Volume 13 Number 7 1985 Nucleic Acids Research

Structure and function of the region of the replication origin of the *Bacillus subtilis* chromosome. IV. Transcription of the *oriC* region and expression of DNA gyrase genes and other open reading frames

Naotake Ogasawara, Shigeki Moriya and Hiroshi Yoshikawa*

Cancer Research Institute, Kanazawa University, 13-1, Takaramachi, Kanazawa 920, Japan

Received 16 January 1985; Revised and Accepted 15 March 1985

ABSTRACT

We have determined nucleotide sequence of some 10,000 base pairs (bp) in the *oriC* region of the *Bacillus subtilis* chromosome. Initiation sites of transcription from this region were determined in vivo by the S1-mapping method. Five major initiation sites were found in the leader sequences of five open reading frames (ORF) deduced from the nucleotide sequence. The sixth site is located inside the ORF323("*recF*"). Putative promoters were found for each transcript. Function of these promoters was demonstrated in *Escherichia coli* by the Maxi-cell method using appropriate fragments cloned in pBR vectors. Based on these results, genes in 10,000 bp *oriC* region are divided into 4 transcriptional units. *GyrB* composes one unit with two other ORFs, while *gyrA* constitutes a single unit by itself. The promoters for ORF446("*dnaA*") and ORF378("*dnaN*") are located within the putative signal sequences for *oriC*. Transcription from these promoters is dependent on a dna-initiation gene, *dnaB*.

INTRODUCTION

Little is known on the regulation of initiation of chromosomal replication in bacteria. A 245 base pair sequence essential for DNA replication was isolated from the replication origin of the chromosome (*oriC*) of *Escherichia coli* (1). However, the initiation of replication of the plasmids containing the *oriC* sequence (mini-chromosome) is no longer under the cell cycle dependent regulation which controls chromosomal replication (2). We have reported previously that formation of a specific protein-DNA complex parallels the cell's ability to initiate a round of replication in *Bacillus subtilis* (3). This process is severely inhibited by inhibitors for DNA gyrase (4). Furthermore, we found near the *oriC* of the *B. subtilis* chromosome two ribosomal RNA operons whose promoters act as strong suppressors for plasmid replication (5,6,7).

These results strongly suggest the importance of genes and inter-gene-sequence organization in the *oriC* region in the initiation of replication.

In the preceding paper we reported nucleotide sequence in the *oriC* region covering some 10,000 base pairs. Analysis of the sequence revealed

seven open reading frames (ORF) including the two structural genes for the DNA gyrase subunits, *gyrA* and *gyrB*. Furthermore, ORF44, ORF446, ORF378 and ORF323 were shown to be homologous with *rpmH* (8), *dnaA* (9), *dnaN* (10), and *recF* (11) of *E. coli*. Hence, we designate these ORFs by adding genetic markers of homologous genes of *E. coli* in parenthesis. These results indicate that the organization of ORFs from ORF44 (*"rpmH"*) to ORF638 (*gyrB*) resembles the organization of genes in the *rpmH-gyrB* region of the *E. coli* chromosome (9). It is now essential to examine how these genes are transcribed and expressed in vivo and to explore possible roles of these genes and inter-gene sequences in initiation of DNA replication. For this purpose we determined initiation sites of transcription from the *oriC* region in vivo by the S1-mapping method and expression of these ORFs in *E. coli* cells using the Maxi--cell method.

The S1-mapping detected 6 major initiation sites for transcription which are consistent with the transcriptional units deduced from expression of the open reading frames in *E. coli*.

MATERIALS AND METHODS

Bacterial strains and plasmids. *B. subtilis* CRK2000 (*leu8, trpC, thyAB*) (12) was used throughout. *E. coli* C600 was used as a host for cloning of *B. subtilis* chromosomal fragments. *E. coli* N1790 (*recA, uvrA*) (13) was used for labeling proteins encoded by plasmid DNA by the Maxi-cell method. Plasmids used in this study are listed in Table 1 and Table 2.

Chemicals. (γ-^{32}P)-ATP (PB10168, 3000 Ci/mmole) and L-(^{35}S)-methionine

Table 1 Plasmids and phages used in this work

Plasmid or phage	inserted[a] DNA	vector (insertion sites)	reference
M13BS501[b]	-573 (PvuII) - 1174 (EcoRI)	M13mp11	This work
pSM2002	201 (EcoRV) - 1426 (EcoRV)	pNO21 (EcoRV)	(14)
pSM1001	1174 (EcoRI) - 2949 (EcoRI)	pBR328 (EcoRI)	(14)
pSM2003	1207 (SalI) - 5204 (SalI)	pNO21 (SalI)	(14)
pNO1003[c]	2949 (EcoRI) - 9011 (EcoRI)	pBR322 (EcoRI)	This work
pMSdR32	6385 (BamHI) - 9723	pMS102' (BamHI)	(5)

(a) Nucleotide numbers and restriction sites of two ends of the fragment are shown. Nucleotide numbers are same as in Fig.3 of the preceding paper.
(b) M13 phage used for sequence determination.
(c) E6 fragment from chBS01 phage (7) was cloned into pBR322.

(SJ204, 1000 Ci/mmole) were purchased from Amersham International Ltd., (Amersham, UK). T4 polynucleotide kinase, T4 DNA ligase and restriction endonucleases were from Takara Shuzo Co. Ltd., (Kyoto, Japan) or from Nippon Gene Co. Ltd., (Toyama, Japan).

Determination of transcripts by the S1-mapping method. Preparation of whole cellular RNA from germinating spores or exponentially growing cells of *B. subtilis* CRK2000 and S1 nuclease mapping with 5'-end labeled probes were as described previously (7).

Detection of proteins encoded in plasmid DNA by the Maxi-cell method. Plasmids containing various fragments from the *oriC* region were introduced in *E. coli* N1790. Proteins synthesized by the plasmid coded genes were specifically labeled with L-(^{35}S)-methionine at 10 μCi/ml as described by Sancar et al (15), and Meadow et al (16). The labeled cells were washed three times with the two fold diluted M9 medium and suspended in 0.2 ml of 50 mM Tris-HCl (pH8.0), 15 mM EDTA. The cells were lysed by three cycles of freezing and thawing with 500 μg/ml lysozyme and addition of 50 μl of five fold concentrated lysis buffer. Samples were boiled for 2 min and analyzed by sodium

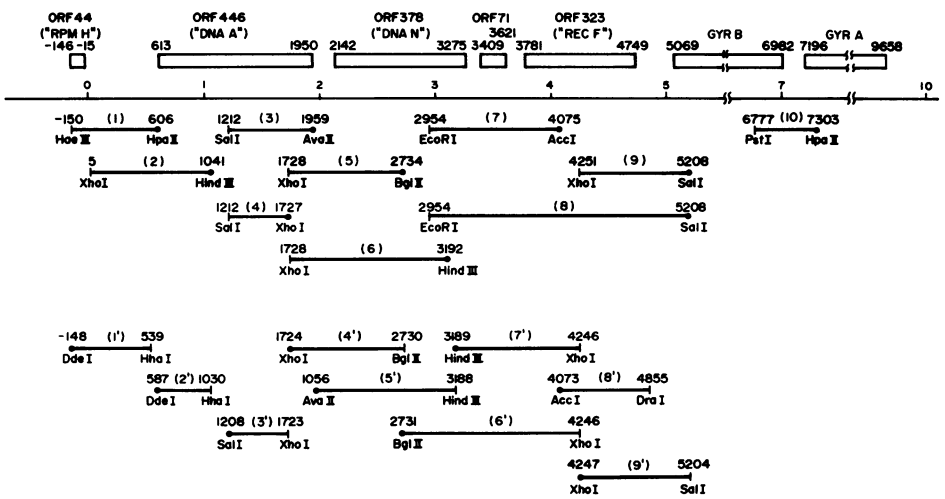


Figure 1. DNA fragments used as probes for the S1-mapping experiments. At the top of the map, open reading frames (ORF), *gyrA* and *gyrB* genes deduced from the nucleotide sequence are shown. Numbers at both ends of the frames are the nucleotide numbers starting from *XhoI* site as described in Fig. 3 of the preceding paper. Labeling of 5'-ends, indicated by closed cireles, are described in MATERIALS AND METHODS. The upper group (1-10) was used to detect transcripts of the left-to-right orientation and the lower group (1'-9') was for the transcripts of the opposite orientation. Restriction enzyme cleavage sites and nucleotide numbers are shown for each fragment.

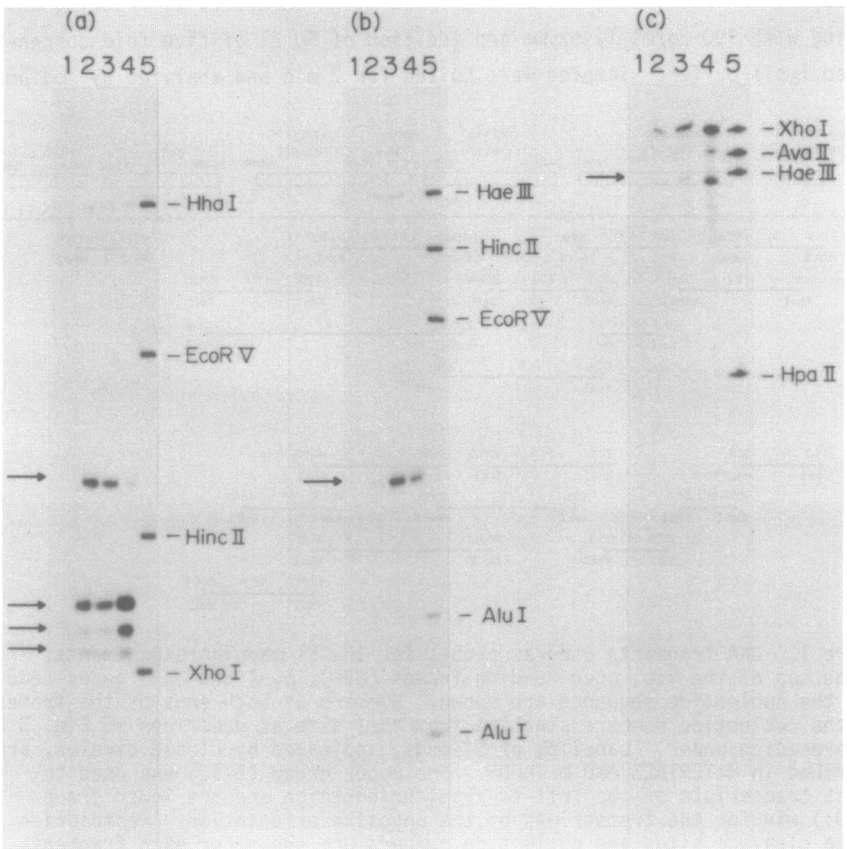
dodecyl sulfate/10 % polyacrylamide gel electrophoresis. The gel was fixed, stained for molecular weight estimation using protein molecular weight standard from BRL (Bethesda Research Lab., Maryland, USA), immersed in the enhancer (New England Nuclear), dried and autoradiographed.

RESULTS

In vivo transcription from the *oriC* region.

Initiation sites of transcription in the *oriC* region of the *B. subtilis* chromosome were determined by S1-mapping method using various cloned DNA fragments as probes to hybridize RNA transcripts from this region (Fig. 1).

Five major initiation sites were detected on one DNA strand, while only one site was found on the complementary strand. Actual transcripts detected by the S1-mapping method are shown in Fig. 2 and summarized schematically in



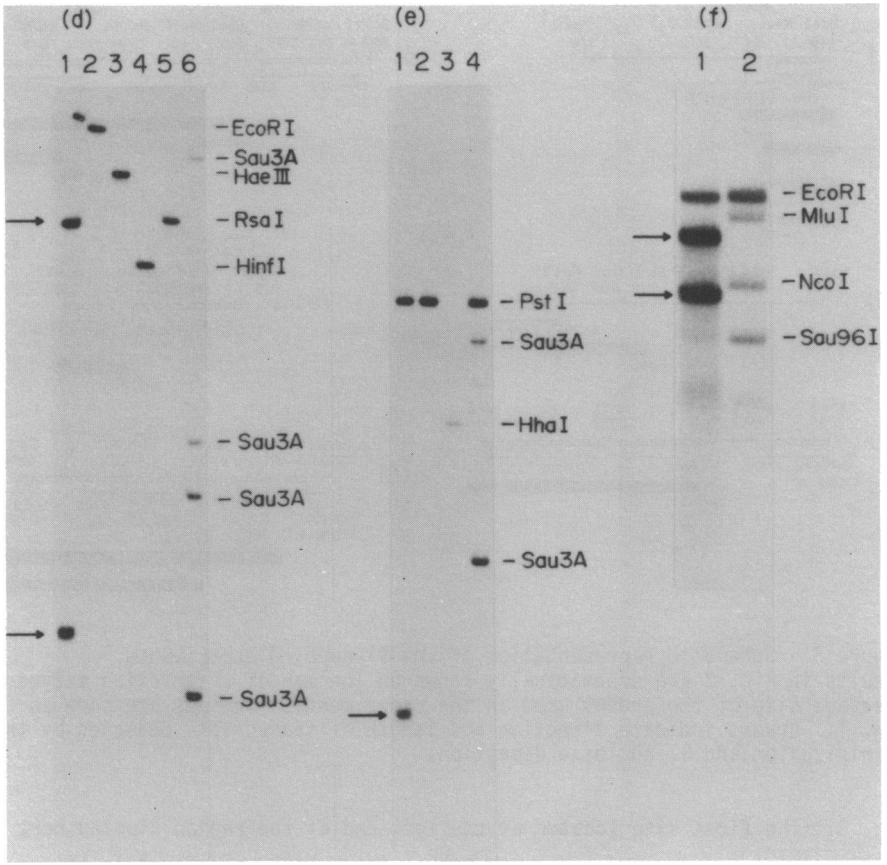


Figure 2. S1-mapping of the *oriC* region.
Experimental procedures are described in MATERIALS AND METHODS. Different probes were used for each set of experiments (a) to (f). Restriction maps of the probes are shown in Fig. 3 with the same alphabetical symbols. Corresponding fragments in Fig. 1 are as follows: a: (1'), b: (1), c: (5), d: (7), e: (10), f: (8). For the set of experiments (a)-(c), lanes 1-5 indicate as follows: 1: no cellular RNA was added or *E. coli* tRNA was added as controls, 2: RNA isolated from spores germinating in thymine-less conditions for 3 hr, 3: RNA from spores germinated in thymine-less culture and then grown for 10 min in the presence of thymidine, 4: RNA from exponentially growing cells, 5: restriction fragments of the labeled probes (cleavage sites at 3'-ends are indicated) as size markers. Maps of the cleavage sites are shown in Fig.3. For the set of experiments (d)-(f), lane 1 is RNA isolated from exponentially growing cells and lanes 2-6 are restriction fragments of probes as markers. Transcripts are pointed out by arrows.

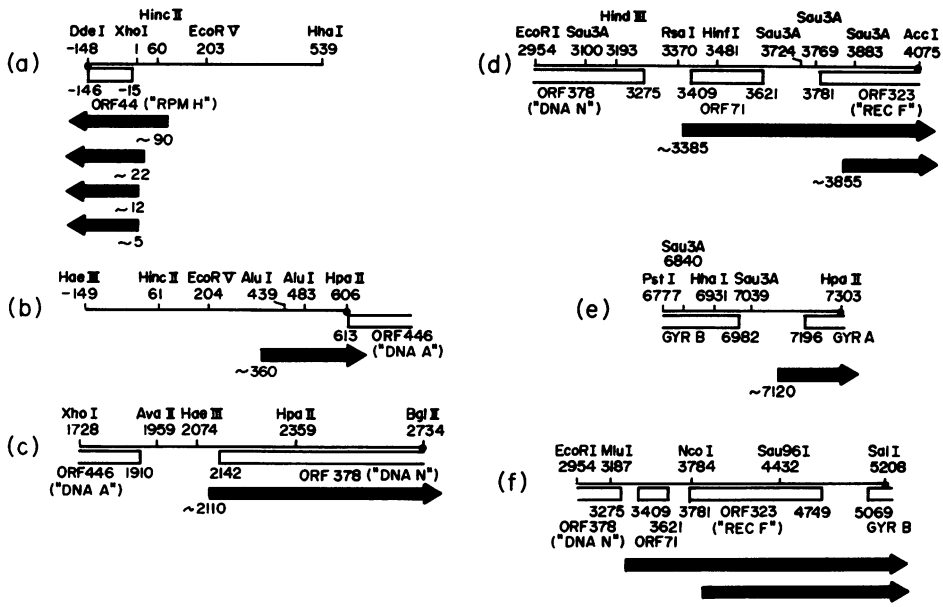


Figure 3. Schematic representation of the S1-mapping experiments. Results in Fig. 2 are schematically shown on the map of restriction enzyme cleavage site of the probes used in the experiments. Numbers are same as Fig. 1. Arrows indicate direction and length of transcripts detected by the hybridization and S1 nuclease digestion.

Fig. 3. The first site located at the left end of the region studied here consists of two major and two minor transcripts which were found to be synthesized from the leader sequence of ORF44(*"rpmH"*) (Fig. 2a & 3a). Typical promoter sequences are found for each of the major transcripts (Fig. 4 TR1 & TR2). Their relative amount of the transcripts varied depending on growth cycle and growth conditions of the cell (see Fig. 2a, lanes 2-4). These transcripts are the only transcripts within the entire *oriC* region analyzed which are synthesized in the right-to-left orientation. All others are synthesized in the opposite left-to-right orientation. The first one from the left end (TR3) starts transcription within the leader sequence of ORF446(*"dnaA"*) (Fig. 2b & 3b) from a site immediately downstream of a typical promoter (Fig. 4, TR3). The second initiation site is in the non-coding spacer sequence between ORF446(*"dnaA"*) and ORF378(*"dnaN"*) (Fig. 2c & 3c). No typical promoter sequence was found except for a possible -10 sequence (Fig. 4, TR4). The third (TR5) and fourth (TR6) transcripts were detected in the *recF-gyrB* region. TR5 starts in the spacer between ORF378(*"dnaN"*) and ORF71

Figure 4. Putative promoters for transcripts in the *oriC* region. TR1-TR7 are described in the text and shown in Fig. 7. Possible -10 and -35 sequences are boxed. Arrows indicate initiation sites of transcripts determined by the S1-mapping experiments.

and transcribes through the *gyrB* gene, while TR6 begins inside of ORF323 ("*recF*") and reads through the *gyrB* gene (Fig. 2d,f & 3d,f). Although there is a large non-coding region between ORF323("*recF*") and *gyrB*, initiation of transcription was not detected in this region. Consistent with the read-through transcription from ORF323*("recF")* into *gyrB*, no signal sequence for termination was detected downstream of ORF323*("recF")*. The fifth transcript (TR7) is initiated in the leader sequence of the *gyrA* gene (Fig. 2e & 3e). The *gyrA* gene is followed by a typical termination signal; further downstream are the two tandem promoters for a ribosomal RNA operon *rrnO*. Transcription from these promoters has been reported (7). For the last three transcripts (TR5, TR6, TR7) no typical -35 sequences were found upstream of their initiation sites, but similar sequences were observed downstream of their -10 sequences (Fig. 4, TR5, TR6 & TR7).

The locations of all transcripts except for TR6 are consistent with the assumption that they are 5'-ends of the messenger RNA for ORFs located downstream of the initiation sites. This assumption was proved by testing the expression of the ORFs in *E. coli*.

Expression of ORF sequences in *E. coli*.

The S1-mapping experiments predicted the possible trancriptional units and the location of their putative promoters as mentioned above. To examine

Table 2. Expression of ORFs in Maxi-cell

plasmid	vector (cloning site)	cloned fragment [a]	prediction [b]	observed [c]
pMS2002	pNO21 (EcoRV)	201 (EcoRV) - 1426 (EcoRV)	36.1 K (ORF446')	37.5 K
pSM2050	pNO21 (EcoRV, EcoRI)	201 (EcoRV) - 2949 (EcoRI)	50.8 (ORF446) 30.0 (ORF378')	- 32
pSM2003	pNO21 (SalI)	1207 (SalI) - 5204 (SalI)	42.1 (ORF378) 36.8 (ORF323)	47 38
pSM2051	pNO21 (SalI, EcoRI)	1207 (SalI) - 4135 (PvuII)	42.1 (ORF378) 14.5 (ORF323')	47 -
pSM2052	pNO21 (SalI)	1723 (XhoI) - 4246 (XhoI)	42.1 (ORF378) 20.9 (ORF323')	47 18.5
pSM2053	pNO21 (SalI)	1723 (XhoI) - 5204 (SalI)	42.1 (ORF378) 36.8 (ORF323)	47 38
pSM2014	pNO21 (EcoRI)	2949 (EcoRI) - 9011 (EcoRI)	36.8 (ORF323) 71.4 (GYR B) 68.1 (GYR A')	38 86 74
pSM1041	pBR322 (BamHI, EcoRI)	2949 (EcoRI) - 9723	36.8 (ORF323) 71.4 (GYR B) 92.0 (GYR A)	38 86 99
pSM1042	pBR322 (BamHI, EcoRI)	2949 (EcoRI) - 3188 (HindIII) 4246 (XhoI) - 9723	92.0 (GYR A)	99
pSM1043	pBR322 (BamHI, EcoRI)	2949 (EcoRI) - 3099 3414 - 9723	71.4 (GYR B) 92.0 (GYR A)	88 105

(a) Nucleotide numbers and restriction sites of two ends of the cloned fragment are indicated.
(b) Molecular weight in kilo dalton (K) of product protein of complete ORF or partial ORF (ORF') was estimated from nucleotide sequence of each cloned fragment.
(c) Data from Fig.5.

whether or not these promoters are indeed functional, the expression of the coding sequences was examined with respect to proteins which might correspond to ORFs deduced from the nucleotide sequence. In order to detect proteins synthesized from the genes in *oriC* region and to correlate them with ORF deduced from the nucleotide sequence, various *oriC* fragments were cloned into plasmids, introduced in *E. coli*, and proteins synthesized in the cell were identified by the Maxi-cell method as described in MATERIALS AND METHODS.

The DNA segments introduced either in pNO21 (14) or pBR322 are listed in Table 2. Fragments were inserted in such a way that no read-through transcription would occur from the vector DNA. Fig. 5 shows that discrete species of proteins are synthesized from the fragments incorporated into the plasmids. The molecular weight of these proteins as estimated and compared with these value predicted from the different ORFs is summarized in Table 2. The results clearly show that ORF378("*dnaN*"), ORF323("*recF*"), *gyrB* and *gyrA* are expressed in *E. coli* cells. Only a partial product of ORF446("*dnaA*") was

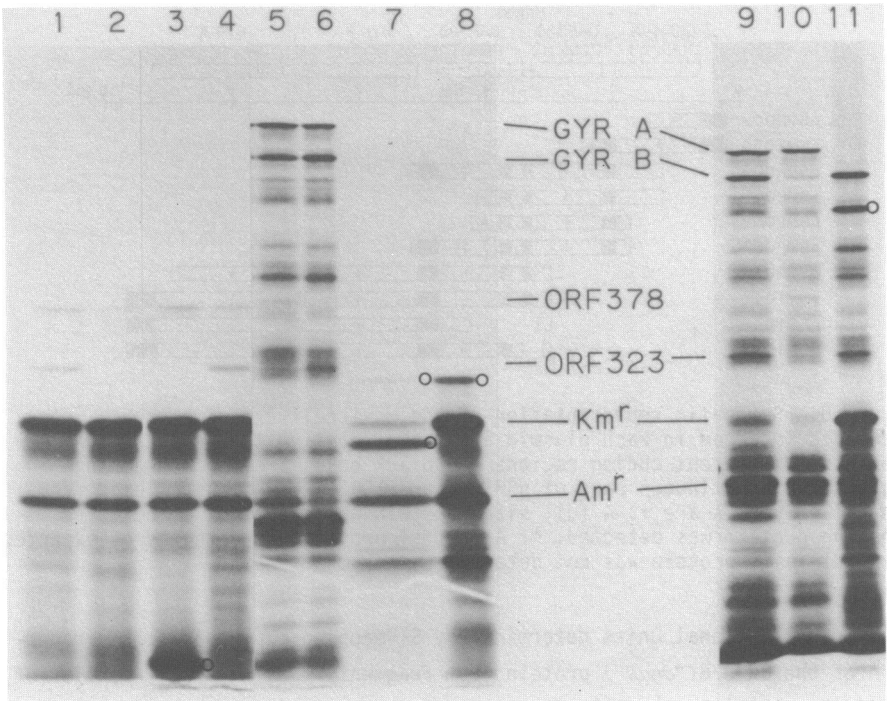


Figure 5. Expression of the ORFs in *E. coli*.
Experimental procedures are described in MATERIALS AND METHODS. Lanes 1-11 correspond to plasmids pSM2053, 2051, 2052, 2003,1043, 1041, 2050, 2002, 1041, 1042 and 2014 respectively. Bands indicated by circle are partial product proteins of ORF446(*"dnaA"*) (lane 8), ORF378(*"dnaN"*) (lane 7) and *gyrA* (lane 11).

identified (see plasmid pSM2002 in Fig. 5 and Table 2). Failure to detect a complete ORF446(*"dnaA"*) product using plasmid pSM2050 which was expected to contain a complete ORF446(*"dnaA"*) may be because the plasmid yielded only a low level of *"dnaA"* protein in the Maxi-cell. Alternatively, nucleotide changes inside the ORF when the plasmid was constructed by fusing two fragments within reading frame may be responsible. Products from ORF44("*rpmH*") and ORF71 are too small to be detected by the present method. It is interesting to note that a higher background of cellular protein was observed when the plasmids contained *gyrA* or *gyrB* genes. An overproduction of gyrase subunits may protect the UV damage to cellular DNA by unknown mechanisms.

Results of the Maxi-cell experiments are summarized schematically in Fig. 6. The mode of expression in *E. coli* is consistent with that predicted

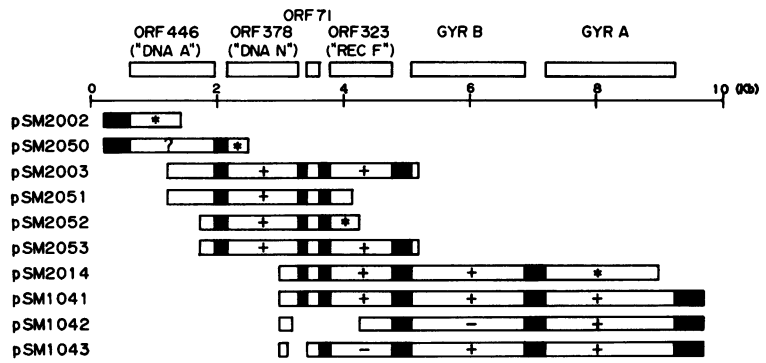


Figure 6. Schematic representation of the Maxi-cell experiments. Fragments inserted in each plasmid are shown relative to the map of ORF. Open boxes represent coding regions and black boxes are for non-coding regions. Discontinuous part of pSM1042 and 1043 are regions deleted from pSM1041. Symbols are +: A full sized protein was detected, -: no corresponding protein was detected, *: A partial protein was detected as expected, ?: A full size protein was not detected by unknown reason.

from transcriptional units determined by S1-mapping experiments: 1) Production of the ORF378(*"dnaN"*) protein from fragments lacking the 5'-portion of ORF446("*dnaA*") (pSM2003,2051,2052 & 2053) indicates an inherent transcriptional signal for ORF378("*dnaN*"). 2) When the promoter region of ORF71 was deleted (pSM1043), expression of ORF323("*recF*") disappeared completely, while both *gyrB* and *gyrA* were produced normally (see results by plasmid pSM1043). When the deletion was extended 1/3 into ORF323("*recF*") (pSM1042), expression of *gyrB* was lost without affecting *gyrA* gene expression (see plasmid pSM-1042). These results are consistent with three transcriptional units, the first starting from promoter for ORF71, the second from inside of ORF323 ("*recF*") and the third from the promoter for *gyrA*.

DISCUSSION

Six transcriptional units are identified in the *oriC* region of the *B. subtilis* chromosome. Experiments by S1-mapping and Maxi-cell are summarized in Fig. 7. The transcription units comprise one or several of ORFs deduced from nucleotide sequence (see preceding paper). ORF44("*rpmH*") constitutes one transcriptional unit with both a typical promoter and terminator; the protein product, however, was too small to be detected by the present method. Multiple start sites of transcription of the ORF44(*"rpmH"*) were detected as was reported also for the transcription of *rpmH* gene in *E. coli* (8).

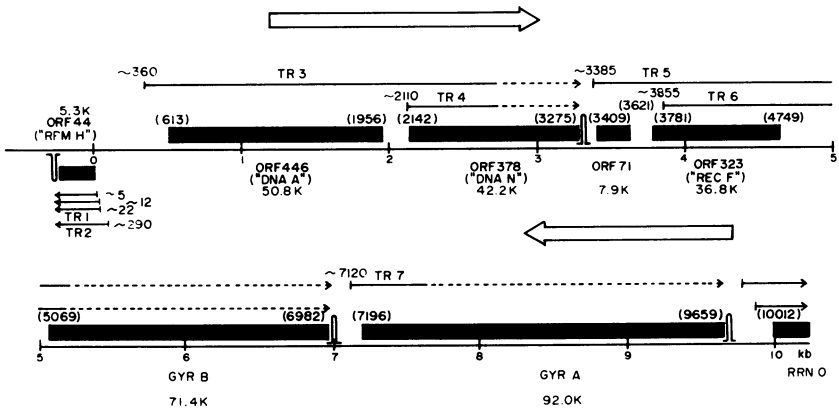


Figure 7. Transcriptional units of the *oriC* region.
A map of ORF (black bars) including *gyrB*, *gyrA* and *rrnO* in the *oriC* region is shown. Nucleotide numbers at two ends of ORF are same as in Fig. 1 and shown in parenthesis. Under each ORF, molecular weight of protein product is estimated from amino acid sequence deduced from nucleotide sequence and shown in kilo daltons (k). Open big arrows indicate location, direction and size of the first DNA strands synthesized during synchronous initiation of replication. Thin arrows indicate RNA transcripts actually detected by the S1-mapping experiments (solid lines) and assumed from the nucleotide sequence and the Maxi-cell experiments (dotted lines). Numbers at the ends of transcripts are approximate nucleotide numbers of 5'-ends of the transcripts. ⎍ at the end of some of the ORFs indicate palindromes typical for termination signals.

Independent expression of ORF446("*dnaA*") and ORF378("*dnaN*") shown by the Maxi-cell method is consistent with the detection of two transcripts, TR3 and TR4. However, these two genes may constitute one transcriptional unit because there is no termination signal in the spacer sequence and substantial read-through transcripts are detected by the S1-mapping and Northern blot experiments (data not shown). In this respect, the mode of transcription of ORF446*("dnaA")* and ORF378 *("dnaN")* resembles that of *dnaA-dnaN* operon in *E. coli* (10). That ORF71 and ORF323("*recF*") and *gyrB* gene constitute a single tanscriptional unit was shown clearly by both transcription and expression experiments. Absence of identifiable termination signals downstream of ORF71 and ORF323("*recF*") and its presence near the 3'-end of *gyrB* are consistent with the single transcriptional unit. Two independent functional initiations, one from the leader sequence of ORF71 and the second from inside of ORF323*("recF")* was confirmed by the Maxi-cell experiment. In contrast to *gyrB*, the *gyrA* gene constitutes a single transcriptional unit by itself.

Expression of ORFs of larger than 15 kd protein was demonstrated in *E.*

coli. Good correlation between transcriptional signals detected by S1-mapping method and those by expression in *E. coli* shows that the same transcriptional signals are used in *B. subtilis* and *E. coli*.

Two transcripts, TR3 and TR4, attract our attention in relation to initiation of DNA replication because: 1) the promoter for TR3 is located close to the site of the synthesis of the first DNA strand (see Fig. 7) (14) and 2) the promoter for TR3 and TR4 are located within the region surrounded by 9 and 4 repeated sequences respectively which are common to the *dnaA* protein binding sequence in *oriC* of *E. coli* (17). Our preliminary results show that the synthesis of both TR3 and TR4 is severely inhibited in a dna-initiation-ts mutant (*dnaB*) of *B. subtilis* (18) at non-permissive temperatures. These transcripts may be involved in synthesis and regulation of the primer RNA for initiation of the replication.

Regulation of transcription of gyrase genes is another interesting feature in this region. Although *gyrB* and *gyrA* genes are located next to each other, they are transcribed independently. Transcription of *gyrB* occurs from two different sites. In one case three genes, ORF71, ORF323*("recF")* and *gyrB* compose a single transcriptional unit. In the other, the transcription starts from within ORF323*("recF")*. The latter initiation is also supported by the Maxi-cell experiment. No typical promoter was found in the leader sequence of *gyrA* except for a Pribnow's box. A similar lack of -35 sequences is found in the leader sequences of *gyrB* and *gyrA* in *E. coli* (19).

Assuming that the leader sequence of ORF446*("dnaA")* containing 9 repeated sequence is a putative *oriC*, the transcriptional unit would be so organized that the transcription proceeds outwards bidirectionally from the origin. Additional genes should be identified in the left side of ORF44 ("*rpmH*") in order to verify this assumption.

ACKNOWLEDGMENTS

We express our appreciation to Dr. M. Gellert for communicating unpublished data. We thank Dr. S. Murakami and other colleagues in our laboratory for discussion and Mrs. K. Terada and Mrs. T. Nishikawa for help.

This work was supported by a Grant-in-aid for Special Project Research, for Cooperative Research and for Scientific Research from Ministry of Education, Science and Culture, Japan.

*To whom correspondence should be addressed

REFERENCES

1. Oka, A., Sugiura, K., Takanami, M. and Hirota, Y. (1980) Mol. Gen. Genet., 178, 9-20
2. Leonard, A. C., Hucul, J. A. and Helmstetter, C. H. (1982) J. Bacteriol., 149, 499-507
3. Yoshikawa, H., Ogasawara, N. and Seiki, M. (1980) Mol. Gen. Genet., 179, 265-272
4. Ogasawara, N., Seiki, M. and Yoshikawa, H. (1981) Mol. Gen. Genet., 181, 332-337
5. Seiki, M., Ogasawara, N. and Yoshikawa, H. (1982) Proc. Natl. Acad. Sci., USA, 79, 4285-4289
6. Ogasawara, N., Seiki. M. and Yoshikawa, H. (1983) J. Bacteriol., 154, 50-57
7. Ogasawara, N., Moriya, S. and Yoshikawa, H. (1983) Nucleic Acids Res., 11, 6301-6318
8. Hansen, F. G., Hansen, E. B. and Atlung, T. A. (1982) EMBO J., 1,1043-1048
9. Hansen, E. B., Hansen, F. G. and von Meyenburg, K. (1982) Nucleic Acids Res., 10, 7373-7385
10. Ohmori, H., Kimura, M. and Sakakibara, Y. (1984) Gene, 28, 159-170
11. Blanar, M. H., Sandler, S. J., Armengod, M. E., Ream, L. W. and Clark, A. J. (1984) Proc. Natl. Acad. Sci., 81, 4622-4626
12. Yoshikawa, H. (1965) Proc. Natl. Acad. Sci., USA, 53, 1476-1483
13. Horii, T., Ogawa, T. and Ogawa, H. (1981) Cell, 23, 689-697
14. Ogasawara, N., Mizumoto, S. and Yoshikawa, H. (1984) Gene, 30, 173-182
15. Sancar, A., Wharton, R. P., Seltzer, S., Kacinski, B. M., Clarke, N. D. and Rupp, W. D. (1981) J. Mol. Biol., 48, 45-62
16. Meadow, N. D., Saffen, D. W., Dottin, R. P.and Roseman, S. (1982) Proc. Natl. Acad. Sci., USA, 79, 2528-2532
17. Fuller, R. S., Funnell, B. E. and Kornberg, A. (1984) Cell, 38, 889-900
18. Murakami, S., Inuzuka, N. Yamaguchi, M., Yamaguchi, K. and Yoshikawa, H. (1976) J. Mol. Biol., 108, 683-704
19. Gellert, M., personal communication