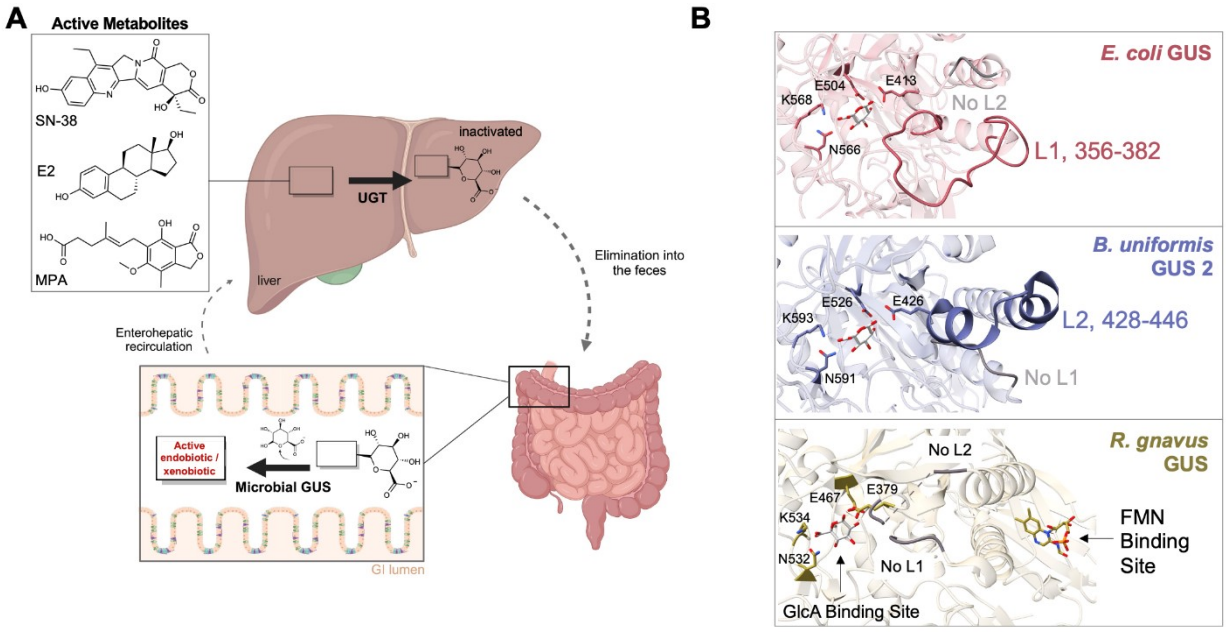
Supplemental Figures

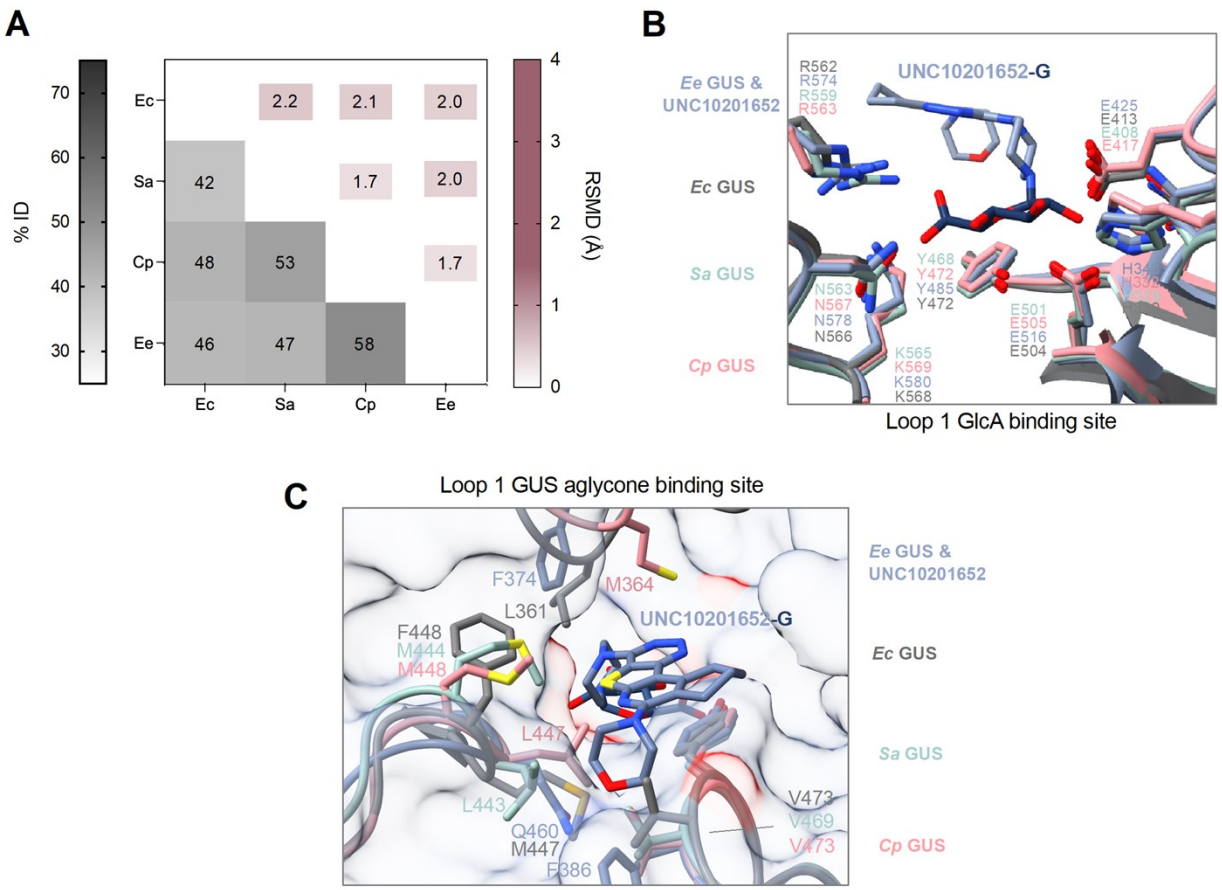


Supplemental Figure 1. Lifecycle of glucuronidated compounds and their reactivation by gut microbial GUS.

(a) Endobiotics and xenobiotics can be glucuronidated in the liver by host UGTs to improve compound solubility and to promote subsequent detoxification and excretion. Glucuronide-conjugates are sent to the GI tract for elimination where they encounter gut microbial GUS that can cleave the glucuronic acid and reactivate the compounds, producing damaging effects.
(b) Active site gating loops that differentiate microbial GUS enzymes and are used to classify sequences into structural classes. (*E. coli* GUS, PDB 3K46; *B. uniformis* GUS, PDB 5UJ6; *R. gnavus* GUS, PDB 6MVG)

SN-38, active form of irinotecan; E2, estradiol; FMN, flavin mononucleotide; MPA, active form of mycophenolate mofetil; L1, Loop 1 GUS; L2, Loop 2 GUS; UGT, human uridine diphosphate glucuronosyl transferases.

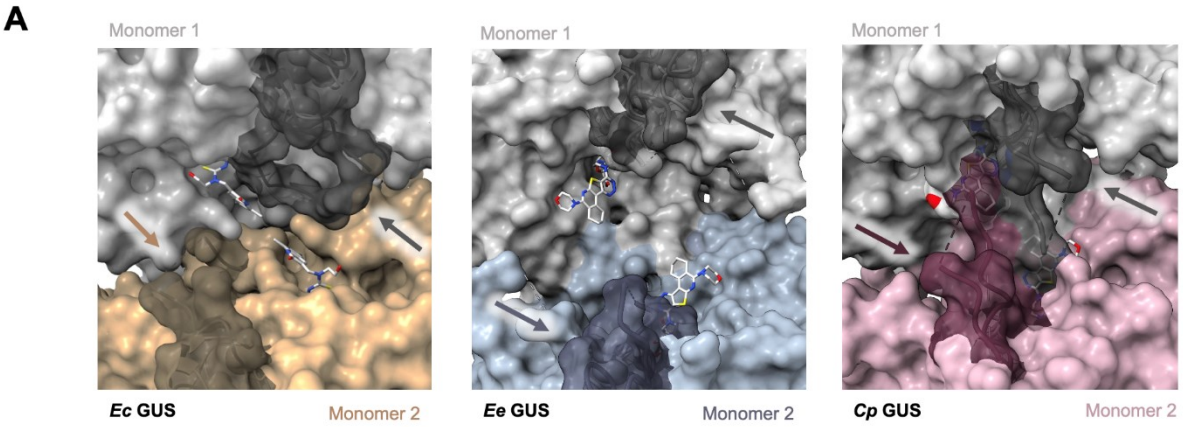
Supplemental Figure 1a was generated using BioRender.



Supplemental Figure 2. L1 GUS structural and c-alpha alignments reveal conserved carbon backbone and GlcA binding site residues.

(a) Carbon backbone alignment and percent sequence identity plot of L1 GUS show conserved 3-dimensional structure while having a relatively low sequence identity within the same loop class.
(b) Structural alignment of L1 GUS GlcA binding site (*Ee* in blue, PDB 8GEN; *Ec* in grey, PDB 3K46; *Sa* in green, PDB 4JKK; *Cp* in pink, PDB 6CXS).
(c) Structural alignment of L1 GUS highlighting the aglycone binding pocket (*Ee* in blue, PDB 8GEN; *Ec* in grey, PDB 3K46; *Sa* in green, PDB 4JKK; *Cp* in pink, PDB 6CXS).

Cp, *Clostridium perfringens; Ec, Escherichia coli; Ee, Eubacterium eligens;* GlcA, glucuronic acid; ID, identity; RMSD, root mean square deviation; *Sa*, *Streptococcus agalactiae*.

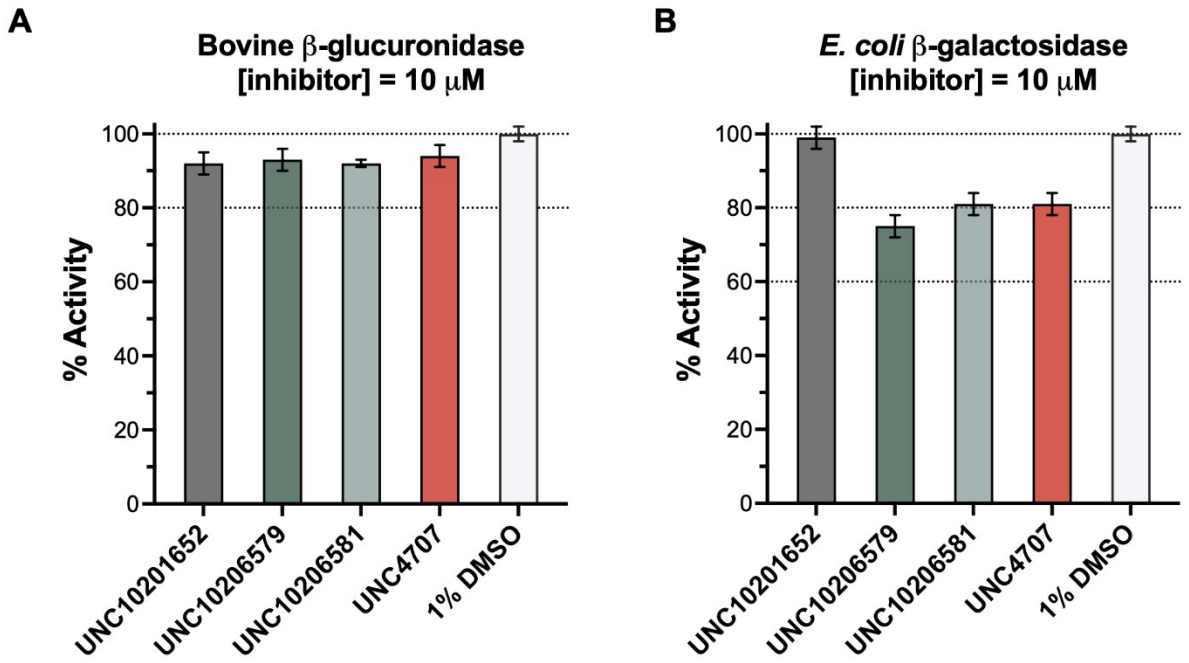


Supplemental Figure 3. L1 region of GUS enzymes reach into the neighboring dimer and make stabilizing contacts with inhibitors.

(a) Dimeric interface of L1 GUSs bound to GUS inhibitors (Ec GUS & Inhibitor 2, PDB 3LPF; Ee GUS & UNC10201652-G, PDB 8GEN; and Cp GUS & UNC10201652, PDB 6CXS), the L1 region is large enough to reach into the aglycone binding pocket and make contacts with inhibitors. Loop regions are rendered as darker, transparent, and have an arrow pointing to them. Inhibitors are rendered in white.

(b) Multiple sequence alignment of L1 GUSs tested in our SAR campaign, comparing the sequences and length of the L1 region.

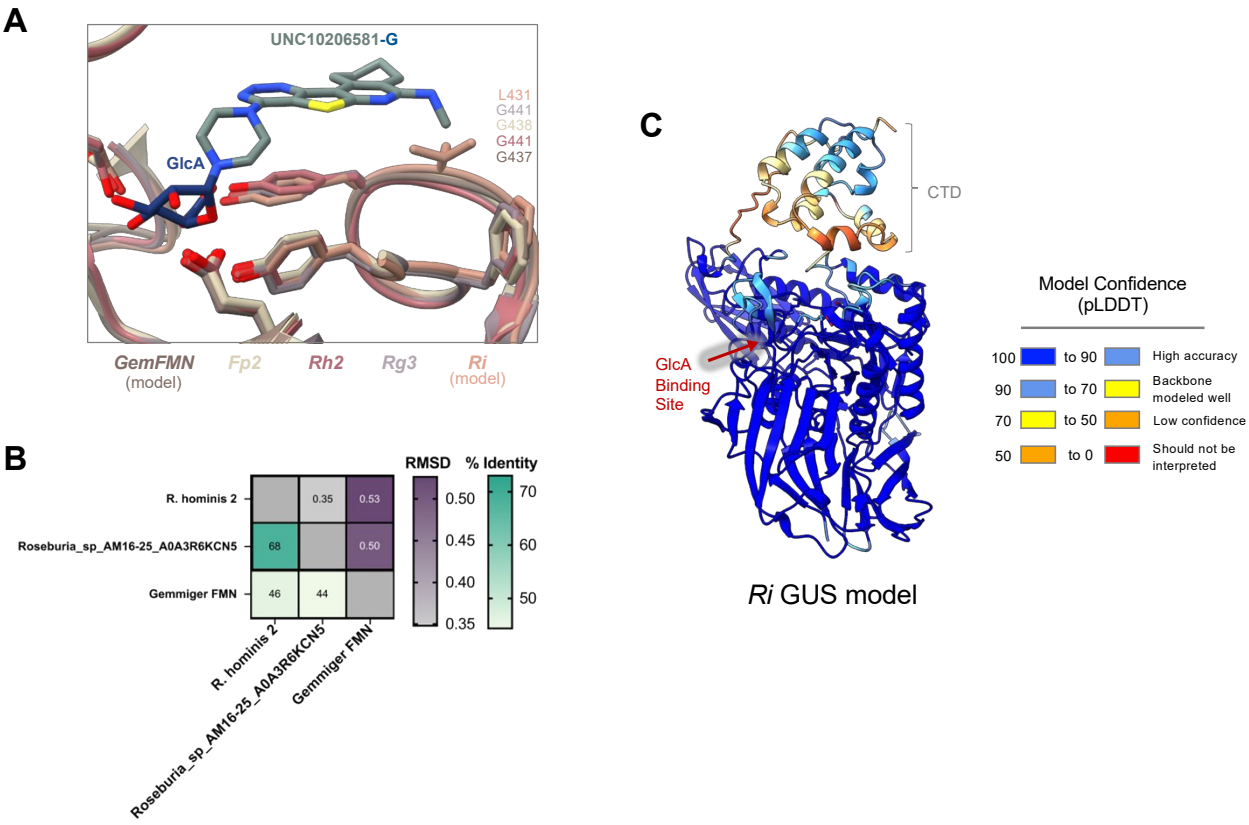
Cp, *Clostridium perfringens; Ec, Escherichia coli; Ee, Eubacterium eligens; Sa*, *Streptococcus agalactiae.*



Supplemental Figure 4. Selectivity assays reveal GUS inhibitors retain strong selectivity for GUS over other closely related orthologs and glycoside hydrolase proteins.

(a) Percent activity assays of lead analogs against the bovine GUS ortholog reveal strong selectivity for microbial GUS. n=3 replicates, (mean ± SEM). *p*-nitrophenyl-β-D-glucuronide was used as a substrate.

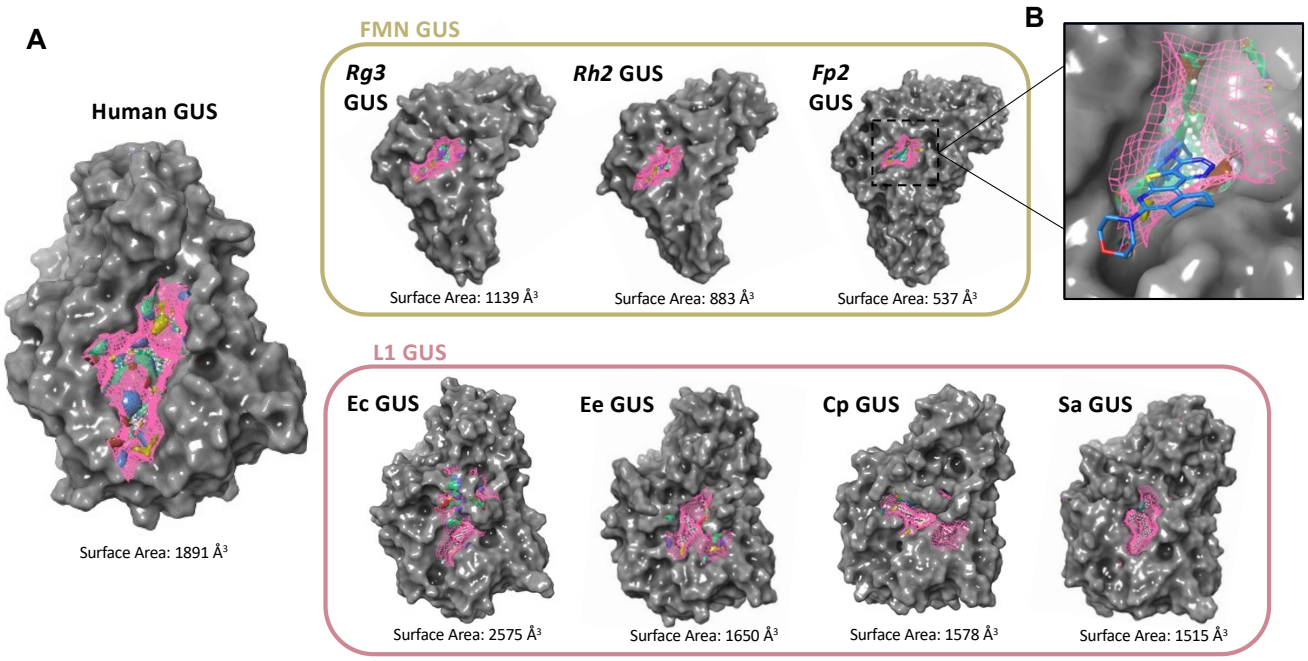
(b) Percent activity assays of lead analogs against closely related glycoside hydrolase family, beta-galactosidase, reveal selectivity for Ec GUS over Ec beta-galactosidase. n=3 replicates, (mean ± SEM). *p*-nitrophenyl-β-D-glucuronide was used as a substrate.



Supplemental Figure 5. Comparison of FMN GUS aglycone binding site and an AlphaFOLD model of Ri GUS.

(a) Alignment of UNC10206581-G in extant FMN GUS structures (*Fp2*, 6MVF; *Rg3*, 6MVG; *Rh2*, 6MVH) and AlphaFOLD models (*Ri* and *GemFMN*) reveal conserved contacts observed in L1 GUSs.
(b) Carbon backbone alignment and percent sequence identity plot of FMN GUSs found in human fecal donors 5, 6, and 8 from Figure 4d. AlphaFold models were used for the carbon backbone alignment of *Gem*FMN and AM16-25.
(c) AlphaFOLD model of *Ri* GUS (FMN) shows moderate to low confidence structure of the C-terminal domain (CTD) region. Uniprot ID: A0A173RPS8

FMN, flavin mononucleotide; GlcA, glucuronic acid; NTL, N-terminal loop; *Fp2, Faecalibacterium prausnitzii L2-6; Rg3, Ruminococcus gnavus; Rh2, Ruminococcus hominis; Ri, Ruminococcus inulinivorans; GemFMN, Gemmiger sp.*.



Supplemental Figure 6. Sitemap analysis of GlcA binding sites show variability in pocket size, depth, and physiochemical properties between microbial and human GUS isoforms.

(a) SiteMap analysis of GlcA binding pockets for human, L1, and FMN GUS. Binding pocket surface is rendered as pink mesh, hydrophobic area is rendered as yellow surface, hydrophilic area is rendered as green surface, hydrogen bond donors are rendered as purple surface, hydrogen bond acceptors are rendered as red surface, and cavities are space filled with white spheres.
(b) Structural overlay of Fp2 GUS sitemap analysis with UNC10291652-G. Binding pocket surface is rendered as pink mesh, hydrophobic area is rendered as yellow surface, hydrophilic area is rendered as green surface, hydrogen bond donors are rendered as purple surface, hydrogen bond acceptors are rendered as red surface, cavities are space filled with white spheres, and the ligand is rendered as blue sticks.

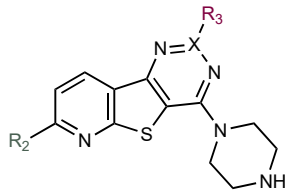
The following PDB entries were used for sitemap analysis: Ec GUS, PDB 3LPF; Ee GUS, PDB 8GEN; Cp GUS, PDB 6CXS; Sa GUS, 4JKK; *Fp2* GUS, 6MVF; *Rg3* GUS, 6MVG; *Rh2* GUS, 6MVH; human GUS, 3nh3.

Cp, *Clostridium perfringens; Ec, Escherichia coli; Ee, Eubacterium eligens;* FMN, flavin mononucleotide; GlcA, glucuronic acid; NTL, N-terminal loop; *Fp2, Faecalibacterium prausnitzii L2-6; Rg3, Ruminococcus gnavus; Rh3, Ruminococcus hominis; Sa*, *Streptococcus agalactiae.*

Supplemental Tables

Supplemental Table 1: Changes to R2, R3, and X of the UNC4917 scaffold are well-tolerated.

Parent scaffold data is highlighted in bold. *p*-nitrophenyl-β-D-glucuronide was used as a substrate. *Cp, Clostridium perfringens*, Ec, *Escherichia coli; Eubacterium eligens* GUS, Sa, *Streptococcus agalactiae.*



					IC_{50} (nM)			
#	Name	R_2	X	R_3	*Ec* GUS	*Ee* GUS	*Cp* GUS	*Sa* GUS
13	**UNC4917**		**C**		**20 ± 5**	**1,100 ± 200**	**50.5 ± 0.4**	**2,400 ± 500**
14	UNC4764		N	H	23 ± 5	800 ± 200	14 ± 3	2,900 ± 600
15	UNC4830		N	H	30 ± 20	800 ± 100	20 ± 10	1,040 ± 80
16	UNC4746		C	H	20 ± 10	600 ± 100	14 ± 8	730 ± 70
17	UNC4785		C		22 ± 7	400 ± 100	30 ± 8	400 ± 70
18	UNC4600		C	H	16 ± 9	270 ± 50	15 ± 8	520 ± 90
19	UNC4708		C		50 ± 20	600 ± 100	90 ± 40	720 ± 80
20	UNC4707		C		50 ± 10	400 ± 80	71 ± 2	260 ± 30
21	UNC4847		C		41 ± 9	500 ± 100	51 ± 1	1,300 ± 200
22	UNC4666	H	C	H	19 ± 3	2,700 ± 600	44 ± 3	13,000 ± 3,000
23	UNC4910	H	C		9 ± 2	500 ± 100	60 ± 1	2,800 ± 600

Supplemental Table 2. Crystallography data collection and refinement statistics.

	***E. eligens* GUS : UNC10206581-G**
PDB Code	8UGT
Wavelength	1.00
Resolution range (Å)	44.26 - 2.65 (2.75 - 2.65)
Space group	P 64 2 2
Unit cell (a, b, c, a, b, g; Å, deg.)	179.9, 179.9, 132.8, 90, 90, 120
Total reflections	74,512 (7,246)
Unique reflections	37,256 (3,623)
Multiplicity	2.0 (2.0)
Completeness (%)	99.95 (99.94)
Mean I/sigma (I)	13.74 (1.36)
Wilson B-factor (Å^2)	64.66
R-merge	0.03524 (0.5127)
R-meas	0.04984 (0.7251)
R-pim	0.03524 (0.5127)
CC1/2	0.999 (0.57)
CC*	1 (0.852)
Reflections used in refinement	37,254 (3,623)
Reflections used for R-free	2,000 (195)
R-work	0.2332 (0.3429)
R-free	0.2630 (0.3817)
CCwork	0.937 (0.650)
CCfree	0.924 (0.525)
Number of non-hydrogen atoms	4,662
macromolecules	4,565
ligands	74
solvent	23
Protein residues	582
RMSbonds (Å)	0.009
RMSangles (deg.)	1.16
Ramachandran favored (%)	91.4
Ramachandran allowed (%)	8.22
Ramachandran outliers (%)	0.35
Rotamer outliers (%)	2.28
Clashscore	18.26
Average B-factor, all atoms (Å^2)	67.5
Average B-factor, macromolecules (Å^2)	67.81
Average B-factor, ligands (Å^2)	50.99
Average B-factor, solvent (Å^2)	67.54

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

Supplemental Table 3. Reference ID and strain origin for GUSs examined.

GUS Abbreviation	Strain of Origin	NCBI Reference ID or Uniprot ID
Cp	*Clostridium perfringens* (strain 13 / Type A)	Q8XP19
Gemmiger FMN	Subdoligranulum sp. APC924/74	WP_114000118.1
Ec	*Escherichia coli K-12*	P05804
Ee	*Lachnospira eligens (strain ATCC 27750 / DSM 3376 / VPI C15-48 / C15-B4) (Eubacterium eligens)*	C4Z6Z2
Fp L2-6	*Faecalibacterium prausnitzii* L2-6	WP_015563823.1
Rh2	*Roseburia hominis* (strain DSM 16839 / JCM 17582 / NCIMB 14029 / A2-183)	WP_014080400.1
Rg3	*Ruminococcus gnavus*	WP_118581144.1
Sa	*Streptococcus agalactiae serotype V (strain ATCC BAA-611 / 2603 V/R)*	Q8E0N2

NCBI reference IDs begin with “WP_”