View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Volume 200, number 1 FEBS 3604 May 1986

The complete amino acid sequence of 3-dehydroquinate synthase of *Escherichia coli* K12

Gary Millar and John R. Coggins

Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, Scotland

Received 5 March 1986

The complete amino acid sequence of the *Escherichia coli* 3-dehydroquinate synthase has been determined by a combined nucleotide and direct amino acid sequencing strategy. *E. coli* 3-dehydroquinate synthase is 362 amino acids long and has a calculated M_r of 38880. Analysis of the *aroB* nucleotide sequence and its 5′- and 3′-flanking regions has identified the *aroB* promoter elements and a possible 3′-terminator site.

3-Dehydroquinate synthase aroB *gene* (Escherichia coli) *Shikimate pathway*

1. INTRODUCTION

The cyclisation of 3-deoxy-D-*arabino*-heptulosonate-7-phosphate (DAHP) to 3-dehydroquinate (DHQ) is catalysed by the enzyme 3-dehydroquinate synthase (EC 4.6.1.3) [1]. Despite the low levels at which it naturally occurs, the enzyme catalysing the formation of this important aromatic intermediate has been investigated [2–5], and purification to homogeneity of monofunctional activities from *Escherichia coli, Bacillus subtilis* and mung bean has been reported [6–9]. In the fungus *Neurospora crassa*, the DHQ synthase activity is the first of 5 sequential steps in the shikimate pathway catalysed by the pentafunctional *arom* enzyme complex [10]. Of particular interest therefore is the evolutionary relationship of prokaryotic monofunctional DHQ synthases to their eukaryotic multifunctional counterparts.

The recent successful cloning of the *E. coli aroB* gene encoding DHQ synthase has led to the construction of overproducing strains [9,11] and has now allowed us to establish the complete amino acid sequence of the enzyme by a combination of direct amino acid sequencing and nucleotide sequencing.

2. MATERIALS AND METHODS

2.1. *Reagents*

The reagents for M13 cloning and sequencing were purchased as kits from Amersham (England). Restriction endonucleases were purchased from Bethesda Research Laboratories (Paisley, Scotland). [α-^{35}S]dATPαS was also purchased from Amersham. Avian myeloma virus reverse transcriptase was from Northumbria Biochemicals (Cramlington, England). Media reagents were from Difco (Detroit).

2.2. *Purification of DHQ synthase*

DHQ synthase was purified from *E. coli* strain AB2826/pGM107 essentially as described by Frost et al. [9].

2.3. *Automatic amino acid sequencing*

10 nmol *E. coli* DHQ synthase, reduced and carboxymethylated as described in [12], was sequenced as in [13] using a Beckman model 890 liquid-phase sequencer. The phenylthiohydantoin derivatives of liberated N-terminal amino acids were separated and identified by chromatography on a Waters Resolve C_{18} reverse-phase column with a pH 5.0 acetate-acetonitrile buffer gradient

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50 © 1986 Federation of European Biochemical Societies

system as in [14]. PTH-methionine and PTH-valine derivatives co-eluted under these conditions and were uniquely identified on a Waters Micro Bondapak column using the same buffer system.

2.4. *Amino acid composition analysis*

Performic acid-oxidised *E. coli* DHQ synthase was hydrolysed in 6 N HCl as in [15] and analysed on an LKB model 440D amino acid analyser as in [12].

2.5. *DNA manipulations*

Plasmid DNA manipulations used to generate, select and characterise chimeric plasmids followed the procedures of Maniatis et al. [16].

2.6. *Construction of pGM107*

Cloning of the *E. coli aroB* gene has been described [9,11].

Construction of the *aroB*-carrying plasmid pJB14 [9] had created two additional *Eco*RI restriction sites flanking the *Eco*RI site already present in genomic DNA used by Duncan and Coggins [11] to clone the *aroB* gene (see fig.1). *Eco*RI restriction digestion of pJB14 was carried out and the fragments separated by electrophoresis on a low-melting-temperature agarose gel. A 1.65 kbp DNA fragment was excised and the DNA purified [16]. This DNA was ligated into the expression plasmid pKK223/3 (Pharmacia), which had been cut with *Eco*RI and treated with calf-intestinal phosphatase, and the ligation mix was used to transform *E. coli* aroB strain AB2826 [18]. Cells were plated on L agar containing ampicillin (50 μg/ml). Following overnight growth, colonies were replica-plated onto minimal medium. Ten colonies growing on minimal medium were picked and plasmid DNA extracted and analysed by restriction enzyme analyses [19]. A construct carrying the 1.65 kbp insert in the orientation shown was selected and designated pGM107 (fig.1). An identical procedure was used to construct plasmid pGM108 except the plasmid vector used was suitably treated pAT153 [21]. The genomic inserts in plasmids pGM107 and pGM108 were in the same orientations with respect to the plasmid origin of replication.

2.7. *Preparation of DNA for sequencing*

The DNA fragments of several *Eco*RI digests of pGM107 were separated by gel electrophoresis on low-melting-temperature agarose. The *aroB*-containing DNA fragments were excised and subjected to a secondary digest with one of the restriction enzymes *Hpa*II, *Taq*I or *Sau*3A. The resultant unfractionated mixes were purified by phenol/chloroform extraction and ethanol precipitation [16]. In addition, an intact 1.65 kbp *Eco*RI fragment and a 280 bp *Hinc*II fragment of insert DNA were purified. Extensive use was made of the Messing series of M13 vectors [20] to clone these genomic DNA fragments. Single-stranded template M13 DNA was prepared as in [22] and stored at $-20°C$ prior to sequencing.

2.8. *DNA sequencing*

Construction of recombinant M13 clones allowed the use of the dideoxy chain-termination method for DNA sequencing [17,23]. Template and primer were annealed and DNA sequencing reactions performed as in [22]. Electrophoresis of reaction mixtures containing incorporated [α-^{35}S]dATPαS was on 6% polyacrylamide gels ($20 \times 40 \times 0.04$ cm) [22]. Gels were dried on Whatman 3 MM paper using a Bio-Rad model 1125 gel dryer and autoradiography was overnight at room temperature with Fuji RX film.

2.9. *Transcript mapping*

RNA was prepared [27] from exponentially growing *E. coli* AB2826 transformed with either pGM107 or pGM108. A synthetic oligonucleotide primer, 25 nucleotides long (sequence 5′-GGTAATTGGGTAACTACGTTCCCCG-3′), complementary to nucleotides 24–48 in the *aroB* coding sequence was annealed to the RNA at 55°C for 30 min in 10 mM Tris-HCl buffer (pH 8.5), containing 10 mM $MgCl_2$. Primer extension synthesis of the run-off transcripts was carried out by adding 10 μCi [α-^{35}S]dATPαS and dGTP, dCTP and dTTP to a final concentration of 0.1 mM and avian myeloma virus reverse transcriptase (23 units). Incubation was at 30°C for 30 min. Run-off transcripts were analysed using polyacrylamide sequencing gels.

3. RESULTS AND DISCUSSION

Takeda et al. [25] have shown that the Clarke and Carbon plasmid pLC29-47 [24] carries the

ponA gene which lies adjacent to *aroB* on the *E. coli* chromosome. Independent sub-cloning from pLC29-47 has further demonstrated that the *aroB* gene lies within a 3.6 kbp *Eco*RI-*Bgl*II fragment (pKD106) [11] and on a 2.5 kbp *Msp*I fragment (pJB14) [9]. Restriction mapping and hybridisation analysis (not shown) established that these two *aroB* clones both contained a common 1.65 kbp region of *E. coli* genomic DNA. This 1.65 kbp fragment was further sub-cloned to yield plasmids pGM107 and pGM108, each of which retained both the ability to relieve the auxotrophic requirements of *E. coli* AB2826 and to overexpress 3-dehydroquinate synthase activity. The 1.65 kbp sub-clone, PGM107 (fig.1), was used as starting material for DNA sequence analysis, and the transformed strain AB2826/pGM107 used as source for protein purification.

The DNA sequencing strategy is summarised in fig.2, and involved the alignment of the sequences obtained from the various *Hpa*II, *Hinc*II, *Sau*3A and *Taq*I M13 sub-clones until all the restriction sites that had been used were overlapped and the DNA sequence had been determined on both strands. The complete sequence, including the *Eco*RI recognition sites used to generate pGM107, totalled 1644 bp in length. The sequence was analysed on both strands for possible open reading frames and one, extending some 1100 bp from positions 438 to 1544 in the nucleotide sequence, was identified. Evidence that this was a coding region was obtained by using a computer program which examines the positional base frequencies in all 3 reading frames [28].

Two methionine (ATG) codons at positions 456 and 557, close to the beginning of the open reading frame were identified as possible translational initiation sites. Direct N-terminal amino acid sequencing of the purified 3-dehydroquinate synthase confirmed that this open reading frame was the *E. coli* DHQ synthase coding sequence and that the Met codon at position 456 was the translational start site. Located 11 bp upstream of this codon is the sequence GGTGG which exhibits homology with the consensus sequence for the *E. coli* ribosome binding site, GGAGG [29].

The complete *E. coli aroB* gene nucleotide sequence is shown in fig.4. The predicted amino acid sequence corresponds to a protein of calculated M_r 38880 containing 362 amino acid residues. SDS-polyacrylamide gel electrophoresis [26] of the purified 3-dehydroquinate synthase demonstrated that the enzyme was homogeneous (not shown) and had a subunit M_r of 38000–39000. Both values are in agreement with the published value for the purified enzyme [9]. The first 45 N-terminal amino acids were identified (fig.3) and agreed exactly with the amino acid sequence predicted from

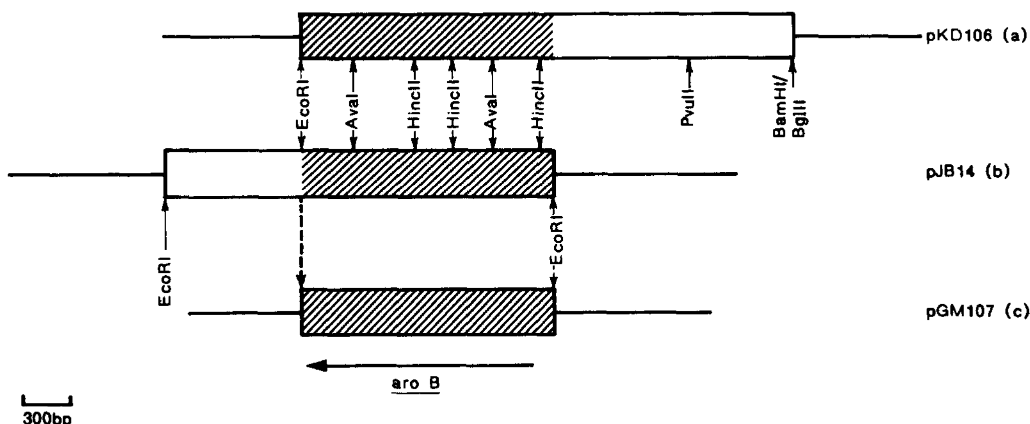

Fig.1. Construction of pGM107. Restriction enzyme digest profiles of both *aroB*-carrying plasmids pKD106 (a) and pJB14 (b) were compared and aligned. Genomic DNA is represented by the boxed areas, solid lines show vector DNA. The hatched area indicates the region of overlap sub-cloned as pGM107 (c). The direction of transcription of the *aroB* is shown. Note the two additional *Eco*RI sites, flanking the genomic *Eco*RI site, in pJB14 generated as cloning artefacts. Plasmid pGM108 contains the same genomic insert as pGM107 except cloned in pAT153.

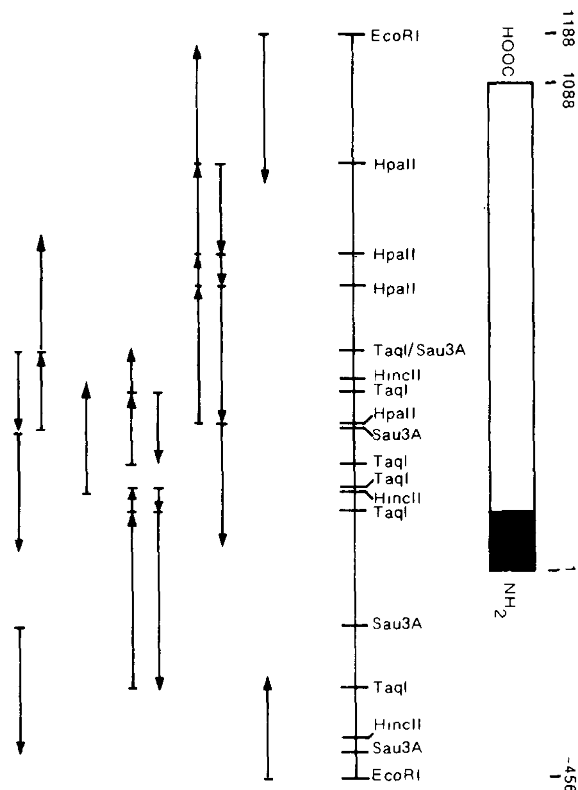

Fig.2. Sequencing strategy for the *aroB* gene. The boxed area above the restriction map represents the 3-dehydroquinate synthase coding region; the shaded region defines the extent of the N-terminal amino acid sequencing. The length and orientation of sub-clones used in the sequence analysis are represented by arrows and the relevant restriction sites are shown.

the nucleotide data. The amino acid composition determined experimentally was also in good agreement with that deduced from the DNA sequence (table 1). Transcript mapping analysis of the *aroB*-encoding mRNA using reverse transcriptase-directed DNA synthesis from a synthetic oligonucleotide primer identified an A residue (position 356), 100 bp upstream of the translational start site, as the 5′-end of the *aroB* mRNA (fig.4). Secondary structure formation at the 5′-end of the mRNA results in a number of premature terminations of run-off transcripts (see, for example, position (ii) in fig.5). However, the longest run-off band (indicated at position (i) in fig.5) maps accurately at position 356. Analysis of RNA prepared from *E. coli* AB2826 transformed

1 10

Met - Glu - Arg - Ile - Val - Val - Thr - Leu - Gly - Glu -

11 20

Arg - Ser - Tyr - Pro - Ile - Thr - Ile - Ala - Ser - Gly -

21 30

Leu - Phe - Asn - Glu - Pro - Ala - Ser - Phe - Leu - Pro -

31 40

Leu - Lys - Ser - Gly - Glu - Gln - Val - Met - Leu - Val -

41 45

(-) - Asn - Glu - (-) - Leu

Fig.3. The N-terminal amino acid sequence of 3-dehydroquinate synthase. The sequence was determined on a liquid-phase sequencer as described in section 2.3. The initial amount of protein sequenced was 5.5 nmol and the repetitive yield from residue 1 to 45, by least-squares regression analysis, was 95%. Residues 41 and 44 could not be unambiguously identified.

with either pGM107 or pGM108 gave the same result. The location of both the full-length reverse run-off transcript band and the major premature termination species are identical for both pGM107- and pGM108-derived *aroB* mRNAs. This identifies the natural *aroB* promoter and precludes the possibility that expression of *aroB* in construct pGM107 may be from an artificially long transcript originating at or near the powerful vector *tac* promoter.

Immediately upstream of the transcriptional start site are sequences with considerable homology to the consensus *E. coli* -35 and -10 promoter sequences elements [30]. The *aroB* promoter -10 region (GATGAT) is separated by 17 bp, the optimal separation, from the -35 region (TTGCCG). Analysis of the *aroB* promoter elements reveals that only 3 strongly conserved base pairs differ from those of the 'standard' *E. coli* promoter [30]. Immediately 3′- to the end of the *aroB* gene, and starting at position 1588 (fig.4), is a sequence capable of forming a stem-loop structure characteristic of a rho-independent terminator (fig.6). A free energy of formation (ΔG) was

```
   1  GAATTCATCGGAGCTGATGTGGGCTGGGTTTTCGATTTAGAAGGCGAAGAAGGCTTCCG

  60  CGATCGCGAAGAAAAGGTCATCAATGAGTTGACCGAGAAACAGGGTATTGTGCTGGCTAC

 120  TGGCGGCGGCTCTGTGAAATCCCGTGAAACGCGTAACCGTCTTTCCGCTCGTGCGTTGTC

 180  GTTTATCTTGAAACGACCATCGAAAAGCAACTTGCACGCACGCAGCGGATAAAAAACGCC

 240  CGTTGCTGCACGTTGAAACACCGCCGCGTGAAGTTCTGGAAGCGTTGGCCAATGAACGAA

                     tcTTGACat   t  .  t  tg  TAtAaT .        +1
 300  TCCGCTGTATGAAGAGATTGCCGACGTGACCATTCGTACTGATGATCAAAGCGCTAAAGT
                             -35                     -10        +---->

 360  GGTTGCAAACCAGATTATTCACATGCTGGAAAGCAACTAATTCTGGCTTTATATACACTC

 420  GTCTGCGGGTACAGTAATTAAGGTGGATGTCGCGTTATGGAGAGGATTGTCGTTACTCTC
                               RBS           MetGluArgIleValValThrLeu [8]

 480  GGGGAACGTAGTTACCCAATTACCATCGCATCTGGTTTGTTTAATGAACCAGCTTCATTC
      GlyGluArgSerTyrProIleThrIleAlaSerGlyLeuPheAsnGluProAlaSerPhe [28]

 540  TTACCGCTGAAATCGGGCGAGCAGGTCATGTTGGTCACCAACGAAACCCTGGCTCCTCTG
      LeuProLeuLysSerGlyGluGlnValMetLeuValThrAsnGluThrLeuAlaProLeu [48]

 600  TATCTCGATAAGGTCCGCGGCGTACTTGAACAGGCGGGTGTTAACGTCGATAGCGTTATC
      TyrLeuAspLysValArgGlyValLeuGluGlnAlaGlyValAsnValAspSerValIle [68]

 660  CTCCCTGACGGCGAGCAGTATAAAAGCCTGGCTGTACTCGATACCGTCTTTACGGCGTTG
      LeuProAspGlyGluGlnTyrLysSerLeuAlaValLeuAspThrValPheThrAlaLeu [88]

 720  TTACAAAAACCGCATGGTCGCGATACTACGCTGGTGGCGCTTGGCGGCGGCGTAGTGGGC
      LeuGlnLysProHisGlyArgAspThrThrLeuValAlaLeuGlyGlyGlyValValGly [108]

 780  GATCTGACCGGCTTCGCGGCGGCGAGTTATCAGCGCGGTGTCCGTTTCATTCAAGTCCCG
      AspLeuThrGlyPheAlaAlaAlaSerTyrGlnArgGlyValArgPheIleGlnValPro [128]

 840  ACGACGTTACTGTCGCAGGTCGATTCCTCCGTTGGCGGCAAAACTGCGGTCAACCATCCC
      ThrThrLeuLeuSerGlnValAspSerSerValGlyGlyLysThrAlaValAsnHisPro [148]

 900  CTCGGTAAAAACATGATTGGCGCGTTCTACCAACCTGCTTCAGTGGTGGTGGATCTCGAC
      LeuGlyLysAsnMetIleGlyAlaPheTyrGlnProAlaSerValValValAspLeuAsp [168]

 960  TGTCTGAAAACGCTTCCCCCGCGTGAGTTAGCGTCGGGGCTGGCAGAAGTCATCAAATAC
      CysLeuLysThrLeuProProArgGluLeuAlaSerGlyLeuAlaGluValIleLysTyr [188]

1020  GGCATTATTCTTGACGGTGCGTTTTTTAACTGGCTGGAAGAGAATCTGGATGCGTTGTTG
      GlyIleIleLeuAspGlyAlaPhePheAsnTrpLeuGluGluAsnLeuAspAlaLeuLeu [208]

1080  CGTCTGGACGGTCCGGCAATGGCGTACTGTATTCGCCGTTGTTGTGAACTGAAGGCAGAA
      ArgLeuAspGlyProAlaMetAlaTyrCysIleArgArgCysCysGluLeuLysAlaGlu [228]

1140  GTTGTCGCCGCCGACGAGCGCGAAACCGGGTTACGTGCTTTACTGAATCTGGGACACACC
      ValValAlaAlaAspGluArgGluThrGlyLeuArgAlaLeuLeuAsnLeuGlyHisThr [248]

1200  TTTGGTCATGCCATTGAAGCTGAAATGGGGTATGGCAATTGGTTACATGGTGAAGCGGTC
      PheGlyHisAlaIleGluAlaGluMetGlyTyrGlyAsnTrpLeuHisGlyGluAlaVal [268]

1260  GCTGCGGGTATGGTGATGGCGGCGCGGACGTCGGAACGTCTCGGGCAGTTTAGTTCTGCC
      AlaAlaGlyMetValMetAlaAlaArgThrSerGluArgLeuGlyGlnPheSerSerAla [288]

1320  GAAACGCAGCGTATTATAACCCTGCTCAAGCGGGCTGGGTTACCGGTCAATGGGCCGCGC
      GluThrGlnArgIleIleThrLeuLeuLysArgAlaGlyLeuProValAsnGlyProArg [308]

1380  GAAATGTCCGCGCAGGCGTATTTACCGCATATGCTGCGTGACAAGAAAGTCCTTGCGGGA
      GluMetSerAlaGlnAlaTyrLeuProHisMetLeuArgAspLysLysValLeuAlaGly [328]

1440  GAGATGCGCTTAATTCTTCCGTTGGCAATTGGTAAGAGTGAAGTTCGCAGCGGCGTTTCG
      GluMetArgLeuIleLeuProLeuAlaIleGlyLysSerGluValArgSerGlyValSer [348]

1500  CACGAGCTTGTTCTTAACGCCATTGCCGATTGTCAATCAGCGTAACAACAAGAAAGGTCA
      HisGluLeuValLeuAsnAlaIleAlaAspCysGlnSerAlaEnd [362]

1560  GGCCGCTTATCAAGCGTCTATTAGCTTCAGGTTAATTGCAACGTGGTAAGCATTAACCTT

1620  TTAGTGGGGTGTTAATGGTGAATTC  1644
```

Fig.4. The complete *aroB* coding sequence together with the 3′- and 5′-flanking sequences. The DNA sequence of the complete 1644 bp *Eco*RI fragment cloned in pGM107 and pGM108 is shown and numbered. The numbered amino acids in the *aroB* coding region are indicated [ ]. The translational start site is preceded by the proposed ribosome binding site (RBS). The *aroB* transcriptional start site is indicated (+1) and the promoter sequences upstream compared with the consensus −35 and −10 sequences. The location of an inverted repeat 3′- to the end of the *aroB* gene is indicated by overlining.

calculated by the rules of Tinoco et al. [31] and the negative value obtained suggests that this stem loop may form a stable secondary structure in vivo. This sequence may form part of an *aroB* terminator. The codon utilisation of the *aroB* gene

Table 1

The amino acid composition of *E. coli* 3-dehydroquinate synthase compared with that predicted from the DNA sequence of the *E. coli aroB* gene

Amino acid	Relative amino acid composition based on Leu = 49 residues	Theoretical amino acid composition deduced from *aroB* DNA sequence
Cys[c]	5.1	5
Asx	26.1	26
Met[a]	9.2	10
Thr[b]	16.2	18
Ser[b]	21.1	20
Glx	40.5	36
Pro	16.5	16
Gly	36.5	34
Ala[b]	36.8	39
Val	30.0	32
Ile	13.7	16
Leu	49	49
Tyr	6.1	9
Phe	9.8	10
His	6.7	7
Lys	12.8	13
Arg	20.6	20
Trp	nd	2

[a] Determined as methionine sulphone
[b] Experimental values extrapolated to zero time
[c] Determined as cysteic acid

nd, not determined

has been examined, and is similar to that of other *E. coli* genes which are weakly expressed [36,37].

Although differing in divalent metal ion requirements [1,32] a common feature of the DHQ synthase activities of *N. crassa* and *E. coli* is a catalytic requirement for NAD^+ [1,32]. It is recognised that many nucleotide-binding proteins contain a region of conserved amino acid sequence [33,38] involved in nucleotide binding. One such structural motif (or 'fingerprint') has been described in detail for several NAD^+- and FAD-linked enzymes [33]. The key structural feature is a conserved amino acid sequence contained within a β-α-β secondary structure [33]. This conserved sequence contains 3 glycine residues arranged Gly-X-Gly-X-X-Gly, an invariant hydrophobic amino acid 4 residues before this sequence, and an invariant negatively charged amino acid 18–20

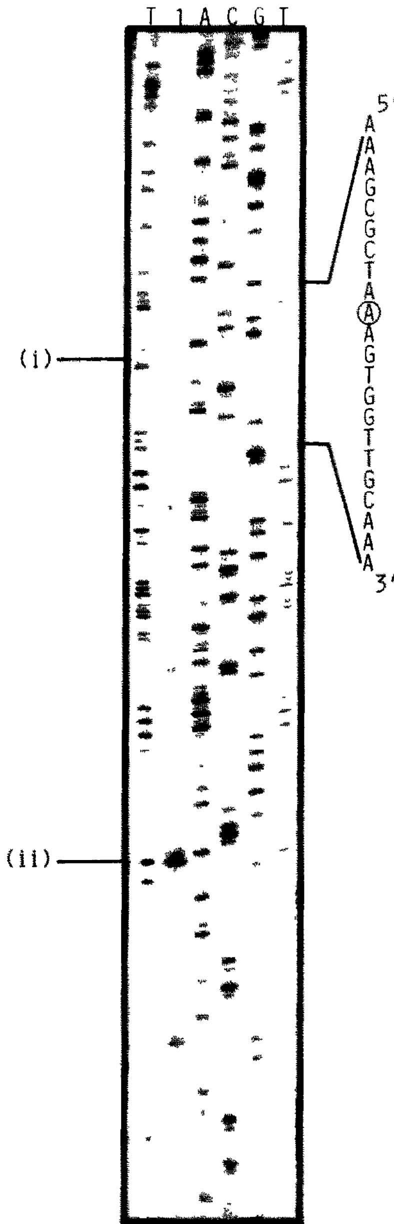

Fig.5. Transcript mapping of the *aroB* gene. Tracks A, C, G, T represent the relevant sequencing ladders of a recombinant M13 clone carrying the *E. coli aroB* gene. Track 1 shows the reverse run-off products of oligonucleotide primer extension of pGM107-derived mRNA. The same oligonucleotide primer was used for both the sequencing and the reverse run-off synthesis reactions. Indicated as (i) is the full-length reverse run-off band; the significance of band (ii) is discussed in the text. The sequence at the transcriptional start site is shown and the exact start encircled.

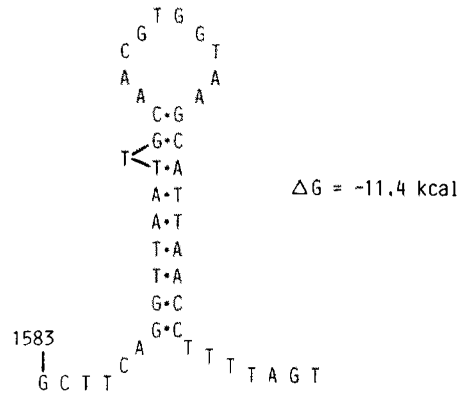

Fig.6. Possible *aroB* terminator sequence. Shown is the sequence 3′- to the end of the *E. coli aroB* gene indicated by overlining in fig.4) as a possible stem loop structure.

residues after the last conserved glycine. Also contained within this 30–32 residue 'domain' are several conserved hydrophobic residues forming the hydrophobic core of the unit [33]. Alignment of this adenine nucleotide-binding site has been demonstrated for several flavoproteins and dehydrogenases that bind FAD or NAD^+, including *E. coli* NADH dehydrogenase, human glutathione reductase, dogfish lactate dehydrogenase, *E. coli* lipoamide dehydrogenase and *E. coli* chorismate mutase/prephenate dehydrogenase [34]. Although the *E. coli* DHQ synthase uses NAD^+ as a prosthetic group and not as a coenzyme [3], a comparison of the enzyme sequence with that of the proposed fingerprint does reveal some interesting similarities. Residues 96–126 of the *E. coli* DHQ synthase sequence have a predicted secondary structure of β-α-β [35] which with the exception of the invariant negatively charged residues (where the *E. coli* DHQ synthase sequence has a Gln residue) contains all of the conserved features of this fingerprint model. The availability of overproducing strains that readily yield 100 mg quantities of *E. coli* 3-dehydroquinate synthase [9] should soon allow X-ray crystallography to be used to confirm the significance of this homology.

ACKNOWLEDGEMENTS

We thank Professor J.R. Knowles, Harvard University, for a gift of plasmid pJB14. We also thank Professor J.G. Fothergill, Dr L.A.

Fothergill-Gilmore and Mr B. Dunbar (University of Aberdeen Protein Sequencing Facility) for help with the amino acid sequencing, and Mr J. Jardine (University of Glasgow) for assistance with amino acid analyses. We also thank The University of Glasgow for a studentship and the Science and Engineering Research Council for financial support.

REFERENCES

[1] Srinivasan, P.R., Rothschild, J. and Sprinson, D.B. (1963) J. Biol. Chem. 238, 3176–3182.
[2] Rotenberg, S.L. and Sprinson, D.B. (1970) Proc. Natl. Acad. Sci. USA 67, 1669–1672.
[3] Maitra, W.I. and Sprinson, D.B. (1978) J. Biol. Chem. 253, 5426–5430.
[4] Le Marechal, P. and Azerad, R. (1976) Biochimie 58, 1123–1128.
[5] Le Marechal, P., Froussios, C., Level, M. and Azerdd, R. (1980) Biochem. Biophys. Res. Commun. 92, 1104–1109.
[6] Maitra, W.I. and Sprinson, D.B. (1978) J. Biol. Chem. 253, 5426–5430.
[7] Hasan, N. and Nester, E.W. (1978) J. Biol. Chem. 253, 4999–5004.
[8] Yamamoto, E. (1980) Phytochemistry 19, 779–781.
[9] Frost, J., Bender, J.L., Kadonaga, J.T. and Knowler, J.R. (1984) Biochemistry 23, 4470–4475.
[10] Giles, N.H., Case, M.E., Partridge, C.W.H. and Ahmed, S.I. (1967) Proc. Natl. Acad. Sci. USA 58, 1453–1460.
[11] Duncan, K. and Coggins, J.R. (1984) Biochem. Soc. Trans. 2, 274–275.
[12] Lumsden, J. and Coggins, J.R. (1978) Biochem. J. 169, 441–444.
[13] Smith, M.A., Gerrie, L.M., Dunbar, B. and Fothergill, J.E. (1982) Biochem. J. 207, 253–260.
[14] Carter, P.E., Dunbar, B. and Fothergill, J.E. (1983) Biochem. J. 215, 565–571.
[15] Hirs, C.H.W. (1967) Methods Enzymol. 11, 197–199.
[16] Maniatis, T., Fritsch, E.F. and Sambrook, T. (1982) Molecular Cloning, Cold Spring Harbour, NY.
[17] Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963–3965.
[18] Pittard, J. and Wallace, B.J. (1966) J. Bacteriol. 91, 1494–1508.
[19] Holmes, D.S. and Quigley, M. (1981) Anal. Biochem. 114, 193–197.
[20] Messing, J. and Viera, J. (1982) Gene 19, 269–276.
[21] Twigg, A.J. and Sherratt, D. (1980) Nature 283, 216–218.
[22] M13 Cloning and Sequencing Handbook (1983) Amersham plc, Amersham, England.
[23] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[24] Clarke, L. and Carbon, J. (1979) Methods Enzymol. 68, 396–408.
[25] Takeda, Y., Nashimura, A., Nashimura, Y., Yamada, M., Yasuda, S., Suzuki, H. and Hirota, Y. (1981) Plasmid 6, 86–98.
[26] Laemmli, U.K. (1970) Nature 227, 680–685.
[27] Aiba, H., Adhya, S. and De Crombrugghe, B. (1981) J. Biol. Chem. 256, 11905–11910.
[28] Staden, R. (1984) Nucleic Acids Res. 12, 505–509.
[29] Shine, J. and Dalgarno, L. (1975) Nature 254, 34–38.
[30] Hawley, C.J. and McClure, W.R. (1983) Nucleic Acids Res. 11, 2237–2255.
[31] Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nat. New Biol. 246, 40–41.
[32] Lambert, J.M., Boocock, M.R. and Coggins, J.R. (1985) Biochem. J. 226, 817–829.
[33] Wierenga, R.K. and Hol, W.G.J. (1983) Nature 302, 842–844.
[34] Hudson, G.S. and Davidson, B.E. (1984) J. Mol. Biol. 180, 1023–1051.
[35] Chou, P.Y. and Fasman, G.D. (1978) Annu. Rev. Biochem. 47, 251–277.
[36] Grosjean, H. and Fiers, W. (1982) Gene 18, 199–209.
[37] Gouy, M. and Gautiev, C. (1982) Nucleic Acids Res. 10, 7055–7074.
[38] Walker, J.E., Saraste, M., Runswick, M.J. and Gray, M.J. (1982) EMBO J. 1, 945–951.