Evolution of Enzymatic Activities in the Enolase Superfamily: L-Talarate/Galactarate Dehydratase from *Salmonella typhimurium* LT2[†,‡]

Wen Shan Yew,[§] Alexander A. Fedorov,[⊥] Elena V. Fedorov,[⊥] Steven C. Almo,[*,⊥] and John A. Gerlt[*,§,#]

Departments of Biochemistry and Chemistry, University of Illinois at Urbana−Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801, and Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461

Received May 10, 2007; Revised Manuscript Received June 4, 2007

ABSTRACT: We assigned L-talarate dehydratase (TalrD) and galactarate dehydratase (GalrD) functions to a group of orthologous proteins in the mechanistically diverse enolase superfamily, focusing our characterization on the protein encoded by the *Salmonella typhimurium* LT2 genome (GI:16766982; STM3697). Like the homologous mandelate racemase, L-fuconate dehydratase, and D-tartrate dehydratase, the active site of TalrD/GalrD contains a general acid/base Lys 197 at the end of the second β-strand in the $(\beta/\alpha)_7\beta$-barrel domain, Asp 226, Glu 252, and Glu 278 as ligands for the essential Mg^{2+} at the ends of the third, fourth, and fifth β-strands, a general acid/base His 328-Asp 301 dyad at the ends of the seventh and sixth β-strands, and an electrophilic Glu 348 at the end of the eighth β-strand. We discovered the function of STM3697 by screening a library of acid sugars; it catalyzes the efficient dehydration of both L-talarate ($k_{cat} = 2.1\ s^{-1}$, $k_{cat}/K_m = 9.1 \times 10^3\ M^{-1}\ s^{-1}$) and galactarate ($k_{cat} = 3.5\ s^{-1}$, $k_{cat}/K_m = 1.1 \times 10^4\ M^{-1}\ s^{-1}$). Because L-talarate is a previously unknown metabolite, we demonstrated that *S. typhimurium* LT2 can utilize L-talarate as carbon source. Insertional disruption of the gene encoding STM3697 abolishes this phenotype; this disruption also diminishes, but does not eliminate, the ability of the organism to utilize galactarate as carbon source. The dehydration of L-talarate is accompanied by competing epimerization to galactarate; little epimerization to L-talarate is observed in the dehydration of galactarate. On the basis of (1) structures of the wild type enzyme complexed with L-lyxarohydroxamate, an analogue of the enolate intermediate, and of the K197A mutant complexed with L-glucarate, a substrate for exchange of the α-proton, and (2) incorporation of solvent deuterium into galactarate in competition with dehydration, we conclude that Lys 197 functions as the galactarate-specific base and His 328 functions as the L-talarate-specific base. The epimerization of L-talarate to galactarate that competes with dehydration can be rationalized by partitioning of the enolate intermediate between dehydration (departure of the 3-OH group catalyzed by the conjugate acid of His 328) and epimerization (protonation on C2 by the conjugate acid of Lys 197). The promiscuous catalytic activities discovered for STM3697 highlight the evolutionary potential of a "conserved" active site architecture.

The reactions catalyzed by members of the mechanistically diverse enolase superfamily are initiated by general base-catalyzed, Mg^{2+}-assisted enolization of a carboxylate anion substrate (*1*−*3*). The stabilized enolate anion intermediate is then directed to product, usually with the assistance of a general acid catalyst. Coordination of the Mg^{2+} to the carboxylate group of the substrate and, subsequently, to the vicinal oxygens of the enolate intermediate allows essential stabilization of the intermediate so that it can be kinetically competent.

The active sites are located at the C-terminal ends of the β-strands in a $(\beta/\alpha)_7\beta$-barrel domain. Conserved (Asp or Glu) ligands for the essential Mg^{2+} are located at the ends of the third, fourth, and fifth β-strands. The general basic catalyst that abstracts the α-proton of the substrate can be located at the end of the second, sixth, or seventh β-strand. The general acid catalyst that directs the enolate intermediate to product can be located at the end of the third, fifth, sixth, or seventh β-strand.

On the basis of the positions and identities of the acid/base catalysts, the members of the superfamily can be partitioned into three major subgroups.

(1) Enolase: The members share a general basic Lys located at the end of the sixth β-strand and a general acid Glu in a loop following the second β-strand. To date, enolase is the only reaction associated with this subgroup.

[†] This research was supported by Program Project Grant GM-71790 from the National Institutes of Health.

[‡] The X-ray coordinates and structure factors for wild type TalrD/GalrD, wild type TalrD/GalrD liganded with Mg^{2+} and L-lyxarohydroxamate, and the K197A mutant liganded with Mg^{2+} and L-glucarate have been deposited in the Protein Data Bank (PDB accession codes, 2PP0, 2PP1, and 2PP3, respectively).

* To whom correspondence should be addressed. J.A.G.: Department of Biochemistry, University of Illinois, 600 S. Mathews Avenue, Urbana, IL 61801. Phone: (217) 244-7414. Fax: (217) 244-6538. E-mail: j-gerlt@uiuc.edu. S.C.A.: Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461. Phone: (718)430-2746. Fax: (718)430-8565. E-mail: almo@aecom.yu.edu.

[§] University of Illinois at Urbana−Champaign.

[⊥] Albert Einstein College of Medicine.

[#] This article is dedicated in memory of Frank H. Westheimer who taught one of us (J.A.G.) that solving a scientific problem often requires new intellectual and experimental approaches.

10.1021/bi7008882 CCC: $37.00
Published on Web 07/25/2007

(2) Muconate lactonizing enzyme (MLE[1]): the members share general acid/base Lys residues located at the ends of both the second and sixth β-strands. To date, four reactions are associated with this subgroup: MLE, *o*-succinylbenzoate synthase, L-Ala-D/L-Glu epimerase (*4*), and N-succinylamino acid racemase (NSAR) (*5*).

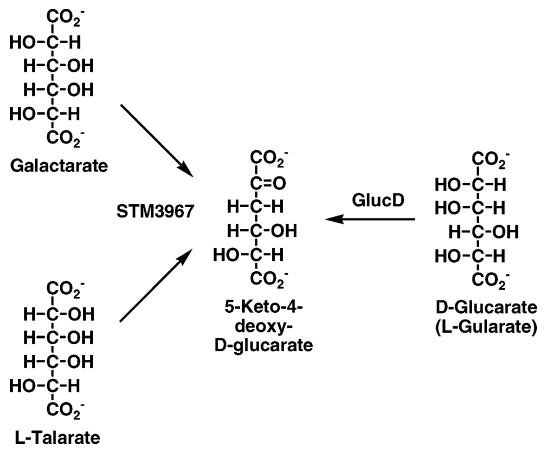
(3) Mandelate racemase (MR): The members share a general acid/base His/Asp dyad located at the ends of the seventh and sixth β-strands, respectively. To date, eight reactions are associated with this subgroup: MR, D-galactonate dehydratase (GalD) (*1*), D-glucarate dehydratase (GlucD; also catalyzes the dehydration of L-idarate and the interconversion of D-glucarate and L-idarate by epimerization at C5 (*6*)), D-gluconate dehydratase (*7*, *8*), L-fuconate dehydratase (FucD) (*9*), and D-tartrate dehydratase (TarD) (*10*).

Despite this abundance of functions, many members of the superfamily have unknown or uncertain functions: their sequences were discovered in genome projects and are highly diverged from those members with rigorously assigned functions. Although most of these are bacterial proteins that are often encoded by operons, assignment of function remains a major challenge. Therefore, we are working toward the goal of using computational methods to predict the substrate specificities of unknown and/or misannotated members so that assignment of function might be facilitated (the reactions catalyzed by the known members of the superfamily are unimolecular, so correct prediction of specificity would be sufficient to predict function). As a complement to these computational approaches, we recently described the successful use of small libraries of possible substrates to assign the NSAR function in the MLE subgroup (*5*) and the FucD and TarD functions in the MR subgroup (*9*, *10*).

In this manuscript we describe screening a library of acid sugars to assign a new function to an orthologous group of several previously uncharacterized bacterial members of the MR subgroup. Like MR, FucD, and TarD, the members of the group were predicted to contain (1) a KxK motif at the end of the second β-strand, where the second Lys is expected to be an acid/base catalyst; (2) Asp, Glu, and Glu ligands for the essential Mg^{2+} at the ends of the third, fourth, and fifth β-strands, respectively; (3) a general acid/base His-Asp dyad located at the ends of the seventh and sixth β-strands, respectively; and (4) a Glu at the end of the eighth β-strand.

Unlike the genome contexts of many other members of the MR subgroup that include sugar kinases, dehydrogenases, aldolases, and/or mutarotases and, therefore, are known or assumed to be acid sugar dehydratases, the genome contexts of the members of this "new" group of uncharacterized proteins include only a conserved, divergently transcribed transcriptional regulator and, occasionally, a protein annotated as a "putative permease". Thus, genome context provides no useful clues for functional assignment. We screened two members of this group, STM3697 from *Salmonella typhimurium* LT2 (GI:16766982) and Bpro_0435 from *Polaromonas* sp. JS666 (GI:9178645), with a library of mono- and diacid sugars and identified galactarate and L-talarate as the only substrates for dehydration. Although we studied the mechanisms of the reactions catalyzed by both enzymes, in this manuscript we describe our studies for STM3697 from *S. typhimurium* LT2. STM3697 catalyzes not only the dehydration of both galactarate and L-talarate but also their interconversion by epimerization (Scheme 1), with Lys 197 located at the end of the second β-strand identified as the galactarate-specific base and His 328 located at the end of the seventh β-strand identified as the L-talarate-specific base. On the basis of these studies with the purified protein, we conclude that STM3697 is a promiscuous L-talarate dehydratase/galactarate dehydratase (TalrD/GalrD) that also catalyzes interconversion of both substrates by epimerization.

Scheme 1



L-Talarate has not been reported as a metabolite for any organism. Thus, we sought evidence that the TalrD function deduced from library screening is physiologically relevant. We demonstrated that *S. typhimurium* LT2 can utilize L-talarate as a carbon source and that disruption of the gene encoding STM3697 abolishes the ability of the organism to utilize L-talarate. We also introduced the genes encoding STM3697 and the proximal transcriptional regulator and "permease" into *Escherichia coli* K-12 which lacks these genes; the transformed cells acquired the ability to utilize L-talarate as carbon source.

We conclude that the TalrD function is physiologically relevant for STM3697, thereby highlighting the use of library screening to identify not only new enzymatic functions in the enolase superfamily but also new metabolites. Furthermore, the discovery of this promiscuous dehydratase/epimerase provides further evidence for the catalytic versatility of active sites in the enolase superfamily.

MATERIALS AND METHODS

^{1}H NMR spectra were recorded using a Varian Unity INOVA 500NB MHz NMR spectrometer.

Cloning, Expression, and Protein Purification of STM3697, a Protein of Previously Unknown Function. The gene encoding STM3697 (GI:16766982) was PCR amplified from genomic DNA isolated from *S. typhimurium* LT2 (ATCC) using Platinum *Pfx* DNA polymerase (Invitrogen). The PCR reaction (100 μL) contained 1 ng of plasmid DNA, 10 μL of 10X *Pfx* amplification buffer, 1 mM $MgSO_4$, 0.4 mM of each dNTP, 40 pmol of each primer (forward primer

[1] Abbreviations: FucD, L-fuconate dehydratase; GalD, D-galactonate dehydratase; GalrD, galactarate dehydratase; GlucD, D-glucarate/L-idarate dehydratase; GK, glycerate kinase; KDGlucA, 5-keto-4-deoxy-D-glucarate aldolase; MLE, muconate lactonizing enzyme; MR, mandelate racemase; NSAR, N-succinylamino acid racemase; TalrD, L-talarate dehydratase; TSAR, tartronate semialdehyde reductase.

5′-GTGATTATCAGGAGAAAACAT-ATGGCTT-TAAGCGCGAATTCCG-3′ and reverse primer 5′-GATTC-CCGCCAGG-ATCCTTAAGGGCGTTTGCCAAATTCAC-3′), and 5 U of platinum *Pfx* DNA polymerase. The gene was amplified using a PTC-200 Gradient Thermal Cycler (MJ Research), with the following parameters: 94 °C for 2 min followed by 40 cycles of: 94 °C for 1 min, a gradient temperature range of 45−60 °C for 1 min and 15 s, 68 °C for 2 min, and a final extension of 68 °C for 10 min. The amplified gene was cloned into a pET-15b (Novagen) vector.

The protein was expressed in *E. coli* strain BL21(DE3). Transformed cells were grown at 25 °C in LB broth (supplemented with 100 μg/mL ampicillin) for 48 h and harvested by centrifugation. No IPTG was added to induce protein expression. The cells were resuspended in binding buffer (5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9, and 5 mM $MgCl_2$) and lysed by sonication. The lysate was cleared by centrifugation, and the His-tagged protein was purified using a column of chelating Sepharose Fast Flow (GE Healthcare Bio-Sciences Corp.) charged with Ni^{2+}. Cell lysate was applied to the column in binding buffer, washed with 5% elute buffer (1 M imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9, and 5 mM $MgCl_2$)−95% wash buffer (60 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9, and 5 mM $MgCl_2$), and eluted with 100 mM L-histidine, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9, and 5 mM $MgCl_2$. The N-terminal His-tag was removed with thrombin (GE Healthcare Bio-Sciences Corp.) according to the manufacturer's instructions, and the protein was purified to homogeneity on a Q Sepharose high performance column (GE Healthcare Bio-Sciences Corp.) equilibrated with binding buffer (25 mM Tris-HCl, pH 7.9, 5 mM $MgCl_2$) and eluted with a linear gradient of 0 to 0.5 M elution buffer (1 M NaCl, 25 mM Tris-HCl, pH 7.9, 5 mM $MgCl_2$).

Site-Directed Mutagenesis and Protein Purification. The site-directed K197A, H328N, and H328A mutants of STM3697 were constructed using the QuikChange kit (Stratagene), verified by sequencing, expressed in the BL21-(DE3) *E. coli* cells, and purified as previously described for the wild-type protein.

Synthesis of a Mixture of L-Lyxarohydroxamate and L-Arabarohydroxamate. Potassium L-lyxarate (1 mmol) was converted to a mixture of L-lyxarolactone and L-arabarolactone by passage through a column of Dowex-50 resin (200−400 mesh, H^+-form) followed by rotary evaporation and storage in a vacuum desiccator in the presence of $Mg(ClO_4)_2$ for 4 days. An excess volume of 50% aqueous hydroxylamine was added to the mixture of lactones, and the solution was allowed to stir for 10 min at room temperature to yield a mixture of L-lyxaro- and L-arabarohydroxamic acids (*11*). Unreacted hydroxylamine and solvent were removed by rotary evaporation. The product mixture was characterized by ^1H and ^{13}C NMR spectroscopy. For L-lyxarohydroxamate: ^1H NMR (500 MHz, D_2O) δ 4.04 (d, $J = 1.3$ Hz, 1H), 3.87 (m, 2H); ^{13}C NMR (500 MHz, D_2O) δ 180.0 (C5), 171.0 (C1), 73.1, 72.0, 63.1 (C2−C4). For L-arabarohydroxamate: ^1H NMR (500 MHz, D_2O) δ 4.06 (d, $J = 1,3$ Hz, 1H), 3.93 (m, 2H); ^{13}C NMR (500 MHz, D_2O) δ 180.0 (C5), 171.0 (C1), 73.3, 72.4, 63.1 (C2−C4).

Crystallization and Data Collection. Three different crystal forms (Table 1) were grown by hanging drop vapor diffusion at room temperature: (1) selenomethionine (SeMet)-substituted wild type STM3697, (2) wild-type STM3697 complexed with Mg^{2+} and L-lyxarohydroxamate, and (3) the K197A mutant of STM3697 complexed with Mg^{2+} and L-glucarate. The following crystallization conditions were used:

Table 1: Kinetic Parameters of TalrD/GalrD

substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})
L-talarate	2.1 ± 0.1	0.23 ± 0.04	9.1×10^3
galactarate	3.6 ± 0.2	0.33 ± 0.07	1.1×10^4

(1) SeMet-substituted wild type STM3697: the protein solution contained SeMet-substituted STM3697 (85 mg/mL) in 20 mM Tris-HCl, pH 7.9, 100 mM NaCl, and 1 mM β-mercaptoethanol; the precipitant contained 1.6 M ammonium sulfate and 0.1 M Tris-HCl, pH 8.5. Crystals appeared in 3 days and exhibited diffraction consistent with space group *P*422, with three molecules of STM3697 per asymmetric unit.

(2) Wild type STM3697 complexed with Mg^{2+} and L-lyxarohydroxamate: the protein solution contained STM3697 (36.2 mg/mL) in 20 mM Tris-HCl, pH 7.9, 100 mM NaCl, 5 mM $MgCl_2$, and 10 mM of a mixture of L-lyxarohydroxamate and L-arabarohydroxamate; the precipitant contained 2.0 M ammonium sulfate, 0.1 M Hepes, pH 7.5, and 5 mM $MgCl_2$. Crystals appeared in 5 days and exhibited diffraction consistent with space group *P*222, with six molecules of STM3697 per asymmetric unit.

(3) The K197A mutant of STM3697 complexed with Mg^{2+} and L-glucarate: the protein solution contained the K197A mutant (47.9 mg/mL) in 20 mM Tris-HCl, pH 7.9, 100 mM NaCl, 5 mM $MgCl_2$, and 20 mM L-glucarate. The precipitant contained 2.0 M ammonium sulfate and 0.1 M Bis-Tris (pH 5.5). Crystals appeared in 7 days and exhibited diffraction consistent with space group *P*422, with three molecules of STM3697 per asymmetric unit.

Prior to data collection, the crystals were transferred to cryoprotectant solutions composed of their mother liquids (including metal and ligands) and 20% glycerol. After incubation for ∼15 s, the crystals were flash-cooled in a nitrogen stream. A single wavelength anomalous dispersion (SAD) data set for a crystal of SeMet-substituted wild type STM3697 (Table 2, column 1) was collected to 2.2 Å resolution at the NSLS X4A beamline (Brookhaven National Laboratory) on an ADSC CCD detector. Data sets for the complexes of wild type STM3697 with Mg^{2+} and L-lyxarohydroxamate (column 2) and of the K197A mutant with Mg^{2+} and L-glucarate (column 3) were collected at the same beamline to the same resolution. Intensities were integrated and scaled with DENZO and SCALEPACK (*12*). Data collection statistics are given in Table 2.

Structure Determination and Refinement. The structure of the SeMet-substituted wild type STM3697 was determined by SAD with SOLVE (*13*); 27 of the 30 selenium sites were identified. These heavy atom sites were used to calculate initial phases which were improved by solvent flattening and NCS-averaging with RESOLVE (*14*), yielding an interpretable map for three monomers in the asymmetric unit for space group *P*422. Iterative cycles of automatic rebuilding with ARP (*15*), manual rebuilding with TOM (*16*), and refinement with CNS (*17*) resulted in a model at 2.2 Å with R_{cryst} of 0.207 and an R_{free} of 0.223. The first three residues

Table 2: Data Collection and Refinement Statistics

	STM3697 (SeMet)	STM3697·Mg^{2+}·LLH[a]	K197A·Mg^{2+}·LGL[b]
data collection			
beamline	NSLS X4A	NSLS X4A	NSLS X4A
wavelength (Å)	0.97915	0.979	0.979
space group	*P*422	*P*222	*P*422
no. of molecules in a.u.	3	6	3
unit cell parameters			
a (Å)	173.89	123.19	174.23
b (Å)	173.89	173.83	174.23
c (Å)	123.20	173.79	123.57
resolution (Å)	2.2	2.2	2.2
no. of unique reflections	86423	163956	83947
completeness (%)	90.3	89.8	87.1
R_{merge}	0.056	0.069	0.074
average I/σ	34.5	29.9	31.2
refinement			
resolution	25.0−2.2	25.0−2.2	25.0−2.2
R_{cryst}	0.207	0.223	0.204
R_{free}	0.223	0.237	0.231
rmsd for bonds (Å)	0.007	0.006	0.006
rmsd for angles (deg)	1.3	1.4	1.3
no. of protein atoms	9243	18386	9231
no. of waters	437	593	304
no. of Mg^{2+} ions	0	6	3
bound ligand	glycerol	Mg^{2+}, LLH	Mg^{2+}, LGL
ligand atoms	18	80	45
PDB entry	2PP0	2PP1	2PP3

[a] L-Lyxarohydroxamate [b] L-Glucarate

in all three monomers in the asymmetric unit are disordered and are not included in the final model. One glycerol molecule is bound in the active site of each monomer.

The structure contains two types of dimers (A and B). Dimer A consists of two monomers connected by a crystallographic 2-fold axis, where residues 116−124 of the N-terminal cappping domain of each monomer are inserted into the cavity between the N-terminal cappping and $(\beta/\alpha)_7\beta$-barrel domains of a neighboring monomer; these dimers are analogous to those previously observed in the structures of MR, FucD, and TarD. Dimer B consists of two monomers with their N-terminal residues 3−26 swapped; these monomers are connected by a noncrystallographic 2-fold axis. As the consequence of additional crystallographic operations, each of these dimers results in the formation of similar octamers with 422 symmetry. The STM3697 octamer is similar to that observed in MR and MLE.

The structure of wild type STM3697 complexed with Mg^{2+} and L-lyxarohydroxamate was determined by molecular replacement with PHASER (*18*), using the SeMet-substituted STM3697 structure as the search model. Iterative cycles of automatic rebuilding with ARP, manual rebuilding with TOM, and refinement with CNS were performed. The model was refined at 2.2 Å with an R_{cryst} of 0.223 and an R_{free} of 0.237. The Mg^{2+} was clearly visible in the electron density maps for all 6 protein molecules in the asymmetric unit. Although the crystallization conditions included a mixture of L-lyxarohydroxamate and L-arabarohydroxamate, a molecule of only L-lyxarohydroxamate molecule was well-defined in each STM3697 monomer.

The structure of the K197A mutant complexed with Mg^{2+} and L-glucarate was determined by molecular replacement using the SeMet-substituted STM3697 structure as the search model. The model was refined at 2.2 Å with an R_{cryst} of 0.204 and an R_{free} of 0.231. The Mg^{2+} ion and L-glucarate molecule were well-defined in all three polypeptides in the asymmetric unit. This structure and that of wild type STM3697 liganded with Mg^{2+} and L-lyxarohydroxamate contain the same dimers described for the SeMet-substituted wild type protein.

Final refinement statistics for all three structures are given in Table 2.

Screening a Library of Acid Sugars for Dehydratase Activity. Dehydration of the members of a library of mono- and diacid sugars was monitored after 16 h by end-point detection of the semicarbazone derivative at 250 nm, as previously described (*9*). The dehydration reactions were performed at 30 °C in 50 mM Tris-HCl (pH 8.0), containing 10 mM $MgCl_2$, 10 mM acid sugar, and 1 μM enzyme.

Screening a Library of Acid Sugars for Epimerization and Exchange of the α*-Proton.* To determine whether STM3697 catalyzes either exchange of the α-proton and/or epimerization of the members of the library of acid sugars, ^{1}H NMR spectra were recorded after STM3697 was incubated for 16 h with the members of the library in a D_2O-containing buffer at 25 °C. A typical reaction (800 μL) contained a mixture of pairs of acid sugars (to reduce the number of NMR spectra) at a concentration of 2 mM for each acid sugar, 50 mM Tris-DCl, pD 7.5, 10 mM $MgCl_2$, and 1 μM of STM3697. STM3697 was exchanged into D_2O buffer (50 mM Tris-DCl, pD 7.5) using an Amicon (10 000 Da) stirred ultrafiltration cell by repeated concentration and dilution.

Aerobic Growths of Bacterial Strains for Carbon Source Testing. For testing possible carbon sources, 3 mL cultures containing 10 mM of the carbon source (e.g., 10 mM L-talarate, galactarate, D-glucarate, and D-glucose) in M9 minimal medium were inoculated with 30 μL of an overnight nutrient broth (LB) culture of bacteria (washed once with M9 minimal salts medium) and grown aerobically at 37 °C. Growth was monitored spectrophotometrically by measuring the absorbance at 600 nm.

Construction of Insertional Deletions in S. typhimurium LT2. The method described by Datsenko and Wanner (*19*) was used to construct insertional deletions of genes *STM3697*, *garL* (encoding KDGlucA), *garD* (encoding GalcD), *STM3698*, and *gudT* (encoding putative D-glucarate transporter) in *S. typhimurium* strain LT2. A 1113 bp region in the middle of the 1197 bp *STM3697* gene was replaced with the kanamycin resistance gene (for neomycin phosphotransferase) from plasmid pKD13 (*STM3697::kan*). A 687 bp region in the middle of the 771 bp *garL* gene was replaced with the kanamycin resistance gene from plasmid pKD13 (*garL::kan*). A 1488 bp region in the middle of the 1572 bp *garD* gene was replaced with the chloramphenicol resistance gene (for chloramphenicol acetyltransferase) from plasmid pKD3 (*garD::cam*). A 1227 bp region in the middle of the 1311 bp *STM3698* gene was replaced with the kanamycin resistance gene (for neomycin phosphotransferase) from plasmid pKD13 (*STM3698::kan*). A 1275 bp region in the middle of the 1359 bp *garD* gene was replaced with the chloramphenicol resistance gene (for chloramphenicol acetyltransferase) from plasmid pKD3 (*gudT::cam*). The positions of the deletions were confirmed by DNA sequence analysis of PCR-amplified regions of the mutant genomes using locus-specific primers flanking the intended deletion sites of the *STM3697*, *garL*, *garD*, *STM3698* and *gudT* genes.

Cloning of the Operon Containing STM3697 in S. typhimurium LT2. The operon encoding STM3696, STM3697,

and STM3698 was PCR amplified from genomic DNA isolated from *S. typhimurium* LT2 (ATCC) using *Pfu* Ultra HF DNA polymerase (Stratagene). The PCR reaction (100 μL) contained 1 ng of plasmid DNA, 10 μL of 10X *Pfu* Ultra HF amplification buffer, 0.4 mM of each dNTP, 100 ng/μL of each primer (forward primer 5′-GTCGGTGGTGTCTAGAGAAGCGTGGCGCCAGTTAG-GATATCC-3′ and reverse primer 5′-CGATGGAGTGCTCGAGCGGAACAGCGG-AGCGAAGAATTCGC-3′), and 5 U of *Pfu* Ultra HF DNA polymerase. The gene was amplified using a PTC-200 gradient thermal cycler (MJ Research), with the following parameters: 95 °C for 2 min followed by 40 cycles of: 95 °C for 30 s, 55 °C for 30 s, 72 °C for 9 min, and a final extension of 72 °C for 10 min. The amplified gene was cloned into a pET-17b (Novagen) vector. The truncated operon encoding *STM3696* and *STM3697* was similarly PCR amplified and cloned into a pET-17b vector using 100 ng/μL of each primer (forward primer 5′-GTCGGTGGTGTCTAGAGAAGCGTGGCGCCAGTTA-GGATATCC-3′ and reverse primer 5′-GATTCCCGCCACTCGAGTTAAGGGC-GTTTGCCAAATTCACATG-3′) and conditions listed above.

Complementation Growths of E. coli K-12 Strain BW25113. The *E. coli* K-12 strain BW25113 was transformed with the T7 RNA polymerase-encoding plasmid, pTara (*20*), and subsequently transformed with either pET17b-STM3696-3698 or pET17b-STM3696-3697. Aerobic growths of these resulting strains on different carbon sources were performed as described before.

Kinetic Assay of L*-Talarate Dehydratase and Galactarate Dehydratase Activity.* L-Talarate and galactarate dehydration activities were assayed by a continuous coupled spectrophotometric assay, using 5-keto-4-deoxy-D-glucarate aldolase (KDGlucA, previously cloned and expressed from the D-glucarate/galactarate pathway in *E. coli* (*21*)) to cleave the dehydrated 5-keto-4-deoxy-D-glucarate product to pyruvate and tartronate semialdehyde and L-lactate dehydrogenase (LDH) to reduce the pyruvate to L-lactate. The oxidation of NADH was quantitated by measuring the decrease in absorbance at 340 nm. The assay (1 mL at 25 °C) contained STM3697, 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 0.16 mM NADH, 30 U of KDGlucA, and 30 U of LDH and L-talarate or galactarate.

Polarimetric Detection of L*-Talarate/Galactarate Epimerization and Dehydration by STM3697.* The epimerization and dehydration of L-talarate and galactarate by TalrD was monitored at 25 °C using 10 mM L-talarate or galactarate in 50 mM Tris-HCl, pH 8.0, 5 mM $MgCl_2$, and 1 μM STM3697. The change in optical rotation was quantitated using the sodium D line at 589 nm using a JASCO P-1010 polarimeter, with a 10 s integration time and a 10 cm path length cuvette.

Dehydration of L*-Talarate and Galactarate Monitored by ^{1}H NMR Spectroscopy.* Dehydration of L-talarate and galactarate by STM3697 was separately determined by ^{1}H NMR spectroscopy. The sample for NMR analysis (800 μL at 25 °C) contained 10 mM L-talarate or galactarate, 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, and 1 μM of STM3697. The mixture was incubated for 16 h at 25 °C, lyophilized, and resuspended in 800 μL of D_2O. A duplicate sample without enzyme was prepared for comparison. L-Talarate and galactarate were also separately dehydrated by STM3697 in a D_2O-containing buffer at 20 °C, with ^{1}H NMR spectra recorded as a function of time after initiation of the reaction. A typical reaction mixture contained 10 mM L-talarate or galactarate, 50 mM Tris-DCl (pD 7.5), 10 mM $MgCl_2$, and 0.5 μM STM3697. All enzymes were exchanged into D_2O buffer [50 mM Tris-DCl (pD 7.5)] using an Amicon (10 000 Da) stirred ultrafiltration cell.

TalrD-Catalyzed Exchange of the α*-Proton of* L*-Glucarate.* L-Glucarate was incubated with STM3697 in D_2O buffer at 20 °C, and ^{1}H NMR spectra were recorded as a function of time. A typical reaction contained 10 mM L-glucarate, 50 mM Tris-DCl (pD 7.5), 10 mM $MgCl_2$, and 0.5 μM STM3697. The rate of exchange (k_{exc}) of the α-proton of L-glucarate was calculated from eq 1, which takes into account the total amount of L-glucarate bound at any time:

$$k_{exc} = k_{obs}[\text{L-glucarate}]_T/[\text{L-glucarate}]_B \qquad (1)$$

where k_{obs} is the observed first-order rate constant, $[\text{L-glucarate}]_T$ is the total L-glucarate concentration, and $[\text{L-glucarate}]_B$ is the concentration of bound L-glucarate. Because the concentration of L-glucarate (10 mM) was significantly higher than the enzyme concentration (0.5 μM), the concentration of bound L-glucarate ($[\text{L-glucarate}]_B$) was equal to the enzyme concentration.

Stereochemical Course of L*-Talarate and Galactarate Dehydration.* To determine the stereochemical course of dehydration, L-talarate and galactarate were separately dehydrated in a D_2O-containing buffer at 20 °C, with ^{1}H NMR spectra recorded upon completion of the reaction (as assayed by semicarbazone formation). A typical reaction (800 μL) contained 10 mM L-talarate or galactarate, 50 mM Tris-DCl, pD 7.5, 10 mM $MgCl_2$, and 1 μM STM3697. STM3697 was exchanged into D_2O buffer (50 mM Tris-DCl, pD 7.5) using an Amicon (10 000 Da) stirred ultrafiltration cell.

RESULTS AND DISCUSSION

Although we have identified several diverse functions in the enolase superfamily, many members remain to be functionally assigned. Our goal is to develop and implement an integrated sequence-structure-computation strategy for predicting the substrate specificities of the uncharacterized members that will allow their functions to be assigned (*22*). So that we might expand the "library" of known enzymatic functions and associated structures, thereby facilitating predictions of function, we have adopted the strategy of screening unknown members with potential substrates to discover function.

The selection of the group of proteins that is the focus of this manuscript was based on the large number of members of the enolase superfamily encoded by the genome of *Polaromonas* JS666. This genome encodes 15 members, 5 of which could be assigned by sequence identity; the genome of *E. coli* K-12 encodes 8 members, 7 of which have assigned functions. We assumed that the uncharacterized members in *Polaromonas* would encode "new" functions, so their genes were cloned, the proteins were purified, and these were screened for dehydration activity using a library of acid sugars. The function described in this manuscript was discovered using one of those proteins (GI:91786345; Bpro_0435). Although we studied the mechanism of its reaction, the studies described in this manuscript focus on

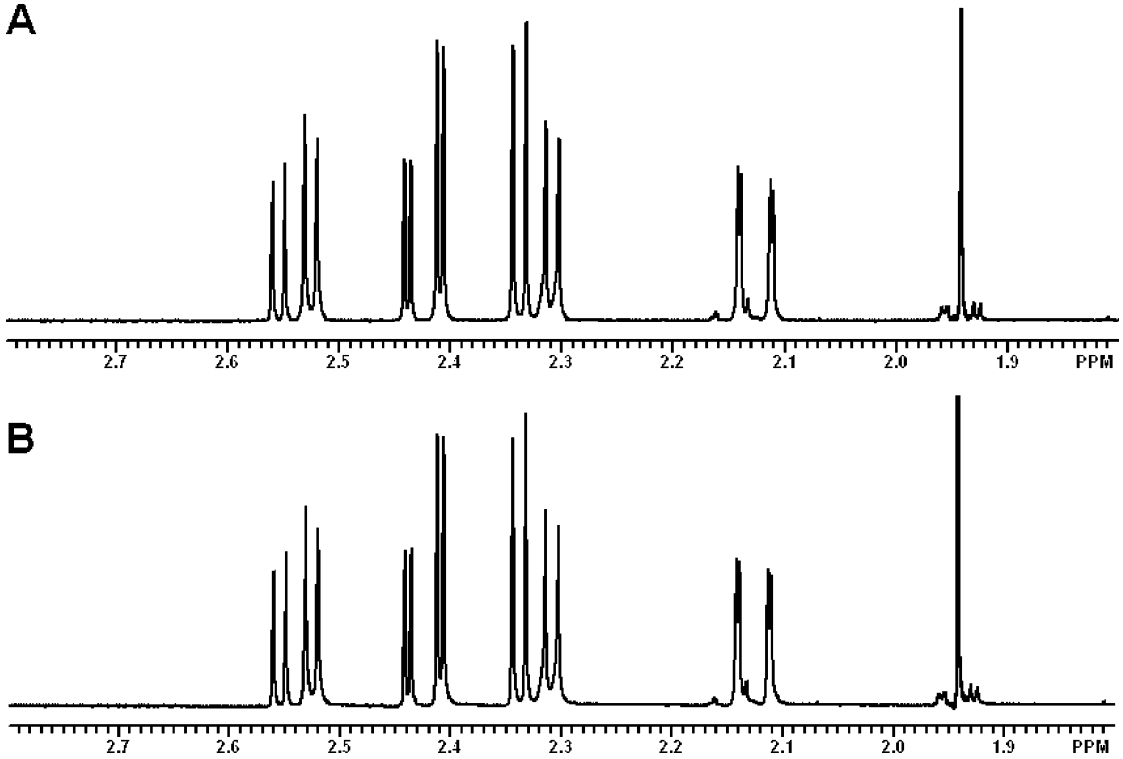


FIGURE 1: (A) Partial ^{1}H NMR spectrum of the dehydrated L-talarate product, 5-keto-4-deoxy-D-glucarate, of TalrD/GalrD. (B) Partial ^{1}H NMR spectrum of the dehydrated galactarate product, 5-keto-4-deoxy-D-glucarate, of TalrD/GalrD.

the orthologue from *S. typhimurium* LT2 (GI:16766982; STM3697), because insertional disruptions of the genes in this organism can be accomplished using the same methodology used for insertional disruption of genes in *E. coli* K-12 (*19*), an experimental approach that was essential for assignment of function.

Identification of L-Talarate and Galactarate as Substrates for STM3697 from S. typhimurium LT2 (GI:16766982). STM3697 is a member of a group of proteins in the MR subgroup[2] thought to be isofunctional on the basis of sequence alignments (>60% identity) and conservation of specificity determining residues in the active site. Although the genome contexts of the orthologues encoding these proteins often include a conserved, divergently transcribed transcriptional regulator and a "permease", no useful clues that would aid specific functional assignment are apparent. The utility of screening a library of acid sugars was previously demonstrated by the functional assignment of L-fuconate dehydratase and D-tartrate dehydratase (*9*, *10*). Hence, we sought to identify substrates for STM3697 by screening the library of acid sugars for dehydratase activity. We discovered that only galactarate and L-talarate were completely dehydrated by STM3697; no other acid sugar resulted in detectable turnover (<1%; results not shown).

The ^{1}H NMR spectra of both dehydration products were identical (Figure 1). Because galactarate and L-talarate are epimers at C2, dehydration of galactarate likely proceeds via an *anti*-elimination initiated by abstraction of the proton from C2 and protonation of the hydroxyl-leaving group at C3; dehydration of L-talarate likely proceeds via a *syn*-elimination mechanism (Scheme 1). These predictions are based on the stereochemical courses of the reactions catalyzed by the homologous FucD and TarD that share the same active site functional groups (*9*, *10*).

Although the ^{1}H NMR spectrum identifies the carbon from which the proton is abstracted from L-talarate, galactarate is a meso compound so the identity of the carbon from which the proton is abstracted is ambiguous (i.e., the products from L-talarate and galactarate could be enantiomers). We measured the optical rotations of both products as well as that obtained from D-glucarate by GlucD (5-keto-4-deoxy-D-glucarate) and determined that all three had the same specific rotation, $\alpha_D^{25} = +45°$, i.e., all three diacids are converted to the same product as shown in Scheme 1.

Because 5-keto-4-deoxy-D-glucarate is a metabolite in the characterized catabolic pathways for D-glucarate and galactarate in *E. coli* (*21*), we could use the aldolase from that pathway (KDGlucA) as the basis for a coupled-enzyme, spectrophotometric assay for the dehydration of both L-talarate and galactarate catalyzed by STM3697. The values of the kinetic constants are listed in Table 1. Because these are comparable to those measured for other sugar acid dehydratases of known function in the enolase superfamily, we assign STM3697 as a L-talarate dehydratase (TalrD)/galactarate dehydratase (GalrD) (Scheme 2).

Assignment of Function: Utilization of L-Talarate as Carbon Source. Although galactarate is a known carbon source for *S. typhimurium* LT2, to the best of our knowledge, L-talarate has not previously been identified as a metabolite for any organism. We determined that *S. typhimurium* LT2 utilizes L-talarate as carbon source (Figure 2, panel A). Using

[2] GI:16766982, *S. typhimurium* LT2; GI:16762612, *Salmonella enterica subsp. enterica serovar Typhi str.* CT18; GI:56415587, *Salmonella enterica subsp. enterica serovar Paratyphi A str.* ATCC 9150; GI:50119938, *Erwinia carotovora subsp. atroseptica* SCRI1043; GI:91786345, *Polaromonas sp.* JS666; GI:73538470, *Ralstonia eutropha* JMP134; GI:66963863, *Arthrobacter sp.* FB24; GI:119964208, *Arthobacter aurescens* TC1; GI:118729863, *Delftia acidovorans* SPH-1; GI:134101131, *Saccahropolyspora erythrea* NRRL 2338; GI:118471936, *Mycobacterium smegmatis* str. MC2 155; GI:116248907, *Rhizobium leguminosarum bv. viciae* 3841.

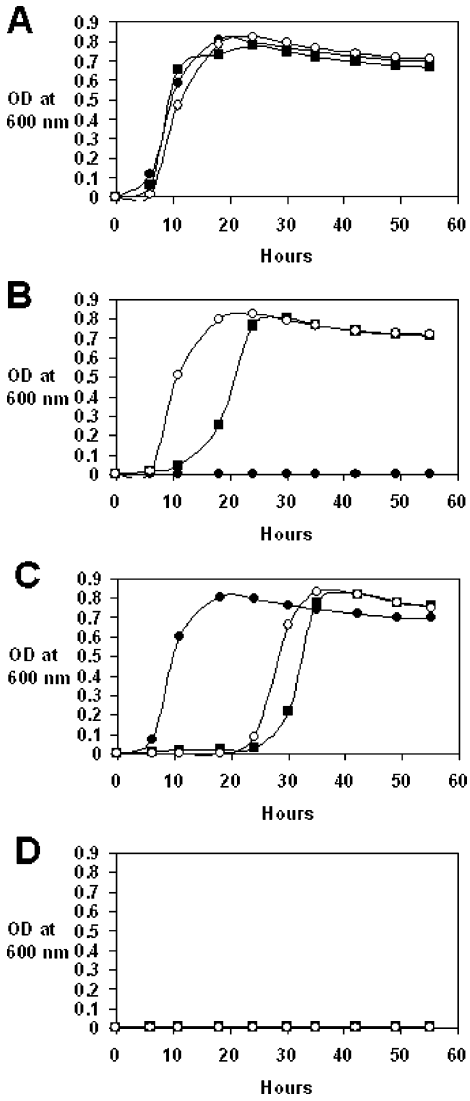
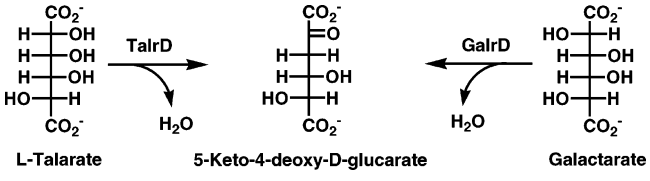


FIGURE 2: (A) Aerobic growth of *S. typhimurium* LT2 in M9 minimal medium with (●) 10 mM L-talarate, (■) 10 mM galactarate, and (○) 10 mM D-glucarate. (B) Aerobic growth of TalrD/GalrD deletion strain (*ultB::kan*) in M9 minimal medium with (●) 10 mM L-talarate, (■) 10 mM galactarate, and (○) 10 mM D-glucarate. (C) Aerobic growth of GalrD/GarD deletion strain (*garD::cam*) in M9 minimal medium with (●) 10 mM L-talarate, (■) 10 mM galactarate, and (○) 10 mM D-glucarate. (D) Aerobic growth of KDGlucA/GarL deletion strain (*garL::kan*) in M9 minimal medium with (●) 10 mM L-talarate, (■) 10 mM galactarate, and (○) 10 mM D-glucarate.

Scheme 2



10 mM concentrations of L-talarate, galactarate, and D-glucarate, comparable rates and extents of growth were observed for all three diacid sugars.

Although studies of the purified protein established dehydration of L-talarate by STM3697, growth on L-talarate does not prove that STM3697 is uniquely responsible for utilization of L-talarate. Therefore, we used the methods described for insertional disruptions of genes in *E. coli* K-12

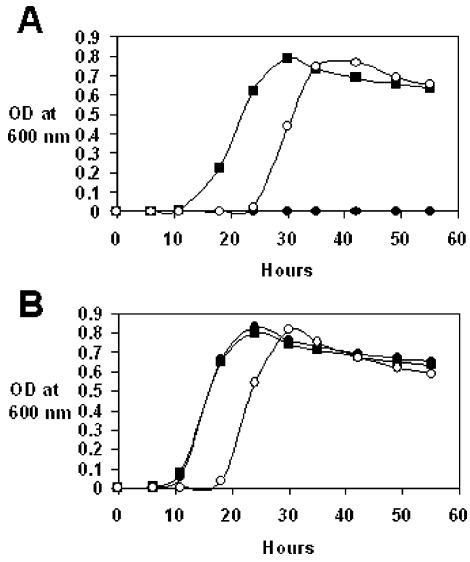


FIGURE 3: (A) Aerobic growth of BW25113 in M9 minimal medium with (●) 10 mM L-talarate, (■) 10 mM galactarate, and (○) 10 mM D-glucarate. (B) Aerobic growth of BW25113, transformed with pET17b-STM3696-3698 and pTara, in M9 minimal medium with (●) 10 mM L-talarate, (■) 10 mM galactarate, and (○) 10 mM D-glucarate.

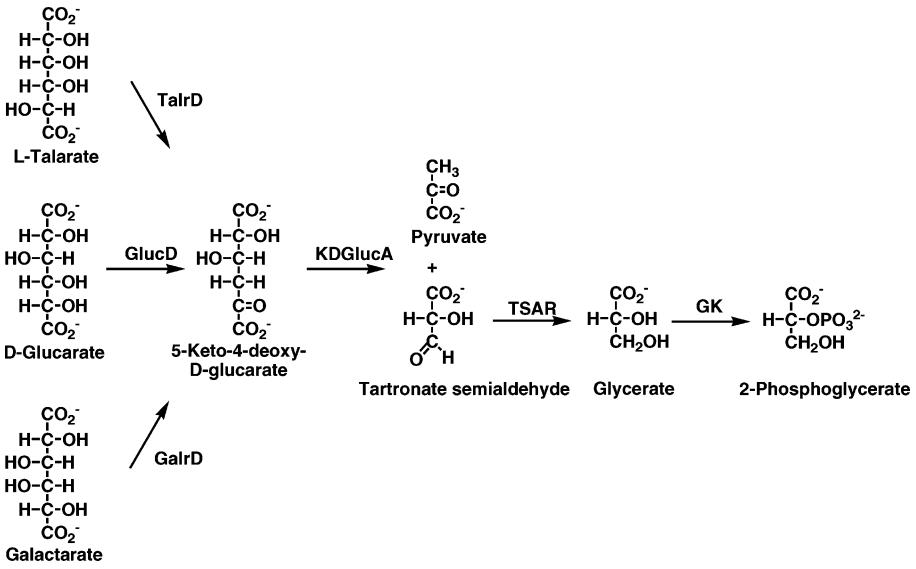
(*19*) to generate an insertional disruption of the gene encoding STM3697. This strain failed to grow on L-talarate as carbon source (Figure 2, panel B), although it retained the ability of the wild type strain to utilize D-glucarate. This phenotype provides persuasive evidence that our *in vitro* assignment of the TalrD function to STM3697 is physiologically relevant.

As shown in Figure 2, panel B, the disruption of the gene encoding STM3697 reduced the ability to utilize galactarate as carbon source. Because STM3697 is also able to dehydrate galactarate, we sought additional evidence that the protein also functions in galactarate utilization. The genome of *S. typhimurium* LT2 encodes a distinct GalrD (encoded by *garD* in the *gar* operon) that is unrelated to the enolase superfamily. The gene encoding this GalrD was insertionally disrupted, and the mutant strain was able to grow on galactarate as a sole carbon source (Figure 2, panel C), although slower than the wild type strain. This observation is consistent with the assignment of GalrD function to STM3697.

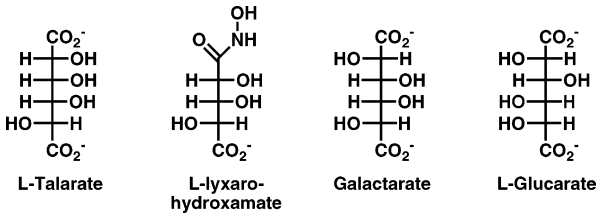
We conclude that our assignment of the TalrD and GalrD functions to STM3697 is correct and physiologically relevant.

The separate GalrD and the GlucD that is a member of the enolase superfamily yield the same product, 5-keto-4-deoxy-D-glucarate. The catabolism of this diacid is completed by an aldolase (KDGlucA), tartronate semialdehyde reductase (TSAR), and a glycerate kinase (GK) that produces 2-phosphoglycerate (*21*). Because STM3697 produces the same dehydration product, we sought evidence that the downstream catabolic pathway for L-talarate is shared with those for galactarate and D-glucarate. The gene encoding KDGlucA was disrupted; the resulting strain failed to utilize L-talarate as carbon source (Figure 2, panel D). As expected, the disruption also abolished growth of *S. typhimurium* LT2 on both D-glucarate and galactarate. Thus, we conclude that the catabolism of L-talarate is completed by the same reactions as those shared in the catabolism of galactarate and D-glucarate (Scheme 3).

Scheme 3



Scheme 4



Enabling E. coli To Utilize L*-Talarate as Carbon Source*. The genome of *E. coli* K-12 does not encode an orthologue of the TalrD/GalrD, although genes encoding GlucD, the unrelated GalrD, and downstream catabolic enzymes are present. Accordingly, *E. coli* utilizes D-glucarate and galactarate but not L-talarate as carbon source (Figure 3, panel A).

We transformed *E. coli* K-12 strain BW25113 with a plasmid containing DNA encoding the bifunctional TalrD/GalrD and the proximal transcriptional regulator and "permease". The transformed strain was able to utilize L-talarate as carbon source (Figure 3, panel B). This "gain-of-function" experiment provides additional persuasive evidence that TalrD is a correct physiological function for STM3697.[3]

Structure of TalrD/GalrD. Structures of wild type TalrD/GalrD were solved both in the absence and in the presence of L-lyxarohydroxamate (Scheme 4), an analogue of the enolate intermediate derived from abstraction of the 2-proton from either galactarate or L-talarate. The structure of the K197A mutant was solved in the presence L-glucarate, a substrate for exchange of the 2-proton but not dehydration (*vide infra*). The biological quaternary organization predicted from the structures is a tetramer of dimers, with the dimers sharing quaternary interactions similar to those previously observed in MR, FucD, and TarD.

The gene encoding TalrD/GalrD was amplified from the ATG codon that provides the longest polypeptide for the open reading frame and purified with an N-terminal His_6-tag that was removed prior to crystallization. The N-terminal three residues are disordered, and the following 27 residues precede Val 31 that is located at the beginning of the first β-strand in the N-terminal capping domain; these N-terminal residues wrap around one monomer of an adjacent dimer in the octameric structure to form dimer B (as defined in Materials and Methods). In MR, FucD, and TarD, the N-terminal residue of the polypeptide is located at the beginning of the first β-strand of the N-terminal capping domain. Perhaps the "natural" N-terminus of TalrD/GalrD is Val 31 (GTG codon) of the protein that was crystallized, with an adventitious N-terminal extension interacting with an adjacent polypeptide in the octamer.

The arrangement of polypeptides in dimer A is similar to those observed in MR, FucD, and TarD. However, in contrast to MR, FucD, and TarD in which a residue from the N-terminal domain of one polypeptide provides an essential interaction that completes the binding site for the substrate in the adjacent polypeptide of the dimer, in TalrD/GalrD the substrate binding site is formed entirely by residues in the polypeptide chain in which the acid/base catalysts and ligands for the essential Mg^{2+} are located. Residues 116−124 of the N-terminal domain of one monomer interact with interface between the N-terminal capping and $(\beta/\alpha)_7\beta$-barrel domains of the other monomer.

The active site of TalrD/GalrD complexed with L-lyxarohydroxamate is shown in Figure 4, panel A; the active site of the inactive K197A mutant complexed with L-glucarate is shown in Figure 4, panel B. As expected from sequence alignments, Asp 226, Glu 252, and Glu 278, located at the ends of the third, fourth, and fifth β-strands of the barrel domain, respectively, are ligands of the Mg^{2+}. Both the hydroxamate enolate anion analogue and the L-glucarate substrate analogue are bidentate ligands of the Mg^{2+}; in addition, the second carboxylate oxygen of L-glucarate is

[3] We performed additional physiological experiments in which the possible role of the "permease" encoded by the gene proximal encoding STM3697 in L-talarate utilization was examined. Disruption of this gene did not eliminate the ability to utilize L-talarate as carbon source, as might have been expected if it were an "L-talarate permease" (in some genomes the product of this gene is annotated as a "permease"). Disruption of the gene encoding the galactarate/D-glucarate transporter partially eliminated the ability to utilize L-talarate as carbon source but abolished the ability to utilize both D-glucarate and galactarate. We conclude that the transport of L-talarate is complex and may involve additional, unidentified transport systems.

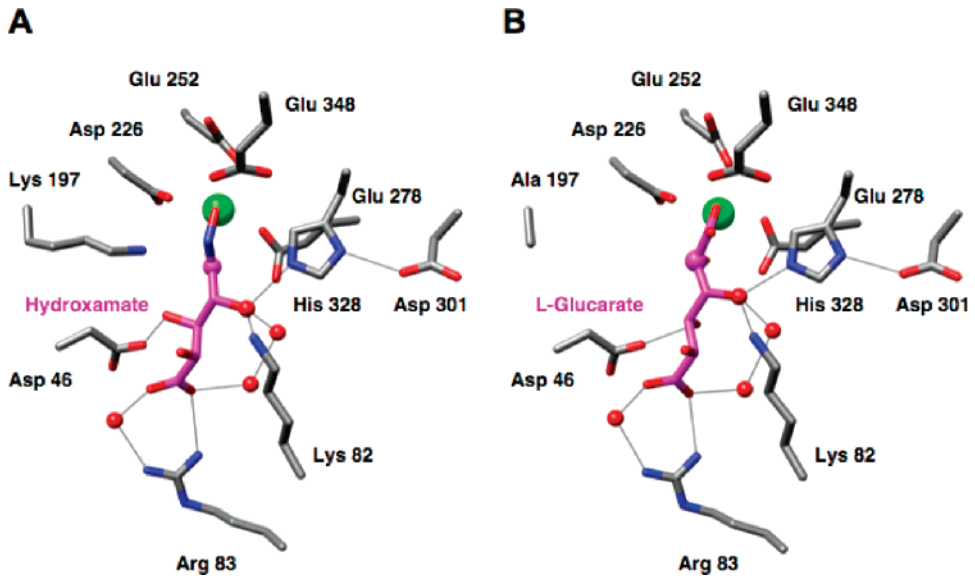


FIGURE 4: Comparison of the active site structures of (A) wild type TalrD/GalrD complexed with L-lyxarohydroxamate and (B) the inactive K197A mutant complexed with L-glucarate.

hydrogen bonded to Glu 348 at the end of the eighth β-strand. Lys 197 at the end of the second β-strand and His 328 at the end of the seventh β-strand are positioned on opposite faces of the active site. From these structures, we hypothesize that Lys 197 is the galactarate-specific acid/base catalyst and His 328 is the L-talarate specific acid/base catalyst (*vide infra*).

As shown in Figure 4, the interactions that determine substrate specificity include the following: (1) the distal carboxylate group is hydrogen bonded to the guanidinium group of Arg 83 as well as a water molecule that is hydrogen bonded to both the carbonyl oxygen of Val 44 and the guanidinium group of Arg 83; (2) the 4-OH group is hydrogen bonded to the carboxylate group of Asp 46 as well as the ammonium group of Lys 197; and (3) the 3-OH group (leaving group) is hydrogen bonded to the imidazolium group of His 328, the ammonium group of Lys 82, and a water molecule. The 5-OH group does not participate in any hydrogen bonding interactions. L-Glucarate, the 4-epimer of galactarate, is a substrate for exchange of the α-proton but does not undergo dehydration; the structure of the "inactive" K197A mutant complexed with L-glucarate does not provide a structural explanation for its lack of reactivity with respect to dehydration.

Mechanism of the TalrD/GalrD-Catalyzed Reaction: Epimerization Occurs in Competition with Dehydration. The time courses for reactions with L-talarate (red) and galactarate (green) monitored by polarimetry are shown in Figure 5 (note that galactarate is a meso compound so its optical rotation is zero). The time-dependencies of the initial rates of change in optical rotation were not consistent with the rates of dehydration measured using the coupled-enzyme assay (Table 1). Notably, the presence of a "lag" for L-talarate but not galactarate suggested that dehydration was not the only reaction occurring during the dehydration of L-talarate.

We also used ^{1}H NMR spectroscopy to study the reactions with L-talarate and galactarate. We observed that L-talarate is significantly epimerized to galactarate in competition with dehydration to 5-keto-4-deoxy-D-glucarate (Figure 6, panel A). In contrast, when galactarate is used as substrate, only

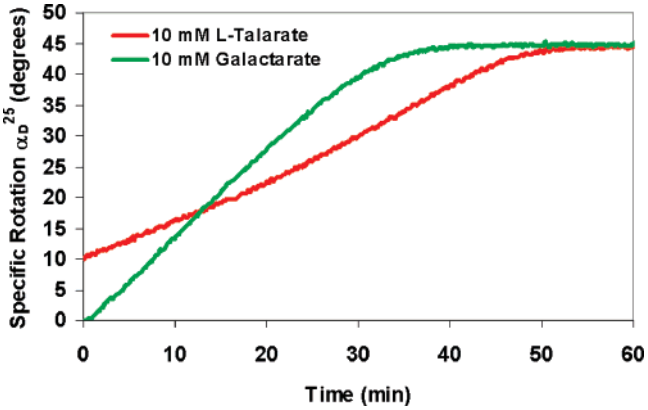


FIGURE 5: Polarimetric profiles of 10 mM L-talarate (red) and 10 mM galactarate (green) incubated with 1 μM TalrD/GalrD, at 25 °C.

small amounts of the epimerized L-talarate are observed (Figure 6, panel B). Using the time dependencies of integrals of the intensities of the resonances associated with L-talarate and galactarate, the rate of epimerization of L-talarate to galactarate is 0.65 s^{-1}, 30% of the rate of its dehydration.

Thus, the time course measured with polarimetry for L-talarate is explained by competitive partitioning of the enolate anion intermediate obtained by proton abstraction between dehydration and epimerization (protonation on the opposite face by an active site acid). When galactarate is the substrate, the enolate anion intermediate partitions almost entirely to dehydration, explaining the absence of a "lag" in the optical rotation.

Identities of the L-Talarate- and Galactarate-Specific Bases. On the basis of both sequence alignments and the experimentally determined structures, we predict that Lys 197 is the galactarate-specific base and His 328 is the L-talarate-specific base. We constructed the K197A, H328N, and H328A mutants, so that the importance of these putative acid/base catalysts could be evaluated. Each totally eliminated both the dehydration and epimerization activities using both L-talarate and galactarate (data not shown). Although these substitutions could not be used to experimentally probe the importance of these residues using polarimetry or ^{1}H

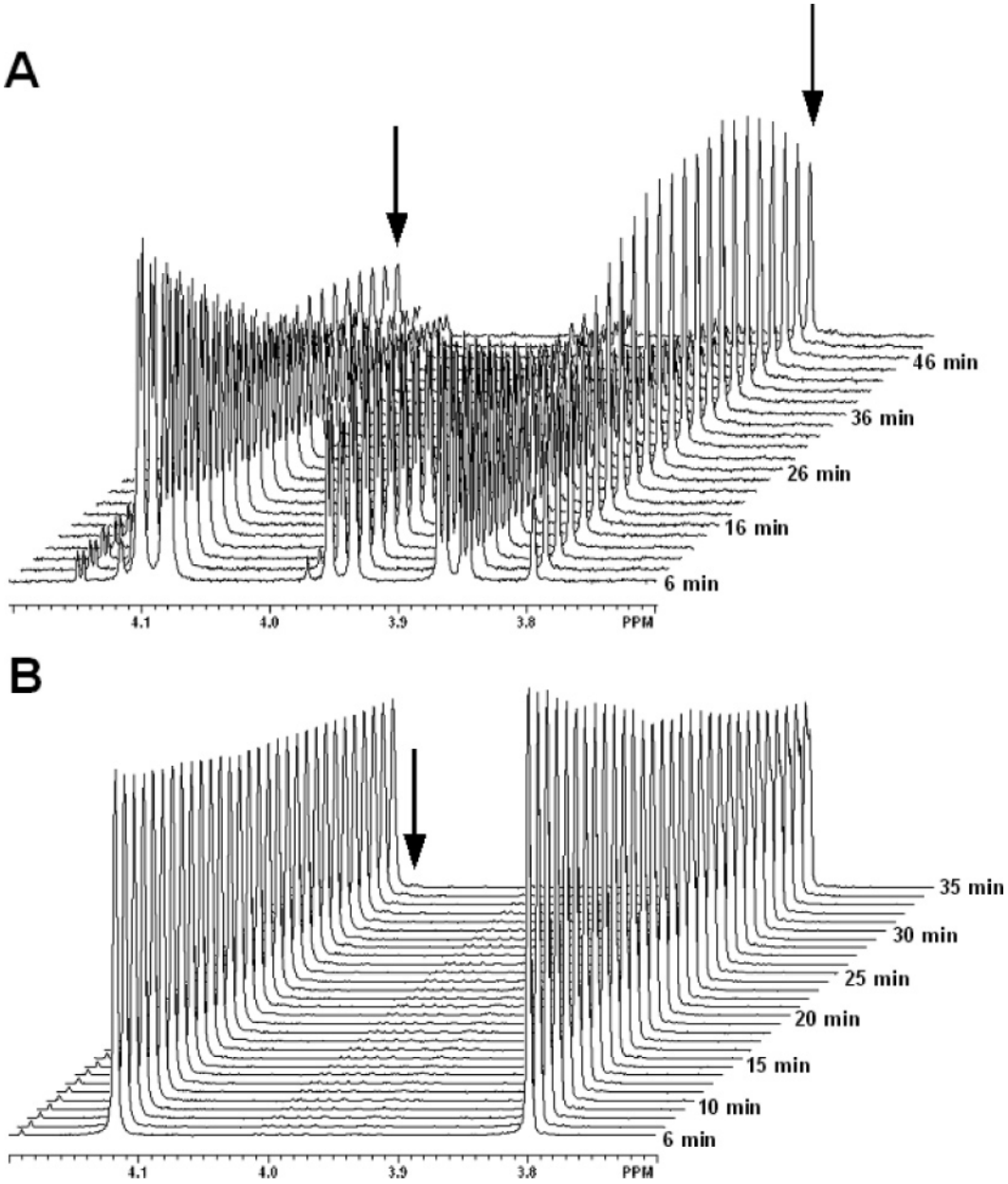


FIGURE 6: (A) ^{1}H NMR spectra of L-talarate, showing the TalrD-catalyzed epimerization to galactarate and dehydration to 5-keto-4-deoxy-D-glucarate. The appearance (TalrD-catalyzed epimerization) and subsequent diminishing (TalrD-catalyzed dehydration) peaks of galactarate are indicated by the arrows. (B) ^{1}H NMR spectra of galactarate, showing minimal GalrD-catalyzed epimerization to L-talarate. The appearance of the C5 proton of [2-^{2}H]-L-talarate (GalrD-catalyzed epimerization) is indicated by the arrow.

NMR spectroscopy, the lack of activity provides evidence for the importance of these functional groups in the dehydration and epimerization reactions catalyzed by STM3697.

When L-talarate was incubated with the wild type TalrD/GalrD and the reaction was monitored using ^{1}H NMR spectroscopy, the intensity of the resonance associated with the α-proton (on C2) of L-talarate decreased as the dehydration product was formed. No exchange of this proton with solvent deuterium was visible: the resonance associated with the proton on C3 (a doublet of doublets) did not collapse to a doublet as the reaction proceeded (Figure 7, panel A). The absence of exchange suggests that the L-talarate-specific base is the monoprotic His 328 at the end of the seventh β-strand. As the reaction proceeded, monodeuteriated galactarate was formed, as expected if the epimerization reaction is accomplished by delivery of a solvent-derived deuteron from the opposite face of the active site, i.e., Lys 197.

Consistent with the expectation that Lys 197 is the galactarate-specific base, we observed exchange of the α-proton of galactarate with solvent deuterium (Figure 7, panel B). As a result of exchange of the proton of galactarate, an upfield-shifted resonance associated with the proton on C3 of [2-^{2}H]-galactarate was observed. The appearance of this resonance is explained by abstraction of the C2 proton by Lys 197 to generate a Mg^{2+}-stabilized enolate intermediate; this intermediate partitions between vinylogous β-elimination (dehydration) and deuteriation by the conjugate acid of the Lys 197 base to yield [2-^{2}H-]-galactarate. Because the rate of vinylogous β-elimination is significantly greater than protonation of the opposite face to give the epimeric L-talarate, the resonances associated with L-talarate are barely observable.

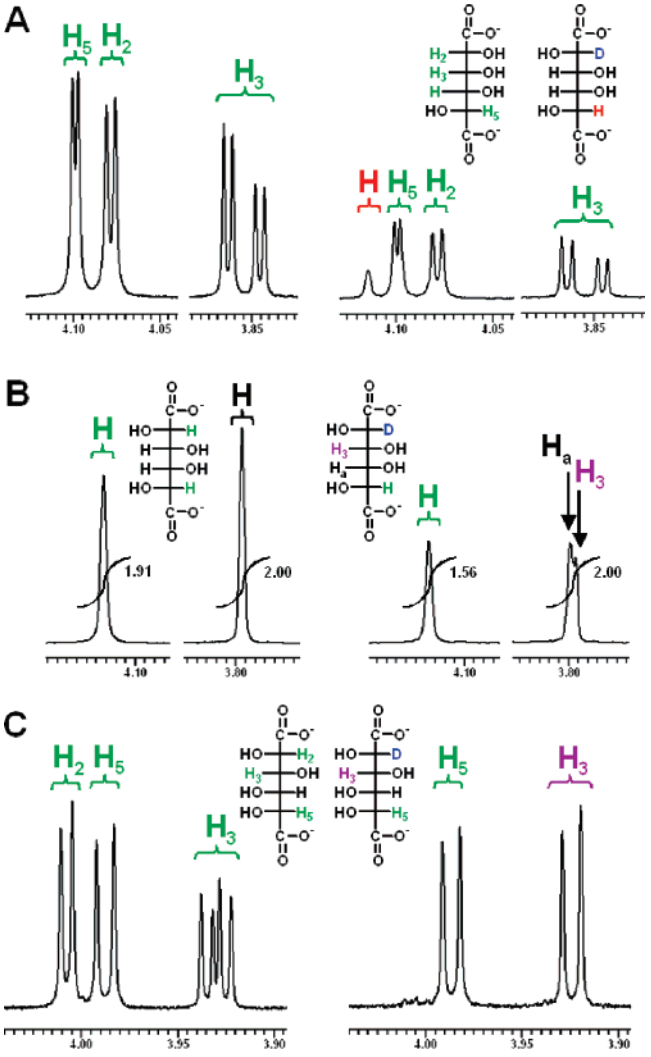


FIGURE 7: Representative partial ^{1}H spectrum of (A) L-talarate, (B) galactarate, and (C) L-glucarate after incubation with TalrD/GalrD in D_2O buffer. The peaks associated with the proton of C3 of [2S-^{2}H]-substrate (B and C) are indicated by magenta H_3.

We also examined the ability of TalrD/GalrD to catalyze exchange of the α-proton of other members of our acid sugar library. Exchange was only observed for L-glucarate (Figure 7, panel C). L-Glucarate (the 4-epimer of galactarate) is a substrate for neither dehydration nor epimerization. However, the rate constant for exchange of the α-proton of L-glucarate with solvent deuterium (5 s^{-1}) exceeds the values for k_{cat} for dehydration of galactarate. This exchange reaction provides additional evidence that Lys 197 is the galactarate-specific base. On the basis of the structure determined for the K197A mutant complexed with L-glucarate (Figure 4, panel B), no obvious explanation is available for the inability of the wild type enzyme to catalyze either the dehydration of L-glucarate or its epimerization to L-mannarate. The 5-OH group does not participate in any hydrogen bonding interactions with the active site, and the structure of the complex, relative to that determined for that with L-lyxarohydroxamate, does not reveal any changes in the orientation of the L-glucarate ligand vis-à-vis the acid/base groups in the active site.

Mechanism of the TalrD/GalrD-Catalyzed Reaction: Stereochemical Course of the Dehydration Reaction. We determined the stereochemical course of replacement of the 3-OH groups of both L-talarate and galactarate with solvent-derived hydrogen so that we could determine the relative orientation of the leaving group and the general acid that catalyzes ketonization of the enol product formed by vinylogous β-elimination of water to yield the 5-keto-4-deoxy-D-glucarate product. The product exists as a mixture of α- and β-furanosyl hemiketals (Figure 8, panel A) as ascertained by both ^{1}H and ^{13}C NMR spectroscopies.

From the magnitude of the vicinal C3−C4 ^{1}H−^{1}H coupling constants as well as NOE measurements, we determined that the larger coupling constant is associated with the 3-proS hydrogen. When both galactarate and L-talarate were dehydrated in D_2O-containing buffer, one prochiral hydrogen of C3 of each anomer was stereospecifically deuteriated in the 3-proR position (Figure 8, panel B). Thus, ketonization of the enol intermediate derived from vinylogous β-elimination is enzyme-catalyzed, and the departing 3-OH group is replaced with solvent deuterium with retention of configuration.

The hydrogen bonding interactions of the 3-OH leaving group with both His 328, the L-talarate specific base, and Lys 82 suggest that these cooperate in both the departure of the leaving group and the subsequent protonation of C3 to generate the 2-keto-3-deoxy dehydration product. In the active site of TarD, His 322 at the end of the seventh β-strand and Lys 102 in the symmetry related polypeptide were hydrogen bonded to the 3-OH leaving group, although in this case the ketonization of the enol intermediate derived from dehydration was not enzyme-catalyzed (*10*).

Mechanism of the TalrD/GalrD Reaction. On the basis of the mechanistic evidence presented in this article, the dehydration of galactarate catalyzed by TalrD/GalrD is initiated by abstraction of the 2-proton by Lys 197 to generate an enolate intermediate that partitions between vinylogous β-elimination, reprotonation to re-form the substrate, and reprotonation on the opposite face to form the epimeric product L-talarate (Figure 9). The dehydration of L-talarate catalyzed by TalrD/GalrD is initiated by abstraction of the 2-proton by His 328, in the His 328-Asp 301 hydrogen-bonded dyad, to generate the enolate intermediate that partitions between vinylogous β-elimination and reprotonation on the opposite face to form the epimeric product galactarate. His 328 and Lys 82 are both positioned to assist departure of the 3-OH group by acid catalysis and, also, deliver a proton to C3 to generate the 2-keto-3-deoxy product in the final partial reaction.

Conclusions: The Limitations of Functional Discovery by Library Screening. We used a library of acid sugars followed by physiological experiments to assign the promiscuous TalrD/GalrD function to a member of the MR subgroup encoded by the *S. typhimurium* LT2 genome. Our conclusion that L-talarate is a carbon source for *S. typhimurium* LT2 and that the TalrD function discovered by library screening is responsible for its catabolism highlights a significant problem in functional assignment of unknown proteins discovered in genome projects, including members of the enolase superfamily.

The KEGG database (http://www.genome.ad.jp/kegg/ligand.html) includes >14000 compounds and assigns many of these to metabolic pathways; L-talarate is not included in this database. A search of the literature using PubMed and Chemical Abstracts also fails to identify any references to

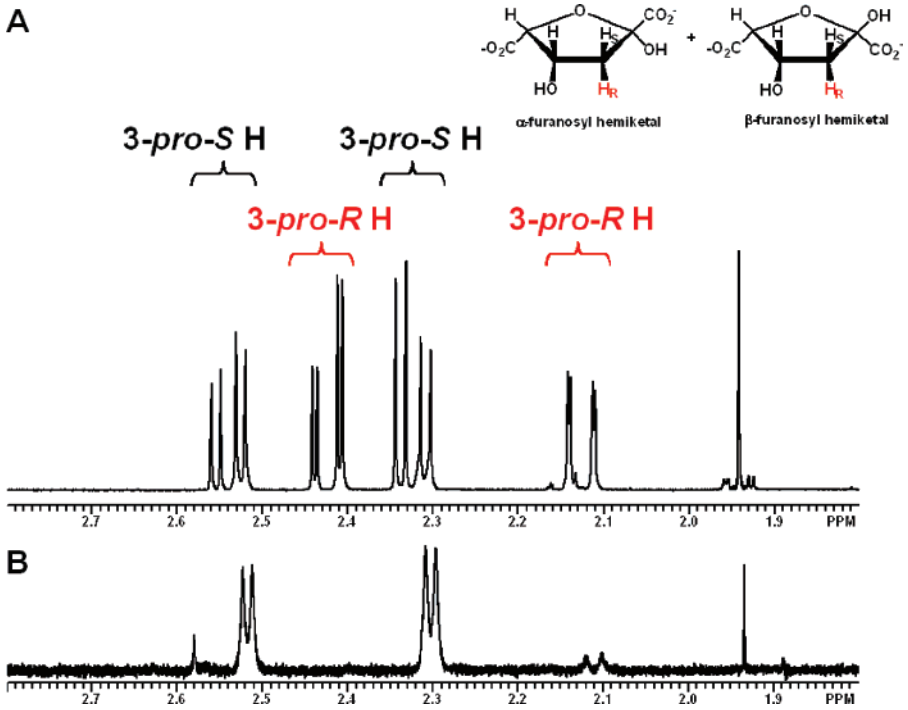


FIGURE 8: Representative partial ^{1}H NMR spectrum of 5-keto-4-deoxy-D-glucarate, (A) obtained by dehydration of either L-talarate or galactarate by TalrD/GalrD in protio, showing resonances associated with the 3-*proS* and 3-*proR* protons; (B) obtained by dehydration of either L-talarate or galactarate by TalrD/GalrD in D_2O buffer.

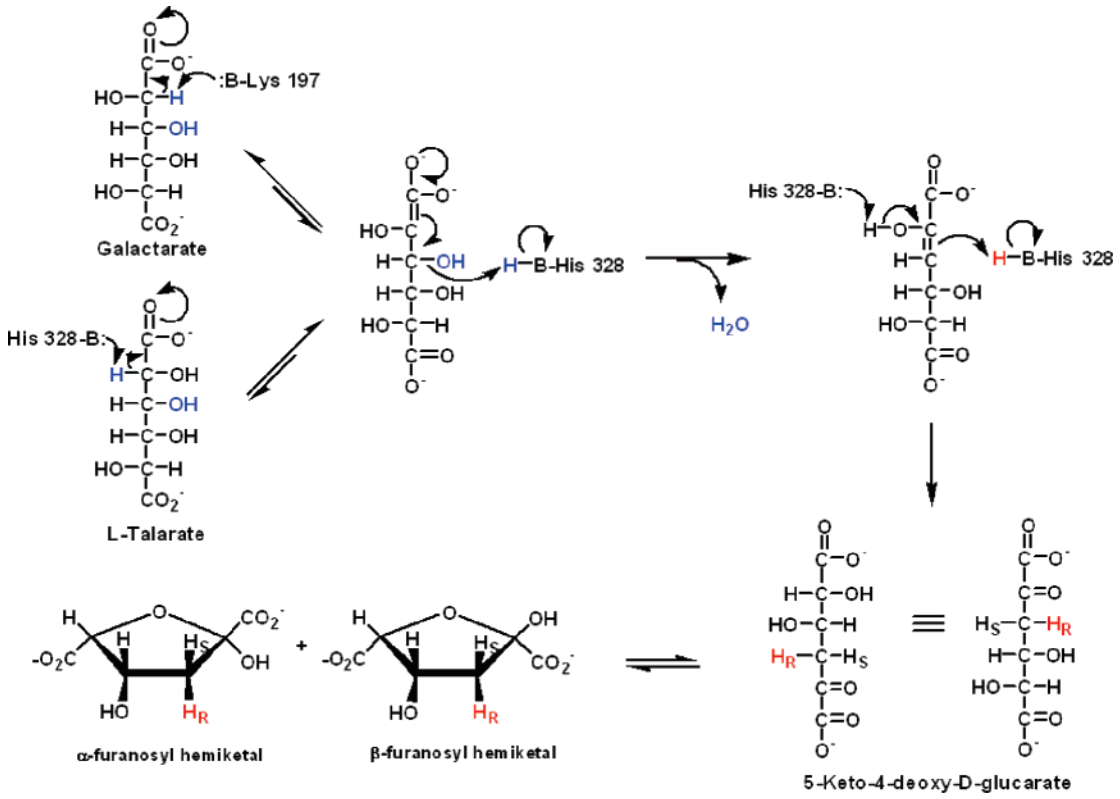


FIGURE 9: Proposed mechanism of the TalrD/GalrD-catalyzed reaction.

L-talarate as a carbon source or metabolite. Thus, the TalrD function we discovered provides a new metabolite and a new metabolic pathway. But, the only reason we made these discoveries is that we assembled a library of acid sugars so that it would contain all of the mono- and diacids derived from all of the D- and L-enantiomers of hexoses, pentoses, and tetroses. If we had restricted our screening to known metabolites, the TalrD function likely would have been missed (perhaps we would have noticed that the GalrD function of STM3697 is accompanied by slow epimerization to L-talarate, although this is unfavorable as shown in Figure 6, panel B). Indeed, as implied earlier in this article, we have used our acid sugar library to screen unknown members from the enolase superfamily thought to be acid sugar dehydratases on the basis of operon context and have failed to identify functions. The simplest explanation is that our library does

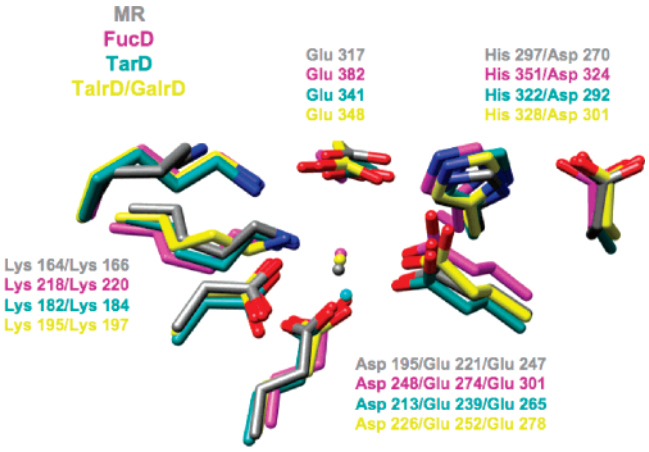


FIGURE 10: Superpositions of the active sites of MR (gray), FucD (magenta), TarD (cyan), and TalrD/GalrD (yellow).

not contain all physiologically relevant acid sugars, i.e., the metabolome of acid sugars is both large and uncertain, suggesting that the successful use of screening approaches will require considerable efforts in organic synthesis. Indeed, this realization highlights the importance of in silico algorithms for predicting the substrate specificities of unknown members of the enolase superfamily.

Conclusions: Evolution of Function in the Enolase Superfamily. The mechanism of the reaction catalyzed by MR has been the intellectual paradigm for understanding the structural bases of mechanistic diversity in the enolase superfamily. MR catalyzes only the interconversion of the enantiomers of mandelate by a 1,1-proton transfer reaction.

We have discovered three homologues of MR that catalyze the dehydration of different acid sugars, FucD (*9*), TarD (*10*), and, now, TalrD/GalrD. Superpositions of the structures of the active sites of MR and the three dehydratases reveal a remarkable overlay of the positions of the acid/base Lys residue at the end of the second β-strand as well as the His-Asp dyad at the ends of the seventh and sixth β-strands, respectively (Figure 10). Thus, as in the functionally divergent members of the MLE subgroup of the enolase superfamily, the active sites of these homologous members of the MR subgroup are "hard-wired" for acid/base chemistry (*23*, *24*).

MR catalyzes only a 1,1-proton transfer reaction. TarD catalyzes only the dehydration of D-tartrate, with no detectable epimerization to meso-tartrate. FucD catalyzes the dehydration of L-fuconate, L-galactonate (6-OH L-fuconate), and D-arabinonate (6-desmethyl L-fuconate) but also the much slower dehydration of L-talonate (the 2-epimer of L-galactonate) and D-ribonate (the 2-epimer of D-arabinonate). The latter reactions are accompanied by epimerization to L-galactonate and D-arabinonate, respectively; given the slow rates, these cannot be physiologically important but reflect the evolutionary potential of the active site for new functions. For TalrD/GalrD, we discovered that L-talarate is dehydrated and epimerized to galactarate at competitive rates, suggesting that its active site has evolved to catalyze proton abstraction of 2-epimeric substrates, thereby allowing both to be metabolites.

However, despite these functional differences, the positions of the Lys and His-Asp dyad acid/base catalysts are conserved. The active sites of TarD and TalrD/GalrD, but not FucD, contain an additional Lys residue that also is hydrogen bonded to the 3-OH leaving group (Lys 102 from the symmetry related polypeptide in TarD and Lys 82 in TalrD/GalrD). This Lys residue cannot be uniquely associated with either catalysis of departure of the leaving group (missing in FucD) or delivery of a proton to C3 to generate the dehydration product (protonation is random in TarD but stereospecific in both FucD and TalrD/GalrD). Thus, the reactivities of the enolate anion intermediate for vinylogous β-elimination vis-à-vis protonation and of the enol intermediate derived from elimination for protonation on C3 apparently are influenced by "remote" effects.

This conclusion implies that the evolution of new functions is more complicated than simply adding or deleting functional groups. The structural factors that modulate the observed range of functional promiscuity remain to be discovered but are important for understanding the relationships between protein structure and function.

REFERENCES

1. Babbitt, P. C., Mrachko, G. T., Hasson, M. S., Huisman, G. W., Kolter, R., Ringe, D., Petsko, G. A., Kenyon, G. L., and Gerlt, J. A. (1995) A functionally diverse enzyme superfamily that abstracts the alpha protons of carboxylic acids, *Science 267*, 1159−1161.
2. Gerlt, J. A., and Babbitt, P. C. (2001) DIVERGENT EVOLUTION OF ENZYMATIC FUNCTION: Mechanistically Diverse Superfamilies and Functionally Distinct Suprafamilies, *Annu. Rev. Biochem. 70*, 209−246.
3. Gerlt, J. A., Babbitt, P. C., and Rayment, I. (2005) Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity, *Arch. Biochem. Biophys. 433*, 59−70.
4. Schmidt, D. M., Hubbard, B. K., and Gerlt, J. A. (2001) Evolution of Enzymatic Activities in the Enolase Superfamily: Functional Assignment of Unknown Proteins in Bacillus subtilis and Escherichia coli as L-Ala-D/L-Glu Epimerases, *Biochemistry 40*, 15707−15715.
5. Sakai, A., Xiang, D. F., Xu, C., Song, L., Yew, W. S., Raushel, F. M., and Gerlt, J. A. (2006) Evolution of enzymatic activities in the enolase superfamily: N-succinylamino acid racemase and a new pathway for the irreversible conversion of D- to L-amino acids, *Biochemistry 45*, 4455−4462.
6. Palmer, D. R., and Gerlt, J. A. (1996) Evolution of Enzymatic Activities: Multiple Pathways for Generating and Partitioning a Common Enolic Intermediate by Glucarate Dehydratase from Pseudomonas putida, *J. Am. Chem. Soc. 118*, 10323−10324.
7. Lamble, H. J., Milburn, C. C., Taylor, G. L., Hough, D. W., and Danson, M. J. (2004) Gluconate dehydratase from the promiscuous Entner-Doudoroff pathway in Sulfolobus solfataricus, *FEBS Lett. 576*, 133−136.
8. Ahmed, H., Ettema, T. J., Tjaden, B., Geerling, A. C., van der Oost, J., and Siebers, B. (2005) The semi-phosphorylative Entner-Doudoroff pathway in hyperthermophilic archaea: a re-evaluation, *Biochem. J. 390*, 529−540.
9. Yew, W. S., Fedorov, A. A., Fedorov, E. V., Rakus, J. F., Pierce, R. W., Almo, S. C., and Gerlt, J. A. (2006) Evolution of Enzymatic Activities in the Enolase Superfamily: L-Fuconate Dehydratase from Xanthomonas campestris, *Biochemistry 43*, 14582−14597.
10. Yew, W. S., Fedorov, A. A., Fedorov, E. V., Wood, B. M., Almo, S. C., and Gerlt, J. A. (2006) Evolution of Enzymatic Activities in the Enolase Superfamily: D-Tartrate Dehydratase from Bradyrhizobium japonicum, *Biochemistry 43*, 14598−14608.
11. Salmon, L., Prost, E., Merienne, C., Hardre, R., and Morgant, G. (2001) A convenient preparation of aldonohydroamic acids in water and crystal structure of L-erythronohydroxamic acid, *Carbohydr. Res. 335*, 195−204.
12. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, in *Methods in Enzymology* (Carter, C. W. J., Sweet, R. M., Abelson, J. N., and Simon, M. I., Eds.) pp 307−326, Academic Press, New York.
13. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 55*, 849−861.

14. Terwilliger, T. C. (2000) Maximum-likelihood density modification, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 56*, 965−972.
15. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol. 6*, 458−463.
16. Jones, A. T. (1985) Interactive computer graphics: FRODO, *Methods Enzymol. 115*, 157−171.
17. Brunger, A. T., Adams, P. D., Clore, G. M., DeLAno, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D54*, 905−921.
18. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005) Likelihood-enhanced fast translation functions, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 61*, 458−464.
19. Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products, *Proc. Natl. Acad. Sci. U.S.A. 97*, 6640−6645.
20. Wycuff, D. R., and Matthews, K. S. (2000) Generation of an AraC-araBAD promoter-regulated T7 expression system, *Anal. Biochem. 277*, 67−73.
21. Hubbard, B. K., Koch, M., Palmer, D. R., 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.
22. Kalyanaraman, C., Bernacki, K., and Jacobson, M. P. (2005) Virtual screening against highly charged active sites: identifying substrates of alpha-beta barrel enzymes, *Biochemistry 44*, 2059−2071.
23. Schmidt, D. M., Mundorff, E. C., Dojka, M., Bermudez, E., Ness, J. E., Govindarajan, S., Babbitt, P. C., Minshull, J., and Gerlt, J. A. (2003) Evolutionary potential of $(\beta/\alpha)_8$-barrels: functional promiscuity produced by single substitutions in the enolase superfamily, *Biochemistry 42*, 8387−8393.
24. Vick, J. E., Schmidt, D. M., and Gerlt, J. A. (2005) Evolutionary Potential of $(\beta/\alpha)_8$-Barrels: In Vitro Enhancement of a "New" Reaction in the Enolase Superfamily, *Biochemistry 44*, 11722−11729.

BI7008882