Volume 13 Number 19 1985 Nucleic Acids Research

The distribution and properties of RNA primed initiation sites of DNA synthesis at the replication origin of *Escherichia coli* chromosome

Yuji Kohara, Naoki Tohdoh*, Xiao-Wang Jiang and Tuneko Okazaki

Institute of Molecular Biology, Faculty of Science, Nagoya University, Nagoya 464, Japan

Received 29 July 1985; Accepted 9 September 1985

SUMMARY RNA-linked DNA molecules were obtained from E. coli dnaC^{ts} cells synchronously initiating a new round of chromosome replication. The deoxynucleotides at the transition from primer RNA to DNA were ^{32}P-labeled, and their positions were located on the nucleotide sequence of 1.4 kb genomic region (position -906 to +493) including the oriC and its leftside flanking region. In the r-strand (the counterclockwise strand), many strong transition sites were mapped in the left half portion of the oriC and a few weak sites in the left outside region. In the l-strand (the clockwise strand), no transition sites were found inside the oriC but many weak sites were found in the left outside region. The results support the initiation mechanism in which the first leading strand synthesis starts with the r-strand counterclockwise from the oriC that is followed by the l-strand synthesis on the displaced template strand on the left of oriC. Primer RNA molecules attached to the strong r-strand transition sites were only a few residues in length. Properties of the transition sites were discussed.

INTRODUCTION

DNA replication of E. coli starts from a fixed origin and proceeds bidirectionally. In order to understand the molecular mechanisms by which initiation of DNA replication is specifically started and regulated at the origin, many studies have been conducted. The organization of the origin region of the DNA and the interaction between the origin region of the DNA and the protein factors which functions in the initiation or the elongation steps of DNA replication are as follows: The minimum essential region of DNA required for the origin function (oriC) was narrowed by Oka et al. to the 245 base-pair (bp) stretch which locates at 83.5 min on the E. coli genetic map (1), the area approximately from where the bidirectional replication of E. coli chromosome had been predicted to start by the classical works of Bird et al. and Masters and Broda (2, 3). After intensive investigations, Oka et al. concluded that the oriC sequence is composed of two types of sequence blocks, those which function in specific recognition in the initiation reaction (the recognition sequences) and the others which keep the recognition sequences in the proper special arrangement (the spacer sequences) (4). An initiator

protein, dnaA gene product, was shown recently to bind a 5'-TTAT$^{C}_{A}$CA$^{C}_{A}$A sequence which is repeated four times within oriC and categorized previously as the recognition sequence (5, 6). The in vitro system, which specifically recognize and replicate oriC plasmid, requires dnaA protein together with RNA polymerase and other replication fork proteins such as dnaB, dnaC, dnaG (DNA primase) SSB, DNA gyrase and DNA polymerase III holoenzyme (7).

Despite these findings, the biochemical mechanism by which the first DNA synthesis is primed is still unclear. The first leading-strand DNA, synthesized at the replication origin, is thought to be primed by RNA synthesized by either RNA polymerase or by DNA primase (8). In both mechanisms of the primer synthesis, the transcription by RNA polymerase may be related to the activation of the replication origin (8). The purpose of this work is to gain insight regarding the priming mechanism for the initiation reaction through the mapping analyses of the RNA-DNA transition sites and the 5' ends of primer RNA molecules in the origin region using RNA-linked DNA molecules isolated from the cells synchronously initiating chromosome replication.

Previous analyses from this laboratory revealed that two RNA-DNA transition sites exist within oriC in the r-strand (counterclockwise strand) but none in the l-strand (clockwise strand) (9). Based on these results, the unidirectional start model for the bidirectional replication of E. coli genome were proposed, in which the first DNA synthesis starts counter-clockwise from the oriC (the leading r-strand synthesis), and the second synthesis starts only after the template for the l-strand is displaced as a result of the r-strand synthesis from the left outside area of the oriC. In this study, we improve the sensitivity of the analyses by use of cloned single-stranded probes, and extend the mapping area of transition sites to 0.9 kb on the left of the oriC. We find in the r-strand, clusters of strong transition sites inside the oriC but only several weak sites in the flanking region left of the oriC. In contrast, we find that, in the l-strand, more than ten weak sites are scattered in the flanking region left of the oriC but none inside the oriC. By also mapping the 5'-terminus of primer RNA in the oriC region, we detect the RNA terminus immediately upstream of each strong junction site. The properties of the transition sites and possible initiation mechanisms are discussed.

MATERIALS AND METHODS

Bacteria, phage and plasmid

E. coli PC2 (leu, thy, dra, str, dnaC2), oriC transducing phage λ asn5

and plasmid pNH14 (9), f1 R199 phage and its host K38 (10) have been described.

Reagents

Carrier free [γ-^{32}P]ATP (10 Ci/μmole) was prepared according to Walseth and Johnson (11). Arabinosyl cytosine (1-β-D-arabinofuranosyl cytosine, ara-C) was from Kohjin Co., Japan. EcoRI linker was from Collaborative Research. T4 polynucleotide kinase and T4 DNA ligase (9) and bacterial alkaline phosphatase (12) were described. Bovine pancreatic DNase I was from Worthington Biochemical Corp. Restriction enzyme HaeIII was purchased from Takara Shuzo Co., Japan. EcoRI and HindIII was from Boehringer Mannheim GmbH, FRG. BamHI, BglII, HgiAI, HhaI, HincII and NciI were from New England Biolabs, Inc. RNase T_1 and A were gifts of Dr. T. Uchida. RNase T_2 and nuclease P1 were from Seikagaku Kogyo, Japan.

Preparation of single stranded probes for oriC and its leftside flanking region

The region analyzed in this study is from HaeIII site at -906 to PstI site at +494, according to the numbering by Sugimoto et al. (13). We have determined the sequence from -104 to -906 by the strategy shown in Fig. 1. The region between -104 and -650 has been determined independently by Buhk and Messer (14). The strand whose 5' to 3' polarity is clockwise in the genetic map, is referred to as l-strand, and the complementary strand as the r-strand. The genomic regions corresponding to three HaeIII fragments, 4 (-906 to -607), 2 (-606 to -159) and 6 (-149 to -42) and a HaeIII-PstI fragment (-41 to +494) in Figure 1 lane b are called region 4, 2, 6 and 1, respectively. These fragments were prepared from λ asn5 or pNH14 DNA, inserted into EcoRI site of f1 R199 RF molecules using EcoRI linker and cloned. Since some chimeric phages frequently delete the insertion segment, phages were propagated by the plate-lysate method. The chimeric phages were purified by banding in CsCl gradient and then the phage DNA were extracted by phenol.

Preparation of RNA-linked DNA molecules

The procedure is essentially the same as described by Hirose et al. (9). E. coli PC2 (dnaC2) cells were cultured at 28°C in 1 l of Tris-buffered medium (16) containing 2 μg thymine/ml until a cell density of 3 x 10^8 per ml was reached. The culture was divided into 10 portions, shifted up to 40°C for 50 min, and then arabinocyl cytocine was added at the final concentration of 1 mg/ml. Ten min later, the temperature of the culture was rapidly lowered to 28°C and incubated for 10 min, then DNA synthesis was stopped by

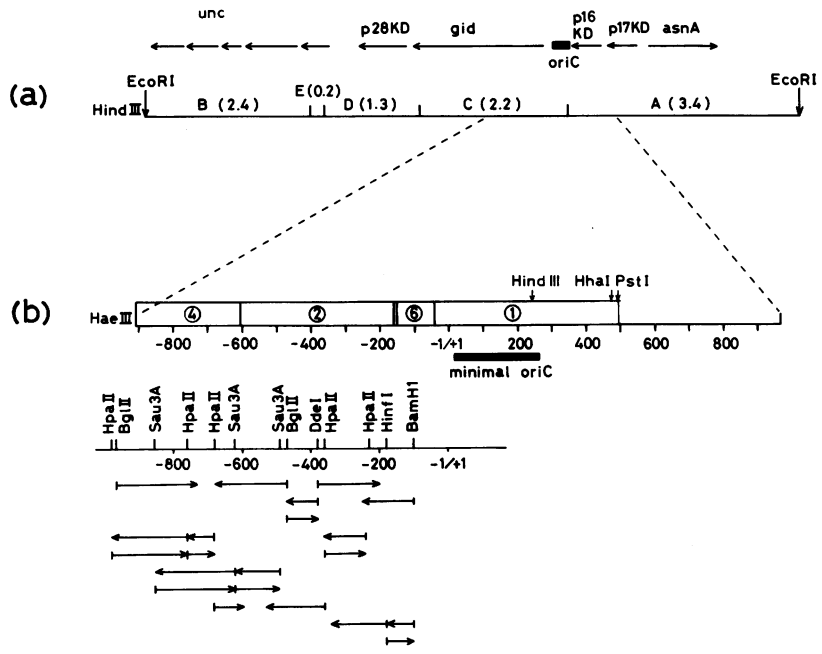


Fig. 1. Physical map around the oriC.
(a) The HindIII map of the oriC containing an 8.6 kb EcoRI fragment (15). The number in parenthesis represents the size of each HindIII segment in kb. The arrows show the coding region and the direction of the transcription (38).
(b) HaeIII map around the oriC. HaeIII segments (regions 2, 4 and 6) and HaeIII-PstI segment (region 1) used in this study, and some restriction enzyme cleavage sites are indicated. The numbers below the map are the nucleotide position numbers according to Sugimoto et al. (13). The arrows below show the sequencing strategy.

mixing the culture with an equal volume of the ethanol-phenol mixture. As described previously (12), after extraction, DNA molecules less than 3000 nucleotides long were purified by: sedimentation through a neutral sucrose gradient; nitrocellulose column chromatography with the exception that the charging buffer contained 1.5 M NaCl, and Cs_2SO_4 equilibrium centrifugation (short DNA fraction). The fraction contained RNA-linked DNA molecules (RNA-DNA molecules) as well as RNA-free molecules.

Methods for mapping RNA-DNA transition sites and 5'-terminus of the linked RNA

(Method A) The short DNA fraction was phosphorylated with cold ATP and T4 polynucleotide kinase, digested in alkali and then the 5'-OH ends of DNA, exposed at RNA-DNA transition sites of RNA-DNA molecules, were selectively phosphorylated with [γ-^{32}P]ATP (10 Ci/μmole) and T4 polynucleotide kinase as

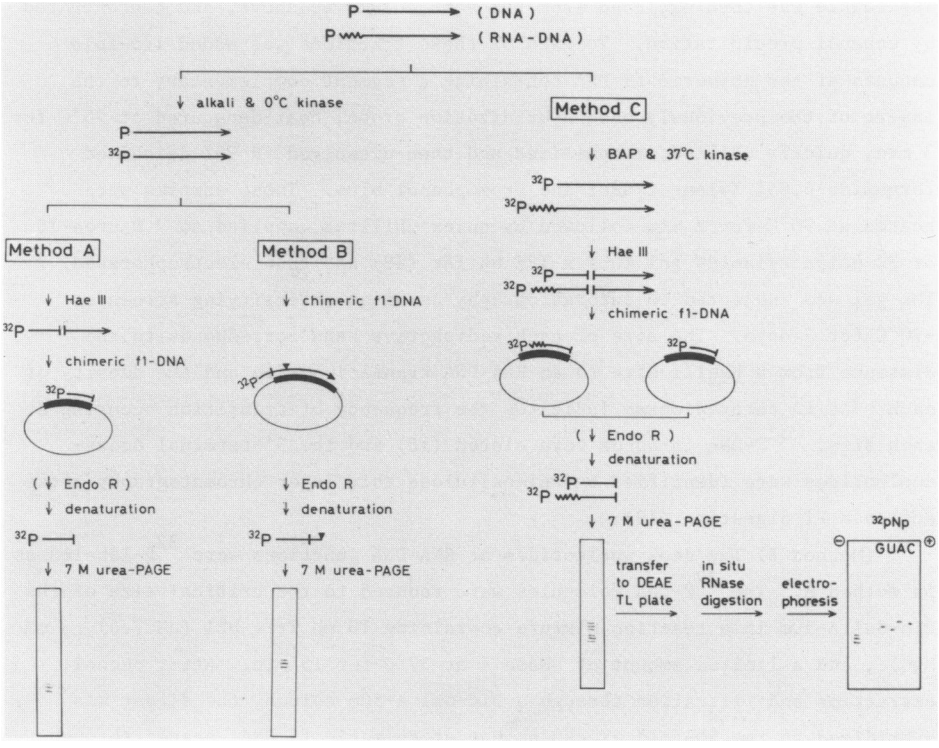


Fig. 2. Strategy for mapping RNA-DNA transition sites (methods A and B) and 5'-terminus of the linked RNA (method C). See Materials and Methods and text for detail.

described previously (17) except that [γ-^{32}P]ATP was used at about 3 µM. Resulting ^{32}P-DNA were digested with 125 units HaeIII/ml for 12 hr to make unique 3'-ends, and when specified with 100 units HhaI/ml. The digests were filtered through Bio-Gel A-50m columns after phenol extraction. The included fractions were pooled, concentrated by ethanol precipitation and then divided into eight portions. Each portion was mixed with 2.5 pmole of one of the chimeric f1 DNA probes (more than 20-fold excess amount), heat-denatured at 95°C for 3 min and quickly chilled. NaCl was then added to a concentration of 0.5 M, volume was adjusted to 300 µl and incubated at 65°C for 10 min for hybridization. The mixtures, after incubation, were filtered through a Bio-Gel A-50m column and the excluded fractions containing hybridized molecules were pooled. To remove the molecules not specifically interacted with the probe DNA, each pooled fraction was concentrated to 300 µl, heat denatured and then hybridization process was repeated at 65°C for 12 min.

The sample was then excluded from the same column as above, and concentrated by ethanol precipitation. To each of these fractions was added 1.5-fold amounts of the chimeric f1 DNA containing a segment complementary to the insert of the previously used hybridization probe, heat-denatured at 95°C for 3 min, quickly chilled, lyophilized and then dissolved in 98% deionized formamide-0.05% Xylene Cyanol and Bromophenol blue. These samples were heated at 90°C for 2 min followed by quick chilling, applied to 7 M urea-10 or 8% polyacrylamide gel in 2 x TBE buffer (18) and then electrophoresed. The gel was subjected to autoradiography using an intensifying screen at -70°C for 3 days. The size of each radioactive band corresponds to the distance from a HaeIII site to an RNA-DNA transition site and the density of each band in autoradiogram indicates the frequency of transition occuring at each site. ^{32}P-DNA in bands were eluted (18) and the 5'-terminal deoxynucleotides were identified by PEI-cellulose thin layer chromatography after nuclease P1 digestion (19).

(Method B) The deoxynucleotides at RNA-DNA junctions were ^{32}P-labeled as in method A. The ^{32}P-DNA molecules were reduced to the critical size of the Bio-Gel A-50m in a reaction mixture containing 10 mM Tris-HCl (pH 7.6), 3 mM $MgCl_2$, and a limited amount of DNase I at 37°C for 15 min. After phenol extraction and filtration through a Bio-Gel A-50m column, the digest was hybridized to the 50-fold excess amount of chimeric f1 DNA, passed through the Bio-Gel A-50m column. The excluded fractions were heat-denatured and the hybridization process was repeated as described in method A. The resulting molecules were cleaved by an appropriate restriction enzyme and processed as in method A for gel electrophoresis and autoradiography.

(Method C) The short DNA fraction was dephosphorylated by bacterial alkaline phosphatase and then ^{32}P-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase as previously described (12), with the exception of a 3 μM concentration of [γ-^{32}P]ATP. The resulting 5'-^{32}P-nucleic acids were further purified by Cs_2SO_4 equilibrium centrifugation, and then digested with 125 units HaeIII and 100 units HhaI/ml in the presence of 1 mg/ml of carrier tRNA for 12 hr. As in method A the digest was processed and hybridized with chimeric f1 DNA containing r- or l-strand of region-1. The 5'-end of primer RNA was then mapped by the two dimensional electrophoresis by the method described below: 1st dimension; The hybridized molecules were digested with BglII or BamHI when indicated and electrophoresed as in method A. 2nd dimension; ^{32}P-nucleic acids were transferred from the gel of the first-dimension electrophoresis to a DEAE-cellulose thin layer plate (Polygram CEL

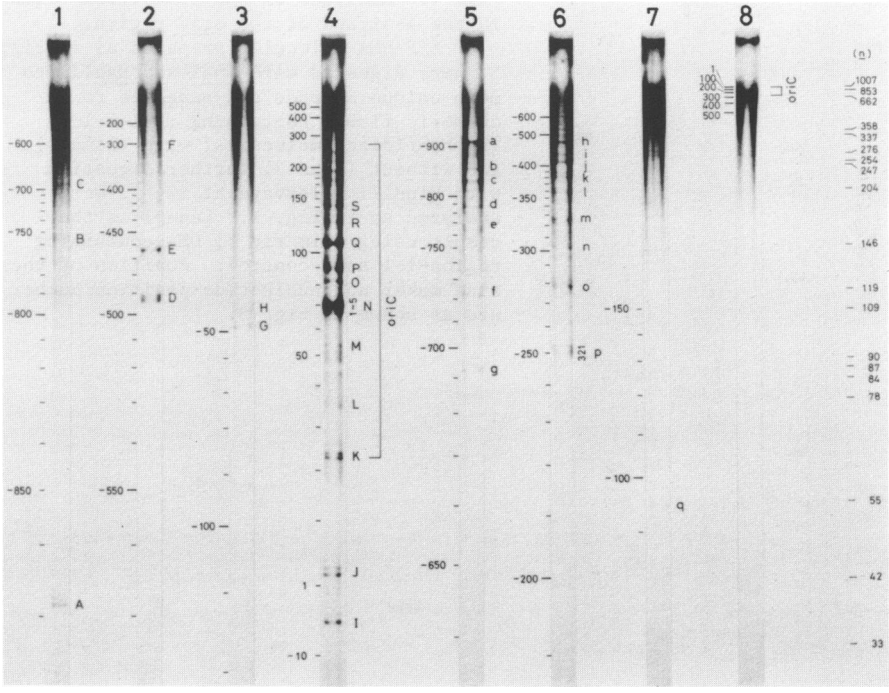


Fig. 3. RNA-DNA transition sites in oriC and its vicinity to the left.
RNA-DNA transition sites in each region were detected by method A as described in Materials and Methods. Positions of size makers (AluI digests of ϕX174 RF DNA) are shown on the right margin. The scale, left of each lane, represents the nucleotide position number deduced from the distance of migration. Capital and small letters represent bands in r-strand (lane 1 to 4) and in l-strand (lane 5 to 8), respectively. Lanes 1, 5 (region-4), lanes 2, 6 (region-2), lanes 3, 7 (region-6) and lanes 4, 8 (region-1). Numbers to the right of lanes 4 and 6 shows the ladder numbers whose 5' terminal nucleotides were analyzed in the experiments shown in Fig. 6.

300 DEAE, 40 x 20 cm) according to Tanaka et al. (20), except for the following modifications; (i) the gel strip was 1 cm wide, (ii) 2 M ammonium acetate buffer (pH 4.5) was used for transfer of nucleic acids longer than 40 nucleotide residues from the gel, and (iii) the origin line of the plate contacted with gel was kept wet through out the process with the same buffer. The blotted samples were digested in situ by mixture of RNase T_2, and T_1 and A, and then thin layer electrophoresis was performed as described (20), except that RNase digestion was carried out for 12 hr.

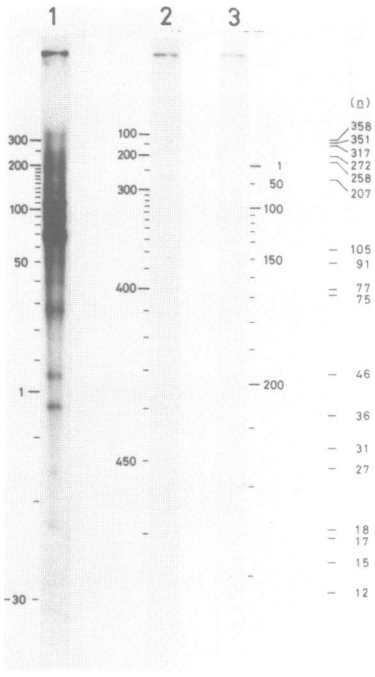


Fig. 4. Analyses of the transition sites in the 1-strand of the oriC region. ^{32}P-DNA molecules prepared as in Fig. 3, were digested with HhaI and HaeIII to make unique 3'-ends and annealed to chimeric f1 DNA containing region-1-r. The hybridized molecules, with (lane 2) and without (lane 3) further digestion with HindIII (cleavage at +245), were analyzed as in Fig. 3. Lane 1 is the result using chimeric f1 DNA containing region-1-1 as a control. Position of the size maker and nucleotide position number are as shown in Fig. 3.

RESULTS

Mapping of RNA-DNA transition sites

To analyze initiation fragments, short DNA molecules, containing also RNA-linked DNA molecules, were purified from E. coli PC2 (dnaC2) cells, which were synchronously initiating the new round of chromosome replication by temperature shifting, in the presence of arabinosyl cytocine (1 mg/ml). By the addition of arabinosyl cytocine, nascent short fragments derived from the oriC region, has been shown to accumulate to about 5-fold (9).

Fig. 3 shows the results of the analysis of RNA-DNA transition sites obtained by method A. In r-strand, strong bands were found clustered in the region-1, especially in the left half of oriC (lane 4), while only several weak bands were detected in the other regions (lanes 1, 2 and 3). The two strongest bands, bands N and Q, correspond to the sites 2 and 1 previously detected by Hirose et al. (9), respectively. In 1-strand, many weak bands were detected in region-4 and -2 (lanes 5 and 6), and very weak bands in the region-6 (lane 7). As for region-1, the ^{32}P-DNA derived from this region, if any, should migrate more slowly than 500 residues DNA in the gel (lane 8), due to having a unique 3'-end at +974 after HaeIII cleavage and the f1

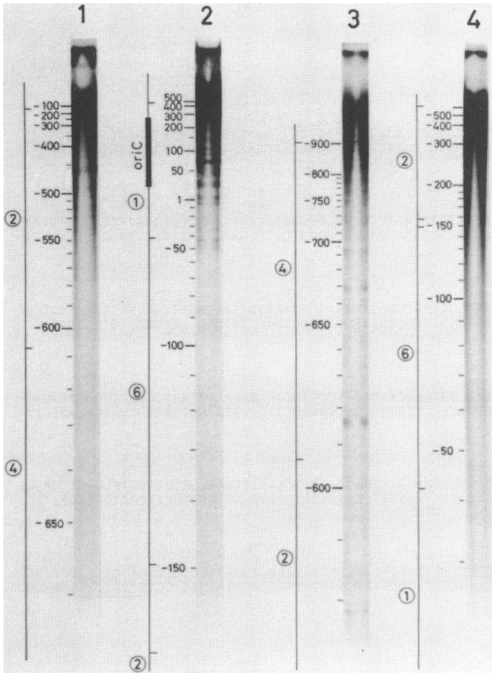


Fig. 5. Detection of junctions in the border areas.
Junctions were analyzed by method B as described in Materials and Methods. Each chimeric f1 DNA sample used in hybridization for lanes 1, 2, 3 and 4 has an insertion of region-4-1, 2-1, 2-r and 1-r, respectively. After hybridization, samples for lanes 1, 2, 3 and 4 were cleaved by NciI (site at -686), HincII (at -176), HgiAI (at -542) and BamHI (at +2), respectively. The nucleotide position numbers, deduced from the distance of migration relative to that of the size makers (not shown), are shown on the left of each lane.

chimera DNA used in hybridization contained DNA segment from -41 (HaeIII) to +494 (PstI) (Fig. 1). To make the 3'-end more proximal, the ^{32}P-DNA molecules were digested with both HhaI (cleavage site at +478) and HaeIII, and then processed as above. As shown in Fig. 4 lane 2, no bands were detected. Even after further shortening of hybridized molecules with HindIII (cleavage site at +245), no bands were visible (Fig. 4, lane 3). On the control gel in which the opposite r-strand was run after the same treatment, the same bands as shown in Fig. 3 were again detected (lane 1).

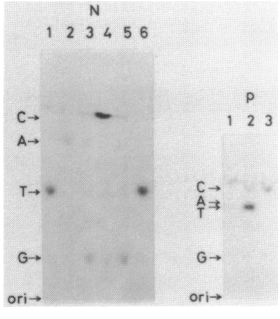
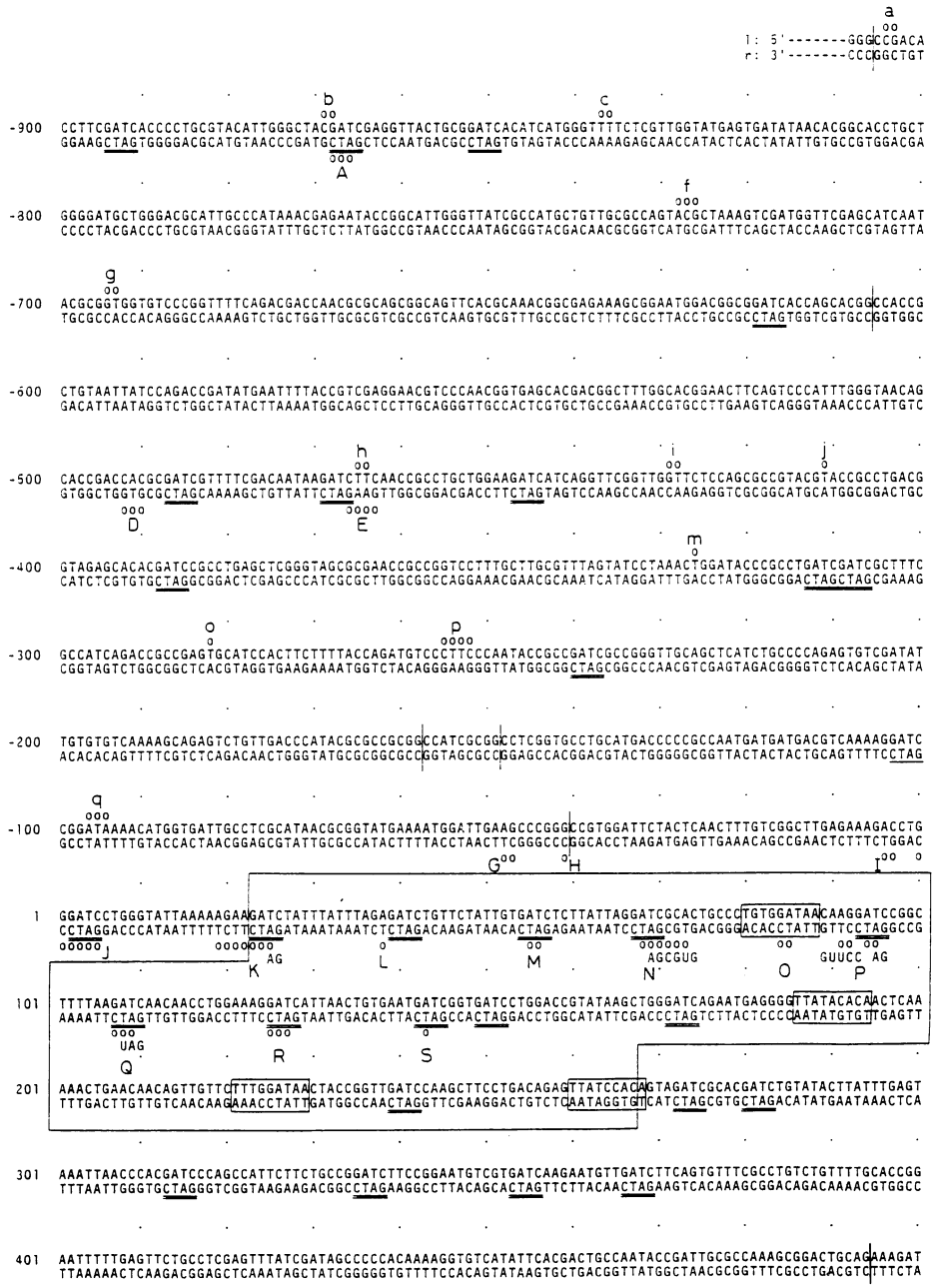


Fig. 6. Identification of deoxynucleotides at the junctions.
^{32}P-DNAs in radioactive ladders (indicated with numbers) at sites p and N in Fig. 3 were eluted separately, digested with nuclease P1 and then subjected to PEI- cellulose thin layer chromatography. The positions of pN are shown by G, A, T and C.



The transition sites in the region very close to HaeIII sites were examined by method B. For example, to map such sites in the r-strand of region-6 and region-1, the ^{32}P-DNA were annealed to chimeric f1 DNA containing l-strand of region-2 and cleaved with HincII (cleavage site at -176). By this procedure, strong and weak bands in region-1 and -6 were confirmed, but no additional strong bands were detected with the exception of a weak band at about -34 (Fig. 5, lane 2). Similar analyses, performed on the border region near the other HaeIII sites, showed no additional strong transition sites (Fig. 5, lanes 1, 3 and 4).

To determine the nucleotide position of each junction site precisely, the ^{32}P-DNA of each band in Fig. 3 were eluted. Since the resolution in the upper part of the gel is low, ^{32}P-DNA were also eluted from the corresponding bands shifted downward by second cleavage with an appropriate restriction enzyme. The ^{32}P-DNA were completely digested with nuclease P1 to mononucleotides, and the 5'-terminal nucleotides were identified by PEI-cellulose thin layer chromatography. Results of representative sites are shown in Fig. 6.

In Fig. 7, the junction sites thus determined are shown in the nucleotide sequence. Transition occurred at two to seven consecutive nucleotides at most sites. The relative frequency of transition occurred at each site is illustrated in Fig. 8. The distribution of junction sites between the complementary strands is very asymmetrical.

Mapping of 5'-terminus of primer RNA in oriC region

The short DNA molecules contain both RNA-linked DNA (<10%) fragments and RNA-free DNA fragments (>90%), which may be generated by removal of the RNA portion of RNA-DNA molecules and/or by in vivo degradation of DNA. When all 5'-terminus of the short DNA molecules are ^{32}P-labeled by T4 polynucleotide kinase after dephosphorylation, a sensitive method to distinguish the molecules terminated with ribonucleotide from those with deoxynucleotide is required for mapping of RNA terminus. For this purpose, we have developed a modified application of the method used for a RNA sequencing (20), two dimensional electrophoresis (Fig. 2, method C). The results of the two

Fig. 7. Location of transition sites and 5'-terminus of the linked-RNA in the nucleotide sequence of oriC and the left vicinity.

The transition sites in l-strand and r-strand are marked with (o) above and below the sequence for each strand. Letters correspond to the bands in Fig. 3. 5'-terminus of the linked-RNA in the oriC region were detected only in the r-strand. The bases identified as ^{32}pNp are shown under the sequence. Minimal oriC from +23 to +267 is boxed. GATC sequence (═══) and dnaA protein recognition sequence, TTAT$^{C}_{A}$CA$^{C}_{A}$A (5), (▭) are marked. ([ ]) shows the borders of genomic segments inserted into the probe chimeric f1-DNA.

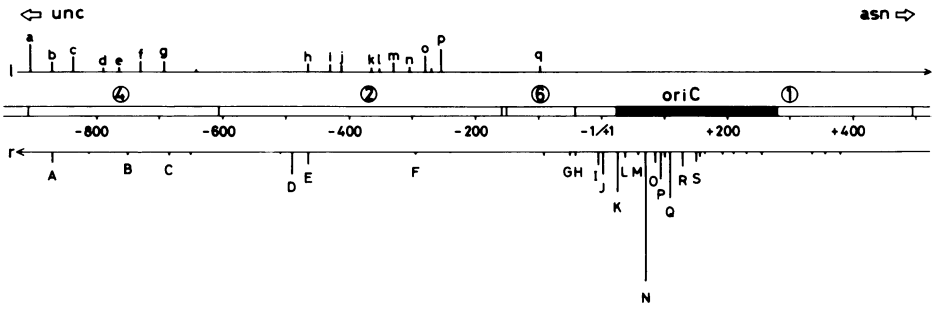


Fig. 8. A histogram of relative frequency of RNA-DNA junction sites.
The length of vertical lines represents the relative frequency of junctions deduced from the tracing of bands in the autoradiograms of Fig. 3 by densitometer. For the sites where RNA-DNA switching occurred at several consecutive nucleotides, the junction frequency of these nucleotides were summed and indicated by one line. Upper: transition sites in the l-strand, whose direction of DNA synthesis is to the right, or clockwise. Lower: those in r-strand, the counterclockwise strand. Letters correspond to the bands in Fig. 3.

dimensional electophoresis are shown in Fig. 9. Several ^{32}pNp spots were detected at or around the nucleotide positions of +5, +25 and +70 (plate a). Since nucleic acids longer than 100 nucleotide residues were less efficiently transferred from gel to the thin-layer plate, the ^{32}P-nucleic acids were shortened by digesting the hybridized molecules with BglII (cleavage site at +23 and +39) (plate b) or BamHI (at +2 and +93) (plate c) resulting in RNA terminus in the oriC region being identified accurately. The 5'-termini of the RNA-linked DNA molecules deduced from the location of ^{32}pNp spots in Figure 9 are depicted in Fig. 7. All RNA termini were located immediately upstream of the RNA-DNA transition sites, suggesting that the RNA portion was very short. To measure the length of the RNA portion of RNA-DNA molecules, ^{32}P-nucleic acids were eluted from the corresponding bands in the gel of the first dimension, digested with 3' to 5' exodeoxynuclease of T4 DNA polymerase to remove DNA moieties, and then analyzed by polyacrylamide gel electrophoresis. No long RNA molecules were detected (data not shown). Thus, the results, together with the location of 5' ends of RNA determined by the mapping experiments, indicate that the RNA are, at most, a few nucleotide residues in length. In the complementary l-strand of this region, no RNA termini were detected (data not shown).

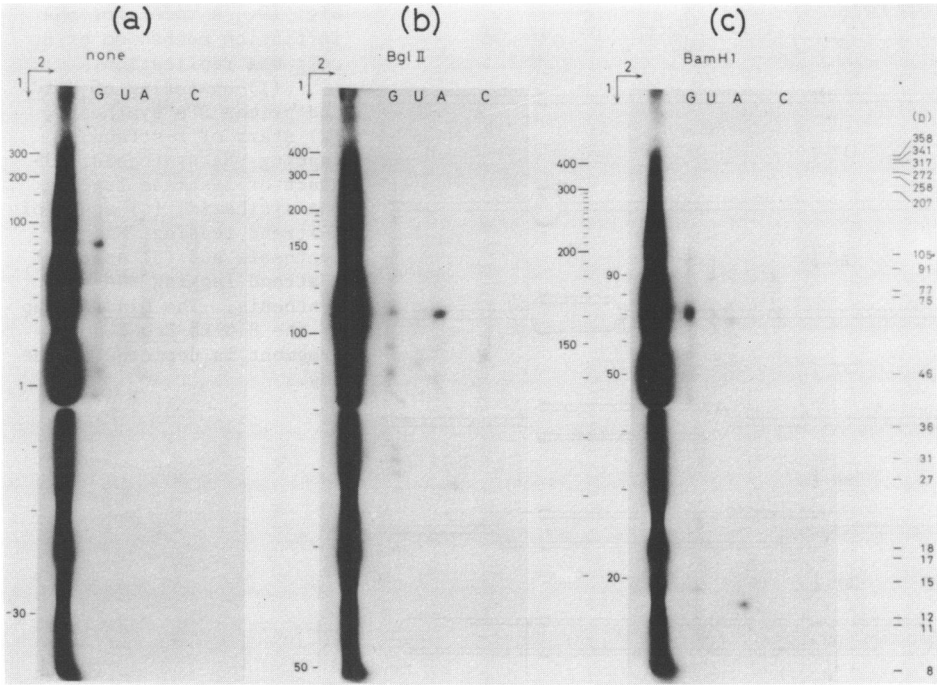


Fig. 9. Detection of 5'-termini of the linked-RNA in the oriC region. RNA termini were detected by method C as described in Materials and Methods. The hybridized molecules were digested in the absence of enzyme (plate a), with BglII (plate b) or with BamHI (plate c). 1st dimension: 7 M urea-10% polyacrylamide gel electrophoresis. 2nd dimension: DEAE-cellulose thin layer electrophoresis. The positions of the size markers are shown on the right. The nucleotide position numbers for each lane are shown as in Fig. 3. Since BamHI cleaved at two sites 91 nucleotides apart, plate c has 2 scales for position numbers. The positions of 4 pNps are shown on the top by G, U, A and C.

DISCUSSION

We have mapped transition sites from primer RNA to DNA synthesis, onto the 1.4 kb region, including oriC and its leftside flanking region of E. coli genome. The distribution of junctions is very asymmetrical between the complementary strands, as shown in Fig. 8. In the r-strand (the counter-clockwise strand), strong junctions are clustered on the left half of oriC in the region-1, while only a few weak junctions were detected in other regions. Hirose et al. had previously detected two junction sites in the r-strand of region-1 (region-1-r) (9): sites N and Q of the present analysis correspond to sites 2 and 1 of their work, respectively. Additional junctions were detected in the region-1-r in this study, probably because of improved

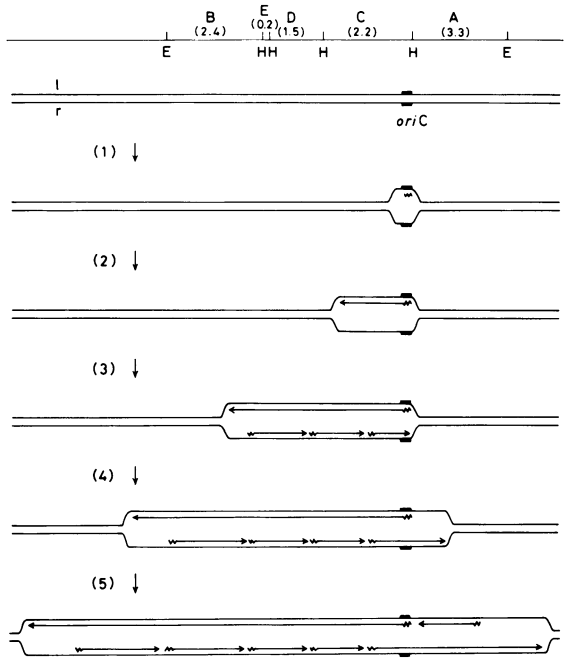


Fig. 10. A model for the initiation mechanism of E. coli DNA replication. (1) unwinding of oriC and primer RNA synthesis, (2) start of r-strand leading DNA synthesis, (3) start of l-strand lagging DNA synthesis, (4) start of l-strand leading DNA synthesis and (5) start of r-strand lagging DNA synthesis. The HindIII map of the 8.6 kb EcoRI fragment is depicted on the top.

sensitivity of the mapping method. Even with this improved method, no junction sites have been detected in the complementary l-strand of region-1 (region-1-1), confirming the previous results of Hirose et al (9). While in the left flanking region (region-4, 2 and 6) of the l-strand, many weak junction sites were scattered.

These results strongly support the unidirectional initiation model proposed by Hirose et al. (Fig. 10), in which the first DNA synthesis starts leftward from one of the r-strand transition sites in oriC (step 2: the synthesis of the leading strand for the counterclockwise replication fork) and is followed by the start of the lagging l-strand synthesis on the displaced template (r-strand) on the outside region, left of oriC (step 3) (9). It is also suggested that this lagging l-strand DNA becomes the leading strand in the clockwise replication fork, since no junction was detected in the l-strand within oriC or its vicinity. Previously, Marsh and Worcel reported that, when E. coli cells were pulse-labeled with ^{3}H-thymidine at the start of synchronous replication, the 1.5 kb HindIII fragment (D fragment in Fig. 1 and 10) had the highest specific radioactivity in the HindIII digests of the pulse-labeled DNA (16). The fragment is on the left of the 2.2 kb fragment containing oriC (fragment C in Fig. 1 and 10). In their analysis,

^{3}H-labeled restriction fragments of the proper size should be produced only from the region where new DNA synthesis traversed both of the adjacent HindIII sites. If DNA synthesis indeed starts from the major RNA-DNA transition sites of the r-strand on the left of the HindIII site (at +245) in the oriC, it is predictable that the 1.5 kb HindIII fragment rather than the 2.2 kb one is the first fragment ^{3}H-labeled in the l-strand and cleaved in the proper size. Similarly, the 1.5 kb fragment should be the first one ^{3}H-labeled in l-strand, if the lagging l-strand synthesis starts first from the region on the left to oriC where we have found many RNA-DNA transition sites and if the leading l-strand synthesis through the oriC to the right (step 4) begins after a short lag. This type of unidirectional start mechanism is also suggested in other bidirectional replication systems such as T7 phage (21) (Sugimoto, Kohara and Okazaki, in preparation) and SV40 (22). In the case of T7 phage, the region essential to ori-function was approximately 800 bp away from the eye-center of early replicative intermediates (23).

The present results suggest that the start site of the leading r-strand is not restricted to a unique site, but may be variable within sites I-S (Fig. 3). The sites at A to F (Fig. 3) might also represent the leading r-strand start sites which are shifted further to the counterclockwise direction on rare occasions. Tabata et al. (24) analyzed the start site of DNA replication in a replication system of oriC plasmid in vitro, which is essentially identical to those discovered by Fuller et al. (25). Inconsistent with our results, the start sites they reported, as judged by the specific radioactivities of the restriction fragments prepared from the early stage reaction products, were at a region near but to the left outside region of the 245 bp-oriC sequence in both strands. They also found that the cell extract, with an insufficient amount of dnaA protein, catalyzed replication dependent on the oriC sequence, but start sites of DNA were apparently dispersed all over the template (24). Also using the in vitro replication system by Fuller et al. (25), we have analyzed RNA-DNA transition sites within oriC. Similar to the results of this work, many sites have been found in the r-strand but none in the l-strand (our unpublished data). Interestingly, the transition sites used in vivo were used also in vitro, but with different relative frequencies and additional sites were detected (our unpublished results). The observation of wide variation of the start sites of DNA synthesis in vitro might reflect the intrinsically versatile nature of the switching point from primer RNA to DNA synthesis and/or of the start site

of primer RNA.

One purpose of this study is to obtain information concerning whether RNA polymerase or DNA primase (dnaG protein) makes the primer for the first leading strand at the replication origin. We have analyzed the 5' ends of the primer RNA remaining on the DNA chains in the oriC region. The primer ends have been detected at the adjacent positions of the strong RNA-DNA transition sites in the r-strand (Fig. 9). The length of the primer RNA at these sites was only a few residues and too short to specify one of the two enzymatic mechanisms.

Results of the analysis of RNA-DNA transition sites have provided some information relevant to this subject. One characteristic of the RNA-DNA transition sites is the presence of many strong sites inside the r-strand of the oriC. This contrasts to the transition site of ColE1 plasmid, from where the leading l-strand synthesis starts in vitro. In the case of ColE1, the primer end, composed of three consecutive nucleotides, is known to be created by RNase H on the precursor primer RNA synthesized by RNA polymerase from the 555 nucleotides upstream of the first cleavage site (26). Multiple primer sites are characteristic of the priming mechanism of discontinuous replication. Recently, Kitani et al. (27) found that the majority of E. coli primer RNA molecules are 10 to 12 nucleotide residues in length, starting with pppPu (A/G ~5), and the most frequent dinucleotide sequence at the 5' ends is pppApPu (~60%). Properties of nucleotide sequence upstream of the transition sites, detected in this work, have been investigated in the light of this information. In the strong transition sites within, or very close to, oriC, such as J, K and N, the transition occurs over a 5 to 7 contiguous nucleotide sequence, whereas in other sites the transition occurs within three adjacent nucleotides. In many of the latter type sites, Pu-Pu sequences, mainly AG and AA, are found at the region between 10 to 12 nucleotides upstream of the transition sites (Fig. 7). In a few sites, only one purine residue is found at the corresponding position. Thus, the sequence property upstream of the transition sites satisfies, at least, the prerequisite for the priming by a discontinuous mechanism. For the sites such as J, K and N, in which transition sites are composed of the relatively long cluster, not every transition nucleotide has a Pu-Pu or Pu sequence at the corresponding upstream position. Therefore, primer RNA with an irregular size would be made at these sites. This might suggest that if the DNA primase is responsible for the priming at these sites, the higher organization of the priming machinery at these sites would be somewhat different from what functions at the established replication forks.

In similar analysis of the RNA-DNA transition sites in the replication origin of bacteriophage lambda *in vivo*, Yoda *et al*. found many sites in and around origin (oriλ) (Yoda, Yasuda, Jiang & Okazaki, manuscript in preparation). Similar to the case of *E. coli*, the sites distribute asymmetrically between the complementary strands and they can be classified into two types with regard to the width of each site. In the sites where RNA-DNA transition occurs on 1 to 3 consecutive nucleotides, the 5'-CAG sequence is most frequently found at the position centered at 11 nucleotide upstream at the transition sites. Tsurimoto and Matsubara reported that in a replication system of oriλ *in vitro* many transition sites are asymmetrically distributed on the flanking region of oriλ (28). Although the exact location of the transition sites *in vitro* does not agree with those found *in vivo*, essential properties of the sites, such as multiple start and asymmetrical distribution, are common, as they were in *E. coli oriC in vivo* and *in vitro*.

If the start of DNA synthesis at the chromosome origin is primed by the so called primosome (29), the higher organization of priming machinery for the discontinuous replication in which DNA primase plays the central role, it may be assembled at, or in close vicinity to, the essential *oriC* segment. Assembly sites of the ϕX174 primosome (30), which had been proposed to be a model for the priming machinery of the discontinuous phase of *E. coli* chromosome replication, had been analyzed but not detected in the *oriC* area (31, 32). These negative results however may not exclude the function of a primosome in the initiation reaction, but might rather indicate that the primosome of *E. coli* is somewhat different in its detailed structure from that of ϕX174. In accordance with this, primer RNA made by *in vitro* ϕX174 primosome is different from that of *E. coli* in chain length (33).

As for the role of RNA polymerase, we must point out that the transcription started from the promoter for 16 kd gene (16 kd-promoter) enters into *oriC* to the counterclockwise direction. Although this promoter can be deleted in a *oriC* plasmid (1), its requirement to the efficient maintenance of the *oriC* plasmid is suggested (34). It is also suggested that the transcript is related to the regulation of the initiation of replication (35, 36). Interestingly, terminator activity was found in the BamHI-HindIII region of the *oriC* (+93 to +245) (37, 38). Our unpublished results show that bulk of the transcripts from the 16-kd promoter go through the *oriC* and a small portion terminate at 4 sites; one in the area between the *oriC* and the 16 kd gene and others inside the *oriC*. The transcription termination sites inside the *oriC* do not coincide with the major r-strand transition sites we

have detected in this work, but locate closer to the core dnaA-protein binding sites (5, 6). It has been suggested that dnaA gene products interact physically with RNA polymerase (39) and act in the attenuation of transcription of tryptophane operon (40). Unfortunately, we could not find a change in the level of the transcripts which were terminated at these sites upon the temperature inactivation of dnaA-gene products in vivo (our unpublished result). Another transcript in the counterclockwise orientation that starts with C at +166 within oriC in vitro was reported by Lother and Messer (41). Recently, a reconstituted oriC-specific DNA replication system was reported which uses RNA polymerase alone for priming of the start of DNA synthesis (7). Neither the transcript, nor the start site of DNA synthesis of the system, has been analyzed. Thus the relationship between the known transcripts around the origin region and the RNA-DNA transition sites we have detected in the r-strand inside the oriC remains unclear.

Many GATC stretches, the sites for dam methylase, are present in oriC and its flanking region (13). Interestingly, many transition sites in r-strand fall in or near the GATC stretches. Moreover, the sites were preceded and/or followed by another GATC stretches. The three weak transition sites (A, D and E) outside the oriC also fall into this type of sequences. Recognition of secondary or tertiary structure of DNA by DNA primase has been proposed previously (42). These GATC stretches could serve to make a secondary structure and might become signals for synthesis of primer RNA by DNA primase or effect switching from RNA to DNA synthesis in priming by RNA polymerase. Alternatively, methylation on the stretch might be recognized in some step of the initiation reaction (43).

ACKNOWLEDGEMENT

We thank Drs. T. Uchida and J. D. Boeke for generous gifts of RNase T_1, A and f1R199 phage, respectively and Dr. W. Messer for communication of the nucleotide sequence of pCM959 before publication. This work was supported by Grants-in-Aid for Special Project Research from the Ministry of Education, Science and Culture of Japan.

*Present address: Biology Division, National Cancer Center Research Institute, Tsukiji 5-1-1, Chuo-ku, Tokyo 104, Japan

REFERENCES

1. Oka, A., Sugimoto, K., Takanami, M. and Hirota, Y. (1980) Mol. Gen. Genet. 178, 9-20.

2. Bird, R. E., Louarn, J., Martuscelli, J. and Caro, L. (1972) J. Mol. Biol. 70, 549-566.
3. Masters, M. & Broda, P. (1971) Nature (New Biol.) 232, 137-140.
4. Oka, A., Sasaki, H., Sugimoto, K. and Takanami, M. (1984) J. Mol. Biol. 176, 443-458.
5. Fuller, R. S. and Kornberg, A. (1984) Cell 38, 889-900.
6. Matsui, M., Oka, A., Takanami, M., Yasuda, S. and Hirota, Y. (1985) J. Mol. Biol. 184, in press.
7. Kaguni, J. M. and Kornberg, A. (1984) Cell 38, 183-190.
8. Tomizawa, J. and Selzer, G. (1979) Ann. Rev. Biochem. 48, 999-1034.
9. Hirose, S., Hiraga, S. and Okazaki, T. (1983) Mol. Gen. Genet. 189, 422-431.
10. Boeke, J. D., Vovis, G. F. and Zinder, N. D. (1979) Proc. Natl. Acad. Sci. USA 76, 2699-2702.
11. Walseth, T. F. and Johnson, R. A. (1979) Biochem. Biophys. Acta 562, 11-31.
12. Ogawa, T., Hirose, S., Okazaki, T. and Okazaki, R. (1977) J. Mol. Biol. 112, 121-140.
13. Sugimoto, K., Oka, A., Sugisaki, H., Takanami, M., Nishimura, A., Yasuda, S. and Hirota, Y. (1979) Proc. Natl. Acad. Sci. USA 76, 575-579.
14. Buhk, H-J. and Messer, W. (1983) Gene 24, 265-279.
15. Yasuda, S. and Hirota, Y. (1977) Proc. Natl. Acad. Sci. USA 74, 5458-5462.
16. Marsh, R. C. and Worcell, A. (1977) Proc. Natl. Acad. Sci. USA 74, 2720-2724.
17. Okazaki, R., Hirose, S., Okazaki, T., Ogawa, T. and Kurosawa, Y. (1975) Biochem. Biophy. Res. Commn. 62, 1018-1024.
18. Maxam, A. and Gilbert, W. (1980) Methods in Enzymology 65, 499-560.
19. Fujiyama, A., Kohara, Y. and Okazaki, T. (1981) Proc. Natl. Acad. Sci. USA 78, 903-907.
20. Tanaka, Y., Dyer, T. A. and Brownlee, G. G. (1980) Nucleic Acids Res. 8, 1259-1272.
21. Richardson, C. C. (1983) Cell 33, 315-317.
22. Hay, R. T. and DePamphilis, M. L. (1982) Cell 28, 767-779.
23. Tamanoi, F., Saito, H. and Richardson, C. C. (1980) Proc. Natl. Acad. Sci. USA 77, 2656-2660.
24. Tabata, S., Oka, A., Sugimoto, K., Takanami, M., Yasuda, S. and Hirota, Y. (1983) Nucleic Acids Res. 11, 2617-2626.
25. Fuller, R. S., Kaguni, J. M. and Kornberg, A. (1981) Proc. Natl. Acad. Sci. USA 78, 7370-7374.
26. Itoh, T. and Tomizawa, J. (1980) Proc. Natl. Acad. Sci. USA 77, 2450-2454.
27. Kitani, T., Yoda, K., Ogawa, T. and Okazaki, T. (1985) J. Mol. Biol. 183, in press.
28. Tsurimoto, T. and Matsubara, K.(1985) Proc. Natl. Acad. Sci. USA 81, 7402-7406.
29. Kornberg, A. (1980) DNA Replication, Freeman, San Francisco.
30. Low, R. L., Arai, K. and Kornberg, A. (1981) Proc. Natl. Acad. Sci. USA 28, 1436-1440.
31. Ray, D. S., Cleary, J. M., Hines, J. C., Kim, M. M., Strathearn, M., Kaguni, L. S. and Roark, M. (1981) in The Initiation of DNA Replication, Ray, D. S. Ed., pp. 169-193, Academic Press, New York.
32. Stuije, A. R., Weisbeek, P. J. and Meijer, M. (1984) Nucleic Acids Res. 12, 3321-3332.
33. Ogawa, T., Arai, K. and Okazaki, T. (1983) J. Biol. Chem. 258, 13353-13358.
34. Stuije, A. R. and Meijer, M. (1983) Nucleic Acids Res. 11, 5775-5791.

35. Tanaka, M. and Hiraga, S. (1985) Mol. Gen. Genet. in press.
36. Lother, H., Kölling, R., Kücherer, C. and Schauzu, M. (1985) EMBO J. 4, 555-560.
37. Morita, M., Sugimoto, K., Oka, A., Takanami, M. and Hirota, Y. (1981) in The Initiation of DNA Replication, ICN-UCLA Symp. Molec. Cell. Biol., Ray, D. Ed., Vol. 22, pp. 29-35, Academic Press, New York.
38. Hansen, F. G., Koefoed, S., von Meyenburg, K. and Atlung, T. (1981) in The Initiation of DNA Replication, ICN-UCLA Symp. Molec. Cell. Biol., Ray, D. Ed., Vol. 22, pp. 1-12, Academic Press, New York.
39. Bagdasarian, M. M., Isakowska, M. and Bagdasarian, M. (1977) J. Bacteriol. 130, 577-582.
40. Atlung, T. and Hansen, F. G. (1983) J. Bacteriol. 156, 985-992.
41. Lother, H. and Messer, W. (1981) Nature 294, 376-378.
42. Sims, J. and Bentz, Jr., E. W. (1980) Proc. Natl. Acad. Sci. USA 77, 900-904.
43. Hughes, P., Sqvali-Houssaini, F-Z., Fovterre, P. and Kohiyama, M. (1984) J. Mol. Biol. 176, 155-159.