APPLIED AND ENVIRONMENTAL MICROBIOLOGY, June 1991, p. 1838–1841
0099-2240/91/061838-04$02.00/0

Vol. 57, No. 6

Molecular Structure of the Replication Origin of a *Bacillus subtilis* (*natto*) Plasmid, pUH1

TOSHIO HARA,* SINICHIRO NAGATOMO, SEIYA OGATA, AND SEINOSUKE UEDA†

Microbial Genetics Division, Institute of Genetic Resources, Faculty of Agriculture, Kyushu University, Hakozaki, Fukuoka 812, Japan

Received 8 November 1990/Accepted 7 March 1991

The structure of a 2.0-kb *Bst*EII DNA sequence necessary and sufficient for the replication of a 5.7-kb Natto plasmid, pUH1, which is responsible for γ-polyglutamate production by *Bacillus subtilis* (*natto*), has been characterized by using a trimethoprim resistance gene derived from *B. subtilis* chromosomal DNA as a selective marker. The 2.0-kb DNA sequence contains an open reading frame, *rep*, stretching for 999 bp; a promotor region for *rep* expression; and a possible replication origin for the plasmid upstream of the promotor. The predicted Rep protein has highly homologous amino acid sequences with rep_{14} of pFTB14 in *B. amyloliquefaciens*, RepB of pUB110, and protein A, which is necessary for pC194 replication in staphylococci throughout the protein molecule, but is not homologous with RepC of staphylococcal plasmid pT181.

In a previous paper (5), we reported that a 5.7-kb plasmid designated pUH1, which encodes the γ-glutamyltranspeptidase gene responsible for γ-polyglutamate production, was detected in a *Bacillus subtilis* (*natto*) strain isolated from a fermented soybean food, natto. Recently, electron microscope studies of the hybridization products of pLS11, which is detected in *B. subtilis* IFO 3022 and is 8.6 kb in size, with pUH1 revealed a heteroduplex molecule containing 1.11- and 2.14-kb double-stranded DNA termini (6). We thought that these double-strand regions might play an important role in polyglutamate production and DNA replication.

This communication reports that the 2.0-kb *Bst*EII DNA fragment corresponds to the replication origin of Natto plasmid pUH1 and contains a 999-bp open reading frame, a promoter for the open reading frame, and a possible replication origin upstream of the promotor. Significant homologies were observed between the amino acid sequence predicted from the 999-bp open reading frame and those of the similar putative proteins encoded by the *B. amyloliquefaciens* plasmid pFTB14 (13) and *Staphylococcus aureus* plasmid pC194 (3, 7).

To facilitate identification of the replication region of

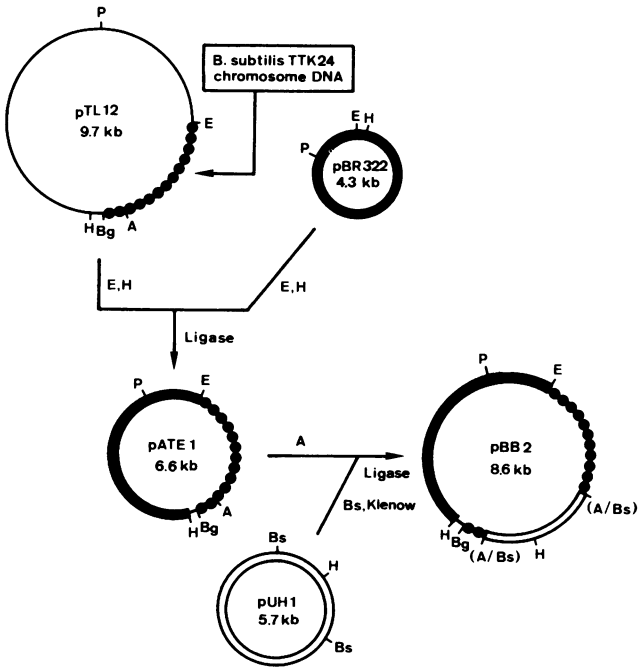

FIG. 1. Derivation of plasmids used in the present study. The chain of solid circles in the diagram indicates the DNA segment containing the Tmp^r gene of *B. subtilis* TTK24. Heavy and thin lines represent the regions of pBR322 and pTL12, respectively. Double lines represent the DNA fragment of pUH1. A, *Aat*I; Bg, *Bgl*II; Bs, *Bst*EII; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I.

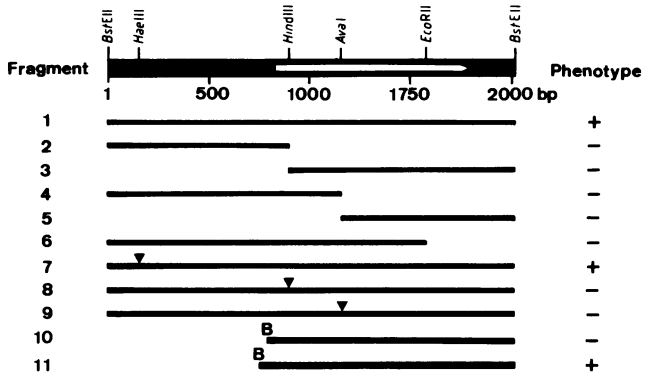

FIG. 2. Structure and replication activity of the modified fragments of the 2.0-kb *ori* fragment. The open arrow in the restriction map indicates an open reading frame found in the sequence data (see Fig. 3). + and − indicate, respectively, ability and inability to replicate in the *B. subtilis* host. Modifications were made by deletion with restriction enzymes or insertion of a few base pairs (closed triangle), using the Klenow fragment, after digestion with the indicated restriction enzyme or by *Bal*31 digestion (B) from the *Bgl*II site of pBB2.

* Corresponding author.

† Present address: Department of Applied Microbiology and Technology, Kumamoto Institute of Technology, Ikeda, Kumamoto 860, Japan.

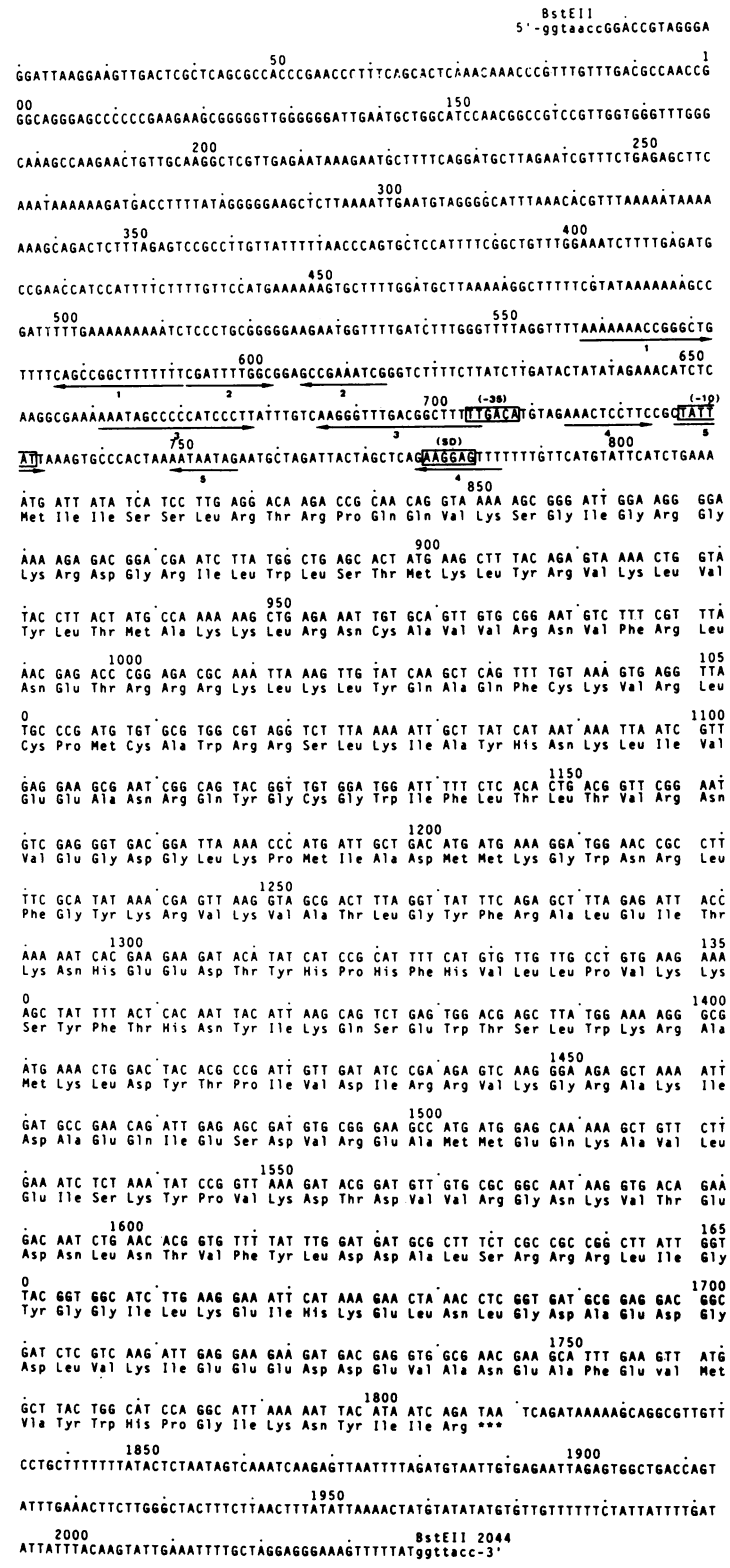

FIG. 3. Nucleotide sequence of the 2.0-kb *Bst*EII DNA fragment. −10, −35, and SD indicate, respectively, plausible sequence for the −10 and −35 sequences and the ribosome binding site. The nucleotide sequences indicated by lowercase latters at the two termini originate from the *Bst*EII site.

pUH1, we used the trimethoprim-resistant (Tmp^r) dihydrofolate reductase gene of *B. subtilis* 168. A schematic presentation of the constructed plasmids is given in Fig. 1. The dihydrofolate reductase-coding gene was derived from a Tmp^r strain, TTK24, of *B. subtilis* 168 (20) and has been cloned in plasmid pBR322 of *Escherichia coli* (2). DNAs from pTL12, carrying the Tmp^r dihydrofolate reductase gene, which was constructed by Tanaka and Kawano (18), and pBR322 were both treated with *Eco*RI and *Hin*dIII, mixed, and ligated with T4 ligase; then pATE1 was constructed. Natto plasmid pUH1 was digested with *Bst*EII, and then the ends were filled in with Klenow fragment to generate a blunt end. The DNA fragments were mixed and ligated at the *Aat*I site of pATE1 with T4 ligase and then added to *B. subtilis* MI112 (19) protoplasts. Several Tmp^r colonies were obtained on AA agar plates (18) containing trimethoprim (1 μg/ml), and one of them was used for further study. Plasmid pBB2 carried in such a Tmp^r colony had a molecular mass of 8.4 kb (Fig. 1). The physical map of pBB2, using various restriction enzymes, is shown in Fig. 1.

To define the boundaries of a functional unit of pUH1 replication, a 2.0-kb *Bst*EII fragment of pUH1 was digested with selected restriction endonucleases to obtain a set of overlapping DNA fragments. Digests modified with Klenow fragment were ligated with pATE1, followed by introduction into *E. coli* by transformation and successive selection for ampicillin resistance. The plasmid DNA preparations containing each generated fragment were tested for replication activity in *B. subtilis*. The results are summarized in Fig. 2. Plasmid pBB2, which corresponds to fragment 1 in Fig. 2, and a derivative containing fragment 7 specified Tmp^r in *B. subtilis*, but attempts to transform *B. subtilis* with recombinant plasmid preparations containing the small fragments (fragments 2 to 6 in Fig. 2) and inserted fragments (fragments 8 and 9) were unsuccessful.

The nucleotide sequence of the 2.0-kb *Bst*EII fragment was determined by the method of Sangar et al. (14). Though the strategy is not shown, the nucleotide sequence was determined for both strands by using numerous restriction fragments to give enough overlapping regions. By examination of possible open reading frames, we found only one large frame (Fig. 2), designated *rep*, which consists of 999 bp and encodes a protein molecule with 333 amino acids and an M_r of 39,074. The 5′ upstream region of *rep* contains a 5′-AAGGAG-3′ sequence (indicated as SD in Fig. 3) complementary to the 3′ end of 16S rRNA (3′-OH-UCUUUC CUCCAGUAG-5′) of *B. subtilis* (11) at nucleotides 777 to 782 for translation initiation. There is a 5′-TATTAT-3′ sequence (−10 in Fig. 3) resembling a Pribnow box (12) at nucleotides 727 to 732, and at a site 24 bp upstream of this −10 sequence, there is a 5′-TTGACA-3′ sequence resembling the −35 consensus sequence (12) of the *B. subtilis* gene. The observed distance (18 bp) between the −10 and −35 sequences accords well with that observed generally in *B. subtilis* genes (17 to 18 bp) (12).

Plasmid pBB2 was digested with *Bgl*II and then treated for 20 min with exonuclease *Bal*31 under conditions in which about 50 bp was removed per min from each end of the DNA molecule. After ligation by T4 ligase, the DNA was introduced into *B. subtilis* by protoplast transformation. Tmp^r transformants, which contained fragment 11 in Fig. 2, were obtained at high efficiencies with *Bal*31-generated deletions of 581 bp (as determined by nucleotide sequencing), while no transformants were obtained with a similar 653-bp deletion (fragment 10 in Fig. 2), suggesting that the 5′ upstream portion is dispensable for replication. The interaction of the initiator protein with its binding site on the DNA is a key step in the replication initiation process (4, 21). Therefore, the structure of the 2.0-kb *ori* sequence that is necessary and sufficient for the replication of the Natto plasmid pUH1 molecule contains a 999-bp open reading frame, *rep*, a promotor region for *rep* expression, and a putative replication origin for the plasmid, which is located upstream of the promoter between positions 581 and 653.

```
                    10         20         30         40         50         60
rep         ------MIISSLRTRPQQVKSGIGRGKRDGRILWLSTMKLYRVKLVYLTMAKKLRNCAVV
rep14       MYSSENDYRILEDKTATGKKRDWRGKKRRANLMAEHYEALEKRIGAPYYGKKAERLSECA
RepB        ------------------------------------------------------------
Protein A   ----------------------MCYNMEKYTEKKQRNQVFQKFIKRHIGENQMDLVEDCN

                    70         80         90         100        110        120
rep         RNVFRLNETRRRKLKLYQAQFCKVRLCPMCAWRRSLKIAYHNKLIVEEANRQYGCGWIFL
rep14       EMLSFKRDPETGRLKLYQAHFCKVRLCPMCAWRRSLKIAYHNKLIIEEANRQYGCGWIFL
RepB        ----------------------------------MKHGIQSQKVVAEVIKQKPTVRWLFL
Protein A   TFLSFVADKTLEKQKLYKANSCKNRFCPVCAWRKARKDALGLSLMMQYIKQQEKKEFIFL

                    130        140        150        160        170        180
rep         TLTVRNV-EGDGLKPMIADMMKGWNRLFGYKRVKVATLGYFRALEITKNHEEDTYHPHFHV
rep14       TLTVRNV-KGERLKPQISEMMEGFRKLFQYKKVKTSVLGFFRALEITKNHEEDTYHPHFHV
RepB        TLTVKNVYDGEELNKSLSDMAQGFRRMMQYKKINKNLVGFMRATEVTINNKDNSYNQHMHV
Protein A   TLTTPNVMSDE-LENEIKRYNNSFRKLIKRKKVGSVIKGYVRKLEITYNKKDNSYNQHMHV

                    190        200        210        220        230        240
rep         LLPVKKSYF--THNYIKQSEWTSLWKRAMKLDYTPIVDIRRVKGRAKIDAEQISDVREAMM
rep14       LLPVKRNYF--GKNYIKQAEWTSLWKRAMKLDYTPIVDIRRVKGRVKIDAEQISDVREAMM
RepB        LVCVEPTYFKNTENYVNGKQWIQFWKKAMKLDYDPNV-----------KVQMIRPKNKYKSD
Protein A   LIAVNKSYFTDKRYYISQQEWLDLW-----RDVTGISEITQVQVQKIRQNNNKELYEMAKYS

                    250        260        270        280        290        300
rep         EQKAVLEISKYPVKDTDVVRGNKVTEDNLNTVFYLDDALSRRRLIGYGGILKEIHKELNL
rep14       EQKAVLEISKYPVKTDDVVRGSKVTDDNLNTVFYLDDALSARRLIGYGGILKEIHKELNL
RepB        IQSAIDETAKYPVKDTD--FMTDDEEKNLKRLSDLEEGLHRKRLISYDGLLKEIHKKLNL
Protein A   GKDSDYLINKSKSL-------------------------------------------------

                    310        320        330
rep         GDAEDGDLVKIEEEDDEVANEAFEVMAYWHPGIKNYIIR-
rep14       GDAEGGDLVKIEEEDDEVANGAFEVMAYWHPGIKNYILK-
RepB        DDTEEGDL-IHTDDDEKADEDGFSIIAMWNWERKNYFIKE
Protein A   -------------------------------------
```

FIG. 4. Comparison of the amino acid sequences of *rep* of pUH1, rep_{14} of pFTB14, RepB of pUB110, and protein A of pC194. Amino acid identities with *rep* of pUH1 are indicated (*). Amino acid numbers follow the sequence of the *rep* protein from the amino-terminal methionine to the carboxyl terminus. Gaps have been inserted to gain maximum matching. The one-letter amino acid code has been used.

The amino acid sequence of the *rep* protein coding region of pUH1 was compared with a number of protein sequences registered in GenBank by using the homology search system of GENAS (10). Homologies between the predicted amino acid sequence of the *rep* protein of pUH1 and those of the *rep* protein of pFTB14, RepB of pUB110, and protein A of pC194 are illustrated in Fig. 4. A high similarity (71.8%) was found between the *rep* protein of pUH1 and rep_{14} of pFTB14 in *B. amyloliquefaciens*. The structure of the replication origin sequence of pFTB14 contains an open reading frame (*rep*), stretching for 1,017 bp, a promoter region for *rep* expression, and a possible replication origin for *rep* expression, which is located upstream of the promoter. The *rep* product is *trans* active and essential for plasmid replication (13). Compared with *rep* of pUH1, a low similarity (39.7%) was found for RepB of pUB110 (1) and for protein A (29.2%) of pC194 (8) in *S. aureus*. Khan et al. (9) determined the start site of pT181 DNA synthesis within a 127-bp segment and showed that a 168-bp segment containing the replication start site is enough to initiate unidirectional replication. Furthermore, like RepC protein of pT181 (9), the protein of *E. coli* plasmid R6K (4, 17, 18) does not have any significant homology in its amino acid sequences with those of the *rep* proteins of pUH1 and pFTB14 deduced from the open reading frame (data not shown). The Rep protein of Natto plasmid pUH1 shows high homology with replication proteins encoded by pFTB14, pC194, and pUB110, which originate from various gram-positive bacteria such as *B. amyloliquefaciens* and *S. aureus*. Taxonomic studies (based

on the 16S rRNA similarities) revealed that *Bacillus* and *S. aureus* strains are more related than *Streptococcus* and *Lactobacillus* strains (16). The plasmid homologies suggest an exchange of plasmid replicons by recent horizontal transfer through the different genera, including "natto" *Bacillus* species.

We are indebted to T. Tanaka for the generous supply of plasmid pTL12 and also to S. Kuhara for the computer search of protein sequences.

REFERENCES

1. **Ano, T., T. Imanaka, and S. Aiba.** 1986. The copy number of *Bacillus* plasmids pRBH1 is negatively controlled by RepB protein. Mol. Gen. Genet. **202**:416–420.
2. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, and H. W. Boyer.** 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene **2**:95–113.
3. **Dagert, M., I. Jones, A. Goze, S. Romac, B. Niaudet, and S. D. Ehrlich.** 1984. Replication functions of pC194 are necessary for efficient plasmid transduction by M13 phage. EMBO J. **3**:81–86.
4. **Germino, J., and D. Bastia.** 1983. Interaction of the plasmid R6K-encoded replication initiator protein with its binding sites on DNA. Cell **34**:125–134.
5. **Hara, T., A. Aumayr, Y. Fujio, and S. Ueda.** 1982. Elimination of plasmid-linked polyglutamate production by *Bacillus subtilis* (*natto*) with acridine orange. Appl. Environ. Microbiol. **44**: 1456–1458.
6. **Hara, T., A. Ishizaki, and S. Ueda.** 1986. Formation of heteroduplex molecules between plasmids pUH1 and pLS11 in polyglutamate-producing *Bacillus* strains. Agric. Biol. Chem. **50**: 2391–2394.
7. **Horinouchi, S., and B. Weisblum.** 1982. Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. J. Bacteriol. **150**:815–825.
8. **Iordanescu, S., and M. Surdeanu.** 1983. Isolation and complementation of temperature-sensitive replication mutants of *Staphylococcus aureus* plasmid pC194. Mol. Gen. Genet. **191**: 201–206.
9. **Khan, S. H., S. M. Carlton, and R. P. Novick.** 1982. Functional origin of replication of pT181 plasmid DNA is contained within a 168-base-pair segment. Proc. Natl. Acad. Sci. USA **79**:4580–4584.
10. **Kuhara, S., F. Matsuo, S. Futamura, A. Fujita, T. Shinohara, T. Takagi, and Y. Sakaki.** 1984. GENAS: a database system for nucleic acid sequence analysis. Nucleic Acids Res. **12**:89–99.
11. **McLaughlin, J. R., C. L. Murray, and J. C. Rabinowitz.** 1981. Unique features in the ribosome binding site sequence of the gram-positive *Staphylococcus aureus* β-lactamase gene. J. Biol. Chem. **256**:11283–11291.
12. **Moran, C. P., Jr, N. Lang, S. F. J. LeGrice, G. Lee, M. Stephans, A. L. Sonenshein, J. Pero, and R. Losick.** 1982. Nucleotide sequence that signal the initiation of transcription and translation in *Bacillus subtilis*. Mol. Gen. Genet. **186**:339–346.
13. **Murai, M., H. Miyashita, H. Araki, T. Seki, and Y. Oshima.** 1987. Molecular structure of the replication origin of a *Bacillus amyloliquefaciens* plasmid pFTB14. Mol. Gen. Genet. **210**:92–100.
14. **Sangar, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5476.
15. **Scott, J. R.** 1984. Regulation of plasmid replication. Microbiol. Rev. **48**:1–23.
16. **Stackebrandt, E., and M. Teuber.** 1988. Molecular taxonomy and phylogenetic position of lactic acid bacteria. Biochimie **70**:317–324.
17. **Stalker, D. M., R. Kolter, and D. R. Helinski.** 1982. Plasmid R6K DNA replication. I. Complete nucleotide sequence of an autonomously replicating segment. J. Mol. Biol. **161**:33–43.
18. **Tanaka, T., and N. Kawano.** 1980. Cloning vehicles for the homologous *Bacillus subtilis* host-vector system. Gene **10**:131–136.
19. **Tanaka, T., and K. Sakaguchi.** 1978. Construction of a recombinant plasmid composed of *B. subtilis* leucine genes and a *B. subtilis* (*natto*) plasmid: its use as cloning vehicle in *B. subtilis* 168. Mol. Gen. Genet. **165**:269–276.
20. **Wainscott, V. J., and J. F. Kane.** 1976. Dihydrofolate reductase in *Bacillus subtilis*, p. 208–213. *In* D. Schlessinger (ed.), Microbiology—1976. American Society for Microbiology, Washington, D.C.
21. **Yamaguchi, K., and M. Yamaguchi.** 1984. The replication origin of pSC101: the nucleotide sequence and replication functions of the ori region. Gene **29**:211–219.