JOURNAL OF BACTERIOLOGY, Apr. 1989, p. 2195–2201
0021-9193/89/042195-07$02.00/0
Copyright © 1989, American Society for Microbiology
Vol. 171, No. 4

Nucleotide Sequence Analysis and Expression of the Minimum REPI Replication Region and Incompatibility Determinants of pColV-K30

JOSE F. PEREZ-CASAL,† ALISON E. GAMMIE, AND JORGE H. CROSA*

Department of Microbiology and Immunology, The Oregon Health Sciences University, Portland, Oregon 97201

Received 24 October 1988/Accepted 18 January 1989

We sequenced the minimum REPI replication region and the incompatibility determinants of pColV-K30. The minimum replication region contains an open reading frame which corresponds to a 35-kilodalton (kDa) protein. Sodium dodecyl sulfate-polyacrylamide gel electrophoretic analysis with maxicells transformed with a number of deletion derivatives demonstrated that this replication region encodes a 39-kDa protein and also established the direction of transcription of the RepI protein gene. The 39-kDa polypeptide was identified as the *trans*-acting factor essential for replication of REPI-containing plasmids. A translated region of the nucleotide sequence of the RepI protein gene showed homology with the helix-turn-helix binding domains of a number of DNA-binding proteins and also with other plasmid replication proteins. Further nucleotide analysis of the REPI region revealed the presence of direct and inverted repeat sequences in the *incE*, *incF*, and *ori* regions. The REPI *ori* also contained a perfect DnaA-binding site in addition to a high frequency of occurrence of the DNA adenine methylation (*dam*) site 5′GATC3′.

Plasmid pColV-K30 encodes the very efficient aerobactin iron assimilation system (51). The aerobactin genes are flanked by two copies of the insertion sequence IS*1* and two distinct replication regions, REPI and REPII (31). Recent analysis demonstrated the conservation of genes encoding the aerobactin iron transport system, together with the upstream REPI replication region, in 11 ColV plasmids for which pColV-K30 is the prototype (50). The conserved aerobactin-REPI sequences are thought to be part of a virulence factor replication unit which might facilitate the preservation and epidemiological spread of the aerobactin iron transport system in bacteria.

Plasmid pColV-K30 belongs to the IncFI incompatibility group (4), whose main representative is plasmid F. F contains two complete replication regions: RepFIA, present in the 10-kilobase (kb) *Eco*RI fragment f5 (25), and RepFIB, located in the 7.6-kb *Eco*RI fragment f7 (22). RepFIA is homologous to REPII (31), and RepFIB is homologous to REPI (32). Bergquist et al. reported that the RepFIB region was conserved among members of the IncFI group to a greater extent than the RepFIA sequences (4, 5). The RepFIB-like replication regions present in f7 and in plasmid P307 have been characterized. In f7, the minimum replication region and the *incE* sequences, claimed to be responsible for the IncFI incompatibility functions of these plasmids, were located in a DNA fragment of about 2.2 kb (14). In P307, the replication sequences mapped within the 4.3-kb *Eco*RI fragment E11 (33). Both replication regions showed a high degree of homology with a 2.2-kb fragment containing the REPI replication region and the associated incompatibility sequences (32).

A more detailed analysis of the REPI replication region confirmed that a DNA fragment of about 1.8 kb contains the entire replicon. Analysis of the replication products indicated that REPI encodes a *trans*-acting factor essential for the replication of REPI-containing plasmids. REPI also possesses two distinct incompatibility sequences, *incE* and the novel incompatibility region *incF*. REPI-par sequences mapped to a DNA fragment located 20 kb downstream of the aerobactin iron transport system of pColV-K30 (32).

In this work, we present the complete nucleotide sequence of the minimum REPI and incompatibility regions *incE* and *incF*. We identified a polypeptide of 39 kilodaltons (kDa) as the *trans*-acting factor essential for replication of REPI plasmids. A translated region of the nucleotide sequence of this protein gene shows homology with the helix-turn-helix binding domains of a number of DNA-binding proteins and with other plasmid replication proteins. Further nucleotide analysis of the REPI region revealed the presence of direct and inverted repeat sequences in the *incE*, *incF*, and *ori* regions which might be involved in the replication mechanisms of REPI-containing plasmids.

(These findings were previously presented in part [A. E. Gammie, J. F. Perez-Casal, D. H. Farrell, and J. H. Crosa, Abstr. Annu. Meet. Am. Soc. Microbiol. 1988, H-117, p. 164].)

MATERIALS AND METHODS

Bacterial strains. The *Escherichia coli* strains used in this study were HB101 (F$^-$ *hsdS20 recA13 ara-14 proA2 lacY1 galK2 rpsL20 xyl-5 mtl-1 supE44* [7]), C2110 (*polA1 rha his* [23]), JM107 (*lac pro thi gyrA96 endA1 hsdR17 relA1 supE44* F′ *traD36 proAB lacI^q Z* ΔM15 [38]), and BN660 (F$^-$ *thr-1 leuB6 proA2 lon-22 his-4 thi-1 ara-14 lacY1 galK2 xyl-5 mtl-1 rpsL31 tsx-33 recA srl-300*::Tn*10* [34]).

Molecular cloning experiments. Purification of plasmid DNA was performed as described previously (32). Restriction endonucleases were used as recommended by the supplier (Bethesda Research Laboratories, Inc., Gaithersburg, Md.). All of the clones used in the maxicell analysis were described previously (32). Cloning of the REPI fragments for sequencing was performed by insertion into the M13mp18 and M13mp19 bacteriophage systems described by Vieira and Messing (47).

* Corresponding author.
† Present address: Department of Microbiology and Immunology, Emory University, Atlanta, GA 30322.

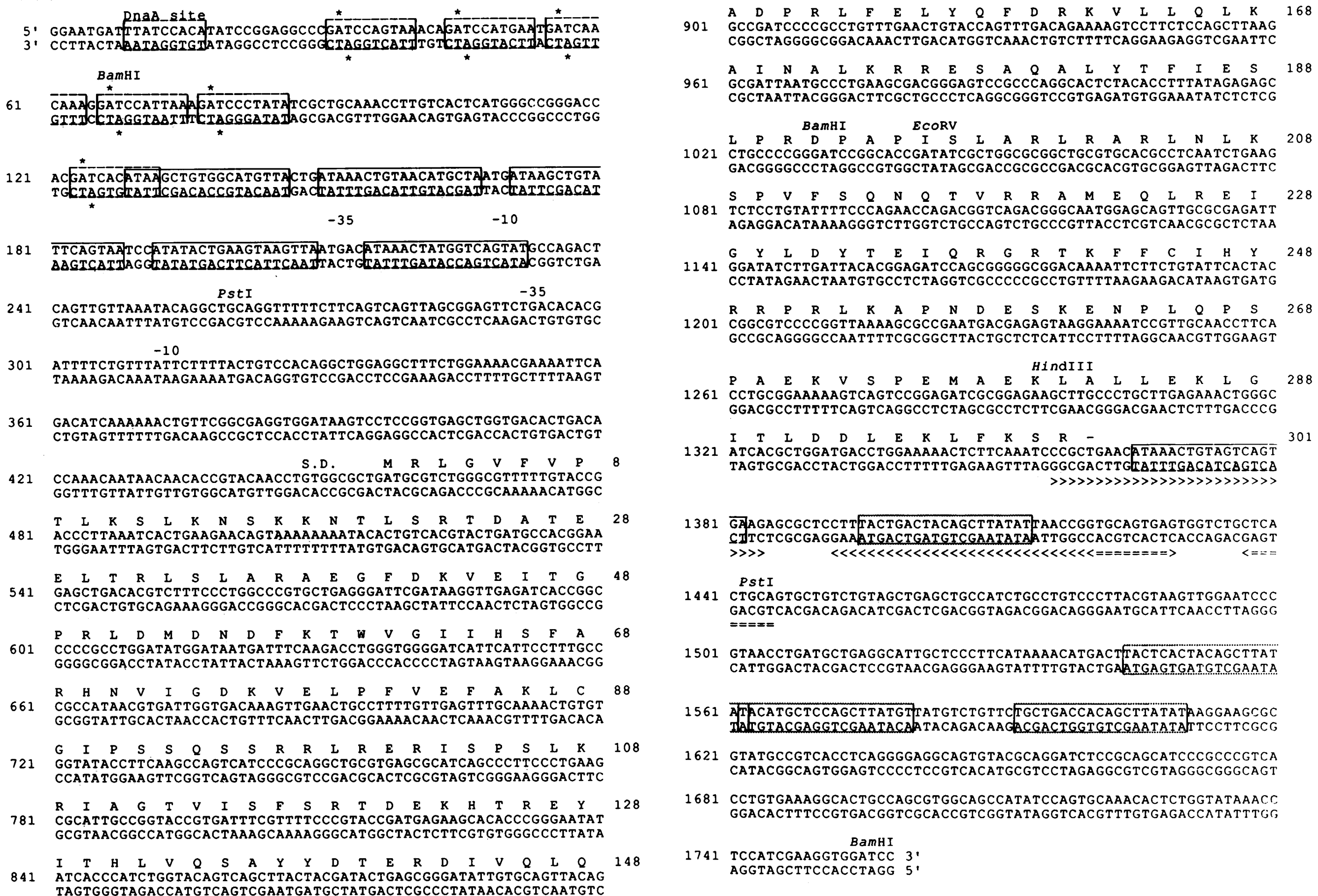

FIG. 1. Nucleotide sequence of REPI and the incompatibility regions. The complete nucleotide sequence of the 1,758-bp DNA fragment containing the minimum REPI and the *incE* and *incF* incompatibility regions is shown. Nucleotide numbers are on the left. The adenine residues around the origin which are methylated (★), the perfect DnaA box, and the various direct repeats (boxes) discussed in the text are highlighted in the sequence. Possible RNA polymerase-binding sites are marked above the lines as −35 and −10. S.D. refers to the putative ribosome-binding site. ORF1 begins at nucleotide position 457, and ORF2 begins at position 613. The amino acid sequence represented by single-letter code is shown above the sequence, and the amino acid numbers are on the right. The ORF is followed by two sets of inverted repeated sequences (>>> <<< and ==> <==).

Maxicell experiments. The maxicell strain used was *E. coli* BN660 (34). Maxicells were prepared as described previously (37).

Sequencing of DNA. We used the sequencing procedure of Sanger et al. (38). The region was sequenced on both strands. Nucleotide analysis was conducted with the Pustell Sequence Analysis Programs, International Biotechnologies, Inc. Homology searches for DNA-binding capacities were done with the Amino Acid Search Program developed at The Institute of Molecular Biology, University of Oregon, Eugene.

RESULTS

Nucleotide sequence of REPI and incompatibility regions *incE* and *incF*. The nucleotide sequence of REPI and the incompatibility regions was determined as described in Materials and Methods. The complete nucleotide sequence of a 1,776-base-pair (bp) fragment containing REPI and the *incE* and *incF* incompatibility regions is shown in Fig. 1. Previous cloning experiments (32) localized the minimum REPI region and the incompatibility regions. The minimum REPI region is within a DNA fragment of 1,299 bp (bases 1 to 1,299), incompatibility region *incE* is found in a DNA fragment of 412 bp (bases 1031 to 1443), and region *incF* is present in a DNA fragment of 332 bp (bases 1444 to 1776).

Computer analysis revealed two major open reading frames (ORFs) within the REPI region when the standard ATG start codon at positions 457 and 613 was used. Both ORFs have the same reading frame but begin with different start codons. Other start codons (GTG) were found at positions 395, 409, and 448. All of the above-described start codons allow for ORFs, but for reasons given below we analyzed only those that began with ATG. The start codons (ATG) for each ORF are preceded by sequences with low homology to the consensus Shine-Dalgarno ribosome-binding sequence (40). ORF1, corresponding to bases 457 to 1359, codes for a 35-kDa protein, while ORF2, spanning bases 613 to 1359, codes for a protein of 29 kDa. The estimated pI of both translated ORFs is 10.69. These data suggest that RepI is a basic protein. The region upstream of these ORFs has sequences which share homology with the consensus for the *E. coli* promoter (35). The REPI ORFs are followed by DNA sequences located between bases 1362 and 1445 which are capable of forming two stem-loop structures. These sequences may act as a signal for termination of transcription or confer stability to the RepI transcript.

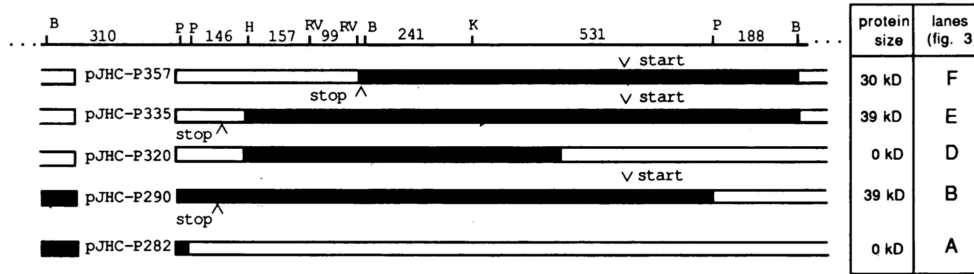

FIG. 2. Maps of clones used in maxicell analysis. Restriction sites: B, *Bam*HI; P, *Pst*I, H, *Hin*dIII; RV, *Eco*RV; K, *Kpn*I. The dark bars represent REPI regions, and the open bars represent pBR322 regions. The arrowheads pointing down correspond to the start positions, and the arrowheads pointing up correspond to the stop sites of the proteins. Estimated protein sizes and lane designations corresponding to those of Fig. 3 are on the right. kD, Kilodaltons.

Analysis of polypeptides encoded by the REPI region. We reported previously that the minimum REPI replication region encodes a *trans*-acting factor that is essential for the replication of REPI-containing plasmids (32). Furthermore, nucleotide sequence analysis revealed the presence of ORFs within the REPI region. To identify the actual polypeptides encoded by this region and to assess which is the essential *trans*-acting product(s), we transformed the maxicell producer strain *E. coli* BN660 with several clones derived from REPI. The results are shown in Fig. 2 and 3. The clones used were described previously (32). As a control we used pBR322 (6), the vector for these clones.

The results (Fig. 2 and 3) indicated that, in addition to the proteins produced by the vector, pJHC-P335 encodes a polypeptide of 39 kDa, whereas pJHC-P357 codes for a truncated protein of about 30 kDa which migrated close to the vector proteins. Plasmid pJHC-P320 produces only the proteins encoded by the vector pBR322.

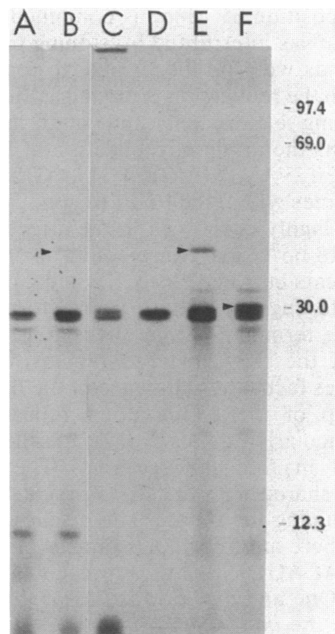

FIG. 3. Proteins encoded by REPI derivatives. Autoradiogram of a sodium dodecyl sulfate–12.5% polyacrylamide gel containing [^{35}S]methionine-labeled plasmid proteins from maxicells. Lanes: A, pJHC-P282; B, pJHC-P290; C, pBR322; D, pJHC-P320; E, pJHC-P335; F, pJHC-P357. The arrowheads indicate the RepI protein and its derivatives.

Since RepI is a basic protein (estimated pI, 10.69), its migration on sodium-dodecyl sulfate-polyacrylamide gel electrophoresis should be slower than that of a neutral protein of the same molecular weight. This could explain why RepI appeared as 39 kDa rather than 35 kDa in the maxicell analysis. Since the truncated protein from pJHC-P357 is also basic, we also expected anomalous migration of this protein so that the 22-kDa protein would appear as 30 kDa. We feel justified in predicting that the start of the RepI protein begins with the ATG start codon at nucleotide 457, since assigning the upstream, nonstandard GTG codon to accommodate the migration on sodium dodecyl sulfate-polyacrylamide gel electrophoresis may not be appropriate because RepI is a basic protein which would be expected to migrate anomalously and would appear larger than the size assessed on the basis of the sequence.

The results discussed in the previous paragraph allowed for prediction of the direction of transcription of the RepI protein gene, which coincided with the direction deduced from the sequencing data. pJHC-P290 and pJHC-P282 are identical clones except for the RepI-coding region contained in the 1,183-kb *Pst*I-*Pst*I fragment found in pJHC-P290. Clone pJHC-P282, which lacks the RepI-coding region, does not produce the 39-kDa protein, while pJHC-P290 does.

Polypeptides corresponding to each of the ORFs (including those initiated at GTG) can be observed in vitro by transcription-translation studies (data not shown), but only the 39-kDa protein corresponding to ORF1 was visualized in the in vivo maxicell system. As shown above, the start codon for ORF1 is preceded by a weak ribosome-binding site which might account, at least in part, for the low levels of in vivo expression observed for the RepI protein (Fig. 3, lane B, clone pJHC-P290). The RepI protein, like other replication proteins, might be autoregulated, which could also account for the low levels of expression in vivo. The fact that clone pJHC-P335 showed higher expression of RepI than did pJHC-P290 (Fig. 3) could be due to the strong promoters in the tetracycline resistance determinants in the vector which are just upstream of the RepI gene in pJHC-P335. These promoters could override the RepI promoter. The RepI gene cloned in pJHC-P290 is in the orientation opposite to that of pJHC-P335 and thus would not be driven by vector promoters. Alternatively, RepI promoter sequences might have been deleted in the cloning of pJHC-P290 but retained in pJHC-P335, resulting in the differences in expression.

Analysis of the DNA sequence indicated that the ORF for RepI ends 60 bp downstream from the *Hin*dIII site. Clone pJHC-P335 was derived by cloning REPI into the *Hin*dIII site of pBR322. Given the orientation of the insert in the

TABLE 1. Estimated pIs of computer-predicted CNBr cleavage fragments of the translated RepI protein gene

DNA fragment (nucleotide no.)	Polypeptide molecular mass (daltons)	Estimated pI
1–53	5,935	10.30
54–222	19,599	11.30
223–277	6,563	9.28
278–301	1,376	7.00

vector, it is likely that this procedure resulted in a fusion protein in which the terminal 22 amino acids of RepI were replaced with 18 amino acids encoded in the pBR322 sequences (43). Computer analysis showed that this RepI fusion protein has about the same molecular weight and estimated pI as the RepI protein. Maxicell analysis showed that the proteins have the same mobility (Fig. 3). Clone pJHC-P335, which codes for the fusion protein, was capable of replicating in a *polA* strain indicating that this protein was functional. A clone in which the insert is cloned in the reverse orientation so that the fusion would not occur produced a truncated protein lacking the terminal 22 amino acids. This clone was still capable of replicating in the *polA* strain (data not shown). Clone pJHC-P357, in which 318 bp of the 3′-OH portion of the RepI-coding region was deleted, resulting in a truncated protein (Fig. 3), is not capable of replication in a *polA* strain unless the intact RepI protein is provided in *trans* (32).

The predicted CNBr cleavage pattern of the translated ORF1 is shown in Table 1. The basic nature of the protein, particularly at the amino-terminal end, is a feature commonly found in DNA-binding proteins (30). Furthermore, by using computer analysis (8) and the Chou and Fasman predictive method (10), we identified a helix-turn-helix motif in the amino-terminal portion of the protein. The predicted amino acid sequence of a region of the RepI protein (amino acids 32 to 51) shares homology at key positions with Cro and cII proteins which have well-characterized DNA-binding capacities (Fig. 4a) (30). The conserved amino acids which are important in the structure of the helix-turn-helix are highlighted in Fig. 4a. This region in RepI is also highly conserved in the replication proteins encoded by Rts1 and P1 (Fig. 4b). The homology with the replication proteins encoded by Rts1 and P1 extends beyond this 20-amino-acid region (data not shown).

Analysis of DNA sequences spanning the *ori*, *incE*, and *incF* regions. The stretch of DNA from bases 1 to 233 possesses several DNA sequences repeated in inverse and direct orientations and specific sequences commonly found in replication regions. All direct or inverse repeats discussed in this paper share 70% or greater homology with their respective consensus sequences. Figure 5 is a schematic representation of the salient features of the sequence shown in Fig. 1.

Two copies of the DnaA-binding site were found. The one beginning at nucleotide 9 is 100% (9/9) homologous with the consensus 5′TTAT(A/C)CA(A/C)A3′ (13), and the other, beginning at nucleotide 54, is 77% (7/9) homologous (not shown in Fig. 5). In addition, the nucleotide sequence 5′GATC3′ was found in a high density between bases 1 and 70. The tetranucleotide sequence 5′GATC3′ is the DNA adenine methylation site (*dam*). The GATC sequences in the REPI replication region form part of a larger direct repeat, 5′GATCCAT(A/T)AA3′, found five times between bases 1 and 85 and once starting at nucleotide 124 (Fig. 1 and 5). The spacing of the repeats between bases 1 and 85 could allow for

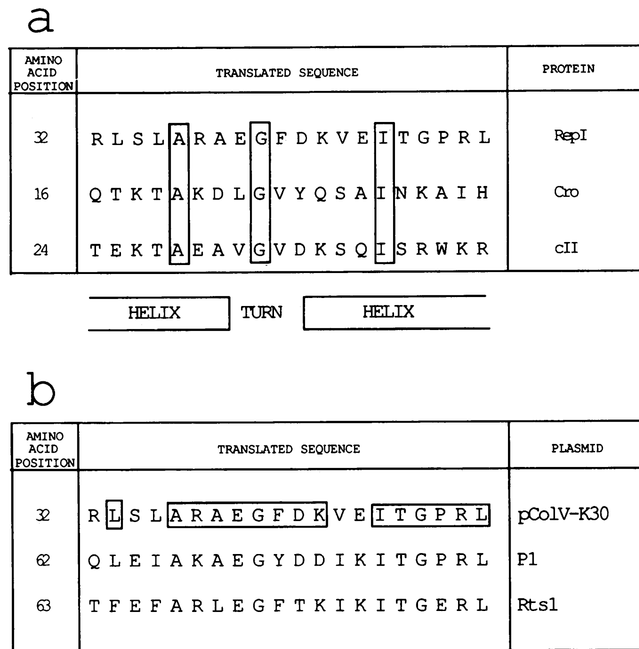

FIG. 4. (a) Amino acid sequence homology with the RepI protein and well-characterized DNA-binding proteins Cro and cII. The conserved amino acids which are important in the structure of the helix-turn-helix are boxed. (b) Homology among regions of the replication proteins from REPI, Rts1, and P1. Amino acid position indicates the first amino acid in the sequence shown. The amino acids in RepI which are conserved in one or both of the other replication proteins are boxed.

alignment of three to four of the repeats on one side of the DNA helix given the 10.4-bp-per-turn average. When the *Bam*HI site at position 65 which is contained within one of these sequences was interrupted by cloning procedures, the resultant plasmids were no longer able to replicate in *polA* cells, even when the replication protein was supplied in *trans* (32). The cloning experiments indicated that this DNA segment contains the origin of replication of REPI.

An 18-bp repeat, 5′**ATAA(A/G)CTGT(A/G)GTAAGTAA**3′, was found six times within the REPI region. The nucleotides in boldface are highly conserved in the direct repeats. Five copies of the 18-bp repeat were found as nearly tandem head-to-tail repeats between bases 129 and 233, which is the putative promoter region for RepI. Copy 6 formed part of the stem-loop at the termination site of the RepI protein gene. The spacing of the five clustered repeats was such that similar sequences faced the same face of the helix. Stretches of 10 to 13 bp of the 18-bp direct repeat were found elsewhere, particularly upstream of and in the RepI protein gene and also in the *incE* regions. The subsequence 5′ATAA3′ was shared with the 10-bp repeats found around the region containing the *Bam*HI site discussed above.

In addition, there are five copies of a 19-bp direct repeat, 5′TACTNACTACAGCTTATAT3′, present in the incompatibility regions. One and one-half copies of the 19-bp repeat were present in the *incE* region, and three were present in the *incF* fragment. The 19-bp direct repeat in the *incE* region formed part of the stem-loop structure at the end of the RepI protein gene by complementing regions in the 18-bp direct repeat described above. Subsequences sharing 10 to 18-bp homology with this consensus were found in the RepI-coding region, *incE* and *incF*.

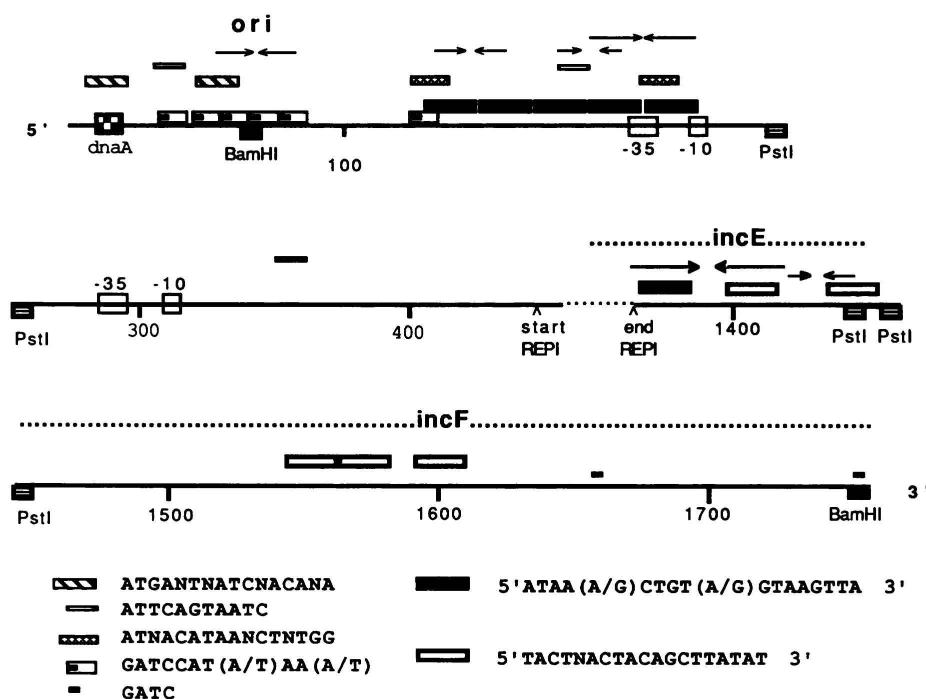

FIG. 5. Schematic representation of *ori*, *incE*, and *incF*, including the repeated sequences, the *dam* sites, and the perfect DnaA (dnaA) box. The second DnaA box is not depicted to minimize symbol crowding around the *Bam*HI site at the *ori* region. The restriction sites and relative positions corresponding to Fig. 1 are shown below the line. The nucleotides represented by the various symbols are shown at the bottom.

DISCUSSION

Our previous results indicated that the REPI region of pColV-K30 encodes a trans-acting factor capable of restoring the replication functions of REPI deletion derivatives (32). Maxicell experiments presented in this paper showed that the REPI region codes for a polypeptide of 39 kDa. Failure to produce this RepI protein correlated with inability to replicate in *polA* strains. We therefore conclude that the 39-kDa protein produced by replication-proficient REPI clones is the *trans*-acting factor essential for replication of REPI-containing plasmids.

The direction of transcription of the RepI protein gene could be determined by polypeptide analysis of deletion derivatives. Clone pJHC-P335 codes for a RepI fusion protein in which 22 amino acids of the carboxy-terminal portion were replaced with 18 amino acids encoded by the vector pBR322. This RepI fusion protein has about the same estimated molecular weight and pI as the RepI protein and is capable of supporting replication in a *polA* strain. However, further deletion of 106 amino acids from the carboxy-terminal end resulted in a loss of replication function, as seen in clone pJHC-P357 (32). This suggests that the 22 amino acids in the carboxy-terminal portion of the RepI protein are not necessary for replicative functions. Further truncation of the RepI protein perhaps results in structural changes or deletion of important functional domains which render the polypeptide inactive.

The translated ORF which we identified as corresponding to the RepI protein has a basic estimated pI which is characteristic of DNA-binding proteins (30). Furthermore, the predicted amino acid sequence and the estimated pI of a specific region of the RepI protein share homology with proteins Cro and cII, which have well-characterized DNA-binding capacities (30). Interestingly, this region, as well as other regions, is conserved in replication proteins such as those encoded by Rts1 (19) and P1 (2).

Cloning experiments localized the origin of replication between nucleotides 1 and 1043 (32). The precise location of *ori* within this stretch is not known, but the region between nucleotides 1 and 233 contains motifs and specific sequences commonly found in origins of replication. For example, two DnaA-binding sites were found starting at nucleotides 9 and 54. DnaA binding to plasmid replicons has been demonstrated for pSC101 (48), F (28), P1 (2), and Rts1 (19). In some cases, DnaA is essential for plasmid replication (16, 17). Furthermore, the tetranucleotide sequence 5'GATC3', which is the DNA adenine methylation site (*dam*), was found in a high density between bases 1 and 85. This sequence occurs with high frequency within the origins of replication of the *E. coli* and *Salmonella typhimurium* chromosomes (27, 52) and within the origins of replication of many plasmids (2, 12, 18). DNA hemimethylated at Dam methylase recognition sites has been shown to play important roles in segregation of the *E. coli* chromosome (29), in mismatch repair (26), and in controlling replication (36). In P1, DNA adenine methylation is required for plasmid replication (1).

The GATC sequences in the REPI replication region form part of a larger direct repeat, 5'GATCCA(A/T)AA3', found five times between bases 1 and 85 and once starting at nucleotide 124. The clustered sequences are present at intervals which would align most of them on the same face of the helix. When the *Bam*HI site at position 65 which is contained within one of these sequences was interrupted by cloning procedures, the resultant plasmids were no longer able to replicate in *polA* cells, even when the replication protein was supplied in *trans* (32). The subsequence 5'ATAA3' of this repeat shares homology with the series of the 18-bp repeat found just downstream in the putative

promoter region for the RepI protein gene. The region beginning at nucleotide position 124 shows how the two direct-repeat sequences actually overlap (Fig. 1 and 5).

This 18-bp directly repeated sequence is found five times in the region between nucleotides 129 and 233. This region shares characteristics with other replicons. Replication regions often contain direct repeats which could align on the same face of the 10.4-bp-per-turn DNA helix (2, 11, 19, 42). In some cases, a plasmid-encoded replication protein binds specifically to these repeated sequences (3, 9, 15, 19).

Because the RepI protein shares characteristics with DNA-binding proteins and with replication proteins, it is tempting to speculate that the mechanism of action of the RepI polypeptide involves an interaction with the *ori* repeated sequences found around the *Bam*HI site (nucleotide 65) and that the 18-bp repeated sequences titrate the RepI protein, possibly by the shared 5′ATAA3′ sequence. The titration of the protein could serve to control the copy number and autoregulate expression of the RepI protein gene by binding to putative RepI promoters present in this repeat region or in areas just downstream containing partial repeats. This proposed mechanism has been found in other replication systems (20, 49). Of course, alternative models exist; for example, the RepI protein might have multiple binding domains, one for the replication function and another for incompatibility and autoregulatory repeats.

Five copies of a 19-bp direct repeat occur in the incompatibility regions; 1.5 copies are in the *incE* segment, and three are in the *incF* fragment. The 19-bp repeat is nearly the reverse complement of the 18-bp repeat found around the *ori* region. The complete copy found in *incE* forms part of the large stem-loop structure at the end of the RepI protein gene by complementing a copy of the 18-bp repeat. The role of this stem-loop is not clear. However, this structure must not be essential for replication, since clone pJHC-P335, which lacks this stem-loop, is still capable of replicating in *polA* strains (32).

Direct repeats are responsible for the incompatibility phenotype of the main replicon of F (28, 41, 45, 46), P1 (2, 9), R6K (3, 15, 21, 39), pSC101 (11), Rts1 (44), and R1162 (24). Incompatibility on these plasmids is expressed by specific binding of the replication proteins to repeated sequences. Results presented in a previous paper indicate that the REPI *incF* region confers incompatibility to REPI, P307, and f7 derivatives (32). The possibility exists that the replication proteins of these plasmids are titrated by the 19-bp direct-repeat sequences present in the *incF* region and thus prevent initiation of replication. Whether the reverse orientation of these direct repeats plays a role in the *incF*-determined phenotype has yet to be determined. Furthermore, the existence of other factors encoded in this region which may play a role in this novel incompatibility phenotype cannot be disregarded.

The mechanism of action of *incE* remains unclear. Plasmids containing *incE* were incompatible only with f7 derivatives (32); this phenomenon is difficult to justify by the presence of 1.5 copies of the 19-bp repeated sequence. One hypothesis is that changes of the nucleotide sequence which generated the *Hin*dIII site on REPI and P307 (32) created a region to which the RepI protein cannot bind efficiently so that the *incE* region of REPI does not express incompatibility toward REPI or P307 derivatives. Experiments are under way to address these questions as well as to characterize the precise role of the RepI protein in the replication and incompatibility properties of IncFI plasmids.

The organization of REPI is similar to that of many plasmid replication systems, such as P1 and Rts1, for example, with respect to a DnaA-binding site(s), a high frequency of DNA adenine methylation sites, five or more tandem direct repeats of 18 to 21 nucleotides just upstream of the gene for a protein essential for replication, and several more repeats in direct or reverse orientation. The similarities allow for the following conjectures about the functions of these similar structures. REPI might have a DnaA and/or *dam* requirement for replication. The removal of some of the direct repeats flanking the RepI protein gene should alter the copy number phenotype. Finally, the five direct repeats upstream of the RepI protein gene should also express incompatibility and might play a role in autoregulation.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation grant DMB-8807735 to J.H.C., J.F.P.-C. and A.E.G. were recipients of Tartar Trust Fund awards from the Medical Research Foundation of Oregon.

We thank R. G. Brennan for homology search computer analysis of the REPI sequence and David H. Farrell for help with sequencing and nucleotide analysis.

LITERATURE CITED

1. **Abeles, A. L., and S. J. Austin.** 1987. P1 plasmid replication requires methylated DNA. EMBO J. **6**:3185–3189.
2. **Abeles, A. L., K. M. Snyder, and D. K. Chattoraj.** 1984. P1 plasmid replication: replicon structure. J. Mol. Biol. **173**:307–324.
3. **Bastia, D., C. Vocke, J. Germino, and J. Gray.** 1985. DNA-protein interaction at the replication origins of plasmid chromosomes, p. 397–414. *In* D. R. Helinski, S. N. Cohen, D. B. Clewell, D. A. Jackson, and A. Hollaender. (ed.), Plasmids in bacteria. Plenum Publishing Corp., New York.
4. **Bergquist, P. L., H. E. D. Lane, L. Malcolm, and R. A. Downward.** 1982. Molecular homology and incompatibility in the IncFI plasmid group. J. Gen. Microbiol. **128**:223–238.
5. **Bergquist, P. L., S. Saadi, and W. K. Maas.** 1986. Distribution of basic replicons having homology with RepFIA, RepFIB, and RepFIC among IncFI group plasmids. Plasmid **15**:19–34.
6. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow.** 1977. Construction and characterization of new cloning vehicles. II. A multiple purpose cloning system. Gene **2**:95–113.
7. **Boyer, H. W., and D. Roulland-Doussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. **41**:459–472.
8. **Brennan, R. G., L. H. Weaver, and B. W. Mathews.** 1986. Use of protein sequence and structure to infer distant evolutionary relationships. Chem. Scr. **26**:251–255.
9. **Chattoraj, D., K. Cordes, and A. Abeles.** 1984. Plasmid P1 replication: negative control by repeated DNA sequences. Proc. Natl. Acad. Sci. USA **81**:6456–6460.
10. **Chou, P. Y., and G. D. Fasman.** 1978. Empirical predictions of protein conformation. Annu. Rev. Biochem. **47**:251–276.
11. **Churchward, G., P. Linder, and L. Caro.** 1983. The nucleotide sequence of replication and maintenance functions encoded by plasmid pSC101. Nucleic Acids Res. **11**:5645–5659.
12. **Froehlich, B. J., and J. R. Scott.** 1988. A single amino acid difference between Rep proteins of P1 and P7 affects plasmid copy number. Plasmid **19**:121–133.
13. **Fuller, R. S., and A. Kornberg.** 1983. Purified *dnaA* protein in initiation of replication at the *Escherichia coli* chromosomal origin of replication. Proc. Natl. Acad. Sci. USA **80**:5817–5821.
14. **Gardner, R., J. McAnulty, E. Feher, and D. Lane.** 1985. Location of *rep* and *inc* sequences in the F-secondary replicon. Plasmid **13**:145–148.
15. **Germino, J., and D. Bastia.** 1983. Interaction of plasmid R6K-encoded replication initiator protein with its binding sites on DNA. Cell **34**:125–134.
16. **Hansen, E. B., and M. B. Yarmolinsky.** 1986. Host participation

in plasmid maintenance: dependence upon *dnaA* of replicons derived from P1 and F. Proc. Natl. Acad. Sci. USA **83**: 4423–4427.
17. **Hasunuma, D., and M. Sekiguchi.** 1977. Replication of plasmid pSC101 in *Escherichia coli* K-12: requirement for *dnaA* function. Mol. Gen. Genet. **154**:225–230.
18. **Itoh, Y., Y. Kamio, and Y. Terawaki.** 1987. Essential DNA sequences for the replication of Rts1. J. Bacteriol. **169**:1153–1160.
19. **Kamio, Y., Y. Itoh, and Y. Terawaki.** 1988. Purification of Rts1 RepA protein and binding of the protein to mini-Rts1 DNA. J. Bacteriol. **170**:4411–4414.
20. **Kelley, W., and D. Bastia.** 1985. Replication initiator protein of plasmid R6K autoregulates its own synthesis at the transcriptional step. Proc. Natl. Acad. Sci. USA **82**:2574–2578.
21. **Kolter, R., and D. R. Helinski.** 1982. Plasmid R6K DNA replication. II. Direct nucleotide sequence repeats are required for an active origin. J. Mol. Biol. **161**:45–56.
22. **Lane, D., and R. C. Gardner.** 1979. Second *Eco*RI fragment of F capable of self-replication. J. Bacteriol. **139**:141–151.
23. **Leong, S., G. S. Ditta, and D. R. Helinski.** 1982. Heme biosynthesis in *Rhizobium*. Identification of a cloned gene coding for α-aminolevulinic acid synthetase from *Rhizobium meliloti*. J. Biol. Chem. **257**:8724–8730.
24. **Lin, L.-S., and R. J. Meyer.** 1986. Directly repeated, 20-bp sequence of plasmid R1162 DNA is required for replication, expression of incompatibility, and copy-number control. Plasmid **15**:35–47.
25. **Lovett, M. A., and D. R. Helinski.** 1976. Method for the isolation of the replication region of a bacterical replicon: construction of a mini F'-*km* plasmid. J. Bacteriol. **127**:982–987.
26. **Lu, A.-L.** 1987. Influence of GATC sequences on *Escherichia coli*. DNA mismatch repair in vitro. J. Bacteriol. **169**:1254–1259.
27. **Meijer, M., E. Beck, F. G. Hansen, H. E. N. Bergmans, W. Messer, K. von Mayenburg, and H. Schaller.** 1979. Nucleotide sequence of the origin of replication of the *Escherichia coli* K12 chromosome. Proc. Natl. Acad. Sci. USA **76**:580–584.
28. **Murotsu, T., K. Matsubara, H. Sugisaki, and M. Takanami.** 1981. Nine unique repeating sequences in a region essential for replication and incompatibility of the mini-F plasmid. Gene **15**:257–271.
29. **Ogden, G. B., M. J. Pratt, and M. Schaechter.** 1988. The replicative origin of the *Escherichia coli* chromosome binds to membrane only when hemimethylated. Cell **54**:127–135.
30. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53**:293–321.
31. **Perez-Casal, J. F., and J. H. Crosa.** 1984. Aerobactin iron uptake sequences in plasmid ColV-K30 are flanked by inverted IS1-like elements and replication regions. J. Bacteriol. **160**: 256–265.
32. **Perez-Casal, J. F., and J. H. Crosa.** 1987. Novel incompatibility and partition loci for the REPI replication region of plasmid ColV-K30. J. Bacteriol. **169**:5078–5086.
33. **Picken, R. N., A. J. Mazaitis, S. Saadi, and W. K. Maas.** 1984. Characterization of the basic replicons of the chimeric R/Ent plasmid pCG86 and the related Ent plasmid P307. Plasmid **12**:10–18.
34. **Prody, C. A., and J. B. Neilands.** 1984. Genetic and biochemical characterization of the *Escherichia coli* K12 *fhuB* mutation. J. Bacteriol. **157**:874–880.
35. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in promotion and termination of DNA transcription. Annu. Rev. Genet. **13**:319–353.
36. **Russell, D. W., and N. D. Zinder.** 1987. Hemimethylation prevents DNA replication in *E. coli*. Cell **50**:1071–1079.
37. **Sancar, A., A. M. Hack, and W. D. Rupp.** 1979. Simple method for identification of plasmid-coded proteins. J. Bacteriol. **137**: 692–693.
38. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
39. **Shafferman, A., R. Kolter, D. Stalker, and D. R. Helinski.** 1982. Plasmid R6K DNA replication. III. Regulatory properties of the initiation protein. J. Mol. Biol. **161**:57–76.
40. **Shine, J., and L. Dalgarno.** 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
41. **Sogaard-Anderson, L., L. A. Rokeach, and S. Molin.** 1984. Regulated expression of a gene important for replication of plasmid F in *Escherichia coli*. EMBO J. **3**:257–262.
42. **Stalker, D. M., R. Kolter, and D. R. Helinski.** 1979. Nucleotide sequence of the region of an origin of replication of the antibiotic resistance plasmid R6K. Proc. Natl. Acad. Sci. USA **76**:1150–1154.
43. **Sutcliffe, J. G.** 1979. Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. **XLIII**:77–90.
44. **Terawaki, Y., and Y. Itoh.** 1985. Copy mutant of mini-Rts1: lowered binding affinity of mutated RepA protein to direct repeats. J. Bacteriol. **162**:72–77.
45. **Tolun, A., and D. R. Helinski.** 1981. Direct repeats of the F plasmid *incC* region express incompatibility. Cell **24**:687–694.
46. **Tsutsui, H., A. Fujiyama, T. Murotsu, and K. Matsubara.** 1983. Role of nine repeating sequences of the mini-F genome for expression of F-specific incompatibility phenotype and copy number control. J. Bacteriol. **155**:337–344.
47. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
48. **Vocke, C., and D. Bastia.** 1984. Primary structure of the essential replicon of the plasmid pSC101. Proc. Natl. Acad. Sci. USA **80**:6557–6561.
49. **Vocke, C., and D. Bastia.** 1985. The replication initiator protein of plasmid pSC101 is a transcriptional repressor of its own cistron. Proc. Natl. Acad. Sci. USA **82**:2252–2256.
50. **Waters, V. L., and J. H. Crosa.** 1986. DNA environment of the aerobactin iron uptake system genes in prototypic ColV plasmids. J. Bacteriol. **167**:647–654.
51. **Williams, P. H.** 1979. Novel iron uptake system specified by ColV plasmids: an important component in the virulence of invasive strains of *Escherichia coli*. Infect. Immun. **26**:925–932.
52. **Zyskind, J. W., and D. W. Smith.** 1980. Nucleotide sequence of the *Salmonella typhimurium* origin of DNA replication. Proc. Natl. Acad. Sci. USA **77**:2460–2464.