E. coli* K-12 to primary bovine epithelial cells while Eha A, B, and D showed biofilm formation ability on polystyrene surfaces [31]. In this chapter, I took the structural biology approach to analyze the important structure features among these closely related proteins in hope that the structure information will provide more details for understanding the pathogenesis of pathogenic *E. coli* mediated through AT systems.

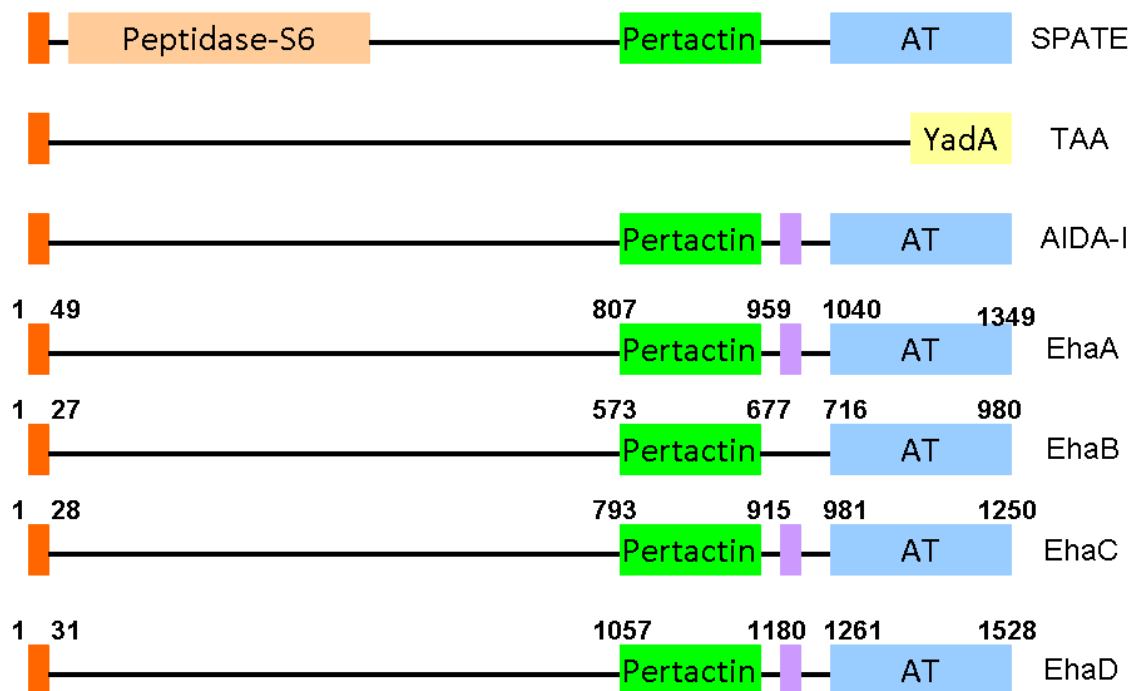


Figure 3.2 Cartoon representation of SPATE, TAA, AIDA-I subfamily of AT in *E. Coli*. EhaA, B, C, D proteins, which belong to AIDA-I are also shown in the picture. Domains are color coded as follow: orange: signal sequence; light orange: peptidase-S6; green: pertactin; lavender: proline-rich repeats; light blue: C-terminal β domain. Numbers above indicate domain boundaries.

3.1.4 The structure of V_a autotransporters

The canonical type V_a AT protein typically contains a cleavable signal sequence, an N-terminal passenger domain and a C-terminal translocator or β domain. The signal sequence on the N-terminus is used to target the protein to the Sec complex [37] and initiates translocation across the bacterial IM while the C-terminal β -barrel structure can be inserted into the OM and assist the transport of the passenger domain across the OM. The passenger domains are the ones that possess the actual enzyme activity and are usually virulence factors, adhesion proteins, and enzymes that may or may not be cleaved and released into the outside medium or inserted into the host cell once they are transported [29]. The C-terminal β -barrel is embedded in the OM and functions as the anchor and translocator of the passenger domain. However, the mechanism of how the passenger domain is transported across the OM remains debatable and will be reviewed in section 3.1.5 in detail.

While sequence alignments of type V_a ATs suggested remarkable variability of the passenger domain, several recently resolved structures [38-40] revealed some common structural features of the N-terminal passenger domains (Figure 3.3). The available evidence indicates that the structure of the passenger domains contain a repeating β -helix motif (yellow) shown in Figure 3.3, that produce a quite elongated structure. Comprehensive sequence analysis and available structures suggested that in most canonical autotransporters, the passenger domain is likely to contain substantial β -helical structure. Given that the β -helical structure is pretty uncommon among protein structures, the prevalence of this structure in the AT family suggests a certain important role of this structure in the function of ATs. In general, the canonical autotransporters seem to use the β -helix as the scaffold for the passenger domain, upon which various loops and domains are appended (Figure 3.3)

that mediate function in each unique passenger domain. Many TAAs also share similar overall architectures with the type V_a ATs. For example, in the trimeric autotransporter YadA structure, the passenger domain forms a β -roll that is composed of three long β -sheets [34]. This also reiterates the importance of the basic β -helix fold in the overall structure of the ATs.

Unlike the highly diverse passenger domains, the C-terminal translocators are very similar in size and overall structure, which are composed of the β barrel porin domains that transverse the bacterial OM. Instead of forming a hollow channel like most of the porin proteins, the C-terminal β barrel of autotransporters usually contains one or more helices inside the channel (magenta α helix in Figure 3.3), which is the extension of the passenger domain and links the two domains together. In cases where the passenger domains will be cleaved and released, the cleavage site is usually presented in this linker region.

3.1.5 The structural implications of autotransporter biogenesis

ATs are synthesized in the cytosol and translocated across the IM via the Sec pathway; the most ATs have a 20-30 amino acid peptide signal that is typical for Sec-dependent proteins [41-43]. After being translocated into the periplasmic space, the β -domain is inserted into the OM and the passenger domain is translocated across the OM. However, the mechanism of the insertion and translocation remain debatable. Until now, three hypotheses have been proposed (shown in Figure 3.4), although none of them completely fit all of the experimental data.

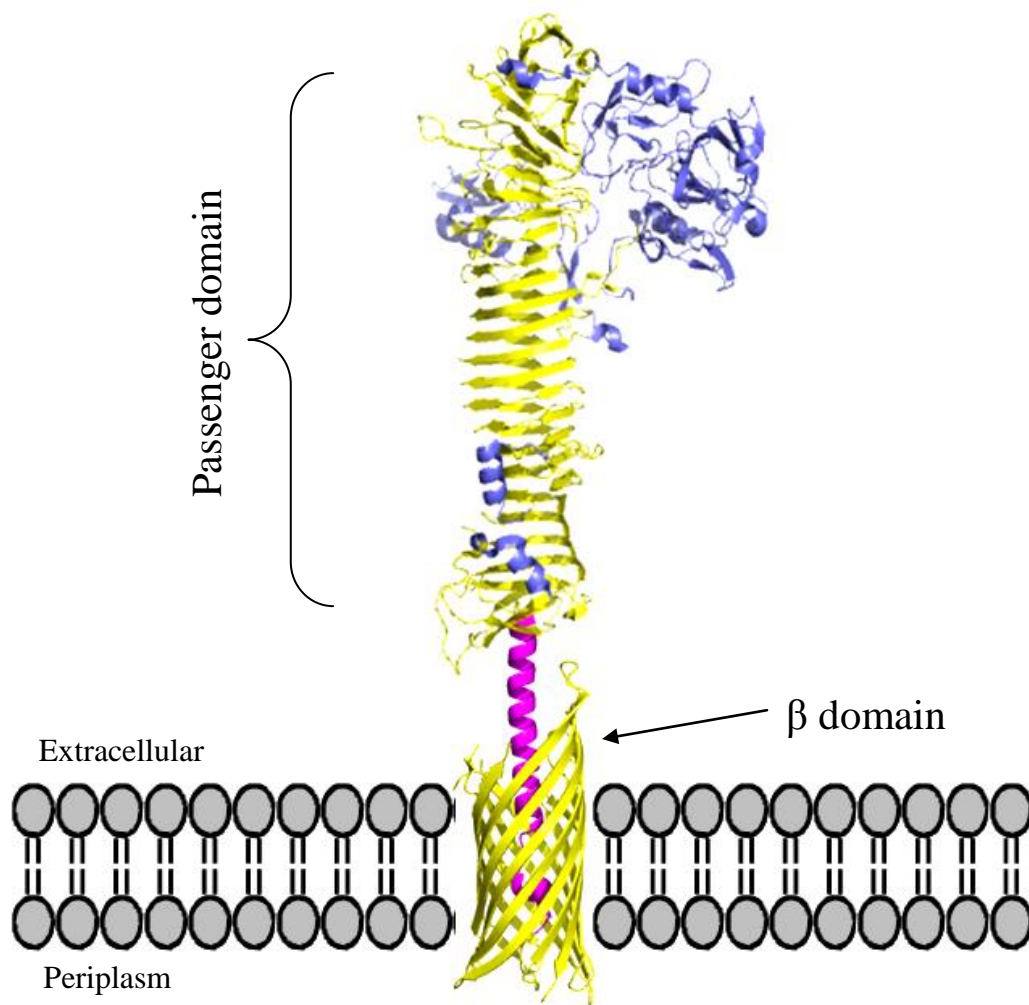


Figure 3.3 A cartoon representation of a classical type V_a autotransporter structure. The structures of the passenger and β domains are taken from *E. coli* Hbp fragment structures PDB 1WXR and PDB 3AEH, respectively. Structure of the full length protein SPATE AT is speculative and is not to scale.

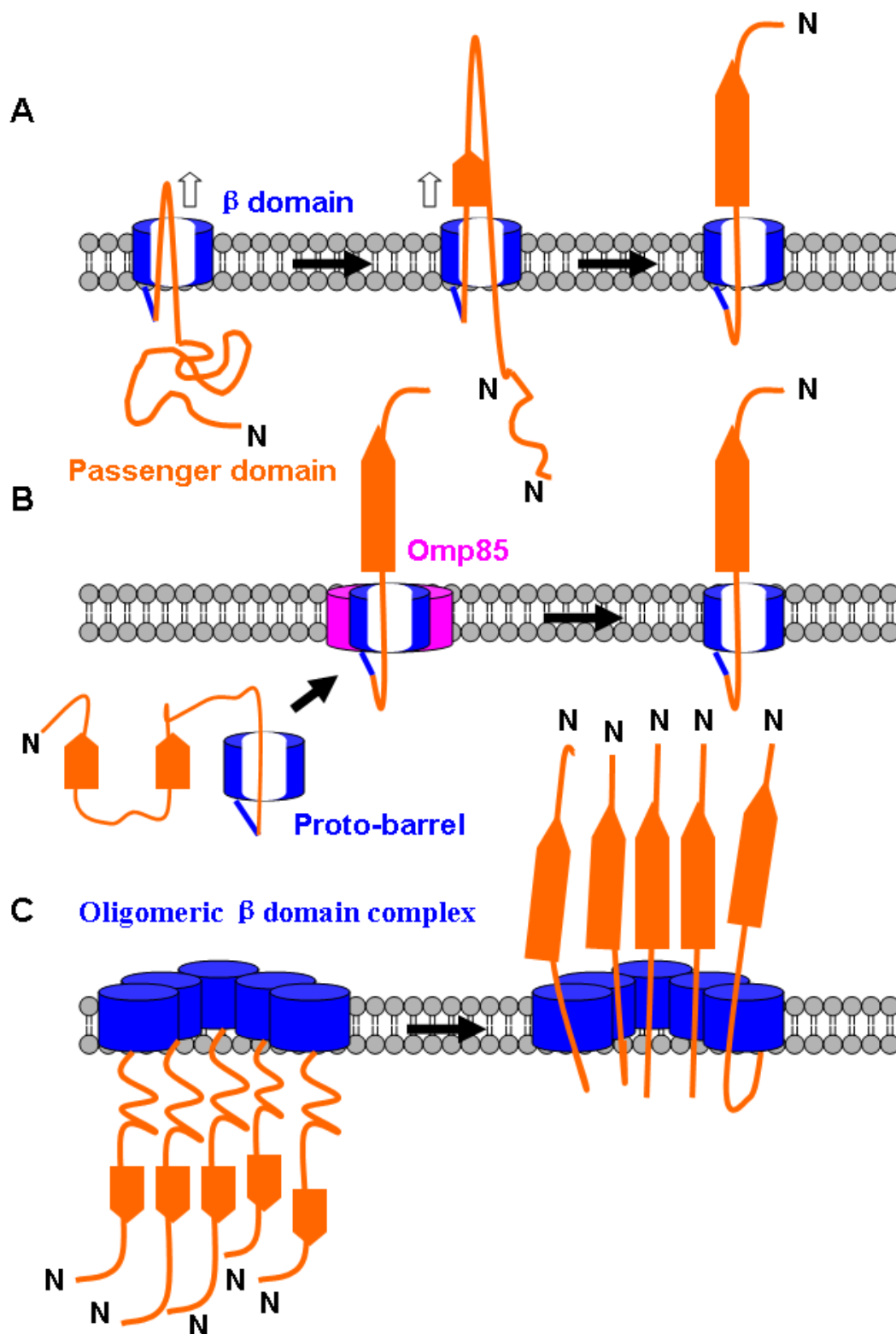


Figure 3.4 Proposed mechanisms of biogenesis of AT. (A): hairpin model; (B): Omp85 model; (C): multimer model. The β -domain is colored in blue and the passenger domain is colored in orange. The Omp85 complex is represented by a pink barrel.

The hairpin model (Figure 3.4A) was first proposed based on the observation that N-terminally truncated passenger domains, as well as passenger domain chimeras, can be effectively translocated across the OM. Therefore, the passenger domain is not threaded through the channel formed by the β -domain from the N-terminus. Alternatively, the hairpin model proposed that the C terminus of the passenger domain initiates the translocation by forming a hairpin structure inside the β domain and this hairpin is maintained during the translocation process while proximal segments of the polypeptide are threaded through the pore in a linear fashion. Available structural information also supported this model in that the diameter of the β -domain of NalP is only wide enough to accompany two polypeptide strands, if they were in a completely extended conformation [44], while the β -domain of TAA Hia can accommodate three hairpins composed of two extended polypeptide chains [45]. However a recent molecular dynamic simulation showed that the pore of NalP would never be wide enough to accommodate a polypeptide chain that contains any tertiary structure [46] , meaning that the passenger domain will have to fold outside the cell completely and the α -helical conformation of the small segment observed inside the β -domain and the hydrogen bonding network between the α -helix and the interior of the β barrel would have to form after the translocation is completed. Also, the diameter measured on the structure of Hia [45] suggested a very tight fit for the three potential hairpins inside the β -domain, which would probably make any movement during the translocation process sterically difficult.

The Omp85 model (Figure 3.4B) proposed that the passenger domain is translocated by an external translocase, potentially the Omp85 protein, instead of the β -domain of itself. Omp85 protein was suggested to be a candidate because it plays an essential but poorly understood role during the insertion of β -barrel proteins, such as the β -domain of ATs [47,

48]. Therefore, in this model, both the passenger domain and the β -domain are translocated and inserted by Omp85 in an at least partially folded state. The β -domain is still essential for targeting the passenger domain to the OM and probably during the initiation of the translocation, but does not play actual catalytic role during the translocation. This model accounts for the observations that the β -domain seems to be pre-folded in the periplasm [49] and small, folded polypeptide segments can be secreted apparently intact [50]. It can also explain the small diameter of the pore formed by the β -domain and avoids the problem of the hydrogen bonding in the hairpin model because the whole structure of the autotransporter is folded simultaneously with the help of the Omp85 complex. However, the Omp85 model does not explain the necessity of the α -helical segment inside the β -domain, if it is not the initiation point of the translocation of the passenger domain.

An alternative multimer model for AT translocation was proposed based on the study of the IgA protease [51]. In this model (Figure 3.4C), a channel wide enough to accommodate a folded (or partially folded) polypeptide chain is formed by an oligomer assembly of ATs, which facilitates the transport of several passenger domains. However, this model was proposed solely based on the observation that the C-terminus of IgA protease purifies as a large oligomeric complex that forms striking ring structures [51]. In another words, this model may be an exception that does not represent the majority of the ATs, even if it is correct. On the other hand, it is unclear whether the multimer is formed before or after the transport of the passenger domain. In the latter case, the oligomerization may play an entirely different role in the function of the IgA protease. Another fundamental flaw of this multimer model lies in the intrinsic structural property of the β -barrel protein. The β -barrel structure forms an aqueous channel in the lipid membrane by placing the hydrophobic residues outside

the channel to interact with the hydrophobic lipid environment while keeping the inside hydrophilic with hydrophilic residues lining the interior of the channel. However, in the multimer model, the presumed aqueous channel required for translocation would require that the β -barrel domains of the multimer to "unzip" to create this macro-pore. After translocation of the passenger domains, the β -barrel domains would then have to zip up again to enclose the α -helical linker. Unless further experiments can prove it in detail, this hypothesis is the least supported by the evidence.

Besides the mystery of the mechanism for passenger domain transport, another important aspect of the type V secretion system remains enigmatic -- the driving force of the translocation. Unlike other secretion systems that utilize the energy released by ATP hydrolysis, type V ATs do not utilize any ATPase activity; the periplasmic space is also devoid of ATP. Moreover, the OM is devoid of electrochemical gradient that can be used as an energy source because OM porins permit free diffusion of small ions, which renders the OM leaky. One possibility is that the electrochemical potential maintained by the IM is somehow coupled and harvested by the ATs located in the OM. However, unlike other secretion systems that have a structure which transverses and couples both membranes, no evidence has been found to support this hypothesis in the type V secretion system.

3.1.6 Research aims on Eha proteins

The main aim of the study in this chapter is to further characterize a sub family of AIDA-I type ATs by using structural, biochemical and biophysical techniques and to gain insight into the mechanism of protein translocation across the outer membrane. The foci of this chapter are: (1) cloning, expressing, and purifying the Eha adhesins, as full-length proteins, as domain fragments, and as chimeras in an effort to characterize their structure-

function relationships; (2) Efforts were made to crystallize the different Eha protein constructs, and crystals were obtained; (3) The structure of EhaB translocation domain was determined and its biophysical behavior was characterized.

3.2 Experimental procedures

3.2.1 Materials

Restriction enzymes and T4 ligase were purchased from New England Biolabs. The pET-28b and pET-22b vectors were purchased from Novagen. Detergents Zwittergent 3-14 and 3-12 were purchased from Anatrace. Crystal Screen I & II were purchased from Hampton Research. Cryo Screen I & II were purchased from Emerald Biosystems. MemGold and Memsys & Memstart were purchased from Molecular Dimensions.

3.2.2 Expression and purification of Eha proteins

The full-length coding sequences of *ehaA*, *ehaB*, and *ehaD* were cloned from *E. coli* O157:H7 EDL933 genomic DNA from the STEC Center (courtesy of Dr. Linda Mansfield, MRU/STEC Principal Investigator). The gene fragment *ehaB_c* (translocation domain, 678-980) was cloned into pET-28b vector between restriction sites NcoI and XhoI. An additional *EhaB_c* construct was created: pRMG_*EhaB_c*_His6 containing the native signal sequence (1-27), the *ehaB_c* translocation domain (678-980), and the introduction of two intervening restriction sites NdeI and SpeI. A final *ehaB* expression construct was created in pET-28b containing the native signal sequence (1-27), the *ehaB* passenger domain (57-573), the *ehaB* pertactin domain (574-677), and the *ehaB_c* translocation domain (678-980). This modified full length version of *ehaB*, designated as pRMG_*EhaB_fl*_His6, differed from the native coding sequence by the introduction of unique restrictions sites between all the putative structural domains. For *ehaA* and *ehaD*, plasmids designated pRMG_*EhaA_c*_His6,

pRMG_EhaA_fl_His6, and pRMG_EhaD_c_His6, were constructed in a similar way, except utilizing the NcoI or NdeI sites on the 3'-end and XhoI at the 5'-end and ligated into a modified pET22b vector. Also, construct pRMG_EhaA_fl_His6 included the passenger and translocation domains to express an intact type V_a autotransporter.

The expression of the Eha translocation domains from *ehaA*, *ehaB*, and *ehaD* were done by transforming the appropriate plasmid into *E. coli* BL21 (DE3) strain. A single colony selected from an LB plate containing 100 µg/mL of ampicillin was scaled up into 1 L of liquid LB media culture and grew at 37 °C until the OD₆₀₀ reached 1.0. Protein expression was induced with 0.5 mM IPTG and the cells were grown at 37 °C for 4~6 h before harvesting. The cell pellets were re-suspended and broken by sonication. In all three cases, the Eha translocation domain proteins were enriched in the inclusion body fraction.

The Eha translocation domains were purified and refolded from inclusion bodies by using a procedure developed from methods described by Buchanan et al. [52] and Gellman and colleagues [53]. The inclusion body fraction was washed twice in 20 mM Tris-HCl, 100 mM NaCl, 1 mM β-EtSH (pH 8.0) with 0.1% deoxycholate added and centrifuged. The washed inclusion bodies were then denatured and solubilized in freshly-made 20 mM Tris-HCl, 100 mM NaCl, 8 M urea, 1 mM β-EtSH (pH 8.0). The denatured protein solution was subjected to sonication and centrifugation at 16,000 × g for 20 min at 4 °C. The supernatant was mixed with an equal volume of 10% detergent zwittergent 3-14 (Z3-14) solution and refolded by a drop wise dilution method in the presence of 0.05 % detergent Zwittergent 3-14 in buffer A (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 2 mM EDTA, 5 mM CaCl₂). Refolded Eha translocation domains were then partially purified by Q Sepherose® fast flow anion exchange chromatography and eluted by a salt gradient up to 1 M NaCl. The eluted

fractions were pooled and subsequently loaded onto a gravity flow Ni-NTA (Qiagen) chelation affinity column pre-equilibrated with buffer B (30 mM Tris-HCl pH 8.0, 200 mM NaCl, 0.02 % Z3-14). The Eha translocation domains were eluted with buffer B supplemented with 150 mM imidazole. The eluted protein was subjected to either gel filtration chromatography (Superose 6) or a second anion exchange chromatography before being concentrated to 10 mg/ml with a 50 kDa MWCO concentrator (Millipore).

The pRMG_EhaB_fl_His6 was transformed into *E. coli* BL21 (DE3) strain. *E. coli* were grown at 37 °C with shaking at 200 rpm until OD₆₀₀ reached 1.0. Expression of recombinant EhaB full length protein was induced by adding 0.1 mM IPTG to the cell culture and the cultures were grown at 20 °C for an additional 16 h. Cells were harvested by centrifugation and stored at -70 °C. The cell pellet was thawed and re-suspended in buffer C (30 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM β-EtSH, 0.1 mM EDTA); the cells were broken by two passes through Emulsiflex -C3 at 20,000 psi. Cell debris was discarded after centrifugation at 12,000 × g for 20 min at 4 °C, and the supernatant with the cell membranes were centrifuged at 170,000 × g for 1 h at 4 °C. The membrane pellet was re-suspended in buffer A and homogenized before the BCA protein assay. The detergent β-decyl maltoside (C₁₀M) was added to 1 % (w/v) to a protein concentration of 10 mg/mL. The solution was mixed and incubated for 20 min at 4 °C and any insoluble material was discarded after the second round of ultracentrifugation at 170,000 × g for 1 h at 4 °C. The supernatant containing EhaB-detergent complexes was immediately loaded onto a pre-equilibrated gravity flow column containing 15 mL Ni-NTA agarose slurry. The column was washed extensively with buffer D (30 mM Tris-HCl pH 8.0, 300 mM NaCl, 25 mM imidazole, 0.17 % C₁₀M) to remove any non-specific binding proteins until the OD_{280nm} reading was below 0.08. Full

length EhaB was eluted with Buffer E (30 mM Tris-HCl pH 8.0, 100 mM NaCl, 200 mM imidazole, 0.17% C₁₀M). The elution was then subjected to Hi-Trap Q anion exchange chromatography (column volume: 1mL) to further purify the target protein. The enriched fractions of EhaB full length protein were pooled and concentrated to 8 mg/mL by a 100 KDa MWCO concentrator (Millipore). The purity of the final protein solution was assessed by SDS-PAGE analysis. The EhaA full length protein was similarly purified as the EhaB full length protein except that Z3-12 was used instead of C₁₀M.

3.2.3 Crystallization and cyroprotection of Eha proteins

Crystallization was done by vapor diffusion using the sitting drop technique. In order to find initial crystallization conditions, several commercially available screens were used according to a standard sparse matrix method, including Crystal Screen I & II, Cryo Screen I & II, MemGold, and Memsys & Memstart. Preliminary screenings were set up by using a Gryphon[®] crystallization robot (ArtRobins) to mix the protein solution with the screening solution in three different volume ratios (e.g. 200 nL : 200 nL, 450 nL: 150 nL, 150 nL : 450 nL). Optimization of the initial hits was carried out around the original condition by varying precipitant concentrations, the buffer pH and types of additives. Once promising crystallization conditions were identified, the larger microliter scale experiments were set up manually to produce crystals of better quality. For most crystallization experiments, the crystallization trays were incubated at 20 °C.

All the crystals obtained were treated before all diffraction experiments by the addition of cryoprotectants to the mother liquor. The final cryoprotectant concentrations varied depending on the compound used: ~32 % (v/v) for methylpentanediol (MPD), ~18 % (v/v)

for ethylene glycol, or ~2.5 M sodium malonate. Crystals were then flash-frozen by rapidly dunking into liquid nitrogen before data collection.

3.2.4 X-ray diffraction data collection and structure determination

The resulting crystals from the purified adhesins were characterized by X-ray diffraction collected at 100 K at LS-CAT 21-ID-G and 21-ID-F beamlines at the Advanced Photon Source, Argonne National Laboratory. When possible, complete sets of X-ray data images were processed and scaled with HKL2000 [54] and analyzed with Xtriage from the Phenix [55] program suite (version 1.6.2). For EhaA_c, EhaB_c, and EhaD_c, initial molecular replacement trials was done by using NalP (PDB 1UYN) or EspP (PDB 2QOM) as starting models in Phaser [56]. For EhaB_c, one solution, derived from EspP, gave reasonable initial R-factor and R-free values in Phaser. Inspection of the initial $2F_o - F_c$ electron density maps revealed continuous density for the α -helical linker that was not part of the phasing model. After several iterative rounds of model building and refinement with Phenix, the EhaB-c structure from residues 682-980 (with only 3 missing extracellular loops) was resolved at 2.2 Å resolution. Coot was used for manual model building [57] with iterative refinement cycles performed in Phenix Refine. The TLS refinement picking was performed in Phenix. The structure quality was assessed with Molprobit [58].

3.2.5 Planar bilayer lipid membrane (BLM) assay for ion channel activity

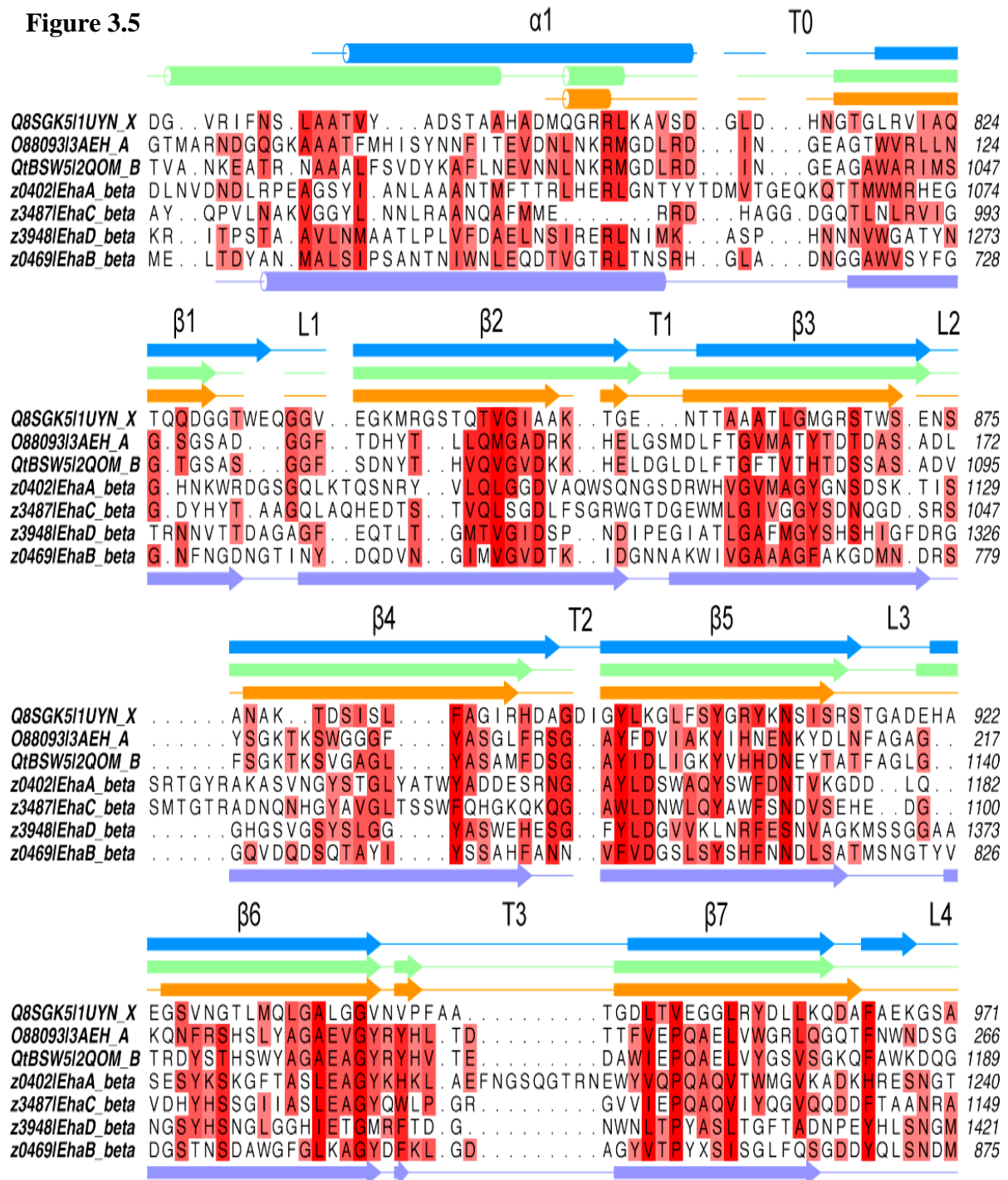
The ion channel activity of the EhaB_c was measured using the BLM technique [59]. Diphytanoylphosphatidylcholine (DPhPC) (20 mg/mL) was dissolved in hexane, and small aliquots were introduced on a 1 mm aperture in the thin Teflon vial separating two compartments containing 1 M KCl in 20 mM HEPES, pH 7.4. Ag–AgCl electrodes were

used to provide electrical contact between the buffer and the operational amplifier (Axopatch 1D amplifier, Axon instrument). In the absence of the protein, the conductance of the lipid membrane was 1–3 pS. The membrane protein to be examined and/or chemicals were added into either compartment under constant stirring. Experiments were carried out at 22–24 °C. Purified and concentrated EhaB_c was first incorporated into C₁₂E₄ micelles by simple mixing and incubation at 4 °C for 48 h and then painted into planar BLM by placing a glass micropipette in contact with the membranes. Single or several channels of the same types can be isolated for channel activity recording.

3.2.6 Multiple sequence alignment of β domain of Eha proteins

The sequences of the β -domain of EhaA, B, C, and D are aligned with NalP, Hbp (N1100D), and EspP with T-Coffee [60]. Figure 3.6 shows the result of the alignment with light blue representing the secondary structure of NalP, light green representing the secondary structure of Hbp, orange representing the secondary structure of EspP on top of the sequences and the lavender representing the secondary structure of EhaB at the bottom of the sequences. The figure was prepared using the program ALINE [61].

Figure 3.5



3.3 Results and Discussion

3.3.1 Purification of Eha full length autotransporter proteins and translocation domains

The expression of Eha full length proteins at 20 °C gave rise to a higher percentage of protein in the membrane fraction. Both EhaA_{fl} and EhaB_{fl} have been obtained in high purity by using Z3-12 and C₁₀M detergents, respectively. The SDS-PAGE analysis showed the EhaA_{fl} protein has a MW of 140 kDa (Figure 3.6A) and the EhaB_{fl} has a MW of 110 kDa (Figure 3.6C). An additional step of purification (i.e., anion exchange chromatography) was very useful to remove many contaminant or potential interacting proteins (Figure 3.7). The final protein yield per liter cell culture was 1 mg for EhaB and 2 mg for EhaA.

Three translocation domain fragments, EhaA_c, EhaB_c, and EhaD_c, were all purified from the inclusion bodies after denaturation and detergent-assisted refolding procedures [52, 53, 62], which has been successfully used to refold and purify many outer membrane porins. To increase the monodispersity of the refolded proteins, gel filtration studies were carried out to eliminate any misfolded aggregates. Figure 3.6B showed that the peak fraction of EhaA_c eluted from a Superose 6 gel filtration column. An unknown protein band with a MW of 49 kDa was present in both EhaA_c and EhaD_c protein solutions. An additional step of anion exchange chromatography was also useful for Eha translocation domain proteins (data not shown). Interestingly, the refolded β -barrel domain is very stable. The Figure showed no pronounced degradation of EhaA_c, EhaB_c, and EhaD_c after half a year of storage at 4 °C (Figure 3.6D).

3.3.2 Crystallization of EhaA and EhaB full length proteins

In order to elucidate the structural basis of the adhesion and the biogenesis of the autotransporters, crystallization of the EhaA_{fl} protein was attempted. Small crystals of EhaA_{fl} were grown by mixing 3 μ L of purified EhaA_{fl} protein (7~8 mg/mL) with 1 μ L of the following well solution: 0.2 M NaAc, 25.5% PEG 4000, 0.1 M Tris-HCl pH 8.5. Crystals of the bi-pyramidal shape with size of 15 x 15 x 15 μ m appeared after 4 weeks of incubation. Unfortunately, these small crystals not only took a month to grow, but also only diffracted to a resolution of 10 Å at best. Indexing the diffraction data with HKL2000 suggested a C2 space group.

The growth of EhaB_{fl} crystals remained elusive until recently. The protein was purified through an additional step of the ion exchange chromatography. Small crystals of cubic shape and bi-pyramidal shape appeared after several months of incubation in two very similar conditions where PEG 200 was the major precipitant: (1) 30% (v/v) PEG 200, 0.2 M (NH₄)₂HPO₄, 0.1 M Tris-HCl pH 8.5; (2) 30% (v/v) PEG 200, 0.2 M (NH₄)₂SO₄, 0.1 M CAPS pH 10.5. Cube-shape crystals have size of 30 x 30 x 30 μ m and bi-pyramidal shape crystals with size of 15 x 15 x 15 μ m. Unlike EhaA_{fl} crystals, the EhaB_{fl} crystals were birefringent in cross-polarized light and gave all indications that they are genuine EhaB_{fl} protein crystals. However, more characterization of these crystals is needed, including their X-ray diffraction behavior.

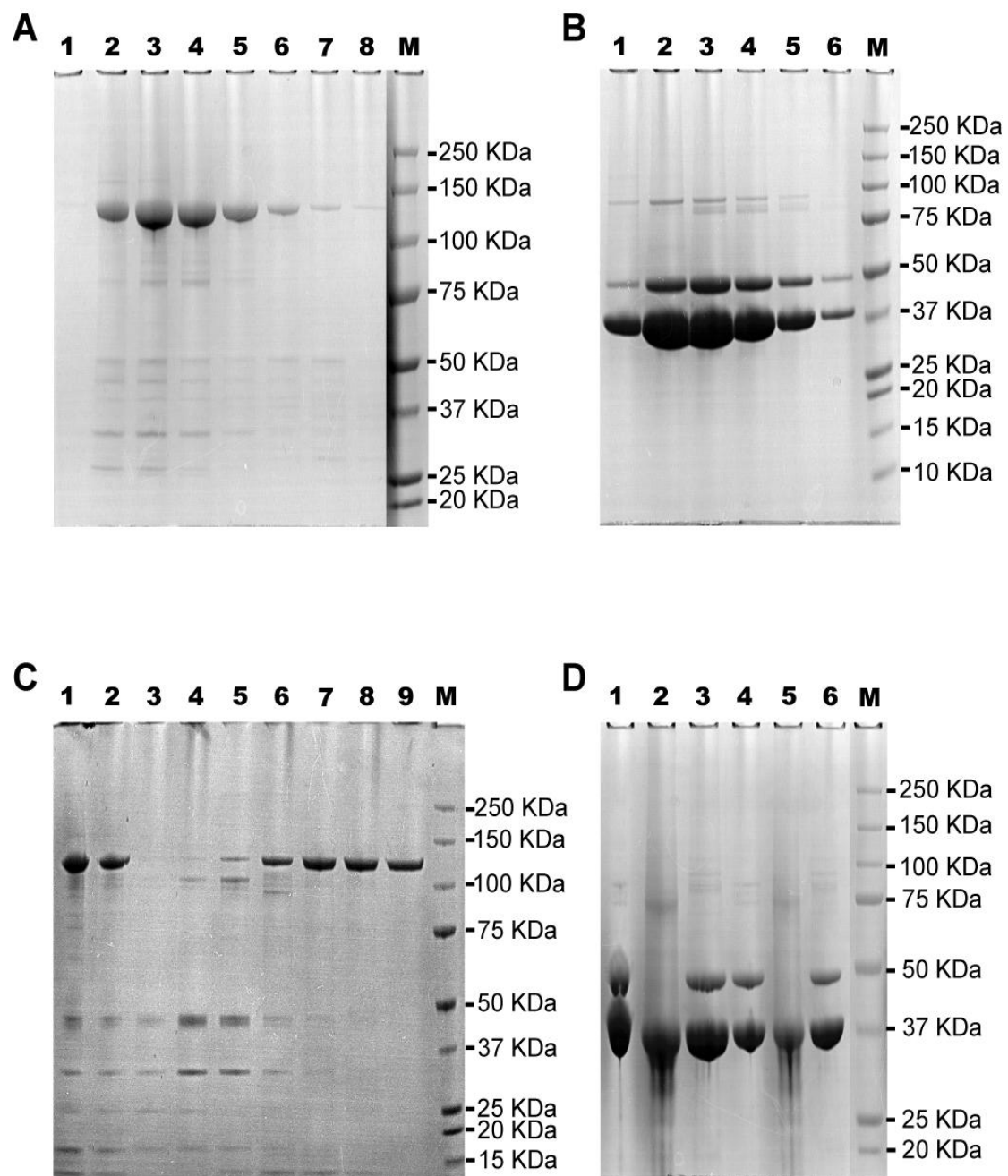


Figure 3.6 SDS-PAGE analysis of purified Eha proteins. (A) Elution fractions of EhaA_{fl} protein from Superose 6 gel filtration. Lanes 1-8, elution fractions 8-16. (B) Elution fractions of EhaA_c protein from Superose 6 gel filtration. Lanes 1-6, elution fractions 13-18. (C) Elution fractions of EhaB_{fl} protein after Hi-Trap Q ion-exchange chromatography. Lanes 1 & 2 fractions 2 & 3, lanes 3- 9, fractions 12, 15, 18, 21, 24, 27, 30, (every third fractions from 10-30). (D) Eha proteins after storage at 4 °C for half a year. Lanes 1-3 & 4-6, 5 μ L and 2.5 μ L, respectively, of EhaA_c (10 mg/mL), EhaB_c (8 mg/mL), and EhaD_c (13 mg/mL). All fractions are 1 mL/fraction. Lane M stands for the protein molecular weight standards.

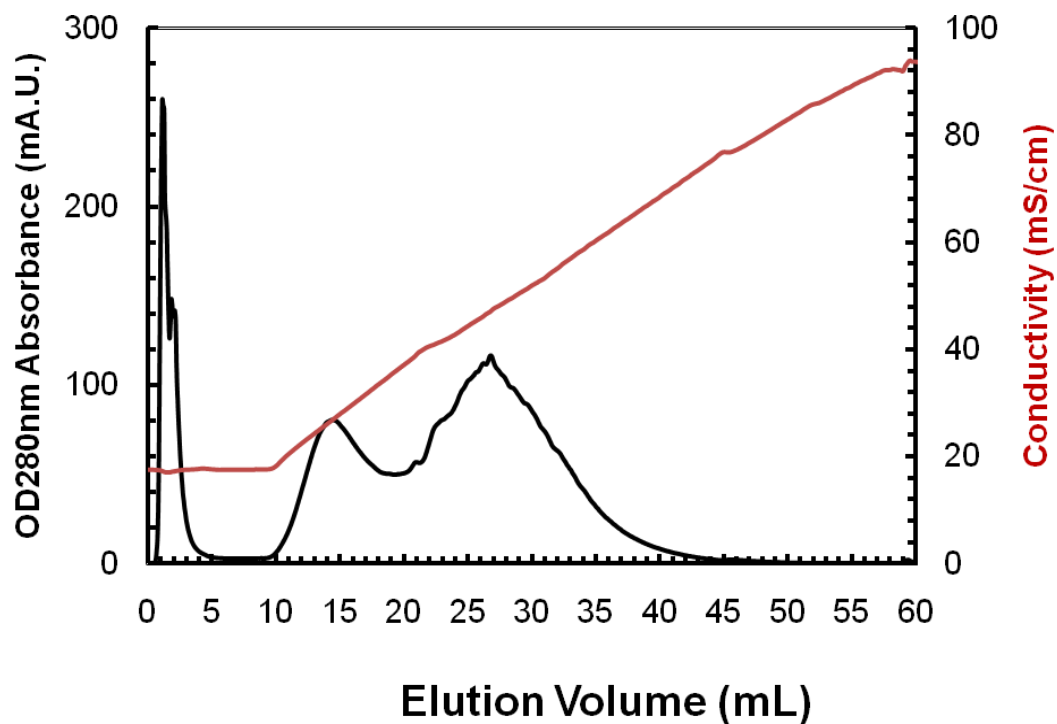


Figure 3.7 Elution profile of ion-exchange chromatography of EhaB_{fl}. SDS-PAGE analysis of the fractions is shown in Figure 3.6 C.

3.3.3 Crystallization of Eha translocation domains

The crystallization of the refolded Eha β -domain proteins in detergent Z3-14 or Z3-12 proved to be quite straightforward. Initial screens revealed a variety of conditions where crystals grew. The three major precipitants that best promoted crystal growth of Eha β -domains are PEG 400, AS, and MPD. Interestingly the growth time of the crystals is positively correlated to the diffraction power of the crystals. Crystals of the Eha β -domain proteins appeared overnight in the PEG 400 conditions, ~4 days for those grown in AS, and 2 ~ 4 weeks for those grown in MPD. The best diffraction was observed in crystals grown in the MPD conditions. Crystals of EhaA_c exhibiting excellent X-ray diffraction were grown by mixing 2 μ L of purified EhaA_c protein (10 mg/mL) with 2 μ L of one of the two following well solutions: (1) 0.2 M magnesium acetate, MPD 30 % (v/v), 100 mM sodium cacodylate pH 6.5; (2) 1.84 M AS, 5% (v/v) iso-propanol. Hexagonal plate shaped crystals with a size of 30 x 30 x 10 μ m appeared in condition 1. The rod shape crystals could grow in size up to 40 x 40 x 200 μ m in condition 2. Crystals of EhaB_c were grown by mixing 2 μ L purified EhaB_c protein (5~6 mg/mL) with 2 μ L of the following well solution: MPD 23~28 % (v/v), PEG 8000 5% (w/v), and 100 mM HEPES at pH 6.99. Rod shaped crystals with a size of 25 x 25 x 100 μ m appeared after 2 ~ 3 weeks of incubation. The best crystals of EhaD_c were grown by mixing 2 μ L of purified EhaD_c protein (13 mg/mL) with 2 μ L of the following well solution: 30% MPD, 0.05 M zinc acetate, and 0.1 M sodium cacodylate at pH 6.5. Hexagonal plate shaped crystals with a size of 30 x 30 x 10 μ m appeared after 4 weeks of incubation. Representative crystal pictures are shown in Figure 3.9. One dataset at 2.0 Å resolution was collected for the EhaB_c protein at beamline 21-ID-G. Similarly datasets of EhaA_c and EhaD_c were collected at resolutions of 2.35 Å and 2.45 Å, respectively. The

structure was solved by molecular replacement by using the NaIP β domain (PDB 1UYN) as the search model. The data collection and refinement statistics are summarized in Table 3.1.

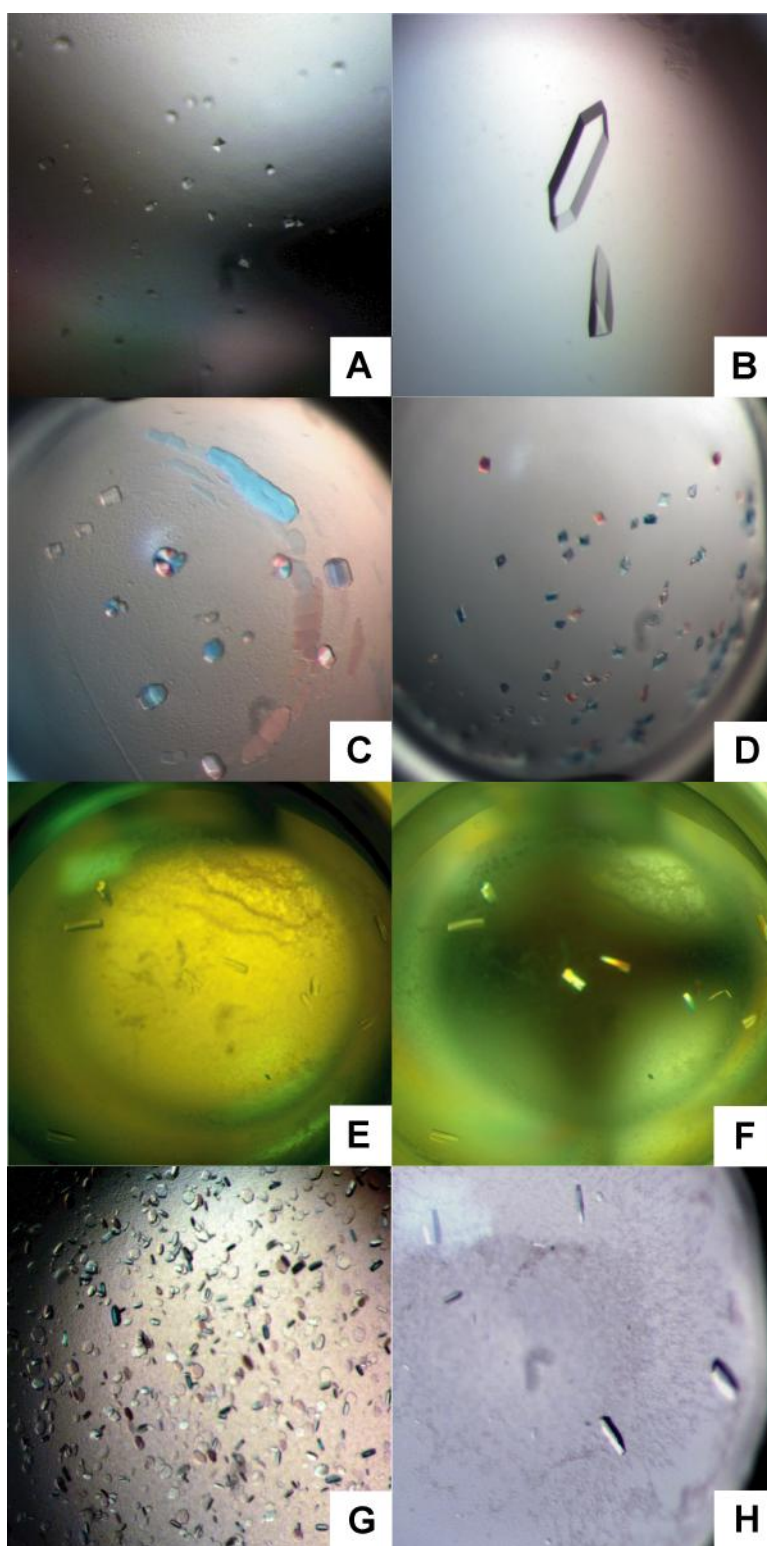


Figure 3.8 Representative pictures of Eha protein crystals. (A): EhaB_{fl} crystals (grown from 25.5% PEG 4000, pH 8.5). (B) EhaA_c crystals (grown from 1.86 M AS and 5% isopropanol). (C) EhaB_{fl} crystals (grown from 30% PEG 200, pH 8.5). (D) EhaB_{fl} crystals (grown from 30 % PEG 200, pH 10.5) (E&F) EhaB_c crystals (grown from 28% MPD and 5% PEG 8000, pH 6.99). (G&H) EhaD_c crystals (grown from 30% MPD, pH 6.5).

Table 3.1 Data collection, phasing and refinement statistics for Eha structures.

	EhaB_c	EhaA_c	EhaD_c
Data collection			
Space group	C 2	P 4 2 2	P 3 2 1
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	127.1 103.2 62.7	147.1 147.1 170.9	137.1 137.1 51.6
α , β , γ (°)	90.0 99.2 90.0	90.0 90.0 90.0	90.0 90.0 120.0
Resolution (Å)	50-2.20 (2.28-2.20) ¹	50.00-2.45 (2.49-2.45)	50.00-2.35 (2.39-2.35)
<i>R</i> _{sym}	0.04(0.19)	0.11(0.47)	0.082(0.34)
<i>I</i> / σI	21.7(4.5)	23.7(1.7)	24.4(4.3)
Completeness (%)	97.8(87.1)	99.7(95.0)	94.3(75.0)
Redundancy	4.1(3.5)	13.7(6.6)	10.2(7.3)
Refinement			
Resolution (Å)	24.03-2.20	-	-
No. reflections	40343	-	-
<i>R</i> _{work} / <i>R</i> _{free}	0.222/0.252	-	-
No. atoms:		-	-
Protein	4200	-	-
Ligand/ion	234	-	-
Water	49	-	-
<i>B</i> -factors		-	-
Protein	-	-	-
Ligand/ion	-	-	-
Water	-	-	-
r.m.s deviations			
Bond lengths (Å)	0.008	-	-
Bond angles (°)	1.254	-	-

¹ Values in parentheses are for highest-resolution shell.

3.3.4 The crystal structure of the EhaB translocation domain (EhaB₆₇₈₋₉₈₀)

Crystals of the EhaB_c protein were obtained in space group *C2*, with two copies in the asymmetric unit. The solution derived from EspP as the phasing model underwent several iterative rounds of model building and refinement with the data in the resolution range of 2.2 ~ 25 Å. Although no NCS restraints were applied, the two copies of the protein are essentially identical. The final refined structure has a high average B factor of 68, after TLS restraints were applied. The overall structure of EhaB_c (Figure 3.9) comprises an α -helix surrounded by a 12-stranded β -barrel structure with a hydrophobic thickness of 25.2 Å. An estimate of the average β -strand inclination to the outer membrane is 40 degrees, assuming a 5 degree tilt according to Lomize and colleagues [63]. Structural alignment of EhaB_c with the β -barrel structures of the NalP (PDB 1UYN) and the Hbp (PDB 3AEH) showed RMSDs of 0.29 Å and 0.15 Å in spite of a very low sequence identity (<20%). According to the standard nomenclature [64], the strand connections are named turn N at the periplasmic end and loop N at the external end, where N= 1, 2, 3,...,12. Loop 4 and loop 5 are disordered and missing from the ehaB_c model. This result is expected because the missing passenger domain may interact with and stabilize the external loops of the β barrel domain. For example, all the loops are clearly resolved in the only structure of a full length AT EstA (PDB 3KVN), where the passenger domain stabilizes these external loops.

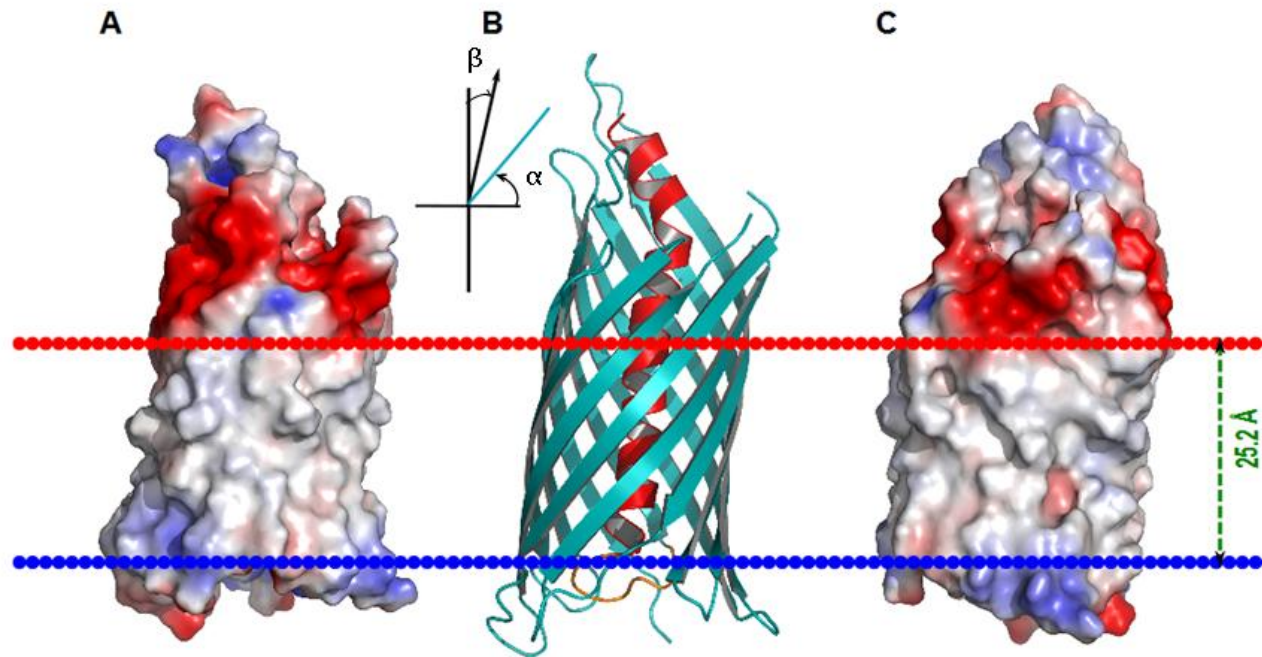


Figure 3.9 Structure of EhaB_c. (A) Surface representation of the EhaB_c monomer viewed from the side. (B) Cartoon model viewed from the side. The α -helix is colored in red, the β barrel is colored in dark cyan, and the turn 1 is colored in orange. (C) Surface representation and viewed from side after 180° rotation. (A) and (C) Electrostatic surface potential calculated using Adaptive Poisson–Boltzmann Solver[65]. Positive potential is colored blue, and negative potential is colored red. Potentials are contoured from +20 kT/e to -20 kT/e. Two lines of dots indicate the outer membrane boundary. The membrane thickness is 25.2 Å. The red dots refer to the extracellular side and the blue ones to the periplasmic side. The inclination of the β strand to the membrane boundary is labeled as angle α at 40°. The tilting of the β barrel to the axis vertical to the membrane is labeled as angle β at 5°.

EhaB belongs to the AIDA-I subfamily of the Type V_a ATs, and its passenger domain is not cleaved after the translocation. The α -helix starts at residue 682 and traverses most of the barrel lumen before exiting as a loop, which is then connected to the first β -strand at the periplasmic end. The continuous α -helix of EhaB_c is analogous to that of NalP β -domain, a member of non-SPATE-like serine protease ATs, but quite different from the one observed in the Hbp β -domain, where a conserved motif (EVN \uparrow NLNKRMGDLRD) is found to be the autocleavage site between the long and short α -helices inside the β -domain. In both NalP and Hbp proteins, the passenger domain can be released from the pathogen, an event not expected to occur in EhaB [66].

Detailed analysis of the distribution of the aromatic residues (Phe, Trp, and Tyr) from three AT β -domains is summarized in Table 3.2 and a side view of the β -domain with the aromatic residues highlighted in NalP, EhaB_c and Hbp is shown in sticks Figure 3.10. One anomalous observation is the inverted distribution of the aromatic residues inside versus outside of the barrel structure. Inside of the EhaB_c barrel there are 10 aromatic residues within the β -barrel pore (i.e., in membrane, but on the inner surface of the pore) and 4 on the extracellular surface, while NalP and Hbp have less within the membrane (3 and 6, respectively) and more outside the membrane (8 and 7, respectively). Although there is no satisfactory explanation for this discrepancy at the moment, our data from molecular dynamic simulation of the EhaB_c (Y. Zheng, M. Feig., and R. M. Garavito, unpublished data) indicate that the α -helix alone may be prone to unfolding and can be “pushed out” of the barrel structure into the periplasmic space. This was not observed in the similar study on

NalP β -domain [44] and prompted us to further analyze the residue distribution within the barrel. Our structure indeed shows significant differences in the aromatic residue distribution, which suggests that the inverted distribution of the aromatic residues may play a role in stabilizing or destabilizing the α -helix inside the barrel. This may have some functional implications as well. Since EhaB has a non-cleavable passenger domain that functions as an adhesin to laminin and collagen I, an additional tug force may be required to enhance or stabilize adhesion, thus preventing adhesin molecules from falling off the outer membrane. However, these are only hypotheses that need to be tested by experiment.

Table 3.2 Aromatic residue distributions in the EhaB, BalP and Hbp β domains

Number of aromatic residues of the β barrel	α -helix	Barrel inner surface		outer surface
		Within ^a	Above ^b	
EhaB	1	10	4	24
NalP	0	3	8	9
Hbp	3	6	7	18

a: refers to the part of β barrel predicted to be within the membrane thickness.

b: refers to the part of β barrel predicted to be above the membrane.

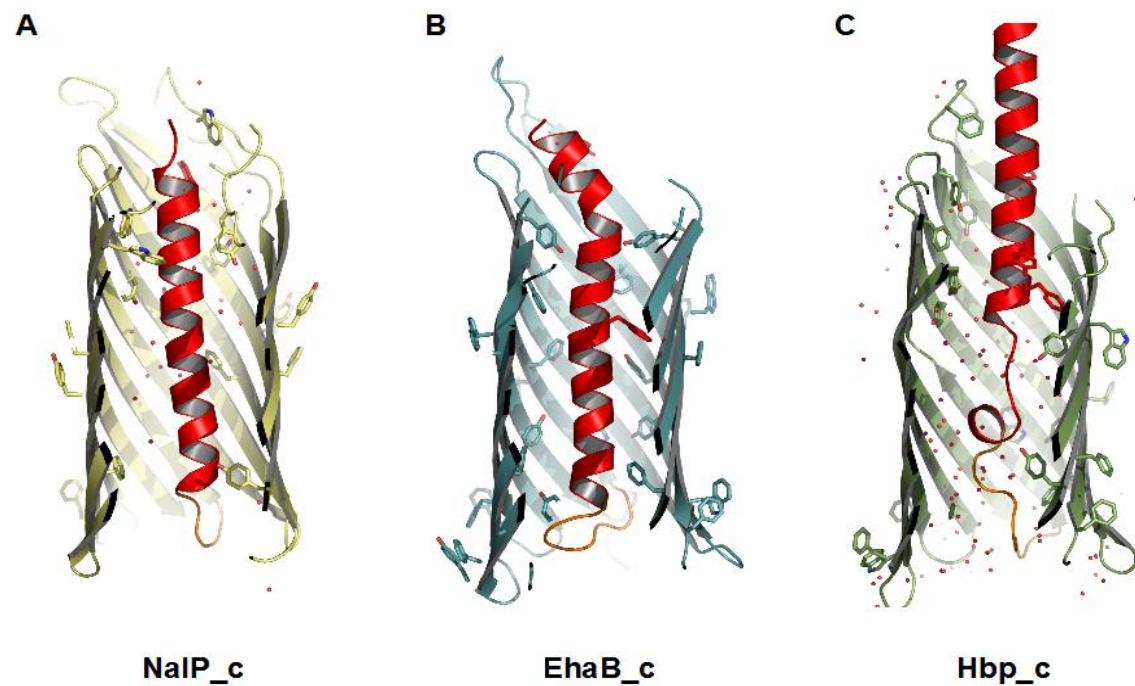


Figure 3.10 Cut-away view of NalP_c, EhaB_c and Hbp_c. (A) NalP_c (1UYN, (B) EhaB_c, (C) Hbp_c (3AEH). Helices inside the barrel are colored in red, link regions between the helix and the barrel are colored in orange and the β barrels are colored in yellow, pale cyan and pale green, respectively.

3.3.5 The planar BLM studies of the EhaB translocation domain

Although the translocation domain of AT proteins is not a pore in a strict sense, its β -barrel structure is highly reminiscent of the OM porins, and the gap between the α -helix and the β -barrel indeed is capable of allowing ion flow across the membrane [44, 67]. Therefore, an investigation of ion conductivity of EhaB_c protein in a planar lipid bilayer (BLM) may provide valuable information concerning the interactions between the α -helix and the wall of the β -barrel, as well as the overall structural dynamics of the translocation domain. In preliminary planar BLM studies, a single EhaB_c protein was incorporated into a DPhPC membrane and its ion conductance was measured at different membrane potentials. The EhaB_c protein is open at positive potentials and stable for several seconds, but less inclined to open at negative potentials (Figure 3.11). As shown in the potential ramping graph (Figure 3.12), the EhaB protein functions as a rectifying voltage-gated channel when inserted into the membranes (i.e., ions flow easier in one direction than the other.). As the absolute orientation of the EhaB_c protein relative to the electrodes is unknown, the relationship of the rectification to the pore's structure is also not known. However, this observation is consistent with the fact that the α -helix folds in a directional manner inside the β -barrel.

The constant alternating between the open state and the closed state (in less than ms) indicates the dynamic nature of the α -helix. At 100 mV, the channel abruptly closed after only a few seconds of opening (Figure 3.11). One explanation could be that there is a potential induced conformational state into which the α -helix can be folded to close the pore

for a long period of time. Given that the absolute orientation of the EhaB_c in the membrane is not yet known, we have designed and expressed EhaB_c fusion protein with a second protein fused to the N-terminus of the EhaB_c protein, tethered by a cleavable linker (data not shown). If these EhaB_c protein chimeras behave well in BLM experiments, conductance measurements can be done in the presence and absence of proteases. Such experiments should provide the opportunity to determine the absolute orientation of the translocation domain in the membrane and allow a wide-ranging array of mutagenic studies to tackle structure function questions regarding the dynamics and folding of the translocator domains in type V_a ATs.

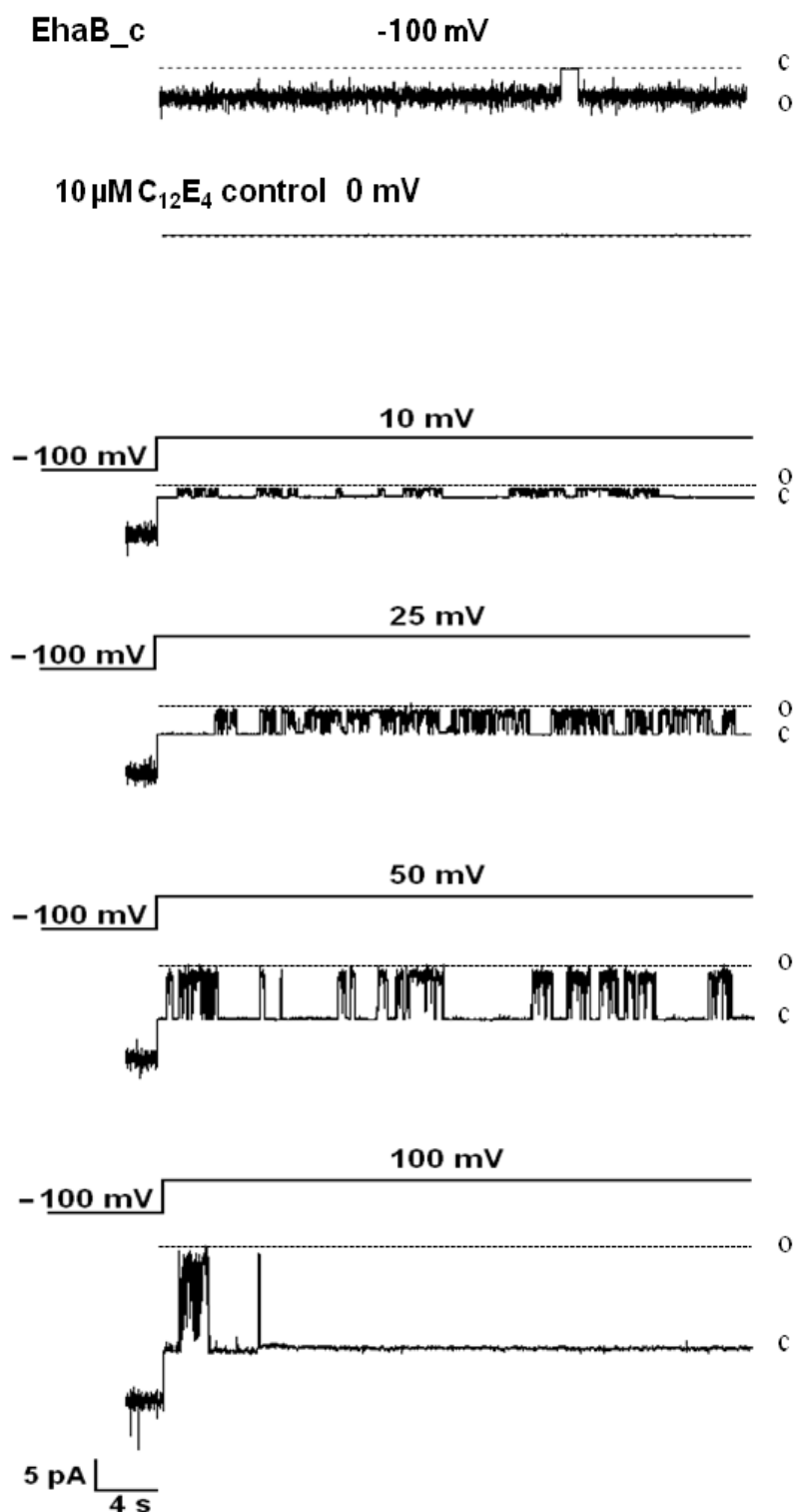


Figure 3.11 Behavior of EhaB_c in planar BLM. No current is detected in the detergent control sample. When single channel is incorporated into the membrane, -100 mV potential is applied until the current is stable. A gradual increase of positive potentials is applied to record the channel activity. The closed state and the open state are labeled as c and o, respectively. Both chambers are filled with buffer (200 mM KCl, 5 mM CaCl₂, 20 mM HEPES pH 7.3).

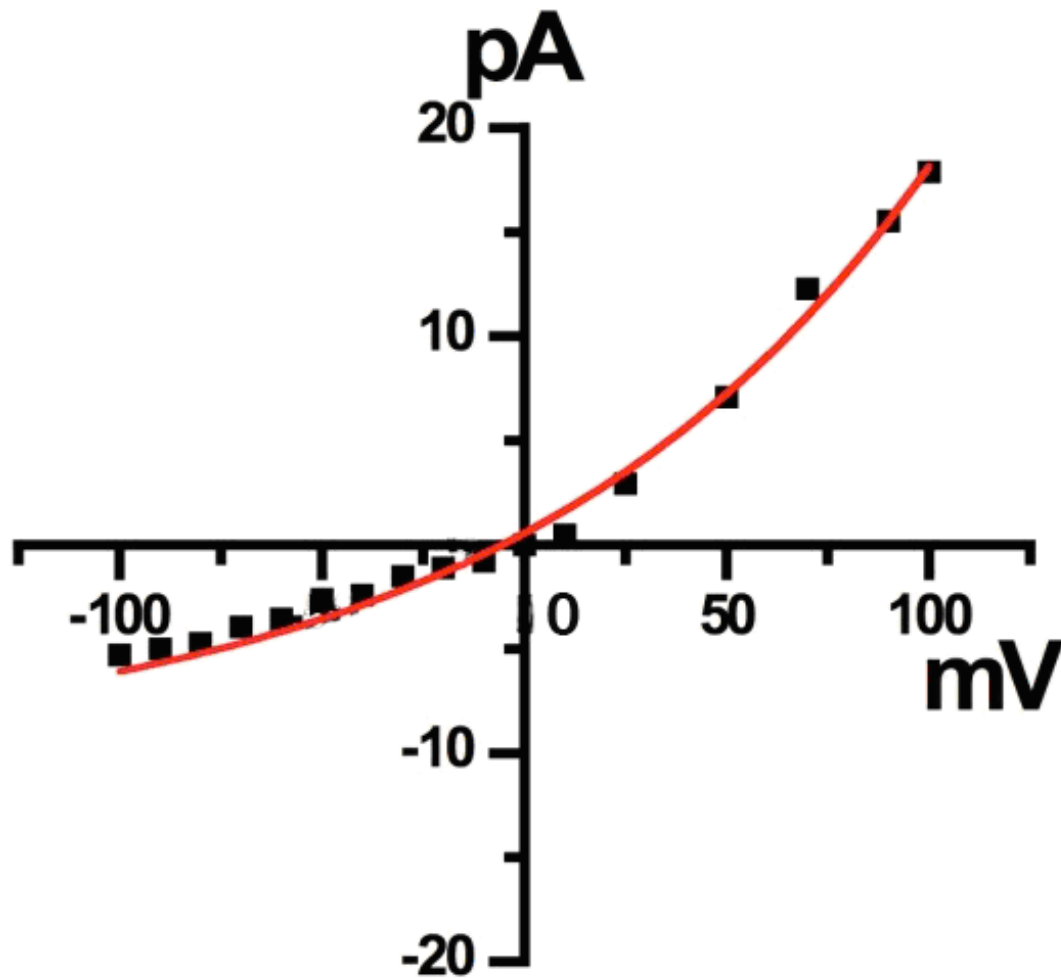


Figure 3.12 EhaB_c is a rectifying voltage-gated channel. Cycles of potential ramping from +100 mV to -100 mV are applied to the single channel incorporated in the planar BLM. The recorded currents are plotted on the Y axis to show the degree of the channel opening at the specific potential.

3.4 Conclusions

Autotransporters are important virulence factors in Gram-negative bacteria. Despite their divergent virulence functions, a highly conserved domain organization suggests the shared biogenesis pathway. The crystal structure of the EhaB translocation domain, an AIDA-I type AT, provides the structure basis for close comparison to the structures from SPATE type ATs. EhaB is an outer membrane protein with a conserved 12-stranded β -barrel and an N-terminus α -helix folding inside the barrel. Some of differences in the aromatic residue distribution highlighted here may play a role in the proper function of the passenger domain.

REFERENCES

REFERENCES

- 1 Nataro, J. P. and Kaper, J. B. (1998) Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev.* **11**, 142-201
- 2 Riley, L. W., Remis, R. S., Helgerson, S. D., Mcgee, H. B., Wells, J. G., Davis, B. R., Hebert, R. J., Olcott, E. S., Johnson, L. M., Hargrett, N. T., Blake, P. A. and Cohen, M. L. (1983) Hemorrhagic Colitis Associated with a Rare *Escherichia-Coli* Serotype. *New England Journal of Medicine.* **308**, 681-685
- 3 Ratnam, S., March, S. B., Ahmed, R., Bezanson, G. S. and Kasatiya, S. (1988) Characterization of *Escherichia coli* serotype O157:H7. *J Clin Microbiol.* **26**, 2006-2012
- 4 Mead, P. S., Slutsker, L., Dietz, V., McCaig, L. F., Bresee, J. S., Shapiro, C., Griffin, P. M. and Tauxe, R. V. (1999) Food-related illness and death in the United States. *Emerging Infectious Diseases.* **5**, 607-625
- 5 Kaper, J. B., Nataro, J. P. and Mobley, H. L. (2004) Pathogenic *Escherichia coli*. *Nat Rev Microbiol.* **2**, 123-140
- 6 Brussow, H., Canchaya, C. and Hardt, W. D. (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev.* **68**, 560-602
- 7 Melton-Celsa, A. R., Rogers, J. E., Schmitt, C. K., Darnell, S. C. and O'Brien, A. D. (1998) Virulence of Shiga toxin-producing *Escherichia coli* (STEC) in orally-infected mice correlates with the type of toxin produced by the infecting strain. *Jpn J Med Sci Biol.* **51 Suppl**, S108-114
- 8 Andreoli, S. P., Trachtman, H., Acheson, D. W., Siegler, R. L. and Obrig, T. G. (2002) Hemolytic uremic syndrome: epidemiology, pathophysiology, and therapy. *Pediatr Nephrol.* **17**, 293-298

- 9 Perna, N. T., Plunkett, G., 3rd, Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E. J., Davis, N. W., Lim, A., Dimalanta, E. T., Potamouisis, K. D., Apodaca, J., Anantharaman, T. S., Lin, J., Yen, G., Schwartz, D. C., Welch, R. A. and Blattner, F. R. (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature*. **409**, 529-533
- 10 Kenny, B. (2002) Mechanism of action of EPEC type III effector molecules. *Int J Med Microbiol*. **291**, 469-477
- 11 Hecht, G. (2001) Microbes and microbial toxins: paradigms for microbial-mucosal interactions. VII. Enteropathogenic *Escherichia coli*: physiological alterations from an extracellular position. *Am J Physiol Gastrointest Liver Physiol*. **281**, G1-7
- 12 Wolf, M. K. (1997) Occurrence, distribution, and associations of O and H serogroups, colonization factor antigens, and toxins of enterotoxigenic *Escherichia coli*. *Clin Microbiol Rev*. **10**, 569-584
- 13 Tauschek, M., Gorrell, R. J., Strugnelli, R. A. and Robins-Browne, R. M. (2002) Identification of a protein secretory pathway for the secretion of heat-labile enterotoxin by an enterotoxigenic strain of *Escherichia coli*. *Proc Natl Acad Sci U S A*. **99**, 7066-7071
- 14 Hueck, C. J. (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev*. **62**, 379-433
- 15 Balakrishnan, L., Hughes, C. and Koronakis, V. (2001) Substrate-triggered recruitment of the TolC channel-tunnel during type I export of hemolysin by *Escherichia coli*. *J Mol Biol*. **313**, 501-510
- 16 Durand, E., Verger, D., Rego, A. T., Chandran, V., Meng, G., Fronzes, R. and Waksman, G. (2009) Structural biology of bacterial secretion systems in gram-negative pathogens--potential for new drug targets. *Infect Disord Drug Targets*. **9**, 518-547
- 17 Delepelaire, P. (2004) Type I secretion in gram-negative bacteria. *Biochim Biophys Acta*. **1694**, 149-161

- 18 Cascales, E. and Christie, P. J. (2003) The versatile bacterial type IV secretion systems. *Nat Rev Microbiol.* **1**, 137-149
- 19 Thanassi, D. G., Stathopoulos, C., Karkal, A. and Li, H. (2005) Protein secretion in the absence of ATP: the autotransporter, two-partner secretion and chaperone/usher pathways of gram-negative bacteria (review). *Mol Membr Biol.* **22**, 63-72
- 20 Kostakioti, M., Newman, C. L., Thanassi, D. G. and Stathopoulos, C. (2005) Mechanisms of protein export across the bacterial outer membrane. *J Bacteriol.* **187**, 4306-4314
- 21 Bingle, L. E., Bailey, C. M. and Pallen, M. J. (2008) Type VI secretion: a beginner's guide. *Curr Opin Microbiol.* **11**, 3-8
- 22 Cascales, E. (2008) The type VI secretion toolkit. *EMBO Rep.* **9**, 735-741
- 23 Ayers, M., Howell, P. L. and Burrows, L. L. Architecture of the type II secretion and type IV pilus machineries. *Future Microbiol.* **5**, 1203-1218
- 24 Alvarez-Martinez, C. E. and Christie, P. J. (2009) Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev.* **73**, 775-808
- 25 Mougous, J. D., Cuff, M. E., Raunser, S., Shen, A., Zhou, M., Gifford, C. A., Goodman, A. L., Joachimiak, G., Ordonez, C. L., Lory, S., Walz, T., Joachimiak, A. and Mekalanos, J. J. (2006) A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science.* **312**, 1526-1530
- 26 Leiman, P. G., Basler, M., Ramagopal, U. A., Bonanno, J. B., Sauder, J. M., Pukatzki, S., Burley, S. K., Almo, S. C. and Mekalanos, J. J. (2009) Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proceedings of the National Academy of Sciences of the United States of America.* **106**, 4154-4159
- 27 Pell, L. G., Kanelis, V., Donaldson, L. W., Howell, P. L. and Davidson, A. R. (2009) The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proceedings of the National Academy of Sciences of the United States of America.* **106**, 4160-4165

- 28 Kajava, A. V., Cheng, N., Cleaver, R., Kessel, M., Simon, M. N., Willery, E., Jacob-Dubuisson, F., Loch, C. and Steven, A. C. (2001) Beta-helix model for the filamentous haemagglutinin adhesin of *Bordetella pertussis* and related bacterial secretory proteins. *Mol Microbiol.* **42**, 279-292
- 29 Dautin, N. and Bernstein, H. D. (2007) Protein secretion in gram-negative bacteria via the autotransporter pathway. *Annual Review of Microbiology.* **61**, 89-112
- 30 Wells, T. J., Totsika, M. and Schembri, M. A. (2010) Autotransporters of *Escherichia coli*: a sequence-based characterization. *Microbiology-Sgm.* **156**, 2459-2469
- 31 Wells, T. J., Sherlock, O., Rivas, L., Mahajan, A., Beatson, S. A., Torpdahl, M., Webb, R. I., Allsopp, L. P., Gobijs, K. S., Gally, D. L. and Schembri, M. A. (2008) EhaA is a novel autotransporter protein of enterohemorrhagic *Escherichia coli* O157 : H7 that contributes to adhesion and biofilm formation. *Environmental Microbiology.* **10**, 589-604
- 32 Nummelin, H., Merckel, M. C., Leo, J. C., Lankinen, H., Skurnik, M. and Goldman, A. (2004) The *Yersinia* adhesin YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. *Embo Journal.* **23**, 701-711
- 33 Yeo, H. J., Cotter, S. E., Laarmann, S., Juehne, T., St Geme, J. W. and Waksman, G. (2004) Structural basis for host recognition by the *Haemophilus influenzae* Hia autotransporter. *Embo Journal.* **23**, 1245-1256
- 34 Linke, D., Riess, T., Autenrieth, I. B., Lupas, A. and Kempf, V. A. J. (2006) Trimeric autotransporter adhesins: variable structure, common function. *Trends in Microbiology.* **14**, 264-270
- 35 Hoiczky, E., Roggenkamp, A., Reichenbecher, M., Lupas, A. and Heesemann, J. (2000) Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins. *Embo Journal.* **19**, 5989-5999
- 36 Girard, V. and Mourez, M. (2006) Adhesion mediated by autotransporters of Gram-negative bacteria: Structural and functional features. *Research in Microbiology.* **157**, 407-416

- 37 Brandon, L. D., Goehring, N., Janakiraman, A., Yan, A. W., Wu, T., Beckwith, J. and Goldberg, M. B. (2003) IcsA, a polarly localized autotransporter with an atypical signal peptide, uses the Sec apparatus for secretion, although the Sec apparatus is circumferentially distributed. *Mol Microbiol.* **50**, 45-60
- 38 Emsley, P., Charles, I. G., Fairweather, N. F. and Isaacs, N. W. (1996) Structure of *Bordetella pertussis* virulence factor P.69 pertactin. *Nature.* **381**, 90-92
- 39 Nummelin, H., Merckel, M. C., Leo, J. C., Lankinen, H., Skurnik, M. and Goldman, A. (2004) The *Yersinia* adhesin YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. *EMBO J.* **23**, 701-711
- 40 Otto, B. R., Sijbrandi, R., Luirink, J., Oudega, B., Heddle, J. G., Mizutani, K., Park, S. Y. and Tame, J. R. (2005) Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic *Escherichia coli*. *J Biol Chem.* **280**, 17339-17345
- 41 Brandon, L. D., Goehring, N., Janakiraman, A., Yan, A. W., Wu, T., Beckwith, J. and Goldberg, M. B. (2003) IcsA, a polarly localized autotransporter with an atypical signal peptide, uses the Sec apparatus for secretion, although the Sec apparatus is circumferentially distributed. *Molecular Microbiology.* **50**, 45-60
- 42 Peterson, J. H., Szabady, R. L. and Bernstein, H. D. (2006) An unusual signal peptide extension inhibits the binding of bacterial presecretory proteins to the signal recognition particle, trigger factor, and the SecYEG complex. *Journal of Biological Chemistry.* **281**, 9038-9048
- 43 Sijbrandi, R., Urbanus, M. L., ten Hagen-Jongman, C. M., Bernstein, H. D., Oudega, B., Otto, B. R. and Luirink, J. (2003) Signal recognition particle (SRP)-mediated targeting and sec-dependent translocation of an extracellular *Escherichia coli* protein. *Journal of Biological Chemistry.* **278**, 4654-4659
- 44 Oomen, C. J., van Ulsen, P., van Gelder, P., Feijen, M., Tommassen, J. and Gros, P. (2004) Structure of the translocator domain of a bacterial autotransporter. *EMBO J.* **23**, 1257-1266
- 45 Meng, G., Surana, N. K., St Geme, J. W., 3rd and Waksman, G. (2006) Structure of the outer membrane translocator domain of the *Haemophilus influenzae* Hia trimeric autotransporter. *EMBO J.* **25**, 2297-2304

- 46 Khalid, S. and Sansom, M. S. (2006) Molecular dynamics simulations of a bacterial autotransporter: NalP from *Neisseria meningitidis*. *Mol Membr Biol.* **23**, 499-508
- 47 Voulhoux, R., Bos, M. P., Geurtsen, J., Mols, M. and Tommassen, J. (2003) Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science.* **299**, 262-265
- 48 Wu, T., Malinverni, J., Ruiz, N., Kim, S., Silhavy, T. J. and Kahne, D. (2005) Identification of a multicomponent complex required for outer membrane biogenesis in *Escherichia coli*. *Cell.* **121**, 235-245
- 49 Ieva, R., Skillman, K. M. and Bernstein, H. D. (2008) Incorporation of a polypeptide segment into the beta-domain pore during the assembly of a bacterial autotransporter. *Mol Microbiol.* **67**, 188-201
- 50 Skillman, K. M., Barnard, T. J., Peterson, J. H., Ghirlando, R. and Bernstein, H. D. (2005) Efficient secretion of a folded protein domain by a monomeric bacterial autotransporter. *Molecular Microbiology.* **58**, 945-958
- 51 Veiga, E., Sugawara, E., Nikaido, H., de Lorenzo, V. and Fernandez, L. A. (2002) Export of autotransported proteins proceeds through an oligomeric ring shaped by C-terminal domains. *EMBO J.* **21**, 2122-2131
- 52 Buchanan, S. K. (1999) Overexpression and refolding of an 80-kDa iron transporter from the outer membrane of *Escherichia coli*. *Biochem Soc Trans.* **27**, 903-908
- 53 Daugherty, D. L., Rozema, D., Hanson, P. E. and Gellman, S. H. (1998) Artificial chaperone-assisted refolding of citrate synthase. *J Biol Chem.* **273**, 33961-33971
- 54 Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A.* **276**, 307-326
- 55 Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallographica Section D-Biological Crystallography.* **58**, 1948-1954

- 56 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) Phaser crystallographic software. *Journal of Applied Crystallography*. **40**, 658-674
- 57 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography*. **60**, 2126-2132
- 58 Davis, I. W., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2004) MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Research*. **32**, W615-W619
- 59 Mironova, G. D., Baumann, M., Kolomytkin, O., Krasichkova, Z., Berdimuratov, A., Sirota, T., Virtanen, I. and Saris, N. E. L. (1994) Purification of the Channel Component of the Mitochondrial Calcium Uniporter and Its Reconstitution into Planar Lipid Bilayers. *Journal of Bioenergetics and Biomembranes*. **26**, 231-238
- 60 Notredame, C., Higgins, D. G. and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*. **302**, 205-217
- 61 Bond, C. S. and Schuttelkopf, A. W. (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr D Biol Crystallogr*. **65**, 510-512
- 62 Qi, H. L., Tai, J. Y. and Blake, M. S. (1994) Expression of Large Amounts of Neisserial Porin Proteins in Escherichia-Coli and Refolding of the Proteins into Native Trimers. *Infection and Immunity*. **62**, 2432-2439
- 63 Lomize, M. A., Lomize, A. L., Pogozheva, I. D. and Mosberg, H. I. (2006) OPM: orientations of proteins in membranes database. *Bioinformatics*. **22**, 623-625
- 64 Schulz, G. E. (2000) beta-Barrel membrane proteins. *Curr Opin Struct Biol*. **10**, 443-447
- 65 Baker, N. A., Sept, D., Joseph, S., Holst, M. J. and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*. **98**, 10037-10041

- 66 Wells, T. J., McNeilly, T. N., Totsika, M., Mahajan, A., Gally, D. L. and Schembri, M. A. (2009) The *Escherichia coli* O157:H7 EhaB autotransporter protein binds to laminin and collagen I and induces a serum IgA response in O157:H7 challenged cattle. *Environ Microbiol.* **11**, 1803-1814
- 67 Shannon, J. L. and Fernandez, R. C. (1999) The C-terminal domain of the *Bordetella pertussis* autotransporter BrkA forms a pore in lipid bilayer membranes. *J Bacteriol.* **181**, 5838-5842

CHAPTER 4

CRYSTAL STRUCTURE DETERMINATION OF SUCROSE SYNTHASE-1

FROM *ARABIDOPSIS THALIANA*

4.1 Introduction

4.1.1 Sucrose metabolism and sucrose synthesis in plants

Sucrose transport and metabolism is the central system for the partitioning of carbon resources in all vascular plants [1, 2]. In oxygenic plants, sucrose is the major product of photosynthesis [2] and is transported to sites of growth, development, and energy storage [1]. Yet, despite the vital nature of carbon allocation via sucrose transport, only 4 enzyme types [2] are needed to mediate sucrose metabolism in plants: sucrose phosphate synthase, sucrose phosphate phosphatase, invertase, and sucrose synthase. By coordinating the levels of activity of these enzyme classes at different times during growth, in different plant organs, and in different cellular locations [3-9], plants create a spatial and temporal system of sucrose sources and sinks that effectively and efficiently transports sucrose from one site to another [3, 4, 10-14]. The balance between both sucrose flux and its metabolic utilization can be achieved simply by controlling the activities of the sucrose-cleaving enzymes sucrose synthase and invertase [1, 6, 15].

Of the sucrose-metabolizing enzymes, only sucrose synthase (SUS) is capable of catalyzing the synthesis and cleavage of sucrose in a reversible and relatively energy neutral manner [1, 2, 15]. The SUS reaction therefore provides a crucial direct, but reversible, link in

sucrose metabolism between respiration, carbohydrate biosynthesis, and carbohydrate utilization [2], as well as allowing a sucrose sink to be rapidly converted into a sucrose source without the synthesis or degradation of enzyme [1, 15]. Although invertases have a clear role in normal plant growth [16, 17], SUS is critically involved in pollen tube growth [18, 19], the establishment of nitrogen fixation [9, 20-22], biomass production [4, 12], and fruit and seed maturation [3, 23-25], particularly during abiotic stresses [17, 26, 27]. SUS participates in the regulation of sucrose flux through its capacity to rapidly alter its cellular location, from the cytosol to sites of cellulose, callose, and starch biosynthesis [4, 6, 10, 11, 18, 28-30]. SUS interacts with various organelles [13, 15, 31, 32], membrane protein targets [6, 18, 19, 22, 33-35], and cytoskeletal actin [5, 36]. However, the mechanism by which SUS binds to cellular targets, the structural aspects that control SUS partitioning, and the structural effects of the partitioning on the catalytic function of SUS are still mostly unknown at a molecular and atomic level.

4.1.2 Sucrose synthase structure biology

SUS is a retaining glycosyltransferase and a member of the GT-4 glycosyltransferase subfamily, within the metal-independent GT-B glycosyltransferase superfamily [37]. From primary sequence and biochemical analyses, SUS is structurally and functionally similar to sucrose phosphate synthases and glycogen synthases [2]. In some plants, more than one SUS isoform is expressed *in vivo* [1, 15, 17]. The SUS1 isoform from *Arabidopsis thaliana* (AtSus1) has the primary structure of a typical sucrose synthase: residues 1 to 277 form an N-

terminal “regulatory” domain involved in cellular targeting [38] and residues 278 to 771 form the GT-B glycosyltransferase domain. Finally, a C-terminal extension, which is the most variable of the SUS domains, is only 37 amino acids long in AtSus1, but it can be over 100 residues in other SUS isoforms [39]. Within the N-terminal regulatory domain, two Ser residues have been identified as sites for phosphorylation [40-42]. In *Z. mays*, the phosphorylation of the SUS1 isoform at Ser15 (Ser13 in AtSus1) diminishes SUS binding to actin [15], increases membrane association [42, 43], and enhances SUS catalytic activity at acidic pH [43]. In contrast, the phosphorylation of Ser170 in maize SUS1 (Ser167 in AtSus1) promotes SUS turnover by enhancing SUS ubiquitinylation and its subsequent degradation [40, 44]. Interestingly, SUS1 from *Z. mays* and *Glycine max* also bind the early nodulin 40 (ENOD40) peptides [40, 45], which are hormone-like peptides involved in root nodule organogenesis in legumes [46].

In this chapter, I present the X-ray crystal structures of AtSus1, which was determined both as a complex with bound UDP-glucose at 2.8 Å resolution and as a complex with bound UDP and fructose at 2.85 Å resolution. The structures shed light on structure-function relationships in sucrose catalysis and cleavage and the structural relationships between retaining and inverting glycosyltransferases in the GTB superfamily. The structure of AtSus1 also reveals structural features that may be involved in SUS interactions with its cellular targets.

4.2 Experimental procedures

4.2.1 Materials

Chemicals were purchased from Sigma-Aldrich. UDP were purchased from Sigma-Aldrich. UDP-glucose was purchased from Crystalchem. *Escherichia coli* strain B843 was purchased from Novagen. Selenomethionine was purchased from Anatrace. Crystallization screens were purchased from Hampton Research, Molecular Dimensions and Emerald Biosystem. The cDNA library of *A. thaliana* was generously provided by Dr. Eric Moellering at Michigan State University.

4.2.2 Methods

4.2.2.1 Cloning and expression of AtSus1

The open reading frame (ORF) of AtSUS1 was PCR-amplified with high-fidelity Taq DNA polymerase (Invitrogen) using the first strand cDNAs prepared from *A. thaliana* (ecotype Columbia), and the gene-specific primers AtSUS1-FW, 5'-ACATGTCAGCAAACGCTGAACGTATG-3', and AtSUS1-RV, 5'-GTCGACATCATCTTGTGCAAGAGGAA-3'. The primer sequences included appropriate restriction sites AflIII and SalI (underlined). The PCR product was cloned into the pGEM®-T Easy vector (Promega). The excised fragment of AtSUS1 ORF was subcloned into the expression vector pET17b (Novagen) using the compatible NcoI and XhoI sites to create a C-terminal fusion protein with a His6-tag. Constructs were confirmed by sequencing and introduced into the *E. coli* expression strain Rosetta λDE3.

4.2.2.2 Protein expression and purification

The transformed bacteria were cultured at 37 °C in 2YT medium containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol until reaching an OD₆₀₀ of ~1.0, and then gene expression was induced with 0.5 mM IPTG for 12–16 h at 20~23 °C. Harvested cells (~4 g from 700 ml of cell culture) were resuspended in 50 ml lysis buffer (30 mM Tris-HCl, pH 8.0, 200 mM NaCl, 150 mM sucrose, 2 mM EDTA, and 1 mM DTT) and were lysed after two passes at 15,000–20,000 p.s.i on an Avestin EmulsiFlex C-3 homogenizer. Following a centrifugation step at 20,000 x g (4 °C) for 20 min, the clarified lysate was fractionated by the addition of (NH₄)₂SO₄; the fraction that precipitated between 45 and 55% (NH₄)₂SO₄ saturation was collected by centrifugation at 20,000 x g for 20 min. The precipitate was redissolved in 50 mL buffer A (lysis buffer without EDTA) and was applied to a 10 mL Ni-NTA (Qiagen) gravity flow column, pre-equilibrated in buffer A. After extensive washing, fractions were eluted using buffer A containing 150 mM imidazole over two column volumes. Fractions containing AtSUS1 were pooled and concentrated to about 1 mL, and 500 µL volumes were applied at 0.4 mL min⁻¹ to a Superose 6 gel filtration column (GE Life Sciences) pre-equilibrated with either buffer A (1mM EDTA) or buffer B (30 mM TrisHCl, pH 8.0, 200 mM NaCl, 20 mM fructose, 1 mM EDTA, and 1 mM DTT). Peak fractions were concentrated to a protein concentration of ~15 mg/mL. The selenomethionyl (SeMet) derivative of AtSUS1 was expressed in the methionine auxotroph cell line B843 (DE3) (Novagen) and purified in the same way as the native protein.

4.2.2.3 Crystallization and cryoprotection

UDP or UDP-glucose was added to the AtSUS1 protein solution to a 5 mM final concentration. Crystals were grown by the hanging-drop method at 20 °C using 2 μ L of protein solution mixed with an equal volume of reservoir solution. AtSus1 crystallized under a number of different conditions in monoclinic and tetragonal habits at 1.8 - 2.0 M ammonium sulfate (AS). The crystals of AtSUS1 in complex with fructose and UDP grew in 80 mM sodium citrate, pH 5.8, 150 mM potassium sodium tartrate, and 1.86 M ammonium sulfate. The crystals of UDP-glucose complex grew in the same condition except the pH of the citrate buffer was 6.1. Cryoprotection was achieved by sequential addition of increments of mother liquor supplemented with 2.0 M sodium malonate pH 5.8 followed by subsequent flash cooling in liquid nitrogen. Native AtSus1 crystals diffracted to 2.6-2.9 Å for C2 space group and 3.4 Å for I422 space group. Interestingly, reducing agents may play a role in determining the crystal space groups. In the absence of DTT, AtSus1 crystals are about equally divided into the two space groups, whereas if 2 mM DTT is added, the population of Sus1 crystals is predominately monoclinic. AtSus1 C2 crystals have the lattice constants: $a=276.75$ Å, $b=261.88$ Å, $c=160.19$ Å, $\alpha = \beta = 90.0^\circ$, $\gamma = 108.7^\circ$.

4.2.2.4 Phasing, structure determination, and refinement

The native and SAD (single-wavelength anomalous dispersion) data were collected at LS-CAT (Advanced Photon Source, Argonne, IL) beamlines 21-ID-D and 21-ID-G and processed using XDS [47] and SCALA [48]. Positions for 94 of the 112 selenium sites in

the asymmetric unit (ASU) of the SeMet-AtSus1/UDP/fructose crystals were located at 3.1 Å using AutoSol in the Phenix program suite [49], version 1.6.2; the overall figure of merit was 0.404 over 168,850 Friedel pairs. In the first SAD-phased, density-modified electron density map, the molecular boundary was clear and secondary structural elements could be identified. Density modification and initial model building were done using AutoBuild, which yielded an initial model with >4800 of possible 6,520 residues built in the first round ($R_{\text{work}} = 0.31$ and $R_{\text{free}} = 0.35$). Using the Se sites as sequence landmarks, a consistent primary sequence assignment was made for 1 chain in the SeMet-AtSus1 structure, and yielded a tentative model from residue 47 to 120 and 153 to 804 for the 8 subunits in the ASU. This initial model was then used to phase native data for the AtSus1/UDP/fructose complex to 2.85 Å resolution. Iterative model building with COOT 6.1 [50] and refinement against the native data using Phenix Refine, with NCS restraints for residues 130-807, yielded consistent atomic models for all 8 subunits in the ASU. After releasing the NCS restraints during the final rounds of model building and refinement, the final model accounted for 6250 of the possible 6,520 residues. The refined model for the AtSus1/UDP/fructose complex was used to phase native data for the AtSus1/UDP-glucose complex to 2.80 Å resolution. Iterative model building with COOT 6.1 [50] and refinement using the same protocols yielded a final model that accounted for 6216 of the possible 6,520 residues. Data collection and refinement statistics for the AtSus1 structures are listed in Table 4.1. For both complexes, ~95.5% of the residues are in the most favored and ~4.4%

are in the additionally allowed region of the Ramachandran plot, and 0~0.1% are in disallowed regions.

4.2.2.5 Multiple sequence alignment of the SUS homologs

The SUS sequences from a variety of plant species were obtained from NCBI, and aligned with T-Coffee [51] with respect to the different folding domains. Figure 4.1 shows the alignment of the CTD (cellular targeting domain) and EPBD (ENOD40 peptide-binding domain) domains (residues 1-276) in AtSus1 with 11 other SUS homologs from 9 different plant species. The figure was prepared using the program ALINE [52].

Table 4.1 Data collection, phasing and refinement statistics for AtSus1 structures

	AtSUS1/Fructose/UDP (SeMet) ¹	AtSUS1/Fructose/ UDP	AtSUS1/UDP- Glucose
Data collection			
Space group	C 2	C 2	C 2
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	276.2 263.7 159.7	277.2 261.5 161.1	276.8 261.9 160.2
α , β , γ (°)	90.0 108.7 90.0	90.0 109.3 90.0	90.0 108.7 90.0
	<i>Peak</i>		
Resolution (Å)	50-2.90	50.00-2.85	50.00-2.80
	(2.95-2.90) ²	(3.0-2.85)	(2.95-2.80)
<i>R</i> _{sym}	0.18(0.72)	0.13(0.68)	0.11(0.62)
<i>I</i> / σI	5.7(1.4)	7.6(1.1)	9.1(1.4)
Completeness (%)	99.9(89.7)	99.4(99.5)	99.6(99.2)
Redundancy	7.3	2.6(2.6)	3.0(3.0)
Refinement			
Resolution (Å)	24.99-2.91	24.96-2.85	24.98-2.80
No. reflections	214680	249285	263089
<i>R</i> _{work} / <i>R</i> _{free}	0.186/0.237	0.185/0.234	0.188/0.236
No. atoms			
Protein	50240	50443	50122
Ligand/ion	480	538	560
Water	563	499	442
<i>B</i> -factors			
Protein	-	46.29	47.90
Ligand/ion	-	48.47	40.93
Water	-	31.67	32.34
R.M.S deviations			
Bond lengths (Å)	0.011	0.011	0.010
Bond angles (°)	1.136	1.079	1.114

1 The structure of Se-Met labeled AtSUS1 complexed with UDP and fructose was only partially refined, due to its lower resolution, and just used to phase the other two complexes.

2 Values in parentheses are for highest-resolution shell.

Figure 4.1

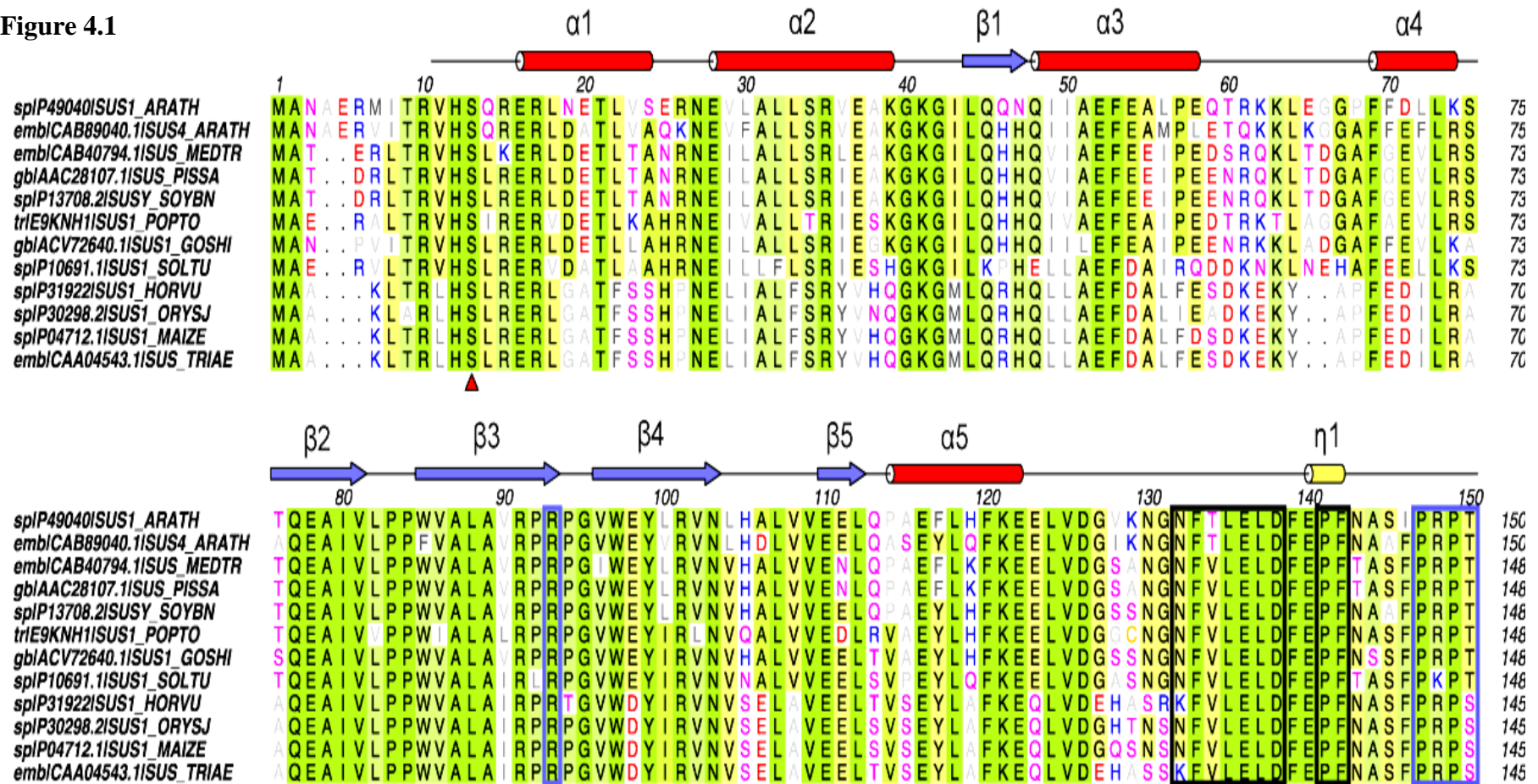


Figure 4.1 continued

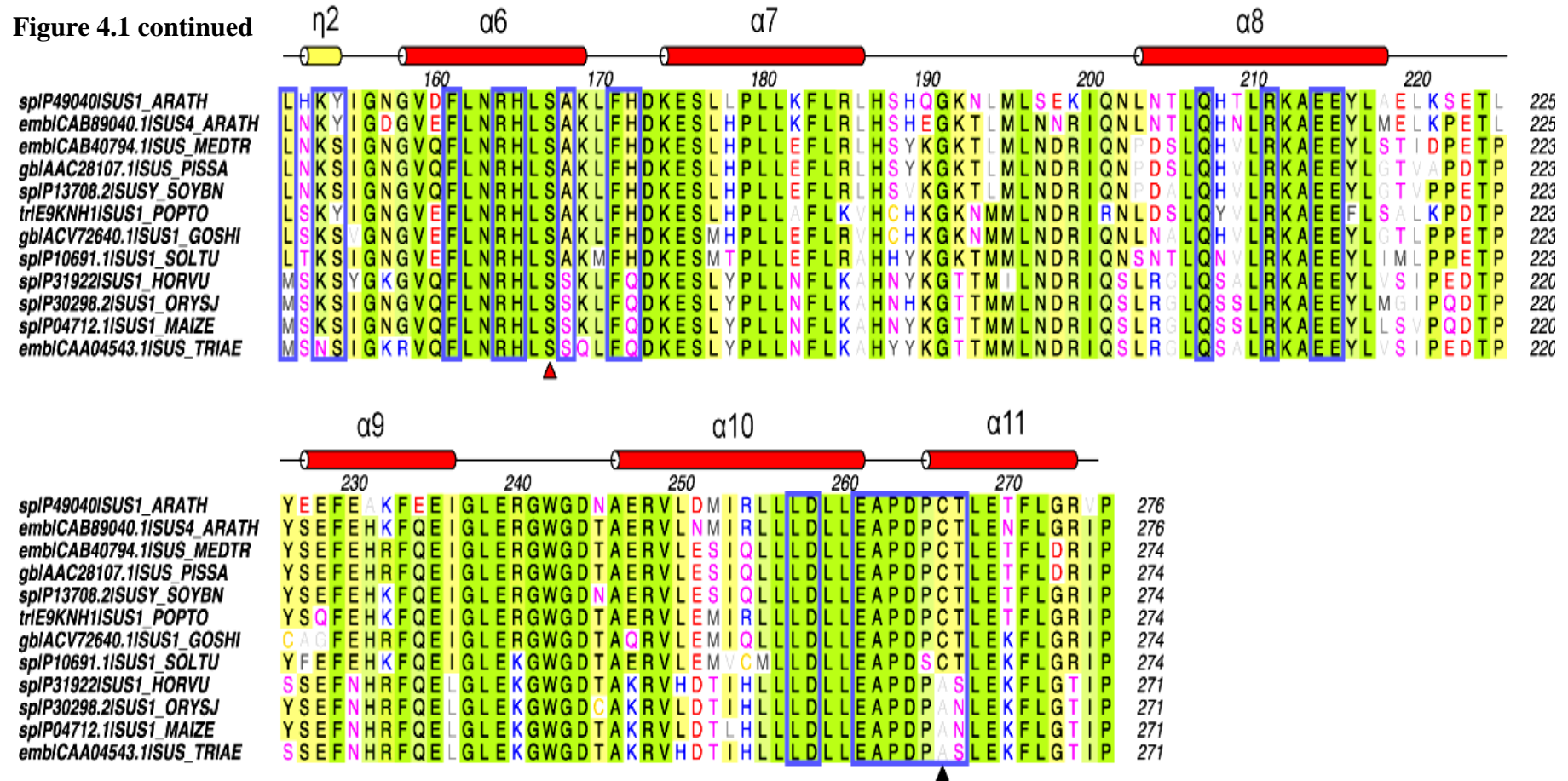


Figure 4.1 Sequence alignment of N terminal regulatory domains of selected SUS enzymes. The numbering scheme and secondary structure profile at the top of the alignment refers to AtSus1. Shown in the alignment are sp|P49040|SUS1_ARATH (AtSus1), emb|CAB89040.1|SUS4_ARATH (*A. thaliana* SUS4), emb|CAB40794.1|SUS_MEDTR (Barrel Clover), gb|AAC28107.1|SUS_PISSA (garden pea), sp|P13708.2|SUSY_SOYBN (soy bean), tr|E9KNH1|SUS1_POPTO (poplar), gb|ACV72640.1|SUS1_GOSHI (cotton), sp|P10691.1|SUS1_SOLTU (potato), sp|P31922|SUS1_HORVU (barley), sp|P30298.2|SUS1_ORYSJ (rice), sp|P04712.1|SUS1_MAIZE (maize), emb|CAA04543.1|SUS_TRIAE (common wheat). Conserved amino acids are shown in four levels of green, the two conserved phosphorylation sites of Ser13 and Ser167 residues is denoted by a red triangle, and site of the thiolation, Cys266, is marked by a black triangle. The interface residues between monomers A and B are highlighted in blue rectangular box. The residues in the black rectangular box are at interfaces between monomers A and D.

4.3 Results and Discussion

4.3.1 Overall structure of AtSus1

Reiterative cycles of model building and refinement yielded well-defined structures for AtSUS1 complexed with either UDP and fructose or UDP-glucose, at 2.85 Å and 2.80 Å resolution, respectively. The resolved contents of the asymmetric unit revealed two 222-symmetric tetramers of AtSUS1 (Figure 4.2; denoted as tetramers ABCD and EFGH) related by a pure translation parallel to the *c*-axis of the unit cell. The overall packing arrangement reveals columns of stacked tetramers (Figure 4.3A), with each column slightly interdigitated with its neighboring columns (Figure 4.3B). This interdigitation apparently impacts the order (Table 4.2) and conformation (Figure 4.3C) of the N-terminal domain of each monomer such that differing portions of the N-termini could not be resolved. For the two tetramers in the asymmetric unit of the AtSUS1/UDP/fructose complex, the model encompassed residues 11 to 807 for subunit H (Figure 4.4), residues 14 to 807 for subunit B, and residues 27 to 807 for the remaining subunits. In the AtSUS1/UDP-glucose complex, the model included residues 28 to 807 for subunits B and H and residues 27 to 807 for the remaining six subunits.

Table 4.2A Average B-factors for all protein atoms (by chain) in SUS1/Fru/UDP complex

Chain	Domain 1 ^a	Domain 2 ^b	Domain 3/4 ^c	Overall
A	73.4 (999) ^d	56.0 (987)	34.8 (4297)	44.3 (6283)
B	82.7 (1035)	62.3 (987)	38.1 (4303)	49.2 (6325)
C	72.5 (1001)	56.8 (987)	41.4 (4303)	48.8 (6291)
D	68.3 (988)	59.6 (987)	35.0 (4303)	44.1 (6278)
E	77.2 (987)	56.9 (987)	36.2 (4297)	45.9 (6271)
F	65.5 (1009)	60.6 (987)	34.3 (4303)	43.4 (6299)
G	66.2 (1008)	55.3 (987)	35.9 (4303)	43.8 (6298)
H	76.3 (1111)	56.7 (984)	38.7 (4303)	48.0 (6398)

a) Residues to 156 (including linker, 129-156)

b) Residues 157-277

c) Residues 278-807

d) Number of atoms

Table 4.2B Average B-factors for all protein atoms (by chain) in SUS1/UDP-glc complex

Chain	Domain 1 ^a	Domain 2 ^b	Domain 3/4 ^c	Overall
A	76.6 (999) ^d	55.7 (987)	33.3 (4303)	43.0 (6289)
B	92.1 (862)	62.7 (987)	40.8 (4303)	51.5 (6152)
C	79.2 (1004)	59.3 (987)	49.8 (4303)	56.0 (6294)
D	71.3 (985)	57.2 (987)	34.7 (4303)	44.0 (6275)
E	82.9 (987)	56.6 (987)	37.3 (4297)	47.5 (6271)
F	66.6 (1006)	59.1 (987)	31.3 (4303)	41.3 (6296)
G	67.2 (1003)	57.3 (987)	37.9 (4303)	45.6 (6293)
H	89.9 (965)	58.3 (984)	42.3 (4303)	52.2 (6252)

a) Residues to 156 (including linker, 129-156)

b) Residues 157-277

c) Residues 278-807

d) Number of atoms

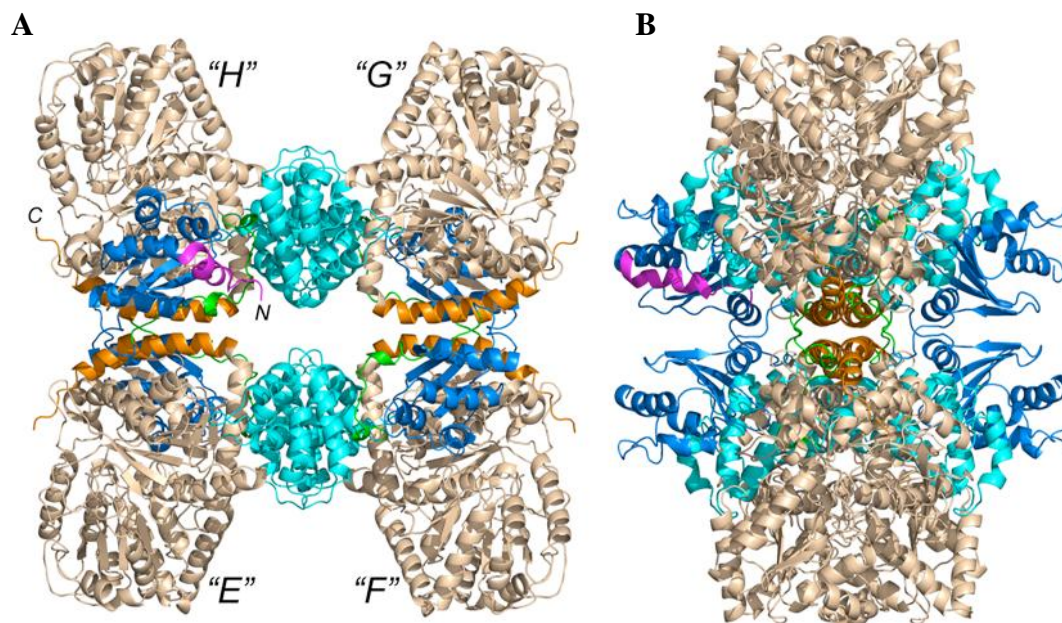


Figure 4.2 A ribbon drawing of the AtSus1 tetramer. Panel A shows the face-on view of the EFGH tetramer, which is equivalent to the ABCD tetramer; panel B shows a side view of the tetramer. The cellular targeting domain (CTD) is colored marine; the ENOD40 peptide binding domain (EPBD) is cyan; the linker between CTD and EPBD is green, the GT-B domains are wheat, and the C-terminal extension helix is orange. The first helix ($\alpha 1$) of the CTD (in magenta) is only ordered in the H and B subunits.

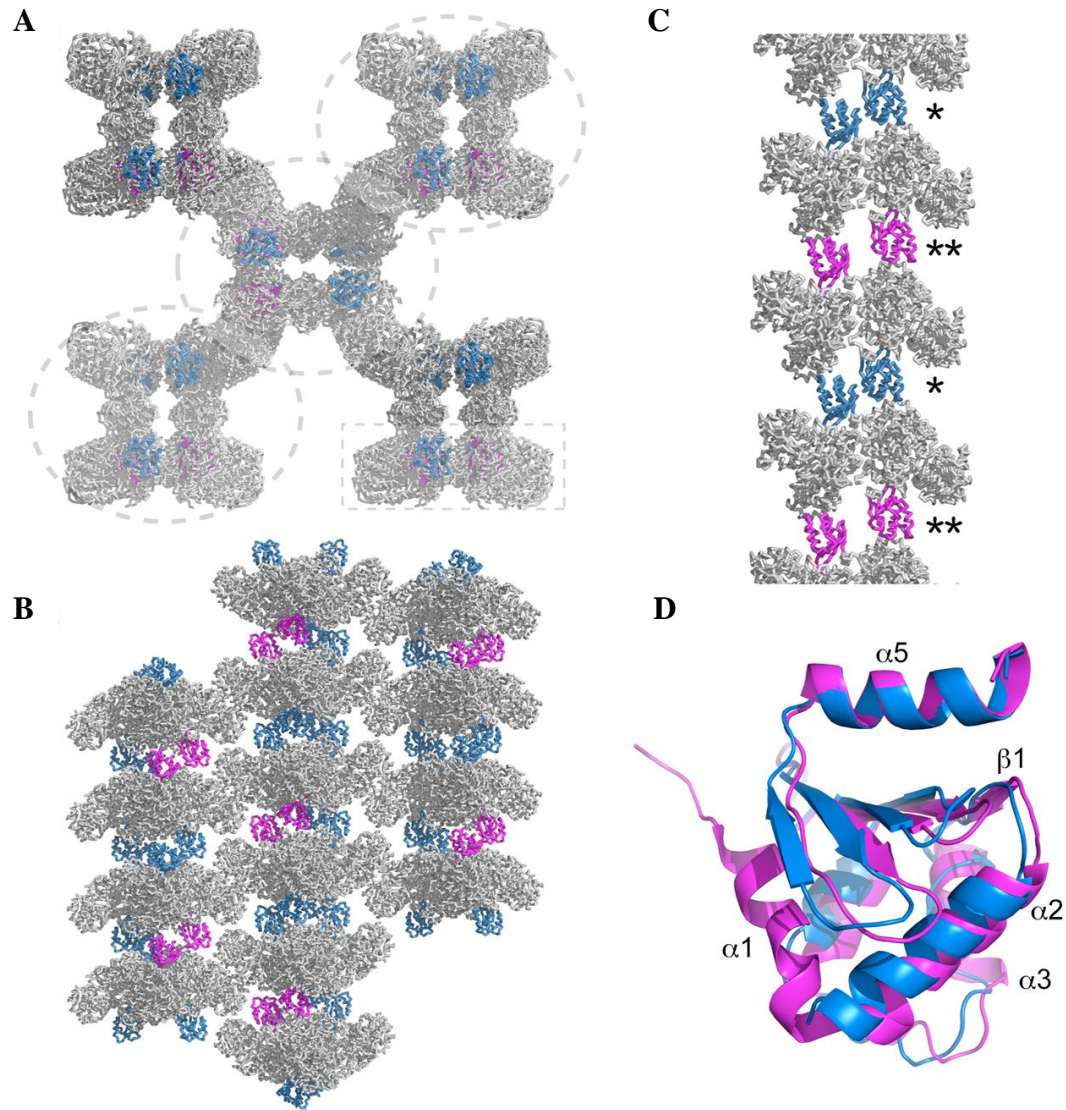


Figure 4.3 A packing diagram of the AtSus1 tetramers in the unit cell. The protein is shown in a line representation. The cellular targeting domains (CTDs) are colored in magenta for subunits B and H, and blue for subunits A and C-G; the remaining portions of the AtSus1 polypeptides are grey. In A, the view down the *c*-axis shows five stacks of tetramers; the stacks bounded by ellipses are shown from the side in B to highlight how each stack interdigitates with its neighbors. The half stack bounded by the rectangle in A is shown from the side in C to highlight how the CTDs change conformation from one layer (*) to the next (**). In D, a ribbon diagram shows the conformational differences in the CTDs (relative to the EPBD and the GT-B domains) between subunits B and H (magenta) and the remaining 6 subunits (blue).

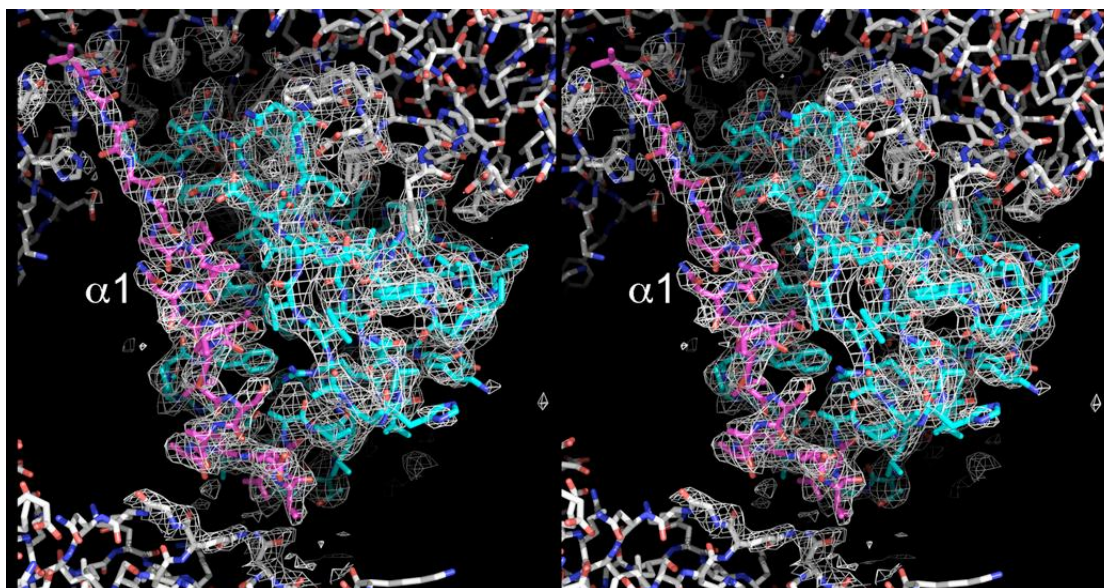


Figure 4.4 A stereo view of the CTD in subunit H in AtSus1 complexed with UDP and fructose. The polypeptide is represented as sticks: the carbon atoms of the helix $\alpha 1$ in the CTD are colored in magenta, the rest of the CTD is cyan, and the polypeptide chain outside of the CTD is grey. The region of the 2Fo-Fc electron density map surrounding the CTD is contoured at 0.95 σ .

The AtSus1 polypeptide chain (residues 11-808) folds into a tri-lobed structure with four distinct domains (Figure 4.5A). The first two domains have been designated a cellular

targeting domain (CTD, residues 11-127) and an ENOD40 peptide-binding domain (EPBD, residues 157-276). The final two domains comprise the two Rossmann-fold domains of a canonical GT-B glycosyltransferase [37]. The observed secondary structure of the CTD starts at residue 16 with two α -helices (helix $\alpha 1$ and $\alpha 2$; Figure. 4.3D), which are oriented roughly perpendicular to each other. Helix $\alpha 2$ is connected to a very short β -strand ($\beta 1$), which is then followed by another two helices (helix $\alpha 3$ and $\alpha 4$). With $\beta 1$, the next four β -strands create a 5 stranded, anti-parallel β -sheet (strands $\beta 1$ - $\beta 5$), which give the appearance of a platform from which two α -helical loops (helices $\alpha 1/\alpha 2$ and $\alpha 3/\alpha 4$) dangle. Although the primary sequence in the helices of the CTD differs between SUS from different plant species, the core of the β -sheet (strands $\beta 2$ - $\beta 5$) shows a high degree of sequence conservation (Figure. 4.1). The AtSus1 monomers exhibit only minor deviations from 222 NCS symmetry within the tetramer, except for the CTD where significant conformational differences are observed (Figure 4.3C). In both crystal structures, the CTD domains in the B and H subunits exhibit a rigid body rotation of ~ 6 degrees with respect to the other domains, compared to their orientation in the remaining 6 subunits. The helix $\alpha 5$ lies across the surface of the β -sheet, which seems to pivot on the helix surface (Figure 4.3D). More precise measurements of domain movement are difficult because of the high level of disorder in the CTD domains (Table 4.2). The overall fold of the CTD bears no resemblance to previously known folds, as judged by the DALI program [53]. From helix $\alpha 5$, the polypeptide meanders for 28 residues (128-156) to the EPBD (Figure 4.5A), a relatively compact bundle of six α -helices. A putative potassium ion binding site (Figure 4.5A) is located at the C-terminal end of helix $\alpha 7$ and coordinated by 5 backbone carbonyl oxygens (L184, R185, H187, L194, L196) and the O $\delta 1$ atom of Asn193. The primary sequence conservation in the linker and EPBD is much higher between SUS from different plants

than that found in the CTD (Figure 4.1). The overall fold of the EPBD shows only a very weak similarity to ubiquitin when analyzed with the DALI program [53].

Figure 4.5

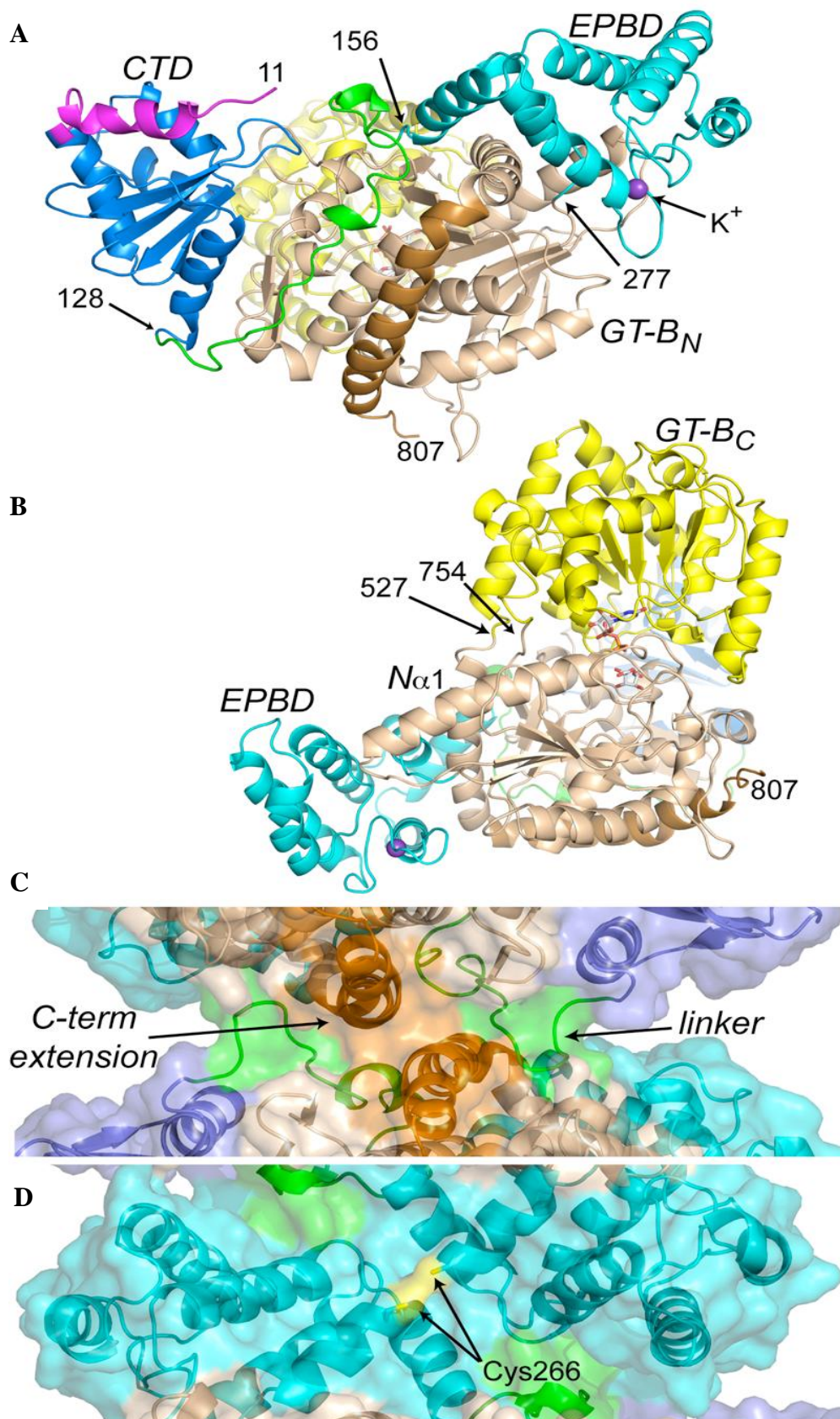


Figure 4.5 Views of the overall fold of AtSus1 and its subunit interfaces. In panel **A**, helix $\alpha 1$ of the subunit H (magenta) is clearly seen as it begins the CTD. At residue 128, the CTD-EPBD linker (green) meanders to residue 156, which starts the helical EPBD (cyan); a putative K^+ ion (violet) is seen bound to the EPBD. At residue 278, the polypeptide enters the N-terminal GT-B domain (GT-B_N, in wheat). A rotated view of the monomer (panel **B**) shows the GT-B domains, GT-B_N (wheat) and GT-B_C (yellow) with UDP and fructose in the active site within the interdomain cleft. The GT-B hinge is located at residues 527 and 754. Note the long N α 1 helix of the GT-B_N domain, which extends from the active site to the EPBD. In panel **C**, the A:D subunit interface is essentially formed by C-terminal extension helix (copper) and part of the CTD-EPBD linker (green). In panel **D**, the A:B subunit interface is formed by EPBD (cyan) and the CTD-EPBD linker (green). Cys266 (yellow), the site of thiolation by the ENOD40 peptides, lies in a groove near the molecular 2-fold axis.

As the polypeptide chain exits helix $\alpha 11$, a 5-residue stretch separates the EPBD from the first β -strand of a GT-B glycosyltransferase, a bi-domain structure where each domain is comprised of a Rossmann fold [37]. As the structural topology and secondary structural elements of the GTB enzymes are quite conserved [37], we will use GT-B_N and GT-B_C to refer to the N- and C-terminal domains, respectively, of the GT-B glycosyltransferase. Likewise, the conserved secondary structural elements of the GT-B domain will be referred to by the nomenclature defined by Ha et al. [54]; e.g., N α 1 would refer to the first α -helix in the GT-B_N domain (Figure 4.6). The initial portion of the GT-B_N domain in AtSus1 extends from residues 278-526 and the GT-B_C domain from residues 527-754. As typical of all known GT-B glycosyltransferases [37], the interdomain hinge in AtSus1 (Figure 4.5B) consists of two linkers, one between the two domains and one between two final C-terminal α -helices (C α 7 and N α 6). The polypeptide then ends with a helical, 37-amino acid C-terminal extension (Figures 4.5A and 4.5B).

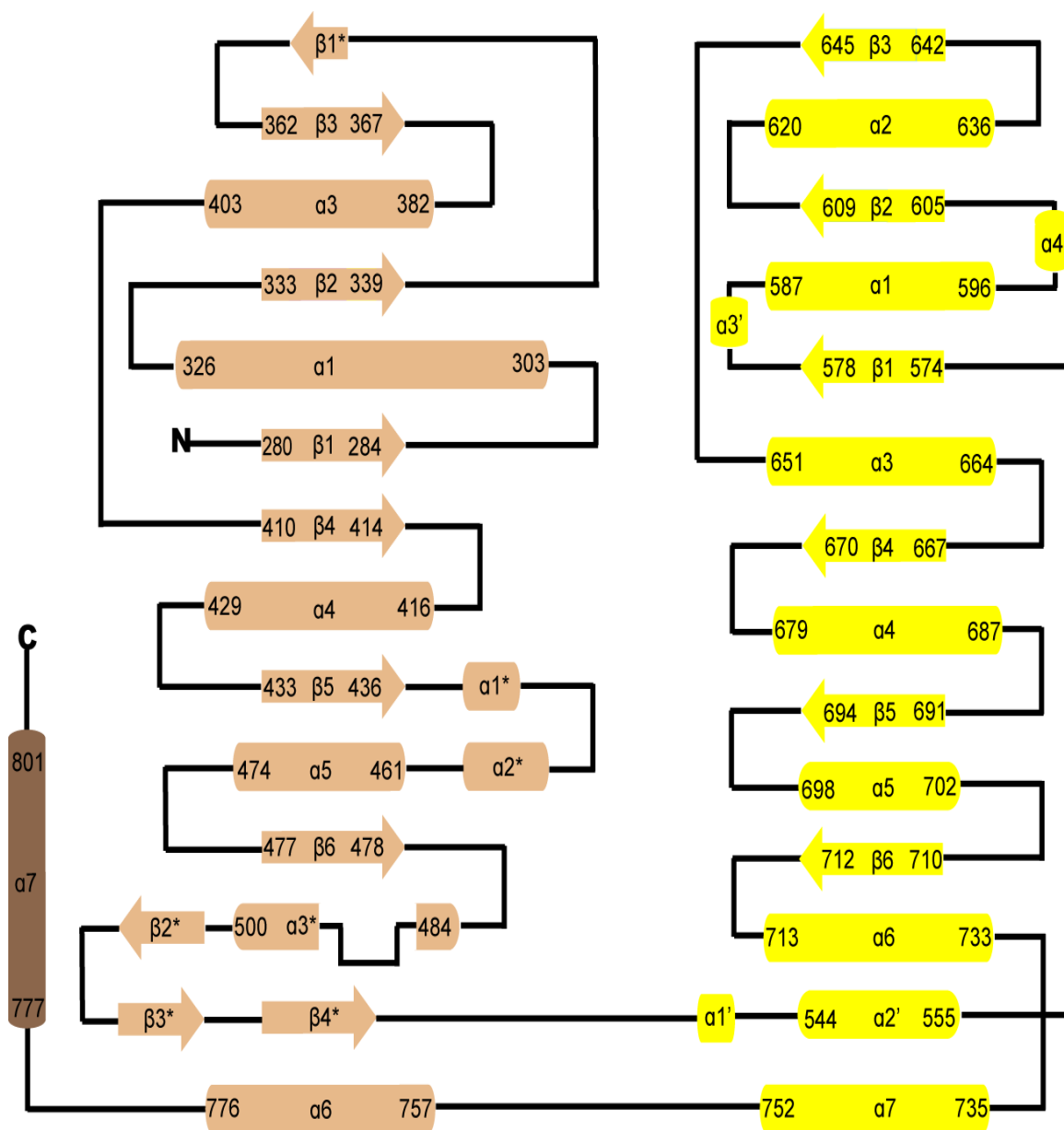


Figure 4.6 Topology diagram of the GT-B glycosyltransferase domain in AtSus1. The secondary structural elements for the GT-B_N domain are shown in wheat and for the GT-B_C domain are in yellow. The nomenclature for denoting the secondary structural elements is derived from Ha et al. [54] for MurG, except for the helical extension (helix $\alpha 7$) shown in brown, which is unique to sucrose synthases.

Although the glycosyltransferase domain has all the conserved secondary structural elements of the GT-B glycosyltransferases, there are several structural elaborations not seen in other GTB enzyme structures. In the GT-B_N domain, the Rossmann-fold deviates from its canonical form by the replacement of the Nα2 helical return by the β-strand Nβ1*, which makes several antiparallel interactions with the linker region between the CTD and the EPBD (Figure. 4.5A). Another unique feature of GTB-N is that helix Nα1 is ~10 Å longer than the homologous helix in other GT-B glycosyltransferases. As a result, the C-terminal end sticks out of the GT-B_N domain and abuts against the EPBD (Figure. 4.5B). One consequence of this fold is that helix Cα7, the helix adjacent to the second hinge site (residues 753-754), is held in place by the Nα1 helix of GT-B_N and helices α6 and α11 in the EPBD domain. In contrast, the GT-B_C domain generally follows the canonical Rossmann fold seen in the other GT-B glycosyltransferases. The active site of AtSus1 lies within the cleft region formed by both GT-B_N and GT-B_C (Figure 4.5B).

4.3.2 Implications of the AtSus1 quaternary structure

The biophysical analyses on SUS from maize [44], mung bean [55], and potato [56] strongly suggested that its native oligomeric state is a tetramer, although SUS may exist in a dimer form [5]. The initial selection of an AtSus1 tetramer created a rather flat oligomer that exhibited excellent noncrystallographic 222-symmetry (Figure 4.2). This tetramer displays only two sets of subunit interfaces (A:B and C:D or A:D and B:C), which results in a large hole in the center of the oligomer (Figure 4.2A). As other tetramer arrangements could be generated, the monomer interactions within the crystal were further analyzed through the use of the PDBePISA server (http://www.ebi.ac.uk/msd-srv/prot_int/). The analysis confirmed that the initial choice of the tetramer produced the most compact and

most plausible oligomer. The average surface area covered by the A:D interface is $\sim 1280 \text{ \AA}^2$ (only about 4% of a monomer's total surface area) with a predicted solvation free energy gain $\Delta^i G$ of $-18.8 \text{ kcal mol}^{-1}$. In contrast, the A:B interface covers less average surface area ($\sim 1076 \text{ \AA}^2$) and is much less hydrophobic ($\Delta^i G = -6.4 \text{ kcal mol}^{-1}$). Interestingly, neither the A:B nor the A:D interfaces involve significant interactions between the GT-B domains.

The subunit interfaces in the proposed tetramer (Figure 4.2) were also examined to see if the tetramer was consistent with the known biochemical data. The N-terminal half of the CTD-EPBD domain linker (residues 131-142) and portions of the C-terminal extension (residues 778-796) create the A:D interface (Figure 4.5C). Although the majority of contacts are between these two polypeptide segments, residues 131-134 of the linker also contact the surfaces of helices N α 3 and N α 4 in the GT-B_N domain. One prediction that can be made is that removal of the C-terminal extension would seriously compromise this subunit interface. Hardin et al. [38] demonstrated that C-terminal truncation of recombinant maize SUS1 resulted in the expressed protein forming dimers, which is consistent with the A:D interface being part of the native AtSus1 tetramer.

The A:B interface (Figure 4.5D) is created almost entirely from interactions between adjacent EPBDs, although the C-terminal half of the CTD-EPBD domain linker (residues 147-154) also makes a few contacts with helix α 10 of a neighboring EPBD. The A:B interface creates a long groove with Cys266 sitting $\sim 3 \text{ \AA}$ away from the dyad axis within this groove. The ENOD40 peptides A and B are small, hormone-like peptides that are involved in root nodule organogenesis in legumes [46]. These peptides bind to SUS from maize [40] and soybean [45], and Röhrig et al. [57] have demonstrated that the dodecapeptide ENOD40-A specifically thiolates Cys264 of soybean SUS, which is

equivalent to Cys266 in AtSus1. Even though the equivalent residue in maize SUS1 is Ala261, Hardin and colleagues [40] observed that ENOD40 peptides not only bound tightly to SUS1, but also inhibited the phosphorylation of Ser170. The phosphorylation of Ser170 in SUS1 promotes SUS turnover by enhancing the ubiquitinylation of SUS, followed by its subsequent degradation [40, 44]. Ser167, the equivalent residue in AtSus1, lies within the A:B interface at the N-terminal end of $\alpha 6$ (Figure 4.1). In the intact native tetramer, Ser167 is inaccessible to phosphorylation, but would be accessible if the tetramer dissociates into dimers upon disruption of the A:B interface, a reasonable hypothesis given that this interface is predicted to be much less hydrophobic than the A:D interface. The dissociation into dimers would then allow its phosphorylation and increase the potential for ubiquitinylation and turnover. Hence, the *in vitro* effects on ENOD40 peptides on SUS phosphorylation can be completely explained by their interaction at A:B interface.

4.3.3 Sucrose synthase active site

As expected for a GT-B family glycosyltransferase, the ligands (UDP-Glc or UDP and fructose) bound in the cleft between the GT-B_N and GT-B_C domains (Figure 4.5B). No significant global conformational differences with respect to the GTB domains were observed between the AtSUS1/UDP/fructose and AtSUS1/UDP-glucose complexes. In the AtSUS1/UDP/fructose complex, UDP and D-fructose were clearly resolved in each active site within the asymmetric unit; what appeared to be UDP-glucose was also well resolved in each active site of the AtSUS1/UDP-glucose complex. As monomer F had the lowest average B-factor for both complexes among the eight monomers (Table 4.2), the discussion will refer to this subunit.

The UDP moiety in both complexes is bound in an identical manner (Figures 4.7 and 4.8), primarily through interactions with the GT-B_C domain (loops C β 1 \rightarrow α 1, C β 3 \rightarrow α 3, and C β 4 \rightarrow α 4, as well as helices C α 3 and C α 4); the only interactions with the GT-B_N domain involve hydrogen bonds from the backbone amide nitrogen of Gly303 to O1B and O3B of the β -phosphate. The face of the uracil ring stacks against the methyl group of Met578, while the edge of the ring makes three hydrogen bonds: the uracil atoms O4, N3, and O2 to the amide N of Gln648, to the carbonyl O of Gln648, and to N δ 2 of Asn654, respectively. The O2' and O3' hydroxyls of the ribose ring hydrogen bond to the carboxylate oxygens of the Glu685. The α -phosphate of UDP is bound to the protein through hydrogen bonds from the O1A and O2A atoms to backbone amide nitrogens of Thr680 and Leu679, respectively. However, the O1A atom also participates in a hydrogen bond chain with the amide nitrogen of Arg580 that is mediated by two conserved water molecules in the active site. Finally, the N η 1 atom of Arg580 contacts the β -phosphate via the O1B and O2B oxygens, while the N ζ nitrogen of Lys585 interacts with O2B and O3A atoms. Overall, the UDP moiety binds in a similar manner to AtSus1 as ADP in *E. coli* glycogen synthase (ecGS), although the ribose ring of ADP is liganded by Asp21 from the GT-B_N domain[58] and not by a residue from the GT-B_C domain.

Figure 4.7

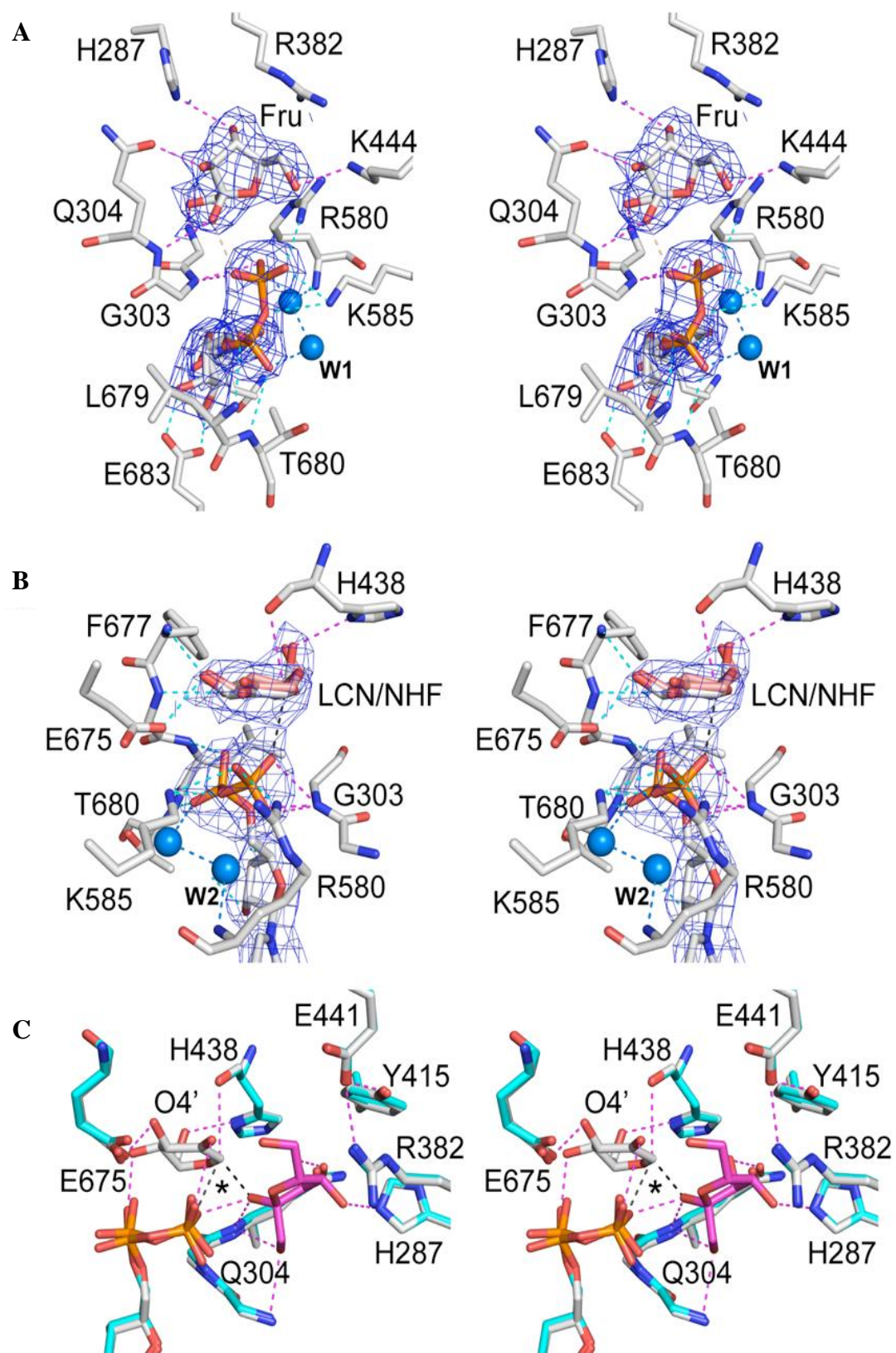


Figure 4.7 Stereo views of the active site in AtSus1. Panel **A** shows the interactions between the protein, UDP, and fructose (Fru) in the UDP/fructose complex. Potential hydrogen bonds from the GT-B_N domain are shown in magenta, from GT-B_C domain in cyan. Note a potential hydrogen bond (wheat) from the O3B phosphate oxygen to the O2 hydroxyl of fructose. Panel **B** shows the interactions between the protein, UDP, and a glucosyl intermediate in the UDP-glucose complex. Potential hydrogen bonds from the protein to the ligands are colored as in **A**. The two analogs for the oxocarbenium ion intermediate, lichenan (LCN) and 1,5-anhydro fructose (NHF), are shown in grey and salmon, respectively. The O3B phosphate oxygen is too far away from the C1 anomeric carbon of LCN for a covalent bond (2.9 Å; black dotted line). The 2F_o-F_c electron density maps (blue) shown in panels **A** and **B** are contoured at 1.8 σ; the two conserved water molecules (blue) are shown as spheres. After superposition of the UDP/fructose (cyan) and UDP-glucose (grey) complexes, panel **C** shows a composite view of the active site; fructose is shown in magenta and LCN in grey. Note that C1 anomeric carbon of LCN (asterisk) is approached on the α face by O2 hydroxyl of fructose (2.3 Å), and O3B of the UDP β-phosphate (2.9 Å), shown as black dotted lines; on the β face, the C1 carbon is in close proximity to the backbone carbonyl of His438 (3.3 Å).

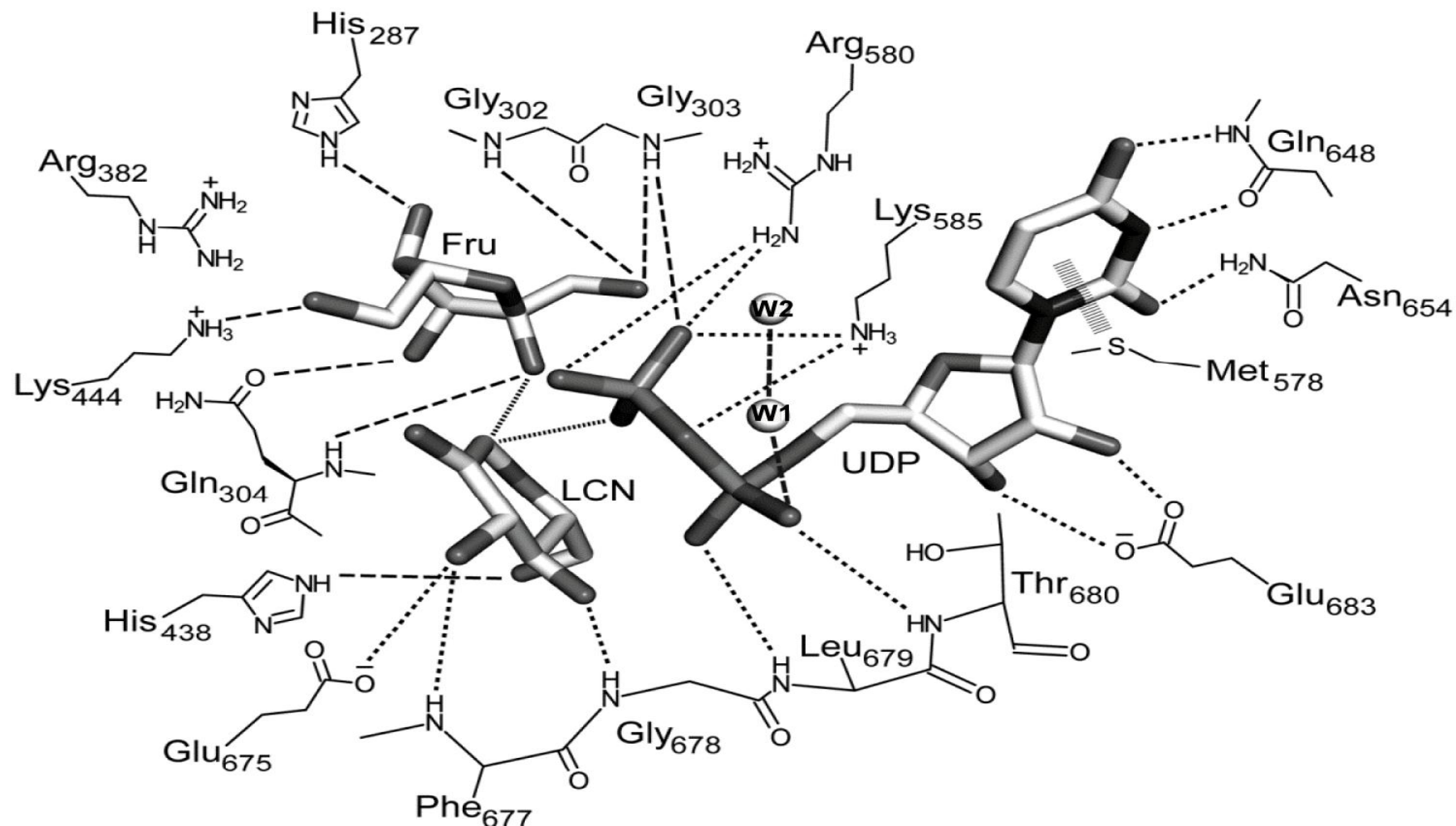


Figure 4.8 Schematic diagram of potential hydrogen-bonding interactions of bound UDP, LCN and fructose with the protein, from the **superposition of the two complexes**. All potential interactions are within 3.3 Å. Circles represent bound water molecules W1 and W2. Hydrogen bonds from GT-B_N domain shown as long dashed lines (----); hydrogen bonds from GT-B_C domain are shorter dashed lines (.....) dotted lines (.....) show the interactions between leaving and attacking groups to the C1 anomeric carbon of LCN.

In the AtSUS1/UDP/fructose complex, fructose is firmly bound in the β -furanose form within a pocket formed exclusively by residues of the GT-B_N domain (Figures 4.7 and 4.8). Each hydroxyl of fructose is hydrogen bonded, which may help maximize substrate specificity. The O1 hydroxyl interacts with the amide nitrogen atoms of Gly302 and Gly304, the O2 hydroxyl binds to the amide nitrogen of Gln304, and the O3 hydroxyl binds to the O ϵ 1 oxygen of Gln304. Finally, the O4 hydroxyl is bound to N ϵ 2 nitrogen of His287. The only interaction with a charged side chain is made between the O6 hydroxyl and the N ζ atom of Lys444. Although no hydrogen bonds are made, the guanidinium group of Arg382 sits within ~ 3.6 Å of O4 and C6 (Figure 4.7A).

The modeling of the glucose moiety in the AtSUS1/UDP-glucose complex was much more challenging, despite the slightly higher resolution and excellent electron density. Cycles of model building and refinement, followed by omit difference map analyses, consistently revealed a significant gap between the O3B β -phosphate atom of the UDP and the C1 anomeric carbon of the glucose moiety in all monomers within the asymmetric unit. Attempts to fit the glucose moiety of an intact UDP-Glc into the rather flat electron density always resulted in marked distortion of the stereochemistry at the anomeric carbon and the appearance of significant negative electron density in $F_o - F_c$ difference maps. The liganded species was then modeled as a mixture of UDP-Glc, UDP, and a separate glucosyl species. The proposed mechanisms for transfer of the donor saccharide in the retaining glycosyltransferases fall into three distinctly different types [37]: (1) a double-displacement mechanism that predicts the formation of a covalent enzyme-glycosyl intermediate, (2) an S_Ni mechanism that predicts a transition state with a transient oxocarbenium ion pair, and (3) an S_Ni -like mechanism that combines transient ion pair intermediate with a conformational

shift as the incoming acceptor attacks (Figure 4.9). Despite the growing amount of structural information on retaining glycosyltransferases, a preponderance of evidence favoring one mechanism over the others is still lacking. The S_Ni -like catalytic mechanism has been proposed for several retaining glycosyltransferases from the metal-dependent GT-A and the metal-independent GT-B enzyme families [37, 59]. However, Soya et al. [60] recently presented evidence showing that the blood group A and B synthesizing glycosyltransferases GTA and GTB form covalent enzyme-glycosyl intermediates, which supports the double-displacement mechanism. Against this background of catalytic mechanisms, we evaluated several different scenarios for UDP and glucose within the AtSus1 active site.

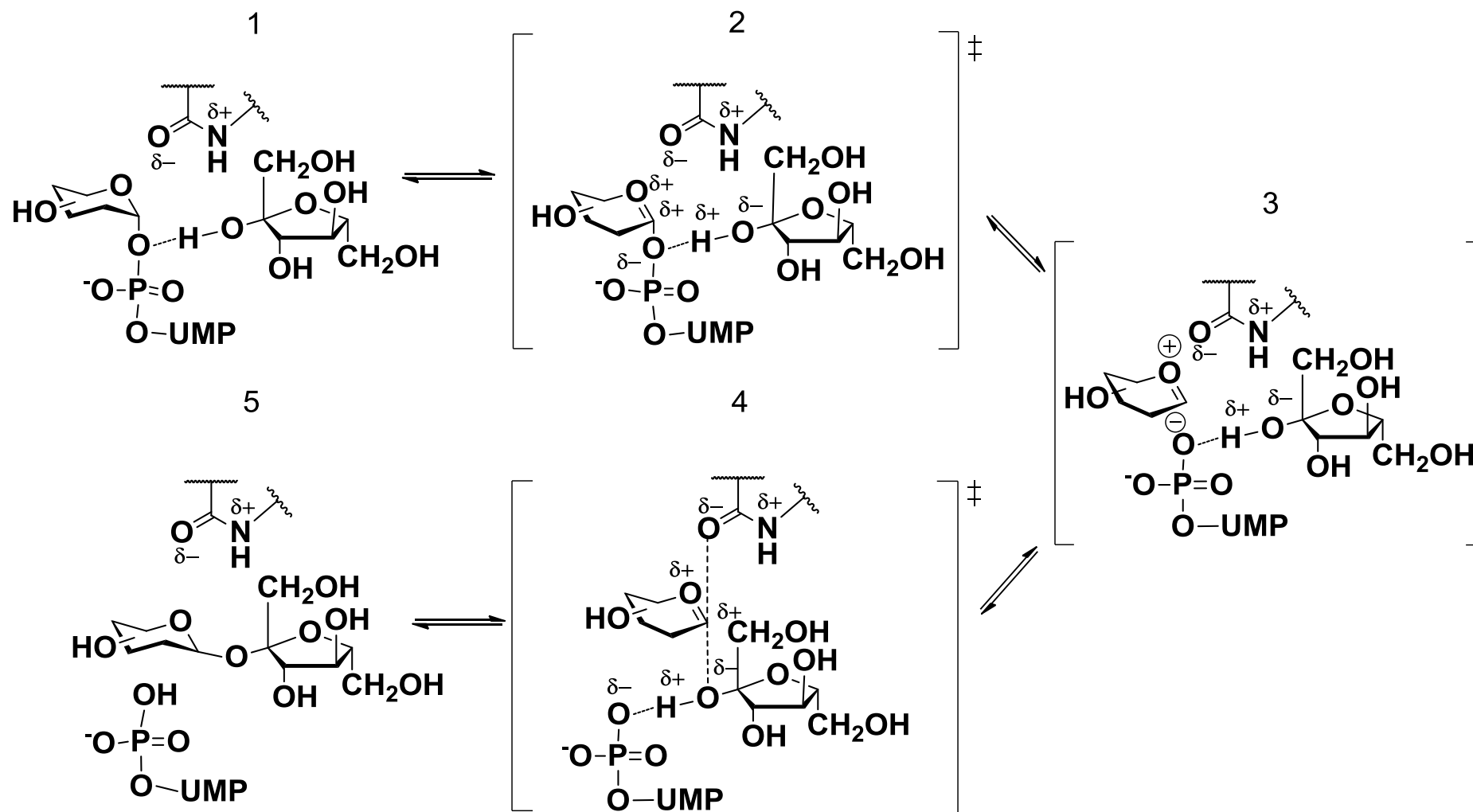


Figure 4.9 A diagram of the S_Ni -like reaction scheme (adapted from Lairson *et al.* [37]). As SUS binds either UDP-glucose and fructose (step 1) or UDP and sucrose (step 5), the ternary ES complex rapidly generates a stabilized oxocarbenium phosphate ion pair intermediate (step 3). The collapse of the intermediate towards either sucrose cleavage (step 1) or sucrose synthesis (step 5) depends on subtle changes in position of the oxocarbenium ion with respect to UDP (step 2) or fructose (step 4).

The initial evaluation unequivocally ruled out the existence of an enzyme-glucosyl intermediate, even at a low level of occupancy. Rather, the flat electron density around the C1 anomeric carbon suggested a somewhat distorted glucosyl species, similar to the glucosyl intermediate observed in EcGS [58]. Two glucose derivatives were chosen that mimic the oxocarbenium-phosphate ion pair intermediate predicted by the S_Ni -like catalytic mechanism (Figure 4.9): 1,5-anhydro fructose (NHF), which is a natural product [61], and the tautomer of NHF, lichenan (LCN), which is a C2 deprotonation product of the D-glucopyranosylium (oxocarbenium ion) and contains a C1-C2 double bond. The primary difference between NHF and LCN is the extent of the planar portion of the ring; i.e., O1-C1-C2-O2-C3 atoms lie in a plane in LCN, but C1-C2-O2-C3 atoms lie in a plane in NHF. After parallel cycles of refinement and model building using various combinations of UDP-Glc and UDP with the individual glucosyl analogs, the results suggest that UDP-Glc or free glucose probably contributes to less than 10% of the electron density of the glucosyl moiety. On the other hand, a mixture of LCN and NHF can account for the observed electron density (Figure 4.7B) when at roughly a 60:40 ratio in each monomer. Thus, in the absence of an acceptor, the glucosyl binding site seemingly traps a mixture of tautomeric intermediates after UDP-Glc hydrolysis (step 3 in Figure 4.9). For simplicity, interactions between the glucose moiety and AtSus1 will be described using LCN, although the NHF displayed the identical intermolecular interactions with the protein and UDP. The protein hydrogen bonds with LCN at four points (Figures 4.7B and 4.4): both the Oε1 oxygen of Glu675 and amide nitrogen of Phe677 ligand to O3' hydroxyl, the amide of Gly678 binds to the O4' hydroxyl, and the Nε2 nitrogen of His 438 binds to the 6-hydroxyl.

Both complexes are in the identical closed, ligand-bound form; the pair-wise superimposition of the C α atoms between the GT-B domains of the two complexes gives R.M.S. deviations of less than 0.23 Å. Given the high degree of conformational homology between the two complexes, the structures were superimposed to provide a composite view of the ligands with respect to each other (Figure 4.7C). The composite view also shows that the C1 carbon is in close proximity to the backbone carbonyl oxygen of His438 (3.3 Å), O2 hydroxyl of fructose (2.3 Å), and O3B of the UDP β -phosphate (2.9 Å). Moreover, the attacking group (O2 hydroxyl of fructose) and the leaving group (O3B phosphate oxygen) are both on the α face of LCN, while the carbonyl of His438, which may stabilize the partial positive charge at the C1 carbon, is on the β face. The AtSUS1/UDP-glucose complex also suggests that the stabilized glucosyl intermediate makes intimate interactions with the pyrophosphate of UDP (Figure 4.7C): the O2' hydroxyl interacts with O2B and the O4' hydroxyl interacts with O2A.

At this limited resolution, it is difficult to provide a more detailed analysis of the AtSus1 complexes. Nonetheless, it is clear that there is no suitable protein residue in the immediate area surrounding LCN and fructose (Figure 4.7C) to act as a glucosyl acceptor, as would be expected for the S_N2 reaction mechanism, and that UDP-glucose is cleaved into UDP and an unknown glucosyl species that is not D- β -glucose (Figure 4.7B). Moreover, this glucosyl intermediate is also well stabilized by interactions with the protein and the pyrophosphate of UDP (Figures 4.7C and 4.8). In fact, the interactions of His438 and Glu675 with LCN are identical to those made by His161 and Glu377 in EcGS to the 1,5-anhydrosorbitol (Figure 4.10), the model for the putative glucosyl intermediate observed in EcGS [37, 59]. Finally, the O3B oxygen of UDP and the C1 carbon of fructose are in a

position to attack the same of face of LCN. These results are consistent with the S_Ni -like reaction mechanism proposed for retaining glycosyltransferases [37, 59](Figure 4.9). What makes these observations surprising is the long-lived nature of this glycosyl intermediate. Few structures of catalytic intermediates have been reported for either the GT-A or GT-B retaining glycosyltransferases. Although a short-lived intermediate was reportedly observed in ternary structures of MalP [62], Sheng and colleagues [58] have provided the only convincing demonstration of a stable glycosyl intermediate in the GT-B retaining glycosyltransferase EcGS. Why SUS would generate a stable glycosyl intermediate in the absence of fructose is an open question. However, SUS can reversibly transfer glucose between UDP and fructose; virtually all other retaining glycosyltransferases catalyze irreversible reactions. Perhaps by stabilizing the glucosyl intermediate in an enzyme complex with UDP, the direction of the reaction can more readily respond to changes in substrate concentrations. Hence, mass action can more effectively shift SUS between sucrose cleavage and sucrose synthesis, as is required for plant development and growth [1, 2].

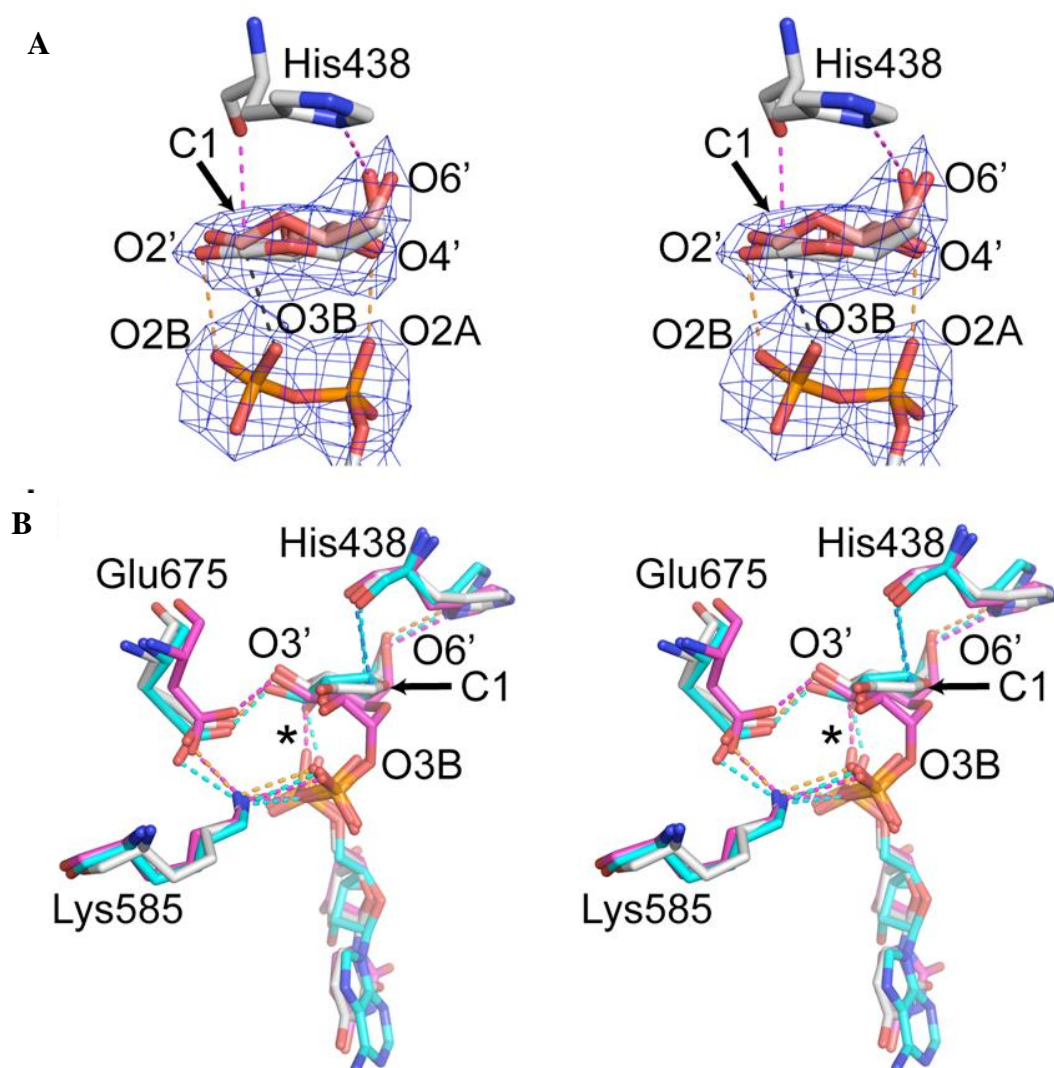


Figure 4.10 Stereoview of the conserved interactions in the active site of AtSus1 and the retaining glycosyltransferases glycogen synthase from *E. coli* (EcGS; PDB 3GUH) [58] and UDP-glucosyltransferase OtsA from *Mycobacterium tuberculosis* (PDB 3C4V) [63]. (A): Electron density of modeled breakdown product of UDP-glucose LCN and NTF. (B): Overlay of active site structure. Each of these enzyme complexes was crystallized in the presence of UDP-glucose. The UDP moieties for AtSus1 (grey) and OtsA (magenta) overlay well with the ADP moiety from EcGS (cyan), but the glucosyl intermediates observed in AtSus1 and EcGS [58] are displaced from the glucosyl moiety in OtsA due to its intact O3B-C1 bond. The glucosyl moieties are liganded by the protein in a conserved manner: the O3' hydroxyl by a carboxylate (Glu675 in AtSus1, Glu377 in EcGS, and Asp361 in OtsA) and the O6' hydroxyl by a histidine (His438 in AtSus1, His161 in EcGS, and His154 in OtsA). Interestingly, in AtSus1 and EcGS, the amide nitrogens of His 438 (AtSus1) and His161 (EcGS) sit less than 3.3 Å away from the C1 anomeric carbon, as expected for a role in stabilizing the intermediate. Also note how a conserved lysine coordinates (Lys585 in AtSus1, Lys305 in EcGS, and Lys267 in OtsA) to the conserved carboxylate and to the pyrophosphate oxygens. Finally, the α-phosphate ligands the O4' hydroxyl (*) in all three enzyme-ligand complexes.

4.3.4 Insights into the evolution of the retaining and inverting GT-B glycosyltransferases

A major unanswered question concerns the evolution of the inverting and retaining reaction mechanisms within the GT-A and GT-B superfamilies [37]. As noted by Martinez-Fleites *et al.* [64], the loop between C β 4 and C α 4 differs markedly between the inverting and retaining glycosyltransferases in the GT-B superfamily (Figure 4.11). In the retaining GT-B glycosyltransferases (blue in Figure 4.11), the C β 4-C α 4 loop is quite long and essentially blocks the pyrophosphate and glycosyl moieties of the NDP-sugar donor from entering this area (Figure 4.6A). For AtSus1, a portion of the C β 4-C α 4 loop (residues 675-680) participates in the binding of the UDP-glucose (Figure 4.8). In contrast, the much shorter C β 4-C α 4 loop in the inverting subfamily allows the NDP-sugar donor to bind across the front of the C α 4 helix (Figure 4.10B), which may allow the positive end of the helix dipole to assist in liganding the pyrophosphate group [65, 66]. An additional feature was noted when comparing AtSus1 with other inverting and retaining glycosyltransferases in the GT-B superfamily. In the inverting GT-B glycosyltransferases, the N β 1-N α 1 loop is also quite short and extends outward from N β 1 before returning to begin the N α 1 helix (Figure 4.11B). As a consequence, the acceptor molecule lies away from the mouth of N α 1 helix. In the retaining GT-B glycosyltransferases, however, the N β 1-N α 1 loop is much longer, but extends up and over N β 1 and N α 1 before beginning the first turn of the N α 1 helix (Figure 4.11A). Now the acceptor molecule can lie adjacent to the mouth of N α 1 helix, and in a position to be affected by the positive end of the N α 1 helix dipole. The end of the N β 1-N α 1 loop (residues 302-302) actually participates in the binding of fructose and the pyrophosphate (Figure 4.8).

The analysis of multiple sequence alignments between GT-B glycosyltransferases consistently reveal a higher degree of conservation in GT-B_C than GT-B_N. One explanation is that the diversity of the sugar acceptors requires a more adaptable binding pocket. When both domains of inverting and retaining GTs are structurally aligned, in most cases, the superpositions give relatively high R.M.S. deviations for equivalenced C α position, which often prevents a more detailed analysis. However, smaller R.M.S. deviations for C α positions can be obtained when each individual GT-B domain is superimposed. This prompted us to use an alternative procedure to superimpose the inverting and retaining GTs. First, the GT-B_C domains of AtSus1 and MurG (PDB 1NLM), an inverting GT, are superimposed to highlight any differences between the two GT-B_N domains. The GT-B_N domain of AtSus1 alone is then superimposed onto the GT-B_N domain of MurG. This manner of comparing the GT-B_N domain of AtSus1 to an inverting GT would allow the elucidation of possible rigid body movements that do not arise simply from the conformational changes required to open and close the active site cleft. The only caveat is the GT-B structures to be superimposed must all be in the catalytically-competent “closed” state.

The resulting comparison is shown in Figure 4.12 with the original AtSus1 structure colored in blue and the AtSus1 structure after rotation colored in dark red. As expected, the UDP moieties in AtSus1 and MurG superimposed very well; the binding the NDP-sugar donor is generally defined by the GT-B_C domain. The question to ask next is how the GT-B_N domain differ with respect to the NDP-sugar donor. I observed striking differences in the placement of the GT-B_N domains, relative to the GT-B_C domains, such that there is the 34 degree rotation of the N α 1 helix between inverting and retaining GTs. As clearly shown in Figure 4.12 the N β 1-N α 1 loop in AtSus1 forms the binding pocket for the fructose and

the N β 5-N α 5 loop, which occupies the β face of the NDP sugar ring to stabilize the C1 anomeric carbon, essentially blocks any attack of a sugar acceptor from the opposite (inverting) direction. Conversely, when the GT-B_N domain of AtSus1 is superimposed directly onto the GT-B_N domain of MurG, the position of the N β 1-N α 1 loop blocks a sugar acceptor from the “retaining” direction, but the N β 5-N α 5 loop is now rotated away to make room for the inverting sugar acceptor. Two conclusions can be drawn from my structural superposition experiments. First, GT-B_N domains in inverting and retaining GTs are quite structurally homologous, an observation made by many others [67]. The second conclusion is that despite this structural homology, the relative rotation of the GT-B_N domain relative to the GT-B_C domain positions the N β 1-N α 1 loop and the N β 5-N α 5 loop and it may be this difference in tertiary structure that determines the direction from which the sugar acceptor will attack the C1 anomeric carbon. The GT-B_N domain rotation in concert with the restriction of the NDP-sugar by the C β 4-C α 4 loop distinguishes the inverting mechanism from retaining mechanism. Further investigation is still needed to determine what structural feature causes the GT-B_N rotation. Detailed comparison of the hinge regions may provide more information concerning the underlying structural differences between the inverting and retaining GTs.

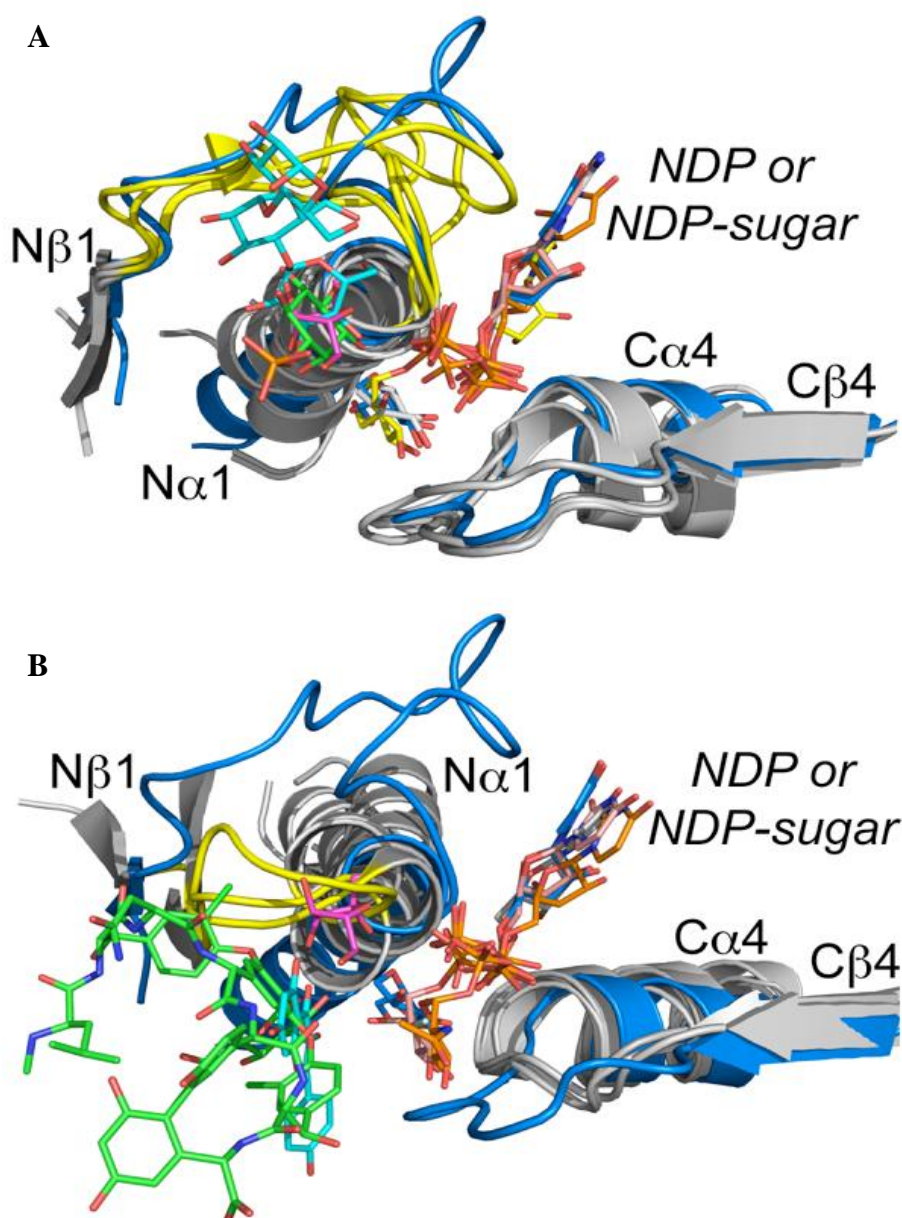


Figure 4.11 Structural comparison of inverting and retaining GTs. In **A**, the retaining glycosyltransferases AtSus1 (blue), WaaG (PDB 2IW1), EcGS (PDB 3GUH, 3CX4), and MshA (PDB 3C4V) were superimposed by using strand Cβ4, helix Cα4, and the nucleotide diphosphate moieties. Note how the Nβ1-Nα1 loops for WaaG, EcGS, and MshA (all yellow) follow the same path, while the sugar acceptors (α -1,4-glucan, cyan; L-myoinositol-1-phosphate, green; fructose, magenta) are clustered near the α face of the sugar ring of the NDP-sugar donor and at the mouth of the Nα1 helix. In **B**, the inverting glycosyltransferases GtfD (PDB 1RRV), MurG (PDB 1NLM), VvGT1 (PDB 2C1Z) were superimposed with AtSus1 with the same protocol. Note how the sugar acceptors (kaempferol, cyan; desvancosaminyl vancomycin, green) are clustered near β face of the sugar ring of the NDP-sugar donor and positioned away from the mouth of the Nα1 helix. The Nβ1-Nα1 loops (yellow) and the Cβ4-Cα4 loops in the inverting glycosyltransferases (**B**) are much shorter than the corresponding loops in the retaining glycosyltransferases (**A**).

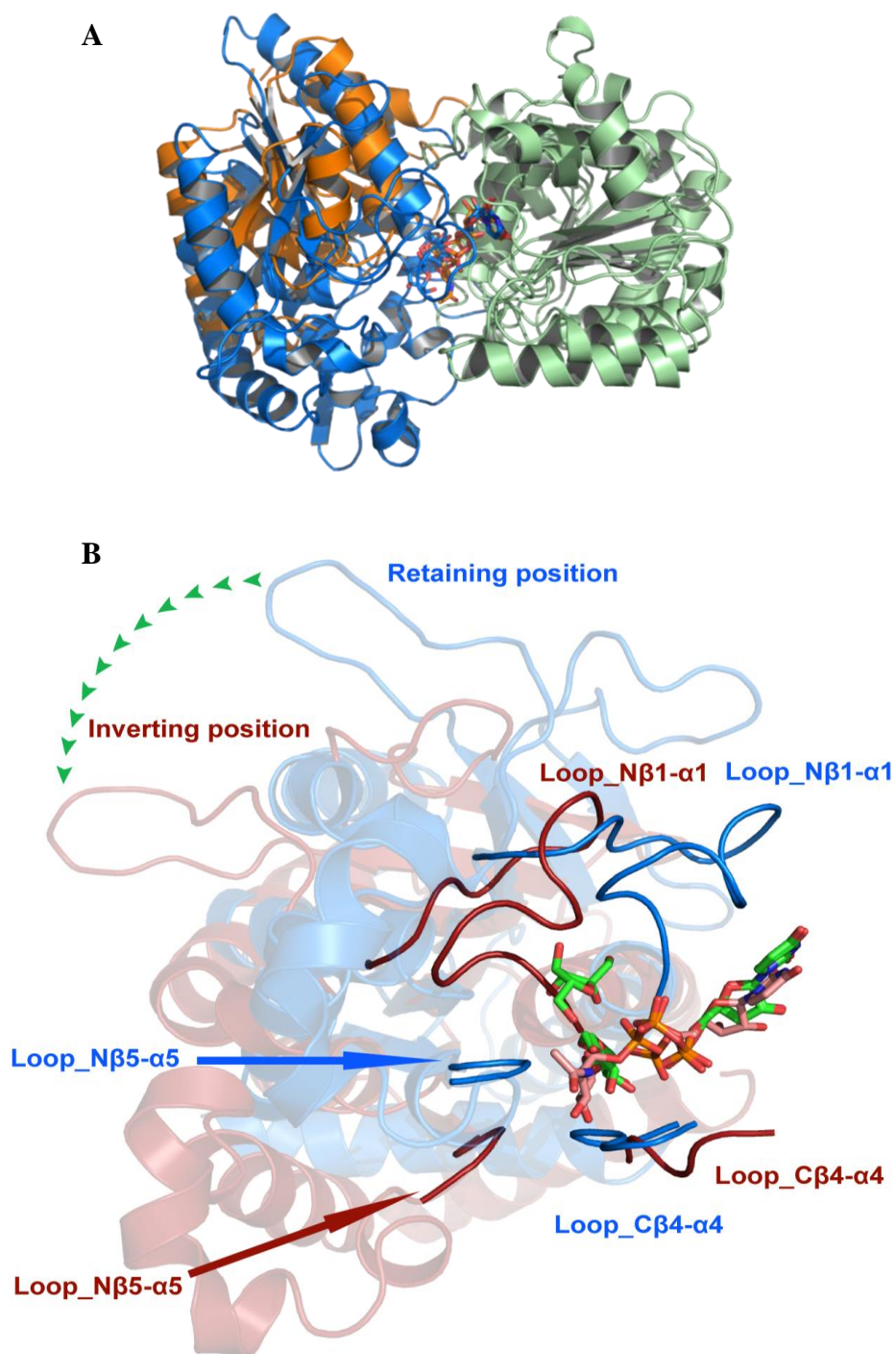


Figure 4.12 Structural comparison of AtSus1 GT-B_N domain in the retaining and inverting positions. (A) Superimposition of the GT-B_C domains (pale green) of the AtSus1 and EcMurG structures. The resulting deviation in the placement of the two GT-B_N domains is shown by coloring AtSus1 (blue) and EcMurG (orange) differently. (B) The original AtSus1 structure is colored in blue while the structure after rotation (see text for details) is colored in dark red. Retaining sugar acceptor fructose as well as sugar donor UDP-glucose are shown in green sticks. Inverting sugar donor UDP-N-acetylglucosamine is shown in salmon stick. The loops around the active site are highlighted. The rotation path is labeled with a chain of green arrows.

4.3.5 Functional implications of the AtSus1 structure

A preponderance of evidence clearly shows that SUS binds to membranes, to membrane-bound complexes for cellulose and callose biosynthesis, and to cytoskeletal elements. How SUS interacts with different cellular targets is unknown, but the presence of sucrose [5, 38], thiolation of Cys266 by ENOD40-A [57], and the phosphorylation of SUS [5, 40, 43, 55] seem to alter its binding to cell membranes and F-actin, as well as its sucrose cleavage activity. In maize SUS1, the phosphorylation of Ser15 [5, 43] shifts the oligomerization of SUS1 from dimers to tetramers and increases SUS1 binding to membranes, but markedly diminishes SUS1 binding to F-actin. The residue equivalent to Ser15 in AtSus1 is Ser 13 (Figure 4.1), which lies in a disordered region in all but one subunit: subunit H in the AtSUS1/UDP/fructose complex (Figure 4.4). Although the main chain is fairly well resolved in this region, the electron density for the side chains is very poor. Nonetheless, Ser 13 in subunit H would lie next to helix $\alpha 8$ in the EPBD of subunit H (Figure 4.13A). Although the level of primary sequence identity is lower in the CTD and EPBD domains relative to the GTB domain, there are clear regions of conservation (Figure 4.1). It is premature to speculate on the detailed interactions between the N-terminus and helix $\alpha 8$, but the phosphorylation of Ser13 would certainly alter the electrostatic environment in this region. The conserved sequence surrounding Ser13 is cationic (¹⁰R-X-H-S-X-R/K-E-R¹⁷), while the neighboring helix $\alpha 8$ has both anionic and cationic patches (²¹¹L-K-R-A-E-E-Y-L²¹⁶). Would a phosphorylated Ser13 promote its interaction with Arg 211 or Lys212 or perhaps disrupt interactions with Glu214 or Glu215? Regardless, $\alpha 1$

easily becomes disordered, and the entire CTD can move as a rigid body (Figure 4.3). Mapping the movement of the CTD onto all the subunits, the overall impact of this conformational change is to move the CTDs closer or farther apart across an approximately 50 Å wide groove on the face of the AtSus1 tetramer (Figure 4.13B); the $\alpha 9$ helices of two EPBDs form the base of the groove. The overall dimensions and shape of the groove are quite complementary to the size and topography of the F-actin fiber (Figure 4.13C). This qualitative model for AtSus1 binding to F-actin proposes that the CTD domains would bind within the groove of the F-actin, where the $\alpha 1$ helix of the CTD and helix $\alpha 9$ in the EPBD may directly mediate contact with the actin fiber.

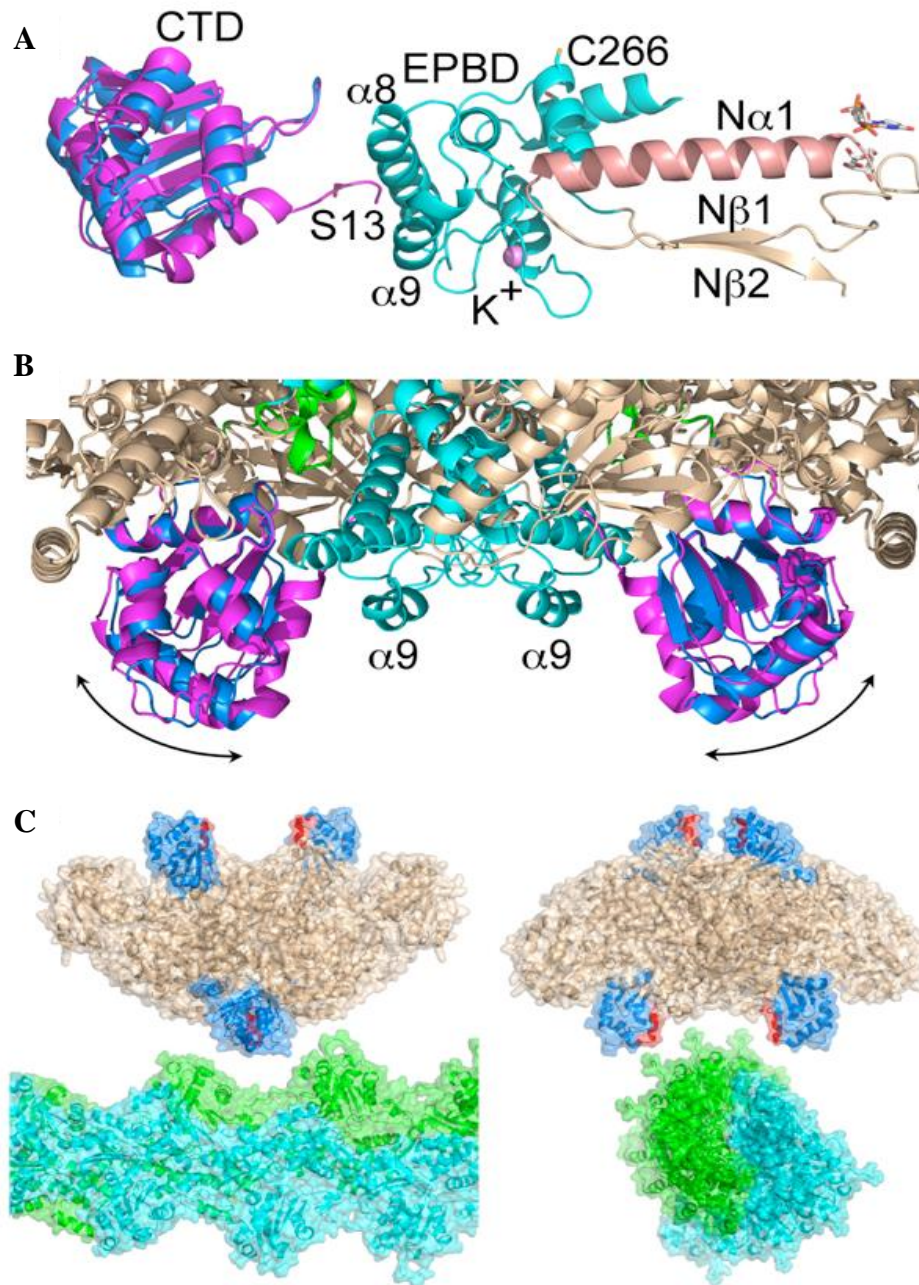


Figure 4.13 Juxtaposition of the CTD and EPBD domains with the AtSus1 active site. (A) Ser13, the site of phosphorylation, lies against helix $\alpha 8$ of the EPBD. Nearby is Cys266, the site of thiolation by the ENOD40 peptides, and the C-terminal end of helix $N\alpha 1$, which starts within the active site. Conformational changes originating at the CTD or A:B interface could be transmitted to the active site via helix $N\alpha 1$. (B) The observed conformational changes in the CTD essentially change the width of a groove on the face of the AtSus1 tetramer; helix $\alpha 9$ of the EPBD makes the top of the groove. (C) The AtSus1 tetramer is juxtaposed with a double-stranded F-actin filament generated from PDB 3MFB (rabbit F-actin). One F-actin strand is green and the other is cyan; the CTD of AtSus1 is colored blue, and the bulk of AtSus1 is wheat. The CTD is the size of the F-actin groove and helix $\alpha 1$ of the CTD (magenta) would lie against the F-actin fiber.

The structure of AtSus1 does not shed light on how SUS interacts with the membrane or other cellular sites or why phosphorylation of Ser13 enhances SUS binding to membranes, but diminishes its binding to actin [5]. However, the structure of AtSus1 suggests how CTD interaction with cellular targets, SUS phosphorylation at Ser13, and the thiolation by ENOD40 peptides may alter SUS catalysis. As shown Figure 4.12A, the unusually long N α 1 helix (303-326) of the GT-B_N extends out from the active site. The N-terminal end of N α 1 interacts with pyrophosphate and fructose (Figures. 4.8 and 4.13A); the C-terminal end of N α 1 abuts against helix α 10 of the EPBD (Figure 4.13A). Adjacent to helix α 10 are helix α 9, which may interact with actin, and helix α 8, which interacts with the region surrounding Ser13; nearby is Cys266, the site of thiolation by ENOD40 peptide A. The juxtaposition of potential regulatory sites around the EPBD suggests that conformational changes in this domain could be readily transmitted into the active site through distortions of the N α 1 helix. In mung bean SUS1, Nakai and colleagues noted that the phosphorylation of Ser11, the equivalent of Ser13 in AtSus1, increases sucrose affinity [55], suggesting that phosphorylation may enhance sucrose cleavage, i.e., UDP-glucose generation. They also found that mutation of Ser13 to glutamate mimics phosphorylation. Therefore, mutagenic analysis of the residues in the EPBD, at the C-terminal end of N α 1, and surrounding Ser13 may provide crucial insights in the regulation of SUS function and enzymatic activity.

4.4 Conclusions

The structure determination of AtSus1 provided the structural basis for the regulation and the assembly of a eukaryotic glycosyltransferase. The structures in complex with substrates also unveiled the evidence for the S_Ni like enzyme mechanism proposed for the retaining glycosyltransferases in the previous studies [37]. Two novel folds revealed in the structures will provide the knowledge base for the rational design of experiments to investigate their roles in Sus1's spatial regulation as well as the effects of stabilization by the tetrameric assembly on the enzyme activity and its physiological rules. The structural comparison of the inverting and retaining glycosyltransferases pinpointed some simple but important structural features, such as the loop variation and the domain rotation, which may be essential for pre-determining the inverting or retaining mechanism and the outcome of the glycosylation. Further experiments are needed on how the two hinges connecting the two GTB sub domains contribute to the reactions. This information would greatly enhance our understanding of the glycosyltransferases and also help the design of compounds or antibodies to modulate the activity of a specific glycosyltransferase.

REFERENCES

REFERENCES

- 1 Koch, K. (2004) Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Curr Opin Plant Biol.* **7**, 235-246
- 2 Salerno, G. L. and Curatti, L. (2003) Origin of sucrose metabolism in higher plants: when, how and why? *Trends Plant Sci.* **8**, 63-69
- 3 Angeles-Nunez, J. G. and Tiessen, A. (2010) Arabidopsis sucrose synthase 2 and 3 modulate metabolic homeostasis and direct carbon towards starch synthesis in developing seeds. *Planta.* **232**, 701-718
- 4 Coleman, H. D., Yan, J. and Mansfield, S. D. (2009) Sucrose synthase affects carbon partitioning to increase cellulose production and altered cell wall ultrastructure. *Proc Natl Acad Sci U S A.* **106**, 13118-13123
- 5 Duncan, K. A. and Huber, S. C. (2007) Sucrose synthase oligomerization and F-actin association are regulated by sucrose concentration and phosphorylation. *Plant Cell Physiol.* **48**, 1612-1623
- 6 Haigler, C. H., Ivanova-Datcheva, M., Hogan, P. S., Salnikov, V. V., Hwang, S., Martin, K. and Delmer, D. P. (2001) Carbon partitioning to cellulose synthesis. *Plant Mol Biol.* **47**, 29-51
- 7 Ruan, Y. L., Chourey, P. S., Delmer, D. P. and Perez-Grau, L. (1997) The Differential Expression of Sucrose Synthase in Relation to Diverse Patterns of Carbon Partitioning in Developing Cotton Seed. *Plant Physiol.* **115**, 375-385
- 8 Salnikov, V. V., Grimson, M. J., Seagull, R. W. and Haigler, C. H. (2003) Localization of sucrose synthase and callose in freeze-substituted secondary-wall-stage cotton fibers. *Protoplasma.* **221**, 175-184
- 9 Schubert, M., Koteyeva, N. K., Wabnitz, P. W., Santos, P., Buttner, M., Sauer, N., Demchenko, K. and Pawlowski, K. (2010) Plasmodesmata distribution and sugar partitioning in nitrogen-fixing root nodules of *Datisca glomerata*. *Planta*

- 10 Albrecht, G. and Mustroph, A. (2003) Localization of sucrose synthase in wheat roots: increased in situ activity of sucrose synthase correlates with cell wall thickening by cellulose deposition under hypoxia. *Planta*. **217**, 252-260
- 11 Baroja-Fernandez, E., Munoz, F. J., Montero, M., Etxeberria, E., Sesma, M. T., Ovecka, M., Bahaji, A., Ezquer, I., Li, J., Prat, S. and Pozueta-Romero, J. (2009) Enhancing sucrose synthase activity in transgenic potato (*Solanum tuberosum* L.) tubers results in increased levels of starch, ADPglucose and UDPglucose and total yield. *Plant Cell Physiol*. **50**, 1651-1662
- 12 Coleman, H. D., Beamish, L., Reid, A., Park, J. Y. and Mansfield, S. D. (2010) Altered sucrose metabolism impacts plant biomass production and flower development. *Transgenic Res*. **19**, 269-283
- 13 Etxeberria, E. and Gonzalez, P. (2003) Evidence for a tonoplast-associated form of sucrose synthase and its potential involvement in sucrose mobilization from the vacuole. *J Exp Bot*. **54**, 1407-1414
- 14 Zrenner, R., Salanoubat, M., Willmitzer, L. and Sonnewald, U. (1995) Evidence of the crucial role of sucrose synthase for sink strength using transgenic potato plants (*Solanum tuberosum* L.). *Plant J*. **7**, 97-107
- 15 Winter, H. and Huber, S. C. (2000) Regulation of sucrose metabolism in higher plants: localization and regulation of activity of key enzymes. *Crit Rev Biochem Mol Biol*. **35**, 253-289
- 16 Barratt, D. H., Derbyshire, P., Findlay, K., Pike, M., Wellner, N., Lunn, J., Feil, R., Simpson, C., Maule, A. J. and Smith, A. M. (2009) Normal growth of *Arabidopsis* requires cytosolic invertase but not sucrose synthase. *Proc Natl Acad Sci U S A*. **106**, 13124-13129
- 17 Bieniawska, Z., Paul Barratt, D. H., Garlick, A. P., Thole, V., Kruger, N. J., Martin, C., Zrenner, R. and Smith, A. M. (2007) Analysis of the sucrose synthase gene family in *Arabidopsis*. *Plant J*. **49**, 810-828
- 18 Cai, G., Faleri, C., Del Casino, C., Emons, A. M. and Cresti, M. (2011) Distribution of callose synthase, cellulose synthase, and sucrose synthase in tobacco pollen tube is controlled in dissimilar ways by actin filaments and microtubules. *Plant Physiol*. **155**, 1169-1190

- 19 Persia, D., Cai, G., Del Casino, C., Faleri, C., Willemse, M. T. and Cresti, M. (2008) Sucrose synthase is associated with the cell wall of tobacco pollen tubes. *Plant Physiol.* **147**, 1603-1618
- 20 Baier, M. C., Barsch, A., Kuster, H. and Hohnjec, N. (2007) Antisense repression of the *Medicago truncatula* nodule-enhanced sucrose synthase leads to a handicapped nitrogen fixation mirrored by specific alterations in the symbiotic transcriptome and metabolome. *Plant Physiol.* **145**, 1600-1618
- 21 Baier, M. C., Keck, M., Godde, V., Niehaus, K., Kuster, H. and Hohnjec, N. (2010) Knockdown of the symbiotic sucrose synthase MtSucS1 affects arbuscule maturation and maintenance in mycorrhizal roots of *Medicago truncatula*. *Plant Physiol.* **152**, 1000-1014
- 22 Wachter, R., Langhans, M., Aloni, R., Gotz, S., Weilmunster, A., Koops, A., Temguia, L., Mistrik, I., Pavlovkin, J., Rascher, U., Schwalm, K., Koch, K. E. and Ullrich, C. I. (2003) Vascularization, high-volume solution flow, and localized roles for enzymes of sucrose metabolism during tumorigenesis by *Agrobacterium tumefaciens*. *Plant Physiol.* **133**, 1024-1037
- 23 Ishimaru, T., Hirose, T., Matsuda, T., Goto, A., Takahashi, K., Sasaki, H., Terao, T., Ishii, R., Ohsugi, R. and Yamagishi, T. (2005) Expression patterns of genes encoding carbohydrate-metabolizing enzymes and their relationship to grain filling in rice (*Oryza sativa* L.): comparison of caryopses located at different positions in a panicle. *Plant Cell Physiol.* **46**, 620-628
- 24 Jiang, Q., Hou, J., Hao, C., Wang, L., Ge, H., Dong, Y. and Zhang, X. (2011) The wheat (*T. aestivum*) sucrose synthase 2 gene (TaSus2) active in endosperm development is associated with yield traits. *Funct Integr Genomics.* **11**, 49-61
- 25 Tang, T., Xie, H., Wang, Y., Lu, B. and Liang, J. (2009) The effect of sucrose and abscisic acid interaction on sucrose synthase and its relationship to grain filling of rice (*Oryza sativa* L.). *J Exp Bot.* **60**, 2641-2652
- 26 Bologa, K. L., Fernie, A. R., Leisse, A., Loureiro, M. E. and Geigenberger, P. (2003) A bypass of sucrose synthase leads to low internal oxygen and impaired metabolic performance in growing potato tubers. *Plant Physiol.* **132**, 2058-2072

- 27 Yang, J., Zhang, J., Wang, Z., Xu, G. and Zhu, Q. (2004) Activities of key enzymes in sucrose-to-starch conversion in wheat grains subjected to water deficit during grain filling. *Plant Physiol.* **135**, 1621-1629
- 28 Bolton, J. J., Soliman, K. M., Wilkins, T. A. and Jenkins, J. N. (2009) Aberrant Expression of Critical Genes during Secondary Cell Wall Biogenesis in a Cotton Mutant, Ligon Lintless-1 (Li-1). *Comp Funct Genomics*, 659301
- 29 Gardiner, J. C., Taylor, N. G. and Turner, S. R. (2003) Control of cellulose synthase complex localization in developing xylem. *Plant Cell.* **15**, 1740-1748
- 30 Hennen-Bierwagen, T. A., Lin, Q., Grimaud, F., Planchot, V., Keeling, P. L., James, M. G. and Myers, A. M. (2009) Proteins from multiple metabolic pathways associate with starch biosynthetic enzymes in high molecular weight complexes: a model for regulation of carbon allocation in maize amyloplasts. *Plant Physiol.* **149**, 1541-1559
- 31 Duncan, K. A., Hardin, S. C. and Huber, S. C. (2006) The three maize sucrose synthase isoforms differ in distribution, localization, and phosphorylation. *Plant Cell Physiol.* **47**, 959-971
- 32 Subbaiah, C. C., Palaniappan, A., Duncan, K., Rhoads, D. M., Huber, S. C. and Sachs, M. M. (2006) Mitochondrial localization and putative signaling function of sucrose synthase in maize. *J Biol Chem.* **281**, 15625-15635
- 33 Amor, Y., Haigler, C. H., Johnson, S., Wainscott, M. and Delmer, D. P. (1995) A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants. *Proc Natl Acad Sci U S A.* **92**, 9353-9357
- 34 Fujii, S., Hayashi, T. and Mizuno, K. (2010) Sucrose synthase is an integral component of the cellulose synthesis machinery. *Plant Cell Physiol.* **51**, 294-301
- 35 Winter, H., Huber, J. L. and Huber, S. C. (1997) Membrane association of sucrose synthase: changes during the graviresponse and possible control by protein phosphorylation. *FEBS Lett.* **420**, 151-155
- 36 Winter, H., Huber, J. L. and Huber, S. C. (1998) Identification of sucrose synthase as an actin-binding protein. *FEBS Lett.* **430**, 205-208

- 37 Lairson, L. L., Henrissat, B., Davies, G. J. and Withers, S. G. (2008) Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem.* **77**, 521-555
- 38 Hardin, S. C., Duncan, K. A. and Huber, S. C. (2006) Determination of structural requirements and probable regulatory effectors for membrane association of maize sucrose synthase 1. *Plant Physiol.* **141**, 1106-1119
- 39 Barrero-Sicilia, C., Hernando-Amado, S., Gonzalez-Melendi, P. and Carbonero, P. (2011) Structure, expression profile and subcellular localisation of four different sucrose synthase genes from barley. *Planta*, in press
- 40 Hardin, S. C., Tang, G. Q., Scholz, A., Holtgraewe, D., Winter, H. and Huber, S. C. (2003) Phosphorylation of sucrose synthase at serine 170: occurrence and possible role as a signal for proteolysis. *Plant J.* **35**, 588-603
- 41 Huber, S. C., Huber, J. L., Liao, P. C., Gage, D. A., McMichael, R. W., Jr., Chourey, P. S., Hannah, L. C. and Koch, K. (1996) Phosphorylation of serine-15 of maize leaf sucrose synthase. Occurrence in vivo and possible regulatory significance. *Plant Physiol.* **112**, 793-802
- 42 Komina, O., Zhou, Y., Sarath, G. and Chollet, R. (2002) In vivo and in vitro phosphorylation of membrane and soluble forms of soybean nodule sucrose synthase. *Plant Physiol.* **129**, 1664-1673
- 43 Hardin, S. C., Winter, H. and Huber, S. C. (2004) Phosphorylation of the amino terminus of maize sucrose synthase in relation to membrane association and enzyme activity. *Plant Physiol.* **134**, 1427-1438
- 44 Hardin, S. C. and Huber, S. C. (2004) Proteasome activity and the post-translational control of sucrose synthase stability in maize leaves. *Plant Physiol Biochem.* **42**, 197-208
- 45 Rohrig, H., Schmidt, J., Miklashevichs, E., Schell, J. and John, M. (2002) Soybean ENOD40 encodes two peptides that bind to sucrose synthase. *Proc Natl Acad Sci U S A.* **99**, 1915-1920

- 46 Batut, J., Mergaert, P. and Masson-Boivin, C. (2010) Peptide signalling in the rhizobium-legume symbiosis. *Curr Opin Microbiol.* **14**, 181-187
- 47 Kabsch, W. (2010) XDS. *Acta Crystallogr D Biol Crystallogr.* **66**, 125-132
- 48 Collaborative Computational Project, N. (1994) The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D.* **50**, 760-763
- 49 Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* **66**, 213-221
- 50 Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr.* **66**, 486-501
- 51 Notredame, C., Higgins, D. G. and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* **302**, 205-217
- 52 Bond, C. S. and Schuttelkopf, A. W. (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr D Biol Crystallogr.* **65**, 510-512
- 53 Holm, L. and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545-549
- 54 Ha, S., Walker, D., Shi, Y. and Walker, S. (2000) The 1.9 Å crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Sci.* **9**, 1045-1052
- 55 Nakai, T., Konishi, T., Zhang, X. Q., Chollet, R., Tonouchi, N., Tsuchida, T., Yoshinaga, F., Mori, H., Sakai, F. and Hayashi, T. (1998) An increase in apparent affinity for sucrose of mung bean sucrose synthase is caused by in vitro phosphorylation or directed mutagenesis of Ser11. *Plant Cell Physiol.* **39**, 1337-1341

- 56 Romer, U., Schrader, H., Gunther, N., Nettelstroth, N., Frommer, W. B. and Elling, L. (2004) Expression, purification and characterization of recombinant sucrose synthase 1 from *Solanum tuberosum* L. for carbohydrate engineering. *J Biotechnol.* **107**, 135-149
- 57 Rohrig, H., John, M. and Schmidt, J. (2004) Modification of soybean sucrose synthase by S-thiolation with ENOD40 peptide A. *Biochem Biophys Res Commun.* **325**, 864-870
- 58 Sheng, F., Jia, X., Yep, A., Preiss, J. and Geiger, J. H. (2009) The crystal structures of the open and catalytically competent closed conformation of *Escherichia coli* glycogen synthase. *J Biol Chem.* **284**, 17796-17807
- 59 Errey, J. C., Lee, S. S., Gibson, R. P., Martinez Fleites, C., Barry, C. S., Jung, P. M., O'Sullivan, A. C., Davis, B. G. and Davies, G. J. (2010) Mechanistic insight into enzymatic glycosyl transfer with retention of configuration through analysis of glycomimetic inhibitors. *Angew Chem Int Ed Engl.* **49**, 1234-1237
- 60 Soya, N., Fang, Y., Palcic, M. M. and Klassen, J. S. (2011) Trapping and characterization of covalent intermediates of mutant retaining glycosyltransferases. *Glycobiology.* **21**, 547-552
- 61 Yu, S. (2008) The anhydrofructose pathway of glycogen catabolism. *IUBMB Life.* **60**, 798-809
- 62 Geremia, S., Campagnolo, M., Demitri, N. and Johnson, L. N. (2006) Simulation of diffusion time of small molecules in protein crystals. *Structure.* **14**, 393-400
- 63 Gibson, R. P., Tarling, C. A., Roberts, S., Withers, S. G. and Davies, G. J. (2004) The donor subsite of trehalose-6-phosphate synthase: binary complexes with UDP-glucose and UDP-2-deoxy-2-fluoro-glucose at 2 Å resolution. *J Biol Chem.* **279**, 1950-1955
- 64 Martinez-Fleites, C., Proctor, M., Roberts, S., Bolam, D. N., Gilbert, H. J. and Davies, G. J. (2006) Insights into the synthesis of lipopolysaccharide and antibiotics through the structures of two retaining glycosyltransferases from family GT4. *Chem Biol.* **13**, 1143-1152

- 65 Mulichak, A. M., Losey, H. C., Lu, W., Wawrzak, Z., Walsh, C. T. and Garavito, R. M. (2003) Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. *Proc Natl Acad Sci U S A.* **100**, 9238-9243
- 66 Mulichak, A. M., Lu, W., Losey, H. C., Walsh, C. T. and Garavito, R. M. (2004) Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: interactions with acceptor and nucleotide ligands. *Biochemistry.* **43**, 5170-5180
- 67 Franco, O. L. and Rigden, D. J. (2003) Fold recognition analysis of glycosyltransferase families: further members of structural superfamilies. *Glycobiology.* **13**, 707-712

CHAPTER 5

CONCLUSION AND FUTURE DIRECTIONS

5.1 Conclusion of the thesis

The major endeavor of this thesis is the crystallization of several important proteins involved in the host-microbe interactions from model organisms and pathogenic bacteria to provide the knowledge base for better understanding of the interactions between host and microbes. To this end, the structures of AtSus1 and EhaB_c were determined. New structures of AtSus1 in complex with different substrates not only helped to elucidate the retaining glycosyl transfer mechanism but also shed light on the potential structural relationships between the retaining and inverting glycosyltransferases. The structure of the EhaB_c protein from pathogenic *E. coli* O157: H7 was determined to 2.0 Å. Interestingly, EhaB_c has a significantly higher amount of bulky aromatic residues (Trp, Tyr, Phe) lining the pore and a considerably lower pore conductance compared to the NalP protein [1]. On the other hand, preliminary results of a molecular dynamic simulation (data not shown) suggested that the helix inside the EhaB β domain has a tendency to fall back to the periplasmic side of the OM. These new features of EhaB_c will potentially provide additional insight on the biogenesis of AT proteins but future characterizations are still needed. A major effort of this thesis was devoted to the crystallization of BsYpfP but no well-diffracting crystal was obtained so far. However, the exploration of the fusion module to facilitate the crystallization of BsYpfP indeed yielded crystals and proved the concept of

the strategy [2-4]. Better yet, these exercises provided valuable experiences in designing customized crystallization strategies for hard-to-study proteins such as *BsYpfP*. At the same time, the *in vitro* and *in vivo* activities as well as the biophysical and biochemical properties of *BsYpfP* were characterized in detail and will undoubtedly facilitate future crystallization and research on *BsYpfP*.

5.2 Future directions

5.2.1 Future research directions on *BsYpfP*

The extensive crystallization screen and manipulation with the *BsYpfP* fusion proteins eventually gave some clues for the crystallization condition of *BsYpfP*. The encouraging crystal obtained with the TagF-*BsYpfP* fusion protein especially suggested that a proper fusion module may play an important role in the crystallization of *BsYpfP*. The next step will be further exploring more fusion modules to facilitate the crystallization of *BsYpfP*. Despite this being still a trial-and-error effort, the new crystallization robot and imaging system here at Michigan State University will significantly reduce the amount of time and resources needed and increase the possibility of obtaining the crystal of *BsYpfP*. Other approaches to increase the possibility of getting the crystal of *BsYpfP* may include cutting off unstable regions of the protein with information from limited proteolysis [5] and predictions [6] as well as including ligands in the drop to stabilize the crystal contacts.

Besides the structure determination of *BsYpfP* that may provide a detailed mechanism of the glycosyl transfer reaction, biochemical characterization of *BsYpfP* will

also lead to better understanding of the processive glucosyltransferases. In this thesis, a giant liposome assay and a hybrid synthesis of the non-reactive substrate UDP-6-deoxy-glucose have been developed. However, due to lack of a proper separation column, analysis of the activity with this substrate did not proceed further. Continuous work on this subject may result in a more detailed understating of the unique processive glycosyl transfer activity. The giant liposome assay combined with the glycolipid anchor will provide a platform for characterization of downstream enzymes in the biosynthesis of LTAs.

5.2.2 Future research directions on Eha proteins

Crystallization of full length Eha proteins is critical for understanding of the mechanism of translocation. Preliminary screening has already identified promising conditions for EhaA_{fl} and EhaB_{fl} but further optimization is still needed to produce diffraction quality crystals. At the same time, diffraction datasets have been collected for EhaA_c and EhaD_c proteins at resolutions of 2.35 Å and 2.45 Å, but their structure solution by molecular replacement has not yet been successful. However, the structure determination EhaA_c and EhaD_c proteins should be readily obtainable through the use of single or multiple anomalous dispersion methods; this aspect of the work is currently underway. The structures of these closely related Eha proteins may provide additional information on the biogenesis of ATs as well as insights on their role in virulence of *E. coli* O157:H7.

The BLM experiment on the MBP_EhaB_c was only a preliminary study, and the quite interesting results need to be explored in more detail. To be more convincing, the BLM experiment needs to be repeated under a variety of conditions and the BLM experiment on other Eha proteins needs to be carried out to better understand the significant differences in conductivity and its relationship with the bulky residues inside the porin domain.

In the longer term, the structural information obtained from high resolution structures of Eha proteins can be used in the further development of the autodisplay system based on *E. coli* ATs. The technique of autodisplay was introduced in 1986 when Freudl *et al.* showed that heterologous OmpA can be expressed and exposed on the surface of an *E. coli* K-12 strain [7]. Since then, a broad number of display systems have been established for a variety of model organisms and have been widely used as the platform for the development of whole cell biocatalysis, live vaccines, biosorbents, biosensors, inhibitor design and protein/peptide library screening as well as matrices for epitope mapping and antigen delivery [8-12] in both research and industry.

Slightly different strategies are used in each organism and reviewed elsewhere [13]. In *E. coli*, the endogenous AT system is utilized for the display where the natural passenger domain of the AT is replaced by the fragment of interest and thus transferred by the β -domain of the AT and subsequently displayed on the cell surface of the bacterial cell. The autodisplay system in *E. coli* has several outstanding advantages in terms of ease of manipulation, simplicity of the design, and abundance of protein molecules displayed [14],

as well as free diffusion environment to facilitate the observation of protein-protein interaction [14, 15]. The autodisplay system based on *E. coli* has been very successful in a number of cases. The additional structural information of the adhesin Eha proteins will offer even more versatility for this system.

5.2.3 Future research directions on AtSus1

The structure of AtSus1 revealed some potential interactions between CTD with cellular targets such as actin [16, 17], as well as the potential SUS phosphorylation at Ser13 [18, 19], and the function of thiolation at ENOD40 peptides [20], but there are still many unanswered questions. For example, what is the mechanism of SUS interacting with actin, the membrane, or other cellular sites? What is the function of the phosphorylation of Ser13? Will the conformational changes in the EPBD domain be transmitted to the active site through distortions of the N α 1 helix? In mung bean SUS1, Nakai and colleagues noted that the phosphorylation of Ser11, the equivalent of Ser13 in AtSus1, increases sucrose affinity [21], suggesting that phosphorylation may enhance sucrose cleavage, i.e., UDP-glucose generation. They also found that mutation of Ser13 to glutamate mimics phosphorylation. Therefore, mutagenic analysis of the residues in the EPBD, at the C-terminal end of N α 1, and surrounding Ser13 may provide crucial insights into the regulation of SUS function and enzymatic activity. Crystallization of these mutants may provide critical information to answer these questions. On the other hand, in order to further assess the influence of conformational changes in the EPBD domain on the global structure as well as the balance between the sucrose synthesis and cleavage activities, characterization of a solution

structure is necessary. With the high resolution structure at hand, small angle X-ray scattering (SAXS) experiments will be very useful to characterize these conformational changes. In addition, the solution structure may reveal some physiological significance of the equilibrium of tetrameric verse other oligomeric states. A SAXS experiment with a different substrate binding at the active site is already under way. Further detailed characterization and analysis will be carried out in the near future.

REFERENCES

REFERENCES

- 1 Oomen, C. J., van Ulsen, P., van Gelder, P., Feijen, M., Tommassen, J. and Gros, P. (2004) Structure of the translocator domain of a bacterial autotransporter. *EMBO J* **23**, 1257-1266
- 2 Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P. and Kobe, B. (2003) Crystal structures of fusion proteins with large-affinity tags. *Protein Science* **12**, 1313-1322
- 3 Nauli, S., Farr, S., Lee, Y. J., Kim, H. Y., Faham, S. and Bowie, J. U. (2007) Polymer-driven crystallization. *Protein Science* **16**, 2542-2551
- 4 Ahamed, T., Ottens, M., van Dedem, G. W. K. and van der Wielen, L. A. M. (2005) Design of self-interaction chromatography as an analytical tool for predicting protein phase behavior. *Journal of Chromatography A* **1089**, 111-124
- 5 Gao, X., Bain, K., Bonanno, J. B., Buchanan, M., Henderson, D., Lorimer, D., Marsh, C., Reynes, J. A., Sauder, J. M., Schwinn, K., Thai, C. and Burley, S. K. (2005) High-throughput limited proteolysis/mass spectrometry for protein domain elucidation. *J Struct Funct Genomics* **6**, 129-134
- 6 Slabinski, L., Jaroszewski, L., Rychlewski, L., Wilson, I. A., Lesley, S. A. and Godzik, A. (2007) XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **23**, 3403-3405
- 7 Freudl, R., MacIntyre, S., Degen, M. and Henning, U. (1986) Cell surface exposure of the outer membrane protein OmpA of *Escherichia coli* K-12. *J Mol Biol* **188**, 491-494
- 8 Benhar, I. (2001) Biotechnological applications of phage and cell display. *Biotechnology Advances* **19**, 1-33

- 9 Lee, H. W. and Byun, S. M. (2003) The pore size of the autotransporter domain is critical for the active translocation of the passenger domain. *Biochem Biophys Res Commun* **307**, 820-825
- 10 Wernerus, H. and Stahl, S. (2004) Biotechnological applications for surface-engineered bacteria. *Biotechnol Appl Biochem* **40**, 209-228
- 11 Georgiou, G., Stathopoulos, C., Daugherty, P. S., Nayak, A. R., Iverson, B. L. and Curtiss, R. (1997) Display of heterologous proteins on the surface of microorganisms: From the screening of combinatorial libraries to live recombinant vaccines. *Nature Biotechnology* **15**, 29-34
- 12 Jose, J. (2006) Autodisplay: efficient bacterial surface display of recombinant proteins. *Applied Microbiology and Biotechnology* **69**, 607-614
- 13 Mattanovich, D. and Borth, N. (2006) Applications of cell sorting in biotechnology. *Microbial Cell Factories* **5**, -
- 14 Jose, J., Bernhardt, R. and Hannemann, F. (2001) Functional display of active bovine adrenodoxin on the surface of E-coli by chemical incorporation of the [2Fe-2S] cluster. *Chembiochem* **2**, 695-701
- 15 Jose, J. and von Schwichow, S. (2004) Autodisplay of active sorbitol dehydrogenase (SDH) yields a whole cell biocatalyst for the synthesis of rare sugars. *Chembiochem* **5**, 491-499
- 16 Duncan, K. A. and Huber, S. C. (2007) Sucrose synthase oligomerization and F-actin association are regulated by sucrose concentration and phosphorylation. *Plant Cell Physiol* **48**, 1612-1623
- 17 Winter, H., Huber, J. L. and Huber, S. C. (1998) Identification of sucrose synthase as an actin-binding protein. *FEBS Lett* **430**, 205-208
- 18 Hardin, S. C., Winter, H. and Huber, S. C. (2004) Phosphorylation of the amino terminus of maize sucrose synthase in relation to membrane association and enzyme activity. *Plant Physiol* **134**, 1427-1438

- 19 Komina, O., Zhou, Y., Sarath, G. and Chollet, R. (2002) In vivo and in vitro phosphorylation of membrane and soluble forms of soybean nodule sucrose synthase. *Plant Physiol* **129**, 1664-1673
- 20 Rohrig, H., John, M. and Schmidt, J. (2004) Modification of soybean sucrose synthase by S-thiolation with ENOD40 peptide A. *Biochem Biophys Res Commun* **325**, 864-870
- 21 Nakai, T., Konishi, T., Zhang, X. Q., Chollet, R., Tonouchi, N., Tsuchida, T., Yoshinaga, F., Mori, H., Sakai, F. and Hayashi, T. (1998) An increase in apparent affinity for sucrose of mung bean sucrose synthase is caused by in vitro phosphorylation or directed mutagenesis of Ser11. *Plant Cell Physiol* **39**, 1337-1341