BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Discovery of Function in the Enolase Superfamily: D-Mannonate and D-Gluconate Dehydratases in the D-Mannonate Dehydratase Subgroup

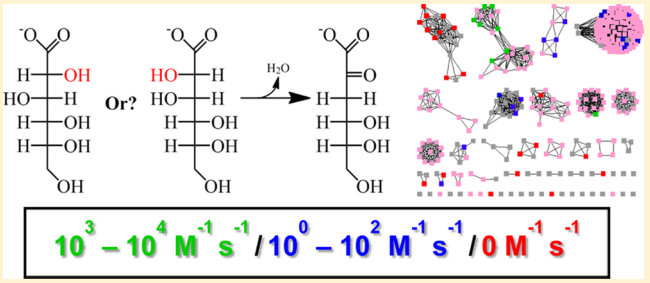
Daniel J. Wichelecki,[§] Bryan M. Balthazor,[§] Anthony C. Chau,[§] Matthew W. Vetting,[‡] Alexander A. Fedorov,[‡] Elena V. Fedorov,[‡] Tiit Lukk,[§] Yury V. Patskovsky,[‡] Mark B. Stead,[‡] Brandan S. Hillerich,[‡] Ronald D. Seidel,[‡] Steven C. Almo,[‡] and John A. Gerlt*[,§]

[§]Departments of Biochemistry and Chemistry and Institute for Genomic Biology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States

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

Supporting Information

ABSTRACT: The continued increase in the size of the protein sequence databases as a result of advances in genome sequencing technology is overwhelming the ability to perform experimental characterization of function. Consequently, functions are assigned to the vast majority of proteins via automated, homology-based methods, with the result that as many as 50% are incorrectly annotated or unannotated (Schnoes et al. *PLoS Comput. Biol.* 2009, *5* (12), e1000605). This manuscript describes a study of the D-mannonate dehydratase (ManD) subgroup of the enolase superfamily (ENS) to investigate how function diverges as sequence diverges. Previously, one member of the subgroup had been experimentally characterized as ManD [dehydration of D-mannonate to 2-keto-3-deoxy-D-mannonate (equivalently, 2-keto-3-deoxy-D-gluconate)]. In this study, 42 additional members were characterized to sample sequence−function space in the ManD subgroup. These were found to differ in both catalytic efficiency and substrate specificity: (1) high efficiency (k_{cat}/K_M = 10^3 to 10^4 M^{-1} s^{-1}) for dehydration of D-mannonate, (2) low efficiency (k_{cat}/K_M = 10^1 to 10^2 M^{-1} s^{-1}) for dehydration of D-mannonate and/or D-gluconate, and 3) no-activity with either D-mannonate or D-gluconate (or any other acid sugar tested). Thus, the ManD subgroup is not isofunctional and includes D-gluconate dehydratases (GlcDs) that are divergent from the GlcDs that have been characterized in the mandelate racemase subgroup of the ENS (Lamble et al. *FEBS Lett.* 2004, *576*, 133−136) (Ahmed et al. *Biochem. J.* 2005, *390*, 529−540). These observations signal caution for functional assignment based on sequence homology and lay the foundation for the studies of the physiological functions of the GlcDs and the promiscuous ManDs/GlcDs.



The massive influx of sequence data since the first bacterial genome sequence was published in 1995 has necessitated a reliance on homology-based annotations of protein function.[1,2] However, because this method assigns the function of the "closest" homologue, an estimated 30−50% of the functional annotations in the databases are incorrect,[3−5] with the magnitude of the problem increasing as the incorrect annotations are propagated in assigning functions to proteins discovered in newly sequenced genomes. In a study of several functionally diverse superfamilies, Schnoes, Babbitt, and co-workers concluded that 85% of misannotations resulted from annotations that are more detailed than justified.[3] Automated methods often are able to achieve high degrees of accuracy in the transfer of the first three Enzyme Commission (EC) code numbers, but accurate transfer of the fourth EC code number (substrate specificity) is much more difficult.[6] This study examines the D-mannonate dehydratase (ManD) subgroup of the enolase superfamily (ENS) to determine experimentally, on a large scale, how function diverges as sequence diverges in highly homologous enzymes. Our results illustrate the difficulty of accurately assigning function via homology-based methods and, also, provide insights into how different functions can arise in highly homologous enzymes.

Two conserved features are shared by members of the ENS: mechanism and structure. The mechanism is general base-catalyzed abstraction of a proton alpha to a carboxylate group of the substrate to form an enediolate intermediate.[7] The enediolate intermediate is stabilized by coordination to an active site divalent metal cation (usually Mg^{2+}). Furthermore, members of the ENS share a common structural motif: an (α + β) capping domain that contains the residues that determine substrate specificity and a modified TIM-barrel domain ((β/

Received: March 3, 2014
Revised: April 2, 2014
Published: April 4, 2014

   dx.doi.org/10.1021/bi500264p | *Biochemistry* 2014, 53, 2722−2731

$\alpha)_7\beta$-barrel) that contains the residues that mediate acid/base chemistry.[8,9] Subgroups are differentiated by the conserved metal-binding residues at the ends of the third, fourth, and fifth β-strands as well as conserved catalytic acid/base residues at the ends of the second, third, sixth, and/or seventh β-strands of the modified TIM-barrel domain. The ENS is particularly interesting because its members share a common mechanism and the same structural motif but are functionally diverse (e.g., β-elimination and 1,1-proton transfer reactions).[10,11] Thus, the ENS is a good model to investigate how function diverges as sequence diverges.

In 2007, Rakus and co-workers discovered the ManD subgroup of the ENS.[12] In that study, a protein from *Novosphingobium aromaticivorans* (*Na*ManD) was structurally characterized and discovered to catalyze the *syn*-dehydration of D-mannonate to 2-keto-3-deoxy-D-mannonate (equivalently, 2-keto-3-deoxy-D-gluconate) (EC 4.2.1.8 and Unitprot ID A4XF23). The Mg^{2+}-binding residues located at the ends of the third, fourth, and fifth β-strands are Asp 210, Glu 236, and Glu 262, respectively. The general base that abstracts the 2-proton is the Tyr 159-Arg 147 dyad in the "150−180s" loop between the second and third β-strands; the general acid that facilitates departure of the 3-hydroxyl group is His 212 located at the end of the third β-strand[12] (Figure 1). His 315, located at

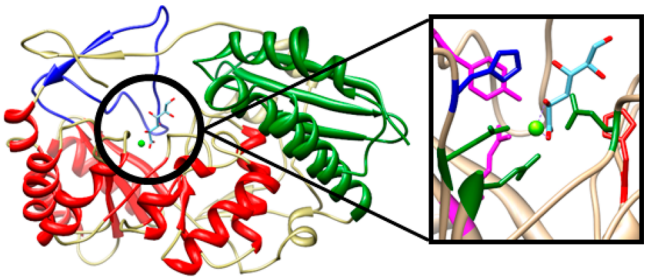
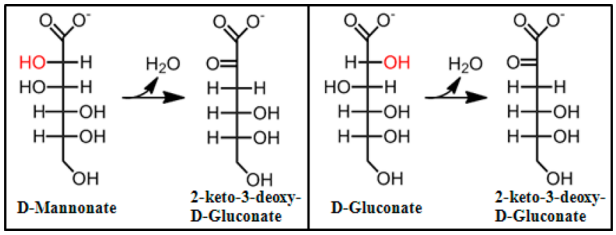


Figure 1. Denotes the structural features of the ManDs (PDB 2QJJ). On the left, the "150−180s" loop (blue), TIM barrel (red), and capping domain (green) are displayed. The right inset shows the active site residues: metal binding Asp210, Glu236, Glu262 (green); Tyr157-Arg149 catalytic dyad (magenta); acidic His212 (blue); and conserved His315 at the end of the 7th β-strand (red). The D-mannonate ligand from the 2QJM structure is shown in light blue.

the end of the seventh β-strand that is hydrogen-bonded to the 5-hydroxyl group of the D-mannonate substrate, is also conserved. Since this study, no other members of the ManD subgroup have been experimentally characterized. Given the conservation of metal-binding residues, catalytic residues, and active site architecture, the assumption was that the entire subgroup is isofunctional.

In this study, we sought to investigate how function diverges as sequence diverges within this subgroup and to determine if the ManD subgroup is, in fact, isofunctional. When the target proteins for this study were selected in April 2011, the ManD subgroup included *Na*ManD[12] and 299 uncharacterized proteins that share ≥35% sequence identity [Structure−Function Linkage Database (http://sfld.rbvi.ucsf.edu/)]. [At the time of submission of this manuscript, the UniProtKB database contained the sequences for 2919 members of the ManD subgroup.] Forty-three members representing the breadth of sequence−function space were produced as soluble proteins and screened for activity using a library of acid sugars. Surprisingly, we found that the ManD subgroup is *not* isofunctional; instead, in addition to ManDs it also contains D-gluconate dehydratases (GlcDs) that catalyze the *anti*-dehydration of D-gluconate to 2-keto-3-deoxy-D-gluconate as well as promiscuous proteins that catalyze both the ManD and GlcD reactions (Scheme 1). In addition, a wide range of

Scheme 1



catalytic efficiencies (values of k_{cat}/K_M) were discovered. Using sequence similarity networks (SSNs),[13] the members with these divergent functions could be separated into isofunctional clusters. Furthermore, 16 unique crystal structures (for a total of 36 unliganded and liganded structures) were solved to survey sequence and structure space; these revealed conserved active site structures but divergent conformations for the "150−180s" loops that contain the general basic Tyr-Arg dyads and close over the active site to sequester the substrate from solvent. Taken together, the functional and structural data provide a comprehensive description of how *in vitro* function diverges as sequence diverges.

MATERIALS AND METHODS

Cloning, Expression, and Purification of Targets (AECOM). pNIC28-BSA4-based expression vectors were transformed into BL21(DE3) *Escherichia coli* containing the pRIL plasmid (Stratagene) and used to inoculate a 10 mL 2xYT culture containing 25 μg/mL kanamycin and 34 μg/mL chloramphenicol. The cultures were allowed to grow overnight at 37 °C in a shaking incubator. The overnight culture was used to inoculate 2 L of PASM-5052 autoinduction media.[14] The culture was placed in a LEX48 airlift fermenter and incubated at 37 °C for 4 h and then at 22 °C overnight. The culture was harvested and pelleted by centrifugation.

Cells were resuspended in lysis buffer (20 mM HEPES, pH 7.5, 500 mM NaCl, 20 mM imidazole, and 10% glycerol) and lysed by sonication. The lysate was clarified by centrifugation at 35000*g* for 30 min. The protein was purified using an AKTAxpress FPLC (GE Healthcare). The lysate was loaded onto a 1 mL His60 column (Clontech), washed with 10 column volumes of lysis buffer, and eluted with buffer containing 20 mM HEPES, pH 7.5, 500 mM NaCl, 500 mM Imidazole, and 10% glycerol. This partially purified protein was loaded onto a HiLoad S200 16/60 PR gel filtration column equilibrated with SECB buffer (20 mM HEPES, pH7.5, 150 mM NaCl, 10% glycerol, and 5 mM DTT). The protein was analyzed by SDS-PAGE, flash frozen in liquid nitrogen, and stored at −80°C.

Expression and Purification of N-Terminal His-Tagged Proteins (UIUC). Genes in pET15b (Novagen) were expressed in *E. coli* strain BL21. Small-scale cultures were grown at 37 °C for 18 h in 5 mL of LB containing 100 μg/mL ampicillin and used to inoculate 1 L LB containing 100 μg/mL ampicillin. IPTG (500 μM) was added at $OD_{600\ nm}$ = 0.6−0.8 to induce expression. The induced cultures then were grown for an

dx.doi.org/10.1021/bi500264p | *Biochemistry* 2014, 53, 2722−2731

additional 18 h at 37 °C. The cells were harvested by centrifugation at 5000 rpm for 10 min and resuspended in 70 mL of binding buffer (6 mM imidazole, 20 mM Tris-HCl, pH 7.9, 5 mM $MgCl_2$, and 500 mM NaCl). The resuspended cells were lysed by sonication and centrifuged at 17 000 rpm for 30 min. The supernatant was loaded onto a column of 50 mL chelating Sepharose Fast Flow (Amersham Biosciences) charged with Ni^{2+} and eluted with a linear gradient of imidazole (0−1 M over 600 mL). Fractions were analyzed using SDS-PAGE. Fractions containing protein at high purity (>90%) were combined and dialyzed against 4 L of buffer containing 100 mM imidazole, 20 mM Tris-HCl, pH 7.9, 10 mM $MgCl_2$, 150 mM NaCl, and 10% glycerol for 2 h at 4 °C. The protein was dialyzed in this manner a total of three times. Finally, the protein was concentrated to a maximum of ∼10 mg/mL (depending on solubility) and flash-frozen using liquid nitrogen and stored at −80 °C.

Expression and Purification of Tagless ManD Constructs (UIUC). The genes in pET17b (Novagen) were expressed in *E. coli* strain BL21. Small-scale cultures were grown at 37 °C for 18 h in 5 mL of LB containing 100 μg/mL ampicillin and used to inoculate 1 L LB containing 100 μg/mL ampicillin. The 1 L cultures were grown for an additional 18 h at 37 °C without induction. The cells were harvested by centrifugation at 5000 rpm for 10 min and resuspended in 70 mL of binding buffer. The resuspended cells were lysed by sonication and centrifuged at 17 000 rpm for 30 min. The supernatant was loaded onto a 300 mL DEAE Sepharose column (Amersham Biosciences) and eluted with a linear gradient of NaCl (0−1 M over 1.6 L) in 10 mM Tris-HCl, pH 7.9, containing 5 mM $MgCl_2$. Fractions containing the protein of interest at high purity were combined and dialyzed against 4 L buffer containing 10 mM Tris-HCl, pH 7.9, and 5 mM $MgCl_2$ for 2 h at 4 °C. The dialyzed protein was then loaded onto a 30 mL Q-Sepharose column (Amersham Biosciences) and eluted with a linear gradient of NaCl (0−1 M over 500 mL) in 10 mM Tris-HCl, pH 7.9, containing 5 mM $MgCl_2$. Fractions containing the protein of interest at high purity were combined and dialyzed in 4 L buffer containing 10 mM Tris-HCl, pH 7.9, containing 5 mM $MgCl_2$ for 2 h at 4 °C. Ammonium sulfate was added to a final concentration of 1 M, and the sample was loaded onto a 30 mL phenylsepharose column (Amersham Biosciences). The protein was eluted with a gradient of ammonium sulfate (1−0 M over 500 mL) in 10 mM Tris-HCl, pH 7.9, containing 5 mM $MgCl_2$. Fractions with pure protein (SDS-PAGE) were combined and dialyzed against 4 L buffer containing 10 mM Tris-HCl, pH 7.9, 5 mM $MgCl_2$, 150 mM NaCl, and 10% glycerol for 2 h at 4 °C. Finally, the protein was concentrated to a maximum of ∼10 mg/mL (depending on solubility) and flash frozen in liquid nitrogen and stored at −80 °C.

Screen for Dehydration. Reactions to test for enzymatic activity were performed in acrylic, UV transparent 96-well plates (Corning Incorporated) using a library of 72 acid sugars (Figure S1, Supporting Information). Reactions (60 μL) contained 50 mM HEPES, pH 7.9, 10 mM $MgCl_2$, 1 μM enzyme, and 1 mM of acid sugar substrate (blanks without enzyme). The plates were incubated at 30 °C for 16 h. A 1% semicarbazide reagent solution (240 μL) was added to each well and incubated for 1 h at room temperature. The absorbancies were measured at 250 nm (ε = 10 200 M^{-1} cm^{-1}) using a Tecan Infinite M200PRO plate reader.

Kinetic Assays. Dehydration of D-mannonate and D-gluconate was quantitated using either a discontinuous assay with the semicarbazide reagent[15,16] or a continuous, coupled-enzyme spectrophotometric assay. In the latter assay, the product was phosphorylated using 2-keto-3-deoxy-D-gluconate kinase (KdgK) and ATP; formation of ADP was measured using pyruvate kinase (PK) and L-lactate dehydrogenase (LDH). The assay (200 μL) at 25 °C contained 50 mM potassium HEPES, pH 7.5, 5 mM $MgCl_2$, 1.5 mM ATP, 1.5 mM PEP, 0.16 mM NADH, 9 units of PK, 9 units of LDH, 18 units of KdgK, and ManD/GlcD. Dehydration was quantitated by measuring the decrease in absorbance at 340 nm (ε = 6220 M^{-1} cm^{-1}). Low-activity enzymes were characterized using the discontinuous assay; high-activity ManDs were characterized using the coupled assay.

Site-Directed Mutagenesis (SDM). Mutants were constructed using primers designed with the Agilent Technologies online webserver (https://www.genomics.agilent.com/) and purchased from Bio-Synthesis, Inc. Forward and reverse primers containing the mutations of interest are listed in Table S1, Supporting Information. PCR reactions (30 μL) contained 1 mM $MgCl_2$, 1× Pfx Amp buffer, 0.33 mM dNTP, 0.33 μM of FOR/REV primer, and 1.25 units Pfx polymerase (Invitrogen Platinum Pfx DNA Polymerase kit). The templates were 50 ng ManD-containing pET15b (*Na*ManD) or pET17b (*Cs*ManD). Amplifications were performed according to the manufacturer's guidelines. After addition of DpnI (10 units), the reactions were incubated for 4 h at 37 °C. The DpnI-digested products were purified by gel electrophoresis, extracted, and transformed (electroporation, Bio-Rad Micropulsar Electroporator) into XL1 Blue competent cells. Finally, plasmids isolated from the transformants were sequenced to confirm the mutations.

Circular Dichroism of NaManD and CsManD Loop Mutants. The circular dichroism spectrum of a 10 μM solution of mutant enzyme in an optically clear borate buffer (50 mM boric acid, 100 mM KCl, 0.7 mM DTT, pH 8.0) was measured from 190 to 260 nm using a Jasco J-715 spectropolarimeter. Five replicate measurements were made.

Protein Crystallization and X-ray Diffraction Data Collection. Proteins were crystallized by the sitting-drop vapor diffusion method. The concentrated (usually 5−40 mg/mL) protein solutions (from 0.3 to 1 μL) were mixed with an equal volume of a precipitant solution and equilibrated at room temperature (∼294 K) against the same precipitant solution in clear tape-sealed 96-well INTELLI-plates (Art Robbins Instruments, Sunnyvale, CA). Crystallization was performed using either a TECAN crystallization robot (TECAN US, Research Triangle Park, NC) or a PHOENIX crystallization robot (Art Robbins Instruments) and four types of commercial crystallization screens: the WIZARD I&II screen (Emerald BioSystems, Bainbridge Island, WA); the INDEX HT and the CRYSTAL SCREEN HT (both from Hampton Research, Aliso Viejo, CA); and the MCSG screen (Microlytic, Woburn, MA). The appearance of protein crystals has been monitored either by visual inspection or using a Rock Imager 1000 (Formulatrix, Waltham, MA) starting within 24 h of incubation and again at weeks 1, 2, 3, 5, 8, and 12. Where necessary, the crystallization conditions were optimized manually using 24-well Cryschem sitting drop plates (Hampton Research). The crystallization conditions for all crystal structures are listed in the PDB methods tab and the Supporting Information. The crystals were either directly frozen in liquid nitrogen or treated with a

cryoprotectant (glycerol or ethylene glycol, 20−30%, vol/vol) before freezing.

The X-ray diffraction data for the frozen crystals were collected at 100 K on the beamline X29A (National Synchrotron Light Source, Brookhaven National Laboratory, Upton, NY) using a wavelength of 1.075 Å or on the beamline 31I-D (LRL-CAT, Advanced Photon Source, Argonne National Laboratory, IL, USA) using a wavelength of 0.9793 Å. The diffraction data were processed and scaled with SCALA[17] (APS data) or HKL[18] (NSLS data). The crystal structures reported here were determine by molecular replacement using coordinates for similar structures from the PDB (listed in each PDB deposition as REMARK 200: STARTING MODEL) and PHASER MR software (the CCP4 program package suite).[19] Each structure was refined using the programs REFMAC[20] or PHENIX,[21] and the resulting models were rebuilt manually using COOT visualization and refinement software.[22] The data collection and refinement statistics for all crystal structures are listed in the Table S2, Supporting Information.

RESULTS AND DISCUSSION

Selection of Targets. In April 2011, the Structure−Function Linkage Database (SFLD; sfld.rbvi.ucsf.edu/) contained 300 sequences for the ManD subgroup of the ENS (*Na*ManD and 299 uncharacterized homologues). These sequences were used to generate a sequence similarity network (SSN) with a BLASTP e-value threshold of 10^{-80} (~35% sequence identity) (Figure 2a).[13,23,24] As the BLASTP e-value threshold is decreased to 10^{-190}, the sequences segregate into several clusters sharing >70% sequence identity (Figure 2c).

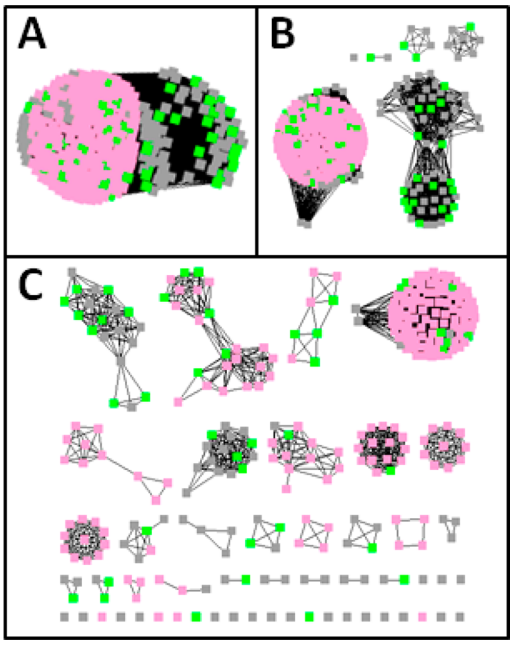


Figure 2. Sequence similarity networks (SSNs) of the ManD subgroup at several e-value thresholds to illustrate the effect of increasing stringency on clustering. Panel A, 10^{-80}, ~35% identity. Panel B, 10^{-120}, ~45% identity. Panel C, 10^{-190}, ~75% identity. Pink coloring indicates proteins predicted to be ManDs by the Structure Function Linkage Database. Green coloring indicates proteins that were purified and subjected to activity screening.

The genome neighborhoods of the genes encoding members of the ManD subgroup (±10 genes) were analyzed to aid in target selection for protein production and structure determination. The genome neighborhoods of some members encode 2-keto-3-deoxy-D-gluconate kinase (KdgK) and 2-keto-3-deoxy-D-gluconate-6-P aldolase (KdgA). KdgK and KdgA metabolize the 2-keto-3-deoxy-D-gluconate product of the ManD reaction to pyruvate and D-glyceraldehyde 3-phosphate, indicating a catabolic role for the proximal member of the ManD subgroup. Alternatively, for some members the genome neighborhoods lack these enzymes but contain, for example, dehydrogenases, suggesting divergent catalytic and metabolic functions. Targets for protein production by the Protein Core of the Enzyme Function Initiative (EFI; enzymefunction.org), functional characterization by the University of Illinois, and structure determination by the Structure Core of the EFI were chosen from both types of genome neighborhoods. A large number of targets (115) were chosen with the anticipation that not all targets would produce soluble, purified proteins.

Substrate Screening. Of the 115 targets, 42 were produced as soluble, purified proteins (sharing less than 95% sequence identity). The ManD SSN in Figure 2 highlights the diversity of purified proteins assayed. The proteins were screened for dehydration activity with a library of 72 acid sugars using a semicarbazide-based assay (Figure S1, Supporting Information).[25−27] The catalytically active proteins (24 of the 42 screened) utilize D-mannonate, D-gluconate, or both as substrates; no other hits were observed with the acid sugar library. Positive hits were verified with ^{1}H NMR spectra of the products (2-keto-3-deoxy-D-mannonate/2-keto-3-deoxy-D-gluconate) before the proteins were subjected to more in-depth analyses to determine kinetic constants. The ability of some members to catalyze the dehydration of D-gluconate was not expected (*vide infra*).

The kinetic characterizations revealed further unexpected divergence in function (Table 1). Among the newly characterized ManDs, seven dehydrate D-mannonate with catalytic efficiencies similar to that of *Na*ManD (k_{cat}/K_M = 10^3 to 10^4 M^{-1} s^{-1}). However, 16 targets showed low catalytic efficiencies (k_{cat}/K_M = 10^1 to 10^2 M^{-1} s^{-1}); 19 showed no detectable activity with any member of the acid sugar library. Three of the 12 proteins that dehydrate D-mannonate with low catalytic efficiency also dehydrate D-gluconate with low catalytic efficiency. Furthermore, 4 of the 23 targets with no activity on D-mannonate dehydrate D-gluconate. Thus, the functionally characterized members were assigned into three categories according to catalytic efficiency and substrate specificity: (1) high-activity (k_{cat}/K_M = 10^3 to 10^4 M^{-1} s^{-1}) and specific for D-mannonate; (2) low-activity (k_{cat}/K_M = 10^1 to 10^2 M^{-1} s^{-1}) and specific for either D-mannonate or D-gluconate or promiscuous for both; and (3) no-activity with either D-mannonate or D-gluconate (or any acid sugar in the library). The SSN constructed with a threshold of 10^{-190} (~75% sequence identity) segregates groups with different catalytic efficiencies and substrate specificities (Figure 3).

Divergence in Activity. Members with different *in vitro* activities are assumed to have different *in vivo* functions. Physiologically, the dehydration of D-mannonate to 2-keto-3-deoxy-D-mannonate is found in the D-glucuronate degradation pathway in which D-glucuronate is isomerized to 5-keto-D-mannonate and then reduced to D-mannonate. The D-mannonate is dehydrated, phosphorylated, and cleaved by an aldolase to form pyruvate and glyceraldehyde 3-phosphate.

dx.doi.org/10.1021/bi500264p | *Biochemistry* 2014, 53, 2722−2731

Table 1. Kinetic Parameters for Members of the ManD Subgroup

Cluster	Uniprot ID	D-mannonate k_{cat} (s^{-1})	D-mannonate k_{cat}/K_M (M^{-1} s^{-1})	D-gluconate k_{cat} (s^{-1})	D-gluconate k_{cat}/K_M (M^{-1} s^{-1})	end of 7th β-strand	UxuA?
1	A5KUH4					Pro	yes
1	C9NUM5					Pro	no
1	C9Y5D5					Pro	yes
1	D0KC90					Pro	yes
1	A4W7D6					Pro	yes
1	D0X4R4					Pro	yes
1	A6AMN2					Pro	yes
1	Q6DAR4					Pro	yes
1	C6DI84					Pro	yes
6	B8HCK2					Pro	no
9	C9CN91					Pro	yes
9	C8ZZN2					Pro	yes
10	C7PW26					Pro	no
Singleton	A6M2W4					Pro	yes
Singleton	Q2CIN0					Pro	yes
Singleton	A8RQK7					Pro	yes
Singleton	C6CVY9					Gly	yes
Singleton	C9A1P5					Pro	yes
Singleton	B5GCP6					Pro	no
3	A6VRA1	0.02 ± 0.001	160			Ala	yes
3	E1V4Y0	0.03 ± 0.006	20	0.05 ± 0.004	20	Pro	yes
3	B3PDB1	0.03 ± 0.002	100			Ala	yes
3	*Cs*ManD/Q1QT89	0.02 ± 0.0005	5	0.04 ± 0.006	40	Pro	yes
4	Q8FHC7	0.02 ± 0.001	10			Pro	yes
4	A4WA78	0.02 ± 0.002	30			Pro	yes
4	B1ELW6	0.02 ± 0.001	20			Pro	yes
4	D8ADB5	0.01 ± 0.002	30			Pro	yes
4	J7KNU2	0.01 ± 0.001	20			Pro	yes
4	B5RAG0	0.01 ± 0.001	50			Pro	yes
5	D4GJ14			0.04 ± 0.003	120	Pro	yes
5	B5R541			0.05 ± 0.003	80	Pro	yes
5	B5QBD4			0.02 ± 0.0005	150	Pro	yes
5	C6CBG9	0.04 ± 0.002	50	0.03 ± 0.002	60	Pro	yes
Singleton	D9UNB2	0.004 ± 0.001	60			Pro	yes
8	D7BPX0			0.01 ± 0.002	40	Pro	no
2	Q1NAJ2	2 ± 0.07	4200			Ala	no
2	Q9AAR4	1 ± 0.006	12300			Ala	no
2	Q9A4L8	0.65 ± 0.02	1200			Ala	no
2	B0T4L2	0.3 ± 0.01	1200			Ala	no
2	B0T0B1	2 ± 0.2	12100	0.003 ± 0.001	5	Ala	no
2	A5V6Z0	4 ± 0.2	2900	0.01 ± 0.001	10	Ala	no
2	*Na*ManD/A4XF23	1.3 ± 0.1	3200			Ala	no
7	G7TAD9	0.8 ± 0.03	4400			Ala	no

When this pathway was discovered in *E. coli* and *Erwinia carotovora*, dehydration of D-mannonate was found to be catalyzed by a dehydratase, UxuA, that is not a member of the ENS.[28,29] Therefore, the discovery that members of the ManD subgroup dehydrate D-mannonate with high catalytic efficiency implies convergent evolution of function in different superfamilies within the D-glucuronate catabolic pathway. Interestingly, the genomes of all of the organisms encoding high-activity ManDs lack the gene encoding UxuA; however, the genomes of the majority of organisms with low- or no-activity ManDs have a gene encoding UxuA (Table 1). This suggests that the high-activity ManDs perform the same role as UxuA in the encoding organisms; however, the low-activity members have a different metabolic function. In those organisms that encode low- or no-activity members of the ManD subgroup but no UxuA, growth on D-glucuronate likely is enabled by an alternate catabolic pathway, such as the uronate dehydrogenase or KduI pathway.[30,31] Therefore, members with low *in vitro* catalytic efficiencies likely have different *in vivo* metabolic functions, even if they can dehydrate D-mannonate. In work that will be described elsewhere, we are characterizing some of these divergent physiological functions.

D-Gluconate Dehydration. Although, in retrospect, the discovery of D-gluconate as a substrate for some members is not surprising because D-gluconate and D-mannonate are epimers at carbon-2 so they yield the same dehydration product, this stereochemical difference requires that a base other than the Tyr-Arg dyad in the "150−180s" loop abstract the 2-proton from D-gluconate. To investigate which residue could function as the D-gluconate specific base, D-mannonate and D-gluconate were modeled into the active site of the member from *Chromohalobacter salexigens* (*Cs*ManD) (Uniprot ID Q1QT89)

dx.doi.org/10.1021/bi500264p | *Biochemistry* 2014, 53, 2722−2731

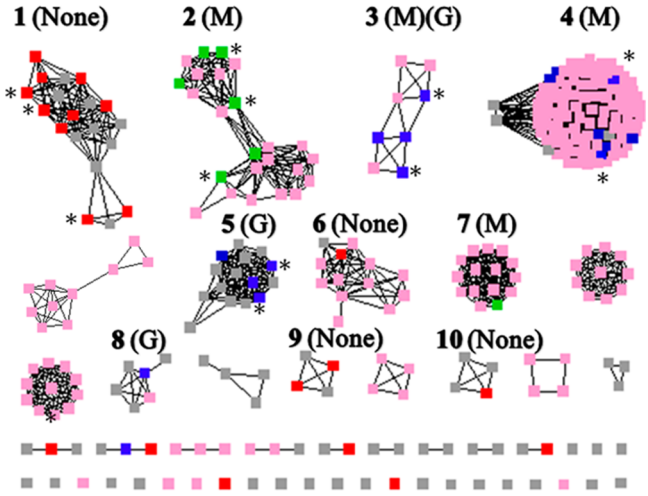


Figure 3. SSN (e-value threshold of at 10^{-190}) showing the distribution of high- (green), low- (blue), and no-activity (red) proteins along with substrate specificities (M, D-mannonate; G, D-gluconate; M/G, D-mannonate and D-gluconate). Proteins for which structures were determined are marked with asterisks. The Pro and Ala residues associated with different substrate specificities for D-mannonate and D-gluconate are located in separate clusters: clusters 1, 4, 5, 6, 8, 9, and 10 contain Pro; clusters 2 and 7 contain Ala; and cluster 3 contains both. Pro-containing clusters exhibit low or no dehydration activity; Ala-containing clusters exhibit high dehydration activity with D-mannonate.

that dehydrates both D-mannonate and D-gluconate (PDB code 3BSM). This was accomplished by superposing the structures of *Na*ManD with D-mannonate in its active site (2QJM), Uniprot ID B5R541 with D-gluconate in its active site (3TWB), and unliganded *Cs*ManD (3BSM) (Figure 4). On the basis of this comparison, we hypothesized that the conserved His after the seventh β-strand is the base in D-gluconate dehydration (His 315 in *Cs*ManD).

Site-directed mutagenesis was performed to convert His 315 in *Cs*ManD to either Asn or Gln. His 315 is conserved in all members of the ManD subgroup, including *Na*ManD, because it hydrogen bonds to the 5-hydroxyl group of D-mannonate

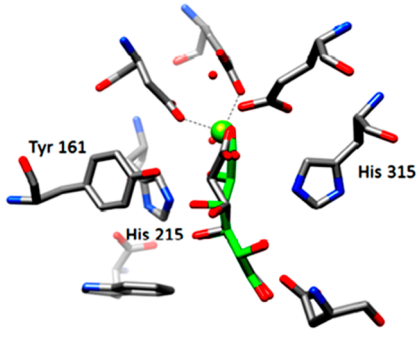


Figure 4. A superposition of a structure with D-mannonate bound in the active site (2QJM, *Na*ManD) with one with D-gluconate bound in the active site (3TWB; *Cs*ManD). In 2QJM, Tyr 161 is the general base that abstracts the 2-proton and His 215 is the general acid that catalyzes the departure of the 3-OH group from D-mannonate. In 3TWB, His 315 is proposed to be the general base that abstracts the 2-proton from D-gluconate or hydrogen bonds with the C5 hydroxyl of D-mannonate. The ϵ-nitrogen of His 315 is 3.0 Å from the C5 hydroxyl of D-mannonate and 3.1 Å from C2 of D-gluconate. Both distances are appropriate for proton abstraction or hydrogen bonding.

(Figure S2, Supporting Information).[12] Both mutants abolished dehydration activity with D-gluconate. However, the H315Q mutant maintained wild-type catalytic efficiency with D-mannonate (Table 2). In contrast, the H315N mutant was

Table 2. Kinetic Parameters for His 315 Mutants of *Cs*ManD[a]

	D-mannonate			D-gluconate		
protein	WT	H315Q	H315N	WT	H315Q	H315N
k_{cat}/K_M (M^{-1} s^{-1})	10	10	NA	40	NA	NA

[a]WT = wild type; NA = no activity.

inactive with D-mannonate, presumably because it is not able to hydrogen bond to the 5-hydroxyl group. These studies support the suggested role of His 315 as the base for dehydration of D-gluconate.

Structure Analysis. "New" crystal structures were solved for 12 sequence diverse members of the subgroup; structures previously were available for four other members. Taken together, a total of 36 unliganded and liganded structures are now available for members of the ManD subgroup (Table S3, Supporting Information). These structures were used to identify the general base that initiates dehydration of D-gluconate and also yielded insights into how structure diverges as sequence diverges. An overlay of one structure from each structure-containing cluster is shown in Figure 5 (the overlay

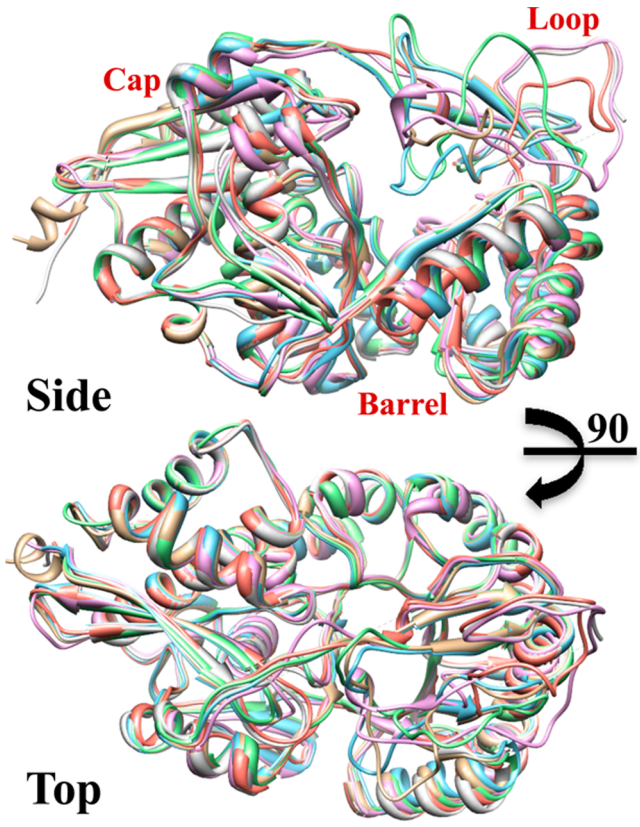


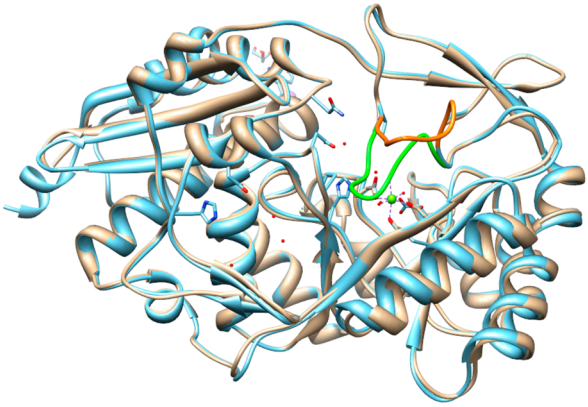
Figure 5. An overlay of *Na*ManD (blue), *Cs*ManD (tan), Uniprot ID Q8FHC7 (green), Uniprot ID B5R541 (magenta), Uniprot ID A5KUH4 (red), and Uniprot ID A6M2W4 (gray) showing the overall structural homology. The "150−180s" loops are conformationally distinct.

dx.doi.org/10.1021/bi500264p | *Biochemistry* 2014, 53, 2722−2731

includes only structures with ordered "150−180s" loops). The structures of the modified TIM-barrel and capping domains are highly conserved, although the conformations of the "150−180s" loop are divergent. The sequences of this loop are also highly variable (Figure S3, Supporting Information).

Initially, the "150−180s" loops were proposed to contain the substrate specificity determinants in analogy with the "20s" loops in other ENS members.[32] Of the 36 structures available (Figure S1, Supporting Information), 9 have an ordered "150−180s" loop covering the active site, 3 with a substrate bound. Seven of the 11 structures with disordered "150−180s" loops also have a substrate bound. The disorder of the loop, even in the presence of substrate, suggests that the substrate and the "150−180s" loop are not interacting strongly. In the structures with a bound substrate and an ordered loop, the only hydrogen bonds to the substrate involve ordered water molecules and backbone amide groups. This suggests a role other than determining substrate specificity for the "150−180s" loops.

To determine the importance of the sequence of the "150−180s" loops, segments of the loops in *Cs*ManD (A164 to E172) and *Na*ManD (V161 to E169) that are proximal to the substrate were mutated to AGAGGAGAG (Figure 6) to



*Na*ManD WT	TGVPGIKDAYGVGRGKLYYEPADASLPSVTG
*Na*ManD GA Loop Mutant	TGVPGIKDAYGAGAGGAGAGPADASLPSVTG
*Cs*ManD WT	VQSGVPGIETTYGVAKTPGERYEPADSSLPA
*Cs*ManD GA Loop Mutant	VQSGVPGIETTYGVAGAGGAGAGPADSSLPA

Figure 6. An overlay of *Cs*ManD (4F4R, blue ribbon, orange loop) and *Na*ManD (2QJJ, tan ribbon, green loop) showing the regions of their "150−180s" loops that were mutated. The sequences of the loops are given with the area of mutagenesis highlighted with yellow.

eliminate any ionic or hydrogen-bonding interactions with the substrate. Both mutants were expressed and purified as soluble proteins; correct folding was verified via circular dichroism and X-ray crystallography (Figures S4 and S5, Supporting Information). Screening of both mutants with the acid sugar library showed a complete loss of activity. We conclude that this loop is important for catalysis, although its exact contribution is unknown.

Structural Differences between D-Mannonate/D-Gluconate Specific ManDs. The structures of proteins with divergent functions are very similar. A superposition of the structures of a high-activity ManD (*Na*ManD), a low activity ManD (Uniprot ID Q8FHC7), a low-activity GlcD (Uniprot ID B5R541), and a member with no activity (Unitprot ID A5KUH4) reveals that the identities and positions of the active site residues are conserved (Figure 7). In addition, the residues within 6 Å of the substrate are conserved. The primary site of

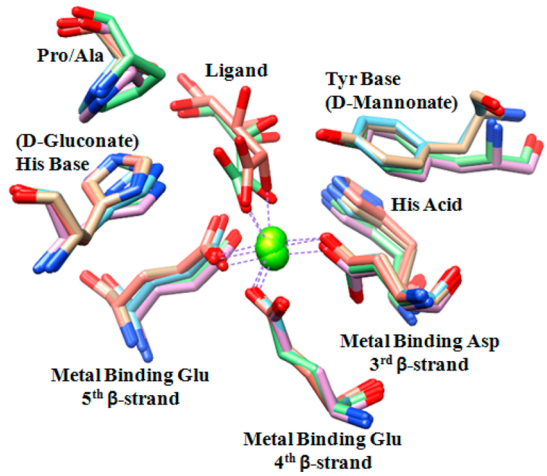


Figure 7. An overlay of the active sites of *Na*ManD (2QJJ, high-activity, D-mannonate specific - red), *Cs*ManD (4F4R, low-activity, promiscuous for D-mannonate/D-gluconate - blue), Uniprot ID D4GJ14 (3T6C, low-activity, D-gluconate specific - green), and Uniprot ID A4W7D6 (3TJI, no-activity - magenta). The metal binding and acid/base residues are superimposable. The Pro/Ala dimorphism is also shown. The ligands are D-mannonate from 2QJM (red) and D-gluconate from 3T6C (green).

divergence in the active site is a Pro/Ala substitution two residues after the conserved His at the end of the seventh β-strand that hydrogen bonds to the 5-hydroxyl group of the substrate in the ManD reaction or is the general base in the GlcD reaction (Figure 4). When this substitution is mapped onto the SSN, low- and no-activity proteins possess Pro at this position, but high-activity ManDs possess an Ala (Figure 3). The single exception is cluster 3, which is recently separated from cluster 2 (i.e., connected at an e-value of 10^{-184}). Cluster 2 includes high-activity ManDs that contain Ala; cluster 3 contains low-activity proteins that contain either Ala or Pro. Members that contain Pro dehydrate both D-mannonate and D-gluconate, but members with Ala dehydrate only D-mannonate. Two of the 8 high-activity ManDs also dehydrate D-gluconate, but with very low catalytic efficiency. We hypothesize that the proteins in cluster 3 may be intermediates in the evolution of the GlcD function.

This hypothesis was investigated by constructing the *Na*ManD A314P and *Cs*ManD P317A mutants. The A314P mutant of *Na*ManD maintained its specificity for D-mannonate, but with a reduced catalytic efficiency (3200 M^{-1} s^{-1} to 60 M^{-1} s^{-1}). The P317A mutant of *Cs*ManD maintained low-activity on D-mannonate and D-gluconate; the catalytic efficiency for D-mannonate is increased (from 5 M^{-1} s^{-1} to 100 M^{-1} s^{-1}), but that for D-gluconate is somewhat decreased (from 40 M^{-1} s^{-1} to 5 M^{-1} s^{-1}). Thus, we conclude that (1) Ala favors D-mannonate dehydration and disfavors D-gluconate dehydration; and (2) Pro disfavors D-mannonate dehydration. A restriction of backbone flexibility may explain the observed change in substrate specificity and catalytic efficiency, with the flexibility associated with Ala allowing hydrogen-bonding to the substrate and the rigidity associated with Pro promoting general-base catalysis.

Proteins with No Activity. The genes encoding many inactive members of the ManD subgroup are neighbors of genes encoding 2-keto-3-deoxy-D-gluconate kinase and 2-keto-3-deoxy-D-gluconate-6-phosphate aldolase that are downstream of ManD in the D-glucuronate catabolic pathway (Figure S6,

dx.doi.org/10.1021/bi500264p | *Biochemistry* 2014, 53, 2722−2731

Supporting Information). However, they have no activity with D-mannonate or the three carbon-2 or -3 epimers that would be dehydrated to 2-keto-3-deoxy-D-mannonate (D-altronate, D-allonate, or D-gluconate). Variations in pH, temperature, salt concentration, osmolarity, osmolytes, and metal cations as well as the presence of dithiothreitol and nucleotide mono-, di-, and triphosphates were tested to determine whether these enzymes may be subject to unanticipated regulation; however, no changes in activity were observed. D-Mannonate 6-phosphate and D-gluconate 6-phosphate also were tested, but no activity was observed. Perhaps these proteins function in multienzyme complexes (channeling) or utilize an acid sugar that is not present in our library.

CONCLUSIONS

The ManD subgroup of the ENS is an excellent example of homologous proteins that have divergent catalytic efficiencies and substrate specificities. Unexpectedly, we discovered that the ManD subgroup is not isofunctional: in addition to the ManDs, some members catalyze the dehydration of D-gluconate, and others are promiscuous for dehydration of both D-mannonate and D-gluconate. Clearly, automated methods would provide misleading/incorrect annotations that would be of limited/no value in deducing their metabolic functions. We have also determined that the structural determinations of substrate specificity are both indirect and subtle: a Pro/Ala substitution appears to be the major determination of specificity.

In addition, the role of the sequence divergent and conformationally flexible "150−180s" loop is uncertain. The side chains of the loop makes no direct contacts with the substrate, so the loop does not appear to be a determinant of substrate specificity as has been well-established for members of the muconate lactonizing enzyme (MLE) and mandelate racemase (MR) subgroups in the ENS. Although we attempted to determine whether the loop is involved in protein−protein interactions/substrate channeling, we could not obtain any conclusive results.

We also discovered that the catalytic efficiencies of members of the subgroup are highly variable, despite conservation of active site residues and structures. These data may provide important insights for the metabolic engineering community. The data illustrates how closely related sequences can perform different reactions, and therefore, may help guide studies which aim to redesign proteins for use in new pathways.

Enzymological dogma has been that enzymes have evolved to achieve catalytic perfection; i.e., the reactions are diffusion-controlled. *In vitro* experiments alone can and will not provide biological insights into why the values of k_{cat}/K_M for some of the members of the subgroup are "low". We have selected several of these for future biological/metabolic characterization so that we might be able to better understand the relationship between catalytic efficiencies and metabolic requirements.

ASSOCIATED CONTENT

Supporting Information

The acid sugar library, circular dichroism data, structural data, sequence alignments, and genome contexts. This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

The X-ray coordinates and structure factors for the following structures have been deposited in the Protein Data Bank: high-activity ManD/Uniprot ID B0T0B1 from *Caulobacter* sp. K31 liganded with Mg^{2+}, Cl^-, and glycerol (entry 4FI4), high-activity ManD/EFI target 502209 from *Caulobacter crescentus* CB15 liganded with Mg^{2+}, Cl^-, D-mannonic acid, CO_3^{2-}, and glycerol (entry 4GME), high-activity ManD/EFI target 502209 from *Caulobacter crescentus* CB15 liganded with Mg^{2+}, Cl^-, CO_3^{2-}, and glycerol (entry 3VCN), low-activity ManD/Uniprot ID B3PDB1 from *Cellvibrio japonicus* Ueda107 liganded with Mg^{2+}, Cl^-, 2-[*N*-cyclohexylamino]ethane sulfonic acid, and glycerol (entry 3V3W), low-activity ManD/Uniprot ID B3PDB1 from *Cellvibrio japonicus* Ueda107 liganded with Mg^{2+}, Cl^-, and L-tartaric acid (entry 3V4B), the P317A mutant of promiscuous ManD/GlcD from *Chromohalobacter salexigens* (Uniprot ID Q1QT89) liganded with Mg^{2+} and D-gluconate (entry 3QKF), promiscuous ManD/GlcD/Uniprot ID Q1QT89 from *Chromohalobacter salexigens* DSM 3043 liganded with Na^+, Cl^-, and glycerol (4F4R), promiscuous ManD/GlcD/Uniprot ID Q1QT89 from *Chromohalobacter salexigens* DSM 3043 liganded with Mg^{2+}, SO_4^{2-}, and glycerol (entry 3OW1), promiscuous ManD/GlcD/Uniprot ID Q1QT89 from *Chromohalobacter salexigens* DSM 3043 liganded with Mg^{2+} and glycerol (entry 3PK7), promiscuous ManD/GlcD/Uniprot ID Q1QT89 from *Chromohalobacter salexigens* DSM 3043 liganded with Mg^{2+}, D-mannonic acid, and 2-keto-3-deoxygluconate (entry 3P93), promiscuous ManD/GlcD/Uniprot ID Q1QT89 from *Chromohalobacter salexigens* liganded with Mg^{2+}, and D-gluconate (entry 3QKE), promiscuous ManD/GlcD/Uniprot ID Q1QT89 from *Chromohalobacter salexigens* liganded with Co^{2+} and D-arabinohydroxamate (entry 3RGT), no-activity protein/Uniprot ID A6M2W4 from *Clostridium beijerinckii* liganded with Mg^{2+} (entry 3S47), Uniprot ID C6CBG9 from *Dickeya dadantii* Ech703 liganded with Mg^{2+}, formic acid, I^-, Cl^-, and glycerol (4IHC), no-activity protein/Uniprot ID A4W7D6 from *Enterobacter* sp. 638 liganded with Mg^{2+}, Cl^-, and glycerol (entry 3TJI), no-activity protein/Uniprot ID C8ZZN2 from *Enterococcus gallinarum* EG2 liganded with Mg^{2+}, Cl^-, and glycerol (entry 4HNL), low-activity ManD/Uniprot ID Q8FHC7 from *Escherichia coli* CFT073 complexed with Mg^{2+} (entry 4IL2), the loop mutant (V161A, R163A, K165G, L166A, Y167G, Y168A, E169G) of high-activity D-mannonate dehydratase (*Na*ManD) from *Novosphingobium aromaticivorans* liganded with Mg^{2+} and glycerol (entry 4K8G), the mutant A314P of high-activity *Na*ManD from *Novosphingoium aromaticivorans* liganded with Mg^{2+} and D-mannonate (entry 3R4E), low-activity GlcD/Uniprot ID D4GJ14 from *Pantoea ananatis* LMG 20103 liganded with Mg^{2+}, Cl^-, 1,2-ethanediol, and gluconic acid (entry 3T6C), EFI target 502240 from *Pectobacterium carotovorum subsp. carotovorum* PC1 liganded with Mg^{2+}, formic acid, Cl^-, and glycerol (entry 4E4F), low-activity GlcD/Uniprot ID B5R541 from *Salmonella enterica subsp. enterica serovar Enteritidis str.* P125109 liganded with Cl^- and glycerol (entry 3TW9), Uniprot ID B5R541 from *Salmonella enterica subsp. enterica serovar Enteritidis str.* P125109 liganded with Mg^{2+}, Cl^-, and glycerol (entry 3TWA), low-activity GlcD/Uniprot ID B5R541 from *Salmonella enterica subsp. enterica serovar Enteritidis str.* P125109 liganded with Cl^- and glycerol (entry 3TWB), high-activity ManD/Uniprot ID Q1NAJ2 from *Sphingomonas sp.* SKA58 liganded with Mg^{2+}, Cl^-, and glycerol (entry 3THU), no-activity protein/Uniprot ID A5KUH4 from *Vibrionales* bacterium liganded with Mg^{2+}, and glycerol (entry 3R25), the P311A mutant of no-activity protein/ Uniprot ID A5KUH4 from *Vibrionales* bacterium liganded with Mg^{2+} and D-arabinonate (entry 3SBF), no-activity protein/Uniprot ID D0X4R4 from *Vibrio harveyi* liganded with

Mg^{2+}, Cl^-, glycerol, maleic acid, and malonic acid (entry 4GIS), no-activity protein/ Uniprot ID D0X4R4 from *Vibrio harveyi* liganded with Mg^{2+}, Cl^-, 1,2-ethanediol, Na^+, and SO_4^{2-} (entry 4GIR), and no-activity protein/ Uniprot ID D0X4R4 from *Vibrio harveyi* liganded with Mg^{2+}, Cl-, 1,2-ethanediol, glycerol, and 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid (entry 4GGH).

This manuscript describes characterization of *in vitro* enzymatic activities of proteins with the following UniProt accession IDs: A4W7D6, A4WA78, A4XF23, A5KUH4, A5V6Z0, A6AMN2, A6M2W4, A6VRA1, A8RQK7, B0T0B1, B0T4L2, B1ELW6, B3PDB1, B5GCP6, B5QBD4, B5R541, B5RAG0, B8HCK2, C6CBG9, C6CVY9, C6DI84, C7PW26, C8ZZN2, C9A1P5, C9CN91, C9NUM5, C9Y5D5, D0KC90, D0X4R4, D4GJ14, D7BPX0, D8ADB5, D9UNB2, E1V4Y0, G7TAD9, J7KNU2, Q1NAJ2, Q1QT89, Q2CIN0, Q6DAR4, Q8FHC7, Q9A4L8, and Q9AAR4.

AUTHOR INFORMATION

Corresponding Author

*Address: Institute for Genomic Biology, University of Illinois, 1206 W. Gregory Dr., Urbana, IL 61801. Phone: (217) 244-7414. Fax: (217) 333-0508. E-mail: j-gerlt@illinois.edu.

Funding

This research was supported by a program project grant and three cooperative agreements from the U.S. National Institutes of Health (P01GM071790, U54GM093342, U54GM074945, and U54GM094662). Molecular graphics and analyses were 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 NIGMS 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 U.S. DOE under Contract No. 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 Company, which operates the facility.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge Drs. Patricia C. Babbitt and Shoshana Brown (University of California, San Francisco) for the SSN used in this study, Dr. Katie Whalen for assistance in circular dichroism and manuscript preparation, Rafael Toro and Rahul Bhosle for aid in running the crystallization facilities at AECOM, and finally National Synchrotron Light Source (NSLS) for shooting of our crystals.

ABBREVIATIONS

EC, enzyme commission; EFI, Enzyme Function Initiative; ENS, enolase superfamily; HMM, hidden Markov model; KdgK, 2-keto-3-deoxy-D-gluconate kinase; KdgpA, 2-keto-3-deoxy-D-gluconate-6-P aldolase; ManD, mannonate dehydratase subgroup of the enolase superfamily; SDM, site-directed mutagenesis; SSN, sequence similarity network

REFERENCES

(1) Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., and Merrick, J. M. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science 269*, 496−512.

(2) Bork, P. (1996) Go hunting in sequence databases but watch out for the traps. *Genetwork 12*, 425−427.

(3) Devos, D., and V., A. (2001) Intrinsic errors in genome annotation. *Trends Genet. 117* (8), 429−431.

(4) Jones, C. E., Brown, A. L., and Baumann U. Estimating the annotation error rate of the curated GO database sequence annotations. *BMC Bioinformatics* 2007, *8*, 170, DOI: 10.1186/1471-2105-8-170.

(5) Schnoes, A. M., Brown, S. D., Dodevski, I., and Babbitt, P. C. (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput. Biol. 5* (12), No. e1000605.

(6) Tian, W., and S., J. (2003) How well is enzyme function conserved as a function of pairwise sequence identity? *J. Mol. Biol. 333*, 863−882.

(7) 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.

(8) 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.

(9) Gerlt, J. A., Babbit, P. C., and Rayment, I. (2005) Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity. *Arch. Biochem. Biophys. 433*, 59−70.

(10) Babbitt, P. C., and Gerlt, J. A. (1997) Understanding enzyme superfamilies. Chemistry as the fundamental deeterminant in the evolution of new catalytic activities. *J. Biol. Chem. 272*, 30591−30594.

(11) Gerlt, J. A., Allen, K. N., Almo, S. C., Armstrong, R. N., Babbitt, P. C., Cronan, J. E., Dunaway-Mariano, D., Imker, H. J., Jacobson, M. P., Minor, W., Poulter, C. D., Raushel, F. M., Sali, A., Shoichet, B. K., and Sweedler, J. V. (2011) The Enzyme Function Initiative. *Biochemistry 50*, 9950−9962.

(12) Rakus, J. F., Fedorov, A. A., Fedorov, E. V., Glasner, M. E., Vick, J. E., Babbitt, P. C., Almo, S. C., and Gerlt, J. A. (2007) Evolution of enzymatic activities in the enolase superfamily: D-mannonate dehydratase from *Novosphingobium aromaticivorans*. *Biochemistry 46* (45), 12896−12908.

(13) Atkinson, H. J., Morris, J. H., Ferrin, T. E., and Babbitt, P. C. (2009) Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One 4*, e4345.

(14) Studier, F. W. (2005) Protein production by suto-induction in high density shaking cultures. *Protein Expr. Purif. 41*, 207−234.

(15) Gulick, A. M., Hubbard, B. K., Gerlt, J. A., and Rayment, I. (2000) Evolution of enzymatic activities in the enolase superfamily: crystallographic and mutagenesis studies of the reaction catalyzed by D-glucarate dehydratase from *Escherichia coli*. *Biochemistry 39*, 4590−4602.

(16) Olson, J. A. (1959) Spectrophotometric measurement of alpha-keto acid semicarbazones. *Arch. Biochem. Biophys. 85*, 225−233.

(17) Evans, P. (2005) Scaling and assessment of data quality. *Acta Crystallogr. D62*, 62−72.

(18) Szebenyi, D. M. E., Arval, A., Ealick, S., Laluppa, J. M., and Nielsen, C. (1997) A system for integrated collection and analysis of crystallographic diffraction data. *J. Synchrotron Radiat. 4*, 128−135.

(19) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr. 40*, 658−674.

(20) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D53*, 240−255.

(21) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D66*, 213−221.

(22) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of COOT. *Acta Crystallogr. D66*, 486−501.
(23) Barber, A. E., 2nd, and Babbitt, P. C. (2012) Pythoscape: a framework for generation of large protein similarity networks. *Bioinformatics 28*, 2845−2846.
(24) Shannon, P. M. A., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res. 13*, 2498−2504.
(25) 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 45*, 14582−14597.
(26) 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 45*, 14598−14608.
(27) 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.
(28) Ashwel, G. (1962) Enzymes of glucuronic and galacturonic acid metabolism in bacteria. *Methods Enzymol. 5*, 190−208.
(29) Kilgore, W. W., and Starr, M. P. (1959) Catabolism of galacturonic and glucuronic acids by *Erwinia carotovora*. *Biol. Chem. 234*, 2227−2236.
(30) Chang, Y. F., and Feingold, D. S. (1969) Hexuronic acid dehydrogenase of Agrobacterium tumefaciens. *J. Bacteriol. 99*, 667−673.
(31) Condemine, G., and Robert-Baudouy, J. (1991) Analysis of an Erwinia chrysanthemi gene cluster involved in pectin degradation. *Mol. Microbiol. 5*, 2191−2202.
(32) Vick, J. E., Schmidt, D. M., and Gerlt, J. A. (2005) Evolutionary potential of (alpha/beta)$_8$-barrels: In vitro enhancement of a "new" reaction in teh enolase superfamily. *Biochemistry 44*, 11722−11729.
(33) 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.
(34) 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 the hyperthermophilic archaea: a reevaluation. *Biochem. J. 390*, 529−540.