Volume 286, number 1,2, 79–82 FEBS 09919 July 1991

ADONIS 001457939100631A

Expression of the envelope antigen F1 of *Yersinia pestis* is mediated by the product of *caf1M* gene having homology with the chaperone protein PapD of *Escherichia coli*

E.E. Galyov[1], A.V. Karlishev[2], T.V. Chernovskaya[1], D.A. Dolgikh[1], O.Yu. Smirnov[1], K.I. Volkovoy[2], V.M. Abramov[1] and V.P. Zav'yalov[1]

[1]*Institute of Immunology, 142380 Lyubuchany, Moscow Region, USSR* and [2]*All-Union Research Institute of Applied Microbiology, 142279 Obolensk, Moscow Region, USSR*

Received 23 April 1991

The effective synthesis of the envelope antigen F1 of *Y. pestis* in *E. coli* HB101 is mediated by the expression of the *caf1M* gene. This gene was sequenced, and the protein encoded was found to have a significant homology with the chaperone protein PapD of uropathogenic *E. coli*. The data presented allow one to suppose Caf1M and PapD proteins perform similar functions in the biogenesis of the *Y. pestis* capsule and *E. coli* P-pili, respectively.

Chaperone protein; Nucleotide sequence; Envelope antigen expression; *Yersinia pestis*

1. INTRODUCTION

Most strains of *Y. pestis* (causative agent of plague) possess the virulence plasmid pFra (110 kb in size) which mediates the expression of the fraction 1 envelope antigen at 37°C [1]. The envelope is associated with virulence, and the protein F1 is able to induce protective immunity against plague [2]. Recently the plasmid-located gene *caf1* encoding the capsular antigen F1 of *Y. pestis* has been cloned and sequenced [3]. The cosmid p153 carrying a segment (of about 40 kb in size) of plasmid pFra mediated the synthesis of the antigen F1, that enabled the detection of the p153 in the library of pFra clones by enzyme immunoassay. At the same time the recombinant plasmid pF18L carrying the *caf1* gene (Fig. 1A) and a fairly long preceding segment (263 bp) involving a possible promoter and ribosome binding site, did not ensure the F1 expression in *E. coli* [3]. These data stimulated us to identify the genetic elements affecting the expression of the F1 and making up a single operon together with the *caf1* gene.

2. MATERIALS AND METHODS

2.1. *Bacterial strains and DNA manipulations*

The *E. coli* strain HB101 was used as transient host for plasmids. Cultures were grown overnight while shaking at 37°C in liquid LB or on solid medium supplemented with 50 μg/ml ampicillin. Isolation of plasmid DNA by alkali-lysis method, DNA cloning and Maxam-Gilbert sequencing were performed essentially as described by Maniatis et al. [4]. The plasmid p12R was used for gene sequencing. Both strands were sequenced.

Correspondence address: V.P. Zav'yalov, 142380 Lyubuchany, Moscow Region, Chekhov District, USSR

2.2. *Construction of plasmid derivatives*

Description of plasmids pUC19, pFS2-13, pF18L and cosmids p153, pFS2 was reported earlier [3,5]. The cosmid pFS2 was digested with *Sal*GI and then partially with *Hin*dIII; 5.7-kb fragment was isolated and cloned into pUC19. The plasmid generated was named p12R. Plasmid p14-34 was constructed by deletion of *Pst*I-*Kpn*I fragment of p12R. Plasmid pCDB carrying a frame-shift mutation at the unique site *Bam*HI in *caf1M* was constructed by linearizing p14-34 with *Bam*HI, filling in the ends with Klenow fragment and blunt-end ligating.

2.3. *Enzyme immunoassay and immunoblot analysis*

Total F1 content was determined by ELISA using peroxidase-conjugated monoclonal antibodies.

Lysates of recombinant *E. coli* strains were electrophoresed under reducing conditions in 15% polyacrylamide gel with SDS, and then electroblotted to nitrocellulose. The F1 antigen was detected using rabbit polyclonal serum to F1 and peroxidase-conjugated goat anti-rabbit IgG.

3. RESULTS AND DISCUSSION

To establish the localization of the possible accessory genetic elements in the *caf1* DNA, we tried to identify the smallest region required for expression of the antigen F1. To achieve this, subclones of p153: pFS2, pFS2-13, p12R, p14-34, pF18L (Fig. 1A) were used for transformation of the *E. coli* HB101 and F1 production mediated by the plasmids was analysed by enzyme immunoassay and immunoblot analysis. As shown in Fig. 1B high level of the F1 expression was mediated by plasmids p12R, p14-34 and cosmid pFS2, whereas the shorter plasmids pF18L and pFS2-13 failed to promote

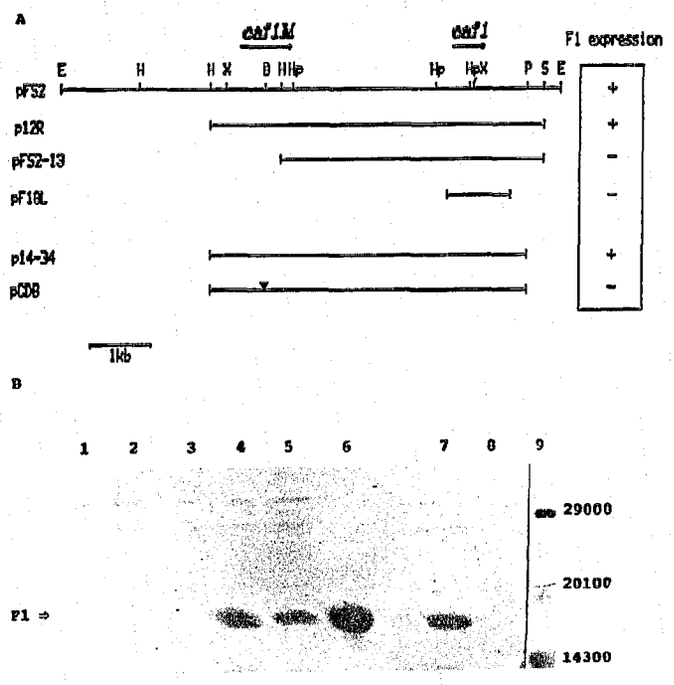

Fig. 1. (A) Restriction map of pFS2 and its derivative plasmids; plasmid-mediated F1 expression as estimated by the enzyme immunoassay. Localization of *caf1* and *caf1M* genes is indicated by arrows at the top. Position of the shift mutation in pCDB is indicated by ▼. *Abbreviations*: H, *Hin*dIII; Hp, *Hpa*I; P, *Pst*I; S, *Sal*GI; E, *Eco*RI; B, *Bam*HI. (B) Immunoblots of the proteins from cell lysates. Lanes: (1) HB101/pUC19; (2) HB101/pF18L; (3) HB101/pFS2-13; (4) HB101/p12R; (5) HB101/pFS2; (6) F1 isolated from *Y. pestis*; (7) HB101/p14-34; (8) HB101/pCDB; (9) biotinylated molecular weight markers (mwm). The proteins were immunostained with rabbit anti-F1 serum. Incubation with goat anti-rabbit peroxidase-conjugated IgG was followed by the reaction with diaminobenzidine. Biotinylated mwm were stained with peroxidase-conjugated avidin.

a detectable extent of the F1 synthesis. It follows from the results that the deletion of the 1.2-kb *Hin*dIII-*Hin*dIII fragment of the plasmid p12R generating the pFS2-13 (Fig. 1A) exerts a dramatic influence on the F1 expression. Thus, at least one of the functionally important (for the F1 expression) regions of Caf1 DNA is completely or partially located inside the fragment.

The 1.1-kb DNA fragment located between the *Xba*I and *Hpa*I sites of the plasmid p12R was sequenced and found to contain an open reading frame 258 amino acid residues long, capable of encoding the 28764-Da protein (Fig. 2). The gene was named *caf1M* (*caf1* mediator). The open reading frame is preceded by well conserved Shine-Dalgarno sequence, TAAGGAGGT, located 7 bp before the initiation codon. The region upstream of the ribosome-binding site has putative promoter sequences similar to the *E. coli* consensus sequences (−35, −10) [6] (Fig. 2). The three-fold repeated sequence, TTAGCTATTTG$^{C}/_{T}$GCA, located upstream of the putative promoter (Fig. 2) might be involved in the regulation of the *caf1M* gene.

```
TCTAGATAGTGGGCCGGCGCCGGGACTTAGCTATTTGCGCATACCCAGCAACACCAATCT   60
TAGCTATTTGTGCACGCGCATCAATATCAAAATTAGCTATTTGCGCAACAAGCAAGTGGA  120
GTGCGCGAAAAGCTAAACTTTGTGTGCATTTTTAAATAAAATTGTTCTCAGTGAGGCTGT  180
                        -35                 -10
GCTACGGATATAAAAATCCCCTTCATTTGTTACCCACCTTTTTACGCATATCGTCGATAT  240
                                                          H
GAAATGATGGGGAGGGGGTGGGAAGGTGTTGTCACCATTCCGTAAGGAGGTTAAGCTCAT  300
 I  L  N  R  L  S  T  L  G  I  I  T  F  G  M  L  S  F  A  A   21
GATTTTAAATAGATTAAGTACGTTAGGAATTATTACTTTCGGCATGCTTAGTTTTGCTGC  360
 N  S  A  Q  P  D  I  K  F  A  S  K  E  Y  G  V  T  I  G  E   41
GAACTCTGCTCAACCAGATATCAAATTCGCAAGCAAAGAGTATGGCGTGACTATAGGTGA  420
 S  R  I  I  Y  P  L  D  A  A  G  V  M  V  S  V  K  N  T  Q   61
GAGTAGGATCATATACCCGTTAGATGCTGCTGGCGTTATGGTCTCGGTGAAAAACACCCA  480
 D  Y  P  V  L  I  Q  S  R  I  Y  D  E  N  K  E  K  E  S  E   81
AGATTATCCGGTTCTCATTCAGTCTAGGATCTACGACGAGAATAAAGAAAAAGAATCAGA  540
 D  P  F  V  V  T  P  P  L  F  R  L  D  A  K  Q  Q  N  S  L  101
GGATCCTTTCGTGGTCACTCCGCCATTGTTTCGATTGGATGCTAAGCAACAAAATTCTTT  600
 R  I  A  Q  A  G  G  V  F  P  R  D  K  E  S  L  K  W  L  C  121
GCGTATAGCTCAGGCTGGAGGTGTTTTCCCGCGAGATAAAGAGAGCCTAAAGTGGTTATG  660
 V  K  G  I  P  P  K  D  E  D  I  W  V  D  D  A  T  N  K  Q  141
CGTAAAAGGGATTCCACCAAAGGATGAAGATATATGGGTTGATGATGCGACAAATAAGCA  720
 K  F  N  P  D  K  D  V  G  V  F  V  Q  F  A  I  N  N  C  I  161
AAAATTCAATCCAGACAAAGATGTGGGAGTGTTCGTGCAATTCGCAATTAATAATTGCAT  780
 K  L  L  V  R  P  N  E  L  K  G  T  P  I  Q  F  A  E  K  L  181
TAAGCTTTTGGTTCGACCGAATGAATTAAAAGGAACCCCTATACAGTTTGCTGAAAAGTT  840
 S  W  K  V  D  G  G  K  L  I  A  E  N  P  S  P  F  Y  M  N  201
AAGCTGGAAAGTTGATGGGGGGAAGCTAATTGCTGAAAACCCCTCACCTTTCTACATGAA  900
 I  G  E  L  T  F  G  G  K  S  I  P  S  H  Y  I  P  P  K  S  221
CATAGGTGAATTAACATTTGGAGGGAAAAGTATTCCTTCTCACTATATTCCACCTAAATC  960
 T  W  A  F  D  L  P  K  G  L  A  G  A  R  N  V  S  W  R  I  241
GACGTGGGCTTTTGATTTGCCAAAAGGACTAGCGGGAGCACGTAATGTTTCGTGGAGAAT 1020
 I  N  D  Q  G  G  L  D  R  L  Y  S  K  N  V  T  L  *       258
AATTAATGATCAGGGAGGGTTGGATCGTTTGTATTCCAAAAATGTGACTTTATGATGATG 1080

TTTAAAGGGGACGGGAATAATGAGGTATTCAAAGCTGTTCCTGTGTGCAGGTTAAC     1136
```

Fig. 2. Nucleotide sequence of the *caf1M* gene and deduced amino acid sequence. The −35 and −10 hexamers of possible promoter are underlined; a probable ribosome-binding site (Shine-Dalgarno sequence) is marked with dashed lines; repeated sequences are boxed. The line above the sequence indicates the possible signal peptide.

To elucidate the importance of the *caf1M* gene product, we constructed plasmid pCDB with the frameshift mutation in *caf1M* by inserting additional 4 bp at the unique *Bam*HI site in p14-34. From the DNA sequence of the *caf1M* gene (Fig. 2) the mutated gene was predicted to give rise to a translation product carrying 82 amino acids of the N-terminal portion (including the signal sequence) of the normally 258 amino acids long Caf1M followed by 14 erroneous amino acids encoded by the additional 4 bp and the +2 frame. This frameshift mutant unlike the wild-type did not produce the F1 antigen detectable by enzyme immunoassay or immunoblot analysis (Fig. 1B). The presented data show that the effective synthesis of the capsular antigen F1 of *Y. pestis* is mediated by the expression of the *caf1M* gene.

The PIR protein sequence library was used to search for homology between Caf1M and other protein se-

```
Caf1M  1MILNRLSTLGIITFGMLSF-AANSAQPDIKFASKEYGVTIGESRIIYPLDAAGVMVSVKNTQD 62

PapD   1MIRKKILMAAIPLFVISGADA---------------AVSLDRTRAVFDGSEKSMTLDISNDNK 48
                                                            A        B

-YPVLIQSRIYDENKEKESEDPFVVTPPLFRLDAKQQNSLRIAQAGGV--FPRDKESLKWLCVKGIPPKD 129

QLPYLAQAWIENENQEKIITGPVIATPPVQRLDPGAKSMVRLSTTPDISKLPQDRESLFYFNLREIPPRS 118
     C                            D      E                  F     G

EDIWVDDATNKQKFNPDKDVGVFVQFAINNCIKLLVRPNELKGTPIQF--AEKLSWKVDGGKLIAENPSP 197

EKANV------------------LQIALQTKIKLFYRPAAIKTRPNEVWQDQLILNKVSGGYRI-ENPTP 170
                                                         A       B

FYMN-IGELTFGGKSIPSHYIPPKSTWAFDLPKGLAGARNVSWRIINDQGGLDRLYSKNVTL         258

YYVTVIGLGGSEKQAEEGEFETVMLSPRSEQTVKSANYNTPYLSYINDYGGRPVLSFICNGSRCSVKKEK 239
  C                    D     E              F      G            H
```

Fig. 3. Primary structure alignment of the chaperone protein PapD of *E. coli* [7,8] and Caf1M of *Y. pestis*, β-sheet sites in the PapD protein are underlined. The signal peptide sequences are marked with italics.

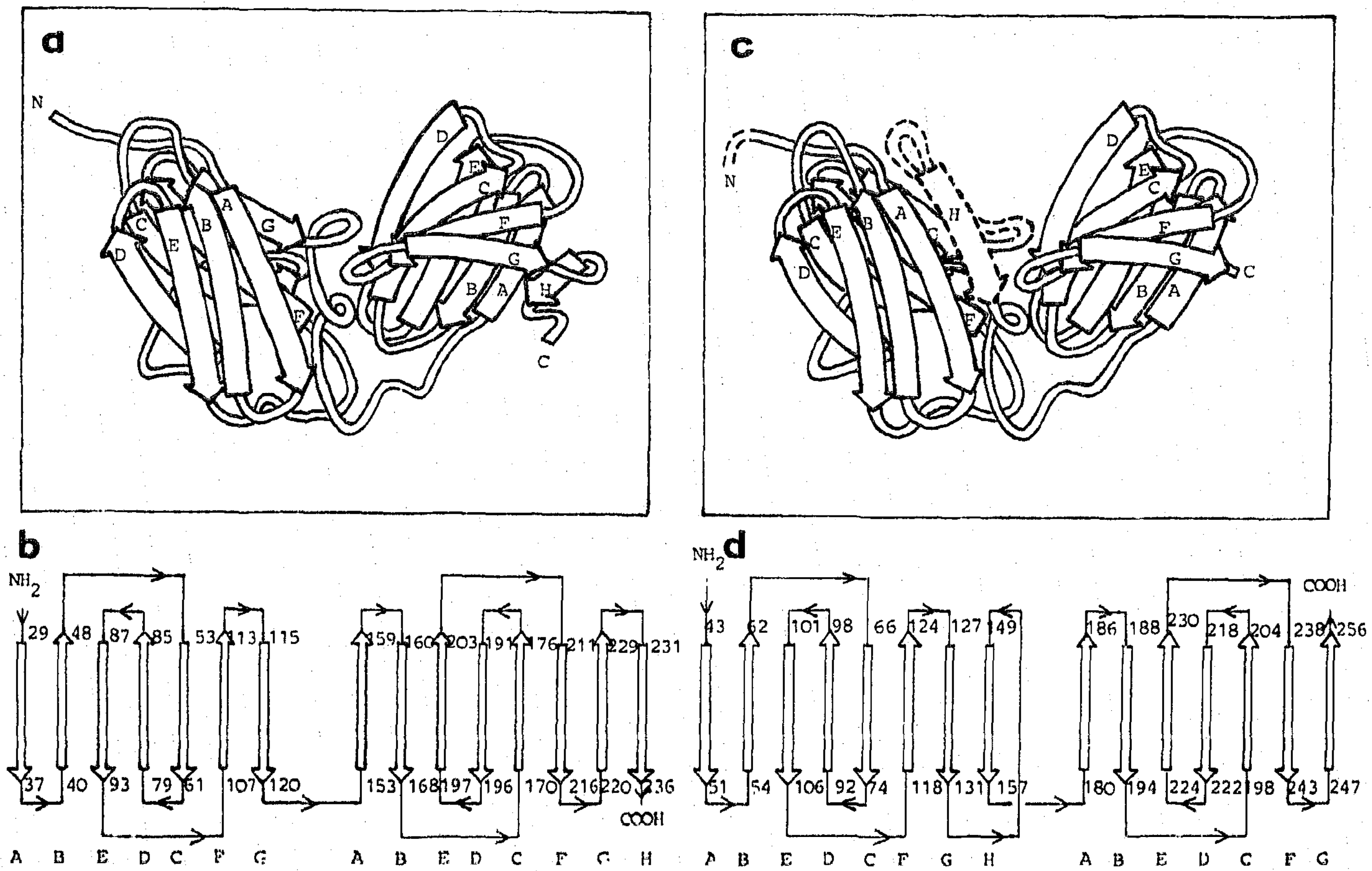

Fig. 4. Tertiary (a) and secondary (b) structures of the chaperone protein PapD of *E. coli* [8] in comparison with the tentative tertiary (c) and secondary structures (d) of the Caf1M protein of *Y. pestis*. Reconstruction of the Caf1M conformation was carried out using as a basis the tertiary structure of PapD, taking into account primary structures of the proteins compared (Fig. 3).

quences. The significant homology of Caf1M with the chaperone protein PapD of uropathogenic *E. coli* [7,8] was found. The PapD is suggested to function in P-pilus biogenesis as a periplasmic transport protein [7]. Fig. 3 shows the amino acid sequence alignment of the proteins compared constructed on the basis of the PapD secondary structure [8]. In comparison with the protein PapD, the Caf1M sequence includes an extended insertion at the beginning of the mature protein and also a rather long insertion between the domains. At the same time the Caf1M lacks the sequence corresponding to the β-strand H in the second domain of PapD. In the Caf1M there is no possibility of formation of a disulfide bridge in the second domain, but this may take place in the first domain between Cys-121 and Cys-160. Taking into account this disulfide bridge and due to the fact that the β-sheet conformation is predicted for the Caf1M sequence 149–157 by the three methods [9–11], we tried to reconstruct the Caf1M spatial structure (Fig. 4c) taking as a basis the known steric structure of PapD [8] (Fig. 4a). As seen, the three-stranded β-sheet of the first domain making up a tentative binding site of the pili subunits in PapD [8] is converted into the four-stranded β-sheet without affecting the general protein architecture and conformation of the hinge region between the domains. In the central part of the putative site of binding of the pili subunits by the PapD protein the solvent-exposed hydrophobic residues V31, L124, I175, I215 were replaced by hydrophobic residues I45, V153, I202, I242 in the protein Caf1M. At the same time the loops forming boundaries of the tentative binding site in the proteins compared significantly differ in their amino acid sequences, which seems to determine specificity of binding of the pili subunits in the case of PapD and F1 antigen in the case of Caf1M.

The data presented above and structural resemblance between Caf1M and PapD allow one to suggest similar functions of the two proteins. By analogy to the PapD protein [7,8,12], the Caf1M seems to be a periplasmic transport protein and act as a chaperone to promote the posttranslational folding and effective export of components of the *Y. pestis* capsule to the bacterial surface.

Acknowledgements: We are grateful to G.A. Zav'yalova for help in the computer estimations and Dr. P.A. Cherepanov for critical discussion of the work.

REFERENCES

[1] Protsenko, O.A., Anisimov, P.I., Mozharov, O.T., Konnov, N.P., Popov, Y.A. and Kokushikin, A.M. (1983) Genetika *(Moscow) 19, 1081–1091.*
[2] Meyer, K.F., Hightower, J.A. and McCrumb, F.R. (1973) J. Infect. Dis. Suppl. 129, 41–45.
[3] Galyov, E.E., Smirnov, O.Yu., Karlishev, A.V., Volkovoy, K.I., Denesyuk, A.I., Nazimov, I.V., Rubtsov, K.S., Abramov, V.M., Dalvadyanz, S.M. and Zav'yalov, V.P. *(1990)* FEBS Lett. 277, 230–232.
[4] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[5] Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103–119.
[6] Rosenberg, M. and Court, D. (1979) Annu. Rev. Genet. 13, 319–353.
[7] Lindberg, F., Tennent, J.M., Hultgren, S.J., Lund, B. and Normark, S. (1989) J. Bacteriol. 171, 6052–6058.
[8] Holmgren, A. and Branden, C.-I. (1989) Nature 342, 248–251.
[9] Zav'yalov, V.P. (1983) Biochim. Biophys. Acta 310, 70–75.
[10] Novotny, J. and Auffray, C. (1984) Nucleic Acids Res. 12, 243–255.
[11] Ptitsyn, O.B., Finkelstein, A.V. and Murzin, A.G. (1985) FEBS Lett. 186, 143–148.
[12] Hultgren, S.J., Lindberg, F., Magnusson, G., Kihlberg, J., Tennent, J.M. and Normark, S. (1989) Proc. Natl. Acad. Sci. USA 86, 4357–4361.