Sequences of the *Escherichia coli dnaG* primase gene and regulation of its expression

(DNA sequence/tryptic peptides/NH₂ terminus/unusual codon usage/antitermination)

BOB L. SMILEY*, JAMES R. LUPSKI*, PAMELA S. SVEC*, ROGER MCMACKEN†, AND G. NIGEL GODSON*

*Biochemistry Department, New York University Medical Center, 550 First Avenue, New York, New York 10016; and †Department of Biochemistry, The Johns Hopkins University School of Hygiene and Public Health, 615 North Wolfe Street, Baltimore, Maryland 21205

ABSTRACT   The nucleotide sequence of a cloned section of the *Escherichia coli* chromosome containing the *dnaG* primase gene [Lupski, J., Smiley, B., Blattner, F. & Godson, G. N. (1982) *Mol. Gen. Genet.* 185, 120–128] has been determined. The region coding for the dnaG primase has been identified by NH₂-terminal and tryptic peptide amino acid analysis of the dnaG protein. The coding region is 1,740 base pairs long (580 amino acids) and is preceded by an unusual ribosome-binding site sequence (G-G-G-G). The *dnaG* gene is read in the same direction as the adjacent *rpoD* gene, but no obvious promoter sequences can be found for either gene within several hundred nucleotides upstream. Other unusual features of the *dnaG* gene that may explain the maintenance of its product at low copy number are the presence of a RNA polymerase terminator 31 nucleotides upstream from the ATG initiator codon and greater use (3–10 times) of certain codons that occur infrequently in other *E. coli* genes. The nucleotide sequence has also been correlated with data from transposon Tn5 insertional inactivation mapping.

The *Escherichia coli dnaG* gene product has long been known from genetic studies to be essential for *E. coli* chromosomal DNA replication (1). This has been confirmed by studies *in vitro* of the initiation of DNA replication. The *dnaG* gene product has been shown to be a species of RNA polymerase called primase that interacts with DNA to synthesize a RNA oligonucleotide that primes DNA synthesis (2–4). In the phage G4 *in vitro* system of DNA synthesis the dnaG primase recognizes a specific region of the single-stranded viral template and in the presence of single-strand-binding protein (SSB) synthesizes a 14- to 29-nucleotide complementary RNA, primer RNA, which is then elongated by DNA polymerase III holoenzyme to synthesize the complete G4 complementary DNA strand (5). The primase recognition site on the G4 chromosome is within a region of single-stranded DNA, the gene *F/G* intergenic region (6), that can form stable secondary structure hairpin loops, and the primer RNA always starts at the same nucleotide (7). In an analogous system, phage α3, the primase protein was shown to interact directly with these loops (8).

In the φ174 *in vitro* system, dnaG primase does not appear to recognize a specific sequence in the DNA, but is involved in a conglomerate of proteins (dnaB/dnaC/i/n/n′,n″) called a primosome (4, 9) that moves around the single-stranded DNA synthesizing a RNA primer at many different sites (10–12). Primase is believed to function similarly in priming synthesis of Okazaki fragments on the lagging DNA strand during *E. coli* chromosomal synthesis (1, 13). The dnaG protein is therefore of considerable interest as a member of the class of proteins that interact directly not only with the DNA double helix itself but also with other proteins.

In the cell the *dnaG* gene product is kept in low copy number (2) and appears to be under tight regulation. Excess amounts of dnaG protein appear to kill the cell (14). The only successful overproduction of the *dnaG* gene product so far observed is with a temperature-sensitive copy number plasmid (Q1/pRLM61) that bears the *dnaG* gene plus several thousand preceding nucleotides (14). This preceding region is believed to contain the controlling regions for *dnaG* gene expression (14, 15), and it is being proposed as a regulatory region for a macromolecular synthesis operon (unpublished) containing *dnaG* and the 3′-adjacent gene *rpoD* (14, 16). This paper reports the nucleotide sequence of the *dnaG* gene, together with peculiarities of its codon usage, its corresponding amino acid composition, and the regulatory region that immediately precedes the gene.

MATERIALS AND METHODS

Enzymes and Biochemicals. Restriction endonucleases were purchased from New England BioLabs; deoxynucleoside and dideoxynucleoside triphosphates, from P-L Biochemicals; DNA polymerase I (Klenow fragment), from Boehringer Mannheim; and phage T4 DNA ligase was a gift from Chris Schindler (New York University Medical Center). The [³²P]dATP was purchased from New England Nuclear.

Bacterial Strains and Plasmids. The plasmid pGL444, which contains the *dnaG* gene (15), was grown in *E. coli* HB101 (17), and the plasmid DNA was prepared and isolated by standard procedures (15). JM103 (18) was used to grow M13mp7 phage and its recombinants. Q1/pRLM61, which contains the temperature-sensitive copy number plasmid pRLM61 that bears the *dnaG* gene, was used as a source of dnaG protein.

DNA Sequence Analysis. The nucleotide sequence was obtained by using the Sanger dideoxynucleotide chain termination method (19) and single-stranded DNA from clones of the filamentous phage cloning vector M13mp7 (20) containing random restriction enzyme fragments of the HindIII/Sac I fragment of plasmid pGL444 (15). The HindIII/Sac I fragment was cleaved out of pGL444 and isolated by electrophoresis on a 0.35% low-melting-point agarose (Bethesda Research Laboratories) gel. The DNA was recovered from the gel by centrifuging the excised portion containing the fragment in an Oak Ridge tube at 40,000 rpm for 10 min in a Beckman Ti 50 rotor. The supernatant was extracted twice with phenol, washed with diethyl ether, and precipitated with ethanol. Methods for the ligation of fragments of this HindIII/Sac I DNA into M13mp7 cloning vector and the transfection and generation of single-stranded DNA for sequencing are standard and described in ref. 18. The final sequence was built up by determining the se-

quences of the first 20 random clones from each of 7 clone banks. *Rsa* I, *Alu* I, *Hae* III, *Hind*II, *Hha* I (S1 nuclease treated), and *Xho* I (filled-in) fragments were inserted into alkaline phosphatase-treated (21) *Hind*II-cleaved M13mp7 replicative form I DNA, so that all plaques contained a DNA insert. The *Sau*3A digest was inserted into *Bam*HI-cut M13mp7 vector. Strategies for turning cloned fragments around and subcloning *in situ* will be described elsewhere. The Staden computer programs (22) and programs developed by B.L.S. were used to assemble and analyze the DNA sequence.

Isolation of dnaG Protein and Peptide Analysis. Primase (dnaG) protein was isolated from cells containing the temperature-sensitive copy number plasmid Q1/pRLM61, which overproduces dnaG protein by 100-fold at 42°C (14). The dnaG protein purification procedures were a modification of those described in ref. 2. Peptide data were kindly provided by William Konigsberg in advance of publication.

RESULTS

Nucleotide Sequence. The *dnaG* gene has been shown by transposon Tn5 mapping to be contained within a 1.9-kilobase *Hind*III/*Sac* I DNA fragment on the *dnaG*-bearing recombinant plasmid pGL444 (15). The sequence of this *Hind*III/*Sac* I fragment has now been determined by using the dideoxynucleotide chain termination method (19) after inserting pieces into the filamentous phage vector M13mp7 and subcloning (20). The strategy of subcloning and sequencing is shown in Fig. 1. Seventy percent of the sequence was determined in both strands of the DNA.

The nucleotide sequence of the *dnaG* gene is shown in Fig. 2, with the numbering starting at the *Hind*III site. The start of the coding sequence, nucleotide 182, was determined from the NH$_2$-terminal amino acid analysis of the dnaG protein (see below). The coding sequence is preceded by no obvious Shine and Dalgarno ribosome-binding site sequence (23), and it is presumed that this function is performed by the G-G-G-G sequence, three nucleotides before the ATG initiation codon (see

Fig. 4). An autoradiograph of the sequencing gel across this region is shown in Fig. 3.

In the 400 nucleotides preceding the *dnaG* gene (the sequence of the region to the left of the *Hind*III site is unpublished), there are no sequences that can reasonably fit the canonical promoter sequence (T-A-T-R-A-T-R) (R, purine nucleoside) (24) or the −35 sequences (T-T-G-A-C-A) (25, 26), and it is considered that the *dnaG* promoter must be far to the left of the *Hind*III site (nucleotide 1), several hundred nucleotides upstream. Preceding the *dnaG* gene by 36 nucleotides, however, is a strong 27-base pair hyphenated palindrome (nucleotides 119–145) followed by a T-T-A-T-T-T-T sequence (see Fig. 4). This structure is a typical RNA polymerase transcription termination signal (28). The coding region of the *dnaG* gene is 1,740 nucleotides long and ends in an amber termination codon TGA at nucleotide 1,922 on the sequence (see Fig. 2). The COOH terminus is therefore 33 nucleotides beyond the *Sac* I site.

NH$_2$-Terminal and Tryptic Peptide Analysis of the dnaG Protein. The sequence of the NH$_2$ terminus of the dnaG protein was determined by automated analysis to be NH$_2$-Ala-Gly-Arg-Ile-Pro-Arg-Val-Phe-Ile-Asn-Asp-Leu-. This corresponded exactly with the amino acid sequence predicted from the nucleotide sequence following the ATG codon beginning at nucleotide 182. This result shows that the methionine residue specified by the initiation codon is cleaved off *in vivo*. The amino acid composition of 15 tryptic peptides of the dnaG protein that were isolated in pure form by HPLC are marked in Fig. 2. These are in complete agreement with compositions based on the nucleotide sequence. Approximately 40% of the structural gene was confirmed in this manner.

Codon Usage. The codon use in the *dnaG* gene is somewhat unusual for an *E. coli* protein in that in comparison with 18 other *E. coli* proteins of known sequence (29–37) it uses the codons AAT (Asn), ACA (Thr), TCG (Ser), CAA (Gln), and ATA (Ile) at a much higher relative frequency than is normal. This is particularly noticeable in the case of AAT, which is used for 15 out of 31 Asn codons (50%) compared with an average of 20% in
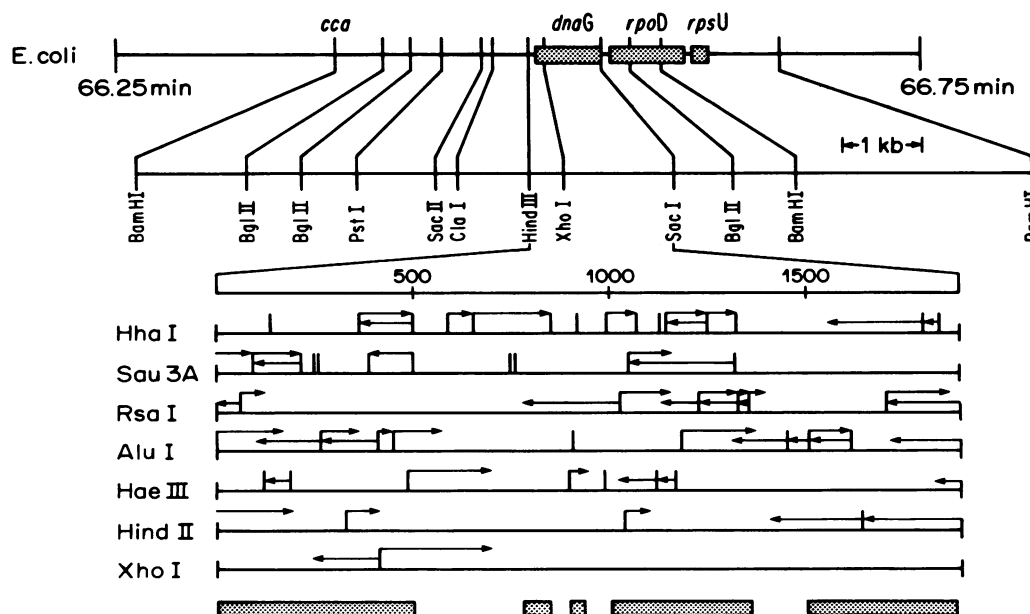
FIG. 1. Restriction map and strategy for determining the sequence of the *dnaG* gene. At the top is a genetic map of the *E. coli* chromosome (kb, kilobase). This is expanded (scale in base pairs) to give a restriction enzyme cleavage map of the chromosomal region surrounding *dnaG* (15). Six different restriction enzyme digests of the *Hind*III/*Sac* I DNA fragment from pGL444 were inserted as mixtures into M13mp7 and then cloned. The arrows show the direction of sequencing and the length of the sequence determined; vertical bars indicate the position where a restriction site for the given enzyme is found. Thick stippled bars at the bottom demonstrate the region for which the sequences of both strands were determined.

```
          10    ◆    20    ◆   30        40        50        60        70    ◆   80 ◆   90
AAGCTTCTGCAGTGAAACGTCACGCGAAGAAACTGGCTCGCGAAAACGCACGCCGCACTCGTCTGTACTAATTCCCCGAGAGCGTTGCTCTCCGATCAGA   100

            ◆
CCGAGTTGTAGTTGTAAGGCCGTGCTTCCGAAAGGAATGCGCGGCTTATTTTCGTTTATGAATTGCTAAAAATCGGGGCCTATGGCTGGACGAATCCCAC   200
                                                                                      MetAlaGlyArgIleProA

GCGTATTCATTAATGATCTGCTGGCACGCACTGACATCGTCGATCTGATCAATGCCCGTGTGAAGCTGAAAAAGCAGGGCAAGAATTTCCACGCGTGTTG   300
rgValPheIleAsnAspLeuLeuAlaArgThrAspIleValAspLeuIleAsnAlaArgValLysLeuLysLysGlnGlyLysAsnPheHisAlaCysCy

TCCATTCCACAACGAGAAAACCCCGTCCTTCACCGTTAACGGTGAGAAACAGTTTTACCACTGCTTTGGATGTGGCGCGCACGGCAACGCGATCGACTTC   400
sProPheHisAsnGluLysThrProSerPheThrValAsnGlyGluLysGlnPheTyrHisCysPheGlyCysGlyAlaHisGlyAsnAlaIleAspPhe

CTGATGAACTACGACAAGCTCGAGTTCGTCGAAACGGTCGAAGAGCTGGCAGCAATGCACAATCTTGAAGTGCCATTTGAAGCAGGCAGCGGCCCCANGC   500
LeuMetAsnTyrAspLysLeuGluPheValGluThrValGluGluLeuAlaAlaMetHisAsnLeuGluValProPheGluAlaGlySerGlyPro???P

CAGATCGAGCGCCATCAGAGGCAAACGCTTTATCAGTTGATGGACGGTCTGAATACGTTTTACCAACAATCTTTACAACAACCTGTTGCCACGTCTGCGC   600
roAspArgAlaProSerGluAlaAsnAlaLeuSerValAspGlyArgSerGluTyrValLeuProThrIlePheThrThrThrCysCysHisValCysAl

GCCAGTATCTGGAAAACGCGGATTAAGCCACGAGGTTATCGCTCGCTTTGGATTGGTTTTGCGCCCCCCGGCTGGACAACGTCTGAAGCGGTTTGGCGGC   700
aProValSerGlyLysArgGlyLeuSerHisGluValIleAlaArgPheGlyLeuValLeuArgProProAlaGlyGlnArgLeuLysArgPheGlyGly

AATCCAGAAAATCGCCAGTCATTGATTGATGCGGGGATGTTGGTCACTAACGATCAGGGACGCAGTTACGATCGTTTCCGCGAGCGGGTGATGTTCCCCA   800
AsnProGluAsnArgGlnSerLeuIleAspAlaGlyMetLeuValThrAsnAspGlnGlyArgSerTyrAspArgPheArgGluArgValMetPheProI

TTCGCGATAAACGCGGTCGGGTGATTGGTTTTGGCGGGCGCGTGCTGGGCAACGATACCCCCAAATACCTGAACTCGCCGGAAACAGACATTTTCCATAA   900
leArgAspLysArgGlyArgValIleGlyPheGlyGlyArgValLeuGlyAsnAspThrProLysTyrLeuAsnSerProGluThrAspIlePheHisLy

AGGCCGCCAGCTTTACGGTCTTTATGAAGCGCAGCAGGATAACGCTGAACCCAATCGTCTGCTTGTGGTCGAAGGCTATATGGACGTGGTGGCCGTGGCG   1000
sGlyArgGlnLeuTyrGlyLeuTyrGluAlaGlnGlnAspAsnAlaGluProAsnArgLeuLeuValValGluGlyTyrMetAspValValAlaValAla

CAATACGGCATTAATTACGCCGTTGCGTCGTTAGGTACGTCAACCACCGCCGATCACATACAACTGTTGTTCCGCGCGACCAACAATGTCATTTGCTGTT   1100
GlnTyrGlyIleAsnTyrAlaValAlaSerLeuGlyThrSerThrThrAlaAspHisIleGlnLeuLeuPheArgAlaThrAsnAsnValIleCysCysT

ATGACGGCGACCGTGCAGGCCGCGATGCCGCCTGGCGAGCGCTGGAAACGGCGCTGCCTTACATGACAGACGGCCGTCAGCTACGCTTTATGTTTTTGCC   1200
yrAspGlyAspArgAlaGlyArgAspAlaAlaTrpArgAlaLeuGluThrAlaLeuProTyrMetThrAspGlyArgGlnLeuArgPheMetPheLeuPr

TGATGGCGAAGACCCTGACACGCTAGTACGAAAAGAAGGTAAAGAAGCGTTTGAAGCGCGGATGGAGCAGGCGATGCCACTCTCCGCATTTCTGTTTAAC   1300
oAspGlyGluAspProAspThrLeuValArgLysGluGlyLysGluAlaPheGluAlaArgMetGluGlnAlaMetProLeuSerAlaPheLeuPheAsn

AGTCTGATGCCGCAAGTTGATCTGAGTACCCCTGACGGGCGCGCACGTTTGAGTACGCTGGCACTACCATTGATATCGCAAGTGCCGGGCGAAACGCTGC   1400
SerLeuMetProGlnValAspLeuSerThrProAspGlyArgAlaArgLeuSerThrLeuAlaLeuProLeuIleSerGlnValProGlyGluThrLeuA

GAATATATCTTCGTCAGGAATTAGGCAACAAATTAGGCATACTTGATGACAGCCAGCTTGAACGATTAATGCCAAAAGCGGCAGAGAGCGGCGTTTCTCG   1500
rgIleTyrLeuArgGlnGluLeuGlyAsnLysLeuGlyIleLeuAspAspSerGlnLeuGluArgLeuMetProLysAlaAlaGluSerGlyValSerAr

CCCTGTTCCGCAGCTAAAACGCACGACCATGCGTATACTTATAGGGTTGCTGGTGCAAAATCCAGAATTAGCGACGTTGGTCCCGCCGCTTGAGAATCTG   1600
gProValProGlnLeuLysArgThrThrMetArgIleLeuIleGlyLeuLeuValGlnAsnProGluLeuAlaThrLeuValProProLeuGluAsnLeu

GATGAAAATAAGCTCCCTGGACTTGGCTTATTCAGAGAACTGGTCAACACTTGTCTCTCCCAGCCAGGTCTGACCACCGGGCAACTTTTAGAGCACTATC   1700
AspGluAsnLysLeuProGlyLeuGlyLeuPheArgGluLeuValAsnThrCysLeuSerGlnProGlyLeuThrThrGlyGlnLeuLeuGluHisTyrA

GTGGTACAAATAATGCTGCCACCCTTGAAAAACTGTCGATGTGGGACGATATAGCAGATAAGAATATTGCTGAGCAAACCTTCACCGACTCACTCAACCA   1800
rgGlyThrAsnAsnAlaAlaThrLeuGluLysLeuSerMetTrpAspAspIleAlaAspLysAsnIleAlaGluGlnThrPheThrAspSerLeuAsnHi

TATGTTTGATTCGCTGCTTGAACTGCGCCAGGAAGAGTTAATCGCTCGTGAGCGCACGCATGGTTTAAGCAACGAAGAACGCCTGGAGCTCTGGACATTA   1900
sMetPheAspSerLeuLeuGluLeuArgGlnGluGluLeuIleAlaArgGluArgThrHisGlyLeuSerAsnGluGluArgLeuGluLeuTrpThrLeu

AACCAGGAGCTGGCGAAAAAGTGATTTAACGGCTTAAGTGCCGAAGAGCGATCGGGAAGCCCCGACAGCCGCACTGAGAGGCAGCGGCAAATATATAAG   2000
AsnGlnGluLeuAlaLysLys***
```
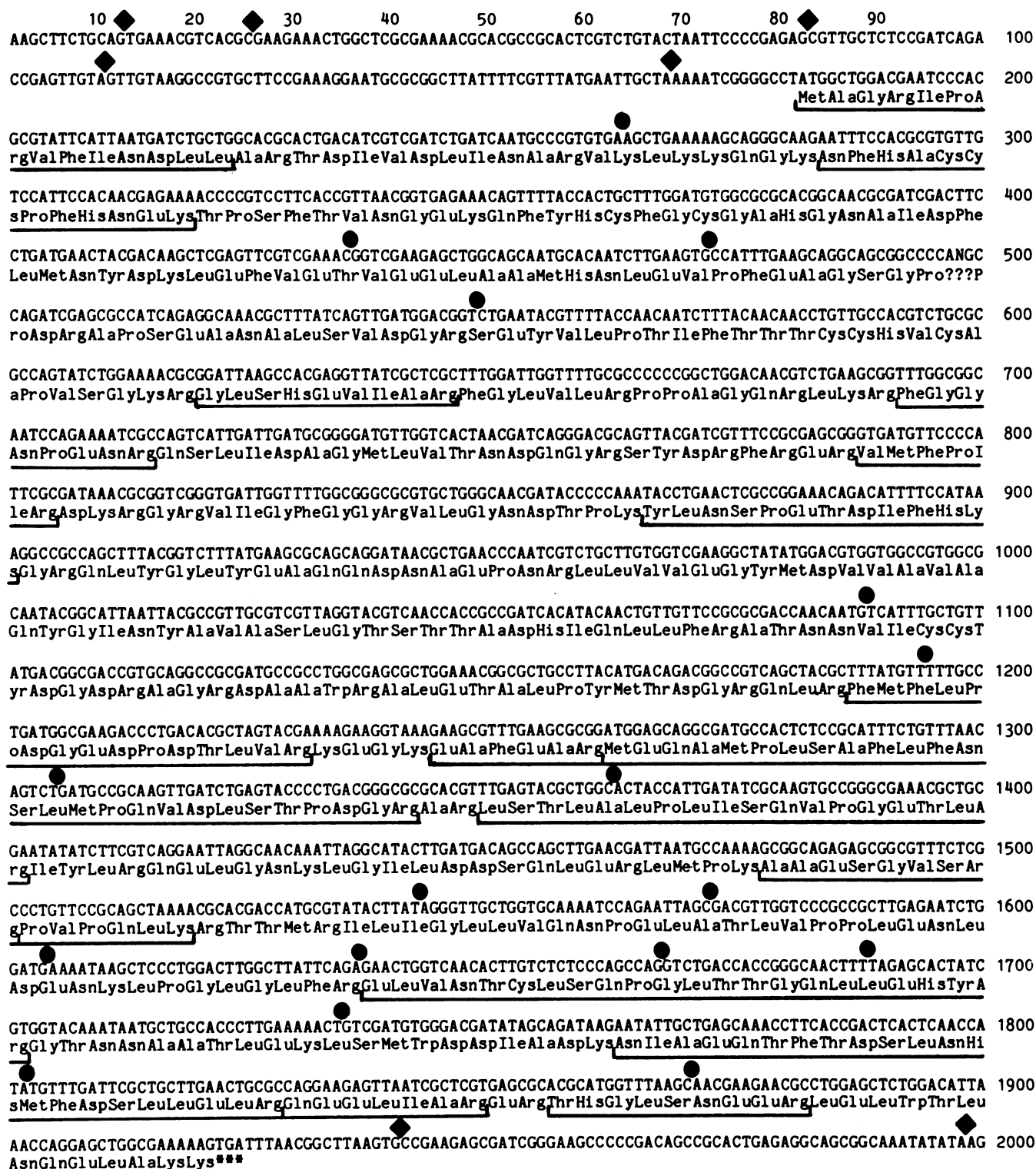
FIG. 2. Sequence of the *dnaG* gene. The nucleotides are numbered from the *Hin*dIII (A↓A-G-C-T-T) cleavage site. The ATG initiation codon is at nucleotide 182. The sequence on the 3' side of the *Sac* I site (1,888–2,000) is taken from a *Sac* I/*Bam*HI clone that contains part of the adjacent *rpoD* gene (see Fig. 1). Lines drawn under the predicted amino acid sequence depict where the tryptic peptides were determined. The ●s above the nucleotide sequence represent the approximate position of the transposon Tn*5* inserts that inactivate the *dnaG* gene, and the ◆s represent the position of Tn*5* inserts that do not inactivate the gene. The positions of these sites were calculated from the published restriction mapping data of Tn*5* insertions into *dnaG* (15).

other *E. coli* proteins, and ATA, which is used in 7 out of 22 Ile codons (33%) versus an average of 5% in other *E. coli* proteins. Six of the 15 AAT codons were confirmed by amino acid analysis of tryptic peptides. These codons are distributed at random throughout the protein.

DISCUSSION

Quality of the *dnaG* Gene Sequence. The complete nucleotide sequence of the *dnaG* gene has been determined. Over 70% of the DNA sequence was confirmed in both strands. The start of *dnaG* was fixed by determining the first 12 amino acids
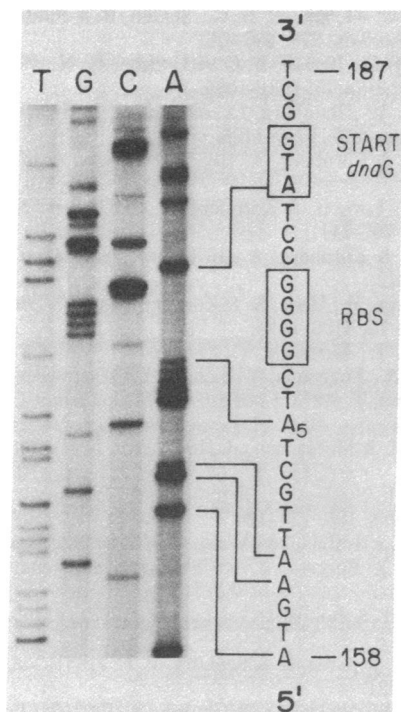
FIG. 3. Autoradiograph of the sequence containing the start of the *dnaG* gene. This is to illustrate clearly the evidence for the presumptive ribosome-binding site (RBS) sequence preceding the ATG initiation codon of the *dnaG* gene. This sequence has also been confirmed on the opposite DNA strand.

of the NH₂ terminus of purified primase protein. Over 40% of the nucleotide sequence has been confirmed by amino acid analysis of tryptic peptides. These peptides are spread randomly throughout the gene and corroborate the reading frame. The reading frame was also monitored by comparing the codon use in each of the three possible reading frames. An analysis of codon usage in 18 sequenced *E. coli* genes showed that the codons AAT (Asn), ACA (Thr), TCG (Ser), CAA (Gln), ATA (Ile), and AGG (Arg) occurred much less frequently (⅓ to ¹⁄₁₀) in the coding frame than in the two noncoding frames. This codon pref-

erence has been proposed as a method for identifying the correct reading frame when no corroborative amino acid sequence data are available (38, 39). The *dnaG* gene sequence follows this rule, despite the overall higher incidence of infrequently used codons (see below). However, in order to preserve this pattern, an N has been inserted at nucleotide number 497 and a choice of A-A-A-A made at nucleotides 613–616 instead of a possible A-C-A-A-A. The spacing of the A-A-A-A was irregular and the presence of a C was ambiguous. In the chosen reading frame, there are no infrequently used codons between nucleotides 496 and 621, but in the possible reading frame (without the N and with the C) there are 10 such codons, one being AGG (Arg), which has never been observed to occur in any *E. coli* gene so far recorded.

The dnaG Protein. The dnaG protein has two strongly basic regions: at the NH₂ terminus there are 13 basic amino acids (Arg, Lys, His) in the first 60 residues (i.e., 22%) and in the region of amino acids 145–214 there are 17 basic residues out of 70 (i.e., 24%). The rest of the polypeptide chain is composed of approximately 13% basic amino acids. The acidic amino acids (Asp, Glu) are distributed fairly randomly except for a strongly acidic COOH terminus, where 14 out of the last 60 amino acids are acidic (23%), compared with an average of 12% throughout the protein. The dnaG protein is also rather high in cysteine residues (10 Cys) compared with most other known *E. coli* proteins, and six of these are present as Cys-Cys doublets (residues 39-40, 135-136, and 305-306) which again is unusual for an *E. coli* protein, though not for a eukaryotic protein. Two of these Cys-Cys doublets were confirmed in an amino acid analysis of tryptic peptides (see Fig. 2).

Regulation of *dnaG* Gene Expression. The copy number of the dnaG protein is kept low in the cell, as might be expected for a DNA initiation protein, and several features of the gene might account for this. The gene is not preceded by a recognizable Shine and Dalgarno (23) ribosome-binding site and it is probable that the G-G-G-G sequence that precedes the ATG initiation codon by three nucleotides serves this function. This, however, may well be an inefficient ribosome-binding site and could exert a translational control by lowering the frequency of ribosome binding and initiation of translation. A second possible translational control could come from the high incidence of infrequently used codons that occur in the gene. The AAT (Asn),
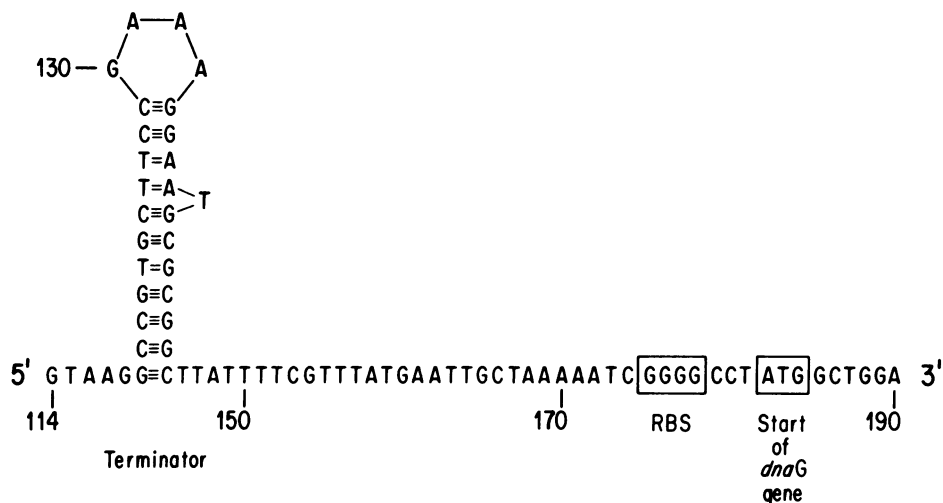

FIG. 4. 5' sequence preceding the coding sequence of the *dnaG* gene. The hyphenated palindrome is drawn as a base-paired loop. The negative free energy of formation of this structure is calculated to be −26.4 kcal/mol (110 kJ/mol), using the rules of Gralla and Crothers (27). The ATG initiation codon was confirmed by NH₂-terminal amino acid sequence analysis, but the designation of the G-G-G-G as a ribosome-binding site (RBS) is speculative, as is the designation of the base-paired loop as a terminator.

ACA (Thr), TCG (Ser), CAA (Gln), and ATA (Ile) codons are used in *E. coli* relatively infrequently compared with their synonyms. But this is not the case in the *dnaG* gene, in which the coding sequence contains 44 such codons, which is 10 times higher than the frequency observed in the average *E. coli* protein. Many of these infrequently used codons have been confirmed by amino acid composition analysis of the tryptic peptides, and one peptide of 18 amino acids (amino acids 390–407) confirmed five such codons, three of them (ATA, TCG, and CAA) occurring consecutively. Codon use within an organism is usually similar among its genes and is believed to reflect the relative abundance of isoaccepting tRNA species (40). Translational modulation by using rare tRNAs has often been suggested as a translational control mechanism, and it is possible that the *dnaG* gene uses it. A combination of unusual codons and a poor ribosome-binding site would reduce both the frequency of initiation of translation and the rate of passage of the ribosomes along the mRNA and account for the low amount of dnaG protein made compared with the amount of adjacent *rpoD* gene product (16, 41).

Another feature of the *dnaG* gene region of the chromosome that must be involved in the regulation of its expression is the presence of a RNA polymerase transcription terminator 31 nucleotides before the ATG initiation codon of the *dnaG* gene. Between this terminator and start of the gene there is no promoter-like sequence, and indeed, there is no sequence that can fit the canonical promoter sequence (T-A-T-R-A-T-R) (24) or the −35 sequence (T-T-G-A-C-A) (25, 26) for at least 400 nucleotides upstream (data not shown). There is also no obvious promoter sequence for the adjacent *rpoD* gene, either in the 197-nucleotide intergenic sequence that separates the two genes (29) or in the *dnaG* gene coding region, and the promoter for both the genes must lie far to the 5′ side of the *dnaG* gene and on the 5′ side of the terminator. This could explain the genetic results of Nakamura (41, 42) that the N protein of λ phage (an antiterminator) increases expression of the *rpoD* gene *in vivo*. Our findings coupled with those of both Nakamura and Wold and McMacken (14) suggest that there must be an *E. coli* antiterminator protein and the *dnaG* system might be used to identify it.

The *dnaG* gene product and the *rpoD* gene product (as part of RNA polymerase) are both involved in initiation of DNA replication and both perform a similar function of synthesizing a primer RNA (pRNA) for initiating DNA synthesis (5, 43). It is reasonable that the control of DNA replication can be exercised through control of expression of these two gene products, and perhaps it is no accident that the *dnaG* gene and the *rpoD* gene are next to each other on the chromosome and that their expression is tightly regulated.

1. Lark, K. G. (1972) *Nature (London) New Biol.* 240, 237–240.
2. Rowen, L. & Kornberg, A. (1978) *J. Biol. Chem.* 253, 758–764.
3. Rowen, L. & Kornberg, A. (1978) *J. Biol. Chem.* 253, 770–774.
4. McMacken, R. & Kornberg, A. (1978) *J. Biol. Chem.* 253, 3313–3319.
5. Bouche, J. P., Rowen, L. & Kornberg, A. (1978) *J. Biol. Chem.* 253, 765–769.
6. Godson, G. N., Barrell, B. G., Staden, R. & Fiddes, J. C. (1978) *Nature (London)* 276, 236–247.
7. Fiddes, J. C., Barrell, B. G. & Godson, G. N. (1978) *Proc. Natl. Acad. Sci. USA* 75, 1081–1085.
8. Benz, E. W., Rainberg, D., Vicuna, R. & Hurwitz, J. (1980) *J. Biol. Chem.* 255, 1096–1106.
9. Arai, K. & Kornberg, A. (1981) *Proc. Natl. Acad. Sci. USA* 78, 69–73.
10. Arai, K., Low, R. & Kornberg, A. (1981) *Proc. Natl. Acad. Sci. USA* 78, 707–711.
11. Arai, K. & Kornberg, A. (1979) *Proc. Natl. Acad. Sci. USA* 76, 4308–4312.
12. McMacken, R., Ueda, K. & Kornberg, A. (1977) *Proc. Natl. Acad. Sci. USA* 74, 4190–4194.
13. Okazaki, T., Kurosawa, T., Ogawa, T., Seki, K., Shinozaki, S., Hirose, A., Fugiyama, A., Kohara, Y., Machida, Y., Tamanoi, F. & Hozumi, T. (1978) *Cold Spring Harbor Symp. Quant. Biol.* 43, 203–219.
14. Wold, M. & McMacken, R. (1982) *Proc. Natl. Acad. Sci. USA* 79, in press.
15. Lupski, J., Smiley, B., Blattner, F. & Godson, G. N. (1982) *Mol. Gen. Genet.* 185, 120–128.
16. Gross, C., Hoffman, J., Wald, C., Hager, D., Burdick, G., Berger, H. & Burgess, R. (1978) *Proc. Natl. Acad. Sci. USA* 75, 427–431.
17. Boyer, H. & Roulland-Dussoix, D. (1969) *J. Mol. Biol.* 41, 459–472.
18. Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* 9, 309–321.
19. Sanger, R., Nicklen, S. & Coulson, A. R. (1971) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
20. Messing, J., Gronenborn, B., Müller-Hill, B. & Hofschneider, P. H. (1977) *Proc. Natl. Acad. Sci. USA* 74, 3642–3646.
21. Heyneker, H. L., Shine, J., Goodman, H. M., Boyer, H. W., Rosenberg, J., Dickerson, R. E., Nurang, S. A., Itakura, K., Lin, S. & Riggs, A. D. (1976) *Nature (London)* 263, 748–752.
22. Staden, R. (1977) *Nucleic Acids Res.* 4, 4037–4051.
23. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* 71, 1342–1346.
24. Pribnow, D. (1975) *Proc. Natl. Acad. Sci. USA* 72, 784–788.
25. Pribnow, D. (1975) *J. Mol. Biol.* 99, 419–443.
26. Takanami, M., Sugimoto, K., Sugisaki, H. & Okamoto, T. (1976) *Nature (London)* 260, 297–302.
27. Gralla, J. & Crothers, D. M. (1975) *J. Mol. Biol.* 73, 497–511.
28. Adhya, S. & Gottesman, M. (1970) *Annu. Rev. Biochem.* 47, 967–996.
29. Burton, Z., Burgess, R. R., Lin, J., Moore, D., Holder, S. & Gross, C. A. (1981) *Nucleic Acids Res.* 9, 2889–2903.
30. Marcoli, R., Iida, S. & Bickle, T. A. (1980) *FEBS Lett.* 110, 11–14.
31. Gay, N. J. & Walker, J. E. (1981) *Nucleic Acids Res.* 9, 3919–3927.
32. Gay, N. J. & Walker, J. E. (1981) *Nucleic Acids Res.* 9, 2187–2193.
33. Nakumura, K., Pirtle, R. M., Takeishi, K. & Inouye, M. (1980) *J. Biol. Chem.* 255, 210–216.
34. Sancar, A., Stachlek, C., Konigsberg, W. & Rupp, W. D. (1980) *Proc. Natl. Acad. Sci. USA* 77, 2611–2615.
35. Nichols, B. P., Blumberg, M. & Yanofsky, C. (1981) *Nucleic Acids Res.* 9, 1743–1755.
36. Sancar, A., William, K. R., Chase, J. & Rupp, W. D. (1981) *Proc. Natl. Acad. Sci. USA* 78, 4274–4278.
37. Beck, E. & Bremer, E. (1980) *Nucleic Acids Res.* 8, 3011–3024.
38. Staden, R. & McLachlan, A. D. (1982) *Nucleic Acids Res.* 10, 141–156.
39. Shepherd, J. C. W. (1981) *Proc. Natl. Acad. Sci. USA* 78, 1596–1600.
40. Post, L. E. & Nomura, M. (1980) *J. Biol. Chem.* 255, 4660–4666.
41. Nakamura, Y. (1980) *Mol. Gen. Genet.* 178, 487–497.
42. Nakamura, Y. & Yura, T. (1976) *Proc. Natl. Acad. Sci. USA* 73, 4405–4409.
43. Fuller, R. S., Kaguni, J. M. & Kornberg, A. (1981) *Proc. Natl. Acad. Sci. USA* 78, 7370–7374.