MICROBIAL & COMPARATIVE GENOMICS
Volume 2, Number 2, 1997
Mary Ann Liebert, Inc.

The AroQ and PheA Domains of the Bifunctional P-Protein from *Xanthomonas campestris* in a Context of Genomic Comparison

WEI GU,[1] DONNA S. WILLIAMS,[1] HENRY C. ALDRICH,[1] GARY XIE,[1] DEAN W. GABRIEL,[2] and ROY A. JENSEN[1]

ABSTRACT

The gene (denoted *aroQ$_p$·pheA*) encoding the bifunctional P-protein (chorismate mutase-P/prephenate dehydratase) from *Xanthomonas campestris* was cloned. *aroQ$_p$·pheA* is essential for L-phenylalanine biosynthesis. DNA sequencing of the smallest subclone capable of functional complementation of an *Escherichia coli* phenylalanine auxotroph revealed a putative open reading frame (ORF) of 1200 bp that would encode a 43,438-Da protein. AroQ$_p$·PheA exhibited 51% amino acid identity with a *Pseudomonas stutzeri* homologue and greater than 30% identities with AroQ$_p$·PheA proteins from *Haemophilus influenzae, Neisseria gonorrhoeae,* and a number of enteric bacteria. AroQ$_p$·PheA from *X. campestris,* when expressed in *E. coli,* possesses a 40-residue amino-terminal extension that is lysine-rich and that is absent in all of the AroQ$_p$·PheA homologues known at present. About 95% of AroQ$_p$·PheA was particulate and readily sedimented by low-speed centrifugation. Soluble preparations of cloned AroQ$_p$·PheA exhibited a native molecular mass of 81,000 Da, indicating that the active enzyme species is a homodimer. These preparations were unstable after purification of about 40-fold, even in the presence of glycerol, which was an effective protectant before fractionation. When AroQ$_p$·PheA was overproduced by a T$_7$ translation vector, unusual inclusion bodies having a macromolecular structure consisting of protein fibrils were observed by electron microscopy. Insoluble protein collected at low-speed centrifugation possessed high catalytic activity. The single band obtained via SDS-PAGE was used to confirm the translational start via N-terminal amino acid sequencing. A perspective on the evolutionary relationships of monofunctional AroQ and PheA proteins and the AroQ$_p$·PheA family of proteins is presented. A *serC* gene located immediately upstream of *X. campestris aroQ$_p$·pheA* appears to reflect a conserved gene organization, and both may belong to a single transcriptional unit.

Rationale for terminology: It is unfortunate that the early *Escherichia coli* auxotrophs for phenylalanine and tyrosine were denoted *pheA* and *tyrA* before it was known that each gene product was bifunctional, carrying two catalytic domains. Each domain corresponds to individual proteins in many other or-

[1]Department of Microbiology and Cell Science and [2]Department of Plant Pathology and Nematology, University of Florida, Gainesville, Florida.

Accession number: The GenBank accession number for the sequence reported in this article is U64518.

ganisms. It is becoming increasingly awkward, especially where multiple molecular-genetic comparisons are being made, to use acronyms whose meaning is different in different organisms. We, therefore, propose the following definitions. *pheA* encodes monofunctional prephenate dehydratase; *·pheA* encodes the prephenate dehydratase domain of a P-protein. *tyrA* encodes monofunctional prephenate dehydrogenase; *·tyrA* encodes the prephenate dehydrogenase domain of a T-protein. The AroQ family of chorismate mutase proteins is encoded by the following genes. Monofunctional chorismate mutase (commonly called chorismate mutase-F) is encoded by $aroQ_f$. The chorismate mutase domain of the bifunctional P-protein is encoded by $aroQ_p\cdot$, whereas the chorismate mutase domain of the bifunctional T-protein is encoded by $aroQ_t\cdot$. The chorismate mutase domain coexisting with a DAHP synthase domain on a bifunctional protein is encoded by $aroQ_d\cdot$. In addition to the AroQ family of chorismate mutase homologues, at least two additional families exist: the AroH family represented by *Bacillus subtilis* and cyanobacteria and the AroR family represented by yeast and higher plants. It is possible that the AroQ and AroR families may belong to a common superfamily (see text). The P-protein is called $AroQ_p$·PheA, and the T-protein is called $AroQ_t$·TyrA. The convention of using a bullet to separate the named domains of a fusion protein or to separate the corresponding coding regions is according to the precedent of Crawford with tryptophan-pathway genes (Crawford, 1989).

INTRODUCTION

The enzymatic chorismate-to-prephenate rearrangement by chorismate mutase reaction is a universal step of initial carbon flow commitment to biosynthesis of L-phenylalanine and L-tyrosine. In nature, subsequent enzymatic steps vary for the L-phenylalanine pathway (Jensen, 1992). Prephenate can be transaminated to L-arogenate via prephenate aminotransferase and then converted to L-phenylalanine via arogenate dehydratase. Alternatively, prephenate can be converted to phenylpyruvate via prephenate dehydratase and then transaminated to L-phenylalanine. Some organisms possess a broad-specificity enzyme, cyclohexadienyl dehydratase, which is competent for catalysis of both the prephenate dehydratase and arogenate dehydratase reactions.

Two of the three major divisions of gram-negative bacteria possess a bifunctional P-protein that carries two catalytic domains, one specifying chorismate mutase ($AroQ_p\cdot$) and another specifying prephenate dehydratase (·PheA). This must have arisen long ago via gene fusion before the divergence of the two gram-negative divisions (Ahmad and Jensen, 1986). Much more recently, a bifunctional T-protein emerged in enteric bacteria. This carries two catalytic domains, one specifying chorismate mutase ($AroQ_t\cdot$) (Subramaniam et al., 1994) and the other prephenate dehydrogenase (·TyrA). In gram-negative bacteria, chorismate mutase may also exist as a monofunctional protein ($AroQ_f$). $AroQ_f$ in at least some genera is processed for translocation to the periplasm (Xia et al., 1993a). Some enteric bacteria possess $AroQ_p\cdot$, $AroQ_t\cdot$, and $AroQ_f$ as coexisting isoenzyme species (Xia and Jensen, 1992).

The functional role of the $AroQ_p$ domain of microbial P-proteins is different depending on whether $AroQ_t$·TyrA is present or absent. In organisms such as *Neisseria gonorrhoeae,* which lack $AroQ_t\cdot$ (Subramaniam et al., 1994), the $AroQ_p$ domain must supply prephenate for both tyrosine and phenylalanine biosynthesis, as it is the only cytoplasmic chorismate mutase present. On the other hand, in the presence of a coexisting $AroQ_t$ ·TyrA (e.g., in enteric bacteria), the $AroQ_p$ domain presumably is dedicated exclusively to phenylalanine biosynthesis. In this context, the P-protein encoded by $aroQ_p$·*pheA* will yield interesting comparative information and an evermore informative evolutionary perspective as representative genes are cloned in apropriately spaced phylogenetic progressions. In addition to our work, $aroQ_p$·*pheA* has been cloned from enteric bacteria [*Escherichia coli* (Hudson and Davidson, 1984) and *Erwinia herbicola* (Xia et al., 1993b)], from the closely related *Haemophilus influenzae* (Fleischmann et al., 1995), and from the more distant *Pseudomonas stutzeri* (Fischer et al., 1991) and *N. gonorrhoeae* (Gonococcal Genome Sequencing Project) (www.genome.ou.edu/gono.html).

In this article, $aroQ_p$·*pheA* from *Xanthomonas campestris* was chosen for cloning because of its phylogenetic distance from other organisms from which $aroQ_p$·*pheA* has been sequenced. In addition, *X. campestris* represents an enzymologic patterning group of aromatic biosynthesis different from any of the

others (Whitaker et al., 1985). It lacks *aroQ·tyrA* (unlike enteric bacteria and *H. influenzae*) but possesses *pheC* (encodes cyclohexadienyl dehydratase) and $aroQ_f$ (unlike *P. stutzeri*).

MATERIALS AND METHODS

Bacterial strains, plasmids, and media

Bacterial strains and plasmids used in this study are listed in Table 1. *Xanthomonas* cultures were grown on tryptone, yeast extract, and MOPS medium at 28°C. Luria-Bertani (LB) medium was used as enriched

TABLE 1. BACTERIAL STRAINS AND PLASMIDS

Strain/plasmid	*Relevant characteristics*	*Source or reference*
Escherichia coli		
DH5α	F^-, ϕ80d*lac*ZΔM15, *rec*A1, *end*A1, *gyr*A96, *sup*E44, *hsd*R17 (r_K^-, m_K^+), Δ (*lac*ZYA-*arg*F) U169, *thi*-1, *rel*A1, *deo*R	Bethesda Research Laboratories
JM83	F^-, *ara*Δ (*pro*AB-*lac*), *rps*L, *thi*-1, ϕ80d*lac*ZΔM15	Bethesda Research Laboratories
KA197	*thi*-1, *phe*A97, *rel*A1, *spo*T1	CGSC 5243
BL21 (DE3)	F , *omp*T, $r_B^-m_B^-$	Novagen
Xanthomonas campestris pv. *translucens*		D. Gabriel
Plasmid		
pUC18	Ap^r, *lac*I′ POZ′	Yanisch-Perron et al., 1985
pUC19	Ap^r, *lac*I′ POZ′	Yanisch-Perron et al., 1985
pUFR043	cos, Nm^r, Gm^r	This study
pGEM-5Zf(+)	Ap^r, *lac*I′ POZ′	Promega
pET-11a	Ap^r, T_7*lac*	Novagen
pka06	Original clone of $aroQ_p$•*pheA* from *X. campestris* in pUFR043	This study
pJG-XP1	2.7-kb *Eco*R I-*Eco*R I fragment from pka06 subcloned in pUC18	This study
pJG-XP2a	1.7-kb *Pst* I-*Eco*R I fragment from pJG-XP1 subcloned in pU18	This study
pJG-XP2b	1.7-kb *Pst* I-*Eco*R I fragment from pJG-XP1 subcloned in pUC19	This study
pJG-XP3	1.0-kb *Pst* I-*Pst* I fragment from pJG-XP1 subcloned in pUC18	This study
pJG-XP4	1.5-kb *Pst* I-*Sac* II fragment from pJG-XP2a subcloned in pGEM-5Zf(+)	This study
pJG-XP5	1.2-kb PCR-generated $aroQ_p$•*pheA* subcloned in *Hinc* II site of pUC18	This study
pJG-XP6	1.2-kb *Nde*I-*Bam*H I fragment from pJG-XP5 subcloned in pET-11a	This study

medium for *E. coli* at 37°C (Silhavy et al., 1984). The minimal salts medium used for *E. coli* strains was the M-9 formulation (Maniatis et al., 1989). Where indicated, ampicillin was added to media at 100 μg/ml, gentamicin at 3 μg/ml, thiamine at 17 μg/ml, and X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) at 40 ng/ml. Medium was solidified with 2% (w/v) agar.

DNA manipulation

Standard molecular biology procedures were performed as described by Maniatis et al. (1989) unless otherwise indicated. A genomic library of *X. campestris* was constructed by partially digesting genomic DNA with *Sau*3A I and cloning into the *Bam*H I site of cosmid pUFR043. Restriction enzymes, ligase, and calf intestine alkaline phosphatase were purchased from New England Biolabs of Promega and were used according to manufacturer's instructions.

DNA sequencing and data analysis

Plasmid pJG-XP2a was purified by the method recommended by Applied Biosystems, Inc. (Applied Biosystems, Inc.). Double-stranded plasmid DNA was sequenced in both directions at the DNA Sequencing Core Facility of the Interdisciplinary Center of Biotechnology Research (ICBR) at the University of Florida. In order to sequence DNA beyond the range of the annealing sites for the M13 universal and reverse primers, which flank the multiple cloning site of the pUC18 vector, 18-mer oligonucleotides were synthesized to regions upstream of the unsequenced regions at the DNA Synthesis Core Laboratory of ICBR. The nucleotide sequence and the deduced amino acid sequence were analyzed by using the updated version of the sequence analysis software package offered by Genetics Computer Group, Inc. (1995).

Crude extract preparation and enzyme assays

Crude extract was prepared as described previously (Gu et al., 1995). Chorismate mutase and prephenate dehydratase activities of the cloned *aroQ$_p$*·pheA gene product were assayed by a modification (Ahmad and Jensen, 1988) of the method of Cotton and Gibson (1965). Protein concentration was determined by the method of Bradford (1976).

DEAE-cellulose chromatography

Crude extract (1.76 g) that had been prepared in a standard buffer containing 50 mM potassium phosphate buffer (pH 7.0), 1 mM L-tyrosine, 1 mM β-mercaptoethanol, 0.1 mM phenylmethylsulfonyl fluoride (PMSF), and 20% (v/v) glycerol was precipitated with $(NH_4)_2SO_4$ at 50% of saturation. Protein taken up in buffer was dialyzed against 2000 ml of buffer overnight with one change of the same buffer. A 640-mg amount of protein was applied to a 2.5 × 50 cm DE-52 column equilibrated in buffer. The column was washed with 280 ml of buffer, and bound protein was then eluted with a 1000-ml linear salt gradient (0–0.3 M KCl) prepared in the same buffer. Fractions (5 ml) were collected and assayed for both chorismate mutase and prephenate dehydratase activities.

Hydroxylapatite chromatography

A 30-mg amount of protein partially purified from DE-52 was loaded onto the bed of a 1.5 × 20 cm column of Bio-Gel HTP previously equilibrated with buffer. The column was washed with 60 ml of starting buffer and then eluted by application of a gradient of between 200 mM and 600 mM of potassium phosphate containing the other supplements of standard buffer except PMSF. Fractions (2 ml) were collected, and both activities were assayed.

Overexpression of aroQ$_p$·pheA

Two 29-mer oligonucleotides were synthesized at the DNA Synthesis Core Laboratory of ICBR (5′-ATGCGCTGCTCAGGGAATGGCGACCGGAT-3′ and 5′-CCC**ATATG**GCTCCCAAGCCCAAGAACACC-3′) that were appropriate for annealing to regions embracing the entire *aroQ$_p$·pheA* open reading

frame (ORF). The latter oligonucleotide contains an *Nde*I site (indicated with bold letters) overlapping the start codon of $aroQ_p$·*pheA*. The 1200-bp $aroQ_p$·*pheA* fragment was amplified by PCR (Perkin Elmer Cetus, Norwalk, CT) and ligated into the *Hinc* II site of pUC18 (pJG-XP5). The $aroQ_p$·*pheA* fragment was then released by double digestion with *Nde* I and *Bam*H I and ligated into pET-11a (pJG-XP6), a translation overexpression vector (Novagen, Inc.). Thus, $aroQ_p$·*pheA* was placed under control of a T_7 promoter and a well-designed Shine-Dalgarno sequence.

Plasmid pJG-XP6 was transformed into *E. coli* BL21 (DE3). A loopful of culture from the transformed plate was inoculated into 10 ml of LB medium containing ampicillin and shaken at 37°C until the OD_{600} reached 0.6–1.0. IPTG was added to a final concentration of 1 mM, and the incubation was continued for 3 h. The resulting culture could be checked directly for $aroQ_p$·*pheA* overexpression by SDS-PAGE.

Western blot analysis

The $AroQ_p$·PheA samples were subjected to SDS-PAGE. After electrophoresis, the gels were blotted onto nitrocellulose membranes (Towbin et al., 1979). The blots were then treated with the antibody raised against the P-protein from *Acinetobacter calcoaceticus* (Ahmad et al., 1988), followed by goat anti-(rabbit IgG) alkaline phosphatase conjugate. Nitroblue tetrazolium chloride (NBT) and 5-bromo-4-chloro-3-indolylphosphate *p*-toluidine salt (BCIP) were used as chromogenic substrates for alkaline phosphatase.

Amino acid sequencing of the cloned P-protein

The IPTG-induced culture was harvested by centrifugation and resuspended in 50 mM potassium phosphate buffer (pH 7.0). The cells were broken by sonication, and 500 μl of the sample was transferred into a microcentrifuge tube and spun for 5 min at 200*g*. The pellet was resuspended in the same phosphate buffer. A 10-μl portion was mixed with 10 μl of SDS-PAGE buffer, boiled for 4 min, and subjected to SDS-PAGE. The protein was then transferred to a polyvinylidene difluoride membrane by a protein miniblotting apparatus and sequenced by using an Applied Biosystems model 470A protein sequencer with an online 120A phenylthiohydantoin analyzer at the Protein Core Facility at ICBR.

Molecular mass determination

The molecular mass of the native enzyme was estimated by gel filtration using an FPLC-connected Superdex 75 HR 10/30 column (Pharmacia Biotech), which had been previously equilibrated with 50 mM potassium phosphate buffer (pH 7.0) containing 1 mM tyrosine and 0.15 M NaCl. The same buffer was used for elution. The column was calibrated with cytochrome C (12,400), carbonic anhydrase (29,000), bovine serum albumin (BSA) (66,000), and alcohol dehydrogenase (150,000). V_o was determined by blue dextran (2,000,000). SDS-PAGE was carried out to determine the subunit molecular mass of the enzyme. Lysozyme (14,400), soybean trypsin inhibitor (21,500), carbonic anhydrase (31,000), ovalbumin (45,000), BSA (66,000), and phosphorylase b (97,400) were used as molecular mass standards.

Electron microscopy

To study the morphology of inclusion bodies within intact cells, an IPTG-induced *E. coli* BL21 (DE3)/pJG-XP6 culture was pelleted in a microfuge, resuspended in 2.5% glutaraldehyde/2.5% formaldehyde in 0.1 M sodium cacodylate buffer (pH 7.2) and immediately subjected to microwave fixation for 30 sec to a maximum temperature of 32°C (Aldrich et al., 1996). The pellet was then fixed for 30 min at 4°C in 1% OsO_4, followed by soaking in 0.5% aqueous uranyl acetate overnight at 4°C and embedding in Spurr's low-viscosity resin. Thin sections were poststained for 5 min on lead citrate.

Pellets obtained following cell lysis by use of a French press were fixed for 30 min at ambient temperature in cacodylate-buffered 2.5% glutaraldehyde/2.5% formaldehyde/0.1% tannic acid (Aldrich et al., 1996) and postfixed for 30 min in OsO_4 and in 1% aqueous uranyl acetate for 30 min at ambient temperature before embedding in Epon. Thin sections were poststained for 5 min on lead citrate.

Cells processed for immunogold labeling were fixed for 15 min at 4°C in 0.5% glutaraldehyde/4% formaldehyde, dehydrated through cold ethanols, and embedded in Lowicryl HM-20 resin, which was poly-

merized under an ultraviolet lamp at −10°C. Thin sections were picked up on Formvar-coated nickel grids, blocked for 10 min on 1% nonfat dry milk, floated on primary rabbit polyclonal antiserum for 12 h at 4°C, and then floated on secondary antibody (sheep antirabbit, Sigma) carrying a 12-nm gold label. All solutions were buffered with either phosphate-buffered saline or high-salt TRIS at pH 7.2 (Aldrich et al., 1992). These buffers were also used for washing. Lowicryl sections require very short poststain times. We used durations of 30 sec on 1% aqueous uranyl acetate and 30 sec on lead citrate. All microscopy was performed on a Zeiss EM-10 CA transmission electron microscope.

Biochemicals and chemicals

Ampicillin, gentamicin, thiamine, α-ketoglutarate, PMSF, amino acids, and goat anti-(rabbit IgG) alkaline phosphatase were purchased from Sigma Chemical Company, NBT and BCIP were from Gibco-BRL, nitrocellulose membranes were from Micron Separations Inc., DEAE-cellulose was from Whatman, and hydroxylapatite was from Bio-Rad. Molecular mass standards for gel filtration and for SDS-PAGE were from Sigma and Bio-Rad, respectively. Growth medium components and agar were from Difco. Prephenate and chorismate were prepared by this laboratory.

RESULTS

Cloning of $aroQ_p$·pheA *in* E. coli

A random cosmid library generated via a partial digestion by *Sau*3AI of *X. campestris* genomic DNA was constructed in *E. coli*. Six clones were obtained that carried plasmid inserts able to complement the L-phenylalanine auxotrophy of *E. coli* KA197 when cultured on minimal medium with gentamicin to select for maintenance of plasmid. The selection strategy could have yielded either *pheC* or $aroQ_p$·*pheA*. Expression of *pheC* (encoding cyclohexadienyl dehydratase) would yield crude extracts showing prephenate dehydratase and arogenate dehydratase activities, whereas expression of $aroQ_p$·*pheA* would produce prephenate dehydratase activity only. Clone pka06, on these criteria, appeared to possess an $aroQ_p$·*pheA* insert. This was readily confirmed when anion-exchange chromatography (DEAE-cellulose) produced a fraction containing coeluting chorismate mutase and prephenate dehydratase activities.

The pka06 clone contained a fragment of about 40 kb inserted in the *Bam*H I site of the cosmid pUFR043. On digestion with *Eco*R I, fragments of 2.7 kb, 4.3 kb, >15 kb, and >23 kb were obtained. When the 2.7-kb fragment was ligated into the *Eco*R I site of pUC18, it was successfully used for functional complementation of *E. coli* KA197. Enzymatic assays confirmed the presence of $aroQ_p$·*pheA* in the cloning strain KA197/pJG-XP1. Further subcloning experiments localized *pheA* within the *Pst* I-*Sac* II fragment of pJG-XP4 (Fig. 1). A comparison of the results obtained with pJG-XP2a and pJG-XP2b, which have opposite insert orientations, indicated that expression depends on the *lac* promoter of the vector. Thus, either a native promoter is absent in the insert or the native promoter is not recognized by *E. coli*.

Sequence analysis of X. campestris $aroQ_p$·pheA

Figure 2 shows the nucleotide sequence of $aroQ_p$·*pheA*, as well as some of the proximal flanking regions. The 1200-bp ORF begins at codon ATG and terminates at codon TGA. The G + C content of $aroQ_p$·*pheA* (62.1%) is just outside the 63%–72% range of *X. campestris*. The codon usage of $aroQ_p$·*pheA* is similar to that of rRNA-homology group I pseudomonads (such as *Pseudomonas aeruginosa*). The deduced amino acid sequence results in a 400-residue protein having a molecular mass of 43,428 Da. The highest pairwise amino acid identity (51%) is with the *P. stutzeri* homologue, a result consonant with the phylogenetic proximity of *X. campestris* and *P. stutzeri*.

A partial ORF encoding 27 residues was found upstream of $aroQ_p$·*pheA*. This gene is *serC* (encodes phosphoserine aminotransferase) in view of the high amino acid sequence identity (>50%) with the 3′-termini from *serC* genes of both *Salmonella gallinarum* (Griffin and Griffin, 1991) and *Yersinia enterocolitica* (O'-Gaora et al., 1989). A factor-independent terminator structure is positioned at the immediate 5′-side of the ri-

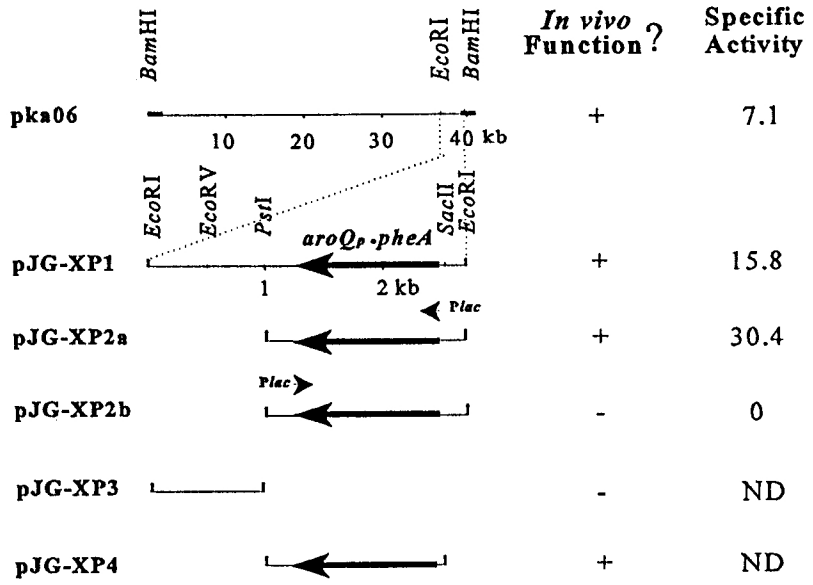


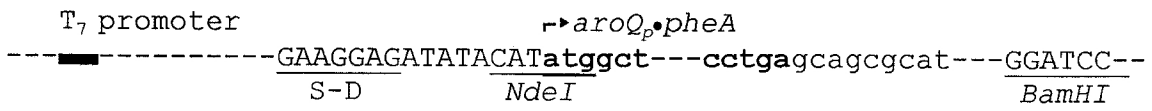
FIG. 1. Physical map and subclone analysis of *X. campestris aroQ$_p$·pheA*. Thin horizontal line, *X. campestris* DNA; thick horizontal line, vector DNA. The orientation of the *lac* promoter in the pka06 vector is shown for the pJG-XP2a and pJG-XP2b subclones. In all other constructs shown (including pka06), the orientation of the *lac* promoter (P*lac*) is the same as shown for pJG-XP2a. Positive (+) in vivo function refers to the ability of a given clone to complement the L-phenylalanine auxotrophy of *E. coli* KA197. Specific activities of prephenate dehydratase are given as nmol min^{-1} mg^{-1} at right.

```
-▸serC
GCGCTTCGCTGTACAACGCCATGCCGGTGGCCGGTGCGCAGGCACTGGTGGCCTTCATGCACGATTTCCAGCAGCGCCACGGCTGAGTCCCCCACAACCGCGGCATTGCCGCGCTGGCGC   120
  A  S  L  Y  N  A  M  P  V  A  G  A  Q  A  L  V  A  F  M  H  D  F  Q  Q  R  H  G  *                    >> >>>>>>> <
                                  ⌐▸aroQp domain
AGTGCGCCGTTTTATTGAAGGAAAGCCCCCCGATGGCTCCCAAGCCCAAGAACACCAACGCCGCTGGCGGCACCGCCAAGTCGGCTACCAAGGCCAGCAGCAAGAAAGCCACCAAGTTGG   240
<<<<<< <<        rbs          M  A  P  K  P  K  N  T  N  A  A  G  G  T  A  K  S  A  T  K  A  S  S  K  K  A  T  K  L  A

CTGGCAAGGACAGCGCCAAGCCACTGGCAACCGCTGCCCCAGTGCTGGCCGATGTGCGTGCCAAGATCGACGAGATCGACCGTGGCATCCAGGCGTTGATCGCCGAGCGCGCCAACTTCG   360
  G  K  D  S  A  K  P  L  A  T  A  A  P  V  L  A  D  V  R  A  K  I  D  E  I  D  R  G  I  Q  A  L  I  A  E  R  A  N  F  A

CGCATCAGGTCGGCAAGGCCAAGGGCAAGCTGGCAGCGGCGGTGGATTACTACCGCCCCGAGCGCGAAGCGCAGGTGCTGCGCATGGTGGTGGACCGCAACGAAGGACCACTCAGCGATG   480
  H  Q  V  G  K  A  K  G  K  L  A  A  A  V  D  Y  Y  R  P  E  R  E  A  Q  V  L  R  M  V  V  D  R  N  E  G  P  L  S  D  E
                                                                ⌐▸pheA domain
AAGTGCTGGTGCATGTGTTCCGCGAAATCATGTCTGCGTGCCTTGCGCAGCAGGAGCCGCTGAAGATCGGGTATCTCGGCCCGGAAGGCACCTTCAGCCAGCAGGCGGTGCTCAAGCATT   600
  V  L  V  H  V  F  R  E  I  M  S  A  C  L  A  Q  Q  E  P  L  K  I  G  Y  L  G  P  E  G  T  F  S  Q  Q  A  V  L  K  H  F

TCGGCCGGTCGGCGGTGGGTCTGCCGATGGCGACCATCGAAGAAGTGTTTCAGGAAGTGGAAGCCGCCAATGCCGATTTCGGCGTGGTGCCGGTAGAAAATTCCGGCCAGGGCACGATCC   720
  G  R  S  A  V  G  L  P  M  A  T  I  E  E  V  F  Q  E  V  E  A  G  N  A  D  F  G  V  V  P  V  E  N  S  G  Q  G  T  I  Q

AGGTCACGCTGGACATGTTCCTCACCTCCAACCTCAAGATCTGCGGCGAAGTCGAGTTGCGTGTGCATCAGTACCTGCTTTCGCGCAACGGGCGCCTGGAAGACATCGAACGCATCTACG   840
  V  T  L  D  M  F  L  T  S  N  L  K  I  C  G  E  V  E  L  R  V  H  Q  Y  L  L  S  R  N  G  R  L  E  D  I  E  R  I  Y  A

CGCATTCGCAGTCGTTTGCGCAGACCGCCGGCTGGCTGCGTTCGCACTTGCCCAAGGTGGAAAAGATTGCGGTCTCCAGCAATGCCGAAGGCGCGCGCCGTGCGCGTAATGCCGAGGACG   960
  H  S  Q  S  F  A  Q  T  A  G  W  L  R  S  H  L  P  K  V  E  K  I  A  V  S  S  N  A  E  G  A  R  R  A  R  N  A  E  D  A

CGGCGGCCATCGGTGGCGAAAGCGCCGCGCACGTGTATGGCTTGAAAAAAGTCATCATGAAATCGATCGAGGACGACGACGACAACACCACGCGCTTTCTGGTGATCGGCCGGCAGATCT  1080
  A  A  I  G  G  E  S  A  A  H  V  Y  G  L  K  K  V  I  M  K  S  I  E  D  D  D  D  N  T  T  R  F  L  V  I  G  R  Q  I  F

TTCCATCGTCCGGGCATGACCGGACCTCGGTGCTGGTGTTCATCCACGACAAGCCGGGCGCGTTGTTCGATGTACTCAGCCCGTTCGCGCGGCATGGCATCAGCATGAACCGCATCGAGT  1200
  P  S  S  G  H  D  R  T  S  V  L  V  F  I  H  D  K  P  G  A  L  F  D  V  L  S  P  F  A  R  H  G  I  S  M  N  R  I  E  S

CGCGGCCGTCGCACCAGGCCAAGTGGGAATACGGCTTCTTTATCGACCTGGCCGGCCATGTGGAGGACGAGTCGATGAAACAGGCACTGGCCGAACTGGAAGCGCATTCGGCACAGATCA  1320
  R  P  S  H  Q  A  K  W  E  Y  G  F  F  I  D  L  A  G  H  V  E  D  E  S  M  K  Q  A  L  A  E  L  E  A  H  S  A  Q  I  K

AGGTGCTGGGCTCGTATCCGGTCGCCATTCCCTGAGCAGCGCATCGGCTGCGGCGCCGCGTCGGCCCTGACGCGGCGCCTGTCTCTACACCGCCTTCCTTCCCACGCGCATTGCTGACCG  1440
  V  L  G  S  Y  P  V  A  I  P  *

CAGCGATGTCGTCGCAATCGGATCATTCGCATGAGCAACAGCACGCAACACTGGATCGCACAGCGCGGCACCGCGCTGCAG                                        1548
```

FIG. 2. Nucleotide sequence of *X. campestris aroQ$_p$·pheA* and its flanking regions. The probable ribosome binding site (rbs) is underlined. The 27-amino acid carboxy end of SerC is shown at the top. Converging arrowheads mark a terminator structure between *serC* and *aroQ$_p$·pheA*. The first 10 amino acids of AroQ$_p$·PheA (bold) were confirmed by N-terminal amino acid sequencing. The span of domains corresponding to *aroQ$_p$* and to *·pheA*, as marked, were estimated through multiple alignment analysis with monofunctional AroQ$_f$ and monofunctional PheA gene products. Nucleotide numbers assigned are shown at right.

A

Overexpression vector (pJG-XP6):



B

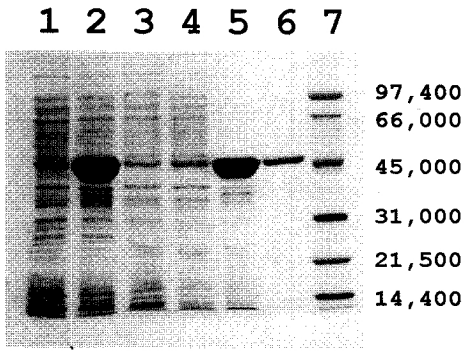


FIG. 3. Overexpression of *aroQ$_p$·pheA* from *X. campestris* in *E. coli* **(A)** Schematic representation of overexpression vector pJG-XP6. Uppercase letters represent the vector, lowercase letters represent the inserted *X. campestris* nucleotides, and bold lowercase letters represent *aroQ$_p$·pheA*. The *Nde*I restriction site connects the vector and the 5′-end of the *aroQ$_p$·pheA* insert, and the *Bam*H I restriction site is provided by the vector. **(B)** SDS-PAGE analysis of *aroQ$_p$·pheA* expression in the following extracts. Lane 1, BL21 (DE3)/pJG-XP6 (−IPTG); lane 2, BL21(DE3)/pJG-XP6 (+IPTG); lane 3, BL21(DE3)/pET11a (+IPTG); lane 4, 200*g* supernatant from BL21(DE3)/pJG-XP6; lane 5, 200*g* pellet from BL21(DE3)/pJG-XP6; lane 6, 200*g* pellet from BL21(DE3)/pJG-XP6 at 10-fold dilution; lane 7, molecular weight standards. Lanes 1 and 2 were loaded with about 20 μg of protein; lanes 3–5 were loaded with 10 μg of protein; lane 6 was loaded with 1 μg of protein.

bosome-binding site for *aroQ$_p$·pheA*. Thus, because of space constraints, the promoter used for *aroQ$_p$·pheA* must be within *serC* or further upstream.

Expression of X. campestris aroQ$_p$·pheA *in* E. coli

AroQ$_p$·PheA proved to be quite labile. Under various conditions where instability was encountered, both chorismate mutase and prephenate dehydratase activities were affected about equally. Crude extracts containing soluble AroQ$_p$·PheA were prepared from KA197/pJG-XP2a. Preliminary purification led to complete loss of activity after the first step of several different fractionation protocols attempted. Therefore, conditions favoring stability were determined by testing additives that might protect AroQ$_p$·PheA activities in crude extract from thermal inactivation. Crude extract in 50 mM K-phosphate buffer at pH 7.0 was found to be subject to rapid inactivation at temperatures over 45°C. On the basis of protection conferred at an assay temperature of 48°C, the components of the standard buffer (Materials and Methods) were adopted for routine use.

AroQ$_p$·PheA activities expressed in the *E. coli* clone were estimated to be 10 times greater than in the native organism. Sequential fractionation steps using $(NH_4)_2SO_4$ at 50% of saturation, DEAE-cellulose chromatography, and hydroxylapatite chromatography produced an additional overall purification of about 42-fold. SDS-PAGE analysis showed multiple bands. Any subsequent fractionation step resulted in total loss

of activity. Attempts to purify $AroQ_p$·PheA with protocols employing a different order of fractionation steps were no more successful. It, therefore, appeared that extreme lability resulted once a purification of about 40-fold was achieved.

In one experiment, a partially purified preparation of $AroQ_p$·PheA from KA197/pJG-XP2a [after steps of $(NH_4)_2SO_4$ precipitation and DEAE-cellulose chromatography] was applied to a Sephadex G-150 column to estimate the native molecular mass (data not shown). The active enzyme species is a homodimer, judging from the molecular mass determined (81,000 Da). In parallel gel filtration runs where either 1 mM L-tyrosine or 1 mM L-phenylalanine was present, 81,000-Da homodimers were again eluted.

Overexpression of $AroQ_p$·PheA in E. coli

PCR was used to generate an *$aroQ_p$·pheA* insert for placement into the translation overexpression vector pET-11a. This construct (pJG-XP6) eliminated all native DNA sequence located upstream of the ATG start codon (Fig. 3A). pJG-XP6 was transformed into *E. coli* BL21(DE3), and a culture was supplemented with IPTG to induce the chromosomally encoded T_7 RNA polymerase. Chorismate mutase and prephenate dehydratase activities in crude extracts prepared from these cultures increased over background only about sevenfold. However, the pellet fraction obtained by the routine ultracentrifugation used to prepare crude extracts exhibited a prominent protein band at a M_r position of 43,000 Da following SDS-PAGE. The particulate $AroQ_p$·PheA could be collected by low-speed centrifugation (200*g*). Growth and IPTG induction at 25°C did not increase the proportion of soluble $AroQ_p$·PheA recovered. When evaluated by SDS-PAGE, $AroQ_p$·PheA in the pellet fractions represented more than 90% of total protein (Fig. 3B). N-terminal amino acid sequencing of the excised band confirmed the identity of the protein as $AroQ_p$·PheA. The enzyme activity assayed in the particulate fractions was extremely high, more than 400-fold greater than that previously obtained with the KA197/pJG-XP2 construct. The enzyme activity assayed in the particulate fractions cannot be attributed to a contaminating aqueous component, as passage through a Gelman Supor Acrodisc-13 syringe filter removed all enzyme activity.

Lack of sensitivity to allosteric effectors

None of our $AroQ_p$·PheA preparations exhibited significant sensitivity to inhibition by L-phenylalanine or to activation by L-tyrosine. Figure 4 shows results of an experiment in which a substrate saturation curve

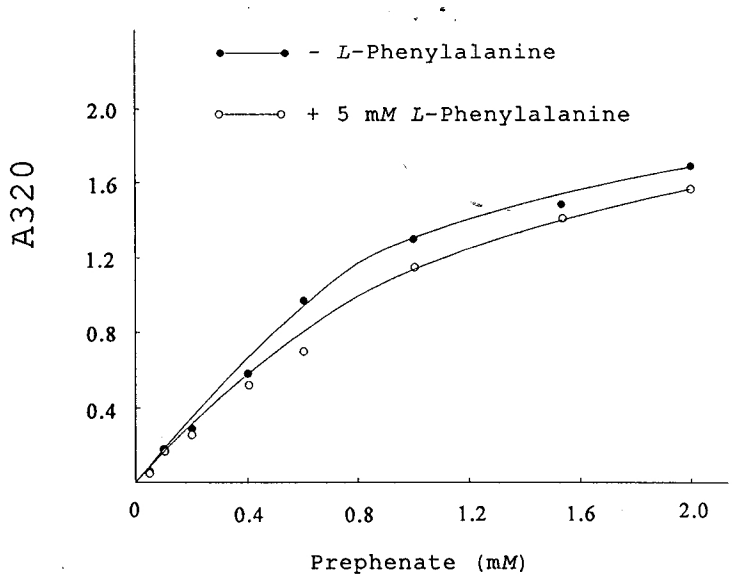


FIG. 4. Substrate saturation curve of prephenate dehydratase activity from *X. campestris* $AroQ_p$·PheA in the presence or absence of L-phenylalanine. Data shown were obtained from the particulate fraction recovered from BL21(DE3)/pJG-XP6.

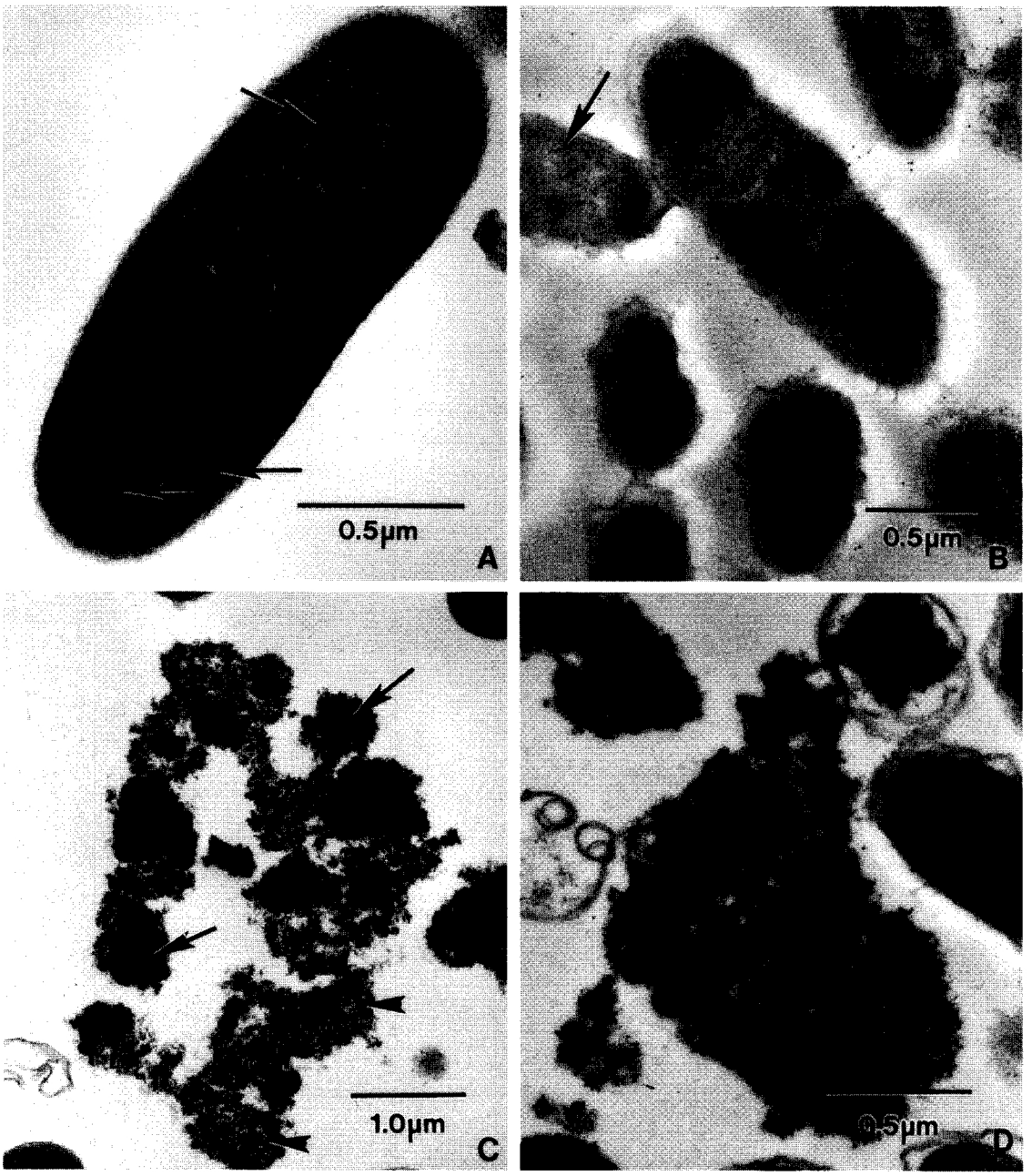


FIG. 5. Visualization of inclusion bodies by electron microscopy in IPTG-induced *E. coli* BL21(DE3)/pJG-XP6. **(A)** Transmission electron micrograph of *E. coli* containing P-protein inclusion bodies from *X. campestris* (arrows). This section was cut intentionally thick (ca. 100 nm) and photographed at 100 kV accelerating voltage to illustrate the convoluted, swirled nature of the protein fibrils that comprise the inclusion bodies. **(B)** Immunogold labeling of dark-staining inclusion bodies within intact cells, demonstrating their identity as P-protein. Heavy gold label appears throughout the dark bodies, and less intense label appears in lighter staining areas. Occasional cells that exhibit little or no labeling (arrow) contain little or no P-protein, probably because of plasmid loss. **(C)** Section of pellet produced by lysis of cells by French press. As in intact cells shown in **B,** some material stains very heavily (arrows), whereas other regions stain lighter (arrowheads). **(D)** Immunogold labeling of preparation similar to that shown in **C.** Gold label appears in both lighter and darker staining areas of isolated inclusion body material in center of field. Wall fragment at left does not label. Intact cells contain label over most of the cytoplasm.

for prephenate dehydratase was constructed in the presence of 5 mM L-phenylalanine. No inhibition was seen, even at low substrate concentration. Chorismate mutase activity was similarly insensitive (data not shown). Soluble enzyme preparations from KA197/pJG-XP2a yielded similar results. Activation of $AroQ_p$·PheA by L-tyrosine was not observed.

Immunologic cross-reactivity

Antibody raised against *A. calcoaceticus* $AroQ_p$·PheA previously had been found (Ahmad et al., 1988) to cross-react strongly with homologues from *P. stutzeri* and *P. aeruginosa* but only marginally with the homologues from several enteric bacteria. We found that anti-*A. calcoaceticus* $AroQ_p$·PheA antibody exhibited considerable cross-reactivity with *X. campestris* $AroQ_p$·PheA. Western blot hybridization visualized a distinctive band at an M_r position of 43,000 Da.

Visualization of inclusion bodies

Although the insoluble character of $AroQ_p$·PheA suggested that it was probably sequestered in inclusion bodies, we considered the possibility that the positively charged amino-terminus might be tightly associated with a negatively charged particulate entity, such as cell wall debris. However, the elaboration of inclusion bodies was readily confirmed by electron microscopy (Fig. 5). The inclusion bodies were atypical in not being round and refractile. They often occupied a substantial fraction of the cell volume, were not membrane bound, and exhibited a macromolecular structure manifesting spaghetti-like formations (Fig. 5A). Immunogold cytolocalization using the *A. calcoaceticus* antibody confirmed the synonymy of $AroQ_p$·PheA and the inclusion bodies (Fig. 5B). Cell-free extracts contained macromolecular aggregations resembling the intracellular inclusion bodies (Fig. 5C). Immunogold cytolocalization again confirmed their identity as $AroQ_p$·PheA (Fig. 5D).

DISCUSSION

Gene organization and regulation

In enteric bacteria (*E. coli* and *E. herbicola*), $AroQ_p$·PheA is regulated by attenuation, which depends on an L-phenylalanine-rich leader peptide encoded by *pheL* (Hudson and Davidson, 1984; Xia et al., 1993). *X. campestris* lacks such an upstream gene encoding a small leader peptide. However, the limited intercistronic space indicates that the promoter for *$aroQ_p$·pheA* must be ahead of or within the upstream gene, *serC*. Alternative stem-loop structures exist between the two genes (Fig. 6), which may reflect a mechanism of reg-

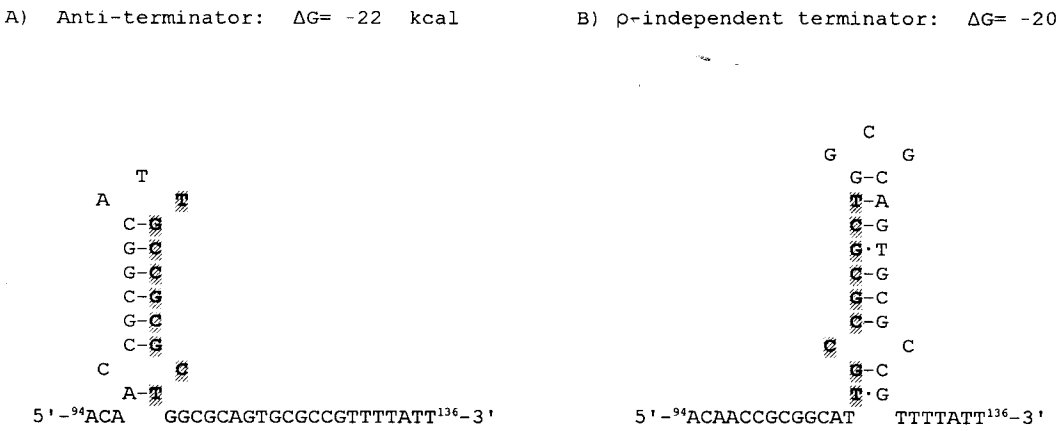


FIG. 6. Alternative stem-loop structures between *X. campestris serC* and *$aroQ_p$·pheA*. The shaded base residues are common to both the antiterminator structure shown in **A**, and to the factor-independent terminator structure shown in **B.**

ulation by attenuation. The *serC-pheA* gene organization is also maintained in the relatively close phylogenetic neighbor, *P. stutzeri*. In the latter case, there is no intercistronic space, and *serC* and *aroQ$_p$·pheA* are translationally coupled (Fischer et al., 1991). Again, alternative stem-loop structures exist, one of which sequesters the *aroQ$_p$·pheA* ribosome binding site. Thus, yet another form of attenuation may be used.

In enteric bacteria, *serC* and *aroE* are organized as a mixed-function operon and transcribed from a common promoter (Duncan and Coggins, 1986; Griffin and Griffin, 1991; O'Gaora et al., 1989). In *H. influenzae, serC* may be organized (Fleischmann et al., 1995) with the neighboring *hisH,* encoding a broad-specificity aminotransferase that can transaminate aromatic amino acids (Jensen and Gu, 1996). It is intriguing that in *P. stutzeri* the gene immediately downstream of *serC/aroQ$_p$·pheA* is *hisH* (T. Xie and R.A. Jensen, unpublished observations). Thus, for unknown reasons, *serC* appears to be commonly organized within mixed-function operons in which the second gene participates in aromatic amino acid biosynthesis.

AroQ$_p$·PheA from X. campestris

When expressed in *E. coli,* AroQ$_p$·PheA from *X. campestris* was obtained as an approximately 1:20 mixture of soluble and particulate forms. SDS-PAGE of both forms indicated an identical subunit M_r. The soluble AroQ$_p$·PheA (see below) migrated during gel filtration as a homodimer in the presence or absence of either L-tyrosine or L-phenylalanine.

Protein aggregates that form inclusion bodies are generally considered to consist of nonnative molecules. If cloned AroQ$_p$·PheA consists of misfolded molecules, competent catalytic centers of both domains must be exposed because high activity is obtained. On the other hand, allosteric sites within the carboxy portion of ·PheA (see below) may be masked, as L-tyrosine or L-phenylalanine failed to affect activity. However, the physical state of insolubility per se does not uniquely correlate with insensitivity to allosteric and aggregative effects of L-phenylalanine or L-tyrosine. In fact, the fraction of total activity that is soluble also exhibited a complete lack of allosteric responsiveness.

The 40-amino acid extension at the amino-terminus of *X. campestris* AroQ$_p$·PheA is so far unique among the fairly substantial family of homologues currently available. Its high lysine content confers a potent positive charge at the amino-terminus. Water molecules retained by this hydrophilic region might provide an aqueous layer that could account for the enzyme activity of the insoluble aggregate of *X. campestris* AroQ$_p$·PheA. The positively charged terminus could also act as a funnel to attract chorismate molecules (with their two negatively charged carboxyl moieties) to the specific microenvironment of the AroQ$_p$ catalytic site. Alternatively, the lysine-rich region may allow the N-terminal domain to remain largely solvated, thereby preventing collapse into inactive, nonnative aggregates dominated by hydrophobic interactions.

The gene fusion responsible for contemporary AroQ$_p$·PheA proteins was an ancient event. A considerable diversity of physical and catalytic properties of AroQ$_p$·PheA proteins throughout the β and γ divisions of gram-negative bacteria has emerged. These include quaternary structure of the native protein and influence of effectors on it. For example, in *E. coli* and *E. herbicola,* L-phenylalanine (but not L-tyrosine) promotes tetramerization of the native homodimer (Hudson and Davidson, 1984). In *A. calcoaceticus,* both L-tyrosine and L-phenylalanine promote tetramerization of a native homodimer (Ahmad et al., 1988; Berry et al., 1985). In *Alcaligenes eutrophus,* a native homotetramer was found to persist in the presence or absence of any effector molecules (Friedrich et al., 1976). The future cloning of *A. calcoaceticus* and *A. eutrophus aroQ$_p$·pheA* genes should be most informative, in part because of their appropriate phylogenetic distance and in part because they have been thoroughly characterized at the enzymologic level.

A second aspect of diversity is that various microbial AroQ$_p$·PheA proteins differ in their ability to interact with L-tyrosine. At one extreme, for example, in enteric bacteria (Hudson and Davidson, 1984), there is no detectable effect of L-tyrosine as an allosteric agent or protectant. At the other extreme, AroQ$_p$·PheA activity is totally dependent on the presence of L-tyrosine, for example, in *Alcaligenes faecalis* and *Nitrosomonas europaea* (Subramaniam et al., 1994). Throughout the β and γ subdivisions of the Proteobacteria, other AroQ$_p$·PheA proteins exhibit a wide range of sensitivity to L-tyrosine as an activator. In such organisms as *N. gonorrhoeae* and *Pseudomonas saccharophila,* dramatic activation on the order of 20-fold is observed (Subramaniam et al., 1994). In contrast, such organisms as *Rhodocyclus tenuis* (Subramaniam et al., 1994) show only modest activation effects of 10% or less.

Allosteric control of *X. campestris* $AroQ_p$·PheA is expected because the residues within the allostery-specific portion of the ·PheA domain that are highly conserved in other $AroQ_p$·PheA proteins have been retained. It thus seems possible that the unique presence of the lysine-rich amino-terminus may reflect a posttranslational processing mechanism in the native organism. Experiments beyond the scope of this work could establish the extent to which excision of the amino-terminus of *X. campestris* $AroQ_p$·PheA might affect sensitivity to inhibition by L-phenylalanine, sensitivity to allosteric activation by L-tyrosine, and effects of L-phenylalanine or L-tyrosine or both on quaternary structure.

The aroQ *gene family*

Chorismate mutase catalyzes a facile reaction that appears to have evolved independently at least two times. X-ray crystallographic analysis has revealed structurally different catalytic centers for *E. coli* $AroQ_p$ (Lee et al., 1995), and *B. subtilis* AroH (Chook et al., 1993). Yeast AroR possesses a catalytic center that is structurally similar to that of *E. coli* $AroQ_p$, but the primary sequences do not exhibit homology, although four dispersed regions of suggestive identity exist (Xue and Lipscomb, 1995). From a vantage point of tertiary structure, the eukaryotic chorismate mutases belong to the AroQ class. In both cases, 4-helix bundles form the catalytic domain and active-site residues are similar. It is possible that common ancestry has been obscured by divergence, as suggested by Xue and Lipscomb (1995). *B. subtilis* AroH has one known homologue (in *Synechocystis* sp.). The AroR family consists of two known homologues, one from *Saccharomyces* and one from *Arabidopsis thaliana*. The current AroQ gene family has fourteen members, as shown by the multiple alignment in Figure 7.

For 11 members of the family, a catalytic domain for chorismate mutase coexists on the same protein with another catalytic domain. The second catalytic domain of these bifunctional proteins is either prephenate dehydratase (six sequences), prephenate dehydrogenase (four sequences), or 3-deoxy-D-arabino-heptulosonate 7-P synthase (one sequence). A monofunctional $AroQ_f$ from *Methanococcus jannaschii* (Bult et al., 1996), an archaeon representative, is remarkably conserved with respect to important catalytic residues. The amino acid sequence at the deduced N-terminus does not exhibit features expected for a cleavable signal peptide. Thus, unlike *E. herbicola* $AroQ_f$ and *Mycobacterium tuberculosis* $AroQ_f$, *M. jannaschii* $AroQ_f$ must be located in the cytoplasm. The monofunctional $AroQ_f$ from *E. herbicola* (and from *M. tuberculosis*) exhibits the least overall identity with the other family members (including two paralog sequences from *E. herbicola*). This apparently reflects a divergence related to the periplasmic location (Xia et al., 1993a) of this species. Signal peptide cleavage (Fig. 7) produces a mature protein that, as judged by the absence of key amino acid residues, appears to lack part of the otherwise conserved N-terminal α-helix that contributes to the four-helix bundle domain forming the active site. Inspection of the three-dimensional structure of the homologous *E. coli* $AroQ_p$ shows that the catalytic site is composed of the N-terminal half of the N-terminal helix of one subunit and three α-helices from the adjacent subunit (Lee et al., 1995). Hence, a different topology of the catalytic chorismate mutase domain must distinguish *E. herbicola* and *M. tuberculosis* species of $AroQ_f$ from other family members. Nevertheless, except for the conserved active-site residues in the N-terminal helix (most notably the strictly conserved arginine 11′, see below), all important catalytic amino acids are present in *E. herbicola* $AroQ_f$, and there is clear primary sequence similarity beyond the missing region (Fig. 7). The same is true for *M. tuberculosis* $AroQ_f$, except for the absence of glutamate at position 52.

Figure 7 includes a schematic diagram illustrating the hydrogen bonding and electrostatic interactions of a transition state analogue (*endo*-oxabicyclic diacid inhibitor) with side chains of eight catalytic-site residues of *E. coli* $AroQ_p$ (Lee et al., 1995). Of these, Arg51, Arg28, Glu52, and Lys39 are conserved throughout the family. Glu52 and Lys39 provide a negative and positive charge, respectively, within a dipolar active site, and it has been proposed that an electrostatic gradient in the active site is a critical element of catalysis (Kast et al., 1996, and references therein). Perhaps for *M. tuberculosis* $AroQ_f$, Glu53 is equivalent to Glu52 in the rest of the family.

As discussed, only $AroQ_f$ from *E. herbicola* and *M. tuberculosis* lacks Arg11′, whose guanidinium group forms a prominent salt bridge with the carboxylate moiety of the enolpyruvyl side chain. This interaction may be essential to fix the flexible chorismate molecule in a conformation primed for the rearrangement. Ser84 and Asp48 are not highly conserved. In those cases where position 88 is not occupied by Gln, it is

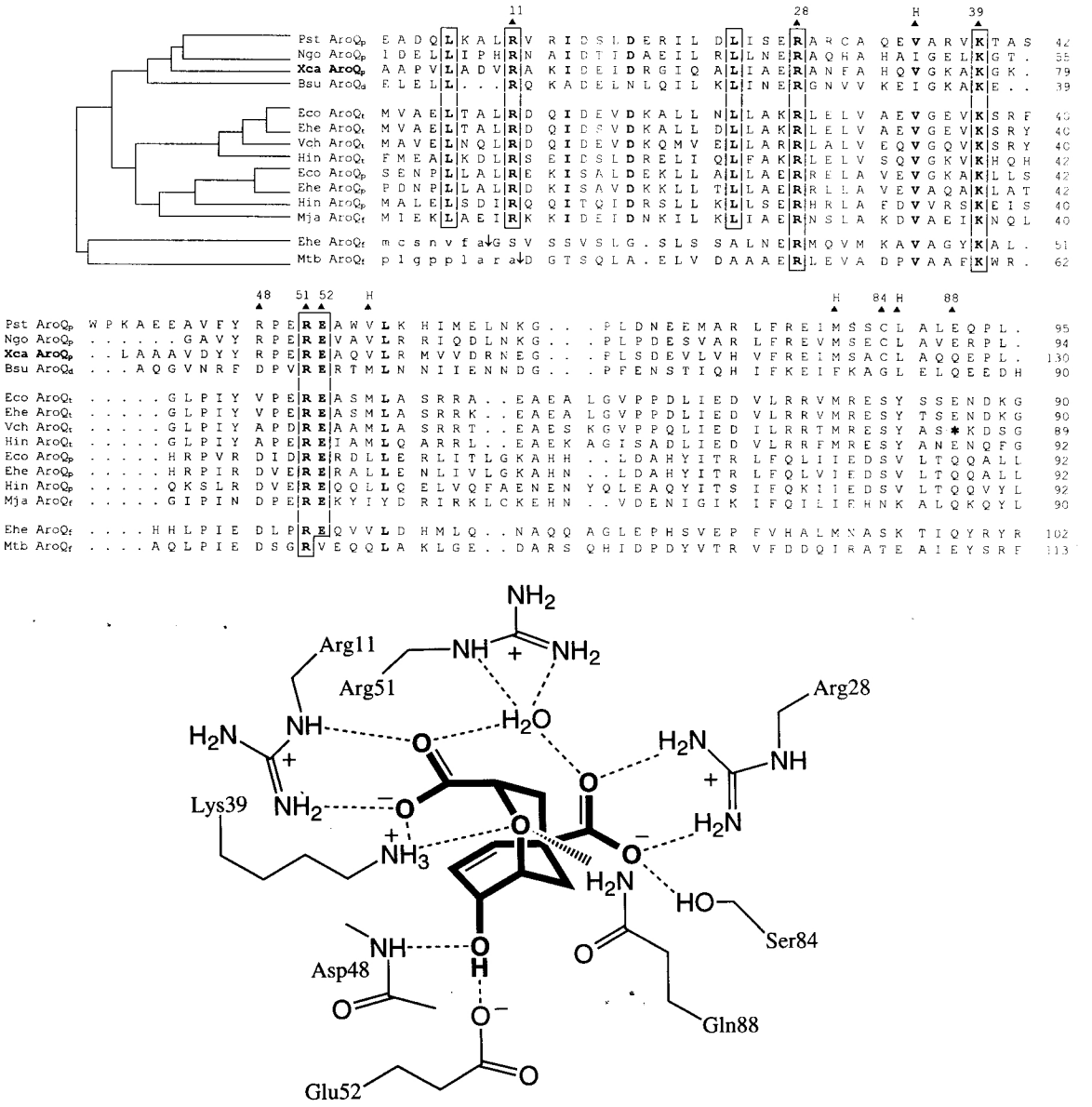


FIG. 7. Multiple alignment of the AroQ protein family. The most divergent members of the family, Ehe $AroQ_f$ and Vch $AroQ_f$, are spaced at the bottom from other family members. Lower-case letters denote the signal sequences that are cleaved from the latter two proteins. Residues located at the active site of Eco $AroQ_p$, as demonstrated by x-ray crystallography (bottom schematic), are indicated by residue numbers at the top. With respect to other indicated residues, residue 11′ is located on the other subunit in the active homodimer. Residues conserved throughout at least the upper major cluster are boxed. Other highly conserved residues are in boldface. Residues marked with an H at the top participate in important hydrophobic interactions in Eco $AroQ_p$. Residue numbers are shown at right. An asterisk marks the position of residue 85 in the Vch AroQt sequence which almost certainly is a sequencing error. The present TAA stop codon is unlikely because of its location within the coding region of other homologs; residues immediately before and after the codon are highly conserved. It is likely that this residue is E (GAA) or Q (CAA). Pst, *P. stutzeri*; Ngo, *N. gonorrhoeae*; Xca, *X. campestris*; Bsu, *Bacillus subtilis*; Eco, *E. coli*; Ehe, *E. herbicola*; Vch, *Vibrio cholerae*; Hin, *H. influenzae*; Mja, *Methanococcus jannaschii*; Mtb, *Mycobacterium tuberculosis*. Reference citations in order from top to bottom documenting the sequences shown are Fischer et al., 1991; Contig 416; this report; Bolotin et al., 1995; Hudson and Davidson, 1984; Xia et al., 1993; TIGR GVCDG49F; Fleischmann et al., 1995; Hudson and Davidson, 1984; Xia et al., 1993b; Fleischmann et al., 1995; Bult et al., 1996; Xia et al., 1993; Accession No. Z97193. (Bottom schematic with permission from Kast et al., 1996.)

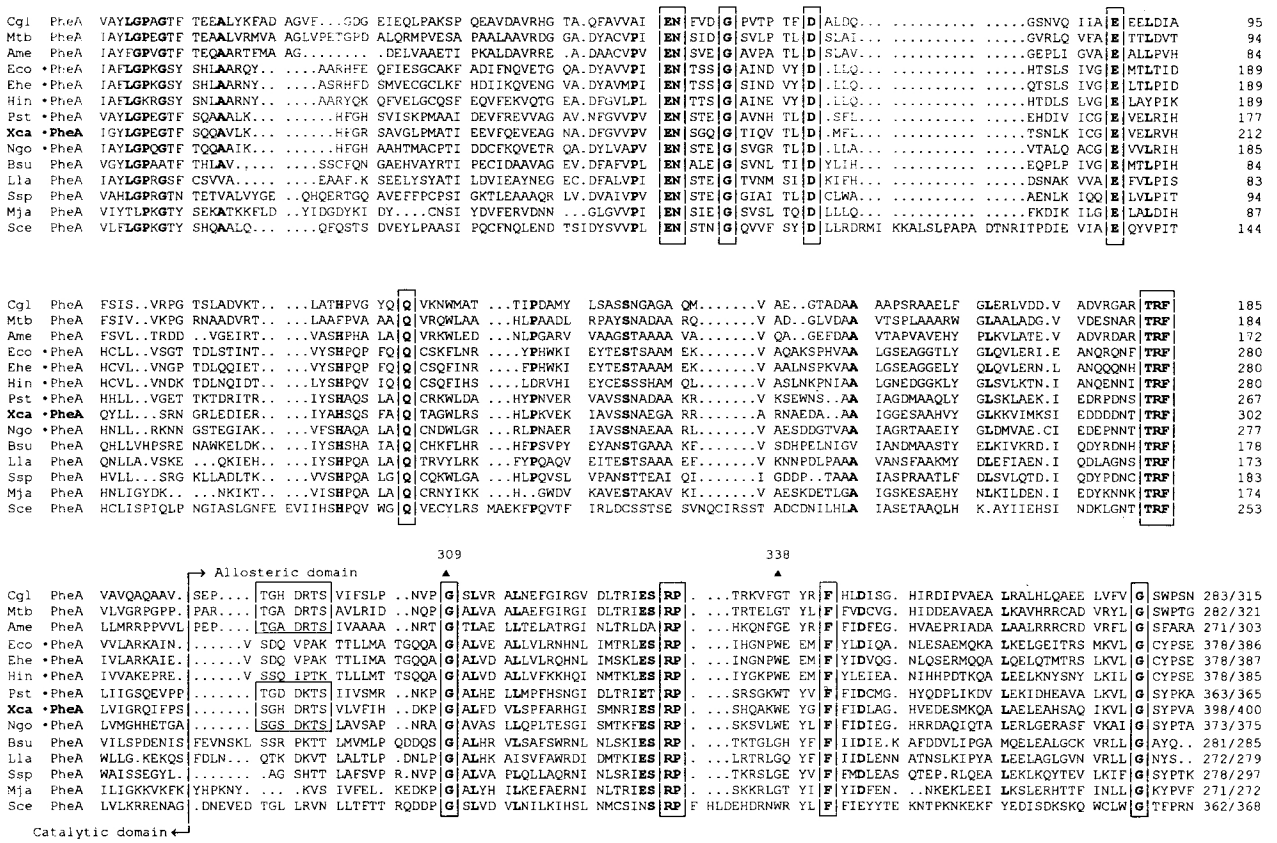


FIG. 8. Multiple alignment of monofunctional prephenate dehydratases (PheA) and P-protein prephenate dehydratase domains (·PheA). Invariant residues are boxed, and highly conserved residues are shown in boldface. The junction separating the catalytic and allosteric regions is indicated. Residues 309 and 338, shown by mutation analysis in *E. coli* to be important for feedback inhibition (Nelms et al., 1992), are marked. The motif (S/TGxDR/KTS) suggested for allosteric activation by tyrosine is boxed. Reference citations in order from top to bottom documenting the sequences shown are Follettie and Sinskey, 1986; Accession No. Z83864; Vrijbloed et al., 1995; Hudson and Davidson, 1984; Xia et al., 1993; Fleischmann et al., 1995; Fischer et al., 1991; this report; Contig 396; Trach and Hoch, 1989; Griffin and Gasson, 1995; Accession No. D90900; Bult et al., 1996; Accession No. P32452.

Glu88. Interestingly, when Eco-AroQ$_p$ Gln88 was replaced by Glu88, 1% of normal catalytic activity was retained at pH 7.8. However, at pH 4.5, the Q88E mutant exhibited 140% of wild-type activity (Zhang et al., 1996). The pH-dependent activity of the same mutant was reported by Liu et al. (1996). Low pH presumably allows the undissociated carboxylic acid of Glu88 to function as a hydrogen bond donor in concert with Lys39 (Zhang et al., 1996). Consistent with this is that yeast chorismate mutase, which has a glutamate residue in the equivalent position (Sträter et al., 1996), exhibits a pH optimum of 5.5 with little activity at pH 7.8.

The pheA *gene family*

Prephenate dehydratase is represented in nature either as a monofunctional protein (PheA) or as one catalytic domain of bifunctional proteins (AroQ$_p$·PheA). A multiple alignment is shown in Figure 8 of 14 sequences, 8 of them monofunctional proteins (PheA) and 6 of them PheA domains (·PheA). All ·PheA domains are from the γ division of gram-negative bacteria, except for Ngo·PheA (β division). AroQ$_p$·PheA proteins from some members of the β division have been well studied (Friedrich et al., 1976; Subramaniam et al., 1994). Sequences for the monofunctional PheA (and AroQ) proteins present in the α division are not yet represented. All dehydratase proteins are feedback inhibited by L-phenylalanine. The carboxy

portions of prephenate dehydratase proteins comprise a discrete location for feedback inhibition, based on deletion and mutant analysis of *aroQ$_p$·pheA* from *E. herbicola* (Xia et al., 1992) and *E. coli* (Nelms et al., 1992). Figure 8 shows an approximate demarcation between the N-terminal catalytic domain and the C-terminal allosteric domain based on the location of known mutations or deletions that affect catalysis or allostery independently of one another. The conservation of various residues throughout the homology family indicates that all share the N-catalytic/C-allostery organization. It has been suggested (Vrijbloed et al., 1995) that the motif PTGxD might correspond to allosteric activation by tyrosine found in Cgl PheA (Follettie and Sinskey, 1986), Ame PheA (Vrijbloed et al., 1995), and *P. stutzeri* PheA (Byng et al., 1983). Based on the additional data from Xca·PheA and Ngo·PheA, we suggest the motif S/TGxDR/KTS. This predicts that the Mtb PheA is subject to activation by tyrosine. However, the cyanobacterial protein (Ssp PheA) exhibits no similarity to this motif, even though it is strongly activated by tyrosine (Ahmad and Jensen, 1986).

Bacillus subtilis PheA is known to exhibit additional allosteric effects, which include inhibition by L-tryptophan and activation by L-methionine and L-leucine. Activator and inhibitor molecules induce transitions of molecular mass to dimers and octamers, respectively (Pierson and Jensen, 1974). The molecular basis of this complexity in relationship to homologues having entirely different effector-mediated patterns would be of interest.

ACKNOWLEDGMENTS

The authors thank the Institute for Genomic Research and the Gonococcal Genome Sequencing Project (University of Oklahoma) for availability of sequence data before publication. We thank Peter Kast for his helpful critical comments and suggestions, as well as for provision of the schematic for chorismate mutase shown in Figure 7. This article is Florida Agriculture Station Journal Series No. R-05767.

REFERENCES

AHMAD, S., and JENSEN, R.A. (1986). The evolutionary history of two bifunctional proteins that emerged in the purple bacteria. Trends Biochem Sci **11,** 108–112.

AHMAD, S., and JENSEN, R.A. (1988). Phylogenetic distribution of components of the overflow pathway to L-phenylalanine within the enteric lineage of bacteria. Curr Microbiol **16,** 295–302.

AHMAD, S., WILSON, A.-T., and JENSEN, R.A. (1988). Chorismate mutase:prephenate dehydratase from *Acinetobacter calcoaceticus:* purification, properties and immunological cross-reactivity. Eur J Biochem **176,** 69–79.

ALDRICH, H.C., McDOWELL, L., BARBOSA, M. de F.S., YOMANO, L.P., SCOPES, R.K., and INGRAM, L.O. (1992). Immunocytochemical localization of glycolytic and fermentative enzymes in *Zymomonas mobilis*. J Bacteriol **174,** 4504–4508.

ALDRICH, H.C., WILLAMS, D.S., ROBERSON, R.W., and HARDING, M. (1996). Improved preservation of fungi and slime molds with microwave-assisted fixation. J Histochem Cytochem **44,** 788.

APPLIED BIOSYSTEMS, Inc. User bulletin 18. Applied Biosystems, Inc., Foster City, CA.

BERRY, A., BYNG, G.S., and JENSEN, R.A. (1985). Interconvertible molecular-weight forms of the bifunctional chorismate mutase-prephenate dehydratase from *Acinetobacter calcoaceticus*. Arch Biochem Biophys **243,** 470–479.

BOLOTIN, A., KHAZAK, V., STOYNOVA, N., RATMANOVA, K., YOMANTAS, Y., and KOZLOV, Y. (1995). Identical amino acid sequence of the aroA(G) gene products of *Bacillus subtilis* 168 and *B. subtilis* Marburg strain. Microbiology **141,** 2219–2222.

BRADFORD, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem **72,** 248–254.

BULT, C.J., WHITE, O., OLSEN, G.J., ZHOU, L., FLEISCHMANN, R.D., SUTTON, G.G., et al. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science **273,** 1058–1075.

BYNG, G.S., WHITAKER, R.J., and JENSEN, R.A. (1983). Evolution of L-phenylalanine biosynthesis in rRNA homology group I of *Pseudomonas*. Arch Microbiol **136,** 163–168.

CHOOK, Y.M., KE, H., and LIPSCOMB, W.N. (1993). Crystal structure of the monofunctional chorismate mutase from *Bacillus subtilis* and its complex with a transition state analog. Proc Natl Acad Sci USA **90,** 8600–8603.

COTTON, R.G.H., and GIBSON, F. (1965). The biosynthesis of phenylalanine and tyrosine: enzymes converting cho-

rismic acid into prephenic acid and their relationship to prephenate dehydratase and prephenate dehydrogenase. Biochim Biophys Acta. **100,** 76–88.

CRAWFORD, I.P. (1989). Evolution of a biosynthetic pathway: the tryptophan paradigm. Annu Rev Microbiol **43,** 567–600.

DUNCAN, K., and COGGINS, J.R. (1986). The *serC-aroA* operon of *Escherichia coli.* Biochem J **234,** 49–57.

FISCHER, R.S., ZHAO, G., and JENSEN, R.A. (1991). Cloning, sequencing, and expression of the P-protein gene (*pheA*) of *Pseudomonas stutzeri* in *Escherichia coli:* implications for evolutionary relationships in phenylalanine biosynthesis. J Gen Microbiol **137,** 1293–1301.

FLEISCHMANN, R.D., ADAMS, M.D., WHITE, O., CLAYTON, R.A., KIRKNESS, E.F., KERIAVAGE, A.K., et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269,** 496–512.

FOLLETTIE, M.T., and SINSKEY, A.J. (1986). Molecular cloning and nucleotide sequence of the *Corynebacterium glutamicum pheA* gene. J Bacteriol **167,** 695–702.

FRIEDRICH, C.G., FRIEDRICH, B., and SCHLEGEL, H.G. (1976). Purification and properties of chorismate mutase-prephenate dehydratase from *Alcaligenes eutrophus.* J Bacteriol **126,** 712–722.

GENETICS COMPUTER GROUP, INC. (1995). Wisconsin sequence analysis package, version 8.0. Genetics Computer Group, Inc., Madison, WI.

GRIFFIN, H.G., and GASSON, M.J. (1995). Genetic aspects of aromatic amino acid biosynthesis in *Lactococcus lactis.* Mol Gen Genet **246,** 119–127.

GRIFFIN, H.G., and GRIFFIN, A.M. (1991). Cloning and DNA sequence analysis of the *serC-aroA* operon from *Salmonella gallinarum:* evolutionary relationships between the prokaryotic and eukaryotic *aroA*-encoded enzymes. J Gen Microbiol **137,** 113–121.

GU, W., ZHAO, G., EDDY, C., and JENSEN, R.A. (1995). Imidazole acetol phosphate aminotransferase in *Zymomonas mobilis:* molecular genetic, biochemical, and evolutionary analyses. J Bacteriol **177,** 1576–1584.

HUDSON, G.S., and DAVIDSON, B.E. (1984). Nucleotide sequence and transcription of the phenylalanine and tyrosine operons of *Escherichia coli* K-12. J Mol Biol **180,** 1023–1051.

JENSEN, R.A. (1992). An emerging outline of the evolutionary history of aromatic amino acid biosynthesis. In *The Evolution of Metabolic Function.* R.P. Mortlock, ed. (CRC Press, Inc., Boca Raton, FL), pp. 205–236.

JENSEN, R.A., and GU, W. (1996). Evolutionary recruitment of biochemically specialized subdivisions of Family-I within the protein superfamily of aminotransferases. J Bacteriol **178,** 2161–2171.

KAST, P., ASIF-ULLAH, M., JIANG, N., and HILVERT, D. (1996). Exploring the active site of chorismate mutase by combinatorial mutagenesis and selection: the importance of electrostatic catalysis. Proc Natl Acad Sci USA **93,** 5043–5048.

LEE, A.Y., KARPLUS, P.A., GANEM, B., and CLARDY, J. (1995). Atomic structure of the buried catalytic pocket of *Escherichia coli* chorismate mutase. J Am Chem Soc **117,** 3627–3628.

LIU, D.R., CLOAD, S.T., PASTOR, R.M., and SCHULTZ, P.G. (1996). Analysis of active site residues in *Escherichia coli* chorismate mutase by site-directed mutagenesis. J Am Chem Soc **118,** 1789–1790.

MANIATIS, T., FRITSCH, E.F., and SAMBROOK, J. (1989). *Molecular Cloning: A Laboratory Manual,* 2nd ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

NELMS, J., EDWARDS, R.M., WARWICK, J., and FOTHERINGHAM, I. (1992). Novel mutations in the *pheA* gene of variants of chorismate mutase/prephenate dehydratase. Appl Environ Microbiol **58,** 2592–2598.

O'GAORA, P., MASKELL, D., COLEMAN, D., CAFFERKEY, M., and DOUGAN, G. (1989). Cloning and characterization of the *serC* and *aroA* genes of *Yersinia enterocolitica* and construction of an *aroA* mutant. Gene **84,** 23–30.

PIERSON, D.L., and JENSEN, R.A. (1974). Metabolic interlock: control of an interconvertible prephenate dehydratase by hydrophobic amino acids in *Bacillus subtilis.* J Mol Biol **90,** 563–580.

SILHAVY, T.J., BERMAN, M.L., and ENQUIST, L.W. (1984). *Experiments with Gene Fusions.* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

STRÄTER, N., HÄKANSSON, K., SCHNAPPAUF, G., BRAUS, G., and LIPSCOMB, W.N. (1996). Crystal structure of the T state of allosteric yeast chorismate mutase and comparison with the R state. Proc Natl Acad Sci USA **93,** 3330–3334.

SUBRAMANIAM, P., BHATNAGAR, R., HOOPER, A., and JENSEN, R.A. (1994). The dynamic progression of evolved character states for aromatic amino acid biosynthesis in gram-negative bacteria. Microbiology **140,** 3431–3440.

TOWBIN, H., STAEHELIN, T., and GORDON, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gel to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci USA **76,** 4350–4354.

TRACH, K., and HOCH, J.A. (1989). The *Bacillus subtilis spoOB* stage 0 sporulation operon encodes an essential GTP-binding protein. J Bacteriol **171,** 1362–1371.

VRIJBLOED, J.W., VLIEG, J.H., PUT, N.M., HESSELS, G.I., and DIJKHUIZEN, L. (1995). Molecular cloning with

a pMEA300-derived shuttle vector and characterization of the *Amycolatopsis methanolica* prephenate dehydratase gene. J Bacteriol **177,** 6666–6669.

WHITAKER, R.J., BERRY, A., BYNG, G.S., FISKE, M.J., and JENSEN, R.A. (1985). Clues from *Xanthomonas campestris* about the evolution of aromatic biosynthesis and its regulation. J Mol Evol **21,** 139–149.

XIA, T., and JENSEN, R.A. (1992). Monofunctional chorismate mutase from *Serratia rubidaea:* a paradigm system for the three-isozyme gene family of enteric bacteria. Arch Biochem Biophys **294,** 147–153.

XIA, T., SONG, T., ZHAO, G., ALDRICH, H., and JENSEN, R.A. (1993a). The *aroQ*-encoded monofunctional chorismate mutase (CM-F) protein is a periplasmic enzyme in *Erwinia herbicola.* J Bacteriol **175,** 4729–4737.

XIA, T., ZHAO, G., and JENSEN, R.A. (1992). Loss of allosteric control but retention of the bifunctional catalytic competence of a fusion protein formed by excision of 260 base pairs from the 3′ terminus of *pheA* from *Erwinia herbicola.* Appl Environ Microbiol **58,** 2792–2798.

XIA, T., ZHAO, G., and JENSEN, R.A. (1993b). The *pheA/tyrA/aroF* region from *Erwinia herbicola:* an emerging comparative basis for analysis of gene organization and regulation in enteric bacteria. J Mol Evol **36,** 107–120.

XUE, Y., and LIPSCOMB, W.N. (1995). Location of the active site of allosteric chorismate mutase from *Saccharomyces cerevisiae,* and comments on the catalytic and regulatory mechanisms. Proc Natl Acad Sci USA **92,** 10595–10598.

YANISCH-PERRON, C., VIEIRA, J., and MESSING, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33,** 103–119.

ZHANG, S., KONGSAEREE, P., CLARDY, J., WILSON, D.B., and GANEM, B. (1996). Site-directed mutagenesis of monofunctional chorismate mutase engineered from the *E. coli* P-protein. Bioorg Med Chem **4,** 1015–1020.

Address reprint requests to:
Roy A. Jensen
Department of Microbiology
and Cell Science
University of Florida
Gainesville, FL 32611