FUNCTIONAL CHARACTERIZATION OF CARBOHYDRATE-UTILIZATION ENZYMES ENCODED BY THE *YCJ* GENE CLUSTER, A WELL CONSERVED ASSEMBLY OF GENES AMONG GRAM-NEGATIVE BACTERIA IN THE GUT

A Dissertation

by

KEYA MUKHERJEE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Frank Raushel
Committee Members,	Tatyana Igumenova
	Gregory Reinhart
	Michael Manson
Head of Department,	Gregory Reinhart

December 2018

Major Subject: Biochemistry

ABSTRACT

Scientists are slowly gaining a better understanding of the composition of the gut microbiota and the role it plays in maintaining metabolic homeostasis in the host. Metagenomic sequencing of various gut microbial species have revealed the correlation between the composition of the gut microbial population and a number of diseases. Although it is acknowledged that the gut microbiome encodes a number of enzymes that aid in the digestion of various dietary components, much is unknown about the underlying biochemical processes. The functional annotation of unknown proteins has not kept pace with the speed at which bacterial genomes from the gut are being sequenced. Sometimes the difficulty in determining functions of these uncharacterized proteins is compounded due to the misannotation of homologues present in various public databases. In this dissertation, efforts to elucidate the functions of enzymes in the *ycj* gene cluster in *Escherichia coli* K-12 are described. This gene cluster, which encodes for enzymes responsible for carbohydrate-utilization, is well conserved among a number of Gram-negative bacteria in the gut which implies that it plays an important role in degrading sugars.

The substrate profiles for six uncharacterized enzymes (YcjM, YcjQ, YcjR, YcjS, YcjT, and YcjU) of unknown function in *Escherichia coli* K-12 have been determined. YcjT was determined to be kojibiose phosphorylase which catalyzes the phosphorolysis of kojibiose into D-glucose and β-D-glucose-1-phosphate. This is the first such enzyme to be reported in *E. coli*. Kojibiose is a component of cell wall lipoteichoic acids in Gram-positive bacteria and is of interest as a potential low-calorie sweetener and prebiotic. YcjU was determined to be a β-phosphoglucomutase with activities against β-D-galactose-1-phosphate, β- D-mannose-1-phosphate, and β-D-allose-1-phosphate. These activities have not been reported previously in the

other known homologues of this enzyme. YcjM catalyzes the phosphorolysis of α-(1,2)-D-glucose-D-glycerate. YcjS was determined to be a NAD-dependent D-glucose dehydrogenase, while YcjQ was found to oxidize D-gulose, making it the first reported dehydrogenase that utilizes this rare hexose. The products of the dehydrogenases are proposed to be substrates for YcjR, a putative epimerase. The structure of selenomethionine-derivatized YcjR was solved and binding of a potential substrate was studied after docking it into the active site.

DEDICATION

To my grandfather, Sanat Kumar Banerjee, my first teacher. To my mother, Dr. Karabi Mukherjee, a pathologist who instilled in me the love for living systems by introducing me to the world of microbes through hours spent using the microscope in her laboratory. To my father, Dr. Ashok Mukherjee, who taught me that there is no substitute for hard work. To Karl, for the unwavering love and support throughout the journey.

ACKNOWLEDGEMENTS

This work would not have been possible without the support of many friends, mentors, and co-workers. I would like to thank my advisor, Dr. Frank Raushel for his valuable guidance and advice throughout my time as a graduate student in his lab and also, for his understanding during a very difficult phase in my personal life. I would also like to thank Dr. Michael Manson for taking time out, on numerous occasions, to look at my experimental results and for his many invaluable suggestions from a different point of view. I would like to extend my gratitude to my other committee members, Dr. Tatyana Igumenova and Dr. Gregory Reinhart for their support and encouragement throughout the course of my research.

A big thanks goes to all the current and former lab members who made life as a Ph.D. student possible. Most of all, I would like to acknowledge Dr. Jamison Huddleston, Dr. Andrew Bigley, and Dr. Dao Feng Xiang for patiently answering all my queries over the years.

To all my friends, a huge thank you! This would not be possible without your support and the strength you gave me to face another day when experiments did not work. A sincere thank you to Karl who has been my rock throughout these years. Lastly, to my family, thank you for always believing in me.

CONTRIBUTORS AND FUNDING SOURCES

The work presented in this dissertation was supported by a dissertation committee consisting of Dr. Frank Raushel (chair), Dr. Tatyana Igumenova, and Dr. Gregory Reinhart of the Department of Biochemistry and Biophysics; and Dr. Michael Manson from the Department of Biology.

All work was completed by the student, with the exception of the chemical synthesis of various substrates was carried out by Dr. Ventakesh Nemmara and Dr. Tamari Narindoshvili; and the crystal structure reported was solved by Dr. Mark Mabanglo using protein purified by the student.

The research work was supported by grants from the Robert A. Welch Foundation (A-840) and the National Institutes of Health (GM-122825).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SCHEMES

LIST OF TABLES

CHAPTER I

INTRODUCTION

The human gut is colonized by a complex community of microorganisms or microbiota (bacteria, archaea, fungi, viruses, and eukaryotic colonizers or parasites) that play an essential role in host nutrition and physiology.[1] The human body contains more microorganisms than human eukaryotic cells (~3.8 x 10^{13} to ~3 x 10^{13}).[2] Of the three kingdoms, bacteria make up the majority of the human gut microbiome. Within bacteria, seven different bacterial phyla predominate: Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria, Verrucomicrobia, Tenericutes, and Fusobacteria, of which species from Firmicutes and Bacteroidetes are the most prevelant.[3] Research has shown that some species in the microbiota are transferred from parent to offspring (vertical gene transmission).[4] In fact, the gut of fetuses, once thought to be sterile, is now believed to be colonized by microbes from the mother.[5] Microbial transmission across generations has been used as a marker to study population migration and human ancestry, utilizing bacteria such as *Helicobacter pylori*, among others.[6] An individual's microbiota is not a static community, it becomes more diverse over the course of a lifetime in response to environmental exposure and other factors such as diet, social interactions, illnesses, and antibiotic usage.[7] Disruption in the composition of the microbiota has been associated with a number of metabolic conditions that include obesity, diabetes, food-related allergies, inflammatory bowel disorder, colorectal cancer, and hepatic diseases.[8-13]

Diet plays an important role in the composition of the gut microbiota. Neurodegenerative diseases, for instance Alzheimer's, Parkinson's, Huntington's, or amyotrophic lateral sclerosis are often associated with chronic inflammation in the central nervous system (CNS). During

inflammation, human microglial cells and astrocytes initiate immune responses to repair damaged tissue. Recent studies have shown that the inflammatory responses by the CNS are regulated by aryl hydrocarbon receptor, AHR. AHR influences microglia and astrocytes by binding to metabolic-derivatives of glucobrassicin, a compound formed from glucose and tryptophan and is found in cruciferous vegetables such as broccoli and cabbage.[14] These metabolic derivatives of glucobrassicin are produced by the microbes in the gut.[15] This has opened up the discussion of how symptoms of diseases such as multiple sclerosis and other neurodegenerative diseases can be ameliorated by modifying diets so as to facilitate preferential growth of certain beneficial gut bacteria.[16]

The collective genome of the microbes in the gut and that of the host is referred to as the hologenome.[17] Over the course of millennia, *Homo* sapiens evolved to eat meat and vegetables when they were hunter-gatherers, as such the human genome encodes for relatively few enzymes that can break down complex carbohydrates from cereal crops which were only introduced into our diet relatively recently in our evolution. The microbiome, on the other hand, codes for metabolic traits that are missing in the host genome, especially that of carbohydrate and lipid metabolism. Apart from this, bacterial symbionts contribute a number of other functions that benefit the host: development of immune and digestive systems, angiogenesis, protection against pathogenic bacteria, synthesis of vitamin K, bioavailability of amino acids by anaerobic fermentation of peptides and proteins, biotransformation of drugs, as well as influencing behavioral traits by aiding in the development of brain.[18-22]

As antibiotic resistance among bacteria grows, a probiotics-based therapy to treating pathogenic infections is gaining traction among clinicians and scientists. This approach includes administering a bacterial species that produces antimicrobials which specifically target the

antibiotic-resistant pathogen, helping to reduce the severity of the infection.[23] Treatment for *Clostridium difficile* infections (CDIs) consists of administering last-line antibiotics such as vancomycin, metronidazole and fidaxomicin or in severe cases, a fecal microbiota transplant. Infections by antibiotic-resistant *C. difficile* are rising, with 35% of patients who undergo antibiotic treatment for a CDI, having recurrent CDIs. A recent study identified the gut bacteria *Lactobacillus reuteri,* as a potential antibiotic, due to the ability to convert glycerol to reuterin, a potent broad-spectrum antimicrobial compound that has been shown to inhibit *C. difficile* growth. *L. reuteri* was also found to be resistant to the antibiotics that are usually used to treat CDIs, making it a good candidate for a probiotics-based therapy against *C. difficile*. Glycerol supplementation in combination with *L. reuteri* was shown to be effective in inhibiting *C. difficile* growth both *in vitro* and *ex vivo.*[24]

Carbohydrate Metabolism by the Gut Microbiome

One of the major beneficial roles of the gut microbiota is aiding in the digestion of complex carbohydrates, which are usually derived from plants. These sugars include cellulose, hemi-cellulose, β-glucan, xylan, mannan, and pectin among others.[25] There are also a number of host-derived glucans like mucins, chondroitin sulfate, and hyaluronate which are further digested in the large intestine by the gut microbiota.[26] Microbial fermentation of these indigestible carbohydrates in the anaerobic large intestine releases short chain fatty acids (SCFA, such as acetate, butyrate, and propionate), organic acids (lactate, pyruvate, and succinate), and gases (H_2, H_2S, CO_2, and CH_4). The intestinal epithelial cells gain most of their energy from the further metabolism of these short chain fatty acids.[27]

Greater than 10^{12} hexasaccharides can be derived from various combinations of just D-hexoses, therefore the degradation of complex and diverse carbohydrates requires an equally vast

and varied set of enzymes.[28] The major polysaccharide-degrading bacteria in the gut belong to the Gram-negative genus *Bacteroides*. The enzymes required to degrade these complex sugars are arranged in discrete gene clusters called Polysaccharide Utilization Loci (PUL). PULs are distributed all across microbial communities in the ecosystem around us (**Figure 1.1**).

Complex polysaccharides are usually cleaved into oligosaccharides either on the outer cell surface by extracellular endoglycanases or transported into the periplasm where they are depolymerized by glycoside hydrolases. A typical PUL also, has outer membrane proteins responsible for the import of sugars into the periplasm. These transporters responsible for the uptake of the oligosaccharides into the cell are outer-membrane sugar-binding proteins, a TonB-dependent receptor/transporter, and sometimes a TonB-dependent transporter system.[29] The component monosaccharides are then transported into the cytoplasm where other enzymes metabolize them to generate energy through the breakdown of SCFA.

Transcriptional regulation within a specific PUL usually falls within three categories – starch utilization system (sus) sensor/regulator, extracytoplasmic function sigma (ECF-σ) factor anti-σ-factor pairs, or by hybrid two component systems (HTCS). Sometimes the expression of PULs can also be controlled by a LacI-type or OmpR-type system among others.[30] Not much is known about how the PUL gene expression is regulated by carbohydrates. Schwalm *et. al*, studied how the breakdown of a polysaccharide and the corresponding monosaccharide can influence regulation of a PUL in different *Bacteroides* spp. in the gut.[31] They found that arabinan, a polymer of arabinose, was a transcriptional activator of a PUL that degraded arabinoglycan, while arabinose played the role of a transcriptional repressor. Another study in 2016, by Cao and co-authors, identified a family of antisense small RNA (sRNA) that downregulates the expression of 15 different PULs from *B. fragilis* when enough of a preferred

carbon source, for instance glucose, is present in the environment. However, most of these PULs, under sRNA regulation, seem to be involved in metabolism of glycans that are host-derived rather than from plants.[32]

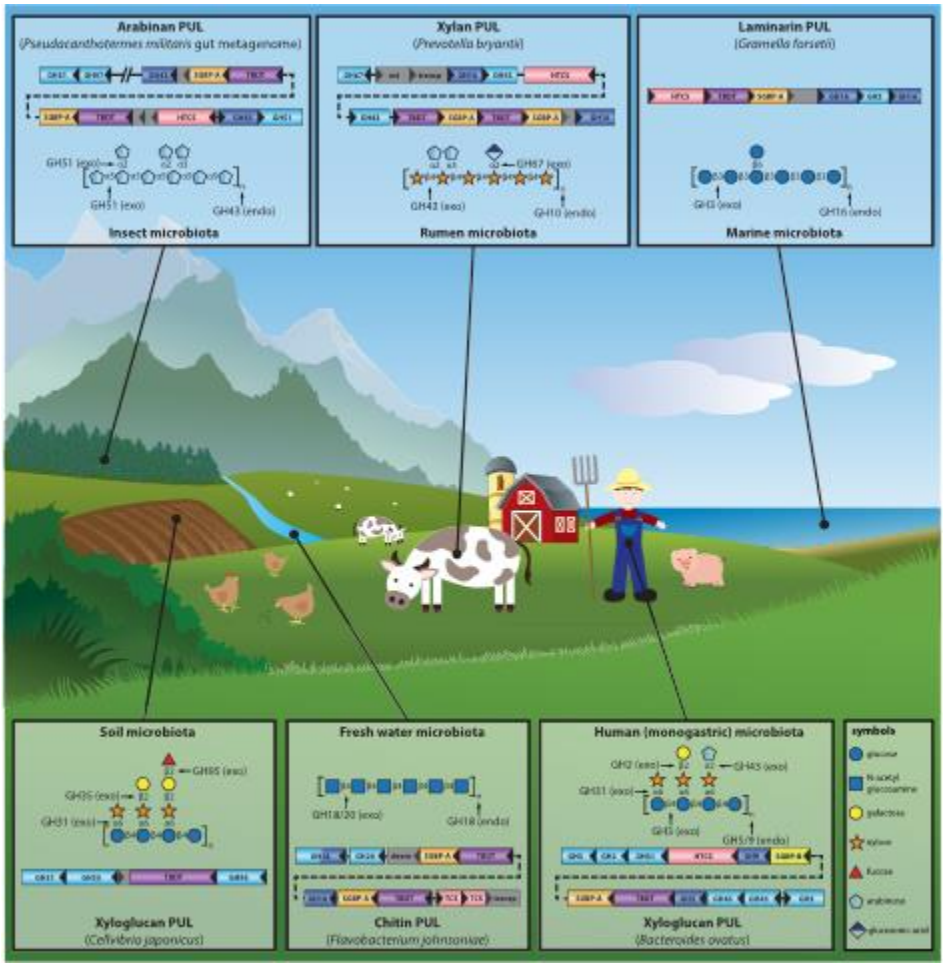


Figure 1.1: Distribution of polysaccharide utilization loci consisting of different glycoside hydrolase families found in nature. (Figure reproduced with permission from Grondin *et. al*, 2017, *J. Bacteriology* 199:e00860-16).

The Carbohydrate Active enZyme (CAZy) database (http://www.cazy.org/) contains a wide distribution of enzymes that are involved in carbohydrate degradation and synthesis.[33] The classification is based on amino acid sequence similarity of various enzymes that are typically involved in carbohydrate metabolism. Within a certain family, the structural folds, active site residues, and catalytic mechanism are highly conserved with few exceptions. There are 153 glycoside hydrolase (GH) families responsible for the hydrolytic cleavage of glycosidic bonds; 28 polysaccharide lyase families (PL) responsible for the non-hydrolytic cleavage of glycosidic bonds; 105 glucosyltransferase families that are involved in the formation of glycosidic bonds; 16 carbohydrate esterase (CE) families responsible for the hydrolytic cleavage of carbohydrate esters; and 15 families classified as having auxiliary activities (AA) that are mostly responsible for degrading lignin.

In 2014, Larsbrinsk *et. al* discovered a PUL in *Bacteroides ovatus* responsible for the degradation of xyloglucans, which are abundant in the cell walls of plants.[34] The structure of these glycans typically consist of a backbone of β-1,4 linked glucose units which are further linked with xylosyl moieties via β-1,6 linkages. These side chains can also have additional monosaccharides at their ends, typically galactose, fucose and/or arabinose. In 2011, the same authors identified the potential pathway responsible for xyloglucan degradation by observing the upregulation of certain genes when *B. ovatus* was grown in the presence of galactoxyloglucan.[35] The PUL consisted of sugar transport components as well as eight glycoside hydrolases (GH) from six different families. Utilizing reducing-sugar assays, mass spectrometry, and gene knockouts they were able to characterize the functions of the xyloglucan degrading enzymes when cloned into *Escherichia coli*. The identified enzymes belonged to GH_5A and GH_9A endo-xyloglucanases which cleave the main chain of the xyloglucans to form xyloglucan

oligosaccharides (units of four glucose molecules). The other enzymes were identified as belonging to GH_2A, a β-galactosidase, GH_3A and GH_3B, β-glucosidases, $GH_{31}A$, an α-xyloside, and $GH_{43}A$ and $GH_{43}B$, α-L-arabinofuranosidases. The complete degradation of the galactoxyloglucan and arabinogalactoxyloglucan are shown in **Figure 1.2**. Interesting, not all members of the *Bacteroides* genus have this xyloglucan degrading PUL.[34] Other studies have shown how some of the simple oligosaccharides formed by *B. ovatus* in this process diffuse into the extracellular environment where they are taken up by other *Bacteroides* spp. which do not have this PUL, for instance *B. adolescentis.*[36]

Due to the large genetic capacity in the hologenome, many novel enzyme families and catalytic activities are being discovered in gut microbes that are not found in their host or other organisms studied thus far. A study published by Ndeh *et. al*, elucidates how rhamnogalacturonate II (RG-II), a glycan found in red wine, is broken down by *Bacteroides thetaiotaomicron*, a well-studied gut microbe.[37] This polysaccharide is the most complex glycan known, with a structure made up of a combination of 13 different sugars connected by 21 different glycosidic bonds (**Figure 1.3**).

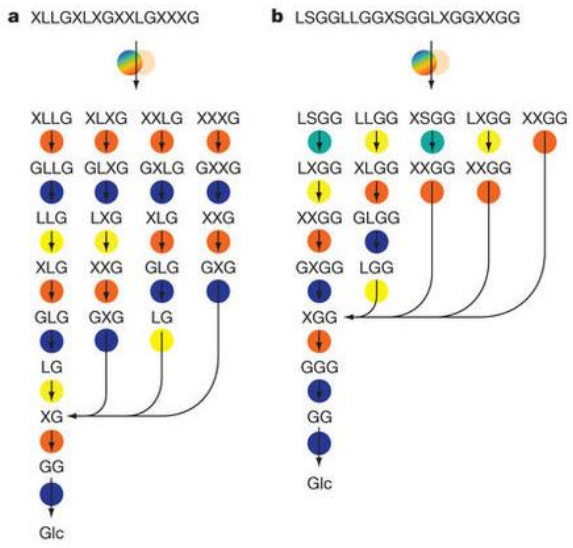


Figure 1.2: Galactoxyloglucan and arabinogalactoxyloglucan from *Bacteroides ovatus*. (a) galactoxyloglucan (b) arabinogalactoxyloglucan. Enzymes are represented in colored circles –rainbow BoGH5A (endo-xyloglucanase); tan BoGH9A (endo-xyloglucanase); orange BoGH31A α-xylosidase; turquoise BoGH43A and BoGH43B (α-L-arabinofuranosides); yellow BoGH2 (β-galactosidase), dark blue BoGH3A and BoGH3B (β-glucosidases). (Adapted with permission from Larsbrink *et. al,* 2014, *Nature* 506: 498-502).

B. thetaiotaomicron was identified by screening a number of *Bacteroides* spp. for growth on RG-II. A transcriptome analysis of *B. thetaiotaomicron* showed upregulation of a number of genes, arranged in three different polysaccharide utilization loci. These genes were cloned into and overexpressed in *E. coli*. The purified enzymes were able to completely degrade all the glyosidic bonds in the RG-II structure into its 13 component sugars with the exception of a single α1-3 bond between 2-*O*-methyl-D-xylose and L-fucose.

The RG-II "degradome" (degrading system) includes seven enzymes that belonged to new glycoside hydrolase (GH) families: BT1012, *endo*-apiosidase (GH140); BT0984, α-2-O-methyl-L-fucosidase (GH139); BT1002, α-L-fucosidase (GH141); BT0997, BT0997, α-galacturonidase (GH138); BT1020 (N-terminus), DHA-hydrolase (GH143); BT1020 (C-

terminus) and BT0996, two β-L-arbinofuranosidases (GH137 and GH142). BT1017 was found to belong to a new pectin methyl esterase family. Some of these enzymes were also found to hydrolyze two different glycosidic bonds in the glycan structure. BT1013 could hydrolyze an α1-5 bond between L-rhamnose and keto-deoxyoctulosonate (KDO) as well as an α2-3 bond between KDO and D-galacturonic acid. Similarly, BT1020 could hydrolyze a β1-5 glycosidic bond between L-arabinofuranose and 3-deoxy-D-lyxo-heptulosaric acid (DHA) and β2-3 bond between DHA and D-galacturonic acid. BT0996 was observed to hydrolyze a β2-1 bond between L-rhamnose and L-arabinofuranose as well as a β1-4 bond between L-fucose and D-glucuronic acid. The study also uncovered substrate specificities not previously reported in existing glycoside hydrolase families.[37]

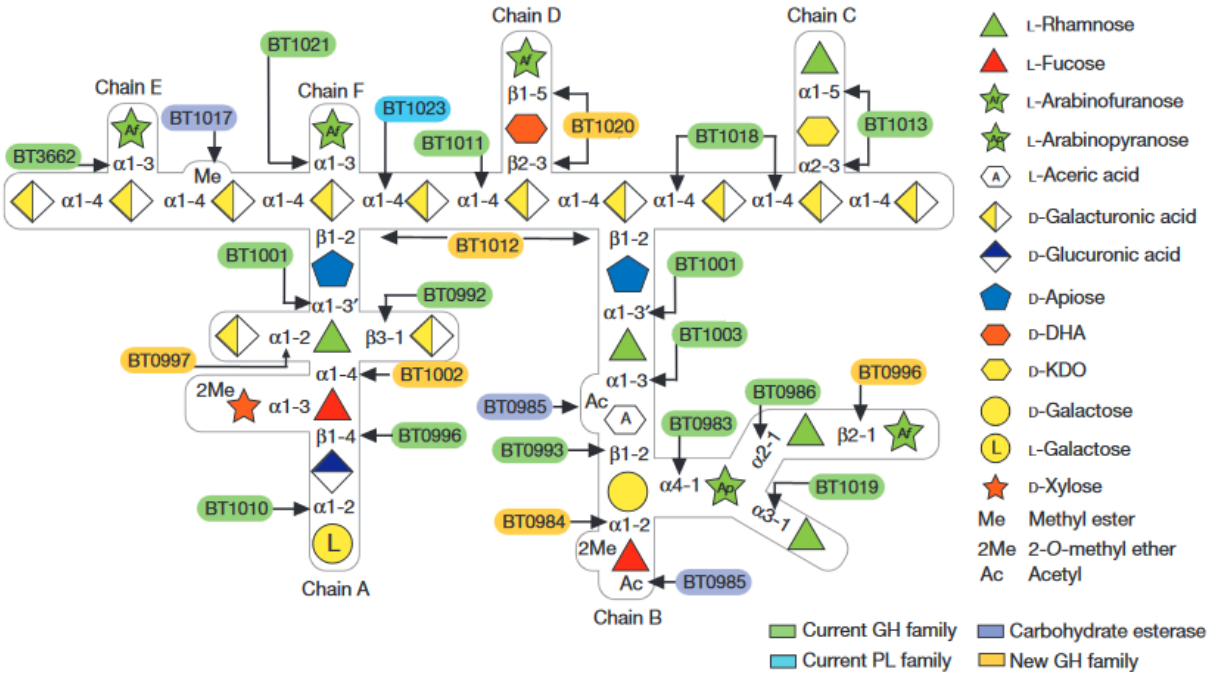


Figure 1.3: The "degradome" complex from *B. thetaiotaomicron* that breaks down rhamnogalacturonate II. (Figure reproduced with permission from Ndeh *et. al*, 2017, *Nature* 544: 65-70).

Munoz-Munoz *et. al*, found another novel enzyme, a polysaccharide lyase from *B. thetaiotaomicron*, that cleaves β1-4 bonds between L-rhamnose (Rha) and D-glucuronic acid (GlcA).[38] This particular disaccharide (Rha-α1,4-GlcA) is found in arabinogalactan proteins (AGPs), which are complex glycoproteins consisting of protein and carbohydrate components and are found in the plant cell walls. They play important roles in the growth and development of plants. AGPs mostly consist of a hydroxyproline-rich protein core with repeating arabinose and galactose monosaccharides that form discrete polysaccharide units attached to the core protein at multiple sites. L-rhamnose, D-mannose, D-xylose or L-fucose can be found on the side chains of the carbohydrate backbone.

Some gut microbes utilize the complex carbohydrate part of these glycoproteins as a carbon source. *B. thetaiotaomicron* and *B. ovatus* have previously been shown to grow on the AGPs.[35] Munoz-Munoz *et. al* identified genes that were upregulated in the presence of AGPs in *B. thetaiotaomicron* and found that the enzyme BT0263 had a role in depolymerizing the plant glycan. The sugar units that cap the ends of side chains in AGPs are sometimes L-rhamnose and there are very few reported glycoside hydrolases (α-L-rhamnosidases) that can cleave a α1-4 bond.[39] BT0263 is however able to non-hydrolytically cleave the glycosidic bond by acting on the D-glucuronic acid to release L-rhamnose from AGPs. The proposed mechanism of this enzyme is shown in **Figure 1.4**. BT0263 showed no sequence similarity with members of any known polysaccharide lyase families and was thus considered a founding member of a new PL family.

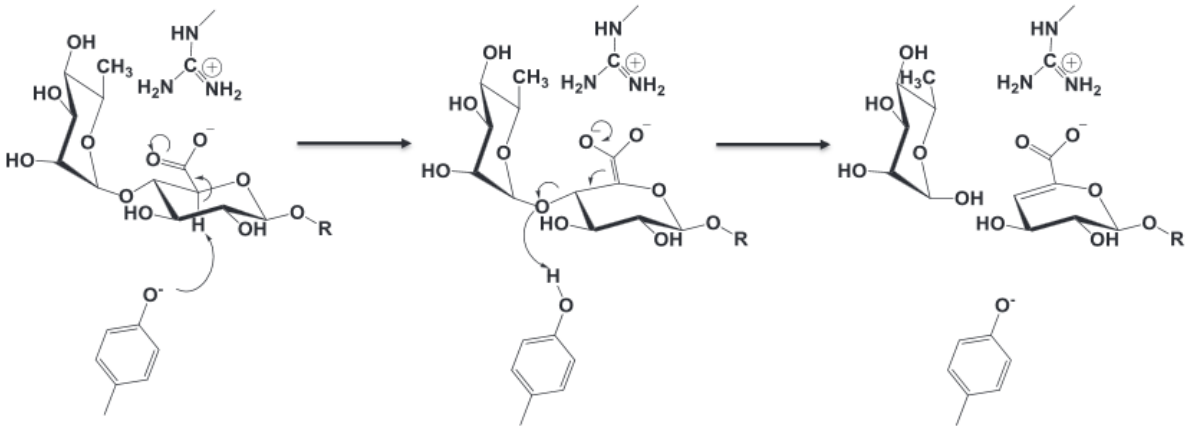
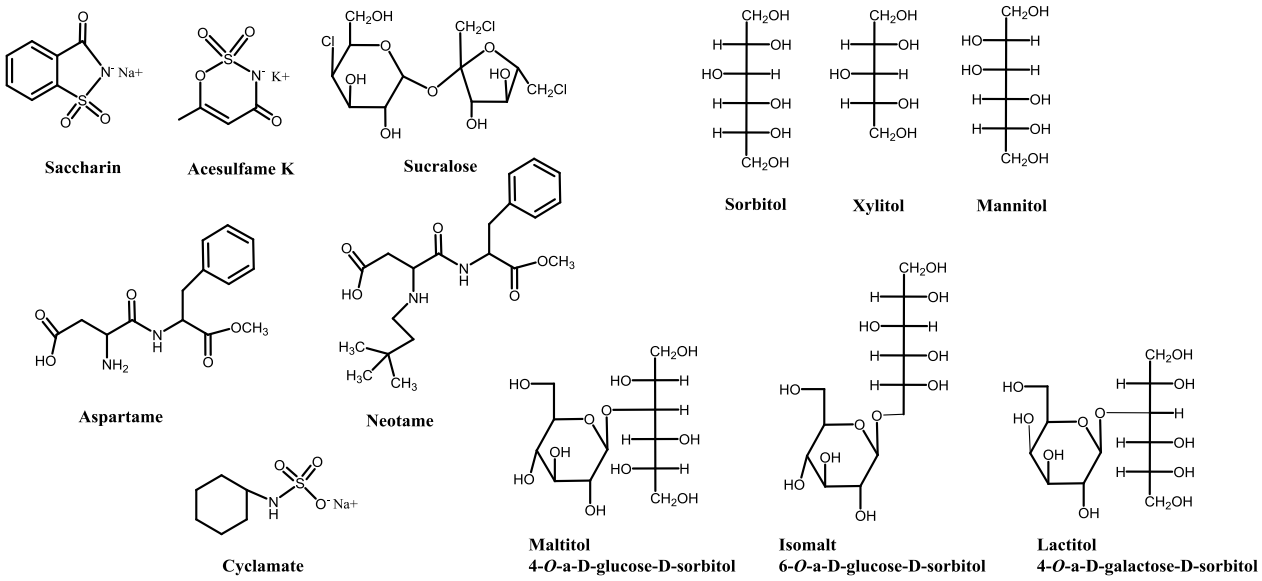


Figure 1.4: Proposed reaction mechanism of BT0263 from *B. thetaiotaomicron.* (Figure reproduced with permission from Munoz-Munoz *et. al*, 2017, *J. Biol. Chem.* 292: 13271-13283).

In the past forty years there has been an increase in the consumption of artificial or non-nutritive sweeteners.[40] The first artificial sweetener discovered was saccharin (Sweet 'n' Low®) in 1879 and was approved for use by the Food and Drug Administration (FDA) in 1970. Since then a number of other artificial sweeteners have also been approved by the FDA, for instance, acesulfame K, aspartame (Equal®), neotame, and sucralose (Splenda®) (**Scheme 1.1**).[41] There are some sugars that are also gaining popularity as sweeteners: D-psicose, D-tagatose, sugar alcohols like sorbitol, mannitol, xylitol (monosaccharide polyols) and isomalt, maltitol, lactitol (disaccharide polyols).[42] Their structures are also shown in **Scheme 1.1**. Due to low availability of these sugars in nature, they are known as 'rare' sugars. There have been a number of studies on the processes of manufacturing these 'rare' sugars either chemically or enzymatically.[43, 44] As these sweeteners are 30 – 13,000 times sweeter than natural sugar (sucrose), only a small amount of the sugar substitute (relative to natural sugar) is usually required to give food sweetness. This allows manufacturers to label such food products as 'reduced-calorie' or 'sugar-free'.[45] The

sweeteners are considered non-caloric as they are not broken down by our digestive enzymes and also because they are added in such small quantities to the food. However, once ingested they still encounter the microbial population in the gut and not much is known about how these sweeteners are metabolized by gut bacteria at the molecular level.

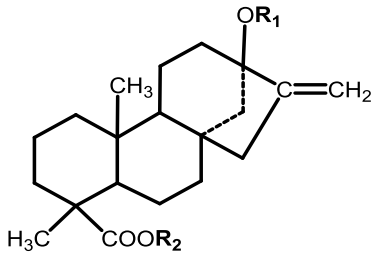


Scheme 1.1: Artificial sweeteners and 'rare' sugars. The six compounds on the left are artificial sweeteners while the six on the right are sweeteners that are also rare sugars.

Studies done on animal models, from as early as 1980, suggested that sweeteners such as saccharin altered the composition of microbial flora and impaired protein digestion.[46] Others studies have linked artificial sweeteners to various metabolic disorders, for instance glucose intolerance, which is a pre-diabetic condition, obesity, as well as cardiovascular diseases and cancer.[47-50]

In the gut, aspartame (L-aspartyl-L-phenylalanine methyl ester) is metabolized in a mode of digestion similar to that of proteins or protein hydrolysates. Aspartame is first broken down by the host's esterases and peptidases into aspartate, phenylalanine, and methanol.[51] Neotame, whose structure resembles aspartame, is digested by esterases to yield de-esterified neotame and methanol.[52] High doses of acesulfame K has been suggested to bind to DNA and cause damage, while a by-product of acesulfame K degradation is acetoacetamide, which can also be toxic at high doses.[52, 53] Cyclamate is metabolized by the gut bacteria to form cyclohexamine which was found to be carcinogenic.[54, 55] This led to cyclamate being banned as an artificial sweetener in the U.K. and the U.S. Disaccharide polyols are hydrolyzed in the small intestine by α-amylases (or β-galactosidases in case of lactitol) to from glucose / galactose and sorbitol. Monosaccharide polyols are then partially absorbed by the small intestine. The unabsorbed portion gets fermented by microbial enzymes to produce volatile fatty acids (VFA) in the large intestine. Part of this VFA is used by the gut microbiota for energy and the rest goes to the liver where they are metabolized further to form fat.[56]

The plant derived reduced-calorie sweetener, Stevia® or Stevioside, was shown to be completely degraded to steviol, its aglycone form, by the human gut microflora in ~10 hours. This degradation occurs through the formation of the intermediate steviolbioside (structures are shown in **Figure 1.5**). The steviol formed appeared to be resistant to further degradation by the gut bacteria.[57] However, nothing else is known about the enzymes involved in this degradative pathway.



Compound	R_1	R_2
Stevioside	-glucose (1-2)-glucose	-glucose
Steviolbioside	-glucose (1-2)-glucose	-H
Steviol	-H	-H

Figure 1.5: Structures of stevioside, stevobioside, and steviol. (Adapted with permission from Gardana et. *al*, 2003, *J. Agric. Food Chem.* 51: 6618 – 6622).

'Rare' sugars can also be defined as polysaccharides that are restricted to certain parts of the world. The ability to digest agarose was once only found among the marine bacteria. In 2010, Hehemann *et. al* were searching for enzymes that broke down porphyran, a constituent of algal polysaccharide, instead they found that the human gut bacteria *B. pleibus* carried an agarase enzyme.[58] Interestingly, the gene encoding this enzyme was only found in a particular gut bacterial species present in people from Japan. Consumption of seaweed has been a staple for centuries among the people living in coastal countries, such as Japan. They also found that there were two genes in the genome of *B. pleibus* that came from a marine bacterium, most likely *Zobellia galactanivorans*. It was concluded that a horizontal gene transfer event occurred after marine bacteria were ingested allowing *B. pleibus* to be able to adapt to a new carbon source.[58]

Recently, Pluvinage *et. al* reported that the ancestral genome of the previously non-agarolytic gut bacteria, *B. uniformis (Bu)* NP1 was found to contain an insertion of a 45 kbp genetic segment flanked by mobile elements.[59] The segment has 33 genes which appear to code for a number of glycoside hydrolases, capable of breaking down algal polysaccharides, outer membrane proteins, responsible for the import of glycans, as well as transcriptional regulators. The authors cloned some of these genes from the *Bu* NP1 genome and expressed them in *E. coli*. With the purified enzymes, they were able to demonstrate the depolymerization of agarose to form 3-*O*-β-D-galactose and 4-*O*-α-3,6-anhydro-L-galactose. Metagenomic analysis of individuals from four separate geographical areas and cultures, North America, Europe, Japan, and China, showed that the agarolytic enzymes were more prevalent in individuals from Japan and less common among Europeans and Chinese. Interestingly, it also present in high numbers among people from North America. The authors speculate that the reason for that is based on immigration and cultural aspects.

β-glucans are naturally occurring β-D-glucose polysaccharides found in plants such as oats and barley as well as in the cell walls of certain fungi, yeast and certain Gram-negative bacteria. β1-3 and β1-6 glycosidic linkages are more prevalent in the backbone of these structures, but β-1,2 glucans are quite rare. Shimizu *et. al*, report the discovery of a 2-β-D-glucooligosaccharide sophorohydrolase (BDI_3064) from *Parabacteroides distasonsis*, a gut bacteria.[60] This enzyme was shown to act on linear β-1,2 glucans to release sophorose (β-(1,2)-D-glucose-D-glucose) from longer chains with kinetic parameters of k_{cat} = 0.5 s^{-1}, K_m = 1 mM, and k_{cat}/ K_m 4.7 x 10^2 $M^{-1}s^{-1}$.

Unknown Function Discovery among Enzymes in Gut Bacteria

It is difficult to study the human gut microflora by *ex vivo* methods owing to a number of limitations: the selectivity of culture media, the need for stringent anaerobic growth conditions in some cases, and the difficulty in stimulating reciprocal interactions between different microbes or between the microbes and the host. Therefore, culture-independent methods based on ribosomal 16S sequencing, fluorescent in situ hybridization (FISH), competitive and quantative PCR, restriction fragment length polymorphism (RFLP), shotgun sequencing, and metagenome analyses are used.[61-64]

With the rapid advent in DNA sequencing technologies, the number of sequences deposited in Uniprot/TrEMBL as well as Uniprot/Swissprot have exploded. As of April 25th 2018, there were 557,275 sequences in the Uniprot/Swissprot database, which is manually curated based on literature, as well as computational analysis that has been reviewed by a curator, and over 114 million sequences in Uniprot/TrEMBL database (**Figure 1.6)** where functions are computationally annotated (http://www.uniprot.org/). Less than 1% of the total deposited protein sequences have their functions annotated based on direct biochemical evidence. In Uniprot/TrEMBL, annotation predictions are made computationally and are based on how close the amino acid sequence of an unknown protein is to the closest homolog whose function has been verified experimentally. However, these annotations are not evaluated by a curator and thus, can be misleading, which is why researchers need to devise ways to parse out the relevant information.

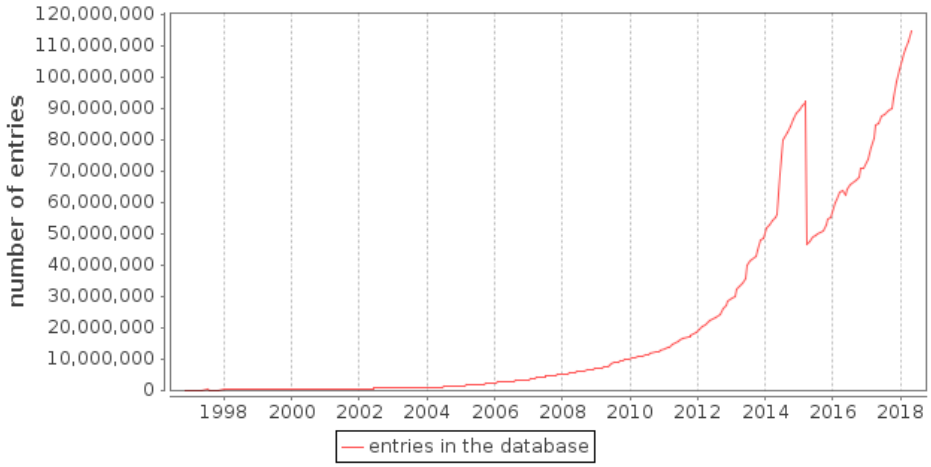


Figure 1.6: Increase in the number of sequences deposited in the Uniprot/TrEMBL database as a function of time. The steep drop in 2015 is due to the purging of redundant proteomes from the database using the Proteome Redundancy Detector which was released that year. (Figure taken from the UniProt/TrEMBL website: https://www.uniprot.org/statistics/TrEMBL).

The misannotation of enzymatic functions is a major issue in public databases. A study published by Schnoes *et. al*, attempted to compare the annotations of 37 different protein families from public, computationally annotated databases such as UniProtKB/TrEMBL and GenBank NR to those from the manually curated database UniProtKB/SwissProt.[65] They found variable levels (5% - 80%) of misannotations across the 37 families from the computationally annotated databases, while the misannotations were close to zero in the manually curated UniProtKB/SwissProt. There was no correlation between annotation errors and protein superfamilies with diverse functionalities. With misannotation also comes error propagation. The authors provided a visualization of how over a time period from 1995-2003, single proteins, which were incorrectly annotated, became connected via ‘edges’ (which denotes a pairwise relationship based on sequence similarity) or lines with other proteins which were also annotated

to the same function based on the first wrongly annotated protein from that family. This error propagation was found in the pathways from KEGG (Kyoto Encyclopedia of Genes and Genomes) which tends to pool functions from databases such as Pfam or InterPro, where annotations based on protein domains only can be misleading.[65]

β-D-Glucuronidases are prevalent in the human gut and are involved in the hydrolysis of β-glucuronides often found in the diet or in compounds produced in the liver after initial metabolism of xenobiotics.[66] Crohn's disease has been associated with a reduced β-D-glucuronidase activity while an increase in the activity of this enzyme is linked to colon cancer.[67, 68] β-D-glucuronidases are often misannotated as β-galactosidases due to their high sequence identity and similar 3D structural fold.

Pollet and co-authors, conducted a thorough bioinformatics search analysis to identify β-D-glucuronidases from the gut microbiome.[69] The Human Microbiome Project (HMP) led to the establishment of Human Microbiome Gene Indices (HMGI) and Human Microbiome Clustered Gene indices (HMGC) databases, which had more than 30 million translated protein sequences between them.[70] Both of these databases were initially screened to identify proteins that had ≥ 25% sequence identity and an *E*-value of <0.05 in pairwise alignments with any of the known β-glucuronidases from *E. coli*, *Clostridium perfringens*, *Streptococcus agalactiae*, and *Bacteroides fragilis*. A total of 1,646,315 and 267,594 proteins were identified from the HMGI and HMGC databases, respectively. β-D-Glucuronidases have a α/β hydrolase fold found in β-galactosidases, but there are a number of differences in the active site residues. Conserved residues in the active site of β-glucuronidases include two catalytic glutamates, an asparagine, and a glycine-asparagine-lysine (G-N-K) motif that interacts with the carboxylic acid moiety of the glucuronate. Multiple sequence alignments with confirmed β-glucuronidases showed that these

residues were highly conserved. The 1,646,315 and 267,594 proteins pooled from the two databases were further screened for these conserved residues. This second step identified 3,013 sequences from the HMGI database and 279 protein sequences from the HMGC database. Multiple sequence alignments within the two data sets gave further information on how the enzymes could be further categorized based on conservation found among residues that made up loop regions in the known structures. [69]

Often genes that encode for enzymes that work sequentially to either synthesize large or complicated biomolecules or degrade complex carbohydrates are clustered together such as polysaccharide utilization loci of the gut bacteria mentioned earlier or the biosynthetic gene clusters (BGC) involved in natural products biosynthesis. Automatic prediction methods can be used to identify these clusters from genomic data based on characteristic features of the gene clusters. A study published by Terrapon *et. al,* in 2015, looked at predicting PULs from 67 different *Bacteroides* genome sequences.[71] The authors designed their prediction software to look for certain class of proteins involved in sugar transport (SusCD-like) or for the presence of gene expression regulators which are present usually either at the beginning or the end of PULs. Using this methodology they were able to identify 3,745 PULs from the different *Bacteroides* genomes. Their designed algorithm now has a web interface, Polysaccharide Utilization Loci Database or PULDB (http://www.cazy.org/PULDB/). The annotations provided on this website are pooled from predictions in the CAZy database.[71]

There are a number of websites that can predict the presence of BGCs in a genome. These include antiSMASH (Antibiotics and Secondary Metabolite Analysis sHell), RODEO (Rapid ORF Description and Evaluation Online), and PRISM (Prediction Informatics for Secondary Metabolome).[72-74] The predictions on all of these databases are dependent on the

classification of proteins under certain protein families based on sequence similarity and therefore this methodology also suffers from problems that arise from misannotations.

Sequence similarity networks (SSN) and genome neighborhood networks (GNN) are powerful tools that an enzymologist can use to predict functions of uncharacterized proteins. SSN and GNN can be built on the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) website (https://efi.igb.illinois.edu/efi-est/index.php) found under the EFI-EST and EFI-GNT tabs respectively. The custom built networks can then be visualized using the Cytoscape program (http://www.cytoscape.org/).[75]

SSN displays a sequence-function relationship within a protein family on a global scale. These networks are based on pairwise relationships between proteins in all-by-all sequence comparisons. The display consists of 'nodes' which are individual protein sequences that are connected to each other by 'edges'. Each edge denotes a pairwise sequence identity between two proteins that share a sequence similarity greater than the *E*-value that has been set by the user. The primary idea behind building the network is to cluster isofunctional proteins together. These networks therefore, provide a broad overview of protein superfamilies that do similar chemistry but utilize different substrates.[76] In **Figure 1.7**, different substrate specificities within the same protein family are marked with different colors and separated into isofunctional clusters as the stringency is increased. If an experimentally verified function is found in the cluster that harbors an uncharacterized protein, it provides a good starting point to assemble a substrate library to screen the protein of interest with. However, if the cluster with the unknown protein does not have an ortholog with a verified function, lowering the stringency helps as it collapses two or more clusters together until a homolog with a verified function joins the cluster. The user can

then analyze the consensus protein sequences and substrate specificities of the other clusters to predict potential substrates for the uncharacterized protein.

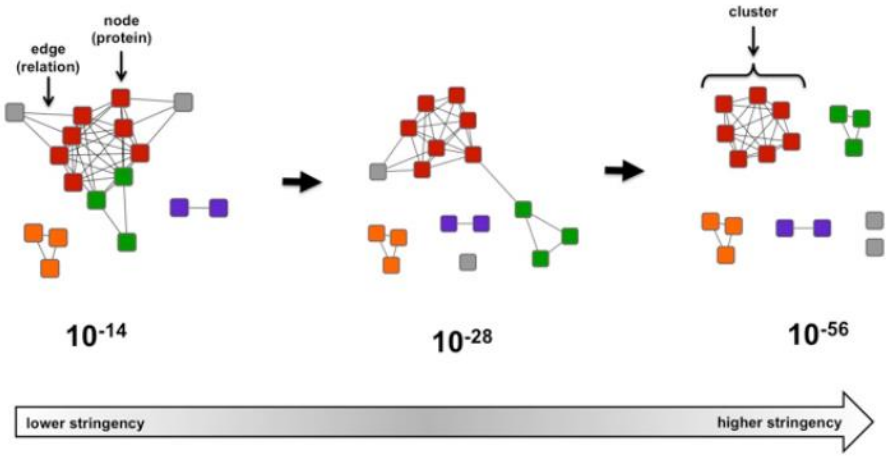


Figure 1.7: As stringency is increased orthologs cluster to form isofunctional groups or cluster of orthologous groups (COGs). Orthologs are colored the same color. (Figure taken from Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) website (https://efi.igb.illinois.edu/efi-est/index.php).

Genome Neighborhood Networks (GNN) allows the user to visualize the genomic context of an unknown protein. Through the EFI-EST Genome Neighborhood Tool (GNT), one can submit an entire cluster containing orthologous proteins for analysis. Since arrangements of operons or gene clusters are not conserved across phylogenetic species, a GNN of all the orthologs allows the user to identify enzymes that might be missing from the pathway of interest based on co-occurrence frequencies. A single protein sequence can also be used as an input to generate GNN for the gene. Integrating SNNs and GNNs provides a useful tool in discovering the function of an enzyme based on homologues as well as through a functional context by identifying roles of neighboring genes.

Vetting *et. al*, outlined an approach to identify potential metabolic pathways using solute binding proteins (SBPs).[77] These proteins are involved in the transport of a metabolite across cell membranes, which is the first step in a catabolic pathway. The SBPs can be broadly classified into three categories: ATP-binding cassette (ABC) transporters; TRipartite ATP-independent Periplasmic transporters (TRAPs); and Tripartite Tricarboxylate Transporters (TTTs). These proteins tend to cluster together with the rest of the genes that make up a metabolic pathway and are also regulated in the same fashion. The 8,240 protein sequences from the InterPro family (IPR018389) of TRAP SBPs were used to construct SSNs. From the SSN 304 protein sequences, which did not cluster with any known experimentally verified homologue, were cloned and expressed in *E. coli*. Out of this, 171 purified, soluble proteins were then screened against a library of 189 compounds of known SBP ligands using differential scanning fluorimetry (DSF). An increase in the melting temperature (T_m) of 5°C or more was found with 82 proteins binding to at least one substrate from the library.[77] Using the same methodology but with ABC transporters from *Mycobacterium smegmatis*, Huang *et. al*, identified three SBPs that bind to D-threitol, L-theritol, and erythritol.[78] Constructing a SSN for a dehydrogenase located next to the SBP for D-threitol followed by building a GNN allowed the authors to slowly piece together the entire catabolic pathway for the sugar.[78] This combinatorial approach is however, limited by the intergenic distances between different genes that make up a pathway and will most likely not work for catabolic pathways where the components are widely dispersed in the genome.[78]

An SSN constructed for an unknown protein from a large superfamily, which consists of a number of different functions, usually shows clusters with identified functions, as well as a number of other clusters with no functional annotations. Metagenomic data can be used to choose targets from these clusters of unknown functions. Levin *et. al*, published a study where

they outlined a "chemically guided" approach to function discovery.[79] They combined SSN analysis of the large glycyl radical enzyme superfamily with quantitative metagenomics data (**Figure 1.8**). From clusters with no functional annotations, segments of protein sequences that do not overlap with the consensus sequence of the cluster were denoted as "marker' sequences. These "marker" sequences were then compared to the translated short reads generated from metagenome sequencing of the human gut microbiota. The clusters with the highest "abundance" of metagenome reads were considered to have biological relevance and hence, good targets to study. They named their metagenomic data analysis tool ShortBRED (Short, Better Representative Extract Dataset).[79]

Transcriptomics can also be used to identify genes that are part of the same pathway but are not clustered together. Ghasempur *et. al*, studied a novel L-lyxonate degradation pathway in *Pseudomonas aeruginosa* PAO1.[80] The authors showed that the microbe could grow using L-lyxonate as a carbon source and through qRT-PCR they were able to identify five genes that were upregulated in the process. In-depth biochemical analysis revealed that three out of the five genes encoded for the following enzymes: L-lyxonate dehydratase (LyxD), 2-keto-3-deoxy-L-arabinonate dehydratase (L-KdlD), and semialdehyde dehydrogenase (KGSDH). These enzymes sequentially converted L-lyxonate to ketoglutarate, an intermediate in the citric acid cycle (**Figure 1.9**). However, the other two upregulated genes encoded for proteins that are in the amidohydrolase superfamily and the 4-hydroxy-threonine-4-phosphate dehydrogenase family and were not part of the L-lyxonate degradation pathway. Attempts to identify the substrates of these two proteins were unsuccessful. [80] There are therefore, limitations to this method as well but it provides the idea of what a potential pathway could resemble.

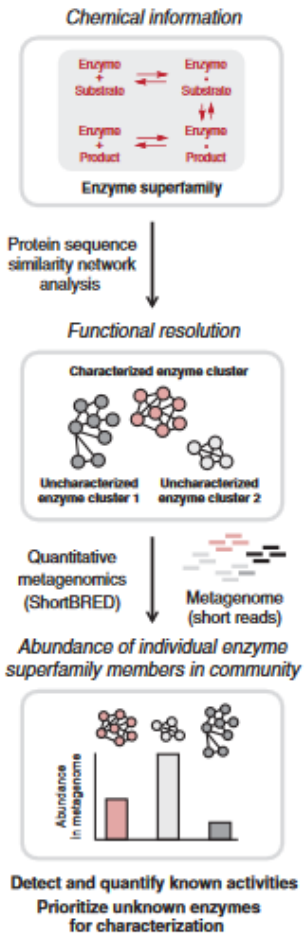


Figure 1.8: Chemically guided approach to find target proteins based on biological relevance in the gut microbiome. (Figure reproduced with permission from Levin *et. al* 2017, *Science* 355: eaai8386).

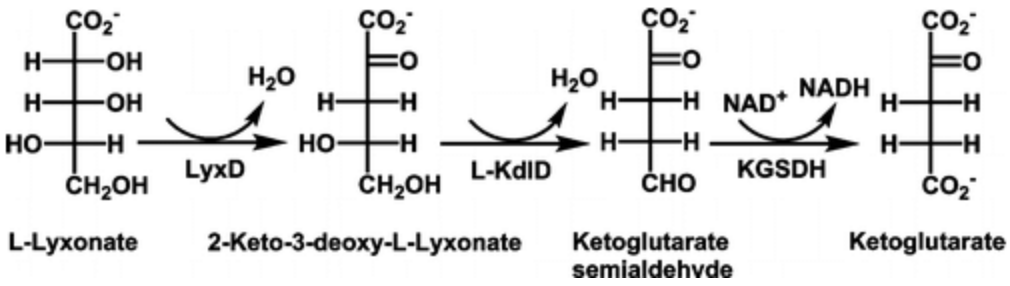


Figure 1.9: L-lyxonate catabolic pathway. (Figure reproduced with permission from Ghasempur *et. al* 2015, *Biochemistry* 53: 3357 – 3366).

Another way to discover potential functions of proteins in a pathway is to dock substrates into either solved x-ray structures of proteins with unknown functions or into homology models based on their homologues.[81, 82] Calhoun *et. al*, published a study where they found the functions of L-gulonate metabolic pathway from *Haemophilus influenza* Rd KW20 using an integrative pathway mapping method where enzyme structure-function relationship played a big role.[83] They identified a solute binding protein (SBP) which was found to bind to L-gulonate or D-mannonate by carrying out differential scanning fluorimetry (DSF) with 189 potential ligands. A GNN constructed for this protein found a dehydratase, two dehydrogenases, a kinase and an aldolase in the genomic neighborhood of the SBP. They then built homology models of these proteins based on the closest structural templates found in the Protein Data Bank where they docked 14,212 metabolites found in the KEGG database. The ligands utilized by the enzymes in a pathway resemble each other. Using this logic they assembled the highest scoring docked ligands from the homology models and a constraint was placed so as to end the pathway with a molecule that enters central metabolism. The closest annotated function to the putative dehydratase was found to be D-mannonate dehydratase.

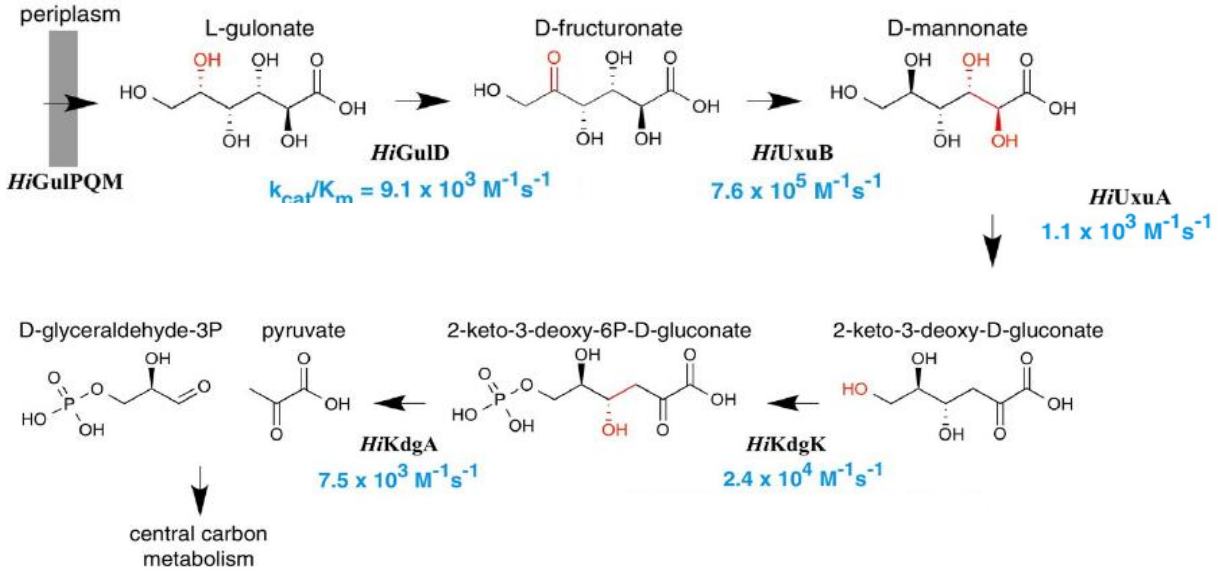


Figure 1.10: Catabolic pathway for L-gulonate in *H. influenza* Rd KW20. (Figure adapted from Calhoun *et. al*, 2018, *eLife* 7: e31097).

With this information they used their integrative pathway mapping algorithm to generate 154 predicted pathways for L-gulonate metabolism. The pathway with the highest score was the one that starting with L-gulonate as the ligand for the SBP and ended with glyceraldehyde-3-phosphate and pyruvate that entered the Entner-Dourdoroff pathway. These enzymes were then cloned, expressed, purified and their functions were experimentally verified (**Figure 1.10**)[83].

Even with the advent of better technologies to screen libraries of substrates, as well as advancements in computational methods used to assign functions to orphan enzymes, there is a vast gap in our function annotation capabilities. The goal of an enzymologist working in the field of function discovery of uncharacterized enzymes is to bridge this gap. In this dissertation, the chapters elucidate the work that has been carried out to characterize the enzymes encoded by the

ycj gene cluster in *Escherichia coli* K12 MG1655. This gene cluster, as well as another one from *Bacteroides thetaiotaomicron* (BT3581 to BT3586), were identified by searching for common sugar dehydrogenase COGs (0673 and 1063) in the genomes of *E. coli* K-12 and *B. thetaiotaomicron*. Both the gene clusters looked like they encoded for pathways that were novel. The gene cluster from *B. thetaiotaomicron* was cloned and expressed in *E. coli*. However, due to insolubility issues with most of the enzymes in the pathway, attention was directed towards the *ycj* gene cluster from *E. coli*. At the beginning of the project, the *ycj* cluster showed the following annotated functions: two sugar dehydrogenases (YcjS, YcjQ), putative sugar transport elements (YcjN, YcjO, YcjP, YcjV, and OmpG), predicted sugar isomerase/ epimerase (YcjR), putative sugar hydrolases/ phosphorylases (YcjM and YcjT), a predicted LacI-type transcriptional repressor (YcjW) and a potential phosphoglucomutase (YcjU). The first chapter in this dissertation addresses the functional characterization of YcjM, YcjT, and YcjU. YcjM was found to be an α-D-glucosyl-2-glycerate phosphorylase while YcjT and YcjU were characterized as kojibiose phosphorylase and a β-phosphoglucomutase respectively. In the second chapter, attempts to determine the substrate specificities and functional characterization of YcjQ, YcjR, and YcjS are discussed.

CHAPTER II

DISCOVERY OF A KOJIBIOSE PHOSPHORYLASE IN *ESCHERICHIA COLI* K-12*

In a typical adult there are more microbial prokaryotic cells (~3.8 x 10^{13}) than there are human eukaryotic cells (~3.0 x 10^{13}).[2] These bacteria reside predominantly in the gastrointestinal tract and are strongly influenced by factors such as diet, disease, and antibiotic usage.[84-86] The impact of these bacterial communities on human health has been investigated and a wide range of diseases such as obesity, diabetes, autoimmune disorders, inflammatory bowel disease, colon cancer, and transmission of retroviruses has been linked to dysbiotic microbiota.[10, 19, 87-94] Since the first bacterial genome was sequenced, there have been rapid technological advancements in the rate of bacterial genome sequencing.[95, 96] Metagenomic sequencing projects have led to the creation of a 'gene catalogue' of the most common microbial genes.[64] The 'microbiome' encode for a significant number of enzymes that are lacking in the host glycobiome. These microbial enzymes often confer metabolic capabilities to the host, such as the breakdown of indigestible polysaccharides.[97] However, the functional annotation of the microbial enzymes found in these organisms has not kept pace with the speed of various DNA sequencing projects.

A cluster of 12 genes (*ycjM-W* and *ompG*) has been identified in *Escherichia coli* K-12 MG1655, which may code for the expression of enzymes used for the metabolism of carbohydrates of unknown structure (**Figure 2.1**).

Most of the enzymes in this putative biochemical pathway bear little similarity to proteins of known metabolic functions. This gene cluster is also conserved in a number of other gut-dwelling, Gram-negative bacterial species, including *Salmonella enterica*, *Shigella dysenteriae*, *Erwinia tasmaniensis*, and *Citrobacter rodentium*, among others (*www.microbesonline.org*). This gene cluster consists of five predicted sugar transporters (YcjN, YcjO, YcjP, YcjV, and OmpG), two sugar dehydrogenases (YcjS and YcjQ), two putative polysaccharide hydrolases/phosphorylases (YcjM and YcjT), an epimerase/isomerase (YcjR), an enzyme that is a presumptive β-D-phosphoglucomutase (YcjU), and a predicted LacI-type repressor (YcjW).

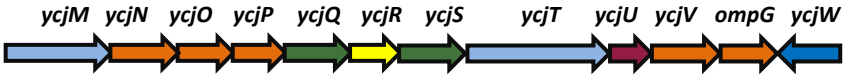


Figure 2.1: Organization of the *ycj* gene cluster in *E. coli*. These genes encode proteins that are predicted to include five sugar transporters (YcjN, YcjO, YcjP, YcjV, and OmpG), two NAD^+-dependent dehydrogenases (YcjS and YcjQ), two polysaccharide hydrolases/phosphorylases (YcjM and YcjT), an epimerase/isomerase (YcjR), a β-phosphoglucomutase (YcjU), and a LacI-type repressor (YcjW).

Here we describe our attempts to functionally characterize the substrate profiles for the two putative polysaccharide hydrolase/phosphorylase enzymes (YcjT and YcjM) and YcjU, the presumptive β-D-phosphoglucomutase. The sequence similarity network (SSN) for YcjM (cog0366) at an *E*-value cutoff of 1×10^{-50} (**Figure 2.2**) shows that the closest functionally characterized homologs include α-(1,2)-D-glucose-D-glycerate phosphorylase (GGP) from *Meiothermus silvanus*, sucrose phosphorylase from *Bifidobacterium adolescentis*, and amylosucrase from *Neisseria polysaccharea.*[76, 98-100] For YcjT (cog1554) the SSN at an E-value

cutoff of 10^{-150} (**Figure 2.3**) indicates that the closest functional homologues of this enzyme include maltose phosphorylase from *Lactobacillus brevis,* trehalose phosphorylase from *Bacillus stearothermophilus,* nigerose phosphorylase from *Lachnoclostridium phytofermentans*, α-(1,2)-D-glucose-glycerol phosphorylase from *Bacillus selenitireducens*, and kojibiose phosphorylase (KBP) from *Thermoanaerobacter brockii* and *Caldicellulosiruptor saccharolyticus*.[101-106] There are two experimentally verified enzymes in the SSN for YcjU (**Figure 2.4**): β-D-phosphoglucomutase (PgmB) from *Lactococcus lactis* and β-D-phosphoglucomutase (YvdM) from *Bacillus subtilis* (PDB id: 3NAS).[107] It appears likely that YcjU is a β-D-phosphoglucomutase (β-PGM). However, to the best of our knowledge, there are no currently known enzymes in *E. coli* K-12 that have been shown to catalyze the formation of β-D-glucose-1-P.

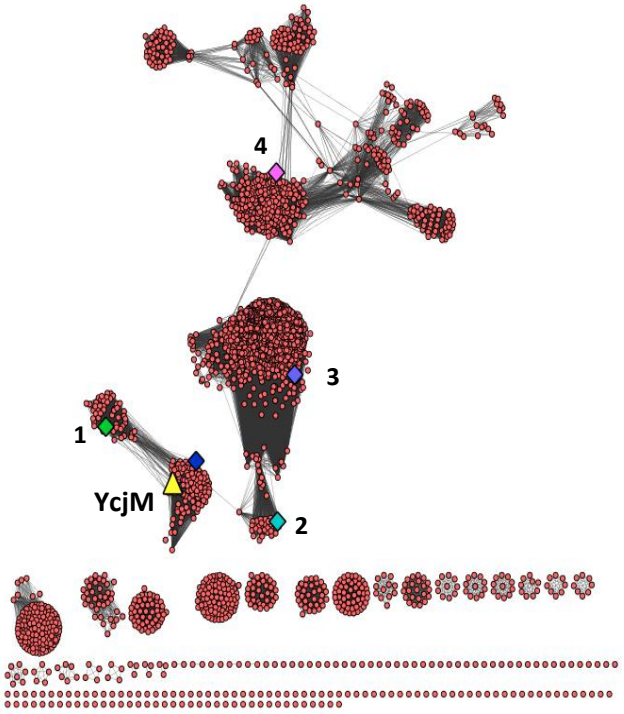


Figure 2.2: Sequence similarity network of cog0366 at an *E*-value cut-off of 1 x 10^{-50}. The network was created using Cytoscape (www.cytoscape.org). Each node represents a non-redundant protein sequence and each edge (or connecting line) represents a BLAST *E*-value between two sequences that is better than the arbitrary value of 1 x 10^{-50}. The lengths of the edges are not significant; in tight clusters the sequences are more closely related as compared to the clusters that contain fewer connections. YcjM is shown as a yellow triangle. Within the same cluster α-(1,2)-glucose-D-glycerate phosphorylase from *Meiothermus silvanus* is indicated with a diamond (dark blue). Diamonds in Groups 1, 2, 3, and 4 contain enzymes that have been experimentally verified. Group 1: sucrose phosphorylase from *Bifidobacterium adolescentis* (green); Group 2: amylosucrase from *Neisseria polysaccharea* (cyan); Group 3: oligo-1,6-glucosidase from *Bacillus subtilis* (blue); and Group 4: neopullulanase from *Geobacillus stearothermophilus* (pink).

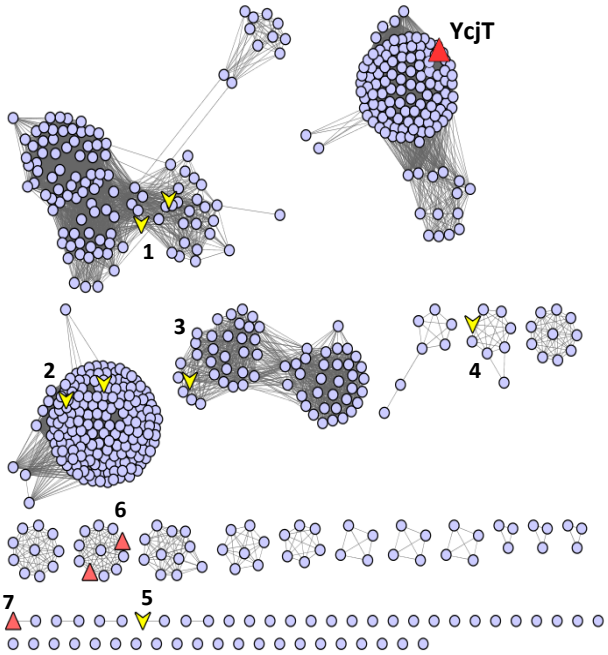


Figure 2.3: Sequence similarity network of cog1554 at an *E*-value cut-off of 1 x 10^{-150}. The network was generated using Cytoscape (www.cytoscape.org). Groups 1, 2, 3, 4, 5, 6, and 7 contain enzymes whose functions are experimentally verified (yellow arrowheads or red triangles). Group 1: α,α- trehalose phosphorylase (*Bacillus stearothermophilus*); Group 2: maltose phosphorylase (*Lactobacillus brevis*); Group 3: trehalose-6-phosphate phosphorylase (*Lactococcus lactis*); Group 4: nigerose phosphorylase (*Lachnoclostridium phytofermentans*); Group 5: α-(1,2)-D-glucose-glycerol phosphorylase (*Bacillus selenitireducens*); Group 6: kojibiose phosphorylase from *Thermoanaerobacter brockii* and *Caldicellulosiruptor saccharolyticus*; Group 7: kojibiose phosphorylase from *Pyrococcus spp*.

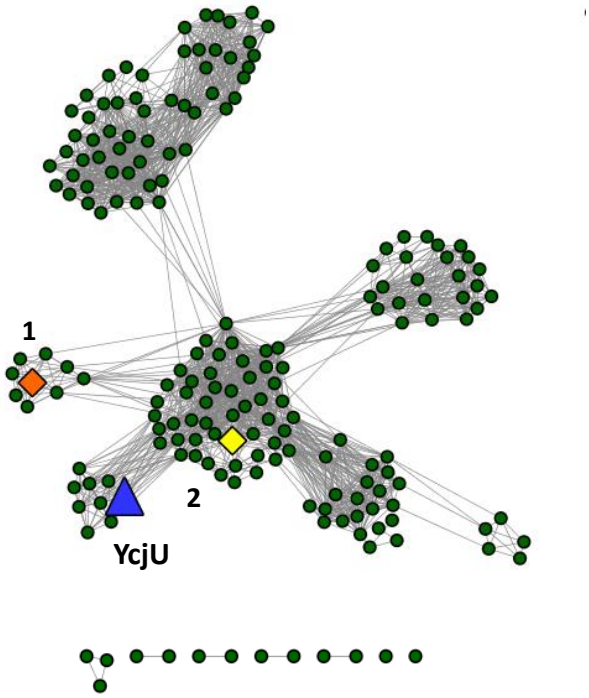


Figure 2.4: Sequence similarity network of InterPro family IPR010972 at an *E*-value cut-off of 1 x 10^{-85}. The network was created using Cytoscape (www.cytoscape.org). Diamonds (orange and yellow) represent two experimentally verified β-D-phosphoglucomutase enzymes: PgmB from *Lactococcus lactis* and YvdM from *Bacillus subtilis*. The blue triangle represents YcjU from *Escherichia coli* K-12 MG1655.

Materials and Methods

Materials. The restriction endonucleases, *pfu* turbo polymerase and T4 DNA ligase used in cloning of the genes *ycjM*, *ycjT*, and *ycjU* from *E. coli* were purchased from New England BioLabs. The PCR cleanup and gel extraction kits were obtained from Qiagen. Isopropyl-β-D-thiogalactopyranoside (IPTG) was acquired from Research Products International Corporation. The buffers, phenylmethylsulfonyl fluoride (PMSF), and the Phosphate Colorimetric Detection Kit were purchased from Sigma Aldrich. The 5-mL HisTrap columns, PD-10 desalting columns, and Vivaspin protein concentrators (molecular weight cut-off of 30 kDa) were bought from GE Healthcare. The M9 minimal media salts and thiamine were obtained from Sigma Aldrich. The carbohydrates used as potential substrates for YcjM, YcjT, and YcjU were obtained from Carbosynth, except for α-D-glucose-1,6-bisphosphate, D-glycerate, and α-D-glucose-1-P, which were obtained from Sigma Aldrich. α-(1,2)-D-Glucose-D-glycerate was synthesized by a modification of a previously published procedure.[108] The following bacterial strains were purchased from the Keio collection: parent strain for the knockouts BW25113 (CGSC 7636), Δ*ycjM* (CGSC 9167), Δ*ycjR* (CGSC 11236) and Δ*ycjT* (CGSC 9173) (https://cgsc2.biology.yale.edu/KeioList.php).

Synthesis of Alternate Substrates for YcjU. 2,3,4,5-Tetra-*O*-acetyl-β-mannose was synthesized from D-mannose as previously described.[109] Similarly, 2,3,4,5-tetra-*O*-acetyl-β-D-allose was synthesized from D-allose.[110] 2,3,4,5-Tetra-*O*-acetyl-β-galactose was synthesized from β-D-galactose pentaacetate by modification of a previously published procedure.[111] The phosphorylation of the acetylated sugars and subsequent isolation of β-D-mannose-1-P, β-D-allose-1-P and β-D-galactose-1-P was conducted according to Zhu *et. al.*[112]

Construction of Sequence Similarity Networks. Sequence similarity networks were created using Cytoscape (https://cytoscape.org). For YcjT, all members of cog1554 were downloaded from NCBI as a FASTA file. Protein sequences within this file that shared ≥ 90% identity with one another were consolidated. After this redundancy check was performed using the CD-HIT website, the custom FASTA file was uploaded to the Enzyme Function Initiative (EFI) enzyme similarity tool website (http://efi.igb.illinois.edu/efi-est/stepa.php).[113, 114] The tool-generated data set was analyzed under an organic layout using the Cytoscape software.[75] With YcjM, members of cog0366 were downloaded from the NCBI database to make the Cytoscape network after a redundancy cut-off of 85%. For YcjU (cog0637), members of the InterPro family IPR010972 were used to generate the network for this protein after a redundancy cut-off of 90%.

Cloning of ycjM from E. coli K-12 MG1655. The following set of primers was used to amplify the DNA corresponding to *ycjM* (gi|90111247; UniProt P76041) from *E. coli* K-12 MG1655 genomic DNA:

5′- AATACTGGATCCATGAAACAGAAAATT -3′

5′- ATAGTACTCGAGTTATTTAATCCACAT -3′

The amplified gene and the pET30a (+) vector were digested with *Bam*HI and *Xho*I restriction enzymes and ligated together to create the recombinant plasmid. After verification of the gene sequence, *E. coli* BL-21 (DE3) cells were transformed with the plasmid and plated on LB agar.

Expression and Purification of YcjM. A single colony was used to inoculate a 10-mL culture of LB, which was grown overnight at 37 °C. The overnight culture was used to initiate two 1-L cultures of LB supplemented with 50 µg/mL kanamycin in 2.8-L Fernbach flasks. The cultures were allowed to grow at 37 ^{o}C until the OD_{600} reached ~0.6, the temperature was

reduced to 25 °C, and the cultures allowed to grow overnight without the addition of IPTG. The cells were harvested by centrifugation at 11,000 g for 12 min and the cell pellet was stored at -80 °C until needed.

The frozen cell pellet was thawed and suspended in 20 mM HEPES/K^+, 0.5 M KCl, pH 8.0 at 4 °C. Prior to sonication, 0.1 mg/mL PMSF and 0.4 mg/mL DNase I were added to the cell suspension and stirred for 15 min. The cells were lysed with a Branson Sonifier 450 for six 5-minute intervals at 50% output. After sonication, the cell suspension was clarified by centrifugation at 12,000 g for 20 min. The supernatant fluid was passed through a 0.45 µm syringe filter, and loaded onto a 5-mL HisTrap column at room temperature, which had previously been equilibrated with binding buffer. The protein of interest was eluted from the column by applying a gradient of the elution buffer (20 mM HEPES/K^+, 0.25 M KCl, 0.5 M imidazole, pH 8.0). The isolated protein was >95% pure based on SDS-PAGE. The fractions were collected and concentrated using Vivaspin protein concentrators. The concentrated protein was then passed through a PD-10 desalting column to remove imidazole. Protein concentration was determined spectrophotometrically at 280 nm using an extinction coefficient of 94,700 M^{-1} cm^{-1} and a molecular weight of 69.6 kDa that includes the 5-kDa linker between the His-tag and the protein (https://web.expasy.org/protparam/). The protein was flash-frozen and stored at -80 °C until needed. Approximately 5 mg of the purified enzyme was obtained from 1.0 L of bacterial cell culture.

Determination of Phosphorylase and Hydrolase Activities of YcjM. A small library of α-D-glucose derivatives was tested as potential substrates with YcjM (**Scheme 2.S1**). The assays for the D-glucose-glucose disaccharides contained the following: 1.0 mM disaccharide, 10 mM inorganic phosphate, 5.0 mM $NADP^+$, 2.0 mM ATP, 5.0 mM $MgCl_2$, 1 U hexokinase, 3 U

glucose-6-phosphate dehydrogenase (G6PDH), and 2.0 μM YcjM in 50 mM HEPES/K^+, pH 8.0. The D-glucose-fructose, D-glucose-galactose, and D-glucose-glycerate disaccharides were assayed in the presence of 1.0 U α-PGM, 3 U G6PDH, 1.0 mM disaccharide, 10 mM inorganic phosphate, 4.0 mM $NADP^+$, 100 μM α-D-glucose-1,6-bisphosphate, and 2.0 μM YcjM in 50 mM HEPES/K^+, pH 8.0. The assays were repeated with 2.0 μM YcjU (β-PGM) instead of α-PGM.

The assays used to screen the reverse phosphorylase activity of YcjM were conducted at 30 °C using either 50 mM HEPES/K^+, pH 8.0, or 50 mM cacodylate, pH 6.5. The reactions contained 1.0 mM monosaccharide (**Schemes 2.S2**, **2.S3**, and **2.S4**), 1.0 mM α-D-glucose-1-P, and 1.0 μM YcjM. Aliquots were removed every hour and the formation of inorganic phosphate was determined using the phosphate detection kit.

The kinetic constants for the phosphorolysis of α-(1,2)-D-glucose-D-glycerate catalyzed by YcjM were measured with varying concentrations (0-350 μM) of α-(1,2)-D-glucose-D-glycerate, 10 mM inorganic phosphate, 0.25 μM YcjM, 4 U α-PGM (activated with 0.25 mM α-D-glucose-1,6-bisphosphate), 8 U G6PDH, 5 mM $MgCl_2$, and 5 mM $NADP^+$ in 50 mM cacodylate/K^+, pH 6.5. Reactions with varying concentrations of inorganic phosphate (0-15 mM) were conducted by keeping the concentration of α-(1,2)-D-glucose-D-glycerate constant at 150 μM. The assays used for measuring the kinetic constants of the reverse reaction contained α-D-glucose-1-phosphate (0-20 mM), 4.0 mM D-glycerate, and 2.5 nM YcjM in 50 cacodylate/K^+, pH 6.5. The reactions were incubated at 30 °C and aliquots were removed every 10 min over a period of 60 min. The concentration of inorganic phosphate was determined using the phosphate detection kit.

Cloning of ycjT from E. coli K-12 MG1655. The gene for *ycjT* (gi|16129277; UniProt P77154) was amplified from the genomic DNA of *E. coli* K-12 MG1655 using the following set of primers:

5′-AATGCGGATCCATGACCAGGCCAGTAACG- 3′

5′-ATAGTCTCGAGTCATTCATCCTCCTGATGTTTGG -3′

The restriction enzymes *Bam*HI and *Xho*I were used to digest the amplified gene, followed by ligation into a pET30a(+) vector, which was also digested with the same restriction enzymes and purified using the gel extraction kit. After verification of the amplified gene sequence, *E. coli* BL-21 (DE3) cells were transformed with the recombinant plasmid by electroporation.

Expression and Purification of YcjT. A single colony was grown overnight in 20 mL of LB containing 50 μg/mL kanamycin at 37 °C. The overnight culture was divided to inoculate four 1 L of LB media supplemented with 50 μg/mL kanamycin in four 2.8-L Fernbach flasks. The cell cultures were grown at 37 °C until an OD_{600} of ~0.6 was reached. Expression was induced with 0.5 mM IPTG and the cell cultures were allowed to grow overnight at a reduced temperature of 25 °C. The cells were harvested by centrifugation at 11,000 g for 12 min and the cell pellet was stored at -80 °C until needed.

The frozen cell pellet was re-suspended in 50 mM HEPES/K^+, 100 mM KCl, 10 mM imidazole, at pH 7.5. Prior to sonication, 0.1 mg/mL PMSF and 0.4 mg/mL of DNase I were added to the re-suspended cells, and the cell suspension was stirred for 20 min at 4 °C. The cells were lysed by sonication with a Branson Sonifier 450 using three 4-minute intervals at 50% output. After sonication, the cell debris was separated from the soluble proteins in the supernatant solution by centrifugation at 12,000 g for 20 min. The supernatant solution was loaded onto a 5-mL HisTrap column, which had previously been equilibrated with binding

buffer. YcjT was eluted from the column by applying a gradient of 50 mM HEPES/K^+, 100 mM KCl, and 500 mM imidazole at pH 7.5. The fractions were analyzed for purity using SDS-PAGE, and the fractions that were judged to be greater than ~90% pure were pooled and dialyzed against a solution of 50 mM HEPES/K^+, pH 7.5, and 100 mM KCl, to remove the imidazole. Protein concentration was determined spectrophotometrically at 280 nm using an extinction coefficient of 145,000 M^{-1} cm^{-1} and molecular weight of 90 kDa, including the 5-kDa linker between the His-tag and the protein (https://web.expasy.org/protparam/). Aliquots of YcjT were flash-frozen and stored at -80 °C until needed. Approximately 25 mg of YcjT was obtained from 1.0 L of bacterial cell culture.

Determination of Phosphorylase and Hydrolase Activity of YcjT. All kinetic assays were carried out at 30 °C in 50 mM HEPES/K^+, 100 mM KCl, at pH 7.5. For potential substrates consisting of a D-glucose-glucose disaccharide (**Scheme 2.S1**), 1.0 mM of the disaccharide, 10 mM inorganic phosphate, 5.0 mM $MgCl_2$, 2.0 mM ATP, 5.0 mM $NADP^+$, and 0.5 μM YcjT was mixed with 1 U hexokinase and 3 U G6PDH in a total volume of 250 μL. The increase in absorbance was monitored at 340 nm ($\Delta\varepsilon$ = 6220 M^{-1} cm^{-1}). To determine the potential catalytic activity with D-glucose-galactose or D-glucose-fructose disaccharides (**Scheme 2.S1**), the reaction mixtures contained 1.0 mM disaccharide, 10 mM inorganic phosphate, 5.0 mM $MgCl_2$, 2.0 mM ATP, 5.0 mM $NADP^+$, 1.0 μM YcjU, 3 U of G6PDH, and 0.5 μM YcjT in a total volume of 250 μL. The formation of NADPH was measured at 340 nm.

Assays designed to identify potential substrates for the reverse phosphorolysis reaction catalyzed by YcjT contained 1.0 mM monosaccharide (**Schemes 2.S2**, **2.S3**, and **2.S4**), 250 μM β-D-glucose-1-phosphate, 50 mM HEPES/K^+, 100 mM KCl, pH 7.5, and 0.5 μM YcjT. The reactions were incubated at room temperature, and aliquots were removed every 30 min for a

total of 4 h and the release of inorganic phosphate was measured using the phosphate detection kit.

The kinetic constants for the phosphorolysis of D-kojibiose by YcjT were conducted in 50 mM HEPES/K^+ and 100 mM KCl at pH 7.5 using an assay containing (0 - 15) mM kojibiose, 15 mM inorganic phosphate, 5.0 mM $MgCl_2$, 5.0 mM $NADP^+$, 1.0 μM YcjU, 3 U G6PDH, and 0.5 μM YcjT in a total volume of 250 μL. To determine the kinetic constants for phosphate, the assays were conducted with 5.0 mM kojibiose, 0.5 μM YcjT, and various concentrations (0 - 25 mM) of inorganic phosphate in the presence of 5.0 mM $MgCl_2$, 5.0 mM ATP, 5.0 mM $NADP^+$, 2 U hexokinase, 3 U G6PDH in 50 mM HEPES/K^+, 100 mM KCl, pH 7.5 at 30 °C. The initial velocity was calculated by monitoring the formation of NADPH at 340 nm.

To determine if YcjT hydrolyzed β-D-mannose-1-P, β-D-allose-1-P, and β-D-galactose-1-P, the reactions were conducted in the presence of 1.0 mM sugar phosphate and 1.0 μM YcjT in 50 mM HEPES/K+, pH 7.5, at 30 °C. The reactions were assayed for the liberation of phosphate every 30 min for 3 h with the phosphate detection kit.

Cloning of ycjU from E. coli K-12 MG1655. The gene *ycjU* (gi|16129278; UniProt P77366) was amplified from the genomic DNA of *E. coli* K-12 MG1655 using the following set of primers:

5′-AATGCGGATCCATGAAACTGCAAGGGGTAATTTTCGATCTGG-3′

5′-ATAGTCTCGAGCTATACGTTTTGCCAGAAGGCCGATAACC-3′

The amplified gene was purified using the PCR cleanup kit and digested with the restriction enzymes *Bam*HI and *Xho*I. The digested gene was ligated into a pET30a(+) vector, which was previously digested with the same two restriction enzymes and purified using the gel extraction

kit. After verification of the gene sequence, *E. coli* BL-21 (DE3) cells were transformed with the plasmid and plated on LB agar.

Expression and Purification of YcjU. A single colony was used to initiate a 10-mL culture of LB, which was grown overnight at 37 °C. The overnight culture was used to inoculate two 1.0-L cultures of LB supplemented with 50 μg/mL kanamycin and 1.0 mM $MgCl_2$ in 2.8-L Fernbach flasks and then allowed to grow at 30 °C until the OD_{600} reached ~0.6. Protein overexpression was induced with 0.5 mM IPTG and the cell cultures were grown overnight at a temperature of 25 °C. The cells were harvested by centrifugation at 11,000 g for 12 min and the cell pellet was stored at -80 °C until needed for the purification of YcjU.

The frozen cell pellet was re-suspended in a binding buffer containing 20 mM HEPES/K^+, 0.25 M KCl, and 15 mM imidazole at pH 7.9 at 4 °C. Prior to sonication, 0.1 mg/mL PMSF and 0.4 mg/mL of DNase I were added to the re-suspended cells and then stirred for 15 min. The cells were lysed with a Branson Sonifier 450 for three 4-min intervals at 50% output. After sonication, the cell debris was separated from the soluble proteins in the supernatant solution by centrifugation at 12,000 g for 20 min. The supernatant solution was passed through a 0.2 μm syringe filter, and loaded onto a 5-mL HisTrap column at room temperature, which had previously been equilibrated with binding buffer. YcjU was eluted from the column by applying a gradient of 20 mM HEPES/K^+, 0.25 M KCl, and 500 mM imidazole, pH 7.9. The column fractions were analyzed for purity using SDS-PAGE, and fractions with more than 90% purity were pooled and dialyzed against a solution of 20 mM HEPES/K^+ and 100 mM KCl at pH 8.0 to remove the imidazole. Protein concentration was determined spectrophotometrically at 280 nm using an extinction coefficient of 19,500 M^{-1} cm^{-1} and molecular weight of 28.5 kDa that includes the 5-kDa linker between the His-tag and the protein

(https://web.expasy.org/protparam/). The protein was flash-frozen and stored at -80 °C until needed. Approximately 35 mg of purified enzyme was obtained from 1.0 L of bacterial cell culture.

Determination of YcjU Activity. The kinetic assays were conducted at 30 °C in 50 mM HEPES/K^+, pH 8.0. The formation of D-glucose-6-phosphate from β-D-glucose-1-phosphate was determined by measuring the increase in absorbance at 340 nm from the G6PDH catalyzed reduction of $NADP^+$. The assays were carried out in the presence of varying amounts (0 - 10 mM) of β-D-glucose-1-P using 100 μM of α-D-glucose-1,6-bisphosphate as the activator, 0.1 μM YcjU, 2.0 units of G6PDH, 2.0 mM $NADP^+$, and 2.0 mM $MgCl_2$.

To determine the equilibrium constant of the YcjU catalyzed reaction, ^{31}P NMR (161 MHz, relaxation delay of 1.0 sec) was used to measure the relative concentration of D-glucose-6-P and β-D-glucose-1-P. The forward reaction contained 2.0 mM β-D-glucose-1-P, 125 μM α-D-glucose-1,6-bisphosphate, 1.0 mM $MgCl_2$, and 0.1 μM YcjU in 50 mM HEPES/K^+, 100 mM KCl, pH 8.0 with 10% D_2O. After incubation with the enzyme for 1.0 h, 2.0 mM EDTA was added to quench the reaction. The reverse reaction was conducted in a similar manner with the exception of adding 2.0 mM D-glucose-6-P instead of β-D-glucose-1-P to initiate the reaction.

^{31}P-NMR experiments were used to determine the activity of YcjU with β-D-mannose-1-P, β-D-allose-1-P, and β-D-galactose-1-P. The reaction mixtures contained ~~containing~~ 3.0 mM sugar phosphate, 100 μM α-D-glucose-1,6-bisphosphate as an activator, and 2.0 mM $MgCl_2$ in 50 mM HEPES/K^+, 100 mM KCl, pH 8.0. Concentration of YcjU used to initiate the reaction with β-D-mannose-1-P, β-D-galactose-1-P, and β-D-allose-1-P was 2.5 μM, 0.1 μM, and 0.05 μM respectively. ^{31}P NMR spectra were recorded before and after addition of YcjU. Scans were taken every 30 min after addition of YcjU for a total period of 2 h. Integration of the NMR

resonances for the substrate and product was used to monitor the change in substrate and product concentrations as a function of time. The equilibrium constants were determined using an initial concentration of either 2.0 mM of the sugar-1-P substrate or 2.0 mM of the corresponding sugar-6-P product. The reactions contained 100 μM α-D-glucose-1,6-bisphosphate, 1.0 mM $MgCl_2$, and 1.0 μM YcjU in 50 mM HEPES/K^+, 100 mM KCl, pH 8.0 and were allowed to proceed for 3-6 h to reach equilibrium. The reactions were quenched by the addition of 3.0 mM EDTA.

Determination of YcjU Activation by Mg^{2+}. To determine the effect of added Mg^{2+} on the catalytic activity of YcjU, assays were initiated with 100 μM β-D-glucose-1-P, 2.0 mM $NADP^+$, 50 μM α-D-glucose-1,6-bisphosphate, 0.2 μM YcjU, 3 U G6PDH, and varying amounts (0 - 10 mM) of Mg^{2+} in 50 mM HEPES/K^+, 100 mM KCl, pH 8.0 at 30 °C. The initial velocity associated with D-glucose-6-phosphate formation was determined by monitoring the increase in absorbance at 340 nm.

Determination of Kinetic Constants. The values of k_{cat} and k_{cat}/K_m were determined from the initial velocity data using Eqn. 1 where v is the velocity, E_t is the enzyme concentration, A is the substrate concentration, k_{cat} is the turnover number, and K_m is the Michaelis constant.

$$v/E_t = k_{cat}\, A/ (K_m + A) \qquad (1)$$

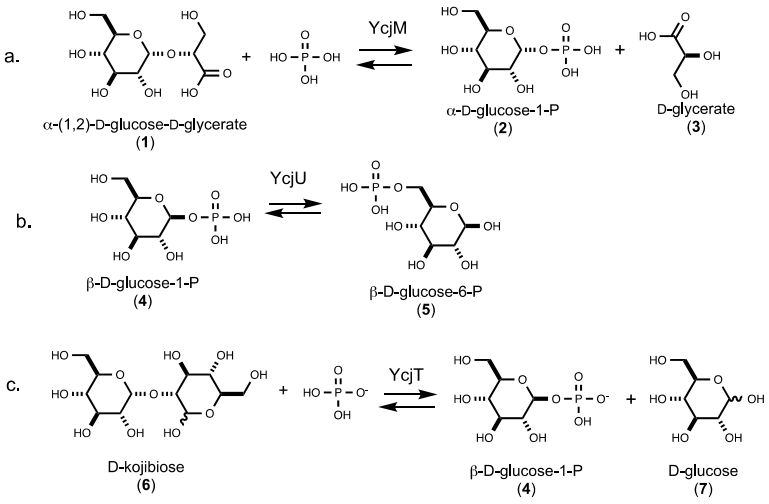
Growth of E. coli on Various Carbon Sources. M9 minimal media was prepared from commercially available M9 minimal media salts (Na_2HPO_4, KH_2PO_4, NH_4Cl, and NaCl). Minimal media plates were prepared from minimal media, 1.5% (w/v) of agar with 100 μM thiamine, and 0.25% w/v carbon source (D-glucose or D-kojibiose) added to the autoclaved media before plating. Single colonies of BW25113, Δ*ycjT*, and Δ*ycjM* were grown in 5 mL of LB media overnight at 37 °C. BL-21 (DE3) cells, transformed with the *ycjT* overexpression

plasmid, were grown in 5 mL of LB at 37 °C until an OD_{600} reached ~0.6. The culture was induced with 0.5 mM IPTG and allowed to grow overnight at 25°C. From each of the 5-mL cultures, 1 mL was removed, centrifuged at 16,000 g for 2 min, and the cell pellets re-suspended in minimal media. The cell suspensions were centrifuged at 16,000 g for 2 min followed by re-suspension in minimal media. This 'washing' was repeated five times to ensure that no LB media was carried over with the cells onto the minimal media plates. In order to determine the number of colonies associated with different dilution factors for the different bacterial strains, the cell suspensions were serially diluted (up to 10^{-6}) and each serial dilution was plated on minimal media plates with either 0.25% w/v glucose or kojibiose. Plates with no added carbon source were used as negative control plates. The plates with BL-21 (DE3) also had 0.25 mM IPTG added to maintain overexpression levels of YcjT. SDS-PAGE was conducted with the BL-21 (DE3) cell culture to confirm overexpression of YcjT before plating. Growth on the plates was monitored for a total of 4 days. To determine growth of different strains of bacteria on various α-D-glucose-glucose disaccharides, BW25113 and BL-21 (DE3) were plated on minimal media plates supplemented with 0.5% (w/v) of α,α-trehalose, kojibiose, sophorose, laminaribiose, nigerose, maltose, cellobiose, isomaltose, and gentiobiose as potential carbon sources. Minimal media plates with various concentrations (0.1% w/v, 0.25% w/v, and 0.5% w/v) of kojibiose, glucose, and sucrose were also made to test potential carbon sources for BW25113 strain.

Results

Catalytic Activity of YcjM. YcjM was purified to homogeneity and subsequently tested as a catalyst for the hydrolysis or phosphorolysis of the library of disaccharides presented in **Scheme 2.S1**. Based on the Cytoscape network constructed for YcjM (**Figure 2.2**), α-(1,2)-D-glucose-D-glycerate phosphorylase and sucrose phosphorylase are the two closest experimentally

verified enzymatic functions for enzymes in cog0366. Of the compounds tested, YcjM exhibited catalytic activity only with α-(1,2)-D-glucose-D-glycerate (**1**) in the presence of phosphate. The two products were determined to be α-D-glucose-1-P (**2**) and D-glycerate (**3**), and the overall reaction is summarized in **Scheme 2.1a**. YcjM was unable to catalyze the hydrolysis of any of the other compounds listed in **Scheme 2.S1**. The kinetic constants for the phosphorolysis of α-(1,2)-D-glucose-D-glycerate are as follows: k_{cat} = 2.1 s^{-1}, K_m = 69 μM, and k_{cat}/K_m = 3.1 x 10^4 M^{-1} s^{-1} at a fixed concentration of 20 mM phosphate at pH 6.5. The K_m for phosphate was determined to be 2.2 mM in the presence of 1.0 mM α-(1,2)-D-glucose-D-glycerate and varying (0 – 20 mM) concentrations of inorganic phosphate.



Scheme 2.1: Reactions catalyzed by YcjT, YcjU, and YcjM.

YcjM catalyzed the hydrolysis of α-D-glucose-1-P with a k_{cat} = 2.4 s^{-1}, K_m = 15.2 mM, and k_{cat}/K_m = 1.6 x 10^2 M^{-1} s^{-1}. Less than 1% of this activity was detected in the presence of α-D-galactose-1-P, α-D-mannose-1-P or β-D-glucose-1-P. A small library of monosaccharides (**Schemes 2.S2, 2 S3,** and **2.S4**) was tested as potential acceptor substrates with α-D-glucose-1-P in the back reaction. Of the compounds tested, catalytic activity was observed with D-glycerate (**3**) and D-erythronate. At a fixed concentration of 4.0 mM D-glycerate, the apparent kinetic parameters for α-D-glucose-1-P (**2**) during the formation of α-(1,2)-D-glucose-D-glycerate are k_{cat} = 350 s^{-1}, K_m = 9.4 mM, and k_{cat}/K_m = 3.7 x 10^4 M^{-1} s^{-1}. In the presence of 10 mM α-D-glucose-1-P, the K_m for D-glycerate was determined to be 4.4 mM. At a fixed concentration of 10 mM α-D-glucose-1-P and varying concentrations (0-20 mM) of D-erythronate, the kinetic parameters of the enzymatic reactions were k_{cat} = 68 s^{-1}, K_m = 8 mM, and k_{cat}/K_m = 8.5 x 10^3 M^{-1} s^{-1}. Less than 5% of this activity was observed with either L-glycerate, D-ribonate, D-allonate or glycerol. The kinetic constants for the reactions catalyzed by YcjM are summarized in **Table 2.1**.

Catalytic Activity of YcjU. Based on the available information from the sequence similarity network diagram (**Figure 2.3**) the proposed catalytic properties of YcjU as a β-phosphoglucomutase were tested with β-D-glucose-1-P (**4**) (**Scheme 2.1b**). To fully activate YcjU by phosphorylation, the enzyme was incubated with 100 μM α-D-glucose-1,6-bisphosphate for 15 min prior to the kinetic assays. In the β-phosphoglucomutase from *Lactococcus lactis*, the addition of α-D-glucose-1,6-bisphosphate to the enzyme facilitates the transfer of the phosphoryl group at C6 to Asp-8 in the active site. The phosphoryl group is subsequently transferred to the substrate β-D-glucose-1-P to form the β-D-glucose-1,6-bisphosphate intermediate.[115, 116] We assume that the activation of YcjU by α-D-glucose-1,6-bisphosphate operates in a similar fashion. The kinetic parameters for the interconversion of β-D-glucose-1-P to D-glucose-6-P are

as follows: k_{cat} = 21 s^{-1}, K_m = 18.5 μM, and k_{cat}/K_m = 1.1 x 10^6 M^{-1} s^{-1}. Substrate inhibition was observed at concentrations of β-D-glucose-1-P greater than ~0.2 mM. When the activity of YcjU was tested in the presence of varying concentrations of Mg^{2+} (0 – 10 mM) with 0.1 mM β-D-glucose-1-P, the activation constant (K_{act}) for Mg^{2+} was determined to be 135 ± 2 μM. In the absence of added Mg^{2+} the observed catalytic activity was less than 2% of the turnover observed in the presence of 2.0 mM Mg^{2+}. YcjU exhibited <1% of the catalytic activity (relative to β-D-glucose-1-P) with α-D-glucose-1-P. When YcjU was not phosphorylated with α-D-glucose-1,6-bisphosphate, less than 5% of the enzyme activity was observed. In the reverse direction, the kinetic parameters for the formation of β-D-glucose-1-P from D-glucose-6-P are as follows: k_{cat} = 0.5 s^{-1}, K_m = 1.19 mM, and k_{cat}/K_m = 4.2 x 10^2 M^{-1} s^{-1}. The kinetic constants for the reactions catalyzed by YcjU are summarized in **Table 2.1**. The equilibrium constant for the YcjU-catalyzed reaction was calculated to be 47 ± 2 from an integration of the ^{31}P NMR resonances for β-D-glucose-1-P and D-glucose-6-P when YcjU was added to either substrate and allowed to reach equilibrium.

The YcjU catalyzed reactions with β-D-allose-1-P, β-D-galactose-1-P, and β-D-mannose-1-P are shown in **Scheme 2.S5**. In the presence of saturating concentrations of β-D-allose-1-P, β-D-galactose-1-P, and β-D-mannose-1-P, apparent k_{cat} was calculated to be 10.5 ± 0.2 s^{-1}, 0.47 ± 0.09 s^{-1} and 0.09 ± 0.01 s^{-1} respectively. The equilibrium constants for the YcjU catalyzed reactions with β-D-allose-1-P, β-D-galactose-1-P and β-D-mannose-1-P were 24 ± 3, 29 ± 2 and 30 ± 3 respectively.

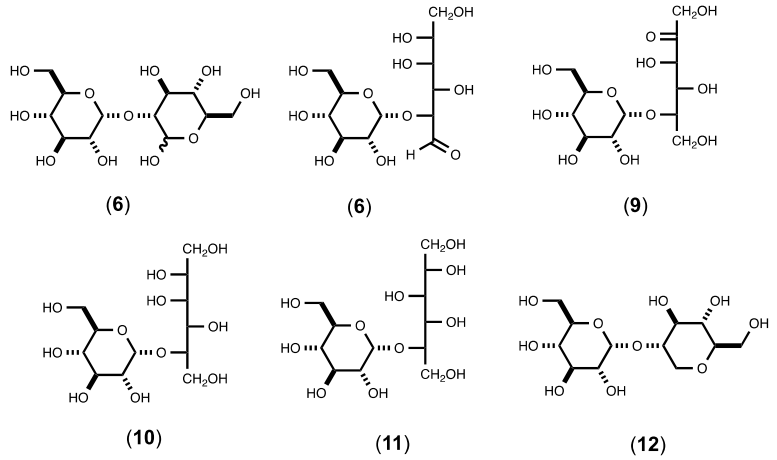
Catalytic Activity of YcjT. Based on the reported substrates for the closest functional homologs of YcjT from the sequence similarity network presented in **Figure 2.4**, a broad list of potential disaccharides was tested for catalytic activity. Among all of the disaccharides that were

tested for catalytic activity (**Scheme 2.S1**), YcjT showed phosphorylase activity only with D-kojibiose (**Scheme 2.1c**). The remaining disaccharides were less than 1% as active as D-kojibiose (**6**) at a concentration of 1.0 mM. At a fixed concentration of 5.0 mM phosphate, the apparent kinetic parameters for the phosphorolysis of D-kojibiose (0 -15 mM) are as follows: k_{cat} = 1.1 s^{-1}, K_m = 1.05 mM, and k_{cat}/K_m = 1.1 x 10^3 M^{-1} s^{-1}. In the presence of varying concentrations of phosphate (0-25 mM) at a fixed concentration of 5.0 mM kojibiose, the K_m for phosphate was determined to be 3.0 mM.

To determine the identity of the phosphorylated product from the phosphorolysis of D-kojibiose, the catalytic activity of YcjT was coupled to either α-phosphoglucomutase or YcjU (β-PGM) and G6PDH. Activity was observed only when the product of the reaction was coupled with YcjU, indicating that the other product (in addition to D-glucose) is β-D-glucose-1-P.

The catalytic properties of the reverse reaction were determined by incubating YcjT with β-D-glucose-1-P and various monosaccharides (**Schemes 2.S2, 2.S3,** and **2.S4**). Catalytic activity (measured by the formation of phosphate) was observed only with D-glucose, L-sorbose, D-sorbitol, L-iditol or 1,5-anhydro-D-glucitol, and the proposed structures of the corresponding disaccharides (compounds **6**, **9**-**12**) formed in these reactions are shown in **Scheme 2.2**. The rate constants for the hydrolysis of β-D-glucose-1-P in the absence of an acceptor substrate were determined to be k_{cat} = 0.014 s^{-1}, K_m = 2.01 mM and k_{cat}/K_m = 7.0 M^{-1} s^{-1}. At a fixed concentration of 8.0 mM β-D-glucose-1-phosphate, the apparent kinetic parameters for the formation of D-kojibiose (k_{cat} = 0.8 s^{-1}, K_m = 1.7 mM, and k_{cat}/K_m = 2.6 x 10^2 M^{-1} s^{-1}), α-(1,2)-D-glucose-anhydro-D-glucitol (**12**; k_{cat} = 1.05 s^{-1}, K_m = 6.1 mM, and k_{cat}/K_m = 1.7 x 10^2 M^{-1} s^{-1}), α-(1,5)-D-glucose-L-sorbose (**9**; k_{cat} = 1.2 s^{-1}, K_m = 7.9 mM, and k_{cat}/K_m = 1.5 x 10^2 M^{-1} s^{-1}), α-(1,2)-glucose-D-sorbitol (**10**; k_{cat} = 1.2 s^{-1}, K_m = 12 mM, and k_{cat}/K_m = 1.0 x 10^2 M^{-1} s^{-1}), and α-(1,2)-D-

glucose-L-iditol (**11**; k_{cat} = 1.3 s^{-1}, K_m = 14 mM, and k_{cat}/K_m = 92 M^{-1} s^{-1}) were determined. In an attempt to determine if YcjT can synthesize higher order oligosaccharides with β-D-glucose-1-P, various D-glucose-glucose disaccharides (**Scheme 2.S1**) were tested. None of the disaccharides showed any significant activity (< 1%) with YcjT when compared to the rate with D-glucose. YcjT did not show any phosphate release when β-D-mannose-1-P, β-D-galactose-1-P, and β-D-allose-1-P were incubated with the enzyme in the presence of D-glucose. The kinetic constants for the reactions catalyzed by YcjT are summarized in **Table 2.1**.



Scheme 2.2: Disaccharides formed by the reverse reaction with YcjT and β-D-glucose-1-P.

Growth on Kojibiose. To determine the ability of kojibiose to function as the sole carbon source for *E. coli*, overnight cultures of various cell lines (BW25113, BW25113 Δ*ycjT*, BW25113 Δ*ycjM*, and BL21 (DE3) with pET30a (+) containing *ycjT*) were washed, serial diluted, and then plated on minimal media with either D-kojibiose or D-glucose. Colony forming units were observed for all strains on glucose plates with 100 µL of the 10^{-6} dilution culture. No

colonies were observed on any of the plates in the absence of a carbon source (negative control). None of the bacterial strains grew on the plates supplemented with kojibiose. Of the various disaccharides tested as potential carbon sources for BW25113 and BL-21 (DE3), only α,α-trehalose and maltose supported growth. No growth was observed on plates with kojibiose, sophorose, laminaribiose, nigerose, cellobiose, isomaltose, gentiobiose or sucrose as the sole carbon source.

Discussion

There are numerous enzymes that are responsible for the construction and disassembly of complex carbohydrate structures. Given the vast diversity in carbohydrate structures and the specificity of enzymes for their respective substrates, characterization of these enzymes has been challenging.[33] Glycoside phosphorylases have garnered interest from the biotechnology industry because of their potential in synthesizing unusual glycosides that have commercial applications in the cosmetic industry, and as food additives.[117-120] Genomic analysis of bacteria that inhabit the human gut show that numerous enzymes are annotated as carbohydrate-active enzymes but a large fraction of these proteins have not been experimentally interrogated.[121] Here we describe our efforts to characterize three enzymes (YcjM, YcjT, and YcjU) from a putative carbohydrate metabolic pathway that is present in a variety of Gram-negative bacteria in the human gut, including various strains of *E. coli.*

Table 2.1: Apparent kinetic constants for YcjM, YcjT, and YcjU at 30 ^{o}C.

enzyme	varied substrate	fixed substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})
YcjM	α-(1,2)-D-glucose-D-glycerate	phosphate (20 mM)	2.1 ± 0.2	0.069 ± 0.002	(3.1 ± 0.1) x 10^4
YcjM	phosphate	α-(1,2)-D-glucose-D-glycerate (1 mM)	2.4 ± 0.4	2.2 ± 0.5	(1.1 ± 0.2) x 10^3
YcjM	α-D-glucose-1-P	D-glycerate (4 mM)	350 ± 2	9.4 ± 1.4	(3.7 ± 0.3) x 10^4
YcjM	D-glycerate	α-D-glucose-1-P (10 mM)	295 ± 4	4.4 ± 0.5	(6.7 ± 0.5) x 10^4
YcjM	α-D-glucose-1-P	H_2O	2.4 ± 0.3	15.2 ± 0.1	(1.6 ± 0.1) x 10^2
YcjM	D-erythronate	α-D-glucose-1-P (10 mM)	68 ± 1	8 ± 2	(8.5 ± 0.5) x 10^3
YcjT	kojibiose	phosphate (5 mM)	1.1 ± 0.1	1.05 ± 0.03	(1.1 ± 0.1) x 10^3
YcjT	phosphate	kojibiose (5 mM)	0.8 ± 0.1	3.0 ± 0.2	(2.7 ± 0.3) x 10^2
YcjT	β-D-glucose-1-P	H_2O	0.014 ± 0.001	2.01 ± 0.31	7.0 ± 0.1
YcjT	D-glucose	β-D-glucose-1-P (8 mM)	0.8 ± 0.1	1.7 ± 0.2	(2.6 ± 0.1) x 10^2
YcjT	L-sorbose	β-D-glucose-1-P (8 mM)	1.2 ± 0.5	7.9 ± 1.0	(1.5 ± 0.7) x 10^2
YcjT	L-iditol	β-D-glucose-1-P (8 mM)	1.3 ± 0.4	14.0 ± 1.0	92 ± 1
YcjT	D-sorbitol	β-D-glucose-1-P (8 mM)	1.2 ± 0.3	12.0 ± 1.0	(1.0 ± 0.1) x 10^2
YcjT	1,5-anhydro-D-glucitol	β-D-glucose-1-P (8 mM)	1.05 ± 0.03	6.1 ± 0.4	(1.7 ± 0.5) x 10^2
YcjU	β-D-glucose-1-P	---	21 ± 1	0.018 ± 0.001	(1.1 ± 0.1) x 10^6
YcjU	D-glucose-6-P	---	0.5 ± 0.1	1.19 ± 0.04	(4.2 ± 0.1) x 10^2

Functional Characterization of YcjM, a D-Glucose-D-Glycerate Phosphorylase. YcjM belongs to cog0366 and shares 46% identity with glucose-glycerate phosphorylase from *Meiothermus silvanus*, and 27% identity with sucrose phosphorylase from *Bifidobacterium adolescentis* (PDB: 1R7A) and amylosucrase from *Neisseria polysaccharea* (PDB: 1JGI).[98-100] All four enzymes belong to the retaining glycoside hydrolase family 13 (GH13), also known as the α-amylase family, according to the CAZy database.[122] Members of this enzyme family catalyze reactions through a double displacement mechanism that proceeds though a covalent glucose-enzyme intermediate and involves a nucleophilic aspartate residue and a glutamate residue which acts as a proton donor.[123]

A homology model of YcjM was built using Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) with sucrose phosphorylase (PDB id: 2GDU) as the template. A structural comparison between sucrose phosphorylase (PDB id: 2GDU), amylosucrase (PDB id: 1JGI) and the YcjM model was assessed using Pymol. Homology modeling of D-glucose-D-glycerate phosphorylase from *M. silvanus,* also built with sucrose phosphorylase as the structural template, followed by docking of α-(1,2)-D-glucose-D-glycerate into the active site showed that Asn275 and Glu383 appeared to be positioned to interact productively with D-glycerate.[98)] Both of the homologous residues in YcjM (Asn273 and Glu382) are conserved (**Figure 2.S1**). In the YcjM model, the docked α-(1,2)-D-glucose-D-glycerate superimposes well with the bound sucrose when overlaid with the sucrose phosphorylase structure (PDB id: 2GDU) from *B. adolescentis*. The catalytic residues, Asp229 and Glu271, are 4.4 Å and 4.3 Å from the anomeric carbon, respectively. Based on this model, Arg443 can potentially interact with the carboxylate group of α-(1,2)-D-glucose-D-glycerate.

There are two other arginine residues in the vicinity (Arg171 and Arg378) that can potentially move in closer to the bound substrate to provide additional electrostatic interactions. α-(1,2)-D-Glucose-D-glycerate, once considered a rare osmoprotectant, is now known to accumulate under conditions of salt stress and nitrogen scarcity in enterobacteria, marine cyanobacteria, and halophilic methanogens.[124-127] Two species of *Streptococcus* have been reported to accumulate D-glucose-D-glycerate, along with trehalose.[128, 129] *Streptomyces caelestis* is also known to excrete D-glucose-D-glycerate into the medium.[128] Glucose-D-glycerate was first identified in the 1960s as part of the reducing ends of the methylglucose lipopolysaccharide (MGLP) in *Mycobacterium phlei.*[130]

There are two common pathways for the biosynthesis of D-glucose-D-glycerate. The first one involves a two-step biosynthetic route using glucose-3-phosphoglycerate synthase and glucose-3-phosphoglycerate phosphatase.[131] The physiological role of MGLP in the formation of mycolic acids has been suggested and since the deletion of glucose-3-phosphoglycerate synthase was shown to affect the biosynthesis of MGLP, it is considered as a potential anti-tuberculosis drug target.[132, 133] The second route is catalyzed by glucose-glycerate synthase which converts D-glycerate and ADP-glucose to D-glucose-D-glycerate and ADP.[134]

Functional Characterization of YcjT, a Kojibiose Phosphorylase. Kojibiose phosphorylase belongs to the inverting glycoside hydrolase family 65 (GH65), which also includes trehalose phosphorylase and maltose phosphorylase.[135] The product of this reaction is β-D-glucose-1-P, which is the apparent substrate for the next enzyme in the pathway, YcjU. YcjT, which belongs to cog1554, shares 30% sequence identity with orthologs from *Thermoanaerobacter brockii*, *Caldicellulosiruptor saccharolyticus*, and an archaea, *Pyrococcus* sp.[105, 106, 136] Kojibiose phosphorylases have previously been shown to transfer β-glucose-1-P to

alternative acceptor substrates, apart from glucose. Our studies showed that among the monosaccharides tested in presence of β-D-glucose-1-P, only 1,5-anhydro-D-glucitol, L-sorbose, D-sorbitol or L-iditol can function as alternative substrates in the back reaction. The formation of a disaccharide with L-sorbose, D-sorbitol, glycerol, and *myo*-inositol has been reported previously with kojibiose phosphorylase from *T. brockii.*[105, 137] As the stereochemistry of the hydroxyl groups at C5, C4, and C3 of L-sorbose correspond to the hydroxyl groups at C2, C3, and C4 of D-glucose, the glycoside bond of the D-glucose-L-sorbose disaccharide is 1,5 rather than 1,2. YcjT did not exhibit significant catalytic activity for the formation of larger oligosaccharides when incubated with β-D-glucose-1-P and kojibiose. On the contrary, the kojibiose phosphorylase from *C. saccharolyticus* prefers kojitriose as an acceptor relative to kojibiose.[106]

Based on the multiple sequence alignment (**Figure 2.S2**) between YcjT, and kojibiose phosphorylases from *C. saccharolyticus* (PDB id: 3WIQ), *T. brockii*, and *P. horikoshii*, the residues involved in the binding of kojibiose (Tyr327, Asp334, Trp333, Trp381, Glu382, Lys573, Gln574, Ser608, and Ser609) are conserved with the exception of the presence of serine (Ser409) in YcjT as compared to threonine in kojibiose phosphorylases from the other three organisms.[138] The proposed general acid residue, Glu473 is also conserved in YcjT.

Attempts to grow BW25113, and deletions strains of Δ*ycjT* and Δ*ycjM* on kojibiose were unsuccessful. BL-21 (DE3) with a YcjT overexpression plasmid also failed to grow on kojibiose in the presence of 0.5 mM IPTG. This result suggests that kojibiose is probably not transported inside the cells. It appears that the *ycj* pathway, in addition to the putative ABC-type transporter complex (YcjNOP), has OmpG, which is a large, monomeric porin protein (PDB id: 2IWW) and is believed to be involved in the unspecific uptake of oligosaccharides.[139] All attempts to express

this protein under laboratory conditions have been unsuccessful.[140] Our experiments indicate that kojibiose is most likely not the inducer for the *ycj* pathway.

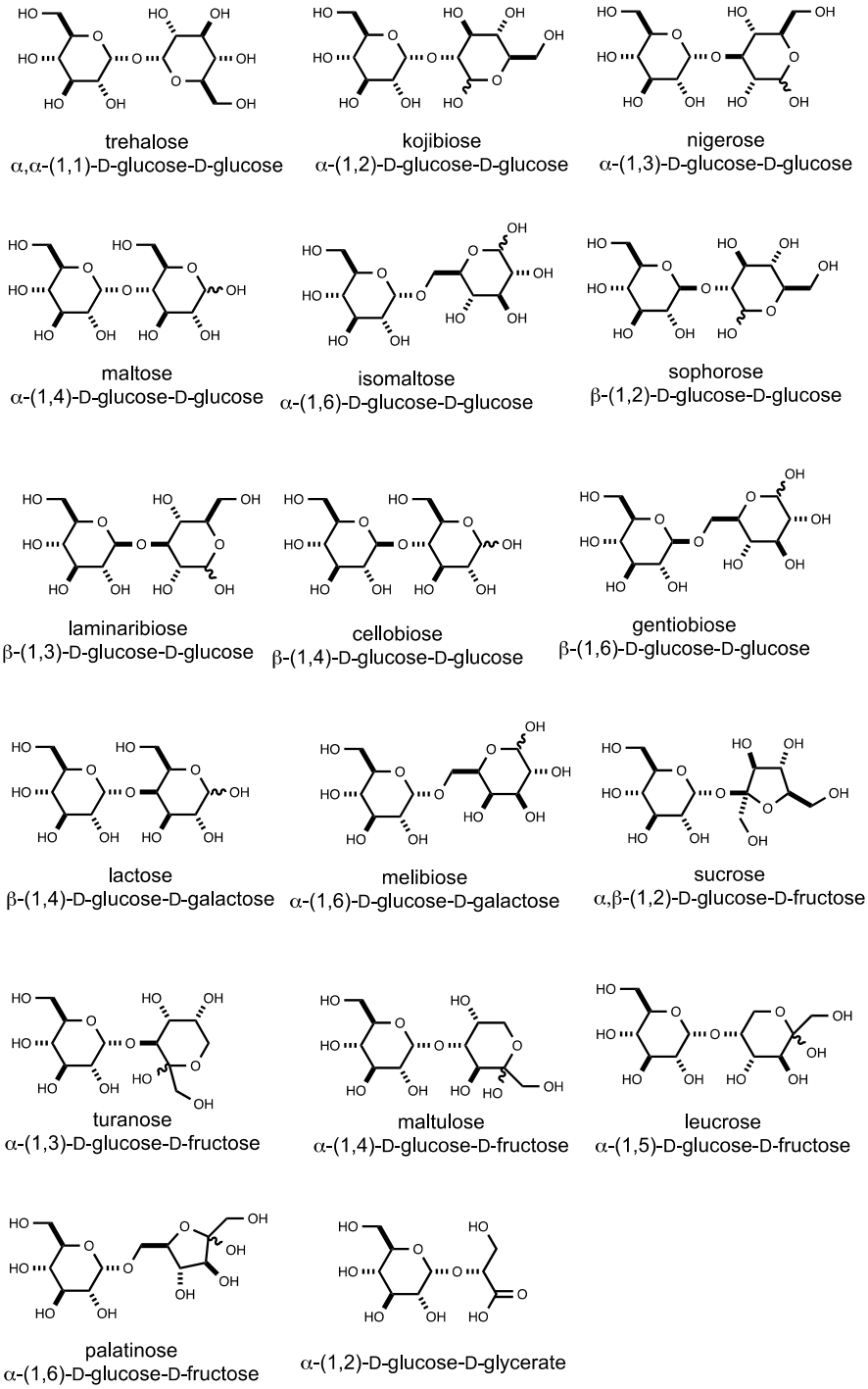
Functional Characterization of YcjU, a β-Phosphoglucomutase. An InterPro search with the protein sequence shows that YcjU belongs to the HAD hydrolase superfamily, subfamily IA and cog0637. In *E. coli*, the source of α-glucose-1-phosphate is primarily from the breakdown of glycogen by glycogen phosphorylase.[141] The catalytic activity of kojibiose phosphorylase releases β-glucose-1-P in *E .coli,* which is a substrate for YcjU. The catalytic efficiencies of β-phosphoglucomutase (β-PGM) from *Lactococcus lactis* and YcjU are comparable.[142] The two enzymes share 44% identity and a comparison of their structures *L. lactis* β-PGM (PDB id: 1LVH) and *E. coli* YcjU (PDB id: 4G9B) show that the active site residues are mostly conserved. YcjU exhibits catalytic activity with β-D-mannose-1-P, β-D-allose-1-P and β-D-galactose-1-P to form the corresponding sugar-6-P product. This activity has not previously been reported before, but these β-sugar-1-P products are not substrates for the back reaction catalyzed by YcjT as a replacement for β-D-glucose-1-P. A derivative of *E. coli* K-12 with a mutation in *ycjU* has been reported to be UV-radiation sensitive showing that this enzyme may have other physiological roles in facilitating survival of *E. coli* cells against agents that damage DNA.[143]

The gene cluster responsible for the metabolism of kojibiose in *C. saccharolyticus, T. brockii*, and *Pyrococcus sp*. consists of genes encoding for sugar transport proteins, a transcription regulator, kojibiose phosphorylase, β-phosphoglucomutase and an additional phosphorylase, which also belongs to cog1554. The *ycj* gene cluster in *E. coli*, and in other organisms where this pathway is conserved, has a few additional genes that encode for proteins with the following putative functions: YcjQ and YcjS are NAD^+-dependent sugar

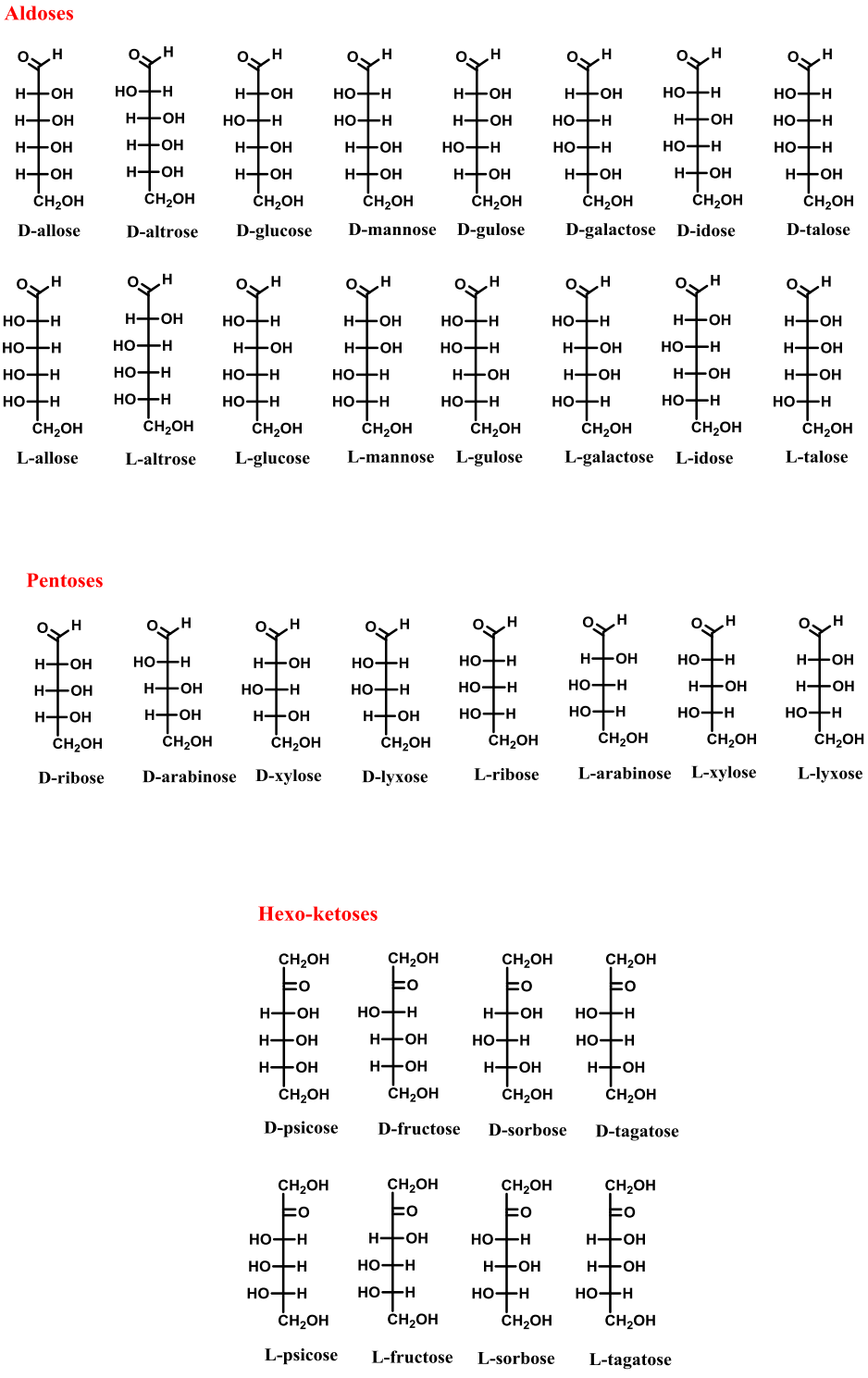
dehydrogenases, YcjR, an epimerase/isomerase, and OmpG, a porin.[144] This observation indicates that the fate of the breakdown products of kojibiose in these organisms might be different from that of *C. saccharolyticus*, *T. brockii*, and *Pyrococcus* spp.

Kojibiose is a rare sugar, which is present in low levels in sake and koji extracts, honey, and beer.[145-147] In the human gut, kojibiose is also found as part of the 'intracellular' teichoic acids in Gram-positive bacteria such as *Streptococci*.[148] Studies have uncovered that these teichoic acids may have a role to play in conferring resistance to β-lactams in Methicillin-resistant *Staphylococcus aureus* (MRSA) and the biogenesis now serves as a target for drug design.[149] Kojibiose has recently gained commercial interest as a potential low calorie sweetener because α-(1,2)-glycosidic bonds are apparently resistant to breakdown by enzymes in the digestive tract.[150-153] Moreover, kojibiose has a prebiotic effect on beneficial colonic bacterial species such as *Bifidobacterium* and *Lactobacillus* in the gut with a high prebiotic index of 22 making it an ideal candidate for an alternate sugar substitute.[150, 153, 154] Additionally, kojibiose has shown inhibitory effects on α-glucosidase I in rat liver microsomes, yeast microsomal extracts, and bovine mammary glands.[155-157] This observation has unveiled the potential for using this disaccharide as part of the pseudo disaccharide class of drugs that targets human immunodeficiency virus (HIV) as well as inhibiting intestinal α-glucosidases as a potential treatment for diabetes mellitus.[150, 158]

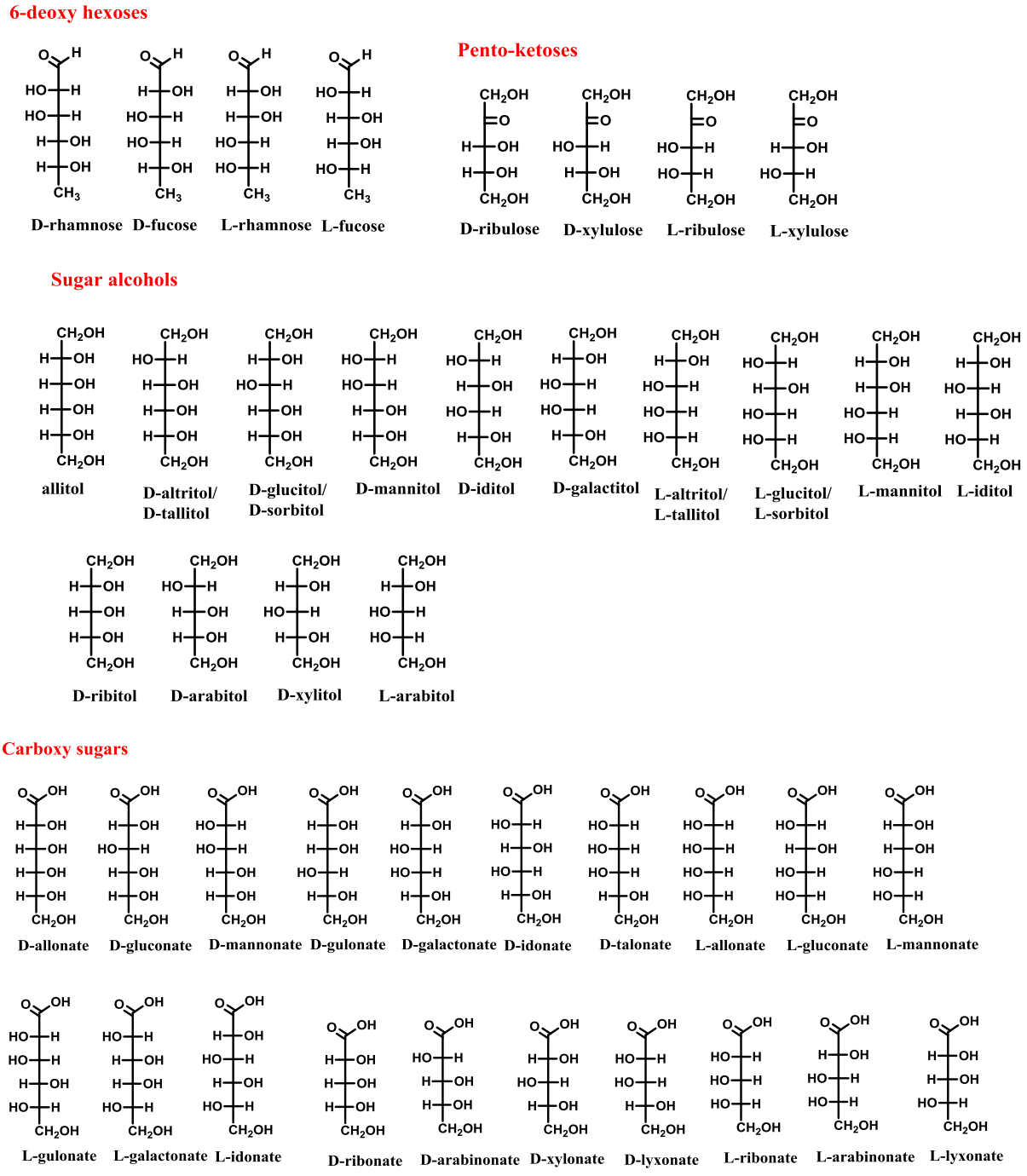
Supplementary Information



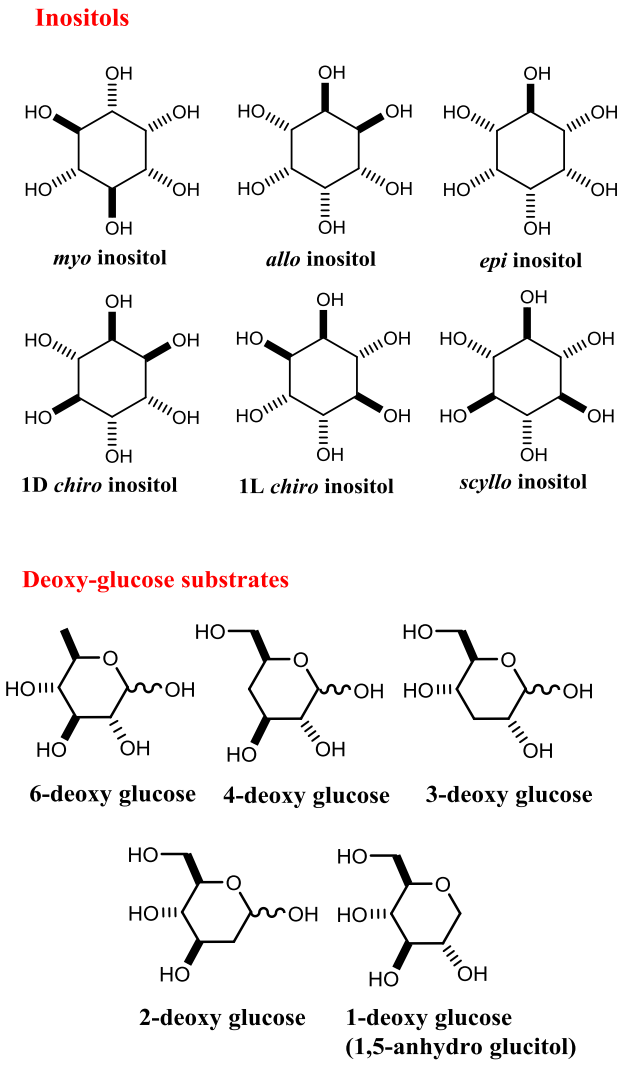
Scheme 2.S1: Structures of disaccharides tested as substrates for YcjM and YcjT.



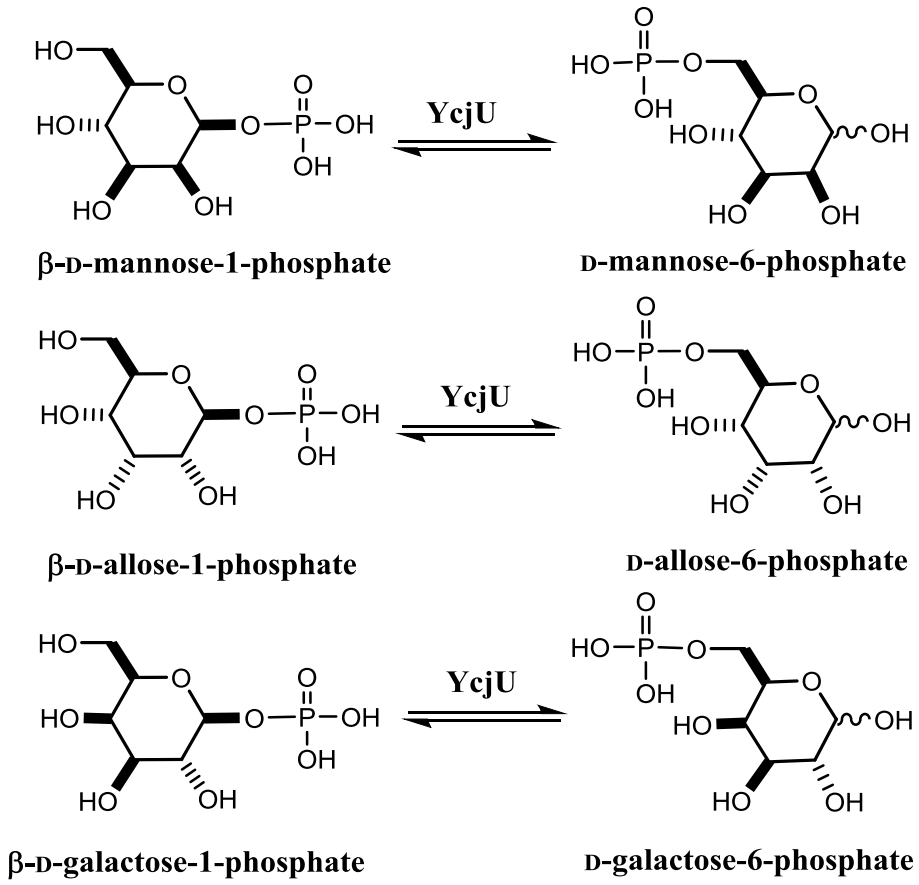
Scheme 2.S2: Structure of aldohexoses, aldopentoses, and ketohexoses tested as substrates for YcjM and YcjT.



Scheme 2.S3: Structures of 6-deoxyaldoses, ketopentoses, sugar carboxylates and sugar alcohols tested as substrates for YcjM and YcjT.



Scheme 2.S4: Structures of deoxy-D-glucose variants and inositols tested as substrates for YcjT.



Scheme 2.S5: Reactions catalyzed by YcjU.

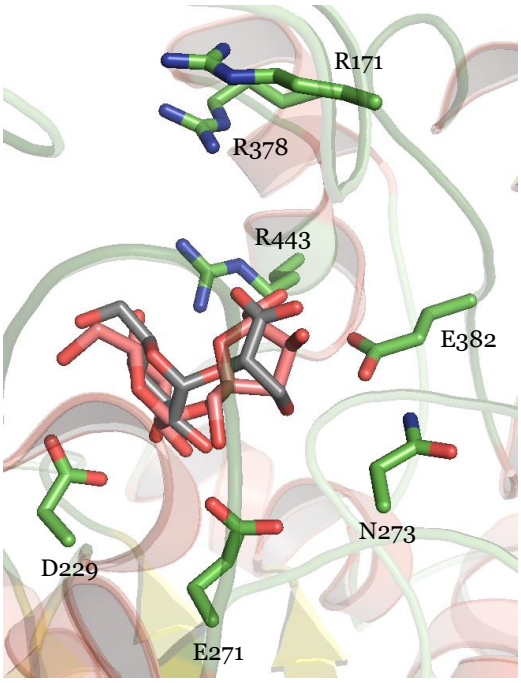


Figure 2.S1: Homology model of YcjM built using Phyre2. The α-(1,2)-D-glucose-D-glycerate was docked in the active using AutoDock Vina. The YcjM model is superimposed with sucrose phosphorylase structure from *B. adolescentis* (2GDU). The bound sucrose is shown in red. In the YcjM model, α-(1,2)-D-glucose-D-glycerate is shown in grey and the residues are shown in green.

Figure 2.S2: Multiple sequence alignment of YcjT, and kojibiose phosphorylases from *Pyrococcus horikoshii* (PsKP), *Caldicellulosiruptor saccharolyticus* (CsKP), and *Thermoanaerobacter brockii* (TbKP) created using Clustal omega. The residues interacting with kojibiose, as observed in the CsKP structure (2IWW), are conserved in YcjT. These residues are indicated in pink. YcjT has Ser 409 in place of conserved Thr residues as indicated in blue. The catalytic glutamate residue is shown in green.

CHAPTER III

FUNCTIONAL CHARACTERIZATION OF *YCJ* QRS FROM A POTENTIAL CATABOLIC PATHWAY OF AN UNKNOWN SUGAR IN THE GUT

We are colonized by billions of microorganisms since birth.[97] Apart from the protective effect the gut microbes exert by preventing colonization by pathogens, microbiota serve many important functions in nutrient and drug metabolism, as well as modulation of the gut immune system.[9, 22, 159] Imbalances in the bacterial composition (or dysbiosis) in the gut can be linked to a number of diseases such as obesity, diabetes, and cancer.[10, 93, 160] Recent studies have also linked Alzheimer's disease, Parkinson's disorder, and psychiatric illnesses to microbial dysbiosis.[161-163] With improved understanding of the human microbiome and the substantial reduction in the cost of whole genome sequencing, analysis of individual microbiomes may eventually be introduced as a routine diagnostic tool to test for wellness. This may lead to adopting a more personalized approach when dealing with chronic diseases as well as advancements in precision medicine.[164-166]

There has been a marked increase in the interest the gut microbiota due, in part, to the technological advancements in DNA sequencing and in gnotobiotics.[96, 167, 168] A study by Qin *et. al,* found that there are an estimated 3.3 million genes present in our gut metagenome, compared to the 23,000 protein-coding genes in humans. Broadly classified, the genes in the metagenome serve either bacterial housekeeping functions or have roles specific to the survival of the bacteria in the gut.[64] One of the most obvious functions required for their survival is aiding in the absorption and metabolism of macromolecules. The microbial 'glycobiome' contains a large number of genes that facilitate in the degradation of complex carbohydrates, which are derived

either from the host diet or from the lining of the intestine.[26, 169, 170] However, functional annotation of these genes has not kept pace with the genome sequencing efforts.

A cluster of 12 genes (*ycjM-W* and *ompG*) that is likely responsible for the catabolism of unknown carbohydrates was identified in *Escherichia coli* K-12 and is shown in **Figure 3.1**. Our previous investigation of this gene cluster indicated that YcjT is a kojibiose phosphorylase, which catalyzes the phosphorolysis of kojibiose, (α-(1,2)-D-glucose-D-glucose), to D-glucose and β-D-glucose-1-phosphate.[171] It was also shown that in the back reaction, YcjT will accept other substrates in place of D-glucose, to form disaccharides containing L-sorbose, D-sorbitol, L-iditol, or 1,5-anhydro-D-glucitol coupled with D-glucose. YcjU is a β-phosphoglucomutase, which converts β-D-glucose-1-phosphate to D-glucose-6-phosphate. The enzyme was also shown to form the corresponding sugar-6-P products in the presence of β-D-allose-1-P, β-D-galactose-1-P, and β-D-mannose-1-P.[171] YcjM was determined to be a α-D-glucosyl-2-glycerate phosphorylase that catalyzes the formation of D-glycerate and α-D-glucose-1-phosphate from α-(1,2)-D-glucose-D-glycerate and phosphate.[171]

The *ycj* operon also encodes putative sugar transport components (YcjN, YcjO, YcjP, YcjV, and OmpG), two NAD-dependent sugar dehydrogenases (YcjS and YcjQ), a predicted sugar isomerase/epimerase, YcjR, and YcjW, annotated as a LacI type repressor. This putative carbohydrate metabolic pathway is also conserved in a number of other Gram-negative bacteria that are found in the human gut including, *Salmonella enterica*, *Shigella dysenteriae*, *Erwinia tasmaniensis*, and *Citrobacter rodentium*, among others (*www.microbesonline.org*). In this paper we describe our attempts to functionally characterize the two NAD-dependent sugar dehydrogenases (YcjS and YcjQ), and the putative isomerase/epimerase, YcjR.

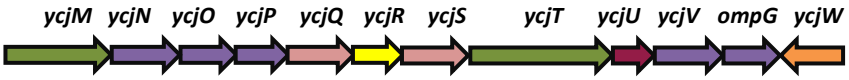


Figure 3.1: The *ycj* gene cluster in *Escherichia coli*. The proteins encoded by this cluster include two polysaccharide hydrolases/phosphorylases (YcjM and YcjT), five sugar transporters (YcjN, YcjO, YcjP, YcjV, and OmpG, a porin), two NAD-dependent dehydrogenases (YcjQ and YcjS), a putative isomerase/epimerase (YcjR), a β- phosphoglucomutase (YcjU, and a LacI-type repressor (YcjW).

Materials and Methods

Materials. The restriction endonucleases, *pfu* turbo polymerase, and T4 DNA ligase used in the cloning of *ycjQ*, *ycjR*, and *ycjS* were purchased from New England Biolabs. The PCR cleanup and gel extraction kits were bought from Qiagen. The plasmid miniprep kit, buffers, phenylmethylsulfonyl fluoride (PMSF), DNase I, and Chelex® resin were obtained from Sigma-Aldrich. Isopropyl-β-D-thiogalactopyranoside (IPTG), nicotinamide adenine dinucleotide (NAD), nicotinamide adenine dinucleotide reduced (NADH), nicotinamide adenine dinucleotide phosphate (NADP), and nicotinamide adenine dinucleotide phosphate reduced (NADPH) were purchased from Research Products International Corporation. Selenomethione was obtained from TCI America. All isotopically-labeled derivatives of D-glucose and D-gulose were purchased from Omicron Biochemicals. The carbohydrates used as potential substrates for YcjS and YcjQ were obtained from Carbosynth, with the exception of methyl-α-D-glucopyranoside and methyl-β-D-glucopyranoside, which were purchased from Sigma Aldrich.

Synthesis of Substrates for YcjS and YcjQ. Methyl-D-[3-^{13}C] glucopyranoside and methyl-β-D-gulopyranoside were synthesized from D-[3-^{13}C] glucose and D-gulose, respectively, by modification of a previously published procedure.[172] α-(1,2)-D-glucose-D-glycerate and 3-keto-D-glucose were synthesized as described previously.[108, 173, 174]

Construction of Sequence Similarity Networks. Cytoscape (www.cytoscape.org) was used to generate the sequence similarity networks. Sequences listed under cog0673 (YcjS), cog1063 (YcjQ), or cog1082 (YcjR) were downloaded as FASTA files from the NCBI database. These files were then uploaded to the Enzyme Function Initiative (EFI) enzyme similarity tool website (https://efi.igb.illinois.edu/efi-est/). The data sets generated were analyzed under an organic layout on Cytoscape.[74]

Cloning of ycjS from E. coli K-12. The DNA corresponding to *ycjS* (gi|16129276; Uniprot P77503) was amplified from *Escherichia coli* K-12 MG1655 genomic DNA using the following pair of primers:

5′-ACCGTGAATTCATGAAAATCGGCACACAGAATCAGGCG-3′

5′-AATCCAAGCTTTTAGCAGGTACGCAACCAGGC-3′

The gene was cloned into pET30a (+) by digestion with *Eco*RI and *Hin*dIII restriction enzymes followed by ligation with T4 DNA ligase to create the recombinant plasmid, which was verified by sequencing. Cloning was carried out to facilitate a His_6-tag at the amino-terminus of the recombinant protein. BL-21 (DE3) cells were transformed with plasmid containing *ycjS* and plated on LB agar supplemented with 50 μg/mL kanamycin.

Expression and Purification of YcjS. A 15-mL culture from a single colony of BL-21 (DE3) was used to inoculate three 1 L cultures of LB with 50 μg/mL kanamycin in 2.8-L Fernbach flasks. The flasks were incubated at 37 ^{o}C with shaking at 150 rpm until the OD_{600}

reached ~0.6. IPTG was added to a final concentration of 0.5 mM to induce protein overexpression. The cell cultures were allowed to grow overnight at 25 ^{o}C. The cells were harvested by centrifugation at 11,000 g for 15 min, and the pellet was stored at -80 ^{o}C until needed.

The frozen cell pellet was suspended in 50 mM HEPES/K^{+}, 100 mM KCl, 10 mM imidazole, pH 8.0. Prior to sonication, 0.4 mg/mL of DNase I and 0.1 mg/mL of PMSF were added to the cell lysate. Following sonication, the cell debris was removed by centrifugation at 12,000 g for 15 min and the supernatant fluid containing the crude protein was passed through a 0.45 μm filter before being loaded onto a 5-mL HisTrap column. The protein was eluted from the column by applying a linear gradient of (0 – 0.5 M) imidazole in 50 mM HEPES/K^{+}, 100 mM KCl, pH 8.0. Individual fractions were analyzed by SDS-PAGE, combined, and dialyzed against 50 mM HEPES/K^{+}, 100 mM KCl, pH 8.0. Aliquots were flash-frozen and stored at -80 ^{o}C. Protein concentration was determined spectrophotometrically at 280 nm using an extinction coefficient of 38,900 M^{-1} cm^{-1} and molecular weight of 43 kDa that includes the 5-kDa linker between the His-tag and the protein (https://web.expasy.org/protparam/). Typical yield of purified YcjS was ~10 mg of protein from 1.0 L of cell culture.

Cloning of ycjQ from E. coli K12. The following set of primers was used to amplify *ycjQ* (gi| 16129274; UniProt: P76043) from the *E. coli* K-12 genomic DNA:

5′-ACCGTGAATTCATGAAAAAGTTAGTAGCCACAGCACCGC-3′

5′-AATCCAAGCTTTTAAAACGTAACGCCCATTTTGATGCTCTGTTCCG-3′

The gene was cloned with a His_6-tag at the N-terminal end. The amplified gene and pET30 a(+) vector were digested with *Eco*RI and *Hin*dIII and ligated with T4 DNA ligase to create the

recombinant plasmid. The DNA sequence of the inserted gene in was verified by sequencing. *E. coli* BL-21 (DE3) cells were transformed with the plasmid by electroporation and plated on LB agar that contained 50 μg/mL kanamycin.

Expression and Purification of YcjQ. A 10-mL culture of LB, started from a single BL-21 (DE3) colony, was grown overnight at 37 °C. This sample used to inoculate two 1 L cultures of LB media supplemented with 50 μg/mL kanamycin and 1.0 mM $ZnCl_2$ in 2.8-L Fernbach flasks. The cultures were allowed to grow at 37 °C until the OD_{600} reached ~0.6. The expression of protein was induced with 0.5 mM IPTG and the cell cultures were allowed to grow overnight at 25 °C. The following day the cells were harvested by centrifugation at 11,000 g for 12 min. The cell pellet was stored at -80 °C until needed.

The frozen cell paste was thawed and suspended in 50 mM HEPES/K^+, 200 mM KCl, 15 mM imidazole, pH 8.0 (binding buffer). DNase (0.4 mg/mL) and PMSF (0.1 mg/mL) were added to the mixture and left to stir for 15 min at 4 °C. The cells were lysed using a Branson Sonifier 450 for three 4-min intervals at 50% output. After sonication, the cell debris was separated from the supernatant fluid by centrifugation at 12,000 g for 20 min. The supernatant solution was passed through a 0.45 μm filter and loaded onto a HisTrap column, which had been equilibrated with binding buffer. The bound protein was then eluted from the column by applying a linear gradient of elution buffer (50 mM HEPES/K^+, 200 mM KCl, 0.5 M imidazole, pH 8.0). The fractions were analyzed by SDS-PAGE, pooled, and the imidazole was removed by dialyzing against 50 mM HEPES/K^+, 100 mM KCl, pH 8.0. Following dialysis, the protein was concentrated using Vivaspin protein concentrators, aliquoted into smaller volumes, flash frozen, and stored at -80 °C. The protein concentration was determined spectrophotometrically

at 280 nm using an extinction coefficient of 40,400 M^{-1} cm^{-1} and molecular weight of 38.2 kDa, which includes the 5-kDa linker between the His-tag and the protein.

Characterization of Dehydrogenase Activities of YcjS and YcjQ. All kinetic assays were conducted at 30 °C. A library of sugars (**Schemes 3.S1** and **3.S2**) was tested as potential substrates for YcjS and YcjQ. The assays contained 2.0 mM substrate, 1.0 mM NAD, 0.5 μM YcjS or YcjQ in 50 mM CHES/K^+, 100 mM KCl, pH 9.5. The increase in absorbance was monitored at 340 nm ($\Delta\varepsilon$ = 6,220 M^{-1} cm^{-1}). The kinetic constants for the oxidation of D-glucose by YcjS were measured with varying concentrations (0 - 30 mM) of the substrate, 2.0 mM NAD, and 0.1 μM YcjS in 50 mM CHES/K^+, 100 mM KCl, pH 9.5. Steady state kinetics with α-methyl-D-glucopyranoside (0 – 20 mM) and β-methyl-D-glucopyranoside (0 – 30 mM) were carried with 2.0 mM NAD+ and 0.5 μM YcjS. Kinetic constants with two other substrates 1,5-anhydro-D-glucitol (0 – 20 mM) and α-(1,2)-D-glucose-D-glycerate (0 – 5 mM) were determined in reactions containing 2.0 mM NAD+ and 2.0 μM and 0.5 μM respectively. Kinetic parameters for the reduction of 3-keto-D-glucose were conducted in 50 mM cacodylate/K^+, pH 6.5, in assays containing varying concentration (0-10 mM) of 3-keto-D-glucose, 250 μM NADH, and 0.1 μM YcjS. Kinetic constants for the oxidation of YcjQ were measured by varying (0 – 25 mM) D-gulose, 2.0 mM NAD+, and 0.05 μM YcjQ in 50 mM CHES/K^+, 100 mM KCl, pH 9.5. The reactions to determine kinetic parameters with β-methyl-D-gulopyranoside contained varying substrate (0 – 40 mM), 4.0 mM NAD+, 2.0 μM YcjQ in 50 mM CHES/K^+, 100 mM KCl, pH 9.5.

Assays for Formate. The NAD-dependent formate dehydrogenase (FDH) catalyzes the oxidation of formate to carbon dioxide. The combined assays with FDH and YcjS initially contained 130 μM D-glucose, 4 mM NAD, and 2.0 μM YcjS in 50 mM CHES/K^+, pH 9.5, at 30

°C. The formation of NADH was monitored at 340 nm ($\Delta\varepsilon = 6220\ M^{-1}\ cm^{-1}$). The enzymatic reaction was allowed to proceed until there was no further increase in the absorbance at 340 nm. A total of 2 U of FDH was added and the reaction was monitored at 340 nm for 90 min. In a complementary assay 250 µM glucose, 4 mM NAD, 1 µM YcjS and 3 U FDH was incubated with 50 mM CHES/K+, pH 9.5, in a volume of 1.0 mL and the change in absorbance was monitored at 340 nm. Formate dehydrogenase was also used to quantify the amount of formate released from the YcjQ-catalyzed reaction with D-gulose. The reactions contained 150 µM D-gulose, 4 mM NAD, and 2.0 µM YcjQ in 50 mM CHES/K^+, pH 9.5, at 30 °C. The formation of NADH was monitored at 340 nm. The enzymatic reaction was allowed to proceed until there was no further increase in the absorbance at 340 nm. A total of 3 U of FDH was added and the reaction was monitored at 340 nm for 90 min.

NMR and Mass Spectrometry. LC-MS experiments, carried out to determine the specific carbon that is oxidized by YcjS contained the following: 250 µM of the deuterium labeled (at C-1, C-2, C-3, C-4 or C-6) D-glucose, 2.0 mM NAD, 1.7 µM YcjS in 50 mM CHES/K^+, pH 9.5. The reactions were incubated for 6 h at 30 °C, and the enzyme was removed using the Vivaspin protein filters. The NADH was isolated using reverse phase chromatography and electrospray ionization in the positive mode combined with tandem mass spectrometry was used to identify the mass of the NADH produced during the oxidation of D-glucose. LC-MS was performed using an Agilent 1260 HPLC system connected to a MicroToF-QII mass spectrometer (Bruker Daltonics). Separation was carried out on a C-18 column (3.0 cm x 100 mm, 2.7 µm particles). Buffers A (5 mM ammonium acetate, pH 6.6) and B (75% methanol and 25% H2O) were used for liquid chromatography.

^{13}C-NMR experiments were utilized to determine the product of the YcjS-catalyzed oxidation of D-glucose. The reaction contained 1.0 mM D-[UL-$^{13}C_6$] glucose, 3.0 mM NAD, 5.0 μM YcjS in 50 mM pyrophosphate buffer, pH 9.5. Similar reactions were also carried out with D-glucose containing a ^{13}C-label at positions C-1, C-2, C-3, C-4, C-5 or C-6. The formation of the oxidized product from ^{13}C-labeled D-glucose were also carried out in reactions that contained 1.0 mM D-[1-^{13}C] glucose, 1.0 mM D-[3-^{13}C] glucose, 300 μM NAD, 20 mM sodium pyruvate, 40 U lactate dehydrogenase, and 10 μM YcjS in 50 mM cacodylate/K^+, pH 6.5. ^{13}C-NMR experiments with enzymatic reactions containing methyl-D-[3-^{13}C] glucopyranoside as the substrate contained 2.0 mM methyl-D-[3-^{13}C] glucopyranoside, 300 μM NAD, 5.0 μM YcjS, 20 mM sodium pyruvate, and 40 U lactate dehydrogenase in 50 mM ammonium bicarbonate/K^+, pH 8.0

^{13}C-NMR experiments with YcjQ and D-[UL-$^{13}C_6$] gulose contained 1.0 mM labeled substrate, 3.0 mM NAD, 5.0 μM YcjQ in 50 mM pyrophosphate buffer, pH 9.5. In an attempt to determine the product of the YcjQ reaction, ^{13}C-NMR experiments were conducted with methyl-β-D-gulopyranoside. These reactions contained 10 mM methyl-β-D-gulopyranoside, 300 μM NAD, 20 μM YcjQ, 20 mM sodium pyruvate, and 40 U lactate dehydrogenase in 50 mM ammonium bicarbonate/K^+, pH 8.0.

Determination of Absorbance at 310 nm. An absorbance at 310 nm has been reported to be indicative of a 3-keto-D-glucose product.[175] To determine if YcjS catalyzed reaction with D-glucose showed absorbance at 310 nm an assay was carried out with the following: 1.0 mM D-glucose, 0.5 μM YcjS, 0.5 mM NAD^+, 10 mM sodium pyruvate, 20 U LDH in 50 mM CHES/K+, 100 mM KCl, pH 9.5. The reaction was monitored at 340 nm as well as 310 nm for 3.0 h. The same reaction was also carried out at pH 8.5 (50 mM HEPES/K^+, 100 mM KCl).

The YcjQ reaction with D-gulose was also carried out under similar conditions to look for the absorbance at 310 nm.

Determination of Kinetic Constants. SigmaPlot 11.0 was used to fit the initial velocity data to Eq. 1 to determine the kinetic constants where v is the velocity, E_t is the enzyme concentration, k_{cat} is the turnover number, A is the substrate concentration, and K_m is the Michaelis constant.

$$v / E_t = k_{cat} A / (K_m + A) \quad (1)$$

Cloning of ycjR from E. coli K12. The gene corresponding to *ycjR* (gi|90111248; UniProt P76044) was amplified from *E. coli* K-12 genomic DNA with the following primer pair:

5′-ACCGTGAATTCATGAAAATCGGCACACAGAATCAGGCG-3′

5′-AATCCAAGCTTTTAGCAGGTACGCAACCAGGC-3′

The vector pET30a(+) and the amplified gene were digested with *Eco*RI and *Hin*dIII restriction enzymes and ligated together with T4 DNA ligase to create the recombinant plasmid. The *ycjR* gene was cloned with an N-terminal His_6 tag. Following verification by DNA sequencing, BL-21 (DE3) cells were transformed with the plasmid and plated on LB agar with 50 μg/mL *kanamycin.*

Expression and Purification of YcjR. A single colony was used to start a 10- mL culture of LB supplemented with 50 μg/mL kanamycin and allowed to grow overnight at 37 °C. The 10-mL LB culture was used to inoculate two 1 L cultures of LB in 2.8-L Fernbach flasks. To each 1 L LB culture, 50 μg/mL kanamycin and 1.0 mM $MnCl_2$ was added. The cultures were allowed to grow at 37 °C until the OD_{600} reached ~0.6, 1.0 mM IPTG was added to induce protein

expression, and then growth was resumed at 25 °C overnight. The cells were harvested by centrifugation at 11,000 g for 12 min. The cell paste was stored at -80 °C until needed.

The frozen cell pellet was thawed and re-suspended in binding buffer that contained 20 mM HEPES/K^+, 200 mM KCl, 20 mM imidazole, pH 8.0. Before sonication 0.1 mg/mL PMSF and 0.4 mg/mL DNase I were added to the cell suspension and stirred for 10 min at 4 °C. The cells were lysed using a Branson Sonifier 450 for four 5-minute intervals at 50% output. The cell suspension was clarified by centrifugation at 12,000 g for 20 min. The supernatant was passed through a 0.45 μm filter and loaded on a 5-mL HisTrap column, which was previously equilibrated with binding buffer. The protein of interest was eluted from the column using a linear gradient of the elution buffer, which contained 20 mM HEPES/K^+, 200 mM KCl, 0.5 M imidazole, pH 8.0. Based on SDS PAGE, the isolated protein was >95% pure. The appropriate fractions were pooled and concentrated. To remove imidazole, the concentrated protein was passed through a PD-10 column, which had been previously equilibrated with 20 mM HEPES/K^+, pH 8.0. Protein concentration was determined spectrophotometrically at 280 nm using an extinction coefficient of 33,000 M^{-1} cm^{-1} and a molecular weight of 29.8 kDa, which includes the 5-kDa linker region between the protein and the His-tag (https://web.expasy.org/protparam/). The collected fractions were pooled, flash frozen, and stored at -80 °C. Typical yield for this protein was ~60 mg of protein per liter of cell culture.

Metal Content Analysis. The metal content of YcjQ and YcjR was determined using a Perkin-Elmer DRCII inductively coupled plasma mass spectrometer (ICP-MS). Samples for ICP-MS were digested with ≥69% (v/v) nitric acid, to prevent precipitation of proteins during measurement, and refluxed for 30 min. The samples were then diluted with deionized water to ensure a final concentration of ~1.0 μM protein and 1% (v/v) nitric acid.

Expression and Purification of YcjR Enriched with Selenomethionine. The pET30a (+) plasmid containing *ycjR* was transformed into the methionine auxotroph, B834 (DE3) pLysS cells, and applied to LB plates supplemented with 50 μg/mL kanamycin. A single colony was picked from the plate and used to inoculate 30 mL of LB media. This starter culture was grown overnight at 37 °C and was used to start three 1-L LB cultures in 2.8-L Fernbach flasks. These cultures were allowed to grow for 24 h at 30 °C. Minimal media stock solutions were made by dissolving 37.5 g K_2HPO_4, 9.0 g KH_2PO_4, 7.5 g NaCl, and 7.5 g NH_4Cl in 300 mL distilled water, adjusting the pH to 7.5 using 5 M KOH, followed by autoclaving. The trace-metals solution was prepared by combining 250 mg $FeSO_4$ ($6H_2O$), 3.0 mg $ZnCl_2$, 3.0 mg $CoCl_2$, 3.0 mg $MnCl_2$, 0.6 g $MgSO_4$, 3.0 mg $CaCl_2$ in 30 mL distilled water and filter sterilized using a 0.22 μm filter. The 20% (w/v) glucose solution was made by dissolving 60 g of D-glucose into 300 mL distilled water and autoclaved. The following chemicals were filter sterilized separately: 50 μg/mL kanamycin, 0.5 M IPTG, 1.0 mM $MnCl_2$, 50 mg/mL thiamine, 50 mg/mL selenomethionine. Separately, 780 mL of water was autoclaved in three 2.8-L Fernbach flasks.

After 24 h the cells growing in LB media from each flask were harvested separately by centrifugation at 9,000 g for 15 min and resuspended in three 1-L 1X minimal media in 2.8-L Fernbach flasks. The 1x minimal media was made by combining 100 mL of 10X minimal media stock with 100 mL 20% glucose solution, 10 mL trace metals solution (freshly made), 50 μg/mL kanamycin, and 50 mg/mL thiamine solutions. The cultures were allowed to grow at 30 °C for 9 h to starve the cells of methionine. 1.0 mL of filter-sterilized selenomethionine (50 mg/mL) was added to each flask and the cultures were left to grow for 45 min. Following this, 1.0 mM $MnCl_2$ was added to the cultures and after 30 min protein expression was induced with 0.5 mM IPTG. After induction, the cells were allowed to grow overnight at 25 °C and harvested the following

day by centrifugation at 8,700 g for 12 min. Purification of the selenomethionine supplemented YcjR was carried out exactly as described earlier with the exception of adding 2.0 mM TCEP to all the buffers in order to keep the selenomethionine reduced. The total yield of YcjR from 3 L of cell culture was 184 mg. The protein was concentrated using Vivaspin concentrators. When the final protein concentration was 54 mg/mL, the protein was flash-frozen and stored at -80 °C.

Crystallization and Data Collection. Suitable conditions for crystallization were determined by setting up a sparse matrix screen (approx. 1000 conditions) using reagents commercially bought from Hampton research. Intelliplates (93-3 LVR) were used to set up the crystallization drops, which contained 10 mg/mL of YcjR (preincubated with 100 µM of $MnCl_2$) and the reservoir solution in a 1:1 ratio. The plates were then incubated at 18 ˚C. Within a week crystals were observed for the following conditions, 0.1 M sodium malonate pH 5.0, 12% w/v PED 3350 (PEG/Ion Screen) and 0.5 M ammonium sulfate, 0.1 M sodium citrate tribasic dehydrate, pH 5.6, 1 M $LiSO_4$ (Crystal Screen). Diffraction quality crystals were produced under the first set of conditions using the hanging drop method by mixing in a 1:1 ratio, 8 mg/mL of selenomethionine-labeled YcjR, which was preincubated with 1.3 mM $MnCl_2$ for 1 h and 0.1 M sodium malonate pH 5.0, 12% w/v PED 3350.

Structure Determination and Refinement. The initial structure of YcjR was determined using single wavelength anomalous diffraction. Prior to X-ray data collection, the selenomethionine YcjR crystals were soaked in a stabilization solution containing 20% ethylene glycol (cryoprotectant) in the original reservoir solution. The crystals were then flash frozen in liquid nitrogen and sent to the Stanford Synchrotron Radiation Lightsource (SSRL). A fluorescence scan of the YcjR crystal identified the selenium peak (12661.38 eV), inflection

(12659.51 eV), and remote (13000.00 eV) wavelengths. Three separate data sets were collected at the peak, inflection and remote wavelengths.

HKL2000 was used to index, scale and integrate the diffraction data.[176] The selenium peak data set were used to determine the YcjR structure in PHENIX autosol.[177] The electron density was determined using PHENIX Autobuild and COOT was used to further construct the structure.[177, 178] The structure refinements were carried out using PHENIX by optimizing x-ray coordinates, TLS parameters, occupancies, and individual B-factors. The refined structure has excellent geometry, with more than 98% of residues in favored Ramachandran regions (**Table 3.1**).

Table 3.1: Crystallographic data collection and refinement statistics for YcjR.

Data collection	
Wavelength (Å)	0.979231
Resolution (Å)	50.00-1.86 (1.89-1.86)
Space group	P 4_1 2_1 2
a, *b*, *c* (Å)	116.6, 116.6, 247.7
α, β, γ (°)	90.0, 90.0, 90.0
Z (molecules/ASU)	4
Total/unique reflections	2,113,171/142,992
Multiplicity	14.8 (14.9)
Completeness (%)	100.0 (100.0)
Mean I/σI	36.7 (3.7)

Table 3.1 Continued.

$R_{sym}/R_{meas}/R_{pim}$	0.101/0.104/0.027
Refinement	
R_{work}/R_{free}	0.1682/0.1937
RMSD bonds (Å)	0.009
RMSD Angles (°)	1.13
Ramachandran favored (%)	98.85
Ramachandran allowed (%)	1.05
Ramachandran outliers (%)	0.10
Average B ($Å^2$)	25.01
PDB	*to be deposited*

[a]Statistics for the highest-resolution bin are given in parenthesis.
[b] $R_{work} = \Sigma_{hkl}||F_{obs}| - |F_{calc}||/\Sigma_{hkl}|F_{obs}|$, calculated over 95% of the data in the working set. R_{free} is equivalent to R_{work} except that it is calculated over the remaining 5% of the data.

Determination of Catalytic Activity of YcjR. YcjR was tested for catalytic activity with D-psicose, D-fructose, D-tagatose, D-sorbose, L-xylulose, and L-ribulose using ^{1}H NMR spectroscopy. The reactions contained 5.0 mM substrate, 0.5 mM $MnCl_2$, and 5.0 µM YcjR in 50 mM ammonium bicarbonate/K^+, pH 8.0. After incubation with the enzyme for 2.0 h at 30 °C, the enzyme was removed using a Vivaspin protein filter (molecular weight cut-off of 10 kDa) and the reactions were passed over Chelex resin, equilibrated with 50 mM ammonium

bicarbonate/K^+, pH 8.0 to chelate any excess manganese. The substrate solutions and buffer contained >90% D_2O.

Owing to the instability of 3-keto-D-glucose, attempts were made to determine epimerase activity of YcjR with α-methyl-3-keto-D-glucopyranoside. The reactions contained 2.0 mM unlabeled α-methyl-D-glucopyranoside, 2.0 mM NAD^+, and 5.0 μM YcjS in ammonium bicarbonate/K^+, pH 8.0, and were incubated for 3 h at 30 °C. To this assay, 5.0 μM YcjR and 0.5 mM $MnCl_2$ were added and the reaction was allowed to proceed for 2 h at 30 °C. The enzymes were removed using the Vivaspin protein filters (molecular cut-off of 10 kDa) and the reactions were passed over Chelex resin, equilibrated with 50 mM ammonium bicarbonate/K^+, pH 8.0 to remove the excess manganese. The reactions were then subjected to ^{1}H-NMR. All the buffers, substrate and co-enzyme solutions were made in >90% D_2O.

Results

Determination of Substrate Profile for YcjS. YcjS belongs to cog0673 and a sequence similarity network (SSN) was constructed at a BLAST *E*-value of 1 x 10^{-45} (**Figure 3.2**). In this figure each node represents a unique protein sequence in cog0637 and within every cluster, nodes are connected by edges (or lines) that indicate that two protein sequences are related to each other by *E*- value of 10-^{45} or less. Based on this SSN, the closest functionally verified homologues to YcjS are UDP *N*-acetyl-D-glucosamine-3-dehydrogenase from *Methanococcus maripaludis* and UDP-*N*-acetyl-D-glucosaminuronic acid dehydrogenase from *Pseudomonas aeruginosa* and *Thermus thermophiles.*[179, 180] These homologues are NAD-dependent enzymes that oxidize the hydroxyl group at C-3 of their respective substrates.

YcjS was purified to homogeneity and initially tested for catalytic activity against a small and focused library of D- and L-hexoses (**Scheme 3.S1**) using NAD as the oxidant at pH 9.5.

YcjS catalyzed the oxidation of D-glucose (**1**) with a k_{cat} = 7.8 s^{-1}, K_m = 1.5 mM, and k_{cat}/K_m = 5.4 x 10^3 M^{-1} s^{-1}. YcjS was subsequently screened against a larger library of mono- and di-saccharides (**Schemes 3.S1** and **3.S2**). No appreciable activity (<2% relative to D-glucose) was observed with UDP-glucose, UDP-*N*-acetyl-D-glucosamine, D-glucosamine, *N*-acetyl-D-glucosamine or UDP-galactose. None of the other substrates in the library showed any activity except for the following compounds (**Scheme 3.1)**: α-methyl-D-glucopyranoside (**2**), β-methyl-D-glucopyranoside (**3**), 1,5-anhydro-D-glucitol (**4**), and α-(1,2)-D-glucose-D-glycerate (**5**). The kinetic parameters for the oxidation of these substrates by YcjS are summarized in **Table 3.2**.

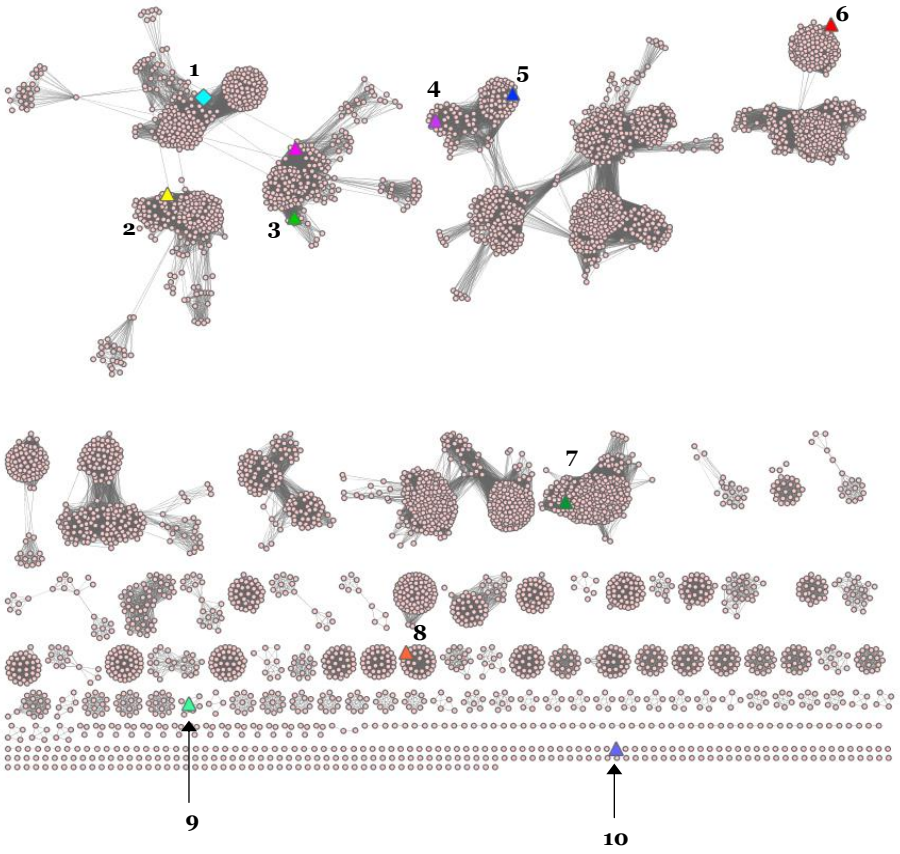
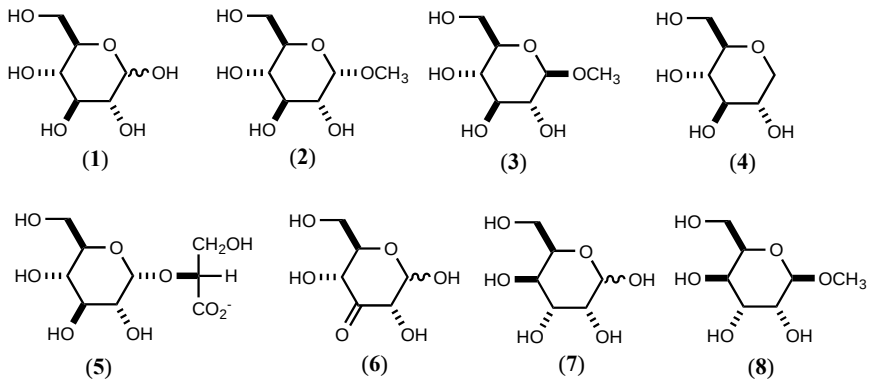


Figure 3.2: Sequence similarity network of cog0673 at an *E*-value cut-off of 1 x 10^{-45}. This network was generated using Cytoscape (www.cytoscape.org). Each node represents a non-redundant protein sequence and every edge (or connecting line) represents a BLAST value *E*-value, between a pair of sequences that are better than 1 x 10^{-45}. The lengths of the edges are not significant; sequences in tight clusters are considered more closely related than clusters with fewer edges. YcjS is depicted as a diamond (cyan) in Group 1. Groups 2 - 10 have experimentally verified functions and are shown as triangles. Group 2: UDP-*N*-acetylglucosamine-3-dehydrogenase from *Methanococcus maripaludis* (yellow); Group 3: UDP-*N*-acetyl-D-glucosaminuronic acid dehydrogenase, from *Pseudomonas aeruginosa* (pink) and from *Thermus thermophiles* (green); Group 4: glucose-fructose oxidoreductase from *Zymomonas mobilis* (pink); Group 5: xylose dehydrogenase from *Haloarcula marismortui* (blue); Group 6: D-*chiro*-inositol-2-dehydrogenase from *Bacillus subtilis* sp. 168 (red); Group 7: *scyllo*-inositol-2-dehydrogenase from *Bacillus subtilis* sp. 168 (green); Group 8: sugar dehydrogenase from a putative trehalose-utilization pathway, in *Agrobacterium tumefaciens* (ThuB); Group 9: D-glucose-6-phosphate dehydrogenase (NtdC) from *Bacillus subtilis* (orange); Group 10: ketoreductase (KijD10) from *Actinomadura kijaniata* (blue).



Scheme 3.1: Substrates for YcjS and YcjQ.

To determine the product of the YcjS-catalyzed reaction with D-glucose, the reaction using [UL-$^{13}C_6$]-D-glucose as the substrate at pH 9.5 was analyzed by ^{13}C-NMR spectroscopy. The NMR spectrum of the reaction products showed a prominent resonance at 171.03 ppm and numerous other resonances in the range of 60-100 ppm (**Figure 3.3**). A similar reaction, using [1-^{13}C]-D-glucose demonstrated that essentially all of the ^{13}C-label from this substrate was found in the product peak with a chemical shift of 171.03 ppm (**Figure 3.S1A**) and subsequently identified as formate. Formate dehydrogenase (FDH) was used to quantify the amount of formate produced in the YcjS-catalyzed reaction. Addition of FDH to a reaction mixture containing YcjS and 130 µM D-glucose and an excess of NAD ultimately produced 257 µM NADH, as determined from the change in absorbance at 340 nm. This result is consistent with the nearly quantitative conversion of C1 from D-glucose to formate after oxidation by YcjS at pH 9.5. ^{13}C NMR experiments with reactions catalyzed by YcjS with NAD and D-glucose containing separate ^{13}C-labels at C-2, C-3, C-4, C-5 or C-6 positions also showed multiple resonances for each carbon in these NMR spectra (**Figures 3.S1B** to **3.S1F**).

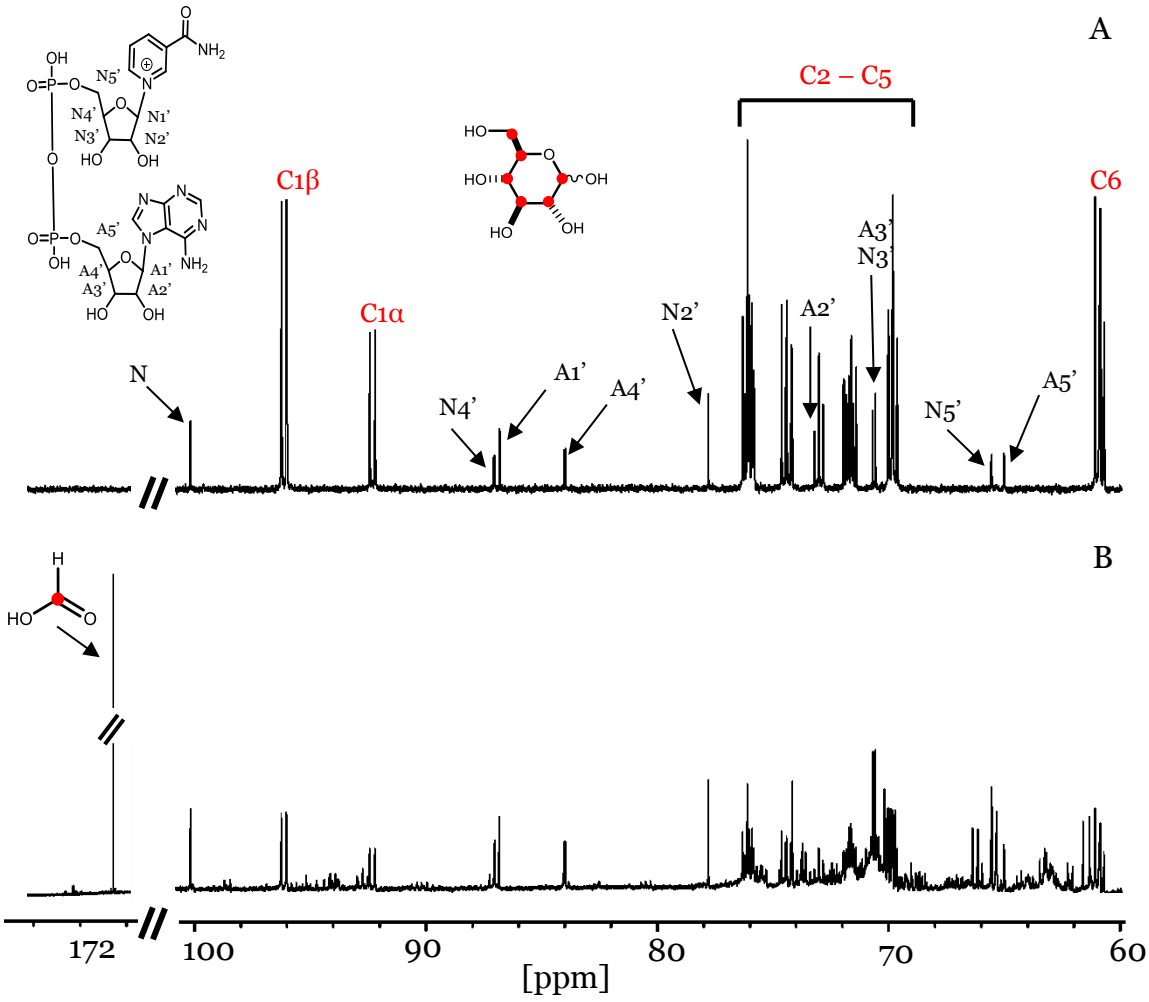


Figure 3.3: ^{13}C-NMR spectra showing [UL-$^{13}C_6$]-D-glucose and NAD^+ with and without addition of YcjS . **(A)** without the addition of YcjS; **(B)** after incubation with the enzyme. The resonance for formate can be observed at 171.03 ppm. The chemical shifts for uniformly labeled glucose are as follows: δ (C1) 96.21, 92.53; (C2) 74.27, 71.62; (C3) 75.94, 72.92; (C4) 69.73, 69.70; (C5) 76.01, 71.49; and (C6) 60.84, 60.68. The resonances for the ribose groups on the coenzyme are indicated. The full ^{13}C-NMR spectrum is shown in **Figure 3.S7**.

To determine which one of the hydroxyl groups contained within D-glucose is oxidized by YcjS, the enzyme-catalyzed reaction was conducted using D-glucose that was mono-deuterated at C-1, C-2, C-3, or C-4. The reduced nucleotide was isolated by ion exchange chromatography and the molecular weight determined by mass spectrometry. Only in the presence of [3-^{2}H]-D-glucose was deuterium transferred from mono-deuterated D-glucose to NAD with the formation of NAD(^{2}H) and an observed *m/z* of 667.144 (**Figure 3.4**). This result clearly demonstrates that the hydroxyl group at C3 of D-glucose is oxidized in the presence of the YcjS and the initial product of the enzymatic reaction is 3-keto-D-glucose (**6**).

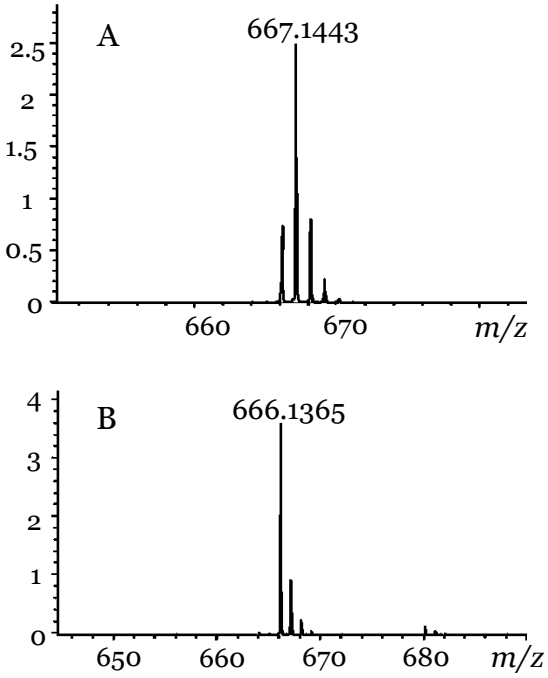
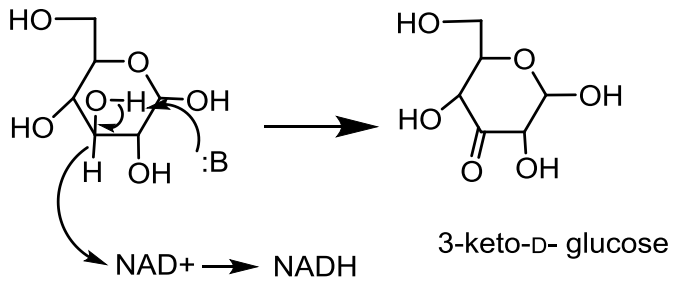


Figure 3.4: LC-MS results showing enzymatic reactions with NAD^+ and D-[3-H^2]-glucose (A) or D-[1-H^2]-glucose (B).

In an attempt to circumvent the apparent instability of the 3-keto-D-glucose product at elevated values of pH, a ^{13}C-NMR experiment was conducted using an equimolar mixture of [1-^{13}C]- and [3-^{13}C]-D-glucose in the presence of YcjS. The formation of NADH was coupled to the reduction of pyruvate at pH 6.5 using lactate dehydrogenase to overcome the unfavorable equilibrium constant for the oxidation of D-glucose at this pH. The NMR spectrum shows the emergence of four new resonances for the oxidized product. The resonances at 94.89 ppm and 97.56 ppm correspond to the resonances for C-1 of the β-pyranose and α-furanose anomers of 3-keto-D-glucose, respectively.[174] The resonances at 207.86 ppm, 206.77 ppm correspond to C3 of the α-D-furanose and β-D-pyranose anomer of 3-keto-D-glucose.[174] Due to the inherent instability of 3-keto glucose, a resonance for formate is also observed at 171.03 ppm (**Figure 3.5**). The ^{13}C-NMR spectrum of chemical synthesized 3-keto-D-glucose is presented in **Figure 3.S2**. The proposed mechanism for the formation of 3-keto-D- glucose is shown in **Scheme 3.2**.



Scheme 3.2: Proposed mechanism for the formation of 3-keto-D-glucose.

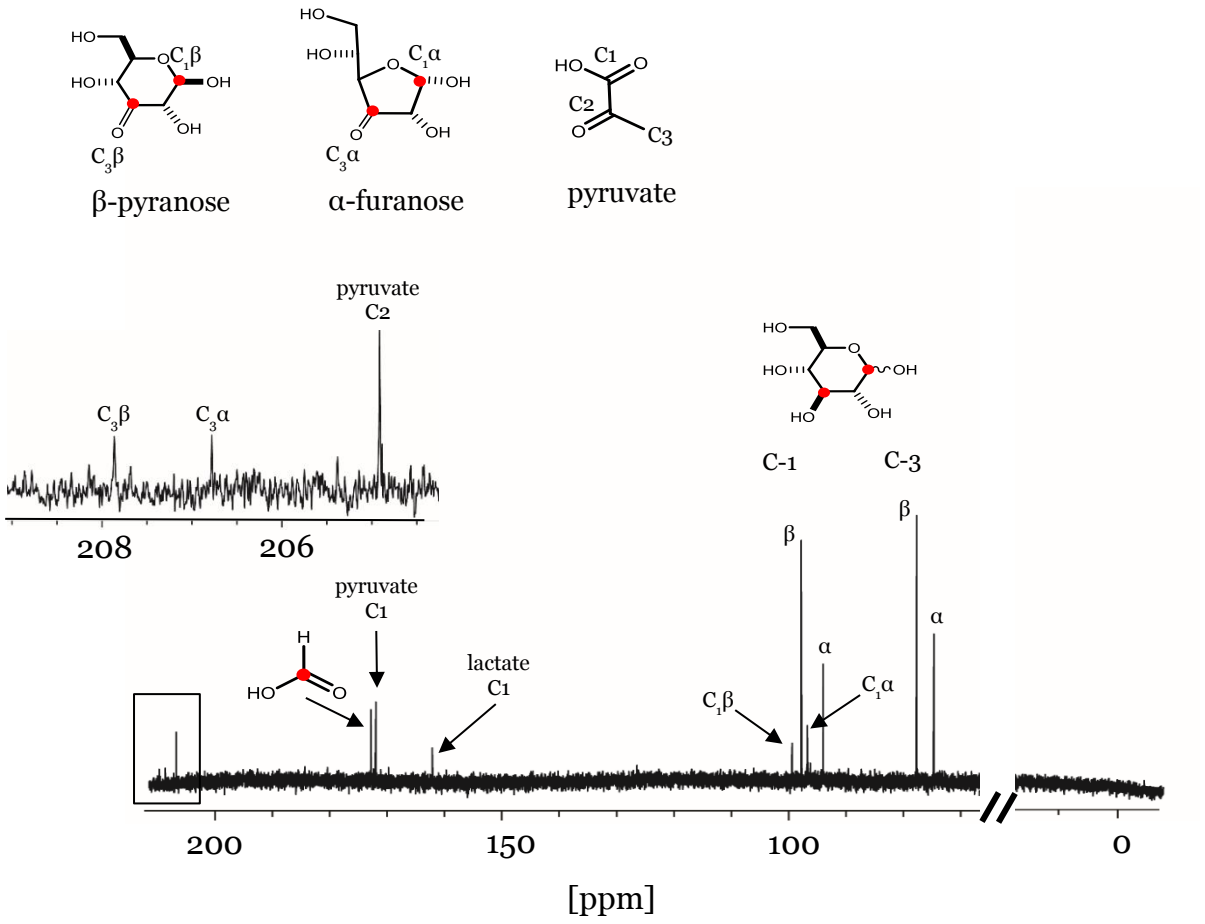


Figure 3.5: ^{13}C-NMR spectra showing the enzymatic reaction with equivalent concentrations of [1-^{13}C]-D-glucose and [3-^{13}C]-D-glucose at pH 6.5 in the presence of pyruvate and lactate dehydrogenase coupling system. The full spectrum is shown in **Figure 3.S8**.

To further work around the stability issues associated with 3-keto-D-glucose, ^{13}C-NMR experiments were conducted with α/β-methyl-[3-^{13}C]-D-glucoside. The NMR spectrum of the product formed from the oxidation of α/β-methyl-[3-^{13}C]-D-glucoside, exhibits two new resonances at 207.25 ppm, which corresponds to the expected chemical shift of the 3-keto product, and at 94.6 ppm, which is most likely the hydrated form of the 3-keto compound (**Figure 3.6**).

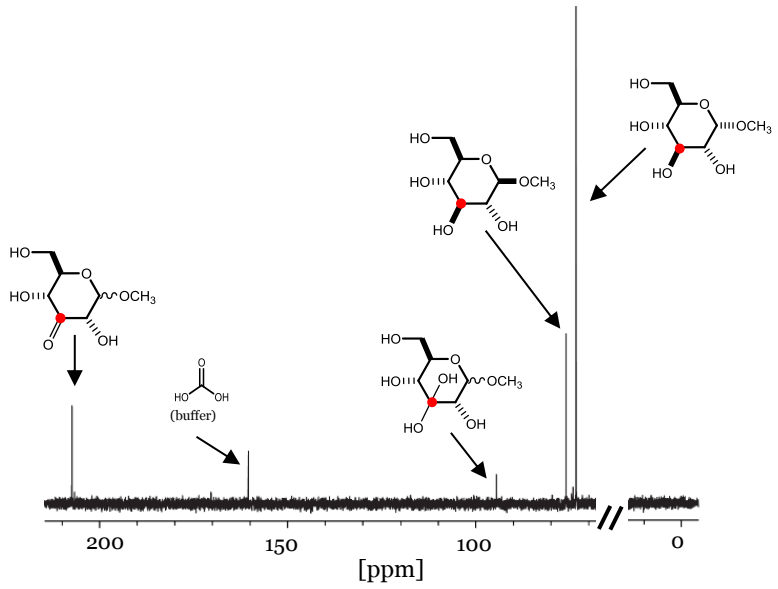


Figure 3.6: ^{13}C-NMR spectra showing the YcjS catalyzed reaction with α/β-methyl-D-[3-^{13}C] glucoside. The full spectrum is shown in **Figure 3.S9**.

Catalytic Activity of YcjQ: The sequence similarity network (**Figure 3.7**) for YcjQ has several proteins whose catalytic properties have been determined previously: *scyllo*-inosose-3-dehydrogenase from *Thermotoga maritima*, L-galactonate dehydrogenase from *Bacteroides vulgatus*, galactitol-1-phosphate-5-dehydrogenase from *E. coli* K-12 MG1655, and sorbitol dehydrogenase from *Bacillus subtilis.*[181-184] However, there were no experimentally verified homologues close to the cluster containing YcjQ. The purified enzyme was found to have a metal content of 0.7 equivalents of zinc/ enzyme. YcjQ was screened for catalytic activity against the library of hexoses (**Scheme 3.S1**) and activity was observed with D-gulose and NAD. The kinetic parameters for the reaction are as follows: k_{cat} = 4.6 s^{-1}, K_m = 4.0 mM, and k_{cat}/K_m = 1.3 x 10^3 M^{-1} s^{-1}. Among the other compounds tested as potential substrates (**Scheme 3.S2**), YcjQ showed activity with β-methyl-D-gulopyranoside (**8**). The kinetic parameters are presented

in **Table 3.2**. A ^{13}C-NMR experiment with [UL-$^{13}C_6$]-D-gulose and NAD with YcjQ at pH 9.5 showed the formation of formate, along with multiple peaks which were most likely degradation products formed because of the instability of the enzymatic product (**Figure 3.S3**). In the presence of formate dehydrogenase, YcjQ, 150 µM D-gulose and an excess of NAD+, approximately 304 µM NADH was formed as observed from an increase in absorbance at 340 nm on the spectrophotometer. This result is consistent with the formate peak observed on the ^{13}C-NMR spectra.

Lability of 3-keto products. An absorbance at 310 nm was observed when YcjS was incubated with D-glucose and 0.5 mM NAD+ in presence of pyruvate and LDH at pH 9.5 (**Figure 3.S4**) and 8.5. As the reactions progressed, absorbance increased with time for ~30 min following which it plateaued. Similar observations were made with the YcjQ reactions with D-gulose. The peak maxima at 310 nm increased in absorbance units as the pH was changed from 8.5 to 9.5. This led us to conclude that the absorbance around 300 nm is most likely because of the enolate degradation products formed following release of formate under alkaline conditions.

Functional characterization of YcjR. The SSN for YcjR (cog1082), at an *E*-value of 1 x 10^{-25} (**Figure 3.8**) indicates that the closest functionally verified homologues are D-tagatose-3-epimerase from *Rhodobacter sphaeroides*, D-psicose-3-epimerase from *Agrobacterium tumefaciens*, and L-ribulose-3-epimerase from *Rhizobium loti*.[185-187] The metal content of YcjR was found to be 0.8 equivalents of manganese/ enzyme.

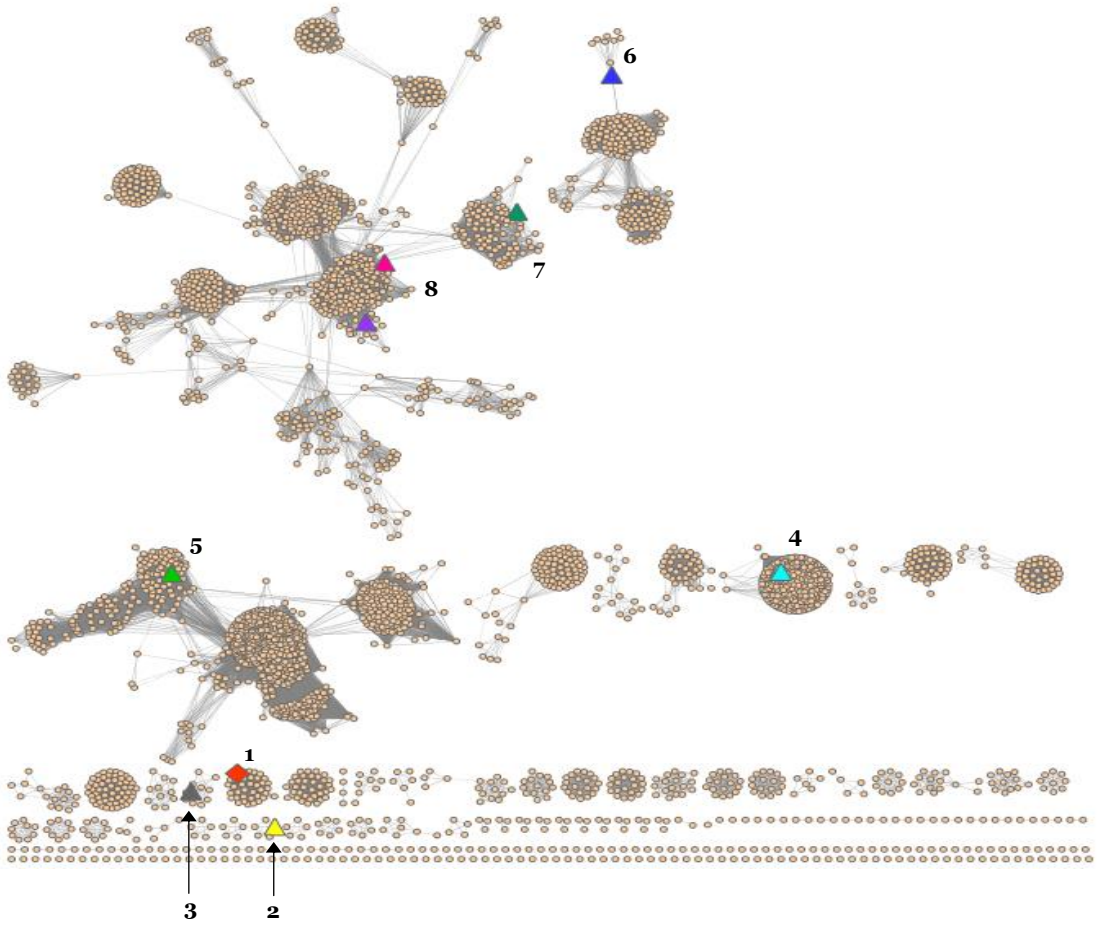


Figure 3.7: Sequence similarity network of cog1063 at an *E*-value cut-off of 1 x 10^{-60}. YcjQ is indicated as a diamond (red) in Group 1. Groups 2-7 contain functions, indicated by triangles, that have been experimentally verified. Group 2: *scyllo*-inosose-3-dehydrogenase from *Thermotoga maritima* (yellow); Group 3: D-glucose-1-dehydrogenase from *Thermoplasma acidophilum* (gray); Group 4: L-threonine-3-dehydrogenase from *Pyrococcus horikoshii* (cyan); Group 5: alcohol dehydrogenase from *Cuprivadus necator* (green); Group 6: L-galactonate dehydrogenase from *Bacteroides vulgatus* (blue); Group 7: galactitol-1-phosphate-5-dehydrogenase from *Escherichia coli* K-12 MG1655 (green); Group 8: L-arabinitol-4-dehydrogenase from *Trichoderma reesei* (pink) and sorbitol dehydrogenase from *Bacillus subtilis* str. 168 (purple).

Table 3.2: Kinetic constants with YcjS and YcjQ and their respective substrates. All the reactions were carried out at pH 9.5, except for the steady state kinetics with 3-keto-D-glucose which was carried out at pH 6.5. All enzymatic incubations were carried out at 30˚C.

enzyme	**varied substrate**	k_{cat} **(s^{-1})**	K_m **(mM)**	k_{cat}/K_m **($M^{-1}\ s^{-1}$)**
YcjS	D-glucose	7.8 ± 0.7	1.5 ± 0.3	(5.4 ± 0.5) x 10^3
	α-methyl-D-glucopyranoside	1.1 ± 0.2	4.8 ± 0.8	(2.3 ± 0.4) x 10^2
	β-methyl-D-glucopyranoside	0.33 ± 0.01	8.2 ± 1	36.6 ± 1.1
	1,5-anhydro-D-glucitol	0.4 ± 0.1	4.8 ± 0.4	79.2 ± 1.3
	α-(1,2)-D-glucose-D-glycerate	0.18 ± 0.02	0.53 ± 0.05	(1.9 ± 0.4) x 10^2
	3-keto-D-glucose	4.7 ± 0.3	0.5 ± 0.1	(8.5 ± 0.2) x 10^3
YcjQ	D-gulose	4.6 ± 0.2	3.6 ± 0.8	(1.3 ± 0.3) x 10^3
	β-methyl-D-gulopyranoside	1.5 ± 0.3	21.5 ± 4.5	71.6 ± 2.2

In an attempt to characterize the putative epimerase activity with YcjR, the enzyme was incubated with the substrates of the closest homologues. No new resonances were observed on the ^{1}H-NMR spectra when YcjR was incubated with D-psicose, D-fructose, D-tagatose, D-sorbose, L-xylulose, or L-ribulose. The assignment of the multiplets observed on ^{1}H-NMR spectrum with α-methyl-D-glucoside (**Figure 3.S5**) was made based on an article by Ceborska *et. al*, 2017. The ^{1}H-NMR spectra when α-methyl-D-glucoside was incubated with YcjS had new resonances near H4 as well as what appeared to be an exchange of the acidic H2 with the

solvent. This was reflected at H1 which was observed mostly as a singlet having lost its coupling to H2. A number of new resonances were also observed near 3.64 ppm. The ^{1}H-NMR spectra after addition of YcjR, showed that the H4 has been exchanged with deuterium. New resonances were observed in the region between 3.7 ppm and 3.8 ppm. Based on our results, YcjR appears to show activity as an epimerase potentially at H4 of α-methyl-D-glucoside. More thorough experiments need to be conducted to verify this.

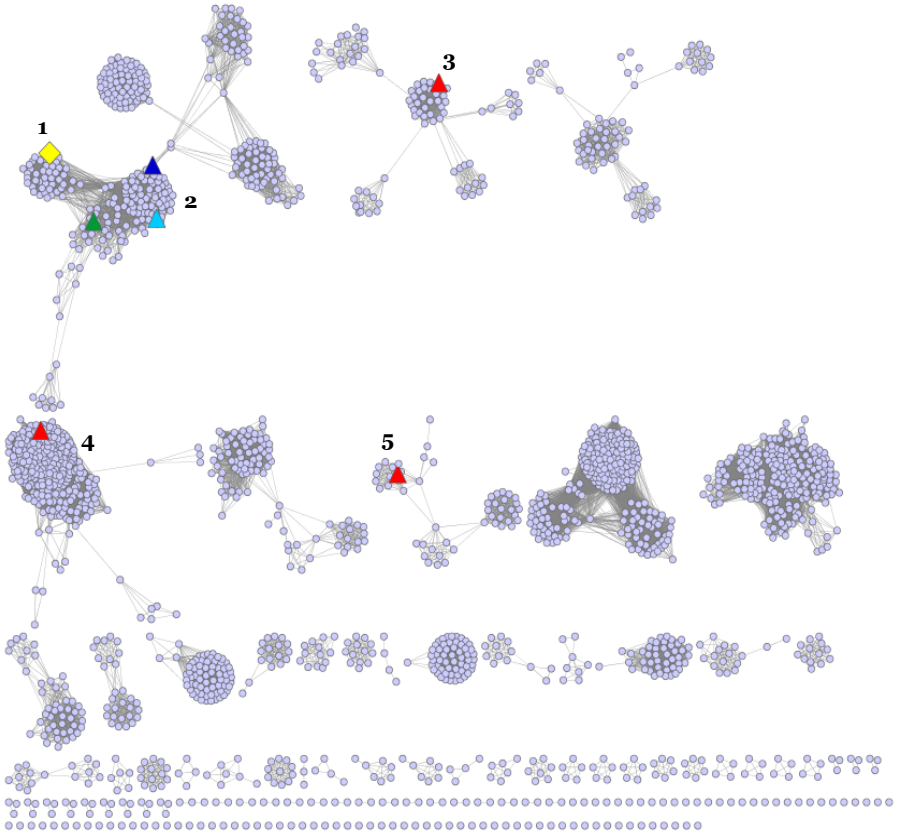


Figure 3.8: Sequence similarity network of cog1082 at an *E*-value cut-off of 1 x 10^{-25}. YcjR is shown as a diamond (yellow) in group 1. Groups 2, 3, 4, and 5 have experimentally verified functions shown as triangles. Group 2 has the three closest homologues; D-tagatose-3-epimerase from *Rhodobacter sphaeroides* (green), D-psicose-3-epimerase from *Agrobacterium tumefaciens* (cyan), and L-ribulose-3-epimerase from *Rhizobium loti* (blue). Functions in Groups 3, 4, and 5 are inosose isomerase from *Bacillus subtilis* (red), inosose dehydratase from *Bacillus subtilis* (pink), and xylose isomerase from *Plantomyces limnophilus* (purple) respectively.

Crystallization and Structure determination of YcjR. YcjR was crystallized in 0.1 M sodium malonate, pH 5.0, and 12% w/v PED 3350. The crystals diffracted to 2.2 Å at the local x-ray source. Due to the lack of a suitable homologous structural model, the selenomethionine derivative of YcjR was constructed, purified and subsequently crystallized under the same conditions. The YcjR crystals belonged to space group $P4_12_12$ with four molecules per asymmetric unit. It diffracted to a resolution of 1.86 Å and refined to a R_{work}/R_{free} of 0.1682/0.1937. The unit cell had the following dimensions: a = 115. 85 Å, b = 115.85 Å, c = 247. 3 Å, alpha = 90.00, and beta = 90.00. The single data set from peak wavelength contained sufficient anomalous information to solve the phases and the entire amino acid sequence of YcjR was built into the resulting electron density.

Overall structure of YcjR. The structure of YcjR was determined using single wavelength anomalous diffraction and was refined to 1.86 Å. Gel filtration results indicated that YcjR elutes as a tetramer at pH 8, under low salt (50 mM KCl) and high salt (0.5 M KCl) conditions. This observation agrees with the crystal structure of YcjR (**Figure 3.9A**) which shows a dimer of dimers. The dimeric interface between chains A and B or C and D has a buried surface area of 1071 $Å^2$. Each monomer adopts a modified $(\alpha/\beta)_8$ barrel fold. There are six β-strands surrounded by ten α-helices (**Figure 3.9B**). The solvent accessible monomeric surface area is 11,994 $Å^2$ and the solvation free energy of folding (ΔG) is -243.7 kcal/ mol. The values were calculated using the PDBePISA program (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver). Subunits A and B form a dimer with a two-fold symmetry. The two monomers come together with the active sites facing each other. The distance between the two active sites, as calculated from the spacing between the two metal ions, is 25 Å. The dimeric interface between chains A and B consists of 28 residues that are mostly present in the loop regions. The

other dimeric interface between chains A and C or B and D has a buried surface area of 386 $Å^2$. Each monomer contributes the following residues to the dimeric interface: Phe-69, Ile-70, Glu-71, Glu-72, Arg-73, Pro-115, and Ser-117. These residues are present at the ends of α4 and α5.

Active site. Multiple sequence alignment with the three homologues, D-tagatose-3-epimerase, D-psicose-3-epimerase, and L-ribulose-3-epimerase helped to identify the potential residues responsible for interacting with the metal ion and substrate as well as the catalytic residues. The bound Zn^{2+} in the active site (**Figure 3.10**) of YcjR is coordinated to the side-chains of Glu-146, His-205, Glu-240, and Asp-179. The interatomic distance between these four residues and the bound metal is 2.0 Å.

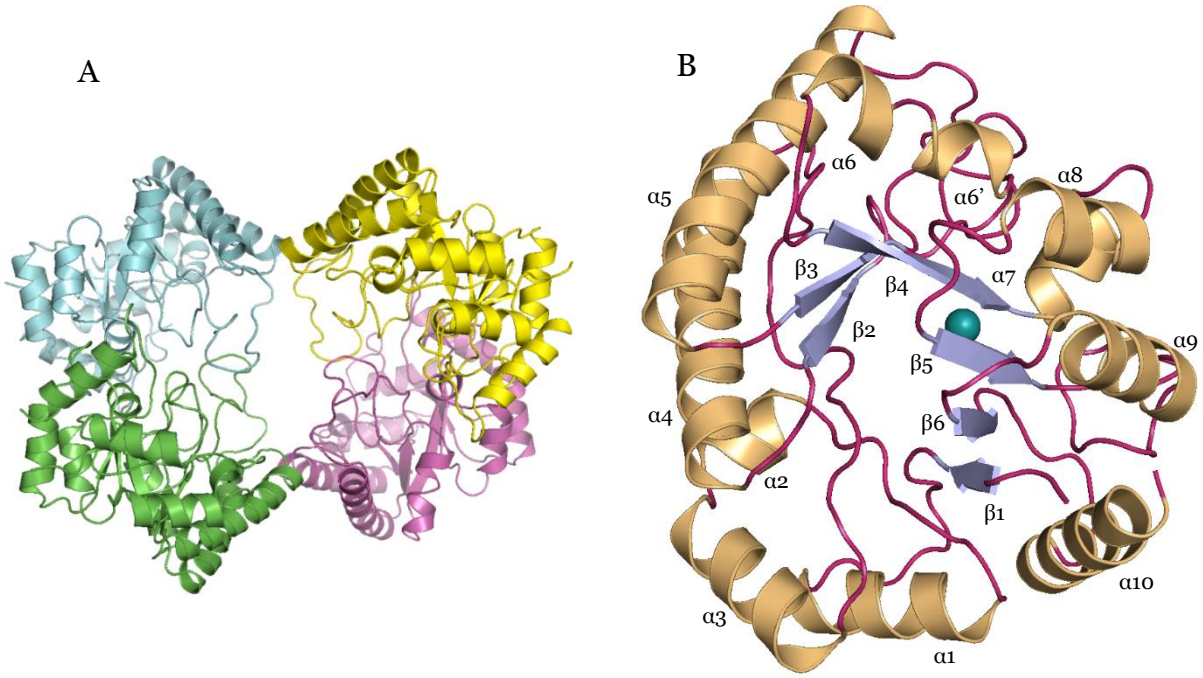


Figure 3.9: Overall structures of YcjR. (A) Tetrameric structure of YcjR showing the four monomers, (B) Ribbon representation of the YcjR monomer which adopts a modified TIM-barrel fold. There are six β-strands (purple) surrounded by ten α-helices (yellow). The bound Zn^{2+} is shown in cyan. The figure was prepared with Pymol.

These residues are highly conserved among the close homologues of YcjR. Attempts to co-crystallize YcjR with D-fructose or 3-keto-D-glucose were unsuccessful. Autodock Vina was used to dock 3-keto D-glucose into the active site (**Figure 3.S5A**).[188] The O3 of 3-keto-D-glucose is 3.8 Å from the bound metal while the hydrogen at C4 of the docked substrate is 3.1 Å away from Glu-146 and Glu-240, which are most likely the catalytic residues based on the multiple sequence alignment as well as their distance from the metal ion and the docked 3-keto D-glucose. The other substrate-binding residues in the active site that are conserved among D-psicose-3-epimerase, D-tagatose-3-epimerase, L-ribulose-3-epimerase and YcjR include His-182 (2.9 Å away from O3), and Arg-211 (2.6 Å from O3). YcjR has a glutamine (Gln-152) instead of a conserved glutamate, which is present in the other homologues. This Gln-152 is 3.6 Å from O4 and interacts with Trp-102 (4.0 Å), another conserved residue (**Figure 3.S5B**).

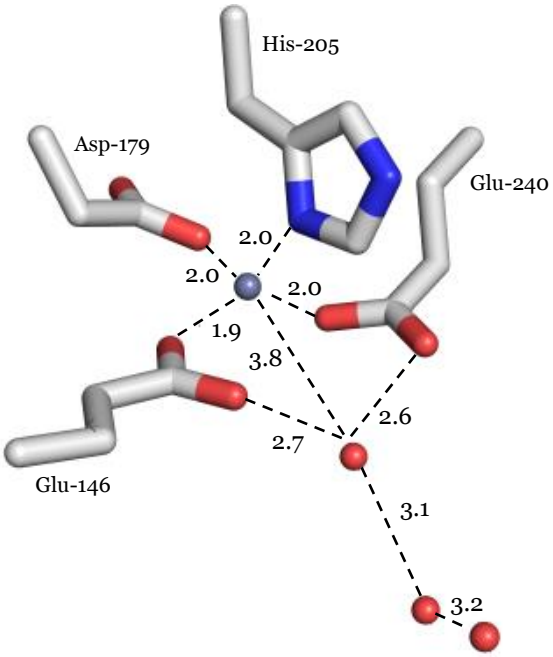
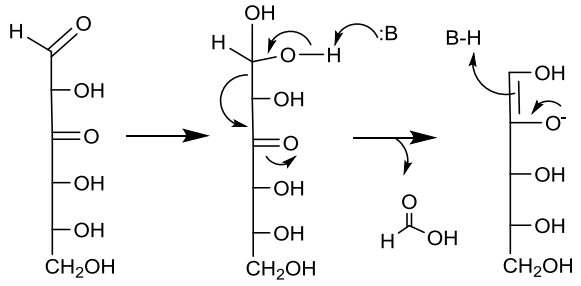


Figure 3.10: Coordination of Zn^{2+} with the active site residues and water molecules. The bound Zn^{2+} is shown as a blue sphere and the water molecules are shown as red spheres.

Discussion

Functional Characterization of YcjS, a D-Glucose-3-Dehydrogenase. YcjS was shown to oxidize D-glucose with a k_{cat} = 7.8 s^{-1}, K_m = 1.5 mM, and k_{cat}/K_m = 5.4 x 10^3 M^{-1} s^{-1} and reduce 3-keto-D-glucose with a k_{cat} = 4.7 s^{-1}, K_m = 0.5 mM, and k_{cat}/K_m = 8.5 x 10^3 M^{-1} s^{-1}. Formation of NAD(^{2}H) from D-[3-H^2]-glucose indicated that it is the hydroxyl group at C3 that gets oxidized to form 3-keto-D-glucose. This was further supported by ^{13}C-NMR experiments with α/β-methyl-D-[3-^{13}C] glucoside which shows a resonance at 207.25 ppm, corresponding to the 3-keto product. In the presence of equivalent concentrations of [1-^{13}C]-D-glucose and [3-^{13}C]-D-glucose at pH 6.5, four new resonances were identified on the ^{13}C-NMR spectra. These resonances corresponded to C-1 and C-3 of the α-furanose and β-pyranose anomers of the chemically synthesized 3-keto-D-glucose. At high pH values, 3-keto-D-glucose was found to be unstable. This was verified by the formation of formate from [UL-$^{13}C_6$]-D-glucose as well as D-glucose containing ^{13}C-label at C-1, C-2, C-3, C-4, C-5, and C-6 positions. The proposed mechanism for the formation of formate from D-glucose is shown in **Scheme 3.3**.



Scheme 3.3: Proposed mechanism for the formation of formate from 3-keto-D-glucose.

YcjS belongs to cog0673 and shares 28% identity (over 68% sequence coverage) with UDP *N*-acetyl-D-glucosamine-3-dehydrogenase and 30% identity (over 35% sequence coverage) with UDP-*N*-acetyl-D-glucosaminuronic acid dehydrogenase (PDB id: 3OA2).[180, 189] These dehydrogenases belong to the glucose-fructose oxidoreductase (GFOR) family.[190] An InterPro search of the YcjS sequence shows a predicted Rossman fold domain, in the *N*-terminal (5 – 158 amino acids) part of the protein, which is responsible for binding to NAD. A sequence alignment of UDP-*N*-acetyl-D-glucosaminuronic acid dehydrogenase from *P. aeruginosa* (PDB id: 3OA2) with YcjS showed that amino acid residues, in the enzyme from *P. aeruginosa,* that bind to UDP (Tyr-12, Lys-169, Arg-160, Thr-243, and Arg-245), the *N*-acetyl group (Tyr-156 and Arg-245), or the C-6' carboxylate group of the hexose (Arg-160, Tyr-164, Lys-169, and Asn-181) are not conserved in YcjS.[180] The hydroxyl group at C3' that gets oxidized, in UDP-*N*-acetyl-D-glucosaminuronic acid dehydrogenase, is 3.0 Å away from C4 of the nicotinamide ring of NAD and hydrogen bonds with Lys-101 and His-185.[180] Both of these residues are conserved in YcjS (Lys-103 and His-189). The lysine residue is part of the characteristic consensus sequence (Glu-Lys-Pro) found in all members of the GFOR family, including YcjS. This motif interacts with the nicotinamide ring of NAD.[191] A second predicted consensus sequence is found in some GFOR family members: Gly-Gly-X_3-Asp-X_3-(Tyr/His), which is present in the active site and has been proposed to bind and orient the substrate for catalysis.[192] In the KijD10 structure (PDB id: 3RC1), Asp-182 from the conserved motif sequence, interacts with the C2 hydroxyl of the nicotinamide ribose and was suggested to have a role in binding with C4" of the bound hexose substrate based on modeling of dTDP-3-keto-6-deoxy-D-galactose in the active site.[193] YcjS has this consensus sequence as well (Gly 180-Gly-Pro-Leu-Ile-Asp-Ile-Gly-Ile-His 189).

The NAD-dependent YcjS was shown to oxidize the hydroxyl group at C3 of D-glucose. In UDP *N*-acetyl-D-glucosamine dehydrogenase, the authors were unable to isolate the enzymatic product (3-keto-UDP-GlcNAc) using HPLC. Instead they detected UDP which was proposed to form because of the inherent instability of the 3-keto product.[189] This observation agrees with our results where the labile 3-keto product leads to the formation of formate under alkaline conditions, followed by further degradation to smaller products. In the study reporting the role of NtdC in the biosynthesis of kanosamine, the authors acknowledge the instability of the enzymatic product, 3-keto-D-glucose-6-phosphate. They reported that 3-keto-D-glucose-6-phosphate exhibits an absorbance at ~300 nm owing to the formation of an enolate and/or enone form of the product. The authors successfully synthesized 3-keto-D-glucose, which at pH 8.0 showed an absorbance at 310 nm.[175] According to our studies, enzymatic reactions with YcjS and D-glucose showed absorbance peaks at 310 nm. Higher the pH value of the enzymatic reaction, more was the change (increase) in absorbance at 310 nm.

Functional Characterization of YcjQ, a D-Gulose-3-Dehydrogenase. YcjQ belongs to cog1063 and the SSN constructed for this cog did not show any functionally verified homologs close to YcjQ. Testing YcjQ against a library of different sugars showed catalytic activity only with D-gulose. Based on our current knowledge, this is the first reported enzyme that oxidizes D-gulose, a very rare hexose. From the degradation pattern of the enzymatic product and the formation of formate, as observed in the ^{13}C-NMR spectra, it is proposed that the product of the YcjQ and D-gulose reaction is 3-keto-D-gulose. Attempts to synthesize the 3-keto product were unsuccessful but reactions with YcjQ and D-gulose at high pH values showed an absorbance at 310 nm, suggesting that the enzymatic product is unstable, much like 3-keto-D-glucose.

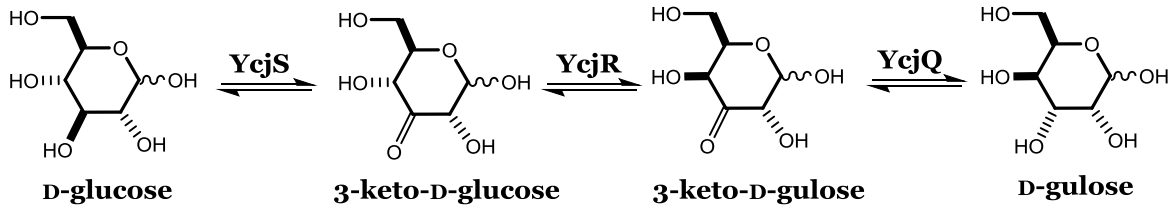
Attempts at Functional Characterization of YcjR. The SSN for cog1082 reveals that the three closest homologues to YcjR are annotated as D-tagatose-3-epimerase, D-psicose-3-epimerase, or L-ribulose-3-epimerase. YcjR shares a sequence identity of ≤30% with the three enzymes. A comparison of the three dimensional structure of YcjR with these homologues (PDB ids: 2HK1, 2QUM, and 3VYL) reveals that while all the three homologues have 8 β-strands, YcjR only has 6.[187, 194, 195] The dimeric interface between two chains has a buried surface area of 1071 $Å^2$. This area is smaller as compared to the buried surface area between the dimers of the closest homologues: D-psicose-3-epimerase (1,353 $Å^2$) from *Agrobacterium tumefaciens* (PDB: 2HK1) and D-tagatose-3-epimerase (1,479 $Å^2$) from *Pseudomonas cichorii* (PDB: 2QUM) as calculated on the PDBePISA website.

All three homologues epimerize the stereochemistry of the hydroxyl group attached to C3 of their respective substrates, which contain a keto-group at the C2 position. Like these homologues, YcjR is a manganese-dependent enzyme as evident by our metal-content studies; however the selenomethionine-derivatized YcjR crystallized with zinc bound in the active site. This can be attributed to the addition of $MnCl_2$ right before induction with IPTG instead of during the inoculation of the cultures. The conserved active site residues, Glu-146, Asp-179, His-205, and Glu-240 are 2.0 Å away from the bound Zn^{2+}. The two glutamate residues are critical for catalysis. In D-psicose-3-epimerase, proton abstraction from C3 of D-fructose by Glu-244 is followed by formation of an ene-diolate intermediate. The negative charge is stabilized by the bound metal ion as well as by the side chain of the conserved Arg-215. Reprotonation is carried out by Glu-150 from the opposite side of the ene-diolate intermediate to form D-psicose. In the active site of YcjR, O3 of the docked 3-keto D-glucose is 3.8 Å away from the metal ion. The other conserved residues that stabilize the substrate during catalysis, in the three homologues, are

also found in YcjR (Trp-102, Arg-211, and His-182), with the exception of having a glutamine (Gln-152) instead of the conserved glutamate in that position.

From our experiments, it can be concluded that YcjS and YcjQ catalyze the oxidation/reduction at the C3 of D-glucose and D-gulose respectively. A number of previous studies as well as our experiments indicate that the 3-keto glycosides are inherently unstable.[175, 179, 189, 196] But the formation of keto sugars is not entirely uncommon.[197-200] Coupling enzymes that catalyze the formation of 3-keto glycosides with the next enzyme in the pathway is an efficient way to transfer the unstable ketone intermediate without allowing it to accumulate.[175, 179]

Based on our analysis of YcjR and the mode of action of its closest relatives, it can be suggested that YcjR most likely epimerizes at the C4 position of 3-keto-D-glucose to form 3-keto-D-gulose and *vice versa*. This provides a potential pathway for the conversion of D-glucose to D-gulose or the other way around (**Scheme 3.4**). Very little is found in literature about the occurrence of D-gulose in nature. 6-deoxy-D-gulose is found in the lipopolysaccharide *O*-antigen of *Yersinia enterocolitica* while D-gulose has been reported to be part of the extracellular polysaccharide of *Caulobacter crescentus*, in the extracellular matrix of *Volvox carteri*, an algae, and is also present in the exoglycolipid of an antigen derived from *Chlamydia trachomatis*.[201-204]



Scheme 3.4: Proposed pathway for the transformation between D-glucose and D-gulose.

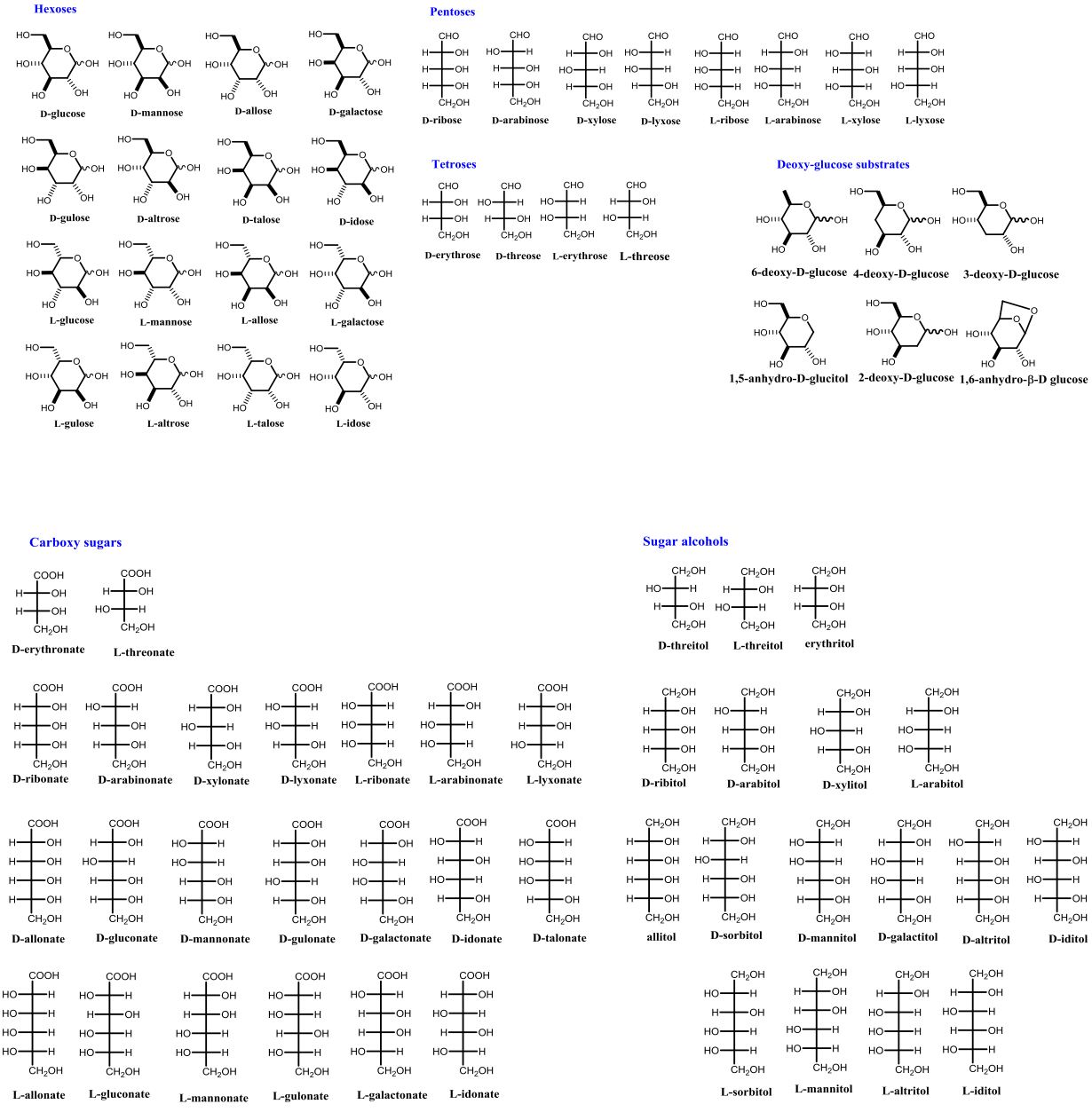
Other dehydrogenases responsible for the formation of 3-keto-D-glucosides. 3-keto glucosides have important applications as food additives, antioxidants, and aminoglycoside antibiotics.[205, 206] An FAD-dependent glucoside-3-dehydrogenase that oxidizes the C3 hydroxyl group of hexoses and its derivatives was first reported in *Agrobacterium tumefaciens* and has since been identified in *Halomonas (Deleya)* sp. α-15, *Agaricus bisporus*, *Flavobacterium saccharophilum*, *Cytophaga marinoflava*, *Stenotrophomonas maltophilia*, and *Sphingobacterium faecium.*[207-213] Studies conducted on *Agrobacterium tumefaciens* showed that the FAD-dependent glucoside-3-dehydrogenase, most likely present in the periplasmic space of the bacterium, converts disaccharides, such as sucrose to 3-keto sucrose, which is then transported inside the cell[214-217]. The cytoplasmic enzyme, 3-ketoglucosidase, catalyzes the hydrolysis of 3-keto sucrose to form fructose and 3-keto-D-glucose, which in turn gets reduced to glucose by the NADP-dependent 3-ketoglucose reductase.[218, 219] Ampomah *et. al*, reported that the *thuEFGKAB* operon found in *A. tumefaciens* was responsible for the transport of various disaccharides in their 3-keto forms.[220] Disruption of *thuAB* genes affected the bacterium's ability to utilize trehalose as a carbon source, but the mutant was able to grow on other disaccharides such as maltose, sucrose, cellobiose, and lactose, indicating that the bacteria has multiple pathways dedicated to the transport and utilization of disaccharides. However, the function of ThuAB, from the putative trehalose-utilization pathway, has not been experimentally verified.[220] While ThuB belongs to cog0673, the other FAD-dependent enzymes from *Halomonas* sp. α-15 and *Agaricus bisporus* belong to cog2303.

There are a number of NAD-dependent glucoside-3-dehydrogenase enzymes that catalyze the conversion of hexoses to their 3-keto derivative. Some of the enzymes that convert sugar-nucleotides to 3-hexulose sugar nucleotides include UDP N-acetyl-D-glucosamine

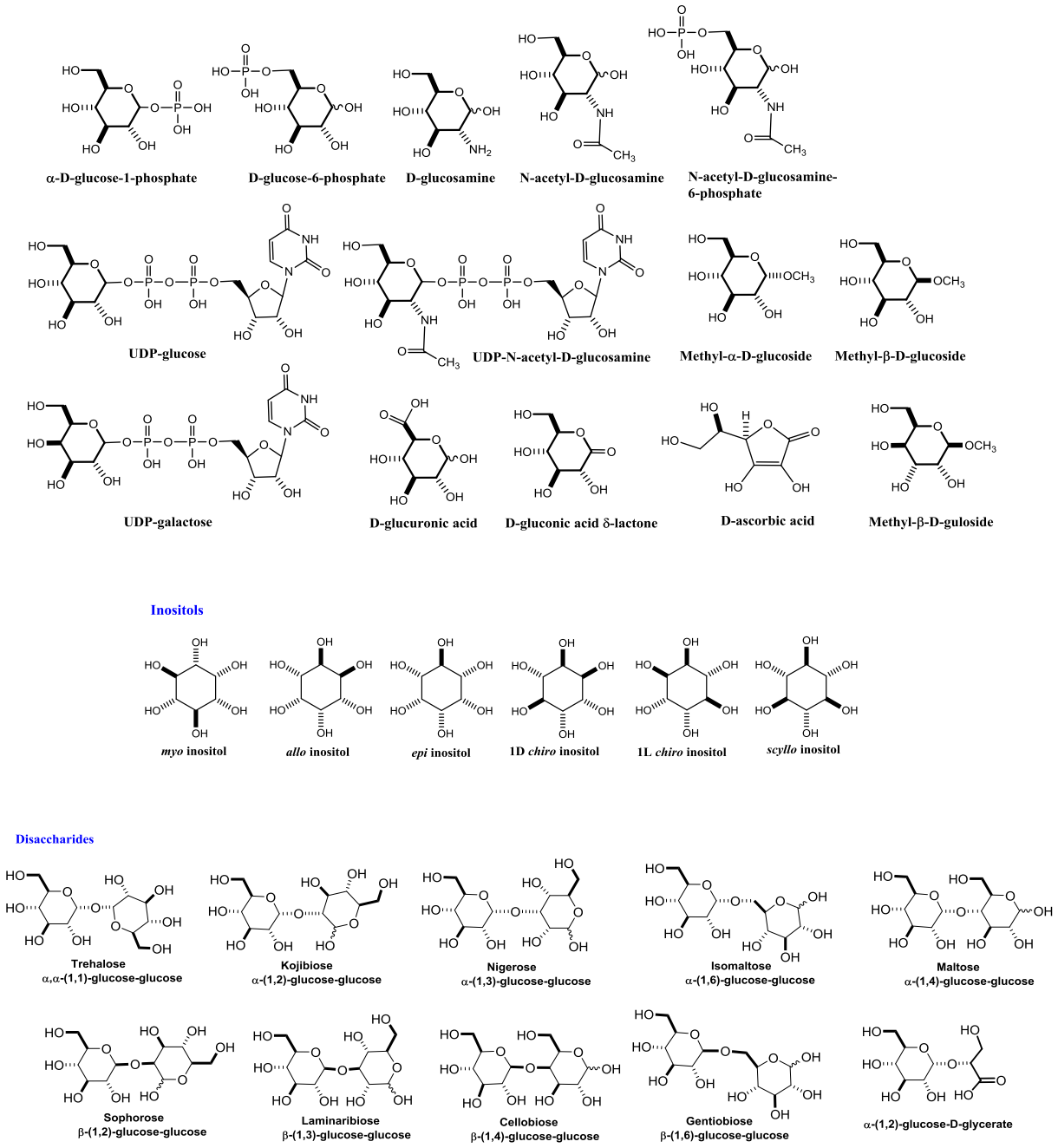
dehydrogenase, and UDP-*N*-acetyl-D-glucosaminuronic acid dehydrogenases, from *Pseudomonas aeruginosa* PAO1 and *Thermus thermophiles*.[180, 189] UDP-*N*-acetyl-D-glucosaminuronic acid dehydrogenases were purified with NAD bound to the enzymes.[180, 189] Levoglucosan (1,6-anhydro-β-D-glucopyranose) dehydrogenase from *Arthrobacter* sp. I-552 that can oxidize levoglucosan, a product of burnt biomass, to 3-keto levoglucosan.[221] In the kanosamine biosynthetic pathway, NtdC, a sugar dehydrogenase, from *Bacillus subtilis* catalyzes the oxidation of the hydroxyl group at C-3 of D-glucose-6-phosphate.[175] KijD10, a C-3" ketoreductase from *Actinomadura kijaniata* is a NADPH dependent enzyme catalyzes the reduction of dTDP-3,4-diketo-2,6-dideoxy-D-glucose to dTDP-4-keto-2,6-dideoxy-D-glucose.[193] Along with ThuB, UDP-*N*-acetyl-D-glucosaminuronic acid dehydrogenase, NtdC, KijD10, and UDP *N*-acetyl-D-glucosamine dehydrogenase belong to cog0673 and are shown in **Figure 3.2**.

So far the *ycj* gene cluster seems to be responsible for the degradation of kojibiose, a rare sugar, to form D-glucose-6-phosphate, that most likely enters glycolysis, and D-glucose which gets converted to D-gulose, another rare sugar.[171] We have not found any evidence of occurrence of D-gulose in *Escherichia coli* K-12 MG1655 and therefore, the purpose of the formation of this hexose via the activity of these enzymes remains unknown. How the function of YcjM, α-D-glucosyl-2-glycerate phosphorylase, is related to the functions of the other Ycj enzymes remains to be determined. Our earlier efforts to grow *E. coli* on kojibiose were unsuccessful and we suggested that it was because kojibiose was most likely not the inducer for the *ycj* gene cluster.[171] In the future, it may be worthwhile to explore if either the substrate (α-(1,2)-D-glucose-D-glycerate) or the products (D-glycerate and α-D-glucose-1-phosphate) of YcjM have any roles in the induction of the gene cluster.

Supplementary Information



Scheme 3.S1: Structures of hexoses, pentoses, tetroses, sugar carboxylates, and sugar alcohols tested as substrates for YcjS and YcjQ.



Scheme 3.S2: Structures of glucose and gulose derivatives, inositols, and disaccharides tested with YcjS and YcjQ.

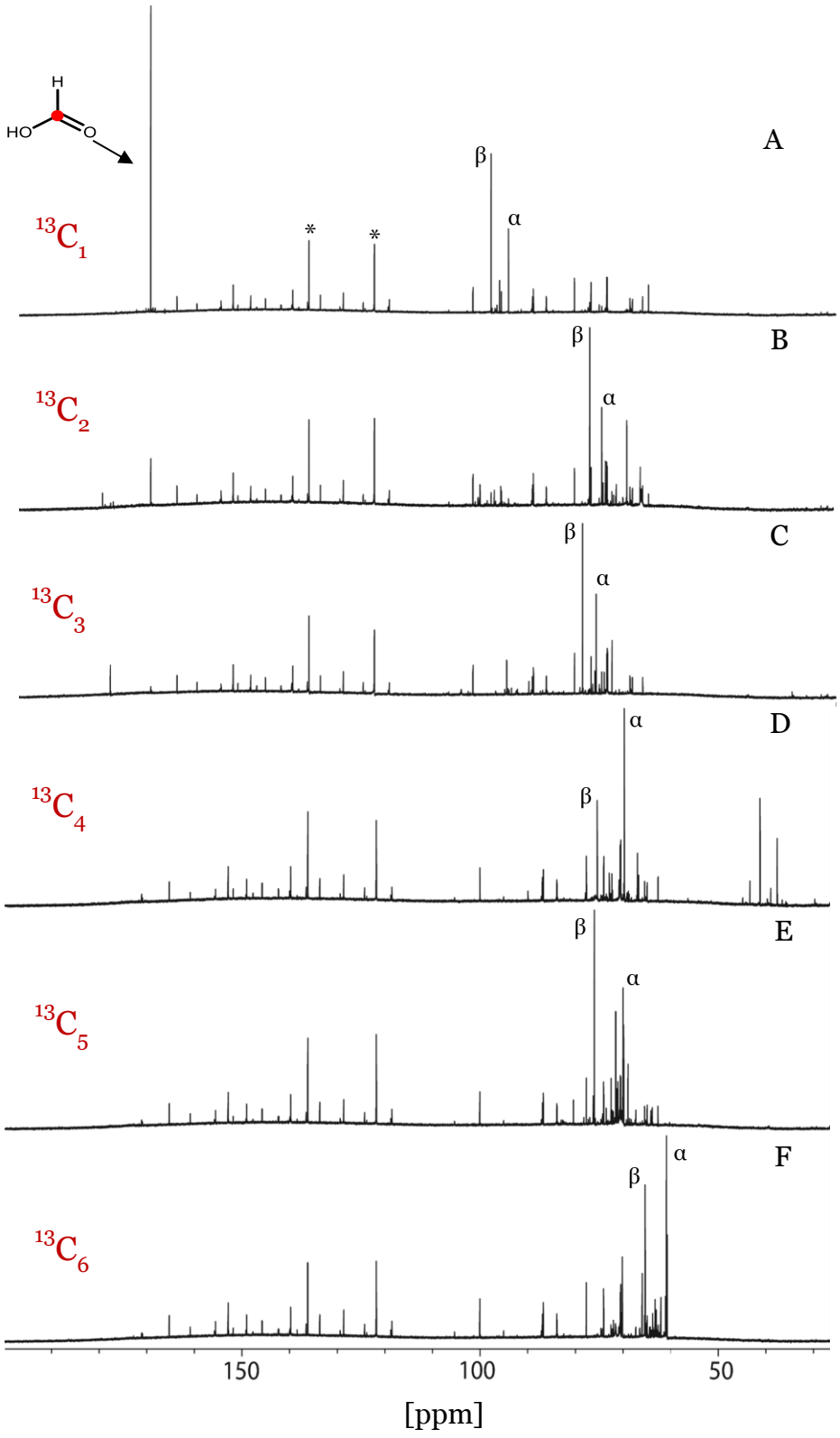


Figure 3.S1: ^{13}C-NMR spectra showing multiple resonances from enzymatic reactions with D-glucose containing ^{13}C-label at positions C1 (A), C2 (B), C3 (C), C4 (D), C5 (E) or C6 (F). * - represents imidazole peaks from protein purification.

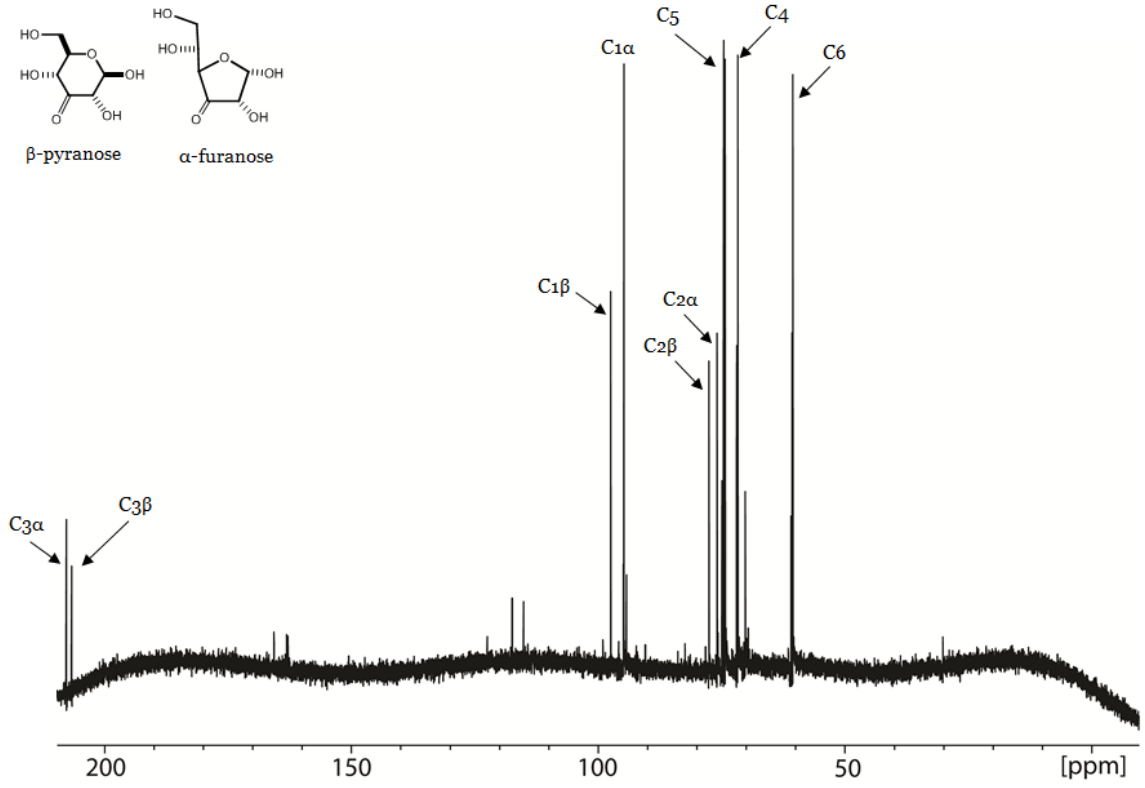


Figure 3.S2: ^{13}C-NMR of the chemically synthesized 3-keto-D-glucose.

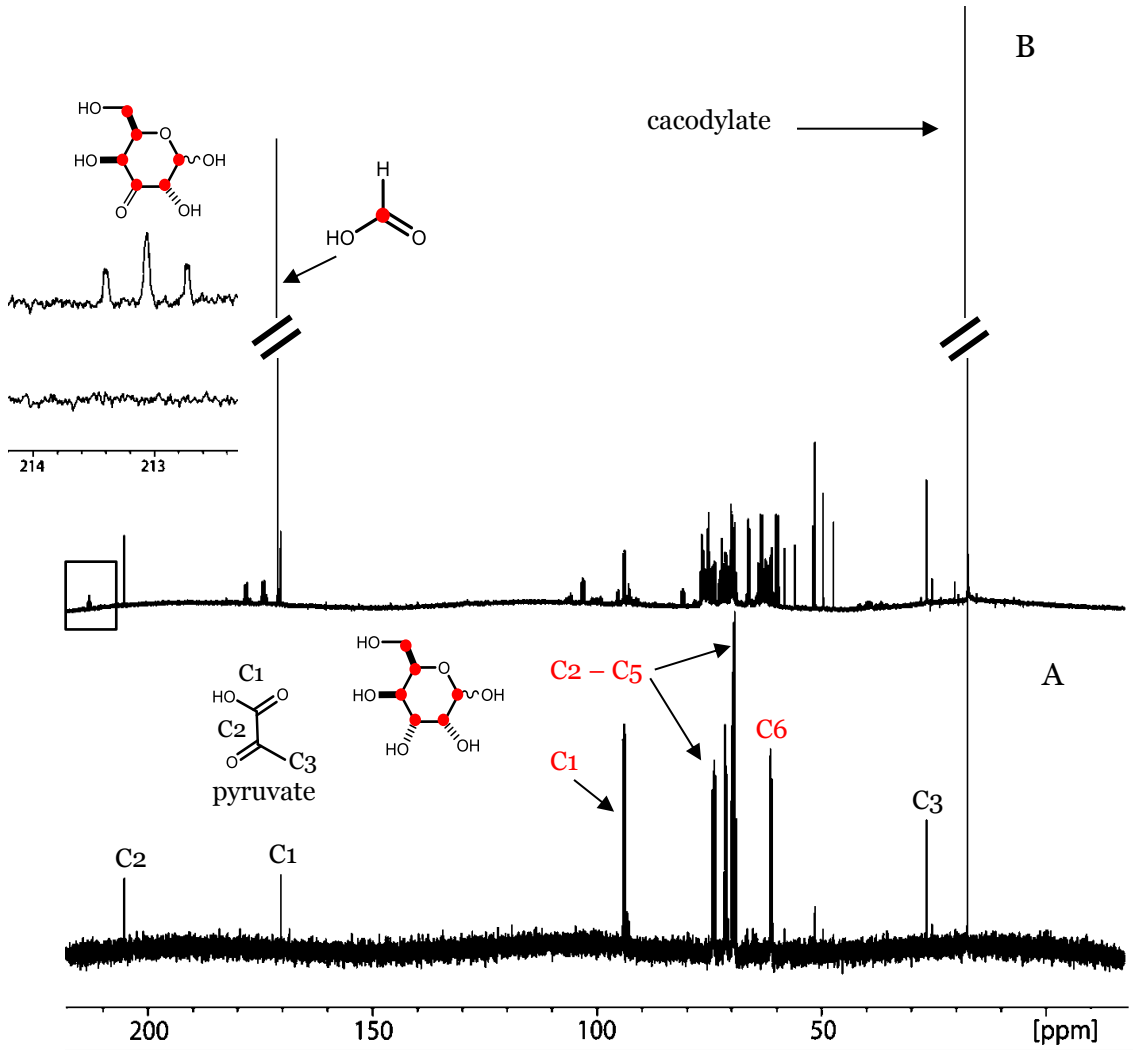


Figure 3.S3: ^{13}C-NMR spectra showing multiple resonances from enzymatic reaction with [UL-$^{13}C_6$]-D-gulose. (A) without the addition of YcjQ; (B) after incubation with YcjQ. The resonance for formate was observed at 171.03 ppm. A triplet is observed around 213 ppm region with *J*(CC) of 42 Hz.

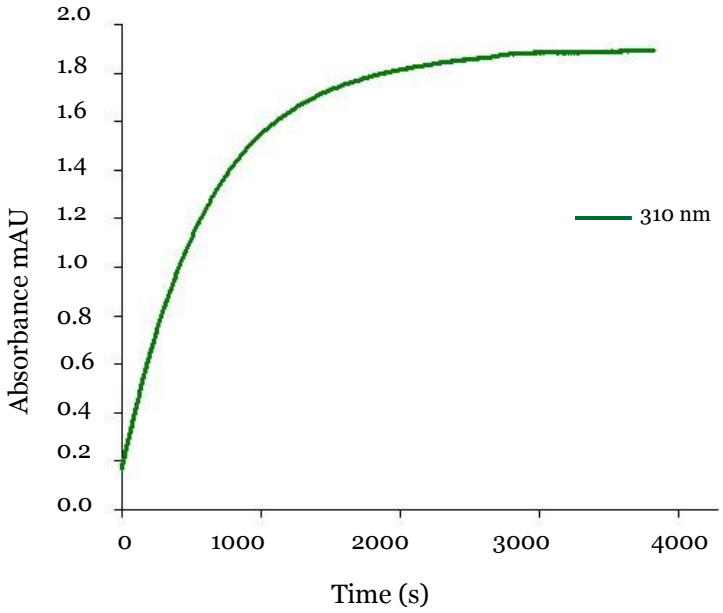


Figure 3.S4: YcjS catalyzed reaction with D-glucose in presence of pyruvate and lactate dehydrogenase at pH 9.5. The reaction had 0.5 mM NAD^+ and was monitored at a wavelength of 310 nm.

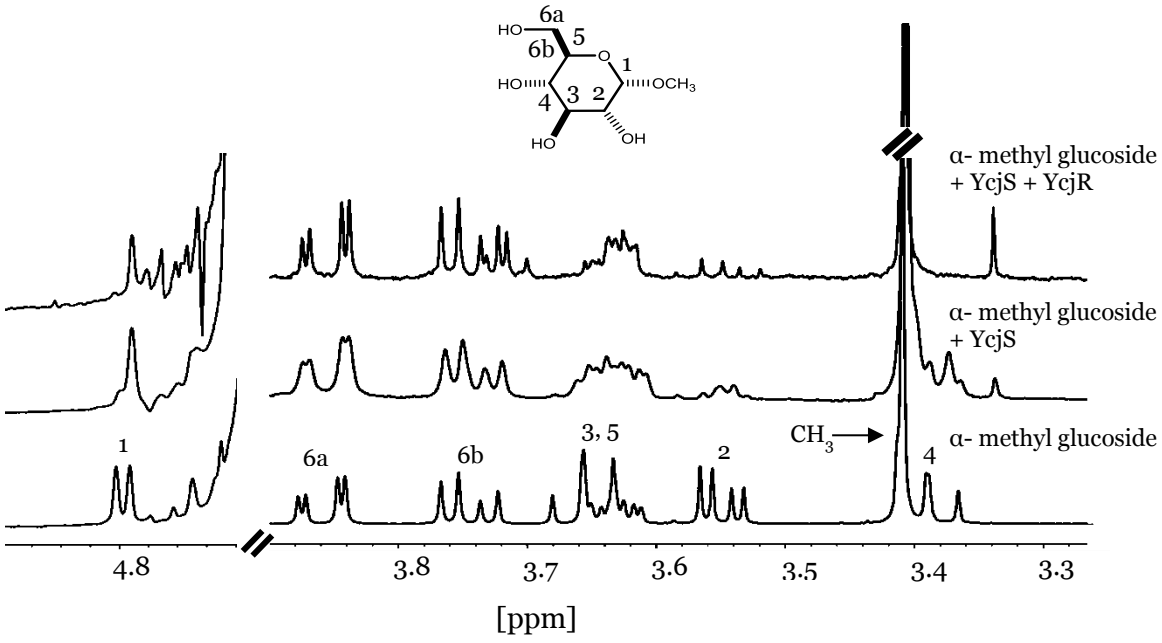


Figure 3.S5: ^{1}H-NMR spectra showing the enzymatic reactions with YcjS and YcjR in the presence of α-methyl-D-glucoside.

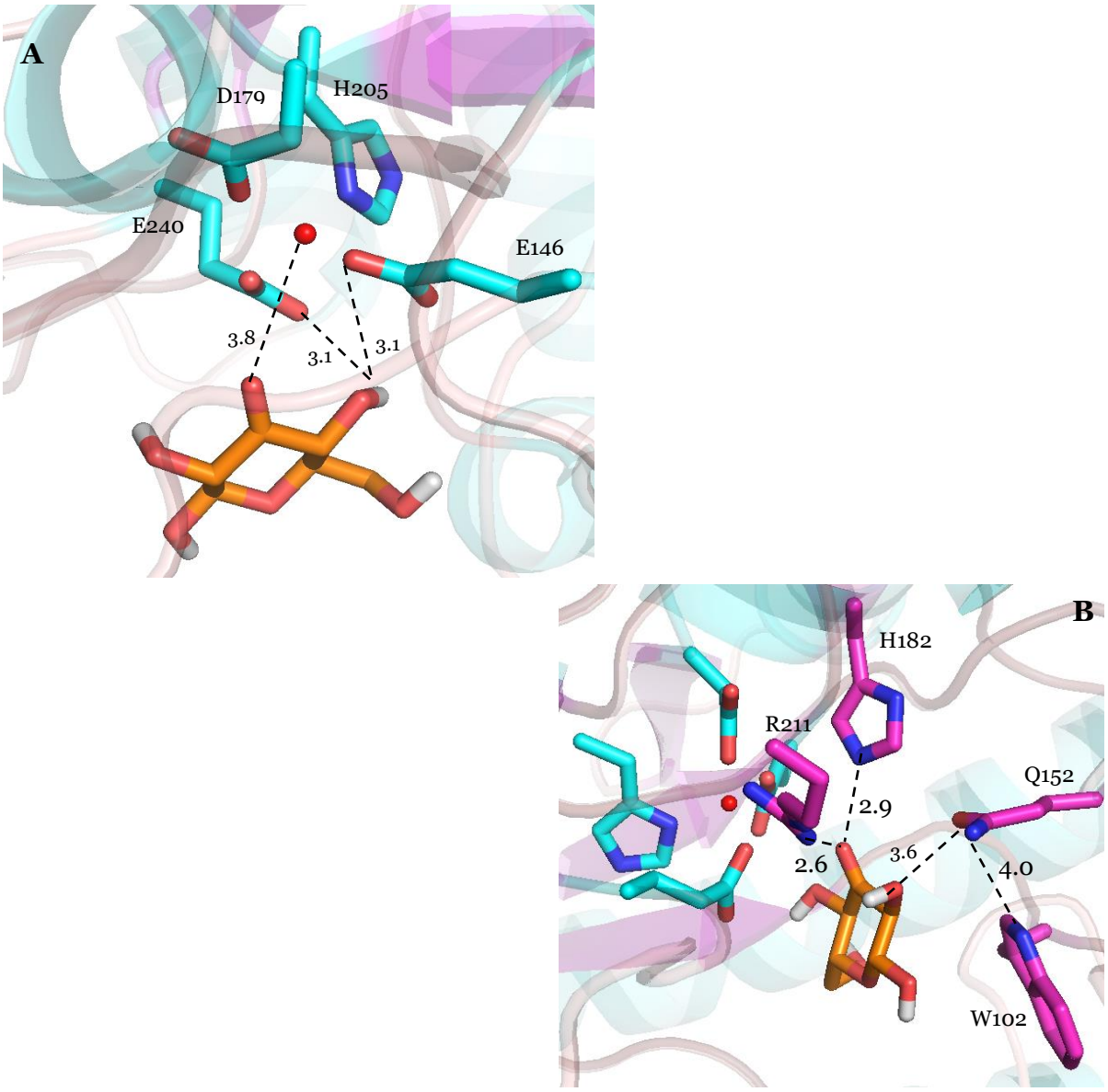


Figure 3.S6: Active site of YcjR showing the docked 3-keto D-glucose molecule. (A) The key active site residues are shown in cyan, the metal ion is shown in red. (B) Conserved substrate binding residues are indicated in pink. The interatomic distances (Å) between residues and the docked molecule are indicated by dashed lines.

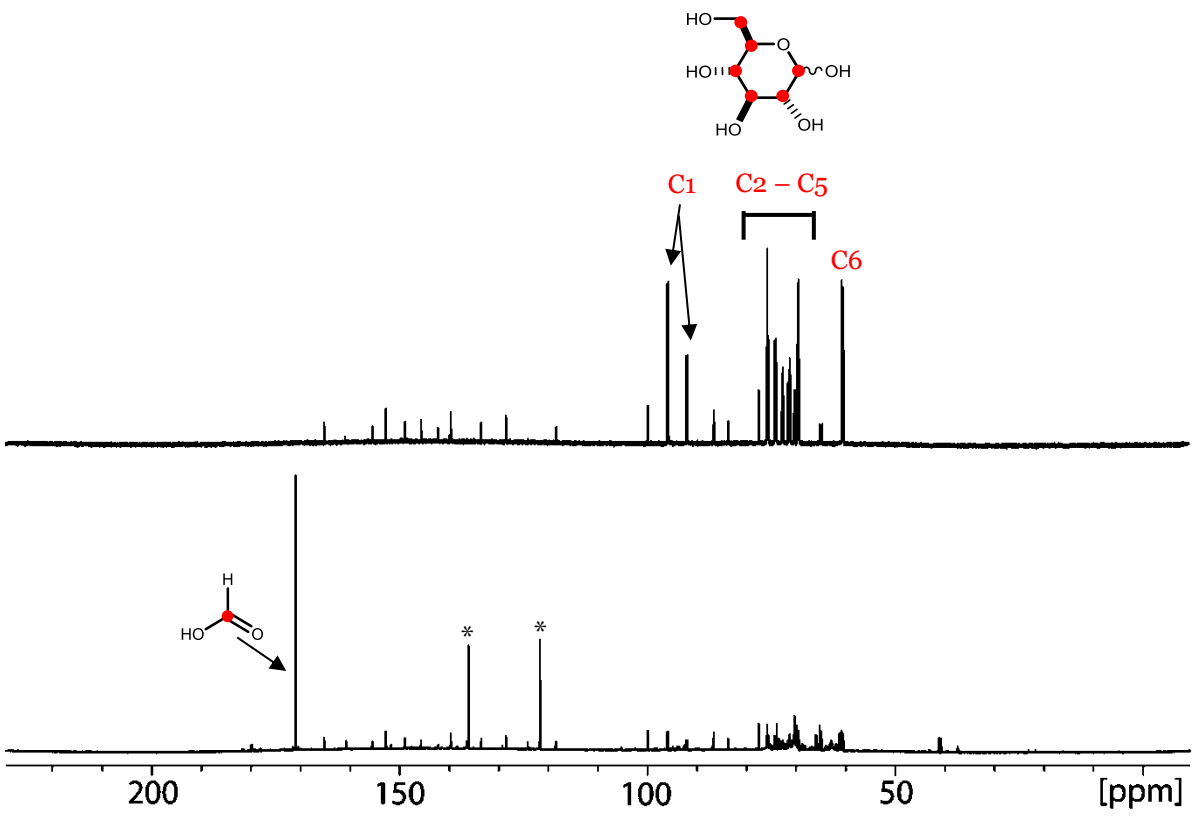


Figure 3.S7: ^{13}C-NMR spectra showing [UL-$^{13}C_6$]-D-glucose and NAD^+ with and without the addition of YcjS. * refers to the imidazole carried over from protein purification.

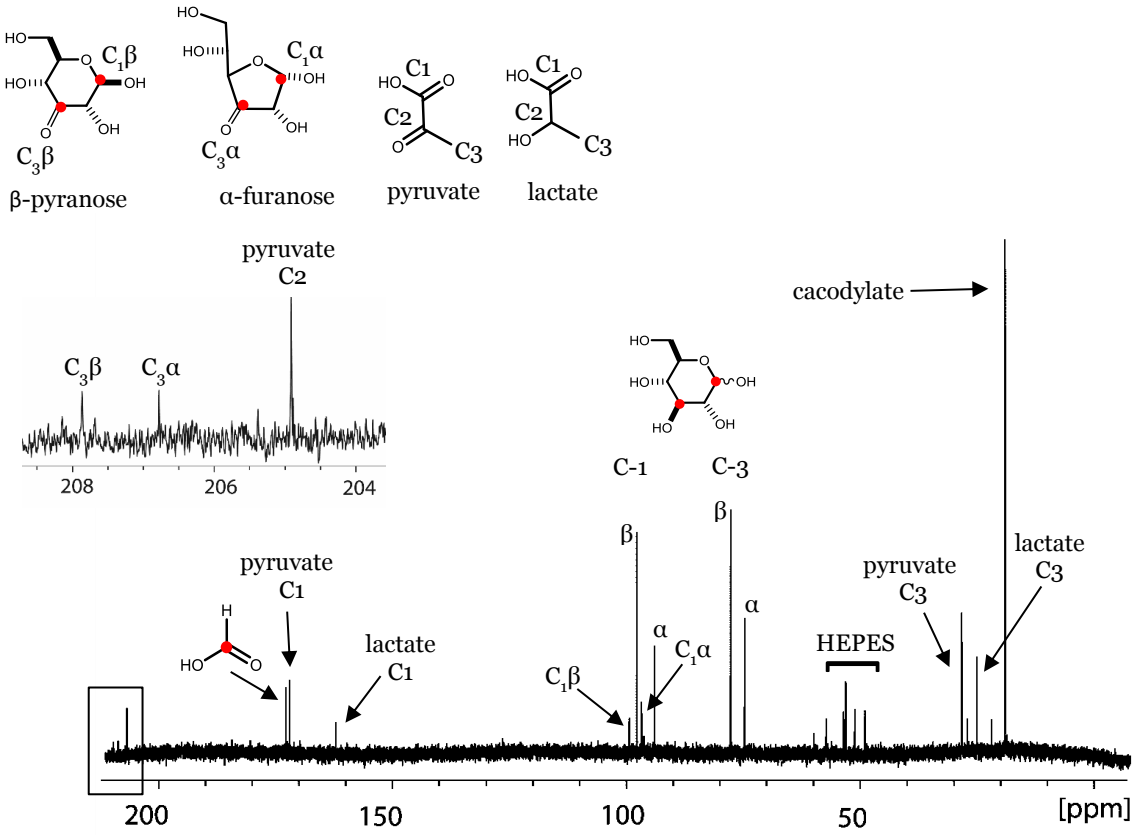


Figure 3.S8: ^{13}C-NMR spectra showing the enzymatic reaction with equivalent concentrations of [1-^{13}C]-D-glucose and [3-^{13}C]-D-glucose at pH 6.5 in the presence of pyruvate and lactate dehydrogenase coupling system.

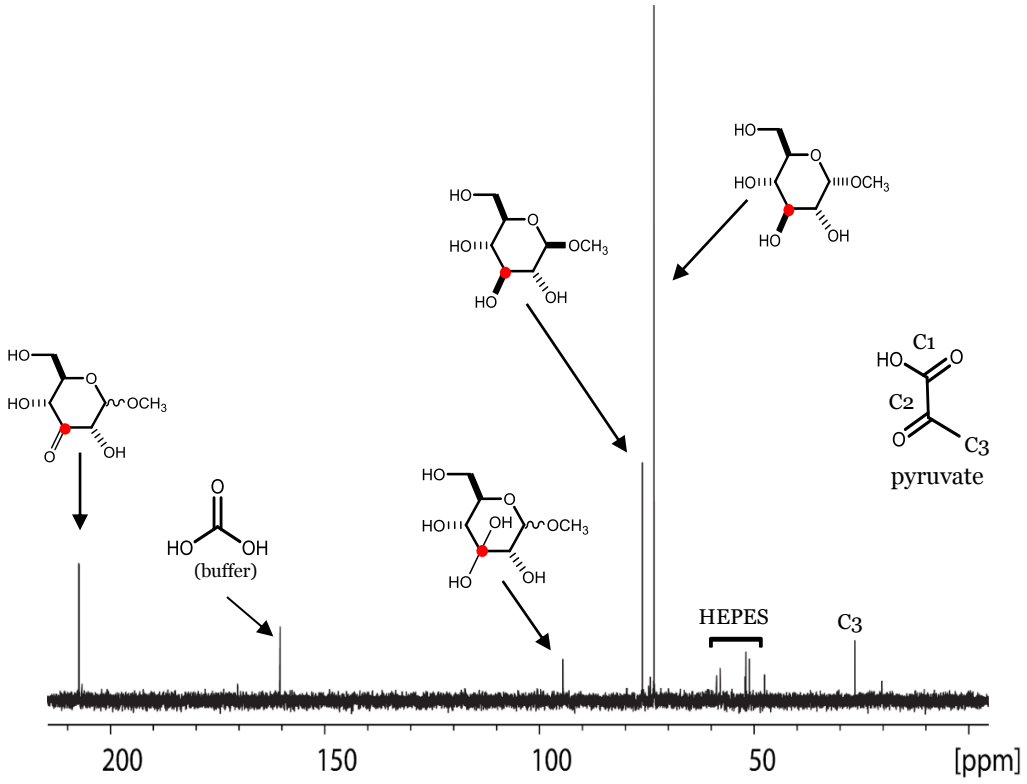


Figure 3.S9: ^{13}C-NMR spectra showing the YcjS catalyzed reaction with α/β methyl-D-[3-^{13}C] glucose.

CHAPTER IV

SUMMARY AND CONCLUSIONS

The gut microbiota plays an important role in the host nutrient metabolism and the introduction of this dissertation mentioned this role with an emphasis on the degradation of carbohydrates, primarily from plant sources. A number of novel enzymes, from different gut bacterial species, are responsible for the breakdown of these complex sugars, and more are frequently being discovered. However, enzyme function discovery has not kept pace with the amount of metagenomics data available today. Bridging this gap is crucial for understanding the complex array of functions available in the human gut. A number of public databases contain protein annotations that are computationally predicted and tend to suffer from misannotations. Misannotations can arise from predicting functions of proteins based on other misannotated proteins or by only considering sequence similarity or domains. Therefore, proper biochemical analyses of the functions of uncharacterized proteins are vital. This is especially true for large superfamilies of proteins that do not have many experimentally verified functions. Wrong annotations can often lead to error propagation as other websites, such as KEGG, pool predicted functions from various sources to give the reader information about a particular pathway. Finding relevant enzyme targets for function discovery can be equally challenging. The introduction chapter also, mentions ways to look for novel enzymatic pathways and some tools that an enzymologist can utilize to assemble substrates for screening against an uncharacterized enzyme.

Chapter II describes the discovery of three enzymes (YcjM, YcjT, and YcjU) from the *ycj* gene cluster in *E. coli* K-12. Based on the sequence similarity network (SSN) for YcjT

(cog1554), it was deduced that the substrate for the enzyme was a glucose-glucose disaccharide. By assembling a library of glucose-glucose, glucose-galactose, and glucose-fructose disaccharides and screening YcjT against it, the enzyme was identified as a kojibiose phosphorylase, the first to be discovered in *E. coli*. It was shown that YcjT could also use other substrates (D-sorbitol, 1,5-anhydro-D-glucitol, L-iditol, and L-sorbose) in the presence of β-D-glucose-1-phosphate to form the corresponding disaccharides. YcjU was determined to be a β-phosphoglucomutase and based on our knowledge, the only β-phosphoglucomutase reported in *E. coli*. YcjU also showed activity with β-D-galactose-1-phosphate, β- D-mannose-1-phosphate, and β-D-allose-1-phosphate. This activity has been not reported in the previously studied homologues of YcjU. The function of YcjM was determined to be α-D-glucosyl-2-glycerate phosphorylase based on bioinformatics analyses and steady state kinetics.

Chapter III summarizes our attempts to characterize YcjQ, YcjR, and YcjS from the *ycj* gene cluster. YcjS was determined to be a NAD-dependent glucose dehydrogenase. Our experiments showed that the product of the enzymatic oxidation is 3-keto-D-glucose, which was found to be extremely labile under alkaline conditions and prone to degradation. 3-keto-D-glucose was chemically synthesized and was shown to be a substrate for YcjS. YcjQ was determined to be a NAD-dependent gulose dehydrogenase, the first ever such reported enzyme based on our knowledge. The primary degradation product (formate) of the YcjQ catalyzed reaction was reminiscent of what was observed on ^{13}C-NMR spectra with YcjS and D-glucose at pH 9.5. The product of the YcjQ catalyzed oxidation of D-gulose was suggested to be 3-keto-D-gulose. Based on our experiments with YcjR and the reactions catalyzed by the closest homolog of YcjR, we propose that the enzyme is a C4-epimerase that requires a substrate with a keto group at the C3 position. Therefore, the proposed function of YcjR is to convert 3-keto-D-

glucose to 3-keto-D-gulose and *vice versa*. The structure of selenomethionine-derivatized YcjR was solved to 1.86 Å resolution. Since attempts at co-crystallization with potential substrates failed, 3-keto-D-glucose was docked into the active site of the YcjR structure.

REFERENCES

[1] Parfrey, L., Walters, W., and Knight, R. (2011) Microbial eukaryotes in the human microbiome: ecology, evolution, and future directions. *Front Microbiol 2*, 153.

[2] Sender, R., Fuchs, S., and Milo, R. (2016) Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol 14*, e1002533.

[3] Costello, E., Lauber, C., Hamady, M., Fierer, N., Gordon, J., and Knight, R. (2009) Bacterial community variation in human body habitats across space and time. *Science 326*, 1694-1697.

[4] Ley, R., Lozupone, C., Hamady, M., Knight, R., and Gordon, J. (2008) Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol 6*, 776-788.

[5] Aagaard, K., Ma, J., Antony, K., Ganu, R., Petrosino, J., and Versalovic, J. (2014) The placenta harbors a unique microbiome. *Sci Transl Med 6*, 237ra265.

[6] Dominguez-Bello, M., and Blaser, M. (2011) The Human Microbiota as a Marker for Migrations of Individuals and Populations. *Annu Rev of Anthropol 40*, 451-474.

[7] Spor, A., Koren, O., and Ley, R. (2011) Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol 9*, 279-290.

[8] Bajaj, J., Heuman, D., Hylemon, P., Sanyal, A., White, M., Monteith, P., Noble, N., Unser, A., Daita, K., Fisher, A., Sikaroodi, M., and Gillevet, P. (2014) Altered profile of human gut microbiome is associated with cirrhosis and its complications. *J Hepatol 60*, 940-947.

[9] Ley, R., Turnbaugh, P., Klein, S., and Gordon, J. (2006) Microbial ecology: human gut microbes associated with obesity. *Nature 444*, 1022-1023.

[10] Wen, L., Ley, R., Volchkov, P., Stranges, P., Avanesyan, L., Stonebraker, A., Hu, C., Wong, F., Szot, G., Bluestone, J., Gordon, J., and Chervonsky, A. (2008) Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature 455*, 1109-1113.

[11] Ling, Z., Li, Z., Liu, X., Cheng, Y., Luo, Y., Tong, X., Yuan, L., Wang, Y., Sun, J., Li, L., and Xiang, C. (2014) Altered fecal microbiota composition associated with food allergy in infants. *Appl Environ Microbiol 80*, 2546-2554.

[12] Cummings, J., Macfarlane, G., and Macfarlane, S. (2003) Intestinal bacteria and ulcerative colitis. *Curr Issues Intest Microbiol 4*, 9-20.

[13] Nakatsu, G., Li, X., Zhou, H., Sheng, J., Wong, S., Wu, W., Ng, S., Tsoi, H., Dong, Y., Zhang, N., He, Y., Kang, Q., Cao, L., Wang, K., Zhang, J., Liang, Q., Yu, J., and Sung, J. (2015) Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat Commun 6*, 8727.

[14] Veldhoen, M., and Brucklacher-Waldert, V. (2012) Dietary influences on intestinal immunity. *Nat Rev Immunol 12*, 696-708.

[15] Watson, G., Beaver, L., Williams, D., Dashwood, R. and Ho, E. (2013) Phytochemicals from cruciferous vegetables, epigenetics, and prostate cancer prevention. *AAPS J 15*, 951-961.

[16] Riemann-Lorenz, K., Eilers, M., von Geldern, G., Schulz, K., Kopke, S., and Heesen, C. (2016) Dietary Interventions in Multiple Sclerosis: Development and Pilot-Testing of an Evidence Based Patient Education Program. *PLoS One 11*, e0165246.

[17] Rosenberg, E., and Zilber-Rosenberg, I. (2016) Microbes Drive Evolution of Animals and Plants: the Hologenome Concept. *MBio 7*, e01395.

[18] Diaz Heijtz, R., Wang, S., Anuar, F., Qian, Y., Bjorkholm, B., Samuelsson, A., Hibberd, M.L., Forssberg, H., and Pettersson, S. (2011) Normal gut microbiota modulates brain development and behavior. *Proc Natl Acad Sci U S A 108*, 3047-3052.

[19] Lee, Y. and Mazmanian, S. (2010) Has the microbiota played a critical role in the evolution of the adaptive immune system? *Science 330*, 1768-1773.

[20] Ley, R., Peterson, D., and Gordon, J. (2006) Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell 124*, 837-848.

[21] Conly, J., and Stein, K. (1992) The production of menaquinones (vitamin K2) by intestinal bacteria and their role in maintaining coagulation homeostasis. *Prog Food Nutr Sci. 16*, 307-343.

[22] Sousa, T., Paterson, R., Moore, V., Carlsson, A., Abrahamsson, B., and Basit, A. (2008) The gastrointestinal microbiota as a site for the biotransformation of drugs. *Int J Pharm 363*, 1-25.

[23] Varankovich, N., Nickerson, M., and Korber, D. (2015) Probiotic-based strategies for therapeutic and prophylactic use against multiple gastrointestinal diseases. *Front Microbiol 6*, 685.

[24] Spinler, J., Auchtung, J., Brown, A., Boonma, P., Oezguen, N., Ross, C., Luna, R., Runge, J., Versalovic, J., Peniche, A., Dann, S., Britton, R., Haag, A., and Savidge, T. (2017) Next-Generation Probiotics Targeting Clostridium difficile through Precursor-Directed Antimicrobial Biosynthesis. *Infect Immun 85*.

[25] Flint, H., Scott, K., Duncan, S., Louis, P., and Forano, E. (2012) Microbial degradation of complex carbohydrates in the gut. *Gut Microbes 3*, 289-306.

[26] Hooper, L., Midtvedt, T., and Gordon, J. (2002) How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu Rev Nutr 22*, 283-307.

[27] Cummings, J. and Macfarlane, G. (1991) The control and consequences of bacterial fermentation in the human colon. *J Appl Bacteriol 70*, 443-459.

[28] Laine, R. (1994) A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 x 10(12) structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology 4*, 759-767.

[29] Noinaj, N., Guillier, M., Barnard, T., and Buchanan, S. (2010) TonB-dependent transporters: regulation, structure, and function. *Annu Rev Microbiol 64*, 43-60.

[30] Grondin, J., Tamura, K., Dejean, G., Abbott, D., and Brumer, H. (2017) Polysaccharide Utilization Loci: Fueling Microbial Communities. *J Bacteriol 199: e00860-16.*

[31] Schwalm, N. III, Townsend, G. II, and Groisman, E. (2016) Multiple Signals Govern Utilization of a Polysaccharide in the Gut Bacterium Bacteroides thetaiotaomicron. *MBio : e01342-16*.

[32] Cao, Y., Forstner, K., Vogel, J., and Smith, C. (2016) cis-Encoded Small RNAs, a Conserved Mechanism for Repression of Polysaccharide Utilization in Bacteroides. *J Bacteriol 198*, 2410-2418.

[33] Cantarel, B., Coutinho, P., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res 37*, D233-238.

[34] Larsbrink, J., Rogers, T., Hemsworth, G., McKee, L., Tauzin, A., Spadiut, O., Klinter, S., Pudlo, N., Urs, K., Koropatkin, N., Creagh, A., Haynes, C., Kelly, A., Cederholm, S., Davies, G., Martens, E., and Brumer, H. (2014) A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature 506*, 498-502.

[35] Martens, E., Lowe, E., Chiang, H., Pudlo, N., Wu, M., McNulty, N., Abbott, D., Henrissat, B., Gilbert, H., Bolam, D., and Gordon, J. (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol 9*, e1001221.

[36] Rogowski, A., Briggs, J., Mortimer, J., Tryfona, T., Terrapon, N., Lowe, E., Basle, A., Morland, C., Day, A., Zheng, H., Rogers, T., Thompson, P., Hawkins, A., Yadav, M., Henrissat, B., Martens, E., Dupree, P., Gilbert, H., and Bolam, D. (2015) Glycan complexity dictates microbial resource allocation in the large intestine. *Nat Commun 6*, 7481.

[37] Ndeh, D., Rogowski, A., Cartmell, A., Luis, A.S., Basle, A., Gray, J., Venditto, I., Briggs, J., Zhang, X., Labourel, A., Terrapon, N., Buffetto, F., Nepogodiev, S., Xiao, Y., Field, R.A., Zhu, Y., O' Neil, M.A., Urbanowicz, B., York, W., Davies, G., Abbott, D., Ralet, M., Martens, E., Henrissat, B., and Gilbert, H. (2017) Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature 544*, 65-70.

[38] Munoz-Munoz, J., Cartmell, A., Terrapon, N., Basle, A., Henrissat, B., and Gilbert, H. (2017) An evolutionarily distinct family of polysaccharide lyases removes rhamnose capping of complex arabinogalactan proteins. *J Biol Chem 292*, 13271-13283.

[39] Yadav, V., Yadav, P., Yadav, S., and Yadav, K. (2010) α-l-Rhamnosidase: A review. *Process Biochem 45*, 1226-1235.

[40] Sylvetsky, A., Welsh, J., Brown, R., and Vos, M. (2012) Low-calorie sweetener consumption is increasing in the United States. *Am J Clin Nutr 96*, 640-646.

[41] Spencer, M., Gupta, A., Dam, L.V., Shannon, C., Menees, S., and Chey, W.D. (2016) Artificial Sweeteners: A Systematic Review and Primer for Gastroenterologists. *J Neurogastroenterol Motil 22*, 168-180.

[42] Shankar, P., Ahuja, S., and Sriram, K. (2013) Non-nutritive sweeteners: review and update. *Nutrition 29*, 1293-1299.

[43] Zhang, W., Zhang, T., Jiang, B., and Mu, W. (2017) Enzymatic approaches to rare sugar production. *Biotechnol Adv 35*, 267-274.

[44] Park, Y., Oh, E., Jo, J., Jin, Y., and Seo, J. (2016) Recent advances in biological production of sugar alcohols. *Curr Opin Biotechnol 37*, 105-113.

[45] Zygler, A., Wasik, A., Kot-Wasik, A., and Namiesnik, J. (2011) Determination of nine high-intensity sweeteners in various foods by high-performance liquid chromatography with mass spectrometric detection. *Anal Bioanal Chem 400*, 2159-2172.

[46] Anderson, R. and Kirkland, J. (1980) The effect of sodium saccharin in the diet on caecal microflora. *Food Cosmet Toxicol 18*, 353-355.

[47] Palmnas, M., Cowan, T., Bomhof, M., Su, J., Reimer, R., Vogel, H., Hittel, D., and Shearer, J. (2014) Low-dose aspartame consumption differentially affects gut microbiota-host metabolic interactions in the diet-induced obese rat. *PLoS One 9*, e109841.

[48] Swithers, S., Laboy, A., Clark, K., Cooper, S., and Davidson, T. (2012) Experience with the high-intensity sweetener saccharin impairs glucose homeostasis and GLP-1 release in rats. *Behav Brain Res 233*, 1-14.

[49] Swithers, S., Baker, C., and Davidson, T. (2009) General and persistent effects of high-intensity sweeteners on body weight gain and caloric compensation in rats. *Behav Neurosci 123*, 772-780.

[50] Vyas, A., Rubenstein, L., Robinson, J., Seguin, R.A., Vitolins, M., Kazlauskaite, R., Shikany, J., Johnson, K., Snetselaar, L., and Wallace, R. (2015) Diet drink consumption and the risk of cardiovascular events: a report from the Women's Health Initiative. *J Gen Intern Med 30*, 462-468.

[51] Filer, L. and Stegink, L. (1989) Aspartame Metabolism in Normal Adults, Phenylketonuric Heterozygotes, and Diabetic Subjects. *Diabetes Care 12*, 67-74.

[52] Whitehouse, C., Boullata, J., and McCauley, L. (2008) The potential toxicity of artificial sweeteners. *AAOHN J 56*, 251-259; quiz 260-251.

[53] Kroger, M., Meister, K., and Kava, R. (2006) Low-calorie Sweeteners and Other Sugar Substitutes: A Review of the Safety Issues. *Compr Rev Food Sci Food Saf 5*, 35-47.

[54] Price, J., Biava, C., Oser, B., Vogin, E., Steinfeld, J., and Ley, H. (1970) Bladder Tumors in Rats Fed Cyclohexylamine or High Doses of a Mixture of Cyclamate and Saccharin. *Science 167*, 1131-1132.

[55] Renwick, A. and Williams, R. (1972) The fate of cyclamate in man and other species. *Biochem J 129*, 869-879.

[56] Rapaille, A., Goosens, J., and Heume, M. (2003) Sugar Alcohols. Encyclopedia of Food Sciences and Nutrition, Second edition, Academic Press. 5665-5671.

[57] Gardana, C., Simonetti, P., Canzi, E., Zanchi, R., and Pietta, P. (2003) Metabolism of stevioside and rebaudioside A from *Stevia rebaudiana* extracts by human microflora. *J Agric Food Chem 51*, 6618-6622.

[58] Hehemann, J., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., and Michel, G. (2010) Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature 464*, 908-912.

[59] Pluvinage, B., Grondin, J.M., Amundsen, C., Klassen, L., Moote, P.E., Xiao, Y., Thomas, D., Pudlo, N.A., Anele, A., Martens, E.C., Inglis, G.D., Uwiera, R.E.R., Boraston, A.B., and Abbott, D.W. (2018) Molecular basis of an agarose metabolic pathway acquired by a human intestinal symbiont. *Nat Commun 9*, 1043.

[60] Shimizu, H., Nakajima, M., Miyanaga, A., Takahashi, Y., Tanaka, N., Kobayashi, K., Sugimoto, N., Nakai, H., and Taguchi, H. (2018) Characterization and structural analysis of a novel exo-type enzyme acting on beta-1,2-glucooligosaccharides from Parabacteroides distasonis. *Biochemistry*, 57, 3849-3860.

[61] Brugger, S., Frei, L., Frey, P., Aebi, S., Muhlemann, K., and Hilty, M. (2012) 16S rRNA terminal restriction fragment length polymorphism for the characterization of the nasopharyngeal microbiota. *PLoS One 7*, e52241.

[62] Zaneveld, J., Lozupone, C., Gordon, J., and Knight, R. (2010) Ribosomal RNA diversity predicts genome diversity in gut bacteria and their relatives. *Nucleic Acids Res 38*, 3869-3879.

[63] Harmsen, H., Raangs, G., He, T., Degener, J., and Welling, G. (2002) Extensive set of 16S rRNA-based probes for detection of bacteria in human feces. *Appl Environ Microbiol 68*, 2982-2990.

[64] Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Meta, H., Bork, P., Ehrlich, S., and Wang, J. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature 464*, 59-65.

[65] Schnoes, A., Brown, S., Dodevski, I., and Babbitt, P. (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput Biol 5*, e1000605.

[66] Claus, S., Guillou, H., and Ellero-Simatos, S. (2016) The gut microbiota: a major player in the toxicity of environmental pollutants? *NPJ Biofilms Microbiomes 2*, 16003.

[67] Carrette, O., Favier, C., Mizon, C., Neut, C., Cortot, A., Colombel, J., and Mizon, J. (1995) Bacterial enzymes used for colon-specific drug delivery are decreased in active Crohn's disease. *Dig Dis Sci 40*, 2641-2646.

[68] Geier, M., Butler, R., and Howarth, G. (2014) Probiotics, prebiotics and synbiotics: A role in chemoprevention for colorectal cancer? *Cancer Biol Ther 5*, 1265-1269.

[69] Pollet, R., D'Agostino, E., Walton, W., Xu, Y., Little, M., Biernat, K., Pellock, S., Patterson, L., Creekmore, B., Isenberg, H., Bahethi, R., Bhatt, A., Liu, J., Gharaibeh, R., and Redinbo, M. (2017) An Atlas of beta-Glucuronidases in the Human Intestinal Microbiome. *Structure 25*, 967-977 e965.

[70] Human Microbiome Project Consortium (2012) A framework for human microbiome research. *Nature 486*, 215-221.

[71] Terrapon, N., Lombard, V., Gilbert, H., and Henrissat, B. (2015) Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics 31*, 647-655.

[72] Blin, K., Wolf, T., Chevrette, M.G., Lu, X., Schwalen, C., Kautsar, S., Suarez Duran, H., de Los Santos, E., Kim, H., Nave, M., Dickschat, J., Mitchell, D., Shelest, E., Breitling, R., Takano, E., Lee, S.Y., Weber, T., and Medema, M. (2017) antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res 45*, W36-W41.

[73] Tietz, J., Schwalen, C., Patel, P., Maxson, T., Blair, P., Tai, H., Zakai, U., and Mitchell, D. (2017) A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nat Chem Biol 13*, 470-478.

[74] Skinnider, M., Merwin, N., Johnston, C., and Magarvey, N. (2017) PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res 45*, W49-W54.

[75] Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res 13*, 2498-2504.

[76] Atkinson, H., Morris, J., Ferrin, T., and Babbitt, P. (2009) Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One 4*, e4345.

[77] Vetting, M., Al-Obaidi, N., Zhao, S., San Francisco, B., Kim, J., Wichelecki, D., Bouvier, J., Solbiati, J., Vu, H., Zhang, X., Rodionov, D., Love, J., Hillerich, B., Seidel, R., Quinn, R., Osterman, A., Cronan, J., Jacobson, M., Gerlt, J., and Almo, S. (2015) Experimental strategies for functional annotation and metabolism discovery: targeted screening of solute binding proteins and unbiased panning of metabolomes. *Biochemistry 54*, 909-931.

[78] Huang, H., Carter, M., Vetting, M., Al-Obaidi, N., Patskovsky, Y., Almo, S., and Gerlt, J. (2015) A General Strategy for the Discovery of Metabolic Pathways: d-Threitol, l-Threitol, and Erythritol Utilization in Mycobacterium smegmatis. *J Am Chem Soc 137*, 14570-14573.

[79] Levin, B., Huang, Y., Peck, S., Wei, Y., Martinez-Del Campo, A., Marks, J., Franzosa, E., Huttenhower, C., and Balskus, E. (2017) A prominent glycyl radical enzyme in human gut microbiomes metabolizes trans-4-hydroxy-l-proline. *Science 355*.

[80] Ghasempur, S., Eswaramoorthy, S., Hillerich, B., Seidel, R., Swaminathan, S., Almo, S., and Gerlt, J. (2014) Discovery of a novel L-lyxonate degradation pathway in Pseudomonas aeruginosa PAO1. *Biochemistry 53*, 3357-3366.

[81] Jacobson, M., Kalyanaraman, C., Zhao, S., and Tian, B. (2014) Leveraging structure for enzyme function prediction: methods, opportunities, and challenges. *Trends Biochem Sci 39*, 363-371.

[82] Hitchcock, D., Fan, H., Kim, J., Vetting, M., Hillerich, B., Seidel, R., Almo, S., Shoichet, B., Sali, A., and Raushel, F. (2013) Structure-guided discovery of new deaminase enzymes. *J Am Chem Soc 135*, 13927-13933.

[83] Calhoun, S., Korczynska, M., Wichelecki, D., San Francisco, B., Zhao, S., Rodionov, D., Vetting, M., Al-Obaidi, N., Lin, H., O'Meara, M., Scott, D., Morris, J., Russel, D., Almo, S.C., Osterman, A., Gerlt, J., Jacobson, M., Shoichet, B., and Sali, A. (2018) Prediction of enzymatic pathways by integrative pathway mapping. *Elife 7*, e31097.

[84] Savage, D. (1977) Microbial ecology of the gastrointestinal tract. *Annu Rev Microbiol 31*, 107-133.

[85] Wu, G., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y., Keilbaugh, S., Bewtra, M., Knights, D., Walters, W., Knight, R., Sinha, R., Gilroy, E., Gupta, K., Baldassano, R., Nessel, L., Li, H., Bushman, F.D., and Lewis, J.D. (2011) Linking long-term dietary patterns with gut microbial enterotypes. *Science 334*, 105-108.

[86] Langdon, A., Crook, N., and Dantas, G. (2016) The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. *Genome Med 8*, 39.

[87] Clemente, J., Ursell, L., Parfrey, L., and Knight, R. (2012) The impact of the gut microbiota on human health: an integrative view. *Cell 148*, 1258-1270.

[88] Janssen, A. and Kersten, S. (2015) The role of the gut microbiota in metabolic health. *FASEB J 29*, 3111-3123.

[89] Neish, A. (2009) Microbes in gastrointestinal health and disease. *Gastroenterology 136*, 65-80.

[90] Virgin, H. and Todd, J. (2011) Metagenomics and personalized medicine. *Cell 147*, 44-56.

[91] Musso, G., Gambino, R., and Cassader, M. (2011) Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. *Annu Rev Med 62*, 361-380.

[92] Cadwell, K., Patel, K., Maloney, N., Liu, T., Ng, A., Storer, C., Head, R., Xavier, R., Stappenbeck, T., and Virgin, H. (2010) Virus-plus-susceptibility gene interaction determines Crohn's disease gene Atg16L1 phenotypes in intestine. *Cell 141*, 1135-1145.

[93] Kostic, A., Gevers, D., Pedamallu, C., Michaud, M., Duke, F., Earl, A., Ojesina, A., Jung, J., Bass, A., Tabernero, J., Baselga, J., Liu, C., Shivdasani, R., Ogino, S., Birren, B., Huttenhower, C., Garrett, W., and Meyerson, M. (2012) Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res 22*, 292-298.

[94] Kane, M., Case, L., Kopaskie, K., Kozlova, A., MacDearmid, C., Chervonsky, A., and Golovkina, T. (2011) Successful transmission of a retrovirus depends on the commensal microbiota. *Science 334*, 245-249.

[95] Fleischmann, R., Adams, M., White, O., Clayton, R., Kirkness, E., Kerlavage, A., Bult, C., Tomb, J., Dougherty, B., Merrick, J., and et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science 269*, 496-512.

[96] Land, M., Hauser, L., Jun, S., Nookaew, I., Leuze, M., Ahn, T., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S., and Ussery, D. (2015) Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics 15*, 141-161.

[97] Gill, S., Pop, M., Deboy, R., Eckburg, P., Turnbaugh, P., Samuel, B., Gordon, J., Relman, D., Fraser-Liggett, C., and Nelson, K. (2006) Metagenomic analysis of the human distal gut microbiome. *Science 312*, 1355-1359.

[98] Franceus, J., Pinel, D., and Desmet, T. (2017) Glucosylglycerate Phosphorylase, an Enzyme with Novel Specificity Involved in Compatible Solute Metabolism. *Appl Environ Microbiol 83*, e01434-17.

[99] Sprogoe, D., van den Broek, L., Mirza, O., Kastrup, J., Voragen, A., Gajhede, M., and Skov, L. (2004) Crystal structure of sucrose phosphorylase from *Bifidobacterium adolescentis*. *Biochemistry 43*, 1156-1162.

[100] Buttcher, V., Welsh, T., Willmitzer, L., and Kossmann, J. (1997) Cloning and characterization of the gene for amylosucrase from *Neisseria polysaccharea*: production of a linear alpha-1,4-glucan. *J Bacteriol 179*, 3324-3330.

[101] Egloff, M., Uppenberg, J., Haalck, L., and van Tilbeurgh, H. (2001) Crystal structure of maltose phosphorylase from *Lactobacillus brevis*: unexpected evolutionary relationship with glucoamylases. *Structure 9*, 689-697.

[102] Inoue, Y., Ishii, K., Tomita, T., Yatake, T., and Fukui, F. (2002) Characterization of trehalose phosphorylase from *Bacillus stearothermophilus* SK-1 and nucleotide sequence of the corresponding gene. *Biosci Biotechnol Biochem 66*, 1835-1843.

[103] Nihira, T., Nakai, H., Chiku, K., and Kitaoka, M. (2012) Discovery of nigerose phosphorylase from *Clostridium phytofermentans*. *Appl Microbiol Biotechnol 93*, 1513-1522.

[104] Nihira, T., Saito, Y., Ohtsubo, K., Nakai, H., and Kitaoka, M. (2014) 2-O-alpha-D-glucosylglycerol phosphorylase from *Bacillus selenitireducens* MLS10 possessing hydrolytic activity on beta-D-glucose 1-phosphate. *PLoS One 9*, e86548.

[105] Yamamoto, T., Maruta, K., Mukai, K., Yamashita, H., Nishimoto, T., Kubota, M., Fukuda, S., Kurimoto, M., and Tsujisaka, Y. (2004) Cloning and sequencing of kojibiose phosphorylase gene from *Thermoanaerobacter brockii* ATCC35047. *J Biosci Bioeng 98*, 99-106.

[106] Yamamoto, T., Nishio-Kosaka, M., Izawa, S., Aga, H., Nishimoto, T., Chaen, H., and Fukuda, S. (2011) Enzymatic properties of recombinant kojibiose phosphorylase from *Caldicellulosiruptor saccharolyticus* ATCC43494. *Biosci Biotechnol Biochem 75*, 1208-1210.

[107] Qian, N., Stanley, G., Bunte, A., and Radstrom, P. (1997) Product formation and phosphoglucomutase activities in *Lactococcus lactis*: cloning and characterization of a novel phosphoglucomutase gene. *Microbiology 143 (Pt 3)*, 855-865.

[108] Lourenco, E., Maycock, C., and Rita Ventura, M. (2009) Synthesis of potassium (2R)-2-O-alpha-d-glucopyranosyl-(1-->6)-alpha-d-glucopyranosyl-2,3-dihydroxypropa noate a natural compatible solute. *Carbohydr Res 344*, 2073-2078.

[109] Bonner, W. (1958) Isomers of Tetra-O-acetyl-D-mannopyranose. *J. Am. Chem. Soc. 80*, 3372-3379.

[110] Namchuk, M., McCarter, J., Becalski, A., Andrews, T., and Withers, S. (2000) The role of sugar substituents in glycoside hydrolysis. *J. Am. Chem. Soc. 122*, 1270-1277.

[111] Pilgrim, W., and Murphy, P. (2010) SnCl(4)- and TiCl(4)-catalyzed anomerization of acylated O- and S-glycosides: analysis of factors that lead to higher alpha:beta anomer ratios and reaction rates. *J Org Chem 75*, 6747-6755.

[112] Zhu, J., McCormick, N., Timmons, S., and Jakeman, D. (2016) Synthesis of alpha-Deoxymono and Difluorohexopyranosyl 1-Phosphates and Kinetic Evaluation with Thymidylyl- and Guanidylyltransferases. *J Org Chem 81*, 8816-8825.

[113] Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics 26*, 680-682.

[114] Gerlt, J., Bouvier, J., Davidson, D., Imker, H., Sadkhin, B., Slater, D., and Whalen, K. (2015) Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim Biophys Acta 1854*, 1019-1037.

[115] Zhang, G., Dai, J., Wang, L., Dunaway-Mariano, D., Tremblay, L., and Allen, K. (2005) Catalytic cycling in beta-phosphoglucomutase: a kinetic and structural analysis. *Biochemistry 44*, 9404-9416.

[116] Lahiri, S., Zhang, G., Dunaway-Mariano, D., and Allen, K. (2003) The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. *Science 299*, 2067-2071.

[117] O'Neill, E. and Field, R. (2015) Enzymatic synthesis using glycoside phosphorylases. *Carbohydr Res 403*, 23-37.

[118] De Winter, K., Dewitte, G., Dirks-Hofmeister, M., De Laet, S., Pelantova, H., Kren, V., and Desmet, T. (2015) Enzymatic Glycosylation of Phenolic Antioxidants: Phosphorylase-Mediated Synthesis and Characterization. *J Agric Food Chem 63*, 10131-10139.

[119] Luley-Goedl, C., Sawangwan, T., Mueller, M., Schwarz, A., and Nidetzky, B. (2010) Biocatalytic Process for Production of α-Glucosylglycerol Using Sucrose Phosphorylase. *Food Technol Biotechl 48*, 276-283.

[120] De Winter, K., Desmet, T., Devlamynck, T., Van Renterghem, L., Verhaeghe, T., Pelantová, H., Křen, V., and Soetaert, W. (2014) Biphasic Catalysis with Disaccharide Phosphorylases: Chemoenzymatic Synthesis of α-d-Glucosides Using Sucrose Phosphorylase. *Organic Process Research & Development 18*, 781-787.

[121] Koppel, N. and Balskus, E. (2016) Exploring and Understanding the Biochemical Diversity of the Human Microbiota. *Cell Chem Biol 23*, 18-30.

[122] Stam, M., Danchin, E., Rancurel, C., Coutinho, P., and Henrissat, B. (2006) Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins. *Protein Eng Des Sel 19*, 555-562.

[123] Uitdehaag, J., Mosi, R., Kalk, K., van der Veen, B., Dijkhuizen, L., Withers, S., and Dijkstra, B. (1999) X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the alpha-amylase family. *Nat Struct Biol 6*, 432-436.

[124] Brown, A. (1976) Microbial water stress. *Bacteriol Rev 40*, 803-846.

[125] Goude, R., Renaud, S., Bonnassie, S., Bernard, T., and Blanco, C. (2004) Glutamine, glutamate, and alpha-glucosylglycerate are the major osmotic solutes accumulated by Erwinia chrysanthemi strain 3937. *Appl Environ Microbiol 70*, 6535-6541.

[126] Klahn, S., Steglich, C., Hess, W., and Hagemann, M. (2010) Glucosylglycerate: a secondary compatible solute common to marine cyanobacteria from nitrogen-poor environments. *Environ Microbiol 12*, 83-94.

[127] Robertson, D., Lai, M., Gunsalus, R., and Roberts, M. (1992) Composition, Variation, and Dynamics of Major Osmotic Solutes in *Methanohalophilus* Strain FDF1. *Appl Environ Microbiol 58*, 2438-2443.

[128] Pospisl, S., Halada, P., Petricek, M., and Sedmera, P. (2007) Glucosylglycerate is an osmotic solute and an extracellular metabolite produced by Streptomyces caelestis. *Folia Microbiol (Praha) 52*, 451-456.

[129] Sedmera, P., Halada, P., and Pospisil, S. (2009) New carbasugars from *Streptomyces lincolnensis*. *Magn Reson Chem 47*, 519-522.

[130] Saier, M., Jr., and Ballou, C. (1968) The 6-O-methylglucose-containing lipopolysaccharide of *Mycobacterium phlei*. Identification of D-glyceric acid and 3-O-methyl-D-glucose in the polysaccharide. *J Biol Chem 243*, 992-1005.

[131] Costa, J., Empadinhas, N., Goncalves, L., Lamosa, P., Santos, H., and da Costa, M. (2006) Characterization of the biosynthetic pathway of glucosylglycerate in the archaeon *Methanococcoides burtonii*. *J Bacteriol 188*, 1022-1030.

[132] Empadinhas, N., and da Costa, M. (2011) Diversity, biological roles and biosynthetic pathways for sugar-glycerate containing compatible solutes in bacteria and archaea. *Environ Microbiol 13*, 2056-2077.

[133] Kaur, D., Pham, H., Larrouy-Maumus, G., Riviere, M., Vissa, V., Guerin, M., Puzo, G., Brennan, P., and Jackson, M. (2009) Initiation of methylglucose lipopolysaccharide biosynthesis in mycobacteria. *PLoS One 4*, e5447.

[134] Fernandes, C., Empadinhas, N., and da Costa, M. (2007) Single-step pathway for synthesis of glucosylglycerate in *Persephonella marina*. *J Bacteriol 189*, 4014-4019.

[135] Luley-Goedl, C., and Nidetzky, B. (2010) Carbohydrate synthesis by disaccharide phosphorylases: reactions, catalytic mechanisms and application in the glycosciences. *Biotechnol J 5*, 1324-1338.

[136] Jung, J., Seo, D., Holden, J., and Park, C. (2014) Identification and characterization of an archaeal kojibiose catabolic pathway in the hyperthermophilic *Pyrococcus sp*. strain ST04. *J Bacteriol 196*, 1122-1131.

[137] Yamamoto, T., Watanabe, H., Nishimoto, T., Aga, H., Kubota, M., Chaen, H., and Fukuda, S. (2006) Acceptor recognition of kojibiose phosphorylase from *Thermoanaerobacter brockii*: syntheses of glycosyl glycerol and myo-inositol. *J Biosci Bioeng 101*, 427-433.

[138] Okada, S., Yamamoto, T., Watanabe, H., Nishimoto, T., Chaen, H., Fukuda, S., Wakagi, T., and Fushinobu, S. (2014) Structural and mutational analysis of substrate recognition in kojibiose phosphorylase. *FEBS J 281*, 778-786.

[139] Yildiz, O., Vinothkumar, K., Goswami, P., and Kuhlbrandt, W. (2006) Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation. *EMBO J 25*, 3702-3713.

[140] Fajardo, D., Cheung, J., Ito, C., Sugawara, E., Nikaido, H., and Misra, R. (1998) Biochemistry and regulation of a novel Escherichia coli K-12 porin protein, OmpG, which produces unusually large channels. *J Bacteriol 180*, 4452-4459.

[141] Jones, S., Jorgensen, M., Chowdhury, F., Rodgers, R., Hartline, J., Leatham, M., Struve, C., Krogfelt, K., Cohen, P., and Conway, T. (2008) Glycogen and maltose utilization by *Escherichia coli* O157:H7 in the mouse intestine. *Infect Immun 76*, 2531-2540.

[142] Dai, J., Wang, L., Allen, K., Radstrom, P., and Dunaway-Mariano, D. (2006) Conformational cycling in beta-phosphoglucomutase catalysis: reorientation of the beta-D-glucose 1,6-(Bis)phosphate intermediate. *Biochemistry 45*, 7818-7824.

[143] Han, X., Dorsey-Oresto, A., Malik, M., Wang, J., Drlica, K., Zhao, X., and Lu, T. (2010) Escherichia coli genes that reduce the lethal effects of stress. *BMC Microbiol 10*, 35.

[144] Misra, R. and Benson, S. (1989) A novel mutation, cog, which results in production of a new porin protein (OmpG) of *Escherichia coli* K-12. *J Bacteriol 171*, 4105-4111.

[145] Matsuda, K. and Aso, K. (1953) Studies on the unfermentable sugars (IV): On the unfermentable disaccharides in sake and koji Juice. *Hakko Kogaku Zasshi 31*, 211-213.

[146] Watanabe, T. and Aso, K. (1959) Isolation of kojibiose from honey. *Nature 183*, 1740.

[147] Aso, K. and Watanabe, T. (1961) Studies on Beer. Part II. *Journal Agric Chem Soc Japan 35*, 1078-1082.

[148] Wicken, A.J. and Baddiley, J. (1963) Structure of intracellular teichoic acids from group D streptococci. *Biochem J 87*, 54-62.

[149] Sewell, E. and Brown, E. (2014) Taking aim at wall teichoic acid synthesis: new biology and new leads for antibiotics. *J Antibiot (Tokyo) 67*, 43-51.

[150] Díez-Municio, M., Montilla, A., Moreno, F., and Herrero, M. (2014) A sustainable biotechnological process for the efficient synthesis of kojibiose. *Green Chem. 16*, 2219-2226.

[151] Beerens, K., De Winter, K., Van de Walle, D., Grootaert, C., Kamiloglu, S., Miclotte, L., Van de Wiele, T., Van Camp, J., Dewettinck, K., and Desmet, T. (2017) Biocatalytic Synthesis of the Rare Sugar Kojibiose: Process Scale-Up and Application Testing. *J Agric Food Chem 65*, 6030-6041.

[152] Hodoniczky, J., Morris, C., and Rae, A. (2012) Oral and intestinal digestion of oligosaccharides as potential sweeteners: A systematic evaluation. *Food Chem 132*, 1951-1958.

[153] Valette, P., Pelenc, V., Djouzi, Z., Andrieux, C., Paul, F., Monsan, P., and Szylit, O. (1993) Bioavailability of new synthesised glucooligosaccharides in the intestinal tract of gnotobiotic rats. *J Sci Food Agric 62*, 121-127.

[154] Sanz, M., Gibson, G., and Rastall, R. (2005) Influence of disaccharide structure on prebiotic selectivity in vitro. *J Agric Food Chem 53*, 5192-5199.

[155] Takeuchi, M., Kamata, K., Yoshida, M., Kameda, Y., and Matsui, K. (1990) Inhibitory effect of pseudo-aminosugars on oligosaccharide glucosidases I and II and on lysosomal alpha-glucosidase from rat liver. *J Biochem 108*, 42-46.

[156] Bause, E., Erkens, R., Schweden, J., and Jaenicke, L. (1986) Purification and characterization of trimming glucosidase I from *Saccharomyces cerevisiae*. *FEBS Letters 206*, 208-212.

[157] Shailubhai, K., Pratta, M., and Vijay, I. (1987) Purification and characterization of glucosidase I involved in N-linked glycoprotein processing in bovine mammary gland. *Biochem J 247*, 555-562.

[158] Ogawa, S., Ashiura, M., and Uchida, C. (1998) Synthesis of alpha-glucosidase inhibitors: kojibiose-type pseudo-disaccharides and a related pseudotrisaccharide. *Carbohydr Res 307*, 83-95.

[159] Kelly, D., Conway, S., and Aminov, R. (2005) Commensal gut bacteria: mechanisms of immune modulation. *Trends Immunol 26*, 326-333.

[160] Walters, W., Xu, Z., and Knight, R. (2014) Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Lett 588*, 4223-4233.

[161] Vogt, N., Kerby, R., Dill-McFarland, K., Harding, S., Merluzzi, A., Johnson, S., Carlsson, C., Asthana, S., Zetterberg, H., Blennow, K., Bendlin, B., and Rey, F. (2017) Gut microbiome alterations in Alzheimer's disease. *Sci Rep 7*, 13537.

[162] Hill-Burns, E., Debelius, J., Morton, J., Wissemann, W., Lewis, M., Wallen, Z., Peddada, S., Factor, S., Molho, E., Zabetian, C., Knight, R., and Payami, H. (2017) Parkinson's disease and Parkinson's disease medications have distinct signatures of the gut microbiome. *Mov Disord 32*, 739-749.

[163] Malan-Muller, S., Valles-Colomer, M., Raes, J., Lowry, C., Seedat, S., and Hemmings, S. (2018) The Gut Microbiome and Mental Health: Implications for Anxiety- and Trauma-Related Disorders. *OMICS 22*, 90-107.

[164] Boyd, S. (2013) Diagnostic applications of high-throughput DNA sequencing. *Annu Rev Pathol 8*, 381-410.

[165] Kuntz, T. and Gilbert, J. (2017) Introducing the Microbiome into Precision Medicine. *Trends Pharmacol Sci 38*, 81-91.

[166] Gilbert, J., Blaser, M., Caporaso, J., Jansson, J., Lynch, S., and Knight, R. (2018) Current understanding of the human microbiome. *Nat Med 24*, 392-400.

[167] Williams, S. (2014) Gnotobiotics. *Proc Natl Acad Sci U S A 111*, 1661.

[168] Martin, R., Bermudez-Humaran, L., and Langella, P. (2016) Gnotobiotic Rodents: An In Vivo Model for the Study of Microbe-Microbe Interactions. *Front Microbiol 7*, 409.

[169] Flint, H., Bayer, E., Rincon, M., Lamed, R., and White, B. (2008) Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Microbiol 6*, 121-131.

[170] Xu, J. (2004) Message from a human gut symbiont: sensitivity is a prerequisite for sharing. *Trends Microbiol 12*, 21-28.

[171] Mukherjee, K., Narindoshvili, T., and Raushel, F.M. (2018) Discovery of a Kojibiose Phosphorylase in Escherichia coli K-12. *Biochemistry 57*, 2857-2867.

[172] Li, C., Dong, H., and Cui, C. (2015) The synthesis and antitumor activity of twelve galloyl glucosides. *Molecules 20*, 2034-2060.

[173] Kartha, K. (1986) Iodine, a novel catalyst in carbohydrate reactions I. -isopropylidination of carbohydrates. *Tetrahedron Lett 27*, 3415-3416.

[174] Morris, P., Hope, K., and Kiely, D. (1989) The Isomeric Composition of D-Ribo-Hexos-3-Ulose (3-Keto-D-Glucose) in Aqueous-Solution. *J Carbohydr Chem 8*, 515-530.

[175] Vetter, N., Langill, D., Anjum, S., Boisvert-Martel, J., Jagdhane, R., Omene, E., Zheng, H., van Straaten, K., Asiamah, I., Krol, E., Sanders, D., and Palmer, D. (2013) A previously unrecognized kanosamine biosynthesis pathway in Bacillus subtilis. *J Am Chem Soc 135*, 5970-5973.

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

[177] Adams, P., Afonine, P., Bunkoczi, G., Chen, V., Davis, I., Echols, N., Headd, J., Hung, L., Kapral, G., Grosse-Kunstleve, R., McCoy, A., Moriarty, N., Oeffner, R., Read, R., Richardson, D., Richardson, J., Terwilliger, T., and Zwart, P. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr 66*, 213-221.

[178] Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr 60*, 2126-2132.

[179] Larkin, A., and Imperiali, B. (2009) Biosynthesis of UDP-GlcNAc(3NAc)A by WbpB, WbpE, and WbpD: enzymes in the Wbp pathway responsible for O-antigen assembly in Pseudomonas aeruginosa PAO1. *Biochemistry 48*, 5446-5455.

[180] Thoden, J. and Holden, H. (2010) Structural and functional studies of WlbA: A dehydrogenase involved in the biosynthesis of 2,3-diacetamido-2,3-dideoxy-D-mannuronic acid. *Biochemistry 49*, 7939-7948.

[181] Rodionova, I., Leyn, S., Burkart, M., Boucher, N., Noll, K, Osterman, A., and Rodionov, D. (2013) Novel inositol catabolic pathway in *Thermotoga maritima*. *Environ Microbiol 15*, 2254-2266.

[182] Hobbs, M., Williams, H., Hillerich, B., Almo, S., and Raushel, F. (2014) l-Galactose metabolism in *Bacteroides vulgatus* from the human gut microbiota. *Biochemistry 53*, 4661-4670.

[183] Benavente, R., Esteban-Torres, M., Kohring, G., Cortes-Cabrera, A., Sanchez-Murcia, P., Gago, F., Acebron, I., de las Rivas, B., Munoz, R., and Mancheno, J. (2015) Enantioselective oxidation of galactitol 1-phosphate by galactitol-1-phosphate 5-dehydrogenase from Escherichia coli. *Acta Crystallogr D Biol Crystallogr 71*, 1540-1554.

[184] Ng, K., Ye, R., Wu, X., and Wong, S. (1992) Sorbitol dehydrogenase from Bacillus subtilis. Purification, characterization, and gene cloning. *J Biol Chem 267*, 24989-24994.

[185] Zhang, L., Mu, W., Jiang, B., and Zhang, T. (2009) Characterization of D-tagatose-3-epimerase from Rhodobacter sphaeroides that converts D-fructose into D-psicose. *Biotechnol Lett 31*, 857-862.

[186] Kim, H., Hyun, E., Kim, Y., Lee, Y., and Oh, D. (2006) Characterization of an Agrobacterium tumefaciens D-psicose 3-epimerase that converts D-fructose to D-psicose. *Appl Environ Microbiol 72*, 981-985.

[187] Uechi, K., Sakuraba, H., Yoshihara, A., Morimoto, K., and Takata, G. (2013) Structural insight into L-ribulose 3-epimerase from Mesorhizobium loti. *Acta Crystallogr D Biol Crystallogr 69*, 2330-2339.

[188] Trott, O., and Olson, A.J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem 31*, 455-461.

[189] Namboori, S. and Graham, D. (2008) Enzymatic analysis of uridine diphosphate N-acetyl-D-glucosamine. *Anal Biochem 381*, 94-100.

[190] Taberman, H., Parkkinen, T., and Rouvinen, J. (2016) Structural and functional features of the NAD(P) dependent Gfo/Idh/MocA protein family oxidoreductases. *Protein Sci 25*, 778-786.

[191] Kingston, R., Scopes, R., and Baker, E. (1996) The structure of glucose-fructose oxidoreductase from Zymomonas mobilis: an osmoprotective periplasmic enzyme containing non-dissociable NADP. *Structure 4*, 1413-1428.

[192] Carbone, V., Hara, A., and El-Kabbani, O. (2008) Structural and functional features of dimeric dihydrodiol dehydrogenase. *Cell Mol Life Sci 65*, 1464-1474.

[193] Kubiak, R. and Holden, H. (2011) Combined structural and functional investigation of a C-3"-ketoreductase involved in the biosynthesis of dTDP-L-digitoxose. *Biochemistry 50*, 5905-5917.

[194] Kim, K., Kim, H., Oh, D., Cha, S., and Rhee, S. (2006) Crystal structure of D-psicose 3-epimerase from Agrobacterium tumefaciens and its complex with true substrate D-fructose: a pivotal role of metal in catalysis, an active site for the non-phosphorylated substrate, and its conformational changes. *J Mol Biol 361*, 920-931.

[195] Yoshida, H., Yamada, M., Nishitani, T., Takada, G., Izumori, K., and Kamitori, S. (2007) Crystal structures of D-tagatose 3-epimerase from Pseudomonas cichorii and its complexes with D-tagatose and D-fructose. *J Mol Biol 374*, 443-453.

[196] Müller, T. and Schmidt, R. (1995) Thymidine Diphospho-6-deoxy-α-D-ribo-3-hexulose—Synthesis of a Central Intermediate in the Biosynthesis of Di- and Trideoxysugars. *Angewandte Chemie International Edition in English 34*, 1328-1329.

[197] Ward, S., Hu, Z., Schirmer, A., Reid, R., Revill, W., Reeves, C., Petrakovsky, O., Dong, S., and Katz, L. (2004) Chalcomycin biosynthesis gene cluster from *Streptomyces bikiniensis*: novel features of an unusual ketolide produced through expression of the chm

polyketide synthase in Streptomyces fradiae. *Antimicrob Agents Chemother 48*, 4703-4712.

[198] Yew, W., Fedorov, A., Fedorov, E., Rakus, J., Pierce, R., Almo, S., and Gerlt, J. (2006) Evolution of enzymatic activities in the enolase superfamily: L-fuconate dehydratase from Xanthomonas campestris. *Biochemistry 45*, 14582-14597.

[199] Elshafei, A., and Abdel-Fatah, O.. (2001) Evidence for a non-phosphorylated route of galactose breakdown in cell-free extracts of *Aspergillus niger*. *Enzyme Microb Technol 29*, 76-83.

[200] Watanabe, S., Yamada, M., Ohtsu, I., and Makino, K. (2007) alpha-ketoglutaric semialdehyde dehydrogenase isozymes involved in metabolic pathways of D-glucarate, D-galactarate, and hydroxy-L-proline. Molecular and metabolic convergent evolution. *J Biol Chem 282*, 6685-6695.

[201] Zhang, L., Radziejewska-Lebrecht, J., Krajewska-Pietrasik, D., Toivanen, P., and Skurnik, M. (1997) Molecular and chemical characterization of the lipopolysaccharide O-antigen and its role in the virulence of Yersinia enterocolitica serotype O:8. *Mol Microbiol 23*, 63-76.

[202] Ravenscroft, N., Walker, S., Dutton, G., and Smit, J. (1991) Identification, isolation, and structural studies of extracellular polysaccharides produced by *Caulobacter crescentus*. *J Bacteriol 173*, 5677-5684.

[203] Mengele, R. and Sumper, M. (1992) Gulose as a constituent of a glycoprotein. *FEBS Letters 298*, 14-16.

[204] Stuart, E., Tirrell, S., and MacDonald, A. (1987) Characterization of an antigen secreted by Chlamydia-infected cell culture. *Immunology 61*, 527-533.

[205] Maeda, A., Adachi, S., and Matsuno, R. (2001) Improvement of selectivity in 3-ketocellobiose production from cellobiose by *Agrobacterium tumefaciens*. *Biochem Eng J 8*, 217-221.

[206] Stoppok, E., Matalla, K., and Buchholz, K. (1992) Microbial modification of sugars as building blocks for chemicals. *Appl Microbiol Biotechnol 36,* 604-610.

[207] Bernaerts, M. and De Ley, J. (1958) 3-Ketoglycosides, new intermediates in the bacterial catabolism of disaccharides. *Biochimica et Biophysica Acta 30,* 661-662.

[208] Bernaerts, M. and De Ley, J. (1960) The structure of 3-ketoglycosides formed from disaccharides by certain bacteria. *J Gen Microbiol 22,* 137-146.

[209] Kojima, K., Tsugawa, W., and Sode, K. (2001) Cloning and expression of glucose 3-dehydrogenase from *Halomonas sp*. alpha-15 in *Escherichia coli*. *Biochem Biophys Res Commun 282,* 21-27.

[210] Morrison, S., Wood, D., and Wood, P. (1999) Characterization of a glucose 3-dehydrogenase from the cultivated mushroom (*Agaricus bisporus*). *Appl Microbiol Biotechnol 51,* 58-64.

[211] Tsugawa, W., Horiuchi, S., Tanaka, M., Wake, H., and Sode, K. (1996) Purification of a marine bacterial glucose dehydrogenase fromCytophaga marinoflava and its application for measurement of 1,5-anhydro-d-glucitol. *Appl Biochem Biotechnol 56,* 301-310.

[212] Zhang, J., Zheng, Y., Xue, Y., and Shen, Y. (2006) Purification and characterization of the glucoside 3-dehydrogenase produced by a newly isolated *Stenotrophomonas maltrophilia* CCTCC M 204024. *Appl Microbiol Biotechnol 71,* 638-645.

[213] Zhang, J., Chen, W., Ke, W., and Chen, H. (2014) Screening of a glucoside 3-dehydrogenase-producing strain, *Sphingobacterium faecium*, based on a high-throughput screening method and optimization of the culture conditions for enzyme production. *Appl Biochem Biotechnol 172,* 3448-3460.

[214] Schuerman, P., Liu, J., Mou, H., and Dandekar, A. (1997) 3-Ketoglycoside-mediated metabolism of sucrose in E. coli as conferred by genes from Agrobacterium tumefaciens. *Appl Microbiol Biotechnol 47,* 560-565.

[215] Fensom, A., Kurowski, W., and Pirt, S. (2007) The use of ferricyanide for the production of 3-keto sugars by non-growing suspensions of *Agrobacterium tumefaciens*. *J Appl Chem Biotechnol 24*, 457-467.

[216] Stoppok, E. and Buchholz, K. (1999) The Production of 3-Keto-Derivatives of Disaccharides. *Carbohydr Biotechnol Prot 10*, 277-289.

[217] Hayano, K. and Fukui, S. (1967) Purification and properties of a 3-ketosucrose forming enzymes from the cells of *A. tumefaciens*. *J Biol Chem. 242*, 3655-3672

.

[218] Hayano, K. and Fukui, S. (1970) Alpha-3-ketoglucosidase of Agrobacterium tumefaciens. *J Bacteriol 101*, 692-697.

[219] Hayano, K., Tsubouchi, Y., and Fukui, S. (1973) 3-Ketoglucose reductase of Agrobacterium tumefaciens. *J Bacteriol 113*, 652-657.

[220] Ampomah, O., Avetisyan, A., Hansen, E., Svenson, J., Huser, T., Jensen, J., and Bhuvaneswari, T. (2013) The thuEFGKAB operon of *Rhizobia* and *Agrobacterium tumefaciens* codes for transport of trehalose, maltitol, and isomers of sucrose and their assimilation through the formation of their 3-keto derivatives. *J Bacteriol 195*, 3797-3807.

[221] Nakahara, K., Kitamura, Y., Yamagishi, Y., Shoun, H., and Yasui, T. (1994) Levoglucosan dehydrogenase involved in the assimilation of levoglucosan in *Arthrobacter sp*. I-552. *Biosci Biotechnol Biochem 58*, 2193-2196.