Supplemental Information

for

Discovering the Microbial Enzymes Driving Drug Toxicity with Activity-Based Protein Profiling

Parth B. Jariwala[1], Samuel J. Pellock[1], Dennis Goldfarb[6,7], Erica W. Cloer[2], Marta Artola[11], Joshua B. Simpson[1], Aadra P. Bhatt[5], William G. Walton[1], Lee R. Roberts[9], Michael B. Major[7,8], Gideon J. Davies[10], Herman S. Overkleeft[11], and Matthew R. Redinbo*[,1,3,4]

[1]Department of Chemistry, [2]Lineberger Comprehensive Cancer Center, [3]Integrated Program for Biological and Genome Sciences, and [4]Departments of Biochemistry and Microbiology, [5]Center for Gastrointestinal Biology and Disease, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States; [6]Institute for Informatics, [7]Department of Cell Biology and Physiology, [8]Department of Otolaryngology, Washington University, St. Louis, Missouri, United States; [9]Exploratory Science Center, Merck & Company Inc, Cambridge, Massachusetts, United States; [10]York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York, UK; [11]Department of Bioorganic Synthesis, Leiden Institute of Chemistry, Leiden University, Leiden, the Netherlands

METHODS

Protein expression, purification and site-directed mutagenesis. All GUS enzymes were expressed and purified as previously described.[1–4] *Bu*GUS-1^{NxK} and *Bu*GUS-2^{NxK} mutants were generated, expressed, and purified as previously described.[4] Type I and Type II β-glucosidases subcloned into pLIC-His vectors were purchased from BioBasic and expressed and purified as previously described.[4] Sequence information for purified β-glucosidases can be found in **Figure S8**. 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.[5] 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 (BHI) 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.[6] 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 (13,000x*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 (300x*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 again 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 (17,000x*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.[7]

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 S12**.

***In vitro* GUS activity and inhibition**. Apparent IC_{50} values for GUS inhibition by compounds **1–4** were calculated using an endpoint format.[8] 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.[8] 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.[7,8] 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.[8] 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 S9**).

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.[9] 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.[10]

Raw LC/MS data processing. Peptides and protein groups were identified by an iterative database strategy within MetaLab (version 1.1.1)[11], which used MaxQuant (version 1.6.2.3)[12]. The database search was performed using the integrated reference catalog of the human gut microbiome (IGC) database[13] combined with the UniProtKB/Swiss-Prot human sequence database (downloaded 1 Feb. 2017)[14] 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.[15] 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 S6**).

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.*[1] 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.[4]

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.[16] 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.

Safety Statement. No unexpected or unusually high safety hazards were encountered.

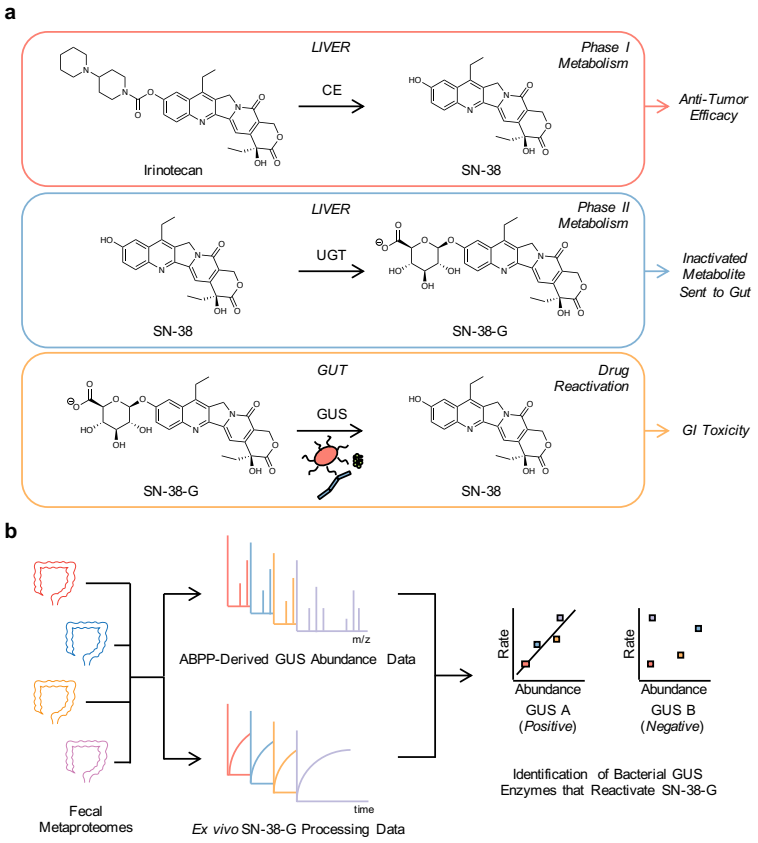


Figure S1. Approach to identify GUS enzymes responsible for drug glucuronide metabolism in the human gut microbiome. (**a**) The anti-cancer pro-drug irinotecan is processed by carboxylesterases (CE) during Phase I metabolism to form the active metabolite, SN-38, a topoisomerase I inhibitor. Uridine-diphosphate 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.

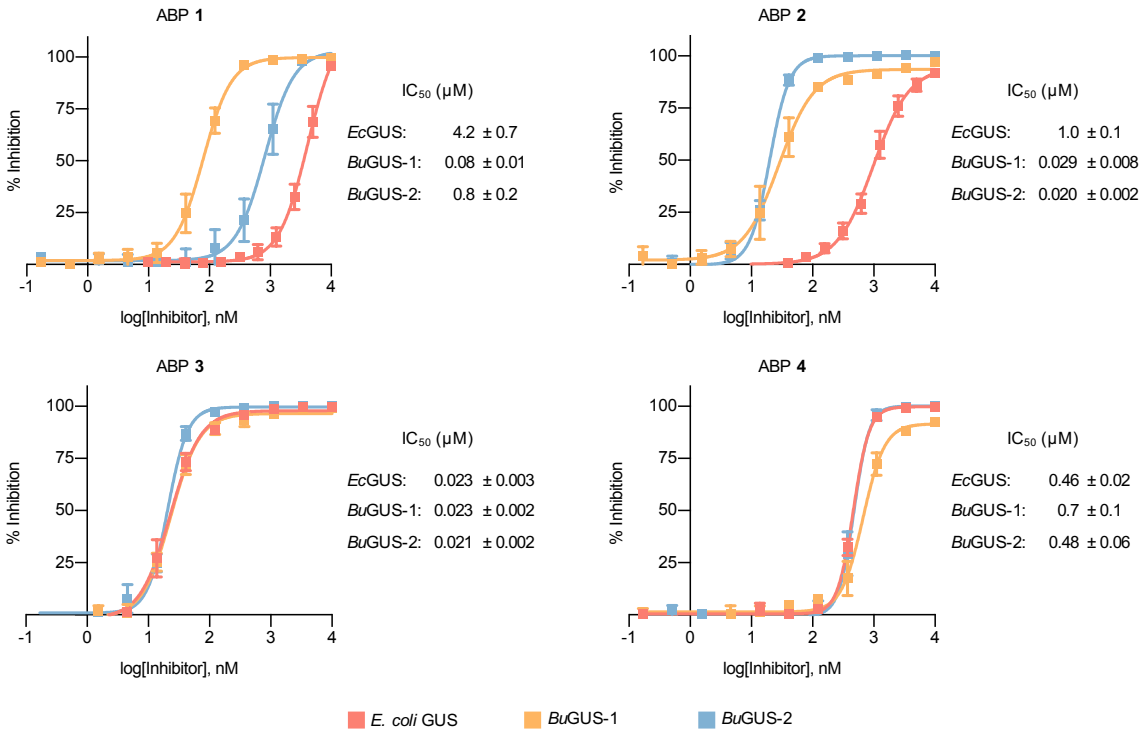


Figure S2. Dose-response plots for inhibition of select gut bacterial GUS enzymes by cyclophellitol-based inhibitors and ABPs. Apparent IC_{50} values shown are the mean ± standard deviation using N=3 biological replicates. Lower values indicate more potent inhibition.

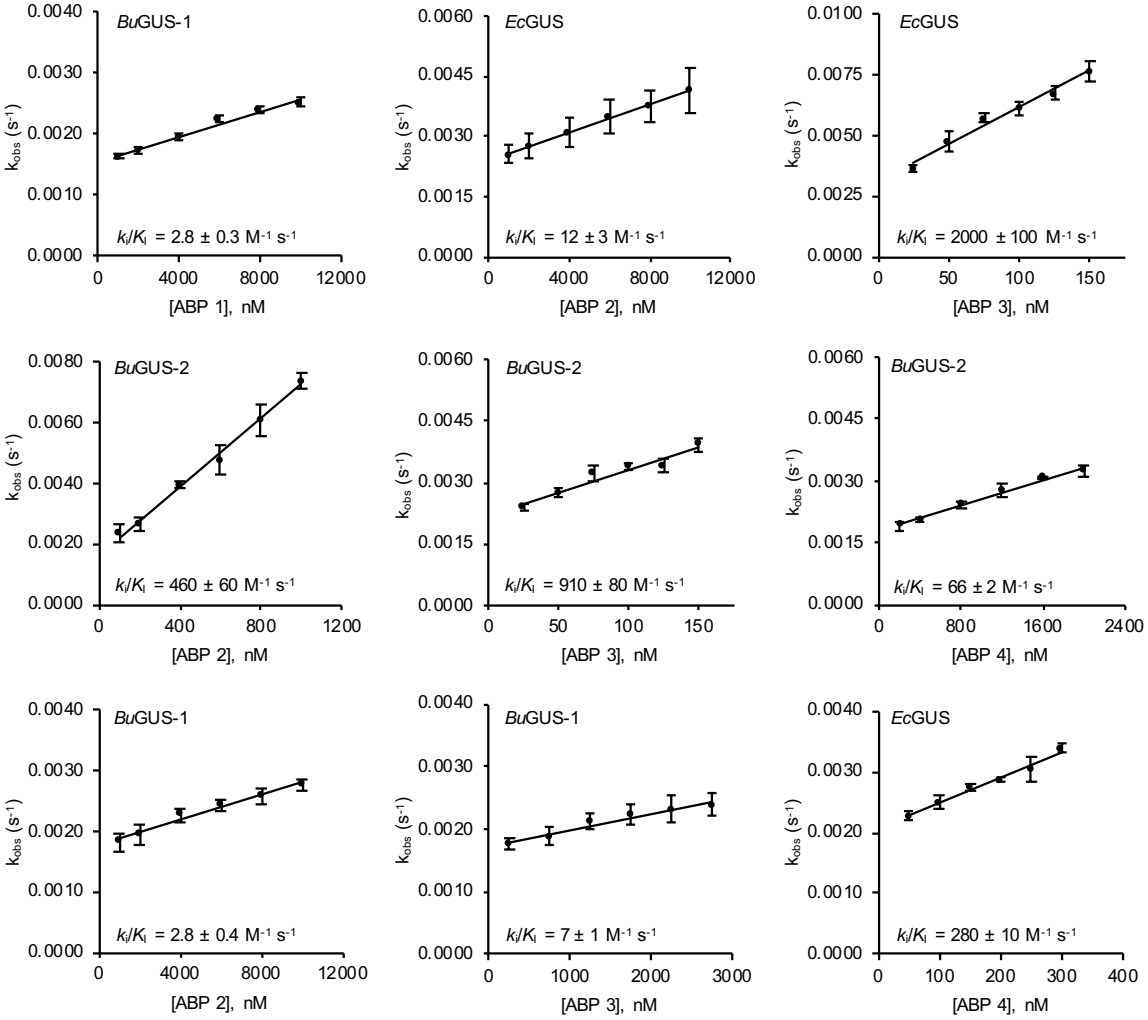


Figure S3. Derivation of kinetic parameters for GUS inhibition by inhibitors and ABPs 1-4. Plots shown are k_{obs} vs. concentration of inhibitors and ABPs **1–4**.

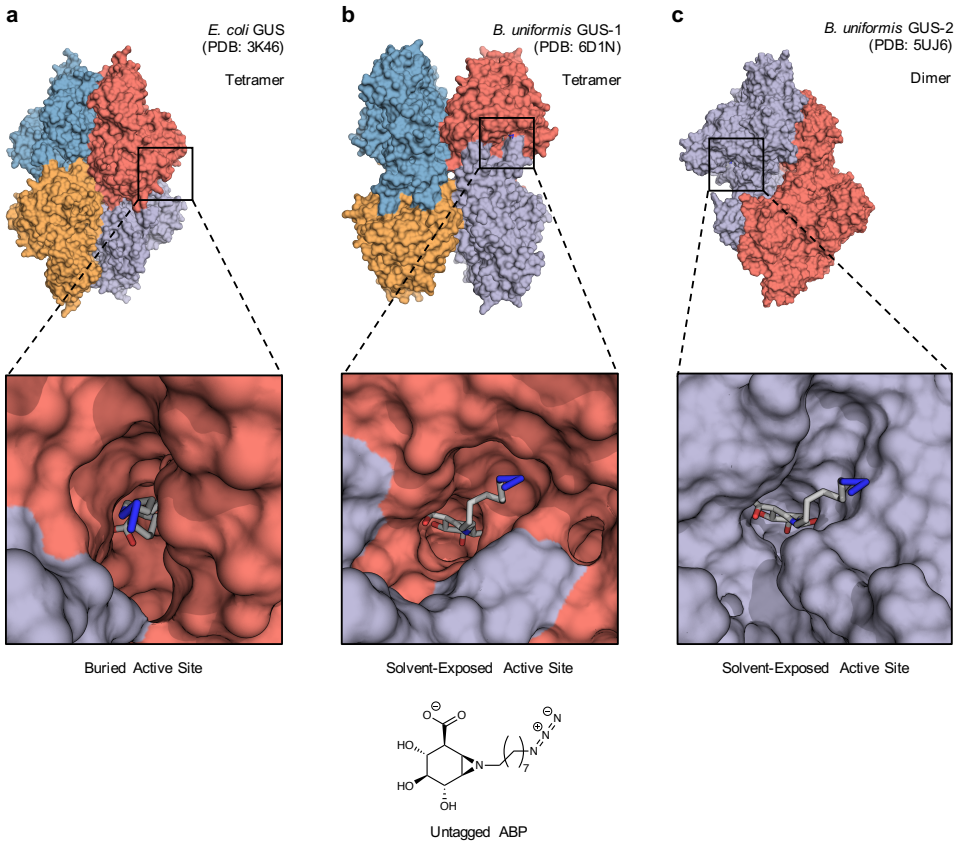


Figure S4. GUS ABPs target structurally diverse active sites. (**a**) *Ec*GUS tetramer, (**b**) *Bu*GUS-1 tetramer, and (**c**) *Bu*GUS-2 dimer with zoom-in of untagged ABP manually docked in active site in PyMol.

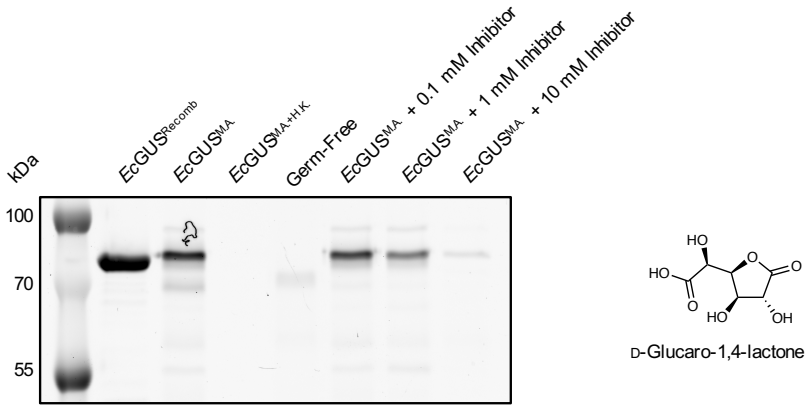


Figure S5. Cyclophellitol-based ABP labels GUS enzymes in mouse fecal lysate. Fecal extracts from germ-free and *E. coli* mono-associated mice were labelled with Cy5-ABP. Labelling was blocked in a dose-dependent manner using the pan-GUS inhibitor, D-glucaro-1,4-lactone. Recombinant *Ec*GUS sample (*Ec*GUS^{Recomb}), mono-associated *Ec*GUS fecal sample (*Ec*GUS$^{M.A.}$), heat-killed, mono-associated *Ec*GUS fecal sample (*Ec*GUS$^{M.A. + H.K.}$).

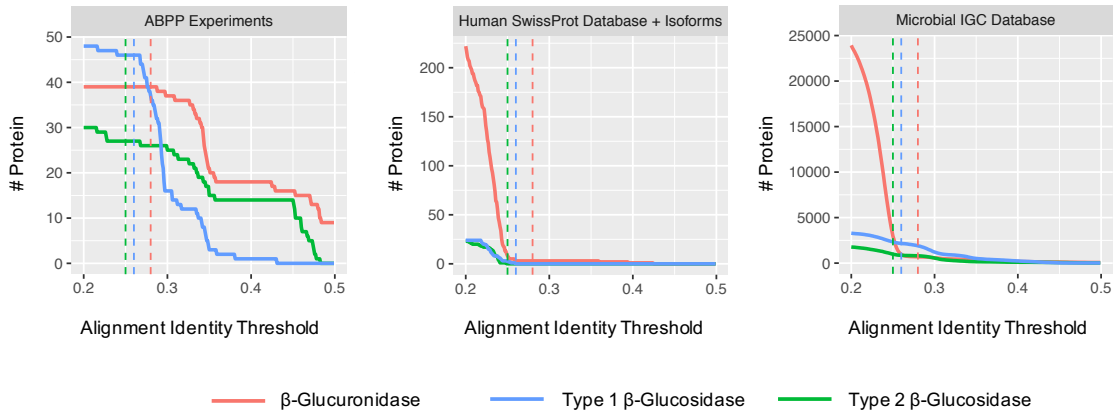


Figure S6. Determining sequence identity thresholds for GUS, Type I β-glucosidases, Type II β-glucosidases. Dashed vertical lines indicate chosen sequence identity thresholds: 0.28 (28%) for GUS, 0.25 (25%) for Type 1 β-glucosidases, and 0.26 (26%) for Type II β-glucosidases.

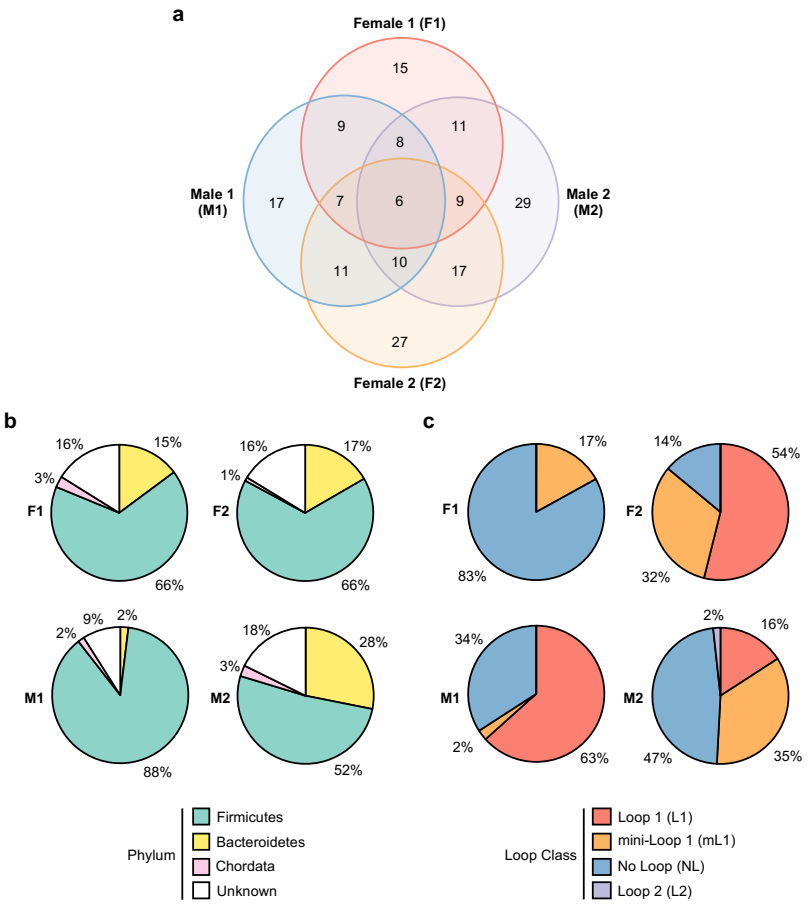


Figure S7. Interindividual variability in bacterial GUS composition. (**a**) Venn diagram of identified GUS protein groups from four individuals analyzed by this study. (**b**) Phylum-level and (**c**) structure-level analysis of GUS abundance from human fecal samples. Phyla-level composition information 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.

“Type I” Glucosidase

IGC sequence identified from fecal sample:

KIMPETKGILKMYICKRRIKIMAKWQRSFFQPVLPLGEDGRRVTGSKEHIALSRMAAGEGMVLLKNEKNTLPIRRGT
KVALFGKGTVDYVKGGGGSGDVTVEYIRNLYEGMKIKEDEGKVEVFDKLAKYYEKDIQKQYADGAVPGMTVEPELPD
ELLNEAREYTDTAVITICRFSGEGWDRKCEAAQDGYVLDGEEKRNSELSAKIFENGDFCLTNAENAMVEKVKKAFPH
VIVVMNVGGIVDTKWFRDCDEIQSVLMAWQGGMEGGLATADILCGDVNPSGKLSDTYAKDLEDYPSTANFHESAFYV
DYTEDIYVGYRYFETIPGAAERVNYPFGFGLSYTDFDWKMTGASEENGVITVLTEVTNTGKTAGKEVIQLYYGAPQG
KLGKPAKVLGAFKKTSILQPGERQILTLKIPVNQMASYDDLGKVCRSAYVLEAGEYAIYIGTNVRDAAKIDFTYVVK
EDTVTEQLSRKAAPYHLQKRMLADGSYEELPQREYVEEEGLPRQDKYAIGLPCPDTRGQKGIDFLDFLDSKGVRFSD
VADGKMTLDEFMDILTLDDCINLLGGQPNTGCANTFGMGNLPEYGVPNVMTADGPAGLRILPKCGVNTTAWPCATLL
ASTWDEELVEKVGKAGAEEVKENNISIWLTPACNIHRSPLCGRNFEYYSEDPYLAGKTGAAMVRGIQSQHIGASVKH
FAANNKETNRKDSDSRVSERALREIYLKQFEIIVKEAHPYTIMSSYNLINGIHASENKELLTGILRDEWGFDGLVTT
DWWTFGEHYRETKAGNDIKMAAGYPERIKEAYEKGFITEAEIRLSARRILNMILKID

Closest UniProt match: C6JEB3 (97.2%)

MAKWQRSFFQPVLPLGEDGRRVTGSKEHIALSRMAAGEGMVLLKNEKNTLPIRRGTKVAFFGKGTVDYVKGGGGSGD
VTVEYIRNLYEGMKIKEDEGKVEVFDKLAKYYEKDIQKQYADGAVPGMTVEPELPDELLNEAREYTDTAVITICRFS
GEGWDRKCEAAQDGYVLDGEEKRNSELSAKIFENGDFCLTNAENAMVEKVKKAFPHVIVVMNVGGIVDTKWFRDCDE
IQSVLMAWQGGMEGGLATADILCGDVNPSGKLSDTYAKDLEDYPSTANFHESAFYVDYTEDIYVGYRYFETIPGAAE
RVNYPFGFGLSYTDFDWKMTGASEENGVITVLTEVTNTGKTAGKEVIQLYYGAPQGKLGKPAKVLGAFKKTSLLQPG
ERQILTLKIPVDQMASYDDLGKVCRSAYVLEAGEYAIYIGTNVRDAAKIDFTYVVKEDTVTEQLSRKAAPYHLQKRM
LADGSYEELPQREYVEEEGLPRQDKYAIGLPCPDTRGQKGIDFLDFLDSKGVRFSDVADGKMTLDEFMDILTLDDCI
NLLGGQPNTGCANTFGMGNLPEYGVPNVMTADGPAGLRILPKCGVNTTAWPCATLLASTWDEELVEKVGKAGAEEVK
ENNISIWLTPACNIHRSPLCGRNFEYYSEDPYLAGKTGAAMVRGIQSQHIGASVKHFAANNKETNRKDSDSRVSERA
LREIYLKQFEIIVKEAHPYTIMSSYNLINGIHASENKELLTGILRDEWGFDGMVTTDWWTFGEHYRETKAGNDIKMA
AGYPERIKEAYEKGFITEGEIRLSARRILNMILKID

“Type I” conserved residues (sequence alignment to 5K6M):

```
5K6M            111 WSQPEMELTDEIVSNASAK--SDVAIVI-I-GR-TAGE--D-K-D-FS-D    149
C6JEB3          129 ---PE--LPDELL-N-EAREYTDTA-VITIC-RFS-GEGWDRKCEA-AQD    167

5K6M            512 AFG-G-VSEYLRKMDIP-AVC-CDDGPSGMR-LDS-GA-T-AFSMP-NGT    552
C6JEB3          553 TFGMGNLPEY----GVPN-VMTAD-GPAGLRILPKCGVNTTAW--PC-AT    593

5K6M            553 MLASTFNPD--VIERMYG-FT-S-LEMI--YNKVECL-L-GPGMNIHRNP    593
C6JEB3          594 LLASTW--DEELVEKV-GK-AGAE-E-VKE-NNIS-IWLT-PACNIHRSP    634

5K6M            594 L-NGRNFEYFSEDPYL-NG-T-IASAMLKGL---HKYG-SDGVAKHFCCN    635
C6JEB3          635 LC-GRNFEYYSEDPYLA-GKTG-A-AMVRGIQSQH-IGAS--V-KHFAAN    676

5K6M            679 Q-VNGM-WTAG-N-YDL-NT-RILRDEWGFKGIVMTDWWAQVNDRGGEPT    722
C6JEB3          719 NLINGIH--ASENK-ELL-TG-ILRDEWGFDGMVTTDWWT-F---G-E--    756
```

Expressed and purified His-tagged sequence:

MHHHHHHSSGVDLGTENLYFQSNASKEHIALSRMAAGEGMVLLKNEKNTLPIRRGTKVAFFGKGTVDYVKGGGGSGD
VTVEYIRNLYEGMKIKEDEGKVEVFDKLAKYYEKDIQKQYADGAVPGMTVEPELPDELLNEAREYTDTAVITICRFS
GEGWDRKCEAAQDGYVLDGEEKRNSELSAKIFENGDFCLTNAENAMVEKVKKAFPHVIVVMNVGGIVDTKWFRDCDE
IQSVLMAWQGGMEGGLATADILCGDVNPSGKLSDTYAKDLEDYPSTANFHESAFYVDYTEDIYVGYRYFETIPGAAE
RVNYPFGFGLSYTDFDWKMTGASEENGVITVLTEVTNTGKTAGKEVIQLYYGAPQGKLGKPAKVLGAFKKTSLLQPG
ERQILTLKIPVDQMASYDDLGKVCRSAYVLEAGEYAIYIGTNVRDAAKIDFTYVVKEDTVTEQLSRKAAPYHLQKRM
LADGSYEELPQREYVEEEGLPRQDKYAIGLPCPDTRGQKGIDFLDFLDSKGVRFSDVADGKMTLDEFMDILTLDDCI
NLLGGQPNTGCANTFGMGNLPEYGVPNVMTADGPAGLRILPKCGVNTTAWPCATLLASTWDEELVEKVGKAGAEEVK
ENNISIWLTPACNIHRSPLCGRNFEYYSEDPYLAGKTGAAMVRGIQSQHIGASVKHFAANNKETNRKDSDSRVSERA
LREIYLKQFEIIVKEAHPYTIMSSYNLINGIHASENKELLTGILRDEWGFDGMVTTDWWTFGEHYRETKAGNDIKMA
AGYPERIKEAYEKGFITEGEIRLSARRILNMILKID

“Type II” Glucosidase

IGC sequence identified from fecal sample:

MKKNIISMAAAMAVLSACGPGVPQLGKASLDEVIGAMTLEEKAHLVVGTGMAGFSGDSAVIGATKKLVPGAAGTTYP
IERLGIPAVVLADGPAGLRIDPTREGDSATYYCTHFPIGTLLASTWDQELVESVGQSIGNEVLEYGADVLLAPALNI
HRNPLCGRNFEYYSEDPVVSGKIAAAYVRGVQSNGVGTSIKHFAVNNQETNRMATDAHVSPRALREIYLKGFEIAVK
ESAPWTVMSSYNYLNGVYTSENKELQTTMLRDEWGFKGMVMTDWFGGKDAVAQMVAGNDMLQPGLPKQYEAIVKGVQ
DGALDEAILNQNVKRILEMILQTPHFKGYKYSNKPDLKAHAAVTRQSATEGMVLLKNDNASLPLAADVKNVALFGCT
SYDFIAGGTGSGNVNRAYTVSLLDGLKNAGYVVDEALKNSYEAYLKAEKGRLSKDKKEWFMPDERPAEMAVSAEVIR
EQAAKADVALVTLGRTSGEFLDRMVADFNLTKEELDMLKAVSDAFHAAGKRVVVVLNIGGVIETASWKSVPDAILCA
WQAGQEGGNSVADVLSGKASPSGKLTMTFPVKFEDAASSANFPIDMRVSIDLVNKGGKKNDVKNVDYTNYEEDIYVG
YRYFDTFGKQVSYPFGYGLSYTTFAYDKAAVKADNGVYTVSVEVKNTGKVAGKEVVQLYVSAPDAADANKPEKELKA
FAKTKELKPGEATVVTLKVSAADLASYDEAASAWVVAPGNYKFLVGASSRDIKATLEAEVAASSEKTNDILKLQEPI
SLLER

Closest UniProt match: A0A285SH81 (98.3%)

MKKNIISMAAAMAVLSACGPGVPQLGKSSLDEVIGAMTLEEKAHLVVGTGMAGFSGDSAVIGATRKLVPGAAGTTYP
IERLGIPAVVLADGPAGLRIDPKREGDSATYYCTHFPIGTLLASTWDQELVESVGQSIGNEVLEYGADVLLAPALNI
HRNPLCGRNFEYYSEDPLVSGKIAAAYVRGVQSNGVGTSIKHFAVNNQETNRMATDAHVSPRALREIYLKGFEIAVK
ESAPWTVMSSYNYLNGVYTSENKELQTTMLRDEWGFKGMVMTDWFGGKDAVAQMVAGNDMLQPGLPKQYEAIVKGVQ
DGALDEAILNQNVKRILEMILQTPHFKGYKYSNKPDLKAHAAVTRQSATEGMVLLKNDNGALPLAADVKNVALFGCT
SYDFIAGGTGSGNVNRAYTVSLLDGLKNAGYVVDEALKNSYEAYLKAEKERLSKDKKEWFMPDTRPAEMAVSAQVIR
EQAAKADVALVTLGRTSGEFLDRMVADFNLTKEEQDMLKAVSDAFHAAGKKVVVVLNIGGVIETASWKSVPDAILCA
WQAGQEGGNSVADVLSGKASPSGKLTMTFPVKFEDAASSDNFPIDMRVSTDLMNKGGKKNDVKNVDYTNYEEDIYVG
YRYFDTFGKQVSYPFGYGLSYTTFAYDKAAVKADNGVYTVSVEVKNTGKVAGKEVVQLYVSAPDAADANKPEKELKA
FAKTKELKPGETTVVTLKVNAADLASYDEAASAWVVTPGNYKFLVGASSRDIKATLEAEVAAATQKTNNILKLQEPM
SLLKR

“Type II” conserved residues (sequence alignment to 5WAB):

```
5WAB              48 GL**R**---KSNSATTGEVDLNNSVPAT--C--FPPAAG-L-SSSWNPELIHQ     88
A0A285SH8         94 GL**R**IDPK-RE---G--D---S--ATYYCTHFP-I-GTLLASTWDQELV-E    129

5WAB              89 -VGEAMAEECI-QEKV-A-VILGPGVNIKRNPL-GG**R**-CFEYWSEDPYL-    131
A0A285SH8        130 SVGQSIGNE-VL-E-YGADVLLAPALNIHRNPLCG-**R**N-FEYYSEDP-LV    173

5WAB             132 AGHE-AVGIVAGVQSKGVGTSL**KH**FAANNQETDRLRVS-ANISQRALREI    179
A0A285SH8        174 SG-KIAAAYVRGVQSNGVGTSI**KH**FAVNNQETNRM-ATDAHVSPRALREI    221

5WAB             224 GYEGIVMS**D**W-GADHDRVASL-NAGLNLEM--P--PS-YTDDQIVYAARD    266
A0A285SH8        266 GFKGMVMT**D**WFGGK-DAVAQMV-AG-N-DMLQPGLPKQY-E-AIV---K-    305

5WAB             410 LPEAAES**E**G-F-DRETL--D--IPAK--QVELLKAV--A--AENKNIVVV    447
A0A285SH8        477 --RT--S-G**E**FLDR--MVADFNL-TKEEQ-DMLKAVSDAFHAAGKKVVVV    517
```

Expressed and purified His-tagged sequence:

MHHHHHHSSGVDLGTENLYFQSNAGVPQLGKSSLDEVIGAMTLEEKAHLVVGTGMAGFSGDSAVIGATRKLVPGAAG
TTYPIERLGIPAVVLADGPAGLRIDPKREGDSATYYCTHFPIGTLLASTWDQELVESVGQSIGNEVLEYGADVLLAP
ALNIHRNPLCGRNFEYYSEDPLVSGKIAAAYVRGVQSNGVGTSIKHFAVNNQETNRMATDAHVSPRALREIYLKGFE
IAVKESAPWTVMSSYNYLNGVYTSENKELQTTMLRDEWGFKGMVMTDWFGGKDAVAQMVAGNDMLQPGLPKQYEAIV
KGVQDGALDEAILNQNVKRILEMILQTPHFKGYKYSNKPDLKAHAAVTRQSATEGMVLLKNDNGALPLAADVKNVAL
FGCTSYDFIAGGTGSGNVNRAYTVSLLDGLKNAGYVVDEALKNSYEAYLKAEKERLSKDKKEWFMPDTRPAEMAVSA
QVIREQAAKADVALVTLGRTSGEFLDRMVADFNLTKEEQDMLKAVSDAFHAAGKKVVVVLNIGGVIETASWKSVPDA
ILCAWQAGQEGGNSVADVLSGKASPSGKLTMTFPVKFEDAASSDNFPIDMRVSTDLMNKGGKKNDVKNVDYTNYEED
IYVGYRYFDTFGKQVSYPFGYGLSYTTFAYDKAAVKADNGVYTVSVEVKNTGKVAGKEVVQLYVSAPDAADANKPEK
ELKAFAKTKELKPGETTVVTLKVNAADLASYDEAASAWVVTPGNYKFLVGASSRDIKATLEAEVAAATQKTNNILKL
QEPMSLLKR

Figure S8. Sequences of Type I and Type II β-glucosidases expressed and purified for *in vitro* analysis. For each type of β-glucosidase, the sequence identified from proteomics data (IGC sequence), the closest UniProt match, and sequence alignment with a published structure are listed. Key conserved residues are bolded in the sequence alignment.

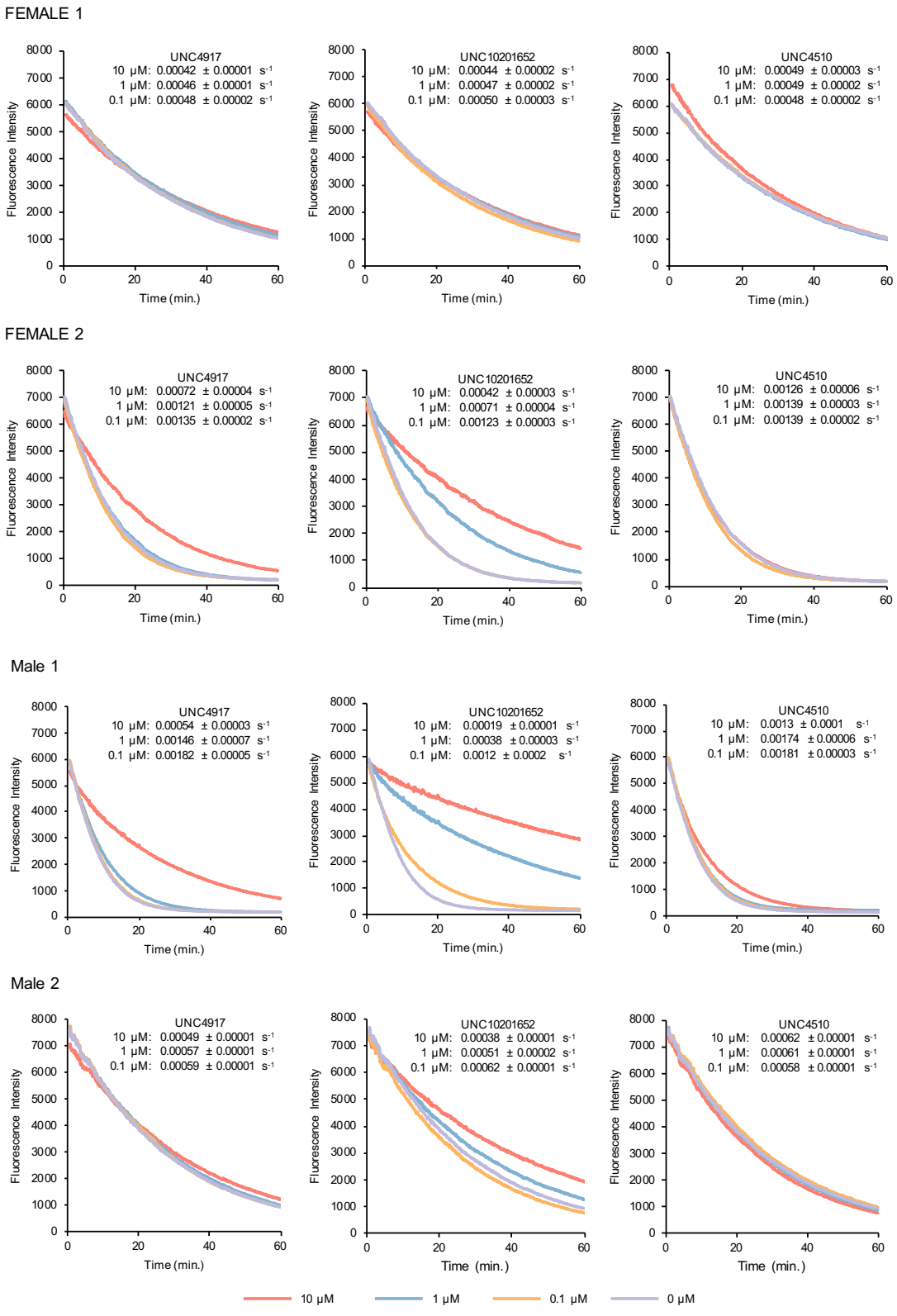


Figure S9. Progress curves showing SN-38-G processing and inhibition in fecal samples.

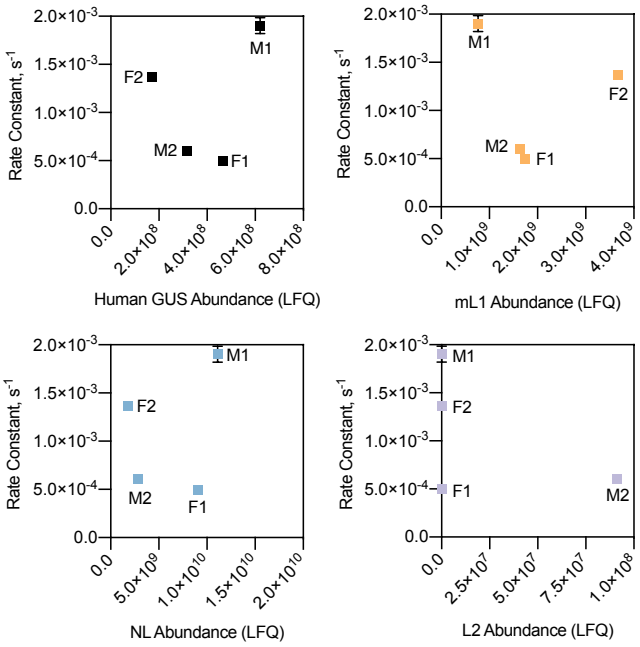


Figure S10. Correlation analysis of abundance data from various GUS structural classes against SN-38-G processing data. Correlation analysis between mini-Loop 1 GUS, No Loop GUS, Loop 2 GUS abundance, human GUS abundance and SN-38-G processing.

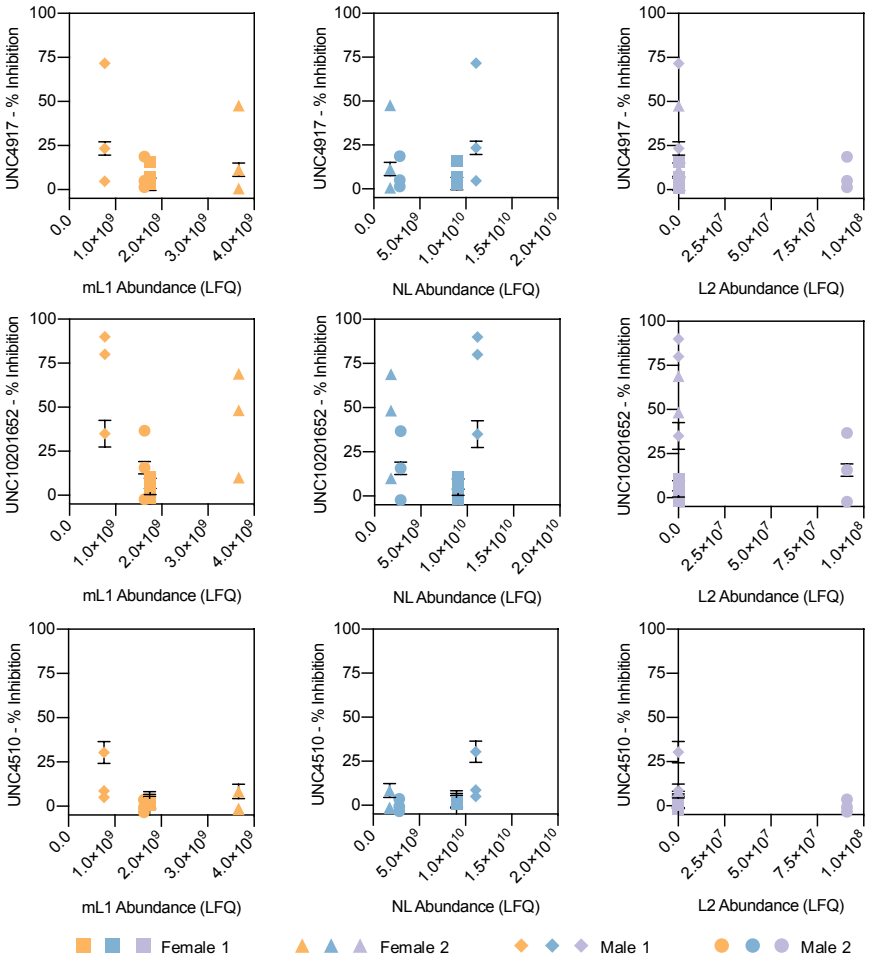


Figure S11. Correlation analysis of abundance data from various GUS structural classes against SN-38-G inhibition data. Correlation analysis between mini-Loop 1 GUS, No Loop GUS, Loop 2 GUS abundance, and inhibition data for Loop 1-specific GUS inhibitors, UNC4917, UNC10201652, and UNC4510.

Figure 1c – Fluorescence Scan

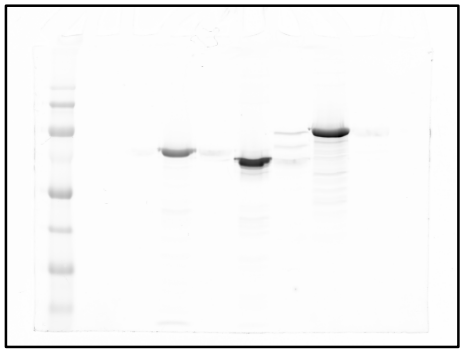


Figure 1d – Fluorescence Scan

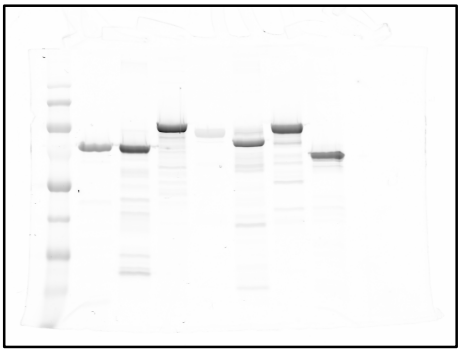


Figure 1c – Coommassie Stain

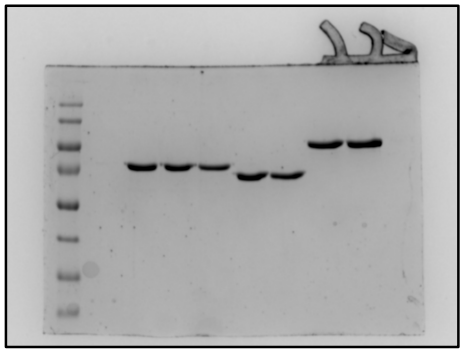
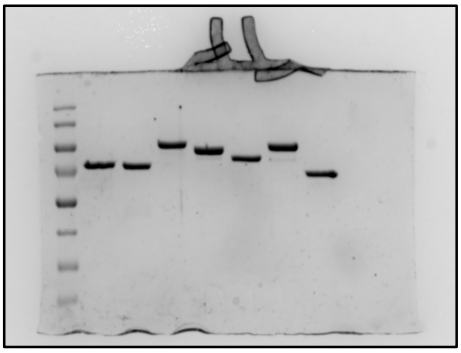


Figure 1d – Coommassie Stain



Suppl. Figure 5 – Fluorescence Scan

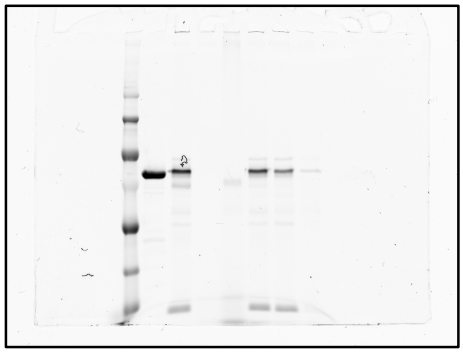


Figure S12. Full gel images.

REFERENCES

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

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

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

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

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

(6) Biernat, K. A., Pellock, S. J., Bhatt, A. P., Bivins, M. M., Walton, W. G., Tran, B. N. T., Wei, L., Snider, M. C., Cesmat, A. P., Tripathy, A., Erie, D. A., and Redinbo, M. R. (2019) Structure, function, and inhibition of drug reactivating human gut microbial β-glucuronidases. *Sci. Rep. 9*, 1–15.

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

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

(9) Davis, S., Charles, P. D., He, L., Mowlds, P., Kessler, B. M., and Fischer, R. (2017) Expanding Proteome Coverage with CHarge Ordered Parallel Ion aNalysis (CHOPIN) Combined with Broad Specificity Proteolysis. *J. Proteome Res. 16*, 1288–1299.

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

(11) Cheng, K., Ning, Z., Zhang, X., Li, L., Liao, B., Mayne, J., Stintzi, A., and Figeys, D. (2017) MetaLab: an automated pipeline for metaproteomic data analysis. *Microbiome 5*, 1–10.
(12) Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol. 26*, 1367–72.
(13) Li, J., Wang, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., Arumugam, M., Kultima, J. R., Prifti, E., Nielsen, T., Juncker, A. S., Manichanh, C., Chen, B., Zhang, W., Levenez, F., Wang, J., Xu, X., Xiao, L., Liang, S., Zhang, D., Zhang, Z., Chen, W., Zhao, H., Al-Aama, J. Y., Edris, S., Yang, H., Wang, J., Hansen, T., Nielsen, H. B., Brunak, S., Kristiansen, K., Guarner, F., Pedersen, O., Doré, J., Ehrlich, S. D., and Bork, P. (2014) An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol. 32*, 834–841.
(14) Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-AJee, H., Cowley, A., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., Castro, L. G., Figueira, L., Garmiri, P., Georghiou, G., Gonzalez, D., Hatton-Ellis, E., Li, W., Liu, W., Lopez, R., Luo, J., Lussi, Y., MacDougall, A., Nightingale, A., Palka, B., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G., Rosanoff, S., Saidi, R., Sawford, T., Shypitsyna, A., Speretta, E., Turner, E., Tyagi, N., Volynkin, V., Wardell, T., Warner, K., Watkins, X., Zaru, R., Zellner, H., Xenarios, I., Bougueleret, L., Bridge, A., Poux, S., Redaschi, N., Aimo, L., ArgoudPuy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M. C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., Casal-Casas, C., De Castro, E., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Jungo, F., Keller, G., Lara, V., Lemercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Nouspikel, N., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A. L., Wu, C. H., Arighi, C. N., Arminski, L., Chen, C., Chen, Y., Garavelli, J. S., Huang, H., Laiho, K., McGarvey, P., Natale, D. A., Ross, K., Vinayaka, C. R., Wang, Q., Wang, Y., Yeh, L. S., and Zhang, J. (2017) UniProt: The universal protein knowledgebase. *Nucleic Acids Res. 45*, D158–D169.
(15) Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet. 16*, 276–277.
(16) Mesuere, B., Willems, T., Van Der Jeugt, F., Devreese, B., Vandamme, P., and Dawyndt, P. (2016) Unipept web services for metaproteomics analysis. *Bioinformatics 32*, 1746–1748.