5773

DNA STRUCTURES CONTRIBUTING TO THE INSTABILITY OF RECOMBINANT PLASMIDS IN *STREPTOCOCCUS THERMOPHILUS*

SUMMARY. Bifunctional shuttle vector (pBN183) and recombinant plasmids (pDCO2 and pDCO3) carrying a *Streptomyces* cholesterol oxidase gene (*choA*) were deleted to varying degrees in *Streptococcus thermophilus*. Restriction mapping of the deleted plasmids led to the identification of deletion prone regions in the transforming DNAs. Sequence analysis revealed that direct repeats and hairpin structures occurred in these regions, suggesting that they are deleterious to the stability of plasmids in *S. thermophilus*.

INTRODUCTION

Streptococcus thermophilus is an important microorganism used in the production of yogurt and certain cheeses. Many research efforts have been directed to genetically engineer this organism. The successful development of a genetically engineered microorganism with commercial feasibility requires that the incoming DNA is stably maintained. The structural feature of the newly introduced DNA often affects its stability in the host organism. This is especially the case with DNAs that replicate by the rolling-circle replication mechanism (Gruss and Ehrlich, 1989; Michel *et al.*, 1989). In this communication, we report the restriction mapping of the deletion-prone regions of recombinant plasmids pBN183 and pDCOs in *S. thermophilus*. DNA sequence analysis of these regions suggest that direct and inverted repeats contribute to DNA structural instability in *S. thermophilus*.

MATERIALS AND METHODS

Competent *Escherichia coli* HB101 and DH5α were obtained from BRL Life Technologies, Inc. (Gaithersburg, MD). *Streptococcus thermophilus* strain ST128 was from our laboratory collection. *E. coli* was grown in LBd by the supplier. *S. thermophilus* was electrotransformed as described previously (Somkuti and Steinberg, 1988).

Figure 1 shows the maps of various plasmids used in this study. The pBN183 was constructed by ligating the *Bam*HI-digested pBR322 and a *Bgl*II-linearized lactic strep-

Competent *E. coli* was transformed by heat shock method as prescribed by the supplier. *S. thermophilus* was electrotransformed as described previously (Somkuti and Steinberg, 1988).

Figure 1 shows the maps of various plasmids used in this study. The pBN183 was constructed by ligating the *Bam*HI-digested pBR322 and a *Bgl*II-linearized lactic streptococcal plasmid pNZ18 (a kind gift from G. Simons, Netherlands Institute for Dairy Research). pDBN183 was derived from pBN183 by *in vitro* excision of the small (1.2 kbp) *Sal*I fragment (Solaiman *et al.*, 1992). The *Streptomyces choA* gene (Ishizaki *et al.*, 1989) was retrieved as a 2.7-kbp *Eco*RI fragment from pUCO196 (Solaiman *et al.*, 1992). pDCO2 and pDCO3 were constructed by inserting this *choA*-containing *Eco*RI piece into the *Eco*RI site of pDBN183.

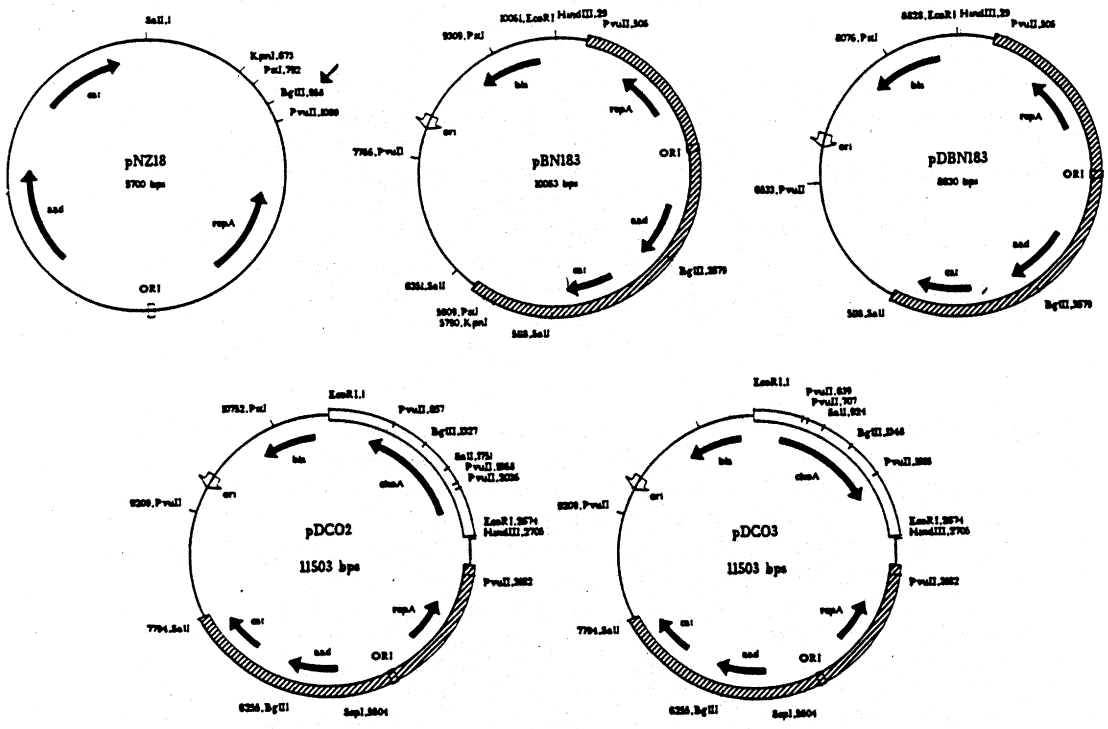

Fig. 1. Plasmid Structures. The *bla, aad* and *cat* genes confer Ap^R, Km^R (kanamycin resistance) and Cm^R, respectively, to the appropriate host. ORI and ori represent the origin of replication from pBR322 and pNZ18, respectively. *choA*, a *Streptomyces* cholesterol oxidase gene. Left-slanted hatched box: sequence originated from pNZ18. Open box: DNA derived from pCO-1 plasmid (Ishizaki *et al.*, 1989). Plasmids were not drawn to size.

Plasmids were isolated from *E. coli* (Ausubel *et al.*, 1987) and *S. thermophilus* (Somkuti and Steinberg, 1986) by established procedures. Restriction endonucleases were obtained from commercial sources and used according to the suppliers' recommendations. Reactions using T4 DNA polymerase and T4 DNA ligase were performed as described by Ausubel *et al.* (1987). Agarose gel electrophoresis was carried out in TBE buffer.

The Pustell DNA homology matrix computer program (IBI-Kodak, New Haven, CT) was used to analyze the DNA sequences of the deletion-prone regions of pBN183, pDCO2 and pDCO3. The nucleotide sequence of pNZ18 was kindly provided by G. Simons (Netherlands Institute for Dairy Research), and that of the *Streptomyces choAP* operon was retrieved from GENBANK nucleic acids data bank (Accession no. M31939 J03356). The free energy change for the formation of secondary structure was estimated as described (Freier *et al.*, 1986).

RESULTS AND DISCUSSION

Isolation of representative deleted plasmid species

When pBN183 was introduced into *S. thermophilus* ST128 by electrotransformation, plasmid preparations obtained from the Cm^R clones consisted of various deleted forms of pBN183. In order to delineate the individual deleted plasmids, we transformed *E. coli* HB101 to Cm^R phenotype using the ST128 plasmid preparations. Subsequent miniplasmid screening of 24 clones allowed us to identify five major deleted-plasmid species (Fig. 2); these were labeled pBN183A (8.7 kbp), pBN183B (7.6 kbp), pBN183C (10 kbp), pBN183D (6 kbp) and pBN183E (4.9 kbp). Reintroduction of these plasmids into *S. thermophilus* and subsequent plasmid characterization of the Cm^R clones showed that the five pBN183 derivatives were structurally stable in ST128 (Fig. 2). Plasmids pBN183B and pBN183C, however, appeared to replicate poorly in *S. thermophilus*, as suggested by the low DNA contents of the plasmid preparations (Fig. 2). Thus, pBN183A with the least deletion and yet still replicating efficiently in ST128 was chosen for detailed analysis in this study.

The recombinant plasmid pDCO2 (11.5 kbp) also experienced various degrees of deletion when introduced into *S. thermophilus*. By employing similar strategy described above, we delineated the pDCO2 deletion derivatives into four major plasmids in *E. coli* DH5α. These plasmids were labeled pdDCO2-16 (8.4 kbp), pdDCO2-18 (8.6 kbp), pdDCO2-19 (9.5 kbp) and pdDCO2-32 (8 kbp). Interestingly, with the pDCO3 recombinant, only one deleted species was recovered using this strategy. This deletion derivative was designated pdDCO3-1 (6.8 kbp).

Mapping of deletion-prone region of the recombinants

Among the five pBN183 deletion progenies, only pBN183A, pBN183D and pBN183E replicated efficiently in *S. themophilus* (Fig. 2). The largest of these 3 plasmids, namely pBN183A, would most likely still retain features (such as *bla*, *cat*, ori and ORI) that were useful in a shuttle vector. Consequently, we chose this plasmid for further study. Another reason for targeting only pBN183A for detailed investigation was that the primary "deletion hot spot" of the parental plasmid pBN183 appeared to have been removed in this derivative, as evidenced by its stability in *S. thermophilus* on subsequent transformation. Restriction analysis of pBN183A showed that the following restrictions sites were no longer present: *Sal*I (at 5.12-kb coordinate of pBN183), *Kpn*I (at 5.79 kb), and *Pst*I (at 5.91 kb). This suggested that a 1.4-kbp sequence had been deleted from a region flanked by coor-

dinates 4.9 and 5.3 kb (Fig. 1) of the parental plasmid pBN183. We subsequently targeted this deletion-susceptible region of pBN183 for sequence analysis.

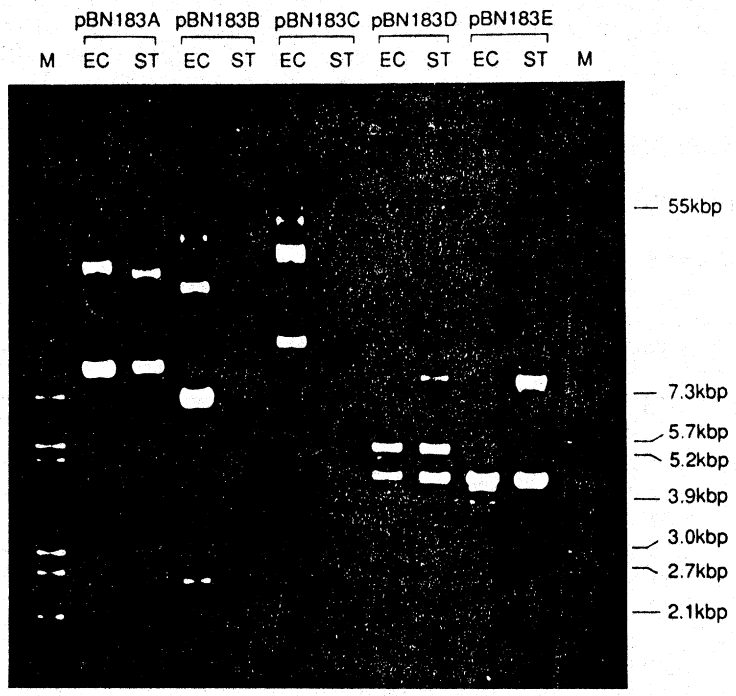

Fig. 2. Gel Electrophoretic Analysis of pBN183 Deletion Derivatives. Electrophoresis was performed on a 0.7% (w/v) agarose gel in TBE buffer. M: *E. coli* V517 plasmids. EC and ST: *E. coli* and *S. thermophilus*, respectively, as the source of the plasmid.

Results of the restriction analysis of the largest deletion derivative of pDCO2, i.e. pdDCO2-19 (9.5 kbp), indicated that a 2-kbp DNA sequence had been deleted from a region bordered by the 2- and 4.7-kb coordinates (Fig. 1) of pDCO2. Restriction sites for *Sal*I (at 1.75-kb coordinate of pDCO2) and *Pvu*II (at 1.97- and 2.04-kb) were still present in pdDCO2-19, while those for *Eco*RI (at 2.67-kb) and *Pvu*II (at 3.18-kb) were now missing. The second largest deletion derivative pdDCO2-18 (8.6 kbp) still retained the *Sal*I site (at 1.75-kb coordinate), but had lost the two *Pvu*II sites (at 1.97- and 2.04-kb). In the third deletion plasmid pdDCO2-16 (8.4 kbp), a 3.1-kbp fragment that contained the *Sal*I site at 1.75-kb as well as those restriction sites absent in pdDCO2-18, was now missing. The smallest deletion species pdDCO2-32 (8 kbp) had lost a 3.5-kbp piece spanning 1.3- to 4.8-kb coordinates that included the *Bgl*II (at 1.33-kb) and *Pvu*II (at 3.2-kb) sites. These results seemed to indicate that the primary "deletion hot spot" was located in the 2- to 4.7-kb coordinates of the parent pDCO2. From this primary site, progressively more extensive deletions occurred leading to the formation of a ladder of deletion

derivatives. Consequently, detailed sequence analysis was performed on the region between 2- and 4.7-kb coordinates of the pDCO2 (Fig. 1).

In the case of pdDCO3-1, restriction analysis showed that the 4.7-kb deleted segment was originally located in the region spanning 0.6- and 5.4-kb coordinates of pDCO3. Consequently, *Pvu*II (at 0.64-kb coordinate of pDCO3) and *Ssp*I (at 4.03-kb) sites were no longer present in pdDCO3-1, while the unique *Eco*RI and the *Ssp*I (at 5.6-kb coordinate of pDCO3) sites remained intact in the deletion derivative.

Sequence analysis of deletion-prone region

DNA sequences that contain direct or inverted repeats had been shown to have deleterious effects on plasmid stability (Gruss and Ehrlich, 1989; Michel *et al.*, 1989). Accordingly, we analyzed the deletion-prone regions of pBN183 and pDCO2 to determine the possible occurrence of these sequence features.

DNA homology matrix analysis of pBN183 sequence between nt #'s 1 and 8000 (the upper limit of the program's data capacity) revealed the occurrence of tandemly repeated sequences between 5- and 7-kb coordinates (Fig. 1). One of the repeating unit started at ca. 5.1-kb coordinate and extended for 0.6-kbp long in a clockwise direction; the other unit (also 0.6 kbp) started at ca. 6.4-kb and proceeded in the same direction. These results appeared to uphold the notion that direct repeats promoted plasmid rearrangement. In a separate study described elsewhere (Solaiman *et al.*, 1992), we demonstrated that the *in vitro* excision of the 1.2-kbp *Sal*I fragment and, consequently, one of the two repeating sequences from pBN183, yielded a stable pDBN183 vector in *S. thermophilus*, further supporting the contention that direct repeats were detrimental to plasmid stability.

When *Streptomyces choA* gene was subcloned into the stable pDBN183, the recombinant plasmid (pDCO2) was unstable and experienced deletion in *S. thermophilus*. Unlike the case with pBN183, sequence analysis of the deletion-prone region between coordinates 2- and 4.7-kb of pDCO2 did not identify any direct repeats. However, it revealed the possible occurrence of two stem-loop/hairpin structures. The first structure, hairpin A (Fig. 3), occurred between nt #'s 2120 and 2155 and located entirely in the subcloned *choA*-containing fragment. The 15-basepair (bp) stem of this hairpin structure contained only three mismatched base pairs. Estimation of the

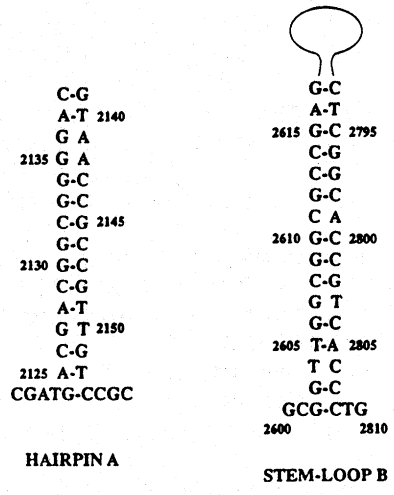

Fig. 3. Stem-Loop/Hair--pin Structures of pDCO2. These DNA 2° structures were identified as des--cribed in the text.

ΔG^{o}(37) for this structure yielded a value of -18 kcal/mol. The other stem-loop feature, structure B (Fig. 3), occurred between nt #'s 2600 and 2810 that spanned both the subcloned DNA fragment and the vector proper. One strand of the 16-bp stem of this structure consisted of nt #'s 2602 - 2617, and the other of nt #'s 2793 -2808. As with the hairpin structure A, the stem region of B contained only 3 mismatched base pairs. The free energy change associated with the formation of this structure was estimated as -11 kcal/mol.

Since the small *Eco*RI segment (nt #'s 1 - 2674) of pDCO3 is oriented opposite to that of pDCO2, the hairpin structure A is situated between nt #'s 500 and 535 in pDCO3. Furthermore, the inverted repeats that constituted the stem-loop structure B in pDCO2 were converted in pDCO3 into direct repeats. One repeating unit (16-bp long) occurred between nt #'s 65 and 80; the other unit remained at nt #'s 2793 - 2808. The locations of these features would predict a deletion region that spanned at least from nt # 65 (where the first repeating unit was encountered) to nt # 2808 (at which point the second homologous sequence ended). This was indeed observed in pdDCO3-1.

In summary, the results of this study identify direct repeats and stem-loop structures in the deletion-susceptible loci of pBN183, pDCO2 and pDCO3. These features apparently contribute to the instability of several recombinant plasmids in *S. thermophilus*. Future plasmid constructs can thus be analysed for the presence of these structures, and appropriate measures such as deletion or site-directed mutagenesis can be taken to remove them. This approach should benefit the development of stable recombinant DNAs suitable for the genetic engineering of this important industrial microbe.

REFERENCES

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1987). Current Protocols in Molecular Biology. John Wiley & Sons, New York, NY.

Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Nielson, T., and Turner, D.H. (1986). *Proc. Natl. Acad. Sci. USA*. 83, 9373-9377.

Gruss, A., and Ehrlich, S.D. (1989). *Microbiol. Rev*. 53, 231-241.

Ishizaki, T., Hirayama, N., Shinkawa, H., Nimi, O., and Murooka, Y. (1989). *J. Bacteriol*. 171, 596-601.

Michel, B., D'Allancon, E., and Ehrlich, S.D. (1989). *J. Bacteriol*. 171, 1846-1853.

Solaiman, D.K.Y., Somkuti, G.A., and Steinberg, D.H. (1992). *Plasmid* (In Press).

Somkuti, G.A., and Steinberg, D.H. (1986). *J. Biotechnol*. 3, 323-332.

Somkuti, G.A., and Steinberg, D.H. (1988). *Biochimie* 70, 579-585.