JOURNAL OF BACTERIOLOGY, Aug. 1987, p. 3829–3832
0021-9193/87/083829-04$02.00/0

Vol. 169, No. 8

oriT Sequence of the Antibiotic Resistance Plasmid R100

SARAH A. McINTIRE[1]* AND WALTER B. DEMPSEY[1,2]

Veterans Administration Medical Center, Dallas, Texas 75216,[1] and Department of Biochemistry, University of Texas Health Science Center, Dallas, Texas 75235[2]

Received 15 December 1986/Accepted 4 May 1987

We present the nucleotide sequence of the *oriT* region from plasmid R100. Comparison to other IncF plasmids revealed homology around the proposed nick sites as well as conservation of inverted repeated sequences in the nonhomologous region. Three areas showed strong homology (eight of nine nucleotides) to the consensus sequence for binding of integration host factor, suggesting a role for this DNA-binding protein in nicking at *oriT*.

Our understanding of bacterial conjugation depends mostly on analyses of the fertility factor F. The genetic information necessary for conjugation of F and other closely related IncF plasmids is encoded in approximately 34 kilobases of plasmid DNA and in several chromosomal genes. The contiguous plasmid genes include the following: (i) an origin of transfer (*oriT*), (ii) more than 20 structural genes for assembly of the pilus, (iii) genes involved in the metabolism of DNA during transfer, (iv) genes necessary for surface exclusion, and (v) genes whose products control expression of the genes mentioned above. Reviews of plasmid genes have been published (20, 42, 43). The chromosomal genes include *dye* (*sfrA*, *fex*), *cpxA*, *cpxB*, *sfrB*, *himA*, and *hip*. The first three chromosomal genes (2, 4, 23, 27) encode proteins which are believed to link expression of transfer genes to the cell membrane (28), and the *sfrB* product is required to prevent premature transcriptional termination of the transfer operon (3, 15). A review of these chromosomal loci has been published recently (39). The chromosomal genes *himA* and *hip* encode integration host factor (IHF) (30–32), recently shown to be required for transfer of IncF plasmid R100 (W. B. Dempsey, submitted for publication).

The transfer (*tra*) genes of IncF plasmids are homologous, both physically (24, 38) and genetically (10, 11, 34), but some of the genes are not interchangeable and are termed plasmid specific. One plasmid-specific function is the processing of plasmid DNA during transfer, which involves both *oriT* and the genes whose products are thought to interact at *oriT* .

Conjugation involves the transfer of a single strand of DNA into a recipient cell; *oriT* is the site on that strand where nicking occurs (7, 40) and where unwinding and consequent transfer begin, with the 5′ end of the nicked DNA leading into the recipient (33, 37) and the transfer genes entering last. Everett and Willetts (6) showed that *traY* and *traZ* are required for in vivo nicking at *oriT* and proposed that a *traYZ*-encoded endonuclease catalyzes the interconversion of covalently closed and open circular plasmid DNA by nicking and ligating at *oriT*. In response to some signal after mating pairs are formed, the nicking-ligating equilibrium is shifted, and nicked, single-stranded DNA is available for transfer. The product of *traM* is proposed as the signal that triggers transfer (42). Both *traM* and *traY* are plasmid specific, and it is proposed (41) that both products bind in the *oriT* region.

* Corresponding author.

Willetts and Maule (41) recently surveyed 12 IncF plasmids in an attempt to correlate various plasmid-specific properties. These plasmids fall into four groups on the basis of their *oriT-traM-traY* specificities: type I includes F, ColV2, R386, ColB2, R124, and ColVB*trp*; type II includes R538-1 and probably ColB4-K98; type III contains R1; type IV includes R100, R6-5, and R136. The *oriT* DNA sequences of F (40) and R1 (35) have been established. A recent paper (9) presented the ColB4 *oriT* sequence and included a prepublication copy of our R100 *oriT* sequence, mistakenly thought to have been included by Fee and Dempsey (8). We present our R100 *oriT* sequence and include some additional observations on features of this region in all four groups of IncF plasmids.

Subcloning of R100 *oriT* after *Hae*III digestion of pBF1 DNA produced pWD36, which was *oriT*$^+$ (8) and was the source of DNA for sequencing *oriT*. Our end-labeling procedures have been described elsewhere (29). The *oriT* sequence was determined by using a modification (1) of the chemical degradation method of Maxam and Gilbert (25, 26), and the sequence data were analyzed by using the Beckman MicroGenie program developed by Queen and Korn (36). Our sequence includes 73 more bases to the left (5′ end) than were presented previously (Fig. 1) (9). At the 3′ end, the R100 DNA in pWD36 ended at the *Hae*III site at base pair 541, and the ATG codon at base pairs 538 to 540 is the beginning of the *traM* sequence, as shown by Fee and Dempsey (8). Finlay et al. (9) compared four IncF types of *oriT* sequences and tabulated the various attributes of this R100 region. These aspects will not be detailed here, but briefly included the following: (i) regions of identity with a type I plasmid F, from base 74 (first F base available [40]) to base 246, with a type II plasmid ColB4, from base 74 (first ColB4 base available [9]) to base 246, and with a type III plasmid R1, from base 98 (first R1 base available [35]) to base 246 (this region contains the nick sites proposed for F [18, 40]); (ii) an A-T-rich region beginning in the homologous area near base 225 and extending into the nonhomologous region to about base 310; and (iii) several inverted repeated sequences (IRs) in similar regions in all four plasmids. Additional IRs in R100 (Fig. 1) occurred at bases 56 to 65 and 96 to 105 ($\Delta G = -11.6$ kcal [1 kcal = 4,184 J]) and at bases 401 to 415 and 447 to 464 ($\Delta G = 21.8$ kcal). An IR in the F sequence homologous to the second IR was overlooked previously (9) at F bases 334 to 344 and 375 to 385 ($\Delta G = 23.4$ kcal).

```
                                                                    70
GATATCATTC TTAGTAAATC CGGGTCTATT TTGTAATCCC GACCTGCAAG ATCAAAGCAA TCGGCGGCTT
CTATAGTAAG AATCATTTAG GCCCAGATAA AACATTAGGG CTGGACGTTC TAGTTTCGTT AGCCGCCGAA

                                                                   140
CGCAGATCCC ATTTATAAAC ATCAGGCAGA TTGCTAACAT CCATTTTTTC ATTTTTCCAC CTCTGGTGAC
GCGTCTAGGG TAAATATTTG TAGTCCGTCT AACGATTGTA GGTAAAAAAG TAAAAAGGTG GAGACCACTG

                                                                   210
TTTATCCGTA AATAATTTAA CCCACTCCAC AAAAAGGCTC AACAGGTTGG TGGTTCTCAC CACCAAAAGC
AAATAGGCAT TTATTAAATT GGGTGAGGTG TTTTTCCGAG TTGTCCAACC ACCAAGAGTG GTGGTTTTCG

                                                                   280
ACCACACACT ACGCAAAAAC AAGTTTTTGC TGATTTACGT TTTAAATCAT GTAATTAACA AACGCTATTT
TGGTGTGTGA TGCGTTTTTG TTCAAAAACG ACTAAATGCA AAATTTAGTA CATTAATTGT TTGCGATAAA

                                                                   350
ATATTTAATA ATTCTGAATT ATTAAATAGA GAGTCGTTGG CGATCCTGTT ACGACCCTAT AGCGACTCTT
TATAAATTAT TAAGACTTAA TAATTTATCT CTCAGCAACC GCTAGGACAA TGCTGGGATA TCGCTGAGAA

                                                                   420
ATTAGGATCA GTACTGACAA TTTTCACGAA AAACATTGAT ATATATTAGG GACGCCCCTA GCGCCGGATC
TAATCCTAGT CATGACTGTT AAAAGTGCTT TTTGTAACTA TATATAATCC CTGCGGGGAT CGCGGCCTAG

                                                                   490
GCATTTTTTT ATAGGATGAT CCTAGAGGCG CTGCTAGGAG TGTCTTTCTT ACTTGTAAAC ACTATATGCA
CGTAAAAAAA TATCCTACTA GGATCTCCGC GACGATCCTC ACAGAAAGAA TGAACATTTG TGATATACGT

TATTTATAGT GAAAAGTTCG TTGCTTAATA ATGAAAGGTT TTATCTTATG GCCAGAGTAA
ATAAATATCA CTTTTCAAGC AACGAATTAT TACTTTCCAA AATAGAATAC CGGTCTCATT
```

FIG. 1. Nucleotide sequence of R100 *oriT* region. Base 74 corresponds to base 1 in Finlay et al. (9), and base 518 corresponds to base 1 in Fee and Dempsey (8). Sequence to the left of the vertical line between bases 246 and 247 is homologous to other IncF plasmids. ↑, Proposed nick sites (18, 40) in plasmid F *oriT*; ×, two known *oriT* mutations (40). IRs not previously identified (9) are shown by arrows above the sequence. Putative IHF-binding sites are underlined (bottom strand) or overlined (top strand).

On the basis of these sequences, the groupings of Willetts and Maule (41), and the available information on *tra* gene products, Finlay et al. (9) proposed that each plasmid-specific *traY* product binds to its respective A-T-rich region and (presumably) orients the nonspecific *traZ* endonuclease for nicking (Fig. 1, bottom strand). They further proposed that each plasmid-specific *traM* product recognizes its own IR located just outside the A-T-rich region. In R100, the IR is at bases 305 to 316 and bases 343 to 353 and has a predicted free energy of −11.2 kcal. While not homologous, this region in all four types contains IRs in similar locations. The reader is referred to Finlay et al. (9) for a detailed discussion of this comparison.

The recent discovery that IHF is required for expression of R100 transfer (Dempsey, submitted) and also that it binds to an A-T-rich sequence in the pSC101 replication region (16) prompted us to examine *oriT* for homology to the IHF-binding-site consensus sequence, C/TAANNNNTTGATA/T (22). IHF is a heterodimeric DNA-binding protein that was first detected as a host factor needed for integrative recombination in bacteriophage λ (13, 21, 32). The protein protects *att* DNA in λ, P22, and φ80 (5, 22) and has been shown to be involved in expression and regulation of several systems (12, 14, 17, 19). In P22 *att*, one site with two deviations binds IHF less well, but does still bind (22). Nevertheless, we initially screened for the consensus sequence and for sites with one deviation from consensus. The three IHF-binding sites in R100 *oriT* region are highlighted in Fig. 1 and shown diagrammatically in Fig. 2. The first site (bases 227 to 239) was on the bottom strand in the region which was homologous in all four types of IncF plasmids; the second site (bases 255 to 267) was also on the bottom strand; the third site (bases 379 to 391) was on the top strand in such a location that it could be involved in *traM* expression. All three of these sites in R100 deviated from consensus by only one base. Examination of *oriT* sequences of the other IncF plasmids revealed that both ColB4 and R1 contained the second IHF site, each with one deviation from consensus.

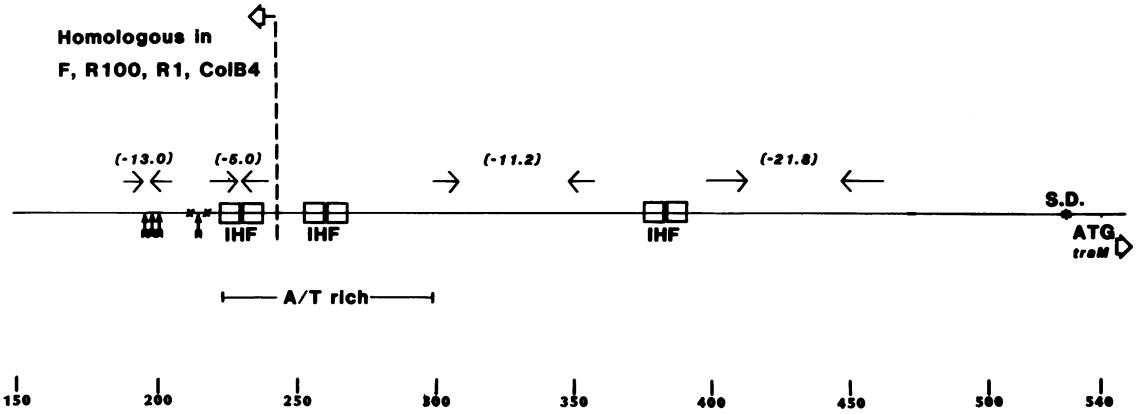

FIG. 2. Features of the *oriT* region of plasmid R100, beginning at base 150 in the region of IncF homology. Some IRs are shown (arrows above the line) with their ΔG values in kcal. The double boxes represent homology with the IHF-binding-site consensus sequence. ×, *oriT* mutation in F; ↑, nick sites in F *oriT*; S.D., homology with Shine-Dalgarno ribosome-binding-site consensus sequence.

This region of F DNA contained three deviations, although a nearer consensus (two deviations) occurred 12 bases downstream. A similar situation occurred at the third site near *traM*; both ColB4 and R1 have the site with two deviations, and F has the site with three deviations.

A cursory search of GenBank by using Beckman MicroGenie revealed that several vegetative origin regions contain one or more sites which are homologous to the IHF consensus. These include the chromosomal origins of *Escherichia coli*, *Enterobacter aerogenes*, *Klebsiella* sp., and *Salmonella typhimurium* and the *rep* regions of plasmids F, R100, R6-5, R1, R6K, RK2, RSF1030, Rts1, pSC101, and pNTP1. There are no data showing either a direct or an indirect role for IHF in either these regions or the *oriT* described here. In λ, however, IHF has been shown to enhance the affinity of Int protein binding to the λ*att* core (5), and it may play a similar role in the *oriT* region of these IncF plasmids, i.e., by increasing the binding of one or more *tra* proteins involved in nicking at *oriT*. In that regard, it may be significant that binding at the first IHF site in *oriT* probably would cover bases 215 to 244 (5, 22), which would include one of the proposed nick sites (18). It has been suggested that IHF may be involved in partial unwinding of the DNA helix (22); this also could be its role at *oriT*. In addition, all of the IncF *oriT* sites lie in the plasmid DNA that would enter the recipient last, so it is also possible that IHF plays a role in the recircularization of a unit length of plasmid DNA.

We thank Brian E. Fee, Julie Kalantar, and Kathryn Sims for technical assistance.

This research was supported by the Research Service of the Veterans Administration and by Public Health Service grant AI-17402 from the National Institutes of Health.

LITERATURE CITED

1. **Bencini, D. A., G. A. O'Donovan, and J. R. Wild.** 1984. Rapid chemical degradation sequencing. Biotechniques **2**:4–5.
2. **Beutin, L., and M. Achtman.** 1979. Two *Escherichia coli* chromosomal cistrons, *sfrA* and *sfrB*, which are needed for expression of F factor *tra* functions. J. Bacteriol. **139**:730–737.
3. **Beutin, L., P. A. Manning, M. Achtman, and N. Willetts.** 1981. *sfrA* and *sfrB* products of *Escherichia coli* K-12 are transcriptional control factors. J. Bacteriol. **145**:840–844.
4. **Buxton, R. S., K. Hammer-Jespersen, and T. D. Hansen.** 1978. Insertion of bacteriophage lambda into the *deo* operon of *Escherichia coil* K-12 and isolation of plaque-forming λ*deo*$^+$ transducing bacteriophages. J. Bacteriol. **136**:668–681.
5. **Craig, N., and H. A. Nash.** 1984. *E. coli* integration host factor binds to specific sites in DNA. Cell **39**:707–716.
6. **Everett, R., and N. Willetts.** 1980. Characterization of an *in vivo* system for nicking at the origin of conjugal DNA transfer of the sex factor F. J. Mol. Biol. **136**:129–150.
7. **Everett, R., and N. S. Willetts.** 1982. Cloning, mutation, and location of the F origin of conjugal transfer. EMBO J. **1**:747–753.
8. **Fee, B. E., and W. B. Dempsey.** 1986. Cloning, mapping, and sequencing of plasmid R100 *traM* and *finP* genes. J. Bacteriol. **167**:336–345.
9. **Finlay, B. B., L. S. Frost, and W. Paranchych.** 1986. Origin of transfer of IncF plasmids and nucleotide sequences of the type II *oriT*, *traM*, and *traY* alleles from ColB4-K98 and the type IV *traY* allele from R100-1. J. Bacteriol. **168**:132–139.
10. **Foster, T. J., and N. S. Willetts.** 1976. Genetic analysis of deletions of R100-1 that are both transfer-deficient and tetracycline-sensitive. J. Gen. Microbiol. **93**:133–140.
11. **Foster, T. J., and N. S. Willetts.** 1977. Characterization of transfer-deficient mutants of the R100-1 Tc^s plasmid pDU202, caused by insertion of Tn10. Mol. Gen. Genet. **156**:107–114.
12. **Friden, P., K. Voelkel, R. Sternglanz, and M. Freundlich.** 1984. Reduced expression of the isoleucine and valine enzymes in integration host factor mutants of *E. coli*. J. Mol. Biol. **172**:573–579.
13. **Friedman, D. I., E. J. Olson, D. Carver, and M. Gellert.** 1984. Synergistic effect of *himA* and *gyrB* mutations: evidence that Him functions control expression of *ilv* and *xyl* genes. J. Bacteriol. **157**:484–489.
14. **Friedman, D. I., E. R. Olson, C. Georgopoulos, K. Tilly, I. Herskowitz, and F. Banuett.** 1984. Interactions of bacteriophage and host macromolecules in the growth of bacteriophage λ. Microbiol. Rev. **48**:299–325.
15. **Gaffney, D., R. Skurray, and N. Willetts.** 1983. Regulation of the F conjugation genes studied by hybridization and *tra-lacZ* fusion. J. Mol. Biol. **168**:103–122.
16. **Gamas, P., A. C. Burger, G. Churchword, L. Caro, D. Galas, and M. Chandler.** 1986. Replication of pSC101: effects of mutations in the *E. coli* DNA binding protein IHF. Mol. Gen. Genet. **204**:85–89.
17. **Goosen, N., and P. van de Putte.** 1984. Regulation of Mu transposition. I. Localization of the presumed recognition sites for HimD and *ner* functions controlling bacteriophage Mu transcription. Gene **30**:41–46.
18. **Horowitz, B., and R. C. Deonier.** 1985. Formation of Δ*F traF'* plasmids: specific recombination at *oriT*. J. Mol. Biol. **186**: 267–274.
19. **Hoyt, M. A., D. M. Knight, A. Das, H. I. Miller, and H. Echols.** 1982. Control of phage λ development by stability and synthesis of cII protein: role of the viral cIII and host *hfl*, *himA*, and *himD* genes. Cell **31**:565–573.
20. **Ippen-Ihler, K. A., and E. G. Minkley, Jr.** 1986. The conjugation system of F, the fertility factor of *E. coli*. Annu. Rev. Genet. **20**:593–624.
21. **Kikuchi, Y., and H. A. Nash.** 1978. Integrative recombination of bacteriophage λ: requirement for supertwisted DNA *in vivo* and characterization of *int*. Cold Spring Harbor Symp. Quant. Biol. **43**:1099–1109.
22. **Leong, J. M., S. Nunes-Duby, C. F. Lesser, P. Youderian, M. M. Susskind, and A. Landy.** 1985. The φ80 and P22 attachment sites. J. Biol. Chem. **260**:4468–4477.
23. **Lerner, T. J., and N. D. Zinder.** 1979. Chromosomal regulation of sexual expression in *Escherichia coli*. J. Bacteriol. **137**:1063–1065.
24. **Manning, P. A., and G. Morelli.** 1982. DNA homology of the promoter-distal regions of the *tra* operons of sex factors F and R100 in *Escherichia coli* K-12. J. Bacteriol. **150**:389–394.
25. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.
26. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
27. **McEwen, J., and P. Silverman.** 1980. Chromosomal mutations of *Escherichia coli* that alter expression of conjugative plasmid functions. Proc. Natl. Acad. Sci. USA **77**:513–517.
28. **McEwen, J., and P. M. Silverman.** 1982. Mutations in genes *cpxA* and *cpxB* alter the protein composition of *Escherichia coli* inner and outer membranes. J. Bacteriol. **151**:1553–1559.
29. **McIntire, S. A., and W. B. Dempsey.** 1987. Fertility inhibition gene of plasmid R100. Nucleic Acids Res. **15**:2029–2042.
30. **Miller, H. I., and D. I. Friedman.** 1980. An *Escherichia coli* gene product required for lambda site specific recombination. Cell **20**:711–719.
31. **Miller, H. I., and H. A. Nash.** 1981. Direct role of the *himA* gene product in phage λ integration. Nature (London) **290**:523–526.
32. **Nash, H. A., and C. A. Robertson.** 1981. Purification and properties of the *E. coli* protein factor required for λ integrative recombination. J. Biol. Chem. **256**:9246–9253.
33. **Ohki, M., and J. Tomizawa.** 1986. Asymmetric transfer of DNA strands in bacterial conjugation. Cold Spring Harbor Symp. Quant. Biol. **33**:651–657.
34. **Ohtsubo, E., Y. Nishimura, and Y. Hirota.** 1970. Transfer-defective mutants of sex factor in *E. coli*. I. Defective mutants and complementation analysis. Genetics **64**:173–188.

35. **Ostermann, E., F. Kricek, and G. Hogenauer.** 1984. Cloning the origin of transfer region of the resistance plasmid R1. EMBO J. **3**:1731–1735.
36. **Queen, C., and L. J. Korn.** 1984. A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids Res. **12**:581–599.
37. **Rupp, W. D., and G. Ihler.** 1968. Strand selection during bacterial mating. Cold Spring Harbor Symp. Quant. Biol. **33**:647–650.
38. **Sharp, P. A., S. N. Cohen, and N. Davidson.** 1973. Electron microscope heteroduplex studies of sequence relations among plasmids of *E. coli*. II. Structure of drug resistance (R) factors and F factors. J. Mol. Biol. **75**:235–255.
39. **Silverman, P. M.** 1985. Host cell-plasmid interactions in the expression of DNA donor activity by F^+ strains of *Escherichia coli* K-12. BioEssays **2**:254–259.
40. **Thompson, R., L. Taylor, K. Kelly, R. Everett, and N. Willetts.** 1984. The F plasmid origin of transfer: DNA sequence of wild-type and mutant origins and location of origin-specific nicks. EMBO J. **3**:1175–1180.
41. **Willetts, N., and J. Maule.** 1986. Specificities of IncF plasmid conjugation genes. Genet. Res. **47**:1–11.
42. **Willetts, N., and B. Wilkins.** 1984. Processing of plasmid DNA during bacterial conjugation. Microbiol. Rev. **48**:24–41.
43. **Willetts, N. S., and R. Skurray.** 1980. The conjugative system of F-like plasmids. Annu. Rev. Genet. **14**:41–76.