Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 4620–4624, June 1990
Biochemistry

The *priA* gene encoding the primosomal replicative n′ protein of *Escherichia coli*

EUI HUM LEE*, HISAO MASAI†‡, GEORGE C. ALLEN, JR.*, AND ARTHUR KORNBERG*

*Department of Biochemistry, Beckman Center, Stanford University, Stanford, CA 94305-5307; and †DNAX Research Institute of Molecular and Cellular Biology, 901 California Street, Palo Alto, CA 94304-1104

Contributed by Arthur Kornberg, April 2, 1990

ABSTRACT **The *Escherichia coli* gene encoding protein n′ has been isolated and named *priA* for *pri*mosomal protein *A*. Protein n′ is absolutely required for the conversion of single-stranded φX174 DNA to the duplex replicative form in an *in vitro*-reconstituted system. The gene maps to 88.7 minutes on the chromosome adjacent to the *cytR* locus. Soluble protein extracts from cells harboring the *priA* gene on a multicopy plasmid contained 45-fold more n′ replication activity than wild-type extracts. Enhanced overproduction of >1000-fold was achieved by replacing the natural Shine–Dalgarno sequence with that of the phage T7 φ10 gene and placing this *priA* under the control of the T7 phage promoter and RNA polymerase. The *priA* sequence reveals a 732-amino acid open reading frame and a nucleotide-binding consensus site consistent with the size and ATPase activity of the purified protein. The gene for protein n has been named *priB* and the putative gene for protein n″, *priC*.**

The conversion of the single-stranded (ss) genome of phage φX174 to its duplex replicative form (RF) depends solely on host proteins (1). This *in vitro* φX ss to RF system allowed the purification and characterization of these *Escherichia coli* replication proteins as well as providing models for the mechanism of discontinuous strand synthesis (1, 2). The majority of the proteins required for reconstituted φX replication are intimately involved in host replication including: DnaB, DnaC, DnaG (primase), DnaT, ssDNA-binding protein, and DNA polymerase III (1, 3–5). The role of the remaining proteins, n, n′, and n″, in host replication is still unclear.

Protein n′ is first to act in converting φX ssDNA to the duplex RF. It binds to a specific hairpin on ssDNA-binding protein-coated φX ssDNA triggering the assembly of a mobile primosome by the concerted action of DnaB, DnaC, DnaT, proteins n and n″, and primase (6). This primosome can translocate bidirectionally on the template through the helicase action of DnaB and protein n′ fueled by ATP hydrolysis (2, 7). Simultaneously, primase in the primosome can lay down an RNA primer that DNA polymerase III can utilize.

We report the isolation and nucleotide sequence§ of the gene encoding protein n′, *priA*, and the features of that sequence. The construction of a plasmid that overproduces PriA activity >1000-fold is also presented. Nurse *et al.* (8) have independently isolated the *priA* gene and determined its sequence.

MATERIALS AND METHODS

Strains, Enzymes, and Other Reagents. *E. coli* strains used were DH5α [F⁻, φ80d, *lacZΔM15, endA1, recA1, hsr−, hsm+, supE44, thi-1, gyrA,* λ−, Δ(*lacZYA-argF*), U169], K38 [HfrC(λ)] (9), and W3110 (F⁻,λ⁻). Restriction endonucleases and modifying enzymes were purchased from either New England Biolabs or Bethesda Research Laboratories and used according to the manufacturer's instructions. Highly purified DNA replication proteins were as described (10). Plasmid pTZ18R was from Pharmacia LKB. The pGP1-2 and pT7-6 plasmids were kindly provided by S. Tabor (Harvard Medical School).

Reagents were: unlabeled deoxynucleotide triphosphates and ribonucleoside triphosphates (Pharmacia LKB); [γ-^{32}P]ATP, deoxyadenosine 5′-[α-[^{35}S]thio]triphosphate, and [α-^{32}P]dTTP (Amersham); bovine serum albumin (Pentax Fr V; Sigma); prokaryotic DNA-directed translation kit (Amersham); Sequenase DNA sequencing kit (United States Biochemical); Erase-a-Base system kit (Promega). Buffer G is 20 mM Tris·HCl (pH 7.5), 150 mM potassium glutamate, 9 mM $MgCl_2$, 4% (wt/vol) sucrose, and bovine serum albumin (0.1 mg/ml).

Replication Assay for Protein n′. The reconstituted φX ss to RF replicative assay was essentially as described (10) with modification to buffer G as noted above. One unit of replication activity promotes the incorporation of 1 pmol of nucleotide (nt) in 1 min at 30°C.

Preparation of Fraction II. Cells were grown in 2 liters of L broth to OD_{600} = 1.0, collected, and lysed as described (10). The supernatant (fraction I) was precipitated with ammonium sulfate (0.35 g/ml). The centrifuged pellets were resuspended in buffer containing 50 mM Hepes·KOH (pH 7.6), 20% (vol/vol) glycerol, 100 mM NaCl, 1 mM dithiothreitol, and 1 mM EDTA (fraction II).

Screening of an *E. coli* Genomic Library with Synthetic Oligonucleotide Probes. Protein n′ was purified to near homogeneity as described (10). N-terminal amino acid sequence of protein n′ performed by Allan Smith (Stanford University) produced the amino acid sequence, NH_2-XXXAXVALPV-PLPRTFDYLLP, from which two oligonucleotide probes were designed. Probe 1 was 5′-CCICG(C,T)AC(C,G,T)TT-(C,T)GA(C,T)TA-3′ (nt 100–116, amino acids 13–18) and probe 2 was 5′-GTIGCICTGCCGGTICCGCTGCCGCGT-ACITTCGACTACCTGCTGCC-3′ (nt 79–125, amino acids 6–21). Oligonucleotides were synthesized by Felix Vega (DNAX Research Institute) on a DNA synthesizer (Applied Biosystems). The *E. coli* genomic library was constructed by inserting *Sau*3A1 partial digests of *E. coli* W3110 DNA at the *Bam*HI site of pTZ18R vector. The screening of the *E. coli* DNA was by plasmid DNA grouping and Southern blot hybridization. From 3000 individual clones, a single clone, pEL042 that hybridizes to both probes was obtained. Additional screening of the *E. coli* library was performed by colony lift, as described (11). Two positive clones were obtained from 12,000 colonies: pEL050 hybridizing to both probes and pEL052 hybridizing to probe 1 only.

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: ss, single-stranded; RF, replicative form; ORF, open reading frame; nt, nucleotide; SD, Shine–Dalgarno.

‡Present address: Department of Molecular Biology, Institute of Medical Science, University of Tokyo, Shirokamedai Minatoku, Tokyo 108, Japan.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M33293).

DNA Sequencing. DNA sequencing was by the dideoxynucleotide sequencing method (11) using the Sequenase system (United States Biochemical) according to the specifications of the manufacturer. Briefly, the 2.7-kilobase (kb) *Hin*dIII–*Eco*RI insert of pEL042 was digested with *Alu* I and the resulting fragments were subcloned into pTZ18R and sequenced. Additionally, nested internal deletions were constructed using the Erase-A-Base system as described by the manufacturer's specifications and then sequenced. Analysis of the nucleotide sequences was carried out using the programs of IntelliGenetics (12). All sequences presented here have been determined on both strands.

Construction of Overexpression Vectors. A plasmid expressing the *priA* gene from the T7 ϕ10 promoter, pRA13, was made by cloning the *Hin*dIII–*Eco*RI fragment of pEL042 into *Hin*dIII–*Eco*RI-cut pT7-6. A plasmid with the ϕ10 Shine–Dalgarno (SD) region upstream of the *priA* coding region was constructed by first subcloning the *Hin*dIII–*Eco*RI fragment of pEL042 into *Hin*dIII–*Eco*RI-cut M13mp-18. Site-directed mutagenesis by the method of Kunkel *et al.* (13) using 5′-GTCAGGATGACATATGGCCCGTTGC-3′ to prime minus-strand synthesis generated an *Nde* I site at the start codon. This mutation was confirmed by restriction digest and DNA sequencing. This *Nde* I mutant of *priA* was then introduced into pT7-6 as with the wild-type sequence. The resulting plasmid was digested with *Nde* I and *Eco*RI and ligated with *Eco*RI/*Xba* I-cut pT7-6 and the SD cassette with *Xba* I and *Nde* I ends (cassette sequence shown in Fig. 1*A*). The resulting plasmid, pRA45, was introduced into K38(pGP1-2). Growth and induction were as described (9).

RESULTS

Isolation of the Gene Encoding Protein n′. The gene encoding the n′ protein was isolated by standard methods. The N-terminal amino acid sequence suggested two synthetic oligonucleotide probes that were used to screen a plasmid-based *E. coli* genomic library. By Southern blot analysis and colony hybridization, three potential clones were identified, pEL042, pEL050, and pEL052. Restriction analysis showed that these contained 2.7-, 4.0-, and 3.0-kb inserts, respectively.

All of these inserts are sufficiently long to encode protein n′ and so were further characterized. In a coupled *in vitro* transcription/translation system, only pEL042 yielded a protein of 76 kDa consistent with the migration of purified n′ protein; the others produced polypeptides of less than 20 kDa (data not shown). Deletion of 1.3 kb of the pEL042 insert (between nt 1 and 1250) abolished synthesis of the 76-kDa protein confirming that its synthesis is insert-derived. Furthermore, a crude ammonium sulfate fraction prepared from cells harboring pEL042 contained 45-fold more n′ activity than cells harboring vector alone (Table 1). Cells bearing either pEL050 and pEL052 overexpressed little n′ activity. The pEL042 clone thus contains the gene for protein n′, named *priA*. Accordingly, the genes for the primosomal proteins, n and n″ (14), have been named *priB* and *priC*, respectively.

Enhanced Overproduction of PriA Protein. The first strategy employed for achieving overexpression of the PriA protein was to improve transcription efficiency by placing it under the control of the T7 promoter and RNA polymerase (9). Even under these conditions in which transcription should not be limiting, PriA activity was overexpressed only 32-fold relative to vector alone (Table 1). This level of overproduction was comparable to that of pEL042 (Table 1), indicating that translation was limiting. Therefore, the natural SD sequence was exchanged for that of the efficiently translated ϕ10 gene of T7 (Fig. 1*A*). The resulting construct, pRA45, overproduced PriA protein as a visible band on SDS/PAGE of total cell protein after induction (Fig. 1*B*) and 1000-fold or more in a reconstituted ϕX ss to RF assay (Table 1).

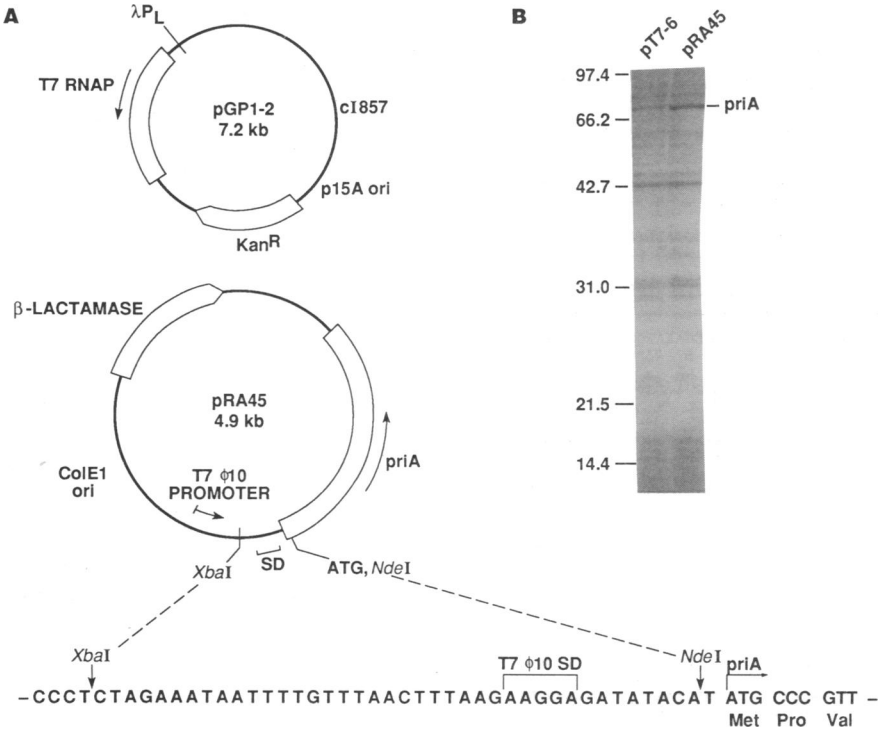

FIG. 1. Overproduction of PriA protein. (*A*) pRA45 contains the *priA* gene downstream of the T7 phage promoter. The 5′ nontranslated region of *priA* has been replaced by the SD and flanking sequences of the T7 gene *10*; pGP1-2 contains the T7 RNA polymerase (RNAP) gene under control of the λ phage P_L promoter and the *c*I857 temperature-sensitive repressor. (*B*) *E. coli* K38 cells harboring pGP1-2 and either pT7-6 or pRA45 were grown and induced as described (9). After induction equal numbers of cells were incubated in 125 mM Tris·HCl, pH 6.8/0.1% SDS/20% glycerol/60 mM 2-mercaptoethanol for 3 min at 100°C and analyzed by SDS/PAGE on 12% gels. Proteins were visualized by Coomassie brilliant blue staining. The mobility of PriA protein on SDS/PAGE is indicated. Molecular masses are indicated in kDa.

Table 1. Overproduction of protein n′ activity

Group	Plasmid	Insert	n′ activity ($\times 10^{-3}$), units/g of cells	Over-production, fold
A	pTZ18R	—	8.5	1
	pEL042	2.7 kb	381	45
	pEL050	4.0 kb	12.9	1.5
	pEL052	3.0 kb	12.1	1.4
B	pT7-6	—	6.4	1
	pRA13	*priA* with native SD	209	32
	pRA45	*priA* with T7 ϕ10 SD	6986	1083

Fraction II was prepared from strains harboring the plasmid indicated. Group A plasmids are either pTZ18R or its derivatives and were propagated in DH5α. Group B plasmids are either pT7-6 or its derivatives and were propagated in K38 also harboring pGP1-2 (Fig. 1*A*). Cells were grown at 30°C as described (9) and induced by adding an equal volume of L broth preheated to 54°C and then were grown again at 37°C for 2 hr.

Nucleotide Sequence of the *priA* Gene. The nucleotide sequence of *priA* (Fig. 2) reveals an open reading frame of 2196 base pairs (nt 64–2259) translating to a protein of 732 amino acids and a calculated mass of 81.8 kDa. The size, though somewhat larger than the mobility of protein n′ on gels (76 kDa), and translated N-terminal amino acid sequence are consistent with those of purified protein n′. A putative promoter containing a −35 and −10 consensus (nt 5–10 and nt 27–32, respectively) and an SD sequence (nt 53–59) were also identified upstream of the *priA* coding region. The *priA* gene uses several rare codons such as CCC (proline), ACA (threonine), and CGG (arginine), 2–4 times more frequently than average (data not shown). Rare codon usage and inefficient translation initiation may contribute to the low cellular abundance of protein n′ (70 molecules per cell; ref. 10). Sequences at the end of the 2.7-kb insert of pEL042 match those flanking and overlapping the chromosomal *cytR* locus (nt 2119–2460). The *cytR* gene encodes a protein that regulates the expression of several proteins involved in nucleoside metabolism. Using the Kohara phage library of the *E. coli* chromosome (17), the *priA* gene was mapped to 88.7 minutes (data not shown) in agreement with the reported location of *cytR* (15).

Structural Features of the *priA* Gene. A consensus for nucleotide binding composed of two motifs has been found in

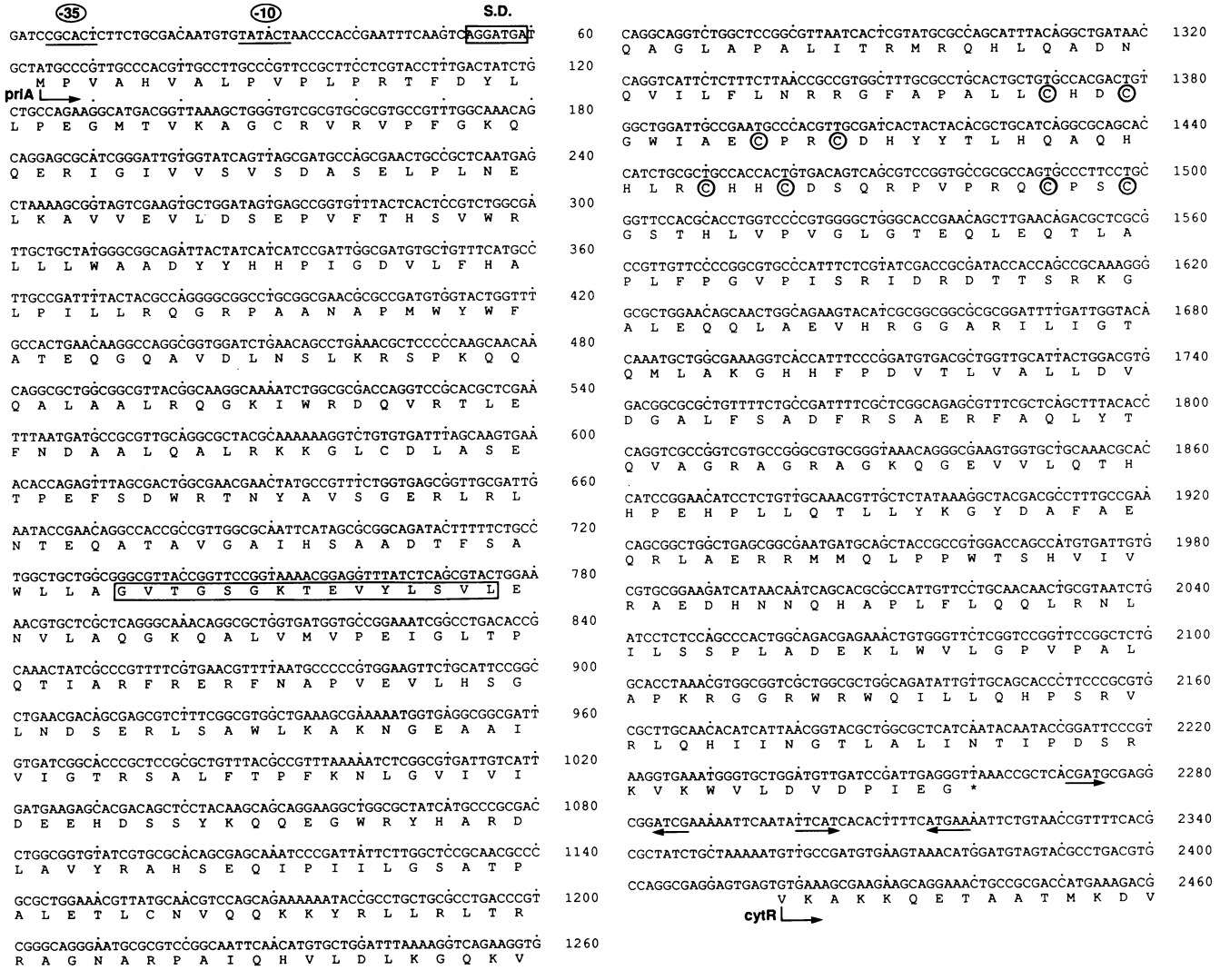

FIG. 2. Nucleotide sequence of the *priA* gene of *E. coli*. The nucleotide sequence of the insert of pEL042 containing the *priA* gene was determined and the amino acid sequence was deduced. The −35 and −10 regions are underlined and the SD sequence for translation initiation is boxed. Two possible stem–loop structures involved in transcription termination are shown by two small arrows pointing toward each other. A portion of the deduced amino acid sequence of the *cytR* gene (15), which overlaps the end of the pEL042 insert, is also shown. The GX_4GKTX_4I nucleotide-binding motif (16) is indicated by a large box. The cysteine residues of the cysteine-rich region are circled. Further details are discussed in the text.

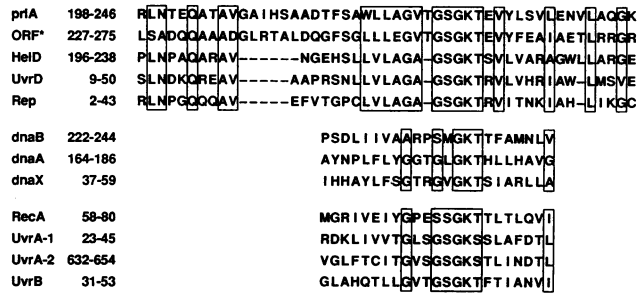

FIG. 3. Nucleotide-binding consensus sites from replicative and repair proteins of *E. coli*. The larger consensus sequence emerging from the comparison of these replicative and repair nucleotide-binding proteins is shown in boxes with the consensus of Walker *et al.* (16), the GX_4GKTX_4I motif, embedded within it. Amino acid positions of the sequences are shown. References: *priA*, this work; ORF, *R. rubrum* (18); HelD (19); UvrD (20); Rep (21); DnaB (22); DnaA (23); DnaX (24); RecA (25); UvrA (26); and UvrB (27).

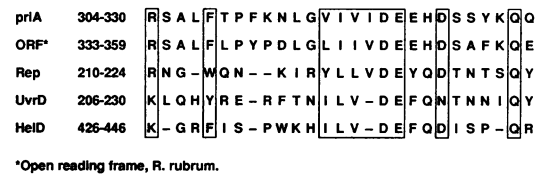

FIG. 4. Nucleotide-binding consensus sites from helicases of *E. coli*. A consensus sequence derived from several helicases with a partial match to the R/KX_3GX_3-(hydrophobic)$_4$-D motif of the nucleotide-binding consensus is shown in the boxes. Amino acid positions of sequences are shown. References: PriA, this work; ORF, *R. rubrum* (18); HelD (19); UvrD (20); and Rep (21).

many ATP-binding proteins and ATPases (16), including the replicative and repair helicases. The *priA* coding region contains such a nucleotide binding consensus consistent with the ATPase and helicase activities of PriA protein (Figs. 2, 3, and 4). The PriA protein shows a strong match to the GX_4GKTX_4I motif (Fig. 3). The homology between several of the prokaryotic helicases including PriA protein extends to flanking regions outside the nucleotide binding site implying conservation of structure in these areas as well. The second motif, R/KX_3GX_3-(hydrophobic)$_4$-D, is not matched as well but the hydrophobic residues thought to interact with the adenine ring and the aspartic acid thought to interact with a Mg^{2+} are present (Fig. 4). Also of note is a cysteine-rich region (8 cysteines of 44 residues; amino acids 435–479; Fig. 2). As has been noted (28), these regions may be involved in metal binding and play a role in either protein–protein or protein–DNA interactions.

A search of the Swiss Protein databank revealed an homologous *priA* gene in the distantly related, photosynthetic bacterium *Rhodospirillum rubrum* (18). The *priA* gene and the unidentified open reading frame (ORF) of *R. rubrum* show a 37% conservation of sequence at the amino acid level spread across the entire ORF including residues in both nucleotide-binding motifs as well as the cysteine-rich region.

DISCUSSION

Protein n′ is an essential component of the *E. coli* primosome. The protein recognizes a specific hairpin sequence on ϕX ssDNA, directs primosome assembly at that site, then functions as a helicase within the primosome. The genes for the majority of the primosomal proteins have been identified, aiding significantly in the genetic and biochemical characterization of their functions (1, 3, 5). As reported, we have identified the gene for n′, the *priA* gene. This gene has allowed the construction of strains that overproduce PriA protein >1000-fold, thereby providing an abundant source for the biochemical analysis of its function. Interruption of the gene as well as its overexpression should be helpful in assessing the physiologic role of *priA*. The *priB* gene, encoding protein n, has been discovered in a ribosomal protein operon (G.C.A., Jr. and A.K., unpublished results), but the gene for protein n″ has yet to be found.

The sequence of the *priA* gene revealed several features consistent with its role in primosome function. The PriA protein contains matches to two motifs of the nucleotide-binding consensus found in most ATP-binding proteins and ATPases (16). More striking is the level of conservation in these regions among several helicases. Of the known *E. coli* helicases, UvrD, Rep, DNA helicase IV (HelD), and PriA are similar in their helicase polarity (3′ → 5′), nucleotide utilization (ATP or dATP), and size (70–80 kDa) (7). As reported (21), Rep and UvrD are highly homologous across the entire polypeptide chain whereas HelD contains only localized homology to Rep and UvrD. PriA has the least overall homology to any of these three helicases. However, all of these helicases have a high degree of homology within and flanking the GX_4GKTX_4I motif including the indeterminate residues, X. These additional conserved residues are not present in other replicative and repair ATP-binding proteins (Fig. 3), suggestive of their importance for the function of these helicases.

The bacterium, *R. rubrum*, contains an unidentified ORF that is homologous to PriA. These two coding regions of nearly identical size show a 37% identity spread over the length of the polypeptide with higher degrees of homology localized to the nucleotide binding motifs and the cysteine-rich region. These localized homologies imply that these structures are critical to PriA function. The conservation of primary sequence between these two distantly related bacteria suggests that the ORF of *R. rubrum* is a functional analogue of PriA and that the two may be evolutionarily related.

Note. The *priA* genes isolated in our laboratory and that of Kenneth Marians (8) are identical in sequence with the exception of nucleotide 2009, which we report as a thymidine and they as a cytidine translating to an alanine at residue 648 of our protein sequence and a valine in their sequence.

We thank LeRoy Bertsch for his help in preparing this manuscript and Dr. Douglas Brutlag for his help in conducting homology searches and sequence comparisons. We are grateful to Dr. Kenneth Marians for the exchange of *priA* sequences with us prior to publication. This work was supported by National Institutes of Health Grant GM07581 and National Science Foundation Grant DMB87-1007945. G.C.A., Jr., was supported by a predoctoral fellowship from the Howard Hughes Medical Institute.

1. Kornberg, A. (1980) *DNA Replication* (Freeman, San Francisco).
2. Lee, M. S. & Marians, K. (1989) *J. Biol. Chem.* **264,** 14531–14542.
3. Lark, C. A., Riazi, J. & Lark, K. G. (1978) *J. Bacteriol.* **136,** 1008–1017.
4. Kornberg, A., Baker, T. A., Bertsch, L. L., Bramhill, D., Funnell, B. E., Lasken, R. S., Maki, H., Maki, S., Sekimizu, K. & Wahle, E. (1987) *DNA Replication and Recombination* (Liss, New York).
5. Sekimizu, K., Bramhill, D. & Kornberg, A. (1988) *J. Biol. Chem.* **263,** 7124–7130.
6. Shlomai, J. & Kornberg, A. (1980) *J. Biol. Chem.* **255,** 6794–6798.
7. Matson, S. W., Kaiser-Rogers, K. A. (1990) *Annu. Rev. Biochem.* **59,** 289–329.
8. Nurse, P., DiGate, R., Zavitz, K. & Marians, K. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 4615–4619.

9. Tabor, S. & Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1074–1078.
10. Shlomai, J. & Kornberg, A. (1980) *J. Biol. Chem.* **255,** 6789–6793.
11. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
12. Brutlag, D. L., Clayton, J., Friedland, P. & Kedes, L. H. (1982) *Nucleic Acids Res.* **10,** 279–294.
13. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987) *Methods Enzymol.* **154,** 367–382.
14. Low, R. L., Shlomai, J. & Kornberg, A. (1982) *J. Biol. Chem.* **257,** 6242–6250.
15. Valentin-Hansen, P., Larsen, J. E. L., Hojrup, P., Short, S. A. & Barbier, C. S. (1986) *Nucleic Acids Res.* **14,** 2215–2228.
16. Walker, J. E., Runswick, M. J. & Gay, N. J. (1982) *EMBO J.* **1,** 945–951.
17. Kohara, Y., Akiyama, K. & Isono, K. (1987) *Cell* **50,** 495–508.
18. Falk, G., Hampe, A. & Walker, J. E. (1985) *Biochem. J.* **228,** 391–407.
19. Wood, E. R. & Matson, S. W. (1989) *J. Biol. Chem.* **264,** 8297–8303.
20. Finch, P. W. & Emmerson, P. T. (1984) *Nucleic Acids Res.* **12,** 5789–5799.
21. Gilchrist, C. A. & Denhardt, D. T. (1987) *Nucleic Acids Res.* **15,** 465–475.
22. Nakayama, N., Arai, N., Bond, M. W., Kasiro, Y. & Arai, K. (1984) *J. Biol. Chem.* **259,** 97–101.
23. Hansen, E. B., Hansen, F. G. & von Meyenburg, K. (1982) *Nucleic Acids Res.* **10,** 7373–7385.
24. Yin, K.-C., Blinkowa, A. & Walker, J. R. (1986) *Nucleic Acids Res.* **14,** 6541–6549.
25. Sancar, A., Stachelek, C., Konigsberg, W. & Rupp, W. D. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 2611–2615.
26. Husain, I., Van Houten, B., Thomas, D. C. & Sancar, A. (1986) *J. Biol. Chem.* **261,** 4895–4901.
27. Arikan, E., Kulkarni, M. S., Thomas, D. C. & Sancar, A. (1986) *Nucleic Acids Res.* **14,** 2637–2650.
28. Berg, J. M. (1986) *Science* **232,** 485–487.