Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 4615–4619, June 1990
Biochemistry

Molecular cloning and DNA sequence analysis of *Escherichia coli priA*, the gene encoding the primosomal protein replication factor Y

(DNA helicase)

PEARL NURSE*, RUSSELL J. DIGATE*, KENTON H. ZAVITZ†, AND KENNETH J. MARIANS*†

*Program in Molecular Biology, Sloan–Kettering Institute, Memorial Sloan–Kettering Cancer Center, New York, NY 10021; and †Graduate Program in Molecular Biology, Cornell University Graduate School of Medical Sciences, New York, NY 10021

Communicated by Jerard Hurwitz, March 21, 1990 (received for review March 5, 1990)

ABSTRACT ***Escherichia coli*** **replication factor Y (protein n') functions in the assembly of a mobile multiprotein replication-priming complex called the primosome. Although the role of factor Y in primosome assembly during replication *in vitro* of bacteriophage ϕX174 and plasmid pBR322 DNA is clear, its role in *E. coli* chromosomal replication is not. To address this issue, the gene for factor Y has been cloned molecularly and its DNA sequence has been determined. The cloned fragment of DNA contained an open reading frame capable of encoding a polypeptide of 81.7 kDa. This open reading frame contains amino acid sequences identical to 13 N-terminal amino acids of purified factor Y, as well as to a 10-amino acid internal sequence (from a cyanogen bromide fragment) as determined by gas-phase microsequencing. Expression of the polypeptide encoded by this open reading frame using a bacteriophage T7 transient expression system resulted in the accumulation of a polypeptide with an apparent molecular mass of 78 kDa that comigrated with bona fide factor Y during SDS/polyacrylamide gel electrophoresis. Soluble extracts made from cells overexpressing the product of the putative factor Y open reading frame showed a 2000-fold increase in factor Y activity during bacteriophage ϕX174 complementary-strand DNA synthesis *in vitro* when compared to control extracts. The gene encoding factor Y, which maps to 88.5 min on the *E. coli* chromosome, has been designated primosome A (*priA*).**

The primosome (1, 2) is a mobile multiprotein DNA replication-priming apparatus that requires seven *Escherichia coli* proteins, factor Y (protein n'), proteins n and n'', and the products of the *dnaT*, *dnaB*, *dnaC*, and *dnaG* (primase) genes, for assembly on a single-stranded DNA binding protein-coated single-stranded DNA template containing a primosome assembly site (PAS). Primosome assembly begins with the specific recognition and binding of factor Y to the PAS. The multiprotein primosome, driven by the DNA helicase activity of the DnaB protein (3) in the 5' → 3' direction or the DNA helicase activity of factor Y (4, 5) in the 3' → 5' direction, can translocate and prime in both directions along the DNA template (6). When present at the replication fork with the complete primosome complex, the *E. coli* DNA polymerase III holoenzyme can catalyze both leading- and lagging-strand DNA synthesis *in vitro* at a rate approaching that estimated for replication forks *in vivo* (approximately 1000 nucleotides per sec) (7).

Factor Y was originally identified as an essential protein for primosome assembly during bacteriophage ϕX174 complementary-strand synthesis *in vitro* (8, 9). Subsequently, it was shown to be required for the initiation of lagging-strand DNA synthesis during pBR322 replication *in vitro* (10). The role of factor Y in *E. coli* replication, however, remains obscure. Factor Y is not required for DNA replication *in vitro* from the *E. coli* origin of DNA replication (*oriC*) (11) and there are no PAS sequences immediately adjacent to *oriC*, although several have been detected within several kilobase pairs of *oriC* (12, 13). These may serve as points of assembly for the primosome as the replication fork passes through them.

To investigate the role of factor Y in cellular replication and to begin a detailed mutational and structural analysis of this multifunctional protein, we have cloned molecularly, sequenced, and mapped the gene encoding factor Y.‡

MATERIALS AND METHODS

***E. coli* Strains and Bacteriophage.** Competent DH5α and DH5αF' cells for transformation were obtained from Bethesda Research Laboratories. *E. coli* strain BL21(DE3) was the gift of F. William Studier (Brookhaven National Laboratory). Bacteriophage λ 538 and λ 329 (14) were the gifts of C. Squires (Columbia University). Transformation of competent *E. coli* DH5α and DH5αF' was performed as described by the manufacturer's recommendations.

DNA. pBS-M13 (BlueScribe) was obtained from Stratagene. M13Mp18 and M13Mp19 form I (supercoiled) DNAs were obtained from Bethesda Research Laboratories. Genomic DNA was prepared from *E. coli* strain HMS-83 (*rha*, *lys*, *thy*, *polA1*, *polB100*, *lacZ*am, *str*^{R}). Plasmid pET-3c was the gift of F. William Studier.

Enzymes and Reagents. Restriction enzymes were from New England Biolabs, Bethesda Research Laboratories, and Boehringer Mannheim. Sequenase was from United States Biochemical. *Thermus aquaticus* polymerase was from Stratagene. Acrylamide, Zeta-Probe membrane, and protein determination kit were from Bio-Rad. SeaKem ME agarose was from FMC. Immobilon was from Millipore.

DNA Sequencing. DNA fragments were cloned molecularly into M13Mp18 and M13Mp19 replicative form DNAs. Bacteriophage DNA was isolated and sequenced using Sequenase and *Thermus aquaticus* polymerase in the presence of dGTP and dITP and in the presence and absence of single-stranded DNA binding protein as suggested by the manufacturer. The products of DNA sequencing reactions were separated through 6% polyacrylamide gels containing 50% (wt/vol) urea and using 100 mM Tris borate/1 mM EDTA, pH 8.0, as the running buffer. Gels were dried before autoradiography.

Southern Analysis. Southern analysis was performed using a VacuGene apparatus (Pharmacia LKB Biotechnology). Transfer to Zeta-Probe was carried out in 0.4 M NaOH. All other procedures were as described (15) except that 2× SST (0.3 M NaCl/0.03 M Tris·HCl, pH 7.5, at 22°C/0.005 M

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ORF, open reading frame; PAS, primosome assembly site.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. M33881).

EDTA) was used in place of 2× SSC (1× SSC = 0.15 M NaCl/0.015 M sodium citrate, pH 7.0).

Amino Acid Microsequencing. *E. coli* factor Y was purified as described (16). N-terminal microsequencing of intact factor Y was performed at The Biomedical Resource Center, University of California at San Francisco (courtesy of J. Kealy). Additional sequence information was obtained after cyanogen bromide cleavage. Factor Y (20 μg) was digested with cyanogen bromide in 30% (vol/vol) formic acid for 18 hr at 23°C. The reaction was terminated by increasing the reaction volume 2.5-fold with 50 mM Tris·HCl (pH 8.4 at 30°C). The products were lyophilized, resuspended in distilled H_2O, and neutralized with 10 M NaOH. Samples were electrophoresed through a 15% polyacrylamide gel containing 0.1% SDS in Tris/glycine/SDS buffer (29). The gel was preelectrophoresed with buffer containing 0.2 mM thioglycolic acid. Protein was transferred to poly(vinylidene difluoride) (Immobilon) membrane as described by Matsudaira (17). Polypeptide fragments of approximately 40, 28, and 20 kDa were identified by Coomassie blue staining, excised, and used for microsequence analysis (courtesy of W. Lane, Harvard University).

RESULTS

Molecular Cloning of the Gene Encoding Factor Y. A 14-amino acid sequence derived from the N-terminal end of intact factor Y and a 10-amino acid sequence derived from the N-terminal end of the 28-kDa isolated cyanogen bromide fragment of factor Y were used to design two "best guess" oligonucleotide probes based on the codon usage frequency in *E. coli* (18) and allowing G·T base pairing in the wobble position (Fig. 1). After sequence analysis of the isolated gene (as described below), the best guess oligonucleotide sequence for the N terminus of the intact protein was shown to be 74% correct, and that for the internal sequence was 88% correct. If one assumes that G·T base pairing neither strengthens nor weakens the DNA duplex, then the N-terminal sequence was 83% correct and the internal sequence was 93.5% correct.

Polymerase chain reaction amplification (19), using the best guess oligonucleotide for the N terminus and the complementary sequence of the best guess oligonucleotide for the internal sequence as primers and *E. coli* HMS-83 genomic DNA as template, resulted in the accumulation of a DNA fragment of approximately 1.4 kilobase pairs (kbp). Southern blots of genomic DNA digested with restriction endonucleases possessing 4-base and 6-base recognition sequences were probed with the best guess oligonucleotides, as well as with intact polymerase chain reaction-amplified DNA. A 2.9-kbp fragment was identified in *Sma* I-digested genomic DNA that hybridized at high stringency to all three probes.

A.	- Asp Ala His Val Ala Leu Pro Val Pro Leu Pro Thr Thr Phe Val
B.	GCT CAT GTT GCT CTG CCG GTT CCG CTG CCG ACT ACT TTT
C.	GCC CAC GTT GCC TTG CCC GTT CCG CTT CCT CGT ACC TTT
D.	Arg Gln His Leu Gln Ala Asp Asn Gln Val
E.	ATG CGT CAG CAT CTG CAG GCT GAT AAC CAG GTT
F.	ATG CGC CAG CAT TTA CAG GCT GAT AAC CAG GTC

FIG. 1. Amino acid sequences and probes of *priA*. (*A*) N-terminal sequence as determined by microsequence analysis. (*B*) Sequence of the best-guess oligonucleotide from the N terminus. (*C*) Nucleotide sequence of the N terminus of the cloned *priA* gene. (*D*) Sequence of the N terminus of the cyanogen bromide fragment as determined by microsequence analysis. (*E*) Sequence of the best-guess oligonucleotide from the cyanogen bromide fragment. (*F*) Nucleotide sequence of the N terminus of the cyanogen bromide fragment from the cloned *priA* gene.

HMS-83 genomic DNA (80 μg) was digested with the *Sma* I restriction endonuclease and fractionated through an agarose gel. The region of the gel surrounding the 2.9-kbp DNA fragment was excised, and the DNA was electroeluted and recovered by ethanol precipitation. This mixture of DNA fragments was ligated with *Sma* I-digested pBS-M13 replicative form III DNA (Bluescribe) and used to transform competent *E. coli* DH5α cells.

Recombinant plasmid DNAs were isolated from cells grown from 34 white colonies. The plasmid DNA was electrophoresed through agarose gels, blotted, and probed with radiolabeled oligonucleotide comprised of the internal sequence. One positive signal was obtained. This clone contained the 2.9-kbp fragment.

The entire 2.9-kbp fragment was subcloned into M13mp18 and M13mp19 replicative form DNA and the insert region in the recombinant M13 single-stranded DNAs was sequenced in both directions using a series of oligonucleotide primers (Fig. 2). The 2.9-kbp *Sma* I DNA fragment contained an open reading frame (ORF) extending from nucleotide 406 to nucleotide 2602. The predicted molecular mass, 81.7 kDa, of the polypeptide that could be translated from the sequence of this ORF was in good agreement with the observed molecular mass of 78 kDa for factor Y (9, 16). An analysis of the DNA sequence upstream of the ORF revealed a Shine–Dalgarno sequence at nucleotides 395–400, a Pribnow box at nucleotides 369–374, and −35 promoter sequence at nucleotides 354–359.

To obtain evidence that this ORF encoded factor Y, a DNA fragment carrying the ORF was inserted into the pET-3c transient expression vector (20) such that it came under the control of the T7 ϕ10 promoter (unpublished data). This construct was transformed into *E. coli* BL-21(DE3) cells that harbor a λ lysogen carrying the T7 RNA polymerase gene under the control of the *lac* UV5 promotor. These cells also harbor a second plasmid, pLys-S, that expresses a low level of T7 lysozyme, a natural inhibitor of T7 RNA polymerase, to control leakage from the *lac* UV5 promotor before induction. Treatment of these cultures with 0.4 mM isopropyl β-D-thiogalactopyranoside resulted in the time-dependent accumulation of a polypeptide with an apparent molecular mass of 78 kDa that comigrated with bona fide factor Y during SDS/PAGE (Fig. 3*A*, lanes 2–7).

Most of the 78-kDa polypeptide was present in the soluble fraction [compare Fig. 3*A*, lane 10 (soluble fraction) with lane 11 (insoluble cell debris)]. The overproduction of this polypeptide was correlated with a 2000-fold increase over control extracts of factor Y activity as scored by the bacteriophage ϕX174 single-stranded circular → replicative form DNA replication assay reconstituted with purified proteins (Fig. 3*B*). Incorporation of nucleotide into acid-insoluble product was completely dependent on the presence of the ϕX174 PAS sequence DNA in the template and on the inclusion of the DnaT protein in the reaction mixture (data not shown), both diagnostic of factor Y-dependent primosome assembly (8, 9).

The putative factor Y ORF contains amino acid sequences identical to 13 N-terminal amino acids of purified factor Y, as well as to a 10-amino acid internal sequence from the cyanogen bromide fragment, as determined by gas-phase microsequencing (Figs. 1 and 2). This finding and the correlation of the overexpression of the 78-kDa polypeptide with the dramatic induction of factor Y replication activity strongly supports the conclusion that this ORF constitutes the gene for factor Y.

The Gene for Factor Y Is Located at 88.5 min on the *E. coli* Chromosome. Kohara *et al.* (14) have established a restriction endonuclease map of the entire *E. coli* chromosome and have constructed a bacteriophage λ library of the *E. coli* genome. It is, therefore, possible to map a gene by performing Southern blot analysis on genomic DNA using the same eight

```
   1 CCCGGGATGTTGAAACGCTTGTTGAAGCGGTCAACACGGCCACCGGTAGCAACATCACGCTGTTTGCCAGTGAAGAACGGGTGGCACTTGCTGCACACGTCGAGGTTCAG
 111 GTCATGACCAACGGTGGAGCGGATTTTCATTACGTTACCGCAAGAGCAGCTAGCAGTAATTTCTTCGTATTTCGGGTGAATATCTTTTTTCATGGGAAAACCTCGGTTTA
 221 AGGCCGCGTCGCTCTTCCAGCCCTAACGCCAGACACCACGCGATGTTAAAAGTATAGCTTCAATACGATCATTTCGTACGAAGGCGCGAAATCATACAGAAATTAACCAG
 331 CGTATGCAAACTGATCCGCACTCTTCTACGGCAATGTGTATACTAACCCACCGAATTTCAAGTCAGGATGATGCTATGCCCGTTGCCCACGTTGCCTTGCCCGTTCCGCT
                             -35         -10                       S.D.      M  P  V  A  H  V  A  L  P  V  P  L→
 441 TCCTCGTACCTTTGACTATCTGCTGCCAGAAGGCATGACGGTTAAAGCTGGGTGTCGCGTGCGCGTGCCGTTTGGCAAACAGCAGGAGCGCATCGGGATTGTGGTATCAG
      P  R  T  F  D  Y  L  L  P  E  G  M  T  V  K  A  G  C  R  V  R  V  P  F  G  K  Q  Q  E  R  I  G  I  V  V  S→
 551 TTAGCGATGCCAGCGAACTGCCGCTCAATGAGCTAAAAGCGGTAGTCGAAGTGCTGGATAGTGAGCCGGTGTTTACTCACTCCGTCTGGCGATTGCTGCTATGGGCGGCA
     V  S  D  A  S  E  L  P  L  N  E  L  K  A  V  V  E  V  L  D  S  E  P  V  F  T  H  S  V  W  R  L  L  L  W  A  A→
 661 GATTACTATCATCATCCGATTGGCGATGTGCTGTTTCATGCCTTGCCGATTTTACTACGCCAGGGGCGGCCTGCGGCGAACGCGCCGATGTGGTACTGGTTTGCCACTGA
      D  Y  Y  H  H  P  I  G  D  V  L  F  H  A  L  P  I  L  L  R  Q  G  R  P  A  A  N  A  P  M  W  Y  W  F  A  T  E→
 771 ACAAGGCCAGGCGGTGGATCTGAACAGCCTGAAACGCTCCCCCAAGCAACAACAGGCGCTGGCGGCGTTACGGCAAGGCAAAATCTGGCGCGACCAGGTCGCCACGCTCG
      Q  G  Q  A  V  D  L  N  S  L  K  R  S  P  K  Q  Q  Q  A  L  A  A  L  R  Q  G  K  I  W  R  D  Q  V  A  T  L→
 881 AATTTAATGATGCCGCGTTGCAGGCGCTACGCAAAAAAGGTCTGTGTGATTTAGCAAGTGAAACACCAGAGTTTAGCGACTGGCGAACGAACTATGCCGTTTCTGGTGAG
     E  F  N  D  A  A  L  Q  A  L  R  K  K  G  L  C  D  L  A  S  E  T  P  E  F  S  D  W  R  T  N  Y  A  V  S  G  E→
 991 CGGTTGCGATTGAATACCGAACAGGCCACCGCCGTTGGCGCAATTCATAGCGCGGCAGATACTTTTTCTGCCTGGCTGCTGGCGGGCGTTACCGGTTCCGGTAAAACGGA
      R  L  R  L  N  T  E  Q  A  T  A  V  G  A  I  H  S  A  A  D  T  F  S  A  W  L  L  A [G  V  T  G  S  G  K  T] E→
1101 GGTTTATCTCAGCGTACTGGAAAACGTGCTCGCTCAGGGCAAACAGGCGCTGGTGATGGTGCCGGAAATCGGCCTGACACCGCAAACTATCGCCCGTTTTCGTGAACGTT
      V  Y  L  S  V  L  E  N  V  L  A  Q  G  K  Q  A  L  V  M  V  P  E  I  G  L  T  P  Q  T  I  A  R  F  R  E  R→
1211 TTAATGCCCCCGTGGAAGTTCTGCATTCCGGCCTGAACGACAGCGAGCGTCTTTCGGCGTGGCTGAAAGCGAAAAATGGTGAGGCGGCGATTGTGATCGGCACCCGCTCC
     F  N  A  P  V  E  V  L  H  S  G  L  N  D  S  E  R  L  S  A  W  L  K  A  K  N  G  E  A  A  I  V  I  G  T  R  S→
1321 GCGCTGTTTACGCCGTTTAAAAATCTCGGCGTGATTGTCATTGATGAAGAGCACGACAGCTCCTACAAGCAGCAGGAAGGCTGGCGCTATCATGCCCGCGACCTGGCGGT
      A  L  F  T  P  F  K  N  L  G  V  I  V  I  D  E  E  H  D  S  S  Y  K  Q  Q  E  G  W  R  Y  H  A  R  D  L  A  V→
1431 GTATCGTGCGCACAGCGAGCAAATCCCGATTATTCTTGGCTCCGCAACGCCCGCGCTGGAAACGTTATGCAACGTCCAGCAGAAAAAATACCGCCTGCTGCGCCTGACCC
      Y  R  A  H  S  E  Q  I  P  I  I  L  G  S  A  T  P  A  L  E  T  L  C  N  V  Q  Q  K  K  Y  R  L  L  R  L  T→
1541 GTCGGGCAGGGAATGCGCGTCCGGCAATTCAACATGTGCTGGATTTAAAAGGTCAGAAGGTGCAGGCAGGTCTGGCTCCGGCGTTAATCACTCGTATGCGCCAGCATTTA
     R  R  A  G  N  A  R  P  A  I  Q  H  V  L  D  L  K  G  Q  K  V  Q  A  G  L  A  P  A  L  I  T  R  M  R  Q  H  L→
1651 CAGGCTGATAACCAGGTCATTCTCTTTCTTAACCGCCGTGGCTTTGCGCCTGCACTGCTGTGCCACGACTGTGGCTGGATTGCCGAATGCCCACGTTGCGATCACTACTA
      Q  A  D  N  Q  V  I  L  F  L  N  R  R  G  F  A  P  A  L  L  C  H  D  C  G  W  I  A  E  C  P  R  C  D  H  Y  Y→
1761 CACGCTGCATCAGGCGCAGCACCATCTGCGCTGCCACCACTGTGACAGTCAGCGTCCGGTGCCGCGCCAGTGCCCTTCCTGCGGTTCCACGCACCTGGTCCCCGTGGGGC
      T  L  H  Q  A  Q  H  H  L  R  C  H  H  C  D  S  Q  R  P  V  P  R  Q  C  P  S  C  G  S  T  H  L  V  P  V  G→
1871 TGGGCACCGAACAGCTTGAACAGACGCTCGCGCCGTTGTTCCCCGGCGTGCCCATTTCTCGTATCGACCGCGATACCACCAGCCGCAAAGGGGCGCTGGAACAGCAACTG
      L  G  T  E  Q  L  E  Q  T  L  A  P  L  F  P  G  V  P  I  S  R  I  D  R  D  T  T  S  R  K  G  A  L  E  Q  Q  L→
1981 GCAGAAGTACATCGCGGCGGCGCGCGGATTTTGATTGGTACACAAATGCTGGCGAAAGGTCACCATTTCCCGGATGTGACGCTGGTTGCATTACTGGACGTGGACGGCGC
      A  E  V  H  R  G  G  A  R  I  L  I  G  T  Q  M  L  A  K  G  H  H  F  P  D  V  T  L  V  A  L  L  D  V  D  G  A→
2091 GCTGTTTTCTGCCGATTTTCGCTCGGCAGAGCGTTTCGCTCAGCTTTACACCCAGGTCGCCGGTCGTGCCGGGCGTGCGGGTAAACAGGGCGAAGTGGTGCTGCAAACGC
      L  F  S  A  D  F  R  S  A  E  R  F  A  Q  L  Y  T  Q  V  A [G  R  A  G  R  A  G  K  Q] G  E  V  V  L  Q  T→
2201 ACCATCCGGAACATCCTCTGTTGCAAACGTTGCTCTATAAAGGCTACGACGCCTTTGCCGAACAGGCGCTGGCTGAGCGGCGAATGATGCAGCTACCGCCGTGGACCAGC
     H  H  P  E  H  P  L  L  Q  T  L  L  Y  K  G  Y  D  A  F  A  E  Q  A  L  A  E  R  R  M  M  Q  L  P  P  W  T  S→
2311 CATGTGATTGTGCGTGCGGAAGATCATAACAATCAGCACGTGCCATTGTTCCTGCAACAACTGCGTAATCTGATCCTCTCCAGCCCACTGGCAGACGAGAAACTGTGGGT
      H  V  I  V  R  A  E  D  H  N  N  Q  H  V  P  L  F  L  Q  Q  L  R  N  L  I  L  S  S  P  L  A  D  E  K  L  W  V→
2421 TCTCGGTCCGGTTCCGGCTCTGGCACCTAAACGTGGCGGTCGCTGGCGCTGGCAGATATTGTTGCAGCACCCTTCCCGCGTGCGCTTGCAACACATCATTAACGGTACGC
      L  G  P  V  P  A  L  A  P  K  R  G  G  R  W  R  W  Q  I  L  L  Q  H  P  S  R  V  R  L  Q  H  I  I  N  G  T→
2531 TGGCGCTCATCAATACAATACCGGATTCCCGTAAGGTGAAATGGGTGCTGGATGTTGATCCGATTGAGGGTTAAACCGCTCACGATGCGAGGCGGATCGAAAAATTCAAT
     L  A  L  I  N  T  I  P  D  S  R  K  V  K  W  V  L  D  V  D  P  I  E  G  *→
2641 ATTCATCACACTTTTCATGAAAATTCTGTAACCGTTTTCACGCGCTATCTGCTAAAAATGTTGCCGATGTGAAGTAAACATGGATGTAGTACGCCTGACGTGCCAGGCGA
2751 GGAGTGAGTGTGAAAGCGAAGAAGCAGGAAACTGCCGCGACCATGAAAGACGTTGCCCTCAAGGCAAAAGTCTCTACAGCGACCGTCTCCCGAGCATTAATGAATCCCGA
2861 TAAAGTCTCCCAGGCCACCCGTAATCGGGTTGAAAAAGCGGCCCGGGG
```

FIG. 2. DNA sequence of the 2.9-kbp *Sma* I DNA fragment. The 732-amino acid ORF (nucleotides 406–2602), capable of encoding an 81.7-kDa polypeptide, has been translated. A potential ribosome binding site (S.D.), a −10 box, and a −35 box are indicated just upstream of the start of the ORF. Two putative nucleotide binding sites have been boxed in the amino acid sequence and an extended "zinc-finger" consensus has been underlined. The amino acid sequence corresponding to the N terminus of the cyanogen bromide fragment extends from nucleotide 1639 to nucleotide 1669. The single-letter amino acid sequence is used.

restriction enzymes used by Kohara *et al.* (14) and comparing the patterns obtained with the existing *E. coli* map. In this manner, the factor Y gene was localized to 88.5 min on the *E. coli* chromosome.

To confirm the location of the factor Y gene, bacteriophage λ 538 containing this region of the genome was obtained and the DNA was isolated and subjected to Southern blot analysis using a factor Y gene-specific probe (Fig. 4). This probe hybridized to identically sized *Sma* I fragments produced from bacteriophage λ 538 DNA, plasmid DNA carrying the factor Y gene, or total *E. coli* genomic DNA, thus confirming the location of the factor Y gene at 88.5 min on the *E. coli* map.

DISCUSSION

The gene encoding the *E. coli* primosomal protein factor Y has been molecularly cloned, sequenced, and mapped in the *E. coli* genome. DNA sequence analysis of the fragment containing the factor Y gene revealed a large ORF capable of encoding a 732-amino acid polypeptide of 81.7 kDa. When this ORF was expressed using a transient T7 expression system, a polypeptide with an apparent molecular mass of 78 kDa was observed. In addition, soluble extracts prepared from cells containing the overproduced polypeptide showed at least a 2000-fold increase over control extracts of factor Y DNA replication activity. We conclude that the cloned ORF corresponds to the authentic gene for factor Y. As proposed by Lee *et al.* (22), who have independently isolated and sequenced this gene, the gene for factor Y has been designated *priA* (primosomal) and we will refer to the protein as the PriA protein.

The *priA* gene has an excellent ribosome-binding site 7 base pairs upstream from the initiator ATG of the ORF. A −10 Pribnow box is located 31 nucleotides upstream of the initiator ATG and 16 nucleotides further upstream is a putative −35 box.

A search of the translated GenBank, Doolittle, and Protein Identification Resource data banks (February 6, 1990) revealed no significant homologies. An examination of the amino acid sequence resulting from the translated *priA* gene revealed a consensus nucleotide-binding domain (23), Gly-Val-Thr-Gly-Ser-Gly-Lys-Thr, from amino acid 124 to amino acid 131 in the N-terminal portion of the polypeptide. A

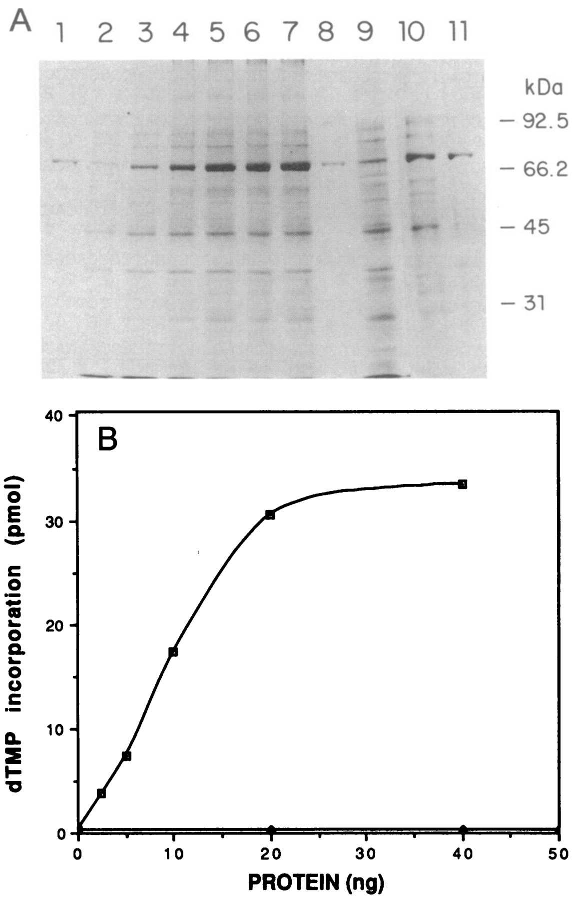

FIG. 3. Expression of the factor Y ORF results in the accumulation of a 78-kDa polypeptide and factor Y-specific DNA replication activity. (*A*) Cells containing a *priA*-bearing plasmid (pET-3c with the putative factor Y ORF cloned into the *Nde* I site) were grown to an OD_{600} of 0.6. Isopropyl β-D-galactopyranoside (0.4 mM) was added and a sample (1 ml) of cells was removed at the indicated times. The cells were pelleted and resuspended in 0.05 vol of TE (10 mM Tris·HCl, pH 7.5/1 mM EDTA). One volume of 2× SDS gel dye [0.25 M Tris·HCl, pH 6.8/1% SDS/5 mM dithiothreitol/10% (vol/vol) glycerol/0.005% bromophenol blue] was added and each sample was heated to 100°C for 5 min. Equal volumes (5 μl) of the samples were then electrophoresed through a 10% polyacrylamide gel containing 0.1% SDS and the protein was visualized by staining with Coomassie blue. Lanes: 1 and 8, 700 ng of bona fide factor Y; 2–7, time course of induction (0 min, 30 min, 1 hr, 2 hr, 3 hr, and 4 hr, respectively); 9, lysate (5 μl) from uninduced cells at 4 hr; 10, 12.5 μg of soluble protein extract (prepared as described below); 11, 12.5 μg of protein from the resuspended insoluble pellet. (*B*) Four hours after induction, cells were harvested by centrifugation, washed in 50 mM Tris·HCl, pH 8.0 (at 4°C)/10% (wt/vol) sucrose, resuspended in lysis buffer (50 mM Tris·HCl, pH 8.5/150 mM NaCl/10 mM EDTA/1 mM dithiothreitol/10% sucrose), and quick frozen in liquid N_2. Thawed cells were lysed on ice in the presence of 0.02% lysozyme and 0.1% Brij 58 for 30 min. The lysate was centrifuged at 37,000 rpm in an A841 rotor for 1 hr at 4°C. The supernatant is the soluble cell extract. The insoluble pellet was resuspended in an equal volume of lysis buffer. The soluble extract was assayed for factor Y activity using the bacteriophage ϕX174 single-stranded circular → replicative form DNA replication system reconstituted with purified proteins as described (21). ⊡, Induced extracts; ●, control extracts made from cells identical except for the absence of T7 RNA polymerase that are, therefore, unable to express factor Y protein. No incorporation of nucleotide was observed until the addition of 850 ng of control extract protein. Based on calculations of the specific activity of factor Y in the two extracts, at least a 2000-fold overexpression of this protein was achieved.

second putative nucleotide-binding domain, Gly-Arg-Ala-Gly-Arg-Ala-Gly-Lys-Gln, is present from amino acid 583 to amino acid 591 in the C-terminal region. The presence of one or more such domains was expected, given that the PriA protein is a potent single-stranded DNA-dependent (d)ATPase and DNA helicase (4, 5, 8, 9).

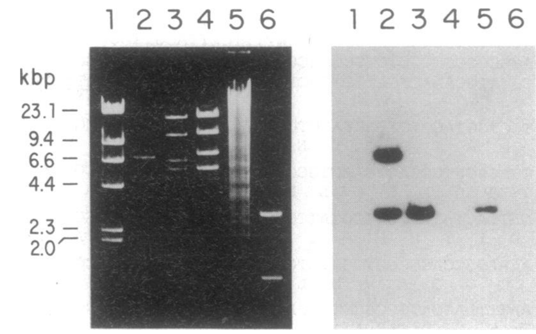

FIG. 4. *priA* gene maps to 88.5 min on the *E. coli* chromosome. Southern blot analysis of the *priA* gene in plasmid DNA, bacteriophage λ 538 DNA, and total genomic *E. coli* DNA. (*Left*) Ethidium bromide-stained agarose gel. (*Right*) Southern blot probed with nick-translated plasmid DNA carrying the *Sma* I fragment containing the *priA* gene. Lanes: 1, size markers; 2, M13mp19 replicative form DNA carrying the *priA* gene digested with *Sma* I; 3, bacteriophage λ 538 DNA, containing the region about 88.5 min on the *E. coli* chromosome, digested with *Sma* I; 4, λ 329 DNA, containing the region about 38 min on the *E. coli* chromosome (14), digested with *Sma* I; 5, *E. coli* genomic DNA digested with *Sma* I; 6, 2.9-kbp and 1.2-kbp size markers. The *priA*-containing plasmid DNA, bacteriophage λ 538, and total *E. coli* genomic DNA all contain an identical 2.9-kbp *Sma* I fragment that is complementary to the *priA* probe. The 7.5-kbp band observed in lane 2 corresponds to M13mp19 vector sequences that hybridize because the *priA* probe DNA was nick-translated in the same plasmid vector.

The ability of the PriA protein to recognize specifically and bind to a variety of PAS sequences that have relatively little homology (16), other than the potential to form extensive secondary structures (9, 24, 25), and subsequently to catalyze assembly of the primosome suggests that the PriA protein participates in a unique and complex set of protein–DNA and DNA–DNA interactions. Identification and analysis of the DNA-binding domain(s) of the PriA protein should, therefore, prove informative.

Preliminary examination of the amino acid sequence inferred from the *priA* ORF reveals a series of 44 amino acids that fit the consensus for one or more metal-binding or "zinc-finger" domains (26) identified as being necessary for the sequence-specific binding of a number of prokaryotic and eukaryotic proteins. The putative PriA zinc-finger domain extends from amino acid 436 to amino acid 480 and contains the sequence Cys-Xaa_2-Cys-Xaa_5-Cys-Xaa_2-Cys-Xaa_{14}-Cys-Xaa_2-Cys-Xaa_9-Cys-Xaa_2-Cys. Three distinct zinc fingers derived from overlapping sequences could be formed from this region as indicated below:

$$CX_2CX_5CX_2CX_{14}CX_2CX_9CX_2C$$

The central zinc finger in this domain is very similar to the zinc-finger consensus within the primase domain of the T7 gene *4* primase–helicase protein (27), as well as that present in the T4 gene *61* primase protein (28). The PriA protein must move from the primarily double-stranded PAS sequence onto the contiguous single-stranded DNA binding protein-coated single-stranded DNA where it can subsequently translocate in the 3′ → 5′ direction. It is interesting to speculate that this could involve a series of ATP hydrolysis-driven conformational changes designed to alternate which zinc finger is occupied.

The location of the *priA* gene has been mapped to 88.5 min on the *E. coli* chromosome, between the *glpK* gene (glycerol kinase) at 88.4 min and the *btuB* gene (thiamine uptake) at 88.7 min. No existing *dna* mutant maps to this position, leaving the role of the PriA protein in chromosomal DNA replication unknown.

We thank Eui-Hum Lee, Hisao Masai, George Allen, Jr., and Arthur Kornberg (Stanford University) for exchanging *priA* sequences in advance of publication and Drs. J. Hurwitz, M. O'Donnell, S. Rabkin, and S. Shuman for their critical reading of the manuscript. These studies were supported by National Institutes of Health Grant GM34557.

1. Arai, K. & Kornberg, A. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 69–73.
2. Marians, K. J. (1984) *CRC Crit. Rev. Biochem.* **17**, 153–215.
3. LeBowitz, J. & McMacken, R. (1986) *J. Biol. Chem.* **261**, 4738–4748.
4. Lee, M. S. & Marians, K. J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8345–8349.
5. Lasken, R. S. & Kornberg, A. (1988) *J. Biol. Chem.* **263**, 5512–5518.
6. Lee, M. S. & Marians, K. J. (1989) *J. Biol. Chem.* **264**, 14531–14542.
7. Mok, M. & Marians, K. J. (1987) *J. Biol. Chem.* **262**, 16644–16654.
8. Wickner, S. & Hurwitz, J. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3342–3346.
9. Shlomai, J. & Kornberg, A. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 799–803.
10. Minden, J. S. & Marians, K. J. (1985) *J. Biol. Chem.* **260**, 9316–9325.
11. van der Ende, A., Baker, T. A., Ogawa, T. & Kornberg, A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3954–3958.
12. Stuitje, A. R., Weisbeck, P. J. & Meijer, M. (1984) *Nucleic Acids Res.* **12**, 3321–3332.
13. van der Ende, A., Teestra, R., van der Avoort, H. G. A. M. & Weisbeck, P. J. (1983) *Nucleic Acids Res.* **11**, 4957–4975.
14. Kohara, Y., Akiyama, K. & Isono, K. (1987) *Cell* **50**, 495–508.
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
16. Greenbaum, J. H. & Marians, K. J. (1984) *J. Biol. Chem.* **259**, 2594–2601.
17. Matsudaira, P. (1987) *J. Biol. Chem.* **262**, 10035–10038.
18. Konigsberg, W. & Godson, G. N. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 687–691.
19. Saiki, K. N., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horr, G. T., Mullis, K. B. & Elrich, H. A. (1988) *Science* **239**, 487–491.
20. Studier, W. F., Rosenberg, A. H. & Dunn, J. J. (1990) *Methods Enzymol.*, in press.
21. Abarzua, P., Soeller, W. & Marians, K. J. (1984) *J. Biol. Chem.* **259**, 14286–14292.
22. Lee, E. H., Masai, H., Allen, G. C., Jr., & Kornberg, A. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4620–4624.
23. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982) *EMBO J.* **1**, 945–951.
24. Soeller, W., Abarzua, P. & Marians, K. J. (1984) *J. Biol. Chem.* **259**, 14293–14300.
25. Greenbaum, J. H. & Marians, K. J. (1985) *J. Biol. Chem.* **260**, 12266–12272.
26. Berg, J. M. (1986) *Science* **232**, 485–487.
27. Bernstein, J. A. & Richardson, C. C. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 396–400.
28. MacDonald, P. M. & Mosig, G. (1984) *EMBO J.* **3**, 2863–2871.
29. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.