Location and nucleotide sequence of the transfer origin of the broad host range plasmid RK2

(bacterial conjugation/relaxation complex/DNA structure)

DONALD G. GUINEY AND EMANUEL YAKOBSON

Department of Medicine H811F, University of California, San Diego Medical Center, 225 Dickinson Street, San Diego, California 92103

Communicated by Donald R. Helinski, March 10, 1983

ABSTRACT The origin of plasmid DNA transfer, *oriT*, has been localized on RK2, a conjugative drug-resistance plasmid of the IncP group with a very broad host range in gram-negative bacteria. The transfer origin is contained in a 760-base-pair *Hae* II restriction fragment that maps in the same region as the single-strand nick made by the RK2 relaxation complex. The functional *oriT* was subcloned as a 112-base-pair *Hpa* II fragment, and the DNA sequence of this region was determined. The dominant structural feature of the *oriT* sequence is a 19-base-pair inverted repeat, with 15 of the 19 bases able to form pairs in a hairpin structure. This inverted repeat may be the recognition site for the relaxation complex proteins, which nick the plasmid DNA molecule and initiate the transfer process.

The transfer of plasmid DNA during bacterial conjugation begins at a specific site on the plasmid DNA molecule: the origin of transfer, *oriT*. The transfer origin was initially defined genetically as a site required in *cis* for plasmid transfer during mating (1, 2). Studies with the F plasmid of *Escherichia coli* have shown that a specific single strand of the DNA molecule is transferred with a leading 5' end (3, 4). DNA transfer is initiated by a strand- and site-specific nick at *oriT*. The *oriT* of the F plasmid has been mapped near the start of the F transfer operon, within a 385-base-pair (bp) DNA fragment, and an *oriT*-specific nicking activity is encoded by the *traY* and *traZ* gene products (5, 6). The discovery of plasmid DNA-protein relaxation complexes has facilitated the study of the nicking process in other plasmids. Many bacterial plasmids can be isolated in the form of a superhelical DNA-protein relaxation complex, including R factors and colicinogenic plasmids (7–9). Upon treatment *in vitro* with detergents, proteases, or ethidium bromide, the supercoiled DNA is nicked in one strand and converted to the open circular form, a process called relaxation. In the case of the ColE1 plasmid, the nick occurs at a specific site within a 320-bp region required in *cis* for transfer, the presumed *oriT* (10). A 60,000-dalton protein remains covalently attached to the 5' end of the nicked strand and probably facilitates transfer to the recipient cell (11–13). The ColE1 plasmid is not self-transmissible and requires the presence of a conjugative plasmid in the same cell for mobilization during mating. However, study of the self-transmissible, broad host range IncP plasmid RK2 has revealed a similar property of the relaxation complex: the site-specific nicking event occurs within a 5-kilobase (kb) region of the plasmid that also contains *oriT* (14). In this paper, *oriT* of RK2 is localized to a 760-bp region that coincides with the relaxation nick. The functional *oriT* is contained within a 112-bp fragment, and the DNA sequence of this region has been determined.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Bacterial Strains and Plasmids. The *E. coli* strains used were HB101 (*pro*, *leu*, *thi*, *lacY*, *str*^R, *hsr*[−], *hsm*[−], *endoI*[−], *recA*), C600 Mu⁺ (*thr*, *leu*, *thi*, *SupE46*, *lacY*, *tonA*), and HB101 *nal*^R. The vector plasmid pDGV1 was derived from pMK2004 (15), a pBR322 derivative, by deletion of the *Hind*III fragment containing the tetracycline-resistance region and part of the kanamycin-resistance gene. pDGV1 is 2.7 kb in size and contains the pBR322 replicon and ampicillin-resistance gene but is missing the *nic/bom* site (*oriT*) of pBR322. RK2, pRK231, and pRK215.1 have been described (14).

Plasmid DNA Methodology. The techniques for plasmid construction, mapping of restriction sites, and detection of relaxation complex have been described (14, 15). All enzymes were from Bethesda Research Laboratories.

Matings. The recombinant derivatives of pDGV1 to be tested for *oriT* function were transformed into HB101 containing pRK231, a *tra*⁺ derivative of RK2 with a *Pst* I deletion in the β -lactamase gene, rendering it penicillin sensitive. The donor strain was crossed with the recipient, HB101 *nal*^R, by the filter mating technique as described (14). Transfer of the pDGV1 derivative was selected on LB plates (14) with penicillin (200 μ g/ml) and nalidixic acid (25 μ g/ml), whereas transfer of the pRK231 helper plasmid was detected on plates containing kanamycin (50 μ g/ml) and nalidixic acid (25 μ g/ml).

DNA Sequence Analysis. The end-labeling and chemical-cleavage method of Maxam and Gilbert was used (16). The DNA sequence was read from 25% and 12% acrylamide sequence determination gels.

RESULTS

Cloning of *oriT*. The physical and genetic map of RK2 is presented in Fig. 1. Previous results had located the site of the relaxation complex nick, *rlx*, at 48 kb on the map, and genetic evidence showed that the 5-kb region surrounding this site was required in *cis* for transfer. Furthermore, the 10-kb region, from 46 to 56 kb, possessed RK2-specific *oriT* activity when cloned onto pBR322. To further localize the functional *oriT* within this 10-kb region, a number of deletions were generated. To avoid any extraneous *oriT* activity from the vector, the 10-kb region was cloned onto pDGV1. This plasmid is a 2.7-kb derivative of pMK2004 (15) in which the pBR322 *nic/bom* site has been removed by a *Hind*III deletion. pDG4 was constructed by ligating the 12-kb *Eco*RI-*Hind*III fragment of pRK215.1 into the *Eco*RI and *Hind*III sites of pDGV1. As shown in Fig. 1, pDG4 contains the 10-kb region of RK2 previously shown to possess *oriT* activity. Deletions of pDG4 were generated by partial *Hae* II digestion, ligation, and transformation of *E. coli* C600 Mu⁺ selecting for ampicillin resistance. The deletions were located

Abbreviations: bp, base pair(s); kb, kilobase(s).

FIG. 1. Map of RK2 showing the regions cloned on pDG4. The location of the vegetative origin of replication, *oriV*, is from Meyer and Helinski (17). The three *tra* regions are taken from the Tn7 insertion map of RP4 (18, 19), a plasmid closely related or identical to RK2 (20). The site of the relaxation complex nick, *rlx*, was mapped previously (14). The direction of transfer from the origin, *oriT*, is as described by Grinter (21).

by reference to a *Hae* II map of pDG4. Two deletion derivatives of pDG4 are shown in Fig. 1. pDG4Δ22 contains a 3-kb region surrounding *rlx* and a 1-kb region adjacent to the *Eco*RI site, whereas pDG4Δ27 contains only the *Hae* II fragment next to the *Eco*RI site. These recombinant plasmids were transformed into HB101 and tested for mobilization by the *tra*⁺ RK2 helper plasmid pRK231. pDG4 and pDG4Δ22 were transferred at high frequency, whereas pDG4Δ27 and the vector pDG5 were transferred at low or below detectable levels (Table 1). These results show that the functional *oriT* lies in the 3-kb region of pDG4Δ22 containing *rlx*. The *oriT* location was refined by a further *Hae* II partial digestion and ligation of pDG4Δ22 to generate pDG5 (Figs. 1 and 2). pDG5 contains a 760-bp *Hae* II fragment from the *rlx* region linked to the 1-kb *Hae* II fragment adjacent to the *Eco*RI site in RK2. As shown in Table 1, pDG5 is efficiently mobilized by pRK231, indicating that the fully functional *oriT* is contained within this 760-bp *Hae* II fragment (designated *Hae* II fragment 9 from pDG4). The RK2 DNA next to the *Eco*RI site, *Hae* II fragment "6" in Fig. 2, is not required for *oriT* activity because the purified fragment 9 was cloned in several plasmid vectors and expresses full *oriT* function (data not shown). A restriction map of this *Hae* II fragment 9 was constructed for the enzymes *Ava* I, *Hinf* I, and *Hpa* II (Fig. 2). The fragment has four internal *Hpa* II sites, producing

FIG. 2. Map of pDG5, with upper and lower scales in kb. The *Hae* II numbers refer to fragments from pDG4 that are retained in pDG5. *Hae* II fragments 1 and 12 contain the replication region and ampicillin-resistance genes from the vector pDG5. The RK2 DNA present in pDG5 is indicated by the heavy line. Fragment 9 maps in the region of the *rlx* site, as shown in Fig. 1. Fragment 6 and the adjacent portion of fragment 1 map just counterclockwise to the *Eco*RI site of RK2. An expanded restriction map of fragment 9 is shown also for the enzymes *Hpa* II, *Ava* I, and *Hinf* I. The letters A, B, C, D₁, and D₂ refer to the five fragments that result from a *Hpa* II digestion of purified fragment 9 DNA.

five subfragments. The D fragments are the same size but can be distinguished by the *Hinf* I site in D₂. To further localize the *oriT* region, purified *Hae* II fragment 9 DNA from pDG5 was digested with *Hpa* II and ligated into the single *Cla* I site of pDG5 (corresponding to the *Cla* I site at 24 bp on the pBR322 map) (22). Plasmids containing *Hpa* II inserts were selected by redigestion with *Cla* I followed by transformation into HB101. These recombinant plasmids were then tested for *oriT* activity by mobilization with pRK231. Only plasmids containing *Hpa* II fragment D₂ were transferred at high frequency by pRK231, and one of these, designated pEYDG59, was chosen for further study. pEYDG59 contains the 112-bp *Hpa* II fragment D₂ inserted in the *Cla* I site of pDG5 between the adjacent *Eco*RI and *Hind*III sites (Fig. 3). pEYDG59 is mobilized by the helper plasmid pRK231, but the transfer frequency is about 1/50 that of pDG4, pDG4Δ22, or pDG5 (Table 1). Because the transfer frequency of pEYDG59 was somewhat lower, the transconjugants were analyzed to rule out mobilization by a non-*oriT*-mediated mechanism such as transposition or cointegrate formation. Mobilization of nonconjugative plasmids can occur by recombination with the self-transmissible plasmid. This mo-

FIG. 3. Restriction map of pEYDG59, containing the 112-bp *Hpa* II D₂ fragment cloned into the *Cla* I site of pDG5, as indicated by the thick line. The strategy for sequence analysis of this fragment is shown by the arrows at the top of the figure. By labeling the 5' ends of the *Eco*RI and *Hind*III sites, the entire sequence of the D₂ fragment could be read in both strands.

Table 1. Mobilization of RK2 derivative plasmids

Plasmid	Mobilization by pRK231,* transfer frequency	Relaxability, [†] %
pDG4	1	20
pDG4Δ22	2	<5
pDG4Δ27	<1 × 10 ⁻⁵	NT
pDG5	2	<5
pEYDG59	4 × 10 ⁻²	NT
pDG5	4 × 10 ⁻⁵	NT

NT, not tested.

* Number of penicillin-resistant transconjugants per input donor cell in 3-hr matings.

[†] Percentage of total plasmid DNA in the open circular form after NaDodSO₄ treatment of the purified relaxation complex as described (13, 14).

FIG. 4. Mobilization of pEYDG59 by pRK231 in *E. coli* HB101. Plasmid DNA was extracted from the donor (lane A), and seven transconjugants (lanes B-H) selected for transfer of pEYDG59. The DNA samples were electrophoresed in a 0.7% agarose gel.

bilization is not dependent on *oriT* and occurs by cointegrate formation as a result of *recA*-dependent homologous recombination or as an intermediate in the transposition of an insertion element (23). Fig. 4 shows the results of a plasmid DNA analysis of the donor, HB101 (a *recA* strain) containing both pRK231 and pEYDG59, and several HB101 *nal^R* transconjugants selected for transfer of pEYDG59. Each transconjugant contains pEYDG59 in an unaltered form, making mobilization by *recA*-independent recombination very unlikely.

Sequence of the *oriT* Region. The nucleotide sequence of the 112-bp *Hpa* II D_2 fragment was obtained by the Maxam and Gilbert technique (16). The strategy for sequence analysis is shown in Fig. 3. By separate end-labeling of the *Eco*RI and *Hind*III sites, the sequence of the entire 112-bp *Hpa* II insert was determined in both strands and is shown in Fig. 5. The major structural feature of the sequence is the presence of a region of 2-fold rotational symmetry around bp 79 and 80. This inverted repeat extends for 19 bp in either direction and gives rise to a potential hairpin or cruciform structure with extensive intrastrand base pairing, as shown in Fig. 6. There is an inner 7-bp perfect inverted repeat that is very G+C rich. In all, 16 of the 19 bases are able to form pairs in this potential hairpin structure, although steric constraints in the loop may not allow formation of the first G-C base pair in the inverted repeat (24). This inverted repeat is the dominant structural feature of the sequence and comprises 40 of the 112 bases in the functional *oriT*. The sequence also contains a direct repeat of 12 bp, as indicated by the dotted lines in Fig. 5. There are only two base changes in these repeats, and one direct repeat overlaps with the inverted repeat structure.

Presence of Relaxation Complex in Deletions of pDG4. In order to localize the coding regions for the relaxation complex proteins, pDG4 and its derivatives containing the *rlx* site were

FIG. 5. DNA sequence of the 112-bp *Hpa* II D_2 fragment containing *oriT*. The position of the inverted repeat is shown by the solid arrows, and the direct repeat, by the dashed arrows. The internal *Hind*III site permitted orientation with respect to the vector.

FIG. 6. Potential hairpin structure involving the inverted repeat region of *oriT*. The first G-C base pair in the inverted repeat (positions 78 and 81) probably does not form because hairpin loops of less than three bases are sterically unfavorable (24, 25). The free energy of formation for this structure, calculated by the method of Tinoco *et al.* (25), is $-10.8 \text{ kcal mol}^{-1}$, if one assumes a hairpin loop of four bases and the internal loops of two and four bases.

tested for relaxation complex formation (Table 1). About 20% of pDG4 plasmids, isolated from Triton X-100-lysed cells, were relaxed by treatment with NaDodSO₄. This level of relaxation complex is comparable to the 30% found for the RK2 parent plasmid (14). Because the vector in pDG4 had been deleted for the pBR322 *nic/bom* site, this result indicates that pDG4 contains all of the information necessary for the RK2 relaxation complex. However, no relaxation complex was detected for pDG4 Δ 22 and pDG5 (Table 1), even though both of these plasmids still contain the *rlx* site. This finding suggests that at least one component of the RK2 relaxation complex is encoded in the region of pDG4 that is missing in pDG4 Δ 22 (see Fig. 1).

DISCUSSION

oriT is the region of principle interest in studying the mechanism of plasmid DNA transfer during conjugation. The major biochemical events of DNA transfer occur at *oriT*, including nicking of the closed circular DNA, conduction of the single strand from donor to recipient, initiation of conjugal DNA synthesis, and recircularization in the recipient. The results pre-

sented in this study show that the functional RK2 transfer origin is contained in a 112-bp DNA fragment. This region fulfills genetic criteria for an origin of transfer including: (i) the *oriT* site is required in *cis* for transfer; (ii) *oriT* confers high-frequency mobilization when the conjugation system is supplied in *trans*; (iii) high-frequency mobilization by the helper plasmid is *recA* independent; and (iv) cointegrate formation is not involved in mobilization. The 760-bp *Hae* II *oriT* fragment in pDG5 increases the mobilization frequency of the vector, pDGV1, by a factor of 10^5 (Table 1), whereas the 112-bp *Hpa* II *oriT* region raises the transfer frequency by 10^3 . This transfer is not due to homologous recombination because all the matings were done in a *recA* strain. In addition, previous studies showed that even large derivatives of RK2, containing extensive regions of homology, are not mobilized in a *recA* strain if they are missing the *oriT* site (14). Transfer of nonconjugative plasmids also can occur by cointegrate formation during the *recA*-independent transposition of an insertion element. When the cointegrate is resolved in the recipient, the mobilized plasmid contains a copy of the insertion element or transposon (23). This form of transfer is not dependent on the presence of *oriT* or mobilization functions in the nonconjugative plasmid. However, such a transfer process was ruled out for pEYDG59 mobilization by the data in Fig. 4: all the transconjugants contained pEYDG59 in an unaltered form. Finally, previous results have shown that high-frequency mobilization of plasmids containing the RK2 *oriT* region is specific for the RK2 transfer system, in that *Flac* and R64*drd* 11 do not mobilize the RK2 *oriT* (14). These results show that the *oriT* sequence on pEYDG59 interacts in *trans* with the RK2 conjugation system on pRK231 to promote high-frequency transfer of pEYDG59.

The dominant feature of the *oriT* sequence is a 19-bp inverted repeat, which could form a potential cruciform structure in the double-stranded supercoiled molecule or a hairpin structure in single-stranded DNA (Fig. 6). Cruciform structures may function as recognition sites for proteins, and the inverted repeat in *oriT* could form the binding site for the RK2 relaxation complex proteins. The supercoiling of plasmid DNA favors the formation of cruciform structures (26), and inverted repeats of only 9 bp in pBR322 have been detected by the S1 nuclease cleavage assay (27). Alternatively, the hairpin depicted in Fig. 6 could form in the single-stranded DNA during transfer and provide a secondary structure that interacts with the mating apparatus during conduction of the DNA from donor to recipient.

Studies of the ColE1 and R6K plasmids also have correlated the site of the relaxation complex nick with the functional origin of transfer. The DNA sequence around the ColE1 relaxation complex nick site is known (28), and this site coincides with the 320-bp basis of mobility, or *bom* region, which is required in *cis* for transfer (10). Further mapping by deletions has located *bom* within a 141-bp *Hha* I fragment, which presumably constitutes the ColE1 *oriT* (29). However, deletion of sequences adjacent to this *Hha* I fragment results in a 95–99% reduction in the frequency of mobilization. This finding is analogous to the RK2 *oriT* data in Table 1: the 112-bp *Hpa* II *D*₂ fragment in pEYDG59 is mobilized about $1/50$ as efficiently as the 760-bp *Hae* II fragment in pDG5. Thus, the functional *oriT* regions for ColE1 and RK2 are contained within small DNA fragments, but sequences adjacent to these regions are required for full *oriT* activity. In the case of R6K, the relaxation complex nick can be located near either the α or β origin of vegetative DNA replication (30). Recently, Shafferman and Helinski have determined the sequence of the entire 1,962-bp β origin of replication and have shown that this region also contains a functional *oriT* (31). The R6K sequence shares a 34-bp region of

homology with the ColE1 and pBR322 *nic/bom* sites. These results led Shafferman and Helinski to propose a consensus sequence for the *oriT* regions of ColE1, pBR322, and R6K. This consensus sequence is not present in the 112-bp *oriT* fragment of RK2. Furthermore, low-stringency hybridizations by the Southern blot technique have failed to reveal homology between the 760-bp RK2 *Hae* II *oriT* fragment and ColE1, pBR322, R6K, or F (data not shown). This lack of homology is not surprising because the latter plasmids all have a narrow host range and are found mainly in *E. coli* or related enteric bacteria, whereas the host range of RK2 is very broad and includes many gram-negative organisms. The identification and characterization of the RK2 *oriT* region will facilitate the use of the broad host range conjugative properties of RK2 for genetic manipulation in a variety of bacterial hosts.

We thank Catherine Lee and Marcin Filutowicz for invaluable help in the sequence determinations. The expert technical assistance of Karen Arnold is appreciated. This work was supported by Grant GM 28924 from the National Institute of General Medical Sciences. D.G.G. is a Teaching and Research Scholar of the American College of Physicians.

- Willetts, N. S. (1972) *J. Bacteriol.* **112**, 773–778.
- Guyer, M. S. & Clark, A. J. (1976) *J. Bacteriol.* **125**, 233–247.
- Rupp, W. & Ihler, G. (1968) *Cold Spring Harbor Symp. Quant. Biol.* **33**, 647–650.
- Ohki, M. & Tomizawa, J. (1968) *Cold Spring Harbor Symp. Quant. Biol.* **33**, 651–658.
- Everett, R. & Willetts, N. S. (1982) *EMBO J.* **1**, 747–753.
- Everett, R. & Willetts, N. S. (1980) *J. Mol. Biol.* **136**, 129–150.
- Clewell, D. & Helinski, D. R. (1969) *Proc. Natl. Acad. Sci. USA* **62**, 1159–1166.
- Morris, C., Hershberger, C. & Rownd, R. (1973) *J. Bacteriol.* **114**, 300–308.
- Kupersztach-Portnoy, Y., Miklos, G. & Helinski, D. R. (1974) *J. Bacteriol.* **120**, 545–548.
- Warren, G., Twigg, A. & Sherratt, D. J. (1978) *Nature (London)* **274**, 259–261.
- Blair, D. & Helinski, D. R. (1975) *J. Biol. Chem.* **250**, 8785–8789.
- Lovett, M. A. & Helinski, D. R. (1975) *J. Biol. Chem.* **250**, 8790–8795.
- Guiney, D. G. & Helinski, D. R. (1975) *J. Biol. Chem.* **250**, 8796–8803.
- Guiney, D. G. & Helinski, D. R. (1979) *Mol. Gen. Genet.* **176**, 183–189.
- Kahn, H., Kolter, R., Thomas, C., Figurski, D., Meyer, R., Remaut, E. & Helinski, D. R. (1979) *Methods Enzymol.* **68**, 368–380.
- Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
- Meyer, R. & Helinski, D. R. (1977) *Biochim. Biophys. Acta* **478**, 109–113.
- Barth, P. T., Grinter, N. J. & Bradley, D. E. (1978) *J. Bacteriol.* **133**, 43–52.
- Barth, P. T. (1978) in *Plasmids of Medical, Environmental, and Commercial Importance*, eds. Timmis, K. N. & Puhler, A. (Elsevier/North-Holland, Amsterdam), pp. 399–410.
- Burkhardt, H., Reiss, G. & Puhler, A. (1979) *J. Gen. Microbiol.* **114**, 341–348.
- Grinter, N. J. (1981) *Plasmid* **5**, 267–276.
- Sutcliffe, J. G. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 77–97.
- Clark, A. & Warren, G. (1979) *Annu. Rev. Genet.* **13**, 99–125.
- Scheffler, I. E., Elson, E. & Baldwin, R. L. (1970) *J. Mol. Biol.* **48**, 145–171.
- Tinoco, I., Uhlenbeck, O. C. & Levine, H. D. (1971) *Nature (London)* **230**, 362–367.
- Mizuuchi, K., Mizuuchi, M. & Gellert, M. (1982) *J. Mol. Biol.* **156**, 229–243.
- Lilley, D. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6468–6472.
- Bastia, D. (1978) *J. Mol. Biol.* **124**, 601–639.
- Finnegan, J. & Sherratt, D. (1982) *Mol. Gen. Genet.* **185**, 344–351.
- Nordheim, A., Hashimoto-Gotoh, T. & Timmis, K. N. (1980) *J. Bacteriol.* **144**, 923–932.
- Shafferman, A. & Helinski, D. R. (1983) *J. Biol. Chem.* **258**, 4083–4090.