APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Aug. 1992, p. 2674–2678
0099-2240/92/082674-05$02.00/0

Vol. 58, No. 8

Use of Degenerate Primers for Polymerase Chain Reaction Cloning and Sequencing of the *Lactococcus lactis* subsp. *lactis recA* Gene

PATRICK DUWAT,* S. DUSKO EHRLICH, AND ALEXANDRA GRUSS

Laboratoire de Génétique Microbienne, Institut de Biotechnologie, Institut National de la Recherche Agronomique, Domaine de Vilvert, 78352 Jouy en Josas Cedex, France

Received 13 December 1991/Accepted 5 May 1992

Two particularly well-conserved stretches in the RecA protein sequences were chosen as templates to synthesize degenerate oligonucleotides, which were used in polymerase chain reaction to amplify an internal *recA* DNA fragment of *Lactococcus lactis* subsp. *lactis* ML3. Using this fragment, we recovered and sequenced the entire lactococcal *recA* gene. The end of an open reading frame present upstream of the *recA* gene shows strong homology with formamidopyrimidine-DNA-glycosylase, a protein involved in DNA repair.

Lactococcus lactis subsp. *lactis* is a gram-positive dairy microorganism with a G+C content of about 37% (18) and is used extensively in the food industry. Recent advances in lactococcal genetics allow the development of strains with superior fermentation properties. The genetic stability and good growth of recombinant strains are important parameters in strain development. To characterize the recombination functions important for genetic stability in *L. lactis*, we considered other organisms in which genes involved in genetic stability are known. In *Escherichia coli*, the key element in homologous recombination, as well as SOS response, DNA repair, mutagenesis, and bacteriophage induction, is the *recA* gene (see reference 25 for a review). More than 40 *recA* analogs have been cloned from gram-negative bacteria, using in nearly all cases complementation of an *E. coli recA* mutant as the means of selection (see references 24 and 25 for reviews). In contrast, only three *recA* genes of gram-positive bacteria (11, 22, 24) have been identified on the molecular level. The underrepresentation of cloned *recA* genes from the gram-positive group may be due to difficulties in gene expression; attempts at complementation of *E. coli recA* mutants by their gram-positive counterpart have failed due to either instability of the recombinant clone (30) or the lack of gene expression (11).

Information about recombination and DNA repair in lactic bacteria is still preliminary. A recombination-deficient *L. lactis* strain showing partial sensitivity to UV and reduced recombination capacity was isolated (1); however, the affected gene(s) has not been identified.

We present the cloning of the *recA* gene of *L. lactis* (called $recA_L$). The presence of a RecA protein in *L. lactis* was confirmed by cross-reaction with antibodies to *E. coli* RecA (2, 12a). Our initial strategy, to clone the lactococcal *recA* gene by complementation of an *E. coli* or *B. subtilis rec* mutant, proved unsuccessful; further attempts to identify the gene by hybridization, using a DNA fragment containing part of the *B. subtilis recA* gene as the probe, gave an ambiguous mapping of the gene (12a). We therefore employed polymerase chain reaction (PCR) (26) to isolate DNA fragments internal to the *recA* genes of lactic bacteria. Although *recA* DNA sequences are not well conserved between species, the protein sequences do have highly conserved domains. Degenerate primers were designed by using the conserved amino acid stretches of known RecA proteins as templates. The PCR amplification of a *recA* internal fragment allowed us to recover the entire *recA* gene of *L. lactis* subsp. *lactis* ML3. DNA sequences of internal fragments of the *recA* genes of *L. lactis* subsp. *lactis* IL1403 and *L. lactis* subsp. *cremoris* IL736 were also determined and compared.

Strategy for cloning an internal fragment of the $recA_L$ gene. Amino acid sequences of RecA proteins, extrapolated from more than 20 known DNA sequences, are highly conserved (56 to 98% homology) (24). Two particularly well-conserved domains, approximately 120 amino acids apart, were selected as the basis for degenerate primer synthesis. The DNA sequence of the coding strand primer is 5′-GA(A,G) CA(C,T)GCNCTNGA(C,T)CC-3′ (128-fold degeneracy); it is based on a RecA protein consensus sequence, 96-Glu-His-Ala-Leu-Asp-Pro-101 (numbering corresponds to the *E. coli* RecA sequence). The DNA sequence of the complementary strand primer, 5′-CC(A,T)CC(A,T)G(G,T)(A,T)GT(A,C,T) GT(C,T)TCNGG-3′ (384-fold degeneracy) corresponds to consensus sequence 211-Glu-Gly-(Thr,Pro)-Thr-Thr-Glu-Pro-206. Codon preference was not considered in the primer design, as DNA sequences of previously cloned lactococcal genes of *L. lactis* showed no clear codon usage bias (DNA sequences data of genes from the histidine, tryptophan, and leucine operons [4, 11a, 13]). We took into account the G+C content (37%) of *L. lactis* DNA in designing the first eight nucleotides (5′ end), but not the remainder of the primer, since perfect homology on the 3′ end is desirable for good elongation. The primer couple was used in PCR to amplify the internal DNA fragment of the lactococcal *recA* gene (called $recA_{L\text{-}i}$) from total DNA of *L. lactis* subsp. *lactis* ML3 (prepared as described in reference 15). The reaction mixtures (prepared by using Perkin-Elmer Cetus instructions) were supplemented with 1.25% formamide (28) (good amplification was contingent upon the presence of formamide) and were subjected to 30 cycles consisting of a 1-min denaturation period at 94°C, a 2-min annealing period at 55°C, and a 3-min extension period at 72°C. The amplified DNA fragment of expected size (about 360 bp) was purified and cloned into the unique *Hin*cII site of the pBluescript KS+ vector (Stratagene), using *E. coli* host strain TG1 (21),

* Corresponding author.

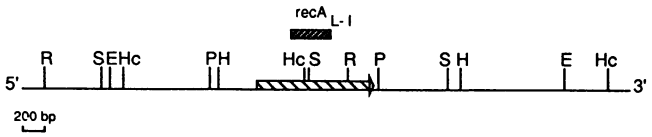

FIG. 1. Physical map of the $recA_L$ locus of *L. lactis* ML3. The map was constructed by hybridization analysis of restricted total chromosomal DNA, using the PCR-amplified DNA fragment $recA_{L\text{-}i}$ (location shown above) as the probe. The arrow corresponds to the direction of translation of the RecA ORF as subsequently derived from the DNA sequence. Restriction endonucleases used are as follows: E, *Eco*RI; H, *Hin*dIII; Hc, *Hin*cII; P, *Pvu*II; R, *Eco*RV; S, *Sty*I.

and sequenced from the plasmid on both strands. An uninterrupted open reading frame (ORF) was detected in the DNA sequence. Comparison of the DNA and amino acid sequences of the ORF with the corresponding sequences of *B. subtilis* and *E. coli* reveals 69 and 60% homology at the DNA level and 79 and 69% homology at the amino acid level, respectively. These results indicate that PCR was successfully used to isolate an internal segment of a lactococcal *recA* gene.

Cloning and sequencing of the entire *recA* gene. Chromosomal DNA of strain ML3 was digested with several restriction endonucleases, separated by agarose gel electrophoresis, and hybridized under stringent conditions (21), using the cloned $recA_{L\text{-}i}$ DNA fragment as the probe. A fine restriction map of the ML3 chromosome in the vicinity of the $recA_L$ gene was thus obtained (Fig. 1). These results were utilized to clone DNA segments encompassing the entire $recA_L$ gene. Considering the size of known *recA* genes (about 1.3 kb) and the localization of the $recA_{L\text{-}i}$ DNA fragment around the *Hin*cII site, two *Pvu*II-*Hin*cII fragments, each approximately 0.8 kb in length, were expected to contain all of $recA_L$. To recover these fragments, chromosomal DNA was digested by *Pvu*II and *Hin*cII, and segments of 0.6 to 0.9 kb were purified and cloned into *Sma*I-linearized pBluescript KS+. Amp-resistant transformants were submitted to colony hybridization (21), using the $recA_{L\text{-}i}$ DNA fragment as the probe, to identify the clones containing the $recA_L$ DNA. Plasmid DNA from positive clones was extracted, analyzed by restriction enzyme digestion, and sequenced on both strands. The data were assembled to obtain the entire sequence of the $recA_L$ gene (Fig. 2). An ORF of 365 amino acids, with a predicted molecular weight of 38,900 is encoded within the sequenced region. No sequence errors were acquired from PCR, since the DNA sequences determined from the PCR-amplified $recA_{L\text{-}i}$ fragment and from the cloned chromosomal *recA* fragments are identical.

Sequence analysis of the *recA* gene. The deduced amino acid sequence of lactococcal $RecA_L$ was compared with RecA sequences from *B. subtilis* and *E. coli* (Fig. 3). Overall, $RecA_L$ is 61% homologous to *B. subtilis* RecA and 56% homologous to *E. coli* RecA; these homologies increase to 73 and 67%, respectively, when allowances are made for conservative amino acid substitutions. The homology is reduced to about 35% in the 50-amino acid carboxyl end of the protein.

DNA sequence analysis revealed signals corresponding to transcriptional and translational start and stop sites. Two regions, −170 and −331 bp upstream of the RecA ORF, showed the best match with the gram-positive consensus promoter TTgaca 16-18N TAtaAT, with nucleotides in capital letters being the most conserved (14) (Fig. 2). In the same region, we found three sequences (Fig. 2) comparable to a consensus sequence identified in various DNA damage-inducible promoters in *B. subtilis*, GAAC-N4-GTTC (7). A ribosome binding site, GAGGA, is present 6 bp upstream of the start codon TTG of the $recA_L$ ORF. An 8-bp hairpin followed by 8 A's was found in the 48 bp after the stop codon TAA of the $recA_L$ ORF (Fig. 2). The DNA sequences of the coding strand just flanking the $recA_L$ coding region are remarkably A rich (52%).

We found the end of an ORF encoded 240 bp upstream of the *recA* gene. Sequence comparison (using the GenBank data base) showed strong homology with the last 80 amino acids of formamidopyrimidine-DNA-glycosylase (Fpg), an enzyme involved in DNA repair (6), found in both *E. coli* and *B. firmus*.

Extension of the PCR method to other *L. lactis* strains. We used the same degenerate primers and conditions to amplify the DNA segment internal to the *recA* gene of two other lactococcal strains, *L. lactis* subsp. *lactis* IL1403 (plasmid-free R^-/M^-) (8) and *L. lactis* subsp. *cremoris* IL736 (AM2 from CNRZ collection number Z380). For each strain, an amplified DNA fragment of approximately 360 bp was cloned and sequenced from two independent clones. Sequence comparison with the corresponding region of the *recA* gene of *L. lactis* subsp. *lactis* ML3 (Fig. 2) showed a very high degree of conservation among the three strains. The DNA sequence differences are translationally silent, as amino acid sequences of the three RecA proteins are identical in this region.

Applications of the system. RecA proteins are closely related, but *recA* gene coding and flanking sequences are divergent, particularly between gram-positive and -negative bacteria. Degenerate primers designed from the most conserved RecA domains, as determined from mostly gram-negative sequences, effectively amplified a lactococcal *recA* fragment by PCR. Since these primers picked up numerous gram-positive *recA* gene fragments (13), we propose this strategy as the most reliable for *recA* gene cloning from gram-positive bacteria.

The isolation of *recA* mutants would be useful in the development of genetic systems in the lactic bacteria. We would expect a *rec* mutant to be defective in homologous recombination, SOS induction (DNA repair), and bacteriophage induction, as seen in the numerous gram-negative and the one gram-positive strain thus far examined. Although direct proof is still being sought, the present information supports the prediction that lactococcal RecA is functionally analogous to other RecA proteins; it is known that *L. lactis* undergoes homologous recombination (9), that lactococcal *recA* mRNA is increased in the presence of SOS inducers (12a), and that conditions which induce SOS functions also induce lactic bacteriophages (16, 23). Through the use of conditional suicide vectors (20), the inactivation of the lactic *recA* gene, which is under way, will allow us to confirm these functions.

In addition, two of the principal problems in lactic fermentation processes can be addressed by the manipulation of the *recA* gene. Despite the economic importance of the cheese industry, there is no cure for bacteriophage contamination; furthermore, strains used for fermentation are poorly defined and have unstable properties. Specific $RecA_L$ engineering, using the high conservation of the RecA protein structure (29) as a guide, can address both problems. Modifications of RecA which specifically inhibit bacteriophage induction are feasible, as the region involved in phage repressor cleavage

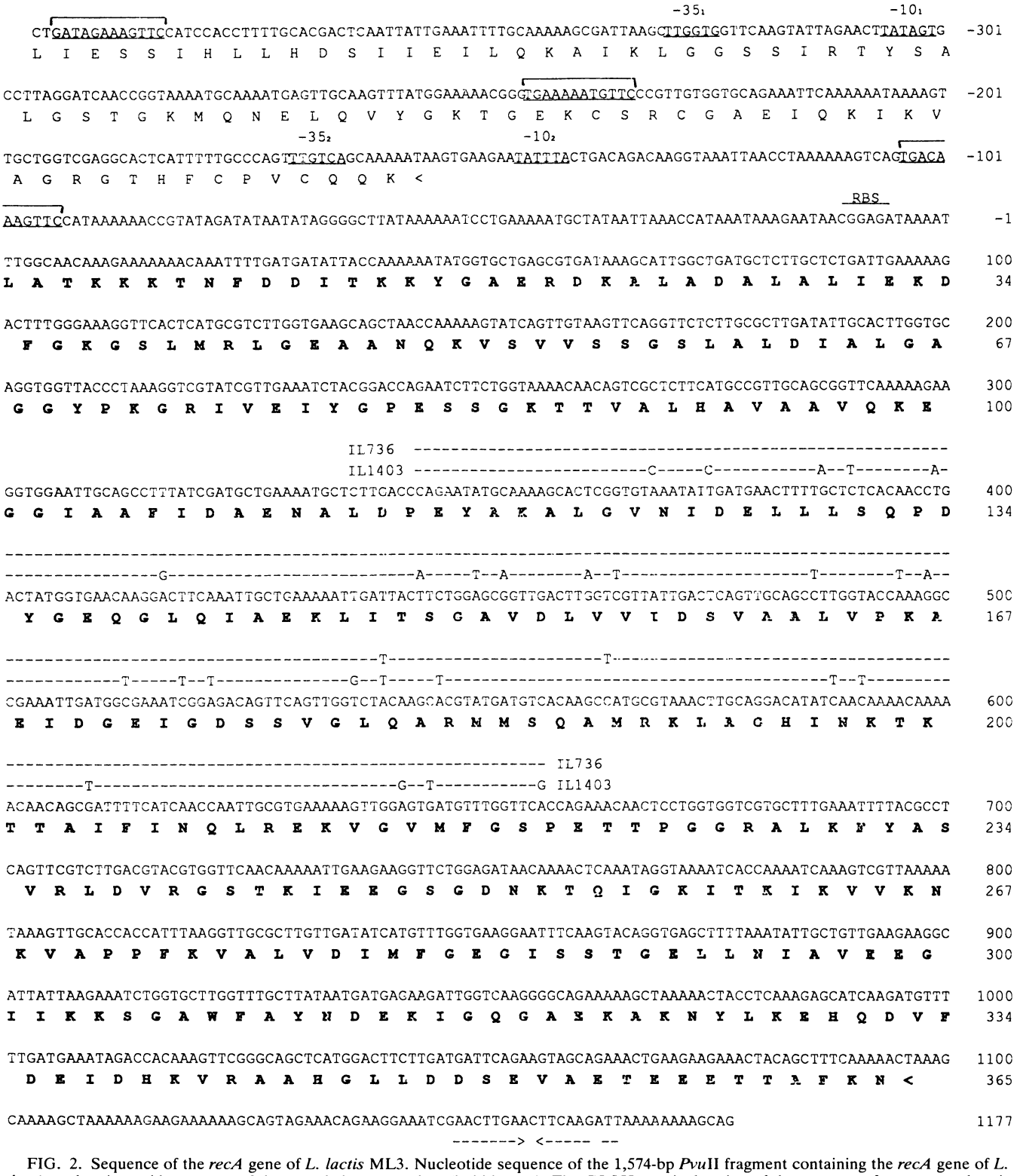

FIG. 2. Sequence of the *recA* gene of *L. lactis* ML3. Nucleotide sequence of the 1,574-bp *Pvu*II fragment containing the *recA* gene of *L. lactis* and amino acid sequence of the encoded gene product (bold letters). The COOH-terminal region of the upstream *fpg* gene product is also shown. Position 1 corresponds to the first nucleotide coding for the RecA protein. Two possible promoters are underlined (positions −301 and −170). A probable ribosomal binding site (RBS) is overlined (position −6). Three potential regions comparable to the *B. subtilis* regulator of damage-inducible genes are bracketed. A palindromic sequence, indicated by arrows, followed by eight A's, may serve as a transcriptional terminator. DNA sequences of the PCR-amplified $rec_{L\text{-}i}$ fragments of strains IL1403 (*L. lactis* subsp. *lactis*) and IL736 (*L. lactis* (subsp. *cremoris*) are given above the ML3 sequence. Identical residues are depicted as dashes, and nucleotide substitutions are indicated.

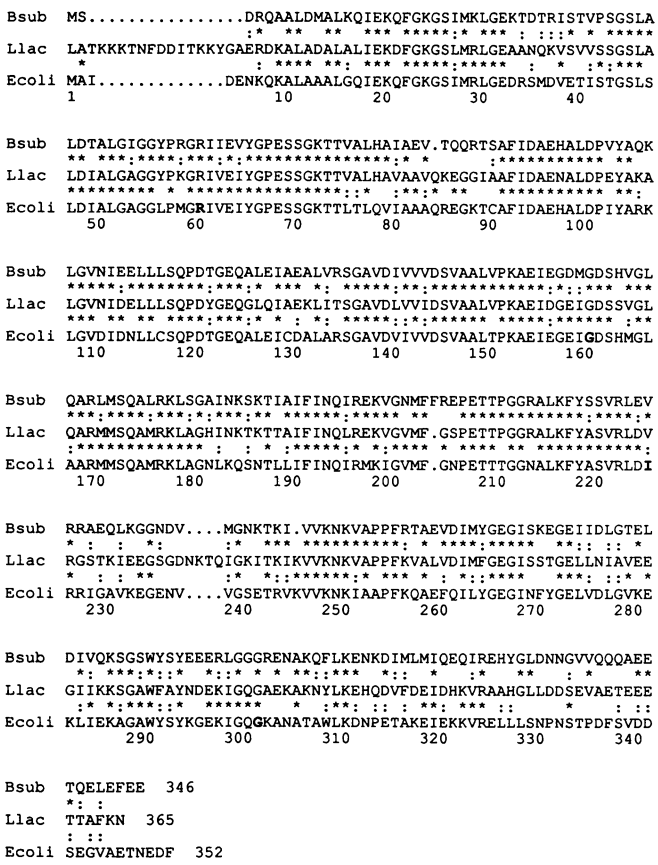

FIG. 3. Comparison of the deduced amino acid sequences of the *recA* genes of *B. subtilis* (Bsub), *L. lactis* (Llac), and *E. coli* (Ecoli). Amino acid positions correspond to the *E. coli* RecA protein. A star means that amino acids are identical, and a colon means that the amino acids are conserved. The four amino acids in bold letters have been studied functionally in *E. coli* (see Discussion).

has been localized on *E. coli* RecA (24); notably, certain RecA proteins can complement homologous recombination defects without catalyzing cleavage of the phage repressor (25). Specific mutations which reduce homologous recombination can also be chosen by comparing the lactococcal and *E. coli* RecA proteins. For example, in *E. coli*, RecA proteins altered at Arg-60, Gly-160, or Gly-301 are unable to promote recombination or SOS induction (24). As these amino acid residues are conserved in $RecA_L$, alterations at these positions are expected to have identical effects.

The gram-negative and -positive RecA proteins contain distinct structural differences. The *recA142* mutant of *E. coli* (12) is recombination and SOS defective and only partially cleaves the bacteriophage λ repressor; it contains a single point mutation resulting in the replacement of Ile-225 by a valine residue. This Ile is conserved in all sequenced gram-negative RecA homologs, but is replaced by Val in the four sequenced gram-positive RecA proteins, suggesting a structural compensation in the gram-positive RecA protein.

The $recA_{L-i}$ sequences of three other lactococcal strains (Fig. 2) show that the *L. lactis* subsp. *lactis* ML3 DNA fragment is closer to that of *L. lactis* subsp. *cremoris* (98%) than to that of *L. lactis* subsp. *lactis* IL1403 (92%). These results are consistent with reports that ML3 is an *L. lactis* subsp. *cremoris* strain rather than an *L. lactis* subsp. *lactis* strain (17, 27).

The upstream region adjacent to the $recA_L$ gene contains the end of a potential ORF. The 80-amino-acid sequence shares strong homology with Fpg, found in *E. coli* and *B. firmus* (5, 6). In *E. coli*, this 269-amino-acid protein is involved in DNA repair, particularly upon damage induced by oxidative stress, and seems not to be regulated by the SOS pathway in *E. coli* (10). The *fpg* gene and $recA_L$ are clustered in *L. lactis*, whereas they are unlinked in *E. coli* (5). The proximity of these genes in *L. lactis* may indicate overlapping regulation, particularly since the putative *recA* promoters lie within the *fpg* ORF (Fig. 2). This organization suggests a possible control circuit linking recombination and repair. A more condensed organization of the lactococcal genes may also be a reflection of its compact genome (2,500 kb compared to 4,750 kb for *E. coli*) (3, 19).

Nucleotide sequence accession number. The nucleotide sequence data reported in this paper will appear in the GenBank Nucleotide Sequence Database under accession number M88106.

We are grateful to P. Dabert, E. Maguin, and P. Langella for constant discussion of this work. The degenerate single-stranded DNA oligonucleotide primers used for PCR were synthesized by Appligene and by J. Buri (this institute). We thank L. McKay for providing us with strain ML3 and the M. C. and A. Chopin laboratory for providing strains and sequence data prior to publication.

Our research was supported in part by a grant from the European Economic Community, BRIDGE grant CT91-0263.

REFERENCES

1. **Anderson, D. G., and L. L. McKay.** 1983. Isolation of a recombination-deficient mutant of *Streptococcus lactis* ML3. J. Bacteriol. **155:**930–932.
2. **Auffray, Y., B. Thammavongs, and P. Boutibonnes.** 1991. Identification of a RecA-like protein in *Lactococcus lactis*. Biochimie **73:**231–233.
3. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54:**130–197.
4. **Bardowski, J., S. D. Ehrlich, and A. Chopin.** Submitted for publication.
5. **Boiteux, S., and O. Huisman.** 1989. Isolation of a formamidopyrimidine-DNA-glycosylase (*fpg*) mutant of *Escherichia coli* K12. Mol. Gen. Genet. **215:**300–305.
6. **Boiteux, S., T. R. O'Connor, and J. Laval.** 1987. Formamidopyrimidine-DNA-glycosylase of *Escherichia coli*: cloning and sequencing of the *fpg* structural gene and overproduction of the protein. EMBO J. **6:**3177–3183.
7. **Cheo, D. L., K. W. Bayles, and R. E. Yasbin.** 1991. Cloning and characterization of DNA damage-inducible promoter regions from *Bacillus subtilis*. J. Bacteriol. **173:**1696–1703.
8. **Chopin, A., M. C. Chopin, A. Moillo-Batt, and P. Langella.** 1984. Two plasmid-directed restriction and modification systems in *Streptococcus lactis*. Plasmid **11:**260–263.
9. **Chopin, M. C., A. Chopin, A. Rouault, and N. Galleron.** 1989. Insertion and amplification of foreign genes in the *Lactococcus lactis* subsp. *lactis* chromosome. J. Bacteriol. **55:**1769–1774.
10. **Czeczot, H., B. Tudek, B. Lambert, J. Laval, and S. Boiteux.** 1991. *Escherichia coli* Fpg protein and UvrABC endonuclease repair DNA damage induced by methylene blue plus visible light in vivo and in vitro. J. Bacteriol. **173:**3419–3424.
11. **Davis, E. O., S. G. Sedwick, and J. M. Colston.** 1991. Novel structure of the *recA* locus of *Mycobacterium tuberculosis* implies processing of the gene product. J. Bacteriol. **173:**5653–5662.

11a. **Delorme, C., S. D. Ehrlich, and P. Renault.** Submitted for publication.

12. **Dutreix, M., P. L. Moreau, A. Bailone, F. Galibert, J. R. Battista, G. C. Walker, and R. Devoret.** 1989. New *recA*

mutations that dissociate the various RecA protein activities in *Escherichia coli* provide evidence for an additional role for the RecA protein in UV mutagenesis. J. Bacteriol. **171:**2415–2423.

12a. **Duwat, P.** Unpublished data.

13. **Duwat, P., S. D. Ehrlich, and A. Gruss.** A general method for cloning *recA* genes of gram-positive bacteria by polymerase chain reaction. J. Bacteriol., in press.

14. **Graves, M. C., and J. C. Rabinowitz.** 1986. In vivo and in vitro transcription of the *Clostridium pasteurianum* ferredoxin gene. Evidence for extended promoter element in Gram-positive organisms. J. Biol. Chem. **261:**11409–11415.

15. **Gruss, A., and S. D. Ehrlich.** 1988. Insertion of foreign DNA into plasmids from gram-positive bacteria induces formation of high-molecular-weight plasmid multimers. J. Bacteriol. **170:**77–83.

16. **Huggins, A. R., and W. E. Sandine.** 1977. Incidence and properties of temperate bacteriophages induced from lactic streptococci. Appl. Environ. Microbiol. **33:**184–191.

17. **Jarvis, A. W., and B. D. W. Jarvis.** 1981. Deoxyribonucleic acid homology among lactic streptococci. Appl. Environ. Microbiol. **41:**77–83.

18. **Kilpper-Balz, R., G. Fischer, and K. H. Schleifer.** 1982. Nucleic acid hybridization of group N and group D streptococci. Curr. Microbiol. **7:**245–250.

19. **LeBourgeois, P., M. Mata, and P. Ritzenthaler.** 1989. Genome comparison of Lactococcus strains by pulse-field electrophoresis. FEMS Microbiol. Lett. **59:**65–70.

20. **Maguin, E., P. Duwat, T. Hege, S. D. Ehrlich, and A. Gruss.** Submitted for publication.

21. **Maniatis, T. E., E. F. Fritsch, and J. Sambrook.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

22. **Marrero, R., and R. E. Yasbin.** 1988. Cloning of the *Bacillus subtilis recE*$^+$ gene and functional expression of *recE*$^+$ in *B. subtilis*. J. Bacteriol. **170:**335–344.

23. **McKay, L. L., and K. A. Baldwin.** 1973. Induction of prophage in *Streptococcus lactis* C2 by ultraviolet radiation. Appl. Environ. Microbiol. **25:**682–684.

24. **Miller, R. V., and T. A. Kokjohn.** 1990. General microbiology of *recA*: environmental and evolutionary significance. Annu. Rev. Microbiol. **44:**365–394.

25. **Roca, A. I., and M. M. Cox.** 1990. The RecA protein: structure and function. Crit. Rev. Biochem. Mol. Biol. **25:**415–456.

26. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Hullis, and H. A. Ehrlich.** 1988. Primer directed enzymatic amplification of DNA with thermostable DNA polymerase. Science **239:**487–491.

27. **Salama, M., W. Sandine, and S. Giovannoni.** 1991. Development and application of oligonucleotide probes for identification of *Lactococcus lactis* subsp. *cremoris*. Appl. Environ. Microbiol. **57:**1313–1318.

28. **Sarkar, G., S. Kapelner, and S. S. Sommer.** 1990. Formamide can dramatically improve the specificity of PCR. Nucleic Acids Res. **18:**7465.

29. **Story, R. M., and T. A. Thomas.** 1992. Structure of the recA protein-ADP complex. Nature (London) **135:**374–376.

30. **Yasbin, R. E., M. Stranathan, and K. W. Bayles.** 1991. The *rec*E(A)+ gene of *B. subtilis* and its gene product: further characterization of this universal protein. Biochimie **73:**245–250.