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. Galyov¹, O.Yu. Smirnov¹, A.V. Katlisiev², K.I. Volkovoy², A.I. Denesyuk¹, I.V. Nazimov³, K.S. Rubtsov³, V.M. Abramov¹, S.M. Dalvadyanz¹ and V.P. Zav’yalov¹

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

Received 22 October 1990

The plasmid-located gene *cafl* 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 M₉ 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 cafl* 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 transient 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.

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 EcoRI, ligated with EcoRI-digested cosmid pHC79 and packaged in vitro. The library was amplified in *E. coli* LE392 and Ap°Tc° 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 EcoRI and an 8.6-kb fragment was cloned into pHC79. The resulting cosmids pFS2 was digested with *SalI* and *HindIII* and a 4.5-kb fragment was cloned into pUC19. The plasmid pFS2-13 generated was used for gene sequencing. The 1.0-kb *AluI* fragment of pFS2-13 was cloned into the *SalI*-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 *cafl* gene

Subcloning experiments from cosmid p153 allowed us to construct plasmid pFS2-13 carrying the *cafl* gene (Fig. 1). Based on the results of transposon mutagenesis
Fig. 1. Sequencing strategy for the \textit{caf} \textit{fl} gene. Overlapping restriction fragments of pFSZ-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 \textit{F1} structural gene was localized near \textit{CiaI} and \textit{PsI I} sites. The sequencing strategy for \textit{caf} \textit{fl} is shown in Fig. 1. Nucleotide sequencing revealed that the \textit{caf} \textit{fl} gene was located on the overlapping 1.0-kb \textit{AluI}–\textit{AluI} fragment (Fig. 2) included in pF18L (Fig. 1). The coding region for the \textit{F1} protein was 510 bp long. Although the 5′-flanking region had putative promoter sequences similar to the \textit{E. coli} consensus sequences (-35, -10) [12], \textit{F1} was not expressed in \textit{E. coli} carrying pF18L. In the 3′-non-coding region, the revealed inverted repeat structure resembled the Rho-independent transcriptional termination signal of \textit{E. coli} [13] but the T-stretch following this terminator sequence was not clearly identified. Thus, the \textit{caf} \textit{fl} 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 \textit{Mr} 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–Thr–Ala–Asp. This yields a leader peptide of 21 residues and a secreted \textit{F1} of 149 residues with a deduced \textit{Mr}, of 15.5 kDa and pl 4.3. The amino acid composition of the \textit{F1} as deduced from the nucleotide sequence shows agreement with the data reported [2,15], except that the \textit{F1} sequence does not contain Cys and Trp.

3.2. Secondary structure and antigenically active epitopes

Fig. 3A demonstrates that the \textit{F1} protein has mainly \textit{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
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