Molecular and Genetic Analysis of a Region of Plasmid pCF10 Containing Positive Control Genes and Structural Genes Encoding Surface Proteins Involved in Pheromone-Inducible Conjugation in Enterococcus faecalis

SU-MEI KAO,† STEPHEN B. OLMSTED,1,2 ANN S. VIKSNINS,1,2 JEAN C. GALLO,† AND GARY M. DUNNY1,2*

Department of Microbiology, Immunology and Parasitology, New York State College of Veterinary Medicine, Cornell University, Ithaca, New York, 14853,1 and Institute for Advanced Studies in Biological Process Technology and Department of Microbiology, University of Minnesota, St. Paul, Minnesota 551082*

Received 8 May 1991/Accepted 2 October 1991

Exposure of Enterococcus faecalis cells carrying the tetracycline resistance plasmid pCF10 to the heptapeptide pheromone cCF10 results in an increase in conjugal transfer frequency by as much as 104-fold. Pheromone-induced donor cells also express at least two plasmid-encoded surface proteins, the 130-kDa Sec 10 protein, which is involved in surface exclusion, and the 150-kDa Asc10 protein, which has been associated with the formation of mating aggregates. Previous subcloning and transposon mutagenesis studies indicated that the adjacent EcoRI c (7.5 kb) and e (4.5 kb) fragments of pCF10 encode the structural genes for these proteins and that the EcoRI c fragment also encodes at least two regulatory genes involved in activation of the expression of the genes encoding Asc10 and Sec10. In this paper, the results of physical and genetic analysis of this region of pCF10, along with the complete DNA sequences of the EcoRI c and e fragments, are reported. The results of the genetic studies indicate the location of the structural genes for the surface proteins and reveal important features of their transcription. In addition, we provide evidence here and in the accompanying paper (S. B. Olmsted, S.-M. Kao, L. J. van Putte, J. C. Gallo, and G. M. Dunny, J. Bacteriol. 173:7665–7672, 1991) for a role of Asc10 in mating aggregate formation. The data also reveal a complex positive control system that acts at distances of at least 3 to 6 kb to activate expression of Asc10. DNA sequence analysis presented here reveals the positions of a number of specific genes, termed prg (pheromone-responsive genes) in this region of pCF10. The genes mapped include prgA (encoding Sec10) and prgB (encoding Asc10), as well as four putative regulatory genes, prgX, -R, -S, and -T. Although the predicted amino acid sequences of Sec10 and Asc10 have some structural features in common with a number of surface proteins of gram-positive cocci, and the Asc10 sequence is highly similar to that of a similar protein encoded by the pheromone-inducible plasmid pAD1 (D. Galli, F. Lottspeich, and R. Wirth, Mol. Microbiol. 4:895–904, 1990), the regulatory genes show relatively little resemblance to any previously sequenced genes from either procaryotes or eucaryotes.

Conjugal transfer of certain plasmids in Enterococcus faecalis is induced by small, hydrophobic peptide pheromones (18; and reviewed in references 13 and 17). Potential recipients can secrete at least five different pheromones to solicit DNA transfer from donor cells carrying the cognate plasmids (13, 19). In response to pheromone induction, the donor cells produce several surface protein antigens (21, 22, 27, 35, 42, 56, 60), including a surface adhesion called aggregation substance that facilitates the formation of mating aggregates (13, 17, 19, 27, 56). We have identified two major pheromone-inducible surface antigens that are associated with the 58-kb tetracycline resistance plasmid pCF10 (17). Similar-size proteins have been identified on cells carrying the hemolysin plasmid pAD1 (22, 27), and the various groups working in this area have recently agreed on a nomenclature system for these proteins (59). In this and the accompanying paper (45), we show that the 150-kDa Asc10 (formerly Tra50) protein, which can be detected only in extracts of pheromone-induced cells carrying pCF10, plays a role in the aggregation of donor and recipient cells. The 130-kDa Sec10 (formerly Tra130) protein is produced constitutively by cells carrying pCF10 and migrates on gels as a series of bands with apparent molecular weights of 1.2 × 105 to 1.3 × 105. Upon pheromone induction, the amount of this protein increases substantially, and there is a shift in the migration pattern such that the 130-kDa form of the protein is predominant (21, 41). Sec10 appears to have no direct role in the formation of mating aggregates. Results of mating experiments utilizing monoclonal antibodies have suggested that this protein mediates surface (entry) exclusion to prevent donor cells from self-mating (21). Recently, we described the cloning and expression of the genes encoding these proteins in Escherichia coli and E. faecalis (10). These two genes reside in two adjacent restriction fragments of pCF10, namely, the 7.5-kb EcoRI c fragment and the 4.5-kb EcoRI e fragment.

Previous analysis of these cloned fragments involved characterization of the proteins encoded by several Tn5-containing derivatives of these clones (10). These studies enabled us to determine the approximate location of the structural genes encoding Asc10 and Sec10 and also resulted in identification of two putative regulatory loci in the c
TABLE 1. Bacterial strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Species/relevant genetic markers*</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>OG1SSp</td>
<td>E. faecalis/st/rif/spc</td>
<td>18</td>
</tr>
<tr>
<td>OG1RF</td>
<td>E. faecalis/rif/fus</td>
<td>18</td>
</tr>
<tr>
<td>VJS470</td>
<td>E. coli/F^- r^- m^- proC^:+Tn5</td>
<td>V. Stewart (Cornell University)</td>
</tr>
<tr>
<td>JM101</td>
<td>E. coli supE thi D(lac-proAB) (F' traD36 proAB lacF'PZDM15)</td>
<td>61</td>
</tr>
<tr>
<td>JM109</td>
<td>E. coli recA1 gyrA96 thi hsdR17 supE44 relA1 D(lac-proAB) (F' traD36 proAB lacF'PZDM15)</td>
<td>61</td>
</tr>
<tr>
<td>DH5α</td>
<td>E. coli F^- F(lacZ∆M15 D(kZYA-argF)U169 recA1 hsdR17 rK^- mK^-)</td>
<td>Bethesda Research Laboratories</td>
</tr>
</tbody>
</table>

*Abbreviations for E. faecalis markers: str, streptomycin resistance; rif, rifampin resistance; cat, 3-fusidic acid resistance.

The R150 fragment (~1 kb) appeared to encode a positive control function required for expression of the 150-kDa Asc10 protein (but not Sec10) in E. faecalis. The R150 region spans a segment of pCF10 where insertion of Tn917 into any of three small, adjacent regions (tra 4 to 6) of the wild-type plasmid abolished pheromone-inducible transfer functions (9, 10). A second regulatory locus encoded by the Eco RI c fragment, R130, was proposed to activate expression of Sec10 in E. faecalis because chimeric plasmids containing Tn5 insertions in this region did not confer expression of the 130-kDa Sec10 protein when transferred to E. coli by protoplast transformation (10). To better characterize this important region of pCF10, we have carried out more extensive transposon mutagenesis and in vitro deletion analysis. In this communication, we also report the complete nucleotide sequences of these two fragments, which represent approximately half of the total pheromone-inducible transfer system of this plasmid. These data have allowed us to assign precise locations of a number of specific genes, denoted prg (pheromone-responsive genes), which are involved in pheromone-inducible conjugal transfer of pCF10.

MATERIALS AND METHODS

Bacterial strains, media, and reagents. The bacterial strains and plasmids used in this study are listed in Tables 1 and 2. E. coli was propagated in LB or M9 medium (40). E. faecalis was grown in BYGT or M9-YE medium (16). The concentrations of antibiotics in selective media were described previously (10). Restriction enzymes, exonuclease III, DNA polymerase Klenow fragment, nick translation kits, and DH5α competent cells were obtained from Bethesda Research Laboratories. T4 DNA ligase, deoxynucleoside triphosphates, and reverse primers were from Boehringer Mannheim. Lysozyme, antibiotics, and S1 nuclease were purchased from Sigma Chemical Co. T7 DNA polymerase and sequencing reagents were from United States Biochemical Corp. Oligonucleotide primers were synthesized at the Oligonucleotide Synthesis Facility of the Cornell University Biotechnology Program and were also synthesized in our laboratory, using an Applied Biosystems 391 automated synthesizer. [α-32P]dCTP (800 Ci/mmol) and [α-35S]dATP (500 Ci/mmol) were acquired from New England Nuclear Corp. Geneclean was from Bio101, Inc.

**Protein extraction and immunoblotting.** All of the immunoblotting studies presented in this report were carried out by using a polyclonal rabbit antiserum raised against pheromone-induced E. faecalis cells carrying pCF11, as described previously (10) and in the accompanying paper (45). Our

---

**TABLE 2. Plasmids**

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pWM402</td>
<td>tet cat (E. coli), cat (E. faecalis)</td>
<td>R. Wirth (University of Munich)</td>
</tr>
<tr>
<td>pDL276</td>
<td>lacZ Kan</td>
<td>20</td>
</tr>
<tr>
<td>pUC18</td>
<td>amp</td>
<td>61</td>
</tr>
<tr>
<td>pUC19</td>
<td>amp</td>
<td>61</td>
</tr>
<tr>
<td>pNY1801</td>
<td>EcoRI-c and -e of pCF10 cloned in the natural orientation into pWM402</td>
<td>10</td>
</tr>
<tr>
<td>pNY1825</td>
<td>EcoRI-c of pCF10 cloned into pWM402</td>
<td>10</td>
</tr>
<tr>
<td>pNY1842</td>
<td>EcoRI-c of pCF10 cloned in the opposite orientation from pNY1825</td>
<td>10</td>
</tr>
<tr>
<td>pNY4001-40</td>
<td>Deletion derivatives of pNY1842</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4501-63</td>
<td>pNY1801 with Tn5 insertions</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4100-74</td>
<td>pNY1825 with Tn5 insertions</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4200-45</td>
<td>pNY1842 with Tn5 insertions</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4401</td>
<td>EcoRI-c of pCF10 cloned into pDL276</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4404</td>
<td>EcoRI-c cloned in the opposite orientation</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4409</td>
<td>EcoRI-c of pCF10 cloned into pDL276</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4416</td>
<td>2.0-kb EcoRI-XbaI of EcoRI-c cloned into pDL276</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4421</td>
<td>5.5-kb XbaI-EcoRI of EcoRI-c cloned into pDL276</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4424</td>
<td>1.2-kb BglII-XbaI of EcoRI-c cloned into pUC18</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4425</td>
<td>1.2-kb BglII-XbaI of EcoRI-c cloned into pUC19</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4426</td>
<td>2.2-kb XbaI-HindIII of EcoRI-c cloned into pUC18</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4427</td>
<td>2.2-kb XbaI-HindIII of EcoRI-c cloned into pUC19</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4428</td>
<td>2.5-kb HindIII-PstI of EcoRI-c cloned into pUC19</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4429</td>
<td>0.7-kb PstI-EcoRI of EcoRI-c cloned into pUC18</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4430</td>
<td>3.6-kb EcoRI-PstI of EcoRI-c cloned into pUC18</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4432</td>
<td>0.88-kb BamHI-EcoRI of EcoRI-c cloned into pUC18</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4435</td>
<td>1.2-kb PstI-PstI fragment of EcoRI-c cloned into pUC18</td>
<td>This study</td>
</tr>
<tr>
<td>pNY4436</td>
<td>1.3-kb PstI-PstI fragment of EcoRI-c cloned into pUC18</td>
<td>This study</td>
</tr>
<tr>
<td>pMSP1001</td>
<td>1.8-kb PstI-PstI EcoRI c-e junction fragment cloned into pUC18</td>
<td>This study</td>
</tr>
</tbody>
</table>
previous work (10, 41) has established that the major proteins recognized by this antisem are Sec10 and Asc10. E. coli proteins were obtained from 3-m1 cultures grown overnight in M9-YE medium. Cells were harvested by centrifugation, washed once with phosphate-buffered saline (0.14 M NaCl, 1.5 mM Na2HPO4, 2.7 mM KCl), and resuspended in 100 ml of 10 mM Tris-HCl (pH 7.4). The cell suspensions were subjected to three 20-s bursts of sonication with a Virsonic model 16-850 cell dismuter. Cell debris was removed by centrifugation in a Fisher Scientific microfuge for 15 min at 4°C. The antigenic material in the sonic extracts was concentrated by immunoprecipitation. An appropriate dilution (as determined by enzyme-linked immunosorbent assays [21, 45]) of the antisem described above was incubated with the sonic extracts for 45 min at room temperature with gentle rocking. One thousand units of goat anti-rabbit immunoglobulin conjugated to horseradish peroxidase (Boehringer Mannheim) was added, and the mixture was incubated as described above for an additional period of 1 h. The immunoprecipitates were harvested by centrifugation in a microfuge for 1.5 h at 4°C. Pellets were resuspended in Laemml’s loading buffer (39) prior to electrophoresis. In interpreting immunoblotting results obtained with E. coli extracts, we assumed that the highest-molecular-weight antigen observed, migrating in the range of 130 to 150 kDa, represented Asc10 or truncated derivatives thereof, and that the largest antigen, migrating in the range of 85 to 130 kDa, represented Sec10 or a truncated derivative. This conclusion is based on the following observations: (i) E. coli strains carrying the cloned EcoRI c and e fragments produce an antigen with an apparent mass of 150 kDa (Asc10), a second antigen of 130 kDa (Sec10), and a ladder of less intense antigens of smaller size migrating below each of the two major antigens (see Fig. 2a, lane 1); (ii) E. coli carrying only the pWM402 vector produces no high-molecular-weight antigens reactive with this serum (10); (iii) the series of minor antigens disappears if the gene encoding the corresponding full-length antigen is inactivated (see below); and (iv) prolonged storage of antigenic preparations results in disappearance of the full-length proteins and an increase in concentration of the smaller proteins. Thus, it is likely that most if not all of the minor antigens result from proteolysis of Sec10 and Asc10. Extraction of surface antigens from E. faecalis cells and Western immunoblot analysis was carried out as described previously (10) and in the accompanying paper (45).

DNA manipulation and recombinant DNA techniques. The alkaline lysis procedure was used to isolate plasmids from E. faecalis (2). Small-scale plasmid isolation from E. coli was performed by the boiling method (33). Preparative-scale isolation of plasmid DNA from E. coli was carried out as described previously (40). Conditions for restriction digestions, ligations, Southern hybridizations, and nick translations have been described elsewhere (9, 10). Where restriction fragments were to be purified, samples were fractionated in Tris-acetate buffer, and the appropriate DNA bands were excised from the agarose gel and extracted with GeneCLean as recommended by the supplier. Nested deletions were performed according to Guo et al. (28). Dideoxychain termination DNA sequencing (48) was carried out by using T7 DNA polymerase (52) and [α-35S]dATP (4) and double-stranded plasmid as a source of template (7). In many cases, internal oligonucleotide primers were used to extend the sequence information from a given clone. The products of the sequencing reactions were resolved on a 6 or 8% polyacrylamide gel. Both strands of the DNA were sequenced, and each sequence of both strands was verified by analysis of at least two gels. The DNA sequence was assembled and analyzed by using the Pustell/IBI computer programs and Intelligenetics Suite/PC Gene programs at the University of Minnesota Molecular Biology Computing Center. Tn5 insertion mutagenesis of cloned pCF10 DNA has been described elsewhere (10), and 10 independent mutagenesis experiments were conducted in the present studies. Transformation of E. faecalis was done on protoplasts (58). E. coli cells were transformed according to Hanahan (31).

Nucleotide sequence accession number. The GenBank accession number is M64978 for the sequences reported here.

RESULTS

Genetic and physical mapping of the EcorI c and e fragments of pCF10. Previous studies indicated that the EcoR I c and e fragments of pCF10 contained structural and regulatory genes required for expression of the Sec10 and Asc10 proteins. To confirm and extend these findings, we generated a large number of insertion mutations of this region by using the transposon Tn5. Tn5 insertions were obtained in plasmids pINY1801 (c and e), pINY1825 (e), and pINY1842 (c; opposite orientation from pINY1825). In the case of insertion mutants derived from the latter two plasmids, we analyzed Sec10 expression, while both Sec10 and Asc10 expression was analyzed in pINY1801: Tn5 derivatives. We also derived a series of nested deletions in the EcorI c fragment in vitro as described in the Materials and Methods. All of these mutations were mapped physically by restriction endonuclease digestion (the positions of the mutations are indicated in Fig. 1). The mutant plasmids were originally characterized in an E. coli background, and a number of them were transformed subsequently into E. faecalis. (Immunoblotting analysis of the proteins produced by E. coli and E. faecalis strains carrying the mutant plasmids is illustrated in Fig. 2 and 3.) These studies resulted in identification of four distinct genetic loci, prgA, prgB, prgR, and prgX, which are described individually below.

The prgA locus. The boundaries and the orientation of the prgA gene encoding the Sec10 protein were determined by analysis of nested deletions of the EcoR I c fragment as well as by transposon mutagenesis. As indicated in the lower portion of Fig. 1, E. coli strains containing either pINY4003, 4014, 4027, or 4039 produced the mature 130-kDa Sec10, while those carrying pINY4019 or -4004 made 88- and 100-kDa truncated Sec10 derivatives, respectively. The size difference between the two truncated proteins predicts about a 300-bp difference in the sizes of the two deleted genes, which is comparable to the distance determined by mapping the DNA. Because the size of Sec10 decreased as deletions extended toward the left end of the c fragment, we infer that transcription of the prgA gene proceeds from left to right. The 5' end of the prgA gene should be located at 3.2 kb and the 3' end should be at 6.2 kb, as estimated from the truncated peptides. This was substantiated by analysis of pINY4039, which contained a deletion ending just to the left of the HindII site at 6.3 kb and encoded full-size Sec10. Western blotting analysis of the insertion mutants showed that all insertions between the positions of pINY4502 and pINY4549 affected the production of Sec10 (Fig. 1 and 2). In particular, pINY4532,-4253, -4261, -4503, and -4552 encoded 100-, 97-, 110-, 118-, and 126-kDa truncated Sec10 derivatives, respectively. Although the insertions in pINY4532 and -4253 were mapped at the same position, the Sec10 encoded by pINY4523 appeared smaller
on sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE). This could be due to the limit of resolution in restriction mapping or the aberrant migration of one of the peptides. Taken together, the sizes of the truncated Sec10 derivatives estimated from the protein analysis of the strains containing Tn5 insertions would place the 5’ end of prgA at about 3.0 kb and the 3’ end of the gene at about 6.2 kb, in agreement with the deletion analysis.

The prgB locus. All insertions of Tn5 between the positions of the insertions in pINY45S7 and -4505, except 4526, abolished synthesis of the 150-kDa Asc10 (Fig. 1 and 2), defining the location of the prgB gene encoding this protein. The 145-kDa truncated Asc10 made by insertion 4526 sug-

![Figure 1](http://jb.asm.org/Downloadedfromhttp://jb.asm.org/)  
**FIG. 1.** Physical map of the EcoRI c and e fragments based on insertional mutagenesis and in vitro deletion analysis. These fragments are located in the central portion of a 25- to 30-kb region of pCF10 encoding pheromone-inducible transfer functions; the region to the left of the e fragment encodes negative control functions, whereas the region to the right of the e fragment is believed to encode genes required for DNA transfer between attached cells, as reviewed in reference 17. Insertions of Tn5 into plasmids pNY1801 (numbers between 4501 and 4563), pNY1825 (numbers between 4101 and 4174), and pNY1842 (numbers between 4200 and 4270) were mapped by single and double digestions with EcoRI (E), BamHI (B), and BglII (Bg). Cleavage sites for other restriction enzymes, including HindIII (H) and PstI (P), are also indicated. As shown in the lower part of the figure, deletions extending from the right into the cloned EcoRI c fragment of pCF10 were generated in vitro, and the proteins encoded by the deletion derivatives in *E. coli* were extracted and analyzed by immunoblotting analysis (Materials and Methods). The sizes of the deletions were determined by analysis of fragments produced by digestion with the restriction enzymes EcoRI (E), HindIII (H), HpaI (Hp), HindII (Hc), and XbaI (X), with confirmation of the mapping by Southern hybridization (50). The size of the deleted EcoRI c fragment in each derivative is indicated by a horizontal line, and immunoblotting results are indicated as follows: +, production of full-length Sec10; −, no Sec10 produced; +*, truncated Sec10 produced. The putative position and orientation of transcription of the prgA gene encoding Sec10 is indicated by the arrow at the top. The designations of the prgX, -R, -A, and -B loci are based on immunoblotting analyses presented in Fig. 2 and 3 and described in the text.

gests that prgB starts at 6.4 kb and ends at 10.1 kb, an assumption supported by analysis of the insertions in pINY45S7 and -4505, which were mapped at 6.4 and 10.4 kb, respectively. The results also indicated that the prgA and prgB genes might form two distinct transcriptional units because insertion of Tn5 in prgA had no effect on prgB expression and vice versa (compare Fig. 2a and b). *E. faecalis*, but not *E. coli*, carrying pNY4515, -4503, or -4505 exhibited a clumpy phenotype in broth culture (data not shown) as well as production of Asc10, similar to the phenotype conferred by pNY1801 (containing the cloned EcoRI c and e fragments of wild-type pCF10) from which they were derived. All three Tn5 insertions in these plasmids were mapped outside of the prgB locus. In contrast, insertions within prgB simultaneously abolished Asc10 production and the clumpy phenotype, indicating the direct association of Asc10 with cellular aggregation.

The prgR locus. Analysis of certain Tn5 insertions into the region to the left of prgA indicated that these insertions abolished expression of prgB (but not prgA) in *E. faecalis* but not in *E. coli* (10). This region originally was designated R150, since it appeared to regulate expression of the 150-kDa Asc10 protein. Analysis of protein profiles (Fig. 3) of *E. faecalis* strains carrying additional derivatives of pNY1801 containing Tn5 insertions provide further evidence for the existence of a regulatory function encoded by this region. Current and previous analysis of insertional mutants identifies a locus between the positions of insertion 4509 on the left and insertions 4501 and 4513 on the right (Fig. 1), designated here prgR, as encoding an activation function for prgB expression in *E. faecalis*. Interestingly, the prgA gene, which is located between prgB and the prgR region, is not dependent on prgR for expression.

The prgX locus. Initial analyses of Tn5 insertions to the left of prgR indicated that a region (originally designated R130) might be required for expression of the 130-kDa Sec10 protein encoded by prgA (10). However, subsequent physical mapping (34) of plasmid DNA obtained from *E. faecalis* strains transformed with pNY1825 derivatives (pINY4141, -4148, and -4245; Fig. 1) containing Tn5 insertions in this region indicated that extensive rearrangement of pCF10 DNA occurred in all cases upon transfer to *E. faecalis*, despite numerous attempts to introduce these plasmids into an enterococcal host. (Interestingly, this is also a region of pCF10 where no Tn917 inserts were obtained [9].) We have designated this locus, whose interruption by mutation seems to be lethal in *E. faecalis*, prgX. Recent subcloning studies (11) suggest that prgX (like the prgR region) is probably required for expression of prgB, as well as for prgA expression (which is independent of the prgR region), in an *E. faecalis* background. It should be noted that none of the Tn5 insertions in either the prgX or prgR region have any observable phenotype in an *E. coli* host, indicating that the *E. coli* transcription/translation apparatus can express prgA and prgB without these positive regulatory functions (possibly by recognizing different promoter sequences).

DNA sequence analysis. Once the genes encoding Sec10 and Asc10 were mapped, as described above, we generated a set of subclones spanning both of these genes and the remainder of Eco RI c and e fragments in pUC18 and -19 (data not shown). We were later able to clone larger DNA segments, including the intact c and e fragments, into the recently developed shuttle vector pDL276 (20). This vector features a pUC-derived polylinker region flanked by transcription terminators, which appear to stabilize high-copy-number plasmids containing cloned DNA from gram-posi-
FIG. 2. Effects of Tn5 insertions in the prgA and prgB loci on expression of Sec10 and Asc10 in E. coli. Immunoblotting analysis of protein extracts from E. coli carrying the plasmids indicated below were analyzed as described in Materials and Methods. The positions of the Tn5 insertions are shown in Fig. 1. The two arrows at the right indicate the positions of migration of the wild-type Sec10 (bottom arrow) and Asc10 (top arrow) proteins. (a) Analysis focusing on Sec10 expression. Lanes: 1, pINY1801; 2, size standards; 3, pINY4501; 4, pINY4169; 5, pINY4513; 6, pINY4502; 7, pINY4162; 8, pINY4524; 9, pINY4548; 10, pINY4516; 11, pINY4252; 12, pINY4270; 13, pINY4515; 14 and 15, pINY4536; 16, pINY4545; 17, pINY4540; 18, pINY4532; 19, pINY4523; 20, pINY4261; 21, pINY4503; 22, pINY4552; 23, pINY4549. Sizes on the left are indicated in kilodaltons. (b) Analysis focusing on Asc10 expression. Lanes: 1, pINY1801; 2, pINY4557; 3, pINY4501; 4, pINY4563; 5, pINY4517; 6, pINY4547; 7, pINY4559; 8, pINY4561; 9, pINY4553; 10, pINY4530; 11, pINY4535; 12, pINY4519; 13, pINY4506; 14, pINY4526; 15, pINY4505; 16, pINY4544; 17, pINY4504; 18, pINY4541.

tive cocci in an E. coli host (8). Chimeric plasmids derived both from pDL276 and from pUC18 and -19 were then used in sequencing reactions, which allowed for the determination of the entire sequence of both strands of both fragments.

The sequence of 11,958-bp c and e fragments (Fig. 4) exhibited an overall G+C ratio of 35%, consistent with the moles percent G+C content of E. faecalis (23). However, some regions, e.g., open reading frame 2 (ORF2) (prgX), showed only 26% G+C. Sixteen potential ORFs were identified, with ORF2, -6, -8, -10, -12, -13, and -16 encoded by the complementary strand. Their locations are shown in Fig. 4. ORF2 coincided with prgX, while ORF4 and ORF5 corresponded to the prgR locus (Fig. 1). Therefore, the prgR locus was divided into two genes, designated as prgR and prgS, respectively. ORF9 corresponded to prgA, and ORF11 corresponded to prgB.

Figure 4 also shows the translated sequence of several ORFs from the top strand plus the translated prgX sequence, transcribed from the complementary strand. ORF1 began at the first base of the fragment and probably represents a gene whose transcription initiates in the adjacent EcoRI b fragment. Two ATG triplets in the same reading frame, at 3633 and 3681, could start ORF9. ORF8 began within ORF9 and ended just before the first potential start codon of ORF9. ORF1 and ORF2 converged at the ends, while ORF6 and ORF7 could encode divergent, overlapping genes. The two
that recent mRNA analysis (11) indicates that transcription of these sequences is pheromone inducible.

**Sequences with potential secondary structure.** Several inverted repeat sequences (IR) were identified (Fig. 4). IR1, which includes the TAA termination codon of *prgX*, is a 24-base perfect inverted repeat followed by six thymidines, a structure resembling a factor-independent transcription terminator. This structure, having an estimated free energy of \(-32\) kcal (1 kcal = 4.184 kJ/mol (53)), would be thermodynamically stable and thus could serve as the transcription termination signal of *prgX*. It is also possible that this region could be involved in termination of transcription of ORF1. Similarly, IR3 may be the transcription terminator of ORF3. A region between 40 and 90 bases upstream of *prgX* spanned both IR2 and a potential promoter sequence of *prgX*. Several dyad structures were found immediately upstream and within the 5′ translated region of *prgX*. The palindromic sequence IR4 (−15.2 kcal/mol) was located upstream of ORF4 and could be involved in the regulation of *prgR* expression, since similar structures have been shown to be involved in other procaryotic regulatory mechanisms (36). IR5 (−12.8 kcal/mol) is just downstream from the end of *prgA* and also includes a portion of a putative *prgB* promoter region identified recently by mRNA analysis (11). IR5 could thus be involved in termination of *prgA* transcription, control of *prgB* transcription, or both. Although the predicted thermodynamic stability of the secondary structure formed by IR6 (−13.8 kcal/mol) is somewhat low, this sequence potentially could serve as the terminator for *prgA*.

**The translated polypeptides.** Table 3 lists properties of the deduced polypeptides. Although several of the ORFs are fairly short, we have genetic evidence and some transcription data (11) indicating that at least some of these may encode important products. Therefore, we have listed all of the potential ORFs from both strands that would encode polypeptides of at least 40 amino acids.

The translated sequence of *prgA* encoding Sec10 was 98 or 100 kDa, depending on the initiation site used (we favor the second site because it is preceded by a potential ribosome binding site). The mature protein would be 95 kDa upon removal of the signal peptide. This is much smaller than the apparent mass of Sec10, which is 130 kDa on an SDS-polyacrylamide gel. The difference could be due to the high content of small amino acid residues. Twenty-nine percent of the translated ORF9 was made up of glycine, proline, and alanine, which are known to cause the slow migration of polypeptides on SDS-PAGE (25). The high proline content near its C terminus might also contribute to the aberrant migration of Sec10. In this respect, the type 6 M protein of *Streptococcus pyogenes* has been shown to migrate normally when the proline-rich C-terminal end is removed (32). However, posttranslational modifications which often result in altered migration of peptides cannot be ruled out.

The carboxy terminus of Sec10 contained a segment of hydrophobic amino acids followed by six basic residues (Fig. 4). This region shared 28 and 31% identity with the corresponding segment of the staphylococcal protein A and streptococcal M6 protein that were proposed to be the membrane-spanning domains (29, 32). Therefore, this hydrophobic segment of Sec10 might also serve the same function, with the positively charged tail presumably protruding into the cytoplasm. A proline-rich region was observed N terminal to this hydrophobic segment (Fig. 4), including the putative wall-spanning sequence LPQTE, a motif highly conserved among many surface proteins from gram-positive cocci (24). This region of M6 protein is involved in the
FIG. 4. Sequence of the EcoRl c and e fragments of PCF10. The sequence shown represents that of the strand that reads from left to right in the 5'-to-3' direction. The putative products of ORF1, -3, -4, -5, -7, -9, -11, and -15 are indicated in single (uppercase)-letter notation. For all of the ORFs for which there is genetic or other evidence indicating a function in pheromone-inducible conjugation, a prg designation has been given, as noted in the text. With the exception of the putative prgX (ORF2) gene product, the sequences of polypeptides potentially encoded by the ORFs in the complementary strand are shown. A single start and stop sites of these ORFs are indicated. The prgX (ORF2) gene product is shown from left to right in single (lowercase)-letter notation; this gene product and the ribosome binding sites are encoded by the complementary strand. The positions of potential IR structures and their estimated free energies are indicated, and potential glycosylation sites (Cho), potential ribosome binding sites (RBS), and significant amino acids are highlighted. See text for further details.
attachment of the protein to the peptidoglycan moiety of the streptococcal cell wall (46). This region of Sec10 probably has a similar role. Immediately after the second initiation methionine of Sec10 were two lysine residues and a region containing mostly hydrophobic amino acids (Fig. 4). These features are often characteristic of signal peptides of pro-
caryotic secreted proteins (57). Sec10 had a helix-breaking amino acid (Pro or Gly) before the presumed signal peptide cleavage sequence QAA. A hydrophilic region at the N-terminal is characteristic of signal peptides in gram-positive organisms (57). A hydropathy analysis (38) and computer prediction of secondary structure, using the Intelligenetics Suite, indicated that Sec10 is very hydrophilic with the exception of the N and C termini and probably contains extended regions of α-helical structure. The presumed membrane anchor could account for the hydrophobicity of the C terminus, while the hydrophobic core of the signal peptide corresponded to the N-terminal hydrophilic domain.

For Asc10, the predicted molecular weight of the translated amino acid sequence is 142,000, although the relative migration in an SDS-polyacrylamide gel is approximately that of a 150-kDa protein. This discrepancy could be due either to the amino acid composition of the protein or to posttranslational modification. As was observed with Sec10, the amino terminus of Asc10 featured a positively charged segment followed by a hydrophobic core and then a polar region. Two possible signal peptide cleavage sites have been identified in Asc10 at the amino-terminal end of the protein, one between positions 43 and 44 and one between positions 90 and 91. Data reported by Galli et al. (26) from the analysis of Asa1, the aggregation substance encoded by the E. faecalis conjugative plasmid pAD1, would support the first cleavage site at position 43. Asc10 also contains a hydrophobic carboxy terminus preceded by a proline-rich region and the putative wall-spanning sequence LPKTGE. As in the case of Sec10, Asc10 is predicted to be fairly hydrophilic except for the two termini. In contrast to Sec10, computer-generated structural predictions suggest that the Asc10 protein is primarily globular (59). The deduced Asc10 protein also contains two sequences, RGDS and RGDV, that could potentially function in binding of the protein to eucaryotic cell receptors (see Discussion).

The deduced gene products of prgX and prgR did not show any extended hydrophobic region and therefore are likely to be cytoplasmic proteins. The putative prgS (ORF5) product could contain two distinct domains, a hydrophobic N terminus and a hydrophilic C terminus. The translated product of ORF7 would be a small basic protein. We recently initiated physical analysis of the transcription of this region of pCF10 (11). The results to date indicate that ORF7 is part of a pheromone-inducible transcript that also includes most of ORF5 (prgS). ORF7 seems to be located in a region (between the Tn5 insertions in pNY4502 and pNY4162, shown in Fig. 2) where no positional mutants were obtained. As noted above, on the basis of transcriptional data (11) and the similarity of the ORF7 product to another transcriptional activator (see Discussion), we have tentatively designated ORF7 as prgT.

The protein product of ORF15 is predicted to be a 30-kDa, acidic protein. In contrast to other ORFs, this region is high in direct repeats. Several encode proline, and as a consequence this protein is rich in proline at 18%. Like Asc10, this protein has a predicted signal sequence with a cleavage site between amino acids 25 and 26. It is also predicted to have a membrane-spanning region from 256 to 280 which would serve as a membrane anchor. On the basis of data (11) demonstrating pheromone-inducible transcription of this region and the results of DNA homology searches described below, this ORF has been designated prgC, although we do not have any genetic data on the function of this gene as yet.

Recent results (11) from our laboratory indicate that the ORF3 region is transcribed constitutively at high levels, and comparison of this sequence with a corresponding region of pAD1 (12) suggests that a portion of this region might be involved in regulation of pheromone production by the host cell, either by the synthesis of an inhibitor peptide or by other mechanisms. More genetic analysis is required to confirm this. None of our data suggest that this region is involved in regulation of expression of the surface proteins. In the case of the other ORFs (ORF6, -8, -10, -12, -13, -14, and -16), there is not yet any evidence for whether or not they have biological functions.


discussion

The cloned E. coli c and e fragments of the conjugative plasmid pCF10 encode the two major pheromone-inducible surface proteins, Sec10 and Asc10, as well as genes required for their expression in E. faecalis. Thus far, our attempts to purify sufficient quantities of either the Sec10 or Asc10 protein for N-terminal sequencing have been hampered by the low amount of intact material that can be extracted from cells and by their sensitivity to endogenous proteases. However, the genetic analysis presented here, including the characterization of truncated derivatives of each protein encoded by plasmids with insertion or deletion mutations near the 3' ends of the cloned prgA and prgB genes, provides very strong evidence that these genes actually encode these two proteins.

The Sec10 protein is very likely made as a precursor containing a signal peptide, which is consistent with its cell surface location (10, 54). With regard to the putative membrane anchor region encoded by the 3' end of prgA, loss of the membrane anchor could explain the failure to detect truncated Sec10 proteins in surface antigen extractions of E. faecalis carrying pNY4004 and -4503. These two mutants

---

**Table 3. Summary of the properties of the deduced polypeptides encoded by the ORFs in the E. coli c and e fragments of pCF10.**

<table>
<thead>
<tr>
<th>ORF</th>
<th>Gene designation*</th>
<th>No. of amino acid residues</th>
<th>Mol wt</th>
<th>Estimated pl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>prgX</td>
<td>220</td>
<td>24,746</td>
<td>9.7</td>
</tr>
<tr>
<td>2</td>
<td>prgX</td>
<td>317</td>
<td>37,095</td>
<td>6.7</td>
</tr>
<tr>
<td>3</td>
<td>prgR</td>
<td>121</td>
<td>14,468</td>
<td>10.8</td>
</tr>
<tr>
<td>4</td>
<td>prgS</td>
<td>90</td>
<td>10,486</td>
<td>11.3</td>
</tr>
<tr>
<td>5</td>
<td>prgT</td>
<td>87</td>
<td>9,937</td>
<td>11.0</td>
</tr>
<tr>
<td>6</td>
<td>prgE</td>
<td>102</td>
<td>12,209</td>
<td>10.1</td>
</tr>
<tr>
<td>7</td>
<td>prgB</td>
<td>1,305</td>
<td>142,307</td>
<td>5.8</td>
</tr>
<tr>
<td>8</td>
<td>prgC</td>
<td>80</td>
<td>8,195</td>
<td>8.0</td>
</tr>
<tr>
<td>9</td>
<td>prgD</td>
<td>94</td>
<td>9,360</td>
<td>3.6</td>
</tr>
<tr>
<td>10</td>
<td>prgE</td>
<td>74</td>
<td>8,634</td>
<td>7.2</td>
</tr>
<tr>
<td>11</td>
<td>prgF</td>
<td>285</td>
<td>30,131</td>
<td>4.3</td>
</tr>
<tr>
<td>12</td>
<td>prgG</td>
<td>87</td>
<td>12,209</td>
<td>10.1</td>
</tr>
</tbody>
</table>

* The prgA gene encodes the Sec10 surface protein, and the prgB gene encodes the Asc10 protein. The rationale for designation of other genes is given in the text.
were mapped 5' to the putative membrane anchor region (Fig. 1). A hydropathy analysis of Sec10 indicated that this protein is very hydrophilic. Since Sec10 is a surface protein, the majority of the molecule probably is extruded out on the cell exterior. Secondary structure predictions of the prgA product indicated a high probability of helical structure. Although the sequence of the prgA product did not show extended similarity to any of the sequences in the data banks, a region between residues 140 to 180 shared 34% identity with the receptor binding domain (55) of colicins (Fig. 5). The significance of the resemblance is unknown. Some of our data (34) indicate that the recipient ability of *E. faecalis* harboring either the cloned Eco RI c or EcoR I b fragment of pCF10 is reduced. One possibility is that Sec10, in conjunction with another gene product (located in the b fragment, for example), is required to bring about the full expression of the surface exclusion function. In the absence of this second protein, the partial surface exclusion activity of Sec10 would be observed. The surface exclusion barrier in *E. coli* conjugation systems consists of two proteins (3, 47). The outer membrane TraT protein exerts its action by interacting with OmpA to prevent plius binding (47). The inner membrane TraS protein is thought to inhibit the conjugal DNA metabolism that follows mating pair formation (3).

Donor cells respond to pheromones by producing an adhesive cell surface aggregation substance. The interaction between aggregation substance and a chromosomally determined surface binding substance is thought to result in clumping of the cells (13, 17). We recently showed that *E. faecalis* harboring the cloned c and e fragments exhibits a clumping phenotype as well as Asc10 production (10). Furthermore, an excellent correlation was observed between donor ability and the appearance of Asc10, suggesting that Asc10 is involved in the formation of mating aggregates (41). In this and the accompanying report (45), we provide direct evidence that Asc10, the prgB gene product, is the aggregation substance encoded by pCF10. The involvement of Sec10 was ruled out because *E. faecalis* strains containing plasmids with insertions in the prgA structural gene for Sec10 (e.g., 4515 and 4503) showed a clumpy phenotype, while strains producing Sec10 but not Asc10 did not clump.

Asc10 was investigated for important sites and signatures. As noted above, a discrepancy exists between the molecular weight of the protein as observed by SDS-PAGE and that predicted by the computer based on the amino acid sequence. Eight possible glycosylation sites were identified (Fig. 4). Of particular interest are two motifs found in the protein; at amino acid positions 600 and 930 are the residues Arg-Gly-Asp-Ser (RGDS) and Arg-Gly-Asp-Val (RGDV), respectively (Fig. 4). These RG D motifs have also been identified in the same positions in the highly related Asa1 protein (26, 59) encoded by the hemolysin plasmid pAD1, which encodes a pheromone-inducible transfer system stimulated by the peptide pheromone cAD1 (13). These motifs have been identified as recognition sequences for eucaryotic membrane receptors known as integrins (33a), suggesting the possibility that the enterococcal aggregation substances could act as virulence factors mediating adherence to eucaryotic cells.

The deduced product of prgC is similar to that of the ORFC sequence from the broad-host-range conjugative plasmid pAMβ1 (14, 51). The two deduced proteins have a protein similarity of 39%, using a PAM-256 similarity matrix (5), with comparable predicted molecular weights. In addition, both exhibit extensive direct repeats and high proline content. Both the putative prgC product and the ORFC-encoded protein exhibit a predicted signal sequence (residues 1 to 24), a possible membrane anchor region, and are predicted to be associated with the cell surface. The strong similarity between these two gene products suggests that they may function in a step of plasmid transfer that is common both to pheromone-inducible plasmids which transfer efficiently in liquid and to broad-host-range plasmids such as pAMβ1, which generally require solid surfaces to transfer (13).

The prgR and prgS genes appear to be required for prgB expression in *E. faecalis* because several Tn5 insertions in this region completely abolished Asc10 synthesis. It might be predicted that prgR and prgS would act in trans to control prgB expression, because prgA is located between these genes and prgB (Fig. 1 and 5). Tn5 insertion in prgX appears to be lethal to *E. faecalis* cells carrying the Eco RI c fragment, because such insertions resulted in deletions or other rearrangements upon transfer from *E. coli* to *E. faecalis*. Recent subcloning studies suggest that, in addition to prgR and prgS, prgX may be required for prgB expression (11). These recent results also indicate that complete deletion of prgX, including the putative promoter region, generates a plasmid that is stable in enterococci. However, construction of partially deleted prgX derivatives seems to result in lethality, similar to that observed with prgX: Tn5 mutants described here. Thus, it appears that certain truncated prgX derivatives may encode a product that is toxic to *E. faecalis* cells. Although genetic disruption or deletion of the prgX, -R, and -S genes have marked effects on expression of surface antigens involved in plasmid transfer, there is little similarity of these gene products to any regulatory proteins in the data bases, with the exception of some marginal relatedness of the prgX product to transcriptional activators of Drosophila (37) and Shigella (1) species.

We assigned the gene designation prgT to ORF7 on
basis of recent evidence from our laboratory indicating that
this region is part of a pheromone-inducible transcript, along
with prgS (11). The deduced prgT product shows some
similarity (25 to 30% residue identity, depending on the
alignment used) to the hutP gene product of B. subtilis (43).
The latter protein has been shown to be required to activate
transcription of genes involved in utilization of histidine (43).
Because prgT is physically linked to, and seems to be
cotranscribed with, a gene (prgS) required for activation of
prgB and also shows similarity to another transcriptional
activator, it probably interacts with the products of other
genes in the prgX-R-S region in activating prgB.

On the basis of the results of our initial mapping and
cloning studies of pCF10 (9, 10), the results reported here,
and those of ongoing complementation and transcriptional
analyses (11), an overall picture of the complex regulatory
circuits controlling pheromone-inducible conjugational
transfer of pCF10 is beginning to emerge. Our current view of
the control mechanism is shown in Fig. 6. In wild-type pCF10,
there is a fairly large negative control region in the Eco RI b
fragment to the left of Eco RI c (9, 17). In the absence of
pheromone, products of one or more genes in this region
may block expression of aggregation and transfer functions,
possibly by interfering with expression of one or more genes
in the prgR-S-T region, or by blocking the activation process
at a later stage. When this inhibition is overcome by addition
of pheromone, by insertion of a transposon into the negative
control genes (9), or by removing the negative control genes
by molecular cloning (10), the positive control system mediates
activation of aggregation and transfer functions. Analysis
of gene fusions to the prgB promoter and mapping of
transcripts (11) indicate that the activation mechanism is at
the level of transcription. The genes encoding transfer func-
tions include prgB and probably other genes to the right of
prgB, as discussed above and in the accompanying paper
(45).

The prgX gene has some regulatory functions that are not
fully understood at the present time. As described in this
report, Tn5 insertions in this gene have no apparent phenotype
in an E. coli host but appear to be lethal in E. faecalis.

In Fig. 6, we have indicated a second regulatory circuit
involving activation of both prgA and prgB by the prgX gene
product. This view is based on very recent results (11) of
mRNA analysis and genetic complementation indicating that
prgX is transcribed constitutively and that it is required for
expression of both prgA and prgB. Thus, despite the fact
that prgA and prgB have distinct control mechanisms, there
is some overlap in the regulatory circuits. One possible
model that could account for the results obtained thus far is
that prgA transcription is constitutive and requires prgX.
In the absence of pheromone, the transcript terminates at the 3'
end of prgA. Pheromone induction could result in a read-
through of this transcript to include prgB, with the prgR-S-T
region possibly providing an antitermination function.
The presence of an inverted repeat (IR5) in the region that would
be a likely site for a termination/antitermination regulatory
mechanism is consistent with this type of model. Alterna-
tively, the prgB transcript could initiate near the start of the
gene and require products or sites from both the prgX and
the prgR-S-T regions. Molecular and genetic experiments to
distinguish between these models are in progress.

ACKNOWLEDGMENTS

We thank Vickie Johncox, Peter Saurugger, and the staff of
the University of Minnesota Molecular Biology Computing Center
for assistance with computer analysis of sequence data. We also thank
Jungwon Chung for communicating unpublished data and for useful
discussions.

This work was supported by Public Health Service grant A119930
of the National Institutes of Health and by a research grant from the Cornell Biotechnology Program. S.-M.K. was the
recipient of a predoctoral fellowship from the Cornell Biotechnology Program.

REFERENCES

1. Adler, B., C. Sasakawa, T. Tobe, S. Makino, K. Komatsu, and
M. Yoshikawa. 1989. A dual transcriptional activation system
for the 230 kb plasmid genes for virulence-associated antigens

method for isolating large plasmid DNA from lactic strepto-
PHEROMONE-RESPONSIVE GENES OF pCF10


49. Shine, J.,


