# Identification of the *in Vivo* Function of the High-Efficiency D-Mannonate Dehydratase in *Caulobacter crescentus* NA1000 from the Enolase Superfamily

Daniel J. Wichelecki,[†] Dylan C. Graff,[†] Nawar Al-Obaidi,[‡] Steven C. Almo,[‡] and John A. Gerlt*,[†]

[†]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States
[‡]Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461, United States

ⓢ *Supporting Information*

**ABSTRACT:** The D-mannonate dehydratase (ManD) subgroup of the enolase superfamily contains members with varying catalytic activities (high-efficiency, low-efficiency, or no activity) that dehydrate D-mannonate and/or D-gluconate to 2-keto-3-deoxy-D-gluconate [Wichelecki, D. J., et al. (2014) *Biochemistry 53*, 2722−2731]. Despite extensive *in vitro* characterization, the *in vivo* physiological role of a ManD has yet to be established. In this study, we report the *in vivo* functional characterization of a high-efficiency ManD from *Caulobacter crescentus* NA1000 (UniProt entry B8GZZ7) by *in vivo* discovery of its essential role in D-glucuronate metabolism. This *in vivo* functional annotation may be extended to ∼50 additional proteins.

The enolase superfamily (ENS) has been the subject of function discovery-based research since it was demonstrated that its homologous members catalyze different chemical reactions.[1] Although all members of the ENS catalyze a common partial reaction (general base-catalyzed abstraction of a proton $\alpha$ to a carboxylate group to form a $Mg^{2+}$-stabilized enediolate intermediate), their overall reactions and chemistries can be quite diverse (e.g., $\beta$-elimination and 1,1-proton transfer reactions).[2,3] Nevertheless, the members of the ENS share significant structural homology: they share a $(\beta/\alpha)_7(\beta)$-barrel domain for acid/base catalysis and an $(\alpha+\beta)$ capping domain for substrate specificity.[4−9]

When the D-mannonate dehydratase (ManD) subgroup of the ENS was discovered in 2007, all of its members were assumed to catalyze the dehydration of D-mannonate to 2-keto-3-deoxy-D-gluconate.[10] The Asp, Glu, and Glu ligands for the essential $Mg^{2+}$ are located at the ends of the third, fourth, and fifth $\beta$-strands, respectively, of the barrel domain. The general base catalyst is a Tyr-Arg dyad located at the start of the "150−180s" loop (between the second and third $\beta$-strands); the general acid catalyst is a His located at the end of the third $\beta$-strand.[10]

A large-scale *in vitro* survey of sequence−function space in the ManD subgroup was supported by the Enzyme Function Initiative (EFI; U54GM093342), a large-scale, collaborative, project focused on devising tools and strategies for addressing and correcting the widespread non- and misannotation present in today's automatically curated protein databases.[3,11] For confident functional and physiological assignment, the EFI seeks not only *in vitro* enzymatic activity (values of $k_{cat}$ and $k_{cat}/K_M$) but also *in vivo* metabolic function.

The ManD subgroup of the ENS (in Pfam families PF02746 and PF13378) initially was thought to be isofunctional, but the survey of the sequence space revealed dehydration activity with D-mannonate and/or D-gluconate.[12] The members of the ManD subgroup have a range of catalytic activities and efficiencies: high-efficiency (specific to D-mannonate with a $k_{cat}/K_M$ of $10^3−10^4$ $M^{-1}$ $s^{-1}$), low-efficiency (D-mannonate and/or D-gluconate with a $k_{cat}/K_M$ of $10−10^2$ $M^{-1}$ $s^{-1}$), and no activity.[12] However, the *in vivo* physiological function had not been assigned to any member of the ManD subgroup. The only metabolic pathway in which dehydration of D-mannonate is known to occur is the catabolism of D-glucuronate (Figure 1), in which the ManD reaction is performed by the well-characterized UxuA (utilization of hexuronate A)[13,14] that is included in Pfam family PF03786.

One high-efficiency ManD (Uniprot entry B8GZZ7; Protein Data Bank entries 3VCN and 4GME) was shown to be specific for D-mannonate with a catalytic efficiency of $1.2 \times 10^4$ $M^{-1}$ $s^{-1}$.[12] The UxuAs in D-glucuronate catabolism have catalytic efficiencies of ∼$10^3$ $M^{-1}$ $s^{-1}$. Interestingly, low-efficiency and no activity ManDs are encoded by genomes that encode a member of the UxuA family; however, genomes that encode a high-efficiency ManD do not encode a UxuA.[12] This suggested that high-efficiency ManDs are analogues for UxuAs, with low-efficiency and no activity ManDs likely having different physiological functions.

Of the high-efficiency ManDs that were characterized *in vitro*, we selected B8GZZ7 for *in vivo* study because its encoding organism, *Caulobacter crescentus* NA1000, is amenable to genetic manipulation.[15] *C. crescentus* is an important organism in the study of asymmetric cell division and prokaryotic development because it can differentiate into mobile swarmer cells and immobile stalk cells.[15] Species of *Caulobacter* utilize aquatic vegetation as a carbon source in freshwater environments;[15] D-glucuronate makes up a large part of plant cell-wall biomass.[16] Therefore, we hypothesized that *C. crescentus* NA1000 could utilize D-glucuronate as a carbon source via B8GZZ7, a high-efficiency ManD, despite the absence of UxuA.
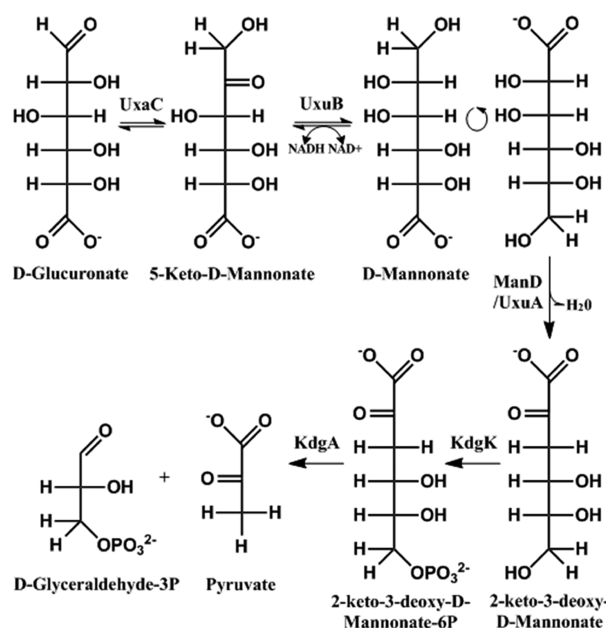
**Figure 1.** Degradation pathway of D-glucuronate in *Escherichia coli.* The dehydration of D-mannonate to 2-keto-3-deoxy-D-gluconate is performed by UxuA.

The gene encoding B8GZZ7 is located in a genome neighborhood that is not indicative of D-glucuronate metabolism. Nearby genes encode catabolic enzymes for *N*-acetylglucosamine 6-phosphate, a mannosidase, and a GntR transcriptional regulator (Figure 2). Nevertheless, the ability of D-glucuronate to upregulate the gene encoding B8GZZ7 was investigated via quantitative reverse transcriptase polymerase chain reaction (qRT-PCR). Expression of the gene encoding B8GZZ7 was induced >1000-fold when *C. crescentus* NA1000 was grown on D-glucuronate as a carbon source (relative to growth on D-xylose) (Figure 3). Importantly, the gene encoding
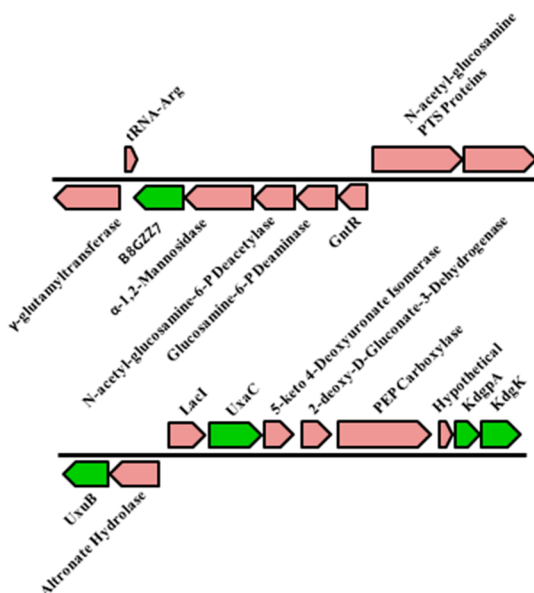


**Figure 2.** Genome neighborhoods of B8GZZ7, a high-efficiency ManD, (top) and canonical D-glucuronate catabolism genes (bottom). The genes directly involved in D-glucuronate metabolism are colored green.
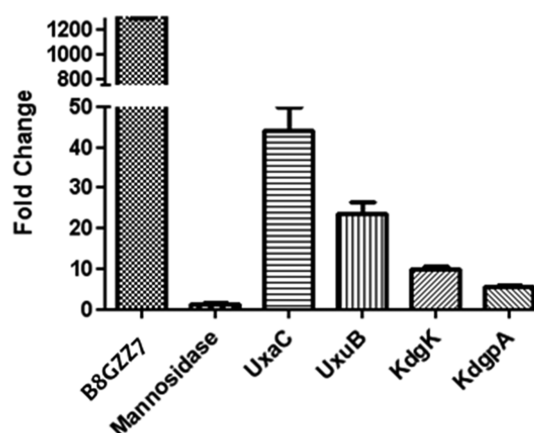


**Figure 3.** Upregulation of the genes encoding B8GZZ7 and the D-glucuronate catabolism genes shown in Figure 2.

the mannosidase immediately upstream of the gene encoding B8GZZ7 was not upregulated (Figure 3); therefore, the gene encoding B8GZZ7 is transcribed separately from its genome neighbors and is likely not involved in *N*-acetylglucosamine catabolism. The absence of cotranscription of the genome neighborhood was further verified by the lack of a PCR product when cDNA was amplified using primers inside of the genes encoding both B8GZZ7 and the mannosidase (such that the potential product would span the gap between the two genes). Thus, we conclude that genomic context can be misleading for functional assignment in the absence of definitive information about regulation and the identities of transcriptional units.

ThermoFluor[17] ligand binding studies were performed on the GntR transcriptional regulator in the genome neighborhood of B8GZZ7 using a library of acid sugars, including D-glucuronate and D-mannonate, as well as other metabolites (Table S2 of the Supporting Information). This assay yielded no stabilizing ligands, providing further evidence that the gene encoding B8GZZ7 is regulated separately.

The genes encoding the remaining enzymes in the canonical D-glucuronate pathway are located elsewhere in the genome, closely clustered with other sugar-metabolizing genes (Figure 2). The genes encoding D-glucuronate isomerase (UxaC), fructuronate reductase (UxuB), 2-keto-3-deoxy-D-gluconate kinase (KdgK), and 2-keto-3-deoxy-D-gluconate-6-phosphate aldolase (KdgA) were upregulated when they were grown on D-glucuronate (relative to growth on D-xylose) (Figure 3). These qRT-PCR data provide compelling evidence that B8GZZ7 is performing the ManD reaction in the canonical D-glucuronate catabolic pathway.

The knockout of the gene encoding B8GZZ7 showed no growth on D-glucuronate (Figure S1 of the Supporting Information). Complementation of the knockout with an isopropyl β-D-1-thiogalactopyranoside-inducible gene encoding B8GZZ7 (pSRK-Kan)[18] restored growth on D-glucuronate. Although the growth rate of the complemented knockout strain was slower than that of wild-type *C. crescentus* NA1000 on D-glucuronate, the wild-type strain transformed with the same complementation vector showed a similar growth defect (Figure S1 of the Supporting Information), likely because of the toxicity of plasmid-based expression of B8GZZ7. The requirement of B8GZZ7 for growth on D-glucuronate and complementation of the growth defect by supplying B8GZZ7 *in trans* verifies that B8GZZ7 is the ManD for D-glucuronate metabolism in *C. crescentus* NA1000.

We visualize functional diversity in the ManD subgroup using sequence similarity networks.[12] Using an $E$-value threshold of $10^{-190}$, corresponding to ~75% identity, the clusters are isofunctional as judged by a sequence alignment.[12] B8GZZ7 is a member of an isofunctional cluster that includes 55 additional proteins (Figure S2 and Table S3 of the Supporting Information). None of the organisms that encode these proteins encode the UxuA in D-glucuronate catabolism, although the genes encoding all of the other enzymes needed for catabolism of D-glucuronate via the pathway shown in Figure 1 are present in the genomes.

The genomes of a few of these organisms, including *C. crescentus* NA1000, encode two high-efficiency ManDs. In *C. crescentus* NA1000, the second ManD (UniProt entry B8H1R9) is not involved in D-glucuronate catabolism: the knockout of the gene encoding the B8H1R9 retains the ability to utilize D-glucuronate as a carbon source (Figure S1 of the Supporting Information); also, the gene encoding B8H1R9 is not upregulated by D-glucuronate. Apparently, the ManD activity of B8H1R9 is essential for another metabolic function that is currently unknown.

We are confident that we can transfer the now verified physiological function of B8GZZ7, the ManD in the catabolic pathway for D-glucuronate, to the high-efficiency ManDs that are encoded by the organisms with a single orthologue (Table S3 of the Supporting Information); all of these organisms encode the other enzymes in D-glucuronate catabolism. For those organisms that encode two high-efficiency ManDs, genome context is sufficient to assign function in some cases, e.g., *Sphingomonas wittichii* and *Sphingomonas* sp. MM-1; in other cases, such as *C. crescentus* NA1000 described here, transcriptomic and genetic experiments are needed to assign the physiological function.

Previously, no physiological function for a ManD in the ENS had been determined. We now have used *in vivo* approaches to establish that B8GZZ7, a high-efficiency ManD in *C. crescentus* NA1000, participates in D-glucuronate catabolism. Thus, we conclude that the high-efficiency ManDs in the ENS are examples of convergent evolution of function with the ManDs in the UxuA family.

This study demonstrates that (1) *in vitro* enzymatic activity alone cannot be used to infer *in vivo* physiological function for an "unknown" enzyme discovered in a genome project and (2) genome context also is not sufficient to infer the *in vivo* physiological function. In due course, we will report the *in vivo* physiological functions for members of the ManD subgroup that have low catalytic efficiencies for dehydration of D-mannonate and D-gluconate.

## ■ ASSOCIATED CONTENT

### Ⓢ Supporting Information

Growth curves, ThermoFluor library, Uniprot entries of orthologous ManDs, and all experimental methods. This material is available free of charge via the Internet at http://pubs.acs.org.

### Accession Codes

This manuscript describes characterization of *in vivo* function of Uniprot entry B8GZZ7.

## ■ AUTHOR INFORMATION

### Corresponding Author

*E-mail: j-gerlt@illinois.edu. Phone: (217) 244-7414. Fax: (217) 244-0508.

### Notes

The authors declare no competing financial interests.

## ■ REFERENCES

(1) Neidhart, D. J., Kenyon, G. L., Gerlt, J. A., and Petsko, G. A. (1990) *Nature* 347, 692−694.

(2) Babbitt, P. C., and Gerlt, J. A. (1997) *J. Biol. Chem.* 272, 30591−30594.

(3) 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) *Biochemistry* 50, 9950−9962.

(4) Gerlt, J. A., Babbit, P. C., and Rayment, I. (2005) *Arch. Biochem. Biophys.* 433, 59−70.

(5) Gulick, A. M., Hubbard, B. K., Gerlt, J. A., and Rayment, I. (2000) *Biochemistry* 39, 4590−4602.

(6) Schmidt, D. M. Z., Hubbard, B. K., and Gerlt, J. A. (2001) *Biochemistry* 40, 15707−15715.

(7) Yew, W. S., Fedorov, A. A., Fedorov, E. V., Rakus, J. F., Pierce, R. W., Almo, S. C., and Gerlt, J. A. (2006) *Biochemistry* 45, 14582−14597.

(8) Yew, W. S., Fedorov, A. A., Fedorov, E. V., Wood, B. M., Almo, S. C., and Gerlt, J. A. (2006) *Biochemistry* 45, 14598−14608.

(9) Yew, W. S., Fedorov, A. A., Fedorov, E. V., Almo, S. C., and Gerlt, J. A. (2007) *Biochemistry* 46, 9564−9577.

(10) 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) *Biochemistry* 46, 12896−12908.

(11) Schnoes, A. M., Brown, S. D., Dodevski, I., and Babbitt, P. C. (2009) *PLoS Comput. Biol.* 5, e1000605.

(12) Wichelecki, D. J., Balthazor, B. M., Chau, A. A., Vetting, M. W., Fedorov, A. A., Fedorov, E. V., Lukk, T., Patskovsky, Y. V., Stead, M. B., Hillerich, B. S., Seidel, R. D., Almo, S. C., and Gerlt, J. A. (2014) *Biochemistry* 53, 2722−2231.

(13) Ashwell, G. (1962) *Methods Enzymol.* 5, 190.

(14) Zhang, Q., Gao, F., Peng, H., Cheng, H., Liu, Y., Tang, J., Thompson, J., Wei, G., Zhang, J., Du, Y., Yan, J., and Gao, G. F. (2009) *J. Bacteriol.* 191, 5832−5837.

(15) Hottes, A. K., Meewan, M., Yang, D., Arana, N., Romero, P., McAdams, H. H., and Stephens, C. (2004) *J. Bacteriol.* 186, 1448−1461.

(16) Seifert, G. J. (2004) *Curr. Opin. Plant Biol.* 7, 277−284.

(17) Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. A., Lane, P., and Salemme, F. R. (2001) *J. Biomol. Screening* 6, 429−440.

(18) Khan, S. R., Gaines, J., Roop, R. M., II, and Farrand, S. K. (2008) *Appl. Environ. Microbiol.* 74, 5053−5062.