ELUCIDATING THE INTERACTIONS BETWEEN GUT BACTERIAL β-GLUCURONIDASES AND APPROVED DRUGS

Parth B. Jariwala

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
2021

Approved by:

Matthew R. Redinbo

Gary J. Pielak

Bo Li

Leslie M. Hicks

Lindsey I. James

ABSTRACT

Parth B. Jariwala: Elucidating the interactions between gut bacterial β-glucuronidases and approved drugs
(Under the direction of Matthew R. Redinbo)

The human gut microbiome contains a plethora of enzymes that metabolize a myriad of diet-derived, host-derived, and therapeutic agents. Metabolism of therapeutic agents by gut microbiota can result in altered drug efficacy and toxicity. An important bacterial enzyme involved in the metabolism of drugs in the gastrointestinal tract is the gut bacterial β-glucuronidase (GUS). One notable example of drug metabolism by GUS enzymes is the reactivation of the active metabolite of the chemotherapeutic irinotecan, SN-38, which causes severe, dose limiting toxicity in patients. Recent analysis of the Human Microbiome Project revealed 279 unique GUS isoforms encoded by gut microbiota. Elucidating the exact GUS isoform reactivating drugs like SN-38 can lead to precise dosing of drugs on an individual basis and reduce gut microbiota-mediated toxic side effects. In this dissertation, a novel activity-based proteomics strategy is used to identify the exact GUS isoform responsible for the reactivation of SN-38. In addition to the metabolism of drugs by GUS enzymes, another important interaction is the inhibition of GUS enzymes by FDA-approved drugs. Inhibition of GUS activity can result in altered homeostasis because GUS enzymes also reactivate host-derived endobiotics and process diet-derived glucuronic acid-containing polysaccharides. We present data showing that approved drugs with a particular chemical scaffold can inhibit the activity of GUS enzymes. Altering the abundance or activity of GUS enzymes can potentially reduce GUS-mediated metabolism of therapeutics. We next present data on how dietary fiber influences the abundance of GUS-

encoding gut bacterial species. These data provide information on how diet can potentially be used to modulate GUS activity in the gut. In addition to diet, another approach to reduce the metabolism of drugs by GUS and other gut microbial enzymes is using small molecule adjuvants. The final chapter outlines a chemoproteomics platform that could be used to discover selective small molecule inhibitors that reduce the metabolism of any drug of interest. Together, the work presented here expands on our knowledge of GUS-drug interactions, discusses methods to alter GUS activity, and outlines a chemoproteomics platform to identify small molecule inhibitors of gut enzymes.

To my family and friends

ACKNOWLEDGEMENTS

To my dear parents, thank you for just about everything. I whole-heartedly appreciate your endless support of my dreams and endeavors. You both are powerful role models of mine and I can only hope to achieve as much as you both have. To my dear brother, thank you for being chill and your unwavering support. Your energy and positive outlook continuously amaze me. You too are a great role model of mine. To my aunt and uncle, thank you for taking me in and raising me in India. You both cared for me as if I were your own son. I will never forget that.

To my dear Redinbo Lab mates, thank you for being amazing work colleagues these past few years. Sam Pellock, thank you for coaching me at the bench and more importantly, teaching me how to prioritize what is important in life. Bill Walton, thank you for your kindness, willingness to help, and upbeat attitude. Marissa Bivins, thank you for your kindness and help. Josh Simpson, your strength to overcome challenges in life always amazes me and thank you for talking science out with me. Morgan Gibbs, thank you for being brutally honest and your help. Amanda Grabowski, thank you for your positive energy and upbeat attitude. Kacey Davis, thank you for reminding me why I do science: to be innovative and rigorous, to have fun, and to build a community of colleagues. Finally, to all the undergraduates, thank you for your support.

To my colleagues, friends, and mentors at UNC Chapel Hill, thank you for your endless support. Sam and Melissa Olson, thank you for being great friends and having me over for dinners and movies. You both got me through some rough patches. Chase and Brittney Weidmann, you both have been great mentors of mine and I truly appreciate your advice, support, and friendship. To my friend Nolan, thank you for being you. You have shown me what

true strength and courage looks like. Thanks for your support. To Joey and Hannah Thole, thanks for having me over and being great supportive friends. And thank you to all my friends in GSB. I will miss all the shenanigans, attitude adjustments, and laughs.

To my friends and mentors outside of Chapel Hill, thank you. There are too many to name, but you all have given me great degrees of support and I truly appreciate it.

Lastly, thank you Matt Redinbo. You took me into your group at my lowest point and gave me the space I needed to grow as a scientist. You have taught me how to be a compassionate leader and I will try my best to carry that forward.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

2-NP-Glc	2-nitrophenyl-β-D-glucopyranoside
4-MU-G	4-methylumbelliferone-β-D-glucuronide
Å	angstrom
ABP	activity-based probes
ABPP	activity-based protein profiling
Abx/PEG	antibiotics and polyethylene glycol
BLASTP	protein basic alignment search tool
*Bd*GUS	*B. dorei* β-glucuronidase
*Bf*GUS	*B. fragilis* β-glucuronidase
*Bo*GUS	*B. ovatus* β-glucuronidase
*Bt*GAL	*B. thetaiotamicron* β-galactosidase
*Bu*GUS-1	*B. uniformis* β-glucuronidase-1
*Bu*GUS-2	*B. uniformis* β-glucuronidase-2
*Bu*GUS-3	*B. uniformis* β-glucuronidase-3
°C	celsius
cgr	cardiac glycoside reductase
*Cp*GUS	*C. perfringens* β-glucuronidase
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
*Ec*GAL	*E. coli* β-galactosidase
*Ec*GUS	*E. coli* β-glucuronidase
*Ec*GUS$^{H.D.}$	heat-denatured *E. coli* β-glucuronidase

*Ec*GUS$^{M.A.}$	mono-associated *E. coli* β-glucuronidase
*Ec*GUS$^{M.A. + H.K.}$	mono-associated and heat-killed *E. coli* β-glucuronidase
*Ec*GUS^{Recomb}	recombinant *E. coli* β-glucuronidase
*Ee*GUS	*E. eligens* β-glucuronidase
*Et*GalAse	*E. tayi* galacturonidase
FPLC	fast protein liquid chromatography
FBDD	fragment-based drug discovery
FMT	fecal material transfer
FDA	food and drug administration
FARMM	food and resulting microbial metabolites
FMN	flavin mononucleotide
GalA	galacturonic acid
GAL	β-galactosidase
GI	gastrointestinal
GlcA	glucuronic acid
GUS	β-glucuronidase
HR-MS	high-resolution mass spectrometers
HTS	high-throughput screening
HMP	human microbiome project
IBD	inflammatory bowel disease
IGC	integrated gene catalog
k_{cat}	catalytic turnover
K_M	Michalis constant

LC-MS	liquid chromatography mass spectrometry
L1	loop 1
L2	loop 2
MS	mass spectrometry
MAC	microbiota accessible carbohydrates
mL1	mini-loop 1
mL2	mini-loop 2
mL1, mL2	mini-loop 1, mini-loop 2
mM	millimolar
MSA	multiple sequence alignment
nM	nanomolar
NGS	next-generation sequencing
NL	no loop
NSAID	nonsteroidal anti-inflammatory drug
NTL	N-terminal loop
NxK	asparagine-x-lysine
*p*NP-GlcA	*p*-nitrophenyl-β-D-glucuronide
PCR	polymerase chain reaction
PDB	protein data bank
PPI	proton-pump inhibitor
PLP	pyridoxal-5'-phosphate
RMSD	root mean square deviation
RNA-seq	RNA sequencing

SCFA	short-chain fatty acids
SNP	single nucleotide polymorphism
SN-38	7-ethyl-10-hydroxyl-camptothecin
SN-38-G	7-ethyl-10-hydroxyl-camptothecin glucuronide
TyrDC	tyrosine decarboxylase
*tyr*DC	tyrosine decarboxylase gene
UGTs	uridine-diphosphate glucuronosyltransferases
μL	microliter
μM	micromolar

CHAPTER 1: PHARMACOMICROBIOMICS

INTRODUCTION

In the Indian epic The *Mahabharata*, the guru Dronacharya leads a group of his students in archery practice. The objective of the lesson is to locate and precisely hit the eye of a wooden bird hidden in a nearby tree. Prior to allowing them to launch their arrows, Dronacharya asks each of them a very simple question: "What do you see?" Many of the students said a tree, some said leaves, a few said branches, and a few of them were simply confused by the question. The last student he asked was the warrior Prince Arjuna who replied, "I only see the eye of the bird." Highly pleased with Arjuna's answer, Dronacharya explains to his pupils the importance of solely focusing on a given target and minimizing any distractions. Arjuna goes on to become one of the most skilled archers in the epic, one whose arrows seldom missed their targets.

The perfect drug is like Arjuna's arrow: when released into the human body, it specifically hits the intended target and is efficiently excreted from the body. However, such a therapeutic arrow is still an epic fiction, as evidenced by high failure rates in clinical trials. In a survey done by the data analytics company Centres for Medicines Research International, nearly 90% of therapeutics were found to fail during phase I clinical trials, 80% failed in phase II, and 40% failed in phase III (*1*). Clinical trials may be terminated because of low patient recruitment or lack of funding, but the two main reasons are low efficacy and unknown toxic side effects (*2*, *3*). Most drugs fail phase I because of safety concerns, like unexpected side effects; meanwhile, in phase II, 50% of candidates fail due to lack of efficacy and another 30% fail because of unknown toxicities (*2*). These unanticipated biological outcomes that lead to reduced efficacy

and toxicity are largely due to poor pre-clinical models that do not accurately represent the complexities of drug interactions and metabolism in the body (*4*).

Current pre-clinical models examine the probability of a therapeutic's success using a combination of *in vitro* and *in vivo* models. The goal is to characterize drug efficacy, pharmacodynamics (*i.e.*, the physiological impact of the drug on the host), pharmacokinetics (*i.e.*, the absorption, distribution, metabolism, and excretion of the drug), toxicity, and dosing requirements prior to human testing (*5*). Recent advances, like the use of three-dimensional cell cultures (spheroids), help to better recapitulate human physiology and cellular biology (*6*). Spheroids, both *in vitro* and those implanted in animal models, enable predictions of drug efficacy and toxicity that better reflect those that will be seen in humans (*6*). Despite these innovations, even our most advanced pre-clinical models lack the biochemical complexity of the human body and thus do not fully predict how a drug will interact with the human body.

Although many factors contribute to the inaccuracy of pre-clinical models, an underappreciated component of drug-body interactions is the effect of the host gut microbiota on drug efficacy and toxicity. Orally and parenterally taken drugs and associated drug metabolite(s) can reach the gastrointestinal (GI) tract in appreciable quantities directly after oral administration or indirectly via biliary excretion from the liver (*7*). For example, the Food and Drug Administration (FDA) reports that 92% of the tyrosine kinase inhibitor ceritinib and 69% of the cyclin-dependent kinase inhibitor ribociclib were recovered in feces as the unmetabolized parent compound (*8*). When in the GI tract, drugs like ceritinib and ribociclib are exposed to the most diverse reservoir of biological factors found in the host: those generated by the intestinal microbiome (*9*). An increasing body of evidence shows that an appreciable number of drugs can be metabolized by the human gut microbiota (**Figure 1.1**) (*10*). These microbiota-derived

metabolites have the potential to re-enter systemic circulation via enterohepatic circulation (*11*). Indeed, the biotransformation of small molecule agents catalyzed by the gut can result in altered drug bioavailability, treatment efficacy, and drug toxicity (*12*).

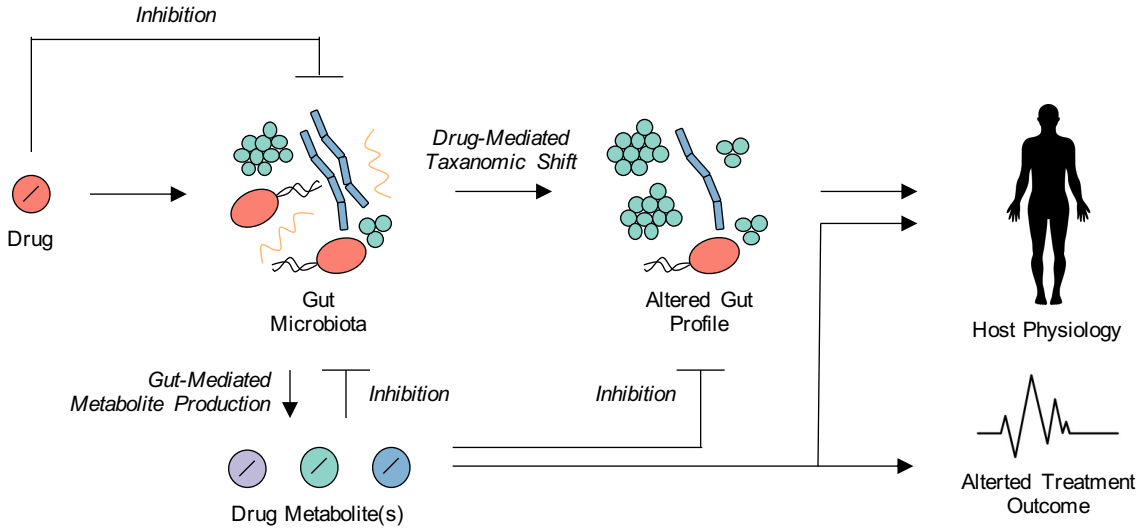


Figure 1.1 Pharmacomicrobiomics. An overview of the various ways drugs and gut flora interact with one another.

The human gut microbiota is home to approximately 1,000,000,000,000 bacterial cells, which is equivalent to the number of human cells in the body (*13*). To date, the genomes of more than 1,000 different gut bacterial isolates have been sequenced and mapped; cumulatively, the gut microbiota is estimated to contain 100 times more genes than the human genome (*13*). This consortium of gut bacteria has the potential to modify a drug through a myriad of enzymatic chemistries (*e.g.*, hydrolysis, decarboxylation, and reduction) (*12*). Biotransformation of drugs in the GI tract not only impact drug pharmacology and efficacy but can also cause dysbiosis in gut bacterial composition through unwarranted metabolite-bacteria interactions (**Figure 1.1**). Many gut bacterial species are beneficial for the host as they produce chemical agents that are important for normal functioning like analgesics, vitamins, antioxidants, anti-inflammatory agents, and short-chain fatty acids (SCFAs) (*e.g.*, acetate, propionate, and butyrate) (*14*). Drug-

mediated dysbiosis and reduction in species-level diversity via metabolite-bacteria interactions can disrupt production and regulation of these beneficial biochemicals, which can lead to pathogenesis of GI disorders such as inflammatory bowel disease and irritable bowel syndrome, as well as extra-intestinal disorders like obesity and cardiovascular disease (*15*).

In addition to the richness of bacterial species present in the gut, there is great variation in gut bacterial composition from individual to individual (*16*). This immense disparity is due to host genomics as well as environmental factors, which include diet, exercise, the host immune system, and even geographic location (*17*, *18*). Remarkably, even monozygotic twins can develop different gut bacterial profiles, which highlights the integral role of the environment in regulating gut bacterial composition (*19*). For drugs that are metabolized by the gut microbiota and have narrow therapeutic indexes, the mechanism and extent to which these compounds are metabolized can be crucial for accurately dosing individuals based upon distinct gut bacterial composition. The emerging field of pharmacomicrobiomics studies the metabolism of drugs by the human gut microbiota (**Figure 1.1**). By combining data on pharmacology and toxicology with pharmacomicrobiomics of a drug, the hunt for new and efficacious drugs is one step closer to the precision of Arjuna's arrows.

In this chapter, the following key questions in the field of pharmacomicrobiomics for assessing drug-gut microbiota interactions will be discussed:

1. *Does the drug and/or its metabolite(s) shift gut bacterial composition?*
2. *What metabolite(s) of the drug are produced by gut bacterial pathways?*
3. *What are the bacterial gene product(s) conducting the biotransformation of the drug?*
4. *Which gut bacterial strains are efficient at metabolizing the drug?*
5. *Does the drug inhibit gut bacterial enzyme activity?*

Current technologies and associated resources to answer these questions will be covered briefly in the next section, followed by a discussion on recent studies that address these questions. Next, an outlook of the pharmacomicrobiomics field will be addressed. Finally, a brief outline of this dissertation will be presented.

CURRENT TECHNOLOGIES AND DATABASES

The key questions of pharmacomicrobiomics can be answered by combining the powerful technologies in next-generation sequencing (NGS), mass spectrometry (MS), and computational biology with biochemical, biophysical, and animal studies. An overview of the types and level of information that can be obtained by nucleic acid NGS sequencing and MS-based experiments are presented below.

16S rRNA Sequencing

NGS technologies offer a culture-independent means to profile the composition of bacterial communities present in an individual's gut using relatively low volumes of fecal material (*14*). Of NGS-based technologies, 16S rRNA sequencing has been the primary method for cataloguing the taxonomy of the fecal microbiota because it is inexpensive, quick, and can be easily outsourced (*14*). To obtain a taxonomic profile using 16S, universal primers are designed to target the conserved base pair regions in the 16S ribosomal gene of gut microbes *en masse*. The hypervariable regions of the resulting amplified 16S genes are analyzed using bioinformatic pipelines like QIIME2 (*20*) and DADA2 (*21*) to discern the taxonomic profile and relative abundances of bacteria present in the analyzed fecal sample. Although conducting 16S rRNA sequencing is extremely facile, it lacks quantitative functional characterization of microbial

communities present in a sample. 16S rRNA sequencing datasets are often capable of discerning the genera of bacteria present, yet it often lacks the phylogenetic power to distinguish bacterial species, let alone strains (*22*). Hence, for pharmacomicrobiomic studies, 16S rRNA sequencing is ideal when analyzing perturbations at the genera level or higher in gut microbial composition in the presence of a drug of interest.

Metagenome Sequencing

Significant reduction in the cost to conduct NGS sequencing ($5,000 per megabyte in 2001 compared to $0.05 per megabyte in 2017) has made shotgun metagenome sequencing a commonplace technique for obtaining profiles of gut bacterial communities using fecal samples. For shotgun metagenome sequencing of fecal microbiotas, DNA is extracted from the sample, sheared using ultrasonication, and sequenced using a massively parallel sequencing strategy. The resulting gigabytes worth of sequencing reads are then aligned either to pre-annotated gut bacterial genomes for gene cataloguing using streamlined bioinformatic packages like MetaPhlAn 3.0 (*23*) or into contigs using *de novo* genome assembly tools like SPAdes (*24*) and Velvet (*25*) to identify and quantify sets of genes that may be sample-specific. Unlike 16S rRNA sequencing, shotgun metagenome sequencing can attain strain-level resolution of bacterial communities present in the analyzed sample (*14*). Additionally, advanced bioinformatics tools like HUMAnN 3.0 (*26*) can use shotgun metagenome sequencing data to profile the presence, absence, and abundance of gut microbial genes involved in metabolic pathways.

The low-cost, high-throughput nature of metagenome sequencing and the availability of streamlined bioinformatic packages for analysis have led to the collection and deposition of terabytes worth of fecal metagenome datasets with associated patient metadata like sex, age, and

chronic disorders to online repositories for public use by multi-institutional consortia like the Human Microbiome Project (HMP) and the European MetaHit Project (**Table 1.1**) (*27*). In pharmacomicrobiomic studies, these extant databases can serve as powerful databanks to delineate patient sub-populations that may be prone to gut microbiota-mediated drug metabolism based on genetic and environmental factors. For example, if the downstream product of a gut bacterial gene has been mechanistically associated to the biotransformation of a drug of interest, then that gene can be used as a molecular biomarker to scan extant metagenomic databases to unveil patient sub-populations that may be capable of metabolizing that drug or how widespread the presence of the gene is in global communities (*28*). Although metagenome sequencing is an indispensable tool for profiling the differences in the composition of bacterial communities and metabolic pathways in individuals with various phenotypes (*e.g.,* diseased, healthy, etc.), it cannot be used to evaluate functional differences since the gene expression levels cannot be ascertained.

Table 1.1 Extant gut microbiome metagenomic sequencing databases.

Source	Descriptor	URL	Ref.
HMP	Healthy	https://portal.hmpdacc.org	(*27*)
HMP	Pregnancy and Pre-term Birth	http://vmc.vcu.edu/momspi	(*27*)
HMP	Type II Diabetes	https://www.ibdmdb.org	(*27*)
HMP	Inflammatory Bowel Disease	http://med.stanford.edu/ipop.html	(*27*)
IGC	Healthy	https://db.cngb.org/microbiome/	(*29*)
CGR	Healthy	-	(*30*)
UHGP	Healthy	https://www.ebi.ac.uk/metagenomics/	(*31*)
Ceylan *et al*.	Diet	-	(*32*)

RNA Sequencing

Advances in RNA sequencing (RNA-seq) have enabled profiling of genes that are actively expressed in gut bacterial communities present in a fecal sample (*33*). Such functional profiling can highlight genes that may be driving factors in phenotypes of interest. For RNA-seq

experiments using fecal microbiotas, global RNA is extracted and rRNA is removed (*33*). The purified mRNA is transcribed into cDNA using reverse transcriptase, which is subsequently sequenced using a massively parallel sequencing strategy as in metagenome sequencing. Bioinformatics pipelines like HUMAnN 3.0 (*26*) and MG-RAST (*34*) can align the resulting sequence reads to reference genomes to obtain gene expression profiles. Although the data obtained from fecal RNA-seq analysis can be extremely powerful, a few major challenges include the short half-lives of bacterial mRNA and the need for high-quality RNA. Thus, RNA degradation and instability make it difficult to effectively capture gene expression for certain phenotypes (*33*).

Proteomics

Although information on gene expression obtained from RNA-seq gives some indication of function, it is not entirely accurate because the presence of mRNA is not always synonymous with the presence of protein or protein activity (*35*). Non-targeted shotgun proteomics can be used to evaluate levels of expressed protein in microbial communities present in fecal microbiota samples. Proteins are first extracted from the fecal matrix and digested using proteolytic enzymes like trypsin. The peptide mixture is analyzed using high-resolution mass spectrometers (HR-MSs) like FT-Orbitraps that can distinguish masses with an accuracy of 0.001 Da (*36*), making it possible to discern the spectra of tens of thousands of digested peptides within a sample (*36*). The resulting peptide spectra dataset are then compared to extant fecal protein sequence databases like the Integrated Gene Catalog (IGC) (*29*), which contains nearly 10 million, non-redundant protein sequences, using software like MetaLab (*37*) and MaxQuant (*38*) to identify and quantify the proteins present in the sample. Yet, a key limitation in non-targeted shotgun

proteomics has been the identification and quantification of low-abundance proteins (*39*). This problem has been circumvented by enriching the target protein(s) of interest using either affinity-based or irreversible small-molecule probes prior to proteolytic cleavage and MS analysis (*39–41*). Continued development in the field of single-cell proteomics will lead to further sensitivity to identify low-abundance proteins from fecal microbiota samples (*42*).

Although proteomic analyses of fecal microbiota are starting to be accomplished, comprehensive repositories of metaproteomic datasets with associated metadata like those in **Table 1.1** do not exist for human fecal proteomes. Academic research groups that have HR-MS, computational expertise in metaproteomic analysis, and a strong understanding of the human gut microbiota remain few, making it difficult to perform large-scale studies. By creating a robust and universal method to acquire fecal proteomic data and using a collaborative approach as was used to form the HMP and MetaHit, a central repository of proteomics data can be assembled with associated metadata. Protein(s) associated with the metabolism of a drug of interest can be used as molecular biomarkers to scan extant proteomic datasets and identify patient sub-populations that are capable of biotransforming a given drug. Such a proteomics approach is a more accurate means of identifying patient sub-populations compared to a metagenomics approach as the presence of a gene is not indicative of expressed protein. Additionally, if proteomic datasets can be collected from individuals who have undergone treatment with a drug of interest, it may be possible to correlate phenotype (*i.e.,* toxicity, reduced efficacy, etc.) to protein abundance data to *potentially* either track down proteins that are associated with the metabolism of the drug, or to validate a protein that is involved in the metabolism of the drug.

Metabolomics

For pharmacomicrobiomics, the advent of HR-MSs has enabled the accurate determination of biotransformed metabolite(s) of drugs using either *ex vivo* fecal mixtures or gut bacterial isolates (*10*, *28*). More importantly, current metabolomic strategies can be used to identify and quantify levels of gut-derived metabolites present in fecal microbiota samples. For example, a recent study examined fecal samples from 786 individuals and identified over 1,000 metabolites (*43*). Additionally, biotransformation reactions can be ascertained by calculating mass differences between parent compounds and associated metabolites (*10*, *28*). The human gut microbiota produces a myriad of chemicals that are undoubtably important for human health as the flux from basal levels has been associated with a variety of disorders like obesity and diabetes. Additionally, disruptions in metabolite homeostasis in the gut effects blood serum concentrations, which can have potential systemic health consequences (*44*). For pharmacomicrobiomic studies, untargeted metabolomics can be adapted to assess how shifts in gut bacterial communities because of the presence of a drug of interest can lead to changes in the gut metabolome. Ultimately, certain GI side effects of drugs can potentially be explained by changes in levels of certain metabolites by the gut microbiota.

RECENT ADVANCES IN PHARMACOMICROBIOMICS

An overview of recent works that use a combination of these technologies to answer the key questions in pharmacomicrobiomics is presented below:

Does the drug and/or its metabolite(s) shift gut bacterial composition?

Humans and gut microbes have co-evolved to increase the overall survival fitness of one another. Although humans provide a slurry of energy-dense nutrients for microbes, the human gut microbiome produces a wide array of bioactive compounds like SCFAs (*45*), vitamins (*46*), and essential amino acids that are important for normal bodily functions (*46*). Flux of these gut-derived beneficial compounds from basal levels can alter host physiology, often leading to either the onset of disease or aiding in disease progression (*15*, *45*). For example, in patients with inflammatory bowel disease (IBD), disease-induced dysbiosis of gut bacterial communities often leads to a decrease in *Firmicutes* species, which are major producers of the SCFA butyrate (*45*). Butyrate serves as the main energy source for colonocytes, maintains intestinal homeostasis through anti-inflammatory actions, and has immunomodulatory roles (*45*). Reduced butyrate production disrupts energy supply to colonocytes and increases mucosal inflammation, which exacerbates IBD symptoms (*45*).

Basal levels of beneficial gut-derived compounds can also be altered by drug-induced shifts in gut bacterial communities (*12*). For example, broad-spectrum antibiotics can affect the abundances of nearly 30% of known gut bacteria and cause a rapid and significant drop in species richness and diversity (*47*). Such a change in community composition promotes the overgrowth of pathogenic species like *Klebsiella pneumoniae*, *Staphylococcus aureus*, and *Clostridium difficile*, which can result in the pathogenesis of disease states like pseudomembranous colitis and decreased abundance of commensal bacterial that produce health-positive compounds like SCFAs (*47*). In addition to antibiotics, non-antibiotic agents can also influence the composition of the gut microbiota. Indeed, a population level analysis of nearly 4,000 individuals found that medication use correlated strongly to specific taxonomic profiles

(*16*). Additionally, studies that utilized 16S rRNA and metagenomic sequencing to profile bacterial communities in stool samples collected from individuals taking commonly prescribed medications, like the type 2 diabetes medication metformin, the anti-psychotic olanzapine, the statin rosuvastatin, the opioid morphine, non-steroidal anti-inflammatory drugs, and proton-pump inhibitors (PPIs), showed significant alteration in taxonomic profiles. A comprehensive review that summarizes the modulation in specific gut bacterial taxa by these drugs can be found (*48*).

Many non-antibiotic drugs modify the composition of gut bacterial communities by altering the gut environment (*e.g.*, pH, blood flow, and composition of the bile acid pool) and/or the host immune system (*e.g.*, inflammatory markers and immune regulation) (*48*). For example, drugs like metformin and morphine alter the composition of the bile acid pool by inhibiting the actions of bile acid re-uptake transporters in gut epithelial cells, as well as the conversion of cholesterol to primary bile acids in the liver (*49*, *50*). Because bile salts are metabolized by select bacteria and inhibit growth of others, alteration of the bile salt pool can lead to shifts in gut bacterial communities (*51*). Additionally, PPIs reduce acidity throughout the GI tract; the reduction in acidity enables oral microbes to colonize the intestine, which can lead to dramatic shifts in gut bacterial communities (*48*).

Non-antibiotic drugs also impact bacterial growth through direct interactions with bacteria. Typas and co-workers systematically assessed the growth inhibitory impact of 1197 off-patent FDA drugs against 40 gut bacterial isolates (*52*). The 40 tested gut bacterial isolates captured 78% of the median assignable relative abundance of the human gut microbiota at the genus level. Of the non-antibiotic drugs tested, 27% were found to inhibit growth for at least one gut bacterial isolate that was tested. Although Typas and co-workers provide an impressive map

of drug agents that have growth inhibitory activities, a key limitation of this study was that it did not assess growth of species in a community setting. Microbe-microbe interactions can heavily influence gut bacterial growth (*53*); hence, future studies incorporating high-throughput screening strategies that use 16S rRNA sequencing or metagenomic sequencing to assess growth in community settings (be it artificial consortia or fecal samples) should be done. Additionally, the growth inhibitory activities of both host-derived and gut-derived metabolite(s) should be assessed.

What metabolite(s) of the drug are produced by gut bacterial pathways?

Drugs and their metabolite(s) can reach the intestine either directly after ingestion or following first-pass metabolism via biliary excretion. The human body has evolved to excrete small molecule agents using a combination of mechanical and biochemical mechanisms. Orally taken drugs must first successfully pass through the gut epithelial barrier to reach the hepatic portal system (*5*). After arriving in the liver, the small molecule agents face their biggest challenge: phase I and II metabolism in hepatocyte cells (*i.e.*, first-pass metabolism). During phase I metabolism, several low-specificity enzyme systems (*e.g.*, cytochrome P450 monooxygenase system and flavin-containing monooxygenase system) introduce chemical modifications like the addition of epoxide, hydroxyl, amino, or thiol groups to increase the overall polarity of molecules (*5*). Typically, modifications in phase I metabolism generate either electrophilic sites or nucleophilic sites on a compound, which then serve as sites for conjugation to hydrophilic endogenous substrates like glucuronic acid (GlcA), sulfate, glycine, acetyl, and glutathione during phase II metabolism (*5*). The increased charge and polarity of the drug

reduces cellular uptake, thus making it easier for the body to purge the drug via urinary or intestinal excretion.

In the GI tract, the large consortia of gut bacterial enzymes can introduce further modifications to drugs (*48*). For example, gut bacterial hydrolases like glycoside hydrolases and sulfatases can hydrolyze phase II-derived glucuronyl and sulfonyl groups from drug compounds (*12*). Bacterial lyases like C-S β-lyase can cleave C-S bonds in phase II-derived cysteine-S-conjugates of drugs via a β-elimination reaction, which results in the production of ammonia, pyruvate, and a thiol-containing analog of the deconjugated drug (*12*). Bacterial reductases can reduce a wide variety of functional groups on drug scaffolds that include alkene, nitro, *N*-oxide, azo, and sulfoxide groups (*12*). Lastly, bacterial transferase enzymes can add or remove functional groups (*e.g.*, methyl and acyl groups) from the chemical scaffold of a drug or its metabolite (*12*). The products of these various reactions can be further metabolized by gut microbial enzymes. Due to the inherent technical challenge of screening drugs against thousands of cultured isolates, the full extent to which drugs and their metabolite(s) are produced is largely unknown.

In an effort to catalogue the various metabolites that are produced by gut microbes, Goodman and co-workers systematically measured the ability of 76 gut bacterial isolates to biotransform 271 orally administered drugs that spanned the chemical drug space (*i.e.*, diverse clinical indications, physiochemical properties, and predicted intestinal concentrations) (*10*). A previously established combinatorial pooling strategy was used to reduce the 20,596 bacteria-drug pairs to 3,840 samples for liquid chromatography-coupled MS (LC-MS) analysis (*54*). Nearly two-thirds of the 271 tested drugs were reported as metabolized by at least one gut bacterial isolate, with each strain metabolizing between 11-95 drugs. Using mass differences

between parent compounds and observed drug-specific metabolites, the investigators found that certain chemical substructures like lactone rings and nitro, azo, and urea groups in a drug scaffold were susceptible to modification by gut bacteria, which corroborates previous findings. Interestingly, functional groups (*e.g.,* esters and amides) that were previously reported to be targets of gut bacterial enzymes were found in drugs that were not metabolized, indicating that the overall scaffold of the drug plays a key role in microbial drug metabolism. Of the drug-microbe positive hits, the corticosteroid dexamethasone was validated in a mouse model to be uniquely metabolized (via desmolytic cleavage) by *Clostridium scindens* to form the androgen version of the drug.

As strain-to-strain variation in the metabolism of drugs has been extensively observed, a strategy to directly screen microbes found in patient stool samples is critical (*55*, *56*). Donia and co-workers conducted a high-throughput screening campaign using patient-derived stool sample to determine the metabolite produced (*28*). Prior to screening, the team developed an *ex vivo* mixed-culturing system that enabled close capture *ex vivo* (70%; when compared with the original patient-derived sample) based on 16S rRNA sequencing. Using untargeted metabolomics, the team tested 575 orally administered drugs in the *ex vivo* patient-derived stool culture and found that 57 compounds were metabolized based on either diminished concentration of parent compound or presence of new metabolite. Akin to the study done by Goodman and co-workers (*10*), the team observed chemistries like nitro-reduction, hydrolysis, and azo-reduction. The findings were validated *in vivo* using a mouse colonized with the patient-derived sample after antibiotic treatment. The chemotherapeutic capecitabine was found to be deglycosylated over time to form deglycocapecitabine in the mouse model. As the investigators note, 100% of the species were not captured in the *ex vivo* mixed-culturing system, which highlights the need to

develop better culturing systems to recapitulate bacterial colonies present in patient-derived fecal samples.

What are the bacterial gene product(s) conducting the biotransformation of the drug?

Knowing the exact gene product(s) conducting the biotransformation of a drug is necessary for the appropriate development of diagnostic biomarkers. These gene product(s) can also serve as potential targets for adjuvants to minimize gut-mediated metabolism of a drug of interest. Unfortunately, pinpointing the gene product(s) involved in the biotransformation of drugs is like finding a needle in a haystack, given that the human gut microbiome genome contains a hundred times more genes than the human genome (*28*). Although identifying gut bacterial enzymes conducting simple deconjugation chemistries on phase II-derived metabolites (*e.g.*, sulfatases and β-glucuronidases) is intuitive, tracking down the enzymes catalyzing general transformations (*i.e.*, reductions and hydrolysis) can be tough. In recent years, investigators have combined various NGS-based and MS-based technologies with biochemical and genetic assays to find the gene product(s) involved in the metabolism of drugs of interest. Three main strategies of note include (i) genome mining approaches, (ii) comparative transcriptomics, and (iii) functional metagenomics screening (*28*).

A genome mining approach is ideal when a homolog of a protein associated with conducting the chemistry of a drug of interest can be found in the literature. Balskus and co-workers elegantly displayed the utility of this approach to identify a previously unknown gut bacterial decarboxylase enzyme involved in the decarboxylation of levodopa to form dopamine (*57*). Levodopa is the primary drug to treat Parkinson's disease. Ideally, levodopa is decarboxylated to form the active agent dopamine after crossing the blood-brain barrier.

However, decarboxylation can occur prematurely in the gut, leading to reduced amounts of levodopa reaching the brain. The team postulated that the decarboxylation in the gut must be orchestrated by a pyridoxal-5'-phosphate (PLP)-containing enzyme; hence, they performed a Protein Basic Alignment Search Tool search against the complete set of HMP reference genome using the PLP-dependent tyrosine decarboxylase (TyrDC) from the food associated strain *Lactobacillus brevis* CGMCC 1.2028 as the query sequence. Among the match sequences, all *Enterococcus faecalis*-associated species were found to turn over levodopa to dopamine in anaerobic culture and thus, Balskus and co-workers did all subsequent structural characterization of TyrDC was done using this species. Importantly, 12 out of 19 human fecal samples tested were found to variably convert levodopa to dopamine. Using quantitative polymerase chain reaction (qPCR), the team observed that TyrDC gene (*tyrDC*) levels were predictive of levodopa metabolism among this tested cohort. Likewise, 16S rRNA sequencing analysis revealed that the abundance of *E. faecalis* also strongly predicted metabolism of levodopa in the cohort. These genome mining approaches highlight the importance of continued cultivation of extant databases like the HMP, as they can be excellent screening platforms to find gut bacterial enzymes of interest.

Comparative transcriptomics is a great strategy for finding genes involved in the metabolism of a drug if no homolog can be found. The strategy assumes that an enzyme is upregulated in the presence of a drug that it metabolizes. Turnbaugh and co-workers used RNA-seq to find the key genes involved in the reduction of digoxin to dihydrodigoxin (*55*). Digoxin is used to treat heart failures and arrythmias by blocking Na^+- and K^+-dependent adenosine triphosphatase. Reduction of digoxin to the inactive form dihydrodigoxin causes treatment efficacy to be unpredictable. Previously, the gut bacterial isolate *Eggerthella lenta* was found to

reduce digoxin *in vitro*. The team incubated *E. lenta* with and without digoxin and calculated mRNA levels across the genome. Upon comparison of digoxin plus and minus samples, the team found that a two-gene operon, *cgr*1 and *cgr*2 (termed cardiac glycoside reductase) was upregulated. Digoxin was found to be variably reduced in a cohort of 20 healthy individuals. Using qPCR, the team found that expression of the *cgr* operon relative to *E. lenta* 16S rRNA gene was significantly increased in high metabolizers. Subsequent structural characterization revealed that the product of *cgr*2, an [Fe-S]-dependent reductase (cgr2), was the acting enzyme that reduces digoxin, whereas the product of *cgr*1 was likely a membrane protein (cgr1) that served to anchor and transfer electrons to cgr2 (*56*). In the future, comparative proteomics should be applied to gut bacterial isolates with the aim of directly identifying the protein products that are upregulated in the presence and absence of a drug of interest.

Finally, functional metagenomic screens are useful when lead information (*i.e.*, the specific enzyme chemistry or a validated gut bacterial species) on the metabolism of a drug of interest cannot be ascertained. In general, such screens involve fragmenting DNA from a source (*i.e.*, gut isolate or a patient-derived fecal sample) already known to metabolize a drug of interest, inserting the resulting fragment library into an expression vector, and then screening to identify clones that can metabolize drugs of interest. Metagenomic sequencing of the hit clone can then pinpoint the gene product involved in the biotransformation. Using this strategy, Goodman and co-workers identified the gene product conducting the biotransformation of the oral calcium channel blocker diltiazem to deacetyldiltiazem by the gut isolate *Bacteroides thetaiotamicron*. The team sheared DNA from *Bacteroides thetaiotamicron* and cloned the resulting fragments (mean insert length = 3.1 kb) into *Escherichia coli* vectors (N = 51,000). Using a high-throughput pooling strategy in 384-well plates, the team found that deacetylation of

diltiazem was dependent on a specific gene, which they named *bt4096*. The product of this gene was found to turnover diltiazem to deacetyldiltiazem *in vitro*. The team also applied this strategy to map gene-drug-metabolite networks for *Bacteroides dorei* and *Collinsella aerofaciens*.

In addition to studying gut bacterial isolates, functional metagenomics can also be a powerful tool to elucidate gene product(s) involved in the biotransformation of therapeutics directly from patient-derived fecal samples. Compared with conducting functional metagenomic screens on specific gut bacterial isolates, working directly from fecal samples enables assessment of gene product(s) that may be from unculturable gut bacterial isolates. Donia and co-workers isolated DNA from a patient-derived fecal sample and cloned the resulting sheered DNA fragments (fragment insert size ranging from 2-4 kb) into *E. coli* expression vectors (80 pools with 6×10^4 unique clones, each!). This library was used to find the specific enzyme involved in the reduction of the steroid hydrocortisone to 20β-hydrocortisone. Reduction of hydrocortisone results in altered bioavailability. The team used an elegant dilution strategy whereby each serial dilution was tested for cortisone transformation activity, iteratively, until the specific clone processing hydrocortisone was isolated. Subsequent sequencing of the resulting clone revealed a gene product, which they named Hyd-red-2, that was likely from a *Bifidobacterium* species. This corroborates previous findings that *Bifidobacterium adolescents* metabolizes hydrocortisone (*58*).

Which gut bacterial strains are efficient at metabolizing the drug?

A healthy adult gut microbiome contains, on average, 1,000 bacterial species and 7,000 unique strains (*59*). Each of these microbes within us differ remarkably from one another in their ability to process small molecules. Considerable inter-species, and remarkably even inter-strain,

variability exist in gut-mediated drug metabolism, and thus, taxonomic identity of gut microbial species alone is not predictive of metabolic functions (*55*, *57*). Evolutionarily speaking, gut bacteria have evolved to resist exposure to foreign agents like drugs using mechanisms including cell permeability (or lack thereof), compound efflux, reduced susceptibility due to biofilm formation and physical location of the microbe within the GI tract (*60*). Differences in these defense barriers make certain drugs more susceptible to metabolism by a subset of gut microbes. For example, Gram-positive bacteria contain no outer membrane and, in general, are more likely to uptake drugs through passive diffusion. Gram-negative bacteria contain an outer membrane which in combination with the inner layer provides an orthogonal membrane system that significantly reduces the uptake of both hydrophobic and hydrophilic small molecules. Drug metabolism by Gram-negative bacteria is dependent on the ability of drugs to be taken up via porin and active transporters (*61*). High-throughput cell uptake studies using LC-MS can help identify isolates that are more or less efficient at intaking drugs of interest and could provide further explanation for interindividual variability in drug processing (*60*).

Defining the exact gene or gene product(s) that drive gut-mediated drug metabolism is essential for precisely predicting which species or strain can metabolize a drug of interest. For example, the type-strain of *E. lenta* reduces the cardiac drug digoxin to the inactive form dihydrodigoxin, whereas other *E. lenta* strains like FAA 1-3-56 and FAA 1-1-60 do not (*55*). Turnbaugh and co-workers explain the inter-strain variability in digoxin processing by *E. lenta* strains by first identifying the gene operon (*cgr*) driving the reduction of digoxin. They show that the type-strain contains the *cgr* operon, whereas the non-reducing strains do not. Another example includes the decarboxylation of dopamine to *m*-tyramine by *E. lenta* species. Balskus and co-workers observed that 10 out of 26 *E. lenta* strains successfully convert dopamine to *m*-

tyramine. They find that strain-to-strain variability in dopamine processing can be explained by a single-nucleotide polymorphism (SNP) found in the *dadh* gene that produced the decarboxylase. *E. lenta* species that can convert dopamine to *m*-tyramine contain an arginine at the 506-residue position, whereas species that cannot mediate the conversion have an SNP that results in a serine at the 506-residue position. Work by Balskus and co-workers highlight the importance of examining drug metabolism at the protein level, which can give insight into strain-level specificity of drug metabolism.

Interspecies variability in drug processing can also be explained by architectural differences in enzymes that process drugs of interest between strains. A great example of this is the gut bacterial β-glucuronidases (GUSs), which play an important role in energy capture by hydrolyzing GlcA from complex carbohydrates using a double displacement mechanism. During phase II metabolism, drugs can be conjugated to GlcA to increase overall polarity of the compound. In the gut, GUS enzymes can hydrolyze the GlcA sugar from the parent drug compound (*62*). Redinbo and co-workers mined the HMP (4.8 million translated sequences) using active site residues found essential for glucuronide recognition and processing and found 279 unique, non-redundant sequences. Significant structural work (over 20 published enzymes) of GUS enzymes reveals extensive variation in higher order structure. Oligomeric states that exist include monomers, dimers, trimers, tetramers and even hexamers. Variation in oligomeric states lead to difference in active site architecture, which can heavily impact glucuronide recognition and processing. Putative GUS sequences mined from the HMP database were categorized into six structural categories based on the size and location of an active site adjacent loop. Previous work has shown that this active site adjacent loop influences glucuronide

recognition and efficient processing. Loop classes have been found to be generally indicative of preferential processing of specific small molecule glucuronides (*63*, *64*).

Although genome mining approaches are an excellent starting point, they can potentially yield false negatives. For example, a few GUS enzymes mined through the HMP using key, conserved residues were found to not process small molecule glucuronides (*64*). Additionally, mining of metagenomic databases like the HMP does not afford any functional information (*i.e.*, gene expression). Thus, approaches that examine the proteins produced and select for proteins that are functionally active towards a target small molecule can be more widely applicable. Advances in activity-based proteomics can help circumvent this problem. Irreversible or affinity-based inhibitors that mimic the small molecule of interest can be used to discover and examine target proteins. By using a substrate mimic, proteins that are true processers can be identified, reducing false negatives. Furthermore, proteomic strategy can help identify proteins that are expressed and functionally active. **Chapter 2** will discuss the development and application of an activity-based profiling strategy to profile GUS enzymes from fecal microbiotas. Models that account for the physiological location of the microbe relative to drugs, cellular uptake and efflux of the drug, and variability in enzyme kinetics can help attain a precise understanding of interindividual variability in drug processing.

Does the drug inhibit gut bacterial enzyme activity?

The human gut microbiota is often described as a forgotten "organ" that has been inextricably linked to human health. As described above, drugs can alter the composition of gut bacterial communities, which can perturb normal functions of the gut microbiota like production

of SCFAs, essential vitamins, and analgesics. Another potential mechanism by which normal gut function can be altered is via the direct inhibition of gut bacterial enzymes by drugs and their metabolites. Remarkably, a survey of the primary literature shows that studies interrogating the direct inhibition of gut bacterial enzymes is lacking.

Of the few studies focusing on examining direct gut bacterial inhibition, Redinbo and co-workers have shown that GUS enzymes are selectively and potently inhibited by synthetic inhibitors (*65*). Using a combination of kinetic assays and x-ray crystallography, the team found that the piperazine and/or piperadine ring in these inhibitors act as warhead that covalently intercepts the catalytic cycle intermediate during GlcA hydrolysis. Interestingly, piperazine and piperadine rings are found widely in currently approved drugs (*65*). These nitrogen-containing rings serve to increase the solubility of drugs and serve as excellent building blocks. Indeed, experiments done by Redinbo and co-workers have shown that GUS enzymes are inhibited by approved drugs that contain a piperazine and piperadine ring. **Chapter 3** will expand on GUS inhibition by piperazine and piperadine containing FDA approved drugs.

OUTLOOK

Candidate drugs are being discovered at an unprecedented scale because of the powerful innovations in early-stage drug discovery. For example, DNA-encoded libraries enable drug hunters to screen upwards to 100 million compounds against a protein target of interest (*66*). Despite the increasing ease of search for candidate drugs, success rates of candidate drugs in human clinical trials nonetheless remain low (*1*). Poor success rates can be attributed to unforeseen toxicities and low drug efficacy. These unforeseen biological outcomes are primarily due to inaccurate recapitulation of host drug metabolism by current pre-clinical models. An

understudied component of drug metabolism is the role of the human gut microbiome in drug metabolism and disposition. Work in the past 20 years has clearly shown that the human gut microbiota heavily impacts the pharmacology of a drug. The burgeoning field of pharmacomicrobiomics is working to understand how the gut influence drug bioavailability and toxicity.

Astonishingly, the current total cost to identify a drug candidate, take it through human clinical trials, and launch it in the market is, on average, $1.2 billion (*67*). A significant portion of this cost is spent on studies in human clinical trials (*67*). To increase success rates in human clinical trials of drug candidates and thereby reduce overall drug costs, AstraZeneca has implemented a new "5R Framework," resulting in an increase from a 4% success rate from 2005-2010 in phase III to 19% in 2012-2016 (*68*). The 5R Framework places more weight on choosing the five Rights: the right target, the right tissue, the right safety, the right patient, and the right commercial potential. Moving forward, it is critical for researchers to assess how incorporation of pharmacomicrobiomics into strategies like the 5R Framework would further improve success rate of candidate drugs in human clinical trials. The vast complexity of the microbiome, and overwhelming technical challenge of testing hundreds of drugs against thousands of cultured isolates under multiple conditions, makes it difficult to incorporate pharmacomicrobiomic studies into existing drug development pipelines. If high-throughput strategies like those developed by the Goodman and co-workers and Donia and co-workers can be incorporated into drug development pipelines, then lead drug agents can be further optimized to ensure that gut-mediated alterations are minimized.

Side effects mediated by the gut microbiota are coming to light for many currently administered therapeutics. For example, gut-mediated reduction of digoxin and the

decarboxylation of levodopa heavily impact the bioavailability of these drugs, which explains the high interindividual variability in treatment outcome observed in patients taking these therapeutics. Reducing gut-mediated side effects can facilitate appropriate dosing and improve overall treatment efficacy. Strategies to reduce gut-mediated biotransformation of drugs include the development of small molecules that inhibit gut enzymes (adjuvants). The gut microbiota can also be modulated by diet (prebiotics and probiotics) to control the level of enzymes that may act upon drugs. **Chapter 4** will highlight how diet may influence the composition of GUS enzymes.

Poor success rates of candidate drugs lead to sunk costs, which are a dead weight loss to society. In economics, dead weight loss is the total welfare loss to society. Investing to improve pre-clinical models can enable drug candidates with greater chances of success, thereby potentially reducing overall drug costs. If the dead weight loss associated with clinical trials (*i.e.*, monetary loss due to resources and personnel associated with running a clinical trial) can be reduced, then saved resources can be allocated to other elements in the drug discovery pipeline. By integrating pharmacomicrobiomics into the drug discovery pipeline, the hunt for new and efficacious drugs is one step closer to the precision of Arjuna's arrows, which will diminish the dead weight loss in the drug development industry.

CHAPTER 2: DISCOVERING THE MICROBIAL ENZYMES DRIVING DRUG TOXICITY WITH ACTIVITY-BASED PROTEIN PROFILING[1]

INTRODUCTION

The gut microbiota are capable of metabolizing a myriad of drugs (*10*), and the biotransformation of these compounds by commensal intestinal bacteria can impact therapeutic outcomes by altering drug efficacy, and in some instances, inducing disease onset (*12*). Since each person harbors a unique set of gut microbes, drug response varies considerably between individuals (*69*). Although key recent reports have profoundly advanced our understanding of the central microbes and genes implicated in the metabolism of drugs (*12*, *69*), only a handful of studies have focused on gut bacterial proteins implicated in the biotransformation of drug metabolites (*56*, *70*, *71*). Pinpointing the exact microbial enzymes that process drugs in the gut could lead to the development of precision biomarkers for the determination of therapeutic efficacy and may serve as drug targets for the modulation of the gut microbiota to optimize drug responses.

The gut bacterial β-glucuronidase (GUS) enzyme mediates drug-induced gastrointestinal (GI) toxicity by reversing glucuronidation, a Phase II transformation that inactivates and detoxifies drugs by conjugating them to glucuronic acid (GlcA) (**Figure 2.1a**) (*62*). Inactive drug glucuronides created in the liver traverse the biliary duct to reach the intestines where they are

[1]This chapter previously appeared as an article in the *ACS Chemical Biology*. The original citation is as follows P. B. Jariwala, S. J. Pellock, D. Goldfarb, E. W. Cloer, M. Artola, J. B. Simpson, A. P. Bhatt, W. G. Walton, L. R. Roberts, M. B. Major, G. J. Davies, H. S. Overkleeft, M. R. Redinbo, Discovering the Microbial Enzymes Driving Drug Toxicity with Activity-Based Protein Profiling. *ACS Chem. Biol.* (2019).

excreted from the body (*72*). However, once in the gut, drug glucuronides have the potential to be reactivated via the hydrolytic removal of the GlcA tag by gut bacterial GUS enzymes. Intestinal reactivation of drug metabolites has been reported to cause acute, dose-limiting GI toxicities (*73*, *74*). The severity of irinotecan-induced GI toxicity varies considerably between patients and may be due to the interindividual variability of the human gut microbiota (*75*, *76*). Previous analysis of the Human Microbiome Project (HMP) stool sample database revealed that the gut microbiota contains hundreds of putative GUS enzymes with seven unique structural classes that display varying catalytic efficiencies against the reporter substrates *p*-nitrophenyl-β-D-glucuronide (*p*NP-GlcA) and 4-methylumbelliferone-β-D-glucuronide (4-MUG) (*64*, *77*). Since gut bacterial GUS enzymes process glucuronide conjugates with varying efficiencies, we hypothesized that interindividual differences in gut bacterial GUS abundance and composition might influence the differential drug response to irinotecan.

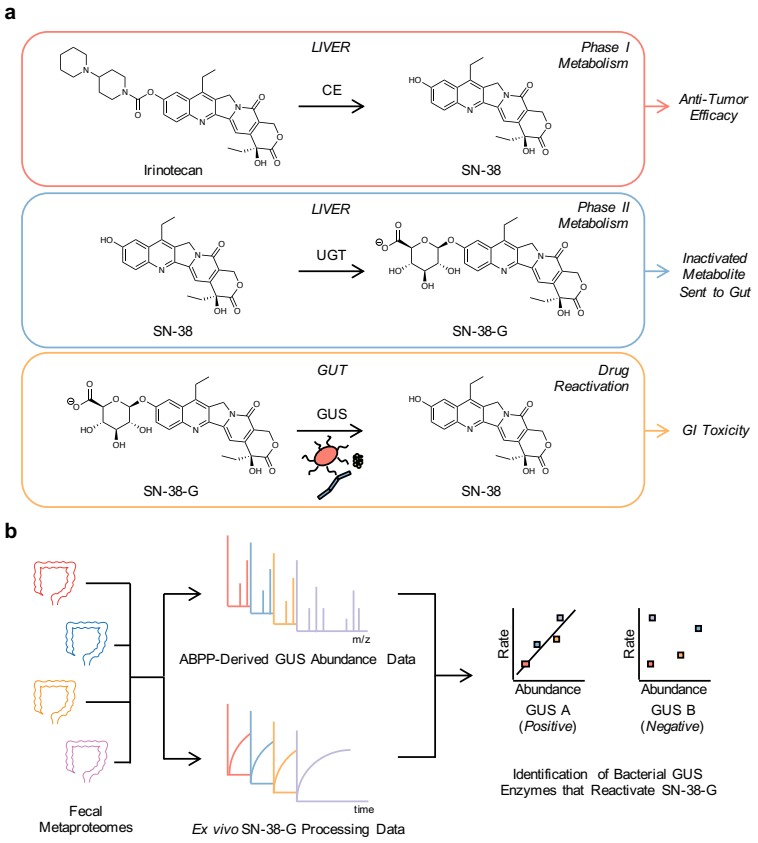


Figure 2.1 Approach to identify GUS enzymes responsible for drug glucuronide metabolism in the human gut microbiome. (**a**) The anti-cancer pro-drug irinotecan is processed by carboxylesterases (CE) during Phase I metabolism to form the active metabolite, SN-38, a topoisomerase I inhibitor. Uridine-diphosphate glucuronosyltransferases (UGTs) in the liver glucuronidate SN-38, which inactivates and promotes its excretion from the body. Bacterial GUS can reactivate SN-38 via hydrolytic removal of the glucuronic acid sugar. Reactivation of SN-38 has been reported to cause severe, dose-limiting GI toxicity. (**b**) To decipher the causative GUS enzymes in SN-38-G reactivation, we employed a strategy that integrates ABPP-enabled GUS abundance data with *ex vivo* SN-38-G processing data obtained from fecal metaproteomes. Correlation analysis between SN-38-G processing data and GUS abundance data can be used to identify specific GUS enzymes involved in SN-38-G processing.

Efficient and facile strategies to identify the exact gut bacterial GUS enzymes that process drug glucuronides of interest from fecal material are lacking. Significant advancements in mass spectrometry (MS) and related bioinformatics software have made the identification and quantification of proteins from complex fecal supernatant possible (*37*, *39*, *78*). However, recent work has shown that shotgun-based metaproteomics cannot accurately identify and quantify low abundance proteins from fecal lysates (*39*). Activity-based probes (ABPs) serve as powerful tools to access low abundance targets and enrich for functionally active proteins from fecal lysate (*39*, *41*). ABPs target the catalytic machinery of specific enzymes and can be outfitted with a chemical handle for target enrichment, enabling identification and quantitation using MS. Activity-based protein profiling (ABPP)-enabled GUS abundance data obtained from fecal metaproteomes can then be correlated with *ex vivo* drug glucuronide processing data to identify the exact GUS enzymes that process drug glucuronides of interest (**Figure 2.1b**).

Using a unique pipeline that integrates ABPP-enabled GUS abundance data with *ex vivo* SN-38-G processing data, we pinpoint, from human feces, the exact bacterial GUS enzymes that reactivate SN-38, the active metabolite of the anti-cancer drug irinotecan. For the first time, we show that cyclophellitol-based ABPs can be used to identify and quantify gut bacterial GUS enzymes from human fecal lysate. We identify Loop 1 (L1) GUS enzymes as key modulators of SN-38 reactivation and verify this finding with *in vitro* kinetic data and structural modeling. Finally, we use the ABPP-enabled pipeline outlined in this study to provide a rationale for differential GUS inhibition between human fecal samples by previously designed piperazine-containing GUS inhibitors.

RESULTS

Cyclophellitol-based inhibitors and ABPs target structurally diverse gut bacterial GUS enzymes

Cyclophellitol-based epoxide and aziridine inhibitors **1** and **2** and activity-based probes (ABPs) **3** and **4** were previously developed to profile GUS in human cells (**Figure 2.2a**) (*79*). Since human and bacterial GUS utilize the same retaining mechanism to catalyze glucuronide hydrolysis, we hypothesized that **1–4** could also be used to target gut bacterial GUS enzymes from the human gut (*80*). To confirm that **1–4** covalently label the catalytic glutamate in bacterial GUS enzymes, we determined the 2.4 Å resolution crystal structure of a bacterial GUS from the human gut commensal strain *B. uniformis* (*Bu*GUS-2) in complex with the unsubstituted cyclophellitol-based aziridine inhibitor (**2**) (**Table 2.1**). Examination of the active site revealed inhibitor **2** covalently linked to the catalytic nucleophile (E526) of *Bu*GUS-2 indicating that it is also an inhibitor of bacterial GUS (**Figure 2.2b**). Key contacts were also observed between the carboxylic acid of inhibitor **2** and N591 and K593, the conserved NxK motif that is essential for recognition of glucuronides by bacterial GUS (**Figure 2.2b**) (*73*). In-gel labelling of wild type and mutant enzymes using Cy5-ABP (**4**) further indicated that a functionally active GUS is necessary for labelling and that the NxK motif is essential for recognition of ABP **4** by bacterial GUS enzymes (**Figure 2.2c**).

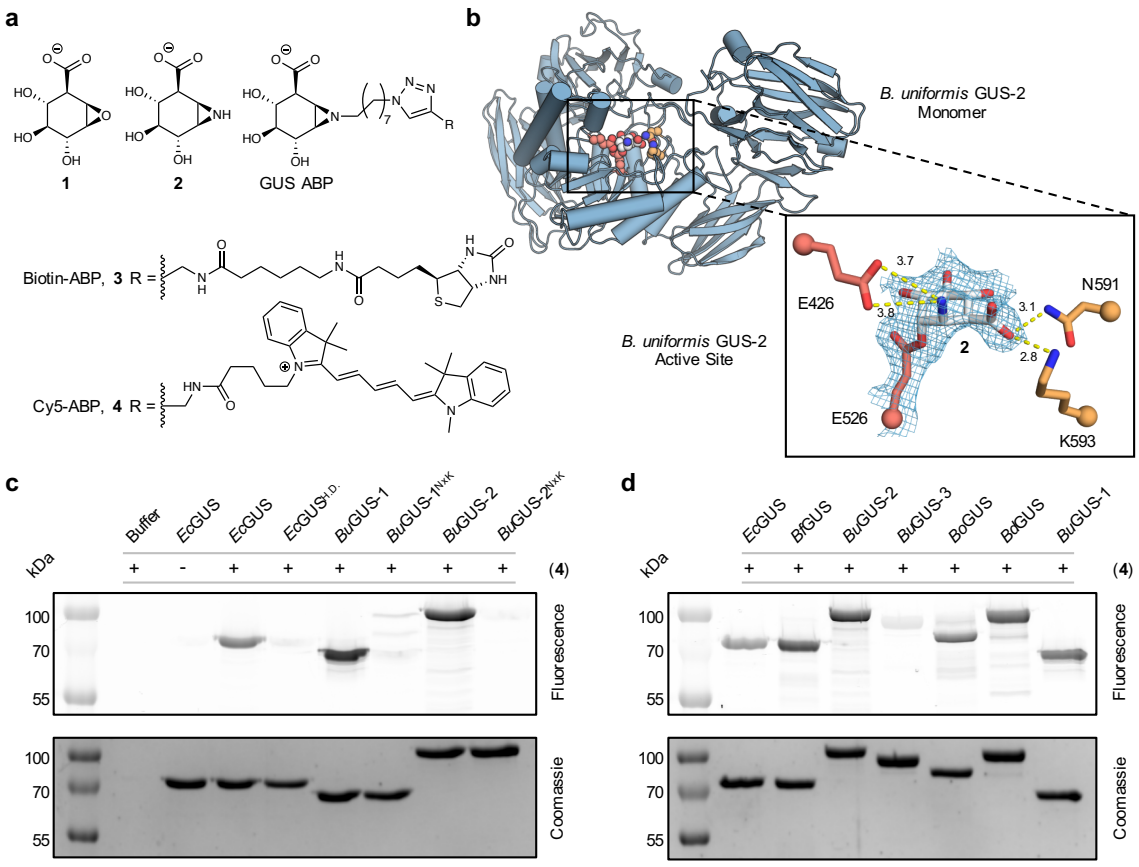


Figure 2.2 Cyclophellitol-based inhibitors and ABPs label structurally diverse gut bacterial GUS enzymes. (**a**) Cyclophellitol-based epoxide and aziridine inhibitors **1** and **2** and ABPs **3** and **4**. (**b**) A 2.4 Å resolution crystal structure (PDB: 6NZG) of inhibitor **2** bound to *Bu*GUS-2. Inset shows $2F_o - F_c$ map (after refinement) at 1 σ and distances are shown in Å. (**c**) In-gel fluorescence labelling of wild type and inactive GUS controls by ABP **4**. *E. coli* GUS (*Ec*GUS), heat-denatured *E. coli* GUS (*Ec*GUS$^{H.D.}$), *B. uniformis* GUS-1 (*Bu*GUS-1), *B. uniformis* GUS-2 (*Bu*GUS-2), and *Bu*GUS-1 and *Bu*GUS-2 mutants (*Bu*GUS-1^{NxK} and *Bu*GUS-2^{NxK}) where the asparagine and lysine residues of the NxK motif have been mutated to alanines. (**d**) In-gel fluorescence labelling of structurally diverse gut bacterial GUS by ABP **4**. *B. fragilis* GUS (*Bf*GUS), *B. uniformis* GUS-3 (*Bu*GUS-3), *B. ovatus* GUS (*Bo*GUS), and *B. dorei* GUS (*Bd*GUS). All wild type and mutant proteins were exogenously purified.

Table 2.1 Crystallographic statistics for *B. uniformis* GUS-2 bound to the unsubstituted cyclophellitol-based aziridine ABP (2).

Resolution range	29.30-2.43 (2.52-2.43)
Space group	$P2_12_12_1$
Unit cell [a, b, c (Å); α, β, γ (°)]	74.5, 142.4, 180.7; 90.0, 90.0, 90.0
Total reflections	952737 (93990)
Unique reflections	73051 (7171)
Multiplicity	13.0 (13.1)
Completeness (%)	99.7 (99.5)
Mean I/sigma(I)	15.9 (2.8)
Wilson B-factor	40.5
R-merge	0.146 (0.859)
R-pim	0.042 (0.247)
CC1/2	0.997 (0.782)
R-work	0.170 (0.221)
R-free	0.221 (0.306)
Number of Non-Hydrogen Atoms	14097
Macromolecules	13548
Ligands	72
Solvent	477
Protein residues	1678
RMS (bonds) (Å)	0.008
RMS (angles) (°)	0.98
Ramachandran outliers (%)	0
Rotamer outliers (%)	1.99
Clash score	4.23
Average B-factor (Å^2)	42.8
Macromolecules	42.9
Ligands	45.8
Solvent	39.9
PDB code	6NZG

The gut microbiota contains a structurally diverse assortment of bacterial GUS enzymes (*77*). Using in-gel labelling studies, we found that ABP **4** labels most exogenously purified GUS enzymes from this structurally and functionally diverse group of enzymes (**Figure 2.2d**). Labelling was not observed for a GUS from *B. uniformis* (*Bu*GUS-3), which corroborates a recent study reporting that *Bu*GUS-3 does not process small molecule glucuronides and poorly

processes GlcA-containing polysaccharides (*64*). *In vitro* apparent IC_{50} values showed that **1–4** inhibit *E. coli* GUS (*Ec*GUS), *B. uniformis* GUS-1 (*Bu*GUS-1), and *Bu*GUS-2 with values ranging from 20 nM to 4 µM (**Figure 2.3**). Further kinetic analysis of GUS inactivation by **1–4** displayed k_i/K_I values that mirrored the IC_{50} values (**Table 2.2 and Figure 2.4**). Taken together, these data establish that cyclophellitol-based inhibitors and ABPs target structurally diverse and functionally active gut bacterial GUS enzymes.

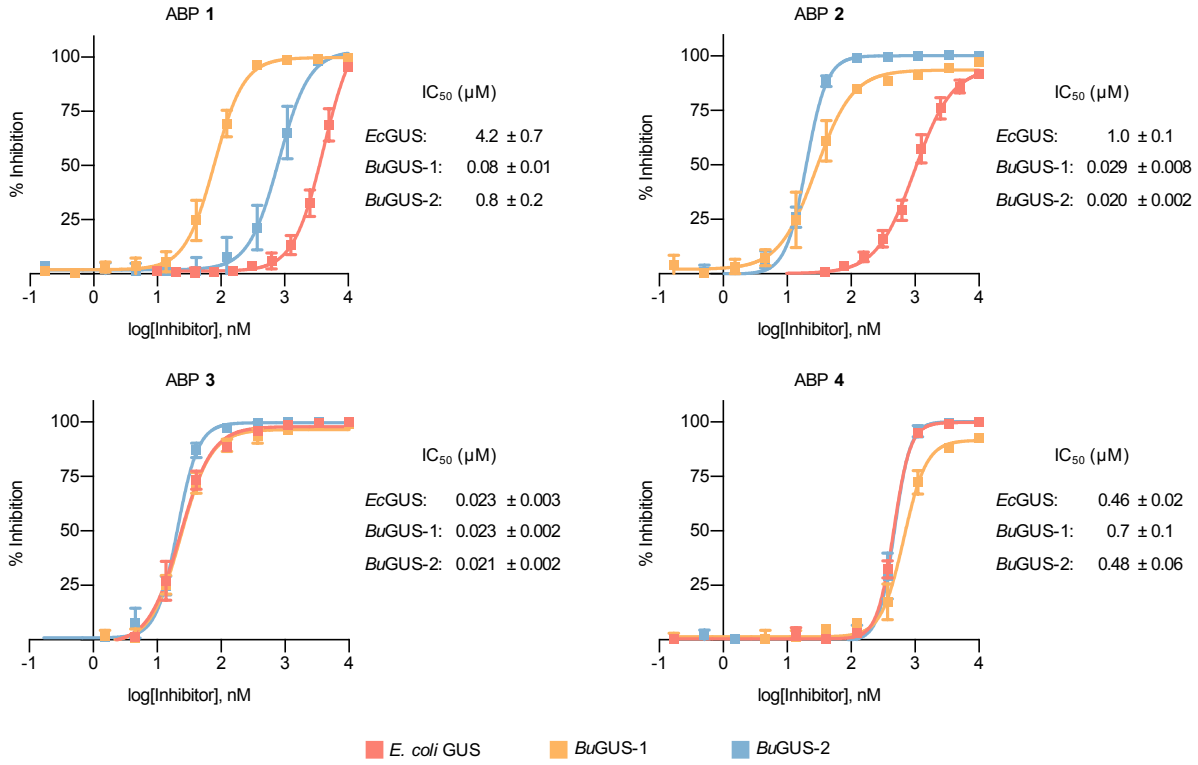


Figure 2.3 Dose-response plots for inhibition of select gut bacterial GUS enzymes by cyclophellitol-based inhibitors and ABPs. Apparent IC_{50} values shown are the mean ± standard deviation using $N = 3$ biological replicates. Lower values indicate more potent inhibition.

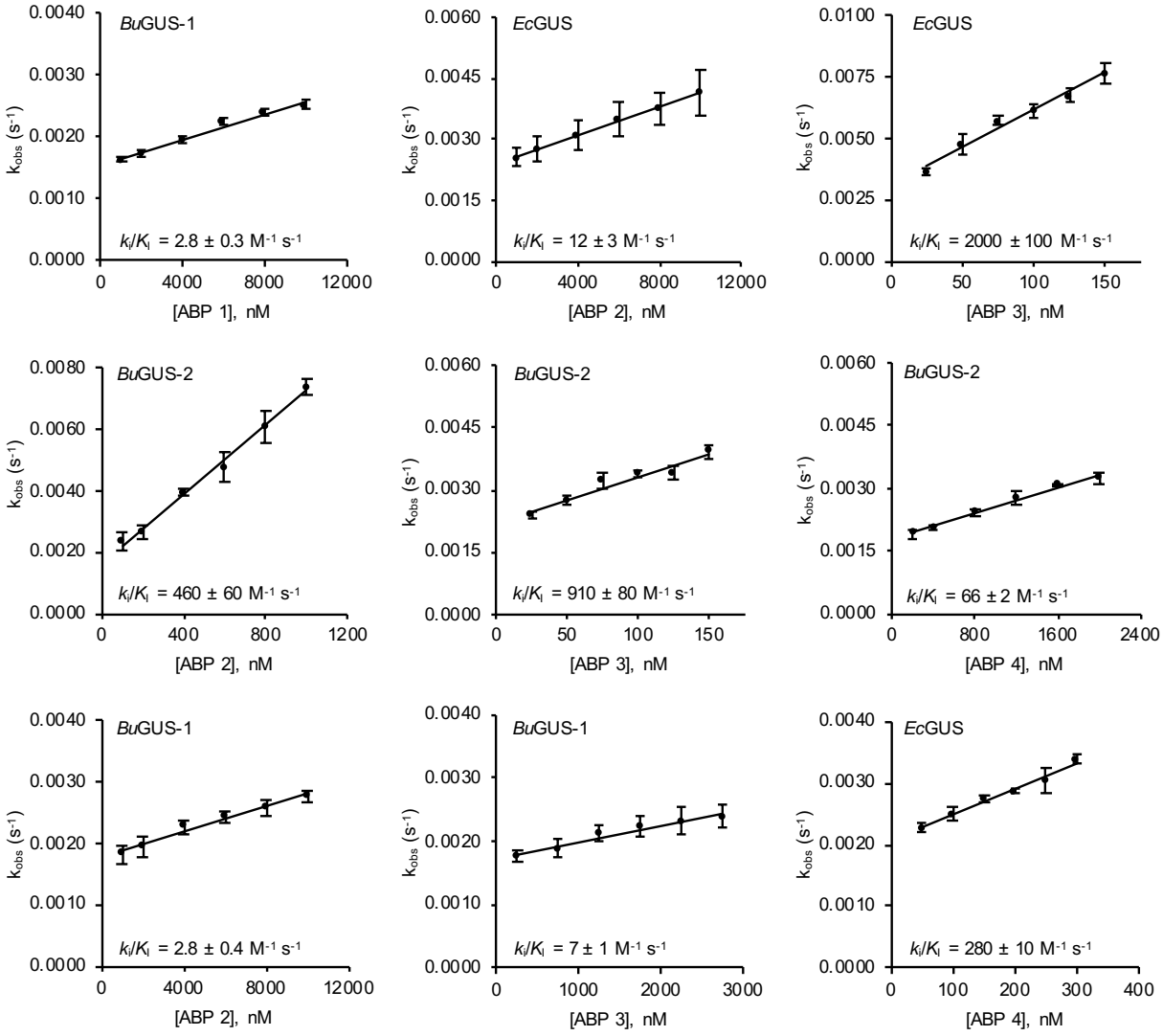


Figure 2.4 Derivation of kinetic parameters for GUS inhibition by inhibitors and ABPs 1-4. Plots shown are k_{obs} vs. concentration of inhibitors and ABPs **1-4**.

Table 2.2 Kinetic parameters for inhibition of select gut bacterial GUS enzymes by cyclophellitol-based inhibitors and ABPs. Values shown are the mean ± standard using $N = 3$ biological replicates. N.M. = not measurable due to slow inhibition. Higher values indicate more potent inhibition.

	k_i/K_I (M^{-1} s^{-1})			
GUS	**Compound 1**	**Compound 2**	**Compound 3**	**Compound 4**
***Ec*GUS**	N.M.	12 ± 3	2000 ± 100	280 ± 10
***Bu*GUS-1**	2.8 ± 0.3	2.8 ± 0.4	7 ± 1	N.M.
***Bu*GUS-2**	N.M.	460 ± 60	910 ± 80	66 ± 2

Cyclophellitol-based ABPs label GUS enzymes in mouse fecal mixtures

As a controlled proof-of-concept for labelling of GUS enzymes by the cyclophellitol-based ABPs, we collected fecal samples from wild-type germ-free mice and mice mono-associated with *gus*$^{+}$ *E. coli* (*Ec*GUS$^{M.A.}$; M.A., mono-associated). Labelling of *Ec*GUS$^{M.A.}$ fecal extracts with ABP **4** revealed a single, prominent band with a molecular weight indicative of recombinant *Ec*GUS (**Figure 2.5**). Heat denaturation of the fecal extracts from the mono-associated mice (*Ec*GUS$^{M.A. + H.K.}$; H.K., heat-killed) resulted in complete loss of labelling, which further establishes that these ABPs only label functionally active GUS enzymes. No significant labelling was observed in the fecal mixtures collected from germ-free mice which indicates that labelling of non-microbial protein is minimal. Finally, we show that labelling of *Ec*GUS by ABP **4** can be blocked in a complex fecal setting in a dose-dependent manner using the pan-GUS inhibitor, D-glucaro-1,4-lactone. These results demonstrate successful labelling of bacterial GUS enzymes in a controlled fecal matrix.

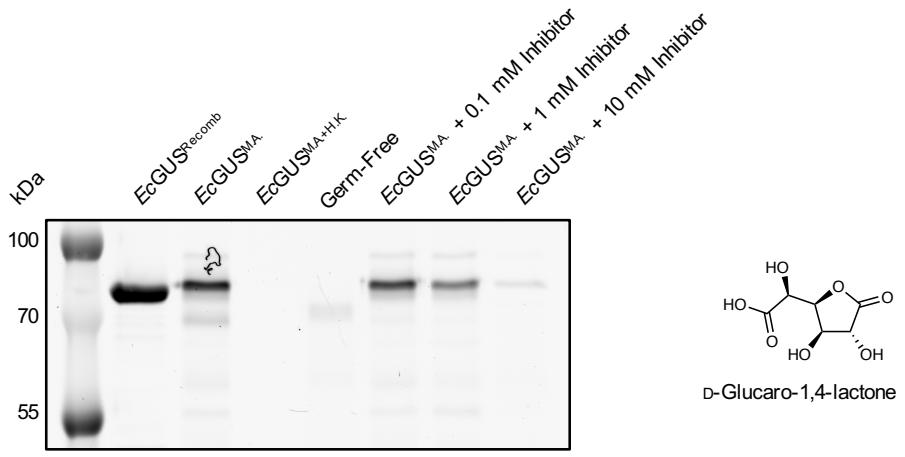


Figure 2.5 Cyclophellitol-based ABP labels GUS enzymes in mouse fecal lysate. Fecal extracts from germ-free and *E. coli* mono-associated mice were labelled with Cy5-ABP. Labelling was blocked in a dose-dependent manner using the pan-GUS inhibitor, D-glucaro-1,4-lactone. Recombinant *Ec*GUS sample (*Ec*GUS^{Recomb}), mono-associated *Ec*GUS fecal sample (*Ec*GUS$^{M.A.}$), heat-killed, mono-associated *Ec*GUS fecal sample (*Ec*GUS$^{M.A. + H.K.}$).

Gut bacterial GUS enzymes can be identified and quantified using cyclophellitol-based aziridine ABPs

After confirming that the cyclophellitol-based inhibitors and ABPs **1–4** label bacterial GUS enzymes *in vitro* and in a controlled mouse model, we performed ABPP to identify and quantify bacterial GUS enzymes present in human fecal samples collected from two females (F1 and F2) and two males (M1 and M2). We extracted total protein from human fecal lysates and enriched for GUS using the biotin-ABP (**3**) (**Figure 2.6a**). Resultant samples were analyzed by liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) and a bioinformatics pipeline that queried the integrated gene catalog (IGC) using MetaLab to assemble and quantify enriched protein groups (**Table 2.3, Table 2.4, and Table 2.5**) (*29*, *37*). Protein groups were defined as GUS enzymes if sequences shared similarity to either *Ec*GUS, *C. perfringens* GUS (*Cp*GUS), *S. agalactiae* GUS (*Sa*GUS), or *B. fragilis* GUS (*Bf*GUS) and contained the catalytic glutamates as well as the NxK motif (**Figure 2.6a and Figure 2.7**). GUS abundance is

represented by LFQ intensities, which are normalized, and combined peptide signal intensities as determined by the MaxLFQ algorithm in MaxQuant. Further taxonomic classifications are shown below the abundance heatmap (**Figure 2.6a**). Unknown refers to protein groups where the phylum assignment was ambiguous due to mapping of GUS peptides to multiple phyla. Analysis of the identified GUS protein groups revealed significant variations in taxa, structure, and abundance of the GUS enzymes present in the four fecal samples (**Figure 2.6b and Figure 2.8**). Individuals contained between 15–29 bacterial GUS protein groups, similar to a recent metagenomic study which showed that individuals harbor between 4–38 bacterial *gus* genes (**Figure 2.8a**) (*77*). Phylum-level analysis revealed that all four individuals predominantly contained GUS enzymes from Firmicutes but displayed substantial variation in GUS composition at lower taxa levels (**Figure 2.6b and Figure 2.8b**). Further examination using a previously defined GUS structure rubric allowed us to analyze the identified GUS protein groups based on three-dimensional structure, which revealed significant structural diversity (**Figure 2.8c**) (*77*). We have developed an ABPP-enabled proteomics pipeline to identify and quantify functionally active GUS enzymes present in human fecal material.

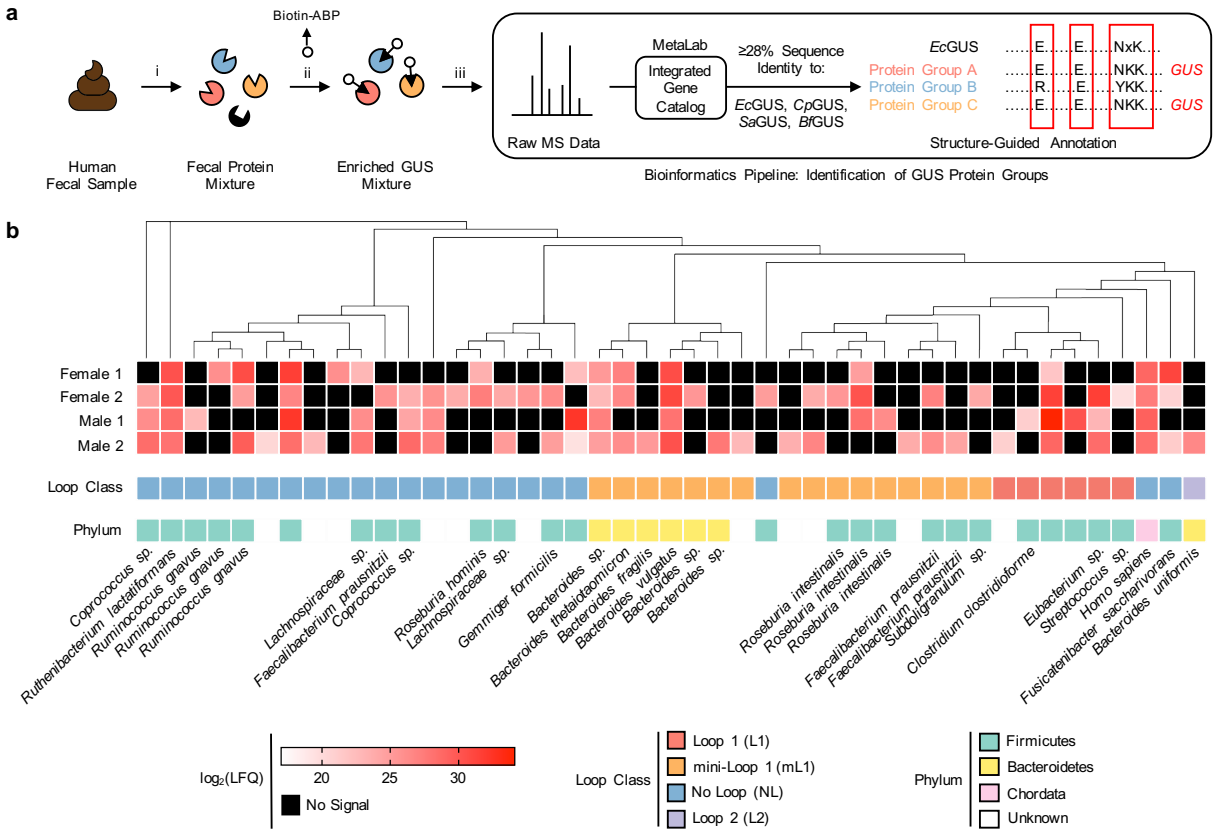


Figure 2.6 Probe-enabled proteomics and structure-guided bioinformatics enable identification and relative quantitation of bacterial GUS enzymes from human fecal samples. (**a**) General schematic of the probe-enabled proteomics pipeline used to identify and quantify GUS from fecal material. In brief, (**i**) proteins are extracted from feces using ultrasonication, (**ii**) GUS enzymes are enriched using the pre-clicked biotin-ABP (**3**) and streptavidin beads, and (**iii**) MetaLab is used to query the integrated gene catalog using raw MS data to assemble and quantify protein groups. Only proteins with the GUS fold and active site features, including the catalytic glutamates (E) and NxK motif, are defined as GUS enzymes. (**b**) Heatmap of identified GUS protein groups organized by sequence similarity and color coded by abundance. Further taxonomic classifications are shown below the abundance heatmap.

Table 2.3 Calculated LFQ intensity for GUS identified protein groups. Protein groups are organized by sequence similarity.

Protein Group #	Female 1 (LFQ intensity)	Female 2 (LFQ intensity)	Male 1 (LFQ intensity)	Male 2 (LFQ intensity)
1	0.00E+00	0.00E+00	0.00E+00	9.10E+07
2	2.35E+09	1.64E+06	0.00E+00	2.49E+06
3	4.65E+08	1.70E+08	6.21E+08	3.16E+08
4	0.00E+00	7.90E+05	0.00E+00	0.00E+00
5	0.00E+00	3.16E+09	8.25E+06	3.78E+08
6	0.00E+00	0.00E+00	1.22E+09	0.00E+00
7	3.33E+06	2.99E+09	1.82E+10	4.60E+08
8	0.00E+00	0.00E+00	2.17E+06	0.00E+00
9	0.00E+00	0.00E+00	0.00E+00	1.93E+06
10	0.00E+00	1.67E+07	0.00E+00	0.00E+00
11	0.00E+00	0.00E+00	0.00E+00	2.17E+07
12	0.00E+00	1.25E+08	0.00E+00	7.11E+07
13	0.00E+00	0.00E+00	0.00E+00	1.43E+07
14	0.00E+00	0.00E+00	7.57E+07	0.00E+00
15	2.79E+07	1.16E+09	3.06E+08	0.00E+00
16	0.00E+00	4.18E+07	0.00E+00	0.00E+00
17	0.00E+00	1.41E+07	0.00E+00	1.05E+08
18	0.00E+00	0.00E+00	0.00E+00	1.15E+07
19	0.00E+00	3.04E+07	0.00E+00	0.00E+00
20	0.00E+00	0.00E+00	0.00E+00	7.15E+06
21	0.00E+00	0.00E+00	0.00E+00	2.08E+08
22	0.00E+00	4.23E+07	0.00E+00	0.00E+00
23	1.55E+09	2.17E+09	2.51E+08	1.28E+09
24	0.00E+00	0.00E+00	0.00E+00	3.72E+07
25	1.32E+08	9.06E+07	0.00E+00	6.74E+07
26	3.94E+07	6.79E+06	1.35E+08	2.80E+07
27	5.27E+06	0.00E+00	5.11E+09	7.30E+05
28	0.00E+00	7.48E+07	0.00E+00	2.96E+07
29	0.00E+00	4.63E+07	0.00E+00	0.00E+00
30	0.00E+00	1.67E+07	0.00E+00	3.58E+07
31	1.10E+07	1.58E+08	0.00E+00	0.00E+00
32	0.00E+00	1.47E+07	0.00E+00	0.00E+00
33	0.00E+00	5.37E+07	8.19E+07	1.51E+08
34	0.00E+00	1.17E+07	2.01E+07	3.99E+08
35	0.00E+00	5.25E+07	0.00E+00	0.00E+00
36	7.85E+06	0.00E+00	6.76E+07	2.19E+08
37	6.37E+07	0.00E+00	0.00E+00	0.00E+00
38	0.00E+00	0.00E+00	0.00E+00	7.03E+06
39	3.12E+09	1.03E+08	4.77E+09	3.38E+08
40	0.00E+00	0.00E+00	0.00E+00	1.28E+06
41	1.63E+09	3.05E+07	0.00E+00	7.18E+08
42	6.37E+07	0.00E+00	0.00E+00	0.00E+00
43	0.00E+00	0.00E+00	6.69E+06	0.00E+00
44	1.31E+09	9.77E+08	3.55E+08	2.61E+08
45	0.00E+00	2.54E+07	6.90E+07	3.52E+08

Table 2.4 Taxa breakdown for GUS identified protein groups. Protein groups are organized by sequence similarity. Blanks refers to protein groups where the phylum assignment was ambiguous due to mapping of GUS peptides to multiple phyla.

Protein Group #	Loop class	Phylum	Family	Genus	Species
1	Loop 2	*Bacteroidetes*	*Bacteroidaceae*	*Bacteroides*	*B. uniformis*
2	No loop	*Firmicutes*	*Lachnospiraceae*	*Fusicatenibacter*	*F. saccharivorans*
3	No loop	*Chordata*	*Hominidae*	*Homo*	*Homo sapiens*
4	Loop 1	*Firmicutes*	*Streptococcaceae*	*Streptococcus*	
5	Loop 1	*Firmicutes*	*Eubacteriaceae*	*Eubacterium*	
6	Loop 1	*Firmicutes*			
7	Loop 1	*Firmicutes*			
8	Loop 1	*Firmicutes*	*Lachnospiraceae*	*Lachnoclostridium*	*L. clostridioforme*
9	Loop 1				
10	Mini-Loop 1	*Firmicutes*	*Ruminococcaceae*	*Subdoligranulum*	
11	Mini-Loop 1	*Firmicutes*	*Ruminococcaceae*	*Faecalibacterium*	*F. prausnitzii*
12	Mini-Loop 1	*Firmicutes*	*Ruminococcaceae*	*Faecalibacterium*	*F. prausnitzii*
13	Mini-Loop 1				
14	Mini-Loop 1	*Firmicutes*	*Lachnospiraceae*	*Roseburia*	*R. intestinalis*
15	Mini-Loop 1	*Firmicutes*	*Lachnospiraceae*	*Roseburia*	*R. intestinalis*
16	Mini-Loop 1	*Firmicutes*	*Lachnospiraceae*	*Roseburia*	*R. intestinalis*
17	Mini-Loop 1				
18	Mini-Loop 1				
19	No loop	*Firmicutes*			
20	Mini-Loop 1				
21	Mini-Loop 1	*Bacteroidetes*	*Bacteroidaceae*	*Bacteroides*	
22	Mini-Loop 1	*Bacteroidetes*	*Bacteroidaceae*	*Bacteroides*	
23	Mini-Loop 1	*Bacteroidetes*	*Bacteroidaceae*	*Bacteroides*	*B. vulgatus*
24	Mini-Loop 1	*Bacteroidetes*	*Bacteroidaceae*	*Bacteroides*	*B. fragilis*
25	Mini-Loop 1	*Bacteroidetes*	*Bacteroidaceae*	*Bacteroides*	*B. thetaiotaomicron*
26	Mini-Loop 1	*Bacteroidetes*	*Bacteroidaceae*	*Bacteroides*	
27	No loop	*Firmicutes*			
28	No loop	*Firmicutes*	*Ruminococcaceae*	*Gemmiger*	*G. formicilis*
29	No loop				
30	No loop	*Firmicutes*	*Lachnospiraceae*		
31	No loop	*Firmicutes*	*Lachnospiraceae*	*Roseburia*	*R. hominis*
32	No loop				
33	No loop				
34	No loop	*Firmicutes*	*Lachnospiraceae*	*Coprococcus*	
35	No loop	*Firmicutes*	*Ruminococcaceae*	*Faecalibacterium*	*F. prausnitzii*
36	No loop	*Firmicutes*	*Lachnospiraceae*		
37	No loop				
38	No loop				
39	No loop	*Firmicutes*			
40	No loop				
41	No loop	*Firmicutes*	*Lachnospiraceae*	*Blautia*	*R. gnavus*
42	No loop	*Firmicutes*	*Lachnospiraceae*	*Blautia*	*R. gnavus*
43	No loop	*Firmicutes*	*Lachnospiraceae*	*Blautia*	*R. gnavus*
44	No loop	*Firmicutes*	*Ruminococcaceae*	*Ruthenibacterium*	*R. lactatiformans*
45	No loop	*Firmicutes*	*Lachnospiraceae*	*Coprococcus*	

Table 2.5 Compiled GUS abundance based on GUS loop class. Bacterial GUS abundance broken down by GUS loop categories.

Bacterial GUS Abundance (LFQ Intensity)	Female 1	Female 2	Male 1	Male 2
Loop 1	3.33E+06 (0%)	6.15E+09 (54%)	1.95E+10 (63%)	8.40E+08 (16%)
mini-Loop 1	1.75E+09 (17%)	3.67E+09 (32%)	7.67E+08 (2%)	1.86E+09 (35%)
No Loop	8.56E+09 (83%)	1.60E+09 (14%)	1.05E+10 (34%)	2.52E+09 (47%)
Loop 2	0.00E+00 (0%)	0.00E+00 (0%)	0.00E+00 (0%)	9.10E+07 (2%)
Total	1.03E+10	1.14E+10	3.07E+10	5.30E+09
Human GUS Abundance (LFQ Intensity)	4.65E+08	1.70E+08	6.21E+08	3.16E+08
Total GUS Abundance (LFQ Intensity)	1.08E+10	1.16E+10	3.13E+10	5.62E+09

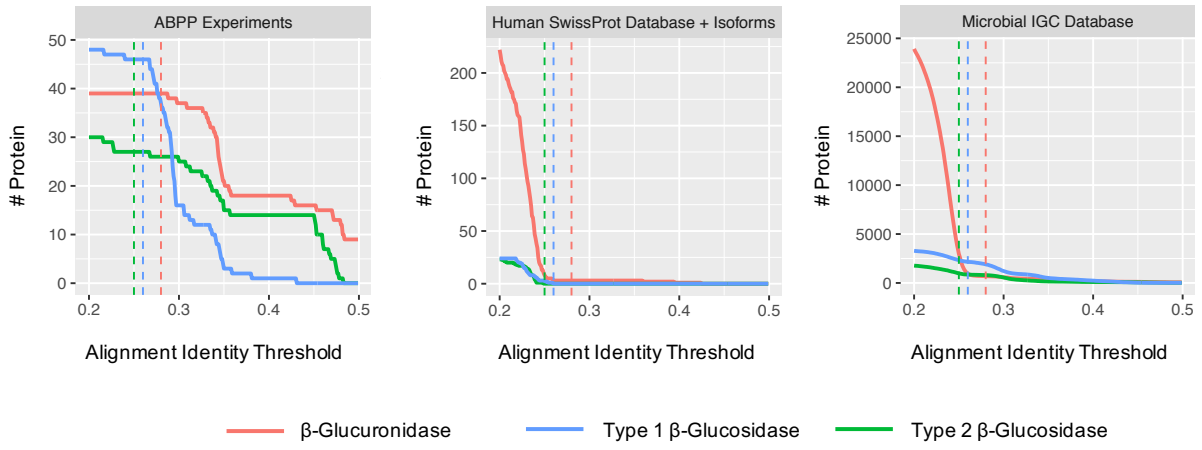


Figure 2.7 Determining sequence identity thresholds for GUS, Type I β-glucosidases, Type II β-glucosidases. Dashed vertical lines indicate chosen sequence identity thresholds: 0.28 (28%) for GUS, 0.25 (25%) for Type 1 β-glucosidases, and 0.26 (26%) for Type II β-glucosidases.

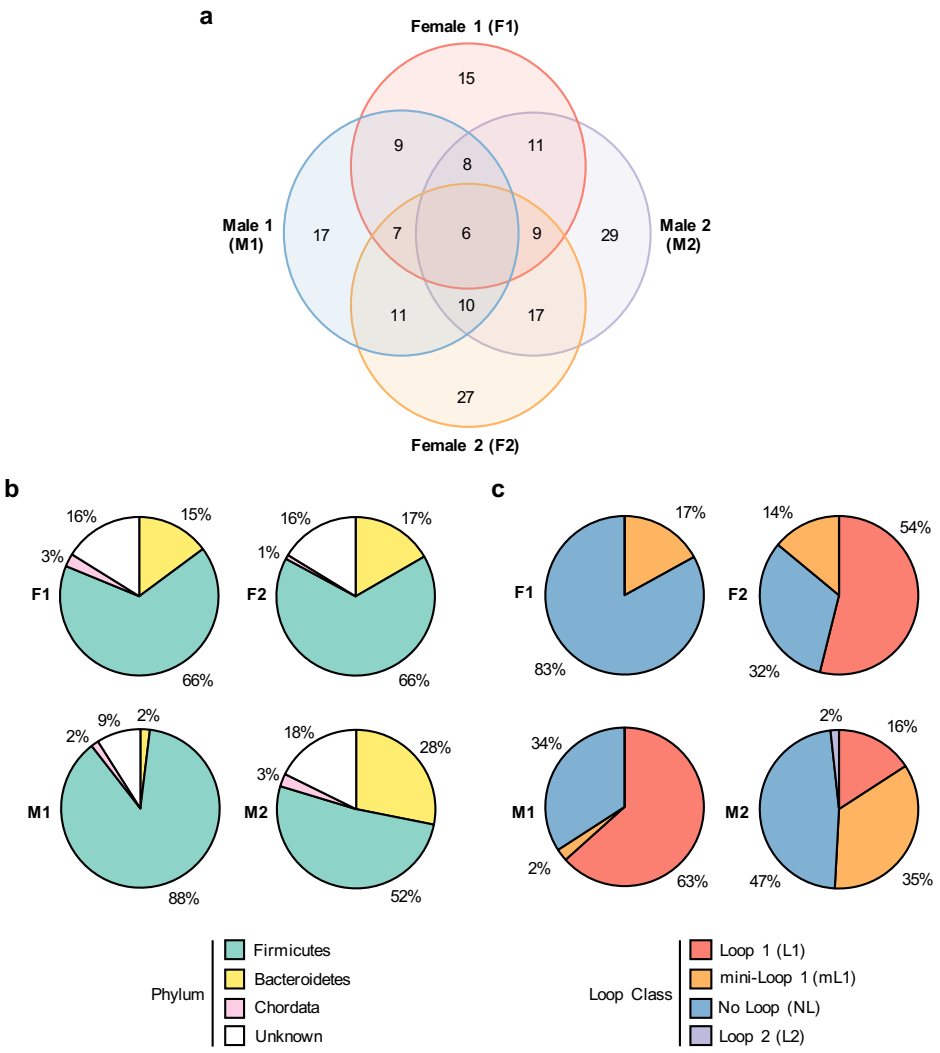


Figure 2.8 Interindividual variability in bacterial GUS composition. (**a**) Venn diagram of identified GUS protein groups from four individuals analyzed by this study. (**b**) Phylum-level and (**c**) structure-level analysis of GUS abundance from human fecal samples. Phyla-level composition information was calculated using peptide abundance information. Structure-level composition information was calculated using protein group abundance information. Human GUS abundance information was not included in structure-level composition analysis.

Cyclophellitol-based aziridine ABPs also target GH3 β-glucosidases

Because the human gut microbiota contains a diverse assortment of glycoside hydrolases (GHs), performing ABPP from fecal material is a veritable test of the selectivity of the GUS ABPs (*9*). Sequence analysis of the protein groups identified from human fecal extracts revealed a major off-target hit, GH3 β-glucosidases (**Figure 2.9a**). The GH3 β-glucosidases enriched are structurally similar but occupy two topologically distinct categories that we have termed "Type I" and "Type II". Manual docking analysis of the untagged ABP in structurally characterized GH3 β-glucosidases revealed favorable positioning of the catalytic nucleophile for attack of the aziridine ring (**Figure 2.9b**). Additionally, an arginine residue was also present (R538 and R50 in Type I and Type II, respectively) that may contact the carboxylic acid moiety of the probe, enabling recognition and subsequent processing of ABPs by GH3 β-glucosidases. We expressed and purified both a Type I and Type II β-glucosidase identified in the fecal samples and confirmed *in vitro* that they are labelled by high concentrations of ABP **3** (**Figure 2.9c**). Despite labelling of the GH3 β-glucosidase by GlcA-like aziridine probes, neither type of β-glucosidase processed *p*NP-GlcA, suggesting that off-target labelling of β-glucosidases is probably due to the reactive aziridine moiety of the GUS ABPs (**Figure 2.9d, e**).

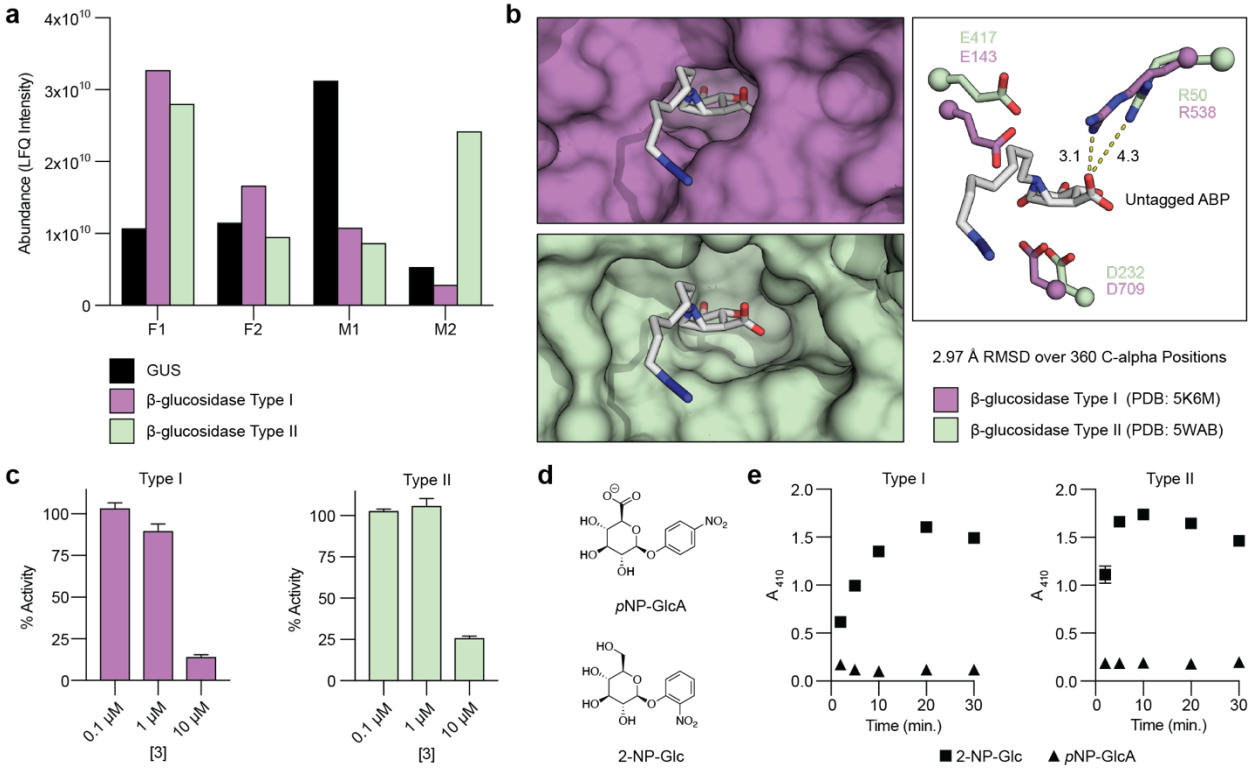


Figure 2.9 β-glucosidase is a specific off-target of GUS ABPs. (**a**) Protein abundance of GUS, Type I β-glucosidase, and Type II β-glucosidase identified from human fecal samples. (**b**) Conserved active sites of topologically distinct Type I (PDB: 5K6M) and Type II (PDB: 5WAB) β-glucosidases with the untagged ABP manually docked in PyMol. Distances are shown in Å. (**c**) Type I and Type II β-glucosidase inhibition by the biotin-ABP (**3**). (**d**) Chemical structures of 2-nitrophenyl β-D-glucopyranoside (2-NP-Glc) and *p*-nitrophenyl-β-D-glucuronide (*p*NP-GlcA). (**e**) *In-vitro* processing of 2-NP-Glc and *p*NP-GlcA by Type I and Type II β-glucosidases. All percent activity and rate values shown are mean values ± standard deviation using *N*=3 biological replicates.

Gut bacterial Loop 1 GUS enzymes are key mediators of SN-38 reactivation

After successfully identifying and quantifying bacterial GUS enzymes from human feces, we investigated whether we could identify the exact bacterial GUS enzymes responsible for SN-38 reactivation in the gut by integrating ABPP-enabled GUS abundance information with *ex vivo* SN-38-G processing data. We measured *ex vivo* SN-38-G hydrolysis by human fecal extracts, which revealed faster processing for F2 and M1 than F1 and M2 (**Figure 2.10a and 2.10b, Figure 2.11**). We found a strong correlation between Loop 1 (L1) GUS abundance and rate of SN-38-G hydrolysis when compared to total bacterial GUS abundance (**Figure 2.10c**). No correlation was found between either human GUS or other GUS structural classes and the rate of SN-38-G hydrolysis (**Figure 2.12**). We validated the correlation by assessing the catalytic efficiency of SN-38-G processing by a panel of purified GUS enzymes from various GUS structural classes and found that bacterial L1 GUS enzymes process SN-38-G most efficiently (**Figure 2.10d**). We also found that F2, M1, and M2 were abundant in L1 GUS enzymes that had sequence identities ≥ 90% to *E. eligens* GUS (*Ee*GUS, PDB: 6BJQ). We expressed and purified *Ee*GUS and found that it processed SN-38-G faster than all other examined GUS enzymes *in vitro*. A close examination of the crystal structure of *Ee*GUS reveals a hydrophobic active site pocket formed at the interface of two monomers in the L1 tetramer (**Figure 2.10e**) (*65*). The hydrophobic pocket formed by the oligomeric interface appears to optimally recognize hydrophobic small molecule glucuronides like SN-38-G. Taken together, correlation analysis between *ex vivo* processing data and ABPP-enabled GUS abundance data, further informed by *in vitro* enzyme kinetics and structural modeling, provides a molecular rationale for interindividual variation in SN-38 reactivation in human fecal samples, and identifies L1 GUS enzymes, particularly *Ee*GUS, as key molecular regulators of efficient SN-38-G reactivation.

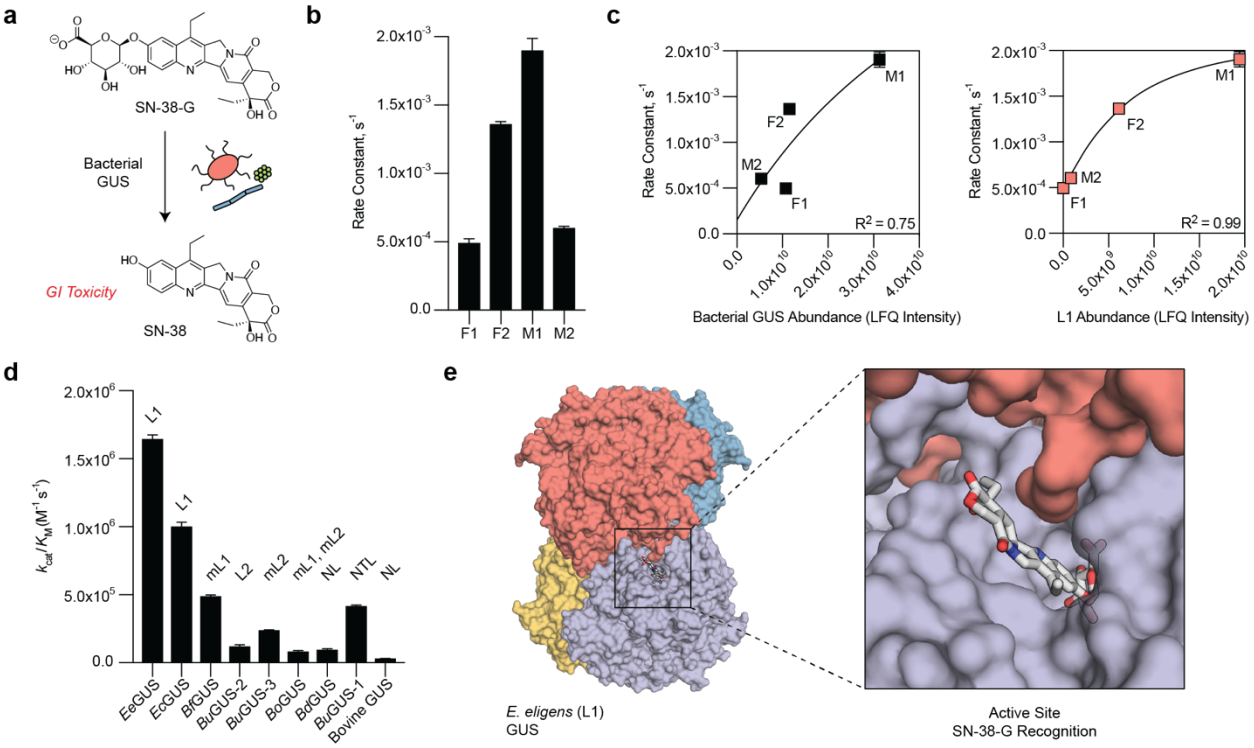


Figure 2.10 ABPP coupled with *ex vivo* processing data provides a molecular rationale for GUS-mediated SN-38 reactivation. (**a**) SN-38 glucuronide (SN-38-G) is the inactive metabolite of the topoisomerase I inhibitor irinotecan and is reactivated to SN-38 in the gut by bacterial GUS enzymes, resulting in acute, dose-limiting GI toxicity. (**b**) *Ex vivo* processing of SN-38-G by human fecal protein extracts. (**c**) Correlation analysis between total bacterial GUS abundance and Loop 1 (L1) GUS abundance against SN-38-G processing. (**d**) *In vitro* catalytic efficiencies of SN-38-G processing for a representative panel of GUS enzymes of different loop types. mini-Loop 1 (mL1); Loop 2 (L2); mini-Loop 2 (mL2); mini-Loop 1, mini-Loop 2 (mL1, mL2); No Loop (NL); N-Terminal Loop (NTL) (**e**) Quaternary structure of *E. eligens* GUS (*Ee*GUS, PDB: 6BJQ) with SN-38-G manually docked in PyMol.

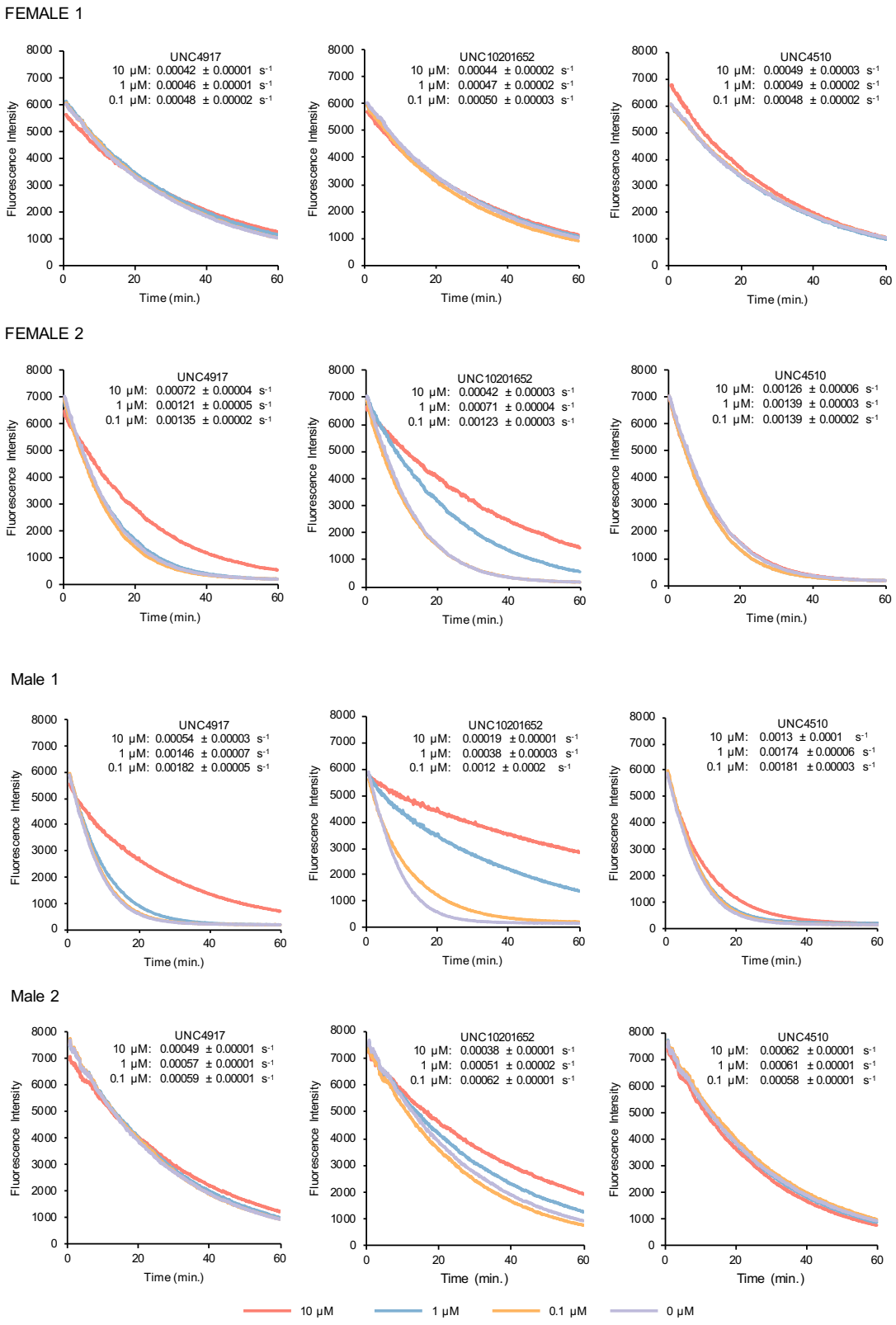


Figure 2.11 Progress curves showing SN-38-G processing and inhibition in fecal samples.

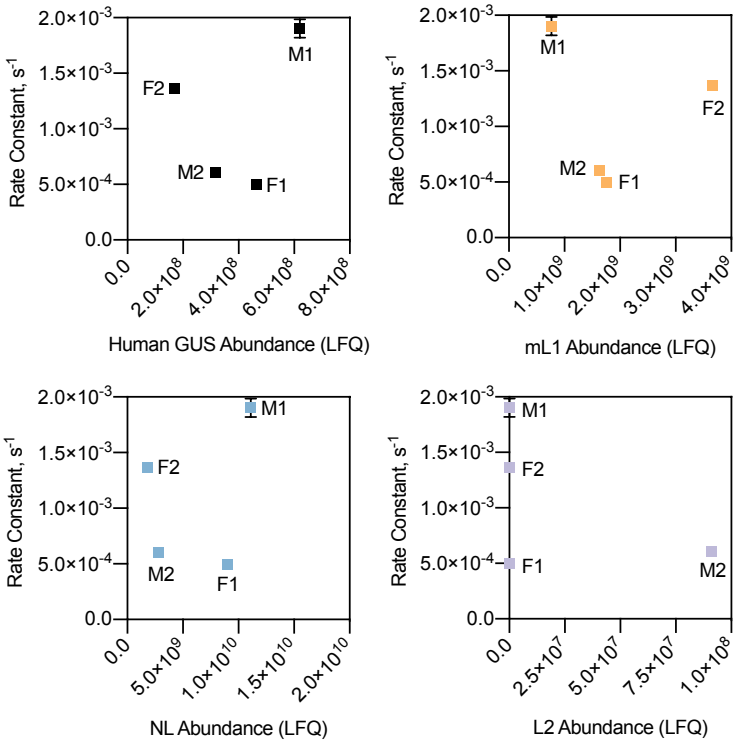


Figure 2.12 Correlation analysis of abundance data from various GUS structural classes against SN-38-G processing data. Correlation analysis between mini-Loop 1 GUS, No Loop GUS, Loop 2 GUS abundance, human GUS abundance and SN-38-G processing.

Piperazine containing small molecules inhibitors target gut bacterial Loop 1 GUS enzymes

Finally, we sought to extend these investigations to explain differential gut bacterial GUS inhibition. We have developed selective, potent, and non-lethal gut bacterial GUS inhibitors that block the reactivation of drug metabolites like SN-38-G (**Figure 2.13a**) (*63*, *64*). The piperazine moiety in both UNC4917 and UNC10201652 acts as a warhead that targets the catalytic machinery of bacterial GUS enzymes by intercepting the catalytic cycle (*65*). We find that SN-38-G processing was differentially inhibited in all four human fecal extracts using these GUS inhibitors (**Figure 2.13b**). Subsequent analyses reveal a strong correlation between inhibition and L1 GUS abundance while no correlation was observed for the other GUS structural classes, confirming previous work that UNC4917 and UNC10201652 act as L1-specific GUS inhibitors (**Figure 2.13c** and **Figure 2.14**) (*63*, *73*). Furthermore, we verified that UNC4510, a negative control analog of UNC10201652 that contains a methylated piperazine moiety, poorly inhibited SN-38-G processing for all GUS enzymes (*65*). These data show that L1-specific GUS inhibitors can block SN-38-G processing only in individuals whose fecal gut microbiota is highly abundant in L1 GUS enzymes.

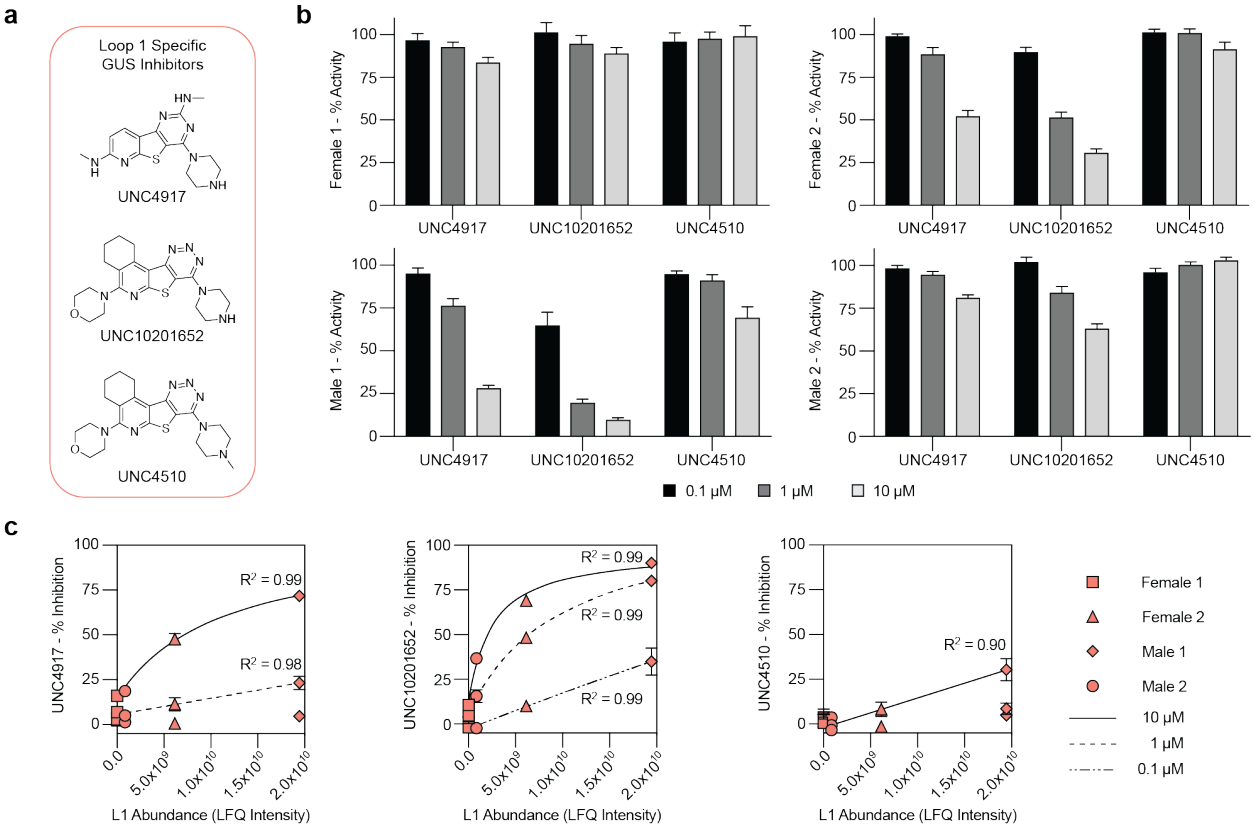


Figure 2.13 ABPP coupled with *ex vivo* processing data explains differential propensities for GUS inhibition. (**a**) Structures of L1 GUS inhibitors, UNC4917, UNC10201652, and the poor inhibitor, UNC4510 (negative control). (**b**) Inhibition of SN-38 reactivation in human fecal samples by selective bacterial GUS inhibitors. All percent activity values shown are mean values ± standard deviation using *N*=3 biological replicates. (**c**) Correlation analysis between L1 GUS abundance and inhibition data for each GUS inhibitor.

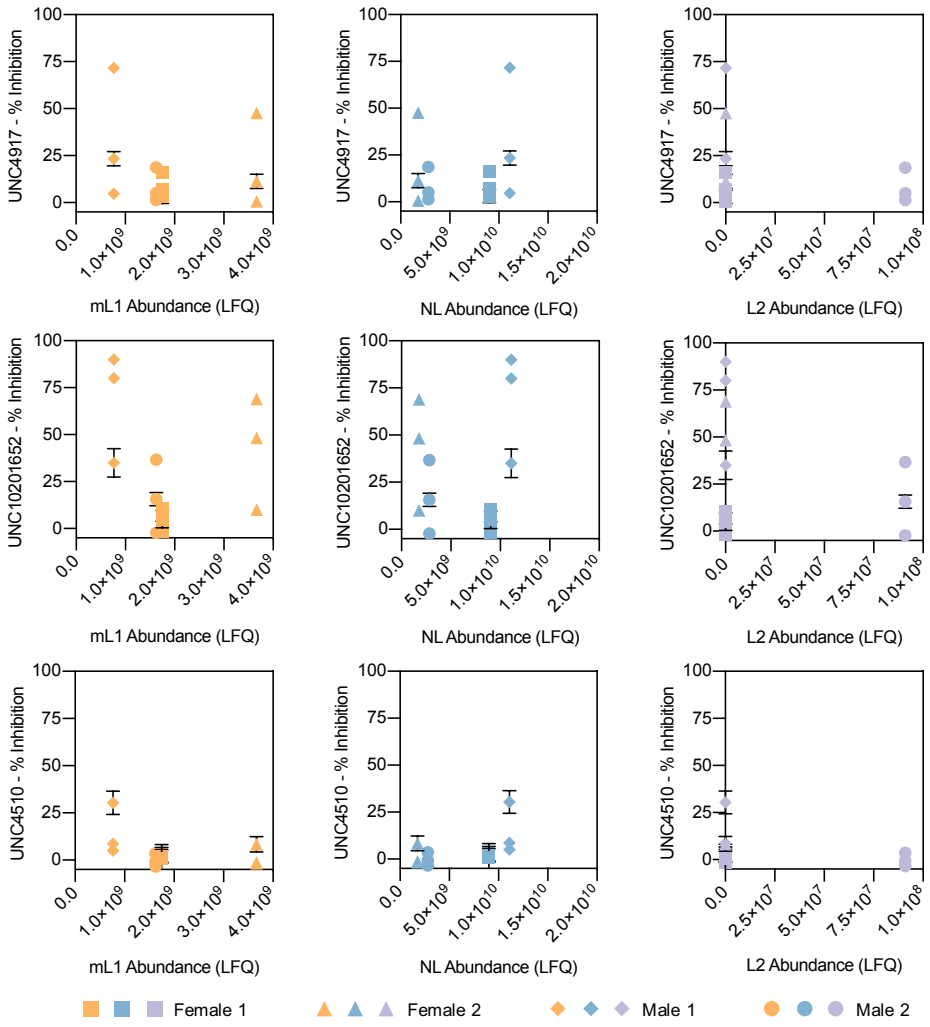


Figure 2.14 Correlation analysis of abundance data from various GUS structural classes against SN-38-G inhibition data. Correlation analysis between mini-Loop 1 GUS, No Loop GUS, Loop 2 GUS abundance, and inhibition data for Loop 1-specific GUS inhibitors, UNC4917, UNC10201652, and UNC4510.

DISCUSSION

Here we show that cyclophellitol-based epoxide and aziridine inhibitors and ABPs can target gut bacterial GUS enzymes. Using a combination of *in vitro* and in-gel assays, we find that **1–4** target structurally diverse GUS enzymes with varying potencies. The variation in GUS inhibition is likely due to differences both in oligomeric states and active site features of the bacterial GUS enzymes examined (**Figure 2.15**). For example, we observe more potent inhibition of *E. coli* GUS by the biotin-ABP (**3**) when compared to the unsubstituted aziridine inhibitor (**2**). Like *E. eligens* GUS, previous structural work has shown that *E. coli* GUS is a tetramer with a hydrophobic active site formed at the interface of its monomers (*70*, *73*). Thus, the increase in inhibition by ABP **3** compared to inhibitor **2** is likely due to hydrophobic interactions between the *E. coli* GUS active site and the nonpolar alkyl chain present in **3**. Furthermore, ABP **3** and **4** displayed notable differences in inhibition for all GUS enzymes. The Cy5-ABP (**4**) is weaker at inhibiting GUS enzymes than the biotin-ABP (**3**), and this is likely due to steric clashes between the bulky fluorophore group and the GUS enzymes examined.

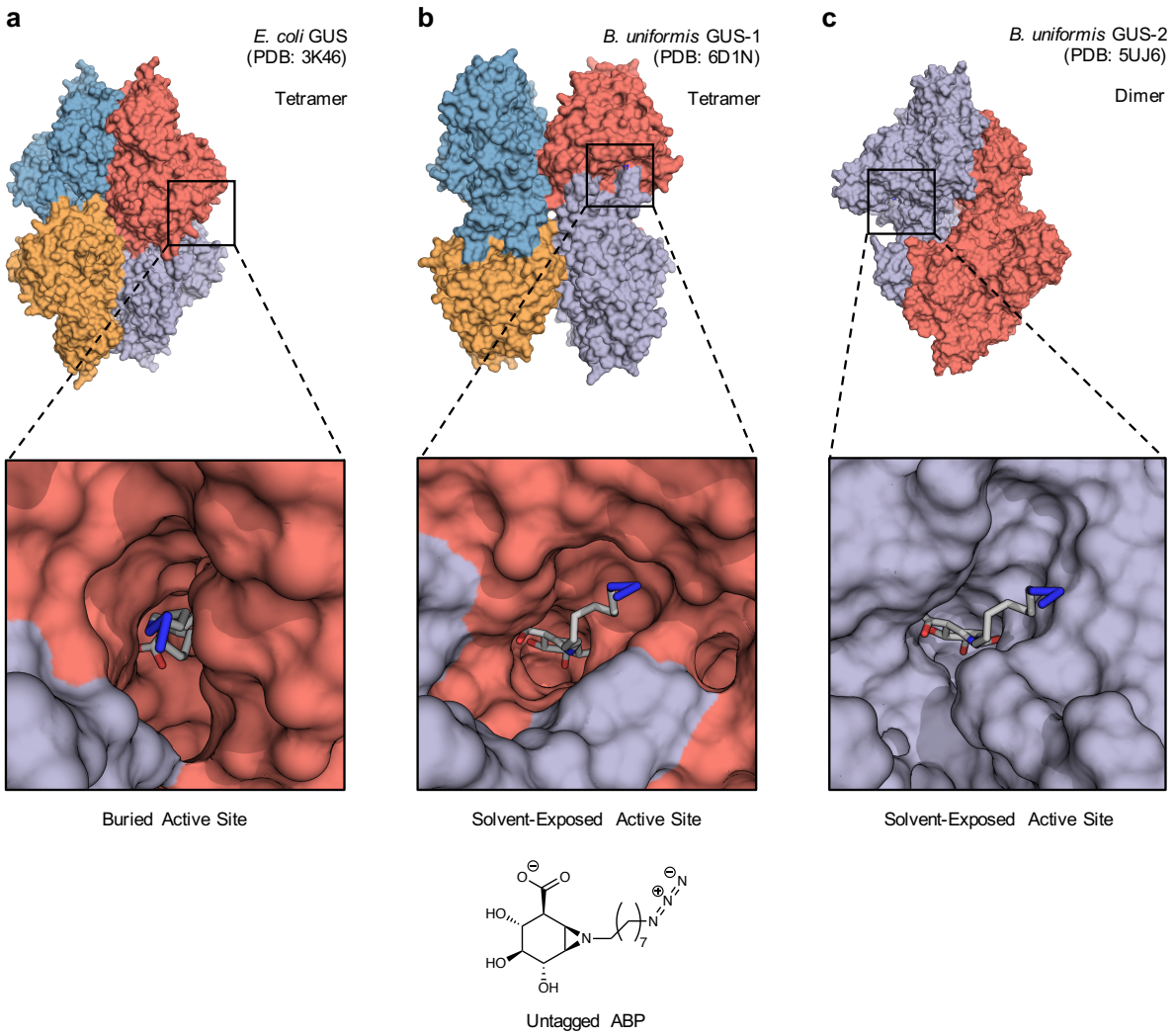


Figure 2.15 GUS ABPs target structurally diverse active sites. (**a**) *Ec*GUS tetramer, (**b**) *Bu*GUS-1 tetramer, and (**c**) *Bu*GUS-2 dimer with zoom-in of untagged ABP manually docked in active site in PyMol.

Most importantly, we show that gut microbial GUS enzymes can be identified and quantified from human feces using ABPP. We were interested in examining GUS sequence information obtained through our ABPP-enabled pipeline to better understand the structural diversity of GUS enzymes present in the gut microbiome and to correlate GUS structure to SN-38-G processing. By using powerful metaproteomic software tools like MetaLab (*37*) and Unipept (*81*), we also show that peptide MS data can be employed to obtain taxon information for GUS-producing bacterial species found in human feces. However, many protein groups could not be assigned to lower taxonomic ranks due to a lack of taxon-specific distinctive peptides. Thus, in the future, strategies that both increase peptide count and yield longer peptides for MS analysis should be explored to improve taxonomy assignment using ABPP. Metagenomic sequencing could be pursued to develop a sample-specific sequence database to query peptides, but this approach may be economically prohibitive (*27*). While other methods have coupled deep-sequencing with ABPs to uncover GUS-producing species (*40*), we provide evidence here that ABPP alone can be used to obtain a strong level of taxa information for GUS-producing bacterial species from human fecal samples.

An unexpected yet exciting finding from our investigation was the identification of GH3 β-glucosidases as an off-target hits. We identified two topologically distinct GH3 β-glucosidases as off-target hits of the GUS ABPs. Since ABPs sample enzyme function, we initially hypothesized that the identified GH3 β-glucosidases may process GlcA-containing substrates, but *in vitro* assays using *p*NP-GlcA revealed that these enzymes do not process glucuronides, and are in fact, off-target hits (**Figure 2.9d, e**). Further assessment of previously published GH3 β-glucosidase structures reveal a solvent exposed active site and an arginine residue that interacts with the carboxylic acid moiety of GlcA. These features combined with the highly reactive

nature of the aziridine moiety in the cyclophellitol-based ABP likely cause labelling of the GH3 β-glucosidases. The identification of only one class of off-target hits is remarkable given that the human gut microbiome is one of the most glycoside hydrolase rich environments found in nature (*9*), and further demonstrates that cyclophellitol-based GUS ABPs are incredibly precise and effective probes.

Integration of ABPP-enabled GUS abundance with *ex vivo* SN-38-G processing data enabled the identification of L1 GUS enzymes as the key molecular regulators of SN-38-G turnover. Importantly, this predictive correlation was validated by both *in vitro* enzyme kinetics and structural modeling. Although we have strongly correlated L1 GUS enzymes to SN-38-G processing, they are lead biomarkers that will need to be further characterized for clinical use. For example, the ABPP methodology outlined here does not examine the bacterial cell uptake of these glucuronide substrates. Further studies analyzing relevant gut bacterial isolates will be needed to assess cellular uptake of SN-38-G. Additionally, the gut microbiota contains hundreds of unique GUS enzymes, all of which are not encompassed by the four fecal samples used in this study. The strategy outlined here provides a foundation on which future proteomics and drug processing can be added to extant datasets to re-run correlation analyses and identify new biomarkers.

We also show that SN-38-G processing can be inhibited in complex metaproteomes using previously designed L1-specific GUS inhibitors and that GUS inhibition can be accurately predicted with probe-derived proteomics data. Interestingly, our data indicates that UNC10201652 is more potent than UNC4917 at inhibiting L1 GUS enzymes in fecal samples, a similar result found in a previous study (*65*). Structure-activity relationships can be conducted against a large assortment of GUS enzymes found in fecal samples using this strategy to identify

the inhibitor chemotypes that block GUS enzymes from reactivating drug glucuronides like SN-38-G. Coupling ABPP-enabled GUS abundance with *ex vivo* inhibition data can serve as a powerful strategy to conduct structure-activity relationships in a high-throughput manner. Since we have a limited understanding of enzyme-substrate pairs in the microbiome, we believe it is imperative that high precision gut bacterial inhibitors be developed in lieu of broad-spectrum drugs like antibiotics or inhibitors that target enzymes classes.

Recent work was published on a distinct GUS ABP composed of a GlcA warhead linked to a quinone methide leaving group at the anomeric position (*40*). The main difference between the quinone methide ABP and the cyclophellitol-based aziridine ABP employed here is target specificity. As noted by Wright and co-workers, the quinone methide ABP, once activated, has the potential to leave the enzyme active site and label off-target macromolecules (*40*). In contrast, the cyclophellitol-based aziridine ABP employed here reacts directly with the GUS active site in a mechanism-based fashion to form a covalent bond with the glutamate nucleophile, likely reducing the number of off-targets. Although labeling live bacteria with a quinone-methide ABP coupled with FACS sorting and 16S rRNA sequencing can give general taxa information on bacterial populations found in feces (*40*), it seems less suitable for sequence-level identification and quantification of active GUS enzymes from fecal supernatant due to the promiscuity of the activated quinone-methide leaving group.

In summary, we determined the composition and relative abundance of bacterial GUS enzymes from human fecal samples using ABPP. We utilized these data to identify the key modulators of SN-38 reactivation and to rationalize differential GUS inhibition across fecal samples. While we focused on SN-38-G metabolism in the present study, the combination of proteomics data and functional assays can be employed to pinpoint specific GUS enzymes

implicated in the reactivation of other drug glucuronides. Furthermore, proteomics-activity correlations provide a universal tool to identify a specific molecular target for any enzyme activity in the microbiome, an approach that is only limited and facilitated by the current set and continued development of ABPs that target gut bacterial enzymes (*39–41*, *79*, *82–84*). Together, the data gained from this ABPP approach enables the identification of potential gut bacterial drug targets for the molecular modulation of the gut microbiota and can be employed to reveal highly precise biomarkers for possible diagnostic development in the era of personalized medicine.

MATERIALS AND METHODS

Protein expression, purification, and site-directed mutagenesis

All GUS enzymes were expressed and purified as previously described (*64*, *70*, *73*, *77*). *Bu*GUS-1^{NxK} and *Bu*GUS-2^{NxK} mutants were generated, expressed, and purified as previously described (*64*). Type I and Type II β-glucosidases subcloned into pLIC-His vectors were purchased from BioBasic and expressed and purified as previously described (*64*). Sequence information for purified β-glucosidases can be found in online (*85*). Briefly, all proteins were expressed with a *N*-terminal 6x histidine tag and subsequently purified using a Ni-NTA HP column (GE Healthcare). Additional purification was performed using a HiLoad 16/60 Superdex 200 gel filtration column. Proteins were eluted and aliquots were flash frozen in liquid nitrogen and stored at –80°C until further use.

Protein crystallography

Crystals of *Bu*GUS-2 bound to the unsubstituted cyclophellitol-based aziridine inhibitor (**2**) were produced via the hanging-drop vapor diffusion method. *Bu*GUS-2 at 10 mg mL^{-1} was preincubated with an equivalent amount of inhibitor **2** prior to addition into the crystalline solution. Crystals were formed by incubating ligand bound *Bu*GUS-2 in 0.2 M KCl and 18% PEG 3350. The crystals were cryoprotected using 0.2 M KCl and 18% PEG 3350 in 20% glycerol. Diffraction data for all crystals were collected on the 23-ID-B beamline at GM/Ca-CAT (Advanced Photon Source, Argonne National Laboratory). Refinements and ligand generation were carried out in Phenix, and ligand fitting was performed in Coot (*86*). Final coordinates and structure factors have been submitted to the RCSB and the assigned accession code 6NZG for ligand bound structure.

Animal study design

All animal studies were approved by the University of North Carolina Institutional Animal Care and Use Committee (IACUC), according to Care and Use of Laboratory Animals guidelines set by the National Institutes of Health.

Germ-free wild-type C57/BL6J mice were bred and maintained in-house at the National Gnotobiotic Rodent Resource Center (NGRRC; University of North Carolina, Chapel Hill, NC). Mice were housed in Green Line cages (Techniplast). At 8–10 weeks of age, mice were colonized by oral gavage and rectal swabbing with viable WT *E. coli* MG1655 or isogenic ΔGUS mutant that were cultured overnight in lysogeny broth in anaerobic conditions as described previously (REF: PMID 29269393). Colonization was monitored by quantitative plating onto brain heart infusion agar plates of serial dilutions of freshly collected feces. Plates

were incubated for 24 hours under aerobic conditions at 37°C, and colonies were enumerated. Freshly voided stools were collected aseptically into sterile tubes, snap frozen, and stored at –80°C until further analysis.

Mouse fecal extract

Mouse fecal pellets were collected and stored at –80°C until further use. Fecal extracts were created as previously described (*63*). In brief, 1–2 pellets were rehydrated with 350 µL cold extraction buffer (pH 6.5, 25 mM HEPES, 25 mM NaCl with Roche cOmplete™ protease inhibitor cocktail) containing autoclaved 0.7 mm garnet beads (Omni International). The mixture was vortexed to break up dense, fibrous material. Bacterial cells were lysed using a Tissuelyzer II (Qiagen) for 2 min. at 30 Hertz. The resulting homogenate was sonicated for 2 min. The sonication was repeated after mixing the homogenate by inversion. After centrifugation (13,000 x *g*, 10 min., 4°C), the supernatant was decanted. The total protein concentration was calculated using a standard Bradford Assay protocol. The mouse fecal extract was aliquoted and snap frozen using liquid nitrogen. The aliquots were stored at –80°C until further use.

Human fecal extract

Human fecal samples were purchased from a commercial vendor, BioIVT, and stored at –80°C until further use. Approximately 5 g of thawed fecal material in a solution containing 25 mL of cold extraction buffer (pH 6.5, 25 mM HEPES, 25 mM NaCl with Roche cOmplete™ protease inhibitor cocktail) and 500 mg of autoclaved garnet beads was vortexed vigorously to break up dense, fibrous material. The suspended sample was centrifuged at low speed (300 x *g*, 5 min., 4°C) to separate out any insoluble fecal material. After decanting the microbial supernatant, an

additional 25 mL of cold extraction buffer was added to the remaining fibrous material and the extraction process was repeated. The combined microbial supernatant (~40–45 mL) was centrifuged at low speed to remove any remaining insoluble debris. This process was repeated with the decanted microbial supernatant. The microbial supernatant was ultrasonicated for 1.5 min. while on ice. The lysate was mixed by inversion and the sonication repeated. The lysed cells were centrifuged at high speed (17,000 x *g*, 20 min., 4°C) to remove cellular debris. The decanted lysate was concentrated, and metabolites were removed by buffer exchanging with fresh extraction buffer. The concentration of total protein in the fecal extract was calculated using a standard Bradford Assay protocol. The human fecal extract was aliquoted and snap frozen using liquid nitrogen. The aliquots were stored at –80°C until further use.

GUS inhibitors and activity-based probes (ABPs)

Cyclophellitol-based inhibitors **1** and **2**, and ABPs **3** and **4** were synthesized and purified as previously described (*79*).

Fluorescence labelling

Select recombinant GUS enzymes were diluted to the appropriate concentration in buffer and pre-incubated for 5 min. at 37°C prior to the addition of Cy5-ABP (**4**). The final reaction volume was 50 µL containing 30 µL water, 10 µL buffer (pH 6.5, 25 mM HEPES, 25 mM NaCl, and 1% DMSO, final), 5 µL GUS (1 µM, final), and 5 µL ABP **4** (100 nM, final). For the heat denatured *Ec*GUS control, *Ec*GUS was incubated at 95°C for 5 min. prior to the addition of ABP **4**. After the addition of ABP **4**, reaction mixtures were incubated at 37°C for 1 hr. and denatured with 50 µL gel loading buffer (pH 6.8, 50 mM Tris-HCl, 100 mM DTT, 2% SDS, and 10% glycerol,

final) at 95°C for 5 min. Samples were cooled on ice, run on a 10% acrylamide gel, and imaged using a fluorescence scanner (Amersham Typhoon) with an excitation wavelength of 649 nm and emission wavelength of 670 nm. Gels were subsequently stained with coomassie blue and imaged.

Mouse fecal extracts from *E. coli* mono-associated and germ-free mice were incubated with Cy5-ABP (**4**) at 37°C for 1 hr. The final reaction volume was 20 µL containing 12 µL water, 4 µL buffer (pH 6.5, 25 mM HEPES, 25 mM NaCl, final), 2 µL mouse fecal extract (0.1 mg mL^{-1}, final), and 2 µL ABP **4** (100 nM, final). For the heat denatured control, *E. coli* mono-associated fecal extract was incubated at 95°C for 5 min. prior to the addition of ABP **4**. For label blocking studies, *E. coli* mono-associated fecal extract was pre-incubated with various concentrations of D-glucaro-1,4-lactone prior to the addition of ABP **4**. Samples were denatured with 20 µL gel loading buffer, ran on gel, and processed as stated above.

Full gel images can be found in **Figure 2.16**.

Figure 1c – Fluorescence Scan

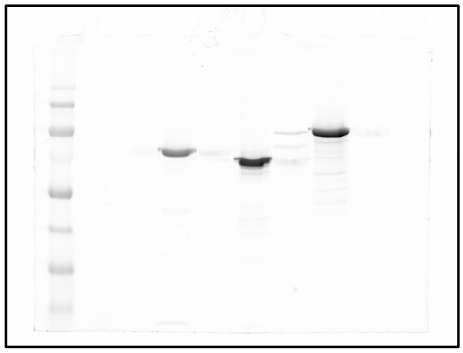


Figure 1d – Fluorescence Scan

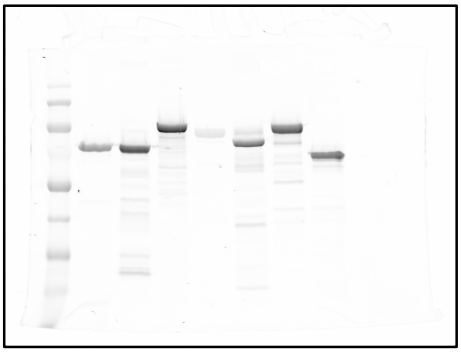


Figure 1c – Coommassie Stain

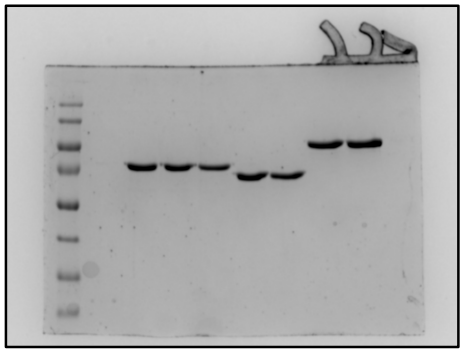
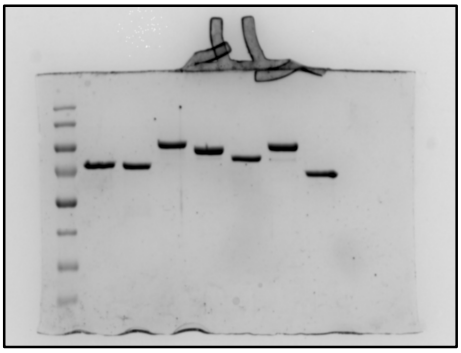


Figure 1d – Coommassie Stain



Suppl. Figure 5 – Fluorescence Scan

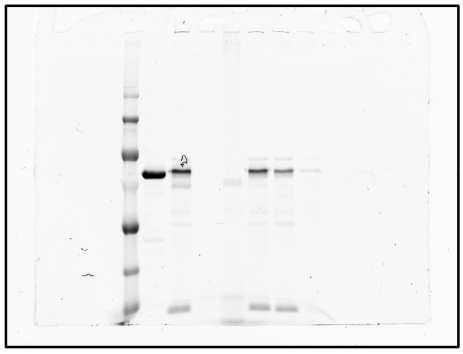


Figure 2.16 Full gel images done in study.

In vitro GUS activity and inhibition

Apparent IC_{50} values for GUS inhibition by compounds **1–4** were calculated using an endpoint format (*65*). Select GUS enzymes were pre-incubated with a range of compound concentrations at 37°C for 30 min. prior to initiating the reaction with the fluorogenic substrate, 4-methylumbelliferone glucuronide (4-MUG). The final reaction volume was 50 µL containing 25 µL water, 10 µL buffer (pH 6.5, 25 mM HEPES, 25 mM NaCl, and 2% DMSO, final), 5 µL enzyme (15 nM, final), 5 µL compound (varying concentrations), and 5 µL 4-MUG (900 µM, final). After the addition of substrate, reaction mixtures were incubated at 37°C for 1 hr. and then quenched with 50 µL 0.2 M sodium carbonate. Fluorescence intensities were measured at an excitation wavelength of 350 nm and an emission wavelength of 450 nm (PHERAStar BMG Labtech). End point fluorescence intensities were converted to percent inhibition as previously described (*65*). Percent inhibition values were subsequently plotted against the log of compound concentration and fit with a four-parameter logistic function in GraphPad Prism to determine the concentration at which 50% inhibition (IC_{50}) is observed.

Kinetic parameters were determined for GUS inhibition by compounds **1–4** as previously described with some modification (*65*, *79*). Briefly, select GUS enzymes were pre-incubated in buffer at 37°C for 5 min. prior to initiating the reaction with the addition of both compound and 4-MUG. The final reaction volume was 50 µL containing 25 µL water, 10 µL buffer (pH 6.5, 25 mM HEPES, 25 mM NaCl, and 2% DMSO, final), 5 µL enzyme (5 nM, final), 5 µL compound (varying concentrations), and 5 µL 4-MUG (900 µM, final). After addition of substrate and compound, reactions were monitored continuously at 37°C at an excitation wavelength of 350 nm and an emission wavelength of 450 nm (PHERAStar BMG Labtech). The first-order rate constant, k_{obs}, was calculated as previously described (*65*). We applied a one-step kinetic scheme

to fit in Excel the resulting k_{obs} vs. [compound] using a linear function, which gives the apparent inhibition parameter k_i/K_I' as the slope. k_i/K_I' parameter was converted to k_i/K_I using the following equation, $K_I' = K_I(1 + [S]/ K_M)$, where [S] = 900 μM and K_M = 64 μM, 25 μM, and 80 μM for *Ec*GUS, *Bu*GUS-1, and *Bu*GUS-2, respectively.

Catalytic efficiencies for SN-38-G processing by GUS enzymes were calculated using a continuous read format. Select GUS enzymes were pre-incubated in buffer at 37°C for 5 min. prior to initiating the reaction with the fluorogenic substrate, SN-38-G. The final reaction volume was 25 μL containing 5 μL water, 12.5 μL buffer (pH 6.5, 25 mM HEPES, and 25 mM NaCl, final), 2.5 μL enzyme (15 nM, final), and 5 μL SN-38-G (varying concentrations). After addition of substrate, reactions were monitored continuously at 37°C at an excitation wavelength of 230 nm and an emission wavelength of 420 nm (Tecan Infinite M1000 Pro). Initial velocities from the resultant progress curves were fitted using a linear regression with a custom MATLAB program, and k_{cat}/K_M was determined in Excel.

In vitro β-glucosidase activity and inhibition

Inhibition of purified Type I and Type II β-glucosidases by biotin-ABP (**3**) was determined using an end-point assay format. The β-glucosidases were pre-incubated with ABP **3** at 37°C for 30 min. prior to initiating the reaction with the colorimetric substrate, 2-nitrophenyl β-D-glucopyranoside (2-NP-Glc). The final reaction volume was 50 μL containing 25 μL water, 10 μL buffer (pH 6.5, 25 mM HEPES, and 25 mM NaCl, final), 5 μL ABP **3** (varying concentrations), 5 μL enzyme (200 nM and 100 nM, final, for Type I and Type II, respectively), and 5 μL 2-NP-Glc (900 μM, final). After addition of substrate, reaction mixtures were

incubated at 37°C for 30 min. and then quenched with 0.2 M sodium carbonate. Absorbance was measured at 410 nm (PHERAStar BMG Labtech).

Activity against *p*-nitrophenyl β-D-glucuronide (*p*NP-GlcA) and 2-NP-Glc hydrolysis by Type I and Type II β-glucosidases was determined using a continuous read format. Both Type I and Type II β-glucosidases were pre-incubated in buffer at 37°C for 5 min. prior to initiating the reaction with either *p*NP-GlcA or 2-NP-Glc. The final reaction volume was 50 µL containing 30 µL water, 10 µL buffer (pH 6.5, 25 mM HEPES, and 25 mM NaCl, final), 5 µL enzyme (200 nM and 100 nM, final, for Type I and Type II, respectively), and 5 µL *p*NP-GlcA or 2-NP-Glc (900 µM, final). After addition of substrate, reactions were stopped at various time points by quenching with 0.2 M sodium carbonate. Absorbance was measured at 410 nm (PHERAStar BMG Labtech).

Ex vivo GUS activity and inhibition

Inhibition of SN-38-G hydrolysis in human fecal extracts was determined using a continuous read format. Inhibitors were diluted to the appropriate concentrations and pre-mixed with human fecal extract prior to initiating the reaction with the fluorogenic substrate, SN-38-G. The final reaction volume was 25 µL containing 12.5 µL water, 5 µL buffer (pH 6.5, 25 mM HEPES, 25 mM NaCl, and 1.3% DMSO, final), 2.5 µL human fecal extract (0.1 mg mL^{-1}, final), 2.5 µL inhibitor (various concentrations), and 2.5 µL substrate (15 µM, final). Reaction mixtures were pre-incubated with inhibitor at 37°C for 5 min. prior to the addition of substrate. After addition of substrate, reactions were monitored continuously at 37°C with an excitation wavelength of 230 nm and an emission wavelength of 420 nm (Tecan Infinite M1000 Pro). The first order rate

constant, k (s^{-1}), was obtained by fitting resulting progress curves using an exponential decay function in MATLAB (**Figure 2.11**).

Proteomics

Human fecal extracts (3.5 mg) were incubated at 37°C for 60 min. with either 10 μM pre-clicked, biotin-ABP (**3**) or 10 μM biotin only in 500 uL (pH 6.5, 25 mM HEPES, 25 mM NaCl, and 1% DMSO, final, containing Roche cOmplete™ protease inhibitor cocktail). Reactions were quenched by adding 125 μL 10% SDS and heating at 95°C for 5 min. After cooling on ice, the samples were washed 3 times with 0.05% SDS buffer (pH 6.5, 25 mM HEPES, and 25 mM NaCl) using 1.5 mL Amicon 10K cutoff spin concentrators to remove unreacted probe. The samples were centrifuged at 14,000x*g* for 5 min. at 4°C between wash steps. The volume in each sample was adjusted to 1 mL using 0.05% SDS buffer (pH 6.5, 25 mM HEPES, and 25 mM NaCl). To each sample, 15 μL streptavidin sepharose beads (GE) were added and incubated at room temperature for 60 min. The beads were subsequently washed with 300 μL 0.1% SDS buffer (pH 6.5, 25 mM HEPES, 25 mM NaCl), 3 times with 300 μL buffer (pH 6.5, 25 mM HEPES, 25 mM NaCl), and 3 times with 300 μL 50 mM NH_4HCO_3. The samples were centrifuged (400x*g*, 2 min., 4°C) between wash steps. The beads in each sample were re-suspended in 100 μL 50 mM NH_4HCO_3 and stored at –20°C until further analysis.

Proteins were eluted using 0.5% RapiGest (Waters; 18600861) in 50 mM NH_4HCO_3 and reduced with dithiothreitol (DTT) at 65°C for 30 min. Proteins were alkylated using 2-chloroacetamide (Acros Organics; 148415000) for 20 min. at room temperature in the dark. Beads were pelleted by centrifugation (200x*g*, 2 min., room temperature). The supernatant was transferred to a new tube and trypsinized overnight for 18 hr. at 37°C with 2.5 μg of trypsin

(Promega; V511C). RapiGest surfactant was quenched using 250 mM HCl for 45 min. at 37°C. Samples were then concentrated to 100 µL using a speedvac followed by C18 desalting columns in accordance with the manufacturer's protocols (ThermoScientific; 89870). Samples were then concentrated using a speedvac and resolubilized in 100 µL of LC-Optima MS grade water (Thermo; W7SK). Ethyl acetate (Thermo; E196SK) extraction followed by speedvac was performed to remove residual detergents. Peptides were quantified and normalized using the Pierce QFP assay (Thermo; 23290) in accordance with the manufacturer's protocol.

Reverse-phase nano-high-performance liquid chromatography (nano-HPLC) coupled with a nanoACQUITY ultraperformance liquid chromatography (UPLC) system (Waters Corporation; Milford, MA) was used to separate trypsinized peptides. Trapping and separation of peptides were performed in a 2 cm column (Pepmap 100; 3-m particle size and 100-Å pore size), and a 25-cm EASYspray analytical column (75-m inside diameter [i.d.], 2.0-m C18 particle size, and 100-Å pore size) at 300 nL/min and 35°C, respectively. Analysis of a 60-min. gradient of 2% to 25% buffer B (0.1% formic acid in acetonitrile) was performed on an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific). The ion source was operated at 2.4 kV and the ion transfer tube was set to 300°C. Full MS scans (350-2000 m/z) were analyzed in the Orbitrap at a resolution of 120,000 and 1e6 AGC target. The MS2 spectra were collected using a 1.6 m/z isolation width and were analyzed either by the Orbitrap or the linear ion trap depending on peak charge and intensity using a 3 s TopSpeed CHOPIN method (*87*). Orbitrap MS2 scans were acquired at 7500 resolution, with a 5e4 AGC, and 22 ms maximum injection time after HCD fragmentation with a normalized energy of 30%. Rapid linear ion trap MS2 scans were acquired using an 4e3 AGC, 250 ms maximum injection time after CID 30 fragmentation. Precursor ions were chosen based on intensity thresholds (>1e3) from the full scan as well as on charge states

(2–7) with a 30-s dynamic exclusion window. Polysiloxane 371.10124 was used as the lock mass. All proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD014864 (*88*).

Raw LC/MS data processing

Peptides and protein groups were identified by an iterative database strategy within MetaLab (version 1.1.1) (*37*), which used MaxQuant (version 1.6.2.3) (*38*). The database search was performed using the integrated reference catalog of the human gut microbiome (IGC) database (*29*) combined with the UniProtKB/Swiss-Prot human sequence database (downloaded 1 Feb. 2017) (*89*) with a total of 9,920,788 sequences. Search parameters were: static carbamidomethyl cysteine modification, specific trypsin digestion with up to two missed cleavages, variable protein N-terminal acetylation and methionine oxidation, match between runs, and label-free quantification (LFQ) with a minimum ratio count of 2. Protein identifications were filtered for a false discovery rate (FDR) of 1%, and potential contaminants and decoys were removed.

Identification of GUS enzymes and β-glucosidases from IGC and UniProt databases

GUS enzymes and β-glucosidases in the IGC and UniProt (SwissProt and TrEMBL) databases were identified by pairwise alignment to representative proteins. Candidate sequences were accepted if they passed a sequence identity threshold and contained conserved residues. For GUS enzymes, ≥ 28% identity was required with at least one of four representative proteins: *Escherichia coli* (*Ec*GUS, UniProt: P05804), *Clostridium perfringens* (*Cp*GUS, UniProt: Q8VNV4), *Streptococcus agalactiae* (*Sa*GUS, UniProt: Q8E0N2), and *Bacteroides fragilis* (*Bf*GUS, PDB: 3CMG). Additionally, all conserved residues had to be present and correctly

aligned to the representative protein that passed the identity threshold. The conserved residues were: *Ec*GUS E413, E504, N566, K568; *Cp*GUS E412, E505, N567, K569; *Sa*GUS E408, E501, N563, K565; and *Bf*GUS E395, E476, N547, K549. For Type I β-glucosidases, ≥ 25% identity was required with GH3 β-glucosidase from the cow rumen metagenome (PDB: 5K6M) with conserved residues E143, R597, K630, H631, D709. For Type II β-glucosidases, ≥ 26% identity was required with GH3 β-glucosidase from *Bifidobacterium adolescentis* (PDB: 5WAB) with conserved residues R120, K153, H154, D232, E417. For E417, an exact alignment was not required, rather an E had to be within ± 4 residues, including gaps. Pairwise alignment was performed by EMBOSS Stretcher with parameters gapopen=1 and gapextend=1 (*90*). Sequence identity thresholds (i.e., 28%, 25% and 26%) were chosen by selecting the smallest value for which no human proteins (except GUSB) were accepted (**Figure 2.7**).

Annotation of GUS loops

GUS loop classes were determined by multiple sequence alignment (MSA) of all GUS enzymes identified in the IGC database along with representative proteins. An initial MSA for IGC GUS enzymes was created using Clustal Omega with parameters --full, --full-iter, --iter=10. To determine Loop 1 and Loop 2 categories, *Ec*GUS (Uniprot: P05804) was aligned to the initial MSA using the same Clustal Omega parameters. Criteria for each class are defined by Pollet *et al* (*77*). To determine the N-Terminal Loop (NTL) class, *Bu*GUS-1 (PDB: 6D1N) was aligned to the initial MSA as before. NTL criteria are defined by Pellock *et al* (*64*).

GUS-specific taxonomy quantification

The relative GUS abundance of a taxon was defined as the summed intensities of GUS peptides distinct to the taxon. GUS peptides were all identified peptides that are present in a GUS protein in the UniProt database. To determine a peptide's distinct taxon, UniProt protein entries containing the peptide were found using UniPept's pept2prot program (*81*). Since most of the proteins identified by ABPP were GUS enzymes and β-glucosidases, we rationalized that the entries could be restricted to only those that belonged to these classes by pairwise alignment. Proteins were further filtered out if their taxon was "uncultured bacterium", "uncultured organism", "human gut metagenome" or a metazoan other than *homo sapiens*. The taxa for these proteins were then input into UniPept's taxa2lca program to determine the least common ancestor of the taxa, which is the most specific taxon for which the peptide is distinct.

Taxonomy identification for protein groups

A protein group's taxon was defined as the least common taxon that contained all the unique and razor peptides of the protein group. For each protein group, its razor and unique peptides were mapped to UniProt protein entries and full taxonomies using pept2prot and filtered as described in the taxonomy quantification section. To be considered the least common taxon, it must be the only taxon at its rank to contain all the peptides.

GUS correlation analyses

All correlation analyses were performed in GraphPad Prism by fitting with a one phase decay function.

CHAPTER 3: INHIBITION OF GUT BACTERIAL β-GLUCURONIDASES BY FDA-APPROVED DRUGS

INTRODUCTION

The human gut microbiome contains on the order of 1,000,000,000,000 bacterial cells, which is roughly equal to the number of human cells that comprise our body (*12*). The large reservoir of microbes living in our distal gut contains a plethora of enzymes that are important for both digestion and gut homeostasis (*12*). Gut bacterial enzymes are also capable of transforming a wide array of small molecule metabolites derived from the host (*91*), diet (*92*, *93*), and ingested drugs (*92*). The products of these biotransformation reactions can drastically impact human health as well as therapeutic outcomes. For example, gut microbiota-mediated metabolism of host-derived metabolites like bile acids and other endobiotics plays a role in the normal homeostasis of these compounds (*94*, *95*). Additionally, gut microbiota-driven transformation of small molecule drugs has been linked to drug toxicity and altered drug efficacy (*55*, *57*, *70*) (see **Chapter 1**).

A class of gut bacterial enzymes that is inextricably linked to digestion and drug efficacy are gut bacterial β-glucuronidases (GUS) (*70*, *73*). As discussed in **Chapter 2**, drugs and drug metabolites (*e.g.*, SN-38 and non-steroidal anti-inflammatory drugs) have the potential to be glucuronidated during Phase II drug metabolism by UDP glucuronosyl transferases (UGT) in hepatocyte cells (*62*). Appendage with GlcA increases the overall polarity and molecular weight of the drug which makes it easier for the body to dispose via either urinary or fecal excretion (*72*). If the GlcA-conjugated drug is excreted into the gastrointestinal (GI) tract, then gut

bacterial GUS enzymes have the potential to cleave the GlcA tag which results in the reactivation of the drug or drug metabolite. Individual GlcA units are further metabolized in the Entner-Doudoroff pathway to generate pyruvate, which enters the citric acid cycle and ultimately results in energy production for the microbe (*65*). Reactivation of drugs and associated metabolites by GUS enzymes in the gut has been linked to drug toxicity and altered drug efficacy; hence, considerable effort has been placed into creating GUS inhibitors to reduce GUS-mediated drug reactivation (*70*, *73*, *95*).

An extensive catalogue of GUS inhibitors has been reported (*96*). A commonly used gut bacterial GUS enzyme to screen for new GUS inhibitors is *Escherichia coli* GUS (*Ec*GUS). Among the growing list of GUS inhibitors, the synthetic compound **UNC10201652** is the most potent against *Ec*GUS (*96*). A detailed study of **UNC10201652** by Pellock and co-workers showed that it is a slow-binding, substrate-dependent inhibitor of GUS enzymes (*65*). Importantly, a combination of x-ray crystallography, mass spectrometry, and *in vitro* kinetic assays show that the secondary piperazine amine in the **UNC10201652** scaffold intercepts the glycosyl-enzyme catalytic intermediate during the hydrolysis of small molecule glucuronides (*65*). The inhibitor forms a covalent bond with a GlcA in the active site which results in near irreversible binding. Furthermore, structure-activity relationship (SAR) studies indicate that the secondary piperazine amine is necessary for potent inhibition of GUS enzymes by **UNC10201652**.

Since the secondary piperazine amine was found to drive the interception of the glycosyl-enzyme catalytic intermediate, Pellock and co-workers hypothesized that approved drugs containing either a secondary piperazine or piperadine amine will also inhibit GUS enzymes (*65*). Indeed, the team found using a combination of *in vitro* and cell-based assays that the tested

drugs in the study containing a secondary piperazine or piperadine amine inhibits *Ec*GUS with IC_{50} values ranging from 0.5 – 10 μM (*65*). While the potency of inhibition against *Ec*GUS for the examined drugs is 5 – 100 fold lower when compared to **UNC10201652** (IC_{50} = 0.117 ± 0.005 μM), the small intestine and colon concentrations of many approved drugs are on average in the mid to high micromolar range, which is high enough for these drugs to block the activity of GUS enzymes (*65*).

Recent reports indicate that GUS enzymes mediate the reactivation of endobiotics such as estrogen and serotonin in the GI tract (*91*, *95*). Like drugs, endobiotics are glucuronidated in hepatocytes and excreted into the GI tract via the biliary ducts. In the GI tract, gut bacterial GUS enzymes can cleave the GlcA tag which results in the reactivation of endobiotics. More importantly, the untagged endobiotic can be reabsorbed via enterohepatic recirculation. Inhibition of GUS-mediated metabolism of complex carbohydrates and endobiotics by approved piperazine and piperadine containing drugs can potentially interfere with gut homeostasis and hormone levels in the body which further highlights the importance of examining these approved drugs for GUS inhibition.

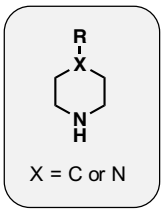
As discussed in **Chapter 2**, gut bacterial GUS enzymes from different gut bacterial species vary widely in structure. In addition, individuals differ in the type (*i.e.*, bacterial species source) and abundance of GUS enzyme that they harbor (*85*). Pellock and co-workers only tested one GUS structural category against a small set of piperazine and piperadine containing approved drugs (*65*). Here, we survey all known drugs containing piperazine and piperadine moieties for GUS inhibition. Importantly, we also examine inhibition against a structurally diverse library of GUS enzymes. Using a combination of *in vitro* kinetic assays, mass spectrometry, and x-ray crystallography, we also interrogate and validate the mechanism of

action of the approved drugs tested here. Finally, we use an *ex vivo* kinetic assay in combination with proteomics-derived GUS abundance data to predict what GUS isoforms are susceptible to inhibition by the tested drugs.

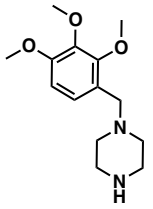
RESULTS AND DISCUSSION

GUS inhibition broad screen

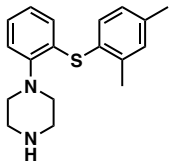
Previously, 5 FDA approved drugs containing either a terminal piperazine or piperadine ring (palbociclib, crizotinib, vortioxetine, amoxapine, and ciprofloxacin) were found to inhibit *Escherichia coli* GUS (*Ec*GUS) *in vitro* with IC_{50} values ranging from 0.5 – 10 μM (*65*). We conducted a substructure search against an online database called e-Drug 3D (last update: July 2020) to expand this list and test all non-antibiotic drugs containing either a terminal piperazine or piperadine ring. e-Drug 3D is a comprehensive database that contains 1993 molecular structures approved between 1939 and 2020 as well as all known metabolites for each drug (*97*). In total, we found 22 non-antibiotic drugs (or associated metabolites) containing either a terminal piperazine or piperadine moiety but only 20 were examined based on commercial availability (**Figure 3.1**). We also included the anti-psychotic loxapine, a methylated analog of amoxapine, in our drug library for the purpose of structure activity relationship (SAR) analysis. Using a previously described formula, the calculated small intestine and colonic concentrations of the drugs in our library range from 5 – 340 μM and 10 – 1000 μM, respectively (**Table 3.1**) (*52*).



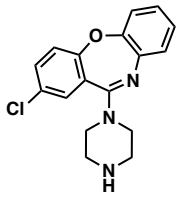
X = C or N



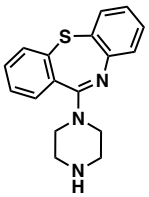
Trimetazidine
Anti-Ischemic



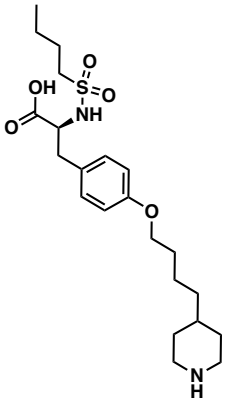
Vortioxetine
Anti-Depressant



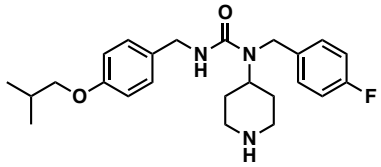
Amoxapine
Anti-Depressant



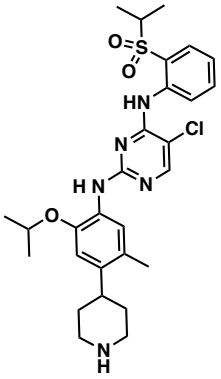
Norquetiapine*
Anti-Psychotic



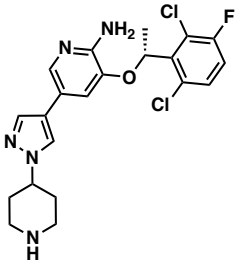
Tirofiban
Anti-Platelet



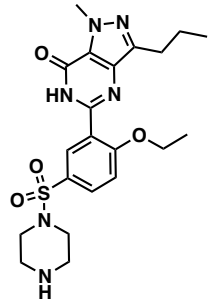
Pimavanserin*
Anti-Psychotic



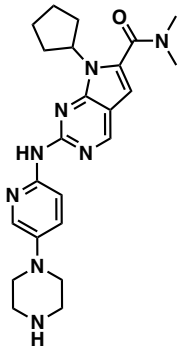
Ceritinib
Anti-Cancer



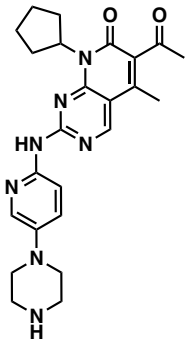
Crizotinib
Anti-Cancer



Sildenafil*
Anti-Hypertensive



Ribociclib
Anti-Cancer



Palbociclib
Anti-Cancer

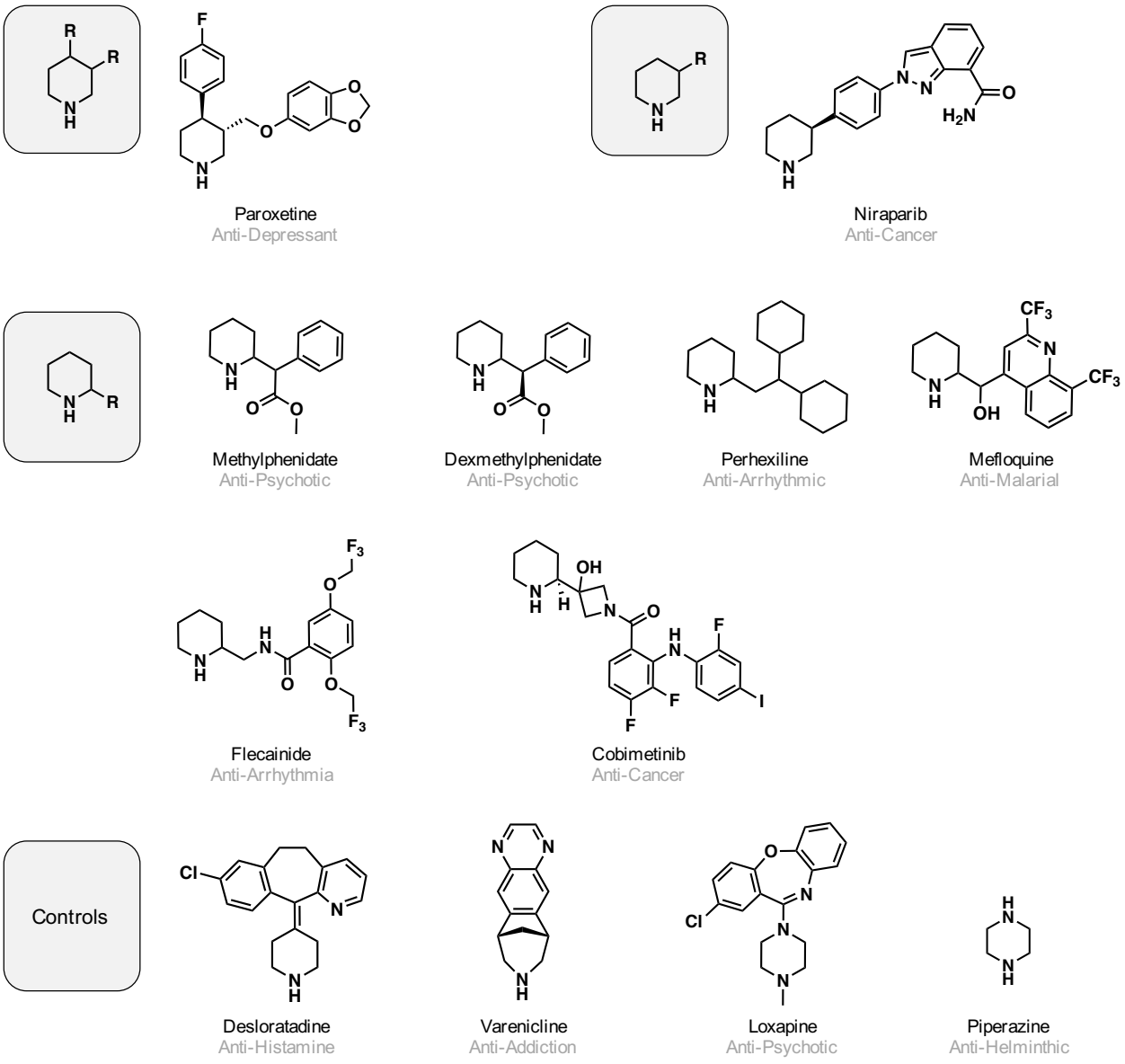


Figure 3.1 Tested drugs and drug metabolites that contain either a terminal piperazine or piperadine moiety. Methylphenidate, dexmethylphenidate, and pimavanserin were not available for purchase and thus not examined.

Table 3.1 Estimated intestinal and colonic concentration of drugs examined in this study. Concentrations were calculated using a previously established formula (*52*). N.A. = not calculated due to insufficient dosing data.

Drug	Drug Type	Intestinal [ ], µM	Colonic [ ], µM
trimetazidine	anti-ischemic	44	N.A.
vortioxetine	anti-psychotic	22	29
amoxapine	anti-psychotic	106	96
norquetiapine	anti-psychotic	339	339
tirofiban	anti-clotting	10	12
ceritinib	anti-cancer	90	414
crizotinib	anti-cancer	185	582
N-desmethyl sildenafil	anti-hypertensive	72	290
ribociclib	anti-cancer	153	529
palbociclib	anti-cancer	75	275
paroxetine	anti-psychotic	40	73
niraparib	anti-cancer	104	151
perhexiline	anti-hypertensive	240	N.A.
cobimetinib	anti-cancer	13	48
mefloquine	anti-malarial	220	991
flecainide	anti-arrhythmic	121	30
desloratadine	anti-histamine	5	11
varenicline	anti-addiction	2	N.A.
loxapine	anti-psychotic	10	N.A.
piperazine	anti-helminthic	N.A.	N.A.

The variation in GUS enzyme structure between gut bacterial species prompted Pollet and co-workers to categorize GUS enzymes into unique structural categories (*77*). GUS enzymes were initially binned into unique structural categories based on the size and location of an adjacent active site loop which plays a key role in recognition and binding of small molecule substrates and inhibitors (*77*). Interestingly, Biernet and co-workers observed that **UNC10201652** and other small molecule inhibitors selectively and most potently inhibit GUS enzymes of the Loop 1 (L1) class (*63*). Thus, we hypothesized that the tested drugs here will uniquely inhibit L1 GUS enzymes. We tested each drug in our library against representative GUS enzymes from each of the previously established GUS structural categories including a category of GUS enzymes that bind the co-factor FMN (**Table 3.2**). Each drug was tested for GUS inhibition using the reporter substrate 4-methylumbelliferone-β-D-glucuronide (4-MU-G)

at the physiologically relevant pH of 7.4. All drugs were tested at 10 μM since standard drug screening campaigns use this concentration and more importantly, all of the tested drugs were found to be soluble at this concentration (*98*).

Table 3.2 Examined GUS enzymes in broad screen. GUS categories: Loop 1 (L1); mini-Loop 1 (mL1); Loop 2 (L2); mini-Loop 2 (mL2); mini-Loop 1, mini-Loop 2 (mL1, mL2); No Loop (NL); FMN-binding (FMN); N-Terminal Loop (NTL).

GUS	Loop Category	Oligomeric State	PDB Code
C. perfringens	L1	Tetramer	4JKM
E. coli	L1	Tetramer	3K46
E. eligens	L1	Tetramer	6BJW
S. agalactiae	L1	Tetramer	4JKL
B. fragilis	mL1	Tetramer	3CMG
B. uniformis-2	L2	Dimer	5UJ6
P. meridae	mL2	Tetramer	6DXU
B. ovatus	mL1, mL2	Tetramer	6D8K
F. saccharivorans	NL	Hexamer	6NCZ
B. dorei	NL	Dimer	6ED1
R. gnavus-3	FMN	Dimer-Tetramer Mix	6MVG
R. hominis-2	FMN	Dimer-Tetramer Mix	6MVH
Faecalibacterium prausnitzii L2-6	FMN	Trimer	6MVF
B. uniformis-1	NTL	Tetramer	6D1N

As expected, we observed inhibition of the tested L1 GUS enzymes: *Ec*GUS, *Clostridium perfringens* GUS (*Cp*GUS), *Streptococcus agalactia* GUS (*Sa*GUS), and *Eubacterium eligens* GUS (*Ee*GUS) (**Figure 3.2**). These L1 GUS enzymes are tetramers that contain a long loop that overlaps adjacent protomers (*70*). The loop overlay results in the formation of a hydrophobic pocket that is ideal for binding of small molecule substrates and inhibitors but limits access to large carbohydrate chains (*70*). Interestingly, we also observed inhibition of FMN-binding GUS enzymes despite prior structural analysis indicating that *Faecalibacterium prausnitzii* L2-6 GUS (*Fp*L2-6GUS) and *Roseburia hominis* GUS (*Rh*2GUS) contain solvent exposed active sites which is suggestive for preferential binding of glucuronate containing polysaccharides (*99*). Further structural characterization will need to be accomplished to rationalize inhibition of FMN-binding GUS enzymes by the tested drugs. We did not observe

potent inhibition of GUS enzymes from other GUS structural classes. Many of these GUS enzymes contain large solvent exposed active sites that are not conducive to binding by small molecule inhibitors, but rather large polysaccharide chains. Indeed, GUS enzymes from the mini-Loop 1 (mL1), Loop 2 (L2), mini-Loop 2 (mL2), and No Loop (NL) classes have been observed to process the carbohydrate heparin *in vitro* while having lower catalytic efficiencies against the small molecule reporter substrate *p*-nitrophenyl-β-D-glucuronide (*77*).

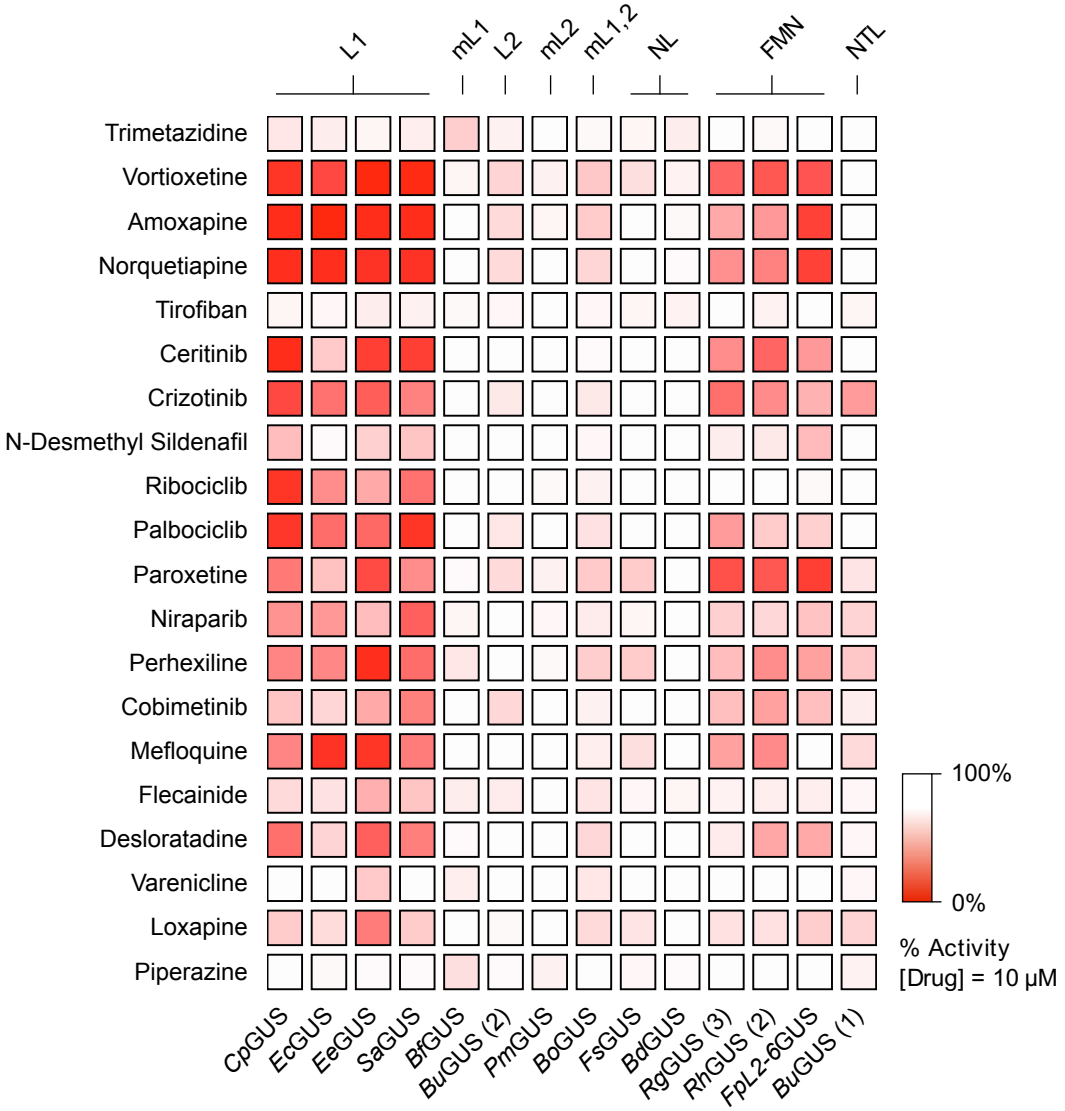
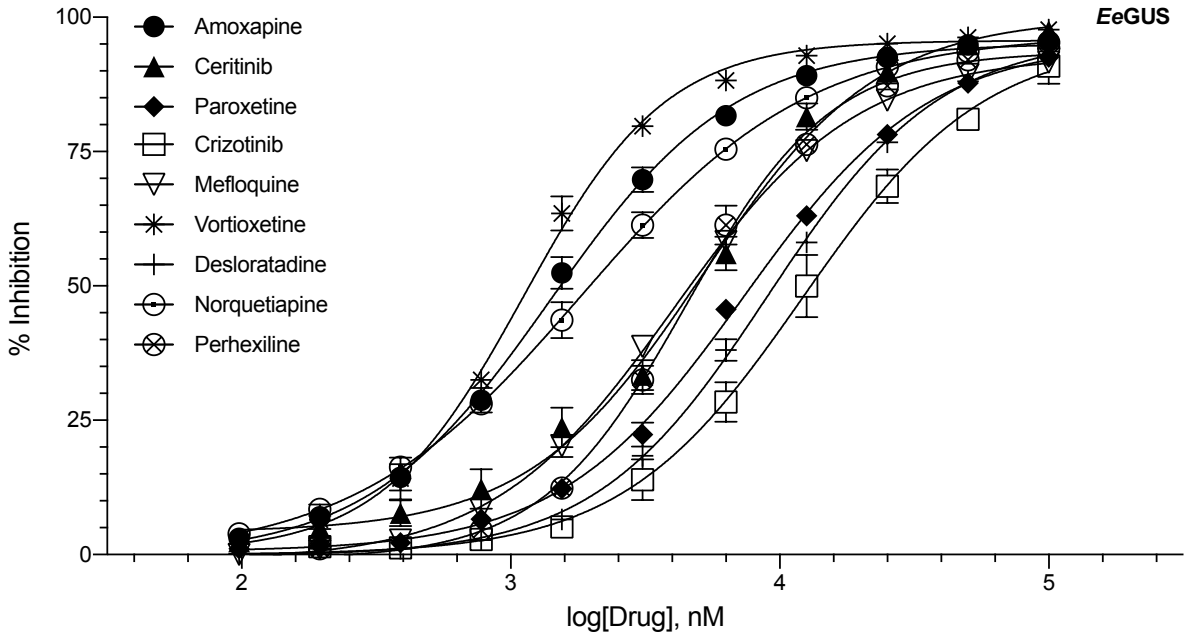


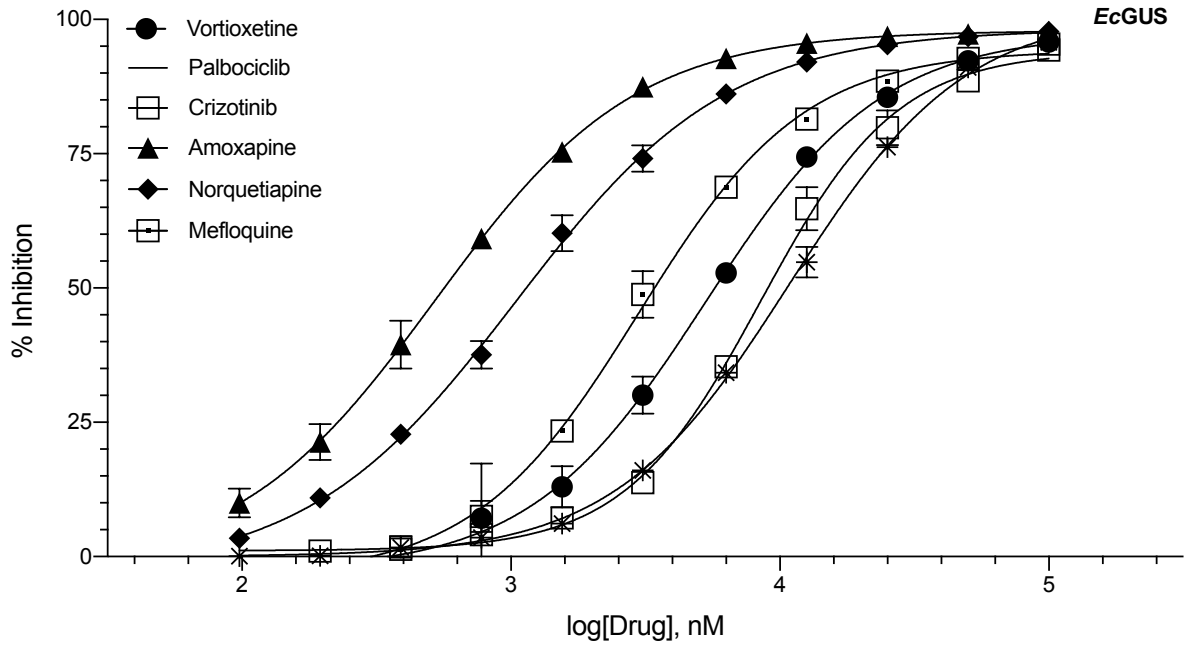
Figure 3.2 GUS inhibition data for tested drugs. Drugs were tested at 10 μM and percent activity remaining is reported.

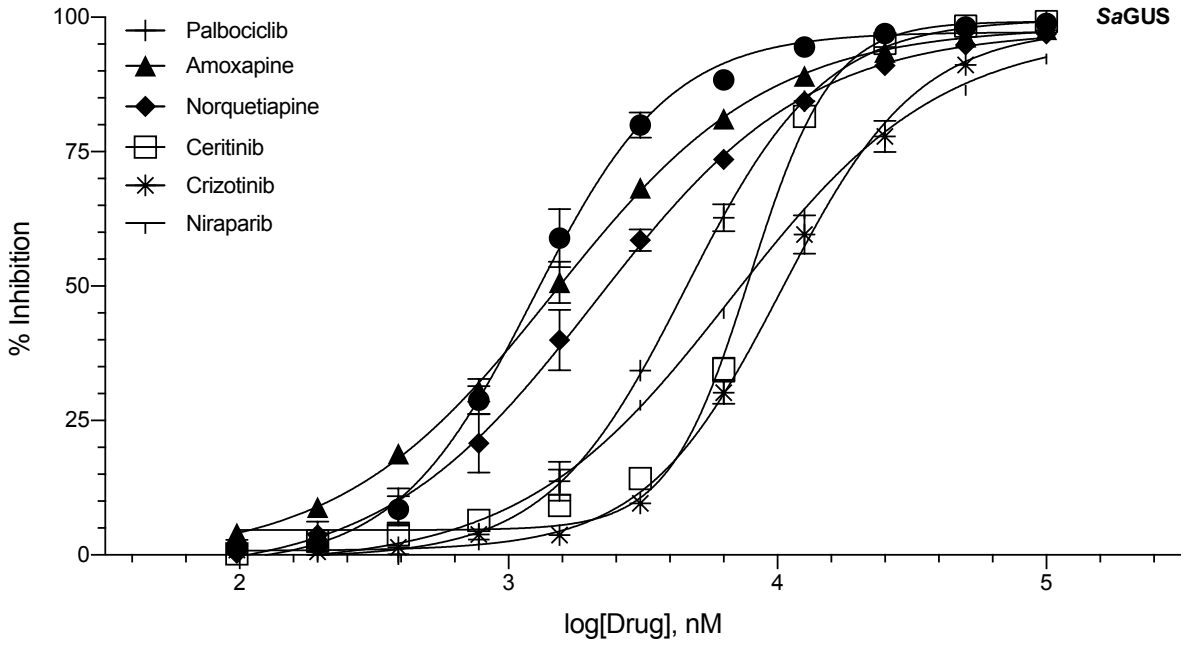
We next tabulated IC_{50} values for inhibition of L1 and FMN-binding GUS enzymes by the tested drugs. Full dose-response plots are only presented for inhibitors with IC_{50} values ≤ 10 μM based on the initial broad screen (**Table 3.3** and **Figure 3.3**). An analysis of the IC_{50} values revealed variability in inhibition between the different tested L1 GUS enzymes which is likely due to differences in the Loop 1 sequence (*63*). Overall, we see less potent inhibition for the FMN-binding GUS enzymes when compared to L1 GUS enzymes except for ceritinib and crizotinib where near equal inhibition across the FMN and L1 GUS enzymes is observed (**Table 3.3**). Taken together, we have found more drugs that inhibit GUS enzymes and in addition to L1 GUS enzymes, FMN-binding GUS enzymes are also inhibited by piperazine and piperadine containing drugs.

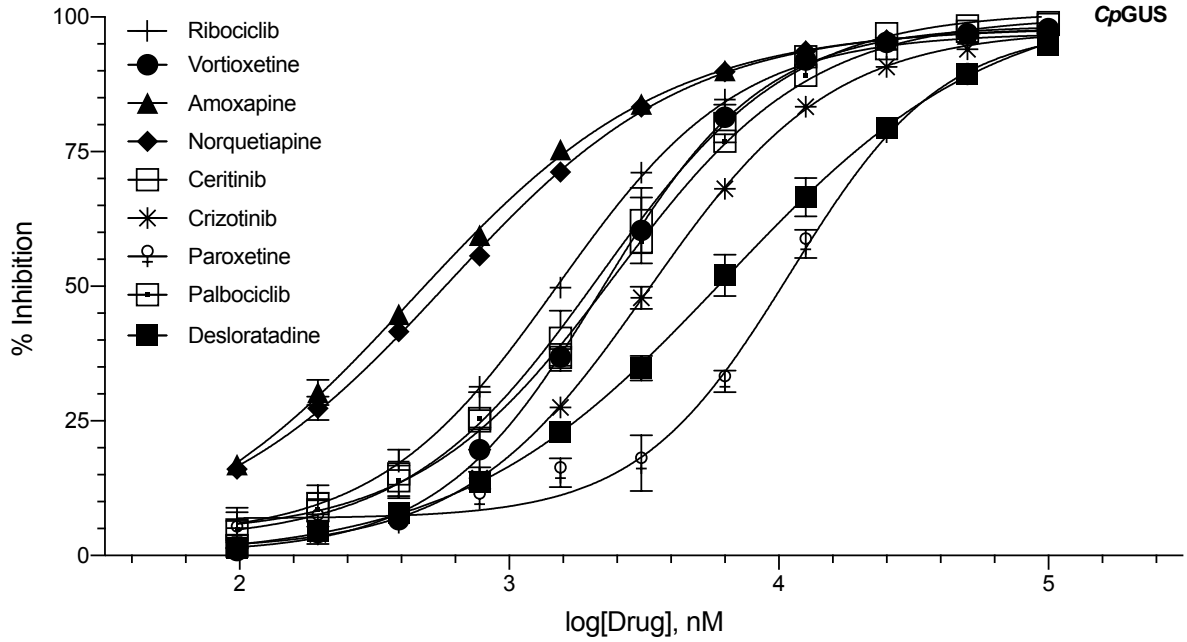
Table 3.3 IC_{50} values for tested drugs against L1 and FMN-binding GUS enzymes.

IC_{50} (μM)	***Ec*GUS**	***Ee*GUS**	***Sa*GUS**	***Cp*GUS**	***Rg*3GUS**	***Rh*2GUS**
trimetazidine	> 10	> 10	> 10	> 10	> 10	> 10
vortioxetine	5.3 ± 0.2	1.1 ± 0.1	1.2 ± 0.1	2.2 ± 0.3	3.2 ± 0.6	> 10
amoxapine	0.52 ± 0.07	1.46 ± 0.09	1.5 ± 0.2	0.44 ± 0.03	> 10	> 10
norquetiapine	1.1 ± 0.2	1.8 ± 0.2	2.1 ± 0.3	0.59 ± 0.05	> 10	> 10
tirofiban	> 10	> 10	> 10	> 10	> 10	> 10
ceritinib	> 10	5.2 ± 0.3	7.9 ± 0.2	2.2 ± 0.4	6.4 ± 0.4	2.4 ± 0.2
crizotinib	8.5 ± 0.7	12 ± 3	10 ± 1	3.27 ± 0.07	6.1 ± 0.6	9.2 ± 0.5
N-desmethyl sildenafil	> 10	> 10	> 10	> 10	> 10	> 10
ribociclib	> 10	> 10	> 10	1.51 ± 0.09	> 10	> 10
palbociclib	11 ± 1	> 10	4.6 ± 0.3	2.46 ± 0.05	> 10	> 10
paroxetine	> 10	7.4 ± 0.4	> 10	11 ± 1	3.1 ± 0.7	1.6 ± 0.3
niraparib	> 10	> 10	6.7 ± 0.6	> 10	> 10	> 10
perhexiline	> 10	4.5 ± 0.4	> 10	> 10	> 10	> 10
cobimetinib	> 10	> 10	> 10	> 10	> 10	> 10
mefloquine	2.98 ± 0.07	4.0 ± 0.2	> 10	> 10	> 10	5 ± 2
flecainide	> 10	> 10	> 10	> 10	> 10	> 10
desloratadine	> 10	9.0 ± 0.6	> 10	6 ± 1	> 10	> 10
varenicline	> 10	> 10	> 10	> 10	> 10	> 10
loxapine	> 10	> 10	> 10	> 10	> 10	> 10
piperazine	> 10	> 10	> 10	> 10	> 10	> 10


*Ee*GUS
Amoxapine
Ceritinib
Paroxetine
Crizotinib
Mefloquine
Vortioxetine
Desloratadine
Norquetiapine
Perhexiline
% Inhibition
log[Drug], nM


*Ec*GUS
Vortioxetine
Palbociclib
Crizotinib
Amoxapine
Norquetiapine
Mefloquine
% Inhibition
log[Drug], nM


SaGUS
Palbociclib
Amoxapine
Norquetiapine
Ceritinib
Crizotinib
Niraparib
% Inhibition
log[Drug], nM


CpGUS
Ribociclib
Vortioxetine
Amoxapine
Norquetiapine
Ceritinib
Crizotinib
Paroxetine
Palbociclib
Desloratadine
% Inhibition
log[Drug], nM

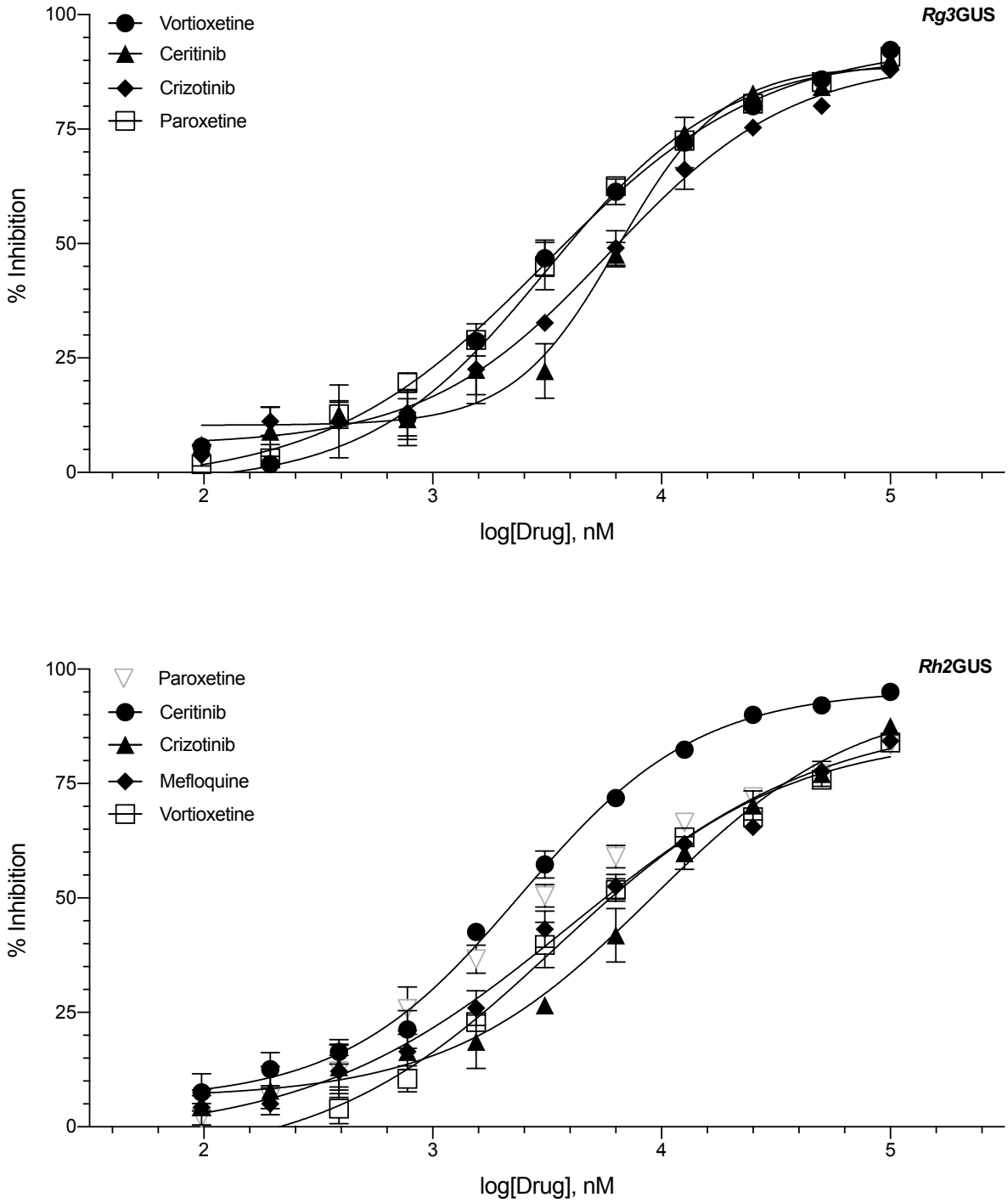
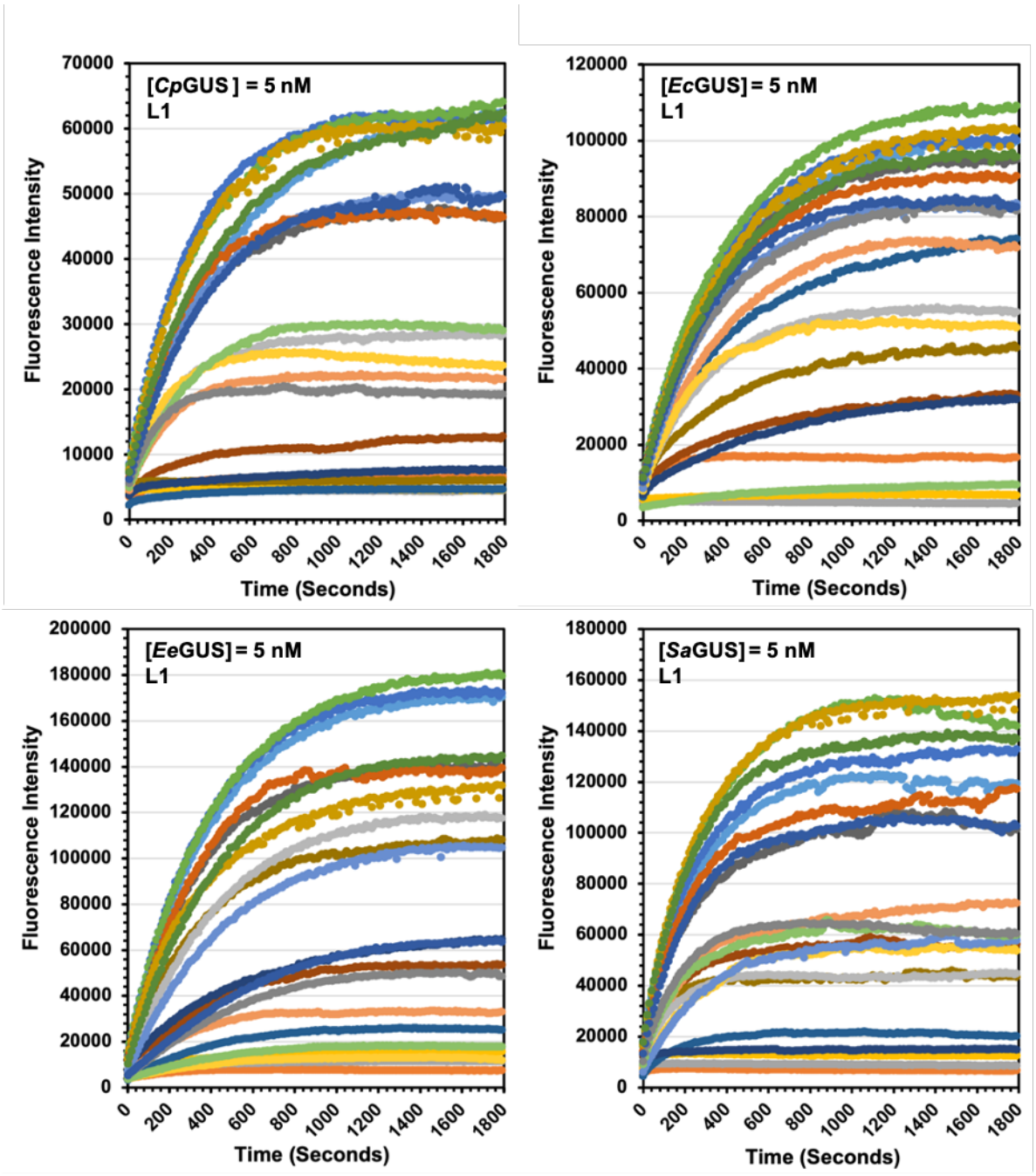


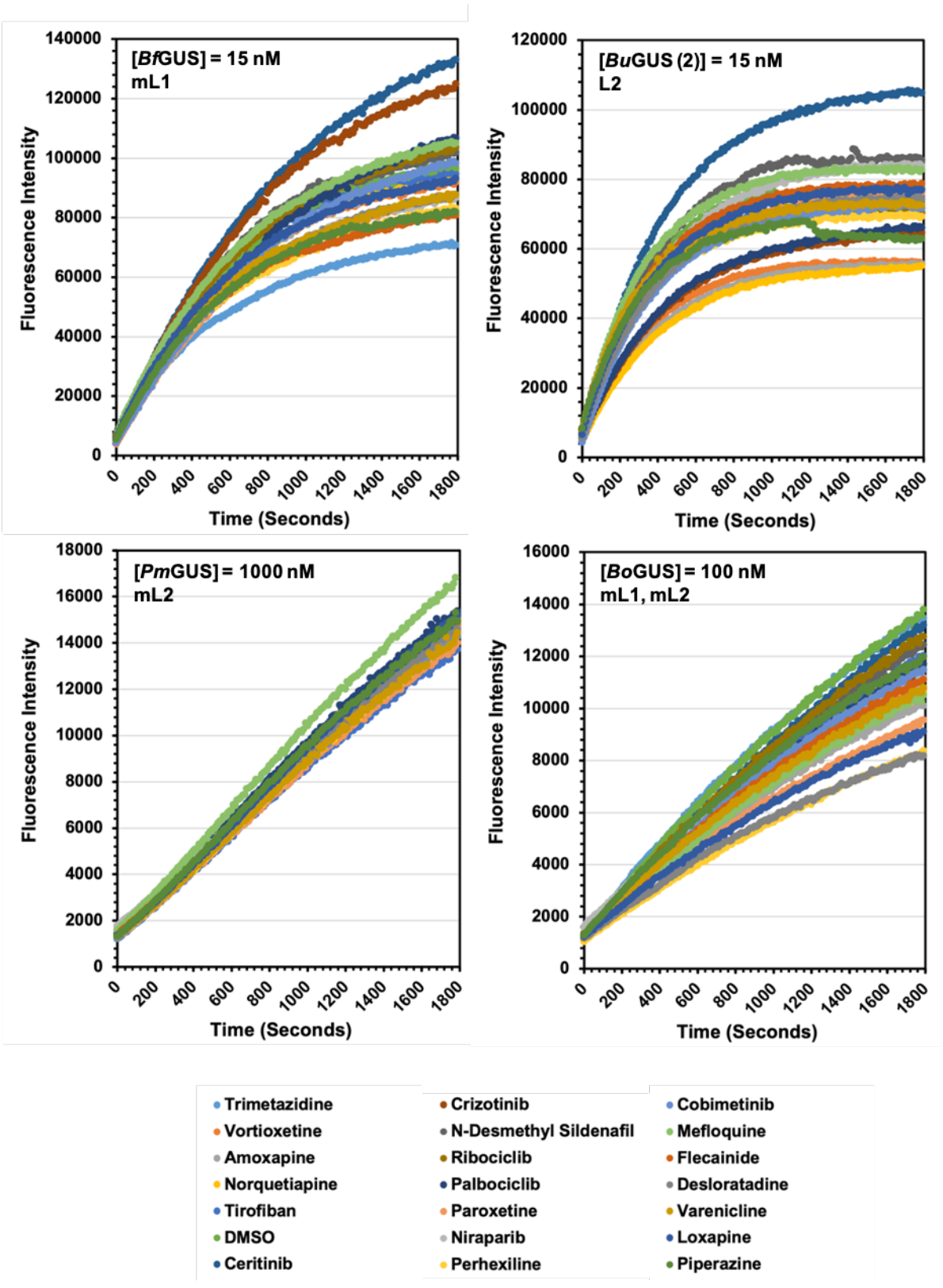
Figure 3.3 Dose response plots for tested drugs against L1 and FMN-binding GUS enzymes. Full curves are reported for drug-GUS pairs that have IC_{50} values ≤ 10 μM.

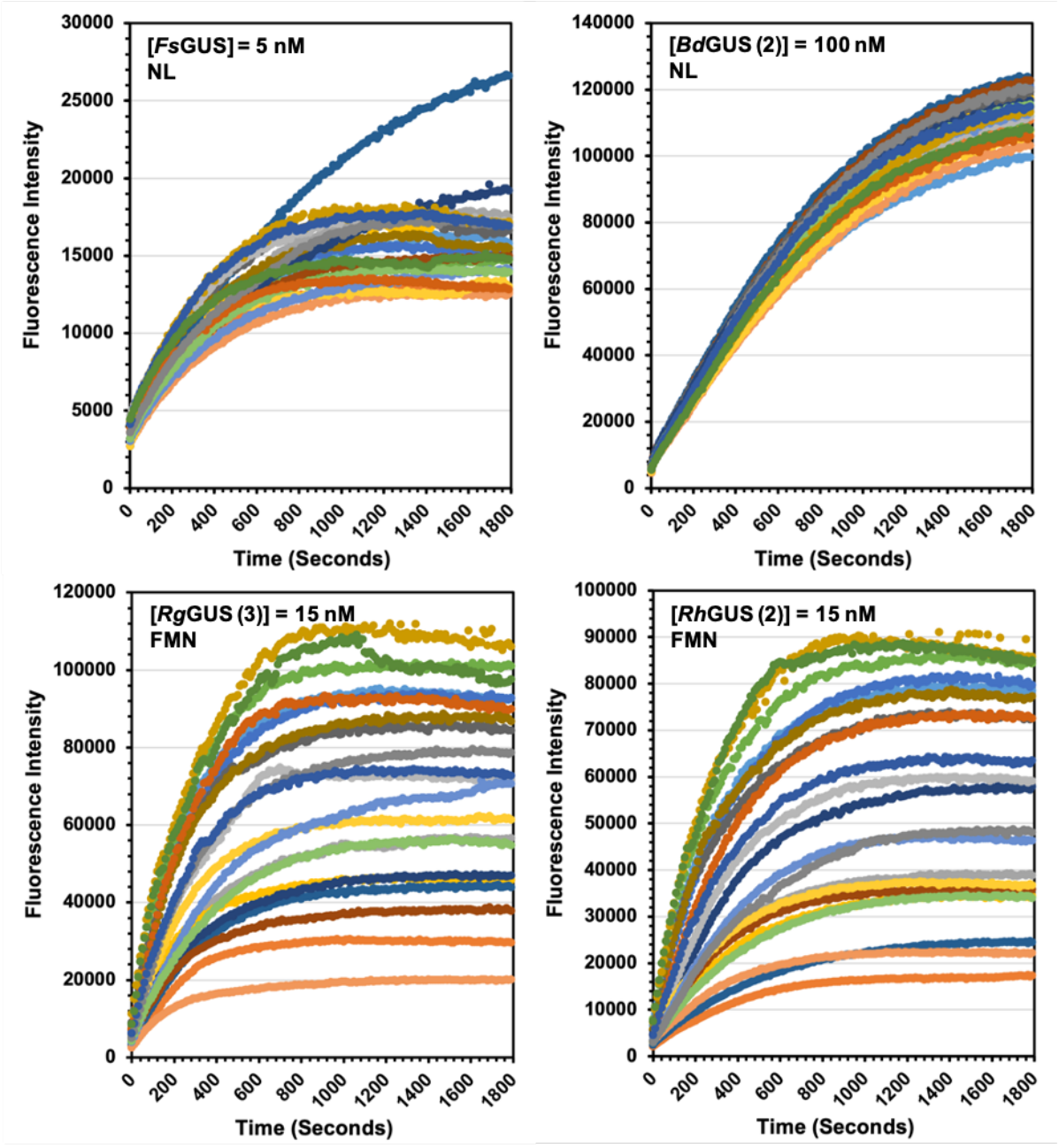
Investigating the mechanism of action

UNC10201652 was previously found to be a slow binding, substrate dependent inhibitor of GUS enzymes (*65*). Substrate turnover is necessary for potent inhibition by this inhibitor. In the absence of **UNC10201652**, the catalytic glutamate in the GUS enzyme attacks the anomeric carbon of the small molecule glucuronide which results in the release of the aglycone moiety and the formation of a covalent bond between the GlcA and the GUS enzyme. The putative catalytic acid/base glutamate residue deprotonates a water molecule which then attacks the anomeric carbon thereby releasing the GlcA from the enzyme active site. In the presence of **UNC10201652**, the catalytic acid/base glutamate residue deprotonates the secondary piperazine amine. The nucleophilic amine then attacks the GlcA-enzyme intermediate which results in the formation of a covalent bond between the piperazine ring and GlcA.

To determine whether the drugs tested here follow this mechanism and are also slow-binding, we first collected progress curves for all drug-GUS pairs (**Figure 3.4**). All drugs that inhibited GUS activity have non-linear progress curves which is indicative of slow binding. Additionally, steady state velocities (v_s) of the tested drugs were either zero or nearly zero which indicates that these drugs work as enzyme inactivators (**Figure 3.4**). To assess whether the newly tested drugs here are substrate dependent, we pre-incubated *Ee*GUS with drugs at various time points and found that the kinetic profile was the same for all which indicates that steady state kinetics of these drugs is not driven by inhibitor-enzymes interactions (**Figure 3.5**). Jump dilution assays show that pre-incubation with substrate results in the onset of steady state inhibition which indicates that these drugs are substrate dependent inhibitors of GUS enzymes (**Figure 3.6**).


[CpGUS] = 5 nM
L1
[EcGUS] = 5 nM
L1
[EeGUS] = 5 nM
L1
[SaGUS] = 5 nM
L1
Fluorescence Intensity
Time (Seconds)
Trimetazidine
Vortioxetine
Amoxapine
Norquetiapine
Tirofiban
DMSO
Ceritinib
Crizotinib
N-Desmethyl Sildenafil
Ribociclib
Palbociclib
Paroxetine
Niraparib
Perhexiline
Cobimetinib
Mefloquine
Flecainide
Desloratadine
Varenicline
Loxapine
Piperazine


[BfGUS] = 15 nM
mL1
[BuGUS (2)] = 15 nM
L2
[PmGUS] = 1000 nM
mL2
[BoGUS] = 100 nM
mL1, mL2
Fluorescence Intensity
Time (Seconds)
Trimetazidine
Vortioxetine
Amoxapine
Norquetiapine
Tirofiban
DMSO
Ceritinib
Crizotinib
N-Desmethyl Sildenafil
Ribociclib
Palbociclib
Paroxetine
Niraparib
Perhexiline
Cobimetinib
Mefloquine
Flecainide
Desloratadine
Varenicline
Loxapine
Piperazine


[FsGUS] = 5 nM
NL
[BdGUS (2)] = 100 nM
NL
[RgGUS (3)] = 15 nM
FMN
[RhGUS (2)] = 15 nM
FMN
Fluorescence Intensity
Time (Seconds)


Trimetazidine
Vortioxetine
Amoxapine
Norquetiapine
Tirofiban
DMSO
Ceritinib
Crizotinib
N-Desmethyl Sildenafil
Ribociclib
Palbociclib
Paroxetine
Niraparib
Perhexiline
Cobimetinib
Mefloquine
Flecainide
Desloratadine
Varenicline
Loxapine
Piperazine

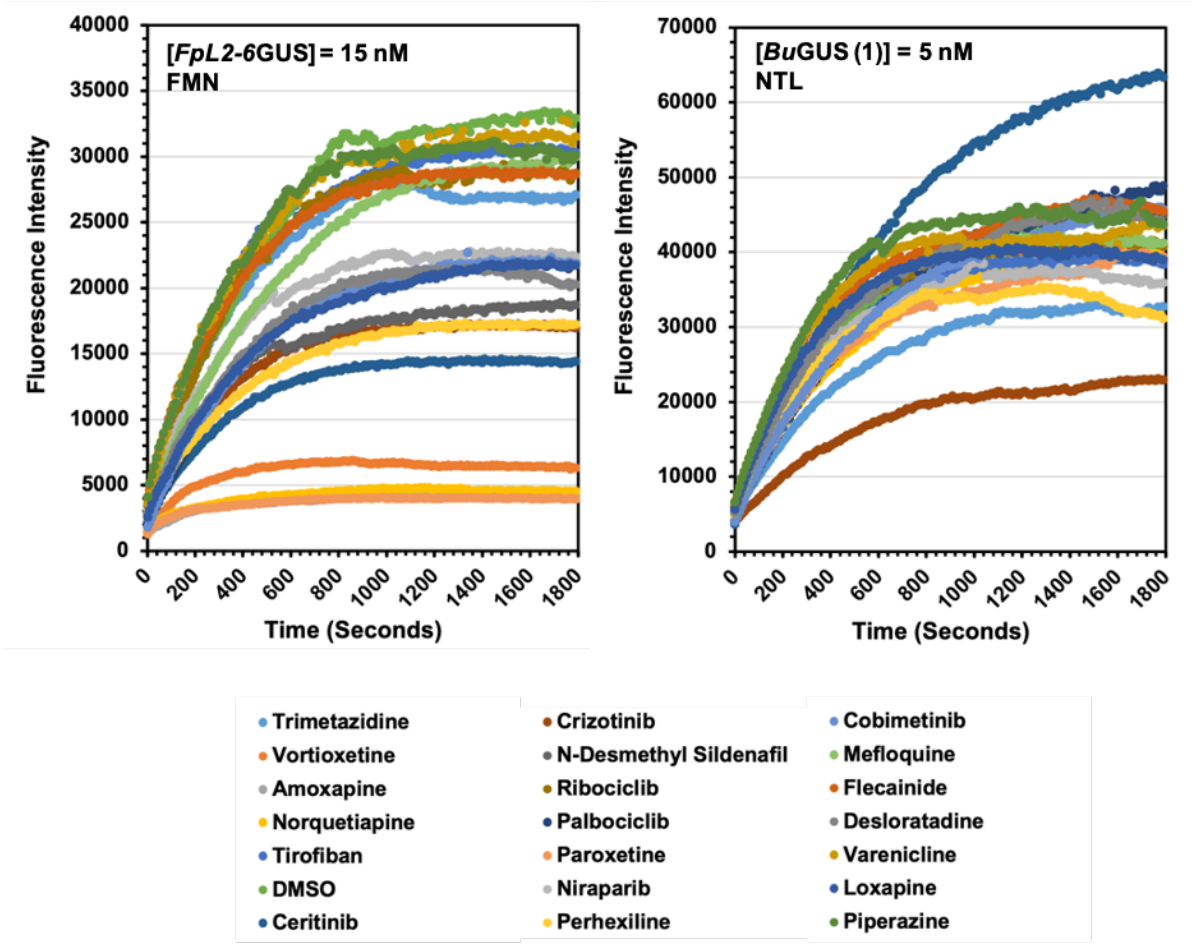
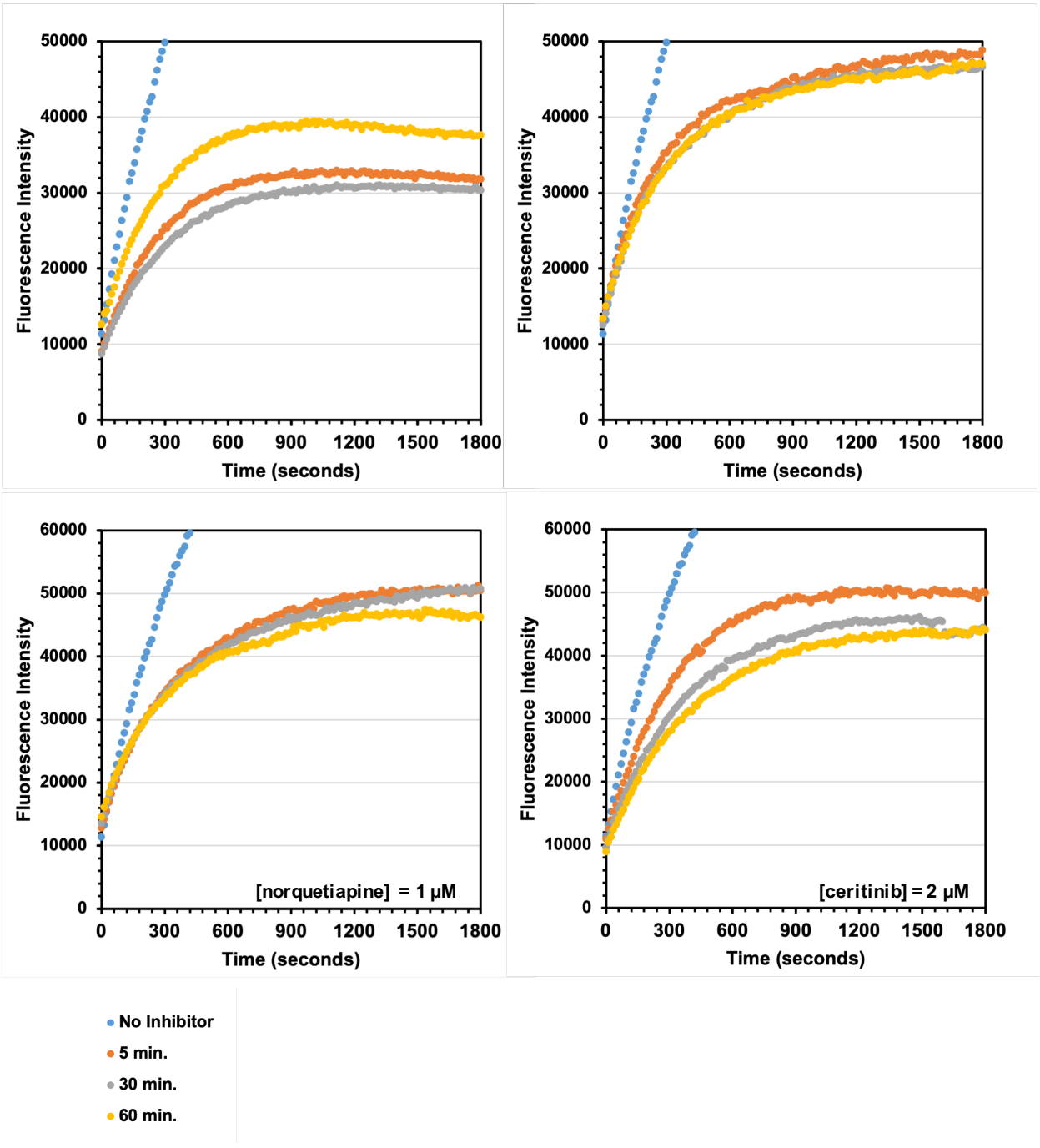
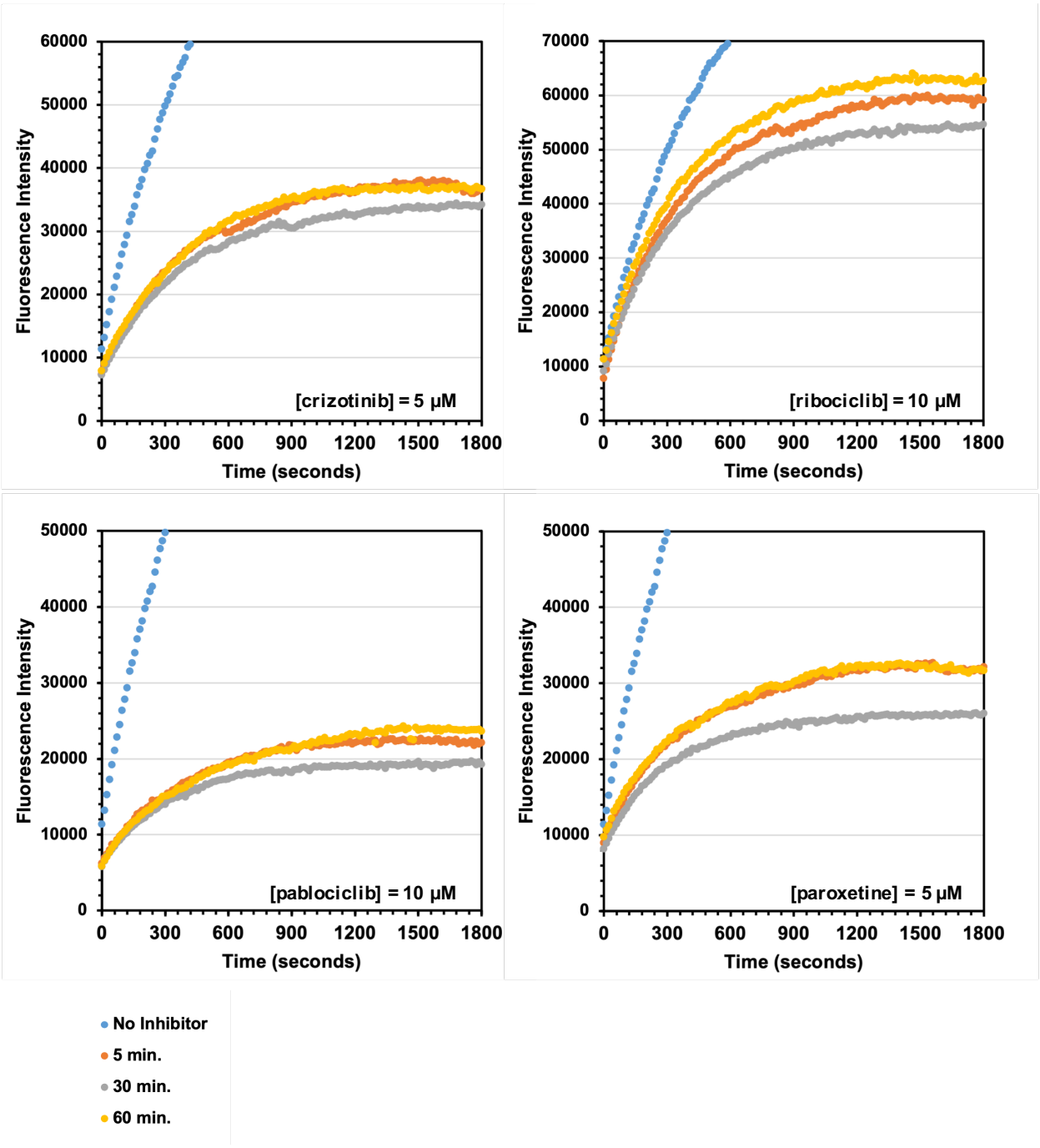


Figure 3.4 Progress curves for all drug-GUS pairs.


Fluorescence Intensity
Time (seconds)
[norquetiapine] = 1 μM
[ceritinib] = 2 μM
No Inhibitor
5 min.
30 min.
60 min.


Fluorescence Intensity
Time (seconds)
[crizotinib] = 5 µM
[ribociclib] = 10 µM
[pablociclib] = 10 µM
[paroxetine] = 5 µM
No Inhibitor
5 min.
30 min.
60 min.

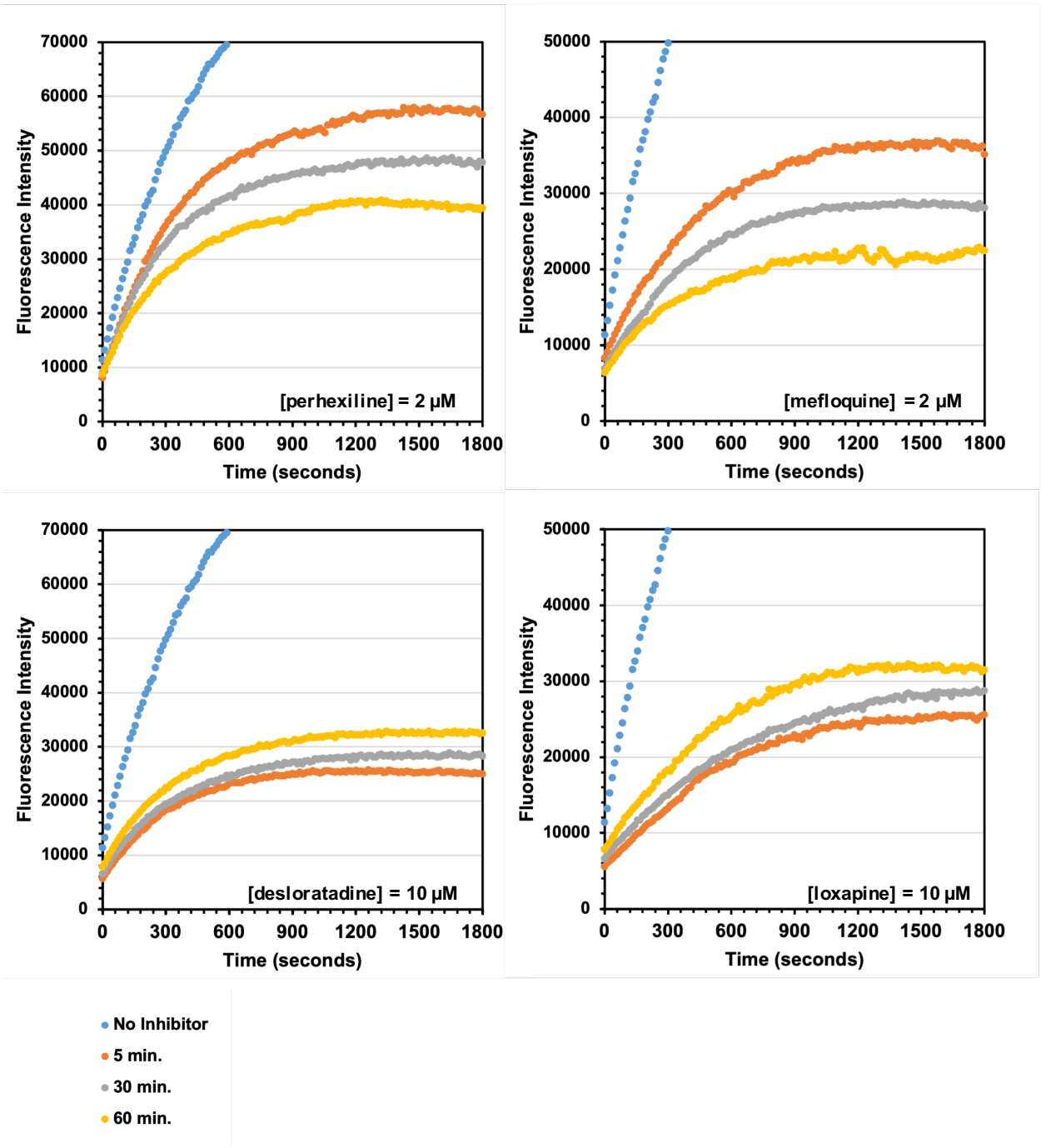
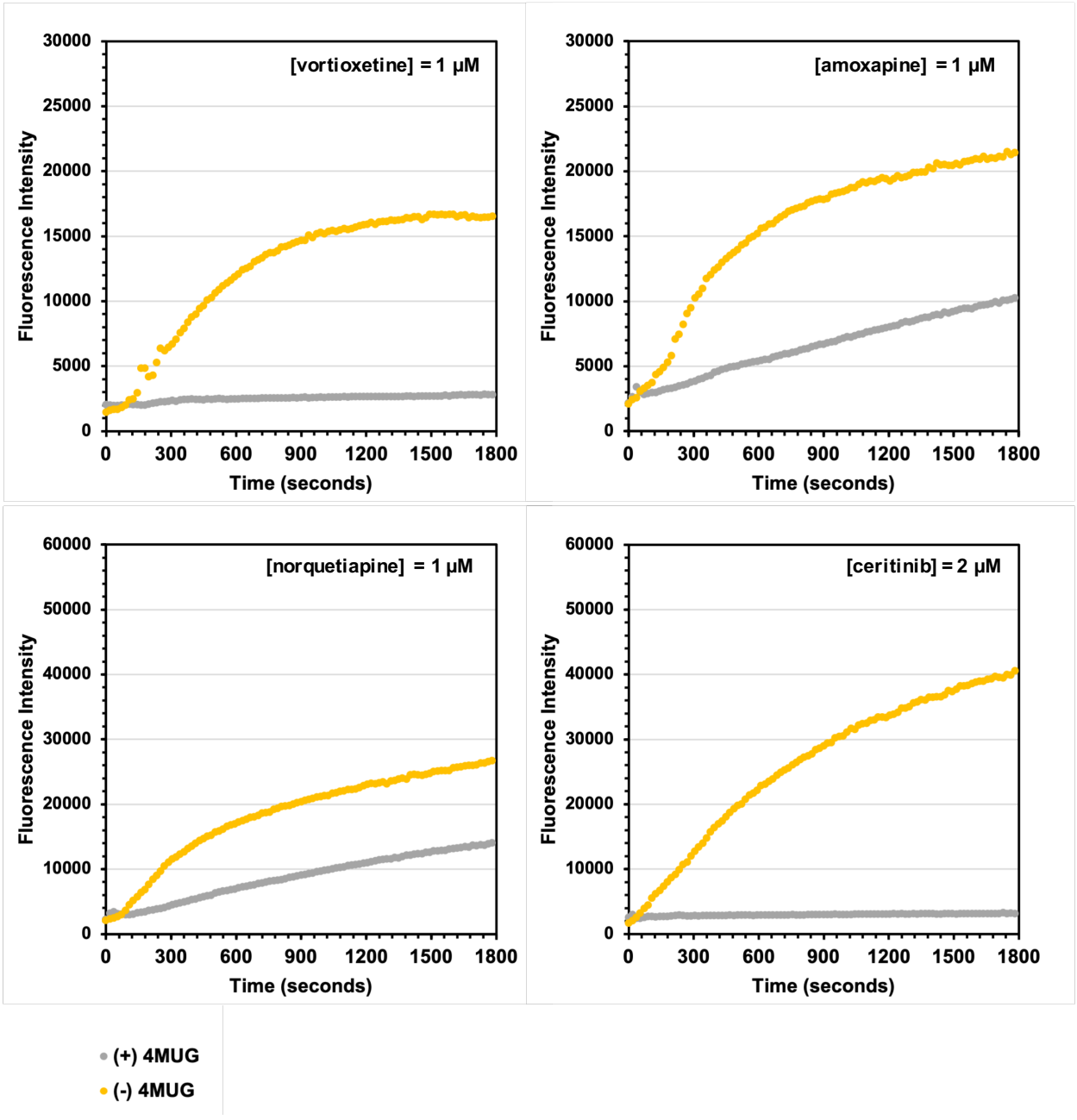
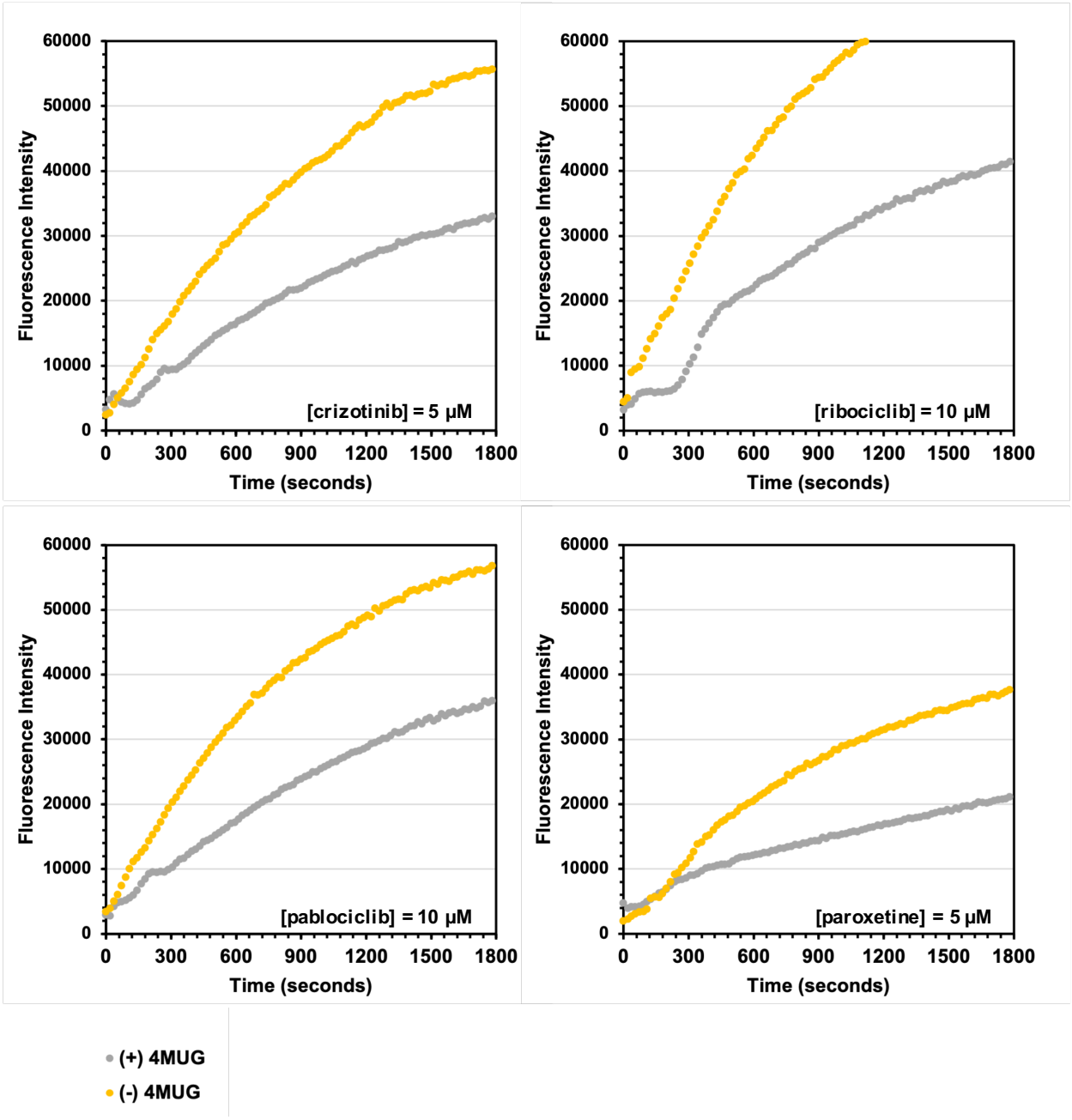


Figure 3.5 Pre-incubation of *Ee*GUS with various drugs and various time points. Only drugs that showed potent inhibition against *Ee*GUS were examined.


[vortioxetine] = 1 µM
[amoxapine] = 1 µM
[norquetiapine] = 1 µM
[ceritinib] = 2 µM
Fluorescence Intensity
Time (seconds)
0
5000
10000
15000
20000
25000
30000
40000
50000
60000
300
600
900
1200
1500
1800
(+) 4MUG
(-) 4MUG


Fluorescence Intensity
Time (seconds)
[crizotinib] = 5 µM
[ribociclib] = 10 µM
[pablociclib] = 10 µM
[paroxetine] = 5 µM
(+) 4MUG
(-) 4MUG

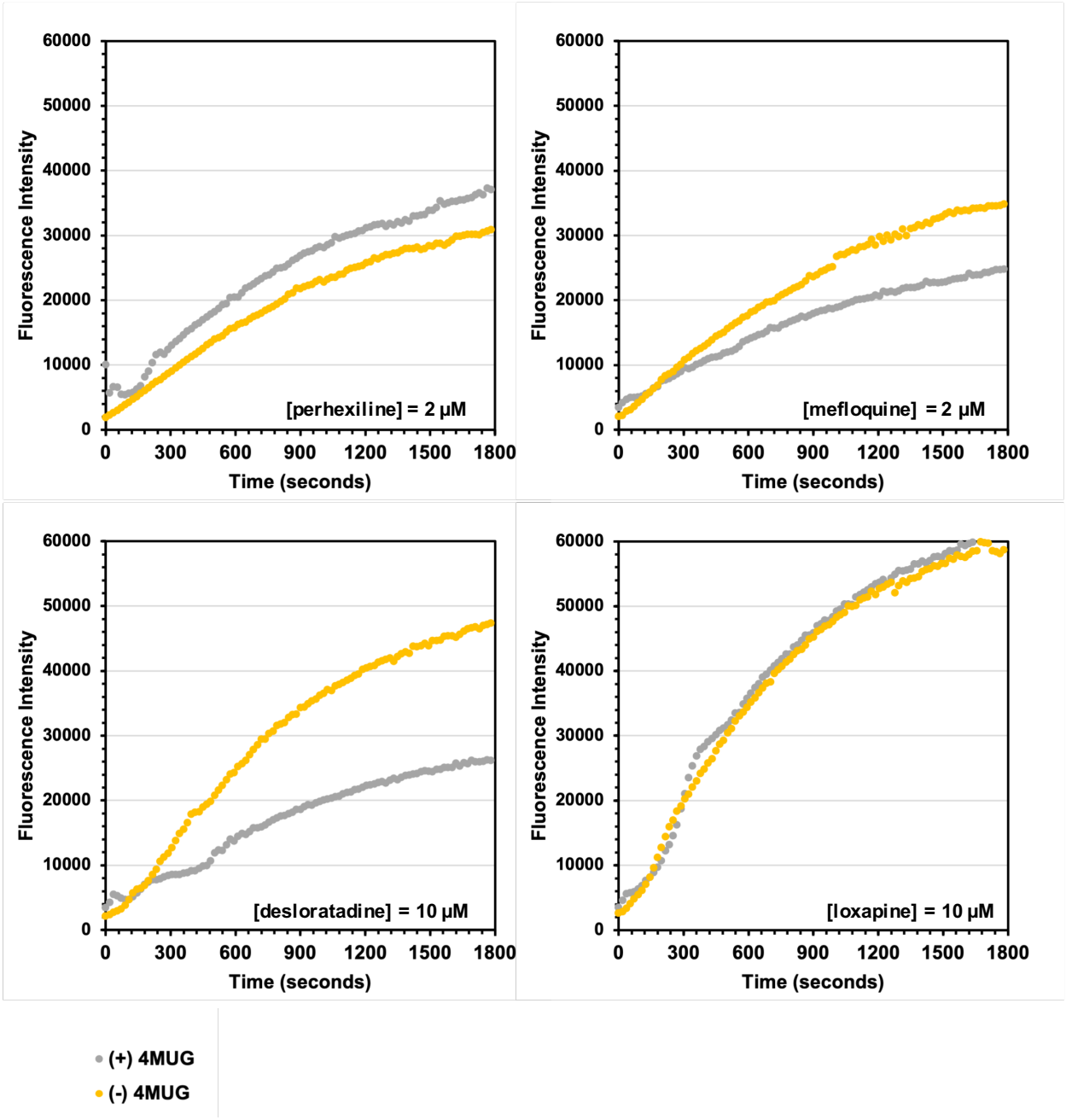


Figure 3.6 Jump dilution assay of *Ee*GUS with various drugs. Only drugs that showed potent inhibition against *Ee*GUS were examined.

We performed a crystallography screening campaign of drugs that were found to inhibit *Ee*GUS to assess whether the drug-GlcA conjugate is forming. Drugs were incubated with *Ee*GUS in the presence of 4-MU-G to capture the drug-GlcA product in the active site. We successfully obtained structures of *Ee*GUS bound to norquetiapine and mefloquine 2.4 Å and 2.0 Å, respectively (**Table 3.4**). Unbiased difference electron density within the active site of *Ee*GUS bound to norquetiapine revealed that norquetiapine was covalently β-linked to the anomeric carbon in GlcA (**Figure 3.7a**). Formation of the norquetiapine-GlcA scaffold was further validated by LC-MS (**Figure 3.7b**). We did not observe a covalent linkage between mefloquine and GlcA in the *Ee*GUS crystal structure (**Figure 3.7c**). Interestingly, LC-MS analysis of *Ee*GUS incubated with mefloquine and 4-MU-G revealed formation of the mefloquine-GlcA conjugate (**Figure 3.7d**). Possible hydrolysis of the mefloquine-GlcA conjugate may have occurred during the crystallization process which may be why the conjugation is not observed in the mefloquine bound *Ee*GUS structure. While *Ee*GUS was examined here, future work should be done to assess whether drug-GlcA conjugates form in FMN-binding GUSs.

Table 3.4 Crystallographic statistics for *E. eligens* GUS bound to norquetiapine and mefloquine.

	Mefloquine	Norquetiapine
Resolution range	29.43-2.04 (2.12-2.04)	33.64-2.40 (2.49-2.40)
Space group	P 64 2 2	C 64 2 2
Unit cell [a, b, c (Å); α, β, γ (°)]	179.80, 179.80, 134.92; 90, 90, 120	179.42, 179.42, 134.54; 90, 90, 120
Total reflections	153949 (11291)	470863 (34768)
Unique reflections	77542 (4872)	50048 (4779)
Multiplicity	2.0 (1.9)	9.4 (7.3)
Completeness (%)	93.8 (60.9)	99.6 (96.8)
Mean I/sigma(I)	14.85 (0.39)	13.67 (1.11)
Wilson B-factor	51.65	57.09
R-merge	0.02621 (1.689)	0.1112 (1.113)
R-pim	0.02621 (1.689)	0.03841 (0.4304)
CC1/2	0.999 (0.143)	0.982 (0.613)
R-work	0.2076 (0.3955)	0.2012 (0.3139)
R-free	0.2410 (0.4069)	0.2283 (0.3460)
# of Non-Hydrogen Atoms	4798	4867
Macromolecules	4612	4723
Ligands	57	33
Solvent	129	111
Protein residues	586	598
RMS (bonds) (Å)	0.015	0.009
RMS (angles) (°)	1.34	1.26
Ramachandran outliers (%)	0.17	0.00
Rotamer outliers (%)	0.00	0.40
Clash score	4.75	7.71
Average B-factor (Å^2)	61.94	88.13
Macromolecules	62.06	88.86
Ligands	73.85	73.33
Solvent	52.74	61.50

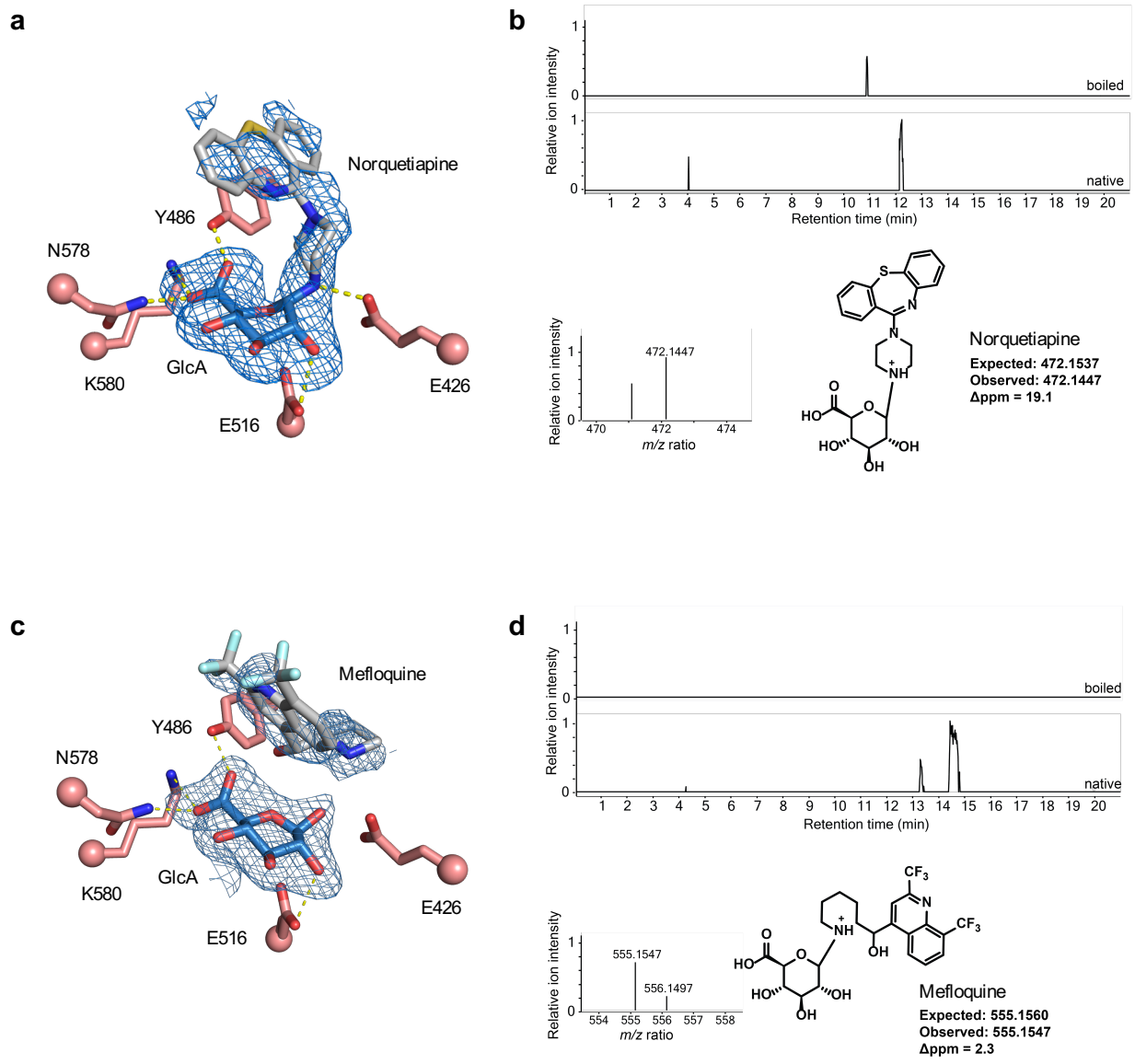


Figure 3.7 Structural analysis of norquetiapine and mefloquine bound to *Ee*GUS. (a) Active site of *Ee*GUS bound to norquetiapine-GlcA conjugate with F_o-F_c density. (b) Mass spectrum of norquetiapine-GlcA conjugate that is generated when incubating *Ee*GUS with norquetiapine and 4-MUG. (c) Active site of *Ee*GUS bound to mefloquine and GlcA with F_o-F_c density. (d) Mass spectrum of mefloquine-GlcA conjugate that is generated when incubating *Ee*GUS with mefloquine and 4-MUG.

In addition to the capture of the GlcA within the active site, previous structural work has shown that interaction with the long loop in L1 GUS enzymes is also important for potent binding by **UNC10201652** (*63*, *65*). The long loop overlaps adjacent protomers which creates a

hydrophobic pocket that is ideal for binding of small molecule inhibitors like **UNC10201652** (*63*, *65*). The amino acid residues in the Loop 1 sequence interacts with the non-piperazine and -piperadine portion (herein termed "R-Group") in **UNC10201652** which stabilizes it within the active site. Indeed, we found that the compound piperazine did not inhibit GUS activity which indicates that R-Group is necessary for binding and inhibition (**Table 3.3**). Furthermore, we did not observe GUS inhibition for tirofiban, trimetazidine, *N*-desmethyl sildenafil, cobimetinib, and flecainide which may be due to suboptimal interactions with the Loop 1 sequence within the active site (**Table 3.3**).

Ex vivo kinetics and activity-based proteomics

In **Chapter 2**, gut bacterial GUS enzymes were identified and quantified using a unique activity-based proteomics approach. Furthermore, percent inhibition of SN-38 glucuronide turnover by **UNC10201652** was shown to correlate strongly with L1 GUS abundance which corroborates previous reports that **UNC10201652** most potently inhibits L1 GUS enzymes (*63*). To assess whether the tested drugs are selective for L1 and FMN-binding GUS enzymes in an *ex vivo* setting, we first determined the GUS profiles of 6 individuals using the proteomics approach outlined in **Chapter 2**. As expected, the diversity and abundance of GUS structural categories varies widely among the profiled individuals (**Figure 3.8a**).

Next, we assessed percent inhibition of 4-MU-G turnover for each drug-fecal lysate set. 4-MU-G is a universal substrate; hence, we can assess correlations for all GUS enzymes. We did not observe inhibition of 4-MU-G turnover by drugs found not to inhibit GUS enzymes *in vitro* like piperazine, trimetazidine, and tirofiban (**Figure 3.8b**). We observed a strong correlation between the sum of L1 and FMN-binding GUS abundance and percent inhibition of 4-MUG

turnover by fecal extracts obtained from the 6 individuals (**Figure 3.8c** and **Figure 3.9**). Taken together, *ex vivo* analysis shows that percent inhibition correlates well with the sum of L1 and FMN-binding GUS abundance. In the future, a higher fecal sample size should be used to further validate correlation.

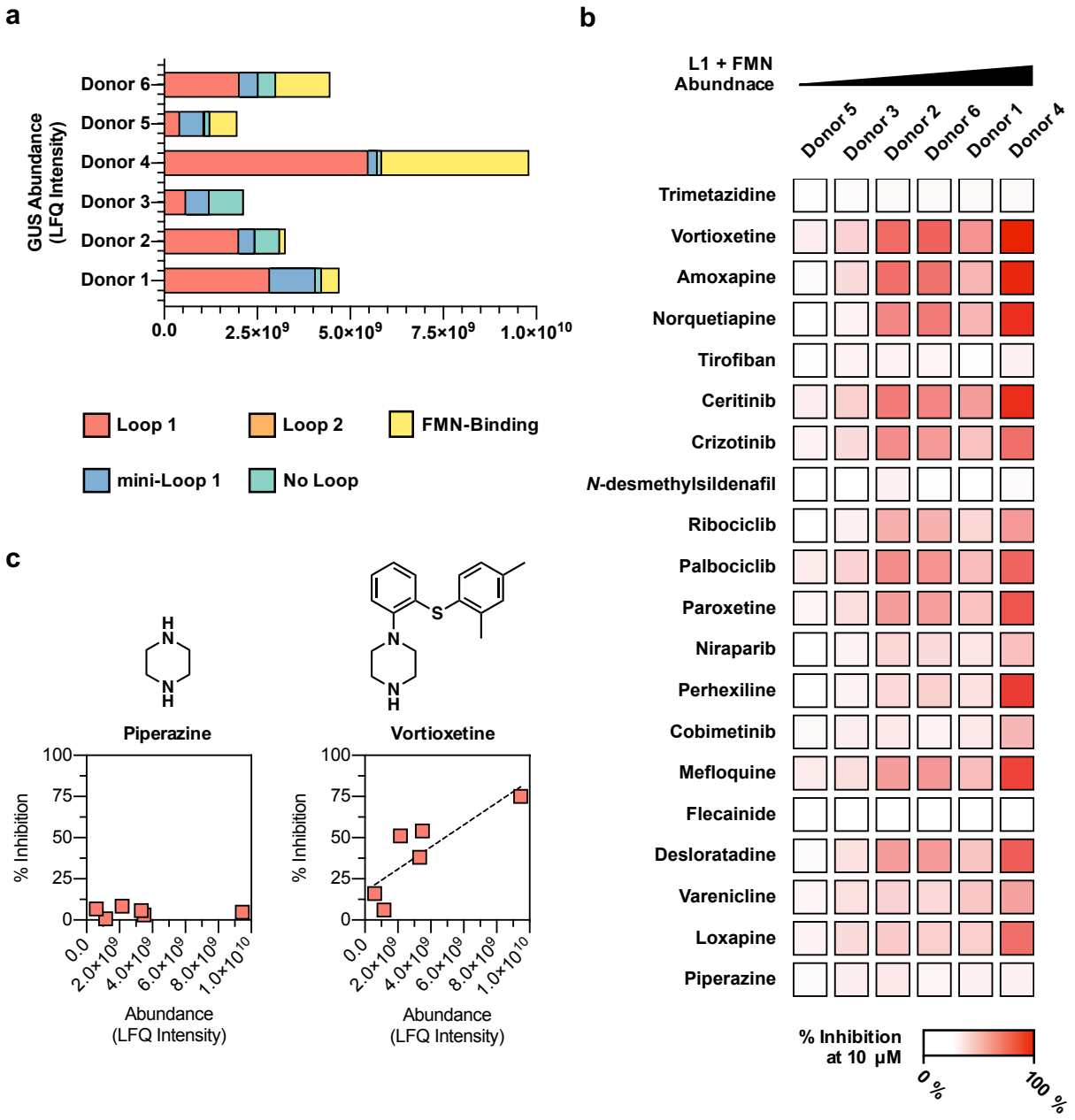


Figure 3.8 ABPP coupled with *ex vivo* inhibition of GUS activity by tested drugs. (a) GUS abundance data (b) Percent inhibition of 4-MU-G turnover by each tested drug. Donors are organized by the sum of L1 and FMN-binding GUS abundances. (c) Correlation between the sum of L1 and FMN-binding GUS abundance and percent inhibition of 4-MU-G turnover.

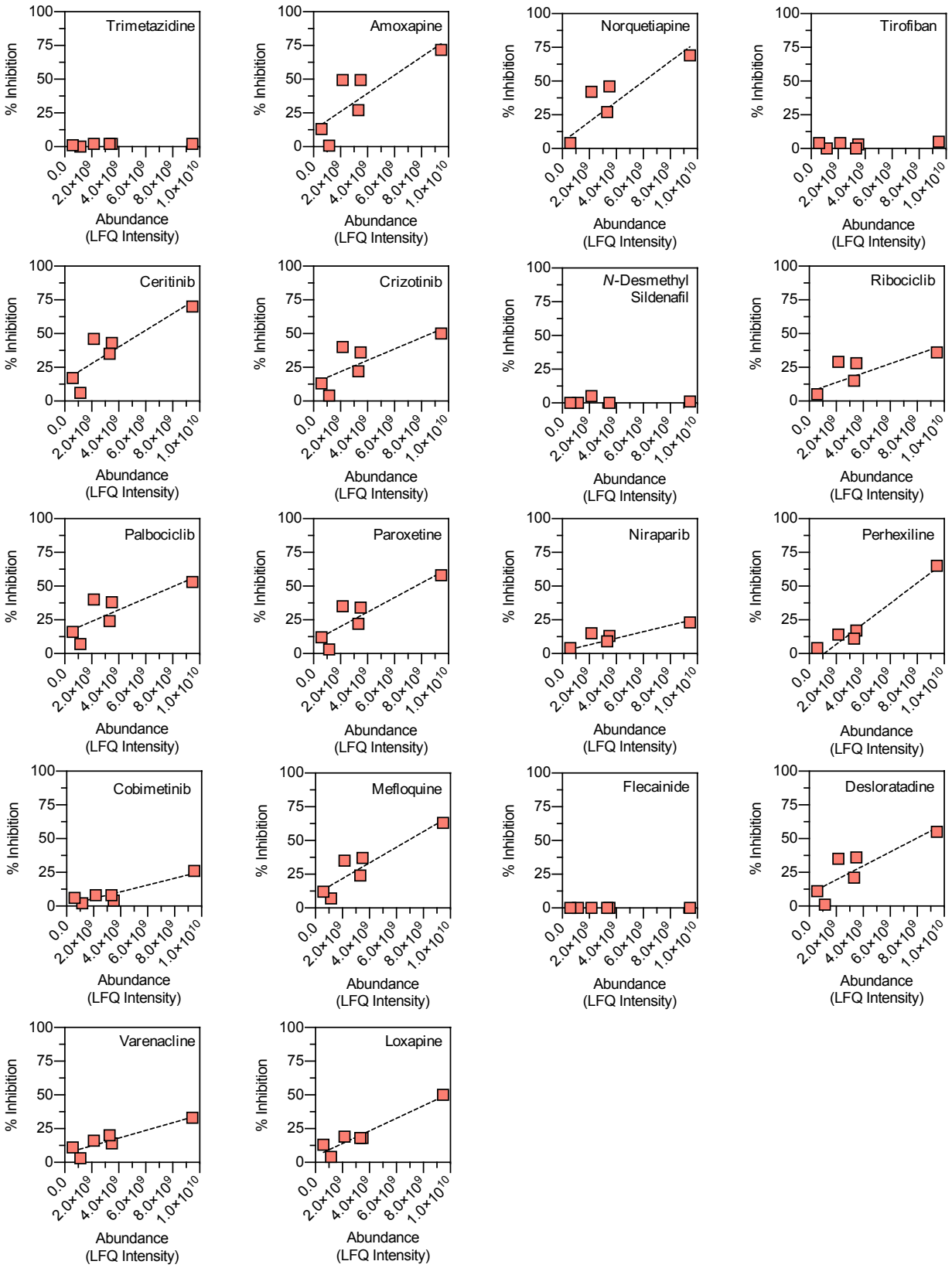


Figure 3.9 Correlation analysis between the sum of L1 and FMN-binding GUS abundance and percent inhibition by tested drugs.

CONCLUSION

Characterization of a selective and potent GUS inhibitor, **UNC10201652**, revealed that the secondary piperazine amine intercepts the glycosyl-enzyme intermediate which results in the formation of a covalent bond with GlcA (*65*). Here, we show that drugs containing a terminal piperazine or piperadine moiety also inhibit GUS enzymes. Importantly, we find that the tested drugs specifically inhibit GUS enzymes from the L1 and FMN-binding GUS structural classes. We validated this finding by correlating *ex vivo* kinetics data with GUS abundance data derived from activity-based proteomics. Finally, using *Ee*GUS as a case study, we show that the drugs tested here are also slow binding and substrate dependent.

The work presented here has shown that the tested drugs can inhibit GUS enzymes *in vitro*. For the few drugs tested, Pellock and co-workers used in-cell assays to show that the drugs can inhibit GUS activity with EC_{50} values ranging from 0.16 – 3.5 μM (*65*). In the future, work needs to be done to study the potency of the drugs tested here using in-cell assays. Importantly, a diverse range of gut microbes that differ in cell wall type (*i.e.,* gram-positive vs. gram negative) need to be tested to better understand what types of microbes are prone to drug uptake and GUS inhibition.

Pellock and co-workers show that the secondary piperazine amine in **UNC10201652** forms a covalent bond with the GlcA intermediate (*65*). The bond forms a salt bridge with a nearby catalytic glutamate which likely aids in binding and stabilization. Examination of synthetic analogs of **UNC10201652** that contain a methylated analog of the piperazine ring revealed inhibition of GUS activity albeit with reduced potency (80-fold weaker). While the tertiary amine in this analog could not serve as a nucleophile, the positive charge likely forms a salt bridge and is stabilized enough to inhibit GUS enzymes (IC_{50} ~10 μM). Drugs containing a

methylated piperazine or piperadine ring should be assessed in the future as their concentrations in the gut may be strong enough to block GUS activity. Additionally, gut bacterial *N*-demethylases may remove the methyl group resulting in the formation of a secondary piperazine amine that will potently inhibit GUS enzymes.

MATERIALS AND METHODS

Protein expression and purification

All recombinant GUS enzymes examined *in vitro* were expressed and purified as previously described (*64*, *77*, *99*). Briefly, all proteins were expressed with a *N*-terminal 6x histidine tag and subsequently purified using a Ni-NTA HP column (GE Healthcare). Additional purification was performed using a HiLoad 16/60 Superdex 200 gel filtration column. Proteins were eluted and aliquots were flash frozen in liquid nitrogen and stored at –80°C until further use.

Protein crystallography

Crystals of *Ee*GUS bound to mefloquine were produced via the hanging-drop vapor diffusion method. *Ee*GUS at 11.5 mg mL^{-1} was preincubated with 10-fold molar excess mefloquine and *p*-nitrophenyl-β-D-glucuronide (*p*NP-G) prior to addition into the crystalline solution. Crystals were formed by incubating ligand bound *Ee*GUS in 0.1 M HEPES: NaOH, pH 7.5 and 20 % (w/v) PEG 8000. The crystals were cryoprotected using 25% propylene glycol.

Crystals of *Ee*GUS bound to norquetiapine were produced via the hanging-drop vapor diffusion method. *Ee*GUS at 11.5 mg mL^{-1} was preincubated with mefloquine (6-fold molar excess) and 4-MU-G (4-fold molar excess) for 1 hour at 37°C prior to addition into the crystalline solution. Crystals were formed by incubating ligand bound *Ee*GUS in 10% PEG400

and 0.1 M sodium acetate pH = 5.0. The crystals were cryoprotected using 35% (w/v) PEG400 and 0.1 M sodium acetate pH = 5.0. Protein crystals were transferred directly from the crystallization drop to the cryoprotectant solution and then quickly flash cooled (< 1 minute) in liquid nitrogen.

All Diffraction data for all crystals were collected on the 23-ID-B beamline at GM/Ca-CAT (Advanced Photon Source, Argonne National Laboratory). Refinements and ligand generation were carried out in Phenix, and ligand fitting was performed in Coot (*86*).

Human fecal extract

Human fecal extracts were made as previously described (see Chapter 2).

GUS activity-based probe (ABPs)

Cyclophellitol-based ABP was synthesized and purified as previously described (*79*).

GUS broad screen

Inhibition of GUS by drugs were calculated using a continuous read format (*65*). Select GUS enzymes were pre-incubated with a 10 μM drug 37°C for 5 min. prior to initiating the reaction with the fluorogenic substrate, 4-methylumbelliferone-β-D-glucuronide (4-MUG). The final reaction volume was 50 μL containing 25 μL water, 10 μL buffer (25 mM HEPES, 25 mM NaCl, 2% DMSO, pH 7.4 final), 5 μL enzyme (varying concentrations), 5 μL drug (10 μM final), and 5 μL 4-MUG (1000 μM final). After addition of 4-MUG, reactions were monitored continuously at 37°C for 30 min. at an excitation wavelength of 350 nm and an emission wavelength of 450 nm (PHERAStar BMG Labtech). Percent activity remaining was calculated using the 30-minute time

point.

Substrate-dependent jump dilution assays

The jump dilution assays to determine the substrate-dependence of slow-binding inhibition were performed by combining 5 µL of 5 µM GUS (500 nM final), 5 µL of various concentrations of inhibitor, 5 µL of 10 mM 4-MUG (1 mM final), and 35 µL of assay buffer (25 mM NaCl, 25 mM HEPES, pH 7.4 final). Reactions were incubated at 37°C for 1 h. After preincubation, 1 µL of the reaction was diluted into 99 µL of 4-MUG-containing buffer (1 mM 4-MUG, 25 mM HEPES, 25 mM NaCl, pH 7.4 final), and the resulting activity was monitored continuously at an excitation wavelength of 350 nm and an emission wavelength of 450 nm (PHERAStar BMG Labtech). Progress curves were plotted in Microsoft Excel.

***In vitro* IC_{50} assay**

IC_{50} values for bacterial GUS inhibition by drugs was assessed by combining 5 µL of 50 nM GUS (5 nM final), 5 µL of various concentrations of inhibitor, 5 µL of 10 mM 4-MUG (1 mM final), and 35 µL of assay buffer (25 mM NaCl, 25 mM HEPES, pH 7.4 final) in a black 96-well Greiner F-bottom plate. Reactions were incubated for approximately 30 min. after the addition of 4-MUG. Completed reactions were quenched with 50 uL 0.2 M sodium carbonate and end point fluorescence was determined at an excitation wavelength of 350 nm and an emission wavelength of 450 nm (PHERAStar BMG Labtech). 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

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

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 GraphPad Prism to determine the IC_{50} as described above.

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 4-MUG (500 µM final), and 35 µL of buffer (10 mM NaCl, 10 mM HEPES, pH 7.4 final). Reactions were quenched by heating the sample at 99°C for 10 min. Samples were centrifuged at 13,000 rpm for 5 min and 10 µL of supernatant was analyzed by LC-MS. Samples were injected into a Kinetex C18 column (Phenomenex, 150 mm length, 2.6 µm particle size and 100 Å pore size) and separated using the following method at a flow rate of 0.4 mL/min. Solvent A consisted of 0.1% formic acid in water, and solvent B consisted of 0.1% formic acid in acetonitrile. Mobile phase was held at 2% B for 2 min, increased to 98% B over 16 min, held at 98% B for 2 min before returning to 2% B over 1 min. Analysis was conducted using Agilent Technologies 6520 Accurate-Mass Quadruple-Time of Flight (Q-TOF) coupled with 1210 high performance liquid chromatography (HPLC). The mass spectrometry (MS) method includes electrospray ionization (ESI) under positive ion mode with the following parameters: gas temperature 300 °C, drying gas 10 L/min, nebulizer 45 lb/in^2, fragmentor 175 V, and skimmer 65V.

Ex vivo GUS activity and inhibition

Inhibition of 4-MU-G hydrolysis in human fecal extracts was determined using an endpoint assay format. Drugs were diluted to the appropriate concentrations and pre-mixed with human fecal extract prior to initiating the reaction with the fluorogenic substrate, 4-MU-G. The final reaction volume was 25 µL containing 12.5 µL water, 5 µL buffer (pH 6.5, 25 mM HEPES, 25 mM NaCl, and 1% DMSO, final), 2.5 µL human fecal extract (0.1 mg mL^{-1}, final), 2.5 µL inhibitor (10 µM, final), and 2.5 µL substrate (900 µM, final). Reaction mixtures were pre-incubated with inhibitor at 37°C for 5 min. prior to the addition of substrate. After addition of substrate, reactions were incubated at 37°C and quenched with 0.2 M sodium carbonate after 1 hour. Fluorescence was measured with an excitation wavelength of 350 nm and an emission wavelength of 450 nm (PHERAStar BMG Labtech).

Proteomics and GUS identification

All proteomics data and GUS identification was done using previous methods (see **Chapter 2**) (*85*).

GUS correlation analyses

All correlation analyses were performed in GraphPad Prism by fitting with percent inhibition.

Safety Statement

No unexpected or unusually high safety hazards were encountered.

CHAPTER 4: DIETERY FIBER AND GUT BACTERIAL β-GLUCURONIDASES

INTRODUCTION

Environmental factors heavily shape the composition of gut bacterial communities in the distal gut (*17*). In terms of frequency and volume, the human gut microbiome is most exposed to diet-derived substances compared to other ingested foreign chemicals like drugs and environmental toxins. Diet-derived substances like proteins, fats, carbohydrates, and polyphenols modulate the abundance of gut bacterial species (*100*). For example, ingestion of whey and pea protein increases the abundance of *Bifidobacterium* and *Lactobacillus* species. These particular species are heavy producers of short chain fatty acids (SCFAs), which have anti-inflammatory properties and strengthen the gut mucosal barrier (*100*). Thus, diet-derived substances play a critical role in maintaining proper gut health. Remarkably, varying the intake of diet-derived substances can cause rapid and reproducible changes in gut bacterial profiles (*18*). Delineating the mechanism by which diet-derived substances modulate gut bacterial communities can better inform the development of tailored diets that can lead to predictable human responses (*32*).

Non-digestible carbohydrate fibers play a pivotal role in maintaining proper gut health by influencing the profiles of gut bacterial communities (*100*). Dietary fibers from soybeans, wheat, and barley as well as oligosaccharides like fructooligosaccharides, galactooligosaccharides, xylooligosaccharides, and arabinooligosaccharides are not digestible by host enzymes (*100*). Fermentation of otherwise non-digestible carbohydrate fibers by gut bacteria produces SCFAs like butyrate, which is a major energy source for human colonocytes (*100*). While the human genome only encodes for ~17 glycoside hydrolase (GH) enzymes, the gut microbiota of a typical

adult encodes more than 9,000 (*101*). The rich arsenal of structurally diverse bacterial GH enzymes breaks down these non-digestible carbohydrate fibers. Fermentation of these microbiota accessible carbohydrates (MACs) also provides energy for gut bacteria, allowing for the blooming of bacteria that efficiently metabolize MACs. Therefore, the dietary intake of MACs increases both total gut bacterial load and species diversity; both of which have been linked to positive health benefits like the suppression of pathogenic species (*100*).

Most studies exploring the role of MACs in gut health have been performed using murine models (*32*). To determine the role of MAC intake on the gut microbiota in humans, Wu and co-workers recruited 30 adult participants as a part of the Food and Resulting Microbial Metabolites (FARMM) study (*32*). The purpose of the FARMM study was to assess how divergent diets including a vegan (high in MACs), omnivore (intermediate levels of MACs), and formula-based (EEN that is devoid of MACs) diets influence the composition of gut bacterial taxa in the distal gut (**Figure 4.1**). Incoming omnivores were randomly assigned to either EEN or omnivore diets in a regulated in-patient setting. Incoming vegans remained as out-patient and reported their dietary intake using provided diet intake surveys. During Days 0 – 5, all participants consumed their respective diets (termed dietary phase). During Days 6 – 8, all participants were given antibiotics and polyethylene glycol (termed Abx/PEG) to transiently reduce the concentration of bacteria in their gut, thereby allowing for a better assessment of the effect of dietary fiber on the recovery of the gut microbiota. Finally, during Day 9 – 15, the three study arms continued their respective diets (termed reconstitution phase). Each day, fecal samples were collected for shotgun metagenome sequencing.

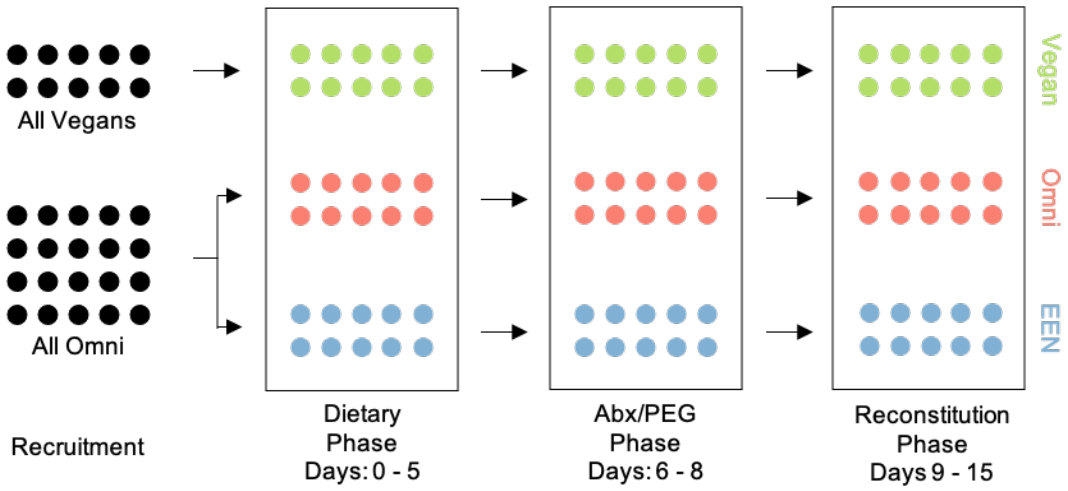


Figure 4.1 FARMM study outline. Participants in the vegan arm were vegans for at least one year prior to recruitment. Twenty omnivores (omni) were randomly assigned into two groups: omni or EEN.

To provide evidence that lack of MACs is a major contributor for diet-associated compositional alterations in the gut microbiota, Wu and co-workers quantified reads that aligned to glycoside hydrolase genes using the KEGG protein database (category 3.2.1, glycoside hydrolases) (*32*). Interestingly, plant structure degrading GH enzymes such as arabinoxylans (xylan 1,4 beta-xylanase, alpha- N-arabinofuranosidase) and pectic polysaccharides (galacturan 1,4-alpha-galacturonidase, arabinan endo-1,5-alpha-L-arabinosidase) were found to be statistically reduced during the initial dietary phase in the EEN study arm, while they observed an increase in the GH enzymes involved in the digestion of sucrose and short fructooligosaccharides (trehalose-6-phosphate hydrolase). These results were expected given the carbohydrate makeup of the formulation used in the EEN study arm, Modulen. Modulen contains simple sugars like sucrose but does not contain complex plant carbohydrates. The findings by Wu and co-workers also corroborate studies in mice where carbohydrate diets influence the expression levels of GH enzymes (*102*).

Here, we examine the FARMM dataset to understand the impact of dietary fiber levels on β-glucuronidase (GUS) gene abundance, and the gene abundance of a related GH enzyme β-galactosidase, using structure-guided approaches. Considerable structural interrogation (over 20 published structures) has shown that GUS enzymes vary widely in the architectural makeup of its active sites between gut bacterial species (*77*). This architectural variation explains interspecies variability in small molecule glucuronide and glucuronic acid (GlcA)-containing polysaccharides. To understand how the diets with varying fiber content used in the FARMM study dictate levels of gut bacterial species harboring GUS genes from a structural perspective, we mine for putative GUS sequences using metagenomic sequencing reads obtained from the original study, then align reads to obtain gene abundance data for GUS sequences, and finally, analyze the gene abundance data to assess which GUS structural classes are impacted by dietary fiber levels.

RESULTS AND DISCUSSION

Redundant GUS identification and structural categorization

Metagenomic fastq files for all samples in the FARMM cohort were obtained directly from the University of Pennsylvania School of Medicine. To interrogate the impact of MACs on GUS gene abundance, we first assembled the raw reads into scaffolds using the de Bruijn-based assembly software SPAdes v3.14.1. SPAdes was chosen because it is ideal for Illumina short sequencing reads, is well-documented, can handle metagenomic datasets, and outperforms other assembly tools (*e.g.*, IDBA-UD and Megahit) in terms of assembly size statistics and species diversity capture (*103*, *104*). Assemblies were successfully created for 328 out of the 380

samples. Assemblies could not be created for 52 samples because of low quality reads or missing data.

Prodigal v.2.6.3 was used to generate a list of predicted genes from the assembled scaffolds (*105*). Predicted genes were subsequently annotated as putative GUS sequences if they shared a sequence identity ≥ 25% identical to either *Escherichia coli* GUS (*Ec*GUS), *Bacteroides fragilis* GUS (*Bf*GUS), *Streptococcus agalactiae* (*Sa*GUS), or *Clostridium perfringens* GUS (*Cp*GUS) based on BLASTp alignment (*77*). Additionally, the query sequence must contain active site residues necessary for glucuronide recognition and processing to be annotated as a GUS sequence (**Figure 4.2a**). The average number of GUS genes identified per sample during the dietary phase is comparable to the number of GUS genes found per individual when the same analysis pipeline was applied to metagenomic sequencing data from healthy individuals from the Human Microbiome Project (HMP) (**Figure 4.2b**) (*77*). Predicted GUS sequences were binned into seven unique structural categories using previously established rubrics (*64*, *77*).

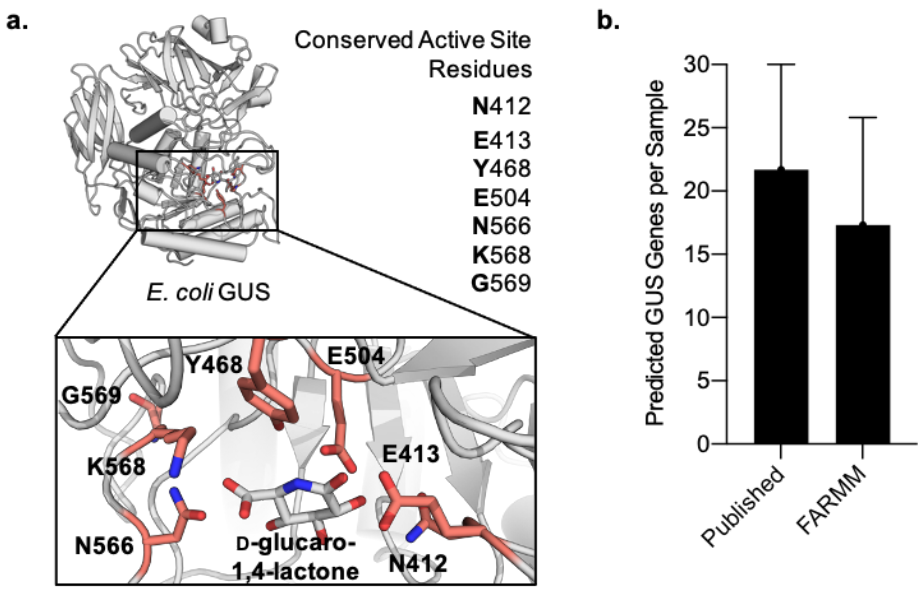


Figure 4.2 GUS rubric and statistics. (a) Key active site residues used to mine for putative GUS sequences in the FARMM study. (b) Comparison of the average number of GUS genes identified per individual in a previous study that examined HMP datasets to that found in the FARMM cohort.

Recent examination of HMP derived putative GUS sequences by Pellock and co-workers revealed that a putative GUS sequence from *Eisenbergiella tayi*, when expressed, did not process the reporter substrate *p*-nitrophenyl-β-D-GlcA *in vitro* (*106*). Instead, the expressed *E. tayi* enzyme was found to turn over the reporter substrate *p*-nitrophenyl-β-D-galacturonide (*106*). Galacturonic acid (GalA) is an epimer of GlcA and has an axial hydroxyl group at the C4 position. Structural interrogation revealed that a single arginine residue at 337 in the *E. tayi* sequence (henceforth, termed *Et*GalAse) hydrogen bonds with the axial hydroxyl group at the C4 position, which confers galacturonidase activity over GUS activity. To account for potential GalAses in the list of putative GUS sequences mined from the FARMM data, putative GUS sequences were aligned to *Et*GalAse and removed if an arginine was present at the equivalent position in *Et*GalAse in the alignment. Very few GalAses were found in the list of identified

GUS sequences. Taken together, GUS sequences were successfully identified and annotated in all samples with high quality reads.

GUS gene abundance analysis in the dietary phase

Filtered reads were mapped to predicted GUS genes to obtain gene abundance during the dietary and reconstitution phases using bowtie2 v2.4.1 and subread v2.0.0 (**Figure 4.3, Figure 4.4, Figure 4.5, Figure 4.6, Figure 4.7, and Figure 4.8**). The liquid formulation, Modulen, given to individuals in the EEN group does not contain any GlcA-containing carbohydrates. Conversely, the administered omnivore diet contains meats (*e.g.*, roasted chicken breast and spaghetti with meatballs) that are rich in GlcA-containing carbohydrates like heparin, hyaluronic acid, and chondroitin. Thus, we hypothesized that the total GUS gene abundance would diminish during the dietary phase in the EEN group when compared to the omnivore group. A statistically significant reduction in total GUS gene abundance was not observed in the EEN group when compared to the omnivore group based on linear mixed effects model. As noted by Wu and co-workers, the study timeframe (5 days) may be too narrow to fully capture changes in gut bacterial profiles during the dietary phase.

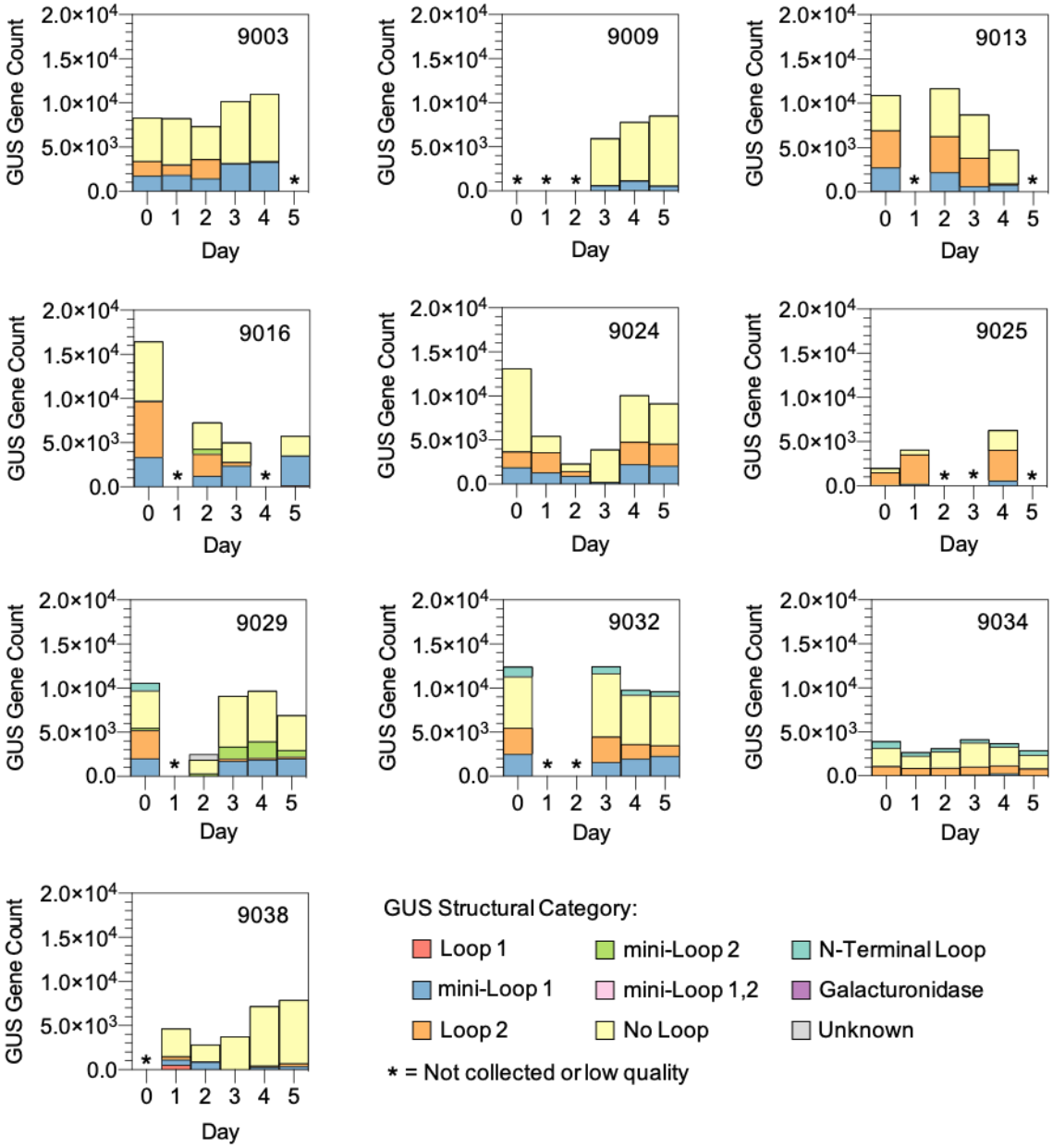


Figure 4.3 GUSome analysis of the EEN group during the dietary phase.

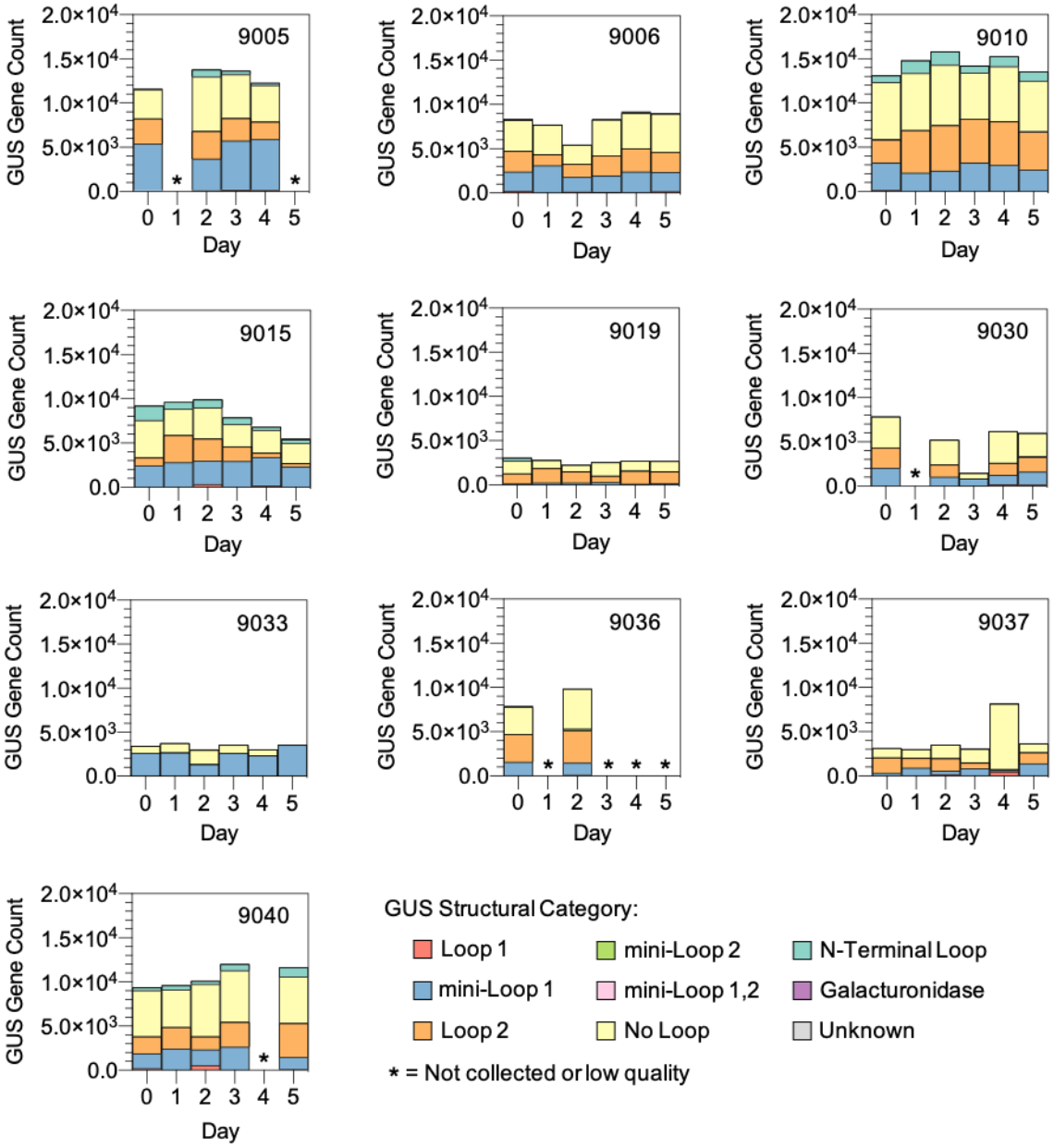


Figure 4.4 GUSome analysis of the omnivore group during the dietary phase.

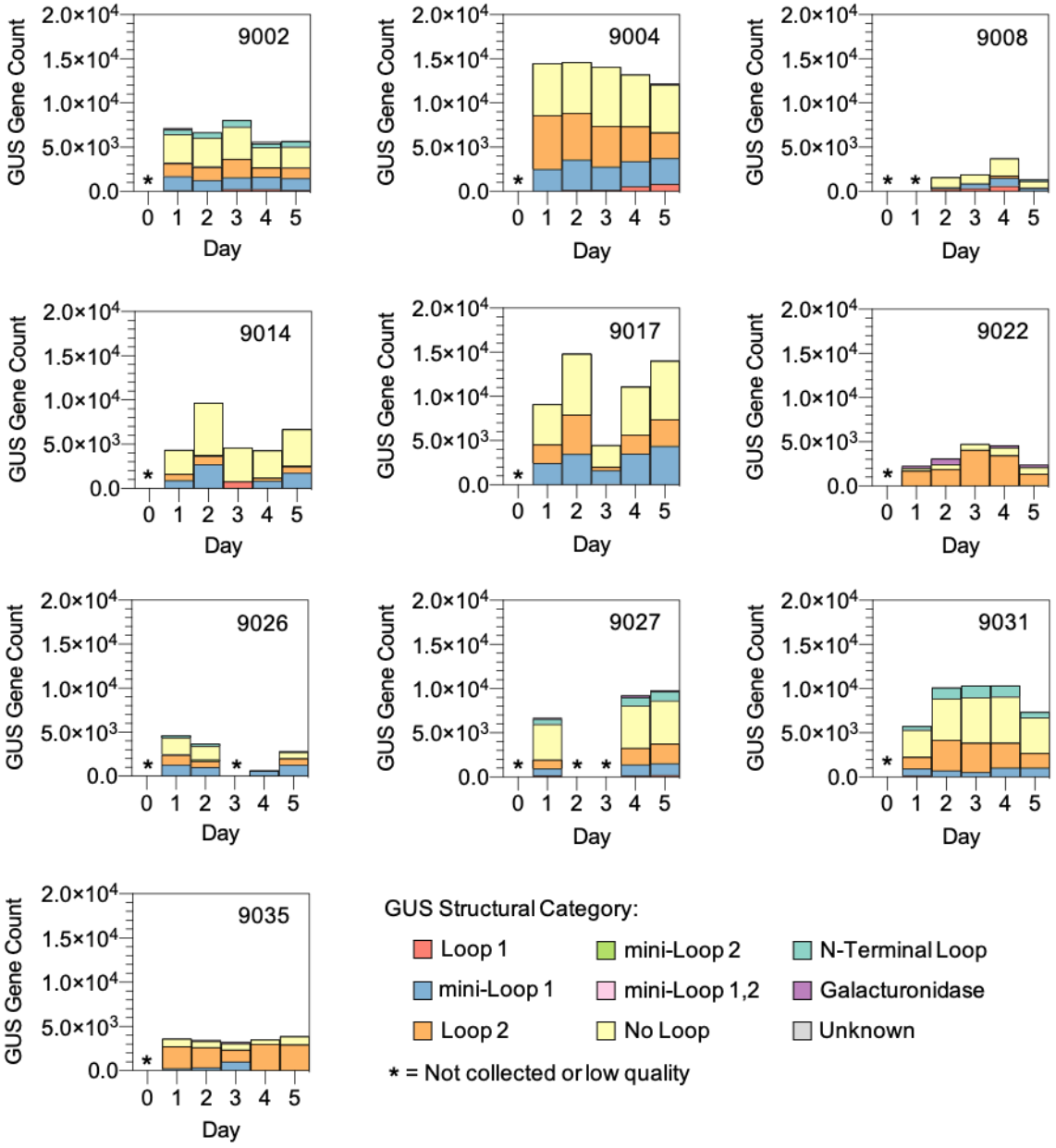


Figure 4.5 GUSome analysis of the vegan group during the dietary phase.

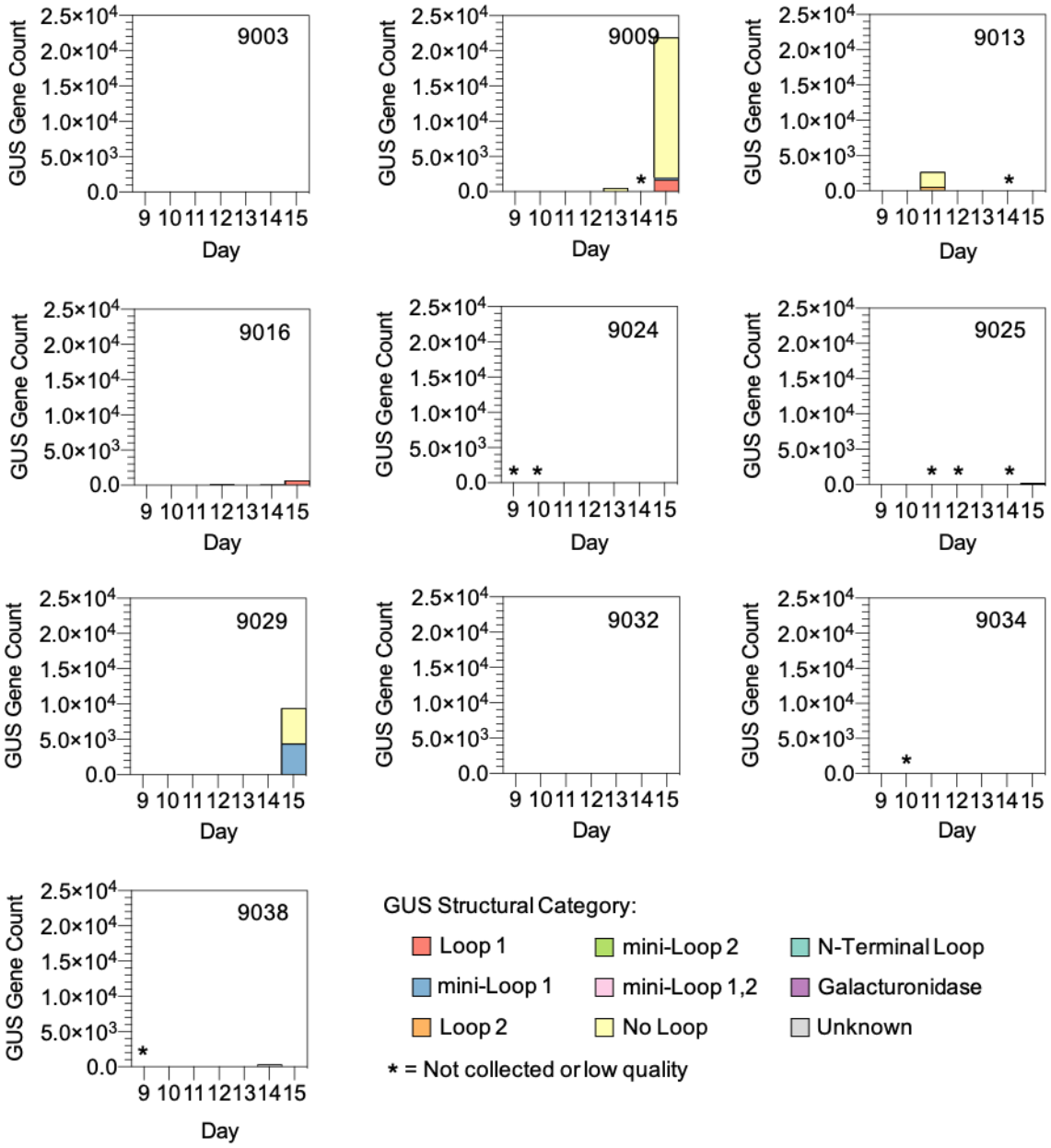


Figure 4.6 GUSome analysis of the EEN group during the reconstitution phase.

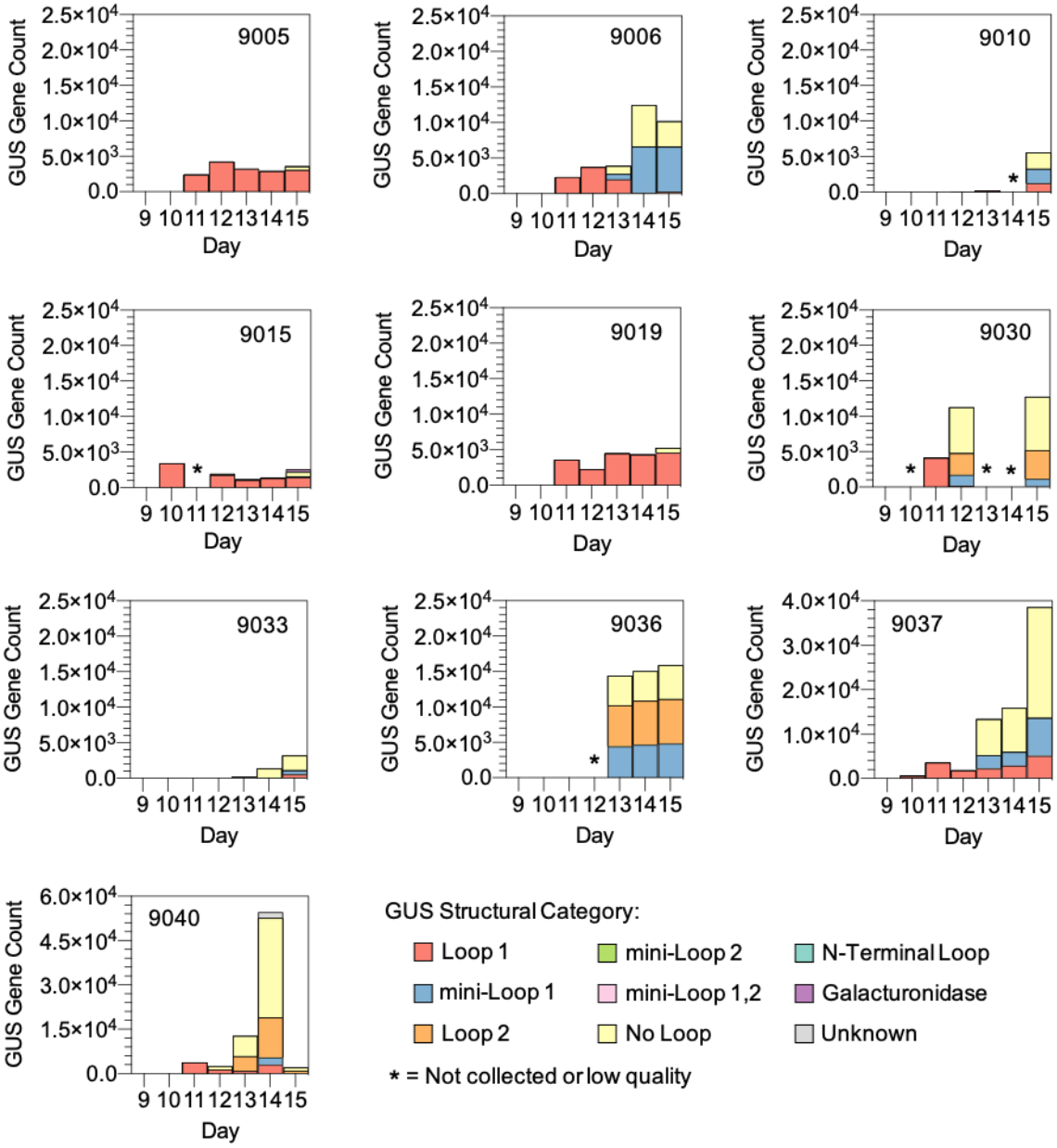


Figure 4.7 GUSome analysis of the omnivore group during the reconstitution phase.

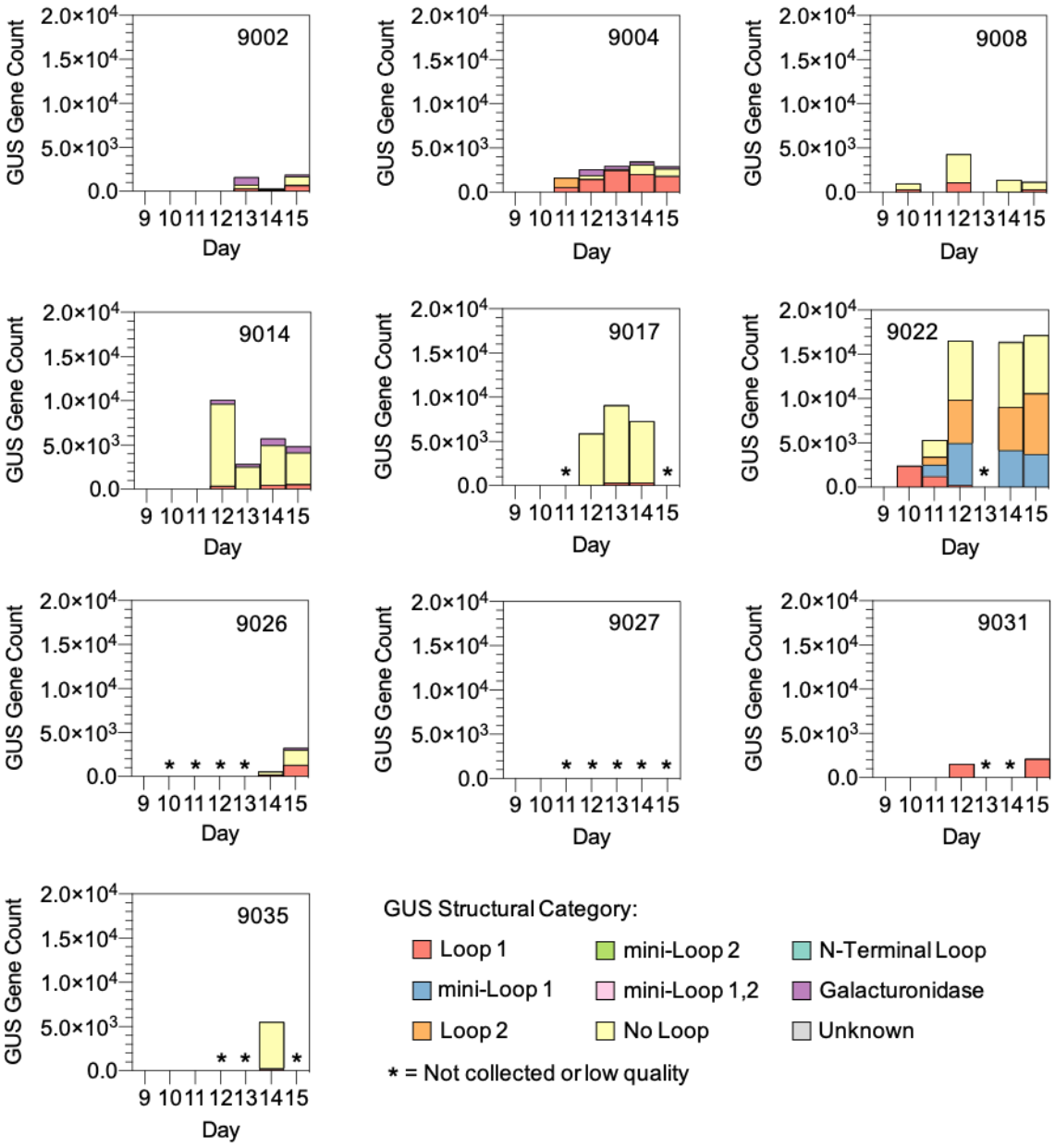


Figure 4.8 GUSome analysis of the vegan group during the reconstitution phase.

While we did not observe a statistically significant reduction in total GUS abundance during the dietary phase, we did observe a statistically significant reduction in the proportion of Loop 2 (L2) GUS in the EEN group when compared to omnivores (by 320 ± 120 reads per day; *p*-value = 0.0064 and FDR = 0.042 based on linear mixed effects model). Interestingly, a well-

characterized GUS from this GUS structural class, *Bacteroides uniformis* GUS-2 (*Bu*GUS-2), has been observed to process the complex GlcA-containing carbohydrate heparin *in vitro* (*64*, *77*). Compared to other GUS structural classes, the dimer *Bu*GUS-2 contains a relatively open, solvent accessible active site that can accommodate large polysaccharide units (*64*). Other structural classes like mini-Loop 1 (mL1), mini-Loop 2 (mL2), and No Loop (NL) have been shown to process heparin *in vitro*, but statistically significant reduction in these structural classes was not observed during the dietary phase in the EEN group. Again, extension of the dietary phase may enable further capture in the reduction of these other loop classes.

GUS gene abundance analysis in the reconstitution phase

Wu and co-workers in previous studies noted that Abx/PEG reduces the total bacterial load in the murine gut and that this is necessary for engraftment of specific bacterial colonies (*107*). This strategy was applied to the FARMM cohort. Each participant was given Abx/PEG during Days 6 – 8 to determine how dietary fiber plays a role in reconstitution of the gut microbiota. Total bacterial load was reduced 5-log units by the end of Abx/PEG treatment (*32*). During the recovery phase, gut bacterial communities that harbor GUS genes were the slowest to come back in the EEN group when compared to vegan and omnivores (**Figure 4.6, Figure 4.7,** and **Figure 4.8**). Total GUS in omnivores increases by 2071 ± 316 reads per day whereas in the EEN group, total GUS increases by 412 reads per day, which is about 25% of the omnivore rate (*p*-value = 0.0007 and FDR = 0.007 based on linear mixed effects model). This corroborates findings by Wu and co-workers, where a statistically significant increase in gut bacterial species harboring GUS genes was not observed in the EEN group during the dietary phase. A slower reconstitution of gut bacterial species harboring GUS genes when compared to an omnivore diet further highlights

the critical role of dietary fiber and its role in controlling the production of GH enzymes and hence, total bacterial load and species diversity.

β-galactosidase analysis in the reconstitution phase of EEN group

To assess whether the slow reconstitution of gut bacterial species harboring GUS genes was not spurious and simply due to slow recovery of total bacterial load, we analyzed the abundance of bacterial species harboring β-galactosidase (GAL) genes. Modulen contains milk, which is composed of simple carbohydrates including lactose, a disaccharide composed of glucose and galactose (*32*). We hypothesized that gut bacterial species harboring GAL genes would reconstitute relatively faster than those containing GUS genes. To test this hypothesis, we designed a rubric to mine for putative GAL genes using previously resolved GAL structures: *E. coli* GAL (*Ec*GAL, PDB ID 4V40) and *B. thetaiotamicron* (*Bt*GAL, PDB ID 3BGA). Predicted genes were annotated as putative GAL sequences if they shared a sequence identity ≥ 25% identical to either *Ec*GAL or *Bt*GAL based on BLASTp alignment (*77*). Additionally, the putative GAL sequence must contain previously elucidated conserved active site residues (*108*). Compared to GUS gene abundance, GAL gene abundance was much higher for all participants receiving the EEN diet (except 9038) (**Figure 4.9**), which supports the initial hypothesis that gut bacterial species with GAL genes will reconstitute faster than those with GUS genes. Other enzymes that process sugars present in Modulen may show the same behavior during the reconstitution phase. Since glucose is also a component of Modulen, glucosidase gene abundance should also be analyzed in the future.

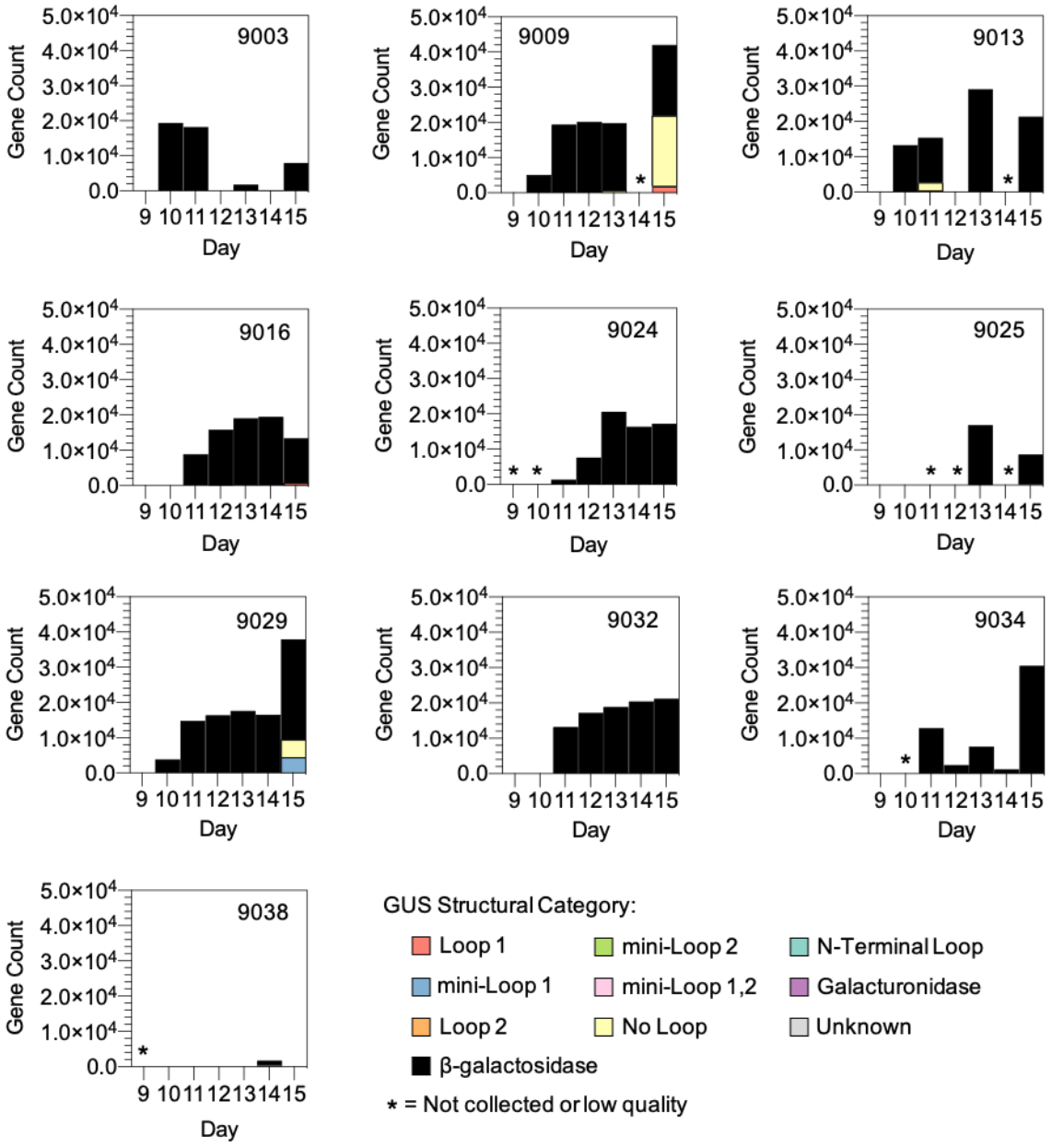


Figure 4.9 β-Galactosidase analysis of the EEN group during the reconstitution phase.

CONCLUSION

Dietary fiber, or microbiota accessible carbohydrates (MACs), plays an integral role in controlling total gut bacterial load, species diversity, and strengthening the gut mucosal lining

(*100*). Most studies to date have been performed using murine models (*32*). Wu and co-workers assessed the role of MACs in gut health using a human cohort in the FARMM study. For a diet lacking MACs (EEN), the team observed a reduction in gut bacterial communities that harbor GH genes that degrade complex plant carbohydrates. We also observed that MACs influence the levels of GUS genes using a structure-guided approach. While a reduction in total GUS gene abundance was not observed during the dietary phase, a statistically significant reduction in gut bacterial species harboring L2 GUS genes was observed in the EEN group. Previous work has shown that a L2 GUS enzyme, *Bu*GUS-2, turns over the complex GlcA-polysaccharide, heparin (*64*). In the future, the dietary phase should be extended to capture further deviations in GUS gene abundance from baseline.

Recent work has shown that select putative GUS sequences identified from the HMP do not process GlcA-containing substrates *in vitro*. For example, a substitution with a single arginine residue confers a previously annotated GUS sequence from *E. tayi* with galacturonidase activity, even though the sequence contains the same $(\alpha/\beta)_8$ TIM barrel-fold and key conserved active site residues found in GUS (*106*). This observation indicates that genome mining approaches that solely use sequence identity and residue conservation as filters often time leads to identification of sequences that are false positives. Thus, the GUS rubric will need to be continuously refined. As an alternative to genome mining approaches, activity-based proteomics is well-positioned to elucidate GUS sequences based on enzyme function and not solely on sequence information, which reduces the possibility of false positives (see **Chapter 2**). Additionally, the presence of a gene is not synonymous with the production of the functional unit of the gene, the protein; hence, proteomics can help unravel GUS enzymes that are expressed. In

the future, studies akin to FARMM should collect proteomics datasets to gain more granularity on how dietary fiber influences the expression of functionally active GUS enzymes.

As discussed in **Chapter 2**, GUS has been linked to the reactivation of the active metabolite of the chemotherapeutic irinotecan, SN-38. Modulation of the production of GUS enzymes by tailored diets could serve as a more economical alternative to small-molecule adjuvants that block GUS enzyme. Although the current dataset does not show a major reduction in gut bacterial communities harboring GUS genes during the dietary phase, future studies should expand the dietary phase to better assess how a diet lacking GlcA-containing carbohydrates impacts levels of GUS production using a proteomics approach. Pre-treatment with a EEN diet can potentially help reduce GUS expression, thereby reducing the chances of GUS-mediated GI toxicity during irinotecan treatment. The population size and duration of the study during the dietary phase needs to be increased to assess if EEN diets do indeed reduce total GUS load.

MATERIALS AND METHODS

Shotgun metagenomic data and code availability

DNA extraction, library preparation, and shotgun metagenomic sequencing for the FARMM cohort was conducted at the University of Pennsylvania School of Medicine on a Illumina HiSeq 2500 using 2 x 125 base pair chemistry (*32*). The shotgun metagenomic sequence data has been previously submitted to the Sequence Read Archive (SRA) and is available under the SRA accession number: PRJNA675301 (*32*). All custom scripts used in this study are available on GitHub at https://github.com/redinbolab.

Filtering and trimming of metagenomic reads

Shotgun metagenomic paired end reads in fastq format were pre-processed to remove adapters, low-quality reads, and human reads at the University of Pennsylvania School of Medicine prior to use in this study (*32*). Trimmomatic was used to remove adapters and low-quality reads using default parameters (*32*, *109*). Sequencing reads from all samples were aligned to the human genome version 38.v4 and the phiX genome and host reads and phiX reads were removed using Burrows-Wheeler Aligner 2 (BWA2) package (*32*, *110*).

***De novo* genome assembly of metagenomic reads**

For each sample, trimmed and filtered metagenomic paired end reads in fastq format were assembled using SPAdes v3.14.1 (*24*). The following parameter settings were used for each SPAdes assembly run: *--meta -t 12 -m 150*. All other parameters were set to default.

Gene prediction and GUS identification

Prodigal v2.6.3 was used to predict genes from the scaffolds fasta file (SPAdes output) for each sample (*105*). The following parameter settings were used for each Prodigal run: *-f gff -p meta*. All other parameters were set to default.

The resulting amino acid fasta file was then used to identify sequences as putative GUS enzymes using a previously established method and custom scripts (*77*). Each predicted sequence was aligned to representative GUS sequences that include *Bacteroides fragilis* GUS (*Bf*GUS, PDB ID 3CMG), *Escherichia coli* GUS (*Ec*GUS, UniProt ID P05804), *Streptococcus agalactiae* (*Sa*GUS, UniProt ID Q8E0N2), or *Clostridium perfringens* GUS (*Cp*GUS, UniProt ID Q8VNV4) using the Protein Basic Alignment Search Tool (BLASTP). The query sequence

must have an e-value ≤ 0.05 and sequence identity ≥ 25% identical to the representative GUS sequence. Additionally, all conserved residues had to be present and correctly aligned to the representative protein that passed the identity threshold. The conserved residues were: (*Ec*GUS) N412, E413, Y468, E504, N566, K568, G569; (*Cp*GUS) N411, E412, Y468, E505, N567, K569, G570; (*Sa*GUS) N407, E408, Y464, E501, N563, K565, G566; and (*Bf*GUS) N394, E395, Y445, E476, N547, K549, G550.

GUS structural annotation

GUS enzymes were binned into various structural categories based on previously established parameters (*77*). To bin predicted GUS sequences into unique structural categories, predicted GUS sequences were concatenated from all samples. The concatenated sequences were aligned to *Ec*GUS, *Bacteroides uniformis* GUS (2) (*Bu*GUS-2, PDB ID 5UJ6), *Bacteroides uniformis* GUS (1) (*Bu*GUS-1, PDB ID 6D1N), and *Eisenbergiella tayi* (*Et*Galase, UniProt ID A0A1E3AEY6) using Clustal Omega v1.2.2 (online webtool) to produce a multiple sequence alignment (MSA) (*111*). A custom script was used to determine the loop classification of each putative GUS sequence based on the number of residues present in the following MSA alignment regions: *Ec*GUS residue region 356-380 and the *Bu*GUS-2 residue region 429-446. See table below:

Loop Category	**Residues in *Ec*GUS region 356-380**	**Residues in *Bu*GUS-2 region 429-446**
Loop 1	> 15	< 9
mini-Loop 1	≤ 15 and ≥ 10	< 9
Loop 2	< 10	≥ 12
mini-Loop 2	< 10	≥ 9 and < 12
mini-Loop 1, mini-Loop 2	≤ 15 and ≥ 10	≥ 9 and < 12
No Loop	< 10	< 9

Putative GUS sequences were annotated as N-Terminal Loop (NTL) if residues in the MSA alignment region contained an equal or greater number of residues in *Bu*GUS-1 residue region 54-67 (*64*). Sequences were denoted as galacturonidases if in the MSA alignment, the query sequence contained an arginine at the 337-arginine position in *Et*Galase (*106*).

β-galactosidase structural annotation

Amino acid fasta file (Prodigal output) was used to identify sequences as putative β-galactosidases using previously identified key residues and custom scripts (*77*). Each predicted sequence was aligned to representative β-galactosidase sequences that include *E. coli* β-galactosidase (*Ec-β*-Gal, PDB ID 4V40) and *Bacteroides thetaiotamicron* (*Bt-β*-Gal, PDB ID 3BGA) using the Protein Basic Alignment Search Tool (BLASTP). The query sequence must have an e-value ≤ 0.05 and a sequence identity ≥ 25% identical to the representative *β*-Gal sequences. Additionally, all conserved residues had to be present and correctly aligned to the representative protein that passed the identity threshold. The conserved residues were: (*Ec-β*-Gal) E461, M502, Y503, E537, and (*Bt-β*-Gal) E484, M524, Y525, E548. Any GUS sequences that may have passed this filter were removed.

Determination of metagenomic depths of coverage for predicted genes

For each sample, trimmed and filtered metagenomic paired end reads in fastq format were mapped against the sample scaffolds fasta file (SPAdes output) using bowtie2 v2.4.1 with default parameters (*112*). The gff file (Prodigal output) was reformatted using a custom script. For each sample, read counts for each predicted gene were calculated using featureCounts, which is a component of subread v2.0.0 (*113*). The bam file (bowtie2 output) and reformatted gff file were

used as inputs and following parameter settings were used for each featureCounts run: *-t CDS -g ID*. Counts for putative GUS sequences and β-galactosidases were tabulated using custom scripts and read counts were normalized using the following equation:

$$\frac{RC}{n} \times \frac{\sum x}{N}$$

where RC is the read count for a gene in a particular sample, *n* is the total number of reads in that sample, *x* is the sum of the total number of reads in all samples, and *N* is the total number of samples (*77*).

GUS Statistical analysis

All statistical analysis was performed at the University of Pennsylvania School of Medicine. Normalized GUS gene abundance levels were analyzed using a linear mixed effects model.

CHAPTER 5: FUTURE DIRECTIONS

INHIBITION OF GUT BACTERIAL ENZYMES USING SMALL MOLECULE ADJUVANTS

Studies in the past decade have clearly shown that drugs can be metabolized by gut bacterial enzymes (**Table 5.1**). The metabolism of drugs by microbiota in the gastrointestinal (GI) tract can lead to altered drug efficacy and toxicity (*55*, *57*, *70*). Reduced drug efficacy due to gut microbial metabolism could be circumvented by measuring the activity of gut bacterial enzymes present and calculating personalized dosages; however, measuring enzymatic activity in an individual on a regular basis to account for shifts over time is labor and cost intensive. Importantly, gut microbiota-mediated metabolism can also lead to the production of toxic metabolites. Thus, strategies to either block the activity or modulate the production of gut bacterial enzymes metabolizing drugs must be developed to address toxicities and possibly to improve personalized dosing. Ongoing development of such strategies include probiotics and prebiotics, phage therapy, fecal material transfer (FMT), and small molecule adjuvants. Compared to other strategies, small molecule adjuvants that inhibit gut bacterial enzymes are ideal because they can be designed to be non-immunogenic (unlike phage therapy), dosing can be controlled (unlike prebiotics), and easy to produce reliably (unlike FMT). While other strategies to either block or modulate the activity of gut bacterial enzymes are still theoretical, small molecule adjuvants have already been successfully shown to reduce gut microbiota-mediated side effects of drugs in animal models.

Table 5.1 Drugs with validated gut bacterial processing using *in vivo* models.

Parent Drug	Primary Indication	Acting Gut Bacterial Enzyme	Chemical Modification	Ref.
Capecitabine	Colorectal Cancer	-	Deglycosylation	(*28*)
Dexamethasone	Anti-inflammatory	Steroid Desmolases	Desmolation	(*10*) (*114*)
Diclofenac	Anti-inflammatory	β-Glucuronidases	Deglycosylation	(*115*)
Digoxin	Anti-arrhythmic	Cgr2	Reduction	(*55*) (*56*)
Diltiazem	Hypertension	BT_4096	Deacetylation	(*10*)
Indomethacin	Anti-inflammatory	β-Glucuronidases	Deglycosylation	(*74*)
Irinotecan	Pancreatic & Colon Cancer	β-Glucuronidases	Deglycosylation	(*70*) (*116*)
Ketoprofen	Anti-inflammatory	β-Glucuronidases	Deglycosylation	(*74*)
Levodopa	Parkinson's Disease	Tyrosine Decarboxylases	Decarboxylation	(*57*)
Mycophenolate	Immunosuppressant	β-Glucuronidases	Deglycosylation	(*117*)
Sulfasalazine	Rheumatoid Arthritis	Azoreductases	Reduction	(*118*)

LESSONS LEARNED FROM INHIBITING GUT BACTERIAL β-GLUCURONIDASES

As discussed in previous chapters, gut bacterial β-glucuronidases (GUS) cleave glucuronic acid (GlcA) from GlcA-conjugated drugs or GlcA-conjugated drug metabolites. For example, GUS enzymes can reactivate the active metabolite of irinotecan, SN-38, by hydrolyzing off the GlcA tag from the inactive form, SN-38 glucuronide. Reactivation of SN-38 has been linked to severe, dose limiting GI toxicity (see **Chapter 2**). In 2010, Redinbo and co-workers identified a series of selective, potent, and non-lethal inhibitors which target gut bacterial GUS enzymes (*e.g.,* K_i ranging from 2 μM – 164 nM against *Ec*GUS) but not human GUS (*70*, *73*). Characterization against a broad set of GUS enzymes from various structural classes revealed that these inhibitors specifically target bacterial GUS enzymes from the Loop 1 (L1) class and not others (*63*). Detailed interrogation of L1 GUS enzymes bound with various inhibitors (*Ec*GUS and **Inhibitor 2,** PDB: 3LPF; *Ec*GUS and **Inhibitor 3**, PDB: 3LPG; and *Ee*GUS and **UNC4917**, PDB: 6BO6) also revealed that the high potency of the inhibitors is due to overlapping loops at the tetrameric interface in L1 GUS enzymes which stabilizes the inhibitors in the active site (*63*, *70*). GUS enzymes from other structural classes do not contain stabilizing units like the loop found in L1

GUS enzymes; hence, these current series of inhibitors do not efficiently inhibit non-L1 GUS enzymes. Selective inhibition of L1 GUS enzymes works well when the drug glucuronide of interest is selectively or most potently processed by L1 GUS enzymes. For example, L1 GUS enzymes most efficiently turn over SN-38 glucuronide to SN-38 when compared to GUS enzymes from other structural classes (see **Chapter 2**). Thus, L1 GUS inhibitors are ideal for reducing the reactivation of SN-38 in the GI lumen. Indeed, L1 GUS inhibitors have been shown to significantly reduce diarrhea and gut epithelial damage caused by treatment with irinotecan, the pro-drug of SN-38 (*70*, *116*). In scenarios where a drug glucuronide of interest is not processed solely or preferentially by L1 GUS enzymes, new inhibitors for other GUS structural classes must be designed.

A small molecule compound that can inhibit all GUS isoforms can ideally be used to universally block the reactivation of any drug glucuronide of interest. Designing such a molecule can be difficult as gut bacterial enzymes from different gut bacterial species vary in structure. For example, considerable structural analysis of GUS enzymes (> 20 published structures) shows that the active site architecture of GUS enzymes differs remarkably between gut bacterial species (*63*, *64*, *77*). Strain-to-strain variability in drug metabolism has been observed for other gut bacterial enzyme families which is likely due to structural variation (*55*, *57*). While GUS enzymes have been heavily discussed thus far, methods for developing such pan inhibitors for any gut bacterial enzyme families that are known to metabolize drugs is warranted (**Table 5.1**).

To successfully find pan inhibitors for a gut bacterial enzyme of interest, appropriate representatives for all known isoforms of the enzyme must be screened against a library of small molecules, which can be tedious, time consuming, and not economical. A screening platform wherein small molecules can capture overlapping chemical spaces of key isoforms of a gut

bacterial enzyme of interest is essential (**Challenge 1**). Additionally, a platform that can counter screen against other gut bacterial products is ideal to create highly selective inhibitors (**Challenge 2**). Finally, the platform must be high-throughput and be able to generate a consistent supply of gut bacterial products for screening (**Challenge 3**). Proposed solutions to these challenges are presented below.

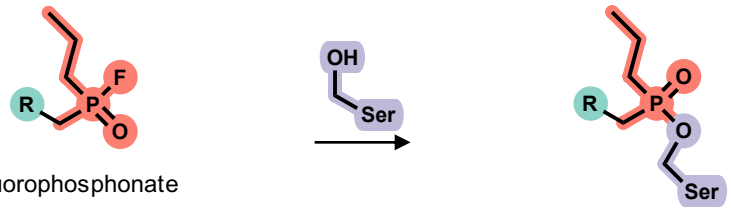
ADDRESSING CHALLENGE 1

In classic high-throughput screening (HTS) campaigns, small molecule libraries with up to millions of compounds with typical molecular weights ~500 Da are screened to find hits with promising, hopefully even nanomolar, levels of potency (*119*). Conversely, fragment-based drug discovery (FBDD) campaigns use small molecule libraries with around ~1,000 compounds with molecular weights ranging from 150-250 Da. The aim in FBDD is to "build up" weak fragment binders into strong binders by merging or elaborating fragment hits (*119*). Compared to HTS screens, FBDD yields ligands with high ligand efficiency, which is the free energy of binding divided by the number of atoms (*119*). Since fragments are small, they can form high-quality interactions within crevices and pockets in proteins. Importantly, FBDD is ideal for designing small molecules that have multi-target capabilities. In principle, fragment-based strategies can sample a larger theoretical "diversity space" than is practical through standard HTS methods. In the case of gut bacterial enzymes where structure can differ remarkably between gut bacterial species, a screening approach that maximizes coverage of chemical space is key. The small nature of fragments enables coverage of more chemical space in a target set of enzymes which is essential to develop an inhibitor that can target multiple isoforms of a gut bacterial enzyme of interest.

ADDRESSING CHALLENGE 2

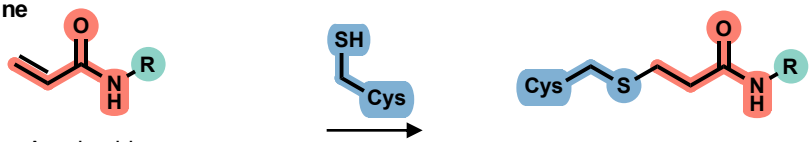
An approach that investigators are incorporating into drug discovery pipelines to assess ligand specificity is screening ligands against entire proteomes using mass spectrometry (MS). In this approach, ligands typically consist of a variable 'R' group which provides specificity, a chemical handle for enrichment (*e.g.,* alkyne tag for subsequent streptavidin enrichment), and a chemical unit for covalent capture of bound proteins. Chemical units include either a photoactivatable group, which when irradiated forms a highly reactive carbene that will react with proximal atoms, or a chemical moiety that can irreversibly form bonds with amino acid side chains (**Figure 5.1**). A succinct review of such covalent and photoactivatable ligands has been published (*120*). Cellular proteomes are incubated with each ligand (or probe, which will be used interchangeably). After cell lysis and enrichment using streptavidin beads, proteins bound to the ligand are identified using LC-MS/MS. The selectively of a ligand can be assessed by screening against an entire proteome. Importantly, many investigators have successfully screened for binders using fragment-based libraries outfitted with either photoactivatable groups or side-chain reactive groups (*121*). In terms of screening for inhibitors for gut bacterial enzymes, using a fragment-based library in tandem with MS is ideal for developing potent and selective inhibitors.

| Serine

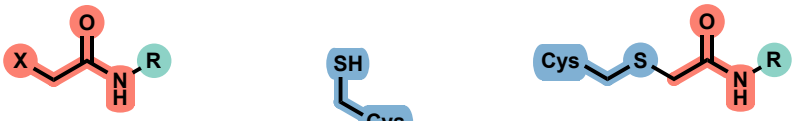


Fluorophosphonate

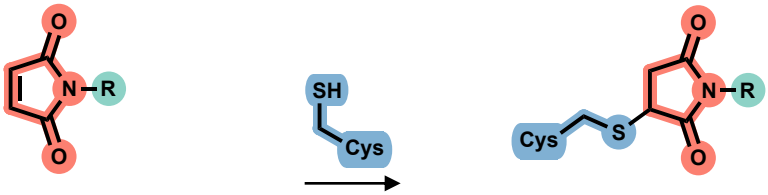
| Cysteine



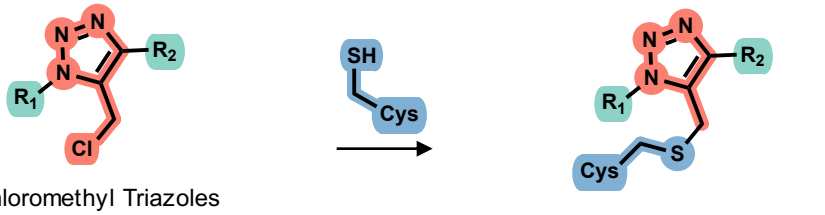
Acrylamide



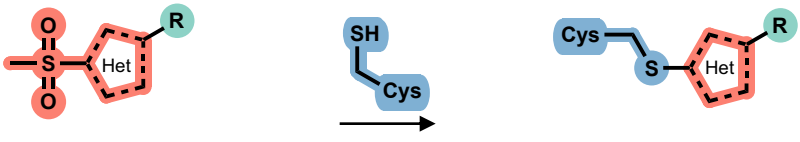
Haloacetamide (X = I, Cl)



Maleimide

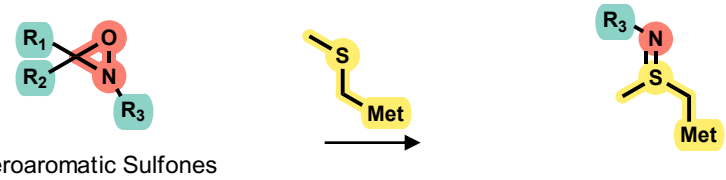


Chloromethyl Triazoles



Heteroaromatic Sulfones

| Methionine



Heteroaromatic Sulfones

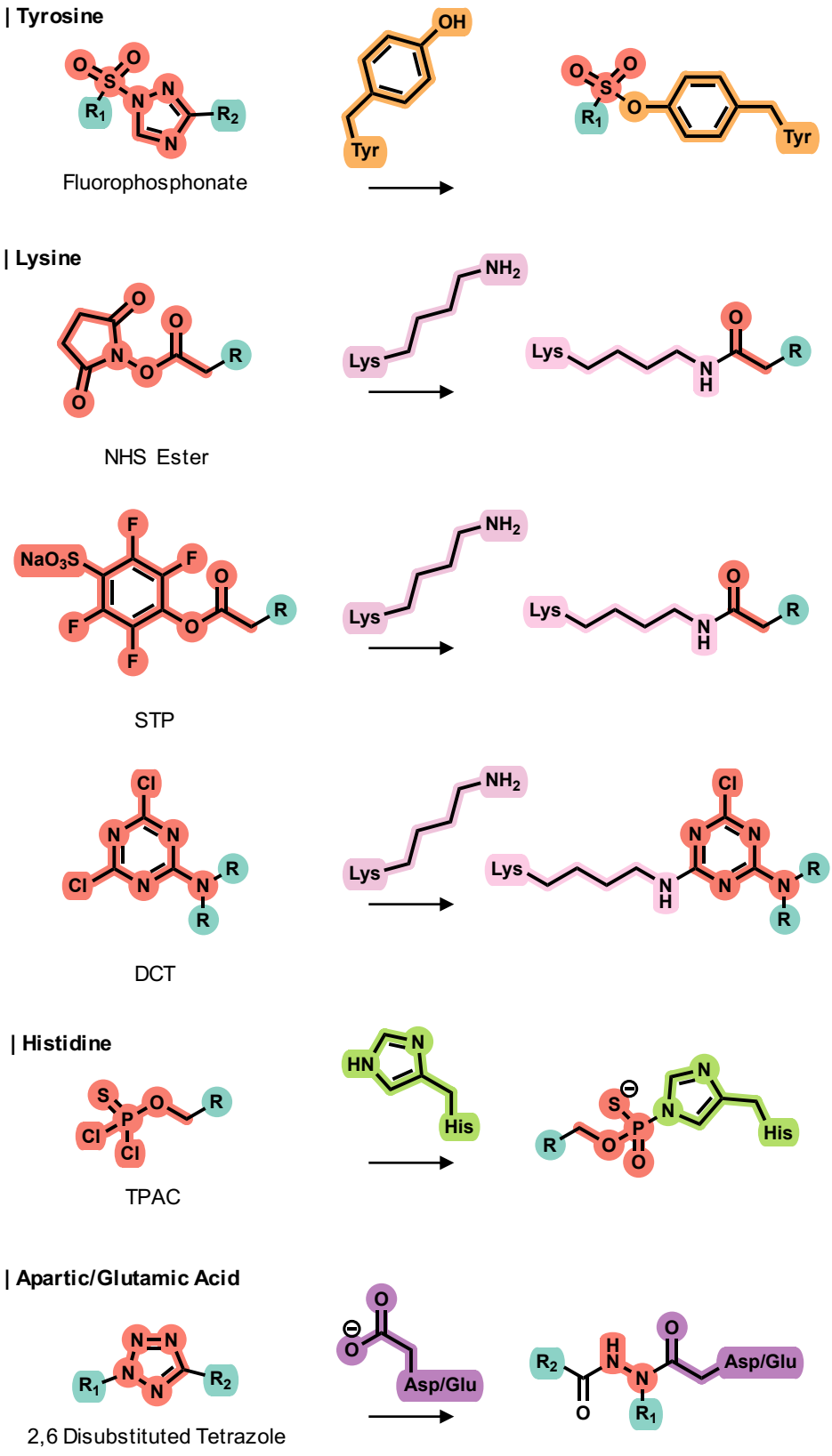


Figure 5.1 List of chemistries for covalent probes targeting specific residues.

ADDRESSING CHALLENGE 3

Unlike screening against proteomes of cell lines, where a consistent supply of biomaterial can be achieved by simply producing more cells, screening against the gut metaproteome is challenging due to the inability to culture many species. To circumvent this problem, an approach like that used by Donia and co-workers can be employed (see **Chapter 1**). DNA can be isolated from fecal material (**Figure 5.2a**). After sheering, resulting fragments are ligated into vectors which are all transformed into *E. coli*. The resulting pool of vectors can serve as an artificial proteome to screen against. More importantly, ligated gene products can be produced in high quantities for screening. After identifying a drug that is metabolized by this milieu (**Figure 5.2b**), a phenotypic screen using a library of fragment-based covalent ligands (with enrichment handles) can be performed to identify a ligand that halts the metabolism of the target drug (**Figure 5.2c**). LC-MS/MS can then be used to assess the specificity of the resulting ligand. Finally, fragments can be elaborated, and the screening process is reiterated to build a more highly specific probe against the gut bacterial enzyme of interest (**Figure 5.2d**).

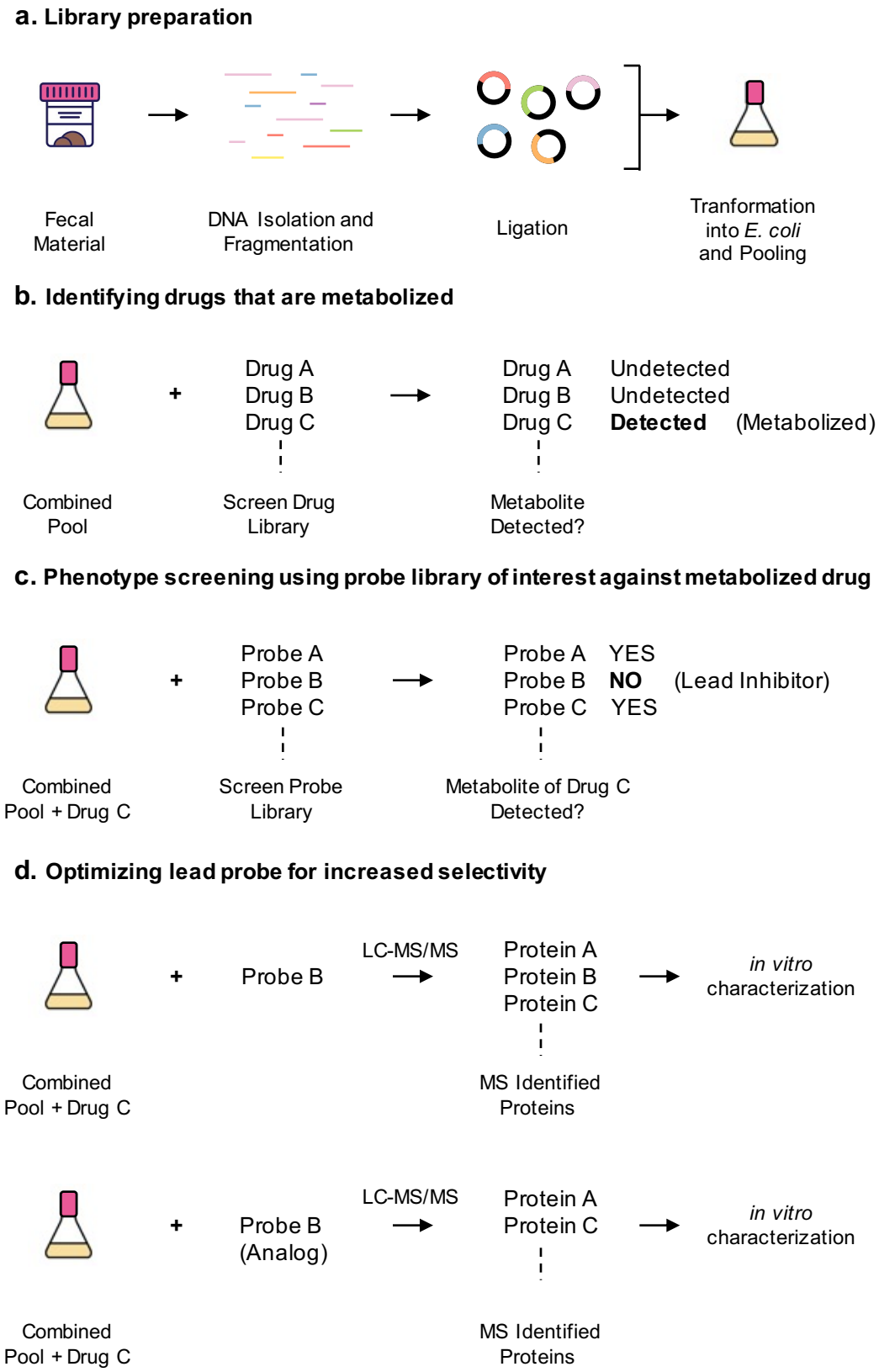


Figure 5.2 Proposed high-throughput screening strategy.

REFERENCES

1. H. Dowden, J. Munro, Trends in clinical success rates and therapeutic focus. *Nat. Rev. Drug Discov.* **18**, 495–496 (2019).

2. G. A. Van Norman, Phase II trials in drug development and adaptive trial design. *JACC Basic to Transl. Sci.* **4**, 428–437 (2019).

3. D. B. Fogel, Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemp. Clin. Trials Commun.* **11**, 156–164 (2018).

4. J. T. Atkins, G. C. George, K. Hess, K. L. Marcelo-Lewis, Y. Yuan, G. Borthakur, S. Khozin, P. Lorusso, D. S. Hong, Pre-clinical animal models are poor predictors of human toxicities in phase 1 oncology clinical trials. *Br. J. Cancer*. **123**, 1496–1501 (2020).

5. D. E. Golan, E. J. Armstrong, A. W. Armstrong, *Principles of pharmacology: the pathophysiologic basis of drug therapy* (Wolters Kluwer, Philadelphia, ed. 4th, 2017).

6. M. T. Santini, G. Rainaldi, Three-dimensional spheroid model in tumor biology. *Pathobiology*. **67**, 148–157 (1999).

7. S. M. Tsunoda, C. Gonzales, A. K. Jarmusch, J. D. Momper, J. D. Ma, Contribution of the gut microbiome to drug disposition, pharmacokinetic and pharmacodynamic variability. *Clin. Pharmacokinet.*, 1–14 (2021).

8. D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).

9. R. Berlemont, A. C. Martiny, Glycoside hydrolases across environmental microbial communities. *PLoS Comput. Biol.* **12**, 1–16 (2016).

10. M. Zimmermann, M. Zimmermann-Kogadeeva, R. Wegmann, A. L. Goodman, Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature*. **570**, 462–467 (2019).

11. M. Okour, R. C. Brundage, Modeling enterohepatic circulation. *Curr. Pharmacol. Reports*. **3**, 301–313 (2017).

12. N. Koppel, V. M. Rekdal, E. P. Balskus, Chemical transformation of xenobiotics by the human gut microbiota. *Science.* **356**, 1–11 (2017).

13. R. Sender, S. Fuchs, R. Milo, Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol.* **14**, 1–14 (2016).

14. A. K. Mangalam, D. N. Krementsov, R. K. Mondal, A. Kumar, M. A. Malla, A. Dubey, S.

Yadav, A. Hashem, E. F. Abd Allah, Exploring the human microbiome: the potential future role of next-generation sequencing in disease diagnosis and treatment. *Front. Immunol.* **9**, 1–23 (2019).

15. S. Carding, K. Verbeke, D. T. Vipond, B. M. Corfe, L. J. Owen, Dysbiosis of the gut microbiota in disease. *Microb. Ecol. Heal. Dis.* **26**, 1–9 (2015).

16. G. Falony, M. Joossens, S. Vieira-Silva, J. Wang, Y. Darzi, K. Faust, A. Kurilshikov, M. J. Bonder, M. Valles-Colomer, D. Vandeputte, R. Y. Tito, S. Chaffron, L. Rymenans, C. Verspecht, L. De Sutter, G. Lima-Mendez, K. D'hoe, K. Jonckheere, D. Homola, R. Garcia, E. F. Tigchelaar, L. Eeckhaudt, J. Fu, L. Henckaerts, A. Zhernakova, C. Wijmenga, J. Raes, Population-level analysis of gut microbiome variation. *Science.* **352**, 560–564 (2016).

17. F. Carbonero, C. Dutta, V. K. Gupta, S. Paul, Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. *Front. Microbiol.* **8**, 1–16 (2017).

18. L. A. David, C. F. Maurice, R. N. Carmody, D. B. Gootenberg, J. E. Button, B. E. Wolfe, A. V Ling, A. S. Devlin, Y. Varma, M. A. Fischbach, S. B. Biddinger, R. J. Dutton, P. J. Turnbaugh, Diet rapidly and reproducibly alters the human gut microbiome. *Nature*. **505**, 559–563 (2014).

19. S. Lee, J. Sung, J. Lee, G. Ko, Comparison of the gut microbiotas of healthy adult twins living in south korea and the united states. *Appl. Environ. Microbiol.* **77**, 7433–7437 (2011).

20. J. Gregory Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. Gonzalez Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, R. Knight, QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*. **7**, 335–336 (2010).

21. A. Prodanid, V. Tremaroli, H. Brolin, A. H. Zwinderman, M. Nieuwdorp, E. Levin, Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *PLoS One*. **15**, 1–19 (2020).

22. J. M. Janda, S. L. Abbott, 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J. Clin. Microbiol.* **45**, 2761–2764 (2007).

23. F. Beghini, L. J. Mciver, A. Blanco-Miguez, L. Dubois, F. Asnicar, S. Maharjan, A. Maillyan, P. Manghi, M. Scholz, A. M. Thomas, M. Valles-Colomer, G. Weingart, Y. Zhang, M. Zolfo, C. Huttenhower, E. A. Franzosa, N. Segata, Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife*. **10**, 1–42 (2021).

24. A. Bankevich, S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V Pyshkin, A. V Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, P. A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

25. K. Sato, Y. Sakakibara, MetaVelvet-SL: an extension of the velvet assembler to a de novo metagenomic assembler utilizing supervised learning. *DNA Res.* **22**, 69–77 (2015).

26. E. A. Franzosa, L. J. Mciver, G. Rahnavard, L. R. Thompson, M. Schirmer, G. Weingart, K. S. Lipson, R. Knight, J. G. Caporaso, N. Segata, C. Huttenhower, Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods*. **15**, 962–968 (2018).

27. L. M. Proctor, H. H. Creasy, J. M. Fettweis, J. Lloyd-Price, A. Mahurkar, W. Zhou, G. A. Buck, M. P. Snyder, J. F. Strauss, G. M. Weinstock, O. White, C. Huttenhower, The integrative human microbiome project. *Nature*. **569**, 641–648 (2019).

28. B. Javdan, J. G. Lopez, P. Chankhamjon, Y. C. J. Lee, R. Hull, Q. Wu, X. Wang, S. Chatterjee, M. S. Donia, Personalized mapping of drug metabolism by the human gut microbiome. *Cell*. **181**, 1661–1679 (2020).

29. J. Li, H. Jia, X. Cai, H. Zhong, Q. Feng, S. Sunagawa, M. Arumugam, J. Roat Kultima, E. Prifti, T. Nielsen, A. Sierakowska Juncker, C. Manichanh, B. Chen, W. Zhang, F. Levenez, J. Wang, X. Xu, L. Xiao, S. Liang, D. Zhang, Z. Zhang, W. Chen, H. Zhao, J. Yousuf Al-Aama, S. Edris, H. Yang, J. Wang, T. Hansen, H. Bjørn Nielsen, S. Brunak, K. Kristiansen, F. Guarner, O. Pedersen, S. Arabia, J. D. Watson, An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).

30. Y. Zou, W. Xue, G. Luo, Z. Deng, P. Qin, R. Guo, H. Sun, Y. Xia, S. Liang, Y. Dai, D. Wan, R. Jiang, L. Su, Q. Feng, Z. Jie, T. Guo, Z. Xia, C. Liu, J. Yu, Y. Lin, S. Tang, G. Huo, X. Xu, Y. Hou, X. Liu, J. Wang, H. Yang, K. Kristiansen, J. Li, H. Jia, L. Xiao, 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. *Nat. Biotechnol.* **37**, 179–185 (2019).

31. A. Almeida, S. Nayfach, M. Boland, F. Strozzi, M. Beracochea, Z. Jason Shi, K. S. Pollard, E. Sakharova, D. H. Parks, P. Hugenholtz, N. Segata, N. C. Kyrpides, R. D. Finn, A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* **39**, 105–114 (2020).

32. C. Tanes, K. Bittinger, Y. Gao, E. S. Friedman, L. Nessel, U. R. Paladhi, L. Chau, E. Panfen, M. A. Fischbach, J. Braun, R. J. Xavier, C. B. Clish, H. Li, F. D. Bushman, J. D. Lewis, G. D. Wu, Role of dietary fiber in the recovery of the human gut microbiome and its metabolome. *Cell Host Microbe*. **29**, 394–407 (2021).

33. S. Bashiardes, G. Zilberman-Schapira, E. Elinav, Use of metatranscriptomics in microbiome research. *Bioinforma. Biol. Insight*. **10**, 19–25 (2016).

34. E. M. Glass, J. Wilkening, A. Wilke, D. Antonopoulos, F. Meyer, Using the

metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb. Protoc.* **2010** (2010).

35. L. Jiang, M. Wang, S. Lin, R. Jian, X. Li, J. Chan, G. Dong, H. Fang, A. E. Robinson, GTEx Consortium, M. P. Snyder, A quantitative proteome map of the human body. *Cell*. **183**, P269–P283 (2020).

36. M. Mann, N. L. Kelleher, Precision proteomics: The case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci.* **105**, 18132–18138 (2008).

37. K. Cheng, Z. Ning, X. Zhang, L. Li, B. Liao, J. Mayne, A. Stintzi, D. Figeys, MetaLab: an automated pipeline for metaproteomic data analysis. *Microbiome*. **5**, 1–10 (2017).

38. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–72 (2008).

39. M. D. Mayers, C. Moon, G. S. Stupp, A. I. Su, D. W. Wolan, Quantitative metaproteomics and activity-based probe enrichment reveals significant alterations in protein expression from a mouse model of inflammatory bowel disease. *J. Proteome Res.* **16**, 1014–1026 (2017).

40. C. Whidbey, N. C. Sadler, R. N. Nair, R. F. Volk, A. J. Deleon, L. M. Bramer, S. J. Fansler, J. R. Hansen, A. K. Shukla, J. K. Jansson, B. D. Thrall, Wrig, A probe-enabled approach for the selective isolation and characterization of functionally active subpopulations in the gut microbiome. *J. Am. Chem. Soc.* **141**, 42–47 (2019).

41. B. Parasar, H. Zhou, X. Xiao, Q. Shi, I. L. Brito, P. V. Chang, Chemoproteomic profiling of gut microbiota-associated bile salt hydrolase activity. *ACS Cent. Sci.* **5**, 867–873 (2019).

42. V. Marx, A dream of single-cell proteomics. *Nat. Methods*. **16**, 809–812 (2019).

43. J. Zierer, M. A. Jackson, G. Kastenmüller, M. Mangino, T. Long, A. Telenti, R. P. Mohney, K. S. Small, J. T. Bell, C. J. Steves, A. M. Valdes, T. D. Spector, C. M., S. Burnham, The fecal metabolome as a functional readout of the gut microbiome. *Nat. Genet.* **50**, 790–795 (2018).

44. H. Krogh Pedersen, V. Gudmundsdottir, H. Bjørn Nielsen, T. Hyotylainen, T. Nielsen, B. A. H. Jensen, K. Forslund, F. Hildebrand, E. Prifti, G. Falony, E. Le Chatelier, F. Levenez, J. Doré, I. Mattila, D. R. Plichta, P. Pöhö, L. I. Hellgren, M. Arumugam, S. Sunagawa, S. Vieira-Silva, T. Jørgensen, J. B. Holm, K. Trošt, Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature*. **535**, 376–381 (2016).

45. E. C. Lavelle, S. Raghavan, C. Patrick McEntee, M. A. Hermoso, D. Parada Venegas, M. K. De la Fuente, G. Landskron, M. Julieta González, R. Quera, G. Dijkstra, H. J. M Harmsen, K. Nico Faber, Short chain fatty acids (SCFAs)-mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. *Front. Immunol.*

10, 1–16 (2019).

46. S. R. Gill, M. Pop, R. T. Deboy, P. B. Eckburg, P. J. Turnbaugh, B. S. Samuel, J. I. Gordon, D. A. Relman, C. M. Fraser-Liggett, K. E. Nelson, Metagenomic analysis of the human distal gut microbiome. *Science.* **312**, 1355–1359 (2006).

47. M. P. Francino, Antibiotics and the human gut microbiome: dysbioses and accumulation of resistances. *Front. Microbiol.* **6**, 1–11 (2016).

48. Q. Le Bastard, G. A. Al-Ghalith, M. Gr Egoire, G. Chapelet, F. Javaudin, E. Dailly, E. Batard, D. Knights, E. Montassier, Systematic review: human gut dysbiosis induced by non-antibiotic prescription medications. *Aliment. Pharmacol. Ther.* **47**, 332–345 (2018).

49. S. Banerjee, G. Sindberg, F. Wang, J. Meng, U. Sharma, L. Zhang, P. Dauer, C. Chen, J. Dalluge, T. Johnson, S. Roy, Opioid-induced gut microbial disruption and bile dysregulation leads to gut barrier compromise and sustained systemic inflammation. *Mucosal Immunol.* **9**, 1418–1428 (2016).

50. A. Napotilano, S. Miller, A. Nicholls, D. Baker, S. Van Horn, E. Thomas, D. Rajpal, A. Spivak, J. R. Brown, D. Nunez, Novel gut-based pharmacology of metformin in patients with type 2 diabetes mellitus. *PLoS One*. **9**, 1–14 (2014).

51. J. M. Ridlon, J. M. P. Alves, P. B. Hylemon, J. S. Bajaj, Cirrhosis, bile acids and gut microbiota: unraveling a complex relationship. *Gut Microbes*. **4**, 382–387 (2013).

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

53. K. Z. Coyte, S. Rakoff-Nahoum, Understanding competition and cooperation within the mammalian gut microbiome. *Cell Curr. Biol.* **29**, R538–R544 (2019).

54. A. L. Goodman, N. P. McNulty, Y. Zhao, D. Leip, R. D. Mitra, C. A. Lozupone, R. Knight, J. I. Gordon, Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe*. **6**, 279–289 (2009).

55. H. J. Haiser, D. B. Gootenberg, K. Chatman, G. Sirasani, E. P. Balskus, P. J. Turnbaugh, Predicting and manipulating cardiac drug inactivation by the human gut bacterium eggerthella lenta. *Science.* **341**, 295–298 (2013).

56. N. Koppel, J. E. Bisanz, M.-E. Pandelia, P. J. Turnbaugh, E. P. Balskus, Discovery and characterization of a prevalent human gut bacterial enzyme sufficient for the inactivation of a family of plant toxins. *Elife*. **7**, 1–32 (2018).

57. V. M. Rekdal, E. N. Bess, J. E. Bisanz, P. J. Turnbaugh, E. P. Balskus, Discovery and inhibition of an interspecies gut bacterial pathway for levodopa metabolism. *Science.* **364**, 1–8 (2019).

58. J. Winter, A. Cerone-McLernon, S. O'Rourke, L. Ponticorvo, V. D. Bokkenheuser, Formation of 20β-dihydrosteroids by anaerobic bacteria. *J. Steroid Biochem.* **17**, 661–667 (1982).

59. M. A. Henson, P. Phalak, Suboptimal community growth mediated through metabolite crossfeeding promotes species diversity in the gut microbiota. *PLoS Comput. Biol.* **14**, 1–21 (2018).

60. Y. Zhou, C. Joubran, L. Miller-Vedam, V. Isabella, A. Nayar, S. Tentarelli, A. Miller, Thinking outside the "bug": a unique assay to measure intracellular drug penetration in gram-negative bacteria. *Anal. Chem.* **87**, 3579–3584 (2015).

61. S. Zhao, J. W. Adamiak, V. Bonifay, J. Mehla, H. I. Zgurskaya, D. S. Tan, Defining new chemical space for drug penetration into gram-negative bacteria. *Nat. Chem. Biol.* **16**, 1293–1302 (2020).

62. G. J. Dutton, *Glucuronic Acid, Free and Combined, Biochemistry, Pharmacology, and Medicine* (Academic Press, New York, 1966).

63. K. A. Biernat, samuel J. Pellock, A. Bhatt, M. M. Bivins, W. G. Walton, B. Ngoc tran, L. Wei, M. C. Snider, A. Cesmat, A. Tripathy, D. A. Erie, M. R. Redinbo, Structure, function, and inhibition of drug reactivating human gut microbial β-glucuronidases. *Sci. Rep.* **9**, 1–15 (2018).

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

65. S. J. Pellock, B. C. Creekmore, W. G. Walton, N. Mehta, K. A. Biernat, A. P. Cesmat, Y. Ariyarathna, Z. D. Dunn, B. Li, J. Jin, L. I. James, M. R. Redinbo, Gut microbial β-glucuronidase inhibition via catalytic cycle interception. *ACS Cent. Sci.* **4**, 868–879 (2018).

66. R. A. Goodnow, C. E. Dumelin, A. D. Keefe, DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat. Rev. Drug Discov.* **16**, 131–147 (2017).

67. J. A. DiMasi, H. G. Grabowski, R. W. Hansen, Innovation in the pharmaceutical industry: new estimates of R&D costs. *J. Health Econ.* **47**, 20–33 (2016).

68. P. Morgan, D. G. Brown, S. Lennard, M. J. Anderton, J. Carl Barrett, U. Eriksson, M. Fidock, B. Hamrén, A. Johnson, R. E. March, J. Matcham, J. Mettetal, D. J. Nicholls, S. Platz, S. Rees, M. A. Snowden, M. N. Pangalos, Impact of a five-dimensional framework on R&D productivity at AstraZeneca. *Nature*. **17**, 167–181 (2018).

69. K. N. Lam, M. Alexander, P. J. Turnbaugh, Precision medicine goes microscopic: engineering the microbiome to improve drug outcomes. *Cell Host Microbe*. **26**, 22–34 (2019).

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

71. S. P. van Kessel, A. K. Frye, A. O. El-Gendy, M. Castejon, A. Keshavarzian, G. van Dijk, S. El Aidy, Gut bacterial tyrosine decarboxylases restrict levels of levodopa in the treatment of parkinson's disease. *Nat. Commun.* **10**, 1–11 (2019).

72. G. J. Dutton, *Glucuronidation of Drugs and Other Compounds* (CRC press, Boca Raton, 1980).

73. B. D. Wallace, A. B. Roberts, R. M. Pollet, J. D. Ingle, K. A. Biernat, S. J. Pellock, M. K. Venkatesh, L. Guthrie, S. K. O'Neal, S. J. Robinson, M. Dollinger, E. Figueroa, S. R. McShane, R. D. Cohen, J. Jin, S. V. Frye, W. C. Zamboni, C. Pepe-Ranney, S. Mani, L. Kelly, M. R. Redinbo, Structure and inhibition of microbiome β-glucuronidases essential to the alleviation of cancer drug toxicity. *Chem. Biol.* **22**, 1238–1249 (2015).

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

75. D. Kweekel, H. J. Guchelaar, H. Gelderblom, Clinical and pharmacogenetic factors associated with irinotecan toxicity. *Cancer Treat. Rev.* **34**, 656–669 (2008).

76. L. Guthrie, S. Gupta, J. Daily, L. Kelly, Human microbiome signatures of differential colorectal cancer drug metabolism. *NPJ Biofilms Microbiomes*. **3**, 1–8 (2017).

77. R. M. Pollet, E. H. D'Agostino, W. G. Walton, Y. Xu, M. S. Little, K. A. Biernat, S. J. Pellock, L. M. Patterson, B. C. Creekmore, H. N. Isenberg, R. R. Bahethi, A. P. Bhatt, J. Liu, R. Z. Gharaibeh, M. R. Redinbo, An atlas of β-glucuronidases in the human intestinal microbiome. *Structure*. **25**, 967–977 (2017).

78. X. Zhang, D. Figeys, Perspective and guidelines for metaproteomics in microbiome studies. *J. Proteome Res.* **18**, 2370–2380 (2019).

79. L. Wu, J. Jiang, Y. Jin, W. W. Kallemeijn, C. L. Kuo, M. Artola, W. Dai, C. Van Elk, M. Van Eijk, G. A. Van Der Marel, J. D. C. Codée, B. I. Florea, J. M. F. G. Aerts, H. S. Overkleeft, G. J. Davies, Activity-based probes for functional interrogation of retaining β-glucuronidases. *Nat. Chem. Biol.* **13**, 867–873 (2017).

80. V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, 490–495 (2014).

81. B. Mesuere, T. Willems, F. Van Der Jeugt, B. Devreese, P. Vandamme, P. Dawyndt, UniPept web services for metaproteomics analysis. *Bioinformatics*. **32**, 1746–1748 (2016).

82. S. P. Schröder, C. De Boer, N. G. S. McGregor, R. J. Rowland, O. Moroz, E. Blagova, J. Reijngoud, M. Arentshorst, D. Osborn, M. D. Morant, E. Abbate, M. A. Stringer, K. B. R. M. Krogh, L. Raich, C. Rovira, J. G. Berrin, G. P. Van Wezel, A. F. J. Ram, B. I. Florea, G. A. Van Der Marel, J. D. C. Codée, K. S. Wilson, L. Wu, G. J. Davies, H. S. Overkleeft, Dynamic and functional profiling of xylan-degrading enzymes in aspergillus secretomes using activity-based probes. *ACS Cent. Sci.* **5**, 1067–1078 (2019).

83. J. Jiang, C. L. Kuo, L. Wu, C. Franke, W. W. Kallemeijn, B. I. Florea, E. Van Meel, G. A. Van Der Marel, J. D. C. Codée, R. G. Boot, G. J. Davies, H. S. Overkleeft, J. M. F. G. Aerts, Detection of active mammalian GH31 α-glucosidases in health and disease using in-class, broad-spectrum activity-based probes. *ACS Cent. Sci.* **2**, 351–358 (2016).

84. K. Y. Li, J. Jiang, M. D. Witte, W. W. Kallemeijn, W. E. Donker-Koopman, R. G. Boot, J. M. F. G. Aerts, J. D. C. Codée, G. A. Van Der Marel, H. S. Overkleeft, Exploring functional cyclophellitol analogues as human retaining beta-glucosidase inhibitors. *Org. Biomol. Chem.* **12**, 7786–7791 (2014).

85. P. B. Jariwala, S. J. Pellock, D. Goldfarb, E. W. Cloer, M. Artola, J. B. Simpson, A. P. Bhatt, W. G. Walton, L. R. Roberts, M. B. Major, G. J. Davies, H. S. Overkleeft, M. R. Redinbo, Discovering the microbial enzymes driving drug toxicity with activity-based protein profiling. *ACS Chem. Biol.* **15**, 217–225 (2019).

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

87. S. Davis, P. D. Charles, L. He, P. Mowlds, B. M. Kessler, R. Fischer, Expanding proteome coverage with CHarge Ordered Parallel Ion aNalysis (CHOPIN) combined with broad specificity proteolysis. *J. Proteome Res.* **16**, 1288–1299 (2017).

88. J. A. Vizcaíno, A. Csordas, N. Del-Toro, J. A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol, F. Reisinger, T. Ternent, Q. W. Xu, R. Wang, H. Hermjakob, 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, D447–D456 (2016).

89. A. Bateman, M. J. Martin, C. O'Donovan, M. Magrane, E. Alpi, R. Antunes, B. Bely, M. Bingley, C. Bonilla, R. Britto, B. Bursteinas, H. Bye-AJee, A. Cowley, A. Da Silva, M. De Giorgi, T. Dogan, F. Fazzini, L. G. Castro, L. Figueira, P. Garmiri, G. Georghiou, D. Gonzalez, E. Hatton-Ellis, W. Li, W. Liu, R. Lopez, J. Luo, Y. Lussi, A. MacDougall, A. Nightingale, B. Palka, K. Pichler, D. Poggioli, S. Pundir, L. Pureza, G. Qi, S. Rosanoff, R. Saidi, T. Sawford, A. Shypitsyna, E. Speretta, E. Turner, N. Tyagi, V. Volynkin, T. Wardell, K. Warner, X. Watkins, R. Zaru, H. Zellner, I. Xenarios, L. Bougueleret, A. Bridge, S. Poux, N. Redaschi, L. Aimo, G. ArgoudPuy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M. C. Blatter, B. Boeckmann, J. Bolleman, E. Boutet, L. Breuza, C. Casal-Casas, E. De Castro, E. Coudert, B. Cuche, M. Doche, D. Dornevil, S. Duvaud, A. Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, F. Jungo, G. Keller, V. Lara, P. Lemercier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. Neto, N. Nouspikel, S. Paesano, I. Pedruzzi, S. Pilbout, M. Pozzato, M. Pruess, C. Rivoire, B. Roechert, M.

Schneider, C. Sigrist, K. Sonesson, S. Staehli, A. Stutz, S. Sundaram, M. Tognolli, L. Verbregue, A. L. Veuthey, C. H. Wu, C. N. Arighi, L. Arminski, C. Chen, Y. Chen, J. S. Garavelli, H. Huang, K. Laiho, P. McGarvey, D. A. Natale, K. Ross, C. R. Vinayaka, Q. Wang, Y. Wang, L. S. Yeh, J. Zhang, UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).

90. P. Rice, I. Longden, A. Bleasby, EMBOSS: the european molecular biology open software suite. *Trends Genet.* **16**, 276–277 (2000).

91. X. M. Samantha Ervin, H. Li, X. Lauren Lim, L. R. Roberts, X. Liang, X. Sridhar Mani, X. R. Matthew Redinbo, Gut microbial β-glucuronidases reactivate estrogens as components of the estrobolome that reactivate estrogens. *J. Biol. Chem.* **294**, 18586–18599 (2019).

92. P. Spanogiannopoulos, E. N. Bess, R. N. Carmody, P. J. Turnbaugh, The microbial pharmacists within us: A metagenomic view of xenobiotic metabolism. *Nat. Rev. Microbiol.* **14**, 273–287 (2016).

93. K. Oliphant, E. Allen-Vercoe, Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health. *Microbiome*. **7**, 1–15 (2019).

94. J. Y. L. Chiang, Bile acid metabolism and signaling. *Compr Physiol*. **3**, 1191–1212 (2013).

95. J. Walsh, L. Olavarria-Ramirez, G. Lach, M. Boehme, T. G. Dinan, J. F. Cryan, B. T. Griffin, N. P. Hyland, G. Clarke, Impact of host and environmental factors on β-glucuronidase enzymatic activity: Implications for gastrointestinal serotonin. *Am. J. Physiol. - Gastrointest. Liver Physiol.* **318**, G816–G826 (2020).

96. P. Awolade, N. Cele, N. Kerru, L. Gummidi, E. Oluwakemi, P. Singh, Therapeutic significance of β-glucuronidase activity and its inhibitors: a review. *Eur. J. Med. Chem.* **187**, 1–46 (2020).

97. D. Douguet, Data sets representative of the structures and experimental properties of FDA-approved drugs. *ACS Med. Chem. Lett.* **9**, 204–209 (2018).

98. L. Ejim, M. A. Farha, S. B. Falconer, J. Wildenhain, B. K. Coombes, M. Tyers, E. D. Brown, G. D. Wright, Combinations of antibiotics and nonantibiotic drugs enhance antimicrobial efficacy. *Nat. Chem. Biol.* **7**, 348–350 (2011).

99. S. J. Pellock, W. G. Walton, S. M. Ervin, D. Torres-Rivera, B. C. Creekmore, G. Bergan, Z. D. Dunn, B. Li, A. Tripathy, M. R. Redinbo, Discovery and characterization of FMN-binding β-glucuronidases in the human gut microbiome. *J. Mol. Biol.* **431**, 970–980 (2019).

100. R. K. Singh, H.-W. Chang, D. Yan, K. M. Lee, D. Ucmak, K. Wong, M. Abrouk, B. Farahnik, M. Nakamura, T. H. Zhu, T. Bhutani, W. Liao, Influence of diet on the gut

microbiome and implications for human health. *J. Transl. Med.* **15**, 1–17 (2017).

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

102. J. L. Sonnenberg, J. Xu, D. Leip, C.-H. Chen, B. P. Westover, J. Weatherford, J. D. Buhler, J. I. Gordon, Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science.* **307**, 1955–1959 (2005).

103. S. Nurk, D. Meleshko, A. Korobeynikov, P. A. Pevzner, metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).

104. J. Vollmers, S. Wiegand, A.-K. Kaster, Comparing and evaluating metagenome assembly tools from a microbiologist's perspective - not only size matters! *PLoS One*. **12**, 1–31 (2017).

105. D. Hyatt, G.-L. Chen, P. F. Locascio, M. L. Land, F. W. Larimer, L. J. Hauser, Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. **11**, 1–11 (2010).

106. S. J. Pellock, W. G. Walton, M. R. Redinbo, Selecting a single stereocenter: the molecular nuances that differentiate β-hexuronidases in the human gut microbiome. *Biochemistry*. **58**, 1311–1317 (2019).

107. J. Ni, T.-C. D. Shen, E. Z. Chen, K. Bittinger, A. Bailey, M. Roggiani, A. Sirota-Madi, E. S. Friedman, L. Chau, A. Lin, I. Nissim, J. Scott, A. Lauder, C. Hoffmann, G. Rivas, L. Albenberg, R. N. Baldassano, J. Braun, R. J. Xavier, C. B. Clish, M. Yudkoff, H. Li, M. Goulian, F. D. Bushman, J. D. Lewis, G. D. Wu, A role for bacterial urease in gut dysbiosis and crohn's disease. *Sci. Transl. Med.* **9**, 1–11 (2017).

108. R. H. Jacobson, X. J. Zhang, R. F. DuBose, B. M. Matthews, Three-dimensional structure of β-galactosidase from E. coli. *Nature*. **369**, 761–766 (1994).

109. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. **30**, 2114–2120 (2014).

110. H. Li, R. Durbin, Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. **25**, 1754–1760 (2009).

111. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, D. G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 1–6 (2011).

112. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with bowtie 2. *Nat. Methods*. **9**, 357–359 (2012).

113. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. **30**, 923–930 (2014).

114. J. M. Ridlon, S. Ikegawa, J. M. P. Alves, B. Zhou, A. Kobayashi, T. Iida, K. Mitamura, G. Tanabe, M. Serrano, A. De Guzman, P. Cooper, G. A. Buck, P. B. Hylemon, Clostridium scindens: a human gut microbe with a high potential to convert glucocorticoids into androgens. *J. Lipid Res.* **54**, 2437–2449 (2013).

115. A. Loguidice, B. D. Wallace, L. Bendel, M. R. Redinbo, U. A. Boelsterli, Pharmacologic targeting of bacterial β-glucuronidase alleviates nonsteroidal anti-inflammatory drug-induced enteropathy in mice. *J. Pharmacol. Exp. Ther.* **341**, 447–454 (2012).

116. A. P. Bhatt, S. J. Pellock, K. A. Biernat, W. G. Walton, B. D. Wallace, B. C. Creekmore, M. M. Letertre, J. R. Swann, I. D. Wilson, J. R. Roques, D. B. Darr, S. T. Bailey, S. A. Montgomery, J. M. Roach, M. A. Azcarate-Peril, R. B. Sartor, R. Z. Gharaibeh, S. J. Bultman, M. R. Redinbo, Targeted inhibition of gut bacterial β-glucuronidase activity enhances anticancer drug efficacy. *Proc. Natl. Acad. Sci.* **117**, 7374–7381 (2020).

117. K. L. Flannigan, M. R. Taylor, S. K. Pereira, J. Rodriguez-Arguello, A. W. Moffat, L. Alston, X. Wang, K. K. Poon, P. L. Beck, K. P. Rioux, M. Jonnalagadda, P. K. Chelikani, H. J. Galipeau, I. A. Lewis, M. L. Workentine, S. C. Greenway, S. A. Hirota, An intact microbiota is required for the gastrointestinal toxicity of the immunosuppressant mycophenolate mofetil. *J. Hear. Lung Transplant.* **37**, 1047–1059 (2018).

118. M. A. Peppercorn, P. Goldman, The role of intestinal bacteria in the metabolism of salicylazosulfapyridine. *J. Pharmacol. Exp. Ther.* **181**, 555–562 (1972).

119. C. W. Murray, D. C. Rees, The rise of fragment-based drug discovery. *Nat. Chem.* **1**, 187–192 (2009).

120. J. N. Spradlin, E. Zhang, D. K. Nomura, Reimagining druggability using chemoproteomic platforms. *Acc. Chem. Res*. **54**, 1801–1813 (2021).

121. W. Lu, M. Kostic, T. Zhang, J. Che, M. P. Patricelli, L. H. Jones, E. T. Chouchani Ae, N. S. Gray, Fragment-based covalent ligand discovery. *RSC Chem. Biol.* **2**, 354–367 (2021).