DECIPHERING THE FUNCTION OF ENZYMES RELATED TO SALMONELLA

INFECTION IN THE GUT OF CATTLE

A Dissertation

by

YUAN ZHI

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 M. Raushel
Committee Members,	David P. Barondeau
	Jennifer K. Herman
	Thomas D. Meek
	Gregory Reinhart
Head of Department,	Josh Wand

May 2020

Major Subject: Biochemistry

ABSTRACT

Non-typhoidal *Salmonella* can colonize gastrointestinal system of cattle and can also cause significant food-borne disease in humans. The use of a library of single-gene deletions (SGD) in *Salmonella* enterica serotype Typhimurium allowed identification of several proteins that are under selection in the intestine of cattle. STM2437 (*yfeJ*) encodes one of these proteins and it is currently annotated as a Type I glutamine amidotransferase. STM2437 was purified to homogeneity and its catalytic properties with a wide range of γ-glutamyl derivatives were determined. Catalytic efficiency towards the hydrolysis of L-glutamine was extremely weak with a value of k_{cat}/K_m of 20 $M^{-1} s^{-1}$. γ-L-Glutamyl hydroxamate was identified as the best substrate for STM2437, with a value of k_{cat}/K_m of 9.6 x 10^4 $M^{-1} s^{-1}$. A homology model of STM2437 was constructed based on the known crystal structure of a protein of unknown function (PDB id: 3L7N) and γ-L-glutamyl hydroxamate was docked into the active site. Acivicin was shown to inactivate the enzyme by reaction with the active site cysteine residue and the subsequent loss of HCl. Mutation of Cys91 to serine completely abolished catalytic activity. Inactivation of STM2437 did not affect the ability of this strain to colonize mice, but it inhibited the growth of *S. typhimurium* in bacteriologic media containing γ-L-glutamyl hydroxamate. Thus, although STM2437 does not appear to play a role in the murine models we tested, it is necessary for growth in the presence of γ-L-glutamyl hydroxamate. Of other genes identified, a cluster of genes is related to carbohydrate metabolism and transportation. It is proposed that the incoming carbohydrate is first

phosphorylated by a hypothetical phosphoenolpyruvate (PEP)-dependent phosphotransferase system (PTS). The metabolite is further phosphorylated by the kinase STM3781, and then cleaved by the aldolase STM3780. The aldolase was purified to homogeneity, and its aldol condensation activity with a wide range of aldehydes was determined. Surprisingly, STM3780 can condense two molecules of dihydroxyacetone phosphate (DHAP) into dendroketose bisphosphate. The retro-aldol catalytic efficiency towards fructose bisphosphate was extremely weak with a value of k_{cat}/K_m of 28 M^{-1} s^{-1}. The complete characterization of this gene cluster remains elusive.

ACKNOWLEDGEMENTS

I am grateful to my advisor, Dr. Frank M. Raushel, for his guidance and support over the last six and half years at Texas A&M University. I am also thankful to Dr. Helene Andrews-Polymenis for her collaboration with the experiments related to Salmonella, Yunyi Wang for her help to set up NMR experiments, Dr. Su Tang for his help with molecular biology, Dr. Gregory Reinhart, Dr. David P. Barondeau, Dr. Jennifer K. Herman, and Dr. Thomas D. Meek for serving as a member on my dissertation committee. In addition, I want to give a big shout-out to my lab mates in the Raushel lab, past and present, especially Dr. Andrew Bigley, Dr. Daofeng Xiang, Dr. Swapnil Ghodge, Dr. Zhongjie Ren, Dr. Jamison Huddleston for the help I received over the years.

I am thankful to my parents for the support and faith, and allowing me being away for so long. Finally, I am glad to have a better me, that I haven't given up and idled through the journey.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Frank M. Raushel, Dr. Gregory Reinhart, Dr. Jennifer K. Herman, and Dr. Thomas D. Meek of the Department of Biochemistry and Biophysics and Professor Dr. David P. Barondeau of the Department of Chemistry.

All the work conducted for the dissertation was completed by the student independently, with the exception of the chemical synthesis of various substrates that was carried out by Dr. Tamari Narindoshvili; and the Salmonella growth experiments designed and conducted by Professor Helene Andrews-Polymenis, Dr. Lydia Bogomolnaya, Marissa Talamantes, and Ahmed El Saadi.

TABLE OF CONTENTS

LIST OF SCHEMES

LIST OF FIGURES

x

xi

LIST OF TABLES

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

1.1 Interaction between Human and Microbes

Humans live in a world of microbes. Archaea, bacteria, unicellular eukaryotes and viruses reside in every possible habitat, including humans (*1*). There are estimated 10^{11} bacteria on the surface of skin, whereas there are even more in body cavities, especially in the gut. It is estimated that human body contains a 1:1 ratio of microbial cells to host cells (*2*). These bacteria may have co-developed with the individual since he or she was born, going through complex interactions with the host. These bacteria are constantly interacting with each other and the host, on the skin and other niches, especially in the gut. The dynamic of the interaction determines the relationship between the microbes and human, and microbes ranges from mutualistic to pathogenic.

1.1.1 Role of Microbes

The determination of being mutualistic or pathogenic is not definitive, as even commensal bacteria can turn to be pathogenic under certain circumstances (*3*). Conversely, passing through the physical barrier in the gut, or even residing intracellularly does not necessarily mean the bacteria are pathogenic. In the gut, *Alcaligenes spp*, commensal bacteria, intracellularly colonize dendritic cells in healthy mice, where they promote immune response and prevent disseminating themselves systemically (*4*).

As colonized bacteria are in close contact with the immune system of the host, they must have developed a way to modulate the immune response. It has been observed that gut microbes can produce immunomodulatory metabolites from aromatic amino acids (*5*), for example, histamine from histidine and indolepropionic acid from tryptophan (*6*).

In addition to escaping immune scrutiny, to successfully colonize in the host gut, bacteria have to be able to obtain required nutrients, including carbohydrates, amino acids, lipids, vitamins and other essential nutrients, and survive competition with other members of microbiota (*7*).

Complex polysaccharides derived from food are perhaps the most extensively studied substrates in the context of the human gut microbiota. Most of these molecules cannot be digested by enzymes secreted by the host but undergo extensive metabolism by gut microbes. The diversity of microbes in human gut is driven by the metabolism of polysaccharides. Complex polysaccharides are broken down by enzymes either from single organisms or a consortia of them (*8*). The final products of polysaccharide fermentation in the gut are short-chain fatty acids (SCFAs). SCFAs, including propionate, butyrate, and acetate, provide a key energy source for colonocytes. Up to 10% of daily caloric requirements are contributed by them (*1*).

With such complex interactions with the host, microbiota as a whole can be beneficial. They can modulate nutrient metabolism of the host, promote mucosal immune system formation, maintain dynamic homoeostasis and defend invading bacteria (*7, 9, 10*).

The other extreme of the interaction spectrum is pathogenic. The infection has been viewed as a linear progression that the exogenous bacteria enter, colonize and disseminate to cause disease. However, when colonization resistance is taken into consideration, there are obstructions in the process, where the invading bacteria have to overcome both the defense from the host and surrounding microbiota in order to establish a population large enough for causing disease. For example, microbiota can induce host mucosal immune defense against the invasion of exogenous bacteria. In the meantime, microbiota can directly interact with the exogenous bacteria through primary metabolites (nutrients) and secondary metabolites (antibiotics). The focus here is to discuss the infection of *Salmonella* enterica and its interaction with the host.

1.2 *Salmonella enterica*

Salmonella enterica are Gram-negative bacteria, and they are close relatives of *Escherichia coli* (*11*). *Salmonella* exist in a range of animals, including humans, as well as in the environment. Most subspecies of *Salmonella* are considered pathogenic (*12*). They can be consumed by animals and become contamination in food(*13*). There are two types of *Salmonella* infection, causing either local intestinal disease in a range of hosts or systemic disease. Both infections start with taking in *Salmonella*-contaminated food or water (*11*). Typically, *Salmonella*-caused gastroenteritis is uncomplicated and disappears without treatment, but disease can be severe in those with weakened immunity.

1.2.1 Infection Overview

Salmonella has to endure stomach acid, bile and intestinal defensins to make the way to the gut (*14*). If the infectious dose of *Salmonella* taken by an individual is large enough, (higher than 1,000 to 100,000 colony forming units), and competition from resident microbiota in the gut is overcome, surviving *Salmonella* can overwhelm the defense and start to cause disease, like enteritis, in the intestine (*11*).

During enteritis, *Salmonella* proliferates mostly in the gut lumen, although some can invade the physical barrier, the intestinal mucosa, and proliferate inside of gut epithelial cells (*11*). Intraepithelial *Salmonella* can lead to inflammation and extrusion of infected cells, which may contribute to water loss and diarrhea (*15*). Released

Salmonella then reach the intestine and ultimately feces, which can start a new cycle of transmission.

1.2.2 Genome of Salmonella

To better understand the life cycle of *Salmonella* and function required for infection, the genome of *Salmonella* has been under close scrutiny. Genome sequencing efforts in *S. enterica* have focused on the most prevalent serovars of subspecies enterica. The genome sequencing of *Salmonella enterica* subspecies I, serovar typhimurium (*S. typhimurium*) was completed in 2001 (*16*). There are ~4.9 Mbps in the genome of *Salmonella* typhimurium LT2, with an estimated ~4500 genes. When compared with the genome of *E. coli* K12, 29% of the genes are missing. Respectively, *E. coli* K12 has ~4.6 Mbps in the genome, with ~4300 genes (*17*).

Salmonella and *E. coli* are both enterobacteria and share high similarity (>70%) in genome. As *E. coli* K12 is considered commensal, with such high similarity, it raises the question what makes one commensal and the other one pathogenic? It was proposed that *Salmonella* is pathogenic because of those different genes from commensal *E. coli* (*18*). An attempt has been made to convert commensal *E. coli* into a pathogen. A plasmid containing virulence genes from *Shigella flexneri* was introduced into a laboratory strain of *E. coli* to make it invasive (*19*). In a separate experiment, the locus of enterocyte effacement from enteropathogenic *E. coli* was transferred to a laboratory strain of *E. coli,* yielding pathogenic phenotype (*20*). Both experiments show that *E. coli* has capacity to become a pathogen upon the acquisition of virulence gene cluster.

Therefore, identifying virulence related genes is essential to understanding pathogenicity of *Salmonella*.

Microbes have to survive not only in complex environment like gut, but also in hostile environment after invasion. Many genes are required for biosynthesis of nutrients that are scarce within host tissues, and for defense against host microbicidal mechanisms. These genes are encoded in commensal *E. coli*, as well as encoded in the *Salmonella* genome. In addition to these genes present in benign microorganisms, there are other *Salmonella*-specific genes related with virulence. Most of them are encoded in pathogenicity islands: large clusters of virulence genes not found in related non-pathogenic species.

These clusters of virulence genes are not only confined to plasmids like those in *Yersinia* and *Shigella*, but also in the chromosome of pathogenic organisms. They are responsible for particular virulence phenotype (*21*). Two major pathogenic islands, SPI-1 and SPI-2, have been identified in *Salmonella*, each contributing to different stages of infection. SPI-1 contains a gene set responsible for invading epithelial cells that encodes components for Type III secretion system. SPI-2 contains genes contributing to intra-macrophage survival and encodes a distinct Type III secretion system and a two-component regulatory system.

It was proposed that at least 60 genes are responsible for virulence of *Salmonella*, the majority of which are present in pathogenicity islands SPI-1 and SPI-2. In addition to these gene islands, there are several smaller pathogenicity islets essential for virulence. Even though many enzymes that are synthesized and active during infection

(based on proteomics data), they contribute weakly to virulence (*22*). This may result from: i) limited relevance of the pathway's product to virulence; ii) presence of isozymes or alternative pathways; iii) availability of needed products or alternative metabolites produced by the host microenvironment. Therefore, to gain a full picture of infection, one cannot only focus on these clustered genes, as there are other unclustered genes contributing to different steps of colonization and infection. This highlights the need for a screening methodology that is able to identify essential genes related to infection.

1.2.3 Targeted Single Gene Deletion to Study the Gene of Salmonella

Genetic screening is a procedure used to introduce mutations to microorganism, which is able to detect the phenotype change caused by a gene or a set of genes. It is one of the most efficient methods to identify virulence factors associated with certain phenotypes of bacteria, in the case here, infection. *Salmonella* can infect a range of animal species, including chicken, cattle, mouse and human (Figure 1).
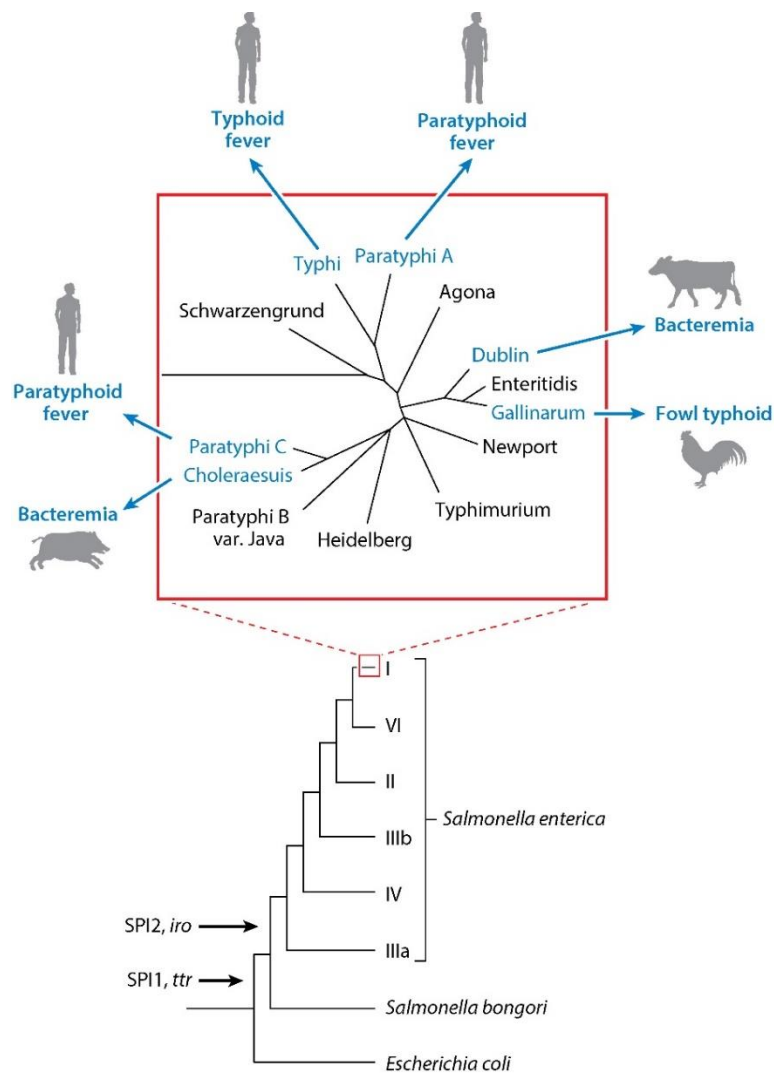
Figure 1 Phylogenetic tree of *Salmonella*. *S. enterica* subspecies I contain strains responsible for infection on a range of hosts, including human. Reprinted from Figure 1 of (*23*).

There are many animal models for salmonellosis, however, the majority of screening related with NTS infection have been done in mice models that do not naturally develop inflammatory diarrhea when infected with NTS. To more closely mimic human disease, mice need to be treated with broad spectrum antibiotics to eliminate natural microflora prior to infection with *Salmonella*. These pre-treated animals do develop neutrophilic inflammation, however, it precludes the opportunity to observe the survival mechanism of *Salmonella* in intact complex microbiota environment of the gastrointestinal tract. Therefore, it is reasonable to monitor the infection development in calf model, which they are naturally susceptible to infection with NTS and developing inflammatory diarrhea that most resembles human. To conduct systematic mutation or deletion across genome to identify genes of interest, transposon mutagenesis of *Salmonella* was employed (*24*). Using transposons to randomly knock out genes in genome is not ideal. It is hard to cover the complete genome, and hard to target short genes with this method. Therefore, it requires a minimum population to be sufficient for complete coverage. However, during infection, the bacterial population may fall below the threshold. In addition, the nature of transposon insertions can have polar effect to tune down downstream gene expression, making it hard to map observed phenotype to the corresponding gene.

1.2.4 Construction of the SGD Library

In order to overcome these mentioned issues, Santiviago *et al.* constructed a library of 1052 gene-targeted mutants, consisting primarily of *Salmonella* genes not found in *E. coli (25)*. In a plasmid, two gene specific primers are located upstream and downstream of an antibiotic resistance gene (Figure 2). Between the primer and the antibiotic resistance gene, there is an FRT element (yeast FLP/FLP recombination target), which is used for site-specific recombination. A T7 RNA polymerase promoter is positioned directly downstream of each mutant to generate a unique transcript from the *Salmonella* genome. A ribosomal binding site (RBS) and a downstream ATG (start codon) are located near the 3` end of the inserted sequence. They were designed to ensure that a 12 amino acid peptide is transcribed to reduce polar effects. The resulting PCR product was designed to have 45-base sequences at the 5` and 3` ends that are homologous to those at each end of the targeted ORF. Transformation of these PCR products into *Salmonella typhimurium* ATCC14028 expressing lambda Red recombinase in trans led to a recombination resulting in the swap-in of the PCR product, and swap-out of the targeted gene. Targeted mutants in many genes were pooled and then used for monitoring with hybridization on microarray.

Figure 2 Construction of the SGD library. Reprinted from Figure 1 of (*25*).

To identify genes that contribute to long-term systemic infection in mice, a transposon based mutant library was used to infect mice, and over 100 genes were shown to contribute to a 28-day infection (*26*).

A similar transposon based mutant library was inoculated into ligated ileal loops of calves. After a period of infection, the loops were harvested and tissues were processed to extract total DNA. After preparation, the DNA fragments were hybridized to a microarray along with control DNA. The relative signal of each mutant was compared to a control group and normalized. This experiment identified 54 genes of interest that are essential for *Salmonella* to survive and colonize in the gut of cattle (Figure 3).
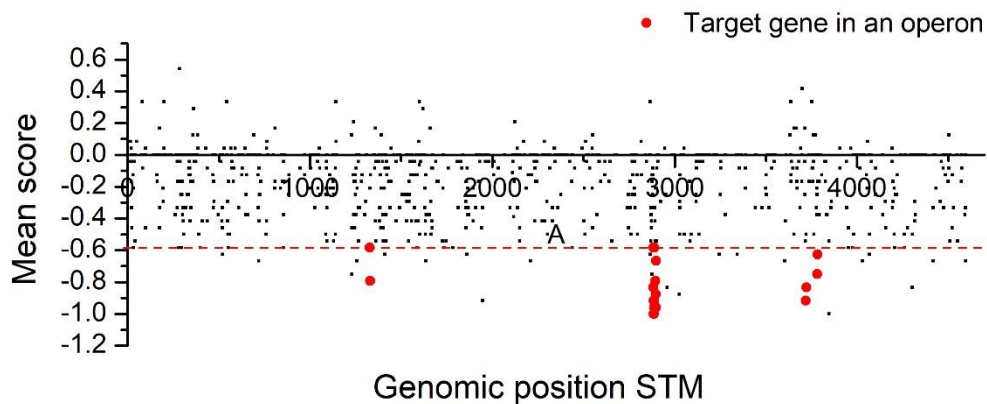
Figure 3 Loss of targeted mutant genes after inoculation into calf ligated ileal loops shown by genomic position. Genes that fall in gene clusters with score < -0.5 were colored in red to show the relative location in the genome, adapted from Figure 2 of (*27*).

1.2.5 Identified Genes of Interest

The candidate genes are annotated to be associated with lipopolysaccharide biosynthesis, metabolism, posttranslational modification, signal transduction, transcription regulation (*27*). Of the genes not previously identified to be important during infection, only seven of them were under selection in both mice and cattle models. This may indicate that different sets of genes are essential for survival under different infection conditions. Identified 54 genes in the cattle model are predicted to belong 42 operons, in which 21 of them are aggregated in 4 operons (Table 1).

Table 1 Description of identified operons related with *Salmonella* colonization.

Operon	Identified from the library	Function
Salmonella pathogenicity island 1	STM2867, STM2875, STM2883, STM2884, STM2885, STM2886, STM2887, STM2888, STM2889, STM2892, STM2893, STM2896, STM2897, STM2898	Type III Protein Secretion System, needed for the entry of *Salmonella* species into mammalian cells
STM1329-STM1331	STM1329, STM1331	Unknown
STM3714-STM3724	STM3719, STM3722, STM3723	Lipopolysaccharide biosynthesis
STM3779-STM3785	STM3781, STM3783	Carbohydrate transportation and catabolism

The 54 genes identified under selection, along with a short description, are listed in Appendix A, Table 14. Out of these candidates, we selected STM2437 *yfeJ* and Operon STM3779-STM3785 for further characterization in order to understand their roles in survival and colonization of *Salmonella* in the gut of cattle. The information collected based on their annotation was not enough to design experiments to characterize their potential functions.

1.2.6 Methodology to Predict and Characterize the Target of Interest

Lacking clear phenotype associated function, the entry point of identifying the unknown function of an enzyme depends on homology-based function prediction. Other information that can be utilized to predict protein function include operon or metabolic context provided by co-localized genes and known protein-protein interactions.

Similarity in sequence or structure implies similar function. By aligning sequence of interest with other sequence with characterized function, one can identify domains and conserved residues. With phylogenetic tree, one can comprehend functional divergences and distance between different subset of proteins. Yet both methods above cannot deal with large sets of evolutionarily related proteins (*28*). With N-1 dimension of data set, where each protein is linked with another protein by similarity, it is necessary to compile them into 2 or 3-dimension to facilitate relationship observation. Therefore, a sequence similarity network (SSN) is built to visualize the data in 2-dimension.

In a sequence similarity network, each node represents a single protein, where similar proteins are connected through an edge. The distance between any nodes, or the length of the edge, is highly correlated with similarity score, as a result of compiling from the N-1 dimension of data into 2-dimension. The similarity score can be BLASTp score or sequence identity percentage. Each node is information-rich, where it can contain detailed attributes, like functional annotation, organism, whether it has been characterized, or whether its structure has been solved. Theoretically, any two proteins can have certain degree of similarity. To facilitate the visualization, a cut-off of similarity score can be defined by user, thus only similarity above a pre-set threshold can be shown in the network. This is a way to relax the similarity from permissive to stringent (Figure 4). In addition, one protein may have thousands of close to identical homologs from the same species or close relatives. To reduce such redundancy in visualization and BLASTp calculation, the raw data set before input can be pre-sorted to remove redundant protein sequence. In other words, proteins with the richest information are kept as representatives of these redundant proteins.
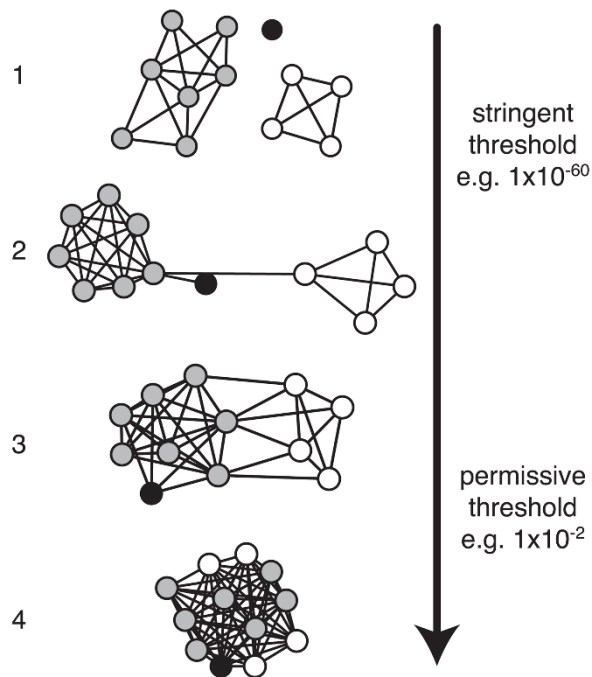
Figure 4 The same sequence similarity networks are represented at different thresholds. Nodes represent sequences. Edges represent pairwise sequence alignments better than a threshold. Reprinted from Figure 1A of (*28*).

Following the above principle, a set of protein sequence from a superfamily or a

COG can be uploaded to EFI tool, where a sequence similarity network can be

generated. A sequence similarity network with selected similarity cut-off is illustrated

by Cytoscape in Figure 4. Proteins with similar sequence, and thus more likely similar

function, are clustered together. To identify the function of an unknown protein, one can

start by checking all available annotations for any proteins from the same cluster. If

there is no available information, one can relax the cutoff to regroup proteins with lower

17

similarity. In this way, one can observe the trend of function variation within a superfamily or a COG. Therefore, one can find a reliable start point to predict the function, either from the perspective of substrate preference or catalysis in terms of type and chemical mechanism. It also can be used to track the functional change within a protein family (*28*). However, the sequence similarity boundaries between functions are unknown. One has to keep in mind that it is risky to assign function based on homology-based approaches alone. It has to be validated by both in *vitro* enzymatic assay and in *vivo* phenotypes.

In addition to sequence similarity network, one can use genomic neighborhood network (GNN). It is used to analyze the genome/operon contexts of bacterial enzymes. Within protein families, GNN can be used to identify other enzymes that are part of the same metabolic pathway to provide additional functional clues. For example, the genomic co-localization of a biosynthetic gene required for non-ribosomal peptide synthesis facilitates pathway discovery and characterization of individual genes in the operon or cluster (*29*).

Predicting the function of a pathway may be easier than predicting the function of an isolated enzyme. Solving any one gene within the pathway may provide guidance about others, as the product of one enzyme is the substrate of another in the pathway (*30*). However, a complete pathway is not always encoded the same way throughout genomes of different species, as genome restructuring and insertion into the operon can happen. By surveying the genomic context of individual protein from the same iso-function cluster at the same time, one has a better chance to identify any potential link to

other functional enzymes in the context of metabolic pathway. Making a genomic neighborhood network enables visualization and comparison of the genomic neighborhood for clusters of similar sequence (*31*). In GNN, each hub node represents a cluster from SSN. Each hub node is surrounded by neighbor genes, which are shaped based on their Pfam family. The value of the edge is calibrated by the co-occurrence frequency of the SSN queries and neighbor genes, in which the threshold of frequency can be preset. The co-occurrence frequency is shown by the size of the spoke node. Combined with the information collected from the sequence similarity network and genomic neighborhood network, we started characterization of STM2437 YfeJ and Operon STM3779-STM3785.

1.3. Glutamyl Amidotransferase Overview

1.3.1 Glutamine Amidotransferases in Salmonella enterica

Glutamine plays an important role in incorporating nitrogen into biological molecules. A class of enzymes, glutamyl amidotransferases, catalyze the transfer of amido group on the side chain of glutamine to an acceptor substrate. The transfer itself does not require the hydrolysis of ATP, however, for some glutamyl amidotransferases, the whole catalyzed reaction involves an activated acceptor substrate, which is formed at the expense of one molecule of ATP. They are involved in the biosynthesis of nucleotides, amino acids, coenzymes and antibiotics (*32*). In the genome of *Salmonella* alone, there are more than 30 genes encoding enzymes belonging to the glutamyl amidotransferase family. (Appendix B. Table 15) To facilitate the incorporation of ammonia into an activated substrate, members of this family of enzymes usually contain one domain for the hydrolysis of glutamine and another domain or subunit to accept the transferred ammonia (*32, 33*).

These glutamyl amidotransferases are annotated based on their function, while their structure can be categorized into 3 types: Type I, Type II and those containing glutamyl amidotransferases like domain. Type I glutamyl amidotransferases have a characteristic catalytic triad Cys-His-Glu, which resembles cysteine protease. The histidine residue is proposed to activate the catalytic cysteine with the help of glutamate. Type II glutamyl amidotransferases have a unique catalytic cysteine which is located at the amino terminus of the peptide chain (N-terminal nucleophile).
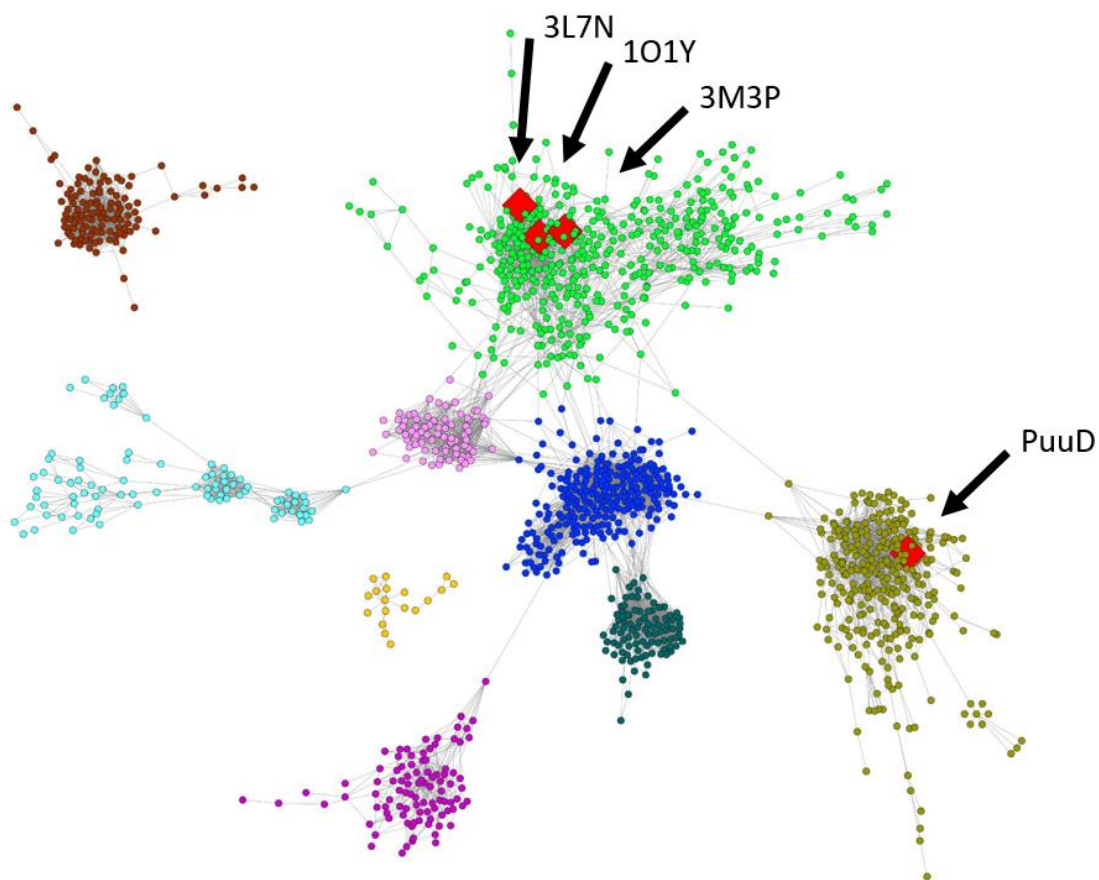
Figure 5 Sequence similarity network of glutamine amidotransferase type I. Each node in the network represents a collection of proteins grouped with 40% sequence identity. PuuD from *E. coli* is a γ-glutamyl-gamma-aminobutyrate hydrolase. 3L7N represents Smu.1228c from *Streptococcus* mutants UA159. 3L7N, 1O1Y, 3M3P are PDB codes of homologs of YfeJ.

1.3.2 YfeJ and Its Homologs

Even though the previous characterization of homologs cannot directly tell the function of YfeJ, it can still provide guidance on the type of reaction that YfeJ is involved. STM2437 (*yfeJ*) was identified as a candidate gene under selection in calves (*27*). A homologue of STM2437 in *Erwinia chrysanthemi* (now *Dickeya dadantii*), Dda3937_03594, is up-regulated after infection of African violet leaves (*34*). The gene *yfeJ* is currently annotated as encoding a putative GMP synthase-glutamine amidotransferase domain, but the specific function of the expressed protein is unknown. GMP synthase (GMPS) uses glutamine as a nitrogen donor and transfers the ammonia derived from the hydrolysis of glutamine to an ATP-activated XMP intermediate for the biosynthesis of GMP in the cell. GMPS belongs to a family of enzymes known as glutamine amidotransferases (GAT) (*32*).

In *Salmonella* enterica, GMP synthase (GuaA, STM2510) is a well-characterized enzyme that contains two functional domains: one for the hydrolysis of glutamine (residues 1-205) and another (residues 206-525) for the ultimate formation of GMP. STM2437, encoding a protein of 239 amino acids, is 28% identical in amino acid sequence to the glutamine hydrolase domain of GMP synthase. It is therefore unlikely that STM2437 functions as part of a GMPS module. In addition, the genomic neighborhood of STM2437 does not contain an obvious candidate to function as a partner for the formation of a heterodimeric enzyme.

p-Aminobenzoate synthase is an enzyme from the GAT family of enzymes, consisting of separate subunits for the hydrolysis of glutamine (PabA) and the synthesis

of *p*-aminobenzoate (PabB) (*35*) . In general, proteins that form heterodimeric complexes are often clustered close to one another in the genome. However, in bacteria such as *E. coli*, the genes for PabA and PabB are separated by ~1.6 million base pairs (*36*). Therefore, it is possible that STM2437 may associate with an unknown partner protein that is not localized to the genomic neighborhood of this gene for the biosynthesis of an unknown metabolite. In *Salmonella* enterica, the functionally annotated Type I glutamine amidotransferases include: STM0066 (small subunit of carbamoyl phosphate synthetase; STM1724 (anthranilate synthase, component II); STM2075 (imidazole glycerol phosphate synthase, subunit HisH); STM2510 (GMP synthase); STM2565 (phosphoribosylformyglycineamide synthetase); STM2953 (CTP synthetase); and STM3469 (p-aminobenzoate synthase, component II) (Figure 6).
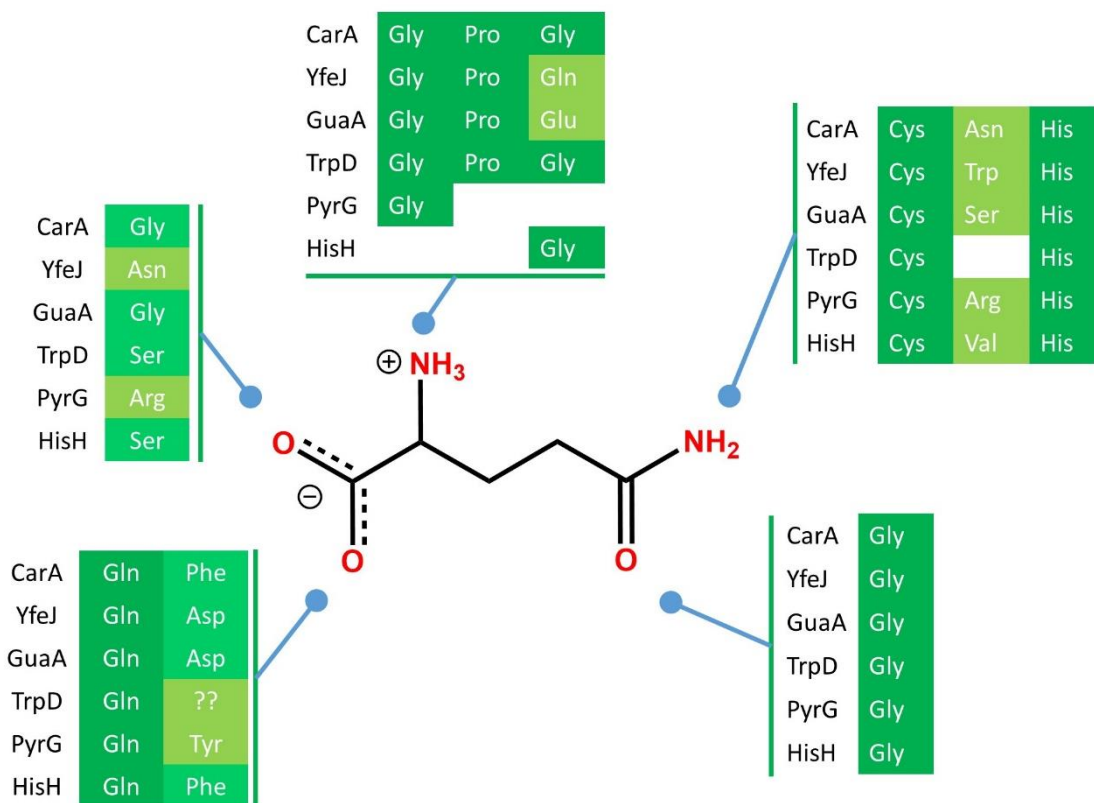
Figure 6 Active site residues conservation throughout GAT Type I enzymes.

1.3.3 Proposed Role of YfeJ

Alternatively, the primary catalytic function of STM2437 may involve the independent hydrolysis of glutamine or a modified derivative of L-glutamine (Scheme 1). Structural homologues of glutamine amidotransferase have previously been shown to catalyze the hydrolysis of γ-L-glutamyl-γ-aminobutyrate (*37*) or L-glutamine-CDP (*38*). Since STM2437 is under selection in the calf model, it is of interest to clearly understand the catalytic properties of this enzyme. Here we describe the isolation of the

purified enzyme and the determination of the substrate profile using a combination of

homology modeling, bioinformatics, and screening with a physical library of potential

substrates. Furthermore, we show that STM2437 is necessary for survival of both

Salmonella and *E. coli* in the presence of γ-L-glutamyl hydroxamate.

γ-Glutamyl amide or ester   Glutamate
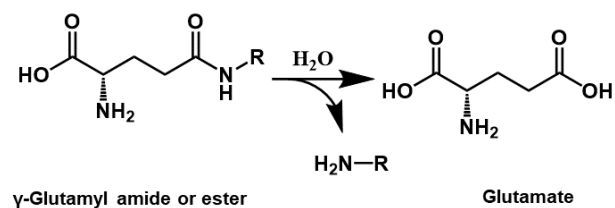
Scheme 1 Hydrolysis of glutamyl derivatives catalyzed by glutamine amidotransferases.

1.4 Carbohydrate Metabolism Control and Transportation

Of those with genomic context, a cluster of genes, STM3781 and STM3783, appear to be related to carbohydrate metabolic and transportation, where they are part of a hypothetical phosphoenolpyruvate (PEP)-dependent phosphotransferase system (PTS) in operon STM3779-STM3785. This operon is present in pathogenic *E. coli* O157:H7, but not in commensal *E. coli* K-12 (Figure 7). In this gene cluster, STM3780 is a putative Class II fructose-bisphosphate aldolase, STM3781 is a kinase belonging to the FGGY carbohydrate kinase family, STM3785 is the regulator, and STM3784-STM3782 corresponds to EIIA, EIIB, EIIC components of the PTS. STM3779 is the phosphate level sensor protein HPr.
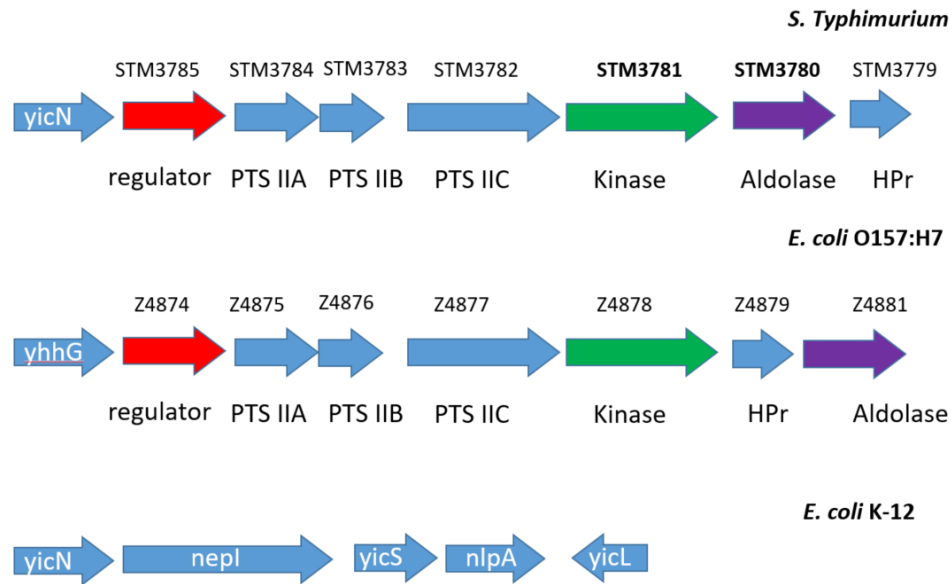
Figure 7 Genomic context comparison between *Salmonella* and *E. coli* O157:H7.
The gene *yhhG* is annotated as encoding a nickel responsive transcriptional regulator.
The gene *nepI* is annotated as encoding a purine ribonucleoside efflux pump.

1.4.1 Bacterial Phosphotransferase System (PTS)

Salmonella, as well as many other bacteria, take up carbohydrates and their

derivatives via the PTS (*39*). Therefore, a gene cluster containing PTS in the bacteria

genome indicates a pathway associated with carbohydrate metabolism, which is

important for the function discovery.

The bacterial PTS catalyzes the concomitant transportation and phosphorylation

of its carbohydrate substrates. Almost all incoming carbohydrates through PTS will first

be phosphorylated, where the phosphate is provided by PEP, and imported into the

cytosol for further transformation and metabolism (Figure 8) (*40*). It contains Enzyme I

(EI) and histidine-phosphorylatable protein (HPr) in cytosol, which senses the energy

level in cytoplasm. Both EI and HPr lack carbohydrate specificity. PTS also contains a

membrane Enzyme II complex, usually three proteins EIIA, EIIB, EIIC. In one of the

PTS family, the Mannose family, it has one more membrane spanning protein or

domain, EIID. Phosphorylation relays sequentially from PEP to EIIB, and finally to the

incoming carbohydrate, which is transported across the membrane by EIIC.

Figure 8 Signal cascade of bacterial PTS depends on the phosphorylation level in the cell. Reprinted from Figure 1 of (*41*).

PTS does more than just transporting and phosphorylating carbohydrates. It can carry out regulatory function, and act as the central regulator of carbon metabolism (*39*). In bacteria, processes that are regulated by the PTS includes transportation of non-PTS carbon sources, net production of energy storage (poly beta hydroxybutyrate and glycogen), the switch between fermentative and respiratory metabolism, flagellar motility and control of transcription of carbon and nitrogen metabolism (*40*). Therefore, PTS plays a central role to regulate the metabolism of bacteria. No PTS homologues have been identified in eukaryotes so far. This may provide new targets to control bacterial metabolism or even infection.

1.4.2 Comparative Genomic Analysis of the Bacterial PTS

Parallel analysis of PTS operons from commensal *E. coli* and *Salmonella* with their phenotype on various single carbon source can be used to identify the difference in the utilization of certain carbon sources and may result in identification of the responsible operon. Of those tested carbon sources, some of them can only be utilized by *Salmonella*, but not by *E. coli*, which can be explained by the presence or absence of certain pathways (Table 2). Of those tested compounds, D-glucosaminic acid, D-psicose, D-tartaric acid and D-galactopyranosyl-D-arabinose were shown with different growth phenotypes between *Salmonella* and *E. coli*. Based on this parallel analysis, an operon STM3779-3785 was identified in *Salmonella*, but not in commensal *E. coli*; it was concluded that this operon may be related to metabolism of D-psicose or D-tartaric acid.

Table 2 Metabolic differences between *E. coli* and *Salmonella* with no clear explanation. Adapted from Table 3 of Ref (*42*).

Compound	*E. coli*	*Salmonella*	Gene	Explanation
D-glucosaminate	-	+	*dgaABCDEF*	*Salmonella* Utilizes D-Glucosaminate via a Mannose Family PTS*
D-Psicose	-	+	?	?
D-Tartaric acid	-	+	*STM3354*	Uniprot: Q7CPN2
D-Galactopyranosyl-D-arabinose	+	-	?	?

*No reports of D-glucosaminate in the intestinal tract contents of any animal; No study identified conditions that resulted in the up-regulation of the *dga* operon or offered clues as to where *Salmonella* might encounter D-glucosaminate.

Since the publication of this analysis, *dga* operon (*dga*A-F) has been shown to be responsible for the catabolism of D-glucosaminate (*43*). The catabolism of D-tartaric acid is predicted be associated with STM3354, since it contains an annotated as an L(+)-tartrate dehydratase subunit. However, the pathway for the catabolism of D-psicose remains unknown. It remains possible that there are untested compounds to account for the presence of different PTSs in the genomes between *Salmonella* and *E. coli*. In *Salmonella*, the utilization of fructoselysine and glucoselysine has been shown to be associated with a PTS encoded by *gfrABCDEF*, which were not tested in the previous experiments (Figure 9) *(44)*.
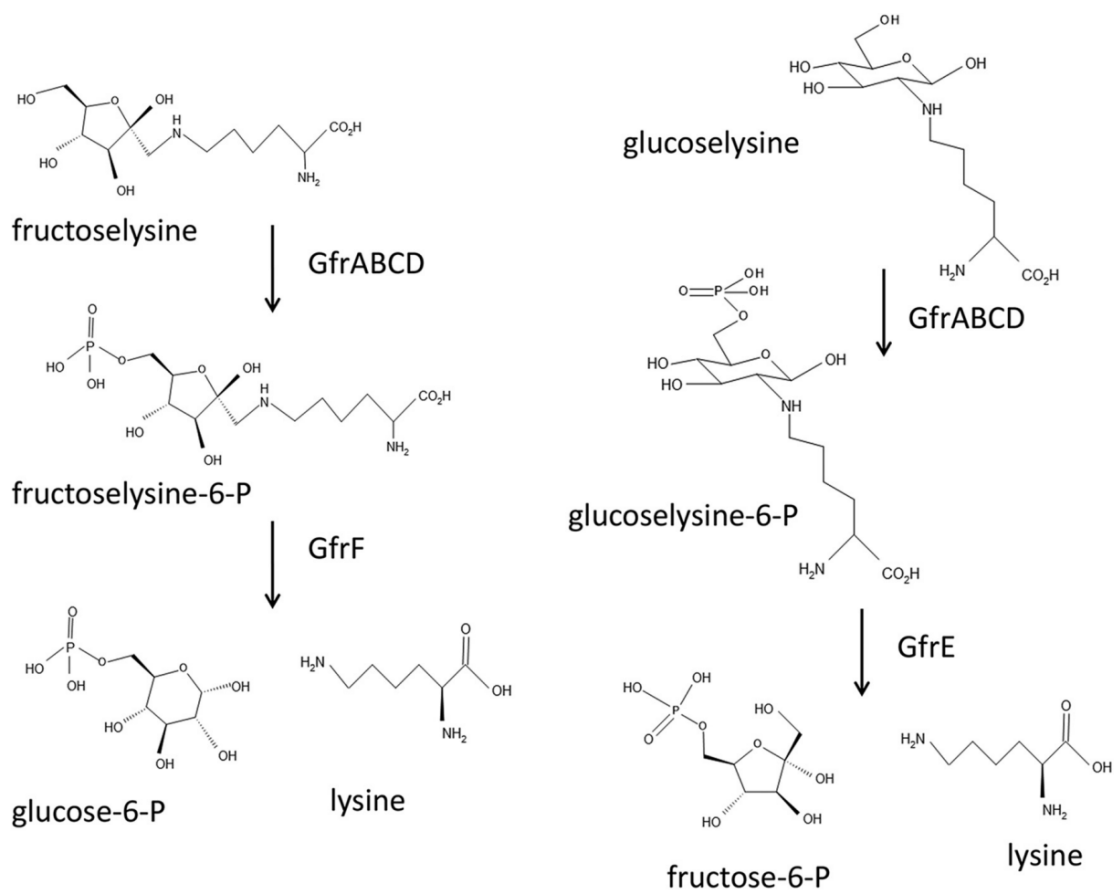
Figure 9 Proposed pathway for utilization of fructoselysine and glucoselysine.
G*frABCD* is a PTS permease complex that transports both fructoselysine and glucoselysine, phosphorylating the sugar moieties of these compounds at the C-6 position as they are transported across the membrane. GfrF and GfrE are deglycases that cleave fructoselysine 6-phosphate and glucoselysine 6-phosphate, respectively, to generate the corresponding hexose-6-phosphates and lysine. Reprinted from Figure 2 of (*44*).

1.4.3 STM3782 PTS EIIC

According to a phylogenetic tree of PTS EIIC, seven families are recognized, including EIIC^{Glc}, EIIC^{Fru}, EIIC^{Lac}, EIIC^{Gut}, EIIC^{Gat}, EIIC^{Man}, and EIIC^{Asc}. Their carbohydrate substrates are listed in Table 3.

Table 3 Carbohydrate substrates transported by various PTS permeases EIIC.
Reprinted from Table 3 of (*40*).

Family	Subfamily	Substrates
Glc	Glucose; Glucoside	Glucose, *N*-acetylglucosamine, maltose, glucosamine, and α-glucosides; β-Glucosides, sucrose, trehalose, and *N*-acetylmuramic acid
Fru		Fructose, tagatose, mannitol, mannose, and 2-*O*-mannosyl-D-glycerate
Lac	Lactose; *N,N*-diacetylchitobiose	Lactose, aromatic-glucosides, and cellobiose *N,N*-Diacetylchitobiose and lichenan oligosaccharides
Gut		Glucitol (sorbitol) and 2-methyl-D-erythritol
Gat		Galactitol and D-arabitol
Man		Glucose, mannose, sorbose, fructose, glucosamine, galactosamine, *N*-acetylgalactosamine, and several other sugars; broad-specificity porters
Asc		L-Ascorbate (anaerobic utilization)

STM3782 is an annotated EIIC family member, identified as a putative galactitol-specific transporter. As the EIIC is the most substrate-specific in this operon, it was hypothesized to be useful to start with STM3782 to guide substrate screening through its sequence similarity network. EIIC contains two domains, the dimerization domain and the substrate recognition domain.

PTS EIIC is the transmembrane component of the transportation complex that transfers the phosphate group from EIIB to the incoming carbohydrate. There are only two exceptions that the entering carbohydrate is either not phosphorylated or is subsequently dephosphorylated once inside of the membrane. Fucosyl-α-1,3-N-acetylglucosamine is transported by a PTS from *Lactobacillus casei* without being phosphorylated, and maltose taken up by a PTS from *Enterococcus faecalis* cells was reported to be first dephosphorylated to maltose by the maltose-6'-phosphate (maltose-6'-P)-specific phosphatase MapP. Within EIICs, the crystal structure of the *N,N'*-diacetylchitobiose-specific EIIC component of *Bacillus cereus* with the disaccharide bound to the active site has been determined. It confirmed that the phosphorylatable hydroxyl group at the C-6' position of the disaccharide is located close to the cytoplasm and can therefore be accessed by the P~EIIB component (*39, 45*).

From the Uniprot database, all EIICs that have been reviewed with function annotation were retrieved and analyzed to generate a sequence similarity network (Figure 10). Since the sequence of EIIC is highly conserved across different families, the similarity cut-off is set as sequence similarity of 25%. STM3782 is categorized in a family called galactitol-specific transporter, which transfers phosphate group onto the C6

of the incoming galactitol. Even though the molecular structure of galactitol is symmetric, this galactitol-1-phosphate is further transformed into D-tagatose-6-phosphate, which indicates that the phosphate group is on the C6 of galactitol as shown in Scheme 2. In the case of D-arabitol, similar rule applies that the C5 of D-arabitol is phosphorylated, in which the PTS EIIC belongs to the galactitol-specific transporter as well. Based on the current knowledge about EIIC, the incoming carbohydrate is likely phosphorylated on C6 by STM3782.
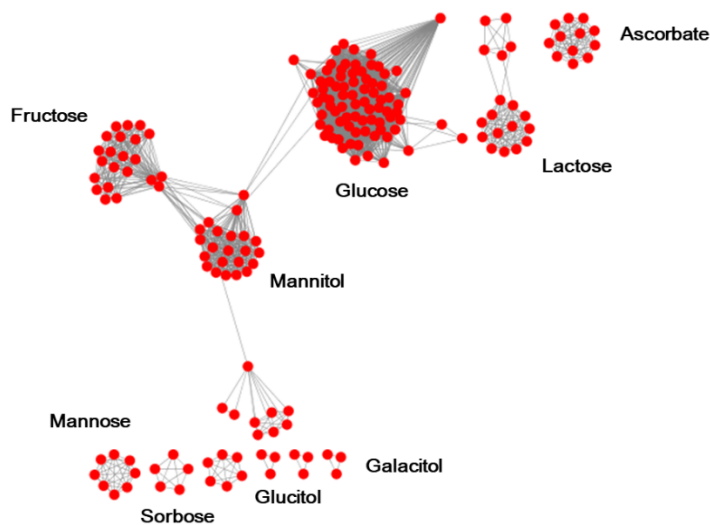
Figure 10 Sequence similarity network of PTS EIIC with reviewed function. Each node represents an entry in Uniprot. The similarity cut-off is set as sequence similarity of 25%.

34

CH₂OH ... (Scheme 2 chemical structures)

D-arabitol → PTS IIC → D-arabitol-1-P → Rotate 180° → D-arabitol-1-P → AtlD, NAD+ / NADH → D-xylulose-5-P

Galactitol → PTS IIC → Galactitol-1-P → Rotate 180° → Galactitol-1-P → GatD, NAD+ / NADH → D-tagatose-6-P
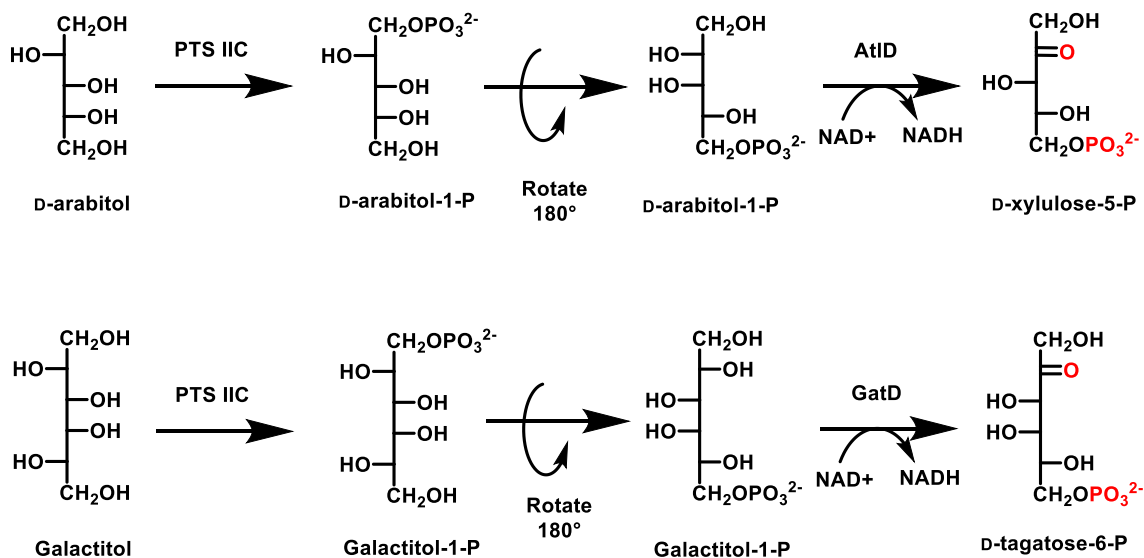
Scheme 2 Transportation and catabolism of D-arabitol and galactitol by PTS. D-arabitol to D-arabitol-1-phosphate then D-xylulose-5-phosphate. Galactitol to galactitol-1-phosphate then D-tagatose-6-phosphate.

1.4.4 STM3780 Aldolase

STM3780 was annotated as a putative fructose 1,6-bisphosphate aldolase. Its closest homolog from pathogenic *E. coli*, Z4881, is proposed as a biofilm related gene (*46*). So far, there is no reported catalytic activity associated with STM3780. The closest homolog is a D-tagatose-1,6-bisphosphate aldolase, KbaY and GatY, with 37% sequence identify to both of them.

KbaY and GatY are both annotated as D-tagatose-1,6-bisphosphate aldolase, and present in the *E. coli* genome. KbaY (former name AgaY) is part of the pathway that is responsible to transport and metabolize *N*-acetylgalactosamine, where KbaY is the last

step to cleave D-tagatose-1,6-bisphosphate into DHAP and D-glyceraldehyde-3-phosphate. The enzymatic step ahead of aldolase is proposed being phosphorylation of tagatose-6-phosphate to produce tagatose-1,6-bisphosphate by KbaZ. The hypothesis is based on the observation that the proposed catabolic pathway was missing a kinase, where a similar pathway composition is seen in the catabolic pathway from *N*-acetylgalactosamine to fructose-6-phosphate and fructose-1,6-bisphosphate (Scheme 3). In addition, no homolog of KbaZ had assigned functions. Yet, no kinase activity of KbaZ has been shown with experimental result, and the sequence of KbaZ indicates there is no ATP binding motif. Later, KbaZ was annotated as D-tagatose-1,6-bisphosphate aldolase subunit II, where in crude experiment, the co-expression of KbaZ with KbaY increased the k_{cat}/K_m of the aldolase, but KbaZ itself has no aldolase activity (*47*). Purified KbaY was shown to have D-fructose-1,6-bisphosphate aldolase activity, where with the presence of DHAP and D-glyceraldehyde-3-phosphate, D-fructose-1,6-bisphosphate is one of the products (unpublished data). It has been observed that in the genome of *E. coli*, there is another D-tagatose-1,6-bisphosphate aldolase, GatY. GatY shares 55% sequence identity with KbaY. Similarly, GatZ, the annotated chaperon of GatY that forms a heterodimer with GatY, shares 54% sequence identity with KbaZ.

A homolog of GatZ from COG4573, Atu3167 sharing 45% sequence identity, was shown that it can catalyze the epimerization between fructose-6-phosphate and tagatose-6-phosphate (*48*). In this way, the epimerase can channel tagatose-6-phosphate into a parallel catabolic pathway (Scheme 3). It is worth noting that there is no tagatose-1,6-bisphosphate aldolase in the genomic context of atu3167.

36

Predicted

N-Acetylglucosamine → N-Acetylglucosamine 6-P → Glucosamine 6-P → Fructose 6-P → Fructose 1,6-P → Fda

PTS   NagA   NagB   PfkA   Aldolase → Triose-P

Deacetylase   Deaminase   Kinase

N-Acetylgalactosamine → N-Acetylgalactosamine 6-P → Galactosamine 6-P → Tagatose 6-P ----→ Tagatose 1,6-P → KbaY

PTS   KbaA   KbaI   KbaZ

Deacetylase   Deaminase   Kinase   Aldolase

Galactitol → Tagatose → Tagatose 6-P → Fructose 6-P ----→ Fructose 1,6-P

PF00107   PF00294   AtuTag6PK / PF08013

Dehydrogenase   Kinase   Epimerase   Kinase

Scheme 3 Proposed metabolic pathways for *N*-acetylglucosamine, *N*-acetylgalactosamine, and galactitol. Unverified reactions are shown in dashed lines.

In many cases, aldolases are used in the final step of carbohydrate metabolic pathways, where they disassemble the imported and metabolized carbohydrates to provide dihydroxyacetone phosphate (DHAP) or pyruvate. For example, fructose bisphosphate aldolase cleaves fructose bisphosphate into DHAP and D-glyceraldehyde-3-phosphate.

Based on their catalytic mechanisms, aldolases can be categorized into Class I (lysine to form Schiff-base) or Class II (metal dependent) (Scheme 4) (*49*). Within Class II, aldolases can be further divided based on their products; one type has pyruvate as product and Mg as cofactor, while the other has DHAP as product and Zn as cofactor. For those having DHAP as product, they can be grouped into two clusters of orthologous groups (COG), COG0191 and COG0235, where they have different stereochemistry at C3 of the product (*50*). COG0191 includes the functionally annotated aldolases,

37

fructose bisphosphate aldolase (FbaA), D-tagatose bisphosphate aldolase (KbaY, GatY), 2-deoxy-5-keto-D-gluconate-6-phosphate aldolase (IolJ), and fructose-1,6-bisphosphate aldolase homologue (YdjI); and COG0235 includes L-rhamnulose-1-phosphate aldolase (RhaD), L-fuculose-1-phosphate aldolase (FucA) and 3-oxo-tetronate 4-phosphate decarboxylase (OtnC or YgbL) (*50-52*). Recently, YdjI was shown to prefer higher-order monosaccharides with the best substrate identified to be L-glycero-L-galacto-octulonate-1-phosphate (*53*).
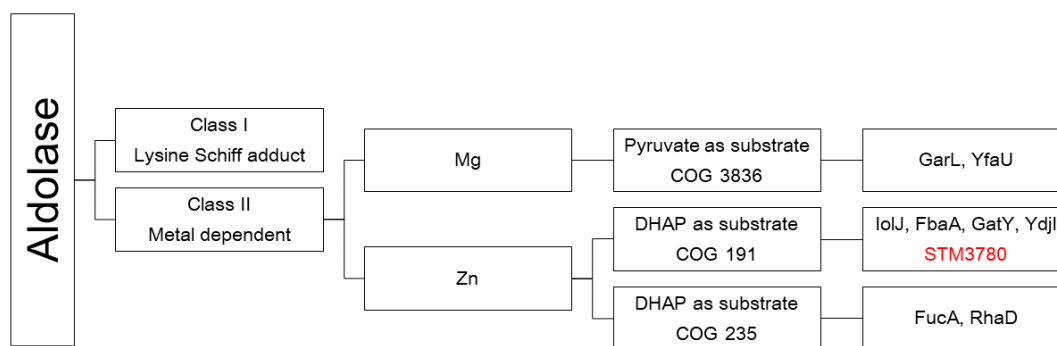
Scheme 4 Categorization of Class I and II aldolases.
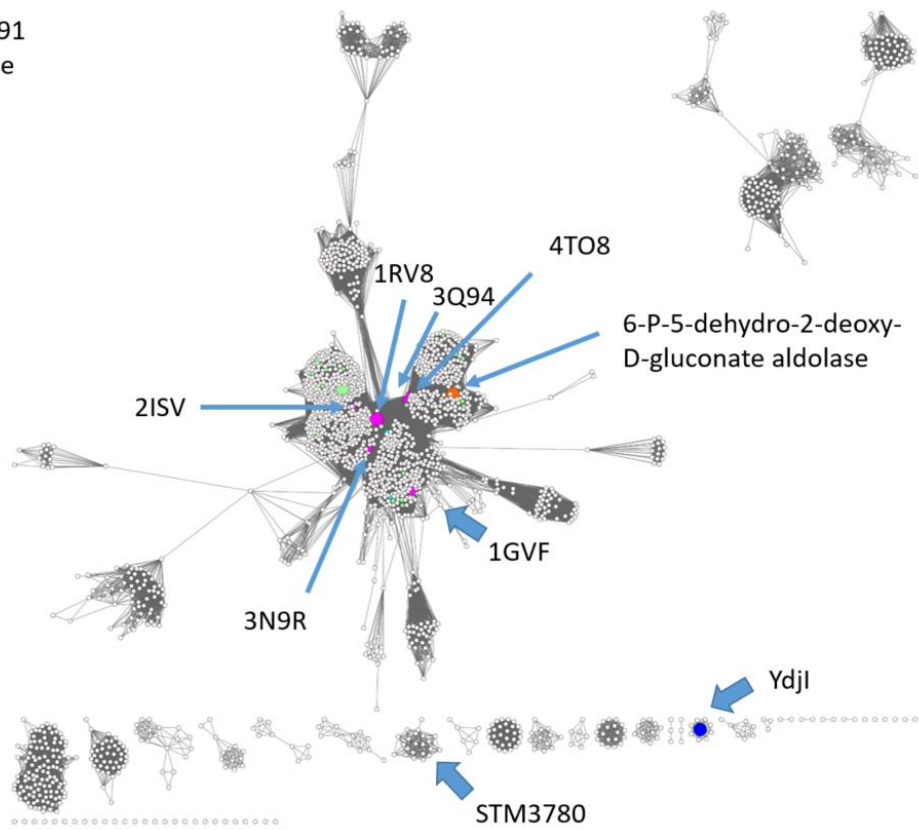
Figure 11 Sequence similarity network of STM3780 was constructed with COG0191. Cut-off with E-value -70. The aldolase STM3780 is in an isolated cluster.

Based on sequence similarity, STM3780 is categorized as a Class II aldolase in COG191, in which all known members use Zn to polarize the carbonyl group and promote the deprotonation on C3 of DHAP (Figure 11). In COG191, fructose bisphosphate aldolase and tagatose bisphosphate aldolase have crystal structure available, in which they all share a Na^+ binding pocket and use DHAP as substrate. For the other part of the condensation, it has to have an aldehyde group. Based on the structural alignment between fructose bisphosphate aldolase and tagatose bisphosphate aldolase, it is hard to tell if there are conserved residues to determine the substrate preference of aldehydes. The following questions remain to be addressed: 1) the orientation of aldehyde addition; 2) the differentiation of the end of the aldehyde between neutral group and phosphate or carboxylic group; 3) the length range of the aldehyde.

1.4.5 STM3781 Kinase

STM3781 was annotated as a putative pentulose or hexulose kinase. It is categorized in COG1069, which contains multiple known carbohydrate kinases, including rhamnulokinase, 3-dehydro-L-gulonate kinase, L-fuculokinase and Autoinducer 2 kinase. Yet, its similarity with these homologs is low, as shown in a phylogenetic tree of kinases from COG1069 and COG1070 in *E. coli* that is built to illustrate similarity within family members (Figure 12.).
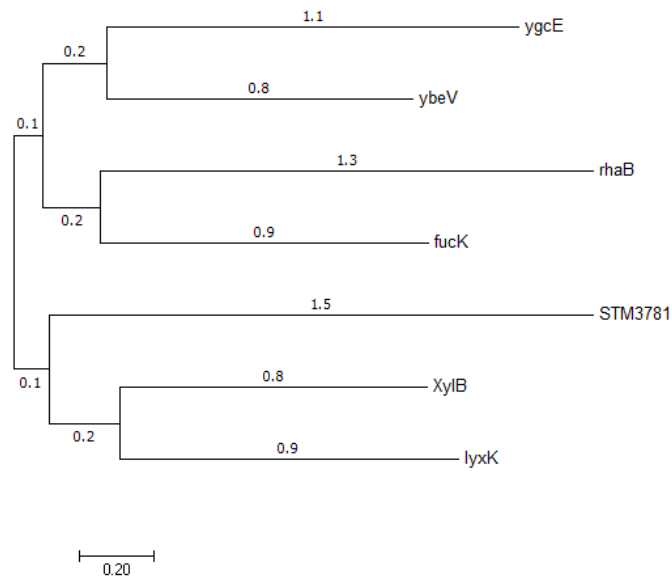
Figure 12 Phylogenetic tree of kinases from COG1069 and COG1070 in *E. coli*.
The function of YgcE remains unknown, and it is not in the genome of *Salmonella*.

Its closest homolog from *E. coli* is XylB with 24% sequence identity, which is a

xylulose kinase that phosphorylate xylulose to give xylulose-5-phosphate. Since the

similarity with its homologs in COG1069 and COG1070 is low, the function of

STM3781 remains unknown. A sequence similarity network of STM3781 was

constructed with COG1069, with E-value cut-off -70 (Figure 13). The encoded kinase is

located distant from the majority proteins in the COG, which is glycerol kinase.
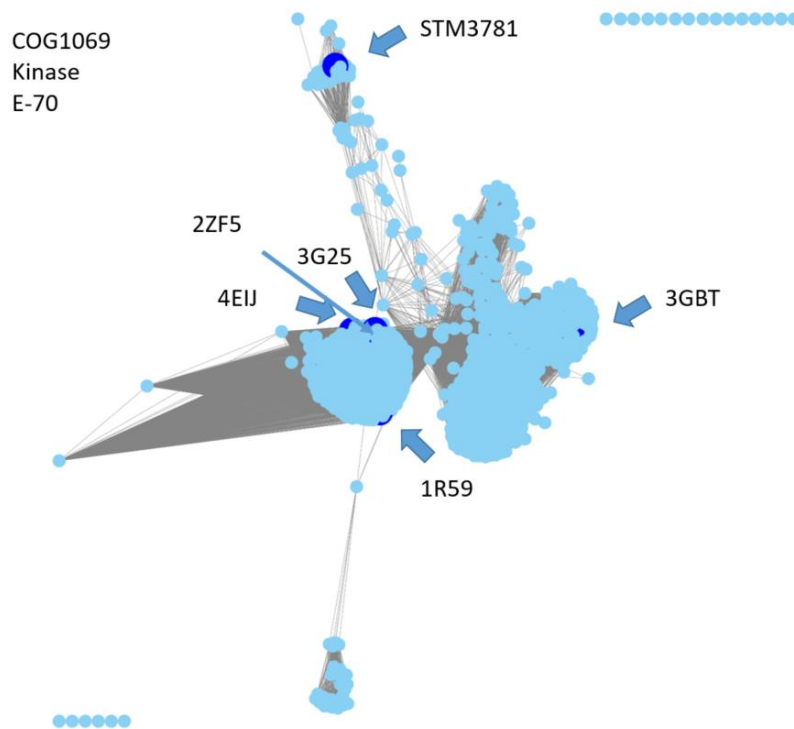
Figure 13 Sequence similarity network of COG 1069. Cutoff with E-value -70. Kinase STM3781 is distant from other characterized kinases in the COG.

STM3781 belongs to the FGGY carbohydrate kinase family. These enzymes consist of two domains. The N-terminal and C-terminal domains both adopt a ribonuclease H-like fold and are structurally related to each other. They play diverse roles in carbohydrate metabolism. An illustrative metabolic pathway map in *Thermotoga maritima* shows carbohydrate metabolism involved with FGGY kinases (Figure 14). Up until now, there has been no identified PTS that is specific for xylulose transportation.
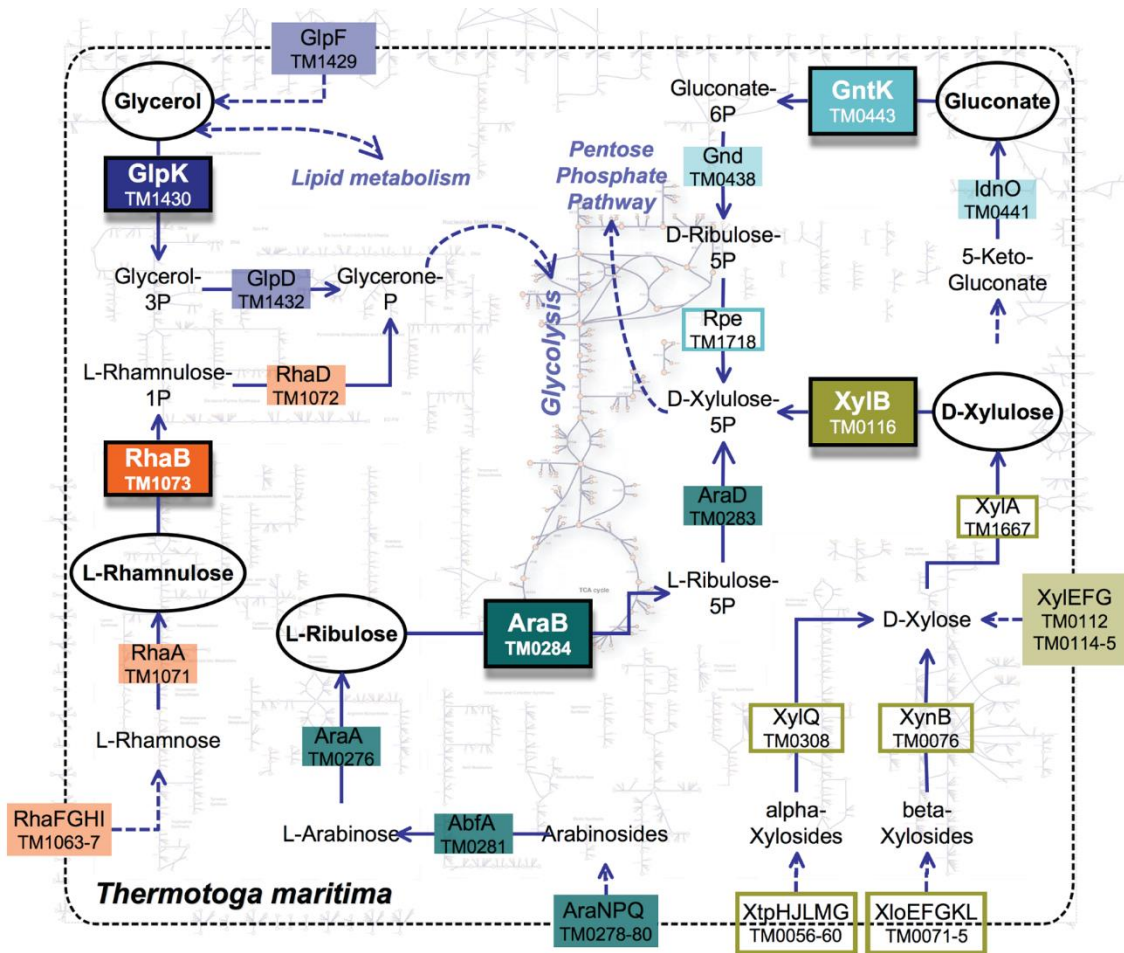
Figure 14 Overview of the metabolic pathways that employ FGGY kinases in *Thermotoga maritima*. Reprinted from Figure 1B of (*54*).

A list of functions annotated for other members in the kinase family is shown in Table 4. The similarity within FGGY kinase family is illustrated in a phylogenetic tree, with substrate and phosphorylated site emphasized (Figure 15). Based on the current observation of their functions, the kinase from FGGY kinase family prefers to phosphorylate the bottom end hydroxyl group, the one far from the carbonyl, independent of the L or D conformation of the carbohydrate substrate. The exceptions are L-rhamnulose and L-fuculose, as their bottom ends are methyl groups (deoxy carbohydrate) rather than hydroxyl groups.

Table 4 The list of function observed for the protein in the FGGY kinase family , adapted from Table 1 of (*54*).

Function	Gene	Product	PDB
L-ribulokinase	*araB*	L-ribulose-5P	3QDK
D-erythritol kinase	*eryA*	D-erythritol-4P	
L-fuculokinase	*fucK*	L-fuculose-1P	
glycerol kinase	*glpK*	D-glycerol-1P	1BO5
D-gluconokinase	*gntK*	6P-D-gluconate	3GBT
L-xylulose kinase	*lyxK*	L-xylulose-5P	
D-ribulokinase	*rbtK*	D-ribulose-5P	
L-rhamnulokinase	*rhaB*	L-rhamnulose-1P	2CGJ
D-xylulose kinase	*xylB*	D-xylulose-5P	2ITM
autoinducer-2 kinase	*lsrK/ydeV*	4,5-dihydroxy-2,3-pentanedione-5P	
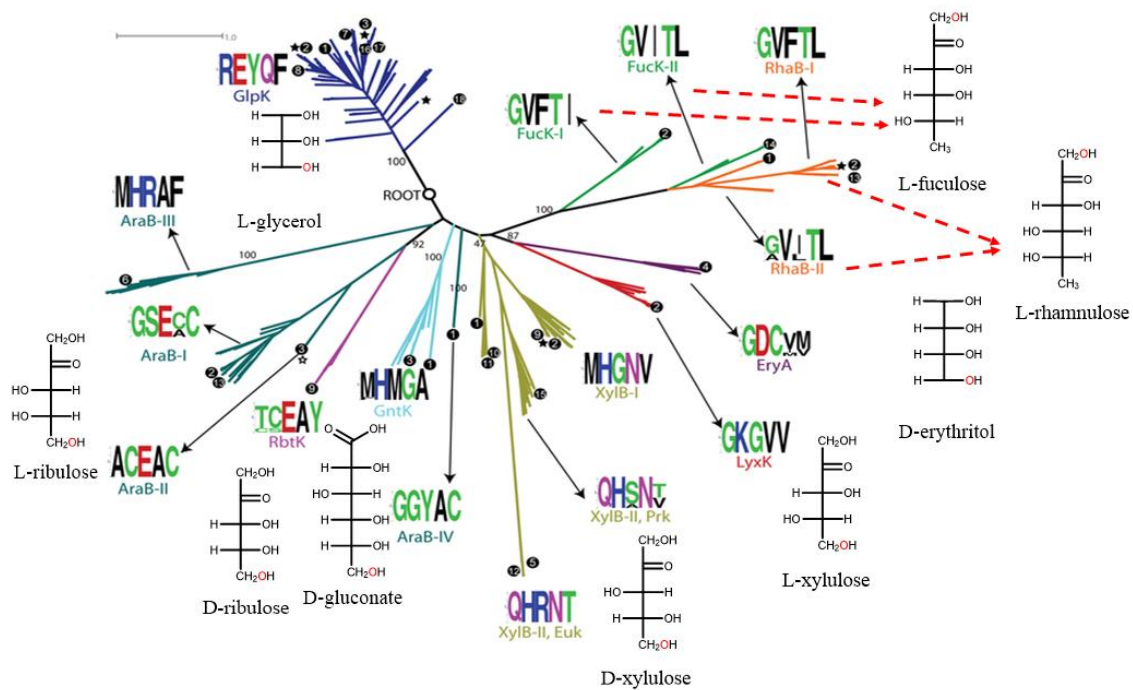
Figure 15 Phylogenetic tree of kinases from the FGGY kinase family, with the phosphorylation site highlighted in the substrates. Adapted from Figure 2 of (*54*).

1.4.6 Proposed Hypothesis of Operon STM3779-STM3785

To solve the puzzle of the metabolism related to the STM3779-STM3785 operon, the structure of preferred substrate can be predicted based on the substrate preference of each enzymes in the operon and their sequence in the pathway. To start with, the aldolase STM3780 prefers a DHAP moiety in the top half of the carbohydrate. It also needs an aldehyde in the bottom half. As EIIC STM3782, the phosphate transferase, is the first step to transform the carbohydrate, the substrate should have a hydroxyl group either on the top or the bottom of structure. Then there is a kinase STM3781, which is predicted to prefer phosphorylation the bottom end of the substrate. Either STM3781 or STM3782 should be responsible for the phosphate group in the DHAP moiety.
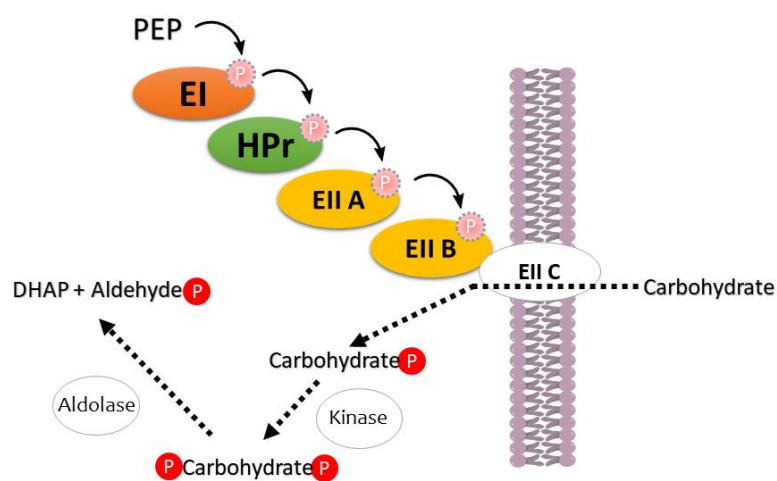
Scheme 5 Proposed catabolism pathway for operon STM3779-3785.

In the case where STM3782 first transfers a phosphate group onto the bottom of the incoming carbohydrate, STM3781 is deemed as the second enzyme in the pathway to phosphorylate the top of the phosphorylated carbohydrate (Scheme 5). The prediction on STM3781 follows the exception of FGGY kinase, where if the bottom cannot be phosphorylated, the kinase shall phosphorylate the other end. In a less probable scenario, where STM3782 first transfer a phosphate group onto the top of the incoming carbohydrate, the kinase can either be the second enzyme to phosphorylate the bottom of the phosphorylated carbohydrate, or the kinase can be the enzyme following the aldolase to phosphorylate the bottom of DHAP or the aldehyde.

CHAPTER II

DECIPHERING THE ENZYMATIC FUNCTION OF THE BOVINE ENTERIC

INFECTION RELATED PROTEIN YFEJ FROM *SALMONELLA ENTERICA*

SEROTYPE TYPHIMURIUM[*]

2.1 Materials and Methods

2.1.1 Materials

All chemicals were purchased from Sigma Aldrich, unless indicated otherwise. Theanine (10) was obtained from TCI. γ-L-Glutamyl methyl ester (2), γ-L-glutamyl ethyl ester (3) and γ-L-glutamyl nitroanilide (7) were purchased from Alfa Aesar. The dipeptide γ-L-glutamyl-L-glutamate (24) was acquired from MP BioMedicals and γ-L-glutamyl analide (9) was obtained from Chem-Impex International. γ-L-Glutamyl hydrazine was purchased from Santa Cruz Biotechnology and 4-aminobutanamide (19) was obtained from UORSY and γ-L-glutamyl methylamide (11) was obtained from Professor Daniel Romo (Baylor University). The structures of the compounds that were found to be substrates are presented in Scheme 6 and those that were found not to be substrates are found in Scheme 7.

Scheme 6 Compounds shown to be substrates for STM2437.
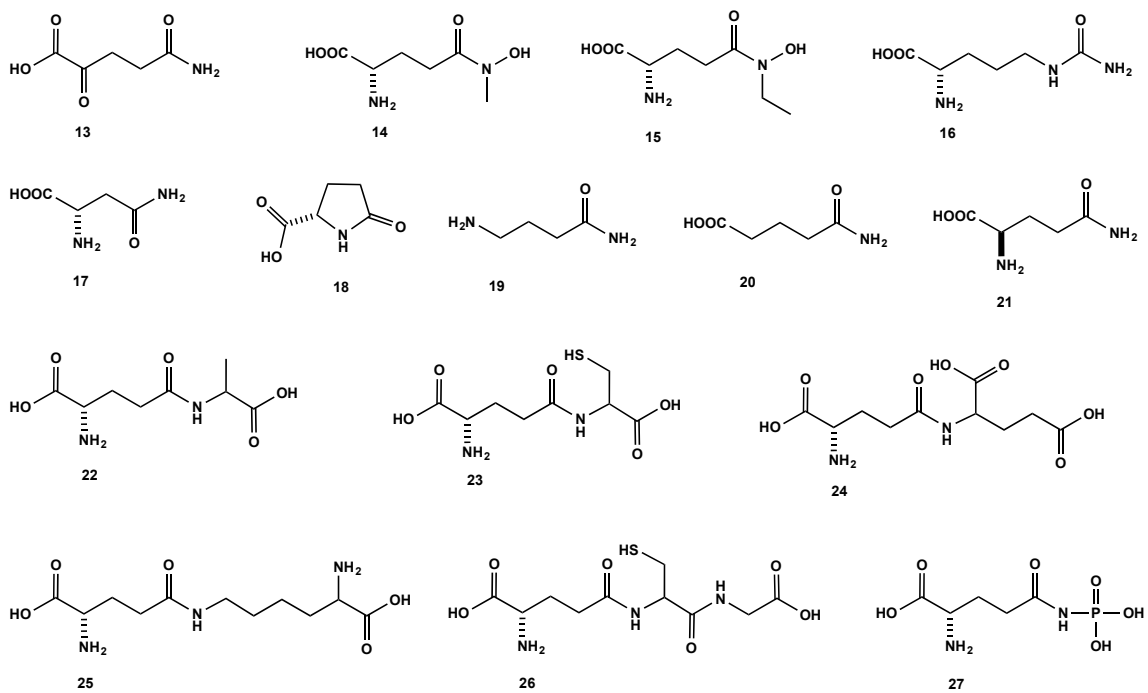
Scheme 7 Compounds tested that were not substrates for STM2437.

2.1.2 Synthesis of Glutamine Derivatives and Analogues

The synthesis of α- ketoglutaramate (13) was catalyzed by the oxidation of L-glutamine (1) with L-amino acid oxidase as described by Van Schaftingen et al. (*55*). L-Glutamine phosphate (27) was synthesized according to the method of Taylor et al. (*38*). Compounds 14 and 15 were synthesized by the procedures of Wittman et al. (*56*) and Teze et al. (*57*) .

2.1.3 Cloning, Expression, and Purification of STM2437

The gene for STM2437 (gi|16765757|; UniProt: P40194) was amplified from S. enterica ATCC 700720 genomic DNA using 5'-TACTTCCAATCCATGGTGCGAGTTCATTTTGTTGTCCATG -3' as the forward primer and 5'- TATCCACCTCTCGAGTTTCTGTGAGTGCGCCAGCG -3' as the reverse primer. PCR was performed using Pfu Turbo Polymerase from Agilent. The reaction conditions were 5 min at 95 °C, followed by 40 cycles of 45 s at 95 °C, 1.5 min at 65 °C, and 3 min at 68 °C. Restriction sites for NcoI and XhoI were inserted into the forward and reverse primers, respectively. The PCR product was purified with a PCR cleanup system from Qiagen and subsequently digested with the restriction enzymes NcoI and XhoI. The resulting DNA product was ligated into a pET-28a(+) vector obtained from Novagen, with hexa histidine-tag encoded in the C terminus.

The ligation reaction mixture was transformed into BL21(DE3) cells via electroporation. Bacterial colonies containing the plasmid of interest (STM2437 in pET-28a(+) and pET-30a(+)) were inoculated into 5-mL cultures of LB and incubated at 37

51

°C, overnight. One-liter cultures of LB containing 50 μg/mL kanamycin were inoculated with the 5-mL culture and then grown at 30 °C until an OD_{600} of 0.4−0.6 was obtained. The cultures were subsequently incubated at 30 °C, and expression was initiated by the addition of 1.0 mM IPTG. The cells were harvested after 20 h via centrifugation at 4 °C and then resuspended in 20 mM HEPES, pH 7.6, 300 mM NaCl, 2.5 mM TCEP (tris(2-carboxyethyl) phosphine), 10% glycerol, and 0.1 mg/mL PMSF. Resuspended cells were lysed via four rounds of sonication and the cell debris was removed via centrifugation at 4 °C. The supernatant solution was collected, filtered through a 0.45 μm membrane, and then applied to a Ni-NTA column, which was equilibrated with binding buffer (20 mM HEPES, pH 7.5, 300 mM NaCl and 10% glycerol). The column was washed with 5 column volumes of wash buffer (20 mM HEPES, pH 7.5, 500 mM NaCl, 20 mM imidazole and 10% glycerol). The protein was eluted with a solution containing 20 mM HEPES, pH 7.5, 500 mM NaCl, 300 mM imidazole and 10% glycerol. The protein fractions were collected based on the absorbance at 280 nm. The concentration of STM2437 (MW = 28.1 kDa) was determined by the absorbance at 280 nm using a calculated molar extinction coefficient of 39,400 M^{-1} cm^{-1}. (web.expasy.org/protparam/) The purity of the protein was confirmed by SDS-PAGE. The purified protein was flash-frozen with liquid nitrogen and stored at -80 °C.

2.1.4 Site-Directed Mutagenesis

All point mutants of STM2437 were constructed using the standard QuikChange PCR protocol provided by Stratagene. Glu12 from STM2437 was mutated to aspartate,

52

asparagine and glycine. Cys91 from STM2437 was mutated to serine. The plasmid

encoding wild-type STM2437 with a C-terminal polyhistidine tag was used as a template

for the site-directed mutagenesis for the expression of the E12D, E12Q, E12G and C91S

variants. Primers used for the site-directed mutagenesis are shown in Table 5. The

protein variants were purified using the Ni-NTA column with the same conditions as the

wild-type enzyme.

Table 5 Primers used to construct variants of STM2437. The base highlighted in red font denotes the site of the mutation.

E12D F	GTCCATGAGTCGTTTGACTCCGCTGGCGCTTATC
E12D R	GATAAGCGCCAGCGGAGTCAAACGACTCATGGAC
E12Q F	GTCCATGAGTCGTTTCAATCCGCTGGCGCTTATC
E12Q R	GATAAGCGCCAGCGGATTGAAACGACTCATGGAC
E12G F	GTCCATGAGTCGTTTGGATCCGCTGGCGCTTATC
E12G R	GATAAGCGCCAGCGGATCCAAACGACTCATGGAC
C91S F	GAATGGTTATCGGCATTAGTCTTGGTTCGCAGCTC
C91S R	GAGCTGCGAACCAAGACTAATGCCGATAACCATTC

2.1.5 Quaternary Structure

Gel filtration was performed with a Superdex 200 10/300 GL column using a solution of 25 mM HEPES, pH 7.5, and 100 mM NaCl as the running buffer at 0.5 mL/min. To estimate the molecular mass of the protein, a calibration plot of partition coefficients vs. log molecular weight was generated using a set of gel filtration standards (Sigma-Aldrich), that included carbonic anhydrase (29 kDa), albumin (66 kDa), alcohol dehydrogenase (150 kDa), β-amylase (200 kDa), apoferritin (443 kDa), and blue dextran (2000 kDa).

2.1.6 Measurement of Enzyme Activity

The enzymatic hydrolysis of L-glutamine (1) and various γ-L-glutamyl derivatives (2-4, 6, 8-11, 14, 15, 18, and 22-27) was monitored at 30 °C by following the formation of L-glutamate in a coupled enzymatic assay with bovine liver glutamate dehydrogenase (GDH) and NAD$^+$ (*58*). The production of L-glutamate was quantified by the increase in absorption at 340 nm with a SpectraMax340 UV-visible spectrophotometer, using an extinction coefficient (Δε) of 6,300 M^{-1} cm^{-1}. Reaction assays were conducted in 50 mM Tricine/KOH buffer, pH 8.0, containing 14 units/mL GDH, 1.0 mM NAD$^+$ and varying amounts of substrate. The hydrolysis of γ-L-glutamyl p-nitroanilide (7) was monitored by following the release of p-nitroaniline at 410 nm, using an extinction coefficient of 8,800 M^{-1} cm^{-1} (*59*). All of the assay mixtures were pre-incubated with GDH for 1 h prior to the addition of STM2437 to remove any L-glutamate impurities in the compounds being tested as potential substrates.

54

The hydrolysis of D-glutamine (21) was monitored by following production of

NH$_3$ in a coupled enzymatic assay with NADH and bovine liver glutamate

dehydrogenase (GDH) as the coupling enzyme (60). The production of NH$_3$ was

quantified by following the decrease in absorption at 340 nm using an extinction

coefficient of 6300 M^{-1} cm^{-1}. The reaction assays were performed in a solution

containing 50 mM Tricine/KOH, pH 8.0, 7 units/mL GDH, 10 mM of D-glutamine, 5.0

mM of α-ketoglutarate and 0.4 mM NADH. The enzymatic hydrolysis of compounds

13, 16, 19, and 20 was monitored by measuring the rate of NH3 formation.

The hydrolysis of β-L-aspartyl hydroxamate (5) and L-asparagine (17) was

monitored by following the production of L-aspartate in a coupled enzymatic assay with

glutamate/oxalacetate transaminase (GOT) and malate dehydrogenase (MDH) as the

coupling enzymes with NADH (61). The production of L-aspartate was quantified by

following the decrease in absorption at 340 nm. Reaction assays were performed in 50

mM Tricine/KOH buffer, pH 8.0, containing 3.2 units/mL MDH, 64 units/mL GOT, 2.0

mM α-ketoglutarate, 0.4 mM NADH and varying concentrations of β-L-aspartyl

hydroxamate

The enzyme-catalyzed synthesis of γ-L-glutamyl hydroxamate (4) was monitored

by following the formation of an adduct after the addition of Fe(III) at 540 nm (58). The

stop-solution was prepared from a 1:1:1 mixture containing 80% trichloroacetic acid, 6.0

M HCl, and 10% FeCl$_3$ in 0.02 M HCl. A standard curve was used to determine the

concentration of γ-glutamyl hydroxamate (58). The reaction was performed in 50 mM

Tricine/KOH buffer, pH 8.0, 12.5 mM L-glutamine, 0.5 M hydroxylamine, and 0.5 μM

STM2437. The lower limit for the detection of glutamyl hydroxamate is 0.05 mM. The range of substrate concentrations used to measure the kinetic constants for wild-type STM2437 and the various mutants are provided in Tables 6 and 7, respectively.

Table 6 Substrate concentration tested with STM2437.

Substrate	Tested conc. range (mM)	Concentration of STM2437 used in assay (uM)
L-glutamine (**1**)	0.1 - 15	3.5
γ-L-glutamyl methyl ester (**2**)	0.08 - 15	1.3
γ-L-glutamyl ethyl ester (**3**)	0.2 - 8	1.3
γ-L-glutamyl hydroxamate (**4**)	0.01 - 0.8	0.08
β-L-aspartyl hydroxamate (**5**)	0.2 - 10	1.0
γ-L-glutamyl hydrazine (**6**)	0.1 - 2	0.52
γ-L-glutamyl nitroanilide (**7**)	0.01 - 1	0.08
γ-L-glutamyl benzyl ester (**8**)	0.01 - 1	0.6
γ-L-glutamyl anilide (**9**)	0.016 - 0.32	1.0
γ-L-glutamyl ethylamide (**10**)	0.4 - 5	1.0
γ-L-glutamyl methylamide (**11**)	0.4 - 5	1.0

Table 7 Substrate concentration range utilized with variants of STM2437.

Variant	Substrate	Tested conc. range (mM)	Concentration of STM2437 used in assay (µM)
E12D		0.04 – 2.0	0.5
E12Q	L-glutamyl	0.04 – 2.0	1.0
E12G	hydroxamate	0.04 – 2.0	0.9
C91S		2.5	3.2
E12D		0.16 - 20	1.7
E12Q		0.16 - 20	4.8
E12G	L-glutamine	0.75 - 18	2.2
C91S		4.0	3.2
E12D		0.01 – 1.0	0.08
E12Q	L-glutamyl	0.01 – 1.0	0.08
E12G	nitroanilide	0.01 – 1.0	0.08
C91S		0.4	3.0

2.1.7 Inactivation by Acivicin

Acivicin (12) was tested as an inactivator of STM2437, as this compound has previously been shown to covalently inactivate other γ-glutamyl transferase enzymes (*62*). To determine the partition ratio for inactivation of STM2437 by acivicin, the reactions were conducted by incubating 2.1 µM STM2437 containing 1.0 mg/mL BSA with varying concentrations of acivicin (0.5 – 6 µM) for 30 min at 30 °C in 80 mM Tricine/KOH, pH 8.0. Aliquots of 100 µL were removed and then added to 150 µL of an assay solution containing 0.8 mM L-glutamyl *p*-nitroanilide and 80 mM Tricine/KOH buffer (pH 8.0). The reaction was monitored by following the release of p-nitroaniline at 410 nm, using an extinction coefficient of 8,800 M^{-1} cm^{-1}.

The rate of enzyme inactivation by acivicin was determined by monitoring the changes in the catalytic activity of STM2437 as a function of time using various concentrations of acivicin (4-100 µM) at an enzyme concentration of 1.0 µM at 30 °C in 62.5 mM Tricine/KOH buffer, pH 8.0. Aliquots of each incubation mixture were removed at various times of incubation and the remaining catalytic activity was determined using 0.9 mM L-glutamyl p-nitroanilide as the substrate. The catalytic activity was plotted against the time of incubation to obtain the apparent first-order rate constant (k_{obs}) for inactivation. The value of k_{obs} was subsequently plotted against the acivicin concentration to determine the dependence of the inactivation rate constant on the concentration of acivicin.

Confirmation of a covalent adduct that forms between acivicin and STM2437 was determined by mass spectrometry using an API QSTAR Pulsar Hybrid QTOF mass

spectrometer (Applied Biosystems/MDS Sciex, Framingham, MA) equipped with an

electrospray ionization (ESI) source. STM2437 (16 μM) was incubated with 80 μM

acivicin in 20 mM HEPES, pH 7.5. After an incubation period of 1 h the reaction

mixture was loaded and eluted from a buffer-exchanged PD-10 column (GE-Healthcare).

The mass spectrometric data were acquired in the positive ion mode (500-2000 Da)

using a spray voltage of +2700 V. BioAnalyst software (Applied Biosystems) was used

for spectral deconvolution with a mass range of m/z 500-2000 and an output range was

27800 to 28600 Da using a step mass of 0.1 Da and a S/N threshold of 2.

2.1.8 Effect of α-L-Glutamyl Hydroxamate and Hydroxylamine on E. coli

A single colony of *E. coli* BL21 cells was picked and inoculated into 5 mL of LB

medium. The culture was incubated at 37 °C and grown overnight to stationary phase.

To each 5 mL of fresh LB medium, a 25-μL culture was inoculated, and γ-L-glutamyl

hydroxamate or hydroxylamine was added to a final concentration, ranging from 0 to 5

mM. Cell growth at 37 °C was monitored hourly by measurement of the absorbance at

600 nm.

To test the effect of STM2437 on E. coil growth in the presence of γ-L-glutamyl

hydroxamate or hydroxylamine, a 25-μL stationary phase culture of *E. coli* with wild-

type STM2437 or the C91S mutant was inoculated in 5 mL of LB medium containing 50

μg/mL kanamycin, with addition of γ-L-glutamyl hydroxamate or hydroxylamine.

Protein expression was induced with 0.1 mM IPTG. Cell growth at 37 °C was

monitored by measuring the optical density at 600 nm of a 250-µL culture using a 96-well plate.

2.1.9 Salmonella *Growth Conditions*

All experiments utilizing *Salmonella* isolates used HA420 (or its derivatives), a spontaneous nalidixic acid resistant isolate of *Salmonella* enterica serotype Typhimurium ATCC14028 (*63*). A mutant with a deletion of STM2437 was generated in ATCC14028 by lambda red recombinase-mediated homologous recombination (*25*). Mutation was moved to HA420 by P22 phage transduction (*64*) and the resulting strain was named [HA1564, ΔSTM2437::Kan]. Antibiotics were used as appropriate at the following concentrations: nalidixic acid (50 mg/mL) and kanamycin (50 mg/mL).

2.1.10 *Effect of γ-L-Glutamyl Hydroxamate on* Salmonella *Growth*

To test the effect of γ-L-glutamyl hydroxamate on the growth of both wild-type *Salmonella* Typhimurium and the ΔSTM2437 mutant, stationary phase cultures of HA420 or HA1564 were diluted 1:100 into 5 mL of LB broth containing a range of concentrations of γ-L-glutamyl hydroxamate (0.1 mM, 0.25 mM, 0.50 mM, 1.0 mM, 2.0 mM, 5.0 mM, and 10 mM). Diluted cultures were incubated at 37 °C with aeration. Growth of both strains was measured at 10 min intervals by measuring the optical density at 600 nm using a Varioskan Flash plate reader (Thermo Scientific). Samples were run in triplicate in each experiment, and experiments were repeated on at least three separate occasions.

To test the effect of the presence of γ-L-glutamyl hydroxamate on *Salmonella* growth in both aerobic and anaerobic conditions, stationary phase cultures of HA420 and HA1564 were diluted 1:100 either in LB broth, or in LB broth containing 5.0 mM γ-L-glutamyl hydroxamate. For experiments carried out in anaerobic conditions with an internal atmosphere of 5% H_2, 5% CO_2, and 90% N_2, bacteriologic media with or without γ-L-glutamyl hydroxamate was pre-incubated in anaerobic conditions for 24 h prior to inoculation. Growth of each strain was evaluated by removing aliquots of culture at hourly intervals for serial dilution and plating to determine colony forming units (CFU) of each strain under each condition. These experiments were repeated on at least three separate occasions. Bacterial generation number was calculated using the following equation: $[\log_{10}(\text{CFU final}) - \log_{10}(\text{CFU start})]/\log_{10}$ (*65*).

2.1.11 Murine Experiments

To evaluate the ability of the ΔSTM2437 mutant to colonize and grow in murine models we used both a murine model for *Salmonella* bacteremia (C57Bl6 mice), and a murine colitis model (C57Bl6 mice pretreated by gavage with 20 mg of streptomycin). In the bacteremia model, separate groups of five C57Bl6 mice between 10 and 12 weeks of age (Jackson Laboratories) were inoculated with approximately 10^8 CFU of either HA420 or with HA1564 (ΔSTM2437::Kan) by gavage. Infected animals were observed twice daily and were euthanized at four days post infection. In the murine colitis model, C57Bl6 mice were treated with 20 mg streptomycin by gavage and were inoculated with approximately 10^8 CFU of either HA420 or with HA1564 (ΔSTM2437::Kan) 24 h post-

streptomycin treatment. Infected animals were monitored twice daily for signs of illness, and were euthanized four days post infection. Cecum, mesenteric lymph nodes, Peyer's patches, liver and spleen were collected from each animal, weighed, homogenized in 3 mL sterile phosphate buffered saline (PBS), serially diluted, and CFU were enumerated by plating.

2.1.12 Data Analysis

The kinetic constants for substrate hydrolysis were determined by fitting the initial velocity data using eqn. 1 using SigmaPlot 11, where v is the initial velocity, Et is the total enzyme concentration, k_{cat}/K_m is the turnover number, [A] is the substrate concentration, and K_m is the Michaelis constant.

$$v/Et = k_{cat} [A]/(K_m + [A]) \tag{1}$$

2.1.13 Sequence Similarity Network for STM2437

A sequence similarity network (*28*) for STM2437 was constructed using the EFI – Enzyme Similarity Tool at the University of Illinois (efi.igb.illinois.edu/efi-est). The nearest 10,000 protein sequences to STM2437 were identified using the BLASTp tool at NCBI. Each node in the resulting network represents a cluster of proteins sequences that are consolidated at a sequence identity of greater than 90%. Each node is linked to another node by an edge if the Blast E-value is better than 10^{-70}.

2.1.14 Homology Model for STM2437

A homology model for STM2437 was built using Swiss-Model

(http://swissmodel.expasy.org/) and the crystal structure of SMU_1228c from

Streptococcus mutans (PDB id: 3L7N). This protein has a 49% sequence identity to that

of STM2437. To position putative substrates into the active site of STM2437, the

homology model was structurally aligned with the C269S mutant of the CarA subunit of

the carbamoyl phosphate synthetase (CPS) from *E. coli* that contains L-glutamine bound

in the active site (PDB id: 1C3O). To approximate the binding of γ-L-glutamyl

hydroxamate, a hydroxyl group was manually added to the amide nitrogen of L-

glutamine in both the *syn* and *anti* orientations.

2.2 Results

2.2.1 Purification and Quaternary Structure

The gene encoding STM2437 from S. enterica was cloned and expressed in *E.

coli*, and the enzyme purified to apparent homogeneity. The estimated molecular mass

of STM2437, based on the elution volume from a calibrated gel filtration column, is

25,600 Da (Figure 19). The monomeric molecular weight of STM2437, based on the

DNA sequence, is 28,115 Da, and thus this protein is a monomer in solution.

2.2.2 Sequence Similarity Network.

Based on the constructed sequence similarity network of Type I glutamine

amidotransferases, STM2437 is located in a cluster distinct from other known glutamine

amidotransferases (Figure 16). The function of the close homologs of STM2437 remain unknown. Diverse groups with edge cutoffs at an E-value of -70 indicates different substrate preferences and functions within these homologs. Cluster 5, which includes STM2437 (~5,000 sequences), contains sequences primarily from various species of *Salmonella*, *Klebsiella* and *Streptococcus*.

2.2.3 Substrate Profile for STM2437

Since STM2437 is apparently homologous to other well-characterized L-glutamyl amidotransferases, the enzyme was incubated with a small library of L-glutamyl derivatives as a preliminary test of catalytic activity. Of the compounds tested, hydrolysis was detected with L-glutamine (1), γ-L-glutamyl methyl ester (2), γ-L-glutamyl ethyl ester (3), γ-L-glutamyl hydroxamate (4), β-L-aspartyl hydroxamate (5), γ-L-glutamyl hydrazine (6), γ-L-glutamyl nitroanilide (7), γ-L-glutamyl benzyl ester (8), γ-L-glutamyl anilide (9), γ-glutamyl ethylamide (10), and γ-L-glutamyl methylamide (11). The kinetic constants for the hydrolysis of these compounds are presented in Table 8 and the structures are shown in Scheme 6. γ-L-Glutamyl hydroxamate (4) was clearly the best substrate identified, with a value of k_{cat}/K_m of 9.6 x 10^4 M^{-1} s^{-1}. Catalytic activity was not detected ($k_{cat} \leq 0.002$ s^{-1}) for 15 other compounds and the structures of these compounds are presented in Scheme 7. STM2437 was unable to catalyze the net synthesis of γ-L-glutamyl hydroxamate (4) starting from L-glutamine (1) and an excess hydroxylamine at pH 8.0.
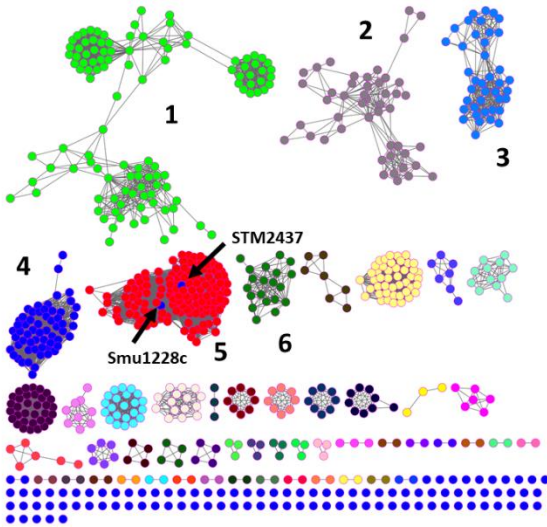
Figure 16 Sequence similarity network for the 10,000 closest homologs of STM2437.
Each node in the network represents a group of proteins that share ≥90% sequence
identity. An edge (represented as a line) is drawn between two nodes with a BLAST E-
value cutoff of better than 10^{-70}.

Table 8 Kinetic constants for the hydrolysis of various substrates by STM2437.

Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})
L-glutamine (**1**)	0.11 ± 0.01	2.5 ± 0.1	44 ± 1
γ-L-glutamyl methyl ester (**2**)	0.12 ± 0.01	1.2 ± 0.2	104 ± 12
γ-L-glutamyl ethyl ester (**3**)	0.047 ± 0.001	1.5 ± 0.1	32 ± 2
γ-L-glutamyl hydroxamate (**4**)	2.74 ± 0.07	0.028 ± 0.003	$(9.6 \pm 0.2) \times 10^4$
β-L-aspartyl hydroxamate (**5**)	0.098 ± 0.004	0.45 ± 0.07	218 ± 35
γ-L-glutamyl hydrazine (**6**)	0.94 ± 0.05	0.50 ± 0.06	$(1.9 \pm 0.2) \times 10^3$
γ-L-glutamyl nitroanilide (**7**)	ND	ND	$(1.3 \pm 0.1) \times 10^4$
γ-L-glutamyl benzyl ester (**8**)	ND	ND	294 ± 20
γ-L-glutamyl anilide (**9**)	ND	ND	176 ± 3
γ-L-glutamyl ethylamide (**10**)	ND	ND	6.8 ± 0.5
γ-L-glutamyl methylamide (**11**)	ND	ND	4.6 ± 0.1

pH 8.0, 30 ^{o}C
ND, not determined because the enzyme was not saturated at the highest substrate concentration used.

Table 9 Activity of STM2437 mutants with three substrates.

	Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})
E12D		1.3 ± 0.1	0.07 ± 0.02	$(2.0 \pm 0.1) \times 10^4$
E12Q	γ-L-glutamyl	0.54 ± 0.03	0.54 ± 0.07	$(1.0 \pm 0.1) \times 10^3$
E12G	hydroxamate	1.3 ± 0.1	0.87 ± 0.07	$(1.5 \pm 0.1) \times 10^3$
C91S		≤ 0.002	--	--
E12D		0.31 ± 0.02	8.9 ± 1.6	35 ± 2
E12Q		0.067 ± 0.003	0.8 ± 0.2	80 ± 4
E12G	γ-L-glutamine	0.18 ± 0.01	2.5 ± 0.5	71 ± 4
C91S		≤ 0.002	--	--
E12D		49 ± 2	1.0 ± 0.1	$(5.1 \pm 0.1) \times 10^4$
E12Q	γ-L-glutamyl	--	--	$(1.8 \pm 0.1) \times 10^3$
E12G	nitroanilide	60 ± 2	0.31 ± 0.03	$(2.0 \pm 0.1) \times 10^5$
C91S		≤ 0.008	--	--

2.2.4 Inactivation of STM2437 by Acivicin

Acivicin (12) has been shown previously to inactivate enzymes from the

glutamine amidotransferase superfamily by covalently reacting with the active site

cysteine residue (Scheme 8) (*62*). When 16 μM STM2437 was incubated with 80 μM L-

acivicin, greater than 90% of the catalytic activity for the hydrolysis of γ-L-glutamyl p-

nitroanilide was lost in 60 min at pH 8.0. The acivicin-inhibited enzyme was subjected

to mass spectrometric analysis in an attempt to provide direct evidence for the formation

of a covalent adduct. Based on the reported amino acid composition, the calculated

molecular mass of STM2437 is 28,115 Da, including the polyhistidine tag at the C-

terminus of the protein. A peak with an m/e of 28,116 was observed in the ESI+ mode

for the native enzyme (Figure 17A). After incubation with L-acivicin (Mw = 178.5), the

observed mass for the protein complex was 28,258, with a mass difference of 142,

relative to the unlabeled protein (Figure 17B). This difference is consistent with the

formation of a covalent adduct between STM2437 and L-acivicin and the loss of HCl as

illustrated in Scheme 8.
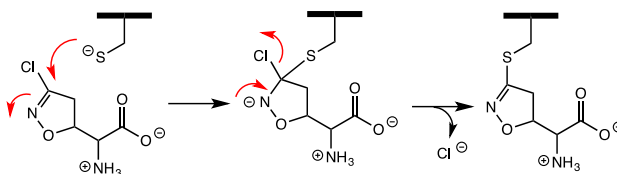
Scheme 8 Inactivation of STM2437 by acivicin.

The partition ratio for the inactivation of STM2437 by acivicin was determined by incubation of a fixed amount of enzyme with variable amounts of the inactivator for 30 min. The fraction of the remaining activity was determined as a function of the relative concentration of inactivator to enzyme and the partition ratio for complete inactivation of the enzyme was determined to be ~1.0 (Figure 18). The rate of inactivation of STM2437 as a function of the initial concentration of acivicin was determined. Up to a concentration of 100 μM acivicin there is a linear increase in the observed rate (k_{obs}) of inactivation (Figure 20). The apparent second-order rate constant (k_{inac}/K_I) is 730 ± 40 M^{-1} s^{-1}.
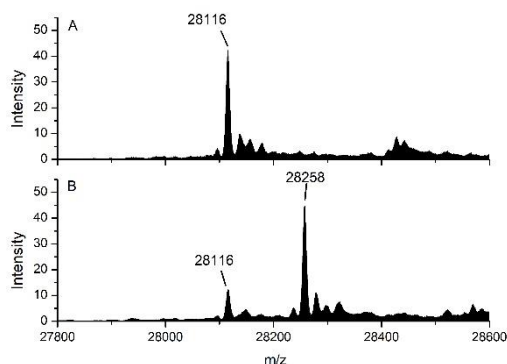
Figure 17 Mass spectra of STM2437 in the presence and absence of the enzyme inactivator L-acivicin. (A) The apparent mass of the native unlabeled enzyme is 28,116. (B) After incubation with 80 μM L-acivicin for 60 min at pH 7.5, the observed mass for the protein increased to 28,258. The mass difference of 142 is consistent with a covalent adduct formed as illustrated in Scheme 8 with the loss of HCl.
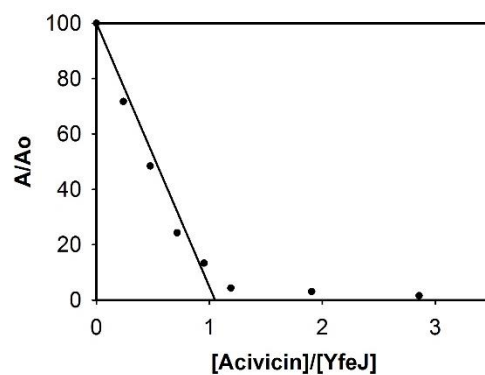
Figure 18 Inactivation of STM2437 by variation of the concentration of acivicin at a fixed concentration of enzyme. The concentration of STM2437 was 2.1 μM and catalytic activity was determined using L-glutamyl nitroanilide as the substrate. The partition ratio for the inactivation of STM2437 by acivicin was determined to be ~1.0.
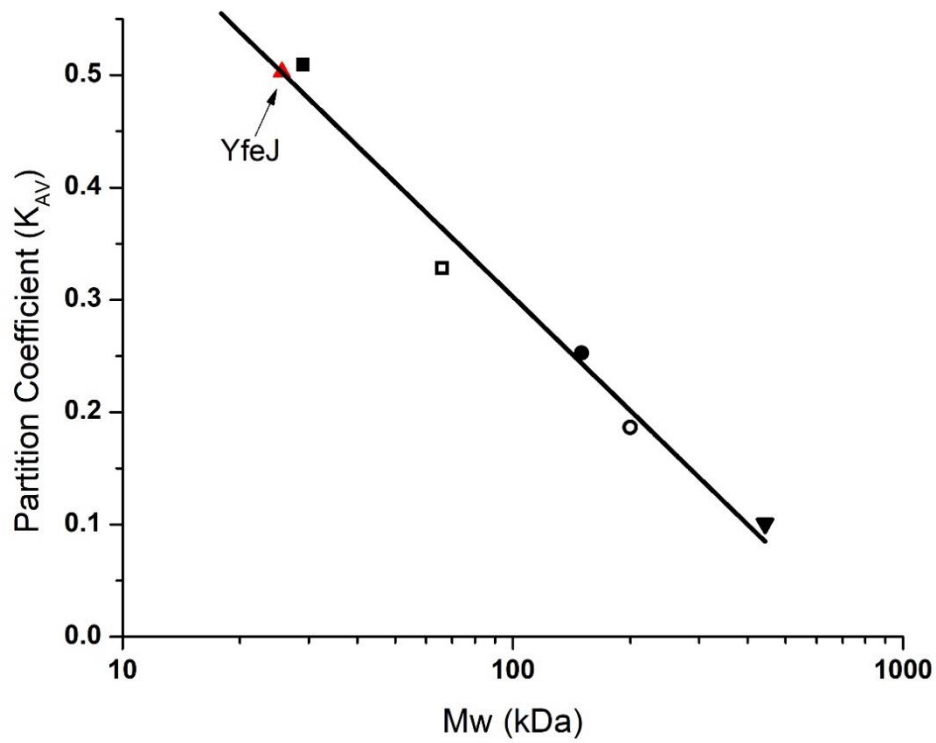
	Protein	Molecular Weight (kDa)
	carbonic anhydrase	29
☐	albumin	66
	alcohol dehydrogenase	150
	β-amylase	200
▼	apoferritin	443
▲	STM2437 (YfeJ)	26

Figure 19 Calibration curve for the determination of the molecular weight of STM2437 using gel filtration chromatography.

2.2.5 Homology Model for STM2437

A homology model of STM2437 (Figure 21) was built based on the three-dimensional crystal structure of SMU_1228c from Streptococcus mutans (PDB id: 3L7N). This protein has a 49% amino acid sequence identity to STM2437 and is the closest structural homologue currently in the Protein Data Bank. An amino acid sequence alignment of STM2437, SMU_1228c, the amidotransferase domain of GMP synthetase, and the CarA subunit of carbamoyl phosphate synthetase illustrates the conserved residues for the hydrolysis and binding of glutamine (Figure 22).
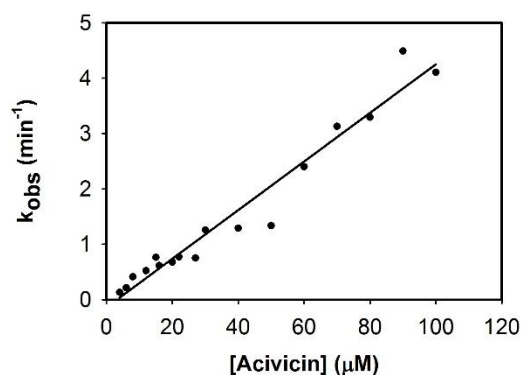
Figure 20 The effect of changing the concentration of acivicin on the apparent rate constant for the inactivation of STM2437. The second-order rate constant (k_{inact}/K_I) is 730 ± 40 M^{-1} s^{-1}.
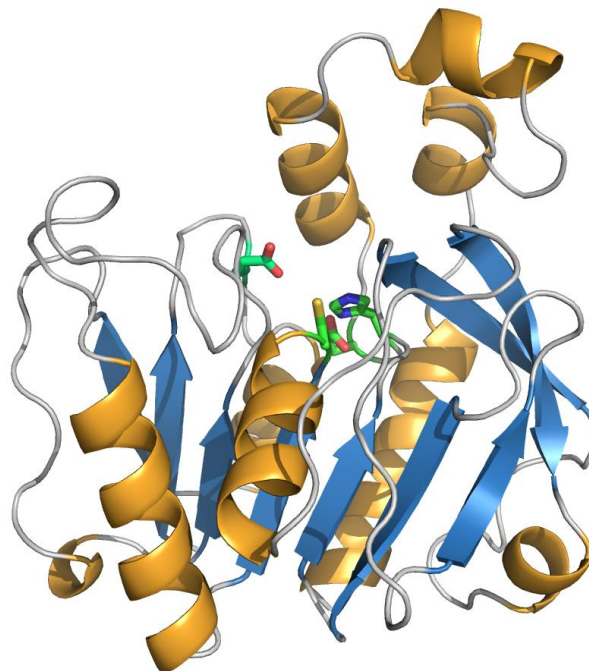
Figure 21 The homology model of STM2437 was built based on the crystal structure of SMU_1228c from Streptococcus mutans (PDB id: 3L7N), which has a three layer α–β–α flavodoxin-like fold. The α-helix is colored orange, the β-sheet colored blue and the β-strand colored gray. Catalytic residues including Glu12, Cys91, His179 and Glu181 are colored green.

```
STM2437     1 - - - - M R V H F V V H E S F E S A G A Y L K W A - - E D R G Y T I S W S R V Y A G E A L P P N A D E F   46
SMU1228C    1 - - - - M R I H F I L H E T F E A P G A Y L A W A - - A L R G H D V S M T K V Y R Y E K L P K D I D D F   46
   GuaA     1 M T E N I H K H R I L I L D F G S Q Y T Q L V A R R V R E L G V Y C - - - E L W A W D V T E A Q I R D F   49
   CarA   193 - - - - - - - - H V V A Y D F G A K R N I L - - R M L V D R G C R L - - - T I V P A Q T S A E D V L K M 231

STM2437    47 D M - - L V V F G G P Q S P R T T R E E C P Y F D S R A E Q H L I N Q A V T A R R M V I G I C L G S Q L   96
SMU1228C   47 D M - - L I L M G G P Q S P S S T K K E F P Y Y D A Q A E V K L I Q K A A K S E K I I V G V C L G A Q L   96
   GuaA    50 N P S G I I L S G G P E S T T E - - - - - - - - - - E N S P R A P Q Y V F E A G V P V F G V C Y G M Q T   91
   CarA   232 N P D G I F L S N G P G D P A P - - - - - - - - - C D Y A I T A I Q K F L E T D I P V F G I C L G H Q L 274

STM2437    97 I G E A L G A A V C Q S P E K E I G H Y - P I T L T - E A G L R H P L I A H F G - - - - S P L T V G H W 142
SMU1228C   97 M G V A Y G A D Y L H S P K K E I G N Y - L I S L T - E A G K M D S Y L S D F S - - - - D D L L V G H W 142
   GuaA    92 M A M Q L G G H V E G S N E R E F G - Y A Q V E V L T D S A L V R G I E D S L T A D G K P L L D V W M S 142
   CarA   275 L A L A S G A K T V K M K F G H H G G N H P V K D V - E K N V V M I T A Q N - - - - - - - - - - - - - 311

STM2437   143 H N D - - - M P G L T D Q A - T V L A E S E G C P R Q I V Q Y G N F V - Y G F Q C H M E F T V E A V E G 189
SMU1228C  143 H G D - - - M P G L P D K A - Q V L A I S Q G C P R Q I I K F G P K Q - Y A F Q C H L E F T P E L V A A 189
   GuaA   143 H G D K V T A - - I P S D F V T V A - S T E S C P F A I M A N E E K R F Y G V Q F H P E V T - - - - - - 185
   CarA   312 H G F A V D E A T L P A N L R V T H K S L F D G T L Q G I H R T D K P A F S F Q G H P E A S P G P - - - 360

STM2437   190 L I Q H S Q Q E L A D A Q G K R F I R S V A E M R A W N Y - Q Q M N E K L W R F L D E L T L A H S Q K   239
SMU1228C  190 L I A Q E D D L D T Q S Q T E T Y V Q T A E E M Q T F D Y - S S M N Q A L Y S F L D R L T E R K - - -   236
   GuaA   186 - - - - - - - - - - - - - - - - - - - - - - - - - H T R Q G M R M L E R F V R D I C Q C E - - - -   205
   CarA   361 - - - - - - - - - - - - - - - - - - - - - - - - - - - H D A A P L F D H F I E L I E Q Y R K T -   380
```

Figure 22 Amino acid sequence alignment comparison of STM2437, the amidotransferase domain of GMP synthetase (GuaA) from *E. coli* and SMU_1228c from *Streptococcus* mutans and CarA from *E. coli*.

The crystal structure of the C269S mutant of the CarA subunit of carbamoyl phosphate synthetase (PDB id:1C3O) has the L-glutamine substrate bound in the active site (*66*). This structure provides the molecular details for the binding of L-glutamine in the active site of glutamine amidotransferases and may provide insight into the potential interactions contained within STM2437 for the binding of the best substrate, γ-L-glutamyl hydroxamate (4). In the active site of the CarA subunit of CPS, the α-amino group of L-glutamine interacts with the backbone carbonyl groups of Gly241 (2.8 Å), Pro242 (3.9 Å) and Gly243 (2.7 Å); the α-carboxylate group interacts with the backbone

74

of Phe314 (2.8 Å) and the side chain amide nitrogen of Gln273 (3.3 Å); the γ-carbonyl

interacts with the backbone amide nitrogen of Gly241 (2.4 Å); and the γ-amide interacts

with the backbone carbonyl of Asn311 (2.6 Å) (Figure 23A).

The model for the binding of γ-L-glutamyl hydroxamate in the active site of

STM2437 is shown in Figure 23B. The model is based on the molecular interactions for

the binding of L-glutamine in the active site of CPS and the corresponding homology

model of STM2437. For this model we have utilized the syn-conformation of the

hydroxamate moiety since the anti-conformation is sterically hindered by Trp142. In

this model the α-amino and α-carboxylate groups of the substrate can make the same

backbone interactions as observed in the crystal structure of L-glutamine bound in the

active site of CPS. Of particular interest are the potential interactions between the

hydroxamate moiety of the substrate and the enzyme. The most likely interaction is

between the hydroxyl group of the hydroxamate moiety and the side chain carboxylate

of Glu12 of STM2437. This glutamate is conserved in Clusters 1, 2, 5, and 6 of the

sequence similarity network (Figure 16), but in CPS and GMP synthetase this residue is

a glycine.
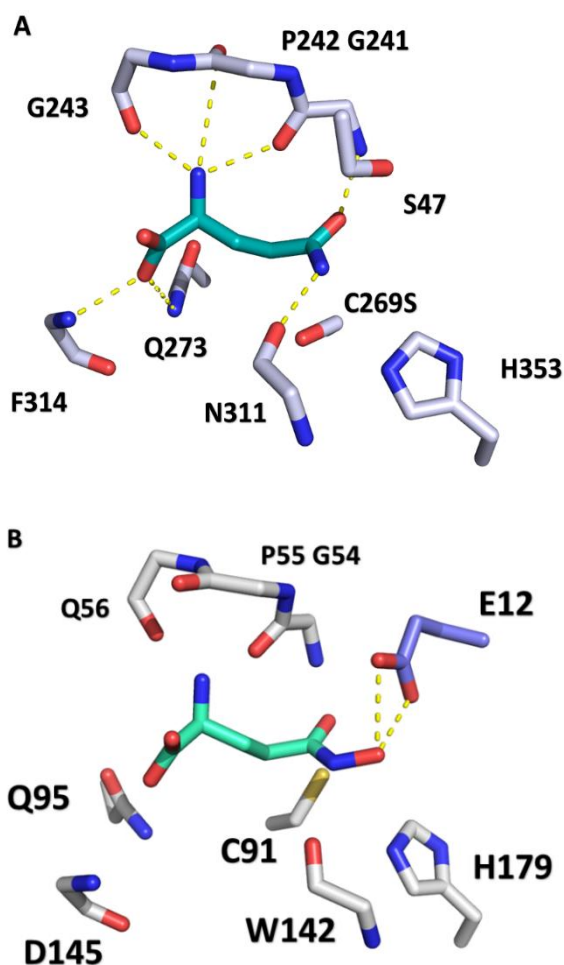
Figure 23 Interaction of glutamine in the active site. (A) L-Glutamine bound in the active site of the small subunit (CarA) of carbamoyl phosphate synthetase in the C269S mutant (PDB id: 1C3O). (B) Proposed binding of γ-L-glutamyl hydroxamate (4) in the active site of the homology model of STM2437. Glu12 is proposed to interact with the hydroxyl group of the hydroxamate moiety of the substrate.

76

2.2.6 Site-Directed Mutagenesis

The substitution of Glu12 with other residues in the active site of STM2437 was performed to assess the importance of this residue for substrate recognition. The kinetic constants for the hydrolysis of three substrates (L-glutamine, γ-L-glutamyl hydroxamate, and γ-L-glutamyl nitroanilide) were determined for the E12D, E12Q, and E12G mutants (Table 7). When Glu12 is mutated to either asparagine or glycine, there is a significant increase in the Km for the hydrolysis of γ-L-glutamyl hydroxamate and a corresponding decrease in the value of k_{cat}/K_m by about two orders of magnitude. Relatively smaller changes are observed in the values of k_{cat}/K_m for the hydrolysis of L-glutamine. However, when Glu12 is mutated to glycine the value of k_{cat}/K_m for the hydrolysis of the nitroanilide derivative increased by an order of magnitude and the K_m decreased to 31 µM, presumably because the active site is better able to accommodate the larger leaving group.

The substitution of Cys91 with serine in the active site of STM2437 was performed to inactivate STM2437. As anticipated, no activity was detected using L-glutamine, γ-L-glutamyl hydroxamate ($k_{cat} \leq 0.002$ s^{-1}) or γ-L-glutamyl nitroanilide ($k_{cat} \leq 0.008$ s^{-1}) as potential substrates. These results are consistent with Cys91 as being the nucleophile that attacks the amide carbonyl of the substrate.

2.2.7 Growth of E. coli *in the Presence of γ-L-Glutamyl Hydroxamate or Hydroxylamine*

The effect of glutamyl hydroxamate and hydroxylamine on the growth of *E. coli* was tested over the concentration range of 0.1- 5.0 mM (Figure 24). Inhibition in the rate of growth was observed at ~0.4 mM γ-L-glutamyl hydroxamate or ~0.3 mM hydroxylamine. To test the effect of expression of STM2437 on the growth of *E. coli* in the presence of γ-L-glutamyl hydroxamate or hydroxylamine in the medium, wild-type STM2437 and the C91S mutant were expressed in *E. coli* in the presence of varying concentrations of glutamyl hydroxamate or hydroxylamine (Figure 25). The results show that the expression of wild-type STM2437 can facilitate the growth of *E. coli* in the presence of glutamyl hydroxamate in the growth medium, when compared with cells expressing the inactive C91S mutant of STM2437. There was no significant difference in cellular growth in the medium containing hydroxylamine, with either wild-type STM2437 or the inactive C91S mutant.
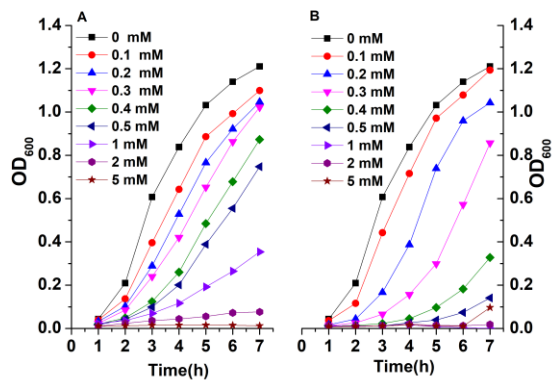
Figure 24 Effect of varying the concentration of γ-L-glutamyl hydroxamate. (A) and hydroxylamine (B) on the growth of *E. coli*. Additional details are provided in the text.
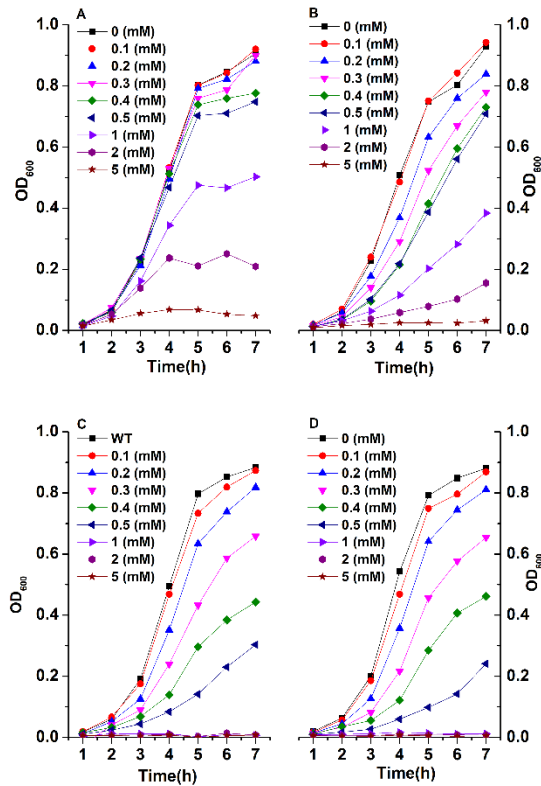
Figure 25 Effect of STM2437 and the C91S mutant on the growth of *E. coli* in the presence of γ-L-glutamyl hydroxamate or hydroxylamine. (A) wild type STM2437 and variable amounts of γ-L-glutamyl hydroxamate; (B) C91S mutant of STM2437 and variable amounts of γ-L-glutamyl hydroxamate; (C) wild type STM2437 and variable amounts of hydroxylamine; (D) C91S mutant of STM2437 and variable amounts of hydroxylamine. Additional details are provided in the text.

80

2.2.8 Effects of γ-L-Glutamyl Hydroxamate on Salmonella *Growth* in vitro

To determine the effect of γ-L-glutamyl hydroxamate on *Salmonella* replication, wild-type *Salmonella Typhimurium* ATCC14028 derivative HA420 was grown in varying concentrations of γ-L-glutamyl hydroxamate and the optical density of these cultures was measured at 600 nm at 10 min intervals (Figure 27). At concentrations of 5 mM γ-L-glutamyl hydroxamate, S. Typhimurium had mild, but statistically significantly reduced growth relative to organisms grown in LB broth alone (Figure 27A). Concentrations of γ-L-glutamyl hydroxamate below 5 mM had no significant detrimental effect on S. Typhimurium growth.

We also evaluated the growth of mutants lacking STM2437 in the presence and absence of γ-L-glutamyl hydroxamate. In bacteriologic media, the STM2437 mutant grew indistinguishably to the otherwise isogenic wild-type organism in both aerobic and anaerobic conditions. Concentrations of γ-L-glutamyl hydroxamate below 2 mM did not affect the growth of the STM2437 mutant. However, at 5 mM γ-L-glutamyl hydroxamate, the growth of the STM2437 mutant was severely inhibited compared to the growth of this mutant in the absence of γ-L-glutamyl hydroxamate and as compared to the growth of the wild-type organism grown in similar conditions (Figure 27A-D). Furthermore, this reduction in growth in the presence of 5 mM γ-L-glutamyl hydroxamate occurred irrespective of whether the growth conditions were aerobic or anaerobic (Figure 27E-F). Thus, our data suggest that mutants lacking STM2437 are very sensitive to growth inhibition by γ-L-glutamyl hydroxamate at concentrations of 5

mM and above, and the presence of an intact STM2437 gene reverses this growth
inhibition.

2.2.9 Murine Experiments

To evaluate the importance of STM2437 during infection, we used murine
models of both bacteremia and colitis. In both models, the STM2437 mutant colonized
and was equally virulent to the otherwise isogenic wild-type organism (Figure 26).
These findings suggest that STM2437 does not play a role in *Salmonella* colonization of
these two infection models.
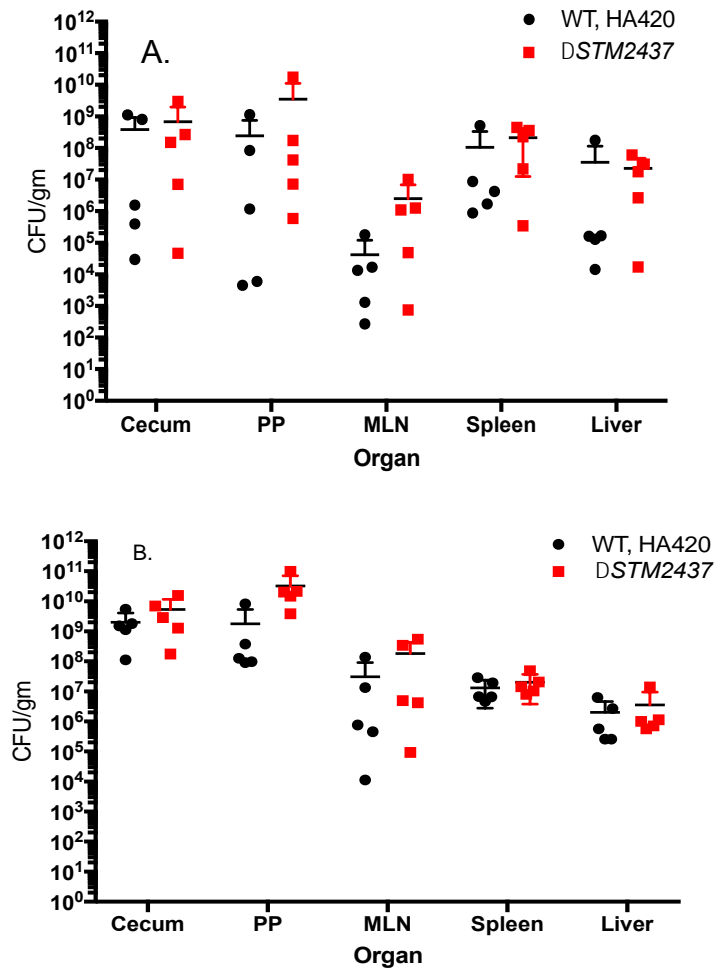
Figure 26 STM2437 does not provide a colonization advantage in murine models of bacteremia and intestinal colonization. (A) Bacteremia model: Groups of C57Bl6 mice were infected with either wild type *Salmonella* (HA420) or the ΔSTM2437 mutant (HA1564, ΔSTM2437::kan) by gavage. Four days post-infection mice were humanely euthanized and bacteria were enumerated from the cecum, Peyer's patches, mesenteric lymph node (MLN), spleen and liver. (B) Murine colitis model: Groups of C57Bl6 mice were treated with streptomycin and infected the following day with either wild type *Salmonella* (HA420) or the ΔSTM2437 mutant (HA1564, ΔSTM2437::kan) by gavage as described in the materials and methods. Four days post-infection mice were humanely euthanized and bacteria were enumerated from the cecum, Peyer's patches, mesenteric lymph node (MLN), spleen and liver. Bars indicate the mean and error bars indicate standard error. Statistical significance was determined by using a Student's two-tail t-test.
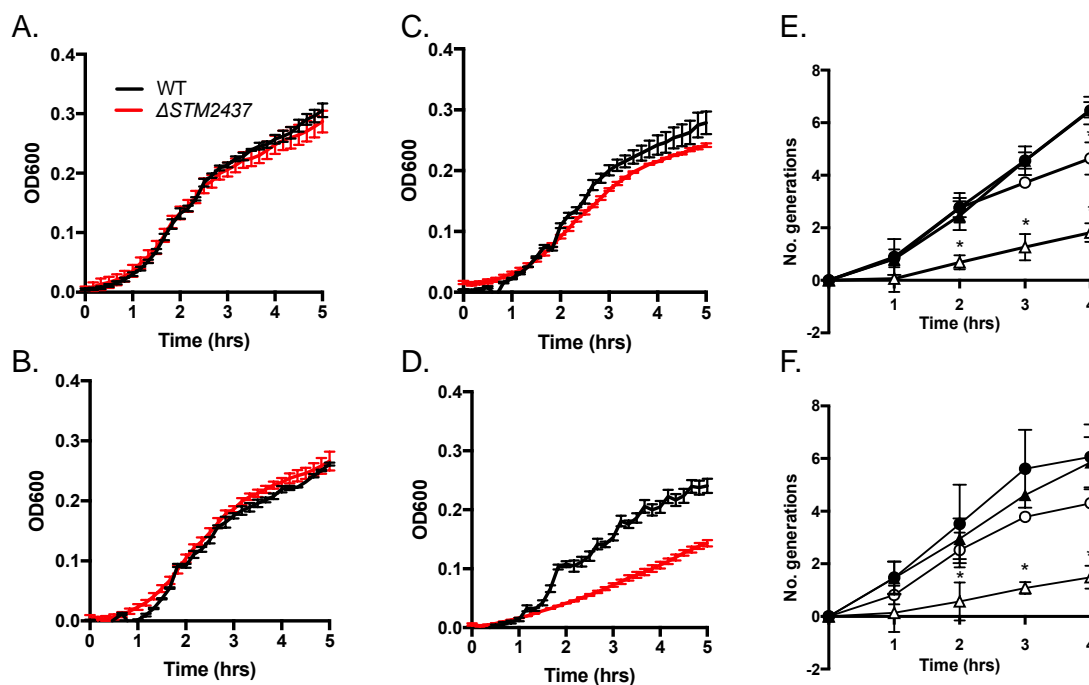
Figure 27 Effect of varying the concentration of γ-L-glutamyl hydroxamate on HA420 and HA1564 (ΔSTM2437::kan) during aerobic growth. (A) 0 γ-L-glutamyl hydroxamate; (B) 1.0 mM γ-L-glutamyl hydroxamate; (C) 2.0 mM γ-L-glutamyl hydroxamate; and (D) 5.0 mM γ-L-glutamyl hydroxamate. Growth of the wild-type (HA420) and the ΔSTM2437 mutant with (open symbols) or without (filled symbols 5.0 mM γ-L-glutamyl hydroxamate was more carefully evaluated in both aerobic (E) and (F) anaerobic conditions. HA420 represented by the circular markers, while HA1564 is represented in the triangular markers.

2.3 Discussion

STM2437 was shown previously to be under selection in ligated ileal loops in calves (*27*), but the catalytic or regulatory function of this protein was not previously studied. In this work we show that STM2437 does not appear to be important for *S.* Typhimurium growth during bacteremia in mice or in the mouse colitis model; two models commonly employed to study *S.* Typhimurium infection biology. These data leave the question of the importance of STM2437 during infection as yet unresolved.

Regarding the potential enzymatic function(s) of STM2437, sequence comparisons with homologous proteins clearly indicate that STM2437 is related to a family of enzymes that catalyze the hydrolysis of either L-glutamine or modified γ-L-glutamyl derivatives (*32*). With most of the enzymes from the amidotransferase family, glutamine is hydrolyzed to generate ammonia that is subsequently captured by a synthetase domain or subunit to react with an activated intermediate and formation of a new carbon-nitrogen bond. Carbamoyl phosphate synthetase and GMP synthetase are two well-characterized examples (*33, 66*). There is no obvious protein partner encoded in the genomic neighborhood of STM2437 that could function as a synthetase domain/protein for the biosynthesis of a metabolite using the ammonia from the hydrolysis of L-glutamine (Figure 28).

Since an obvious protein partner was not identified for STM2437 we concluded that this protein most likely catalyzes the hydrolysis of L-glutamine or a γ-L-glutamyl derivative. It is unlikely that the physiological substrate for STM2437 is L-glutamine since the value of k_{cat}/K_m is only 20 M^{-1} s^{-1}. Previously, enzymes have been identified

that catalyze the hydrolysis of 4-(γ-L-glutamylamino)butanoate (PuuD), NDP-L-glutamine (Cj1417c), γ-glutamyl-alaninol (IpuF), γ-L-glutamyl-hydroxamate (TsnB9), γ-L-glutamyl-glycine (Atu2144) and γ-L-glutamyl-ε-lysine (GGACT) (*38, 67-70*).

Of the compounds tested for catalytic activity, the best substrate was found to be γ-L-glutamyl hydroxamate. This compound is hydrolyzed by STM2437 with a k_{cat}/K_m of nearly 10^5 M^{-1} s^{-1}, which is significantly faster than any of the other compounds tested. We constructed a homology model of the three-dimensional structure of STM2437 based on the sequence identity of a protein of unknown function whose three-dimensional structure was previously deposited in the PDB. We were able to successfully dock the best substrate into the active site based on the structure of L-glutamine bound in the active site of the CarA subunit of carbamoyl phosphate synthetase from *E. coli* (*66*). This exercise identified a candidate residue (Glu12) that appeared positioned to interact with the hydroxamate moiety of the substrate. This residue is conserved in all of the closest relatives identified in the sequence similarity network (Figure 16) and in those proteins that function to hydrolyze L-glutamine as the physiological substrate this residue is conserved as a glycine.

For all of the enzymes from the amidotransferase family of enzymes, L-glutamine is hydrolyzed via a thioester intermediate(*32*). For STM2437 we propose that the conserved Cys91 is activated by His179 to make a nucleophilic attack at the amide carbonyl forming a tetrahedral intermediate that collapses with the formation of the thioester intermediate and hydroxylamine. The thiolate anion is regenerated by the nucleophilic attack of water, which is activated via proton abstraction by His179.

Support for this proposal is provided by the inactivation of STM2437 by mutation of Cys91 to serine, and the inactivation of STM2437 by the mechanism-based inhibitor acivicin. The C91S mutant of STM2437 mutant exhibited undetectable catalytic activity toward γ-L-glutamyl hydroxamate and L-glutamine. Acivicin was also found to be a potent inactivator of STM2437 and the proposed covalent confirmed by mass spectrometry.

Recently, Kuzuyama et al. identified γ-L-glutamyl hydroxamate as involved in the biosynthesis of trichostatin A (*69*). In this pathway L-glutamine is oxidized by a monooxygenase to generate γ-glutamyl hydroxamate. The γ-L-glutamyl hydroxamate is further hydrolyzed by a type II amidotransferase, where the hydroxylamine is transferred to a synthase domain to condense with an adenylated carboxylate group of trichostatin acid (*69*). Kaysser et al. identified the hydroxamate moiety in actinonin as being responsible for the inhibition of metallo-proteases because of its ability to chelate divalent metals in the active site (*71, 72*). The proposed monooxygenase is a homolog of the enzyme involved in the biosynthesis of trichostatin A. However, no homologue of these proposed monooxygenases could be identified within the genome of *Salmonella*. Recently, van der Donk et al. identified γ-L-glutamyl hydrazine being involved in the biosynthesis of fosfazinomycin and kinamycin (*73*). They showed that one of the nitrogen atoms in hydrazine originates from nitrous acid. The hydrazine moiety is further acetylated and subsequently transferred through a glutamyl carrier into the biosynthetic pathways of fosfazinomycin and kinamycin.

To understand the effect of γ-L-glutamyl hydroxamate on *Salmonella* growth, we evaluated *Salmonella* growth in vitro in different concentrations of γ-L-glutamyl hydroxamate. Although the growth of wild-type *Salmonella* is mildly inhibited by γ-L-glutamyl hydroxamate in vitro, mutants lacking STM2437 are severely inhibited. This result suggests that STM2437 allows *Salmonella* to grow in the presence of γ-L-glutamyl hydroxamate, likely by hydrolyzing this compound. Although we can only speculate as to why *Salmonella* would possess a mechanism to hydrolyze γ-L-glutamyl hydroxamate, an analog of glutamine, *Salmonella* senses environmental nitrogen limitation via limitations in its internal glutamine pool (*74*). Perhaps excess glutamyl hydroxamate perturbs this sensing, somehow interfering with the regulation of nitrogen metabolism in the cell. Furthermore, work in other organisms suggests that glutamyl hydroxamate can also interfere with differentiation and motility normally mediated by glutamine (*75*). Thus it is formally possible that γ-L-glutamyl hydroxamate may interfere with glutamine-mediated phenotypes in *Salmonella*. These hypotheses remain to be explored in future studies.

Figure 28 Genomic context of *Salmonella* typhimurium LT2, aligned by the sequence similarity of YfeJ in different organisms. (http://www.microbesonline.org/). The genomic neighborhood of yfeJ is constructed through Microbesonline. The gene yfeJ (STM2437) of *Salmonella* typhimurium LT2 is selected and shown in the tree mode. Homologs are colored in the same color. Gray color indicates no homolog in the current view of the genomic neighborhood. The gene yfeJ is colored in cyan.

89

CHAPTER III

DECIPHERING THE ENZYMATIC FUNCTION OF THE BOVINE ENTERIC

INFECTION RELATED PROTEIN STM3780 FROM *SALMONELLA ENTERICA*

SEROTYPE TYPHIMURIUM

To understand the function of the unknown aldolase and its potential role in the carbohydrate metabolism, here we describe the functional characterization of STM3780. We also show an unusual catalytic activity associated with STM3780 that condensates two molecules of DHAP to form dendroketose bisphosphate, a branched-chain monosaccharide.

3.1 Materials and Methods

3.1.1 Materials

All chemicals were purchased from Sigma Aldrich, or Carbosynth, unless indicated otherwise. **Compound** (**8, 9, 11, 12**) was synthesized based on a previously reported method (*53*). The structures of the compounds that were found to be substrates are presented in **Scheme 9** and those that were found not to be substrates are found in **Scheme 10**. The methods for the synthesis of compounds that are not commercially available are found in the Supplementary Information.

glyoxylic acid
1

glycoaldehyde
phosphate
2

dihydroxyacetone
phosphate
3

D/L-glyceraldehyde
4

D-glyceraldehyde-
3-phosphate
5

L-glyceraldehyde-
3-phosphate
6

methyl glyoxal
7

L-threoteturonate
8

L-threose-4-P
9

D-xyluluronic
11

L-araburonate
12

D-erythrose-4-P
10

Scheme 9 Substrate list.

13
CHO
CH₃ → CHO, CH_3

Scheme structures (Fischer projections):

13: CHO / CH₃

14: CHO / CH₂OH

15: CHO / —OH / CH₃

16: CHO / HO— / CH₃

17: CHO / O= / CH₃

18: CHO / ⁓OH / COOH

19: CHO / —OH / —OH / CH₂OH

20: CHO / HO— / HO— / CH₂OH

21: CHO / —OH / HO— / CH₂OH

22: CHO / —OH / —OH / COOH

23: CHO / HO— / —OH / COOH

24: CHO / —OH / —OH / —OH / CH₂OH

25: CHO / HO— / —OH / —OH / CH₂OH

26: CHO / —OH / HO— / —OH / CH₂OH

27: CHO / —OH / —OH / HO— / CH₂OH

28: CHO / —OH / HO— / HO— / CH₂OH

29: CHO / HO— / —OH / HO— / CH₂OH

30: CHO / —OH / HO— / HO— / CH₂OH

31: CHO / —OH / —OH / —OH / CH₂OPO₃H₂

32: CHO / HO— / —OH / —OH / CH₂OPO₃H₂

33: CHO / —OH / —OH / —OH / COOH

34: CHO / HO— / —OH / —OH / COOH

35: CHO / —OH / HO— / —OH / COOH

36: CHO / —OH / —OH / HO— / COOH

37: CHO / —OH / HO— / —OH / —OH / CH₂OH

38: CHO / —OH / HO— / HO— / —OH / CH₂OH

39: CHO / HO— / HO— / —OH / —OH / CH₂OH

40: CHO / HO— / —OH / HO— / —OH / CH₂OH

41: CHO / —OH / —OH / HO— / —OH / CH₂OH

42: H—C=O / HO—H / H—OH / H—OH / HO—H / CH₂OH

43: H—C=O / H—OH / H—OH / HO—H / HO—H / CH₂OH

44: H—C=O / H—OH / HO—H / H—OH / HO—H / CH₂OH

45: H—C=O / HO—H / HO—H / H—OH / HO—H / CH₂OH

46: CHO / HO—H / HO—H / H—OH / H—OH / CH₃

47: CHO / H—OH / HO—H / HO—H / H—OH / CH₃

48: CHO / HO—H / H—OH / H—OH / HO—H / CH₃

49: H—C=O / H—OH / HO—H / H—OH / H—OH / COOH

50: H—C=O / HO—H / H—OH / HO—H / HO—H / COOH

51: H—C=O / HO—H / HO—H / H—OH / H—OH / COOH

52: H—C=O / H—OH / HO—H / H—OH / HO—H / COOH

53: H—C=O / H—OH / HO—H / HO—H / H—OH / COOH

54: CHO / —OH / HO— / —OH / —OH / CH₂OPO₃H₂

55: CHO / —OH / HO— / HO— / —OH / CH₂OPO₃H₂

56: CHO / HO— / HO— / —OH / —OH / CH₂OPO₃H₂

57: CHO / —OH / —OH / —OH / —OH / CH₂OPO₃H₂

Scheme 10 Aldehydes tested that were not substrates for STM3780.

3.1.2 Cloning, Expression, and Purification of STM3780

The gene for STM3780 (gi|1255304|; UniProt: Q8ZL21) was amplified from *S. enterica* ATCC 700720 genomic DNA using 5'-GTCTCTCCCATGCCTTTAGTCAATGGTAGGATCCTGC -3' as the forward primer and 5'- GGTTCTCCCCAGCAGAGACTTTGCGGAATGCTGAGTGA -3' as the reverse primer. Ligation independent cloning sites that are compatible with the pMCSG28 vector were incorporated into both primers. PCR was performed using Pfu Turbo Polymerase from Agilent. The reaction conditions were 5 min at 95 °C, followed by 40 cycles of 45 s at 95 °C, 1 min at 58.3 °C, and 1 min at 68 °C. The PCR product was purified with a PCR cleanup system from Qiagen and subsequently treated with the T4 polymerase in the presence of dTTP. The resulting DNA product was ligated into a pMCSG28 vector obtained from the laboratory of Dr. James Sacchettini (Texas A&M University). The pMCSG28 vector was amplified with 5'-GGGGAGAACCTGTACTTCCAATCCG -3' and 5'-GGGAGAGACTCCTTCTTAAAGTTAAAC -3', both of which were phosphorylated on the 5' end through a customized order with IDT. PCR was performed using Pfu Turbo Polymerase. The reaction conditions were 5 min at 95 °C, followed by 20 cycles of 45 s at 95 °C, 1 min at 55 °C, and 7 min at 72 °C. The amplified product was subsequently treated with DpnI to remove the original template. Following purification, the PCR product was treated with the T4 polymerase in the presence of dATP. Both the vector and the insert were mixed and annealed on ice for 30 min before being added to chemically competent DH5α cells (Novagen). The mixture was transformed into

competent DH5α cells via heat shock. The plasmid with the gene of interest (*STM3780*

in pMCSG28) was transformed into BL21 (DE3) cell via electroporation. Bacterial

colonies were inoculated into 5-mL cultures of LB and incubated at 37 °C, overnight.

One-liter cultures of LB medium containing 100 μg/mL ampicillin were inoculated with

the 5-mL culture and then grown at 37 °C. After 1 h of growth, 1.0 mM zinc acetate was

added to the culture. The cultures were incubated until an OD_{600} of 0.4−0.6 was

obtained. The cultures were subsequently incubated at 20 °C, and expression was

initiated by the addition of 1.0 mM IPTG. The cells were harvested after 20 h via

centrifugation at 4 °C and then resuspended in 50 mM triethanolamine, pH 7.7, 100 mM

KCl, 1.0 mM DTT. Resuspended cells were lysed via seven rounds of sonication and

the cell debris was removed via centrifugation at 4 °C. The supernatant solution was

collected, filtered through a 0.45 μm membrane, and then applied to a Ni-NTA column,

which was equilibrated with binding buffer (50 mM triethanolamine, pH 7.7, 100 mM

KCl, 1.0 mM DTT). The column was washed with 5 column volumes of buffer (50 mM

triethanolamine, pH 7.7, 100 mM KCl, 20 mM imidazole and 1.0 mM DTT). The

protein was eluted with a solution containing 50 mM triethanolamine, pH 7.7, 100 mM

KCl, 300 mM imidazole and 1.0 mM DTT. The protein fractions were collected based

on the absorbance at 280 nm. The concentration of STM3780 (MW = 33.2 kDa) was

determined by the absorbance at 280 nm using a calculated molar extinction coefficient

of 14,400 M^{-1} cm^{-1}. The purity of the protein was confirmed by SDS-PAGE. The

purified protein was stored at 4 °C immediately. The enzyme used for dendroketose

94

bisphosphate biosynthesis was purified in 50 mM NH_4HCO_3, pH 7.2, 100 mM KCl, 1.0 mM $ZnSO_4$ eluted with a gradient of imidazole from 30 mM to 300 mM.

3.1.3 STM3780-Catalyzed Hydrogen-Deuterium Exchange

STM3780 was exchanged into 25 mM imidazole-D_2O buffer using Vivaspin 500 concentrators with 10 kDa cut-off filters (GE Healthcare Bio-Sciences) and multiple washes of 25 mM imidazole-D_2O buffer, pH 7.4. Potential substrates (10 mM) for STM3780 were mixed with enzyme (2.0 µM) in 20 mM imidazole-D_2O buffer, pH 7.4. The compounds assayed for catalytic activity included dihydroxyacetone phosphate, dihydroxyacetone and pyruvate. Potential exchange of the proton with deuterium from solvent was monitored by ^{1}H NMR spectroscopy. For measurement of the rate of solvent exchange with DHAP catalyzed by STM3780, 10 mM DHAP was mixed with 2.0 µM STM3780 and the ^{1}H NMR spectrum was collected every 3 min for 60 min. The extent of the substrate exchanged with solvent was calculated from the integrated resonance area.

3.1.4 Stereochemistry Preference of H/D Exchange in DHAP Catalyzed by STM3780

10 mM DHAP was added to STM3780 (10 µM) in 25 mM imidazole-D_2O buffer, pH 7.4. The reaction was incubated for 2 h, until the DHAP was completely exchanged with deuterium from solvent. Commercial rabbit muscle aldolase class I (Sigma-Aldrich) was diluted into 25 mM imidazole buffer to a concentration of 1 U/µL. Toward

the reaction above, 1 U of aldolase was added, and the reaction was monitored by ^{1}H

NMR spectroscopy for another 2 h.

3.1.5 Catalytic Activity of STM3780 Monitored by ^{31}P NMR Spectroscopy

Potential aldehyde substrates were screened with DHAP and STM3780 using ^{31}P

NMR spectroscopy. The aldol condensation reaction with potential substrates was

monitored by mixing 10 mM DHAP, 15 µM STM3780, and 20 mM aldehyde substrates

in 25 mM HEPES buffer, pH 7.4, and incubated for 2 h at 30 C.

Kinetic experiments were conducted under the same reaction conditions as the

original assays with a variable concentrations of STM3780 (2-20 µM), except in 50 mM

triethanolamine buffer, pH 7.7. For identified aldehyde substrates with good resolution

in ^{31}P NMR spectra, the reaction was monitored as a function of time by the ^{31}P NMR

spectroscopy to obtain the kinetic parameters. For substrates with poor resolution in ^{31}P

NMR spectra, aliquots were removed every 20 min, quenched with 1.0 mM EDTA, and

the pH adjusted to 8.4. Spectra were collected for each aliquot up to 2 h with the

product resonances quantified by integration with respect to total phosphate

concentration. Concentration of product was plotted as a function of time and fit to a

linear equation or a single exponential. The retro aldol reaction was monitored by UV

spectroscopy using a coupled assay containing 12 µM STM3780, 0.4-5 mM fructose

bisphosphate, glycerol-3-phosphate dehydrogenase (G3PD, 0.1 U), triose isomerase

(TIM, 0.1 U) and 0.4 mM NADH, in 50 mM triethanolamine buffer, pH 7.7.

3.1.6 Determination of STM3780 Stereochemistry at C4

STM3780 was used to enzymatically synthesize D-xylulonate-1-P. To synthesize

D-xylulonate-1-P, 10 mg DHAP (0.06 mmol) and 14 mg sodium glyoxylate (**1**) (0.12

mmol) were dissolved in 1.4 mL H_2O and 400 µL of 250 mM NH_4HCO_3 (pH 7.3).

Then, 0.2 mL of 500 µM STM3780 (buffer exchanged in 25 mM NH_4HCO_3) was added

to the reaction. The reaction was incubated at room temperature for 2 h. The reaction

mixture was lyophilized under reduced pressure until the volume reached ~ 600 µL. The

reaction was then combined with a similar reaction catalyzed by YdjI (*53*). The ^{13}C

NMR spectrum of the mixture was collected.

3.1.7 Stereoselective Deuterium Exchange of DHAP Protons

STM3780 was used to selectively exchange the pro-*S* hydrogen on C3 of DHAP

with deuterium. In this way, deuterium eventually was incorporated at C6 of

dendroketose bisphosphate. By comparing the ^1H NMR spectra of the product

dendroketose bisphosphate with or without deuterium, the resonance in the ^1H NMR

spectrum was assigned to pro-*S* and pro-*R* hydrogens at C6 of dendroketose

bisphosphate. The rate of hydrogen-deuterium exchange of DHAP is observed to be

much faster than the condensation rate of DHAP. A higher concentration of STM3780

was added initially to transform DHAP to deuterium-substituted DHAP. 10 mM DHAP

was added to 18 µM STM3780 in 25 mM imidazole-D_2O buffer, pH 7.4. This reaction

was incubated for 2 h. The ^1H NMR spectrum of the reaction was then collected.

3.1.8 Purification of STM3780 Products

Selected products of the STM3780 catalyzed reactions were purified so they could be further characterized or assayed. To synthesize dendroketose bisphosphate, 55 mg dihydroxyacetone (0.6 mmol) and 103 mg phosphoenolpyruvate (0.5 mmol) were dissolved in 3.0 mL of 500 mM NH_4HCO_3. After adding 250 µL of 50 mM ATP and 12.5 µL of 1.0 M $MgCl_2$, the pH was adjusted to 7.4. Then, the mixture was added to a solution containing 25 U pyruvate kinase, 2.5 U glycerol kinase, and 8.0 mL of 50 µM STM3780. The reaction was incubated at room temperature overnight. Aliquots were removed to monitor the progress of the reaction by ^{1}H NMR spectroscopy. After the reaction had reaction equilibrium, the sample was centrifuged at $21130 \times g$ for 2 min to remove precipitation. Afterwards, the reaction mixture was loaded onto an anion-exchange column (DEAE-Sephadex A-25 chloride form, equilibrated with bicarbonate). The resin was first washed with water until the pH was neutral. Then, the product was eluted with a gradient of ammonium bicarbonate (pH 8.3) up to 1.0 M. The fractions (300 mM - 500 mM ammonium bicarbonate) containing dendroketose bisphosphate were identified by the phenol-sulfuric acid test and confirmed with ^{1}H NMR spectroscopy (*76*).

To determine the structure of dendroketose bisphosphate, a 1D transient NOESY experiment was attempted (*77, 78*). However, due to the proximity between the two hydrogens on C6, the NOE cannot be used to determine the stereochemistry at C5. To obtain the ^{13}C NMR spectrum of dendroketose bisphosphate, 10 mg DHAP was added to

150 µM STM3780 in 50 mM NH_4HCO_3, pH 7.3. This reaction was incubated for three hours. The ^{13}C NMR spectrum of the reaction was collected.

3.1.9 Data Analysis

The kinetic constants for the retro-aldol reaction were determined by fitting the initial velocity data using eqn. 2 with SigmaPlot 11, where v is the initial velocity, E_t is the total enzyme concentration, k_{cat} is the turnover number, [A] is the substrate concentration, and K_m is the Michaelis constant.

$$v/E_t = k_{cat}\,[A]/(K_m+[A]) \tag{2}$$

3.1.10 Sequence Similarity Network for STM3780

A sequence similarity network for STM3780 was constructed using the EFI – Enzyme Similarity Tool at the University of Illinois (efi.igb.illinois.edu/efi-est) (*31*). Each node in the resulting network represents a cluster of proteins sequences that are consolidated at a sequence identity of greater than 90%. Each node is linked to another node by an edge if the E-value is better than 10^{-70}.

The gene *STM3781* was cloned from the genome of *Salmonella* into various vectors. The gene was also codon-optimized for expression in *E. coli* based on the tool provided in the IDT website (https://www.idtdna.com/CodonOpt). Constructs and corresponding primers used for amplification are listed in Table 10. The expression of STM3781 in *E. coli* was tested under various conditions including varied induction

temperatures and IPTG concentrations. Auto-induction was tested in AMRESCO ZYP-5052 media (*79*). Since STM3781 has a CXXXCXXC motif based on the translation of the gene, it indicates that STM3781 could potentially bind with an Fe-S cluster. The expression of the gene was also tested in media supplemented with $(NH_4)_2Fe(SO_4)_2$ and of L-cysteine (*80*). To express STM3781 in *E. coli*, an ampicillin resistant plasmid (pMCSG28) containing gene *STM3780* followed by a His_6 sequence at the C terminus were co-transformed with a kanamycin resistant plasmid (pET28a) containing codon optimized gene STM3781 following a SUMO sequence at the N terminus into *E. coli* BL21 strain via electroporation. The expression was tested in varied IPTG concentrations with the presence of both antibiotics.

Table 10 Primers used to clone gene STM3780 into different vectors.

Vector	Gene	Construct with Tag		Forward primer	Cloning method
pMCSG28	STM3781	No	F	CGCTTATTGGTGAAGGGGAGAACCTG	QuikChange
			R	CAGGTTCTCCCCTTCACCAATAAGCG	
pSCG23	STM3781	N-His		N/A	Obtained
pMCSG7	STM3781	N-His	F	TACTTCCAATCCAATGCCATGCCTGATT ACCACGCAGCGC	LIC
			R	TTATCCACTTCCAATGTTATTCACCAAT AAGCGCCCCTTC	
pMCSG9	STM3781	N-MBP	F	TACTTCCAATCCAATGCCATGCCTGATT ACCACGCAGCGC	LIC
			R	TTATCCACTTCCAATGTTATTCACCAAT AAGCGCCCCTTC	
pET28a	STM3781	N-SUMO	F	TCTAGGTACCGAGAACCTGTACTTCCAA TCC	KpnI & XhoI
			R	TGACCTCGAGTTATTCACCAATAAGCGC CCCTTC	
pMCSG7	STM3781	N-Smt3	F	CTGACCATGGCGACTCAGAAGTCAATC AAGAAGCTAAG	NcoI
			R	CTAGCCATGGCCAGAATGATGATGATG ATGGTGCATATG	
pMCSG7	STM3781	N-His, N-Smt3		N/A	Synthesized
pMCSG28	STM3781	C-His	F	GTCTCTCCCATGCCTGATTACCACGCAG CGC	LIC
			R	GGTTCTCCCCAGCTTCACCAATAAGCGC CCCTTCTTG	
pET28a	STM3781op	N-SUMO	F	TCTACATATGCCGGATTATCATGCCGCC	KpnI & XhoI
			R	TGACCTCGAGTTACTCCCCGATAAGTGC GCC	
pMCSG7	STM3781op	N-His		N/A	Synthesized
pMCSG9	STM3781op	N-MBP	F	TACTTCCAATCCAATGCCATGCCGGATT ATCATGCCGCC	LIC
			R	TTATCCACTTCCAATGTTACTCCCCGAT AAGTGCGCCTTC	
pMCSG28	STM3781op	C-His	F	GTCTCTCCCATGCCGGATTATCATGCCG CC	LIC
			R	GGTTCTCCCCAGCCTCCCCGATAAGTGC GCCTTC	

The gene followed by "op" indicate that the codon of the sequence has been optimized. QuikChange is a mutagenesis technique protocol provided by Agilent. Some of the constructs are synthesized by VectorBuilder. LIC is a cloning protocol developed by Mamie Z Li et al. (*81*) .

3.2 Results

3.2.1 Predication of STM3780 Function

STM3780 is assigned to COG0191, which contains class II zinc-dependent aldolases that prefer to use DHAP as one of the substrates. In a sequence similarity network generated with proteins from COG0191, STM3780 is found in a cluster with no known functions at an E-value of -70, and it is not in the same cluster with fructose bisphosphate aldolase (FbaA), tagatose bisphosphate aldolase (KbaYZ, GatYZ), 2-deoxy-5-keto-D-gluconate-6-phosphate aldolase (IolJ), or YdjI (Figure 11). Considering the genomic context with the PTS, STM3780 was predicted as a carbohydrate aldolase that forms DHAP as one of the products.

3.2.2 Stereospecific Hydrogen-Deuterium Exchange

The gene encoding STM3780 from *Salmonella* was cloned and expressed, and the enzyme was purified to homogeneity. Dihydroxyacetone phosphate, dihydroxyacetone and pyruvate were tested as potential substrates for STM3780. The hydrogen-deuterium exchange reactions were monitored via ^{1}H NMR spectroscopy, but only DHAP was observed with measurable activity. The hydrogen-deuterium exchange at C3 of DHAP was followed by the disappearance of the hydrogen resonance for one of the two prochiral hydrogens at 3.55 ppm and the appearance of a broader resonance at 3.53 ppm (Figure 29). The apparent rate constant for the hydrogen exchange reaction was 2.9 s^{-1}. To determine which of the two prochiral hydrogens at C3 of DHAP is exchanged by STM3780, rabbit muscle aldolase was added to the reaction mixture after

the hydrogen exchange was complete with STM3780. After an incubation of 2 h, no

additional hydrogen exchange with the solvent was observed. Therefore, STM3780

catalyzes the abstraction of the same prochiral hydrogen as rabbit FBP aldolase, which is

the pro-*S* hydrogen at C3 of DHAP (Scheme 11).
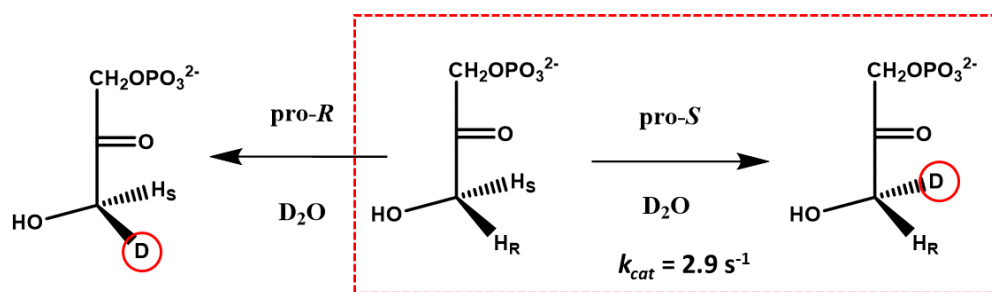
Scheme 11 H abstraction of DHAP catalyzed by STM3780.

103
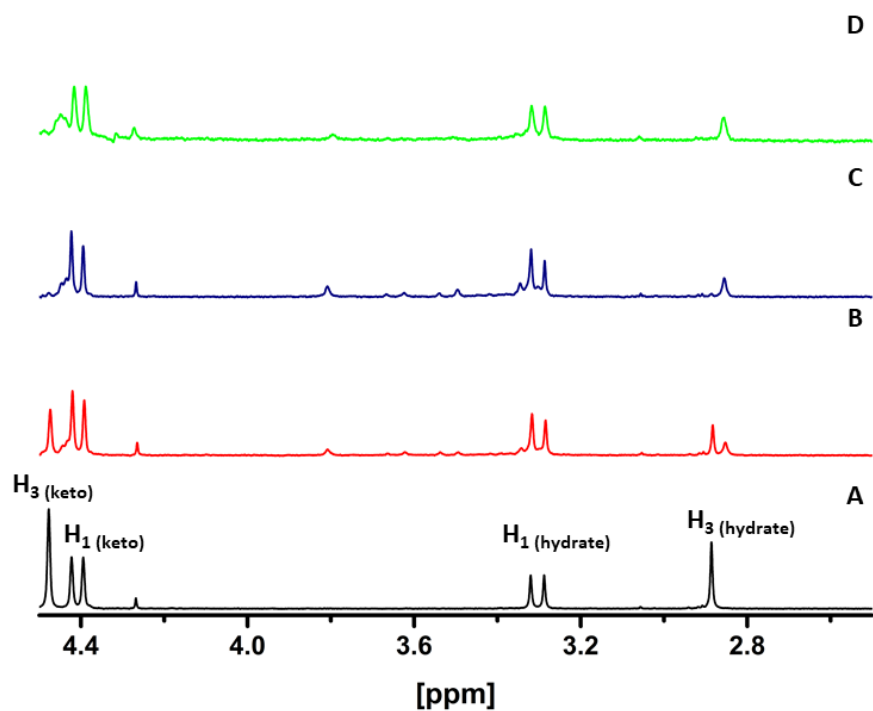
Figure 29 Solvent deuterium exchange of DHAP. (A) ^{1}H-NMR spectrum of DHAP in 25 mM imidazole-D$_2$O buffer, pH 7.4. (B) ^{1}H-NMR spectrum of DHAP (10 mM) after 27 min incubation with STM3780 (2.0 µM). (C) ^{1}H-NMR spectrum of DHAP (10 mM) after 51 min incubation with STM3780 (2.0 µM). (D) ^{1}H-NMR spectrum of DHAP (10 mM) after 2 h incubation with STM3780 (10 µM) and another 2 h incubation with FBP aldolase (1.0 U).

3.2.3 Stereochemistry at C4 of the Product after Aldol Condensation

The stereochemistry at C4 after aldol condensation with an aldehyde remained to be determined. In the direction of aldol condensation, the chirality at C4 of the product is distinct in condensations catalyzed by fructose bisphosphate aldolase or tagatose bisphosphate aldolase. In order to determine the stereochemistry at C4 of the product, STM3780 was mixed with DHAP and glyoxylic acid to generate either D-xylulonate-1-P or L-ribulonate-1-P. Only one major product was shown in the ^{31}P NMR spectrum. The ^{13}C NMR spectrum was compared with the spectrum of D-xylulonate. To demonstrate D-xylulonate-1-P is the product, the reaction with DHAP and glyoxylic acid catalyzed by STM3780 was mixed with the reaction catalyzed by YdjI. The ^{13}C NMR spectrum did not show any different resonances (Figure 30). The product of the reaction was further dephosphorylated to compare its ^{13}C NMR spectrum with the spectrum that was previously reported for D-xylulonate (*82*). The result indicated that the initial product was D-xylulonate-1-P. Therefore, the STM3780 catalyzed the condensation of DHAP and glyoxylic acid to produce D-xylulonate-1-P, with formation of an *S* stereocenter at C3 and an *S* stereocenter at C4, in a manner similar to YdjI and fructose bisphosphate aldolase.
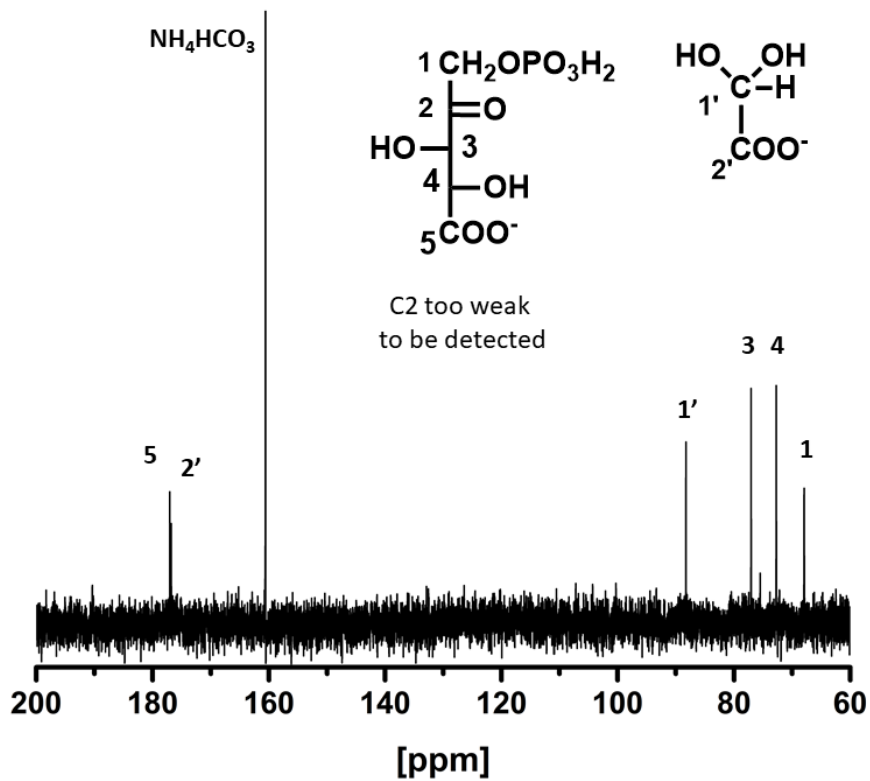
Figure 30 Products of DHAP and glyoxylic acid catalyzed by STM3780 or YdjI were mixed in a ratio of 2:1. The ^{13}C NMR spectrum was acquired with 1600 scans, setting AQ=0.5 s. Resonance belong to the carbonyl group C2 was not observed. Resonances at 177 ppm and 88 ppm belong to the excess of glyoxylic acid. Resonances at 68 ppm is a doublet.

3.2.4 Screening for Aldehyde Substrates in Condensation Catalyzed by STM3780 and

Rate Comparison

After establishing DHAP as the ketose part of the STM3780 catalyzed reaction, DHAP was mixed with other aldehydes in the presence of STM3780 to screen potential substrate through ^{31}P NMR spectroscopy. However, in the control experiment with only DHAP and STM3780, multiple product resonances were observed. One resonance was initially determined as 6-deoxy-5-keto-fructose-1-P, which was the product of DHAP and methyl glyoxal (**7**), a degradation product of DHAP commonly found in commercially available DHAP. Two resonances in the control experiment were unassigned.

There were 56 aldehydes and 5 ketones tested, in which 11 aldehydes and 1 ketone were identified as substrates for STM3780. Significant promiscuity was observed since the identified substrates with varied structure ranging from 2 carbons up to 5 carbons, in addition to the self-condensation of two molecules of DHAP. Turnover numbers were measured for compounds with robust activity by following changes in the ^{1}H NMR or ^{31}P NMR spectra (Table 11). The reactions were also subjected to mass spectrometry and confirmed to have expected molecular weight for the product. In the retro-aldol reaction, FBP was shown to be a relatively poor substrate for STM3780 with a k_{cat}/K_m of 28 M^{-1} s^{-1}.

Table 11 Turnover numbers for substrates with STM3780 and DHAP.[a]

Substrate	Apparent k_{cat} (s[-1])	Expected Product	Expected MW	Actual MW
dihydroxyacetone phosphate (3)	0.04	dendroketose bis-P	338.99	338.99
glyoxylic acid (1)	0.20	D-xylulonate-1-P	242.99	242.99
glycoaldehyde phosphate (2)	0.05	D-xylulose-1,5-bis-P	308.98	308.98
D-glyceraldehyde-3-phosphate (5)	0.008	D-fructose-1,6-bis-P	338.99	338.99
L-glyceraldehyde-3-phosphate (6)	0.003	L-sorbose-1,6-bis-P	338.99	338.99
methyl glyoxal (7)	0.10	6-deoxy-5-keto-D-fructose-1-P	241.01	241.01
D/L-glyceraldehyde (4)	0.02	D-fructose-1-P/ L-sorbose-1-P	259.02	259.02
L-threose-4-P (9)	0.082	L-galactoheptulose-1,7-bis-P	369.00	369.00
D-erythrose-4-P (10)	<0.01[b]	D-galactoheptulose-1,7-bis-P	369.00	369.00
L-threoteturonic (8)	0.65	L-galactoheptulonate-1-P	303.01	
L-araburonic (12)	0.01	L-*glycero*-L-*galacto*-octulonate-1-P	333.02	
D-xyluluronic (11)	0.004	D-*glycero*-L-*galacto*-octulonate-1-P	333.02	

[a]Reactions were monitored by ^{1}H NMR or ^{31}P NMR spectroscopy and quantified by integration with an internal standard or total concentration of substrates. [b]Overlap between substrate and product resonances prevented accurate integrations.

3.2.5 Structural Characterization of Selected Substrates

DHAP can form a dimer when it is crystalized (*83*). It is unlikely to form the dimer in solution, since in the ^{13}C NMR, ^{31}P NMR, and ^{1}H NMR spectra, only DHAP and its hydrated derivates were observed without the addition of STM3780. The incubation of DHAP with FBP aldolase did not result in any new resonances in the ^{31}P NMR spectrum. For DHAP dimers, one isoform is symmetric and the other isoform is asymmetric (Scheme 12). The symmetric isoform will have only a single resonance in the ^{31}P NMR spectrum, and the asymmetric isoform will have two resonances of equal intensity.
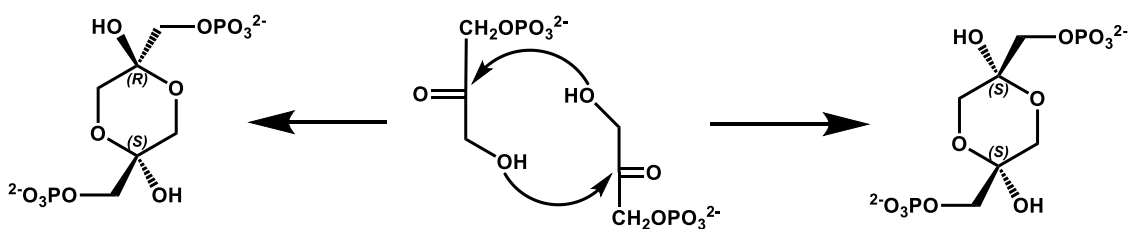

Scheme 12 DHAP can cyclize to form dimers when it is crystalized, which has two isoforms.

This pattern was not observed. Therefore, dendroketose bisphosphate was proposed as the product resulting from self-condensation of two molecules of DHAP (Scheme 13). Given the uncommon structure of branched monosaccharide dendroketose, it was of interest to more fully characterize the conformational structure of this molecule by ^{13}C NMR, ^{31}P NMR, and ^{1}H NMR spectroscopy. To obtain the ^{13}C NMR spectrum of the self-condensation product, STM3780 was exchanged into ammonia bicarbonate buffer. DHAP and STM3780 were incubated in ammonia bicarbonate buffer, and the ^{13}C NMR spectrum of the reaction mixture was acquired (Figure 31).
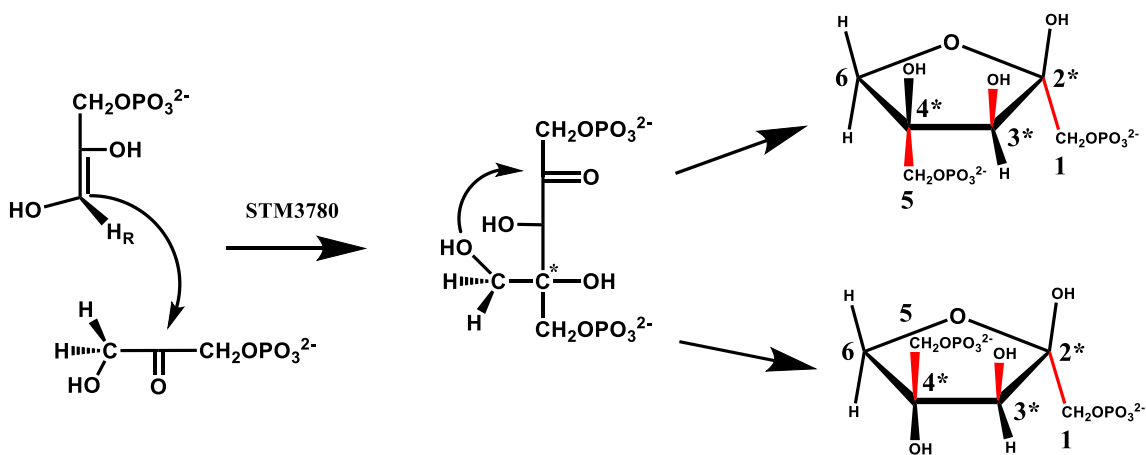
Scheme 13 The self-condensation of two molecules of DHAP catalyzed by STM3780. The product dendroketose bisphosphate is further cyclized to form a furanose.
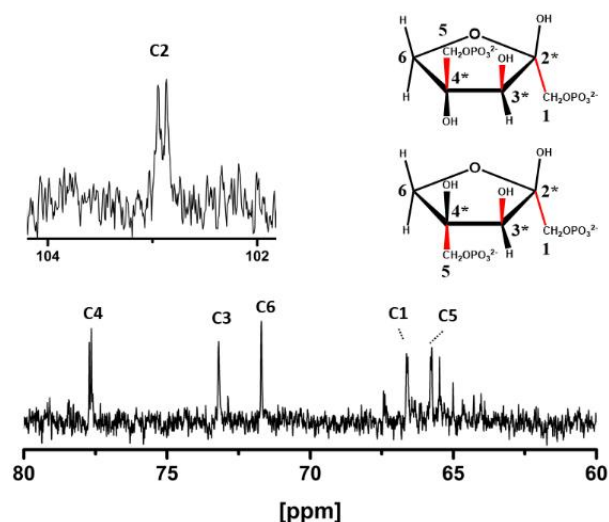
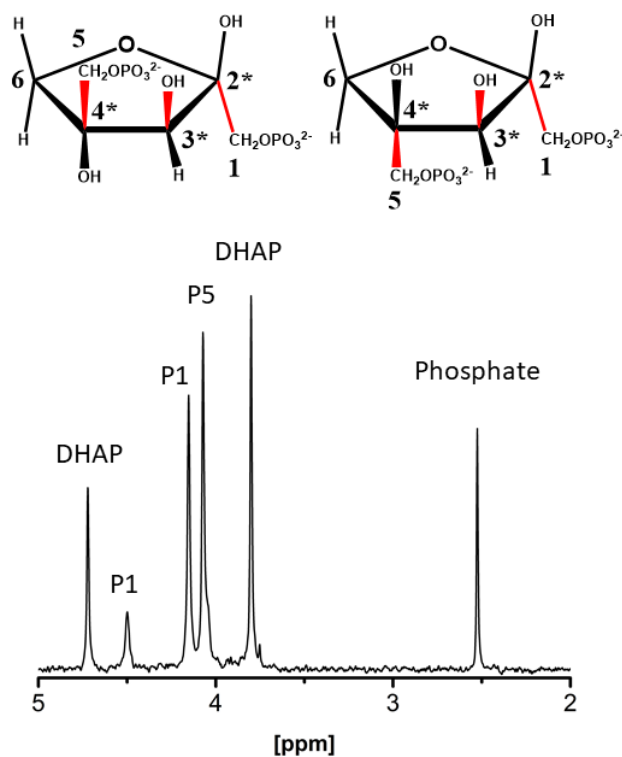**Figure 31 ^{13}C NMR spectrum of DHAP self-condensation.** Inlet shows the resonance of C2.

The assignment of resonances in the ^{13}C NMR spectrum was based on their chemical shifts and coupling constant with the attached phosphate groups (Table 12). C1 and C5 are two bonds away from P, which should have a C-P coupling constant of ~ 5 Hz. C2 and C4 are 3 bonds away from P, which should have a C-P coupling constant of ~ 9 Hz, $^{3}J_{C2,P1}$=9 Hz (*84*). The resonance at 103 ppm was assigned to C2 based on its anomeric position with 2 C-O linkages. Therefore, the resonance at 77 ppm was assigned to C4. The remaining resonances were assigned to C3 and C6. Since C6 has 2 hydrogens attached, its chemical shift should be less than C3. The resonance at 71 ppm was assigned to C6, and the resonance at 73 ppm was assigned to C3. The assignment

111

for C1 and C5 of dendroketose bisphosphate cannot be confidently determined with current data. They are tentatively assigned in the Table 12.

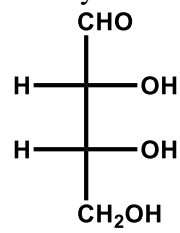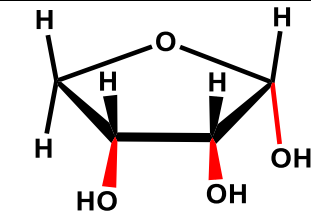**Table 12 ^{13}C NMR chemical shifts. (part per million)**
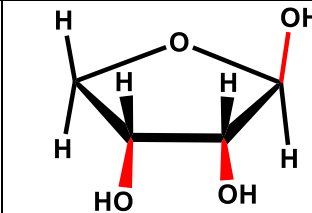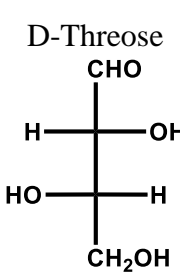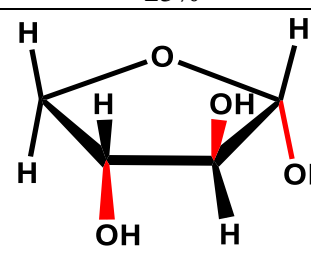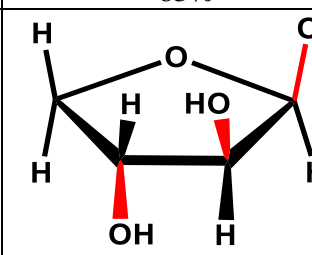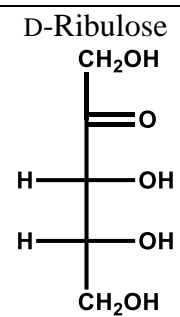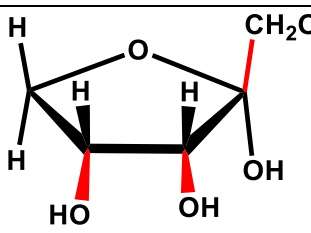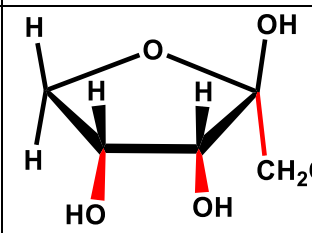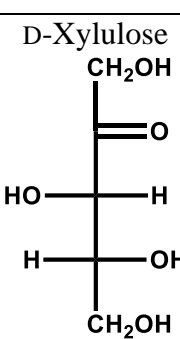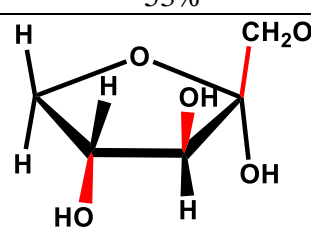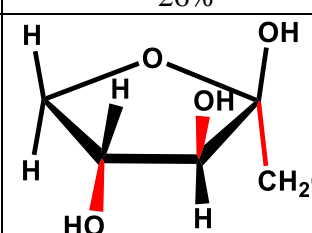
Dendroketose

^{13}C NMR	0 phosphate	1-phosphate	C-P coupling	bisphosphate	C-P coupling
C1	62.77	65.67	4 Hz	66.59*	5 Hz
C2	103.2	102.4	9 Hz	102.82	9 Hz
C3	73.12	70.96		73.2	
C4	78.27	78.01		77.64	9 Hz
C5	63.96	63.88		65.75*	5 Hz
C6	70.61	72.86		71.7	

In the ^{31}P NMR spectrum, the integration indicated there are two major forms of dendroketose bisphosphate, with a ratio of 80%:20% (Figure 32). Based on the chiral center of fructose C3, the stereochemistry at C3 of dendroketose bisphosphate is assigned as *S*, which is illustrated in Figure 32. Therefore, the dominant form of dendroketose bisphosphate is predicted as β-furanose, where the $CH_2OPO_3^{2-}$ group is oriented *trans* to the -OH at C3. This prediction is consistent with the trend observed with furanoses (Table 13).
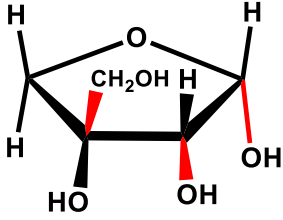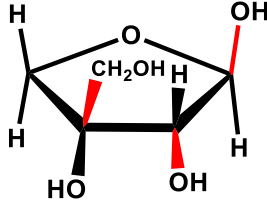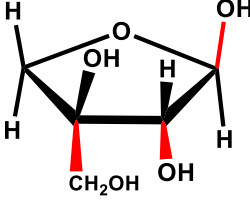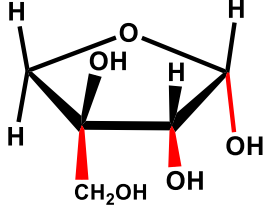
Figure 32 ^{31}P NMR spectrum of DHAP self-condensation.

113

Table 13 Anomer distribution in different furanose rings.

Compound	Distribution		
D-Erythrose CHO H——OH H——OH CH₂OH	α-ᴅ-Erythrose	β-ᴅ-Erythrose	Hydrate
	25%	63%	12%
D-Threose CHO H——OH HO——H CH₂OH	α-ᴅ-Threose	β-ᴅ-Threose	Hydrate
	51%	37%	12%
D-Ribulose CH₂OH ——O H——OH H——OH CH₂OH	α-ᴅ-Ribulose	β-ᴅ-Ribulose	Hydrate
	53%	26%	21%
D-Xylulose CH₂OH ——O HO——H H——OH CH₂OH	α-ᴅ-Xylulose	β-ᴅ-Xylulose	Hydrate
	18%	55%	27%

114

Table 13 Continued.

Compound	Distribution	
D-Apiose CHO H——OH HOH$_2$C——OH CH$_2$OH	**3−C-hydroxymethyl-α-D-Erythrose**	**3−C-hydroxymethyl-β-D-Erythrose**
	22%	54%
	3−C-hydroxymethyl-α-L-Threose	**3−C-hydroxymethyl-β-L-Threose**
	9%	13%

In either 4*S* or 4*R* configuration of dendroketose bisphosphate, the chemical environment of the P5 oriented *cis* or *trans* to the -OH at C3 is close to identical. The resonance at 4.05 ppm with the biggest integration area thus is assigned to P5, where one resonance is overlapped of one another. The resonance at 4.10 ppm is assigned to P1 oriented *trans* to the -OH at C3, which is the favored configuration. The less abundant resonance at 4.57 ppm is assigned to P1, which is oriented *cis* to the -OH at C3.

In the ^{13}C NMR spectrum of D-apiose, the difference in chemical shift of branched 3'C either oriented *cis* or *trans* to the -OH at C2 is ~ 0.6 ppm, which indicates the chemical environment is quite similar (*85*). Therefore, it is hard to predict whether C4 of dendroketose bisphosphate is *R* or *S* based on the ^{13}C and ^{31}P NMR spectra without a reference compound.

Dendroketose bisphosphate has 7 hydrogens attached to the carbons, two hydrogens each for C1, C5 and C6, and one hydrogen for C3. Since hydrogens on C1 and C5 have H-P coupling, with a constant of ~6-8 Hz, and 2 hydrogens on C6 are not equivalent to each other, the only hydrogen with no coupling in the ^{1}H NMR spectrum is assigned to hydrogen on C3, which is at 4.00 ppm (Figure 33). The H1 and H1' may or may not be magnetically equivalent to each other. A similar situation applies to the H5 and H5'. It is also hard to differentiate between H1 and H5 without further information. 1D NOESY experiment was attempted to determine the relative configuration by measuring the distance between H3 or H6 to H1 and H5 on a 500 MHz NMR instrument, but the resolution is not enough to distinguish H1 and H5 (*78*).
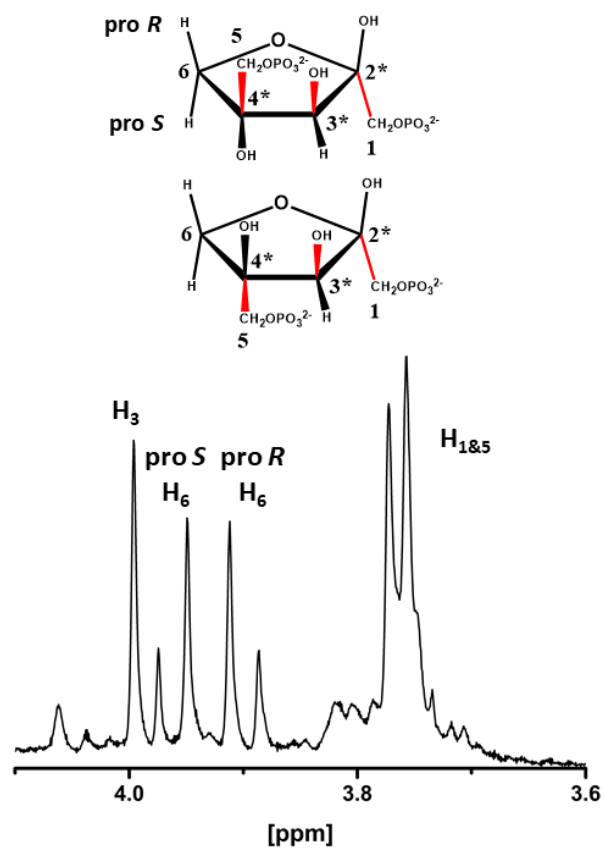
Figure 33 ¹H NMR spectrum of dendroketose bisphosphate, purified through a DEAE column.

3.2.6 Selective Deuteration of Dendroketose Bisphosphate

One of the hydrogens on C6 of dendroketose bisphosphate was stereo-selectively exchanged with deuterium. In methyl-β-D-erythro-furanoside, after the pro-*S* hydrogen at C4 was deuterated, the pro-*R* hydrogen at C4 shifted from 4.11 to 4.08 ppm (Scheme 14) (*86*). During the condensation of two molecules of DHAP in 100% D_2O, the pro-*S* hydrogen at C3 of DHAP was exchanged with deuterium. Sequentially, this deuterium on DHAP was selectively incorporated into C6 of dendroketose bisphosphate. In the ^{1}H NMR spectrum, the resonance of one hydrogen on C6 decreased, and the resonance of the remaining hydrogen had an upfield shift (Figure 34). Therefore, the resonance at 3.95 ppm is assigned to the pro-*S* hydrogen at C6, and the resonance at 3.91 ppm is assigned to the pro-*R* hydrogen at C6.
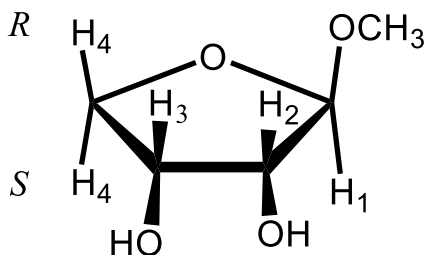
Scheme 14 Structure of methyl-β-D-erythro-furanoside.
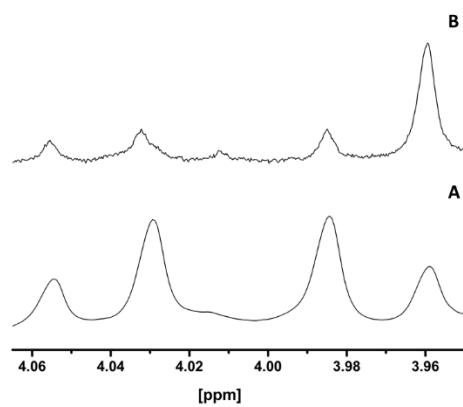
Figure 34 The resonances of Hs at C6 of dendroketose bisphosphate. (A) Dendroketose bisphosphate synthesized from H_2O. (B) Dendroketose bisphosphate synthesized from 100% D_2O.

3.3 Discussion

STM3780 is found in a gene cluster that is essential for colonization in ligated ileal loops in beef calves; the function related to this PTS gene cluster was unknown. There are more than 20 complete PTS permeases (enzyme II complexes consist of EIIA, EIIB, EIIC) in *Salmonella*, which are responsible for the transportation of carbohydrates across the membrane, including fructose, galactosamine, galactitol, glucose, and maltose (*40*). Since the PTS permeases dictate which carbohydrate to transport, among the enzyme II complex components, EIIC has the highest substrate specificity (*39*). STM3782, is an annotated PTS EIIC, and belongs to a family of galactitol-specific transporters. So far, only galactitol and D-arabitol have been identified to be transported by this family of permeases (Scheme 2) (*87, 88*). Another characteristic of the PTS EIIC is that it transfers a phosphate group from EIIB to the incoming carbohydrate. Therefore, it is expected that the substrate for the pathway is a phosphorylated carbohydrate.

Instead of screening dozens of carbohydrates, the functions of the aldolase STM3780 and the kinase STM3781 were first predicted based on sequence similarity analysis. STM3780 belongs to COG0191, in which all previously characterized aldolases were shown to have DHAP as a product of their reaction. In addition, they abstracted the pro-*S* hydrogen from DHAP. The experimental results supported this prediction, in which STM3780 has exclusive activity toward DHAP. For the intact carbohydrate substrate, the stereochemical preference at C4 and the modification remain to be determined.

Depending on how the enol-form of DHAP attacks the aldehyde in the condensation reaction, there are two possible stereocenters at C4. The condensation between DHAP and glyoxylic acid catalyzed by STM3780 yields D-xylulonate-1-P as product, which is the same as with fructose bisphosphate aldolase. This result indicates that STM3780 prefers an *R* stereocenter at C4 of ketose sugars, which is the same as fructose bisphosphate aldolase (*89*). Therefore, STM3780 is predicted to cleave a ketose with phosphorylation at C1 and the stereocenter at C3 and C4 being *S* and *R*, respectively.

Potential aldehyde substrates were tested, from two carbons up to six carbons. Among the compounds tested, STM3780 has a strong preference for aldehydes with terminal negative charges versus a hydroxyl group. The best substrate identified, L-threoteturonic acid (**8**), however, only has k_{cat} of 0.65 s^{-1}. In contrast, the H-D exchange rate is 2.9 s^{-1}. Therefore, the physiological substrate remains unknown.

Surprisingly, DHAP was shown to undergo a self-condensation reaction catalyzed by STM3780. The product is dendroketose bisphosphate. Therefore, it was of interest to characterize the structure of this branched monosaccharide bisphosphate. The self-condensation follows the abstraction of the pro-*S* hydrogen from C3 of DHAP, therefore, the chiral center at C3 of the product dendroketose bisphosphate was assigned as *S*. The enol-form of DHAP then attacks the ketone of another molecule of DHAP, presumably, with the same *si*-face attack as FBP aldolase (Scheme 15). Since STM3780 prefers aldehydes with negative charges, it is predicted to position the phosphate group of DHAP similar to the carboxylic group of glyoxylic acid. Thus, the C4 of dendroketose

bisphosphate is predicted to be in *R* configuration. The stereochemistry of C4 was

attempted to be explicitly solved with 1D NOESY experiment (*84*). However, the

adjacent hydrogens, both H3 and H6A or B, that could have Nuclear Overhauser Effect

on H5 have resonances close with each other in the ^{1}H NMR spectrum, which prevents

the accurate measuring of their relative distance.
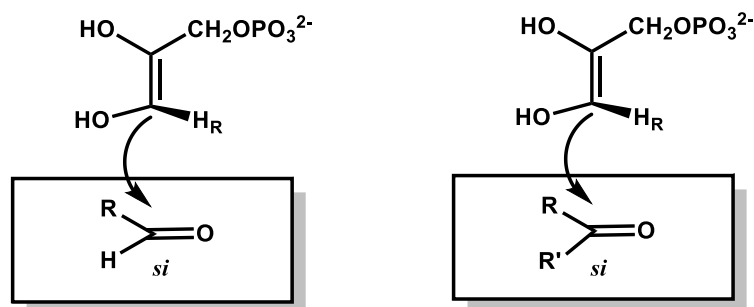
Scheme 15 Keto-group is attacked by the enol-form of DHAP. In the self-
condensation of DHAP, the enol-form of DHAP is proposed to attack the ketone of
another DHAP molecule in *si*-face, the same with attacking glyoxylic acid. R group has
negative charges.

Dendroketose bisphosphate cyclizes to form a furanose ring structure, in which the hydroxyl group at C6 will attack the ketone at C2, forming both α and β-furanoses. In the furanose conformation and distribution of D-erythrose, D-threose, D-ribulose, D-xylulose, D-apiose, dendroketose and dendroketose-1P, the preference of α and β-furanose is determined by the relative position of bulky group on C2 and C3 (Table 13). As for (3*S*, 4*R*) dendroketose bisphosphate, the β-furanose is predicted to be the dominant configuration, since C1 oriented *trans* with the -OH group at C3. In the ^{31}P NMR spectrum, it is shown to have two sets of resonances overlapping with each other.

Branched-chain monosaccharides are unusual among the naturally occurring sugars. D-Apiose is a branched pentose found in plant cell wall polysaccharides, as well as secondary metabolites (*90*). The biosynthesis of apiose remains unclear, and the function of apiose in polysaccharides or secondary metabolites is unknown (*91*). Only three branched-chain monosaccharides have been identified in plants: apiose, hamamelose and aceric acid (*91*). There are other branched-chain monosaccharides found in microbes, including, yersinioses and mycarose (Scheme 16) (*92, 93*).

CHO
H —— OH
HOH₂C —— OH
CH₂OH

Apiose

CHO
H —— OH
H —— H
          CH₃
HO ——
          OH
H —— OH
CH₃

Yersiniose

CHO
HOH₂C —— OH
H —— OH
HO —— H
CH₂OH

Hamamelose

CH₂OH
H —— H
H —— CH₃
O =
H —— OH
CH₃

Mycarose

CHO
HO —— H
HOOC —— OH
HO —— H
CH₃
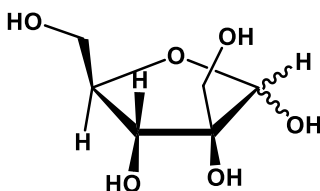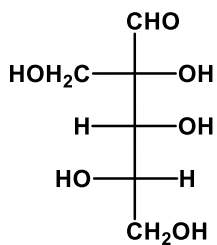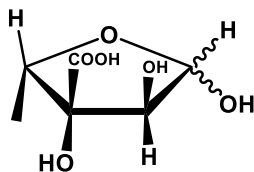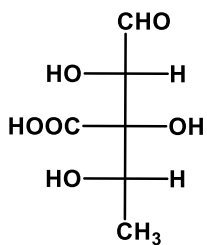
Aceric acid

Scheme 16 Identified branched-chain monosaccharides, including apiose, hamamelose, aceric acid, yersiniose, and mycarose.

The synthesis of racemic dendroketose was first achieved by Utkin by treating DHA with NaOH. Subsequently, the L isomer was utilized by a fungus, while leaving the D isomer (*94*). A remote homolog of rhamnulose aldolase from *Bacteroides thetaiotaomicron* has been shown to condense DHAP and DHA to form dendroketose-1-phosphate (*84*). The dendroketose-1-phosphate was further dephosphorylated to yield dendroketose.

The remaining question is the order of enzyme functions with the kinase and the aldolase. In the cases of *N*-acetylglucosamine and galactitol, the aldolase is the last step in the catabolism, following oxidation, isomerization, and phosphorylation. With this hypothesis, the kinase STM3781 would need to phosphorylate a phosphorylated carbohydrate. However, among homologs in the FGGY kinase family, there are no precedents of phosphorylation on a phosphorylated carbohydrate. The substrates of the FGGY kinase family was mapped on a phylogenetic tree (Figure 15). Within this family, the kinases prefer to phosphorylate on the terminus away from C1, unless the terminus is deoxy.

The function of the kinase STM3781 remains unknown. Different expression constructs were tested with different growth and induction conditions, but soluble protein was not obtained. In the genomic context of tagatose bisphosphate aldolase, KbaY, a neighborhood gene *kbaZ* was initially proposed to encode a kinase (Scheme 3) (*47*). However, the kinase activity of KbaZ was never observed. Later, KbaZ was proposed to have a chaperone-like function, which is required for the full activity and stability of the tagatose bisphosphate aldolase. A homolog of KbaZ, AtuTag6PK

125

(PF08013), was later shown as an epimerase that isomerizes tagatose-6-phosphate to fructose-6-phosphate (*48*). However, the sequence similarity between STM3781 and either KbaZ or AtuTag6PK is low. Co-expression of STM3780 and STM3781 was tested, however, the expression of STM3781 under tested condition was not observed.

CHAPTER IV

CONCLUSIONS

4.1 Link Characterized Functions to *Salmonella* Colonization

In this study, it was proposed that γ-L-glutamyl hydroxamate may interfere with glutamine-mediated metabolism in *Salmonella*, which may further have an effect on its colonization. However, the presence of γ-L-glutamyl hydroxamate has not been established in the gut of cattle. It remains possible that there is an unidentified compound that is a better substrate for STM2437. The glutamyl hydrolase activity of STM2437 is surprisingly different with other homologues in the glutamine amidotransferase family, since no other homologues have reported activity toward glutamyl hydroxamate.

The catabolism of the carbohydrate transported by the identified PTS gene cluster STM3779-3785 was hypothesized and characterized. It remains unclear how the catabolism of the best substrate L-galactoheptulonate-1-phosphate contributes to the fitness of *Salmonella*. The aldolase STM3780 was shown to catalyze the self-condensation of DHAP to yield dendroketose bisphosphate, a branched monosaccharide. Further experiments are warranted to link the characterized functions to their role in *Salmonella* colonization and infection.

Identified gene clusters may play roles not only in infection of animal hosts, but also in plants. Bacterial infection of rice can be particularly destructive. The main pathogen that causes rice blight is *Xanthomonas oryzae* pv. *oryzae*. The *Xoo* pathogen

127

produces targeted effector molecules that binds to sugar transporters to facilitate the

transport of carbohydrates that is required for the bacterial growth (*95*). Elevated sugar

transportation enhances the virulence of bacteria and leading to the disease susceptibility

of rice. Therefore, it may be worthwhile to screen genes in *Salmonella* related to

infection of plants in future.

4.2 Methodology to Predict Gene Cluster Function

Previously, there have been several attempts to classify genes into categories

including cell motility, transcriptional regulations, etc., based on the annotation of

individual genes. However, the ambiguous and broad classifications seldomly provide

valuable information to guide their function discovery. The ultimate goal of gene cluster

function prediction is to recognize gene clusters from an unknown genome and predict

their functions.

There have been established workflows for the function discovery of non-

ribosomal peptide synthetase (NRPS), and the function discovery of catabolic pathways

in which solute binding protein determines which substrate to be carried across the

membrane by the ATP binding cassette transport system. The underlying rationale is that

a cluster of genes with related function are often located in close proximity and can have

a pattern which include a series of reaction modules/COGs, where the function

prediction is based on the pattern of the gene cluster together, rather than individual

ones. Here I propose a generalized workflow to identify gene clusters and predict their

128

function (Scheme 17). This workflow is exemplified by function prediction of gene clusters associated with carbohydrate metabolism and transportation.
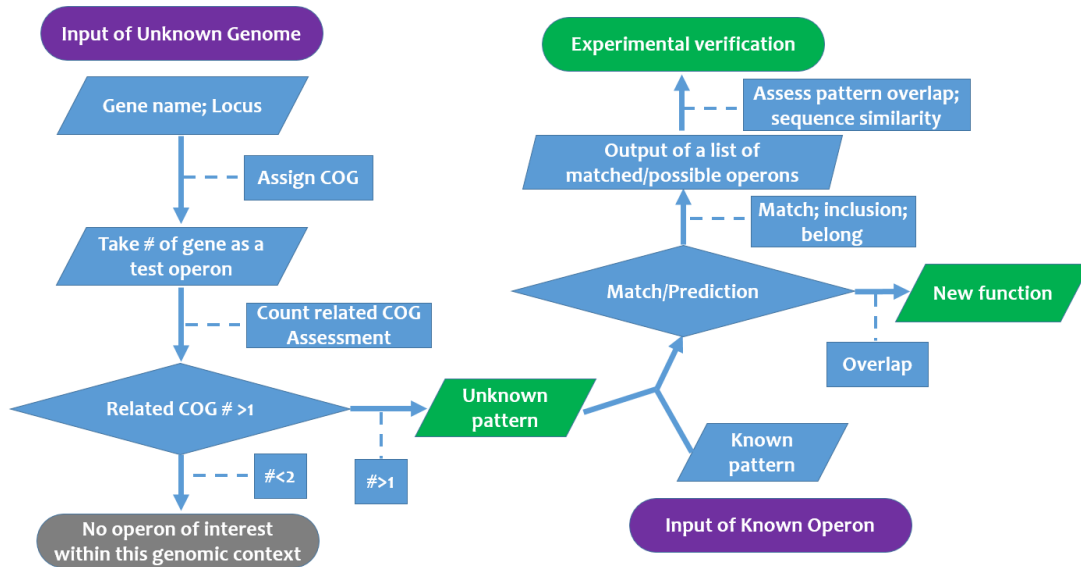
Scheme 17 Flow chart of gene cluster function prediction.

It takes several steps to realize this function prediction.

First, the patterns of known pathways need to be recorded, which will be used as template to predict or to assess an unknown gene cluster. Reviewed carbohydrate related pathways can be downloaded from MetaCyc, where the components of the pathway, individual enzymes and their COGs, are analyzed and assembled as functional modules. For example, a subset of the gene cluster containing PTS EIIABC is characteristic for a

transportation system related to carbohydrate, which often follows a set of enzymes for the catabolism of transported carbohydrate.

Second, it needs a strategy to define a gene cluster within the certain genomic context. When an unknown genome is the input, without the promoter/regulator analysis, it needs reasonable hypothesis to define an optimal number as the size of the gene cluster. Alternatively, one can start with a known gene or COG, and search genes in the upstream and downstream to assemble functional related genes together to have a functional gene cluster in an iterative way. In each iteration, a new gene is added to the current functional pathway only if it meets pre-set criteria.

Third, after analyzing an unknown gene cluster, a program should identify the pattern and find approximate matches from known pathways. Such unknown pattern will either match a combination of reaction modules/COGs, or have an overlap with known patterns.

Finally, a list of assessment or prediction result will be the output, and it takes experimental validation to support the function prediction of the unknown gene cluster. This in turn provides feedback to the pattern recognition and expand the pattern database.

This strategy of gene cluster function prediction will be validated by function prediction of carbohydrate related gene cluster from unknown genomes. A fragment of genome will be input into the program, in which gene clusters will be assessed and predicted with potential function based on gene composition. Once such workflow for

function prediction is validated, the input could be larger genome fragments, even a

complete genome.

REFERENCES

1.   Chittim, C. L., Irwin, S. M., and Balskus, E. P. (2018) Deciphering Human Gut
        Microbiota–Nutrient Interactions: A Role for Biochemistry, *Biochemistry 57*,
        2567-2577.

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

3.   Chen, Y. E., Fischbach, M. A., and Belkaid, Y. (2018) Skin microbiota–host
        interactions, *Nature 553*, 427.

4.   Sonnenberg, G. F., Monticelli, L. A., Alenghat, T., Fung, T. C., Hutnick, N. A.,
        Kunisawa, J., Shibata, N., Grunberg, S., Sinha, R., Zahm, A. M., Tardif, M. R.,
        Sathaliyawala, T., Kubota, M., Farber, D. L., Collman, R. G., Shaked, A.,
        Fouser, L. A., Weiner, D. B., Tessier, P. A., Friedman, J. R., Kiyono, H.,
        Bushman, F. D., Chang, K.-M., and Artis, D. (2012) Innate Lymphoid Cells
        Promote Anatomical Containment of Lymphoid-Resident Commensal Bacteria,
        Science 336, 1321-1325.

5.   Dodd, D., Spitzer, M. H., Van Treuren, W., Merrill, B. D., Hryckowian, A. J.,
        Higginbottom, S. K., Le, A., Cowan, T. M., Nolan, G. P., Fischbach, M. A., and
        Sonnenburg, J. L. (2017) A gut bacterial pathway metabolizes aromatic amino
        acids into nine circulating metabolites, *Nature 551*, 648.

6.   Thomas, C. M., Hong, T., van Pijkeren, J. P., Hemarajata, P., Trinh, D. V., Hu,
        W., Britton, R. A., Kalkum, M., and Versalovic, J. (2012) Histamine Derived
        from Probiotic Lactobacillus reuteri Suppresses TNF via Modulation of PKA and
        ERK Signaling, *PLoS ONE 7*, e31951.

7.   Nicholson, J. K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., and
        Pettersson, S. (2012) Host-Gut Microbiota Metabolic Interactions, *Science 336*,
        1262-1267.

8.   Ndeh, D., Rogowski, A., Cartmell, A., Luis, A. S., Baslé, 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'Neill, M. A., Urbanowicz, B. R., York, W.
        S., Davies, G. J., Abbott, D. W., Ralet, M.-C., Martens, E. C., Henrissat, B., and

Gilbert, H. J. (2017) Complex pectin metabolism by gut bacteria reveals novel catalytic functions, *Nature 544*, 65.

9.   Libertucci, J., and Young, V. B. (2019) The role of the microbiota in infectious diseases, *Nature Microbiology 4*, 35-45.

10.   Lawley, T. D., and Walker, A. W. (2013) Intestinal colonization resistance, *Immunology 138*, 1-11.

11.   Bumann, D., and Schothorst, J. (2017) Intracellular Salmonella metabolism, *Cellular Microbiology 19*, e12766.

12.   Desai, P. T., Porwollik, S., Long, F., Cheng, P., Wollam, A., Clifton, S. W., Weinstock, G. M., and McClelland, M. (2013) Evolutionary genomics of salmonella enterica subspecies, *mBio 4*, e00579-00512.

13.   Dandekar, T., Fieselmann, A., Fischer, E., Popp, J., Hensel, M., and Noster, J. (2015) Salmonella—how a metabolic generalist adopts an intracellular lifestyle during infection, *Frontiers in Cellular and Infection Microbiology 4*, 1-11.

14.   Wotzka, S. Y., Nguyen, B. D., and Hardt, W.-D. (2017) Salmonella Typhimurium Diarrhea Reveals Basic Principles of Enteropathogen Infection and Disease-Promoted DNA Exchange, *Cell Host & Microbe 21*, 443-454.

15.   Sellin, Mikael E., Müller, Anna A., Felmy, B., Dolowschiak, T., Diard, M., Tardivel, A., Maslowski, Kendle M., and Hardt, W.-D. (2014) Epithelium-Intrinsic NAIP/NLRC4 Inflammasome Drives Infected Enterocyte Expulsion to Restrict Salmonella Replication in the Intestinal Mucosa, *Cell Host & Microbe 16*, 237-248.

16.   McClelland, M., Sanderson, K. E., Spieth, J., Clifton, S. W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R., and Wilson, R. K. (2001) Complete genome sequence of Salmonella enterica serovar Typhimurium LT2, *Nature 413*, 852-856.

17.   Rasko, D. A., Rosovitz, M. J., Myers, G. S. A., Mongodin, E. F., Fricke, W. F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N. R., Chaudhuri, R., Henderson, I. R., Sperandio, V., and Ravel, J. (2008) The Pangenome Structure of

Escherichia coli Comparative Genomic Analysis of E. coli Commensal and Pathogenic Isolates, *Journal of Bacteriology 190*, 6881.

18.   Groisman, E. A., and Ochman, H. (1997) How Salmonella became a pathogen, *Trends in Microbiology 5*, 343-349.

19.   Sansonetti, P. J., Kopecko, D. J., and Formal, S. B. (1982) Involvement of a plasmid in the invasive ability of Shigella flexneri, *Infection and Immunity 35*, 852.

20.   McDaniel, T. K., and Kaper, J. B. (1997) A cloned pathogenicity island from enteropathogenic Escherichia coli confers the attaching and effacing phenotype on E. coli K-12, *Molecular Microbiology 23*, 399-407.

21.   Groisman, E. A., and Ochman, H. (1996) Pathogenicity Islands: Bacterial Evolution in Quantum Leaps, *Cell 87*, 791-794.

22.   Bumann, D. (2009) System-level analysis of Salmonella metabolism during infection, *Current Opinion in Microbiology 12*, 559-567.

23.   Rivera-Chávez, F., and Bäumler, A. J. (2015) The Pyromaniac Inside You: Salmonella Metabolism in the Host Gut, *Annual Review of Microbiology 69*, 31-48.

24.   Plasterk, R. H. A., Izsvák, Z., and Ivics, Z. (1999) Resident aliens: the Tc1/mariner superfamily of transposable elements, *Trends in Genetics 15*, 326-332.

25.   Santiviago, C. A., Reynolds, M. M., Porwollik, S., Choi, S. H., Long, F., Andrews-Polymenis, H. L., and McClelland, M. (2009) Analysis of pools of targeted Salmonella deletion mutants identifies novel genes affecting fitness during competitive infection in mice, *PLoS pathogens 5*, e1000477.

26.   Lawley, T. D., Chan, K., Thompson, L. J., Kim, C. C., Govoni, G. R., and Monack, D. M. (2006) Genome-Wide Screen for Salmonella Genes Required for Long-Term Systemic Infection of the Mouse, *PLoS pathogens 2*, e11.

27.   Elfenbein, J. R., Endicott-Yazdani, T., Porwollik, S., Bogomolnaya, L. M., Cheng, P., Guo, J., Zheng, Y., Yang, H.-J., Talamantes, M., Shields, C., Maple, A., Ragoza, Y., DeAtley, K., Tatsch, T., Cui, P., Andrews, K. D., McClelland, M., Lawhon, S. D., and Andrews-Polymenis, H. (2013) Novel Determinants of

Intestinal Colonization of Salmonella enterica Serotype Typhimurium Identified in Bovine Enteric Infection, *Infection and Immunity 81*, 4311-4320.

28.   Atkinson, H. J., Morris, J. H., Ferrin, T. E., and Babbitt, P. C. (2009) Using Sequence Similarity Networks for Visualization of Relationships Across Diverse Protein Superfamilies, *PLoS ONE 4*, e4345.

29.   Ikeda, H., Nonomiya, T., Usami, M., Ohta, T., and Ōmura, S. (1999) Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in Streptomyces avermitilis, *Proceedings of the National Academy of Sciences 96*, 9509-9514.

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

31.   Gerlt, J. A. (2017) Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence–Function Space and Genome Context to Discover Novel Functions, *Biochemistry 56*, 4293-4308.

32.   Massière, F., and Badet-Denisot, M. A. (1998) The mechanism of glutamine-dependent amidotransferases, *CMLS, Cell. Mol. Life Sci. 54*, 205-222.

33.   Tesmer, J. J., Klem, T. J., Deras, M. L., Davisson, V. J., and Smith, J. L. (1996) The crystal structure of GMP synthetase reveals a novel catalytic triad and is a structural paradigm for two enzyme families, *Nature Structural Biology 3*, 74-86.

34.   Okinaka, Y., Yang, C. H., Perna, N. T., and Keen, N. T. (2002) Microarray profiling of Erwinia chrysanthemi 3937 genes that are regulated during plant infection, *Molecular Plant-Microbe Interactions 15*, 619-629.

35.   Basset, G. J. C., Quinlivan, E. P., Ravanel, S., Rébeillé, F., Nichols, B. P., Shinozaki, K., Seki, M., Adams-Phillips, L. C., Giovannoni, J. J., Gregory, J. F., and Hanson, A. D. (2004) Folate synthesis in plants: The p-aminobenzoate branch is initiated by a bifunctional PabA-PabB protein that is targeted to plastids, *Proceedings of the National Academy of Sciences of the United States of America 101*, 1496-1501.

36. Green, J. M., and Nichols, B. P. (1991) p-Aminobenzoate biosynthesis in Escherichia coli. Purification of aminodeoxychorismate lyase and cloning of pabC, *Journal of Biological Chemistry 266*, 12971-12975.

37. Kurihara, S., Oda, S., Tsuboi, Y., Kim, H. G., Oshida, M., Kumagai, H., and Suzuki, H. (2008) γ-Glutamylputrescine Synthetase in the Putrescine Utilization Pathway of Escherichia coli K-12, *Journal of Biological Chemistry 283*, 19981-19990.

38. Taylor, Z. W., Brown, H. A., Holden, H. M., and Raushel, F. M. (2017) Biosynthesis of Nucleoside Diphosphoramidates in Campylobacter jejuni, *Biochemistry 56*, 6079-6082.

39. Deutscher, J., Aké, F. M. D., Derkaoui, M., Zébré, A. C., Cao, T. N., Bouraoui, H., Kentache, T., Mokhtari, A., Milohanic, E., and Joyet, P. (2014) The Bacterial Phosphoenolpyruvate:Carbohydrate Phosphotransferase System: Regulation by Protein Phosphorylation and Phosphorylation-Dependent Protein-Protein Interactions, *Microbiology and Molecular Biology Reviews 78*, 231.

40. Barabote, R. D., and Saier, M. H. (2005) Comparative Genomic Analyses of the Bacterial Phosphotransferase System, *Microbiology and Molecular Biology Reviews 69*, 608.

41. Deutscher, J., Francke, C., and Postma, P. W. (2006) How Phosphotransferase System-Related Protein Phosphorylation Regulates Carbohydrate Metabolism in Bacteria, *Microbiology and Molecular Biology Reviews 70*, 939.

42. AbuOun, M., Suthers, P. F., Jones, G. I., Carter, B. R., Saunders, M. P., Maranas, C. D., Woodward, M. J., and Anjum, M. F. (2009) Genome Scale Reconstruction of a Salmonella Metabolic Model: Comparison of Similarity and Differences with a Commensal Escherichia coli Strain, *Journal of Biological Chemistry 284*, 29480-29488.

43. Miller, K. A., Phillips, R. S., Mrázek, J., and Hoover, T. R. (2013) Salmonella Utilizes D-Glucosaminate via a Mannose Family Phosphotransferase System Permease and Associated Enzymes, *Journal of Bacteriology 195*, 4057.

44. Miller, K. A., Phillips, R. S., Kilgore, P. B., Smith, G. L., and Hoover, T. R. (2015) A Mannose Family Phosphotransferase System Permease and Associated Enzymes Are Required for Utilization of Fructoselysine and Glucoselysine in Salmonella enterica Serovar Typhimurium, *Journal of Bacteriology 197*, 2831.

45.   Cao, Y., Jin, X., Levin, E. J., Huang, H., Zong, Y., Quick, M., Weng, J., Pan, Y., Love, J., Punta, M., Rost, B., Hendrickson, W. A., Javitch, J. A., Rajashankar, K. R., and Zhou, M. (2011) Crystal structure of a phosphorylation-coupled saccharide transporter, *Nature 473*, 50-54.

46.   Puttamreddy, S., Cornick, N. A., and Minion, F. C. (2010) Genome-Wide Transposon Mutagenesis Reveals a Role for pO157 Genes in Biofilm Development in Escherichia col O157:H7 EDL933, *Infection and Immunity 78*, 2377.

47.   Brinkkötter, A., Shakeri-Garakani, A., and Lengeler, J. W. (2002) Two class II D-tagatose-bisphosphate aldolases from enteric bacteria, *Archives of Microbiology 177*, 410-419.

48.   Wichelecki, D. J., Vetting, M. W., Chou, L., Al-Obaidi, N., Bouvier, J. T., Almo, S. C., and Gerlt, J. A. (2015) ATP-binding Cassette (ABC) Transport System Solute-binding Protein-guided Identification of Novel d-Altritol and Galactitol Catabolic Pathways in Agrobacterium tumefaciens C58, *Journal of Biological Chemistry 290*, 28963-28976.

49.   Qamar, S., Marsh, K., and Berry, A. (1996) Identification of arginine 331 as an important active site residue in the Class II fructose-1,6-bisphosphate aldolase of Escherichia coli, *Protein Science 5*, 154-161.

50.   Kroemer, M., Merkel, I., and Schulz, G. E. (2003) Structure and Catalytic Mechanism of l-Rhamnulose-1-phosphate Aldolase, *Biochemistry 42*, 10560-10568.

51.   Joerger, A. C., Gosse, C., Fessner, W.-D., and Schulz, G. E. (2000) Catalytic Action of Fuculose 1-Phosphate Aldolase (Class II) As Derived from Structure-Directed Mutagenesis, *Biochemistry 39*, 6033-6041.

52.   Zhang, X., Carter, M. S., Vetting, M. W., San Francisco, B., Zhao, S., Al-Obaidi, N. F., Solbiati, J. O., Thiaville, J. J., de Crécy-Lagard, V., Jacobson, M. P., Almo, S. C., and Gerlt, J. A. (2016) Assignment of function to a domain of unknown function: DUF1537 is a new kinase family in catabolic pathways for acid sugars, *Proceedings of the National Academy of Sciences 113*, E4161-E4169.

53.   Huddleston, J. P., Thoden, J. B., Dopkins, B. J., Narindoshvili, T., Fose, B. J., Holden, H. M., and Raushel, F. M. (2019) Structural and Functional

Characterization of YdjI, an Aldolase of Unknown Specificity in Escherichia coli K12, *Biochemistry 58*, 3340-3353.

54.   Zhang, Y., Zagnitko, O., Rodionova, I., Osterman, A., and Godzik, A. (2011) The FGGY Carbohydrate Kinase Family: Insights into the Evolution of Functional Specificities, *PLOS Computational Biology 7*, e1002318.

55.   Jaisson, S., Veiga-da-Cunha, M., and Van Schaftingen, E. (2009) Molecular identification of ω-amidase, the enzyme that is functionally coupled with glutamine transaminases, as the putative tumor suppressor Nit2, *Biochimie 91*, 1066-1071.

56.   Wittmann, S., Schnabelrauch, M., Scherlitz-Hofmann, I., Möllmann, U., Ankel-Fuchs, D., and Heinisch, L. (2002) New synthetic siderophores and Their β-Lactam conjugates based on diamino acids and dipeptides, *Bioorganic & Medicinal Chemistry 10*, 1659-1670.

57.   Teze, D., Dion, M., Daligault, F., Tran, V., André-Miral, C., and Tellier, C. (2013) Alkoxyamino glycoside acceptors for the regioselective synthesis of oligosaccharides using glycosynthases and transglycosidases, *Bioorganic & Medicinal Chemistry Letters 23*, 448-451.

58.   List, F., Bocola, M., Haeger, M. C., and Sterner, R. (2012) Constitutively active glutaminase variants provide insights into the activation mechanism of anthranilate synthase, *Biochemistry 51*, 2812-2818.

59.   Barrett, A. J. (1981) [42] Cathepsin G, In *Methods in Enzymology*, pp 561-565, Academic Press.

60.   Marti-Arbona, R., Xu, C., Steele, S., Weeks, A., Kuty, G. F., Seibert, C. M., and Raushel, F. M. (2006) Annotating enzymes of unknown function: N-formimino-L-glutamate deiminase is a member of the amidohydrolase superfamily, *Biochemistry 45*, 1997-2005.

61.   Cummings, J. A., Nguyen, T. T., Fedorov, A. A., Kolb, P., Xu, C., Fedorov, E. V., Shoichet, B. K., Barondeau, D. P., Almo, S. C., and Raushel, F. M. (2010) Structure, mechanism, and substrate profile for Sco3058: the closest bacterial homologue to human renal dipeptidase, *Biochemistry 49*, 611-622.

62.   Chittur, S. V., Klem, T. J., Shafer, C. M., and Davisson, V. J. (2001) Mechanism for Acivicin Inactivation of Triad Glutamine Amidotransferases, *Biochemistry 40*, 876-887.

63.   Bogomolnaya, L. M., Santiviago, C. A., Yang, H.-J., Baumler, A. J., and Andrews-Polymenis, H. L. (2008) 'Form variation' of the O12 antigen is critical for persistence of Salmonella Typhimurium in the murine intestine, *Molecular Microbiology 70*, 1105-1119.

64.   Thierauf, A., Perez, G., Maloy, and Stanley. (2009) Generalized Transduction, In *Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions* (Clokie, M. R. J., and Kropinski, A. M., Eds.), pp 267-286, Humana Press, Totowa, NJ.

65.   Yang, H.-J., Bogomolnaya, L. M., Elfenbein, J. R., Endicott-Yazdani, T., Reynolds, M. M., Porwollik, S., Cheng, P., Xia, X.-Q., McClelland, M., and Andrews-Polymenis, H. (2016) Novel Two-Step Hierarchical Screening of Mutant Pools Reveals Mutants under Selection in Chicks, *Infection and Immunity 84*, 1226.

66.   Thoden, J. B., Huang, X., Raushel, F. M., and Holden, H. M. (1999) The Small Subunit of Carbamoyl Phosphate Synthetase: Snapshots along the Reaction Pathway, *Biochemistry 38*, 16158-16166.

67.   Chen, Y., Scanlan, J., Song, L., Crombie, A., Rahman, M. T., Schafer, H., and Murrell, J. C. (2010) γ-Glutamylmethylamide is an essential intermediate in the metabolism of methylamine by Methylocella silvestris, *Applied and Environmental Microbiology 76*, 4530-4537.

68.   de Azevedo Wäsch, S. I., van der Ploeg, J. R., Maire, T., Lebreton, A., Kiener, A., and Leisinger, T. (2002) Transformation of Isopropylamine to l-Alaninol by Pseudomonas sp. Strain KIE171 Involves N-Glutamylated Intermediates, *Applied and Environmental Microbiology 68*, 2368-2375.

69.   Kudo, K., Ozaki, T., Shin-ya, K., Nishiyama, M., and Kuzuyama, T. (2017) Biosynthetic Origin of the Hydroxamic Acid Moiety of Trichostatin A: Identification of Unprecedented Enzymatic Machinery Involved in Hydroxylamine Transfer, *Journal of the American Chemical Society 139*, 6799-6802.

70.   Oakley, A. J., Coggan, M., and Board, P. G. (2010) Identification and Characterization of γ-Glutamylamine Cyclotransferase, an Enzyme Responsible for γ-Glutamyl-ε-lysine Catabolism, *Journal of Biological Chemistry 285*, 9642-9648.

71.   Wolf, F., Leipoldt, F., Kulik, A., Wibberg, D., Kalinowski, J., and Kaysser, L. (2018) Characterization of the Actinonin Biosynthetic Gene Cluster, *ChemBioChem 19*, 1189-1195.

72.   Leipoldt, F., Santos-Aberturas, J., Stegmann, D. P., Wolf, F., Kulik, A., Lacret, R., Popadić, D., Keinhörster, D., Kirchner, N., Bekiesch, P., Gross, H., Truman, A. W., and Kaysser, L. (2017) Warhead biosynthesis and the origin of structural diversity in hydroxamate metalloproteinase inhibitors, *Nature Communications 8*, 1965.

73.   Wang, K.-K. A., Ng, T. L., Wang, P., Huang, Z., Balskus, E. P., and van der Donk, W. A. (2018) Glutamic acid is a carrier for hydrazine during the biosyntheses of fosfazinomycin and kinamycin, *Nature Communications 9*, 3687.

74.   Ikeda, T. P., Shauger, A. E., and Kustu, S. (1996) Salmonella typhimuriumApparently Perceives External Nitrogen Limitation as Internal Glutamine Limitation, *Journal of Molecular Biology 259*, 589-607.

75.   Allison, C., Lai, H.-C., Gygi, D., and Hughes, C. (1993) Cell differentiation of Proteus mirabilis is initiated by glutamine, a specific chemoattractant for swarming cells, *Molecular Microbiology 8*, 53-60.

76.   Nielsen, S. S. (2010) Phenol-sulfuric acid method for total carbohydrates., In *Food analysis laboratory manual*, pp 47-53, Purdue University, West Lafayette, USA.

77.   Stott, K., Keeler, J., Van, Q. N., and Shaka, A. J. (1997) One-Dimensional NOE Experiments Using Pulsed Field Gradients, *Journal of Magnetic Resonance 125*, 302-324.

78.   Hu, H., and Krishnamurthy, K. (2006) Revisiting the initial rate approximation in kinetic NOE measurements, *Journal of Magnetic Resonance 182*, 173-177.

79.   Studier, F. W. (2005) Protein production by auto-induction in high-density shaking cultures, *Protein Expression and Purification 41*, 207-234.

80.   Kamat, S. S., Williams, H. J., Dangott, L. J., Chakrabarti, M., and Raushel, F. M. (2013) The catalytic mechanism for aerobic formation of methane by bacteria, *Nature 497*, 132.

81.   Li, M. Z., and Elledge, S. J. (2007) Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC, *Nature Methods 4*, 251-256.

82.   Adachi, O., Hours, R. A., Shinagawa, E., Akakabe, Y., Yakushi, T., and Matsushita, K. (2011) Enzymatic Synthesis of 4-Pentulosonate (4-Keto-D-pentonate) from D-Aldopentose and D-Pentonate by Two Different Pathways Using Membrane Enzymes of Acetic Acid Bacteria, *Bioscience, Biotechnology, and Biochemistry 75*, 2418-2420.

83.   Ślepokura, K., and Lis, T. (2010) Dihydroxyacetone phosphate, DHAP, in the crystalline state: monomeric and dimeric forms, *Carbohydrate Research 345*, 512-529.

84.   Laurent, V., Darii, E., Aujon, A., Debacker, M., Petit, J.-L., Hélaine, V., Liptaj, T., Breza, M., Mariage, A., Nauton, L., Traïkia, M., Salanoubat, M., Lemaire, M., Guérard-Hélaine, C., and de Berardinis, V. (2018) Synthesis of Branched-Chain Sugars with a DHAP-Dependent Aldolase: Ketones are Electrophile Substrates of Rhamnulose-1-phosphate Aldolases, *Angewandte Chemie International Edition 57*, 5467-5471.

85.   Snyder, J. R., and Serianni, A. S. (1987) dl-apiose substituted with stable isotopes: Synthesis, N.M.R.-spectral analysis, and furanose anomerization, *Carbohydrate Research 166*, 85-99.

86.   Wu, G. D., Serianni, A. S., and Barker, R. (1983) Stereoselective deuterium exchange of methylene protons in methyl tetrofuranosides: hydroxymethyl group conformations in methyl pentofuranosides, *The Journal of Organic Chemistry 48*, 1750-1757.

87.   Shakeri-Garakani, A., Brinkkötter, A., Schmid, K., Turgut, S., and Lengeler, J. W. (2004) The genes and enzymes for the catabolism of galactitol, D-tagatose, and related carbohydrates in Klebsiella oxytoca M5a1 and other enteric bacteria display convergent evolution, *Molecular Genetics and Genomics 271*, 717-728.

88.   López, M. G., Irla, M., Brito, L. F., and Wendisch, V. F. (2019) Characterization of D-Arabitol as Newly Discovered Carbon Source of Bacillus methanolicus, *Frontiers in Microbiology 10*, 1-14.

89.   Williams, G. J., Domann, S., Nelson, A., and Berry, A. (2003) Modifying the stereochemistry of an enzyme-catalyzed reaction by directed evolution, *Proceedings of the National Academy of Sciences 100*, 3143.

90.   Ndeh, D., Rogowski, A., Cartmell, A., Luis, A. S., Baslé, 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'Neill, M. A., Urbanowicz, B. R., York, W. S., Davies, G. J., Abbott, D. W., Ralet, M.-C., Martens, E. C., Henrissat, B., and Gilbert, H. J. (2017) Complex pectin metabolism by gut bacteria reveals novel catalytic functions, *Nature 544*, 65-70.

91.   Pičmanová, M., and Møller, B. L. (2016) Apiose: one of nature's witty games, *Glycobiology 26*, 430-442.

92.   Gorshkova, R. P., Zubkov, V. A., Isakov, V. V., and Ovodov, Y. S. (1984) Yersiniose, a new branched-chain sugar, *Carbohydrate Research 126*, 308-312.

93.   Takahashi, H., Liu, Y.-n., Chen, H., and Liu, H.-w. (2005) Biosynthesis of TDP-l-Mycarose: The Specificity of a Single Enzyme Governs the Outcome of the Pathway, *Journal of the American Chemical Society 127*, 9340-9341.

94.   Rathbone, E. B., and Woolard, G. R. (1976) A stereospecific synthesis of l-dendroketose derivatives, *Carbohydrate Research 46*, 183-187.

95.   Buscaill, P., Chandrasekar, B., Sanguankiattichai, N., Kourelis, J., Kaschani, F., Thomas, E. L., Morimoto, K., Kaiser, M., Preston, G. M., Ichinose, Y., and van der Hoorn, R. A. L. (2019) Glycosidase and glycan polymorphism control hydrolytic release of immunogenic flagellin peptides, *Science 364*, eaav0748.

Table 14 Identified 54 genes under selection in the cattle model.

Locus tag and gene name	Description
STM2867, hilC	Transcription regulator on pathogenicity island 1
STM2875, hilD	Transcription regulator on pathogenicity island 1
STM2883, sipD	Essential for bacterial entry into cultured mammalian cells, may modulate the secretion of other proteins
STM2884, sipC	Actin-binding protein that interferes with host cell actin cytoskeleton
STM2885, sipB	Direct heterotypic fusion
STM2886, sicA	Chaperone of SipBC complex
STM2887, spaS	Involved in a secretory pathway responsible for the surface presentation of effectors needed for the entry of Salmonella species into mammalian cells.
STM2888, spaR	Involved in a secretory pathway responsible for the surface presentation of effectors needed for the entry of Salmonella species into mammalian cells.
STM2889, spaQ	Proposed for surface presentation of antigens
STM2892, invJ	Controls secretion substrate switching and needle length in Salmonella type III secretion system
STM2893, invI	Part of sorting platform that determines the order of protein secretion
STM2896, invA	Export apparatus component
STM2897, invE	Regulate cell invasion complex SipBC
STM2898, invG	Form the neck and outer rings structure of the secretion complex
STM1441, ssaK	Annotated efflux pump related with bile resistance
STM4261, siiE	Host-specific colonization factor
STM0296	
STM0719	UDP-galactopyranose mutase
STM1737, tonB	Interacts with outer membrane receptor proteins that carry out high-affinity binding and energy dependent uptake into the periplasmic space of specific substrates such as cobalamin, and various iron complex
STM3719, rfaB	Lipopolysaccharide 1,6-galactosyltransferase
STM3722, rfaG	α-1,3 glucosyltransferase, which transfers a glucose from UDP-Glc onto L-glycero-D-manno-heptose II, contributing to the core structure of LPS

Table 14 Continued.

Locus tag and gene name	Description
STM3723, rfaQ	3-deoxy- D-manno-octulosonic acid transferase, the enzyme responsible for attachment of the two 3-deoxy-D-manno-octulosonic acid residues that constitute the link between lipid A and the core oligosaccharide of the lipopolysaccharide.
STM3975, tatC	Signal peptide-binding component of the Tat transport system
STM3846, rrtT	Retron reverse transcriptase Typhimurium
STM0522, allP	Putative nucleic acid transporter
STM1636	D-Alanine ABC transporter
STM2437, yfeJ	Glutamyl hydrolase
STM3781	Putative carbohydrate phosphate kinase
STM3342, sspA	RNA polymerase-associated Transcription Factor
STM4067	Putative ADP-ribosylglycohydrolase
STM0398, phoR	Phosphate regulon, sensor protein
STM1230, phoQ	Extracellular Me dependent inner membrane sensor
STM1947, sirA or uvrY	Member of the two-component regulatory system UvrY/BarA involved in the regulation of carbon metabolism via the CsrA/CsrB regulatory system. UvrY activates the transcription of the untranslated csrB RNA and of barA, in an autoregulatory loop. Mediates the effects of CsrA on csrB RNA by BarA-dependent and BarA-independent mechanisms.
STM2958, barA	Member of the two-component regulatory system UvrY/BarA involved in the regulation of carbon metabolism via the CsrA/CsrB regulatory system. Phosphorylates UvrY, probably via a four-step phosphorelay.
STM0031	Putative transcriptional activator
STM0552, fimW	Putative transcription regulator
STM1588, yncC	YncC increases biofilm formation by repressing overproduction of the exopolysaccharide colanic acid
STM3245, tdcA	Putative transcriptional activator of threonine metabolism operon
STM3602	Putative transcriptional regulator of type 3 secretion system and phosphonoacetic metabolism
STM4417, iolR	Negative regulator of Salmonella myo-inositol utilization
STM0278	Putative, swarming related
STM0285	Putative type six secretion system

Table 14 Continued.

Locus tag and gene name	Description
STM1258	Putative ATPase component of ABC-type transport system related with nickel transport
STM1329	Putative inner membrane protein
STM1331	Putative inner membrane protein (obsolete from microbesonline.org)
STM1785	Putative cytoplasmic protein
STM1861	Putative cytoplasmic protein
STM2209	Putative inner membrane protein, related with lipopolysaccharide O-antigen side chain length
STM3026	Putative outer membrane protein
STM3954, yigG	Putative inner membrane protein
STM4030	Putative toxic component of a type II toxin-antitoxin (TA) system.
STM4206	Putatively involved in O antigen modification. Involved in the translocation of bactoprenol-linked glucose across the cytoplasmic membrane.
STM4302	Putative cytoplasmic protein, may related to anaerobic fumarate metabolism
STM4596	Putative inner membrane protein

Table 15 Glutamine amidotransferases in *Salmonella* enterica.

Gene name	Function	Type	COG
COG-GuaA	GMP synthase - Glutamine amidotransferase domain	**Glutamine amidotransferase** type 1	COG518
guaA	GMP synthase, PP-ATPase domain/subunit	**Glutamine amidotransferase** type 1	COG519
hisH	Imidazole glycerol phosphate synthase, subunit H	**Glutamine amidotransferase** type 1	COG118
pabA	para-aminobenzoate synthase, glutamine amidotransferase component II	**Glutamine amidotransferase** type 1	COG512
glmS	glucosamine--fructose-6-phosphate aminotransferase	**Glutamine amidotransferase** type 2	COG449
asnB	asparagine synthetase B	**Glutamine amidotransferase** type 2	COG367
purF	amidophosphoribosyltransferase	**Glutamine amidotransferase** type 2	COG34
carA	carbamoyl-phosphate synthase small chain	**Glutamine amidotransferase** type 1	COG505
yafJ	conserved hypothetical protein, Predicted glutamine amidotransferase	**Glutamine amidotransferase** type 2	COG121
trpD	anthranilate synthase component II	**Glutamine amidotransferase** type 1	COG547
cbiP	putative cobyric acid synthase	**non-classified Glutamine amidotransferase**	COG1492
purG/L	phosphoribosylformylglycineamide synthetase	**Glutamine amidotransferase** type 1	COG46
cbiA	cobyrinic acid A,C-diamide synthase	**non-classified Glutamine amidotransferase**	COG1797
pyrG	CTP synthetase	**Glutamine amidotransferase** type 1	COG504
gltB	glutamate synthase NADPH large chain precursor	**Glutamine amidotransferase** type 2	COG69
COG-ThiJ	Putative intracellular protease/amidase	**Contain Glutamine amidotransferase like domain**	COG693
LdcA	Uncharacterized proteins, homologs of microcin C7 resistance protein MccF; Peptidase S66, LD-carboxypeptidase A	**Contain Glutamine amidotransferase like domain**	COG1619
katE	catalase HPII	**Contain Glutamine amidotransferase like domain**	COG753
Orf 245 protein	Uncharacterized membrane protein	**Contain Glutamine amidotransferase like domain**	COG5426
thiJ	4-methyl-5(b-hydroxyethyl)-thiazole monophosphate biosynthesis protein	**Contain Glutamine amidotransferase like domain**	COG693

Table 15 Continued.

Gene name	Function	Type	COG
COG-ThiJ	Putative intracellular protease/amidase	**Contain Glutamine amidotransferase like domain**	COG693
elbB	Uncharacterized protein involved in an early stage of isoprenoid biosynthesis	**Contain Glutamine amidotransferase like domain**	COG3155
COG5426	Uncharacterized membrane protein	**Contain Glutamine amidotransferase like domain**	COG5426
metA	homoserine O-succinyltransferase	**Contain Glutamine amidotransferase like domain**	COG1897
pepE	peptidase E	**Contain Glutamine amidotransferase like domain**	COG3340