Molecular Analysis of the Rolling-Circle Replicating Plasmid pA1 of Lactobacillus plantarum A112

M. VUJIC and L. TOPISIROVIC

Institute of Molecular Genetics and Genetic Engineering, Vojvode Stepe 283, P.O. Box 794, 11001 Belgrade, Yugoslavia

Received 28 May 1992/Accepted 3 November 1992

Lactobacillus plantarum A112 has four different plasmids. Plus-origin-specific probes were used to determine that the smallest, cryptic plasmid, pA1 (2,820 bp), showed homology to the pE194 plasmid family. This subclass of plasmids uses the rolling-circle mode of replication. Subsequent analysis of plasmid pA1 demonstrated that it generates single-stranded DNA intermediates, and sequence analysis revealed that it contains three putative open reading frames (ORFs): ORF1, ORF2, and ORF3, which could encode proteins designated RepA (47 amino acids [aa]) and RepB (196 aa) and a protein of 103 aa, respectively. Two of these proteins, RepA (5.6 kDa) and RepB (26 kDa), were identified in in vitro transcription translation assays. The RepA protein contains a characteristic α-helix-turn-α-helix motif typical of DNA-binding proteins that act as DNA-binding repressors. The RepB protein shows a significant similarity with replication initiation proteins of the pE194 family of plasmids that use the rolling-circle mode of replication. Plasmid pA1 is able to replicate in Escherichia coli and Lactobacillus lactis subsp. lactis as well as in other L. plantarum strains.

Lactobacilli have a great industrial significance, since they are widely used as starter cultures for a variety of fermented products. Many Lactobacillus species have one or more natural resident plasmids of various sizes (8, 24, 26, 35, 36). Analysis of the organization and replication of various small plasmids isolated from gram-positive bacteria, such as pC194, pE194, pT181, and pUB110, has shown that they replicate by the rolling-circle mode of replication (RCR), generating single-stranded DNA (ssDNA) intermediates (4, 30, 31; for a review, see reference 15). All RCR plasmids contain a gene encoding a plasmid replication initiation protein (Rep) and its target site (plus-origin), as well as a specific sequence which enables the conversion of ssDNA intermediates into double-stranded DNA plasmid molecules, known as a minus-origin.

Classification of RCR plasmids was proposed according to plus-origin sequence homologies. For this purpose, staphylococcal ssDNA-generating plasmids pE194, pC194, and pT181 were used as reference plasmids (15). In addition, the Rep proteins within the same classification group were found to have an extensive homology. Much less sequence homology could be found among the minus-origin sites, even within the same group of plasmids. All minus-origins of RCR plasmids are located at DNA fragments ranging in size from 130 to 300 bp and contain imperfect palindromic structures (4, 6, 12, 13).

Only the few small Lactobacillus plasmid replicons have been analyzed at the molecular level so far. The results obtained demonstrate that all of them replicate by RCR. Comparative analysis of plus-origin sequences revealed that plasmids Lactobacillus plantarum CCM1904 plasmid pLP1 (5) and Lactobacillus hilgardii plasmid pLAB1000 (18) belong to the pC194 plasmid family, whereas L. plantarum NCD01088 plasmid pLB4 (1) belongs to the pE194 family of RCR plasmids. On the basis of these three examples, no structural elements specific only to lactobacillus plasmids could be recognized.

In the present study, we used plus-origin-specific probes of RCR plasmids to detect such plasmids in L. plantarum A112. A cryptic plasmid, pA1, was preliminarily identified as an RCR plasmid by this method; subsequently, we determined the complete nucleotide sequence and structural organization of this plasmid (2.8 kb), the smallest of the four plasmids present in L. plantarum A112.

MATERIALS AND METHODS

Bacterial strains, media, and reagents. L. plantarum A112 was obtained from George Sakellaris, Escherichia coli C600 (thr-1 leuB6 supU44 tonA21) from the laboratory collection was used. Lactobacilli were grown in MRS medium (Difco, Detroit, Mich.), and E. coli was grown in Luria broth (Difco). All media were solidified with 1.5% agar (Difco) when used as agar plates. Ampicillin (Sigma Chemie GmbH, Deisenhofen, Germany) was added to a final concentration of 50 μg/ml. Restriction enzymes, T4 DNA ligase, and nuclease S1 were supplied by Gibco-BRL (DIFRO, Vienna, Austria) and were used according to the manufacturer’s instructions.

DNA manipulations and transformation. Plasmid DNA from E. coli was isolated according to the alkaline lysis method described by Birnboim and Doly (3). Plasmids were isolated from lactobacilli according to the alkaline lysis method described above, but with the following modifications: as lysis solution, PP buffer (40 mM Na2HPO4, 1 mM Mg acetate, 0.5 M sucrose [pH 7.0]) with 7 mg of lysozyme per ml was used. Samples were incubated for 15 min at 37°C. TE1-sodium dodecyl sulfate (SDS) solution (100 mM Tris-HCl, 10 mM EDTA [pH 10.5], 1% [wt/vol] SDS) was used instead of NaOH-SDS solution to complete lysis. After the addition of 100 μl of glacial acetic acid and 125 μl of 5 M NaCl, the samples were mixed gently and centrifuged. The supernatant was removed, and the DNA was precipitated with ethanol. For large-scale plasmid preparations, the pellet was resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA...)
N-lauryl-sarcosyl was used. After centrifugation, the pellets were washed with TEN buffer (50 mM Tris-HCl [pH 8], 10 mM EDTA [pH 8], 50 mM NaCl), and resuspended in PP buffer (described above) with 7 mg of lysozyme per ml. After 15 min of incubation at 37°C, N-lauryl-sarcosyl was added to a final concentration of 1%.

Southern hybridization experiments were performed with a biotin-labeled probe by using the Bio-Nick labeling and detection system (GIBCO-BRL) according to the manufacturer's instructions.

Analysis of ssDNA intermediate production. L. plantarum cells were lysed, and the nuclease S1 treatment of plasmid DNA was according to the method described by Rele et al. (32), with the following modifications. After centrifugation, the pellet was washed with TEN buffer (50 mM Tris-HCl [pH 8], 10 mM EDTA [pH 8], 50 mM NaCl). Gels and Western blots were run in a vacuum oven and then UV irradiated (254 nm) for 1 min. Hybridization was carried out in hybridization buffer (0.5 M sodium phosphate, 7% SDS, 1 mM EDTA) at restrictive temperatures calculated for each probe. The sequences of oligonucleotide probes were specific for the plus-origin sites of plasmids pC194 (PC probe), pE194 (PE probe), and pT181 (PT probe), respectively, were taken from Gruss and Ehrlich (15) and purchased from Genosys Biotechnologies, Inc., The Woodlands, Tex. The same plasmids, as well as pGKV210 (34), were used as positive controls, whereas plasmid pH253 (29) was used as a negative control. The probes were labeled with [γ-32P]ATP (Amersham, Buckinghamshire, United Kingdom) by using T4 polynucleotide kinase (Boehringer, Mannheim, Germany) according to the instructions of the manufacturer.

DNA-directed transcription translation. Protein synthesis with purified double-stranded plasmid DNA as a template was analyzed in an in vitro transcription-translation assay with the Proacyrtic DNA-Directed Translation System kit (Amersham). [35S]methionine-labeled proteins were analyzed by SDS-polyacrylamide gel electrophoresis (PAGE) by the method of Laemmli (20). The acrylamide concentration in the running gel was 15% (wt/vol), and in the stacking gel was 6% (wt/vol).

DNA sequencing and sequence analysis. The deoxyribonucleotide chain termination procedure (28) was applied for DNA sequencing. Double-stranded DNA templates were prepared by cloning fragments of plasmid pA1 into plasmids pUC18 and pUC19 (22). Standard 17-mer universal primers were used to sequence the insert in pUC18 and pUC19 in both directions, and Sequenase version 2.0 (U.S. Biochemicals, Lecerne, Switzerland) was used as DNA polymerase in a procedure following the manufacturer's protocol. DNA and protein sequences were analyzed by using the PC/GENE program (Intelligenetics, Mountain View, Calif.) and the amino acid sequences were also compared with those in the SwissProt data base (September 1991 release).

Nucleotide sequence accession number. The sequence of plasmid pA1 is deposited in the EMBL data base library under the accession number Z11717.

RESULTS

Analysis of plasmid origins of replication by DNA hybridization. According to the analysis of two-dimensional agarose gels, L. plantarum A112 has four plasmids, pA1 (2.8 kb), pA2 (3.8 kb), pA3 (approximately 11 kb), and pA4 (approximately 14 kb) (data not shown). In order to study whether L. plantarum A112 contains plasmids which replicate via the RCR mode of replication, purified plasmid DNAs were hybridized in dot blot experiments with oligonucleotide probes specific for plus-origin sequences of plasmid pC194, pE194, and pT181. The two smallest plasmids, pA1 and pA2, showed hybridization signals with the oligonucleotide probe specific for the plus-origin of plasmid pE194 (Fig. 1). The pC194- and pT181-specific probes showed no hybridization with any of the plasmids in L. plantarum A112, whereas they did with the positive controls. None of the three probes hybridized with negative-control plasmid pH253 (29), which is known to have a mode of replication essentially different from that of the RCR plasmid family. Therefore, these results suggest that plasmids pA1 and pA2 belong to the pE194 family of ssDNA-generating plasmids (15). Hybridization experiments also showed that plasmid pA1 had no homology with other plasmids in L. plantarum A112, since a hybridization signal was observed with plasmid pA1 only by using linearized pA1 plasmid DNA as a probe (data not shown). On the basis of all of these results, the smallest plasmid (pA1) was chosen for further analysis.

Detection of ssDNA intermediates. The presence of ssDNA intermediates of pA1 in L. plantarum A112 strongly supports the idea that this plasmid replicates by RCR. Therefore, plasmids were isolated by using neutral lysis, and two agarose gels were loaded with nuclease S1-treated and
FIG. 2. Analysis of ssDNA production. The blots were made from a denatured gel (A) and a non-denatured gel (B). Plasmids, isolated after neutral lysis of L. plantarum A112 cells, were loaded on agarose gels without (lanes 1) and with nuclease S1 treatment for 15 min (lanes 2) or 30 min (lanes 3). Hybridization was carried out with biotin-labeled, EcoRI-linearized pAl DNA as a probe. OC, open circular DNA; CC, covalently closed circular DNA; SS, ssDNA.

nontreated plasmid samples. One of the gels was subjected to denaturing conditions after electrophoresis, and plasmid DNAs were transferred from both gels to membrane filters. In hybridization experiments, pAl DNA was used as a probe. These experiments clearly showed that ssDNA intermediates of pAl could be detected (Fig. 2).

Host range of the pAl replicon. A restriction enzyme map of plasmid pAl was constructed after restriction analysis and subcloning in E. coli (Fig. 3). Additional cloning experiments revealed that the unique EcoRI restriction site could be used for cloning without interfering with plasmid replication functions. In contrast, the HindIII restriction site had to be intact; otherwise, the replicative ability of the plasmid was lost (data not shown). To study a host range of plasmid pAl, the cat gene from staphylococcal plasmid pCI94 was cloned into the EcoRI restriction site. The resulting construct was used for transformation of E. coli C600, Lactobacillus lactis subsp. lactis MG1363, and L. plantarum NCD01193. The chloramphenicol-resistant transformants obtained in each host showed that the pAl replicon is functional at least in chosen hosts. The segregational and structural stability of the pAl replicon was monitored by growing transformed hosts in the medium lacking chloramphenicol for 60 generations. It was found that segregational stability was rather high in L. lactis (91%) and L. plantarum (95%), in contrast to E. coli, in which after 60 generations only 1% of cells retained the plasmid. When judged by agarose gel mobility and restriction pattern, retained plasmids were structurally unaltered, regardless of the host from which they were prepared for analysis.

DNA sequence analysis. In order to investigate the structural organization of plasmid pAl in more detail, the complete DNA sequence of this plasmid was determined (Fig. 4). Plasmid pAl contains 2,820 bp and has a calculated GC content of 34.9%. Three open reading frames (ORFs) were identified within this sequence. ORF1 (located between nucleotides [nt] 969 and 1112) can encode a polypeptide of 47 amino acids and is followed by ORF2 (nt 1188 to 1778), which is able to encode a polypeptide of 196 amino acids. The two ORFs are separated by 75 bp. ORF3 (located between nt 2028 and 2339) can encode a polypeptide of 103 amino acids and is located downstream of the previous two ORFs with the spacing of 249 bp (Fig. 4). All protein-coding regions have an ATG codon as the putative initiation codon and are oriented in the same direction. A potential ribosomal-binding site (Shine-Dalgarno sequence) was identified upstream of each ORF. The Shine-Dalgarno sequences for ORF1 (GGAG; ΔG° = −9.4 kcal [ca. −39.3 kJ/mol]), ORF2 (AGGAG; ΔG° = −11.6 kcal [ca. −10.8 kJ/mol]), and ORF3 (AGGAG; ΔG° = −11.6 kcal [ca. −10.8 kJ/mol]) are located 96, 9, and 7 nt upstream from the initiation codons, respectively. Only one potential promoter sequence, which is located immediately upstream of ORF1, could be identified. The putative promoter preceding ORF1 has a weak −35 region (5′-TCTAGA-3′) and a nearly consensus −10 region (5′-TATTAT-3′) with one base difference compared with the canonical E. coli promoter sequences (Fig. 4). In addition, an AT-rich sequence and the TC dinucleotide preceding a potential −35 region were located at the same positions in which they were found in other promoters of gram-positive bacteria (23).

Identification of pAl-encoded polypeptides. The functionality of the three putative ORFs was tested by in vitro transcription translation assays as described in Materials and Methods. EcoRI-linearized pAl plasmid DNA was used as a template. Only two proteins could be clearly recognized as translational products of this plasmid (Fig. 5). The sizes of these two proteins corresponded well to the predicted sizes of the proteins encoded by ORF1 (5.6 kDa) and ORF2 (26 kDa), now designated RepA and RepB, respectively. However, expression of ORF3 was not detected, at least under these experimental conditions.

Comparative analysis of protein sequences. The proteins encoded by the three putative ORFs of plasmid pAl were screened for homology with other proteins against the SwissProt bank of the EMBL data base. A high degree of similarity was observed between the RepA protein of pAl and Rep proteins encoded by other RCR plasmids, which included the RepA protein of pLB4 (L. plantarum [42.6% similarity]), the RepC protein of pWV01 (L. lactis subsp. cremoris [46.8% similarity]), the RepA protein of pLS1...
VOL. 59, and gram-positive plasmids, of protein regions initiation plasmids and this a ty]) (1, 7, 9, 21) (Fig. 6). 494 amino (1, 7, 10, 21) (Fig. 6).

FIG. 4. DNA sequence of plasmid pA1 and deduced amino acid sequences of the protein products. Putative ribosomal binding sites are boxed, and a transcription regulatory region (putative −35 and −10 sequences) is underlined. DR, direct repeats.

(Streptococcus agalactiae [57.8% similarity]), and the CopE protein of pE194 (Staphylococcus aureus [46.2% similarity]) (1, 7, 10, 21) (Fig. 6). The last two proteins have been shown to be involved in the copy number control of these plasmids, and this suggests a function analogous to that of RepA of pA1.

The RepB protein was found to be similar to the replication initiation proteins of the pE194 plasmid family (the RepA protein of pWV01 [39.8% similarity], the RepB protein of pLB4 [45.9% similarity], the RepB of pLS1 [29.6% similarity], and the RepF protein of pE194 [35.1% similarity]) (1, 7, 9, 21) (Fig. 6).

The putative protein, which could be encoded by ORF3, showed a high level of homology (53.8 to 72.5%) with N-terminal regions of the protein family known as mobilization (Mob)-plasmid recombination (Pre) proteins of various gram-positive plasmids, such as pE194 (S. aureus), pMV158 (S. agalactiae), or pLAB1000 (L. plantarum) (14, 18, 33). However, ORF3 of pA1 has the potential to encode a protein of 103 amino acids only, whereas the Mob-Pre proteins of the plasmids mentioned above have sizes ranging from 361 to 494 amino acids.

Potential secondary structures in plasmid pA1. It has been shown that both plus- and minus-origin regions of RCR plasmids from gram-positive bacteria have the potential to form secondary hairpin structures (12, 16). Five potential hairpin structures (inverted repeats [IR], designated IR I to IR V, could be identified in the sequence of plasmid pA1 (Fig. 4). IR II, located between nt 771 and 804 and with a free energy of −23.4 kcal (ca. −97.9 KJ/mol, contains a 22-bp sequence, 5′-GGGGGGCATACGACACCCCCC-3′, which is almost identical to that of the known plus-origin of plasmid pE194 (30). According to these data, it is most likely that IR II consists of the plus-origin-specific region.

IR IV (from positions 1799 to 1814) is located downstream of ORF2, immediately after the TAA stop codon (Fig. 4). This IR, with a free energy of −14.2 kcal (ca. −59.4 KJ/mol), is highly homologous to the translation terminator sequence present downstream of the gene encoding the replication initiation protein of the lactococcal plasmid pWV01 (21). This homology suggests an analogous function for IR IV of pA1. Upstream of ORF3 is located IR V (from positions 1970 to 1995, with a free energy of −10.2 kcal (ca. −42.6 KJ/mol) (Fig. 4), which has a high level of homology with the consensus RS4 sequence, a sequence found in other plasmids from gram-positive bacteria (18).

DISCUSSION

L. plantarum A112 contains four different plasmids. Plasmid pA1 was identified as a member of the pE194 family of RCR plasmids by hybridization with a plus-origin-specific probe. Plasmids belonging to the E194 family of RCR plasmids have a high degree of similarity in the genetic organization of the region responsible for replication. They have two rep genes organized in tandem in one transcription unit which is under the control of a promoter preceding the first rep gene. One of the genes encodes a transcriptional repressor, and the other encodes a replication initiation protein. In close vicinity to this transcriptional unit, there is a short sequence, 5′-GGGGGGCATACGACACCCCCC-3′, which is almost identical to that of the known plus-origin of plasmid pE194 (30). According to these data, it is most likely that IR II consists of the plus-origin-specific region.

IR IV (from positions 1799 to 1814) is located downstream of ORF2, immediately after the TAA stop codon (Fig. 4). This IR, with a free energy of −14.2 kcal (ca. −59.4 KJ/mol), is highly homologous to the translation terminator sequence present downstream of the gene encoding the replication initiation protein of the lactococcal plasmid pWV01 (21). This homology suggests an analogous function for IR IV of pA1. Upstream of ORF3 is located IR V (from positions 1970 to 1995, with a free energy of −10.2 kcal (ca. −42.6 KJ/mol) (Fig. 4), which has a high level of homology with the consensus RS4 sequence, a sequence found in other plasmids from gram-positive bacteria (18).
unit, upstream of the rep genes, a plus-origin is located (1, 2, 7, 10, 37). Identical genetic organization of a region involved in replication was identified in plasmid pA1 (Fig. 4). The two rep genes were located most likely downstream of the plus-origin, forming one transcriptional unit. The ORF1-encoded RepA protein has substantial similarity (42.6 to 57.8%) with the CopE protein of pE194, the RepA protein of pWV01, the RepA protein of pLB4, and the RepA protein of pLS1 (1, 7, 10, 21). In addition, the putative replication initiation protein of this plasmid, RepB (encoded by ORF2), also showed significant similarity (29.6 to 45.9%) with corresponding proteins of the pE194 family of RCR plasmids (1, 7, 9, 21).

It has been shown that the RepA protein of plasmid pLS1 acts as a transcriptional repressor on the rep gene, which regulates the copy number of this plasmid. The promoter governing the expression of this transcriptional unit contained a target site for the RepA protein (10, 11). The idea of a similar role for RepA of pA1 is strongly supported by the presence of a characteristic α-helix-turn-α-helix motif (data not shown) typical for proteins that act as DNA-binding repressors (25).

It is interesting to point out that all plus-origin sequences of the pE194 plasmid family so far determined have an identical 5′-TACTACGAC-3′ sequence in the loop of the potential hairpin structure (9, 15). The corresponding region of pA1 differs in a single base: 5′-CACTACGAC-3′ (Fig. 4). This unique feature of the pA1 plus-origin site is possibly involved in the specificity of the RepB protein for its target site.

Sequence analysis also revealed the presence of ORF3, which could encode a protein of 103 amino acids (Mob′ protein). This protein has significant similarity (53.8 to 72.5%) with the N-terminal regions of mobilization proteins of ssDNA-generating plasmids. However, no promoter upstream of ORF3 could be identified, which could be the reason why expression of this ORF was not detected by in vitro assays. Furthermore, the ORF3 protein has an unusually small length, because the lengths of the characterized Mob proteins vary between 361 and 494 amino acids (14, 18).

In addition, two direct repeats of 55 bp were identified within a coding region of ORF3 followed by five 5′-TAAAGA-3′ direct repeats. This rather peculiar organization is probably

**FIG. 5.** Identification of proteins encoded by plasmid pA1 by an in vitro transcription-translation assay, separated by SDS-PAGE (15% polyacrylamide). Lanes: 1, molecular mass standards (sizes indicated in kilodaltons); 2, EcoRI-linearized plasmid pA1; 3, control without DNA.

**FIG. 6.** Comparison of the predicted amino acid sequences of RepA (A) and RepB (B) proteins encoded by plasmid pA1 with those encoded by RCR plasmids of other gram-positive bacteria. Replication proteins of plasmids pLB4, pLS1, pWV01, and pE194 have been aligned, and the original names of these proteins have been replaced with REP followed by a letter and the name of the plasmid from which they originate. The length of each protein (in amino acids) is also indicated. Gaps were introduced to maximize the similarity. Identical and similar amino acid residues are boxed.
the result of recombination events, thereby deleting a large part of the putative mob gene of pPA1. Although no functional mob gene seemed to be present in pPA1, an IR homologous to the consensus RSα site, which is the binding site of Mob-Pre protein, is still present in this plasmid (Fig. 4). Recently, a similar organization for the lactococcal plasmid pFX2 has been described. Plasmid pFX2 carries the RSα site, but an ORF which could encode a Mob-Pre protein is missing (37).

No sequence homology with any of the already known minus-origin sites (4, 6, 15, 19) was detected within the sequence of plasmid pPA1. Therefore, determination of functional minus-origin site of plasmid pPA1 awaits further deletion analysis.

In conclusion, a high sequence similarity between the RepA and RepB proteins of plasmid pPA1 and the related proteins of other gram-positive plasmids, together with the presence of a sequence almost identical to the plus-origin sites of the same plasmids, argues in favor of pPA1 replicating by RCR. This conclusion is strongly supported by the presence of detectable ssDNA intermediates in _L. plantarum_ cells.

ACKNOWLEDGMENTS

We are indebted to Kees Leenhouts from the University of Groningen, The Netherlands, for computer analysis of the plasmid pPA1 sequence and for critical reading of the manuscript. This work was supported by RFN grant 1407 and partially by the ICGB Collaborative Research Programme (grant CRP/YUG88-05).

REFERENCES


