Plasmid Rolling-Circle Replication

J. A. RUIZ-MASÓ,1 C. MACHÓN,2,3 L. BORDANABA-RUISÉCO,1 M. ESPINOSA,1 M. COLL,2,3 and G. DEL SOLAR1

1Centro de Investigaciones Biológicas (CSIC), Ramiro de Maeztu 9, 28040 Madrid, Spain; 2Institute for Research in Biomedicine (IRB-Barcelona), Baldiri Reixac 10-12, 08028 Barcelona, Spain; 3Institut de Biologia Molecular de Barcelona (CSIC), Baldiri Reixac 10-12, 08028 Barcelona, Spain

ABSTRACT Plasmids are DNA entities that undergo controlled replication independent of the chromosomal DNA, a crucial step that guarantees the prevalence of the plasmid in its host. DNA replication has to cope with the incapacity of the DNA polymerases to start de novo DNA synthesis, and different replication mechanisms offer diverse solutions to this problem. Rolling-circle replication (RCR) is a mechanism adopted by certain plasmids, among other genetic elements, that represents one of the simplest initiation strategies, that is, the nicking by a replication initiator protein on one parental strand to generate the primer for leading-strand initiation and a single priming site for lagging-strand synthesis. All RCR plasmid genomes consist of a number of basic elements: leading strand initiation and control, lagging strand origin, phenotypic determinants, and mobilization, generally in that order of frequency. RCR has been mainly characterized in Gram-positive bacterial plasmids, although it has also been described in Gram-negative bacterial or archaeal plasmids. Here we aim to provide an overview of the RCR plasmids’ lifestyle, with emphasis on their characteristic traits, promiscuity, stability, utility as vectors, etc. While RCR is one of the best-characterized plasmid replication mechanisms, there are still many questions left unanswered, which will be pointed out along the way in this review.

GENERAL ASPECTS OF PLASMID ROLLING-CIRCLE REPLICATION

The main features that characterize rolling-circle replication (RCR) (see Fig. 1A) derive from its singular initiation mechanism, which relies on the sequence-specific cleavage, at the nick site of the double-strand origin (dso), of one of the parental DNA strands by an initiator Rep protein. This cleavage generates a 3'-OH end that allows the host DNA polymerases to initiate the leading strand replication. Therefore, the RCR initiation circumvents the synthesis of a primer RNA that is required in all other modes of replication of circular double-stranded DNA (dsDNA). Elongation of the leading strand takes place as the parental double helix is unwound by a host DNA helicase and the cleaved non-template strand is covered with the single-stranded DNA binding protein. Since the nascent DNA is covalently attached to the parental DNA, termination of a round of leading-strand replication implies a new cleavage event at the reconstituted nick site. This reaction is assumed to be catalyzed by the same Rep molecule that carried out the initiation cleavage and remained bound to the 5' end of the parental strand while traveling along with the replication fork. A trans-esterification then occurs that joins this 5' end to the 3' end generated in the termination cleavage, releasing the displaced parental strand as a circular single-stranded DNA (ssDNA). This replicative intermediate serves as the template for the synthesis of the lagging strand, which depends solely on host-encoded enzymes and is initiated from a highly structured region of the ssDNA, termed the single-strand origin (sso).
Thus, the entire process of asymmetric RCR yields, in two separate steps (this is what asymmetric refers to), two circular dsDNAs containing either the newly synthesized leading or lagging strand and the complementary parental template strand. The DNA ligase and gyrase of the host cell next convert the new daughter
DNA molecules in supercoiled forms indistinguishable from the rest of the plasmid pool. Generation of the ssDNA replicative intermediates is the hallmark of RCR, and detection of intracellular strand-specific plasmid ssDNA provides valuable clues about whether a given plasmid replicates by the rolling-circle mechanism (1, 2).

The basic catalytic mechanism operating in initiation and termination of RCR, i.e., the cleavage and rejoining of ssDNA using an active-site Tyr that forms a transient 5'-phosphotyrosine bond with the cleaved DNA, is involved in a range of processes that take place in mobile genetic elements in all three domains of life. The enzymes that exhibit this catalytic mechanism are mainly included in the widespread His-bulky hydrophobic residue-His (HUH) endonuclease superfamily and have key roles in the replication of plasmids, bacteriophages, and plant and animal viruses; in plasmid conjugative transfer; and in transposition (3). RCR was discovered in ssDNA coliphage ΦX174 some 45 years ago (4–6). The pioneer characterization of gene A protein made the initiator of ΦX174 RCR the first member of the HUH endonuclease superfamily (7–10).

Plasmid RCR was first evidenced for the Staphylococcus aureus plasmid pT181 based on the characterization of the origin-specific nicking-closing activity of the purified pT181-encoded RepC protein (11). Shortly afterward, several other small plasmids from staphylococci, bacilli, streptococci, and streptomyces were also found to replicate by the RCR mechanism (12–14), which led to the assumption that most, if not all, small multicopy plasmids in Gram-positive bacteria use RCR. However, this premise proved inaccurate, as some small plasmids isolated from Gram-positive organisms were later reported to replicate by the theta mode (1). Moreover, although RCR plasmids are particularly abundant in Gram-positive bacteria, they have also been identified in various Gram-negative organisms, in archaea, and in mitochondria of the higher plant Chenopodium album (1, 2, 15).

Natural RCR plasmids range in size from as low as the 846 bp of the Thermotoga plasmids pRQ7, pMC24, and pRKU1 (16–18) to the almost 30 kb of pCG4 from Corynebacterium glutamicum (19). The nearly identical plasmids pRQ7, pMC24, and pRKU1 are the smallest found so far and consist of only the basic replicon, i.e., the backbone regions involved in replication and copy-number control. The basic replicon of RCR plasmids should include an essential module containing the dso and the genes that encode the initiator Rep protein and the replication control element(s), as well as at least one host-recognized sso, which, although not strictly essential, provides efficient synthesis of the lagging strand and hence is present in all natural RCR plasmids (Fig. 2). Homology in the essential module of the basic replicon has been the criterion used to classify RCR plasmids into replicon families (see below).

Apart from the basic replicon, some larger RCR plasmids contain additional backbone genes and elements that contribute to their maintenance or help them transfer between host cells (Fig. 2). Of special relevance, because of its frequent presence in RCR plasmids, is the MOB module, which is involved in the conjugal mobilization of the plasmid and consists of the transfer origin (oriT) and the mob gene(s) that encode the relaxase protein and, in some cases, auxiliary proteins (20). The apparent lack of active partition systems in RCR plasmids is consistent with the medium copy number (10 to 30 per chromosome equivalent) that they exhibit in their natural hosts. This feature ensures the stable inheritance of RCR plasmids by only random segregation to the daughter cells, providing that the replication control system efficiently corrects fluctuations of the plasmid copy number in single cells and that the plasmid molecules are maintained as individual copies. In this sense, the presence of homologs to components of toxin-antitoxin (TA) systems in some RCR plasmids is intriguing (21). It is noteworthy that whereas the TA systems were first proposed to play a role in plasmid stability through postsegregational killing of plasmid-free cells, the more recent competition hypothesis postulates that acquisition of these modules allows plasmids to exclude competing TA-free plasmids (22–24).

Some RCR plasmids also carry accessory genes that encode functions that can benefit the host cell under special conditions, thus reflecting the adaptation of the bacteria to their environment (Fig. 2). Antibiotic resistance determinants are among the most frequent traits encoded by RCR plasmids isolated from a variety of bacteria (25). Other accessory genes have been found to be relatively abundant in RCR plasmids from a given host. This is the case of small heat shock protein (shsp) genes carried by Streptococcus thermophilus plasmids belonging to the pC194 replicon family (26, 27). The presence of shsp-containing plasmids has been reported to increase cell survival at the high temperatures reached during different stages of fermentation in the dairy industry (27). Another striking example is the presence, in some Bacillus thuringiensis plasmids, of open reading frames encoding collagen-like proteins that are thought to play a role in aggregation formation or in adherence to other cells or substrates (28).
RCR plasmids are considered to contain promiscuous replicons, as many of them have been shown to replicate in species, genera, or even phyla other than those from which they were isolated (25). The simplicity of the RCR initiation, with only the plasmid-encoded Rep protein participating in recognition of the origin and priming of the leading strand synthesis, may underlie the usual promiscuity of these plasmids. The broadness of the host range of RCR plasmids would depend on the balanced expression of their essential genes involved in initiation and control of replication as well as on the formation of a functional Rep-host helicase complex that can extensively unwind the plasmid DNA in a variety of bacteria (29–31). The broad host range of RCR plasmids is best exemplified by the pMV158-family prototype, which was initially isolated from *Streptococcus agalactiae* and subsequently transferred to a variety of *Firmicutes* (several *Streptococcus* and *Bacillus* species, *Listeria*, *S. aureus*, *Lactococcus lactis*, *Enterococcus faecalis*, *Clostridium*), *Actinobacteria* (*C. glutamicum*, *Brevibacterium*), and the γ-proteobacterium *Escherichia coli*. Moreover, the fact that members of each replicon family have been isolated from a variety of bacteria suggests the promiscuity of the ancestors from which these plasmids derive. In turn, plasmid adaptation to a new host can lead to the narrowing of the host range of the adapted plasmid. This seems to be the case for two *Mycoplasma mycoides* plasmids of the pMV158-replicon family, namely pADB201 and pKMK1, whose rep genes contain at least one UGA codon, which encodes tryptophan in this bacterium but is a stop codon in other bacteria, so that the host range of these plasmids is restricted to *Mycoplasma* species (32).

Due to their general smallness, high copy number, and promiscuity, RCR plasmids appear to be well suited...
Plasmid Rolling-Circle Replication

for the construction of vectors for gene cloning and expression, provided a functional sso is present to minimize the generation of the recombinogenic ssDNA intermediates, which can lead to structural and segregational plasmid instability (33–37). Nevertheless, it has been reported that cloning of heterologous DNA in RCR plasmid vectors can result in the generation of linear high-molecular-weight (HMW) plasmid multimers in relative amounts that correlate positively with the size of the DNA insert (38, 39). The formation of HMW by RCR plasmids has also been implicated in both structural (40) and segregational (41) instability. The generation of HMW plasmid DNA was at first related to a replication defect, as plasmids lacking sso were prone to accumulate HMW DNA (42). Accumulation of HMW plasmid DNA was enhanced in the absence of the ExoV enzyme (RecBCD in Gram-negative or AddAB in Gram-positive bacteria) (41). Despite the potential instability problems, vectors based on RCR plasmids have been developed and successfully used in pneumococci, enterococci, lactococci, and corynebacteria (43–45), for which genetic and biotechnological tools are scarce and hence welcome. It is worth mentioning that most of the nonintegrative plasmid vectors available in Streptococcus pneumoniae are based on pMV158 and that inducible expression vector pLS1ROM and recombinant pLS1ROM-GFP (containing the gfp gene, encoding the Aequorea victoria green fluorescent protein, cloned under control of the maltose-inducible Pm promoter) have proved to be structurally and segregationally stable in pneumococcus, even under induction conditions (45). Similarly, most of the autonomously replicating vectors for the industrial microorganism C. glutamicum are based on plasmids pBL1, pCG1, and pGA1 from C. glutamicum or on the broad-host-range plasmid pNG2 from Corynebacterium diphtheriae, all of them replicating by the rolling circle mode (46). These RCR plasmid vectors were found to be stably maintained in C. glutamicum cells grown under nonselective conditions (47).

An aspect of recognized relevance when pursuing the biotechnological use of plasmid vectors is the metabolic cost that carriage of these extrachromosomal elements imposes on the host, since a significant burden can lead to the overgrowth of the culture by plasmid-free cells even though plasmid inheritance is quite stable. Little information is available on the burden caused by RCR plasmids, as this subject has only been analyzed for the pMV158 replicon. Small (4.4 kb), medium-copy-number (~20 copies per chromosome equivalent) pMV158 derivatives that are stably inherited in pneumococcus and harbor an sso element efficiently recognized in this host slightly burden the S. pneumoniae cells, causing a 7 to 8% increase in the bacterial doubling time (48). Nevertheless, fitness impairment of pneumococcal cells harboring pMV158 derivatives has not been found to negatively affect the segregational stability of pLS1ROM and pLS1ROM-GFP (45).

This chapter aims to provide an updated review of the major findings in the study of the RCR plasmids and to highlight the pending questions and challenges for the detailed understanding of this kind of plasmid replication. Most of these issues have been dealt with in previous reviews on this subject (1, 23, 32, 49, 50).

Apart from the above-referenced asymmetric RCR, which is initiated by the Rep-mediated cleavage of one parental strand, a different, recombination-dependent replication mechanism that also leads to θ-shaped circular intermediates consisting of a circular DNA attached to a growing linear DNA has been reported to play an essential role during the replication cycle of many dsDNA viruses. Single origin-dependent replication of bacterial genomes and of many dsDNA viruses with circular genomes proceeds by the θ (circle to circle) mechanism. The trade-off between different DNA transactions could lead to the stall or collapse of the replication machinery, so that origin-independent remodeling and assembly of a new replisome at the stalled fork is required to restart the replication process. In dsDNA viruses (e.g., bacteriophage lambda, SPP1, etc.), replication restart becomes dependent on recombination proteins with a switch from the origin-mediated θ type to a σ type recombination-dependent replication. The replication shift from θ to σ generates the concatemeric viral DNA substrate needed to produce mature viral particles. This RCR-like σ mode has been reviewed by Lo Piano et al. (51) and will not be addressed here.

**THE DOUBLE-STRAND ORIGIN**

Replication of the leading strand of RCR plasmids initiates and proceeds in a unidirectional manner from their dso, a plasmid DNA region highly specific for its cognate initiator protein that contains the sequences involved in the initiation and termination of the leading strand. The dso, along with the rep gene and the control elements, is part of an essential module that harbors the functions for plasmid replication. Based on the homologies found in this essential module, up to 17 RCR plasmid families have been defined. Only three of these plasmid families have been studied in depth, their prototypes being the staphylococcal plasmids pT181/pC221 (2 and references therein; 52) and pC194/pUB110 (53)
and the streptococcal plasmid pMV158 (1). The following plasmid families have also been studied although less thoroughly: the staphylococcal plasmid pSN2 family (54), the pBL1 and pCG1 plasmid families from C. glutamicum (55 and references therein), the pSTK-1 and pTX14-2 plasmid families from B. thuringiensis (28 and references therein), and the pGRB1 (56) and pGTS (57) plasmid families from archaea.

The dsos of RCR plasmids can be found located upstream of the rep gene (pC194, pMV158, and pSN2 families), embedded within the 5′ portion (pT181 family) or the 3′ portion (pCG1 family) of the sequence coding their respective Rep proteins, or even downstream from the rep gene stop codon (pTX14-2 family). The dsos can be physically and functionally divided into two regions, namely bind, which contains the specific binding sequence for the initiator protein, and nic, where Rep specifically cleaves the DNA at the nick site. The two loci can be either adjacent to each other (pT181 and pC194 families) or separated by a spacer region of up to 100 bp (pMV158 family) (Fig. 1B). The dsos of plasmids of the same family are characterized by a high degree of conservation in the nic region and by the presence of a less well-conserved bind region. In fact, Rep proteins encoded by different plasmids of the same family can perform in vitro the nicking-closing reaction on the dsos of all the plasmids belonging to the same family, but there is little or no cross-interaction with the bind region, which is indicative of the replicon-specificity of the bind locus. Interestingly, the pT181-encoded RepC initiator has been shown to drive in vitro replication of plasmid pC221, although this was greatly reduced if a competing pT181-dso was present (58). In spite of such in vitro recognition and extensive homologies of the Rep proteins and the dsos of pT181 and pC221, there is no cross-reactivity between the Rep proteins and the dsos of these plasmids in vivo, unless the Rep proteins are overproduced (59).

In the case of the pMV158 family, the DNA sequence of the bind locus was reported to consist of two or three direct repeats (DRs), whose lengths ranged from 5 to 21 bp (35), separated from the nick sequence by an intervening sequence of variable length (Fig. 1B). The dso of pJB01, a member of the pE194 subfamily, contains as the Rep-binding site three 7-bp nontandem DRs located 77 bp downstream from the nick site (60). Interestingly, the existence of distant DRs has not been elucidated in some plasmids of this subfamily (unpublished observation). The role of the different regions of the pMV158-dso in the interaction with the plasmid-encoded RepB initiator protein has been addressed in a systematic study (35, 61–64). RepB binds with high affinity to the bind locus, which is made up of three 11-bp tandem DRs located 84 bp downstream from the nick site. These repeats do not constitute an incompatibility determinant toward pMV158 and seem to be essential for plasmid in vivo replication but not for in vitro relaxation of supercoiled DNA mediated by RepB. A second RepB binding site is located in a region around the nick site, within the nic locus. Characterization of the relative affinity of RepB for the bind and nic loci revealed that the three DRs of the bind locus constitute the primary binding site, whereas the weaker binding of RepB to the nic locus could be involved in recognition of the nick site during initiation of replication (64). In plasmids of the pT181 and pC194 families, the DNA sequences of the bind (IRIII) and nic (IRII) loci are located in contiguous inverted repeats (IR) (Fig. 1B). In pT181, both the spacing and the phasing of IRII to IRIII are crucial for origin functionality (65). In addition, the proximal arm and the central part of the IRIII are important for sequence-specific recognition (65). A similar picture is found in plasmids of the pC194 family.

A typical feature of the nic regions is the presence of secondary structures such as hairpins and cruciform. The Rep nick sequence is generally located on an unpaired region within these hairpins, as exemplified by IRI of pT181 and IR-I of pMV158, which accounts for the requirement of plasmid DNA supercoiling to render the cleavage sequence a suitable ssDNA substrate for replication (66–68). The presence of secondary structures is likely to be involved in efficient recruitment and utilization of the initiator protein. Additionally, binding of the initiator protein to the nic locus could promote the melting of the substrate nick sequence. This seems to be the case in pMV158, where the extrusion frequency of the cruciform involving IR-I is very low at the growth temperature of the plasmid host (37°C) (69). In vitro footprinting experiments performed with supercoiled pMV158 DNA showed that binding of RepB to the nic locus promotes the extrusion of the IR-I cruciform, which in turn indicates that initiation of replication would take place only when specific binding of RepB occurs (64). Genetic analysis of the pC194 dso pointed to the existence of a hairpin located downstream of the nick site (70) that was shown to be important for replication of the plasmid (71). In contrast, RepU, the initiator protein of pUB110, does not require the presence of hairpins for efficient recognition of the oriU. Hairpin II, located downstream from the nick site, seems to be dispensable for initiation of replication of pUB110, although its absence provokes the
accumulation of multimers, which is indicative of the involvement of this structure in termination of replication (72).

Out of the three plasmid family prototypes that have been studied in more detail (pT181, pC194, and pMV158), available information regarding the characteristics of the dso is limited to a few plasmids of different families. In the dso sequences of pJV1, plJ101, and pSN22, three plasmids belonging to the same subfamily inside the pC194 family, three conserved regions were identified: the conserved region I of about 100 bp, located upstream of the nick sequence, which is essential for replication, the nick sequence in region II, and region III, which overlaps with the rep start codon. A co-integration experiment between pJV1 and a plJ101-derived vector allowed the identification of the nick site within the sequence 5′-CTAGGTA-3′ of pJV1, located 159 bp upstream of the start codon of the corresponding rep gene (73). In the case of plJ101 and pSN22, the putative nick sequence identified in region II was 5′-CTTGGGA-3′, which is not identical to that of the pC194 group (5′-CTTGATA-3′) (74, 75).

Plasmid pGA1 isolated from C. glutamicum is the best-studied plasmid of the pCG1 family. The location of the dso was analyzed using the runoff DNA synthesis assay. The site- and strand-specific breakage of pGA1 dsDNA occurred within the nucleotide sequence 5′-CTTGGAT/AT-3′ (where / indicates the nick site) in the distal part of the pGA1 rep gene, which is an atypical position among RCR plasmids (76).

The group of RCR plasmids isolated from hyperthermophilic archaea is constituted of pGT5, pRT1, and pTN1. Plasmid pGT5 is the first plasmid to be isolated from a hyperthermophilic organism, Pyrococcus abyssi, and presents similarities to plasmids from the pC194 family. A sequence of 11 nucleotides (nt) identical to that in the dso of pC194 and related plasmids is located at the 5′ region of the rep gene. The presence of pGT5 ssDNA (corresponding to the putative plus strand) in cell extracts of P. abyssi strongly suggests that pGT5 replicates via an RCR mechanism (57). Plasmid pTN1, isolated from Thermococcus, shows an identical nick sequence to that of pGT5 (5′-TTATCCTTGATA-3′) and is also located in the 5′ region of the rep74 gene (77). Similarly, the dso of pRT1, isolated from Pyrococcus sp. strain JT1, exhibited significant identity to the dso of both pGT5 and pC194, further suggesting that the replication mode of this plasmid is via the RCR mechanism, which was confirmed with the detection of ssDNA replication intermediates (78). The position of the dso nick site in the plasmid pZMX201, considered the prototype of a family of plasmids isolated from halophilic archaea, was precisely determined in the sequence 5′-TCTCGGC-3′ (where / denotes the nick site), which is conserved among the members of the family. Although the heptameric sequence is usually located in the stem region of an imperfect hairpin structure that, in turn, could serve as a target for recognition by the Rep protein, the nick site lies in an unpaired position or near an unpaired nucleotide (56). In addition, the use of a hybrid plasmid system revealed the role of the nucleotides of the conserved nick sequence in the RCR initiation and termination process.

**THE REPLICATION INITIATOR REP PROTEINS**

As mentioned above, Rep proteins involved in plasmid RCR initiation are mainly included in a vast superfamily of HUH endonucleases that catalyze cleavage and ligation of ssDNA by using particular recognition and reaction mechanisms. Besides the Rep class, which also includes Reps from ssDNA coliphages and animal and plant viruses, proteins involved in conjugative plasmid transfer (Mob class or relaxases) and in DNA transposition (transposases) also belong to this superfamily, and all of them exhibit a familial relationship based on several conserved protein motifs (79). The two most relevant motifs are the metal-binding HUH motif, composed of two His residues separated by a bulky hydrophobic residue, and the catalytic motif containing either one or two Tyr residues separated by several amino acids (Fig. 3). Characterization of the biochemical activities of several plasmid RCR initiator proteins has contributed to a better understanding of the molecular events during the initiation and termination of RCR. Initiation of plasmid RCR requires Rep-mediated nicking within the unpaired nick sequence of the nic locus in supercoiled DNA. Rep endonucleases exhibit DNA strand-transfer enzymatic activity and catalyze cleavage and rejoining of ssDNA using an active-site Tyr residue to make a transient 3′-phosphotyrosine bond with the DNA substrate and a free 3′-OH at the cleavage site. Moreover, the resultant 3′-OH not only serves to prime replication but also can act as the nucleophile for strand transfer to resolve the phosphotyrosine intermediate in the termination step of RCR.

The divalent metal ion required for the activity of the HUH enzymes probably coordinates one of the oxygen atoms of the scissile DNA phosphate, polarizing it and facilitating the nucleophilic attack of the hydroxyl group of the catalytic Tyr (80, 81). Curiously, the HUH motif for metal binding is not present in plasmids of the pT181 family or in ssDNA filamentous phages, although in both cases the presence of a divalent metal ion is required for
FIGURE 3  Domain structure of the Rep proteins from RCR plasmids. Predicted and observed secondary structures of the replication proteins of different RCR plasmids and of the Rep proteins from the adeno associated virus (AAV) and bovine papillomavirus (BPV). The amino-terminal end (N) and the number of amino acids are indicated for each of the proteins analyzed. The predicted or observed α-helices and β-strands are represented as red and green bars, respectively. The 310-helices are represented as blue bars. Conserved amino acid residues of the active site involved in metal binding (HUH) and in the endonucleolytic activity are indicated in the protein maps. The conserved Walker A, B, and C motifs are indicated in the proteins with a helicase domain. The limits of the origin binding domain (OBD) and of the oligomerization domain (OD) are indicated in the protein maps of RepB, Rep68, and E1. The additional line below the sequence of RepB, Rep68, and E1 shows the secondary structure present in the crystal structure of the protein (PDB entry code is given in the figure).

Plasmidic Rep proteins were aligned by the metal binding HUH motif. However, viral Reps were aligned with RepB by the all-helical OD domain due to the structural similarity found in this region.  

the enzymatic activity. In the case of the pT181 family, it has been shown that the reactive Tyr188 of RepD, the initiator protein of pC221, cleaves the phosphodiester bond 5'-ApT-3' and remains covalently attached to the 5'-P end generated by the cleavage reaction. The importance of the tyrosyl hydroxyl group was confirmed by substitution of Tyr188 by Phe, since this protein variant retains the sequence-specific DNA-binding activities of wild-type RepD but is unable to attach covalently to the replication origin or participate in the nicking-closing reaction in vitro (32). Similarly, the initiator protein RepA of pC194 forms a 5'-phosphotyrosyl DNA link at the initiation step, and mutations in the catalytic Tyr214 drastically reduced its catalytic activity without affecting RepA binding affinity. RepB of pMV158 does not seem to generate a stable covalent tyrosylphosphodiester bond with its DNA target (62). However, there is experimental evidence showing that RepB, like the filamentous phage gpII, forms a transient covalent complex with the 5'-P end of the cleaved DNA (63, 82). A singular case among Rep proteins of RCR plasmids is represented by Rep75 of pGT5, a protein that exhibits a highly thermophilic nicking-closing activity in vitro combined with an unusual site-specific nucleotidyl transferase (NTT) activity which has not been described for proteins of this type (83). Substitution of the catalytic Tyr448 by Phe caused a severe reduction of the nicking-closing activity in vitro and prevented the formation of the 5'-phosphotyrosyl DNA link without affecting the dsDNA binding activity of Rep75 (84). A second critical residue for the activity of Rep75 is Arg451, as the protein variant Rep75-Arg451Leu exhibits a reduced closing activity and has completely lost the NTT activity (84).

A modular structure based on the presence of at least two domains involved in origin nicking and specific recognition of the dsO sequence has been assumed to be a general feature of the initiator proteins of RCR plasmids (85–88). This assumption could be valid for plasmids of the pT181 family, as the sequence-specific DNA binding and DNA relaxation activities of RepC, the initiator of pT181, are mutationally separable and lie on distant protein regions (86). Taking advantage of this property, the identification of the role of individual monomers of pT181-RepC in RCR was addressed by generating heterodimers of the initiator containing a combination of wild-type, DNA binding, and nicking mutants (89). The results demonstrated that a single monomer of RepC is sufficient for origin-specific binding and nicking. In addition, the monomer involved in sequence-specific binding to the dsO must also nick the DNA to initiate replication (89).

In plasmids of the pMV158 family, a similar assumption was proposed on the basis of the higher degree of amino acid identity found at the N-terminal region than at the C terminus of their Rep proteins, which suggests that their conserved N-terminal moiety would be involved in endonuclease activities, whereas the C termini would be involved in the specific dsO recognition (32). However, resolution of the three-dimensional structure of RepB, the initiator protein of pMV158, by X-ray crystallography and by image reconstruction methods showed a different functional domain organization from that initially proposed. RepB is the first and to date the only published example of an atomic structure of a Rep protein from RCR plasmids or bacteriophages (3, 90). Purified full-length native RepB behaves as a hexamer in solution, as observed in analytical ultracentrifugation assays (91), and is crystallized in the same oligomeric state, forming a toroidal homohexameric ring (90). Each RepB protomer comprises an N-terminal origin binding domain (OBD), which retains the DNA binding capabilities as well as the nuclease and strand-transfer activities of RepB, and a C-terminal oligomerization domain (OD) that forms a cylinder with a 6-fold symmetry in the hexamer (Fig. 4A). Separate expression and purification of the RepB OD and OBD domains demonstrated that the enzymatic and dsDNA binding activities and the oligomerization potential can be uncoupled and confirmed the essentiality of OD for hexamerization (D. R. Boer, J. A. Ruiz-Masó, M. Rueda, M. Pethoukov, D. I. Svergun, M. Espinosa, M. Orozco, G. del Solar, and M. Coll, submitted for publication).

Resolution of the three-dimensional structure of the catalytic N-terminal domain of RepB confirmed the involvement of the conserved motifs in the enzymatic activity of the protein. Based on the protein sequence alignment of Reps of the pMV158 family, a set of five conserved motifs were designated for this particular family of proteins (32). Motifs I, III, and IV correspond to the conserved motifs 1, 2, and 3 of Ilyina and Koonin (79), whereas motifs II and V were new. A divalent metal ion, Mn^{2+}, required for the catalytic activity of RepB, is located at the active site coordinated by four amino acid ligands and a single-solvent molecule. Residues His39 and Asp42, included in motif II and with no function assigned initially, together with the His55 and His57 residues of motif III (HUH), provide the four amino acid ligands necessary for metal ion coordination. The conserved catalytic Tyr99 (motif IV) and the conserved Tyr115 (motif V), which interacts with the metal ligand Asp42 of motif II, complete the list of the conserved residues located in the active center. Residues in motif I
are outside the active center and seem to play a structural role by forming part of the strand β1.

In spite of the presence of the bind locus dsDNA sequence in the cocrystals of pMV158-RepB, it was not possible to extract structural information on the DNA or on the protein elements interacting with it. Nevertheless, the presence of a single and large electropositive region that covers the outer surface and crevice at the N-terminal of the RepB hexamer was consistent with binding to only one bind locus-containing DNA molecule (Fig. 4B). In addition, by using electron microscopy reconstruction methods, it was possible to observe a density that occludes the N-terminal region crevice in RepB hexameric particles exposed to the bind locus (90). These observations point to the location of the DNA-interacting surface in the N-terminal outer region and crevice of the hexamer. In fact, site-directed mutagenesis to some positively charged side chains of helix α2 resulted in OBD protein variants with a reduced dsDNA binding affinity but with an intact nicking-closing activity (90).

Two structures of the pMV158-RepB hexamer were obtained from crystals belonging to a trigonal and a tetragonal form, respectively. In both structures, a short hinge region connects the two protein domains. However, the OBDs do not follow the 6-fold symmetry of the ODs and the positions of the OBDs with respect to the ODs change significantly when the two crystal forms are compared. In fact, analysis of the two structures revealed that the N-terminal OBD domains are found in nine distinct orientations relative to the OD. The conformation plasticity of RepB has been explored by combining techniques such as X-ray crystallography, small-angle

**FIGURE 4** Cartoon representation of the structure of RepB obtained by X-ray crystallography. (A) Top (left) and side (right) views of the RepB hexamer. The locations of the OBD (continuous line) and of the OD (dotted line) are also indicated in the two views. The position of the hinge connecting both domains is indicated in the side view. (B) Top (left) and side (right) views of the electrostatic potential on the solvent-accessible surface of the RepB hexamer structure. The location of the crevice is indicated. doi:10.1128/microbiolspec.PLAS-0035-2014.f4
X-ray scattering, sedimentation experiments, and molecular simulations (D. R. Boer, J. A. Ruiz-Masó, M. Rueda, M. Pethoukov, D. I. Svergun, M. Espinosa, M. Orozco, G. del Solar, and M. Coll, submitted for publication). These studies revealed that the position and freedom of movement of the OBDs is mainly determined by the OBD-OD hinge region, since contacts between adjacent OBDs do not seem to play an important role in fixing their positions. A consequence of the loosely coupled domain arrangement observed in the RepB hexamer is the high level of conformational freedom of the OBDs, which is probably important for formation of a functional replisome.

The hexameric state of purified RepB is unique among plasmid replication initiators characterized so far. Most of these proteins are purified as monomers, as in the case of the Rep proteins of the pC194 family (53), or dimers, such as Reps of the pT181 family (52, 54, 92, 93). Interestingly, some of these proteins tend to form multimers upon binding to the ori DNA, suggesting that oligomerization could be involved in some of the biochemical activities of the initiators (93, 94).

The existence of tight interactions between the pMV158-RepB all-helical ODs enables the formation of a toroidal ring with near 6-fold rotational symmetry and an inner surface narrowing down from a maximum diameter of ∼20 Å to a minimum diameter of ∼13 Å (90). The search for fold similarities in the ODs from different atomic structures, in combination with a comparative analysis of Rep sequences, confirmed the existence of a RepB-like all-helical OD domain responsible for oligomerization in viral Rep proteins and replication initiators from plasmids of the pMV158 family (D. R. Boer, J. A. Ruiz-Masó, M. Rueda, M. Pethoukov, D. I. Svergun, M. Espinosa, M. Orozco, G. del Solar, and M. Coll, submitted for publication). The combination of hexameric OD ring and hinge-connected OBDs has been proposed as a general feature of hexameric replication initiators of the pMV158 family, although this configuration can also be found among hexameric initiators of animal and plant viruses (Fig. 3) (D. R. Boer, J. A. Ruiz-Masó, M. Rueda, M. Pethoukov, D. I. Svergun, M. Espinosa, M. Orozco, G. del Solar, and M. Coll, submitted for publication). In fact, the domain organization of RepB resembles that of the viral replication initiators, which suggests an evolutionary link between the two protein families. Structural similarities between the α-helical hexamerization domain of RepB and the equivalent domain of the papillomavirus E1 helicase have been reported (90). The crystal structure of E1 comprises the OD and helicase domains and was obtained in complex with a short ssDNA oligonucleotide in the central channel (Fig. 3) (95). Therefore, the resemblance of the E1 and RepB hexameric rings favors a mechanistic model in which the RepB ring might close around a DNA region that could have melted on assembly of the protein and/or cleavage at the nick site, thus encircling one of the plasmid strands within the central channel. Subsequent recruitment of a host helicase (perhaps PcrA) would allow further unwinding of the DNA and the concomitant progression of the hexamer along the plasmid. Enclosure of one plasmid strand may confer high processivity to the RepB/helicase/DNA polymerase replisome complex, thereby allowing replication of pMV158 in a broad range of bacterial hosts.

Ilyina and Koonin (79) hypothesized that geminiviruses would descend from bacterial replicons according to the limited sequence similarity of the three conserved motifs of the RCR Reps of geminiviruses and certain plasmids of Gram-positive bacteria. However, more recent phylogenetic analyses of various RCR Reps suggest that Rep proteins of geminiviruses share a most recent common ancestor with Reps encoded on plasmids of phytoplasmas (96). Several plasmids from phytoplasma have been sequenced and characterized to some extent. The Rep protein of pOYW, a plasmid isolated from onion yellow phytoplasma, is characterized by a chimerical nature containing an RCR plasmid Rep-like domain in the N-terminal region, which displays four out of the five conserved motifs characteristic of the pMV158-Rep family of proteins, and a virus-like helicase domain in the C-terminal region, which also includes Walker A and B nucleotide-binding motifs and shows great similarity to Reps from eukaryotic small DNA viruses or some RNA viruses (97). Similarly, the Rep protein of plasmid pCPa, isolated from Candidatus Phytoplasma australiense, also shows four out of the five protein motifs characteristic of the pMV158-Rep family of proteins in its N-terminal region and Walker A and B motifs typically found in geminiviruses in its C-terminal region (Fig. 3) (98).

**CONTROL OF RCR**

**Mechanisms for Controlling Synthesis of the Rep Protein**

Control of plasmid RCR is exerted via regulation of the synthesis of the replication initiator. Other general control mechanisms such as the origin inactivation by handcuffing, which involves coupling between plasmid molecules through Rep proteins bound to multiple initiator binding sites (iterons), has not been reported in RCR.
plasmids. Availability of the Rep protein determines the frequency of the leading-strand initiation, which is rate-limiting for the plasmid RCR process, and therefore the rep gene encoding the initiator is subjected to strict control. Transcriptional and translational inhibition mechanisms of rep expression are not specific to RCR plasmids and have been reviewed (1, 99, 100). Two classes of replication control systems have been described in RCR plasmids: those that use only antisense RNAs and those involving an antisense RNA in combination with a transcriptional repressor protein. An example of the first class is found in the replication control of pT181 by an antisense RNA-mediated transcriptional attenuation mechanism (101), which is also expected to control replication in most of the plasmids of the family. In this system, the antisense RNA transcribed from the cop region targets the rep-mRNA encoding the initiator protein and blocks its expression. The main consequence of the antisense-target interaction is the premature termination of the rep-mRNA transcription due to the formation of a ρ-independent transcriptional terminator (attenuator) located just 5′ to the repC start codon. Release of the antisense RNA from the mRNA target permits refolding and, therefore, silencing of the transcriptional terminator (101).

A second class of copy-number control system reported in RCR plasmids involves two trans-acting plasmid elements, namely a transcriptional repressor and an antisense RNA, that are involved in controlling the synthesis of the initiator at the transcriptional and translational levels, respectively. This dual mechanism was first discovered in pMV158 and seems to be widespread among plasmids from Gram-positive bacteria, including the RCR plasmids of the pMV158 family (99, 102). In pMV158, the transcriptional repressor CopG binds to its own promoter and represses the transcription of the copG-repB operon. The mechanism of transcriptional repression mediated by CopG has been elucidated by analyzing the interactions among the RNA polymerase (RNAP), CopG, and the promoter (103). CopG is able to bind specifically and cooperatively to its operator, which overlaps with the regulated promoter, in such a way that it not only prevents the binding of the RNAP, but also efficiently displaces the polymerase bound to the promoter. The second control element, the rnalII gene encoding RNAII, overlaps the intergenic region of the pMV158 copG-repB operon and is transcribed in the opposite direction to it. Therefore, the entire sequence of antisense RNAII is complementary to a copG-repB mRNA target region that contains the translation initiation signals for the essential repB gene (104). Small antisense RNAII (48-nt long) consists of a single-stranded 5′ tail followed by a hairpin and a 3′ poly(U) tail, the latter two elements constituting a very efficient intrinsic transcriptional terminator (105, 106). Additionally, RNAII is able to inhibit plasmid replication in trans, it has a short (1 to 2 min) half-life (P. Acebo et al., unpublished results), and it determines a strong incompatibility against pMV158, thus matching the features required for an efficient plasmid replication control element (102, 107).

In the antisense RNA-mediated systems of RCR plasmid replication control where the entire process has been dissected, the formation of the RNA/RNA duplex seems to be initiated by a “kissing” step that involves reversible base-pairing between complementary hairpin loops (108). However, wild-type antisense RNAII supplied in trans retains its inhibitory capacity on derivatives of pMV158 that encode a mutant copG-repB mRNA lacking the hairpin complementary to that in RNAII. These findings suggested that formation of a kissing complex was not strictly required for the RNA pairing mechanism of the pMV158 control system (107) and led to the proposal that copG-repB mRNA/RNAII binding initiates via a loop-linear pairing scheme (109). The existence of an antisense RNA that controls replication of pJB01 (a member of the pMV158 family) has also been proved, and the involvement of different regions of this RNA in translation inhibition of the rep gene was studied by mutational analysis. Copy number inspection of the various mutant plasmids suggested that the entire secondary structure of the antisense RNA was important for interaction with the target mRNA (110).

In addition to this dual control of rep expression that senses and corrects fluctuations in plasmid copy number, proper availability of Rep also relies on the adequate functionality of the transcription and translation initiation regulatory signals. In pMV158, translation of repB was proposed to be initiated from what was termed an atypical ribosome binding site (ARBS) located in the small intergenic region of the copG-repB operon encoding the antisense RNAII. The initiation signals that regulate repB translation were identified and characterized in reference 111. Translation of repB relies on its own initiation signals, which rules out a possible mechanism of translational coupling to the upstream copG gene. Only changes in the sequences involving the ARBS proximal box, thus named because of its proximity to the repB start codon, and the region 3′-adjacent to the ARBS proximal box result in a significant reduction of in vitro synthesis of repB. The results of this study demonstrate the importance of the region immediately upstream of the repB start codon in the efficiency of
translation of repB and call into question the functionality of the postulated ARBS. Moreover, the conclusions of this study could be applicable to the majority of the plasmids of the pMV158 family due to the high degree of identity found at the rep translation initiation regions. The possible involvement of those features of the translation initiation signals of the rep genes in an additional mechanism to regulate the level of the Rep protein remains to be determined.

The existence of a singular element that positively influences the plasmid copy number and stability has been reported in plasmid pGA1 from C. glutamicum. On one hand, the copy number of pGA1 is negatively regulated by an antisense RNA at the translational level (112). Additionally, the IR1 sequence, located in the leader region of the rep mRNA, negatively influences the rep gene expression. On the other hand, the per gene, which encodes a positive effector of replication, positively affects the copy number and segregation stability of pGA1, though it was found to be dispensable for replication (47). In fact, deletion of the per gene results in a significant decrease of the pGA1 copy number in C. glutamicum. Furthermore, the related plasmid pSR1 encodes a similar gene product that can act in trans on pGA1 derivatives (47). Per protein has no effect on the expression from the rep promoter, and it has been hypothesized that it counteracts the inhibitory effect of the antisense RNA on the rep gene expression by interacting with it (112). Interestingly, pGA1 still codes for another accessory effector, the small aes gene, which was shown to increase the segregational stability of pGA1 derivatives in the presence of per (113). The genes per and aes are located nearby and transcribed convergently. The authors propose that a cooperative regulation of expression of per and aes genes could be on the basis of the control of the pGA1 plasmid copy number and of its stable maintenance in C. glutamicum.

In plasmid pUB110, expression of the replication initiator gene (repU) is controlled by two small and unstable antisense RNAs, transcribed from a major incompatibility region, that interfere with rep-mRNA translation by targeting the repU translation initiation signals (114). Interestingly, an additional control of plasmid copy number seems to exist in pUB110. In this plasmid, inactivation of a RepU molecule after a replication event has been proposed to occur by incorporation of a short oligonucleotide, in a way analogous to that described for pT181-RepC (see below). The inactive RepU-RepU* hetero-oligomer complex could form a large nucleoprotein structure at the dso region that interferes with transcription from the repU promoter (94). Therefore, even in the absence of a Cop-like transcriptional repressor, the amount of active Rep-like transcriptional repressor available for replication initiation could be subjected to a dual control at the translational and transcriptional level.

**Mechanisms That Restrict the Use of Rep Molecules to a Single Replication Event**

The replication and copy number of RCR plasmids are tightly regulated, and the mechanisms ensuring that the initiator proteins are unable to catalyze multiple rounds of DNA synthesis are critical to achieve such efficient control. In pT181, the RepC dimer is inactivated after the completion of a round of replication by the attachment of an approximately 10-nt oligonucleotide, representing sequences immediately 3’ to the initiation nick site, to the active tyrosine residue of one of its subunits, thus generating a heterodimer termed RepC/C*. Generation of RepC/C* occurs because once the replication fork reaches the reconstituted origin, the nascent leading strand is extended by ~10 additional nucleotides beyond the Rep nick site. Two site-specific transsterification reactions take place then: the first leads to the release of the circular ssDNA intermediate, and the second yields the RepC/C* heterodimer containing the short 3’ leading-strand extension as well as a dsDNA plasmid molecule that has the sealed new leading strand paired to the parental template strand (115, 116). Thus, the second transsterification allows completion of the termination process by avoiding recycling, i.e., continuous synthesis of the leading strand driven by the same initiator molecule (117). Analysis of the different interaction patterns of RepC/C and RepC/C* with the pT181-dso DNA by in vitro footprinting and binding-bending assays revealed that, although RepC/C* retains the ability to bind to DNA, it is unable to promote cruciform extrusion to expose the nick sequence in ssDNA form (118). This may explain why the formation of a RepC/ RepC* heterodimer inactivates the protein in addition to uncoupling termination of leading-strand replication and initiation of a new replicative round.

A different mechanism preventing recycling upon termination of one replication round has been proposed for plasmid pC194. The pC194-encoded RepA protein has two active-site catalytic residues, namely Tyr214 and Glu210, which catalyze the DNA cleavage at the initiation and termination steps, respectively. In this plasmid, the termination reaction catalyzed by the glutamate residue is hydrolysis that does not generate a covalent complex, thus preventing continuous synthesis of the leading strand (53). By contrast, in gene A protein of
phage ΦX174 two catalytic Tyr of a single protomer would perform alternative nicking and nicking-closing reactions following a so called “flip-flop” mechanism that would allow recycling of the initiator protein (119). These two alternative pathways at the termination step of leading-strand synthesis reflect the different lifestyles of phages and plasmids, with replication control featuring the latter. It is worth noting that the lack of formation of a Rep-DNA adduct in the cleavage reaction of the termination stage of the pC194 leading-strand replication implies that the yet-unknown mechanism of inactivation of the used RepA molecules would differ from that reported for the Rep proteins of the pT181/pC221-family plasmids. Remarkably, although the catalytic residues Glu and Tyr of RepA are conserved in the RepU initiator of pUB110, a replication-dependent modification of RepU implying the loss of its catalytic activity has been reported, and the modification has been suggested to consist, as in the case of RepC*, in the covalent attachment of a single-stranded oligonucleotide (94). In pMV158, an inactivation mechanism similar to that described for pT181-RepC does not seem to be applicable, since RepB has not been shown to form a stable covalent complex with the 5′-P end of the cleaved DNA (62). However, generation of the circular ssDNA intermediate during termination of the plasmid leading-strand synthesis requires that the 5′-P end to be ligated is covalently attached to RepB (62). Since this is the 5′-P end generated in the initiation cleavage, the feasibility of preservation of the RepB-DNA bond through the entire leading-strand replication process under certain circumstances is suggested. Therefore, a mechanism analogous to the flip-flop scheme, which involves alternating nicking and nicking-closing by two catalytic residues of a single protomer, is plausible for RepB. As an alternative, the second catalytic residue could be provided by the active site of an adjacent monomer of the RepB hexamer, which would imply that substrates are transferred between OBDS during the termination reactions. In either case, and despite the biochemical and structural data available about RepB, the identification of the second catalytic residue remains undisclosed.

The unusual NTT activity displayed by the pGT5-encoded Rep75 initiator appears to be directly related to the mechanism that prevents plasmid overreplication. The three reactions catalyzed by Rep75, namely nicking, closing, and NTT, can be uncoupled in vitro, although they share part of their mechanisms. Interestingly, replication initiation activities mediated by Rep75 are inhibited at the concentrations of ATP or dATP that promote the NTT activity of the protein. According to the proposed model, an adenine residue could be transferred to the 3′-OH generated after specific cleavage by the NTT of another Rep molecule, thus re-creating a putative nick site. After cleavage at this site, Rep75 becomes attached to an adenine residue through its active Tyr and, therefore, is inactivated. Hence, the role of the NTT activity might be to reduce the intracellular level of active Rep molecules. The authors propose that the levels of pGT5 replication could be regulated by an equilibrium between active and inactive Rep proteins, itself determined by the intracellular ATP and dATP pools and the availability of free 3′-OH and host replication proteins (84).

THE SINGLE-STRAND ORIGIN

Synthesis of the lagging strand of RCR plasmids initiates from the so-called single-strand origin (sso), a noncoding region containing long and imperfect inverted repeats that form complex secondary structures as the parental leading strand is displaced and becomes an ssDNA replicative intermediate (12, 120). In general, the ssos are located a short distance upstream of the dos and, hence, these elements are among the latest plasmid regions to become single-stranded during replication of the leading strand. This relative position might play a role in preventing run-off synthesis of the lagging strand before closing of the displaced parental strand upon termination of the leading-strand synthesis. The sso is recognized by host factors (usually RNAP) that make a small RNA for priming the lagging-strand synthesis. Functionality of the sso is orientation-dependent, which points to a crucial role of unpaired sequences within the secondary structure of these elements.

Five main types of sso (ssoU, ssoA, ssoT, ssoW, and ssoL) have been reported that differ from each other in structure, in sequence motifs highly conserved among the members of the same group, and in the host range in which they are functional.

The ssoA-type origins consist of a single and long (∼150-nt) hairpin structure containing internal and bulge loops, in addition to the terminal loop. These origins display a high degree of sequence heterogeneity, with only two well-conserved regions: the previously termed recombination site B (RSb), which is present in the lower stem of the hairpin of all the ssoAs, and a 6-nt consensus sequence (CS-6) that is located in the terminal loop of the hairpin of most but not all of these origins (12, 120, 121). The RSb has been shown to be involved in binding of RNAP to the ssoA (122). Mutations in the RSb abolish almost completely the ssoA-dependent synthesis of the
lagging strand both in vitro and in vivo, giving rise to the accumulation of very high numbers of plasmid ssDNA intermediates in the host cell (121, 123). The CS-6-containing terminal loop seems to function as a terminator of the primer RNA synthesis, and changes in this region lead to a moderate increase in the intracellular number of ssDNA plasmid forms without impairing ssoA binding by RNAP (121–123). Sequences similar to CS-6 have been identified in the terminal loops of hairpins located in DNA regions that have been either proposed or shown to act as sso sites in plasmids from Actinobacteria (124–128). In addition to the RsB and CS-6 conserved motifs, sequences resembling the consensus -35 and -10 promoters regions have been identified in various ssoAs (129). A given ssoA only functions efficiently in its natural host or in a few closely related species (12, 120, 129).

The ssoW was identified in lactococcal plasmid pWVO1 (130, 131). This origin is located in a 250-bp DNA fragment that contains two inverted repeats, IR I and IR II. IR I shows homology to the ssoA-type origins since it harbors the CS-6 sequence in the terminal loop as well as a sequence similar to the RSb at the lower stem. Also, the upper stem of IR I shares remarkable sequence similarity with the ΦX174 minus-strand origin, which is recognized by the primosome for priming synthesis of the complementary strand (132). Full ssoW activity requires both IR I and IR II, and conversion of ssDNA into dsDNA from the entire element is only partially inhibited by rifampicin. IR II has no activity on its own, whereas IR I has a partial, RNAP-independent activity for complementary strand synthesis. Thus, priming of the lagging strand from ssoW seems to occur through two different pathways: one is catalyzed by RNAP and requires the entire origin, and the other, which was suggested to involve a primosomal complex, only requires IR I. Efficient functionality of the pWVO1-ssoW seems to be confined to lactococci (131).

The ssoT-type origins have been found in plasmids isolated from Bacillus spp. (133–135). In fact, the majority of the RCR plasmids from Bacillus subtilis harbor an ssoT (136). The minimal ssoT, as defined for the B. subtilis plasmid pBAA1, spans 120 to 190 bp, encompassing three imperfect palindromes that would give rise to hairpins I, II, and III on ssDNA. Results from mutation analysis suggest that the structure of hairpin III and both the structure and sequence of hairpin I are required for full activity of the ssoT (135). Comparison of the ssoT of pBAA1 with the homologous region of the B. thuringiensis pG2 plasmid showed the existence of three conserved sequence motifs located in the loop of hairpin I (motifs 1 and 2) and in the intervening DNA between hairpins II and III (motif 3). Motif 1 has been shown to play an important role in the activity of the ssoT, while the role of motif 3 is controversial and that of motif 2 has not been proved. Initiation of lagging-strand synthesis from the ssoT is RNAP dependent, as can be inferred from the intracellular accumulation of ssDNA plasmid forms upon addition of rifampicin (135). It is also worth noting that the pBAA1-ssoT functions in both B. subtilis and S. aureus, hence showing a broader host range than the ssoA- and ssoW-type origins (135). Based on sequence similarity, the ssoTs of a number of B. subtilis plasmids have been classified into two groups, palT1 and palT2, each including almost identical ssoT origins. The palT1 group includes the ssoTs of plasmids pTA1015, pTA1020, pTA1060, pLS11, and pBAA1, whereas the palT2 group includes those of pTA1030, pTA1040, and pTA1050 (136). The DNA region involved in initiation of the lagging-strand synthesis in B. thuringiensis plasmid pTX14-3 was found to be homologous to the pBAA1 ssoT (137). Actually, sequences highly similar to motifs 1, 2, and 3 are present in the pTX14-3 origin. Curiously enough, the activity of the sso of pTX14-3 is at least partially resistant to rifampicin, which suggests the existence of a lagging-strand priming mechanism independent of host RNAP (138).

The ssoU is the most promiscuous sso origin characterized so far, as it seems to be fully functional in many, if not all, Firmicutes (121, 139–142). The ssoU origin was first identified in staphylococcal plasmid pUB110 (142), within a ~250-bp DNA fragment with the potential to form several hairpin structures containing symmetric and asymmetric internal loops in addition to the terminal one (139). Sequences nearly identical to the pUB110-ssoU were subsequently found in the streptococcal pMV158 and Bacillus pTB913 plasmids, and the involvement of these elements in lagging-strand synthesis was also proved (143). The high level of sequence identity among the ssoUs of plasmids isolated from different bacterial genera is consistent with this kind of sso being efficiently recognized in a broad range of hosts. The promiscuous activity of the ssoU has been suggested to be accounted for by the proven ability of this origin to bind efficiently to RNAP from different bacteria (144). No sequences with significant similarity to the canonical -35 and -10 promoter regions have been found in the ssoU. In vitro, RNAP binds to the left one of the two large hairpins of the ssoU origin, whereas the transcription initiation of the primer RNA and the transition from RNA to DNA synthesis occur, respectively, at the 3′ end of the right large hairpin and at the 3′ arm of its stem (144).
In spite of the sequence heterogeneity between the four types of sso element described above, they all exhibit an RSB-like motif partially unpaired in the bottom part of the stem of a hairpin structure. Remarkably, the RSB-like motif of the ssoU origin overlaps the region of the hairpin contacted by RNAP (144).

Apart from the sso elements mentioned above, a number of RCR Lactobacillus plasmids carry lagging-strand origins showing high levels of similarity among each other but not with the ssoU element found in plasmids from other bacteria genera. This kind of origin (ssoL) spans ~100 bp, being considerably smaller than those of the other groups (145).

Unlike the dsoL, the ssoU is not essential for plasmid replication, provided that an alternative pathway exists for priming lagging-strand synthesis. In general, alternative plasmid-borne signals and/or bacterial mechanisms seem to exist that can partially overcome the lack of the genuine ssoU (12, 121, 139). Especially efficient appears to be the predominant ssoR- and RNAP-independent priming system revealed in Streptomyces lividans and proposed to result from the stabilization of RCR plasmids lacking ssoU in this bacterium (146). In a few cases, however, the in dispensability of the genuine ssoU origin has been reported, as is the case with the Nocardia plasmid pYS1 (127).

Although, as stated above, the ssoU are usually not strictly required for replication, removal of a DNA region encompassing them not only leads to an increase of the ssDNA replicative intermediates, but frequently results in a reduction of the copy number (measured as dsDNA forms) and in segregational instability of the plasmids (12, 147, 148). The stability function linked to the ssoU does not seem, however, to rely on the efficient conversion of ssDNA to dsDNA, since stable inheritance is observed for plasmids whose ssoU lacks activity in a given host, in spite of the inefficient conversion of ssDNA replicative intermediates into dsDNA plasmid forms. Removal of the genuine ssoU region usually leads to unstable plasmid inheritance even in hosts where the lagging-strand origin is not functional at all (148). Thus, although efficient conversion of ssDNA to dsDNA does not guarantee plasmid stability (149), the presence of an sso element appears to contribute to the efficient replication and accurate inheritance of the plasmid molecules and, therefore, to plasmid fitness. Remarkably, all natural RCR plasmids contain at least one sso element that functions efficiently in their natural host, and plasmids isolated from different bacteria and belonging to the same replicon family (i.e., sharing homology at their dsoU and rep genes) show no conservation at their ssoU. These observations suggest that acquisition of an active sso upon plasmid entrance in a new host may improve not only plasmid fitness itself, but also fitness of the plasmid-containing bacteria, hence enabling them to overgrow cells that contain plasmids without a functional ssoU. In this sense, it is noteworthy that intracellular accumulation of ssDNA, which can arise from the presence of a plasmid lacking a functional ssoU, has been reported to induce bacterial stress responses such as the E. coli SOS system (150, 151). This could result, in turn, in a decreased growth rate of the cells and hence impaired bacterial fitness.

Streptococcal plasmid pMV158 is unique, among many other things that have appeared throughout this chapter, in that it contains two lagging-strand origins, namely streptococcal-specific ssoA and promiscuous ssoU. It has been shown that whereas the pMV158-ssoU participates in the plasmid mobilization between different bacterial species, the pMV158-ssoA would be involved mainly in intraspecific transfer (140). The presence of both ssos might reflect the evolutionary story and lifestyle of this mobilizable, highly promiscuous plasmid.

**HOST PROTEINS INVOLVED IN RCR**

**Participation of DNA and RNA Polymerases**

In addition to the critical role frequently played by bacterial RNAP in recognition of the ssoU and priming of the lagging-strand synthesis (see above), several pieces of experimental evidence point to the direct participation of DNA polymerases (Pol) I and III in plasmid RCR.

Analyses of the involvement of Pol I in leading- and lagging-strand synthesis during RCR were performed on the pMV158 model system, by employing plasmid derivatives that carried the ssoA and lacked the ssoU. These studies took advantage of the previous characterization of Pol I of S. pneumoniae and the construction of pneumococcal polA mutants lacking the polymerase activity of this enzyme (152, 153). Mutant strains depleted in the 5’-3’ exonuclease function of Pol I could not be obtained since this activity was found to be essential for cell viability of S. pneumoniae (152). Therefore, only the involvement of the polymerase function of Pol I in plasmid RCR could be tested.

Participation of the polymerase activity of Pol I in the initiation of the lagging-strand synthesis was inferred from the increased fraction of ssDNA plasmid forms accumulated within the cells of the polA mutant strains compared to the wild-type strain (154). The same conclusion was drawn from the analysis of the in vitro replication of ssoA-containing ssDNA in cell-free extracts.
prepared from wild-type or polA-deficient pneumococcal strains. Wild-type levels of in vitro replication of plasmid ssDNA in extracts from the polA-depleted strain were obtained only upon complementation with the entire pneumococcal Pol I but not when a protein variant lacking the 5′-3′ exonuclease activity was added, indicating that both the polymerizing and the exonuclease domains are required for efficient lagging-strand synthesis (122). A role of Pol I during the termination of the lagging-strand synthesis was suggested from the comparative analysis of single-strand discontinuities detected in the DNA of the ssoA-containing pMV158 derivative extracted from the wild-type and polA mutant strains of S. pneumoniae (154). In the polymerase-deficient mutants, a discontinuity in the vicinity of the ssoA origin was observed that could arise from plasmid molecules in which the conversion of ssDNA to dsDNA was not completed as a consequence of the defective replacement of the primer of the lagging strand.

An additional DNA discontinuity only detected in the pMV158 derivative replicating in the pneumococcal polA mutants mapped at around the nick site of the plasmid dso, thus pointing to the participation of the polymerase activity of Pol I in an early step of the synthesis of the leading strand (154). Finally, a critical role of the host Pol III replicase (the one containing the PolC polymerase) in plasmid RCR was inferred early from the inhibition of the in vitro replication of the pT181 DNA in S. aureus by hydroxyphenylhydrazinouracil, an antimicrobial agent that acts specifically on the Gram-positive bacterial PolC polymerase (155).

Role of Superfamily Group 1 Helicases and Their Interaction with the Plasmid-Encoded Rep Initiator

As previously described, plasmid RCR starts with the recognition of a dso by the replication initiator protein, which binds, cleaves, and remains covalently bound to the 5′-end of the nicked DNA strand. After these initial steps, several cellular-encoded proteins are recruited to continue the replication of the plasmid. The versatility of the interaction between these proteins and the replication initiator protein determines whether this plasmid is successfully replicated and maintained in a broad range of hosts or whether it is lost. One of the required proteins for the replication of the leading strand is a helicase, which is capable of unwinding the dsDNA ahead of the replication fork. The helicases responsible for RCR are PcrA for Gram-positive bacteria and UvrD in Gram-negative bacteria, both of which belong to the superfamily 1 group of helicases (156, 157). These proteins share 42% sequence identity and are structurally similar (158–161), both possessing the seven conserved helicase motifs. In both cases, there is some controversy regarding the active oligomeric state of the protein: some authors have suggested that PcrA and UvrD are monomeric helicases (162, 163), while others have shown that they are active only as a dimer (164, 165).

The involvement of PcrA in RCR was first reported by Iordanescu and Basheer (166), who found that plasmid pT181 was unable to replicate in a strain of S. aureus carrying a mutation (pcrA3) within the pcrA gene. Subsequently, a mutation in the gene encoding the replication initiator protein RepC (Asp57Tyr) was identified, which suppressed the pcrA3 mutation and restored pT181 replication in the pcrA3 mutant (167). From these data it was proposed that the pcrA3 mutation may impede replication by disrupting the PcrA3-RepC wild-type interaction, thereby preventing separation of the dsDNA, while the interaction is restored by the RepC-Asp57Tyr mutant, hence rescuing plasmid replication. However, it has since been shown by pull-down experiments that the PcrA3 mutant protein was able to interact with both RepC wild type and the Asp57Tyr mutant (168). Rather, the authors showed that the pcrA3 mutation leads to a threonine to isoleucine change in residue 61, which is located in the conserved motif Ia of the superfamily 1 helicases and results in weak ATPase activity, which prevents the unwinding of the pT181 DNA in the presence of both RepC wild type or Asp57Tyr mutant (168). The authors postulated that the viability of the pcrA3 mutant and the capability of the double mutant pcrA3-repCAsp57Tyr to replicate the pT181 plasmid in vivo may be explained by additional cellular factors required for the replication to take place in this strain or by the role of an alternative cellular helicase which may replace PcrA. However, these hypotheses have yet to be confirmed.

PcrA from S. aureus is able to hydrolyze ATP, dATP, dGTP, dCTP, and TTP, and its NTPase activity is increased in the presence of either ssDNA or RepC covalently bound to the oriC of pT181 (169). In addition, the unwinding of supercoiled plasmid pT181 is only achieved when RepC is covalently attached to the origin of replication and in the presence of ATP in the reaction (169). PcrA is only capable of unwinding a dsDNA when there is a 3′ or a 5′ single-stranded tail exposed. Hence, PcrA from S. aureus has a bipolar 3′-5′ and 5′-3′ helicase activity (169, 170).
Two hypotheses have been postulated for the direction of unwinding achieved by PcrA. As mentioned above, PcrA from *S. aureus* may have a dual helicase activity, whereas PcrA from *Bacillus stea rothermophilus* has been shown to only unwind dsDNA in a 3′-5′ helicase direction (158). This disparity may be explained by several factors, such as the use of different constructs/protocols for protein purification, the absence in *vitro* of additional cofactors that would favor the unwinding in a specific direction, or the differences in sequence between PcrA from the two bacteria (59% sequence identity), among others.

PcrA from *B. stea rothermophilus* is able to bind to a nicked dsDNA containing the *oriD* sequence from the pC221 plasmid, then engage the 3′-OH end and translocate in a 3′-5′ direction along the nicked strand, thus unwinding the DNA. However, in the presence of RepD, the helicase is loaded on the opposing strand (the continuous strand), translocating in the same direction (171). The directionality of PcrA translocation was also confirmed to be 3′-5′, using atomic force microscopy of a linearized plasmid containing the *oriD* sequence at different positions relative to the DNA ends. The *oriD* was nicked by RepD, and the unwinding of the DNA was followed by the appearance of condensed ssDNA (171). Later, the kinetic parameters for PcrA helicase activity were determined in bulk and single-molecule experiments, using linear and supercoiled DNAs containing the *oriD* sequence, in the presence of RepD and saturating concentrations of ATP (172). Under these experimental conditions, the unwinding speed of PcrA was 30 bp s−1, while the translocation rate on ssDNA was 99 bases s−1. The unwinding rate is dependent on the amount of ATP in the reaction and on the presence of RepD: in the absence of the replication initiator protein the number of unwinding events was reduced by more than 10-fold (162). Moreover, the affinity of PcrA for partial duplex DNA increased by one order of magnitude when RepD was already bound to the DNA, from a *Kd* of 22 nM to 170 nM (172).

The recruitment of the helicase onto the replication initiation site has also been shown by footprinting experiments using exonuclease (Exo) III (173). RepD binding to *oriD* creates an area of protection to ExoIII digestion which extends beyond the *oriD* region: ~74 to 80 bp upstream of ICR I for the continuous strand and ~46 to 50 bp downstream of ICR III for the nicked strand (173), although further resistance points can be found within the *oriD*, after longer digestion times. When PcrA was incorporated into the reaction, the helicase was recruited upstream of ICR I, which served to stabilize the complex. The RepD-PcrA complex covered the region spanning from 80 bp upstream of ICR I on the continuous strand to the limit of ICR III in the nicked strand, and this complex was not displaced by ExoIII. However, a drastic change in the ternary complex was observed when a nonhydrolyzable nucleotide (ADPNP) was included in the reaction. Under these circumstances, the resistance to ExoIII digestion was located only on ICR II, indicating that there were important conformational changes at the *oriD* once the helicase had begun to unwind the plasmid. Furthermore, these three stages of protein loading on the DNA have been studied by atomic force microscopy, where RepD appears as a globular particle which bends the DNA fragment around 90° in 39% of the *oriD* fragments analyzed. However, when both proteins were bound to the DNA, the proportion of bent DNA increased up to 60%, while ADPNP decreased this percentage to 41% (173).

PcrA helicase has also been identified and characterized in other Gram-positive microorganisms such as *S. pneumoniae*, *Bacillus anthracis*, or *Bacillus cereus*. In all these strains, the role of PcrA in RCR has been studied using the broad host range pT181 plasmid as a reference (31, 174). A clear interaction of the helicase with the replication initiation protein was successfully observed by pull-down for all three PcrAs. However, they displayed different unwinding activity on the plasmid DNA in the presence of RepC: while PcrA from *B. anthracis* and *B. cereus* fully unwound the DNA, PcrA from *S. pneumoniae* failed to do so and produced only partial unwinding. This may indicate that the interaction between RepC from *S. aureus* and PcrA from *S. pneumoniae* is not sufficiently stable and, hence, the helicase is unable to continue DNA unwinding. In fact, the authors reported that it was not possible to maintain the pT181 plasmid in the *S. pneumoniae* strain (31).

In Gram-negative bacteria, a role of UvrD in plasmid RCR has been postulated, although its role has been much less studied than that of PcrA in Gram-positive bacteria. It has been shown that deletion of *uvrD* in *E. coli* results in the accumulation of nicked pC194 plasmid DNA in the cell and a lack of ssDNA intermediates of plasmid replication (175). Nevertheless, a direct interaction between the replication initiator protein and UvrD or its effect on the helicase activity has yet to be studied.

In summary, the effect of the replication initiator protein on PcrA ATPase/helicase activity has been thoroughly studied at the biochemical and molecular biology level. However, there are still some unanswered questions regarding how the replication machinery is loaded onto the replication initiation origin. For example, it has been
reported that PcrA interacts with RNAP through its disordered but highly conserved C-terminal region (176), but it is not known which domains of PcrA and the replication initiator protein are implicated in the interaction between both proteins. In addition, the stability of the helicase-Rep complex is not known, nor is it known whether there are cycles of loading/unloading of the helicase during the entire round of plasmid replication. Furthermore, unlike pT181, plasmids belonging to the pC194 or pMV158 families are able to replicate in pcrA3 bacterial mutants (177; Ruiz-Masó et al., unpublished results). This raises the question of whether the helicase activity of PcrA is responsible for the replication of these plasmids or if another helicase can cope with this task. Finally, despite a great deal of work on the helicase activity of UvrD alone (160, 163, 164, 178, 179), as previously indicated, little is known about its role in RCR and in the putative interaction with the replication initiator protein.

CONCLUDING REMARKS

In this article we have tried to compile several aspects of the biology of the small RCR plasmids that constitute, per se, an extremely interesting family of replicons. Indeed, they link the primitive forms of self-replicating molecules harboring just the information needed for their replication in the host to more sophisticated beings that are interconnected to the viral world. Although much is known, as we have reflected here, about the replicative mechanism and control of RCR plasmids, still a full horizon expands before we can truly assert that we do know these molecules. We envisage several approaches to the understanding of the biology of these molecules: (i) mechanistic studies; (ii) solution of three-dimensional structures of Rep proteins; (iii) characterization of regulatory nucleoprotein complexes; (iv) solution of ternary complexes such as DNA-Rep-PcrA or DNA-Rep-DNA polymerase; (v) involvement of host-encoded factors, other than the obvious roles of DNA- and RNA-polymerase, DNA helicases, and ssDNA binding proteins (with an assumed but not demonstrated participation in plasmid replication); and (vi) comprehension of the mechanisms of adaptation of plasmids to a new host. We could speculate that we have just started uncovering the biology of RCR plasmids. Their role in bacterial adaptation to a changing world manipulated by humans and their contribution to the fitness of bacteria to their niches and to the biodiversity in the microbial world is totally unknown. Let’s not allow the funding agencies to say that there is no new relevant information in the plasmid world: it is fully untrue.

ACKNOWLEDGMENTS

We are very grateful to Luis Blanco, Juan Alonso, and the members of the REDEEX consortium for fruitful discussions and help.

Funding by the Spanish Ministry of Economy and Competitive-ness (grants CSD2008/00013-INTERMODS to M.E.; BFU2011-22588 to M.C.; Ramón and Cajal subprogramme RYC-2011-09071 to C.M.; BFU2010-19597 to G.S.; Complementary Action BFU2008-00179-E-REDEEX to G.S.) are acknowledged.

Conflicts of interest: We disclose no conflicts.

REFERENCES


ASMscience.org/MicrobiolSpectrum


