

Rolling-circle plasmids from *Bacillus subtilis*: complete nucleotide sequences and analyses of genes of pTA1015, pTA1040, pTA1050 and pTA1060, and comparisons with related plasmids from Gram-positive bacteria

Wilfried J.J. Meijer^{1,a}, G. Bea A. Wisman^a, Peter Terpstra^b, Peter B. Thorsted^c,
Chris M. Thomas^c, S. Holsappel^a, Gerard Venema^a, Sierd Bron^{a,*}

^a Department of Genetics, Groningen Biomolecular Sciences and Biotechnology Institute, Kerklaan 30, 9751 NN Haren, The Netherlands

^b BioMedical Technology (BMTc), University of Groningen, Hanzeplein 1, Building 25, FO-wing, 9713 GZ Groningen, The Netherlands

^c School of Biological Sciences, University of Birmingham, P.O. Box 363, Edgbaston, Birmingham B15 2TT, UK

Received 7 July 1997; revised 3 November 1997; accepted 30 November 1997

Abstract

Most small plasmids of Gram-positive bacteria use the rolling-circle mechanism of replication and several of these have been studied in considerable detail at the DNA level and for the function of their genes. Although most of the common laboratory *Bacillus subtilis* 168 strains do not contain plasmids, several industrial strains and natural soil isolates do contain rolling-circle replicating (RCR) plasmids. So far, knowledge about these plasmids was mainly limited to: (i) a classification into seven groups, based on size and restriction patterns; and (ii) DNA sequences of the replication region of a limited number of them. To increase the knowledge, also with respect to other functions specified by these plasmids, we have determined the complete DNA sequence of four plasmids, representing different groups, and performed computer-assisted and experimental analyses on the possible function of their genes. The plasmids analyzed are pTA1015 (5.8 kbp), pTA1040 (7.8 kbp), pTA1050 (8.4 kbp), and pTA1060 (8.7 kbp). These plasmids have a structural organization similar to most other known RCR plasmids. They contain highly related replication functions, both for leading and lagging strand synthesis. pTA1015 and pTA1060 contain a mobilization gene enabling their conjugative transfer. Strikingly, in addition to the conserved replication modules, these plasmids contain unique module(s) with genes which are not present on known RCR plasmids of other Gram-positive bacteria. Examples are genes encoding a type I signal peptidase and genes encoding proteins belonging to the family of response regulator aspartate phosphatases. The latter are likely to be involved in the regulation of post-exponential phase processes. The presence of these modules on plasmids may reflect an adaptation to the special conditions to which the host cells were exposed. © 1998 Federation of European Microbiological Societies. Published by Elsevier Science B.V.

Keywords: *Bacillus subtilis*; Rolling-circle plasmid; Mobilization; Response aspartate phosphatase; Signal peptidase; Replication module

* Corresponding author. Tel.: +31 (50) 363 2105; Fax: +31 (50) 363 2348; E-mail: S.Bron@BIOL.RUG.NL

¹ Present address: Centro de Biología Molecular, "Severo Ochoa" (CSIC-UAM), Universidad Autónoma, Canto Blanco, 28049 Madrid, Spain

Contents

1. Introduction	338
2. Classification of <i>B. subtilis</i> RCR plasmids	339
3. Sequences and structural organization of pTA1015, pTA1040, pTA1050, and pTA1060	340
4. Replication modules	342
4.1. <i>Rep</i> genes, Rep proteins and DSOs	342
4.1.1. Characterization of <i>rep</i> genes and Rep proteins	342
4.1.2. Promoter regions	345
4.1.3. Double-strand origins	345
4.1.4. Minimal replication regions	345
4.2. Plasmids within one group are highly homologous and may be identical	346
5. Single-strand origins	346
5.1. PalT1- and palT2-type SSOs	346
5.2. Delineation of palT SSOs	348
6. Mobilization modules (pTA1015 and pTA1060)	349
6.1. Absence of γ - <i>gtp</i> genes (pTA1015 and pTA1060)	349
6.2. Functional <i>mob</i> genes (pTA1015 and pTA1060)	351
6.2.1. Relatedness of Mob proteins	352
6.2.2. Conserved modules in Mob proteins	353
6.3. OriT/RS _A sites	353
7. ORF2C modules (pTA1015, pTA1040 and pTA1060)	355
8. <i>ORF1/Sip-P</i> modules (pTA1015 and pTA1040)	355
8.1. Orf1 encodes a putative export protein	355
8.2. Analysis of SipP	355
8.3. Possible function of the <i>ORF1/sipP</i> modules	356
9. <i>Rap</i> modules (pTA1040, pTA1050, and pTA1060)	357
9.1. Rap family of proteins	357
9.2. Expression and function of <i>Rap</i> genes	357
9.3. ComA boxes	357
9.4. Translationally coupled small genes	360
10. Unique modules (pTA1040, pTA1050 and pTA1060)	363
10.1. ParABC (pTA1050)	363
10.2. ORF3C40 (pTA1040)	363
10.3. ORF7C60 (pTA1060)	363
10.4. ORF4.60 and ORF5.60 (pTA1060)	364
11. Conclusions	364
Acknowledgements	364
References	365

1. Introduction

Numerous plasmids of relatively small size have been identified in a wide range of Gram-positive bacteria. Most of these use the rolling-circle mechanism of replication which is characterized by the synthesis of single-stranded (ss) DNA intermediates. Several of these plasmids have been studied extensively (for reviews see [1–4]). Naturally occurring rolling-circle replicating (RCR) plasmids form the basis of many

cloning and expression vectors for Gram-positive bacteria. Although the highly transformable *Bacillus subtilis* strain 168, which has become the paradigm for most studies, does not harbor endogenous plasmids, the existence of *B. subtilis* plasmids has been known for quite some time [5–11]. Since these plasmids do not confer easily selectable phenotypes to their hosts, most cloning vectors used today for *B. subtilis* are based on RCR plasmids from other Gram-positive bacteria, like staphylococci and strep-

tococci [1–3,12–15]. These plasmids can replicate and express their antibiotic resistance genes in *B. subtilis*. However, vectors based on such plasmids are frequently unstable [1,3,16–19].

So far, small cryptic plasmids from *B. subtilis* have attracted relatively little attention. Such plasmids have raised our interest for several reasons: first, for studies on plasmid replication and plasmid stability, and the construction of stable cloning vectors for *B. subtilis*. One of the factors underlying plasmid instability of non-native plasmid vectors is non-optimal plasmid–host interaction [18,20–23]. Thus, vectors based on endogenous *B. subtilis* RCR plasmids are expected to be superior to vectors based on non-native plasmids. Second, interesting novel traits may be present on cryptic *B. subtilis* plasmids. Such studies on the gene functions on plasmids would nicely complement ongoing research on the function analysis of *B. subtilis* chromosomal genes. In the latter project, coordinated by S.D. Ehrlich (INRA, Jouy-en-Josas, France) and N. Ogasawara (Nara Institute of Science and Technology, Nara, Japan), about 25 groups in Europe and Japan cooperate to identify the function of as many chromosomal genes as possible. This project is a follow-up of the recently finished combined European–Japanese project aimed at the determination of the entire sequence of the *B. subtilis* chromosome [24], which was coordinated by F. Kunst (Institut Pasteur, Paris, France) and N. Ogasawara (Nara Institute of Science and Technology, Nara, Japan).

Of the plasmids studied here, several are mobilizable and, hence, can contribute not only to the total gene pool of the host but also to the horizontal gene pool of the bacterial population. Since *B. subtilis* plasmids are generally several kilobasepairs larger than most RCR plasmids from other Gram-positive bacteria, they can easily accommodate genes other than those required for replication and mobilization. Such genes may facilitate molecular adaptation of the host to certain conditions. These genes might even be of industrial relevance, since many of the endogenous plasmids of *B. subtilis* known today were isolated from industrial strains. It is known that several valuable traits for dairy industry are located on plasmids, in that case from lactic acid bacteria [25]. Third, the comparison of several *B. subtilis* plasmids and the determination of their com-

plete nucleotide sequence will give insight in their structural organization and evolutionary relationship.

In the present work we compare the complete DNA sequences of pTA1015 (5.8 kbp), pTA1040 (7.8 kbp), pTA1050 (8.4 kbp; also known under the name pPOD2000 [10]), and pTA1060 (8.6 kbp), and describe computer-based and experimental analyses of the function of the identified putative genes. By including information available from other *B. subtilis* RCR plasmids, we present an overview of our current knowledge on *B. subtilis* RCR plasmids.

2. Classification of *B. subtilis* RCR plasmids

Uozumi et al. [7] analyzed twenty *B. subtilis* strains that harbor relatively small plasmids. Based on size and restriction profiles these plasmids have been classified into six groups, represented by pTA1015, pTA1020, pTA1030, pTA1040, pTA1050, and pTA1060. The production of trace amounts of ssDNA replication intermediates by each of these plasmids indicates that they all use the rolling-circle mechanism of replication [26]. Plasmids isolated in independent studies by Hara et al. [9] (pUH series); Tanaka et al. [5,6] (pLS series); Darabi et al. [27] (pBS2); Devine et al. [28] (pBAA1), and by Poluektova et al. [29] (p1410) can also be classified into one of these six groups. An RCR plasmid belonging to a seventh group, represented by pFTB14, was isolated from the *Bacillus amyloliquefaciens* strain S294 [8]. *B. subtilis* and *B. amyloliquefaciens* are highly related. In fact, based on DNA homology studies, the *B. subtilis* strain IFO3022 harboring pTA1060 is now considered to be a *B. amyloliquefaciens* strain [30]. Whereas most of these plasmids were identified in industrial isolates, Nezametdinova et al. [31] identified small cryptic plasmids in 21 out of the 32 natural *B. subtilis* strains isolated from soil by Kozlovsky and Prozorov [32]. Six of these, p1410 and others [29], belong to the pTA1020 group. An overview of the classification of the *B. subtilis* RCR plasmids is shown in Table 1. Plasmids of identical size and restriction profile are likely to be identical or at least highly similar to each other, and are placed in the same group. In addition, plasmids have been isolated from other *Bacillus* strains that cannot be placed in

Table 1
Classification of cryptic RCR plasmids from *B. subtilis*

Family	Type of SSO ^a	Size ^b	Plasmids with identical sizes and restriction profiles
pTA1015	1	5.8	pTA1015, pTA1010 (=pLS15), pTA1011 (=pLS17), pTA1012, pTA1013 (=pLS19), pTA1014, pTA1016, pTA1017 (=pLS24), pTA1018 (=pLS26), pTA1019, pUH1 through pUH8
pTA1020	1	6.6	pTA1020, pTA1021 (=pLS28), pTA1022 (pLS30), pTA1023, pBAA1, p1410
pTA1060	1	8.7	pTA1060 (=pLS11), pTA1061 (=pLS12), pBS2
pTA1040	2	7.7	pTA1040 (=pLS13)
pTA1030	2	7.2	pTA1030, pTA1031
pTA1050	2	8.4	pTA1050 (=pLS14)
pFTB14	–	8.2	–

^a1 and 2 represent the palT1- and palT2-type SSOs, described in Section 5.

^bPlasmid size in kbp. pTA series according to Uozumi et al. [7]; pLS series according to Tanaka and coworkers [5,6]; pUH series are according to Hara et al. [9].

one of these groups [33,34]. Since only limited information is available for the latter plasmids they are not discussed here.

Tanaka et al. [5] noticed that several *B. subtilis* strains contained, in addition to a relatively small plasmid (the pLS series), a larger plasmid with an average size of approximately 55 or 77 kbp. The minimal replicon of one of these, pLS20, appeared to represent a novel type of theta replicon [35]. As judged from their size, it is likely that also the other large plasmids use theta replication. Thus, the co-residence of two plasmids, one using the theta mode and the other the rolling-circle mode of replication, appears to be common in *B. subtilis*. To our knowledge, the co-residence of two RCR plasmids or two theta-replicating plasmids in cells of natural *B. subtilis* or *B. amyloliquefaciens* strains is rare. A possible exception has been observed by E. Poluektova who noticed the presence of two related replicons in one and the same *B. subtilis* strain isolated from soil (personal communication; A.A. Prozorov). This is in contrast to other Gram-positive bacteria, like lactococci, which can harbor several theta-type plasmids within one cell [36,37].

For a number of *B. subtilis* RCR plasmids additional information is available. First, the replication regions of pBAA1 (pTA1020 group; [28]); pBS2 (pTA1060 group; [27]); pLS11 (pTA1060 group; [38]); pUH1 (pTA1015 group; [39]); and pFTB14 [40] have been sequenced. Second, the single-strand origins (SSOs) (see Section 5) of pTA1060 [26]; pTA1015 [26]; pLS11 (pTA1060 group; [41]); pBAA1 (pTA1020 group; [28,42]), and pTA1040

[26] have been analyzed. Third, a 1.73-kbp region of plasmid pUH1 (pTA1015 group), claimed to contain the γ -glutamyltranspeptidase (γ -*gtp*) gene, has been sequenced [43].

3. Sequences and structural organization of pTA1015, pTA1040, pTA1050, and pTA1060

RCR plasmids consist of several interchangeable modules which frequently show considerable homology at the DNA and/or deduced protein level [1,3,44]. An essential module comprises the replication initiation gene (*rep*) and its cognate target site, the double-strand origin (DSO). Another module contains the SSO which functions as the major initiation site for lagging-strand synthesis. Several RCR plasmids contain, in addition, a module with a gene, denoted *mob* or *pre*, which is involved in conjugative mobilization and site-specific recombination, respectively.

For the analysis of the structural organization of the *B. subtilis* plasmids pTA1015, pTA1040, pTA1050 and pTA1060, we determined their complete nucleotide sequence. The sizes of pTA1015, pTA1040, pTA1050, and pTA1060 are 5807 bp, 7837 bp, 8397 bp, and 8737 bp, respectively. The G+C contents are 40.7% (pTA1015), 36.9% (pTA1040), 38.1% (pTA1060), and 40.0% (pTA1050). The sequences were deposited in the EMBL/GenBank/DBJ database under the accession numbers U32379 (pTA1015), U32378 (pTA1040), U32380 (pTA1060), and U55043

Table 2

Characteristics of the ORFs on pTA plasmids

ORF/gene	Location ^a	Size (codons)	Mol. mass (kDa) ^b	Putative start codon and RBS ^c	Spacing ^d (bp)
pTA1015					
1. <i>rep15</i>	166–1182	339	39.5	<u>cagaaggag</u> gtttttttgttcATG	12
2. ORF1.15	1401–1838	146	15.9	tgct <u>aggagg</u> gaaagttttATG	12
3. <i>sip15</i>	1945–2502	186	21.2	atata <u>gagg</u> gaaatttctATG	10
4. <i>mob15</i>	2840–4245	482	56.6	ctgaat <u>ggggg</u> gtttttctcATG	11
5. ORF2C15	4778–4302 (C) ^e	159	18.8	tttttac <u>aggtg</u> atcgttATG	11
pTA1040					
1. <i>rep40</i>	165–1181	339	39.7	gtcaga <u>ggg</u> gttttctcATG	11
2. ORF3C40	1829–1227 (C)	201	22.2	tta <u>aggagg</u> gtttgaacaatATG	13
3. ORF1.40	2826–3251	142	15.8	<u>gaaaggat</u> ggaagaactATG	16
4. <i>sip40</i>	3372–3926	185	21.5	cca <u>agcggg</u> ggaagcgtaaGTG	10
5. ORF2C40	4587–4117 (C)	157	18.3	cata <u>aaaggtg</u> aacccctATG	12
6. <i>rap40</i>	5943–7067	375	44.3	tgct <u>gaagg</u> agagatgtATG	10
7. <i>phr40</i>	7060–7176	39	4.1	tc <u>ggagg</u> gcgattcttctATG	15
pTA1060					
1. <i>rep60</i>	164–1183	340	39.5	gctcaga <u>agg</u> gttttttTTG	9
2. ORF7C60	1759–1220 (C)	180	20.1	<u>aaagaagg</u> gttttttaaTTG	14
3. <i>mob60</i>	2438–3883	482	56.8	ctgaat <u>cgggg</u> gtttttgtcATG	10
4. ORF2C60	4387–3911 (C)	159	18.8	tttttac <u>aggtg</u> atcgtgATG	11
5. ORF4.60	4825–5739	305	35.6	catacata <u>agg</u> gagatttATG	8
6. ORF5.60	5763–6158	132	15.1	<u>aggggat</u> gggggagtattATG	9 or 15
7. <i>rap60</i>	7199–8323	375	44.3	ta <u>ggggg</u> aggttactcgaATG	13
8. <i>phr60</i>	8316–8426	37	3.9	ttca <u>agggg</u> cgatttccgtATG	12
pTA1050					
1. <i>rep50</i>	164–1189	341	39.7	ggcaga <u>agg</u> gttttttctTTG	10
2. T7C	1724–1404	106	11.7	atata <u>gaaagg</u> gagatgaaTTG	9
3. T5	1863–2387	174	19.0	atattagaa <u>agg</u> gagattTTG	7
4. T8C	2781–2518	87	9.8	tta <u>aaaaa</u> gaaagccgctcATG	10
5. <i>parA</i>	2940–3774	277	30.6	ctttgt <u>agg</u> attgtccaaATG	10
6. <i>parB</i>	3821–4384	187	20.9	tttaaaat <u>ttggg</u> gatattATG	7
7. <i>parC</i> [C]	4632–4426	68	7.7	aaat <u>ggaag</u> ttgttctgTTG	14
8. ORF4C.50	5231–4899	110	12.7	tga <u>acgag</u> gtggaattatATG	11
9. <i>rap50</i>	6660–7778	372	44.0	atgtat <u>ggggg</u> gattgatgATG	11
10. <i>phr50</i>	7768–7884	38	4.0	ttca <u>agggg</u> cgatttgaatATG	12

^aCoordinate numbers for pTA1015, pTA1040, pTA1060 and pTA1050 correspond to those in the EMBL/GenBank Data Library; accession numbers U32379, U32378, U32380, and U55043, respectively. The origin nick sites were taken as position 3 for all four plasmids. In the case of pTA1050, these coordinates differ from those under accession number U55043, since in the database submission a different starting point, on the opposite strand, was used. For reasons of comparison we have used here a coordinate numbering for pTA1050 which corresponds to that of the other plasmids.

^bMolecular mass was calculated based on the deduced amino acid sequence.

^cPutative initiation codons (capital letters) and the 22-bp upstream sequences are shown; nucleotides complementary to the 3'-end of *B. subtilis* 16S rRNA (UCUUUCCUCCACUAG) region [119] are underlined.

^dThe spacing is calculated as the distance from the first nucleotide to the right of the AGGA sequence (or the equivalent) to the nucleotide at the 5'-side of the initiation codon.

^e'C' indicates that the ORF is located on the opposite strand.

(pTA1050). For pTA1015, pTA1040 and pTA1060 the putative nick sites for replication initiation were given coordinate number 3; in the database entry of

pTA1050 (U55043) this position corresponds to coordinate 2776 and is located in the opposite strand relative to the other plasmids.

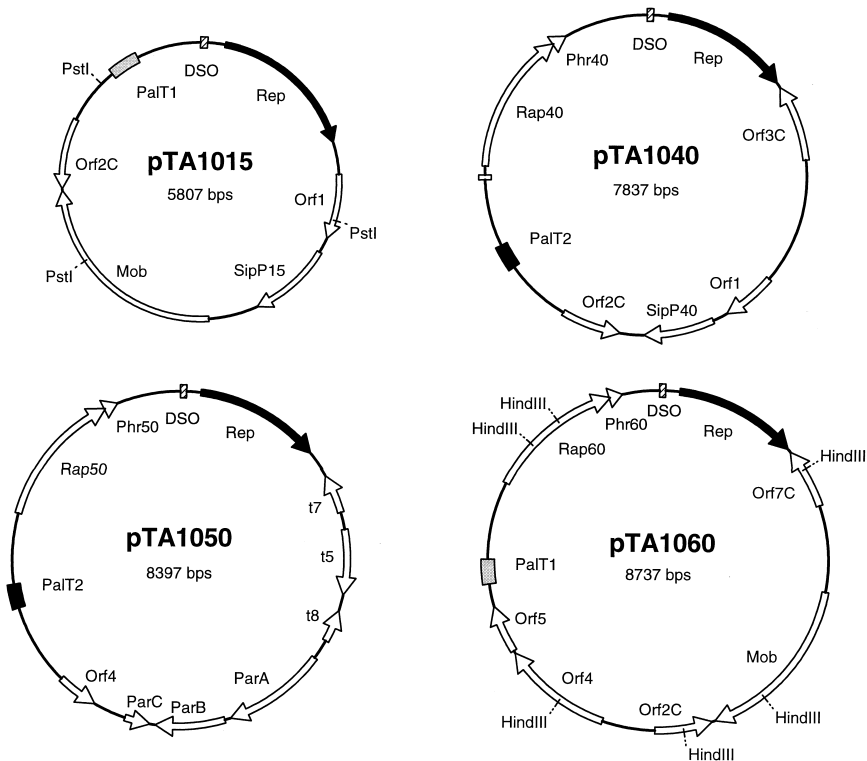


Fig. 1. Structural organization of plasmids pTA1015, pTA1040, pTA1050, and pTA1060. The positions of the primary and secondary replication origins (DSO and SSO; the latter are indicated as palT1 or palT2) are shown with rectangles; those of the ORFs and their direction of transcription are indicated with arrows. In addition, the position of the *Pst*I sites of pTA1015 and the *Hind*III sites of pTA1060, used for delineation of the replication region, is shown.

The DNA sequences and deduced amino acid sequences of identified open reading frames (ORFs) of the four pTA plasmids were compared to known sequences in the databases. This revealed the following putative genes and non-coding DNA sequences, which are also commonly found in other RCR plasmids: (i) a *rep* gene and cognate DSO located upstream of the gene (pTA1015, pTA1040, pTA1060, and pTA1050); (ii) an SSO (pTA1015, pTA1040, pTA1060, and pTA1050); and (iii) a *mob* gene (pTA1015 and pTA1060). In addition, other putative genes were identified, some of which showed significant homology to previously identified genes. These plasmid-located genes were given names according to the nomenclature of their homologues. The directions of transcription and the localization of the genes, as well as the positions of the DSOs and the SSOs (indicated as palT1 or palT2), are shown in Fig. 1. Characteristics of the identified genes are

shown in Table 2. The structural organization of the four pTA plasmids, i.e. Rep-encoding gene with an upstream located DSO, a separated SSO, and the majority of the genes being transcribed in the direction coinciding with the direction of replication, is similar to most other RCR plasmids. In the following sections comparisons and experimental analyses of the different genes and modules are discussed.

4. Replication modules

4.1. Rep genes, Rep proteins and DSOs

4.1.1. Characterization of rep genes and Rep proteins

All RCR plasmids studied so far contain a *rep* gene encoding a Rep protein that is essential for the initiation of replication. Through their DNA-binding and nicking/closing activity the Rep proteins

introduce a strand- and site-specific nick in the plasmid at the DSO. The 3'-OH end of the nick site is subsequently used for the initiation of leading strand synthesis.

Based on sequence similarity of their DSOs and homology of the replication proteins, RCR plasmids from Gram-positive bacteria can be divided into five classes, represented by pT181, pC194, pE194, pSN2, and pTX14-3 [1,45]. Based on these criteria the four pTA plasmids belong to the pC194 class of plasmids. Also the other *B. subtilis*/*B. amyloliquefaciens* RCR plasmids of which the replication regions have been sequenced so far belong to the pC194 class [46]. An alignment of the deduced Rep protein sequences of the four pTA plasmids described here with those of published sequences of other *Bacillus* plasmids is shown in Fig. 2A. This alignment shows that the four pTA plasmids are highly homologous to each other and to the known Rep proteins of the other *B. subtilis*/*B. amyloliquefaciens* RCR plasmids.

Despite their strong overall homologies, the published amino acid sequences of the Rep proteins specified by pTA1060, pLS11, and pBS2, all belonging to the pTA1060 group, seem to deviate in their termini (Fig. 2A). These differences can, at least in part, be attributed to sequence errors. The 3'-end of *rep* (pBS2), compared to the corresponding region of *rep* (pTA1060), contains three differences (one additional C at position 1943, an A/T substitution at position 1985, and a deletion of one T at position 2011 [27]), which are likely to result from sequence errors in pBS2. Even more variation seems to be present in the N-termini of these proteins. The differences between pLS11, pBS2 and pTA1060 in this region can be traced back to a few basepairs. Repeated sequencing gives us confidence that the pTA1060 sequence is correct. Introduction of the corresponding corrections in pBS2 and pLS11 results in N-termini of their Rep proteins identical to that of Rep (pTA1060). The published sequence of the Rep protein of pBAA1 (pTA1020 group) is at its N-terminus more than 30 amino acids shorter than the Rep proteins of pTA1060, pTA1015, pTA1040, and pFTB14 [28]. However, the reading frame of *rep* (pBAA1) is open upstream of its published ATG start codon. In fact, in this region the pBAA1 sequence differs at only one position from that of pTA1060. It is, therefore, likely that also the Rep

protein specified by pBAA1 is identical to the other Rep proteins discussed here. Similarly, the published sequence of the Rep protein of pUH1, belonging to the pTA1015 group, deviates at the N-terminus from that of pTA1015. Also in this case the differences can be traced back to a few basepairs. Since the DNA sequences of pTA1015 and pUH1 are 99.3% identical over 5807 bp, we consider it likely that these differences also result from sequence errors and that the Rep proteins of these plasmids are identical.

A consequence of the corrections following from these considerations is that the actual size of the Rep proteins encoded by pLS11, pBS2, pBAA1, and pUH1 is larger than their published sequence (all contain from 339 to 341 amino acids). Additional support for the corrections is the observation that in all cases a potential ribosomal binding site (RBS) is located upstream of the proposed new start codons. Another implication is that only the *rep* genes of pTA1015, pUH1 and pFTB14 have an ATG start codon; the others have TTG as start codon.

The fact that the various Rep proteins encoded by plasmids of different groups are highly homologous to each other suggests that these plasmids are derived from a common ancestor. The dendrogram (Fig. 2B) shows that the Rep proteins of pBAA1 and pTA1060 are the most closely related. These plasmids also contain an almost identical gene downstream of their *rep* gene (see Section 10). Among the Rep proteins compared in Fig. 2B, those of pTA1040 and pTA1050 deviate most from the others and these can be grouped separately. In agreement with this classification into a subgroup is the observation that pTA1040 and pTA1050 contain a palT2-type SSO (see Section 5), whereas the other plasmids (the SSO of pFTB14 is not known) contain a palT1-type SSO [26].

Ilyina and Koonin [47] compared a considerable number of replication initiation proteins required for rolling-circle replication from bacteria, eukarya and archaea. This revealed three conserved amino acid sequence motifs which are also conserved in the deduced sequences of the Rep proteins of the four pTA plasmids and the other plasmids shown in Fig. 2A. The middle motif (II) consists of the sequence His-Hy-His-Hy-Hy-Hy (His and Hy representing a histidine and a bulky hydrophobic residue). Based on

Rep1040	M-YSSSESYSHFDQDKTATGKKRDKWGGK - VRSLLMAAHYEGLEKRRGAPYYGKKAEKVCCCAECL - APFK - RDLETRRLRLYQAIFCKV	84
Rep1050	MFYSDENYSILDDKTATGKKRDRGKK - VRSLLVAAHYDGLKRRGAPYYGKKAEKVCCCAECL - VFR - RDVETGRRLRLSNAYIFCKV	85
RepFTB14	M-YSSSENDYILEDKTATGKKRDRGKKR - RANLMAEHYEALEKRRGAPYYGKKAERLSCAEHLS - PK - RDPETGRKLKLYQAHFCKV	84
Rep1015	M-YSSSENDYILEDKTATGKKRDRGKKR - RANLMAEHYEAALQSKRPPYYGKKAELSCAECLS - PK - RDPETGKLLYQAQIFCKV	84
Rep1060	M-YSSSENDYILEDKTATGKKRDRGKKR - RANLMAEHYEALESKIGAPYYGKKAELISCAEYLS - PK - RDPETGKLLYQAHFCKV	85
RepBA1	MAEHYEALESKIGAPYYGKKAELISCAEYLS - PK - RDPETGKLLYQAHFCKV	52
RepBS2	MAEHYEALESKIGAPYYGKKAELISCAEYLS - PK - RDPETGKLLYQAHFCKV	52
RepLS11	MIHISLILKIKPQVKSGIGRGRDGRSRT - VEALESKIGAPYYGKKAELISCAEYLS - PK - RDRRPGLKLLYQAHFCKV	78
RepUH1	MIHISLILRTPQVKSGIGRGRDGRSRT - TMLKLYRKLVLITMLKLRNCAVVRNVELETRRRKLLKLYQAQIFCKV	78
<i>motif I</i>		
Rep1040	RLCPMCAMRRSLKIAHNKLIIVEANRQYKPAWIFLTLTVKNVEGDDLKQTIISDMMKGFNKLMMKYKKVDSAVRGYFRALEITKNH	146
Rep1050	RLC[S]MCAMRRSLKIA[S]HNKLIIVEANRQYKESWIFLTLTVRNVGDDLKQTIITDMMKAMNRLFYKVTSLVGFRALEITKNH	147
RepFTB14	RLCPMCAMRRSLKIAIYHNKLIIEANRQYGGWIFLTLTVRNVKGERLKPQIISMMEGFRKLFQYKVKTSVGFRALEITKNH	146
Rep1015	RLCPMCAMRRSLKIAIYHNKLIIVEANRQYGGWIFLTLTVRNVGDDLKPMIADMMKGMNRLFYKRVKVA[A]LGYFRALEITKNH	147
Rep1060	RLCPMCAMRRSLKIAIYHNKLIIVEANRQYGGWIFLTLTVRNVKGERLKPQISMMEGFRKLFQYKVKTSVGFRALEITKNH	147
RepBA1	RLCPMCAMRRSLKIAIYHNKLIIVEANRQYGGWIFLTLTVRNVKGERLKPQIISMMEGFRKLFQYKVKTSVGFRALEITKNH	114
RepBS2	RLCPMCAMRRSLKIAIYHNKLIIVEANRQYGGWIFLTLTVRNVKGERLKPQIISMMEGFRKLFQYKVKTSVGFRALEITKNH	114
RepLS11	RLCPMC[R]MRRSLKIAIYHNKLIIVEANRQY[C]CGWIFLTLTVRNVKGERLKH[H]ISMMEGFRKLFQYKVKTSVGFRALEITKNH	140
RepUH1	RLCPMCAMRRSLKIAIYHNKLIIVEANRQYGGWIFLTLTVRNVGDDLKPMIADMMKGMNRLFYKRVKVA[A]LGYFRALEITKNH	140
<i>motif II</i>		
Rep1040	EEN[T]YHHPHVLIPVRI[T]GYF - GKNYIN[Q]KEWTS[F]WKRAMKLDYTPIV[H]VQRVYKGGKIDAE[A]EQEYRKE[M]ENKAI[L]EISKYPVK	245
Rep1050	E[Q]NTYHHPHVLIPVRSY[F]GHE[F]YIKQAEWTS[L]WKRAMKLDYTPIV[H]VQRVYKGGKIDAE[E]IEKEV[K]E[M]EQKAI[L]EISKYPVK	247
RepFTB14	EEDTYHHPHVLIPVRRNYF - GKNYIKQAEWTS[L]WKRAMKLDYTPIV[D]IRRVYKGRV[K]IDAEQIESDYREAMMEQKAVLEISKYPVK	245
Rep1015	EEDTYHHPHVLIPVKK[S]YF - [T]HNYIKQSEWTS[L]WKRAMKLDYTPIV[D]IRRVYKGRAKIDAEQIESDYREAMMEQKAVLEISKYPVK	245
Rep1060	EEDTYHHPHVLIPVKNYF - GKNYIKQAEWTS[L]WKRAMKLDYTPIV[D]IRRVYKGRAKIDAEQIESDYREAMMEQKAVLEISKYPVK	246
RepBA1	EEDTYHHPHVLIPVKNYF - GKNYIKQAEWTS[L]WKRAMKLDYTPIV[D]IRRVYKGRAKIDAE[L]IENDVREAMMEQKAVLEISKYPVK	213
RepBS2	EEDTYHHPHVLIPVKNYF - GKNYIKQAEWTS[L]WKRAMKLDYTPIV[D]IRRVYKGRAKIDAE[L]IENDVREAMMEQKAVLEISKYPVK	213
RepLS11	EEDTYHHPHVLIPVKNYF - GKNYIKQAEWTS[L]WKRAMKLDYTPIV[D]IRRVYKGRV[K]IDAEQIESDYREAMMEQKAVLEISKYPVK	239
RepUH1	EEDTYHHPHVLIPVKK[S]YF - [T]HNYIKQSEWTS[L]WKRAMKLDYTPIV[D]IRRVYKGRAKIDAEQIESDYREAMMEQKAVLEISKYPVK	239
<i>motif III</i>		
Rep1040	DTDVIRGN[V]TEENL[T]VYLLDALSARRLIGYGGILKEIHKELNLTDAEDGLIRIEEDDDVANE[T]FEVMA[H]HVGIKNYI[T	339
Rep1050	DTDVIRGN[V]TEENL[T]VYLLDALSARRLIGYGGILKEIHKELNLTDAEDGLVRIEEDDDVANE[T]FEVMA[R]WHVGIKNYI[Q	341
RepFTB14	DTDVVR[S]KVTDDNLT[V]YLLDALSARRLIGYGGILKEIHKELNLTDAEDGLVRIEEDDDVANGAFEFVMA[W]HHPGIKNYI[L	339
Rep1015	DTDVVRGNKVTEDNLT[V]YLLDALSARRLIGYGGILKEIHKELNLTDAEDGLVRIEEDDDVANGAFEFVMA[Y]HHPGIKNYI[R	339
Rep1060	DTDVVRGNKVTEDNLT[V]YLLDALSARRLIGYGGILKEIHKELNLTDAEDGLVRIEEDDDVANGAFEFVMA[Y]HHPGIKNYI[R	340
RepBA1	DTDVVRGNKVTEDNLT[V]YLLDALSARRLIGYGGILKEIHKELNLTDAEDGLVRIEEDDDVANGAFEFVMA[Y]HHPGIKNYI[R	307
RepBS2	DTDVVRGNKVTEDNLT[V]YLLDALSARRLIGYGGILKEIHKELNLTDAEDGLVRIEEDDDVANGAFEFVMA[Y]HHPGIKNYI[R	308
RepLS11	DTDVVRGNKVTEDNLT[V]YLLDALSARRLIGYGGMLKEIHKELNLTDAEDGLVRIEEDDDVANGAFEFVMA[Y]HHPGIKNYI[R	333
RepUH1	DTDVVRGNKVTEDNLT[V]YLLDALS[R]RRRLIGYGGILKEIHKELNLTDAEDGLVRIEEDDDVANGAFEFVMA[Y]HHPGIKNYI[R	333

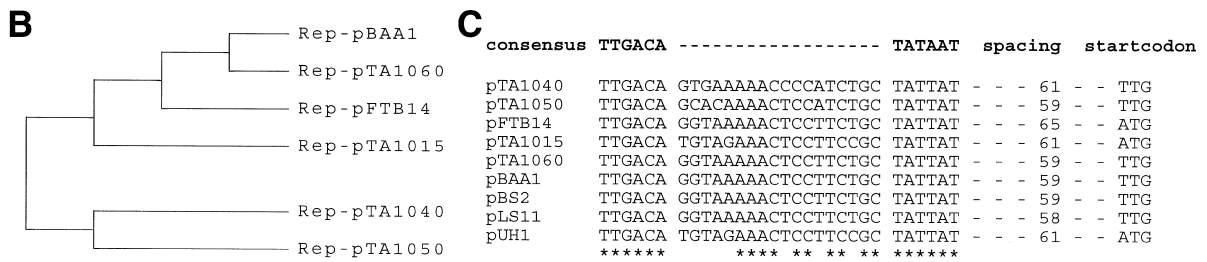


Fig. 2. Comparison of the replication regions of *Bacillus* RCR plasmids. Sequence information was taken from: pLS11, Hara et al. [38]; pBS2, Darabi et al. [27]; pUH1, Hara et al. [39]; pBAA1, Devine et al. [28]; and pFTB14, Murai et al. [40]. A: Alignment of the deduced amino acid sequences. Rep proteins of pTA1040, pTA1050, pFTB14, pTA1015, pTA1060, pBAA1, pBS2, pLS11, and pUH1 (for pLS11, pBS2, pUH1, pBAA1, and pFTB14, the proteins are according to the published data). Conserved amino acid residues, when present in at least five of these Rep proteins, are boxed. The three conserved amino acid motifs present in most replication initiation proteins of rolling-circle plasmids [47] are indicated with bars. The positions of the catalytic Tyr and Glu in the pC194 RepA protein [49] are indicated with asterisks. B: Dendrogram of Rep proteins. The relatedness between the Rep proteins of plasmids pBAA1 (pTA1020 group), pTA1060, pFTB14, pTA1015, pTA1040, and pTA1050 is shown. The dendrogram is based on the 'average linkage cluster analysis' as described by Sneath and Sokal [106]. Based on the levels of identity the Rep proteins can be divided into two groups. Group 1 contains the Rep proteins of pBAA1 (pTA1020 group), pTA1060, pFTB14 and pTA1015; group 2 those of pTA1040 and pTA1050. Whereas the levels of identity between plasmids within one group vary between 83% and 93%, those between Rep proteins from group 1 and group 2 vary between 72% and 77%. C: Alignment of putative promoters of the *rep* genes. The consensus *B. subtilis* σ^A promoter sequence is shown in the top line. Nucleotides that are conserved in all the DNA sequences shown are indicated with asterisks. In addition, the number of nucleotides separating the putative promoters from their start codon (spacing) and the start codons themselves are shown (for pBAA1, pBS2, pLS11, and pUH1, these features are based on the proposed corrections for start codons as discussed in the corresponding part of the text).

analogies with metalloenzymes, the authors hypothesized that the two conserved histidine residues may be involved in the binding of metal ions required for Rep activity. The tyrosine residue present in the C-terminal motif (III) corresponds to a tyrosine residue in the Rep proteins of the RCR plasmids pT181 and pC221. This Tyr residue has been shown to become covalently linked to the nicked DNA at the DSO [48]. Gros et al. [49] showed that at least three amino acid residues of the pC194 replication protein have a catalytic role. One of these is the conserved tyrosine residue in motif III; the other two are glutamates. The corresponding amino acids (indicated in Fig. 2A) are, except for one glutamate in pLS11, conserved in all plasmids included in the present comparison.

4.1.2. Promoter regions

Murai et al. [40] delineated the promoter region of the pFTB14 *rep* gene to about 95 bp. Within this region, sequences were identified resembling the -35 and -10 consensus sequences of *B. subtilis* σ^A promoters. Highly homologous sequences are present upstream of the *rep* genes of the four pTA

plasmids and the other listed plasmids (Fig. 2C), indicating that the expression of these *rep* genes is driven by highly homologous promoters.

4.1.3. Double-strand origins

Characteristically, DSOs are present in a plasmid region that has the potential to form secondary structures. The sequence 5'-TCTTGATA-3' is found at the original nick site of most members of the pC194 group of plasmids [1,46]. As shown in Fig. 3, this consensus nick-site sequence is also conserved in the four pTA plasmids and present within a region of dyad symmetry located upstream (about 160 to 170 bp) from the start codon of its cognate *rep* gene.

4.1.4. Minimal replication regions

The region containing the *rep* gene and its cognate DSO is sufficient to drive replication of plasmids pUH1 [39], pFTB14 [40], pLS11 [38], and pBAA1 [28]. Similarly, the replication functions of pTA1060, present on a 2.2-kbp *Hind*III fragment [50], which contains the *rep* gene and its DSO (Fig. 1), are sufficient for replication. Likewise, the 2.4-kbp *Pst*I region of pTA1015, containing its *rep*

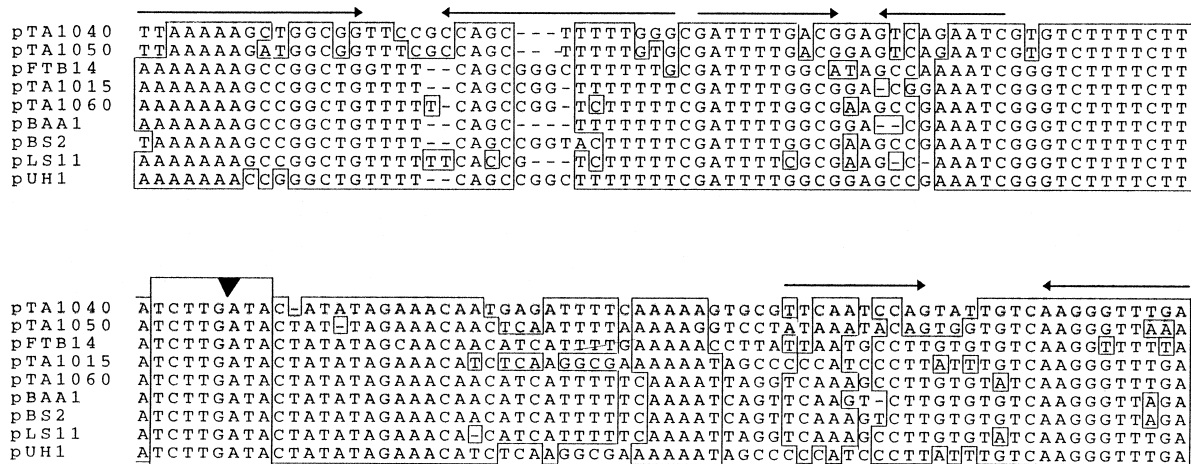


Fig. 3. Alignment of DSO regions. Conserved nucleotides are boxed. Inverted repeated sequences are indicated with arrows. The highly conserved 5'-TCTTGATA-3' sequence, present in almost all plasmids belonging to the pC194 family, is indicated by an enlarged box, and the putative origin nick site within this sequence is indicated with a triangle.

gene and DSO (Fig. 1), is sufficient to drive autonomous replication in *B. subtilis* [51]. As described in more detail elsewhere (Bron et al., J. Biotechnol., in preparation) the replication region of pTA1060 was used to construct the versatile *B. subtilis*/*E. coli* shuttle vector pHB201, which can be obtained from the *Bacillus* Genetic Stock Center (Dr. R. Zeigler, Ohio State University, Department of Biochemistry, 484 West 12th Ave, Columbus, OH; E-mail: dzeigler@magnus.acs.ohio-state.edu).

4.2. Plasmids within one group are highly homologous and may be identical

As described in Section 2, most *B. subtilis* plasmids can be classified into 7 groups based on their size and restriction profile. This suggests that plasmids within one group may be highly homologous to each other. Comparison of known DNA sequences of corresponding regions of various plasmids belonging to the same group supports this idea. This is clearly illustrated with pTA1015 and pUH1, the entire sequences of which are known (this work; [52]). The level of sequence identity of these plasmids is 99.3% over 5807 bp. The few discrepancies between the two plasmid sequences are predominantly located in regions that have the potential to form secondary structures and are likely to be due to sequence errors. Very high levels of DNA homology are also

observed when the available DNA sequences of pTA1060, pBS2 and pLS11 (i.e. the *rep* gene region), all belonging to the pTA1060 group, are compared: pTA1060 and pBS2 are 97.5% identical over 2279 bp; pTA1060 and pLS11 96.9% over 1606 bp; and pBS2 and pLS11 93% over 1606 bp. Again, the differences between these plasmids are mainly located in regions that have the potential to form secondary structures. In addition, DNA sequences upstream of the *rep* gene of pBS2 [27] and pTA1060 (this work) are almost identical.

These observations show that, at least in the cases discussed here, plasmids within one group are highly homologous and possibly even identical.

Taken together, the data described in this section on replication modules show that the minimal replicons of all *B. subtilis*/*B. amyloliquefaciens* plasmids analyzed so far are highly conserved, indicating that they are derived from a common ancestor.

5. Single-strand origins

5.1. *PalT1*- and *palT2*-type SSOs

Efficient conversion of ssDNA into duplex plasmid DNA of RCR plasmids is initiated from specific, non-coding plasmid regions, the SSOs, which have a high potential to form secondary structures. Previ-

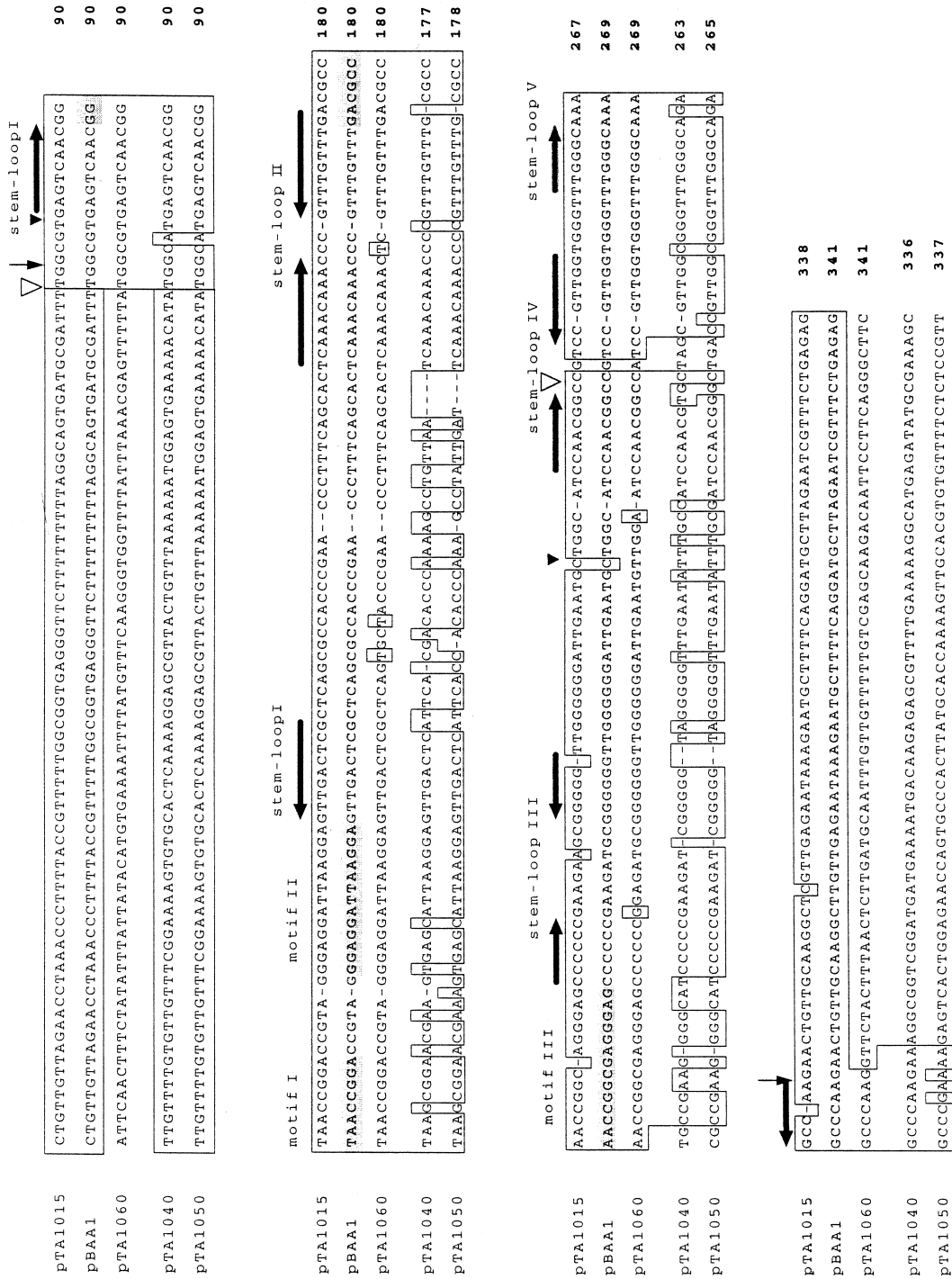


Fig. 4. Alignment of SSOs. The regions comprising the SSOs of plasmids pTA1015, pBAA1 (pTA1020 family), pTA1060, pTA1040, and pTA1050, are shown. In each of the four blocks shown, the upper three lines represent the palTI subclass of SSOs (present on pTA1015, pBAA1 and pTA1060), and the lower two lines the palT2 subclass of SSOs (present on pTA1040 and pTA1050). Potential stem-loop structures are indicated as filled arrows. The experimentally determined minimal regions required for SSO activity are indicated with open triangles (pLS11), closed (pBAA1) triangles, and vertical arrows (pTA1060). Conserved motifs (see text) are shaded. The sequence of the pLS11 palTI is not shown because it is nearly identical to that of pTA1060. The highly conserved region extends from coordinates 74 to about 270.

ously, we have reported the cloning and sequencing of the SSOs of pTA1015, pTA1040, and pTA1060 using a specially designed vector, pWM100 [26]. The results showed that whereas the SSOs of pTA1015 and pTA1060, designated palT1, are almost identical to each other, the SSO of pTA1040, designated palT2, is less homologous (77%; see Fig. 4). Despite the differences between palT1 and palT2 both are highly efficient ssDNA conversion signals in *B. subtilis* [26]. DNA sequences of SSO-containing regions of the following plasmids have been published: pBAA1 (pTA1020 group; [28,42]); pLS11 (pTA1060 group; [41]), and pUH1 (pTA1015 group; [52]). The SSO of pBAA1 is almost identical to those of pTA1015 and pTA1060 and, therefore, belongs to the palT1-type. As expected, pLS11 (pTA1060 group) and pUH1 (pTA1015 group) also contain an SSO of the palT1-type [26]. To study the nature of the SSOs of pTA1030 and pTA1050, the six representative pTA plasmids were compared in Southern hybridization studies using either palT1 of pTA1015 or palT2 of pTA1040 as a probe under stringent conditions. The results showed that pTA1030 and pTA1050, like pTA1040, contain a palT2-type SSO [26]. This conclusion has now been confirmed by sequencing pTA1050 which showed that pTA1050 contains a palT2-type SSO (Fig. 4). This indicates that all plasmids in the groups represented by pTA1015, pTA1020, and pTA1060 contain the palT1-type SSO, whereas plasmids represented by pTA1040, pTA1030, and pTA1050 contain a palT2-type SSO.

5.2. Delineation of palT SSOs

The alignment of the SSO-containing regions of

pTA1015, pBAA1 (pTA1020 group), pTA1060, pTA1040, and pTA1050 reveals a highly conserved region of about 200 bp (Fig. 4, coordinates 74 to about 270). This suggests that this region constitutes the functional SSO. This idea is supported by the fact that this region encompasses the experimentally determined minimal sequence required for full SSO activity determined for pTA1060 [53]. The experimentally determined SSO sequences of pLS11 (pTA1060 group; [41]), and pBAA1 (pTA1020 group; [42]) are about 30 bp shorter at the 3'-end. Since the contribution of this additional 30-bp conserved region to SSO activity is also relatively low in pTA1060 [53], it may have been below the level of detection in the other plasmids studied.

Seery et al. [42] identified three motifs (marked as motif 1, 2, and 3 in Fig. 4) which are highly conserved between palT1 of pBAA1 and the putative SSO of pGI2, a plasmid isolated from *Bacillus thuringiensis* [54]. Madsen et al. [55] noticed that the same motifs were also highly conserved in the SSO of plasmid pTX14-3, another endogenous plasmid from *B. thuringiensis*. Within motifs 1 and 2, only two basepairs differ between palT1 of pBAA1 and the corresponding sequences of pGI2 and pTX14-3. Interestingly, these bases are conserved between palT2 of pTA1040/pTA1050 and pGI2 and pTX14-3 (bp 94 and 109 in Fig. 4).

Whereas in pTA1015, pTA1020, and pUH1 (pTA1015 group), the SSO and DSO sequences are located in close proximity, these primary and secondary replication origins are separated in pTA1040 and pTA1060 by several kilobasepairs (Fig. 1). We showed that the conversion of ssDNA occurred with high efficiency in pTA1015, pTA1040 and pTA1060 [26]. This corroborates the general view that the po-

Table 3
Mobilization frequencies

Plasmid	Genotype			Frequency ^a
	ORF2C	ORF1	<i>mob15</i> ^b	
pTAB11A/pTAB11B	–	+	+	8.4×10^{-5}
pTAB13	–	+	–	$< 1 \times 10^{-9}$
pTAB31	+	–	+	9.0×10^{-6}
pUB110	Absent	Absent	– ^b	2.9×10^{-4}

^aFrequencies are given as the number of transconjugants per donor cell and represent the mean of five independent experiments. (–) denotes interruption of the ORF by the *Km*^R gene (ORF2c, ORF1), or deletion of an internal fragment from the ORF (*mob15*).

^bInstead of *mob15*, pUB110 carries another *mob* gene enabling conjugative transfer.

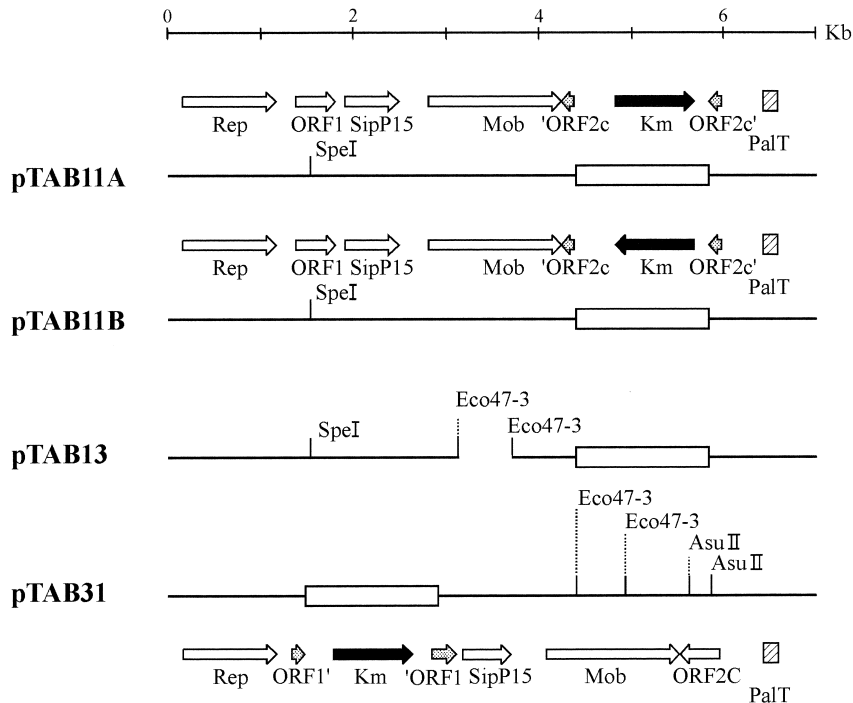


Fig. 5. Linear maps of the pTA1015 derivatives pTAB11A, pTAB11B, pTAB13, and pTAB31. Plasmids pTAB11A and pTAB11B were constructed by replacing the 0.24-kbp *AsuII* fragment of pTA1015, located in the ORF2 reading frame, by the 1.4-kbp *AclI* fragment of pKM1, which contains the Km^R gene from the *Streptococcus faecalis* plasmid pJH1. The difference between pTAB11A and pTAB11B is the orientation of the insert. pTAB13 was obtained by deleting the 0.5-kbp *Eco47^{III}* fragment from pTAB11A, located internally in the *mob* gene, which resulted in a frameshift at the *Eco47^{III}* fusion site (the first 115 codons are still intact and a stop codon is present after 170 codons). pTAB31 was constructed by cloning the 1.4-kbp *XbaI* fragment of pKM2, containing the Km^R gene, into the unique *SpeI* site in ORF1 of pTA1015. In the latter construct the direction of transcription of the Km^R gene is the same as that of the *rep* gene. Relevant features of pTAB11A/B and pTAB13 are shown above the linear maps; those for pTAB31 below the map. Genes are indicated with arrows (according to the relative sizes of the genes). The inserts containing the Km^R gene are indicated as open rectangles. Positions of relevant restriction sites are indicated. Nomenclature of the genes and ORFs is as in Fig. 1. The region deleted from pTAB13 is indicated as a gap. Interrupted ORF2c (pTAB11A/B, pTAB13) and ORF1 (pTAB31) are indicated with shaded arrows. The top line in the figure indicates sizes (in kbp).

sition of SSOs relative to their DSOs usually does not affect their functionality.

6. Mobilization modules (pTA1015 and pTA1060)

6.1. Absence of γ -gtp genes (pTA1015 and pTA1060)

γ -Glutamyl transpeptidase (γ -GTP) catalyzes the hydrolysis of glutathione to glutamic acid and the transfer of the γ -glutamyl group of glutathione to an amino acid or peptide [56,57]. Hara et al. [9,33] reported that some *B. subtilis* strains produce active γ -GTP and claim to have identified a γ -gtp gene on

plasmid pUH1 (pTA1015 group [43]). It is doubtful, however, whether the latter claim is correct. First, the proposed γ -gtp gene on pUH1 has neither significant homology with the *B. subtilis* SJ138 chromosomal γ -gtp gene [57], nor with that of *E. coli* [58]. Second, pTA1015 and pTA1060 encode a putative protein which is almost identical to that of the postulated γ -GTP protein of pUH1. This protein, indicated as Mob in Fig. 1, is required for the conjugative mobilization of pTA1015 and pTA1060 (see Section 6.2).

To prove that the *mob* genes of pTA15 and pTA1060 do not specify γ -GTP, the pTA1015 derivatives pTAB11A, pTAB11B, pTAB31 and pTAB13

A

motif III
★

```

pTA1015  MANVAVIRMERKRYKRD-RLNNGTQKHNQRERFQKSKNENIDRERTHLNYDLVN-ERKPTSYSKATHEKTEGR--VK-R 69
pTA1060  MANVAVIRMERKRYKRD-RLNNGTQKHNQRERFQKSKNENIDRERTYLNLDLVN-ERPIISYSKAVHHEKIEGR--VK-R 69
pGI2     MNKFAIITHMQKFIQIS-DVQGIQKHNQRQGKSKNLDIDYFQSKSEQNYDDLN-QQKIRYESTIKQEIISER--VK-R 69
pTB913  MS-FAVVRMOKMKSY-DLKGIOFHNRERESKTNMEDIKERSHENYDLVN-DEPIDYNERVVKKEIIESQKVG-TR 71
pTB19   MS-YAVCRMQKVKSA-GLKGMQFHNQRERKSRNTDDIDHERTRBNYDLKRN-DKNIDYNERVVKKEIIESQKTG-TR 71
pUB110  MS-YAVCRMQKVKSA-GLKGMQFHNQRERKSRNTDDIDHERTRBNYDLKRN-DKNIDYNERVVKKEIIESQKTG-TR 71
pKH6    MS-YSIVRVSKVKSGTNTTGIQKHVQRENNNYENEDIDHSKTYLNLDLVN-ANKQNFNNLIIDEKIEQNYTGKR-R 71
pT181   MS-YSIVRVSKVKSGTNTTGIQKHVQRENNNYENEDIDHSKTYLNLDLVN-ANKQNFNNLIIDEKIEQNYTGKR-R 71
pE194   MS-HSILRVARVKGSSNTNGIQRHNQRBNKKNYNNKIDINHETTYKNYDLIN-AQNKKYKDKIDETIDENYSQK-R 71
pLC88   MS-YLVANMQKVKAD-NLVGLGNHDDQRRTQHKKNPDIIVDRSALNYDLVA-GRTHGFKTDIEAYINENHKTSG-R 70
pLB4    MS-FAVARMTKCLKAD-NLVGIGNHDDQRKTNNHSEDIIVSHSLNYDLVA-GRTDNFKTDIEAYINENKASK-R 70
pAB1000 MS-FAVARMTKCLKAD-NLVGIGNHDDQRKTNNHSEDIIVSHSLNYDLVA-GRTDNFKTDIEAYINENKASK-R 70
pVA380_1 MS-YMVARMQMKKAG-NLGGAFKHNERRVFEETHSNKIDINPSSRSHLNYELTDRDRSVSYEKQIKKDYVMENKVSNT-R 72
pMV158  MS-YMVARMQMKKAG-NLGGAFKHNERRVFEETHSNKIDINPSSRSHLNYELTDRDRSVSYEKQIKKDYVMENKVSNT-R 71
    
```

motif II

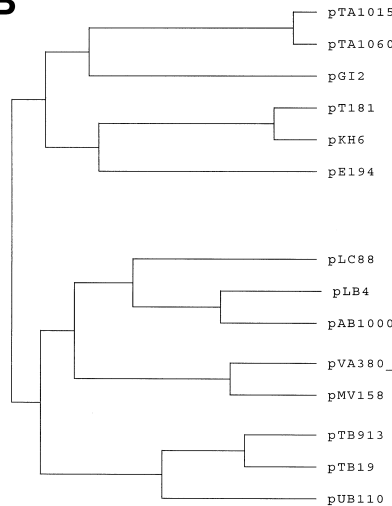
```

pTA1015  KVRADAVLVSEFLITASPDYMNGLSDEEQRRYFETAVDHLKEKYSANMLYATVHMDEATPHMHGIVPITIEDG 143
pTA1060  KVRADAVLVSEFLITASPDYMNGLSDEEQRRYFETAVDHLKEKYSANMLYATVHMDEATPHMHGIVPITIEDG 143
pGI2     KPRANSVVLSEFVVTA SPDYMHSLSLEREQKRYFESSLDPIQKRYGKQNTLYAMVHMDEATPHMHGIVPITIEDN 143
pTB913  KTRKDAVLVNEELIVTSDRHFFRLDPPDEQKRFFEEESYKLFADRYGKQNTLYATVHVDKTPPHMHLGVVPMRDG- 143
pTB19   KTRKDAVLVNEELIVTSDRDFFEQLDPEEQKRFFEEESYKLFSERYGKQNTLYATVHNDEQTPPHMHLGVVPMRDG- 143
pUB110  KTRKDAVLVNEELIVTSDRDFFEQLDPEEQKRFFEEESYKLFSERYGKQNTLYATVHNDEQTPPHMHLGVVPMRDG- 143
pKH6    KIRTDALKHIDGLITSDNDFFDQTPEDTKQFFEYAKFLEQYEGKDNLLYATVHMDEKTPPHMHGIVVPIITDDG 145
pT181   KIRTDALKHIDGLITSDNDFFDQTPEDTKQFFEYAKFLEQYEGKDNLLYATVHMDEKTPPHMHGIVVPIITDDG 145
pE194   KIRSDAIRHVDGLVTSDDFFDQLSGEETERFPKDSLEFLENEYGKENMLYATVHLDERVPHMHGIVVPIITEDG 145
pLC88   AVRKDAVLVNEWIISDSHFADLTAADMRRKYFETAQDYFAEKFGENIRYAVIHLDESTPHMHGIVPFDDDEH 144
pLB4    AVRKDAVLVNEWIITSDKDFEQLDDEAETRRKYFETAQDYFADNYG DENIRYAVVHMDEKTPPHMHGIVPFDDDK 144
pAB1000 AVRKDAVLVNEWIITSDKDFEQLDDEAETRRKYFETAQDYFADNYG DENIRYAVVHMDEKTPPHMHGIVPFDDDK 144
pVA380_1 AIRKDAVLCDIEWIITSDKDFEQLDDEBETRAFFETAKNYFAENYGESN IAYASVHLDESTPHMHMGVVPFENG- 145
pMV158  AIRKDAVLCDIEWIITSDKDFEQLDDEBQTRTFETAKNYFAENYGESN IAYASVHLDESTPHMHMGVVPFENG- 144
    
```

```

pTA1015  RLSAKDFPFGKRLKMKAIQDDFHRHMVVENGFDLVIRGEPSE---KKHENVHQYKINQR-Q-AELE 201
pTA1060  RLSAKDFPFGKRLKMKAIQDDFHRHMVVENGFDLVIRGEPSE---KKHENVHQYKINQR-Q-AELE 201
pGI2     RLSAKDMFTRK-ELISLQDDFPLEMRERKGFDFVDRGEGSE---KKHLSQAFKFKQDLE-VEVE 201
pTB913  KLGKKNVFN-RQELLWLQDKFPPEHMKKQGFELKRGGERG---SDRKHIEETAKFKKQT--LEKEID 201
pTB19   KLGKKNVFN-RQELLWLQDKFPPEHMKKQGFELKRGGERG---SDRKHIEETAKFKKQT--LEKEID 201
pUB110  KLGKKNVFN-RQELLWLQDKFPPEHMKKQGFELKRGGERG---SDRKHIEETAKFKKQT--LEKEID 201
pKH6    RLSAKEVVGNKKALTAFODRFNEHVKQRYGDLERGSQSVTNAKHHEQISQYKQKTEYHKQEYVE 208
pT181   RLSAKEVVGNKKALTAFODRFNEHVKQRYGDLERGSQSVTNAKHHEQISQYKQKTEYHKQEYVE 208
pE194   RLSAKEQLGNKKIDFTQLQDRFNEVYVNEKGYELERGTSEKVTTERHKKAMDQYKDKDTVPHKQELQ 208
pLC88   RLSAKRVFN-RTALRDVQDQLFAYLPOHGLNIORGVOE---SGRKS LAVPGYK-----AMR 196
pLB4    RLSAKRIFN-REALQHIQEELPOYLKENGFDVQRGNKN---KERKNLSVPYK-----AMR 196
pAB1000 RLSAKRIFN-REALQRIQEELPOYLKENGFDVQRGNKN---KERKNLSVPYK-----AMR 196
pVA380_1 KLSKSMFD-REELKHIQEDLPRYMSDHFELERKLN---SEAKHKKTVAEFKRAM--ADMELK 203
pMV158  KLSKSMFD-REELKHIQEDLPRYMSDHFELERKLN---SEAKHKKTVAEFKRAM--ADMELK 202
    
```

B



were constructed, which contain either an intact (the first three constructs) or a disrupted *mob* gene (the latter construct; Fig. 5). In these constructs, genes flanking the *mob* gene were also inactivated by disruption. These derivatives were introduced into *B. subtilis* strain BD630 and γ -GTP assays of growing cells were carried out. Various media were tested: Luria broth, minimal medium, sporulation medium [59] and DGYP (described to be optimal for γ -GTP assays [57]). As a positive control, commercially available γ -GTP was used. Whereas γ -GTP activity was clearly detectable with the commercial enzyme, no γ -GTP activity was detectable in any of the *B. subtilis* samples tested (cell fractions and culture supernatants were assayed as a function of growth time). Koehler and Thorne [60] have reported that pLS19, another plasmid belonging to the pTA1015 group, was also unable to endow its host with γ -GTP activity. Together, these observations make it highly unlikely that the region of pUH1 analyzed by Hara et al. [43] encodes a γ -GTP enzyme.

6.2. Functional *mob* genes (pTA1015 and pTA1060)

The N-terminal parts of the deduced protein sequences of *mob* genes in pTA1015 and pTA1060 reveal significant homology with corresponding regions deduced from *mob* genes present on several RCR plasmids (Fig. 6A). In several cases these genes have been shown to be required for conjugative mobilization [61–64]. To study the function of the pTA1015-located *mob*, the constructs pTAB11A, pTAB11B, pTAB13 and pTAB31, described in the foregoing section, were assayed for their ability to

be mobilized by the conjugative plasmid pLS20. The latter enables co-resident RCR plasmids, like pUB110, pTB913, pBC16, and pLS19 (pTA1015 group), to be transferred to other cells during matings [60,61]. pTAB11A, pTAB11B, pTAB13, pTAB31, and pUB110 (positive control) were introduced into *B. subtilis* strain 3335 UM4, which contains pLS20, and the resulting strains were used as donors in matings with the chloramphenicol-resistant *B. subtilis* strain 1012Cm. The results of the matings (Table 3) demonstrated that pTA1015 derivatives can be mobilized in a process requiring the intact *mob* gene. A derivative of pTA1060, pBB2, in which the homologous ORF is intact, was also shown to be mobilizable (results not shown). Together, these results show that these ORFs of pTA1015 and pTA1060 are required for mobilization. The corresponding genes were designated *mob15* and *mob60*, respectively. These are the first mobilization genes from endogenous *B. subtilis* plasmids which have been sequenced and functionally analyzed.

The results presented in Table 3 show that the frequency of mobilization was significantly higher (almost ten-fold) with pTAB11A (interrupted ORF2C) than with pTAB31 (intact ORF2C). The possibility that readthrough transcription activity from the promoter of ORF2C in pTAB31 interferes with the expression of the convergently transcribed *mob* gene is unlikely because pTAB11A and pTAB11B, which differ only by the orientation of the cloned Km^R gene in ORF2C, showed similar mobilization frequencies. Since neither clear differences in the copy numbers of these plasmids, nor obvious differences in growth rate or number of col-

←
 Fig. 6. Comparison of Mob and Pre proteins. EMBL/GenBank/DBJ accession numbers are given between square brackets. The *S. aureus* plasmids pT181 [J01764] and pE194 [V01278] encode a Pre protein [69], and the other plasmids a Mob protein. pKH6 ([U38428] [107]), which is nearly identical to pNS1 ([M16217] [108]), and pUB110 ([M19465] [70]) originate also from *S. aureus*. pLC88 ([U31333] [109]), pLB4 ([M33531] [110]), and pLAB1000 ([M55222] [111]) originate from *Lactobacillus casei*, *Lactobacillus plantarum* and *Lactobacillus hilgardii*, respectively. pMV158 ([X15669] [64]) and pVA380-1 ([L23803] [112]) originate from *Streptococcus agalactiae* and *Streptococcus ferus*, respectively. pGI2 ([X13481] [54]) is from *B. thuringiensis*, and pTB19 [M63891] and pTB913 ([M63891] [70]) from a thermophilic *Bacillus* strain. A: Alignment of the N-terminal 200 amino acids. Conserved amino acids are boxed when present in at least seven of these proteins. The two conserved amino acid motifs, which are also present in replication initiation proteins for rolling-circle replication (see text), are indicated with filled bars. B: Dendrogram of various Mob/Pre proteins. The dendrogram is based on the 'average linkage cluster analysis' described by Sneath and Sokal [106]. pT181 and pE194 encode a Pre protein; the other plasmids a Mob protein. The pairwise levels of identity of the Mob/Pre proteins shown vary from high to moderate. Examples of levels of identity between closely related Mob/Pre proteins are: pTA1015–pTA1060: 95%; pLC88–pAB1000: 59%; pVA380/1–pMV158: 85%; pTB913–pUB110: 60%. Examples of such levels between more distantly related Mob/Pre proteins are: pTA1015–pT181: 34%; pTA1015–pLC88: 31%; pTA1015–pUB110: 34%. In none of the pairwise comparisons, levels of less than 26% identity were obtained.

		→		←	
<i>S. aureus</i>	pT181	t a a a a a t a a g t C t a	g t g t g t	t a G a c t t a a a	60
<i>S. aureus</i>	pKH6	t a a a a a t a a g t C t a	g t g t g t	t a G a c t t a a a	60
<i>S. aureus</i>	pE194	a g t a a a t a a g t C t a	g t g t g t	t a G a c t t t t a t	60
<i>S. aureus</i>	pT48	a g t a a a t a a g t C t a	g t g t g t	* t a G a c t t t t a a	-
<i>S. epidermis</i>	pNE131	t g a a a a t a a g t C t a	g t g t g t	t a G a c t t t t a a	-
<i>L. lactis</i>	pFX2	a g t a a a t a a g t C t a	g t g t g t	t a G a c t t t t a t	-
<i>S. aureus</i>	pUB110	g a a t a t a a a g t A t a	g t g t g t	t a T a c t t t t a c	84
<i>L. casei</i>	pLC88	c a a g g t a a a g t A t a	t t g g g c	t a T a c c t t t a c	81
<i>L. plantarum</i>	pLB4	c g a a g t a a a g t A t a	g t g g g t	t a T a c t t t t a c	47
<i>L. hilgardii</i>	pLAB1000	t a a c a t a a a g t A t a	g t g g g t	t a T a c t t t t a c	47
<i>S. ferus</i>	pVA380_1	c t a t g t a a a g c t a	g t a c g t	t a C a c t t t t a c	140
<i>S. agalactiae</i>	pMV158	g a a t a t a a a g t A t a	g t g t g t	t a T a c t t t t a c	140
tp <i>Bacillus sp.</i>	pTB913	g a a t a t a a a g t A t a	g t g t g c	t t a T a c t t t t a c	68
tp <i>Bacillus sp.</i>	pTB19	g a a t a t a a a g t A t a	g t g t g t	t a T a c t t t t a c	85
<i>B. thuringiensis</i>	pGI2	t g t a a t t t g g t A t a	a t g g g c	t a T a c c a a a a	93
<i>B. subtilis</i>	pTA1015	c a t a a t t t g g t G t a	g t g c g t	t a C a c c a a a g	71
<i>B. subtilis</i>	pTA1060	c a t a a t t t g g t G t a	g t g c g t	t a C a c c a a a g	71

Fig. 7. *OriTRS_A* sites. An alignment of *RS_AoriT* sites is shown for the *S. aureus* plasmids pE194 [69], pT181 [69], pT48 [71], and pUB110 [70], the *S. epidermis* plasmid pNE131 [113], the *L. hilgardii* plasmid pLAB1000 [111], the *L. plantarum* plasmid pLB4 [110], the *L. casei* plasmid pLC88 [109], the *S. agalactiae* plasmid pMV158 [70], the *S. ferus* plasmid pVA380-1 [112], the *L. lactis* plasmid pFX2 [73], the *B. thuringiensis* plasmid pGI2 [54], the thermophilic (indicated as tp) *Bacillus* plasmids pTB913 and pTB19 [70], and the *B. subtilis* plasmids pTA1015 and pTA1060. The inverted repeat is indicated with arrows above the sequences. Although some differences in the sequence of this region exist, the dyad symmetry in all these sites is maintained in most cases (non-complementary nucleotides in the stems are boxed). The consensus sequence of the loop in the potential stem-loop structure is 5'-GTGTGT-3'; deviations from this sequence are boxed. The distance (in nucleotides) from the first 'G' of this consensus loop-sequence, or the corresponding nucleotide, to the ATG-start codon of the *pre/mob* gene is indicated in the right column. The asterisk indicates the position of two T-residues in the pT48 sequence that are omitted for optimal alignment. The nucleotides located at the third position relative to the loop are shown in capital letters.

ony-forming units per OD₆₀₀ were observed between the various plasmid-containing *B. subtilis* cultures, it is unlikely that these factors underlie the observed differences in mobilization frequencies. An alternative explanation, which implies a direct effect of the gene product of ORF2C on mobilization frequencies, is described in Section 7.

6.2.1. Relatedness of Mob proteins

To obtain insight in the relatedness of the Mob

proteins encoded by pTA1015, pTA1060, and other RCR plasmids, their deduced amino acid sequences were compared. The Mob proteins of pTA1015 and pTA1060 are highly homologous to each other (95% identity), and somewhat more distantly related to those of the other groups (see also legend to Fig. 6B). Although the homology between all sequences of the 14 Mob proteins shown here is rather low (19% identity), pairwise identities are considerably higher (legend to Fig. 6B). Also, much higher levels

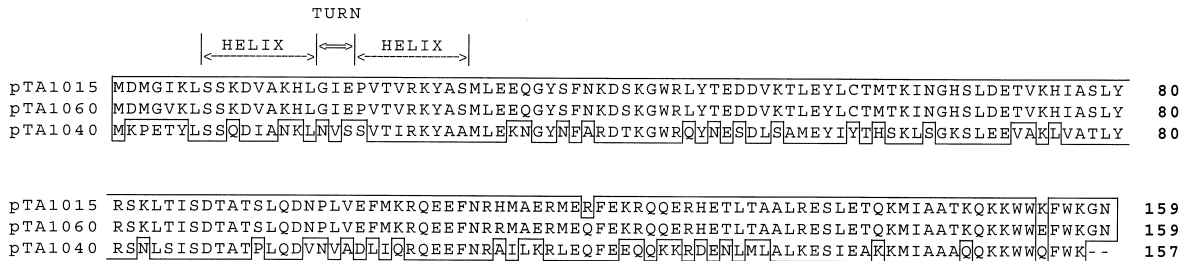


Fig. 8. Alignment of the deduced protein sequences of ORF2C from pTA1015, pTA1040 and pTA1060. Conserved amino acid residues are boxed. The position of the helix-turn-helix motif is indicated.

of identity are observed within the first 200 amino acids of these proteins (46% identity), suggesting that this may constitute an important functional region of the proteins. A schematic overview of the relatedness between the various Mob proteins is presented in the dendrogram of Fig. 6B.

6.2.2. Conserved modules in Mob proteins

Koonin and Ilyina [47,65] compared the deduced amino acid sequences of a considerable number of proteins involved in plasmid transfer. This comparison included, in addition to the mobilization proteins encoded by several RCR plasmids of Gram-positive bacteria, also proteins involved in conjugation that are encoded by plasmids from Gram-negative bacteria (including virulence proteins specified by the *Agrobacterium tumefaciens* pTi plasmid). Interestingly, two of the three conserved amino acid sequence motifs identified by the same authors [47,66] in the replication initiation proteins of RCR plasmids (motifs II and III in Fig. 2A) were also present in proteins involved in plasmid transfer, albeit in inverse orientation (i.e. the motif containing the putative DNA-binding tyrosine residue of motif III is located N-terminally from motif II). These two motifs are also conserved in Mob15 and Mob60 (Fig. 6A). The observation that replication initiation proteins of RCR plasmids and proteins involved in plasmid transfer share amino acid motifs suggests a functional similarity. Most likely, the basis for this similarity is that in both cases the mechanism of DNA transfer involves a rolling-circle-like mechanism. Recently, an alignment of the regions encompassing motifs II and III present in various Mob proteins was also published by Guzmán and Espinosa [67].

6.3. *OriT/RS_A* sites

pE194 [60,68] and pT181 [68] specify proteins with N-termini showing strong homology with the N-termini of Mob proteins (Fig. 6A). The corresponding genes are involved in site-specific, *recA*-independent, plasmid cointegrate formation [69] and were named *pre* (for: plasmid recombination enzyme). Pre proteins act at specific recombination sites (called RS_A), which overlap with the -10 promoter sequence of the *pre* genes [69]. Sequences similar to RS_A were detected upstream of all *mob/pre* genes shown in Fig. 7. Selinger et al. [63] showed that a plasmid containing the RS_A site of pUB110 can be mobilized by the pUB110 Mob protein provided *in trans*. This indicates that, at least in this case, the RS_A -like sequence is the target site for the Mob protein. These sites have been called *oriT* [63]. Recently, Guzmán and Espinosa [67] have demonstrated unambiguously that the Mob protein of the streptococcal plasmid pMV158, MobM, attacks the 5'-GpT-3-dinucleotide within its RS_A -like sequence 5'-TAGTGTG/TTA-3'. A list of *oriT/RS_A* sites present on various RCR plasmids has also been provided in that paper [67].

Analysis of pTA1015 and pTA1060 revealed sequences that are highly homologous to known *oriT/RS_A* sites at a position 57 bp upstream of the potential start codons of the *mob* genes. These sequences, which are identical in these plasmids, are likely to constitute the target sites for the pTA1015- and pTA1060-encoded Mob proteins. Like other *oriT/RS_A* sites [63,69,70], the putative *oriT* sites of pTA1015 and pTA1060 contain an inverted repeat (Fig. 7). The stem of the possible stem-loop structure of $RS_A/oriT$ sites varies from 7 to 10

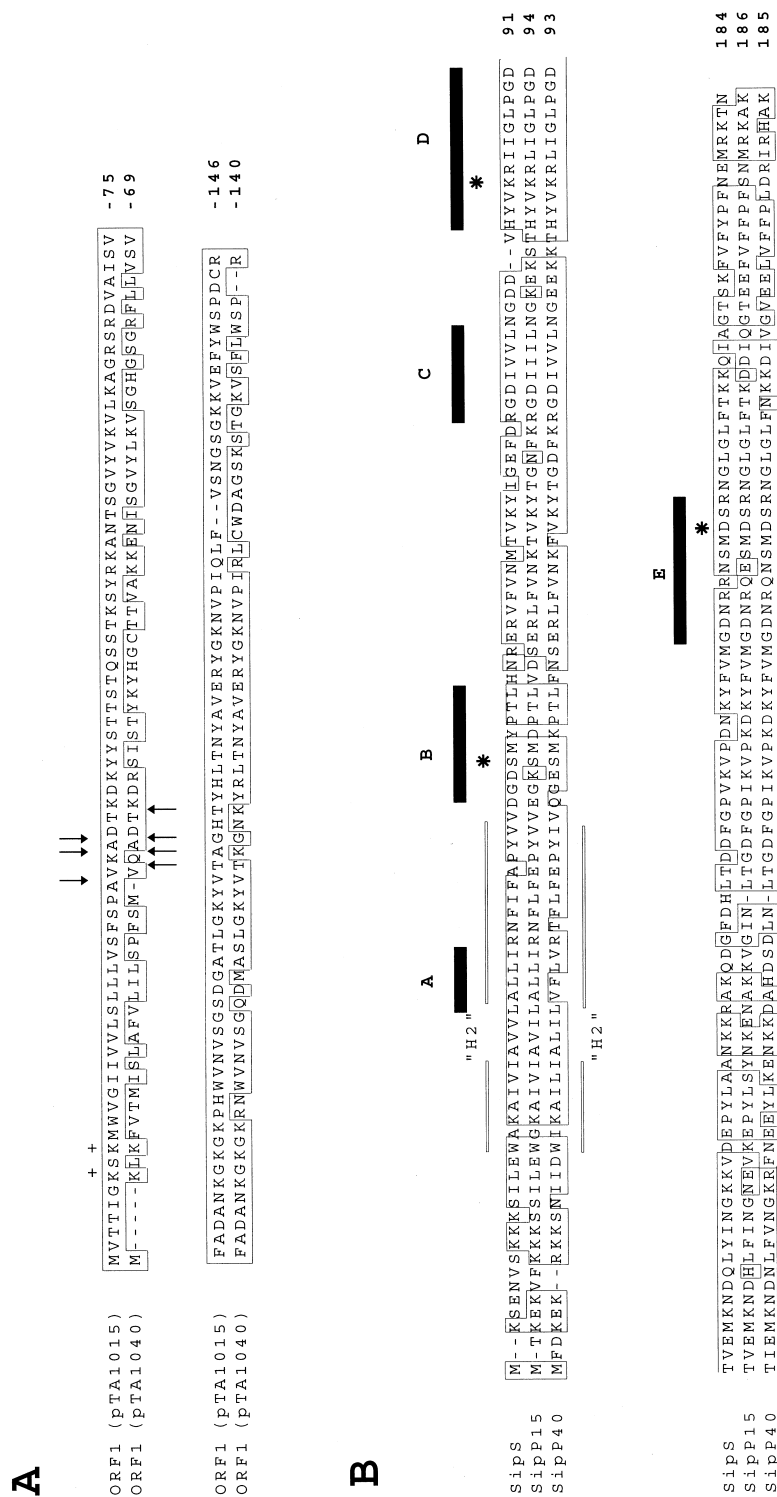


Fig. 9. ORF1/*sipP* modules on pTA1015 and pTA1040. Identical amino acids or conserved changes are boxed. A: Comparison of the ORF1-encoded proteins. Positively charged residues (+) in the putative signal peptide n-regions are indicated. Potential SPase I cleavage sites, indicated by arrows, were predicted using algorithms of von Heijne [114]. B: Comparison of plasmid-specified SipP proteins and the chromosomally-encoded SipS. The comparison includes (EMBL/GenBank/DDBI accession numbers in each of square brackets): SipP (pTA1015) ([L26258] [74]); SipP (pTA1040) ([Z36269] [74]) and SipS (Bsu) of *B. subtilis* ([Z11847] [75]). A putative transmembrane segment in each of these proteins was predicted according to von Heijne [115]. By analogy to the corresponding membrane spanning domain of Lep of *E. coli*, these were denoted 'H2' [87]. Conserved regions A, B, C, D, and E, which are present in all known type I SPases of prokaryotic and eukaryotic origin [75], are indicated. Conserved residues (Ser⁴³, Lys⁸³, and Asp¹⁸³), which are important for the activity of SipS [89], are marked (*).

nucleotides and the loop usually consists of 6 nucleotides. Based on the complementary nucleotide pairs located at the third position relative to the loop sequence, the $RS_A/oriT$ sites can be classified in the three groups indicated in Fig. 7. Although direct evidence is lacking, it is conceivable that these differences are important for the correct recognition by the corresponding Mob proteins.

Since sequences highly homologous to $oriT/RS_A$ sites are also present on plasmids pT48 [71], pNE131 [72], and pFX2 [73], which do not contain *mob* or *pre* genes, we screened pTA1040 and pTA1050 (which, likewise, do not contain a *mob* gene) for the presence of sequences homologous to $oriT/RS_A$. Such sequences were not identified on these plasmids.

The high level of homology in the Mob-encoding regions and $oriT$ of pTA1015 and pTA1060 extends for approximately 150 bp further upstream of the *mob* genes. These homologous sequences have the potential to form secondary structures, but a possible function of these regions is not known.

7. ORF2C modules (pTA1015, pTA1040 and pTA1060)

Except pTA1050, the pTA plasmids contain a highly homologous ORF2C with an orientation of transcription opposite to that of most other ORFs/genes on these plasmids (Fig. 1). In pTA1015 and pTA1060, the stop codons of the convergently oriented *mob* gene and ORF2C are separated by only 10 and 21 basepairs, respectively. These short intergenic regions are part of sequences that have the potential to form secondary structures which could function as transcriptional terminators. An alignment of the deduced ORF2C protein sequences is shown in Fig. 8. A potential DNA-binding helix-turn-helix motif is present in the N-terminal part of these proteins. Such motifs are often present in bacterial transcription regulator proteins. Since in at least two of these plasmids (pTA1015 and pTA1060) ORF2C is located adjacent to the *mob* genes, and interruption of ORF2C of pTA1015 resulted in increased mobilization frequencies (Table 3), we considered the possibility that the ORF2C products might be negative regulators of *mob* gene

expression. Support for this idea was not obtained, however, from an analysis on the effects of intact and interrupted ORF2C on β -galactosidase expression driven by the promoter of the *mob* gene of pTA1015 (results not shown). Interruption of ORF2C had no clear effect on plasmid copy number or plasmid maintenance (results not shown). Therefore, the biological function of the ORF2C remains obscure. In pTA1040 ORF2C is convergently transcribed with the ORF1/*sipP* module (see Section 8). We have no indications that the expression of this module is affected by ORF2C.

8. ORF1/*Sip-P* modules (pTA1015 and pTA1040)

8.1. *Orf1* encodes a putative export protein

Plasmids pTA1015 and pTA1040 contain a homologous region encompassing ORF1 and *sipP* [74]. An alignment of the deduced ORF1 products, which show 56% identity, is presented in Fig. 9A. Since they contain a putative signal peptide with a positively charged n-region, a hydrophobic h-region, and a c-region containing several potential type I signal peptidase (SPase I) cleavage sites, it is likely that these products are export proteins. These proteins show no similarity to previously identified proteins and they have a remarkably high calculated isoelectric point (10.4 and 10.7, respectively). Despite intensive searches, we have not been able to detect ORF1-encoded proteins in the growth medium, cytoplasmic membranes, or unfractionated cellular extracts of *B. subtilis* cells harboring pTA1015 or pTA1040. Consequently, the biological function of ORF1-encoded proteins remains obscure.

8.2. Analysis of *SipP*

Interestingly, downstream of ORF1, pTA1015 and pTA1040 another homologous ORF is present [74], which is highly homologous to the chromosomally-located *sipS* gene of *B. subtilis* [denoted: *sipS* (Bsu)], which encodes a type I signal peptidase (SPase, [75]). Type I SPases are membrane-bound proteases which specifically remove signal peptides of export proteins. For reviews on protein secretion in *B. subtilis* we refer to Simonen and Palva [76]; van Dijk et al.

[77]; and Nagarajan [78]. SPases have been reviewed by von Heijne [79,80].

An alignment between SipS (Bsu) and the homologous products of pTA1015 and pTA1040 is shown in Fig. 9B. It should be realized that, in addition to SipS and the two plasmid-specified SPases, at least four other type I SPases, SipT, SipV, SipW [81] and SipU [82], are encoded by chromosomal *B. subtilis* genes. All these SPases are related but slightly different from each other [81].

Functional activity of the pTA1015- and pTA1040-encoded SPase homologues was demonstrated in vivo, in both *B. subtilis* and *E. coli*, using a hybrid β -lactamase precursor that is processed by SipS(Bsu) but not by its homologue, the leader peptidase of *E. coli* [74,75,83]. The plasmid-located genes were named *sipP* (pTA1015) and *sipP* (pTA1040); *sipP* stands for signal peptidase of plasmid.

Several features are shared by *sipS* and the plasmid-encoded *sipP* genes. First, all three *sip* genes possess an atypical start codon (TTG for *sipS* and *sipP* [pTA1015], and GTG for *sipP* [pTA1040]). Second, no sequences corresponding to the major classes of *B. subtilis* promoters could be detected upstream of their RBS. And third, inverted repeats are present downstream of all three *sip* genes that are potential rho-independent transcriptional terminators. The alignment presented in Fig. 9B shows that also on the amino acid level SipS and the plasmid-encoded SipP proteins have several features in common. All three SPases possess only one putative membrane spanning domain (indicated as 'H2' in Fig. 9B). In this respect, the Sip proteins are distinct from the leader peptidases of the Gram-negative bacteria *Pseudomonas fluorescens*, *Salmonella typhimurium*, and *E. coli* [84–86], which possess two membrane spanning domains [87] and are considerably larger (284, 324 and 323 amino acids, respectively) than SipS, SipP (pTA1015) and SipP (pTA1040).

Patterns of conserved amino acids are present in all known type I SPases of prokaryotic and eukaryotic origin [74,75,77,87,88]. These patterns are also conserved in the plasmid-encoded SPases (Fig. 9B). Finally, in SipS only three out of 30 amino acid residues which are conserved within all known type I SPases are important for functionality [89]. These

amino acid residues are also conserved in the two SipP enzymes (Fig. 9B).

8.3. Possible function of the ORF1/*sipP* modules

In pTA1015 and pTA1040 the homologous region comprising ORF1 and *sipP* extends from about 100 bp upstream of the start codon of ORF1 to 10 bp downstream of the stop codon of *sipP* (overall identity: 68%). The intergenic region between these genes is short (102 bp for pTA1015 and 124 bp for pTA1040) and contains an inverted repeat. Moreover, single transcription products have been identified from which both the ORF1 product and SipP can be translated, indicating that these genes are organized in an operon (H. Tjalsma, S. Bron, J.M. van Dijk, unpublished results). Together, these findings suggest that ORF1 and *sipP* form one structural module on pTA1015 and pTA1040 and that a functional relationship between the products of these genes may exist. Southern hybridization showed that pTA1020 is also likely to possess a *sip* gene but, based on Southern hybridization and PCR analyses, none of the plasmid-located *sip* genes are present on the *B. subtilis* chromosome (results not shown).

The fact that the ORF1/*sipP* module is present on several of the plasmids tested, suggests that this module may occur rather frequently on RCR plasmids of *B. subtilis* and may provide the cell with a specific advantageous trait. We can only speculate about this possible function. The pTA series of plasmids are all isolated from *B. subtilis* strains that are used for the production of natto, a traditional Japanese food product based on fermented soybeans. This process requires the production of large amounts of extracellular polymers, such as polyglutamate and levan. In particular plasmids of the pTA1015 group have been implicated in this process [9]. Since the ORF1-encoded proteins are likely to be exported from the cytoplasm, it is conceivable that these are directly or indirectly involved in the production of the extracellular polymers. A possible function of the plasmid-encoded SPases might be the processing of the ORF1-specified product, which may be a preferred substrate. In fact, substrate preference by SPases of *B. subtilis* does occur, as was

recently demonstrated for the chromosomally-encoded SipS of this bacterium [81,90].

9. Rap modules (pTA1040, pTA1050, and pTA1060)

9.1. Rap family of proteins

pTA1040, pTA1050, and pTA1060 contain an ORF showing significant homology to the chromosomally-located *rapA* and *rapB* genes of *B. subtilis* [91–93]. In the deduced amino acid sequences, the levels of identity amount to about 40%. In addition, we have identified a gene homologous to *rapA/rapB* on yet another plasmid from *B. subtilis*, pLS20, which replicates according to the theta mechanism [35]. We denote these genes from pTA1040, pTA1050, and pTA1060 as *rap40*, *rap50* and *rap60*.

The chromosomal *rapA* gene was initially called *gsiA* (glucose starvation induced protein A; [94,95]). Expression of this gene delays the onset of sporulation [95], and its gene product is a protein-aspartate phosphatase acting on phosphorylated SpoOF (SpoOF~P), a response regulator component of the phosphorelay system [92,93]. Hence, the gene was renamed *rapA* (response regulator aspartate phosphatase protein A). Dephosphorylation of SpoOF~P prevents the transfer of phosphate to SpoOA and lowers the level of SpoOA~P, which is a key factor in the induction of sporulation [96]. Like RapA, RapB is a protein-aspartate phosphatase active on SpoOF~P and it also delays the onset of sporulation [92].

Within the framework of the *B. subtilis* genome sequencing project [24], nine additional members of this gene family were recently identified [91,92,97,98], including *rapC*, *rapD*, *rapE*, *rapF* and *rapG*. An alignment of the deduced protein sequences of the eleven currently known Rap proteins is shown in Fig. 10A, and a dendrogram showing their relatedness in Fig. 10B. The RapA and RapB proteins are clearly highly related to each other (about 50% identity), as are the three plasmid-encoded proteins Rap40, Rap50, and Rap60 (pairwise identities 50% to 60%). The group of plasmid-specified Rap proteins showed the highest similarities with RapA and RapB from the chromosome (about 40% identity). The other chromosomally specified Rap pro-

teins show pairwise levels of identity ranging from 20% (RapD–RapE; RapD–RapG) to about 40% (examples: RapB–RapC; RapB–RapE; RapC–RapE).

9.2. Expression and function of Rap genes

Although, so far, phosphatase activity has only been demonstrated for RapA and RapB, it is tempting to speculate that all Rap proteins belong to a large family of phosphatases which may be expressed under different physiological conditions and may act on different targets. In addition to the demonstrated effects of *rapA* and *rapB* on the initiation of sporulation, *rapC* has been shown to act as a negative regulator of *srfA* expression [99], which is required for the initiation of competence development in *B. subtilis*. Conceivably, other members of the Rap family also have roles in the signal transduction pathways leading to, for instance, chemotaxis, competence, synthesis of secretory proteins, synthesis of antibiotic peptides, and sporulation. The plasmid location, resulting in increased copy numbers, of some of these genes may provide the cell with an additional tool for controlling these global gene expression systems, and extend its adaptative capacity to react to changing conditions.

Evidence that different *rap* genes, including the plasmid-located *rap40*, are expressed under different physiological conditions is already available. Whereas *rapA* is controlled by the ComP–ComA two-component signal transduction system [94], *rapB* seems to be regulated by the AbrB transition state regulator [92,100]. Therefore, the synthesis of these proteins is likely to be induced under different physiological conditions. Analysis in our group [101] showed that the regulation of *rap40* expression differs from that of the chromosomal *rapA* gene: whereas *rapA* is induced under starvation conditions, in particular glucose deprivation [94,95], similar conditions did not induce the *rap40* gene. Moreover, maximal expression of this plasmid-located gene occurred at the transition between the exponential and stationary phase, approximately 2 h before maximal expression of *rapA*.

9.3. ComA boxes

In response to nutrient exhaustion, several devel-

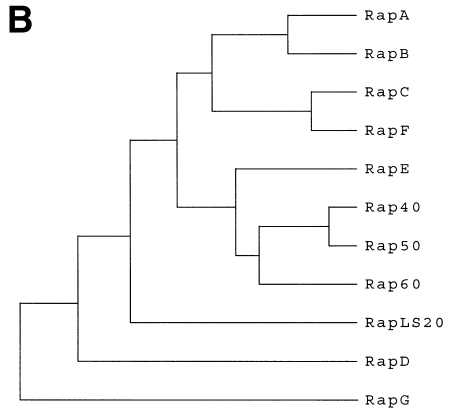


Fig. 10. Comparison of Rap proteins. The EMBL/GenBank/ DDBJ accession numbers are indicated between square brackets: RapA ([X56679] [94]); RapE ([D32216] [98]); RapG ([D78193] [116]); RapC [91]; RapLS20 ([U26059] [35]); Rap40 [U32378], Rap50 [U55043] and Rap60 [U32380] (this work); RapB, RapD and RapF [91–93,97]. Rap40, Rap50, Rap60, and RapLS20 represent the deduced protein sequences specified by the plasmid-located *rap* homologues of pTA1040, pTA1050, pTA1060, and pLS20, respectively; the other Rap proteins are specified by the chromosome. A: Alignment of deduced amino acid sequences. Conserved amino acid residues are boxed only when they are present in at least six of the eleven proteins shown. B: Dendrogram of Rap proteins. The dendrogram is based on the ‘average linkage cluster analysis’ described by Sneath and Sokal [106].

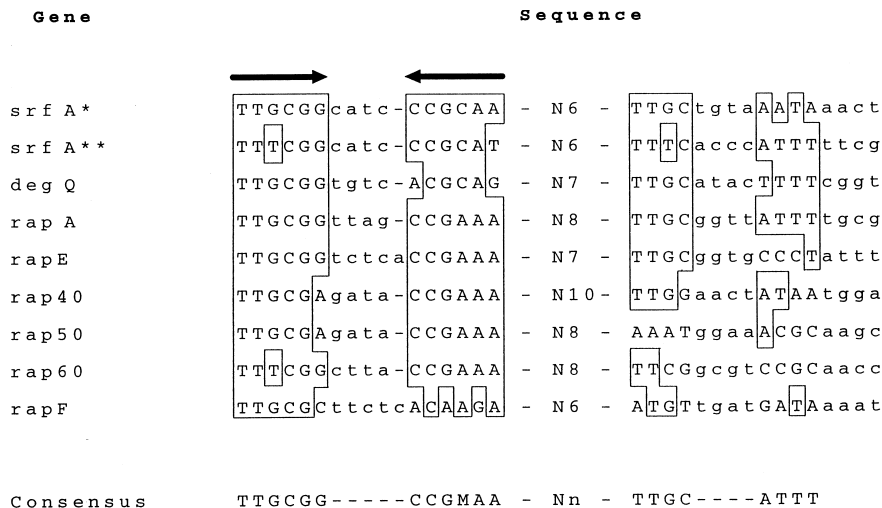


Fig. 11. Comparison of ComA boxes. Sequence information was taken from (EMBL/GenBank/DBJ accession numbers are given between square brackets): *degQ*, Msadek et al. [117]; *srfA*, Nakano et al. [118]; *rapA*, Mueller et al. [X56679] [94]; *rapE*, ([D32216] [98]); *rapF* [97], *rapG* ([D78193] [116]). Nucleotides identical to the consensus sequence shown at the bottom are boxed. Dyad symmetry is indicated with filled arrows. M: A or C residue. The two separate ComA-boxes upstream of the transcription start point of the *srfA* gene are marked as * and **, respectively. Distances of ComA boxes to transcription start sites are known in only a limited number of the genes shown here. These are (in bp): *srfA*, -73 and -117; *degQ*, -70; and *rapA*, -75. Start sites of transcription for the other genes indicated in this figure are not known. DNA sequences upstream of *rapB*, *rapD*, and *rapF* were kindly provided by P. Glaser.

opmental programs such as development of genetic competence, motility and chemotaxis, degradative enzyme synthesis, antibiotic production and sporulation can be initiated by the *B. subtilis* cell. The various interconnected programs are generally regulated by so-called two-component regulatory systems consisting of a histidine protein kinase and a cognate response regulator. Histidine kinases can undergo autophosphorylation by certain environmental or intracellular stimuli. Subsequently, the phosphate group is transferred to the cognate response regulator which, in many cases, modulates the activity of that protein to a transcriptional activator or repressor (for review see [102]). For the development of genetic competence the ComP and ComA proteins are important as histidine kinase and response regulator, respectively [103].

Like the expression of *rapA*, expression of *rapC* also depends on ComA (mentioned in reference [91]), and a putative ComA~P binding site (ComA-box) was identified upstream of the *rapA* gene [94]. Analysis of the DNA sequences upstream of other *rap*-like genes revealed sequences homologous to ComA~P binding sites for *rap40*, *rap50*, *rap60*, *rapE* and *rapF*.

An alignment of putative ComA boxes is shown in Fig. 11. The presence of ComA boxes upstream of the other mentioned *rap* genes suggests that they are also under control of the ComA-ComP system. In addition, we noticed that the consensus SpoOA~P recognition sequence, 5'-TGNCGAA-3', is present just upstream of the putative RBS of *rap40* and thus this gene may be under the control of SpoOA~P. This so-called 'OA box' is not present upstream of *rap60* or *rap50*.

9.4. Translationally coupled small genes

A small ORF, called *gsiAB* by Mueller et al. [94], which could encode a peptide of 44 amino acids, is located at the 3'-end of the *rapA* gene. Expression of this ORF is translationally coupled to that of *rapA*. Small translationally-coupled putative genes are also present at the 3'-ends of the members of the *rap* family shown in Fig. 12, including those on pTA1040, pTA1050, and pTA1060. The deduced protein sequences of these genes contain a potential signal peptide with appropriate SPase I cleavage sites (the small peptide associated with *rapB* could be an

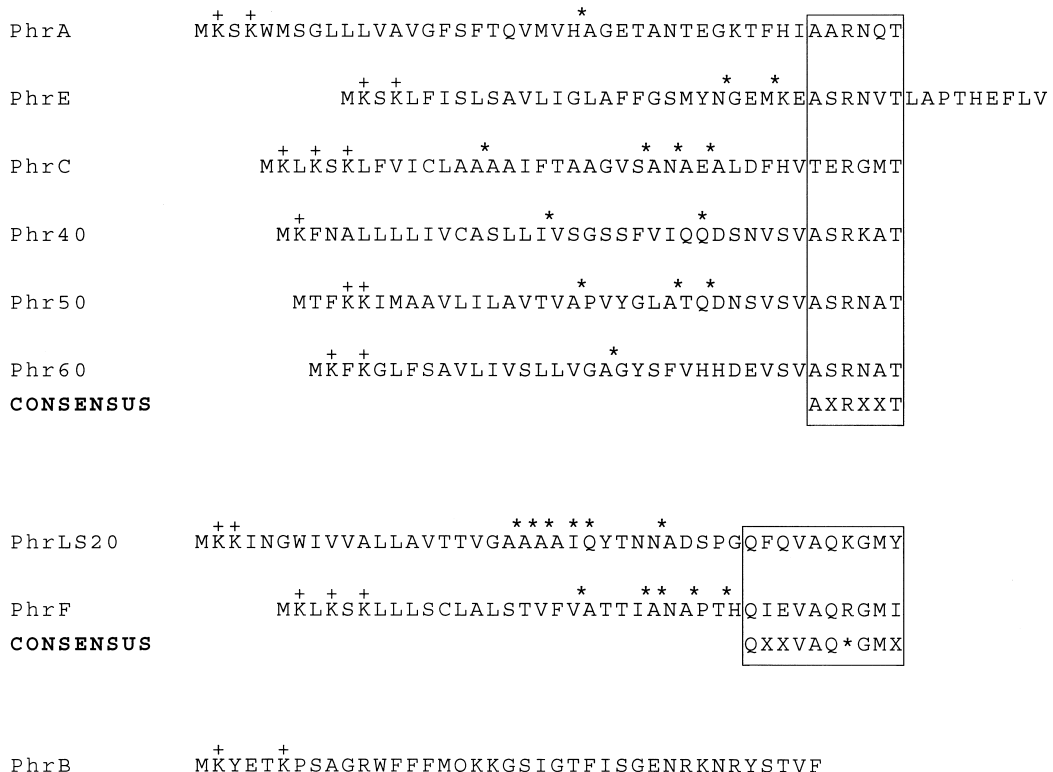


Fig. 12. Deduced amino acid sequences of the small genes that are translationally coupled to the *rap* genes. Positively charged amino acid residues at the N-terminus (+) and potential SPase cleavage sites (*) are indicated above the sequence. Possible conserved motifs are boxed.

exception, since no clear SPase I cleavage site was found in this case).

The presence of potential signal peptides in most of the small gene products indicates that these are secreted into the medium. Support for this hypothesis has recently been obtained for the products of the small genes at the 3'-ends of *rapA* [93] and *rapC* [99]. In the case of *rapA* it was shown that RapA phosphatase activity is negatively controlled by the product of the small gene at its 3'-end, and that the control requires the uptake from the medium via the oligopeptide transport system of at least part of the cognate peptide (minimally the last six amino acid residues from the C-terminus of the peptide are required [91,93]). The *rapA*-associated small gene was, therefore, renamed *phrA* (for phosphatase regulator). As with RapA, the activity of RapC (negative regulator of *srfA*) is antagonized by expression of the small gene located at its 3'-end [99] and, by analogy,

this gene was named *phrC*. The *phrC* gene appeared to specify a precursor of the extracellular CSF (competence and sporation stimulating factor), which serves as a cell density signal for both competence development and sporulation [99]. CSF turned out to be a pentapeptide, corresponding to the carboxy terminus of PhrC which, like PhrA, requires the SpoOK oligopeptide permease for its uptake by the cell [99].

Although no clear homology exists between the deduced amino acid sequences of the various *phr* gene products, an 'AxRxxT' motif was noticed by Perego and Hoch [93] in the C-terminal ends of PhrA, PhrC, and PhrE. As indicated in Fig. 12 this motif is also present in Phr40, Phr50 and Phr60 and can be more accurately described as an 'ASRxAT' motif in the latter cases. In addition, a different C-terminal motif, 'QxxVAQxx[R/K]GMx', may be present in PhrF and PhrLS20. Since the conserved

motifs in PhrA and PhrC are of functional significance [91,99], it can be conceived that Phr peptides sharing a certain motif can down-regulate the activity of one group of Rap proteins, and peptides sharing another motif the activity of other Rap protein(s).

Intriguing questions are why *rap/phr* modules are located on several *B. subtilis* plasmids, and why the Phr peptides are secreted. Since the *rap/phr* modules appear to play important roles in global gene regulation in *B. subtilis*, in particular of regulons that are expressed in the post-exponential growth phase, the increase in copy number of these genes once they are plasmid-located may well extend the cells' ability to adapt to certain conditions. Concerning the secretion of the PhrA peptide, Perego and Hoch [93] proposed that these could serve as 'quorum sensors'. They envisage that if only a small portion of the total cell population produces the peptide, the internal peptide concentration in producing cells will be too low to down-regulate the Rap phosphatase and sporulation will not be initiated. However, if a substantial fraction of the cells produces the peptide, it will be secreted and re-internalized by other cells which can then prevent the synthesis or activity of the Rap phosphatase, resulting in the initiation of sporulation. The peptide could thus be considered to form a check point for sporulation [93].

10. Unique modules (pTA1040, pTA1050 and pTA1060)

10.1. *ParABC* (pTA1050)

Insertional mutagenesis on pPOD2000 (=pTA1050) implicated that the *parABC* region is involved in segregational stability [10]. The predicted products of all three ORFs have signal peptides. ParA has significant similarity to the serine protease subtilisin, while ParB has a probable transmembrane domain. ParC appears to regulate the effect of these polypeptides at a level different from transcription or translation. It may thus provide some sort of immunity function [104]. It is possible that the genes act as a post-segregational killing system acting against plasmid-free bacteria.

10.2. *ORF3C40* (pTA1040)

Downstream of *rep*(pTA1040) ORF3C40 is located which could encode a protein of 201 amino acids. The *rep* gene and ORF3C40 are convergently transcribed and the reading frames are separated by 39 bp in a region showing dyad symmetry. No significant homology of the deduced protein product was found with other proteins, but it contains a putative signal peptide, indicating that it is secreted into the medium. The deduced amino acid sequence, together with the potential SPase cleavage sites, is shown in Fig. 13A.

10.3. *ORF7C60* (pTA1060)

As in pTA1040, in pTA1060 an ORF (ORF7C60) is located downstream of and convergently transcribed with the *rep* gene. This ORF could encode a protein of 180 amino acids. Although in the two plasmids these ORFs and *rep* have a very similar structural organization, including a short intergenic region, the deduced protein sequences of ORF3C40 and ORF7C60 show little homology. The only similarity is that the ORF7C60 product also contains a putative signal peptide. Unlike the putative ORF3C40 product, that of ORF7C60 has a potential transmembrane spanning segment (amino acids 95 to 115), the topology of which is likely to be such that its N-terminal part is located at the outer side of the membrane and the C-terminal part in the cytoplasm (according to the positive-inside-out-rule of Claros and von Heijne [105]). No significant homology was found with other protein sequences. However, an ORF is present downstream of the *rep* gene of pBAA1, the deduced amino acid sequence of which is almost identical to the C-terminal part of the ORF7C60 product (further sequence information is not available for pBAA1). As in pTA1060, in pBAA1 the directions of transcription of this ORF and *rep* are convergent and in both plasmids a short intergenic region containing dyad symmetry separates the genes. An alignment of the deduced amino acid sequences of the two genes is shown in Fig. 13B. So far, the biological function of ORF7C60 is unknown.

10.4. ORF4.60 and ORF5.60 (pTA1060)

pTA1060 contains unique ORFs between the SSO and ORF2C, designated ORF4.60 and ORF5.60. ORF4.60 could encode a protein of 305 amino acids and ORF5.60 a protein of 132 amino acids (Fig. 13C). The reading frames of ORF4.60 and ORF5.60 are separated by only 20 bp on which two possible RBSs for ORF5.60 are located. This organization suggests that the two ORFs form an operon. A potential σ^A promoter is located 29 bp upstream of the start codon of ORF4.60 (TTGAGT–17 nt–TATGAT). Neither the deduced protein sequence of ORF4.60, nor that of ORF5.60 shows significant homology with sequences available in the databases. Until now, the biological function of these putative genes is unknown.

11. Conclusions

Until recently, available data for small endogenous *B. subtilis* plasmids was limited to size, restriction maps and DNA sequences of specific regions. Now, the DNA sequences of the regions encompassing the DSO and cognate *rep* gene of at least one member of the following groups of *B. subtilis* are known: pTA1015, pTA1020, pTA1040, pTA1050, pTA1060 and pFTB14. All plasmids in these groups belong to the pC194 family of RCR plasmids.

The six representative pTA plasmids contain either a palT1- (pTA1015, pTA1020, and pTA1060) or a palT2-type SSO (pTA1030, pTA1040 and pTA1050). The similarities in the DNA sequences involved in replication indicate that all the plasmids described here are highly related to each other, and are likely to be derived from a common ancestor. Like several other RCR plasmids, pTA1015 and pTA1060 contain a *mob* gene and its putative target sequence (*oriT*), enabling the conjugative transfer of the plasmids.

In addition to the replication and mobilization modules, other interesting modules were identified on pTA1015, pTA1040, pTA1050, and pTA1060. These may be unique for RCR plasmids from *B. subtilis* since analogous modules have not been identified on RCR plasmids from other Gram-positive

bacteria. At least three of these modules contain genes which are homologous to chromosomally-located genes. One of these, *sip*, present on pTA1015 and pTA1040, encodes a functional type I SPase. A second gene, *rap*, present on pTA1040, pTA1050, and pTA1060, is likely to encode a phosphatase involved in the regulation of stationary phase processes, and a third gene, *parA*, present on pTA1050, has significant similarity to the serine protease subtilisin and seems to be involved in the stability of the plasmid. Although the actual function of these modules and the reason why they are located on plasmids is, at present, not fully clear, we consider it likely that these modules extend the possibilities of the host to adapt to specific conditions. Since the *B. subtilis* strains harboring these plasmids were industrial isolates, selection for specific (industrial) traits may have been the driving force for the establishment of these genes on plasmids.

In conclusion, the systematic sequence analyses of four cryptic RCR plasmids of *B. subtilis* presented here, has proven to be a valuable approach for increasing our knowledge of RCR plasmids in general and of *B. subtilis* RCR plasmids in particular. These studies provided detailed insight in the structural organization and relatedness of these plasmids and allowed a classification of many *B. subtilis* RCR plasmids.

Acknowledgments

We thank Henk Mulder for preparing the figures. We are indebted to Leendert Hamoen, Bert-Jan Haijema, Rob Meima, Jan Maarten van Dijl, Albert Bolhuis, Harold Tjalsma, and Steven de Jong for helpful discussions. Philippe Glaser, Marta Perego, and James Hoch provided us with information on the *rap/phr* genes prior to publication. W.J.J.M. was supported by STW (Stichting Technische Wetenschappen, The Netherlands) and Gist-Brocades BV (Delft, The Netherlands). S.B. was supported by CEU Grants BRIDGE Program BIOT-CT910268 and BIO2 CT93 0254. P.B.T. was supported by the UK Biotechnology and Biological Sciences Research Council Grant GR/J62838.

References

- [1] Gruss, A. and Ehrlich, S.D. (1989) The family of highly interrelated single-stranded deoxyribonucleic acid plasmids. *Microbiol. Rev.* 53, 231–241.
- [2] Novick, R.P. (1989) Staphylococcal plasmids and their replication. *Annu. Rev. Microbiol.* 43, 537–565.
- [3] Bron, S. (1990) Plasmids. In: *Molecular Biological Methods for Bacillus* (Harwood, C.R. and Cutting, S.M., Eds.) pp. 75–174. John Wiley and Sons Ltd., Chichester, UK.
- [4] Espinosa, M., del Solar, G., Rojo, F. and Alonso, J.C. (1995) Plasmid rolling circle replication and its control. *FEMS Microbiol. Lett.* 130, 111–120.
- [5] Tanaka, T. and Koshikawa, T. (1977) Isolation and characterization of four types of plasmids from *Bacillus subtilis* (*natto*). *J. Bacteriol.* 131, 699–701.
- [6] Tanaka, T., Kuroda, M. and Sakaguchi, K. (1977) Isolation and characterization of four plasmids from *Bacillus subtilis*. *J. Bacteriol.* 129, 1487–1494.
- [7] Uozumi, T., Ozaki, A., Beppu, T. and Arima, K. (1980) New cryptic plasmid of *Bacillus subtilis* and restriction analysis of other plasmids found by general screening. *J. Bacteriol.* 142, 315–318.
- [8] Yoshimura, K., Yamamoto, O., Seki, T., and Oshima, Y. (1983) Distribution of heterogeneous and homologous plasmids in *Bacillus* spp. *Appl. Environ. Microbiol.* 46, 1268–1275.
- [9] Hara, T., Zhang, J. and Ueda, S. (1983) Identification of plasmids linked with polyglutamate production in *Bacillus subtilis* (*natto*). *J. Gen. Appl. Microbiol.* 29, 345–354.
- [10] Gleave, A.P., Mountain, A. and Thomas, C.M. (1990) Use of a novel cassette to label phenotypically a cryptic plasmid of *Bacillus subtilis* and map loci involved in its stable maintenance. *J. Gen. Microbiol.* 136, 905–912.
- [11] Kanapina, A.S., Kanapin, A.A. and Prozorov, A.A. (1995) Determination and comparative analysis of nucleotide sequences in fragments of the minireplicon of p1414 cryptic plasmid from the *Bacillus subtilis* soil strain. *Russ. J. Genet.* 31, 1023–1030.
- [12] Ehrlich, S.D. (1977) Replication and expression of plasmids from *Staphylococcus aureus* in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* 74, 1680–1682.
- [13] Dubnau, D. (1983) Molecular cloning in *Bacillus subtilis*. In: *Experimental Manipulation of Gene Expression* (Inouye, E., Ed.) pp. 33–51. Academic Press Inc., New York.
- [14] Lacks, S.A., Lopez, P., Greenberg, B. and Espinosa, M. (1986) Identification and analysis of genes for tetracycline resistance and replication functions in the broad-host-range plasmid pLS1. *J. Mol. Biol.* 192, 753–765.
- [15] Janni re, L., Gruss, A. and Ehrlich, S.D. (1993) Plasmids. In: *Bacillus subtilis and Other Gram-positive Bacteria; Biochemistry, Physiology and Molecular Genetics* (Sonenshein, A.L., Hoch, J.A. and Losick, R., Eds.) pp. 625–644. American Society for Microbiology, Washington.
- [16] Ehrlich, S.D., Noirot, P., Petit, M.A., Janni re, L., Michel, B. and te Riele, H. (1986) Structural instability of *Bacillus subtilis* plasmids. In: *Genetic Engineering* (Setlow, J.K. and Holander, A., Eds.) pp. 71–83. Plenum Publishing Corp., New York.
- [17] Bron, S., Holsappel, S., Venema, G. and Peeters, B.P.H. (1991) Plasmid deletion formation between short direct repeats in *Bacillus subtilis* is stimulated by single-stranded rolling-circle replication intermediates. *Mol. Gen. Genet.* 226, 88–96.
- [18] Bron, S., Meijer, W., Holsappel, S. and Haima, P. (1991) Plasmid instability and molecular cloning in *Bacillus subtilis*. *Res. Microbiol.* 142, 875–883.
- [19] Ehrlich, S.D., Bierne, H., D’Alençon, E., Vilette, D., Petranovic, M., Noirot, P. and Michel, B. (1993) Mechanisms of illegitimate recombination. *Gene* 135, 161–166.
- [20] del Solar, G.H., Puyet, A. and Espinosa, M. (1987) Initiation signals for the conversion of single stranded to double stranded DNA forms in the streptococcal plasmid pLS1. *Nucleic Acids Res.* 15, 5561–5580.
- [21] Meijer, W.J.J., van der Lelie, D., Venema, G. and Bron, S. (1995) Effects of the generation of single-stranded DNA on the maintenance of plasmid pMV158 and derivatives in *Lactococcus lactis*. *Plasmid* 33, 91–99.
- [22] Meijer, W.J.J., van der Lelie, D., Venema, G. and Bron, S. (1995) Effects of the generation of single-stranded DNA on the maintenance of plasmid pMV158 and derivatives in different *Bacillus subtilis* strains. *Plasmid* 33, 79–89.
- [23] Viret, J. and Alonso, J.C. (1988) A DNA sequence outside the pUB110 minimal replicon is required for normal replication in *Bacillus subtilis*. *Nucleic Acids Res.* 16, 4389–4406.
- [24] Kunst, F., Ogasawara, N., and 150 others (1997) The complete genome sequence of the Gram-positive model organism *Bacillus subtilis* (strain 168). *Nature* 390, 249–256.
- [25] McKay, L.L. (1983) Functional properties of plasmids in lactic streptococci. *Antonie van Leeuwenhoek* 49, 259–274.
- [26] Meijer, W.J.J., Venema, G. and Bron, S. (1995) Characterization of single strand origins of cryptic rolling-circle plasmids from *Bacillus subtilis*. *Nucleic Acids Res.* 23, 612–619.
- [27] Darabi, A., Forough, R., Bhardwaj, G., Watabe, M., Goodarzi, G., Gross, S.C. and Watabe, K. (1989) Identification and nucleotide sequence of the minimal replicon of the low copy number plasmid pBS2. *Plasmid* 22, 281–286.
- [28] Devine, K.M., Hogan, S.T., Higgins, D.G. and McConnell, D.J. (1989) Replication and segregational stability of the *Bacillus* plasmid pBAA1. *J. Bacteriol.* 171, 1166–1172.
- [29] Poluektova, E.U., Karandashova, I.V., Sapogova, E.Y. and Prozorov, A.A. (1996) Study of homology of *Bacillus subtilis* cryptic plasmids. *Genetika* 32, 1498–1503.
- [30] Seki, T., Oshima, T. and Oshima, Y. (1975) Taxonomic study of *Bacillus* by deoxyribonucleic acid-deoxyribonucleic acid hybridization and interspecific transformation. *Int. J. Syst. Bacteriol.* 25, 258–270.
- [31] Nezametdinova, V.Z., Kanapina, A.S., Kozlovski, Y.E. and Prozorov, A.A. (1992) Cryptic plasmids from soil strains of *Bacilli*. *Genetika* 27, 49–55.
- [32] Kozlovski, Y.E. and Prozorov, A.A. (1981) Restriction and modification systems in bacilli strains related to *Bacillus subtilis*. *Dokl. Akad. Nauk SSSR* 258, 1457–1459.
- [33] Hara, T., Ogata, S. and Ueda, S. (1993) Plasmid distribution

- in gamma-polyglutamate-producing *Bacillus* strains isolated from 'Dan-Douchi', a 'natto'-like non-salty fermented soybean food in China. *J. Gen. Appl. Microbiol.* 39, 75–82.
- [34] Zawadzki, P., Riley, M.A. and Cohan, F.M. (1996) Homology among nearly all plasmids infecting three *Bacillus* species. *J. Bacteriol.* 178, 191–198.
- [35] Meijer, W.J.J., de Boer, A., van Tongeren, S., Venema, G. and Bron, S. (1995) Characterization of the replication region of the *Bacillus subtilis* plasmid pLS20: a novel type of replicon. *Nucleic Acids Res.* 23, 3214–3223.
- [36] Kok, J. and Venema, G. (1988) Genetics of proteinases of lactic acid bacteria. *Biochimie* 70, 475–488.
- [37] Seegers, J.F.M.L., Bron, S., Franke, C.M., Venema, G. and Kiewiet, R. (1994) The majority of lactococcal plasmids carry a highly related replicon. *Microbiology* 140, 1291–1300.
- [38] Hara, T., Nakajima, K., Saito, H., Ishizaki, A., Ogata, S. and Ueda, S. (1992) Sequence analysis of replication origin of plasmid pLS11 of *Bacillus subtilis* IFO3022. *Biosci. Biotech. Biochem.* 56, 223–227.
- [39] Hara, T., Nagatomo, S., Ogata, S. and Ueda, S. (1991) Molecular structure of the replication origin of a *Bacillus subtilis* (natto) plasmid, pUH1. *Appl. Environ. Microbiol.* 57, 1838–1841.
- [40] Murai, M., Miyashita, H., Araki, H., Seki, T. and Oshima, Y. (1987) Molecular structure of the replication origin of a *Bacillus amyloliquefaciens* plasmid pFTB14. *Mol. Gen. Genet.* 210, 92–100.
- [41] Chang, S., Chang, S. and Gray, O. (1987) Structural and genetic analysis of a *par* locus that regulates plasmid partition in *Bacillus subtilis*. *J. Bacteriol.* 169, 3952–3962.
- [42] Seery, L. and Devine, K.M. (1993) Analysis of features contributing to activity of the single-stranded origin of *Bacillus* plasmid pBAA1. *J. Bacteriol.* 175, 1988–1994.
- [43] Hara, T., Nagatomo, S., Ogata, S. and Ueda, S. (1992) The DNA sequence of gamma-glutamyltranspeptidase gene of *Bacillus subtilis* (natto) plasmid pUH1. *Appl. Environ. Microbiol.* 37, 211–215.
- [44] Projan, S. and Novick, R. (1988) Comparative analysis of five related staphylococcal plasmids. *Plasmid* 19, 203–221.
- [45] Meijer, W.J.J. (1995) Replication and maintenance of plasmids in *Bacillus subtilis*. Academic Thesis, State University of Groningen, The Netherlands.
- [46] Seery, L.T., Nolan, N.C., Sharp, P.M. and Devine, K.M. (1993) Comparative analysis of the pC194 group of rolling circle plasmids. *Plasmid* 30, 185–196.
- [47] Ilyina, T.V. and Koonin, E.V. (1992) Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaeobacteria. *Nucleic Acids Res.* 20, 3279–3285.
- [48] Thomas, C.D., Balson, D. and Shaw, W. (1988) Identification of the tyrosine residue involved in bond formation between replication origin and initiator protein of plasmid pC221. *Biochem. Soc. Trans.* 16, 758–759.
- [49] Gros, M.F., Bیدنکو, V. and Ehrlich, S.D. (1994) Active site of the replication protein of the rolling circle plasmid pC194. *EMBO J.* 13, 4412–4420.
- [50] Haima, P., Bron, S. and Venema, G. (1987) The effect of restriction on shotgun cloning and plasmid stability in *Bacillus subtilis* Marburg. *Mol. Gen. Genet.* 209, 335–342.
- [51] Meijer, W.J.J., Venema, G. and Bron, S. (1997) Unpublished results.
- [52] Hara, T., Yasuda, S., Ogata, S., Soeda, E. and Ueda, S. (1992) Complete nucleotide sequence of *Bacillus subtilis* (natto) plasmid responsible for gamma-polyglutamate synthesis. *J. Fac. Agr. Kyushu Univ.* 36, 209–218.
- [53] Bron, S. and Holsappel, S. (1997) Unpublished results.
- [54] Mahillon, J. and Seurinck, J. (1988) Complete nucleotide sequence of pGI2, a *Bacillus thuringiensis* plasmid containing Tn4430. *Nucleic Acids Res.* 16, 11827–11828.
- [55] Madsen, S.M., Andrup, L. and Boe, L. (1993) Fine mapping and DNA sequence of replication functions of *Bacillus thuringiensis* plasmid pTX14-3. *Plasmid* 30, 119–130.
- [56] Meister, A. and Tate, S. (1976) Glutathione and related gamma-glutamyl compounds: biosynthesis and utilization. *Annu. Rev. Biochem.* 45, 559–604.
- [57] Ryoichi, K., Toru, M., Shigenori, O., Moriyuki, S. and Kazuo, Y. (1991) Process for the production of gamma-glutamyl transpeptidase. European Patent 91101620.2, 1–22.
- [58] Suzuki, H., Kumagai, H., Echigo, T. and Tochikura, T. (1989) DNA sequence of the *Escherichia coli* K-12 gamma-glutamyltranspeptidase gene, *ggt*. *J. Bacteriol.* 171, 5169–5172.
- [59] Schaeffer, P., Millet, I. and Aubert, J. (1965) Catabolite repression of bacterial sporulation. *Proc. Natl. Acad. Sci. USA* 54, 704–711.
- [60] Koehler, T.M. and Thorne, C.B. (1987) *Bacillus subtilis* (natto) plasmid pLS20 mediates interspecies plasmid transfer. *J. Bacteriol.* 169, 5271–5278.
- [61] Oskam, L., Venema, G. and Bron, S. (1991) The large *Bacillus* plasmid pTB19 contains two integrated rolling-circle plasmids carrying mobilization functions. *Plasmid* 26, 30–39.
- [62] Priebe, S.D. and Lacks, S.A. (1989) Region of the streptococcal plasmid pMV158 required for conjugative mobilization. *J. Bacteriol.* 171, 4778–4784.
- [63] Selinger, L.B., McGregor, N.F., Khachatourians, G.G. and Hynes, M.F. (1990) Mobilization of closely related plasmids pUB110 and pBC16 by *Bacillus* plasmid pXO503 requires *trans*-acting open reading frame β . *J. Bacteriol.* 172, 3290–3297.
- [64] van der Lelie, D., Wösten, H.A.B., Bron, S., Oskam, L. and Venema, G. (1990) Conjugal mobilization of streptococcal plasmid pMV158 between strains of *Lactococcus lactis* subsp. *lactis*. *J. Bacteriol.* 172, 47–52.
- [65] Koonin, E.V. and Ilyina, T.V. (1993) Computer-assisted dissection of rolling circle DNA replication. *Biosystems* 30, 241–268.
- [66] Koonin, E.V. (1993) A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif. *J. Mol. Biol.* 232, 1013–1013.
- [67] Guzmán, L.M. and Espinosa, M. (1997) The mobilization protein, MobM, of the streptococcal plasmid pMV158 specifically cleaves supercoiled DNA at the plasmid *oriT*. *J. Mol. Biol.* 266, 688–702.
- [68] Projan, S.J. and Archer, G.L. (1989) Mobilization of the relaxable *Staphylococcus aureus* plasmid pC221 by the conjuga-

- tive plasmid pGO1 involves three pC221 loci. *J. Bacteriol.* 171, 1841–1845.
- [69] Gennaro, M.L., Kornblum, J. and Novick, R.P. (1987) A site-specific recombination function in *Staphylococcus aureus* plasmids. *J. Bacteriol.* 169, 2601–2610.
- [70] van der Lelie, D., Bron, S., Venema, G. and Oskam, L. (1989) Similarity of minus origins of replication and flanking open reading frames of plasmids pUB110, pTB913 and pMV158. *Nucleic Acids Res.* 17, 7283–7294.
- [71] Catchpole, I., Thomas, C., Davies, A. and Dyke, K.G.H. (1988) The nucleotide sequence of *Staphylococcus aureus* plasmid pT48 conferring inducible macrolide-lincosamide-streptogramin B resistance and comparison with similar plasmids expressing constitutive resistance. *J. Gen. Microbiol.* 134, 697–709.
- [72] Lampson, B.C. and Parisi, J.T. (1986) Nucleotide sequence of the constitutive macrolide-lincosamide-streptogramin B resistance plasmid pNE131 from *Staphylococcus epidermidis* and homologies with *Staphylococcus aureus* plasmids pE194 and pSN2. *J. Bacteriol.* 167, 888–892.
- [73] Xu, F., Pierce, L.E. and Yu, P.-L. (1991) Genetic analysis of a lactococcal plasmid replicon. *Mol. Gen. Genet.* 227, 33–39.
- [74] Meijer, W.J.J., de Jong, A., Wisman, G.B.A., Tjalsma, H., Venema, G., Bron, S. and van Dijl, J.M. (1995) The endogenous *Bacillus subtilis* (natto) plasmids pTA1015 and pTA1040 contain signal peptidase-encoding genes: identification of a new structural module on cryptic plasmids. *Mol. Microbiol.* 17, 621–631.
- [75] van Dijl, J.M., de Jong, A., Vehmaanperä, J., Venema, G. and Bron, S. (1992) Signal peptidase I of *Bacillus subtilis*: patterns of conserved amino acids in prokaryotic and eukaryotic type I signal peptidases. *EMBO J.* 11, 2819–2828.
- [76] Simonen, M. and Palva, I. (1993) Protein secretion in *Bacillus* species. *Microbiol. Rev.* 57, 109–137.
- [77] van Dijl, J.M., de Jong, A., Smith, H., Venema, G. and Bron, S. (1993) Protein secretion in *Bacillus subtilis*. In: *DNA Transfer and Gene Expression in Microorganisms* (Balla, E., Berencsi, G., Farkas, J. and Szentirmai, A., Eds.) pp. 251–261. Intercept Ltd., Andover, UK.
- [78] Nagarajan, V. (1993) Protein secretion. In: *Bacillus subtilis and Other Gram-positive Bacteria; Biochemistry, Physiology and Molecular Genetics* (Sonenshein, A.L., Hoch, J.A. and Losick, R., Eds.) Vol. 1, pp. 713–728. American Society for Microbiology, Washington, DC.
- [79] von Heijne, G. (1994) Signal Peptidase. R.G. Landes Company, Austin, TX.
- [80] von Heijne, G. and Abrahmsén, L. (1989) Species-specific variation in signal peptide design. *FEBS Lett.* 244, 439–446.
- [81] Tjalsma, H., Noback, M.A., Bron, S., Venema, G., Yamane, K. and van Dijl, J.M. (1997) *Bacillus subtilis* contains four closely related type I signal peptidases with overlapping substrate specificities: constitutive and temporally controlled expression of different *sip* genes. *J. Biol. Chem.* 272, 25983–25992.
- [82] Akagawe, E., Kurita, K., Sugawara, T., Nakamura, K., Kasakara, Y., Ogasawara, N. and Yamane, K. (1995) Determination of a 17,484 bp nucleotide sequence around the 39 degrees region of the *Bacillus subtilis* chromosome and similarity analysis of the products of putative ORFs. *Microbiology* 141, 3341–3345.
- [83] van Dijl, J.M., de Jong, A., Smith, H., Bron, S. and Venema, G. (1991) Signal peptidase I overproduction results in increased efficiencies of export and maturation of hybrid secretory proteins in *Escherichia coli*. *Mol. Gen. Genet.* 227, 40–48.
- [84] Black, M.T., Munn, J.G.R. and Allsop, A. (1992) On the catalytic mechanism of prokaryotic leader peptidase I. *Biochem. J.* 282, 539–543.
- [85] van Dijl, J.M., van den Bergh, R., Reversma, T., Smith, H., Bron, S. and Venema, G. (1990) Molecular cloning of the *Salmonella typhimurium* *lep* gene in *Escherichia coli*. *Mol. Gen. Genet.* 223, 233–240.
- [86] Wolfe, P.B., Wickner, W. and Goodman, J.M. (1983) Sequence of the leader peptidase gene of *Escherichia coli* and the orientation of leader peptidase in the bacterial envelope. *J. Biol. Chem.* 258, 12073–12080.
- [87] Dalbey, R.E. and von Heijne, G. (1992) Signal peptidases in prokaryotes and eukaryotes: a new protease family. *Trends Biochem. Sci.* 17, 474–478.
- [88] Packer, J.C.L., André, D. and Howe, C.J. (1995) Cloning and sequence analysis of a signal peptidase I from the thermophilic cyanobacterium *Phormidium laminosum*. *Plant Mol. Biol.* 27, 199–204.
- [89] van Dijl, J.M., de Jong, A., Venema, G. and Bron, S. (1995) Identification of the potential active site of the signal peptidase SipS of *Bacillus subtilis*. Structural and functional similarities with LexA-like proteases. *J. Biol. Chem.* 270, 3611–3618.
- [90] Bolhuis, A., Sorokin, A., Azevedo, V., Ehrlich, S.D., Braun, P.G., de Jong, A., Venema, G., Bron, S. and van Dijl, J.M. (1996) *Bacillus subtilis* can modulate its capacity and specificity for protein secretion through temporally controlled expression of the *sipS* gene for signal peptidase I. *Mol. Microbiol.* 22, 605–618.
- [91] Perego, M., Glaser, P. and Hoch, J.A. (1996) Aspartyl-phosphate phosphatases deactivate the response regulator components of the sporulation signal transduction system in *Bacillus subtilis*. *Mol. Microbiol.* 19, 1151–1157.
- [92] Perego, M., Hanstein, C., Welsh, K.M., Djavakhishvili, T., Glaser, P. and Hoch, J.A. (1994) Multiple protein aspartate phosphatases provide a mechanism for the integration of diverse signals in the control of development in *Bacillus subtilis*. *Cell* 79, 1047–1055.
- [93] Perego, M. and Hoch, J.A. (1996) Cell-cell communication regulates the effects of protein aspartate phosphatases on the phosphorelay controlling development in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* 93, 1549–1553.
- [94] Mueller, J.P., Bukusoglu, G. and Sonenshein, A.L. (1992) Transcriptional regulation of *Bacillus subtilis* glucose starvation-inducible genes: control of *gsiA* by the ComP-ComA signal transduction system. *J. Bacteriol.* 174, 4361–4373.
- [95] Mueller, J.P. and Sonenshein, A.L. (1992) Role of the *Bacillus subtilis* *gsiA* gene in regulation of early sporulation gene expression. *J. Bacteriol.* 174, 4374–4383.
- [96] Hoch, J.A. (1993) Regulation of the phosphorelay and the

- initiation of sporulation in *Bacillus subtilis*. *Annu. Rev. Microbiol.* 47, 441–465.
- [97] Glaser, P. (1997) Unpublished results.
- [98] Takemaru, K.-I., Mizuno, M., Sato, T., Takeuchi, M. and Kobayashi, Y. (1995) Complete nucleotide sequence of a *skin* element excised by DNA rearrangement during sporulation in *Bacillus subtilis*. *Microbiology* 141, 323–327.
- [99] Solomon, J.M., Lazazzera, B.A. and Grossman, A.D. (1996) Purification and characterization of an extracellular peptide factor that affects different developmental pathways in *Bacillus subtilis*. *Genes Develop.* 10, 2014–2024.
- [100] Strauch, M.A. and Hoch, J.A. (1992) Transition state regulators: sentinels of *Bacillus subtilis* post-exponential gene expression. *Mol. Microbiol.* 7, 337–342.
- [101] Sejalón, N., Bron, S. and Meijer, W.J.J. (1997) Unpublished results.
- [102] Msadek, T., Kunst, F. and Rapoport, G. (1993) Two-component regulatory systems. In: *Bacillus subtilis* and Other Gram-positive Bacteria: Biochemistry, Physiology and Molecular Genetics (Sonenshein, A.L., Hoch, J.A. and Losick, R., Eds.) Vol. 1, pp. 729–746. American Society for Microbiology, Washington, DC.
- [103] Dubnau, D. (1991) Genetic competence in *Bacillus subtilis*. *Microbiol. Rev.* 55, 395–424.
- [104] Thorsted, P.B., Rowlinson, C.A., Gleave, A.P. and Thomas, C.M. (1997) Sequence and molecular analysis of a small cryptic *Bacillus subtilis* plasmid pPOD2000 reveals: a putative serin protease helps to ensure stable inheritance of the plasmid in *Bacillus subtilis* (unpublished results).
- [105] Claros, M.G. and von Heijne, G. (1994) TopPred II: an improved software for membrane protein structure predictions. *Comput. Appl. Biosci.* 10, 685–686.
- [106] Sneath, P.H.A. and Sokal, R.R. (1973) Numerical taxonomy. W.H. Freeman, San Francisco.
- [107] Moon, K.H., Lee, D.W., Yoon, S.J., Kim, W.K., Shin, C.K. and Im, S.H. (1995) Complete nucleotide sequence of the *Staphylococcus aureus* plasmid pKH6. Thesis Kyungsoong University, 110-1, Deayon-dong, Nam-gu, Pusan, 608-736 Korea.
- [108] Noguchi, N., Aoki, T., Sasatsu, M., Kono, M., Shishido, K. and Ando, T. (1986) Determination of the complete nucleotide sequence of pNS1, a staphylococcal tetracycline-resistance plasmid propagated in *Bacillus subtilis*. *FEMS Microbiol. Lett.* 37, 283–288.
- [109] Hong, S.H. (1995) Enhancing survival of lactic acid bacteria in ice cream by natural encapsulation and gene transfer. Academic Thesis, Food Science and Human Nutrition, University of Missouri, Columbia (MO, USA).
- [110] Bates, E.E.M. and Gilbert, H.J. (1989) Characterization of a cryptic plasmid from *Lactobacillus plantarum*. *Gene* 85, 253–258.
- [111] Jossón, K., Soetaert, P., Michiels, F., Joos, H. and Mahillon, J. (1990) *Lactobacillus hilgardii* plasmid pLAB1000 consists of two functional cassettes commonly found in other Gram-positive organisms. *J. Bacteriol.* 172, 3089–3099.
- [112] LeBlanc, D., Chen, Y.M. and Lee, L. (1993) Identification and characterization of a mobilization gene in the streptococcal plasmid, pVA380-1. *Plasmid* 30, 296–302.
- [113] Lampton, B.C. and Parisi, J.T. (1986) Naturally occurring *Staphylococcus epidermidis* plasmid expressing constitutive mocolide-lincosamide-streptogramin B resistance contains a deleted attenuator. *J. Bacteriol.* 166, 479–483.
- [114] von Heijne, G. (1986) A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* 14, 4683–4690.
- [115] von Heijne, G. (1992) Membrane protein structure prediction: hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* 225, 487–494.
- [116] Ogasawara, N., Nakai, S. and Yoshikawa, H. (1994) Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* chromosome containing the replication region. *DNA Res.* 1, 1–14.
- [117] Msadek, T., Kunst, F., Klier, A., Rapoport, G. and Dedonder, R. (1990) Signal transduction pathway controlling synthesis of a class of degradative enzymes in *Bacillus subtilis*: expression of the regulatory genes and analysis of mutations in *degS* and *degU*. *J. Bacteriol.* 172, 824–834.
- [118] Nakano, M.M., Magnuson, R., Myers, A., Curry, J., Grossman, A.D. and Zuber, P. (1991) *SrfA* is an operon required for surfactin production, competence development, and efficient sporulation in *Bacillus subtilis*. *J. Bacteriol.* 173, 1770–1778.
- [119] Moran, C.P., Lang, N., LeGrice, S.F.J., Lee, G., Stephans, M., Sonenshein, A.L., Pero, J. and Losick, R. (1982) Nucleotide sequence that signal the initiation of transcription and translation in *Bacillus subtilis*. *Mol. Gen. Genet.* 186, 339–346.