Volume 277, number 1,2, 230-232 FEBS 09271 December 1990

Nucleotide sequence of the *Yersinia pestis* gene encoding F1 antigen and the primary structure of the protein

Putative T and B cell epitopes

E.E. Galyov[1], O.Yu. Smirnov[1], A.V. Karlishev[2], K.I. Volkovoy[2], A.I. Denesyuk[1], I.V. Nazimov[3], K.S. Rubtsov[3], V.M. Abramov[1], S.M. Dalvadyanz[4] and V.P. Zav'yalov[1]

[1]*Institute of Immunology, USSR Ministry of Medical Industry, 142380 Lyubuchany, Moscow Region,* [2]*All-Union Research Institute of Applied Microbiology, USSR Ministry of Medical Industry, 142279 Obolensk, Moscow Region,* [3]*Shemyakin Institute of Bioorganic Chemistry, USSR Academy of Sciences, Miklukho-Maklay Street 16/10, Moscow* and [4]*All-Union Antiplague Institute 'Microb', USSR Ministry of Health, Universitetskaya Street 46, 410071 Saratov, USSR*

Received 22 October 1990

The plasmid-located gene *caf1* encoding the capsular antigen fraction 1 (F1) of *Yersinia pestis* was cloned and sequenced. The gene codes for a 170 amino acid peptide with a deduced M_r of 17.6 kDa. The signal peptide sequence was highly homologous to the *E. coli* consensus signal sequence. The F1 was assumed to have β-sheet structure for the most part. The region located between amino acids 100 and 150 was suggested to contain putative antigenic determinants and to stimulate T cells.

Capsular antigen; Nucleotide sequence; Signal sequence; Antigenic determinant; *Yersinia pestis*

1. INTRODUCTION

The ultimate goal of infectious disease research is their prevention. Vaccination is one of the most effective ways in which that goal can be attained. It is necessary to know the gene structure and putative immunogenic surface structures of antigens to create recombinant vaccines.

More than 10 antigens have been isolated from *Yersinia pestis*. The capsular antigen fraction 1 (F1) was shown to be a highly protective antigen among such thermolabile antigens as D, F1, T, V and W [1]. Some properties of the F1 structure have been studied recently [2], but the nucleotide and amino acid sequences have been unknown so far. Here, we report the cloning and sequencing of the *Y. pestis caf1* gene coding for the F1, and the predicted secondary structure with potential antigenic determinants.

2. MATERIALS AND METHODS

2.1. *Bacterial strains, plasmids and DNA manipulations*

The *E. coli* strains LE392 and HB101 were used as transistent hosts for cosmid pHC79 [3], and pUC18 or 19 [4], respectively. *Y. pestis* F1 positive vaccine strain EV was obtained from the Culture Collection, All-Union Antiplague Institute 'Microb', USSR. Cultures were grown overnight while shaking at 37°C in liquid LB or on solid medium supplemented with the relevant antibiotics for plasmid selection.

Correspondence address: O.Yu. Smirnov, Institute of Immunology, 142380 Lyubuchany, Chekhov District, Moscow Region, USSR

Isolation of plasmid DNA by alkali-lysis method, DNA cloning and Maxam-Gilbert sequencing were performed essentially as described by Maniatis et al. [5].

2.2. *Construction of a gene library, screening, and subcloning*

The *Y. pestis* plasmid pFra DNA (about 110 kb in size) was partially digested with *Eco*RI, ligated with *Eco*RI-digested cosmid pHC79 and packaged in vitro. The library was amplified in *E. coli* LE392 and Ap^RTc^R colonies selected were further screened for F1 production by enzyme immunoassay. The isolated cosmid p153 containing a 40-kb fragment of pFra DNA was then digested with *Eco*RI and an 8.6-kb fragment was cloned into pHC79. The resulting cosmid pFS2 was digested with *Sal*I and *Hin*dIII and a 4.5-kb fragment was cloned into pUC19. The plasmid pFS2-13 generated was used for gene sequencing. The 1.0-kb *Alu*I fragment of pFS2-13 was cloned into the *Sma*I-digested pUC18 (plasmid pF18L) and sequenced.

2.3. *Protein sequencing*

The F1 protein was isolated from culture medium and purified by polyacrylamide gel electrophoresis as described [2]. The N-terminus of the mature protein was identified by a PTH-amino acid analyzer (Model 120A, Applied Biosystems).

2.4. *Secondary structure and antigenic determinant analysis*

Secondary structure of the F1 protein was predicted from the amino acid composition as described in [6], and by the amino acid sequence analyses [7,8]. Antigenic regions were predicted according to Hopp and Woods [9] or Karplus and Schulz [10]. The T cell antigenicity was predicted as reported in [11].

3. RESULTS AND DISCUSSION

3.1. *Sequence of the* caf1 *gene*

Subcloning experiments from cosmid p153 allowed us to construct plasmid pFS2-13 carrying the *caf1* gene (Fig. 1). Based on the results of transposon mutagenesis

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

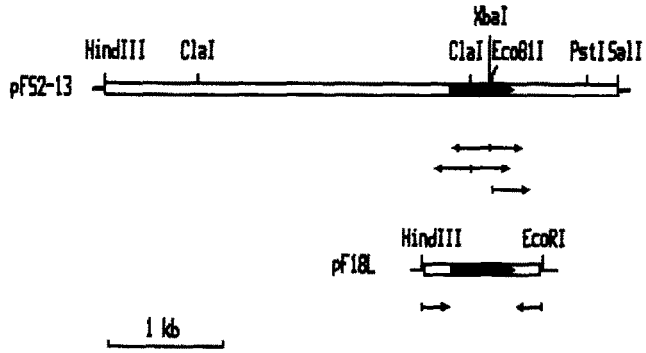

Fig. 1. Sequencing strategy for the *caf1* gene. Overlapping restriction fragments of pFS2-13 and pF18L were sequenced as indicated by arrows. Open boxes indicate cloned DNA fragments. The length of the gene and the direction of transcription are indicated by a hatched arrow.

(data not shown), the F1 structural gene was localized near *Cla*I and *Pst*I sites. The sequencing strategy for *caf1* is shown in Fig. 1. Nucleotide sequencing revealed that the *caf1* gene was located on the overlapping 1.0-kb *Alu*I–*Alu*I fragment (Fig. 2) included in pF18L (Fig. 1). The coding region for the F1 protein was 510 bp long. Although the 5′-flanking region had putative promoter

```
      AluI
  1 AGCTCTGGCATTGTCGGAGATAATAGCGGTGTCTATTTGACTGGACTACCTAAAAAATCA
 61 AAAATACTTGTTAAGTGGGGGAGAGATAAAAATCAATCATGTTCATCTAATGTAGTTCTA
121 CCAGAAAAAACGGATATTTCTGGTGCTTATAGGTTATCCACAACCTGCATCTTAAATAAC
                  -35                          -10
181 TGAAACGGATGTTTATTTCAAACAGGACACAAGCCCTCTCTACGAATTTGTTCGTGGATT
241 GGATTATTCGATAGAGGTAATATATGAAAAAAATCAGTTCCGTTATCGCCATTGCATTAT
  1                          M  K  K  I  S  S  V  I  A  I  A  L

301 TTGGAACTATTGCAACTGCTAATGCGGCAGATTTAACTGCAAGCACCACTGCAACGGCAA
 13 F  G  T  I  A  T  A  N  A  A  D  L  T  A  S  T  T  A  T  A
                           ↑
361 CTCTTGTTGAACCAGCCCGCATCACTCTTACATATAAGGAAGGCGCTCCAATTACAATTA
 33 T  L  V  E  P  A  R  I  T  L  T  Y  K  E  G  A  P  I  T  I
                  ClaI.
421 TGGACAATGGAAACATCGATACAGAATTACTTGTTGGTACGCTTACTCTTGGCGGCTATA
 53 M  D  N  G  N  I  D  T  E  L  L  V  G  T  L  T  L  G  G  Y
                           HpaI
481 AAACAGGAACCACTAGCACATCTGTTAACTTTACAGATGCCGCGGGTGATCCCATGTACT
 73 K  T  G  T  T  S  T  S  V  N  F  T  D  A  A  G  D  P  M  Y

541 TAACATTTACTTCTCAGGATGGAAATAACCACCAATTCACTACAAAAGTGATTGGCAAGG
 93 L  T  F  T  S  Q  D  G  N  N  H  Q  F  T  T  K  V  I  G  K
      XbaI  .     EcoRV.   Eco81I.
601 ATTCTAGAGATTTTGATATCTCTCCTAAGGTAAACGGTGAGAACCTTGTGGGGGATGACG
113 D  S  R  D  F  D  I  S  P  K  V  N  G  E  N  L  V  G  D  D

661 TCGTCTTGGCTACGGGCAGCCAGGATTTCTTTGTTCGCTCAATTGGTTCCAAAGGCGGTA
133 V  V  L  A  T  G  S  Q  D  F  F  V  R  S  I  G  S  K  G  G

721 AACTTGCAGCAGGTAAATACACTGATGCTGTAACCGTAACCGTATCTAACCAATAATCCA
153 K  L  A  A  G  K  Y  T  D  A  V  T  V  T  V  S  N  Q  End

781 TATAGATAATAGATAAAGGAGGGCTATTATGCCCTCCTTTAATATTTATGAATTATCCTA
841 CTTTGAGCCTAACCCTCGCTTTTCTTAATCACGGCATTGATAGCAAGACTGACAAATTTA
901 TGTGAAGATCAATGTTAGGAACTAATGCAGAAAGCCACGCCCTCAATAGATTTCACATAA
      AluI
961 TACACTATTAGCTAAGAATAGAGAGCGCGAAGCAATATAATAGGTTCATATTTATAACTC
1021 TCACCTTAA
```

Fig. 2. Nucleotide and deduced amino acid sequences of the *Yersinia pestis caf1* gene. The presumed promoter (−35, −10), Shine-Dalgarno and terminator sequences are underlined. An arrow indicates the cleavage site.

sequences similar to the *E. coli* consensus sequences (−35, −10) [12], F1 was not expressed in *E. coli* carrying pF18L. In the 3′-non-coding region, the revealed inverted repeat structure resembled the Rho-independent transcriptional termination signal of *E. coli* [13] but the T-stretch following this terminator sequence was not clearly identified. Thus, the *caf1* expression can be regulated in a complex manner and the gene may have a distinct transcriptional termination of its own.

A putative ribosome binding site GAGGT was localized at 5 bp upstream from the ATG codon. An open reading frame from nucleotides 264 to 773 codes for a 170-aa polypeptide with a deduced M_r of 17.6 kDa. A potential signal peptide sequence has been identified with a typical cleavage site (/):
Ala-Thr-Ala-Asn-Ala/Ala-Asp which resembled the procaryotic consensus signal sequence [14]: Ala-$^{Ser}_{Ala}$-Ala-$^{Phe}_{His}$-Ala/Ala-$^{Asp}_{Glu}$. This yields a leader peptide of 21 residues and a secreted F1 of 149 residues with a deduced M_r of 15.5 kDa and pI 4.3. The amino acid composition of the F1 as deduced from the nucleotide sequence shows agreement with the data reported [2,15], except that the F1 sequence does not contain Cys and Trp.

3.2. *Secondary structure and antigenically active epitopes*

Fig. 3A demonstrates that the F1 protein has mainly β-sheet structure (at least 50%) and may have 3 or 4 short strips of α-helix. Hydrophilicity and antigenicity profiles are presented in Fig. 3B. The region between residues 105/120 was supposed to be located on the protein surface, since it contains β-turns, and furthermore, may constitute antigenic determinants. The second putative B cell epitope was assumed to be located in the C-terminus (149/161) whereas there was no clearly

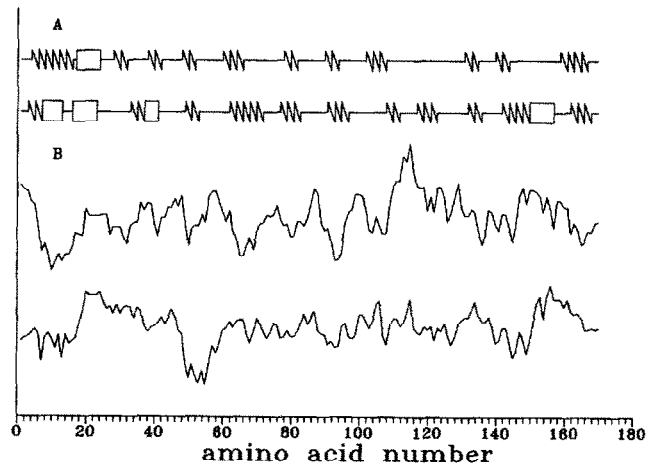

Fig. 3. Analysis of the F1 sequence. (A) The protein secondary structure determined according to [7] (top) and [8] (bottom). The boxes indicate α-helix and the zig-zags indicate β-sheet structure. (B) Hydrophilicity (top) and antigenicity (bottom) profiles of the F1 protein determined according to [9] and [10], respectively.

identified hydrophilic maximum in this region. DeLisi and Berzofsky [17] concluded that amphipathic structures (peptides with non-random secondary structures with opposing hydrophobic and hydrophilic surfaces) correlated with T cell antigenicity. Predicted T cell epitopes were localized between residues 106/115, 127/135, and 137/148.

Thus, the C-terminal region of the F1 antigen (100/150 residues) was supposed to be of great importance for the immunogenicity of the F1.

Acknowledgements: We are grateful to G.A. Zav'yalova for her help in preparing the figures.

REFERENCES

[1] Englesberg, E.V. and Lewy, J. (1954) J. Bact. 67, 438-449.
[2] Anisimov, P.I., Serdobintsev, L.N., Ivanov, Yu.V., Taranenko, T.M. and Naumov, A.V. (1987) Mol. Genet. Microbiol. Virol. (USSR), 2, 24-27.
[3] Hohn, B. and Collins, J. (1980) Gene 11, 291-298.
[4] Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103-119.
[5] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[6] Zav'yalov, V.P. and Denesyuk, A.I. (1982) Immunol. Lett. 4, 7-14.
[7] Novotny, J. and Anffray, C. (1984) Nucleic Acids Res. 12, 243-255.
[8] Zav'yalov, V.P. (1973) Biochim. Biophys. Acta 310, 70-75.
[9] Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824-3828.
[10] Karplus, P.A. and Schulz, G.E. (1985) Naturwissenschaften 72, 212-218.
[11] Rothbard, J. and Taylor, W. (1988) EMBO J. 138, 99-106.
[12] Rosenberg, M. and Court, D. (1979) Annu. Rev. Genet. 13, 319-353.
[13] Von Hippel, P.H., Bear, D.G., Morgan, W.D. and McSwiggen, J.A. (1984) Annu. Rev. Biochem. 53, 389-446.
[14] Von Heijne, G. (1986) Nucleic Acids Res. 14, 4683-4690.
[15] Bennet, L.G. and Tornabene, T.G. (1974) J. Bact. 117, 48-55.
[16] DeLisi, C. and Berzofsky, J.A. (1985) Proc. Natl. Acad. Sci. USA 82, 7048-7052.