

UC Riverside

UC Riverside Previously Published Works

Title

Structure of galactarate dehydratase, a new fold in an enolase involved in bacterial fitness after antibiotic treatment

Permalink

<https://escholarship.org/uc/item/2t98c57g>

Journal

Protein Science, 29(3)

ISSN

0961-8368

Authors

Rosas-Lemus, Monica
Minasov, George
Shuvalova, Ludmilla
et al.

Publication Date

2020-03-01


DOI

10.1002/pro.3796

Peer reviewed

ARTICLE

Structure of galactarate dehydratase, a new fold in an enolase involved in bacterial fitness after antibiotic treatment

Monica Rosas-Lemus^{1,2} | George Minasov^{1,2} | Ludmilla Shuvalova^{1,2} |
 Zdzislaw Wawrzak³ | Olga Kiryukhina^{1,2} | Nathan Mih⁴ |
 Lukasz Jaroszewski^{2,5} | Bernhard Palsson^{4,6} | Adam Godzik^{2,5} |
 Karla J. F. Satchell^{1,2} 

¹Department of Microbiology-Immunology, Northwestern University, Feinberg School of Medicine, Chicago, Illinois

²Center for Structural Genomics of Infectious Diseases, Northwestern University, Feinberg School of Medicine, Chicago, Illinois

³Northwestern Synchrotron Research Center-LS-CAT, Northwestern University, Argonne, Illinois

⁴Department of Bioengineering, University of California San Diego, La Jolla, California

⁵Department of Biomedical Sciences, University of California at Riverside, Riverside, California

⁶Systems Biology Center for Antibiotic Resistance, University of California San Diego, La Jolla, California

Correspondence

Karla J. F. Satchell, Northwestern University, Feinberg School of Medicine, Chicago, IL 60611.

Email: k-satchell@northwestern.edu

Funding information

National Institute of Allergy and Infectious Diseases, Grant/Award Numbers: HHSN272201200026C, HHSN272201700060C, U01AI124316; National Institute of General Medical Sciences, Grant/Award Numbers: GM05789, GM118187

Abstract

Galactarate dehydratase (GarD) is the first enzyme in the galactarate/glucarate pathway and catalyzes the dehydration of galactarate to 3-keto-5-dehydroxygalactarate. This protein is known to increase colonization fitness of intestinal pathogens in antibiotic-treated mice and to promote bacterial survival during stress. The galactarate/glucarate pathway is widespread in bacteria, but not in humans, and thus could be a target to develop new inhibitors for use in combination therapy to combat antibiotic resistance. The structure of almost all the enzymes of the galactarate/glucarate pathway were solved previously, except for GarD, for which only the structure of the N-terminal domain was determined previously. Herein, we report the first crystal structure of full-length GarD solved using a selenomethoionine derivative revealing a new protein fold. The protein consists of three domains, each presenting a novel twist as compared to their distant homologs. GarD in the crystal structure forms dimers and each monomer consists of three domains. The N-terminal domain is comprised of a β -clip fold, connected to the second domain by a long unstructured linker. The second domain serves as a dimerization interface between two monomers. The C-terminal domain forms an unusual variant of a Rossmann fold with a crossover and is built around a seven-stranded parallel β -sheet supported by nine α -helices. A metal binding site in the C-terminal domain is occupied by Ca^{2+} . The activity of GarD was corroborated by the production of 5-keto-4-deoxy-D-glucarate under reducing conditions and in the presence of iron. Thus, GarD is an unusual enolase with a novel protein fold never previously seen in this class of enzymes.

KEYWORDS

antibiotic treatment, bacterial fitness, enolase, *Escherichia coli*, galactarate dehydratase, intestinal pathogens, novel fold

1 | INTRODUCTION

The recent increase of antibiotic resistant bacteria is a serious public health problem that requires finding new therapeutic targets.¹ Bacteria can adapt to almost any environment through metabolic changes regulated by the availability of nutrients, metals, and oxygen.^{2,3} Pathogenic bacteria also take advantage of the metabolites produced within the human host by using metabolic pathways that are completely distinct, or even absent in the host.^{3–6} These unique pathways in pathogens could serve as specific targets for new therapeutics.

Several studies recently described the interaction of pathogens with its host, as well as the metabolic pathways that facilitate bacterial colonization of specific tissues. The use of broad-spectrum antibiotics alters the composition of the microbiota and causes dysbiosis, which disturbs the redox potential and can promote colonization by opportunistic pathogens due to the availability of substrates and electron acceptors.^{2,3,7–10} This process complicates the treatment of infections. For example, the concentration of glucarate and galactarate in the caecum increases significantly after treatment of mice with streptomycin. These dicarboxylic sugars are produced by nitrosylation of glucose and galactose induced by the immune system response.¹¹

The fermentative metabolic glucarate/galactarate pathway is widespread in bacteria although, depending on the organism, galactarate or glucarate can be metabolized by different enzymes. In *Escherichia coli*, galactarate dehydratase (GarD) and glucarate dehydratase (GudD) are the first enzymes of this pathway (Figure 1).^{12,13} The enzymes dehydrate galactarate (mucic acid) and glucarate (saccharic acid), respectively, to produce D-5-keto-4-deoxyglucarate (5-KDG), which is then hydrolyzed by 5-keto-4-dehydroxy-D-glucarate aldolase (GarL) into pyruvic acid and tartronate semialdehyde. The pyruvic acid may be used directly for

fermentation, but tartronate semialdehyde is further reduced by tartronate semialdehyde dehydrogenase into glycerate. Finally, glycerate kinase (GK) phosphorylates glycerate to produce 2-phosphoglycerate. Interestingly, GarD is specific for galactarate, while GudD uses both glucarate and galactarate as substrates to produce a secondary isomer of 5-keto-3-deoxyglucarate from glucarate.¹²

The deletion of *garD* or *gudD* from native *E. coli* and *Salmonella enterica subsp. enterica* serovar Typhimurium (*S. Typhimurium*) demonstrated a significant decrease in the invasiveness to the gut of antibiotic-treated mice, indicating that the metabolism of these sugars is advantageous for the fitness of bacteria when colonizing the intestine.¹¹ Furthermore, adaptive laboratory evolution experiments¹⁴ have shown that, in *E. coli*, the gene *garD* acquired an E207D mutation when exposed to ionic stress. Since the activity of the enzyme is important for growth in the gut of antibiotic-treated mice, solving its structure could be useful to understand its mechanism of action and probably facilitate development of inhibitors in silico to target the galactarate pathway.

Sequence analysis of GarD suggests that it is composed of two domains: the N-terminal Sialic Acid Synthetases, Type III Anti-freeze proteins and bacterial Flagellar Chaperone (SAF) domain (Pfam family SAF [PF08666])¹⁵ and a C-terminal D-galactarate dehydratase/altronate hydrolase C terminus domain (Pfam family GD_AH_C [PF04295]).^{13,15} Several structures of the SAF domain have been solved, including one from the GarD protein, but no structure from the GD_AH_C (PF04295) family have been solved so far.

Here, we report the first crystal structure of full-length GarD, supporting the sequence analysis and revealing additional information. As expected, the N-terminus SAF domain has a β -clip fold. However, the sequence defined GD_AH_C (PF04295) domain was shown to consist of two separate structural domains; one domain formed by three parallel β -strands surrounded by three alpha helices, and

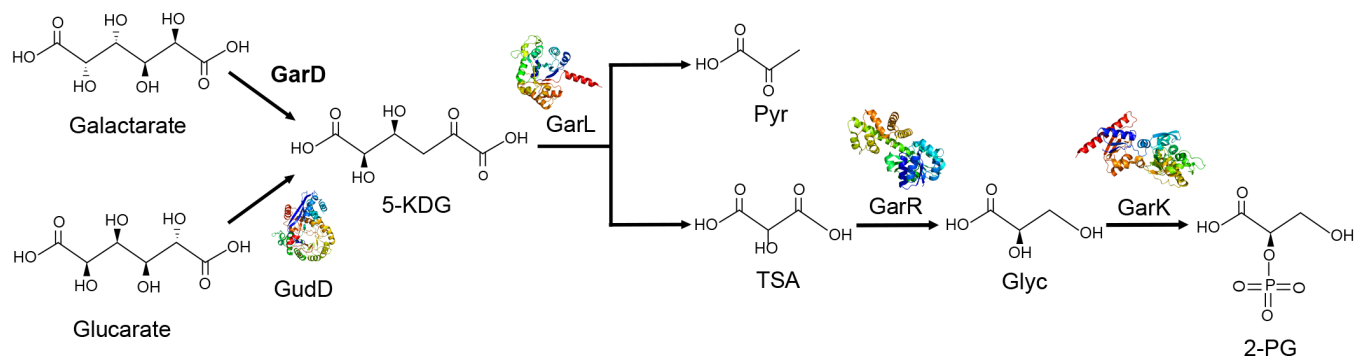


FIGURE 1 The glucarate/galactarate pathway from *E. coli*. The name of the substrates and products are shown below each structure: 5-keto-4-deoxy-D-glucarate (5-KDG), glycerate (Glyc), and 2-phosphoglycerate (2-PG). The corresponding enzymes of each step of the pathway are shown as cartoon models from structures deposited to PDB: glucarate dehydratase, (GudD, PDB ID:1EC7), 5-keto-4-deoxy-D-glucarate aldolase (GarL, PDB ID: 1DXE), tartronate semialdehyde dehydrogenase (GarR, PDB ID:1VDP), and glycerate kinase (GarK, PDB ID: 3CWC)

the C-terminal α/β domain with a topology resembling a Rossmann fold, but with an unusual helix crossover and a possibly catalytic metal binding site. GarD homologs are broadly distributed in bacteria and archaea, with close homologs found in several pathogenic bacteria known for their antibiotic resistance. The high degree of conservation of metal binding motifs suggests that this novel structure could be exploited for development of drugs to treat intestinal infections.

2 | RESULTS

2.1 | GarD is a protein with a novel fold

GarD is a 523 aa protein with a molecular weight of 56.4 kDa. The crystal structure was determined and refined at 2.75 Å resolution (PDB entry 6U7L). Crystals belong to monoclinic space group $P2_1$ with four polypeptide chains in the asymmetric unit. These four chains form two homodimers, predicted by PISA¹⁶ to be stable in solution with estimated of 4,400 Å² buried surface area (Figure 2c). This was supported by size exclusion chromatography as well as clear native gel electrophoresis, where the dimer was

observed as a main stable oligomeric state in solution with the estimated molecular weight of 130 kDa (Figure 2d,e).

Each monomer could be divided into three domains. The N-terminal SAF domain (residues 1–95) has a β -clip fold formed by seven β -strands. There is a long linker (residues 96–119) that connects the N-terminal domain with the second domain, wrapping itself around the third domain. The second domain (residues 120–278) has three β -strands surrounded by three long α -helices finished with a β -strand that follows to the C-terminal from the same chain. Also, this domain interacts with the C-terminal domain from the second chain to create the dimerization surface. The C-terminal domain (residues 279–523), which represents the core of the protein, consists of a seven parallel-stranded β -sheet, surrounded by eight α -helices in a Rossmann-like fold and has a metal binding site.

Coordinates of chain A were uploaded to the DALI¹⁷ and FATCAT¹⁸ servers to search for proteins with similar structure. No structures matching the full-length proteins were found. However, when individual domains were analyzed separately, each found several similar structures in the Protein Data Bank (PDB). Overall this supports the observation that GarD has a novel fold, or at least a novel combination of structural domains.

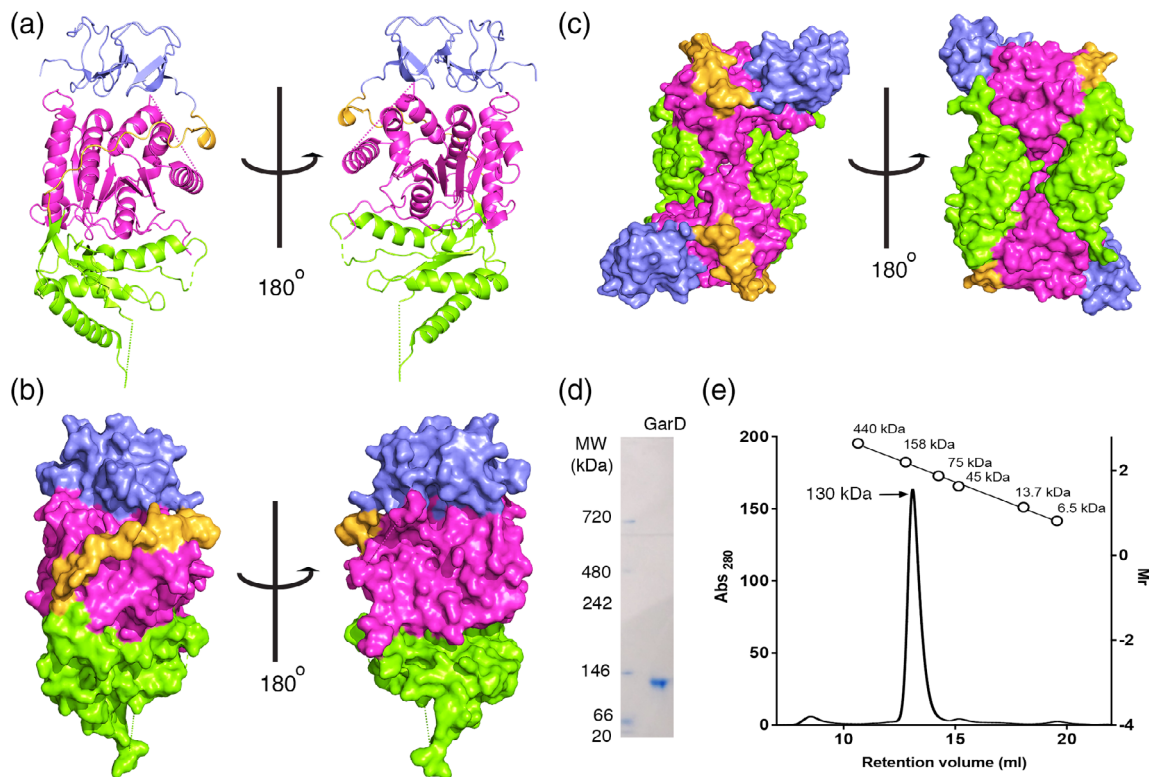


FIGURE 2 The crystal structure and oligomerization state of galactarate dehydratase (GarD). (a) Cartoon representation of GarD monomer: N-terminal domain in blue, linker in orange, second domain in green, and C-terminal domain in magenta. Surface representation of GarD monomer (b) and dimer (c). (d) Clear native gel electrophoresis of GarD; MW, molecular weight. (e) Size exclusion chromatography profile of GarD and calibration curve. The peak corresponds to the elution of the dimer and the estimated molecular weight (130 kDa) is indicated with the arrow. Abs, absorbance, Mr, relative molecular weight

2.2 | The N-terminal SAF domain of GarD forms a β -clip fold

The N-terminal SAF domain of GarD was solved previously by the Midwest Center for Structural Genomics (PDB ID: 3LAZ). The SAF domain is broadly distributed in all kingdoms of life. Structures of several representatives of the SAF domain have been solved, but the GarD structures represent a specific subtype of the SAF domain.¹⁵ In the GarD structure, β 1 and β 2 are antiparallel to each other and followed by a 26-residue long loop (aa 28–35). The β 3 strand runs in the antiparallel direction to β 2 forming the first half of the β -clip. Antiparallel strands β 4 and β 5 pair with β 7 to complete the β -clip fold. Strand β 6 is formed within the 14-residue long loop between β 5 and β 7 and it pairs with β 3 to generate a fourth strand in the open half-barrel of the β -clip (Figure 3b). Contact between the two halves of the fold also occurs via the highly conserved Gly-His-Lys (sHK) tripeptide sequence

motif of the hexuronate dehydratase SAF family. The motif is present in the turn just before β 3 with Lys-54 in the β 3 strand forming hydrogen bonds with Glu-88 between strands β 6 and β 7. From this motif, His-53 is exposed to the surface, where it is adjacent to Trp-85 of the aromatic-hydrophobic-charge (ahc) conserved motif (Trp-Ile-Asp in GarD). A third conserved basic-Tyr-Gly (bYG) tripeptide motif has Arg-68 and Tyr-69 exposed to the surface (Figure 3d). These residues from each of these three motifs form half of a deep channel between the β -clip fold and the core of the protein.

The *E. coli* GarD SAF domain is closely homologous to another example of the SAF β -clip domain, the N-terminal domain of the UxaA altronate dehydratase from *S. flexneri* (PDB ID: 3K3S, RMSD = 1.438) (Figure 3d,e), also solved by the Midwest Center for Structural Genomics. The overlapped structures show the highly conserved barrel-like motif with a β -clip architecture. The first triplet motif in the aligned sequence is the sHK motif

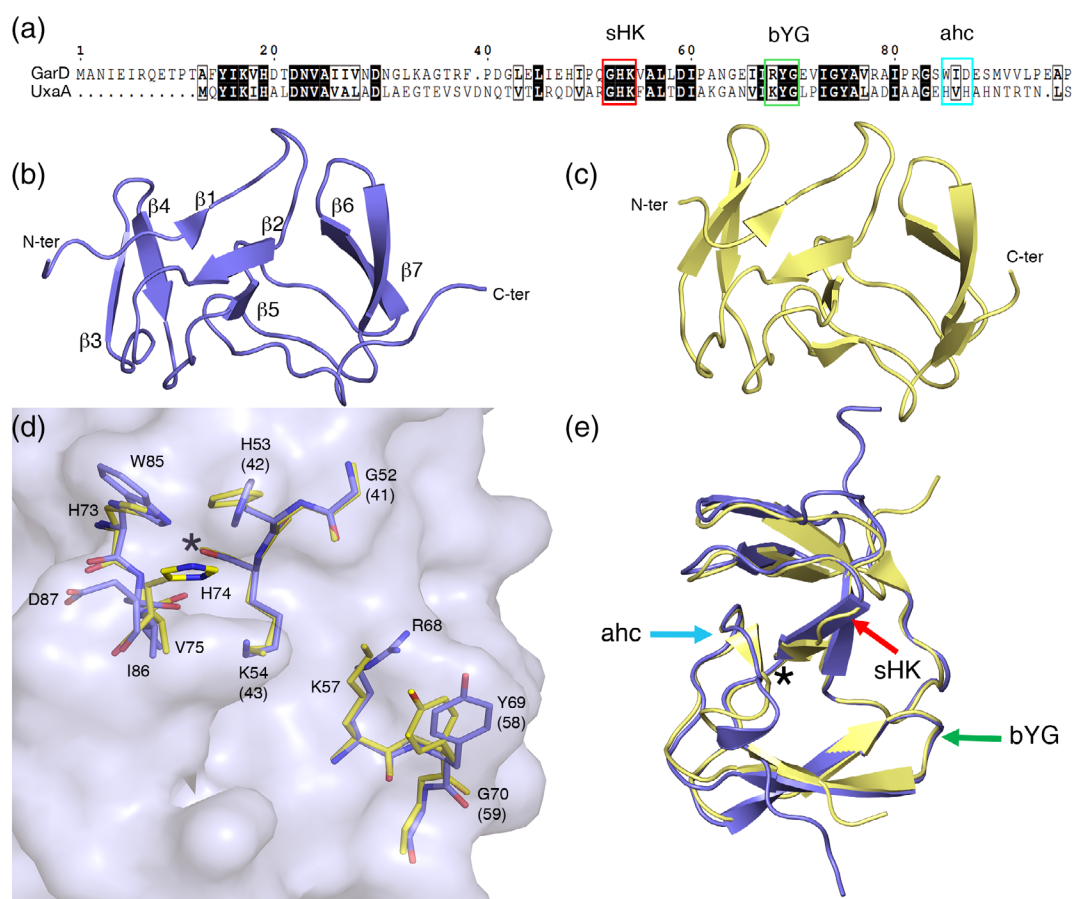


FIGURE 3 The N-terminal domain of GarD. (a) Primary sequence alignment of N-terminal domains of GarD from *E. coli* with UxaA from *S. flexneri*. Three β -clip tripeptide motifs are indicated by the boxes, sHK in red, bYG, green and ahc in cyan. Ribbon diagram representation of β -clips for GarD (b) and UxaA (c). (d) Superposition of tripeptide motifs of GarD and UxaA. Residues are shown as a sticks overlaid on the surface of GarD (translucent blue). Oxygens in red, nitrogens in dark blue, and carbons in blue (GarD) and yellow (UxaA). Residue labeling is based on the sequence of GarD, corresponding UxaA residues are shown in parentheses. The asterisk indicates a proposed metal binding site for UxaA. (e) Ribbon representation of the superposed N-terminal domains of GarD (blue) and UxaA (yellow). Conserved tripeptide motifs are indicated by arrows with colors matching boxes in panel (a)

and the residues and their positions in the structure are identical. Although, in GarD, the side chain of the Lys from this motif forms hydrogen bond interaction with the side chain of Glu-88, while in UxaA this interaction is missing because the position of this glutamate is occupied by an alanine. The second triplet bYG motif from UxaA (Lys-Tyr-Gly) structurally overlaps with the Arg-Tyr-Gly tripeptide from GarD. The ahc motif in the two structures is quite distinct. Unlike UxaA where it is suggested two histidines in this motif could participate in the chelation of metals,¹⁵ in GarD, both histidines are substituted. These differences suggest that the ahc motif in GarD could not be involved in metal binding, however, it is possible that in GarD, this domain participates in substrate recognition, as previously described.¹⁵

2.3 | A long linker connects the β -clip with a second domain

The second domain of GarD is located on the opposite side of the core from the N-terminus, connected by a long linker formed by amino acids 93–118 that is wrapped around the C-terminal core of GarD, opposite the dimer interface (Figure 2c). The linker forms a small α -helix at residues 98–102, but it is mostly unstructured and is tethered to the core domain by the sidechains of valine and leucine residues directed toward the core domain (Figure 2).

The second domain is organized as a three parallel-stranded β -sheet (10–12), wrapped by helices α 2–4 (Figure 4a). Residues 178–185 from the loops between β 10 and α 3 and 32 residues (aa 213–244) from the second domain

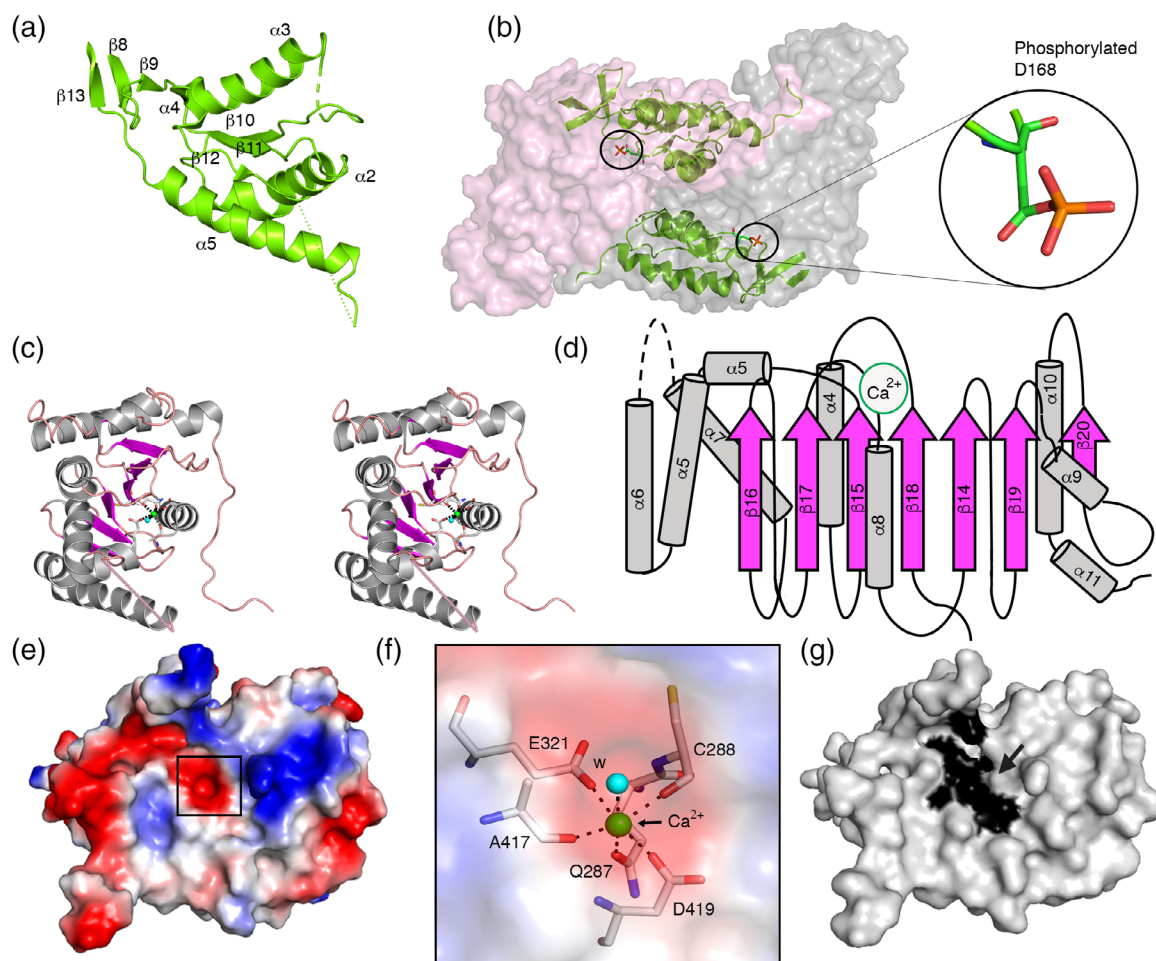


FIGURE 4 The second and C-terminal domains of GarD. (a) Ribbon representation of the dimerization domain in with secondary structure elements labeled. (b) Localization of the second domains (ribbon) and phosphorylated D168 (stick) overlaid on the surface of the GarD dimer. (c). Ribbon wall eye stereo view of the C-terminal domain showing the catalytic core with the β -strands in magenta and α -helices in gray. (d) The structure diagram of the C-terminal domain. The central parallel seven-stranded β -sheet is supported by nine α -helices. Strands and helices are labeled and Ca^{2+} is indicated as a green outlined circle. (e) Vacuum electrostatic surface representation of C-terminal domain. The square indicates the cavity of the Ca^{2+} binding site. (f) The metal binding site. Ca^{2+} (green sphere), water (cyan sphere) and residues involved in metal coordination (in sticks) overlaid on the electrostatic surface. Oxygen, nitrogen and carbon atoms are in red, blue, and gray, respectively. (g) Localization of highly conserved residues on the surface of the C-terminal domain of GarD. GarD, galactarate dehydratase

had poor electron density and are not included in this structure, suggesting some flexibility in this portion of the protein. In the loop between $\alpha 2$ and $\beta 11$, Asp-168 was phosphorylated in two out of four polypeptide chains. This modified side chain of the residue is directed outward from the surface near the interface between the second and core of the C-terminal domain (Figure 4b). The biological significance of this phosphorylation in GarD is unknown since no information is currently available about the regulation of this protein by phosphorylation.

2.4 | The catalytic core of the C-terminal domain has a novel structure, resembling a Rossmann fold

The C-terminal domain is built around a central parallel β -sheet, that starts from the $\beta 14$, coming from the second domain directly to the core of the C-terminal domain. The core is formed by seven parallel β -strands with the 3-4-2-1-5-6 order, surrounded by nine α -helices (Figure 4c,d). Database searches with the coordinates of the C-terminal domain of GarD found a similar arrangement in several proteins, including the C-terminal CMP-Kdo binding domain of WaaA from *Acinetobacter baumannii* (PDB ID: 4BFC), which is classified by the ECOD¹⁹ as a subtype of the Rossmann fold. The arrangement of strands in both structures is identical (Figure 4d) and more distant similarities can be found to multiple structures with the Rossmann-fold. This structure seems to form a catalytic domain, because it has highly conserved residues (Figure 4g) and a metal binding site formed by the $\beta 14$ and $\beta 15$, the loops that connect the $\beta 14$ and $\alpha 6$, as well as the loop between $\alpha 10$ and $\beta 17$ (Figure 4e,f). This arrangement creates a negatively charged cavity in which metal binding was observed (Figure 4e,f). It is known that iron is the preferred metal for enzymatic activity of GarD, however, in this structure we found a metal binding site exposed to the solvent, which was occupied by Ca^{2+} . The calcium ion is coordinated by side chain oxygens of Gln-278, Glu-321, Asp-419, main chain carbonyl oxygens of Cys-228 and Ala-417, and a water molecule (Figure 4f). However, none of these ligands are the preferred for iron coordination.²⁰ Thus, this metal binding site was occupied by calcium because calcium was present in the crystallization solution.

The structure of the C-terminal domain of GarD was surprising as almost all previously described iron dependent dehydratases have a TIM barrel fold. In these structures the metal binding site is exposed to the solvent and is occupied by Mg^{2+} , Mn^{2+} or iron, which facilitates the substrate-metal interaction and catalysis.^{13,21–24} This consistency was the basis for a prediction of GarD as a TIM barrel protein.¹⁵ Interestingly, proteins with GarD activity can be encoded by two

classes on nonhomologous genes in a classic example of a convergent functional evolution or nonhomologous replacement.²⁵ This can be illustrated by a comparison between the structures and sequences of known TIM-barrel fold glucarate/galactarate dehydratases from *Azospirillum brasiliense* (PDB ID: 3FKK), *S. Typhimurium* (PDB ID: 2PP0), *Agrobacterium tumefaciens* (PDB ID: 4YR7), and *Oceanobacillus iheyensis* (PDB ID: 3HPF) with the structure of the Rossmann-like fold of GarD.

2.5 | Activity of GarD

As the GarD structure was revealed to be different from the TIM barrel found in other enzymes with the same activity, we considered that the activity may have been misassigned in the literature. Although the activity of GarD was previously described, this assay has been challenging since both the enzyme and its product are highly sensitive to oxygen. Furthermore, the detection of the product (5-KDG) under UV absorbance has strong interference due to the reactivity of iron with the substrate.²⁶ Thus, we performed the assay in degassed buffer and under reduced conditions, which improved the detection of the product. Here we demonstrated that in absence of iron the enzyme is inactive, but upon iron addition the enzyme produced 4 $\mu\text{mol}/\text{min}$ of 5-keto-4-D-dehydroxyglucarate per mg of protein (Figure 5). Furthermore, we showed that the nonspecific reaction from the substrate and iron is significantly lower than the activity of GarD. These data support that the structure described here corresponds to the iron-dependent GarD previously reported.^{12,13}

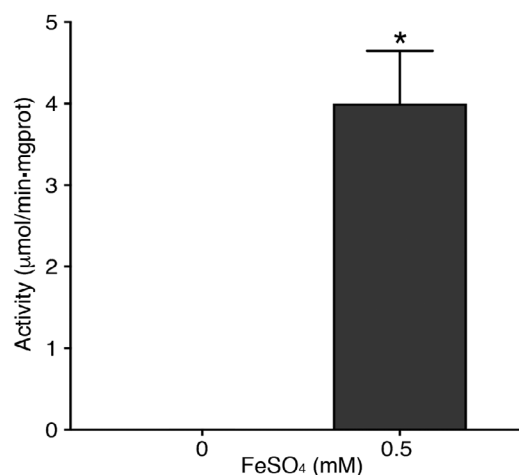


FIGURE 5 Enzymatic activity of GarD. The production of 5-KDG was measured in presence or absence of FeSO_4 as indicated. Results were analyzed by Welch's *t* test from six independent measurements. Asterisks indicate $p < .0001$. GarD, galactarate dehydratase

2.6 | The gene for GarD is highly conserved in diverse antibiotic resistant pathogenic bacteria

The C-terminal D-galactarate dehydratase/altronate hydrolase C terminus domain (Pfam family GD_AH_C

(PF04295)) has broad distribution in bacteria and archaea and even several homologs in protists. A BLAST search of the National Center for Biotechnology Information (NCBI) protein database and UniProt revealed that close homologs of *E. coli* GarD are encoded by the genome of many bacteria, including intestinal pathogens such as

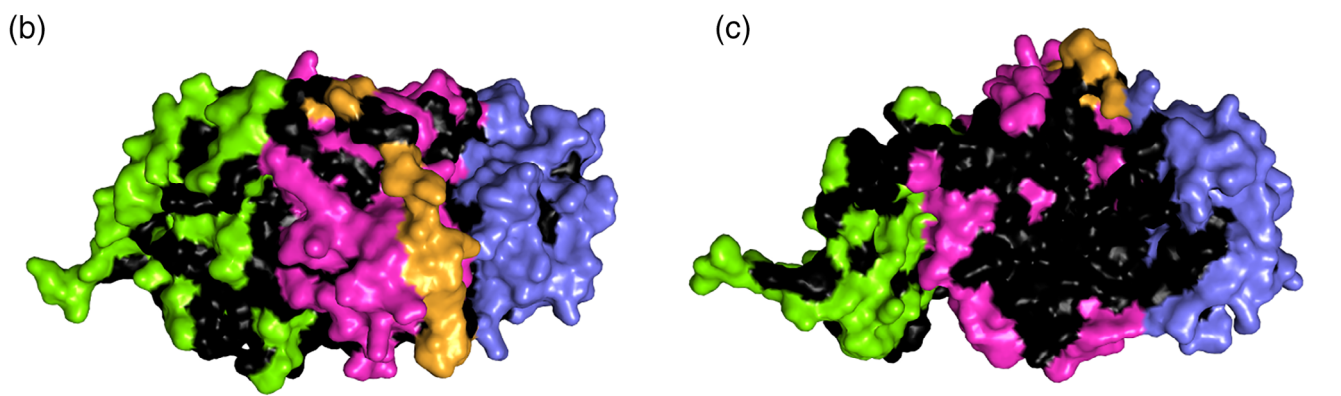
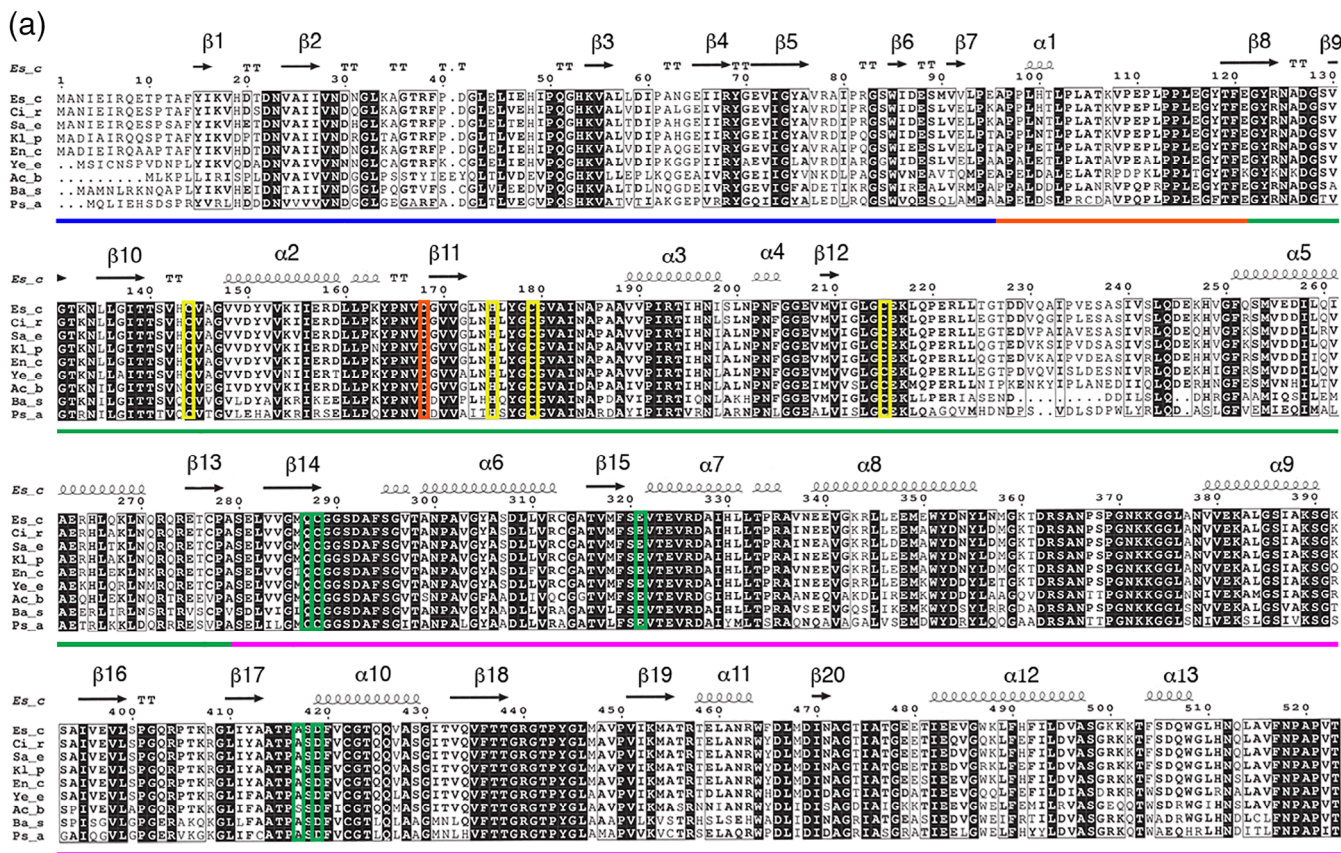


FIGURE 6 Multiple sequence alignment of GarD from different organisms. (a) Aligned GarD sequences from: *E. coli* (*Es_c*), *C. rodentium* (*Cl_r*), *S. Typhimurium* (*Sa_e*), *K. pneumoniae* (*Kl_p*), *E. cloacae* (*En_c*), *Y. enterocolica* (*Ye_e*), *A. baumannii* (*Ac_b*), *B. subtilis* (*Ba_s*), and *P. aeruginosa* (*Ps_a*). Lines below sequences indicate N-terminal β -clip domain (blue), linker (orange), second dimerization domain (green), and C-terminal core domain (magenta). The boxes enclose the conserved and similar amino acids, from the lower similarity in light gray to high similarity bold and 100% identity are white letters shaded in black. The orange box indicates the conserved D168 and yellow boxes indicate the conserved cysteines and histidine predicted for the proposed iron binding site. Green boxes indicate the calcium binding site. Numbering, sequence and the secondary structure elements (α -helices and β -strands) are from *E. coli* GarD. (b, c) Surface representation of the overall structure of GarD. The N-terminal domain, the linker, the second, C-terminal domain and highly conserved residues are shown on blue, yellow, green, magenta, and black, respectively

Gram-negative facultative anaerobes *Citrobacter rodentium* (95% percent identity), *S. Typhimurium* (94%), *Klebsiella pneumoniae* (93%), *Enterobacter cloacae* (93%), and *Yersinia enterocolitica* (84%). Other pathogens with concern for antimicrobial resistance,¹ although not generally associated with the intestine, also have a *garD* gene including *Acinetobacter baumannii* (69%) and *Pseudomonas aeruginosa* (59%). Notably, many of these bacteria are considered critical priority for development of new drugs to treat rising incidence of antimicrobial resistance. The gene was not unique to Gram-negative bacteria, as Gram-positive facultative anaerobe *Bacillus subtilis* (68%) was also found to have a gene that encodes for GarD (Figure 6a). Further, assessment of the location of highly conserved residues revealed strong conservation the three motifs in the β -clip and the catalytic core of the C-terminal domain (Figures 4g and 6b,c).

3 | DISCUSSION

The glucarate/galactarate pathway has been well studied and characterized, although its role in bacterial pathogenesis was only recently proposed as promoting colonization of the gut in antibiotic-treated mice. The availability of glucarate and galactarate as preferred carbon sources in the environment of the inflamed gut could explain the over-expression of the enzymes from this metabolic pathway during envelope stress. Under envelope stress, the BaeSR two component system is often upregulated and some genes controlled in the regulon confer resistance against β -lactams, novobiocin, sodium dodecyl sulfate, and bile salts.²⁷ It is possible that the use of an alternative source of carbon could play an important bioenergetic role and contribute to the proliferation and transmembrane potential energization for the antibiotic efflux.²⁸

Here we present the first structure of GarD, which represents also the first structure from the large altronate/galactarate superfamily of proteins. This structure reveals a β -clip followed by two domains that together form a catalytic core with a new fold. The N-terminal domain seems to be highly conserved, specific for the GarDs, and is likely involved in the substrate recognition. The C-terminus forms a Rossmann-like fold and a solvent-exposed metal binding site cavity that was occupied by Ca^{2+} in our crystallization conditions. Studies of altronate and mannonate hydratases suggest that iron-dependent-hydratases have two metal binding sites, one for iron and one for manganese. Indeed these enzymes are synergistically activated by manganese and iron.²⁹ We did attempt to find an alternative metal binding site in the GarD structure. Three histidines from the sHK and ahc motifs of the *S. flexneri* SAF domain could

have potential to bind metal, but two of the histidines are substituted in GarD and thus not likely to form a metal binding site. The amino acid sequence was also searched using METALDETECTOR V2 software, which predicts possible iron binding sites.³⁰ The software detected four conserved residues (Cys-144, 179 and 215 and His-175) that could bind iron. Three of these residues are located in the second domain, but their positions in the structure of GarD do not cluster and are not arranged to form a metal binding site. It is possible that iron binding could involve residues 181–184 and 211–247 missing in the structure, especially Cys-215, which could bind iron and reduce the flexibility of these loops. Thus, GarD could undergo conformational changes upon iron addition that are not obvious from this structure. Attempts to generate the iron-bound form of GarD were not successful and may require anaerobic conditions. The protein however has iron-dependent GarD activity, as previously reported.^{12,13} Determination of the catalytic mechanism could also benefit from the structure in presence of substrate as part of further studies.

Although the iron binding site is still unknown, the structure of GarD was found to be quite interesting with the novel arrangement of three domains that includes a peculiar Rossmann-like fold with a crossover. This structure of GarD could help for structure prediction of additional proteins in this family. The impact could also extend beyond structure studies. Since GarD is the first enzyme of the metabolism of galactarate and is highly conserved among intestinal pathogens, the inhibition of this pathway could be used for synthetic lethality inhibition to decrease invasiveness of pathogens during dysbiosis caused by antibiotic treatment.

4 | MATERIALS AND METHODS

4.1 | Cloning, protein expression, and purification

The gene *garD* (also known as *b3218* or *yhaG*) was cloned from genomic DNA by ligation independent cloning, using the primers: Fw: 5'-TACTTCCAATCCAATGCGATGGCC AACATCGAAATCAGA-3' and Rv:5'- TTATCCACTTCC AATGTCAGGTACCGGTGCCGG-3' in the pMCSG53 vector,³¹ which encodes ampicillin resistance and a N-terminal 6xHis-tag with a tobacco etch virus (TEV) protease cleavage site. The plasmid was transformed into kanamycin resistant competent cells *E. coli* BL21(DE3)(pMagic).³² The bacteria were cultured in M9 SeMET High-Yield Growth media kit (MD045004, Medicilon) in presence of kanamycin and ampicillin at 37°C and 220 rpm agitation until $\text{OD}_{600} = 0.2$, then placed at 4°C for 18 hr and the growth continued at 37°C, 220 rpm. Protein expression was induced

with 0.5 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG) at OD₆₀₀ of 1.8–2.0 during 14 hr at 25°C and 200 rpm agitation.³³ For structure determination, 0.160 g/L of seleno-methionine (Se-Met) was added right before IPTG induction. For some protein preparations, the protein was produced unlabeled and, 30 min before collection, 1 mM of galactarate-Tris pH 7.5 was added to each culture to prevent oxidation, as reported previously.¹² Cells were collected and washed twice with 50 mM Tris-HCl pH 7.5, 150 mM NaCl, and 1 mM galactarate. Cells were then resuspended in lysis buffer (10 mM Tris, 0.5 M NaCl, 10% glycerol, 0.5 M arginine-HCl, 1 mM galactarate, 0.1% IGEPAL CA-630 (Sigma)). One tablet Roche cOmplete™ EDTA-Free protease inhibitor cocktail was added for each 100 ml of buffer. The pellet was stored at –30°C until purification.

The bacterial suspension was thawed and sonicated using a pulse of 5 s \times 10 s off during 10 min, then the lysate was centrifuged at 39,000g for 40 min, the supernatant was collected, and the protein was purified as previously reported.³⁴ Briefly, the clarified lysate was loaded onto a His-Trap FF (NiNTA) column using a GE Healthcare ÅKTA Purifier system in loading buffer (10 mM Tris-HCl pH 8.5, 500 mM NaCl, 1 mM Tris (2-carboxyethyl)phosphine (TCEP)). The column was washed with loading buffer with 25 mM imidazole and eluted with loading buffer with 500 mM imidazole. The eluant was loaded into a Superdex 200 26/600 column and ran within loading buffer.³⁴ The protein was collected, and the 6xHis-tag was removed with TEV protease, followed by NiNTA-affinity chromatography.

4.2 | Size exclusion chromatography and native gel electrophoresis

Analytical size exclusion chromatography was performed using a Superdex 200 10/30 column and ran with buffer: of 10 mM Tris-HCl, pH 7.5, 150 NaCl, and 1 mM TCEP. The standard calibration curve was obtained using a combined low molecular weight (LMW), and high molecular weight (HMW) Gel filtration Calibration kits (GE Healthcare). The dimers of GarD were detected by native gel electrophoresis. The protein was diluted to 4 mg/ml in loading solution of 50 mM NaCl, 50 mM Imidazole-HCl pH 7 and 0.2 mM aminocaproic acid.³⁵ Samples (1 μ l) and commercial standards (NativeMARK, Life Technologies) were separated in a 10–15% gradient gel using native gel strips (Phast System, GE Healthcare). The gel was fixed in solution (40% Ethanol, 10% acetic acid), stained with Bio-Safe Coomassie G-250 stain (Bio-Rad) and the background was washed out with water.

4.3 | Crystallization, data collection, and structure solution

Purified Se-Met derivative of GarD (7.5 mg/ml in 10 mM Tris-HCl pH 8.3 0.5 M NaCl, 1 mM TCEP) was set up as 2 μ l crystallization drops in (1 μ l protein:1 μ l) reservoir solution in 96-well plates and equilibrated using commercially available Classics II, PACT, PEG's II and ComPas Suites. Diffraction quality crystals grew from, 0.2 M CaCl₂, 0.1 M Tris pH 8.0, 20% PEG6000 (PACT, Qiagen). For cryoprotection, crystals of GarD were transferred into a 5 ml drop of reservoir solution before flash-freezing. The amount of PEG in the reservoir solution was sufficient to serve as a cryoprotectant. Data were collected at selenium absorption peak ($\lambda = 0.97856$) on the LS-CAT 21-ID-G beamline at the Advanced Photon Source (APS) at Argonne National Laboratory. A total of 600 frames were indexed, integrated, and scaled using HKL-3000.³⁶ Data collection and processing statistics are listed in Table 1.

The structure of GarD was solved by the Single Anomalous Dispersion Method (SAD) using a seleno-methionine derivative. Total of 39 out of 44 Se sites were found and 1,851 out of 2,116 residues were automatically built using Automated Structure Solution in Phenix.³⁸ The initial solution underwent several rounds of

TABLE 1 Data collection and processing (Se-Met)

Diffraction source	Beamline 21-ID-G, APS
Wavelength (Å)	0.97856
Temperature (K)	100
Detector	MAR mosaic 300 mm CCD
Space group	P2 ₁
<i>a</i> , <i>b</i> , <i>c</i> (Å)	58.46, 167.57, 117.07
α , β , γ (°)	90.00, 103.97, 90.00
Resolution range (Å)	30.00–2.75 (2.80–2.75)
No. of unique reflections	56,199 (2,851)
Completeness (%)	99.2 (100.0)
Multiplicity	6.1 (6.2)
$\langle I/\sigma(I) \rangle$	16.7 (2.2)
$R_{r.i.m.}^a$	0.044 (0.349)
CC _{1/2} ^b	(0.779)
Overall B factor from Wilson plot (Å ²)	59.2

Notes: Values in parentheses are for the outer shell.

^aEstimated $R_{r.i.m.} = R_{merge}[N/(N-1)]^{1/2}$, where *N* is the data multiplicity.

^bPearson's correlation coefficient.³⁷

refinement in REFMAC v.5.5³⁹ and manual model building and corrections using Coot.⁴⁰ Water molecules were generated using ARP/wARP,⁴¹ Ca²⁺ and Cl⁻ were added to the model manually. Translation–libration–screw (TLS) groups⁴² were created and TLS corrections were applied during the final stages of refinement. MolProbity⁴³ was used to monitor the quality of the model during refinement and for final validation of the structure. The final model and diffraction data were deposited in the PDB (<https://www.rcsb.org/>) with PDB code 6U7L. The final model consisted of four polypeptide chains with the lowest *RMS* deviations for C α atoms between chains A and B (0.34 Å) and highest *RMS* deviations between chains A and C (0.51 Å). Chains A and C have missing nine N-terminal residues of the polypeptide chain and three remaining residues from purification tag, which were disordered and not included in the final model. Chains B and D have missing the tag residues and 13 (chain B) or 11 (chain D) residues at the N-terminal part of the structure. There are three loop regions disordered and not included in the final model. In chain A 181–184, 213–246, and 358–373, in chain B 178–185, 213–244, and 358–376, in chain C 177–185, 213–245, and 358–376, and in chain D 178–184, 215–247, and 358–377. There are 9 Ca²⁺ and 7 Cl⁻ ions and 150 water molecules in the final model. Refinement statistics and the quality of the final model are summarized in Table 2.

4.4 | Sequence and structural alignment

The protein sequence of GarD (aka, GalacD/YhaG/b3128) was analyzed using the HMM profiles of Pfam families on the Pfam server⁴⁴ and also used as template in Smart-Blast (<https://blast.ncbi.nlm.nih.gov/smartblast/smartBlast.cgi>). Other genes were searched manually by organism of interest on the NCBI (<https://www.ncbi.nlm.nih.gov>). The multiple sequence alignment was performed using the locus of the genes: *Citrobacter rodentium* (WP_012908629.1) *E. coli*, *S. Typhimurium* (NP_462163.1), *K. pneumoniae* (WP_016947461.1), *E. cloacae* (WP_063152498.1), *Y. enterocolitica* (WP_083161024.1), *A. baumannii* (WP_000917361.1), *B. subtilis* (WP_076458362.1), and *P. aeruginosa* (WP_029886750.1) in Clustalo (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) and the multi-alignment sequence was merged with the PDB file using ESPript 3.x.⁴⁵ The PDB coordinates of GarD were analyzed on the DALI,¹⁷ CLICK (http://mspc.bii.a-star.edu.sg/minhn/DNA_protein.html) and FATCAT¹⁸ servers to perform structural and sequence alignment. Structural alignments were conducted using Matchmaker in Pymol open source V 2.1.⁴⁶

TABLE 2 Structure refinement (Se-Met)

Resolution range (Å)	29.23–2.75 (2.82–2.75)
Completeness (%)	98.8 (97.3)
No. of reflections, working set	53,337 (3,849)
No. of reflections, test set	2,734 (210)
Final R_{work}	0.207 (0.301)
Final R_{free}	0.241 (0.340)
No. of non-H atoms	
Protein	13,869
Ligand	9 (Ca ²⁺), 7 (Cl ⁻)
Water	150
Total	14,035
<i>RMS</i> deviations	
Bonds (Å)	0.004
Angles (°)	1.198
Average B factors (Å ²)	
Protein	78.0
Ligand	82.0
Water	56.0
Ramachandran plot ^a	
Favored regions (%)	96.0
Additionally allowed (%)	4.0
Outliers (%)	0.0

Notes: Values in parentheses are for the outer shell.

^aCalculated with the program MolProbity.⁴³

4.5 | Enzymatic activity of GarD

The activity of the enzyme was performed with native protein as previously reported,^{12,13,29} with some modifications: The buffer was degassed, and the reaction was performed under semi-anaerobic conditions at 30°C. One mg/ml of protein was diluted in 100 mM Tris–HCl pH 7.0, 90 mM β -mercaptoethanol, 30 mM galactarate, and 0.5 mM FeSO₄ for 10 min. Then the reaction was stopped with 2% formic acid and the precipitate was removed by centrifugation at 14,760g for 5 min at 4°C. The product 5-KDG was detected by the formation of its semicarbazone⁴⁷ by diluting 50 μ l of the reaction to 950 μ l in a solution of 1% semicarbazide hydrochloride in 1.5% of sodium acetate. The reaction was developed for 15 min and the absorbance of the 5-KDG-semicarbazone was detected at 250 nm. As the galactarate may form complexes with FeSO₄,²⁶ which have a strong absorbance at UV wavelength and may cause interference with the 5-KDG-semicarbazone detection, the reaction was

performed in presence and in absence of FeSO₄ and or enzyme as control. In order to eliminate nonspecific interference, the absorbance of the control without enzyme was subtracted to the absorbance in presence of the enzyme, then the concentration of the semicarbazone was calculated using an extinction molar coefficient of 7,500 M⁻¹. The results were plotted as the enzymatic activity in presence or in absence of iron and the statistical analysis was performed with Welch's *t* test in Graph Pad Prism version 8.1.

ACKNOWLEDGMENTS

This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases under Contract Nos. HHSN272201200026C and HHSN272201700060C. Grant Nos. U01AI124316 and R01 GM05789 and from the National Institute of General Medical Sciences grant GM118187, both from National Institutes of Health, Department of Health and Human Services.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

M.R.-L., G.M., L.S., and O.K. designed and conducted experiments and collected data. G.M. and Z.W. solved the structure of GarD. N.M., B.P., and L.J. participated in bioinformatics and design of experiments. M.R.-L., G.M., A.G., and K.S. analyzed data and wrote the manuscript.

DATA AVAILABILITY STATEMENT

Coordinates and structure factors have been deposited in the Protein Data Bank of the Research Collaboratory for Structural Bioinformatics with PDB ID 6U7L.

ORCID

Karla J. F. Satchell  <https://orcid.org/0000-0003-3274-7611>

REFERENCES

- Centers for Disease Control. Antibiotic resistance threats in the United States. Report No. CS298822-B; 2019.
- Marteyn B, West NP, Browning DF, et al. Modulation of *Shigella* virulence in response to available oxygen in vivo. *Nature*. 2010;465:355–358.
- Winter SE, Thiennimitr P, Winter MG, et al. Gut inflammation provides a respiratory electron acceptor for *Salmonella*. *Nature*. 2010;467:426–429.
- Chang D-E, Smalley DJ, Tucker DL, et al. Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proc Natl Acad Sci USA*. 2004;101:7427–7432.
- Karlsson S, Burman LG, Åkerlund T. Induction of toxins in *Clostridium difficile* is associated with dramatic changes of its metabolism. *Microbiology*. 2008;154:3430–3436.
- Pedersen LL, Turco SJ. Galactofuranose metabolism: A potential target for antimicrobial chemotherapy. *Cell Mol Life Sci*. 2003;60:259–266.
- Lazar V, Ditu LM, Pircalabioru GG, et al. Aspects of gut microbiota and immune system interactions in infectious diseases, immunopathology, and cancer. *Front Immunol*. 2018;9:1–18.
- Reeves AE, Theriot CM, Bergin IL, Huffnagle GB, Schloss PD, Young VB. The interplay between microbiome dynamics and pathogen dynamics in a murine model of *Clostridium difficile* infection. *Gut Microbes*. 2011;2:145–158.
- Spiga L, Winter MG, Furtado de Carvalho T, et al. An oxidative central metabolism enables *Salmonella* to utilize microbiota-derived succinate. *Cell Host Microbe*. 2017;22:291–301.
- Bohnhoff M, Miller CP. Enhanced susceptibility to *Salmonella* infection in streptomycin-treated mice. *J Infect Dis*. 1962;111:117–127.
- Faber F, Tran L, Byndloss MX, et al. Host-mediated sugar oxidation promotes post-antibiotic pathogen expansion. *Nature*. 2016;534:697–699.
- Blumenthal HJ, Jepson T. Asymmetric dehydration of galactarate by bacterial galactarate dehydratase. *Biochem Biophys Res Commun*. 1964;17:282–287.
- Hubbard BK, Koch M, Palmer DR, Babbitt PC, Gerlt JA. Evolution of enzymatic activities in the enolase superfamily: Characterization of the (D)-glucarate/galactarate catabolic pathway in *Escherichia coli*. *Biochemistry*. 1998;37:14369–14375.
- LaCroix RA, Sandberg TE, O'Brien EJ, et al. Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of *Escherichia coli* K-12 MG1655 on glucose minimal medium. *Appl Environ Microbiol*. 2015;81:17–30.
- Iyer LM, Aravind L. The emergence of catalytic and structural diversity within the beta-clip fold. *Proteins*. 2004;55:977–991.
- Krissinel E, Henrick K. Detection of protein assemblies in crystals. In: Berthold MR, et al eds, 2005, *CompLife*. Berlin, Heidelberg: Springer-Verlag, pp. 163–174.
- Holm L, Laakso LM. Dali server update. *Nucleic Acids Res*. 2016;44:351–355.
- Ye Y, Godzik A. FATCAT: A web server for flexible structure comparison and structure similarity searching. *Nucleic Acids Res*. 2004;1:582–585.
- Schaeffer RD, Liao Y, Cheng H, Grishin NV. ECOD: New developments in the evolutionary classification of domains. *Nucleic Acids Res*. 2017;4:D296–D302.
- Barber-Zucker S, Shaanan B, Zarivach R. Transition metal binding selectivity in proteins and its correlation with the phylogenomic classification of the cation diffusion facilitator protein family. *Sci Rep*. 2017;7:1–12.
- Goldman AD, Beatty JT, Landweber LF. The TIM barrel architecture facilitated the early evolution of protein-mediated metabolism. *J Mol Evol*. 2016;82:17–26.
- Groninger-Poe FP, Bouvier JT, Vetting MW, et al. Evolution of enzymatic activities in the enolase superfamily: Galactarate dehydratase III from *Agrobacterium tumefaciens* C58. *Biochemistry*. 2014;53:4192–4203.
- Palmer DR, Hubbard BK, Gerlt JA. Evolution of enzymatic activities in the enolase superfamily: Partitioning of reactive

- intermediates by (D)-glucarate dehydratase from *Pseudomonas putida*. *Biochemistry*. 1998;37:14350–14357.
24. Yew WS, Fedorov AA, Fedorov EV, Almo SC, Gerlt JA. Evolution of enzymatic activities in the enolase superfamily: L-talarate/galactarate dehydratase from *Salmonella typhimurium* LT2. *Biochemistry*. 2007;46:9564–9577.
 25. Omelchenko MV, Galperin MY, Wolf YI, Koonin EV. Non-homologous isofunctional enzymes: A systematic analysis of alternative solutions in enzyme evolution. *Biol Direct*. 2010;5:1–20.
 26. Abrahamson HB, Rezvani AB, Brushmiller JG. Photochemical and spectroscopic studies of complexes, of iron(III) with citric acid and other carboxylic acids. *Inorg Chim Acta*. 1994;226:117–127.
 27. Nishino K, Honda T, Yamaguchi A. Genome-wide analyses of *Escherichia coli* gene expression responsive to the BaeSR two-component regulatory system. *J Bacteriol*. 2005;187:1763–1772.
 28. Webber MA, Piddock LJ. The importance of efflux pumps in bacterial antibiotic resistance. *J Antimicrob Chemother*. 2003;51:9–11.
 29. Dreyer JL. The role of iron in the activation of mannonic and altronic acid hydratases, two Fe-requiring hydrolyases. *Eur J Biochem*. 1987;166:623–630.
 30. Passerini A, Lippi M, Frasconi P. MetalDetector v2.0: Predicting the geometry of metal binding sites from protein sequence. *Nucleic Acids Res*. 2011;39:288–292.
 31. Eschenfeldt WH, Makowska-Grzyska M, Stols L, Donnelly MI, Jedrzejczak R, Joachimiak A. New LIC vectors for production of proteins from genes containing rare codons. *J Struct Funct Genomics*. 2013;14:135–144.
 32. Kwon K, Peterson SN. High-throughput cloning for biophysical applications. In: Anderson WF, editor. 2014, *Structural genomics and drug discovery: Methods and protocols*. New York, NY: Springer, pp. 61–74.
 33. Millard CS, Stols L, Quartey P, Kim Y, Dementieva I, Donnelly MI. A less laborious approach to the high-throughput production of recombinant proteins in *Escherichia coli* using 2-liter plastic bottles. *Protein Expr Purif*. 2003;29:311–320.
 34. Shuvalova L. Parallel Protein Purification. In: Anderson W., editor. 2014, *Structural Genomics and Drug Discovery. Methods in Molecular Biology (Methods and Protocols)*. New York, NY: Humana Press, pp. 137–143.
 35. Wittig I, Karas M, Schagger H. High resolution clear native electrophoresis for in-gel functional assays and fluorescence studies of membrane protein complexes. *Mol Cell Proteomics*. 2007;6:1215–1225.
 36. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: The integration of data reduction and structure solution—From diffraction images to an initial model in minutes. *Acta Cryst D*. 2006;62:859–866.
 37. Karplus PA, Diederichs K. Linking crystallographic model and data quality. *Science*. 2012;336:1030–1033.
 38. Adams PDAP, Bunkóczi G, Chen VB, et al. PHENIX: A comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr*. 2010;D66:213–221.
 39. Murshudov GN, Skubak P, Lebedev AA, et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr*. 2011;D67:355–367.
 40. Emsley P, Cowtan K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr*. 2004;D60:2126–2132.
 41. Morris RJ, Perrakis Anastassis, Lamzin VS ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol*. 2003;374:229–244.
 42. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Cryst D*. 2006;62:439–450.
 43. Chen VB, Arendall WB 3rd, Headd JJ, et al. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Cryst D*. 2010;66:12–21.
 44. El-Gebali S, Mistry J, Bateman A, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47:427–432.
 45. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDScript server. *Nucleic Acids Res*. 2014; 42:W320–W324.
 46. DeLano WL. Pymol: An open-source molecular graphics tool CCP4. *Newslett Protein Cryst*. 2002;40:82–92.
 47. Olson JA. Spectrophotometric measurement of α -keto acid semicarbazones. *Arch Biochem Biophys*. 1959;85:225–233.

How to cite this article: Rosas-Lemus M, Minasov G, Shuvalova L, et al. Structure of galactarate dehydratase, a new fold in an enolase involved in bacterial fitness after antibiotic treatment. *Protein Science*. 2020;29:711–722. <https://doi.org/10.1002/pro.3796>