

A Type IC Restriction-Modification System in *Lactococcus lactis*

CATHERINE SCHOULER, FLORENCE CLIER, ALDA LUISA LERAYER,†
S. DUSKO EHRlich, AND MARIE-CHRISTINE CHOPIN*

INRA, Laboratoire de Génétique Microbienne, CRJJ,
78352 Jouy-en-Josas, France

Received 25 July 1997/Accepted 7 November 1997

Three genes coding for the endonuclease, methylase, and specificity subunits of a type I restriction-modification (R-M) system in the *Lactococcus lactis* plasmid pIL2614 have been characterized. Plasmid location, sequence homologies, and inactivation studies indicated that this R-M system is most probably of type IC.

Restriction-modification systems (R-M) are an effective barrier to protect strains from phage infection. As expected for bacteria which are under strong selective pressure in the dairy environment, due to the presence of bacteriophages, a number of R-M systems have been found in *Lactococcus lactis* strains (18). However, only four have been studied in detail. Three are of type II (5, 9, 31, 33, 34, 41) and one, composed of three genes associated with restriction activity and a type II's methylase, is unclassified (35). We previously established that plasmid pIL2614 codes for both the Abi420 phage abortive infection and for an R-M activity (efficiency of 10^{-4}) (36). To characterize the R-M system, plasmid pIL2614, extracted from strain IL1403 (4), was sequenced by chromosome walking in cycle extension reactions, using appropriate primers, *Taq* polymerase, fluorescent dye-coupled dideoxynucleotides and an Applied Biosystems sequencer ABI-373. The DNA and protein sequences were analyzed with the Genetics Computer Group software (6), Genmark (3), and Blast (1) programs. A sequence of 8.6 kb (accession no. U90222), localized upstream of the *abi420* genes, revealed five open reading frames (ORFs). The organization of this sequence and the general features of the ORFs are summarized in Fig. 1.

The *orf1*-specified protein has 66 to 78% identity with replication proteins designated RepB from lactococcal plasmids pSK11 (19), pSL2 (20), pCI528 (29), pUCL22 (11), pSV40 (44), pCI305 (17), and pWV02 (22). All these plasmids belong to a family replicating via theta intermediates. They have structural and DNA sequence similarities at the replication origins (11). This origin is composed of two 10-bp repeats ([T/A]TATATTT) spaced by 3 bp and followed by an AT-rich core containing CG clusters. This core is followed by three 22-bp repeats (TATAN₇AAAAAnCn₂TG [where n stands for any base pair]) and one that is truncated (11). The -35 box of the promoter is located immediately downstream of this origin. All these features are present upstream of *orf1*. There are two 10-bp repeats (ATTATTATTTn₃ TTATATATTT), an AT-rich core, three 22-bp repeats (CTTATACCTAGAAAAAC AATG), one truncated repeat (CTTATACCTAGAAA), and a putative promoter sequence (TTGTATn₁₇TATAAT). Therefore, *orf1* and the upstream DNA sequence are most probably involved in pIL2614 replication that proceeds as described previously for plasmid pUCL22 (11). RepB initiates replica-

tion by binding to the origin. This replication is bidirectional and would be under the control of RepB itself (11).

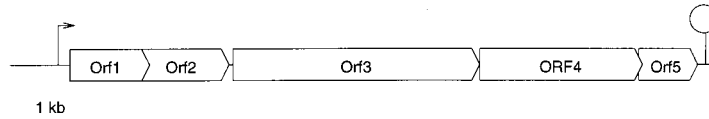
The *orf2*-specified protein has 53 and 57% identity with proteins encoded by *orfX* genes present downstream of *repB* on plasmids pUCL22 (11) and pJW563 (15), respectively. OrfX does not participate in plasmid replication, and its function remains unknown (11). The homology lies in two domains localized at the N and C terminal parts of the *orf2*-encoded protein. Between these two domains, this protein comprises a long helical domain that contains 10 repeats of 11 amino acids (aa) ([N/D]SLEDKQEKKA). These repeats are hydrophilic, and the protein is acidic (P_i 4.5). Therefore, Orf2 is probably neither involved in DNA binding nor membrane anchored as expected for proteins with such repeats.

The *orf3*-specified protein has 29.8, 28.4, and 24% identity with the *Escherichia coli* EcoR124II (8), *Mycoplasmma pulmonis* (7), and *Haemophilus influenzae* (10) endonuclease (R) subunits of type IC R-M systems, respectively. This homology is particularly high at the level of seven helicase-like domains (14, 32) (Fig. 2). Conservation of these domains suggested that the R subunits of type I R-M enzymes may possess helicase activity, playing a role in local unwinding of DNA at the cleavage sites and in DNA translocation (14). Recently, Titheradge et al. (40) identified two additional domains (X and Y) which were well conserved among the four enzymes presented in Fig. 2. Moreover, a 10th domain, localized between domains IA and II, designated Z, is well conserved in R subunits of type IC (Fig. 2). Conserved domains and sequence homologies suggest that *orf3* codes for a type IC endonuclease subunit.

The *orf4*-specified protein has 36.9, 36.9, and 35.5% identity with the *E. coli* EcoR124II (21) and *pr*r (42) and *M. pulmonis* (7) methylase (M) subunits respectively. This identity is in agreement with those (32% [38]) usually found for M polypeptides, thus indicating that Orf4 could be part of a type IC R-M system. An alignment of the sequences is shown in Fig. 3 (because of the identity between EcoR124II- and *pr*r-encoded M subunits, only that from EcoR124II is shown). Two sequence motifs conserved in the adenine methyltransferases (MTases), motif CMI ([D/E/S]X[F/A]XGXG) and motif CMII ([L/I/V/M/A/C]X[D/N]PP[Y/F]) (24, 28, 45) are present in Orf4 (Fig. 3). Motif CMI, found in both N and C MTases is the binding site for the cofactor S-adenosylmethionine. This has been shown by mutational analysis of EcoKI N⁶-adenine MTase (45) and by crystallographic studies of C⁵-cytosine HhaI MTase (23). Motif II probably plays a role in catalysis (45). The aromatic residue has been shown to be essential for methyl group transfer (45). Moreover, the nature of the conserved amino acid residue preceding PP(Y/F) is characteristic for different classes of MTases and correlates with the base

* Corresponding author. Mailing address: INRA, Laboratoire de Génétique Microbienne, CRJJ, 78352 Jouy-en-Josas, France. Phone: 33 01 34 65 25 30. Fax: 33 01 34 65 25 21. E-mail: mchopin@biotec.jouy.inra.fr.

† Present address: Instituto de Tecnologia de Alimentos, 1880 Campinas, S. P. Brazil.



Gene designation	Start	Stop	Product size (aa)	Similarities	Translation start
1	593 (AUG)	1753 (UGA)	386	RepB	AAGGAGcaacttctc ATG GAA ATT
2	1750 (AUG)	2628 (UGA)	292		AAAGGcAGAtttatc ATG AGT GAA
3	2639 (AUG)	5716 (UAA)	1025	HsdR	AGGgGaatcaa ATG AGT CAT
4	5716 (AUG)	7311 (UAA)	531	HsdM	GGAaaGaaGaatata ATG GCG ACA
5	7301 (AUG)	8518 (UAG)	405	HsdS	AGGcGtcoatg ATG AGT AAA

FIG. 1. (Top) Organization of the pIL2614 sequenced fragment. A putative promoter sequence is indicated by an arrow, and a transcription terminator is indicated by a circle atop a vertical line. (Bottom) Features of the ORFs. Start and stop numbers refer to positions in the sequence, the corresponding codons are shown in parentheses. The putative ribosome binding site and the beginning of the ORFs are shown in capital letters.

methylation specificity of the enzymes (39, 45). Therefore, depending on the nature of this amino acid (D, N, or S) and on consensus sequences at two additional conserved motifs (CMIs and CMIII), the subdivision of the *N*⁴-cytosine and *N*⁶-adenine MTases into five classes has been proposed (39). The presence of the NPPY motif together with sequence conservation at the two other domains (Fig. 3) suggests that Orf4 may be an *N*⁶-adenine MTase (N12 class). Moreover, for the three en-

zymes compared in Fig. 3, the CMIs motif defined by Timinkas et al. (39) can be extended to the 22 upstream amino acids.

Four additional domains, conserved among DNA methylases of type II R-M systems, have been identified by structure-guided analysis (30). These domains, not clearly apparent in the MTases compared in Fig. 3, could be absent from type I MTases. Two other motifs are well conserved among the three MTases aligned on Fig. 3: GOEX₄TXNLARMNX₂L located

RpIL2614 EcoR124II Rmycoplasma Rhaemophil	MSHSEQMIE--NQFIQLSEKENQW---YRPDLKSEELQCNFRGHLNRINLAVALVEBQL MTHQHTHTAENNFIVLVDKYTKABPTGDSVQSSEDLRELILQDLNRGG--YEPISVKSQ MSRNELLYE---KFPVDDLKVNQKY---VKLDIKNEEKIFELIPENIGRILNNIELTQQ MLNENDIEQLTLQRLQLSGWEYRYGKDLVPHGKFAAGDLSGVVFEQLREAVRKLNPQ	RpIL2614 EcoR124II Rmycoplasma Rhaemophil	RSVVKFKVKNQFPTMSALLTTHSIAQAKHYIRLIKEMKDNG---TLLNGRFDERHRLLID ----TFPGSKGNAMLAHSVSDAARAYATPKRLQBEAANKSATYKPLRIATIFPSFAAN -----FTNNRN--YNSIAPDTIQDALTYDEPFI---KMDVG---DIFATPIFSSYSN -----SQE-RLLELAADFVQHFARNEVVDV--KAMVV--SRQICVLDLNYQIIA
RpIL2614 EcoR124II Rmycoplasma Rhaemophil	LFDKPFQVQVVEFSRLTCTFPPLASQWLRG---ENGVAQVLLEREDGKRVTELEAFRND --SAMLAVREQLQNLANGVVFNDESERRRTEQYLDNPDSDILGKFRKTHIDYICDFIFDD ---NLLDRELLSNGSASSVRFAQYVW---PDTVKYKNDGPKKIKLKPFVDWE PRSAVDSVVKSATKSDZGLDVLVRNQTFYK----LLRDGVREYVTQNGEKIEMVRLVD	RpIL2614 EcoR124II Rmycoplasma Rhaemophil	KDFPRVAIFFTNPNDRLEKNEQDDE--LVEIMTKYKQFSDASFPQDBKLYNQINIKRLAK EQWALIGELISDEFTDTSAMDSSAKPELDAARREYNSEKFTNFSTDS--NQFQVYRLDAQ EE-----KNEKFPNLKHEKKE---ILKRYEKFPNTSFKRQD--FDRVYDQWR LH-----PEWSDNINEGAIK---IVMTGSASDTPMQKH--IYSQEKQKQLERR
RpIL2614 EcoR124II Rmycoplasma Rhaemophil	-----ISGG---TSSYEVVHVQVVP---DSSRVDRG-DVSLLINGLPIIHIELQKES- ERLBNYILDKRNLNRKMQVYIQQPEQ---AGSHANRY-DVLTLLVNGPLVQIELKKR-- -----NAN---NEFYVVLQPPTRDAANMGRF-DSLLINGPLLLPFKDK-- -----FEHWG---NNRFVAVVQLEIR--SRGGGRIPDIIIGFVNGLPVIVELKNPLR	RpIL2614 EcoR124II Rmycoplasma Rhaemophil	-----SQE-RLLELAADFVQHFARNEVVDV--KAMVV--SRQICVLDLNYQIIA -----SQE-RLLELAADFVQHFARNEVVDV--KAMVV--SRQICVLDLNYQIIA
RpIL2614 EcoR124II Rmycoplasma Rhaemophil	AKDGFMAQQYIQOR--YAEDGFFKQ---IYATTGIMVIPNKVDTRYRPAEEDTAEAYARKM -GVAIREAFNQIHR--YSKESFNSNSLFKYLQLEVISNGTDTRYFAN---TTKRDKNPF -SENINKAINQIDESYRGSVLNRK---IFRFIQLIGSNPEEVKFLAN---NKRTNNNK ESADLLQAFNPET--YKDEIAE---LFFVYNGALIISDGIVARIGLSL---ADFORFTPW	RpIL2614 EcoR124II Rmycoplasma Rhaemophil	DSGLLVGFRPFPTMKNVQNTFRIPNSNEN-----QNFQDLIP---REYEEVKPFIECSF TFGNIVTFRDLERSTIDATLFGDKNTKVVLEKSYTEYMEGFDAATGEAKRGMFTVVS KRGIIVNFSLSEQSINDAFKIYANSSDKE---IQQLVYGE---KYEQVVEDPINFWN NGGLIVDYVGLAKELRAATQYVNTS*GKG---QLAEDVQSVFFPKMKQLEPITLTFAT
RpIL2614 EcoR124II Rmycoplasma Rhaemophil	KELFNWRTEFN--QTVSDFDFTRTLVLRIPDAHELIISOYTLILDVDPKPKFMALRPFYQIH DFTWAKSDN--TLIKLDKDFATCFQKHTLNAVLYNSVFDSQ---TLLVMRPYQIA ILSGFTSENGSSKELLIDPLAVNALLREYLRVYVFGSDEK---TILRAPYQQR KVUDEKNKARLYPDDLEQLSLANGLMQPEDLLDVIYRVLFERDSVGRKTKKTAAYQY	RpIL2614 EcoR124II Rmycoplasma Rhaemophil	LYKQSEADLSENP---HDLKMTIAQVSAQKLGKSYKAFPSYDQYED----- ELEQRFFDPTIESEKKEKDFKLFQELIARINLQNDYDFATLKAQQID---LSDPVA SLKISKNTYDEKN--NEIFPNTSLLENKKVKIKNLSQVSNIPGSLKTEFYG--KVEK-- PIEGKTDFVQAALE--KDNPNLDMARIPAAHIIISLDQLSPDKAHQHWFKKETEPR--
RpIL2614 EcoR124II Rmycoplasma Rhaemophil	AIRKIRPKAA-----QHEGGFIWAHTGSGKTTIS--FVATKLLAQAIGAIVDRTVM ATERILWTKRSFTA--KNWSKPESSGGYIWHHTGSGKTTIS--FKAARLATELDP--IDKVVV AIKKAINFVEKQLRNDLAKHNINNAIYIWHHTGSGKTTIS--YKIAEILSKNSD--IDHVVV GVNEAVDSTWATSE---KGRRRGVMWHHTGSGKTSIMLFYAGKLLAQPELKNPTVVV	RpIL2614 EcoR124II Rmycoplasma Rhaemophil	FEAFSEVVEQLPQYRG--KTNVNTKIKEMIEDEHPEEDEFKLLQE---IAFSSQLNA VKFKPAEHYVDEDEKA--BLQITRILPADRIQDYRSAYNDRDQWR---EKBAEKKE ISDFSLQELNQYQWANEIKNLSSTNEKEKISYEVLNISDISNIKFA---YKEMIDIEI KKAPLKTAGLVKRYGMCGTLAEVFPYNGQIFAFYDAVRAILITKRQKGTGNERQILLKK
RpIL2614 EcoR124II Rmycoplasma Rhaemophil	VVDRDLDAQQTDEFKFASEYHTGQTTGNSVANTLIVGKKNQQLARNLLSKNNNTIL VDRKDLDTQTKKEYQRF--PDSVN-----GSENTAGLKNRNL-DKDNKII LVDRNDLNDQTSQTTFQKLSM-----SKNEK--IDPLNDQTSKLDYELFLKKEKLI VDRNDLNDGQLFTFSSGK-----DLIKQTF--QVEDRDLQRLQLA--NEVGGVF	RpIL2614 EcoR124II Rmycoplasma Rhaemophil	THKVDVDFYINQLLAKIQLENEAGAVEKPEKEIQQ---KDPQIQMYHTLLKQDLVNT SPTMDVVPEVDLKSQENLDYILGLTPEHNRQNG--KGEIMEVKKLIRSSLGNR YLENLFNNKIKSKYPNNRLYEDTLESIDKHQLTKNNYNQKLNQKSYEIPLLVQVK LVNQTVYSBGVIDLDFLLEKQPQIGLSISEEFLQTVKNSPTKMLVSMAMERYLAESIKVK
RpIL2614 EcoR124II Rmycoplasma Rhaemophil	Z VFTIQKLSAAMRSAAQSEEEKGNQFEKLRQEHIVFIVDEAHRADVDEMR----- VFTIQKLSAAMRSAAQSEEEKGNQFEKLRQEHIVFIVDEAHRADVDEMR----- VFTIQKLSAAMRSAAQSEEEKGNQFEKLRQEHIVFIVDEAHRADVDEMR----- FTTIQKLSAAMRSAAQSEEEKGNQFEKLRQEHIVFIVDEAHRADVDEMR----- FTTIQKLSAAMRSAAQSEEEKGNQFEKLRQEHIVFIVDEAHRADVDEMR-----	RpIL2614 EcoR124II Rmycoplasma Rhaemophil	TEEIDVAQ-----LKETS-----IQNEI-----QRLI--QKEAEFFGLSF AREGLVVD-----FIQOT-----NLDDLPPKASIIDAFFTPAQRE--QREABALIKE KNEIKNFF-----IKKDK-----SLDEKFIIDYGRKILKSVQKVNQIEAMWLEKIL SGTNTLQKDFERRLEKALNQHMNLVTVVLELDELFKMSQDFQBRALGKLLGTEEL
RpIL2614 EcoR124II Rmycoplasma Rhaemophil	II VFTIQKLSAAMRSAAQSEEEKGNQFEKLRQEHIVFIVDEAHRADVDEMR----- VFTIQKLSAAMRSAAQSEEEKGNQFEKLRQEHIVFIVDEAHRADVDEMR----- VFTIQKLSAAMRSAAQSEEEKGNQFEKLRQEHIVFIVDEAHRADVDEMR----- FTTIQKLSAAMRSAAQSEEEKGNQFEKLRQEHIVFIVDEAHRADVDEMR----- FTTIQKLSAAMRSAAQSEEEKGNQFEKLRQEHIVFIVDEAHRADVDEMR-----	RpIL2614 EcoR124II Rmycoplasma Rhaemophil	DFLOQAM-NEYQDGGKAIPLYTLHLLDSMTLS-----KEEPEPKAGE NLNEDAA--KRYIRTSLKREYATENGVDLNET-----LPKLSPIN--PQ HKGHGIN--NDQIRKDKKRIKINDKLDLDEKSE-----FIKKWRSREKVDKDIIDK AFYEALSQNQSAKLDMDGDEVLSKLAKETITLRKSVTIDMQYKEAVRARILRLVRRALQK
RpIL2614 EcoR124II Rmycoplasma Rhaemophil	III ---IKKILPNSTWFLGTQTFEFAANKQENGTARTTSQQYQGLHYSYTKNAMDGAV ---LKKPKRYYQFGFTQTFEFAANKQENGTARTTSQQYQGLHYSYTKNAMDGAV ---IKKILPNSTWFLGTQTFEFAANKQENGTARTTSQQYQGLHYSYTKNAMDGAV ---LKKPKRYYQFGFTQTFEFAANKQENGTARTTSQQYQGLHYSYTKNAMDGAV ---IKKILPNSTWFLGTQTFEFAANKQENGTARTTSQQYQGLHYSYTKNAMDGAV	RpIL2614 EcoR124II Rmycoplasma Rhaemophil	YRRFPKV--LEERLRNFQEQKWKBEEL YTKK---QAVFRKSSRLRSLKA--- LSIEYKESTETFLDPTKMMKLESKI--- YKYPPDK--QBAVTVYVKAQAEIADLTLG
RpIL2614 EcoR124II Rmycoplasma Rhaemophil	IV LGFQVYHSLISEEDLIVTVQLNKGKLPGDALQQEELLPAELYEKDEHIRTMLQKIFNR LKFVQVYNDV---PQFKSLETEDEKLLSAAEQQAFHPMIREITQYILNNFRQKTRH LGF-----KVMYQETRIFFREN-----NSNLGKIKSIINIVKSKHLD VFIVYDAR-----QIRLKKRHDFALFAEID-----ALLEEPTSLRLEKRLG	RpIL2614 EcoR124II Rmycoplasma Rhaemophil	

FIG. 2. Alignment of the predicted amino acids of the pIL2614 HsdR peptide with the R subunits of EcoR124II (8), *M. pulmonis* (7), and *H. influenzae* (10). Sequence accession numbers are U90222, X13145, L25415, and L45919, respectively. Helicase-like domains I to VI (14) and X and Y domains (40) as well as the additional conserved domain Z are shown in boldface letters. Conserved amino acids and conservative or semiconservative substitutions are indicated by an asterisk and a period, respectively.

MLlac2614 MA-----TGLNQLWASADILRGKMDASEYKNVLLGLIFPKYLSDAQLREVVEQENK
 MEcoR124II MKMTSQQRAELHRQWITANDVRSVGVGDFKQVVLGALFVYFISEN--PSSVLEAGDD
 Mycoplasma MS-----NSKRELTAVKTKCDQLRSKRVETVRYIYIMGLHLEFFKYLSE-QSKNPFSEPKR
 * * * * *
 MLlac2614 TDTFPERSTQYAGFMWEYEDKDDLEINIQPKQGVYIQPDQLFYSYRIKADNYEENLTDL
 MEcoR124II SICVARKLDDSVIT-----DDIKDDAIKT---KGYFVYPSQLPCNVAARANTDRLNADL
 Mycoplasma VD-YIK---YSEFDENHEQFKIKELIINQDDDFFLAYKYSFQNVVDMQKGVNVIPTI
 * * * * *
 MLlac2614 QAGFNELERQGEFS-----GLFADIDLNSKLGSNALLRNVTITEVLRALDEIDL-
 MEcoR124II NSIPVAIESSAYGVYSEAD--IKGLFADFDTTNSRLGNVTKDKNARLAALVKGVEGLKLG
 Mycoplasma EESFNKIESINSELNDEKKEFFKDLFTNIDFSNKNLGNIDEEKKTQLIITKEINTLNLS
 * * * * *
 CMI
 MLlac2614 -P-EHNGDVIQDAVEYLLGKFPASSAGKKAAGEFYTPQAVSKIMSEITSGQETRAPPHIVD
 MEcoR124II DFNEHQIDLFGDAYEFLISNYAANAGSKSGGFEFTPHQVSKLIAQLAMHQTHVN--KIYD
 Mycoplasma --MDEVDFHGMTEYELLSEFASDTGKKAAGEFYTPSKVSELLVKIVSHGKNKIN--KAYD
 * * * * *
 CMI
 MLlac2614 PAMSGSLMLNTR--VLNPPQVHYHQQLLNTTETLARMNLILHGDKERMYLANAGDPL
 MEcoR124II PAAGSGLLLQAKKQFDNHIIEGFGQIINHTTYNLRMMPLHININVDPKILGNL
 Mycoplasma PACGSGSLILKLANKGVKY---NKIYQEVKTKATYNLRMMNLILGVPFSLDRSGDPL
 * * * * *
 CMI
 MLlac2614 -DADWSEEPYQFDSVCMNPPYSAKMSAADQ--FLSDPRFRFGKLPAPSKADFAFLHGG
 MEcoR124II TEPHFRDEKP--PDAIVSNPFYKVMKIGSDDEPTLINDERFAPAGVLPAPSKADFAFLVHA
 Mycoplasma INP-LHIEEGSFGDVIKAVPFYKQKMPQGE--LSKDRYNSYSLAPSKADFAFLVQH
 * * * * *
 CMII
 MLlac2614 FYHL-KEGTMGIVLPHGVLFPGA--AEGTIRQALLEMGDAIDVGLPANIFFGSPKPT
 MEcoR124II LNYL-SARGRALIVCFPIGIPYRGG--AQKIRQYLVNDDNYVETVLSLAANLFFGPTIAVN
 Mycoplasma LPHVNDKMGILASVFLSILGSRKSPKABEDIRKYIDKNYIDTIIFLFPNLFYNTSIESC
 * * * * *
 MLlac2614 VIIKRNRS--RRDVLFDASQDFEKRNQNVLLDEHIDKIVSIHKREDIERYAHVAF
 MEcoR124II ILVLSKHT--DTNVQFIDASELFFKETTNNLITDAHIEQIMQVFAKEDVAHLAKSVAF
 Mycoplasma IIVARKNKPTNDKRIFMINATKEIQNAKKNQVLSDENINRIFSAAKKEKREEMFNKYISY
 * * * * *
 MLlac2614 DEIQENDENLNIIPRY-VDTPEEE-EPVDLIVAVNTNLLKINEELVQQGVLLSMIDNFAES
 MEcoR124II ETWVANDYNSVSSY-VEAKDNK-EIIDIABLNAELKTTVSKIDQLKIDDAIVAEIEG
 Mycoplasma EDIVKNEYLSMRFYDLNDPESEEDIDIDFVSEIWKINEELLYENEFKMLNPELN-
 * * * * *
 MLlac2614 EENQALIESMRLRLRGHDE
 MEcoR124II EVOK
 Mycoplasma -----KKN

N12 class G-ffTP--aschhhp-h hLnPscCGsGsha-sh CMII hDhhhgPPY h---hphLpsGG-LshahP
 All --s-P--Lhp-hhph hLnPFGsGsh--sh hDhhhgPPY h---h-p-Lpsg-G-hhh-h-s

FIG. 3. Alignment of the predicted amino acids of the pIL2614 HsdM peptide with the M subunits of *EcoR124II* (21) and *M. pulmonis* (7) (GenBank accession no. U90222, X13145, and L25415, respectively). Conserved amino acids and conservative or semiconservative substitutions are indicated by an asterisk and a period, respectively. Conserved motifs (CMIs, CMI, CMII, and CMIII) are shown in boldface letters, and the proposed consensus for the N12 class and all MTases (39) are indicated below. Different groups of amino acids are indicated as follows: p, polar (D, E, N, H, K, R, S, Q, and G); h, hydrophobic (W, F, I, L, M, V, A, P, Y, C, and T); n, negatively charged (D and E); f, aromatic (F, W, Y, and H); a, aliphatic (I, L, V, and M); c, charged (D, E, K, R, and H); and s, small, nonbulky (G, A, S, T, D, N, P, and V).

downstream of CMI and LAPKSKADFAF located just upstream of CMIII. They could be significant in relation to special properties of MTases of the type IC R-M systems.

The *orf5*-specified protein has 26% identity (45% homology) with the putative specificity (S) subunit of a *Spiroplasma citri* type I R-M system (27). Moreover, Orf5 presents structural organization characteristics of S subunits. Two repeats of 38 aa (designated A and A'), present in the central and the C-terminal part of the protein, respectively, have 87% identity. Parts of these repeats are homologous to the repeats (24 aa) identified in all S subunits from type I restriction enzymes (21) (Fig. 4). Two split repeats (designated D and D'), characteristic of type IC S subunits (25), are present in the N-terminal and the central parts of Orf5. Homologies between the central conserved domain and sequences near the N and C termini were proposed to favor a circular organization of the domains of the S subunit, which provides the required symmetry for interactivity with the M subunits and the target DNA sequence (40).

These sequence homologies and gene structure suggest that *orf5* codes for an S subunit which is part of an R-M system including the R (Orf3) and M (Orf4) subunits described above. Based on amino acid identities observed for both R and M, this system must be of type IC. However, the *L. lactis* HsdS protein lacks the TAEI direct repeats characteristic of S subunits of

type IC enzymes (37), the number of which has been shown to determine the length of the nonspecific spacer between the specific domains of the recognition sequence (2). Nevertheless, these repeats are absent in the S subunit of the type IC R-M system of *M. pulmonis* (7) as well as in S subunits of other type I enzymes.

In order to confirm that the region from Orf3 to Orf5 confers the R-M phenotype, a 5,358-bp *EcoRV-SacI* segment, from position 2968 (309 bp downstream of the start codon of *hsdR*) to position 8326 (192 bp upstream of the stop codon of *hsdS*), was deleted from plasmid pIL2614. This segment was replaced by a chloramphenicol resistance cassette recovered from plasmid pGKV259 (43) and previously cloned in pBlue-script plasmid (pIL1388) (1a). The construct was designated pIL1032. Phage bIL170 propagated on strain IL1403 showed efficiencies of plating of 3×10^{-3} and 1 when plated on strains IL1403(pIL2614) and IL1403(pIL1032), respectively. Phages picked up from plaques formed on the pIL2614-harboring strain were no longer restricted by this strain. In contrast, phages picked up on the pIL1032-harboring strain were still restricted by IL1403(pIL2614) with an efficiency of plating of 5×10^{-3} . The loss by pIL1032 of the aptitude to restrict and/or modify the growth of phage bIL170 indicated that the region from Orf3 to Orf5 confers the R-M phenotype. In contrast, pIL1032 still conferred the Abi420 phenotype active on the phage bIL41 (35).

Genes of type I R-M systems of enterobacteria are arranged into two contiguous transcription units, with *hsdM* and *hsdS* forming an operon and *hsdR* being transcribed from its own promoter. The order of the two transcriptional units is different for different families (46), and this has been proposed as an additional evidence for a horizontal transfer of the *hsd* genes (40). This organization differs in *M. pulmonis*, in which the gene order is *hsdS hsdR hsdM*, with only one promoter upstream of *hsdS* and the expression of the genes being con-

D
 SLlac2614 **MMSKSPQLRFEQFTDDWEERKFGVWVKSSERNLNLEYSPKQVLSVAQMKLNPNRNNEQ**
 SLlac2614 DDYMKTYNVLKKGDIAFEGNKSFSFAGFRVLDLDDQGVISHVFXVYRICKMDDTDFMIV
 SLlac2614 YINNESVMKYLVLKATTKTLMTTLNKTDKIVPKPLNPSLE**EQQKIGSFFKQLDAPIALH**
 SLlac2614 **QRKLDLLKEQKQYFQKMF**PKNGAKV**PELRFAGFADDWEDRKL**GELASFGKNGYTKNDL
 SLlac2614 VEFGDPIILYGRLYTKYETVIEKVDTFVNKKDKSIIISGGSEVIVPASGESSEDISRASVV
 SLlac2614 GRSGIILGGDLNIIKPVNYIDSIFLALPISNIGSQQKEMSKRAQKGSVVHLHNSDLKQVNI
 SLlac2614 LYPK**LEEQKIGSFFKQLDNTIVLHQRKLDLFLKEQKQFLQKMF***
 * * * * *
 Repeats of 24 amino-acids
 consensus iP*ppI*EQ-rIv**Id*1-al*d
 Central domain SEcoR124II **NPEKSLAIQSEIVRILDKFTALTA**
 SMycoplasma **LP--NLKIQSAILKIEPEKEDLFF**
 SSpiroplasma **IP--SLQEQTKIVNFFSIIIDRKIE**
 SLlac2614 **LP--SLEEQKIGSFFKQLDAPIA**
 C terminal domain SEcoR124II **IPVFNINEQQRIVEILDKFDTLTN**
 SMycoplasma **IP--NLKIQSAILKIEPEKEDLFF**
 SSpiroplasma **FVSLNYKQEQTKIANFSIIDR-KIE**
 SLlac2614 **LP--SLEEQKIGSFFKQLDAPIA**

FIG. 4. Predicted amino acid sequence of the pIL2614 HsdS peptide. Thirty-nine-amino-acid repeats (A and A') and split repeats (D and D') (25) are shown in boldface letters. The consensus for the 24-aa repeats present in all S peptides (21) is indicated below, together with sequences from *EcoR124II* (21), *M. pulmonis* (7), and *S. citri* (27).

trolled by inversion of a DNA element (7). In *L. lactis*, the absence of consensus sequences for a promoter upstream of *hsdR* and *hsdM* together with gene organization and the presence of a putative terminator structure downstream of *hsdS* (GCCCTAAGATCTAACCTTTATATCTTAGGGGCTATTT TTTT) suggests that the five genes identified could be transcribed from the promoter located upstream of *repB*. However, as weak promoters transcribing type I genes are difficult to spot in DNA sequences, functional analysis will be needed to identify transcriptional units. It has been proposed that autoregulation of the RepB protein could be under the control of heat-shock proteins (11). If this were true, *hsd* genes would be activated under stress conditions and therefore perhaps after phage infection.

Type I R-M systems are able to evolve rapidly. A single subunit that concomitantly confers sequence specificity to both restriction and modification facilitates the acquisition of new specificities. Moreover, an S polypeptide has two recognition domains, each specifying one component of the bipartite target sequence (2). In a given family of S polypeptides, the two variable recognition domains are separated by a conserved core sequence. It has been established *in vivo* (12) and *in vitro* (13, 16) that the *hsdS* genes can recombine at the level of the conserved domain, creating a functional R-M system with an entirely new specificity.

In conclusion, our report describes the second functionally characterized (7) and the third (47) type I R-M system described for gram-positive bacteria. Its location on a plasmid and probably under the control of its replication machinery could both increase plasmid stability by postsegregational killing of plasmid-free cells (26) and possibly allow activation of the R-M system by the stress due to phage infection. This, in addition to the facility to acquire new specificities, confers an obvious selective advantage. Therefore, plasmid-encoded type I R-M systems are likely to be widespread in the *L. lactis* species and possibly other bacteria exposed to phage-abundant environments.

We thank C. Anagnostopoulos for critically reading the manuscript and J. Anba for the gift of pIL1388.

Alda Luisa Lerayer was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (Brasília, Brazil).

REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Anba, J. Unpublished data.
- Bickle, T. A., and D. H. Krüger. 1993. Biology of DNA restriction. *Microbiol. Rev.* **57**:434–450.
- Bodorovsky, M., and J. M. Initch. 1993. Genmark: parallel gene recognition for both DNA strands. *Comput. Chem.* **17**:123–133.
- Chopin, A., M. C. Chopin, A. Moillo-Batt, and P. Langella. 1984. Two plasmid determined restriction and modification systems in *Streptococcus lactis*. *Plasmid* **11**:260–263.
- Davis, R., D. van der Lelie, A. Mercenier, C. Daly, and G. F. Fitzgerald. 1993. *ScrFI* restriction-modification system of *L. lactis* subsp. *cremoris* UC503: cloning and characterization of two *ScrFI* methylase genes. *Appl. Environ. Microbiol.* **59**:777–785.
- Devereux, J., P. Haerberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the vax. *Nucleic Acids Res.* **12**:387–395.
- Dybvig, K., and Y. Huilan. 1994. Regulation of a restriction and modification system via DNA inversion in *Mycoplasma pulmonis*. *Mol. Microbiol.* **12**:547–560.
- Firman, K., and W. Glover. 1989. Basis for changes in DNA recognition by the *EcoR124* and *EcoR124/3* type I DNA restriction and modification enzymes. *J. Mol. Biol.* **205**:115–125.
- Fitzgerald, G. F., C. Daly, L. R. Brown, and T. R. Gingeras. 1982. *ScrFI*: a new sequence-specific endonuclease from *Streptococcus cremoris*. *Nucleic Acids Res.* **10**:8171–8179.
- Fleishmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J.-F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. A. Fields, J. D. Gocayne, J. D. Scott, R. Shirley, L.-I. Liu, A. Glodeck, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Uterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J.-L. Fritchman, J.-L. Fuhrmann, N. S. M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter. 1995. Whole-genome sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**:496–512.
- Frère, J., M. Novel, and G. Novel. 1993. Molecular analysis of the *Lactococcus lactis* subspecies *lactis* CNRZ270 bidirectional theta replicating lactose plasmid pUCL22. *Mol. Microbiol.* **10**:1113–1124.
- Fuller-Pace, F. V., L. R. Bullas, H. Delius, and N. E. Murray. 1984. Genetic recombination can generate altered restriction specificity. *Proc. Natl. Acad. Sci. USA* **81**:6095–6099.
- Gann, A. A. F., A. J. B. Campbell, F. F. Collins, A. F. W. Coulson, and N. E. Murray. 1987. Reassortment of DNA recognition domains and the evolution of new specificities. *Mol. Microbiol.* **1**:13–22.
- Gorbalenya, A. E., and E. V. Koonin. 1991. Endonuclease (R) subunits of type-I and type-III restriction-modification enzymes contain a helicase-like domain. *FEBS Lett.* **291**:277–281.
- Gravesen, A., J. Josephsen, A. von Wright, and F. K. Vogensen. 1995. Characterization of the replicon from the lactococcal theta-replicating plasmid pJW563. *Plasmid* **34**:105–118.
- Gubler, M., D. Bragaglia, J. Meyer, A. Piekawicz, and T. A. Bickle. 1992. Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. *EMBO J.* **11**:233–240.
- Hayes, F., C. Daly, and G. F. Fitzgerald. 1990. Identification of the minimal replicon of *Lactococcus lactis* subsp. *lactis* UC317 plasmid pCI305. *Appl. Environ. Microbiol.* **56**:202–209.
- Hill, C. 1993. Bacteriophage and bacteriophage resistance in lactic acid bacteria. *FEMS Microbiol. Lett.* **12**:87–108.
- Hornig, J. S., K. M. Polzin, and L. L. McKay. 1991. Replication and temperature-sensitive maintenance functions of lactose plasmid pSK11L from *Lactococcus lactis* subsp. *cremoris*. *J. Bacteriol.* **173**:7573–7581.
- Jahns, A., A. Schafer, A. Geis, and M. Teuber. 1991. Identification, cloning and sequencing of the replication region of *Lactococcus lactis* subsp. *lactis* biovar *diacetylactis* Bu2 citrate plasmid pSL2. *FEMS Microbiol. Lett.* **80**:253–258.
- Kannan, P., G. M. Cowan, A. S. Daniel, A. A. F. Gann, and N. E. Murray. 1989. Conservation of organization in the specificity polypeptides of two families of type I restriction enzymes. *J. Mol. Biol.* **209**:335–344.
- Kiewet, R., S. Bron, K. de Jonge, G. Venema, and J. F. M. L. Seegers. 1993. Theta replication of the lactococcal plasmid pWV02. *Mol. Microbiol.* **10**:319–327.
- Klimasauskas, S., S. Kumar, R. J. Roberts, and X. Cheng. 1994. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **7**:357–369.
- Klimasauskas, S., A. Timinkas, S. Menkevicius, D. Butkiene, V. Butkus, and A. Janulaitis. 1990. Sequence motifs characteristic of DNA [cytosine-N¹] methyltransferases: the two domains of global similarity within DNA-methylases. *Exp. Biol.* **1**:4–12.
- Kneale, G. G. 1994. A symmetrical model for the domain structure of type I DNA methyltransferases. *J. Mol. Biol.* **243**:1–5.
- Kulakauskas, S., A. Lubys, and S. D. Ehrlich. 1995. DNA restriction-modification systems mediate plasmid maintenance. *J. Bacteriol.* **177**:3451–3454.
- Laigret, F., P. Gaurivaud, and J.-M. Bove. 1996. The unique organization of the *rpoB* region of *Spiroplasma citri*: a restriction and modification system gene is adjacent to *rpoB*. *Gene* **171**:95–98.
- Lauster, R. 1989. Evolution of type II DNA methyltransferases. A gene duplication model. *J. Mol. Biol.* **206**:313–321.
- Lucey, M., C. Daly, and G. F. Fitzgerald. 1993. Identification and sequence analysis of the replication region of the phage resistance plasmid pCI528 from *Lactococcus lactis* ssp. *cremoris* UC503. *FEMS Microbiol. Lett.* **110**:249–256.
- Malone, T., R. M. Blumenthal, and X. Cheng. 1995. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* **253**:618–632.
- Moineau, S., S. A. Walker, E. R. Vedamuthu, and P. A. Vandenberg. 1995. Cloning and sequencing of *LlaII* restriction/modification genes from *Lactococcus lactis* and relatedness of this system to the *Streptococcus pneumoniae DpnII* system. *Appl. Environ. Microbiol.* **61**:2193–2202.
- Murray, N. E., A. S. Daniel, G. M. Cowan, and P. M. Sharp. 1993. Conservation of motifs within the unusually variable polypeptide sequences of type I restriction and modification enzymes. *Mol. Microbiol.* **9**:133–143.
- Nyengaard, N., F. K. Vogensen, and J. Josephsen. 1992. *LlaAI* and *LlaBI*, two type II restriction endonucleases from *Lactococcus lactis* subsp. *cremoris* W9 and W56 recognizing, respectively, 5′-/GATC-3′ and 5′-/C/TRYAG-3′. *Gene* **136**:371–372.
- Nyengaard, N., F. K. Vogensen, and J. Josephsen. 1995. Restriction-modification systems in *Lactococcus lactis*. *Gene* **157**:13–18.
- O’Sullivan, D. J., K. Zagula, and T. R. Klaenhammer. 1995. *In vivo* restriction by *LlaI* is encoded by three genes, arranged in an operon with *llaIM*, on

- the conjugative *Lactococcus* plasmid pTR2030. *J. Bacteriol.* **177**:134–143.
36. **Portugal, C., F. Clier, R. Valyasevi, S. D. Ehrlich, and M.-C. Chopin.** Characterization of the lactococcal Abi420 determinant for abortive infection. Unpublished data.
 37. **Price, C., J. Lingner, T. A. Bickle, K. Firman, and S. W. Glover.** 1989. Basis for changes in DNA recognition by the *EcoR*124 and *EcoR*124/3 type I DNA restriction and modification enzymes. *J. Mol. Biol.* **205**:115–125.
 38. **Sharp, P. M., J. E. Kelleher, A. S. Daniel, G. M. Cowan, and N. E. Murray.** 1992. Roles of selection and recombination in the evolution of type I restriction-modification systems in enterobacteria. *Proc. Natl. Acad. Sci. USA* **89**:9836–9840.
 39. **Timinkas, A., V. Butkus, and A. Janulaitis.** 1995. Sequence motifs characteristic for DNA [cytosine-N4] and DNA [adenine-N6] methyltransferases. Classification of all DNA methyltransferases. *Gene* **157**:3–11.
 40. **Titheradge, A. J. B., D. Ternent, and N. E. Murray.** 1996. A third family of allelic *hsd* genes in *Salmonella enterica*: sequence comparisons with regions implicated in restriction of DNA. *Mol. Microbiol.* **22**:437–447.
 41. **Towney, D. P., R. Davis, C. Daly, and G. F. Fitzgerald.** 1993. Sequence of the gene encoding a second *ScrFI* m⁵C methyltransferase of *Lactococcus lactis*. *Gene* **136**:205–209.
 42. **Tyndall, C., J. Meister, and T. A. Bickle.** 1994. The *Escherichia coli* *prf* region encodes a functional type IC DNA restriction system closely integrated with an anticodon nuclease gene. *J. Mol. Biol.* **237**:266–274.
 43. **Van der Vossen, J., D. Van der Lelie, and G. Venema.** 1987. Isolation and characterization of *Lactococcus lactis* subsp. *cremoris* Wg2-specific promoters. *Appl. Environ. Microbiol.* **53**:2452–2457.
 44. **von Wright, A., S. Wessels, S. Tynkkynen, and M. Saarela.** 1990. Isolation of a replication region of a large lactococcal plasmid and use in cloning of a nisin resistance determinant. *Appl. Environ. Microbiol.* **56**:2029–2035.
 45. **Willcock, D. F., D. T. F. Dryden, and N. E. Murray.** 1994. A mutational analysis of the two motifs common to adenine methyltransferases. *EMBO J.* **13**:3902–3908.
 46. **Wilson, G. G.** 1991. Organization of restriction-modification systems. *Nucleic Acids Res.* **19**:2539–2566.
 47. **Xu, G., J. Willert, W. Kapfer, and T. A. Trautner.** 1995. *Bsu*CI, a type I restriction-modification system in *Bacillus subtilis*. *Gene* **157**:59.