Molecular cloning and characterization of the aroD gene encoding 3-dehydroquinase from Salmonella typhi

SPIROS SERVOS,1,4 STEVE CHATFIELD,1 DAVID HONE,2,3 MYRON LEVINE,2 GEORGE Dimitriadis,4 DEREK PICKARD,1 GORDON DOUGAN,1 NEIL FAIRWEATHER1 and IAN CHARLES1*

1 Department of Molecular Biology, Wellcome Biotech, Langley Park, South Eden Park Road, Beckenham, Kent BR3 3BS, UK
2 Center for Vaccine Development, School of Medicine, and 3 The Medical Biotechnology Center, University of Maryland, 10 South Pine St, Baltimore, Maryland 21201, USA
4 University of Patras, Faculty of Sciences, Department of Biology, Division of Genetics, Cell and Development Biology, Patras, Greece

(Received 3 July 1990; accepted 19 September 1990)

The aroD gene from Salmonella typhi, encoding 5-dehydroquininate hydrolyase (3-dehydroquinase), has been cloned into Escherichia coli and the DNA sequence determined. The aroD gene was isolated from a cosmid gene bank by complementation of an S. typhimurium aroD mutant. Analysis of the DNA sequence revealed the presence of an open reading frame capable of encoding a protein of 252 amino acids with a calculated M, of 27706. Comparison of the deduced S. typhi 3-dehydroquinase protein sequence with that elucidated for E. coli revealed 69 % homology. Alignment of the S. typhi sequence and equivalent Aspergillus nidulans and Saccharomyces cerevisiae sequences showed that homology was lower, at 24 %, but still significant. Use of a minicell expression system demonstrated that a polyclonal antibody raised against E. coli 3-dehydroquinase cross-reacted with its S. typhi counterpart.

Introduction

The only biosynthetic route for the synthesis of aromatic compounds possessed by micro-organisms is the shikimate pathway, leading to the synthesis of chorismic acid, a central precursor for other aromatic compounds (Giles, 1978; Pittard, 1987). The products of the pathway include the aromatic amino acids p-aminobenzoic acid and 2,3-dihydroxybenzoate. The enzyme 3-dehydroquinase catalyses step three of the prechorismate part of the pathway and in bacteria is encoded by the gene aroD. In Escherichia coli, DNA sequence analysis of aroD has revealed the presence of a monocistronic gene capable of encoding a protein of 240 amino acid residues with a calculated M, of 26377 (Duncan et al., 1986). By contrast, the plant enzyme consists of 3-dehydroquinase linked with shikimate dehydrogenase as a bifunctional polypeptide (Polley, 1978). The 3-dehydroquinase activity found in Aspergillus nidulans (Kinghorn & Hawkins, 1982; Charles et al., 1986; Hawkins, 1987), Neurospora crassa (Lumsden & Coggins, 1977; Catcheside et al., 1985) and Saccharomyces cerevisiae (Duncan et al., 1987) has been identified as part of a pentafunctional polypeptide. In each case, the large polypeptide carries the enzyme activities associated with steps two to six of the prechorismate part of the pathway. In addition to these biosynthetic activities an inducible catabolic dehydroquinase involved in the quinic acid catabolic pathway has been characterized from A. nidulans (Hawkins et al., 1982) and N. crassa (Giles et al., 1985). In A. nidulans the quinate-inducible 3-dehydroquinase is the product of the QuTE gene and has an M, of 16505. A similarly sized protein from N. crassa, the product of the QA2 locus, has considerable homology to the QuTE gene product (Da Silva et al., 1986). Neither of these two catabolic dehydroquinase genes or their deduced protein sequences shows any significant homology with their biosynthetic counterparts.

In our laboratories, interest in the shikimate pathway has come from the observation that pathogenic organisms with genetic lesions in genes encoding enzymes involved in the prechorismate pathway grow poorly in vivo (Dougan et al., 1987; Levine et al., 1987; O'Callaghan et al., 1988) and are of potential use as live vaccines (Bacon et al., 1951; Hoiseth & Stocker, 1981).
Shikimate pathway enzymes are also absent from mammalian species, and consequently have been the target for chemotherapeutic agents and broad-range herbicides (Kishore & Shah, 1988). As part of our research programme to generate aromatic amino acid dependent mutants of pathogenic organisms in order to evaluate their utility as live-attenuated vaccines (Dougan et al., 1987), we have cloned, sequenced and characterized a number of aro genes from different microorganisms (Charles et al., 1985, 1986, 1990; Maskell et al., 1988). In this paper we report the cloning and characterization of the aroD gene from Salmonella typhi, and demonstrate homology between its deduced protein sequence and the protein sequences of other biosynthetic 3-dehydroquinase enzymes.

Methods

**Strains, bacteriophage and plasmids.** E. coli K12 strain TG1 (lanE-pro supE thi hsdS2F r434m2 proA2B2 lacZM15 (Carter et al., 1985)] was the kind gift of Dr T. Gibson, Laboratory of Molecular Biology, Cambridge, UK. E. coli SK2881 (F- dcm A734 m 245 tfrB403 lacZM15) has been described by Kushner (1978). S. typhimurium strain S68 [his-6165 fla-452 metA22 metE51 trpB2 galE496 xyl-404 rpsL20 flaA6 hsdL6 hsdS2A9 lamB(E. coli) zja : Tn10] harbours the lamB gene from E. coli integrated into the chromosome and is susceptible to bacteriophage λ infection when grown on 0.2% maltose; it was the kind gift of Professor B. Stocker, Department of Medical Microbiology, Stanford University, California, USA. S. typhimurium CU038 (aroD), S. typhi Ty2 (aroA : Tn10) and bacteriophage P22 have been described previously (Dougan et al., 1987; Miller et al., 1989).

Cosmid pH79 (Hohn & Collins, 1980) was obtained from Amersham. Transposon Rts1: Tn725 has been described (Ubben & Schmitt, 1986). M13mp18 and M13mp19 (Messing & Vieira, 1982) were supplied by Pharmacia. Minicells were prepared using E. coli strain DS410 (Dougan & Sherrat, 1977).

**Media and reagents.** S. typhimurium, S. typhi and E. coli were grown in Luria broth (LB) or LB solidified with 1.6% (w/v) Noble agar (Difco) was used (Miller, 1972). Restriction endonucleases were supplied by BRL Gibco and Pharmacia. The T7 and Klenow DNA sequencing kits were supplied by Pharmacia and Cambridge Bioscience. T4 DNA ligase was from Boehringer. Oligonucleotide primers were synthesized on a Beckmanpolymerase (Tabor & Richardson, 1987), using kits supplied by Pharmacia and Cambridge Bioscience. Labelled fragments were separated on buffer gradient polyacrylamide/urea gels (Biggin et al., 1983). DNA sequencing was initiated from sites within the S. typhi aroD gene by cloning EcoRI fragments from four independent aroD-associated Tn725 transposon insertions (Ubben & Schmitt, 1986). Gaps in the sequence were filled in using synthetic oligonucleotides as specific primers (Charles et al., 1985, 1986; Strauss et al., 1986).

Oligonucleotides: Oligonucleotide primers were synthesized on a SAMI oligonucleotide synthesizer (Biolabs, UK).

**DNA isolation and Tn725 mutagenesis.** Plasmid DNA was isolated by the alkaline-SDS lysis method (Maniatis et al., 1982). DNA restriction endonuclease fragments were isolated by the method of Tautz & Renz (1983). Chromosomal DNA was prepared by the method of Hull et al. (1981) as modified by Maskell et al. (1988). Tn725 mutagenesis and mapping was carried out by the method of Ubben & Schmitt (1986).

**Construction of an S. typhi chromosomal gene bank in phC79.** Chromosomal DNA from S. typhi aroA : Tn10 was partially digested with Sau3A and fragments in the size range 23-40 kb were size-selected as described previously (Maniatis et al., 1982; Morrissey & Dougan, 1986) and ligated into BamHI-digested pH79 (Hohn & Collins, 1980). Cosmids were packaged in vitro with a kit supplied by Amersham using conditions recommended by the manufacturer, and phage plated out as described previously (Miller et al., 1989). A high titre stock of packaged recombinant phage was mixed with S. typhimurium AS68 at an м.o.i. of 0.1, and recombinants selected on LB plates containing ampicillin at a concentration of 50 μg ml−1. Several thousand recombinants were pooled, and a P22 stock was generated by standard methods (Davis et al., 1980). This P22 bank was subsequently used to complement S. typhimurium strain CU038 (aroD).

**Minicells experiments and Western blotting.** Minicell experiments and SDS-PAGE were carried out as described by Dougan & Sherrat (1977). The procedure used for Western blotting was as described by Towbin et al. (1979).

**DNA sequencing.** Single-stranded DNA templates were sequenced using Klenow (Sanger et al., 1977) or T7 DNA polymerase (Tabor & Richardson, 1987), using kits supplied by Pharmacia and Cambridge Bioscience. Labelled fragments were separated on buffer gradient polyacrylamide/urea gels (Biggin et al., 1983). DNA sequencing was initiated from sites within the S. typhi aroD gene by cloning EcoRI fragments from four independent aroD-associated Tn725 transposon insertions (Ubben & Schmitt, 1986). Gaps in the sequence were filled in using synthetic oligonucleotides as specific primers (Charles et al., 1985, 1986; Strauss et al., 1986).

Oligonucleotides: Oligonucleotide primers were synthesized on a SAMI oligonucleotide synthesizer (Biolabs, UK).

**Results and Discussion**

Cloning and mapping of the S. typhi gene for aroD in E. coli

A cosmide gene bank generated in pHCH79 was used to complement an aroD lesion in S. typhimurium CU038. Transformants were selected on MM containing ampicillin at a concentration of 50 μg ml−1, but lacking aromatic compounds. Colonies were visible after 48 h incubation at 37 °C. Of the initial six transformants that were picked to complement an aroD lesion, a single recombinant cosmid, pWBTYD1, that demonstrated recomplementation was selected for further analysis by transposon mapping, DNA sequencing and minicell-directed protein synthesis.
Subcloning and transposon mutagenesis of \( S. \) typhi aroD

Initial restriction endonuclease mapping of cosmid pWBTD1 demonstrated that it carried an insert of 22 kb. In order to characterize a smaller fragment that still complemented aroD, HindIII, EcoRI and BglII fragments were subcloned into pUC18 and the ligation mixtures used to transform \( E. \) coli SK2881. Analysis of aroD-complementing recombinants by plasmid mini-preps (Maniatis et al., 1982) showed that one recombinant, pCVD1004, carried an 8.2 kb BglII fragment. This plasmid was used as a substrate for Tn1725 mapping as described previously (Ubben & Schmitt, 1986). Four pCVD1004:Tn1725 isolates were identified, three of which no longer complemented the \( E. \) coli aroD strain SK2881, presumably because of gene disruption. As a consequence of these transposon insertions, additional EcoRI sites were introduced into the gene for aroD and these new sites were used to conveniently clone aroD-associated fragments into M13mpl8 and M13mp19 for DNA sequencing (see Fig. 1). Gaps in the sequence were filled in using synthetic oligonucleotides as specific sequencing primers. Analysis of the DNA sequence (Fig. 2) reveals the presence of an open reading frame capable of encoding a protein of 252 amino acids with a calculated \( M_r \) of 27706.

Nucleotide sequence of \( S. \) typhi aroD

A 2 kb HindIII fragment, identified by transposon mapping of the 8-2 kb BglII fragment as covering the region corresponding to the \( S. \) typhi aroD gene, was cloned into M13mp18 in both orientations. These M13 templates were used as the substrates for the synthetic oligonucleotide primers. Tn1725 mutagenesis generated four plasmids, pCVD1005, pCVD1006, pCVD1007 and pCVD1008, the first three of which no longer complemented the \( E. \) coli aroD strain SK2881, presumably because of gene disruption. As a consequence of these transposon insertions, additional EcoRI sites were introduced into the gene for aroD and these new sites were used to conveniently clone aroD-associated fragments into M13mp18 and M13mp19 for DNA sequencing (see Fig. 1). Gaps in the sequence were filled in using synthetic oligonucleotides as specific sequencing primers. Analysis of the DNA sequence (Fig. 2) reveals the presence of an open reading frame capable of encoding a protein of 252 amino acids with a calculated \( M_r \) of 27706.

Putative regulatory elements for the \( S. \) typhi aroD gene

Examination of the DNA sequence (Fig. 2) reveals the presence of two regions with a good fit to the consensus
Fig. 3. Amino acid homologies between the *S. typhi* *aroD* gene product and other published 3-dehydroquinase sequences. Sequences that are common to three or more of the species are boxed. Dots refer to regions of variable length that are introduced in the sequences to produce a best fit for regions of homology.

−10 and −35 promoter elements (Rosenberg & Court, 1979). The sequence 5'-TATAAT-3' from nucleotides 65 to 70 is an exact match with the −10 consensus sequence, while the sequence 5'-TTTACT-3' from nucleotides 41 to 46 with a spacing of 18 bp has a 4/6 match with the consensus −35 region.

Comparison of the DNA sequence flanking the structural gene for *aroD* from *S. typhi* with the *E. coli aroD* flanking regions shows little homology, although both sequences have a candidate AGG ribosome-binding site (Shine & Dalgarno, 1974). In the *E. coli aroD* sequence an inverted repeat has been noted 48 bp after the end of the structural gene but this feature is apparently absent from the *S. typhi aroD* sequence.

Comparison of the *S. typhi* and *E. coli aroD* sequences

Comparison of the sequences shows a high degree of homology, 69.6% at the protein level, and 68.7% when DNA sequences were aligned. Many of the observed protein changes are, however, conservative substitutions. Interestingly, the sequence from *S. typhi* comprises
252 amino acid residues, while the sequence from *E. coli* has 240 residues. This sequence difference appears to be the consequence of an error in the published *E. coli* sequence for *aroD*. Resequencing the *E. coli* gene indicates that it also comprises 252 amino acid residues (J. Coggins & L. Graham, personal communication).

Comparison of the codon usage table of *S. typhi* *aroD* with that derived for another gene encoding a protein involved in aromatic amino acid biosynthesis from *S. typhi*, the *aroC* gene (Charles et al., 1990), shows a high degree of similarity in the choice of which codons are most frequently used. The pattern is consistent for those genes that are poorly expressed in *E. coli* (Grantham et al., 1981). There are some minor differences in codon preference, e.g. in the *S. typhi* *aroC* gene the most popular codon for alanine is GCG, while in the *S. typhi* *aroD* it is GCC.

**Alignment of four biosynthetic dehydroquinase enzymes**

Homology between four biosynthetic dehydroquinase enzymes is displayed in Fig. 3. The figure was obtained by comparing the deduced *S. typhi* 3-dehydroquinase protein sequence individually with those for *E. coli*, *Aspergillus nidulans* and *Saccharomyces cerevisiae* using the program ALIGN (Staden, 1982). The four sequences were then compiled together by eye. Boxes have been drawn around sequences that are conserved in three or more of the proteins.

**Minicell-directed polypeptide expression from an *aroD* recombinant**

Cosmid pWBTYD1 encoding the *S. typhi* *aroD* gene was transferred to the minicell-producing strain DS410. Plasmid-harbouring minicells were purified, labelled with [35S]methionine and analysed using SDS-PAGE followed by Western blotting. Proteins were transferred to nitrocellulose, and reacted with antibody raised against *E. coli* 3-dehydroquinase (kindly supplied by J. Coggins, University of Glasgow, UK). The results are shown in Fig. 4. The *aroD*-complementing cosmid (track 3) directs the expression of a protein of M, approximately 28,000 (marked with an arrow), which is in good agreement with the value of 27,706 deduced from the DNA sequence. This polypeptide is absent from the control (track 2) sample prepared from minicells harbouring pHc79 vector. Cross-reacting bands of lower mobility, present in both the negative control (track 2) and track 1, are incidental to 3-dehydroquinase expression.

We thank Professor J. Coggins for supplying antibody to *E. coli* 3-dehydroquinase, and L. Graham for supplying us with information about the revised *E. coli* *aroD* sequence prior to publication. We also thank Hugh Spence for making oligonucleotides, Steve Nicholls for advice on computing and Tina Silva for typing the manuscript.

**References**


