FINAL REPORT
Grant # DE-FG02-88ER13947

Organization and Control of Genes Encoding Catabolic Enzymes in
Rhizobiaceae

D. Parke and L. N. Ornston

Department of Molecular, Cellular and Developmental Biology
Yale University, New Haven, CT 06520-8103

RECEIVED

SEP 13 1999

OSTI

A diverse array of phenolic compounds are measurable at low levels in the soil, the rhizosphere, and plant tissues. Specific phenolic compounds released from plant roots serve as regulatory signals in the nodulation of plant roots by rhizobia, and certain phenolic compounds released during plant wounding trigger initiation of tumorigenesis by related, pathogenic agrobacteria. The contribution that bioconversion processes make to the generation or depletion of these signal molecules by the target microbes or associated micro-organisms is unclear. In addition, beyond their role as co-inducers of nodulation or virulence genes, phenolic compounds may serve as carbon and energy sources that are significant in defining the niche of a given microbe. The great assortment of phenolic compounds in the natural environment translates into their assimilation by bacteria through a diverse array of metabolic reactions. Many of these reactions converge on a few metabolic pathways. Our research has focused on one of these, the protocatechuate branch of the aerobic, β -ketoadipate pathway, which we have found to be universally distributed among the bacterial group, the *Rhizobiaceae*, comprised of agrobacteria and rhizobia.

Our studies of the protocatechuate pathway in the *Rhizobiaceae* contribute to an understanding of bacterial energy metabolism in a number of respects. (1) Our research relates to bioconversion of phenolic signal molecules because, of the phenolic co-inducers that have been identified as being dissimilated, they are broken down via the protocatechuate pathway. (2) It casts light on the selective forces and evolutionary processes that give rise to a particular pattern of genetic organization and regulation. (3) It underscores the elaborate regulatory mechanisms that distinguish members of this group of bacteria from other, more divergent bacteria. (4) The combination of understanding the regulation of the pathway at a molecular level and having

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

in hand appropriate mutant strains provides the background for research exploring the role of the *pca* genes in this group of bacteria. (5) It also allows us to pursue the question of how bacteria perceive and respond to the plant root or the plant wound environment: under what circumstances are the genes for phenolic compound dissimilation expressed in these environments and are there common substrates that act as preferred carbon and energy sources to repress expression of each unit of transcription required for bioconversion of phenolic compounds?

Two publications supported by this grant reported on methods which were subsequently used throughout our project. The first was the construction of a set of mobilizable pUC-based vectors carrying different antibiotic resistance markers. In addition to having many requests for these plasmids, we deposited them with the American Type Culture Collection at their request. The second "methods" publication communicated the discovery and application of a compound that serves as a chromophore for diphenolic compounds. Beyond its utility in our studies involving the diphenolics protocatechuate and catechol, the chromogenic compound serves as an indicator specific for diverse, substituted diphenolics. As such, it is relevant to studies on degradation of a range of pollutants found at D.O.E. waste sites.

Another early publication from this grant reported on the regulation of genes of the protocatechuate pathway in *Rhizobium leguminosarum* biovar trifolii. Work carried out in collaboration with the laboratory of Andrew Glenn of Murdoch University in Australia employed Tn5 mutants of *R. leguminosarum* biovar trifolii to elucidate the regulation of the β -ketoadipate pathway in this strain. The study revealed that at least two *pca* genes were induced by β -carboxy-*cis,cis*-muconate, and the *pcaIJ* genes were induced by β -ketoadipate. The utilization of β -carboxy-*cis,cis*-muconate as a co-inducer of *pca* genes appears to be an unusual trait, as evaluated by the characterization of molecules that induce *pca* genes in bacteria that are phylogenetically distinct.

Our further investigations largely focused on an analysis of DNA encoding enzymes for the final steps for conversion of phenolic compounds into tricarboxylic acid cycle intermediates. Supraoperonic clustering of genes required for this pathway in *Agrobacterium tumefaciens* A348 helped to facilitate the study. Characterization of a 9-kb segment of DNA has been completed and has revealed that it contains the protocatechuate (*pca*) and 4-hydroxybenzoate (*pob*) genes for the conversion of the phenolic compound 4-

hydroxybenzoate to β -ketoadipate. There is evidence for conservation of nucleotides in the 9-kb segment. A single recombinational event could eliminate the *pob-pca* cassette altogether. The fact that the genes continue to be maintained, and that protocatechuate appears to be a universal carbon source for members of the *Rhizobiaceae* as well as their relatives, the bradyrhizobia, indicates that the genes confer a selective benefit to the bacteria.

The accompanying MiniReview, "Acquisition, reorganization, and merger of genes: novel management of the β -ketoadipate pathway in *Agrobacterium tumefaciens*", provides a particularly succinct summary of research results as well as a figure for reference. Consequently, this final report will be brief on the particulars of research findings that are contained in the reprints of publications and will emphasize unpublished research results on the characterization of two additional regulatory genes found within the 9-kb region. It should be noted that preliminary, partial sequencing data on cloned DNA flanking the 9-kb region has revealed a set of ORF's for genes encoding a transport system on one side and a fourth regulatory gene, with a deduced product homologous to a family of repressor proteins, on the other side. Unfortunately, follow-up analysis of these regions has not been possible.

Figure 1. Partial restriction map of a 4.35-kb region of the *pca-pob* genetic cluster in *A. tumefaciens* A348. Structural genes are shown as gray bars; regulatory genes are patterned. Arrows show directions of transcription. The map does not show the *pcaDCHGB* operon which is transcribed in the opposite direction; its translational start site is 95 nucleotides from the *pcaQ* ORF.

The *pca* structural genes which encode enzymes for dissimilation of protocatechuate to β -ketoadipate are located in two divergently transcribed and independently regulated operons, *pcaDCHGB* and *pcaIJ*, located 4-kb apart. A regulatory gene, *pcaQ*, lies within the 4-kb region, next to the *pcaDCHGB* operon which it controls. Downstream of *pcaQ* and convergently transcribed from it is the *pobA* structural gene, encoding the hydroxylase which converts 4-hydroxybenzoate to protocatechuate. In the remaining 2-kb region between *pobA* and the *pcaIJ* genes lie two more regulatory genes, *pobR* and *pcaR*, that control the respective, adjacent structural genes (Fig. 1).

Analysis of the sequence of *pcaR* from *A. tumefaciens* revealed that its deduced polypeptide product is homologous to a regulator from *Pseudomonas putida*. In both species, β -ketoadipate serves as the coinducer of PcaIJ. Mutation of the *A. tumefaciens* *pcaR* was accomplished by introduction of an omega (Ω) element, which terminates transcription and translation, just upstream of a sequence likely to encode a helix-turn-helix DNA binding motif in *pcaR*. The mutant strain, ADO2069, is unable to grow at the expense of protocatechuate or carbon sources that are precursors of it. It is complemented by a full length *pcaR* gene in *trans*, and curing of the plasmid carrying the gene in *trans* is correlated with restoration of the mutant phenotype. The evidence points to PcaR being an activator of *pcaIJ* expression in *A. tumefaciens*.

Reversion of the ADO2069 mutant phenotype is very infrequent. Isolated revertants grow slowly at the expense of the protocatechuate precursor quinate. PCR amplifications of the *pcaR* gene along with the *pcaIJ-pcaR* intergenic region from three independent mutant strains were carried out. Gel analyses of the amplified DNA show that the strains still carry the *pcaR::* Ω mutation. It is likely that the secondary mutation carried by each pseudorevertant strain is an up promoter mutation. Enzymes assays of the strains and cycle sequencing of the *pcaI-pcaR* intergenic region of the prepared PCR reaction products should reveal the site of the mutation(s) and contribute information on the promoter region.

A common hallmark of regulatory genes is the ability to regulate their expression autogenously. Ligation of a promoterless *lacZ* reporter gene into *pcaR* resulted in a *pcaR::lacZ* transcriptional fusion in a broad host range plasmid, pARO177, which allowed quantitation of *pcaR* expression. The plasmid was introduced into the wild type strain A348 and *pcaR::* Ω strain

ADO2069. Figure 2 shows that *pcaR* is expressed constitutively at a low level, and its expression is repressed in the presence of a full length PcaR to about 30% of the level expressed in its absence.

Figure 2. Autoregulation of *pcaR*. A348 (parental strain) or ADO2069 (*pcaR::Ω*), carrying pARO177 (*pcaR::lacZ*) were grown in minimal medium under inducing and noninducing conditions for 21.5 hr. *pcaR::lacZ* expression was measured by β -galactosidase activity. Values, in Miller units, are the averages from three independent cultures.

As discussed in the MiniReview, the PcaR homologs are members of a small family of regulators. Two of these proteins also regulate enzymes involved in 4-hydroxybenzoate dissimilation: PobR and PcaU from *Acinetobacter calcoaceticus*. Their respective coeffectors are the phenolic compounds, 4-hydroxybenzoate and protocatechuate. Within the PobR subgroup, PobR and PcaU are more closely related to each other than the two PcaR proteins are to each other. All members of the PobR family share a highly conserved, relatively lengthy inverted repetition in the intergenic region, which has been proposed to be a promoter/operator site. The *A. tumefaciens* *pcaI-pcaR* intergenic region has elements of this motif, but the sequence is more divergent.

Situated between the *pcaR* and *pobA* genes is an open reading frame (ORF) with a deduced amino acid sequence which places it in the AraC/XylS family of regulators. This ORF has been analyzed in parallel to the analysis of

pcaR: an omega element was introduced near the N-terminal end of *pobR* in *A. tumefaciens*, autogenous regulation is being studied by introduction of a *pobR::lacZ* fusion in *trans*, and presumptive up promoter mutations have been isolated by PCR amplification.

Mutation of this locus eliminates the ability of cells to grow at the expense of 4-hydroxybenzoate although they are still able to grow using quinate or protocatechuate as carbon sources. *Trans* complementation experiments similar to those carried out with *pcaR* have been completed, and it appears that the ORF encodes an activator of *pobA* expression. Sequence analysis indicates that the AraC-type regulator of *PobA* synthesis has an independent evolutionary origin from the *Acinetobacter* *PobR* although both polypeptides interact with the same coeffector, 4-hydroxybenzoate. A small peptide shared by both polypeptides, but of independent origin, may have a common function in the two regulators. The regulator from *A. tumefaciens* is, however, homologous to presumed regulators of *PobA* synthesis identified in *Pseudomonas aeruginosa* and in *R. leguminosarum*, both of which have been called *PobR*.

A broad host range *pobA::lacZ* fusion plasmid has been constructed to study activation of the agrobacterial *pobA* promoter by *PobR*. One use of this plasmid will be to look at the distribution of the AraC-type *PobR* homologs in rhizobial strains. Although it is clear that *R. leguminosarum* has such a protein, it is not known if it will activate the agrobacterial promoter; if it does, extension of the AraC-type *PobR* to more divergent members of the *Rhizobiaceae* will solidify the body of information that points to conservation of regulatory molecules governing the *pca* and *pob* structural genes in members of this group.

The 4.2-kb region between the *pca* structural genes contains an unusual, compact constellation of genes encoding regulators belonging to three different families. Contrasting with the terminal palindrome structure that is predicted from the *pcaQ-pobA* intergenic sequence, there are no significant inverted repetitions in the *pcaR* or *pobR* sequences. The ORF's for *pcaR* and *pobR* have the further distinction of sharing a 28 nucleotide stretch of DNA at their 3'ends. This finding was initially based on the results of cycle sequencing of plasmid DNA cloned in *Escherichia coli*. Since it was possible that a deletion might have occurred in a plasmid cloned in *E. coli* or secondary structure(s) might have caused sequencing errors, symmetrical PCR in the

presence of 7-deaza-dGTP was used to amplify a region of DNA which encompassed *pcaR* and the 3' portion of the *pobR* gene. The nucleotide sequence of the deaza-dGTP-containing product was identical to the sequence of the plasmid-borne agrobacterial DNA obtained in its absence.

The PcaR protein from *A. tumefaciens* aligns closely with its *P. putida* homolog at the C-terminal end. By contrast, the PobR protein from *A. tumefaciens* contains 11 extra amino acids at its C-terminal end when it is aligned with AraC from *E. coli* K12 or more closely related homologs. Up to the point of divergence, the homologs share a highly conserved sequence. To help to dissect the 3' ends of the *pcaR* and *pobR* ORF's, an omega element is being introduced into a site that is 6 nucleotides beyond the 3' end of the *pcaR* gene.

Although gene rearrangements have occurred, DNA sequence comparisons show conservation of primary sequences within rearranged structural genes. *Agrobacterium* has been placed in the alpha subdivision of *Proteobacteria*, and comparison of translated sequences has revealed varying degrees of homology with the corresponding sequences from bacteria of the gamma subdivision (*A. calcoaceticus* and *P. putida*). Different degrees of constraints on protein structure are reflected in extremes of amino acid identity of aligned residues, exemplified by comparison of PobA and PcaD from *A. tumefaciens* and *A. calcoaceticus* which share 62% and 35% identity, respectively.

Members of *Bradyrhizobium* species, unlike their distant, faster growing relatives, express most of the enzymes associated with the protocatechuate pathway constitutively. Bradyrhizobial species appear to be particularly well-adapted to scavenge a diverse array of aromatic compounds, many of which have been identified in root exudates of their legume host plants. Given the diversity of aromatic compounds that support bradyrhizobial growth and the unusual regulatory pattern that characterizes the group, it was of interest to define more clearly the primary growth substrates that feed into the protocatechuate pathway and those that are dissimilated by independent routes. To this end, a spontaneous mutant strain of *Bradyrhizobium japonicum* I110 was isolated and characterized. The mutant strain, BDO100, is blocked in the first step of protocatechuate degradation, the dioxygenase encoded by *pcaHG*. Analysis of comparative growth yields of wild type and mutant strains revealed that cleavage of the

aromatic ring of *p*-coumarate, ferulate, 4-hydroxybenzoate, and vanillate occurs via the protocatechuate pathway, while independent pathways mediate catabolism of other aromatic compounds such as gallate, gentisate, and syringate. Our investigations also revealed that protocatechuate enzymes, synthesized constitutively in free-living bradyrhizobia, are expressed at a similar level in bacteroids found in root nodules. The sole regulated enzyme, protocatechuate dioxygenase, however, is expressed at a low level in bacteroids. Thus, enzymes of protocatechuate dissimilation do not appear to play a significant role in carbon and energy flux in bacteroids during symbiosis. Rather, it is likely that constitutivity of many of the enzymes evolved as a result of selective pressures exerted during saprophytic times.

Several complications made the above studies of *Bradyrhizobium* unexpectedly challenging. These included the discovery and definition of a secondary mutation that existed in strain BDO100, the necessity of developing novel methods to measure growth yields quantitatively in the face of variably toxic phenolic growth substrates, and other unanticipated results that apparently reflect unusual regulatory loops. In spite of these difficulties, a coherent analysis of pathways of phenolic catabolism in *Bradyrhizobium japonicum* was presented at the 1997 American Society of Microbiology meeting. This work was followed up with the isolation of additional mutant strains and the construction and screening of several complete gene libraries.

Research publications and presentations:

Parke, D. 1990. Construction of mobilizable vectors derived from plasmids RP4, pUC18 and pUC19. *Gene* 93: 135-137.

Parke, D., F. Rynne, and A. Glenn. 1991. Regulation of phenolic catabolism in *Rhizobium leguminosarum* biovar trifolii. *J. Bacteriol.* 173: 5546-5550.

Parke, D. 1992. Application of *p*-toluidine in chromogenic detection of catechol and protocatechuate, diphenolic intermediates in catabolism of aromatic compounds. *Appl. and Environ. Microbiol.* 58: 2694-2697.

Parke, D. 1993. Positive regulation of phenolic catabolism in *Agrobacterium tumefaciens* by the *pcaQ* gene in response to β -carboxy-*cis,cis*-muconate. J. Bacteriol. 175: 3529-3535.

Parke, D. 1993. Conservation of *pcaQ*, a regulatory gene required for phenolic catabolism, in *Rhizobium*, *Bradyrhizobium*, and *Agrobacterium*. 14th North American Conference on Symbiotic Nitrogen Fixation.

Parke, D. 1995. Supraoperonic clustering of *pca* genes for catabolism of the phenolic compound protocatechuate in *Agrobacterium tumefaciens*. J. Bacteriol. 177: 3808-3817.

Parke, D. 1995. Localization of *pobR*, a gene regulating synthesis of *p*-hydroxybenzoate hydroxylase, within a supraoperonic cluster of the *pca* genes in *Agrobacterium tumefaciens*. Abstract 95th Annual Meeting American Society for Microbiology. p. 27

Parke, D. 1996. Characterization of PcaQ, a LysR-type transcriptional activator required for catabolism of phenolic compounds, from *Agrobacterium tumefaciens*. J. Bacteriol. 178: 266-272.

Parke, D. 1996. Conservation of PcaQ, a transcriptional activator of *pca* genes for catabolism of phenolic compounds, in *Agrobacterium tumefaciens* and *Rhizobium* species. J. Bacteriol. 178: 3671-3675.

Parke, D. 1996. Characterization of the role of the *pcaR* gene in the expression of *pcaI*, the structural genes for β -ketoadipyl CoA transferase, in *Agrobacterium tumefaciens*. Abstract 96th Annual Meeting American Society for Microbiology. p. 558.

Parke, D. 1996. Organization and regulation of genes for catabolism of the phenolic compound, 4-hydroxybenzoate, in *Agrobacterium tumefaciens*. Abstract 17th Crown Gall Conference; Urbana, Illinois.

Parke, D. 1997. Acquisition, reorganization, and merger of genes: novel management of the β -ketoadipate pathway in *Agrobacterium tumefaciens*. FEMS Microbiol. Letts. 146: 3-12. (MiniReview)

Parke, D. 1997. Pathways for dissimilation of aromatic compounds in *Bradyrhizobium japonicum*. Abstract 97th Annual Meeting American Society for Microbiology. p. 349.

Manuscripts in preparation:

Parke, D. Characterization of PcaR, a transcriptional activator of *pcaI* expression in *Agrobacterium tumefaciens*.

Parke, D. Divergence of *pcaD*, the structural gene for β -ketoadipate enol-lactone hydrolase in *Agrobacterium tumefaciens*, from *pcaD* and *catD* encoding isofunctional enzymes in *Acinetobacter calcoaceticus*.

Parke, D. Independent evolutionary origins of PcaC and CatC, enzymes catalyzing analogous steps in parallel branches of the β -ketoadipate pathway.

Parke, D. Contiguity and 3' overlap of convergent ORF's for two regulatory genes, *pcaR* and *pobR*, required for dissimilation of 4-hydroxybenzoate in *Agrobacterium tumefaciens*.

Parke, D. Pathways of aromatic catabolism in *Bradyrhizobium japonicum*.