Proc. Natl. Acad. Sci. USA
Vol. 76, No. 2, pp. 580–584, February 1979
Biochemistry

Nucleotide sequence of the origin of replication of the *Escherichia coli* K-12 chromosome

(initiation of DNA replication/minichromosomes/phage λ/phage G4/*dnaG* protein)

MICHIEL MEIJER*†, EWALD BECK‡, FLEMMING G. HANSEN§, HANS E. N. BERGMANS*¶, WALTER MESSER*‖, KASPAR VON MEYENBURG§, AND HEINZ SCHALLER‡

*Max-Planck-Institut für molekulare Genetik, 1000 Berlin 33, Germany; ‡Mikrobiologie, Universität Heidelberg, 6900 Heidelberg, Germany; and §University Institute of Microbiology, Copenhagen, Denmark

Communicated by Donald A. Glaser, November 2, 1978

ABSTRACT **The origin of replication, *oriC*, of the *Escherichia coli* chromosome was mapped within a DNA segment of 422 base pairs. The nucleotide sequence of this segment was determined. The source of DNA for the sequence analysis was a minichromosome constructed *in vivo*, consisting exclusively of chromosomal DNA and a minichromosome constructed by cloning *in vitro*. The nucleotide sequence of the replication origin is characterized by a high degree of repetitiveness due to both inverted and direct repeats. Sequence homologies were found between portions of the replication origins of *E. coli* and phages λ and G4. This suggests similarities in some steps in the initiation of replication of the different replicons.**

Replication of the circular chromosome of *Escherichia coli* is initiated at a fixed site, the origin of replication, for which the symbol *oriC* has been introduced (cf. refs. 1, 2). We define *oriC* here as the minimal segment of DNA that is able to promote *E. coli* specific initiation of replication. Isolation and characterization of specialized transducing phages λ*asn* carrying the origin of replication of the *E. coli* chromosome has led to the allocation of *oriC* between the genetic markers *uncB* and *asn* at 82.5 min on the circular genetic map (1, 2). This is in accordance with the localization of *oriC* obtained by other genetic techniques (3, 4). The replication origin can be placed within a segment of 1500 base pairs of DNA counterclockwise of the *asn* gene (1, 2). A 5.9-megadalton (9.0 kilobase pairs) *Eco*RI fragment from this region of the chromosome (ref. 5; Fig. 1*A*) has been cloned in minichromosomes, plasmids that contain *oriC* as the sole replication origin (3, 6). The position of *oriC* has been narrowed down further to a segment of ≈500 base pairs by means of deletions introduced into minichromosomes *in vitro* by various restriction endonucleases (6, 7).

Here we present the nucleotide sequence of the DNA segment containing *oriC*. As a substrate for the sequence analysis, we constructed by genetic techniques (*in vivo*) a minichromosome 2.65 megadaltons (4.0 kilobase pairs) in size. This consisted exclusively of chromosomal DNA that included, besides *oriC*, the *asn* gene. As an alternative source of DNA we used minichromosomes constructed independently by means of *in vitro* techniques.

Results complementary to our data are described in an accompanying paper (ref. 8; see also refs. 7 and 9).

MATERIALS AND METHODS

Bacterial Strains. All strains are *E. coli* K-12. CM924 (F^+, *thi, asn, recA*), CM927 (F^+, *thi, asn, recA*, λ*cI*857*S*7), CM987 (F^-, *thi, asn, recA*), and CM993 (F^-, *thi, asn, recA*, pCM959) are derivatives of strain ER (F^+, *asn, thi*). For details of strain construction (except for CM993) and of the isolation of the specialized transducing phages λ*asn* see ref. 1. The chromosomal DNA, carried by the λ*asn* and the minichromosomes derived from them, originates from a Δ*att*λ strain originally derived from strain W3110 (10). The strains carrying pOC-plasmids are derivatives of JC7516 (F^-, *recA56, recB21, argE, ara, galK, his, lacY, leu, mtl, phx, proA, rpsL, supE, thi, thr, tsx, xyl*); see ref. 6.

For the large-scale preparation of pCM959 DNA for nucleotide sequence analysis, strain CM993 was constructed by introduction of pCM959 into the F^- strain CM987 (*recA, asn*) by transformation. Twenty to 50 μg of pCM959 DNA could easily be prepared from 1 liter of culture of CM993 grown in minimal medium supplemented with glucose, casamino acids, and thiamine. Maintenance of selective growth conditions was necessary because these plasmids are lost with a frequency of 0.05–0.10 per division in spite of a high copy number.

Growth Conditions and the Methods for Preparation of Specialized Transducing Phages and Plasmid DNA. These were as described (1, 5, 6).

Restriction Analysis. Restriction nucleases *Alu* I, *Hae* III, *Hha* I, *Hin* fI, and *Hpa* II were prepared as described (11); *Bam*HI, *Bgl* II, *Hin*dIII, *Pst* I, *Taq* I, and *Xho* I were from New England BioLabs. Digestion was in 20 mM Tris·HCl/5 mM $MgCl_2$/0.1 mM EDTA/50 mM KCl/0.5 mM dithioerythritol/5% (vol/vol) glycerol, pH 8.0, or in 20 mM Tris·HCl/10 mM NaCl/10 mM $MgCl_2$, pH 7.8.

Fine mapping of restriction sites closest to the *Hin*dIII or *Bgl* II sites in the origin region was done by digestion of fragments, which were labeled at these 5′ termini with [γ-^{32}P]ATP and T4 polynucleotide kinase (Boehringer Mannheim), with additional restriction endonucleases. The lengths of the resulting labeled DNA fragments were determined by electrophoresis on polyacrylamide gels, using restriction fragments from phage fd replicative form DNA of known sequence (12) as size standards.

DNA Sequencing. Sequences were determined as described by Maxam and Gilbert (13). For purine-specific cleavage, a modification of the Maxam and Gilbert technique was used (14). The degradation products were separated on 10% and 20% polyacrylamide gels, which occasionally were rerun after autoradiography (−20°C). This allows determination of sequences up to >250 nucleotides from a single restriction site.

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.

† Present address: Laboratory of Electron Microscopy, University of Amsterdam, Amsterdam, Netherlands.
¶ Present address: Department of Molecular Cell Biology, University of Utrecht, Netherlands.
‖ To whom reprint requests should be addressed.

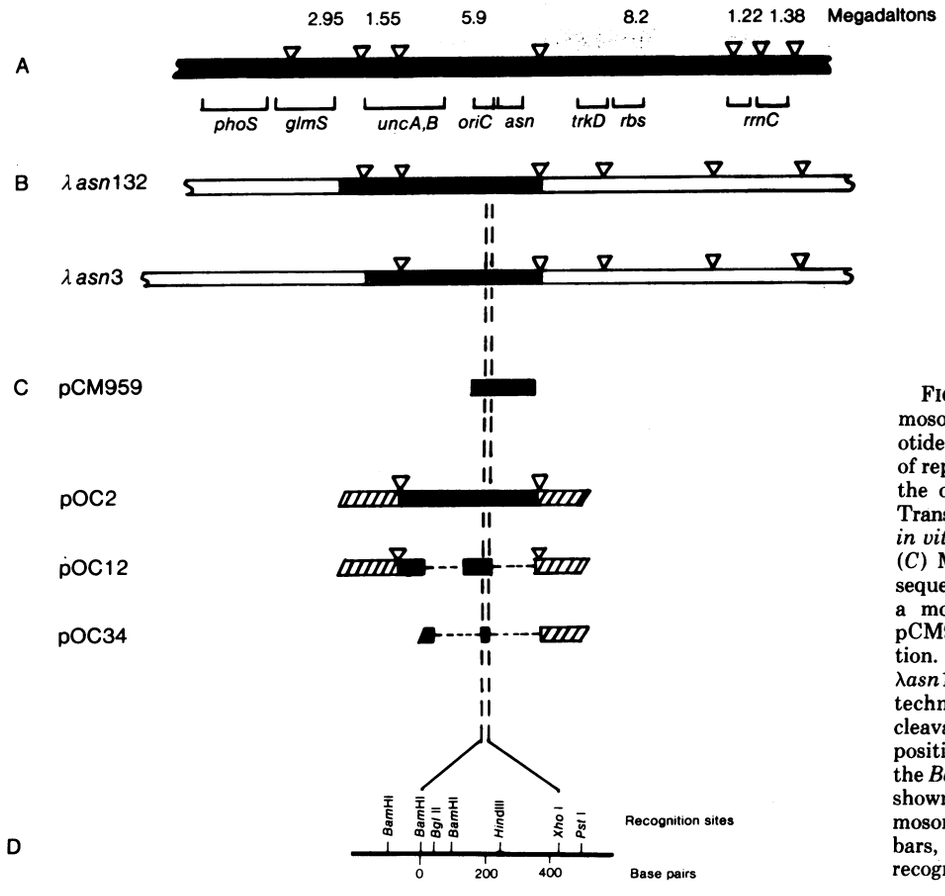

FIG. 1. Transducing phages and minichromosomes used for the definition and the nucleotide sequence analysis of the *E. coli* K-12 origin of replication. (*A*) Genetic and physical map of the origin region of *E. coli* K-12 (1, 5). (*B*) Transducing λ*asn* phages used in the *in vivo* and *in vitro* construction of minichromosomes (1). (*C*) Minichromosomes used in the nucleotide sequence analysis (pCM959 and pOC12) and for a more precise mapping of *oriC* (pOC34). pCM959 is derived from λ*asn*3 by *in vivo* deletion. pOC12 and pOC34 are derived from λ*asn*132 via pOC2 (6, 7) by *in vitro* cloning techniques. (*D*) Restriction endonuclease cleavage map of the origin region of *E. coli*. The position 0 in the base pair scale was assigned to the *Bam*HI site at at the left end of the segment shown to contain *oriC* (see *C*). Solid bars, chromosomal DNA; empty bars, λ DNA; hatched bars, Ap (ampicillin) fragment; ▽, R·*Eco*RI recognition sites.

RESULTS

***In Vivo* Construction of Minichromosomes Consisting Exclusively of Chromosomal DNA.** Specialized transducing phases λ*asn* carrying *oriC* yield, upon infection into a λ lysogenic *recA asn* strain (CM927), asparagine-independent transductants at a high frequency. This is due to their ability to establish themselves as autonomously replicating plasmids, pλ*asn oriC* (1). A high proportion of asparagine-independent transductants from the nonlysogenic strain CM924 (*recA asn*) also appear to harbor the infecting λ*asn oriC* as a plasmid. These single lysogenic transductants are heat-sensitive due to the presence of the right arm of the λ genome containing the *cI*857 allele coding for the heat-labile repressor. This gives the opportunity to isolate mutants of such plasmid phages that have deleted part or all of the λ DNA by selection for heat resistance. Amongst 100 clones, isolated from such asparagine-independent transductants of strain CM924 containing pλ*asn oriC* at 42°C in the absence of asparagine, we found 3 mutants that harbored very small plasmids. One of these plasmids, denoted pCM959 (derived from λ*asn*3, Fig. 1*B*), has a molecular weight of 2,650,000 (4.0 kilobase pairs). Determination of restriction endonuclease cleavage sites led to the conclusion that pCM959

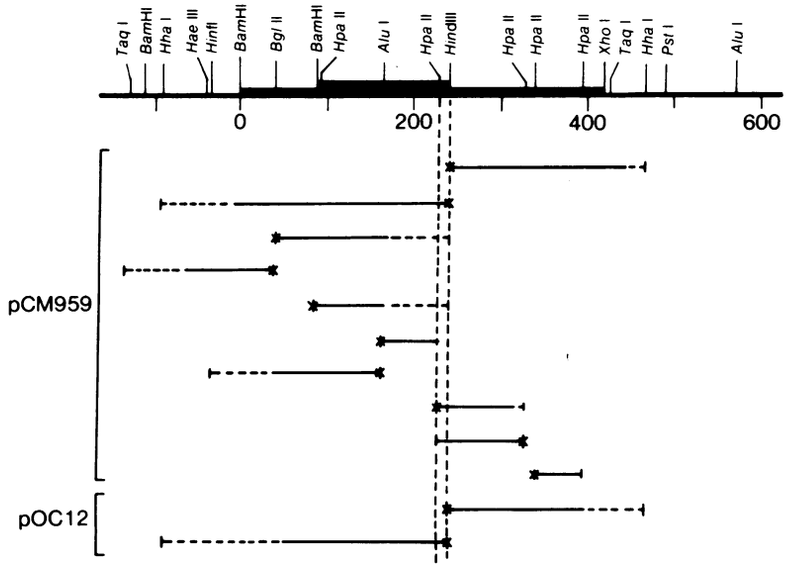

FIG. 2. Restriction endonuclease recognition sites in the *oriC* region and fragments used for nucleotide sequence analysis. The minimum and maximum sizes of the *oriC* segment are indicated by thick lines. Polynucleotides used for the sequence analysis are shown with an * at their labeled 5′ ends and a bar at the site of secondary cleavage. The extent to which the sequence could be read is indicated by solid lines.

(Fig. 1*C*) had deleted all of the λ DNA and thus consists exclusively of chromosomal DNA carrying the *asn* gene and the *oriC* site. The copy number of these small minichromosomes was estimated to be equal to or more than 20 per chromosome equivalent (2).

pOC12 Plasmid. pOC12 was used as an alternative source of DNA for the nucleotide sequence analysis. This plasmid was derived from pOC2, a plasmid that consists of the *Eco*RI fragment carrying the origin region and an *Eco*RI fragment that carries an ampicillin resistance gene (ref. 6; Fig. 1*C*); pOC2 DNA was digested with *Pst* I, and the fragments were religated and transformed into the *recA recB* strain JC7516. pOC12 was isolated from one of the ampicillin-resistant clones selected. This plasmid has retained the *Pst* I fragment carrying the *oriC* site and the *Pst* I fragment that carries the ampicillin-resistance gene (refs. 6 and 7; Fig. 1*C*). Like pCM959, pOC12 DNA can readily be isolated in large amounts, but similarly selective conditions have to be imposed for plasmid maintenance.

Fine Mapping of the Region Containing *oriC*. Fig. 1*D* shows a map of the restriction sites found in the *oriC* region. *oriC* has been located between a *Bam*HI site around coordinate 0 and the *Xho* I site at coordinate 420 (6). There are two additional *Bam*HI sites in close proximity (coordinates −110 and +100, see also Fig. 2). To define precisely the origin region, deletions were made from a derivative of pOC2, by means of *Bam*HI, analogously to the construction of pOC12. Analysis of the deletions generated (e.g. pOC34 in Fig. 1*C*) shows that all DNA to the left of the middle *Bam*HI site (coordinate 0 in Fig. 1*D* and Fig. 2) can be deleted. None out of 19 deletions analyzed deleted material to the right of this *Bam*HI site. pOC34 has deleted DNA both to the left of the *Bam*HI site and to the right of the *Xho* I site. Thus we conclude that *oriC* is located between this *Bam*HI site, to which we assign the coordinate 0, and the *Xho* I site at coordinate 420 (Fig. 1*D*, Fig. 2) and that this segment is sufficient for replication of adjacent DNA.

Nucleotide Sequence of the Replication Origin. The nucleotide sequence of the region containing *oriC* from pCM959 was determined (13). The polynucleotides used for the sequence determination, the restriction sites used for 5′-terminal labeling and for secondary cleavage, and the extent and direction of sequence determination are indicated in Fig. 2. Sequences between positions 39 and 333 were determined from both DNA strands and are considered to contain no mistakes. The outside sequences (positions −51 to +38 and 334 to 470) determined from only one strand may contain 1–2% mistakes.

Part of the sequence (positions 60 to 400) was also determined from pOC12 DNA (Fig. 2). Within the area in which nucleotide sequences from both pCM959 and pOC12 were obtained the sequences of both origins were identical. These two *oriC* plasmids have been derived by entirely different routes from different *asn* transducing phages, λ*asn*3 and λ*asn*132, respectively (Fig. 1*B*). Fig. 3 shows the entire sequence of 519 base pairs.

DISCUSSION

We have located the origin of replication of the *E. coli* K-12 chromosome within a segment of 422 base pairs that is flanked by a *Bam*HI site at position 1 and a *Xho* I site at position 422. This segment is contained in DNA cloned from the region between the *uncB* and *asn* genes (Fig. 1 *A* and *D*) of the chromosome of a sibling of strain W3110. The nucleotide sequence of this region plus some adjacent DNA (519 base pairs total, Fig. 3) was determined by using pCM959 DNA. Part of the sequence (340 base pairs) was also determined for pOC12 and found to be identical.

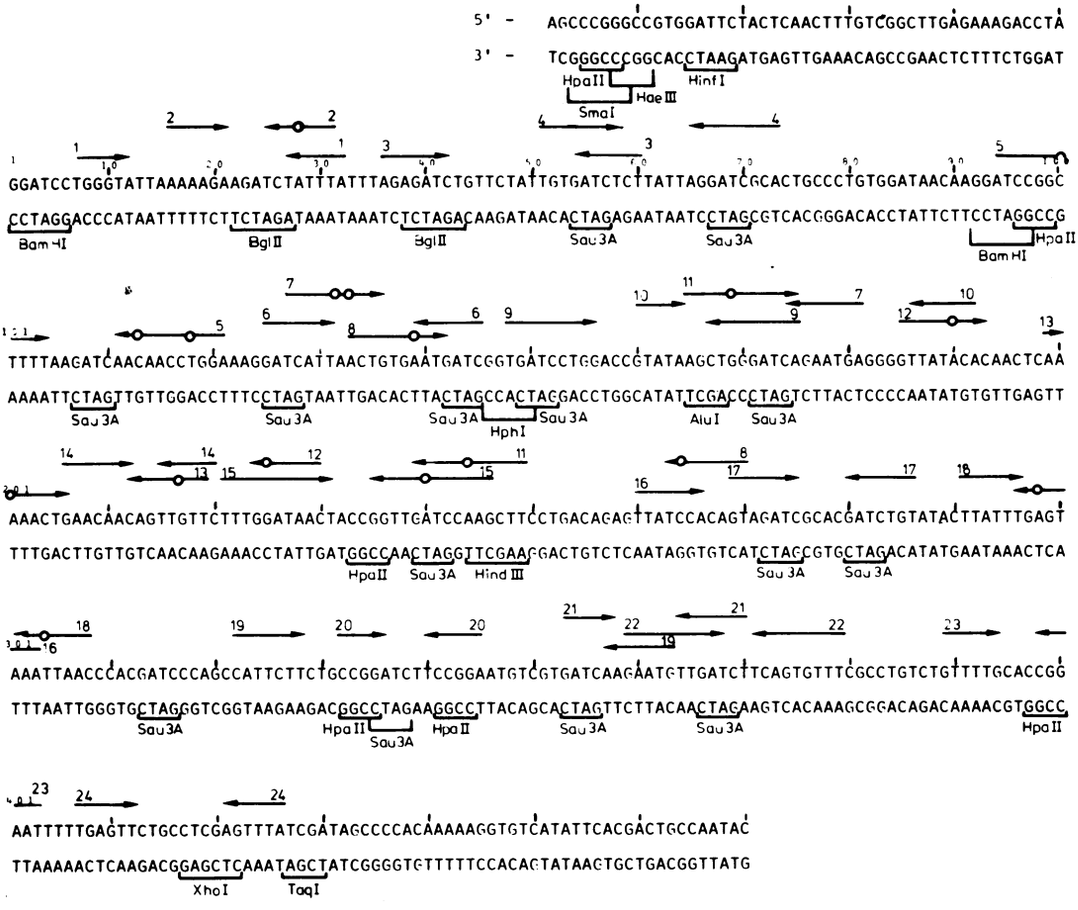

FIG. 3. Nucleotide sequence of the replication origin of *E. coli* K-12. Inverted repeats are indicated by arrows. Mismatches within inverted repeats (O) that are due to inserted bases are shown in only one arrow. The base pair scale is as in Fig. 2; the orientation is as in the *E. coli* genetic map (15). See also Fig. 1*A*.

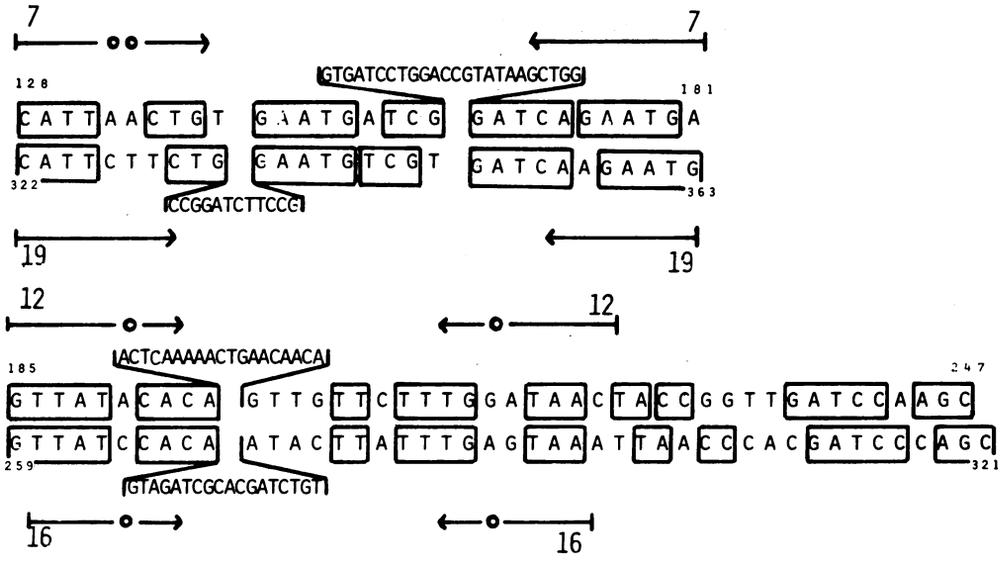

FIG. 4. Alignment of directly repeated sequences from the "upper" strand. Numbers refer to the base pairs and to the inverted repeats in Fig. 3.

The precise limits of the DNA segment comprising *oriC* have not yet been established. However, we may conclude that a core segment that extends from position 92 (*Bam*HI site) to position 249 (*Hin*dIII site) is required for *oriC* function because in the course of the construction of deletions in minichromosomes *in vitro* (7) we have never observed clones in which the context of the chromosomal DNA either at this *Bam*HI site (position 92) or at the *Hin*dIII site (position 249) had been disrupted. To what extent the sequences from positions 1 to 92 and 249 to 422 form part of the replication origin is not yet established. We conclude that the size of the *oriC* segment is at least 158 and at most 422 base pairs, positions 92–249 and 1–422, respectively (see Fig. 2).

In the accompanying paper (8) the nucleotide sequence of a portion of the *E. coli* chromosome, cloned as the *oriC*-containing region, is presented. This portion of the chromosome includes *oriC* as defined by our work (6, 7). Within the segment shown to contain *oriC* (positions 1–422) the nucleotide sequences are identical despite the fact that the respective DNA fragments have been cloned from derivatives of two lines of *E. coli* K-12 (W208, ref. 3; W3110, ref. 1) that have been separate since at least 1944 (10).

Replication of *oriC* plasmids has been found to cease rapidly after inhibition of transcription by rifampicin or after a shift to nonpermissive temperature of *dnaC* and *dnaA* mutants (2).

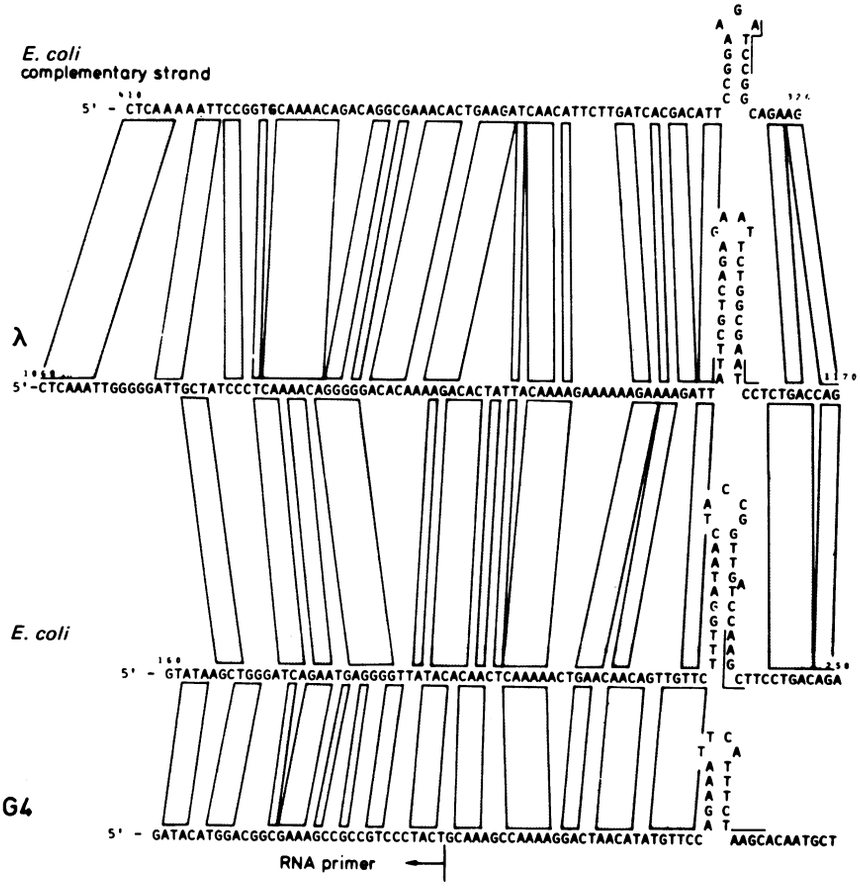

FIG. 5. Homologies between nucleotide sequences of the replication origins from *E. coli*, λ, and from the origin of complementary strand synthesis in G4. The λ sequence is from ref. 20, and G4 is from refs. 21 and 22.

This shows the dependence of the replication of these plasmids on functional RNA polymerase, *dnaC* and *dnaA* gene products, as is the case for the replication of the chromosome.

The complete origin segment from coordinates 1–422 has a GC content of 42%. There are extended regions rich in A·T pairs—e.g., at the base pair positions 10–35, 101–123, 198–223, and 286–307, with 89%, 74%, 72%, and 86% A·T pairs, respectively. There are no extended GC-rich regions. In some cases blocks of a few G·C pairs are adjacent to blocks of A·T pairs—e.g., around positions 100, 185, 308, and 400.

Most of the origin segment is probably not translated into proteins. There is only one initiation codon at the right end of the sequence (position 361) for which no termination codon is found in this reading frame within the determined sequence. It also appears to be the only initiation codon that is preceded by a potential ribosome-binding site at positions 345–356 (16–18). The right-hand part of the sequence could also code for the COOH-terminal portion of a polypeptide reading in from the right and terminating at position 293. The longest possible polypeptide that could be coded by the other parts of the sequence is 49 amino acids long (positions 279–132).

The nucleotide sequence of the *E. coli* replication origin is characterized by a high degree of repetitiveness. On the one hand this is evidenced by numerous inverted repeats, which are indicated in Fig. 3. Inverted repeats that are probably too distant for a successful folding of the single strands into secondary structure are included as well. On the other hand, we found extended stretches of direct, though not perfectly matching, repeats. Two different blocks of repeated sequences become apparent from the alignment of the sequences presented in Fig. 4, which also reveals that the homology is interrupted at several sites by "inserted" sequences, two of which contain inverted repeats. These repeated sequences are arranged approximately symmetrically around the *Hin*dIII site (position 244). The high degree of symmetry, direct and inverted repeats, in this DNA segment is indicative of the presence of several similar sequences which may represent repetitious sites for interaction with proteins involved in the initiation of replication.

Portions of the *oriC* segment show a remarkable degree of homology with the sequence of the replication origin of phage λ (19, 20). Such a region of homology is present twice in the *E. coli* origin, once in the strand reading 5′ → 3′ in the orientation of the *E. coli* genetic map ("upper" strand in Fig. 3, base pairs 160–258) and once in the complementary strand (3′ → 5′, "lower" strand in Fig. 3, base pairs 430–326)—i.e., oriented in an antiparallel way (Fig. 5). Again these segments are arranged roughly symmetrically around the central part of the sequence containing *oriC*. The homologous segment is the same region in which the λ replication origin, and thus also *oriC*, shows homology to the origin of complementary strand synthesis in phage G4 (refs. 21 and 22; Fig. 5). There are additional, though less extensive, regions of homology between *oriC* and the λ origin. No obvious homologies exist between the DNA sequence shown here and the sequences of the replication origins of ColE1 plasmids (23, 24), phage fd (14), or phage φX174 (25).

Sequence homology between the G4 and λ replication origins has been taken to indicate recognition sites for identical or homologous initiation factors, in this specific case for primase (*dnaG*; refs. 21 and 22). In extension, it is tempting to speculate that there are two primase interaction sites in the *E. coli* replication origin, implying that the *dnaG* gene product is responsible for the origin-RNA synthesis (26). The involvement of RNA polymerase in the initiation of replication (2, 27, 28) may then be an indirect one—e.g., as has been suggested in the case of λ initiation through transcriptional activation (29).

Previously determined nucleotide sequences of replication origins have been for phages; the sequence presented here is of a cellular organism. The knowledge of the primary and potential secondary structure of *oriC* can now serve as a basis for a detailed biochemical and genetic analysis of the initiation of replication of the *E. coli* chromosome.

We thank Susanne Bayer, Sylvia Ebel, Michael Hearne, Lissie D. Nielsen, and Barbara Zink for expert and dedicated technical assistance. H.E.N.B. was supported by a European Molecular Biology Organization long-term fellowship. Part of this work was supported by a grant from the Stiftung Volkswagenwerk (I/34748) and a grant from the Deutsche Forchungsgemeinschaft (Scha 134/7).

1. von Meyenburg, K., Hansen, F. G., Nielsen, L. D. & Riise, E. (1978) *Mol. Gen. Genet.* **160,** 287–295.
2. von Meyenburg, K., Hansen, F. G., Riise, E., Bergmans, H. E. N., Meijer, M. & Messer, W. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43,** in press.
3. Yasuda, S. & Hirota, Y. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5458–5462.
4. Fayet, O. & Louarn, J. (1978) *Mol. Gen. Genet.* **162,** 109–111.
5. von Meyenburg, K., Hansen, F. G., Nielsen, L. D. & Jørgensen, P. (1977) *Mol. Gen. Genet.* **158,** 101–109.
6. Messer, W., Bergmans, H. E. N., Meijer, M., Womack, J. E., Hansen, F. G. & von Meyenburg, K. (1978) *Mol. Gen. Genet.* **162,** 269–275.
7. Messer, W., Meijer, M., Bergmans, H. E. N., Hansen, F. G., von Meyenburg, K., Beck, E. & Schaller, H. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43,** in press.
8. Sugimoto, K., Oka, A., Sugisaki, H., Takanami, M., Nishimura, A., Yasuda, Y. & Hirota, Y. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 575–579.
9. Hirota, Y., Yasuda, A., Nishimura, A., Takeda, Y., Yamada, M., Sugimoto, K., Sugisaki, H., Oka, A. & Takanami, M. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43,** in press.
10. Bachmann, B. J. (1972) *Bacteriol. Rev.* **36,** 525–527.
11. Roberts, R. J., Breitmeyer, J. B., Tabachnik, N. F. & Meyers, P. A. (1975) *J. Mol. Biol.* **91,** 121–123.
12. Schaller, H., Beck, E. & Takanami, M. (1978) in *The Single-Stranded DNA Phages,* eds. Denhardt, D. T., Dressler, D. H. & Ray, D. S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York), in press.
13. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
14. Gray, C. P., Sommer, R., Polke, C., Beck, E. & Schaller, H. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 50–53.
15. Bachmann, B. J., Low, K. B. & Taylor, A. L. (1976) *Bacteriol. Rev.* **40,** 116–167.
16. Shine, J. & Dalgarno, L. (1975) *Nature (London)* **254,** 34–38.
17. Steitz, J. A. & Jakes, K. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 4734–4788.
18. Taniguchi, T. & Weissmann, C. (1978) *J. Mol. Biol.* **118,** 533–565.
19. Denniston-Thompson, K., Moore, D. D., Kruger, K. E., Furth, M. E. & Blattner, F. R. (1977) *Science* **198,** 1051–1056.
20. Hobom, G., Grosschedl, R., Lusky, M., Scherer, G., Schwarz, E. & Kössel, H. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43,** in press.
21. Fiddes, J. C., Barrell, B. G. & Godson, G. N. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 1081–1085.
22. Sims, J. & Dressler, D. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3094–3098.
23. Bolivar, F., Betlach, M. C., Heyneker, H. L., Shine, J., Rodriguez, R. L. & Boyer, H. W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5265–5269.
24. Tomizawa, J., Ohmori, H. & Bird, R. E. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 1865–1869.
25. Langeveld, S. A., van Mansfeld, A. D. M., Baas, P. D., Jansz, H. S., van Arkel, G. A. & Weisbeek, P. J. (1978) *Nature (London)* **271,** 417–420.
26. Messer, W., Dankwarth, L., Tippe-Schindler, R., Womack, J. E. & Zahn, G. (1975) in *ICN-UCLA Symposia on Molecular and Cellular Biology,* eds. Goulian, M., Hanawalt, P. & Fox, C. F. (Benjamin, Menlo Park, CA), pp. 602–617.
27. Lark, K. G. (1972) *J. Mol. Biol.* **64,** 47–60.
28. Messer, W. (1972) *J. Bacteriol.* **112,** 7–12.
29. Dove, W. F., Inokuchi, H. & Stevens, F. (1971) in *The Bacteriophage Lambda,* ed. Hershey, A. D. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 747–771.