JOURNAL OF BACTERIOLOGY, June 1988, p. 2749–2757
0021-9193/88/062749-09$02.00/0
Copyright © 1988, American Society for Microbiology
Vol. 170, No. 6

Identification and Characterization of the Products from the *traJ* and *traY* Genes of Plasmid R100

SUSUMU INAMOTO, YASUSHI YOSHIOKA, AND EIICHI OHTSUBO*

Institute of Applied Microbiology, University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan

Received 10 December 1987/Accepted 21 March 1988

The nucleotide sequence of part of the *tra* region of R100 including *traJ* and *traY* was determined, and the products of several *tra* genes were identified. The nucleotide sequence of *traJ*, encoding a protein of 223 amino acids, showed poor homology with the corresponding segments of other plasmids related to R100, but the deduced amino acid sequences showed low but significant homology. The first four amino acids at the N-terminal region of the TraJ protein were not essential for positive regulation of expression of *traY*, the first gene of the *traYZ* operon. The nucleotide sequence of *traY* shows that this gene may use TTG as the initiation codon and that it encodes a protein of 75 amino acids. Analysis of the *traY* gene product, which was obtained as the fusion protein with β-galactosidase, showed that the N-terminal region of the product has an amino acid sequence identical to that deduced from the assigned frame but lacks formylmethionine. *traY* of plasmid F, which encodes a larger protein than the TraY protein of R100, is thought to use ATG as an initiation codon. However, a TTG initiation codon was found in the preceding region of the previously assigned *traY* coding frame of F. Interestingly, when translation of *traY* of F was initiated from TTG, the amino acid sequence homologous to the TraY protein of R100 appeared in tandem in the TraY protein of F. This may suggest that *traY* of F has undergone duplication of a gene like the *traY* gene of R100.

Resistance (R) plasmids such as R100 and R1, belonging to the FII incompatibility group (4), confer conjugal transfer ability, a process in which DNA is transmitted from one bacterial host to another; this requires cell-to-cell contact and synthesis of sex pili by donor bacteria (for reviews, see references 21, 47, and 48). These plasmids share homology with the fertility factor, F, in the *tra* region, which is responsible for DNA transfer (41) and which is about 33 kilobase pairs in length. The *tra* region of F contains at least 26 genes which are organized into three main operons, *traM*, *traJ*, and *traYZ* (2, 19; for a review, see reference 21). Transcription of the *traYZ* operon in F and R1 has been shown to be positively regulated at promoter p_{YZ} by the product of *traJ* (8, 12, 13, 16, 32). Expression of *traJ* itself is repressed by the fertility inhibition complex, FinOP, which is composed of the products of two genes, *finO* and *finP*, encoded by the R plasmids and F (11, 16, 32). The initial event in DNA transfer upon expression of the *traYZ* operon is strand- and site-specific nicking at the origin of transfer, *oriT*, by the plasmid-specified endonuclease, which is thought to be a complex of the products of *traY* and *traZ* (5).

In this paper, we report the nucleotide sequence of the promoter-proximal region of *tra* of R100, including *traJ* and *traY*, and the identification of gene products of this region. The TraJ protein of R100 may positively regulate expression of the *traY* gene of R100, which is shown to use TTG as the initiation codon. We point out that *traY* of F is a gene consisting of the two DNA segments, each of which is homologous to *traY* of R100. This may indicate a simple but clear example for gene duplication giving rise to a polypeptide with two identical domains.

MATERIALS AND METHODS

Bacterial strains and plasmids. The *Escherichia coli* K-12 strains used were JM109 [*recA1* Δ(*lac-pro*) *endA1 gyrA96 thi-1 hsdR17 supE44 relA1* λ⁻/F' *traD36 proAB lacI*^{q}*ZΔM15*] (49), CSR603 [F⁻ *uvrA6 recA1 phr-1 thr-1 leuB6 proA2 thi-1 argE3 lacY1 galK2 ara-14 xyl-5 mtl-1 rpsL31* (Str^r) *tsx-33 supE33*] (39), and MC1000 [F⁻ *araD139* Δ(*ara-leu*)*7697* Δ*lac X74 galU galK strA*] (3).

The plasmids used were R100 (33), pUC19 (49), and pJG200 (17). The pSI plasmids used, except pSI200, are listed in Fig. 1. They were constructed as described below.

Construction of pSI plasmids. pSI28 and pSI87 were constructed by cloning the *Eco*RI *r*4 fragment of R100 into pUC19. They were isolated from Lac⁻ and Amp^r (ampicillin resistance) transformants of JM109. The other pSI plasmids were obtained from pSI28 and pSI87 as follows. pSI28-S2 was constructed by ligation of pSI28 after it was partially digested with *Sma*I. pSI28-B3 and pSI28-B4 were constructed by ligation of pSI28 after it was partially digested with *Bam*HI. pSI87-B19 and pSI87-B18 were constructed by ligation of pSI87, which was partially digested with *Bam*HI. pSI87-S5 was constructed by ligation of pSI87 after it was partially digested with *Sma*I. pSI87-BB was obtained by ligation of pSI87-B18 after it was digested with *Bgl*II and *Bam*HI. pSI87-A2 and pSI87-A9 were constructed by ligation of pSI87 after it was partially digested with *Acc*I and then treated with DNA polymerase I (Klenow) to blunt the cohesive ends. pSI87-ES3 was constructed by ligation of pSI87-B19 after it was partially digested with *Eco*RV, completely digested with *Sal*I, and then treated with DNA polymerase I (Klenow). pSI87-XS5 was constructed by ligation of pSI87-B19 after it was partially digested with *Xmn*I, completely digested with *Sal*I, and then treated with DNA polymerase I (Klenow). pSI87-HE was constructed by ligation of pSI87-B18 after it was digested with *Hpa*I and *Eco*RI and then treated with DNA polymerase I (Klenow). pSI87-BE was constructed by ligation of pSI87-B19 after it was digested with *Bgl*II and *Eco*RI and then treated with DNA polymerase I (Klenow). pSI87-AE was constructed by ligation of pSI87-A9 after it was digested with *Acc*I and *Eco*RI and then treated with DNA polymerase I (Klenow). pSI87-XE1 and pSI87-XE2 were obtained by ligation of

* Corresponding author.

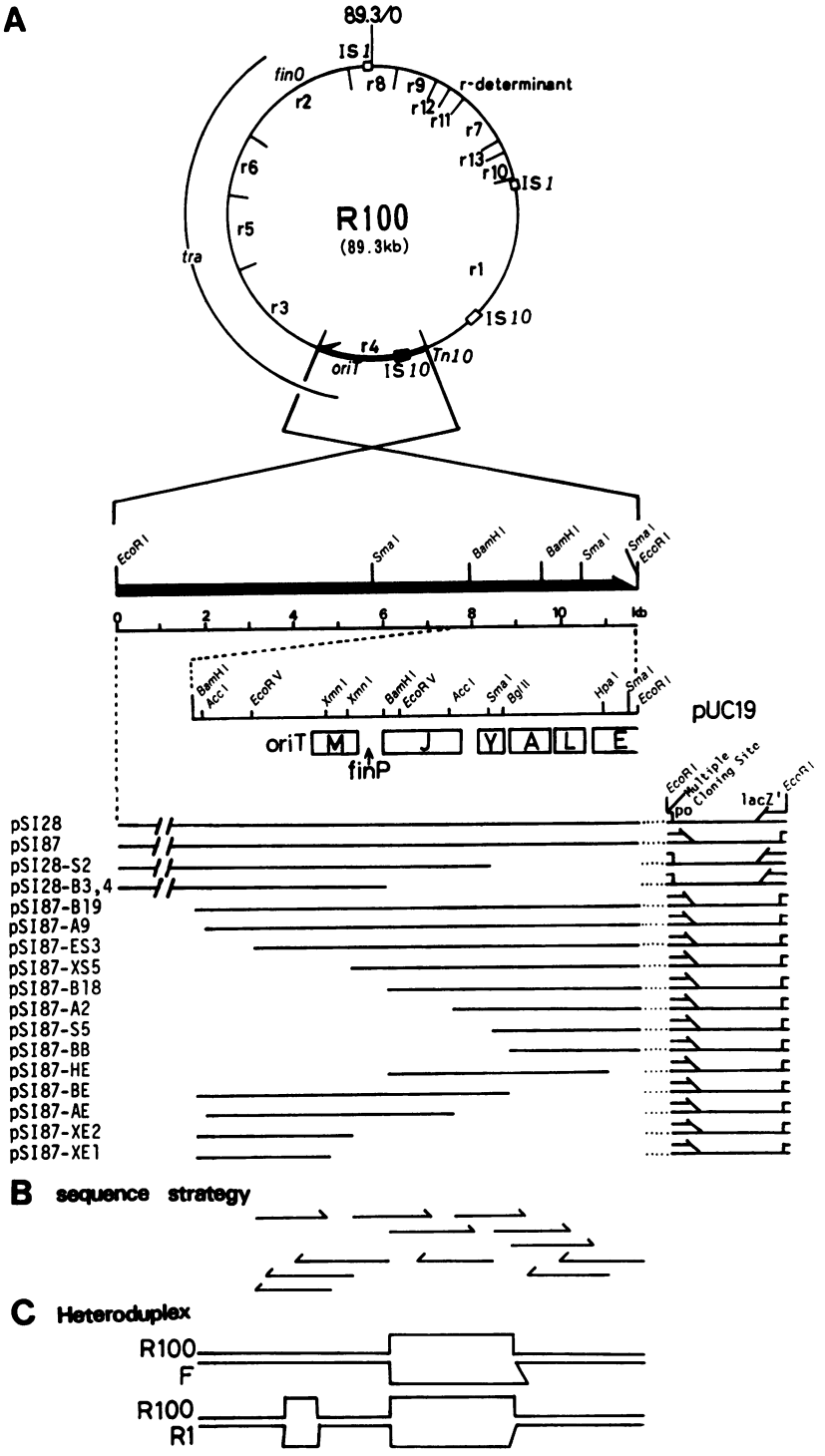

FIG. 1. Structures of R100 and pSI plasmids. (A) A circular map of R100 with the kilobase-pair coordinate 89.3/0 (34; see also reference 50). Positions of the *Eco*RI fragments (35, 42) are shown inside the circle. The resistance determinant region, flanked by the insertion sequence IS*1* (small open boxes), and the tetracycline resistance transposon Tn*10*, flanked by IS*10* (large open boxes), are shown. *tra*, *oriT*, and *finO* are explained in the text. The *Eco*RI fragment *r*4, approximately 11.7 kilobase pairs in length, and a second region of about 4 kilobase pairs are expanded. The structures of the pSI plasmids are shown below the expanded area. The expanded region also shows major restriction sites and the *tra* genes which were deduced from the nucleotide sequence relations between F and R100. Solid lines show portions of the *r*4 fragment carried by the plasmids. In each pSI plasmid, the orientation of the sequence of the vector plasmid pUC19, carrying the *lacZ'* gene (open arrow) and the *lacPO* region, which contains multiple cloning sites, such as *Eco*RI, *Sma*I, *Bam*HI, *Acc*I, and *Sal*I, is shown. Note that all the plasmids derived from pSI28 and pSI87 lack one of the two *Eco*RI sites located between the *lacPO* and R100 sequences in pSI28 and pSI87, while the last five derivatives shown at the bottom of panel A also lack the other *Eco*RI site located between the *lacZ'* and R100 sequences. This occurred during the construction of these plasmids from pSI28 and pSI87, as described in Materials and Methods. The

pSI87-B19 after it was partially digested with *Xmn*I, completely digested with *Eco*RI, and then treated with DNA polymerase I (Klenow). All the pSI plasmids were obtained from Amp^r transformants of JM109.

The construction of pSI200 is described in Results.

Media. The culture media used were LB broth (31), L-rich broth (50), and φ medium (50). φ medium was used for transformation of plasmid DNA. For transformation of plasmid DNA, we used the procedures previously described (50). L-agar plates contained 1.5% Bacto-Agar (Difco Laboratories) in LB broth. The agar plates used to select Lac$^-$ and Amp^r transformants contained 50 μg of ampicillin (Sigma Chemical Co.) per ml, 40 μg of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (Sigma) per ml, and 140 mM isopropyl-β-D-thiogalactopyranoside (IPTG) (Sigma).

Enzymes. Restriction endonucleases (*Eco*RI, *Hpa*I, *Eco*RV, *Acc*I, *Bam*HI, *Bgl*II, *Sal*I, and *Sma*I), T4 DNA ligase, and the Klenow fragment of DNA polymerase I were obtained from Takara Shuzo Co. Restriction endonuclease *Xmn*I was obtained from New England BioLabs, Inc. Ribonuclease A was purchased from Sigma. All enzymes were used as recommended by their suppliers.

DNA preparation. The *E. coli* strain harboring a plasmid was grown in L-rich broth. Plasmid DNA was isolated as described by Ohtsubo et al. (34). The crude lysis method was used to check and isolate a small amount of plasmid DNA from large numbers of cell cultures (27). An alkaline lysis method (28, 50) was used to prepare plasmid DNA for nucleotide sequencing.

Nucleotide sequencing. Nucleotide sequences were determined by the dideoxynucleotide method (30, 40). We used sequencing kits (Takara Shuzo), DNA polymerase I (Klenow), a 15-mer M13 primer, M1 (Takara Shuzo), and a 17-mer reverse primer (Amersham Corp.). The DNA chains were labeled with [α-^{32}P]dCTP (Amersham) and were separated on 8, 6, and 4% polyacrylamide gels containing 8 M urea.

Identification of the *tra* gene products. Plasmid-carrying derivatives of *E. coli* CSR603 were irradiated with UV light according to the maxicell procedure (39), and proteins were labeled with [^{3}H]lysine (40 μCi in a 1-ml culture; Amersham). The labeled proteins produced were analyzed on an 11% polyacrylamide gel containing 0.5% (wt/vol) sodium dodecyl sulfate (SDS) (20). Protein molecular weight standards (Bethesda Research Laboratories, Inc.) were identified by staining with Coomassie brilliant blue (Sigma) according to Fairbanks et al. (6). The SDS sample buffer (×2) contained 20 mM sodium phosphate (pH 7.2), 4% (wt/vol) SDS, 20% (wt/vol) glycerol, 20% (wt/vol) 2-mercaptoethanol, and 0.01% (wt/vol) bromophenol blue.

Preparation of the β-galactosidase-specific affinity matrix. *p*-amino-phenyl-β-D-thiogalactoside (TPEG) (Sigma) was attached to ECH Sepharose 4B (Pharmacia Fine Chemicals) with *N*-ethyl-*N'*-(3-dimethyl-aminopropyl)carbodiimide hydrochloride (Sigma). The gel was washed four times in turn with 0.1 M Tris hydrochloride, 0.5 M NaCl (pH 8.0), and 0.1 M acetate buffer (pH 4.0) containing 0.5 M NaCl. It was then washed with distilled water and AF buffer (20 mM Tris hydrochloride, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 1.6 M NaCl [pH 7.4]) and stored at 4°C.

Purification of the tripartite fusion protein. MC1000 harboring pSI200 was grown with aeration in 250 ml of L-rich broth containing 0.2% (wt/vol) glucose and 50 μg of ampicillin per ml at 30°C to an optical density of 0.5 to 0.6 at A_{600}, and the culture was then incubated with aeration at 42°C for 1 h and chilled in an ice bath. (Unless otherwise indicated, all subsequent operations were carried out at 0 to 4°C.) Cells were harvested, washed with ice-cold AF buffer with 250 μM phenylmethylsulfonyl fluoride, frozen in dry ice, and stored at −70°C. The frozen cells were thawed, suspended uniformly in 15 ml of ice-cold AF buffer with 250 μM phenylmethylsulfonyl fluoride, sonicated in an ice bath by using a Branson sonic oscillator until the cells were disrupted, and centrifuged at 11,000 × *g* for 10 min to remove unbroken cells. The membrane fraction was then removed by centrifugation of the clarified extract at 70,000 × *g* for 1 h. The tripartite fusion protein was purified from the supernatant by using TPEG-Sepharose according to the procedure described by Ullmann (45). Borate buffer (100 mM sodium borate, 10 mM 2-mercaptoethanol [pH 10.0]) was used to elute the protein and was promptly replaced by buffer B (10 mM Tris hydrochloride, 250 mM NaCl, 10 mM $CaCl_2$, 10 mM 2-mercaptoethanol [pH 7.4]) by using column PD-10 (Pharmacia). The presence of the tripartite fusion protein was monitored by assaying β-galactosidase activity as described elsewhere (31) or by electrophoresis of the protein in an SDS-8% polyacrylamide gel according to Laemmli (26).

Amino acid sequencing. A sample of the purified tripartite fusion protein solution was dialyzed against several changes of distilled water at 0°C and concentrated by using a centrifugal evaporator (model EC-57; Sakuma) under reduced pressure at room temperature until the volume was appropriate to apply to the sequencer. The sample was analyzed for the N-terminal amino acid sequence with a gas phase sequencer (model 470A Gas Phase Sequenator with an on-line model 120A PTH Amino Acid Analyzer; Applied Biosystems).

Computer analysis. Dot matrix analyses were performed with the program developed by Y. Maeda in our laboratory, using the PC-9801 personal computer system (Nippon Electric Co.). The amino acid sequences of TraJ proteins were aligned by using the SEQHP program according to the algorithm described by Goad and Kanehisa (18) on the VAX/VMS at the Institute of Medical Science, University of Tokyo.

RESULTS

Nucleotide sequence of the promoter-proximal region of the *tra* region including *traJ* and *traY*. Fig. 1A represents a map of R100 showing the location of *Eco*RI restriction fragments. It has been shown that the promoter-proximal *tra* region was present within the *Eco*RI fragment *r*4 of R100 (1, 7, 22). We cloned the *r*4 fragment into plasmid pUC19 and obtained

structures of the plasmids were checked by sequencing. (B) The strategy used to determine the nucleotide sequence of the region carried by pSI87-ES3 is shown by arrows. The nucleotide sequences in the two regions containing *finP* and *traA* of R100 were identical to those of its derivative, R100-1, which have been previously reported (10, 14). McIntire and Dempsey (29) have reported the nucleotide sequence containing *oriT*, which shows the following differences from ours: A at position 7 in McIntire and Dempsey (29) was G in our sequence; T at position 8 was C; T at 12 was C; C at 377 was T; G at 510 was A. The nucleotide sequences of the *oriT*, *traM*, *finP*, and *traA-E* regions of R100 were highly homologous (more than 76%) with those of F (data not shown). (C) A schematic representation of the heteroduplexes of R100 with F and R1. These heteroduplexes were drawn on the basis of the electron microscopic data obtained by Sharp et al. (41).

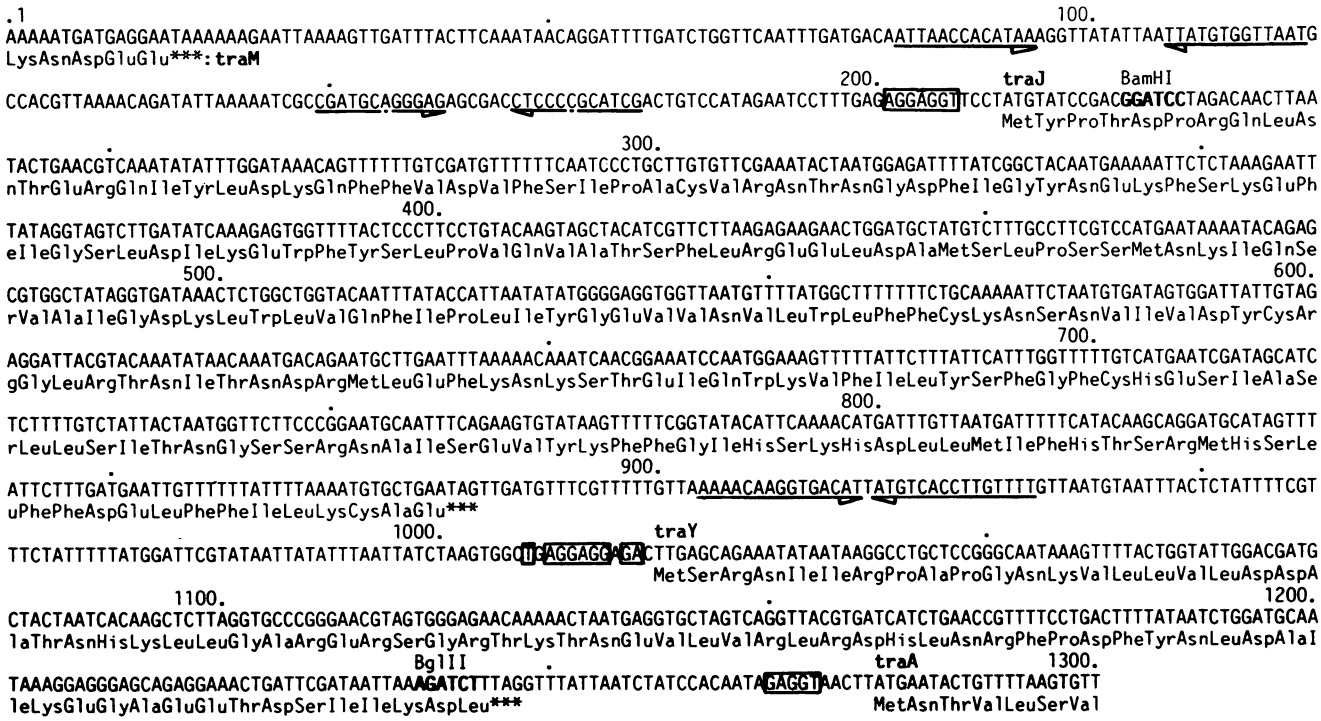

FIG. 2. Nucleotide sequence of the region of R100 containing *traJ* and *traY*. The two possible reading frames for the TraJ and TraY proteins are shown. They are flanked by the C-terminal region of the TraM protein and the N-terminal region of the TraA protein, both of which show a strong homology with those of F and R1 (R1-19), which have been reported previously (8, 14, 15, 25). Asterisks show the termination codons of *tra* genes. The nucleotide sequence of R100 shown may also contain *finP*, but it is not assigned here. Shine-Dalgarno sequences are boxed. Inverted repeats are shown by arrows. The *Bam*HI and *Bgl*II sites used to construct pSI200 (see Fig. 4) are shown. The nucleotide sequence from position 1 to 379 is identical with that previously reported (7). The nucleotide sequence from position 990 to 1289 is identical with that of R100-1, a derivative of R100 (9).

plasmids pSI28 and pSI87 (Fig. 1A). We then reconstructed a series of derivatives having deletions in various portions of the *r4* fragment, as depicted in Fig. 1A. These plasmids were very useful for the experiments described below.

We determined the nucleotide sequence of the 3,267-base-pair region carried by plasmid pSI87-ES3 by using the pSI plasmids obtained above. On the basis of the homology between the nucleotide sequence of R100 with that of F, which has been previously reported (12, 15, 43, 44), the region sequenced was found to contain *traM*, *finP*, *traA*, *traL*, part of *traE*, the origin of transfer (*oriT*), and possibly the *traJ* and *traY* genes, which are of interest in the present study. The regions are schematically represented in Fig. 1A. (See the legend to Fig. 1 for information on these nucleotide sequences, except for those of *traJ* and *traY*.)

Figure 2 shows the nucleotide sequence of the region presumably containing *traJ* and *traY* of R100. Portions of this sequence have been reported previously (7, 9; see the legend to Fig. 2). There are two open reading frames in this region. One is from nucleotide 212 to 880, covering 669 base pairs and encoding 223 amino acids (Fig. 2). This coding region must be the *traJ* gene of R100, since it is located immediately downstream of the *traM* coding sequence identified in this region (Fig. 2). The nucleotide sequence of this region showed poor homology (about 50%) with the *traJ* coding sequences of F (12) and R1 (8, 24). Presumably the *traJ* sequence is lying in a small nonhomologous region that was found within the long homologous region between R100 and F or R1 shown by electron microscope heteroduplex analysis (see the mismatched regions depicted in Fig. 1C). Downstream of *traJ* of R100, there is a 15-base-pair inverted repeat sequence which is GC rich and is flanked by a stretch of four adenines and four thymines (Fig. 2), a typical Rho-independent terminator for transcription starting from either direction (36).

The second open reading frame located downstream of *traJ* must be the *traY* gene of R100 and covers 225 base pairs from nucleotide 1020 to 1244, encoding 75 amino acids to yield a protein of molecular weight 8,542. *traY* of R100 is likely to start with the rare initiation codon TTG, which is preceded by a very good Shine-Dalgarno sequence (Fig. 2). The coding sequence is identical with that of R100-1, a derivative of R100, which has been reported previously (9). The promoter p_{YZ} for the *traYZ* operon of F and R1 has been mapped between the distal region of the *traJ* coding frame and the region preceding *traY* (12, 13, 24). The possible p_{YZ} of R100 did not show much homology with the p_{YZ} sequences reported for F or R1 (data not shown).

Identification of the TraJ protein. To identify *tra* proteins, several pSI plasmids, all of which carried a portion of the R100 sequence determined above, were introduced into *E. coli* CSR603 and the proteins produced were analyzed according to the maxicell procedure. Maxicells harboring pSI87-ES3, which carries the R100 sequence containing all the *tra* genes in the *r4* fragment (Fig. 1A), produced several protein bands in a polyacrylamide gel (Fig. 3). Maxicells harboring another pSI plasmid with a deletion in the R100 sequence produced fewer bands. No protein bands were produced in maxicells harboring plasmid pUC19 or no plasmid, except for the β-lactamase band encoded by

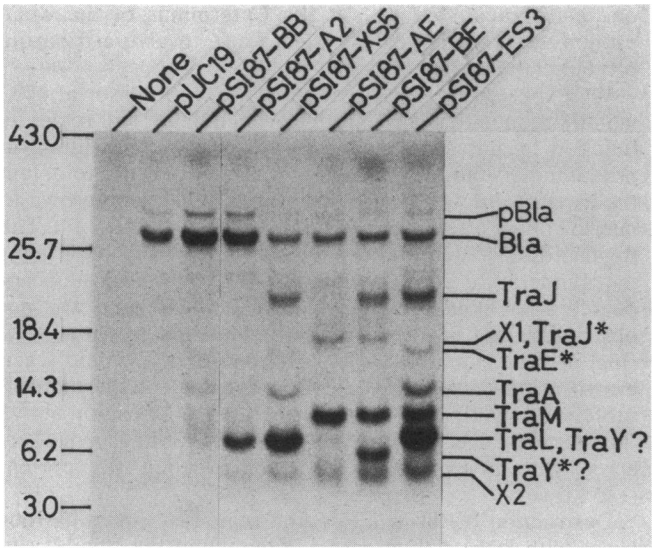

FIG. 3. Autoradiogram of an SDS-11% polyacrylamide gel showing the *tra* gene products encoded by pSI plasmids. On the top are indicated the plasmids used in each lane. Samples were labeled with [^{3}H]lysine according to the maxicell procedure described in Materials and Methods. Identified protein bands are shown on the right side of the gel. Possible truncated proteins from the genes of *tra* fused with the pUC sequence are marked with an asterisk. Sizes of markers are indicated in kilodaltons on the left side of the gel. pBla and Bla denote pre-β-lactamase and mature β-lactamase, respectively, produced from the ampicillin resistance gene *bla* carried by pUC19. X1 indicates the protein produced by a gene which may exist upstream from the *oriT* sequence. X2 might represent degradation products of some of the proteins.

pUC19. Considering the sequences present in pSI plasmids and the molecular weights of *tra* proteins expected from the presumptive coding regions and from the presumptive *tra* genes fused with the pUC19 sequence, most of the bands could be assigned, as shown in Fig. 3. The TraJ protein, which was of interest to us, could be identified, because maxicells harboring a plasmid carrying the *traJ* sequence, for example, pSI87-BE (Fig. 1), produced the TraJ protein band, while the plasmid carrying no or an incomplete *traJ* coding region, for example, pSI87-AE (Fig. 1), did not (Fig. 3).

Identification of the TraY protein, which is also of interest to us, was not possible, however. Plasmid pSI87-A2 (or pSI87-XS5) carrying *traY* of R100 seemed to produce a band of a protein which comigrated with the TraL protein (10,527 daltons), significantly larger than the expected size of the TraY protein (8,542 daltons), although plasmid pSI87-BB, deleted for most of *traY* (Fig. 1), did not seem to produce such a protein (Fig. 3). Also, a possible truncated protein encoded by the fusion gene between *traY* and the pUC sequence was identified in the maxicells carrying pSI87-BE (Fig. 3), although the size observed (5.8 kilodaltons [kDa]) was significantly smaller than the expected one (9611 daltons).

Identification of the TraY protein. To determine that *traY* of R100 produces a protein which is translated by using TTG as the initiation codon, we constructed a fusion gene between *traY* and *lacZ* encoding β-galactosidase (LacZ) and analyzed the product of the gene. Figure 4 depicts the strategy for construction of the plasmid carrying the fusion gene. We first isolated the 1,015-base-pair *Bam*HI-*Bgl*II fragment containing a *traJ-traY* region of pSI87-B19 in which *traJ* is deleted for its N-terminal region and *traY* is deleted for the C-terminal region (Fig. 4; see also Fig. 2). This fragment was cloned into the *Bam*HI site of vector plasmid pJG200, which carries a gene encoding a fusion protein

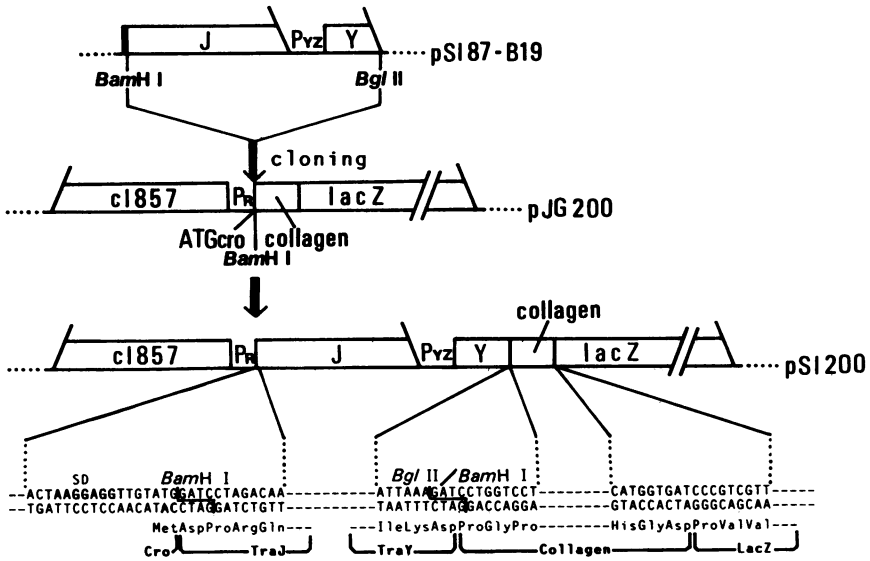

FIG. 4. Construction of pSI200. Only portions of the circular plasmids important for the construction of pSI200 are shown. A detailed description of the structures of the plasmids and of the procedure for construction of pSI200 is given in Results. Genes such as *traJ* (J), *traY* (Y), *cI857*, the collagen gene, and *lacZ* are shown by thick open arrows to indicate the orientations of their reading frames. The initiation triplet of the *cro* gene is also shown. The promoter for *cro* (p_R) and the promoter for the *traYZ* operon (p_{YZ}) are shown. The nucleotide sequences and the amino acid sequences of the three junctions (between *cro* and *traJ*, between *traY* and the collagen gene, and between the collagen gene and *lacZ*) are shown below pSI200. Shine-Dalgarno sequences are shown by boldface letters.

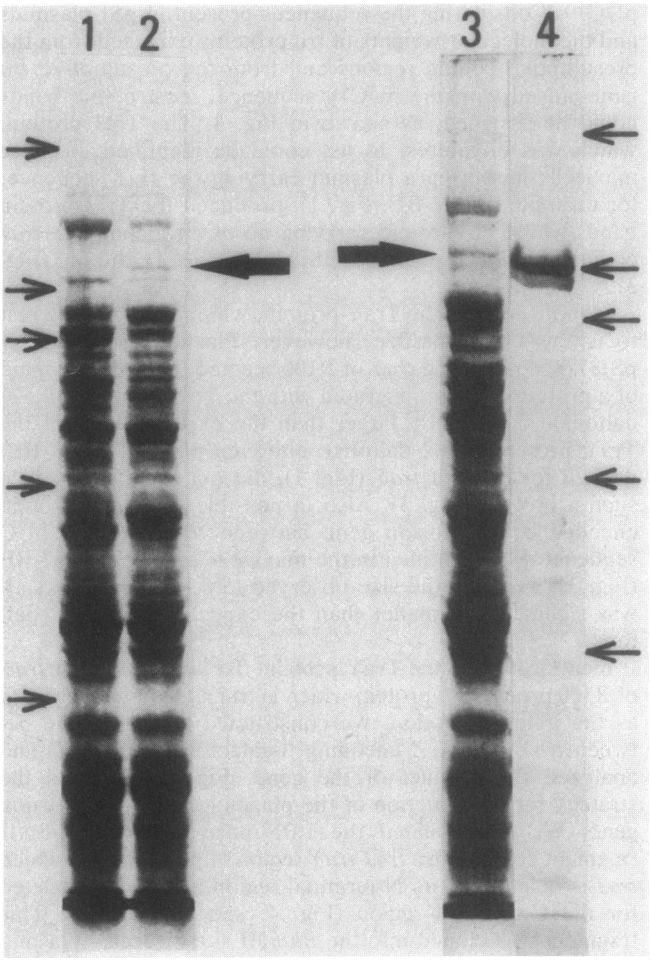

FIG. 5. An SDS-8% polyacrylamide gel, stained with Coomassie blue, showing the tripartite fusion protein. Lanes: 1, total proteins from MC1000, harboring plasmid pSI200, before heat induction; 2, total proteins from MC1000, harboring pSI200, after heat induction; 3, proteins in the supernatant fraction obtained by sedimentation at 70,000 × *g* from total proteins prepared after heat induction; 4, proteins retained in a TPEG-Sepharose column from samples used in lane 3. The position of the band of the tripartite protein is indicated by thick arrows. The molecular mass standards used were myosin (200 kDa), β-galactosidase (116.25 kDa), phosphorylase *b* (92.5 kDa), bovine serum albumin (66.2 kDa), and ovalbumin (45 kDa), whose electrophoretic band positions are shown by thin arrows in the same order from top to bottom.

consisting of three segments, i.e., Cro of bacteriophage lambda, collagen, and β-galactosidase, and which has a *Bam*HI site between the initiation codon of *cro* and the coding frame for collagen-LacZ (17) (Fig. 4). Note that in this vector plasmid, the promoter for the gene has been derived from the *cro* gene and transcription of the gene from the λ p_R promoter therefore is under control of the temperature-sensitive repressor, the product of *c*I857, which is also carried by pJG200 (17) (Fig. 4). The resulting recombinant plasmid, pSI200, contains two fusion genes; one consists of the ATG of *cro* and the *traJ* coding region in frame, and the other consists of *traY* and the collagen-*lacZ* coding region, also in frame (Fig. 4). In the former fusion gene, the amino acid sequence Met-Tyr-Pro-Thr of the N terminus of the TraJ protein was replaced by one amino acid, methionine, to give an altered TraJ protein (called TraJ'), while in the latter, one amino acid, leucine, at the C terminus of the TraY protein was replaced by collagen-LacZ to give a tripartite protein, TraY-collagen-LacZ.

We expected that inactivation of the λ repressor at 42°C would produce the TraJ' protein, which might still retain its function in activating transcription from p_{YZ}, resulting in production of the tripartite protein. Thus, in order to obtain the tripartite protein, cells harboring pSI200 were grown to mid-log phase at 30°C and then incubated at 42°C for 1 h. The β-galactosidase activity increased from 54 to 5,405 U after the temperature shift. A new band of a large protein appeared after induction (Fig. 5, lane 2). This band was not observed in the host harboring pSI200 when grown at 30°C (Fig. 5, lane 1) and in the host harboring no plasmid when grown at 42°C (data not shown). The size of the induced protein (about 123 kDa), which migrates at a position above the intact β-galactosidase (about 116 kDa) used as a marker, agrees with the expected size of the tripartite fusion protein (130 kDa).

Destruction of the *traJ'* coding region by inserting four bases with DNA polymerase I (Klenow) at the *Bam*HI site located at the N-terminal region of *traJ'* in pSI200 resulted in no overproduction of the tripartite protein, because the β-galactosidase activity was increased to only 128 from 13 U after induction of the product of the mutagenized *traJ'* gene. This suggests that transcription from the promoter for *traJ'* does not read through into *traY*, probably because of the presence of the terminatorlike sequence located immediately downstream of *traJ*.

The tripartite protein could be purified to homogeneity by selective binding to and elution from a β-galactosidase specific-affinity column, as shown in Fig. 5, lane 4. Ten amino acids in the N-terminal region of the purified protein were determined. The 1st amino acid was not detected clearly, but the 2nd to 10th amino acid residues were clearly identified and corresponded to the 3rd to 11th residues deduced from the nucleotide sequence of *traY*, although some signals for other amino acids, possibly due to the presence of contaminating polypeptides, were detected in each degradation cycle during amino acid sequencing. The second amino acid, which is serine in the deduced amino acid sequence of the TraY protein, was not clearly detected as the first amino acid in the purified tripartite protein, since detection of the N-terminal amino acid is often difficult and serine, particularly, gives a weak signal in amino acid sequencing. The results described above indicate that the putative *traY* coding region of R100 actually produces a protein which starts with formylmethioninez, corresponding to the TTG triplet, but subsequently loses it, perhaps during maturation of the protein.

Amino acid sequence relations among the TraJ and TraY proteins of different plasmids. As described in an earlier section, the region containing the possible *traJ* gene of R100 showed poor homology with the *traJ* coding region of plasmids F and R1 at the nucleotide sequence level. The deduced amino acid sequence of the TraJ protein of R100, however, showed low but significant homology with the TraJ proteins of F and R1 (Fig. 6A).

The nucleotide sequences of the presumptive *traY* coding regions of R1-19 (8) and of ColB4-K98 (9) have been reported elsewhere. Proteins from these regions, however, have not been identified yet. The deduced amino acid sequences of these TraY proteins show considerable homology with those of *traY* of R100 and R100-1 (Fig. 6B). All of these TraY proteins are composed of 75 amino acids. An interesting observation is that just as *traY* of R100 uses a rare initiation

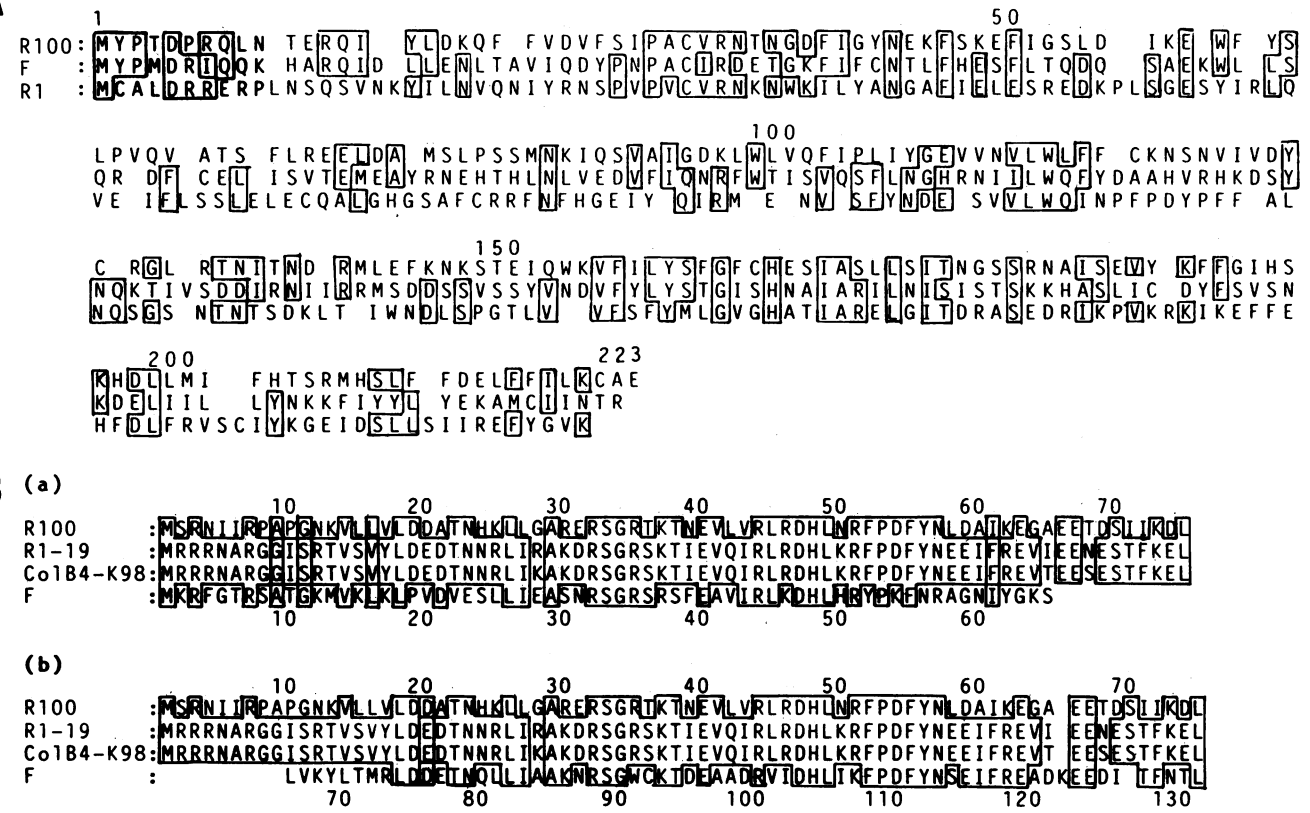

FIG. 6. Amino acid alignments. (A) *traJ* gene products of R100, F, and R1. Numbers above the sequences indicate the positions of amino acid residues of the TraJ protein of R100. Identical amino acids are boxed. The TraJ amino acid sequence of R100 and the TraJ sequences of F and R1, previously reported (12, 24), were aligned first by computer (with a program that compared pairs of sequences by using the algorithm described by Goad and Kanehisa [18]) and then manually, to give the best matches among three sequences. *traJ* of plasmid R1-19 has been reported to encode a 201-amino-acid polypeptide whose sequence is the same as the TraJ amino acid sequence of R1 from residues 1 to 200 (8). (B) *traY* gene products of R100, R1-19, ColB4-K98, and F. Identical amino acids are boxed. The first half (residues 1 to 65) (a) and the second half (residues 66 to 131) (b) of the TraY protein of F, previously reported (12), are compared with the TraY protein of R100 or R100-1 and those of R1-19 and ColB4-K98, previously reported (8, 9). Numbers above the R100 sequence indicate the positions of amino acid residues of the TraY proteins of R100, R1-19, and ColB4-K98, while numbers below the F sequence indicate the positions of amino acid residues of the TraY protein of F. Gaps were introduced to give the best matches.

codon, TTG, *traY* of R1-19 and *traY* of ColB4-K98 use another rare initiation codon, GTG.

The nucleotide sequence containing *traY* of F has also been reported (12). The coding frame of *traY* of F deduced from the nucleotide sequence was postulated to start from ATG as an initiation codon and to encode a protein of 119 amino acids, which is 44 amino acids larger than the products from *traY* of R100 (R100-1), R1-19, and ColB4-K98. We noticed, however, that the open reading frame of *traY* of F could be extended into the preceding region, with the assumption that TTG was used as the initiation codon for *traY* of F, as for *traY* of R100. This TTG in F, like that in the *traY* gene of R100, is preceded by a very good Shine-Dalgarno sequence. With the assumption that TTG is the initiation codon for *traY* of F, the molecular weight of the postulated protein is 15,183 (131 amino acids). Interestingly, a dot matrix analysis indicated that the nucleotide sequence of *traY* of R100 appeared twice in that of the TTG-initiated *traY* of F in a tandem array (Fig. 7A). This is also true for a dot matrix analysis of amino acid sequences (Fig. 7B). The dot matrix comparison of the coding region of *traY* of F to itself also showed the duplication of a segment within the coding region in tandem (data not shown). Figure 6B also shows the homology between each amino acid sequence of the two segments in the TraY protein of F and the TraY proteins of the other plasmids.

DISCUSSION

We have presented the nucleotide sequence of *traJ* of R100 and identified its product. We have then shown that induction of an altered TraJ protein, called TraJ′, that lacks four amino acids at the N terminus of the TraJ protein results in overproduction of the TraY-collagen-LacZ tripartite protein. Transcription from the promoter for *traJ′* does not read through into *traY*. The result described above therefore suggests that the four amino acids at the N-terminal end of the TraJ protein are not essential for activation of the expression of the *traYZ* operon. The deduced amino acid sequence shows low but significant homology with those of the TraJ proteins of F and R1. This low homology is in agreement with the observation that *traJ* of R100 does not complement the *traJ90* mutation of F (46), although both TraJ proteins are thought to have the same function. We have also observed that *traJ* of R100 has been subjected to a drastic change coordinately with a drastic change in p_{YZ}, which is the site of action of the TraJ protein, when these regions are compared with the corresponding regions in F

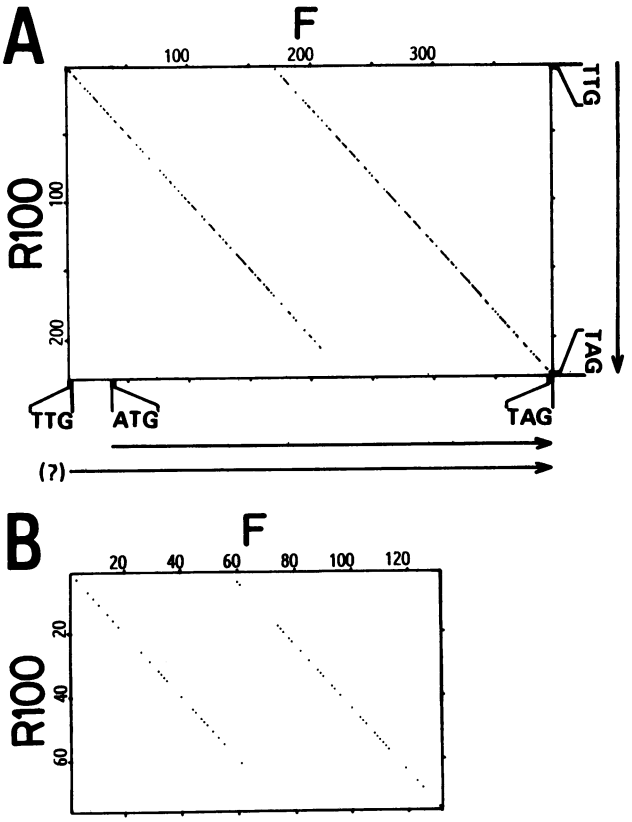

FIG. 7. Dot matrix analysis. (A) Matrix between the nucleotide sequences of the *traY* genes of F and R100. Dots are placed at locations of the identical nucleotides when more than 33 nucleotides of 70 are identical. Arrows indicate the direction of the coding regions. The initiation and termination codons are shown. (B) Matrix between the amino acid sequences of the TraY proteins of F and R100. Dots are placed at locations of identical amino acids when more than 11 amino acids out of 60 are identical.

and R1. In studies on the *repA2* genes of R100 and R1, Ryder et al. (38) have observed similar drastic changes in two loci in which a product of one locus acts on the other locus. There is a Rho-independent terminatorlike sequence downstream of the *traJ* coding frame of R100 which could play an important role in inhibiting the readthrough transcription from the *traJ* promoter into the distally placed promoter p_{YZ}. Such a terminatorlike sequence is also seen in the region downstream of *traJ* in F and R1-19 (8, 12, 24).

We have shown that *traY* of R100 uses TTG as the initiation codon by determining the N-terminal amino acid sequence of the tripartite protein TraY-collagen-LacZ. The method employed here has been very useful for purification of the tripartite protein. The method would also be very advantageous for separating the TraY portion from the β-galactosidase portion by cleavage of the collagen linker with collagenase. In fact, digestion of the hybrid protein with collagenase has given a new polypeptide whose molecular weight agrees well with that of the TraY portion (our unpublished data). This protein is found to be basic and can thus be readily purified to homogeneity by using a cation-exchange column (our unpublished data). This procedure for purification of the TraY protein and the properties of the purified protein will be described elsewhere.

We have pointed out the possibility that *traY* of F uses TTG as the initiation codon and not the previously assigned ATG. Kennedy et al. (23) have reported that *traY* of F produces a 16-kDa protein. Fowler et al. (12) have stated in their paper that the size is slightly larger than that expected from the open reading frame. Note that the *traY* reading frame of F which uses TTG as the initiation codon encodes a protein closer to the size of the protein that has been observed by Kennedy et al. (23). It is therefore quite likely that the discrepancy in size of the TraY protein of F is due to the wrong assignment of the initiation codon for the *traY* reading frame of F.

If *traY* of F uses the TTG triplet as the initiation codon, all four *traY* genes previously reported use rare initiation codons, TTG and GTG. This appears to be an interesting feature conserved in the evolutionary process. Reddy et al. (37) have demonstrated that the efficiency in initiation of translation increases as the initiation codon is changed from TTG to GTG and from GTG to ATG by using the *cya* gene which initially uses TTG as the initiation codon. The use of rare initiation codons in *traY* may therefore lower the cellular level of the TraY protein.

Interestingly, also, *traY* of F contains an apparent tandem duplication of a segment which shows homology with *traY* of R100. It has been suggested that the TraY protein recognizes the plasmid-specific *oriT* sequence (46). We assume that the TraY protein of R100 could function as a dimeric or tetrameric form and that the TraY protein of F by itself has a dimeric form within a single polypeptide. *traY* of F could have evolved by an illegitimate recombination between two identical *traY* genes encoding a small polypeptide composed of 75 amino acids, like the TraY proteins of the other plasmids, such that the recombination has occurred between the distal region of one copy of *traY* and a proximal region of the other copy to give rise to a larger new gene having a tandemly duplicated sequence.

ACKNOWLEDGMENTS

We thank Deepak Bastia for the generous gift of pJG200. We also thank Hiroko Nashimoto for useful advice on the purification of the tripartite fusion protein, Hiromichi Nagasawa for help with the amino acid sequencing, and Yoshimi Maeda for help with dot matrix analyses. We thank Werner Maas for critical reading of the manuscript and for correcting the English.

This research was supported by a grant-in-aid for scientific research from the Ministry of Education, Science, and Culture of Japan.

LITERATURE CITED

1. **Achtman, M., B. Kušeček, and K. N. Timmis.** 1978. *tra* cistrons and proteins encoded by the *Escherichia coli* antibiotic resistance plasmid R6-5. Mol. Gen. Genet. **163**:169–179.
2. **Achtman, M., R. A. Skurray, R. Thompson, R. Helmuth, S. Hall, L. Beutin, and A. J. Clark.** 1978. Assignment of *tra* cistrons to *Eco*RI fragments of F sex factor DNA. J. Bacteriol. **133**:1383–1392.
3. **Casadaban, M. J., and S. N. Cohen.** 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J. Mol. Biol. **138**:179–207.
4. **Datta, N.** 1974. Epidemiology and classification of plasmids, p. 9–15. *In* D. Schlessinger (ed.), Microbiology—1974. American Society for Microbiology, Washington, D.C.
5. **Everett, R., and N. Willetts.** 1980. Characterisation 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.
6. **Fairbanks, G., T. L. Steck, and D. F. H. Wallach.** 1971. Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane. Biochemistry **10**:2606–2617.
7. **Fee, B. E., and W. B. Dempsey.** 1986. Cloning, mapping, and sequencing of plasmid R100 *traM* and *finP* genes. J. Bacteriol.

167:336–345.
8. **Finlay, B. B., L. S. Frost, and W. Paranchych.** 1986. Nucleotide sequences of the R1-19 plasmid transfer genes *traM*, *finP*, *traJ*, and *traY* and the *traYZ* promoter. J. Bacteriol. **166**:368–374.
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. **Finlay, B. B., L. S. Frost, W. Paranchych, and N. S. Willetts.** 1986. Nucleotide sequences of five IncF plasmid *finP* alleles. J. Bacteriol. **167**:754–757.
11. **Finnegan, D., and N. Willetts.** 1973. The site of action of the F transfer inhibitor. Mol. Gen. Genet. **127**:307–316.
12. **Fowler, T., L. Taylor, and R. Thompson.** 1983. The control region of the F plasmid transfer operon: DNA sequence of the *traJ* and *traY* genes and characterisation of the *traY→Z* promoter. Gene **26**:79–89.
13. **Fowler, T., and R. Thompson.** 1986. Shadow promoters in the F plasmid transfer operon. Mol. Gen. Genet. **202**:509–511.
14. **Frost, L. S., B. B. Finlay, A. Opgenorth, W. Paranchych, and J. S. Lee.** 1985. Characterization and sequence analysis of pilin from F-like plasmids. J. Bacteriol. **164**:1238–1247.
15. **Frost, L. S., W. Paranchych, and N. S. Willetts.** 1984. DNA sequence of the F *traALE* region that includes the gene for F pilin. J. Bacteriol. **160**:395–401.
16. **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.
17. **Germino, J., and D. Bastia.** 1984. Rapid purification of a cloned gene product by genetic fusion and site-specific proteolysis. Proc. Natl. Acad. Sci. USA **81**:4692–4696.
18. **Goad, W. B., and M. Kanehisa.** 1982. Pattern recognition in nucleic acid sequences. I. A general method for finding local homologies and symmetries. Nucleic Acids Res. **10**:247–263.
19. **Helmuth, R., and M. Achtman.** 1975. Operon structure of DNA transfer cistrons on the F sex factor. Nature (London) **257**: 652–656.
20. **Hussain, M., S. Ichihara, and S. Mizushima.** 1980. Accumulation of glyceride-containing precursor of the outer membrane lipoprotein in the cytoplasmic membrane of *Escherichia coli* treated with globomycin. J. Biol. Chem. **255**:3707–3712.
21. **Ippen-Ihler, K. A., and E. G. Minkley, Jr.** 1986. The conjugation system of F, the fertility factor of *Escherichia coli*. Annu. Rev. Genet. **20**:593–624.
22. **Kehoe, M. A., and T. J. Foster.** 1979. Transposon 10 promoted deletions and inversions in the transfer genes of R100-1. Mol. Gen. Genet. **176**:113–120.
23. **Kennedy, N., L. Beutin, M. Achtman, R. Skurray, U. Rahmsdorf, and P. Herrlich.** 1977. Conjugation proteins encoded by the F sex factor. Nature (London) **270**:580–585.
24. **Koronakis, V., and G. Högenauer.** 1986. The sequences of the *traJ* gene and the 5′ end of the *traY* gene of the resistance plasmid R1. Mol. Gen. Genet. **203**:137–142.
25. **Koronakis, V. E., E. Bauer, and G. Högenauer.** 1985. The *traM* gene of the resistance plasmid R1: comparison with the corresponding sequence of the *Escherichia coli* F factor. Gene **36**:79–86.
26. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
27. **Machida, Y., C. Machida, and E. Ohtsubo.** 1982. A novel type of transposon generated by insertion element IS*102* present in a pSC101 derivative. Cell **30**:29–36.
28. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 90–91. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **McIntire, S. A., and W. B. Dempsey.** 1987. *oriT* sequence of the antibiotic resistance plasmid R100. J. Bacteriol. **169**:3829–3832.
30. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
31. **Miller, J. H.** 1972. Experiments in molecular genetics, p. 433, 352–355. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. **Mullineaux, P., and N. Willetts.** 1985. Promoters in the transfer region of plasmid F, p. 605–614. *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.
33. **Nakaya, R., A. Nakamura, and Y. Murata.** 1960. Resistance transfer agents in *Shigella*. Biochem. Biophys. Res. Commun. **3**:654–659.
34. **Ohtsubo, E., M. Rosenbloom, H. Schrempf, W. Goebel, and J. Rosen.** 1978. Site specific recombination involved in the generation of small plasmids. Mol. Gen. Genet. **159**:131–141.
35. **Ohtsubo, H., and E. Ohtsubo.** 1977. Repeated DNA sequences in plasmids, phages, and bacterial chromosomes, p. 49–63. *In* A. I. Bukhari, J. A. Shapiro, and S. L. Adhya (ed.), DNA insertion elements, plasmids, and episomes. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
36. **Platt, T.** 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. **55**:339–372.
37. **Reddy, P., A. Peterkofsky, and K. McKenney.** 1985. Translational efficiency of the *Escherichia coli* adenylate cyclase gene: mutating the UUG initiation codon to GUG or AUG results in increased gene expression. Proc. Natl. Acad. Sci. USA **82**:5656–5660.
38. **Ryder, T. B., D. B. Davison, J. I. Rosen, E. Ohtsubo, and H. Ohtsubo.** 1982. Analysis of plasmid genome evolution based on nucleotide-sequence comparison of two related plasmids of *Escherichia coli*. Gene **17**:299–310.
39. **Sancar, A., A. M. Hack, and W. D. Rupp.** 1979. Simple method for identification of plasmid-coded proteins. J. Bacteriol. **137**: 692–693.
40. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
41. **Sharp, P. A., S. N. Cohen, and N. Davidson.** 1973. Electron microscope heteroduplex studies of sequence relations among plasmids of *Escherichia coli*. II. Structure of drug resistance (R) factors and F factors. J. Mol. Biol. **75**:235–255.
42. **Tanaka, N., J. H. Cramer, and R. H. Rownd.** 1976. *Eco*RI restriction endonuclease map of the composite R plasmid NR1. J. Bacteriol. **127**:619–636.
43. **Thompson, R., and L. Taylor.** 1982. Promoter mapping and DNA sequencing of the F plasmid transfer genes *traM* and *traJ*. Mol. Gen. Genet. **188**:513–518.
44. **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.
45. **Ullmann, A.** 1984. One-step purification of hybrid proteins which have β-galactosidase activity. Gene **29**:27–31.
46. **Willetts, N., and J. Maule.** 1986. Specificities of IncF plasmid conjugation genes. Genet. Res. **47**:1–11.
47. **Willetts, N., and R. Skurray.** 1980. The conjugation system of F-like plasmids. Annu. Rev. Genet. **14**:41–76.
48. **Willetts, N., and B. Wilkins.** 1984. Processing of plasmid DNA during bacterial conjugation. Microbiol. Rev. **48**:24–41.
49. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
50. **Yoshioka, Y., H. Ohtsubo, and E. Ohtsubo.** 1987. Repressor gene *finO* in plasmids R100 and F: constitutive transfer of plasmid F is caused by insertion of IS*3* into F *finO*. J. Bacteriol. **169**:619–623.