Structural and functional relationships between fumarase and aspartase

Nucleotide sequences of the fumarase (*fumC*) and aspartase (*aspA*) genes of *Escherichia coli* K12

Sally A. WOODS, John S. MILES, Ruth E. ROBERTS and John R. GUEST*
Department of Microbiology, University of Sheffield, Western Bank, Sheffield S10 2TN, U.K.

1. The nucleotide sequences of two segments of DNA (2250 and 2921 base-pairs) containing the functionally related fumarase (*fumC*) and aspartase (*aspA*) genes of *Escherichia coli* K12 were determined. 2. The *fumC* structural gene comprises 1398 base-pairs (466 codons, excluding the initiation codon), and it encodes a polypeptide of M_r 50353 that resembles the fumarases of *Bacillus subtilis* 168 (*citG*-gene product), rat liver and pig heart. The *fumC* gene starts 140 base-pairs downstream of the structurally-unrelated *fumA* gene, but there is no evidence that both genes form part of the same operon. 3. The *aspA* structural gene comprises 1431 base-pairs (477 codons excluding the initiation codon), and it encodes a polypeptide of M_r 52190, similar to that predicted from maxicell studies and for the enzyme from *E. coli* W. 4. Remarkable homologies were found between the primary structures of the fumarase (*fumC* and *citG*) and aspartase (*aspA*) genes and their products, suggesting close structural and evolutionary relationships.

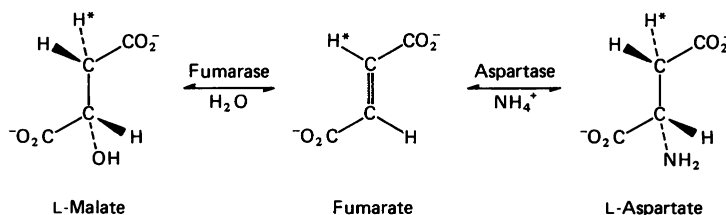
INTRODUCTION

Fumarase (EC 4.2.1.2) and aspartase (EC 4.3.1.1) catalyse analogous reactions involving the reversible hydration or amination of fumarate with the formation of L-malate or L-aspartate respectively (Scheme 1).

Fumarase is widely distributed in animals, plants and micro-organisms. It functions in the citric acid cycle during aerobic respiratory metabolism, and in facultative bacteria (e.g. *Escherichia coli*) it also functions in the reductive conversion of oxaloacetate into succinate during glucose fermentation. Most is known about the mammalian fumarases such as the pig heart and rat liver enzymes, which are extensively characterized tetramers (M_r approx. 194000) of identical subunits (Kanarek *et al.*, 1964; Kobayashi *et al.*, 1981; Kobayashi & Tuboi, 1983).

The existence of an *E. coli* aspartase was first suggested by Harden (1901), and Quastel & Woolf (1926) later established the stoichiometry of the reaction. Since then the enzyme has been found in various bacteria, plants and some animal tissues (Suzuki *et al.*, 1973). The enzymes from *E. coli* B, *E. coli* W and *Pseudomonas fluorescens* have been studied the most (Rudolph & Fromm, 1971; Suzuki *et al.*, 1973; Takagi *et al.*, 1984). They resemble the fumarases in size and quaternary structure. In *E. coli* aspartase participates in glutamate catabolism (Marcus & Halpern, 1969), and in the anaerobic production of fumarate by an alternative and differentially regulated route to that involving fumarase (Gray *et al.*, 1966; Courtright & Henning, 1970; Smith & Neidhardt, 1983).

Studies with transition-state analogues suggest that the reactions catalysed by fumarase and aspartase (Scheme 1) may be mechanistically related (Porter & Bright, 1980). Furthermore, the observations that the enzymes perform equivalent and potentially evolutionarily related functions in anaerobic metabolism suggest that they might also be structurally related. This has prompted a detailed molecular analysis of the corresponding genes. As a result, three fumarase genes from *E. coli* K12 (*fumA*, *fumB* and *fumC*) and one from *Bacillus subtilis* 168 (*citG*) have been cloned (Guest & Roberts, 1983; Moir *et al.*, 1984; Guest *et al.*, 1985). The *fumA* and *fumB* genes are homologous genes that are located at 35.5 and 93.5 min respectively in the *E. coli* linkage map. They encode analogous products of M_r 61000 and complement the deficiency of a fumarase mutant (EJ1535, *fumA1*) in single-copy (*fumA*) and multicopy situations (*fumA* and *fumB*). The *citG* gene encodes an unrelated product of M_r 50425, which nevertheless complements the *E. coli fumA1* strain and probably represents the *B. subtilis* fumarase.



Scheme 1. Stereospecificities of the analogous reactions catalysed by fumarase and aspartase

The *trans* nature of the processes is highlighted by the asterisks.

Abbreviation used: kb, kilobase.
* To whom correspondence should be addressed.

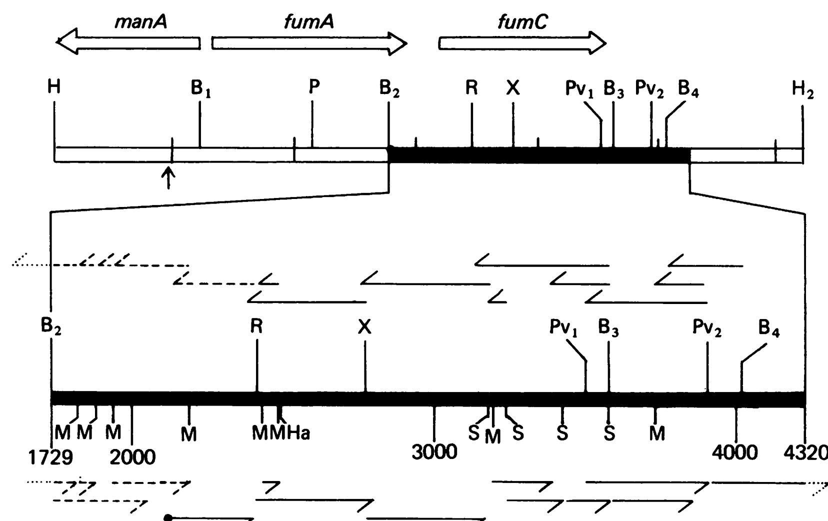


Fig. 1. Organization of two fumarase genes (*fumA* and *fumC*) of *E. coli* K12 and summary of the DNA sequence data derived from M13 clones

A restriction map of the 6.2 kb *Hin*dIII fragment encoding the fumarase (*fumA* and *fumC*) and phosphomannose isomerase (*manA*) genes is shown: scale in kilobase-pairs (kb). Left to right corresponds to anticlockwise in the *E. coli* linkage map and the positions and polarities of the structural genes are indicated by the open arrows. The expanded section illustrates the sequencing strategy for the *fumC* region. The arrows denote the positions and extents of sequences derived from both DNA strands. Dashed arrows indicate sequences that have been described previously (Miles & Guest, 1984*a*). The nucleotide co-ordinates (in base-pairs) are numbered from the first base of an *Msp*I site that is upstream of the *fumA* gene and marked by the small vertical arrow. The sequence derived with a specific primer is indicated by the black dot. The restriction sites are abbreviated as follows: *Bam*HI, Ba; *Bcl*I, B; *Eco*RI, R; *Hae*III, Ha; *Hin*dIII, H; *Msp*I, M; *Pst*I, P; *Pvu*II, Pv; *Sau*3A, S; *Xho*I, X.

Hybridization studies with the *citG* gene have recently revealed the presence of an analogous *E. coli* gene (*fumC*) lying adjacent to the *fumA* gene and formerly designated *g*48 because it encodes a product of M_r 48000 (Guest *et al.*, 1985). The nucleotide sequences of the *fumA* and *citG* genes have been defined, as has the 5′ end of the *fumC* (*g*48) coding region (Miles & Guest, 1984*a*, 1985). The functional significance of the apparent multiplicity of *E. coli* fumarase genes is not known.

The aspartase gene of *E. coli* K12 has been located at 94.1 min in the *E. coli* linkage map, quite close to the fumarase (*fumB*) and fumarate reductase (*frdABCD*) genes at 93.5 and 94.4 min respectively (Spencer *et al.*, 1976). The *aspA* gene was subsequently cloned from a ColE1-*frd*$^+$ hybrid plasmid (Guest *et al.*, 1984).

The present paper reports the complete nucleotide sequence of the *fumC* gene and confirms its homology with the corresponding *B. subtilis* gene (*citG*): the nucleotide sequence of the *aspA* gene of *E. coli* K12 is also presented. This has not only revealed remarkable primary structural homologies between the *fumC*- and *citG*-encoded fumarases and aspartase, but also some highly significant features relating to the catalytic functions of the two types of enzyme.

EXPERIMENTAL

Sources of DNA and cloning in M13

The sole source of DNA for sequencing the *fumC* gene was the pBR322-derived plasmid pGS54 (Guest & Roberts, 1983). This contains a 6.2 kb *Hin*dIII fragment encoding the *manA*, *fumA* and *fumC* (*g*48) genes (Fig. 1). The sequencing strategy involved directed cloning of specific fragments at appropriate sites in M13mp10 or M13mp11 (Messing & Vieira, 1982). These included the *Xho*I–*Bcl*I ($X–B_2$), *Eco*RI–*Hin*dIII ($R–H_2$), *Bcl*I ($B_2–B_3$ and $B_3–B_4$), *Pvu*II ($Pv_1–Pv_2$) and *Bcl*I–*Hin*dIII ($B_4–H_2$) fragments (Fig. 1). Several *Msp*I and *Sau*3A clones were obtained by the 'shot-gun' method starting with the $B_2–B_3$ fragment. The sequencing of the *Bcl*I–*Eco*RI segment (B_2–R) has been described previously (Miles & Guest, 1984*a*) except for the use of a synthetic oligonucleotide primer (S21) with the *Bcl*I clone ($B_2–B_3$) for completing the sequence on both strands.

The sources of DNA used for sequencing the *aspA* gene were primarily the 2.9 kb *Bcl*I fragment ($B_1–B_2$ in Fig. 3) of plasmid pGS73, and the 3.4 kb *Hin*dIII–*Sph*I fragment (H–Sp in Fig. 2) of plasmid pGS94, in which the $B_1–B_2$ fragment is flanked by two segments of pBR322 DNA as

Fig. 2. Nucleotide sequence of the *fumC* gene and primary structure of its product

The nucleotide sequence of 2250 base-pairs of the non-transcribed strand of the fumarase structural gene (*fumC*) and flanking regions, including the 3′ end of the *fumA* gene, is presented in the 5′-to-3′ direction. The nucleotide co-ordinates are assigned relative to the first base of an arbitrary *Msp*I site upstream of the *fumA* gene in accordance with Miles & Guest (1984*a*). The primary structures of the *fumC* gene product and the *C*-terminal end of the *fumA*-gene product are shown above the nucleotide sequence. A potential ribosome-binding site for *fumC* is boxed, and putative −35 and −10 (Pribnow) promoter sequences are indicated by lines and letters (A–C) above the nucleotide sequence. The translation initiation site is underlined, and stop sites are denoted by asterisks. Regions of hyphenated dyad symmetry are underlined by converging arrows and key restriction sites are indicated. Sequences shared with *frd* and *fnr* are denoted by wavy lines.

```
fumA
                                                            A
  C   T   R   C   V   K   *                                 B
AATGCACCCGCTGTGTGAAATAAACAGAGCGCCTTCGGGGCGTTTTTTTACATGGCACGAAAGACCAAACATTTGTTATCAAATGGTAAA
      1990      2000      2010      2020      2030      2040      2050      2060      2070
                                    C                              fumC        fM  N   T   V   R   S
TAATAAGTGAGCTAAAAGTTGCTTAACGAAAGCAAAACAGAAAGAAAAAATTAATCAGGTGAGGAGCAGGTCATGAATACAGTACGCAGC
      2080      2090      2100      2110      2120      2130      2140      2150      2160
          10                            20                            30
 E   K   D   S   M   G   A   I   D   V   P   A   D   K   L   W   G   A   Q   T   Q   R   S   L   E   H   F   R   I   S
GAAAAAGATTCGATGGGGGCGATTGATGTCCCGGCAGATAAGCTGTGGGGCGCACAAACTCAACGCTCGCTGGAGCATTTCCGCATTTCG
      2170      2180      2190      2200      2210      2220      2230      2240      2250
          40                            50                            60
 T   E   K   M   P   T   S   L   I   H   A   L   A   L   T   K   R   A   A   A   K   V   N   E   D   L   G   L   L   S
ACGGAGAAAATGCCCACCTCACTGATTCATGCGCTGGCGCTAACCAAGCGTGCAGCGGCAAAAGTTAATGAAGATTTAGGCTTGTTGTCT
      2260      2270      2280      2290      2300      2310      2320      2330      2340
          70                            80                   EcoRI    90
 E   E   K   A   S   A   I   R   Q   A   A   D   E   V   L   A   G   Q   H   D   D   E   F   P   L   A   I   W   Q   T
GAAGAGAAAGCGAGCGCCATTCGTCAGGCGGCGGATGAAGTACTGGCAGGACAGCATGACGACGAATTCCCGCTGGCTATCTGGCAGACC
      2350      2360      2370      2380      2390      2400      2410      2420      2430
          100                           110                           120
 G   S   G   T   Q   S   N   M   N   M   N   E   V   L   A   N   R   A   S   E   L   L   G   G   V   R   G   M   E   R
GGCTCCGGCACGCAAAGTAACATGAACATGAACGAAGTGCTGGCTAACCGGGCCAGTGAATTACTCGGCGGTGTGCGCGGGATGGAACGT
      2440      2450      2460      2470      2480      2490      2500      2510      2520
          130                           140                           150
 K   V   H   P   N   D   D   V   N   K   S   Q   S   S   N   D   V   F   P   T   A   M   H   V   A   A   L   L   A   L
AAAGTTCACCCTAACGACGACGTGAACAAAAGCCAAAGTTCCAACGATGTCTTTCCGACGGCGATGCACGTTGCGGCGCTGCTGGCGCTG
      2530      2540      2550      2560      2570      2580      2590      2600      2610
          160                           170                           180
 R   K   Q   L   I   P   Q   L   K   T   L   T   Q   T   L   N   E   K   S   R   A   F   A   D   I   V   K   I   G   R
CGCAAGCAACTCATTCCTCAGCTTAAAACCCTGACACAGACACTGAATGAGAAATCCCGTGCTTTTGCCGATATCGTCAAAATTGGTCGT
      2620      2630      2640      2650      2660      2670      2680      2690      2700
          190                           200                  XhoI     210
 T   H   L   Q   D   A   T   P   L   T   L   G   Q   E   I   S   G   W   V   A   M   L   E   H   N   L   K   H   I   E
ACTCACTTGCAGGATGCCACGCCGTTAACGCTGGGGCAGGAGATTTCCGGCTGGGTAGCGATGCTCGAGCATAATCTCAAACATATCGAA
      2710      2720      2730      2740      2750      2760      2770      2780      2790
          220                           230                           240
 Y   S   L   P   H   V   A   E   L   A   L   G   G   T   A   V   G   T   G   L   N   T   H   P   E   Y   A   R   R   V
TACAGCCTGCCTCACGTAGCGGAACTGGCTCTTGGCGGTACAGCGGTGGGTACTGGACTAAATACCCATCCGGAGTATGCGCGTCGCGTA
      2800      2810      2820      2830      2840      2850      2860      2870      2880
          250                           260                           270
 A   D   E   L   A   V   I   T   C   A   P   F   V   T   A   P   N   K   F   E   A   L   A   T   C   D   A   L   V   Q
GCAGATGAACTGGCAGTCATTACCTGTGCACCGTTTGTTACCGCGCCGAACAAATTTGAAGCGCTGGCGACCTGTGATGCCCTGGTTCAG
      2890      2900      2910      2920      2930      2940      2950      2960      2970
          280                           290                           300
 A   H   G   A   L   K   G   L   A   A   S   L   M   K   I   A   N   D   V   R   W   L   A   S   G   P   R   C   G   I
GCGCACGGCGCGTTGAAAGGGTTGGCTGCGTCACTGATGAAAATCGCCAATGATGTCCGCTGGCTGGCCTCTGGCCCGCGCTGCGGAATT
      2980      2990      3000      3010      3020      3030      3040      3050      3060
          310                           320                           330
 G   E   I   S   I   P   E   N   E   P   G   S   S   I   M   P   G   K   V   N   P   T   Q   C   E   A   L   T   M   L
GGTGAAATCTCAATCCCGGAAAATGAGCCGGGCAGCTCAATCATGCCGGGGAAAGTGAACCCAACACAGTGTGAGGCATTAACCATGCTC
      3070      3080      3090      3100      3110      3120      3130      3140      3150
          340                           350                           360
 C   C   Q   V   M   G   N   D   V   A   I   N   M   G   G   A   S   G   N   F   E   L   N   V   F   R   P   M   V   I
TGCTGTCAGGTGATGGGGAACGACGTGGCGATCAACATGGGGGGCGCTTCCGGTAACTTTGAACTGAACGTCTTCCGTCCAATGGTGATC
      3160      3170      3180      3190      3200      3210      3220      3230      3240
          370                           380                           390
 H   N   F   L   Q   S   V   R   L   L   A   D   G   M   E   S   F   N   K   H   C   A   V   G   I   E   P   N   R   E
CACAATTTCCTGCAATCGGTGCGCTTGCTGGCAGATGGCATGGAAAGTTTTAACAAACACTGCGCAGTGGGTATTGAACCGAATCGTGAG
      3250      3260      3270      3280      3290      3300      3310      3320      3330
          400                           410                           420
 R   I   N   Q   L   L   N   E   S   L   M   L   V   T   A   L   N   T   H   I   G   Y   D   K   A   A   E   I   A   K
CGAATCAATCAATTACTCAATGAATCGCTGATGCTGGTGACTGCGCTTAACACCCACATTGGTTATGACAAAGCCGCCGAGATCGCCAAA
      3340      3350      3360      3370      3380      3390      3400      3410      3420
          430                           440                               PvuII
 K   A   H   K   E   G   L   T   L   K   A   A   A   L   A   L   G   Y   L   S   E   A   E   F   D   S   W   V   R   P
AAAGCGCATAAAGAAGGGCTGACCTTAAAAGCTGCGGCCCTTGCGCTGGGGTATCTTAGCGAAGCCGAGTTTGACAGCTGGGTACGGCCA
      3430      3440      3450      3460      3470      3480      3490      3500      3510
          460
 E   Q   M   V   G   S   M   K   A   G   R   *                           BclI
GAACAGATGGTCGGCAGTATGAAAGCCGGGCGTTAATCTGCAACATACAGGTGCAGCCGTGGAATGATCAAACGAAGCGGCTGCGCCTGA
      3520      3530      3540      3550      3560      3570      3580      3590      3600

GGTTTATAACGGTGCTGCACATTGTCGGCATCGTAATTTAACAACTCACCAACGTCCGGCACGCCCGCGCCATTATCCCGATTAATCAGT
      3610      3620      3630      3640      3650      3660      3670      3680      3690

GCAATCAGTGGTGTAGGGCAGGCGTGTTGGACTTGTTTTGATCTCCTTTGTACCAGACGCGGGCAATCGGCTGCACCTTCACCGGACGT
      3700      3710      3720      3730      3740      3750      3760      3770      3780

TTGATTTTTAACTTCGCGTTCTGTGGCAGGGCAGCGATATCCTGATACTCTCGCTCCAGTTTTCTTTGCCACTCCTCGCGCGTCCACGGT
      3790      3800      3810      3820      3830      3840      3850      3860      3870
                                   PvuII
GCGACACTGCGTGGTGATTTCAGGCTTTTTCCAGCTGTGCCAGGACTTCATCACGATGTAAATTCTTAATGATATGTTTATTAGCCCAA
      3880      3890      3900      3910      3920      3930      3940      3950      3960
                                                             BclI
CCAAAGCGTAAAGTGGCGGGGTCGTGCAGAACGGTGAGCGTGCGGTAAGCATTAAGGGTGATCACGCCCGGCAAATGACGATGCACCCAT
      3970      3980      3990      4000      4010      4020      4030      4040      4050

TCAAAACGTGCCGCGGTGGGGATTTCTGATTCAACCGTGACGATATGCTCGAACGTGGCTTCGAGTTTATTGATGTGCTGAATATGACTG
      4060      4070      4080      4090      4100      4110      4120      4130      4140

ACCAACGCTGCTTGCGAAAGGTTATCCGAACCGGAACACAACAGCCAGGCAGACGGACAGTGGCCTTGCTGCTGCGATTCGGACTGTTGT
      4150      4160      4170      4180      4190      4200      4210      4220      4230
```

Table 1. Amino acid compositions of four fumarases and one aspartase

The amino acid compositions predicted for the fumarase (FumC) and aspartase (AspA) of *E. coli* K12 are compared with the DNA-derived composition for the fumarase (CitG) of *B. subtilis* (Miles & Guest, 1985) and the analytical values for the pig heart and rat liver enzymes (Kanarek *et al.*, 1964; Kobayashi *et al.*, 1981). The initiating formylmethionine residues are not included.

Amino acid	FumC (*E. coli*)	AspA (*E. coli*)	CitG (*B. subtilis*)	Fumarase (pig heart)	Fumarase (rat liver)
	Composition (residues/monomer)				
Asp	19	19	24 }	44.3	40.5
Asn	26	36	30 }		
Thr	25	24	28	25.0	29.0
Ser	28	21	26	23.0	22.5
Glu	33	37	35 }	43.5	46.0
Gln	18	19	18 }		
Pro	19	19	19	20.0	24.5
Gly	34	34	29	35.8	37.5
Ala	57	42	49	50.0	53.8
Val	29	44	30	32.8	33.8
Met	17	15	17	15.5	19.0
Ile	22	32	27	24.3	24.0
Leu	52	44	42	36.8	37.8
Tyr	4	16	7	10.0	9.0
Phe	11	13	14	15.8	16.0
Lys	24	27	34	32.3	31.8
His	15	8	11	13.5	12.0
Arg	21	16	14	13.5	12.5
Cys	7	11	3	3.0	4.0
Trp	5	0	4	2.0	1.3
Total	466	477	461	441.1	454.8
M_r . . .	50353	52190	50425	48600	49000

a result of subcloning into the *Bam*HI site of this vector (Guest *et al.*, 1984). A few specific clones were derived from a plasmid (pGS71) which contains only part of the *aspA* gene (Guest *et al.*, 1984). The sequencing strategy for the *aspA* gene involved a combination of 'shot-gun' and directed cloning with appropriately cut M13 vectors (M13mp8, M13mp10, M13mp18 and M11mp19; Norrander *et al.*, 1983). The 'shot-gun' clones were generated from *Msp*I, *Sau*3A and *Taq*I digests of the 3.4 kb hybrid H–Sp fragment (Fig. 2). The specific clones (and their sources) were as follows: B_1–K, K–B_2, B_2–B_1 and K–B_1 (from the isolated *Bcl*I fragment, B_1–B_2); P_2–P_3 and P_1–H (from a *Pst*I digest of H–Sp); P_2–P_1, P_3–P_2, P_3–P_4, P_4–P_3 and P_4–B_2 (from a *Pst*I plus *Sal*I digest of pGS71; B_2 is equivalent to the *Sal*I site on pBR322). Some sequence was obtained from clones containing the B_2–B_1 and B_1–K fragments by using synthetic oligonucleotide primers S6 and S7 respectively (Fig. 3).

Nucleotide sequence analysis

Single-stranded M13 templates were prepared and sequenced by the dideoxy chain-termination method by using a 17-nucleotide 'universal' primer, [α-^{35}S]dATP and salt-gradient gels (Sanger *et al.*, 1977, 1980; Biggin *et al.*, 1983). Nucleotide sequences were compiled and analysed by using the Staden computer programs (Stadan, 1977, 1980, 1982, 1984; Staden & McLachlan, 1982).

Sequence comparisons

Amino acid and nucleotide sequences were compared by using the proportional matching option of the graphics program DIAGON (Staden, 1982). This incorporates a scoring system for amino acids based on the MDM_{78} matrix, which was calculated from the accepted point mutations in 71 families of related proteins (Schwartz & Dayhoff, 1978).

Materials

Restriction endonucleases, bacteriophage-T_4 DNA ligase and DNA polymerase (Klenow fragment) were purchased from Bethesda Research Laboratories and Boehringer Corp. The 17-nucleotide 'universal' primer was from Celltech, the M13mp8, M13mp10, M13mp18 and M13mp19 replicative-form DNAs were from Pharmacia P-L Biochemicals, and [α-^{35}S]dATP was supplied by Amersham International. Two synthetic primers (S6 and S7, 21-mers) were kindly provided by G. D. Searle, Ltd., and another (S21, 17-mer) was a gift from the New Brunswick Instrument Corp.

RESULTS AND DISCUSSION

Nucleotide sequence and other features of the *fumC* region

The *fumC* gene (formerly *g48*) lies adjacent to the *fumA* gene in a 6.2 kb *Hin*dIII fragment of *E. coli* DNA (Fig. 1). It has been identified by hybridization studies with the *B. subtilis* fumarase gene (*citG*) and by *N*-terminal sequence comparisons between the *fumC* and *citG* gene products (Guest *et al.*, 1985). The strategy for sequencing the *fumC* region mainly involved the directed cloning of specific fragments from the *manA–fumA–fumC* plasmid pGS54 (see the Experimental section). The relevant M13

Table 2. Codon usage in the *fumC* and *aspA* genes

The AUG initiation codons are not included with the methionine codons. The codon pairs enclosed in boxes are those whose use varies between strongly and weakly expressed genes, and the asterisks denote potential modulatory codons (Grosjean & Fiers, 1982).

Gene:	*fumC*	*aspA*		*fumC*	*aspA*		*fumC*	*aspA*		*fumC*	*aspA*
UUU Phe	7	2	UCU Ser	2	6	UAU Tyr	3	5	UGU Cys	4	4
UUC Phe	4	11	UCC Ser	5	7	UAC Tyr	1	11	UGC Cys	3	7
UUA Leu	6	0	UCA Ser	4	2	UAA End	1	1	UGA End	0	0
UUG Leu	6	4	UCG Ser	5	0	UAG End	0	0	UGG Trp	5	0
CUU Leu	5	2	CCU Pro	3	2	CAU His	7	1	CGU Arg	9	8
CUC Leu	6	2	CCC Pro	1	0	CAC His	8	7	CGC Arg	9	6
*CUA Leu	2	0	CCA Pro	3	4	CAA Gln	7	4	*CGA Arg	1	0
CUG Leu	27	36	CCG Pro	12	13	CAG Gln	11	15	*CGG Arg	2	0
AUU Ile	11	11	ACU Thr	4	10	AAU Asn	11	3	AGU Ser	5	3
AUC Ile	11	21	ACC Thr	11	12	AAC Asn	15	33	AGC Ser	7	3
*AUA Ile	0	0	ACA Thr	5	1	AAA Lys	21	24	*AGA Arg	0	1
AUG Met	18	16	ACG Thr	5	1	AAG Lys	3	3	*AGG Arg	0	1
GUU Val	5	19	GCU Ala	7	13	GAU Asp	12	8	GGU Gly	8	22
GUC Val	7	6	GCC Ala	13	9	GAC Asp	7	11	GGC Gly	13	11
GUA Val	6	12	GCA Ala	11	13	GAA Glu	21	33	*GGA Gly	3	1
GUG Val	11	7	GCG Ala	26	7	GAG Glu	12	4	*GGG Gly	10	0

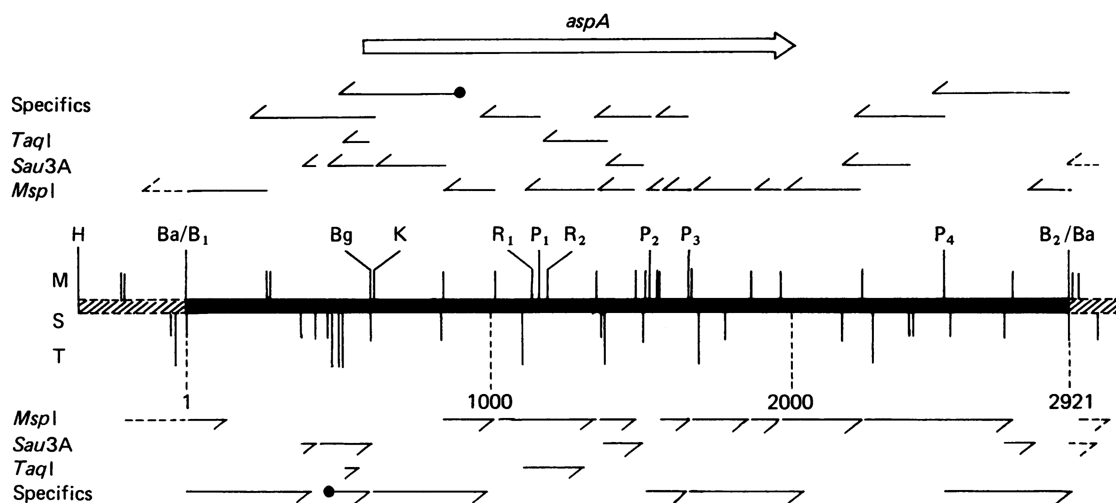


Fig. 3. Sequencing strategy for the aspartase gene (*aspA*) of *E. coli* K12

A restriction map is shown, with the relative position of the aspartase gene (*aspA*) in the 3.4 kb *Hin*dIII–*Sph*I fragment of plasmid pGS94. The 2.9 kb *Bcl*I fragment of bacterial DNA is cloned into the *Bam*HI site of plasmid pBR322 and is thus flanked by two segments of vector DNA (hatched regions). Left to right corresponds to anticlockwise in the *E. coli* linkage map. The arrows denote the positions and extents of sequences obtained from both strands of DNA, and the black dots indicate sequences derived with the aid of specific primers. The nucleotide co-ordinates are numbered in base-pairs from the first base of the *Bcl*I site in the bacterial DNA. Restriction sites identified in 'shot-gun' cloning are: *Msp*I, M; *Sau*3A, S; *Taq*I, T. Those used for directed cloning are: *Bcl*I, B; *Bam*HI, Ba; *Bgl*II, Bg; *Eco*RI, R; *Hin*dIII, H; *Kpn*I, K; *Pst*I, P; *Sph*I, Sp.

clones are summarized in Fig. 1 and the complete nucleotide sequence of a 2250-base-pair segment is shown in Fig. 2. The sequence is fully overlapped, it extends the previous sequence of the *manA–fumA* region by 1821 base-pairs to a total of 5285 base-pairs (Miles & Guest, 1984*a*,*b*), and it provides the complete sequence of the *fumC* gene on both DNA strands.

Potential coding regions were detected with FRAMESCAN (Staden & McLachlan, 1982), with the *E. coli* pyruvate dehydrogenase complex genes (Stephens *et al.*,

```
BclI
TGATCAGCGAAACACTTTTAATCATCTCCGCCGCTGGGTTTTCACCCGCCGCCATTTTTTGCTGCATCAGCACGAAATTCTTAAAGCCCTG
     10        20        30        40        50        60        70        80        90
        GTA      GA
GTTACGTACCAGTGACATACCGATAACTGACGTGAATATAACCAGCACGAGGGTCAGCAATACCCCCAATACATGGGCAACCTGAATAAA
    100       110       120       130       140       150       160       170       180
  TTATAT      A      TTATTT
GATTGAAATCTCAATATAGACATAAAGGAAAATGGCAATAAAAGGTAACCAGCGCAAAGGTTTCTCCTGTAATAGCAGCCGGTTAACCCC
    190       200       210       220       230       240       250       260       270
                      GAATAT      B       AAAGTT
GGCTACCTGAATGGGTTGCGAATCGCGTTTAGCTTATATTGTGGTCATTAGCAAAATTTCAAGATGTTTGCGCAACTATTTTTGGTAGTA
    280       290       300       310       320       330       340       350       360
                              C
ATCCCAAAGCGGTGATCTATTTCACAAATTAATAATTAAGGGGTAAAAACCGACACTTAAAGTGATCCAGATTACGGTAGAAATCCTCAA
    370       380       390       400       410       420       430       440       450
    AvaI                                              D
GCAGCATATGATCTCGGGTATTCGGTCGATGCAGGGGATAATCGTCGGTCGAAAAACATTCGAAACCACATATATTCTGTGTGTTTAAAG
    460       470       480       490       500       510       520       530       540
      aspA                                    BglII     KpnI
                              fM  S  N  N  I  R  I  E  E  D  L  L  G  T  R  E  V  P  A
CAAATCATTGGCAGCTTGAAAAAGAAGGTTCACATGTCAAACAACATTCGTATCGAAGAAGATCTGTTGGGTACCAGGGAAGTTCCAGCT
    550       560       570       580       590       600       610       620       630
 20                            30                            40
 D  A  Y  Y  G  V  H  T  L  R  A  I  V  N  F  Y  I  S  N  N  K  I  S  D  I  P  E  F  V  R
GATGCCTACTATGGTGTTCACACTCTGAGAGCGATTGTAAACTTCTATATCAGCAACAACAAAATCAGTGATATTCCTGAATTTGTTCGC
    640       650       660       670       680       690       700       710       720
 50                            60                            70
 G  M  V  M  V  K  K  A  A  A  M  A  N  K  E  L  Q  T  I  P  K  S  V  A  N  A  I  I  A  A
GGTATGGTAATGGTTAAAAAAGCCGCAGCTATGGCAAACAAAGAGCTGCAAACCATTCCTAAAAGTGTAGCGAATGCCATCATTGCCGCA
    730       740       750       760       770       780       790       800       810
 80                            90                            100
 C  D  E  V  L  N  N  G  K  C  M  D  Q  F  P  V  D  V  Y  Q  G  G  A  G  T  S  V  N  M  N
TGTGATGAAGTCCTGAACAACGGAAAATGCATGGATCAGTTCCCGGTAGACGTCTACCAGGGCGGCGCAGGTACTTCCGTAAACATGAAC
    820       830       840       850       860       870       880       890       900
 110                           120                           130
 T  N  E  V  L  A  N  I  G  L  E  L  M  G  H  Q  K  G  E  Y  Q  Y  L  N  P  N  D  H  V  N
ACCAACGAAGTGCTGGCCAATATCGGTCTGGAACTGATGGGTCACCAAAAAGGTGAATATCAGTACCTGAACCCGAACGACCATGTTAAC
    910       920       930       940       950       960       970       980       990
 140                           150                           160
 K  C  Q  S  T  N  D  A  Y  P  T  G  F  R  I  A  V  Y  S  S  L  I  K  L  V  D  A  I  N  Q
AAATGTCAGTCCACTAACGACGCCTACCCGACCGGTTTCCGTATCGCAGTTTACTCTTCCCTGATTAAGCTGGTAGATGCGATTAACCAA
   1000      1010      1020      1030      1040      1050      1060      1070      1080
 170                        EcoRI                            190 PstI
 L  R  E  G  F  E  R  K  A  V  E  F  Q  D  I  L  K  M  G  R  T  Q  L  Q  D  A  V  P  M  T
CTGCGTGAAGGCTTTGAACGTAAAGCTGTCGAATTCCAGGACATCCTGAAAATGGGTCGTACCCAGCTGCAGGACGCAGTACCGATGACC
   1090      1100      1110      1120      1130      1140      1150      1160      1170
 200   EcoRI                   210                           220
 L  G  Q  E  F  R  A  F  S  I  L  L  K  E  E  V  K  N  I  Q  R  T  A  E  L  L  L  E  V  N
CTCGGTCAGGAATTCCGCGCTTTCAGCATCCTGCTGAAAGAAGAAGTGAAAAACATCCAACGTACCGCTGAACTGCTGCTGGAAGTTAAC
   1180      1190      1200      1210      1220      1230      1240      1250      1260
 230                           240                           250
 L  G  A  T  A  I  G  T  G  L  N  T  P  K  E  Y  S  P  L  A  V  K  K  L  A  E  V  T  G  F
CTTGGTGCAACAGCAATCGGTACTGGTCTGAACACGCCGAAAGAGTACTCTCCGCTGGCAGTGAAAAAACTGGCTGAAGTTACTGGCTTC
   1270      1280      1290      1300      1310      1320      1330      1340      1350
 260                           270                           280
 P  C  V  P  A  E  D  L  I  E  A  T  S  D  C  G  A  Y  V  M  V  H  G  A  L  K  R  L  A  V
CCATGCGTACCGGCTGAAGACCTGATCGAAGCGACCTCTGACTGCGGCGCTTATGTTATGGTTCACGGCGCGCTGAAACGCCTGGCTGTG
   1360      1370      1380      1390      1400      1410      1420      1430      1440
 290                           300                           310             PstI
 K  M  S  K  I  C  N  D  L  R  L  L  S  S  G  P  R  A  G  L  N  E  I  N  L  P  E  L  Q  A
AAGATGTCCAAAATCTGTAACGACCTGCGCTTGCTCTCTTCAGGCCCACGTGCCGGCCTGAACGAGATCAACCTGCCGGAACTGCAGGCG
   1450      1460      1470      1480      1490      1500      1510      1520      1530
 320                           330                           340
 G  S  S  I  M  P  A  K  V  N  P  V  V  P  E  V  V  N  Q  V  C  F  K  V  I  G  N  D  T  T
GGCTCTTCCATCATGCCAGCTAAAGTAAACCCGGTTGTTCCGGAAGTGGTTAACCAGGTATGCTTCAAAGTCATCGGTAACGACACCACT
   1540      1550      1560      1570      1580      1590      1600      1610      1620
 350                     PstI                                370
 V  T  M  A  A  E  A  G  Q  L  Q  L  N  V  M  E  P  V  I  G  Q  A  M  F  E  S  V  H  I  L
GTTACCATGGCAGCAGAAGCAGGTCAGCTGCAGTTGAACGTTATGGAGCCGGTCATTGGCCAGGCCATGTTCGAATCCGTTCACATTCTG
   1630      1640      1650      1660      1670      1680      1690      1700      1710
 380                           390                           400
 T  N  A  C  Y  N  L  L  E  K  C  I  N  G  I  T  A  N  K  E  V  C  E  G  Y  V  Y  N  S  I
ACCAACGCTTGCTACAACCTGCTGGAAAAATGCATTAACGGCATCACTGCTAACAAAGAAGTGTGCGAAGGTTACGTTTACAACTCTATC
   1720      1730      1740      1750      1760      1770      1780      1790      1800
 410                           420                           430
 G  I  V  T  Y  L  N  P  F  I  G  H  H  N  G  D  I  V  G  K  I  C  A  E  T  G  K  S  V  R
GGTATCGTTACTTACCTGAACCCGTTCATCGGTCACCACAACGGTGACATCGTGGGTAAAATCTGTGCCGAAACCGGTAAGAGTGTACGT
   1810      1820      1830      1840      1850      1860      1870      1880      1890
 440                           450 E                         460
 E  V  V  L  E  R  G  L  L  T  E  A  E  L  D  D  I  F  S  V  Q  N  L  M  H  P  A  Y  K  A
GAAGTCGTTCTGGAACGCGGTCTGTTGACTGAAGCGGAACTTGACGATATTTTCTCCGTACAGAATCTGATGCACCCGGCTTACAAAGCA
   1900      1910      1920      1930      1940      1950      1960      1970      1980
 470
 K  R  Y  T  D  E  S  E  Q  *
AAACGCTATACTGATGAAAGCGAACAGTAATCGTACAGGGTAGTACAAATAAAAAAGGCACGTCAGATGACGTGCCTTTTTTCTTGTGAG
   1990      2000      2010      2020      2030      2040      2050      2060      2070
      F                                                                          10
                                           fM  L  V  V  E  L  I  I  V  L  L
CAGTAACTTAAAAATAACAATCTAATATCAACTTGTTAAAAAACAAGGAAGGCTAATATGCTAGTTGTAGAACTCATCATAGTTTTGCTG
   2080      2090      2100      2110      2120      2130      2140      2150      2160
                            20                            30                            40
 A  I  F  L  G  A  R  L  G  G  I  G  I  G  F  A  G  G  L  G  V  L  V  L  A  A  I  G  V  K
GCGATCTTCTTGGGCGCCAGATTGGGGGGAATAGGTATTGGTTTTGCAGGCGGATTGGGGGTGCTGGTTCTTGCCGCTATTGGCGTTAAA
   2170      2180      2190      2200      2210      2220      2230      2240      2250
```

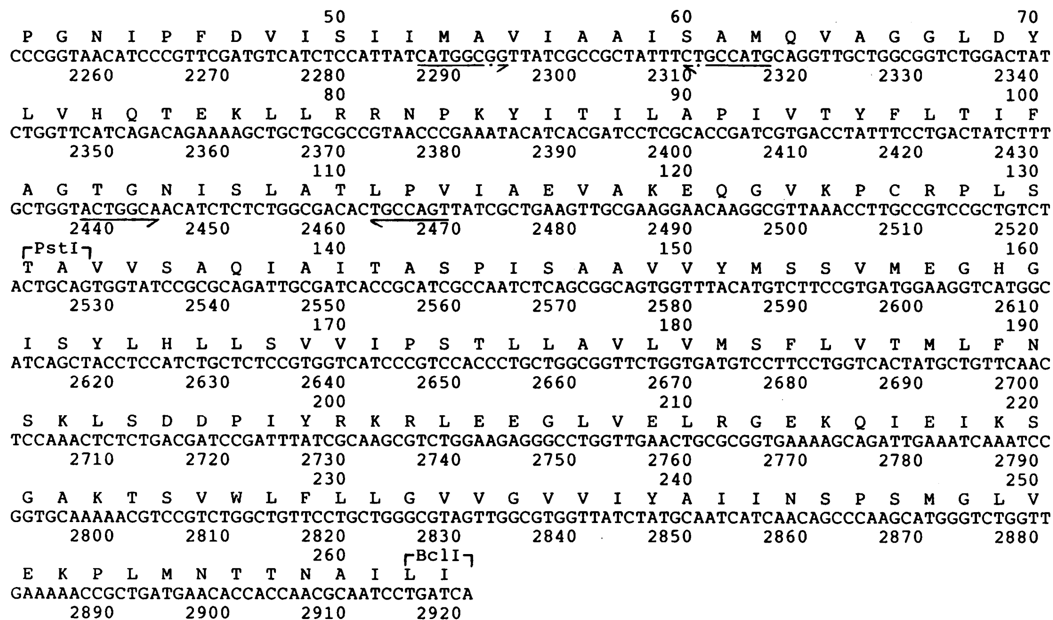


Fig. 4. Nucleotide sequence of the *aspA* gene and primary structure of aspartase

The nucleotide sequence of 2921 base-pairs of the non-transcribed strand of the aspartase (*aspA*) structural genes and flanking regions is presented in the 5′-to-3′ direction. The nucleotide co-ordinates are assigned relative to the first base of the *Bcl*I (B_1) site upstream of *aspA*. The primary structures of the *aspA*-gene product and part of an unidentified downstream gene are shown above the nucleotide sequence. Other features are indicated as in Fig. 2.

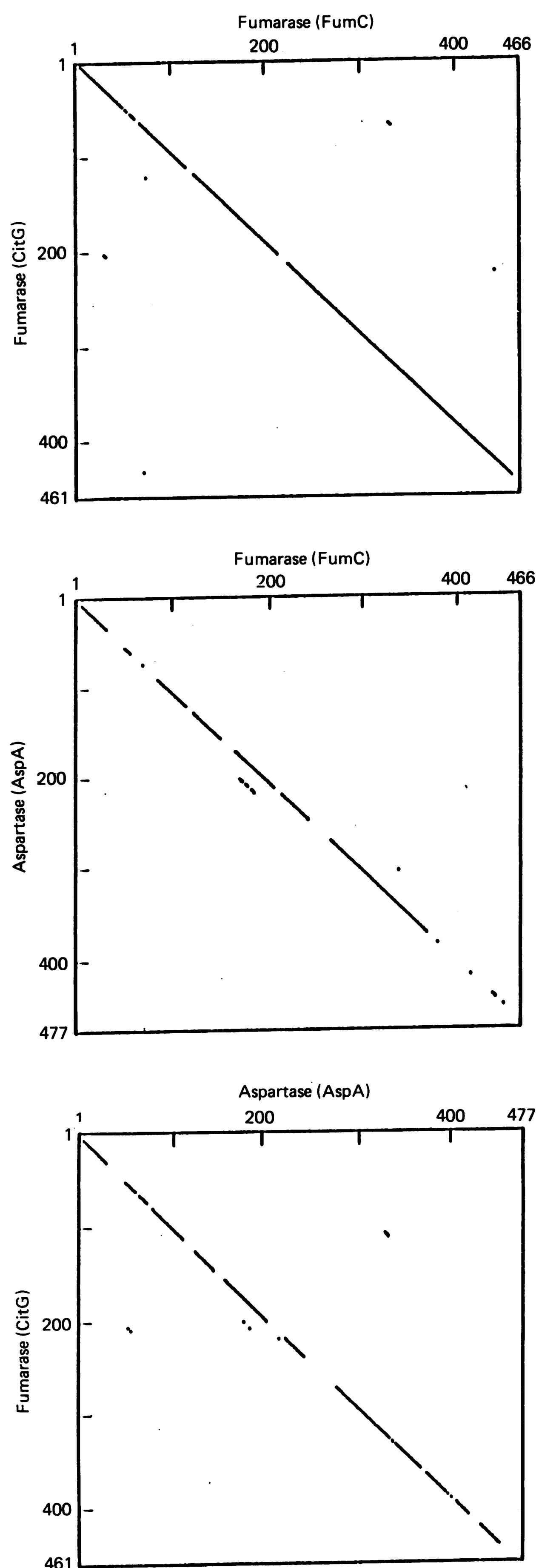
1983) as standards. The *fumC* coding region begins with a methionine coding at a position 2143 and extends for 1401 base-pairs to a stop codon at position 3544 (Fig. 2). It encodes a polypeptide of M_r 50353 (466 amino acid residues, excluding the initiating methionine), which is in good agreement with the value of 48000 estimated by SDS/polyacrylamide-gel electrophoresis (Guest & Roberts, 1983). The amino acid composition of the *fumC*-gene product (FumC) is listed in Table 1. It clearly resembles that predicted for the *B. subtilis* fumarase (CitG) and to a lesser extent those of two mammalian fumarases. The overall composition of the FumC protein differs from that of the FumA protein reported by Miles & Guest (1984*a*). Another coding region of unknown identity is predicted by FRAMESCAN. It has the opposite polarity to *fumC* and extends for at least 227 codons from a point outside the sequenced region up to a stop codon at position 3547 (Fig. 2).

The *fumC* coding region is preceded by a correctly placed ribosome-binding site (Shine–Dalgarno sequence; Gold *et al.*, 1981) and the proposed translational initiation site scores well with the PERCEPTRON algorithm of Stormo *et al.* (1982). The *fumC* codon usage is shown in Table 2. The proportion of modulatory codons (3.9%) and optimal energy codons in a diagnostic set of eight pairs (46%) suggests that *fumC* is only moderately expressed (Grosjean & Fiers, 1982).

The *fumA*–*fumC* intergenic region contains several putative promoter sequences, and three possibilities, which closely match the consensus sequence given by Hawley & McClure (1983) when analysed by the method of Spencer & Guest (1985), are shown in Fig. 2. There are several regions of hyphenated dyad symmetry that could form stable stem-and-loop structures in RNA transcripts. Many occur in presumptive coding regions, and only the most significant are indicated in Fig. 2. The strongest potential hairpin [$\Delta G = -110.9$ kJ/mol (-26.0 kcal/mol); Tinoco *et al.*, 1973] is located just downstream of the *fumC* structural gene (position 3558–3599), where it could function as a transcriptional terminator (Fig. 2). It has the characteristic G+C-rich sequence of *rho*-independent terminators but lacks the typical run of T (U) nucleotides (Rosenberg & Court, 1979). The hairpin at position 2006–2026 has been tentatively identified as the transcriptional terminator for the *fumA* gene (Miles & Guest, 1984*a*), and another in the *fumC* coding region (position 3434–3462) could be the terminator for the unidentified coding region that converges on *fumC* (Fig. 2). These observations suggest that *fumC* can function as an independent transcriptional unit. No sequences resembling CRP-binding sites (Chapon & Kolb, 1983) were detected. However, the promoter region contains sequences resembling those found upstream of the fumarate reductase operon (*frdABCD*) and the *fnr* gene encoding the anaerobic gene activator Fnr (Cole *et al.*, 1982; Shaw & Guest, 1982). The conserved residues are indicated in Fig. 2 and their presence may signify that *fumC* expression is controlled by Fnr.

Nucleotide sequence and other features of the *aspA* region

The *aspA* gene is located in a 2.9 kb *Bcl*I fragment of bacterial DNA that has been cloned into the *Bam*HI site of pBR322 to generate the $AspA^+$ hybrid plasmid pGS94 (Guest *et al.*, 1984). The sequencing strategy for the *aspA* gene was based primarily on a combination of ‘shot-gun’ and directed cloning of subfragments of the 3.4 kb hybrid fragment released from plasmid pGS94 with *Hin*dIII and *Sph*I (see the Experimental section). The M13 clones used in the sequencing analysis are summarized in Fig. 3, and the complete nucleotide sequence of the 2921-base-pair



segment of bacterial DNA (B_1–B_2) is presented in Fig. 4. The sequence is fully overlapped, and 99.4% has been sequenced on both DNA strands.

The sequence was examined for coding regions and other features as described for the *fumC* region. The coding region extending from an initiation codon at position 575 to a stop codon at position 2009 corresponds to the *aspA* structural gene. It encodes a polypeptide of M_r 52190 (477 amino acid residues, excluding the initiating methionine), which is in reasonable agreement with the value of 55000 obtained previously in maxicell studies (Guest *et al.*, 1984). The amino acid composition of the AspA protein (aspartase) resembles those of the bacterial fumarases except for having notably higher tyrosine and valine and lower histidine and tryptophan contents (Table 1). The *aspA* codon usage (Table 2) contains only a small proportion of modulatory codons (0.6%), and there is a preference for codons with optimum codon–anticodon interaction energies in the diagnostic set (72%), suggesting that the *aspA* gene is strongly expressed.

The *aspA* coding region is preceded by a putative ribosome-binding site, and the corresponding translational initiation site is the only one confirmed when the PERCEPTRON algorithm of Stormo *et al.* (1982) is applied. A search for potential promoter sequences that match a consensus sequence (Spencer & Guest, 1985) revealed two likely possibilities some 165 and 48 base-pairs upstream of the initiation codon (C and D; Fig. 4). The *aspA* coding region is followed by a striking potential stem-and-loop structure [$\Delta G = -106.7$ kJ/mol (-25.4 kcal/mol); position 2032–2063] with all the features of a *rho*-independent transcriptional terminator (Fig. 4).

The sequence contains parts of two additional coding regions. One having the opposite polarity to the aspartase gene starts at position 109 and extends for at least 36 codons before reaching the unsequenced region (Fig. 4). It is preceded by a putative ribosome-binding site, and the surrounding sequence strongly resembles a translational initiation site. There are also two potential promoter sequences (A and B) in the upstream region, separated from the proposed *aspA* promoters by 47 base-pairs (Fig. 4). The other coding region extends for at least 264 codons from position 2129 (Fig. 4). It is preceded by two potential ribosome-binding sites and two promoter sequences (E and F), the former being located in the distal segment of the *aspA* gene (Fig. 4). No CRP-binding sites (Chapon & Kolb, 1983) were detected in the intergenic regions, but sequences resembling those found upstream of the *fumC*, *frd* and *fnr* genes occur in the promoter region of the second unidentified gene (Fig. 4).

During the course of the present work the *aspA* genes

Fig. 5. Comparison matrices for the amino acid sequences of the *fumC*-, *citG*- and *aspA*-gene products

The matrices depict pairwise comparisons for the fumarases of *E. coli* K12 (FumC), *B. subtilis* 168 (CitG) and the aspartase of *E. coli* K12 (AspA). The proportional option of the DIAGON program (Staden, 1982) was used and the points correspond to the midpoints of 21-residue spans giving a double matching probability of $\geqslant 0.0001$ (McLachlan, 1971).

Fig. 6. Alignments of amino acid sequences for the FumC, CitG and AspA proteins

The sequences have been aligned for maximum homology based on the DIAGON comparisons (Fig. 5). Identical residues shared by two or more of the sequences are boxed. The asterisks mark the methionine, histidine and cysteine residues that are conserved in all three sequences; the arrows denote the reactive cysteine residues in AspA. The polypeptides are numbered from the residue immediately following the *N*-terminal methionine.

of *E. coli* W and *Serratia marcescens* were cloned and the nucleotide sequence of a 2901-base-pair *Ava*I fragment containing the *E. coli* W gene was defined (Takagi *et al.*, 1985; Takagi & Kisumi, 1985). The *E. coli* W sequence overlaps that of *E. coli* K12 by 2458 base-pairs starting at the *Ava*I site (position 464 in Fig. 4), and, although it lacks some of the upstream intergenic region (including a potential *aspA* promoter), it contains all of the downstream gene encoding an unidentified polypeptide of M_r 38521. The segments shared by *E. coli* K12 and *E. coli* W are remarkably similar, there being only nine nucleotide substitutions. Six of these are in the *aspA* structural gene at positions 669 (T→A), 949 (A→G), 1051 (C→T), 1267 (T→C), 1483 (A→T) and 1687 (C→T), and they all generate synonymous codons except for the first, which changes the valine (GTA) residue at position 31 in *E. coli* K12 to glutamate (GAA) in *E. coli* W (Fig. 4). There are two substitutions in the unidentified gene downstream of *aspA* at positions 2278 (C→T) and 2680 (C→T). The remaining substitution (A→G at position 2035) reduces the stability of the hairpin in the putative *aspA* terminator of *E. coli* W.

Relationships between the *fumC*-, *citG*- and *aspA*-gene products

The computer program DIAGON (Staden, 1982) was used to compare the amino acid sequences of the *fumC*-, *citG*- and *aspA*-gene products (FumC, CitG and AspA respectively) in pairwise fashion (Fig. 5). The diagonal lines indicate a very high degree of conservation among all three sequences. Alignments based on the DIAGON comparisons are shown in Fig. 6. Very few insertions or deletions were needed to maximize the alignments. The amino acid sequences of FumC and CitG have 62.8% identical residues, compared with FumC and AspA (37.5%) and CitG and AspA (37.8%). The homologues increase to 75.9%, 54.9% and 57.6% (respectively) when conservative substitutions scoring $\geqslant 0.12$ in the MDM_{78} mutation data matrix (Schwartz & Dayhoff, 1978) are included. The amino acid sequence homologies are reflected by comparable homologies at the nucleotide level in the coding regions: 59.2% (*fumC–citG*), 48.1% (*fumC–aspA*) and 48.9% (*citG–aspA*). The degree of nucleotide sequence homology diminishes outside the structural genes. It may also be noted that the *fumC* coding sequence to the left of the *Eco*RI site has 53.1% homology with the comparable 260-base-pair *citG* segment, whereas to the right of this site there is 60.2% homology over 1120 base-pairs. These differences in degree and extent of homology probably account for the observed hybridization of a *citG* probe to the right of the *Eco*RI site and not to the left (Guest *et al.*, 1985). Computer predictions of secondary structure and hydropathy profile analyses of all three proteins (results not shown) strengthen the view that the remarkable sequence homologies reflect a high degree of structural conservation at the tertiary level.

Active-site residues and catalytic mechanism

Chemical modification of pig heart fumarase by iodoacetate has suggested that there is a histidine residue at the active site (Bradshaw *et al.*, 1969). Two methionine residues are also carboxymethylated, but it is not clear whether these are at the active site. Analogous studies with the aspartase from *E. coli* W involving modification with diethyl pyrocarbonate and protection by substrates and substrate analogues have likewise suggested that one or two histidine residues are at the active site (Ida & Tokushige, 1985*a*). It is therefore of interest to note that, of all the histidine residues (15 in FumC, 11 in CitG and

eight in AspA), only one is conserved in all three proteins (Fig. 6). Similarly, only two methionine residues out of 17 in FumC, 17 in CitG and 15 in AspA are likewise conserved.

The thiol groups of pig heart fumarase (three cysteine residues per monomer) are not thought to be associated with the catalytic site (Robinson *et al.*, 1967), but studies with *E. coli* aspartase involving modification with *N*-ethylmaleimide have suggested that one or two thiol groups are essential for enzyme activity, and that thiol-group modification is restricted to the active site (Mizuta & Tokushige, 1975). More recently, Cys-140 and Cys-430 have been specifically labelled with *N*-(7-dimethylamino-4-methylcoumaryl)maleimide (Ida & Tokushige, 1985*b*). These cysteine residues are not conserved in either FumC or CitG, but are replaced by a serine and an alanine residue (respectively) in both fumarases (Fig. 6). Aspartase is activated by L-aspartate and is in consequence thought to have a second L-aspartate-binding site in addition to the catalytic site (Ida & Tokushige, 1985*a*). These authors postulate that the labelled cysteine residues may be associated with either the activation site or the catalytic site. Nevertheless, from the sequence comparisons, it is tempting to speculate that the Cys-140→Ser and Cys-430→Ala substitutions are important in determining the substrate specificity of the respective enzymes (aspartase→fumarase). It should be noted that the cysteine residue at position 386 in FumC is conserved in CitG and AspA, but no significance can so far be attached to it.

In view of the strong homologies surrounding the histidine and methionine residues that are conserved in all three enzymes, and the reactivities of these residue types, it is speculated that the critical residues may be located in their active sites. Indeed, the active sites and the reaction mechanisms of fumarase and aspartase may be similar. Support for this view comes from the tight binding to fumarase (pig heart) and aspartase (*Bacterium cadaveris*) of 3-carbanionic substrate analogues, which could mimic the substrate transition state if the reactions proceed via carbanionic intermediates (Porter & Bright, 1980). The chemical-modification data, the pH-dependence of the fumarase reaction and the use of isotope effects have led Blanchard & Cleland (1980) to propose a chemical mechanism involving a histidine residue and a carbanion-like substrate transition state, and they suggest that this mechanism could equally apply to enzymes such as aspartase. However, Jones *et al.* (1980) have studied the mechanism of pig heart fumarase similarly, using isotope effects, and conclude that the reaction may proceed in a concerted fashion via a transition state in which the breaking of the C–OH bond is much further advanced than that of the C–H bond; these mechanisms could both still involve an active-site histidine residue. It is not clear whether the conserved methionine (and cysteine) residues play an important part in the reaction mechanism, participate in substrate binding, or play an essential structural role.

Conclusion

It is clear from the sequence homologies that aspartase and the two fumarases are not only structurally related but may well use analogous chemical mechanisms for catalysing the corresponding reactions. The homologies also suggest that the enzymes are related in evolution. It is likely that both enzymic reactions were important in the metabolism of primitive anaerobes, and in *E. coli* both enzymes retain anaerobic as well as aerobic functions. The sequence analysis has not resolved questions concerning the functional specificities of the several fumarase genes (*fumA*, *fumB* and *fumC*). However, the presence of specific sequence homologies in the upstream regions of *fumC* and genes that are controlled by Fnr (an anaerobic gene activator protein) suggests that *fumC* may encode an anaerobic fumarase. Future studies aimed at over-expressing the enzymes, crystallization and structural determination should permit the elucidation of their structure–function relationships, particularly the molecular mechanisms by which they can distinguish water and ammonia in their respective reactions.

We thank G. D. Searle Ltd, and the New Brunswick Instrument Corp. for gifts of synthetic primers, and Ian K. Duckenfield for assistance with computer analyses. This work was supported by the Science and Engineering Research Council (J. R. G.) and by studentships from the Medical Research Council (J. S. M.) and G. D. Searle Ltd. (S. A. W.).

REFERENCES

Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 3963–3965

Blanchard, J. S. & Cleland, W. W. (1980) Biochemistry **19**, 4506–4513

Bradshaw, R. A., Robinson, G. W., Hass, G. M. & Hill, R. L. (1969) J. Biol. Chem. **244**, 1755–1763

Chapon, C. & Kolb, A. (1983) J. Bacteriol. **156**, 1135–1143

Cole, S. T., Grundstrom, T., Jaurin, B., Robinson, J. J. & Weiner, J. H. (1982) Eur. J. Biochem. **126**, 211–216

Courtright, J. B. & Henning, U. (1970) J. Bacteriol. **102**, 722–728

Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. S. & Stormo, G. (1981) Annu. Rev. Microbiol. **35**, 365–403

Gray, C. T., Wimpenny, J. W. T. & Mossman, M. R. (1966) Biochim. Biophys. Acta **117**, 33–41

Grosjean, H. & Fiers, W. (1982) Gene **18**, 199–209

Guest, J. R. & Roberts, R. E. (1983) J. Bacteriol. **153**, 588–596

Guest, J. R., Roberts, R. E. & Wilde, R. J. (1984) J. Gen. Microbiol. **130**, 1271–1278

Guest, J. R., Miles, J. S., Roberts, R. E. & Woods, S. A. (1985) J. Gen. Microbiol. **131**, 2971–2984

Harden, A. (1901) J. Chem. Soc. **79**, 610–628

Hawley, D. K. & McClure, W. R. (1983) Nucleic Acids Res. **11**, 2237–2255

Ida, N. & Tokushige, M. (1985*a*) J. Biochem. (Tokyo) **98**, 35–39

Ida, N. & Tokushige, M. (1985*b*) J. Biochem. (Tokyo) **98**, 793–797

Jones, V. T., Lowe, G. & Potter, B. V. L. (1980) Eur. J. Biochem. **108**, 433–437

Kanarek, L., Marler, E., Bradshaw, R. A., Fellows, R. E. & Hill, R. L. (1964) J. Biol. Chem. **239**, 4207–4211

Kobayashi, K. & Tuboi, S. (1983) J. Biochem. (Tokyo) **94**, 707–713

Kobayashi, K., Yamanishi, T. & Tuboi, S. (1981) J. Biochem. (Tokyo) **89**, 1923–1931

Marcus, M. & Halpern, Y. S. (1969) Biochim. Biophys. Acta **177**, 314–320

McLachlan, A. D. (1971) J. Mol. Biol. **61**, 409–424

Messing, J. & Vieira, J. (1982) Gene **19**, 269–276

Miles, J. S. & Guest, J. R. (1984*a*) Nucleic Acids Res. **12**, 3631–3642

Miles, J. S. & Guest, J. R. (1984*b*) Gene **32**, 41–48

Miles, J. S. & Guest, J. R. (1985) Nucleic Acids Res. **13**, 131–140

Mizuta, K. & Tokushige, M. (1975) Biochim. Biophys. Acta **403**, 221–231

Moir, A., Feavers, I. M. & Guest, J. R. (1984) J. Gen. Microbiol. **130**, 3009–3017
Norrander, J., Kempe, T. & Messing, J. (1983) Gene **26**, 101–106
Porter, D. J. T. & Bright, H. J. (1980) J. Biol. Chem. **255**, 4772–4780
Quastel, J. H. & Woolf, B. (1926) Biochem. J. **20**, 545–555
Robinson, G. W., Bradshaw, R. A., Kanarek, L. & Hill, R. L. (1967) J. Biol. Chem. **242**, 2709–2718
Rosenberg, M. & Court, D. (1979) Annu. Rev. Genet. **13**, 319–353
Rudolph, F. D. & Fromm, H. J. (1971) Arch. Biochem. Biophys. **147**, 92–98
Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) J. Mol. Biol. **143**, 161–178
Schwartz, R. M. & Dayhoff, M. O. (1978) in Atlas of Protein Sequence and Structure (Dayhoff, M. O. ed.), vol. 5, supplement 3, pp. 353–358, National Biomedical Research Foundation, Washington
Shaw, D. J. & Guest, J. R. (1982) Nucleic Acids Res. **10**, 6119–6130
Smith, M. W. & Neidhardt, F. C. (1983) J. Bacteriol. **154**, 336–343
Spencer, M. E. & Guest, J. R. (1985) Mol. Gen. Genet. **200**, 145–154
Spencer, M. E., Lebeter, V. M. & Guest, J. R. (1976) J. Gen. Microbiol. **97**, 73–82
Staden, R. (1977) Nucleic Acids Res. **4**, 4037–4051
Staden, R. (1980) Nucleic Acids Res. **8**, 3673–3694
Staden, R. (1982) Nucleic Acids Res. **10**, 2951–2961
Staden, R. (1984) Nucleic Acids Res. **12**, 499–504
Staden, R. & McLachlan, A. D. (1982) Nucleic Acids Res. **10**, 141–156
Stephens, P. E., Lewis, H. M., Darlison, M. G. & Guest, J. R. (1983) Eur. J. Biochem. **135**, 519–527
Stormo, G. D., Schneider, T. D., Gold, L. M. & Ehrenfeucht, A. (1982) Nucleic Acids Res. **10**, 2997–3011
Suzuki, S., Yamaguchi, J. & Tokushige, M. (1973) Biochim. Biophys. Acta **321**, 369–381
Takagi, J. S., Fukunaga, R., Tokushige, M. & Katzuki, H. (1984) J. Biochem. (Tokyo) **96**, 545–542
Takagi, J. S., Ida, N., Tokushige, M., Sakamoto, H. & Shimura, Y. (1985) Nucleic Acids Res. **13**, 2063–2074
Takagi, T. & Kisumi, M. (1985) J. Bacteriol. **161**, 1–6
Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O., Crothers, D. M. & Gralla, J. (1973) Nature (London) New Biol. **246**, 40–41

Received 30 December 1985/10 February 1986; accepted 1 April 1986