Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Gene, 148 (1994) 23–32

GENE 08147

Permissible peptide insertions surrounding the signal peptide-mature protein junction of the ClpG prepilin: CS31A fimbriae of *Escherichia coli* as carriers of foreign sequences

(Recombinant DNA; random mutagenesis; signal sequence; genetic fusions; hybrid protein secretion; chaperone; transmissible gastroenteritis virus; epitope; antigen display; pilin)

Maurice Der Vartanian, Marie-Claire Méchin, Bernard Jaffeux, Yolande Bertin, Isabelle Félix and Brigitte Gaillard-Martinie

Laboratoire de Microbiologie, Institut National de la Recherche Agronomique, Centre de Recherches de Clermont-Ferrand-Theix, 63122 Saint-Genès-Champanelle, France. Tel. (33-73) 624-241

Received by M. Bagdasarian: 21 February 1994; Revised/Accepted: 26 April/8 May 1994; Received at publishers: 6 June 1994

SUMMARY

The *clpG* gene, expressing the *Escherichia coli* major CS31A fimbrial subunit ClpG, was subjected to random mutagenesis by insertion of an *Eco*RI linker and a kanamycin-resistance (Km^R) cassette into the multiple newly generated *Eco*RI sites. The Km^R gene was then excised by *Pst*I, which left a 48-bp linker representing the heterologous sequence. The same procedure was followed to introduce a synthetic oligodeoxyribonucleotide (oligo) corresponding to epitope C from the spike protein S from the porcine transmissible gastroenteritis coronavirus (TGEV). Nine insertion/deletion mutants (indels) that contained long foreign peptides variously located around the ClpG signal peptide (SP) processing site were characterized. A striking feature of this study is the variety of amino acid (aa) insertions in the ClpG prepilin that have little or no effect on CS31A fimbria biogenesis. These 'permissive' sites tolerate inserts of 18 or 19 aa and accept sequences of different natures in view of their aa composition, charge and hydrophobicity. The results obtained here are also interesting in light of the high level of aa sequence conservation seen in the SP and N-terminal domains of the ClpG-related subunits. The structure-function relationship of the ClpG SP is discussed. The TGEV-C epitope fused to the N-terminal end of the mature ClpG protein was cell-surface exposed, as observed on immuno-electron microscopy. Therefore, the CS31A fimbria seems to be a potent tool for the presentation of foreign antigenic determinants or the production of heterologous polypeptides in *E. coli*.

INTRODUCTION

Polymer CS31A is a plasmid-encoded fibrillar protein associated with animal and human pathogenic isolates of *Escherichia coli* (Contrepois et al., 1989; Cherifi et al., 1990). The *clp* gene cluster, which contains seven structural genes (*clpC, clpD, clpE, clpF, clpG, clpH* and *clpI*), encodes all the secretory protein products required for

Correspondence to: Dr. M. Der Vartanian, Laboratoire de Microbiologie, INRA de Clermont-Ferrand-Theix, 63122 Saint-Genès-Champanelle, France. Tel. (33-73) 624-243; Fax (33-73) 624-581.

Abbreviations: *A*, absorbance (1 cm); aa, amino acid(s); Ab, antibody (ies); Ap, ampicillin; bp, base pair(s); ClpG, major CS31A fimbrial subunit; Cm, chloramphenicol; DNase I, bovine pancreatic deoxyribonuclease; dNTP, deoxynucleotide triphosphate; EM, electron microscopy; IgG, immunoglobulin G; indel, insertion/deletion mutant; IPTG, isopropyl-β-D-thiogalactopyranoside; kb, kilobase(s) or 1000 bp; Km, kanamycin; mAb, monoclonal Ab; MCS, multiple cloning site(s); nt, nucleotide(s); oligo, oligodeoxyribonucleotide; PA, polyacrylamide; pAb, polyclonal Ab; PAGE, PA-gel electrophoresis; PBS, 0.14 M NaCl/2.7 mM KCl/1.8 mM KH_2PO_4/10 mM Na_2HPO_4 pH 7.2; PolIk, Klenow (large) fragment of *E. coli* DNA polymerase I; ^{R}, resistance/resistant; SDS, sodium dodecyl sulfate; SP, signal peptide; TGEV, transmissible gastroenteritis virus; wt, wild type; X, any aa; ::, novel junction (fusion or insertion); [], denotes plasmid-carrier state.

SSDI 0378-1119(94)00358-Y

CS31A biogenesis (Martin et al., 1991). They include the major ClpG monomer and several accessory proteins involved in the stabilization, transport and assembly of ClpG. Extensive nt sequence homology throughout the accessory protein-encoding genes of *CS31A*, *K88* and *F41* operons has been demonstrated by Southern hybridization analysis (Casey et al., 1990; Martin et al., 1991; Korth et al., 1991). The accessory systems mediating expression of these operons may be functionally interchangeable, indicating that CS31A, K88 and F41 fimbriae are members of a closely related family (Korth et al., 1992). The *clpG* gene codes for the ClpG precursor containing a typical Gram$^-$ bacterial 21-aa SP, whose processing results in a mature polypeptide of 257 aa with a deduced M_r of 26 777, and migrating as a 29-kDa protein in SDS-PAGE. The nt sequence of *clpG* has a homology of 60 and 30%, and its corresponding aa sequence an identity of 46 and 24% with those of *faeG* and *f41*, which encode the major subunits FacG and F41 composing the K88 and F41 fimbriae, respectively (Girardeau et al., 1991). By analogy with the known K88 fimbria biogenesis (Bakker, 1991), it seems likely that the ClpG precursor is processed by signal peptidase I, and exported across the cytoplasmic membrane through the Sec-dependent part of the general secretory pathway (see, for a review, Pugsley, 1993). In contrast to the general diversity found in fimbrial signal sequences, a highly conserved (80%) SP was observed for ClpG, FaeG and F41 prepilins, with an identical cleavage site for ClpG and FaeG. Conservation of the SP in concert with high homology in N-terminal domains of the ClpG, FaeG and F41 subunits suggests that extreme functional pressure is applied to conserve this sequence (Girardeau et al., 1991).

The aims of this study were to show that in spite of the extensive modifications engineered in ClpG SP, the resulting mutated prepilins are normally processed, and that a viral antigenic determinant inserted into ClpG prepilin between the residues -1 and $+1$ is presented at the *E. coli* cell surface.

RESULTS AND DISCUSSION

(a) Construction of a *trans*-complementation system

To perform the DNA manipulations in the *clpG* subunit gene without affecting the rest of the operon carrying the helper-genes, we subcloned *clpG* and *clp* helper-genes into two separate compatible plasmids (Fig. 1). One of the two resulting recombinant plasmids, pDSPH524, contained all the accessory genes, and the other, pDEV41155, carried only *clpG*. In complementation experiments, pDSPH524 and pDEV41155 were co-selected on the basis of their Cm and Ap resistance, respectively. The *clp* helper-genes encode *trans*-acting proteins allowing the stable expression of *clpG*, and therefore capable of mediating the biogenesis of CS31A fimbriae. No production of CS31A occurs in *E. coli* containing only either pDSPH524 or pDEV41155.

(b) Localization of permissive regions of the ClpG protein

A DNA fragment from a cassette insertion was joined to the plasmid pDEV41155 to introduce in-frame 48 to 54 bp into the *clpG* gene (Fig. 2). The location of the permissive sites in *clpG* was determined by restriction analysis. Each of the permissive insertions, obtained after *Pst*I-Km^R cassette excision, contained two close *Sal*I sites (Fig. 2) which assisted in mapping the mutations. Another single *Sal*I site was located near *clpG* in pDEV41155 (Fig. 1). The examination of the distribution and location of the insertion sites, screened along the ClpG preprotein (Fig. 3), revealed four hot regions for permissivity which included a part of the SP (region I), the N-terminal end of the mature ClpG protein (region II), the central part of ClpG (region III), and the C-terminal end of ClpG (region IV). Only region IV of pilin seemed to be both permissive, variable, hydrophilic and non-secondary structured.

Because no studies have been made to characterize the effects of signal sequence mutations on the production of *E. coli* fimbriae, the only heterologous insertions accurately located by DNA sequencing were those confined to the signal sequence-surrounding region (region I). The properties of these mutants, ClpG 9/10.1, ClpG 15/29.1, ClpG 16/21.1 and ClpG 21/22.1, are described in detail in section **d**.

(c) Insertion of the TGEV-C epitope and immunoreactivity of the hybrid ClpG proteins

Synthetic oligo encoding the C epitope processing site (aa 361–372) from surface protein S of porcine coronavirus was flanked by *Eco*RI cohesive ends (Fig. 4A). Plasmid DNAs isolated from mutants ClpG 9/10.1, ClpG 15/29.1, ClpG 16/21.1 and ClpG 21/22.1 (section **b**), were restricted by *Eco*RI, religated in the presence of an excess of the epitope C-oligo insert, and then complemented by pDSPH524 in *E. coli* DH5α. Colonies with a TGEV-C epitope insertion were screened and analysed by in situ immunoblotting for CS31A fimbria production and C epitope antigenicity. All of the colonies reacted with anti-ClpG pAb but only 2% with the neutralizing site-C-specific 3b.5 mAb. The sequencing of plasmid-DNAs from clones immunoreactive against mAb 3b.5 indicated that all TGEV-C epitope insertions were located precisely between SP and mature ClpG. One of the clones, ClpG 21/22.3C, was selected for Western blot analysis with an anti-ClpG pAb and the mAb 3b.5, using PAGE both

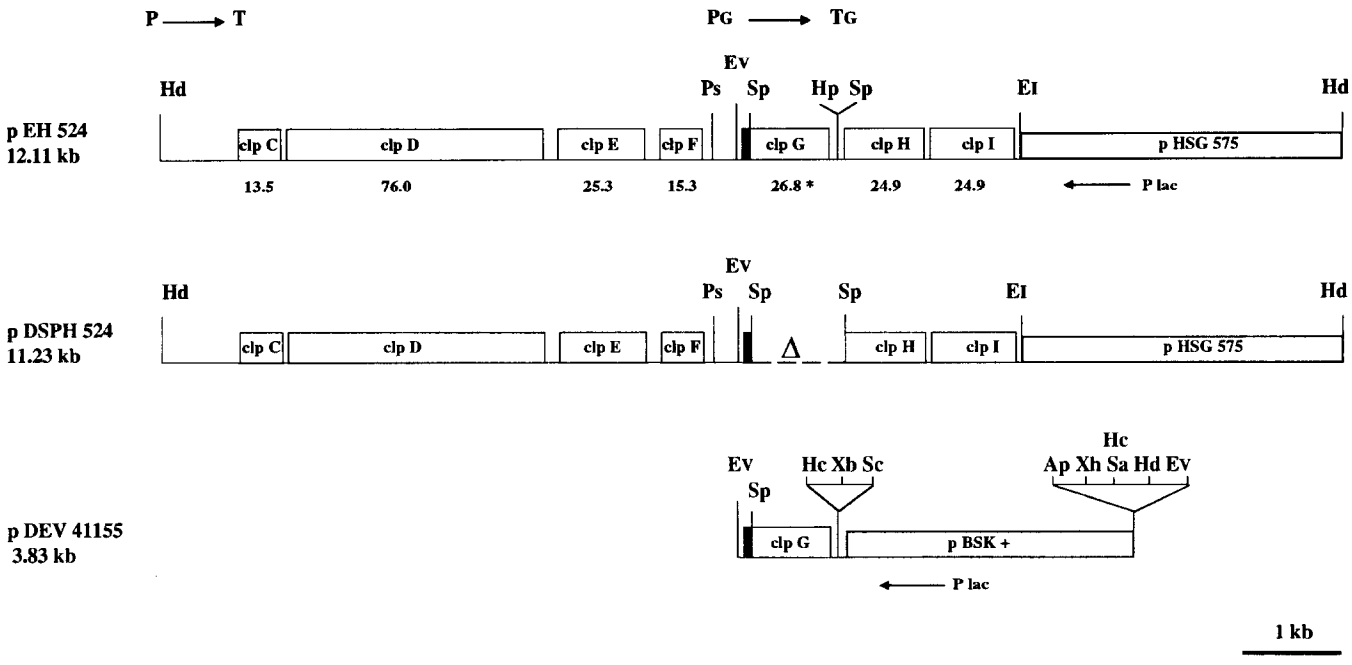

Fig. 1. Construction of the CS31A complementation system. The genetic and physical maps of the hybrid plasmids pEH524, pDSPH524 and pDEV41155 are shown. Plasmids pEH524 and pDSPH524 are low-copy-number plasmids containing a pSC101 replicon, whereas pDEV41155 is a high-copy-number plasmid carrying a pColE1 replicon. Plasmid pEH524 was previously constructed by cloning a 8.5-kb *Eco*RI-*Hin*dIII DNA fragment containing the *CS31A* operon with its own promoters between the *Eco*RI and *Hin*dIII sites of the pHSG575 vector (Martin et al., 1991). To construct a plasmid containing only the *clp* helper genes, *clpG* was deleted from the *CS31A* operon by *Sph*I restriction of pEH524 to give pDSPH524. *clpG* was isolated from pEH524 as a 1.12-kb *Pst*I-*Hpa*I fragment and inserted into *Pst*I + *Hin*cII-digested pUC19. A 0.91-kb *Eco*RV-*Xba*I fragment from this new construct was inserted into *Eco*RV + *Xba*I-digested pBSK + vector to give pDEV41155. The *Eco*RV-*Xba*I fragment that was placed downstream from a strong vector promoter encoded the ClpG precursor protein and contained the three stop codons terminating the open reading frame of *clpG*. Boxes indicate sizes and locations of the various structural *clp* genes. The names of the genes are given in the boxes and the sizes (in kDa) for the mature proteins are beneath the boxes. The black area on the left side of the *clpG* box represents the part of *clpG* coding for the SP. The major fimbrial subunit (pilin) of CS31A fimbria is indicated by an asterisk (after the kDa value). The arrows indicate the direction of transcription: *P*, promoter of the *clp* gene cluster; *PG*, weak promoter of *clpG*; *P*lac, promoter of the lactose operon; *T*, terminator of transcription in the *clp* gene cluster; *TG*, terminator of transcription of the *clpG* gene. Ap, *Apa*I; EI, *Eco*RI; EV, *Eco*RV; Hc, *Hin*cII; Hd, *Hin*dIII; Hp, *Hpa*I; Ps, *Pst*I; Sa, *Sal*I; Sc, *Sac*I; Sp, *Sph*I; Xb, *Xba*I; Xh, *Xho*I; Δ, deletion; pBSK +, pBluescript SK(+) (Stratagene, La Jolla, CA, USA).

under SDS-denaturing and native conditions. Under denaturing conditions, mutant ClpG 21/22.3C produced a protein reacting with both Ab and exhibited a slower migration than wt ClpG (Fig. 4B). The M_r of the monomeric form of the hybrid corresponded to the 257 aa of ClpG plus the 18-aa extension, which includes 12 aa corresponding to the TGEV-C epitope. The ability of hybrid fimbriae to oligomerize was evidenced under non-denaturing conditions with specific anti-ClpG pAb. A difference in migration between the multimeric forms of the native ClpG and native hybrid ClpG::TGEV-C epitope proteins was observed (Fig. 4B), presumably due to the negative net charge (−3) of the oligo insert. Immuno-EM confirmed the presence of the TGEV-C epitope in hybrid fimbriae surrounding the ClpG 21/22.3C cells (Fig. 4C). The method involved the epitope recognition by the 3b.5 mAb and its indirect labeling with goat anti-mouse immunoglobulin G (IgG) conjugated with gold particles. Taken together, these findings demonstrate that epitope C is properly exposed on the monomeric and oligomeric forms of the native hybrid protein at the *E. coli* cell-surface.

Of the remaining 98% of clones that reacted only with anti-ClpG pAb, some were analysed by plasmid-DNA sequencing. In all cases, the TGEV-C epitope-encoding oligo was inserted in-phase in the incorrect orientation and gave mutants ClpG 9/10.2, ClpG 15/29.2, ClpG 16/21.2 and ClpG 21/22.2. This may have been because insufficient clones were analysed to exhaust all the possibilities.

(d) Characterization of preClpG mutants

The examination of the aa sequence immediately upstream from the experimentally determined processing site (Girardeau et al., 1988; Fig. 5) shows that the 21-aa SP of the preClpG has a perfect similarity to typical Gram$^-$ bacterial, signal peptidase-I-processed SP sequences normally present at the N-termini of proteins destined to be translocated through the cytoplasmic membrane by the SecA-mediated pathway (Inouye and

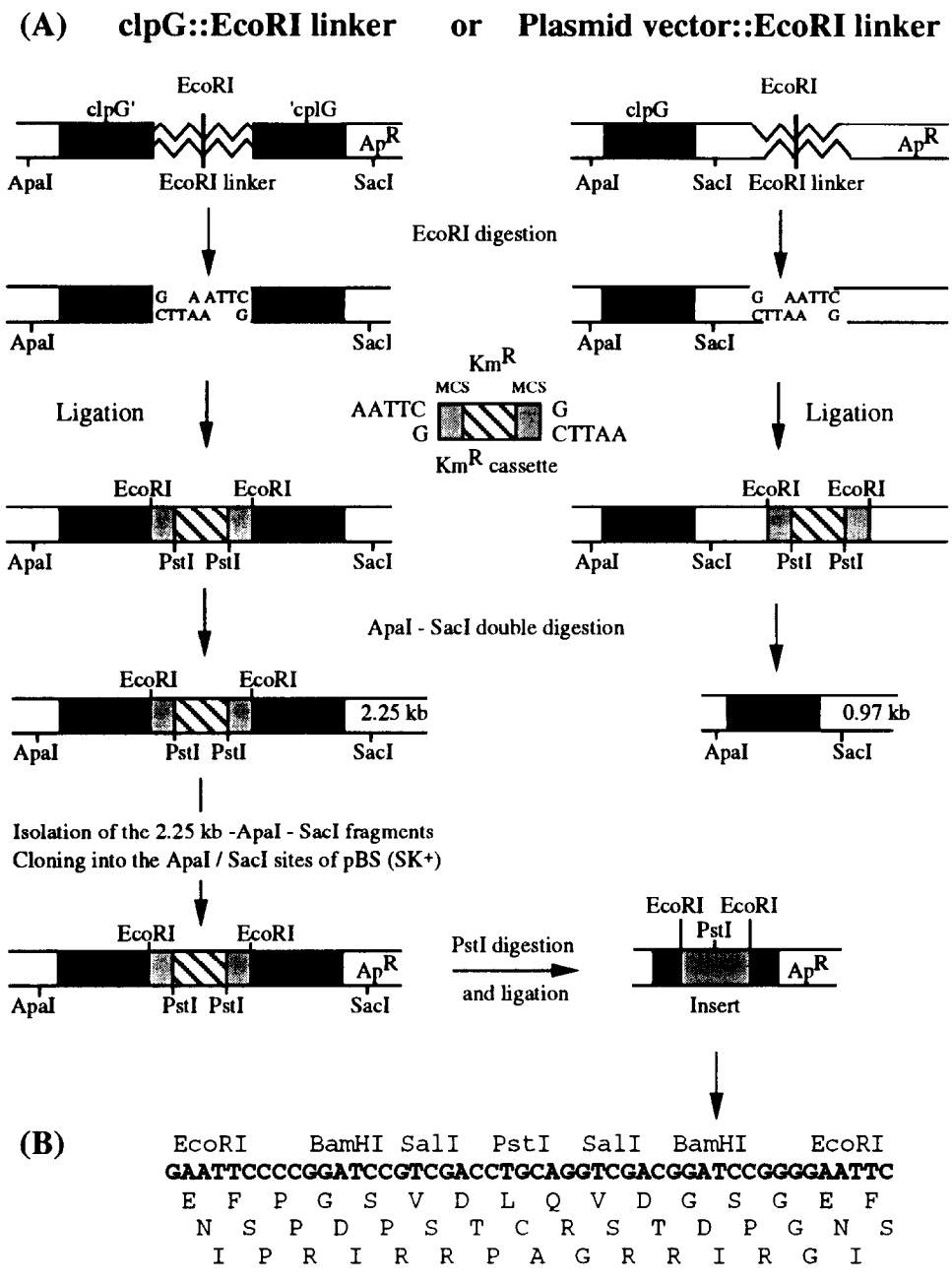

Fig. 2. Mutagenesis by random insertion into *clpG* gene. (**A**) To create random polylinker insertions, pDEV41155 was cleaved with DNase I (BRL) in the presence of Mn^{2+}, as described by Smith and Crouse (1989), to favour the production of singly cut plasmids. The ends were filled-in with T4 DNA polymerase and PolIk plus four dNTPs (Pharmacia). Full-length blunted fragments of the appropriate size were then ligated overnight at 4 C with an excess of unphosphorylated *Eco*RI linkers as described by Lathe et al. (1984). To maintain the correct reading frame, five double-stranded *Eco*RI linkers were used in five simultaneous independent ligation-transformation experiments: d(GGAATTCC), d(CCGAATTCGG), d(CGGAATTCCG), d(CCCGAATTCGGG) and d(CCGGAATTCCGG). After ligation, the T4 DNA ligase was inactivated at 65 C for 15 min and the ligation mixture was used to transform *E. coli* DH5α (BRL). Transformants were grown overnight in a liquid medium instead of the usual plating on solid medium. Resulting plasmids were isolated from stationary phase cultures by alkaline lysis (Birnboim and Doly, 1979), linearized by *Eco*RI, and ligated overnight at 4 C with the Km^R cassette derived from plasmid pUC4K (Vieira and Messing, 1982) as described by Pheiffer and Zimmerman (1983). *E. coli* DH5α transformants were selected on LB plates containing Ap and Km, at concentrations of 100 and 50 μg/ml, respectively, and Ap^RKm^R clones harboring pDEV41155 with Km^R cassette were screened. To select Km^R cassette insertions located only in the *clpG* gene, a pool of plasmid DNAs was prepared from 100 clones Ap^RKm^R and restricted with both *Apa*I and *Sac*I. *Apa*I-*Sac*I fragments of 2.25-kb containing the mutated *clpG* gene were cloned into *Apa*I-*Sac*I sites of the phagemid pBSK + vector (Stratagene). After transformation a pool of plasmid DNAs was prepared from 100 Ap^RKm^R clones. Heterologous sequence insertions of desired size were created in *clpG* by *Pst*I-excision of the Km^R cassette. Linearized plasmids devoid of the Km^R cartridge were purified, religated and transferred into *E. coli* DH5α[pDSPH524]. 12.5% of the $Ap^RCm^RKm^S$ clones produced cell surface CS31A fimbriae as determined by colony immunoblotting (Guesdon et al., 1985) with a specific anti-ClpG pAb (Girardeau et al., 1988). These clones represent insertion sites that were permissive for the introduction of 16–18 extra aa in the ClpG protein. The distribution and the location of these permissive insertions were approximatively determined by *Sal*I restriction analysis of a pool of plasmid DNAs from 100 immunoreacting clones. (**B**) The nt and aa sequences of MCS inserts in the three reading frames, depending upon the type of *Eco*RI linker previously used.

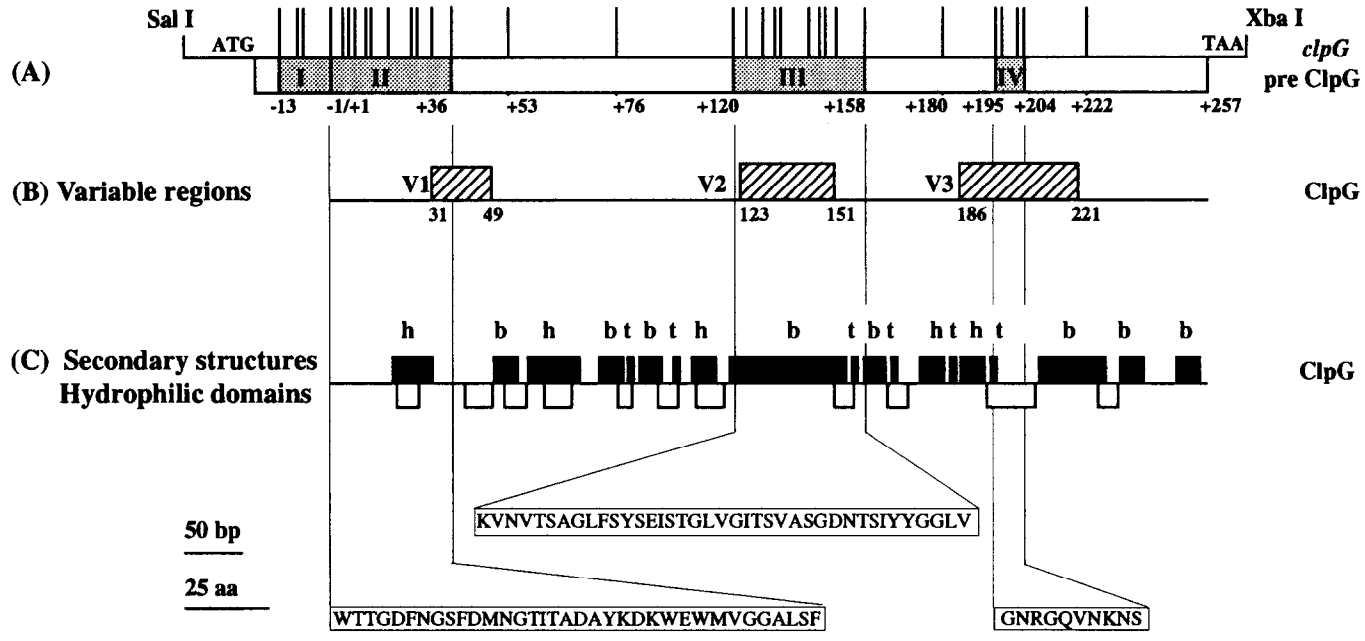

Fig. 3. Schematic diagram showing the mutated *clpG* gene in plasmid pDEV41155 and location of permissible polylinker insertion sites. (**A**) The *clpG* is specified by the start ATG and stop TAA codons. Permissible insertion sites are represented by the upward-extending bars, while the aa sequence of ClpG prepilin is represented by the large downward-extending box, in which permissive regions are identified by shading and symbolized by Roman numerals. Approximate positions of permissive regions are numbered beneath the box representing preClpG according to the aa residues of the mature ClpG sequence. (**B**) The aa sequence of mature ClpG protein is shown. Dashed boxes designate the three variable regions V1, V2 and V3, which were defined by comparison after alignment of the aa sequences of kindred ClpG, FaeG and F41 fimbrial subunits (Girardeau et al., 1991). The number on either side of each box corresponds to the aa residue starting and ending each of these variable regions. (**C**) Diagram showing the aa sequence of mature ClpG protein. The upward-extending black boxes represent secondary structures predicted by the program GORIII using the algorithm of Gibrat et al. (1987). The letters **h**, **b** and **t** correspond to α-helix, β-strand and turn, respectively. Only regions with high potential predicted secondary structures are noted. The downward-extending open boxes designate hydrophilic domains determined from hydropathy profiles analysis by the algorithm of Hopp and Woods (1981). Boxed aa refer to the permissive regions II, III and IV (in panel **A**).

Holegona, 1980). The ClpG SP is considerably different from the SP of type-IV pilins or homologues that are transported through a specialized Sec-independent export machinery and processed by a cognate prepilin peptidase (see, for a review, Hobbs and Mattick, 1993).

We located by DNA sequencing nine mutations generated by random oligo-insertion mutagenesis which specifically altered regions surrounding the signal sequence of *clpG* without drastically affecting the CS31A fimbria biogenesis on the basis of colony- and Western immunoblotting (data not shown) or immuno-EM (only for ClpG 21/22.3C mutant; Fig. 4C). In all cases, depending on the reading frame and surrounding nt, 18 or 19 aa were inserted in-phase into the different sites at positions −13/−12, −7/+8, −6/−1 and −1/+1 (Fig. 5). In some cases the random insertions were accompanied by deletions corresponding to the loss of 4 aa (mutants ClpG 16/21.1 and ClpG 16/21.2) or 13 aa (mutants ClpG 15/29.1 and ClpG 15/29.2; Fig. 5 and Table I). The final modifications are insertions with or without deletions, which are referred to as indels (Betton et al., 1993). Overall, these heterologous insertions introduced 4–7 charged aa residues, with an average net charge of −3 to +1, and several hydrophilic aa, with an average hydrophobicity of −0.25 to −3.55 (Table I).

The indels ClpG 9/10.1 and ClpG 9/10.2, which contain a large disrupting aa sequence in the central hydrophobic core of SP, produced an anti-ClpG pAb-reacting 29-kDa protein comparable to the wt mature ClpG. In contrast, the hybrid ClpG monomers extracted from mutants ClpG 15/29.1 and ClpG 15/29.2, which are affected in the C-domain of the SP plus its 7 downstream aa, migrated slightly slower on SDS-PAGE than did the wt ClpG. In the indels ClpG 16/21.1 and ClpG 16/21.2 the four aa immediately preceding the SP-processing site, were deleted. These mutants revealed two protein bands of 29-kDa and 31-kDa that reacted with the specific anti-ClpG pAb on Western immunoblots. The amount of pilin in these mutants was low. The indels ClpG 21/22.1, ClpG 21/22.2 and ClpG 21/22.3C, containing a foreign peptide at the SP-mature ClpG junction, synthesized hybrid proteins whose sizes, as expected, were slightly greater than that of wt ClpG (Table I).

Since in-frame substitutions in the wt cleavage site do not appear to affect the capacity of some indel pilins to reach the bacterial surface, processing must occur at some

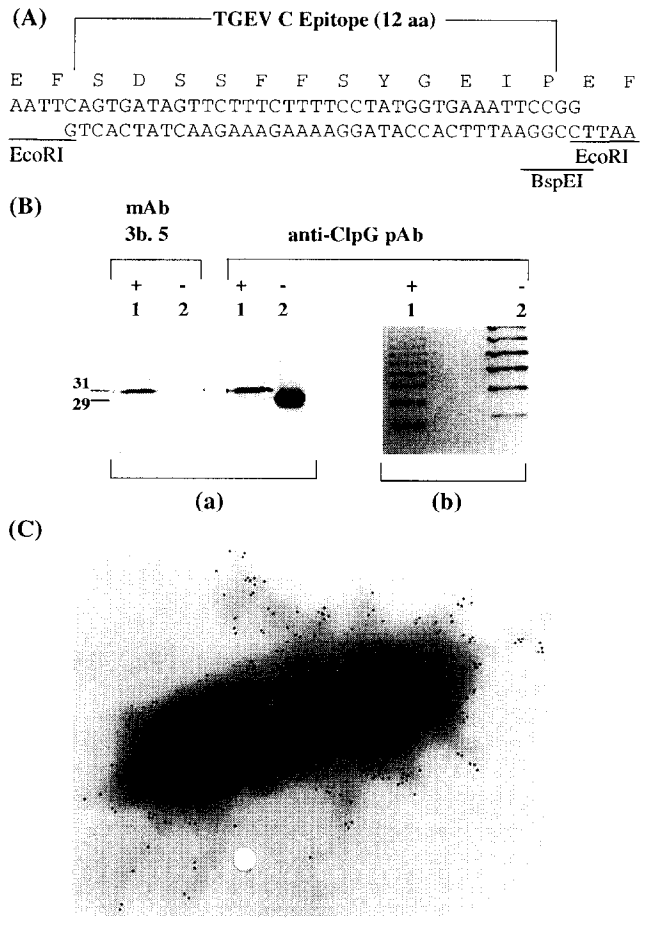

Fig. 4. Analysis of the hybrid ClpG::TGEV-C epitope protein. (**A**) Synthetic oligo encoding the C epitope comprising aa residues 361–372 of the spike S protein of the porcine Purdue-115 strain of TGEV. The site C is a linear neutralizing epitope (Delmas et al., 1990). The mAb 3b.5 raised against TGEV and which recognized the C epitope on native virion (Laude et al., 1986) was used for the accurate screening of the hybrid ClpG::TGEV-C epitope construct. (**B**) Western blot analysis. Plasmids pGETA2, carrying the gene fusion encoding the hybrid ClpG::TGEV-C epitope protein, and pDEV41155, containing only the wt *clpG* gene, were expressed in *E. coli* DH5α[pDSPH524]. Cotransformants were grown overnight on LB agar medium with the appropriate antibiotics. Growing cells were scraped and resuspended in 1 ml of PBS (pH 7.2). The bacterial suspension was then vigorously agitated for 1 min with a top mix shaker, and placed at 60° C for 20 min. After centrifugation the supernatant containing the thermo-eluted fimbriae was collected. Proteins present in the supernatant fraction were separated by 10% PAGE under denaturing (**a**) or native (**b**) conditions. Samples in Laemmli sample buffer were either boiled for 5 min in the presence of SDS and β-mercaptoethanol or incubated at room temperature for 15 min in the absence of SDS and β-mercaptoethanol to obtain the monomeric (**a**) and multimeric (**b**) forms of the proteins, respectively. Western blots were performed by the method of Towbin et al. (1979), with either the murine mAb 3b.5 or a rabbit anti-ClpG pAb, as primary Ab, and either a goat anti-mouse or a goat anti-rabbit IgG conjugated to peroxidase, as secondary Ab. Sheets were stained with 4-chloronaphtol. Lanes: 1, hybrid ClpG proteins; 2, wt ClpG proteins; (+) and (−) refer to the presence or the absence of TGEV-C epitope, respectively. The 29-kDa and 31-kDa bands, corresponding to the wt ClpG and hybrid ClpG proteins, respectively, are indicated. (**C**) immuno-EM of DH5α[pGETA2, pDSPH524] producing CS31A::TGEV-C hybrid fimbriae. Cells were cultured to the mid-exponential phase of growth, collected by brief centrifugation and resuspended in PBS (pH 7.2). Cells were then labeled on EM grids with mAb 3b.5 and 10-nm colloidal gold-labeled goat anti-mouse Ab. Gold labeling of intact cells was carried out essentially as described by Girardeau et al. (1988). The grids were EM examined with a Philips EM400.

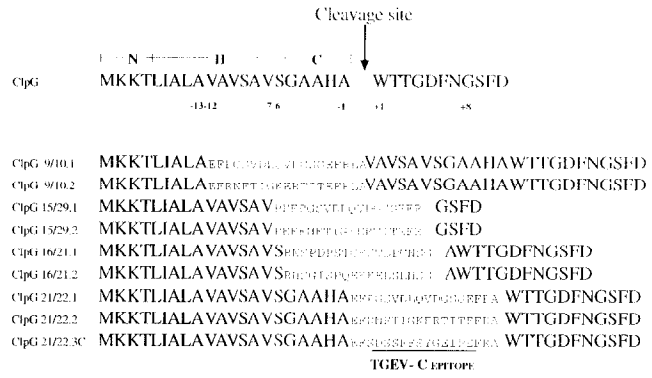

Fig. 5. Sequences of wt and mutant ClpG SP. aa sequences of the wt and mutant SP are shown with their designations on the *left*. Underneath the wt SP sequence is the position of aa relative to the prepilin cleavage site, which is denoted by a vertical arrow. The single capitalized letters N, H and C above the wt SP sequence indicate the polar N-terminal, hydrophobic H-middle and C-terminal parts of the wt SP, respectively. The aa residues from the ClpG protein are in large bold characters and those from the foreign inserted peptides are in small characters. The aa sequence of TGEV-C epitope is underlined.

unknown second site, probably in or upstream from the foreign sequences of modified preClpG, as supported by the finding that in some mutants the mature species clearly migrated slower on SDS-PAGE than did the wt mature ClpG. As suggested by Fikes et al. (1990), there may be circumstances in which the presence of alternate cleavage sites ensures processing in the event of a mutational alteration elsewhere in the signal which shifts the proximity of the core to the normal cleavage site.

(e) Specificity of ClpG SP

To investigate whether ClpG SP specificity is essential for CS31A biogenesis, we fused in-frame the *ompA* signal sequence to the structural *clpG* gene devoid of its signal sequence (Fig. 7A). The choice of the *E. coli* OmpA outer membrane protein was based on the high sequence homology between its SP and those of ClpG, FaeG and F41 prepilins (Fig. 6A). This new construct, carried by pOPA31, expressed a hybrid polypeptide detectable only after total lysis of the cells, suggesting an incorrect fimbrial assembly of hybrid proteins across the outer membrane. Western blot analysis with anti-ClpG pAb (Fig. 7B) revealed that cells harboring complemented pDEV41155 produced only the wt mature ClpG, whereas cells bearing complemented pOPA31 significantly accumulated only the hybrid precursor protein in view of its expected size of about 31-kDa. The conclusion that the

TABLE I

Characterization of ClpG indels

Mutant ClpG[a]	Indel length[b]	ClpG size[c]	Net charge[d]	Average hydrophobicity[e]
ClpG 9/10.1	+19	29	−3 (−4; +1)	−0.25
ClpG 9/10.2	+19	29	+1 (−3; +4)	−2.00
ClpG 15/29.1	+5 (−13/+18)	30	−3 (−4; +1)	−0.75
ClpG 15/29.2	+5 (−13/+18)	30	+1 (−3; +4)	−2.71
ClpG 16/21.1	+14 (−4/+18)	29; 31	0 (−2; +2)	−3.55
ClpG 16/21.2	+14 (−4/+18)	29; 31	+1 (−2; +3)	−2.88
ClpG 21/22.1	+18	32	−3 (−4; +1)	−0.93
ClpG 21/22.2	+18	32	+1 (−3; +4)	−2.76
ClpG 21/22.3C	+18	32	−3 (−4; +1)	−0.59

[a] Mutants ClpG are designated by two slash-separated aa numbers corresponding to the N-terminal (first number) and C-terminal (second number) positions of the indel sites in the ClpG aa sequence relative to the N-terminal Met. The dot-following number indicates the identification order of mutants having a foreign sequence inserted at the same positions in ClpG (for example, ClpG 21/22.3 designates the third mutant selected for an insertion between the aa 21 and the aa 22 of ClpG). The presence of TGEV-C epitope is indicated by the letter C after the dot-following number (ClpG 21/22.3C).

[b] Indel mutation represents insertion with or without deletion, and indel length, i.e., the net length resulting from these insertions. The deletion (−) and insertion (+) lengths, in parentheses, are expressed in the number of aa with respect to wt ClpG.

[c] The size of the mature ClpG protein expressed in kDa, as determined by SDS-PAGE and Western immunoblotting with specific ClpG pAb, as described in the legend to Fig. 4.

[d] Net charge exhibited by the heterologous aa sequence is given by counting +1 for Arg (R) and Lys (K), and −1 for Asp (D) and Glu (E). The numbers of acidic (−) and basic (+) aa residues distributed in the inserted peptide are indicated in parentheses on the left and on the right, respectively.

[e] Average hydrophobic index of inserted peptides in the ClpG sequence environment. Hydrophobic indices were calculated from the hydrophobicity scale of Eisenberg et al. (1984), with an 11-aa window and vectors projecting radially every 100°. The average hydrophobic index was calculated as the arithmetic mean of the hydrophobic index of the 11-aa-long peptides.

OmpA SP did not mediate its proper translocation across the cytoplasmic membrane was strengthened by Western blot analysis with the anti-FLAG peptide M1mAb. This mAb specifically binds to the eight aa FLAG peptide (DYKDDDDK; Fig. 7A) only when it is located at the free N terminus of a FLAG fusion protein, but does not bind to an unprocessed N-terminal FLAG fusion protein (Prickett et al., 1989). Thus, given that the fusion OmpA::ClpG protein did not react against M1mAb

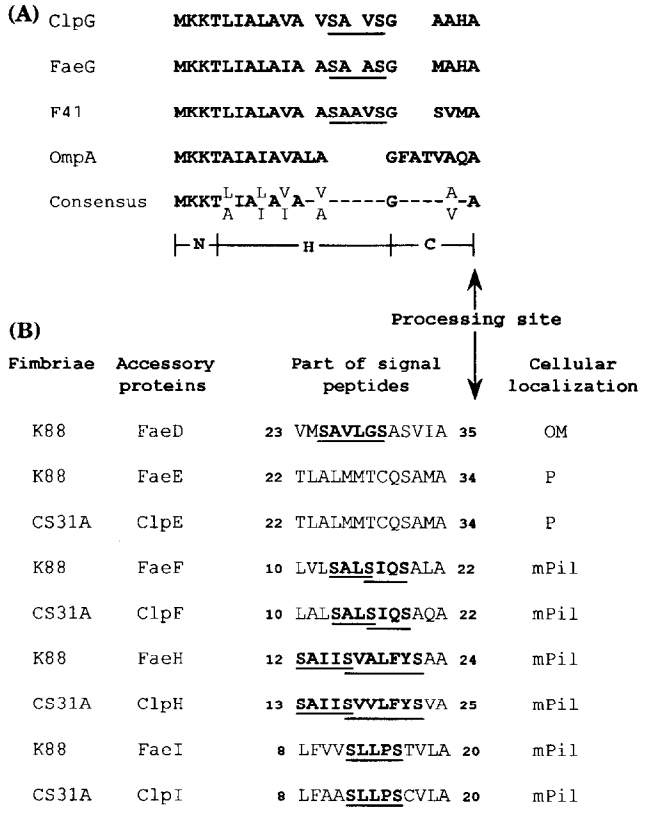

Fig. 6. SP aa sequences of fimbrial components members of the CS31A-related family. (**A**) Comparison and alignment of the SP aa sequences of the major ClpG, FaeG and F41 prepilins, and of the OmpA outer membrane protein. Gaps were introduced to obtain maximal fitting. Identical aa and functionally similar aa from consensus sequence are in bold and normal characters, respectively. The meaning of the capitalized letters N, H and C below the consensus sequence is indicated in the Fig. 5 legend. The cryptic consensus Ser-flanked sequences are underlined. (**B**) Analysis of the C-terminal part of the SP of nine K88 and CS31A accessory proteins. FaeD, FaeE, FaeF, FaeH and FaeI are involved in the K88 fimbria biosynthesis (Bakker, 1991), whereas ClpE, ClpF, ClpH and ClpI refer to the CS31A system (Martin et al., 1991). FaeD functions in the translocation of the fimbrial subunits across the outer membrane (OM) and it may serve as an anchor protein. FaeE and ClpE are periplasmic (P) chaperones. FaeF, ClpF, FaeH, ClpH, FaeI and ClpI are minor fimbrial subunits (mPil) which function in initiation, elongation and/or termination of fimbrial formation. The serine-flanked sequences in large bold characters are underlined. The small bold numbers indicate the aa positions with respect to the full-signal aa sequences.

(Fig. 7B), and from the data reported above, we conclude that the OmpA SP was unprocessed.

Alignment and comparison of the SP aa sequences of ClpG, FaeG and F41 prepilins with that of OmpA (Fig. 6A) showed a cryptic consensus sequence, namely S-A-(A)-V/A-S, only at the C-terminal end of the ClpG, FaeG and F41 hydrophobic cores. It contains polar Ser, hydrophobic Ala and Val residues, which form an α-helix. Previous studies have pointed out the importance of an α-helical conformation in the hydrophobic region for SP function (Pugsley, 1993). In particular, the controlling factor seems to be the overall propensity for helix forma-

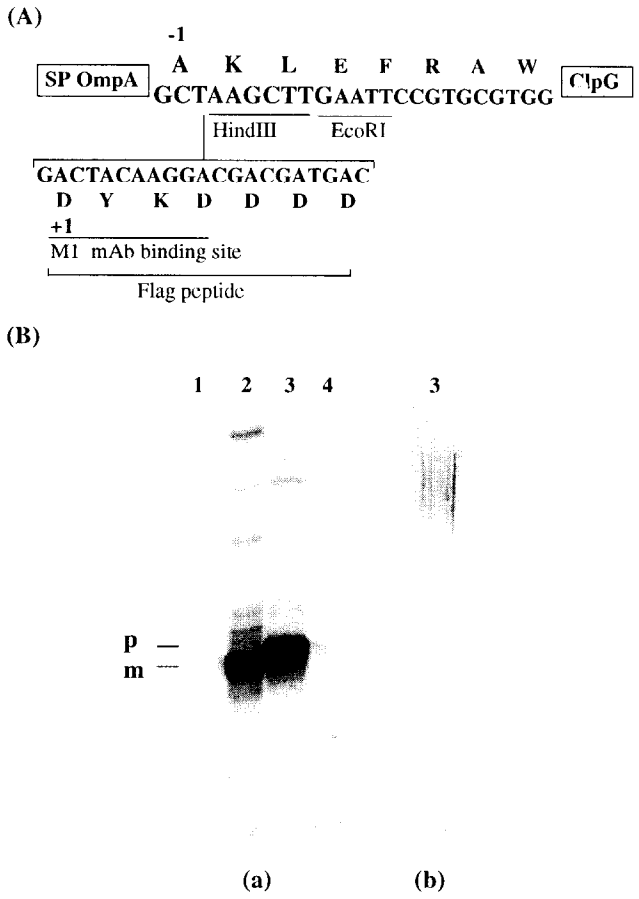

Fig. 7. Construction and expression of the hybrid *ompA::clpG* gene. (**A**) In-frame fusion between *ompA* signal sequence and *clpG* gene is shown. The wt mature ClpG-encoding sequence was removed by *Eco*RI-*Xba*I double digestion of pGETA2 from ClpG 21/22.3C mutant (section **c**, legend to Fig. 4 legend and Fig. 5), and inserted into EcoRI + XbaI-linearized pFLAG-1™ (International Biotechnologies) to give the plasmid pOPA31. The nt and aa from the *clpG* and ClpG sequences are in small characters and those from the *ompA* and OmpA sequences are in large characters. The boxed designations, SPOmpA and ClpG, symbolize the nt and aa sequences corresponding to OmpA SP and ClpG mature protein, respectively. (**B**) *E. coli* DH5α cells harboring either, pDEV41155 (Lane 1), or complemented pDEV41155 (Lane 2), or complemented pOPA31 (Lane 3), or pDSPH524 (Lane 4), were grown at 37°C in LB medium (5 ml) supplemented with appropriate drugs and 1 mM IPTG if necessary (only for the fusion construct). When they reached an A_{600} of 2.0 (for cells containing fusion, IPTG was added at an A_{600} of 0.7), the cells were collected by centrifugation, resuspended in water (20 µl), and diluted twofold in loading buffer (60 mM Tris pH 6.8/10% glycerol/2% SDS/0.05% bromophenol blue/0.25% β-mercaptoethanol). The suspension was then boiled for 5 min and run on a 0.1% SDS-10% PA-gel. Western immunoblot analysis, with either a rabbit anti-ClpG pAb (**a**) or the murine anti-Flag peptide M1 mAb (**b**) was performed as described in the legend to Fig. 4B. m and p indicate wt mature protein and hybrid precursor protein, respectively.

tion rather than the length of the helical region (Lehnhardt et al., 1987). That this phenomena is not applicable to OmpA::ClpG fusion, since it was unprocessed in spite of a normal α-helix-structured OmpA SP, implies that conformation of the SP may be influenced by the mature region. An earlier study (Girardeau et al., 1991) showed a substantial sequence similarity both between the mature regions of ClpG, FaeG and F41 fimbrial subunits and between their corresponding SP. These findings suggest that the importance of a given aspect of SP structure depends on the associated mature region. Thus, in ClpG-related pilins, the conformational properties of aa residues composing the S-A-(A)-V/A-S sequence might either (*i*), play a role in the above association by making the normal cleavage site more accessible to the peptidase, depending on the aa sequences in the vicinity of the site, or (*ii*), act as a topologic signal by which nascent presubunit molecules are recognized for exportation across the cytoplasmic membrane by a specialized export machinery homologous to CS31A, K88 and F41 systems and capable of positioning the prepilin for the cleavage event within the membrane in association with one of CS31A accessory proteins. The CS31A fimbrial ClpE chaperone, required for the transport of ClpG across the periplasm (Bertin et al., 1993), appears as a good candidate capable of interacting with both ClpG SP and mature ClpG. Indeed, Hultgren et al. (1989) reported that processing of the PapG adhesin by signal peptidase I is enhanced by its interaction with the PapD pilin chaperone, and that the last 13 aa at the C-terminal part of the mature PapG were directly involved in the PapG-PapD complex formation. Fimbriae are usually composed of several different subunits, all of which are recognized by the same molecular chaperone. If our second hypothesis is true, the specific mechanism by which the membrane translocation of the major fimbrial subunits occurs through the Ser-flanked sequence recognition by ClpE, must be also applicable to other exported proteins members of the fimbrial assembly machinery. In agreement with this reasoning, most SP aa sequences of nine CS31A and K88 accessory proteins revealed a Ser-flanked segment of 4–7 aa residues near the peptidase cleavage site (Fig. 6B). This segment was missing only for FaeE and ClpE chaperones, strengthening the view that ClpE could be involved in the Ser-flanked sequence recognition. Since the Ser-flanked part of SP is always very closed to the normal peptidase recognition sequence, ClpE binding to this part of SP would authorize the accessibility of cleavage site to signal peptidase I.

(f) Conclusions

(*1*) Integrity of the CS31A fimbrial ClpG subunit-related SP is not strictly required for pilin assembly. Indeed, despite the extensive modifications in the signal sequence region, the mutant ClpG pilin was processed in every case. In particular, the normal peptidase recognition sequence is dispensable since major substitutions of residues across this processing junction were permissive for CS31A fimbria production. Although this is the first

time these findings have been reported for an *E. coli* pilin, similar observations were made with *Pseudomonas aeruginosa* type-IV pilin leader-peptides mutants (Strom and Lory, 1991).

(*2*) In agreement with the conclusions of Hemilä et al. (1992), fusions containing intact SP, which are common when producing heterologous proteins, are not necessarily the most suitable.

(*3*) The work reported here suggests a specific functional interrelationship of the ClpG SP with its associated mature region. A possibility is that the mature region and the SP interact either directly through definite aa zones of recognition, or indirectly in association with other components of the CS31A fimbria assembly machinery.

(*4*) It is possible to insert a viral antigenic determinant into CS31A fimbriae in such a way that it is cell-surface exposed and that intact recombinant cells are recognized by a mAb directed against the viral epitope. This opens up the possibility of creating diagnostic reagents based on whole cells, of developing new live vaccins and of producing foreign polypeptides.

ACKNOWLEDGEMENTS

We are grateful to Dr H. Laude (INRA, Laboratoire de Virologie et Immunologie Moléculaires, France) for kindly providing the sequence of the antigenic TGEV-C epitope and the mAb 3b.5. This work was supported by grants AGRE-0008-C from European Economic Community (ECLAIR program).

REFERENCES

Bakker, D.: Studies on the K88 Fimbriae of Enteropathogenic *Escherichia coli*. Thesis. Vrije Universiteit, Amsterdam, 1991.

Bertin, Y., Girardeau, J.-P., Der Vartanian, M. and Martin, C: The ClpE protein involved in biogenesis of the CS31A capsule-like antigen is a member of a periplasmic chaperone family in Gram-negative bacteria. FEMS Microbiol. Lett. 108 (1993) 59–68.

Betton, J.M., Martineau, P., Saurin, W. and Hofnung, M.: Location of tolerated insertions/deletions in the structure of the maltose binding protein. FEBS Lett. 325 (1993) 34–38.

Birnboim, H.C. and Doly, J.: Rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7 (1979) 1513–1523.

Casey, T.A., Moseley, S.L. and Moon, H.W.: Characterization of bovine septicemic, bovine diarrheal, and human enteroinvasive *Escherichia coli* that hybridize with K88 and F41 accessory gene probes but do not express these adhesins. Microb. Pathog. 8 (1990) 383–392.

Cherifi, A., Contrepois, M., Picard, P., Goullet, P., De Rycke, J., Fairbrother, J.M. and Barnouin, J.: Factors and marker of virulence in *Escherichia coli* from human septicemia. FEMS Microbiol. Lett. 70 (1990) 279–284.

Contrepois, M., Fairbrother, J.M., Kaura, Y.K. and Girardeau, J.-P.: Prevalence of CS31A and F16S surface antigens in *Escherichia coli* isolates from animals in France, Canada and India. FEMS Microbiol. Lett. 59 (1989) 319–324.

Delmas, B., Rasschaert, D., Godet, M., Gelfi, J. and Laude, H.: Four major antigenic sites of the coronavirus transmissible gastroenteritis virus are located on the N-terminal half of spike glycoprotein S. J. Gen. Virol. 71 (1990) 1313–1323.

Eisenberg, D., Schwartz, M., Kamarsmy, M. and Wall, R.: Analysis of membrane and surface protein sequences with hydrophobic moment plot. J. Mol. Biol. 179 (1984) 125–142.

Fikes, J.D., Barkocy-Gallagher, G.A., Klapper, D.G. and Bassford Jr., P.J.: Maturation of *Escherichia coli* maltose-binding protein by signal peptidase I in vivo. J. Biol. Chem. 265 (1990) 3417–3423.

Gibrat, J.F., Garnier, J. and Rosbon, B.: Further developments of proteins secondary structure prediction using information theory. J. Mol. Biol. 198 (1987) 425–443.

Girardeau, J.-P., Bertin, Y., Martin, C., Der Vartanian, M. and Boeuf, C.: Sequence analysis of the *clpG* gene, which codes for surface antigen CS31A subunit: evidence of an evolutionary relationship between CS31A, K88 and F41 subunit genes. J. Bacteriol. 173 (1991) 7676–7683.

Girardeau, J.-P., Der Vartanian, M., Ollier, J.L. and Contrepois, M.: CS31A, a new K88 related fimbrial antigen on bovine enterotoxigenic and septicemic *Escherichia coli* strains. Infect. Immun. 56 (1988) 2180–2188.

Guesdon, J.L., Bouges-Bocquet, B., Debarbouillé, M. and Hofnung, M.: In situ enzyme immunodetection of surface or intracellular bacterial antigens using nitrocellulose sheets. J. Immunol. Methods 84 (1985) 53–63.

Hemilä, H., Pakkanen, R., Heikinheimo, R., Palva, E.T. and Palva, I.: Expression of the *Erwinia carotovora* polygalacturonase-encoding gene in *Bacillus subtilis*: role of signal peptide fusions on production of a heterologous protein. Gene 116 (1992) 27–33.

Hobbs, M. and Mattick, J.S.: Common components in the assembly of type 4 fimbriae, DNA transfer systems, filamentous phage and protein-secretion apparatus: a general system for the formation of surface-associated protein complexes. Mol. Microbiol. 10 (1993) 233–243.

Hopp, T.P. and Woods, K.R.: Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA 78 (1981) 3824–3828.

Hultgren, S.J., Linberg, F., Magnusson, G., Kihlberg, J., Tennent, J.M. and Normark, S.: The PapG adhesin of uropathogenic *Escherichia coli* contains separate regions for receptor binding and for the incorporation into the pilus. Proc. Natl. Acad. Sci. USA 86 (1989) 4357–4361.

Inouye, M. and Halegoua, S.: Secretion and membrane localization of proteins in *Escherichia coli*. Crit. Rev. Biochem. 7 (1980) 339–371.

Korth, M.J., Apostol, J.M. and Moseley, S.L.: Functional expression of heterologous fimbrial subunits mediated by the F41, K88 and CS31A determinants of *Escherichia coli*. Infect. Immun. 60 (1992) 2500–2505.

Korth, M.J., Schneider, R.A. and Moseley, S.L.: An F41-K88-related genetic determinant of bovine septicemic *Escherichia coli* mediates expression of CS31A fimbriae and adherence to epithelial cells. Infect. Immun. 59 (1991) 2333–2340.

Lathe, R., Kieny, M.P., Skory, S. and Lecocq, J.-P.: Laboratory methods linker tailing: unphospho-rylated linker oligonucleotides for joining DNA termini. DNA 3 (1984) 173–182.

Laude, H., Chapsal, J.M., Gelfi, J., Labiau, S. and Grosclaude, J.: Antigenic structure of transmissible gastroenteritis virus, I. Properties of monoclonal antibodies directed against virion proteins. J. Gen. Virol. 67 (1986) 119–130.

Lehnhardt, S., Pollitt, S. and Inouye, M.: The differential effect of two hybrid proteins of deletion mutations within the hydrophobic region of the *Escherichia coli* OmpA signal peptide. J. Biol. Chem. 262 (1987) 1716–1719.

Martin, C., Boeuf, C. and Bousquet, F.: *Escherichia coli* CS31A fimbriae: molecular cloning, expression and homology with the K88 determinant. Microb. Pathog. 10 (1991) 429–442.

Pheiffer, B.H. and Zimmerman, S.B.: Polymer-stimulated ligation: enhanced blunt- or cohesive-end ligation of DNA or deoxyribooligonucleotides by T4 DNA ligase in polymer solutions. Nucleic Acids Res. 11 (1983) 7853.

Prickett, K.S., Amberg, D.C. and Hopp, T.P.: A calcium dependent antibody for identification and purification of recombinant proteins. BioTechniques 7 (1989) 580–589.

Pugsley, A.P.: The complete general secretory pathway in Gram-negative bacteria. Microbiol. Rev. 57 (1993) 50–108.

Smith, M.L. and Crouse, G.F.: Construction of linker-scanning mutations using a kanamycin-resistance cassette with multiple symmetric restriction sites. Gene 84 (1989) 159–164.

Strom, M.S. and Lory, S.: Amino acid substitutions in pilin of *Pseudomonas aeruginosa*. J. Biol. Chem. 266 (1991) 1656–1664.

Towbin, H., Staehelin, T. and Gordon, J.: Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheet: procedure and some applications. Proc. Natl. Acad. Sci. USA 76 (1979) 4350–4354.

Vieira, J. and Messing, J.: The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19 (1982) 259–268.