博士論文

Studies on structures of novel sugar metabolic enzymes

(新規な糖代謝経路で働く酵素の構造生物学的研究)

平成 24年進学　　南　煐祐

指導教員　　伏信進矢

Index

Abbreviations

AA	Auxiliary Activities
bGalE	UDP-galactose 4'-epimerase from *bifidobacterium longum*
bGalT	Galactose-1-phosphate uridylyltransferase from *bifidobacterium longum*
BGL	β-Glucosidase
CbA	Cellobionic acid (4-O-β-D-glucopyranosyl-D-gluconic acid)
CBAP	Cellobionic acid phosphorylase
CBH	Cellobiohydrolase
CBM	Carbohydrate-binding modules
CDH	Cellobiose dehydrogenase
EG	Endoglucanase
GalE	UDP-galactose 4-epimerase
GalK	Galactokinase
GalM	Galactose mutarotase
GalT	UDP-glucose hexose 1-phosphate uridylyltransferase
GH	Glycoside hydrolase
Glc-*β* 1,3-GlcUA	3-*O*-β-D-glucopyranosyl-D-glucuronic acid
GlcA	Gluconic acid
GlcA	α-D-Gluconic acid
GlcUA	Glucuronic acid
GNB	Gal-*β* 1,3-GalNAc
HMO	Human milk oligosaccharide
LGC	Glucono 1,5-lactone
LNB	Gal-*β* 1,3-GlcNAc
LPMO	Lytic polysaccharide monooxigenase
NahK	*N*-Acetylhexosamine 1-kinase
PDB	Protein Data Bank
PDB	Protein Data Bank
PEG	Polyethylene glycerol

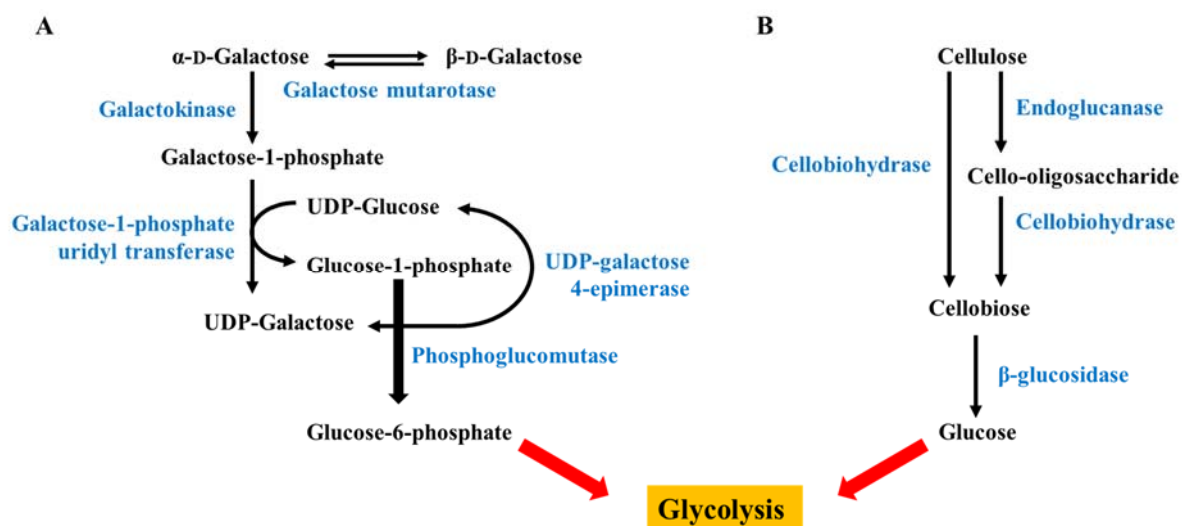|	
RMSD	Root Mean Square Diviation
SAD	Single wavelength Anomalous Diffraction
αGlc1P	α-D-Glucose 1-phosphate

Chapter 1 General Introduction

1-1　Sugar metabolic pathways in microbes

  In biochemistry, sugar metabolic pathways are series of chemical reactions related to extra- and intra-cellular enzymes in organisms. The sugar metabolic pathways play important roles as the maintenance of homeostasis and the creation of new biomolecules by catabolism and anabolism. In all organisms, the metabolic pathways are linked as metabolic networks. Glycolysis, one of the most well-known sugar metabolic pathways in all aerobic and anaerobic organisms, also links to both Leloir pathway (galactose degradation) and cellulose degradation pathway. In these linked pathways, glucose-6-phosphate produced by last step of Leloir pathway and glucose produced by the cellulose degradation pathway can be utilized in glycolysis for generating energy (Fig. 1-1). Both intracellular pathways, glycolysis and Leloir pathway, consist of several enzymes acting as dehydrogenases, transferases, isomeraes, or epimerases. Each enzyme has been characterized and analyzed to understand reaction mechanisms in the pathways though the investigation of functional and structural properties.

  In addition, the discovery of novel enzymes drives advances in understanding new sugar metabolic pathways in organisms. The enzymes in the new sugar metabolic pathways would catalyze specific reactions that involve specific substrates. It is possible that the research is more progressing, the metabolic networks can be more complicated.

Fig. 1-1 Shematic representation of sugar metabolisms linked to glycolysis. (A) Leloir pathway and (B) Cellulose degradation pathway

1-1-1   Cellulose degradation enzymes in fungi

Cellulose is a linear polysaccharide that constitutes plant cell wall (about 50%) as well as the most abundant organic compound in nature. Cellulose consists of β-1,4-linked D-glucopyranosides which form long biopolymer chains. Studies for enzymatic degradation of cellulose have been expected to play a critical role in applications for biofuels and chemicals when considering the production of renewable resources in the future.

The cellulases are wildly found in organisms such as fungi, bacteria, and actinomycetes. Cellulases are often composed of a catalytic domain and one or more carbohydrate-binding modules (CBMs), previously named carbohydrate-binding or cellulose-binding domains (1, 2). A cellulose degradation system in fugi typically consists of three enzymes: cellobiohydrolase (CBH, EC3.2.1.91), endoglucanase (EG, EC 3.2.1.4), and β-glucosidase (BGL, EC 3.2.1.21), as shown in Fig. 1-2. The cellobiohydrolases liberate disaccharide cellbiose from the ends of exposed cellulose polysaccharide while the endoglucanases cleave β-1,4 glycosidic bonds within the cellulose polymer chains (3). Moreover, β-glucosidases catalyze the hydrolysis of the β-glucosidic linkage with release of monosaccharide glucose (4, 5). Additionally, cellobiose produced by the CBH is oxidized by the major oxidoreductase, cellobiose dehydrogenase (CDH, EC 1.1.99.18), producing cellobiono-1,5-lactone. The lactone is hydrolyzed spontaneously in solution to generate cellobionic acid (6, 7). Although this concept is typically accepted, it still remains difficult to understand how the glycoside hydrolases could act on crystalline polysaccharide chains.

The previously classified enzymes as GH61 in the CAZy database of carbohydrate-active enzymes were unknown for hydrolytic activity. The first obtained structure of GH61 was not able to explain the catalytic mechanism, because the carboxylate groups containing amino acid residues (Asp or Glu) were missing (8). However, the genome, transcriptome, proteomic and structural analyses of cellulolytic fungi have identified it as an oxidative enzyme (9-11). The GH61 protein structures have a conserved arrangement of the N-terminal amino group and two histidines which coordinate a copper ion (12, 13). Finally, the functions of these enzyme were revealed as lytic polysaccharide monooxigenases (LPMOs) and led to a reclassification in the CAZy database from GH61 to "Auxiliary Activity (AA)" family 9 (Table 1-1) (14). In fungi, copper-dependent LPMOs have been shown to catalyze the oxidative cleavage of glycosidic bonds on the surface of cellulose without requiring separation of a glucan chain and increase the substrate accessibility for hydrolytic enzymes such as CBHs, and EGs (15). Thus, LPMOs create an entry clefts for GHs, with the hydroxylation of the C1 position of pyranose, an aldonolactone.

Table 1-1 Classification of structure-known LPMOs.

Organisms		Protein Name	Family	PDB
Fungi	*Hypcrea jecorina*	GH61B	AA9	2VTC
	Thielavia terrestris	GH61E	AA9	3EII, 3EJA
	Thermoaucus aurantiacus	GH61A	AA9	2YET, 3ZUD
	Neurospora crassa	PMO-2, PMO-3	AA9	4EIR, 4EIS
Bacteria	*Serratia marcescens*	CBP21	AA10	2LHS, 2BEM, 2BEN
	Vibrio cholera	CBM33	AA10	2XWX
	Enterococcus faecalis	CBM33	AA10	4A02, 4ALC[E,O,R,S,T]
	Burkholderia pseudomallei	CBM33	AA10	3UAM

1-1-2 *Enzymes in Leloir pathway*

Galactose is an essential component of glycoproteins and glycolipids in metazoans and a constituent of the milk sugar, lactose for mammalian infants as a key nutrient. The main pathway of galactose metabolism is the Leloir pathway which was found by Dr. Luis Leloir and was named after the Nobel Prize-winning in part for his contribution to the understanding of galactose metabolism (16). The pathway requires five enzymes to sequentially convert galactose into glucose 6-phosphate (Fig. 1-2). In the reactions of Leloir pathway, the initial step is the conversion of β-D-galactose to α-D-galactose by galactose mutarotase (GalM, EC 5.1.3.3) since this is the active state in this pathway. Next, galactokinase (GalK, EC 2.7.1.6) phosphorylates α-D-galactose to form galactose-1-phosphate. In the third step, galactose-1-phosphate uridylyltransferase (GalT, EC 2.7.7.12) transfers uridine monophosphate from uridine diphosphoglucose to galactose-1-phosphate which is catalyzed by phosphoglucomutase (EC 5.4.2.2) and then linked to the glycolytic pathway. Finally, UDP-galactose 4-epimerase (GalE, EC 5.1.3.2) interconverts UDP-galactose to UDP-glucose.

The GalE enzyme from *E. coli* generally interconverts UDP-galactose to UDP-glucose whereas the human GalE enzyme interconverts both UDP-galactose and UDP-*N*-acetylgalactosamine (UDP-GalNAc) to UDP-glucose and UDP-*N*-acetylglucosamine (UDP-GlcNAc) (17-19). The human GalE not only contributes to the catabolism of dietary galactose, but also reversibly enables the endogenous biosynthesis of both UDP-galactose and UDP-*N*-acetylgalactosamine when exogenous sources are limited. Indeed, the enzymes of the Leloir pathway have attracted significant research attention for well over 30-40 years and the three-dimensional structures of all enzymes have now been defined (Table 1-2 and Fig. 1-2). In detail, the first structure of GalM and GalK from *Lactococcus lactis* was determined by Thoden *et al.* (20-22). The structures of GalT and GalE from *E. coli* were also elucidated (23, 24). Recent advanced researches provide understanding of the function and structure of these enzymes in
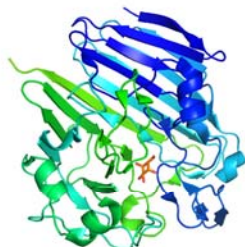
Leloir pathway.

Table 1-2 The three-dimensional structures of enzymes involved in the Leloir pathway from Protein Data Bank

Enzymes	EC Number	Organisms	PDB code[a]
GalM	EC 5.1.3.3	*Lactococcus lactis*	1MMU, 1MMX, 1MMY, 1MMZ, 1L7K
		Homo sapiens	1SNA, 1SNZ, 1SO0
		Bacillus subtilis	4BZE, 4BZF, 4BZG, 4BZH
GalK	EC 2.7.1.6	*Lactococcus lactis*	1PIE
		Homo sapiens	1WUU
		Saccharomyces cerevisiae	2AJ4
		Pyrococcus furiosus	1S4E
GalT	EC 2.7.7.12	*Escherichia coli*	1HXQ
GalE	EC 5.1.3.2	*Escherichia coli*	1LRJ, 1LRK, 1LRL, 1A9Y, 1A9Z
		Homo sapiens	1EK5, 1EK6, 1HZJ
		Archaeoglobus fulgidus	3EHE
		Pseudomonas aeruginosa	1SB8, 1SB9
		Brucella abortus	4TWR
		Aspergillus nidulans	4LIS
		Burkholderia pseudomallei	3ENK

[a]All PDB codes are not shown in the table

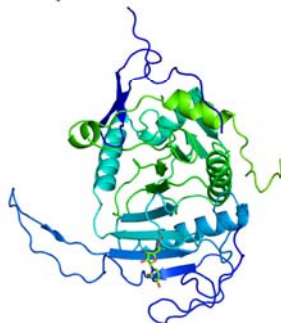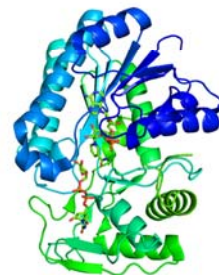
A GalM from *L. lactis* in complex with galactose (PDB 1L7K)

B GalK from *L. lactis* in complex with galactose and phosphate (PDB 1PIE)

C GalT from *E. coli* in complex with UDP (PDB 1HXQ)

D GalE from *H. sapiens* in complex with NADH and UDP-glucose (PDB 1EK6)

Fig. 1-2 The structures of enzymes involved in Leloir pathway.

1-2 Novel sugar metabolic pathways

1-2-1 Intracellular degradation pathway in microbes that is linked to oxidative cellulose degradation
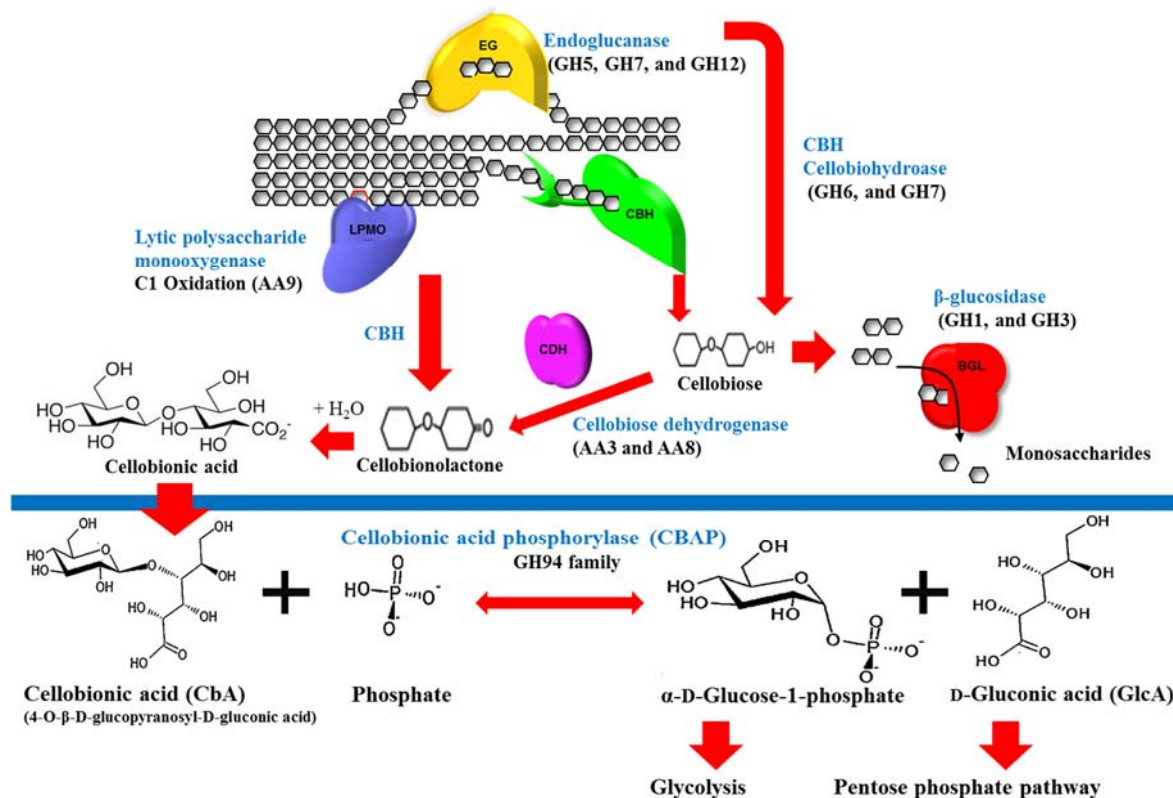
Although studies on cellulose degradating enzymes have been successfully achieved so far, it still remains question that how microbes metabolize cellobionic lactones produced by the combination of extracellular LPMOs and CBHs or CDH in the oxidative cellulose degradation pathway. Furthermore, how cellobionic lactones are connected to glycolysis has been unknown. In the old days, some results suggested that BGL hydrolyzes cellobionic acid (25, 26). However, cellobionic acid is an unsuitable and ineffective for BGL. It was also reported that the potential hydrolysate D-gluconic acid acted as a non-competitive inhibitor of BGL (27, 28).

After the function of LPMOs were clearly identified, a novel enzymecellobionic acid phosphorylase (CBAP) which can utilize cellobionic acid as the substrate was discovered by our collaborators (29). The intracellular CBAPs were characterized from the plant pathogenic bacterium *Xanthomonas compestris* and the cellulolytic fungus *Neurospora crassa*. The CBAPs were classified as new members of glycoside hydrolase family 94 (GH94 family). GH94 phosphorylases share $(\alpha/\alpha)_6$ structural fold and catalyze reversible phosphorolysis of di- or oligosaccharides to produce sugar 1-phosphate and a sugar with particular substrate specificities in the intracellular catabolisms (Table 1-3). CBAPs (EC 2.4.1.321) catalyze reversible phosphorolysis of cellobionic acid (CbA) forming α-D-glucose 1-phosphate (αGlc1P) and α-D-gluconic acid (GlcA).

In summary, cellulose is hydrolyzed into cellobiose by the combination of EG and CBH (Fig. 1-3). The resultant cellobiose is catalyzed into monosaccharides by BGL or also catalyzed into cellobiono-1,5-lactone by CDH. Moreover, the LPMO oxidizes the C1 of crystalline cellulose and then the cellobiono-1,5-lactone is released by CBH from oxidized cello-oligosaccharide. The cellobiono-1,5-lactone changes spontaneously to form cellobionic acid in water. The cellobionic acid is transported into the cytoplasm and phosphorolyzed into αGlc1P and GlcA by CBAP. The released αGlc1P is converted into D-glucose 6-phosphate by α-phosphoglucomutase (EC 2.7.5.1) for entering the glycolysis. Another released GlcA is converted into 6-phosphogluconate by the combination of gluconokinase (EC 2.7.1.12) and 6- phosphogluconate dehydrogenase (EC 1.1.1.44) for entering the pentose phosphate pathway. Therefore, CBAP is the new key enzyme that connects a "missing link" between the oxidative cellulose degradation and glycolysis and pentose phosphate pathway. The discovery of this pathway is contributed to understand the whole picture of the oxidative cellulose degradation pathway together with the discovery of the function of LPMOs.

Table 1-3 GH94 phosphorylase in CAZy database

EC	Name	Cleaved bond	Product
2.4.1.20	Cellobiose phosphorylase	D-Glc-β1,4-D-Glc	α-D-Glc-1P
2.4.1.31	Laminaribiose phosphorylase	D-Glc- β1,3-D-Glc	α-D-Glc-1P
2.4.1.49	Cellodextrin phosphorylase	D-Glc-(β1,4-D-Glc)$_n$	α-D-Glc-1P
2.4.1.280	*N,N'*–Diacetylchitobiose (chitobiose) phosphorylase	D-GlcNAc- β1,4-D-GlcNAc	α-D-GlcNac-1P

Fig 1-3 Proposed overall cellulose degradation pathway. Enzymes are shown by blue characters.

1-2-2 Intracellular GNB/LNB pathway in Bifidobacterium longum

Bifidobacteria are strictly anaerobic Gram-positive bacteria and usually found in the gastrointestinal tract (GIT) of humans and animals. It is well known that they significantly play a role in human health with the actions of various transporters, glycosidases, and metabolic enzymes (30). Bifidobacteria are the predominant intestinal bacteria in breastfed infants and grow in the human lower gastrointestinal tract where common mono- and disaccharides are insufficient (31). Additionally, bovine milk contains lactose, digested by lactase in the small intestine for infant, as the single saccharide, whereas human milk contains large amount of various oligosaccharides. HMOs (human milk oligosaccharides) are not able to be digested by enzymes in large intestine intact, where the HMO are potentially utilized by bifidobacteria. Therefore, Bifidobacteria is considered to have developed alternative pathways that utilize various oligosaccharides such as, the two type of human milk oligosaccharides, lactose, and mucin glycans (32-34) (Table 1-4). More than 130 types of HMOs have been isolated and they can be classified based on core structures type I and type II. Type I HMO, such as lacto-*N*-tetraose (Galβ1,3GlcNAcβ1,3Galβ1,4Glc), lacto-*N*-fucopentaose I (Fucα1,2Galβ1,3GlcNAcβ1,3Galβ1,4Glc), and lacto-*N*-difucohexaose I (Fucα1,2Galβ1,3[Fucα1,4]GlcNAcβ1,3Galβ1,4Glc) contain LNB (Gal-*β* 1,3-GlcNAc), whereas type II HMO contain LacNAc (Gal-*β* 1,4-GlcNAc) (35, 36). Type I oligosaccharides are predominant in HMO but type II oligosaccharides are minor constituents. Interestingly, type I milk oligosaccharides are only found in human milk but not in other mammals' milk (37).

Table 1-4 Type of disaccharides metabolized by Bifidobacteria

Name	Structure	Description
Galacto-*N*-biose (GNB)	Gal-*β* 1,3-GalNAc	Mucin-type *O*-glycans
Lacto-*N*-biose (LNB)	Gal-*β* 1,3-GlcNAc	Type I HMO
N-Acetyllactosamine (LacNAc)	Gal-*β* 1,4-GlcNAc	Type II HMO
Lactose	Gal-*β* 1,4-Glc	Major carbohydrate in mammal milk

In 1999, Derensy-Dron et al. reported that lacto-*N*-biose phosphorylase (LNBP) from *B. bifidum* reversibly phosphorolyzes LNB to α-D-galactopyranose 1-phosphate (Gal1P) and *N*-acetyl-D-glucosamine (GlcNAc) (38). The enzyme also phosphorolyzes galacto-*N*-biose (GNB, Gal-*β* 1,3-GalNAc) to Gal1P and *N*-acetyl-D-galactosamine (GalNAc). In 2005, Kitaoka et al., found a novel putative operon from *bifidobacterium longum* JCM1217 containing the *LNBP* gene (*lnpA*). The *lnpA*

encodes LNBP enzyme which is a part of a putative operon. This operon was also identified in several species of bificobacteria, indicating that bifidobacteria possess a unique intracellular sugar metabolic pathway specific for LNB and GNB (39). Especially, it is well-known that LNB is a selective growth factor and a building block of the type 1 HMOs that is rich in human milk. Accordingly, a total of 208 strains were investigated for the presence of the *GLNBP* gene and tested for growth in the presence of LNB as the sole carbohydrate source (40). Interestingly, all bifidobacterial strains do not grow and do not have the *GLNBP* gene. Only *Bifidobacterium longum* subsp. l*ongum, B. longum* subsp. *infantis, B. breve,* and *B. bifidum* strains were shown to have potentially the GLNBP gene (Table 1-5). The GLNBP gene (BL1641) is located in a putative operon in the *B. longum* NCC2905. In the operon, BL1642, BL1643, and BL1644 are annotated as *N*-acetylhexosamine 1-kinase (NahK, EC 2.7.1.162), galactose-1-phosphate uridylyltransferase (GalT, EC 2.7.7.10), and UDP-galactose 4-epimerase (GalE, EC 5.1.3.2), respectively (Fig. 1-4A) (32).

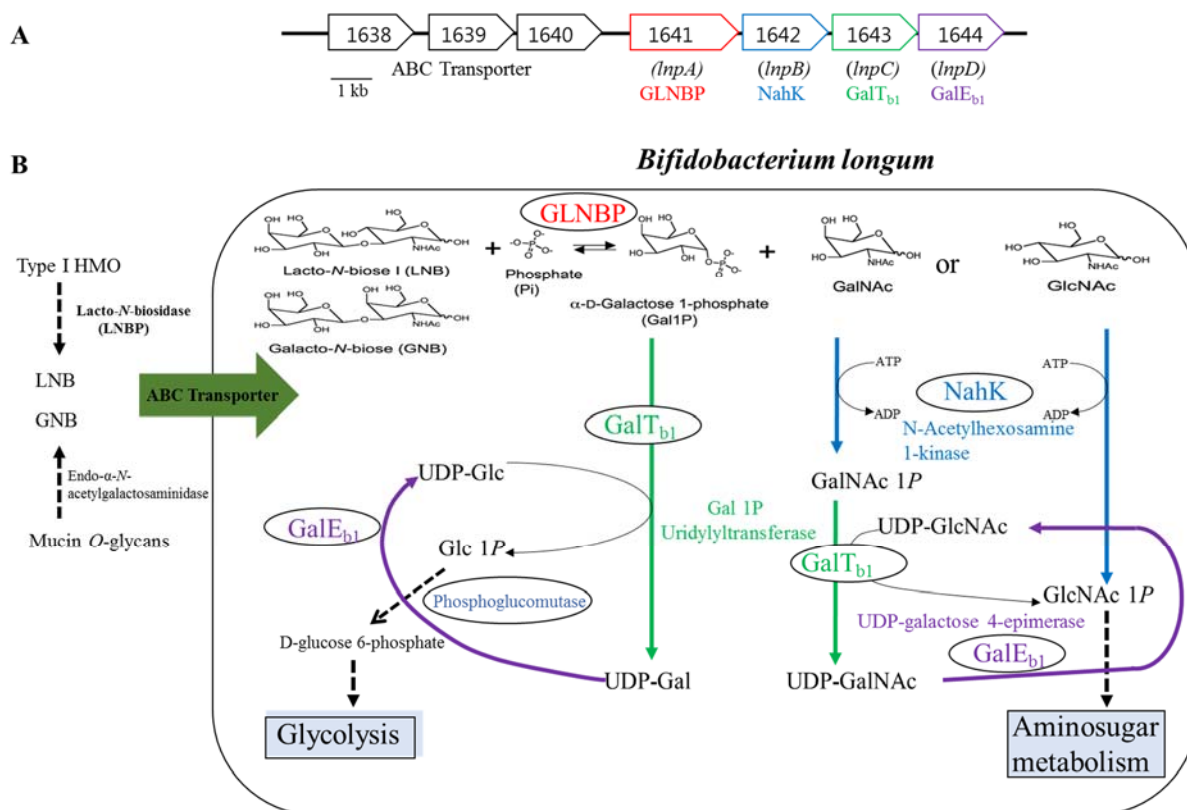**Table 1-5 The growth patterns in bifidobacterium (40).** (Xiao, J. *et al., Appl Environ Microbiol 76*, 54-59, 2010)

Species	Strains	Origin	Growth pattern[a]		
			LNB	Lactulose	Glucose
B. longum subsp. *longum*	IV-51 (same as ATTCC 15707[T])	Feces of adult	5	4	5
	ATCC 15708	Feces of infant	2	2	4
	MCC strains	Feces of adult or infant	2-5	5	5
B. longum subsp. *infantis*	IV-8 (same as ATCC 15697[T])	Infant	5	5	5
	MCC strains	Adult, infant	2-5	1-5	2-5
B. breve	IV-14 (same as ATCC 15700[T])	Infant	5	5	5
	MCC strains	Infant or from yogurt	2-5	3-5	1-5
B. bifidum	ATCC 15696	Intestine of infant	5	5	4
	MCC strains	Feces of infant	5	5	3-5
B. adolescentis	ATCC 15706	Feces of adult	0	1	2
	ATCC 15705	Feces of adult	0	2	2
	MCC strains	Feces of adult or Infant	0	0-3	0-4
B. catenulatum	ATCC 27675	Feces of human	0	2	2
	MCC strains	Feces of adult	0	1-5	2-5
B. longum subsp. *animalis*	IV-58 (same as ATCC 25527[T])	Feces of rat	0	4	4
B. longum subsp. *lactis*	DSM 10140	Yogurt	0	4	5
	MCC strains	Yogurt	0	4-5	5
B. Thermophilum	MCC strinas	Feces of animals	0	0-5	4-5

[a]The growth pattern was categorized from 0 to 5 based on OD_{600} measured at 24h ($OD_{24 h}$) and 144h ($OD_{144 h}$).0, no growth ($OD_{144 h} < 0.1$); 1, slow and mild growth ($OD_{24 h} < 0.1$; $OD_{144 h} = 0.1$ to 0.5); 2, slow but strong growth ($OD_{24 h} < 0.1$; $OD_{144 h} > 0.5$); 3, intermediate and mild growth ($OD_{24 h} = 0.1$ to 0.5; $OD_{144 h} = 0.1$ to 0.5); 4, intermediate and strong growth ($OD_{24 h} = 0.1$ to 0.5; $OD_{144 h} > 0.5$); 5, quick and strong growth($OD_{24 h} > 0.5$)

The downstream genes (*LnpB-LnpD*) were cloned and characterized for the properties of enzymes in *B. longum* JCM1217 (32). NahK encoded by *LnpB* can phosphorylate the anomeric1-position of both GlcNAc and GalNAc. GalT encoded by *LnpC* and GalE encoded by *LnpD* are involved not only in the general glucose/galactose metabolism, but also in GlcNAc/GalNAc metabolism. In the view of enzyme specificity, GalT and GalE can utilize both hexose 1-phosphate and *N*-acetylhexosamine 1-phosphate, and both UDP-Gal and UDP-GalNAC as substrates, respectively. The metabolic pathway encoded by this operon was called as GNB/LNB pathway. These enzymes involved in GNB/LNB pathway can send GNB and LNB molecules to glycolysis and aminosugar metabolism (Fig. 1-4B)

The GNB/LNB pathway can be considered to an energy-saving pathway compared to Leloir pathway. In the Leloir pathway, galactose is phosphorylated by galactokinase, comsuming ATP. However, GLNBP produces galactose 1-phosphate without consuming ATP. The GNB/LNB pathway is the main novel sugar metabolic pathway for HMO and mucin *O*-glycans as an energy source in *B. longum*.

Fig 1-4 Scheme of the GNB/LNB pathway in *Bifidobacterium longum* (A) The schematic structure of the gene cluster encoding the pathway in *b. longum* JCM1217, (B) Schematic representation of the GNB/LNB pathway.

1-3 The aim of this thesis

Glycolysis, the major pathways of carbohydrate metabolism, have evolved to process the hexose monosaccharide glucose. Many of the enzymes in the glycolytic pathway are so specific for monosaccharide glucose that even other sugars are not processed at any appreciable rate. To get over this problem, microorganisms would use number of unknown short pathways for converting other common sugars. When some novel enzymes would be discovered and characterized though biochemical analysis, the short pathways could be revealed and provided to information for understanding biological mechanisms.

In this general introduction, I described two novel sugar metabolic pathway finally connected to the glycolytic pathway. However, some enzymes related to each pathway have been still required for understanding biochemical process and the detailed molecular mechanism.

1) In the oxidative cellulose degradation pathway, hydrolytic enzymes (cellulases) have been well studied in aspect of functional and structural features. Even though cellobionolactone yielded by the combination of LPMOs and CBHs was recently revealed to be catalyzed by intracellular cellobionic acid phosphorylase, there was no information about substrate recognition residues and catalytic mechanism based on crystal structures. In chapter 2, therefore, X-ray crystallography was conducted on CBAP from the marine bacterium *Saccharophagus degradans* 2-40 for identifying the catalytic residues and understanding the phosphorolysis reaction mechanism. This study would provide basic information for the further investigation of cellulose degradation pathway and the future design of phosphorylases with altered substrate specificity.

2) The GNB/LNB pathway is a unique pathway for the metabolism of GNB and LNB as the sole carbohydrate source, only found in some bifidobacteria as described in Fig 1-4. The four intracellular enzymes sequentially convert GNB and LNB and send the products to downstream pathways. The crystal structures of intracellular enzymes belonged to GNB/LNB pathway have been only reported in GLNB phosphorylase (41) and NahK (42). The crystal structures of GalT and GalE related to GNB/LNB pathway are not reported yet. Both of these enzymes have broad substrate specificity compared to general GalT and GalE enzymes belonged to Leloir pathway because they can catalyze UDP-GlcNAc/GalNAc in in addition to UDP-Glc/Gal. It needs to be discussed about the reasons why the both enzymes have different substrate specificity from homologous enzymes in Leloir pathway.

Research of these two sugar metabolic enzymes provide beneficial tools in the investigations of relationship between structure and function of oligosaccharide biocatalytic enzymes and for wide range of applications in biomass utilization, nutrition, and medicine development.

1-4 References

1. **Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J.** (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition, *Biochem J 382*, 769-781.

2. **Shoseyov, O., Shani, Z., and Levy, I.** (2006) Carbohydrate binding modules: biochemical properties and novel applications, *Microbiol Mol Biol Rev 70*, 283-295.

3. **Karlsson, J., Medve, J., and Tjerneld, F.** (1999) Hydrolysis of steam-pretreated lignocellulose: synergism and adsorption for cellobiohydrolase I and endoglucanase II of Trichoderma reesei, *Appl Biochem Biotechnol 82*, 243-258.

4. **Beguin, P.** (1990) Molecular biology of cellulose degradation, *Annu Rev Microbiol 44*, 219-248.

5. **Lin, J., Pillay, B., and Singh, S.** (1999) Purification and biochemical characteristics of beta-D-glucosidase from a thermophilic fungus, Thermomyces lanuginosus-SSBP, *Biotechnol Appl Biochem 30 (Pt 1)*, 81-87.

6. **Westermark, U., and Eriksson, K. E.** (1975) Purification and properties of cellobiose: quinone oxidoreductase from Sporotrichum pulverulentum, *Acta Chem Scand B 29*, 419-424.

7. **Beeson, W. T. t., Iavarone, A. T., Hausmann, C. D., Cate, J. H., and Marletta, M. A.** (2011) Extracellular aldonolactonase from Myceliophthora thermophila, *Appl Environ Microbiol 77*, 650-656.

8. **Li, X., Beeson, W. T. t., Phillips, C. M., Marletta, M. A., and Cate, J. H.** (2012) Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases, *Structure 20*, 1051-1061.

9. **Berka, R. M., Grigoriev, I. V., Otillar, R., Salamov, A., Grimwood, J., Reid, I., Ishmael, N., John, T., Darmond, C., Moisan, M. C., Henrissat, B., Coutinho, P. M., Lombard, V., Natvig, D. O., Lindquist, E., Schmutz, J., Lucas, S., Harris, P., Powlowski, J., Bellemare, A., Taylor, D., Butler, G., de Vries, R. P., Allijn, I. E., van den Brink, J., Ushinsky, S., Storms, R., Powell, A. J., Paulsen, I. T., Elbourne, L. D., Baker, S. E., Magnuson, J., Laboissiere, S., Clutterbuck, A. J., Martinez, D., Wogulis, M., de Leon, A. L., Rey, M. W., and Tsang, A.** (2011) Comparative genomic analysis of the thermophilic biomass-degrading fungi Myceliophthora thermophila and Thielavia terrestris, *Nat Biotechnol 29*, 922-927.

10. **MacDonald, J., Doering, M., Canam, T., Gong, Y., Guttman, D. S., Campbell, M. M., and Master, E. R.** (2011) Transcriptomic responses of the softwood-degrading white-rot fungus Phanerochaete carnosa during growth on coniferous and deciduous wood, *Appl Environ Microbiol 77*, 3211-3218.

11. **Tian, C., Beeson, W. T., Iavarone, A. T., Sun, J., Marletta, M. A., Cate, J. H., and Glass, N. L.** (2009) Systems analysis of plant cell wall degradation by the model filamentous fungus

Neurospora crassa, *Proc Natl Acad Sci U S A 106*, 22157-22162.

12. **Harris, P. V., Welner, D., McFarland, K. C., Re, E., Navarro Poulsen, J. C., Brown, K., Salbo, R., Ding, H., Vlasenko, E., Merino, S., Xu, F., Cherry, J., Larsen, S., and Lo Leggio, L.** (2010) Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family, *Biochemistry 49*, 3305-3316.

13. **Karkehabadi, S., Hansson, H., Kim, S., Piens, K., Mitchinson, C., and Sandgren, M.** (2008) The first structure of a glycoside hydrolase family 61 member, Cel61B from Hypocrea jecorina, at 1.6 A resolution, *J Mol Biol 383*, 144-154.

14. **Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., and Henrissat, B.** (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes, *Biotechnol Biofuels 6*, 41.

15. **Beeson, W. T., Phillips, C. M., Cate, J. H., and Marletta, M. A.** (2012) Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases, *J Am Chem Soc 134*, 890-892.

16. **Cabib, E.** (1970) Research on sugar nucleotides brings honor to Argentinian biochemist (Luis Leloir), *Science 170*, 608-609.

17. **Kingsley, D. M., Kozarsky, K. F., Hobbie, L., and Krieger, M.** (1986) Reversible defects in O-linked glycosylation and LDL receptor expression in a UDP-Gal/UDP-GalNAc 4-epimerase deficient mutant, *Cell 44*, 749-759.

18. **Wohlers, T. M., Christacos, N. C., Harreman, M. T., and Fridovich-Keil, J. L.** (1999) Identification and characterization of a mutation, in the human UDP-galactose-4-epimerase gene, associated with generalized epimerase-deficiency galactosemia, *Am J Hum Genet 64*, 462-470.

19. **Berry, G. T., Moate, P. J., Reynolds, R. A., Yager, C. T., Ning, C., Boston, R. C., and Segal, S.** (2004) The rate of de novo galactose synthesis in patients with galactose-1-phosphate uridyltransferase deficiency, *Mol Genet Metab 81*, 22-30.

20. **Thoden, J. B., and Holden, H. M.** (2002) High resolution X-ray structure of galactose mutarotase from Lactococcus lactis, *J Biol Chem 277*, 20854-20861.

21. **Thoden, J. B., Kim, J., Raushel, F. M., and Holden, H. M.** (2002) Structural and kinetic studies of sugar binding to galactose mutarotase from Lactococcus lactis, *J Biol Chem 277*, 45458-45465.

22. **Thoden, J. B., and Holden, H. M.** (2003) Molecular structure of galactokinase, *J Biol Chem 278*, 33305-33311.

23. **Wedekind, J. E., Frey, P. A., and Rayment, I.** (1995) Three-dimensional structure of galactose-1-phosphate uridylyltransferase from Escherichia coli at 1.8 A resolution,

Biochemistry 34, 11049-11061.

24. **Thoden, J. B., and Holden, H. M.** (1998) Dramatic differences in the binding of UDP-galactose and UDP-glucose to UDP-galactose 4-epimerase from Escherichia coli, *Biochemistry 37*, 11469-11477.

25. **Moeller, P.** (1973) [On the enzymatic hydrolysis of cellobionic acid. I. Synergistic effect of beta-glucosidase and lactonase fractions from Aspergillus cellulase (author's transl)], *Hoppe Seylers Z Physiol Chem 354*, 1271-1276.

26. **Westerma.U, and Eriksson, K. E.** (1974) Carbohydrate-Dependent Enzymic Quinone Reduction during Lignin Degradation, *Acta Chemica Scandinavica Series B-Organic Chemistry and Biochemistry B 28*, 204-208.

27. **Cannella, D., Hsieh, C. W., Felby, C., and Jorgensen, H.** (2012) Production and effect of aldonic acids during enzymatic hydrolysis of lignocellulose at high dry matter content, *Biotechnol Biofuels 5*, 26.

28. **Dale, M. P., Ensley, H. E., Kern, K., Sastry, K. A., and Byers, L. D.** (1985) Reversible inhibitors of beta-glucosidase, *Biochemistry 24*, 3530-3539.

29. **Nihira, T., Saito, Y., Nishimoto, M., Kitaoka, M., Igarashi, K., Ohtsubo, K., and Nakai, H.** (2013) Discovery of cellobionic acid phosphorylase in cellulolytic bacteria and fungi, *FEBS Lett 587*, 3556-3561.

30. **Guarner, F., and Malagelada, J. R.** (2003) Gut flora in health and disease, *Lancet 361*, 512-519.

31. **Turroni, F., Foroni, E., Pizzetti, P., Giubellini, V., Ribbera, A., Merusi, P., Cagnasso, P., Bizzarri, B., de'Angelis, G. L., Shanahan, F., van Sinderen, D., and Ventura, M.** (2009) Exploring the Diversity of the Bifidobacterial Population in the Human Intestinal Tract, *Appl Environ Microb 75*, 1534-1545.

32. **Nishimoto, M., and Kitaoka, M.** (2007) Identification of N-acetylhexosamine 1-kinase in the complete lacto-N-biose I/galacto-N-biose metabolic pathway in Bifidobacterium longum, *Appl Environ Microbiol 73*, 6444-6449.

33. **Fushinobu, S.** (2010) Unique sugar metabolic pathways of bifidobacteria, *Biosci Biotechnol Biochem 74*, 2374-2384.

34. **Kitaoka, M.** (2012) Bifidobacterial enzymes involved in the metabolism of human milk oligosaccharides, *Adv Nutr 3*, 422S-429S.

35. **Asakuma, S., Urashima, T., Akahori, M., Obayashi, H., Nakamura, T., Kimura, K., Watanabe, Y., Arai, I., and Sanai, Y.** (2008) Variation of major neutral oligosaccharides levels in human colostrum, *Eur J Clin Nutr 62*, 488-494.

36. **Erney, R. M., Malone, W. T., Skelding, M. B., Marcon, A. A., Kleman-Leyer, K. M., O'Ryan, M. L., Ruiz-Palacios, G., Hilty, M. D., Pickering, L. K., and Prieto, P. A.** (2000) Variability of human milk neutral oligosaccharides in a diverse population, *J Pediatr Gastroenterol Nutr 30*, 181-192.

37. **Urashima, T., Odaka, G., Asakuma, S., Uemura, Y., Goto, K., Senda, A., Saito, T., Fukuda, K., Messer, M., and Oftedal, O. T.** (2009) Chemical characterization of oligosaccharides in chimpanzee, bonobo, gorilla, orangutan, and siamang milk or colostrum, *Glycobiology 19*, 499-508.

38. **Derensy-Dron, D., Krzewinski, F., Brassart, C., and Bouquelet, S.** (1999) Beta-1,3-galactosyl-N-acetylhexosamine phosphorylase from Bifidobacterium bifidum DSM 20082: characterization, partial purification and relation to mucin degradation, *Biotechnol Appl Biochem 29 (Pt 1)*, 3-10.

39. **Kitaoka, M., Tian, J., and Nishimoto, M.** (2005) Novel putative galactose operon involving lacto-N-biose phosphorylase in Bifidobacterium longum, *Appl Environ Microbiol 71*, 3158-3162.

40. **Xiao, J. Z., Takahashi, S., Nishimoto, M., Odamaki, T., Yaeshima, T., Iwatsuki, K., and Kitaoka, M.** (2010) Distribution of in vitro fermentation ability of lacto-N-biose I, a major building block of human milk oligosaccharides, in bifidobacterial strains, *Appl Environ Microbiol 76*, 54-59.

41. **Hidaka, M., Nishimoto, M., Kitaoka, M., Wakagi, T., Shoun, H., and Fushinobu, S.** (2009) The crystal structure of galacto-N-biose/lacto-N-biose I phosphorylase: a large deformation of a TIM barrel scaffold, *J Biol Chem 284*, 7273-7283.

42. **Wang, K. C., Lyu, S. Y., Liu, Y. C., Chang, C. Y., Wu, C. J., and Li, T. L.** (2014) Insights into the binding specificity and catalytic mechanism of N-acetylhexosamine 1-phosphate kinases through multiple reaction complexes, *Acta Crystallogr D 70*, 1401-1410.

Chapter 2 については　単行本として出版する計画があるため公表できない。

Chapter 2 X-ray crystallography of cellobionic acid phosphorylase from *S.*
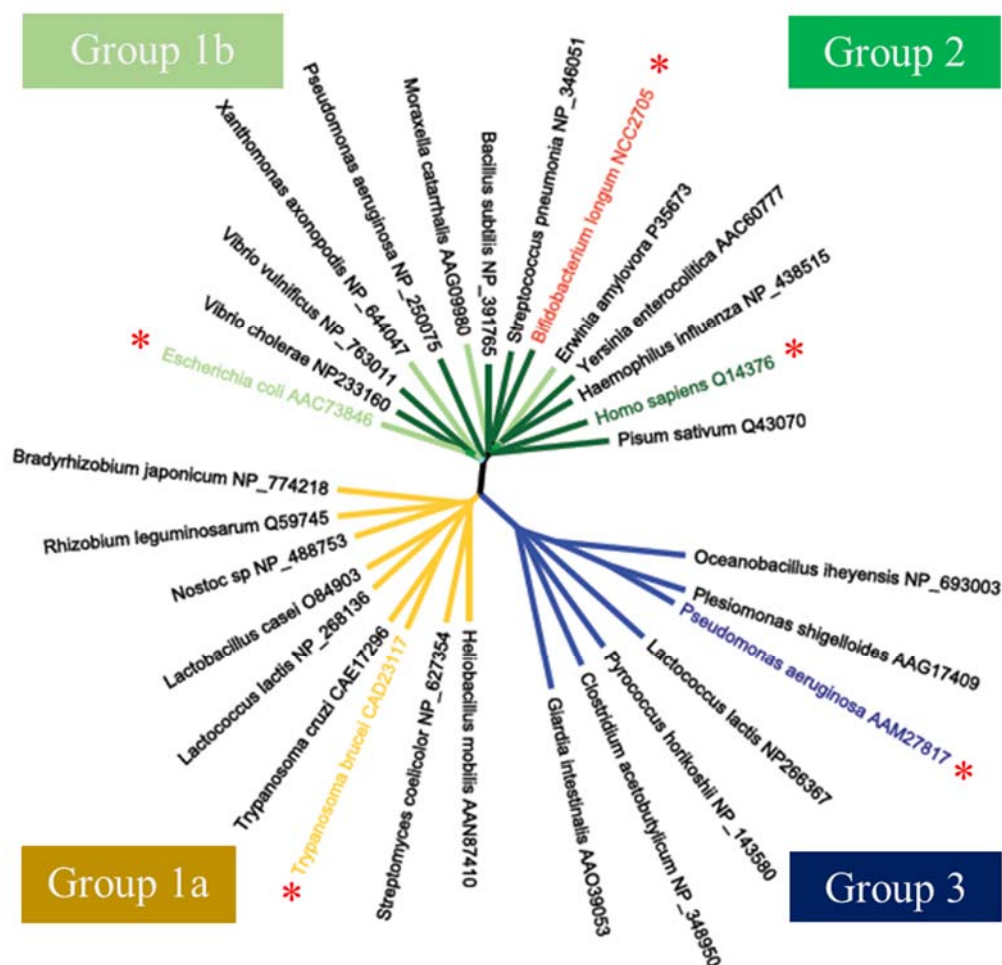***degradans* 2-40**

Chapter 3 X-ray crystallography of UDP- glucose hexose 1-uridylytransferase and UDP-galactose 4-epimerase in GNB/LNB pathway

3-1 Introduction

The intracellular GNB/LNB pathway is a novel sugar metabolic pathway in *B. longum* for human milk oligosaccharides and intestinal mucin glycoproteins. UDP-glucose hexose 1-phosphate uridylyltransferase (GalT) and UDP-galactose 4-epimerase have been conserved (GalE) from *Escherichia coli* to humans. The enzymes involved in the GNB/LNB pathway sequentially convert GNB/LNB into αG1P and GlcNAc 1-phosphate. These enzymes were characterized and distinguished in the aspect to substrate specificity with homologue enzymes involved in Leloir pathway (1). Interestingly, UDP-glucose hexose 1-phosphate uridylyltransferase (bGalT) and UDP-galactose 4-epimerase (bGalE) from *B. logum* are involved not only in typical glucose/galactose metabolism, but also in the GlcNAc/GalNAc metabolism. It was reported that impairment of these enzymes in the Leloir pathway in yeast causes the disease galactosemia and metabolic disorder (2). They also regulates the intracellular homeostasis of the UDP-glucose/UDP-GlcNAc and UDP-galactose/UDP-GalNAc for aminosugar metabolism. As described in the general introduction, the crystal structures of GNB/LNB phosphorylase and *N*-acetylhexosamine 1-kinase (NahK) have been reported (3, 4). On the other hand, the structural studies of bGalT and bGalE have not been accomplished yet. Therefore, I conducted crystallization screening of bGalE and bGalT. A preliminary structure of bGalT was solved at 2.4 Å resolution by a molecular replacement method, but crystallographic refinement is not sufficiently done yet. In this chapter, I mainly describe about X-ray crystallography of bGalE focusing on its wide substrate specificity.

GalE from *E. coli* (eGalE), which has been extensively studied, catalyzes only the conversion between UDP-Gal and UDP-Glc (5-7). On the other hand, human GalE (hGalE) shows a broad substrate specificity similar to bGalE, catalyzing the conversion of both UDP-Gal/UDP-Glc and UDP-GalNAc/UDP-GlcNAc (8, 9). Moreover WbpP from *Pseudomonas aeruginosa* preferentially catalyzes the conversion between UDP-GalNAc and UDP-GlcNAc (10). The GalE enzymes can be classified into three main groups based on their substrate specificity but they do not make clear clusters in the phylogenetic tree (Fig. 3-1) (10, 11). Group 1 represents epimerases which catalyze the conversion between UDP-glucose and UDP-galactose, such as the eGalE, and *Trypanosoma brucei* GalE (tGalE). The difference between group 1a and group 1b is only the size of the sugar binding pocket (discussed after). tGalE has too small saccharide binding pocket to bind acetylated UDP-hexose (group 1a). Group 2 represents epimerases which can catalyze both UDP-Glc/UDP-GlcNAc into UDP-Gal/UDP-GalNAc, such as hGalE and bGalE. Group 3 represents epimerase which can strongly catalyze acetylated substrate, such as WbpP.

Fig. 3-1 Phylogenetic tree based on substrate specificity for UDP-hexose 4-epimeraes.
A phylogenetic tree was prepared using 30 sequences, which were aligned using ClustalW2 (12) and the tree was drawn using the FigTree program. Asterisks indicate structure-known members.

  Even though many numbers of epimerases classified in group 2 have been characterized for biochemical properties (13-15), the structural approaches are accomplished in only hGalE. In this chapter, the structural characterization of bGalE in complex with UDP-GlcNAc and UDP-Glc could support the biochemical information to understand the structural basis of substrate specificity, comparing with the structural properties of group 1, 2, or 3 epimerases (8-10, 16).

3-2 Materials and Methods

3-2-1 Protein expression check

   LnpC (bGalT) and *lnpD* (bGalE) plasmids were transformed into *E. coli* BL21-CodonPlus (DE3) and incubated in 3 ml of Luria-Bertani (LB) medium containing 50 ug/ml kanamycin and chloramphenicol, until the absorbance reached OD 0.6 at 600 nm. The proteins were expressed by range from 0 to 1 mM of isopropyl β-D-thiogalactopyranoside (IPTG) and incubated on different temperatures: 37 ºC for 3 hours, 30 ºC for 8 hours, and 25 ºC for 20 hours. The expressed proteins were confirmed by SDS-PAGE.
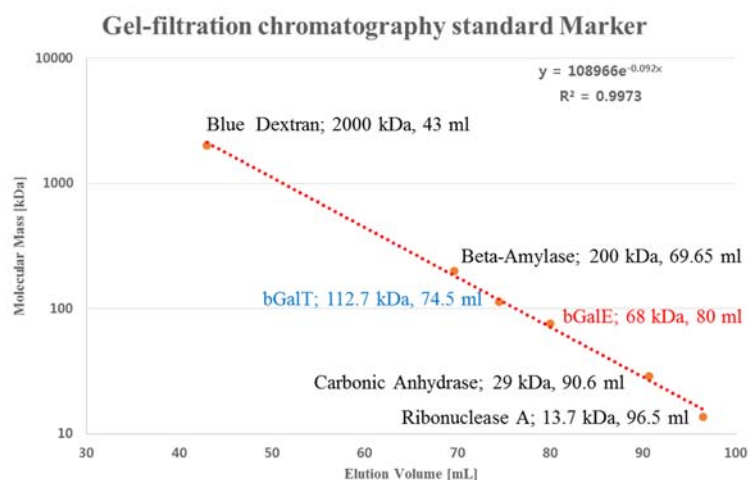

Fig 3-2 Protein expression check for bGalT and bGalE

3-2-2 Protein Preparation

   The nonlabeled bGalT was expressed at 25 ºC for 20 hours after adding 0.1 mM IPTG in 1 L of LB medium. The cells were collected and suspended in 50 mM Tris-HCl buffer, pH 8.0 and then disrupted by sonication (Cosmo Bio). The cell debris was removed by ultracentrifugation at 10,000 × g for 30 min. The cell-free extract was subjected to Ni-affinity chromatography column (HisTrap™ FF crude 5ml, GE Healthcare). The enzyme was eluted with a two steps of 20 mM and 250 mM imidazole in 50 mM Tris-HCL buffer, pH 8.0. The enzyme was then further purified on a column of Mono Q colum (GE Healthcare) with an increasing linear gradient of 0-100% saturation of sodium chloride by using an ÄKTA (GE Healthcare). The molecular mass of purified bGalT was estimated by SDS-PAGE and by gel filtration chromatography (HiLoad 16/60 Superdex 200 pg; GE Healthcare) equilibrated with 50 mM Tris-HCl buffer, pH 8.0 containing 150 mM NaCl at a flow rate of 1.0 ml/min. The

selenomethionine-labeled protein was expressed in methionine auxotroph *E. coli* strain B834 (DE3) (Novagen) in medium containing selenomethionine. The procedures of purification were performed as the same as those for non-labeled bGalT protein.

The bGalE was expressed at 25 ºC for 20 hours after adding 0.1 mM IPTG in 1 L of LB medium. The cells were collected and supended in 50 mM Tris-HCl buffer, pH 7.0 and 0.1 mM PMSF. The supernatant was sequentially purified on Ni-affinity chromatography with a two steps of 20 mM and 250 mM imidazole in 50 mM Tris-HCL buffer, pH 7.0 and gel filtration chromatography equilibrated with 50 mM Tris-HCL buffer, pH 7.0 containing 150 mM NaCl at a flow rate of 1.0 ml/min.

Gel-filtration chromatography standard Marker

$y = 108966e^{-0.092x}$
$R^2 = 0.9973$

Blue Dextran; 2000 kDa, 43 ml
Beta-Amylase; 200 kDa, 69.65 ml
bGalT; 112.7 kDa, 74.5 ml
bGalE; 68 kDa, 80 ml
Carbonic Anhydrase; 29 kDa, 90.6 ml
Ribonuclease A; 13.7 kDa, 96.5 ml

Fig. 3-3. Gel-filtration chromatography. Both of bGalT and bGalE proteins exist as dimer in solution.

3-2-3 Crystallization

The proper concentrations of bGalT and bGalE for crystallization screenings were determined by Pre-Crystallization Test (Hampton Research). The initial screenings for crystallization of bGalT and bGalE was performed using Nextal JCSG⁺ suit, JCSG Core I-IV suit, Crystal screen I&II, and Wizard I&II. Crystals were prepared by sitting drop vapor diffusion method using mixing 1:1 volume of enzyme and reservoir solutions at 4ºC.

The bGalT were crystallized with 0.1 M ammonium acetate, 0.1M Bis-Tris buffer, pH 5.5, 17% (w/v) PEG 10,000, 10 mg/ml protein and 10 mM UDP. After crystal condition was optimized to 0.14-0.18 M Bis-Tris buffer, pH 5.5, 17% (w/v) PEG 1,000, 10 mg/ml protein and 10 mM UDP. The optimized crystal condition was tested as the same method for selenomethionine-labeled bGalT. 20% (v/v) glycerol was used as cryoprotectant before data collection.

For successful crystallization of bGalE, the reservoir solution consisted of 0.1 M HEPES buffer, pH 7.5, 0.2M magnesium chloride, and 30% (v/v) PEG 400, whereas the drop consisted of 7 mg/ml bGalE, either 10 mM UDP-GlcNAc or 10 mM UDP-galactose. The crystals were not obtained without co-crystallization with UDP-substrates.

3-2-4 Data collection

The diffraction data set of native bGalT were collected at a wavelength of 0.97319 Å on beam line BL17A of the Photon Factory, KEK. The single wavelength anomalous dispersion (SAD) data sets for the selenomethionine-labeled crystal, 20 mM $HgCl_2$ soaked crystal, and 20 mM $K_2Pt(NO_2)_4$ soaked crystal were collected at wavelength of 0.97876 Å, 1.00847 Å, and 1.0000 Å on beam line of NE3A, BL17A. All data sets were carried out and scaled using HKL2000 (17).

X-ray diffraction data sets of bGalE were collected using synchrotron radiation at beam lines BL17A, BL5A, and NW12A of the photon Factory, and data were processed and scaled using the HKL2000 program.

3-2-5 Phase calculation of bGalT

The anomalous signal of selenium atom from 8 methionine residues was evaluated. The initial phase calculation and phase improvement was performed by experimental phasing with program the PHENIX *AutoSol* and *AutoBuild* (18). The quality of the phases was insufficient (figure of merit = 0.175 for all reflections) to obtain an ambiguous map at 3.25 Å resolution

3-2-6 Molecular replacement and Refinement of bGalE

The native structure was solved by molecular replacement (Molrep software) using human GalE (hGalE; PDB code 1EK6) as a template (19). Automated model building and refinement was performed using the MR and Refmac5 suite in CCP4, respectively. Manual model rebuilding and refinement were performed using Coot (20). All figures were prepared using PyMol [http://www.pymol.org] and ESPript (21).

3-3 Results

3-3-1 Crystallization and data collection of bGalT and bGalE

Crystals of bGalT were obtained in 0.1 M ammonium acetate, 0.1 M Bis-Tris pH 5.5, 17% (w/v) PEG 10,000 and 10 mg/ml of protein at 4 ℃ (Fig. 3-4A). After optimized crystallization condition with 0.14~0.18 M ammonium acetate, 0.1 M Bis-Tris pH 5.5, 17% (w/v) PEG 1,000 and 10 mg/ml of protein, the crystals of bGalT were bigger in size (Fig. 3-4B). The crystallization of selenomethionie-labeled (SeMet) bGalT was performed on the same condition of optimized for the native crystals. However, the crystals of SeMet bGalT were smaller than native crystals in size and the resolution was poor for determination of phase (Fig. 3-4C). The biggest problem was that the native and SeMet crystals were twinned. In case of native crystals, the advanced beam line and the big size of crystals helped to get an enough resolution in spite of forming twin crystals. On the other hand, the SeMet were too small to collect a data set with enough resolution for complete structure determination. Therefore, heavy atoms were soaked into big native crystals. Even though the resolution was improved, cell constants were different with the native crystals. Crystallographic refinement is currently in progress with SeMet data sets. The all data collections of bGalT are described in Table 3-1.
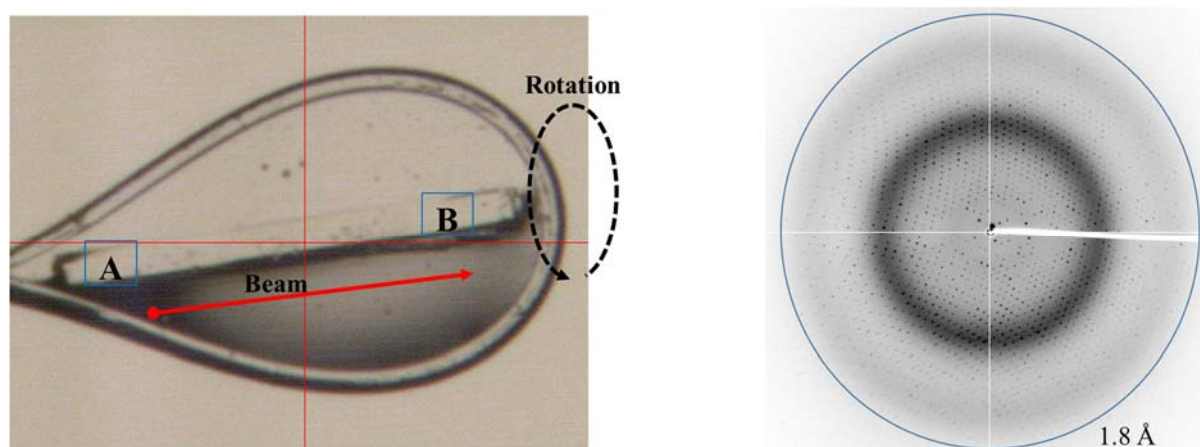

Fig. 3-4 Crystals of bGalT and bGalE. (A) First obtained bGalT crystals. (B) Optimized bGalT crystals. (C) Selenomethionie-labeled bGalT crystlas. (D) bGalE crystals.

Table 3-1 Data collection of bGalT

Data collection	Native	Se_Met (Peak)	Hg	Pt
Beam line	BL17A	NE3A	BL17A	BL17A
Wavelength (Å)	0.97319	0.97876	1.00847	1.0000
Space group	P21	P21	P21	P21
Unit cell (Å)	a=84.848, b=159.433, c=108.867 α=90.000, β=107.329 γ=90.000	a=84.482, b=93.950, c=158.886 α=90.000, β=90.183 γ=90.000	a=58.204, b=160.792, c=78.314 α=90.000, β=105.089 γ=90.000	a=58.190, b=159.475, c=78.218 α=90.000, β=104.820 γ=90.000
Resolution (Å)[a]	50.00-2.4 (2.44-2.40)	50.00-3.25 (3.31-3.25)	50.00-2.6 (2.64-2.60)	50.00-2.80 (2.85-2.80)
Total reflections	395,697	269,254	75,263	68,467
Unique reflections	106,617	39,706	42,781	38,114
Completeness (%)[a]	99.4 (99.9)	99.7 (99.7)	60.3 (62.6)	71.0 (75.6)
Redundancy[a]	3.7 (3.7)	6.8 (6.8)	2.4 (2.4)	2.8 (2.8)
Mean I/σ (I)[a]	12.7 (2.3)	12.6 (2.3)	12.6 (2.0)	8.0 (2.3)
R_{merge} (%)[a]	7.7(45.0)	9.5 (45.4)	12.1(73.9)	20.8 (73.6)

[a] Values in parentheses correspond to the highest resolution shell
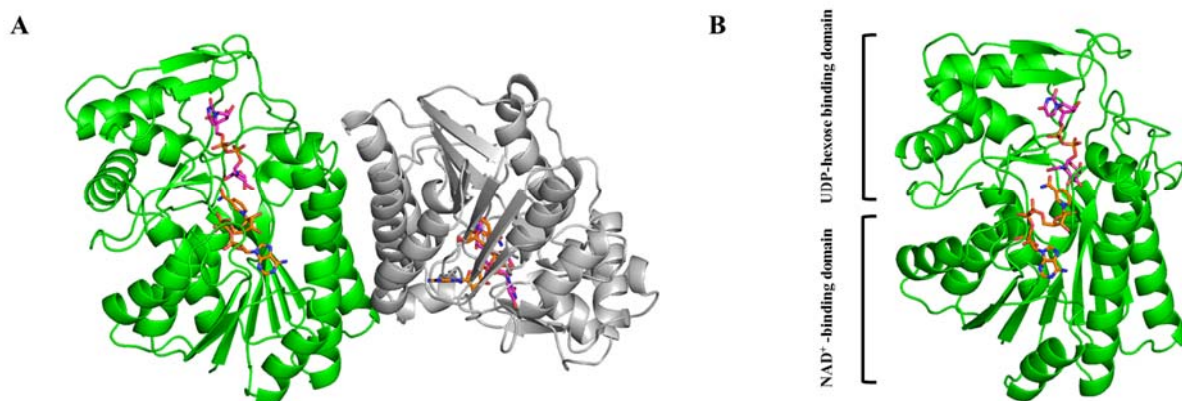
   The crystals of bGalE were pillar-shaped (Fig. 3-4D). The first data collection of bGalE was obtain to 2.4 Å resolution. The crystal got radiation damage easily because of the shape of bGalE. The beamline in Photon Factory supported the helical mode, which could help to collect data with low radiation damage on crystals (Fig. 3-5).

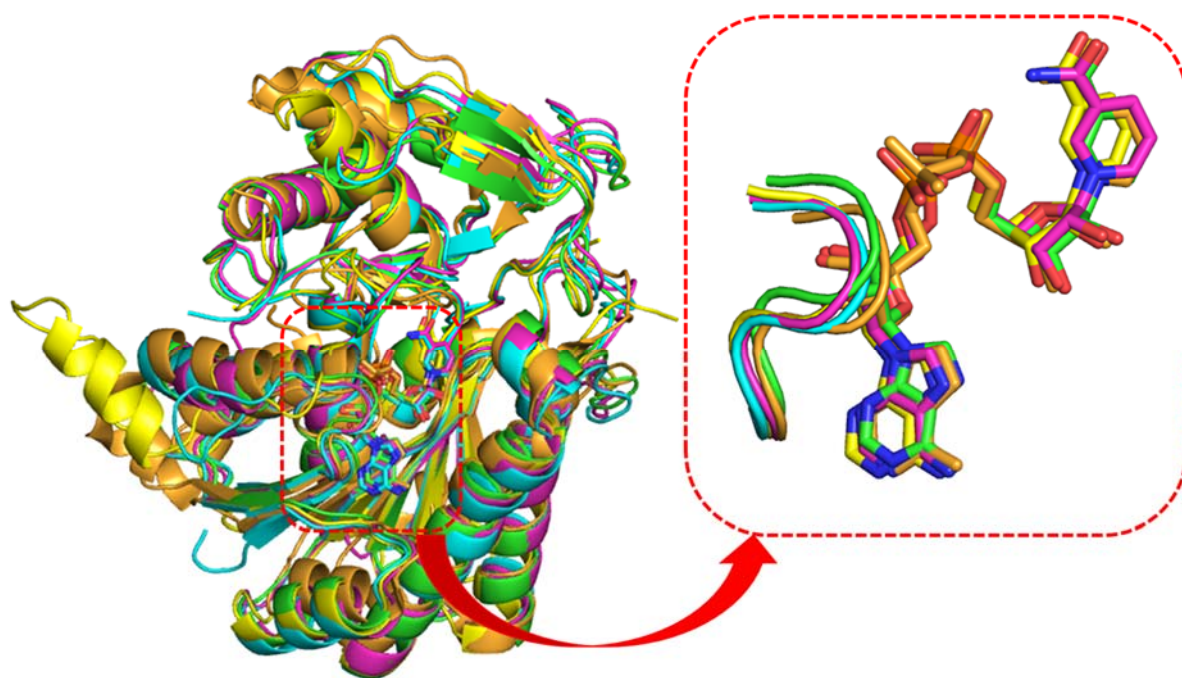**Fig. 3-5 Data collection used by helical mode.**

3-3-2 Overall structure of bGalE

All crystals of bGalE belong to the $P6_522$ space group. The structures of bGalE in complex with UDP, UDP-GlcNAc, and UDP-Glc were determined at 1.8, 2.0, and 1.8 Å resolution to R factor (R_{free}) of 16.4% (18.9%), 19.3% (23.0%), and 16.9% (19.6%), respectively (Table 3-2). The UDP-glucose complex was obtained by co-crystallization with 10 mM UDP-Gal, indicating that bGalE catalyzes the interconversion of the 4'-hydroxyl configuration. One molecule contained in an asymmetric unit of the crystal, whereas bGalE exists as a homodimer in solution (Fig. 3-3). The bGalE forms a homodimer by the formation of a four-helix bundle with each two α-helices. The overall protein structure can be roughly divided into an NAD^+ binding N-terminal domain (residues 1-177 and 236-262) and an UDP-hexose binding C-terminal domain (residues 178-235 and 263-340) (Fig. 3-6). The structure of bGalE adopts a typical Rossmann fold composed of a seven-stranded parallel β-sheet flanked by nine α-helices. The cofactor NAD^+ was intrinsically bound to all complexes. It was reported that all crystal structures of GalE contained NAD^+ or NADH molecule (8, 22). The UDP-hexose binding C-terminal domain possesses an α/β motif consisting of four α-helices and four β-strands. According to superimpositions of α-carbons with homologous GalEs, root mean square deviation (RMSD) showed 0.870 Å, 0.819 Å, 1.372 Å, and 0.759 Å in hGalE (PDB 1E6K), eGalE (PDB 1XEL), WbpP (PDB 1SB8), and GalE from *Trypanosoma brucel* (tGalE; PDB 2CNB), respectively (23). The bGalE structure is basically similar to those of homologous GalEs (Fig. 3-7), except for WbpP from *Pseudomonas aeruginosa*, which has an additional α-helix in the N-terminal domain (residues 3-16) (10).

Fig. 3-6 Overall structure of bGalE. (A) Homodimer (B) Monomer; The magenta and orange sticks indicate UDP-GlcNAc and NAD$^+$, respectively.

Fig. 3-7 Superimposition with homologous GalEs and NAD binding loops. bGalE (green), hGalE (cyan), eGalE (magenta), tGalE (yellow), and WbpP (orange) are shown. Inset, NAD$^+$ binding loop (residues 31-37 in bGalE).
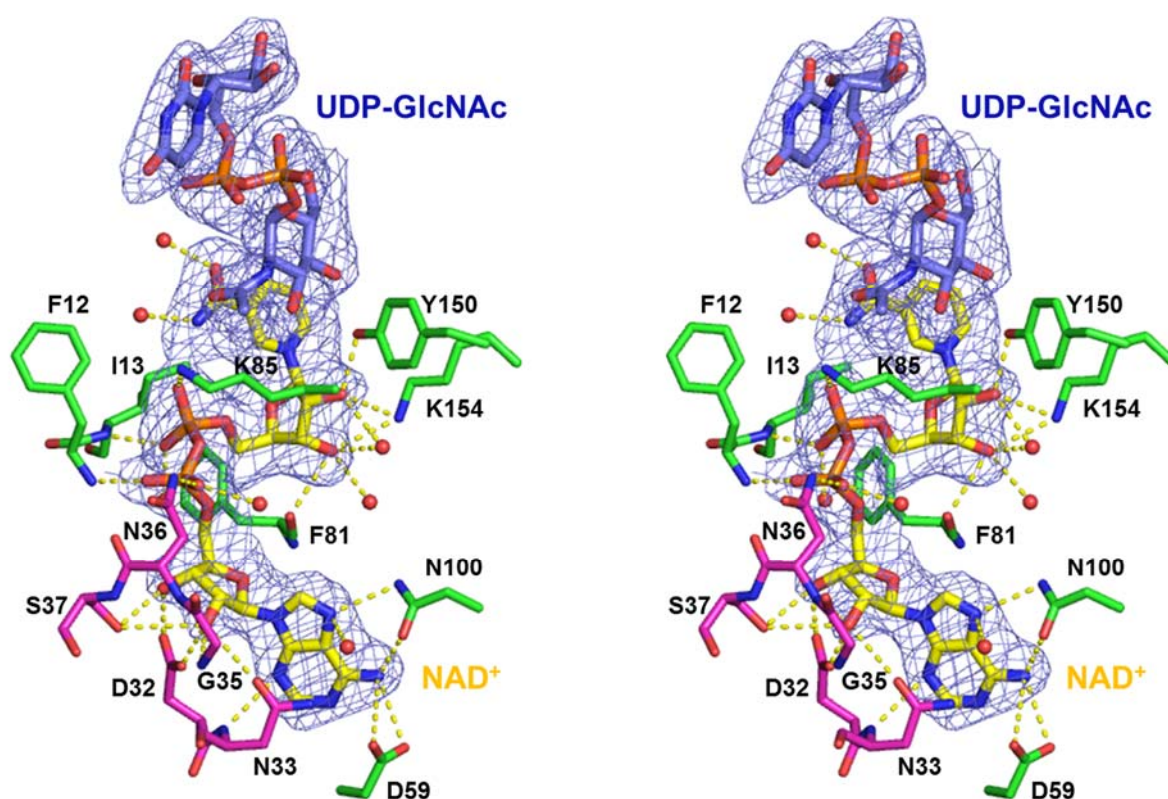
Table 3-2 Data collection and refinement statistics.

Data collection	UDP and NAD⁺	UDP-GlcNAc and NAD⁺	UDP-glucose and NAD⁺
Beam line	BL17A	BL5A	NW12A
Wavelength (Å)	0.9788	1.0	1.0
Space group	$P6_522$	$P6_522$	$P6_522$
Unit cell (Å)	a=70.235, b=70.235, c=320.578 $\alpha=\beta=90.000,\ \gamma=120.00$	a=69.45, b=69.45, c=321.986 $\alpha=\beta=90.000,\ \gamma=120.00$	a=70.359, b=70.359, c=322.553 $\alpha=\beta=90.000,\ \gamma=120.00$
Resolution (Å)[a]	50.00-1.80 (1.83-1.80)	50.00-2.0 (2.03-2.0)	50.00-1.8 (1.83-1.80)
Total reflections	809,9206	552,819	800,543
Unique reflections	44,831	32,725	45,385
Completeness (%)[a]	100.0 (100.0)	100.0 (100.0)	98.1 (97.0)
Redundancy[a]	18.1 (20.6)	16.9 (17.6)	18.0 (18.2)
Mean I/σ (I)[a]	31.9 (6.7)	28.1 (3.5)	37.8 (5.4)
R_{merge} (%)[a]	8.9 (52.0)	7.3 (57.8)	7.2 (40.5)
Refinement			
Resolution (Å)	60.825-1.801	60.145-1.996	60.933-1.800
No. of reflections	42,417	30,845	42,233
R-factor / R_{free} (%)	16.4 /18.9	19.3/23.0	16.9/19.6
No. of atoms	6,072	5,761	6,121
No. of solvents	1 (UDP), 1 (NAD⁺), 273 (water)	1 (UDP-GlcNAc), 1 (NAD⁺), 158(water)	1 (UDP-Glucose), 1 (NAD⁺), 279 (water)
RMSD from ideal values			
Bond lengths (Å)	0.0183	0.0184	0.0187
Bond angles (°)	1.92	1.96	1.96
Ramachandran plot (%)			
Favored residues	99.1	98.8	99.4
Allowed residues	0.6	0.9	0.3
Outlier residues	0.3	0.3	0.3

[a]Values in parentheses correspond to the highest resolution shell.

3-3-3 NAD⁺ binding site

The electron density of the NAD⁺ coenzyme bound in active site was clearly observed, which helped us to place the ligand with reasonable accuracy (Fig. 3-8). Moreover, the position of the nicotinamide ring of the dinucleotide was bound in the *syn* conformation. The nicotinamide ring of NAD⁺ in all complexes of GalEs in the oxidized form adopts the *syn* conformation (7, 24). The N6 atom f the adenine base interacts with the side chain of Asp59 and Asn100. The N7 atom forms a hydrogen bond with the side chain of Asn100, which corresponds to Val107 in hGalE and Asn 99 in eGalE. The hydroxyl groups of the adenine ribose interact with the NAD⁺ binding loop formed by Asp32, Asn33, Gly35, Asn36, and Ser37. The NAD binding loop exists in all homologue GalE (Fig. 3-7). The adenine phosphate interacts with the side chain Asn36, the backbone amide hydrogen of Phe12, and water molecule. On the other hand, the nicotinamide phosphate interacts with the backbone amide hydrogen of Ile13, the side chain of Lys85, and water molecule. The hydroxyl groups of the nicotinamide ribose interact with the side chain of Tyr150, Lys154, with the main chain of Phe81, and two water molecules. The nicotinamide ring interacts with only two water molecules.
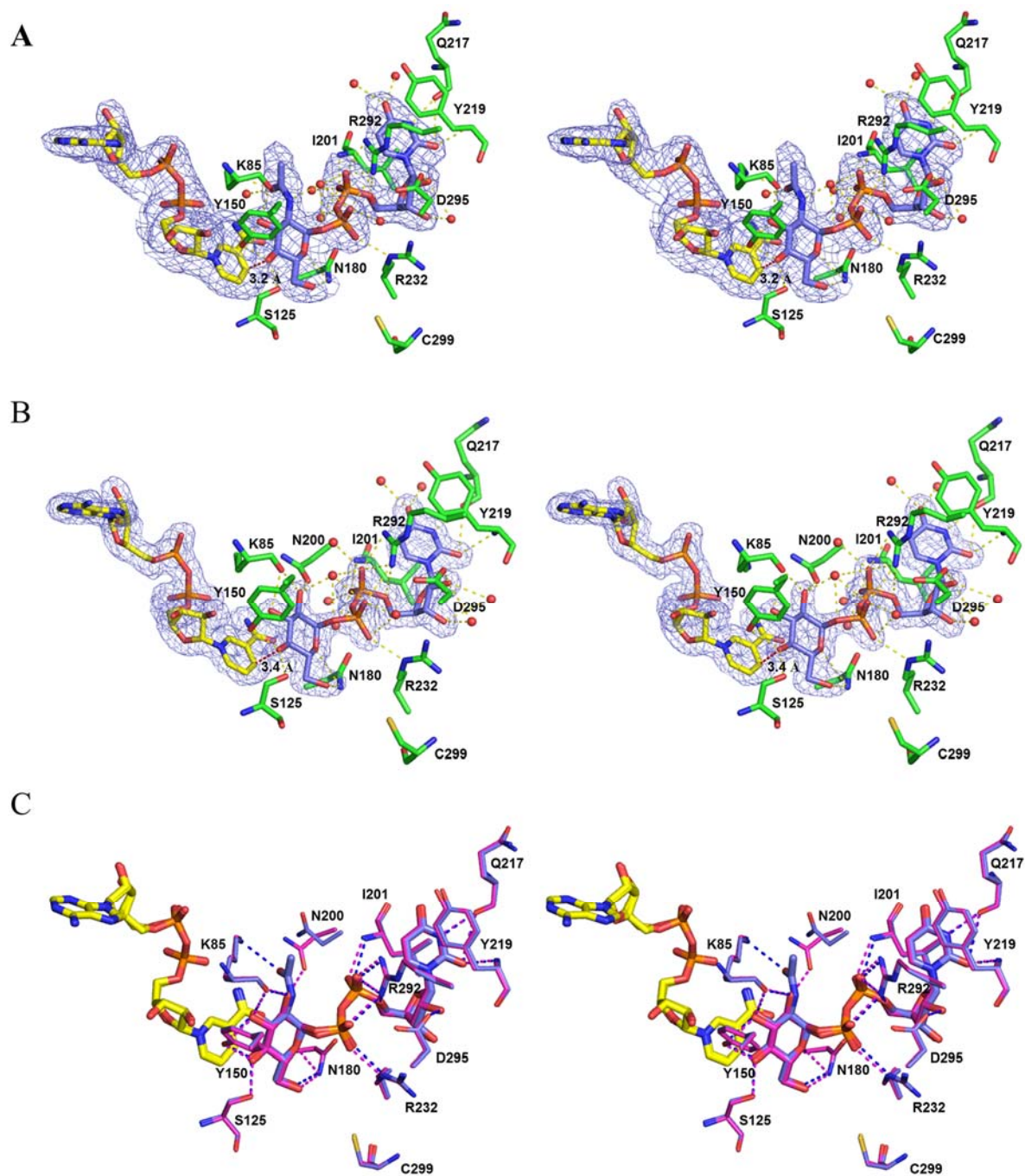

Fig. 3-8 NAD binding site (stereoview). Blue and yellow sticks indicate UDP-GlcNAc and NAD⁺, respectively. Magenta residues represents NAD binding loop. Hydrogen bonds are shown in yellow dash lines. The *2Fo-Fc* electron density map is shown as blue mesh and contoured at 1.5 σ. Red spheres indicate water molecules.

3-3-4 UDP-hexose binding site

Crystal structures of bGalE were determined also in complex with UDP-GlcNAc and UDP-glucose. Hydrogen bonding interactions between the enzyme and UDP-hexose ligand are indicated by the dashed lines in Fig. 3-9. There are 10 electrostatic interactions within 3.5 Å between the enzyme and the UDP hexose ligand. The uracil ring is recognized by the backbone carbonyl and amide groups of Gln217 and Tyr219. The 2'-hydroxyl group of the ribose interacts with the carboxylate side chain of Asp295 and this interaction was conserved on all homologous GalE (Fig. 3-10). Phosphate groups are recognized by backbone carbonyl Leu201 and the side chain of Arg232 and Arg292. In Fig 3-9 A, *N*-acetyl group of UDP-GlcNAc interacts with the side chain of Lys85 and one water molecule. The 2', 3', and 4'-hydroxyl groups of GlcNAc interact with the backbone carbonyl of Lys85, the side chain of Tyr150, and the side chain of Ser125. These interactions were are present in UDP-glucose complex (Fig. 3-9 B). On the other hand, there is a unique interaction in only UDP-glucose complex compared to UDP-GlcNAc complex. In structural superimposition between UDP-GlcNAc and UDP-glucose complex, the side chain conformation of Asn200 at the UDP-hexose binding site is different; it swings "in" or "out" in the UDP-glucose and UDP-GlcNAc complex structures, respectively (Fig. 3-9 C). This phenomenon was also reported for the orientation of Asn207 in UDP-GlcNAc complex as shown in one of subunits in the hGalE (8).

In both complex structures, the C4 atom of the UDP-GlcNAc/UDP-Glc was positioned at distances of 3.2 Å and 3.4 Å from the C4 atom of the nicotinamide ring of the dinucleotide, respectively. In the proposed reaction mechanism of GalE, a hydride transfer is thought to occur between these atoms (7). Additionally, the catalytic base Tyr150 which is conserved in all GalE enzymes, was positioned at 2.6 Å from the C4 atom of the UDP-GlcNAc/Glc. Therefore, these observation in UDP-hexose binding site was clearly elucidated that GlcNAc and glucose moieties in the complex structure are in the catalytically competent substrate orientation.
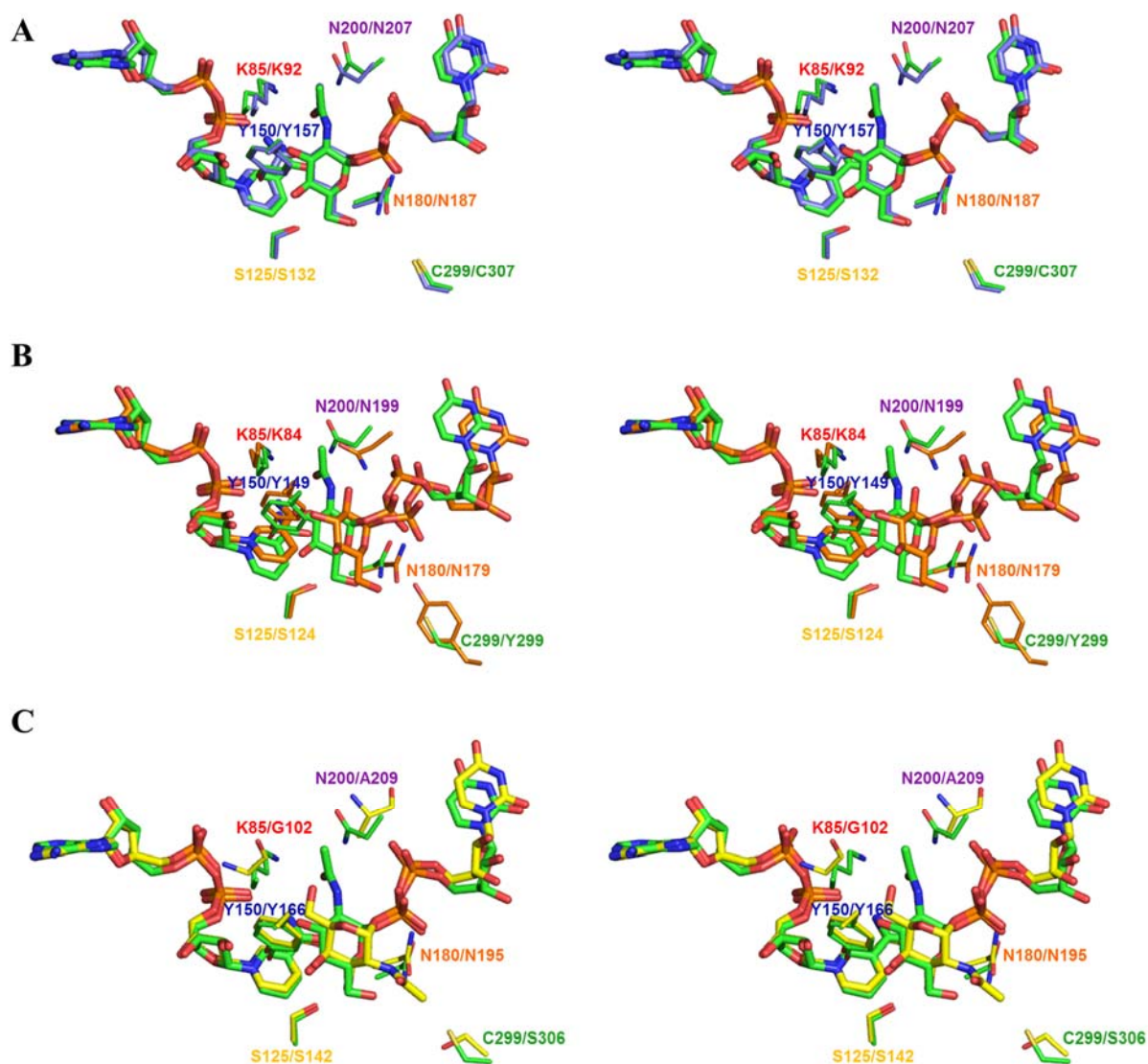
Fig. 3-9 UDP-hexose binding site of bGalE (Stereoview). (A) UDP-GlcNAc complex. (B) UDP-Glc complex. (C) Superimposition of UDP-GlcNAc and UDP-Glc complex. Hydrogen bonding interactions between the enzyme and the UDP-hexose ligand are indicated by dashed lines. Red dash lines indicate the distance between the C4 atom of the UDP-hexose and the C4 atom of the nicotinamide ring.

```
                    α4                α5                              β5      η1
                 ϙϙϙϙϙ  ϙϙϙϙϙϙϙϙϙϙϙϙϙϙϙϙϙϙ                          ➙➙➙➙   ϙϙϙϙ
bGalE            90          100         110                    120         130
                 ·           ·           ·                      ·           ·
                 KAVG                                                       SAT
bGalE   WVIHFAGLKAVGESVAKPIEYYDNNLYSTLVLLKVMKHNVKKIIFSSSATVYGT
hGalE   AVIHFAGLKAVGESVQKPLDYYRVNLTGTIQLLEIMKAHGVKNLVFSSSATVYGN
eGalE   TVIHFAGLKAVGESVQKPLEYYDNNVNGTLRLISAMRAANVKNFIFSSSATVYGD
tGalE   AVVHMCAFLAVGESVRDPLKYYDNNVVGILRLLQAMLLHKCDKIIFSSSAAIFGN
aGalE   EVWHIAANPDVRIGAENPDEIYRNNVLATYRLLEAMRKAGVSRIVFTSTSTVYGE
pGalE   VVFHFAANPEVRLSTTBPIVHFNENVVATFNVLEWARQTGVRTVVFASSSTVYGD
WbpP    YVLHQAALGSVPRSINDPITSNATNIDGFLNMLIAARDAKVQSFTYAASSSTYGD

                       α6                       β6      β7
                 ϙϙϙϙϙϙϙϙϙϙϙϙϙϙϙϙϙϙϙϙϙ       ➙➙➙➙➙  ➙➙➙➙          TT
bGalE            150         160         170         180
                 ·           ·           ·           ·
                 NPYG                                YFNP
bGalE   TNPYGTSKLFQEQILRDVHVADPSWTIVLLRYFNPVGAHES
hGalE   TNPYGKSKFFIEEMIRDLCQADKTWNAVLLRYFNPTGAHAS
eGalE   QSPYGKSKLMVEQILTDLQKAQPDWSIALLRYFNPVGAHPS
tGalE   ESPYGESKLIAERMIRDCAEAYG.IKGICLRYFNACGAHED
aGalE   ISLYGASKLACEALIESYCHTFD.MQAWIYRFANVIGRRST
pGalE   ISVYGAAKAAGEVMCATYARLFG.VRCLAVRYANVVGPRLR
WbpP    LSPYAVTKYVNELYADVFSRCYG.FSTIGLRYFNVFGRRQD

                              α7                T            α10
                          ϙϙϙϙϙϙϙϙϙ                     ϙϙϙϙϙϙϙ
bGalE   190           200         210           300         310
        ·             ·           ·             ·           ·
                 DPKG       I PAN              IAAC
bGalE   GLLGEDPKG....I PANLTPYVAKVAVG    GDIAACYADASKAEKELGWK
hGalE   GCIGEDPQG....I PNNLMPYVSQVAIG    GDVAACYANPSLAQEELGWT
eGalE   GDMGEDPQG....I PNNLMPYIAQVAVG    GDLPAYWADASKADRELNWR
tGalE   GDIGEHYQGSTHLI PIILGRVMSDIAPD    GDPAYLVAASDKAREVLGWK
aGalE   .......HG.......VIYDFIMKLKRN     GDVPVMLLSIEKLKRL.GWK
pGalE   .......HG.......VIYDFIMKLRRN     GDVKYMTLAVTKLMKLTGWR
WbpP    .......PNG...AYAAVIPKWTSSMIQG    GDVRHSLADISKAAKLLGYA
```

Fig. 3-10 Partial amino acid sequence alignment of GalE enzymes. Structure based aminoacid sequence alignment of the bGalE from *B. longum*, hGalE from *human*, eGalE from *E. Coli*, tGalE from *Trypanosoma brucei*, aGalE from *Archaeoglobus fulgidus*, pGalE from *Pyrobaculum. Calidifontis,* and WbpP from *Pseudomonas aeruginosa.* The residues in square box indicate one of the six side walls composed of substate binding pocket. Blue stars indicate the SYK catalytic triad.

3-3-5 Structural basis for substrate specificity in bGalE

According to the superimposition of substrate binding pocket in represented group 1, 2, and 3 enzymes (Fig. 3-1), all GalE enzymes conserved the residues Ser, Tyr, and Asn (Fig. 3-11; S125, Y150, and N180 in bGalE). They are part of the conserved SYK catalytic triad (tGalE: S142, Y173, and K177; eGalE: S124, Y149, and K153; hGalE: S132, Y157, and K161; bGalE: S125, Y150, and K154; WbpP: S142, Y166, and K170 shown in Fig. 3-10). In the superimposition of bGalE and hGale in complex with UDP-GlcNAc, the residues were conserved well and N200/N207 adapts the "swing out" conformation to accommodate *N*-acetylated sugar (Fig. 3-11A). The asparagine in bGalE moves to the "swing in" conformation, when UDP-Glc sugar is bound to active site, as described in Fig. 3-9C. In the case of eGalE, the Asn199 was fixed in the "swing in" conformation, whatever substrates were bound to active site (25). The bulky side chain of Tyr299 in eGalE makes the saccharide binding pocket narrower, in comparison with Cys299 in bGalE. It was discussed that the bulky Try299 of eGalE interrupts the binding of flipped UDP-4-keto-GlcNAc as intermediate (25). In eGalE in complex with UDP-GlcNAc (PDB 1LRJ), it was reported that acetylated moiety of sugar was bound in a wrong position and the interaction between 2'-hydroxyl group of UDP-GlcNAc and Asn199 was missing, because of the bulky Tyr299 (25). However, Y299C mutant resulted in an overall 20% increase in the volume of the saccharide binding pocket and a loss of epimerase activity to UDP-Gal but an increase of activity to UDP-GalNAc (Table 3-3) (25). In comparison between bGalE and WbpP in active site, Ala209 had the same orientation with Asn200 in bGalE structure, even though UDP-GalNAc is in a "flipped" conformation. The Ala209 residue does not show any "swing in or out" in the substrate binding pocket. The small Ala209 and Sre306 residues make a bigger saccharide binding pocket comparing with other homologous GalE enzymes that have Tyr here. According to these results, the small size residues (Cys or Ser) allowed accommodating the large *N*-acetylated sugar into the saccharide binding pocket of GalE enzymes.

Fig. 3-11 Substrate binding pockets in GalE enzymes. (A) Superimposition of bGalE (green) and hGalE (blue). (B) Superimposition of bGalE (green) and eGalE (orange). (C) Superimposition of bGalE (green) and WbpP (flipped; yellow).

3-4 Discussion

The cofactor NAD^+ was intrinsically bound to bGalE without an addition in solutions throughout my experimental processes in all complexes, indicating that NAD^+ is tightly bound to GalE. It was reported that when NAD was removed out from GalE, the protein was irreversibly denatured (7, 9). In contrast to GalE enzymes, NAD-dependent dehydrogenases, L-ThrDHs from the psychrophilic bacterium *Flavobacterium frigidimaris* KUC-1, can easily release NAD^+ from the enzymes (26). According to the structural comparison of L-ThrDHs with eGalE and hGalE, overall structure was similar but topology was different in the loop around the NAD binding site and the NAD binding loop was absent in L-ThrDHs (26). Moreover, Sakuraba *et all*. constructed GalE loop deletion mutant for investigating the role of the NAD binding loop (26). The mutant formed inclusion body during the purification and lost epimerase activity. The all GalE homologues have the NAD binding loop and the interactions between NAD and residues around the NAD binding loop are conserved. Therefore, the NAD binding loop plays a key role in preventing the release of the cofactor from GalE enzymes and contributing to the stability of protein.
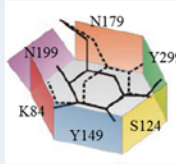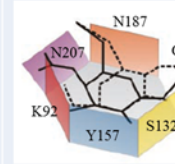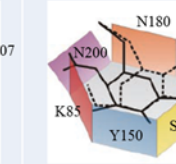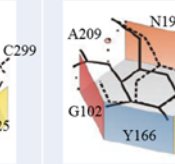
In enzyme activity of GalE enzymes (Table 3-3), it was reported that bGalE catalyzes both acetylated and non-acetylated substrates with comparable activities. The hGalE can also equally epimerize both UDP-GalNAc and UDP-gal. On the other hand, eGalE is highly specific for UDP-gal, whereas WbpP shows strong preference for acetylated substrates

Table 3-3. Enzyme activities of GalE enzymes.

	UDP-Glc	UDP-Gal	UDP-GlcNAc	UDP-GalNAc	Note	*Ref.*
bGalE	–	158	147	–	Specific acitivity (U/mg) at 1 mM UDP-sugar	(1)
hGalE	–	33.8	–	26		
eGalE	–	23.9	–	0.003	Specific acitivity (U/mg) at 0.41 mM UDP-Gal or 0.66 mM UDP-GalNAc	(25)
eGalE (Y299C)	–	5.1	–	0.69		
tGalE	–	5.9	–	–	Specific activity (U/mg) at 20 mM UDP-Gal	(27)
WbpP	0.124	0.188	120	271	k_{cat} (min^{-1}). K_m values were similar for all substrates (197~251 uM)	(28)

These GalE enzymes have been well characterized, and their catalytic mechanisms were studied well with their structural properties (Table 3-4) (1, 25, 27, 28). However, Ishiyama and Demendi *et al*. mentioned (10, 11) that the most studies of GalE enzymes have focused on refining the catalytic mechanism (5-7, 9, 29-32). Only few have studied the molecular basis for substrate specificity (8, 13, 15, 25, 33). Thus, they constructed a conceptual model to facilitate a better understanding of the factors

governing the substrate specificity spectra in GalEs based on the analysis of the relationship between amino acid sequence, three-dimensional structure, and substrate preference (Fig. 3-12) (10, 11).
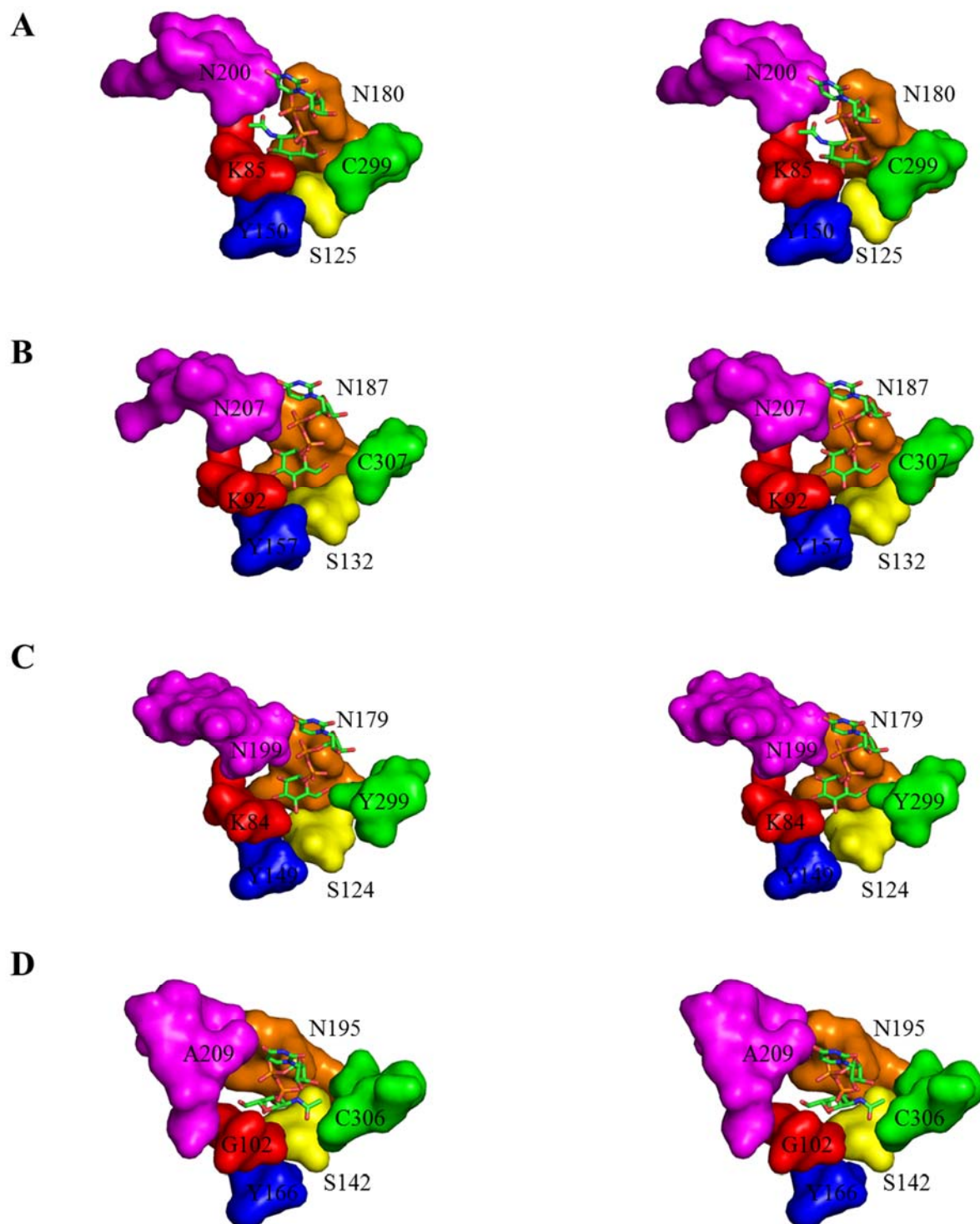
	tGalE	eGalE	hGalE	bGalE	WbpP
	(Trypanosoma brucei)	*(Escherichia coli)*	*(Homo sapiens)*	*(Bifidobacterium longum)*	*(Pseudomonas aeruginosa)*
Key Residues					
	LSYNHL	KSYNNY	KSYNNc	KSYNNc	GSYNas
UDP-GlcNAc	–	–	+	+	+
UDP-GalNAc	–	–	+	+	+
UDP-Glc	+	+	+	+	–
UDP-Gal	+	+	+	+	–
Substrate spectrum	Group 1a	Group 1b	Group 2	Group 2	Group 3

Fig. 3-12 A conceptual model for saccharide-binding pockets and substrate specificity spectra in GalEs. The substrate specificity spectra of tGalE, eGalE, hGalE, bGalE, and WbpP are represented by modifying Fig. 5 in reference (10). The sugar moieties of UDP-Glc(NAc) bound in the catalytically productive conformation ("standard" orientation) are represented as black solid lines. Similarly, the sugar moieties of UDP-Gal(NAc) bound in the "flipped" orientation are represented as dashed lines.

In the conceptual model, tGalE, eGalE, hGalE, bGalE and WbpP are classified into Group 1a, 1b, 2, 2, and 3 based on their substrate specificity, respectively. The bottom of the hexagonally shaped box is formed by the nicotinamide ring of the NAD cofactor and six walls are formed by different regions of the enzymes. The sugar moieties of UDP-Glc(NAc) or UDP-Gal(NAc) bound in the saccharide binding pocket in the "standard" orientation that was observed in bGalE, hGalE, and eGalE (Fig. 3-11AB). On the other hand, the "flipped" orientation that was observed in WbpP (Fig. 3-11C, yellow) is shown as dashed lines. A magenta wall in eGalE, hGalE, and bGalE are open by the "swing out" rotation of Asn side chain to accommodate the rotation of the saccharide ring for the reaction mechanism. The saccharide binding pocket in tGalE is fully closed by six walls, allowing to bind only non-acetylated UDP-hexose whereas the saccharide binding pocket in eGalE is semi-open by Asn199, indicating that acetylated hexose could bind and the 4-keto intermediate could be formed. However, flip of sugar
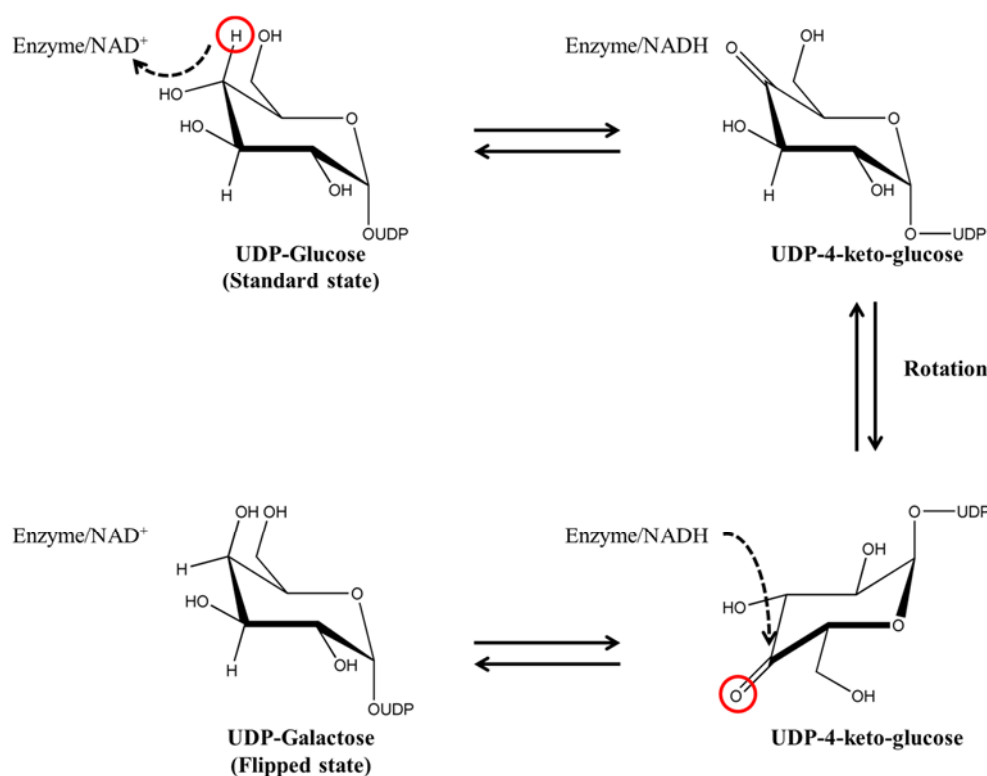
moiety is blocked by the bulky residue Y299 for epimerase reaction. It was reported that when the structural equivalent of Tyr299 in *E. coli* GalE was substituted with Cys, resulting in a loss of epimerase activity to UDP-Gal by almost 5-fold and resulting in a gain of activity against UDP-GalNAc by more than 230-fold (Table 3-3) (25). Thus, this mutation of eGalE changed its substrate spectrum from group 1 to group 2. Actually, the saccharide binding pocket of bGalE is basically the same as that of hGalE. In group 2 and 3 enzymes, two walls adopt a semi-open or open form to accommodate the acetylated substrate. Especially, the small side chains of cysteine and serine make enough space to bind *N*-acetylated sugar compared to leucine and tyrosine in group 1.It also allow to occur the flipped rotation of the *N*-acetylated sugar in the saccharide binding pocket. Moreover, in group 3 WbpP, two of the walls are removed by shorter side chain alanine and serine, contributing to make additional space for binding ordered water molecules in active site, which can be enabled to bind *N*- acetylated sugar. Water molecules in this site (shown as spheres in Fig. 3-12) are thought to contribute to the strong preference to UDP-GlcNAc/GalNAc than UDP-Glc/Gal (10).

For easily understanding the substrate binding pocket size, the active sites in GalE enzymes are visualized by a molecular surface model (Fig. 3-13). In the substrate binding pockets of both bGalE and hGalE, magenta and green walls make an enough space to bind the *N*-acetyl group in both of the "standard" and "flipped" orientations. On the other hand, in eGalE, the green wall is obviously protruded toward the active site, causing interruption of sugar moiety rotation. WbpP has a biggest substrate binding pocket in GalE enzymes. The magenta and green wall are far away from active site and allow to accommodate acetylated sugars. In conclusion, these different shapes and sizes of the saccharide binding pocket in GalE enzymes lead to the distinctive substrate specificity.

Fig. 3-13 Visualized substrate binding pockets in GalE enzymes. (A) bGalE in complex with UDP-GlcNAc. (B) hGalE in complex with UDP-Glc. (C) eGalE in complex with UDP-Glc. (D) WbpP in complex with UDP-GalNAc (flipped). The residues of colored surface are described in Fig. 3-13.

In the proposed epimerase reaction mechanism of GalE (Fig. 3-14), the first step is an abstraction of a hydride from the C4 hydroxyl group of UDP-Glc(NAc), leading to hydride transfer to the C4 atom on the *si*-face NAD$^+$, and formation of UDP-4-keto-Glc(NAc) and NADH as an intermediate state (Fig. 3-14). Next, UDP-4-keto-Glc(NAc) in the active site rotates and is positioned the opposite face of the sugar to the reduced dinucleotide. Finally, the nicotinamide ring of NADH transfers the hydride back to the opposite face of C4 keto-intermediate and the intermediate sugar forms the epimer.

Fig. 3-14 Mechanism of UDP-galactose 4-epimerase.

In general introduction, I described that all bifidobacteria do not have the GLNBP and do not utilize GNB/LNB pathway. LNB and GNB are present in human milk and human gastrointestinal tract. Especially, type I oligosaccharides are predominant in HMO. bGalE involved in GNB/LNB pathway could be specialized for GNB and LNB saccharide during molecular evolution. Moreover, the substrate specificity of GalEs can not be easily predicted by amino acid sequence by a phylogenetic tree analysis (Fig. 3-1). But a few amino acids (ex N200 and C299 in bGalE) are responsible for that. Therefore, many studies about GalE enzymes have showed that substrate specificity can be modified by simply changing one or two residues at active site (5, 25, 31, 34). Thus, Demendi *et al.* proposed substrate specificity modification of WbpP enzyme by the binding pocket residues have co-evolved with early evolution of group 1a, 1b, 2, and 3 (11).

A survey for GalE structures complexed with UDP-hexose substrates in catalytically competent conformations is shown in Table 3-4. The "flipped" orientation was not easily observed with natural substrates and wild-type enzymes, except for WbpP complexed with GlcNAc. In eGalE, a flipped orientation structure was only observed by using a S124A/Y149F double mutant (5). For tGalE, a flipped structure was determined by using a substrate analog UDP-4-deoxy-4-fluoro-α-galactose (UDP-4fGal) (34).

In conclusion, the crystal structure of bGalE involved in the GNB/LNB pathway showed a structural basis of a UDP-galactose 4-epimerase that catalyzes the conversion between UDP-GlcNAc and UDP-GalNAc or UDP-Glc and UDP-Gal. According to the structural basis for the substrate specificity, bGalE has the saccharide binding pocket for catalyzing *N*-acetylated sugar or non-acetylated sugar similar to substrate binding of hGalE, which is the result of the evolution for utilizing various oligosaccharides in the human gastrointestinal tract, where common mono-and disaccharides are insufficient.

Table 3-4 A survey for the structure of GalE enzymes up to date.

Organism	Org. note	UniProt	Phyl. Group	Activity on	85	125	150	180	200	299	PDBID	Ligand	C4-C4 dist. (A)	(B)	C2 res.	C6 res.	Mutation	Conf.	Resol.	Year	PMID	Cofactor	Note
Bifidobacterium longum	Bacteria	E8MF10	2	Glc & GlcNAc	K	S	Y	N	N	c	–	UDP							1.80			NAD+	
Bifidobacterium longum											–	UDP-Glc	3.1		N200	C299		Std	1.80			NAD+	UDP-Gal soaked
Bifidobacterium longum											–	UDP-GlcNAc	3.0		N200	C299		Std	2.00			NAD+	
Escherichia coli	Bacteria	P09147	1b	Glc	K	S	Y	N	N	Y	1XEL	UDP-Glc	3.7		N199	Y299		Std	1.80	1996	8611497	NADH	
Escherichia coli											1A9Z	UDP-Gal	3.5		Y299	F(Y)149	S124A/Y149F	Flip	1.90	1998	9708982	NAD+	
Homo sapiens	Human	Q14376	2	Glc & GlcNAc	K	S	Y	N	N	Y	1EK6	UDP-Glc	3.5	3.5	N207	Y157		Std	1.80	2000	10801319	NADH	
Homo sapiens											1HZJ	UDP-GlcNAc	(4.6)	3.0	N207	Y157		Std (B)	1.50	2001	11279032	NADH	A chain was in a bad orientation.
Pseudomonas aeruginosa	Bacteria	Q8KN66	3	GlcNAc	G	S	Y	N	A	S	1SB8	UDP-GalNAc	3.2		S306	A209		Flip	2.1	2004	15016816	NAD+	WbpP
Trypanosoma brucei	Protist	Q8T8E9	1a	Glc	L	S	Y	N	H	L	2CNB	UDP-4fGal	3.0	3.1	L342	H221		Flip	2.70	2003	12615316	NAD+	
Aspergillus nidulans	Fungi	C8VAU8	1b	Glc	K	S	Y	N	N	L	4LIS	UDP-Glc	3.5		N219	L320		Std	2.80	2013	24116166	NAD+	

3-5 References

1. **Nishimoto, M., and Kitaoka, M.** (2007) Identification of N-acetylhexosamine 1-kinase in the complete lacto-N-biose I/galacto-N-biose metabolic pathway in Bifidobacterium longum, *Appl Environ Microbiol 73*, 6444-6449.

2. **Ross, K. L., Davis, C. N., and Fridovich-Keil, J. L.** (2004) Differential roles of the Leloir pathway enzymes and metabolites in defining galactose sensitivity in yeast, *Mol Genet Metab 83*, 103-116.

3. **Hidaka, M., Nishimoto, M., Kitaoka, M., Wakagi, T., Shoun, H., and Fushinobu, S.** (2009) The crystal structure of galacto-N-biose/lacto-N-biose I phosphorylase: a large deformation of a TIM barrel scaffold, *J Biol Chem 284*, 7273-7283.

4. **Wang, K. C., Lyu, S. Y., Liu, Y. C., Chang, C. Y., Wu, C. J., and Li, T. L.** (2014) Insights into the binding specificity and catalytic mechanism of N-acetylhexosamine 1-phosphate kinases through multiple reaction complexes, *Acta Crystallogr D 70*, 1401-1410.

5. **Thoden, J. B., and Holden, H. M.** (1998) Dramatic differences in the binding of UDP-galactose and UDP-glucose to UDP-galactose 4-epimerase from Escherichia coli, *Biochemistry 37*, 11469-11477.

6. **Thoden, J. B., Hegeman, A. D., Wesenberg, G., Chapeau, M. C., Frey, P. A., and Holden, H. M.** (1997) Structural analysis of UDP-sugar binding to UDP-galactose 4-epimerase from Escherichia coli, *Biochemistry 36*, 6294-6304.

7. **Thoden, J. B., Frey, P. A., and Holden, H. M.** (1996) Crystal structures of the oxidized and reduced forms of UDP-galactose 4-epimerase isolated from Escherichia coli, *Biochemistry 35*, 2557-2566.

8. **Thoden, J. B., Wohlers, T. M., Fridovich-Keil, J. L., and Holden, H. M.** (2001) Human UDP-galactose 4-epimerase. Accommodation of UDP-N-acetylglucosamine within the active site, *J Biol Chem 276*, 15131-15136.

9. **Thoden, J. B., Wohlers, T. M., Fridovich-Keil, J. L., and Holden, H. M.** (2000) Crystallographic evidence for Tyr 157 functioning as the active site base in human UDP-galactose 4-epimerase, *Biochemistry 39*, 5691-5701.

10. **Ishiyama, N., Creuzenet, C., Lam, J. S., and Berghuis, A. M.** (2004) Crystal structure of WbpP, a genuine UDP-N-acetylglucosamine 4-epimerase from Pseudomonas aeruginosa: substrate specificity in udp-hexose 4-epimerases, *J Biol Chem 279*, 22635-22642.

11. **Demendi, M., Ishiyama, N., Lam, J. S., Berghuis, A. M., and Creuzenet, C.** (2005) Towards a better understanding of the substrate specificity of the UDP-N-acetylglucosamine C4 epimerase WbpP, *Biochem J 389*, 173-180.

12.	**Li, K. B.** (2003) ClustalW-MPI: ClustalW analysis using distributed and parallel computing, *Bioinformatics 19*, 1585-1586.

13.	**Bernatchez, S., Szymanski, C. M., Ishiyama, N., Li, J., Jarrell, H. C., Lau, P. C., Berghuis, A. M., Young, N. M., and Wakarchuk, W. W.** (2005) A single bifunctional UDP-GlcNAc/Glc 4-epimerase supports the synthesis of three cell surface glycoconjugates in Campylobacter jejuni, *J Biol Chem 280*, 4792-4802.

14.	**Soldo, B., Scotti, C., Karamata, D., and Lazarevic, V.** (2003) The Bacillus subtilis Gne (GneA, GalE) protein can catalyse UDP-glucose as well as UDP-N-acetylglucosamine 4-epimerisation, *Gene 319*, 65-69.

15.	**Bengoechea, J. A., Pinta, E., Salminen, T., Oertelt, C., Holst, O., Radziejewska-Lebrecht, J., Piotrowska-Seget, Z., Venho, R., and Skurnik, M.** (2002) Functional characterization of Gne (UDP-N-acetylglucosamine-4-epimerase), Wzz (chain length determinant), and Wzy (O-antigen polymerase) of Yersinia enterocolitica serotype O:8, *J Bacteriol 184*, 4277-4287.

16.	**Bauer, A. J., Rayment, I., Frey, P. A., and Holden, H. M.** (1992) The molecular structure of UDP-galactose 4-epimerase from Escherichia coli determined at 2.5 A resolution, *Proteins 12*, 372-381.

17.	**Otwinowski, Z., and Minor, W.** (1997) Processing of X-ray diffraction data collected in oscillation mode, *Method Enzymol 276*, 307-326.

18.	**Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H., and Hung, L. W.** (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard, *Acta Crystallogr D Biol Crystallogr 65*, 582-601.

19.	**Cowtan, K., Emsley, P., and Wilson, K. S.** (2011) From crystal to structure with CCP4, *Acta Crystallogr D Biol Crystallogr 67*, 233-234.

20.	**Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K.** (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr 66*, 486-501.

21.	**Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F.** (1999) ESPript: analysis of multiple sequence alignments in PostScript, *Bioinformatics 15*, 305-308.

22.	**Inoue, T., Nishio, N., Suzuki, S., Kataoka, K., Kohzuma, T., and Kai, Y.** (1999) Crystal structure determinations of oxidized and reduced pseudoazurins from Achromobacter cycloclastes. Concerted movement of copper site in redox forms with the rearrangement of hydrogen bond at a remote histidine, *J Biol Chem 274*, 17845-17852.

23.	**Alphey, M. S., Burton, A., Urbaniak, M. D., Boons, G. J., Ferguson, M. A., and Hunter, W. N.** (2006) Trypanosoma brucei UDP-galactose-4'-epimerase in ternary complex with NAD+ and the substrate analogue UDP-4-deoxy-4-fluoro-alpha-D-galactose, *Acta Crystallogr Sect F Struct Biol Cryst Commun 62*, 829-834.

24. **Bauer, A. J., Rayment, I., Frey, P. A., and Holden, H. M.** (1991) The isolation, purification, and preliminary crystallographic characterization of UDP-galactose-4-epimerase from Escherichia coli, *Proteins 9*, 135-142.

25. **Thoden, J. B., Henderson, J. M., Fridovich-Keil, J. L., and Holden, H. M.** (2002) Structural analysis of the Y299C mutant of Escherichia coli UDP-galactose 4-epimerase. Teaching an old dog new tricks, *J Biol Chem 277*, 27528-27534.

26. **Sakuraba, H., Kawai, T., Yoneda, K., and Ohshima, T.** (2011) Crystal structure of UDP-galactose 4-epimerase from the hyperthermophilic archaeon Pyrobaculum calidifontis, *Arch Biochem Biophys 512*, 126-134.

27. **Roper, J. R., Guther, M. L. S., Milne, K. G., and Ferguson, M. A. J.** (2002) Galactose metabolism is essential for the African sleeping sickness parasite Trypanosoma brucei, *P Natl Acad Sci USA 99*, 5884-5889.

28. **Creuzenet, C., Belanger, M., Wakarchuk, W. W., and Lam, J. S.** (2000) Expression, purification, and biochemical characterization of WbpP, a new UDP-GlcNAc C4 epimerase from Pseudomonas aeruginosa serotype O6, *J Biol Chem 275*, 19060-19067.

29. **Thoden, J. B., Frey, P. A., and Holden, H. M.** (1996) High-resolution X-ray structure of UDP-galactose 4-epimerase complexed with UDP-phenol, *Protein Sci 5*, 2149-2161.

30. **Thoden, J. B., Frey, P. A., and Holden, H. M.** (1996) Molecular structure of the NADH/UDP-glucose abortive complex of UDP-galactose 4-epimerase from Escherichia coli: implications for the catalytic mechanism, *Biochemistry 35*, 5137-5144.

31. **Thoden, J. B., Gulick, A. M., and Holden, H. M.** (1997) Molecular structures of the S124A, S124T, and S124V site-directed mutants of UDP-galactose 4-epimerase from Escherichia coli, *Biochemistry 36*, 10685-10695.

32. **Liu, Y., Thoden, J. B., Kim, J., Berger, E., Gulick, A. M., Ruzicka, F. J., Holden, H. M., and Frey, P. A.** (1997) Mechanistic roles of tyrosine 149 and serine 124 in UDP-galactose 4-epimerase from Escherichia coli, *Biochemistry 36*, 10675-10684.

33. **Schulz, J. M., Watson, A. L., Sanders, R., Ross, K. L., Thoden, J. B., Holden, H. M., and Fridovich-Keil, J. L.** (2004) Determinants of function and substrate specificity in human UDP-galactose 4'-epimerase, *J Biol Chem 279*, 32796-32803.

34. **Thoden, J. B., Wohlers, T. M., Fridovich-Keil, J. L., and Holden, H. M.** (2001) Molecular basis for severe epimerase deficiency galactosemia. X-ray structure of the human V94m-substituted UDP-galactose 4-epimerase, *J Biol Chem 276*, 20617-20623.

Chapter 4 Concluding Remarks

The studies of metabolism are basic for understanding life sciences. Sugar metabolic pathways, one of major metabolisms, are essential for life. The sugar metabolic pathways are based on sequential actions of various enzymes, such as hydrolases, oxidoreductases, transferases, lyases, or isomerases. The researches of functional and structural features for novel enzymes contribute to understand reaction mechanisms and to discover new sugar metabolic pathways. In this study, I study enzymes in two novel alternative sugar metabolic pathways from microbes; intracellular pathways linked to oxidative cellulose degradation and bifidobacteral consumption of human milk oligo saccharides and intestinal mucin glycoproteins. These two novel pathways finally linked to the glycolysis pathway. However, CBAP and bGalE involved in these pathways had been still unknown about catalytic mechanism or the reasons why the enzymes have broad substrate specificity with a concrete structural basis. Therefore, X-ray crystallography and a mutational analysis of CBAP and bGalE were performed to study their reaction mechanism, the catalytic residues, and substrate specificity.

In chapter 2, X-ray crystallography was performed on CBAP from *S. degradans* with the selenomethionine-SAD phasing method. CBAP catalyzes reversible phosphorolysis of CbA or Glc-β1,3-GlcUA. To analyze the catalytic mechanism, crystal structures of CBAP in complex with Apo, CbA, GlcA, and Glc-β1,3-GlcUA was determined. In the CbA complex, the glucose and the GlcA moietes of CbA are bound at subsite -1 and +1, respectively, exhibiting a catalytically plausible structure for the phosphorolytic reaction. The C1 carboxyl group of the GlcA moiety of CbA interacts with R609 and K613 at subsite +1. In the GlcA complex, a dehydrated compound LGC and GlcA are bound at subsite -1 and +1, respectively. The C1 carboxyl group of the GlcA interacts with R609, K613, and Q190', advocating that they are key residues for the molecular recognition of the aldonic acid. On the other hand, in the Glc-β1,3-GlcUA complex, the GlcUA moiety forms as a pyranose and binds at a displaced position. The C6 carboxyl group is not recognized by R609, K613, and Q190'. The conformation of Glc-β1,3-GlcUA was different from that of CbA at active site. To examine the importance of these residues on binding of the aldonic acid and the uronic acid, a mutational analysis was conducted. According to the result, R609, K613, and Q190' residues are required for both aldonic acid and uronic acid, indicating that Glc-β1,3-GlcUA complex was an artificial structure due to the high concentration of sulfate ion compared to the concentration of the disaccharide ligand used for the soaking experiment. Moreover, a possible binding model of GlcUA at subsite +1 elucidates that C6 carboxyl group can be placed at the site surrounded by R609, K613, and Q190'. According to these results, I conclude that both CbA and Glc-β1,3-GlcUA are catalyzed by the CBAP in the same manner. The active site in CBAP was compared with those of GH94 enzymes, showing that CBAP has a unique subsite +1 in the aspect of substrate specificity.

In chapter 3, X-ray crystallography was performed on GalE from *B. longum* for substrate specificity. The bGalE structure in complex with UDP, UDP-GlcNAc, and UDP-Glc was determined with molecular replacement. The overall structure can be roughly divided into an NAD^+ binding N-terminal domain and an UDP-hexose binding C-terminal domain. The structure of bGalE is typical Rossmann fold composed of a seven-strand parallel β-sheet flanked by nine α-helices. The cofactor NAD^+ was intrinsically bound to all complexes. The nicotinamide ring of NAD^+ in all complexes forms an oxidized form adopting *syn* conformation. In structural superimposition between UDP-GlcNAc and UDP-Glc complex, the orientation of side chain of Asn200 at the UDP-hexose binding site was different; it swings "in" and "out" in the complex structures with UDP-Glc and UDP-GlcNAc, respectively. In both complexes, C4 of UDP-GlcNAc/UDP-Glc was positioned at 3.2 Å and 3.4 Å from C4 of the nicotinamide ring, respectively, indicating that C4 hydroxyl group of UDP-hexose transfers the C4 hydride to cofactor. According to the superimposition of substrate binding pocket in represented group 1, 2, and 3 enzymes, Asn200 in bGalE and Asn207 in hGalE move "swing in" or "swing out" depending on binding of *N*-acetylated or non-acetylated sugars. On the other hand, Asn199 in eGalE was fixed to be "swing in" and bulky Tyr299 side chain makes the substrate binding pocket narrower compared to Cys299 in bGalE. In comparison between bGalE and WbpP at the active site, even though, Ala209 in WbpP has the same orientation with Asn200 in bGalE, it does not show any "swing in or swing out". The movement of Asn in group 2 is significant for recognizing and binding *N*-acetylated or non-acetylated sugars. For better understanding for substrate specificity, a conceptual model originally proposed by Ishiyama *et al.* (10) was extended to bGalE. As a result, the configuration of six walls in bGalE are similar to that in hGalE. Therefore, bGalE was shown to allow *N*-acetylated or non-acetylated sugars to accommodate in substrate binding pocket with a concrete structural basis.

In this study, I determined the crystal structures of CBAP and bGalE involved in novel sugar metabolic pathways. According to the structural analysis, these enzymes have been shown to be specialized for their substrates through evolution. CBAP has a unique subsite +1 compared to GH94 emzymes, and bGalE has a broad substrate specificity competed with homologous GalE enzymes. This is because organisms in specialized environments sometimes face to live or survive. Therefore, they have to modify their enzymes to utilize substances in the environment. Studies of these two sugar metabolic enzymes also provide beneficial tools in the investigations of relationship between structure and function of enzyms for oligosaccharide synthesis, and for wide range of applications in biomass utilization and development of novel functional food for human health.

Acknowledgments

December, 2014

Young Woo Nam