Structural basis for the regulation of β-glucuronidase expression by human gut Enterobacteriaceae

Michael S. Little[a], Samuel J. Pellock[a], William G. Walton[a], Ashutosh Tripathy[b], and Matthew R. Redinbo[a,b,c,d,1]

[a]Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3290; [b]Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3290; [c]Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3290; and [d]Integrative Program for Biological and Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3290

Edited by Michael A. Fischbach, Stanford University, Stanford, CA, and accepted by Editorial Board Member Brenda A. Schulman November 15, 2017 (received for review September 14, 2017)

The gut microbiota harbor diverse β-glucuronidase (GUS) enzymes that liberate glucuronic acid (GlcA) sugars from small-molecule conjugates and complex carbohydrates. However, only the Enterobacteriaceae family of human gut-associated Proteobacteria maintain a GUS operon under the transcriptional control of a glucuronide repressor, GusR. Despite its potential importance in *Escherichia*, *Salmonella*, *Klebsiella*, *Shigella*, and *Yersinia* opportunistic pathogens, the structure of GusR has not been examined. Here, we explore the molecular basis for GusR-mediated regulation of GUS expression in response to small-molecule glucuronides. Presented are 2.1-Å-resolution crystal structures of GusRs from *Escherichia coli* and *Salmonella enterica* in complexes with a glucuronide ligand. The GusR-specific DNA operator site in the regulatory region of the *E. coli* GUS operon is identified, and structure-guided GusR mutants pinpoint the residues essential for DNA binding and glucuronide recognition. Interestingly, the endobiotic estradiol-17-glucuronide and the xenobiotic indomethacin-acyl-glucuronide are found to exhibit markedly differential binding to these GusR orthologs. Using structure-guided mutations, we are able to transfer *E. coli* GusR's preferential DNA and glucuronide binding affinity to *S. enterica* GusR. Structures of putative GusR orthologs from GUS-encoding Firmicutes species also reveal functionally unique features of the Enterobacteriaceae GusRs. Finally, dominant-negative GusR variants are validated in cell-based studies. These data provide a molecular framework toward understanding the control of glucuronide utilization by opportunistic pathogens in the human gut.

biochemistry | molecular biology | structural biology | gut microbiota | transcriptional regulation

Microorganisms compete in the gastrointestinal (GI) tract for sources of carbon in the form of simple and complex carbohydrates and have been shown to work in synergy to process dietary fiber that cannot be degraded by the host (1, 2). Microbial carbohydrate utilization plays a critical role in the diversity, abundance, and metabolic activity of both commensal and potentially pathogenic bacteria in the mammalian intestine (3, 4). As researchers have shown, the ability to access the energy present in oligosaccharides provides a competitive advantage to the members of the Bacteroidetes that harbor polysaccharide-utilization loci (PULs), which encode enzymatic and membrane-spanning machinery that catabolizes a range of complex carbohydrates (5–10). This leads to the question of how other members of the microbiota that lack PULs are able to compete for energy within the GI tract.

The GUS operon, which was first described more than 30 y ago in *Escherichia coli*, provides a potential answer to this question (11–15). This operon encodes proteins involved in processing glucuronidated ligands, including β-glucuronidase (GUS). GUS enzymes are glycosyl hydrolases that remove glucuronic acid (GlcA) sugars linked to endobiotic and xenobiotic compounds by phase II drug-metabolizing UDP-glucuronosyltransferase (UGT) enzymes in protective host tissues (e.g., liver and intestines) (16–21). A wide range of chemicals are conjugated to GlcA, including hormones, neurotransmitters, environmental pollutants, and drugs like cancer chemotherapeutics, immunosuppressants, and nonsteroidal antiinflammatory drugs (NSAIDs) (22–25). The microbial GUS-mediated reactivation of these compounds in the gut may play a role in their serum exposure via enterohepatic recirculation (26–29). This pathway also causes the intestinal damage and dose-limiting toxicities of the anticancer drug irinotecan and several NSAIDs (30–32). Microbe-selective GUS inhibitors have been shown to alleviate these toxicities in mice, providing an early demonstration of nonlethal drugs specific to the microbiome (30–34).

GUS is the product of the *gusA* gene, which in the GUS operon is followed by the inner-membrane GlcA-specific transporter *gusB* and nonspecific outer-membrane channel *gusC* genes (Fig. 1) (12, 15). As we show below, similar operons are found only in other Enterobacteriaceae, including *Salmonella*, *Klebsiella*, *Yersinia*, and *Shigella* taxa, all of which are potential intestinal and systemic pathogens. Importantly, the Enterobacteriaceae lack PULs, suggesting that they might rely on systems like the GUS operon to harness available forms of carbon (35, 36). The GUS operons in the Enterobacteriaceae are under the control of the transcriptional repressor GusR, which is expected to respond to the presence of glucuronidated ligands by dissociating from the regulatory region of the operon and thus allowing operon transcription. Similar to the *lac* and other *E. coli* operons, the GUS operon is also sensitive to catabolite repression by glucose, the global metabolic regulator (13, 37, 38). In *E. coli*, GusR and a related repressor, UxuR, were previously found to bind to two operator elements, termed sites 1 and 2, in the regulatory region of the GUS operon (14–16, 39, 40). To date, however, the structural and biochemical

Significance

Commensal microbiota establish nutrient-utilization niches in the gastrointestinal tract. While the large intestine is dominated by the Bacteroidetes that degrade complex carbohydrates, the small intestine contains Proteobacteria and Firmicutes that compete with host tissues for small-molecule sources of carbon. Here, we show that the Enterobacteriaceae family of Proteobacteria, including *Escherichia*, *Salmonella*, *Klebsiella*, *Shigella*, and *Yersinia* pathobionts, maintains DNA operator- and glucuronidated ligand-specific glucuronide repressor (GusR) transcription factors that uniquely respond to glucuronidated ligands.

Author contributions: M.S.L., S.J.P., and M.R.R. designed research; M.S.L., S.J.P., and W.G.W. performed research; A.T. contributed new reagents/analytic tools; M.S.L. analyzed data; and M.S.L. and M.R.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. M.A.F. is a guest editor invited by the Editorial Board.

Data deposition: The atomic coordinates and structure factors reported in this paper have been deposited in the Protein Data Bank, www.wwpdb.org (PDB ID codes 6AYI, 6AYH, 6AZ6, and 6AZH).

[1]To whom correspondence should be addressed. Email: redinbo@unc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1716241115/-/DCSupplemental.

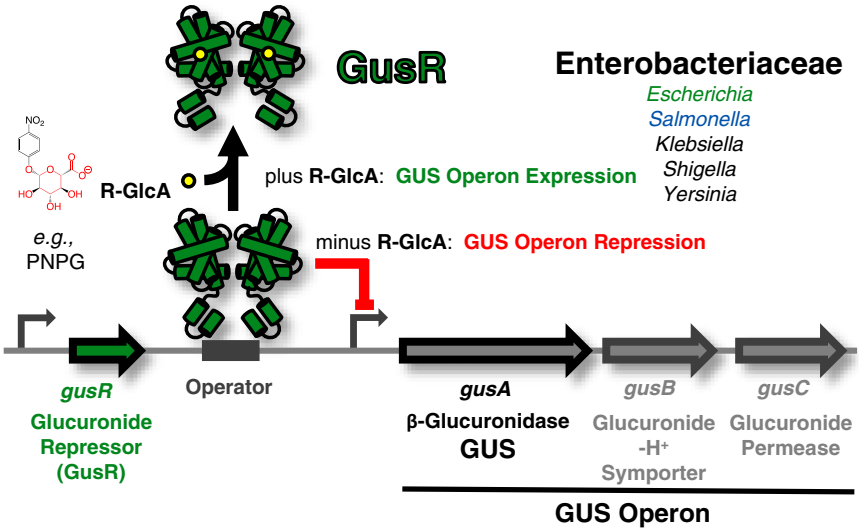


Fig. 1. Schematic of the GUS operon in Enterobacteriaceae. In the absence of a glucuronide ligand (R-GlcA; yellow), GusR (green) is expected to repress (red) the downstream transcription of the GUS operon proteins GusA, GusB, and GusC (gray) by binding to a specific DNA operator site. In the presence of a glucuronide ligand (e.g., *p*-nitrophenyl glucuronide), GusR disassociates from the operator to allow GUS operon expression. As shown here, only the Enterobacteriaceae, including several opportunistic bacterial pathogens, contain a GUS operon and GusR.

bases of the interaction between GusR and the operator DNA, as well as the binding of glucuronides to GusR, have remained undefined.

Here, we describe the crystal structures of two Enterobacteriaceae GusR proteins, from *E. coli* and from *S. enterica*, and outline the DNA binding characteristics of *E. coli* GusR within the regulatory region of its GUS operon. We further pinpoint the molecular determinants of GusR recognition of glucuronidated ligands both in vitro and in cell-based studies. Together, these data advance our understanding of how GI microbiota that lack the capacity to process energy-dense oligosaccharides up-regulate a system to scavenge GlcA from available glucuronides.

Results

GusR Crystal Structures. We determined the 2.1-Å-resolution crystal structures of the GusR proteins from the gut microbial Enterobacteriaceae species *E. coli* (EcGusR) and *S. enterica* (SeGusR) in complexes with *p*-nitrophenyl-β-D-glucuronide (PNPG) (Table S1). PNPG is a standard in vitro GUS assay substrate, that provided cocomplex crystals with the two GusRs examined. Both EcGusR and SeGusR are α-helical homodimers, with each monomer composed of DNA- (α1 to α3) and effector-binding domains (α4 to α10) (Fig. 2). While the DNA-binding domain (DBD) of one monomer is disordered in the EcGusR structure (Fig. 2*A*), both DBDs are fully ordered and visualized in the SeGusR homodimer (Fig. 2*B*). EcGusR and SeGusR share 59% sequence identity and 0.85-Å root-mean-square deviation (rmsd) across 176 equivalent Cα positions. EcGusR and SeGusR also share 3.4- and 3.5-Å rmsd and 17 and 16% sequence identity, respectively, with the *E. coli* TetR protein that defines this family of ligand-regulated transcriptional repressors. Thus, GusR appears to use a TetR-like fold to create separate DBDs and effector-binding domains (EBDs) to recognize its DNA binding site in a manner controlled by glucuronide effector ligands.

DNA-Binding Domain. The GusR DBDs exhibit a helix-turn-helix (HTH) DNA-binding motif composed of α2 and α3 (Fig. 3*A*). The HTH is the most common DNA binding fold in bacterial transcriptional factors, and is highly conserved within the TetR family of ligand-controlled regulators (41). Previous crystal structures of *E. coli* TetR [Protein Data Bank (PDB) ID code 1QPI] and *Staphylococcus aureus* QacR (SaQacR; PDB ID code 1JT0), which both share a 3.4-Å rmsd and 17% sequence identity with EcGusR, reveal that these repressor proteins utilize their HTH motifs to bind as homodimers to their palindromic operator sites (42, 43). In 1987, Blanco reported that the GUS operon in *E. coli* was under the control of GusR and UxuR, a second TetR-like repressor that shares 14% sequence identity with EcGusR, and that both repressors appeared to bind to two operator elements, sites 1 and 2 (44). Site 1 is 30 base pairs (bp) in length and is located 200 bp upstream from the GUS operon's ribosome binding site (rbs), while site 2 (40 bp) is only 50 bp from the same rbs (Fig. S1). The 1987 report employed cell-based *lac* gene fusion experiments to examine operator site interactions (44). Here, we studied the specificity of *E. coli* GusR and UxuR interactions with predicted operator sites in vitro using isothermal titration calorimetry (ITC).

We cloned the gene for *E. coli* UxuR (EcUxuR), recombinantly overexpressed the protein in *E. coli*, and purified it to homogeneity. We then compared the abilities of EcGusR and EcUxuR to bind to operator sites 1 and 2 using ITC. We found that EcGusR bound the 30 bp distal site 1 with 0.2 μM affinity but failed to bind to the 40 bp site 2 located in closer proximity to the start of the GUS operon (Fig. 3 *B* and *C* and Fig. S1). Furthermore, we found that EcUxuR failed to bind to either site in the conditions tested (Fig. 3*B*). Thus, for *E. coli*, while EcUxuR showed no affinity for these DNA duplexes, site 1 was validated as an element capable of binding EcGusR in vitro.

Next, we sought to understand the effect that a specific duplex DNA element might have on EcGusR's ability to bind a glucuronide ligand. We incubated EcGusR with excess operator site 1 or site 2 and then utilized increasing concentrations of the glucuronide ligand PNPG as the titrant for ITC. Without DNA, EcGusR bound to PNPG with a K_d of 0.2 μM (Fig. 4*B* and Fig. S2). However, in the presence of 18-fold molar excess site 1, EcGusR's affinity for PNPG is reduced more than 20-fold, to 4.1 μM. By contrast, 18-fold molar excess site 2 did not change EcGusR's affinity for PNPG, which remained at 0.2 μM. Therefore, *E. coli* GusR's affinity for the effector ligand decreases when its cognate DNA element, site 1, is present. The inverse experiment was also performed, in which saturating levels of PNPG were added to EcGusR, and operator site 1 or site 2 DNA duplexes were then titrated in ITC studies. These

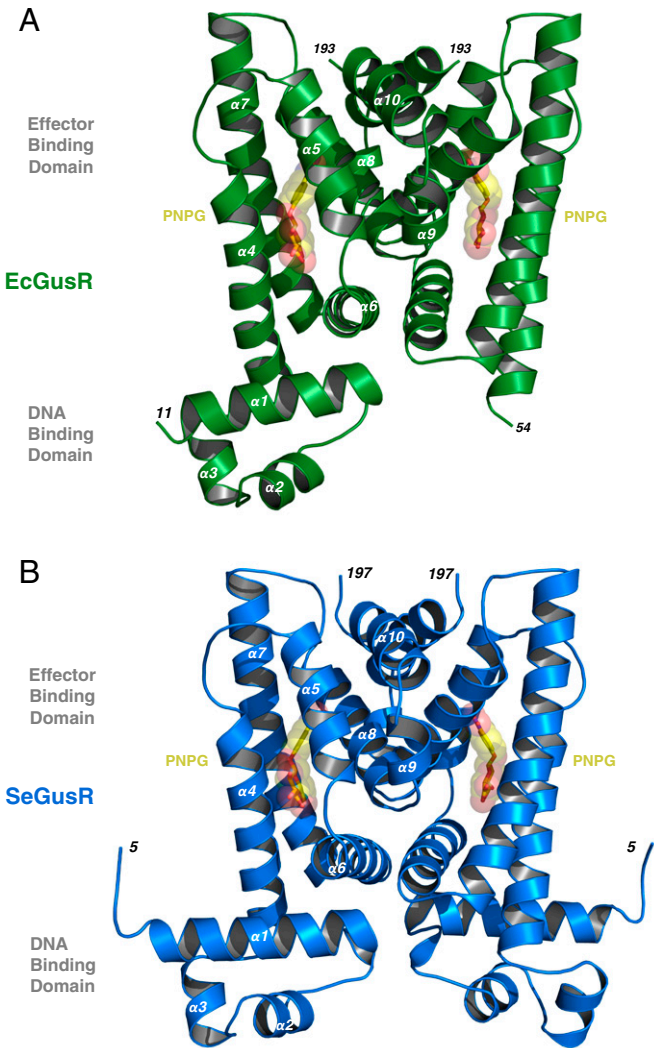


Fig. 2. *E. coli* and *S. enterica* GusR crystal structures. (*A*) Ribbon representation of the 2.1-Å crystal structure of the EcGusR (green; PDB ID code 6AYI) homodimer bound to PNPG (yellow and red), revealing residues 11 to 193 in one monomer and residues 54 to 193 in the other monomer. EcGusR's secondary structure is composed of 10 α-helices. The N-terminal DNA-binding domain is formed by α1 to α3, while the effector-binding domain is framed by α4 to α10. One *E. coli* GusR DBD is not visualized, likely due to motion within the crystal. (*B*) Ribbon representation of the 2.1-Å crystal structure of the SeGusR (blue; PDB ID code 6AYH) homodimer bound to PNPG (yellow and red) composed of residues 5 to 195 and intact DBDs and EBDs in both monomers.

experiments revealed no binding by EcGusR to either operator site when glucuronide ligand is present (Fig. 3*D*). These observations support the conclusion that GusR acts as a DNA-bound transcriptional repressor in the absence of ligand and releases from DNA once an appropriate glucuronide ligand is present.

A multiple sequence alignment of 2,353 TetR family members identified Y40 in SaQacR as highly conserved, and this amino acid side chain was found to form both base-specific and phosphate backbone contacts in the SaQacR–DNA complex structure (41). The equivalent residue in EcGusR is Y49 (Fig. 3*A*). To determine whether this residue is important for GusR's association with its regulatory element, we replaced Y49 with alanine in EcGusR. The EcGusR Y49A mutant failed to bind either DNA operator site 1 or 2, validating the important role this residue plays in GusR DBD function (Fig. 3*B*). The GusR binding sites in the regulatory region of the *S. enterica* GUS operon are not known, and we found that SeGusR does not bind to operator sites 1 or 2 from the *E. coli* operon's regulatory region (Fig. 3*A*). However, we noted that the DBDs from EcGusR and SeGusR are highly similar in sequence and deviate within the HTH region by only four amino acid positions (SCAI in EcGusR; ASDM in SeGusR). Replacement of "ASDM" in SeGusR with "SCAI" creates a variant SeGusR protein that is able to bind to *E. coli's* operator site 1 with 0.3 μM affinity (Fig. 3*B*), highlighting the importance of these residues for site-specific DNA interactions. Together, these data establish that EcGusR binds with high affinity to DNA operator site 1 in the regulatory region of the GUS operon in a fashion that is dependent on a conserved TetR family residue (Y49) and specific sequence of amino acids (SCAI) in the HTH motif of the DBD. Furthermore, EcGusR exhibits reduced affinity for its DNA binding site when the effector ligand is in excess.

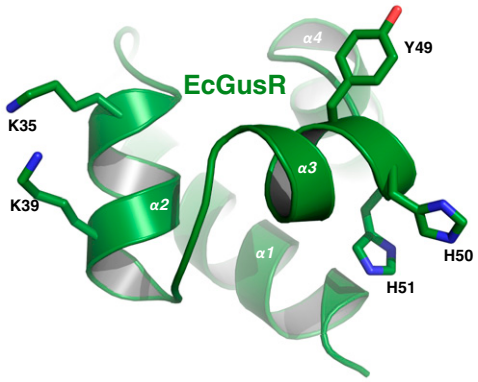
Effector-Binding Domain. The GusR effector-binding domain in the structures of both Enterobacteriaceae proteins is framed by α-helices 4 to 10 and creates a cavity suited for glucuronide recognition (Fig. 4*A*). The carboxylate unique to GlcA, relative to the isostructural glucopyranoside, is located within 2.6 and 2.8 Å from the lysine and tyrosine residues, respectively, that are conserved in sequences of GusR proteins (Fig. 4*A* and Fig. S4). Interestingly, the GUS enzyme also uses a lysine and tyrosine to recognize the same GlcA carboxylate within the enzyme's active site (33). In the GusR effector-binding pocket, each hydroxyl of the glucuronide sugar forms a hydrogen bond with three polar GusR residues that are highly conserved (Fig. 4*A*). Thus, the sugar moiety of the bound glucuronide makes six contacts with a total of five GusR residues. The *p*-nitrophenol group, by contrast, interacts with only three protein side chains, two of which utilize relatively less specific van der Waals contacts. Because this portion of the ligand varies depending on the specific glucuronide bound, it is perhaps not surprising that fewer side chains and less specific contacts are formed with this group (Fig. 4*A*). Taken together, these structural data reveal that GusR uses intimate polar contacts in recognizing the GlcA moiety of a bound effector ligand (Fig. 4*A*).

Next, we compared the ability of EcGusR, SeGusR, and various mutant proteins to bind to glucuronide effector ligand (PNPG) in vitro by ITC. We found that EcGusR and SeGusR bound PNPG with a K_d of 0.2 ± 0.07 μM and 2.7 ± 0.4 μM, respectively (Fig. 4*B* and Fig. S2). The ligand-binding pockets of the two receptors are identical except at three positions—R73/H72, M87/L86, and H126/Y125 for EcGusR/SeGusR, respectively (Fig. 4*A*). An M87L mutation in EcGusR reduced PNPG binding by 8-fold (Table S2), and an EcGusR M87A mutation led to a greater than 100-fold reduction in PNPG affinity compared with wild type (Fig. 4*B*).

We find that mutations of the carboxylate-contacting residues in EcGusR or SeGusR (K125/124; Y164/163) eliminate or significantly reduce PNPG binding (Fig. 4*B*). Replacement of lysine with alanine in both EcGusR and SeGusR produces GusR variants with no binding to PNPG, while mutation of the tyrosine to alanine decreases ligand binding by 70- or 200-fold for SeGusR and EcGusR, respectively (Fig. 4*B*). Eliminating only the hydroxyl group of the tyrosine side chain via phenylalanine mutations still reduces binding by 6- to 50-fold for the Se and Ec receptors, respectively, highlighting the importance of this polar contact with the glucuronide carboxylate (Fig. 4*B*). Reductions in binding affinities of similar magnitudes are also observed when glutamate, arginine, or histidine side chains that form hydrogen bonds with the glucuronide's hydroxyl groups are replaced with alanine (Fig. 4*B*). Thus, GusR not only employs key electrostatic contacts but also relies on a series of hydrogen bonds to precisely recognize the sugar moiety of glucuronides. Indeed, we find that neither *p*-nitrophenyl-glucopyranoside (PNP-Gluco) nor free GlcA binds to either receptor (Fig. S2).

Finally, we find that the endobiotic conjugate estradiol-17-β-D-glucuronide (E_{17}-glucuronide) binds only to EcGusR (K_d 11 μM),

A



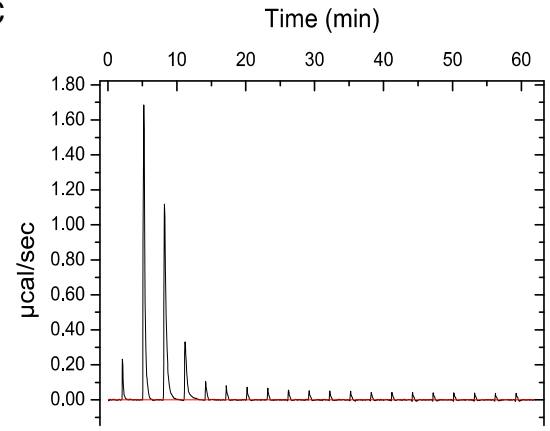
EcGusR	ASM**K**AIC**K**SCAISPGTL**YHH**FI
SeGusR	TSM**K**TIC**K**ASDMSPGTL**YHH**FP
SaQacR	TTTGEIV**K**LSESSKGNL**YYH**FK
EcUxuR	LIMLEIKGLVEVRRGAGIYVLD

B

DNA Binding Affinity by ITC (μM)

	EcGusR WT	EcGusR Y49A	EcUxuR WT	SeGusR WT	SeGusR 'SCAI'
Operator Site 1	0.24 *(0.04)*	*NB*	*NB*	*NB*	0.33 *(0.2)*
Operator Site 2	*NB*	*NB*	*NB*	*NB*	*NB*

C



D

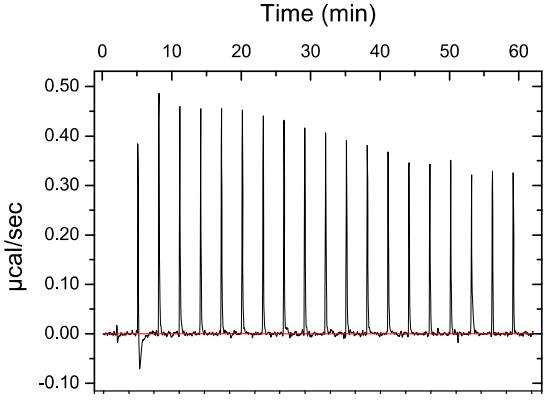


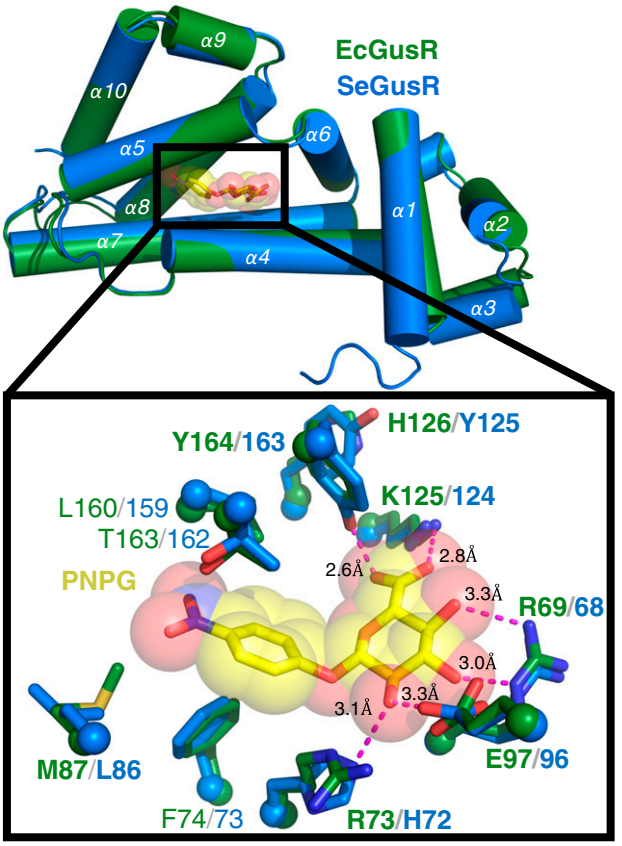
Fig. 3. GusR DNA-binding domain structure and operator site binding. (*A*) Ribbon representation of the EcGusR DBD (green) highlighting DNA-binding residues, and an alignment of GusR sequencing in this region with other TetR family members. Residues important for DNA binding are highlighted in black and gray. (*B*) DNA binding affinities measured by isothermal titration calorimetry for wild-type and variant factors and using two predicted DNA operator sites, 1 and 2 (Fig. S1). Error values are reported in parentheses and italicized, and represent SDs. NB, no binding. (*C*) An ITC heat-map trace of EcGusR binding to DNA operator site 1. (*D*) An ITC heat-map trace of EcGusR preincubated with PNPG and DNA operator site 1 as the ligand titrant.

while the NSAID metabolite indomethacin-acyl-glucuronide (Indo-glucuronide) binds only to SeGusR (K_d 15 μM) (Fig. 5). However, a single ligand-binding pocket L86M mutation in SeGusR confers 15 μM affinity to E_{17}-glucuronide, and improves SeGusR's affinity for Indo-glucuronide to 3.0 μM (Fig. 5). The corresponding mutation in EcGusR (M87L) reduces E_{17}-glucuronide affinity to 19 μM but confers indomethacin-glucuronide binding capabilities to this variant protein (Fig. 5). An alanine mutation in this position (M87A) eliminates EcGusR's affinity to E_{17}-glucuronide and Indo-glucuronide (Fig. 5). Taken together, these data reveal that GusRs bind to glucuronide conjugates primarily via contacts with the sugar moiety but also appear to exhibit species-specific preferences for distinct nonsugar groups, perhaps reflecting differences in glucuronide utilization within the human gut. Additionally, we show that the affinity for a particular ligand can be adjusted by single-residue changes within the GusR effector-binding pocket. Future studies will address GusR binding by other glucuronidated ligands.

Potential Firmicutes GusR Functional Orthologs. We next sought to determine whether GUS expression is controlled by a potential GusR ortholog in gut microbes from the Firmicutes phylum. In Bacteroidetes, the other dominant phylum in the human gut, GUS enzymes and other glycosyl hydrolases are part of well-characterized PULs identifiable by proximal starch-utilization system factors (36). Analogous gram-positive PULs have been described, specific to butyrate-producing bacteria of the Firmicutes phyla (7). Previously, we reported the crystal structures of the GUS proteins from the non–butyrate-producing Firmicutes taxa *Clostridium perfringens* and *Streptococcus agalactiae*; thus, we focused on these two Firmicutes species (33). We found through sequence analysis that Firmicutes do not maintain a GUS operon akin to that observed within the Enterobacteriaceae family of the Proteobacteria phylum. We hypothesized, however, that a glucuronide-responsive transcriptional regulator similar to GusR may be encoded either near the Firmicutes *gusA* genes or elsewhere on the chromosomes of these two species. Thus, we searched locally around the *gusA* gene in *C. perfringens* and *S. agalactiae* to identify a putative GusR ortholog, and searched each organism's reference genome globally to find the closest homolog by sequence identity to the confirmed GusRs outlined above (Fig. S3).

In the global search of the *C. perfringens* reference genome (1152363755), we identified a predicted FadR protein (CpFadR) that shares 20 and 21% sequence homology with EcGusR and SeGusR, respectively. Importantly, it also maintained a lysine in the same sequence position as the lysine residues shown in GusR to be critical for glucuronide ligand binding (Fig. 6*A* and Fig. S4). Thus, we synthesized the gene for CpFadR, overexpressed, purified, and crystallized the CpFadR protein, and determined its structure to 1.9-Å resolution (Fig. 6*A*, Fig. S5, and Table S1). The structure reveals that, like EcGusR and SeGusR, CpFadR is a homodimer that folds into a distinct DBD and EBD, and exhibits a similar overall structure to the Enterobacteriaceae GusR proteins, with 3.4-Å rmsd over 168 and 176 equivalent Cα positions with EcGusR and SeGusR, respectively (Fig. 6*A*). However, despite their fold similarities, the effector-binding pockets of CpFadR and the GusRs are distinct. CpFadR places four amino acid residues into the pocket the GusRs employ to accommodate the ligand in the PNPG–GusR structures; the corresponding CpFadR residues would appear to sterically block glucuronide binding (Fig. 6*A*). Because CpFadR is an apo (ligand-free) state in the crystal structure we resolved, the protein could potentially reposition these residues upon binding to ligand. Therefore, we tested the ability of CpFadR to bind

A



B

GusR Ligand Binding Affinity by ITC (μM)

	EcGusR	SeGusR
WT	0.21 *(0.07)*	2.7 *(0.5)*
K125A / K124A	*NB*	*NB*
Y164A / Y163A	34 *(4)*	106 *(1)*
Y164F / Y163F	10 *(0.8)*	17 *(3)*
H126A / Y125A	0.94 *(0.1)*	2.1 *(0.06)*
E97A / E96A	9.6 *(3)*	35 *(5)*
R73A / H72A	4.5 *(0.4)*	12 *(4)*
R69A / R68A	9.2 *(2)*	34 *(7)*
M87A / L86A	23 *(3)*	2.3 *(0.4)*

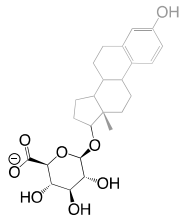
Fig. 4. GusR effector-binding domain structure and function. (*A*) Superposition of PNPG (yellow)-bound EcGusR and SeGusR monomers (green and blue, respectively), with a close-up highlighting side chains that contact PNPG. Residues examined by mutagenesis are indicated (bold). (*B*) PNPG binding affinities measured by ITC for GusR wild-type (WT) and mutant GusR proteins. Error values are reported in parentheses and italicized, and represent SDs. NB, no binding.

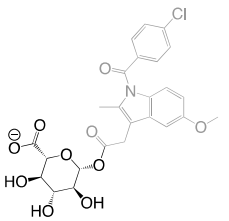
PNPG, PNP-Gluco, or free GlcA in vitro using ITC. We found that CpFadR exhibited no binding to the ligands tested (Fig. S2). Thus, we conclude that, while CpFadR shares the same fold as GusR, it is not a GusR-like glucuronide-responsive repressor in the Firmicutes species *C. perfringens*.

In both *C. perfringens* and *S. agalactiae*, a putative GntR family transcriptional regulator was identified locally in the region adjacent to the *gusA* genes (reference genomes 1152363755 and 674113006, respectively). This was also the most significant hit in the global search of the *S. agalactiae* genome. The GntRs identified share less than 13% sequence identity with EcGusR and SeGusR. CpGntR retains none of the six residues identified in the GusRs to be important for ligand binding, while two residues are conserved in SaGntR (Fig. S4). Despite these differences, we hypothesized, given their proximity to *gusA* genes, that CpGntR and SaGntR may be structural or functional orthologs of the Enterobacteriaceae GusRs. Thus, we synthesized both GntR genes, overexpressed both proteins recombinantly in *E. coli*, and purified them to homogeneity. While we failed to obtain crystals of CpGntR, the apo structure of SaGntR was determined to 1.9-Å resolution (Fig. 6*B* and Table S1). We found that SaGntR is structurally distinct from GusR, forming a homodimer but exhibiting little structural similarity with the glucuronide repressor proteins. SaGntR utilizes a winged helix-turn-helix DBD motif, and its EBD is composed of six α-helices packed to create a barrel-like structure with no evident ligand-binding pocket (Fig. 6*B*). Furthermore, neither CpGntR nor SaGntR exhibited binding to PNPG, PNP-Gluco, or free GlcA in ITC studies (Fig. S2). Taken together, these structure–function data from the Firmicutes taxa *C. perfringens* and *S. agalactiae* indicate that their putative GusR orthologs are not glucuronide-responsive factors. Thus, these species appear to regulate *gusA* gene expression in a manner distinct from the Enterobacteriaceae.

Molecular Determinants of GusR-Mediated Regulation in *E. coli*. We sought to validate the molecular contacts necessary for GusR function in living *E. coli* cells. Because the *gusA* gene is under the control of GusR in this member of the Enterobacteriaceae family, we started by examining changes in GUS activity in cultured *E. coli* BL21 cells in the presence of different potential effector ligands (Fig. 7*A*). We treated cells with potential effectors for 2 h, and then effectors were removed by three rounds of cell pelleting and washing before cell lysis and the subsequent assays of GUS activity (13). We found that free GlcA did not increase GUS activity at concentrations up to 10 mM (Fig. 7*A*). However, increasing concentrations of the GusR ligand PNPG led to a dose-dependent increase in GUS activity, starting with

GusR Endogenous/Exogenous Ligand Binding by ITC (μM)





	E_{17}-Glucuronide	Indo-Glucuronide
EcGusR	11 *(2)*	*NB*
SeGusR	*NB*	15 *(3)*
$EcGusR_{M87L}$	19 *(0.8)*	41 *(17)*
$EcGusR_{M87A}$	*NB*	*NB*
$SeGusR_{L86M}$	15 *(0.9)*	3.0 *(0.5)*

Fig. 5. Binding affinities measured by ITC for GusR WT and variants to estradiol-17-β-D-glucuronide and indomethacin-acyl-β-D-glucuronide. The estradiol and indomethacin scaffolds of the glucuronides are outlined in gray. All values are reported in micromolar concentration. Error values are reported in parentheses and italicized, and represent SDs. NB, no binding.

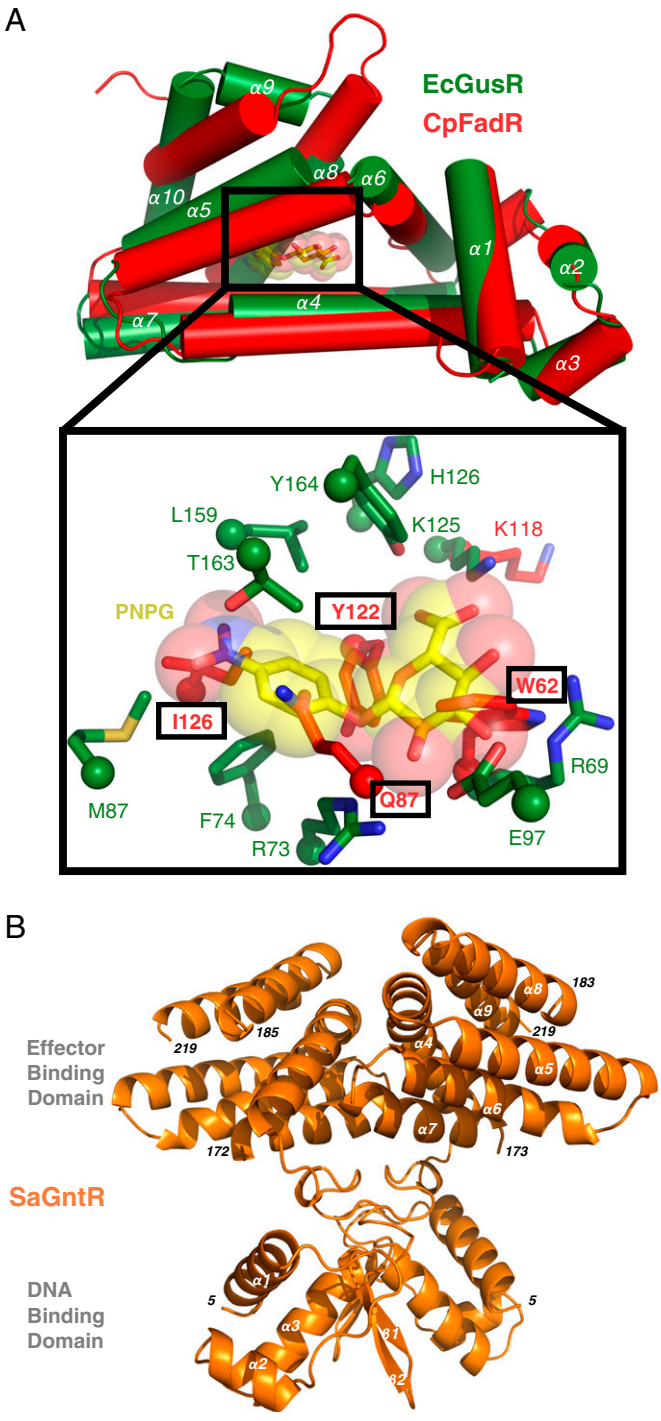


Fig. 6. Putative GusR orthologs from *C. perfringes* and *S. agalactiae*. (*A*) Superposition of EcGusR and *C. perfringes* FadR monomers (green and red, respectively), with a close-up of residues of the putative effector-binding pockets. CpFadR side chains that sterically clash with PNPG are in bold and boxed. (*B*) Ribbon representation of the 1.9-Å crystal structure of the *S. agalactiae* GntR (orange; PDB ID code 6AZ6) homodimer.

basal activity at 1.95 μM PNPG and rising to robust activity at 1 mM PNPG (Fig. 7*A*). By contrast, PNP-Gluco, the near-isostructural ligand to PNPG (Fig. S2), did not increase GUS expression even at 10 mM (Fig. 7*A*). These cell-based results support the in vitro ITC data above indicating that EcGusR binds to PNPG but not to GlcA or PNP-Gluco. Thus, we conclude that EcGusR responds to the presence of PNPG to derepress the GUS operon, allowing for *gusA* expression and an increase of GUS enzyme activity in cultured *E. coli* cells.

Last, we examined how specific GusR residues affect the production of GUS activity in living *E. coli* cells. We introduced plasmids into *E. coli* that contained the gene for wild-type GusR or the K125A GusR variant that failed to bind ligand in vitro (Fig. 4*B*). *E. coli* with an empty plasmid [BL21(DE3)] showed robust GUS activity in the presence of PNPG, likely resulting from endogenous GusR (Fig. 7*B*). Cells containing an expression plasmid for wild-type GusR failed to show GUS activity when PNPG was withheld but exhibited GUS activity when induced with 1 mM PNPG (Fig. 7*B*). By contrast, cells containing an expression plasmid for the K125A GusR variant exhibited no GUS activity even when induced with 1 mM PNPG, indicating that this form of GusR acted as a dominant negative in *E. coli* cells that still retained their endogenous GusR protein (Fig. 7*B*). Additionally, the expression plasmid for *E. coli* UxuR yielded moderate GUS activity with PNPG, similar to levels seen when expression plasmids for wild-type and key mutants of *S. enterica*

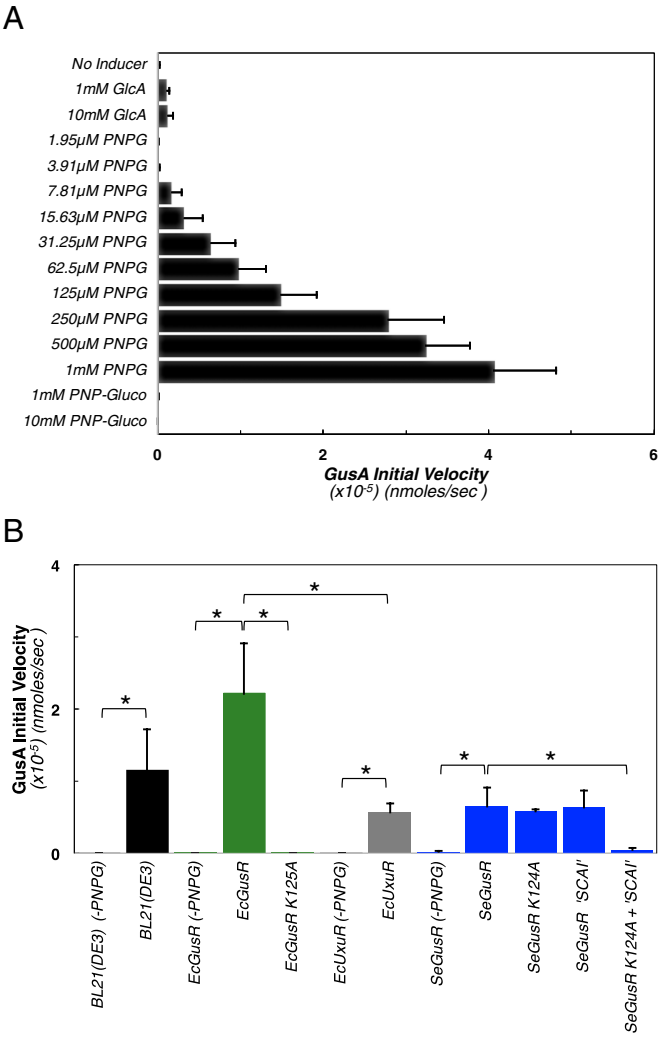


Fig. 7. *E. coli* cell-based GUS activity assays. (*A*) Initial GUS enzyme velocities in *E. coli* BL21 cells in the presence of different potential effector ligands (PNP-Gluco and GlcA) or increasing concentrations of PNPG (1.95 μM to 1 mM). (*B*) Initial GUS enzyme velocities in *E. coli* BL21 cells harboring either an empty vector or expression vectors for EcGusR, EcUxuR, and SeGusR genes induced with 1 mM PNPG. Values are the mean of biological triplicates, run in triplicate for each experiment, and error bars are SEMs. *$P < 0.05$.

GusR were added to *E. coli* cells (Fig. 7*B*). These results indicate that EcUxuR is unable to repress GUS expression in *E. coli*. However, we tested a form of SeGusR in which we combined both the ligand-insensitive mutant K124A and the *E. coli* DNA operator targeting SCAI mutations into a single SeGusR (SCAI + K124A) variant protein. This variant exhibited robust repression of GUS activity in *E. coli* cells, indicating that specific contacts in the *E. coli* GUS operon were formed with this mutant of SeGusR to block GUS expression in cells (Fig. 7*B*). Taken together, these cell-based studies define the molecular determinants important for glucuronide recognition by the GusR transcriptional repressor that controls GUS activity levels in potential gut pathogens like *E. coli* and *S. enterica* (Fig. 8). Furthermore, they show that DNA-binding site-specific but ligand-insensitive receptors can act in a dominant-negative fashion in living *E. coli* cells.

Discussion

The structural basis of GusR-mediated glucuronide recognition by members of the human gut Enterobacteriaceae pathobionts *E. coli* and *S. enterica* is described. By sequence analysis, we find that additional human GI Enterobacteriaceae pathobionts *Shigella*, *Klebsiella*, and *Yersinia* appear to encode GusR orthologs adjacent to GUS operons (Fig. 8). Each of these GusR proteins maintains the conserved residues established here to be functionally important for glucuronide recognition in vitro and in vivo (Fig. 8). In the National Center for Biotechnology Information (NCBI) database, we also found GusR orthologs in *Buttiauxella*, *Erwinia*, and *Raoultella* taxa—Proteobacteria, typically associated with aquatic and soil environments but known to be opportunistic human pathogens. The GusRs from these taxa also maintain the residues shown here to be necessary for glucuronide binding (Fig. 8). The Proteobacteria, and particularly Enterobacteriaceae, can act as condition-specific opportunistic pathogens (45–47). Indeed, a dysbiotic state of microbial imbalance has been associated with the increase in abundance of the Enterobacteriaceae in patients with chronic inflammation and colorectal cancer (45, 48, 49). We speculate that GUS operons provide some Enterobacteriaceae with the ability to utilize intestinal endobiotic and xenobiotic glucuronides as nutrients. UGT enzymes expressed throughout the GI tract have been shown to be efficient producers of a range of phenolic glucuronides (50, 51), some of which we show here to be GusR ligands (Fig. 5 and Fig. S2). Thus, in the presence of suitable glucuronides, GUS operon-containing Enterobacteriaceae may be poised to use this unique source of carbon for colonization and potential expansion in opportunistic conditions.

We find that all microbes encoding a bona fide GusR also maintain elements of a GUS operon, but not all retain a complete GUS operon composed of *gusA*, *gusB*, and *gusC* genes. The NCBI database contains whole-genome sequences for six strains of *E. coli*, and the genomes of five of these six strains encode full GUS operons; by contrast, the pathogenic O157:H7 strain lacks the *gusA* gene that encodes the GUS enzyme. Commensal *E. coli* has been shown to limit the colonization of pathogenic O157:H7 (52, 53), suggesting that an intact GUS operon may provide a competitive advantage within the GI tract. All sequenced strains of GusR-encoding Enterobacteriaceae taxa maintain a complete operon, with the exception of one *Klebsiella* and two *Shigella* strains, which harbor truncated GUS operon genes (Table S3). Taken together, these observations support the conclusion that

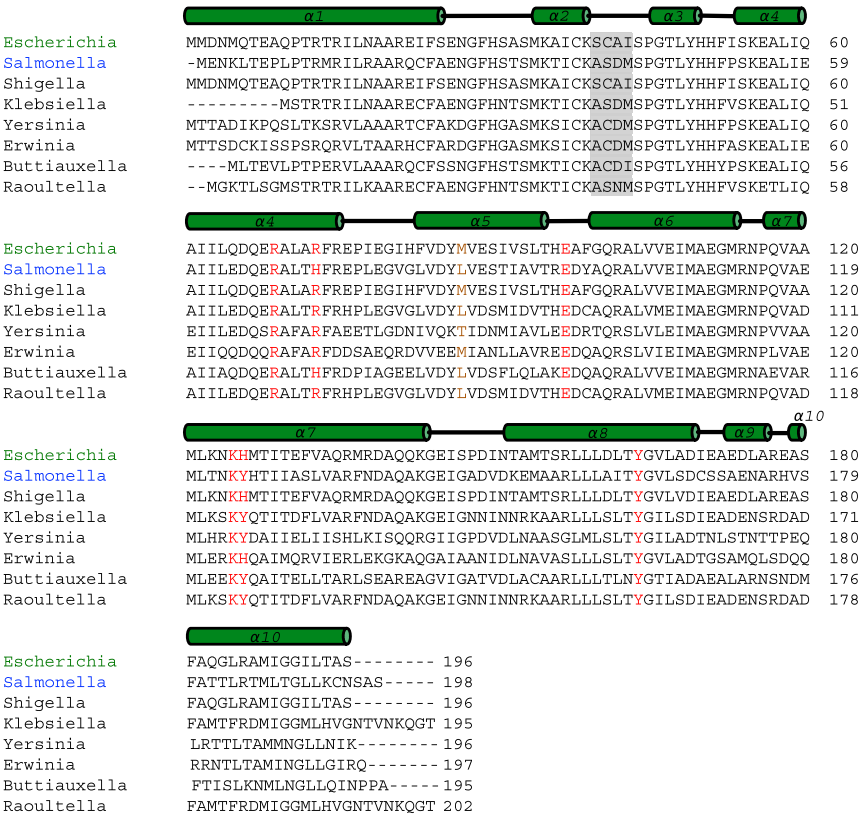


Fig. 8. Sequence alignment of GusRs identified from the NCBI database, with the secondary structure of EcGusR shown above the alignment. Residues important for glucuronide and nonglucuronide recognition are highlighted in red and brown, respectively. Residues important for DNA binding are highlighted in gray.

GUS operons are largely, but not universally, retained in human gut-associated Enterobacteriaceae.

The Proteobacteria are at low abundance in the large intestine, where the primary nutrient source is complex carbohydrates (9, 45, 54, 55). The dominant phyla of the large intestine is Bacteroidetes, which utilize their PUL-encoded machinery to degrade complex carbohydrates (9, 36, 56). The Proteobacteria are associated with the small intestine in the healthy mammalian gut, and appear to compete with host tissues for sugars and other simple carbohydrates in this region of the GI tract (54, 55, 57). Intact GUS operons within the Enterobacteriaceae pathobionts would appear to give them a competitive advantage for nutrient consumption over other Proteobacteria and Firmicutes microbes in the small intestine, particularly downstream from the bile duct that delivers glucuronides from the liver (19, 22, 23).

As we have shown with our estrogen- and indomethacin-glucuronides, different microbial species are equipped to respond to distinct endobiotic and xenobiotic glucuronides, perhaps reflecting differences in glucuronide utilization in the gut (Fig. 5). The molecular basis for this distinction involves a key effector-binding pocket residue, M87 in EcGusR and L86 in SeGusR, and mutational swaps in this position alter ligand binding specificity (Fig. 5). We note that, of the residues that contact the *p*-nitrophenol group in our crystal structures (Fig. 4*A*), this is the only position that varies in the sequences of the Enterobacteriaceae GusRs (Fig. 8). It samples Leu, Met, and Thr amino acids, while the F74/73 (EcGusR/SeGusR), L160/159, and T163/162 residues near the nonglucuronide moiety are almost completely conserved; only Buttiauxella GusR harbors a change in one of these positions, replacing T163/162 with an asparagine (Fig. 8). These observations support the conclusion that the M87/L86 position is an important site for ligand binding specificity in the GusR family of transcription factors.

Although we were able to express, purify, and study EcUxuR, we were unable to identify its DNA-binding element or determine its crystal structure. Furthermore, EcUxuR failed to repress GUS activity in our cell-based studies (Fig. 7*B*). Previous research has shown that this GusR homolog binds to a glucuronide metabolite, D-fructuronate, which may stabilize the protein and perhaps make it amenable for future crystallographic analysis (58). Finally, because different ligands (e.g., indomethacin- and estrogen-glucuronides) show differential binding to distinct GusR proteins (e.g., *E. coli* and *S. enterica*), it is possible that distinct GUS enzymes will also show marked preferences for differing chemical classes of glucuronide substrates. Examining this possibility will be the subject of future GUS structure, function, and inhibition studies.

Methods

Cloning and Expression. *E. coli* GusR and UxuR were cloned from genomic DNA, all other genes were synthesized by GenScript. Genes were cloned into a pLIC vector with an N-terminal 6×histidine tag with a tobacco etch virus (TEV) protease-cleavage site. BL21-Gold competent cells (Agilent Technologies) were transformed with the expression plasmids for each protein and cultured in the presence of antifoam (50 μL) and ampicillin (100 μg/mL). Protein expression was carried out in autoinducing ZYP-5052 media shaking at 325 rpm at 37 °C; when an $OD_{600\ nm}$ of ∼1.0 was reached the temperature was reduced to 18 °C (59). Selenomethionine-substituted protein was expressed in PASM-5052 media (59).

Protein expression for SaGntR and CpFadR was carried out in LB media shaking at 250 rpm at 37 °C until an OD_{600} of 0.6 was attained. Expression was induced with the addition of 0.1 mM isopropyl-1-thio-D-galactopyranoside (IPTG), temperature was decreased to 18 °C, and bacteria were cultured overnight. Selenomethionine-substituted protein was expressed in SelenoMet medium (Molecular Dimensions).

Protein Purification. Cell pellets were collected by centrifugation at 4,500 × *g* for 20 min at 4 °C in a Sorvall (model RC-3B) swinging-bucket centrifuge. Cell pellets were resuspended in buffer A [20 mM potassium phosphate, pH 7.4, 50 mM imidazole, 500 mM NaCl, 1 mM Tris(2-carboxyethyl)phosphine (TCEP)]. along with lysozyme, DNase1, and protease inhibitor tablets. Cells were sonicated, and cell lysate was separated into insoluble and soluble fractions by centrifugation at 17,000 × *g* for 60 min in a Sorvall (model RC-5B) centrifuge. Soluble fractions were syringe-filtered through a sterilized 22-μm polyethersulfone (PES) membrane, applied to an Ni-NTA HisTrap gravity column (GE Healthcare Life Sciences), and washed with buffer A. The bound protein was eluted with buffer B (20 mM potassium phosphate, pH 7.4, 500 mM imidazole, 500 mM NaCl, 1 mM TCEP) and applied to an S200 gel-filtration column in buffer C [20 mM 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (Hepes), 300 mM NaCl, 1 mM TCEP] on an ÄKTAxpress (GE Healthcare Life Sciences). GusR proteins eluted from the S200 column as a single peak in buffer C. The resultant affinity-tagged GusR proteins were incubated overnight with TEV protease at 4 °C to remove the hexahistidine tag, leaving only serine, asparagine, and alanine as nonnative amino acid residues on the N terminus of the protein. This sample was again applied to an Ni-NTA HisTrap gravity column, which retains the histidine tag, and the GusR protein was collected in the flowthrough. Resultant GusR proteins were concentrated to a volume of <5 mL with 10-kDa molecular weight cutoff centrifugal concentrators (EMD Millipore) and separated using an S200 gel-filtration column. Fractions were analyzed by SDS/PAGE and those with >95% purity were combined, concentrated, and snap-frozen using liquid nitrogen and stored at −80 °C.

Crystallization. Initial crystallization conditions were identified in 96-well sitting-drop trays using a Rigaku Phoenix Liquid Handler with 15 mg/mL GusR and 10 mM PNPG (greater than 10-fold molar excess). Crystallization trays were monitored and tracked by Rigaku Gallery 700 plate hotels held at 20 °C. The initial crystal hit for EcGusR plus PNPG (20% PEG 8000, 0.1 M Hepes, pH 7.5; with 10-fold molar excess PNPG) could not be reproduced in the laboratory in hanging-drop vapor-diffusion crystallization trays (Qiagen). However, EcGusR crystals could be reproduced in 96-well sitting-drop trays, and could be looped and cryoprotected in 20% glycerol. Data from these crystals were collected to 2.1-Å resolution at the Advanced Photon Source (APS) using General Medical Sciences and National Cancer Institute Collaborative Access Team (GM/CA CAT) beamline 23ID-B at 100 K. SeGusR was crystallized in 25% PEG 3350, 0.2 M ammonium acetate, 0.1 M Bis-Tris (pH 6.5) (with 10-fold molar excess PNPG) and readily reproduced in the laboratory using hanging-drop vapor-diffusion crystallization trays. These crystals were cryoprotected in 20% glycerol and employed for data collection at GM/CA CAT beamline 23ID-D at the APS. Selenomethionine-substituted SeGusR was crystallized as described above and could be employed for single-wavelength anomalous dispersion phasing, as outlined below.

Selenomethionine-substituted SaGntR was screened in the same way as described above for EcGusR and SeGusR. Screening hits were refined in 15-well Qiagen vapor-diffusion hanging-drop trays. Final crystals were grown overnight at room temperature in 0.2 M triammonium citrate, 12% PEG 3350, and SaGntR at 15 mg/mL. Crystals were cryoprotected in the same condition as the crystallant with 20% glycerol.

Selenomethionine-substituted CpGntR was screened in the same way described above for EcGusR and SeGusR. Screening hits were refined in 15-well Qiagen vapor-diffusion hanging-drop trays. Final crystals were grown overnight at 20 °C in 0.2 M sodium formate, 20% PEG 3350, and CpGntR at 6.8 mg/mL. Crystals were cryoprotected with Fomblin (Sigma-Aldrich).

Structure Determination. An inverse-beam collection method at a wavelength of 0.97939 Å was applied to selenomethionine-substituted SeGusR cocrystallized with PNPG, which diffracted to 2.1-Å resolution. Collection data were autoprocessed by X-ray Detector Software (XDS) (60, 61). The structure of SeGusR was determined using the AutoSol tool in PHENIX (62). After phasing, initial model building was performed in PHENIX via the AutoBuild function. Subsequent refinements and manual building were performed in PHENIX and Coot (63), respectively. The final model consisting of 192 amino acids was built with one molecule in the asymmetric unit (ASU). The final model was built and refined with an R_{work} and R_{free} of 0.1833 and 0.2220, respectively (Table S1). Atomic coordinates and structure factors have been deposited in the Protein Data Bank with ID code 6AYH.

EcGusR cocrystallized with PNPG diffracted to 2.1-Å resolution using a native beam collection method at a wavelength of 1.0332 Å. Collection data were autoprocessed by XDS. The structure of SeGusR was used to determine the correct phase of EcGusR, using the Phaser molecular replacement tool in PHENIX. After phasing, initial model building was performed in PHENIX via the AutoBuild function. Subsequent refinements and manual building were performed in PHENIX and Coot, respectively. The final model consisting of 648 amino acids was built with four molecules in the ASU. The final model was refined with an R_{work} and R_{free} of 0.1823 and 0.2280, respectively (Table

S1). Atomic coordinates and structure factors have been deposited in the Protein Data Bank with ID code 6AYI.

An inverse-beam collection method at a wavelength of 0.97935 Å was applied to selenomethionine-substituted SaGntR crystals, which diffracted to 1.9-Å resolution. The structure of SaGntR was determined using the AutoSol tool in PHENIX. After phasing, initial model building was performed in PHENIX via the AutoBuild function. Subsequent refinements and manual building were performed in PHENIX and Coot, respectively. The final model consisting of 404 amino acids was built with two molecules in the ASU. The final model refined with an R_{work} and R_{free} of 0.1940 and 0.2289, respectively (Table S1). Atomic coordinates and structure factors have been deposited in the Protein Data Bank with ID code 6AZ6.

An inverse-beam collection method at a wavelength of 0.97935 Å was applied to selenomethionine-substituted CpGntR crystals, which diffracted to 1.8-Å resolution. The structure of CpGntR was determined using the AutoSol tool in PHENIX. After phasing, initial model building was performed in PHENIX via the AutoBuild function. Subsequent refinements and manual building were performed in PHENIX and Coot, respectively. The final model consisting of 373 amino acids was built with two molecules in the ASU. The final model refined with an R_{work} and R_{free} of 0.1964 and 0.2253, respectively (Table S1). Atomic coordinates and structure factors have been deposited in the Protein Data Bank with ID code 6AZH.

Preparation of dsDNA. Single-stranded DNA (ssDNA) sequences and complements were ordered from Integrated DNA Technologies (IDT) of predicted operator sites. ssDNAs were dissolved in annealing buffer (100 mM potassium acetate, 30 mM Hepes, pH 7.5). Complementary ssDNAs were mixed together in a 1:1 molar ratio (1 mM). The mix of ssDNAs was heated to 95 °C for 5 min and then gradually cooled back to room temperature. The resultant dsDNAs were buffer-exchanged using Bio-Spin 6 columns (Bio-Rad) into GusR S200 buffer (buffer C).

Isothermal Titration Calorimetry Binding Studies. All ITC measurements were performed at 25 °C using an Auto-ITC200 microcalorimeter (MicroCal/GE Healthcare). The calorimetry cell (volume 200 mL) was loaded with GusR wild-type or mutant protein at a concentration of 50 μM or 100 μM for weak binding mutants. The syringe was loaded with a substrate (dsDNA or glucuronide conjugate) concentration of 1 to 2 mM in a buffer identical to that employed for the protein. A typical injection protocol included a single 0.2-μL first injection followed by 26 1.5-μL injections of the substrate into the calorimetry cell. The spacing between injections was kept at 180 s and the reference power at 8 μcal/s. A control experiment was performed by titrating ligand (dsDNA or glucuronide conjugate) into buffer under identical settings to determine the heat signals that arose from compound dilution; these were subtracted from the heat signals of protein–compound interaction. The data were analyzed using Origin for ITC, version 7.0, software supplied by the manufacturer, and fit well to a one-site binding model.

In Vivo Induction Assay. A 5-mL overnight culture of BL21-Gold competent cells harboring pLIC-His empty vector, UxuR, GusR, or a mutant thereof was grown in LB broth with 100 μg/mL ampicillin. Twenty microliters of the overnight culture was added to 2 mL of fresh LB broth and incubated at 37 °C for 1 h while shaking. PNPG or PNP-Gluco at a final concentration of 1 mM was employed to induce the GUS operon of each culture, along with a no-inducer control. Additionally, cultures harboring protein expression plasmids were induced with 0.1 mM IPTG to induce protein expression. These cultures were incubated for another 3 h at 37 °C while shaking. One-milliliter samples of each culture were transferred to a microcentrifuge tube and spun down at 13,000 × *g* for 10 min in a Sorvall (model Legend Micro 17). The supernatant was decanted and the cells were suspended in 1 mL of fresh LB broth with 100 μg/mL chloramphenicol (Cam), and then centrifuged again for 10 min and resuspended in another 1 mL LB-Cam mix. These twice-washed cell samples were permeabilized by adding a single drop of 0.1% SDS and two drops of chloroform and vortexing for 30 s. Twenty microliters of the aforementioned washed cell samples was added to each well in a 96-well Corning flat clear-bottom black polystyrene microplate. Finally, 80 μL of buffer (1.25 mM PNPG, 150 mM NaCl, 20 mM Hepes, pH 7.5) preincubated at 37 °C was added before starting the assay in the plate reader to monitor absorbance at 410 nm (PHERAstar) (13).

1. Ng KM, et al. (2013) Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* 502:96–99.
2. Kovatcheva-Datchary P, et al. (2015) Dietary fiber-induced improvement in glucose metabolism is associated with increased abundance of Prevotella. *Cell Metab* 22: 971–982.
3. Koropatkin NM, Cameron EA, Martens EC (2012) How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* 10:323–335.
4. Turnbaugh PJ, et al. (2009) The effect of diet on the human gut microbiome: A metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med* 1:6ra14.
5. Xu X, Xu P, Ma C, Tang J, Zhang X (2013) Gut microbiota, host health, and polysaccharides. *Biotechnol Adv* 31:318–337.
6. Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA (2008) Polysaccharide utilization by gut bacteria: Potential for new insights from genomic analysis. *Nat Rev Microbiol* 6: 121–131.
7. Sheridan PO, et al. (2016) Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic Firmicutes. *Microb Genom* 2:e000043.
8. Terrapon N, Lombard V, Gilbert HJ, Henrissat B (2015) Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* 31:647–655.
9. Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H (2017) Polysaccharide utilization loci: Fueling microbial communities. *J Bacteriol* 199:e00860-16.
10. Glenwright AJ, et al. (2017) Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* 541:407–411.
11. Blanco C, Ritzenthaler P, Mata-Gilsinger M (1982) Cloning and endonuclease restriction analysis of uidA and uidR genes in *Escherichia coli* K-12: Determination of transcription direction for the uidA gene. *J Bacteriol* 149:587–594.
12. Liang WJ, et al. (2005) The gusBC genes of *Escherichia coli* encode a glucuronide transport system. *J Bacteriol* 187:2377–2385.
13. Wilson KJ, Hughes SG, Jefferson RA (1992) The *Escherichia coli* gus operon: Induction and expression of the gus operon in *E. coli* and the occurrence and use of GUS in other bacteria. *GUS Protocols. Using the GUS Gene as Reporter of Gene Expression*, ed Gallagher SR (Academic, San Diego), pp 7–22.
14. Blanco C, Ritzenthaler P, Mata-Gilsinger M (1985) Nucleotide sequence of a regulatory region of the uidA gene in *Escherichia coli* K12. *Mol Gen Genet* 199:101–105.
15. Novel M, Novel G (1976) Regulation of beta-glucuronidase synthesis in *Escherichia coli* K-12: Pleiotropic constitutive mutations affecting uxu and uidA expression. *J Bacteriol* 127:418–432.
16. Novel G, Didier-Fichet ML, Stoeber F (1974) Inducibility of beta-glucuronidase in wild-type and hexuronate-negative mutants of *Escherichia coli* K-12. *J Bacteriol* 120:89–95.
17. Tephly TR, Burchell B (1990) UDP-glucuronosyltransferases: A family of detoxifying enzymes. *Trends Pharmacol Sci* 11:276–279.
18. Rowland A, Miners JO, Mackenzie PI (2013) The UDP-glucuronosyltransferases: Their role in drug metabolism and detoxification. *Int J Biochem Cell Biol* 45:1121–1132.
19. Miners JO, Mackenzie PI (1991) Drug glucuronidation in humans. *Pharmacol Ther* 51: 347–369.
20. Gloux K, et al. (2011) A metagenomic β-glucuronidase uncovers a core adaptive function of the human intestinal microbiome. *Proc Natl Acad Sci USA* 108:4539–4546.
21. Pollet RM, et al. (2017) An atlas of β-glucuronidases in the human intestinal microbiome. *Structure* 25:967–977.e5.
22. Kreek MJ, Guggenheim FG, Ross JE, Tapley DF (1963) Glucuronide formation in the transport of testosterone and androstenedione by rat intestine. *Biochim Biophys Acta* 74:418–427.
23. Pellock SJ, Redinbo MR (2017) Glucuronides in the gut: Sugar-driven symbioses between microbe and host. *J Biol Chem* 292:8569–8576.
24. Boelsterli UA, Redinbo MR, Saitta KS (2013) Multiple NSAID-induced hits injure the small intestine: Underlying mechanisms and novel strategies. *Toxicol Sci* 131:654–667.
25. Mani S, Boelsterli UA, Redinbo MR (2014) Understanding and modulating mammalian-microbial communication for improved human health. *Annu Rev Pharmacol Toxicol* 54:559–580.
26. Flores R, et al. (2012) Fecal microbial determinants of fecal and systemic estrogens and estrogen metabolites: A cross-sectional study. *J Transl Med* 10:253.
27. Hullar MAJ, Burnett-Hartman AN, Lampe JW (2014) Gut microbes, diet, and cancer. *Cancer Treat Res* 159:377–399.
28. Plotnikoff GA (2014) Three measurable and modifiable enteric microbial biotransformations relevant to cancer prevention and treatment. *Glob Adv Health Med* 3: 33–43.
29. Koppel N, Maini Rekdal V, Balskus EP (2017) Chemical transformation of xenobiotics by the human gut microbiota. *Science* 356:eaag2770.
30. LoGuidice A, Wallace BD, Bendel L, Redinbo MR, Boelsterli UA (2012) Pharmacologic targeting of bacterial β-glucuronidase alleviates nonsteroidal anti-inflammatory drug-induced enteropathy in mice. *J Pharmacol Exp Ther* 341:447–454.
31. Roberts AB, Wallace BD, Venkatesh MK, Mani S, Redinbo MR (2013) Molecular insights into microbial β-glucuronidase inhibition to abrogate CPT-11 toxicity. *Mol Pharmacol* 84:208–217.
32. Wallace BD, et al. (2010) Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science* 330:831–835.
33. Wallace BD, et al. (2015) Structure and inhibition of microbiome β-glucuronidases essential to the alleviation of cancer drug toxicity. *Chem Biol* 22:1238–1249.
34. Saitta KS, et al. (2014) Bacterial β-glucuronidase inhibition protects mice against enteropathy induced by indomethacin, ketoprofen or diclofenac: Mode of action and pharmacokinetics. *Xenobiotica* 44:28–35.
35. Fabich AJ, et al. (2008) Comparison of carbon nutrition for pathogenic and commensal *Escherichia coli* strains in the mouse intestine. *Infect Immun* 76:1143–1152.
36. Ravcheev DA, Godzik A, Osterman AL, Rodionov DA (2013) Polysaccharides utilization in human gut bacterium Bacteroides thetaiotaomicron: Comparative genomics reconstruction of metabolic and regulatory networks. *BMC Genomics* 14:873.

37. Görke B, Stülke J (2008) Carbon catabolite repression in bacteria: Many ways to make the most out of nutrients. *Nat Rev Microbiol* 6:613–624.
38. Kremling A, Geiselmann J, Ropers D, de Jong H (2015) Understanding carbon catabolite repression in *Escherichia coli* using quantitative models. *Trends Microbiol* 23: 99–109.
39. Ritzenthaler P, Blanco C, Mata-Gilsinger M (1983) Interchangeability of repressors for the control of the uxu and uid operons in *E. coli* K12. *Mol Gen Genet* 191:263–270.
40. Ritzenthaler P, Mata-Gilsinger M (1982) Use of in vitro gene fusions to study the uxuR regulatory gene in *Escherichia coli* K-12: Direction of transcription and regulation of its expression. *J Bacteriol* 150:1040–1047.
41. Ramos JL, et al. (2005) The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev* 69:326–356.
42. Schumacher MA, et al. (2002) Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. *EMBO J* 21:1210–1218.
43. Orth P, Schnappinger D, Hillen W, Saenger W, Hinrichs W (2000) Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat Struct Biol* 7:215–219.
44. Blanco C (1987) Transcriptional and translational signals of the uidA gene in *Escherichia coli* K12. *Mol Gen Genet* 208:490–498.
45. Shin NR, Whon TW, Bae JW (2015) Proteobacteria: Microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol* 33:496–503.
46. Chow J, Tang H, Mazmanian SK (2011) Pathobionts of the gastrointestinal microbiota and inflammatory disease. *Curr Opin Immunol* 23:473–480.
47. Bäumler AJ, Sperandio V (2016) Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* 535:85–93.
48. Yang Y, Jobin C (2014) Microbial imbalance and intestinal pathologies: Connections and contributions. *Dis Model Mech* 7:1131–1142.
49. Carding S, Verbeke K, Vipond DT, Corfe BM, Owen LJ (2015) Dysbiosis of the gut microbiota in disease. *Microb Ecol Health Dis* 26:26191.
50. Cheng Z, Radominska-Pandya A, Tephly TR (1999) Studies on the substrate specificity of human intestinal UDP-glucuronosyltransferases 1A8 and 1A10. *Drug Metab Dispos* 27:1165–1170.
51. Ritter JK (2007) Intestinal UGTs as potential modifiers of pharmacokinetics and biological responses to drugs and xenobiotics. *Expert Opin Drug Metab Toxicol* 3: 93–107.
52. Gamage SD, Patton AK, Strasser JE, Chalk CL, Weiss AA (2006) Commensal bacteria influence *Escherichia coli* O157:H7 persistence and Shiga toxin production in the mouse intestine. *Infect Immun* 74:1977–1983.
53. Leatham MP, et al. (2009) Precolonized human commensal *Escherichia coli* strains serve as a barrier to *E. coli* O157:H7 growth in the streptomycin-treated mouse intestine. *Infect Immun* 77:2876–2886.
54. Gu S, et al. (2013) Bacterial community mapping of the mouse gastrointestinal tract. *PLoS One* 8:e74957.
55. Kamada N, Chen GY, Inohara N, Núñez G (2013) Control of pathogens and pathobionts by the gut microbiota. *Nat Immunol* 14:685–690.
56. White BA, Lamed R, Bayer EA, Flint HJ (2014) Biomass utilization by gut microbiomes. *Annu Rev Microbiol* 68:279–296.
57. Zoetendal EG, et al. (2012) The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *ISME J* 6:1415–1426.
58. Bates Utz C, Nguyen AB, Smalley DJ, Anderson AB, Conway T (2004) GntP is the *Escherichia coli* fructuronic acid transporter and belongs to the UxuR regulon. *J Bacteriol* 186:7690–7696.
59. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41:207–234.
60. Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr Sect D Biol Crystallogr* 66:133–144.
61. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125–132.
62. Adams PD, et al. (2010) PHENIX : A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* 66: 213–221.
63. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12):2126–2132.