JOURNAL OF BACTERIOLOGY, Aug. 1990, p. 4631–4640
0021-9193/90/084631-10$02.00/0

Vol. 172, No. 8

Nucleotide Sequence and Analysis of the *speA* Gene Encoding Biosynthetic Arginine Decarboxylase in *Escherichia coli*

ROBERT C. MOORE AND STEPHEN M. BOYLE*

Department of Pathobiology, Virginia-Maryland Regional College of Veterinary Medicine, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

Received 6 February 1990/Accepted 14 May 1990

The DNA sequence of a 3.23-kilobase fragment of the *Escherichia coli* chromosome encoding biosynthetic arginine decarboxylase (ADC) was determined. This sequence contained the *speA* open reading frame (ORF) as well as partial *speB* and *metK* ORFs. The ADC ORF is 1,974 nucleotides long; the deduced polypeptide contains 658 amino acids with a molecular size of 73,980 daltons. The molecular weight and predicted ADC amino acid composition are nearly identical to the amino acid analysis of purified ADC performed by Wu and Morris (J. Biol. Chem. 248:1687–1695, 1973). A translational *speA-lacZ* fusion, pRM65, including 1,389 base pairs (463 amino acids) of the 5′ end of *speA* was constructed. Western blots (immunoblots) with β-galactosidase antisera revealed two ADC::β-galactosidase fusion proteins in *E. coli* bearing pRM65: 160,000 and 156,000 daltons representing precursor and mature hybrid proteins, respectively. The predicted amino acid sequence of ADC contains a region of six amino acid residues found in two bacterial diaminopimelic acid decarboxylases and three eucaryotic ornithine decarboxylases. This conserved sequence is located approximately eight amino acids from the putative pyridoxal phosphate-binding site of ADC and is predicted to be involved in substrate binding.

Putrescine (1,4-diaminobutane), a diamine required for optimal growth, is produced in *Escherichia coli* and many plants by either of two pathways related to arginine biosynthesis (10, 35). In *E. coli*, biosynthetic arginine decarboxylase (ADC), encoded by the *speA* gene, acts on arginine to produce agmatine, which is converted into putrescine by agmatine ureohydrolase (AUH). A second pathway uses ornithine decarboxylase (ODC) to produce putrescine directly from ornithine. While ODC is common to all cells, ADC appears to be unique to bacteria and plants. These two parallel pathways remain distinct in *E. coli* as the species lacks arginase and is therefore unable to convert arginine to ornithine. In the absence of exogenous arginine, putrescine is produced primarily from the ODC pathway, but when arginine is added to the growth medium, ornithine levels decline due to inhibition of arginine biosynthesis (37). The ADC pathway thereby ensures that putrescine is produced in the presence of exogenous arginine (25), e.g., in the intestinal tract.

ADC can be found in two forms: biodegradative and biosynthetic. Biodegradative ADC is induced in some strains of *E. coli* when they are grown in an acidic enriched medium containing arginine (9, 22). Biodegradative amino acid decarboxylases appear to play a role in regulating pH by consuming protons and thus neutralizing the acidic products of carbohydrate fermentation (9). Biosynthetic ADC is produced when *E. coli* is grown in minimal media at neutral pH; approximately 0.07% of the protein found in *E. coli* UW44 crude extracts is biosynthetic ADC (43). Native biosynthetic ADC is a tetramer of 280,000 daltons which requires both Mg^{2+} and pyridoxal phosphate as cofactors (23, 43). Monomers of ADC are synthesized as 74,000-dalton precursor polypeptides which are posttranslationally processed to a 70,000-dalton mature form localized within the inner periplasmic space (4). The periplasmic location of ADC explains the finding that more putrescine is produced from exogenous arginine than from endogenous arginine, since arginine can be preferentially channeled into putrescine as it is being transported into the cell (38). The *speA* gene and ADC are subject to metabolic control. Putrescine and spermidine repress the *speA* gene and feedback inhibit ADC (37). Cyclic AMP (cAMP) also negatively regulates *speA* gene expression either directly or indirectly via binding to cAMP receptor protein (CRP) (42). The experiments described in this report determined the nucleotide sequence of the *speA* gene and compared the properties of the deduced ADC amino acid sequence with those of purified ADC (43) and other decarboxylases.

MATERIALS AND METHODS

Bacterial strains, plasmids, and media. The bacterial strains and plasmids used in this study are described in Table 1. *E. coli* strains were grown in Luria broth (LB) (19) (1% tryptone, 0.5% yeast extract, 1% NaCl), Terrific broth (1.4% tryptone, 2.7% yeast extract, 0.4% glycerol, 17 mM KH_2PO_4, 72 mM K_2HPO_4) (39), or MOPS (morpholinepropanesulfonic acid) medium (26) supplemented with 1 μg of thiamine per ml, 0.2% glucose, and 50 μg of each naturally occurring L-amino acid per ml except arginine. Ampicillin was used at 100 μg/ml.

Enzymes and reagents. Restriction endonucleases, Klenow fragment of DNA polymerase I, T4 DNA ligase, exonucleases III and VII, and halogenated indolyl-β-D-galactoside. (Bluo-Gal) were purchased from Bethesda Research Laboratories, Inc. (Gaithersburg, Md.). Sequenase was purchased from U.S. Biochemical Corp. (Cleveland, Ohio). Avian myeloblastosis virus reverse transcriptase was purchased from Boehringer Mannheim Biochemicals (Indianapolis, Ind.). [α-^{35}S]dATP and L-[1-^{14}C]arginine were from Dupont, NEN Research Products (Boston, Mass.). β-Galactosidase affinity column and antisera were purchased from Sigma Chemical Co. (St. Louis, Mo.). SeaPlaque low-melting-temperature agarose was from FMC BioProducts (Rockland, Maine).

* Corresponding author.

TABLE 1. *E. coli* strains and plasmids

Strain or plasmid	Genotype or description	Source or reference
E. coli		
DH5α	F⁻ φ80 Δ*lacZ*ΔM15 Δ(*lacZYA-argF*)*U169 recA1 endA1 hsdR17* (r_K^- m_K^+) *supE44 thi-1* lambda *gyrA96 relA1*	Bethesda Research Laboratories
CB806	F⁻ Δ*lacZ lacY*⁺ *galK rpsL thi recA56 phoA8*	31
Plasmids		
pKA5	pBR322 with an 8.0 kb *Eco*RI fragment	2
pGEM-3Z	High-copy-number sequencing plasmid	Promega Biotec
pRM15	3.2-kb *Bal*I-*Acc*I fragment in pGEM-3Z	This study
pRM59	3.2-kb *Acc*I-*Bal*I in reverse orientation in pGEM-3Z	This study
pMC1403	High-copy-number *lacZ* translational fusion vector	5
pRM65	2.2-kb *Bam*HI fragment in pMC1403	This study

Plasmid screening. Plasmids were isolated and purified by standard methods (19). The size of supercoiled plasmid DNA was determined by a phenol-chloroform quick screen method. Bacterial colonies were picked from a plate, suspended in 40 μl of 1× STE (20 mm Tris [pH 7.5], 100 mM NaCl, 10 mM EDTA), lysed by the addition of 0.5 volume of phenol-chloroform (1:1), and vortexed. RNase A (50 μg/ml, final concentration) was mixed with supernatants, incubated at 37°C for 15 min, and electrophoresed on a 0.8% agarose gel to separate chromosomal and plasmid DNA. Plasmid size was estimated by comparing the mobility of supercoiled forms of deletion plasmids with that of the vector (pGEM-3Z) and undeleted plasmid (pRM15).

Subcloning of *speA* gene. The *speA* gene is located on a 8.0-kilobase (kb) *Eco*RI fragment on plasmid pKA5 (2). Plasmid pKA5 was cleaved with *Bal*I and *Acc*I to produce a 3,236-base-pair (bp) fragment (Fig. 1). Restriction fragments were blunt ended by treatment with the Klenow fragment of DNA polymerase and separated by electrophoresis on a low-melting-temperature SeaPlaque 0.7% agarose gel (33). The agarose containing the 3,236-bp *Bal*I-*Acc*I fragment was excised and melted at 60°C, and the 3,236-bp fragment was ligated into the *Sma*I site of pGEM-3Z. Ligation products were transformed (12) into *E. coli* DH5α. Clones containing plasmids bearing the *speA* gene were selected by their ability to overexpress ADC (see Enzyme Assays, below).

To obtain deletions in the opposite orientation, we removed the 3,282-bp insert from pRM15 by digestion with restriction endonucleases *Hin*dIII and *Eco*RI. Restriction products were blunt ended by treatment with the Klenow fragment of DNA polymerase and separated by electrophoresis on a SeaPlaque 0.7% agarose gel. The blunt-ended 3,282-bp *Hin*dIII-*Eco*RI fragment was excised from the gel, ligated into the *Sma*I site of a new pGEM-3Z, and transformed into *E. coli* DH5α. Clones were assayed for their ability to overproduce ADC. Clone pRM59 expressed levels of ADC comparable to that produced by pRM15 and was selected for further study.

Exonuclease deletions of *speA*. Overlapping deletions of plasmid pRM15 were generated with exonuclease III (14). Plasmid pRM15 was cleaved with *Sal*I and *Sph*I, and digested with exonucleases III and VII, followed by treatment with the Klenow fragment of DNA polymerase to create blunt ends. Deleted DNA fragments were separated by electrophoresis in SeaPlaque 0.7% agarose; the DNA bands were excised, recircularized with T4 DNA ligase, and transformed into *E. coli* DH5α. Deletion plasmids were sized by

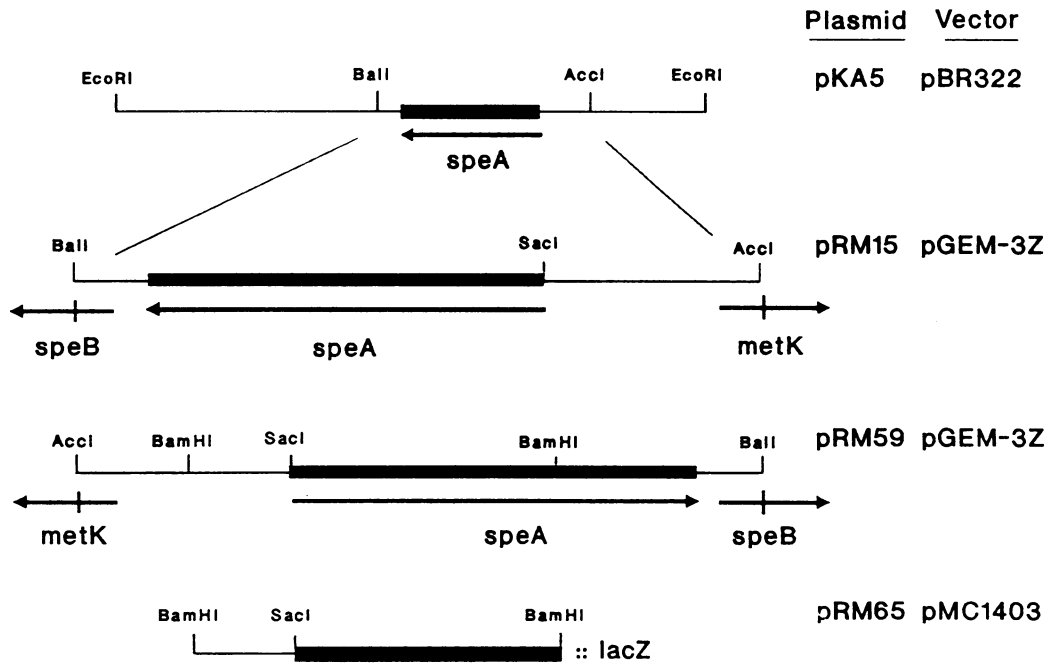


FIG. 1. Physical map of plasmids. Thick lines indicate *speA* (ADC) gene; arrows show direction of transcription; arrow with bar shows truncated *speB* (AUH) and *metK* (methionine adenosyltransferase) genes.

the phenol-chloroform quick screen method, and clones were screened for ADC enzyme activity to determine the approximate location of the *speA* gene within pRM15. Deletions of pRM59 were constructed as previously described except that the deletions were not separated on a SeaPlaque gel after treatment with Klenow fragment. Plasmids were sized by the phenol-chloroform quick screen method, and clones were analyzed with an ADC enzyme assay. All plasmids were restriction mapped to confirm the extent of deletions.

Dideoxy sequencing. Overlapping deletions were sequenced by the dideoxy-chain termination method of Sanger et al. (30) with SP6 and T7 sequencing primers (Promega Biotec, Madison, Wis.). The DNA sequence was determined by using either avian myeloblastosis virus reverse transcriptase or Sequenase following the protocols supplied by the manufacturers, except sequencing primers were hybridized after an alkaline denaturation. Plasmid DNA for sequencing reactions with avian myeloblastosis virus reverse transcriptase was purified by banding on CsCl gradients. Plasmid templates sequenced with Sequenase were isolated by a modified alkaline lysis method (17) from overnight 3-ml cultures. Both strands of overlapping DNA fragments were sequenced.

Construction of a Φ(*speA-lacZ'*) fusion. A translational Φ(*speA-lacZ'*) fusion was constructed by digesting pRM15 with *Bam*HI. The restriction fragments were separated by electrophoresis on a SeaPlaque 0.7% agarose gel. A 2,119-bp *Bam*HI fragment was excised, ligated into the *Bam*HI site of pMC1403, and transformed into *E. coli* DH5α. *E. coli* strains were screened for β-galactosidase production on LB plates containing 100 μg of ampicillin per ml and spread with 100 μl of a 20-mg/ml solution of Bluo-Gal. Plasmids containing Φ(*speA-lacZ'*) fusions were isolated and mapped with *Pst*I and *Sac*I to determine the orientation of the 2,119-bp insert within pMC1403. The orientation of the *Bam*HI insert was confirmed by sequencing the *lacZ* junction. Plasmid pRM65 was chosen for further study.

Western blots (immunoblots). *E. coli* transformed with pRM65 was grown in a shaking water bath at 37°C overnight in 5 ml of LB medium. Cells were recovered by centrifugation and lysed by sonication. ADC::β-galactosidase fusion proteins were separated with a Protosorb *lacZ* immunoaffinity column. Approximately 50 μg of protein was loaded into each well of a 3% stacking gel with a 5% polyacrylamide gel. Samples were loaded a second time in the reverse order to create a mirror image for an immunoblot. After electrophoresis, the gel was cut in two; one half was stained with Coomassie blue and the second half was blotted to a nitrocellulose membrane. The membrane was probed with β-galactosidase antisera to identify β-galactosidase and ADC::β-galactosidase hybrid proteins. The size of the fusion proteins was estimated by comparison with the mobility of molecular weight standards.

Protein determinations. Protein concentrations were determined by the Bradford method (3) (Bio-Rad Laboratories).

Enzyme assays. *E. coli* strains were grown in a shaking water bath at 37°C to a density of 70 Klett units either in LB medium or in MOPS medium supplemented with thiamine (1 μg/ml), glucose (0.2%), and all amino acids (50 μg/ml) except arginine. β-Galactosidase activity was assayed by the method of Miller (21). The lactose operon of *E. coli* strains was induced by the addition of isopropyl-β-D-thiogalactopyranoside (5 mM, final concentration). ADC activity was assayed as previously described (42).

Computer analysis. The DNA sequence was analyzed with the sequence analysis software from ICR (Fox Chase Cancer Institute) and the University of Wisconsin Genetic Computer Group sequence analysis software package installed on a VAX8800. The nucleic acid and predicted amino acid sequences of ADC were used to search for similar proteins in the GenEMBL and NBRF databases with the Wordsearch program.

Nucleotide sequence accession number. The nucleotide sequence data reported in this paper have been submitted to GenBank (accession no. M31770).

RESULTS

Subcloning of *speA* gene. A plasmid (pLC2-5) that overproduces ADC was identified from the Clarke-Carbon *E. coli* genomic library (6). An 8,000-bp *Eco*RI fragment was ligated into the *Eco*RI site of pBR322, yielding the hybrid plasmid pKA5 (36). This 8.0-kb *Eco*RI fragment carries the *speA*, *speB*, and *metK* genes, which overexpress ADC, AUH, and methionine adenosyltransferase, respectively (2). Deletion of a *Bal*I fragment from pKA5 was shown to effect AUH but not ADC or methionine adenosyltransferase activity. A computer search of the *metK* gene sequence (20) revealed an *Acc*I site approximately 3.2 kb downstream from the *Bal*I site in pKA5, within the *metK* open reading frame (ORF) (Fig. 1). A 3,236-bp *Bal*I-*Acc*I fragment from pKA5 was subcloned into pGEM-3Z, and when clones containing recombinant plasmids were assayed for ADC activity, 8 of 15 clones overproduced ADC. A single clone, pRM15, was chosen for further use. Crude extracts of *E. coli* DH5α cells containing pRM15 produced a 4.3-fold increase in ADC specific activity compared with *E. coli* and 1.5 times more activity than *E. coli* bearing pKA5. Crude extracts of *E. coli* DH5α transformed with plasmid pRM59 produced a 6.2-fold increase in ADC specific activity compared with *E. coli* and 2.8 times more than *E. coli* bearing pKA5. The apparent increase in the level of ADC expression of *E. coli* containing either pRM15 or pRM59 compared with cells containing pKA5 may be due to the loss of a regulatory element present on pKA5 or may simply be a reflection of the difference in the plasmid copy number. A slight increase in ADC activity of cells bearing pRM59 is probably due to a transcriptional fusion between the *speA* gene and *lacZ* promoter in the vector pGEM-3Z.

Sequence analysis. The sequence of the 3,236-bp *Bal*I-*Acc*I fragment carried by pRM15 was determined for both strands from overlapping DNA fragments (Fig. 2). A large ORF (ORF1) contains 1,974 nucleotides, begins with an AUG initiation codon (Met) at nucleotide 987, and ends with a GAG codon (Glu) at nucleotide 2960. ORF1 is capable of encoding a 73,980-dalton polypeptide consisting of 658 amino acids with a predicted isoelectric point of 4.70.

Sequences resembling the *E. coli* consensus for −35 (TTCACA) and −10 (AATAAT) promoter regions are located at nucleotides 811 to 816 and 839 to 844, respectively. A possible CRP-binding consensus sequence (TGTGC) was found at nucleotides 831 to 835; a ribosome-binding site, GAGG, was present at nucleotides 977 to 980; and a potential transcription terminator was located at nucleotides 3030 to 3067. A probable pyridoxal phosphate-binding site, H-K-L, is located in ADC from residues 298 to 300.

Partial ORFs corresponding to the *speB* and *metK* genes were also identified (Fig. 1). The *speA* and *speB* (ORF2) ORFs are found on the same DNA strand. ORF2 begins 140 nucleotides downstream from ORF1, continues for 135 nucleotides (45 amino acids), and terminates at a *Bal*I site

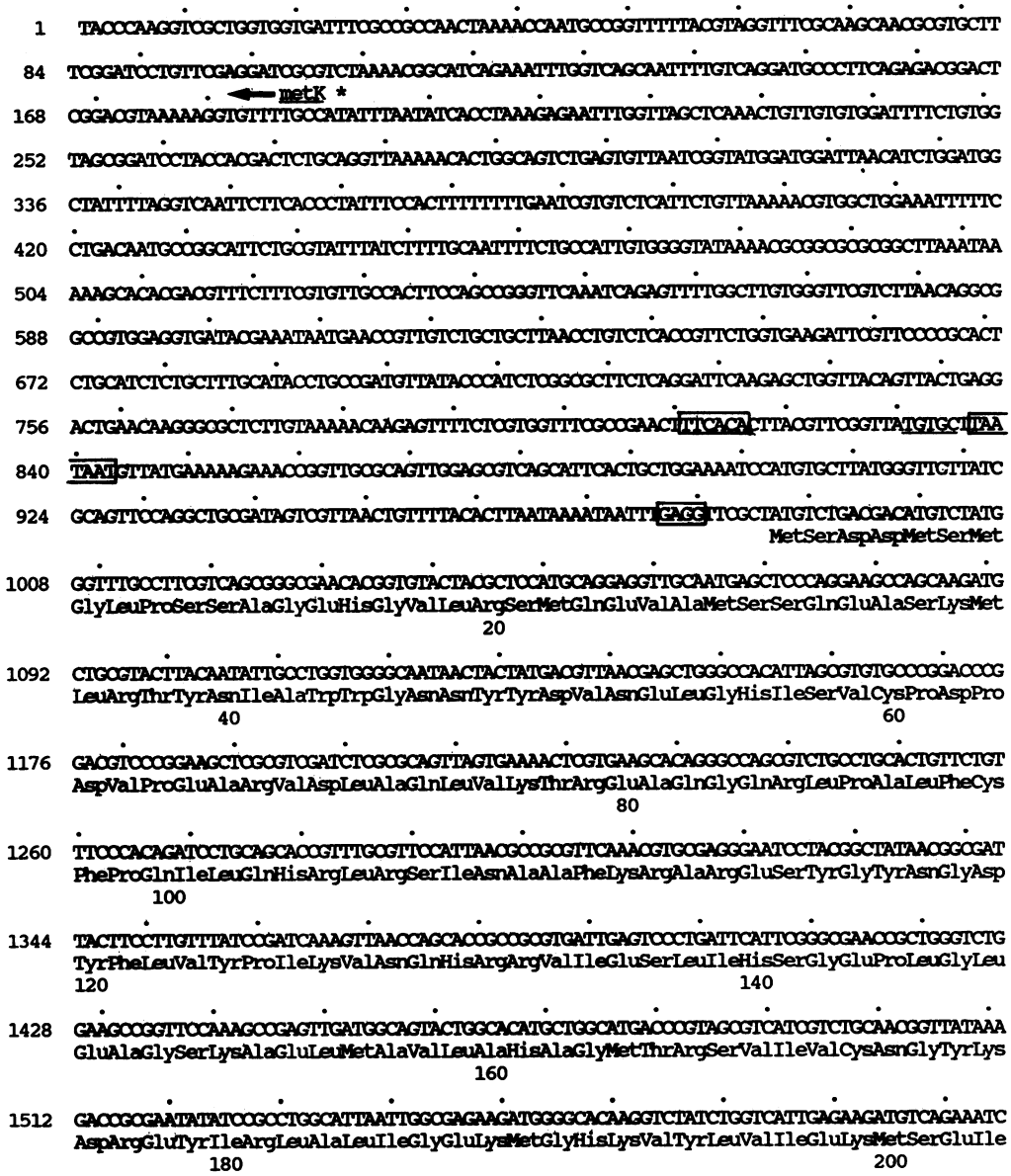
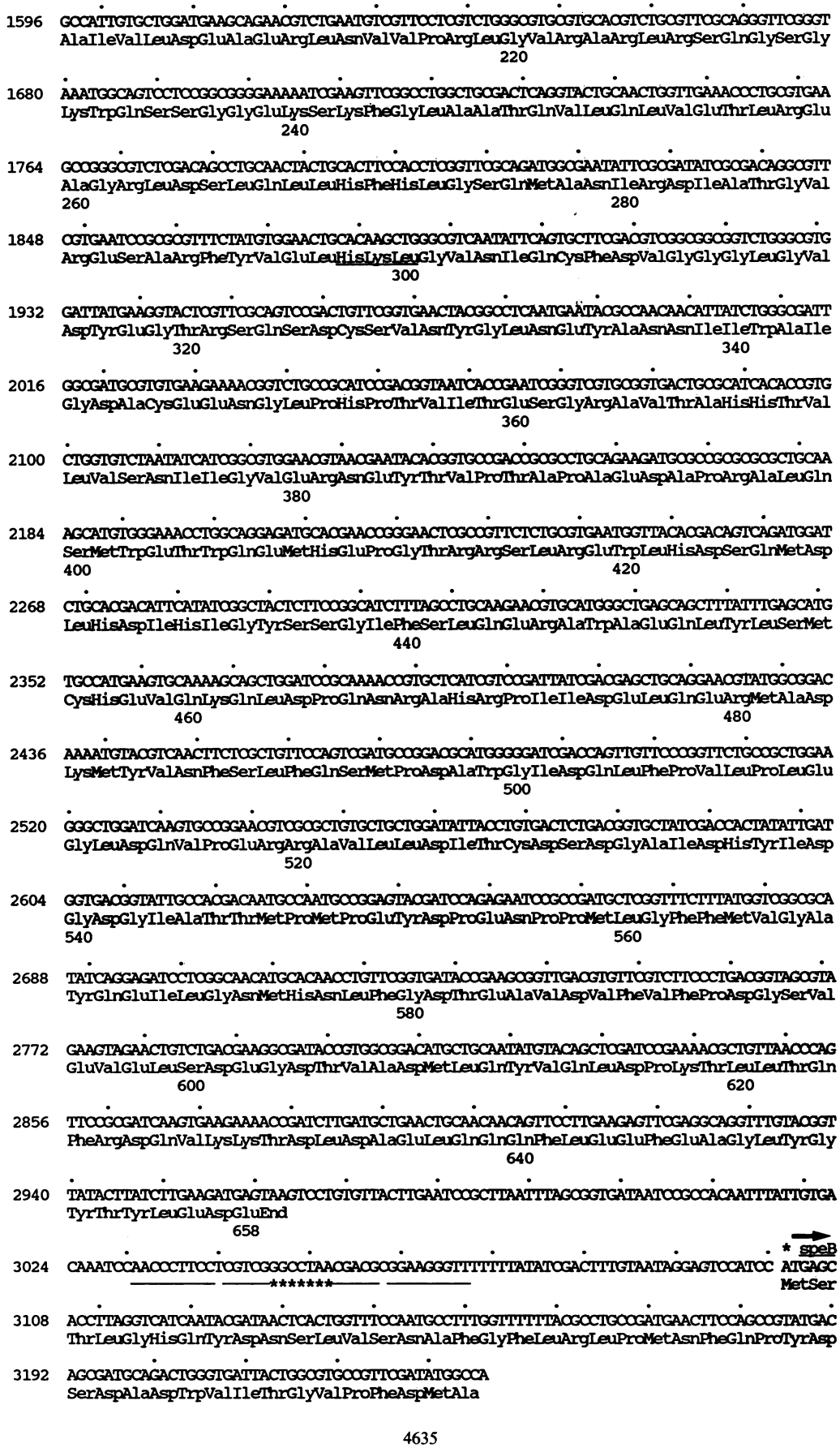


FIG. 2. Nucleotide and deduced amino acid sequences of *speA*. Promoter and ribosome-binding consensus sequences (boxed), CRP-binding site and pyridoxal phosphate-binding consensus (underlined), translation initiation sites of *speB* and *metK* (*), and rho-independent transcriptional terminator (———***———) are indicated.

within the AUH (*speB*) ORF. A 37-bp stem-loop structure, located at nucleotides 3031 to 3067, is present in the 140-bp intergenic region separating the *speA* and *speB* genes. The existence of this stem-loop structure was confirmed by S1 nuclease analysis (34). The *metK* ORF (ORF3), divergent from *speA*, begins 794 nucleotides upstream from ORF1, continues for 192 nucleotides (64 amino acids), and terminates at an *Acc*I site within the methionine adenosyltransferase (*metK*) ORF (20).

Identification of *speA* gene and promoter. The location of the *speA* gene was determined by assessing the ability of clones bearing deletions in pRM15 or pRM59 to overexpress ADC (Fig. 3). ADC expression was maintained when 564 nucleotides (424 bases upstream of ORF1) of the 5′ end of the *Ball*-*Acc*I fragment were removed. This treatment results in the loss of the *metK* ORF and a large portion of the intergenic region between the *speA* and *metK* genes. The ability to express ADC is lost when 355 nucleotides of the 3′ end of the *Ball*-*Acc*I fragment are removed. In this case, the *speB* ORF and 80 bp of ORF1, which encodes ADC, have been removed.

The *speA* promoter and 1,389 nucleotides (463 amino acids) of ORF1 were ligated to the *lacZ* gene in the vector pMC1403 to produce a Φ(*speA-lacZ′*) translation fusion, pRM65. *E. coli* CB806 bearing pRM65 produced a 42-fold increase in β-galactosidase activity relative to *E. coli* CB806 containing pMC1403 (1,250 U in pRM65 and 30 U in pMC1403). Western blots probed with β-galactosidase antisera identified three proteins in *E. coli* DH5α containing pRM65: 160,000- and 156,000-dalton ADC::β-galactosidase fusion proteins and a 116,353-dalton β-galactosidase monomer (8) (Fig. 4). The 160,000-dalton protein results from the fusion of 44,000 daltons of ADC to the 116,353 daltons of β-galactosidase. The presence of an additional 156,000-


1596 GCCATTGTGCTGGATGAAGCAGAACGTCTGAATGTCGTTCCTCGTCTGGGCGTGCGTGCACGTCTGCGTTCGCAGGGTTCGGGT
AlaIleValLeuAspGluAlaGluArgLeuAsnValValProArgLeuGlyValArgAlaArgLeuArgSerGlnGlySerGly
220
1680 AAATGGCAGTCCTCCGGCGGGGAAAAATCGAAGTTCGGCCTGGCTGCGACTCAGGTACTGCAACTGGTTGAAACCCTGCGTGAA
LysTrpGlnSerSerGlyGlyGluLysSerLysPheGlyLeuAlaAlaThrGlnValLeuGlnLeuValGluThrLeuArgGlu
240
1764 GCCGGGCGTCTCGACAGCCTGCAACTACTGCACTTCCACCTCGGTTCGCAGATGGCGAATATTCGCGATATGCGCGACAGGCGTT
AlaGlyArgLeuAspSerLeuGlnLeuLeuHisPheHisLeuGlySerGlnMetAlaAsnIleArgAspIleAlaThrGlyVal
260 280
1848 CGTGAATCCGCGCGTTTCTATGTGGAACTGCACAAGCTGGGCGTCAATATTCAGTGCTTCGACGTCGGCGGCGGTCTGGGCGTG
ArgGluSerAlaArgPheTyrValGluLeuHisLysLeuGlyValAsnIleGlnCysPheAspValGlyGlyGlyLeuGlyVal
300
1932 GATTATGAAGGTACTCGTTCGCAGTCCGACTGTTCGGTGAACTACGGCCTCAATGAATACGCCAACAACATTATCTGGGCGATT
AspTyrGluGlyThrArgSerGlnSerAspCysSerValAsnTyrGlyLeuAsnGluTyrAlaAsnAsnIleIleTrpAlaIle
320 340
2016 GGCGATGCGTGTGAAGAAAACGGTCTGCCGCATCCGACGGTAATCACCGAATCGGGTCGTGCGGTGACTGCGCATCACACCGTG
GlyAspAlaCysGluGluAsnGlyLeuProHisProThrValIleThrGluSerGlyArgAlaValThrAlaHisHisThrVal
360
2100 CTGGTGTCTAATATCATCGGCGTGGAACGTAACGAATACACGGTGCCGACCGCGCCTGCAGAAGATGCGCCGCGCGCGCTGCAA
LeuValSerAsnIleIleGlyValGluArgAsnGluTyrThrValProThrAlaProAlaGluAspAlaProArgAlaLeuGln
380
2184 AGCATGTGGGAAACCTGGCAGGAGATGCACGAACCGGGAACTCGCCGTTCTCTGCGTGAATGGTTACACGACAGTCAGATGGAT
SerMetTrpGluThrTrpGlnGluMetHisGluProGlyThrArgArgSerLeuArgGluTrpLeuHisAspSerGlnMetAsp
400 420
2268 CTGCACGACATTCATATCGGCTACTCTTCCGGCATCTTTAGCCTGCAAGAACGTGCATGGGCTGAGCAGCTTTATTTGAGCATG
LeuHisAspIleHisIleGlyTyrSerSerGlyIlePheSerLeuGlnGluArgAlaTrpAlaGluGlnLeuTyrLeuSerMet
440
2352 TGCCATGAAGTGCAAAAGCAGCTGGATCCGCAAAACCGTGCTCATCGTCCGATTATCGACGAGCTGCAGGAACGTATGGCGGAC
CysHisGluValGlnLysGlnLeuAspProGlnAsnArgAlaHisArgProIleIleAspGluLeuGlnGluArgMetAlaAsp
460 480
2436 AAAATGTACGTCAACTTCTCGCTGTTCCAGTCGATGCCGGACGCATGGGGGATCGACCAGTTGTTCCCGGTTCTGCCGCTGGAA
LysMetTyrValAsnPheSerLeuPheGlnSerMetProAspAlaTrpGlyIleAspGlnLeuPheProValLeuProLeuGlu
500
2520 GGGCTGGATCAAGTGCCGGAACGTCGCGCTGTGCTGCTGGATATTACCTGTGACTCTGACGGTGCTATCGACCACTATATTGAT
GlyLeuAspGlnValProGluArgArgAlaValLeuLeuAspIleThrCysAspSerAspGlyAlaIleAspHisTyrIleAsp
520
2604 GGTGACGGTATTGCCACGACAATGCCAATGCCGGAGTACGATCCAGAGAATCCGCCGATGCTCGGTTTCTTTATGGTCGGCGCA
GlyAspGlyIleAlaThrThrMetProMetProGluTyrAspProGluAsnProProMetLeuGlyPhePheMetValGlyAla
540 560
2688 TATCAGGAGATCCTCGGCAACATGCACAACCTGTTCGGTGATACCGAAGCGGTTGACGTGTTCGTCTTCCCTGACGGTAGCGTA
TyrGlnGluIleLeuGlyAsnMetHisAsnLeuPheGlyAspThrGluAlaValAspValPheValPheProAspGlySerVal
580
2772 GAAGTAGAACTGTCTGACGAAGGCGATACCGTGGCGGACATGCTGCAATATGTACAGCTCGATCCGAAAACGCTGTTAACCCAG
GluValGluLeuSerAspGluGlyAspThrValAlaAspMetLeuGlnTyrValGlnLeuAspProLysThrLeuLeuThrGln
600 620
2856 TTCCGCGATCAAGTGAAGAAAACCGATCTTGATGCTGAACTGCAACAACAGTTCCTTGAAGAGTTCGAGGCAGGTTTGTACGGT
PheArgAspGlnValLysLysThrAspLeuAspAlaGluLeuGlnGlnGlnPheLeuGluGluPheGluAlaGlyLeuTyrGly
640
2940 TATACTTATCTTGAAGATGAGTAAGTCCTGTGTTACTTGAATCCGCTTAATTTAGCGGTGATAATCCGCCACAATTTATTGTGA
TyrThrTyrLeuGluAspGluEnd
658
speB
3024 CAAATCCAACCCTTCCTCGTCGGGCCTAACGACGCGGAAGGGTTTTTTTATATCGACTTTGTAATAGGAGTCCATCC ATGAGC
MetSer
3108 ACCTTAGGTCATCAATACGATAACTCACTGGTTTCCAATGCCTTTGGTTTTTTACGCCTGCCGATGAACTTCCAGCCGTATGAC
ThrLeuGlyHisGlnTyrAspAsnSerLeuValSerAsnAlaPheGlyPheLeuArgLeuProMetAsnPheGlnProTyrAsp
3192 AGCGATGCAGACTGGGTGATTACTGGCGTGCCGTTCGATATGGCCA
SerAspAlaAspTrpValIleThrGlyValProPheAspMetAla

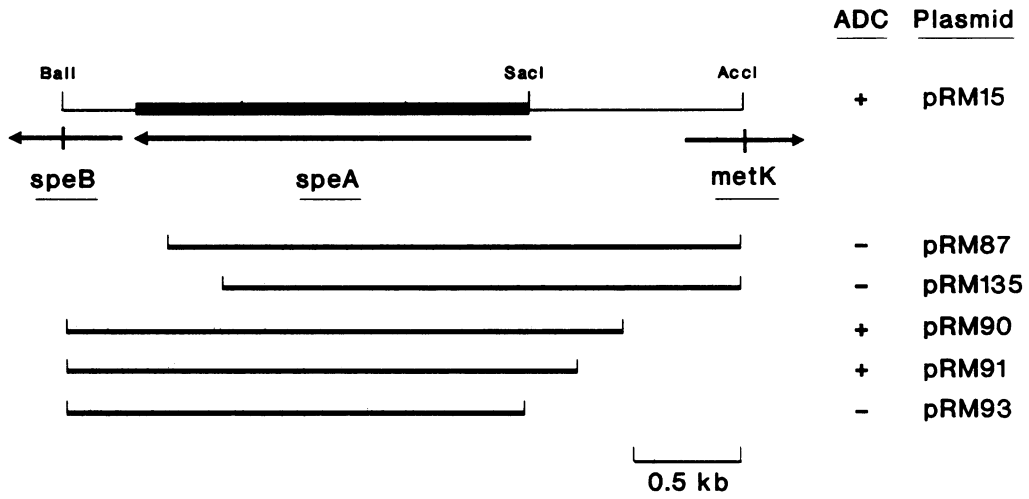


FIG. 3. Location of *speA* by deletion analysis. Region contained in deletion plasmids is indicated by thick lines. The *speB* and *metK* ORFs are truncated in pRM15. ADC activity and designation of plasmids are shown on the right.

dalton fusion protein containing 41,000 daltons of ADC represents the processed form of ADC fused to β-galactosidase; a 74,000-dalton ADC precursor is processed to the 70,000-dalton form (4). The size of these two fusion proteins further suggests that translation initiates at or near the beginning of ORF1. The difference of 4,000 daltons in molecular mass of these two fusion proteins also suggests that the 160,000-dalton form contains a signal peptide that is removed to produce the 156,000-dalton form (4).

The predicted molecular mass of an ADC monomer, 73,980 daltons, is almost identical to the 74,000 daltons estimated from denaturing gels (4, 43). The deduced amino acid composition of ORF1 is also in good agreement with the amino acid analysis of ADC performed by Wu and Morris (43) (Table 2). The results of deletion analysis, immunoblots of ADC::β-galactosidase fusion proteins, amino acid composition, and molecular weight of a protein predicted by ORF1 all indicate that ORF1 encodes ADC.

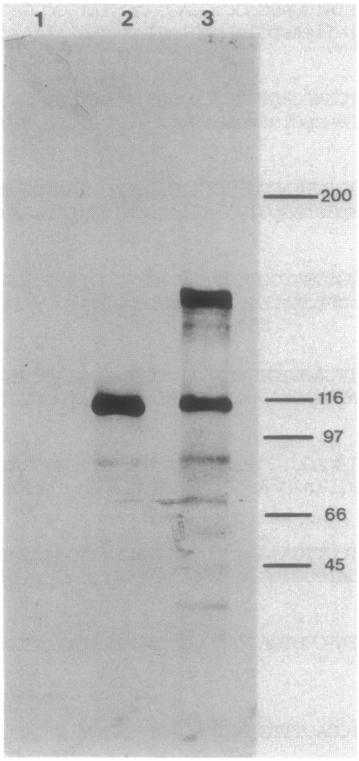


FIG. 4. Identification of ADC::β-galactosidase fusion proteins. Immunoblots were done as described in Materials and Methods. Lane 1, Protosorb column flowthrough; lane 2, β-galactosidase; lane 3, β-galactosidase and ADC::β-galactosidase fusion proteins. Sizes (in kilodaltons) of molecular mass standards are indicated on the right.

TABLE 2. Amino acid composition of *E. coli* ADC

Amino acid	Amino acid analysis[a]		DNA sequence	
	No.	%	No.	%
Ala	52	7.6	48	7.3
Arg	37	5.4	38	5.8
Asx	72	10.5	66	10.0
Asn			23	3.5
Asp			43	6.5
Cys	7	1.0	8	1.2
Glx	99	14.4	90	13.7
Gln			36	5.5
Glu			54	8.2
Gly	53	7.7	51	7.8
His	17	2.5	21	3.2
Ile	28	4.1	33	5.0
Leu	68	9.9	67	10.2
Lys	19	2.7	18	2.7
Met	25	3.6	24	3.7
Phe	21	3.1	20	3.0
Pro	33	4.8	28	4.3
Ser	43	6.3	41	6.2
Thr	27	3.9	24	3.7
Trp	12	1.7	9	1.4
Tyr	27	3.9	25	3.8
Val	47	6.8	47	7.1
Total	687		658	
Molecular mass (daltons)	74,000		73,980	

[a] From Wu and Morris (41).

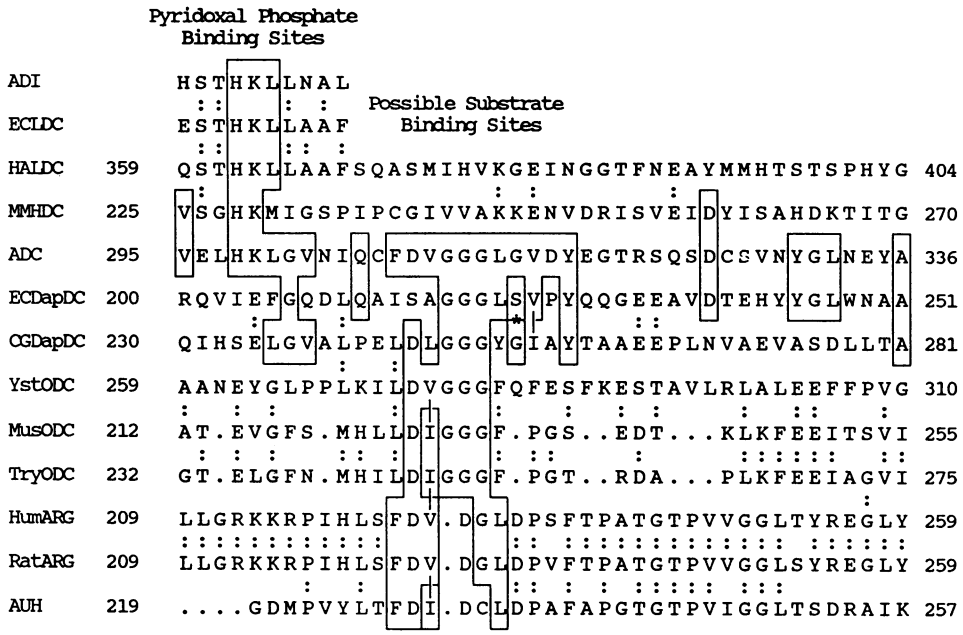


FIG. 5. Comparison of ADC and other enzymes which recognize structurally similar substrates. Gaps have been introduced to achieve the optimum alignment. Amino acids matching ADC are boxed, conservative changes relative to ADC (|), other matching residues (:), and mismatches (*) are indicated. Shown are *E. coli* biodegradative ADC (inducible) (ADI), LDC (ECLDC), and DapDC (ECDapDC), *H. alvei* LDC (HALDC), *M. morganii* HDC (MMHDC), *C. glutamicum* DapDC (CGDapDC), yeast (YstODC), mouse (MusODC), and trypanosome (TryODC) ODCs, human arginase (HumARG), rat arginase (RatARG), and *E. coli* AUH.

Sequence comparisons of ADC and amino acid decarboxylases. The results of a computer search of the GenEMBL and NBRF databases revealed that ADC contains small regions of identity with five other decarboxylases and two arginases. A six-amino-acid region, D-V-G-G-G-L, is conserved in diaminopimelic acid decarboxylases (DapDC) from *E. coli* and *Corynebacterium glutamicum* (Fig. 5). This conserved region appears in the same relative position in each enzyme (53 to 57% toward the carboxy terminus) and is located eight amino acids from the presumed pyridoxal phosphate-binding site in ADC, H-K-L. Both *E. coli* and *C. glutamicum* DapDCs were approximately 67% identical in this sequence. Yeast (D-V-G-G-G-F), mouse (D-I-G-G-G-F), and trypanosome (D-I-G-G-G-F) ODCs also appear to contain sequences with homology to ADC (D-V-G-G-G-L). These regions occur from 53 to 61% toward the carboxy end of each enzyme and were 83, 67, and 67% identical to the ADC sequences, respectively. Human and rat arginase both contain the sequence F-D-V-D-G-L near their carboxy termini. This arginase sequence is 71% identical to the region conserved in ADC (F-D-V-G-G-G-L).

Mapping of *speA* gene within *E. coli* chromosome. The *speA* gene has been mapped by cotransduction to 62.8 (11) minutes on the *E. coli* chromosome (11). The *speB*, *speA*, and *metK* genes, in this gene order, have been identified on plasmid pKA5, which contains an 8.0-kb *Eco*RI fragment derived from the *E. coli* chromosome (2). The DNA sequence reported here contains the *speA* gene and intergenic regions as well as partial sequences of the *speB* and *metK* genes. Since *speB* and *metK* have been sequenced (20, 34), it was possible to produce a 6,530-bp continuous DNA sequence of this region in pKA5. A restriction map of this 6,530-bp sequence was generated and compared with the restriction map of the *E. coli* chromosome constructed by Kohara et al. (18). The *speA* gene is located at 62.9 minutes on the map of Kohara et al. (18) and is contained in lambda clones 1H10 and 23G45. The presence of the *speA* gene on these lambda clones has been confirmed by Southern analysis with a *speA* probe (C. Satishchandran, personal communication).

DISCUSSION

This report describes the subcloning and nucleotide sequence of the *speA* gene encoding *E. coli* ADC. When ADC was originally purified and analyzed by Wu and Morris (43), they reported that purified ADC resolved into two major and three minor bands when electrophoresed on a native gel. On a denaturing gel, these two major bands migrated with molecular masses corresponding to 74,000 and 70,000 daltons. When either the 74- or 70-kilodalton band was assayed for ADC activity, each species decarboxylated arginine. Isolation of ADC by immunoprecipitation of pulse-labeled *E. coli* also demonstrated two major bands corresponding to 74,000 and 70,000 daltons, as well as minor bands in the 30- to 40,000-dalton range (4). These minor bands are probably nascent chains of ADC because they disappeared when the label was chased. The DNA sequence of the 3.2-kb *Bal*I-*Acc*I fragment in pRM15 revealed only one ORF (ORF1) capable of encoding a 74,000- or 70,000-dalton polypeptide.

The loss of approximately 4,000 daltons from a precursor is characteristic of the processing of a signal peptide during translocation (27, 29). The presence of a signal peptide in ADC was supported by the finding that the 74,000-dalton species, but not the 70,000-dalton species, accumulated in an *E. coli* strain unable to process signal sequences (4). Furthermore, selective disruption of the cell envelope showed that the 70,000-dalton species is localized within the inner periplasmic space (4). A review of procaryotic signal peptide characteristics (27, 29) indicates that the predicted amino terminus of ADC does not possess a typical signal sequence. Signal peptides usually contain three distinct regions: a

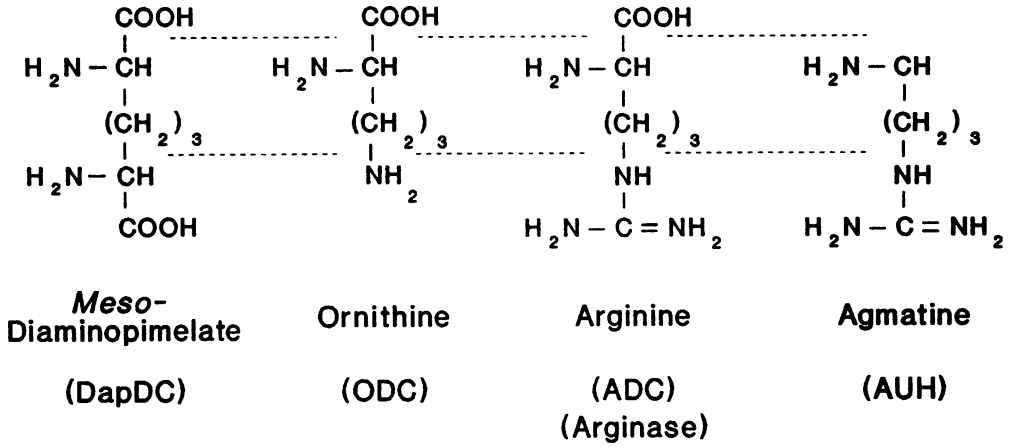


FIG. 6. Structurally similar substrates used by enzymes containing the conserved sequence F-D-V-G-G-G-L (Fig. 5). Dashed lines indicate the common structural skeleton recognized by each enzyme. Enzymes are in parentheses. All decarboxylases remove the carboxyl group in the C-1 position. Arginase and AUH remove the urea group from their respective substrates.

positively charged amino terminus, a core of 12 to 20 primarily hydrophobic amino acid residues, followed by a signal peptidase processing site. Kyte-Doolittle hydropathy plots (not shown) indicate that a signal peptide starting at the beginning of ORF1 would have a hydrophobic core composed of two short hydrophobic regions but would possess a negatively charged amino terminus. However, a net positive charge on a signal peptide is not an absolute requirement for export to the periplasm (15, 16, 28, 41).

The rate of processing of precursor ADC to the mature form occurs with kinetics atypical for most signal sequences (4). Signal peptides are usually processed very rapidly, usually within 30 s to 1 min. Studies employing site-directed mutagenesis on various *E. coli* signal peptides have shown that progressively decreasing the net charge on the amino terminus results in a corresponding decrease in the efficiency of translocation and a reduced rate of synthesis (16, 28, 41). When the net charge of the signal peptide of *E. coli* prolipoprotein was decreased from +2 to −2, 12 min were required to completely convert prolipoprotein into lipoprotein, resulting in an accumulation of the precursor in the cytoplasm. Pulse-chase experiments in *E. coli* showed that the ADC precursor requires approximately 6 min to be converted into mature ADC (4). The slow processing rate suggests that ADC accumulates in the cytoplasm after synthesis and is processed posttranslationally. This appears to be the case as Morris and Koffron (24) found that 20% of the putrescine synthesized in an *E. coli* K-12 strain grown in the absence of arginine utilized the ADC-AUH pathway. Thus, a temporary distribution of active ADC in the periplasm versus the cytoplasm might explain the partitioning of endogenously synthesized arginine between polyamine and protein synthesis.

Most decarboxylases require pyridoxal phosphate as a cofactor for catalytic activity, and many of these enzymes contain a pyridoxal phosphate-binding consensus sequence, S-X-H-K. *E. coli* ADC and lysine decarboxylase (LDC) and *Morganella morganii* histidine decarboxylase (HDC) contain either the amino acid sequence S-X-H-K or simply H-K in their pyridoxal phosphate-binding sites (1, 7, 40). The function of the histidine and lysine residues in the pyridoxal phosphate-binding consensus sequence has been investigated in *M. morganii* HDC by site-directed mutagenesis (40). Studies indicated that Lys-232, homologous to Lys-299 in ADC, was involved in decarboxylation but not in pyridoxal phosphate or histidine binding. His-231, homologous to His-298 in ADC, appeared to play a role in hydrogen bonding of HDC to histidine or pyridoxal phosphate. Biodegradative ADC and LDC are similar in subunit size, molecular weight, and their pyridoxal phosphate-binding sequences (1, 22). In these enzymes, pyridoxal phosphate binds to the ε-amino group of the lysine residue within the conserved region, S-T-H-K-L-L (22). The predicted pyridoxal phosphate-binding site of *Hafnia alvei* LDC is similar (9 of 10 residues match) to the *E. coli* LDC pyridoxal phosphate-binding site and also contains the amino acid sequence S-T-H-K-L-L (7) (Fig. 5). Three of these conserved amino acids, H-K-L, are retained in biosynthetic ADC. The conservation of the amino acid sequence H-K-L in the pyridoxal phosphate-binding sites of several amino acid decarboxylases and their similar relative positions within each enzyme (45 to 49% toward the carboxy end) suggest that H-K-L serves as the pyridoxal phosphate-binding site for biosynthetic ADC.

Pyridoxal phosphate-binding sites are usually followed by a hydrophobic sequence that is thought to form a hydrophobic pocket in which the cofactor or the substrate binds (22). Eight residues from the presumed pyridoxal phosphate-binding site of ADC begins an amino acid sequence, D-V-G-G-G-G-L, which is conserved (67 to 71% identical) in *E. coli* and *Corynebacterium* DapDCs and three eucaryotic ODCs. Hydropathy plots of each of these amino acid sequences (data not shown) reveal that all are predicted to produce a hydrophobic region. The possibility that this region is involved in substrate binding is supported by the observation that ADC, DapDCs, and ODCs each recognize structurally similar substrates and catalyze similar reactions (each removes the C-1 carboxyl group [Fig. 6]). In contrast, *H. alvei* LDC and *M. morganii* HDC pyridoxal phosphate-binding sites are followed by a hydrophobic sequence, but neither LDC or HDC shows any homology with ADC in this region (Fig. 5). Human and rat arginases, which also use arginine as a substrate, contain a similar sequence, F-D-V-D-G-L, that is 71% identical to the conserved region of ADC. Interestingly, this sequence is within one of the most highly conserved regions between arginase and *E. coli* AUH, the second enzyme in the ADC pathway (Fig. 5 and 6).

The *speA* −10 and −35 promoter regions are separated by 21 bp, suggesting that the ADC promoter is weakly tran-

scribed (13). Preliminary experiments indicate that the *speA* promoter is indeed weakly transcribed since a Φ(*speA-lacZ'*) fusion borne on a multicopy plasmid produces less β-galactosidase activity than does a single-copy chromosomal *lacZ* gene in *E. coli*.

A polycistronic mRNA containing *speA* and *speB* is produced from the *speA* promoter. The *speB* gene is also capable of transcription from an additional promoter located within the 5' end of the *speA* ORF to produce a monocistronic *speB* mRNA (34). A classic rho-independent terminator structure begins 69 bp downstream from the end of the *speA* ORF and ends 34 bp upstream of the *speB* ORF. This 37-bp structure has a 15-bp G+C-rich stem (with one mismatch), a seven-base loop, and ends with seven consecutive U residues. The structure does not appear to function as an attenuator as it lacks a suitable ribosome-binding site and contains a translation stop codon within the stem. How transcription of the *speB* gene occurs if this terminatorlike structure exists in vivo is unknown, although it may play a role in regulating the expression of the *speB* gene. The *speB* gene, but not *speA*, is induced by agmatine, while both the *speA* and *speB* genes are repressed by cAMP.

The presence of a CRP-binding consensus sequence in the promoter region of *speA* is consistent with its inhibition by cAMP (42). Presumably the cAMP-CRP complex would bind to this site and interfere with the initiation of transcription of the *speA* gene. However, preliminary experiments with *E. coli* bearing the Φ(*speA-lacZ'*) translational fusion plasmid pRM65 demonstrated that the production of ADC::β-galactosidase fusion protein is not inhibited by cAMP, in contrast to the induction of β-galactosidase synthesis by cAMP in control cells. Additional experiments indicate that two factors contribute to this apparent insensitivity of the plasmid-borne Φ(*speA-lacZ'*) fusion to cAMP-mediated regulation. First, the high copy number of plasmid pRM65 influences the regulation of the *speA* gene by cAMP. Second, the degree of repression of ADC by cAMP appears to be strain dependent. This finding is not surprising, as the *speA* genes of some strains of *E. coli* are known to be insensitive to cAMP-mediated inhibition (32). We are currently attempting to clarify the role of cAMP-mediated regulation of the *speA* gene.

ACKNOWLEDGMENTS

We thank C. Satishchandran for communicating his results to us prior to publication, D. Dean and N. Sriranganathan for critically reading the manuscript, M. Szumanski for helpful discussion, and J. Kromhout for technical assistance in preparing the immunoblot.

This work was supported by grant DMB-8508917 from the National Science Foundation (S.M.B.), by graduate fellowships from the College of Veterinary Medicine and Virginia Polytechnic Institute and State University (R.C.M.), and by a grant-in-aid from Sigma Xi.

LITERATURE CITED

1. **Boeker, E. A., E. H. Fischer, and E. E. Snell.** 1971. Arginine decarboxylase from *Escherichia coli*. IV. Structure of the pyridoxal phosphate binding site. J. Biol. Chem. **246**:6776–6781.
2. **Boyle, S. M., G. D. Markham, E. W. Hafner, J. M. Wright, H. Tabor, and C. W. Tabor.** 1984. Expression of the cloned genes encoding the putrescine biosynthetic enzymes and methionine adenosyltransferase of *Escherichia coli* (*speA*, *speB*, *speC* and *metK*). Gene **30**:129–136.
3. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248–254.
4. **Buch, J. K., and S. M. Boyle.** 1985. Biosynthetic arginine decarboxylase in *Escherichia coli* is synthesized as a precursor and located in the cell envelope. J. Bacteriol. **163**:522–527.
5. **Casadaban, M. J., J. Chou, and S. N. Cohen.** 1980. In vitro fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals. J. Bacteriol. **143**:971–980.
6. **Clarke, L., and J. Carbon.** 1976. A colony bank containing synthetic Col E1 hybrid plasmids representative of the entire *E. coli* genome. Cell **9**:91–99.
7. **Fecker, L. F., H. Beier, and J. Berlin.** 1986. Cloning and characterization of a lysine decarboxylase gene from *Hafnia alvei*. Mol. Gen. Genet. **203**:177–184.
8. **Fowler, A. V., and I. Zabin.** 1978. Amino acid sequence of β-galactosidase. XI. Peptide ordering and the complete sequence. J. Biol. Chem. **253**:5521–5525.
9. **Gale, E. F.** 1946. The bacterial amino acid decarboxylases. Adv. Enzymol. **6**:1–32.
10. **Glansdorff, N.** 1987. Biosynthesis of arginine and polyamines, p. 336–339. *In* J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
11. **Hafner, E. W., C. W. Tabor, and H. Tabor.** 1977. Isolation of a *metK* mutant with a temperature-sensitive *S*-adenosylmethionine synthetase. J. Bacteriol. **132**:832–840.
12. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
13. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
14. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**: 351–359.
15. **Iino, T., M. Takahashi, and T. Sako.** 1987. Role of amino-terminal positive charge on signal peptide in staphylokinase export across the cytoplasmic membrane of *Escherichia coli*. J. Biol. Chem. **262**:7412–7417.
16. **Inouye, S., X. Soberon, T. Franceschini, K. Nakamura, K. Itakura, and M. Inouye.** 1982. Role of positive charge on the amino-terminal region of the signal peptide secretion across the membrane. Proc. Natl. Acad. Sci. USA **79**:3438–3441.
17. **Ish-Horowicz, D., and J. F. Burke.** 1981. Rapid and efficient cosmid cloning. Nucleic Acids Res. **9**:2989–2998.
18. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for the rapid analysis and sorting of a large genomic library. Cell **50**:495–508.
19. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Markham, G. D., J. DeParasis, and J. Gatmaitan.** 1984. The sequence of *metK*, the structural gene for S-adenosylmethionine synthetase in *Escherichia coli*. J. Biol. Chem. **259**:14505–14507.
21. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Morris, D. R., and E. A. Boeker.** 1983. Biosynthetic and biodegradative ornithine and arginine decarboxylases from *Escherichia coli*. Methods Enzymol. **94**:125–134.
23. **Morris, D. R., and R. H. Fillingame.** 1974. Regulation of amino acid decarboxylation. Annu. Rev. Biochem. **43**:303–325.
24. **Morris, D. R., and K. L. Koffron.** 1969. Putrescine biosynthesis in *Escherichia coli* (regulation through pathway selection). J. Biol. Chem. **244**:6094–6099.
25. **Morris, D. R., W. H. Wu, D. Appelbaum, and K. L. Koffron.** 1970. Regulatory pattern in putrescine biosynthesis in *Escherichia coli*. Ann. N.Y. Acad. Sci. **171**:968–976.
26. **Niedhardt, F. C., P. L. Bloch, and D. F. Smith.** 1974. Culture medium for enterobacteria. J. Bacteriol. **119**:736–747.
27. **Oliver, D.** 1985. Protein secretion in *Escherichia coli*. Annu. Rev. Microbiol. **39**:615–648.
28. **Puziss, J. W., J. D. Fikes, and P. J. Bassford, Jr.** 1989. Analysis of mutational alterations in the hydrophilic segment of the

maltose-binding protein signal peptide. J. Bacteriol. **171**:2303–2311.
29. **Randall, L. L., S. J. S. Hardy, and J. R. Thom.** 1987. Export of protein: a biochemical view. Annu. Rev. Microbiol. **41**:507–541.
30. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
31. **Schneider, K., and C. F. Beck.** 1986. Promoter-probe vectors for the analysis of divergently arranged promoters. Gene **42**:37–48.
32. **Shaibe, E., E. Metzer, and Y. S. Halpern.** 1985. Control of utilization of L-arginine, L-ornithine, agmatine, and putrescine as nitrogen sources in *Escherichia coli* K-12. J. Bacteriol. **163**:938–942.
33. **Struhl, K.** 1986. A rapid method for creating recombinant DNA molecules. BioTechniques **3**:452–453.
34. **Szumanski, M. B. W., and S. M. Boyle.** 1990. Analysis and sequence of the *speB* gene encoding agmatine ureohydrolase, a putrescine biosynthetic enzyme in *Escherichia coli*. J. Bacteriol. **172**:538–547.
35. **Tabor, C. W., and H. Tabor.** 1985. Polyamines in microorganisms. Microbiol. Rev. **49**:81–99.
36. **Tabor, C. W., H. Tabor, E. W. Hafner, G. D. Markham, and S. M. Boyle.** 1983. Cloning of the *Escherichia coli* genes for the biosynthetic enzymes for polyamines. Methods Enzymol. **94**: 117–121.
37. **Tabor, H., and C. W. Tabor.** 1969. Formation of 1,4-diaminobutane and of spermidine by an ornithine auxotroph of *Escherichia coli* grown on limiting ornithine or arginine. J. Biol. Chem. **244**:2286–2292.
38. **Tabor, H., and C. W. Tabor.** 1969. Partial separation of two pools of arginine in *Escherichia coli*: preferential use of exogenous rather than endogenous arginine for the biosynthesis of 1,4-diaminobutane. J. Biol. Chem. **244**:6383–6387.
39. **Tarrof, C. W., and C. A. Hobbs.** 1987. Improved media for growing plasmid and cosmid clones. Focus **9(2)**:12.
40. **Vaaler, G. L., and E. E. Snell.** 1989. Pyridoxal 5′-phosphate dependent histidine decarboxylase: overproduction, purification, biosynthesis of soluble site-directed mutant proteins, and replacement of conserved residues. Biochemistry **28**:7306–7313.
41. **Vlasuk, G. P., S. Inouye, H. Ito, K. Itakura, and M. Inouye.** 1983. Effects of the complete removal of basic amino acid residues from the signal peptide on secretion of lipoprotein in *Escherichia coli*. J. Biol. Chem. **258**:7141–7148.
42. **Wright, J. W., and S. M. Boyle.** 1982. Negative control of ornithine decarboxylase and arginine decarboxylase by adenosine-3′:5′-cyclic monophosphate in *Escherichia coli*. Mol. Gen. Genet. **186**:482–487.
43. **Wu, W. H., and D. R. Morris.** 1973. Biosynthetic arginine decarboxylase from *Escherichia coli* (purification and properties). J. Biol. Chem. **248**:1687–1695.