BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Open Access on 06/13/2015

# Evolution of Enzymatic Activities in the Enolase Superfamily: Galactarate Dehydratase III from *Agrobacterium tumefaciens* C58

Fiona P. Groninger-Poe,[†,§] Jason T. Bouvier,[†,§] Matthew W. Vetting,[∥] Chakrapani Kalyanaraman,[⊥] Ritesh Kumar,[§] Steven C. Almo,[∥] Matthew P. Jacobson,[⊥] and John A. Gerlt*[,†,§,‡]

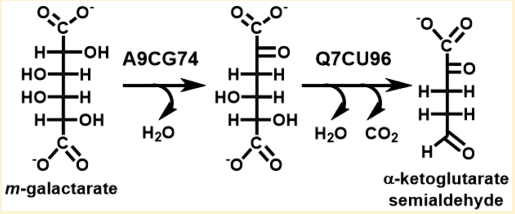
[†]Department of Biochemistry, [‡]Department of Chemistry, and [§]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801, United States

[∥]Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, United States

[⊥]Department of Pharmaceutical Chemistry, School of Pharmacy, and California Institute for Quantitative Biomedical Research, University of California, 1700 4th Street, San Francisco, California 94158, United States

Supporting Information

**ABSTRACT:** The genome of *Agrobacterium tumefaciens* C58 encodes 12 members of the enolase superfamily (ENS), eight of which are members of the mandelate racemase (MR) subgroup and, therefore, likely to be acid sugar dehydratases. Using a library of 77 acid sugars for high-throughput screening, one protein (UniProt entry A9CG74; locus tag Atu4196) showed activity with both *m*-galactarate and D-galacturonate. Two families of galactarate dehydratases had been discovered previously in the ENS, GalrD/TalrD [Yew, W. S., et al. (2007) *Biochemistry 46*, 9564−9577] and GalrD-II [Rakus, J. F., et al. (2009) *Biochemistry 48*, 11546−11558]; these have different active site acid/base catalysis and have no activity with D-galacturonate. A9CG74 dehydrates *m*-galactarate to form 2-keto-3-deoxy-galactarate but does not dehydrate D-galacturonate as expected. Instead, when A9CG74 is incubated with D-galacturonate, 3-deoxy-D-*xylo*-hexarate or 3-deoxy-D-*lyxo*-hexarate is formed. In this reaction, instead of abstracting the C5 proton $\alpha$ to the carboxylate group, the expected reaction for a member of the ENS, the enzyme apparently abstracts the proton $\alpha$ to the aldehyde group to form 3-deoxy-D-*threo*-hexulosuronate that undergoes a 1,2-hydride shift similar to the benzylic acid rearrangement to form the observed product. *A. tumefaciens* C58 does not utilize *m*-galactarate as a carbon source under the conditions tested in this study, although it does utilize D-galacturonate, which is a likely precursor to *m*-galactarate. The gene encoding A9CG74 and several genome proximal genes were upregulated with D-galacturonate as the carbon source. One of these, a member of the dihydrodipicolinate synthase superfamily, catalyzes the dehydration and subsequent decarboxylation of 2-keto-3-deoxy-D-galactarate to $\alpha$-ketoglutarate semialdehyde, thereby providing a pathway for the conversion of *m*-galactarate to $\alpha$-ketoglutarate semialdehyde.




*Agrobacterium tumefaciens* C58 is a plant pathogen responsible for the formation of crown gall tumors and known for its use in agricultural biotechnology. The close physical relationship between *A. tumefaciens* C58 and plants (rich sources of diverse sugars) combined with the fact that its genome encodes five members of the mandelate racemase (MR) subgroup of the enolase superfamily (ENS) with unknown functions and substrate specificities suggests that *A. tumefaciens* C58 is a source of novel sugar catabolic pathways. For example, Gci (UniProt entry A9CEQ8; locus tag Atu3139), a member of the MR subgroup of the ENS, recently was shown to catalyze the novel cycloisomerization of D-galactaro-1,4-lactone ($\gamma$-galactarolactone) to 5-keto-4-deoxy-D-galactarate in a pathway for degradation of D-galacturonate, the major component of pectin found in plant cell walls. This pathway is initiated by a uronate dehydrogenase that oxidizes D-galacturonate to D-galactaro-1,5-lactone ($\delta$-galactarolactone);[1] a lactone isomerase (GLI; UniProt entry A9CEQ7; locus tag Atu3138) then converts $\delta$-galactarolactone to $\gamma$-galactarolactone,[2] the substrate for Gci.

The ENS has >30000 members in addition to the ubiquitous enolases.[3] All members of the ENS share a common overall structure: an N-terminal capping domain that contains the substrate specificity-determining residues and a C-terminal barrel domain that contains the catalytic residues. The active site is located at the interface between the capping and barrel domains and is protected from the solvent by flexible loops from the capping domain. In addition to these structural similarities, members of the ENS share a common mechanistic step, abstraction of a proton $\alpha$ to a carboxylate group by an active site general base. Proton abstraction generates an enolate intermediate that is stabilized by coordination to a conserved magnesium ion.[4−7]

In this work, we describe the assignment of the *in vitro* enzymatic activity to UniProt entry A9CG74 (locus tag Atu4196), a functionally uncharacterized member of the MR subgroup, which is encoded by the *A. tumefaciens* C58 genome. Screening of A9CG74 with a library of acid sugars and a

**Received:** May 6, 2014
**Revised:** June 11, 2014
**Published:** June 13, 2014

semicarbazide assay revealed dehydration of *m*-galactarate as well as an activity with D-galacturonate. Previously reported *m*-galactarate dehydratases from the ENS, GalrD/TalrD and GalrD-II, did not show activity with D-galacturonate,[8,9] nor did previously discovered D-galacturonate dehydratases, GalurDs, show activity with *m*-galactarate (unpublished work). Although details of the reaction mechanism for the reaction with D-galacturonate remain to be established, the structure of the observed product, 3-deoxy-D-*xylo*-hexarate or 3-deoxy-D-*lyxo*-hexarate, suggests an unexpected reaction for the ENS, abstraction of the proton $\alpha$ to the aldehyde group to form 3-deoxy-D-*threo*-hexulosuronate that undergoes a 1,2-hydride transfer similar to the benzylic acid rearrangement to form the observed 3-deoxy-D-*xylo*-hexarate or 3-deoxy-D-*lyxo*-hexarate.

## MATERIALS AND METHODS

**Cloning, Expression, and Protein Purification.** The gene encoding A9CG74 was amplified via polymerase chain reaction (PCR) from genomic DNA isolated from *A. tumefaciens* C58 using Platinum Pfx DNA Polymerase (Invitrogen). The PCR mixture contained 5 $\mu$L of 10× Pfx amplification buffer, 0.3 mM dNTPs, 1 mM $MgSO_4$, primers at 0.3 $\mu$M each (forward primer, 5′-CAT GAG GAA GAC TGA CAT ATG AAA ATC GAT CGC ATG C-3′; reverse primer, 5′-CGA TGA AGC TCG AGT CAG GCG AAG GCA TAA GAA CC-3′), 1 unit of Pfx DNA polymerase, and 50 ng of genomic DNA in a total volume of 50 $\mu$L. The amplification was performed using a PTC-200 gradient cycler (MJ Research) with the following cycling profile: 94 °C for 5 min; followed by 35 cycles of 94 °C for 15 s, 60 °C for 30 s, and 68 °C for 1 min and 30 s; followed by a final extension at 68 °C for 10 min. The amplified gene was cloned into the pET-17b vector using conventional cloning methods. Proteins were expressed by growing 8 L of *Escherichia coli* strain BL21(DE3) cells in LB broth (supplemented with 100 $\mu$g/mL ampicillin) at 37 °C while the sample was shaken at 220 rpm for 24 h. IPTG was not used to induce expression.

The genes encoding A9CG74 orthologs UniProt entry B3Q5L5 (locus tag RHECIAT_PC0000418 from *Rhizobium etli* strain CIAT 652) and UniProt entry B9JNP7 (locus tag Arad_7740 from *Agrobacterium radiobacter* strain K84) were cloned using ligation-independent cloning into a pAVITAG tagless vector. The vector amplification reaction mixture contained 5 $\mu$L of 10× KOD buffer, 0.2 mM dNTPs, 2 mM $MgCl_2$, primers at 0.3 $\mu$M each (forward primer, 5′-AAC CTC TAC TTC CAA TCG CAC CAT CAT CAC CAC CAT TG-3′; reverse primer, 5′-TAT ATC TCC TTC TTA AGG TTA AAC AAA ATT ATT TCT AG-3′), 10 ng of vector template, and 1 unit of KOD polymerase in a total volume of 50 $\mu$L. The vector was amplified with the following cycling profile: 95 °C for 5 min followed by 40 cycles of 95 °C for 30 s, 66 °C for 30 s, and 72 °C for 45 s; the products were purified using a Qiagen PCR purification kit and eluted with 40 $\mu$L of water. The reaction mixture for amplifying the genes contained 5 $\mu$L of 10× KOD buffer, 0.2 mM dNTPs, 2 mM $MgCl_2$, primers at 0.3 $\mu$M each (forward primer, 5′-TTA AGA AGG AGA TAT ACC ATG GTG N-3′, where N denotes 12 complementary nucleotides to the gene; reverse primer, 5′-GAT TGG AAG TAG AGG TTC TCT GCN-3′, where N denotes 12 complementary nucleotides to the gene), 10 ng of vector template, and 1 unit of KOD polymerase in a total volume of 50 $\mu$L. Gene inserts and vectors then were digested in separate reaction mixtures (10 $\mu$L) containing 5 $\mu$L 10× buffer 2, 0.1 $\mu$L of 100× bovine serum albumin, 2.5 mM dNTPs, and 1 unit of T4 DNA polymerase. The digestions were incubated at 22 °C for 60 min followed by 75 °C for 20 min. The digested vector (15 ng) was then combined with 2 $\mu$L of the digested insert and incubated at 22 °C for 15 min. The reaction was stopped by the addition of 10 mM EDTA. The plasmids were transformed into *E. coli* DH10B cells using heat shock. Proteins were expressed by growing 8 L of *E. coli* strain DH10B cells grown in 8 L of LB broth (supplemented with 100 $\mu$g/mL ampicillin) at 37 °C while the sample was being shaken at 220 rpm rpm for 24 h. IPTG was not used to induce expression.

Cells were harvested by centrifugation (4650*g* and 4 °C) and resuspended in 30−40 mL of low-salt buffer [20 mM Tris-HCl (pH 7.9) and 5 mM $MgCl_2$]. Cells were lysed by sonication and pelleted by centrifugation (31000*g* and 4 °C) to remove cell debris. The supernatant was loaded onto a 125 mL Dowex DEAE column equilibrated with 1250 mL of low-salt buffer. The column was washed with 800 mL of low-salt buffer, and the protein was eluted with a linear 1800 mL gradient of 0 to 50% high-salt buffer [1 M NaCl, 20 mM Tris-HCl (pH 7.9), and 5 mM $MgCl_2$] followed by 300 mL of 100% high-salt buffer. The purity was confirmed by sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE). Fractions containing the protein of the appropriate size were pooled and loaded onto a 20 mL Q-Sepharose column equilibrated with 200 mL of low-salt buffer. The column was washed with 100 mL of high-salt buffer, and the protein was eluted with a linear 700 mL gradient of 100 to 0% high-salt buffer. The purity was checked by SDS−PAGE. Fractions containing the protein of the appropriate size were pooled and brought to a final concentration of 1 M $(NH_3)_2SO_4$ before being loaded onto a 50 mL Phenyl Sepharose column. The column was washed with 100 mL of ammonium sulfate buffer [1 M $(NH_3)_2SO_4$, 20 mM Tris-HCl (pH 7.9), and 5 mM $MgCl_2$], and protein was eluted with a 200 mL linear gradient of 100 to 0% ammonium sulfate buffer. The protein purity was checked via SDS−PAGE.

The gene encoding DHDPS Pfam family member Q7CU96 (locus tag Atu4189 from *A. tumefaciens* C58) was cloned into the pMAL-c2x vector containing a maltose binding protein (MBP) tag using Gibson assembly[10] because of incompatible restriction enzyme sites. Separate amplifications of the vector and gene were performed to incorporate regions of homology, followed by incorporation of the gene into the vector using the homologous regions. Primers were designed using a 25 bp overlap with the terminal ends of the gene encoding Q7CU96 and with a 30 bp overlap with the pMAL-c2x vector. The reaction mixture for amplification of the vector (50 $\mu$L) contained 20 ng of pMAL-c2x vector template, vector amplification primers at 0.5 $\mu$M each (forward primer, 5′-TGA AAT CCT TCC CTC GAT CCC GAG GTT GTT G-3′; reverse primer, 5′-GAA TTC GGA TCC TCT AGA GTC GAC CTG CAG GCA AGC-3′), 0.4 mM dNTPs, 10 $\mu$L of 5× HF buffer, 3% DMSO, and 1 unit of Phusion DNA polymerase. The vector reaction was amplified using the following cycling profile: 95 °C for 30 s; followed by three cycles of 95 °C for 10 s, 58 °C for 30 s, and 72 °C for 5 min; followed by three cycles of 95 °C for 10 s, 57 °C for 30 s, and 72 °C for 5 min; followed by 26 cycles of 95 °C for 10 s, 55 °C for 30 s, and 72 °C for 5 min; followed by a final extension at 72 °C for 10 min. The reaction mixture for amplification of the gene (50 $\mu$L) contained 15 ng of Q7CU96 template in pET28a vector, insert amplification primers at 0.5 $\mu$M each (forward primer, 5′-CAA


dx.doi.org/10.1021/bi5005377 | *Biochemistry* 2014, 53, 4192−4203

CAA CCT CGG GAT CGA GGG AAG GAT TTC AAT GAC GAC ATT TGA TAT TCG CCA G-3′; reverse primer, 5′-GCT TGC CTG CAG GTC GAC TCT AGA GGA TCC GAA TTC TTA TTT CCA GCT GGC CAG CAG G-3′). The reaction was amplified as follows: 98 °C for 4 min; followed by 35 cycles of 98 °C for 20 s, 55 °C for 20 s, and 72 °C for 30 s; followed by a final extension at 72 °C for 7 min. After amplification, 25 μL of each PCR product was digested for 16 h at 37 °C with 20 units of DpnI enzyme. Assembly of the full-length DNA construct (20 μL total) was achieved using 35 ng of pMAL vector, 35 ng of the gene of interest, and 15 μL of 1.33× assembly mix. The assembly mix was incubated at 60 °C for 5 min, 4 °C for 5 min, and finally 50 °C for 60 min. The solution was then dialyzed and electroporated into *E. coli* BL21(DE3) cells. Protein was expressed in *E. coli* BL21(DE3) cells grown in LB at 37 °C while the sample was being shaken at 220 rpm. When the $OD_{600}$ reached 0.5, the cells were induced by addition of IPTG to a final concentration of 1 mM. Growth was continued at 20 °C with shaking at 220 rpm for an additional 16 h, at which point the cells were harvested as described above. The cells were sonicated, clarified, and loaded onto a 60 mL amylose column equilibrated with binding buffer [20 mM Tris-HCl (pH 7.9), 5 mM $MgCl_2$, and 0.2 M NaCl]. The column was then washed with 720 mL of binding buffer [20 mM Tris-HCl (pH 7.9), 5 mM $MgCl_2$, and 0.2 M NaCl]. The MBP-tagged protein was eluted with 240 mL of maltose elution buffer [20 mM maltose, 20 mM Tris-HCl (pH 7.9), and 0.2 M NaCl]. The column was then washed with 180 mL of water, 180 mL of 0.1% sodium dodecyl sulfate, and 180 mL of water and finally equilibrated with 360 mL of binding buffer. Fractions containing the MBP-tagged protein were pooled, and Factor Xa was used to cleave the MBP tag in 10 kDa molecular weight cutoff (MWCO) dialysis tubing during dialysis against 20 mM Tris-HCl (pH 7.9), 2 mM $CaCl_2$, and 0.15 M NaCl for 16 h at 4 °C. The protein was then loaded onto a 125 mL DEAE column equilibrated with 20 mM Tris-HCl (pH 7.9) and 5 mM $MgCl_2$. The column was washed with 600 mL of binding buffer followed by a linear gradient over 750 mL to 20 mM Tris-HCl (pH 7.9), 5 mM $MgCl_2$, and 1 M NaCl. Fractions containing the enzyme (32 kDa) were pooled and loaded on the amylose column to remove the MBP tag (42 kDa) using the same method that is described above but instead by collecting the flow-through.

**Screening A9CG74, B3Q5L5, and B9JNP7 with a Library of Acid Sugars.** Proteins were screened for dehydration using a library of 77 mono- and diacid sugars as previously described.[8] Reactions (50 μL) were performed in UV-transparent 96-well plates (Corning) containing 20 mM Tris-HCl (pH 7.9), 5 mM $MgCl_2$, 1 mM substrate, and 1 μM enzyme. The reaction mixtures were incubated at 30 °C for 16 h before the addition of 250 μL of a 1% sodium acetate/1% semicarbazide mixture (semicarbazide solution); the plates were incubated for at least 1 h before the absorbance was recorded at 250 nm ($\varepsilon_{250}$ = 10200 $M^{-1}$ $cm^{-1}$) with a Tecan Plate Reader.

**Kinetic Assay for *m*-Galactarate Dehydratase Activity.** The rate of *m*-galactarate dehydration was determined by performing an end-point semicarbazide assay. For each reaction, 1 μM A9CG74 was incubated with 0.1−10 mM *m*-galactarate in 20 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]-ethanesulfonic acid (HEPES) (pH 7.9) and 5 mM $MgCl_2$. Aliquots were taken at 1, 3, 5, 7, and 9 min and quenched with 5 volumes of a semicarbazide solution. After incubation for 1 h at 22 °C, the absorbance at 250 nm was recorded using a PerkinElmer Lambda14 UV−vis spectrophotometer. Kinetic constants were determined using OriginPro.

**Polarimetric Assay for Determining the Regiospecificity of A9CG74.** The regiochemical preference of A9CG74 for dehydration of *m*-galactarate was determined by combining 5 mM *m*-galactarate, 10 μM enzyme, and 5 mM $MgCl_2$ in 50 mM HEPES (pH 8.0) and monitoring the progress of the reaction using a Jasco P-1010 polarimeter. The change in optical rotation was recorded at 20 °C using a sodium line filter (589 nm), a 10 cm path-length cuvette, and a 1 s integration time for 30 min. The specific optical rotation was calculated for each sample.

**$^1$H NMR Spectra of the *m*-Galactarate Dehydration and the D-Galacturonate Reaction Products.** All $^1$H NMR spectra were recorded with a Varian Unity INOVA 500NB MHz spectrometer unless otherwise noted. Dehydration reactions were performed in $H_2O$ solvent (800 μL total) containing 20 mM potassium phosphate (pH 7.9), 1 mM $MgCl_2$, 2 mM sugar, and 10 μM enzyme. Reaction mixtures in which *m*-galactarate was the substrate were incubated for 16 h at 22 °C, and reaction mixtures in which D-galacturonate was the substrate were incubated for 1 week at 22 °C. The reaction mixtures were lyophilized; the residue was dissolved in 800 μL of $D_2O$ before the $^1$H NMR spectra were recorded. For the reaction with *m*-galactarate, the pD was adjusted to 2 with 5 μL of 1 M DCl.

**Stereochemical Assignment of Solvent Hydrogen Incorporation.** A9CG74 was exchanged into a deuterated buffer by diluting 1 mL of 200 μM enzyme into 9 mL of potassium phosphate buffer at pD 8.1 made using 99.9% $D_2O$ and concentrating the enzyme to 1 mL using a 10 mL Amicon filter with a Millipore 10000 MWCO poly(ether sulfone) ultrafiltration membrane; this procedure was repeated five times. A reaction mixture (800 μL) containing 2 mM substrate, 5 mM $MgCl_2$, and 10 μM enzyme in 20 mM potassium phosphate buffer (pD 8.1) was incubated at 22 °C overnight. For the reaction in which galactarate was the substrate, the pD was adjusted to 2.0 with DCl before the $^1$H NMR spectra were recorded.

**Site-Directed Mutagenesis of A9CG74 and Activity Assays.** Mutants were constructed using the single-overlap extension procedure. The first PCR mixture to generate the megaprimers (50 μL) contained 50 ng of the pET-17b vector containing the gene encoding A9CG74, primers at 0.5 mM each, 10 μL of 5× GC buffer, 3% DMSO, 0.4 mM dNTPs, and 1 unit of Phusion DNA polymerase. The 5′-megaprimer was generated by pairing the T7pro primer with an antisense primer containing the desired mutation (Table S1 of the Supporting Information). The 3′-megaprimer was generated by pairing the T7term primer with a sense primer containing the desired mutation. The PCR cycle was as follows: 98 °C for 4 min; followed by 35 cycles of 98 °C for 20 s, 55 °C for 20 s, and 72 °C for 30 s; followed by a final extension at 72 °C for 7 min. The megaprimers were gel-purified (Qiagen) and used to generate the full-length mutant gene. The PCR mixture (50 μL) for amplifying the full-length mutant gene contained 40 ng of each megaprimer, 10 μL of 5× GC buffer, 3% DMSO, 0.4 mM dNTPs, and 1 unit of Phusion DNA polymerase. An initial PCR program to extend the megaprimers was performed as follows: 98 °C for 4 min followed by five cycles of 98 °C for 20 s, 55 °C for 25 s, and 72 °C for 20 s. Next, T7pro and T7term primers were added to bring the final primer concentration to


dx.doi.org/10.1021/bi5005377 | *Biochemistry* 2014, 53, 4192−4203

0.5 mM, and the PCR program was continued as follows: 35 cycles of 98 °C for 20 s, 55 °C for 20 s, and 72 °C for 25 s followed by a final extension step of 72 °C for 5 min. The full-length mutant genes were then gel-purified and ligated into the pET17b vector. Proteins were expressed and purified as described for the wild-type A9CG74 protein.

The H191N and H292Q mutant proteins were assayed by incubating 1 μM enzyme with 2−10 mM *m*-galactarate using the same method described for the wild-type enzyme. No turnover was detected. These were also assayed for D-galacturonate activity using the $^1$H NMR assay described above, and no reaction was observed.

**Determination of the Activity of Q7CU96 (DHDPS member).** Q7CU96 was tested for activity on the product of the A9CG74 reaction, 2-keto-3-deoxy-D-galactarate (2k3dgalr), and its enantiomer, 5-keto-4-deoxy-D-glucarate (KDG). To synthesize 2k3dgalr, a 2 mL reaction mixture containing 20 μM A9CG74, 40 mM *m*-galactarate, 20 mM potassium phosphate (pH 7.9), and 2 mM $MgCl_2$ was incubated at 22 °C for 16 h. When the reaction was complete, the product was verified by $^1$H NMR spectroscopy. The solution was stored in 0.5 mL aliquots at −20 °C. For the synthesis of KDG, a 2 mL reaction mixture containing 100 μM D-glucarate dehydratase from *E. coli*, 50 mM D-glucarate, 20 mM potassium phosphate buffer (pH 7.9), and 2 mM $MgCl_2$ was incubated at 22 °C for 16 h. The product was verified and stored as described for 2k3dgalr.

Q7CU96 was tested for aldolase activity with both 2k3dgalr and KDG in a coupled enzyme assay containing lactate dehydrogenase (LDH). The reaction mixture (1 mL) contained 10 mM substrate, 100 μM Q7CU96, 2 units of LDH, 0.1 mM NADH, 75 mM potassium phosphate buffer (pH 7.9), and 15 mM $MgSO_4$. The decrease in absorbance at 340 nm was monitored for 10 min.

Q7CU96 was tested for dehydratase/decarboxylase activity with both 2k3dgalr and KDG in a coupled enzyme assay containing α-ketoglutarate semialdehyde dehydrogenase (αKGSDH, UniProt entry Q9I1Q0 from *Pseudomonas aeruginosa*) and diaphorase. In this reaction, Q7CU96 decarboxylates 2k3dgalr to form α-ketoglutarate semialdehyde that is oxidized by αKGSDH with $NAD^+$ to form α-ketoglutarate and NADH. The reduction of iodonitrotetrazolium (INT) by diaphorase using NADH was detected at 500 nm. The reaction mixture (1 mL) contained 2 mM substrate, 1 μM Q7CU96, 0.1 μM αKGSDH, 4 units of diaphorase, 1.6 mM INT, 0.16 mM $NAD^+$, 75 mM HEPES (pH 7.9), and 2 mM $MgSO_4$; 10 μM Q7CU96 was required to detect activity when KDG was used as the substrate. Reaction rates were measured for both substrates using the same assay. For 2k3dgalr, the substrate concentrations varied from 2 to 500 μM and 0.1 μM Q7CU96 was used; for KDG, substrate concentrations varied from 10 μM to 5 mM and 10 μM Q7CU96 was used. Kinetic constants were determined using OriginPro.

A $^1$H NMR spectrum of the reaction mixture (800 μL) was obtained using the following conditions: 2 mM 2k3dgalr, 10 μM Q7CU96, 50 mM potassium phosphate buffer (pH 7.9), and 2 mM MgCl.

**Crystallization and Structure Determination.** B9JNP7, an orthologue of A9CG74, was crystallized by sitting drop vapor diffusion using 96-well Intelliplates (Douglas Instruments). Crystallization drops were assembled by combining 0.5 μL of protein [11 mg/mL in 20 mM Tris-HCl (pH 7.9), 5 mM $MgCl_2$, and 0.1 M NaCl] with 0.5 μL of reservoir buffer. Crystals grew from several conditions, yielding a sulfate-liganded structure [Protein Data Bank (PDB) entry 4JN7] that crystallized in 2 M ammonium sulfate and 100 mM Bis-Tris-HCl (pH 5.5) and an L-malate/$Na^{2+}$-liganded structure (PDB entry 4JN8) that crystallized in 20% PEG3350 and a 0.15 mM mixture of racemic malate. Dodecahedron-shaped crystals grew from both conditions with dimensions of 100 μm on a side and were of the same space group (*I*422), although they had slightly different unit cell dimensions. Single crystals were extracted, transferred to the reservoir condition supplemented with either 20% glycerol (4JN8) or 20% ethylene glycol (4JN7), and vitrified by being plunged in liquid nitrogen. Data were collected at the Advanced Photon Source (beamline 31-ID of Argonne National Laboratory) using a wavelength of 0.9788 Å and a Rayonix 225 HE detector (Rayonix, LLC). Data were integrated in MOSFLM[11] and scaled using SCALA.[12] The automated molecular replacement program BALBES[13] calculated a number of molecular replacement models and output a potential solution utilizing 3RR1 (31% identical sequence) as the search model. The automated building program ARP/WARP[14] was used to fit the majority of the structure, followed by iterative rounds of model building within COOT[15] and refinement within PHENIX.[16] The structures contained a single polypeptide per asymmetric unit with a solvent content of 50%. The identity of the metal ion in the L-malate structure was based on the average distance to the protein ligands and the malate ion, all of which were approximately 2.4 Å, consistent with $Na^{2+}$ instead of $Mg^{2+}$. The quality and stereochemistry of the final structure were verified using MOLPROBITY[17] (refinement statistics listed in Table 1).

**Modeling *m*-Galactarate and D-Galacturonate into the Crystal Structure of PDB Entry 4NJ7.** To examine the potential modes of binding of *m*-galactarate and D-galacturonate, both compounds were docked in the active site of the crystal structure (PDB entry 4JN7) of B9JNP7 liganded with L-malate. PRIME docking, an anchor-and-grow docking method, was used. Briefly, a ligand was built from a core fragment (L-malate), and the conformations were sampled to identify the low-energy pose of the ligand. PRIME docking uses an all-atom OPLS force field with a generalized Born implicit solvent model for all energy evaluations and samples ligand conformations in dihedral angle space.[18] All atoms from L-malate other than the carboxylate group that coordinates to Arg382 and the carbon atom α to the carboxylate group were removed to create the core fragment. Both ligands were built using the Maestro (version 9.3.35) Edit/Build panel. The ligands were subsequently energy minimized and converted into dockable forms using LigPrep (version 4.0515). Parameter files were generated for the ligands using the utility script hetgrp_ffgen (Schrodinger Inc.). PRIME docking then built the ligand from the core fragment in an arbitrary conformation in the active site. The conformations of the ligand were sampled by gradually varying the dihedral angles in 10° increments. The conformations were clustered into 50 different bins, and one conformation from each bin was energy minimized in the active site. The conformation with the lowest energy was selected. The core fragment atoms of each ligand were docked in two different binding modes, i.e., alternate ends of the ligand coordinating to Arg382. After the low-energy poses were identified, the relative binding energy of the ligands corresponding to each pose was evaluated. The relative binding energies corresponding to two alternate poses of *m*-galactarate and D-galacturonate are listed in Table 1.

**Table 1. Data Collection and Refinement Statistics**

| | sulfate (4JN8) | L-malate (4JN7) |
|---|---|---|
| | Data Set Statistics[a] | |
| space group | *I*422 | *I*422 |
| unit cell dimensions (Å) | $a = b = 124.0$, $c = 114.6$ | $a = b = 131.0$, $c = 102.2$ |
| resolution (Å) | 21.2−1.4 (1.48−1.40) | 18.5−1.15 (1.21−1.15) |
| completeness (%) | 100.0 (100.0) | 100.0 (100.0) |
| redundancy | 14.7 (14.4) | 14.1 (14.1) |
| mean(*I*)/sd(*I*) | 14.1 (4.4) | 13.0 (3.9) |
| $R_{sym}$ | 0.136 (0.724) | 0.119 (0.732) |
| | Structure Statistics[a] | |
| resolution (Å) | 21.2−1.4 (1.416−1.40) | 18.5−1.15 (1.163−1.15) |
| no. of unique reflections | 87337 (2291) | 155622 (5136) |
| $R_{cryst}$ (%) | 12.9 (18.1) | 13.6 (20.4) |
| $R_{free}$ (%, 5% of data) | 14.6 (19.4) | 14.5 (21.0) |
| residues in model | A1−395 | A1−395 |
| no. of residues | 395 | 395 |
| no. of waters | 570 | 522 |
| total no. of atoms | 3810 | 3813 |
| average *B* factor (Å$^2$) | | |
| protein | 8.25 | 10.5 |
| waters | 22.9 | 23.5 |
| malate | − | 12.4 |
| $Na^{2+}$ | − | 7.7 |
| root-mean-square deviation | | |
| bond lengths (Å) | 0.008 | 0.008 |
| angles (deg) | 1.30 | 1.30 |
| | MOLPROBITY Statistics | |
| Ramachandran plot (%) | | |
| favored | 97.3 | 97.0 |
| outliers | 0.0 | 0.0 |
| Clashscore[b] | 2.5 (98th percentile) | 2.0 (98th percentile) |
| overall score[b] | 1.17 (97th percentile) | 1.16 (95th percentile) |

[a]Statistics in parentheses are for the highest-resolution bin. [b]Scores are ranked according to structures of similar resolution as formulated in MOLPROBITY.

**RT-qPCR of RNA Isolated from *A. tumefaciens* C58.** A single colony of *A. tumefaciens* C58 from a plated culture was used to inoculate 5 mL of *Agrobacterium* minimal medium (ABM) [3 g/L $K_2HPO_4$, 1 g/L $NaH_2PO_4$, 1 g/L $NH_4Cl$, 0.3 g/L $MgSO_4{\cdot}7H_2O$, 0.15 g/L KCl, 0.01 g/L $CaCl_2{\cdot}2H_2O$, and 0.0025 g/L $FeSO_4{\cdot}7H_2O$ (pH 7.0)] containing 0.4% D-glucose. After the $OD_{600}$ reached 1, the cells were washed five times as follows. The culture was dispensed into 1.5 mL Eppendorf tubes and pelleted in a table-top Eppendorf centrifuge at 8000 rpm for 1 min; the cell pellet was then resuspended in 0.5 mL of ABM lacking a carbon source. After five washes, the cells were resuspended in 0.5 mL of ABM lacking a carbon source and used to inoculate ABM containing either 0.4% D-glucose or D-galacturonate as the sole carbon source. The cell density of each culture was then brought to an $OD_{600}$ of 0.1 by dilution into *Agrobacterium* minimal medium. Cells were harvested when the $OD_{600}$ reached 0.2 by adding an equal volume of RNAlater solution (Qiagen), incubating at 22 °C for 5 min before pelleting, and then removing the supernatant. The cell pellets were frozen at −20 °C until all pellets were harvested. Total RNA was isolated from the cells using an RNeasy kit (Qiagen).

cDNA was reverse-transcribed using the ProtoScript M-MuLV First Strand cDNA Synthesis Kit (NEB) as described by the manufacturer. Briefly, a mixture of 300 ng of RNA and 2 μL of 50 mM random hexamers was brought to a final volume of 8 μL and incubated for 10 min at 65 °C to denature the nucleic acid; this was performed for RNA harvested from cells grown on glucose and cells grown on D-galacturonate. Ten microliters of 2× ProtoScript II Reaction Mix and 2 μL of 20X ProtoScript II Enzyme mix were added to each reaction mixture, and the mixtures were incubated at 42 °C for 1 h followed by 85 °C for 5 min. A negative control containing water instead of reverse transcriptase was generated using the same procedure.

RT-qPCR was performed using a LightCycler 480 (Roche). Reaction mixtures (10 μL) contained primers at 0.5 μM each, 15 ng of cDNA, and 5 μL of 2× SYBR Green I Master Mix (Roche). Reaction mixtures were cycled with an initial hold at 95 °C for 5 min followed by 45 cycles of 95 °C for 10 s, 50 °C for 10 s, and 72 °C for 10 s during which SYBR green fluorescence was monitored. $\Delta C_p$ (change in crossing point) and $\Delta\Delta C_p$ were calculated for each carbon source. The fold change compared to the glucose control was determined using the formula $2^{-\Delta\Delta C_p}$.

**RNA-Seq of RNA Isolated from *A. tumefaciens* C58.** A single colony of *A. tumefaciens* C58 was used to inoculate 5 mL of 0.4% D-glucose-supplemented *Agrobacterium* minimal medium and grown to an $OD_{600}$ of 0.5. Cells were washed twice to remove the carbon source and then inoculated into 0.4% D-galacturonate or 0.4% D-glucose minimal medium and grown for 2 h. RNA was then harvested as described above. Samples were submitted to the Roy J. Carver Biotechnology Center for library preparation, data collection, and analysis.

## RESULTS AND DISCUSSION

Eight of the 12 ENS members encoded by the *A. tumefaciens* C58 genome are members of the MR subgroup. A9CG74 has a DxH motif at the end of β-strand 3 similar to the catalytic residues in D-galactonate dehydratases (Figure 1) but does not catalyze the dehydration of D-galactonate; D-galactonate dehydratases do not catalyze the dehydration of *m*-galactarate or D-galacturonate.[7] A9CG74's *in vitro* activity was determined using biochemical assays, and its biological function was established by transcriptomics.

**Bioinformatics of A9CG74 and Its Homologues.** In a sequence similarity network, A9CG74 clusters with ∼25 other proteins at a BLASTP *e*-value cutoff of $10^{-85}$ (Figure 2). The sequences of members of this cluster are >60% identical, and they are from either *Agrobacterium* or *Rhizobium*, each of which is a member of the Rhizobiaceae family. An alignment of the sequences in this cluster shows that both the active site catalytic residues and substrate binding residues in the capping domain are conserved (Figure 1). The enzymes in this cluster likely are orthologues based on sequence identity and shared catalytic/substrate specificity-determining residues.

**RNA-Seq and RT-qPCR on Neighboring Genes.** The A9CG74 genome neighborhood encodes eight genes that are conserved in the genome neighborhoods of other members of the cluster: Q9CU96 (locus tag Atu4189), a member of the dihydrodipicolinate synthase (DHDPS) superfamily; Q7CU97 (locus tag Atu4190), a potential lactonase; Q7CU98 (locus tag

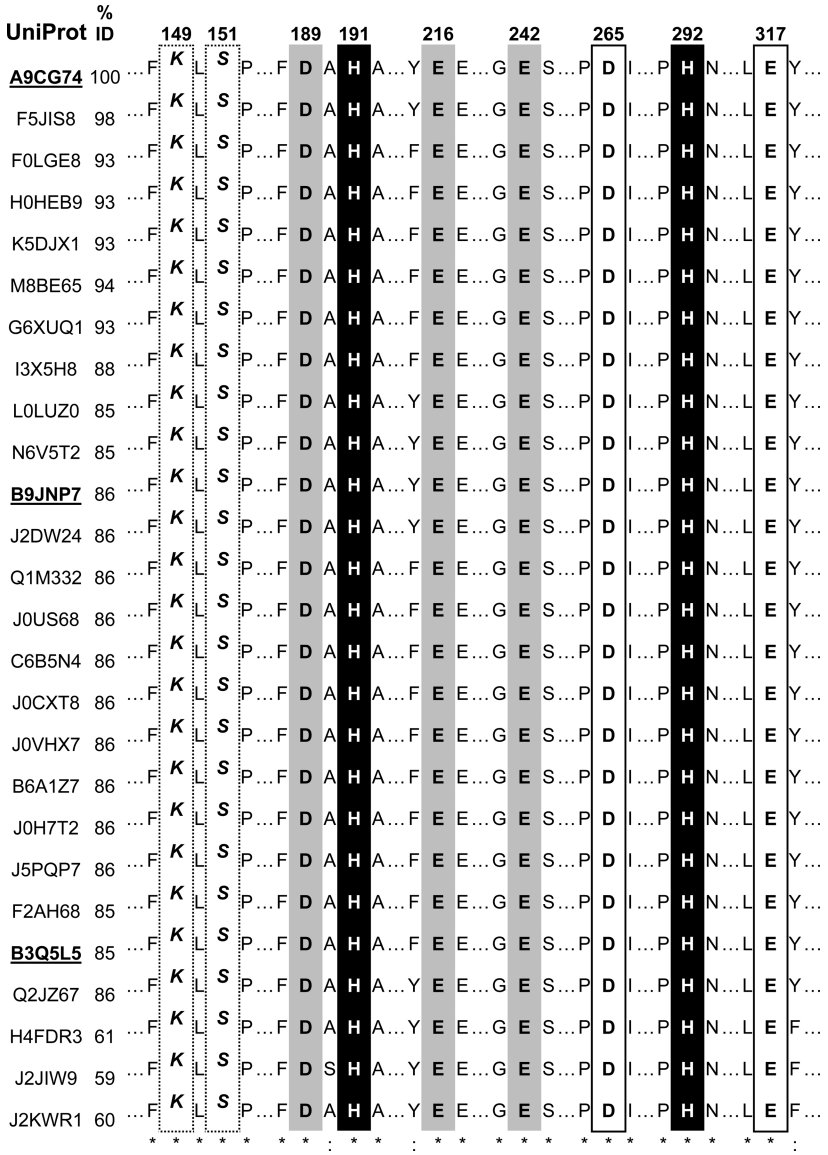



**Figure 1.** Partial sequence alignment of A9CG74 and close homologues sharing at least 60% identical sequences. Catalytic residues are highlighted in black and metal binding residues in gray, and electrophilic catalysts are outlined in solid black. The KxS motif at $\beta$-strand 2 is shown in italics. The numbering is based on A9CG74's sequence. Asterisks denote 100% conserved residues; colons denote chemically similar residues.

Atu4191), GntR regulator; Q7CU99 (locus tag Atu4192), Q7CUA0 (locus tag Atu4193), A9CG73 (locus tag Atu4194), and Q7CUA2 (locus tag Atu4195), ABC transporters; and A9CG75 (locus tag Atu4197), mutarotase (Figure 3). The species of *Agrobacterium* in this cluster share most of these genes, except that a LysM gene, which encodes a cell wall lytic protein, replaces the mutarotase in the *A. radiobacter* K84 genome. Additional differences are present in the Rhizobia members, e.g., the absence of a mutarotase and the presence of an additional ENS member (Figure 3, gene A). An additional $NAD^+$-dependent dehydrogenase is located adjacent to the GntR regulator (Figure 3, gene B). Rhizobia also contain an additional dipeptide transport system and conserved genes, including a flavoprotein and a nitroreductase. Because these genes are not conserved, they likely are not necessary for the metabolism of *m*-galactarate.

We sought to determine which of the neighboring genes are upregulated when *A. tumefaciens* C58 is grown on D-galacturonate as a carbon source. Transcriptomic analysis confirmed that A9CG74 and its surrounding genes were upregulated (Figure 4); however, the genome encodes at least two additional pathways for D-galacturonate degradation, so it is possible that the genes in the genome neighborhood of A9CG74 are part of a much larger set of coregulated genes involved in D-galacturonate metabolism, of which those involved in *m*-galactarate degradation are a subset.

Transcript analysis was performed to determine which genes are cotranscribed. The genes encoding Q7CU96 (locus tag Atu4189, a member of the DHDHPS superfamily) and the potential lactonase Q7CU97 (locus tag Atu4190) were found to be on the same transcript (Figure 5, lane 3); furthermore, the genes encoding the mutarotase A9CG75 (locus tag Atu4197) and GalrD-III A9CG74 (locus tag Atu4196) were found to be on the same transcript (Figure 5, lane 4). Although the mutarotase could not utilize *m*-galactarate, D-galacturonate is a substrate (data not shown); the mutarotase could be involved in the conversion of D-galacturonate to galactarate by catalyzing interconversion of the hemiacetals of D-galacturo-


dx.doi.org/10.1021/bi5005377 | *Biochemistry* 2014, 53, 4192−4203

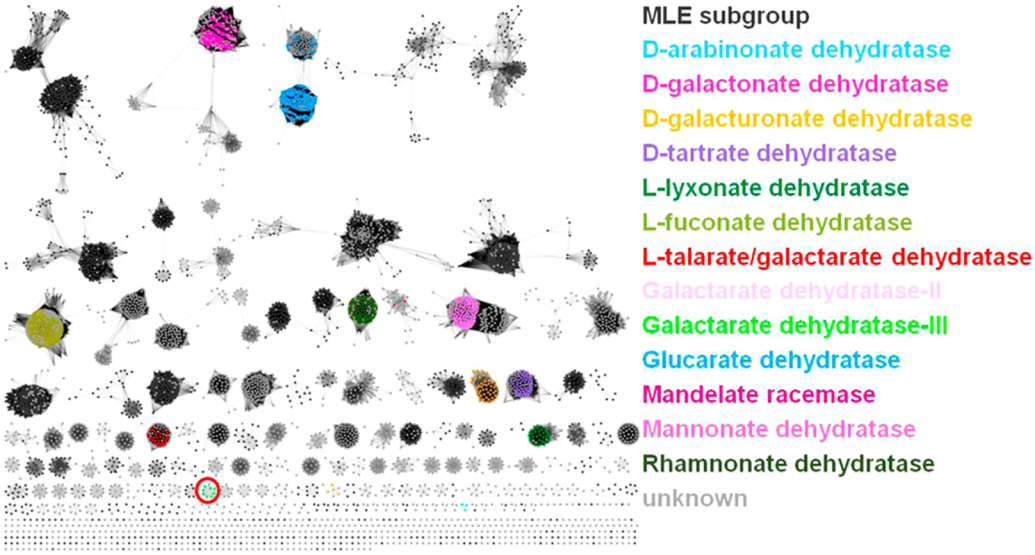



**Figure 2.** Sequence similarity representative node network (95% identical sequences) for the ENS superfamily except for the enolase subgroup clustered at a BLASTP *e*-value of $10^{-85}$. The cluster containing A9CG74 is denoted with a red circle, and its sequence is >60% identical with those of other members of this cluster.

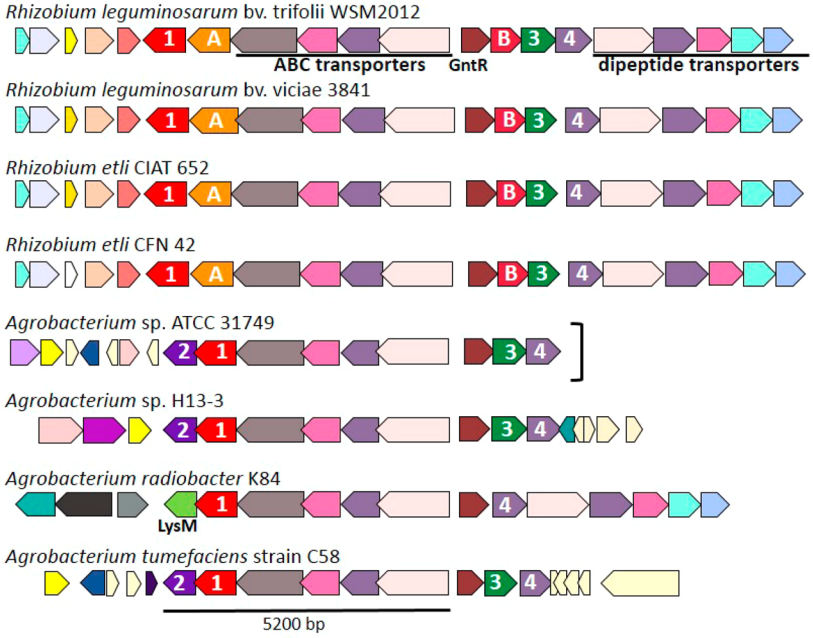



**Figure 3.** Genome neighborhoods of A9CG74 and its orthologues. Annotations: (1) galactarate dehydratase III, (2) mutarotase, (3) a potential lactonase, (4) dehydratase/decarboxylase (DHDPS), (A) nonorthologous MR subgroup members of unknown function, and (B) $NAD^+$-dependent dehydrogenase. The bracket denotes a partially sequenced genome.

nate, one of which likely is a substrate for a uronate dehydrogenase.[19,20] Additional experiments would be needed to determine the physiological role of the mutarotase encoded by A9CG75.

RT-qPCR and transcriptional analysis allowed us to establish the neighboring genes that are upregulated and likely involved in *m*-galactarate metabolism, focusing the possible genes involved in a metabolic pathway in *Agrobacterium* to A9CG74, Q7CU97, and Q7CU96. Both A9CG74 and Q7CU96 were successfully purified and functionally characterized. Unfortunately, Q7CU97 was insoluble.

**A9CG74 (GalrD-III) and Its Orthologues Are *m*-Galactarate Dehydratases.** Screening A9CG74 with a library of 77 acid sugars revealed dehydration of *m*-galactarate as well as a semicarbazide "hit" for D-galacturonate (*vide infra*). This laboratory previously reported two orthologous families of galactarate dehydratases in the enolase superfamily, GalrD/TalrD and GalrD-II; however, the active site residues in those enzymes differ from those found in A9CG74.[8,9] Kinetic constants for A9CG74 with *m*-galactarate were determined with an end point assay ($k_{cat}$ = 0.12 ± 0.05 $s^{-1}$; $K_m$ = 80 ± 30 $\mu$M; $k_{cat}/K_m$ = 1.5 × $10^3$ $M^{-1}$ $s^{-1}$) and are similar to those previously reported for GalrD/TalrD and GalrD-II.[8,9] B3Q5L5


dx.doi.org/10.1021/bi5005377 | *Biochemistry* 2014, 53, 4192−4203

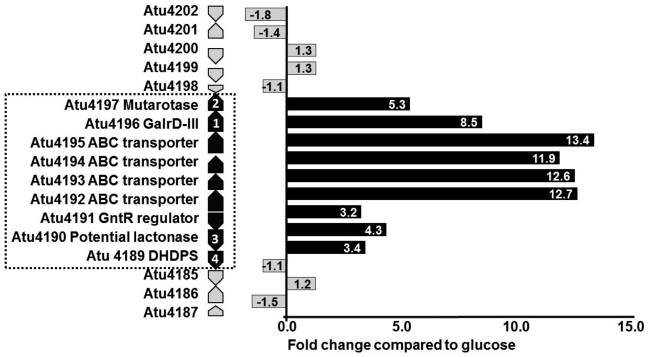



**Figure 4.** Upregulation of the A9CG74 genome neighborhood genes when *A. tumefaciens* C58 is grown on D-galacturonate compared to D-glucose. Gene annotations: (1) galactarate dehydratase III, (2) mutarotase, (3) a potential lactonase, and (4) dehydratase/decarboxylase (DHDPS).

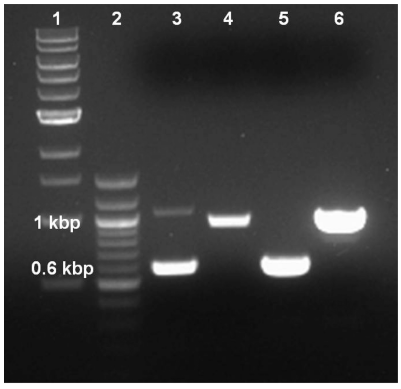



**Figure 5.** Transcript analysis of the amplified product from cDNA constructed starting from cells grown on D-galacturonate vs *Agrobacterium* gDNA. Amplification from cDNA shows adjacent genes A9CG74 and A9CG75 are cotranscribed as are adjacent genes Q7CU96 and Q7CU97: lane 1, NEB 1 kb ladder; lane 2, NEB 100 bp ladder; lane 3, amplicon from primers spanning the intergenic region between A9CG74 and A9CG75 from cDNA (expected size of 550 bp); lane 4, amplicon from primers spanning the intergenic region between Q7CU96 and Q7CU97 (expected size of 950 bp); lane 5, same as lane 3 but from gDNA (expected size of 550 bp); lane 6, same as lane 4 but from gDNA (expected size of 950 bp).

and B9JNP7, orthologues of A9GC74, also showed similar screening results (Table 2). Thus, we conclude that A9CG74, B9JNP7, and B5Q5L5 are orthologous galactarate dehydratases with catalytic residues unlike those previously described for galactarate dehydratases in the ENS.

**X-ray Structure.** An X-ray structure of GalrD-III from *A. radiobacter* K84 (UniProt entry B9JNP7, with L-malate and $Na^+$; PDB entry 4JN7) was determined that showed an $(\alpha+\beta)$ capping domain and a $(\beta/\alpha)_7$ $\beta$-barrel domain characteristic of enolase superfamily members. The $\alpha+\beta$ capping domain contains residues 1−115 at the N-terminus and residues 322−395 at C-terminus of the polypeptide; the barrel domain contains residues 116−321.

Residues located in 20s and 50s loops in the capping domain interact with the substrate, determining substrate specificity. In GalrD-III, both the 20s loop and 50s loops are short, containing residues 14−16 and 35−39, respectively; Arg 16 from the 20s loop forms H-bonds with the nearby carboxylate group of the substrate, and Y36 from the 50s loop forms H-bonds with both the carboxylate and the hydroxyl oxygen on the adjacent carbon. The C-terminus containing residues 375−395 extends near the active site, allowing Arg 382 to form H-bonds with the proximal carboxylate group of the substrate (Figure 6A).

**Table 2. Kinetic Constants of Previously Characterized Galactarate Dehydratases (GalrD/TalrD and GalrD-II) and GalrD-III Characterized Herein**

| enzyme | $k_{cat}$ ($s^{-1}$) | $K_m$ ($\mu$M) | $k_{cat}/K_m$ ($M^{-1}$ $s^{-1}$) |
|---|---|---|---|
| GalrD/TalrD | 3.5 | 320 | $1.1 \times 10^4$ |
| GalrD-II | 6.8 | 620 | $1.1 \times 10^4$ |
| A9CG74 (GalrD-III) | 0.12 ± 0.05 | 80 ± 30 | $1.5 \times 10^3$ |
| B9JNP7 | 0.95 ± 0.05 | 250 ± 50 | $3.8 \times 10^3$ |
| B3Q5L5 | 1.2 ± 0.05 | 360 ± 100 | $3.3 \times 10^3$ |

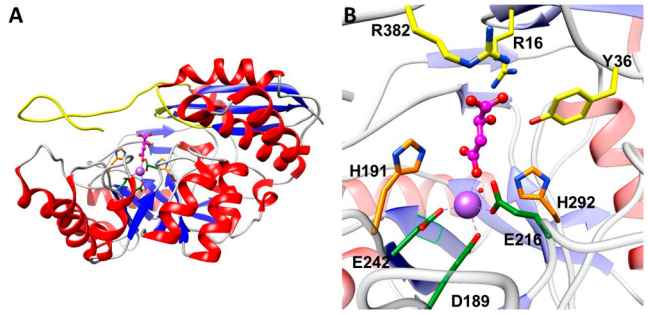



**Figure 6.** Structure of GalrD-III (PDB entry 4JN7) liganded with L-malate (magenta) and $Na^+$ (purple). (A) Overall structure. Loops containing residues for substrate specificity are colored yellow. (B) Active site view with L-malate colored magenta, substrate specificity residues (R382, R16, and Y36) colored yellow, catalytic residues (H191 and H292) colored orange, and metal binding residues (D189, E216, and E242) colored green.

The ligands that coordinate the essential $Mg^{2+}$ ion are Asp 189, Glu 216, and Glu 242 located at the C-terminal ends of $\beta$-strands 3−5, respectively (Figure 6B). As expected for a member of the MR subgroup, His 292, located at the C-terminal end of $\beta$-strand 7, forms a hydrogen-bonded dyad with Asp 265, located at the C-terminal end of $\beta$-strand 6; together, these residues act as the general basic catalyst that initiates the reaction.

Unlike other MR members that contain two Lys residues at the C-terminal end of $\beta$-strand 2, GalrD-III contains a KxS motif consisting of Lys 149 and Ser 151 at $\beta$-strand 2 that would not be catalytic. Instead, GalrD-III shows a catalytic DxH motif at the C-terminal end of $\beta$-strand 3 as observed in D-galactonate dehydratase,[7] containing Asp 189 that coordinates the essential $Mg^{2+}$ and His 191 that would be the general acid catalyst for dehydration (Figure 6B).

**Regiochemistry of Galactarate Dehydration.** As a diacid, *m*-galactarate has two stereochemically distinct $\alpha$-protons available for abstraction. Thus, two enantiomeric "2-keto-3-deoxy galactarate" products are possible. Polarimetry was used to determine the configuration of the A9CG74 product by comparing its optical rotation to those reported for GalrD/TalrD and GalrD-II that abstract the $\alpha$ proton from different "ends" of *m*-galactarate.[8,9] The specific optical rotations ($[\alpha]_{589}^{20}$) for the products of the GalrD/TalrD- and GalrD-II-catalyzed reactions are +4° and −4°, respectively.[8,9] The specific optical rotations for the products produced by A9CG74, B9JNP7, and B5Q5L5 are −4°; i.e., the orthologues


dx.doi.org/10.1021/bi5005377 | *Biochemistry* 2014, 53, 4192−4203

produce 2-keto-D-*threo*-4,5-dihydroxyadipate as observed for the GalrD-II-catalyzed reaction.

**Stereochemistry of Galactarate Dehydration.** A sequence alignment of members of the cluster containing A9CG74 revealed a conserved KxS motif at the end of $\beta$-strand 2, a conserved DxH motif at the end of $\beta$-strand 3, and the His-Asp dyad at the ends of $\beta$-strands 7 and 6, respectively. These catalytic residues are similar to those in D-galactonate dehydratase (Figure 1). To determine whether His residues in the DxH motif and His-Asp dyad are important, H191N ($\beta$-strand 3) and H292Q ($\beta$-strand 7) were characterized. Neither mutant was active with either *m*-galactarate or D-galacturonate, confirming their importance.

The stereochemical course of dehydration of *m*-galactarate catalyzed by A9CG74 was determined by comparing the $^1$H NMR spectra of the products obtained in $D_2O$ and $H_2O$; solvent-derived deuterium is incorporated into the 3-pro-*S* position (Figure 7). Using the model with *m*-galactarate in the active site (Figure 8A), we propose the mechanism shown in Figure 9. On the basis of its proximity to the $\alpha$-proton, His 292 abstracts the proton from the $\alpha$-carbon. His 191, which forms an H-bond with the departing hydroxide at C3, facilitates the departure of the hydroxide leaving group. This model shows that Lys 149 from the KxS motif at the C-terminal end of $\beta$-strand 2 is not involved in catalysis but rather is involved in coordinating the substrate. Lys 149 forms an H-bond with one carboxylate of the *m*-galactarate substrate; similarly, when D-galacturonate is modeled into the active site, Lys 149 forms an H-bond with the aldehyde.

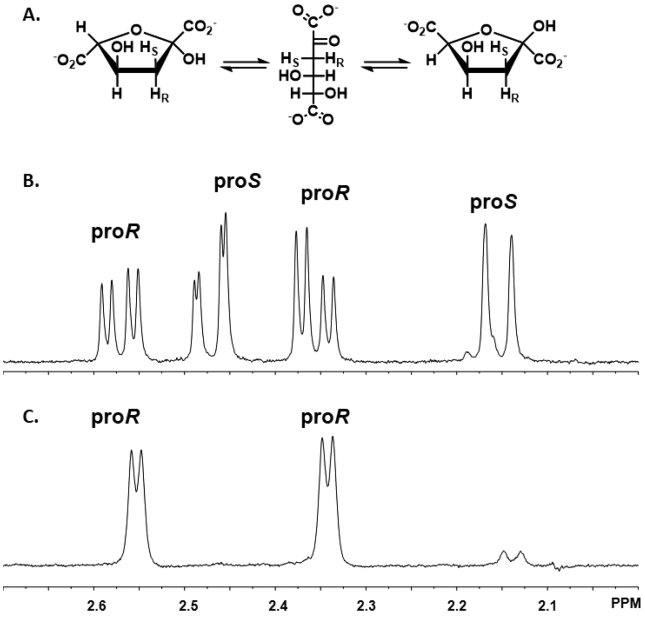



**Figure 7.** Representative $^1$H NMR spectra of the 2-keto-3-deoxy-galactarate product of the reaction of A9CG74 with *m*-galactarate. (A) Hemiketal and linear forms of the 2-keto-3-deoxy-galactarate product. (B) Reaction performed in $H_2O$. Resonances corresponding to the pro-*R* and pro-*S* protons of the hemiketal forms of the 2-keto-3-deoxy-galactarate product are labeled. (C) Reaction performed in $D_2O$. Only the resonances corresponding to the pro-*R* proton are visible.

**A9CG74 Catalyzes a Novel Dehydration Reaction with D-Galacturonate.** The acid sugar library screening of A9CG74 revealed low-level activity with D-galacturonate. However, the $^1$H NMR spectrum of the product formed after incubation for several days was not that expected for dehydration initiated by abstraction of the proton $\alpha$ to the carboxylate group (Figure 10). Instead, the product was identified as either 3-deoxy-D-*xylo*-hexarate or 3-deoxy-D-*lyxo*-hexarate on the basis of the $^1$H−$^{13}$C HSQC and $^{13}$C NMR spectra (Figures S1 and S2 of the Supporting Information). Because neither of these products will react with semicarbazide, we hypothesized that an intermediate that would react with semicarbazide (such as a dehydration product) was formed and later consumed. When B9JNP7 and B5Q5L5, other members of this cluster, were incubated with D-galacturonate, the same product was formed.

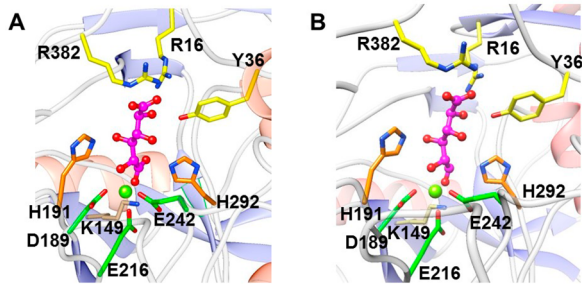



**Figure 8.** Structure of PDB entry 4JN7 with *m*-galactarate or D-galacturonate modeled into the active site. (A) PDB entry 4NJ7 with *m*-galactarate (magenta) modeled into the active site in the lowest-energy configuration. Residues involved in catalysis (His 191 and His 292) are colored orange, metal binding residues (Asp 189, Glu 216, and Glu 242) dark green, and residues involved in substrate stabilization in the active site (Arg 382, Arg 16, Tyr 326, and Tyr 36) yellow. Lys 149 that forms an H-bond with the substrate is colored tan and the essential $Mg^{2+}$ light green. His 292 is positioned to abstract the proton located $\alpha$ to the carboxylate. (B) PDB entry 4NJ7 with D-galacturonate (magenta) modeled in the active site in the lowest-energy configuration. The aldehyde functional group instead of the carboxylate of D-galacturonate is coordinated to the $Mg^{2+}$ ion. His 292 is positioned to abstract the proton located $\alpha$ to the aldehyde group. Residue colors are the same as in panel A.

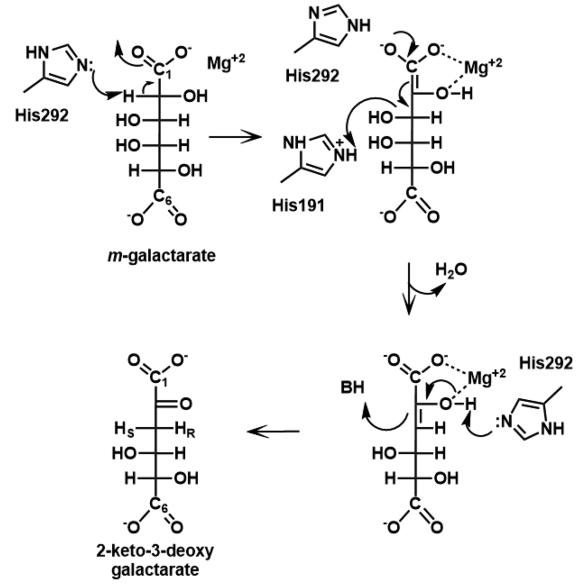



**Figure 9.** Proposed mechanism for the dehydration of *m*-galactarate by GalrD-III. Proton abstraction is performed by His 292 followed by general acid-catalyzed dehydration by His 191. H292 then facilitates tautomerization to the final product, 2-keto-3-deoxy-galactarate.


dx.doi.org/10.1021/bi5005377 | *Biochemistry* 2014, 53, 4192−4203

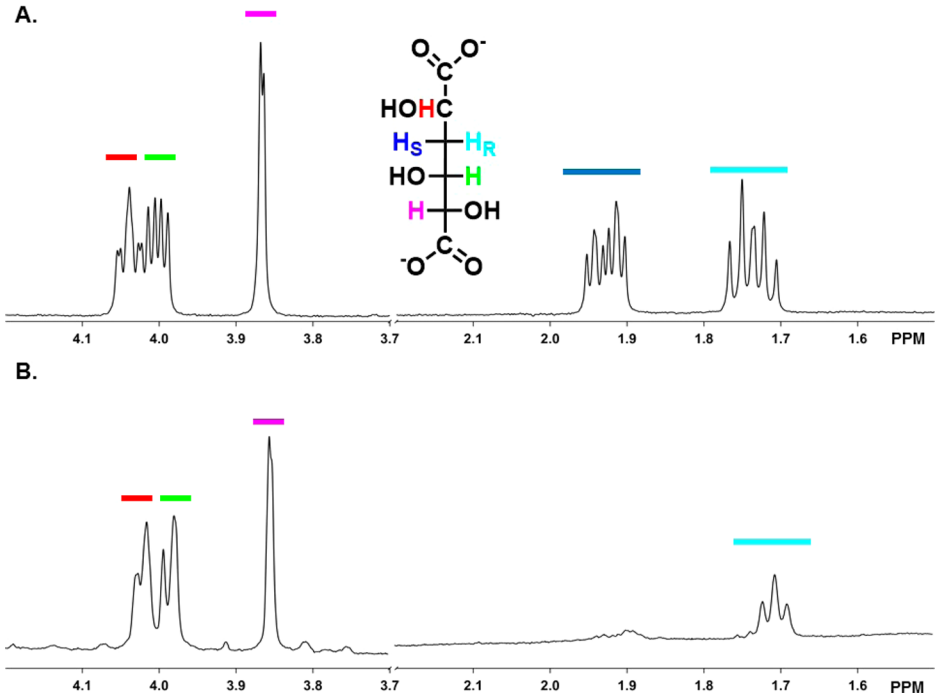



**Figure 10.** Representative $^{1}$H NMR spectra of the product of the reaction of A9CG74 with D-galacturonate as a substrate. (A) Spectrum of the product when D-galacturonate is incubated with GalrD-III for 1 week in $H_2O$ at pH 7.9. (B) Spectrum of the product when D-galacturonate is incubated with GalrD-III for 1 week in $D_2O$ at pH 7.9.

We rationalize the formation of this unexpected product as follows. Instead of catalysis of dehydration by abstraction of the proton from the carbon $\alpha$ to the carboxylate group, the proton $\alpha$ to the aldehyde group is abstracted, resulting in a semicarbazide active product; then, the dehydration product undergoes a reaction similar to the benzilic acid rearrangement (a 1,2-hydride transfer) in which the aldehyde group is oxidized to a carboxylate group and the adjacent carbonyl group is reduced to an alcohol (Figure 11).

The ability of A9CG74 and orthologues to catalyze this reaction reveals the ability of members of the ENS to catalyze an unexpected reaction, although we propose that the active site stabilizes the enolate anion obtained by abstraction of the proton $\alpha$ to the aldehyde group by bidentate coordination to the active site $Mg^{2+}$ as well-established for reactions catalyzed by MR and previously characterized acid sugar dehydratases in the MR subgroup (including dehydration of *m*-galactarate by A9CG74). On the basis of the lowest-energy modeled structure of D-galacturonate in the active site of B9JNP7 (PDB entry 4NJ7), the aldehyde oxygen instead of one of the carboxylate oxygens coordinates to the $Mg^{2+}$ (Figure 8B). On the basis of this model, we propose that His 292 catalyzes abstraction of the proton from the carbon $\alpha$ to the aldehyde group. Dehydration would then occur by His 191 functioning as the general acid that facilitates departure of the hydroxide leaving group, yielding compound **2** (Figure 11). Compound **2** then undergoes a 1,2-hydride transfer similar to the benzylic acid rearrangement to produce 3-deoxy-D-*xylo*-hexarate or 3-deoxy-D-*lyxo*-hexarate (Figure 11, compound **3**).[21] The mechanism of this reaction was not investigated.

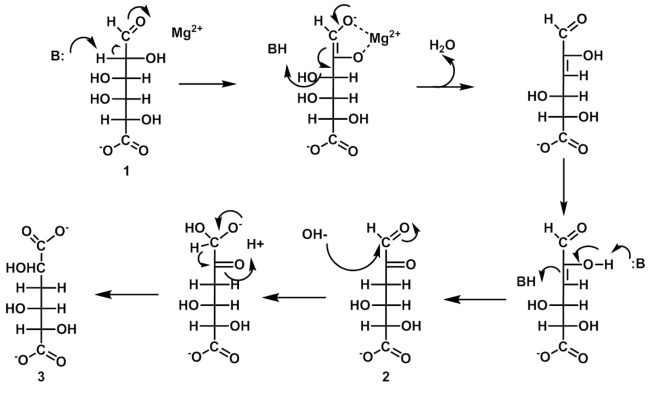



**Figure 11.** Proposed mechanism for the reaction of GalrD-III using D-galacturonate. Rather than abstraction of the proton $\alpha$ to the carboxylate from D-galacturonate (**1**), the proton $\alpha$ to the aldehyde is abstracted and dehydration occurs to yield 3-deoxy-D-*lyxo*-hexulosuronate (**2**). A rearrangement similar to the benzylic acid rearrangement (1,2-hydride shift) yields 3-deoxy-D-*xylo*-hexarate or 3-deoxy-D-*lyxo*-hexarate (**3**) (uncertain stereochemistry at C2).

**Q7CU96 Is a Dehydratase/Decarboxylase That Acts on 2-Keto-3-deoxy-galactarate.** Both RNA-Seq and RT-qPCR experiments confirmed that the gene encoding Q7CU96, a member of the dihydrodipicolinate synthase (DHDPS) superfamily and proximal to the gene encoding A9CG74, is upregulated when *A. tumefaciens* C58 is grown on D-galacturonate (a possible metabolic precursor to *m*-galactarate) as a carbon source (Figure 4). The upregulation and proximity of these genes suggest that Q7CU96 may participate downstream of A9CG74 in a metabolic pathway for *m*-galactarate degradation. Q7CU96 was tested for activity on the product of the A9CG74 reaction, 2-keto-3-deoxy-galactarate, as well as its enantiomer, 5-keto-4-deoxy-D-glucarate.[22,23] Q7CU96 is similar in sequence to other DHDPS members known to be aldolases, suggesting that it is the aldolase that cleaves 2-keto-3-deoxy-galactarate to pyruvate and tartronate semialdehyde. This activity was investigated using a lactate


dx.doi.org/10.1021/bi5005377 | *Biochemistry* 2014, 53, 4192−4203

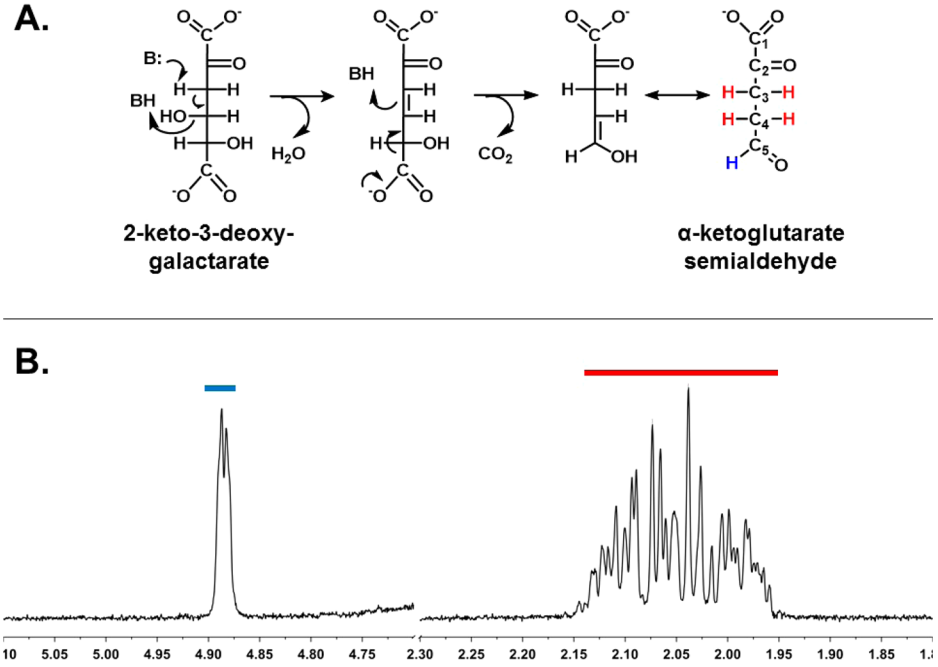



**Figure 12.** Reaction of Q7CU96 with 2-keto-3-deoxy-galactarate as the substrate. (A) Reaction scheme of dehydration and decarboxylation by Q7CU96 to form $\alpha$-ketoglutarate semialdehyde, which is cyclic in solution. (B) Partial $^1$H NMR spectrum of the product of the reaction of Q7CU96 with 2-keto-3-deoxy-galactarate as the substrate. The peak at 4.87 ppm corresponds to the proton on C5; the multiplet at 1.95−2.15 ppm corresponds to the remaining protons on C3 and C4.

dehydrogenase coupled assay that would reduce pyruvate to lactate using NADH; however, no activity was observed.

Next, the ability of Q7CU96 to catalyze dehydration of 2-keto-3-deoxy-galactarate followed by vinylogous decarboxylation was tested by monitoring the formation of $\alpha$-ketoglutarate semialdehyde by coupling the reaction with $\alpha$-ketoglutarate semialdehyde dehydrogenase ($\alpha$KGSDH), an enzyme that oxidizes $\alpha$-ketoglutarate semialdehyde to $\alpha$-ketoglutarate using $NAD^+$ (Figure 12A). Activity was detected, and the kinetic constants were determined: $k_{cat}$ = 1.6 ± 0.1 $s^{-1}$, $K_m$ = 27 ± 9 $\mu$M, and $k_{cat}/K_m$ = 5.9 × $10^4$ $M^{-1}$ $s^{-1}$. The formation of $\alpha$-ketoglutarate semialdehyde was confirmed using $^1$H NMR spectroscopy (Figure 12B). Thus, together, A9CG74 and Q7CU96 transform *m*-galactarate to $\alpha$-ketoglutarate semialdehyde that would be oxidized to $\alpha$-ketoglutarate.

## CONCLUSIONS

From both *in vitro* enzymatic activity and *in vivo* transcriptomics, we propose that the dehydration of *m*-galactarate is the physiological function of A9CG74 and that the product of this reaction, 2-keto-3-deoxy-galactarate, is converted to $\alpha$-ketoglutarate semialdehyde and then $\alpha$-ketoglutarate. We propose the name GalrD-III to indicate both the physiological role and its mechanistic uniqueness from previously reported galactarate dehydratases that do not contain the same active site residues.[8,9] The unexpected, unprecedented reaction observed with D-galacturonate, abstraction of the proton $\alpha$ to an aldehyde group, suggests that aldoses could be substrates for uncharacterized members of the ENS.

## ASSOCIATED CONTENT

### Supporting Information

Sequences of the mutagenic primers and the NMR spectra supporting the identification of 3-deoxy-D-*xylo*-hexarate or 3-deoxy-D-*lyxo*-hexarate as the product of GalrD-III on D-galacturonate. This material is available free of charge via the Internet at http://pubs.acs.org.

### Accession Codes

The atomic coordinates and structure factors for UniProt entry B9JNP7 from *A. radiobacter* strain K84 bound with L-malate and sulfate described herein have been deposited as Protein Data Bank entries 4JN7 and 4JN8, respectively. This work describes characterization of *in vitro* enzymatic activities of UniProt entries A9CG74, B3Q5L5, B9JNP7, and Q7CU96.

## AUTHOR INFORMATION

### Corresponding Author

*Department of Biochemistry, University of Illinois, 1206 W. Gregory Dr., Urbana, IL 61801. E-mail: j-gerlt@illinois.edu.

### Funding

This research was supported by a program project grant and two cooperative agreements from the National Institutes of Health (P01GM071790, U54GM074945, and U54GM094662). Molecular graphics were created and analyses performed with the UCSF Chimera package; Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by National Institute of General Medical Sciences Grant P41-GM103311). Use of the Advanced Photon Source, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the DOE via Contract DE-AC02-06CH11357. Use of the Lilly Research Laboratories Collaborative Access Team (LRL-CAT) beamline at Sector 31 of the Advanced Photon Source was provided by Eli Lilly Co., which operates the facility.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We acknowledge Dr. Alvaro Hernandez in the Roy J. Carver Biotechnology Center for performing the RNA-seq analyses and Drs. Jenny Drnevich and Christopher Fields of High-Performance Biological Computing for performing the bioinformatic analyses. We thank Salehe Ghasempur for providing purified D-glucarate dehydratase for the synthesis of KDG and the αKGSDH used in the coupled assay. M.P.J. is a consultant to Schrodinger LLC.

## ■ ABBREVIATIONS

MR, mandelate racemase; ENS, enolase superfamily; GalrD/TalrD, L-talarate/*m*-galactarate dehydratase; GalrD-II, *m*-galactarate dehydratase II; GalurD, D-galacturonate dehydratase; DHDPS, dihydrodipicolinate synthase; 2k3dgalr, 2-keto-3-deoxy-D-galactarate; KDG, 5-keto-4-deoxy-D-glucarate; αKGSDH, α-ketoglutarate semialdehyde dehydrogenase.

## ■ REFERENCES

(1) Boer, H., Maaheimo, H., Koivula, A., Penttila, M., and Richard, P. (2010) Identification in *Agrobacterium tumefaciens* of the D-galacturonic acid dehydrogenase gene. *Appl. Microbiol. Biotechnol. 86*, 901−909.

(2) Bouvier, J. T., Groninger-Poe, F. P., Vetting, M., Almo, S. C., and Gerlt, J. A. (2014) Galactaro δ-Lactone Isomerase: Lactone Isomerization by a Member of the Amidohydrolase Superfamily. *Biochemistry 53*, 614−616.

(3) Akiva, E., Brown, S., Almonacid, D. E., Barber, A. E., II, Custer, A. F., Hicks, M. A., Huang, C. C., Lauck, F., Mashiyama, S. T., Meng, E. C., Mischel, D., Morris, J. H., Ojha, S., Schnoes, A. M., Stryke, D., Yunes, J. M., Ferrin, T. E., Holliday, G. L., and Babbitt, P. C. (2014) The Structure-Function Linkage Database. *Nucleic Acids Res. 42*, D521−D530.

(4) Gerlt, J. A., Babbitt, P. C., Jacobson, M. P., and Almo, S. C. (2012) Divergent Evolution in Enolase Superfamily: Strategies for Assigning Functions. *J. Biol. Chem. 287*, 29−34.

(5) Gerlt, J. A. (2006) Discovering new functions in the enolase superfamily. *Abstracts of Papers*, p 884, 232nd National Meeting of the American Chemical Society, San Francisco, CA, Sept 10−14, 2006, American Chemical Society, Washington, DC.

(6) Gerlt, J. A., and Babbitt, P. C. (1997) The enolase superfamily: Different reactions catalyzed by similar active sites. *FASEB J. 11*, A1007.

(7) Babbitt, P. C., Hasson, M. S., Wedekind, J. E., Palmer, D. R. J., Barrett, W. C., Reed, G. H., Rayment, I., Ringe, D., Kenyon, G. L., and Gerlt, J. A. (1996) The enolase superfamily: A general strategy for enzyme-catalyzed abstraction of the α-protons of carboxylic acids. *Biochemistry 35*, 16489−16501.

(8) Rakus, J. F., Kalyanaraman, C., Fedorov, A. A., Fedorov, E. V., Mills-Groninger, F. P., Toro, R., Bonanno, J., Bain, K., Sauder, J. M., Burley, S. K., Almo, S. C., Jacobson, M. P., and Gerlt, J. A. (2009) Computation-facilitated assignment of the function in the enolase superfamily: A regiochemically distinct galactarate dehydratase from *Oceanobacillus iheyensis*. *Biochemistry 48*, 11546−11558.

(9) Yew, W. S., Fedorov, A. A., Fedorov, E. V., Almo, S. C., and Gerlt, J. A. (2007) Evolution of enzymatic activities in the enolase superfamily: L-Talarate/galactarate dehydratase from *Salmonella typhimurium* LT2. *Biochemistry 46*, 9564−9577.

(10) Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., III, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods 6*, 343−345.

(11) Leslie, A. G. (2006) The integration of macromolecular diffraction data. *Acta Crystallogr. D62*, 48−57.

(12) Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D62*, 72−82.

(13) Long, F., Vagin, A. A., Young, P., and Murshudov, G. N. (2008) BALBES: A molecular-replacement pipeline. *Acta Crystallogr. 64*, 125−132.

(14) Morris, R. J., Perrakis, A., and Lamzin, V. S. (2003) ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol. 374*, 229−244.

(15) Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126−2132.

(16) Adams, P. D., Gopal, K., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Pai, R. K., Read, R. J., Romo, T. D., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., and Terwilliger, T. C. (2004) Recent developments in the PHENIX software for automated crystallographic structure determination. *J. Synchrotron Radiat. 11*, 53−55.

(17) Davis, I. W., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2004) MOLPROBITY: Structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res. 32*, W615−W619.

(18) Rapp, C., Kalyanaraman, C., Schiffmiller, A., Schoenbrun, E. L., and Jacobson, M. P. (2011) A molecular mechanics approach to modeling protein-ligand interactions: Relative binding affinities in congeneric series. *J. Chem. Inf. Model. 51*, 2082−2089.

(19) Chang, Y. F., and Feingold, D. S. (1970) D-Glucaric acid and galactaric acid catabolism by *Agrobacterium tumefaciens*. *J. Bacteriol. 102*, 85−96.

(20) Chang, Y. F., and Feingold, D. S. (1969) Hexuronic acid dehydrogenase of *Agrobacterium tumefaciens*. *J. Bacteriol. 99*, 667−673.

(21) Niemelä, K., and Sjöström, E. (1985) Non-oxidative and oxidative degradation of D-galacturonic acid with alkali. *Carbohydr. Res. 144*, 93−99.

(22) Hubbard, B. K., Koch, M., Palmer, D. R. J., Babbitt, P. C., and Gerlt, J. A. (1998) Evolution of enzymatic activities in the enolase superfamily: Characterization of the D-glucarate/galactarate catabolic pathway in *Escherichia coli*. *Biochemistry 37*, 14369−14375.

(23) Gulick, A. M., Palmer, D. R. J., Babbitt, P. C., Gerlt, J. A., and Rayment, I. (1998) Evolution of enzymatic activities in the enolase superfamily: Crystal structure of D-glucarate dehydratase from *Pseudomonas putida*. *Biochemistry 37*, 14358−14368.