Sequence Analysis of the *Escherichia coli dnaE* Gene

HENRY G. TOMASIEWICZ AND CHARLES S. McHENRY*

Department of Biochemistry, Biophysics and Genetics, University of Colorado Health Sciences Center, Denver, Colorado 80262

We have determined the sequence of a 4,350-nucleotide region of the *Escherichia coli* chromosome that contains *dnaE*, the structural gene for the α subunit of DNA polymerase III holoenzyme. The *dnaE* gene appeared to be part of an operon containing at least three other genes: 5'-*lpxB*-ORF$_{23}$-*dnaE*-ORF$_{37}$-3' (ORF, open reading frame). The *lpxB* gene encodes lipid A disaccharide synthase, an enzyme essential for cell growth and division (M. Nishijima, C. E. Bulawa, and C. R. H. Raetz, J. Bacteriol. 145:113–121, 1981). The termination codons of *lpxB* and ORF$_{23}$ overlapped the initiation codons of ORF$_{23}$ and *dnaE*, respectively, suggesting translational coupling. No rho-independent transcription termination sequences were observed. A potential internal transcriptional promoter was found preceding *dnaE*. Deletion of the −35 region of this promoter abolished *dnaE* expression in plasmids lacking additional upstream sequences. From the deduced amino acid sequence, α had a molecular weight of 129,920 and an isoelectric point of 4.93 for the denatured protein. ORF$_{23}$ encoded a more basic protein (pI 7.11) with a molecular weight of 23,228. In the accompanying paper (D. N. Crowell, W. S. Reznikoff, and C. R. H. Raetz, J. Bacteriol. 169:5727–5734, 1987), the sequence of the upstream region that contains *lpxA* and *lpxB* is reported.

The DNA polymerase III holoenzyme is the replicative complex responsible for the synthesis of most of the *Escherichia coli* chromosome (for a review, see reference 33). It contains at least two each of seven different subunits, α, β, γ, δ, ε, τ, and θ. α, the catalytic subunit and *dnaE* gene product, can be isolated in a complex with ε, the 3' → 5' exonuclease (10, 51), and θ in the enzyme form termed DNA polymerase III. DNA polymerase III, although as active as holoenzyme on duplex DNA containing short gaps, is inert on the *E. coli* chromosome and on other natural templates such as those provided by bacteriophages (33, 35, 61) or plasmids (38, 56). The addition of the remaining auxiliary subunits confers on the core polymerase those special properties required for it to function on natural templates. These properties include processivity, rapidity, the ability to participate in specific protein-protein interactions with other DNA replication components, and the ability to coordinate leading- with lagging-strand replication mediated through allosteric interactions between two halves of a dimeric enzyme (25, 33).

Genetic studies underscore the biochemical identification of the DNA polymerase III holoenzyme as the replicative complex. Temperature-sensitive mutations in *dnaE* and the structural genes for the other holoenzyme subunits are lethal under nonpermissive conditions, indicative of the essential role of this enzyme in replication. The structural genes that encode the various subunits of DNA polymerase III holoenzyme are located at widely disparate regions on the *E. coli* chromosome. The β subunit is encoded by *dnaN*, which is located distal to the *dnaA* gene in an operon at 83 min on the *E. coli* genetic map (6). The *dnaZ* and *dnaX* genes map adjacent to one another at 10.5 min (19). The sequences of *dnaZ* and *dnaX* indicate that they reside in one long open reading frame (ORF) that apparently encodes two holoenzyme subunits, τ and γ, by a mechanism not yet understood (11, 21, 22, 26, 39, 65). The structural gene for the ε subunit, *dnaQ* (*mutD*), maps at 5 min (10, 20); the *dnaE* gene is

located at 4 min (15). The gene that encodes the θ subunit is unknown.

A clone containing the *dnaE* gene was identified by its ability to complement *dnaE* temperature-sensitive mutations and to direct the synthesis of α in the maxicell system (60). Deletion analysis of this clone established the minimal sequence required for expression of the *dnaE* gene and the direction of transcription as clockwise on the *E. coli* chromosome (53). As part of an ongoing study of DNA replication, our laboratory is interested in the genetic regulation of replication protein synthesis and the coordination of this process with the global control of cell growth and division. Towards this goal, we have determined the nucleotide sequence of a representative replication gene, *dnaE*, and the flanking regions. These data have yielded valuable information that suggests testable hypotheses about the regulation of this essential gene.

MATERIALS AND METHODS

Bacterial strains and media. *E. coli* strains used in this study were CSR603 (*recA1 uvrA6 phr-1*) (45) and JM109 [*recA1 endA1 gyrA96 thi supE44 relA1* Δ(*lac-proAB*) (F' *traD36 proAB*) *lacI^{q}ZΔM15*)] (61). CSR603 cells carrying *dnaE*-containing plasmids (53) were grown on L broth (37) supplemented with thymine (50 μg/ml) and ampicillin (40 μg/ml). JM109 cells were grown on YT broth (37).

Enzymes. All enzymes were obtained from New England Biolabs, except RNase T$_1$ (Boehringer-Mannheim) and were used as specified by the manufacturer.

Preparation of template DNA for sequencing. Cells possessing *dnaE*-containing plasmids were grown overnight (L plus thymine broth with ampicillin), pelleted by centrifugation (6,000 × g), and suspended (2 ml/100 ml of original culture) in a solution containing 25 mM Tris hydrochloride (pH 8.0), 50 mM EDTA, and 1% (wt/vol) glucose. All procedures were performed at 4°C unless otherwise stated. The cells were lysed by adding 2 volumes of a solution containing 0.2 M NaOH and 1% (wt/vol) sodium dodecyl sulfate for 10 min. One-half total volume of 3 M potassium acetate (pH 4.8, final concentration 1 M) was then added (5

---

* Corresponding author.

min), and the solution was clarified by centrifugation (12,000 × g, 30 min). The plasmid DNA was precipitated from the supernatant by addition of 2 volumes of ethanol (−20°C, 2 h). The DNA, pelleted by centrifugation (12,000 × g, 20 min), was suspended (10 mM Tris hydrochloride [pH 7.5], 1 mM EDTA, 1% of original culture). Ammonium acetate was added (0.19 g/ml, final concentration, 4°C, 30 min); the resulting solution was clarified by centrifugation (12,000 × g, 20 min), and the plasmid DNA was precipitated with ethanol as before. The resulting pellet was suspended in 400 μl of 50 mM Tris hydrochloride–1 mM EDTA–50 mM NaCl. RNA was digested with RNase T_1 (6 to 8 U per A_{260} unit, 37° C, 30 min). The solution was extracted once with an equal volume of phenol and then an equal volume of phenol-chloroform (1:1). The plasmid DNA was ethanol precipitated and suspended in 20 mM Tris hydrochloride (pH 7.5)–0.1 mM EDTA.

The dnaE-containing inserts from the plasmids were excised by digesting plasmid DNA with the restriction endonucleases EcoRI and ClaI (Fig. 1). The replicative form DNA of M13mp18 and M13mp19 was digested with the restriction endonucleases EcoRI and AccI. The dnaE inserts (0.8 pmol/ml) and EcoRI-AccI-digested M13mp18 and M13mp19 replicative form DNA (4 pmol/ml) were ligated by using T4 DNA ligase (8,000 U/ml, 14°C, 4 h). Resultant molecules were used to transform E. coli JM109, which were then plated on YT medium containing 5-bromo-4-chloro-3-indolyl-β-D-galactoside and isopropyl-thiogalactoside. Phages containing inserts within the lacZ gene formed colorless plaques. Recombinant phage DNA was analyzed for the presence of appropriate size inserts by comparing their mobility with that of phage DNA containing inserts of known length in agarose gels.

Since the presence of dnaE in the phage genome is nonessential for phage survival and may even be detrimental to phage growth, the following steps were taken to avoid overgrowth by phage containing mutations in dnaE: (i) the amount of phage in supernatants was amplified by only one passage through E. coli, starting with phage isolated from a single plaque, (ii) E. coli JM109 cells were infected at a multiplicity of infection of 2 to 3 to ensure that all cells were infected immediately upon the addition of the phage, and (iii) the phage were harvested after only 6.5 h of growth.

Phage were suspended in 50 mM Tris hydrochloride (pH 7.5)–1 mM EDTA–0.5% Sarkosyl. The solution was clarified by centrifugation (8,000 × g, 5 min). Phage were precipitated by adding NaCl and polyethylene glycol 6000 (final concentration, 0.5 M and 2% [wt/vol], respectively) (63). This was repeated twice. In the final step, phage were precipitated from the same solutions lacking Sarkosyl. Template DNA was prepared from the purified phage by phenol extraction followed by ethanol precipitation.

DNA sequencing. The dnaE gene was sequenced by the dideoxyribonucleotide chain termination procedure of Sanger et al. (50) and the incorporation of [α-^{35}S]dATP (New England Nuclear) as described by New England Biolabs, except the reactions were incubated at 37°C and polymerase was included with the second addition of deoxyribonucleoside triphosphates.

Sequencing gels (8% [wt/vol] acrylamide, 8 M urea, 90 mM Tris, 90 mM borate, 4 mM EDTA, pH 8.3) (49) were run at constant power to maintain the surface temperature at 60°C. Samples were run for 45 min after the bromophenol blue reached the bottom of the gel, whereupon a second sample was added in adjacent lanes. The gels were fixed and dried to plates as described (14), and the bands were visualized by autoradiography with Kodak XAR-5 film.

Synthetic oligodeoxyribonucleotides were prepared on a BioSearch oligodeoxyribonucleotide synthesizer, purified on 20% (wt/vol) acrylamide–8 M urea gels, and detected by shadowing on a UV fluorescence plate. The oligodeoxyribonucleotides were cut out, pulverized by being forced through a wire mesh, and eluted in 10 mM Tris hydrochloride–1 mM EDTA (pH 7.5). The resulting oligodeoxyribonucleotides were purified on DE52 cellulose with elution by 3 M potassium acetate (1; P. Hagerman, personal communication).

Quantitative values for the protein sequence homologies were determined by the structure-genetic matrix program of PCGene by Intelligenetics. The isoelectric points of α, β, and the protein encoded by ORF_{23} were calculated by the PCGene program.

Two-dimensional polyacrylamide gel electrophoresis. Two-dimensional gel electrophoresis of purified DNA polymerase III holoenzyme (1.74 μg) was performed as described by O'Farrell (42). Isoelectric focusing was performed with ampholines (LKB) in the pH ranges 3 to 10 (1%) and 5 to 7 (1%) and run for 7,000 V-h. Holoenzyme subunits were
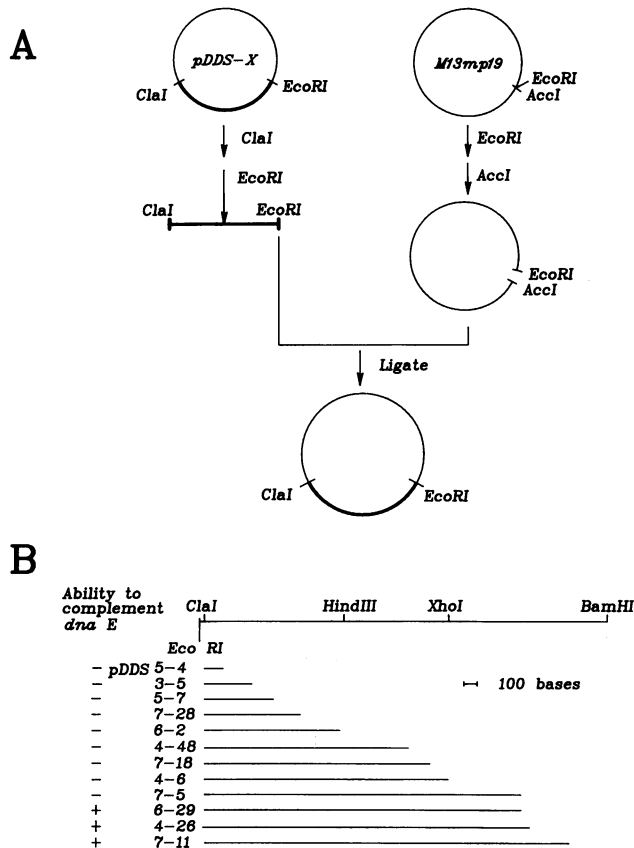
FIG. 1. (A) Cloning of inserts containing deletions in dnaE into M13mp18 and mp19. Plasmids containing inserts with deletions from the 5' terminus of dnaE were digested with the restriction endonucleases ClaI and EcoRI and ligated with either M13mp18 or mp19 replicative form DNA digested with the restriction endonucleases AccI and EcoRI. Two additional ClaI restriction endonuclease recognition sites were identified from the sequence of dnaE. Since the recognition sequence for the dam methylase overlaps these ClaI sites, dnaE-containing plasmids grown in Dam+ cells are not cleaved at these sites. (B) Functionality and size of deletions from the 5' terminus of dnaE that were cloned into M13mp19. pDDS7-11 was cloned into both M13mp18 and mp19.

separated in the second dimension by electrophoresis on a sodium dodecyl sulfate–7.5 to 17.5% polyacrylamide gel. The polypeptides were transferred to nitrocellulose and visualized with peroxidase-conjugated antiholoenzyme antibodies (62).

RESULTS

The region of the *E. coli* chromosome containing the *dnaE* gene was sequenced by the strategy outlined in Fig. 2. Inserts from clones containing a nested set of deletions from the amino-terminal coding region of the *dnaE* gene (Fig. 1B) were inserted into either M13mp18 or M13mp19 replicative form DNA. Since each deletion was defined by the cleavage recognition sequence for the restriction endonuclease *Eco*RI and other terminus contained a *Cla*I site, they could readily be cloned into the proper orientation for sequencing. An oligodeoxyribonucleotide complementary to a sequence adjacent to the *Eco*RI site of the inserts was used as a primer to sequence the *dnaE* gene by the method of Sanger et al. (50). By sequencing the first 250 to 300 nucleotides of the different inserts, the nucleotide sequence of the gene was determined. In places where the distance was too great between deletions, additional primers were synthesized that were complementary to the terminus of the preceding sequence. The second strand was sequenced almost entirely by the latter strategy.

Sequence analysis. We sequenced 4,350 nucleotides of the *E. coli* chromosome containing the *dnaE* gene. The *dnaE* gene, extending from nucleotides 796 to 4275 (Fig. 3), was 3,480 nucleotides long and encoded a 129,920-dalton protein. This is in agreement with the previously reported molecular weight of α (140,000) (34) determined by electrophoresis on sodium dodecyl sulfate-polyacrylamide gels. The *dnaE* gene possessed a relatively weak potential ribosome-binding site that contained a Shine-Dalgarno sequence (Fig. 3) (54), an adenosine at position −3, a conserved residue in the ribosome-binding site for many genes (for a review, see reference 57), and the ATG initiation codon.

Nucleotides 589 to 594 and 612 to 617 preceding the *dnaE* gene (Fig. 3) were homologous to the −35 boxes and −10 boxes, respectively, of known promoters (32, 63). That this promoter can function in vivo was demonstrated by deletion analysis. Plasmid pDDS6-29 (Fig. 3), which lacks sequences upstream of the −35 box, was able to complement cells containing a temperature-sensitive lesion in the *dnaE* gene and direct the synthesis of α in the maxicell system; plasmid pDDS7-5 (Fig. 3), lacking the proposed −35 box, was unable to support either of these functions.

Upstream of the putative promoter were two sequences in opposite orientation located at nucleotides 176 to 184 and

203 to 211 (Fig. 3) which were homologous (eight of nine nucleotides) to the *dnaA* boxes located in *oriC* and in the promoter region of the *dnaA* operon (2, 13).

In addition to *dnaE*, there were three other ORFs in the region we sequenced. The open reading frame from nucleotide 1 to 162 (Fig. 3) encoded the carboxyl terminus of lipid A disaccharide synthase, the *lpxB* gene product. This was shown by direct sequence comparison with the sequence of *lpxB* recently determined by Crowell et al. (9). Immediately following *lpxB* was an ORF, ORF23 (Fig. 3). ORF23 could code for a protein of molecular weight 23,228 with an isoelectric point of 7.11. A protein of this size has been observed in minicell plasmid-encoded protein systems containing this DNA fragment (8).

The third ORF, ORF37 (Fig. 3), followed *dnaE*. ORF37 encoded a 37,000-dalton protein (data not shown). Previous results obtained from maxicells in this laboratory indicated that the region after *dnaE* encoded a protein of molecular weight 37,000 (60).

All of these genes were transcribed from the same strand of DNA. The termination codons of *lpxB* and ORF23 overlapped the initiation codons of ORF23 and *dnaE*, respectively (Fig. 4). These results indicate that this region has the following organization: 5'-*lpxB*-ORF23-*dnaE*-ORF37-3'.

Sequence analysis of α. To check the accuracy of our nucleotide sequence of the *dnaE* gene, we compared the isoelectric point of α determined from its amino acid composition with the value obtained from its migration on two-dimensional polyacrylamide gels. α had a calculated pI of 4.93. Large coding regions in the other reading frames within *dnaE* coded for very basic polypeptides. We used the β subunit of DNA polymerase III holoenzyme, which had a calculated isoelectric point of 5.04, as a reference point. Both α and β exhibited an isoelectric point of 5.6 on two-dimensional gels (Fig. 5). The discrepancy between the calculated and observed isoelectric point could be due to urea diffusing out of the acidic end of the gel and altering the pK_as of the carboxyl groups (42). That the observed isoelectric point of α was similar to that of β, as predicted from its amino acid sequence, supports the accuracy of the *dnaE* sequence.

The amino acid sequence of α was examined for homology with DNA polymerase I. Initial results with a computer program (Microgenie) that used a matrix indicated that α had weak homology to polymerase I (data not shown). Thus, we limited more detailed examination to those regions demonstrated by Steitz and co-workers to be conserved between DNA polymerase I and T7 DNA polymerase (43). Using a computer program that identifies regions of homology by the method of McLachlan (36), we located sequences in α that demonstrated weak homology to five of the nine regions that
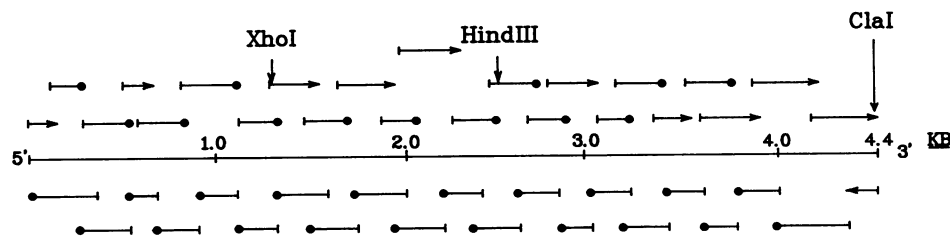
FIG. 2. Strategy for sequencing the *dnaE* gene. The region was sequenced by using either a universal primer complementary to M13 sequences located adjacent to the *dnaE*-containing inserts or synthetic oligodeoxynucleotide primers complementary to the terminus of the preceding region. Arrows represent regions sequenced by using the universal primer. Lines terminated with solid circles represent regions sequenced by using synthetic oligodeoxyribonucleotide primers.

```
lpxB.      .          .          .          .          50         .          .          .
CAGGAAGAGTGTGAGCCGCAAAAACTGGCTGCGGCGCTGTTACCGCTGTTGGCGAACGGGAAAACCAGCCACGCGATGCACGAT
GlnGluGluCysGluProGlnLysLeuAlaAlaAlaLeuLeuProLeuLeuAlaAsnGlyLysThrSerHisAlaMetHisAsp
           100           .          .          .          .          ****     *.
ACCTTCCGTGAACTGCATCAGCAGATCCGCTGCAATGCCGATGAGCAGGCGGCACAAGCCGTTCTGGAGTTAGCACAATGATCG
ThrPheArgGluLeuHisGlnGlnIleArgCysAsnAlaAspGluGlnAlaAlaGlnAlaValLeuGluLeuAlaGlnEnd
                                                          (orf23kDal)MetIleG
      ++++++++    .          200  ++++++++    .          .          .          250
AATTTGTTTATCCGCACACGCAGCTGGTTGCGGGTGTGGATGAAGTCGGACGCGGGCCGTTAGTTGGCGCGGTCGTCACCGCTG
luPheValTyrProHisThrGlnLeuValAlaGlyValAspGluValGlyArgGlyProLeuValGlyAlaValValThrAlaA
      .          .          .          300        .          .          .
CGGTGATCCTTGACCCGGCGCGCCCGATTGCCGGGCTGAATGATTCCAAAAAGCTGAGCGAAAAACGCCGTCTGGCGCTCTATG
laValIleLeuAspProAlaArgProIleAlaGlyLeuAsnAspSerLysLysLeuSerGluLysArgArgLeuAlaLeuTyrG
      .          350          .          .          400        .          .
AAGAGATCAAAGAGAAAGCGTTGAGCTGGAGTCTGGGCCGCGCGGAACCCCACGAAATCGACGAGCTGAACATTCTTCATGCGA
luGluIleLysGluLysAlaLeuSerTrpSerLeuGlyArgAlaGluProHisGluIleAspGluLeuAsnIleLeuHisAlaT
      .          .          450        .          .          500        .
CCATGCTGGCGATGCAGCGTGCCGTCGCTGGGCTGCATATTGCGCCGGAATATGTGTTGATTGATGGTAACCGCTGCCCGAAAT
hrMetLeuAlaMetGlnArgAlaValAlaGlyLeuHisIleAlaProGluTyrValLeuIleAspGlyAsnArgCysProLysL
      .          .          .          550        .          .          . 6-29/
TACCGATGCCTGCGATGGCTGTGGTGAAAGGCGATAGCCGCGTACCGGAAATCAGTGCCGCGTCTATCCTGGCGAAAGTGACGC
euProMetProAlaMetAlaValValLysGlyAspSerArgValProGluIleSerAlaAlaSerIleLeuAlaLysValThrA
      .          /7-5 .          .          .          650        .          .
GTGACGCCGAAATGGCGGCGCTGGATATTGTTTTCCCGCAATATGGTTTTGCCCAACACAAAGGGTACCCAACCGCTTTTCATC
rgAspAlaGluMetAlaAlaLeuAspIleValPheProGlnTyrGlyPheAlaGlnHisLysGlyTyrProThrAlaPheHisL
      .          700          .          .          .          .          750
TGGAAAAACTGGCTGAACACGGCGCGACCGAACACCATCGGCGCAGCTTTGGGCCTGTCAAACGCGCACTTGGGACTTGCGTCC
euGluLysLeuAlaGluHisGlyAlaThrGluHisHisArgArgSerPheGlyProValLysArgAlaLeuGlyThrCysValL
      .          .          .          ***     .          *          .          .
TGATTCTTGTGTCGAGATTAAGTAAACCGGAATCTGAAGATGTCTGAACCACGTTTCGTACACCTGCGGGTGCACAGCGACTAC
euIleLeuValSerArgLeuSerLysProGluSerGluAspValEnd
                               dnaE MetSerGluProArgPheValHisLeuArgValHisSerAspTyr
      850        .          .          .          900        .          .
TCGATGATCGATGGCCTGGCCAAAACCGCACCGTTGGTAAAAAAGGCGGCGGCGTTGGGTATGCCAGCACTGGCGATCACCGAT
SerMetIleAspGlyLeuAlaLysThrAlaProLeuValLysLysAlaAlaAlaLeuGlyMetProAlaLeuAlaIleThrAsp
      .          .          950        .          .          .          1000
TTCACCAACCTTTGTGGTCTGGTGAAGTTCTACGGAGCGGGACATGGCGCAGGGATTAAGCCTATCGTCGGGGCAGATTTTAAC
PheThrAsnLeuCysGlyLeuValLysPheTyrGlyAlaGlyHisGlyAlaGlyIleLysProIleValGlyAlaAspPheAsn
      .          .          .          1050       .          .          .          .
GTCCAGTGCGACCTGCTGGGTGATGAGTTAACCCACCTGACGGTACTGGCGGCGAACAATACCGGCTATCAGAATCTGACGTTG
ValGlnCysAspLeuLeuGlyAspGluLeuThrHisLeuThrValLeuAlaAlaAsnAsnThrGlyTyrGlnAsnLeuThrLeu


      1100       .          .          .          .          1150       .          .
CTGATCTCAAAAGCGTATCAGCGCGGGTACGGTGCCGCCGGGCCGATCATCGATCGCGACTGGCTTATCGAATTAAACGAAGGG
LeuIleSerLysAlaTyrGlnArgGlyTyrGlyAlaAlaGlyProIleIleAspArgAspTrpLeuIleGluLeuAsnGluGly
      .          .          1200       .          .          .          1250
TTGATCCTTCTTTCCGGCGGACGCATGGGCGACGTCGGACGCAGTCTTTTGCGTGGTAACAGCGCGCTGGTAGATGAGTGTGTC
LeuIleLeuLeuSerGlyGlyArgMetGlyAspValGlyArgSerLeuLeuArgGlyAsnSerAlaLeuValAspGluCysVal
      .          .          .          1300       .          .          .
GCGTTTTATGAACACTTCCCGGATCGCTATTTTCTCGAGCTGATCCGCACCGGCAGGCCGGATGAAGAAAGCGAATATCTGCAC
AlaPheTyrGluGluHisPheProAspArgTyrPheLeuGluLeuIleArgThrGlyArgProAspGluGluSerTyrLeuHis
      1350       .          .          .          .          1400       .          .
GCGGCGGTGGAACTGGCGGAAGCGCGCGGTTTGCCCGTCGTGGCGACCAACGACGTGCGCTTTATCGACAGCAGCGACTTTGAC
AlaAlaValGluLeuAlaGluAlaArgGlyLeuProValValAlaThrASnAspValArgPheIleAspSerSerAspPheAsp
      .          .          1450       .          .          .          1500       .
GCACACGAAATCCGCGTCGCGATCCACGACGGCTTTACCCTCGACGATCCTAAACGCCCGCGTAACTATTCGCCGCAGCAATAT
AlaHisGluIleArgValAlaIleHisAspGlyPheThrLeuAspAspProLysArgProArgAsnTyrSerProGlnGlnTyr
      .          .          .          1550       .          .          .
ATGCGTAGCGAAGAGGAGATGTGTGAGCTGTTTGCCGACATCCCCGAAGCCCTTGCCAACACCGTTGAGATCGCCAAACGCTGT
MetArgSerGluGluGluMetCysGluLeuPheAlaAspIleProGluAlaLeuAlaAsnThrValGluIleAlaLysArgCys
      1600       .          .          .          1650       .          .          .
AACGTAACCGTGCGTCTTGGTGAATACTTCCTGCCGCAGTTCCCGACCGGGGACATGAGCACCGAAGATTATCTGGTCAAGCGT
AsnValThrValArgLeuGlyGluTyrPheLeuProGlnPheProThrGlyAspMetSerThrGluAspTyrLeuValLysArg
      .          .          1700       .          .          .          1750       .
GCAAAAGAGGGCCTGGAAGAGCGTCTGGCCTTTTTATTCCCTGATGAGGAAGAACGTCTTAAGCGCCGCCCGGAATATGACGAA
AlaLysGluGlyLeuGluGluArgLeuAlaPheLeuPheProAspGluGluGluArgLeuLysArgArgProGluTyrAspGlu
      .          .          .          1800       .          .          .          18
CGTCTGGAGACTGAACTTCAGGTTATCAACCAGATGGGCTTCCCGGGCTACTTCCTCATCGTTATGGAATTTATCCAGTGGTCG
ArgLeuGluThrGluLeuGlnValIleAsnGlnMetGlyPheProGlyTyrPheLeuIleValMetGluPheIleGlnTrpSer
50         .          .          .          .          1900       .          .
AAAGATAACGGCGTACCGGTAGGGCCAGGCCGTGGCTCCGGTGCGGGTTCACTGGTGGCCTACGCGCTGAAAAATCACCGACCTC
LysAspAsnGlyValProValGlyProGlyArgGlySerGlyAlaGlySerLeuValAlaTyrAlaLeuLysIleThrAspLeu
      .          1950       .          .          .          .          2000       .
GATCCGCTGGAATTTGACCTGCTGTTCGAACGTTTCCTTAACCCGGAACGTGTCTCCATGCCTGACTTCGACGTTGACTTCTGT
AspProLeuGluPheAspLeuLeuPheGluArgPheLeuAsnProGluArgValSerMetProAspPheAspValAspPheCys
      .          .          2050       .          .          .          2100
ATGGAGAAACGCGATCAGGTTATCGAGCACGTAGCGGACATGTACGGTCGTGATGCGGTATCCAGATCATCACCTTCGGTACA
MetGluLysArgAspGlnValIleGluHisValAlaAspMetTyrGlyArgAspAlaValSerGlnIleIleThrPheGlyThr
      .          .          .          2150       .          .          .
ATGGCGGCGAAAGCGGTGATCCGCGACGTAGGCCGCGTGCTGGGGCATCCGTACGGCTTTGTCGATCGTATCTCGAAACTGATC
MetAlaAlaLysAlaValIleArgAspValGlyArgValLeuGlyHisProTyrGlyPheValAspArgIleSerLysLeuIle
      .          2200       .          .          .          .          2250
CCGCCCGATCCGGGGATGACGCTGGCGAAAGCGTTTGAAGCCGAGCCGCAGCTGCCGGAAATCTACGAAGCGGATGAAGAAGTT
ProProAspProGlyMetThrLeuAlaLysAlaPheGluAlaGluProGlnLeuProGluIleTyrGluAlaAspGluGluVal
```

```
        .         .         .         2300         .         .         .         .         2350
AAGGCGCTGATCGACATGGCGCGCAAACTGGAAGGGGTCACCCGTAACGCCGGTAAGCACGCCGGTGGGGTGGTTATCGCGCCG
LysAlaLeuIleAspMetAlaArgLysLeuGluGlyValThrArgAsnAlaGlyLysHisAlaGlyGlyValValIleAlaPro
        .         .         .         .         2400         .         .         .
ACCAAAATTACCGATTTTGCGCCGCTTTACTGCGATGAAGAGGGCAAACATCCGGTCACCCAGTTTGATAAAAGCGACGTTGAA
ThrLysIleThrAspPheAlaProLeuTyrCysAspGluGluGlyLysHisProValThrGlnPheAspLysSerAspValGlu
        .         2450         .         .         .         .         2500         .         .
TACGCCGGACTGGTGAAGTTCGACTTCCTTGGTTTGCGTACGCTCACCATCATCAACTGGGCGCTGGAGATGATCAACAAGCGG
TyrAlaGlyLeuValLysPheAspPheLeuGlyLeuArgThrLeuThrIleIleAsnTrpAlaLeuGluMetIleAsnLysArg
        .         .         2550         .         .         .         .         2600
CGGGCGAAGAATGGCGAGCCGCCGCTGGATATCGCTGCGATCCCGCTGGATGATAAGAAAAGCTTCGACATGCTGCAACGCTCG
ArgAlaLysAsnGlyGluProProLeuAspIleAlaAlaIleProLeuAspAspLysLysSerPheAspMetLeuGlnArgSer
        .         .         .         2650         .         .         .         .
GAAACCACGGCGGTATTCCAGCTTGAATCGCGCGGCATGAAGGACCTGATCAAGCGTCTACAACCTGACTGCTTCGAAGATATG
GluThrThrAlaValPheGlnLeuGluSerArgGlyMetLysAspLeuIleLysArgLeuGlnProAspCysPheGluAspMet
        .         2700         .         .         .         .         2750         .         .
ATCGCCCTAGTGGCACTGTTCCGCCCCGGTCCGTTGCAATCAGGGATGGTGGATAACTTTATCGACCGTAAACATGGTCGTGAA
IleAlaLeuValAlaLeuPheArgProGlyProLeuGlnSerGlyMetValAspAsnPheIleAspArgLysHisGlyArgGlu
        .         .         2800         .         .         .         .         2850
GAGATCTCCTATCCGGACGTACAGTGGCAGCATGAAGCCTGAAACCGGTACTGGAGCCAACCTACGGCATTATCCTGTATCAG
GluIleSerTyrProAspValGlnTrpGlnHisGluSerLeuLysProValLeuGluProThrTyrGlyIleIleLeuTyrGln
        .         .         .         2900         .         .         .         .
GAACAGGTCATGCAGATTGCGCAGGTGCTTTCTGGTTATACCCTCGGTGGCGCGGATATGCTGCGTCGTGCGATGGGTAAGAAA
GluGlnValMetGlnIleAlaGlnValLeuSerGlyTyrThrLeuGlyGlyAlaAspMetLeuArgArgAlaMetGlyLysLys
        2950         .         .         .         .         3000         .         .
AAGCCGGAAGAGATGGCTAAGCAACGTTCTGTATTTGCTGAAGGTGCAGAAAAGAACGGAATCAACGCTGAACTGGCGATGAAA
LysProGluGluMetAlaLysGlnArgSerValPheAlaGluGlyAlaGluLysAsnGlyIleAsnAlaGluLeuAlaMetLys
        .         .         3050         .         .         .         .         3100
ATCTTCGACCTGGTGGAGAAATTCGCTGGTTACGGGATTTAACAAATCGCACTCTGCGGCCTATGCTTTGGTGTCATATCAAACG
IlePheAspLeuValGluLysPheAlaGlyTyrGlyPheAsnLysSerHisSerAlaAlaTyrAlaLeuValSerTyrGlnThr
        .         .         .         3150         .         .         .         .
TTATGGCTGAAAGCGCACTATCCTGCGGAGTTTATGGCGGCGGTAATGACCGCCGATATGGACAACACCGAGAAGGTGGTGGGT
LeuTrpLeuLysAlaHisTyrProAlaGluPheMetAlaAlaValMetThrAlaAspMetAspAsnThrGluLysValValGly
        3200         .         .         .         .         3250         .         .
CTGGTGGATGAGTGCTGGCGGATGGGGCTGAAAATCCTGCCACCAGATATAAACTCCGGTCTTTACCATTTCCACGTCAACGAC
LeuValAspGluCysTrpArgMetGlyLeuLysIleLeuProProAspIleAsnSerGlyLeuTyrHisPheHisValAsnAsp
        .         .         3300         .         .         .         .         3350
GACGGCGAAATCGTGTATGGTATTGGCGCGATCAAAGGGGTCGGTGAAGGTCCGATTGAGGCCATCATCGAAGCCCGTAATAAA
AspGlyGluIleValTyrGlyIleGlyAlaIleLysGlyValGlyGluGlyProIleGluAlaIleIleGluAlaArgAsnLys
        .         .         .         3400         .         .         .         .
GGCGGCTACTTCCGCGAACTGTTTGATCTCTGCGCCCGTACCGACACCAAAAAGTTGAACCGTCGCGTGCTGGAAAAACTGATC
GlyGlyTyrPheArgGluLeuPheAspLeuCysAlaArgThrAspThrLysLysLeuAsnArgArgValLeuGluLysLeuIle


    3450         .         .         .         .         3500.         .         .
ATGTCCGGGGCGTTTGACCGTCTTGGGCCACATCGCGCAGCGCTGATGAACTCGCTGGGCGATGCGTTAAAAGCGGCAGATCAA
MetSerGlyAlaPheAspArgLeuGlyProHisArgAlaAlaLeuMetAsnSerLeuGlyAspAlaLeuLysAlaAlaAspGln
        .         .         3550         .         .         .         .         3600         .
CACGCGAAAGCGGAAGCTATCGGTCAGGCCGATATGTTCGGCGTGCTGGCCGAAGAGCCGGAACAAATTGAACAATCCTACGCC
HisAlaLysAlaGluAlaIleGlyGlnAlaAspMetPheGlyValLeuAlaGluGluProGluGlnIleGluGlnSerTyrAla
        .         .         .         3650         .         .         .         .
AGCTGCCAACCGTGGCCGGAGCAGGTGGTATTAGATGGGGAACGTGAAACGTTAGGCCTGTACCTGACCGGACACCCTATCAAC
SerCysGlnProTrpProGluGlnValValLeuAspGlyGluArgGluThrLeuGlyLeuTyrLeuThrGlyHisProIleAsn
3700         .         .         .         3750         .         .         .
CAGTATTTAAAAGAGATTGAGCGTTATGTCGGAGGCGTAAGGCTGAAAGACATGCACCCGACAGAACGTGGTAAAGTCATCACG
GlnTyrLeuLysGluIleGluArgTyrValGlyGlyValArgLeuLysAspMetHisProThrGluArgGlyLysValIleThr
        .         3800         .         .         .         .         3850
GCTGCGGGGCTCGTTGTTGCCGCGCGGGTTATGGTCACCAAGCGCGGCAATCGTATCGGTATCTGCACGCTGGATGACCGTTCC
AlaAlaGlyLeuValValAlaAlaArgValMetValThrLysArgGlyAsnArgIleGlyIleCysThrLeuAspAspArgSer
        .         .         .         3900         .         .         .         .         39
GGGCGGCTGGAAGTGATGTTGTTTACTGACGCCCTGGATAAATACCAGCAATTGCTGGAAAAAGACCGCATACTTATCGTCAGC
GlyArgLeuGluValMetLeuPheThrAspAlaLeuAspLysTyrGlnGlnLeuLeuGluLysAspArgIleLeuIleValSer
50         .         .         .         .         4000         .         .
GGACAGGTCAGCTTTGATGACTTCAGCGGTGGGCTTAAAATGACCGCTCGCGAAGTGATGGATATTGACGAAGCCCGGGAAAAA
GlyGlnValSerPheAspAspPheSerGlyGlyLeuLysMetThrAlaArgGluValMetAspIleAspGluAlaArgGluLys
        4050         .         .         .         .         4100
TATGCTCGCGGGCTTGCTATCTCGCTGACGGACAGGCAAATTGATGACCAGCTTTTAAACCGACTCCGTCAGTCTCTGGAACCC
TyrAlaArgGlyLeuAlaIleSerLeuThrAspArgGlnIleAspAspGlnLeuLeuAsnArgLeuArgGlnSerLeuGluPro
        .         .         4150         .         .         .         .         4200
CACCGCTCTCGGGACAATTCCAGTACATCTCTACTATCAGAGGGCGGATGCACGCGCGCGGTTGCGTTTTGGCGCGACGTGGCGT
HisArgSerGlyThrIleProValHisLeuTyrTyrGlnArgAlaAspAlaArgAlaArgLeuArgPheGlyAlaThrTrpArg
        .         .         .         4250         .         .         .         ***
GTCTCTCCGAGCGATCGTTTATTAAACGATCTCCGTGGCCTCATTGGTTCGGGAGCAGGTGGAACTGGAGTTTGACTAATACAGG
ValSerProSerAspArgLeuLeuAsnAspLeuArgGlyLeuIleGlySerGluGlnValGluLeuGluPheAspEnd
*   *   .  (orf37kDa)         .         .         .         .         4350
AATACTATGAGTCTGAATTTCCTTGATTTTGAACAGCCGATTGCAGAGCTGGAAGCGAAAATCGAT
          MetSerLeuAsnPheLeuAspPheGluGlnProIleAlaGluLeuGluAlaLysIleAsp
```

FIG. 3. Nucleotide sequence of *dnaE* and adjacent genes. The nucleotide sequences are shown with the corresponding amino acid sequence below them. Sequences with asterisks over them may be involved in ribosome binding, those with plus signs over them are homologous with *dnaA* boxes, and those that are underlined represent the promoter. 6-29 and 7-5 indicates the 5′-terminal nucleotide of clones pDDS6-29 and pDDS7-5, respectively.

```
               ****      *  .
       GTTCTGGAGTTAGCACAATGATCGAATTTGTT
(lpxB) ValLeuGluLeuAlaGlnEnd
              (orf23kDa) MetIleGluPheVal


               ***       *
       AGTAAACCGGAATCTGAAGATGTCTGAACCACGT
(orf23kDa) SerLysProGluSerGluAspValEnd
                       (dnaE) MetSerGluProArg
```

FIG. 4. Reading frames of lpxB and ORF₂₃ overlap with those of ORF₂₃ and dnaE. kDa, kilodaltons.

constitute the DNA-binding domain of DNA polymerase I (Table 1). Within these homologous sequences, the fraction of identical residues ranged from 16 to 30%; the relative score, which takes into account conservative amino acid changes, ranged from 64 to 71%. Other regions in α were found that had weak homology to DNA polymerase I, but these were rejected because they were not in the same linear order as they occurred within DNA polymerase I and T7 DNA polymerase. DNA polymerase I had greater homology with T7 DNA polymerase than it did with α.

DISCUSSION

In a previous study by this laboratory, the direction of transcription and the minimal size of the dnaE gene were delineated by using Bal-31 exonuclease to generate deletions from the termini of plasmids containing dnaE inserts (53). To gain insight into the possible mechanism of regulation of the dnaE gene, we sequenced a 4,350-nucleotide segment (Fig. 3) of the E. coli chromosome that contained this gene. This region was analyzed for the presence of known regulatory

sequences, ORFs, and potential secondary structures that may be of regulatory significance.

The dnaE gene codes for a 129,920-dalton protein, in agreement with the previously published molecular weight of α (140,000) (34), which was determined by using a value of 135,000 daltons for the β-galactosidase standard. Since then the molecular weight of β-galactosidase has been determined by DNA sequence analysis to be only 116,000 (12).

We calculated the isoelectric points of both the α and β subunits of DNA polymerase III holoenzyme from their amino acid composition as 4.93 and 5.04, respectively. Although the isoelectric point of α might appear to be rather acidic for a protein that interacts with DNA, we calculated the isoelectric point for DNA polymerase I to be 5.21. Since polymerases, unlike many other DNA-binding proteins, must glide along DNA, it might be advantageous for them not to bind DNA so strongly that translocation would be hindered.

The isoelectric point of α supports the accuracy of the nucleotide sequence we determined for the dnaE gene. As predicted from the amino acid sequence, both α and β migrated to similar positions on the isoelectric-focusing dimension of two-dimensional gels. That we did not make a frameshift error during the sequencing of dnaE is supported by the other long out-of-phase ORF coding more basic polypeptides. Additionally, the alternative coding regions use a greater percentage of codons corresponding to rare isoaccepting tRNAs than the same region of dnaE (data not shown). Together, these results indicate that it is unlikely that we have shifted reading frames during the sequencing of dnaE.

Recently Ollis et al. (43) showed that T7 DNA polymerase possesses several regions that have homology to the proposed DNA-binding site of E. coli DNA polymerase I. We found sections of α that demonstrated very weak homology
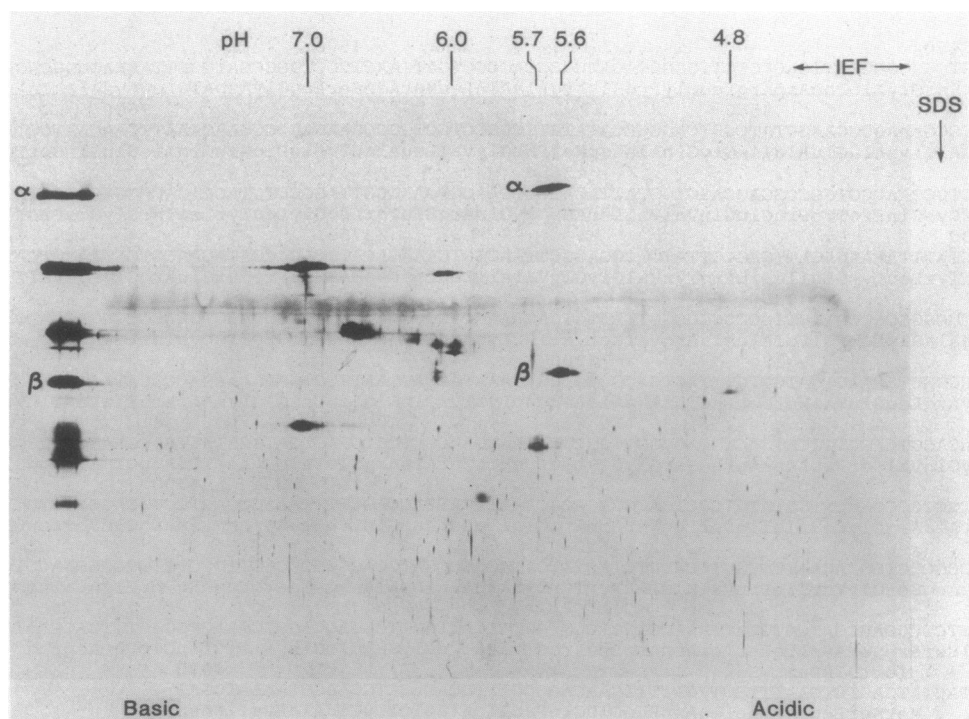

FIG. 5. Separation of DNA polymerase III holoenzyme subunits by two-dimensional polyacrylamide gel electrophoresis. The left lane contains purified holoenzyme subunits fractionated on a sodium dodecyl sulfate (SDS)–7.5 to 17.5% polyacrylamide gel. On the right, purified holoenzyme subunits (1.74 μg) were separated by two-dimensional polyacrylamide gel electrophoresis. α and β subunits are indicated. The basic end is at the left, and the acidic end is at the right, with a pH range of 4.2 to 8.8. IEF, Isoelectric focusing.

TABLE 1. Sequence homology between T7 DNA polymerase (T7 POL), alpha subunit of DNA polymerase III holoenzyme (ALPHA), and DNA polymerase I (POL I)

Protein	Sequences	Fraction identical	Relative score[a]	Polymerase I region[b]
T7 POL	LLEKLLSDKH	0.50	72.7	
POL I	LLKPLLEDEK			I
ALPHA	LFERFLNPER	0.30	71.2	
T7 POL	FNPSS RDHIQKKLQE AGWVPTKYTDKGAPVVDDEVLEGVRVDDP	0.50	68.1	
POL I	FNLSS TKQLQTILFEKQGIKPLKKTPGGAPSTSEEVLEELALDYP			II
ALPHA	YGIILYQEQVMQIAQVLSGTT LGGADMLRRAMGKKKPEEMAKQRS	0.16	65.2	
T7 POL	DASGLELRCLAH	0.58	80.5	
POL I	DYSQIELRIMAH			IV
ALPHA	ECWRMGLKILPP	0.17	70.8	
T7 POL	YIKGLDGRKVHV	0.42	77.7	
POL I	YVETLDGRRLYL			VII
ALPHA	AFDRLGPHRAAL	0.25	65.2	
T7 POL	HAALNTLLQSAGALICKLWII	0.38	77.8	
POL I	RAAINAPMQGTAADIIKRAMI			VIII
ALPHA	GDALKAADQHAKAEAIGQADM	0.32	65.0	

[a] Actual homology score as determined by the method of McLachlan (36) divided by the score for a perfectly homologous sequence.

[b] Regions of DNA polymerase I defined by Ollis et al. as being homologous to T7 DNA polymerase (43).

to five of the nine regions identified by Ollis et al. (43) in DNA polymerase I. Inclusion of the other potential reading frames in this analysis did not improve the homology. Although the sequences of α and DNA polymerase I are different, this does not exclude the possibility that their secondary and tertiary structures may be similar.

We identified a sequence with homology to *E. coli* promoters 5' to the *dnaE* gene. An earlier study in this laboratory used Bal-31 exonuclease to generate deletions from the termini of the *dnaE* gene (53). Two clones containing *dnaE*, pDDS6-29 and pDDS7-5, were identified that differed by only nine nucleotides 5' to *dnaE* (Fig. 3). This sequence was necessary for the expression of *dnaE*, as a plasmid containing this sequence, pDDS6-29, was able to both complement an *E. coli* strain containing a temperature-sensitive defect in *dnaE* and direct the synthesis of α in the maxicell system, while a plasmid lacking this sequence, pDDS7-5, could do neither. The −35 region of the putative promoter is contained within the 9-nucleotide segment. These results indicate that the promoter we have identified from sequence data can function in vivo.

In addition to *dnaE*, the segment we sequenced contains three other ORFs. The first was identified as the 3'-terminal portion of *lpxB* from comparison with the nucleotide sequence of *lpxB* (9). The *lpxB* gene encodes lipid A disaccharide synthase, which catalyzes the synthesis of the disaccharide core of an outer membrane lipopolysaccharide (8). Lipid A disaccharide synthase is an essential enzyme; cells possessing a mutation in *lpxB* are temperature sensitive for growth and cell division (39).

Distal to *lpxB* is an ORF, ORF$_{23}$, capable of coding for a 23,228-dalton protein with an isoelectric point of 7.11. Plasmids with a 3.6-kb segment of DNA containing *lpxB*, ORF$_{23}$, and the 5'-terminal region of *dnaE* were able to direct the synthesis of a protein of the size predicted for ORF$_{23}$, while those lacking ORF$_{23}$ did not (see Fig. 6, lanes pCR9 and pDC29, in reference 8). This protein has sequence homology with the β recombination protein of bacteriophage λ (18). Amino acids 16 to 55 of the β recombination protein have homology to amino acids 123 to 158 of ORF$_{23}$ (relative score of 65%). The β recombination protein promotes the renaturation of complementary single-stranded DNA during λ infection (46).

Immediately following *dnaE* is an ORF, ORF$_{37}$, encoding

a 37,000-dalton protein. Additional evidence supports the notion that this region does encode this protein. (i) Analysis of proteins encoded by plasmids containing ORF$_{37}$ by the maxicell technique reveal the synthesis of a 37,000-dalton protein (see Fig. 4, lanes C and D, in reference 60). Plasmids lacking ORF$_{37}$ are unable to direct the synthesis of a protein of this size. (ii) Both α and the 37,000-dalton protein were overproduced when the λ p_L promoter was inserted upstream from *dnaE* (31), indicating that sequences downstream from *dnaE* coded for a 37,000-dalton protein that was transcribed clockwise on the *E. coli* chromosome, the same direction as *dnaE*. Thus, we conclude that the region we have sequenced has the following organization: 5'-*lpxB*-ORF$_{23}$-*dnaE*-ORF$_{37}$-3'.

Since genes adjacent to *dnaE* are transcribed in the same direction, it is possible that they reside on a common operon. Evidence supporting this hypothesis was derived from the nucleotide sequence. The entire 4,350-nucleotide sequence lacks G+C-rich regions of dyad symmetry followed by a string of T's characteristic of rho-independent transcription termination signals (45, 47). Although we have not ruled out the presence of rho-dependent termination, the absence of an obvious termination signal is consistent with these genes being in an operon. Second, the termination codons of *lpxB* and ORF$_{23}$ overlap the initiation codons of ORF$_{23}$ and *dnaE*, respectively. Those genes that overlap in the *gal* and *trp* operons are translationally coupled (44, 52, 57). We are not aware of any case in which overlapping genes exist on separate *E. coli* transcripts.

The operon that contains the genes coding for the σ subunit of RNA polymerase, the DNA primase, and the S21 ribosomal protein has been referred to as the macromolecular synthesis operon because of their role in the synthesis of macromolecules (7, 30). Since both lipid A disaccharide synthase and α are also involved in the synthesis of macromolecules, we refer to this putative operon as macromolecular synthesis II. Studies are in progress to test this hypothesis by several techniques, including transcription mapping experiments.

α is present at 10 to 20 copies per cell (62). To determine how α is maintained at this low level, we examined the sequence for potential regulatory sequences. There were two regions, in opposite orientations, located at nucleotides 180 and 207 which had sequence homology (eight of nine

nucleotides) with *dnaA* boxes located within *oriC* and in the promoter region of the *dnaA* operon (2, 13). Those located near the promoters for *dnaA* are involved in the autoregulation of *dnaA* expression (2, 4, 16). Flower and McHenry (11) have reported the presence of sequences with homology to *dnaA* boxes in the promoter region of the *E. coli dnaZX* gene, which codes for the γ and τ subunits of DNA polymerase III. The *dnaA* protein may regulate the expression of most of the DNA polymerase III holoenzyme genes.

We identified a promoter that is downstream from the *dnaA* boxes. This promoter is located within the proposed operon. It is not unusual for a complex operon to contain an internal promoter. For example, the macromolecular synthesis operon of *E. coli* possesses several internal promoters (29, 30, 58). Internal promoters allow the genes in this operon, 5′-*rpsU-dnaG-rpoD*-3′, to be coordinately regulated with changes in growth rate (24) and discoordinately regulated during amino acid starvation (3) and heat shock (58). The *rpsU*, *dnaG*, and *rpoD* genes encode the 30S ribosome protein S21, the DNA replication primase, and the σ subunit of RNA polymerase, respectively (7, 55). By using the internal promoter, the level of σ in the cell may be increased without changing that of S21 and primase (29, 58). Similarly, under conditions of rapid growth both the *lpxB* gene product and α are needed. Since α has been implicated in the SOS response (5, 16), only the level of α may need to be increased during times when the chromosome has been damaged and cell division has been inhibited.

Since secondary structures in DNA and RNA have been shown to be of regulatory importance, we searched the region proximal to *dnaE* for sequences capable of forming such structures. There were two sequences which could form hairpin-loop structures within regions involved in the expression of ORF$_{23}$ and *dnaE*. The first was located from nucleotides 176 to 212 and contained the *dnaA* boxes. Using Tinoco's rules (59), we calculate that this hairpin-loop structure in single-stranded RNA would have a $\Delta G^0_{formation}$ of -17 kcal/mol. Whether the *dnaA* protein can bind to this site and either terminate transcription or halt translation remains to be determined. The sequence from nucleotides 613 to 632 was also capable of forming a hairpin-loop structure. Since these sequences contain the -10 box of the promoter, they may be of regulatory importance.

We have found sequences that may affect the translation of *dnaE* mRNA. According to the rules proposed for determining the strength of a translational initiation site (57), both *dnaE* and ORF$_{23}$ possess weak ribosome-binding sites. Both *dnaE* and ORF$_{23}$ have Shine-Dalgarno sequences (GGAG and GGA, respectively) and an AUG initiation codon. However, in both *dnaE* and ORF$_{23}$ the region between the Shine-Dalgarno sequence and the initiation codon contains three C and G nucleotides. Ideally this region should consist of only A and U nucleotides. Also, the initiation codon of these genes is not followed by either GCUA or AAAA, sequences characteristic of highly expressed genes (57). The presence of a weak ribosome-binding site could reduce the translation of this transcript. The initiation codon for *dnaE* was flanked by direct repeats of the sequence UCUGAA that could constitute a recognition sequence for a regulatory protein. The binding of a protein at this site on the mRNA could cause a decrease in translation.

It has recently become apparent that codon usage can affect the translation of an mRNA (for a review, see reference 22). Genes that encode abundant proteins use codons which correspond to the most prevalent tRNAs; those genes encoding poorly expressed proteins use codons correspond-

ing to the rare tRNAs. Both *dnaE* and ORF$_{23}$ use a significant number of rare codons. The codon usage of these genes is similar to that of *dnaG*, which encodes a primase, a protein present at approximately 50 copies per cell (27). Approximately half of the codons used by these genes correspond to rare tRNAs. The fraction of abundant codons (22) (abundant codons divided by total codons), used by *dnaG*, ORF$_{23}$, and *dnaE* was 0.47, 0.51, and 0.54, respectively. Codon usage could contribute to the low level of α in the cell.

The results from analysis of the nucleotide sequence of *dnaE* and surrounding genes provide testable hypotheses pertaining to the regulation of the synthesis of α. Since the *dnaA* protein may be involved in the regulation of the synthesis of many replication proteins, the effect of the *dnaA* protein on *dnaE* expression will be examined. To determine whether *dnaE* is in an operon with *lpxB*, we are mapping the termini of their transcripts. We are also examining conditions under which the internal promoter may be used to transcribe *dnaE* independently of *lpxB*. The occurrence of the genes encoding these two essential genes in an operon may have broad implications regarding the coordination of membrane biosynthesis and DNA replication with cell growth and division.

LITERATURE CITED

1. **Atkinson, T., and M. Smith.** 1984. Solid-phase synthesis of oligodeoxyribonucleotides by the phosphite-triester method, p. 35–81. *In* M. J. Gait (ed.), Oligonucleotide synthesis, a practical approach. IRL Press, Washington, D.C.
2. **Atlung, T., E. S. Clausen, and F. G. Hansen.** 1985. Autoregulation of the *dnaA* gene of *Escherichia coli* K12. Mol. Gen. Genet. **200**:442–450.
3. **Blumenthal, R. M., P. G. Lemaux, F. C. Neidhart, and P. P. Dennis.** 1976. The effects of the *relA* gene on the synthesis of aminoacyl-tRNA synthetases and other transcription and translation proteins in *Escherichia coli* B. Mol. Gen. Genet. **149**:291–296.
4. **Braun, R. E., K. O'Day, and A. Wright.** 1985. Autoregulation of the DNA replication gene *dnaA* in *E. coli* K-12. Cell **40**:159–169.
5. **Bridges, B. A., and R. P. Mottershead.** 1978. Mutagenic DNA repair. VIII. Involvement of DNA polymerase III in constitutive and inducible mutagenic repair after ultraviolet and γ irradiation. Mol. Gen. Genet. **162**:35–41.
6. **Burgers, P. M. J., A. Kornberg, and Y. Sakakibara.** 1981. The *dnaN* gene codes for the β subunit of DNA polymerase III holoenzyme of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **78**:5391–5395.
7. **Burton, Z. F., C. A. Gross, K. K. Watanabe, and R. R. Burgess.** 1983. The operon that encodes the σ subunit of RNA polymerase also encodes ribosome protein S21 and DNA primase in *E. coli* K12. Cell **32**:335–349.
8. **Crowell, D. N., M. S. Anderson, and C. R. H. Raetz.** 1986. Molecular cloning of genes for lipid A disaccharide synthase and UDP-*N*-acetylglucosamine acyltransferase in *Escherichia coli*. J. Bacteriol. **168**:335–349.
9. **Crowell, D. N., W. S. Reznikoff, and C. R. H. Raetz.** 1987. Nucleotide sequence of the *Escherichia coli* gene for lipid A

disaccharide synthase. J. Bacteriol. **169**:5727–5734.

10. **Echols, H., C. Lu, and P. M. J. Burgers.** 1983. Mutator strains of *Escherichia coli*, *mutD* and *dnaQ*, with defective exonucleolytic editing by DNA polymerase III holoenzyme. Proc. Natl. Acad. Sci. USA **80**:2189–2192.

11. **Flower, A. M., and C. S. McHenry.** 1986. The adjacent *dnaZ* and *dnaX* genes of *Escherichia coli* are contained within one continuous open reading frame. Nucleic Acids Res. **14**:8091–8101.

12. **Fowler, A. V., and I. Zabin.** 1978. Amino acid sequence of β-galactosidase. XI. Peptide ordering procedures and the complete sequence. J. Biol. Chem. **253**:5521–5525.

13. **Fuller, R. S., B. E. Funnell, and A. Kornberg.** 1984. The *dnaA* protein complex with the *E. coli* chromosomal replication origin (*oriC*) and other DNA sites. Cell **38**:889–900.

14. **Garoff, H., and W. Ansorge.** 1981. Improvement on DNA sequencing gels. Anal. Biochem. **115**:450–457.

15. **Gefter, M. L., Y. Hirota, T. Kornberg, J. A. Weshler, and C. Barnoux.** 1971. Analysis of DNA polymerase II and III in mutants of *Escherichia coli* thermosensitive for DNA synthesis. Proc. Natl. Acad. Sci. USA **68**:3150–3153.

16. **Hagensee, M., T. Timme, S. Bryan, and R. Moses.** 1987. DNA polymerase III of *Escherichia coli* is required for UV and ethyl methanesulfonate mutagenesis. Proc. Natl. Acad. Sci. USA **78**:5391–5395.

17. **Hansen, F. G., E. B. Hansen, and K. von Meyenberg.** 1982. The nucleotide sequence of the *dnaA* gene and the first part of the *dnaN* gene of *Escherichia coli* K-12. Nucleic Acids Res. **10**:7373–7385.

18. **Hendrix, R. W., J. W. Roberts, F. W. Stahl, and R. A. Weisberg (ed.).** 1983. Lambda II, p. 498–499. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

19. **Henson, J. H., H. Chu, C. A. Irwin, and J. R. Walker.** 1979. Isolation and characterization of *dnaX* and *dnaY* temperature-sensitive mutants of *Escherichia coli*. Genetics **92**:1041–1059.

20. **Horiuchi, T., H. Maki, M. Maruyama, and M. Sekiguchi.** 1981. Identification of the *dnaQ* gene product and location of the structural gene for RNase H of *Escherichia coli* by cloning of the genes. Proc. Natl. Acad. Sci. USA **78**:3770–3774.

21. **Hübscher, U., and A. Kornberg.** 1979. The δ subunit of *Escherichia coli* DNA polymerase III holoenzyme is the *dnaX* gene product. Proc. Natl. Acad. Sci. USA **76**:6284–6288.

22. **Hübscher, U., and A. Kornberg.** 1980. The *dnaZ* protein, the γ subunit of DNA polymerase III holoenzyme of *Escherichia coli*. J. Biol. Chem. **255**:11698–11703.

23. **Ikemura, T.** 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. **2**:13–34.

24. **Iwakura, Y., and A. Ishihama.** 1975. Biosynthesis of RNA polymerase in *Escherichia coli* II. Mol. Gen. Genet. **142**:67–84.

25. **Johanson, K. O., and C. S. McHenry.** 1984. Adenosine 5'-*O*-(3-thiotriphosphate) can support the formation of an initiation complex between the DNA polymerase III holoenzyme and primed DNA. J. Biol. Chem. **259**:4589–4595.

26. **Kodairi, M., S. Biswas, and A. Kornberg.** 1983. The *dnaX* gene encodes the DNA polymerase III holoenzyme τ subunit, precursor of the γ subunit. Mol. Gen. Genet. **192**:80–86.

27. **Konigsberg, W., and G. N. Godson.** 1983. Evidence for use of rare codons in the *dnaG* gene and other regulatory genes of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **80**:687–691.

28. **Lother, H., R. Kolling, C. Kucherer, and M. Schsuzu.** 1985. *dnaA* protein-regulated transcription: effects on the *in vitro* replication of *Escherichia coli* minichromosomes. EMBO J. **4**:555–560.

29. **Lupski, J. R., A. A. Ruiz, and G. N. Godson.** 1984. Promotion, termination and anti-termination in the *rpsU-dnaG-rpoD* macromolecular synthesis operon of *E. coli* K-12. Mol. Gen. Genet. **195**:391–401.

30. **Lupski, J. R., B. L. Smiley, and G. N. Godson.** 1983. Regulation of the *rpsU-dnaG-rpoD* macromolecular synthesis operon and the initiation of DNA replication in *Escherichia coli* K-12. Mol. Gen. Genet. **189**:48–57.

31. **Maki, H., T. Horiuchi, and A. Kornberg.** 1985. The polymerase subunit of DNA polymerase III of *Escherichia coli*. I. Amplification of the *dnaE* gene product and polymerase activity of the

32. **McClure, W. R.** 1985. Mechanisms and control of transcription initiation in prokaryotes. Annu. Rev. Biochem. **54**:171–204.

33. **McHenry, C. S.** 1985. DNA polymerase III holoenzyme of *Escherichia coli*: components and function of a true replicative complex. Mol. Cell. Biochem. **66**:71–85.

34. **McHenry, C. S., and W. Crow.** 1979. DNA polymerase III of *Escherichia coli*. J. Biol. Chem. **254**:1748–1753.

35. **McHenry, C. S., and A. Kornberg.** 1977. DNA polymerase III holoenzyme of *Escherichia coli* purification and resolution into subunits. J. Biol. Chem. **249**:6478–6484.

36. **McLachlan, A. D.** 1971. Tests for comparing related amino-acid sequences. Cytochrome C and cytochrome C$_{557}$. J. Mol. Biol. **61**:409–424.

37. **Miller, J. H.** 1972. Experiments in molecular genetics, p. 431–435. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

38. **Minden, J. S., and K. J. Marians.** 1985. Replication of pBR322 DNA *in vitro* with purified proteins: requirement for topoisomerase I in the maintenance of template specificity. J. Biol. Chem. **260**:9316–9325.

39. **Mullin, D. A., C. L. Woldringh, J. M. Henson, and J. R. Walker.** 1983. Cloning of the *Escherichia coli* *dnaZX* region and identification of its products. Mol. Gen. Genet. **192**:73–79.

40. **Nishijima, M., C. E. Bulana, and C. R. H. Raetz.** 1981. Two interacting mutations causing temperature-sensitive phosphatidylglycerol synthesis in *Escherichia coli* membranes. J. Bacteriol. **145**:113–121.

41. **Normark, S., S. Bergstrom, T. Edlund, T. Grundstrom, B. Jaurin, F. P. Lindberg, and O. Olson.** 1983. Overlapping genes. Annu. Rev. Genet. **17**:499–525.

42. **O'Farrell, P. H.** 1975. High-resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. **250**:4007–4021.

43. **Ollis, D. L., C. Kline, and T. A. Steitz.** 1985. Domain of *E. coli* DNA polymerase I showing sequence homology to T7 DNA polymerase. Nature (London) **313**:818–819.

44. **Oppenheim, D. S., and C. Yanofsky.** 1980. Translational coupling during expression of the tryptophan operon of *Escherichia coli*. Genetics **95**:785–795.

45. **Platt, T.** 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. **55**:339–372.

46. **Radding, C. M., J. Rosenweig, F. Richards, and E. Cassuto.** 1971. Separation and characterization of exonuclease, β protein and a complex of both. J. Biol. Chem. **246**:2510–2518.

47. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.

48. **Sancar, A., A. M. Hack, and W. D. Rupp.** 1979. Simple method for identification of plasmid-coded proteins. J. Bacteriol. **137**:692–693.

49. **Sanger, F., and A. R. Coulson.** 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. **87**:107–110.

50. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

51. **Scheuermann, R., S. Tam, P. M. J. Burgers, C. Lu, and H. Echols.** 1983. Identification of the ε subunit of *Escherichia coli* DNA polymerase III holoenzyme as the *dnaQ* gene product. A fidelity subunit for DNA replication. Proc. Natl. Acad. Sci. USA **80**:7085–7089.

52. **Schumperli, D., K. McKenney, D. A. Sobieski, and M. Rosenberg.** 1982. Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon. Cell **30**:865–871.

53. **Shepard, D., R. W. Oberfelder, M. W. Welch, and C. S. McHenry.** 1984. Determination of the precise location and orientation of the *Escherichia coli* *dnaE* gene. J. Bacteriol. **158**:455–459.

54. **Shine, J., and L. Dalgarno.** 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosome RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.

55. **Smiley, B. L., J. R. Lupski, P. S. Svec, R. McMacken, and G. N. Godson.** 1982. Sequences of the *Escherichia coli* *dnaG* primase

gene and regulation of its expression. Proc. Natl. Acad. Sci. USA 79:4550–4554.

56. Staudenbauer, W. 1976. Replication of small plasmids in extracts of E. coli. Mol. Gen. Genet. 145:273–280.

57. Stormo, G. D. 1986. Translation initiation, p. 195–224. In W. Reznikoff and L. Gold (ed.). Maximizing gene expression. Buttersworth Press, Boston.

58. Taylor, W. E., D. B. Straus, A. D. Grossman, Z. F. Burton, C. A. Gross, and R. R. Burgess. 1984. Transcription from a heat-inducible promoter causes heat shock regulation of the sigma subunit of E. coli RNA polymerase. Cell 38:371–381.

59. Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla. 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) 246:40–41.

60. Welch, M. W., and C. S. McHenry. 1982. Cloning and identification of the product of the dnaE gene of Escherichia coli. J. Bacteriol. 152:351–356.

61. Wold, M. S., J. Mallory, J. Roberts, J. LeBowitz, and R.

McMacken. 1982. λ DNA replication in vitro with purified λ replication proteins. Proc. Natl. Acad. Sci. USA 79:6176–6180.

62. Wu, Y. H., M. A. Franden, J. R. Hawker, and C. S. McHenry. 1984. Monoclonal antibodies specific for the α subunit of the Escherichia coli DNA polymerase III holoenzyme. J. Biol. Chem. 259:12117–12122.

63. Yamamoto, K., B. M. Alberts, R. Benzinger, L. Lawhorne, and G. Treiber. 1970. Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale purification. Virology 40:734–744.

64. Yannisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene 33:103–119.

65. Yin, K.-C., A. Blinkowa, and J. R. Walker. 1986. Nucleotide sequence of the Escherichia coli replication gene dnaZX. Nucleic Acids Res. 14:6541–6550.

66. Youderian, P., S. Bouvier, and M. M. Susskind. 1982. Sequence determinants of promoter activity. Cell 30:843–853.