JOURNAL OF BACTERIOLOGY, July 1990, p. 4127–4131
0021-9193/90/074127-05$02.00/0

Vol. 172, No. 7

Nucleotide Sequence of the *traI* (Helicase I) Gene from the Sex Factor F

HARVEY D. BRADSHAW, JR.,[1] BETH A. TRAXLER,[2]† EDWIN G. MINKLEY, JR.,[3]* EUGENE W. NESTER,[4] AND MILTON P. GORDON[1]

Departments of Biochemistry[1] and Microbiology[4] University of Washington, Seattle, Washington 98195, and Department of Biological Sciences,[2] and Biotechnology Center, Mellon Institute,[3] Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213

Received 9 March 1990/Accepted 30 April 1990

A 6.9-kilobase region of the *Escherichia coli* F plasmid containing the 3′ half of the *traD* gene and the entire *traI* gene (encodes the TraI protein, DNA helicase I, and TraI*, a polypeptide arising from an internal in-frame translational start in *traI*) has been sequenced. A previously unidentified open reading frame (tentatively *trbH*) lies between *traD* and *traI*.

The conjugal transfer of the *Escherichia coli* F plasmid is mediated by the gene products of the *tra* regulon (reviewed in references 9 and 21). TraI protein is required for conjugation of F, is now known to be a DNA helicase (2), and has been purified and partially characterized (1, 10). Recent genetic and biochemical studies have shown that TraI protein also plays a role in nicking at the origin of transfer (19), presumably supplying the TraZ function required for nicking at F *oriT*. Mapping experiments (18) have suggested that TraI* (which is probably the same as protein 2b [12], for which an independent map locus, *traZ*, was once postulated [5]), is the result of an internal in-frame translational start in *traI*. Consequently, in the revised genetic map of F there is a region between *traD* and *traI* sufficiently large to encode additional *tra* polypeptides (18). To facilitate a biochemical analysis of TraI and TraI* function and to investigate the possibility that an unrecognized *tra* gene lies between *traD* and *traI*, we have sequenced the relevant region from the F plasmid. The TraI amino acid composition and N-terminal amino acid sequence deduced from the nucleotide sequence were verified by amino acid composition and N-terminal sequence analyses performed on purified TraI protein.

The DNA to be sequenced was taken from pPD1 (18), which spans the *traDI* locus. Two *Eco*RV (all enzymes except Sequenase [United States Biochemical Corp., Cleveland, Ohio] were from Bethesda Research Laboratories, Inc., Gaithersburg, Md.) fragments (6.4 kilobases and 511 base pairs) were sequenced, and their junction was confirmed by sequencing the overlapping 1-kilobase *Sma*I fragment. The gel-purified fragments were electroeluted and passed over NACS columns (Bethesda Research Laboratories), self-ligated, sonicated to randomly shear the DNA (4), end repaired with T4 DNA polymerase, and cloned into *Sma*I-cleaved M13mp18. Sequencing was by the dideoxy method (17) by using either the Klenow fragment of DNA polymerase I or Sequenase with [α-^{35}S]dATP (Dupont, NEN Research Products, Boston, Mass.) and buffer gradient gels (3). We routinely used dITP in place of dGTP with Sequenase to eliminate band compression artifacts. Shotgun sequence data were assembled by using SEQMAN and SEQMANED software (DNASTAR, Madison, Wis.). Subsequent analysis was performed with GENEPRO (Riverside Scientific, Seattle, Wash.) or FASTP (11). TraI protein was purified as previously described (19), and its amino acid composition was determined on a Durrum D-500 automated amino acid analyzer following hydrolysis in 6 N HCl at 110°C in vacuo for 24 h. Primary sequence determination was made on a 96-kilodalton amino-terminal fragment of TraI protein purified from an *E. coli* strain harboring the TraI chain-terminating derivative pEM10 (19) by using a gas-phase sequencer (model 470A; Applied Biosystems, Inc., Foster City, Calif.) and the standard Edman degradation sequencing program of Hunkapiller et al. (8).

The nucleotide sequence of a 6,881-base-pair region from the F plasmid containing the 3′ half of *traD* (1 to 768), a previously unknown open reading frame (tentatively *trbH* [771 to 1487]), and *traI* (1487 to 6754), along with the predicted translation products, is shown in Fig. 1. The N-terminal amino acid sequence of purified TraI protein, MetMet???IleAlaGlnVal, confirms the initiation of translation derived from nucleotide sequence analysis. (The Ser in position 3 would not be expected to be recovered.) Further support for the validity of the DNA sequence comes from a comparison between the amino acid composition of TraI protein predicted by the nucleotide sequence and that determined experimentally (Table 1). The amino acid analysis is the result of a single 24-h hydrolysis, uncorrected for destruction (would increase Ser and Thr and, to a lesser extent, Tyr and Met) or incomplete hydrolysis (would increase Leu and Ile).

Inspection of the nucleotide sequence shows that the *traD* reading frame is terminated by consecutive opal codons, TGATGA, whose internal ATG (underlined) is the probable initiation codon for an open reading frame designated *trbH*. The *trbH* reading frame ends with two opal codons preceded by an A (ATGATGA), to form the start sequence (underlined) consistent with the two amino terminal residues of TraI protein (we have discounted the first methionine codon in the *traI* reading frame at 1466 to 1468 as a likely translation start because it does not agree with the experimentally determined N-terminal amino acid sequence and because it lacks a consensus ribosome-binding site [Fig. 1]). This arrangement of stop-start codons implies that *traD*, *trbH*, and *traI* are translationally coupled (7) and explains the polar

* Corresponding author.
† Present address: Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115.

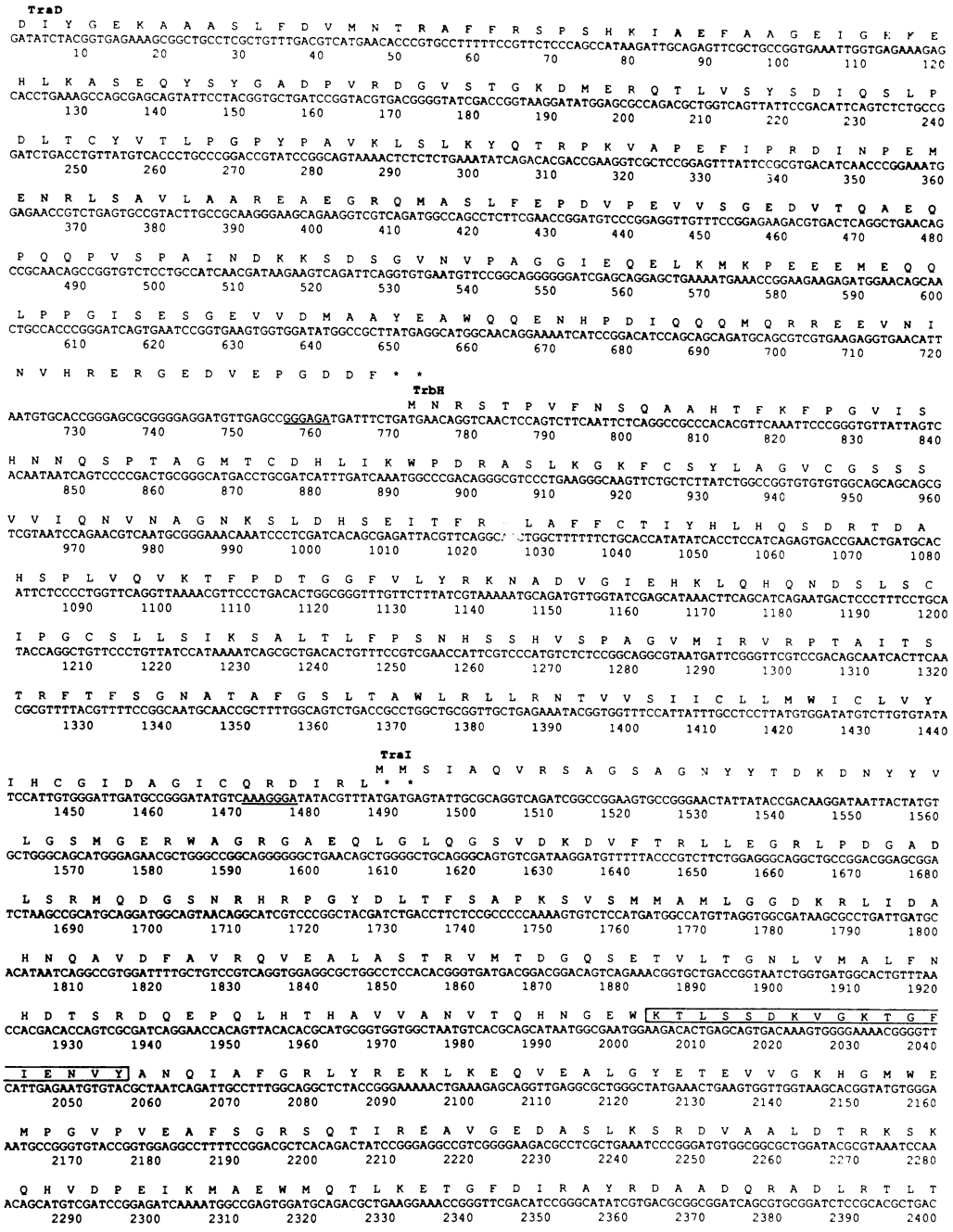

FIG. 1. Nucleotide sequence and predicted translation products of the 6.9-kilobase *Eco*RV fragment containing the *traDI* locus. The putative translation initiation sites of *trbH* (771), *traI* (1487), and *traI** (4349) are shown. Only the 3′ end of the *traD* gene is found on this *Eco*RV fragment. Probable ribosome-binding sites preceding *trbH*, *traI*, and *traI** are underlined. DNA helicase I is a potent DNA-dependent ATPase, and two regions of TraI with amino acid sequence similarity to adenine nucleotide-binding sites (20) are boxed.

effect that nonsense mutations in *traD* have on the production of TraI protein (5, 18).

Although expression of *traI* is mostly dependent upon positive activation of the *tra* operon by the TraJ protein, biochemical experiments have demonstrated that there is significant *traJ*-independent transcription of *traI* that occurs separately from *traD* (6, 16). Based on sequence similarity to the *E. coli* sigma-70 promoter consensus (15), a possible location for this in vivo promoter would be approximately 350 base pairs upstream of the TraI translation startpoint (−10 region at positions 1134 to 1139). Electron microscopy has been used to map a strong RNA polymerase-binding site in this region of *tra* DNA (13).

If the 88-kilodalton TraI* protein derives from an internal translation start in *traI*, as has been supposed (18), the most likely initiation codon is at position 4349 (Fig. 1). There is an acceptable ribosome-binding site nearby, and the deduced TraI* polypeptide has an M_r of 88,000. The predicted TraI helicase has an M_r of 192,000, reasonably close to the 180-kilodalton estimate from sodium dodecyl sulfate-poly-

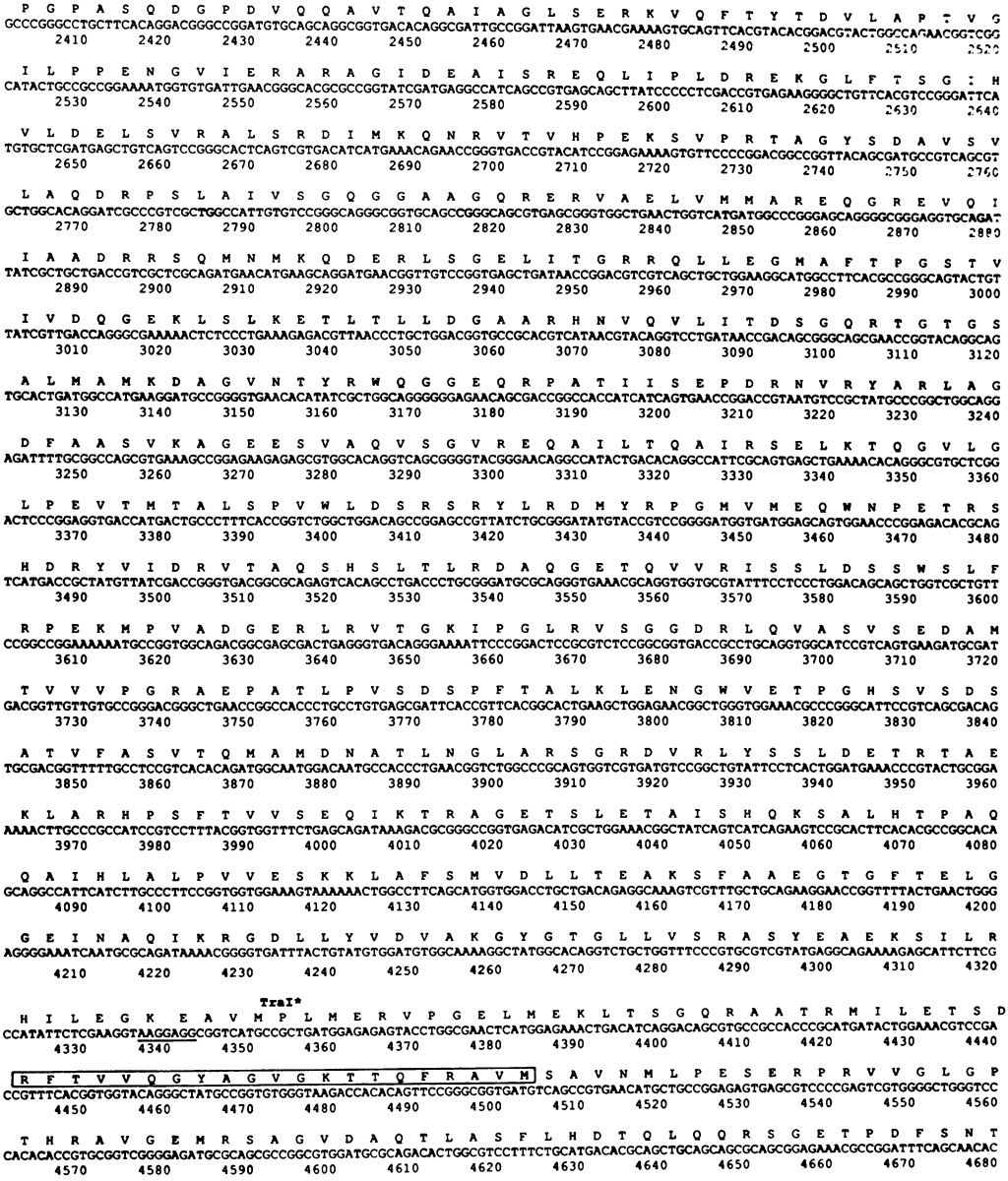

FIG. 1—*Continued.*

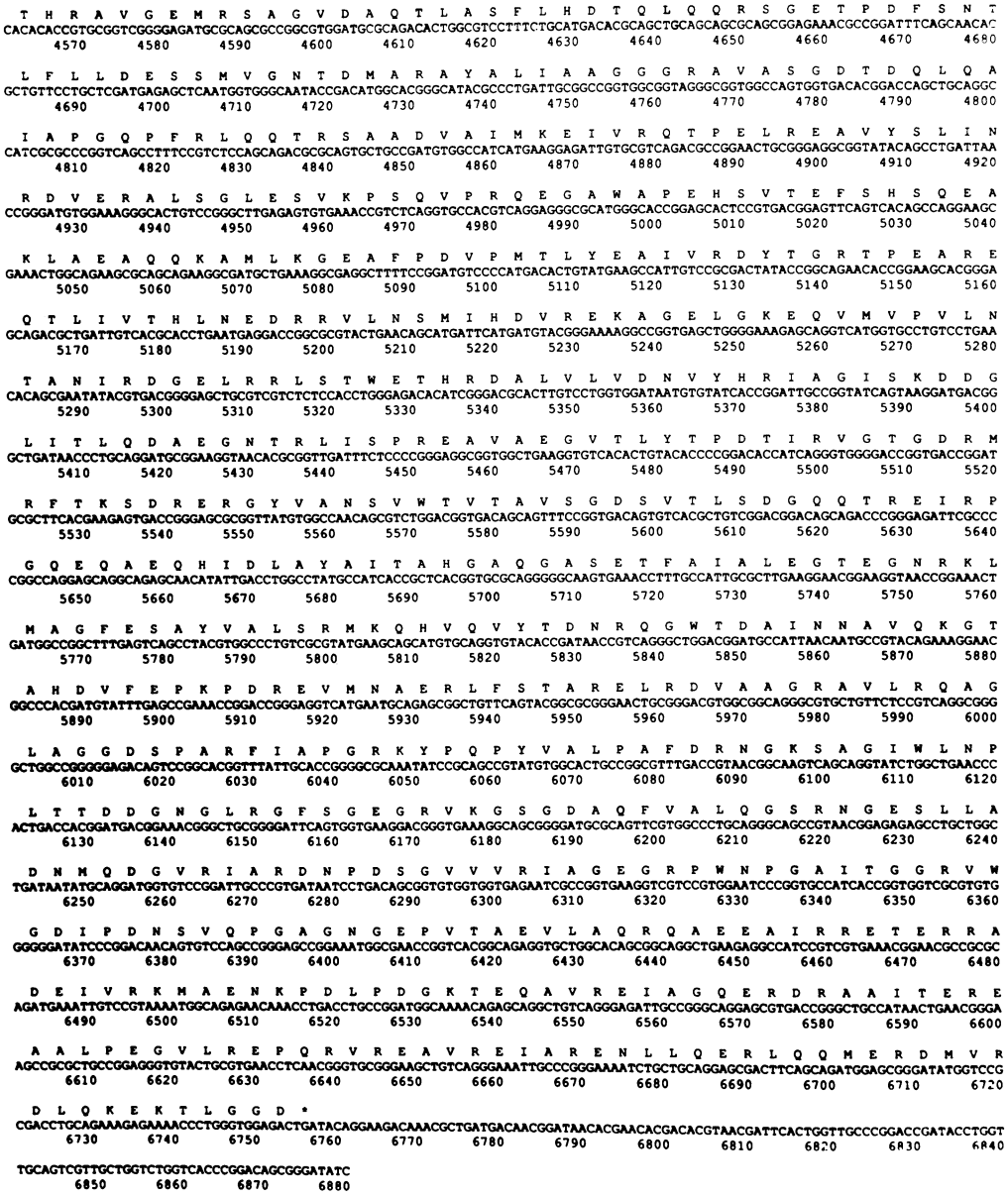

FIG. 1—*Continued.*

TABLE 1. Amino acid composition of TraI determined experimentally (see text) and deduced from the nucleotide sequence

Residue	mol% exptl	mol% deduced
Ala	10.5	10.3
Arg	8.8	8.6
Asx	8.8	8.7
Cys	ND[a]	0
Glx	12.7	13.1
Gly	8.7	8.6
His	2.0	1.7
Ile	3.4	3.7
Leu	8.1	8.1
Lys	3.6	3.5
Met	2.2	3.0
Phe	2.0	2.0
Pro	4.1	4.0
Ser	5.6	6.6
Thr	5.9	6.4
Trp	ND	0.9
Tyr	1.4	1.9
Val	7.9	8.2

[a] ND, Not determined.

acrylamide gel electrophoresis analysis of the authentic protein (1). The predicted TrbH polypeptide has an M_r of 26,000 and is therefore unlikely to be a 12-kilodalton membrane protein which has been observed and is thought to map in this region (12, 14).

Protein analyses were performed at the Protein Primary Structure Determination Facility at Carnegie-Mellon University, William E. Brown, Director.

This work was supported by a postdoctoral fellowship from the Helen Hay Whitney Foundation (H.D.B.), American Cancer Society grant NP336D (M.P.G.), and Public Health Service grants 2RO1 GM32618-17 (E.W.N.) and GM28925 (E.G.M.) from the National Institute of General Medical Sciences.

LITERATURE CITED

1. **Abdel-Monem, M., and H. Hoffman-Berling.** 1976. Enzymic unwinding of DNA. I. Purification and characterization of a DNA-dependent ATPase from *Escherichia coli*. Eur. J. Biochem. **65**:431–440.
2. **Abdel-Monem, M., G. Taucher-Scholz, and M.-Q. Klinkert.** 1983. Identification of *Escherichia coli* helicase I as the *traI* gene product of the F sex factor. Proc. Natl. Acad. Sci. USA **80**:4659–4663.
3. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
4. **Deininger, P. L.** 1983. Random cloning of sheared DNA: applications to shotgun sequence analysis. Anal. Biochem. **129**: 216–223.
5. **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.
6. **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.
7. **Gold, L., and G. Stormo.** 1987. Translational initiation, p. 1302–1307. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
8. **Hunkapiller, M. W., R. M. Hewick, W. J. Dreyer, and L. E. Hood.** 1983. High-sensitivity sequencing with a gas-phase sequenator. Methods Enzymol. **91**:399–413.
9. **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.
10. **Lahue, E. E., and S. W. Matson.** 1988. *Escherichia coli* DNA helicase I catalyzes a unidirectional and highly processive unwinding reaction. J. Biol. Chem. **263**:3208–3215.
11. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
12. **Manning, P. A., B. Kusecek, G. Morelli, C. Fisseau, and M. Achtman.** 1982. Analysis of the promoter-distal region of the *tra* operon of the sex factor of *Escherichia coli* K-12 encoded by *Eco*RI restriction fragments f17, f19, and f2. J. Bacteriol. **150**:76–88.
13. **Manning, P. A., G. Morelli, and C. Fisseau.** 1984. RNA-polymerase binding sites within the *tra* region of the F factor of *Escherichia coli* K-12. Gene **27**:121–123.
14. **Moore, D., B. A. Sowa, and K. Ippen-Ihler.** 1982. A new activity in the F *tra* operon which is required for F-pilin synthesis. Mol. Gen. Genet. **188**:459–464.
15. **Mulligan, M. E., and W. R. McClure.** 1986. Analysis of the occurrence of promoter-sites in DNA. Nucleic Acids Res. **14**:109–126.
16. **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.
17. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
18. **Traxler, B. A., and E. G. Minkley, Jr.** 1987. Revised genetic map of the distal end of the F transfer operon: implications for DNA helicase I, nicking at *oriT*, and conjugal DNA transport. J. Bacteriol. **169**:3251–3259.
19. **Traxler, B. A., and E. G. Minkley, Jr.** 1988. Evidence that DNA helicase I and *oriT* site-specific nicking are both functions of the F TraI protein. J. Mol. Biol. **204**:205–209.
20. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the *a*- and *B*- subunits of ATP synthase, myosin, kinases, and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1**:945–951.
21. **Willetts, N., and B. Wilkins.** 1984. Processing of plasmid DNA during bacterial conjugation. Microbiol. Rev. **48**:24–41.