九州大学学術情報リポジトリ
Kyushu University Institutional Repository

Complete Nucleotide Sequence of Bacillus subtilis (natto) Plasmid Responsible for γ-Polyglutamate Synthesis

Hara, Toshio
Microbial Genetics Division, Institute of Genetic Resources, Faculuty of Agriculture, Kyushu University

Yasuda, Shuhey
The Riken Institute

Ogata, Seiya
Microbial Genetics Division, Institute of Genetic Resources, Faculuty of Agriculture, Kyushu University

Soeda, Eiichi
The Riken Institute

他

https://doi.org/10.5109/23986

出版情報：九州大学大学院農学研究院紀要. 36 (3/4), pp.209-218, 1992-02. Kyushu University
バージョン：
権利関係：

KYUSHU UNIVERSITY

J. Fac. Agr., Kyushu Univ., 36 (3•4), 209-218 (1992)

Complete Nucleotide Sequence of *Bacillus sub tilis (na tto)* Plasmid Responsible for γ-Polyglutamate Synthesis

Toshio Hara*, Shuhey Yasuda[1], Seiya Ogata, Eiichi Soeda[1], and Seinosuke Ueda[2]

Microbial Genetics Division, Institute of Genetic Resources, Faculuty of Agriculture,
Kyushu University, Hakozaki, Higashi-ku, Fukuoka 812,
[1]The Riken Institute, Koyadai, Tsukuba, Ibarak 305, and
[2]Department of Applied Microbial Technology,
Kumamoto Institute of Technology, Ikeda, Kumamoto 860, Japan.

(Received October 25, 1991)

The complete 5,812 base pairs nucleotide sequence has been determined for *Bacillus subtilis (natto)* plasmid pUH1, which encodes a ymglutamyltranspeptidase (EC 2.3.2.2) for synthesis of y-polyglutamate. Nucleotide sequence analysis reveals two long open reading frames coding for the gene for replication *(rep)* protein and y-glutamyltranspeptidase gene. Several putative regulatory sequences conserved were found upstream from these frames including the Shine-Dalgarno sequence, Pribnow box and the "-35 region" at preferred distances for efficient transcription. The amino acid sequences predicted from the rep and ymglutamyltranspeptidase coding frames showed a significant homology, in each small segment, with those of E229 coded from the plasmid pC194 and of C403 from the plasmid pE194 of *Staphylococcus aureus*, respectively. However, no convincing homology of the pUH1 predicted ymglutamyltranspeptidase protein could be found with *E. coli* and mammalian γ-glutamyltranspeptidases.

INTRODUCTION

"Natto" is one of the traditional non-salty fermented foods which is often served at breakfast in Japan. It is manufactured by growing *Bacillus subtilis (natto)* on steamed soybeans. It is an adhesive, and consists of polysaccharide (levan-form fructan) and y-polyglutamate. The adhesive materials are mainly composed of γ-polyglutamate containing D- and L-glutamate in various proportions (Fujii, 1963). A plausible mechanism of the biosynthetic pathway of y-polyglutamate has been proposed by Thorne *et al.* (1955) for one of the strains producing y-polyglutamate (B. *licheniformis* ATCC 9945A). However, since it is known that various strains differ a great deal in the basic requirements for synthesis of the capsule, it is to be assumed that there is more than one biosynthetic pathway.

We showed that a 5.7-kb plasmid designated pUH1, which encodes γ-glutamyltranspeptidase (EC 2.3.2.2) gene responsible for y-polyglutamate synthesis, is distributed widely in *Bacillus subtilis (natto)* isolated from a commercial product, natto (Hara *et al.*, 1983). In a previous paper (Hara et *al.*, 1991), we reported that the 2.0-kb

* Corresponding author

Bst EII DNA fragment of pUH1 contains a gene for replication protein (*rep*) of 999-bp in size. To investigate the participation of a plasmid in γ-polyglutamate synthesis, we performed DNA sequencing of natto plasmid pUH1. In the present work, we report the complete nucleotide sequence of the plasmid pUH1, including data pertinent to the mapping of biological functions and determinants of their expression and regulation.

MATERIALS AND METHODS

Bacterial strain and plasmids

Escherichia coli JM 101 was used for a cloning host. Plasmids pUH1 (Hara *et al.*, 1982) and pATE1 (Hara *et al.*, 1991) have been previously described, respectively.

Enzyme assay

Gamma-glutamyltranspeptidase activity was assayed as previously described (Aumayr et *al.*, 1981).

DNA manipulation

The plasmid pUH1 from *Bacillus subtilis (natto)* was prepared and purified as described previously (Hara *et al.*, 1983). Plasmid pUH1 was sonicated to generate suitable sizes, processed with nuclease P1 and T4 DNA polymerase, and separated by electrophoresis on agarose gels. After elution from gels, the fragments were inserted at *Sma*I cloning site of Ml3 mp8 DNA. After transformation into *E. coli* JM101, a shotgun library was constructed by the modified method (Yasuda *et al.*, 1984) of Deininger (Deininger, 1983). The recombinant phages selected from library were sequenced.

DNA sequencing

DNA sequencing was carried out by means of the dideoxy chain termination method (Sanger *et al.*, 1977) with Ml3 sequencing kit. Nucleotide and deduced amino acid sequences were analyzed by the Hitachi DNASIS system.

Chemicals

Restriction endonucleases, T4 DNA polymerase, and Ml3 cloning and sequencing kits were obtained from Takara Shuzo Co., Ltd., and used according to the suppliers specifications. Nuclease P1 was from Yamasa Shoyu Co., Ltd. (α-32P)dCTP was purchased from RCC Amersham.

RESULTS AND DISCUSSION

Location of biological functions on pUH1

As reported previously (Hara *et al.*, 1982), y-glutamyltranspeptidase gene, which is responsible for polyglutamate synthesis, might be encoded on an endogeneous plasmid, pUH1. To define the bounds of a biological functional unit of pUH1, we constructed a set of plasmid derivatives using pUB110 and pATE1 (Hara *et al.*, 1991). Recombinant plasmids were selected in *B. subtilis* host by using resistance to kanamycin (Km) for pUB110 and trimethoprim (Tmp) for pATE1. The generated BstEII and Hind111 fragments of pUH1 were filled in with Klenow fragment, and cloned into P*vu*II site of pUB110 and *Aat*I site of pATE1, respectively. The results are summarized in Fig. 1. The composite plasmid pBB2 contains a 2.0-kb BstEII fragment (column BsS in Fig. 1) of pUH1 which confers autonomous replication on the plasmid

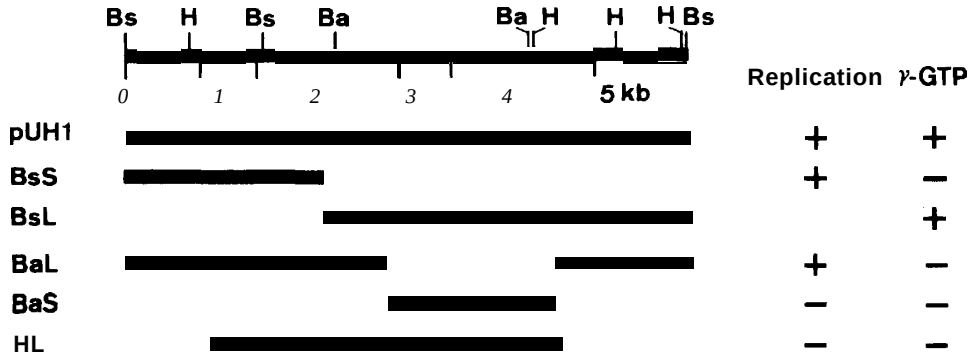

Fig. **1.** Structure and location of biological functions of 5.7-kb plasmid pUH1.
+ and -indicate, respectively, ability and inability to replicate and to synthesis γ-glutamyltranspeptidase in the *B.subtilis* host. Modifications were made by deletion with restriction enzymes.

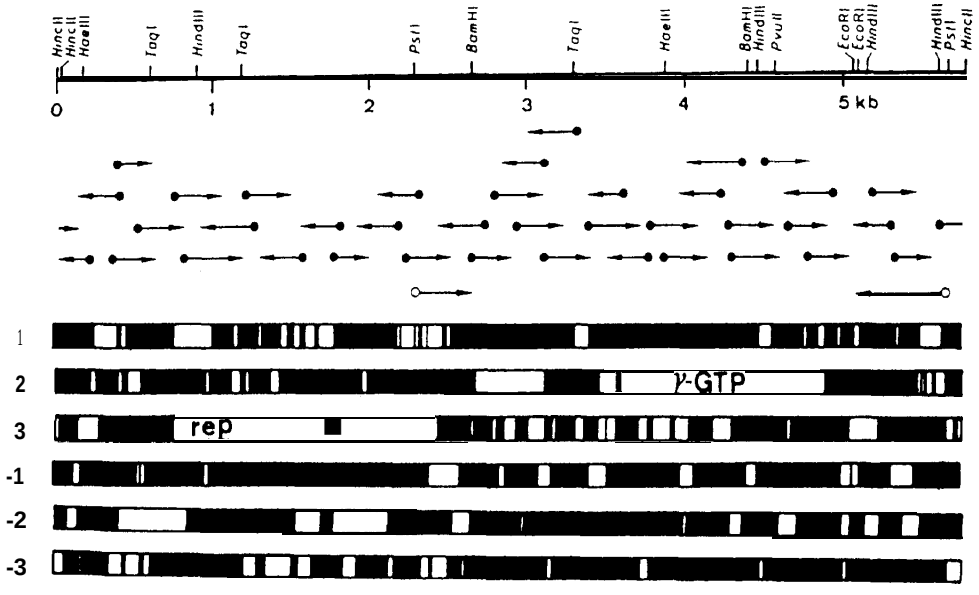

Fig. 2. Strategy for sequencing the plasmid pUH1, and the coding frames.
(Upper) The physical map of pUH1 and strategy for nucleotide sequence. The 5' end of *Hinc*II site of pUH1 is taken as the start point for numbering the nucleotide sequence. The extent of sequencing is specified by arrows. The open and closed circles represent the 5' termini of the cloned fragments generated by restriction enzyme digestion and sonication, respectively. (Lower) The coding frame of the presented sequences: □, open coding frame; ■ , noncoding region.

in the *B. subtilis* host. The 2.0-kb *Bst*EII fragment is necessary and sufficient for the plasmid replication, whose molecular structure has been previously characterized (Hara *et al.*, 1991). The Km' transformants harboring plasmid pPB1, which contain a 3.7-kb *Bst*EII fragment (column BsL in Fig. 1) of pUH1, can produce γ-glutamyltranspeptidase. The enzyme activity of Km' transformants harbaring pUU1, which contain the 3.6kb *Hin*dIII fragment (column HL in Fig. 1), could not detected in *B. subtilis* host.

DNA sequence analysis of pUH1

The strategy for sequencing is outlined in Fig. 2. Restriction sites determined preliminarily by enzyme digestions were confirmed by sequence analysis. The complete nucleotide sequence of pUH1 DNA comprises 5,812 bp (Fig. 3). The DNA sequence contains two open translation reading frames, whose directions are same. Of the two, the longer open reading frame which corresponds to γ-glutamyltranspeptidase, starting with ATG at nucleotide 3,650 and terminating at the stop codon at nucleotide 4,911, is capable of coding for polypeptide of 420 amino acid residues with a *M*r of 49,356 daltons. Upstream from the y-glutamyltranspeptidase, putative regulatory sequences can be identified, including a Shine-Dalgarno sequence (AAC-GAG), which is complementary to the 3' end of 16S rRNA (3'-OH-UCUUUCCUC-CAGUAG-5') of *B. subtilis* (McLaughlin *et al.*, 1981), at nucleotides 3,616 to 3,621, a σ^{43} RNA polymerase recognition site (TTCAAA) at nucleotides 3,555 to 3,560, and a Pribnow box (TATTAT) at nucleotides 3,578 to 3,682. The distance (17 bp) between the recognition site and a Pribnow box is shorter than the preferred one in *B. subtilis* (Moran *et al.*, 1982). It is well known that y-polyglutamate productivity of *B. subtilis (natto)* is decreased during stock on a nutrient agar slant over long period. The instability of y-polyglutamate synthesis might be due to inefficient regulation on transcriptional level of y-glutamyltranspeptidase.

The shorter frame designated rep, starting with ATG at nucleotide 816 and terminating at the stop codon at nucleotide 1,815, is capable of coding for a polypeptide of 333 amino acid residues with a *M*r of 39,074 daltons. The nucleotide sequence upstream from the rep gene is characterized by the presence of typical regulatory signals. They include a Shine-Dalgarno sequence (AAGGAG) at nucleotides 783 to 788 for translation initiation, the sequence (TTGACA) at nucleotides 709 to 714 for a σ^{43} RNA polymerase recognition site and a Pribnow box (TATTAT) at nucleotides 733 to 738, at preferred distances (17-18 bp) from each other. To discover the sequences essential for replication, a homology search was performed to determine whether there are sequences within this region conserved among other *Bacillus* spp. and *S. aureus* plasmids known to replicate in *B. subtilis.* Results showed that a 26-bp sequence (TTTCTTATCTTGATACTATATAGAAA) at nucleotides 626 to 651 was conserved in pUH1, pBAA1 (Devine *et al.*, 1989), pFTB14 (Murai *et al.*, 1987), pUB110 (Muller *et al.*, 1986), and pC194 (Dagert *et al., 1984;* Horinouchi and Weisblum, 1982b). Especially, the consensus sequence (CTTGATA) is found in the hairpin region of ϕX174 shown by Shlomai and Kornberg (1980) to have origin activity. However, no significant homology was found with pE194 (Horinouchi and Weisblum, 1982a; Villafane *et al.*, 1987), and pT181 (Khan *et al., 1982).*

Amino acid sequence comparison

The amino acid sequence of the predicted proteins encoded on pUH1 was compar-

```
                                                  50                                                  90
GTCAACGGTAACCGGACCGTAGGGAGGATTAAGGAAGTTGACTCGCTCAGCGCCACCCGAACCCTTTCAGCACTCAAACAAACCCGTTTG

       100                                                150
TTTGACGCCAACCGGGCAGGGAGCCCCCCGAAGAAGCGGGGGTTGGGGGGATTGAATGCTGGCATCCAACGGCCGTCCGTTGGTGGGTTT

          .        200         .         .         .         .        250         .         .
GGGCAAAGCCAAGAACTGTTGCAAGGCTCGTTGAGAATAAAGAATGCTTTTCAGGATGCTTAGAATCGTTTCTGAGAGCTTCAAATAAAA

                                 300                                                 350
AAGATGACCTTTTATAGGGGGAAGCTCTTAAAATTGAATGTAGGGGCATTTAAACACGTTTAAAAATAAAAAAAGCAGACTCTTTAGAGT

                                                 400                                                 450
CCGCCTTGTTATTTTTAACCCAGTGCTCCATTTTCGGCTGTTTGGAAATCTTTTGAGATGCCGAACCATCCATTTTCTTTTGTTCCATGA

          .         .         .         .         .         500         .         .         .         .
AAAAAGTGCTTTTGGATGCTTAAAAAGGCTTTTTCGTATAAAAAAAGCCGATTTTTGAAAAAAAAATCTCCCTGCGGGGGAAGAATGGTT

       550                                                600
TTGATCTTTGGGTTTTAGGTTTTAAAAAAACCGGGCTGTTTTCAGCCGGCTTTTTTTCGATTTTGGCGGAGCCGAAATCGGGTCTTTTCT

                 650                                                 700         (-35)      .
TATCTTGATACTATATAGAAACATCTCAAGGCGAAAAAATAGCCCCCATCCCTTATTTGTCAAGGGTTTGACGGCTTTTTGACATGTAGA

          .  (-10)  .        750                                 .  (SD)   .        800
AACTCCTTCCGCTATTATTAAAGTGCCCACTAAAATAATAGAATGCTAGATTACTAGCTCAGAAGGAGTTTTTTTGTTCATGTATTCATC

          .         .         .         .         850         .         .                                 900
TGAAAATGATTATATCATCCTTGAGGACAAGACCGCAACAGGTAAAAAGCGGGATTGGAAGGGGAAAAAGAGACGGACGAATCTTATGGC
     MetIleIleSerSerLeuArgThrArgProGlnGlnValLysSerGlyIleGlyArgGlyLysArgAspGlyArgIleLeuTrpL
          .         .         .         .         .         950         .         .         .         .
TGAGCACTATGAAGCTTTACAGAGTAAAACTGGTATACCTTACTATGGCAAAAAAGCTGAGAAATTGTGCAGTTGTGCGGAATGTCTTTC
euSerThrMetLysLeuTyrArgValLysLeuValTyrLeuThrMetAlaLysLysLeuArgAsnCysAlaValValArgAsnValPheA
      1000         .         .         .         .         1050         .         .         .
GTTTAAACGAGACCCGGAGACGCAAATTAAAGTTGTATCAAGCTCAGTTTTGTAAAGTGAGGTTATGCCCGATGTGTGCGTGGCGTAGGT
rgLeuAsnGluThrArgArgArgLysLeuLysLeuTyrGlnAlaGlnPheCysLysValArgLeuCysProMetCysAlaTrpArgArgS
          .        1100         .         .         .         .        1150         .         .
CTTTAAAAATTGCTTATCATAATAAATTAATCGTTGAGGAAGCGAATCGGCAGTACGGTTGTGGATGGATTTTTCTCACACTGACGGTTC
erLeuLysIleAlaTyrHisAsnLysLeuIleValGluGluAlaAsnArgGlnTyrGlyCysGlyTrpIlePheLeuThrLeuThrValA
          .         .        1200         .         .         .         .        1250         .
GGAATGTCGAGGGTGACGGATTAAAACCCATGATTGCTGACATGATGAAAGGATGGAACCGCCTTTTCGGATATAAACGAGTTAAGGTAG
rgAsnValGluGlyAspGlyLeuLysProMetIleAlaAspMetMetLysGlyTrpAsnArgLeuPheGlyTyrLysArgValLysValA
          .         .         .        1300         .         .         .                                 1350
CGACTTTAGGTTATTTCAGAGCTTTAGAGATTACCAAAAATCACGAAGAAGATACATATCATCCGCATTTTCATGTGTTGTTGCCTGTGA
laThrLeuGlyTyrPheArgAlaLeuGluIleThrLysAsnHisGluGluAspThrTyrHisProHisPheHisValLeuLeuProValL
                                                 1400         .         .         .         .         .
AGAAAAGCTATTTTACTCACAATTACATTAAGCAGTCTGAGTGGACGAGCTTATGGAAAAGGGCGATGAAACTGGACTACACGCCGATTG
ysLysSerTyrPheThrHisAsnTyrIleLysGlnSerGluTrpThrSerLeuTrpLysArgAlaMetLysLeuAspTyrThrProIleV
      1450         .         .         .         .         1500         .         .         .
TTGATATCCGAAGAGTCAAGGGAAGAGCTAAAATTGATGCCGAACAGATTGAGAGCGATGTGCGGGAAGCCATGATGGAGCAAAAAGCTG
alAspIleArgArgValLysGlyArgAlaLysIleAspAlaGluGlnIleGluSerAspValArgGluAlaMetMetGluGlnLysAlaV
                1550                                                1600
TTCTTGAAATCTCTAAATATCCGGTTAAAGATACGGATGTTGTGCGCGGCAATAAGGTGACAGAAGACAATCTGAACACGGTGTTTTATT
alLeuGluIleSerLysTyrProValLysAspThrAspValValArgGlyAsnLysValThrGluAspAsnLeuAsnThrValPheTyrL
                                 1650                                                1700
TGGATGATGCGCTTTCTCGCCGCCGGCTTATTGGTTACGGTGGCATCTTGAAGGAAATTCATAAAGAACTAAACCTCGGTGATGCGGAGG
euAspAspAlaLeuSerArgArgArgLeuIleGlyTyrGlyGlyIleLeuLysGluIleHisLysGluLeuAsnLeuGlyAspAlaGluA
          .         .         .        1750         .         .         .         .        1800
ACGGCGATCTCGTCAAGATTGAGGAAGAAGATGACGAGGTGGCGAACGAAGCATTTGAAGTTATGGCTTACTGGCATCCAGGCATTAAAA
spGlyAspLeuValLysIleGluGluGluAspAspGluValAlaAsnGluAlaPheGluValMetAlaTyrTrpHisProGlyIleLysA
                                                 1850
ATTACATAATCAGATAAAAAGCAGGCGTTGTTCCTGCTTTTTTATACTCTAATAGTCAAATCAAGAGTTAATTTTAGATGTAATTGTGA
snTyrIleIleArg***  ------>          <------
```

Fig. 3. Complete nucleotide sequence of pUH1.

Nucleotides are numbered from the 5' end of *Hin*cII site of the DNA strand of the pUH1 DNA with the same polarity as mRNA. Amino acid sequences are also shown on the coding frames. Regulatory signal sequences are underlined. Putative termination signals are indicated by arrows.

```
      1900                                                1950
GAATTAGAGTGGCTGACCAGTATTTGAAACTTCTTGGGCTACTTTCTTAACTTTATATTAAAACTATGTATATATGTGTTGTTTTTTCTA

                 2000                                                2050
TTATTTTGATATTATTTACAAGTATTGAAATTTTGCTAGGAGGGAAAGTTTTTATGGTTACCACGATTGGTAAAAGTAAGATGTGGGTAG

          .                2100                                                2150
GTATTATTGTTGTATTATCTTTATTATTGGTATCTTTTTCGCCTGCTGTAAAGGCTGATACTAAGGATAAATATTATTCTACAACTTCTA

          .          .                2200                                                2250
CCCAATCTTCAACAAAAAGTTATCGTAAAGCTAATACTAGTGGGGTCTATGTGAAAGTACTTAAAGCTGGACGTTCTCGTGATGTTGCTA

          .          .                                2300
TTTCAGTTTTTGCTGATGCGAACAAAGGGAAAGGAAAGCCACATTGGGTAAATGTTTCTGGTAGCGATGGCGCTACTCTGGGAAAATACG

      2350                                                2400
TGACTGCAGGGCATACATATCATCTTACAAACTATGCTGTAGAACGTTATGGAAAGAATGTTCCTATACAATTATTTGTTTCTAATGGTT

          .      2450                                                2500
CTGGCAAAAAAGTTGAATTTTATTGGAGTCCCGATTGTAGATAGTCCAAATAGCAGATGATGAAAAAGCAGGATTAATCCTCTTTTTTAT

                           2550                                                2600
TTTTGTTTTGGTAAAATGTGATAAGCGGGTTTTGAAATATAGAGGAGGAAATTTTCTTTTGACAAAAGCAGAAGTTTTCAAAAAAGAAAA

          .                2650
GCTCTATTTTAGAGTGGGGCAAGGCTATTGTGATAGCTGTTATACTTGCTCTTCTTATTCGGATTTTTTGTTTGAGCCATATGTGGTAGA

                                                2150
AGGGAAGTCTATGGATCCCACTTTGGTTGATTCTGAAAGATTATTTGTAAATAAGACTGTGAAGTATACAGGTAATTTTAAACGAGGGGA

      2800                                                2850
TATAATAATTTTAAACGGAAAGGAAAAAAGCACACATTATGTGAAGCGATTAATTGGTTTACCTGGAGACACTGTAGAAATGAAGAATGA

          .      2900                                                2950
CCACCTTTTTATTAATGGAAATGAAGTTAAGGAACCATATCTTTCTTATAATAAAGAAAATGCTAAGAAAGTGGGTATAAACCTTACAGG

                           3000                                                3050
AGATTTTGGACCAATTAAAGTTCCAAAAGATAAATATTTTGTTATGGGCGATAACCGACAAGAATCAATGGATAGTCGTAATGGGCTTGG

          .                3100                                                3150
ACTCTTTACTAAAGATGATATTCAGGGAACCGAAGAGTTCGTATTTTTTCCATTTAGTAATATGCGAAAAGCTAAATAATTTTTAGACCG

                                                3200
AAACAGGCTATAAGGTCTGTTTTTTCATTTTGTTATGAACATTTAGCACACAGATCAAAGTTTTCATAGTTTGAATGCTTTGATAGCAGC

      3250                                                3300
AAAGGGTATTTCTGATTTTCTGCCGATCTCTCATCGGCGGAAAAGTCGGGTCGGCGGACAGCCGACAAGTGGCACGAACTTTCGATGCGA

          .      3350          .          .          .          .      3400          .          .
CAGCGAGAATGAGAGACGCACCAGCACCGCAGGTCGCACGTCCAAATTTGCCATGGCATAATTTGGTGTAGTGCGTTACACCAAAGATAA

                           3450                                                3500
ACTTTGTGTTACCATAACCCCTATACAGTGGTCTGAATCGGGGGTTTTTCTCATGGCAAATTATGCAGTCATCAGGATGGAAAAATACAA

          .          .                3550      (-35)          .          (-10)      .      3600
AAAAGATAGATTGAATGGAACGCAAAAACACAATCAGCGGGAGTTTCAAAAAAGCAAAAATGAAAAGTATTATCGGGAGCGGACGACTTA

                 (SD)          .                3650
AATTATGATCTAGT GAACGAGAAACCGATTAGCTATTCAAAAGCGATTCATGAAAAAATTGAGGGGCGAGTCAAACGGAAGGTCCGAGCG
                                                MetLysLysLeuArgGlyGluSerAsnGlyArgSerGluAr
      3700                                                3750
GATGCATGTTTTGGTCAGCGAATTTTTGATCACGGCAAGTCCTGACTATATGAATGGGCTGAGCGATGAGGAGCAGCGGCGCTATTTTGA
gMetHisValLeuValSerGluPheLeuIleThrAlaSerProAspTyrMetAsnGlyLeuSerAspGluGluGlnArgArgTyrPheGl
```

```
                3800                                                3850
AACAGCGGTTGATCATTTGAAAGAGAAATACAGCGCTGAAAACATGCTTTATGCTACAGTCCATATGGATGAAGCGACTCCTCATATGCA
uThrAlaValAspHisLeuLysGluLysTyrSerAlaGluAsnMetLeuTyrAlaThrValHisMetAspGluAlaThrProHisMetHi
                                3900                                                3950
TGTTGGTATTGTACCGATCACAGAGGACGGCCGACTCTCTGCGAAAGATTTTTTTAATGGCAAATTGAAGATGAAAGCCATTCAAGATGA
sValGlyIleValProIleThrGluAspGlyArgLeuSerAlaLysAspPhePheAsnGlyLysLeuLysMetLysAlaIleGlnAspAs
                                        4000                                                4050
TTTTCATCGGCACATGGTTGAAAACGGTTTTGACCTGGTGCGCGGCGAACCAAGCGAAAAGAAGCATGAGAATGTTCACCAGTATAAAAT
pPheHisArgHisMetValGluAsnGlyPheAspLeuValArgGlyGluProSerGluLysLysHisGluAsnValHisGlnTyrLysIl
                                                4100
AAATCAGCGGGAACCGGAGCTTGAGCGGCTTAATGCTGAAATTGCTTTAAAGGAAAAGCAGAGAGAGGAACTGGAAAAGCAAAACAAAGC
eAsnGlnArgGluProGluLeuGluArgLeuAsnAlaGluIleAlaLeuLysGluLysGlnArgGluGluLeuGluLysGlnAsnLysAl
      4150                                                4200
TGTTCAAGCAGTTATAGAAGTGAAAAAAGAATCGCTGACAGCTAAGGCTGAAGAGTTGAAAATGCCGACTATTGAACATGAAAAAGCGTG
aValGlnAlaValIleGluValLysLysGluSerLeuThrAlaLysAlaGluGluLeuLysMetProThrIleGluHisGluLysAlaTr
                4250                                                4300
GCTCAAAAAGGATAAAGTCATTGTGCCAGAGCGGGAACTCCATGCTTTGTATGCCTATGCGGAGCAGAAAACTAAAACGGCAGCCGAGTT
pLeuLysLysAspLysValIleValProGluArgGluLeuHisAlaLeuTyrAlaTyrAlaGluGlnLysThrLysThrAlaAlaGluLe
                                4350                                                4400
GGCGGGGCAATTGAAGTCGGAAACGCAGGAAAAGGAGCGCTGGCAGTCTATCGCCCGGCAGAAGCAGATCGGGCGGATGAAAAAGACCAA
uAlaGlyGlnLeuLysSerGluThrGlnGluLysGluArgTrpGlnSerIleAlaArgGlnLysGlnIleGlyArgMetLysLysThrAs
                                        4450                                                4500
CGGCTTCAGGAACTGCAGAGTAGGATCCATTCAGAAGTTGAAGCGTCCAAAAAGGAAATGCGGCGCAAGCTTGCAAAGGAATTTACGGAA
nGlyPheArgAsnCysArgValGlySerIleGlnLysLeuLysArgProLysArgLysCysGlyAlaSerLeuGlnArgAsnLeuArgLy
                                                4550
GAACAAGCGTCAGGATCTTCGGCAGGAAGTGAAAGAGGAACTGACGACTTTACGAACGGAAAACGAGGAACTGTCAGCTGAAAATAAAGT
sAsnLysArgGlnAspLeuArgGlnGluValLysGluGluLeuThrThrLeuArgThrGluAsnGluGluLeuSerAlaGluAsnLysVa
      4600                                                4650
TTTGATCATTCAAAGAAATAGCGAAGCTGCGGAGAGCCTAAAACTAAAACAGGAACTTGATAAGAGAAACGGGCAGTATGCTGAGGTTTT
lLeuIleIleGlnArgAsnSerGluAlaAlaGluSerLeuLysLeuLysGlnGluLeuAspLysArgAsnGlyGlnTyrAlaGluValLe
                4700                                                4750
GAGTTTCGCCAAGAAGCAGAATCAAACGCTTGAAAAAGTGGCTGGAGAAAACAAGGCGTTAAAAAAAGAAAATAAGACACTAAAAGAGAG
uSerPheAlaLysLysGlnAsnGlnThrLeuGluLysValAlaGlyGluAsnLysAlaLeuLysLysGluAsnLysThrLeuLysGluAr
                                4800                                                4850
AGTTGCCGTACTGGAACAATGGAAAGACAAAATGGTTCAGTGGGCTAAAGAAAAATTACCAAAGATGCGGAAATTAGCGGCATCGTTTTT
gValAlaValLeuGluGlnTrpLysAspLysMetValGlnTrpAlaLysGluLysLeuProLysMetArgLysLeuAlaAlaSerPhePh
                                        4900                                                4950
CGTACGGCTGGAATGCCTAGAGAAGCCAATAAATACAAGGACAATGAATTAGAGCGGTGAAAACGGGCAATCAATTGCCCTTCCAAAATT
eValArgLeuGluCysLeuGluLysProIleAsnThrArgThrMetAsn***
                                                5000
TCCACCACTTTTTTTGTTTGGTGGCTGCGATCATTTTTTGTGTCTCAAGGGACTCACGTAATGCAGCGGTGAGTGTCTCGTGCCTTTCTT

      5050                                                5100
GCTGTCTTTTTTCGAATCGTTCCATCCGTTCTGCCATGTGGCGGTTGAATTCTTCCTGTCGTTTCATGAATTCGACTAAAGGATTGTCCT

                3150                                                3200
GTAGCGATGTAGCGGTATCCGATATGGTCAGTTTAGAACGATATAAGCTTGCTATGTGCTTTACCGTTTCGTCGAGTGAGTGTCCATTGA

                                5250                                                5300
TCTTAGTCATTGTACATAGATACTCTAAAGTCTTTACGTCATCCTCGGTATAGAGTCGCCATCCTTTCGAATCTTTATTGAACGAGTAGC

                                        5350                                                5400
CTTGTTCTTCAAGCATACTGGCATACTTGCGGACAGTTACTGGCTCTATACCGAGGTGTTTTGCGACGTCCTTTGATGATAATTTGATTC

                                                5450
CCATATCCATAACGTATCACCTCGTAAAAAGGGTTCGCTATAGTGAGGGCAAACCCGTTTATTTAGAGCCATTTAAAAAGAAAGCTGTTC

      5500                                                5550
CGTTCTTTTATCAGGTTGGGACTGTATTTAGTTTATGGGGGAGTGCCAATGAGTGATTTTGAAACAGGAATGAGATATGTAAGAGCAACC

                3600                                                3650
CTTGGCTTTGAGGGTTTAGTGCGTACGGAAGAAGAGGAAAAGCTTTTAGAAAGACGGTTTCATGGAGAAATCACAGAGGAAGAATACATA
```

```
                    .         .         5700         .         .         .         .         5750         .
CAAAAAGCGTTTGAGCTTTCTTTGATGTGATGTACCTGCAGGGGCTTATAATATTCATAGCTGTTGTTAGAACCTAAACCCTTTTACCGT

                                                  5800
TTTTGGCGGTGAGGGTTCTTTTTTTTAGGCAGTGATGCGATTTTGCCGTGA
```

ed with a number of protein sequences registered in GenBank with use of the homology search system of GENAS (Kuhara et *al., 1984).* Two cassettes could be identified with staphylococcal plasmids. As shown in Fig. 4, sequences homologous to the pUH1 rep protein was found in protein A (229 amino acid residues) with 27,800 daltons encoded in pC194 (Horinouchi and Weisblum, 1982a), and that of y-glutamyltranspeptidase was in putative protein C403 (403 residues) with 48,400 daltons encoded in pE194 (Horinouchi and Weisblum, 1982a). Approximately, 45.7% amino acid homologies were obser-

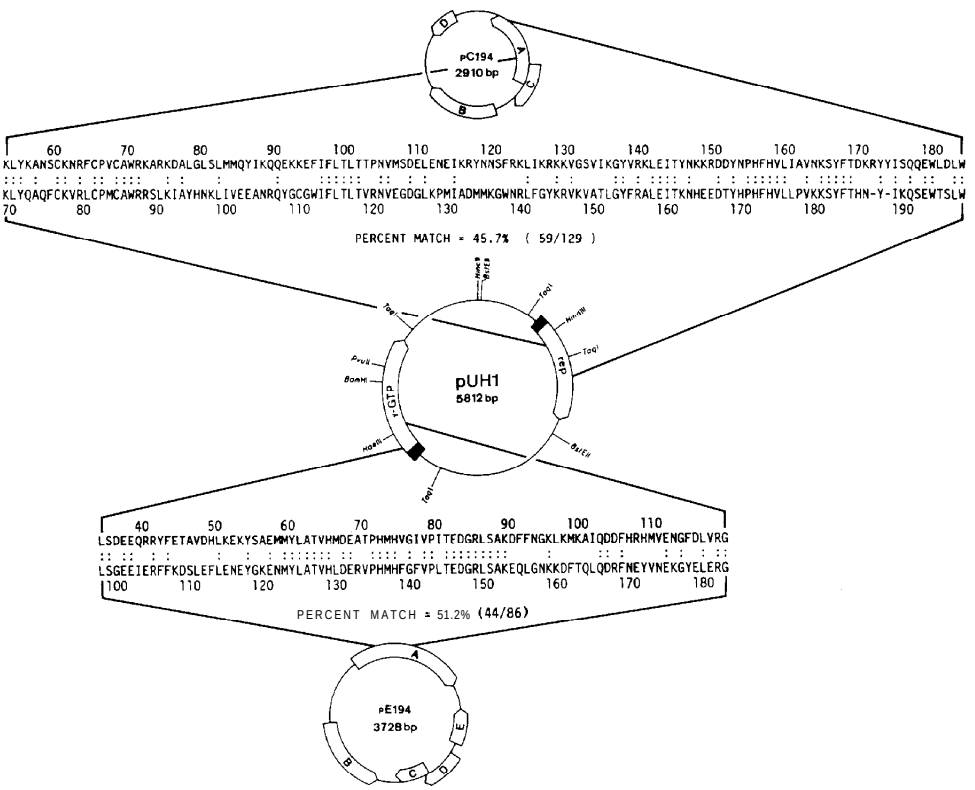

Fig. 4. Schematic comparison of the amino acid sequence deduced from nucleotide sequences of pUH1 and staphylococcal plasmids pC194 and pE194.

Homologous amino acid and deleted sequences are shown by dots and dashes, respectively.

ved in a segment of 129 residues of E229 with pUH1 rep, and 51.2% in 86 residures of C403 with y-glutamyltranspeptidase, but homology scores, in its entirety, were quite low value such as 11.9% with pUH1 rep and 37.4% with y-glutamyltranspeptidase, respectively. The cDNAs of rat renal (Laperche et *al.*, 1986) and human hepatic (Sakamuro et *al.*, 1988), hepatoma (Goodspeed et al., 1989), and placental (Meytset *al.*, 1988) y-glutamyltranspeptidase were cloned, and their nucleotide sequences were determined. The mammalian y-glutamyltranspeptidase, whose amino acid sequences were essentially the same, do not show any similarity at all with pUH1 γ-glutamyltranspeptidase (data not shown). More recently, Suzuki et al. (1989) performed DNA sequencing of *E. coli* y-glutamyltranspeptidase, but no convincing homology could be found between pUH1 and *E. coli* y-glutamyltranspeptidases, for which complete sequence data is available.

REFERENCES

Aumayr, A., T. Hara and S. Ueda 1981 Transformation of *Bacillus subtilis* in polyglutamate production by deoxyribonucleic acid from *B. natto. J. Gen. Appl. Microbiol. 27:* 115-123.

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 Ml3 phage. *EMBO J. 3: 81-86.*

Deininger, P. L. 1983 Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Anal. Biochem.* 129: 216-223.

Devine, K. M., S. T. Hogan, D. G. Higgins and D. J. McConnell 1989 Replication and segregational stability of *Bacillus* plasmid pBAA1. *J. Bacteriol.* 171: 1166~1172.

Fujii, H. 1963 On the formation of mucilage by *Bacillus natto.* Part II. Chemical constituents of mucilage in natto. *Nippon Nogeikagaku Kaishi 37:* 407-411.

Goodspeed, D. C., T. J. Dunn, C. D. Miller and H. C. Pitot 1989 Human y-glutamyltranspeptidase cDNA: comparison of hepatoma and kidney mRNA in the human and rat. Gene 76: 1-9.

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.*

Hara, T., S. Nagatomo, S. Ogata and S. Ueda 1991 Molecular structure of the replication origin of a *Bacillus subtilis (natto)* plasmid, pUH1. *Appl. Environ. Microbiol. 57: 1838-1841.*

Hara, T., J. R. Zhang and S. Ueda 1983 Identification of plasmids linked with polyglutamate production in *Bacillus subtilis (natto). J. Gen. Appl. Microbiol. 29: 345-354.*

Horinouchi, S. and Weisblum B. 1982a Nucleotide sequence and functional map of pE194, a plasmid that specifies inducible resistance to macrolide, lincosamide, and streptogramin type B antibiotics. *J. Bacteriol. 150: 804-814.*

Horinouchi, S. and Weisblum B. 1982b Nucleotide sequence and functinal map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J. Bacteriol. 150: 815-825.*

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.*

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.

Laperche, Y., F. Bulle, T. Aissani, M. -N. Chobert, M. Aggerbeck, J. Hanoune and G. Guellaen 1986 Molecular cloning of rat kidney gammamglutamyltranspeptidase cDNA. *Proc. Natl. Acad. Sci., USA* **83**: *937-941.*

McLaughlin, J. R., C. L. Murray and J. C. Rabinowitz 1981 Uique features in the ribosome binding site sequence of the gram-positive *Staphylococcus aureus* β-lactamase gene. *J. Biol. Chem. 256:* 11283-11291.

Meyts, E. R. -D., N. Heisterkamp and J. Groffen 1988 Cloning and nucleotide sequence of human γ-glutamyltranspeptidase. *Proc. Natl. Acad. Sci., USA 85: 8840-8844.*

Moran, C. P., Jr., N. Lang, S. F. L. LeGrice, G. Len, M. Stephens, 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.*

Muller, R. E., T. Ano, T. Imanaka and S. Aiba 1986 Complete nucleotide sequences of *Bacillus* plasmids pUB110dB, pRBH1 and its copy mutants. *Mol. Gen. Genet.* 202: 169-171.

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.

Sakamuro, D., M. Yamazoe, Y. Matsuda, K. Kangawa, N. Taniguchi, H. Matsuo, H. Yoshikawa and N. Ogasawara 1988 The primary structure of human gamma-glutamyl transpeptidase. Gene. 73: l-9.

Sanger, F., S. Nicklen and A. R. Coulson 1977 DNA sequencing with chainterminating inhibitors. *Proc. Natl. Acad. Sci., USA 74:* 5463-5467.

Shlomai, J. and Kornberg A. 1980 An *E.coli* replication protein that recognizes a unique sequence with a hairpin region of φX174 DNA. *Proc. Natl. Acad. Sci., USA 77: 799-803.*

Suzuki, H. H. Kumagai, T. Echigo and T. Tochikura 1989 DNA sequence of the *Escherichia coli* K-12 y-glutamyltranspeptidase gene, *ggt. J. Bacteriol. 171:* 5169-5172.

Thorne, C. B., C. G. Gomez and R. D. Housewright 1955 Further studies on the biosynthesis of γ-glutamyl peptides by transfer reactions. *J. Biol. Chem. 212:* 427-438.

Villafane, R., D. H. Bechhofer, C. S. Narayanan and D. Dubnau 1987 Replication control genes of plasmid pE194. *J. Bacteriol.* 54: 4822-4829.

Yasuda, S., N. Nakayama, H. Gikuya and E. Soeda 1984 An improved method for shotgun DNA sequencing. *Agric. Biol. Chem. 49:* 1525-1526.