Isolation, characterization and nucleotide sequences of the aroC genes encoding chorismate synthase from Salmonella typhi and Escherichia coli

IAN G. CHARLES,1 HEATHER K. LAMB,2 DEREK PICKARD,1 GORDON DOUGAN1 AND ALASTAIR R. HAWKINS2*

1 Department of Molecular Biology, Wellcome Research Laboratories, Langley Park, South Eden Park Road, Beckenham, Kent BR3 3BS, UK  
2 Department of Biochemistry and Genetics, University of Newcastle upon Tyne, Framlington Place, Newcastle upon Tyne NE2 4HH, UK

(Received 3 August 1989; revised 20 October 1989; accepted 3 November 1989)

The aroC genes from Salmonella typhi and Escherichia coli, encoding 5-enolpyruvylshikimate-3-phosphate phospholoyase (chorismate synthase) were cloned in E. coli and their DNA sequences were determined. The aroC gene from S. typhi was isolated from a cosmid gene bank by complementation of an E. coli aroC mutant. The corresponding E. coli gene was isolated from a pBR322 gene bank by colony hybridization using DNA encoding the aroC gene from S. typhi as a hybridization probe. Analysis of the nucleotide sequence revealed that both genes have an open reading frame capable of encoding proteins comprising 361 amino acids. The calculated molecular mass of the protein from S. typhi is 39,108 Da while that of the protein from E. coli is 39,138 Da. Homology is particularly strong between the coding regions of the genes: 95% when protein sequences are compared, and 83% when DNA sequences are examined. Use of a deletion variant of the E. coli aroC gene demonstrates that the C-terminal 36 amino acids are not essential for the correct folding or functional activity of the chorismate synthase enzyme.

Introduction

The biosynthesis of aromatic compounds including the aromatic amino acids p-aminobenzoic acid and 2,3-dihydroxybenzoate is of great importance for the growth and survival of bacterial cells. The only biosynthetic route for aromatic compounds possessed by microorganisms is the shikimate pathway which leads to the synthesis of chorismic acid, a central precursor for other aromatic compounds (Giles, 1978; Pittard, 1987). The shikimate pathway is present in bacteria, yeasts and in plant cells, but is absent from mammalian cells. As a consequence, enzymes in the shikimate pathway have been identified as potential targets for chemotherapeutic agents (Kishore & Shah, 1988). Because of the importance of the shikimate pathway, attention has turned to the individual enzymes responsible for aromatic compound biosynthesis. At least seven enzymes are required for the synthesis of chorismic acid from non-aromatic precursors. Chorismate synthase (5-enolpyruvylshikimate-3-phosphate phospholoyase; EC 4.6.1.4) is the final enzyme in the pathway and catalyses the conversion of 5-enolpyruvylshikimate 3-phosphate (EPSP) to chorismic acid. The enzyme chorismate synthase has been purified and characterized from a variety of micro-organisms (White et al., 1988) and plant sources (Mousdale & Coggins, 1986).

Pathogenic bacteria carrying mutations in genes encoding enzymes in the shikimate pathway (Bacon et al., 1951) such as aroA (Dougan et al., 1987; O'Callaghan et al., 1988) grow poorly in vivo and are of potential use as live vaccines. A long-term objective of this research is to use the deduced primary amino acid sequences of the chorismate synthase enzymes in a pharmacological study of the interaction of potential inhibitors with the protein. Significantly, the five monofunctional aro enzymes producing EPSP in E. coli have previously been shown to have conserved regions when compared with the equivalent pentafunctional AROM (Giles, 1978) locus of the eukaryote Aspergillus nidulans (Hawkins, 1987). To obtain some measure of the conservation of the protein sequences of the chorismate synthase enzymes we

Abbreviation: EPSP, 5-enolpyruvylshikimate 3-phosphate.

The nucleotide sequence data for E. coli and S. typhi reported in this paper have been submitted to GenBank and have been assigned the accession numbers M27714 and M27715, respectively.

0001-5751 © 1990 SGM
describe the cloning and sequencing of the aroC genes of S. typhi and E. coli. In addition, we present initial data on the size of functional domains within the chorismate synthase enzyme using deletion variants of the E. coli aroC gene.

Methods

Strains, bacteriophage and plasmids. E. coli TG1 [K12 Δ(lac-pro) supE thi hsdS2F-1 traD36 proA+ B+ lacY1 lacZΔM15 (Carter et al., 1985)] and λ b221 cl857 Oam29 Pam80 recA::Tn5 (de Bruijn & Lupski, 1984) were the kind gifts of Dr T. Gibson, Laboratory of Molecular Biology, Cambridge, UK, and Dr T. Foster, Trinity College Dublin, Ireland, respectively. E. coli GLW40 (aroC thi recA13) was the kind gift of Dr I. Hunter, Institute of Genetics, Glasgow University, UK. E. coli HB101 and S. typhi Ty2 aroA::Tn10 (Dougan et al., 1987) were as previously described. BRD049 is an E. coli K12 strain harbouring a stable mutation in aroC (White et al., 1988). Cosmid pH79 (Hohn & Collins, 1980) was from Amersham; plasmid pBR322 was as described previously (Bolivar et al., 1977) and M13mp18 and M13mp19 (Messing & Vieira, 1982) were supplied by Pharmacia. Minicells were prepared using E. coli strain DS410 (Dougan & Sherratt, 1977). E. coli strain C600 was as previously described (Appleyard, 1954).

Media and reagents. S. typhi and E. coli strains were grown in Luria broth (LB) or on LB solidified with 1.6% (w/v) agar (Miller, 1972) (Difco). Minimal medium (MM) was made as described by Miller (1972) and was solidified with 2% (w/v) Noble Agar (Difco). Deoxy- and dideoxynucleotides, ampicillin, tetracycline, dithiothreitol and (Difco). Minimal medium (MM) was made as described by Miller (1972) and was solidified with 2% (w/v) Noble Agar (Difco). Deoxy- and dideoxynucleotides, ampicillin, tetracycline, dithiothreitol and aromatic amino acids were from Sigma. Restriction endonucleases were from BRL Gibco and Pharmacia. T4 DNA ligase was from Boehringer and Klenow fragment of DNA polymerase I was from Pharmacia.

DNA isolation, Tn5 mutagenesis and minicells. Chromosomal DNA was isolated using the method of Hult et al. (1981) as outlined by Maskell et al. (1988). Tn5 mutagenesis and mapping was done with λ::Tn5 using the method described previously by de Bruijn & Lupski (1984). Minicells and SDS-PAGE were as described previously (Dougan & Sherratt, 1977).

Cloning of S. typhi chromosomal DNA in cosmids pH79. Cosmid vector pH79 (Hohn & Collins, 1980) was digested with BamHI and ligated with Sau3A-digested S. typhi DNA in the 40–50 kb size-range, as outlined previously (Maniatis et al., 1982; Morrissey & Dougan, 1986). In vitro packaging of ligated molecules was done with a Gigapack kit (Strategene; Northumbria Biologicals) using the manufacturer’s recommended conditions. E. coli HB101 recipient cells were infected with recombinant phage and plated on LB-ampicillin (50 μg ml⁻¹). Putative recombinants, identified by their Ampr Tetr phenotype, were picked into microtiter plates containing, in each well, LB + 50% (v/v) glycerol + 50 μg ampicillin ml⁻¹. Plates were incubated at 37 °C, then stored frozen at −70 °C.

DNA sequencing. DNA sequencing of the aroC genes from S. typhi and E. coli was done using universal primer, [α-35S]dATP (deoxyadenosine 5'-[α-35S]dtriphosphate) dideoxynucleotide triphosphates, and both gradient and wedge gels (Biggin et al., 1983; Sanger et al., 1977).

Some clones were sequenced with modified T7 DNA polymerase (Tabor & Richardson, 1987) using the Sequenase kit (United States Biochemical Corporation; marketed by Cambridge BioScience). Gaps in the sequences were filled by using synthetic oligonucleotides as specific primers (Charles et al., 1985, 1986). Some master templates for use with specific oligonucleotide primers were prepared from 100 ml cultures, the phage DNA being purified further by a CsCl gradient step (Strauss et al., 1986).

Oligonucleotides. Oligonucleotides, for use as specific sequence primers, were made on a SAM I oligonucleotide synthesizer (Biolabs), or manually by the paper-disc method (Matthes et al., 1984) with modifications to the wash-cycle as described by Sproat & Gait (1985).

Results and Discussion

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

Cosmid recombinants containing S. typhi inserts were packaged in vitro and used to infect E. coli HB101. A master gene bank of 600 individual recombinants was picked into microtitre plates and frozen at −70 °C. In addition, approximately 600 individual colonies were harvested from a plate and used to prepare plasmid DNA. The resulting DNA was used to transform an E. coli aroC host (BRD049) and transformants were selected on MM containing ampicillin but lacking aromatic compounds. A recombinant that grew well was selected for further analysis. After demonstrating retransformation and complementation of the aroC lesion of BRD049, plasmid DNA was prepared and designated pWAC6. pWAC6 was subjected to restriction-enzyme mapping, transposon mutagenesis and DNA-sequence analysis.

The recombinant pWAC6 was only 11·0 kb in size, with an insert of 4·5 kb of S. typhi-derived DNA, suggesting that pWAC6 had arisen by deletion of a larger cosmid. Transposon mutagenesis identified a 3·8 kb SphI–EcoRI fragment, which was subcloned from pWAC6 into pUC8 to form the recombinant plasmid pSE7; it was also able to complement the aroC lesion in BRD049.

Cloning of the E. coli C600 aroC gene

Samples of chromosomal DNA from E. coli C600 were digested separately with the restriction endonucleases ClaI and SalI, separated on a 0·7% (w/v) agarose gel and transferred to nitrocellulose by the method of Southern (1975). The resulting filter was probed with a Psit–NruI fragment from the S. typhi aroC gene that had been labelled with [α-32P]dCTP (Maniatis et al., 1982).
S. typhi and E. coli aroC DNA sequences
Analysis of the Southern blot (data not shown) demonstrated hybridization to a 3.0 kb ClaI fragment, and a 4.7 kb SalI fragment. Size-selected SalI restriction fragments of E. coli C600 DNA in the range of 4–5 kb were purified from a gel by the freeze-squeeze method (Tautz & Renz, 1983) and ligated into pBR322 digested with SalI.

After transformation of E. coli C600 approximately 2000 Ampr colonies were recovered, lysed in situ on nitrocellulose, and probed with the α-32P-labelled PstI–NruI fragment from the S. typhi aroC gene. A single colony showing hybridization to the probe contained a plasmid designated pAN10. This plasmid contained a 4.7 kb SalI fragment, and was capable of complementing the E. coli aroC mutation following re-transformation.

DNA sequencing of the S. typhi and E. coli aroC genes

The DNA sequences of the two aroC genes were determined using universal primer, and specific oligonucleotides to extend and overlap the sequences on both strands. Analysis of the DNA sequences revealed open reading frames encoding proteins of 361 amino acids of molecular mass 39 108 Da and 39 138 Da for the S. typhi and E. coli genes, respectively. Fig. 1 compares the S. typhi and E. coli DNA sequences, and the two deduced protein sequences.

Polypeptide expression from aroC recombinants in minicells

Plasmid pWAC6 encoding the S. typhi aroC gene was transferred into the minicell-producing strain DS410. Plasmid-harbouring minicells were purified, labelled with [35S]methionine and analysed using SDS-PAGE and autoradiography, following immunoprecipitation with antibody raised against E. coli chorismate synthase (kindly supplied by J. Coggins, University of Glasgow, UK). The results are shown in Fig. 2. The aroC-complementing plasmid pWAC6 produced a 40 kDa polypeptide (in addition to some other non-specific proteins that were also immunoprecipitated). This polypeptide is absent from the control minicell preparation containing pHCT9 vector alone (track 2) and is in good agreement with the value of 39.1 kDa for the deduced protein encoded by the S. typhi aroC gene, suggesting strongly that this polypeptide is the S. typhi chorismate synthase enzyme. The 40 kDa polypeptide could not be distinguished when whole-cell preparations of BRD049 harbouring aroC-positive recombinant plasmids were analysed using SDS-polyacrylamide gels stained with Coomassie blue (data not shown). Thus the chorismate synthase of S. typhi was expressed at a relatively low level in E. coli.

Putative regulatory elements for the aroC genes

Neither gene has exact matches with the consensus −10 and −35 promoter elements (Rosenberg & Court, 1979). However, putative −10 elements from positions 201–205 (S. typhi) and 406–411 (E. coli), and −35 elements from positions 178–183 (S. typhi) and 383–388 (E. coli) were identified by eye. The sequence GGAG is found at positions 292–295 (S. typhi) and 446–449 (E. coli) and shows good agreement with the consensus E. coli ribosome-binding site (Shine & Dalgarno, 1974). Imperfect inverted repeats are found at 3'-positions 1408–1436 (S. typhi) and 1562–1590 (E. coli), which have calculated ΔGº values (Tinoco et al., 1973) of −47.7 kJ mol⁻¹ (S. typhi), and −85.4 kJ mol⁻¹ (E. coli), but differ from typical E. coli rho-independent terminators in that they are not followed by a run of uridine residues.
Comparison of the S. typhi and E. coli aroC sequences

Alignment of the two sequences demonstrates very strong homology between the two coding regions of the genes, 95% at the protein level and 84% at the DNA level (Fig. 1). Many of the observed protein changes are conservative substitutions, both genes comprising 361 codons and using TAA as the termination codon.

Examination of the codon-usage table (data not shown) for the two genes shows a high degree of similarity, and is consistent with those E. coli genes that are weakly expressed (Granatham et al., 1981). There are some differences in codon preference, however; the codon for valine most often used in the E. coli aroC is GGT, while this is infrequently used in S. typhi. The G + C ratio for the S. typhi aroC gene is 59 mol% while for the E. coli aroC gene the G + C ratio is 57 mol%. There is less apparent homology when the 5' and 3' non-coding regions of the two sequences are compared. Both genes exhibit a region of imperfect inverted repeat sequence 22 bp after the TAA stop codon, although there are base variations in this 22 bp region, and in the inverted repeat itself. When the 5' non-coding regions of the two genes are compared, the direct base-for-base homology observed in the coding region disappears, although both sequences have a candidate ribosome-binding site with the sequence 5'-GGAG-3'.

There is considerable divergence between the deduced C-terminal sequences of the aroC-encoded proteins reported here and a recently reported E. coli aroC sequence (White et al., 1988). The differences arise from the fact that there is an additional A-base (at position 1457 in Fig. 1) in the sequence of White et al. (1988). This additional base leads to a change in the reading frame, producing a difference in the sequence of the 27 C-terminal amino acids. There are two other DNA sequence differences; at positions 126 and 151 an extra base leads to a change in the reading frame, codon for valine most often used in the E. coli aroC.

Comparison of the S. typhi and E. coli aroC sequences

Alignment of the two sequences demonstrates very strong homology between the two coding regions of the genes, 95% at the protein level and 84% at the DNA level (Fig. 1). Many of the observed protein changes are conservative substitutions, both genes comprising 361 codons and using TAA as the termination codon.

Examination of the codon-usage table (data not shown) for the two genes shows a high degree of similarity, and is consistent with those E. coli genes that are weakly expressed (Granatham et al., 1981). There are some differences in codon preference, however; the codon for valine most often used in the E. coli aroC is GGT, while this is infrequently used in S. typhi. The G + C ratio for the S. typhi aroC gene is 59 mol% while for the E. coli aroC gene the G + C ratio is 57 mol%.

There is less apparent homology when the 5' and 3' non-coding regions of the two sequences are compared. Both genes exhibit a region of imperfect inverted repeat sequence 22 bp after the TAA stop codon, although there are base variations in this 22 bp region, and in the inverted repeat itself. When the 5' non-coding regions of the two genes are compared, the direct base-for-base homology observed in the coding region disappears, although both sequences have a candidate ribosome-binding site with the sequence 5'-GGAG-3'.

There is considerable divergence between the deduced C-terminal sequences of the aroC-encoded proteins reported here and a recently reported E. coli aroC sequence (White et al., 1988). The differences arise from the fact that there is an additional A-base (at position 1457 in Fig. 1) in the sequence of White et al. (1988). This additional base leads to a change in the reading frame, producing a difference in the sequence of the 27 C-terminal amino acids. There are two other DNA sequence differences; at positions 126 and 151 an extra base leads to a change in the reading frame, codon for valine most often used in the E. coli aroC.

References


