A Functionally Diverse Enzyme Superfamily That Abstracts the α Protons of Carboxylic Acids

Patricia Clement Babbitt,* Gregory T. Mrachko, Miriam Sarah Hasson, Gjalt W. Huisman, Roberto Kolter, Dagmar Ringe, Gregory A. Petsko, George L. Kenyon, John A. Gerlt*

Mandelate racemase and muconate lactonizing enzyme are structurally homologous but catalyze different reactions, each initiated by proton abstraction from carbon. The structural similarity to mandelate racemase of a previously unidentified gene product was used to deduce its function as a galactonate dehydratase. In this enzyme superfamily that has evolved to catalyze proton abstraction from carbon, three variations of homologous active site architectures are now represented: lysine and histidine bases in the active site of mandelate racemase, only a lysine base in the active site of muconate lactonizing enzyme, and only a histidine base in the active site of galactonate dehydratase. This discovery supports the hypothesis that new enzymatic activities evolve by recruitment of a protein catalyzing the same type of chemical reaction.

New enzymatic activities likely evolve by the recruitment of a protein catalyzing the same type of chemical reaction but possessing a different substrate specificity (*1*, *2*). Support for this hypothesis has been provided by the discovery that four enzymes in the mandelate pathway in *Pseudomonas putida* are homologous to enzymes in ubiquitous metabolic pathways (*2*, *3*). The mandelate pathway produces benzoate from *R*-mandelate, thereby allowing mandelate to serve as the sole carbon source.

Mandelate racemase (MR) equilibrates the enantiomers of mandelate by a Mg^{2+}-dependent mechanism (Fig. 1A) that involves the abstraction of the proton adjacent to the carboxylate group (α proton) by a general base (*4*, *5*). Lys^{166} (the *S*-specific base) abstracts the proton from *S*-mandelate, and His^{297} (the *R*-specific base) abstracts the proton from *R*-mandelate. [The active site of MR is shown in Fig. 2 (*6*).] The immediate product of proton abstraction from either enantiomer is an enolic intermediate. Protonation of the intermediate by the conjugate acid of either Lys^{166} or His^{297} yields the product enantiomer.

Both the primary (*7*) and three-dimensional structures (*8*) of MR are homologous to the primary (*9*) and three-dimensional (*10*, *11*) structures of muconate lactonizing enzyme (MLE). MLE, an enzyme involved in benzoate catabolism, catalyzes a Mn^{2+}-dependent reaction in which *cis*,*cis*-muconate is equilibrated with muconolactone. Although the reaction catalyzed by MLE may not appear to resemble the reaction catalyzed by MR, in the reverse reaction the α proton of muconolactone is abstracted by a homolog of Lys^{166} (Lys^{169}) to generate an enolic intermediate (Fig. 1B). This intermediate undergoes vinylogous β-elimination of the remote carboxylate group to produce *cis*,*cis*-muconate (*12*). Although the active site of MLE does not contain a homolog for His^{297}, the overall structural similarities establish that MR and MLE are members of a novel enzyme superfamily and led to the proposal that MLE may have been recruited to catalyze the racemization of mandelate enantiomers (*11*).

We have found five additional open reading frames (ORFs) of unassigned or uncertain function whose predicted protein products are homologous to MR (*13*, *14*). Each of these proteins contains a homolog of His^{297} (the *R*-specific base); only one unequivocally contains a homolog of Lys^{166} (the *S*-specific base). One of these ORFs, *f587*, was sequenced in the *Escherichia coli* genome project (*13*). The COOH-terminal two-thirds of the predicted protein encoded by *f587* is homologous to MR but does not contain a homolog of Lys^{166}. To identify the function of this protein, we assumed that its activity involves abstraction of the α proton from a carboxylic acid, a common chemical step catalyzed by MR and MLE.

An operon (*dgo*) for the utilization of D-galactonate (*15*) (Fig. 3A) had previously been located in the vicinity of *f587* (82 min) on the *E. coli* linkage map (*16*). Because the reaction catalyzed by galactonate dehydratase (GalD) is initiated by abstraction of the α proton from galactonate, we hypothesized that the COOH-terminal two-thirds of *f587* encodes GalD.

Three additional lines of evidence supported this hypothesis. First, the GalD activity associated with this pathway is dependent on Mg^{2+} (*17*). A Mg^{2+} binding motif

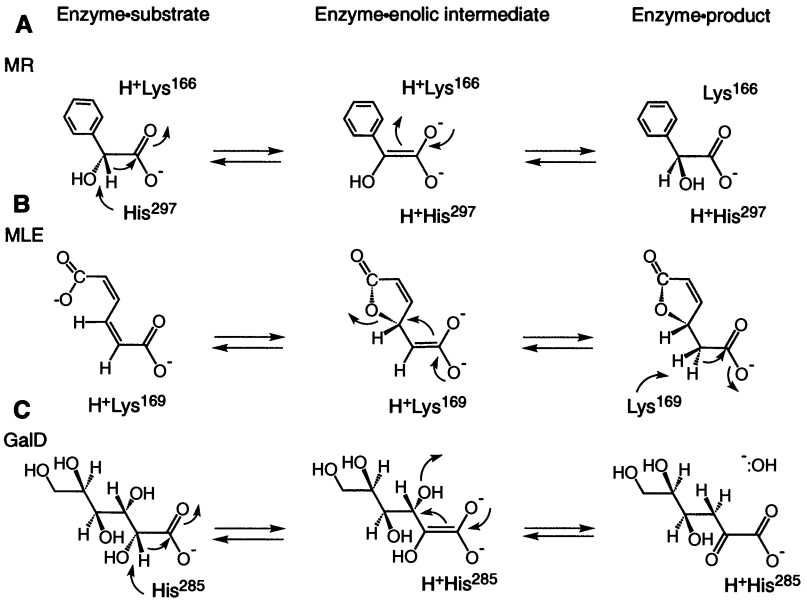

Fig. 1. Mechanisms of the reactions catalyzed by MR (**A**), MLE (**B**), and GalD (**C**). For clarity, only the general base catalysts involved in abstraction of the α protons of the carboxylic acid substrates are shown. In each reaction, the enolic intermediate resulting from abstraction of the α proton is stabilized by interactions with electrophilic groups in the active site, including a divalent metal ion, the γ carboxylic acid group of a glutamic acid, and the ε ammonium group of a lysine residue (*6*). In the active site of MLE, Lys^{169} is the homolog of Lys^{166}, and no homolog is present for His^{297} in MR. In the active site of GalD, His^{285} is the homolog of His^{297}, and no homolog is present for Lys^{166} in MR.

P. C. Babbitt and G. L. Kenyon, Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, CA 94143, USA.
G. T. Mrachko and J. A. Gerlt, Department of Biochemistry, University of Illinois, Urbana, IL 61801, USA.
M. S. Hasson, D. Ringe, G. A. Petsko, Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02254, USA.
G. W. Huisman and R. Kolter, Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115, USA.
*To whom correspondence should be addressed.

is present in the predicted protein encoded by *f587* and aligns with the metal binding motifs of MR and MLE (*18*).

Second, the absolute configuration of the 2-carbon of galactonate is *R*. If galactonate binds to Mg^{2+} via its 2-hydroxyl group and a carboxylate oxygen in analogy to the observed binding of mandelate to Mg^{2+} in MR (6), the general base in GalD should be a homolog of His^{297}, the *R*-specific base in MR. There is such a candidate base, His^{285}, in the GalD sequence.

Third, the *dgo* operon was proposed (*15*) to include genes for a galactonate permease (*dgoT*), a galactonate dehydratase (*dgoD*), a 2-oxo-3-deoxygalactonate kinase (*dgoK*), a 2-oxo-3-deoxygalactonate 6-phosphate aldolase (*dgoA*), and a regulatory protein (*dgoR*). The DNA sequence suggests that *f587* is the third in a series of four predicted ORFs (designated *f177*, *f292*, *f587*, and *f445*) that could constitute a single transcriptional unit (*13*) (Fig. 3B). Although five, not four, proteins were proposed to be involved in galactonate utilization (*15*, *17*), the region of the chromosome containing these ORFs could encode the five expected proteins of the galactonate utilization pathway.

From sequence analysis, the predicted *f445* gene product is similar to bacterial permeases (*19*), which suggests that it encodes galactonate permease (*dgoT*). The predicted *f177* gene product is similar to the gntR family of repressors (*20*), which suggests that it encodes the regulatory protein (*dgoR*). Because the *dgo* operon is subject to catabolite repression, the transcriptional unit should be and is preceded by a possible binding site for the catabolite activator protein (*21*).

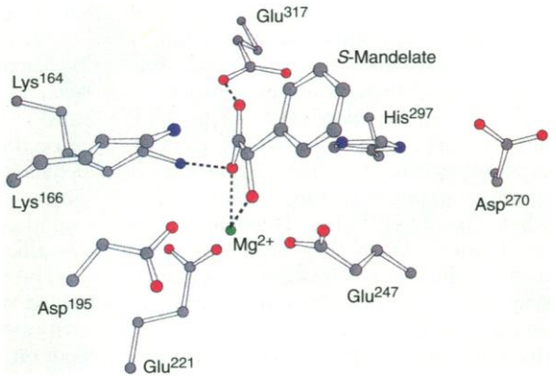

Fig. 2. A model of the active site of MR with the substrate *S*-mandelate bound. The coordinates used to generate this model were obtained from the structure of the covalent adduct between MR and *R*-α-phenylglycidate by deletion of the carbon atom bridging the ε nitrogen of Lys^{166} with the inhibitor. The dotted lines represent the modeled noncovalent interactions between the bound substrate and active-site electrophilic groups. The metal binding ligands in MR are the residues Asp^{195}, Glu^{221}, and Glu^{247}. Their respective homologs in GalD are the residues Asp^{183}, Glu^{209}, and Glu^{235} (numbered from the corrected NH_2-terminus of GalD as reported here).

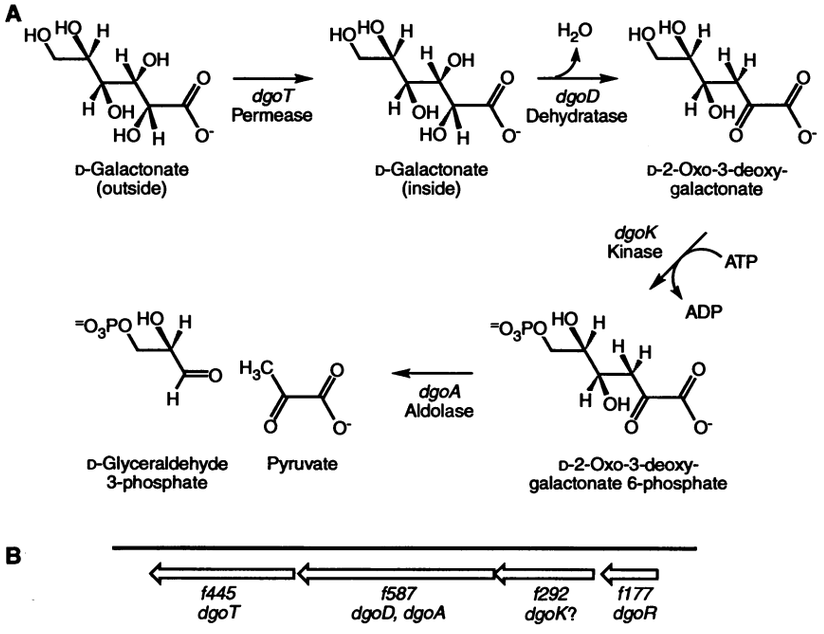

Fig. 3. (**A**) The pathway for utilization of galactonate in *E. coli* (*17*). (**B**) The organization of the putative *dgo* operon (*13*, *15*).

To test the hypothesis that *f587* encodes GalD, a cell-free extract of *E. coli* DH5α transformed with plasmid pGTM1 that contained only *f587* (*22*) was assayed for GalD activity (*17*, *23*). An α ketoacid was produced from galactonate (specific activity 0.042 μmol min^{-1} mg^{-1}) but not from gluconate or galactarate (<0.002 μmol min^{-1} mg^{-1}). No detectable activity (<0.002 μmol min^{-1} mg^{-1}) was detected in a control experiment in which a cell-free extract of DH5α transformed with pUC18 (pGTM1 lacking the *f587* insert) was assayed. We conclude that *f587* does encode GalD and that the position of the *dgo* operon in the physical map (*16*) can be correlated with the sequence of the chromosome (*13*).

We purified the GalD activity to homogeneity from this cell-free extract (*24*). Its mass was 43 kD as determined by SDS–polyacrylamide gel electrophoresis (PAGE), not 64 kD as predicted from the sequence of *f587*. The NH_2-terminal sequence of GalD, Met-Lys-Ile-Thr-Lys-Ile-Thr, was determined by automated Edman degradation and was contained within the predicted gene product of *f587* (the NH_2-terminal Met was codon 206 of *f587*). On this basis, we predict that GalD contains 382 amino acids and has a molecular mass of 43 kD, as experimentally determined. For comparison, MR contains 358 amino acids (*7*), and MLE contains 372 amino acids (*9*).

We hypothesized that the smaller than expected size of GalD is the result of sequence errors after codon 115 of *f587*. The first 115 residues are similar to analogous sequences in a family of bacterial aldolases (*13*, *25*), but the similarity does not extend through an active-site lysine residue that would be expected near codon 125. We resequenced this region of *f587* and found three discrepancies between codons 122 and 204. In the protein encoded by the corrected sequence (GenBank accession number U19577), the homolog of the active-site lysine is present at codon 126, and a TGA stop codon (codon 206) shares its TG with the initiation codon for GalD. We conclude that the DNA included in *f587* encodes both GalD (the COOH-terminal two-thirds; *dgoD*) and 2-oxo-3-deoxygalactonate 6-phosphate aldolase (the NH_2-terminal one-third; *dgoA*; Fig. 3B).

Sequence analyses have not revealed any significant homology of the predicted protein encoded by *f292* with any kinase. However, with the other ORFs in this cluster assigned to specific genes in the *dgo* locus, *f292* most likely encodes 2-oxo-3-deoxygalactonate kinase (*dgo*K). Accordingly, the expected five enzymes in the galactonate utilization pathway are arranged contiguously as five distinct genes in a *dgo* operon (Fig. 3B), as predicted (*15*).

In the reaction catalyzed by GalD, we propose that the active-site base (His^{285}, the homolog of His^{297} in MR) abstracts the 2-R proton of galactonate (Fig. 1C). The immediate product of proton abstraction is an enolic intermediate. This intermediate undergoes vinylogous elimination (*12*) of the 3-hydroxyl group to generate an α,β-unsaturated enol that tautomerizes to the 2-oxo-3-deoxygalactonate product.

With the identification of GalD, we realize that the lysine and histidine general bases found in the active site of MR evolved independently to generate the three different reactions now associated with the MR-MLE-GalD superfamily: racemization catalyzed by MR (requiring both Lys^{166} and His^{297}), β-elimination catalyzed by MLE (with a homolog of Lys^{166} but not of His^{297}), and β-elimination catalyzed by GalD (with a homolog of His^{297} but not of Lys^{166}).

These findings mean that we no longer need to assume that MLE is the immediate evolutionary precursor to MR. Although MR and MLE are both necessary for the catabolism of *R*-mandelate by *P. putida*, the identification of a homolog in carbohydrate metabolism suggests that this superfamily of enzymes may be ubiquitous in nature. In particular, any member of this superfamily of enzymes may have been recruited when the evolution of a metabolic pathway required a reaction in which the α proton of a carboxylic acid must be abstracted (*26*).

REFERENCES AND NOTES

1. R. A. Jensen, *Annu. Rev. Microbiol.* **30**, 409 (1976); *Cold Spring Harbor Symp. Quant. Biol.* **52** (1987).
2. G. A. Petkso, G. L. Kenyon, J. A. Gerlt, D. Ringe, J. W. Kozarich, *Trends Biochem. Sci.* **18**, 372 (1993).
3. A. Y. Tsou *et al.*, *Biochemistry* **29**, 9856 (1990); S. L. Shafer, P. C. Babbitt, J. A. Gerlt, G. L. Kenyon, unpublished material.
4. J. A. Gerlt, G. L. Kenyon, J. W. Kozarich, G. A. Petsko, V. M. Powers, *Curr. Opin. Struct. Biol.* **2**, 736 (1992).
5. J. A. Gerlt and P. G. Gassman, *Biochemistry* **32**, 11943 (1993).
6. J. A. Landro *et al.*, *ibid.* **33**, 635 (1994).
7. S. C. Ransom, J. A. Gerlt, V. Powers, G. L. Kenyon, *ibid.* **27**, 540 (1988).
8. D. J. Neidhart *et al.*, *ibid.* **30**, 9264 (1991).
9. MLE I was from *P. putida* [T. L. Aldrich, B. Frantz, J. F. Gill, J. J. Kilbane, A. M. Chakrabarty, *Gene* **52**, 185 (1987); K. T. Taylor, thesis, University of Maryland, College Park (1991)].
10. A. Goldman, B. L. Ollis, T. A. Steitz, *J. Mol. Biol.* **194**, 143 (1987); M. Hasson, I. Schlichting, J. Moulai, G. Petsko, D. Ringe, unpublished results.
11. D. J. Neidhart, G. L. Kenyon, J. A. Gerlt, G. A. Petsko, *Nature* **327**, 692 (1990).
12. J. A. Gerlt and P. G. Gassman, *J. Am. Chem. Soc.* **114**, 5928 (1992).
13. *f587* was from *E. coli* [V. Burland, G. Plunkett III, D. L. Daniels, F. R. Blattner, *Genomics* **16**, 551 (1993)].
14. *spa2* was from *Streptomyces ambofaciens* [D. Schneider, B. Aigle, P. Leblond, J.-M. Simonet, B. Decaris, *J. Gen. Microbiol.* **139**, 2559 (1993)]; *rspA* was from *E. coli* [G. W. Huisman and R. Kolter, *Science* **265**, 537 (1994)]; *gdh2* was from *P. putida* [R. P. Burlingame *et al.*, GenBank accession number M69160]; *rTS* was from *Homo sapiens* [B. J. Dolnick, *Nucleic Acids Res.* **21**, 1747 (1993); personal communication].
15. R. A. Cooper, *Arch. Microbiol.* **118**, 199 (1978).
16. B. J. Bachman, *Microbiol. Rev.* **54**, 130 (1990).
17. J. Deacon and R. A. Cooper, *FEBS Lett.* **77**, 201 (1977).
18. P. C. Babbitt *et al.*, data not shown.
19. K. Eiglmeier, W. Boos, S. T. Cole, *Mol. Microbiol.* **1**, 251 (1987).
20. D. J. Haydon and J. R. Guest, *FEMS Microbiol. Lett.* **63**, 291 (1991).
21. The sequence TTTTGTGATCT·AAATTGTTAGTA is 24 bp upstream of a putative promoter sequence (−35, TTCAC; −10, CGTCAT) that was identified by Burland *et al.* (*13*). The palindromic consensus sequence for a CAP binding site is AAATGTGATCT·AGATCACATTT [A. Kolb, S. Busby, H. Buc, S. Garges, S. Adhya, *Annu. Rev. Biochem.* **62**, 749 (1993)]. This sequence is 81 base pairs upstream of the initiation codon for *f177* in the putative operon.
22. A 2.453 kb Sac I–Sal I restriction fragment containing *f587* was subcloned from lambda phage Kohara 166 [Y. Kohara, K. Akiyama, K. Isono, *Cell* **50**, 495 (1987)] into pUC18; the resulting plasmid, pGTM1, was used to transform *E. coli* DH5α.
23. A. Donald, D. Sibley, D. E. Lyons, A. S. Dahms, *J. Biol. Chem.* **254**, 2132 (1979). The enzymatic assays (25°C) contained 25 mM substrate, 10 mM $MgCl_2$, and 50 mM glycylglycine.
24. Nucleic acids were precipitated with streptomycin sulfate; protein was fractionated with ammonium sulfate. GalD was purified by ion-exchange chromatography (DEAE-Sephacel), gel filtration (Sephacryl S-200 HR), and finally ion-exchange chromatography (Mono Q). As assessed by SDS-PAGE, the resulting GalD was more than 98% pure.
25. S. E. Egan *et al.*, *J. Bacteriol.* **174**, 4638 (1992).
26. For example, a structural homology between enolase and MLE has been noted [L. Lebioda and B. Stec, *Nature* **333**, 683 (1988); J. H. Wedekind, G. H. Reed, I. Rayment, *Biochemistry*, in press].

27. We thank R. Armstrong, D. Dunaway-Mariano, R. Fletterick, and R. Stroud for comments regarding the manuscript. Supported by NIH grants GM-34572 (to J.A.G.) and GM-40570 (to J.A.G., G.L.K., and G.A.P.), by NSF grant MCB-9207323 (to R.K.), by The Medical Foundation, Inc., Boston, MA (to G.W.H.), and by the Cancer Research Fund of the Damon Runyon–Walter Winchell Foundation grant DRG-1194 (to M.S.H.).

1 September 1994; accepted 21 December 1994

Circadian Clock Mutants in *Arabidopsis* Identified by Luciferase Imaging

Andrew J. Millar, Isabelle A. Carré, Carl A. Strayer, Nam-Hai Chua, Steve A. Kay*

The cycling bioluminescence of *Arabidopsis* plants carrying a firefly luciferase fusion construct was used to identify mutant individuals with aberrant cycling patterns. Both long- and short-period mutants were recovered. A semidominant short-period mutation, timing of *CAB* expression (*toc1*), was mapped to chromosome 5. The *toc1* mutation shortens the period of two distinct circadian rhythms, the expression of chlorophyll *a/b*–binding protein (*CAB*) genes and the movements of primary leaves, although *toc1* mutants do not show extensive pleiotropy for other phenotypes.

Many aspects of plant growth and metabolism are regulated by the circadian clock. Genes that affect circadian rhythmicity have been isolated in *Drosophila melanogaster* (*period* or *per*) and *Neurospora crassa* (*frequency* or *frq*) (*1*). Allelic series at each locus include long- and short-period alleles, whereas the null alleles are largely arrhythmic. Although there is no clear homology of *per* or *frq* sequences to the *Arabidopsis thaliana* genome (*2*), genetic screens for period mutants should nevertheless identify genes required for oscillator function.

A transgenic *Arabidopsis* line was previously constructed in which a fragment of the *Arabidopsis CAB2* promoter mediates circadian-regulated transcription of the firefly luciferase (*Luc*) reporter gene (*3*). We used this *cab2::Ω::Luc* transgenic line to screen for mutant seedlings, in which the first peak of the free-running luminescence cycle occurs either earlier or later than the wild type. M2 seedlings (*4*) were grown for 5 days in 12-hour light:12-hour dark (LD) cycles and then transferred to continuous white light (LL). Candidate mutants were initially selected from M2 populations on the basis of a three-time-point screen (*5*). The mutant candidates were entrained to two LD cycles and assayed again in LL to confirm the mutant phenotype (*6*). We retained 26 "timing of *CAB* expression" (*toc*) lines, which represent at least 21 independent mutations. The mutant phenotypes under LL included seven short-period lines (periods of 21 to 22.5 hours, Fig. 1A), 11 long-period lines (periods of 26 to 28 hours, Fig. 1B), one line with the wild-type period but reduced amplitude (Fig. 1C), and two

A. J. Millar, National Science Foundation (NSF) Center for Biological Timing, Department of Biology, Gilmer Hall, University of Virginia, Charlottesville, VA 22903, USA, and Laboratory of Plant Molecular Biology, Rockefeller University, 1230 York Avenue, New York, NY 10021–6399, USA.
I. A. Carré, C. A. Strayer, S. A. Kay, NSF Center for Biological Timing, Department of Biology, Gilmer Hall, University of Virginia, Charlottesville, VA 22903, USA.
N.-H. Chua, Laboratory of Plant Molecular Biology, Rockefeller University, 1230 York Avenue, New York, NY 10021–6399, USA.
*To whom correspondence should be addressed.