THE STRUCTURE, FUNCTION, AND INHIBITION OF GUT BACTERIAL β-GLUCURONIDASES

Samuel J. Pellock

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry

Chapel Hill
2019

Approved by:

Matthew R. Redinbo

Ian M. Carroll

Gary J. Pielak

Kevin C. Slep

Marcey L. Waters

ABSTRACT

Samuel J. Pellock: The Structure, Function, and Inhibition of Gut Bacterial β-Glucuronidases
(Under the direction of Matthew R. Redinbo**)**

The human gut microbiome is one of the most biochemically rich ecosystems in nature, housing approximately 1000 bacterial species, tens of trillions of cells, and millions of genes. Gut microbes are intimately associated with health outcomes that range from diabetes to depression, yet we have only begun to understand the chemical and biological mechanisms the gut microbiome utilizes to impact host health. One space of this biochemical dark matter is gut bacterial β-glucuronidases (GUSs), glycoside hydrolases that metabolize a myriad of glucuronides in the human gut associated with the dose-limiting toxicities of essential therapeutics. Here we show our efforts to characterize and inhibit gut bacterial GUSs. Structure- and function-guided analysis of GUS genes from the Human Microbiome Project Stool Sample Database revealed three GUSs in a single gut microbe, enabled the discovery of a family of GH2 β-galacturonidases (GalAses) and hybrid GH2 GUS/GalAses, and unearthed a family of novel FMN-binding GUSs in the human gut. Structurally, gut bacterial GUSs demonstrate remarkable diversity for a single enzyme family, in which tertiary structure is conserved, but quaternary structure is highly diverse and a key predictor of substrate specificity. Lastly, we determined the mechanism and performed a structure-activity-relationship of piperazine-containing GUS inhibitors. Inhibition by piperazine-containing inhibitors proceeds via a unique substrate-dependent mechanism that appears to trap GUS during catalysis. We further show that piperazine-containing approved drugs act via the same mechanism, suggesting that approved drugs have off targets in the gut

microbiome. Taken together, the work outlined in this dissertation advances our understanding of the structure, function, and inhibition of gut bacterial GUSs and raises many questions about the core function of bacterial GUS in host physiology.

To anyone reading this, I hope it helps.

ACKNOWLEDGEMENTS

This dissertation was completed with the help of my lab mates, mentors, friends, and family. To Kristina, thanks for being there for me. While I may act like a lone wolf, your support was key for me the past five years, providing a sense of reality and stability during times that felt uncertain and chaotic. To my parents, thank you for your endless support of my obsessions and providing an environment that encouraged reason and curiosity, I would not have been a scientist (or anything at all) without you. To Bill Walton, thanks for sharing your thoughts on data and experiments with me nearly every day. In many ways this dissertation is shared with you, thanks for your help and generosity. To the rest of my lab mates, thanks for the following: Rebecca Pollet for your organization, Aadra Bhatt for your keen advice and wisdom (and cat sitting), Mike Little for your humor and arguments, Kristen Biernat for your caution and skepticism, Sam Ervin for keeping me humble, Marissa Bivins for, again, your caution and skepticism, and Parth, thanks for reminding me of why I do science: to discover, to be innovative, and to have fun. To my undergrads: Ben Creekmore and Andrew Cesmat, thanks for pulling me out of the weeds with your curiosity, I needed it. Lastly, thanks to my mentor, Matt. Thanks for giving me the space to discover on my own. You let me do the science I wanted, which takes trust that I am not sure I deserved, but I gained so much from it. Lastly, thanks for understanding that your students are human beings, you have helped us more in 30 minutes than others have in 5 years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

4-MUG	4-methylumbelliferyl-β-D-glucuronide
Å	Angstrom
Asn	Asparagine
AOM	Azoxymethane
Bac. capillosus	*Bacteroides capillosus*
Bac. fragilis	*Bacteroides fragilis*
Bac. ovatus	*Bacteroides ovatus*
Bac. thetaiotaomicron	*Bacteroides theatiotaomicron*
Bac. uniformis	*Bacteroides uniformis*
*Bu*GUS-1	*Bacteroides uniformis* β-glucuronidase-1
*Bu*GUS-2	*Bacteroides uniformis* β-glucuronidase-2
*Bu*GUS-3	*Bacteroides uniformis* β-glucuronidase-3
Bac. vulgatus	*Bacteroides vulgatus*
Bif. adolescentis	*Bifidobacterium adolescentis*
Bif. angulatum	*Bifidobacterium angulatum*
Bif. bifidum	*Bifidobacterium bifidum*
Bif. breve	*Bifidobacterium breve*
Bif. longum	*Bifidobacterium longum*
Bif. pseudolongum	*Bifidobacterium pseudolongum*
BMSP	*Bacteroides multiple species*
Bry. formatexigens	*Bryantella formatexigens*
*Bv*GUS	*Butyrivibrio fibrisolvens* β-glucuronidase

℃	Celsius
Col. aerofaciens	*Collinsella aerofaciens*
CAZymes	Carbohydrate active enzymes
CBM	Carbohydrate binding module
CNS	Central nervous system
CD	Circular dichroism
C. bartlettii	*Clostridium bartlettii*
C. bifermentans	*Clostridium bifermentans*
C. butyricum	*Clostridium butyricum*
C. clostridioforme	*Clostridium clostridioforme*
C. paraputrificum	*Clostridium paraputrificum*
C. perfringens	*Clostridium perfringens*
*Cp*GUS	*Clostridium perfringens* β-glucuronidase
CYP	Cytochrome P450
DFT	Density functional theory
DMSO	Dimethyl sulfoxide
DUF	Domain of unknown function
E. coli	*Escherichia coli*
*Ec*GUS	*Escherichia coli* β-glucuronidase
EDTA	Ethylenediamine tetra acetic acid
*Ee*GUS	*Eubacterium eligens* β-glucurondase
EFI-EST	Enzyme function initiative-enzyme similarity tool
EFI-GNT	Enzyme function initiative-genome neighborhood tool

Ent. faecalis	*Enterobacter faecalis*
Ent. faecium	*Enterobacter faecium*
*Et*GalAse	*Eisenbergiella tayi* β-galacturonidase
FPLC	Fast protein liquid chromatography
*Fc*GalAse	*Faecalibacterium sp. CAG:74_58_120 β-galacturonidase*
Fp2GUS	*Faecalibacterium prausnitzii L2-6 β-glucuronidase*
FMN	Flavin mononucleotide
*Fs*GUS	*Fusicatenibacter saccharivorans* β-glucuronidase
GalA	Galacturonic acid
GalAse	β-galacturonidase
GUS	β-glucuronidase
GI	Gastrointestinal tract
GND	Genome neighborhood diagram
GlcNAc	N-acetylglucosamine
GlcNS	N-sulfoglucosamine
GlcNS6S	N-sulfoglucosamine-6-sulfate
GlcA	Glucuronic acid
Glu	Glutamic acid
GH	Glycoside hydrolase
GH2	Glycoside hydrolase family 2
HDAC	Histone deacetylase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HGM	Human gut microbiota

HMP	Human microbiome project
HTCS	Hybrid two component system
IBD	Inflammatory bowel disease
IC_{50}	Inhibitor concentration at 50% inhibition
ITC	Isothermal Titration Calorimetry
k_{cat}	Catalytic turnover
K_d	Dissociation constant
K_M	Michaelis constant
L. acidophilus	*Lactobacillus acidophilus*
L. gasseri	*Lactobacillus gasseri*
LB	Lysogeny broth
LIC	Ligation independent cloning
LC-MS	Liquid chromatography-mass spectrometry
L2	Loop 2
Lys	Lysine
MD	Mannonate dehydratase
MO	Mannonate oxidase
MS	Mass-spectrometry
MAM	Methylazoxymethanol
MAM-G	Methylazoxymethanol-glucuronide
MES	2-(N-morpholino)ethanesulfonic acid
mM	millimolar
mL2	Mini-loop 2

nM	nanomolar
NL	No Loop
NSAID	Nonsteroidal anti-inflammatory drug
NTL	N-terminal loop
NxK	Asparagine-x-Lysine
PEG	Polyethylene glycol
p-NP	*p*-nitrophenol
*p*NP-GalA	*p*-nitrophenyl-β-D-galacturonide
*p*NP-GlcA	*p*-nitrophenyl-β-D-glucuronide
P. johnsonii	*Parabacteroides johnsonii*
P. merdae	*Parabacteroides merdae*
PTG	Phenyl-thio-β-D-glucuronide
PL	Polysaccharide lyase
PUL	Polysaccharide utilization loci
PDB	Protein data bank
RG-II	Rhamnogalacturonan II
RMSD	root mean square deviation
*Ri*GUS	*Roseburia inulinivorans* β-glucuronidase
*Rh2*GUS	*Roseburia hominis* β-glucuronidase
*Rg3*GUS	*Ruminococcus gnavus* β-glucuronidase
Ros. inulinivorans	*Roseburia inulinivorans*
Rum. gnavus	*Ruminococcus gnavus*
*Sa*GUS	*Streptococcus agalactiae* β-glucuronidase

SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SD	standard deviation
SEC-MALS	Size-exclusion chromatography-multi-angle light scattering
SN-38	7-ethyl-10-hydroxyl-camptothecin
SN-38-G	7-ethyl-10-hydroxyl-camptothecin glucuronide
SSN	sequence similarity network
Sub. variabile	*Subdilogranulum variabile*
Sus	Starch utilization system
TCA	Tricarboxylic acid cycle
TIM	Triose phosphate isomerase
Trp	Tryptophan
Tyr	Tyrosine
UGT	Uridine diphosphate glucuronosyl transferase
YW motif	Tyrosine-Tryptophan motif
μL	microliter
μM	micromolar

CHAPTER 1: GLUCURONIDES IN THE GUT: SUGAR DRIVEN SYMBIOSES BETWEEN MICROBE AND HOST[1]

β-glucuronidase (GUS) enzymes expressed by the GI microbiota are at the interface of a metabolic symbiosis between microbe and host where they mediate the reactivation of molecules important in host health and disease. Microbial GUS enzymes regenerate toxic drugs and carcinogens in the mammalian GI [1], and their activities are associated with higher incidence of colon cancer and to diets that promote intestinal cancer [2]. Endogenous molecules are also processed by GI GUS proteins, including glucuronides of hormones and neurotransmitters [3–5]. These observations have led to hypotheses linking microbial GUS enzymes to the GI toxicity of drugs, the development of cancer, and increased incidence of Crohn's disease and colitis [2,6–9]. Thus, bacterial GUS enzymes appear to play an important role in health and disease by metabolizing glucuronides in the gut.

GUS proteins catalyze the hydrolysis of glycosidic bonds between glucuronic acid and either small molecules or the terminal ends of polysaccharides. For the purposes of this review, we will focus on small molecule glucuronides generated by Phase II drug metabolism to mark compounds for excretion. Glucuronides are produced by mammalian uridine diphosphate (UDP)-glucuronosyl transferase (UGT) enzymes that append glucuronic acid, derived from UDP-glucuronate, to hydroxyl, carboxylate, and other nucleophilic functional groups of

[1]This chapter previously appeared as an article in the *Journal of Biological Chemistry*. The original citation is as follows: Pellock, S. J., Redinbo, M. R. (2017) Glucuronides in the gut: Sugar-driven symbioses between microbe and host. *J. Biol. Chem. 292*, 8569-8576.

aglycones [10]. Glucuronidation almost exclusively inactivates and detoxifies molecules by increasing their water solubility, which promotes their removal from the body via the kidneys or GI tract [11]. Once in the GI tract, these glucuronides serve as substrates for bacterial GUS proteins that remove the inactivating glucuronic acid moiety. Glucuronic acid then enters the Entner-Doudoroff pathway, a bacterial alternative to glycolysis that catabolizes sugar acids and shunts the resulting pyruvate into the TCA cycle [12]. Mammals also express a GUS enzyme ortholog that is localized to lysosomes in first-pass tissues like liver and intestines, and plays an essential role in degrading endogenous glycosaminoglycans [13]. Germ-line mutations in human GUS cause Sly syndrome, a fatal lysosomal storage disease [14]. Human GUS has also been shown to hydrolyze small-molecule glucuronides, a function that has been leveraged in drug design by attaching drugs to glucuronic acid such that they will be activated at a site of interest upon hydrolysis [15].

As a by-product of glucuronide hydrolysis, bacteria regenerate the original molecule that was eliminated by the host, facilitating reuptake by the GI epithelia and recirculation in the bloodstream [16]. Glucuronidation in the liver, delivery to the GI lumen via the bile duct, reactivation and absorption via the intestinal epithelia, and transport back to the liver is termed enterohepatic circulation (**Figure 1.1**) [17], and it can significantly affect the pharmacokinetics of many drugs and also regulates the levels of endogenous compounds[4,17,18]. Thus, GI microbial GUS enzymes have the capability of directly regulating local and systemic levels of exogenous and endogenous compounds involved in mammalian homeostasis.

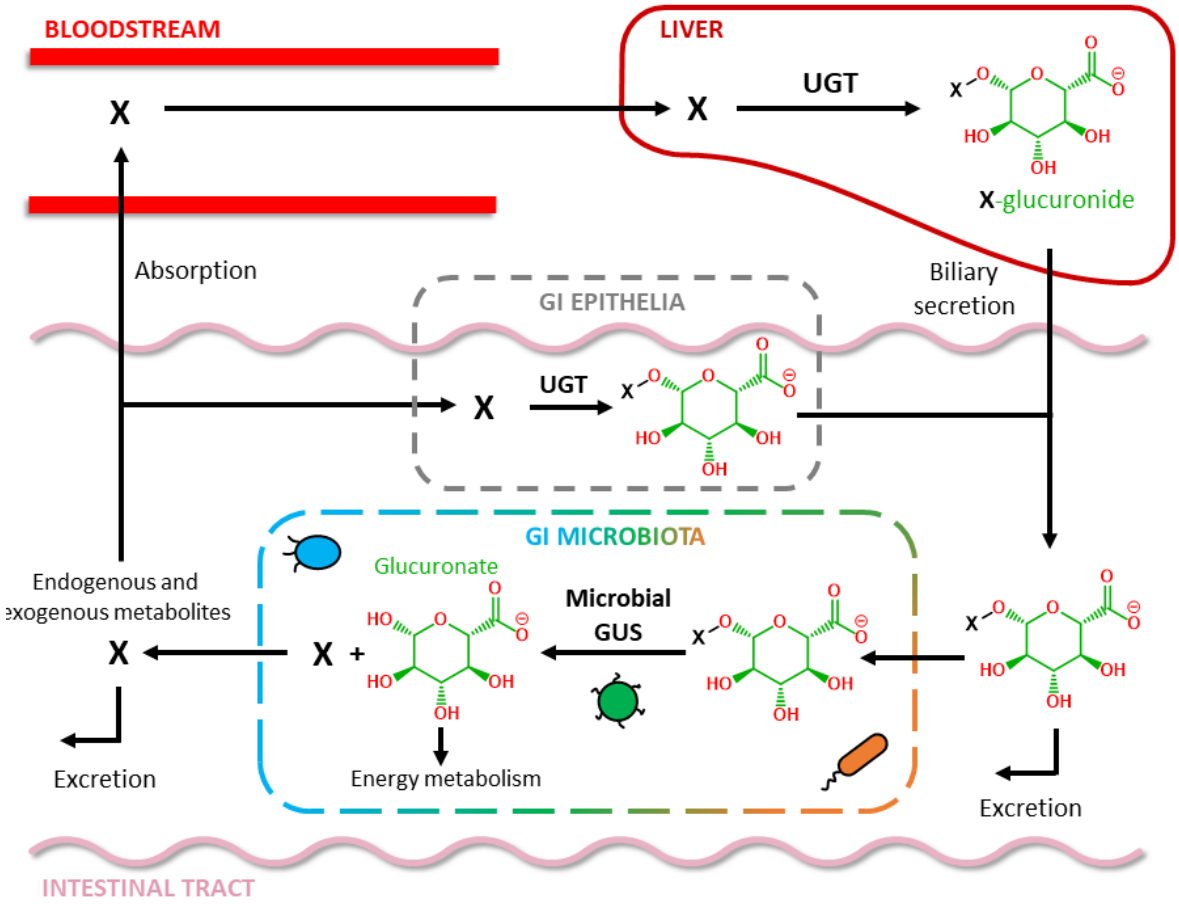


Figure 1.1 - Enterohepatic circulation of chemically distinct molecules (denoted as X) is mediated by the host and microbiota.

Endogenous Glucuronides in the Gut

Endogenous glucuronides were clearly the driving force for the symbiotic evolution of host-associated bacterial GUS enzymes. Glucuronidated endogenous compounds include bilirubin, hormones, neurotransmitters, bile acids, and fatty acids, all of which influence host homeostasis (**Table 1.1** and **Figure 1.2**). As such, GI microbial GUS enzymes participate in a nearly constant mutual symbiosis via the regulation of local and systemic levels of endogenous molecules.

One of the most heavily glucuronidated endogenous molecules is bilirubin, a breakdown product of heme [19]. While it is generally considered a waste product and toxin that contributes to hyperbilirubinemia and neonatal jaundice, normal levels of bilirubin have more recently been shown to have preventative antioxidant activities [20,21]. Approximately 16% and 80% of bilirubin exists as the monoglucuronide and diglucuronide conjugates, respectively, in the bile of healthy humans [19]. Bilirubin glucuronides are generated in the liver by UGT1A1 and enter the GI tract from the bile duct. In the GI, bilirubin glucuronides are heavily metabolized by the intestinal microbiota into stercobilin, which gives feces its brown color, and urobilin, which is responsible for the yellow color of urine and the yellow complexion of jaundiced subjects [22]. The deconjugated bilirubin that manages to escape further metabolism by bacteria is reabsorbed through the GI epithelia and undergoes enterohepatic circulation [18]. However, enterohepatic circulation of bilirubin in healthy humans is relatively low due to bilirubin's nearly complete glucuronidation by the host and substantial subsequent metabolism to stercobilin and urobilin by the GI microbiota. In certain neonates or subjects with Gilbert's syndrome, though, bilirubin is significantly recycled, which contributes to CNS-toxic hyperbilirubinemia [19,20]. Thus, bacterial GUS and human UGT enzymes appear to have co-

Table 1.1 - Examples of molecules subject to glucuronidation in mammals.

Aglycone	Aglycone's Effect	Disease/health
ENDOGENOUS		
Arachidonic acid	Signaling molecule	Inflammation
Bilirubin	Neurotoxin, antioxidant	Gall stones, jaundice
Chenodeoxycholate	Digestion	Cholestasis
Chondroitin sulfate	Glycosaminoglycan	Cancer
Dopamine	GI motility, water absorption	IBD, constipation
Estradiol	Sex hormone, development	Breast cancer
Hyaluronic acid	Glycosaminoglycan	Cancer
Norepinephrine	GI motility	IBD
Serotonin	GI motility	IBD
Testosterone	Sex hormone	Prostate cancer
Thyroxine	Thyroid regulation	Metabolic disorder
EXOGENOUS		
AOM	Alkylating agent	Cancer
Belinostat	HDAC inhibitor	GI toxicity
Benzo[a]pyrene	DNA adduct formation	Cancer
Diclofenac	NSAID	GI toxicity
Ethanol	Depressant	Liver toxicity
Indomethacin	NSAID	GI toxicity
Ketoprofen	NSAID	GI toxicity
Nicotine	Stimulant	Addiction
Panobinostat	HDAC inhibitor	GI toxicity
PhIP	Alkylating agent	Cancer
SN-38	Topoisomerase I inhibitor	GI toxicity

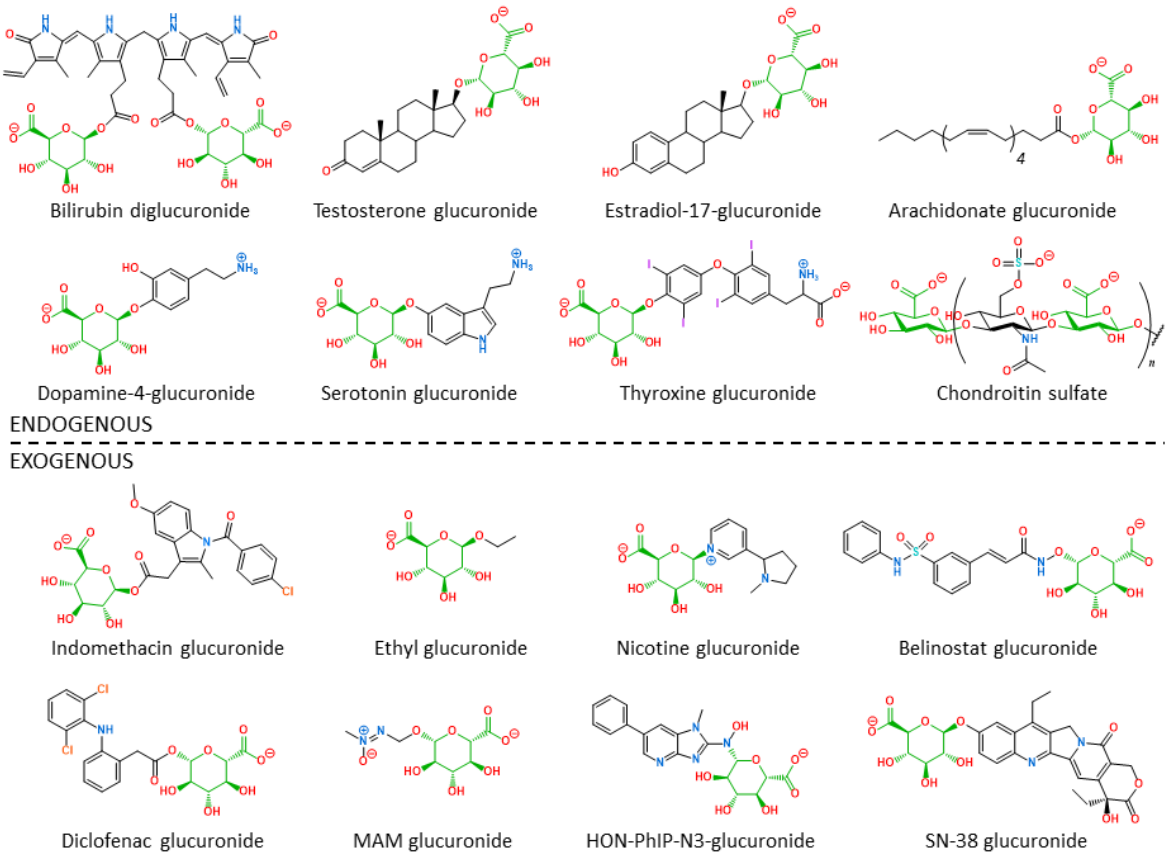


Figure 1.2 - Examples of chemically diverse endogenous and exogenous glucuronides (glucuronic acid shown in green) generated by mammalian UGT enzymes and metabolized by GI microbial GUS enzymes.

evolved a mutually symbiotic heme catabolism pathway to rid the host of high levels of toxic bilirubin and to provide the GI microbiota with a source of energy in the form of glucuronic acid.Unconjugated bilirubin is also capable of forming insoluble calcium salts that contribute to the generation of brown pigment stones in the gallbladder and the biliary ductal system, which reduce bile flow and can impair liver function [23]. Interestingly, the generation of these stones is concomitant with the presence of GUS-expressing Proteobacteria like *Escherichia coli* and *Klebsiella pneumoniae*, suggesting that bacterial GUS activity may promote the formation of the unconjugated bilirubin salts found in gall stones [23,24]. Bacteria of the family *Enterobacteriaceae*, which include *E. coli* and *K. pneumoniae*, are more abundant in the bile [25]. The low affinity GUS inhibitor glucaro-1,4-lactone blocked calcium bilirubinate precipitation *in vitro* [23].

Hormones are also subject to glucuronidation. The estrogen metabolites estradiol, estrone, and estriol are glucuronidated by multiple UGT isoforms [26]. *In vitro* studies have shown that *E. coli* GUS is capable of hydrolyzing a glucuronide metabolite of estriol, and does so with much greater activity than human GUS [27]. Furthermore, estrogen metabolites exhibit significant enterohepatic circulation, suggesting that the regeneration of estrogen aglycones by the GI microbiota may play an important role in regulating plasma levels of this hormone [28]. Radiolabeling studies reveal that enterohepatic circulation of estrone and estriol varies by host species, which suggests that species differences in UGT expression or microbial composition may impact hormone metabolism [29]. While bacterial GUS has been demonstrated *in vitro* to hydrolyze estrogenic glucuronides, a definitive role for the GI microbiota in the enterohepatic circulation of estrogenic metabolites *in vivo* has not been established. However, as posited recently, the reactivation of estrogenic metabolites by the GI microbiota may promote the

enterohepatic circulation of estrogenic metabolites, which may subsequently foster the growth of estrogen-responsive tumors [30]. It is important to note, though, that estrogen metabolites are also heavily sulfated through the action of mammalian sulfotransferases, another set of Phase 2 drug metabolizing enzymes that perform a role analogous to the glucuronidating UGTs [26]. GI bacteria also harbor a variety of sulfatases to process highly sulfated polysaccharides and sulfated small molecules, including estrogen metabolites [26,31]. Overall, mammalian hormone inactivation is likely closely mirrored, and reversed, by enzymes in the GI microbiota.

Other glucuronidated hormones include the androgen testosterone and the thyroid hormone thyroxine [32,33]. Both androgen and thyroxine glucuronides can be hydrolyzed by bacterial GUS enzymes [27,34,35]. Androgens are key drivers of prostate cancer, resulting in therapies primarily focused on androgen deprivation in the form of surgical or chemical castration, although a more recent approach is the enhancement of androgen glucuronidation by UGTs [7]. Thyroxine is a primary thyroid hormone that impacts a variety of processes including metabolic regulation [36]. *In vivo* radiolabeling and *ex vivo* fecal assays indicate that bacteria play a key role in the enterohepatic circulation of thyroxine in mammals [34]. As such, GI microbial GUS proteins could participate in the regulation of metabolism and development by promoting the enterohepatic circulation of thyroxine.

The neurotransmitters dopamine, norepinephrine, and serotonin are glucuronidated in the body and metabolized by bacterial GUS. Roughly 50% of all dopamine is generated in the GI [37], where it acts as a regulator of GI motility and water absorption [38,39]. Microbes were recently shown to have a significant role in the processing of dopamine glucuronide in the GI lumen of mice [5]. This study utilized germ-free mouse models and GUS knockout strains of bacteria to demonstrate that microbial GUS activity is primarily responsible for dopamine

glucuronide hydrolysis. The neurotransmitter norepinephrine, a chemical cousin of dopamine, is also glucuronidated and exhibited microbe-mediated glucuronide hydrolysis in the GI lumen [5]. Similarly, serotonin is subject to glucuronidation, and plasma levels of serotonin in mice fluctuate based on the presence or absence of the microbiota [40,41].

Bile acids are important to gut health and are significantly processed by the microbiota. Bile acids are primarily considered detergents that solubilize dietary components for digestion [42]. Much like bilirubin, bile acids are heavily metabolized by the microbiota, which can dehydrate, oxidize, and deconjugate bile acid variants generated by the liver [43]. In the liver, bile acids are conjugated to sulfate, taurine, and glycine moieties, all of which can be removed by GI microbial sulfatases and bile salt hydrolases. Bile acids are also glucuronidated in the liver [44], and the resulting conjugates account for between 12-36% of the bile acids excreted in the urine. By contrast, sulfate, glycine, and taurine conjugates make up 50-63%, 1.8-28%, and 4.1-8.3% of excreted bile acids in the urine, respectively [45]. Thus, glucuronidated bile acids likely provide a significant energy source to bacteria capable of processing such compounds. Unraveling the connections between host and microbial bile acid metabolism pathways will likely reveal new insights into the co-evolution of mammals and microbes.

Fatty acids are another class of biological detergents processed by liver UGTs. Fatty acids play roles in mammalian biology that range from cell signaling to membrane integrity [46]. *Ex vivo* and *in vitro* analyses show that a variety of fatty acids can be glucuronidated, including arachidonic acid, retinoic acid, prostaglandins, and derivatives of linoleic acid [47–49], although further studies are needed to determine whether fatty acid glucuronides are processed by bacterial GUS enzymes.

Finally, endogenous polysaccharides are a critical source of glucuronides in the gut. Chondroitin sulfate and hyaluronic acid are glucuronic acid-containing polysaccharides present in the GI tract [10]. Bacteria express a wealth of endo- and exo-glycosidases that work in concert to break down complex polysaccharides. Analogous to human GUS, which catabolizes extracellular matrix polysaccharides in lysosomes, bacterial GI GUS enzymes play similar roles with substrates like chondroitin sulfate that enter the GI from host cells sloughed from the epithelia [50]. An excellent review of microbial polysaccharide processing enzymes in the mammalian GI tract has recently been provided [51].

Exogenous Glucuronides in the Gut

Glucuronides of drugs and other exogenous molecules have been a primary focus of research because of their potential importance to therapeutic efficacy and tolerance. Many drugs exhibit GI and liver toxicity that is mediated in part by bacterial GUS activity in the gut, resulting in a parasitic symbiosis in which bacteria receive sugar from drug glucuronides and the host retains toxic metabolites. Carcinogens and other dietary metabolites are also metabolized in the body via glucuronidation and processed by our microbial counterparts, providing a link between bacterial GUS enzymes and carcinogenesis. Exogenous glucuronides that reach the GI are diverse in chemical structure, suggesting that a proportional breadth of functional diversity may be present in the collection of microbial GUS enzymes in the GI.

The anticancer agent SN-38 is the archetype of how metabolism by bacterial GUS can lead to drug toxicity. SN-38 is the active form of the prodrug irinotecan, which is commonly used to treat colorectal and pancreatic cancers [52,53]. SN-38 is inactivated in the liver by conversion to SN-38-glucuronide (SN-38-G); in the GI lumen, however, microbial GUS enzymes recreate SN-38 and cause severe GI toxicity in the form of dose-limiting diarrhea. The authors' laboratory showed that potent, selective, and non-lethal inhibition of bacterial

GUS enzymes reduces the GI toxicity of SN-38 in mice [54–56]. This approach may improve the efficacy and tolerance of other anticancer drugs. Indeed, from a list of 155 anticancer agents, 24 are known to be glucuronidated, and of those that are glucuronidated, 21 (89% of 24) cause GI toxicity. Two such drugs are the histone deacetylase (HDAC) inhibitors belinostat and panobinostat, used to treat lymphoma [57,58]. Metabolites of lapatinib, a GI toxic drug used to treat hormone receptor positive breast cancer, are glucuronidated and their reactivation may damage the liver as well as the GI tract [59]. GI microbial GUS enzymes contribute to hepatotoxicity via the enhancement of enterohepatic circulation, which leads to repeated liver exposure to toxic metabolites [17]. Together, these examples highlight the role that bacterial GUS plays in cancer treatment, efficacy, and toxicity.

Non-steroidal anti-inflammatory drugs (NSAIDs), some of the most widely used therapeutics in the world, are also glucuronidated. NSAIDs inhibit cyclooxygenase enzymes and prostaglandin synthesis and contain a carboxylic acid group that is readily glucuronidated [60]. The NSAID diclofenac is conjugated to glucuronic acid by UGT2B7 in the liver, delivered to the GI tract via the bile duct, and hydrolyzed by bacterial GUS enzymes in the GI [61,62]. The regeneration of diclofenac causes ulceration of the GI epithelia via an unclear mechanism that may involve disruption of mitochondrial function [63]. Similar to SN-38, prevention of diclofenac regeneration by a selective bacterial GUS inhibitor reduced GI ulceration in mice [61]. The GI damage of the NSAIDs ketoprofen and indomethacin can also be ameliorated by selective inhibition of bacterial GUS [62]. Interestingly, the GI toxicity caused by NSAIDs is primarily localized in mice to the distal end of the small intestine, while the damage most often associated with irinotecan is located in the large intestine [64]. It is possible that bacteria that

thrive in the distal small intestine may have a greater capability to hydrolyze NSAID glucuronides than microbes in the proximal small intestine and colon [64].

Certain carcinogens are also glucuronidated. One of the most potent is the alkylating agent methylazoxymethanol (MAM), the active metabolite of azoxymethanol (AOM) that is used to model carcinogenesis in rodents [65]. AOM is converted by cytochrome P450 2E1 and UGTs in the liver to generate MAM-glucuronide (MAM-G), and evidence exists that bacteria in the GI tract reactivate MAM-G to MAM and promote colon carcinogenesis [66,67]. The low-affinity bacterial GUS inhibitor C-GAL has been shown to reduce colon carcinogenesis caused by AOM [68]. Other carcinogens, like the polyaromatic hydrocarbons and heterocyclic aromatic amines, are also metabolized by the CYP-to-UGT pathway, and it has been suggested that microbes hydrolyze those glucuronide metabolites as well [69,70]. Interestingly, colon cancer patients exhibit higher fecal GUS activities than controls [2]. Together, these results support the conclusion that the release of active carcinogens in the GI tract involves microbial GUS enzymes.

Two widely used lifestyle drugs metabolized by host UGTs and bacterial GUSs are ethanol and nicotine. While the majority of ingested ethanol is converted to acetaldehyde by alcohol dehydrogenase, a small fraction of ethanol is glucuronidated [71]. In humans, ethanol glucuronide has been detected in the liver, bile and urine [72]. *Escherichia coli* and *Clostridum sordellii* have both been shown to hydrolyze ethyl glucuronide *in vitro*, which may contribute to a greater retention of ethanol-derived metabolites in the body [73]. Detection of ethyl glucuronide in hair has been employed as a biomarker to diagnose alcohol abuse [74]. Nicotine and its metabolites are primarily processed in humans by oxidation, but they are also glucuronidated [75]. Nicotine is unique among the aglycones discussed here in that it is

conjugated to glucuronic acid through a nitrogen-linkage, and microbial GUS enzymes have been shown to cleave nicotine glucuronide [76]. The glucuronides of ethanol and nicotine highlight the chemical diversity of exogenous compounds that serve as substrates for GUS proteins of the GI microbiota.

While not the primary focus of this review, a small number of plant polysaccharides that contain glucuronic acid are mentioned here. Gum Arabic is a plant-derived secretion that is predominantly composed of glucuronic-acid containing polysaccharides, and is widely utilized in the food and drug industry as a stabilizer [10,77]. This complex polysaccharide is indigestible to animals, but can be fermented by bacteria in the colon and is associated with weight loss in humans [78]. The xylan hemicelluloses, which are heteropolymers of various sugars and components of the plant cell wall, also contain glucuronic acid [10]. Like Gum Arabic, xylan polysaccharides are indigestible by human enzymes, but can be catabolized by GI microbes. Xylan complexity appears to require a diverse set of microbial xylanases to catabolize them to release smaller, glucuronic acid-containing sugars further processed by intestinal bacteria [79].

Microbial β-Glucuronidases in the Gut

Several investigations have detected in vitro GUS activity, ex vivo fecal GUS activity, and in vivo correlations between GI GUS enzyme activity and health. These studies have resulted in the identification of bacteria related to Crohn's disease, the discovery of increased GUS activity in patients with colorectal cancer and subjects on high fat diets, and the mechanistic elucidation of how bacterial GUS promotes drug toxicity [1,2,9,55]. To test the relationship between microbial GUS activity and disease, total fecal proteins have been extracted and GUS assays conducted [80]. This approach yields an overall view of the fecal microbiota's GUS activity, but provides little granularity about the specific microbial

enzymes involved. Other approaches involve culturing bacteria obtained from human fecal samples, and then assessing the GUS activity in these pure cultures [6,81–88]. A tabulation of strains analyzed in culture-based GUS activity assays reveals that bacteria from all the major phyla in the mammalian GI microbiota, including Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria, harbor enzymes that process glucuronides (**Table 1.2**). The conservation of GUS across all major GI bacterial phyla reinforces the hypothesis that GUS proteins may play key roles in chemical dynamics across the intestinal epithelium and serve as a competitive growth advantage for bacteria in the crowded and unforgiving milieu of the mammalian gut.

Table 1.2 - Bacterial strains from the human microbiota that have been shown to exhibit GUS activity in culture.

Strain
Actinobacteria
Bif. adolescentis JCM 1275
Bif. angulatum NCFB 2237
Bif. bifidum NCFB 2454
Bif. breve NCFB 2257
Bif. longum JCM 1217
Bif. pseudolongum NCFB 2244
Col. aerofaciens JCM 7790
Bacteroidetes
Bac. capillosus ATCC 29799
Bac. fragilis NCFB 2217
Bac. ovatus ATCC 8483
Bac. thetaiotaomicron
Bac. uniformis JCM 5828
Bac. vulgatus DCNC 23
P. johnsonii DSM 18315
P. merdae ATCC 43184
Firmicutes
Bry. formatexigens DSM 14469
C. bartlettii DSM 16795
C. bifermentans NCFB 2189
C. butyricum DCNC 19
C. clostridioforme JCM 1291
C. paraputrificum JCM 1293
C. perfringens NCTC 8679
Ent. faecalis DCNC 24
Ent. faecium DCNC 26
Eubacterium L-8
F. prausnitzii M21/2
L. acidophilus DCNC 1237
L. gasseri ADH
Ros. inulinivorans DSM 16841
Rum. gnavus ATCC 29149
Rum. gnavus E1
Sub. variabile DSM 15176
Streptococcus LJ-22
Proteobacteria
E. coli HGU-3

CHAPTER 2: THREE STRUCTURALLY AND FUNCTIONALLY DISTINCT β-GLUCURONIDASES FROM THE HUMAN GUT MICROBE BACTEROIDES UNIFORMIS[2]

INTRODUCTION

The carbohydrates and glycoconjugates that reach the human gastrointestinal (GI) tract are remarkably complex and sample a wide range of structural diversity. Despite the numerous and diverse carbohydrates humans consume, most of the enzymes required to process these molecules are not encoded by the human genome[89]. Fortunately, a mutually beneficial relationship exists between the microbial inhabitants of the GI tract and the human host, in which the human gut microbiota (HGM) expand the host's metabolic capabilities via carbohydrate-active enzymes (CAZymes) [90]. These CAZymes include glycoside hydrolases (GHs) and polysaccharide lyases (PLs) that mediate the fermentation of non-digestible carbohydrates and glycosides [91,92]. The major products of these processes, short chain fatty acids, account for up to 10% of the dietary energy in humans [93] and have been associated with a myriad of health benefits [94–96]. In return, the HGM gain a stable energy source, which is crucial for microbial survival and maintaining balance within the HGM.

The gram-negative phylum Bacteroidetes, one of two dominant bacterial phyla in the human gut microbiome, is a key metabolizer of diverse glycans in the GI tract. Members of

[2]This chapter previously appeared as an article in the *Journal of Biological Chemistry*. The original citation is as follows Pellock, S. J., Walton, W. G., Biernat, K. A., Torres-Rivera, D., Creekmore, B. C., Xu, Y., Liu J., Tripathy, A., Stewart, L.J., Redinbo, M. R. (2018) Three structurally and functionally distinct β-glucuronidases from the human gut microbe *Bacteroides uniformis, J. Biol. Chem. 293,* 18559-18573.

the Bacteroidetes, the majority belonging to the genus *Bacteroides*, degrade both dietary and host-derived carbohydrates, and many species ferment multiple different polysaccharides [89,97]. *Bacteroides thetaiotaomicron*, for example, forages both host mucus glycans and plant polysaccharides, depending on their availability [98]. Accordingly, *Bacteroides* encode genes for large numbers of CAZymes, particularly GHs [91], that are organized in polysaccharide utilization loci (PULs), a distinctive feature of their genomes [99].

Given the thousands of CAZymes that occur in *Bacteroides*, the functional and structural diversity of GHs within individual *Bacteroides* species remain largely unexplored. While the CAZyme classification system, which groups GHs into families based on their amino acid sequences, is reliable for the prediction of catalytic mechanisms and overall structural folds, substrate specificity and unique structural features are difficult to predict. For example, the GH2 family comprises β-glucuronidases, β-glucosidases, β-galactosidases, and β-mannosidases, all of which possess an $(\alpha/\beta)_8$ TIM barrel fold [91]. Thus, it is important to experimentally characterize GHs to understand their structure and to assign function.

Recently, we presented a structure-guided approach to differentiate β-glucuronidase (GUS) proteins from their GH2 family members [100]. As reported, 279 unique GUS enzymes were identified from the 4.8 million unique genes present in the stool sample database of the Human Microbiome Project (HMP) [100]. This provided the first atlas of GUS enzymes in the human gut microbiome. Within that effort, we identified and characterized a GUS from the human gut bacterium *Bacteroides uniformis*, which has been reported to be highly abundant in the human GI tract [101]. We demonstrated that it acts as a β-glucuronidase and is able to process both a small-molecule glucuronide and a polysaccharide with a terminal glucuronic acid moiety [100]. As outlined below, in an attempt to gain further insight into its role in

polysaccharide degradation, we searched the genomic region surrounding this GUS (**Fig. 1*a***). We found two additional GH2 enzymes in the same PUL that retain sequence features previously identified as unique to GUS enzymes (**Fig. 1*a***). To our knowledge, no previously characterized PUL contains three potential GUS enzymes. For this reason, we were interested in their differential structural properties and their abilities to cleave diverse glucuronic acid (GlcA)-containing substrates.

Here we demonstrate that three putative GUS proteins from a single *B. uniformis* microbe share the TIM barrel structural fold but exhibit distinct tertiary and quaternary structures, not obvious from sequence analysis, and harbor unique structural features within their actives sites that likely afford them specific substrate processing capabilities. Indeed, these GUS enzymes displayed differential activities towards a variety of glucuronide substrates, including GlcA-containing polysaccharides and SN-38-G, a metabolite of the cancer drug irinotecan. Additionally, we tested the ability of both selective bacterial GUS inhibitors and a pan GUS inhibitor to inhibit the three GUSs from this microbe, which reveal distinct propensities for inhibition. We further examined the potential for these glucuronidases to act on other sugar acid-containing substrates, such as those that contain galacturonic acid, iduronic acid or mannuronic acid. These results highlight the broad structural and functional diversity among GUS enzymes within a single human gut microbe. Furthermore, the data presented here provide a foundation for understanding the specialized roles of GUS enzymes in the deconstruction of a sugar acid-containing carbohydrate and the ability of the HGM to reactivate drug-glucuronide conjugates.

RESULTS

Discovery and sequence analysis of GUS enzymes from a *B. uniformis* PUL

A GUS from the human gut bacterium *B. uniformis* strain 3978 T3 i (*Bu*GUS) was previously discovered in the HMP database [100]. Further inspection of the genomic region flanking this GUS gene revealed a hallmark of PULs, a nearby *susC*/*susD*-like gene pair. These two genes are involved in the binding of polysaccharides on the outer membrane (SusD) and transport into the periplasm (SusC) (**Figure 2.1*b***) [99]. The presence of the *susC/susD* homologs indicates that *Bu*GUS is located in a PUL, which means *Bu*GUS likely contributes to the orchestrated degradation of a GlcA-containing polysaccharide. Two additional enzymes predicted to belong to the GH2 family were also identified adjacent to *Bu*GUS and the *susC/susD*-like pair (**Figure 2.1*b***). Each of these proteins possess key sequence features that are characteristic of GUS enzymes, including the asparagine-x-lysine (NxK) motif and catalytic glutamates that recognize and cleave glucuronides, respectively [100,102] (**Figure 2.1*a***). Only *Bu*GUS-1 and *Bu*GUS-2, however, possess the GUS-specific tyrosine residue (Y480 and Y495, respectively) that hydrogen bonds with the nucleophilic glutamate. In *Bu*GUS-3, a tryptophan (W483) replaces the tyrosine (**Figure 2.1*a*** and **2.2**).

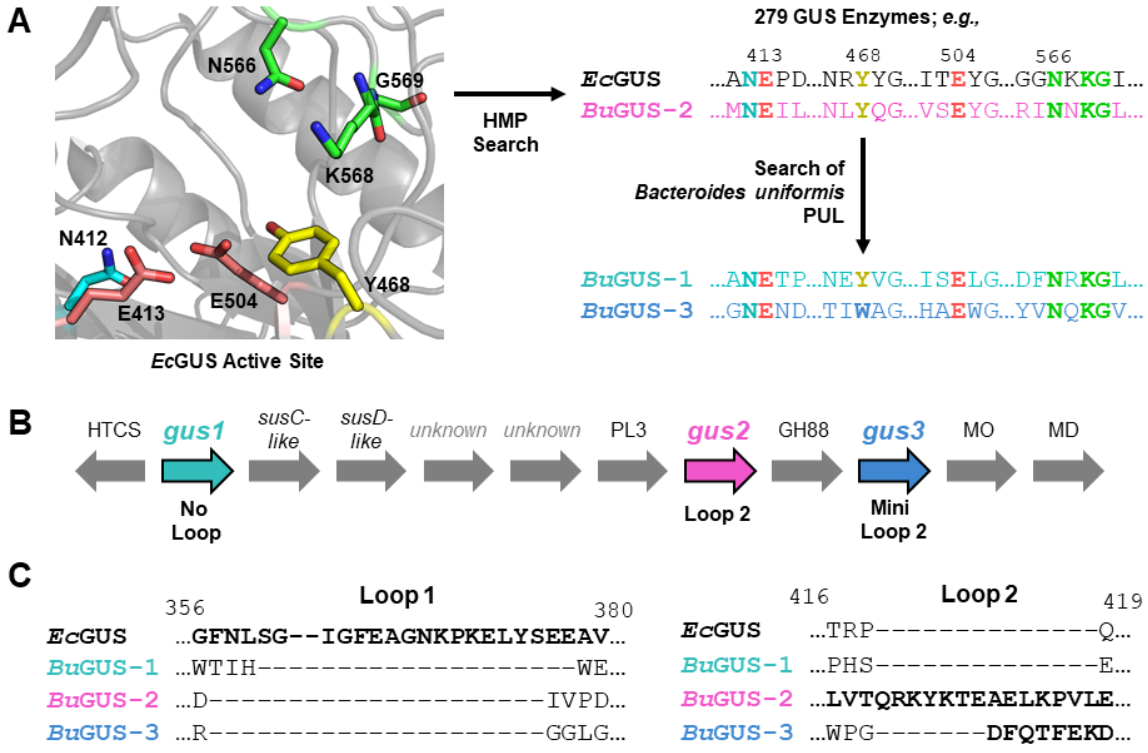


Figure 2.1 - Discovery and analysis of GUS genes in *B. uniformis* strain 3978 T3 i. (**A**) Schematic for the discovery of GUS enzymes from *B. uniformis b*, genetic organization of a PUL from *B. uniformis* reveals 3 glycoside hydrolases with sequence features indicating GUS function, as well as other glucuronic acid metabolizing enzymes. *c*, sequence alignment of *Ec*GUS, *Bu*GUS-1, *Bu*GUS-2, and *Bu*GUS-3 reveals distinct loop classes between these putative GUS enzymes. HTCS: hybrid two-component system, PL: polysaccharide lyase, MO: mannonate oxidase, MD: mannonate dehydratase, susC/D: starch utilization system C/D.

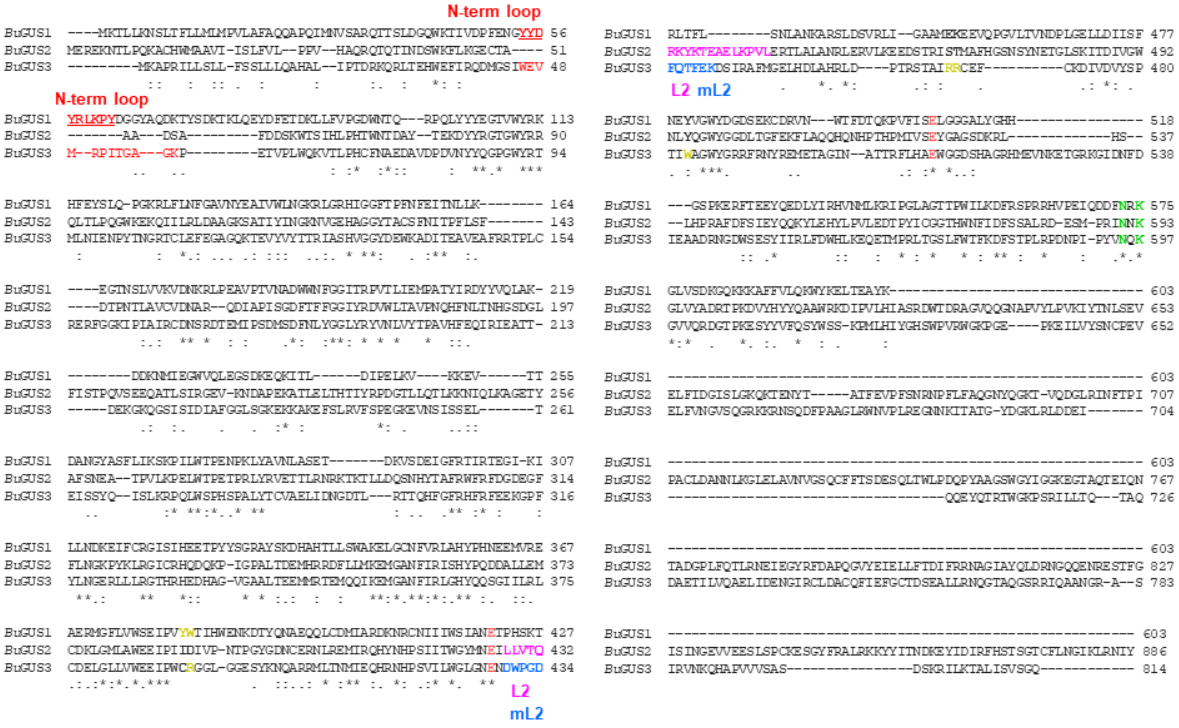


Figure 2.2 - Multiple sequence alignment of *Bu*GUS-1, *Bu*GUS-2, and *Bu*GUS-3.

Regions of interest discussed in the text are bolded and highlighted in various colors. Unique active site residues highlighted in yellow, N-terminal loop and catalytic glutamates in red, NxK motif in green, Loop 2 (L2) in magenta, mini-loop 2 (mL2) in blue.

Sequence alignments with the previously characterized *Bu*GUS (now termed *Bu*GUS-2) and these two new GUS enzymes, termed *Bu*GUS-1 and *Bu*GUS-3, revealed a sequence identity of 27% and 29%, respectively, while *Bu*GUS-1 and *Bu*GUS-3 share 18% sequence identity (**Figure 2.2**). Sequence analysis also revealed that *Bu*GUS-1, *Bu*GUS-2, and *Bu*GUS-3 fall into the previously defined No Loop (NL), Loop 2 (L2), and Mini Loop 2 (mL2) classes, respectively; these classifications are related to the size and location of loops at the active site of gut microbial GUS enzymes, and have been shown to play key roles in substrate specificity [100] (**Figure 2.1*c***). Utilizing the signal peptide prediction tool, SignalP 4.1 Server [103], we found that *Bu*GUS-1, *Bu*GUS-2, and *Bu*GUS-3 have a signal peptide and are thus expected to be periplasmic. Together, this sequence analysis indicates that a PUL from *B. uniformis* contains three putative GUS enzymes with distinct sequence features.

*Bu*GUS-1 and *Bu*GUS-2 exhibit β-glucuronidase activity with 4-MUG

To begin to elucidate the substrate specificities of these three putative GUS enzymes, we performed *in vitro* activity assays with their purified protein products. *Bu*GUS-2 has been previously shown to exhibit GUS activity [100]. To confirm that *Bu*GUS-1 and *Bu*GUS-3 are also GUS enzymes, we synthesized, cloned, expressed, and purified their protein products. We then utilized the standard substrates 4-methylumbelliferyl-β-D-glucuronide (4-MUG) and *p*-nitrophenyl-β-D-glucuronide to assess the pH profile and kinetic parameters of GUS activity, respectively. *Bu*GUS-1 ($k_{cat}/K_m = 3.4\text{x}10^5$ s^{-1} M^{-1}) and *Bu*GUS-2 ($k_{cat}/K_m = 3.8\text{x}10^5$ s^{-1} M^{-1}) both efficiently processed these standard substrates, indicating that they are GUS enzymes (**Table 2.1** and **Figure 2.3**). However, *Bu*GUS-3 was unable to catalyze the hydrolysis of these substrates (**Table 2.1**). Thus, while *Bu*GUS-1 and *Bu*GUS-2 can

Table 2.1 - Kinetic parameters of 4-MUG hydrolysis by *Bu*GUS enzymes and BMSP GUS.

Protein	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (s^{-1} M^{-1})
***Bu*GUS-1**	8.6 ± 0.5	25 ± 8	$3.4x10^5$
***Bu*GUS-2**	31 ± 5	80 ± 20	$3.8x10^5$
***Bu*GUS-3**	NA	NA	NA
BMSP GUS	1.2 ± 0.2	120 ± 30	$1.0x10^4$
***Bu*GUS-1 Δloop**	1.2 ± 0.2	74 ± 8	$1.7x10^4$

Values are an average of n = 3 biological replicates ± SD.
NA = No Activity

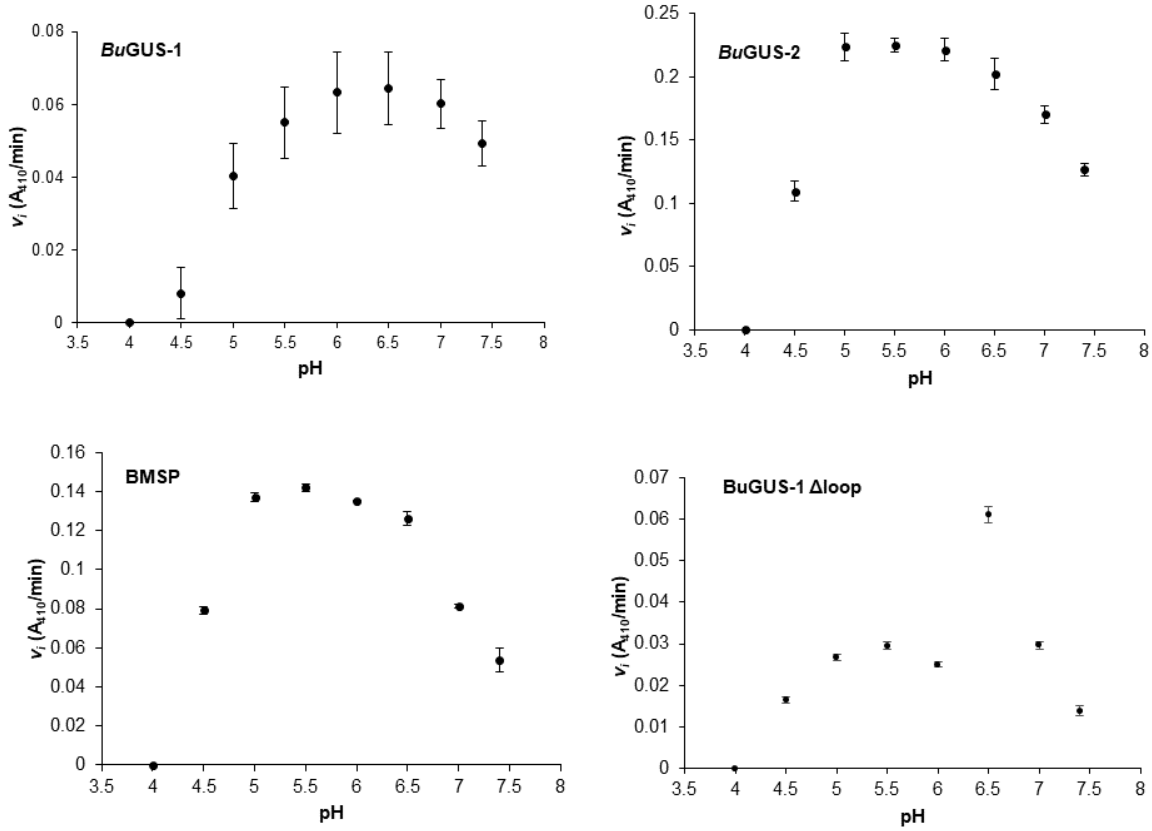


Figure 2.3 - pH screen for *Bu*GUS-1, *Bu*GUS-2, and *Bu*GUS-3. Error bars represent SD of n = 3 biological replicates.

hydrolyze glucuronides, *Bu*GUS-3 may have a distinct but related activity despite its GUS-like sequence features.

_Bu_GUS enzymes exhibit distinct structural features

To evaluate the structural diversity of the three GUS enzymes present in *B. uniformis*, we determined the crystal structures of *Bu*GUS-1 (space group: C2, molecules in asymmetric unit: 2) and *Bu*GUS-3 (space group: I2, molecules in asymmetric unit: 2) and compared them to the previously reported *Bu*GUS-2 structure (**Table 2.2**) [100]. *Bu*GUS-1 and *Bu*GUS-3 share a similar TIM barrel core fold with *Bu*GUS-2, with 3.2 Å root-mean-square deviation (rmsd) across 512 Cα positions and 3.4 Å rmsd over 648 Cα positions, respectively (**Figure 2.4** and **Figure 2.5**). *Bu*GUS-2 and *Bu*GUS-3 encode C-terminal extensions not found in the shorter *Bu*GUS-1 enzyme (**Figure 2.4A, B, C**).

While *Bu*GUS-1 maintains a similar tertiary structure to other GUS enzymes, it possesses unique active site residues, particularly Y382 and W383 (**Figure 2.4A**, highlighted in yellow). These positions are generally occupied by smaller residues in previously characterized GUS enzymes, such as *Bu*GUS-2 (**Figure 2.4B**). In addition to these unique active site residues, *Bu*GUS-1 is only the second tetrameric bacterial GUS characterized that does not contain a Loop 1 by sequence analysis. Instead, remarkably, and unpredictably by sequence analysis alone, an N-terminal loop (NTL; **Figure 2.2**) is donated from an adjacent protomer and resembles the loop-based active sites of previously characterized Loop 1 GUS structures, as outlined below.

Compared to previously characterized GUS enzymes, *Bu*GUS-3 deviates the most in its active site composition, containing five unique residues (yellow) (**Figure 2.4C**). Most

Table 2.2 - X-ray data collection and refinement statistics.

Protein	*Bu*GUS-1	*Bu*GUS-3	BMSP	*Bu*GUS-1 + G-1,5-L	*Bu*GUS-2 + G-1,5-L	*Bu*GUS-1 + GlcA	*Bu*GUS-1 Δloop	*Bu*GUS-2 ΔCa^{2+}	*Bu*GUS-1 + TPG
Resolution (Å)	29.6 – 2.20	29.48 – 2.35	29.43 – 2.65	29.93-1.90	29.48-2.50	29.98-1.80	29.34-2.00	29.55-3.00	29.30-2.40
Resolution in highest shell (Å)	2.25 – 2.20	2.39 – 2.35	2.71 – 2.65	1.93 -1.90	2.56-2.50	1.83-1.80	2.03-2.00	3.12-3.00	2.49-2.40
Space group	C2	C2	$I4_1$	C2	$P2_12_12_1$	C2	$P1 2_1 1$	$P2_12_12_1$	$P3_221$
Unit cell dimensions a,b,c (Å); α,β,γ (°)	125.8,130.2, 94.2; 90, 100.3, 90	209.7, 68.2, 167.4; 90, 122.5, 90	213.5, 213.5, 112.6; 90, 90, 90	126.1, 130.2, 94.5; 90, 100.2, 90	74.5, 141.9, 180.8; 90, 90, 90	134.2, 133.9, 163.4; 90,100.3, 90	76.7,199.2, 84.6; 90,102.8, 90	75.0,142.0, 181.2; 90, 90, 90	252.2, 252.2, 107.3 90, 90, 120
Total reflections (F>0)	517245	568681	305377	797212	446802	1761568	995295	525999	3085366
Unique reflections	75592	83443	73245	114979	67106	261960	163378	39506	152030
Multiplicity	6.8 (7.0)	6.8 (6.9)	4.2 (4.2)	6.9 (6.9)	6.7 (6.8)	6.7 (6.9)	6.1 (5.8)	13.3 (14.0)	20.3 (21.0)
Completeness, %(highest shell)	99.9 (100)	99.9 (99.9)	99.7 (100)	97.7 (96.3)	99.9 (100.0)	100 (100)	98.4 (97.7)	99.8 (100)	99.8 (99.7)
Mean I/sigma (I) (highest shell)	18.6 (4.7)	21.0 (4.9)	16.1 (2.6)	16.6 (4.4)	15.9 (4.1)	11.8 (4.3)	12.1 (3.7)	20.9 (3.9)	23.4 (7.0)
Wilson B-factor ($Å^2$)	21.4	31.9	61.8	14.9	36.5	8.3	25.6	67.7	33.1
R_{merge} (highest shell)	0.078 (0.387)	0.073 (0.445)	0.060 (0.50)	0.081 (0.351)	0.084 (0.438)	0.126 (0.471)	0.118 (0.496)	0.120 (0.69)	0.111 (0.511)
R_{work}	0.1494	0.1755	0.2069	0.1377	0.1703	0.18	0.176	0.1670	0.1739
R_{free}	0.1841	0.2224	0.2827	0.1694	0.2277	0.229	0.23	0.2409	0.2467
Number of molecules in asymmetric unit	2	2	4	2	2	4	4	2	6
Number of water molecules in asymmetric unit	1075	815	110	1549	535	4087	1808	0	1561
rms bonds (Å)	0.003	0.009	0.008	0.012	0.004	0.006	0.007	0.009	0.007
rms angles (°)	0.631	1.24	1.000	1.072	0.669	0.874	0.849	1.030	0.94
Ramachandran favored (%)	96.53	95.34	92.85	96.79	96.23	96.71	96.30	92.91	92.23
Ramachandran outliers (%)	0.00	0.39	0.19	0.00	0.24	0.00	0.09	0.18	0.20
Clash score	3.04	6.87	13	1.96	2.12	7.05	5	8.33	16.84
Average B-factor ($Å^2$)	28.3	35.1	67.0	21.2	41.0	14.6	30.0	57.9	37.9
RCSB ID	6D1N	6D1P	6D8K	6D41	6D5O	6D6W	6D89	6D8G	6D7F

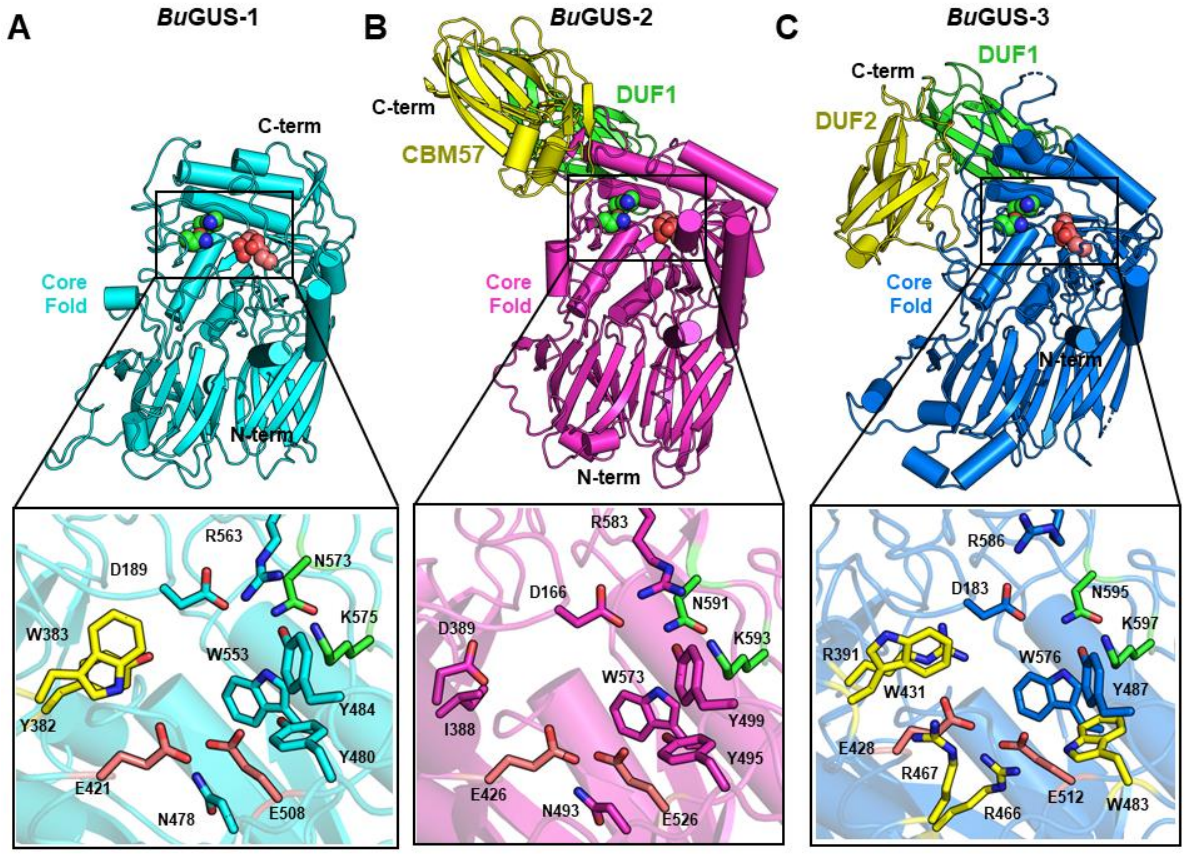


Figure 2.4 - Structural analysis of *Bu*GUS-1, *Bu*GUS-2, and *Bu*GUS-3 reveals distinct tertiary and active site structure. (**A**) tertiary structure of BuGUS-1 with the sugar acid-recognizing NxK motif highlighted as green spheres and catalytic glutamates as deep salmon spheres and zoom-in of active site with unique active site residues highlighted in yellow. (**B)**, tertiary structure of BuGUS-2 with core fold highlighted in magenta and additional C-terminal domains in green (DUF1) and yellow (CBM57) and zoom-in of the active site. (**C)**, tertiary structure of *Bu*GUS-3 with core fold in blue, the sugar acid-recognizing NxK motif highlighted as green spheres and catalytic glutamates as deep salmon spheres and additional C-terminal domains in green (DUF1) and yellow (DUF2) and zoom-in of active site with unique active side residues highlighted in yellow.

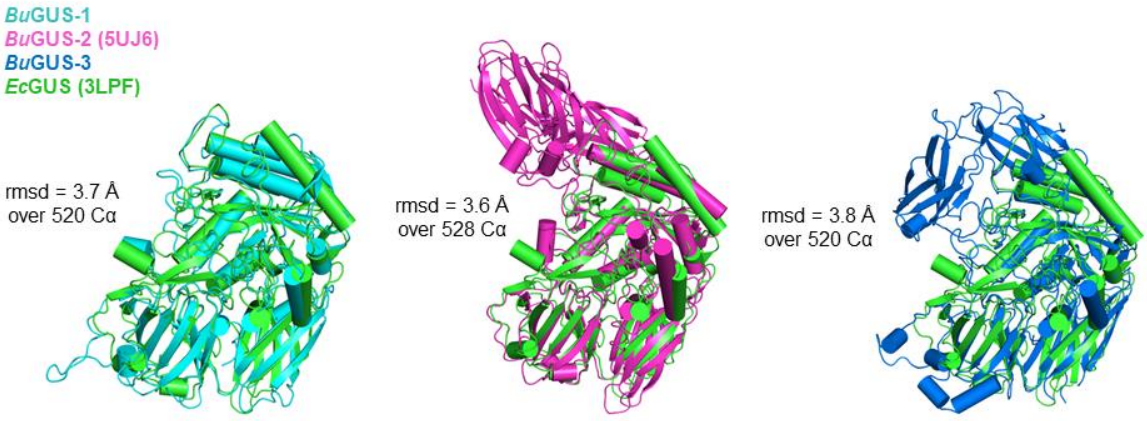


Figure 2.5 - Structural alignments of *Bu*GUS-1 (cyan), *Bu*GUS-2 (magenta), and *Bu*GUS-3 (blue) with previously elucidated structure of *E. coli* GUS (green, *Ec*GUS, PDB: 3LPF).

notably, three arginines in *Bu*GUS-3 replace the small, polar residues, such as asparagine, that are conserved in previously characterized GUS enzymes, and R391 and R466 are positioned to form ionic interactions with the catalytic glutamates (**Figure 2.4C**). Furthermore, the *Bu*GUS-3 active site contains W483, which replaces the conserved tyrosine in all other GUS enzymes characterized. Finally, W431 and R391, localized across the active site from the NxK motif (green), have not been observed in any other GUS enzymes (**Figure 2.4C**). These distinct active site features may explain why *Bu*GUS-3 does not process the standard glucuronide substrates despite the presence of the NxK motif and catalytic glutamates necessary for the recognition and cleavage of glucuronides.

*Bu*GUS-1 displays a unique quaternary structure despite having similar tertiary structure to *Ec*GUS (**Figure 2.6A** and **2.5**). *Bu*GUS-1 forms a unique inverted tetramer in comparison to the previously determined structures of *E. coli* GUS (*Ec*GUS), *Streptococcus agalactiae* GUS, *Clostridium perfringens* GUS, and the human GUS ortholog (**Figure 2.6A, B**) [13,55,102]. In *Bu*GUS-1, individual protomers interact via their N-termini in comparison to the previously examined GUS enzymes outlined above, in which the interface of protomers is formed by their C-termini (**Figure 2.6A, B**). The consequence of this oligomeric organization is a solvent exposed active site.

Unlike the tetrameric *Bu*GUS-1, *Bu*GUS-2 and *Bu*GUS-3 both form dimers (**Figure 2.7A,B** and **2.8**) and contain extra domains at their C-termini (**Figs**. **2.4B, C** and **2.8**). Excluding these additional domains, the core tertiary structures of *Bu*GUS-2 and *Bu*GUS-3 are TIM barrel folds with two beta-sandwich-like domains, similar to previously characterized GUS enzymes (**Figure 2.4B, 2C** and **2.5**). While *Bu*GUS-2 and *Bu*GUS-3 share similar core folds, they have distinct quaternary structures, likely a result of how

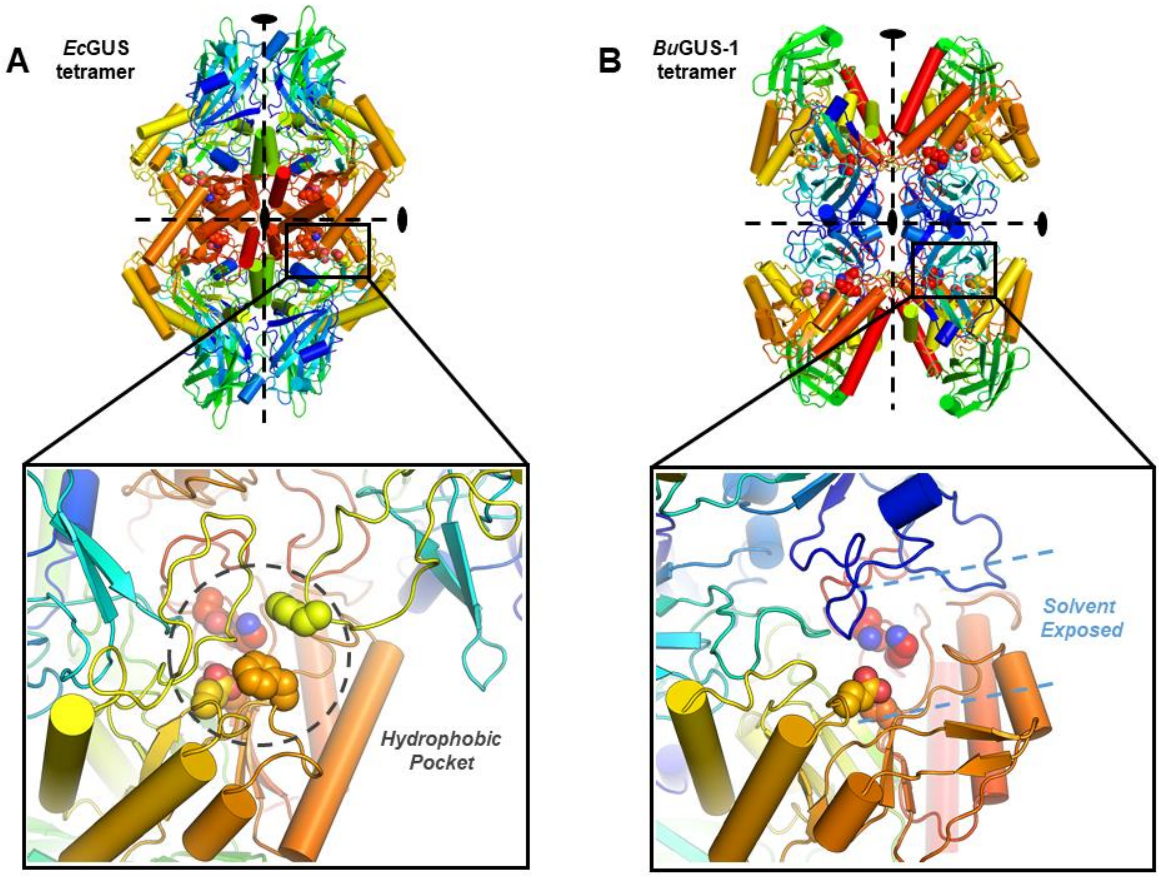


Figure 2.6 - Analysis of quaternary structures and their influence on the active site architecture of *Ec*GUS and *Bu*GUS-1. (A) Tetramer of *Ec*GUS with zoom-in of tetramer interface reveals hydrophobic pocket around the active site situated at the interface of C-terminal regions. (**B**) Tetramer of *Bu*GUS-1 with zoom-in of tetramer interface reveals a solvent exposed active site.

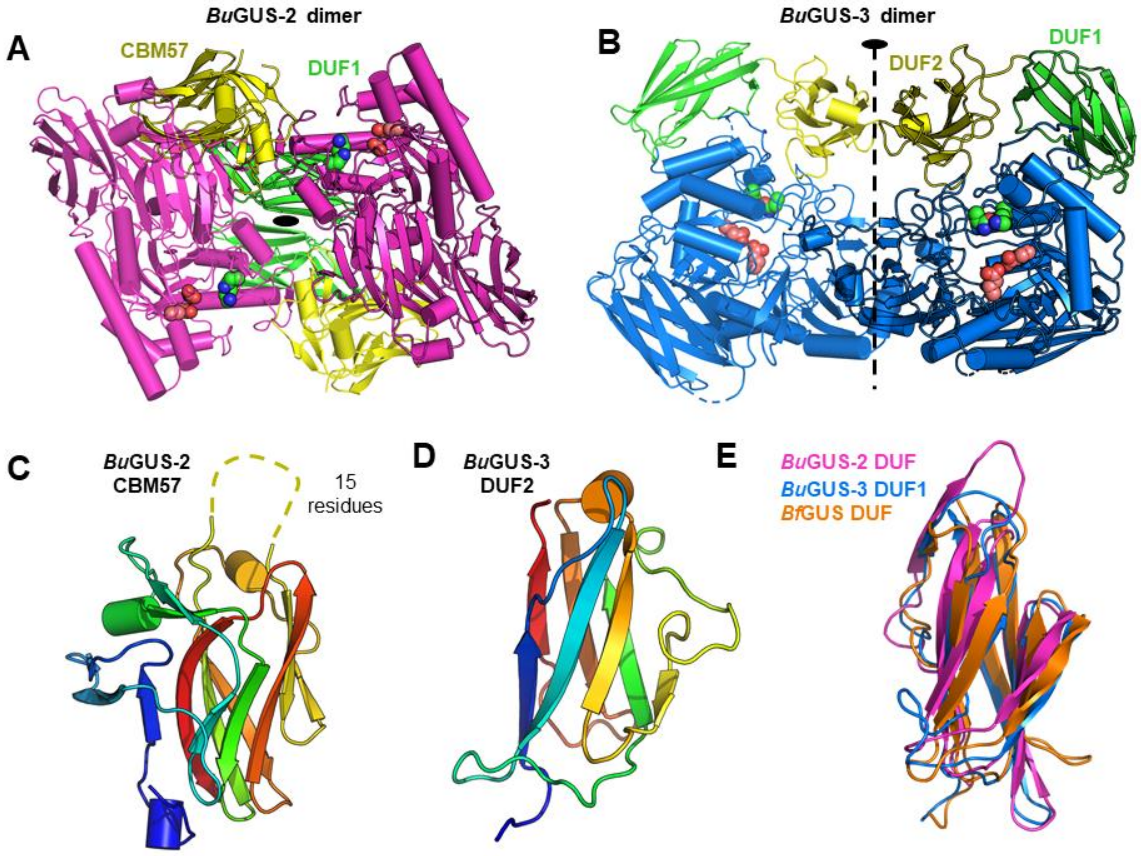


Figure 2.7 - Quaternary structure of BuGUS-2 and BuGUS-3 and structural analysis of C-terminal domains. (**A**) *Bu*GUS-2 dimer with core fold shown in magenta, DUF1 in green, and CBM57 domains in yellow with active site glutamates and NxK motif shown as deep salmon and green spheres, respectively. (**B**) *Bu*GUS-3 dimer with core fold in blue, DUF1 in green, and DUF2 in yellow with catalytic glutamates and NxK motif in deep salmon and green spheres, respectively. (**C**) CBM57 of *Bu*GUS-2 shown with disordered loop shown as dotted line. (**D**) Structure of *Bu*GUS-3 DUF2. € Structural alignment of *Bu*GUS-2 DUF, *Bu*GUS-2 DUF1, and *Bf*GUS DUF.

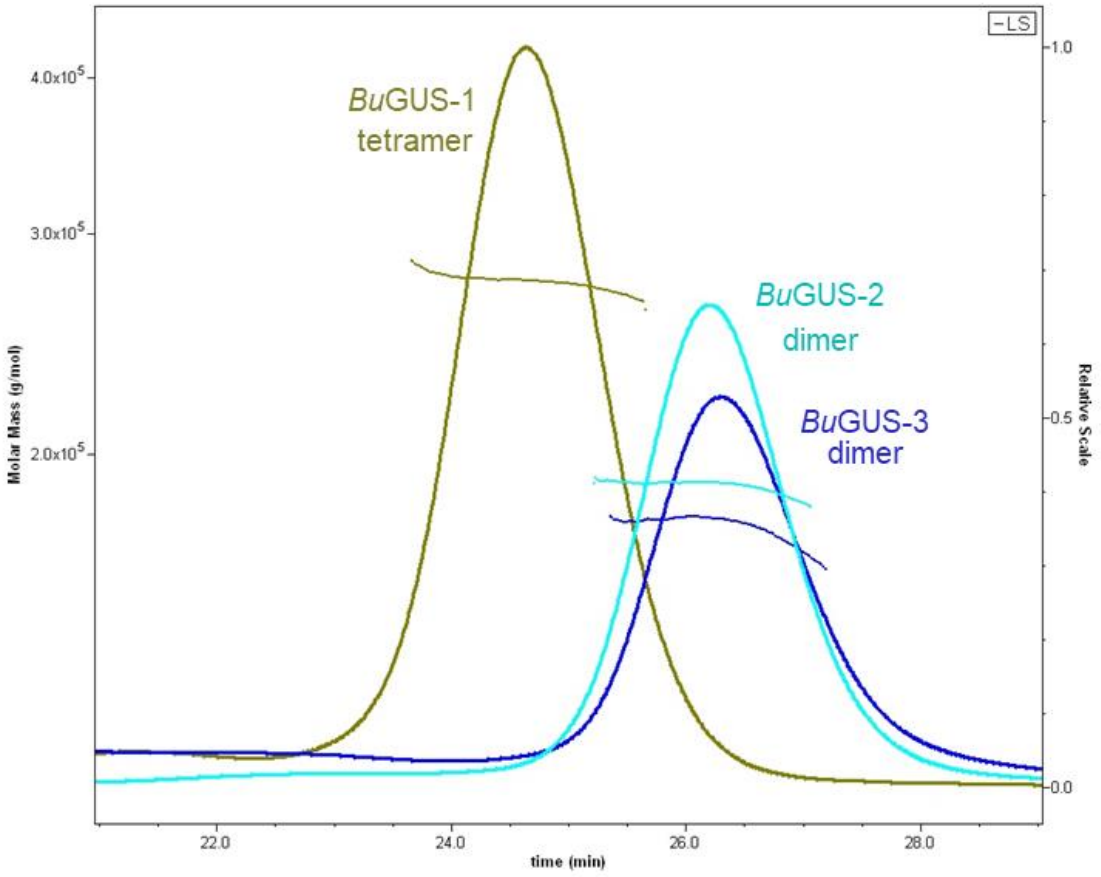


Figure 2.8 - Size-exclusion chromatography multi-angle light scattering analysis (SEC-MALS) of *Bu*GUS-1, *Bu*GUS-2, and *Bu*GUS-3 confirms oligomeric states predicted from crystal structures.

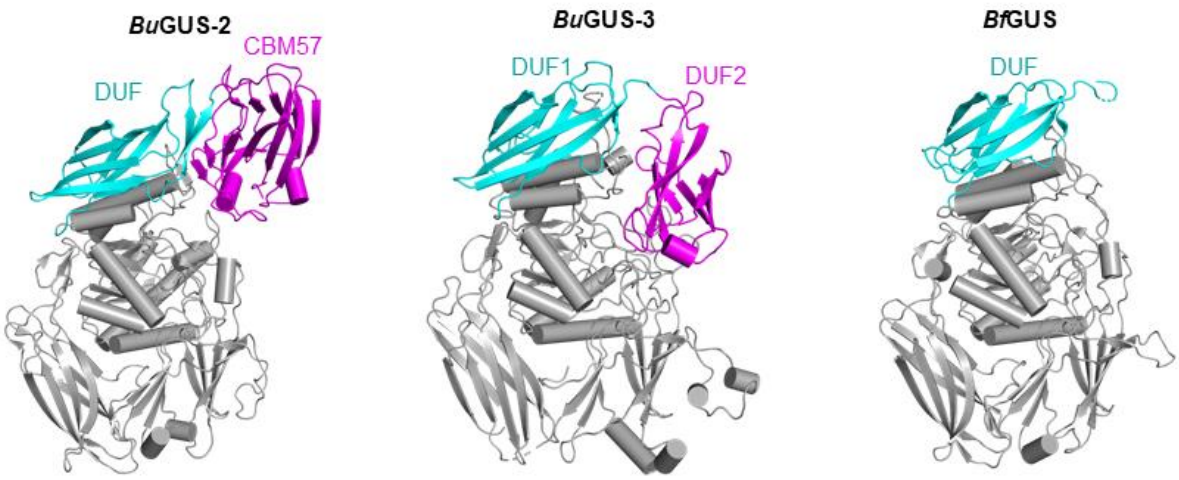


Figure 2.9 - Comparison of CBMs of BuGUS-2, BuGUS-3, and BfGUS. Conserved core fold of GUS shown in grey with DUF-1 and CBM 57 or DUF2 (for *Bu*GUS-2 and *Bu*GUS-3) highlighted in cyan and magenta, respectively.

their additional C-terminal domains are positioned that preclude similar homodimeric organizations (**Figure 2.4B, 2.4C, 2.7A, 2.7B,** and **2.9**). Sequence and structural analysis of the C-terminal domains of *Bu*GUS-2 revealed that the most C-terminal (yellow) is a member of the CBM 57 family, based on malectin that binds to developing glycans in the endoplasmic reticulum (**Figure 2.7C**) [104]. The remaining domains in *Bu*GUS-2 and *Bu*GUS-3 are "domains of unknown function" (DUF) and are not formally defined as, and simply may not be, CBMs (**Figure 2.4B, 2.4C,** and **2.9).** Both sequence (NCBI BLAST) and structure-based (PDBeFold) searches of these additional C-terminal DUFs revealed hits for the C-domains of antibodies. These domains may only serve a role in the oligomeric organization of these proteins. The DUFs from the *Bu*GUS enzymes are similar in structure and have been observed once previously in the structure of *Bacteroides fragilis* GUS (*Bf*GUS), which was also previously designated as a DUF (**Figure 2.9**). Collectively, the unique C-terminal domains of *Bu*GUS-2 and *Bu*GUS-3 may play roles in carbohydrate binding and quaternary structure.

We also find that *Bu*GUS-2 contains a well-organized predicted calcium binding site (**Figure 2.10A, B**) that is unique to this GUS both in *B. uniformis* and in GUS enzymes of known structure to date. Approximately 24 Å from the active site of *Bu*GUS-2 are three aspartic acids and three ordered water molecules that coordinate a predicted calcium ion (**Figure 2.10A, B**). Site-directed mutagenesis of D341 and D367 to alanines led to a complete loss of GUS activity and the crystal structure of this mutant (space group: $P2_12_12_1$, molecules in asymmetric unit: 2, **Table S1**) revealed significant structural changes at the enzyme active site (**Figure 2.10 C, D).** Circular dichroism analysis also revealed a small loss

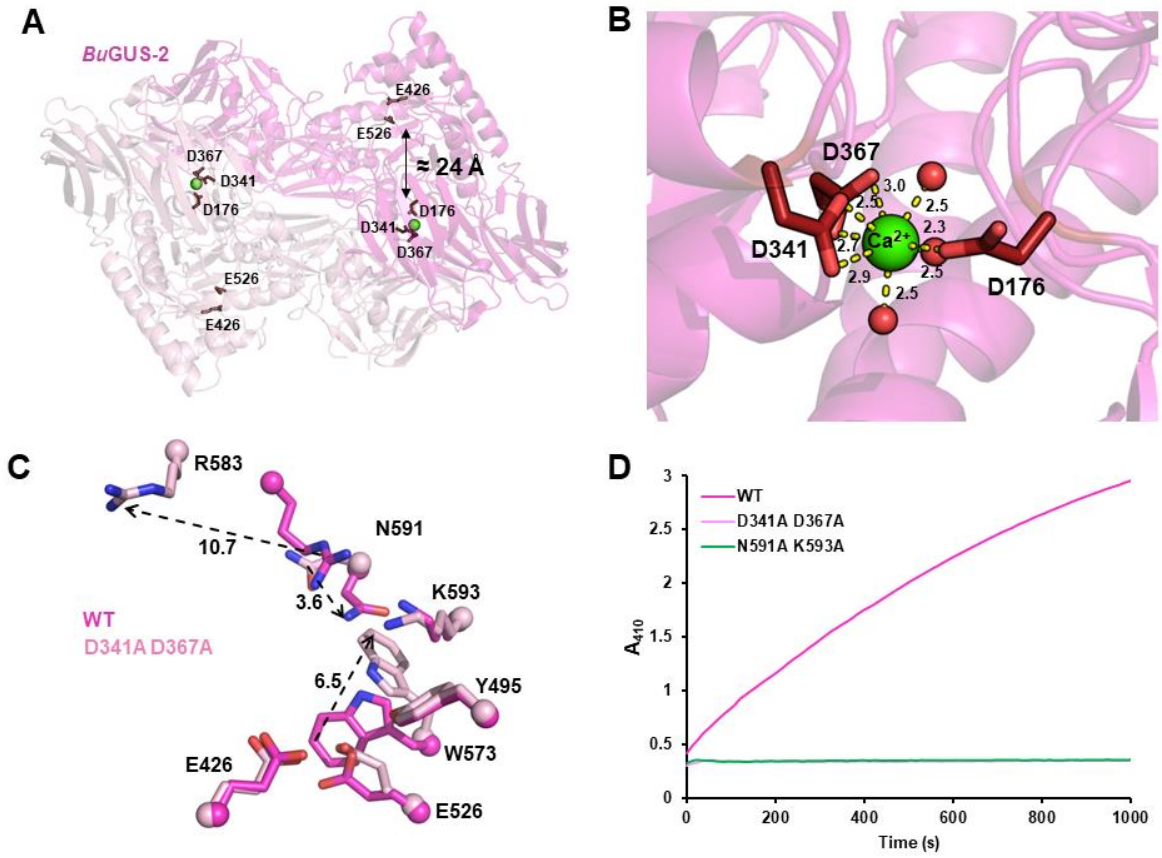


Figure 2.10 - Predicted calcium binding site key for structural and functional integrity of BuGUS-2. (**A**) *Bu*GUS-2 dimer with predicted calcium binding site (green sphere) 24 Å away from active site glutamates. (**B**) the predicted calcium ion is contacted by D176, D341, D367, and three water molecules. (**C**) active site overlay of WT and calcium-binding mutant of *Bu*GUS-2 reveals conformational changes that preclude functional activity. (**D**) progress curves of *Bu*GUS-2 activity reveal that mutation of the predicted calcium binding site results in the same loss of function as mutation of essential active site residues.

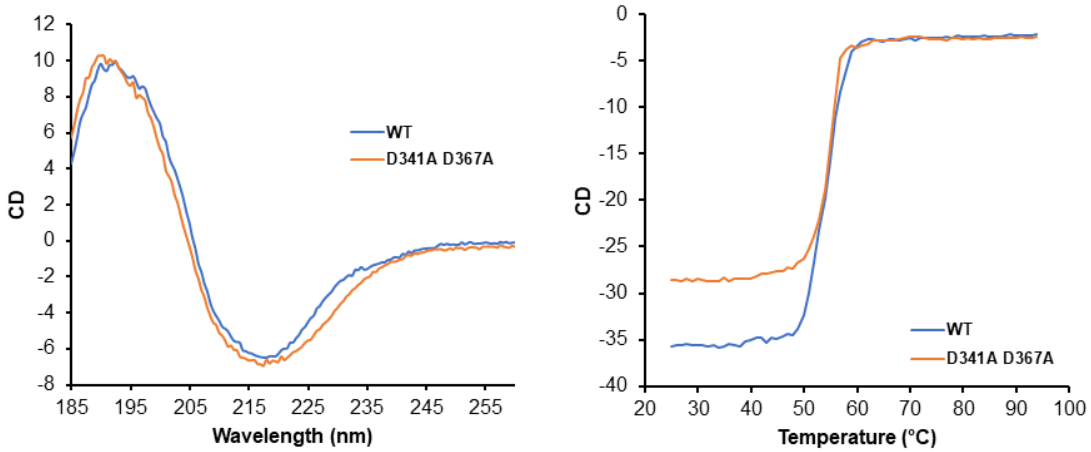


Figure 2.11 - Circular dichroism wavelength scan and melting temperature for WT and D341A D367A calcium binding mutant of BuGUS-2.

in structural order for the predicted calcium-binding mutant in comparison to the wild-type (WT), but an equivalent melting temperature indicated no significant change in overall protein stability (**Figure 2.11).** Sequence analysis of the 279 previously discovered GUS enzymes [100] revealed 17 additional GUS proteins with a predicted calcium binding site (**Table 2.3** and **Figure 2.12**). Thus, it appears that the predicted calcium binding site plays a key role in the structure and function of *Bu*GUS-2 and is conserved among other GUS proteins in the human gut microbiome.

BuGUS enzymes differentially process GlcA-containing polysaccharides

Given the distinct active site architectures of the three GUS enzymes examined here, as well as their differential processing of standard glucuronide substrates, we examined a set of pure synthetic polysaccharide substrates (**Figure 2.13A, B**). We chose heparin-like nonamers (9-mers) that contain GlcA and are either acetylated or sulfated. We also examined shorter polysaccharides (5-mers) and a substrate with GlcA at the penultimate rather than the terminal (non-reducing end) position (NAc 4-mer) (**Figure 2.13B**). Both *Bu*GUS-1 and *Bu*GUS-2 were able to process the acetylated heparin-like nonamer substrate (NAc 9-mer), but *Bu*GUS-3 showed no activity (**Figure 2.13A**). However, all three GUS enzymes, including *Bu*GUS-3, were able to process the terminal ends of a sulfated heparin-like substrate (NS 9-mer; **Figure 2.13A**). We next examined a 9-mer with a doubly sulfated glucosamine moiety at the penultimate position (NS6S 9-mer). We found, though, that this change eliminated activity with all three enzymes (**Figure 2.13A**).

We tested the effect of polysaccharide length on activity by examining shorter 5-mer substrates. Our results were similar to the 9-mer data outlined above *Bu*GUS-1 and *Bu*GUS-2

Table 2.3 - Bioinformatic analysis of the HMP GUS enzymes identified 18 GUS enzymes with calcium binding sites.

HMGC ID	Length	Kingdom	Phylum	Class	Order	Family	Genus	Species	NCBI ID
SRS011529.64867	855	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Paraprevotella		
SRS049959.38447	872	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	copri	CUO96334.1
SRS018575.34501	885	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides	clarus	WP_009122035.1
SRS050422.73127	891	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides	finegoldii	OKZ23312.1
SRS020869.18805	891	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides	intestinalis	WP_007661425.1
SRS011405.14712	861	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides	intestinalis	CCY84748.1
SRS015217.11766	926	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides	ovatus	WP_004304963.1
SRS019601.88766	840	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides	vulgatus	WP_011965330.1
SRS012902.13258	891	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides		
SRS016203.125956	891	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides		
SRS045713.76924	888	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides		
SRS048870.28850	883	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides		
SRS015190.42469	871	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides		
SRS056259.118375	871	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides		
SRS019787.27227	861	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides		
SRS017433.24596	858	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides		
SRS013476.24854	693	Bacteria	Firmicutes	Clostridia	Clostridiales	Eubacteriaceae	Eubacterium		

Figure 2.12 - Sequence alignment of putative calcium binding GUS enzymes from the Human Microbiome Project stool sample database.

A

	*Bu*GUS-1	*Bu*GUS-2	*Bu*GUS-3
NAc 9-mer	62 ± 5	68 ± 4	NA
NS 9-mer	100 ± 0	40 ± 20	50 ± 20
NS6S 9-mer	NA	NA	NA
NAc 5-mer	9 ± 4	66.8 ± 0.9	NA
NS 5-mer	96.6 ± 0.2	57 ± 4	0.7 ± 0.2
NAc 4-mer	NA	NA	NA

B

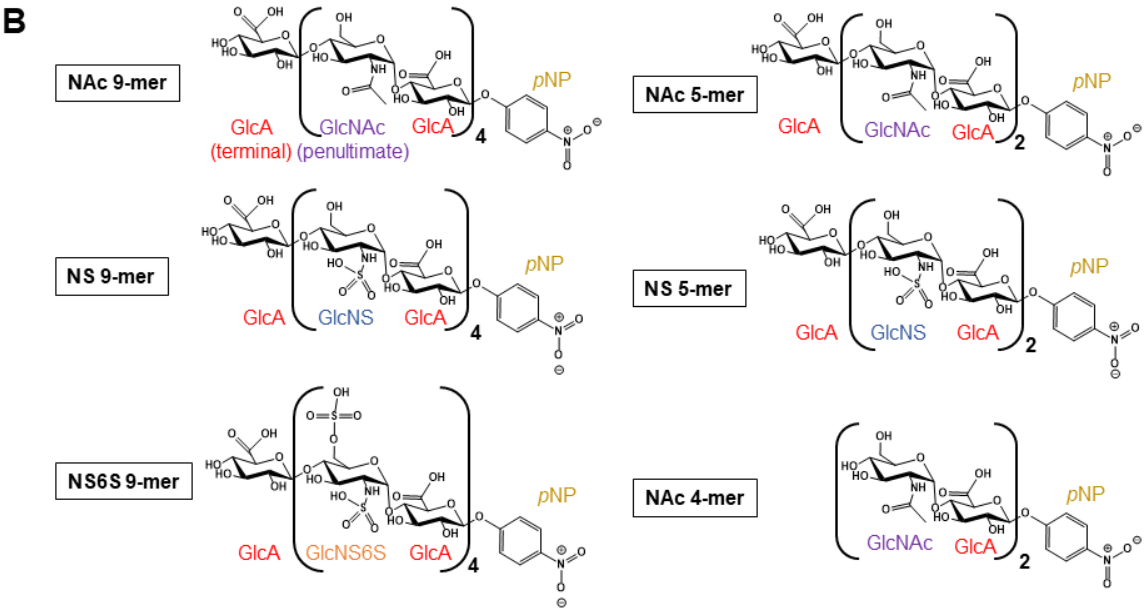


Figure 2.13 - Polysaccharide cleavage by *Bu*GUS-1, *Bu*GUS-2, and *Bu*GUS-3. (**A**) percent cleavage for *Bu*GUS-1, *Bu*GUS-2, and *Bu*GUS-3 with an acetylated and sulfated heparin-like substrates after 3 hours at pH 6.5. N=3, ± standard deviation (SD). (**B**), schematic structures of the pure synthetic polysaccharides utilized to measure polysaccharide processing by *Bu*GUS enzymes. GlcA, glucuronic acid; GlcNAc, N-acetylglucosamine; GlcNS, N-sulfoglucosamine; GlcNS6S, N-sulfoglucosamine-6-sulfate; *p*-NP, *p*-nitrophenol.

processed the NAc 5-mer, while *Bu*GUS-3 did not, and all three GUS enzymes processed the NS 5-mer (**Figure 2.13A**). Interestingly, though, *Bu*GUS-3 displayed much weaker activity with the NS 5-mer than it did with the NS 9-mer (**Figure 2.13A**). Finally, to confirm that these proteins act as exolytic enzymes toward substrates with terminal GlcA moieties, we examined a 4-mer polysaccharide with GlcA at the penultimate position (**Figure 2.13B**). As expected, the three enzymes examined failed to process this compound, indicating that they do not act as endolytic enzymes toward this particular substrate (**Figure 2.13A**). Taken together, these data using six distinct polysaccharide substrates related to compounds found in humans reveal that all three *Bu*GUS enzymes are able to process sulfated 9-mers and sulfated 5-mers, while only *Bu*GUS-1 and *Bu*GUS-2 cleaved the acetylated heparin-like 9-mer and 5-mers. Moreover, the enzyme activity appears limited to removing terminal GlcA groups. Such data provide an initial molecular framework to understand the potential for microbial GUS enzymes to utilize polysaccharide substrates within the human GI tract.

BuGUS enzymes may process additional uronic acid-containing substrates

Given the diversity of uronate-containing polysaccharides, we considered the possibility that these GUS enzymes would process uronic acid conjugates beyond glucuronides. Thus, we docked into the three *Bu*GUS enzymes the following four uronic acids: glucuronic acid (GlcA), galacturonic acid (GalA), mannuronic acid (ManA), and iduronic acid (IdoA). These sugar monosaccharides were identified from the PDB and docked manually in PyMOL based on the glucuronate-bound structure of *Bu*GUS-1 (PDB: 6D6W). Despite the differences in stereochemistry between these sugar acids, docking suggests that each may be accommodated within all three GUS active sites (**Figure 2.14**). Galacturonate appeared to be the most sterically strained sugar, which has an axial hydroxyl

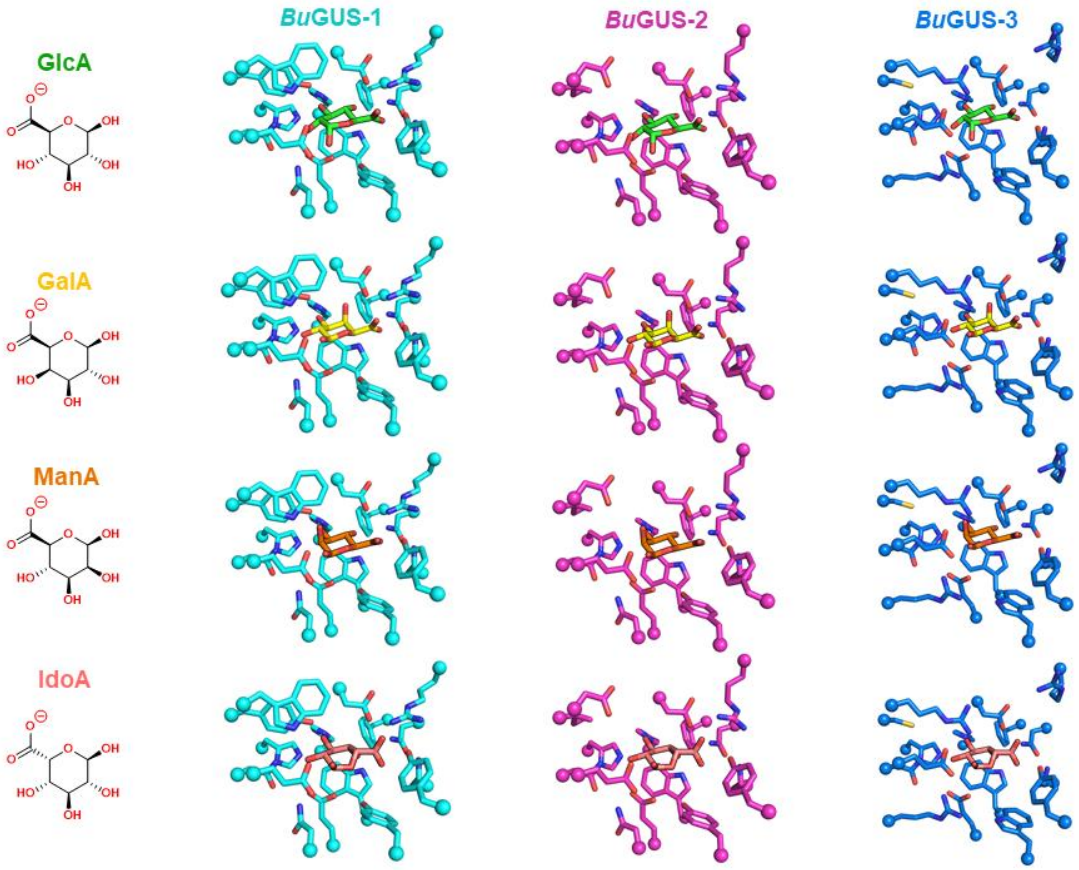
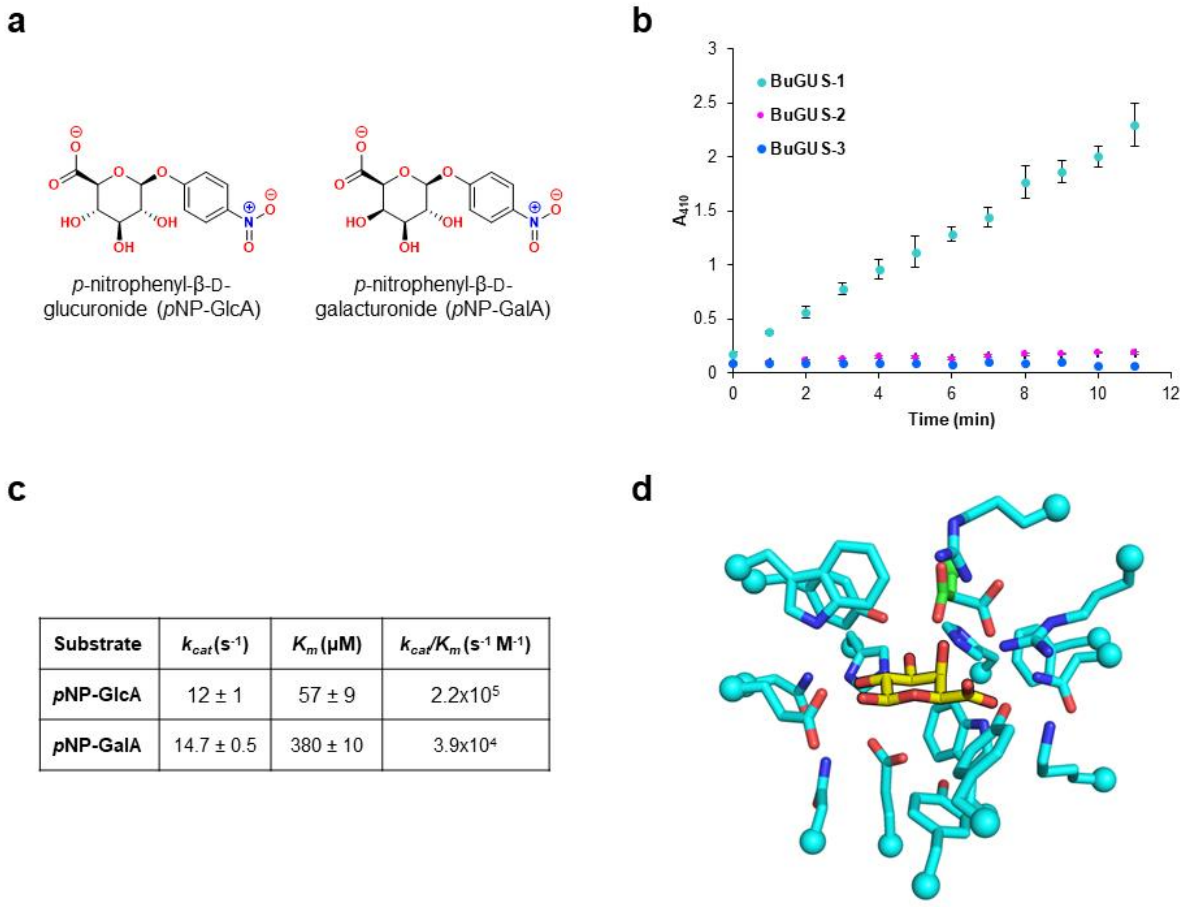


Figure 2.14 - Modeling of sugar monosaccharides of glucuronate (GlcA), galacturonate (GalA), mannuronate (ManA), and iduronate (IdoA) in the active sites of BuGUS-1, BuGUS-2, and BuGUS-3.

at the 4-position that could clash with the aspartic acid side chain conserved in all three *Bu*GUS enzymes (**Figure 2.14**). To test the hypothesis that substrates with terminal sugar acids beyond GlcA could be utilized as substrates, we examined the ability of *p*-nitrophenyl-β-D-galacturonide (*p*NP-GalA) to act as a substrate for *Bu*GUS-1, *Bu*GUS-2, and *Bu*GUS-3 (**Figure 2.15A**). We found that only *Bu*GUS-1 was able to process this galacturonide (**Figure 2.15B**). Kinetic analysis of *Bu*GUS-1 with both *p*NP-GlcA and *p*NP-GalA revealed catalytic efficiencies (k_{cat}/K_m) of $2.2x10^5$ and $3.1x10^4$, respectively, suggesting that while *Bu*GUS-1 can hydrolyze galacturonides, it does so less efficiently than the analogous glucuronide (**Figure 2.15C**). A model of galacturonic acid docked in the active site of BuGUS-1 shows that the aspartic acid (green) that could clash with the hydroxyl at 4 position may cause this weaker efficiency (**Figure 2.15C**). Taken together, docking studies and kinetics suggest that the GUS enzymes considered here may act on polysaccharide substrates containing terminal sugar acids beyond glucuronate, including mannuronate, iduronate, and galacturonate.

*Bu*GUS structures in complex with substrate analogs

To gain a better understanding of substrate recognition by these novel GUS enzymes, we incubated them with the non-hydrolyzable substrate analog phenyl-thio-β-D-glucuronide (PTG) and attempted co-crystallization. Co-crystallization of a PTG-*Bu*GUS-1 complex was successful (space group: $P12_11$, molecules in asymmetric unit: 6, **Table 2.2**), and the crystal structure revealed a conformational shift in the active site in which the catalytic acid/base E421 shifts away from the active site (**Figure 2.16A**, **B**). This conformational change is accommodated by additional shifts adjacent to the active site, in which E453 and K454 undergo 7.8 and 5.9 Å changes in position, respectively, relative to the unliganded structure



Substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (s^{-1} M^{-1})
*p*NP-GlcA	12 ± 1	57 ± 9	$2.2x10^5$
*p*NP-GalA	14.7 ± 0.5	380 ± 10	$3.9x10^4$

Figure 2.15 - *Bu*GUS-1, but not *Bu*GUS-2 or *Bu*GUS-3, can process both glucuronides and galacturonides. *a*, chemical structures of *p*NP-GlcA and *p*NP-GalA. *b*, progress curves for *Bu*GUS-1, *Bu*GUS-2, and *Bu*GUS-3 with *p*NP-GalA. *c*, kinetic parameters of *p*NP-GlcA and *p*NP-GalA hydrolysis by *Bu*GUS-1. *d*, active site of *Bu*GUS-1 with GalA manually docked and potentially important aspartic acid residue highlighted in green. This aspartic acid likely undergoes a rotamer shift, shown in cyan, to accommodate the GalA sugar.

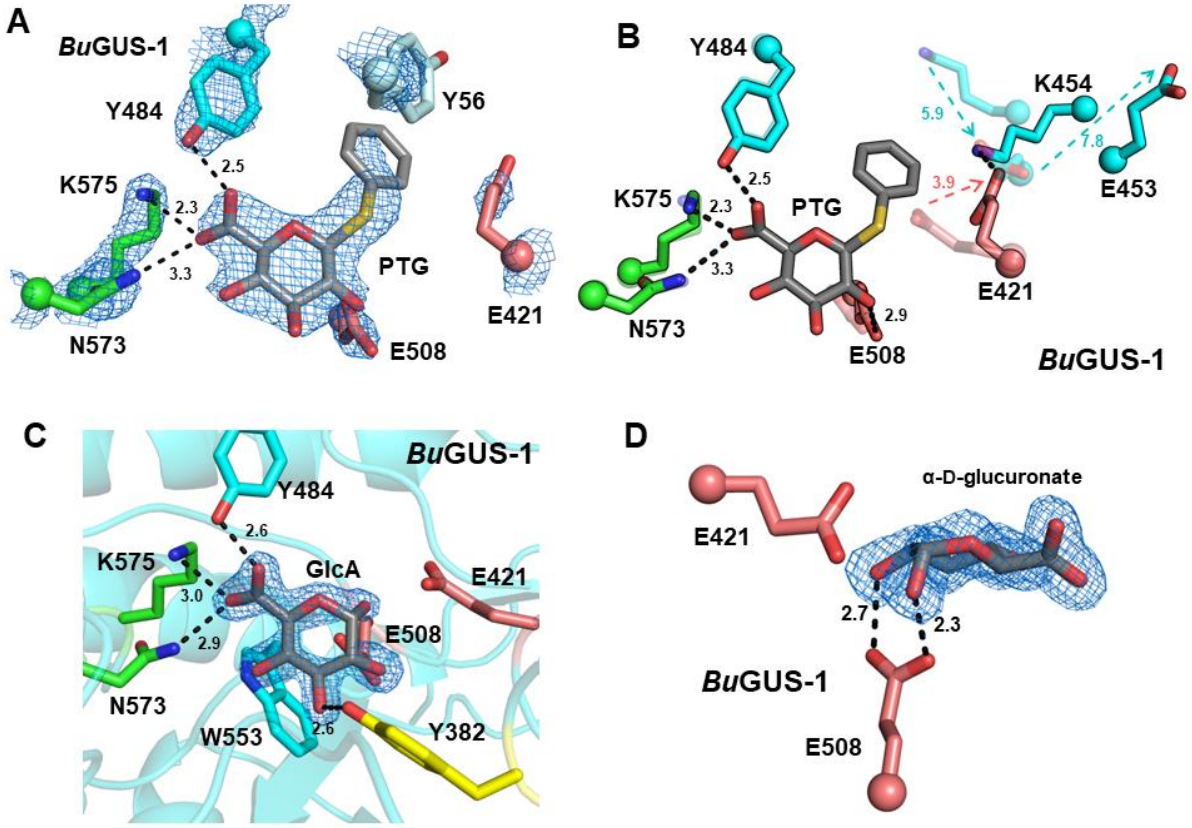


Figure 2.16 - Structural analysis of liganded BuGUS-1 and BuGUS-2 reveal chemical complementarity to GlcA. *a*, PTG bound to *Bu*GUS-1 with mFo-DFc simple omit density shown at 2.5 σ with NxK motif shown in green, catalytic glutamates in deep salmon, and Y56 from an adjacent monomer in pale cyan. *b*, overlay of PTG bound (opaque) and apo (transparent) *Bu*GUS-1 active site reveals significant conformational shifts to catalytic acid/base E421 as well as two nearby residues K454 and E453 to accommodate the large sulfur atom present in PTG. *c*, *Bu*GUS-1 bound to GlcA with mFo-DFc simple omit density shown at 2.5 σ. *d*, *Bu*GUS-1 active site with GlcA shown in the plane of the ring reveals alpha configuration that forms a hydrogen bond with the catalytic nucleophile E508.

(**Figure 2.16 B**). In line with previous studies, the carboxylate of PTG is recognized by N573 and K575 (NxK motif), as well as Y484 (**Figure 2.16A**). In addition to a PTG complex, we also determined the structure of *Bu*GUS-1 in complex with GlcA (space group: C2, molecules in asymmetric unit: 4, **Table 2.2**). GlcA was bound to *Bu*GUS-1 as the alpha anomer (**Figure 2.16C, D**), and much like PTG, the carboxylate of GlcA is recognized by the NxK motif and other residues that contact its hydroxyl groups (**Figure 2.16C**). Additionally, W533 participates in C-H-π interactions with the non-polar face of GlcA (**Figure 2.16C**). The anomeric hydroxyl group forms a hydrogen bond with E508, the catalytic nucleophile (**Figure 2.16D**). Together, these structural data highlight how GUS specifically recognizes its glucuronide substrate.

Differential SN-38-G processing by *Bu*GUS enzymes

GUS enzymes are promiscuous and can hydrolyze a variety of glucuronides related to mammalian gut toxicity [55,62,64,100,102,105–109]. Thus, we sought to determine whether these GUS enzymes are capable of reactivating the inactive metabolite SN-38-G of the cancer drug irinotecan. Despite their localization in a PUL, *Bu*GUS-1 and *Bu*GUS-2 hydrolyzed the small-molecule glucuronide SN-38-G (**Figure 2.17**). Strikingly, *Bu*GUS-1 hydrolyzed SN-38-G with an efficiency that rivals previously characterized Loop 1 GUS enzymes that are not located in PULs and have been shown to prefer only small molecule glucuronides over polysaccharides (**Figure 2.17B**) [100]. We hypothesized that the NTL identified in the structure of *Bu*GUS-1 may play a key role in recognizing the aglycone moiety of SN-38-G (**Figure 2.17A**). The NTL is defined as residues Y54 through A67 and forms a loop that sits by the active site of an adjacent protomer (**Figure 2.17C**). Indeed, the NTL loop deletion (Δloop *Bu*GUS-1), displayed much slower processing with both 4-MUG and SN-38-G in

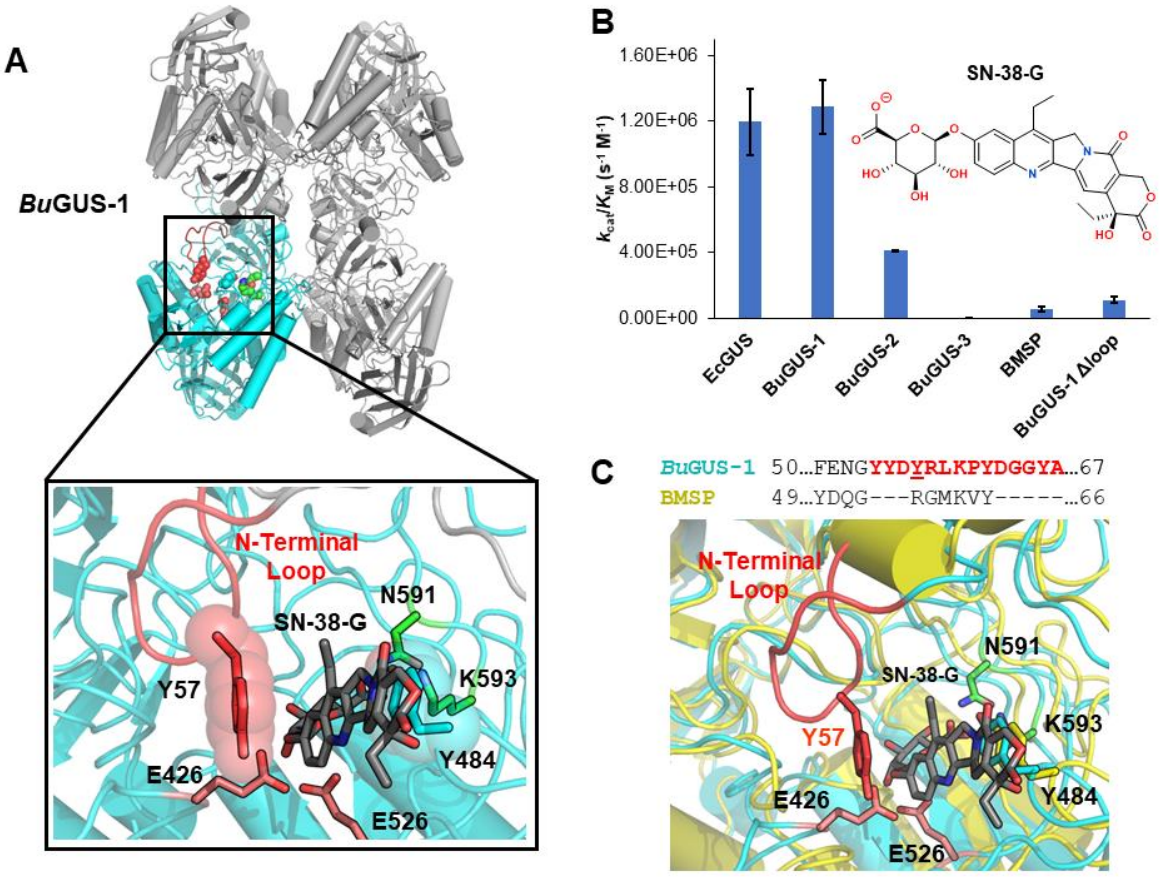


Figure 2.17 - Kinetic and structural analysis of SN-38-G hydrolysis reveals importance of N-terminal loop in *Bu*GUS-1. (A), *Bu*GUS-1 tetramer with adjacent N-Terminal Loop highlighted in red, catalytic glutamates in deep salmon, and NxK motif in green. Zoom-in of active site with SN-38-G manually docked in the active site of *Bu*GUS-1 based on PTG-bound structure. (**B**), catalytic efficiencies k_{cat}/K_M for *Ec*GUS, *Bu*GUS-1, *Bu*GUS-2, *Bu*GUS-3, BMSP, and *Bu*GUS-1 Δloop with the substrate SN-38-G. (**C**), sequence alignment of *Bu*GUS-1 and BMSP GUS N-Terminal Loop regions and overlay of *Bu*GUS-1 and BMSP active sites with SN-38-G manually docked. Error bars represent SD of n = 3 biological replicates.

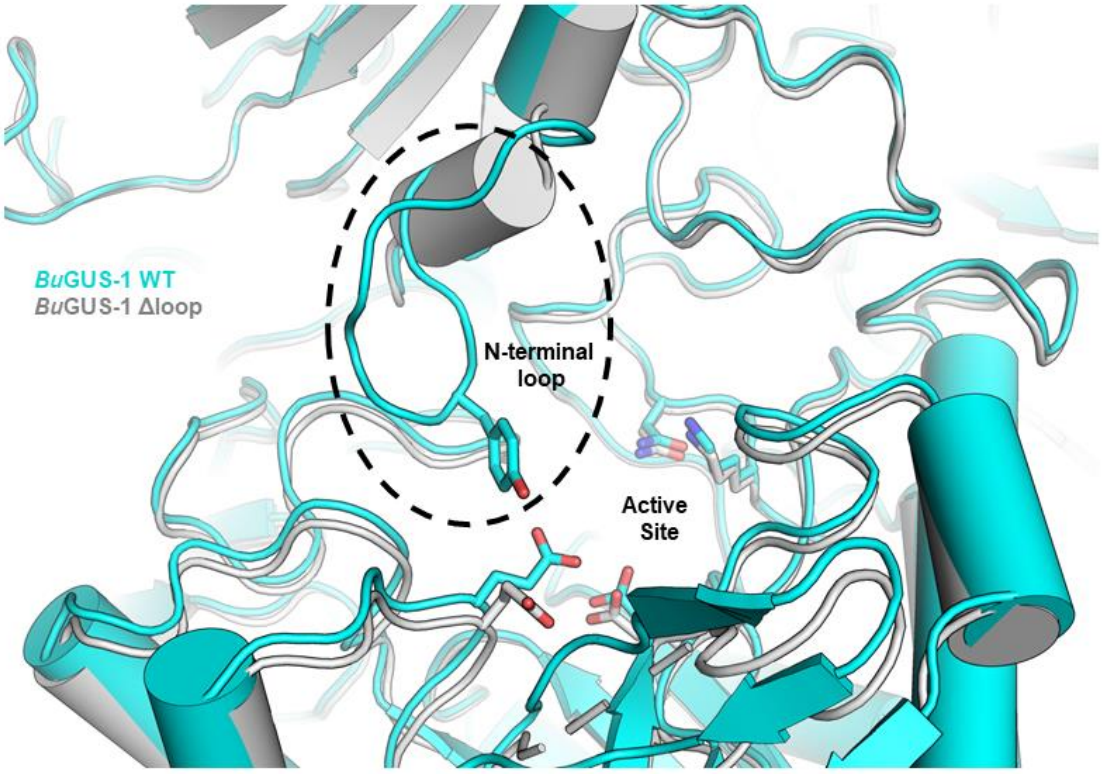


Figure 2.18 - Structural overlay of BuGUS-1 WT (cyan) and BuGUS-1 Δloop (grey) reveals absence of N-terminal loop necessary for efficient processing of SN-38-G.

comparison to the WT *Bu*GUS-1 (**Table 2.1** and **Figure 2.17B**). We solved the structure of Δloop *Bu*GUS-1 (space group: $P12_11$, molecules in asymmetric unit: 4, **Table 2.2**) which shows the absence of this key loop structure (**Figure 2.18**).

As an additional control to test the importance of the NTL for SN-38-G processing by *Bu*GUS-1, we cloned, expressed, and purified a Bacteroides multispecies (BMSP) GUS that is similar to *Bu*GUS-1 but, importantly, lacks the NTL sequence necessary for efficient processing of small molecule glucuronides (**Figure 2.17C**). The 2.65 Å structure of BMSP GUS (space group: $I4_1$, molecules in asymmetric unit: 4, **Table 2.2**) reveals the same tetrameric organization as *Bu*GUS-1 but lacks the N-terminal loop that forms the aglycone binding site of *Bu*GUS-1 (**Figure 2.17C** and **S11*a***). Importantly, BMSP displayed similar 4-MUG and SN-38-G processing efficiencies in comparison to the Δloop variant of *Bu*GUS-1 (**Fig. 8*b*** and **Table 1**). These data suggest that an N-terminal sequence feature in the previously identified No Loop *Bu*GUS-1 allows it to process SN-38-G with activity similar to Loop 1 GUS enzymes.

D-glucaro-1,4-lactone inhibits *Bu*GUS-1 and *Bu*GUS-2

Due to their ability to process SN-38-G, we tested whether we could inhibit *Bu*GUS-1 and *Bu*GUS-2 with our selective bacterial GUS inhibitors as well as the non-specific GUS inhibitor D-glucaro-1,4-lactone (**Figure 2.19C** and **2.22**). The GUS-specific inhibitors Inh1 and UNC10201652 did not inhibit either *Bu*GUS-1 or *Bu*GUS-2 up to 100 μM. However, D-glucaro-1,4-lactone displayed mid micromolar potency against both *Bu*GUS-1 and *Bu*GUS-2 (**Table 2.4**). Together, these data show that previously characterized selective inhibitors of GUS are not effective against *Bu*GUS-1 and *Bu*GUS-2, while the non-specific GUS inhibitor D-glucaro-1,4-lactone is a mid-micromolar inhibitor of *Bu*GUS-1 and *Bu*GUS-2.

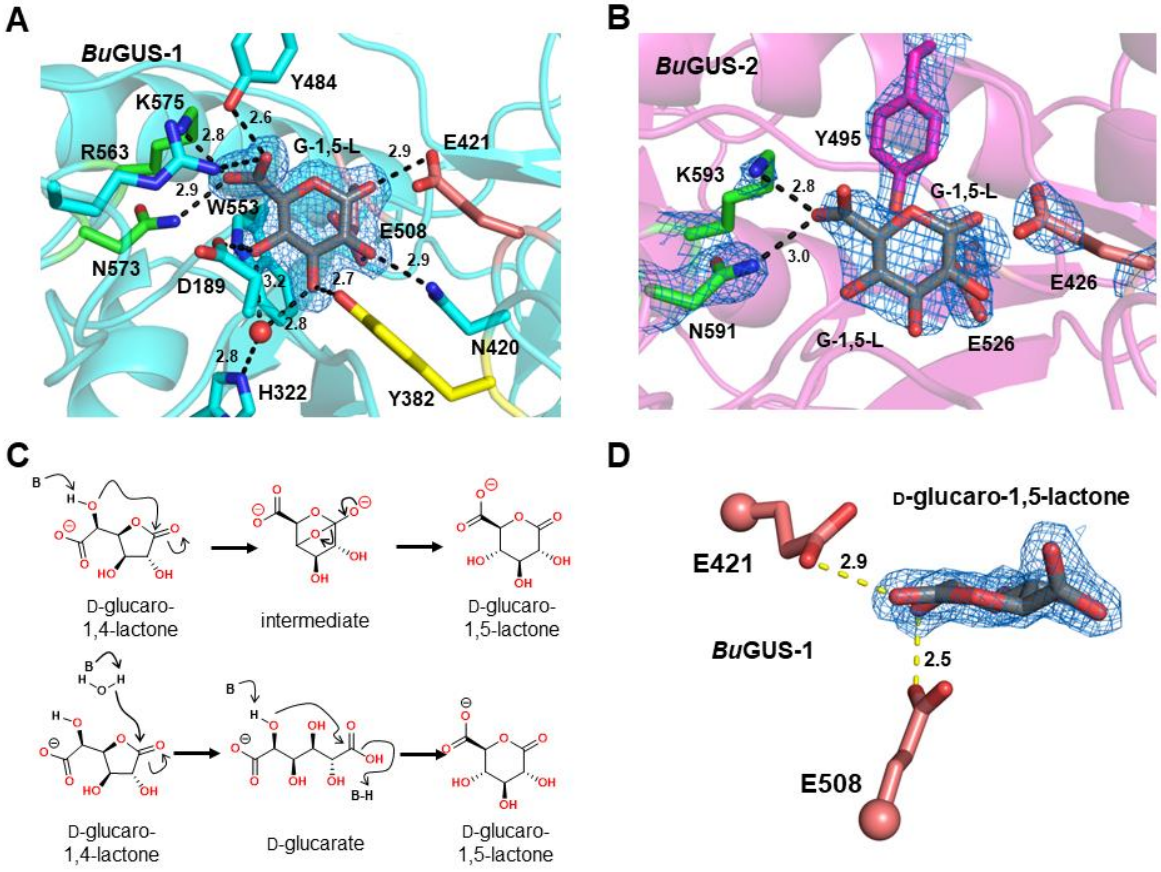


Figure 2.19 - Structural analysis of *Bu*GUS-1 and *Bu*GUS-2 inhibition by D-glucaro-1,4-lactone reveals D-glucaro-1,5-lactone bound instead. *a*, *Bu*GUS-1 bound to D-glucaro-1,5-lactone with mFo-DFc simple omit density shown at 2.5 σ. *b*, *Bu*GUS-2 bound to D-glucaro-1,5-lactone with mFo-DFc simple omit density shown at 1.5 σ. *c*, proposed mechanisms for the conversion of D-glucaro-1,4-lactone to D-glucaro-1,5-lactone. *d*, close-up view in the plane of D-glucaro-1,5-lactone reveals planarity at alpha carbon and key contact with catalytic acid/base E421.

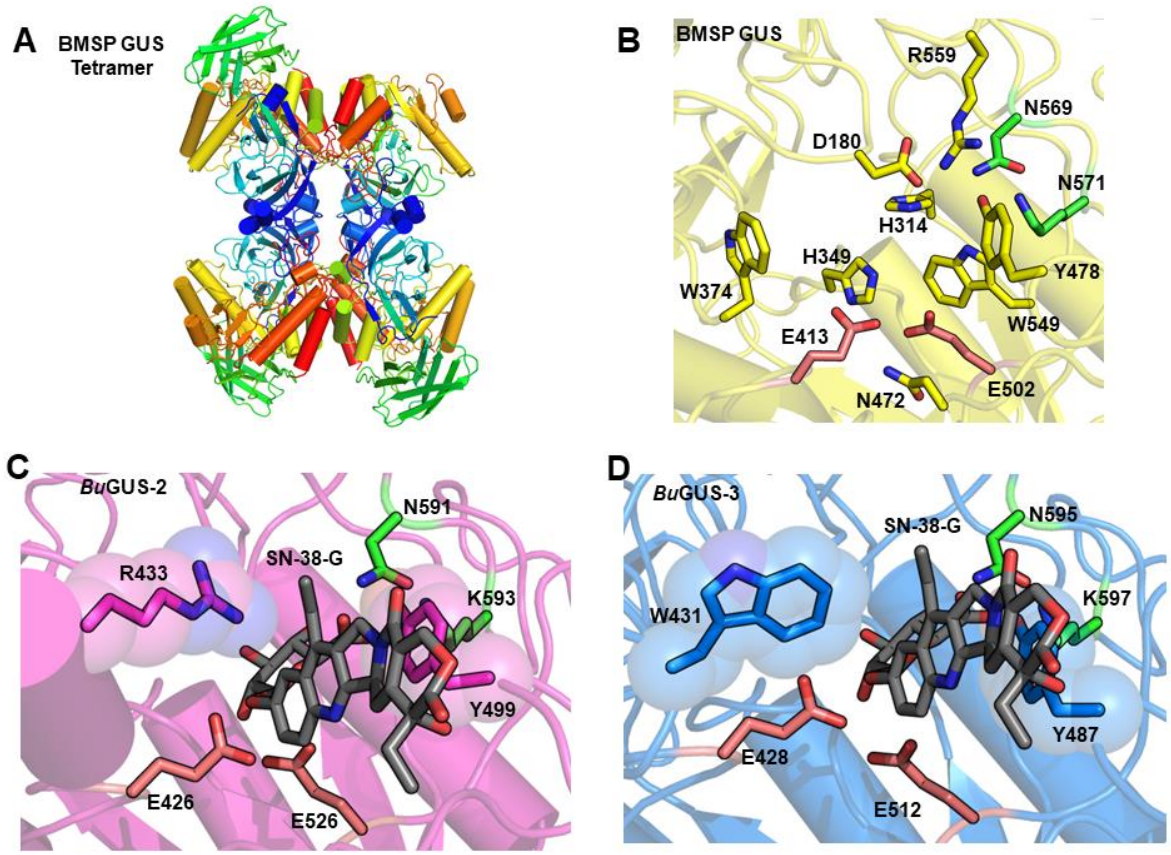


Figure 2.20 - Structural analysis of BMSP GUS and docking studies of SN-38-G in the active sites of *Bu*GUS-2 and *Bu*GUS-3. *a*, tetrameric biological assembly of BMSP GUS. *b*, active site of BMSP GUS with NxK motif in green and catalytic glutamates in deep salmon. Manually docked SN-38-G in active site of *c*, *Bu*GUS-2 and *d*, *Bu*GUS-3 based on the PTG-bound structure of *Bu*GUS-1.

Table 2.4 - IC_{50} values of Inh1, Inh9, and D-glucaro-1,4-lactone for BuGUS-1, BuGUS-2, BuGUS-3, and BMSP GUS

Protein	Inh1	UNC10201652	D-glucaro-1,4-lactone
***Bu*GUS-1**	>100	>100	8.3 ± 0.6
***Bu*GUS-2**	>100	>100	9 ± 2
***Bu*GUS-3**	NA	NA	NA
BMSP	>100	>100	27 ± 1

Values shown are in units of μM and are averages of n = 3 biological replicates ± SD. NA = no activity

Next, we co-crystalized *Bu*GUS-1 and *Bu*GUS-2 in the presence of D-glucaro-1,4-lactone. Successful crystals were grown for both *Bu*GUS-1 and *Bu*GUS-2 upon incubation with D-glucaro-1,4-lactone, and structures at 2.0 Å (space group: C2, molecules in asymmetric unit: 2) and 2.5 Å (space group: $P2_12_12_1$, molecules in asymmetric unit: 2) were determined, respectively (**Table 2.2**). Analysis of the active site of these structures revealed D-glucaro-1,5-lactone bound instead of D-glucaro-1,4-lactone (**Figure 2.19A, B, C**). Previous studies have shown the spontaneous conversion of D-glucaro-1,5-lactone to D-glucaro-1,4-lactone, as it is the thermodynamic product [110]. The reverse process is also chemically possible (**Figure 2.19C**). As expected, the carboxylate of D-glucaro-1,5-lactone interacts with the NxK motif of both *Bu*GUS-1 and *Bu*GUS-2 (**Figure 2.19A, B**). The remaining hydroxyl groups are recognized by other residues, including the catalytic acid/base and nucleophile glutamates (**Figure 2.19A, B, D**).

DISCUSSION

The GHs encoded by *Bacteroides* species play key roles in the processing of carbohydrates and glycosides that reach the GI tract. Here we present three unique GUS enzymes from the human gut microbe *Bacteroides uniformis* that advance our understanding of the structural and functional diversity within this GH family. By analyzing the genes adjacent to a previously characterized *B. uniformis* GUS [100], we discovered two additional GUS enzymes from a *B. uniformis* PUL (**Figure 2.1A, B**). One of these GHs we termed *Bu*GUS-1, as it retained the GUS-specific features previously used to identify GUS enzymes in the HMP [100]. We demonstrated that it is a GUS capable of processing a variety of GlcA-containing substrates (**Table 2.1, Figure 2.13, and Figure 2.3**). *Bu*GUS-3 also possessed several GUS-specific features, including the core fold, catalytic residues, and NxK motif;

however, a tryptophan replaced the tyrosine that hydrogen bonds to and structurally stabilizes the nucleophilic glutamate (**Figure 2.4C**). While *Bu*GUS-3 was unable to process 4-MUG, it did exhibit GUS activity toward the heparin sulfate 9-mer (**Figure 2.13A**)**.** This discovery suggests that the initial GUS rubric defined previously could allow either a tyrosine or a tryptophan at this sequence position[100]. Indeed, a tryptophan is present at this position in the GUS module of BT0996, one of the enzymes responsible for the degradation of rhamnogalacturonan-II in *B. thetaiotaomicron* [111]. This information indicates that the 279 GUS proteins previously identified represent an initial GUS atlas and should be reexamined and updated as new structural and functional data are determined regarding this enzyme family. Indeed, a preliminary analysis of the HMP identified 10 additional GUS proteins with a tryptophan residue in this position (**Table 2.5**); these novel proteins will be the subject of future studies.

As previously discussed by Pollet *et al*., GUS enzymes with longer loops adjacent to the active site (e.g. Loop 1 GUS enzymes) were shown to process small glucuronides, and those possessing open active sites were able to process larger GlcA-containing polysaccharides [100]. In previously determined GUS structures, the tetrameric interface between GUS protomers have been formed by their C-termini, and active site adjacent loop structures (Loop 1) from these adjacent protomers formed the aglycone binding site (**Fig. 3*a***), limiting the access of larger substrates [55,100,102] In contrast, *Bu*GUS-2 was shown to form a dimer, leaving its active site open and solvent exposed to accommodate larger polysaccharides [100] (**Figure 2.4B** and **2.7A**). *Bu*GUS-1, which exhibits an open active site via a unique N-to-C-term-mediated tetrameric arrangement (**Figure 2.6B**), processed 4-MUG with a higher efficiency than *Bu*GUS-2 and *Bu*GUS-3, which are both dimers and contain

Table 2.5 - Bioinformatic analysis of the HMP identified 11 GUS enzymes containing the variable active site tryptophan observed in *Bu*GUS-3.

HMGC ID	Length	Kingdom	Phylum	Class	Order	Family	Genus	Species	NCBI Annotation	NCBI ID
SRS052697.50102	1427	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides	eggerthii	GH2 protein	WP_017140867.1
SRS022137.24481	689	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides	MSP	GH2 protein	WP_007840523.1
SRS055982.32560	875	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides	salyersiae	GUS/b-gal	CUM88360.1
SRS024388.29889	796	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides	sartorii	GH2 protein	WP_016277419.1
SRS011405.4923	1427	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides	thetaiotaomicron	b-gal	WP_081030062.1
SRS014613.5859	1118	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides	thetaiotaomicron	GUS/b-gal	WP_080973891.1
SRS018133.31535	814	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides	uniformis	GH2 protein	WP_044467710.1
SRS017433.149978	1429	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides		b-gal	
SRS011586.60073	826	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Porphyromonadaceae	Parabacteroides	merdae	GH2 protein	WP_005639106.1
SRS014923.34218	642	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Faecalibacterium	CAG:74	b-gal	OLA21514.1
SRS023914.87176	647	Bacteria	Firmicutes	Clostridia	Clostridiales					

more solvent exposed active sites (**Table 2.1** and **Figure 2.7A** and **B**).

*Bu*GUS-1 was also shown to process SN-38-G faster than both *Bu*GUS-2 and *Bu*GUS-3 (**Fig. 8*b***). Surprisingly, the NL *Bu*GUS-1 processed SN-38-G at an efficiency that rivaled that of L1 *Ec*GUS, despite the lack of an active site loop at the canonical position in its amino acid sequence (**Figure 2.1C**). Further examination of the crystal structure of *Bu*GUS-1 revealed the presence of an N-terminal loop (NTL) donated from an adjacent protomer (**Figure 2.17A**). This donated loop mimics the Loop 1 present in L1 GUS enzymes and appears to enable *Bu*GUS-1 to process the small molecule glucuronide 4-MUG (**Table 1**) and SN-38-G at efficiencies similar to those of characterized Loop 1 GUS enzymes (**Figure 2.17B)** [55,102]. Indeed, kinetic analysis of BMSP GUS and the Δloop variant of *Bu*GUS-1 suggest that SN-38-G and 4-MUG processing by *Bu*GUS-1 is greatly facilitated by its N-terminal loop (**Table 2.1** and **Figure 2.17B**). Examination of the GUS proteins present in the GI tracts of healthy humans [100] revealed that six additional enzymes beyond *Bu*GUS-1 maintain an NTL (**Table 2.6**). Collectively, these data indicate that the determination of novel crystal structures of GUS enzymes will continue to enhance our understanding of the structural and functional variations present in this family of proteins.

To further investigate how the *Bu*GUS-1 active site may interact with SN-38-G, we manually docked SN-38-G based on the PTG-bound structure of *Bu*GUS-1. Our analysis shows that the planar, non-polar aglycone of SN-38-G could interact favorably with the *Bu*GUS-1 active site (**Figure 2.17A**). Notably, Y57 located in the donated loop participates in π-π interactions with the aromatic scaffold of SN-38-G in the binding mode modeled (**Figure 2.17A**). This may explain its ability to efficiently hydrolyze this substrate. Docking of SN-38-G into the active sites of *Bu*GUS-2 and *Bu*GUS-3 demonstrates that they do not

Table 2.6 - Bioinformatics analysis of the HMP identified 6 GUS enzymes containing an N-terminal loop like that observed in BuGUS-1.

HMGC ID	Length	Kingdom	Phylum	Class	Order	Family	Genus	Species	NCBI Annotation	NCBI ID
SRS020869.249001	613	Bacteria	Bacteroidetes							
SRS015190.3291	603	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Bacteroidaceae	Bacteroides	uniformis	GUS/b-gal	CUO58783
SRS017307.68457	600	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Rikenellaceae	Alistipes	senagalensis	GUS	WP_019151945
SRS015264.145530	600	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Barnesiellaceae	Coprobacter	secundus	GUS	WP_021929483
SRS049995.49420	609	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidiales	Rikenellaceae	Alistipes	senagalensis	GUS	WP_019149660
SRS017701.130646	318	Bacteria	Bacteroidetes							

harbor the same active site features of *Bu*GUS-1 that would allow them to recognize SN-38-G (**Figure 2.15)**. Specifically, the tyrosine in *Bu*GUS-1 is replaced by an arginine and a tryptophan in *Bu*GUS-2 and *Bu*GUS-3, respectively, which do not appear to favorably interact with the aromatic scaffold of SN-38-G (**Figure 2.15C, D**).

In addition to small glucuronides, we demonstrated that all three GUS enzymes differentially processed GlcA-containing polysaccharides. While bioinformatic analysis of the genes in this PUL do not reveal a clear polysaccharide substrate for these enzymes to act on, we showed that *Bu*GUS-2 was capable of processing a sulfated heparin-like 9-mer and an acetylated heparin-like 9-mer, and *Bu*GUS-3 processed the sulfated heparin-like 9-mer (**Figure 2.13A**). Given the unique nature of the *Bu*GUS-3 active site (**Figure 2.4C**) in comparison to previously characterized GUS enzymes, this likely is key feature that leads to its lack of activity with most of the glucuronide-containing polysaccharides. The differences in polysaccharide processing may also be explained by differences in quaternary structures. Although both *Bu*GUS-2 and *Bu*GUS-3 are dimers and contain extra C-terminal domains, the positioning of these domains is distinct and influence protomer organization (**Figure 2.7A, B**). Taken together, a combination of unique active site residues and quaternary structures likely dictate the specific substrates of these GHs.

Interestingly, *Bu*GUS-1 was also shown to process both GlcA-containing polysaccharides tested (**Figure 2.13A**). Compared to traditional L1 GUS enzymes, the active site of *Bu*GUS-1 is more open due to its N-term-mediated tetrameric interface (**Figure 2.6B**), which allows larger polysaccharides to access the active site. In addition to its unique tetrameric state, the flexible nature of its active site, as evidenced by the PTG-bound structure, may also explain the ability of *Bu*GUS-1 to process bulkier polysaccharides. Upon

PTG binding, several conformational shifts occur, including that exhibited by the catalytic base E421, which appears to conflict with the large sulfur atom of PTG (**Figure 2.16A, B**). Although this conformation would preclude function, as E421 is far from the position it would need to be to serve as an acid/base, the structure demonstrates that there is enough mobility in the active site to accommodate this shift and suggests that the active site is also capable of accommodating larger polysaccharide substrates.

We further found via docking that other sugar acids, like galacturonate, mannuronate and iduronate, are likely to be accommodated in the active sites of these GUS enzymes (**Figure 2.14**), and we confirm that *Bu*GUS-1 can utilize a small-molecule galacturonide as a substrate (**Figure 2.15**). This finding expands our understanding of the substrate-utilization capacities of the gut microbial GUS enzymes, and suggests that these enzymes may coordinate the degradation of polysaccharides that contain uronic acids beyond glucuronate.

Given the importance of quaternary structure relative to GUS function, we were interested in whether computational approaches would provide this critical information. We used Rosetta modeling to predict the tertiary and quaternary structures of the three GUS enzymes reported here. Although the core fold was predicted with a high degree of accuracy for all GUS enzymes analyzed, the critical loop structures as well as the orientation of C-terminal domains were more difficult to position and were heavily influenced by extant structures (**Figure 2.21**). These results highlight the importance of using experimental structures to further refine modeling approaches to accurately predict protein quaternary structures.

Upon determining that *Bu*GUS-1 and *Bu*GUS-2 are targets to prevent GI-side effects via SN-38-G processing, we tested whether they are susceptible to inhibition. Our GUS-

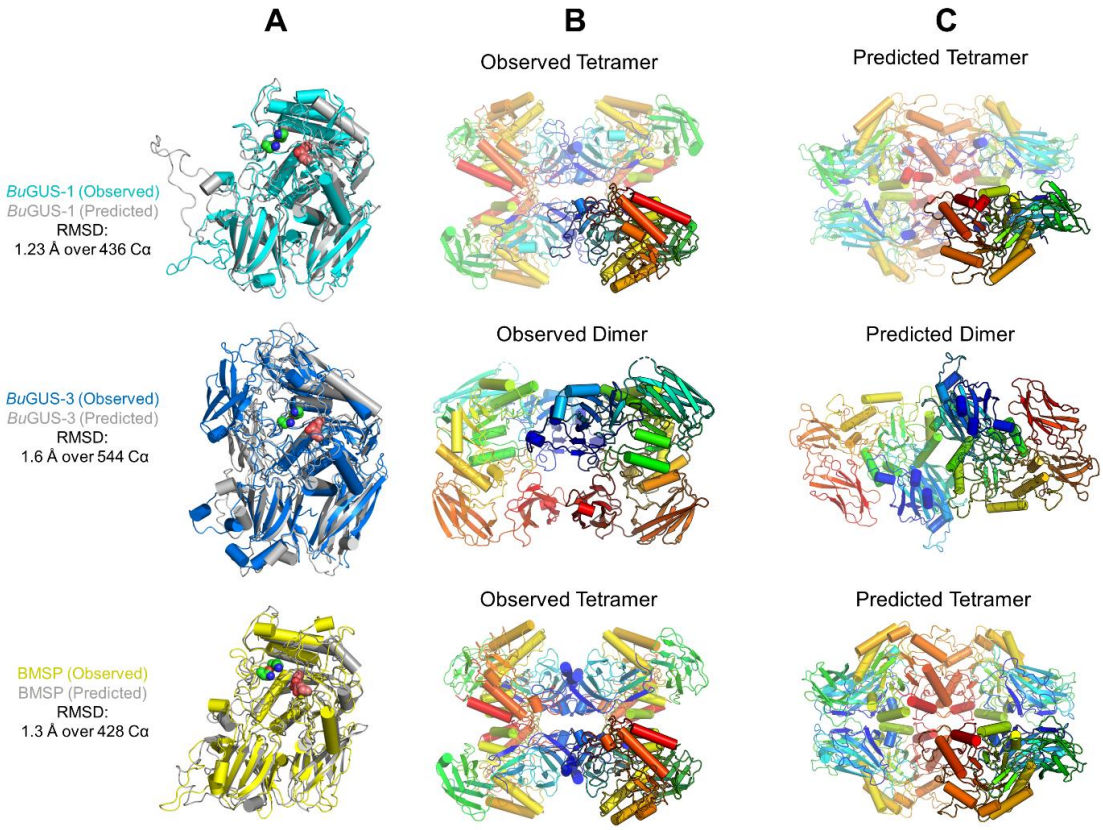


Figure 2.21 - Predicted structures by Rosetta-based modeling of unique tertiary and quaternary structures discovered by crystallography. (A) Overlay of observed and Rosetta-predicted tertiary structure for *Bu*GUS-1, *Bu*GUS-3, and BMSP GUS. (**B**) Observed quaternary structures for *Bu*GUS-1, *Bu*GUS-3, and BMSP. (**C**) Rosetta prediction of quaternary structure for *Bu*GUS-1, *Bu*GUS-3, and BMSP.

Figure 2.22 - Chemical structures of Inhibitor 1 and UNC10201652.

specific inhibitors Inh1 and UNC10201652 did not inhibit *Bu*GUS-1 and *Bu*GUS-3 (**Table 2.4** and **Figure 2.22**). Previously, we have shown that the loop in L1 GUS enzymes stabilizes Inh1 [55]. Although *Bu*GUS-1 contains an N-terminal loop that replaces L1 in the active site, it is distinct from classic Loop 1 GUS enzymes which form deep hydrophobic pockets constructed from two loops from adjacent monomers (**Figure 2.6A, B**) [55]. Thus, the active site in *Bu*GUS-1 is more hydrophilic and solvent accessible, making it unfavorable for binding to the hydrophobic scaffold of Inh1 and UNC10201652.

Although Inh1 and UNC10201652 did not inhibit GUS activity, we showed that the non-specific GUS inhibitor D-glucaro-1,4-lactone did inhibit *Bu*GUS-1 and *Bu*GUS-2 (**Table 2.4**). The crystal structure of *Bu*GUS-1 incubated with the inhibitor revealed D-glucaro-1,5-lactone bound instead of D-glucaro-1,4-lactone (**Figure 2.19A, B**). We hypothesize that D-glucaro-1,5-lactone is spontaneously generated in solution over the time scale of crystal formation, upon which it is stabilized by binding to the GUS active site. Previous studies indicate that hydrolases in general, and GUS specifically, binds more tightly to D-glucaro-1,5-lactone than D-glucaro-1,4-lactone [112] and may explain its presence in the active site. The same result was observed for *Bu*GUS-2, with D-glucaro-1,5-lactone apparent in the active site instead of the administered D-glucaro-1,4-lactone (PDB: 6D5O). Importantly, this pan-GUS inhibitor exhibited mid-micromolar potency against *Bu*GUS-1 and *Bu*GUS-2. These data suggest that other inhibitor chemotypes could be employed to prevent the actions of non-Loop 1 GUS enzyme-mediated reactivation of SN-38-G in the intestinal lumen.

The presence of three structurally and functionally unique GUS enzymes within a single *B. uniformis* PUL suggests that they have evolved to cleave distinct bonds in a uronic acid-rich polysaccharide. However, the action of GUS enzymes is not sufficient to carry out

the complete catabolism of a complex polysaccharide. Thus, it is likely that these GUS enzymes act in concert with the GH88 enzyme, mannonate oxidase (MO), mannonate dehydratase (MD), and polysaccharide lyase (PL) enzymes found in the same PUL to deconstruct a complex uronate-containing glycan found in the human gut. Additionally, the hallmark SusC/SusD proteins likely mediate the transportation of the polysaccharide into the periplasmic space of *B. uniformis* for subsequent catabolism. Further studies are needed to determine the true polysaccharide associated with this PUL, but the data presented provides a basis for understanding the roles these GUS enzymes play in polysaccharide processing as well as their more established roles in drug-glucuronide reactivation.

EXPERIMENTAL PROCEDURES

Enzyme Cloning

The full-length *Bu*GUS-1, *Bu*GUS-3, and BMSP genes were purchased from BioBasic in the pUC57 vector. Protein sequences were analyzed for signal peptide cleavage sites using the online SignalP 4.1 server [103]. The mature gene lacking the signal peptide was amplified and inserted into the pLIC-His vector using the primers in **Table 2.7**.

Site-Directed Mutagenesis

The *Bu*GUS-1 NTL deletion, *Bu*GUS-2 D341A/D367A (*Bu*GUS-2 ΔCa^{2+}), and *Bu*GUS-2 N591A/K593A mutants were created using site-directed mutagenesis. Primers were synthesized by Integrated DNA Technologies and are shown in **Table S5**. The mutant plasmids were sequenced to confirm the mutations. The mutants were produced and purified using *E. coli* BL21 (DE3) Gold as described above. The NTL deletion for *Bu*GUS-1 encompassed residues Y54-A67, which were replaced by a 6-residue linker (RGMKVY) based on the structure of BMSP GUS to maintain protein stability.

Table 2.7 - Primer sequences used in this study.

Primer Name	Sequence
*Bu*GUS-1 Fwd	TACTTCCAATCCAATGCGTTCGCACAGCAGGCACCGCAGATTATGAAC
*Bu*GUS-1 Rev	TTATCCACTTCCAATGCGCTATTTATAGGCTTCAGTCAGTTCTTTATACCATTTCTG
*Bu*GUS-3 Fwd	TACTTCCAATCCAATGCGCTGATCCCGACCGATCGTAAACAGCGTC
*Bu*GUS-3 Rev	TTATCCACTTCCAATGCGCTACTGGCCGCTCACAGAGATCAGCGC
BMSP Fwd	TACTTCCAATCCAATGCGCAGAACCTGATTACCAATGTTTACGGTCGTGACATCC
BMSP Rev	TTATCCACTTCCAATGCGCTATTCACCGAATTTGGTTTTGTAGTACTCACGCATCAGGTA
*Bu*GUS-1 ΔN-termloop Fwd	GTTGATCCGTTCGAAAACGGTCGTGGCATGAAAGTTTACCAGGATAAAACCTATAGTGAT
*Bu*GUS-1 ΔN-termloop Rev	ATCACTATAGGTTTTATCCTGGTAAACTTTCATGCCACGACCGTTTTCGAACGGATCAAC
*Bu*GUS-2 D341A Fwd	GTCCGGCTCTGACGGCTGAAATGCATCGTCG
*Bu*GUS-2 D341A Rev	CGACGATGCATTTCAGCCGTCAGAGCCGGAC
*Bu*GUS-2 D367A Fwd	CACTACCCGCAGGCTGACGCTCTGCTG
*Bu*GUS-2 D367A Rev	CAGCAGAGCGTCAGCCTGCGGGTAGTG
BuGUS-2 N591A/K593A Fwd	CTGCGTGATGAAAGCATGCCGCGTATTGCCAACGCAGGCCTGGTTTAC
BuGUS-2 N591A/K593A Rev	CGTAAACCAGGCCTGCGTTGGCAATACGCGGCATGCTTTCATCACGCAG

Protein Expression and Purification

Each β-glucuronidase expression plasmid was transformed into BL21 DE3 Gold cells for enzyme expression. Cells were grown in the presence of ampicillin in LB medium with shaking at 225 rpm at 37 °C to an OD_{600} of 0.5, at which point the temperature was reduced to 18 °C. At OD_{600} of 0.8, protein expression was induced by the addition of 0.1 mM isopropyl-1-thio-D-galactopyranoside (IPTG) and incubation continued overnight. Cells were collected by centrifugation at 4500xg for 20 min at 4 °C in a Sorvall (model RC-3B) swinging bucket centrifuge. Cell pellets were resuspended in Buffer A (20 mM Potassium Phosphate pH 7.4, 50 mM imidazole, 500 mM NaCl), DNase, lysozyme, and a Roche complete-EDTA free protease inhibitor tablet. Resuspended cells were sonicated and clarified via centrifugation at 17,000xg for 60 min in a Sorvall (model RC-5B). The lysate was flowed over a Ni-NTA HP column (GE Healthcare) loaded onto the Aktaxpress FPLC system (Amersham Bioscience) and washed with Buffer A. Protein was eluted with Buffer B (20 mM Potassium Phosphate pH 7.4, 500 mM Imidazole, 500 mM NaCl). For *Bu*GUS-1 used for crystallography, the His-tag was removed by TEV cleavage in the presence of 1 mM DTT and incubated overnight at 4 °C. This sample was then applied to the Ni-NTA column again and the flow through was collected. Fractions containing the protein of interest were combined and passed over a HiLoadTM 16/60 SuperdexTM 200 gel filtration column. Proteins were eluted in S200 Buffer (20 mM HEPES pH 7.4, 50 mM NaCl), except for BMSP and the *Bu*GUS-1 ΔN-term loop mutant, which were eluted in S200 buffer that contained 300 mM NaCl. Fractions were analyzed by SDS-PAGE and those with >95%

purity were combined and concentrated for long-term storage at -80 °C.

Crystallization and Structure Determination

Protein crystals were formed at 20 °C via the hanging drop vapor diffusion method with 15-well EasyXtal Qiagen trays with 300 µL of well solution in the reservoir and 3 µL total drop volumes. Drop conditions were as follows: *Bu*GUS-1-apo, 2 µL of 15 mg/mL *Bu*GUS-1 and 1 µL of 20% PEG 1000 and 0.1 M Tris pH 8.5; *Bu*GUS-1-G-1,5-L crystals, 2 µL of 10 mg/mL *Bu*GUS-1 in 1 µL of 16% PEG 1000, Tris pH 8.5, and 2 mM glucaro-1,4-lactone; *Bu*GUS-1-GlcA, 2 µL of 10 mg/mL *Bu*GUS-1 in 1 µL of 0.1 M sodium citrate pH 5.5, 20% PEG 3000, and 20 mM glucuronic acid; *Bu*GUS-2-G-1,5-L crystals, 2 µL of 12 mg/mL *Bu*GUS-2 in 1 µL of 0.2 M potassium chloride, 20% PEG 3350, and 10 mM glucaro-1,4-lactone; *Bu*GUS-2 ΔCa^{2+}, 2 µL of 12 mg/ml of the *Bu*GUS-2 D341A/D367A mutant in 1 µL of 0.2 M potassium chloride and 20% PEG 3350; *Bu*GUS-3, 2 µL 15 mg/ml *Bu*GUS-3 in 1 µL 8% PEG 3350, 0.2 M potassium thiocyanate and 0.1 M HEPES pH 8.0; The following three structures were derived from crystals made at 20 °C via the sitting drop method in Hampton Research 3-well Midi Crystallization Plates (Swissci) by an Art Robbins Instruments Crystal Phoenix robot with the following drop conditions: BMSP, 100 nL of 8 mg/ml BMSP in 100 nL of 0.2 M magnesium chloride, 0.1 M MES:NaOH pH 5.5, and 20% PEG 4000; *Bu*GUS-1-PTG, 100 nL of 10 mg/mL *Bu*GUS-1 in 100 nL of 0.2 M potassium sulfate, 20% PEG 3350 and 10 mM phenyl-thio-β-D-glucuronide; *Bu*GUS-1 ΔN-term loop, 100 nL of 12 mg/mL of the *Bu*GUS-1 ΔN-term loop mutant in 100 nL of 0.1 M sodium citrate pH 5.5 and 20% PEG 3000. Prior to crystallization, the N-terminal histidine tag was removed from *Bu*GUS-1 as described in the purification methods above.

Crystal specimens were cryo-protected in the crystallization conditions as described

above containing 20% glycerol, and diffraction data were collected for all crystals at 100 °K at APS Beamline 23-ID-D, except for *Bu*GUS-2 ΔCa^{2+}, which was collected at APS Beamline 23-ID-B. The data were processed with XDS and all structures were solved via molecular replacement in Phenix [113] using the *E. coli* GUS structure (5CZK) as a search model for *Bu*GUS-1-apo, the *B. uniformis* GUS structure (5UJ6) as a search model for *Bu*GUS-2- ΔCa^{2+}, the *B. fragilis* structure (3CMG) as a search model for *Bu*GUS-3, and the *Bu*GUS-1-apo structure (6D1N) for the remaining structures. The resulting starting model and maps from molecular replacement were then used in the AutoBuild function of Phenix. Structures were refined in Phenix and visually inspected and manually built using COOT [114]. Final PDB coordinates for all structures have been deposited to the RCSB Protein Data Bank with corresponding PDB codes in parentheses: *Bu*GUS-1 (6D1N), *Bu*GUS-3 (6D1P), BMSP (6D8K), *Bu*GUS-1 + G-1,5-L complex (6D41), *Bu*GUS-2 + G-1,5-L complex (6D5O), *Bu*GUS-1 + GlcA complex (6D6W), *Bu*GUS-1 Δloop (6D89), *Bu*GUS-2 calcium binding mutant (6D8G), and *Bu*GUS-1 + TPG complex (6D7F).

GUS activity assay of 4-MUG hydrolysis

Initial pH screening was performed with PNPG, as described previously. Because PNPG is not amenable for continuous kinetic studies at pH below 6.5, we utilized an analogous but fluorescent GUS substrate 4-methylumbelliferyl-β-D-glucuronide (4-MUG) for subsequent kinetic investigations. In vitro assays of GUS activity with 4-MUG were carried out in costar black 96-well clear flat bottom plates. Total reaction volume was 50 µL with 5 µL of GUS and 5 µL of 10x buffer (250 mM HEPES, 250 mM NaCl, pH 7.0) mixed and pre-incubated at 37 °C before reaction initiation by addition of 40 µL 4-MUG. Concentration of enzyme was specific to each GUS: 5 nM *Ec*GUS, 5 nM *Bu*GUS-1, 20 nM

*Bu*GUS-2, 40 nM *Bu*GUS-1 Δloop, 80 nM BMSP GUS, and 320 nM BuGUS-3. Reactions were monitored continuously in a BMG lab tech PHERAstar plate reader with an excitation wavelength of 350 nm and an emission wavelength of 450 nm. Resultant progress curves were fit by a custom linear regression analysis program in MATLAB. Initial velocities were then analyzed in the enzyme kinetics module of SigmaPlot 13.0 by Michaelis-Menten fit to determine the catalytic turnover (k_{cat}) and Michaelis constant (K_m).

GUS activity assay of SN-38-G hydrolysis

In vitro assays of GUS activity with the substrate SN-38-G were carried out in costar black 96-well clear flat bottom plates. Total reaction volume was 50 µL with 5 µL of SN-38-G at a range of low substrate concentrations (15, 10, 7.5, 5, and 2.5 µM final), 5 µL of 10x buffer (250 mM HEPES, 250 mM NaCl, pH 7.0), and 35 µL of water mixed and pre-incubated at 37 °C before reaction initiation by addition of 5 µL of GUS. Concentration of enzyme was specific to each GUS: 5 nM *Ec*GUS, 5 nM *Bu*GUS-1, 20 nM *Bu*GUS-2, 40 nM *Bu*GUS-1 Δloop, 80 nM BMSP GUS, and 320 nM BuGUS-3. Reactions were monitored continuously by fluorescence with an emission wavelength of 420 nm and an excitation wavelength of 230 nm. Resultant progress curves were fit by a custom linear regression analysis program in MATLAB. Initial velocities were then plotted against substrate concentration and fit with linear regression in Microsoft Excel to determine catalytic efficiency (k_{cat}/K_m).

GlcA-Containing Polysaccharide Processing Assay

The sulfated heparin-like nonasaccharide [GlcA-(GlcNS-GlcA)$_4$-PNP (where GlcA is glucuronic acid, GlcNS is N-sulfated glucosamine)] and the acetylated heparin-like nonasaccharide [GlcA-(GlcNAc-GlcA)$_4$-PNP (where GlcA is glucuronic acid, GlcNAc is N-

acetylglucosamine)] substrates were from Glycantherapeutics. The additional polysaccharides employed were synthesized in-house (Xu et al., 2014). Putative polysaccharide substrates were digested with each GUS enzyme for 3 hours. Digestion reactions were composed of 0.5 µM GUS enzyme and 10 µg oligosaccharide. Reactions were terminated by heating for five minutes at 95°C. Aliquots of the resultant solutions were analyzed by polyamine-based anion exchange (PAMN)-HPLC. Sugars were eluted from the PAMN column (0.46 cm x 25 cm from Waters) with a linear gradient of KH_2PO_4 from 0 to 1 M in 40 min at a flow rate of 0.5 ml/min. The eluent was monitored by a UV detector at 310 nm. Aliquots of the digestion reactions were analyzed by electrospray ionization mass spectrometry (ESI-MS) by first purifying the reaction mixture by C18 column eluted with a linear gradient of methanol with 1% trifluoroacetic acid from 0 to 1 M in 60 min at a flow rate of 0.5 ml/ min. The purified oligosaccharides were then dried. ESI-MS analysis was performed on a Thermo LCQ-Deca in negative ionization mode. A syringe pump (Harvard Apparatus) was used to introduce the sample by direct infusion (50 ml/min). The purified oligosaccharides were diluted in 200 ml of H_2O with the electrospray source set to 3 KV and 150C. The automatic gain control was set to 1×10^7 for full scan MS. The MS data were acquired and processed using Xcalibur 1.3.

GUS Inhibition Assay

In vitro assays of GUS activity with the substrate 4-MUG were carried out in costar black 96-well clear flat bottom plates. Total reaction volume was 50 µL with 5 µL of GUS, 10 µL of 5x buffer (125 mM HEPES, 125 mM NaCl, pH 7.0), and 5 µL of inhibitor mixed and pre-incubated at 37 °C before reaction initiation by addition of 30 µL 4-MUG. Reactions were monitored continuously in a PHERAstar plate reader at 410 nm. End point absorbance

values after one hour were converted to % inhibition values via the following equation:

$$\%\ inhibition = \left[1 - \left(\frac{A_{exp} - A_{bg}}{A_{max} - A_{bg}}\right)\right] \times\ 100$$

where A_{exp} is the end point absorbance at a particular inhibitor concentration, A_{max} is the absorbance of the uninhibited reaction, and A_{bg} is the background absorbance. Percent inhibition values were subsequently plotted against the log of inhibitor concentration and fit with a four-parameter logistic function in SigmaPlot 13.0 to determine the concentration at which 50% inhibition (IC_{50}) is observed.

SEC-MALS Analysis of BuGUS enzymes

*Bu*GUS-1, *Bu*GUS-2, and *Bu*GUS-3 were analyzed on a Superdex 200 size exclusion column connected to an Agilent FPLC system, Wyatt DAWN HELEOS II multi-angle light scattering instrument and a Trex refractometer. The injection volume was 50 µL, and each protein was assessed at 10 mg/mL in 50 mM HEPES and 150 mM NaCl pH 7.4 buffer. A flow rate of 0.5 mL/min was used. Light scattering and refractive index data were collected and analyzed using Wyatt ASTRA (Ver. 6.1) software. A dn/dc value of 0.185 was used for calculations. Approximately 99% of *Bu*GUS-1 eluted in a single peak with a molar mass of 275 kDa, indicating that it forms a tetramer in solution. In contrast, 99% of *Bu*GUS-2 and 95% *Bu*GUS-3 eluted in single peaks with molar masses of 189 kDa and 175 kDa, respectively, indicating that they form dimers in solution.

CD Analysis of BuGUS-2 Calcium Binding Mutant

The protein stabilities of WT *Bu*GUS-2 and the calcium binding mutant were determined using the Circular Dichroism method [115]. Enzyme (1 µM) in CD buffer containing 10 mM potassium phosphate (pH 7.4) and 100 mM potassium fluoride was loaded

into a 1-mm cuvette. Using a Chirascan-plus instrument (Applied Photophysics Limited), spectra from 185 to 280 nM were recorded at 20 ± 1.0 °C. Measurements were corrected for background signal using a CD buffer sample. The melting profile of the sample (5 µM) was monitored at 218 nm from 25 °C to 94 °C.

Manual Docking of Monosaccharide in PyMOL

Galacturonate, mannonate, and iduronate monosaccharides were accessed from the PDB in previously solved crystal structures (PDB: 1KCC for galacturonate, PDB: 3VLW for mannuronate, and PDB: 4OBR for iduronate). These were then imported into PyMOL and manually aligned to the GlcA-bound *Bu*GUS-1 structure (6DW6) with the 3-button editing tool. After manual alignment of the sugar monosaccharides, structures of *Bu*GUS-1 and *Bu*GUS-3 were aligned to the GlcA-bound *Bu*GUS-2 structure. Visual inspection and final figures after alignment were generated in PyMOL.

Rosetta Modeling

The full length amino acid sequences *Bu*GUS-1, *Bu*GUS-3, and BMSP GUS were submitted to the Robetta modeling server [116–118] to produce 3D homology models of these proteins, including their oligomeric complexes, based on template protein structures available in the Protein Data Bank (PDB) from December 2017 to January 2018. The *Bu*GUS-1 and BMSP GUS Robetta homology models were based on the *E. coli* β-glucuronidase structure (PDB ID: 3LPF). For the *Bu*GUS-3 homology model, the Robetta selected template was a beta-galactosidase from *Bacillus circulans* ATCC 31382 (PDB ID: 4YPJ). Backbone C-alpha coordinates of the homology model protein structures were then superimposed onto X-

ray crystal structures using TM-align algorithms [119].

Identification of Predicted Calcium Binding Sites

To identify calcium binding sites in GUS enzymes from the HMP dataset, the 279 GUS protein sequences previously identified [100] were aligned pairwise to *Bu*GUS-2 using NCBI BLASTp [120]. These alignments were then probed for the three aspartic acid residues in *Bu*GUS-2 (D176, D341, and D367) deemed necessary for calcium binding.

Identification of Tryptophan Substitutions

To identify additional GUS enzymes in the HMP Clustered genes (HMGC) dataset that possess a tryptophan rather than a tyrosine at position W483 in *Bu*GUS-3, the ~267,000 sequences previously determined to share 25% identify with *Ec*GUS, *Sa*GUS, *Cp*GUS, and *Bf*GUS [100] were aligned pairwise to the these GUS enzymes and *Bu*GUS-3 using NCBI BLASTp [120]. The sequences were then probed for the presence of the NxK motif, catalytic E residues, and N and W motifs.

Identification of N-terminal Loops

To identify N-terminal loops in GUS enzymes from the HMP dataset, the 279 GUS protein sequences previously identified [100] were aligned pairwise to *Bu*GUS-1 using NCBI BLASTp [120]. These alignments were then probed for the N-terminal loop in *Bu*GUS-1.

CHAPTER 3: SELECTING A SINGLE STEREOCENTER: THE MOLECULAR NUANCES THAT DIFFERENTIATE β-HEXURONIDASES IN THE HUMAN GUT MICROBIOME.[3]

The human gut microbiome encodes a wide range of glycoside hydrolases (GHs) that play key roles in host health and disease. A prominent group of GHs in the gut microbiome is GH family 2 (GH2), which includes the well-studied β-glucosidases and β-galactosidases, as well as their sugar acid-cleaving counterparts β-glucuronidases (GUS) and β-galacturonidases (GalAse)[121]. Analogous to glucose and galactose released by β-galactosidases and β-glucosidases, glucuronate (GlcA) and galacturonate (GalA) released by bacterial GUSs and GalAses also serve as a source of energy. Bacteria can catabolize GlcA and GalA via either an isomerase or oxidative catabolic pathway[122]. GlcA and GalA are present in host-derived and plant-based polysaccharides, respectively, including glycosaminoglycans in animals and pectin from plants[10,51]. In addition to polysaccharide utilization, bacterial GUS enzymes are important because they play key roles in reversing mammalian drug metabolism[109]. For example, a variety of drugs are glucuronidated in the liver, a process by which uridine diphosphate glucuronosyl transferases (UGTs) attach GlcA to available nucleophilic moieties[11]. Bacterial GUS enzymes can metabolize drug glucuronides generated by the liver and release active and sometimes toxic drug into the gastrointestinal tract. Indeed, the toxic side effects of cancer drugs and NSAIDs are alleviated by selective bacterial GUS

[3] This chapter previously appeared as an article in *Biochemistry*. The original citation is as follows: Pellock, S. J., Walton, W. G., Redinbo, M. R. (2019) Selecting a Single Stereocenter: The molecular nuances that differentiate β-Hexuronidases in the human gut microbiome. *Biochemistry*. *58*, 1311-1317.

inhibition[56,62,123,124]. Together, the key roles that gut microbial GUSs and GalAses play in polysaccharide and drug metabolism in the human gut make them important proteins for structural and functional characterization.

We previously screened the human microbiome project (HMP) stool sample metagenomic database for putative GH2 GUSs[100]. This bioinformatic effort yielded an atlas of 279 putative GUSs, of which only a handful have been characterized either structurally or functionally to date (**Figure 3.1**)[100,125,126]. During the characterization of *Bu*GUS-1 from *Bacteroides uniformis*, we found that, in addition to processing β-glucuronides, *Bu*GUS-1 also hydrolyzed β-galacturonides[126], suggesting that this gut microbial GUS, and others yet to be identified, may process sugar conjugates beyond glucuronides. Indeed, using novel crystal structures and site-directed mutagenesis as a guide, here we show that there are 13 GalAses and 10 hybrid GUS/GalAses in the 279 previously annotated GUSs from the HMP stool sample database. Further kinetic and structural studies revealed that variations on a shared active site structure allow GUSs and GalAses to selectively process epimeric substrates.

MATERIALS AND METHODS

Gene synthesis, mutagenesis, protein expression, and protein purification

All genes were synthesized by BioBasic and cloned into a pUC57 vector. The genes encoding *Et*GalAse, *Fs*GUS, and *Fc*GalAse were then sub-cloned into a pET His6 LIC cloning vector (2Bc-T) for expression with a C-terminal His tag. *Fs*GUS was also cloned into the pLIC-

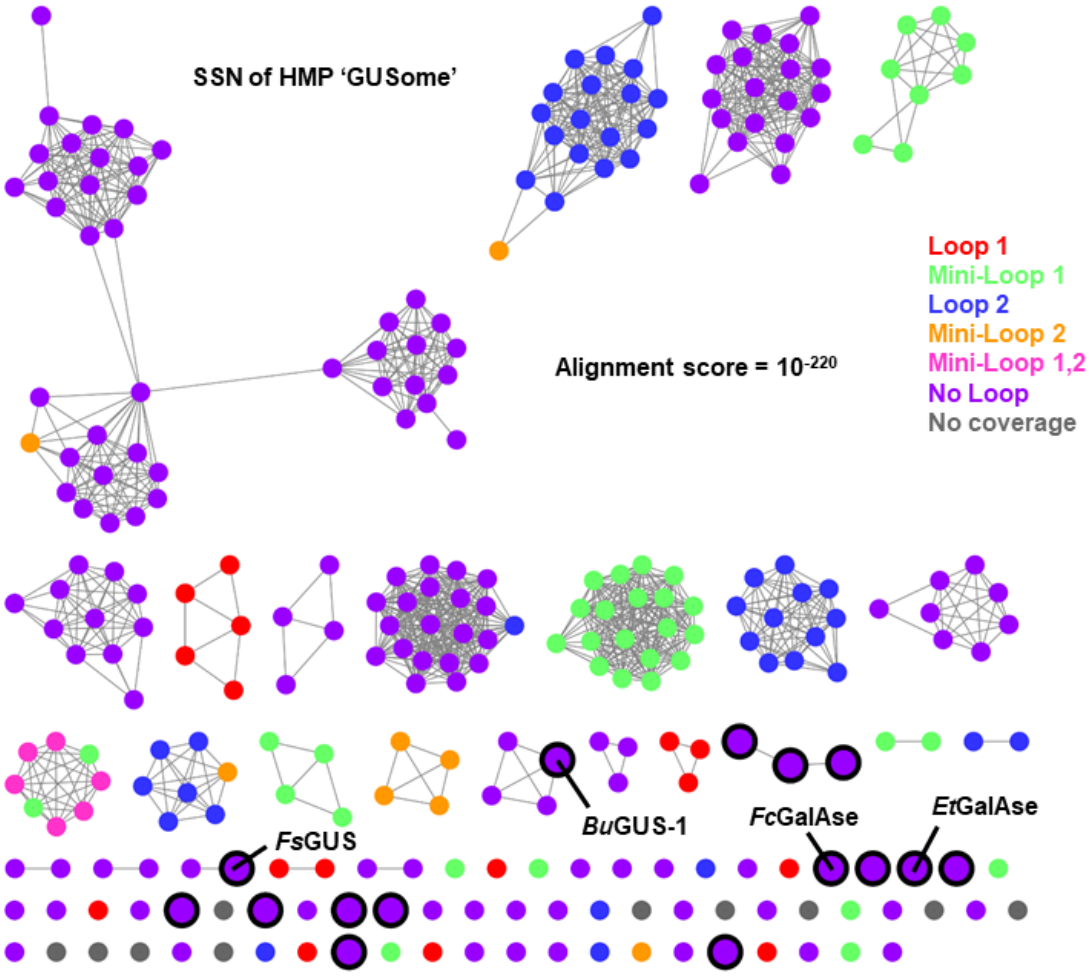


Figure 3.1 - Sequence similarity network (SSN) of putative GUS enzymes from the HMP stool sample metagenomic database. Proteins analyzed in this study are labelled and colors designate loop class.

Table 3.1 - Ligation independent cloning primer sequences.

Primer name	Primer sequence
*Et*GalAse C-term His-tag fw	5' TTTAAGAAGGAGATATAGTTCATGATCCGTACCTTCGAAACCCATAAAATCCGTAAAACC 3'
*Et*GalAse C-term His-tag rv	5' GGATTGGAAGTAGAGGTTCTCGAAGTAGTTACCCAGGGAACGGTAGGAATCTTTAACTAC 3'
*Fc*GalAse C-term His-tag fw	5' TTTAAGAAGGAGATATAGTTCATGTCTGATCGTACCTTCGC 3'
*Fc*GalAse C-term His-tag rv	5' GGATTGGAAGTAGAGGTTCTCACCTTTCTGTTTCTGGAACA 3'
*Fs*GUS C-term His-tag fw	5' TTTAAGAAGGAGATATAGTTCATGGAAGCGAAAAAAGAGAAAAAATACATGTCTGATATT 3'
*Fs*GUS C-term His-tag rv	5' GGATTGGAAGTAGAGGTTCTCCTGTTCTTTACGTTTCAGTTCTTCGTAGAATTTCTGCAG 3'
*Fs*GUS C-term His-tag Δ16 fw	5' TTTAAGAAGGAGATATAGTTCATGGATTACACCGAACAGTACGAAACCGGTTTCGCTAC 3'

His vector pMCSG7 and expressed with an N-terminal tag for crystallography. Cloning primers are listed in **Table 3.1**. All mutants were generated by standard site-directed mutagenesis and primers utilized for mutagenesis were purchased from Integrated DNA Technologies (IDT) and are listed in **Table 3.2**. All cloning and mutagenesis experiments were confirmed by DNA sequencing by Eton Bioscience Inc. Proteins were purified as described previously[100]. Briefly, all proteins were expressed via standard IPTG induction with growth overnight at 18 °C. Cell lysates from expression were clarified and subject to Ni-based affinity chromatography and size exclusion chromatography. Final protein concentrations were determined with a ND-1000 spectrophotometer. Proteins were snap frozen in liquid nitrogen and stored at -80 °C.

Crystallography

All crystals were grown by sitting drop vapor diffusion in Hampton Research 3-well Midi Crystallization Plates (Swissci) by an Art Robbins Instruments Crystal Phoenix. All crystals were grown at 20 °C in the following crystallants: ***Et*GalAse** in 1 M NaH_2PO_4/K_2HPO_4, pH 8.2 and 1 mM GalA, ***Et*GalAse bound to GalA** in 0.2 M ammonium sulfate, 0.1 M Bis-Tris:HCl, pH 6.5, 25% (w/v) PEG 3350, 1 mM GalA, **apo *Fs*GUS (N-terminal His tag)** in 0.15 M DL-malic acid, pH 7.0, 20% w/v PEG 3350, and ***Fs*GUS (C-terminal His tag) bound to phenyl-thio-β-D-glucuronide (PTG)** in 0.2 M sodium tartrate dibasic, 20% (w/v) PEG 3350, and 1 mM PTG. Resultant crystals were cryoprotected in their corresponding crystallants with 20% glycerol added. For liganded structures, 1 mM of ligand was also present in the cryoprotectant solution. Diffraction data were collected for all crystals at 100 K at APS Beamline 23-ID-D. Data were reduced in XDS and scaled in Aimless[127,128]. The *Et*GalAse structures were solved by molecular replacement in Phenix using a truncated

Table 3.2 - Site-directed mutagenesis primer sequences.

Primer name	Primer sequence
*Et*GalAse R337A fw	5' GAAGAAAACCACGCCGCTGGCCTGAGCGAAGA 3'
*Et*GalAse R337A rv	5' TCTTCGCTCAGGCCAGCGGCGTGGTTTTCTTC 3'
*Fs*GUS Y377F fw	5' GTGGGAAGAGATTCCGGTTTTCTGGGCAATCC 3'
*Fs*GUS Y377F rv	5' GGATTGCCCAGAAAACCGGAATCTCTTCCCAC 3'
*Fc*GalAse R335A fw	5' GTTTGGGAAGAAGCGCACGCTGCTGGTCTTCAAGAAG 3'
*Fc*GalAse R335A rv	5' CTTCTTGAAGACCAGCAGCGTGCGCTTCTTCCCAAAC 3'
*Bu*GUS-1 Y382F fw	5' CCAGTGGATGGTCCAGAAAACCGGAATTTCAGACC 3'
*Bu*GUS-1 Y382F rv	5' GGTCTGAAATTCCGGTTTTCTGGACCATCCACTGG 3'

model of *E. coli* GUS (PDB: 3LPF) as the search model. The *Fs*GUS structures were also solved via molecular replacement in Phenix and *Bu*GUS-1 (PDB: 6D1N) was used as the search model. Subsequent refinement and ligand placements were performed in Phenix. Final coordinates and maps were deposited in the PDB for *Et*GalAse (PDB: 6NCW), *Et*GalAse bound to GalA (PDB: 6NCX), *Fs*GUS (PDB: 6NCY), and *Fs*GUS bound to PTG (PDB: 6NCZ) (**Table 3.3**).

Kinetic assays for *p*NP-GalA and *p*NP-GlcA processing

Assays to determine the catalytic parameters of GUS and GalAse activity were performed with the chromogenic substrates *p*-nitrophenyl-β-D-galacturonide (*p*NP-GalA) and *p*-nitrophenyl-β-D-glucuronide (*p*NP-GlcA). Briefly, assays contained 5 µL enzyme, 5 µL 250 mM NaCl and 250 mM HEPES (pH 6.5 and 7) or 5 µL 250 mM NaCl and 250 mM sodium acetate (pH 5.5), and 40 µL substrate. For reactions that showed no activity, enzyme concentrations up to 300 nM were tested. For reactions with measurable activity, enzyme concentrations were as follows: *Et*GalAse WT (5 nM), *Et*GalAse R337A (300 nM), *Fc*GalAse WT (5 nM), *Fs*GUS WT (5 nM), *Fs*GUS Y377F (20 nM), and *Fs*GUS Δ16 (20 nM). Given the lower water solubility of *p*NP-GalA, all kinetic assays contained 1.6% DMSO final. For pH values greater than or equal to 7, reactions were monitored continuously at 410 nm and incubated at 37 °C in Costar 96-well black, flat, clear bottom plates in a BMG lab tech PHERAstar plate reader. For enzyme reactions below pH 6.5, reactions were quenched with 100 µL of 0.2 M sodium carbonate and measured as described for continuous monitoring. Initial velocities were determined in MATLAB and Michaelis-Menten fits were performed in SigmaPlot 13.0.

Table 3.3 - Collection and refinement statistics of crystal structures.

Protein	*Et*GalAse	*Et*GalAse + GalA	*Fs*GUS	*Fs*GUS + PTG
Resolution range	29.54-2.10 (2.18-2.10)	29.88-2.25 (2.33-2.25)	29.71-2.05 (2.12-2.05)	29.67-2.20 (2.28-2.20)
Space group	P1	$P12_11$	$P2_13$	P1
Unit cell [a, b, c (Å); α,β,γ (°)]	65.9, 87.7, 124.5; 93.3, 101.5, 90.4	65.3, 156.0, 124.3; 90.0, 101.3, 90.0	192.6, 192.6, 192.6; 90.0, 90.0, 90.0	109.1, 116.9, 117.6; 60.3, 62.5, 87.1
Total reflections	549589 (53411)	783043 (65156)	5130192 (211530)	767388 (77103)
Unique reflections	155133 (15342)	114891 (11092)	147997 (14676)	215731 (21435)
Multiplicity	3.5 (3.5)	6.8 (5.9)	34.7 (14.4)	3.6 (3.6)
Completeness (%)	97.5 (96.4)	99.3 (96.3)	99.9 (99.7)	97.4 (96.8)
Mean I/sigma(I)	8.4 (1.3)	6.7 (1.1)	21.6 (1.4)	8.3 (1.1)
Wilson B-factor	29.3	25.1	31.4	32.4
R-merge	0.111 (0.896)	0.214 (1.363)	0.168 (1.711)	0.116 (1.093)
R-pim	0.069 (0.564)	0.088 (0.609)	0.028 (0.460)	0.071 (0.664)
CC1/2	0.995 (0.545)	0.995 (0.674)	0.999 (0.578)	0.995 (0.498)
R-work	0.165 (0.273)	0.221 (0.330)	0.153 (0.248)	0.181 (0.303)
R-free	0.207 (0.324)	0.275 (0.348)	0.181 (0.267)	0.220 (0.320)
# of non-hydrogen atoms	19641	19071	10830	30846
macromolecules	18091	17961	9643	28928
ligands	64	56	49	150
solvent	1486	1054	1138	1768
Protein residues	2240	2227	1170	3532
RMS(bonds)	0.009	0.008	0.009	0.008
RMS(angles)	1.24	0.92	1.22	0.95
Ramachandran outliers (%)	0.18	0.32	0.00	0.00
Rotamer outliers (%)	0.77	2.48	1.16	0.13
Clashscore	2.45	4.43	3.85	4.84
Average B-factor	29.2	26.7	33.3	31.2
macromolecules	28.9	26.6	32.1	31.2
ligands	30.4	22.3	48.6	32.1
solvent	33.2	27.9	42.1	32.7
PDB code	6NCW	6NCX	6NCY	6NCZ

Statistics for the highest-resolution shell are shown in parentheses.

Kinetic assays for SN-38-G processing

SN-38-G processing was measured by a fluorescence-based assay as described previously[126]. Briefly, assays were performed in Corning 96-well clear, half-area, UV-transparent plates that contained 5 µL enzyme, 5 µL 250 mM NaCl and 250 mM HEPES (pH 7), 5 µL substrate, and 35 µL water. Reactions were pre-incubated at 37 °C for ten minutes before addition of enzyme. Final DMSO concentrations were 4% for all reactions. Data were fit as described for the *p*NP-GlcA and *p*NP-GalA assays above.

Circular dichroism

Circular dichroism experiments were performed as previously described[126]. In brief, WT and mutant proteins (2 µM) were prepared in CD buffer (10 mM potassium phosphate, 100 mM potassium fluoride, pH 7.4) and loaded into a 1-mm cuvette. Both spectra scans and melting temperatures were performed using a Chirascan-plus instrument (Applied Photophysics Limited).

Size exclusion chromatography-multi-angle light scattering

Oligomeric states of purified proteins in solution were determined as previously described with slight modifications[126]. Briefly, proteins were analyzed on a Superdex 200 size exclusion column connected to an Agilent FPLC system, Wyatt DAWN HELEOS II multi-angle light scattering instrument and a Trex refractometer. Elution buffers were the same as employed for size exclusion chromatography (20 mM HEPES, 50 mM NaCl, pH 7.4). All proteins were run at approximately 5 mg/ml. Light scattering and refractive index data were collected and analyzed using Wyatt ASTRA (Ver. 6.1) software.

Sequence similarity network and genome neighborhood diagram generation

Sequence similarity networks (SSNs) were generated using the 279 putative GUS protein sequences previously identified by a structure- and function-guided search of the HMP stool sample metagenomic database[100]. The 279 sequences were used as input into the FASTA (Option C) section of the web-based Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST)[129] to determine initial alignment scores for the generation of SSNs. Resultant SSNs at multiple different alignment scores were visualized in Cytoscape 3.7.0[130]. Genome neighborhood diagrams (GNDs) were generated with the web-based EFI-Genome Neighborhood Network Tool[129]. UniProt IDs of the proteins of interest were used as input to retrieve the associated GNDs. For protein sequences without a definitive match in either the UniProtKB or NCBI databases, the highest identity match in the UniProtKB database was utilized to generate a GND.

NA = no activity. [a]Catalytic efficiency determined from non-saturating substrate concentrations. Values shown are an average of n = 3 replicates ± SD.

RESULTS

Putative GUS from *Eisenbergiella tayi* is instead a GH2 GalAse

The vast majority of the previously annotated 279 GUS enzymes in the HMP stool sample database remain uncharacterized (**Figure 3.1**)[100]. Thus, we have performed gene synthesis, expression, and purification of a range of these putative GUSs to determine their structure and function. We expressed and purified the putative GUS from the gut microbe *Eisenbergiella tayi* (UniProtKB ID: A0A1E3AEY6), but it did not process the standard reporter glucuronide *p*NP-GlcA (**Figure 3.2** and **Table 3.4**). However, we recently showed that a GUS from the gut microbe *Bacteroides uniformis*, termed *Bu*GUS-1, processed *p*-

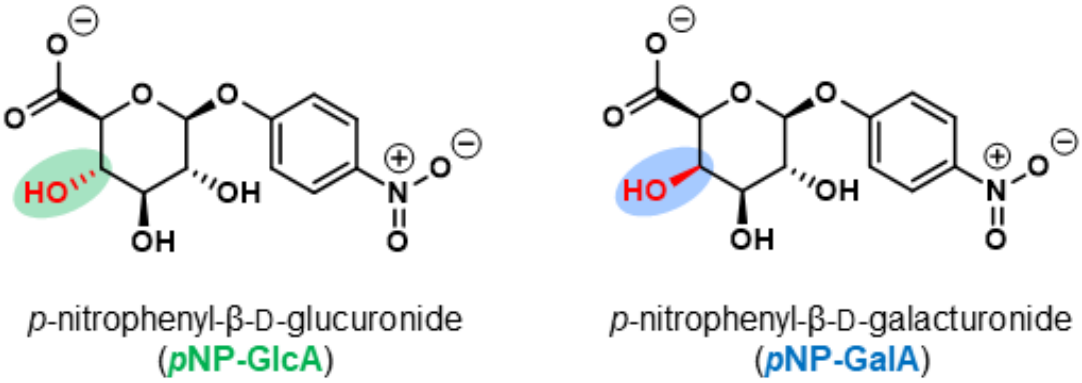


Figure 3.2 - Structures of p-nitrophenyl-β-D-glucuronide (pNP-GlcA) and p-nitrophenyl-β-D-galacturonide (pNP-GalA) with stereochemical difference highlighted in red.

Table 3.4 - Kinetic parameters of *p*NP-GlcA and *p*NP-GalA processing.

	*p*NP-GlcA			*p*NP-GalA		
Enzyme	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (s^{-1} M^{-1})	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (s^{-1} M^{-1})
*Et*GalAse	NA	NA	NA	98 ± 3	160 ± 6	$6.2x10^5$
*Et*GalAse R337A	1.0 ± 0.1	110 ± 20	$5.2x10^3$	NA	NA	NA
*Fc*GalAse	NA	NA	NA	42 ± 3	58.8 ± 0.2	$7.1x10^5$
*Fc*GalAse R335A	NA	NA	NA	NA	NA	NA
*Fs*GUS	15 ± 1	30 ± 6	$5.0x10^5$	34 ± 2	360 ± 20	$9.4x10^4$
*Fs*GUS Y377F	8.9 ± 0.4	55 ± 4	$1.6x10^5$	--	--	[a]$8.9x10^3$
*Fs*GUS Δ16	3.6 ± 0.6	70 ± 10	$5.1x10^4$	14 ± 1	360 ± 60	$3.9x10^4$

NA = no activity. [a]Catalytic efficiency determined from non-saturating substrate concentrations. Values shown are an average of n = 3 replicates ± SD.

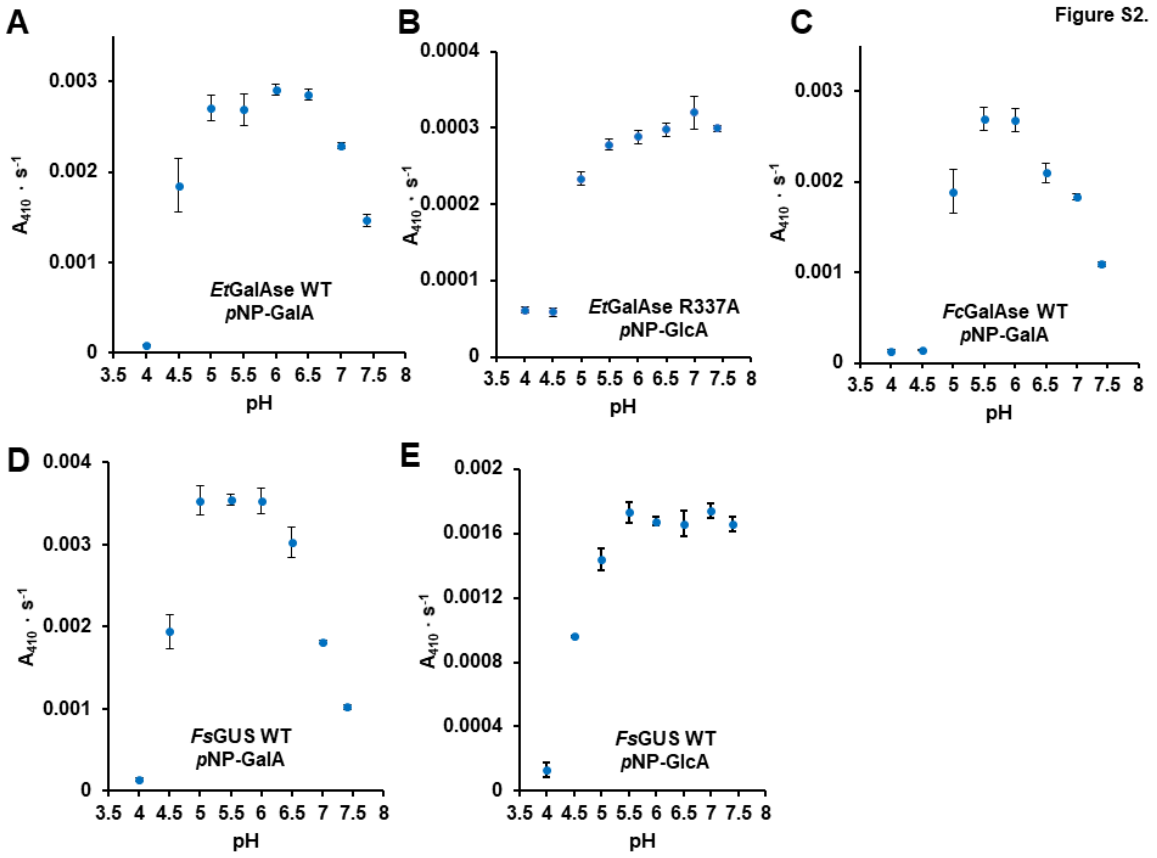


Figure 3.3 - pH profiles of GUS enzymes characterized in this study. pH profiles of (**A**) *Et*GalAse with *p*NP-GalA, (**B**) *Et*GalAse R337A with *p*NP-GlcA, (**C**) *Fc*GalAse WT with *p*NP-GalA, (**D**) *Fs*GUS WT with *p*NP-GalA, and (**E**) *Fs*GUS WT with *p*NP-GlcA.

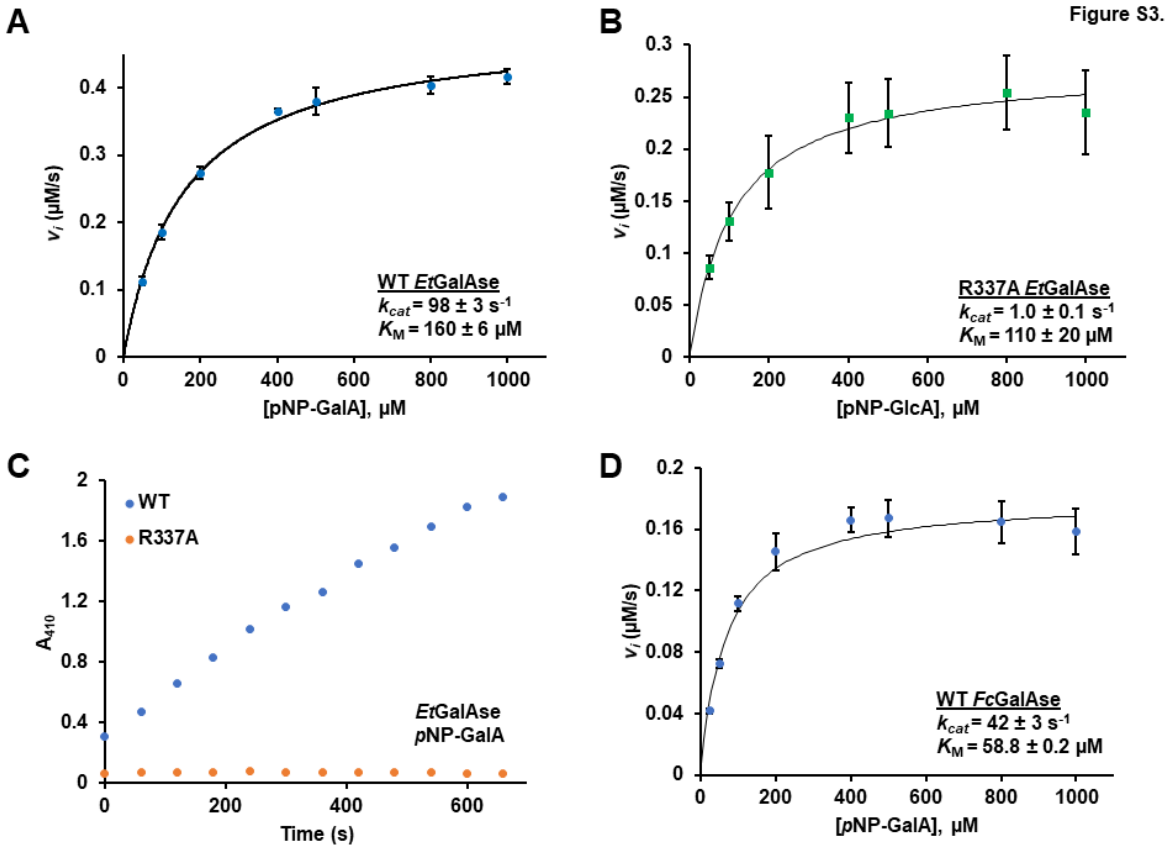


Figure 3.4 - Michaelis-Menten plots and progress cures for EtGalAse and FcGalAse.

(**A**) Michaelis-Menten plot for WT *Et*GalAse with *p*NP-GalA, (**B**) Michaelis-Menten plot for R337A *Et*GalAse mutant with pNP-GlcA, (**C**) Progress curves of WT and R337A mutant with *p*NP-GalA, and (**D**) Michaelis-Menten plot of WT *Fc*GalAse with *p*NP-GalA.

nitrophenyl-β-D-galacturonide (*p*NP-GalA) as well as *p*NP-GlcA (**Figure 3.2**)[126]. Thus, we tested if the putative GUS from *E. tayi* could also process *p*NP-GalA. Indeed, the *E. tayi* protein processed *p*NP-GalA efficiently with optimal activity at pH 6 (k_{cat} = 98 ± 3 s^{-1}, K_M = 160 ± 6 µM, k_{cat}/K_M = 6.2x10^5 M^{-1} s^{-1}) (**Table 3.4**, **Figure 3.3A**, and **Figure 3.4A**). The catalytic efficiency of the *E. tayi* enzyme (*Et*GalAse) with *p*NP-GalA is similar to *p*NP-GlcA cleavage by *bona fide* GUS enzymes[56]. The results presented here demonstrate that the putative GUS from E. tayi, annotated as a GUS by sequence analysis, is instead a GH2 GalAse.

Mutagenesis of a single residue converts EtGalAse into a GUS

We determined the 2.10 Å and 2.25 Å resolution crystal structures of glycerol-bound and GalA-bound *Et*GalAse (**Table 3.3**). *Et*GalAse is a tetramer (**Figure 3.5A**) similar to previously characterized GUS enzymes[56,123]. The active site also resembles previously characterized GUSs, as it contains the key catalytic glutamates, Glu-378 (acid/base) and Glu-465 (nucleophile), and the NxK motif previously identified as essential for GUS activity (**Figure 3.5B**)[126]. A unique feature of the *Et*GalAse active site, however, is Arg-337, which hydrogen bonds to the axial 4-hydroxyl of GalA, the stereocenter that differentiates the epimers GalA and GlcA (**Figure 3.2** and **Figure 3.5B**). To assess the role of Arg-337 in GalAse function, we mutated Arg-337 to alanine in *Et*GalAse. The equivalent residue in established GUS enzymes is observed to be valine (*Ec*GUS, *Cp*GUS, *Sa*GUS, and *Ee*GUS) or isoleucine (*Bu*GUS-2)[56,100,123,131]. Remarkably, the R337A mutant of *Et*GalAse both abolished GalAse activity (**Figure 3.4C**) and conferred GUS activity (k_{cat} = 1.0 ± 0.1 s^{-1}, K_M = 110 ± 20 µM, k_{cat}/K_M = 5.2x10^3 M^{-1} s^{-1}) (**Table 3.4**, **Figure 3.3B**, and **Figure 3.4B**).

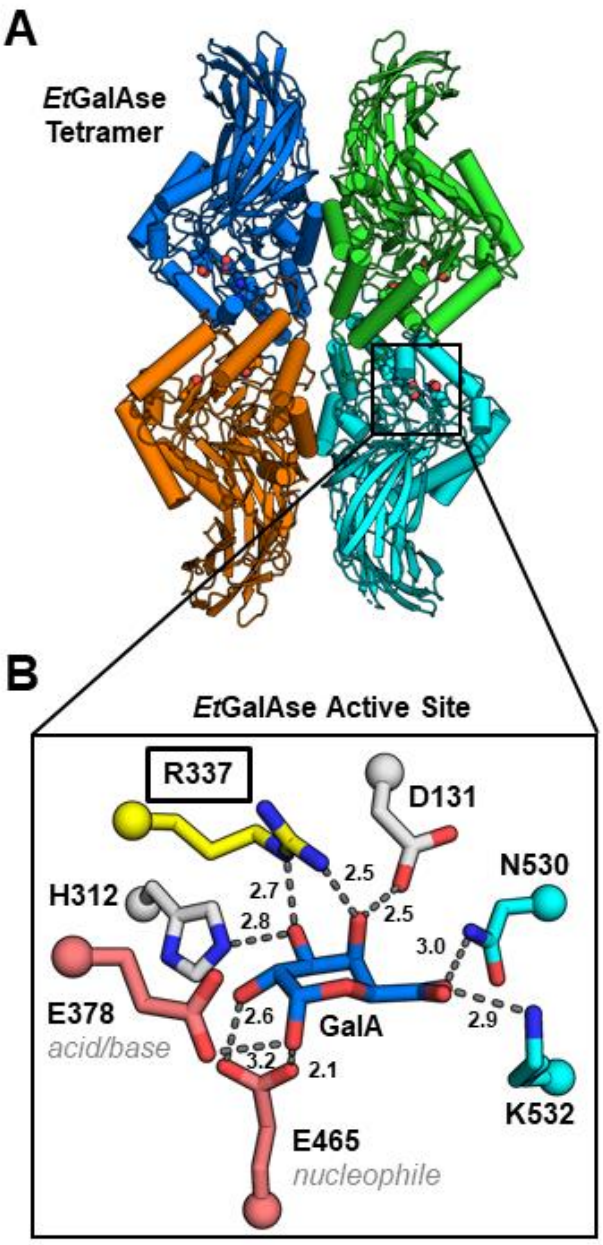


Figure 3.5 - Crystal structure of the GH2 β-galacturonidase *Et*GalAse. (**A**) Quaternary structure of *Et*GalAse shown in cartoon style with active site residues shown as spheres. (**B**) Active site of *Et*GalAse bound to GalA (blue) with catalytic glutamates shown in deep salmon, NxK motif in cyan, and the unique R337 in yellow. All distances shown are in units of angstroms.

Together, these structural and functional data pinpoint Arg-337 as the key active site feature that differentiates a GH2 GalAse from a GH2 GUS.

Structure-guided analysis of putative GUSs from the HMP reveals 12 additional putative GH2 GalAses

With structural and functional data describing key features of *Et*GalAse in-hand, we then screened the 278 additional GUS sequences from the HMP stool sample database to determine if other enzymes previously annotated as GUSs may instead be GH2 GalAses. We performed sequence alignments with putative GUS sequences from the HMP against *Et*GalAse and selected sequences that contained an arginine at the equivalent sequence position. Structure-guided sequence alignments revealed 12 additional putative GalAses out of the 279 previously annotated GUS enzymes (**Figure 3.6**, **Table 3.5**). Together with *Et*GalAse, these 13 putative GalAses represent 4.7% of the total GUS proteins originally identified. Generation of a set of sequence similarity networks using the 279 putative GUS sequences from the HMP showed that the putative GalAses cluster together at an alignment score of 10^{-150}, suggesting they share a common function **(Figure 3.6** and **Figure 3.7**)[129]. This is a lower alignment score (less stringent) than utilized previously to create an SSN of the human gut microbial GUSome (10^{-220}; **Figure 3.1**)[132]. While this lower alignment (10^{-150}) score yielded SSNs that did not differentiate many of the putative GUS enzymes as effectively, it did group the putative GalAses together that were mostly singletons at an alignment score of 10^{-220}, suggesting they are likely isofunctional (**Figure 3.6**, **Figure 3.1**).

To experimentally validate our structure- and function-guided GalAse discovery, we characterized a putative GalAse from *Faecalibacterium sp. CAG:74_58_120* (*Fc*GalAse, UniProtKB: A0A1Q6Q230). *Fc*GalAse hydrolyzed *p*NP-GalA efficiently (k_{cat} = 42 ± 3 s^{-1}, K_M

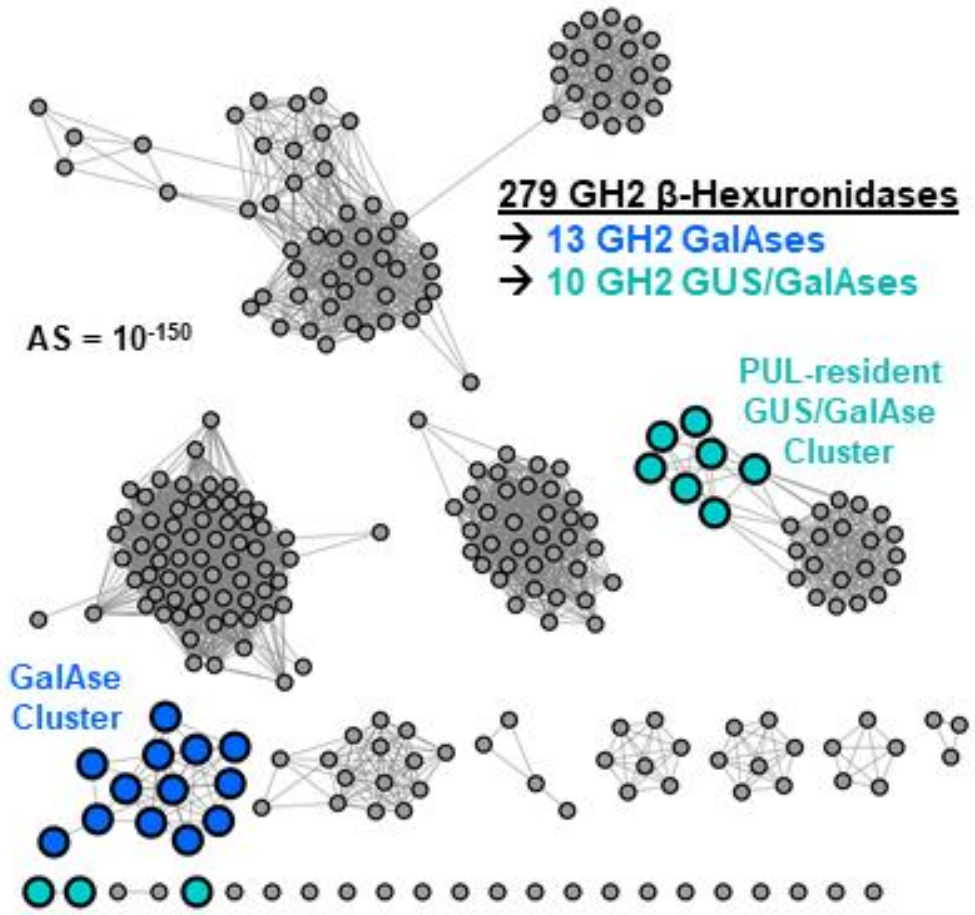


Figure 3.6 - Sequence similarity network (SSN) generated from putative GUS sequences from the HMP stool sample metagenomic database reveals distinct clusters for the 13 putative GH2 GalAses and 10 putative hybrid GUS/GalAses. SSN of the 279 putative GUS enzymes identified in the HMP stool sample database with the GH2 GalAse cluster proteins colored blue and hybrid GUS/GalAses in cyan. The SSN was generated with the EFI-EST web tool with an alignment score of 10^{-150}.

Table 3.5 - Organism and sequence ID for 13 putative GH2 GalAses identified from the HMP 'GUSome'.

Protein	Organism	UniProtKB ID
*Et*GalAse	*Eisenbergiella tayi*	A0A1E3AEY6
*Cs*GalAse	*Clostridium sp. CAG:253*	R6LZJ1
*Ls*GalAse	*Lachnospiraceae bacterium TF01-11*	A0A0Q1BGM3
*Bv*GalAse	*Butyrivibrio sp. CAG:318*	R6QJ22
GalAse-1	Unknown	NA[a]
GalAse-2	Unknown	NA[a]
*La*GalAse	*uncultured Lachnospira sp.*	A0A1C6CM03
*Cs*GalAse-2	*Clostridium sp. 42_12*	A0A1Q6K944
*Fs*GalAse	*Fusicatenibacter sp. 2789STDY5834925*	A0A174X2P5
*Ls*GalAse-2	*Lachnospiraceae bacterium*	A0A358RC24
*Rs*GalAse	*Roseburia sp. CAG:10041_57*	A0A1Q6SLW7
GalAse-3	Unknown	NA[b]
*Fc*GalAse	*Faecalibacterium sp. CAG:74_58_120*	A0A1Q6Q230

[a]GalAse-1 and GalAse-2 share the most identical sequence from *Clostridiales bacterium* KLE1615 (UniProtKB ID: A0A176UAH1) with 87.4% and 82.3% sequence identity, respectively.

[b]*GalAse*-3 is most identical to *Firmicutes bacterium AM55-24TS* (UniProtKB ID: A0A373QB78) with 60% sequence identity.

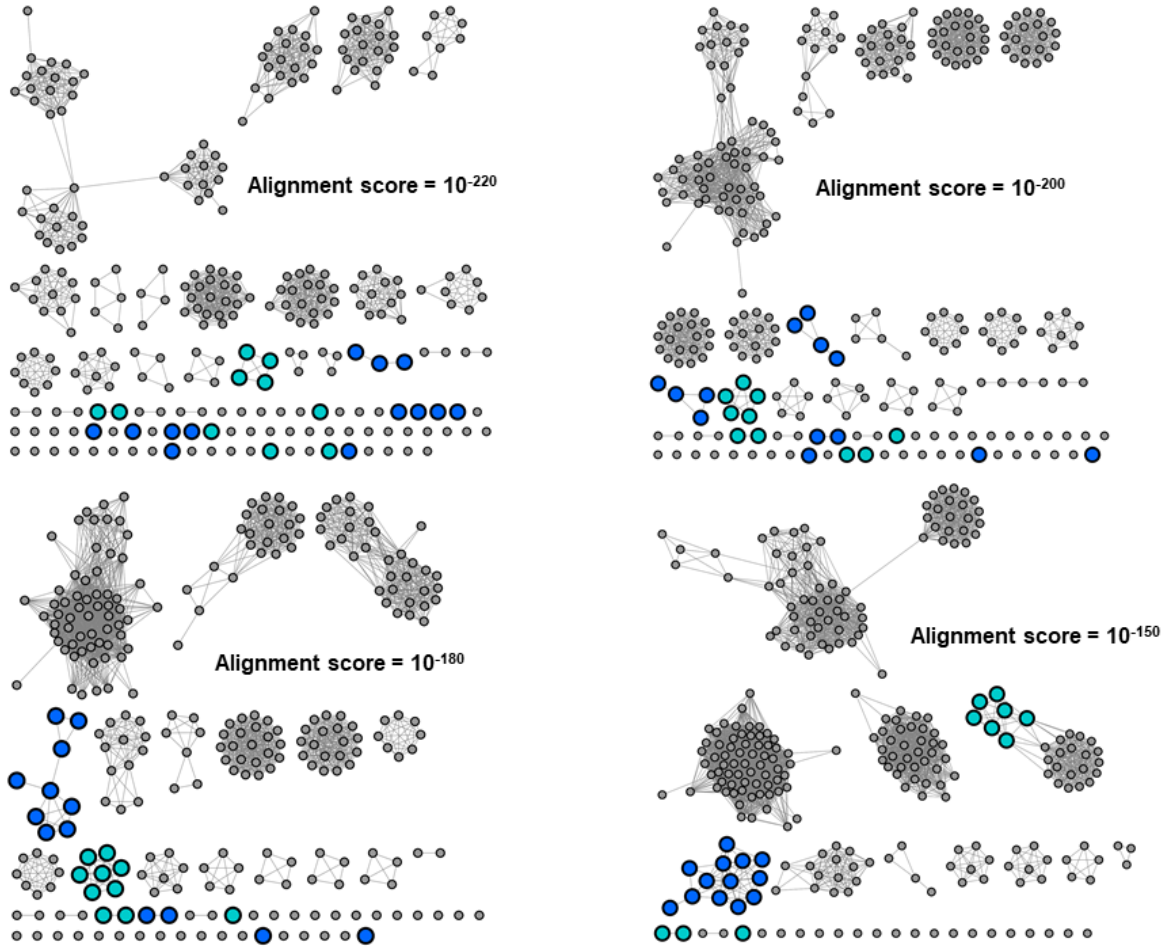


Figure 3.7. Sequence similarity networks of HMP 'GUSome' at different alignment scores with putative GalAses highlighted in blue and GUS/GalAses in cyan.

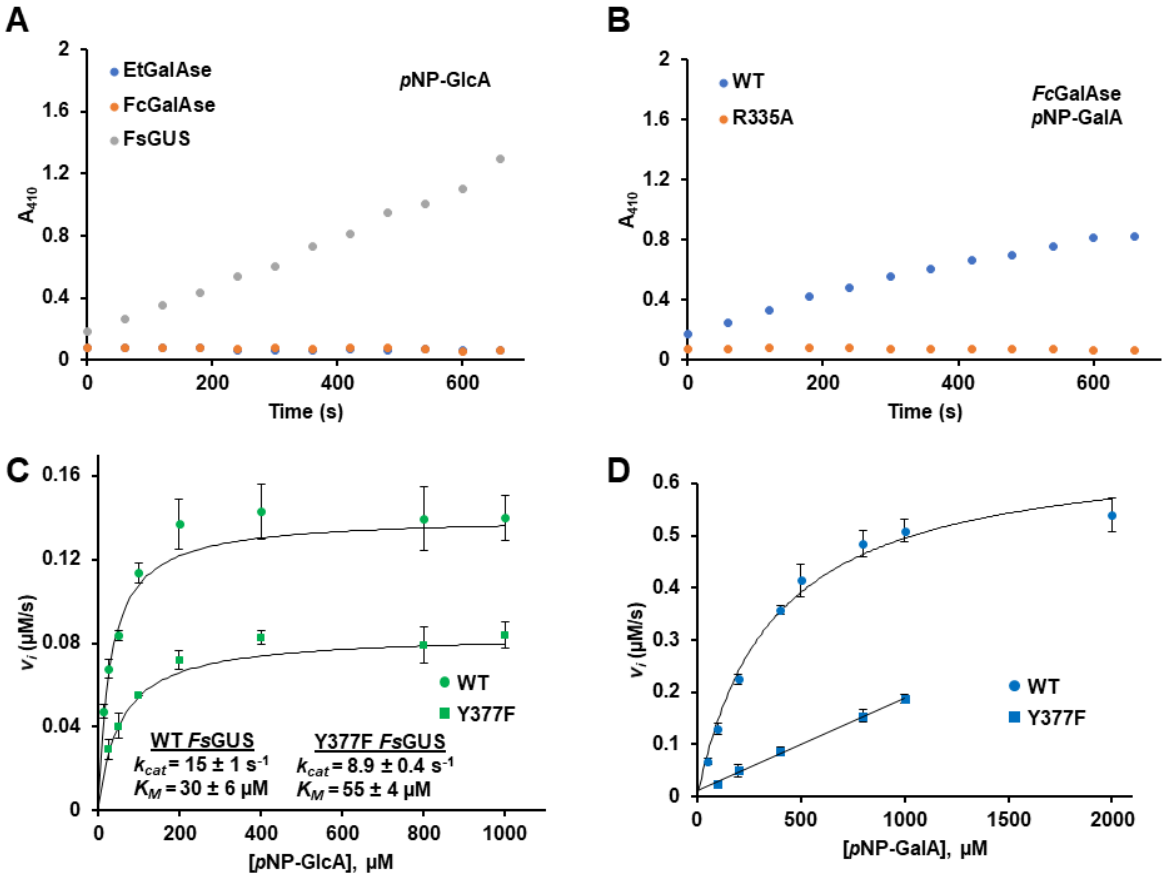


Figure 3.8 - Progress curves and Michaelis-Menten plots for *Et*GalAse, *Fc*GalAse, and *Fs*GUS. (**A**) Progress curves for *Et*GalAse, *Fc*GalAse, and *Fs*GUS with *p*NP-GlcA. (**B**) Progress curves of WT and R335A mutant of *Fc*GalAse with *p*NP-GalA. (**C**) Michaelis-Menten plots of WT and Y377F mutant of *Fs*GUS with *p*NP-GlcA. (**D**) Michaelis-Menten plot of *Fs*GUS WT and Y377F mutant with *p*NP-GalA.

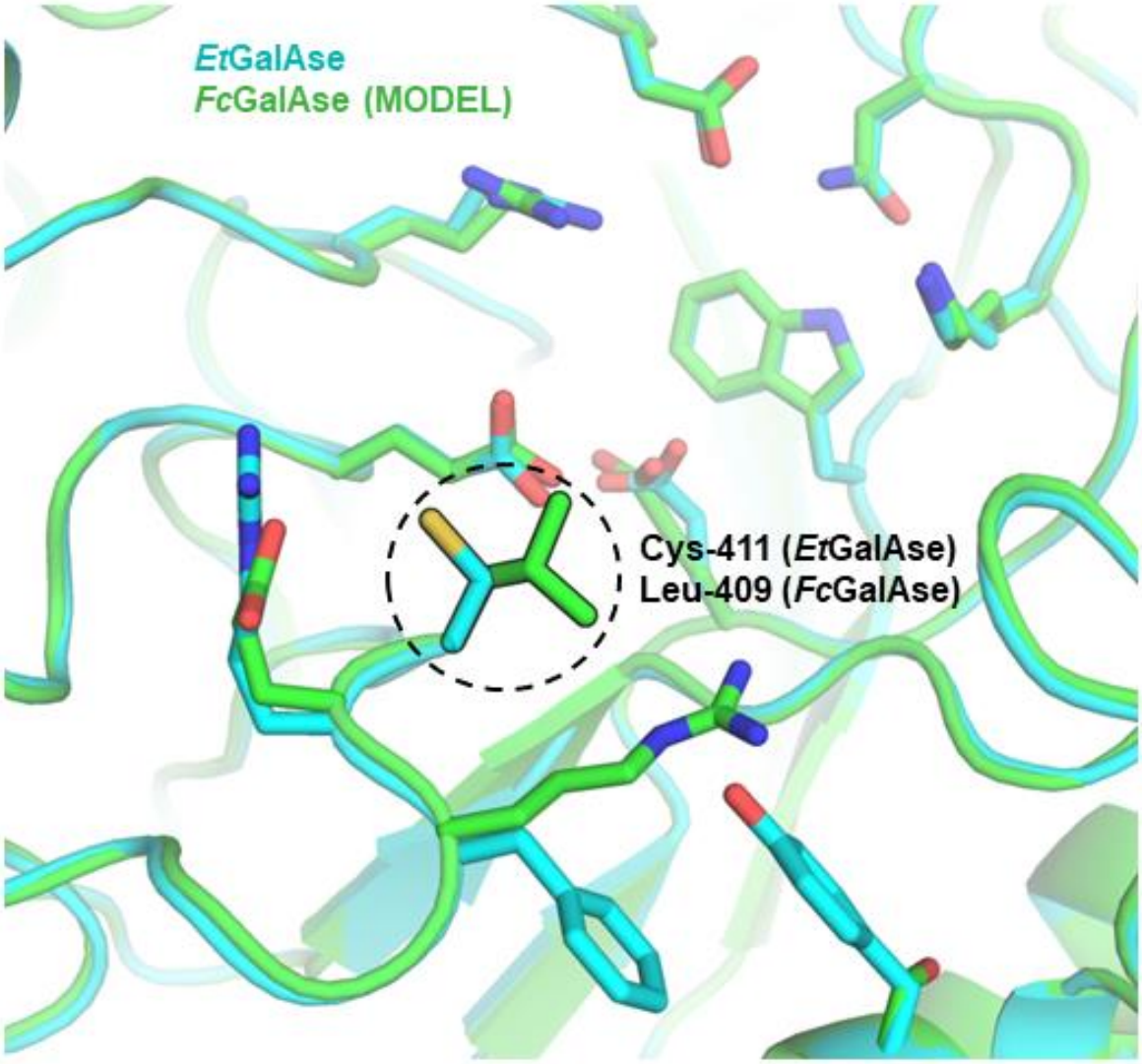


Figure 3.9 - Structural overlay of *Fc*GalAse model and *Et*GalAse crystal structure reveal active site difference that may impact arginine mutant activity.

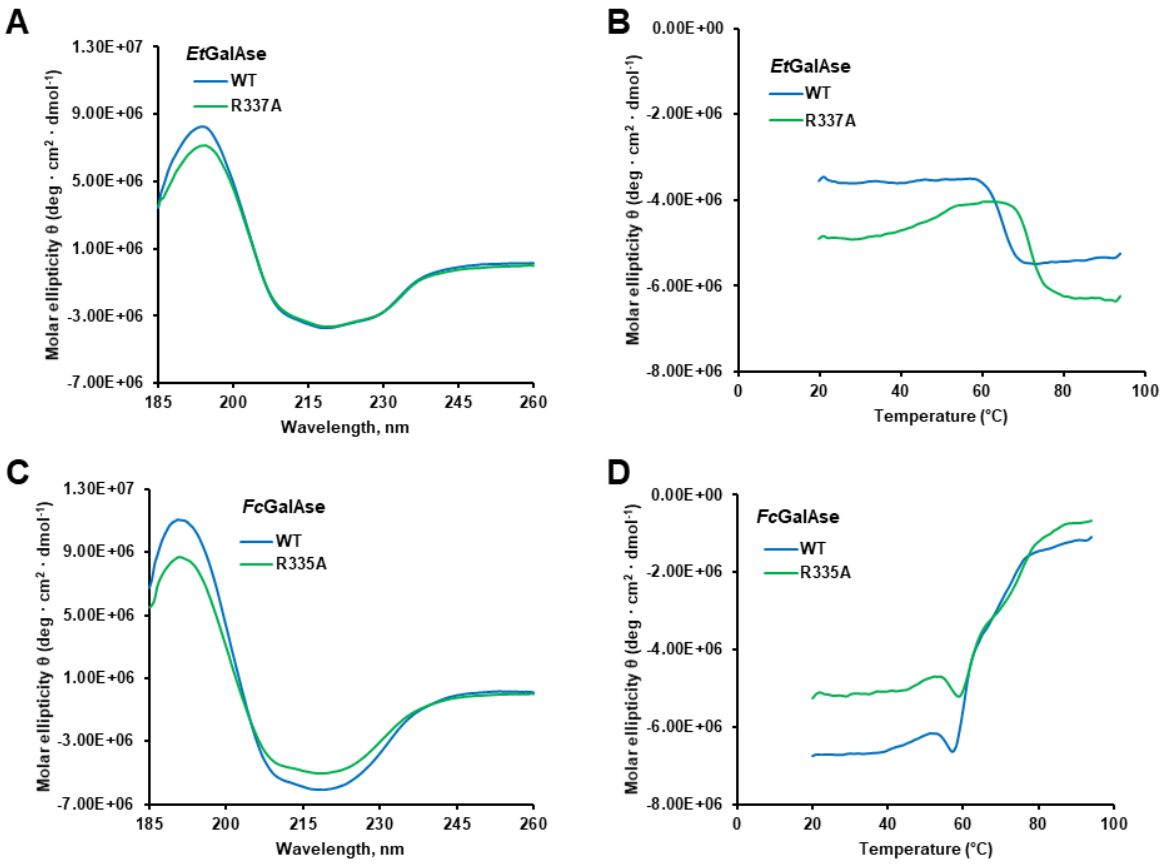


Figure 3.10 - Circular dichroism spectra and melting temperatures of WT and mutant proteins of *Et*GalAse and *Fc*GalAse.

= 58.8 ± 0.2 µM, k_{cat}/K_M = 7.1x10^5 M^{-1} s^{-1}; **Table 3.4**, **Figure 3.3C**, and **Figure 3.4D**) and was unable to process *p*NP-GlcA (**Figure 3.8A**), validating our bioinformatic identification of GH2 GalAses. The R335A mutant of *Fc*GalAse abolished galacturonide processing; however, unlike *Et*GalAse, this single residue mutation did not convert *Fc*GalAse into a GUS (**Table 3.4**). A structural model of *Fc*GalAse derived from the *Et*GalAse structure suggests that the absence of GUS activity for the R335A variant of *Fc*GalAse may be due to the presence of Leu-409 in *Fc*GalAse instead of Cys-411 in *Et*GalAse, which may contribute to a less stable active site in the absence of Arg-335 (**Figure 3.9**). Further structural analysis by circular dichroism (CD) showed that the R335A mutant of *Fc*GalAse displays reduced secondary structure in comparison to WT protein (**Figure 3.10C**). No difference was observed in secondary structure for WT and R337A *Et*GalAse (**Figure 3.10A**). WT *Fc*GalAse and the R335A mutant do not have significantly different melting temperatures, but the *Et*GalAse proteins do not appear to melt at all (**Figure 3.10B, D**). While future studies will be needed to pinpoint the molecular details that enable some GalAses to be converted into GUSs, CD analysis suggests that structural differences between *Fc*GalAse and *Et*GalAse may mediate these functional differences.

Structure and function a hexameric hybrid GUS/GalAse from *Fusicatenibacter saccharivorans*

Previous analysis of a GUS from *Bacteroides uniformis*, *Bu*GUS-1, revealed that it processes both glucuronides and galacturonides, and that it possessed a distinct tyrosine and tryptophan residue in its active site (YW motif) in comparison to previously characterized GUS enzymes[126]. Using the YW motif as a sequence guide, we identified 9 additional hybrid GUS/GalAses from the 279 putative GUS sequences from the HMP stool sample database, 6

Table 3.6 - Organism and sequence ID for 10 putative GH2 GUS/GalAses identified from the HMP 'GUSome'.

Protein	Organism	UniProtKB ID
GUS/GalAse-1	Unknown	NA[a]
*At*GUS/GalAse-2	*Alistipes timonensis JC136*	A0A1H3ZQ67
*Bu*GUS/GalAse	*Bacteroides uniformis dnLKV2*	R9I4I0
*Bu*GUS/GalAse-2	*Bacteroides sp. AF25-5LB*	A0A3A5YFK1
GUS/GalAse-2	Unknown	NA[b]
*Ba*GUS/GalAse-2	*uncultured Bacteroides sp.*	A0A1C5LVQ0
*At*GUS/GalAse-3	*Alistipes timonensis JC136*	A0A1H4BBT7
*Cs*GUS/GalAse	*Coprobacter secundus*	A0A0B2JHJ5
*Fs*GUS/GalAse	*Fusicatenibacter saccharivorans*	A0A174EHD1
GUS/GalAse-3	Unknown	NA[c]

[a]GUS/GalAse-1 is most identical to a protein from *Alistipes timonensis JC136* (UniProtKB ID: A0A1H3ZQ67) with 87.1% sequence identity.

[b]GUS/*GalAse*-2 is most identical to a protein from *Bacteroides sp. D20* (UniProtKB ID: D2EWN8) with 78.7% sequence identity.

[c]GUS/GalAse-3 is most identical to a protein from *Eubacterium sp. CAG:786* (UniProtKB ID: R5GKW6) with 75.2% sequence identity.

of which had been previously annotated as containing the N-terminal loop (NTL) discovered in *Bu*GUS-1 from *Bacteroides uniformis* (**Figure 3.6** and **Table 3.6**)[126]. We selected the putative GUS/GalAse from *Fusicatenibacter saccharivorans* (*Fs*GUS, UniProtKB: A0A174EHD1) for further characterization. Like *Bu*GUS-1, *Fs*GUS processed both *p*NP-GlcA (k_{cat} = 15 ± 1 s^{-1}, K_M = 30 ± 6 µM, k_{cat}/K_M = 5.0x10^5 M^{-1} s^{-1}) and *p*NP-GalA (k_{cat} = 34 ± 2 s^{-1}, K_M = 360 ± 20 µM, k_{cat}/K_M = 9.4x10^4 M^{-1} s^{-1}, **Table 3.4**, **Figure 3.8C**, and **Figure 3.8D**), validating our bioinformatic search.

We next determined the crystal structure of the apo and phenyl-thio-β-D-glucuronide (PTG)-bound structures of *Fs*GUS to 2.05 Å and 2.20 Å resolution, respectively (**Table 3.3**). The crystal structure of *Fs*GUS revealed a unique hexamer among characterized GUS enzymes (**Figure 3.13A**), which we validated in solution by size exclusion chromatography-multi-angle light scattering (SEC-MALS) (**Figure 3.11**). The PTG-bound structure of *Fs*GUS reveals a shift in the catalytic acid/base Glu-416, presumably to avoid a steric clash with the thioether linkage in PTG (**Figure 3.12A**). The *Fs*GUS active site also reveals Tyr-377 and Trp-378 previously identified in *Bu*GUS-1 (**Figure 3.13B**). Tyr-377 is key in recognizing both GlcA and GalA because it can hydrogen bond to the 3-hydroxyl of GlcA (**Figure 3.13B**) and can also hydrogen bond to the axial 4-hydroxyl of GalA when docked in the *Fs*GUS active site (**Figure 3.14**). Tyr-377 in *Fs*GUS occupies nearly the same position as Arg-337 in *Et*GalAse; thus, we were curious if it aided hydrolysis of glucuronides and galacturonides. Indeed, mutation of Tyr-377 to phenylalanine in *Fs*GUS results in a ten-fold lower catalytic efficiency for *p*NP-GalA (k_{cat}/K_M = 8.9x10^3 M^{-1} s^{-1}) and 3-fold lower catalytic efficiency for *p*NP-GlcA (k_{cat}/K_M = 1.6x10^5 M^{-1} s^{-1}) compared to WT *Fs*GUS (*p*NP-GalA: k_{cat}/K_M = 9.4x10^4 M^{-1} s^{-1}, *p*NP-GlcA: k_{cat}/K_M = 5.0x10^5 M^{-1} s^{-1}) (**Table 3.4**, **Figure 3.8C**, **Figure 3.8D,** and **Figure 3.15**). The

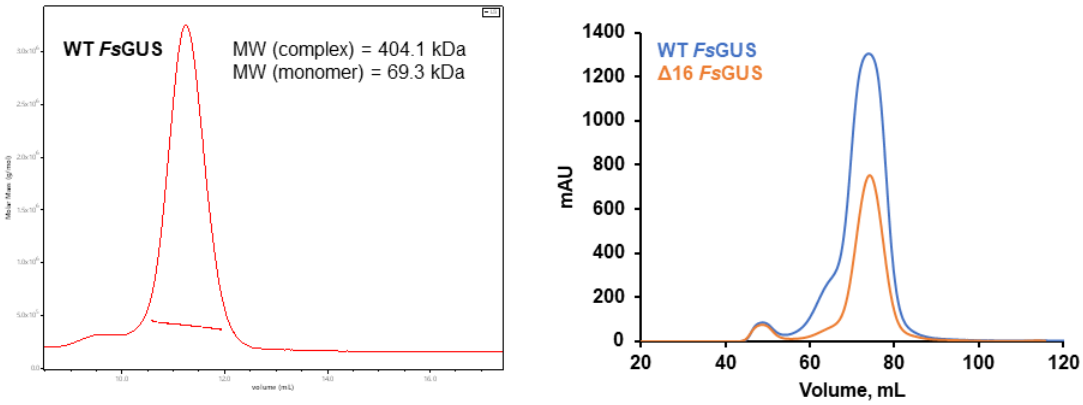


Figure 3.11 - SEC-MALS trace of purified WT *Fs*GUS (left) and SEC traces of WT and Δ16 *Fs*GUS (right).

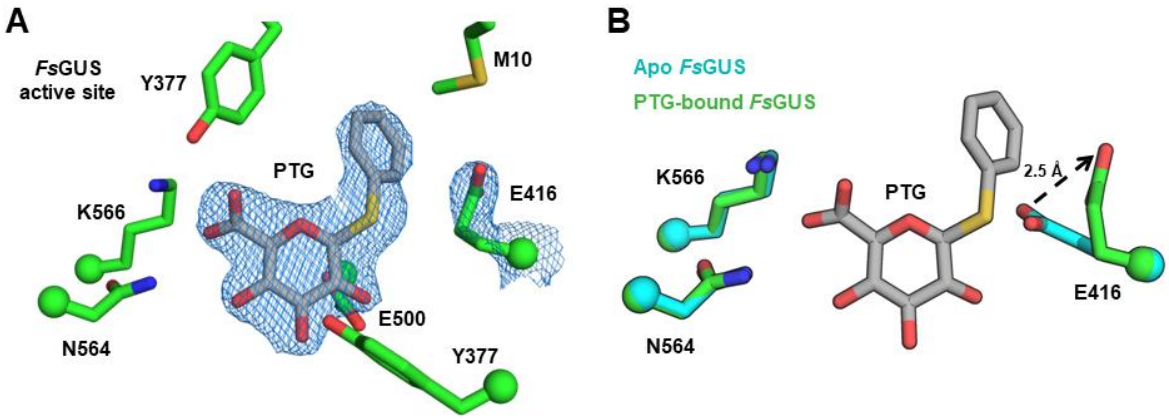


Figure 3.12 - Active site structure of PTG-bound *Fs*GUS. (**A**) PTG bound to the *Fs*GUS active site with 2Fo-Fc density shown at 1.5 σ for the ligand and acid/base Glu-416. (**B**) Overlay of the apo (cyan) and PTG-bound (green) *Fs*GUS active site reveals the movement of the acid/base Glu-416.

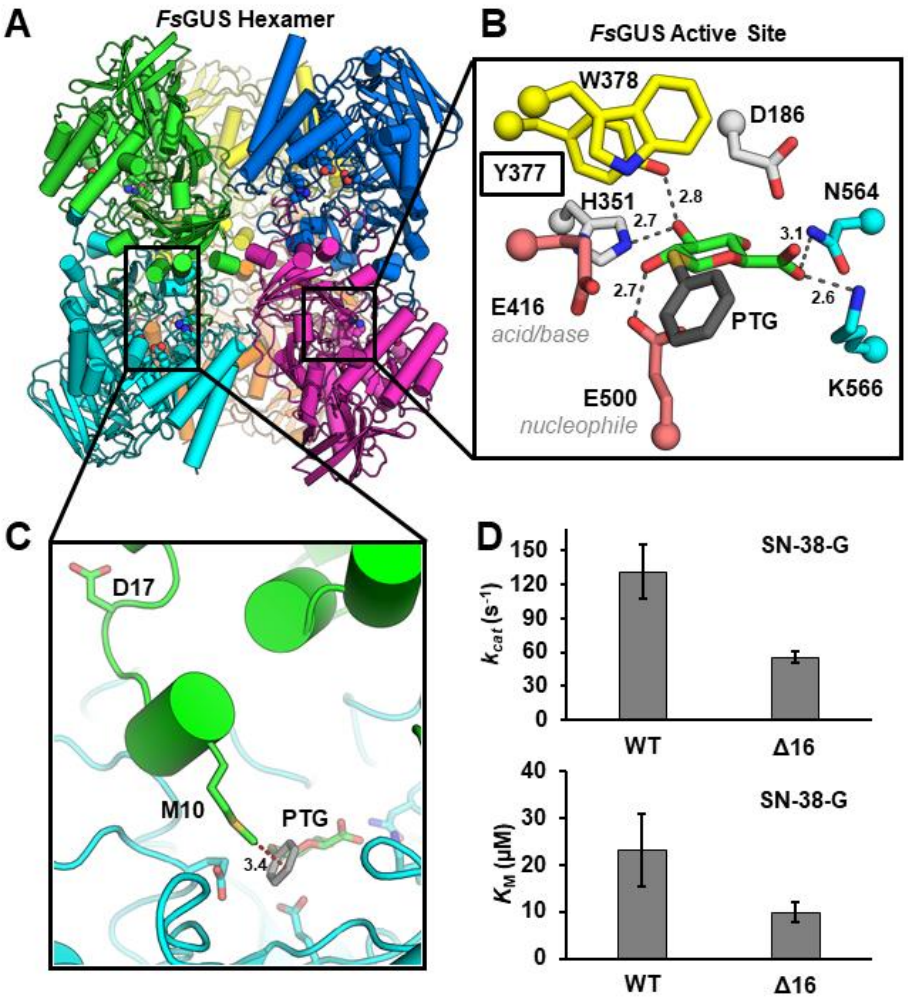


Figure 3.13 - Structure and function of *Fs*GUS reveal Y377 and N-terminus as important structural features for substrate processing. (**A**) The *Fs*GUS hexamer with each chain in a distinct color and active site residues shown as spheres. (**B**) Zoom-in of active site of *Fs*GUS bound to phenyl-thio-β-D-glucuronide (PTG, phenyl in dark grey, GlcA in green) with catalytic glutamates colored deep salmon, NxK motif in cyan, and unique YW motif in yellow. All distances shown are in units of angstroms. (**C**) Active site of *Fs*GUS with PTG bound and N-terminus of adjacent protomer shown in green. (**D**) Catalytic turnovers and Michaelis constants for WT and Δ16 mutant of *Fs*GUS with SN-38-G. Values represent an average of three replicates ± SD.

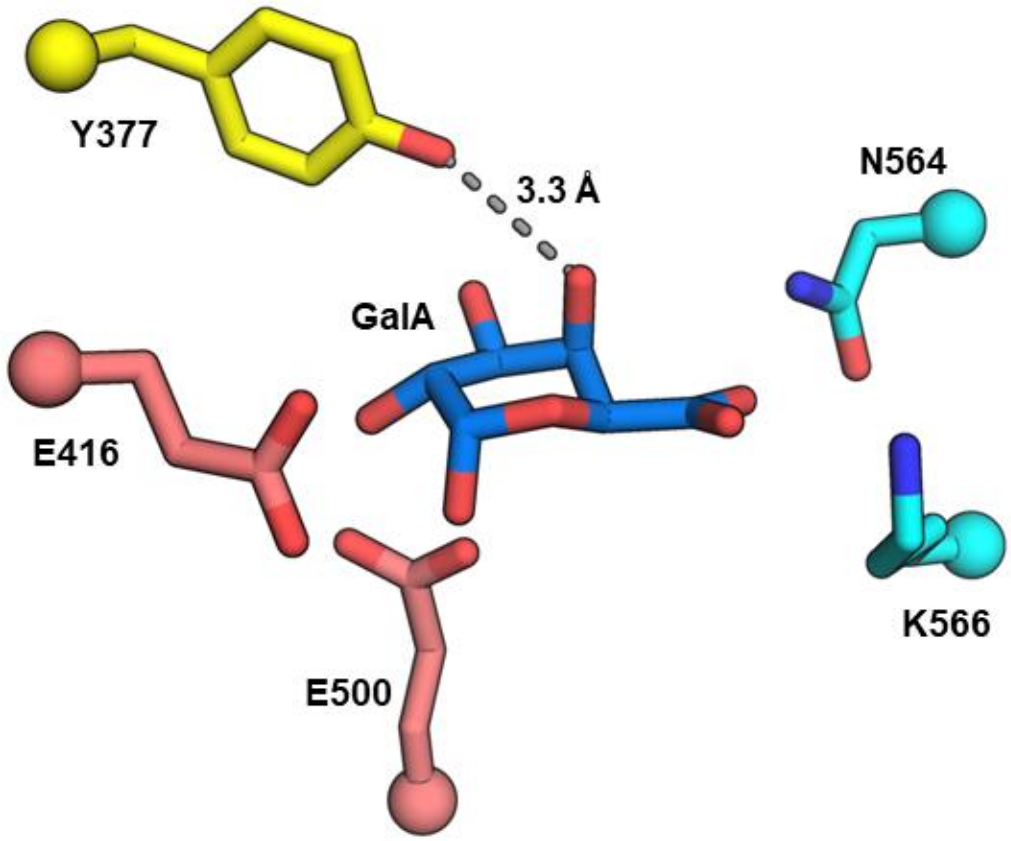


Figure 3.14 - GalA docked in the active site of *Fs*GUS.

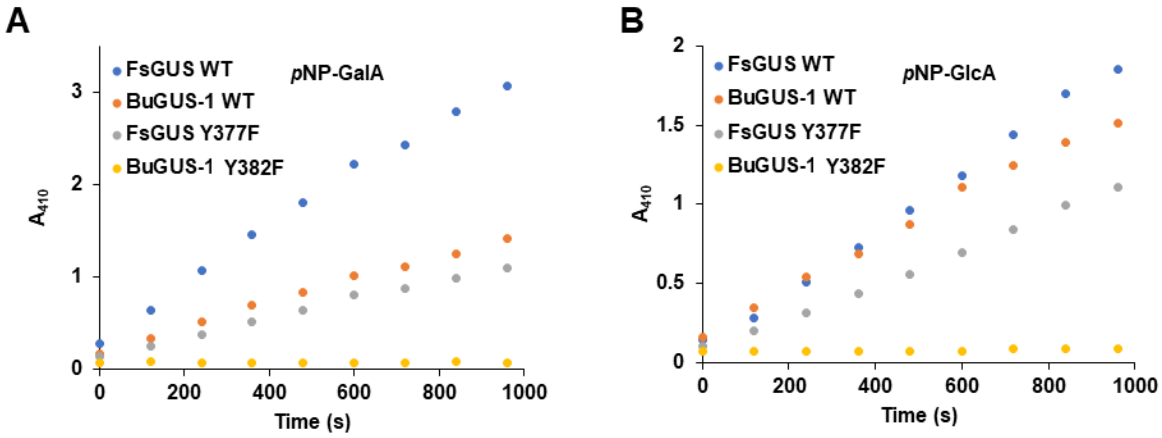


Figure 3.15 - Progress curves of *Fs*GUS, *Bu*GUS-1, and mutants with (A) *p*NP-GalA and (B) *p*NP-GlcA.

greater loss of catalytic efficiency in GalAse function upon mutation of Tyr-377 suggests that it is particularly important for the processing of GalA-containing substrates. Furthermore, we performed the analogous Y382F mutation in the previously characterized *Bu*GUS-1, which resulted in a complete loss of activity against both *p*NP-GlcA and *p*NP-GalA (**Figure 3.15**). Together, this structure-function analysis of *Fs*GUS reveals the first hexameric GUS and identifies Tyr-377 as an important residue for efficient processing of both glucuronides and galacturonides.

The N-termini of adjacent protomers form the aglycone binding sites of the FsGUS hexamer

Further inspection of the *Fs*GUS hexamer revealed that the N-termini of adjacent protomers swap into the active site of each monomer (**Figure 3.13C**). For example, in the PTG-bound structure of *Fs*GUS, Met-10 from an adjacent protomer (green) is 3.4 Å from the phenyl ring of PTG, suggesting that it may be important to substrate recognition of the aglycone moieties of potential glucuronide or galacturonide substrates (**Figure 3.13C**). We performed mutagenesis to remove the first 16 residues (Δ16) of *Fs*GUS to assess the role the N-terminus plays in substrate processing (**Figure 3.13C**). The Δ16 *Fs*GUS mutant elutes at the same time as WT *Fs*GUS by size-exclusion chromatography, suggesting that this mutant occupies the same hexamer as WT *Fs*GUS (**Figure 3.11**). Kinetic analysis of Δ16 *Fs*GUS revealed a ten-fold and two-fold reduction in catalytic efficiency for *p*NP-GlcA and *p*NP-GalA, respectively (**Table 3.4**). Thus, it appears the N-terminus of *Fs*GUS plays a key role in efficient substrate processing.

Because *Fs*GUS was able to efficiently process the small-molecule glucuronide *p*NP-GlcA, we tested whether it could also process the therapeutically relevant glucuronide SN-38-

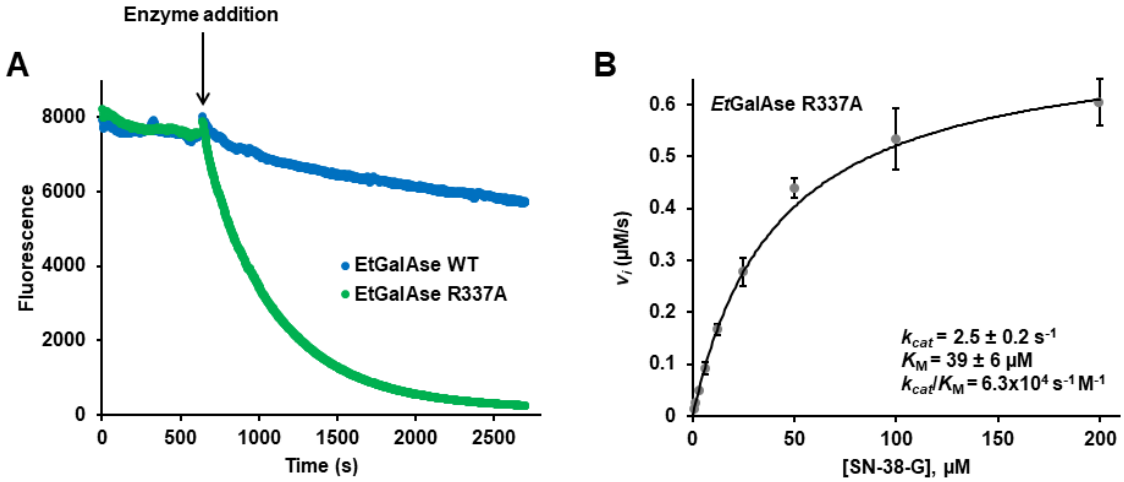


Figure 3.16 - SN-38-G processing by *Et*GalAse R337A mutant. (**A**) Progress curves for WT and R337A mutant of *Et*GalAse with SN-38-G. (**B**) Michaelis-Menten plot of SN-38-G processing by *Et*GalAse R337A.

G, the inactive metabolite of the anticancer drug irinotecan, and our original impetus for studying bacterial GUSs from the gut microbiome[123]. *Fs*GUS was able to process SN-38-G efficiently (k_{cat} = 130 ± 20 s^{-1}, K_M = 23 ± 8 μM, **Figure 3.13D**). We also tested the Δ16 mutant of *Fs*GUS with SN-38-G, which displayed a reduced k_{cat} and K_M (k_{cat} = 56 ± 5 s^{-1}, K_M = 10 ± 2 μM) in comparison to WT *Fs*GUS (**Figure 3.13D**). Lastly, we examined if the WT and R337A variant of *Et*GalAse could process SN-38-G. As expected, WT *Et*GalAse was unable to process SN-38-G efficiently but the R337A mutant displayed activity (k_{cat} = 2.3 ± 0.2 s^{-1}, K_M = 39 ± 6 μM, k_{cat}/K_M = 6.3x10^4 s^{-1} M^{-1}) (**Figure 3.16**). Given the unique structural features of *Fs*GUS, we performed a final analysis of the 279 putative GUS enzymes to determine the potential distribution and frequency of *Fs*GUS-like proteins in the gut microbiome. Unlike our search for putative GalAses or GUS/GalAses, which were predicated on a conserved arginine and tyrosine-tryptophan motif, the unique *Fs*GUS hexamer and N-terminus do not appear to be conserved sequence features. That is, we were unable to identify common sequence motifs that are responsible for these unique structural features. Thus, we performed pairwise sequence alignments with the complete *Fs*GUS sequence against the remaining 278 putative β-hexuronidases to identify potential *Fs*GUS-like proteins. Alignments revealed only one close relative to *Fs*GUS, a sequence with no clear match to a genome termed GUS/GalAse-3 (**Table 3.6**) that shares approximately 68% sequence identity with *Fs*GUS. Based on this analysis, there only appear to be two unique *Fs*GUS-like proteins in the human gut microbiome.

DISCUSSION

The GalAses and GUS/GalAses characterized here reveal modifications on a shared active site structure to differentiate the epimers GalA and GlcA. The enzymes characterized here were originally discovered using two features thought to be specific for GUS activity, the

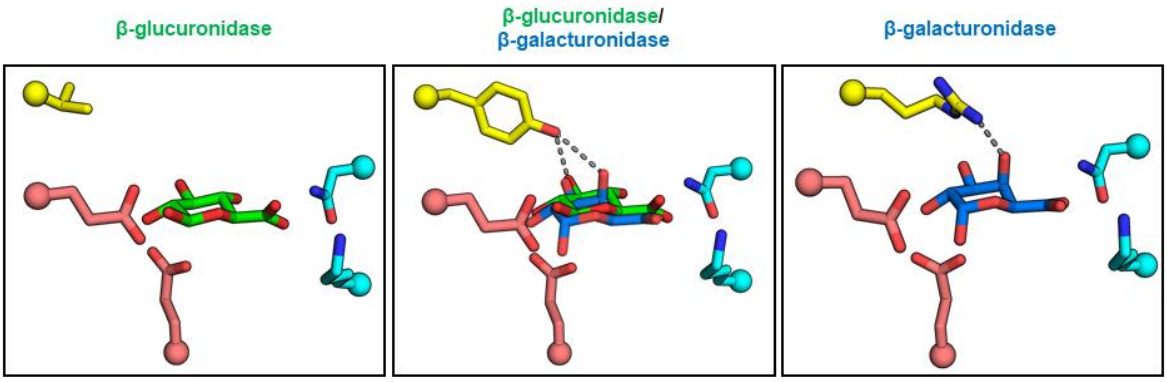


Figure 3.17 - Active site nuances differentiate gut microbial GUSs, hybrid GUS/GalAses, and GalAses. Representative active site structures of GUS (PDB: 4JKL [*Sa*GUS]), GUS/GalAse (*Fs*GUS), and GalAse (*Et*GalAse) with catalytic glutamates in deep salmon, NxK motif in cyan, key differentiating residues in yellow, GlcA in green, and GalA in blue.

catalytic glutamates and NxK motif[11]. We show here that this structure- and function-guided bioinformatic analysis yielded β-hexuronidases in general, which modify this active site template to yield similar, yet unique activities. For example, the GalAses and GUS/GalAses characterized here also utilize the catalytic glutamates and NxK motif, but the remaining active site residues recognize the other hydroxyls present in GlcA and GalA and are what differentiates GUS, hybrid GUS/GalAses, and GalAses (**Figure 3.17**). In bacterial GUS enzymes, the equatorial 4-hydroxyl of GlcA is recognized by an aspartic acid and a tryptophan[126]. The hybrid GUS/GalAses and GalAses characterized here, in which the 4-hydroxyl of their GalA substrate is axial, tyrosine (GUS/GalAse) or arginine (GalAse) residues are positioned above GalA and can hydrogen bond to the 4-hydroxyl (**Figure 3.17**). The active site architectures and 'rules' of selectivity observed could be used to discover β-hexuronidases in the gut microbiome, to design novel β-hexuronidases, and/or to change the function of existing β-hexuronidases.

The crystal structures presented here reveal active site and quaternary structures unique among characterized GUS enzymes. The tertiary structures of *Et*GalAse and *Fs*GUS exhibit the same GH2 fold observed for *E. coli* GUS (*Ec*GUS) and the other GUS enzymes of known structure, with a core TIM-barrel fold and two β-sandwich domains[56,100,123]. Structural alignments of *Et*GalAse and *Fs*GUS with *Ec*GUS reveal root mean square deviations of 2.6 Å over 520 Cα carbons and 2.9 Å over 528 Cα carbons, respectively (**Figure 3.18**). SEC-MALS shows that *Et*GalAse is a tetramer in solution (**Figure 3.19**), and the crystal structure reveals that the tetramer interface is mediated by the C-termini of individual protomers like that observed for *Ec*GUS (**Figure 3.20**). While *Fs*GUS maintains a similar tertiary structure to *Ec*GUS, it has a hexameric quaternary structure not observed in previously characterized GUSs

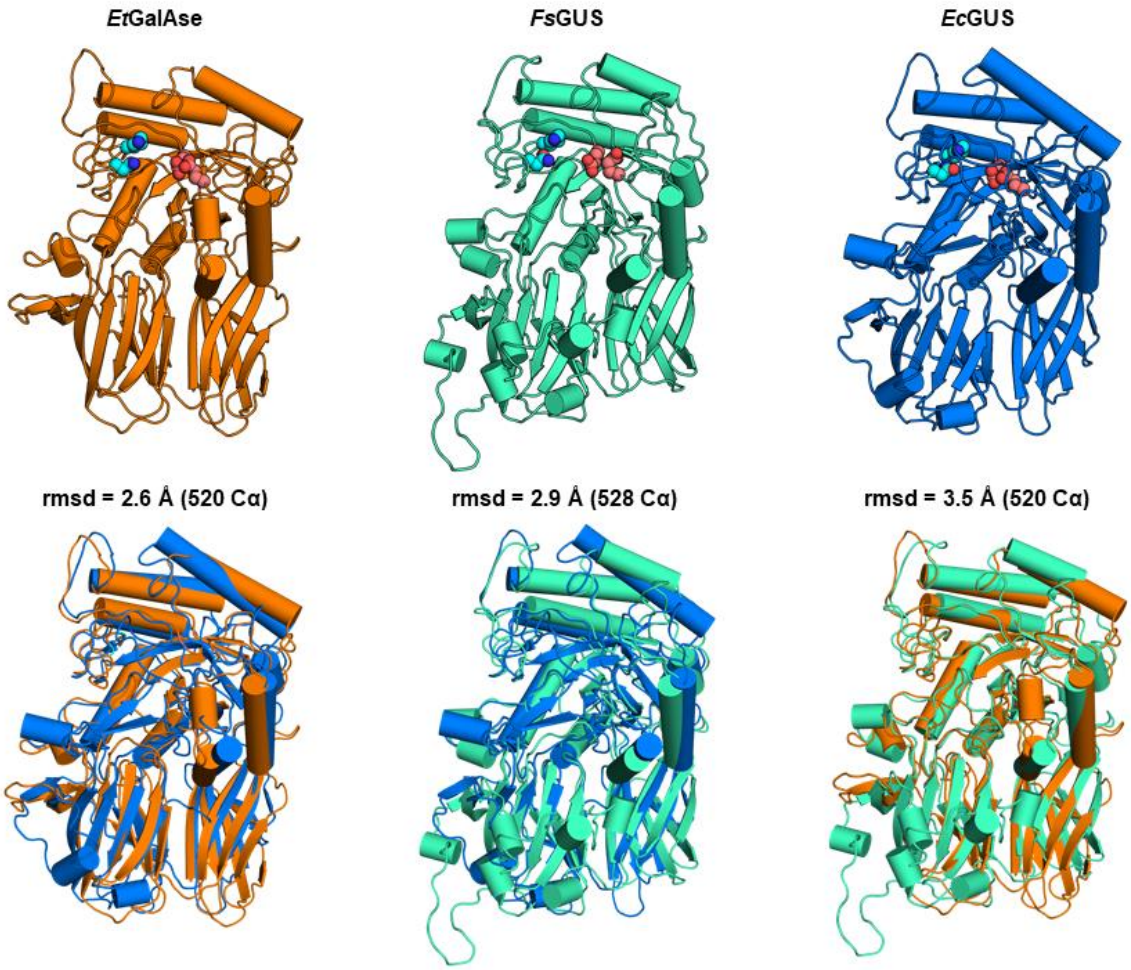


Figure 3.18 - Tertiary structures and structural alignments of *Et*GalAse, *Fs*GUS, and *Ec*GUS.

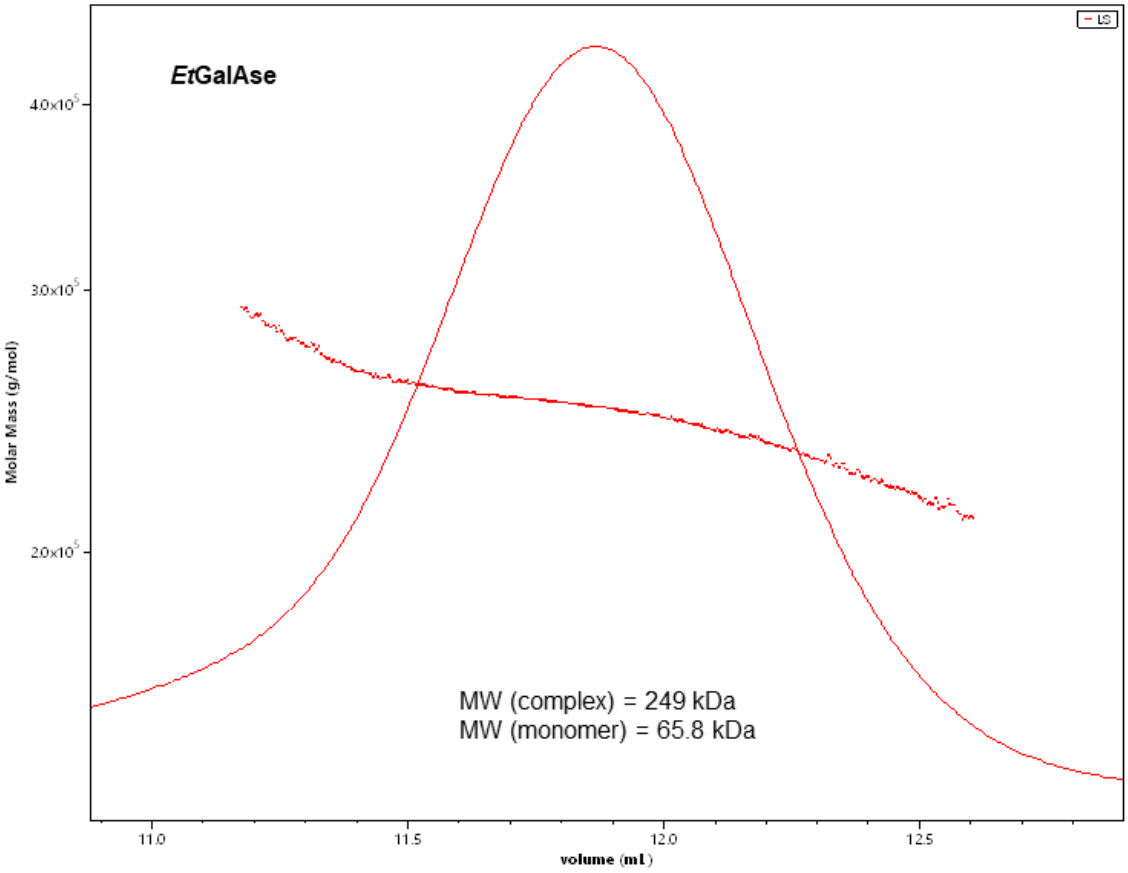


Figure 3.19 - SEC-MALS trace of purified *Et*GalAse.

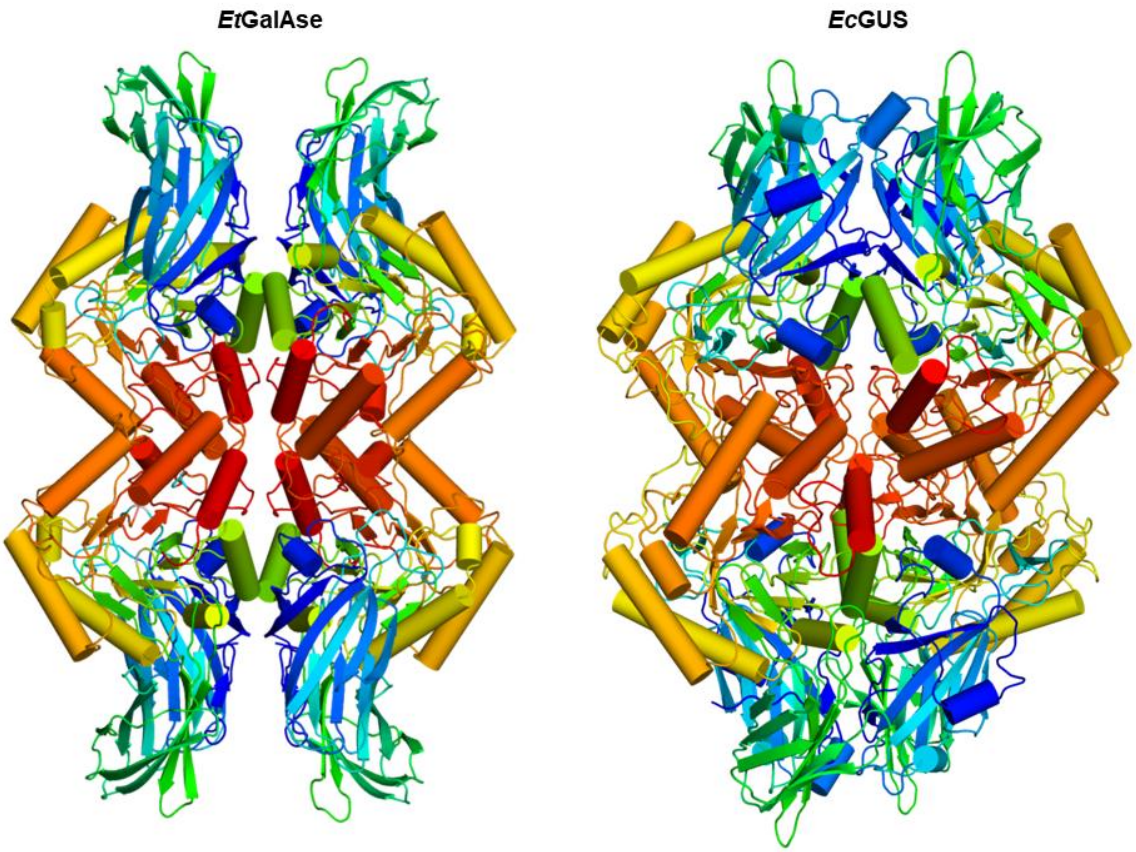


Figure 3.20. Comparison of *Et*GalAse and *Ec*GUS tetramer with each chain colored rainbow style.

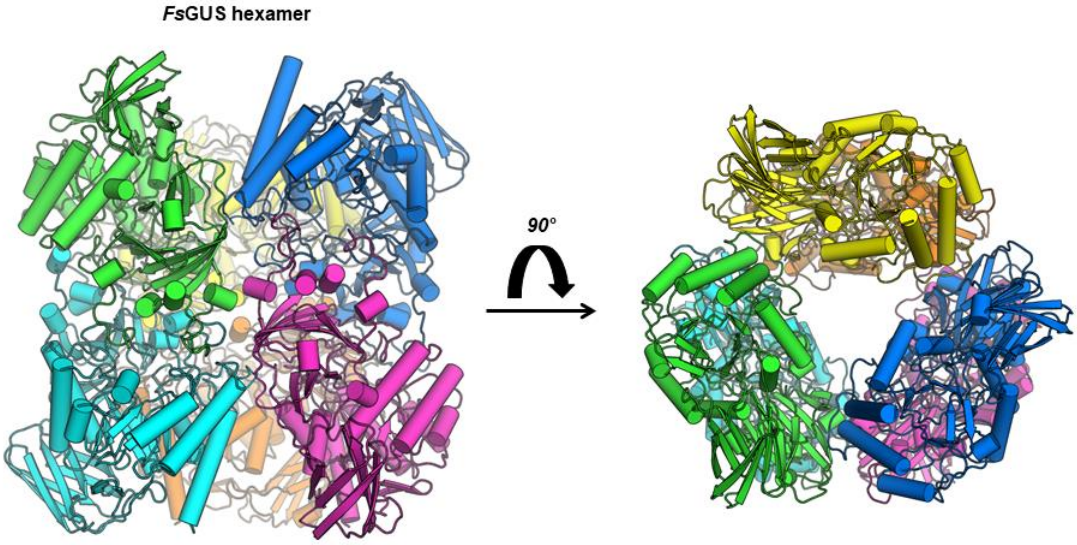


Figure 3.21 - Quaternary structure of *Fs*GUS.

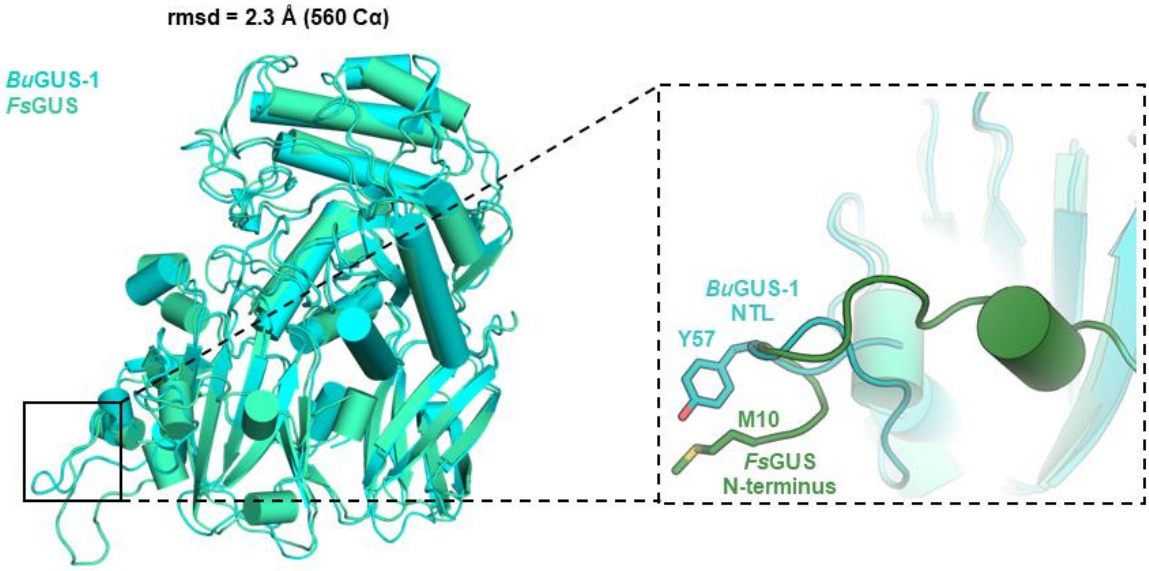


Figure 3.22 - Structural overlay of *Bu*GUS-1 and *Fs*GUS reveals distinct N-terminal structural elements that contribute to the formation of the aglycone binding site.

or any other GH2 enzymes to our knowledge (**Figure 3.13A** and **Figure 3.21**). As observed for other GUSs of distinct structures, the quaternary structure of this enzyme plays a key role in forming the active site and thus affects activity (**Figure 3.13D**). Like the NTL discovered in *Bu*GUS-1, the N-terminus of adjacent protomers in *Fs*GUS contribute to the formation of the aglycone binding site. An overlay of *Bu*GUS-1 and *Fs*GUS reveals that these enzymes utilize distinct structural elements to form the aglycone binding site (**Figure 3.22**). Thus, while *Bu*GUS-1 and *Fs*GUS both share the YW motif that appears to enable hybrid GUS/GalAse activity, they display distinct oligomeric states (*Bu*GUS-1 tetramer, *Fs*GUS hexamer) and utilize distinct protomer-donated structural elements to form the aglycone binding sites of their active sites (**Figure 3.22**). Taken together, quaternary structure is critical to understanding the function of bacterial GUS enzymes. The hybrid GUS/GalAses characterized here likely process glucuronate and galacturonate-containing polysaccharides in the gut. We performed further bioinformatic analysis of the GUS/GalAses identified from the HMP stool sample database using the EFI-GNT web tool[129]. This revealed that 7 of the 9 identified GUS/GalAses are embedded in polysaccharide utilization loci (PUL), gene clusters that coordinate the degradation of complex carbohydrates (**Figure 3.23**)[51]. These observations support the conclusion that the putative GUS/GalAses discovered here process polysaccharides that contain β-linked glucuronate or galacturonate moieties.

*Et*GalAse is functionally related to the previously characterized GH2 GalAse from *B. thetaiotaomicron*, BT_0992, which was shown to process the β-GalA linkage present in rhamnogalacturonan-II (RG-II)[133]. While *Et*GalAse and BT_0992 share 26% sequence identity (**Figure 3.24**), BT_0992 does not encode the conserved NxK motif and is much longer than the GH2 GalAses characterized here, suggesting that BT_0992 likely utilizes a distinct active

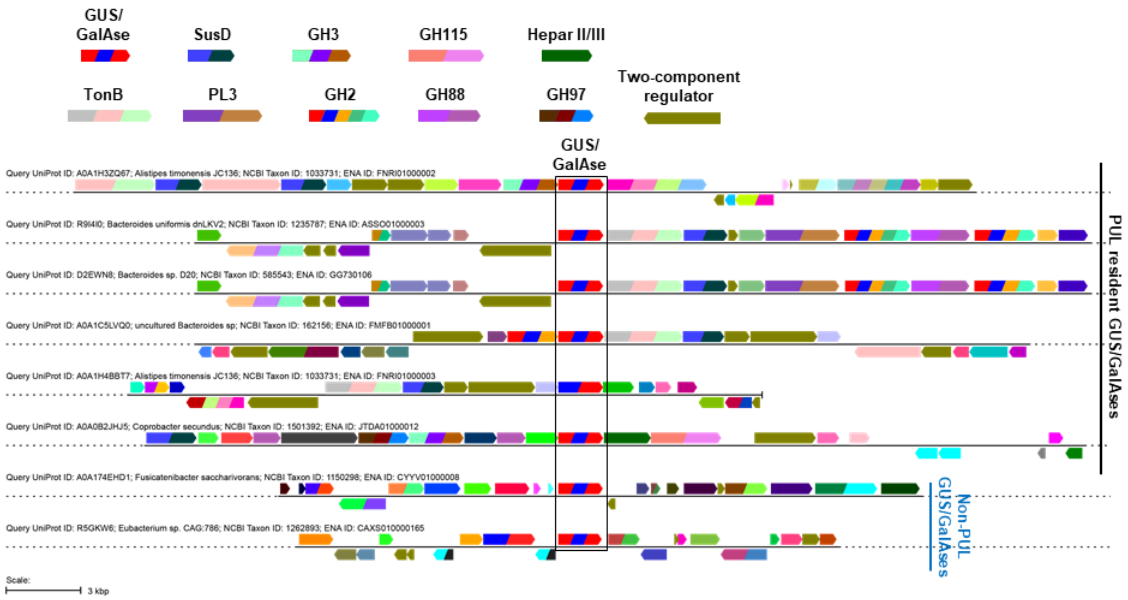


Figure 3.23 - Genome neighborhood diagrams of putative GUS/GalAses identified in the HMP stool sample database. GUS/GalAses are outlined in a black box, other potential coordinating enzymes are labeled in the legend.

Figure 3.24 - Portion of sequence alignment for BT_0992, FsGUS, EtGalAse, and FcGalAse with key residues highlighted.

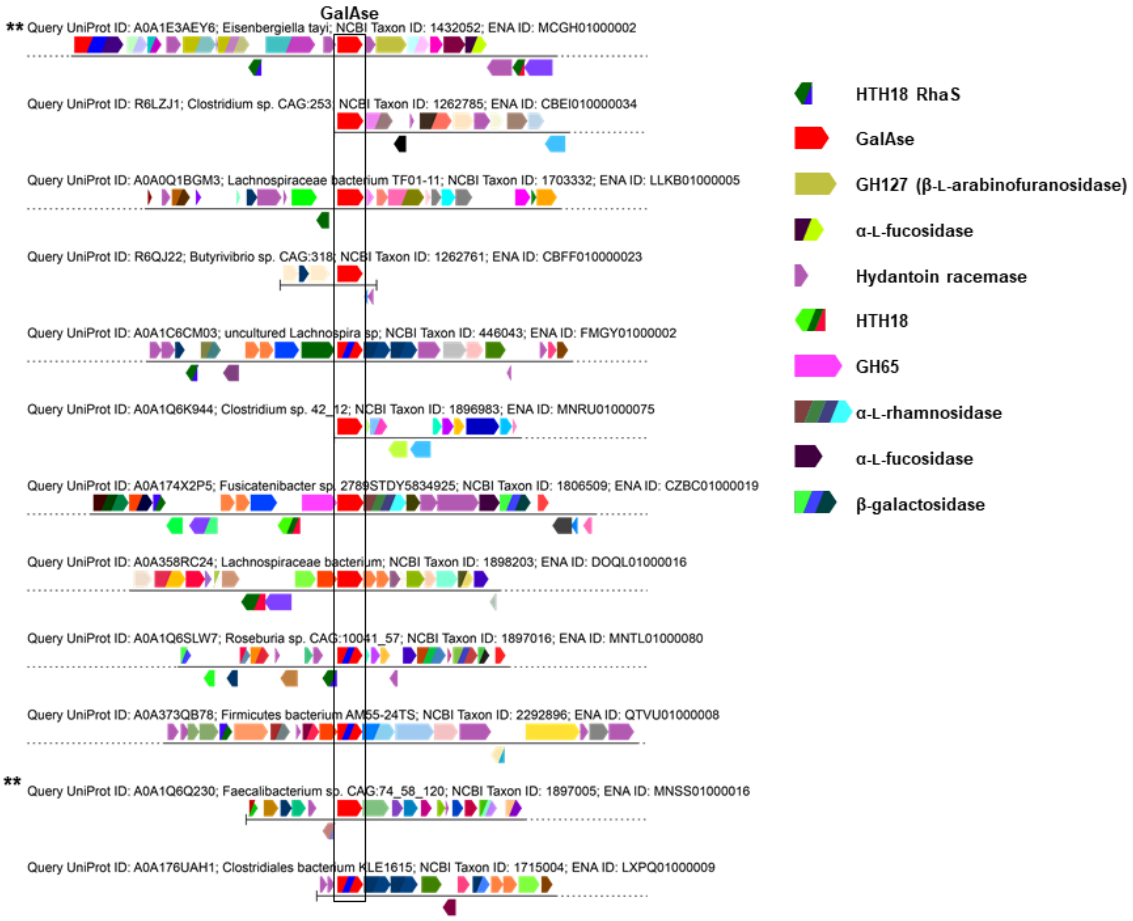


Figure 3.25. Genome neighborhood diagrams of putative GalAses. GalAses are outlined with a box, other potential coordinating genes are labeled in the legend, and asterisks denote the genome neighborhoods from which the GalAses were experimentally characterized.

site architecture to recognize and cleave β-GalA linkages. Analysis of the genome neighborhoods of the GH2 GalAses identified here revealed a variable genetic context. The genomic neighborhoods of the GH2 GalAses from *E. tayi* and *Fusicatenibacter sp. 2789STDY5834925A* contain carbohydrate active enzymes, including β-L-arbinofuranosidase, α-L-fucosidase, α-L-rhamnosidase, and β-galactosidase, but are not located in a polysaccharide utilization locus (**Figure 3.25**). While the linkages that these enzymes cleave are present in RG-II, it is not clear that RG-II is the native substrate of the GH2 GalAses identified in this study without extensive experimental validation. The genomic neighborhoods of the remaining 11 GH2 GalAses do not point to obvious potential substrates (**Figure 3.25**). Future studies on the gut microbial GH2 GalAses identified will be required to determine cognate substrates.

CONCLUSION

Here we show that gut bacterial GH2 enzymes utilize subtle active site changes to differentiate between the epimers glucuronate and galacturonate. We present the first structure of a GH2 GalAse and show that mutating a single residue in *Et*GalAse transforms it into a GUS. Using the structural and functional data from *Et*GalAse and the previously characterized *Bu*GUS-1, we discovered 12 additional GalAse genes and 9 additional hybrid GUS/GalAses in the previously defined gut bacterial GUSome. Through these efforts, we identified the molecular determinants that differentiate bacterial GUSs from GalAses

CHAPTER 4: DISCOVERY AND CHARACTERIZATION OF FMN-BINDING β-GLUCURONIDASES IN THE HUMAN GUT MICROBIOME.[4]

The human gut microbiome encodes about 5 million genes, outnumbering the human genome by 150-fold [134]. Among the millions of genes in the gut microbiota are those that encode carbohydrate active enzymes (CAZymes), which serve key roles in the metabolism of dietary and endogenous polysaccharides in the human gut [91]. Microbes utilize CAZymes to scavenge sugars from complex carbohydrates in the gut, and the metabolism of these sugars leads to the generation of short chain fatty acids that have been shown to play key roles in human physiology [135]. Thus, an understanding of the structure and function of these microbial enzymes is essential for elucidating their roles in human health and disease.

One group of gut bacterial CAZymes are β-glucuronidases (GUSs). Microbial GUS enzymes are unique among CAZymes because they play roles in the metabolism of both polysaccharides and drug metabolites. For example, GUSs are capable of catalyzing the hydrolysis of GlcA-containing polysaccharides, such as heparin and hyaluronate, as well as small molecule drug glucuronides like SN-38-G, the active metabolite of the anticancer drug irinotecan, and NSAID glucuronides [62,64,126,136]. Drug glucuronides are generated in the liver by uridine diphosphate glucuronosyl transferases (UGTs) and then secreted into the GI tract where they are processed by bacterial GUSs, which reverse the actions carried out by the host

[4]This chapter previously appeared as an article in the *Journal of Molecular Biology*. The original citation is as follows: Pellock, S. J., Walton, W. G., Ervin, S. M., Torres-Rivera, D., Creekmore, B. C., Bergan, G., Dunn, Z. D., Li, B., Tripathy, A., Redinbo, M. R. (2019) Discovery and characterization of FMN-binding β-glucuronidases in the human gut microbiome, *J. Mol. Biol. 431*, 970-980.

(**Figure 4.1**) [10]. Many other glucuronides, primarily glucuronate-containing polysaccharides, are present in the gut, including glycosaminoglycans, such as heparin and hyaluronate, as well dietary, bacterial, and plant polysaccharides like pectin, sphingans, and xylans [137]. Given the importance of these diverse molecules in therapeutics, nutrition, and homeostasis of the gut microbiota, it is critical to understand the structure and function of the microbial enzymes that process them.

Utilizing metagenomic data and structural analyses, we recently catalogued 279 unique GUS enzymes from the Human Microbiome Project (HMP) stool sample database [136]. Only a few bacterial GUSs from the human gut have a characterized structure and function, and most of these characterized GUSs do not have a clear function in the gut microbiota. Here we characterize the GUS from *Faecalibacterium prausnitzii* L2-6 and show that it binds flavin mononucleotide (FMN) at a distant surface site. While glycoside hydrolases are among the most well characterized enzymes in biochemistry, this is the first observation of a flavin-bound glycoside hydrolase. Utilizing this structural and functional data, we screened GUSs found in the HMP stool sample database for key FMN-binding residues and identified 14 additional FMN-binding GUSs. We characterized four of these biochemically and determined the crystal structures of two, which confirm a structurally conserved FMN-binding site. We further show by a comprehensive review of the PDB that the GUSs characterized here bind FMN unlike any previously characterized FMN-binding proteins. These data reveal the first association between FMN and a glycoside hydrolase, suggesting a link between FMN and carbohydrate metabolism in the human gut microbiota.

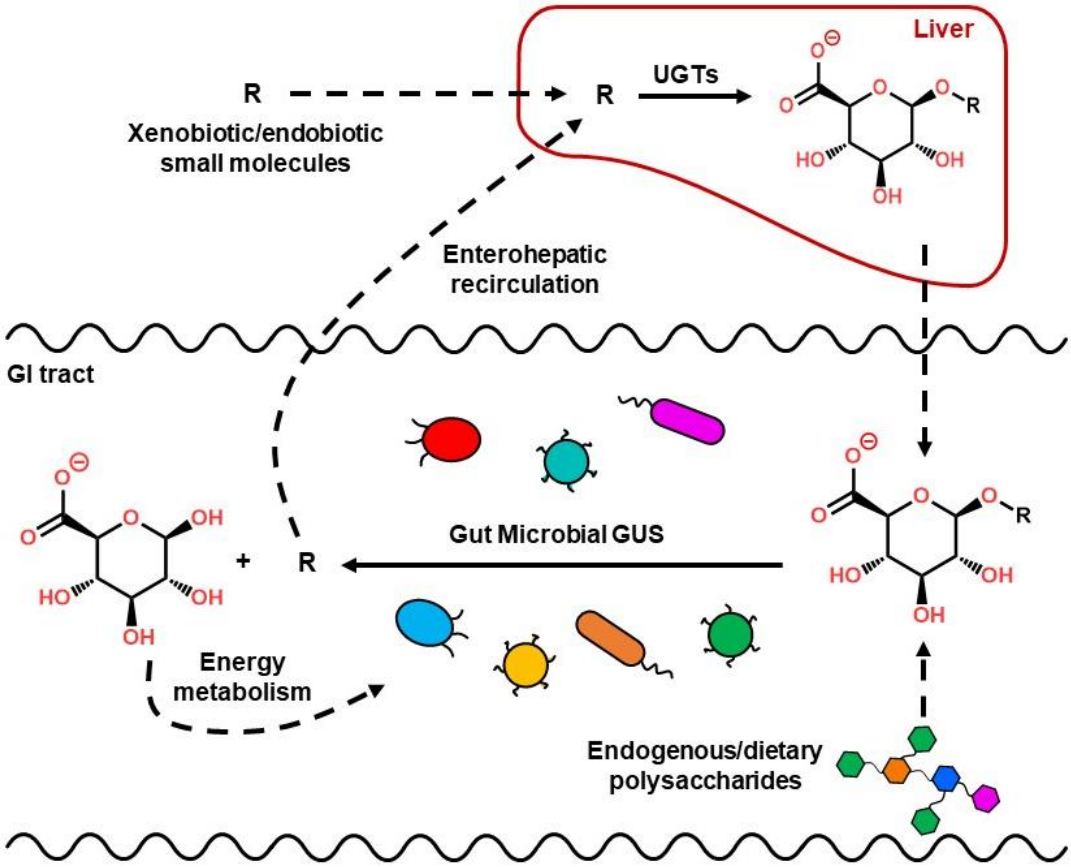


Figure 4.1 - Overview of the potential roles of bacterial GUS in the gut microbiome.

RESULTS

Discovery of an FMN-binding GUS from the Human Gut Microbe *F. prausnitzii* L2-6

Utilizing structural, functional, and sequence data on characterized bacterial GUS enzymes [55,56,107], metagenomic analysis of the HMP stool sample database revealed 279 unique GUS enzymes in the human gut microbiome [136]. Most of these proteins remain uncharacterized; thus, we initiated an effort to express and examine representative GUSs from the distinct classes discovered. Surprisingly, the gene synthesis, protein expression, and purification of a GUS from the human gut bacterium *F. prausnitzii* L2-6 (*Fp2*GUS) yielded a yellow protein product (**Figure 4.2A**). Absorbance scan of purified *Fp2*GUS displayed a profile characteristic of a flavin-binding protein (**Figure 4.2A**), and LC-MS analysis of purified *Fp2*GUS revealed the mass for flavin mononucleotide (FMN) (**Figure 4.2B**). While FMN is bound upon expression and purification of *Fp2*GUS, absorbance scans revealed that the stoichiometry of binding was approximately 0.42:1 (FMN:*Fp2*GUS), suggesting that vacant FMN-binding sites may be present (**Figure 4.3**). Thus, we utilized isothermal titration calorimetry (ITC) to measure the binding affinity of *Fp2*GUS for FMN using this partially occupied sample, which revealed a K_d of 1.05 μM (**Figure 4.2C**). Together, these data establish that *Fp2*GUS is an FMN-binding glycoside hydrolase.

Crystal Structure of *Fp2*GUS Reveals an FMN-Binding Site

To further understand the FMN-binding nature of *Fp2*GUS, we determined its crystal structure to 2.55 Å resolution (**Table 4.1**). The *Fp2*GUS crystal structure revealed an FMN-binding site approximately 30 Å away from the active site, formed by two alpha helices from the core TIM barrel fold (cyan) and two jellyroll-like β-sandwich domains (blue and green)

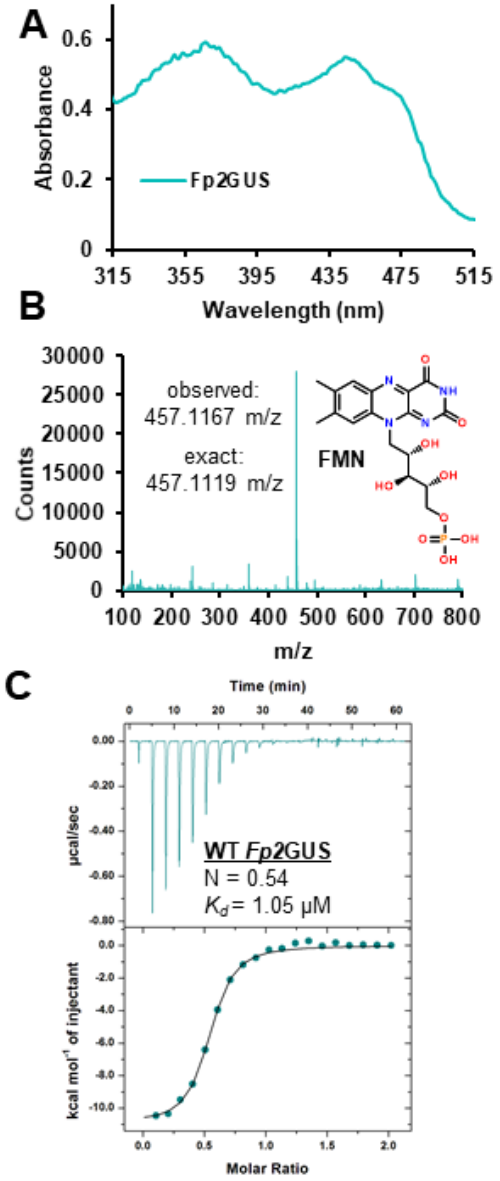
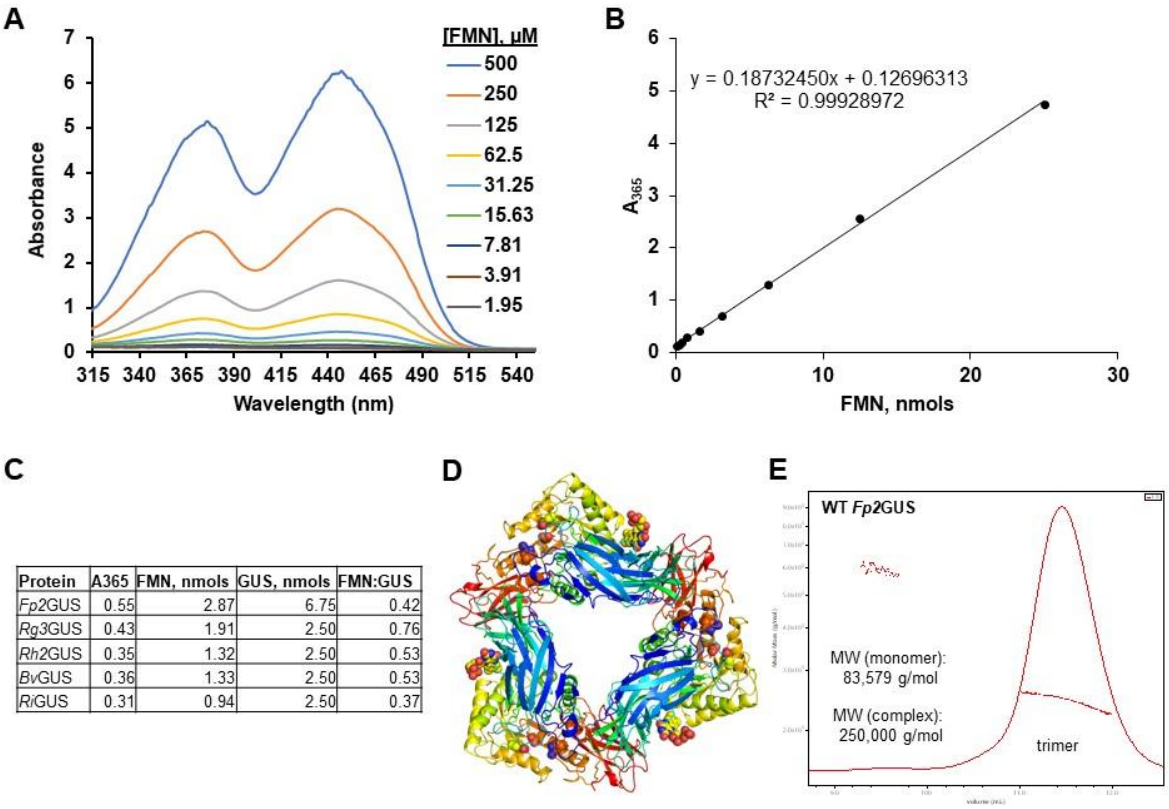


Figure 4.2 - Discovery of an FMN-binding GUS from the human gut microbe *F. prausnitzii* L2-6. (**A**) Purified *Fp2*GUS is yellow and absorbance scan reveals a UV profile characteristic of a flavin-binding protein (**B**) Mass spectrum of purified *Fp2*GUS contains mass for FMN (observed mass: 457.1167 m/z, exact mass: 457.1119 m/z). (**C**) Titration of WT *Fp2*GUS with FMN monitored by isothermal titration calorimetry (ITC) reveals binding constant of 1.05 ± 0.06 µM and an FMN occupancy for WT *Fp2*GUS of 46% (N = 0.54).



Protein	A365	FMN, nmols	GUS, nmols	FMN:GUS
*Fp2*GUS	0.55	2.87	6.75	0.42
*Rg3*GUS	0.43	1.91	2.50	0.76
*Rh2*GUS	0.35	1.32	2.50	0.53
*Bv*GUS	0.36	1.33	2.50	0.53
*Ri*GUS	0.31	0.94	2.50	0.37

Figure 4.3 – FMN stoichiometry and oligomeric state analysis. (**A**) Absorbance scans of free FMN at varying concentrations. (**B**) Standard curve of FMN at absorbance of 365 nm. (**C**) Apparent stoichiometries of FMN:GUS based on protein concentrations and FMN concentrations as determined by standard curve analysis. (**D**) Trimer of Fp2GUS with active site and FMN shown as spheres, each chain rainbow colored from N-term (blue) to C-term (red). (**E**) Size-exclusion chromatography multi angle light scattering analysis of *Fp2*GUS confirms trimeric state in solution.

Table 4.1 - Crystallography data collection and refinement statistics

Protein	*Fp2*GUS (PDB: 6MVF)	*Rg3*GUS (PDB: 6MVG)	*Rh2*GUS (PDB: 6MVH)
Resolution range	29.52 - 2.55 (2.641 - 2.55)	29.9 - 2.8 (2.9 - 2.8)	29.3 - 2.4 (2.5 - 2.4)
Space group	$P2_1$	C2	$P2_1$
Unit cell [a, b, c (Å); α, β, γ (°)]	129.6, 106.8, 183.8; 90.0, 90.3, 90.0	236.0, 127.1, 88.8; 90, 98.5, 90	94.4, 137.4, 108.8; 90, 91.9, 90
Total reflections	819862 (71386)	219322 (22412)	262591 (26323)
Unique reflections	163078 (16164)	63747 (6333)	106453 (10707)
Multiplicity	5.0 (4.4)	3.4 (3.5)	2.5 (2.5)
Completeness (%)	99.58 (99.53)	99.74 (99.91)	98.27 (99.07)
Mean I/sigma(I)	11.22 (2.58)	14.03 (2.36)	14.04 (3.87)
Wilson B-factor	33.66	57.98	32.01
R-merge	0.1212 (0.5703)	0.07687 (0.6069)	0.04908 (0.2139)
R-work	0.1673 (0.2309)	0.1854 (0.3146)	0.2098 (0.2311)
R-free	0.2262 (0.3257)	0.2457 (0.4071)	0.2652 (0.3219)
Number of non-hydrogen atoms	32076	15654	18966
protein	30492	15460	18003
ligands	186	96	128
solvent	1398	104	835
RMS(bonds)	0.008	0.009	0.009
RMS(angles)	0.98	1.08	1.29
Ramachandran favored (%)	96.49	93.99	96.03
Ramachandran allowed (%)	3.51	5.69	3.97
Ramachandran outliers (%)	0	0.32	0
Rotamer outliers (%)	1.81	6.09	0.42
Clashscore	3.15	9.16	6.75
Average B-factor	34.18	63.57	35.27
protein	33.99	63.61	35.25
ligands	63.21	69.05	55.79
solvent	34.52	52.81	32.45

(**Figure 4.4A**). The FMN-binding site is primarily formed by Y154 and F179, which make π-π stacking interactions with the isoalloxazine ring of FMN (**Figure 4.4B, C**). In addition to these two aromatic residues, K356 forms an ionic interaction with the phosphate of FMN, D151 forms a hydrogen bond with the amide nitrogen of the isoalloxazine ring, M161 participates in hydrophobic interactions with the isoalloxazine ring, and Y363 makes an edge-to-face π interaction with the isoalloxazine ring (**Figure 4.4B, C**). Computational generation (DFT theory: wB97x-D 6-31G*) of the electrostatic potential map of FMN reveals that the most electron poor region of the isoalloxazine ring makes π-π stacking interactions with the negative faces of Y154 and F179 (**Figure 4.4C**). In addition to the unprecedented FMN-binding site, *Fp2*GUS is also a trimer as identified from the crystal structure and size exclusion chromatography with multi-angle light scattering (**Figure 4.3D, E**). Together, these structural data reveal that *Fp2*GUS binds to FMN via numerous π-π interactions at a site located 30 Å from the active site.

FMN-Binding Site of *Fp2*GUS Integral to Protein Stability

To determine the role of the FMN-binding site in the structure and function of *Fp2*GUS, we performed site-directed mutagenesis of the following residues that contact FMN in the crystal structure: D151, Y154, F179, K356, and Y363 (**Figure 4.4B, C**). Size-exclusion chromatography and SDS-PAGE analysis revealed that each *Fp2*GUS mutant yielded unstable protein products, with a smaller amount of full-length and soluble protein relative to wild-type (WT), suggesting that the FMN-binding site helps stabilize *Fp2*GUS (**Figure 4.5A, B**). Absorbance scans of the intact peaks of the *Fp2*GUS mutants revealed either reduced or no flavin-binding profiles, suggesting reduced FMN binding (**Figure 4.5C**). The K356A, D151A, and Y363A mutants of *Fp2*GUS still bind FMN with similar potency to

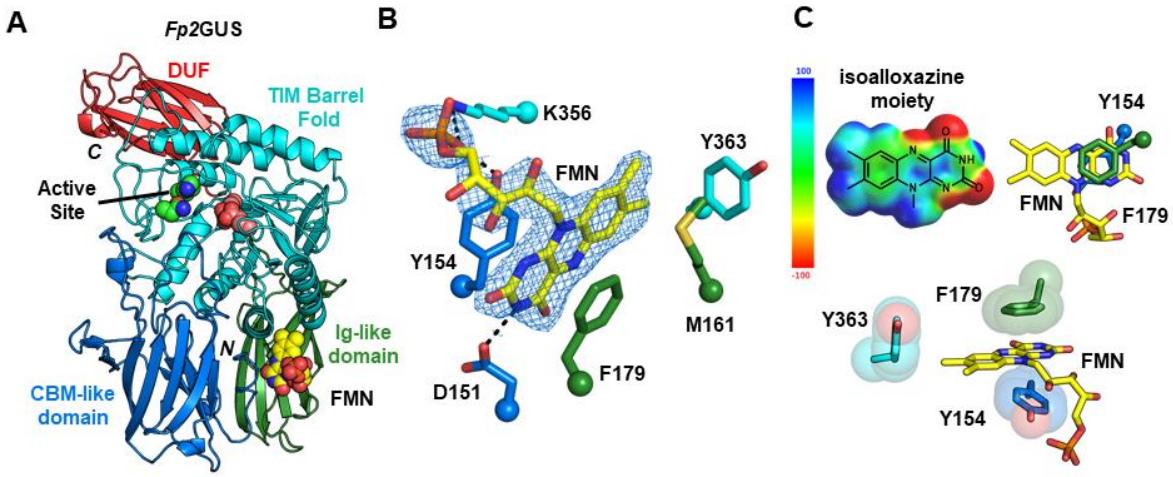
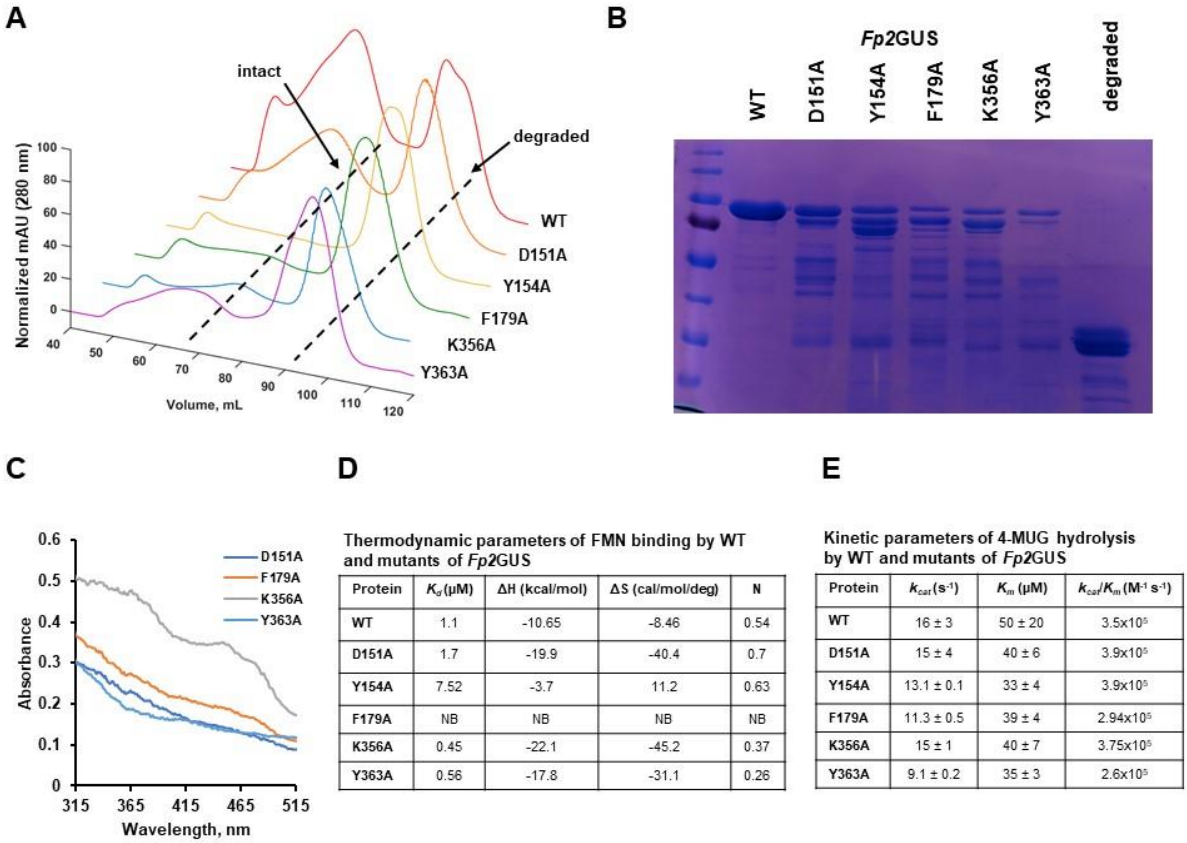


Figure 4.4 - Crystal structure of *Fp2*GUS reveals an FMN-binding site. (**A**) Monomer of *Fp2*GUS with active site glutamates shown as deep salmon spheres, NxK motif shown as green spheres, and FMN shown as yellow spheres. (**B**) FMN-binding site of *Fp2*GUS with 2Fo-Fc density shown at 1.0 σ. (**C**) Electrostatic potential map (DFT theory: wB97x-D 6-31G*) of FMN highlights an array of π interactions between FMN and *Fp2*GUS, including π-π stacking between the electron poor region of the isoalloxazine moiety and the electronegative faces of Y154 and F179 (right and bottom), as well as an edge-to-face interaction with Y363.



Thermodynamic parameters of FMN binding by WT and mutants of *Fp2*GUS

Protein	K_d (μM)	ΔH (kcal/mol)	ΔS (cal/mol/deg)	N
WT	1.1	-10.65	-8.46	0.54
D151A	1.7	-19.9	-40.4	0.7
Y154A	7.52	-3.7	11.2	0.63
F179A	NB	NB	NB	NB
K356A	0.45	-22.1	-45.2	0.37
Y363A	0.56	-17.8	-31.1	0.26

Kinetic parameters of 4-MUG hydrolysis by WT and mutants of *Fp2*GUS

Protein	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})
WT	16 ± 3	50 ± 20	$3.5x10^5$
D151A	15 ± 4	40 ± 6	$3.9x10^5$
Y154A	13.1 ± 0.1	33 ± 4	$3.9x10^5$
F179A	11.3 ± 0.5	39 ± 4	$2.94x10^5$
K356A	15 ± 1	40 ± 7	$3.75x10^5$
Y363A	9.1 ± 0.2	35 ± 3	$2.6x10^5$

Figure 4.5 - Characterization of WT and FMN-binding site mutants of *Fp2*GUS. (**A**) FPLC traces of *Fp2*GUS FMN-binding mutants. (**B**) SDS-PAGE of WT and FMN-binding mutants of *Fp2*GUS (**C**) Absorbance scans of purified *Fp2*GUS FMN-binding mutants. (**D**) ITC binding parameters of WT and FMN-binding site mutants of Fp2GUS for FMN. (**E**) Kinetic parameters of WT and FMN-binding site mutants of *Fp2*GUS.

the WT enzyme, suggesting that while these residues appear to form key contacts with FMN in the crystal structure, their overall contribution to binding FMN is negligible (**Figure 4.5D**). Thus, it would appear that, for the K356A, D151A, and Y363A mutant forms of *Fp2*GUS, less of the protein was bound to FMN, but for the fraction that did bind the cofactor, binding affinity remained similar to wild-type. In contrast, Y154A and F179A displayed significantly reduced binding to FMN, along with reduced stability (**Figure 4.5D**). Finally, interestingly, while stability is lost upon mutation of the FMN-binding site, the catalytic activities of the *Fp2*GUS mutants are not significantly different from the WT protein (**Figure 4.5E**). Together, these data suggest that the FMN-binding site of *Fp2*GUS plays a key role in the stability of the protein but does not affect catalysis.

Bioinformatic Analysis Identifies 14 Additional FMN-binding GUSs in the Human Gut

Using the structural and mutagenesis data from the *Fp2*GUS FMN-binding site as a guide, we analyzed GUS sequences from the HMP stool sample database to determine if other GUSs may be FMN-binders. Out of the 278 proteins examined, a total of 14 sequences, in addition to *Fp2*GUS, met these criteria (**Figure 4.6A** and **4.7**). Generation and analysis of a sequence similarity network (SSN) of GUS enzymes from the HMP stool sample database revealed that the one confirmed and 14 putative FMN-binding GUSs cluster into three clades, all of which contain GUSs exclusively from the previously defined No Loop (NL) structural category, which refers to the absence of an active site adjacent loop shown to play key roles in substrate specificity (**Figure 4.6A**) [136]. BLAST and SignalP 4.1 analysis of the putative FMN-binding GUS sequences reveal that they all come from bacteria in the class Clostridiales and do not contain signal peptide sequences, suggesting they are intracellular

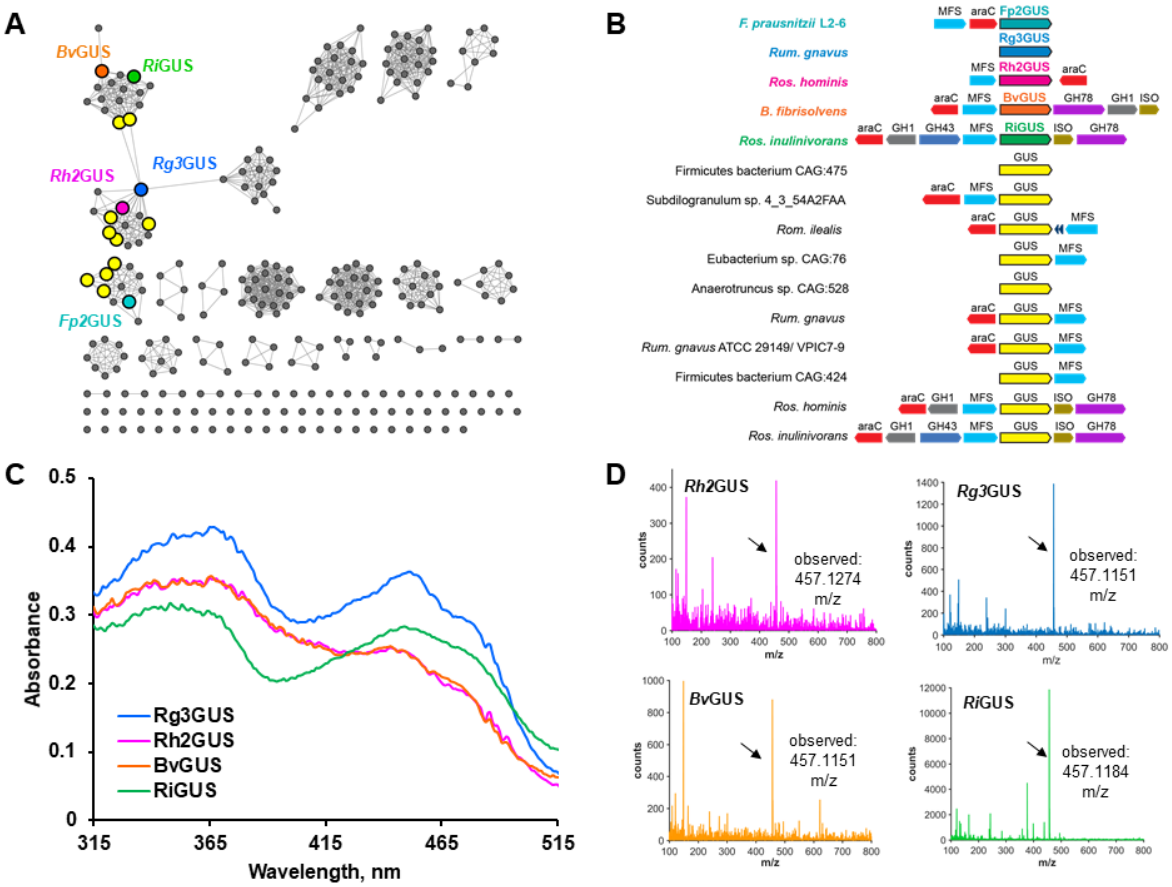


Figure 4.6 - Bioinformatic analysis predicts 14 additional FMN-binding GUS enzymes from the human gut microbiota. (**A**) SSN of the HMP GUSome with putative FMN-binders as larger, outlined circles and those that were synthesized and characterized further are labeled. BLAST/e-value/cut off value utilized to generate this SSN is 10^{-220} and sequences were acquired from the HMP stool sample database. (**B**) Genome neighborhood diagrams of putative FMN-binding GUSs (**C**) Absorbance scans of purified *Rh2*GUS, *Rg3*GUS, *Bv*GUS, and *Ri*GUS reveal flavin binding profile. (**D**) Mass spectra of purified *Rh2*GUS, *Rg3*GUS, *Bv*GUS, and *Ri*GUS reveal mass for FMN.

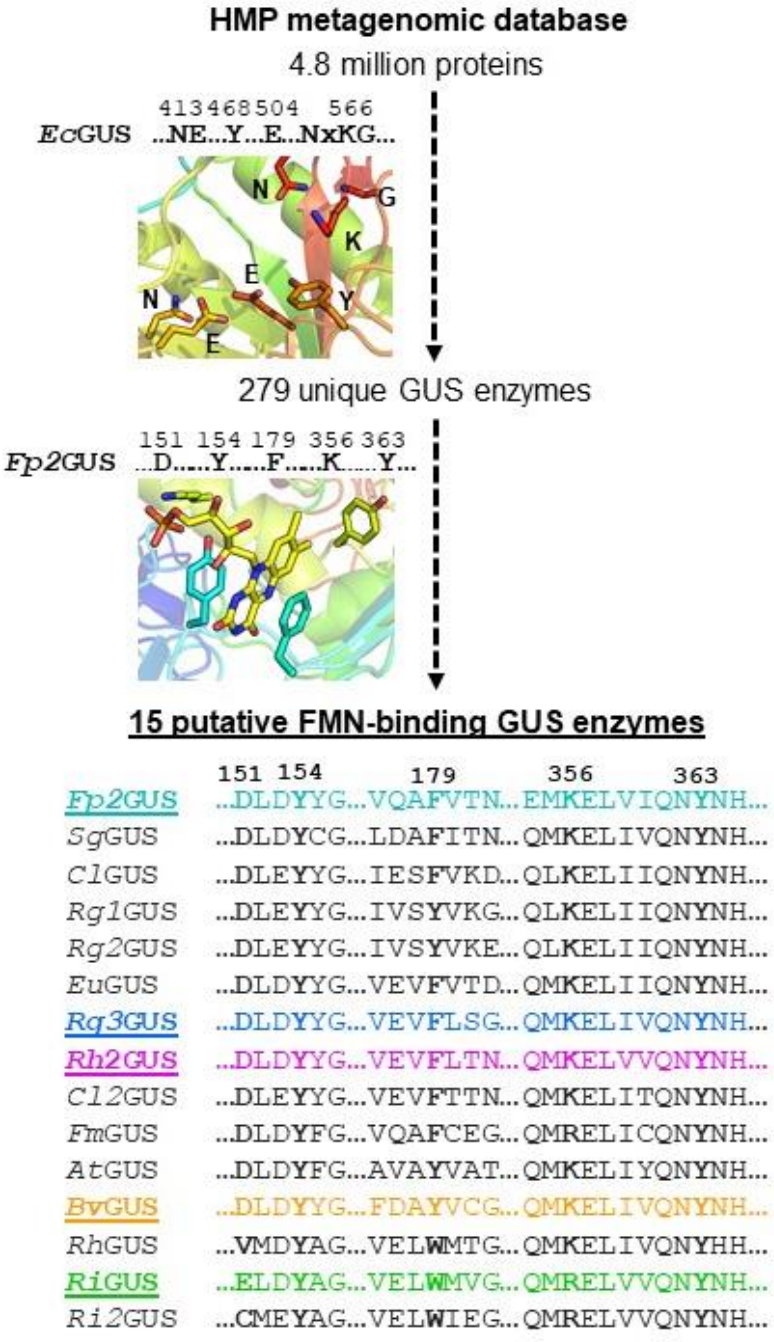


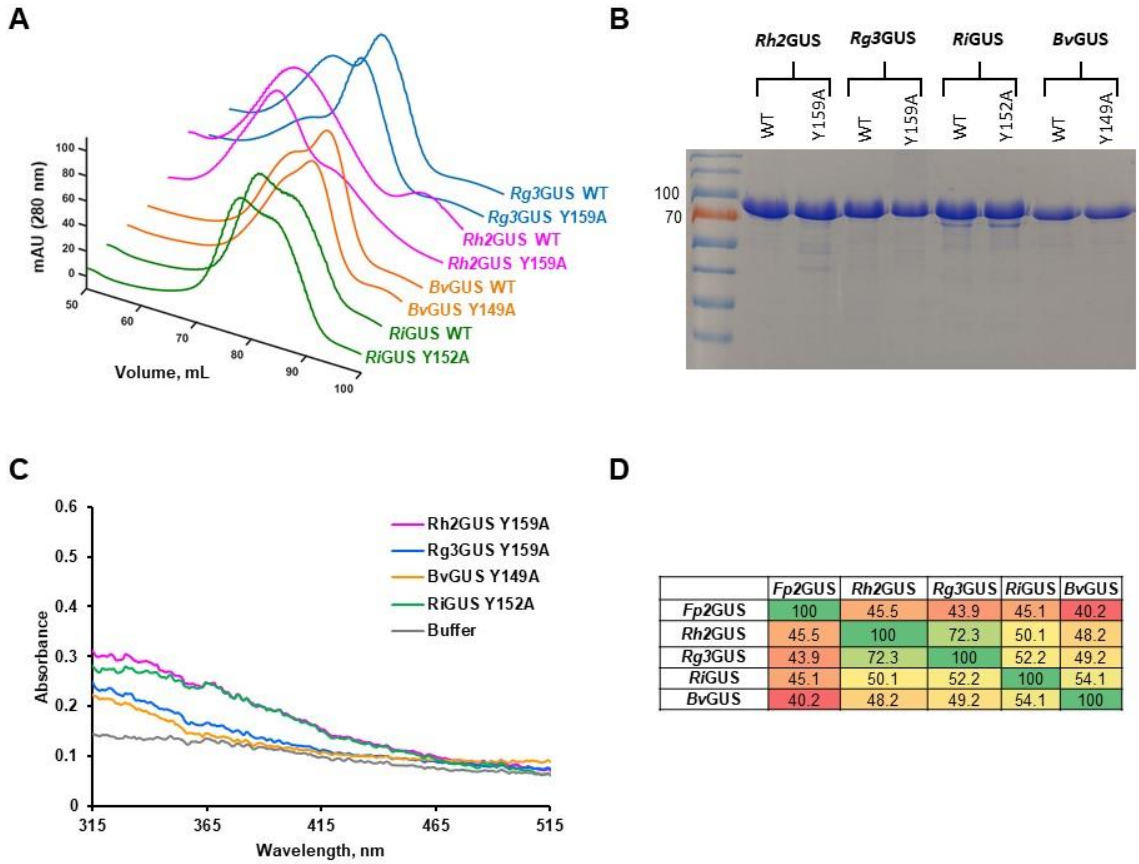
Figure 4.7 - Bioinformatic search of the HMP stool sample database reveals 15 putative FMN-binding GUS enzymes.

[103,138]. Further bioinformatic analysis utilizing the Enzyme Function Initiative Genome Neighborhood Tool (EFI-GNT) revealed two distinct genetic neighborhoods surrounding the FMN-binding GUS genes (**Figure 4.6B**) [129]. Four of the fifteen putative FMN-binders, including *Butyrivibrio fibrisolvens*, two strains of *Roseburia inulinivorans*, and *Roseburia hominis* are flanked by genes encoding a β-glucosidase (GH1), an α-L-rhamnosidase (GH78), and a xylose isomerase (ISO) (**Figure 4.6B**). The remaining 11 FMN-binding GUS genes are flanked by AraC transcriptional regulators and MFS transporters, proteins that likely sense and transport glucuronate-containing molecules (**Figure 4.6B**). Taken together, a family of FMN-binding GUS enzymes appears to be encoded by the human gut microbiome.

Biochemical Characterization Confirms FMN-Binding of Bioinformatic Hits

From the 14 additional putative FMN-binding GUS sequences identified, we selected the following four GUS genes to synthesize, express, and purify: *Roseburia inulinivorans* (*Ri*GUS), *Roseburia hominis* (*Rh2*GUS), *B. fibrisolvens* (*Bv*GUS), and *Ruminococcus gnavus* (*Rg3GUS*). These four enzymes share between 40-46% sequence identity with *Fp2*GUS (**Figure 4.8D**). Upon expression and purification, all four of the selected sequences yielded yellow protein products with a flavin-binding absorbance profile (**Figure 4.6C**), displayed the mass for FMN by LC-MS (**Figure 4.6D**), and bound FMN with dissociation constants that range from 60 nM (*Rg3*GUS) to 1.27 μM (*Bv*GUS) (**Figure 4.9** and **Table 4.2**). These results validate our structure-guided bioinformatic identification of FMN-binding GUSs from the gut microbiome.

To confirm that the FMN-binding sites of *Rh2*GUS, *Rg3*GUS, *Bv*GUS, and *Ri*GUS were similar in molecular nature to that identified in *Fp2*GUS, we mutated the residue



	*Fp2*GUS	*Rh2*GUS	*Rg3*GUS	*Ri*GUS	*Bv*GUS
*Fp2*GUS	100	45.5	43.9	45.1	40.2
*Rh2*GUS	45.5	100	72.3	50.1	48.2
*Rg3*GUS	43.9	72.3	100	52.2	49.2
*Ri*GUS	45.1	50.1	52.2	100	54.1
*Bv*GUS	40.2	48.2	49.2	54.1	100

Figure 4.8 - Biochemical characterization of WT and FMN-binding site mutants of *Rg3*GUS, *Rh2*GUS, *Bv*GUS, and *Ri*GUS. (**A**) SEC traces for FMN-binding site mutants Y159A *Rh2*GUS, Y159A *Rg3*GUS, Y152A *Ri*GUS, and Y149A *Bv*GUS. (**B**) SDS-PAGE gel of WT and FMN-binding site mutants of *Rh2*GUS, *Rg3*GUS, *Ri*GUS, and *Bv*GUS. (**C**) Absorbance scans of WT and FMN-binding site mutants of *Rh2*GUS, *Rg3*GUS, *Ri*GUS, and *Bv*GUS. (**D**) Sequence identity matrix of the five characterized FMN-binding site mutants.

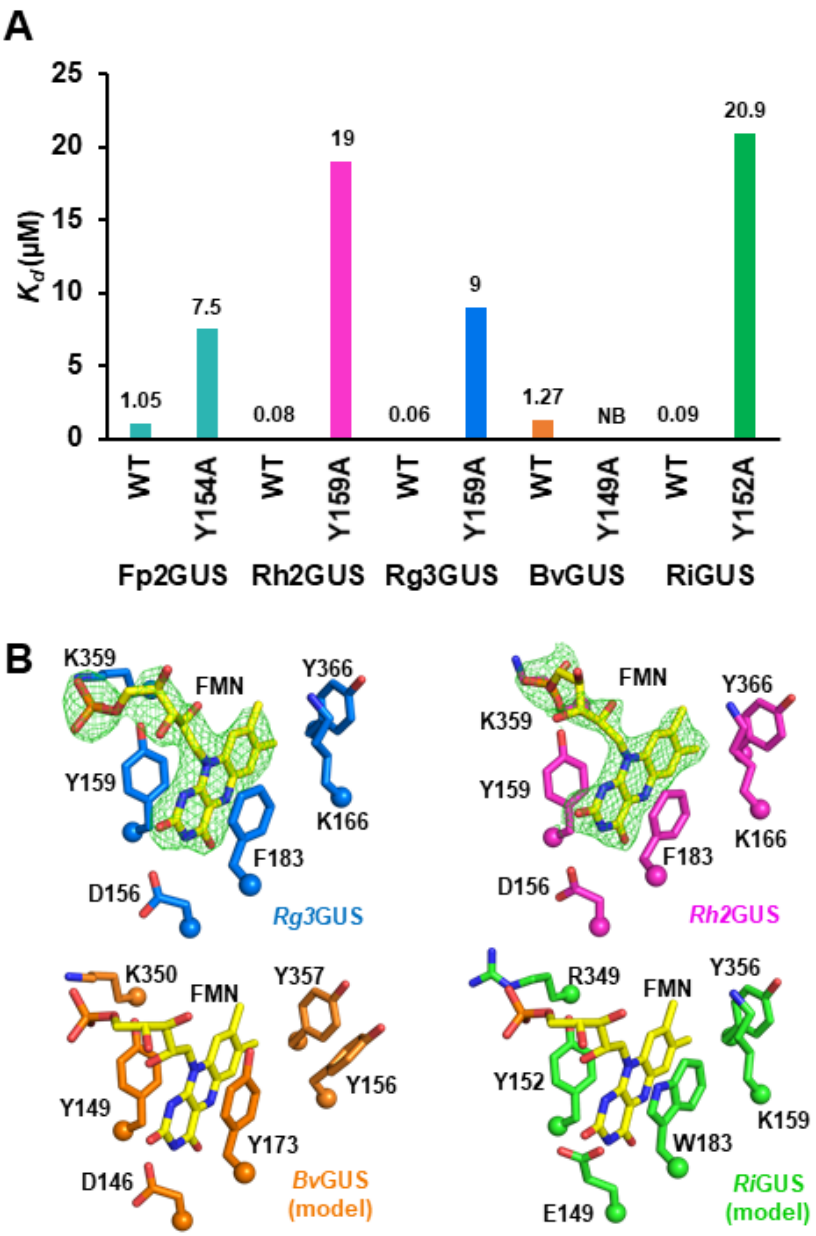


Figure 4.9 - Binding studies and structural analysis demonstrate that GUS enzymes utilize a conserved motif to bind FMN. (**A**) Binding affinities (K_d) of GUS enzymes and FMN-binding site mutants for FMN as determined by isothermal titration calorimetry (**B**) Structures of FMN-binding sites as determined by crystallography for *Rg3*GUS (blue) and *Rh2*GUS (magenta) and by template-guided modeling for *Bv*GUS (orange) and *Ri*GUS (green).

Table 4.2 - Thermodynamic parameters of FMN binding by WT and mutant GUS proteins.

Protein	K_d(μM)	ΔH (kcal/mol)	ΔS (cal/mol/deg)	N
*Fp2*GUS WT	1.05	-10.9	-9.1	0.53
*Fp2*GUS Y154A	7.6	-5	7	0.5
*Rh2*GUS WT	0.080	-19	-30	0.56
*Rh2*GUS Y159A	20	-4	9	0.87
*Rg3*GUS WT	0.060	-16.56	-22.5	0.263
*Rg3*GUS Y159A	9	-9.6	-9	0.82
*Bv*GUS WT	1.27	-10.95	-9.7	0.54
*Bv*GUS Y149A	NB	--	--	--
*Ri*GUS WT	0.09	-18.1	-28	0.33
*Ri*GUS Y152A	20.9	-6.9	-1.6	0.89

Values represent the average of n = 2 replicates.

corresponding to Y154 in *Fp2*GUS in each protein to alanine (**Fig. 2B, 4B**). Each of these mutants were no longer yellow or displayed significantly reduced yellow color, and no longer displayed a flavin-binding profile (**Figure 4.8C**). Interestingly, unlike the FMN-binding site mutants of *Fp2*GUS, these mutants yielded stable protein products (**Figure 4.8A**, **B**). The mutant GUSs Y152A *Ri*GUS, Y159A *Rg3*GUS, and Y159A *Rh2*GUS were still capable of binding FMN, although they did so with much weaker affinities (**Figure 4.9A** and **Table 4.2**). Together, these data demonstrate that FMN-binding GUSs utilize a conserved motif to bind FMN, and that mutation of FMN binding residues has differential effects on distinct GUS enzymes.

Structural Analysis of *Rh2*GUS and *Rg3*GUS Reveals a Conserved FMN-Binding Site

We next determined the crystal structures of *Rh2*GUS and *Rg3*GUS to 2.4 and 2.8 Å, respectively. *Rh2*GUS and *Rg3*GUS share high structural similarity with *Fp2GUS*, aligning with RMSD values of 1.6 Å and 1.8 Å over 624 Cα positions, respectively (**Figure 4.10A**). Similar to *Fp2*GUS, *Rh2*GUS and *Rg3GUS* both contain an FMN-binding site that is approximately 30 Å from the active site (**Figure 4.9B** and **4.10B**). The FMN-binding site of *Rh2*GUS and *Rg3*GUS is similar to *Fp2*GUS, except that a lysine replaces the methionine interacting with the isoalloxazine ring (**Figure 4.9B**). While the quaternary structure of *Rg3*GUS is unclear based on the crystal structure (**Figure 4.10C**), it appears that *Rh2*GUS may form either a unique dimer or tetramer based on its crystal structure (**Figure 4.10D**). SEC-MALS analysis supports this conclusion, with predicted dimer-tetramer mixtures for both *Rh2*GUS and *Rg3*GUS (**Figure 4.11**). Therefore, we conclude that FMN-binding GUSs in the gut microbiota contain similar FMN-binding sites, similar tertiary structures, but have distinct quaternary structures.

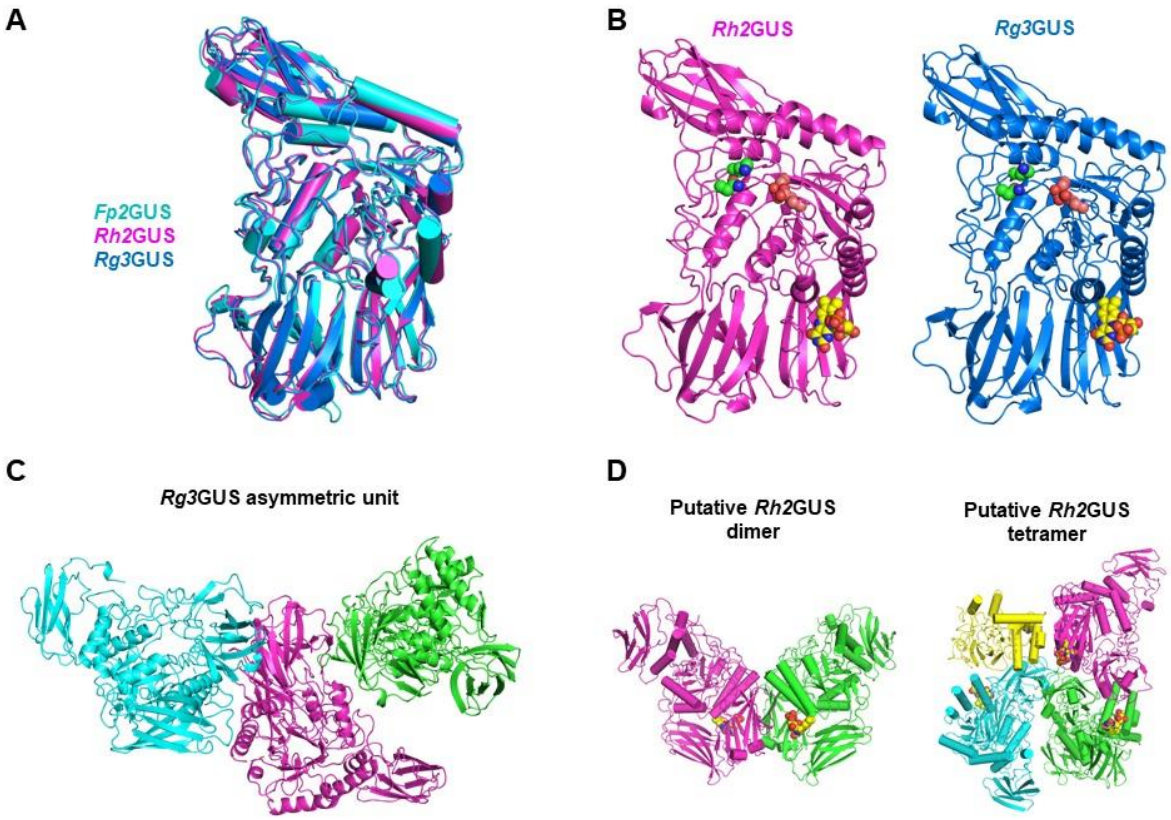


Figure 4.10 - Structural comparison of FMN-binding GUSs and quaternary structure analysis. (**A**) Structural overlay of *Fp2*GUS, *Rh2*GUS, and *Rg3*GUS. (**B**) Tertiary structure analysis of *Rh2*GUS and *Rg3*GUS with catalytic residues and FMN shown as spheres. (**C**) Analysis of *Rg3*GUS asymmetric unit does not reveal obvious dimer or tetramer. (**D**) Putative dimer and tetrameric states of Rh2GUS.

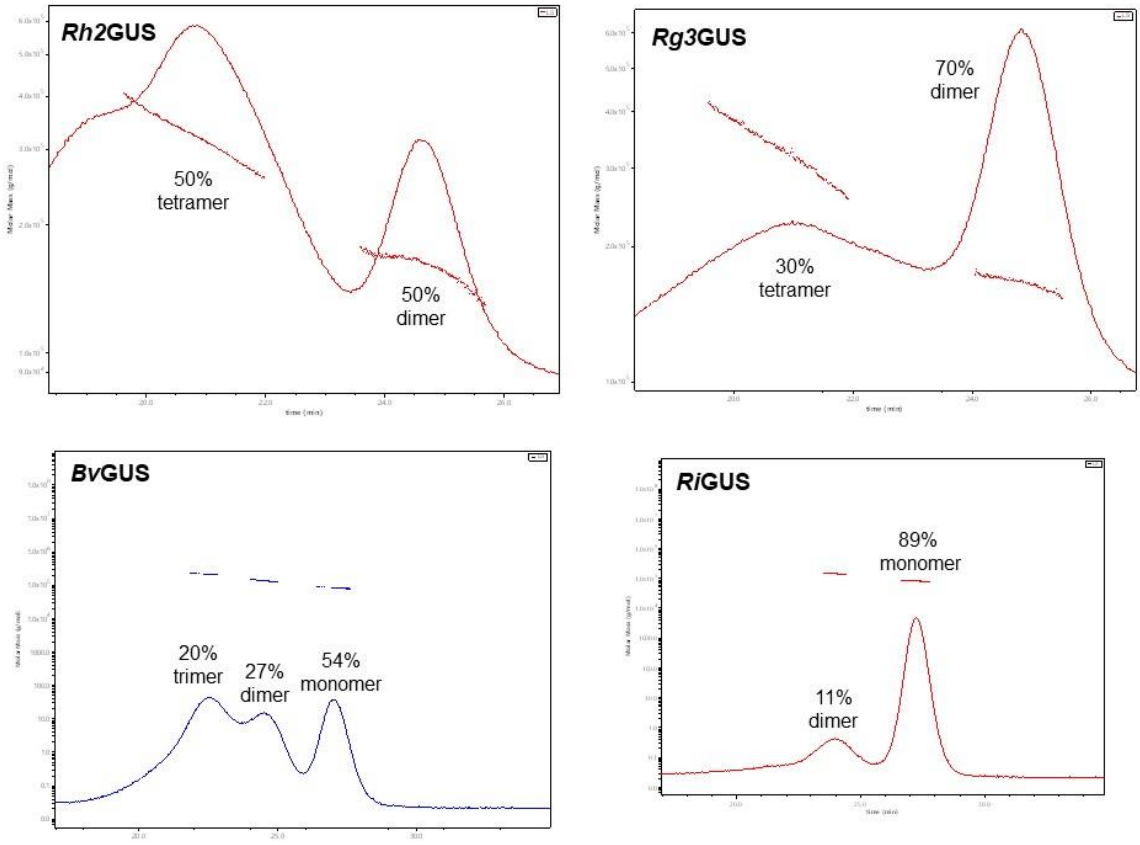


Figure 4.11. SEC-MALS results of FMN-binding GUS enzymes.

The FMN-Binding Site is Not Required for GUS Activity

To assess the role of the FMN-binding site in the catalytic function of these GUSs, we determined the catalytic properties of the WT and FMN-binding site mutants. Each GUS was able to hydrolyze the fluorescent reporter substrate 4-methylumbelliferyl glucuronide (4-MUG) (**Table 4.3**). The WT and FMN-binding mutants displayed nearly identical k_{cat} values (**Figure 4.12A**), but the FMN-binding mutants generally displayed higher K_M values (**Figure 4.12B**). For example, the FMN-binding mutants of *Rh2*GUS, *Rg3*GUS, and *Bv*GUS all display significantly higher K_M values than WT (**Figure 4.12B** and **Table 4.3**). In contrast, the K_M values of the FMN-binding mutants of *Fp2*GUS and *Ri*GUS were not significantly different from the WT enzymes. Thus, these data confirm that the FMN-binding site is not necessary for catalytic function but does increase the K_M for some of these FMN-binding GUSs.

The GUS FMN-binding Site is Unique Among Characterized FMN-binding Proteins

Because the FMN-binding site does not affect catalysis, we performed a comprehensive search of the PDB in an attempt to deduce a function for this site by looking for similar structures. We examined the PDB for FMN-bound structures, which identified 1,056 deposited structures, with 438 non-redundant entries. We then visually inspected these 438 non-redundant structures in PyMOL to determine if any other FMN-binding proteins bind flavin in the same manner as the GUSs characterized in this work. In agreement with previous analyses, the most common fold we encountered was the TIM-barrel fold, and we use Old Yellow Enzyme (PDB: 1OYB) as an example of how FMN binds at the β-barrel core of the fold (**Figure 4.13**). Interestingly, while the FMN-binding GUS enzymes characterized here are also TIM-barrel-containing proteins, they bind FMN on the exterior of this fold, not

Table 4.3 - Kinetic parameters of 4-MUG hydrolysis for WT and mutant GUSs.

Protein	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($s^{-1} M^{-1}$)
*Fp2*GUS WT	16 ± 3	50 ± 20	$4.57x10^5$
*Fp2*GUS Y154A	13.1 ± 0.1	33 ± 4	$3.95x10^5$
*Rh2*GUS WT	33 ± 5	150 ± 20	$2.21x10^5$
*Rh2*GUS Y159A	35 ± 5	220 ± 10	$1.59x10^5$
*Rg3*GUS WT	47 ± 4	87 ± 8	$5.4x10^5$
*Rg3*GUS Y159A	46 ± 5	140 ± 10	$3.22x10^5$
*Bv*GUS WT	6 ± 2	40 ± 20	$1.48x10^5$
*Bv*GUS Y149A	6.5 ± 0.2	84 ± 8	$7.82x10^4$
*Ri*GUS WT	21 ± 3	500 ± 100	$4.40x10^4$
*Ri*GUS Y152A	22 ± 1	370 ± 80	$5.97x10^4$

Values are an average of n = 3 replicates ± SD.

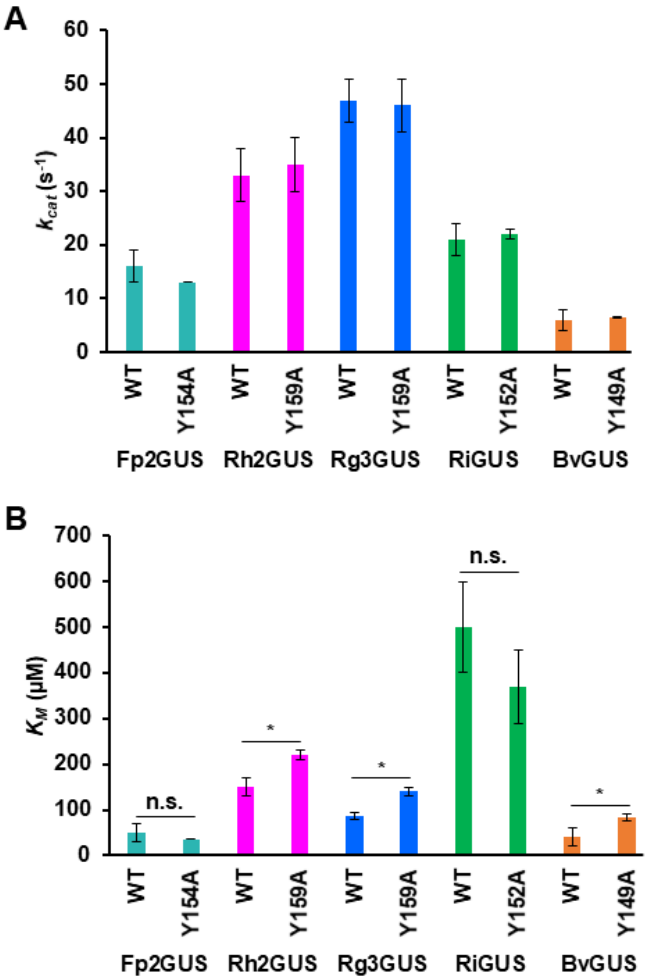


Figure 4.12 - Kinetic studies of FMN-binding GUSs show that FMN-binding site is not required for GUS activity. (**A**) Catalytic turnovers (k_{cat}) of FMN-binding GUSs and their respective FMN-binding site mutants. (**B**) Michaelis constants (K_M) of FMN-binding GUSs and their respective FMN-binding site mutants.

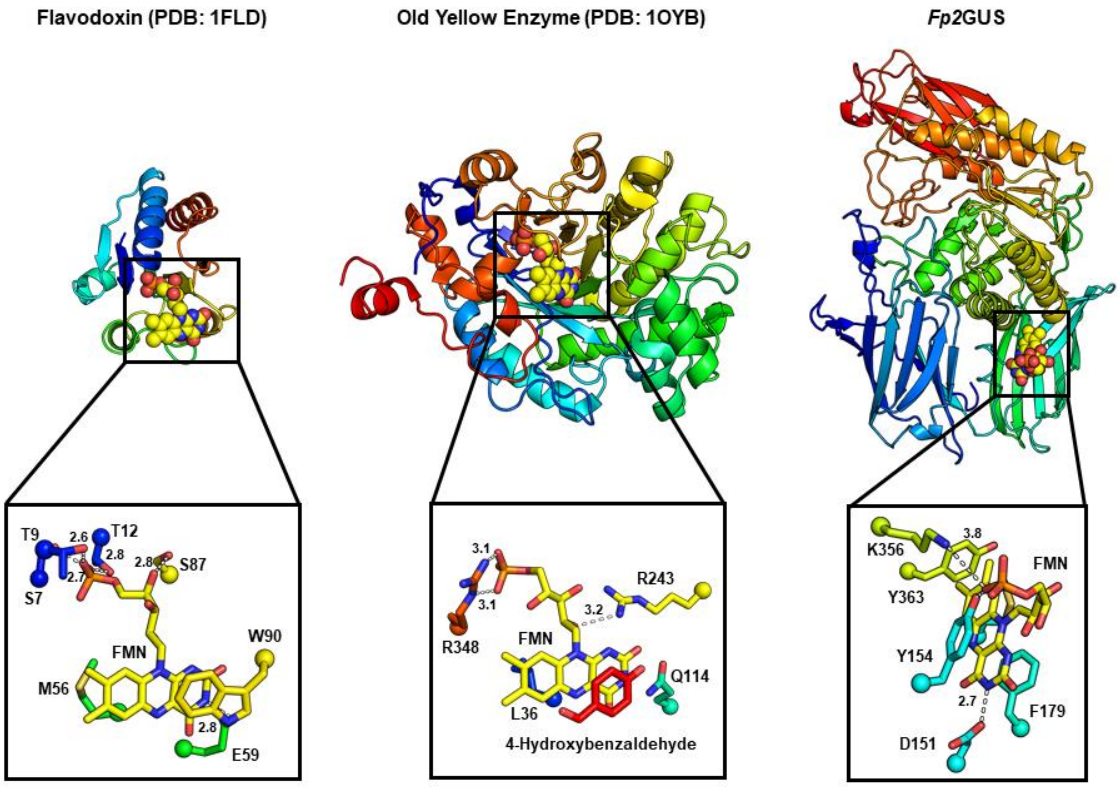


Figure 4.13 - Structural comparison of FMN-binding proteins reveals that FMN-binding GUSs identified here bind FMN in a unique manner. (**A**) Structure of flavodoxin (PDB: 1FLD) (**B**) Old Yellow Enzyme (PDB: 1OYB), and (**C**) *Fp2*GUS with zoom-in of their respective FMN-binding sites.

the interior, and two adjacent β-sandwich domains contribute to the FMN-binding site as well (**Figure 4.4A** and **4.13**). The second most common FMN-binding fold is the flavodoxin fold, named after the protein flavodoxin (PDB: 1FLD), which binds FMN at the edge of a β-sheet flanked on both sides by α-helices (**Figure 4.13**). In addition to the TIM-barrel fold and flavodoxin fold, we identified 18 other structurally distinct FMN-binders out of the 438 non-redundant structures (**Figure 4.14**). These other folds sample a wide variety of structural motifs to bind FMN, but none bind FMN outside the TIM barrel core as observed for the GUS enzymes characterized here (**Figure 4.14)**. Thus, the FMN-binding site in gut microbial GUS enzymes is unique.

As an additional, distinct screen to confirm the novelty of this FMN-binding site, we performed a PDBeFold search to find similar structures to the GUS enzymes characterized here. The top hits were all previously characterized GUS enzymes, with the most similar being *Bf*GUS (PDB: 3CMG) and *Bu*GUS-3 (PDB: 6D1P). Inspection of these structures revealed the common GH2 GUS fold, but the absence of the key FMN-binding residues identified in the GUSs characterized here (**Figure 4.15B**). Taken together, this structural analysis demonstrates that the GUSs characterized here are unique among FMN-binding proteins characterized to date.

Discussion

Here we structurally and functionally characterize a family of FMN-binding GUS enzymes. Characterization of the GUS from the human gut bacterium *F. prausnitzii* L2-6 revealed a novel FMN-binding glycoside hydrolase and its crystal structure guided the search for other FMN-binding GUS enzymes in the HMP stool sample database. We identified 14 more unique FMN-binders, four of which were characterized and confirmed to bind FMN.

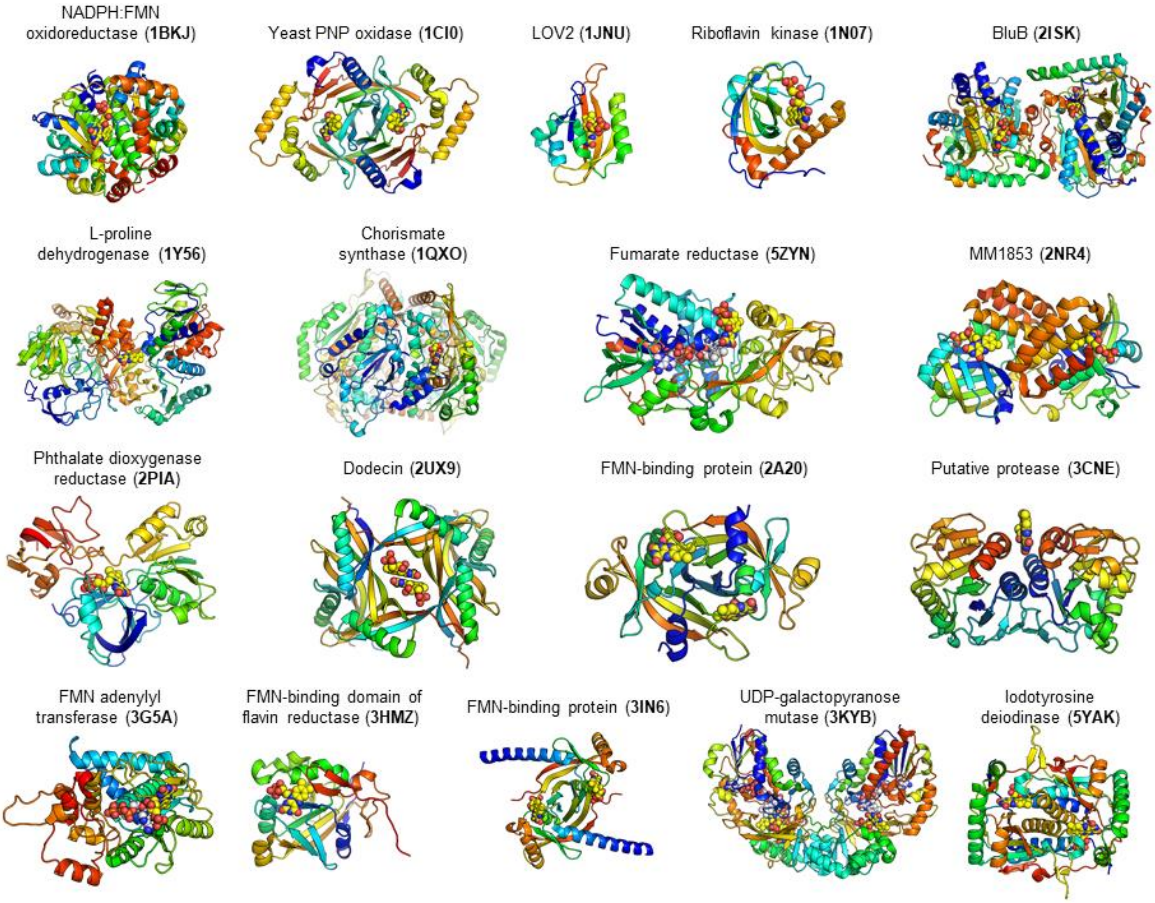


Figure 4.14 - Distinct FMN-binding proteins identified in the Protein Data Bank.

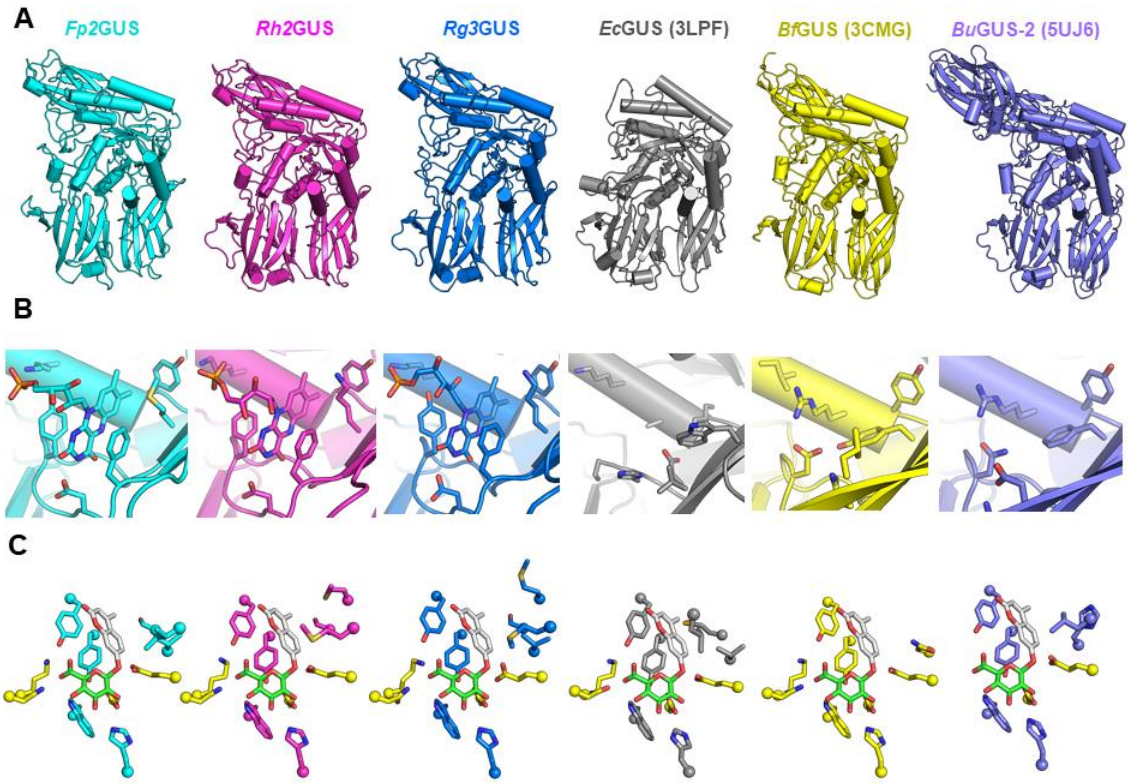


Figure 4.15 - Structural comparison of FMN-binding GUSs with previously characterized GUS enzymes.

Our characterization of the WT and FMN-binding site mutants of these GUSs suggest that FMN plays a key role in the stability of *Fp2*GUS, but not for the other FMN-binders (**Figure 4.5A, B** and **4.8A**, **B**). Furthermore, we show that mutating the FMN-binding site does not significantly affect enzyme activity, suggesting that the FMN-binding site is not essential for GUS function (**Figure 4.12**). Together, these data reveal the first FMN-binding glycoside hydrolase, show that there is a family of FMN-binding glycoside hydrolases in the gut microbiota, the FMN-binding site is not required for the function of the enzyme, and reveal a novel FMN-binding site among FMN-binding proteins structurally characterized to date.

While previously characterized bacterial GUSs share similar tertiary structures to the FMN-binding GUSs discovered here, none contain the residues necessary to form the FMN-binding site (**Figure 4.15B**). *Fp2*GUS shares RMSDs of 3.4 Å (across 528 Cα positions), 2.1 Å (624 Cα), and 2.7 Å (632 Cα) with *E. coli* GUS (*Ec*GUS, PDB: 3LPF), *B. fragilis* GUS (*Bf*GUS, PDB: 3CMG), and *B. uniformis* GUS 2 (*Bu*GUS-2, PDB:5UJ6), respectively. The C-terminal domain of *Fp2*GUS, *Rh2*GUS, and *Rg3*GUS was disordered in the crystal structures elucidated here, with approximately 100 residues missing from each individual chain. This unresolved region may form a carbohydrate binding module (CBM), a structural feature seen at the C-terminus of two GUSs previously characterized from *B. uniformis* [126,136]. The active site of the FMN-binding GUSs characterized here is similar to previously characterized GUS enzymes, containing both the two conserved catalytic glutamates and the NxK motif (**Figure 4.15C**) [136].

The FMN-binding GUSs characterized here add to the diverse quaternary structures discovered recently among gut bacterial GUSs. Five distinct oligomerization states have been previously observed, including three distinct tetramers and two unique dimers (**Figure 4.16**).

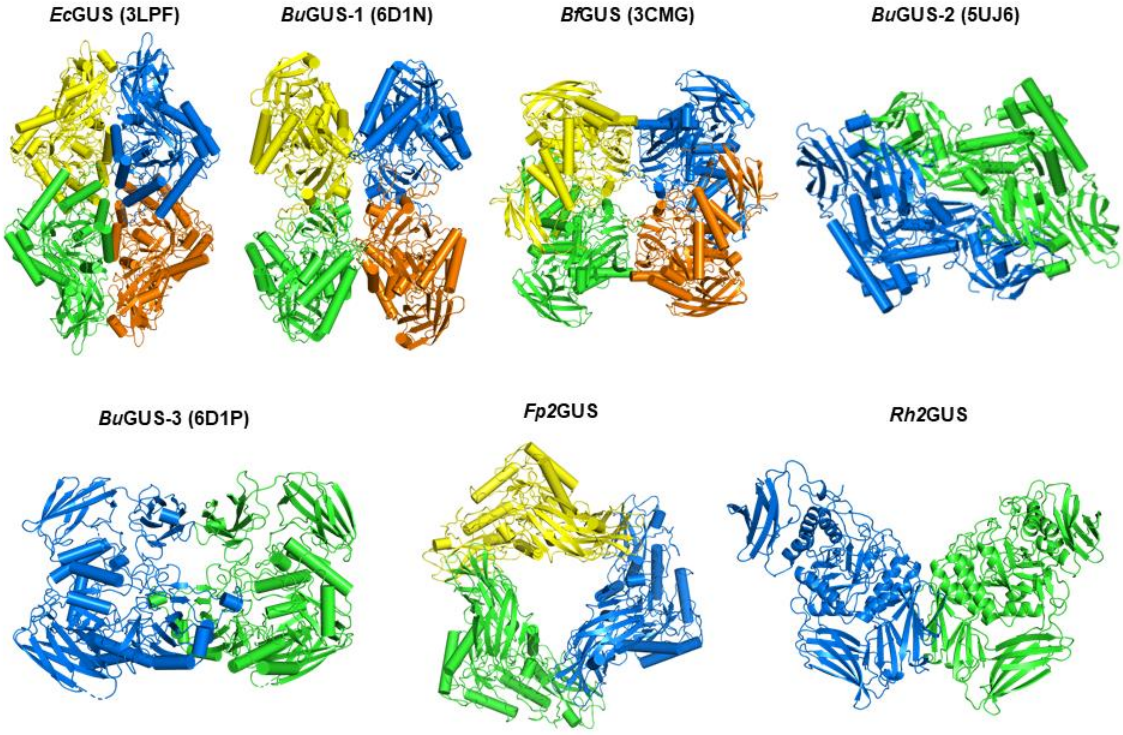


Figure 4.16 - Structural comparison of quaternary structures of microbial GUS enzymes.

Here we uncover two additional tetrameric states, a trimer (*Fp2*GUS) and an apparent dimer-tetramer mix (*Rh2*GUS) (**Figure 4.3D** and **4.10D**). The oligomeric states of bacterial GUSs have previously been shown to play key roles in GUS function. For example, the tetrameric Loop 1 (L1) GUS enzymes (*E. coli* GUS, PDB: 3LPF, **Figure 4.15A**) have small hydrophobic pockets around their active sites due to overlapping loops of adjacent protomers, limiting their substrate scope to lipophilic, small-molecule glucuronides [136]. Most other bacterial GUSs have open, solvent-exposed active sites, allowing access to large, polar substrates like glucuronate-containing polysaccharides [126,136]. Both the trimer of *Fp2*GUS and the dimer-tetramer mix of *Rh2*GUS fit within this latter group of GUSs with solvent exposed active sites, suggesting that their preferred substrates are likely glucuronate-containing polysaccharides.

The cognate substrates of some of these FMN-binding GUSs may be extracellular polysaccharides (EPS). The EFI-GNT analysis of the GUS genes from *B. fibirisolvens*, two strains of *R. inulinivorans*, and *R. hominis* revealed nearby CAZymes, including a β-glucosidase (GH1), an α-L-rhamnosidase (GH78), and a xylose isomerase (ISO) (**Figure 4.6B**). Together, these genes may coordinate the degradation of an EPS within the group known as sphingans. The repeating unit of sphingans contains α-linked rhamnose, β-linked glucose, and β-linked glucuronate moieties, all of which could be processed by the genes present in these genetic loci (**Figure 4.17**) [139]. Indeed, previous studies have associated orthologs of these enzymes with the catabolism of sphingans and related exopolysaccharides [140,141]. Detailed future studies with relevant polysaccharide substrates will be required to identify the cognate substrates of these FMN-binding GUSs.

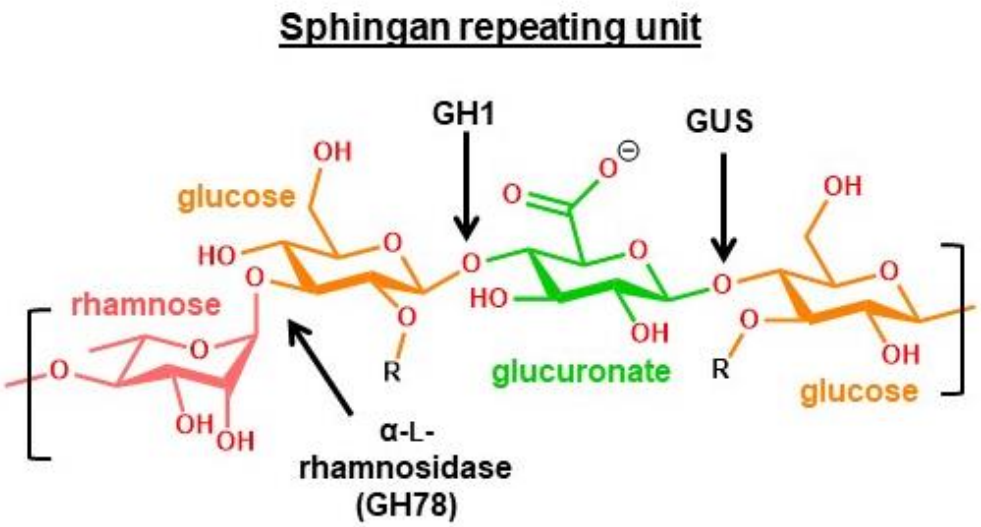


Figure 4.17 - Chemical structure of a sphingan repeating unit, a potential substrate of the FMN-binding GUS enzymes.

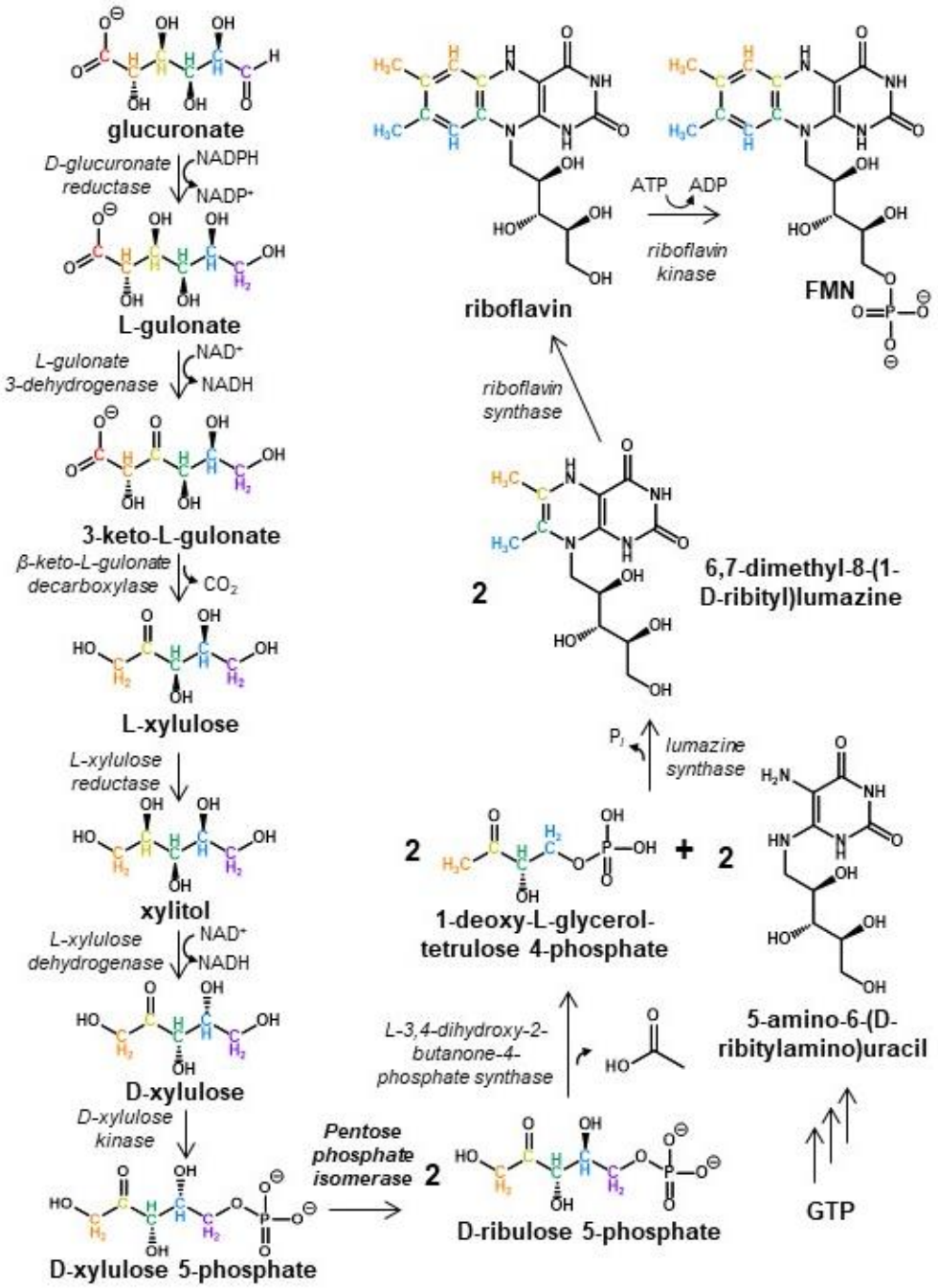


Figure 4.18 - Possible metabolic pathway from glucuronate to FMN.

The discovery of FMN-binding GUSs suggests a link between polysaccharide metabolism and FMN in the human gut. Interestingly, the pentose and glucuronate interconversion pathway contains a biosynthetic route that shunts glucuronate into the riboflavin metabolism pathway, responsible for the generation of FMN and FAD in bacteria (**Figure 4.18**). A myriad of enzymes is needed to transform glucuronate to D-ribulose-5-phosphate, which can generate FMN in combination with GTP (**Figure 4.18**). If glucuronate released by the FMN-binding GUSs is converted into FMN, it could serve as a positive feedback loop for GUS activity. However, this seems unlikely as we observed that FMN has only a small impact on the activity of these enzymes.

We note in our three structures of FMN-binding gut microbial GUS enzymes that a channel exists on the surface of each protein that provides access to N5 of the bound flavin, which is a key site for redox chemistry (**Figure 4.19**). It is possible that a small molecule could access this site to utilize FMN for oxidation-reduction reactions. While this is speculative, two features of the FMN-bound GUS enzymes make this potentially accurate. First, the non-FMN binding GUS proteins have residues that block this channel, while the channel remains open in the FMN-bound GUS enzymes. Second, while many GUS enzymes have signal sequences and may be exported to the periplasm, all the FMN-binding GUS enzymes detected to date lack a signal sequence and thus are expected to remain intracellularly localized, a place where redox chemistry can be better controlled and utilized by the cell. Future work will be required to determine whether this channel provides functionally relevant access to the bound FMN of these particular GUS enzymes.

Structural and functional analysis reported here leaves the role of FMN in GUS function unclear. While FMN was a key factor for stability in *Fp2*GUS, it did not have a

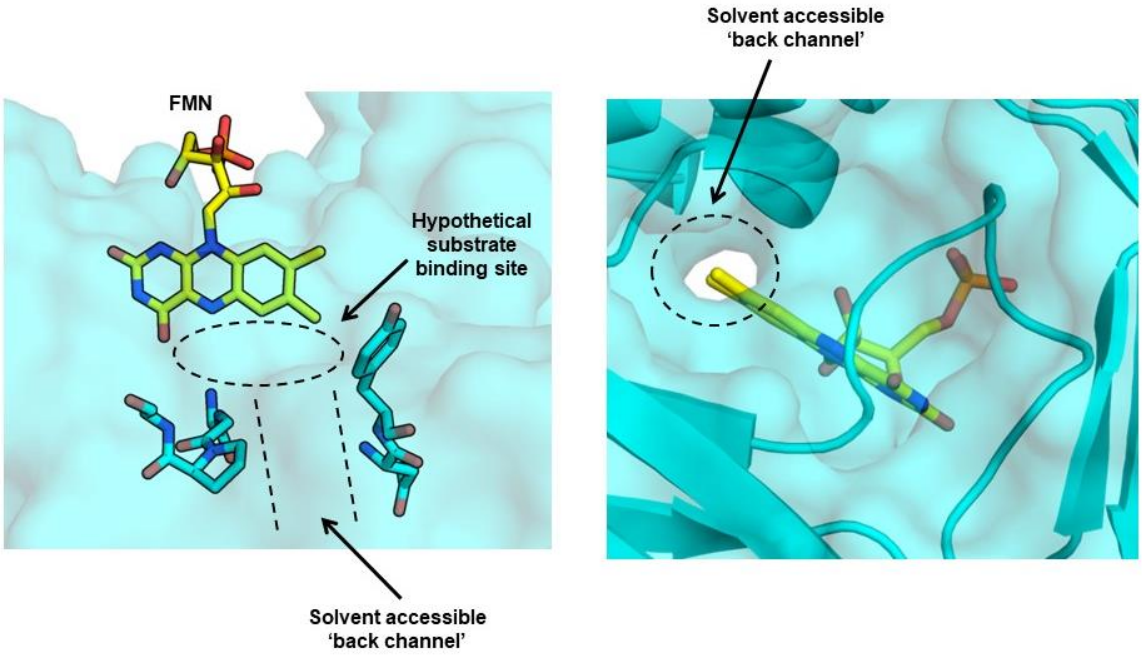


Figure 4.19 - Hypothetical substrate binding for redox chemistry at FMN-binding site.

major effect on the activity or stability of the other four FMN-binding GUSs characterized. Based on structural comparison with previously characterized FMN-binding proteins, it also does not appear to be similar to any oxidoreductases or electron transport-like proteins characterized to date. Future studies will be necessary to unravel the role this cofactor plays in GUS function, whether it is present for stability like that observed for *Fp2*GUS or impacts catalytic efficiency like that observed for some of the other FMN-binding GUSs characterized. Another possibility is that FMN may perform a completely different function from GUS activity. For example, these enzymes may have binding partners that can utilize FMN in ways that are not clear from studying these GUSs in an isolated system.

Conclusion

Here we characterize a unique set of FMN-binding GUS enzymes from the human gut microbiome. We determined the crystal structure of a GUS from *F. prausnitzii* and show that it binds FMN at a surface site approximately 30 Å from the active site. Using these structural data, we screened the HMP stool sample metagenomic database and identified 14 additional putative FMN-binders. We characterized four of these putative FMN-binding GUS enzymes in vitro and confirmed that they are bona fide FMN-binders, with binding affinities as low as 60 nanomolar. Site-directed mutagenesis of all five FMN-binders, and crystals structures of the FMN-binding GUS from *R. hominis* and *R. gnavus*, reveal a conserved FMN-binding site. Kinetic studies of the FMN-binding mutants suggest that the FMN-binding site is not necessary for GUS function, but mutations to this site can impact the K_M. Lastly, a structural bioinformatic search demonstrates that no other characterized FMN-binders interact with FMN like that observed with these FMN-binding glycoside hydrolases.

Methods

Gene synthesis, expression, and purification of FMN-binding GUSs

Genes for *Fp2*GUS, *Rh2*GUS, *Rg3*GUS, *Bv*GUS, and *Ri*GUS were synthesized by BioBasic, incorporated into a pLIC-His vector via ligation independent cloning, and resultant plasmids were transformed into BL21-G *E. coli* cells. Glycerol stocks were made from overnights and snap frozen and stored at -80 °C. Verification of successful transformation and sample integrity were determined by DNA sequencing.

Cultures of 100 mL LB with ampicillin were inoculated with glycerol stock and incubated overnight at 37 °C with shaking at 225 rpm. For protein expression, 50 mL of the overnight, approximately 40 µL Antifoam 204, and 750 µL of 2000x ampicillin were added to 1.5 L LB in a 2.5 L Erlenmeyer flask and incubated at 37 °C at 225 rpm. At an OD was approximately 0.6, the temperature was reduced to 18 °C and induced with IPTG (100 mM) at an OD of approximately 0.8 and incubated overnight with shaking at 225 rpm. Cultures were spun down in a Sorvall Instruments RC-3B centrifuge at 4500 x g for 25 minutes in 1 L round, flat bottom plastic bottles. Cultures were resuspended in LB and transferred to a 50 mL falcon tube and spun down in a ThermoScientific Sorvall ST 40R centrifuge for 15 minutes at 5000 x g. Supernatant was discarded and proteins were stored at -80 °C until purification.

Cell pellets were lysed in 30 mL Nickel A buffer (20 mM KH_2PO_4, 500 mM NaCl, 50 mM imidazole, pH 7.4) with DNase, lysozyme, and a Roche EDTA-free protease inhibitor tablet. The resultant cell slurry was sonicated on a Fischer Scientific Sonic dismembrator model 500 twice with 1 s pulses for 1.5 minutes. The resultant lysate was subsequently spun down on a Beckman Coulter J2-HC centrifuge for 1 hour at 17000 rpm. The supernatant was subject to filtration with a 0.22 µm filter prior to purification.

Protein was first purified with an Aktaxpress FPLC (Amersham Bioscience) via a Ni NTA column. Protein was eluted in one step using Nickel B buffer (20 mM KH_2PO_4, 500 mM NaCl, 500 mM imidazole, pH 7.4). Fractions were collected and concentrated with a 50K centrifilter at 3000 x g for 15 minutes if necessary before size exclusion chromatography. The eluent was then subject to size exclusion chromatography on a HiLoadTM 16/60 Superdex 200 gel filtration column. Size exclusion buffer was utilized for elution (20 mM HEPES, 50 mM NaCl, pH 7.4). Fractions were collected and an SDS-PAGE gel was performed to assess purity and stability of the enzyme. Protein concentration was determined on a ND-1000 spectrophotometer and then snap frozen in liquid nitrogen and stored at -80 °C.

Side-directed mutagenesis of FMN-binding GUSs

All mutants were created via site-directed mutagenesis. Mutagenesis primers were synthesized by Integrated DNA technologies. Mutant plasmids were sequenced by Eton Bioscience to confirm successful mutagenesis.

Absorbance scans of WT and mutant GUSs and stoichiometry determination

Absorbance scans of WT and mutant GUSs were determined in a BMG labtech PHERAstar plate reader. All proteins were analyzed at 50 µM in 96-well Costar half area, clear, flat bottom UV-transparent plates. Resultant absorbance profiles were plotted in Microsoft Excel 2013.

Liquid chromatography-mass spectrometry

Protein samples were diluted in sizing buffer to approximately 10 µM and applied to a 0.22 um filter prior to analysis. Separation was carried out on a Viva C4 5µm 150x2.1 mm column. Solvent A was 0.1% formic acid in water and solvent B was 0.1% formic acid in acetonitrile. FMN-binding proteins were eluted using a linear gradient of 5% solvent B to

60% B over 20 minutes, held for one minute, and then an additional linear gradient from 60% to 95% B for 17 minutes. Samples were analyzed using an Agilent Technologies 6520 Accurate-Mass Q-TOF LC-MS instrument in positive-ion mode and the resultant data were analyzed in MassHunter Qualitative Analysis B0.06.00 software.

Isothermal titration calorimetry to determine FMN binding affinity

ITC experiments were performed on a MicroCal AutoITC-200. All experiments were performed at 25 °C with 120 s intervals, reference heat of 7 kcals/s, and 20 injections total. Protein and ligand were prepared in size exclusion buffer and all ITC data were corrected with a control experiment of ligand dilution into buffer. Protein and ligand concentrations were varied depending on the amount of FMN present as determined from absorbance scans.

Crystal formation, preparation, and data collection of FMN-binding GUSs

*Fp2*GUS was crystallized via the sitting drop method in Hampton Research 3-well Crystallization Plates (Swissci) at 11.1 mg/ml in 0.2 M magnesium chloride, 10% w/v PEG 3000, and 0.1 M sodium cacodylate. Incubation at 20 °C resulted in crystal formation after 11 days. *Rg3*GUS was crystallized by the hanging drop vapor diffusion method at 20.4 mg/mL in 0.13 M magnesium acetate and 10% PEG 8000. Incubation at room temperature yielded crystals of *Rg3*GUS after 2-3 days. *Rh2*GUS was crystallized by the hanging drop vapor diffusion method at 14.9 mg/ml in 0.2 M calcium acetate and 50% PEG 8000 at room temperature after approximately 30 days. All crystals were looped and cryoprotected in their crystallant plus 20% glycerol before storing in liquid nitrogen.

X-ray diffraction experiments were performed at GM/CA ID-D and ID-B beam sources. Standard collection methods were followed and resultant data were reduced in XDS and scaled in aimless. The structure of *Fp2*GUS was solved via molecular replacement in Phenix using the single component MR-Phaser program with *B. fragilis* GUS (3CMG) as the search

model. *Rg3*GUS and *Rh2*GUS were solved by molecular replacement using the *Fp2*GUS structure as the search model. Final coordinates were deposited in the RCSB PDB with PDB codes 6MVF, 6MVG, and 6MVH for *Fp2*GUS, *Rg3*GUS, and *Rh2*GUS, respectively.

Bioinformatic analysis of HMP stool sample database for identification of FMN-binding GUSs

We utilized the previously generated database of GUS sequences from the HMP stool sample database to screen for putative FMN-binding GUSs [136]. To identify other FMN-binding GUSs, we first performed pairwise sequence alignments of each GUS sequence against *Fp2*GUS and rejected those with less than 25% sequence identity. The remaining sequences were screened to with a length requirement of 700-800 residues and contained the FMN-binding site residues (or similar residues) identified in *Fp2*GUS: D, E, or other small residue at position 151, Y, F or W at position 154, F, Y, or W at position 179, K or R at position 356, and Y, F, or W at position 363 (**Fig. S3**).

***In vitro* kinetic assay for k_{cat} and K_M determination of FMN-binding GUSs**

To assess the activities of WT and mutant GUSs, we measured their ability to hydrolyze the fluorescent reporter substrate 4-methylumbelliferyl glucuronide (4-MUG). Reactions were performed in black Costar 96-well plates with a flat, clear bottom and reaction volumes were as follows: 5 µL water, 5 µL buffer (25 mM HEPES, 25 mM NaCl pH 6.5 or 25 mM $NaCH_3CO_2^-$ and 25 mM NaCl pH 5.5), 5 µL FMN (25 µM for WT protein) or water (FMN-binding mutants), 5 µL GUS, and 30 µL 4-MUG (varying concentration). Reactions were initiated by addition of substrate and reactions were continuously monitored with excitation at 350 nm and emission at 450 nm in a BMG labtech PHERAstar plate reader. Initial velocities from the resultant data were fit by linear regression with a custom MATLAB program and k_{cat}, K_M, and k_{cat}/K_M were determined in SigmaPlot 13.0.

Accession numbers

Coordinates and structure factors have been deposited in the Protein Data Bank with accession numbers 6MVF, 6MVG, and 6MVH for *Fp2*GUS, *Rg3*GUS, and *Rh2*GUS, respectively.

CHAPTER 5: GUT MICROBIAL β-GLUCURONIDASE INHIBITION VIA CATALYTIC CYCLE INTERCEPTION.[5]

Glycosyl hydrolases (GHs) are abundant in bacterial and human systems and process a diverse set of molecules, ranging from sugar conjugates to complex polysaccharides. Mammalian GHs are associated with lysosomal storage disorders, viral infections, and Alzheimer's disease[14,142]. While human GHs play important roles in disease, the majority of GHs present in humans are located in the microbiota. Indeed, the gut microbiome encodes thousands of glycosyl hydrolases, whereas the human genome encodes only 97[89,136]. β-glucuronidases (GUSs) and other microbial enzymes are emerging as potential drug targets that can be selectively and potently modulated to improve cancer therapy and prevent heart disease[55,56,143]. The abundance and therapeutic importance of microbial enzymes in the mammalian host yields a rich space for drug discovery.

Bacterial β-glucuronidases (GUS) are key mediators of drug toxicity in the mammalian gut. The archetype of GUS-mediated drug toxicity is the dose-limiting diarrhea caused by irinotecan, a key anticancer drug primarily used to treat colon and pancreas cancers. Bacterial GUS enzymes in the GI tract catalyze the hydrolysis of SN-38-G, a glucuronic acid (GlcA) conjugate of the active form of irinotecan, (**Figure 5.1a**) generated

[5]This chapter previously appeared as an article in *ACS Central Science*. The original citation is as follows: **Pellock, S. J.**, Creekmore, B. C., Walton, W.G., Mehta, N., Biernat, K. A., Cesmat, A. P., Ariyarathna, Y., Dunn, Z. D., Li, B., Jin, J., James, L. I., Redinbo, M. R. (2018) Gut microbial β-glucuronidase inhibition via catalytic cycle interception, *ACS Cent. Sci. 4*, 868-879.

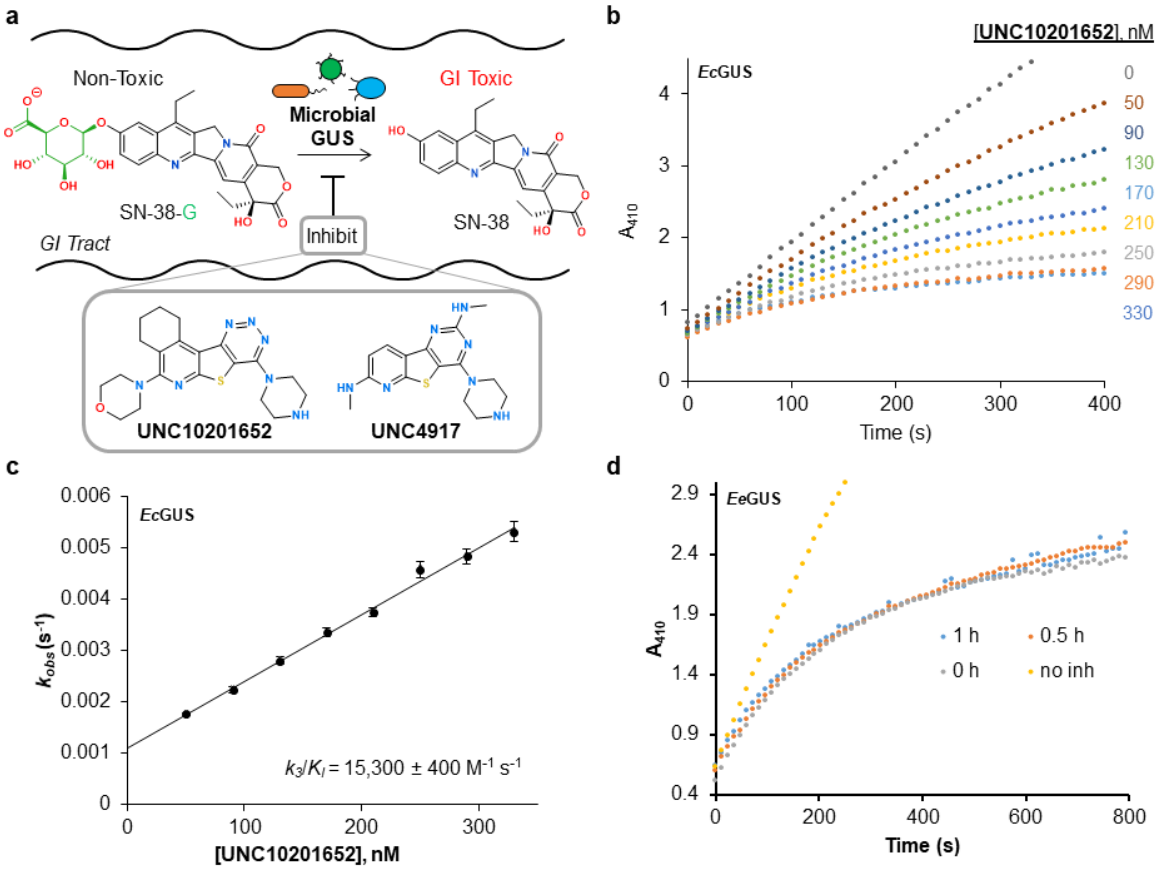


Figure 5.1 - Kinetic analysis of piperazine-containing GUS inhibitors reveals substrate-dependent slow-binding inhibition. (a) Conversion of SN-38-G to SN-38 is mediated by gut microbial GUS enzymes and promotes toxic side effects of this essential cancer therapeutic. Structures of piperazine-containing GUS inhibitors UNC10201652 and UNC4917 characterized in the present study. (b) Nonlinear progress curves of EcGUS activity in the presence of increasing concentrations of UNC10201652. (c) Secondary plot of kobs vs [UNC10201652] for EcGUS reveals one-step inhibition. (d) Preincubation of EeGUS with UNC4917 does not yield steady-state kinetics. Error bars represent SEM of N = 3 biological replicates, and progress curve plots are representative of N = 3 technical replicates.

by uridine-diphosphate glucuronosyl transferases (UGTs) in the liver and other first-pass protective tissues[10]. Glucuronides are generally non-toxic metabolites marked for excretion. However, when microbial GUS enzymes hydrolyze these glucuronides, they release the active drug (**Figure5.1a**, SN-38) into the intestinal lumen that can cause acute and dose-limiting GI toxicity[11]. Intestinal microbes utilize glucuronides as a carbon source; free GlcA can be metabolized via the Entner-Doudoroff pathway to generate pyruvate that enters the citric acid cycle[12]. While bacterial GUS enzymes have been linked to the GI toxicity of chemotherapeutics and NSAIDs (**Figure 5.1a**)[55,61], they may also be involved in carcinogenesis, inflammatory bowel diseases, and gall stone formation[67,9,23]. Thus, inhibiting microbial GUS enzymes may improve the tolerance and efficacy of current drugs, while also enabling the treatment or prevention of human disease.

Inhibitors of bacterial GUS have been developed to block the toxic GI side effects of important drugs. The natural product D-glucaro-1,4-lactone was the first reported GUS inhibitor, with a K_i of 19 μM against *Escherichia coli* GUS[144]. Sugar analogs resembling D-glucaro-1,4-lactone have also been synthesized, the most potent of which is *uronic*-noeurostegine (K_i = 60 nM against *E. coli* GUS)[145]. However, D-glucaro-1,4-lactone, *uronic*-noeurostegine, and similar synthetic analogs are also potent inhibitors of the essential human GUS ortholog, mutations of which cause the lethal lysosomal storage disease Sly syndrome[14,144].

The first inhibitors selective for bacterial GUS were reported in 2010 and have been further developed more recently, and they exhibit K_i values ranging from 2 μM to 164 nM against *E. coli* GUS[55,56,107]. These studies revealed that several previously-described inhibitors blocked GUS activity by binding to overlapping loops at the tetramer interface of

Loop 1 GUS enzymes that are absent in the human ortholog. These compounds have been shown to significantly reduce the diarrhea and ulcers caused by the anti-cancer drug irinotecan and NSAIDs, respectively[55,61,64,62]. Thus, gut microbial GUS enzymes can be inhibited both potently and selectively for therapeutic gain.

Here we describe piperazine-containing GUS inhibitors that are selective for microbial GUS enzymes and inhibit GUS via a striking mechanism – by intercepting the glycosyl-enzyme catalytic intermediate. Using kinetic studies, chemical biology, x-ray crystallography, and mass spectrometry, we demonstrate that these inhibitors intercept the covalent GUS-GlcA catalytic intermediate and are capable of forming covalent inhibitor-GlcA complexes in the GUS active site. Furthermore, we show that a range of clinically-approved piperazine-containing drugs of various therapeutic classes also inhibit bacterial GUS enzymes via the same mechanism-based interception. Taken together, these results advance our understanding of bacterial GUS inhibition and suggest that piperazine-containing drugs may affect non-human targets in the gut microbiome.

Results

UNC10201652 and UNC4917 are substrate-dependent slow-binding GUS inhibitors

UNC10201652 was identified in a high throughput screen using *E. coli* GUS[55], and **UNC4917** is a synthetic **UNC10201652** derivative (**Figure 5.1a, appendix 1**). We employed *in vitro* kinetic analysis to evaluate the potency and mechanism-of-action of **UNC10201652** and **UNC4917** against GUS enzymes from the gut microbiome. *In vitro* assays that assess the ability of GUS enzymes to cleave *p*-nitrophenyl-β-D-glucuronide (PNPG), producing chromogenic *p*-nitrophenol, were performed with purified GUS enzymes from four human GI-resident bacteria: *Escherichia coli* (*Ec*GUS), *Streptococcus agalactiae* (*Sa*GUS),

Clostridium perfringens (*Cp*GUS), and *Eubacterium eligens* (*Ee*GUS)[136]. *Ec*GUS, *Sa*GUS, and *Cp*GUS have been previously characterized[55,56], and are present in the GI microbiota, as is *Ee*GUS[136]. Each of these GUS enzymes are in the Loop 1 class, a group that makes up approximately 5% of unique GUS enzymes found in the human microbiome project (HMP) metagenomic database[136]. Sequence identities between these Loop 1 GUS enzymes range from 43 to 58% (**Figure 5.2a**).

GUS activities in the presence of nanomolar concentrations of **UNC10201652** and **UNC4917** displayed non-linear progress curves over the time course in which the uninhibited reaction remained linear (**Figure 5.1b**, **5.3**, and **5.4**). By contrast, linear progress curves under the same reaction conditions were observed with the previously characterized GUS inhibitor, Inhibitor 1 (**Figure 5.4d**). Non-linear progress curves indicate that **UNC10201652** and **UNC4917** are slow-binding inhibitors of microbial GUS enzymes[146]. Furthermore, steady-state velocities (v_s) in the presence of **UNC10201652** and **UNC4917** were either zero or nearly zero (*i.e.*, v_s approaches zero in **Figure 5.1b**, **5.3**, and **5.4a**, **b**, **c**), demonstrating that some enzyme-inhibitor pairs display enzyme inactivation. Taken together, these data reveal that **UNC10201652** and **UNC4917** display slow-binding kinetics and are capable of inactivating microbial GUS enzymes.

We extended our kinetic analysis to quantitate the onset of steady-state inhibition of bacterial GUS enzymes by **UNC10201652** and **UNC4917**. Plots of k_{obs} versus [**UNC10201652**] and [**UNC4917**] displayed one-step inhibition for all GUS enzymes tested (**Figure 5.1c**, **5.5a, 5.6a**, **Table 5.1**). One-step kinetics suggest that K_I, the equilibrium constant for initial enzyme-inhibitor complex formation, greatly exceeds the concentration of **UNC10201652** and **UNC4917** tested[146]. Indeed, the data fit well to a one-step inhibition

a

	*Ec*GUS	*Sa*GUS	*Cp*GUS	*Ee*GUS
***Ec*GUS**	100	42.9	47.7	46.1
***Sa*GUS**	42.9	100	53.0	47.3
***Cp*GUS**	47.7	53.0	100	58.1
***Ee*GUS**	46.1	47.3	58.1	100

b

```
          356                                  378
EcGUS   …GFNLSLGIGFEAGNKPKELYSEE…
SaGUS   …GLFQNFNASLDLSP---KDNGT-…
CpGUS   …GLHLNFMAT-GFGGDAP-KRDT-…
EeGUS   …GVNLQFGGGANFGGERIGTFDK-…
```

c

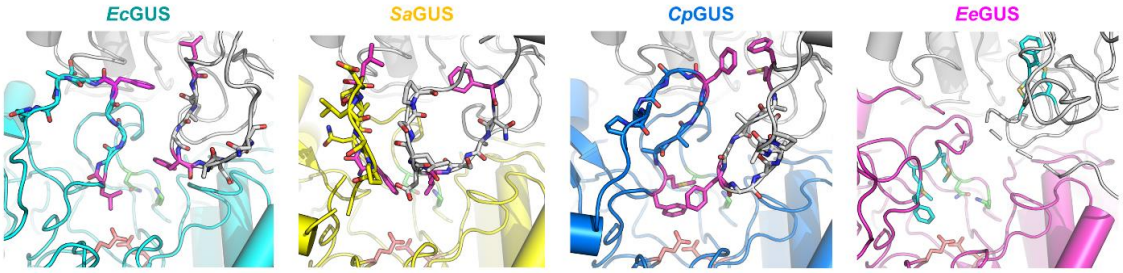


Figure 5.2 - Comparison of GUS enzymes analyzed in the present study. (**a**) Percent identity matrix of *Ec*GUS, *Sa*GUS, *Cp*GUS, and *Ee*GUS. (**b**) Sequence alignment of loop 1 sequences with key residues highlighted in magenta, glycines highlighted in orange, polar residues in blue, and start and end of disordered loop in *Sa*GUS italicized. (**c**) Active sites of *Ec*GUS, *Sa*GUS, *Cp*GUS, and *Ee*GUS with adjacent monomer loop in grey, key residues for contact of inhibitors highlighted in magenta. The loops shown are modeled for SaGUS, as this loop is disordered in the crystal structure.

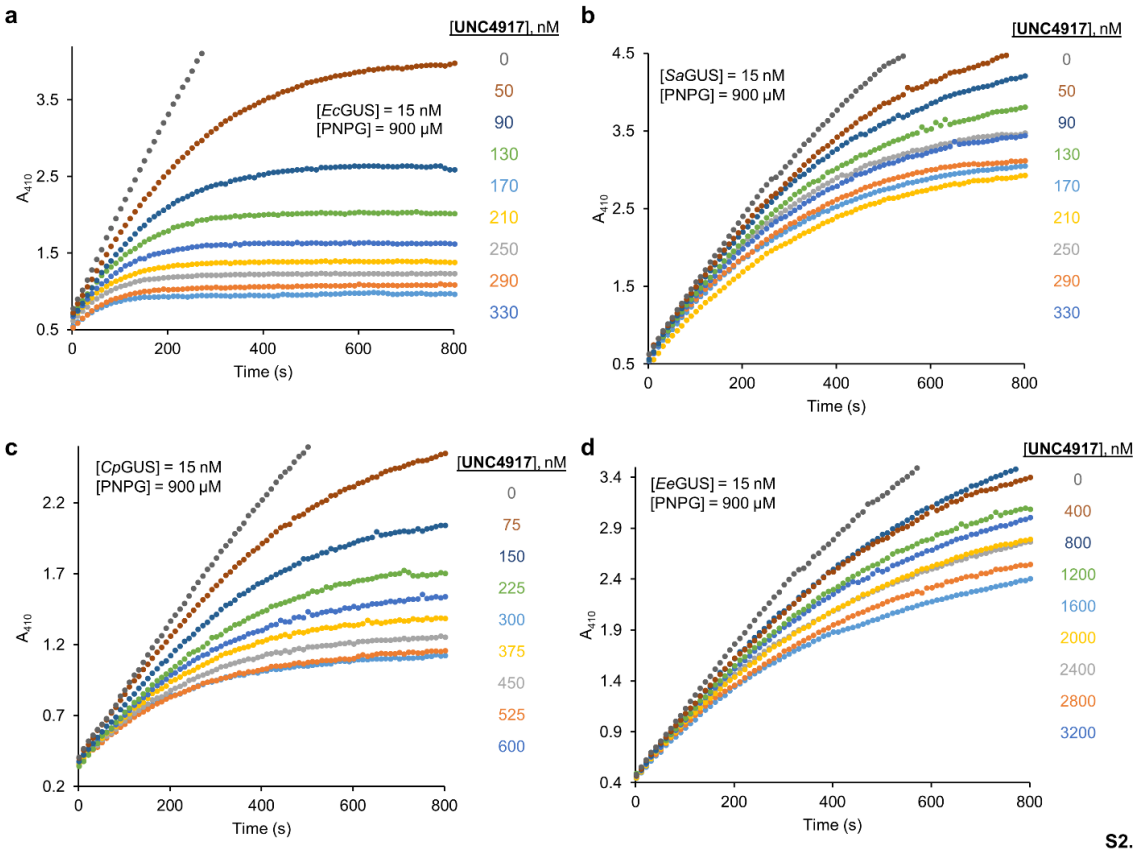


Figure 5.3. Bacterial GUS enzymes display non-linear progress curves in the presence of UNC4917. (**a**) Progress curves for *Ec*GUS in the presence of increasing concentrations of **UNC4917**. (**b**) Progress curves for *Sa*GUS in the presence of increasing concentrations of **UNC4917**. (**c**) Progress curves for *Cp*GUS in the presence of increasing concentrations of **UNC4917**. (**d**) Progress curves for *Ee*GUS in the presence of increasing concentrations of **UNC4917**. Progress curve graphs are representative of N = 3 biological replicates.

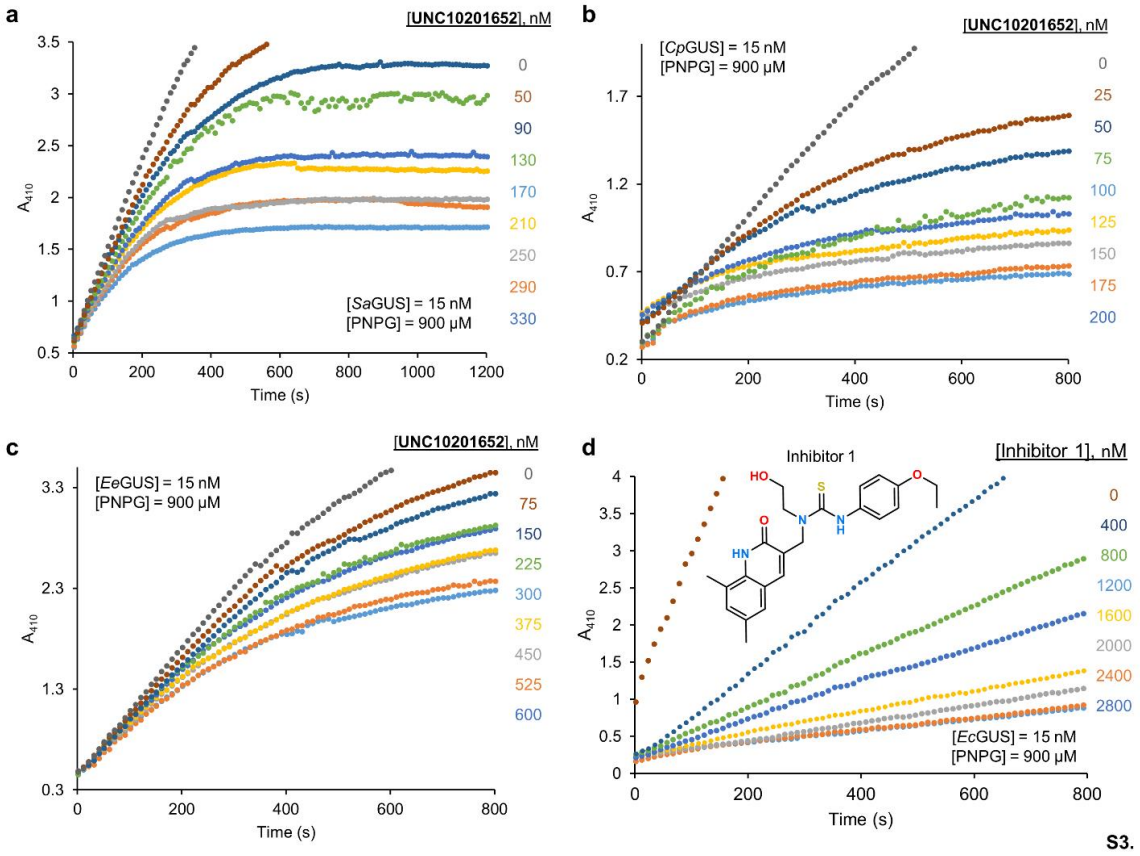


Figure 5.4 - Bacterial GUS enzymes display non-linear progress curves in the presence of UNC10201652. (**a**) Progress curves for *Sa*GUS in the presence of increasing concentrations of **UNC10201652**. (**b**) Progress curves for *Sa*GUS in the presence of increasing concentrations of **UNC10201652**. (**c**) Progress curves for *Cp*GUS in the presence of increasing concentrations of **UNC10201652**. (**d**) Progress curves for *Ec*GUS in the presence of increasing concentrations of Inhibitor 1. Progress curve graphs are representative of N = 3 biological replicates.

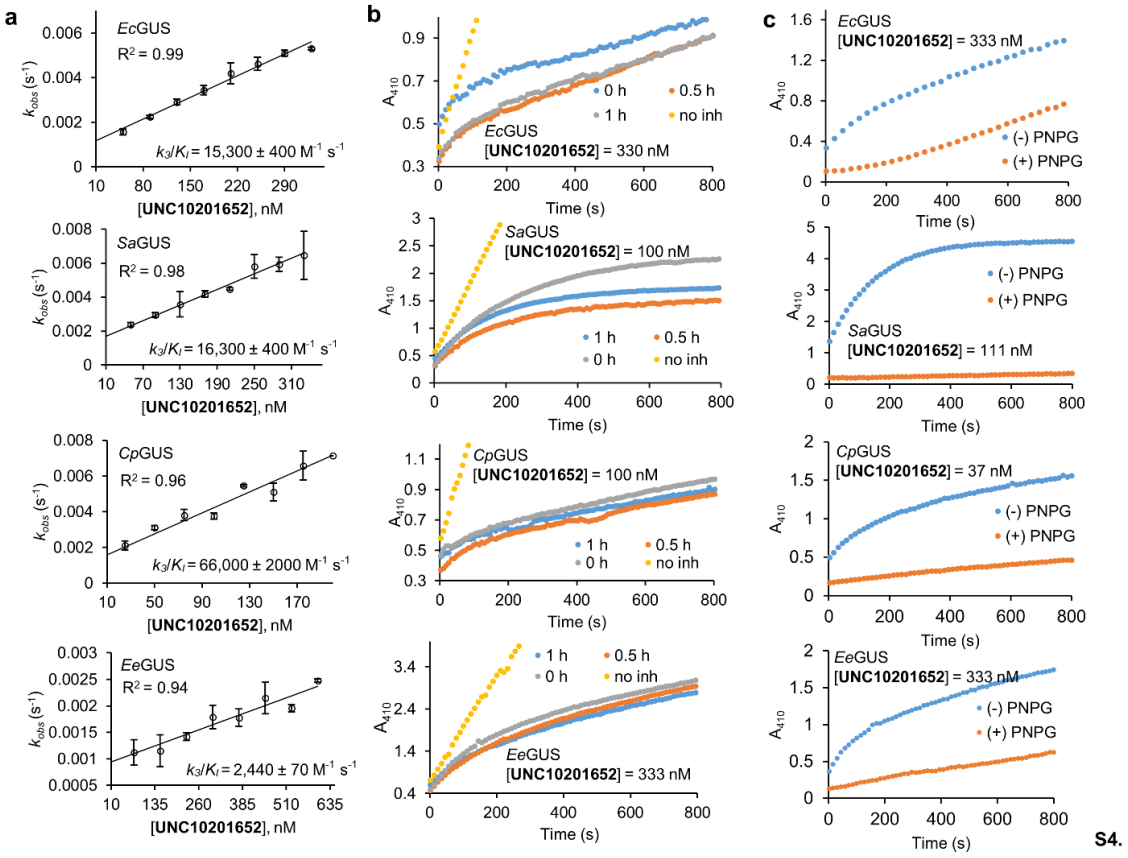


Figure 5.5 - Kinetic analysis of UNC10201652. (a) Plots of k_{obs} versus [**UNC10201652**] for *Ec*GUS, *Sa*GUS, *Cp*GUS, and *Ee*GUS. **(b)** Progress curves of *Ec*GUS, *Sa*GUS, *Cp*GUS, and *Ee*GUS after preincubation with **UNC10201652** for 0, 0.5, and 1h. **(c)** Progress curves of *Ec*GUS, *Sa*GUS, *Cp*GUS, and *Ee*GUS activity after jump-dilution of samples treated for 1 h with **UNC10201652** in the presence (+ PNPG) or absence (- PNPG) of substrate. Error bars represent SEM of n = 3 biological replicates and progress curves graphs are representative of n = 3 biological replicates.

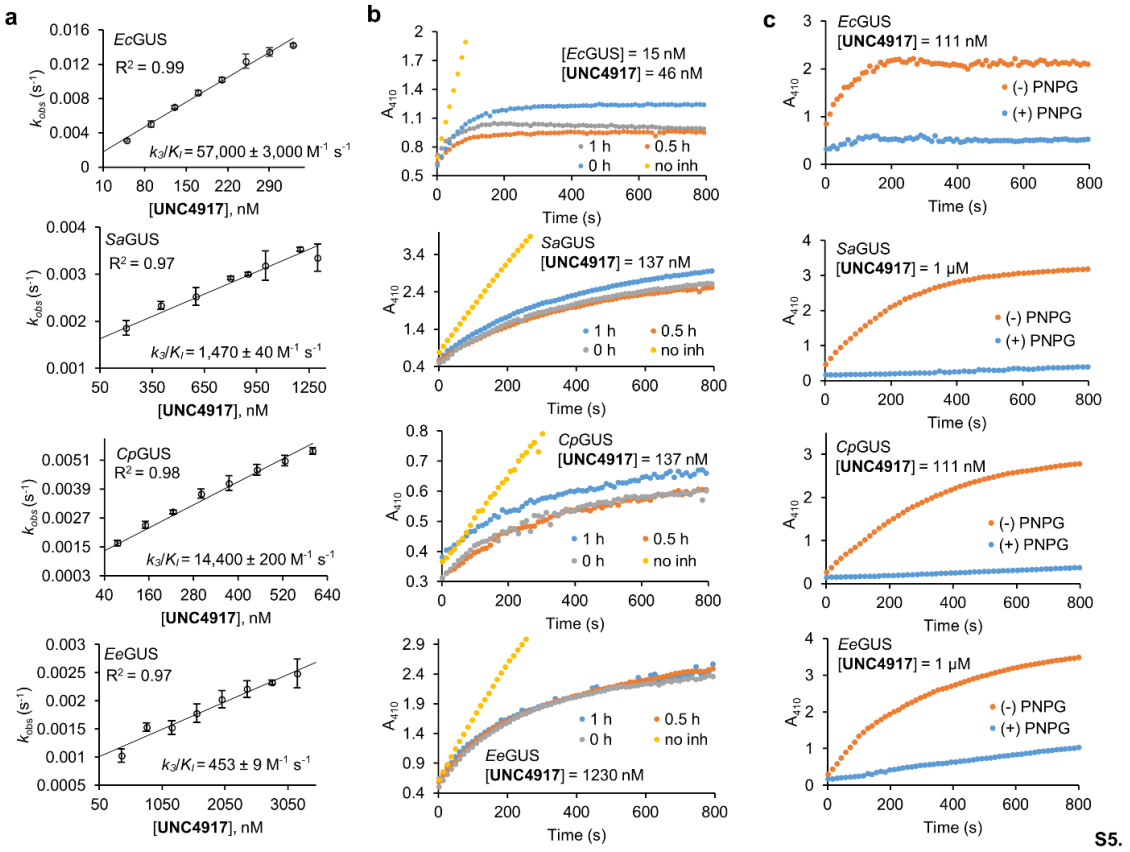


Figure 5.6 - Kinetic analysis of UNC4917. (a) Plots of k_{obs} versus [**UN4917**] for *Ec*GUS, *Sa*GUS, *Cp*GUS, and *Ee*GUS. **(b)** Progress curves of *Ec*GUS, *Sa*GUS, *Cp*GUS, and *Ee*GUS after preincubation with **UNC4917** for 0, 0.5, and 1h. **(c)** Progress curves of *Ec*GUS, *Sa*GUS, *Cp*GUS, and *Ee*GUS activity after jump-dilution of samples treated for 1 h with **UNC4917** in the presence (+ PNPG) or absence (- PNPG) of substrate. Error bars represent SEM of n = 3 biological replicates and progress curve graphs are representative of n = 3 biological replicates.

Table 5.1 - Kinetics of slow-binding inhibition for UNC10201652 and UNC4917 against *Ec*GUS, *Sa*GUS, *Cp*GUS, and *Ee*GUS.

Inhibitor	Enzyme	k_3/K_I ($M^{-1}s^{-1}$)	k_4 (s^{-1})
UNC10201652	*Ec*GUS	15300 ± 400	0.00114 ± 0.00007
	*Sa*GUS	16300 ± 400	0.00164 ± 0.00005
	*Cp*GUS	66000 ± 2000	0.0002 ± 0.0001
	*Ee*GUS	2440 ± 70	0.00085 ± 0.00001
UNC4917	*Ec*GUS	57000 ± 3000	0.00008 ± 0.00001
	*Sa*GUS	1470 ± 40	0.00123 ± 0.00008
	*Cp*GUS	14400 ± 200	0.00053 ± 0.00005
	*Ee*GUS	453 ± 9	0.00117 ± 0.00009

model that allowed us to determine the pseudo-second-order rate constant, k_3/K_I (see kinetic scheme in Methods). The resultant k_3/K_I values revealed that **UNC10201652** and **UNC4917** most efficiently inhibited *Cp*GUS (k_3/K_I = 66,000 ± 2,000 M^{-1} s^{-1}) and *Ec*GUS (k_3/K_I = 57,000 ± 3,000 M^{-1} s^{-1}), respectively (**Table 5.1**), and were weakest against *Ee*GUS, with k_3/K_I values of 2,440 ± 70 M^{-1} s^{-1} and 453 ± 9 M^{-1} s^{-1}, respectively. Rates of reactivation (k_4) ranged from 0.00164 ± 0.00005 s^{-1}(*Sa*GUS with **UNC10201652**; **Table 5.1**) to 0.00008 ± 0.00001 s^{-1}(*Ec*GUS and **UNC4917**; **Table 5.1**). Such rates corroborate the slow steady-state velocities observed in the presence of **UNC10201652** and **UNC4917** (*e.g.*, **Fig. 5.3** and **5.4a**, **b**, **c**). Collectively, these results demonstrate that the potency and onset of steady-state kinetics for **UNC10201652** and **UNC4917** vary with respect to the GUS enzyme examined, likely due to their different Loop 1 sequences (**Figure 5.2**).

Classically, non-linear progress curves indicate slow-binding or time-dependent inhibition, and slow-binding compounds typically yield enhanced potency when pre-incubated with their target[146]. Thus, we examined the time-dependent onset of steady-state inhibition by pre-incubating each GUS enzyme with **UNC10201652** and **UNC4917** for 0 h, 0.5 h, or 1 h before reaction initiation with PNPG. Surprisingly, in contrast to classic slow-binding inhibitors, which display a slower v_i followed by a faster v_s upon pre-incubation[146], we found that pre-incubation with **UNC10201652** and **UNC4917** displayed the same kinetic profile as seen in the absence of pre-incubation (**Figure 5.1d**, **5.5b**, **5.6b**). Thus, we conclude that the onset of steady-state kinetics by **UNC10201652** and **UNC4917** is not driven by inhibitor-enzyme interactions that occur prior to the addition of substrate.

The absence of pre-incubation effects has only been observed to date in cases where inhibitors require co-factor or substrate to initiate slow-binding[147,148]. Since characterized

bacterial GUS enzymes are not known to employ a co-factor, we considered that slow-binding inhibition by **UNC10201652** and **UNC4917** may be substrate-dependent. We pre-incubated each GUS for 1 h with **UNC10201652** or **UNC4917** either with or without PNPG, then jump diluted into PNPG-containing buffer to measure the enzyme activity. Indeed, we found that incubation of GUS with both inhibitor and PNPG resulted in the onset of steady-state inhibition, while incubation of GUS plus inhibitor without PNPG did not (**Figure 5.5c** and **5.6c**). These kinetic analyses indicate that **UNC10201652** and **UNC4917** are substrate-dependent inhibitors of gut microbial GUS enzymes.

Crystal structure reveals UNC4917-glucuronide conjugate in GUS active site

We next employed x-ray crystallography to determine the structural basis of the substrate-dependent onset of steady-state inhibition. First, we crystallized the GUS from *E. eligens* (*Ee*GUS) in both its apo (unliganded) and GlcA-bound states and refined the resultant structures to 2.9 Å and 2.7 Å resolution, respectively (**Figure 5.7a**, **b**). The *Ee*GUS-GlcA structure revealed that GlcA is well recognized by the enzyme's active site, with each sugar hydroxyl group contacting at least one protein side chain directly or via a bridging water molecule (**Figure 5.7b**). Second, we co-crystallized *Ee*GUS with both **UNC4917** and PNPG to mimic the *in vitro* assay conditions in which we observed substrate-dependent inhibition. Unbiased difference electron density within the *Ee*GUS active site of the resultant 2.7 Å resolution crystal structure indicated that both **UNC4917** and GlcA were bound to the enzyme (**Fig. 2a,b**). However, attempts to fit **UNC4917** and GlcA as separate entities within the electron density at the active site were unsuccessful due to significant clashes between the anomeric hydroxyl group of GlcA and the piperazine of **UNC4917**. Interestingly, a **UNC4917**-GlcA conjugate, in which the secondary nitrogen of the piperazine of **UNC4917**

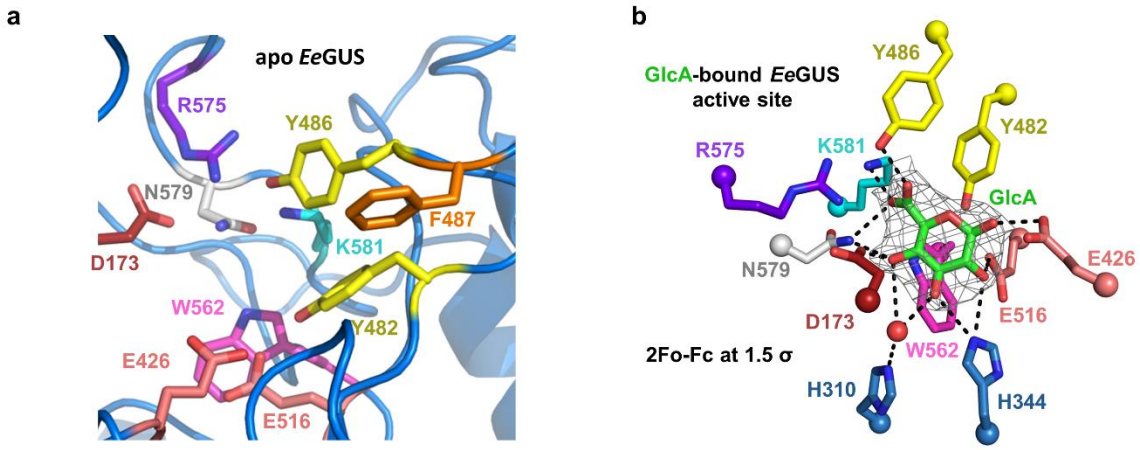


Figure 5.7 - Active site of apo EeGUS and EeGUS bound to GlcA. (**a**) Active site of *Ee*GUS (**b**) *Ee*GUS bound to GlcA showing 2Fo-Fc density at 1.5 σ.

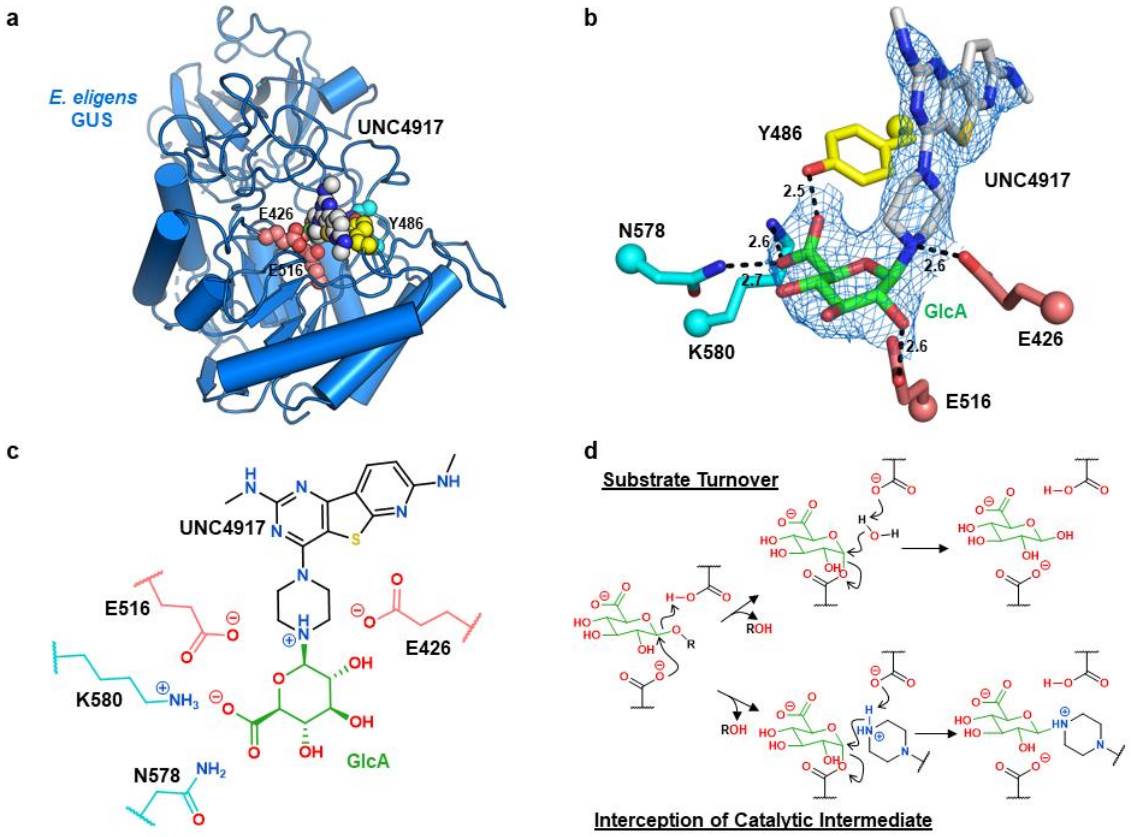


Figure 5.8 - Structural analysis of substrate-dependent slow-binding inhibition by UNC4917. (**a**) Overall structure of EeGUS−UNC4917−GlcA complex with inhibitor and key active site residues shown as spheres. (**b**) Active site of EeGUS bound to a UNC4917−GlcA conjugate with 2Fo-Fc density shown at 1.5 σ. Key contacts represented with black dotted lines and distances labeled in angstroms. (**c**) Chemical structure representation of UNC4917−GlcA conjugate bound to EeGUS active site. (**d**) Mechanism of substrate turnover (top) and proposed mechanism of inhibition by piperazine-containing GUS inhibitors (bottom).

was covalently β-linked to the anomeric carbon of GlcA, fit the density and refined well (**Figure 5.8b**). Thus, it appears that **UNC4917** is able to form a covalent bond with GlcA in the GUS active site (**Figure 5.8c**).

The **UNC4917**-GlcA-bound structure reveals a range of specific contacts formed between GlcA, **UNC4917**, and the *Ee*GUS active site. The secondary piperazine amine of **UNC4917** that appears to covalently link to GlcA forms a salt bridge with E426, the putative catalytic acid/base of bacterial GUS (**Figure 5.8b**, **c**). The aromatic scaffold of **UNC4917** participates in a π-π interaction with Y486, which is highly conserved in bacterial GUS enzymes (**Figure 5.8b**). In addition, as observed previously, the carboxylate of GlcA interacts with N578 and K580[56], as well as Y486 in *Ee*GUS (**Figure 5.8b**). Taken together, these structural data reveal that piperazine-containing microbial GUS inhibitors target the GUS-GlcA catalytic intermediate.

Based on the substrate-dependent onset of steady-state kinetics and the presence of a **UNC4917**-GlcA conjugate in the *Ee*GUS active site, we hypothesized that the piperazine-containing compounds **UNC10201652** and **UNC4917** may function as mechanism-based inhibitors of bacterial GUS. During the GUS catalytic cycle, a GUS-GlcA covalent intermediate is formed between the anomeric carbon of GlcA and the catalytic glutamate nucleophile (E516 in *Ee*GUS; **Figure 5.8d**). The catalytic acid/base (E426 in *Ee*GUS) then deprotonates a water molecule that subsequently hydrolyzes the E516-GlcA bond, releasing GlcA and regenerating GUS (**Fig 5.8d**, top). We propose that **UNC10201652** and **UNC4917** disrupt substrate turnover by intercepting the GUS-GlcA catalytic intermediate (**Figure 5.8d**, bottom), and that these compounds are deprotonated by the catalytic acid/base (*e.g.*, E426) and attack the anomeric carbon of the GUS-GlcA intermediate. This mechanism would yield

the inhibitor-GlcA conjugate observed in the crystal structure of *Ee*GUS outlined above (**Figure 5.8d**) and explain why the onset of steady-state inhibition by **UNC10201652** and **UNC4917** is substrate-dependent.

LC-MS confirms GUS-dependent formation of inhibitor glucuronides

To confirm the formation of an inhibitor-glucuronide conjugate, GUS enzymes were incubated with **UNC10201652** and PNPG, then heat denatured to promote the release of tightly-bound glucuronide conjugates. The products were analyzed by liquid chromatography-mass spectrometry (LC-MS). The mass for the covalent **UNC10201652**-GlcA conjugate was observed with each of the four enzymes tested, *Ee*GUS, *Ec*GUS, *Sa*GUS, and *Cp*GUS, when incubated with both **UNC10201652** and PNPG (**Figure 5.9a, b**). Similarly, incubation of the same four GUS enzymes with **UNC4917** and PNPG also yielded the corresponding glucuronide conjugate (**Figure 5.9c**). Importantly, we did not observe glucuronide formation in the absence of GUS, suggesting that glucuronide formation is GUS-dependent (**Figure 5.9b**, **c**, **d**). Lastly, each GUS enzyme also generated a **UNC10201652**-GlcA conjugate when incubated with SN-38-G (**Figure 5.1a**), the glucuronide of irinotecan's active metabolite (**Figure 5.9d**). These data indicate that GUS-mediated formation of **UNC10201652**-GlcA is aglycone-independent. Thus, LC-MS supports the conclusion that piperazine-containing compounds **UNC10201652** and **UNC4917** are capable of forming covalent inhibitor-GlcA conjugates within the active sites of GUS enzymes from the human gut.

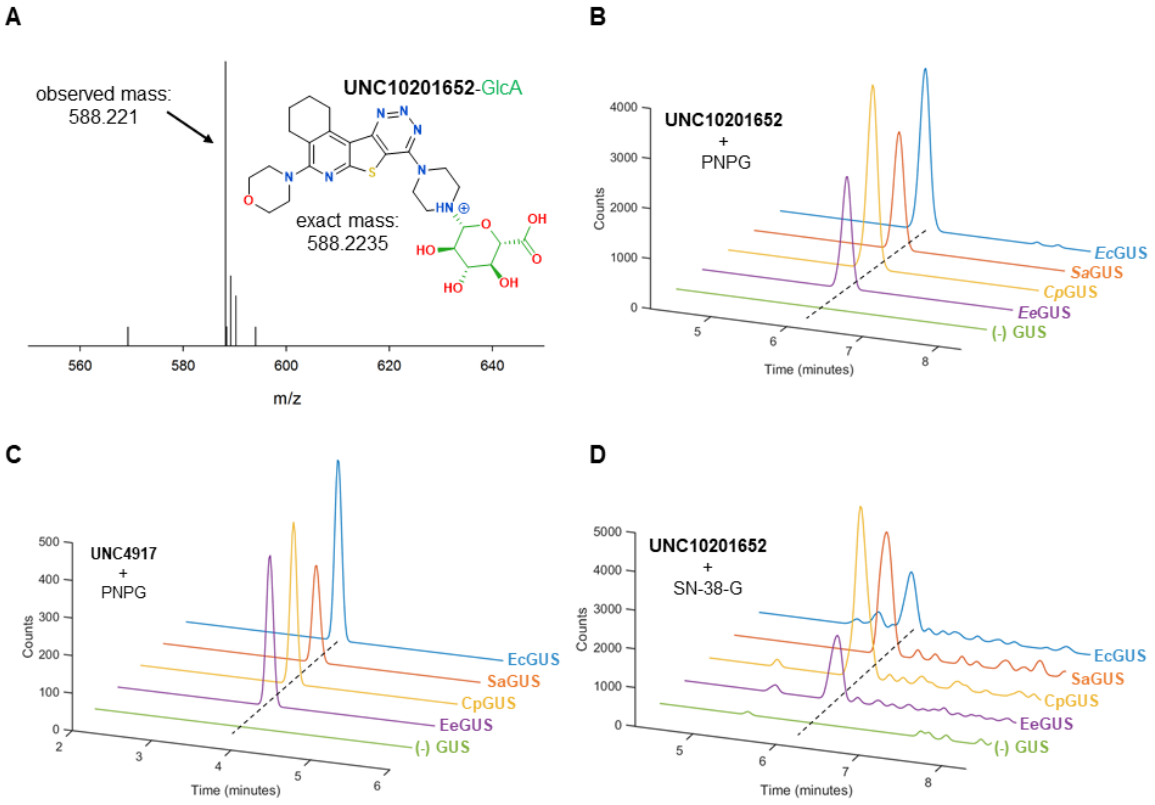


Figure 5.9 - LC−MS confirms GUS-dependent generation of inhibitor glucuronide conjugates. (**A**) Mass spectrum of a UNC10201652−GlcA conjugate (exact mass, 588.2235 m/z; observed mass, 588.221 m/z) generated by incubation of EeGUS with PNPG and UNC10201652. (**B**) Extracted ion chromatograms (588.2235 m/z) of each GUS treated with both UNC10201652 and PNPG as well as a no GUS control. (**C**) Extracted ion chromatograms (506.1816 m/z) of each GUS treated with UNC4917 and PNPG as well as a (−) GUS control. (**D**) Extracted ion chromatograms (588.2235 m/z) of each GUS treated with UNC10201652 and SN-38-G as well as a (−) GUS control. Plots are representative of N = 2 biological replicates.

Chemically synthesized UNC10201652-GlcA conjugate is a weak GUS inhibitor

Crystallographic and LC-MS data indicated that **UNC4917**-GlcA and **UNC10201652**-GlcA conjugates are capable of forming in the GUS active site, in turn serving as a potent GUS inhibitor. Thus, we tested whether administration of chemically synthesized **UNC10201652**-GlcA would potently inhibit bacterial GUS enzymes. **UNC10201652**-GlcA was synthesized from **UNC10201652** and GlcA in the presence of a catalytic amount of glacial acetic acid in methanol (**Figure 5.10a**). This afforded the product (**UNC5670**) as an inseparable, 1:1 mixture of α:β diastereomers (**Supplementary Synthesis and Characterization**). This anomeric mixture of **UNC5670** yielded weaker inhibition than **UNC10201652** against all GUS enzymes tested, exhibiting potencies between 1.8 μM and 18 μM (**Figure 5.10b**). Interestingly, **UNC5670** still displayed non-linear progress curves, suggesting that glucuronide formation is not the rate-limiting step for the onset of steady-state inhibition (**Figure 5.10c**). Furthermore, we found that **UNC5670** is not cleaved by *E. coli* GUS (**Figure 5.10d**). Together, these data reveal that an anomerically impure **UNC10201652**-GlcA conjugate, **UNC5670**, displays slow-onset of steady-state inhibition and is a much weaker inhibitor than the aglycone **UNC10201652**.

Piperazine amine is essential for potent bacterial GUS inhibition

To determine the role of the piperazine for both potency and kinetics of GUS inhibition, we performed a focused structure activity relationship on the secondary piperazine amine that appears to covalently link to GlcA in the GUS active site. First, we synthesized a dimethylated analog of **UNC10201652** to maintain the positive charge but remove its ability to act as a nucleophile (**UNC5671**; **Figure 5.11a**). **UNC5671** exhibited an IC_{50} of 9.6 ± 0.2 μM μM, **Figure 5.11b**) and 120-fold weaker than **UNC4917** (IC_{50} = 80 ± 1 nM).

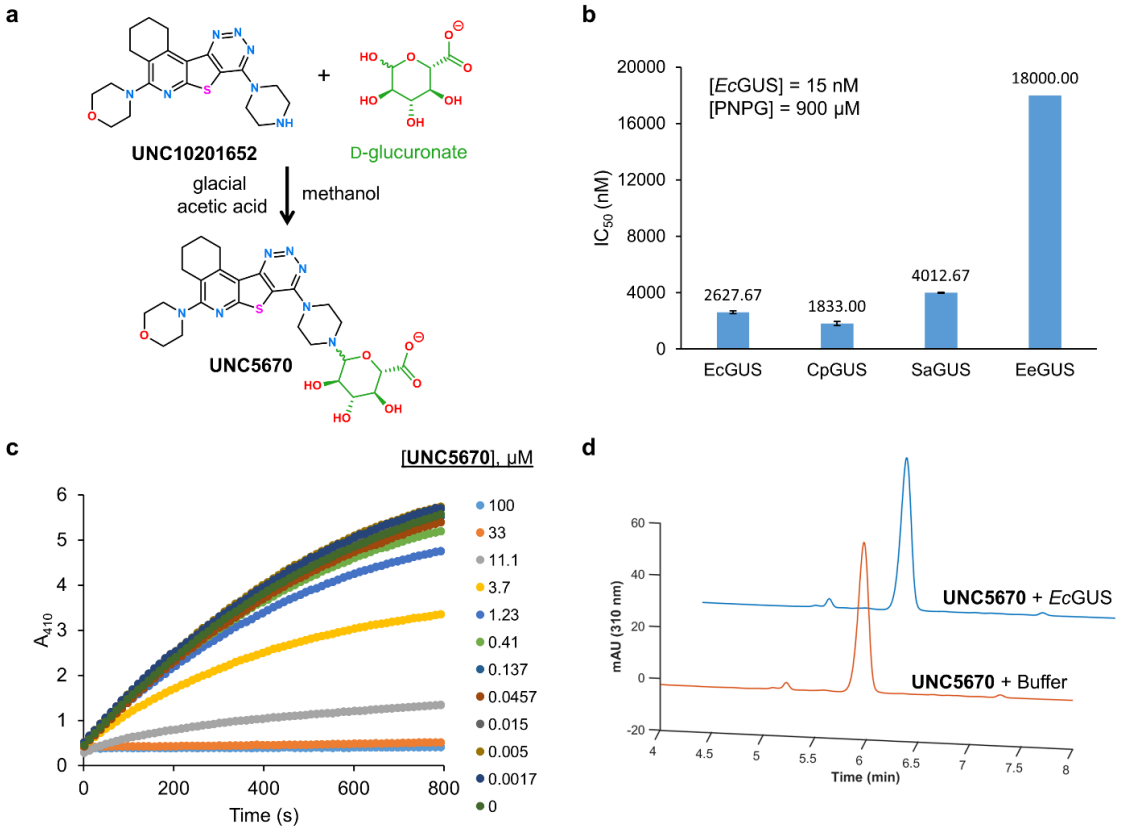


Figure 5.10 - Synthesis and characterization of synthetic UNC10201652-GlcA conjugate (UNC5670). (**a**) Synthetic scheme for generation of glucuronide conjugate of **UNC10201652** (**UNC5670**). (**b**) IC_{50} values for inhibition of each GUS enzyme with **UNC5670**. (**c**) Progress curves of *Ec*GUS in the presence of increasing concentration of UNC5670 reveals non-linear progress curves. (**d**) LC traces of **UNC5670** treated with EcGUS and a buffer control reveal absence of **UNC5670** hydrolysis by *Ec*GUS.

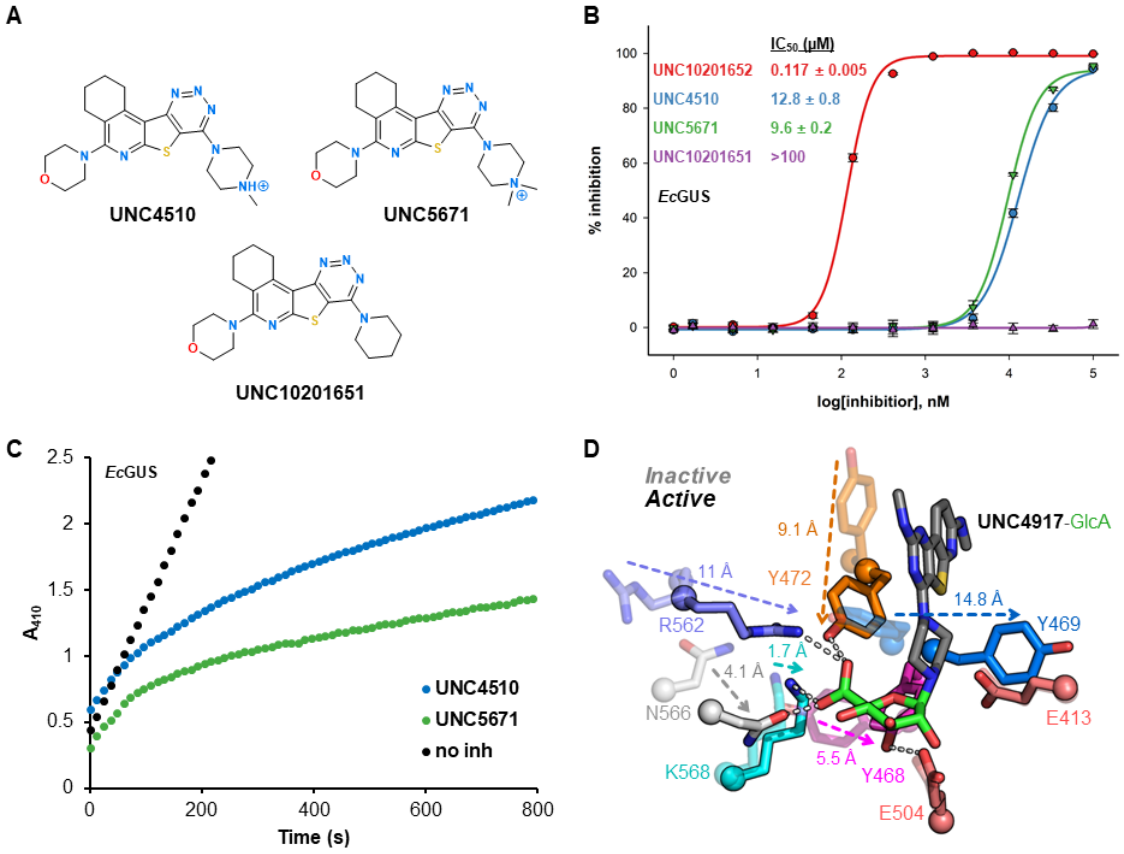


Figure 5.11 Focused SAR reveals key role of piperazine for potent GUS inhibition and demonstrates that glucuronide formation is not necessary to yield slow-binding inhibition. (**A**) Structures of piperazine analogs UNC4510, UNC5671, and UNC10201651. (**B**) IC50 plots for inhibition of EcGUS by parent compound (UNC10201652) and piperazine analogs reveal significantly reduced potencies. (**C**) EcGUS displays nonlinear progress curves in the presence of piperazine analogs UNC4510 and UNC5671. (**D**) UNC4917−GlcA conjugate observed in EeGUS modeled in the Active conformation (PDB: 3LPF) and Inactive conformation (PDB: 3K46) of EcGUS. Plots are representative of N = 3 biological replicates.

Second, a less sterically demanding monomethyl analog of **UNC10201652** was synthesized (**UNC4510; Figure 5.11a**); this compound displayed a ~100-fold weaker IC_{50} than **UNC10201652** against *Ec*GUS (IC_{50} = 12.8 ± 0.8 μM; **Figure 5.11b)**. While these analogs displayed markedly weaker potency than **UNC10201652**, they were similar in potency to the previously characterized Inhibitor 1 (IC_{50} = 8.5 ± 0.7 μM). Finally, a piperidine analog of **UNC10201652** that replaces the secondary nitrogen with a carbon (**UNC10201651**; **Figure 5.11a**) yielded no inhibition up to 100 μM, the maximum concentration we could test (**Figure 5.11b**). Together, these analogs pinpoint the piperazine amine as the essential warhead for potent inhibition of bacterial GUS enzymes.

While the analogs outlined above display markedly reduced potency, they still yield non-linear progress curves (**Figure 5.11c**, **5.12**). **UNC4510** and **UNC5671** display slow-binding efficiencies (k_3/K_I) of 800 ± 100 M^{-1} s^{-1} and 960 ± 70 M^{-1} s^{-1}, respectively (**Figure 5.12b**, **d**), compared to 15,300 ± 400 M^{-1} s^{-1} for **UNC10201652**. Together, these data suggest that the ability of the piperazine to act as a nucleophile on the glycosyl-enzyme catalytic intermediate is not necessary to yield the slow-onset of steady-state GUS inhibition, but is crucial for potent inhibition of gut microbial GUS enzymes.

Slow-onset steady-state kinetics and active site conformational changes

Since both non-nucleophilic piperazine analogs (**UNC4510** and **UNC5671**) and the glucuronide of **UNC10201652** (**UNC5670**) displayed slow-binding inhibition of GUS, we considered that conformational changes at the GUS active site may be responsible for the

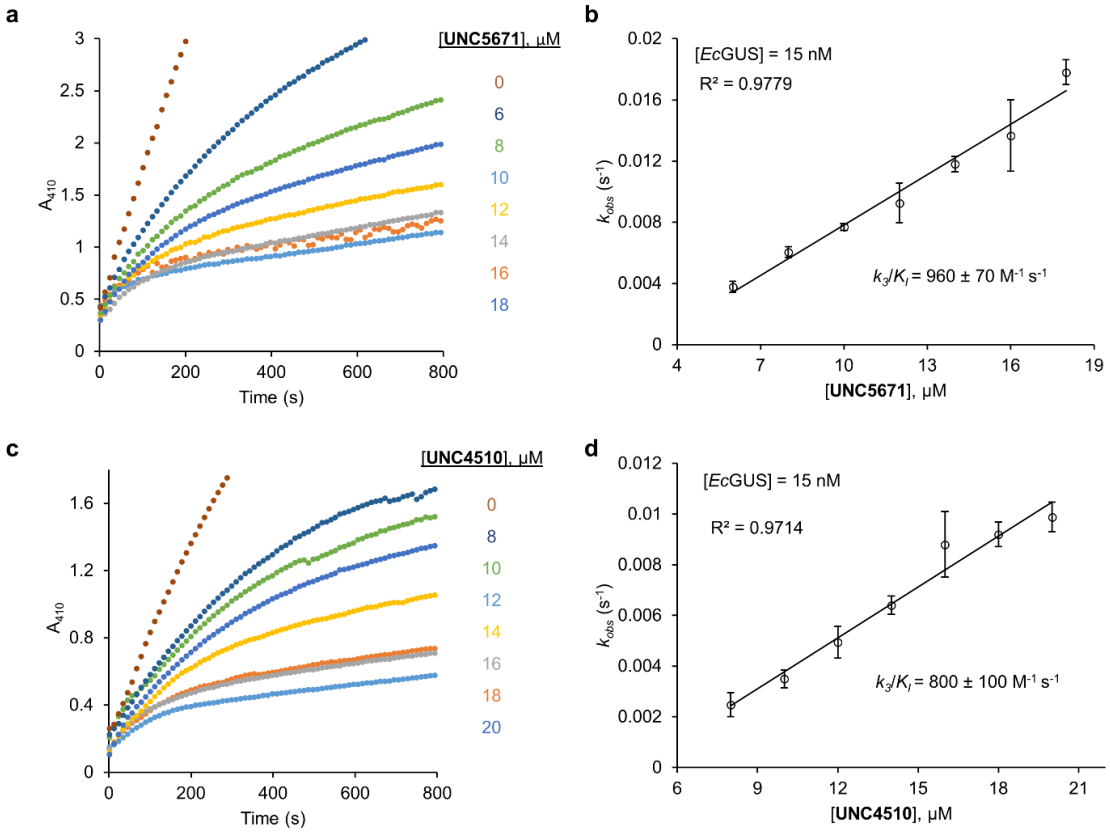


Figure 5.12 - Slow-binding kinetic analysis of UNC5671 and UNC4510. (a) Progress curves for *E. coli* GUS in the presence of increasing concentrations of **UNC5671**. **(b)** Plot of k_{obs} versus [**UNC5671**] **(c)** Progress curves for *E. coli* GUS in the presence of increasing concentrations of **UNC4510**. **(d)** Secondary plot of k_{obs} versus [**UNC4510**]. Error bars represent SEM of n = 3 biological replicates.

slow-binding behavior observed. Previously elucidated structures of *E. coli* GUS reveal that two conformations are available to the GUS active site, Active and Inactive (**Figure 5.11d**)[56]. In the Active conformation, Y472, R562, and the N-K motif of N566 and K568 form direct contacts with the GlcA carboxylate. In the Inactive state, all four contacts are lost (**Figure 5.11d**). Additional active site changes observed between the Active and Inactive conformations include 15 Å, 9 Å, and 6 Å shifts in position by three active site tyrosine residues, Y469, Y472, and Y468, respectively (**Figure 5.11d**). The loss of key contacts with the substrate in the Inactive conformation suggests that glucuronides are only recognized and hydrolyzed when bacterial GUS adopts the Active conformation. The Active state is also more favorable for the recognition of the planar, non-polar scaffold of **UNC10201652** and **UNC4917** (**Figure 5.11d)**. Indeed, *E. eligens* GUS is in the Active conformation in the GlcA-complexed structures presented here (**Fig. 2b**), and **UNC4917** forms edge-face π-π interactions with Y472 (**Figure 5.11d**). Thus, we propose that substrate binding and catalysis induces a conformational change at the GUS active site to form the Active state, to which **UNC10201652** and **UNC4917** preferentially bind.

To test this conformational hypothesis, we mutated Y472 and Y485 in *Ec*GUS and *Ee*GUS, respectively, to either alanine or phenylalanine. The resultant variant proteins, however, displayed such weak activity that we were unable to assess GUS inhibition (**Figure 5.13**). Indeed, these mutations highlight the essential role played by this conserved tyrosine in GUS activity, likely due to its hydrogen bond to the lysine of NxK motif as well as its direct contact to the carboxylate of glucuronic acid (**Figure 5.7b**).

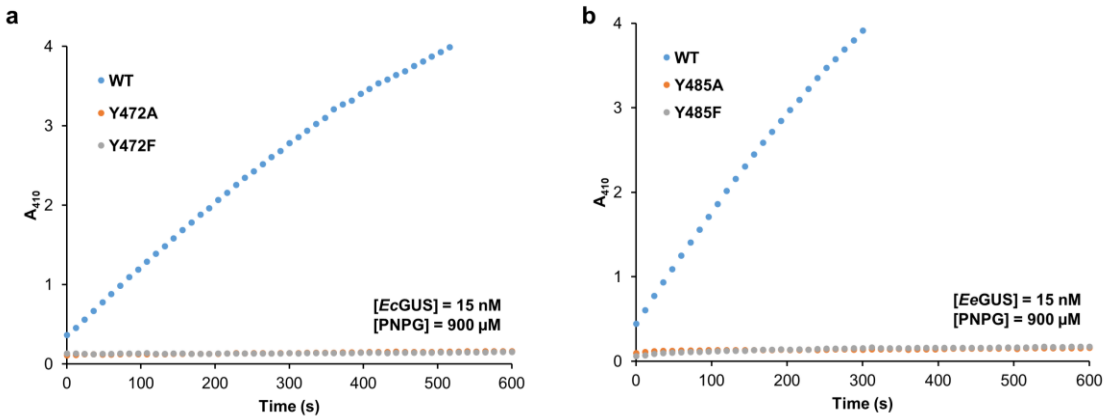


Figure 5.13 - Kinetic analysis of Y472A/F and Y485A/F mutants of EcGUS and EeGUS. (**a**) Progress curves of WT, Y472A, and Y472F *Ec*GUS. (**b**) Progress curves of WT, Y485A, and Y485F *Ee*GUS.

Piperazine and piperidine-containing drugs act as substrate-dependent GUS inhibitors

We have demonstrated that the secondary piperazine amine of **UNC10201652** and related compounds is essential for potent bacterial GUS inhibition. Furthermore, previous studies showed that two clinically-approved piperazine-containing drugs, the antipsychotic amoxapine and the antibiotic ciprofloxacin (**Figure 5.14a**), were capable of inhibiting bacterial GUS and were effective *in vivo* at reducing the toxic side effects of irinotecan[149,150]. Thus, we hypothesized that a range of structurally distinct piperazine-containing therapeutics may function as microbial GUS inhibitors by intercepting the catalytic cycle as outlined above. Five drugs were selected for evaluation: the previously reported amoxapine and ciprofloxacin, as well as palbociclib, a CDK4 inhibitor for ER-positive breast cancer, crizotinib, an anaplastic lymphoma kinase (ALK) and ROS1 kinase inhibitor for non-small cell lung carcinoma and lymphoma that contains a piperidine instead of a piperazine, and the antidepressant vortioxetine (**Figure 5.14a**)[151,152,153]. We found that all five drugs inhibited *Ec*GUS in a substrate-dependent manner (**Figure 5.14c**, **5.15**, **5.16**). A range of potencies were observed, with amoxapine demonstrating the strongest inhibition (IC_{50} = 0.53 ± 0.01 μM) and ciprofloxacin the weakest (IC_{50} = 9 ± 1 μM) (**Figure 5.14b**). We also determined a 2.9 Å resolution crystal structure of *Ee*GUS crystallized in the presence of amoxapine and PNPG and observed a covalent amoxapine-GlcA conjugate at the active site (**Figure 5.14d**). Taken together, these results reveal that diverse chemical scaffolds containing a piperazine or piperidine with a secondary amine inhibit bacterial GUS by intercepting a catalytic intermediate. They further demonstrate that a range of currently-approved human therapeutics may have significant off-target effects through their ability to inhibit bacterial GUS enzymes expressed by the human gut microbiota.

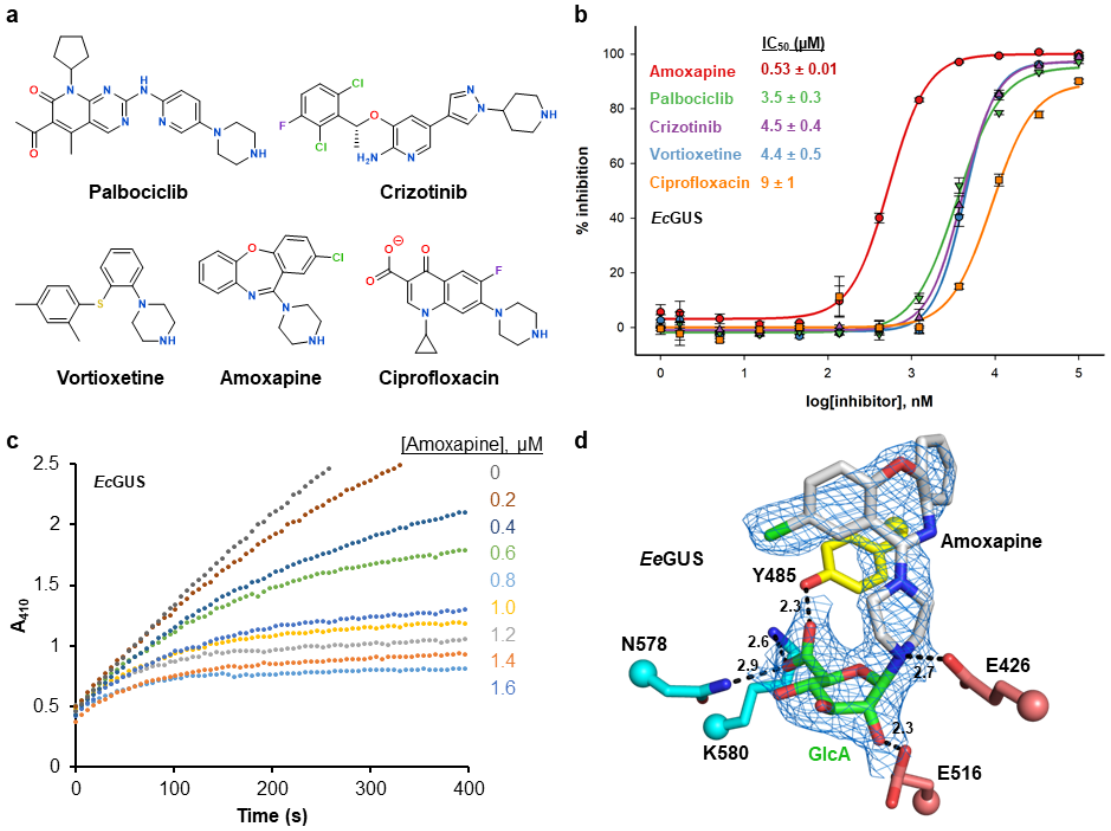


Figure 5.14 - Approved piperazine-/piperidine-containing drugs inhibit GUS in a substrate-dependent slow-binding manner. (**a**) Structures of piperazine- and piperidine-containing drugs tested for substrate-dependent slow-binding inhibition. (**b**) IC50 plots for inhibition of EcGUS by approved piperazine- and piperidine-containing drugs. (**c**) Progress curves of EcGUS in the presence of increasing concentrations of amoxapine display slow-binding characteristics. (**d**) Active site of EeGUS bound to an amoxapine−GlcA conjugate with 2Fo-Fc density shown at 1 σ. Progress curve plots are representative of N = 3 biological replicates.

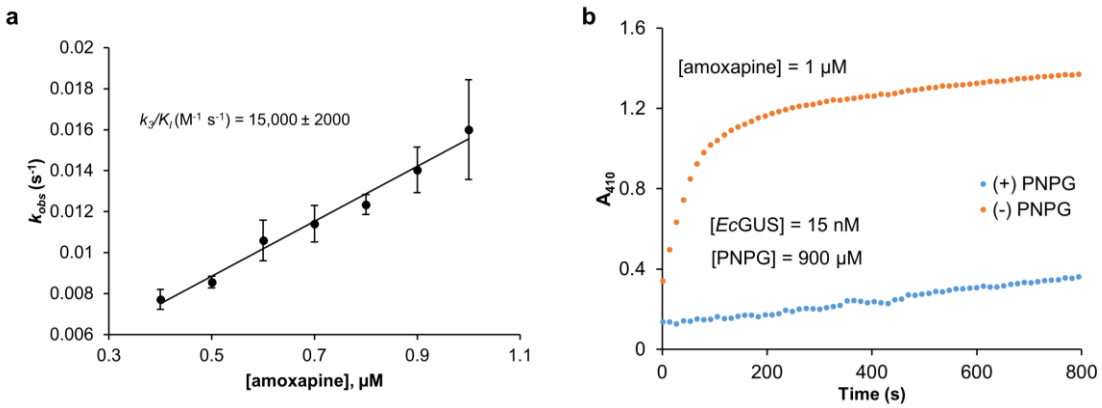
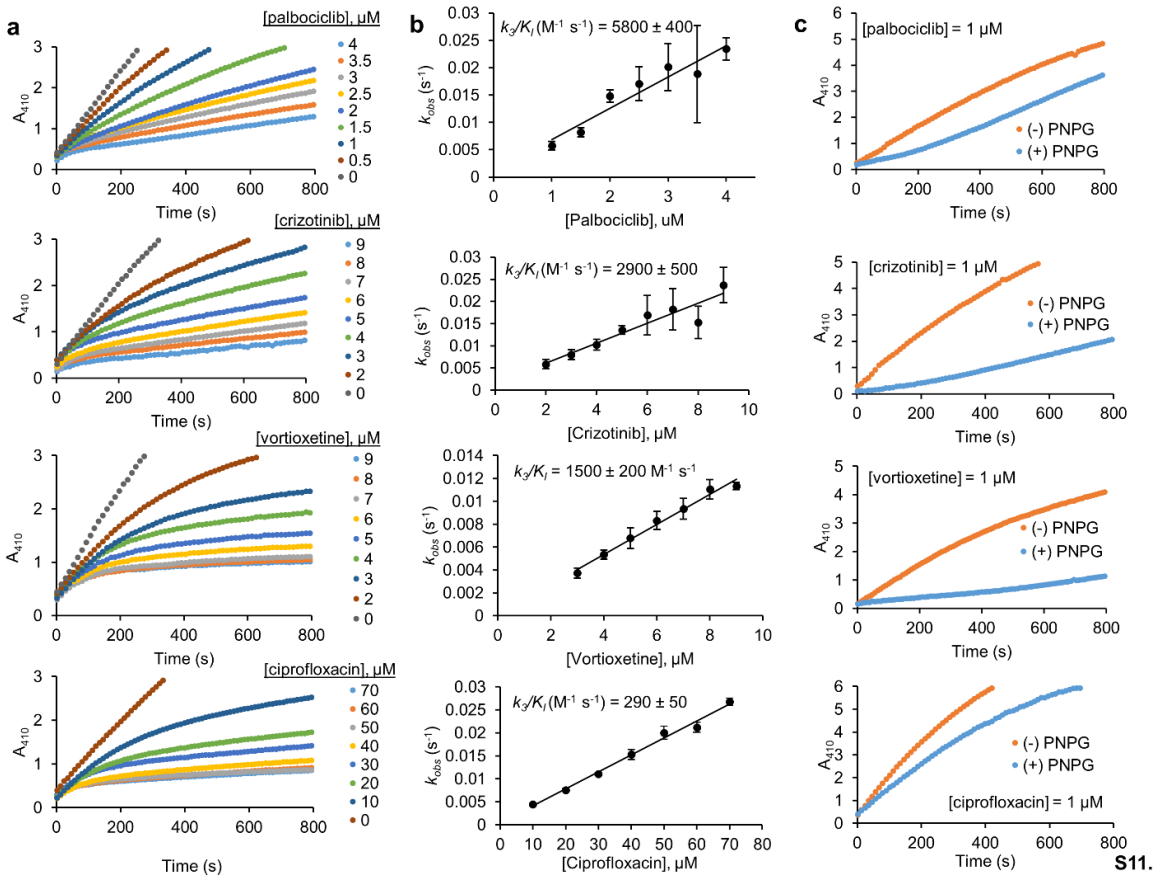


Figure 5.15 - Kinetic analysis of amoxapine with *Ec*GUS. (**a**) Secondary plot of k_{obs} vs [amoxapine]. (**b**) Jump dilution of pre-incubation with or without PNPG.



S11.

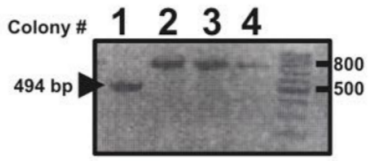
Figure 5.16 - Kinetic analysis of approved piperazine/piperidine-containing drugs. (**a**) Progress curves of *Ec*GUS activity in the presence of increasing concentrations of approved drugs. (**b**) Secondary plots of k_{obs} vs inhibitior concentration for *Ec*GUS with approved drugs. (**c**) Progress curves of GUS activity after incubation with drug for one hour in the presence (+ PNPG) or absence (- PNPG) of substrate. Plots are representative of N =3 biological replicates and error bars represent SEM of N = 3 biological replicates.

In-cell potency and selectivity of UNC10201652 and UNC4917

To determine if **UNC10201652** and **UNC4917** demonstrate potent on-target activity in cells, we examined GUS inhibition in wild-type (WT) *E. coli* K-12 MG1655 cells and in a variant of this strain in which the *gus* gene was truncated to remove the amino acids between the two conserved catalytic glutamates, E413 and E504 (GUSΔ413-504) (**Figure 5.17a**, **b**). This *gus* gene truncation was created using Lambda-Red with CRISPR/Cas9 counterselection[154]. Truncation of *gus* in *E. coli* did not affect cell viability in standard media; no differences were observed in growth curves between WT K-12 MG1655 and GUSΔ413-504 **(Figure 5.17c**). We also determined that the piperazine-containing inhibitors display no toxicity against WT *E. coli* K-12 MG1655 at up to 10 μM (**Figure 5.17d**).

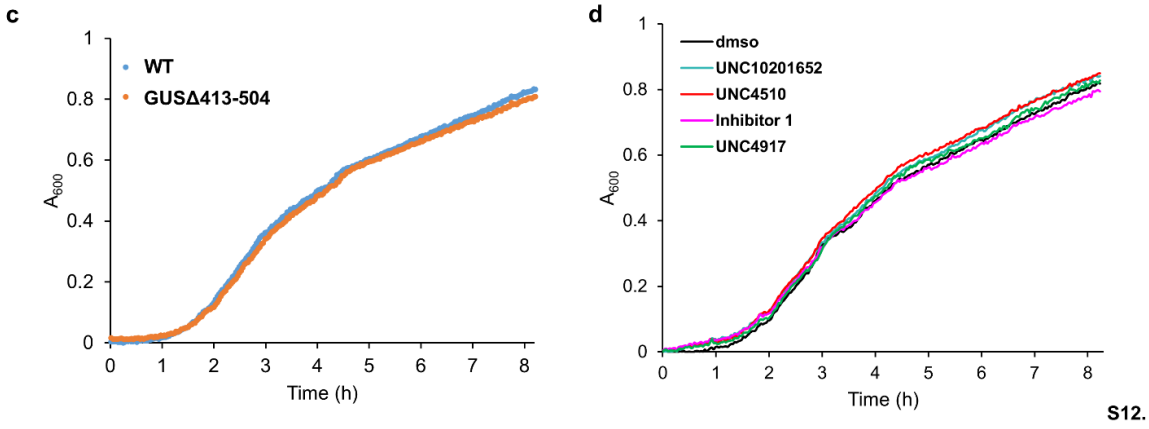
We then evaluated GUS activity in living *E. coli* cells by measuring PNPG cleavage[55,107]. WT *E. coli* K-12 MG1655 displays robust GUS activity, while GUSΔ413-504 *E. coli* lacks GUS activity, as expected (**Figure 5.18a**, **c**). Both WT and GUSΔ413-504 *E. coli* were then treated with the potent *in vitro* inhibitors **UNC10201652** and **UNC4917** (**Figure 5.1a**), the much weaker analogs **UNC4510** and **UNC10201651** (**Figure 5.11a**), and the previously characterized Inhibitor 1 that does not display slow-binding kinetics (**Figure 5.3d**). With WT *E. coli*, the EC_{50} values of the inhibitors directly mirrored their *in vitro* efficacies, with **UNC10201652** and **UNC4917** exhibiting potent inhibition at 74 ± 7 nM and 8 ± 4 nM, respectively, **UNC4510** showing weaker inhibition at 2,300 ± 500 nM, akin to Inhibitor 1 (3,400 ± 400 nM), and **UNC10201651** demonstrating no inhibition up to 10 μM (**Table 5.2**). In the GUSΔ413-504 *E. coli* strain, no GUS activity was observed (**Figure 5.18c**). Indeed, the GUSΔ413-504 *E. coli* strain gave the same level of signal as the WT *E. coli* strain when incubated with our potent GUS inhibitors (**Figure 5.18a**, **c**). Similar

a



b

Primer Name	Oligonucleotide sequence
sgRNA-EcGUSE413-E5041F	CCAAAGCCAGTAAAGTAGAAgttttagagctagaaatagcaag
gamR	tttataacctccttagagctcga
CPEC2F	cggcgtcacactttgctat
sgRNA-EcGUSE413-E5041R	TTCTACTTTACTGGCTTTGGgtgctcagtatctctatcactga
gusdelF	AAT ATG CCT TGC GAG GTC GCA
gusdelR	GGA TTC GAT AAC GTG CTG ATG GTG



S12.

Figure 5.17 - Generation of GUSΔ413-504 strain and analysis of growth. (**A**) Confirmed *E. coli* GUS Δ413-504 (colony 1). The 464 bp band is 276 bp smaller than the intact *gus* fragments at 770 bp (colonies 2-4). (**B**) Primers utilized for construction of pKDsgRNA-gus plasmid and amplification of the *gus* fragment in *E. coli* K-12 MG1655. (**C**) Growth curves of WT and GUS Δ413-504 *E. coli* strains. (**D**) Growth of WT *E. coli* in the presence of 10 µM of various compounds and 1% DMSO control.

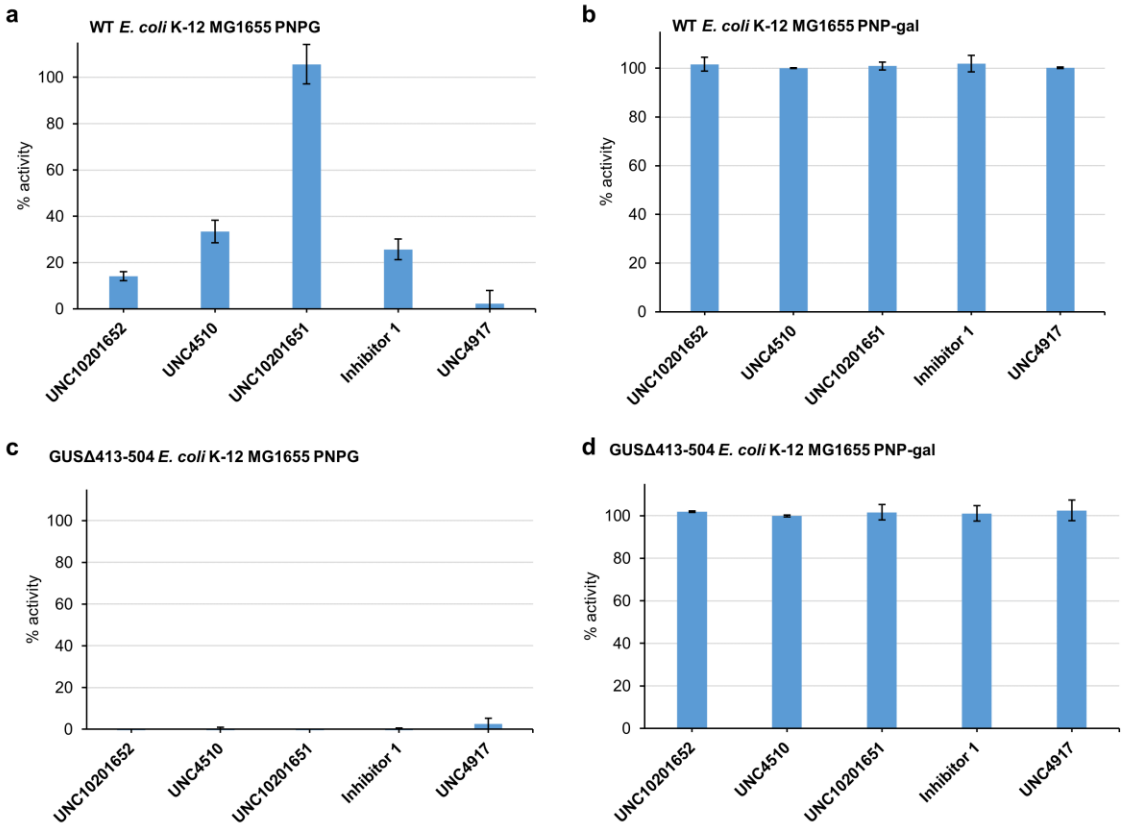


Figure 5.18 - Cell-based studies of potency and selectivity of piperazine-containing inhibitors. (**a**) Plot of % activity for WT *E. coli* K-12 MG1655 cells in the presence of PNPG and 10 μM of listed inhibitors. (**b**) Plot of % activity for WT *E. coli* K-12 MG1655 cells in the presence of PNP-gal and 10 μM of listed inhibitors. (**c**) Plot of % activity for GUSΔ413-504 *E. coli* K-12 MG1655 cells in the presence of PNPG and 10 μM of listed inhibitors. (**d**) Plot of % activity for GUSΔ413-504 *E. coli* K-12 MG1655 cells in the presence of PNP-gal and 10 μM of listed inhibitors. Error bars represent SD of n = 3 biological replicates.

results were observed for the approved drugs, with EC_{50} values ranging from 160 nM to 3.5 μM (**Table 5.2**). These results establish that potent GUS inhibition phenocopies a catalytically inactive *gus* gene in living *E. coli*, and that **UNC10201652**, **UNC4917**, as well as approved piperazine and piperidine-containing drugs, are efficacious in cells. Furthermore, the absence of inhibition by **UNC10201651** in WT *E. coli* pinpoints the secondary piperazine amine as the essential warhead for potent in-cell GUS inhibition.

Finally, to address in-cell selectivity, we examined the activity of *E. coli* □-galactosidase, a closely related glycosyl hydrolase that shares 15% sequence identity with *Ec*GUS, in both the WT and GUSΔ413-504 *E. coli* strains by using *p*-nitrophenyl-β-D-galactopyranoside (PNP-gal), a □-galactosidase substrate. We found that both WT and GUSΔ413-504 *E. coli* strains display robust cleavage of PNP-gal, and that neither strain is affected by GUS inhibitors (**Figure 5.18b**, **d**). Thus, the compounds tested are selective for GUS over the related glycosyl hydrolase β-galactosidase in living *E. coli* cells. To further address selectivity, we examined the *in vitro* inhibition of the mammalian bovine liver GUS that shares 42% sequence identity with *Ec*GUS. All inhibitors failed to yield any inhibition at up to 10 μM (**Figure 5.19**). This indicates that, like the microbial GUS-specific inhibitors previously reported, the piperazine-containing inhibitors described here are selective for bacterial GUS over the human GUS ortholog.

Discussion

We present a set of piperazine-containing compounds that act as inhibitors of microbiome GUS enzymes by intercepting the glycosyl-enzyme catalytic intermediate.

Table 5.2 - EC_{50} values in living WT E. coli MG1655 cells.

	UNC10201652	UNC4510	UNC10201651	Inhibitor 1	UNC4917	Amoxapine	Vortioxetine	Palbociclib	Crizotinib
PNPG	74 ± 7	2,300 ± 500	> 10,000	3,400 ± 400	8 ± 4	160 ± 30	3300 ± 300	2200 ± 400	3500 ± 300
PNP-gal	NI	NI	NI	NI	NI	NI	NI	NI	NI

EC50 values are in units of nanomolar and derived from n = 3 biological replicates ± SEM. NI: no inhibition.

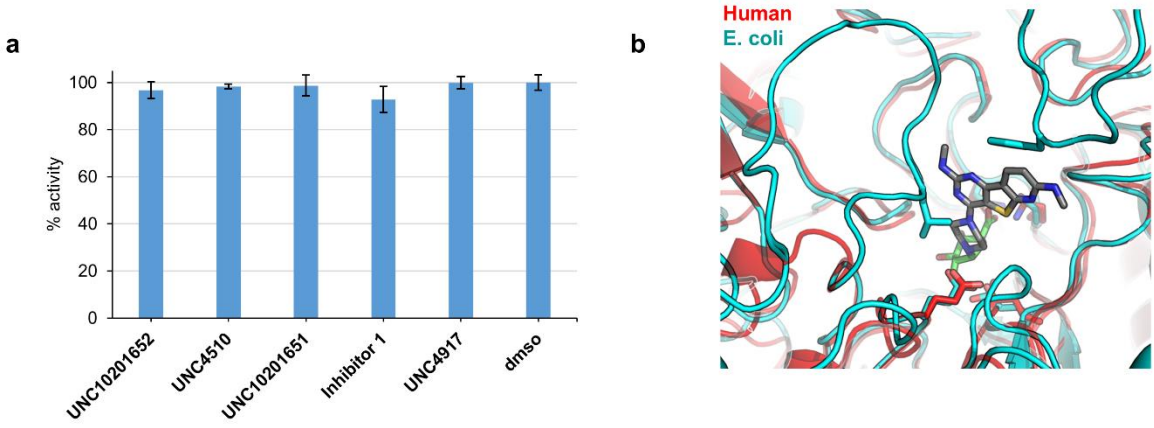


Figure 5.19 – Inhibition of mammalian GUS. (**a**) Percent activity of bovine liver GUS after treatment with 10 μM of inhibitor fails to yield inhibition. (**b**) Overlay of *E. coli* GUS (PDB: 3LPF) with Human GUS (PDB: 1BHG) and UNC4917-GlcA modeled demonstrates that the loop structure of E. coli GUS and other loop 1 GUS enzymes characterized here form the binding site for the piperazine-containing inhibitors.

Because they contact the Loop 1 region unique to bacterial GUS enzymes[136], these compounds are highly selective for these microbial proteins over the human GUS protein orthologue, as has been observed previously[55]. The GUS inhibitors characterized here emulate other studies where the onset of slow-binding only occurs in the presence of a cofactor, such as the binding of finasteride and dutasteride to NADPH-bound 5α-reductase, as well as the inhibition of NAD-bound enoyl reductase by various diazaborines[155,148]. An important distinction in the present study is that instead of covalently linking to the cofactor of an enzyme, UNC10201652 and UNC4917 target a catalytic intermediate, a unique observation among this type of slow-binding inhibitor in general and GUS inhibitors specifically.

The structural data presented here lend insight into how GUS enzymes may recognize their cognate substrates. Structures of GUS-bound GlcA conjugates, **UNC4917**-GlcA and amoxapine-GlcA, resemble GUS substrates such as testosterone, estrogen, and bile acid glucuronides[10]. The non-polar scaffold of **UNC4917** and amoxapine make contacts with an aromatic tyrosine residue conserved in microbial GUS enzymes identified to date, including those from the Human Microbiome Project stool sample database[136]. From the structures of the inhibitor-GlcA conjugates, it also appears that the inhibitor glucuronides occupy a strained, quasi axial β-linkage (**Figure 5.8b**, **5.14d)**. This may emulate the conformation of true glucuronide substrates prior to the hydrolysis of their glycosidic bonds. Due to the lower resolution of the structures elucidated here, analysis of the sugar ring conformations is purely speculative. However, it is likely that the GlcA ring is strained when covalently linked to the piperazine-containing inhibitors in the GUS active site.

Despite observing a well-recognized **UNC4917**-GlcA conjugate in the *Ee*GUS active site, exogenously-synthesized **UNC5670** proved to be a weak inhibitor of all bacterial GUS enzymes tested (**Figure 5.10b**). At least in part, this is due to the anomeric impurity of **UNC5670** (**Figure 5.10a**). That is, since the glucuronide of **UNC10201652** observed in the *Ee*GUS active site appeared exclusively β-linked in the crystal structure, as expected due to the specificity of GUS for this anomeric configuration, the approximately 50% of **UNC5670** in the α configuration is likely a poor inhibitor that may be unable to bind to GUS. Another potential contribution to reduced potency is the enthalpic cost of de-solvating the GlcA of **UNC5670**. Interestingly, **UNC5670** still displays slow-binding progress curves. In the same manner as **UNC4917** and **UNC10201652**, **UNC5670** only displays steady-state kinetics in the presence of substrate. This finding supports the hypothesis that a substrate-induced conformational change promotes the binding of these piperazine-containing compounds.

We propose that the initial state of inhibition is characterized by the interaction of inhibitor with GUS, while the steady-state kinetics are described by the interaction of inhibitor with a GUS-GlcA catalytic intermediate (**Figure 5.8d**). Since both non-nucleophilic analogs (**UNC4510** and **UNC5671**) and a piperazine-glucuronide conjugate (**UNC5670**) displayed non-linear progress curves, it appears that glucuronide formation is not the rate-limiting step for steady-state kinetics. Thus, we propose that substrate-induced isomerization of the GUS active site limits the onset of steady-state inhibition. That is, the active conformation is required for substrate entry and catalytic initiation in the GUS active site, which is also the conformation that the piperazine-containing inhibitors prefer to bind to, whether or not they are capable of forming glucuronide conjugates in the active site (**Figure**

5.11d). This ternary complex of GlcA and inhibitor bound in the active form is long-lived, and results in the slower steady-state observed.

Finally, we show that five different human-targeted drugs, including compounds for depression, infection, and cancer, inhibit gut microbial GUS enzymes via the same mechanism described for **UNC10201652** and **UNC4917**. In vitro and cell-based studies show that these approved drugs yield potent inhibition of GUS (**Figure 5.14b** and **Table 5.2**). A recent study on the effect of drugs on the gut microbiota revealed that the small intestine and colon concentrations of many drugs are on average in the mid- to high-micromolar range[156]. Indeed, using this model, Maier *et al*. calculated a small intestinal concentration of 106 µM and colon concentration of 138 µM for amoxapine, both of which are well above the EC_{50} calculated for amoxapine in *E. coli* cultures (**Table 5.2**)[156]. Using the same method for the other piperazine- and piperidine-containing drugs, we found that the respective small intestinal and colonic concentrations are 18 µM and 23 µM for vortioxetine, 216 µM and 378 µM for ciprofloxacin, 74 µM and 276 µM for palbociclib, and 148 µM and 466 µM for crizotinib. These predicted small intestine and colon concentrations are all greater than their EC_{50} in WT *E. coli* (**Table 5.2**), suggesting that the activity of loop 1 GUS enzymes may be completely blocked in patients taking these drugs.

Conclusions

The results herein show that compounds with terminal piperazines are substrate-dependent inhibitors of bacterial GUS. Furthermore, slow-binding inhibition only occurs when GUS is actively hydrolyzing substrate. Crystallographic analysis reveals that the substrate-dependence of slow-binding inhibition is likely due to an enhanced interaction with a catalytic intermediate where GlcA is covalently linked to GUS. Chemical analogs with

methylated piperazines demonstrate its importance for potent GUS inhibition, and further supports that these inhibitors target a GlcA-bound GUS. Lastly, approved drugs with terminal piperazines also inhibit bacterial GUS in a slow-binding manner. This final result highlights the potential for human therapeutics to exert off-target effects on the gut microbiota that may impact human health.

Methods

In vitro inhibition of bacterial GUS was assessed by combining 5 μL of 150 nM GUS (15 nM final), 5 μL of various concentrations of inhibitor, 30 μL of 1.5 mM PNPG (900 μM final), and 10 μL of assay buffer (25 mM NaCl, 25 mM HEPES, pH 7.5) in a 96-well Costar clear bottom plate. Reactions were initiated by addition of PNPG and then incubated for approximately 1 hour, after which the end point absorbance was determined at 410 nm in a BMG lab tech PHERAstar plate reader. The IC_{50} was determined as the inhibitor concentration that yielded a 50% reduction in the max absorbance of the uninhibited reaction, where percent inhibition was calculated as:

$$\%\ inhibition = \left[1 - \left(\frac{A_{exp} - A_{bg}}{A_{max} - A_{bg}}\right)\right] \times\ 100$$

where A_{exp} is the end point absorbance at a particular inhibitor concentration, A_{max} is the absorbance of the uninhibited reaction, and A_{bg} is the background absorbance. Percent inhibition values were subsequently plotted against the log of inhibitor concentration and fit with a four-parameter logistic function in SigmaPlot 13.0 to determine the IC_{50} as described above.

Slow-binding continuous kinetic assay

The same procedure as outlined for the IC_{50} assay was followed for reaction volumes and concentrations. Product formation was monitored continuously at 410 nm in a BMG lab tech PHERAstar plate reader. Resulting progress curves were truncated such that only data where the uninhibited reaction was linear were utilized to eliminate any potential of non-linear artifacts from substrate depletion. The resultant progress curves were fit by non-linear regression analysis in MATLAB with the following equation[146]:

$$[P] = v_s t + \frac{v_i - v_s}{k_{obs}}[1 - exp(-k_{obs}t)] + A_o \quad (1)$$

where v_i is the initial velocity, v_s is the steady-state velocity, k_{obs} is the first order rate constant for the transition from v_i to v_s, t is time, and A_o is the initial absorbance. In instances where v_s was zero, the following form of equation 1 was utilized:

$$[P] = \frac{v_i}{k_{obs}}[1 - exp(-k_{obs}t)] + A_o \quad (2)$$

The general kinetic scheme used to describe two-step slow-binding is shown below:

$$E + I \underset{k_2}{\overset{k_1}{\rightleftharpoons}} EI \underset{k_4}{\overset{k_3}{\rightleftharpoons}} EI^*$$

where E is enzyme and I is inhibitor. Since the resultant k_{obs} versus [I] plots were linear, we assumed that the initial isomerization was kinetically insignificant (*i.e.* $[I] << K_I$) and utilized a one-step kinetic scheme to fit the linear data of k_{obs} versus inhibitor concentration:

$$k_{obs} = \frac{k_3}{K_I} \cdot [I] + k_4 \ (3)$$

where [I] is the concentration of inhibitor and K_I is the equilibrium that describes the intial binding complex[157].

Substrate-dependent jump dilution assays

The jump dilution assays to determine the substrate-dependence of slow-binding inhibition were performed by mixing 5 µL of 15 µM GUS (1.5 µM final), 5 µL of various inhibitor concentrations, 30 µL of 1.5 mM PNPG (900 µM final) or 30 µL of water, and 10 µL of assay buffer (25 mM HEPES, 25 mM NaCl, pH 7.5). This initial reaction was incubated at 37 °C for 1 hour. After pre-incubation, 1 µL of the reaction was diluted into 99 µL of PNPG-containing buffer (900 µM PNPG, 25 mM HEPES, 25 mM NaCl, pH 7.5), and the resulting activity was monitored continuously at 410 nm. Progress curves were plotted in Microsoft Excel.

Bovine liver GUS selectivity assay

Bovine liver GUS was obtained from Sigma-Aldrich as a lyophilized powder and dissolved in a 10 mM sodium acetate and 10 mM sodium chloride pH 5.0 buffer and stored at 4 °C. Final assay contained 5 µL of bovine liver GUS (0.132 mg/mL), 10 µL of pH 5 buffer (25 mM sodium chloride, 25 sodium acetate), 5 µL inhibitor (10 µM final), and 30 µL of PNPG. Assays were initiated by addition of PNPG and incubated for 1 h at 37 °C. Reactions were quenched by addition of 0.2 M sodium carbonate and absorbance at 410 nm was measured in a BMG lab tech PHERAstar plate reader. Percent inhibition was calculated as described for the in vitro IC50 assay.

Crystallography

Crystals of *Ee*GUS were produced via the hanging-drop vapor diffusion method. Apo-*Ee*GUS crystals were formed by incubation of 13 mg/mL *Ee*GUS in 35% PEG 400 and

0.1 M Bis-tris pH 6.5. The same conditions were used for the *Ee*GUS-GlcA complex crystals, except 10-fold molar excess of GlcA was mixed with *Ee*GUS before addition to the crystallant. For crystals that contained both inhibitor and PNPG as ligands, *Ee*GUS was incubated with inhibitor (10-fold molar excess) and PNPG (30-fold molar excess) for 30 minutes prior to addition to the crystallant solution. Since the crystallant served as a cryoprotectant, no additional cryoprotectant was utilized prior to flash-freezing in liquid nitrogen. Diffraction data for all crystals were collected on the 23-ID-B beamline at GM/CA-CAT (Advanced Photon Source, Argonne National Laboratory). Phasing for the apo structure of *Ee*GUS was performed in Phenix via molecular replacement with *Cp*GUS (PDB: 4JKM)[158]. The apo *Ee*GUS structure was subsequently utilized to perform molecular replacement for both the GlcA-bound and UNC4917-GlcA *Ee*GUS structures. Refinements and ligand generation were carried out in Phenix and ligand fitting was performed in Coot[114]. Final coordinates and structure factors have been submitted to the RCSB and assigned accession codes of 6BJW, 6BJQ, 6BO6, and 6D4O for the apo, GlcA-bound, **UNC4917**-GlcA, and amoxapine-GlcA *Ee*GUS structures, respectively. Statistics for all structures are listed in **Table 5.3**.

Liquid Chromatography Mass Spectrometry

For LC-MS analysis, 50 µL reactions were performed with 5 µL of 100 µM GUS (10 µM final), 5 µL of 10 mM inhibitor (1 mM final), 5 µL of 5 mM PNPG (500 µM final), and 35 µL of buffer (10 mM NaCl, 10 mM HEPES, pH 7.5). Reactions were quenched by heating the sample at 100 ˚C for 5 minutes and subsequent addition of 50 µL of acetonitrile. Samples were then centrifuged at 13000 rpm for 5 minutes and supernatant was utilized for LC-MS analysis. Separation was carried out on a 50 mm Phenomenex Gemini C18 column

Table 5.3 - Crystallographic table for structures of *Ee*GUS, *Ee*GUS + GlcA, *Ee*GUS + UNC4917 + PNPG, and *Ee*GUS + Amoxapine + PNPG.

Protein	*Ee*GUS	*Ee*GUS + GlcA	*EeGUS* + **UNC4917** + PNPG	EeGUS + amoxapine + PNPG
PDB code	6BJW	6BJQ	6BO6	6D4O
Space group	$P6_422$	$P6_422$	$P6_422$	$P6_422$
Unit cell: a, b, c (Å); α, β, γ (°)	179.87, 179.87, 134.88; 90.00, 90.00, 120.00	179.79, 179.79, 134.49; 90.00, 90.00, 120.00	179.44, 179.44, 133.78; 90.00, 90.00, 120.00	179.95, 179.95, 134.94; 90.00, 90.00, 120.00
Resolution range (Å)	29.43-3.00	29.96-2.70	29.37-2.80	29.45-2.90
Wavelength (Å)	1.03	1.03	1.03	1.03
Unique reflections	56356	35651	31671	29069
Multiplicity	12.3	13.6	17.5	19.3
Completeness (%)	99.9	99.9	98.2	99.9
Average I/σ	20.0	15.5	29.2	17.0
Wilson B-factor	65.2	54.3	65.1	46.5
R_{work}	0.2266	0.2179	0.258	0.199
R_{free}	0.2570	0.2413	0.292	0.244
Molecules in AU	1	1	1	1
Waters in AU	0	66	19	54
Residues in AU	579	597	493	612
Average B-factor	70.0	72.0	70.0	53.64
RMS (bonds)	0.003	0.004	0.005	0.007
RMS (angles)	0.655	0.746	0.74	1.293
Ramachandran favored (%)	89.12	93.24	93.37	93.83
Ramachandran outliers (%)	0.88	0.17	0.21	0.00

with 5 μm particle size and 110 Å pore size. Solvent A was 0.1% formic acid in water, and solvent B was 0.1% formic acid in acetonitrile. Compounds were eluted by 2% B for 2 min followed by a linear gradient to 95% B over 10 min and held at 95% B for 2 min. Supernatant was analyzed using an Agilent Technologies 6520 Accurate-Mass Q-TOF LC-MS in positive-ion mode. Results were analyzed in MassHunter Qualitative Analysis B.06.00.

Generation of *E. coli* K-12 MG1655 GUSΔ413-504 Strain

The E413-E504 region of the *gus* gene in *E. coli* strain MG1655 was deleted using Lambda-Red with CRISPR/Cas9 counter-selection to create an *E. coli* K-12 MG1655 GUSΔ413-504 strain[154]. This region was chosen because it contains both catalytic glutamates, E413 and E504; thus, this deletion would be expected to inactivate the enzyme.

The pCas9-CR4 and pKDsgRNA-p15 plasmids used to construct the *E. coli* K-12 MG1655 GUSΔ413-504 strain were purchased from Addgene. Circular polymerase extension cloning (CPEC) was used to replace the 20 bp targeting sequence of the sgRNA in pKDsgRNA-15 with a 20 bp sequence that targets the *gus* gene. The primers listed in **Figure 5.17b** were used to generate two PCR fragments containing overlapping protospacer sequences. The primer pair sgRNA-gus-F and gamR yielded a ~3 kb product, and the primer pair sgRNA-gus-R and CPEC2F yielded a ~4 kb product. PCR products were gel purified and cloned by CPEC with the Q5 HF polymerase to create the pKDsgRNA-gus plasmid. The PCR mixture was transformed into chemically competent DH5α cells, plated on 50 mg/L spectinomycin, and incubated at 30 ºC.

Upon transformation of the pCas9-CR4 and pKDsgRNA-gus plasmids into electrocompetent *E. coli* K-12 MG1655 cells, cells were grown to an OD of ~0.4, and λ-red

was induced with the addition of L-arabinose at a final concentration of 0.2% and incubated at 30ºC for 20 minutes. The oligo designed to incorporate the desired deletion (5'T*G*TACATTGAGTGCAGCCCGGCTAACGTATCCACGCCGTAGTTGGCAATACT CCACATCACCACGCTTGGGTGGTTTT*T*G 3', where * is a phosphorothioate bond) was added to 50 µL of electrocompetent cells at a final concentration of 10 µM. Electroporation was performed and cells were recovered in SOC for 1 hour before plating on 34 mg/L chloramphenicol, 50 mg/L spectinomycin, and 100 µg/L anhydrotetracycline plates at 30ºC overnight.

To confirm successful deletion of the *gus* gene, genomic DNA was isolated from the *E. coli* MG1655 K-12 strain using the PureLink Genomic Isolation Kit (Invitrogen). The region surrounding the *gus* deletion was amplified using the primers listed in **Figure 5.17b** and the Q5 HF polymerase. PCR products were then sequenced to confirm the 276 bp deletion in the *gus* gene. Upon verifying the *gus* deletion, the pCas9-CR4 and pKDsgRNA-GUS plasmids were cured according to the protocol by Reisch et al[154].

Cell-Based Assays

WT and GUSΔ413-504 *E. coli* K-12 MG1655 were grown overnight in 10 mL LB and 100 µL were sub-cultured the following morning in 5 mL of fresh LB. Cells were grown to an OD of approximately 0.6 and used for the cell-based assay. Reactions were carried out in costar 96-well black clear bottom plates. Reaction volumes consisted of 90 µL of cells pre-mixed with 700 µM PNPG and various concentrations of 10 µL inhibitor. This reaction was incubated for 24 h at 37 °C with a low evaporation lid. Incubations were quenched by addition of 50 µL of 0.2 M sodium carbonate. Absorbance values were measured at 410 nm

in a BMG lab tech PHERAstar plate reader. Percent inhibition and EC_{50} values were determined as described previously for the *in vitro* IC_{50} assay.

Growth curve assay for WT and GUS Δ413-504 E. coli K-12 MG1655 cells

Glycerol stocks of WT and GUSΔ413-504 *E. coli* K-12 MG1655 were used to inoculate 10 mL of LB broth and shaken overnight at 37 °C and 225 rpm. From these overnights, 100 µL were sub-cultured into 5 mL of fresh LB, of which 100 µL was added to a 96-well black clear flat bottom plate. The plate was covered with a low evaporation lid and incubated at 37 °C in a PHERAstar plate reader with shaking at two-minute intervals for 30 seconds at 700 rpm. The absorbance at 600 nm was also measured with orbital averaging at two-minute intervals over the course of approximately 8 h.

CHAPTER 6: STRUCTURE ACTIVITY RELATIONSHIP OF A PIPERAZINE-CONTAINING INHIBITOR OF GUT BACTERIAL β-GLUCURONIDASES.

The human gut microbiome plays key roles in drug metabolism, including the reversal of detoxification reactions performed by the host liver. Many drugs are glucuronidated in the liver[11], a phase II reaction mediated by uridine diphosphate glucuronosyl transferases (UGTs) that attach glucuronic acid (GlcA) to various functional groups present on xenobiotics. When the resulting drug-GlcA conjugates, or glucuronides, are secreted into the gastrointestinal tract, they encounter the gut microbiota and an arsenal of bacterial glycoside hydrolases (GHs). Gut bacterial β-glucuronidases (GUSs) hydrolyze glucuronides generated by the liver, reversing drug inactivation and detoxification[159]. The ramifications of this bacterial GUS activity include dose-limiting diarrhea of the chemotherapeutic irinotecan and small-intestinal ulceration caused by NSAIDs[105,123]. Thus, development of small-molecule bacterial GUS inhibitors is of considerable interest as adjuvants for toxic therapeutics.

Selective small-molecule inhibition of gut bacterial GUSs has been shown to alleviate the toxic side effects of anticancer and anti-inflammatory drugs[55,56,124]. Over the past nine years, numerous bacterial GUS inhibitors have been discovered to alleviate the GI toxic side effects of irinotecan. The first selective bacterial GUS inhibitors were identified via an HTS with *E. coli* GUS[160]. These initial inhibitors were selective for bacterial GUS and non-toxic, setting the stage for non-antibiotic approaches to drug the gut microbiome. Other inhibitors have been developed (ref). While these inhibitors are potent against *E. coli* GUS, their mechanism of action and binding modes remain unclear[161]. More recently, the mechanism of

piperazine-containing GUS inhibitors was elucidated via kinetic and structural analysis, demonstrating a unique slow-binding and substrate-dependent mechanism[131]. Approved drugs that contain piperazine moieties are also gut bacterial GUS inhibitors which utilize the same mechanism. These studies have begun to refine our understanding of gut bacterial GUS inhibition, and will hopefully accelerate the development of adjuvant therapies.

Here we detail a structure-activity relationship (SAR) of the piperazine-containing GUS inhibitor **UNC10201652** (**Figure 6.1A**). We synthesized 28 analogs of **UNC10201652** and measured their efficacies against four distinct bacterial GUSs present in the human gut microbiome. We confirmed that the piperazine ring is essential to potent GUS inhibition, and that any modification to this moiety reduces potency against all bacterial GUSs examined. Modification to other moieties of **UNC10201652** generally yielded moderate improvements in potency as well as enabled selectivity among bacterial GUS enzymes. Analysis of existing crystal structures combined with molecular docking suggest that differences in GUS loop structures are key determinants of potency and selectivity. Furthermore, kinetic and pH studies further support the unique substrate-dependent slow-binding mechanism previously proposed for piperazine-containing GUS inhibitors. Lastly, analysis of select analogs reveal selective, potent, and non-toxic inhibition in *E. coli* cells as well reduction of GUS activity in human fecal samples. Taken together, the SAR reported here reveals the molecular features that enable potent GUS inhibition and revealed several promising therapeutic candidates.

RESULTS

Piperazine is Essential for Potent GUS Inhibition

We determined a structure-activity relationship for the previously characterized GUS inhibitor **UNC10201652** by modifying four moieties: the piperazine, morpholine, triazine, and cyclohexyl (**Figure 6.1A**). A key feature of **UNC10201652** is the piperazine that

interacts with the catalytic glutamates of GUS in a substrate-dependent manner (**ref**) (**Figure 6.1A**, **B**). We further explored the effect of the piperazine on bacterial GUS inhibition by synthesizing ten analogs with a modified piperazine (**Table 6.1**). To analyze differences in potency, IC_{50} values for all synthesized analogs were determined with a *p*-nitrophenyl-β-D-glucuronide (PNPG) assay and four distinct gut bacterial GUSs: *Escherichia coli* GUS (*Ec*GUS), *Streptococcus agalactiae* GUS (*Sa*GUS), *Clostridium perfringens* GUS (*Cp*GUS), and *Eubacterium eligens* GUS (*Ee*GUS) (**Table 6.1**). Changes to the secondary amine of the piperazine moiety, including the addition of a methyl group to the secondary amine (**UNC4510**) and the replacement of the secondary amine with an oxygen (**UNC4365**) or carbon (**UNC10201651**), yielded at least 100-fold reduction in potency (**UNC4510**) or complete loss of inhibition (**UNC4365** and **UNC10201651**) in comparison to **UNC10201652** (**Table 6.1**). The complete loss of inhibition observed for **UNC4365** and **UNC10201651** demonstrates that the previously observed ionic interaction between the piperazine and catalytic acid/base is essential for potent GUS inhibition by this chemotype (**Figure 6.1A**)[131]. Manual docking (see **Experimental Section**) of **UNC4510** suggests that reduced affinity is likely a result of the limited space available near the piperazine upon binding to the GlcA-linked enzyme, as the added methyl group would either clash with GlcA or prevent ideal

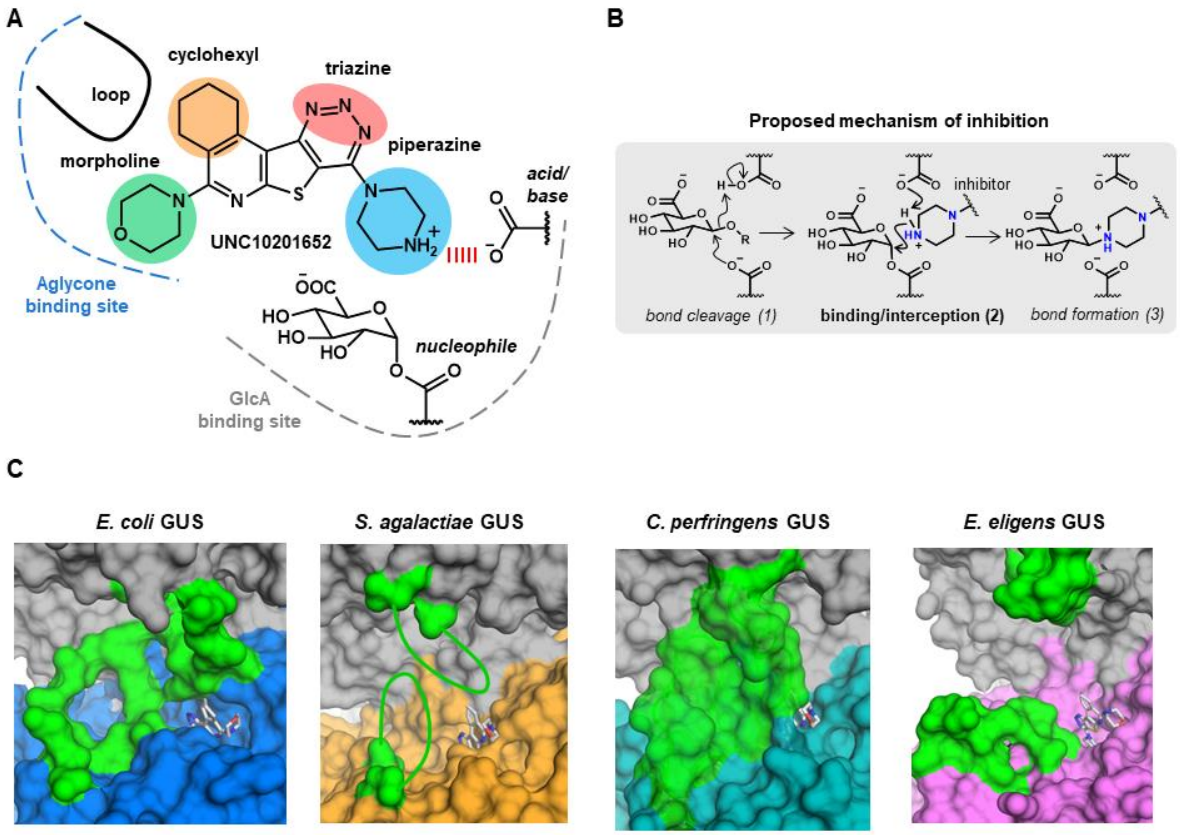
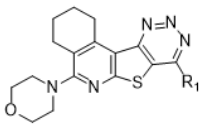
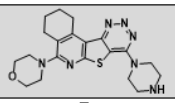
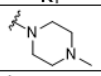
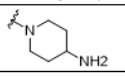
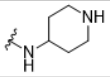
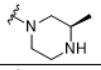
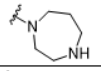
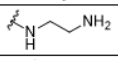


Figure 6.1 – Chemical structure, mechanism, and binding mode of UNC10201652. (**A**) Chemical structure of UNC10201652 with moieties investigated in this study highlighted. (**B**) Proposed mechanism of piperazine-containing GUS inhibitors. (**C**) Structures of the four gut bacterial GUS enzymes analyzed in this study with key loop structures show in green and adjuvant monomers in grey with inhibitor docked.

Table 6.1 - Effect of varying the piperazine amine on in vitro potency against *E. coli*, *S. agalactiae*, *C. perfringens*, and *E. eligens* GUS.



1	**UNC10201652**		**0.108 ± 0.008**	**0.127 ± 0.009**	**0.06 ± 0.01**	**0.6 ± 0.2**
ID	**UNC ID**	**R_1**	***Ec*GUS IC_{50} (μM)**	***Sa*GUS IC_{50} (μM)**	***Cp*GUS IC_{50} (μM)**	***Ee*GUS IC_{50} (μM)**
2	UNC4510		10 ± 2	14 ± 1	5 ± 1	30 ± 2
3	UNC4541		2.2 ± 0.4	6.9 ± 0.7	2 ± 1	8 ± 3
4	UNC4351		10.2 ± 0.8	40 ± 10	6 ± 1	8 ± 1
5	UNC4513		3.3 ± 0.5	2.8 ± 0.3	9.0 ± 0.2	4 ± 1
6	UNC4601		0.8 ± 0.1	0.82 ± 0.04	0.9 ± 0.1	2.4 ± 0.4
7	UNC4684		600 ± 100	640 ± 60	800 ± 200	1.9 ± 0.4
8	UNC4511		1.2 ± 0.8	3.8 ± 0.3	2.0 ± 0.1	11.2 ± 0.8
9	UNC4540		2.6 ± 0.4	21 ± 1	3.2 ± 0.6	11.0 ± 0.3
10	UNC4365		>100	>100	>100	>100
11	UNC10201651		>100	>100	>100	>100

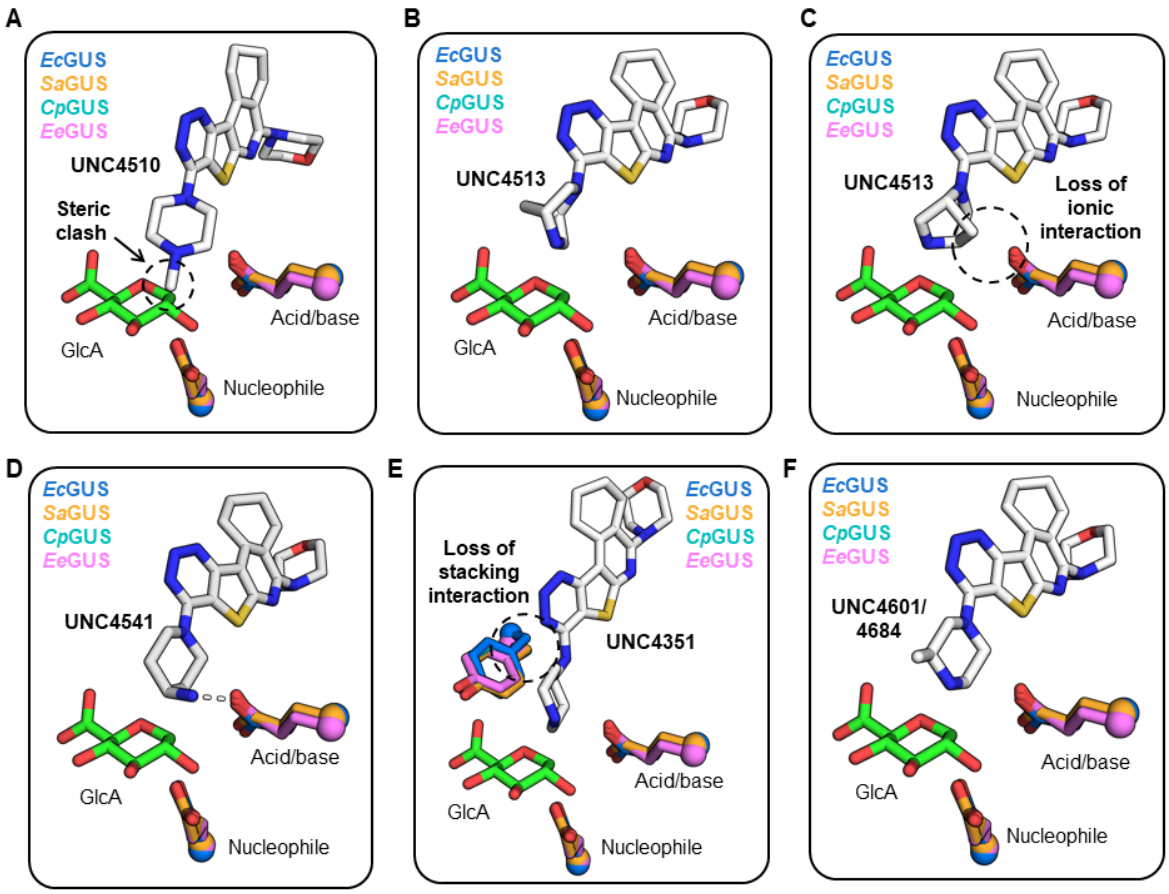


Figure 6.2 – Docking analysis of UNC10201652 piperazine analogs.

interaction with the catalytic acid/base (**Figure 6.2 A**). Taken together, these results confirm the essential nature of the piperazine for potent GUS inhibition and further support the substrate-dependent mechanism of inhibition for the piperazine chemotype.

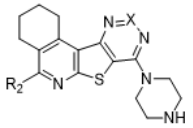
In addition to direct changes to the piperazine secondary amine, changes to any part of the piperazine moiety reduced potency against the GUSs examined. The introduction of a bridging methylene (**UNC4513**) reduces potency by at least ten-fold (*Ec*GUS) and up to 30-fold (*Sa*GUS) in comparison to **UNC10201652** (**Table 6.1**). The bridged ring system of **UNC4513** results in only one favorable binding orientation in comparison to two for the unmodified piperazine and may explain its reduced potency (**Figure6.2B, C**). The replacement of the piperazine with a seven-membered ring (**UNC4511**) displays at least 10-fold weaker potency in comparison to the parent compound with all enzymes (**Table 1**). Moving the nitrogen to an exocyclic position (**UNC4541**) also reduces potency against all GUS enzymes analyzed, which may be a result of poor interaction with the catalytic acid/base glutamate or steric clash with GlcA (**Figure 6.2D**). Replacing the piperazine with an ethyl amine (**UNC4540**) reduces potency by at least 20-fold across all GUS enzymes tested, demonstrating that even an amine in the same position as the corresponding piperazine is unable to potently inhibit GUS (**Table 6.1**). **UNC4351** replaces the piperazine with a piperidine and is linked to the core scaffold via an exocyclic nitrogen, a change that resulted in at least a ten-fold loss in potency against all GUS enzymes tested (**Table 6.1**). Docking analysis of **UNC4351** reveals the loss of a key stacking interaction with a conserved tyrosine to accommodate the exocyclic amine (**Figure 6.2E**). The addition of methyl groups to the carbon adjacent to the secondary amine also reduces potency (**UNC4601** and **UNC4684**), but much less so than the analogs discussed so far (**Table 6.1**). Docking analysis

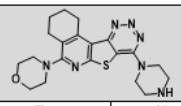
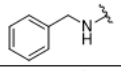
of **UNC4601** does not reveal any obvious steric clashes (**Figure 6.2F**). Taken together, these analogs show that the piperazine of **UNC10201652** is essential for potent inhibition of gut microbial GUS enzymes.

Morpholine modifications yield equivalent or improved potency

We next investigated the morpholine ring of **UNC10201652** and its role in GUS inhibition. **UNC10206579** replaces the morpholine with a second piperazine and is either equipotent (*Ec*GUS and *Sa*GUS) or 2-fold better (*Cp*GUS and *Ee*GUS) than **UNC10201652** (**Table 6.2**). The improvements for *Cp*GUS and *Ee*GUS may be a result of improved interaction with the second piperazine, where the cyclohexyl group can interact with the numerous non-polar residues in *Cp*GUS (F363 and M364) and *Ee*GUS (M150) in the active site (**Figure 6.3A**, **B**). Replacement of the morpholine with a monomethyl amine (**UNC10206581**) improves the potency against *Ec*GUS (0.029 ± 0.006 μM), *Sa*GUS (0.088 ± 0.008 μM), and *Cp*GUS (0.034 ± 0.002 μM) in comparison to **UNC10201652** (**Table 6.2**). The improved potency of **UNC10206581** may be due to fewer steric constraints encountered from loop structures that drape over the active site, allowing for a better fit in the active sites of *Ec*GUS, *Sa*GUS, and *Cp*GUS (**Figure 6.1C, 6.3C**). Furthermore, *Ec*GUS, *Sa*GUS, and *Ee*GUS have larger hydrophobic groups near the monomethyl amine that may cause steric issues with the morpholine that are not present in **UNC10206581** (**Figure 6.3C**). In contrast, replacement of the morpholine with a longer, aromatic benzyl amine moiety (**UNC10206607**) does not affect the potency with the GUS enzymes analyzed (**Table 6.2**), suggesting that the active site loops are mobile and can accommodate extended moieties in the aglycone binding region of the GUS active site. Replacement of the triazine with a diazine (**UNC4628**) results in equipotent inhibition for *Ec*GUS, *Sa*GUS, and *Cp*GUS (**Table**

Table 6.2 - Effect of varying the morpholine and triazine moieties on in vitro potency against *E. coli*, *S. agalactiae*, *C. perfringens*, and *E. eligens* GUS.



1	**UNC10201652**			**0.108 ± 0.008**	**0.127 ± 0.009**	**0.06 ± 0.01**	**0.6 ± 0.2**
ID	**UNC ID**	**R_2**	**X**	***Ec*GUS IC_{50} (μM)**	***Sa*GUS IC_{50} (μM)**	***Cp*GUS IC_{50} (μM)**	***Ee*GUS IC_{50} (μM)**
12	UNC10206579		N	0.10 ± 0.02	0.123 ± 0.004	0.032 ± 0.002	0.30 ± 0.03
13	UNC10206577		N	0.07 ± 0.01	0.118 ± 0.004	0.049 ± 0.002	0.70 ± 0.09
14	UNC10206581		N	0.029 ± 0.006	0.088 ± 0.008	0.034 ± 0.002	0.77 ± 0.09
15	UNC10206607		N	0.10 ± 0.02	0.12 ± 0.01	0.06 ± 0.02	560 ± 80
16	UNC4628		C	0.09 ± 0.05	0.13 ± 0.03	0.05 ± 0.02	0.24 ± 0.06

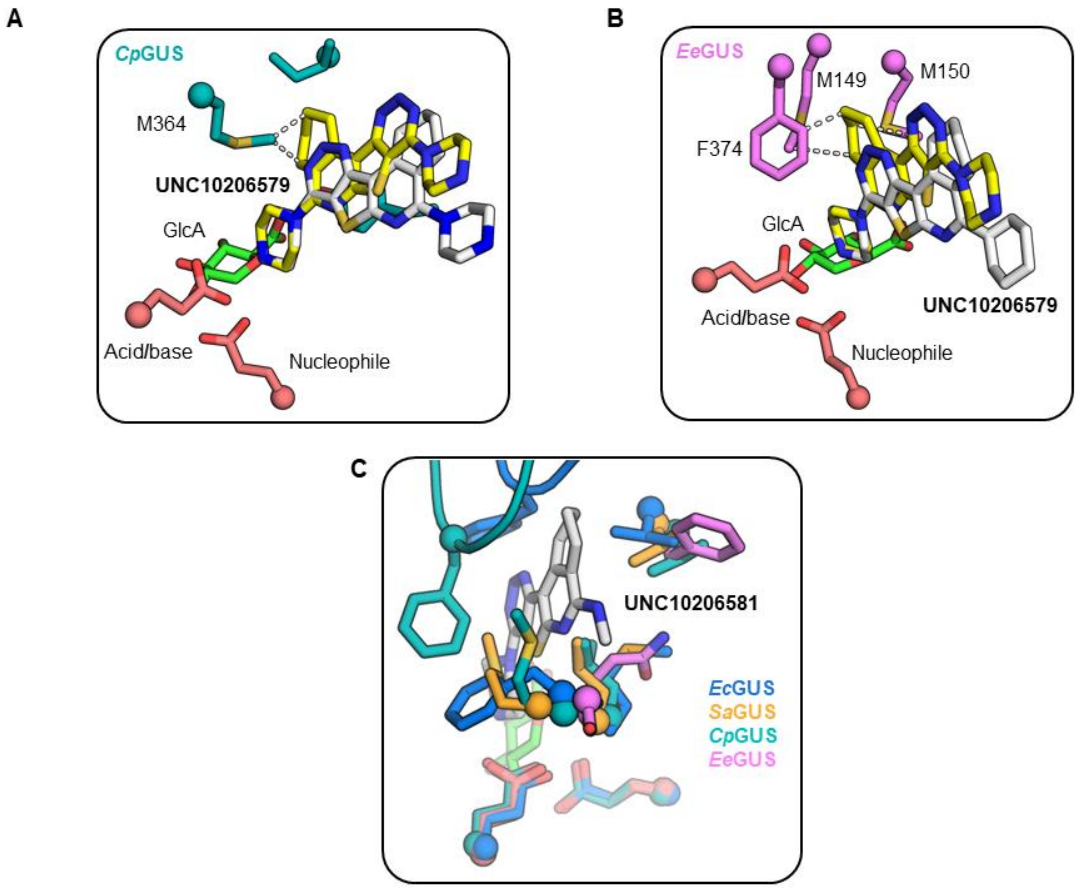


Figure 6.3 – Docking studies of morpholine analogs.

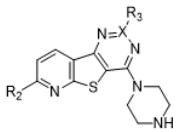
6.2). Interestingly, **UNC4628** is approximately 2-fold more potent against *Ee*GUS, which may be due to more favorable interactions with M150, a unique residue in the *Ee*GUS active site in comparison to the three other GUSs examined. Taken together, modifications to the morpholine either improve potency or are equipotent to **UNC10201652**.

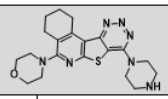
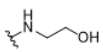
Cyclohexyl removal enables selectivity towards *Ec*GUS and *Cp*GUS

Removal of the cyclohexyl group (**UNC4830**), a moiety near the aglycone binding region of the GUS active site (**Figure 6.1A**), improved potency approximately 4-fold and 3-fold with *Ec*GUS and *Cp*GUS, respectively, but reduced potency 6-fold and 2-fold against *Sa*GUS and *Ee*GUS, respectively (**Table 6.3**). Both *Ec*GUS and *Cp*GUS have long loop structures that contain large aromatic residues at the active site, and thus the removal of the cyclohexyl group in **UNC4380** appears to enable a better fit in their occluded active sites (**Figure 6.1** and **6.4**). In comparison, *Sa*GUS and *Ee*GUS have shorter loop sequences that may accommodate the cyclohexyl moiety and aid in binding. Taken together, removal of the cyclohexyl group enables selectivity among similar GUSs, a potentially useful trait for future inhibitor development.

In addition to removal of the cyclohexyl moiety, we also explored attachments to the modified diazine to explore the space above the GlcA binding site (**Figure 6.1A**). Removal of the cyclohexyl and addition of a monomethyl amine (**UNC4785**) has relatively little impact on potency in comparison to **UNC4830** for *Ec*GUS and *Cp*GUS (**Table 6.3**). For *Sa*GUS and *Ee*GUS, addition of the monomethyl amine is an improvement over **UNC4830**, which suggests that the monomethyl amine may form favorable contacts (**Table 6.3**). Indeed, *Ee*GUS does have methionines near the monomethyl amine moiety, and may contribute to the improved binding of **UNC4785** with *Ee*GUS (**Figure 6.4B**). A combination of

Table 6.3 - Effect of varying the morpholine, cyclohexyl, and triazine moieties on in vitro potency against *E. coli*, *S. agalactiae*, *C. perfringens*, and *E. eligens* GUS.



1	UNC10201652				0.108 ± 0.008	0.127 ± 0.009	0.06 ± 0.01	0.6 ± 0.2
ID	**UNC ID**	**R_2**	**R_3**	**X**	***Ec*GUS IC_{50} (μM)**	***Sa*GUS IC_{50} (μM)**	***Cp*GUS IC_{50} (μM)**	***Ee*GUS IC_{50} (μM)**
17	UNC4830		--	N	0.03 ± 0.02	0.8 ± 0.1	0.02 ± 0.01	1.04 ± 0.08
18	UNC4746		H	C	0.02 ± 0.01	0.6 ± 0.1	0.014 ± 0.008	0.73 ± 0.07
19	UNC4785		NHMe	C	0.022 ± 0.007	0.4 ± 0.1	0.030 ± 0.008	0.40 ± 0.07
20	UNC4652		--	N	0.03 ± 0.01	0.72 ± 0.08	0.02 ± 0.01	1.38 ± 0.09
21	UNC4600		H	C	0.016 ± 0.009	0.27 ± 0.05	0.015 ± 0.008	0.52 ± 0.09
22	UNC4708		NHMe	C	0.05 ± 0.02	0.6 ± 0.1	0.09 ± 0.04	0.72 ± 0.08
23	UNC4707		NMe2	C	0.05 ± 0.01	0.40 ± 0.08	0.071 ± 0.002	0.26 ± 0.03
24	UNC4847			C	0.041 ± 0.009	0.5 ± 0.1	0.051 ± 0.001	1.3 ± 0.2
25	UNC4764	NMe2	--	N	0.023 ± 0.005	0.8 ± 0.2	0.014 ± 0.003	2.9 ± 0.6
26	UNC4911	NMe2	NHMe	C	0.016 ± 0.004	0.5 ± 0.1	0.026 ± 0.001	0.7 ± 0.1
27	UNC4917	NHMe	NHMe	C	0.020 ± 0.005	1.1 ± 0.2	0.0505 ± 0.0004	2.4 ± 0.5
28	UNC4666	H	H	C	0.019 ± 0.003	2.7 ± 0.6	0.044 ± 0.003	13 ± 3
29	UNC4910	H	NHMe	C	0.009 ± 0.002	0.500 ± 0.1	0.060 ± 0.001	2.8 ± 0.6

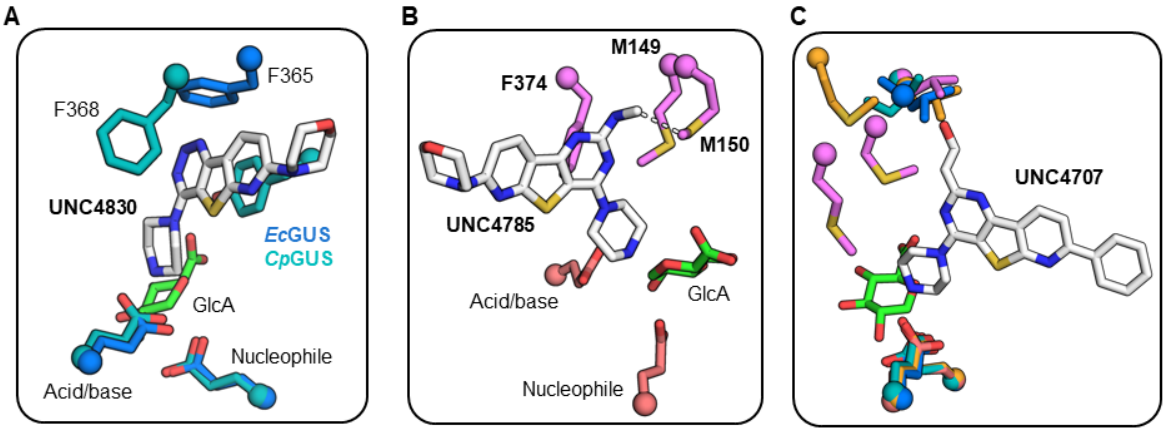


Figure 6.4 – Docking analysis analogs with cyclohexyl moiety removed.

cyclohexyl removal and addition of a phenyl ring in place of the morpholine (**UNC4652**) is similar in potency to removal of the cyclohexyl alone (**UNC4830**) for all GUS enzymes examined (**Table 6.3**). This result corroborates other morpholine analogs that show that changes to this moiety do not greatly impact potency (**Table 6.2**).

Combined modifications of cyclohexyl removal, replacement of the morpholine ring with a phenyl ring at position R_2, and modifications of the triazine to a diazine (**UNC4600**) display improved potency against *Ec*GUS and *Cp*GUS over **UNC4830** (**Table 6.3**). Other changes to the **UNC4600** scaffold at position R_3 do not yield major differences in potency, suggesting that the addition of small, non-polar moieties on the diazine do not significantly enhance or diminish their ability to bind to the four GUS enzymes examined (**Table 6.3**). While the addition of an ethoxy substituent on the diazine (**UNC4847**) does not impact potency significantly of EcGUS SaGUS and CpGUS, it does show significantly worse inhibition against *Ee*GUS (**Table 6.3**). **UNC4707** was approximately two-fold more potent than the parent compound against *Ee*GUS (**Table 6.3**). As discussed previously for **UNC4785**, the unique dimethylamine of **UNC4707** can interact with the methionine-containing loop insert that hangs over the active site of *Ee*GUS (**Figure 6.4C**).

Lastly, we synthesized analogs with modifications to both the R_2 and R_3 position in addition to removal of the cyclohexyl moiety. These analogs are among the most potent examined against *Ec*GUS (**UNC4911**, **UNC4910**). These are the smallest compounds tested and may fit better in the more occluded active site of both *Ec*GUS and *Cp*GUS (**Figure 6.1C**). While these smaller analogs are potent against EcGUS and CpGUS, they are much weaker against *Sa*GUS and *Ee*GUS. UNC4666 display nanomolar potency against *Ec*GUS and *Cp*GUS, but micromolar potency against *Sa*GUS and *Ee*GUS, making it the most

selective of all analogs characterized (**Table 6.3**). Taken together, it appears that the smaller analogs are more potent with *Ec*GUS and *Cp*GUS but worse against *Sa*GUS and *Ee*GUS. Thus, these analogs may prove useful in the future to selectively target specific bacterial GUS enzymes.

Kinetic analysis and pH dependence of piperazine-containing GUS inhibitors

We selected five compounds from the SAR of **UNC10201652** for further analysis, including the parent compound **UNC10201652** and the following analogs: **UNC10206579**, **UNC10206581**, **UNC4707**, and **UNC4911** (**Table 6.1, 6.2,** and **6.3**). These compounds were selected for their potent inhibition against all GUS enzymes tested and their structural diversity. Previous studies with **UNC10201652** and **UNC4917** revealed that they inhibit GUS via substrate-dependent slow-binding inhibition[131]. Thus, we tested if other piperazine-containing inhibitors also display substrate-dependent slow-binding. Indeed, each inhibitor displayed non-linear progress curves indicative of slow-binding inhibition (**Figure 6.5** and **Figure 6.6**)[146]. Furthermore, quantitative analysis of slow-binding inhibition revealed slow-binding efficiencies that mirror the IC50s observed *in vitro* (**Figure 6.6B**). To determine if slow-binding was substrate-dependent, we performed jump dilution experiments in which GUS was pre-incubated with either inhibitor alone or a combination of inhibitor and substrate. Indeed, the onset of steady-state inhibition was only observed when GUS was incubated with both inhibitor and substrate (**Figure 6.6C**), suggesting that the analogs of **UNC10201652** also inhibit GUS via substrate-dependent slow-binding.

Previous work has shown that **UNC10201652** is pH dependent, where inhibition is significantly reduced at lower pH values[162]. We analyzed the pH-dependence of piperazine-containing GUS inhibitors in the present study to further understand the

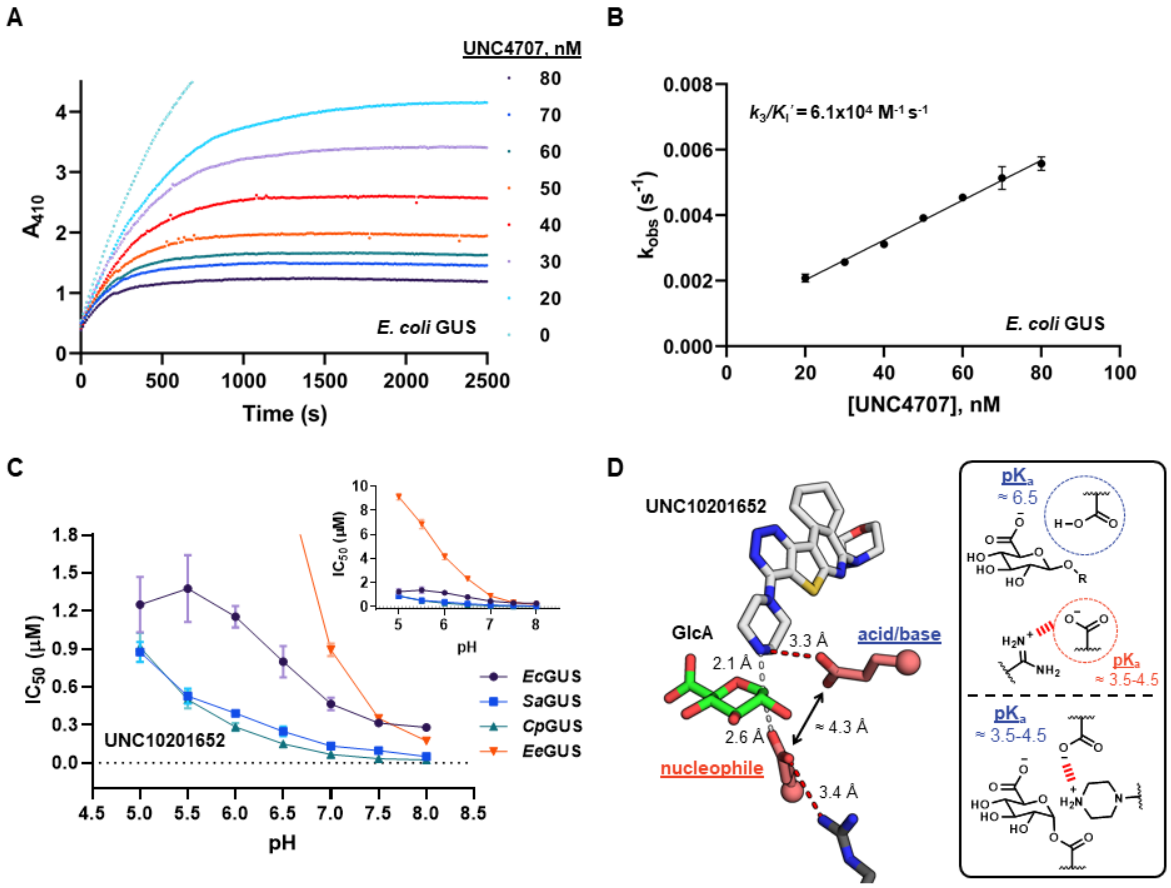


Figure 6.5 – Kinetic analysis and pH dependence studies of piperazine-containing GUS inhibitors.

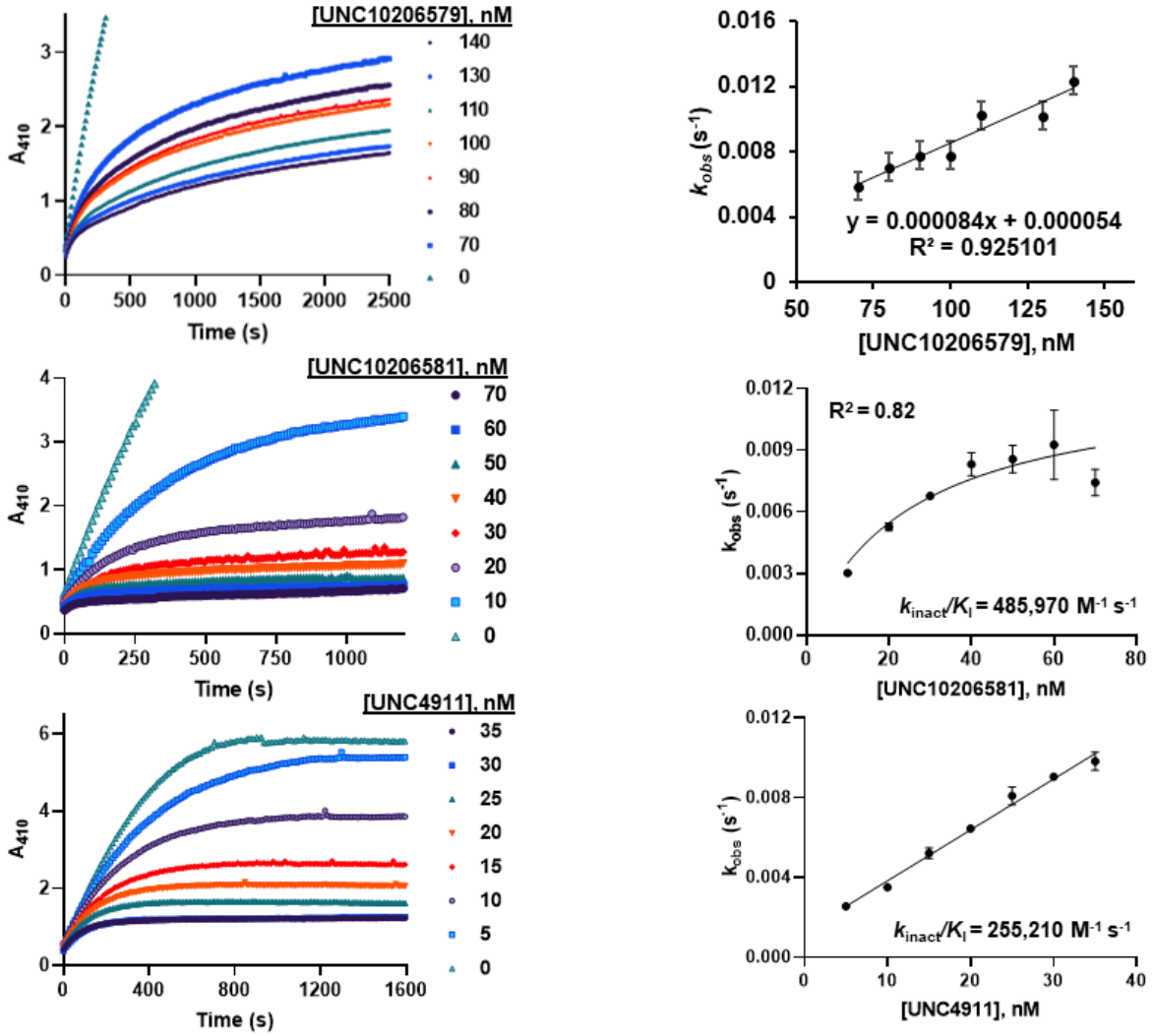


Figure 6.6-Kinetic analysis of select analogs.

mechanism of these gut bacterial GUS inhibitors. Potency was measured from pH 5 to 8 to assess the effect of pH on GUS inhibition by **UNC10201652**. IC_{50}s differ by 4-fold (***Ec*GUS**) to as much as 600-fold (*Ee*GUS) across the pH range tested (**Figure 6.5C** and **6.7**). The piperazine secondary amine has a pKa values of about 9, which suggests that the pH dependence of inhibition is likely due to titration of the catalytic acid/base glutamate of GUS. Previous studies of retaining glycoside hydrolases have revealed a significantly perturbed pKa of the catalytic acid/base (**Figure 6.5D**)[163]. Thus, a pKa of 6.5 for the catalytic acid/base of GHs has precedent. Indeed, the acid/base of retaining GHs has been observed to cycle between high and low pKa states during catalysis, which may control the substrate-dependent slow-binding observed for piperazine-containing inhibitors[163]. We hypothesize that the titration of this catalytic acid base significantly impacts the ionic interaction between the piperazine, thus drastically impacting potency (**Figure 6.5D**). The pH dependence outlined here corroborates previous work and suggests that these inhibitors likely display distinct potencies depending on where they act in the human GI tract.

Selectivity of piperazine-containing GUS inhibitors

To determine the selectivity of piperazine-containing inhibitors identified in the SAR effort, we tested inhibition with the closely related GH2 β-galactosidase from *E. coli* and the mammalian GUS ortholog from bovine liver **UNC10201652**. Partial inhibition of *E. coli* β-galactosidase is evident at 10 μM of select analogs (**Figure 6.8A**), suggesting that glycoside hydrolases beyond GUSs may be inhibited by piperazine-containing inhibitors. *E. coli* β-galactosidase is a tetramer like the Loop 1 GUS enzymes described here but does not appear to have loop structures near the active site that would aid in inhibitor binding. In contrast to *E. coli* β-galactosidase, bovine liver GUS displayed no inhibition at 10 μM for all

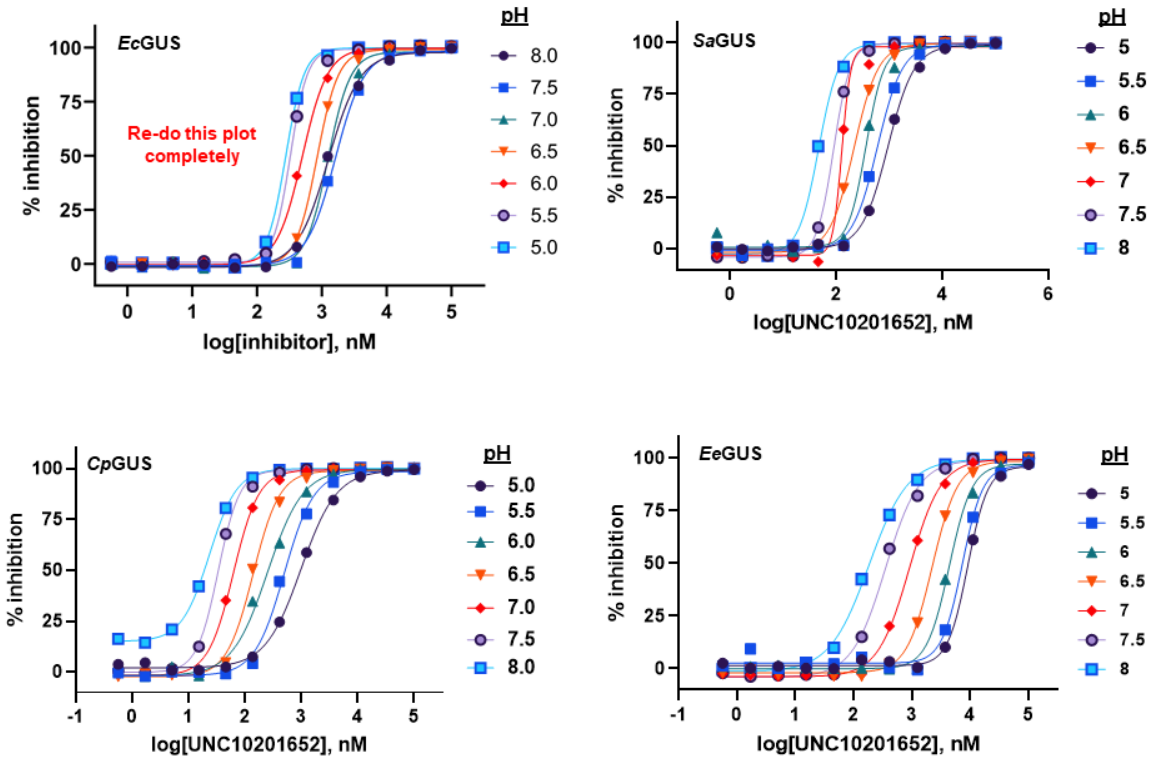


Figure 6.7 - IC50 plots of pH dependence studies.

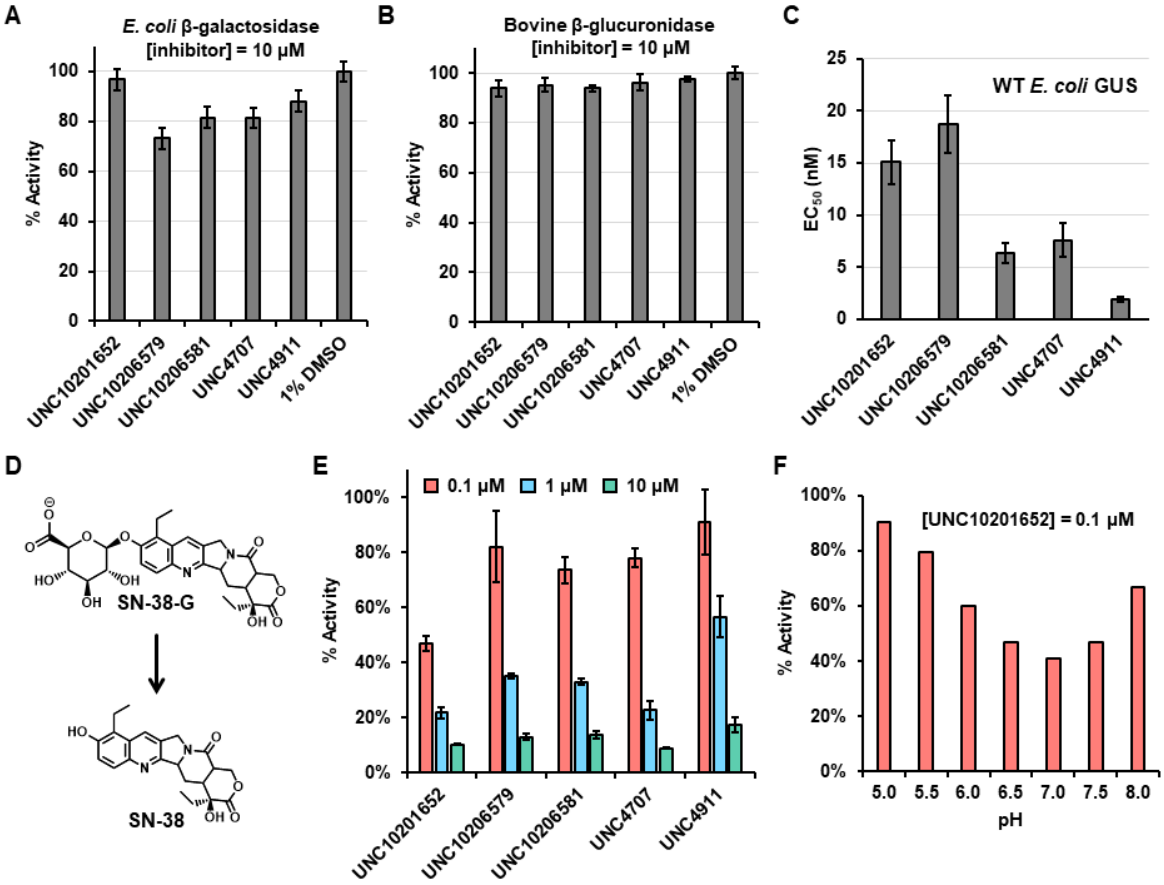


Figure 6.8 – Selectivity, in-cell potency, and in fimo inhibition by select analogs.

inhibitors tested (**Figure 6.8B**). This corroborates previous studies that have shown that other bacterial GUS inhibitors fail to inhibit the mammalian ortholog[123]. The rationale for selectivity for bacterial over human GUS is well-established to be the absence of the loop structures present in the Loop 1 GUS enzymes detailed in this study and in previous work[123].

Potency and cell toxicity of piperazine-containing inhibitors in *E. coli* cells

To assess the in-cell potency of select **UNC10201652** analogs, we performed cell-based inhibition assays with a WT and a previously generated inactive GUS strain of *E. coli* K12 MG1655[131]. All analogs revealed potent inhibition in *E. coli* cells (**Figure 6.8C**). The EC_{50} values largely resemble the SAR revealed from in vitro IC_{50} values with purified *E. coli* GUS (**Figure 6.8C** and **6.9**). Given that *E. coli* GUS is co-expressed with a system for glucuronide sensing and transport to shuttle substrate into the cell[164], it appears that the inhibitors tested penetrate *E. coli* cells. To confirm that inhibition observed in *E. coli* cells was due to selective inhibition of GUS and not cell toxicity, we tested cell growth of WT *E. coli* K12 MG1655 cells in the presence of 10 μM of select analogs and the parent compound **UNC10201652**. No effect was observed for *E. coli* cell growth with all compounds tested (**Figure 6.10**), suggesting that inhibition observed is due to selective targeting of GUS.

Dose-dependent and pH-dependent inhibition of SN-38-G turnover *in fimo*

Lastly, to assess GUS inhibition by piperazine-containing inhibitors in the biological complexity of the human gut microbiome, we tested inhibition of GUS-mediated SN-38-G turnover in human fecal supernatants, or *in fimo*[165]. In short, a human fecal sample lysate was prepared to remove fibrous debris and extract protein, the resultant extract was mixed with the relevant substrate SN-38-G and activity was measured continuously with a previously

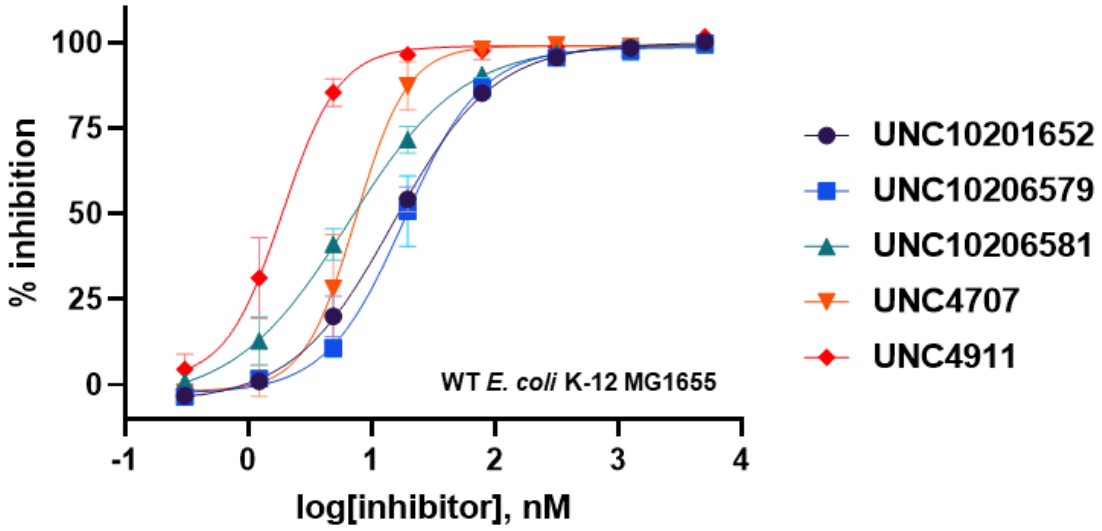


Figure 6.9 – Dose-dependent inhibition of GUS activity in *E. coli*.

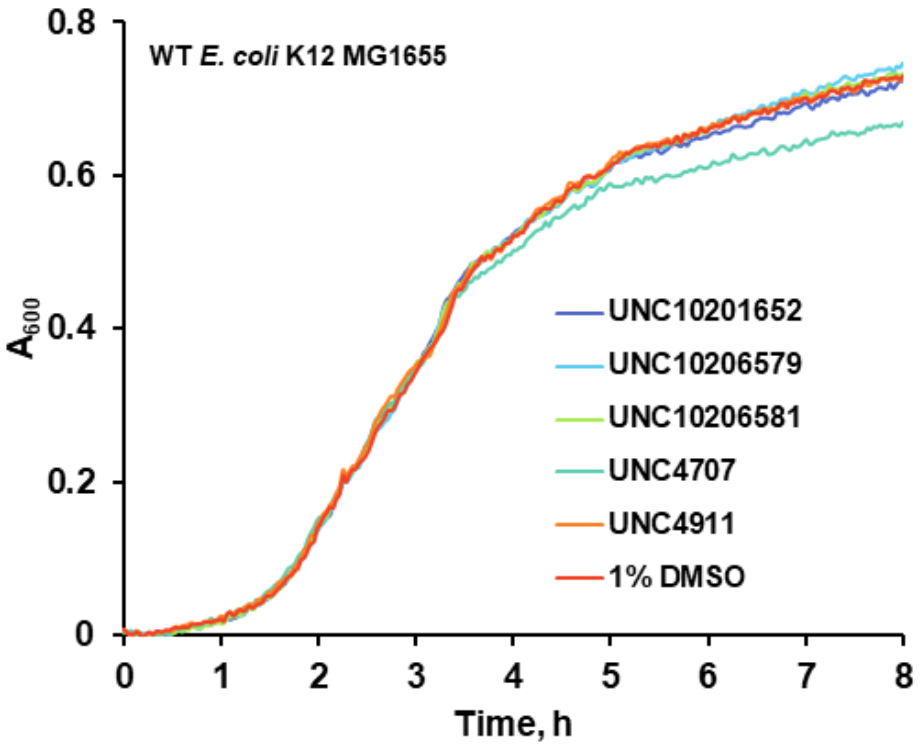


Figure 6.10 – Growth curves in the presence of 10 µM of select analogs.

characterized fluorescent assay[126]. We tested **UNC10201652** and select analogs for their ability to block SN-38-G turnover *in fimo*. All inhibitors display dose-dependent inhibition of SN-38-G turnover *(***Figure 6.8F***).* Interestingly, despite marked improvement of potency in vitro for selected analogs, all analogs reduced GUS activity to a similar extent, but UNC10201652 displayed the greatest reduction in activity at 100 nM. Since inhibition by the piperazine-containing chemotype is pH-dependent, we also measured how inhibition by **UNC10201652** varied with pH *in fimo*. Like *in vitro* pH studies, potency generally increased with increasing pH (**Figure 6.8F**). Unlike our in vitro findings, there is a reduction in potency at pH 8, which may be due to properties of other GUSs present in the fecal sample that we did not characterize in our in vitro studies. Taken together, this analysis shows that the compounds developed herein display potent inhibition of GUS-mediated SN-38-G turnover in complex human fecal sample. Given that the composition of the gut microbiome plays key roles in phenotypic outcomes, future studies will be needed to test inhibition against a range of human samples to assess the utility of these compounds in diverse biological samples.

CONCLUSION

Here we show the structure activity relationship of a piperazine-containing GUS inhibitor and assess the ability of these inhibitors to potently inhibit GUS activity in the complex setting of the gut microbiome. The SAR effort revealed the piperazine as an essential warhead for GUS inhibition, and the other moieties enable selectivity among similar GUS enzymes. Like previously characterized bacterial GUS inhibitors, the piperazine-containing inhibitors are also selective for bacterial GUS over the human ortholog and are potent in cell-based assays. Lastly, the inhibitors outlined here are also potent inhibitors of

GUS activity in a complex fecal extract, suggesting that piperazine-containing GUS inhibitors may be viable adjuvants for the alleviation of the toxic side effects of essential therapeutics.

EXPERIMENTAL SECTION

Protein expression and purification

All proteins were purified as previously described[131]. In brief, all proteins were expressed as N-terminal His-tag fusions in *E. coli* and purified by both Ni-affinity chromatography and size exclusion chromatography. All proteins displayed at least 95% purity by SDS-PAGE and were snap frozen in liquid nitrogen and stored at -80 °C.

In vitro GUS inhibition assay

Inhibition of GUS was determined with an endpoint assay utilizing the standard PNPG substrate. Inhibitors were diluted 3-fold, starting at 100 μM and pre-mixed with enzyme (15 nM) prior to initiating the reaction with PNPG (900 μM). Final reaction volume was 50 μL with 5 μL enzyme, 5 μL inhibitor, 5 μL water, 5 μL buffer (pH 7.5, 25 mM HEPES and NaCl), and 30 μL substrate. Reactions were incubated for 1 hour at 37 ℃ and then quenched with 50 μL of 0.2 M sodium carbonate. Endpoint absorbance values were then determined in a BMG labtech PHERAstar plate reader. Resultant data were converted to percent inhibition and fit with a 4-parameter logistic function for determination of the IC_{50} values.

Structure-guided docking of GUS inhibitors

Docking of novel analogs were based on the crystal structure of *Ee*GUS bound to UNC4917 (PDB: 6BO6). Ligand generation was performed in Phenix software suite and manual placement of resultant ligands were performed in PyMOL. Ligands were initially placed to match UNC4917 binding mode, and then conservatively modified to account for difference in the structures of the distinct GUS enzymes analyzed.

Bacterial cell toxicity assay

Bacterial cell toxicity was assessed as described previously[131]. In brief, overnight cultures of WT E. coli MG1655 K-12 were grown and sub-cultured the following morning. Sub-cultures were subject to 10 μM of various inhibitors and growth was monitored over the course of 8 hours to monitor inhibitor toxicity.

Compound synthesis

See **appendix**.

CHAPTER 7: FUTURE DIRECTIONS TO ELUCIDATE AND CONTROL THE CORE FUNCTIONS OF GUT BACTERIAL β-GLUCURONIDASE

While the structural landscape of bacterial GUS enzymes has become much clearer, we still know very little about the core function of GUS with respect to human health. Part of understanding what role GUS plays in human health will be the identification of the gut microbes primarily responsible for GUS activity. We now know the full complement of potential GUSs based on metagenomic analysis[136], but it remains unclear which GUSs are transcribed, expressed, and functionally active in the gut microbiome. One approach to determining the active GUSs in the gut microbiome is activity-based protein profiling (ABPP). Probes of retaining GUS enzymes have recently been developed to target the human GUS ortholog as a potential diagnostic for host diseases[166]. Since these probes target the mechanism of GUS that is shared between both human and bacterial orthologs, they can likely be utilized to profile functionally active GUS enzymes from the microbiome as well. Utilization of ABPP for the analysis of GUS has the potential to move the field from metagenomes to proteomes, a much-needed scientific advance given the massive genetic content of the gut microbiome. Even for bacterial GUS enzymes alone, 279 proteins has proved difficult to even begin to characterize, and while progress has been made over the past 9 years, it is not clear that the GUS enzymes that have been characterized are the ones that are expressed and processing medically relevant substrates.

Beyond the identification of relevant GUS enzymes, understanding how microbes interact with glucuronides will be critical to understanding the core function of GUS in the

gut microbiome. The most basic questions surrounding bacterial GUS remain unanswered. For example, why do microbes metabolize glucuronides? The current hypothesis is that microbes utilize glucuronic acid as a carbon source, and the ability to access glucuronides may serve as a competitive advantage in the gut microbiome. Testing this hypothesis will require a strong microbiology effort to assess the ability of numerous strains, possibly identified via ABPP of bacterial GUS, to grow and compete for glucuronide substrates. In addition to probing why microbes process glucuronides, we can also begin to answer questions posed in this dissertation by taking a microbiology approach. What is the role of the FMN-binding site in those GUSs that contain it? Finding a genetically tractable microbe that encodes an FMN-binding GUS may allow us to begin to probe the activity of these GUS enzymes in the context of their function in the gut microbiome. What are the cognate GlcA-containing substrates of gut microbes? Exposing microbes to a panel of substrates and measuring expression of relevant GUS enzymes could help accelerate the identification of relevant substrates. Once these basic questions are answered, select microbes can be utilized to colonize gnotobiotic mice to elucidate their effects on host physiology.

Another approach to identifying the core function of GUS is development of pan inhibitors of bacterial GUS. Given the genetic complexity of the gut microbiome and its location in the human gut, genetic approaches to knocking out bacterial GUS activity remain out of reach. The lack of genetic approaches to probing the microbiome puts chemical biology approaches at the forefront in this field. As has been shown in this dissertation, small molecules can serve as powerful tools to inhibit gut microbial enzymes. Unfortunately, the selective bacterial GUS inhibitors we have developed only target a small slice of the entire 'GUSome'. Thus, our current panel of inhibitors are not great tool compounds for assessing

the global role of bacterial GUS activity in the structure and function of the gut microbiome or host physiology. Future work, potentially utilizing technology developed for ABPP of GUS [166], could be used to block GUS activity. Coupling these chemical biology approaches to metagenomics, metabolomics, and proteomics may catalyze the discovery of relevant substrates, enzymes, and microbes, and ultimately reveal the core function of bacterial GUS in host physiology.

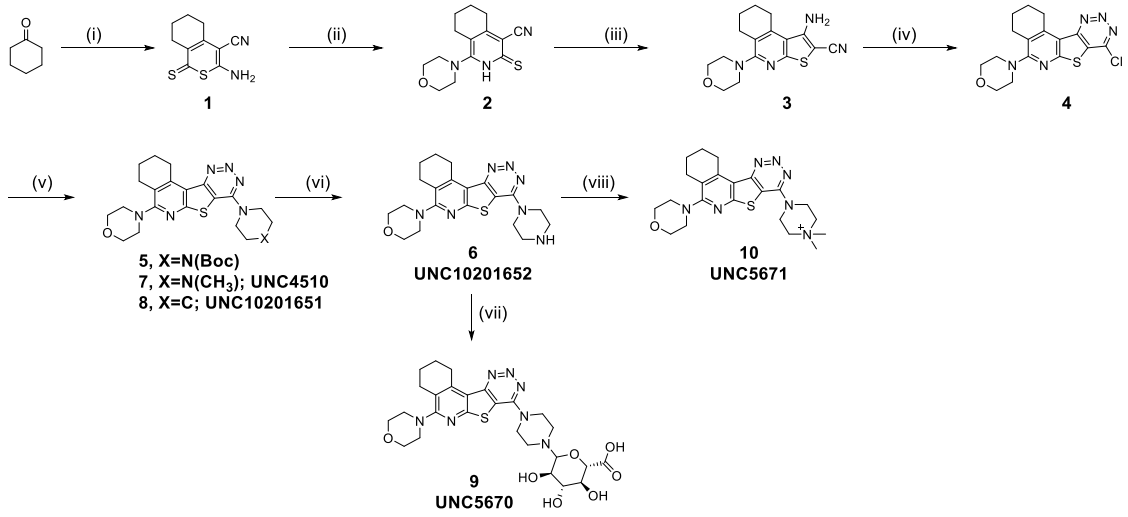
APPENDIX: CHEMICAL SYNTHESIS AND CHARACTERIZATION

Synthesis and Characterization of UNC10201652, UNC4917, UNC4510, UNC5671, UNC10201651, and UNC5670

All starting materials were commercially procured and were used without further purification, unless specified. Analytical LCMS data for all compounds were acquired using an Agilent 6110 Series system with the UV detector set to 220 nm. Samples were injected (<10 µL) onto an Agilent Eclipse Plus 4.6 × 50 mm, 1.8 µm, C18 column at room temperature. Mobile phases A (H2O + 0.1% acetic acid) and B (MeOH + 0.1% acetic acid) were used with a linear gradient from 10% to 100% B in 5.0 min, followed by a flush at 100% B for another 2 minutes with a flow rate of 1.0 mL/min. Mass spectra (MS) data were acquired in positive ion mode using an Agilent 6110 single quadrupole mass spectrometer with an electrospray ionization (ESI) source. Nuclear Magnetic Resonance (NMR) spectra were recorded on a Varian Mercury spectrometer at 400 MHz for proton (1H NMR); chemical shifts are reported in ppm (δ) relative to residual protons in deuterated solvent peaks. Due to intramolecular hydrogen-bonding, hydrogendeuterium exchange between the amide protons of the molecule and the deuterated solvent is slow and requires overnight equilibration for complete exchange. Normal phase column chromatography was performed with a Teledyne Isco CombiFlash®Rf using silica RediSep®Rf columns with the UV detector set to 220 nm and 254 nm. The mobile phases used are indicated for each compound. Reverse phase column chromatography was performed with a Teledyne Isco CombiFlash®Rf 200 using C18 RediSep®Rf Gold columns with the UV detector set to 220 nm and 254 nm. Mobile phases of A (H2O + 0.1% TFA) and B (MeCN) were used with default column gradients. Preparative HPLC was performed using an Agilent Prep 1200 series with the UV detector set to 220 nm and 254 nm. Samples were injected onto

a Phenomenex Luna 250 □ 30 mm, 5 µm, C18 column at room temperature. Mobile phases of A (H2O + 0.1% TFA) and B (MeOH or MeCN) were used with a flow rate of 40 mL/min. A general gradient of 0-15 minutes increasing from 10 to 100% B, followed by a 100% B flush for another 5 minutes. Small variations in this purification method were made as needed to achieve ideal separation for each compound. Analytical LCMS (at 220 nm) and NMR were used to establish the purity of targeted compounds. All compounds that were evaluated in biochemical and biophysical assays had >95% purity as determined by 1HNMR and LC-MS.

Scheme 1. Synthesis of **UNC10201652** and its analogs.



Reagents and conditions: (i) malononitrile, CS_2, Et_3N, MeOH, reflux, 40% (ii) morpholine, EtOH, reflux, 76% (iii) 2-chloroacetonitrile, KOH, DMF, 69% (iv) conc. HCl, $NaNO_2$, H_2O, rt, 92% (v) NHR^1R^2, K_2CO_3, CH_3CN, reflux, 19-43% (vi) 20% TFA in DCM , 0 ^oC, 99% (vii) AcOH, MeOH 59% (viii) MeI, MeOH, 70 ^oC microwave, 14%

General procedure A

To a solution of **4** (85 mg, 0.23 mmol, 1.0 eq.) in CH_3CN (2 mL) was added K_2CO_3 (300 mg) and the respective secondary amine reactant (0.59 mmol, 2.5 eq.), and the reaction mixture was

heated under reflux overnight. Upon completion, the reaction was quenched by addition of 20 mL of sat. $NaHCO_3$ solution. The aqueous phase was extracted with CH_2Cl_2 (3×20 mL) and the organic layers were combined, washed with brine, dried over anhydrous Na_2SO_4, and filtered. Solvent was removed by rotary evaporation to obtain a dark brown crude material. The crude material was adsorbed onto silica gel and purified by silica gel flash column chromatography (0 - 20% (5% NH_4OH in MeOH) in DCM) to afford desired product in quantitative yields.

3-amino-1-thioxo-5,6,7,8-tetrahydro-1H-isothiochromene-4-carbonitrile (1): To a solution of cyclohexanone (5.0 g, 5.3 mL, 51 mmol, 1.0 eq.) in 16 mL of methanol was slowly added carbon disulfide (7.8 g, 6.1 mL, 102 mmol, 2.0 eq.) and malononitrile (3.4 g, 51 mmol, 1.0 eq.), and the mixture was stirred for 5 minutes while maintaining the temperature below 20°C. Triethylamine (2.5 mL) was added, and the reaction mixture was stirred overnight at rt. Precipitated product was filtered, washed with methanol, and vacuum dried to provide **1** as an orange solid (4.3 g, 40%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.69 (s, 2H), 2.68 (m, 2H), 2.59 (m, 2H), 1.68 (m, 4H). LC-MS (λ = 254 nm): 99%, t_R = 5.9 min. MS (ESI+): 223 $[M+H]^+$.

1-morpholino-3-thioxo-2,3,5,6,7,8-hexahydroisoquinoline-4-carbonitrile (2): To a suspension of **1** (4.0 g, 18 mmol, 1.0 eq.) in ethanol (15 mL) was added morpholine (8 mL, 90 mmol, 5 eq.) and the mixture was heated under reflux overnight. Precipitated product was cooled to rt, degassed with nitrogen, filtered, washed with ethanol, and vacuum dried to obtain **2** as an orange solid (3.8 g, 76%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 4.33 – 4.27 (m, 1H), 3.74 (m, 4H), 3.67 (m, 4H), 2.61 (t, J = 5.8 Hz, 2H), 2.41 (t, J = 5.8 Hz, 2H), 1.77 – 1.65 (m, 2H), 1.62 – 1.52 (m, 2H). LC-MS (λ = 254 nm): 99%, t_R = 4.9 min. MS (ESI+): 276 $[M+H]^+$.

1-amino-5-morpholino-6,7,8,9-tetrahydrothieno[2,3-c]isoquinoline-2-carbonitrile (3): To a solution of **2** (3.0 g, 11 mmol, 1.0 eq.) in 22 mL of DMF was added 2-chloroacetonitrile (0.8 mL, 11.99 mmol, 1.1 eq.), and the mixture was stirred at rt for 1 hr. Then the first portion of aq. KOH (10% w/v, 5.5 mL) was added to the reaction mixture and it was stirred at rt overnight, after which a second portion of aq. KOH (10% w/v, 5.5 mL) was added to the reaction mixture and stirred for another 4 hr at room temperature. Water (50 mL) was added to the precipitated solid product after which it was filtered and vacuum dried to obtain **3** as a pinkish yellow solid (2.36 g, 69%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 6.35 (s, 2H), 3.84 – 3.70 (m, 4H), 3.27 (t, *J* = 6.5 Hz, 2H), 3.19 – 3.10 (m, 4H), 2.64 (t, *J* = 5.7 Hz, 2H), 1.88 – 1.75 (m, 2H), 1.70 – 1.58 (m, 2H). LC-MS (λ = 254 nm): 99%, t_R = 5.9 min. MS (ESI+): 315 $[M+H]^+$.

4-(8-chloro-1,2,3,4-tetrahydro-[1,2,3]triazino[4',5':4,5]thieno[2,3-c]isoquinolin-5-yl)morpholine (4): A solution of sodium nitrite (1.5 g, 22 mmol, 3.0 eq.) in water (7.3 mL) was added dropwise to a suspension of **3** (2.3 g, 7.3 mmol, 1.0 eq.) in conc. HCl acid (15 mL) at 0-5 °C over a period of 30min. The mixture was stirred for 1 hr at 0-5 °C and then allowed to stir at rt overnight. Water (100 mL) was added to the precipitated product after which it was filtered, washed with water, and vacuum dried to obtain **4** as a yellow solid (2.4 g, 92%). ^{1}H NMR (400MHz ,$CDCl_3$-d) □ = 3.88 (t, *J* = 4.7 Hz, 4 H), 3.73 (t, *J* = 6.7 Hz, 2 H), 3.43 (t, *J* = 4.7 Hz, 4 H), 2.72 (t, *J* = 5.7 Hz, 2 H), 2.07 - 1.96 (m, 2 H), 1.88 - 1.76 (m, 2 H)LC-MS (λ = 254 nm): 99%, t_R = 6.4 min. MS (ESI+): 362 $[M+H]^+$.

Tert-butyl 4-(5-morpholino-1,2,3,4-tetrahydro-[1,2,3]triazino[4',5':4,5]thieno[2,3-c]isoquinolin-8-yl)piperazine-1-carboxylate (5): To a solution of **4** (85 mg, 0.23 mmol, 1.0 eq.) in CH_3CN (2 mL) was added K_2CO_3 (300 mg) and 1-Boc-piperazine (110 mg, 0.59 mmol,

2.5 eq.) and the reaction mixture was heated under reflux overnight. Upon completion, the reaction was quenched by addition of 20 mL of sat. $NaHCO_3$ solution. The aqueous phase was extracted with CH_2Cl_2 (3×20 mL) and the organic layers were combined, washed with brine, dried over anhydrous Na_2SO_4, and filtered. Solvent was removed by rotary evaporation to obtain a dark brown crude material. The crude material was adsorbed onto silica gel and purified by silica gel flash column chromatography (0 - 20% (5% NH_4OH in MeOH) in DCM) to afford **5** as a pale yellow solid (50 mg, 43%).^{1}H NMR (400 MHz, $CDCl_3$-*d*) δ 4.08 – 4.02 (m, 4H), 3.92 – 3.83 (m, 4H), 3.77 (t, *J* = 6.7 Hz, 2H), 3.68 – 3.60 (m, 4H), 3.38 – 3.30 (m, 4H), 2.73 (t, *J* = 5.8 Hz, 2H), 2.03 – 1.94 (m, 2H), 1.85 – 1.76 (m, 2H), 1.49 (s, 9H). LC-MS (λ = 254 nm): 99%, t_R = 6.5 min. MS (ESI+): 512 $[M+H]^+$.

4-(8-(piperazin-1-yl)-1,2,3,4-tetrahydro-[1,2,3]triazino[4',5':4,5]thieno[2,3-c]isoquinolin-5-yl)morpholine (UNC10201652, 6): To a solution of **5** (50 mg, 0.098 mmol) in CH_2Cl_2 (1 mL) at 0 °C was added trifluoroacetic acid (1 mL) and the reaction mixture was stirred until completion at room temperature. The solution was washed with saturated $NaHCO_3$ solution, dried over anhydrous Na_2SO_4, filtered, and vacuum concentrated to obtain **6** as a yellow solid (40 mg, 99%). ^{1}H NMR (400 MHz, $CDCl_3$-*d*) δ 9.91 (s, 1H), 4.36 (s, 4H), 3.92 – 3.79 (m, 4H), 3.68 (t, *J* = 6.2 Hz, 2H), 3.48 (s, 4H), 3.40 – 3.31 (m, 4H), 2.71 (t, *J* = 5.6 Hz, 2H), 1.97 (m, 2H), 1.79 (m, 2H). LC-MS (λ = 254 nm): 99%, t_R = 5.0 min. MS (ESI+): 412 $[M+H]^+$.

4-(8-(4-methylpiperazin-1-yl)-1,2,3,4-tetrahydro-[1,2,3]triazino[4',5':4,5]thieno[2,3-c]isoquinolin-5-yl)morpholine (UNC4510, 7): Compound **7** was synthesized using 1-methylpiperazine by general procedure A to afford a pale yellow solid (74% yield). ^{1}H NMR (400 MHz, $CDCl_3$-*d*) *δ* 4.12 – 4.05 (m, 4H), 3.90 – 3.82 (m, 4H), 3.76 (t, *J* = 6.7 Hz, 2H), 3.37

– 3.28 (m, 4H), 2.72 (t, *J* = 5.9 Hz, 2H), 2.63 – 2.55 (m, 4H), 2.35 (s, 3H), 2.03 – 1.91 (m, 2H), 1.84 – 1.74 (m, 2H). LC-MS (λ = 254 nm): 99%, t_R = 5.7 min. MS (ESI+): 426 $[M+H]^+$.

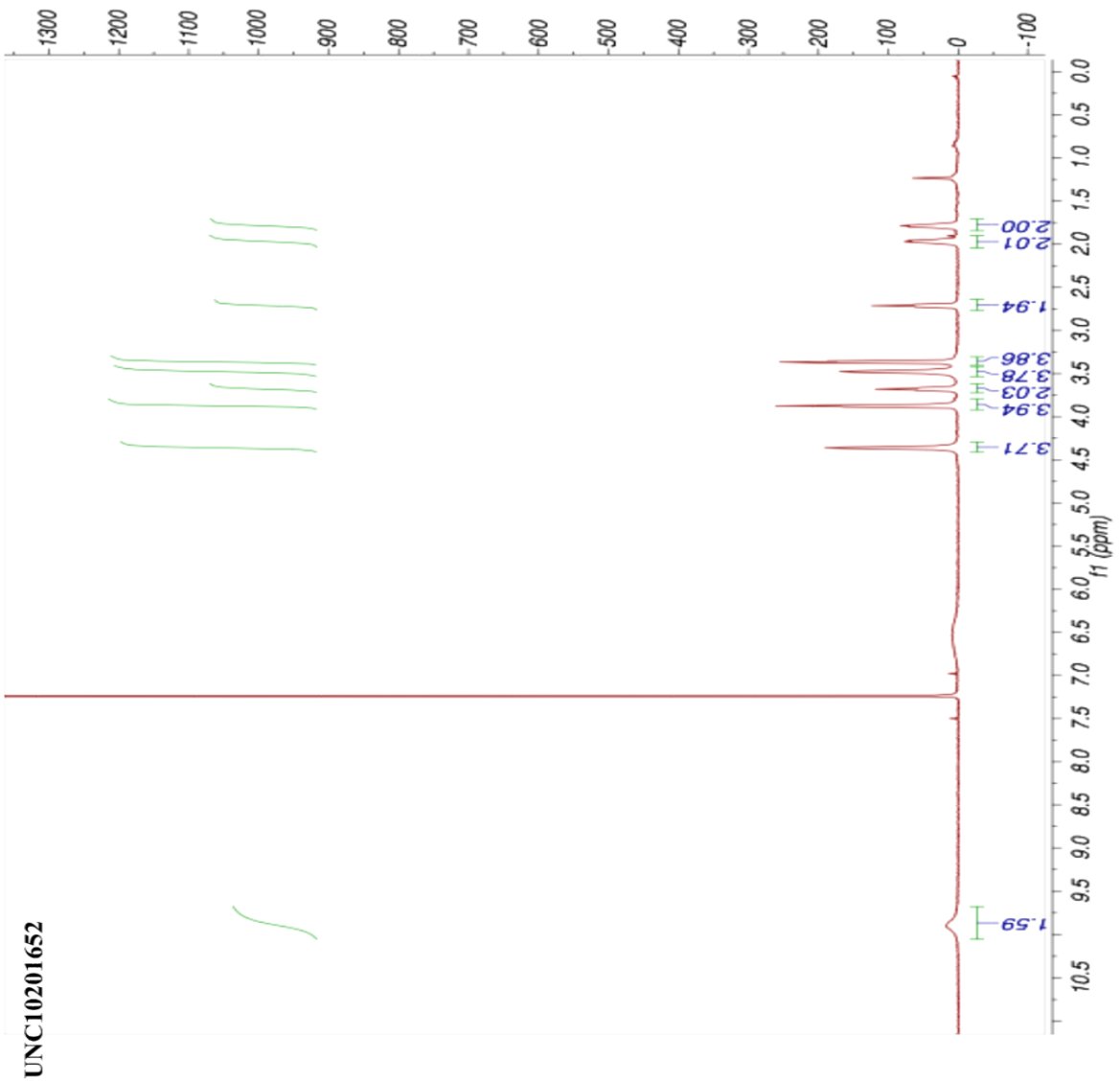
4-(8-(piperidin-1-yl)-1,2,3,4-tetrahydro-[1,2,3]triazino[4',5':4,5]thieno[2,3-c]isoquinolin-5-yl)morpholine (UNC10201651, 8) Compound **8** was synthesized using piperidine by general procedure A to afford a pale yellow solid (Yield 54%). ^{1}H NMR (400 MHz, $CDCl_3$-*d*) δ 4.02 (dd, *J* = 6.0, 3.8 Hz, 4H), 3.90 – 3.82 (m, 4H), 3.76 (t, *J* = 6.6 Hz, 2H), 3.35 – 3.27 (m, 4H), 2.72 (t, *J* = 5.9 Hz, 2H), 2.02 – 1.91 (m, 2H), 1.77 (m, 8H). LC-MS (λ = 254 nm): 99%, t_R = 7.2 min. MS (ESI+): 411 $[M+H]^+$.

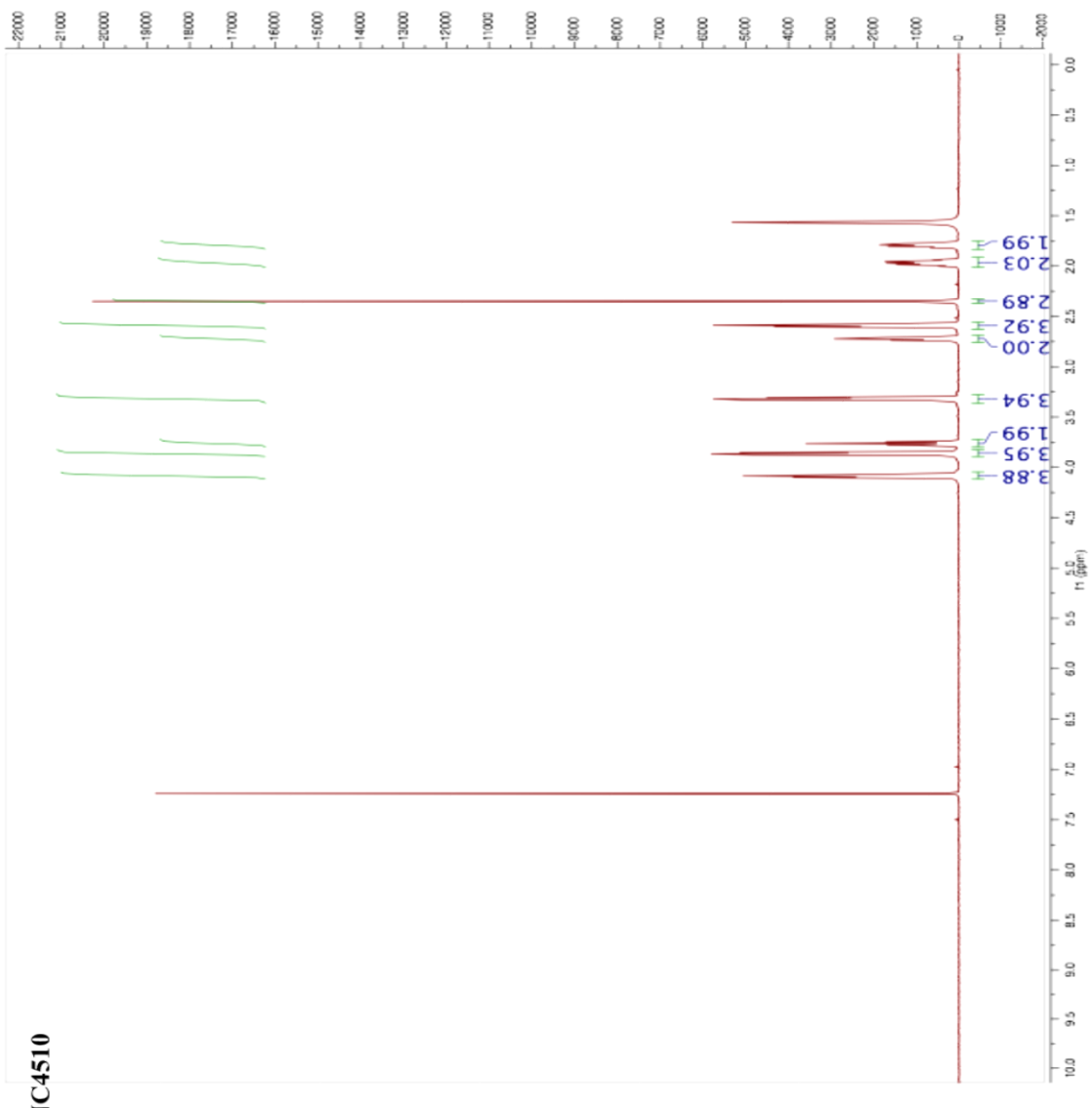
(2S,3S,4S,5R)-3,4,5-trihydroxy-6-(4-(5-morpholino-1,2,3,4-tetrahydro-[1,2,3]triazino[4',5':4,5]thieno[2,3-c]isoquinolin-8-yl)piperazin-1-yl)tetrahydro-2H-pyran-2-carboxylic acid (UNC5670, 9): In a flame-dried 100 mL flask equipped with a stir bar were added **6** (200 mg, 0.485 mmol), sodium (2R,3R,4R,5S)-3,4,5,6-tetrahydroxytetrahydro-2H-pyran-2-carboxylate (0.315 g, 1.46 mmol), and an acetic acid and methanol mixture (ratio 1:8; 15ml). The resultant mixture was allowed to stir at rt for 5 min before heating to 40 ^{o}C and stirred overnight. Upon consumption of **6** as indicated by the LC/MS, the mixture was cooled to rt and concentrated under reduced pressure. The mixture was azeotroped with toluene (5x) to evaporate residual acetic acid. A precipitate was formed upon addition of 1M aqueous sodium hydroxide (2ml) to the crude mixture. The crude mixture was filtered and the precipitate was subsequently washed with acetone (15ml) and DCM (15ml). The precipitate was dried under high-vacuum overnight to afford **9** as a pale white amorphous solid (170 mg, 59.52%). Compound **9** was isolated as an inseparable 1:1 mixture of α and β diastereomers. The diasteromeric ratio was determined using 1H NMR based on anomeric protons for α (5.13 (d, *J* = 3.5 Hz, 1 H)) and β (4.51 (d, J = 7.8 Hz, 1 H)). ^{1}H NMR

(400MHz ,METHANOL-d_4) d = 5.13 (d, *J* = 3.5 Hz, 1 H), 4.71 (d, *J* = 1.2 Hz, 1 H), 4.51 (d, *J* = 7.8 Hz, 1 H), 4.34 (s, 1 H), 4.31 - 4.26 (m, 5 H), 4.25 (s, 1 H), 4.06 (d, *J* = 1.6 Hz, 1 H), 3.86 (t, *J* = 4.3 Hz, 8 H), 3.79 (d, *J* = 9.8 Hz, 2 H), 3.76 - 3.63 (m, 4 H), 3.57 - 3.49 (m, 10 H), 3.42 (d, *J* = 3.5 Hz, 1 H), 3.40 - 3.36 (m, 2 H), 3.20 - 3.15 (m, 1 H), 2.76 (t, *J* = 5.7 Hz, 4 H), 2.01 - 1.91 (m, 4 H), 1.83 - 1.72 (m, 4 H). LC-MS (λ = 254 nm): 99%, t_R = 1.33 min. MS (ESI+): 588.50 $[M+H]^+$.

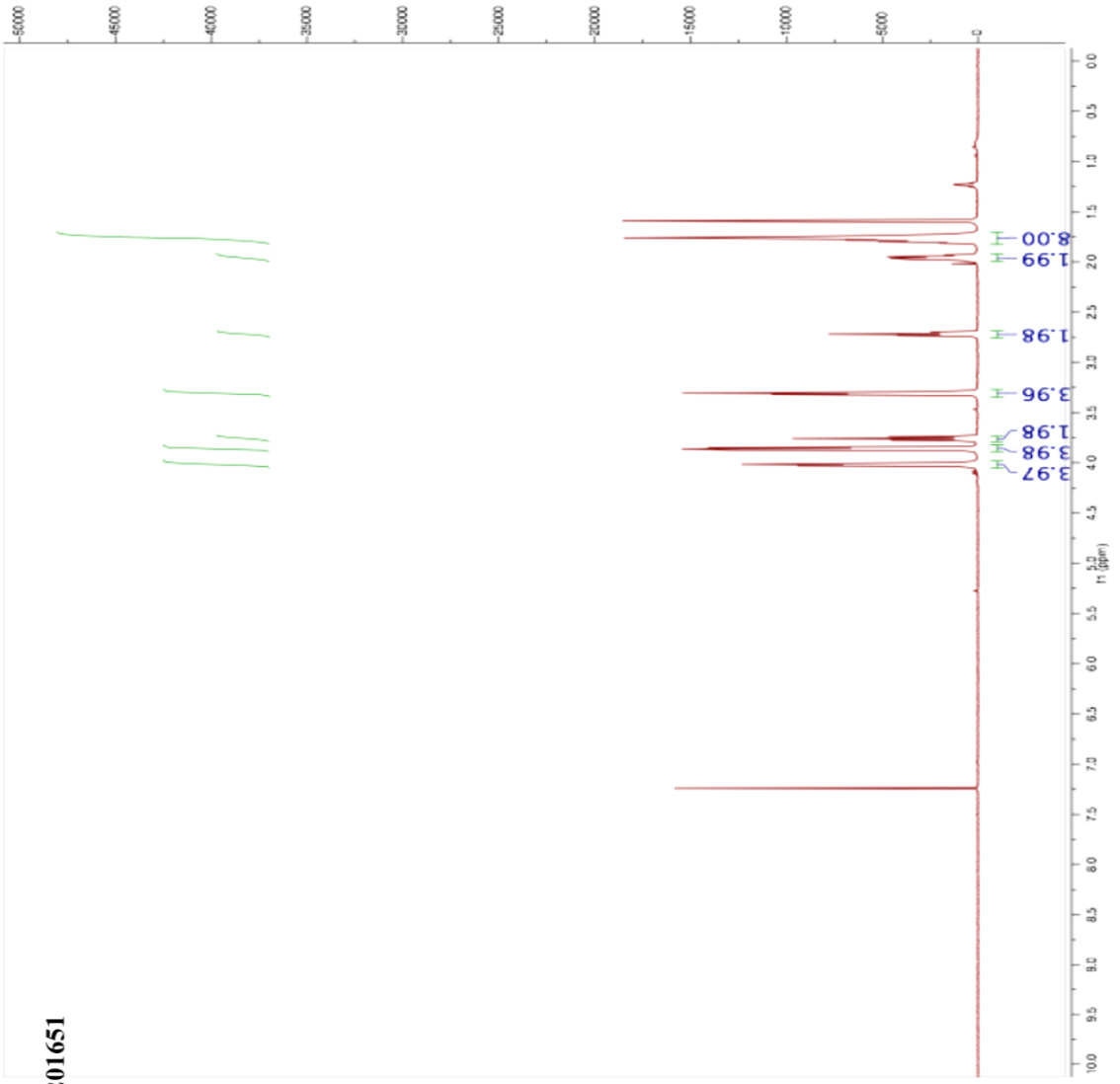
1,1-dimethyl-4-(5-morpholino-1,2,3,4-tetrahydro-[1,2,3]triazino[4',5':4,5]thieno[2,3-c]isoquinolin-8-yl)piperazin-1-ium (UNC5671, 10): In a flame dried 10 mL microwave vial equipped with a stir bar were added **6** (100 mg, 0.242 mmol) and K_2CO_3 (0.033 g, 0.242 mmol) followed by methanol (2 mL) and methyl iodide (0.103 g, 0.045 mL, 0.728 mmol). The resultant mixture was irradiated at 70 ^oC for 30 min, 250 W. The solvent was evaporated under reduced pressure and the crude mixture was first purified using silica gel flash column chromatography (0 - 20% (5% NH4OH in MeOH) in DCM)). The desired material was isolated, concentrated, and re-dissolved in methanol (0.5ml) and then purified further by preparative-HPLC ((H2O + 0.1% TFA)/MeCN) to afford **9** as a pale white solid (15.1 mg, 14.2%). ^1H NMR (400 MHz, CD_3OD-d4) δ = 4.42 (t, *J* = 4.7 Hz, 4H), 3.87 (t, *J* = 4.3 Hz, 4H), 3.78 (t, *J* = 4.7 Hz, 4H), 3.59 (t, *J* = 6.3 Hz, 2H), 3.38 (s, 6H), 3.34 (t, *J* = 4.3 Hz, 4H), 2.77 (t, *J* = 5.5 Hz, 2H), 2.04 - 1.94 (m, 2H), 1.84 - 1.75 (m, 2H). LC-MS (λ = 254 nm): 99%, t_R = 4.56 min. MS (ESI+): 440.25 $[M+H]^+$.

1H NMR Spectra:

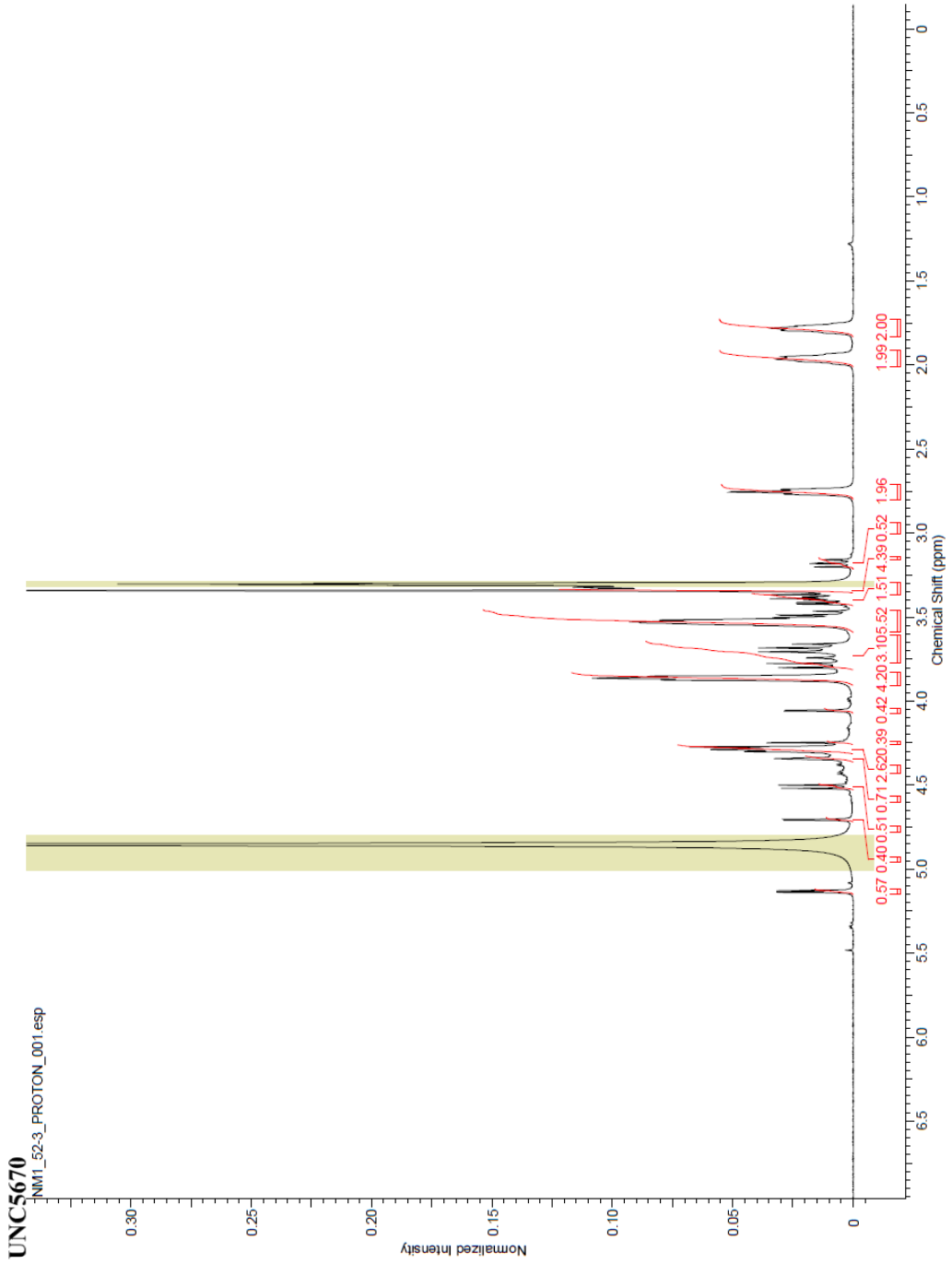




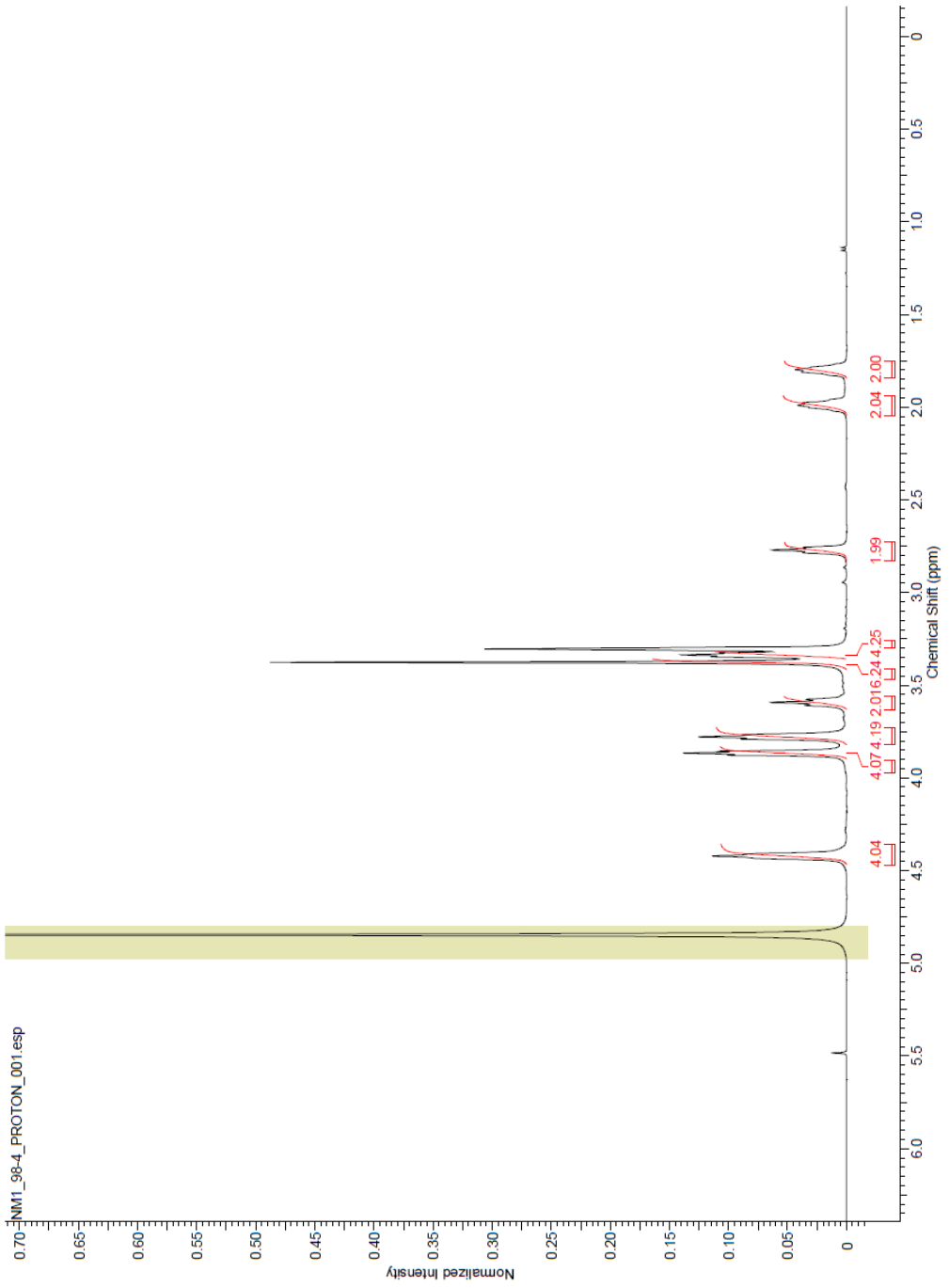
UNC4510



UNC10201651


UNC5670
NM1_52-3_PROTON_001.esp
Normalized Intensity
Chemical Shift (ppm)
0.57 0.40 0.51 0.71 2.62 0.39 0.42 4.20 3.10 5.52 1.51 4.39 0.52 1.96 1.99 2.00

UNC5671


NM1_98-4_PROTON_001.esp
Normalized Intensity
Chemical Shift (ppm)
4.04
4.07
4.19
2.01
6.24
4.25
1.99
2.04
2.00

REFERENCES

(1) Goldin, B. R. (1986) In situ bacterial metabolism and colon mutagens. *Annu. Rev. Microbiol. 40*, 367–93.

(2) Goldin, B. R., and Gorbach, S. L. (1976) The relationship between diet and rat fecal bacterial enzymes implicated in colon cancer. *J. Natl. Cancer Inst. 57*, 371–5.

(3) Lombardi, P., Goldin, B., Boutin, E., and Gorbach, S. L. (1978) Metabolism of androgens and estrogens by human fecal microorganisms. *J. Steroid Biochem. 9*, 795–801.

(4) Winter, J., and Bokkenheuser, V. D. (1987) Bacterial metabolism of natural and synthetic sex hormones undergoing enterohepatic circulation. *J. Steroid Biochem. 27*, 1145–9.

(5) Asano, Y., Hiramoto, T., Nishino, R., Aiba, Y., Kimura, T., Yoshihara, K., Koga, Y., and Sudo, N. (2012) Critical role of gut microbiota in the production of biologically active, free catecholamines in the gut lumen of mice. *Am. J. Physiol. Gastrointest. Liver Physiol. 303*, G1288–G1295.

(6) Kim, D. H., and Jin, Y. H. (2001) Intestinal bacterial beta-glucuronidase activity of patients with colon cancer. *Arch. Pharm. Res. 24*, 564–7.

(7) Grosse, L., Paquet, S., Caron, P., Fazli, L., Rennie, P. S., Belanger, A., and Barbier, O. (2013) Androgen Glucuronidation: An Unexpected Target for Androgen Deprivation Therapy, with Prognosis and Diagnostic Implications. *Cancer Res. 73*, 6963–6971.

(8) Plotnikoff, G. a. (2014) Three Measurable and Modifiable Enteric Microbial Biotransformations Relevant to Cancer Prevention and Treatment. *Glob. Adv. Heal. Med. 3*, 33–43.

(9) Gloux, K., and Anba-Mondoloni, J. (2016) Unique beta-glucuronidase locus in gut microbiomes of Crohn's disease patients and unaffected first-degree relatives. *PLoS One 11*, 1–17.

(10) Dutton, G. J. (1966) Glucuronic Acid, free and combined, biochemistry, pharmacology, and medicine. *Acad. Press*. Academic Press, New York.

(11) Dutton, G. J. (1980) Glucuronidation of drugs and other compounds. CRC press, Boca Raton.

(12) Peekhaus, N., and Conway, T. (1998) What's for Dinner?: Entner-Doudoroff Metabolism in Escherichia coli MINIREVIEW *180*, 3495–3502.

(13) Jain, S., Drendel, W. B., Chen, Z. W., Mathews, F. S., Sly, W. S., and Grubb, J. H. (1996) Structure of human beta-glucuronidase reveals candidate lysosomal targeting and active-site motifs. *Nat. Struct. Biol. 3*, 375–381.

(14) Sly, W. S., Quinton, B. A., McAlister, W. H., and Rimoin, D. L. (1973) Beta glucuronidase deficiency: Report of clinical, radiologic, and biochemical features of a new mucopolysaccharidosis. *J. Pediatr. 82*, 249–257.

(15) Sperker, B., Backman, J. T., and Kroemer, H. K. (1997) The role of beta-glucuronidase

in drug disposition and drug targeting in humans. *Clin.Pharmacokinet. 33*, 18–31.

(16) Wilson, K. J., Hughes, S. G., and Jefferson, R. a. (1992) The Escherichia coli gus Operon: Induction and Expression of the gus Operon in E. coli and the Occurrence and Use of GUS in Other Bacteria. *Gus Protoc.* ACADEMIC PRESS, INC.

(17) Roberts, M. S., Magnusson, B. M., Burczynski, F. J., and Weiss, M. (2002) Enterohepatic circulation: physiological, pharmacokinetic and clinical implications. *Clin. Pharmacokinet. 41*, 751–790.

(18) Vítek, L., and Carey, M. C. (2003) Enterohepatic cycling of bilirubin as a cause of "black" pigment gallstones in adult life. *Eur. J. Clin. Invest. 33*, 799–810.

(19) Fevery, J., Blanckaert, N., Leroy, P., Michiels, R., and Heirwegh, K. P. (1983) Analysis of bilirubins in biological fluids by extraction and thin-layer chromatography of the intact tetrapyrroles: application to bile of patients with Gilbert's syndrome, hemolysis, or cholelithiasis. *Hepatology 3*, 177–183.

(20) Vítek, L., Kotal, P., Jirsa, M., Malina, J., Cerná, M., Chmelar, D., and Fevery, J. (2000) Intestinal colonization leading to fecal urobilinoid excretion may play a role in the pathogenesis of neonatal jaundice. *J. Pediatr. Gastroenterol. Nutr. 30*, 294–298.

(21) Stocker, R., Yamamoto, Y., McDonagh, A. F., Glazer, A. N., and Ames, B. N. (1987) Bilirubin is an antioxidant of possible physiological importance. *Science 235*, 1043–6.

(22) Vítek, L., and Ostrow, J. D. (2009) Bilirubin chemistry and metabolism; harmful and protective aspects. *Curr. Pharm. Des. 15*, 2869–2883.

(23) Maki, T. (1966) Pathogenesis of calcium bilirubinate gallstone: role of E. coli, beta-glucuronidase and coagulation by inorganic ions, polyelectrolytes and agitation. *Ann. Surg. 164*, 90–100.

(24) Shen, H., Ye, F., Xie, L., Yang, J., Li, Z., Xu, P., Meng, F., Li, L., Chen, Y., Bo, X., Ni, M., and Zhang, X. (2015) Metagenomic sequencing of bile from gallstone patients to identify different microbial community patterns and novel biliary bacteria. *Sci. Rep. 5*, 17450.

(25) Ye, F., Shen, H., Li, Z., Meng, F., Li, L., Yang, J., Chen, Y., Bo, X., Zhang, X., and Ni, M. (2016) Influence of the biliary system on biliary bacteria revealed by bacterial communities of the human biliary and upper digestive tracts. *PLoS One 11*.

(26) Raftogianis, R., Creveling, C., Weinshilboum, R., and Weisz, J. (2000) Estrogen metabolism by conjugation. *J. Natl. Cancer Inst. Monogr.* 113–124.

(27) Graef, V., Furuya, E., and Nishikaze, O. (1977) Hydrolysis of Steroid Glucuronides with Beta-Glucuronidase Preparations from Bovine Liver, Helix pomatia, and E. coli. *Clin. Chem. 23*, 532–535.

(28) Sher, A., and Rahman, M. A. (2000) Enterohepatic Recycling of Estrogen and its Relevance with Female Fertility. *Arch Pharm Res 23*, 513–517.

(29) Sandberg, A. A., Kirdani, R. Y., Back, N., Weyman, P., and Slaunwhite, W. R. (1967) Biliary excretion and enterohepatic circulation of estrone and estriol in rodents. *Am. J.*

Physiol. 213, 1138–1142.

(30) Kwa, M., Plottel, C. S., Blaser, M. J., and Adams, S. (2016) The Intestinal Microbiome and Estrogen Receptor – Positive Female Breast Cancer. *J. Natl. Cancer Inst. 108*, 1–10.

(31) Ulmer, J. E., Vilén, E. M., Namburi, R. B., Benjdia, A., Beneteau, J., Malleron, A., Bonnaffé, D., Driguez, P. A., Descroix, K., Lassalle, G., Le Narvor, C., Sandström, C., Spillmann, D., and Berteau, O. (2014) Characterization of glycosaminoglycan (GAG) sulfatases from the human gut symbiont Bacteroides thetaiotaomicron reveals the first GAG-specific bacterial endosulfatase. *J. Biol. Chem. 289*, 24289–24303.

(32) Kreek, M. J., Guggenheim, F. G., Ross, J. E., and Tapley, D. F. (1963) Glucuronide formation in the transport of testosterone and androstenedione by rat intestine. *Biochim. Biophys. Acta 74*, 418–427.

(33) Yamanaka, H., Nakajima, M., Katoh, M., and Yokoi, T. (2007) Glucuronidation of thyroxine in human liver, jejunum, and kidney microsomes. *Drug Metab. Dispos. 35*, 1642–1648.

(34) Hazenberg, M. P., de Herder, W. W., and Visser, T. J. (1988) Hydrolysis of iodothyronine conjugates by intestinal bacteria. *FEMS Microbiol. Rev. 4*, 9–16.

(35) DiStefano, J. J., de Luze, A., and Nguyen, T. T. (1993) Binding and degradation of 3,5,3'-triiodothyronine and thyroxine by rat intestinal bacteria. *Am. J. Physiol. 264*, E966-72.

(36) Zhang, J., and Lazar, M. a. (2000) The mechanism of action of thyroid hormones. *Annu. Rev. Physiol. 62*, 439–66.

(37) Eisenhofer, G., Aneman, Å., Friberg, P., Hooper, D., Fåndriks, L., Lonroth, H., Hunyady, B., and Mezey, E. (1997) Substantial production of Dopamine in the human gastrointestinal tract. *J. Clin. Endocrinol. Metab. 82*, 3864–3871.

(38) Flemström, G., Säfsten, B., and Jedstedt, G. (1993) Stimulation of mucosal alkaline secretion in rat duodenum by dopamine and dopaminergic compounds. *Gastroenterology 104*, 825–833.

(39) Haskel, Y., and Hanani, M. (1994) Inhibition of gastrointestinal motility by MPTP via adrenergic and dopaminergic mechanisms. *Dig. Dis. Sci. 39*, 2364–2367.

(40) Krishnaswamy, S., Duan, S. X., Von Moltke, L. L., Greenblatt, D. J., Sudmeier, J. L., Bachovchin, W. W., and Court, M. H. (2003) Serotonin (5-hydroxytryptamine) glucuronidation in vitro: assay development, human liver microsome activities and species differences. *Xenobiotica. 33*, 169–80.

(41) Wikoff, W. R., Anfora, A. T., Liu, J., Schultz, P. G., Lesley, S. a, Peters, E. C., and Siuzdak, G. (2009) Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl. Acad. Sci. U. S. A. 106*, 3698–3703.

(42) Hofmann, A. F. (2009) The enterohepatic circulation of bile acids in mammals: form and functions. *Front. Biosci. 14*, 2584–2598.

(43) Devlin, A. S., and Fischbach, M. A. (2015) A biosynthetic pathway for a prominent

class of microbiota-derived bile acids. *Nat. Chem. Biol. 11*, 685–690.

(44) Matern, S., Matern, H., Farthmann, E. H., and Gerok, W. (1984) Hepatic and extrahepatic glucuronidation of bile acids in man. Characterization of bile acid uridine 5'-diphosphate-glucuronosyltransferase in hepatic, renal, and intestinal microsomes. *J. Clin. Invest. 74*, 402–10.

(45) Alme, B., and Sjovall, J. (1980) Analysis of bile acid glucuronides in urine. identification of 3alpha,6alpha,12alpha-trihydroxy-5beta-cholanoic acid. *J. Steroid Biochem. 13*, 907–916.

(46) Kremmyda, L.-S., Tvrzicka, E., Stankova, B., and Zak, A. (2011) Fatty acids as biocompounds: their role in human metabolism, health and disease – a review. part 2: fatty acid physiological roles and applications in human health and disease. *Biomed. Pap. 155*, 195–218.

(47) Prakash, C., Zhang, J. Y., Falck, J. R., Chauhan, K., and Blair, I. A. (1992) 20-Hydroxyeicosatetraenoic acid is excreted as a glucuronide conjugate in human urine. *Biochem. Biophys. Res. Commun. 185*, 728–733.

(48) Barua, A. B., and Olson, J. A. (1986) Retinoyl beta-glucuronide: an endogenous compound of human blood. *Am J Clin Nutr 43*, 481–485.

(49) Little, J. M., Kurkela, M., Sonka, J., Jäntti, S., Ketola, R., Bratton, S., Finel, M., and Radominska-Pandya, A. (2004) Glucuronidation of oxidized fatty acids and prostaglandins B1 and E2 by human hepatic and recombinant UDP-glucuronosyltransferases. *J. Lipid Res. 45*, 1694–703.

(50) Salyers, A. A., and O'Brien, M. (1980) Cellular location of enzymes involved in chondroitin sulfate breakdown by Bacteroides thetaiotaomicron. *J. Bacteriol. 143*, 772–780.

(51) Koropatkin, N. M., Cameron, E. a., and Martens, E. C. (2012) How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol. 10*, 323–335.

(52) Pommier, Y. (2006) Topoisomerase I inhibitors: camptothecins and beyond. *Nat. Rev. Cancer 6*, 789–802.

(53) Cantore, M., Rabbi, C., Fiorentini, G., Oliani, C., Zamagni, D., Iacono, C., Mambrini, A., Del Freo, A., and Manni, A. (2004) Combined irinotecan and oxaliplatin in patients with advanced pre-treated pancreatic cancer. *Oncology 67*, 93–97.

(54) Robert, J., and Rivory, L. (1998) Pharmacology of irinotecan. *Drugs of Today*.

(55) Wallace, B. D., Wang, H., Lane, K. T., Scott, J. E., Orans, J., Koo, J. S., Venkatesh, M., Jobin, C., Yeh, L.-A., Mani, S., and Redinbo, M. R. (2010) Alleviating Cancer Drug Toxicity by Inhibiting a Bacterial Enzyme. *Science (80-.). 330*, 831–835.

(56) Wallace, B. D., Roberts, A. B., Pollet, R. M., Ingle, J. D., Biernat, K. A., Pellock, S. J., Venkatesh, M. K., Guthrie, L., O'Neal, S. K., Robinson, S. J., Dollinger, M., Figueroa, E., McShane, S. R., Cohen, R. D., Jin, J., Frye, S. V., Zamboni, W. C., Pepe-Ranney, C., Mani, S., Kelly, L., and Redinbo, M. R. (2015) Structure and Inhibition of Microbiome β-Glucuronidases Essential to the Alleviation of Cancer Drug Toxicity. *Chem. Biol. 22*, 1238–

1249.

(57) Wang, L. Z., Ramírez, J., Yeo, W., Chan, M. Y. M., Thuya, W. L., Lau, J. Y. A., Wan, S. C., Wong, A. L. A., Zee, Y. K., Lim, R., Lee, S. C., Ho, P. C., Lee, H. S., Chan, A., Ansher, S., Ratain, M. J., and Goh, B. C. (2013) Glucuronidation by UGT1A1 Is the Dominant Pathway of the Metabolic Disposition of Belinostat in Liver Cancer Patients. *PLoS One 8*.

(58) Anne, M., Sammartino, D., Barginear, M. F., and Budman, D. (2013) Profile of panobinostat and its potential for treatment in solid tumors: An update. *Onco. Targets. Ther.*

(59) Castellino, S., O'Mara, M., Koch, K., Borts, D. J., Bowers, G. D., and MacLauchlin, C. (2012) Human metabolism of lapatinib, a dual kinase inhibitor: Implications for hepatotoxicity. *Drug Metab. Dispos. 40*, 139–150.

(60) Regan, S. L., Maggs, J. L., Hammond, T. G., Lambert, C., and Williams, D. P. (2010) Acyl Glucuronides : The Good , The Bad and The Ugly *395*, 367–395.

(61) LoGuidice A, Wallace BD, Bendel L, Redinbo MR, and Boelsterli UA. (2012) Pharmacologic targeting of bacterial beta-glucuronidase alleviates nonsteroidal anti-inflammatory drug-induced enteropathy in mice. *J. Pharmacol. Exp. Ther. 341*, 447–454.

(62) Saitta, K. S., Zhang, C., Lee, K. K., Fujimoto, K., Redinbo, M. R., and Boelsterli, U. A. (2014) Bacterial β-glucuronidase inhibition protects mice against enteropathy induced by indomethacin, ketoprofen or diclofenac: Mode of action and pharmacokinetics. *Xenobiotica 44*, 28–35.

(63) Somasundaram, S., Rafi, S., Hayllar, J., Sigthorsson, G., Jacob, M., Price, a B., Macpherson, a, Mahmod, T., Scott, D., Wrigglesworth, J. M., and Bjarnason, I. (1997) Mitochondrial damage: a possible mechanism of the "topical" phase of NSAID induced injury to the rat intestine. *Gut 41*, 344–53.

(64) Boelsterli, U. A., Redinbo, M. R., and Saitta, K. S. (2013) Multiple NSAID-induced hits injure the small intestine: Underlying mechanisms and novel strategies. *Toxicol. Sci. 131*, 654–667.

(65) Femia, A. M., and Caderni, G. (2008) Rodent models of colon carcinogenesis for the study of chemopreventive activity of natural products. *Planta Med.*

(66) Megaraj, V., Ding, X., Fang, C., Kovalchuk, N., Zhu, Y., and Zhang, Q. Y. (2014) Role of hepatic and intestinal P450 enzymes in the metabolic activation of the colon carcinogen azoxymethane in mice. *Chem. Res. Toxicol. 27*, 656–662.

(67) Rats, G. S., Gert, L., and Matsumoto, H. (1981) Comparison of the Carcinogenicity of Methylazoxymethanol-beta-D-glucosiduronic Acid in Conventional and Germfree Sprague-Dawley Rats *67*, 1053–1055.

(68) Takada, H., Hirooka, T., Hiramatsu, Y., and Yamamoto, M. (1982) Effect of beta-glucuronidase inhibitor on azoxymethane-induced colonic carcinogenesis in rats. *Cancer Res. 42*, 331–334.

(69) Gu, D., and McNaughton, L. (2010) A Comprehensive Approach to the Profiling of

Cooked Meat Carcinogens 2-Amino-3,8-Dimethylimidazo[4,5-f]quinoxaline, 2-Amino-1-methyl-6-Phenylimidazo[4,5-b]pyridine, and Their Metabolites in Human Urine. *Chem. Res. Toxicol. 23*, 788–801.

(70) Zheng, Z., Fang, J. L., and Lazarus, P. (2002) Glucuronidation: An important mechanism for detoxification of benzo[a]pyrene metabolites in aerodigestive tract tissues. *Drug Metab. Dispos. 30*, 397–403.

(71) Schmitt, G., Aderjan, R., Keller, T., and Wu, M. (1995) Ethyl glucuronide: An unusual ethanol metabolite in humans. Synthesis, analytical data, and determination in serum and urine. *J. Anal. Toxicol. 19*, 91–94.

(72) Palmer, R. B. (2009) A review of the use of ethyl glucuronide as a marker for ethanol consumption in forensic and clinical medicine. *Semin. Diagn. Pathol. 26*, 18–27.

(73) Baranowski, S., Serr, A., Thierauf, A., Weinmann, W., Grosse Perdekamp, M., Wurst, F. M., and Halter, C. C. (2008) In vitro study of bacterial degradation of ethyl glucuronide and ethyl sulphate. *Int. J. Legal Med. 122*, 389–393.

(74) Morini, L., Politi, L., Groppi, A., Stramesi, C., and Polettini, A. (2006) Determination of ethyl glucuronide in hair samples by liquid chromatography/electrospray tandem mass spectrometry. *J. Mass Spectrom. 41*, 34–42.

(75) Chen, G., Giambrone, N. E., Dluzen, D. F., Muscat, J. E., Berg, A., Gallagher, C. J., and Lazarus, P. (2010) Glucuronidation genotypes and nicotine metabolic phenotypes: Importance of functional UGT2B10 and UGT2B17 polymorphisms. *Cancer Res. 70*, 7543–7552.

(76) Byrd, G. D., Chang, K. M., Greene, J. M., and DeBethizy, J. D. (1992) Evidence for urinary excretion of glucuronide conjugates of nicotine, cotinine, and trans-3'-hydroxycotinine in smokers. *Drug Metab. Dispos. 20*, 192–197.

(77) Montenegro, M. a., Boiero, M. L., Valle, L., and Borsarelli, C. D. (2012) Gum Arabic : More Than an Edible Emulsifier. *Prod. Appl. Biopolym. 17*, 220.

(78) Babiker, R., Merghani, T. H., Elmusharaf, K., Badi, R. M., Lang, F., and Saeed, A. M. (2012) Effects of Gum Arabic ingestion on body mass index and body fat percentage in healthy adult females: two-arm randomized, placebo controlled, double-blind trial. *Nutr. J. 11*, 111.

(79) Chakdar, H., Kumar, M., Pandiyan, K., Singh, A., Nanjappan, K., Kashyap, P. L., and Srivastava, A. K. (2016) Bacterial xylanases: biology to biotechnology. *3 Biotech 6*, 150.

(80) Goldin, B. R., and Gorbach, S. L. (1984) The effect of milk and lactobacillus feeding on human intestinal bacterial enzyme activity. *Am. J. Clin. Nutr. 39*, 756–761.

(81) Beaud, D. (2005) Genetic characterization of the beta-glucuronidase enzyme from a human intestinal bacterium, Ruminococcus gnavus. *Microbiology 151*, 2323–2330.

(82) Dabek, M., McCrae, S. I., Stevens, V. J., Duncan, S. H., and Louis, P. (2008) Distribution of β-glucosidase and β-glucuronidase activity and of β-glucuronidase gene gus in human colonic bacteria. *FEMS Microbiol. Ecol. 66*, 487–495.

(83) Gadelle, D., Raibaud, P., and Sacquet, E. (1985) β-Glucuronidase activities of intestinal bacteria determined both in vitro and in vivo in gnotobiotic rats. *Appl. Environ. Microbiol. 49*, 682–685.

(84) Gloux, K., Berteau, O., El oumami, H., Beguet, F., Leclerc, M., and Dore, J. (2011) A metagenomic β-glucuronidase uncovers a core adaptive function of the human intestinal microbiome. *Proc. Natl. Acad. Sci. 108*, 4539–4546.

(85) Kim, D. H., Hong, S. W., Kim, B. T., Bae, E. A., Park, H. Y., and Han, M. J. (2000) Biotransformation of glycyrrhizin by human intestinal bacteria and its relation to biological activities. *Arch Pharm Res 23*, 172–177.

(86) McBain, A. J., and Macfarlane, G. T. (1998) Ecological and physiological studies on large intestinal bacteria in relation to production of hydrolytic and reductive enzymes involved in formation of genotoxic metabolites. *J. Med. Microbiol. 47*, 407–416.

(87) Nakamura, J., Kubota, Y., Miyaoka, M., Saitoh, T., Mizuno, F., and Benno, Y. (2002) Comparison of four microbial enzymes in Clostridia and Bacteroides isolated from human feces. *Microbiol. Immunol. 46*, 487–90.

(88) Russell, W. M., and Klaenhammer, T. R. (2001) Identification and Cloning of gusA, encoding a new β-glucuronidase from Lactobacillus gasseri ADH. *Appl. Environ. Microbiol. 67*, 1253–1261.

(89) Kaoutari, A. El, Armougom, F., Gordon, J. I., Raoult, D., and Henrissat, B. (2013) The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol. 11*, 497–504.

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

(91) Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.*

(92) Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P. M., and Henrissat, B. (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem. J. 432*, 437–444.

(93) McNeil, N. I. (1984) The contribution of the large intestine to energy supplies in man. *Am. J. Clin. Nutr. 39*, 338–342.

(94) Maslowski, K. ., Vieira, a. ., Ng, a, Kranich, J., Sierro, F., Yu, D., Schilter, H. ., Rolph, M. ., Mackay, F., Artis, D., Xavier, R. ., Teixeira, M. ., and Mackay, C. (2009) Regulation of inflammatory responses by gut microbiota and chemoattractan receptor GPR43. *Nature 461*, 1282–1286.

(95) Pluznick, J. L., Protzko, R. J., Gevorgyan, H., Peterlin, Z., Sipos, A., and Han, J. (2013) Olfactory receptor responding to gut microbiota- derived signals plays a role in renin secretion and blood pressure regulation. *Proc. Natl. Acad. Sci. 110*, 4410–4415.

(96) Samuel, B. S., Shaito, A., Motoike, T., Rey, F. E., Backhed, F., Manchester, J. K., Hammer, R. E., Williams, S. C., Crowley, J., Yanagisawa, M., and Gordon, J. I. (2008)

Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. *Proc. Natl. Acad. Sci. 105*, 16767–16772.

(97) Salyers, A. A., Vercellotti, J. R., West, S. E. H., and Wilkins, T. D. (1977) Fermentation of mucin and plant polysaccharides by strains of Bacteroides from the human colon. *Appl. Environ. Microbiol. 33*, 319–322.

(98) Sonnenburg, J. L., Xu, J., Leip, D. D., Chen, C., Benjamin, P., Weatherford, J., Buhler, J. D., and Gordon, J. I. (2005) Glycan Foragin in Vivo by an Intestine-Adapted Bacterial Symbiont. *Science (80-.). 307*, 1955–1959.

(99) Martens, E. C., Koropatkin, N. M., Smith, T. J., and Gordon, J. I. (2009) Complex glycan catabolism by the human gut microbiota: The bacteroidetes sus-like paradigm. *J. Biol. Chem. 284*, 24673–24677.

(100) Pollet, R. M., D'Agostino, E. H., Walton, W. G., Xu, Y., Little, M. S., Biernat, K. A., Pellock, S. J., Patterson, L. M., Creekmore, B. C., Isenberg, H. N., Bahethi, R. R., Bhatt, A. P., Liu, J., Gharaibeh, R. Z., and Redinbo, M. R. (2017) An Atlas of β-Glucuronidases in the Human Intestinal Microbiome. *Structure 25*, 967–977.

(101) Zitomersky, N. L., Coyne, M. J., and Comstock, L. E. (2011) Longitudinal analysis of the prevalence, maintenance, and IgA response to species of the order Bacteroidales in the human gut. *Infect. Immun. 79*, 2012–2020.

(102) Wallace, B. D., Roberts, A. B., Pollet, R. M., Ingle, J. D., Biernat, K. A., Pellock, S. J., Venkatesh, M. K., Guthrie, L., O'Neal, S. K., Robinson, S. J., Dollinger, M., Figueroa, E., McShane, S. R., Cohen, R. D., Jin, J., Frye, S. V., Zamboni, W. C., Pepe-Ranney, C., Mani, S., Kelly, L., and Redinbo, M. R. (2015) Structure and Inhibition of Microbiome β-Glucuronidases Essential to the Alleviation of Cancer Drug Toxicity. *Chem. Biol. 22*, 1238–1249.

(103) Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*.

(104) Schallus, T., Jaeckh, C., Fehér, K., Palma, A. S., Liu, Y., Simpson, J. C., Mackeen, M., Stier, G., Gibson, T. J., Feizi, T., Pieler, T., and Muhle-goll, C. (2008) Malectin : A Novel Carbohydrate-binding Protein of the Endoplasmic Reticulum and a Candidate Player in the Early Steps of Protein N -Glycosylation *19*, 3404–3414.

(105) LoGuidice, A., Wallace, B. D., Bendel, L., Redinbo, M. R., and Boelsterli, U. a. (2012) Pharmacologic Targeting of Bacterial beta-Glucuronidase Alleviates Nonsteroidal Anti-Inflammatory Drug-Induced Enteropathy in Mice. *J. Pharmacol. Exp. Ther. 341*, 447–454.

(106) Mani, S., Boelsterli, U. a, and Redinbo, M. R. (2014) Understanding and modulating mammalian-microbial communication for improved human health. *Annu. Rev. Pharmacol. Toxicol. 54*, 559–80.

(107) Roberts, A. B., Wallace, B. D., Venkatesh, M. K., Mani, S., and Redinbo, M. R. (2013) Molecular insights into microbial β-glucuronidase inhibition to abrogate CPT-11 toxicity. *Mol. Pharmacol. 84*, 208–17.

(108) Wallace, B. D., and Redinbo, M. R. (2013) The human microbiome is a source of

therapeutic drug targets. *Curr. Opin. Chem. Biol. 17*, 379–384.

(109) Pellock, S. J., and Redinbo, M. R. (2017) Glucuronides in the gut: Sugar-driven symbioses between microbe and host. *J. Biol. Chem. 292*, 8569–8576.

(110) Andberg, M., Maaheimo, H., Boer, H., Penttilä, M., Koivula, A., and Richard, P. (2012) Characterization of a novel Agrobacterium tumefaciens Galactarolactone Cycloisomerase Enzyme for Direct Conversion of D -Galactarolactone to 3-Deoxy-2-keto-L-threo-hexarate. *J. Biol. Chem. 287*, 17662–17671.

(111) 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.

(112) Conchie, J., Hay, A. J., Strachan, I., and Levvy, G. A. (1967) Inhibition of glycosidases by aldonolactones of corresponding configuration: Preparation of (1→5)-lactones by catalytic oxidation of pyranoses and study of their inhibitory properties. *Biochem. J. 102*, 929–941.

(113) Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr. Sect. D Biol. Crystallogr. 58*, 1948–1954.

(114) Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr. 60*, 2126–2132.

(115) Sreerama, N., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem. 287*, 252–260.

(116) Park, H., Ovchinnikov, S., Kim, D. E., DiMaio, F., and Baker, D. (2018) Protein homology model refinement by large-scale energy optimization. *Proc. Natl. Acad. Sci. 115*, 201719115.

(117) Park, H., Kim, D. E., Ovchinnikov, S., Baker, D., and DiMaio, F. (2018) Automatic structure prediction of oligomeric assemblies using Robetta in CASP12. *Proteins Struct. Funct. Bioinforma. 86*, 283–291.

(118) Ovchinnikov, S., Park, H., Kim, D. E., DiMaio, F., and Baker, D. (2018) Protein structure prediction using Rosetta in CASP12. *Proteins Struct. Funct. Bioinforma. 86*, 113–121.

(119) Dong, R., Peng, Z., Zhang, Y., and Yang, J. (2018) mTM-align: an algorithm for fast and accurate multiple protein structure alignment. *Bioinformatics* 1–7.

(120) Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol. 215*, 403–10.

(121) Cantarel, B. I., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat,

B. (2009) The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res.*

(122) Bouvier, J. T., Sernova, N. V., Ghasempur, S., Rodionova, I. A., Vetting, M. W., Al-Obaidi, N. F., Almo, S. C., Gerlt, J. A., and Rodionov, D. A. (2018) Novel Metabolic Pathways and Regulons for Hexuronate Utilization in Proteobacteria. *J. Bacteriol. 201*, e00431-18.

(123) Wallace, B. D., Wang, H., Lane, K. T., Scott, J. E., Orans, J., Koo, J. S., Venkatesh, M., Jobin, C., Yeh, L. A., Mani, S., and Redinbo, M. R. (2010) Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science (80-.). 330*, 831–835.

(124) LoGuidice, A., Wallace, B. D., Bendel, L., Redinbo, M. R., and Boelsterli, U. A. (2012) Pharmacologic Targeting of Bacterial β-Glucuronidase Alleviates Nonsteroidal Anti-Inflammatory Drug-Induced Enteropathy in Mice. *J. Pharmacol. Exp. Ther. 341*, 447–454.

(125) Little, M. S., Ervin, S. M., Walton, W. G., Tripathy, A., Xu, Y., Liu, J., and Redinbo, M. R. (2018) Active site flexibility revealed in crystal structures of Parabacteroides merdae β-glucuronidase from the human gut microbiome. *Protein Sci. 27*, 2010–2022.

(126) Pellock, S. J., Walton, W. G., Biernat, K. A., Torres-Rivera, D., Creekmore, B. C., Xu, Y., Liu, J., Tripathy, A., Stewart, L. J., and Redinbo, M. R. (2018) Three structurally and functionally distinct β-glucuronidases from the human gut microbe Bacteroides uniformis. *J. Biol. Chem. 293*, 18559–18573.

(127) Kabsch, W. (2010) XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr. 66*, 125–132.

(128) Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr. 69*, 1204–1214.

(129) Gerlt, J. A., Bouvier, J. T., Davidson, D. B., Imker, H. J., Sadkhin, B., Slater, D. R., and Whalen, K. L. (2015) Enzyme function initiative-enzyme similarity tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta - Proteins Proteomics 1854*, 1019–1037.

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

(131) Pellock, S. J., Creekmore, B. C., Walton, W. G., Mehta, N., Biernat, K. A., Cesmat, A. P., Ariyarathna, Y., Dunn, Z. D., Li, B., Jin, J., James, L. I., and Redinbo, M. R. (2018) Gut Microbial β-Glucuronidase Inhibition via Catalytic Cycle Interception. *ACS Cent. Sci. 4*, 868–879.

(132) Biernat, K. A., Pellock, S. J., Bhatt, A. P., Bivins, M. M., Walton, W. G., Tran, B. N. T., Wei, L., Snider, M. C., Cesmat, A. P., Tripathy, A., Erie, D. A., and Redinbo, M. R. (2019) Structure, function, and inhibition of drug reactivating human gut microbial β-glucuronidases. *Sci. Rep. 9*, 825.

(133) 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.

(134) Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J. M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H. B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S. D., Wang, J., Antolin, M., Artiguenave, F., Blottiere, H., Borruel, N., Bruls, T., Casellas, F., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Forte, M., Friss, C., Van De Guchte, M., Guedon, E., Haimet, F., Jamet, A., Juste, C., Kaci, G., Kleerebezem, M., Knol, J., Kristensen, M., Layec, S., Le Roux, K., Leclerc, M., Maguin, E., Melo Minardi, R., Oozeer, R., Rescigno, M., Sanchez, N., Tims, S., Torrejon, T., Varela, E., De Vos, W., Winogradsky, Y., and Zoetendal, E. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*.

(135) den Besten, G., van Eunen, K., Groen, A. K., Venema, K., Reijngoud, D.-J., and Bakker, B. M. (2013) The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J. Lipid Res.*

(136) Pollet, R. M., D'Agostino, E. H., Walton, W. G., Xu, Y., Little, M. S., Biernat, K. A., Pellock, S. J., Patterson, L. M., Creekmore, B. C., Isenberg, H. N., Bahethi, R. R., Bhatt, A. P., Liu, J., Gharaibeh, R. Z., and Redinbo, M. R. (2017) An Atlas of β-Glucuronidases in the Human Intestinal Microbiome. *Structure 25*, 967-977.e5.

(137) Porter, N. T., and Martens, E. C. (2017) The Critical Roles of Polysaccharides in Gut Microbial Ecology and Physiology. *Annu. Rev. Microbiol.*

(138) Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.*

(139) Schmid, J., Sperl, N., and Sieber, V. (2014) A comparison of genes involved in sphingan biosynthesis brought up to date. *Appl. Microbiol. Biotechnol.*

(140) Pham, P. L., Dupont, I., Roy, D., Lapointe, G., and Cerning, J. (2000) Production of exopolysaccharide by Lactobacillus rhamnosus R and analysis of its enzymatic degradation during prolonged fermentation. *Appl. Environ. Microbiol.*

(141) HASHIMOTO, W., and MURATA, K. (1998) α- <scp>L</scp> -Rhamnosidase of *Sphingomonas* sp. R1 Producing an Unusual Exopolysaccharide of Sphingan. *Biosci. Biotechnol. Biochem.*

(142) Yuzwa, S. a, and Vocadlo, D. J. (2014) O-GlcNAc and neurodegeneration: biochemical mechanisms and potential roles in Alzheimer's disease and beyond. *Chem. Soc. Rev. 43*, 6839–6858.

(143) Craciun, S., and Balskus, E. P. (2012) Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc. Natl. Acad. Sci. U. S. A. 109*, 21307–12.

(144) Levvy, G. A. A. (1952) The Preparation and Properties of Beta-Glucuronidase. 4. Inhibition by sugar acids and their lactones. *J. Biol. Chem. 52*, 464–472.

(145) Rasmussen, T. S., Koldsø, H., Nakagawa, S., Kato, A., Schiøtt, B., and Jensen, H. H. (2011) Synthesis of uronic-Noeurostegine – a potent bacterial β-glucuronidase inhibitor. *Org. Biomol. Chem. 9*, 7807–7813.

(146) Morrison, J. F., and Walsh, C. T. (1988) The Behavior and Significance of Slow-Binding Enzyme Inhibitors, in *Advances in Enzymology and Related Areas of Molecular Biology*, pp 201–301. Academic Press, New York.

(147) Morrison, J. F., and Cleland, W. W. (1983) Lanthanide-Adenosine 5′-Triphosphate Complexes: Determination of Their Dissociation Constants and Mechanism of Action as Inhibitors of Yeast Hexokinase. *Biochemistry 22*, 5507–5513.

(148) Bull, H. G., Garcia-Calvo, M., Andersson, S., Baginsky, W. F., Chan, H. K., Ellsworth, D. E., Miller, R. R., Stearns, R. A., Bakshi, R. K., Rasmusson, G. H., Tolman, R. L., Myers, R. W., Kozarich, J. W., and Harris, G. S. (1996) Mechanism-Based Inhibition of Human Steroid 5R-Reductase by Finasteride: Enzyme-Catalyzed Formation of NADP-Dihydrofinasteride, a Potent Bisubstrate Analog Inhibitor. *J. Am. Chem. Soc. 118*, 2359–2365.

(149) Ahmad, S., Hughes, M. A., Yeh, L.-A., and Scott, J. E. (2012) Potential Repurposing of Known Drugs as Potent Bacterial -Glucuronidase Inhibitors. *J. Biomol. Screen. 17*, 957–965.

(150) Kodawara, T., Masuda, S., Yano, Y., Matsubara, K., Nakamura, T., and Masada, M. (2014) Inhibitory effect of ciprofloxacin on beta-glucuronidase-mediated deconjugation of mycophenolic acid glucuronide. *Biopharm. Drug Dispos. 35*, 275–283.

(151) Finn, R. S., Crown, J. P., Lang, I., Boer, K., Bondarenko, I. M., Kulyk, S. O., Ettl, J., Patel, R., Pinter, T., Schmidt, M., Shparyk, Y., Thummala, A. R., Voytko, N. L., Fowst, C., Huang, X., Kim, S. T., Randolph, S., and Slamon, D. J. (2015) The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): A randomised phase 2 study. *Lancet Oncol. 16*, 25–35.

(152) Shaw, A. T., Kim, D. W., Nakagawa, K., Seto, T., Crino, L., Ahn, M. J., De Pas, T., Besse, B., Solomon, B. J., Blackhall, F., Wu, Y. L., Thomas, M., O'Byrne, K. J., Moro-Sibilot, D., Camidge, D. R., Mok, T., Hirsh, V., Riely, G. J., Iyer, S., Tassell, V., Polli, A., Wilner, K. D., and Janne, P. A. (2013) Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med 368*, 2385–2394.

(153) Bang-Andersen, B., Ruhland, T., Jorgensen, M., Smith, G., Frederiksen, K., and Jensen, K. G. (2011) Discovery of 1-[2-(2,4- dimethylphenylsulfanyl)phenyl]piperazine (Lu AA21004): a mul- timodal compound for the treatment of major depressive disorder. *J. Med. Chem. 54*, 3206–21.

(154) Reisch, C. R., and Prather, K. L. J. (2017) Scarless Cas9 assisted recombineering (no-SCAR) in Escherichia coli, an easy-to-use system for genome editing. *Curr. Protoc. Mol. Biol. 2017*, 117:31.8.1–31.8.20.

(155) Tian, G., Mook, R., Moss, M. L., and Frye, S. V. (1995) Mechanism of Time-Dependent Inhibition of 5.alpha.-Reductases by .DELTA.1-4-Azasteroids: Toward Perfection of Rates of Time-Dependent Inhibition by Using Ligand-Binding Energies. *Biochemistry 34*, 13453–13459.

(156) Maier, L., Pruteanu, M., Kuhn, M., Zeller, G., Telzerow, A., Anderson, E. E., Brochado, A. R., Fernandez, K. C., Dose, H., Mori, H., Patil, K. R., Bork, P., and Typas, A. (2018) Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature 555*, 623–628.

(157) Domínguez, J. M., Fuertes, A., Orozco, L., Monte-Millán, M. Del, Delgado, E., and Medina, M. (2012) Evidence for irreversible inhibition of glycogen synthase kinase-3β by tideglusib. *J. Biol. Chem. 287*, 893–904.

(158) Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr. 66*, 213–221.

(159) Pellock, S. J., and Redinbo, M. R. (2017) Glucuronides in the gut: Sugar-driven symbioses between microbe and host. *J. Biol. Chem. 292*.

(160) Ahmad, S., Hughes, M. a, Lane, K. T., Redinbo, M. R., Yeh, L.-A., and Scott, J. E. (2011) A High Throughput Assay for Discovery of Bacterial β-Glucuronidase Inhibitors. *Curr. Chem. Genomics 5*, 13–20.

(161) Cheng, K. W., Tseng, C. H., Yang, C. N., Tzeng, C. C., Cheng, T. C., Leu, Y. L., Chuang, Y. C., Wang, J. Y., Lu, Y. C., Chen, Y. L., and Cheng, T. L. (2017) Specific Inhibition of Bacterial β-Glucuronidase by Pyrazolo[4,3-c]quinoline Derivatives via a pH-Dependent Manner to Suppress Chemotherapy-Induced Intestinal Toxicity. *J. Med. Chem.*

(162) Biernat, K. A., Pellock, S. J., Bhatt, A. P., Bivins, M. M., Walton, W. G., Tran, B. N. T., Wei, L., Snider, M. C., Cesmat, A. P., Tripathy, A., Erie, D. A., and Redinbo, M. R. (2019) Structure, function, and inhibition of drug reactivating human gut microbial β-glucuronidases. *Sci. Rep. 9*.

(163) McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Korner, M., Plesniak, L. a., Ziser, L., Wakarchuk, W. W., and Withers, S. G. (1996) The pK a of the General Acid / Base Carboxyl Group of a Glycosidase Cycles during Catalysis : A 13 C-NMR Study of Bacillus circulans Xylanase †. *Biochemistry 35*, 9958–9966.

(164) Little, M. S., Pellock, S. J., Walton, W. G., Tripathy, A., and Redinbo, M. R. (2017) Structural basis for the regulation of β-glucuronidase expression by human gut Enterobacteriaceae. *Proc. Natl. Acad. Sci.*

(165) Bhatt, A. P., Grillo, L., and Redinbo, M. R. (2019) In Fimo: A Term Proposed for Excrement Examined Experimentally. *Gastroenterology*.

(166) Wu, L., Jiang, J., Jin, Y., Kallemeijn, W. W., Kuo, C. L., Artola, M., Dai, W., Van Elk, C., Van Eijk, M., Van Der Marel, G. A., Codée, J. D. C., Florea, B. I., Aerts, J. M. F. G.,

Overkleeft, H. S., and Davies, G. J. (2017) Activity-based probes for functional interrogation of retaining β-glucuronidases. *Nat. Chem. Biol. 13*, 867–873.