Volume 305, number 1, 37–40 FEBS 11185 June 1992

Caf1R gene and its role in the regulation of capsule formation of *Y. pestis*

A.V. Karlyshev, E.E. Galyov, V.M. Abramov and V.P. Zav'yalov

Institute of Immunology, State Concern 'Biopreparation', 142380 Lyubuchany, Moscow Region, Russia

Received 30 March 1992

A new transcription unit of the *f1* gene cluster was found. The DNA sequencing revealed one long open reading frame. Deletion and frame shift mutation analyses have demonstrated the importance of a corresponding gene product for the F1 antigen biosynthesis. A homology of the deduced amino acid sequence with that of AraC family DNA-binding regulators was shown. A potential regulatory DNA region is discussed.

Gene regulation; Capsule antigen; Nucleotide sequence; *Y. pestis*

1. INTRODUCTION

Previously we have shown that at least three gene products are required for the formation of *Y. pestis* capsule: the *caf1* gene product (F1 antigen), which is a major capsule compound [1], the *caf1M* gene product, which plays a chaperone-like function during transport of the F1 subunit [2], and the *caf1A* gene product, which is important for anchoring of the capsule to the cell surface [3]. The genes are assembled into two transcription units oriented in the same direction: one for the *caf1* gene and the other for the *caf1M* and *caf1A* genes. Later another transcription unit, lying in the opposite orientation to these transcription units, was found. In the present study we show that this region contains a gene which is very important for the regulation of capsule biogenesis.

2. MATERIALS AND METHODS

The recombinant plasmids, containing the whole *f1* operon (pFS2) or its fragment (p12R) have been described previously [1,2]. The plasmids, pFSRH7 and pO6, were obtained by cloning of 1,110 and 1,054 bp fragments of pFS2 into pUC19. DNA cloning and sequencing by the Maxam–Gilbert and Sanger–Coulson methods were performed essentially as described by Maniatis et al. [4]. DNA and protein sequence analyses were made using the GENEPRO software package from Riverside Scientific Enterprises and PCGENE package from Genofit. Anti-plague agglutinating antiserum and erythrocytes conjugated with polyclonal anti-F1 antibodies were obtained from the Anti-Plague Institute 'Microb' (Saratov, Russia). The internal primers were obtained from Immune-Biotech (Lyubuchany, Russia).

3. RESULTS AND DISCUSSION

The only long open reading frame was found 1.6 kb upstream from the *caf1M* gene (Fig. 1). A potential gene product, designated Caf1R, was found to be 301 amino acid residues long with M_r = 36,052. A PIR'89 sequence database search revealed a homology between Caf1R and proteins of the AraC family (see references in [5]) (Fig. 2) which are known to be DNA-binding transcription regulators. These proteins are proposed to consist of helix-turn-helix (or homeodomain in eukaryotic analogs) regions containing positively charged amino acid residues. Recently the crystal structure of a homeodomain region of the Matα2–DNA complex has been determined [6]. As shown in Fig. 2, the Matα2 homeodomain displays about 20% primary structure homology with the corresponding region of VirF-Y., with 8 of 12 conserved hydrophobic positions in the AraC family proteins coincident with those of the Matα2 homeodomain, and the sole hyperconserved Arg residue of the AraC proteins coincident with Arg of the Matα2 protein. This Arg is a conserved residue in homeodomains of eukaryotic regulatory proteins, and is shown to be a component of the third homeodomain helix directly interacting with a large DNA groove [6].

The Caf1R protein seems to be highly basic, like other DNA-binding proteins. An isoelectric point of Caf1R was calculated to be 9.5.

A regulatory area of AraC was shown to be located in the C-terminal region of the protein [7]. We found that the plasmid, p12R, codes for a truncated form of Caf1R containing only 81 N-terminal amino acid residues with an additional tail of 19 amino acids encoded by the vector DNA (Figs. 1 and 3). A predicted isoelectric point of this Caf1R derivative was found to be 10.5.

Bacterial cells harboring p12R can produce very large amounts of the F1 antigen and are able to form a capsule. In order to prove that Caf1R and its N-terminal peptide possess positive regulatory activity we have introduced a frame-shift mutation into the plasmids, pFS2 and p12R, by digestion of DNA with *Xho*I, followed by T7 DNA polymerase filling-in of cohesive

Correspondence address: A.V. Karlyshev, Institute of Immunology, 142380 Lyubuchany, Chekhov District, Moscow Region, Russia.

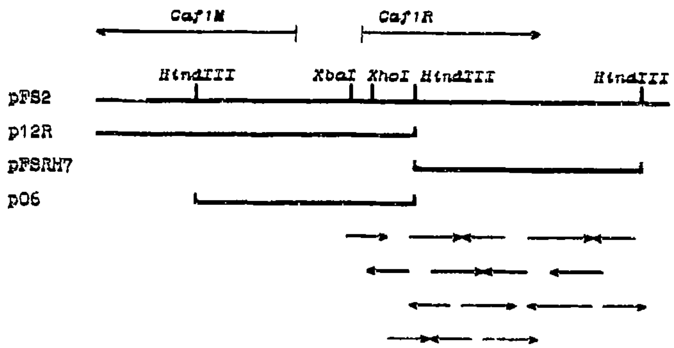

Fig. 1. Restriction map and sequencing strategy of the *caf1R* gene region. Bold bars indicate the cloned fragments. Arrows above the restriction sites indicate the open reading frames. Short arrows below indicate lengths and directions of the sequences determined.

ends and blunt-end ligating. The *E. coli* HB101 clones, containing plasmids lacking *Xho*I sites, failed to produce a detectable amount of the F1 antigen at 37°C. This frame-shift mutation leads to a truncated protein containing only the 18 N-terminal amino acids of Caf1R. These data clearly demonstrate that the regulatory region is included in the N-terminal part of Caf1R 81 amino acid-long region.

We can speculate about a possible Caf1R-determined regulation mechanism involved in the capsule formation, proceeding from the data on the other positive regulators sharing homology with Caf1R. The regulatory regions were shown to be located in the closest vicinity to the genes encoding such regulators. Therefore, the most probable regulatory region affected by

```
                      +           +       + +  +       +     +     +
Caf1R       1 MLKQMTVNSI IQYIEENLES -KFINIDCLV LYSGFSRRYL QISFKEYVG-
Soxs        1 MsgQkiiqdl IanIdEhidq -p-lNIDvva kkSGySknYL QrmFrtvtG-
AraC-E.c  173 ppndnrVrea cQYIsdhLad -snfdIasva qhvclSpsrL shlFrqqlG-
RhaR      202 tssetlldkl ItrlaasLkS -p-falDkfc deascSeRvL rqqFrqqtG-
RhaS      167 ensasrlNll lanlEdhpad -e-vNnDava dqfslSlRtL grqlKqqtG-
MelR      187 rgaQfyVsqm lgfIaENydq -a-ltIndva sevklnanYa mqiFqrvmq-
VirF-Y.   165 gLgnrpeerl qkfmEENylq -g-nklskfa refGhqltff kelFqtvyG-
                  .:   :    .:  ::        :          .           :
Matα2       1    tkpy-rg hrftkENvri lesnfaknie npy-ldtkgL enlmKntsls

                                    1                      2
                 +     o+      *           +   +     ++         *    +
Caf1R         -MPIGTYIRV RRASRAAALL RLTRLTIIEI SAKLFYDSQQ TFTREFKKIF
Soxs          -qtlGdYIRq RRlllAAveL RtTerpIfdI amdLgYvSQQ TFsRvFrrqF
AraC-E.c      -isvlswred qRiSqAklLL stTRmpIatv qrnvgfDdQl yFsRvFKKct
RhaR          -MtInqYlRq vRvchAqyLL qhsRLlIsdI Stecgfedsn yFsvvFtret
RhaS          -ltpqrYlnr lRlmkARqLL RhseasvtdI ayrcgfsdsn gFstlFrreF
MelR          -ltmkqYIta mRinhvrALL sdTdksIldI altagfrSss rFystFgKyv
VirF-Y.       -isprawIse RRilyAgqLL lngkmsIvdI ameagfsSQs yFTqsyrrrF
               :  . :.:  ::       .
Matα2         riqIknnvsn RRr-kektI  63

                    3                                       α%     β%
                  *    +
Caf1R         GYTPRQYRMI PFW 110                            57     25
Soxs          drTPsdYRhr l   107  -- 3 10^-24               48      1
AraC-E.c      GasPsefRaq cee 282  -- 7 10^-13               47      8
RhaR          GhTPsQhRhl nsq 310  -- 2 10^-11               53     16
RhaS          nnsPRdiRqg rdg 275  -- 5 10^-6                64      3
MelR          GhsPqQYRkl sqq 295  -- 6 10^-9                58     27
VirF-Y.       GcTPsQaRlt kia 273  -- 4 10^-8                74     17

Average                                                     57.3   13.9
```

Fig. 2. Amino acid sequence alignment of the Caf1R protein with the AraC and Matα2 protein. Numbers indicate positions in the corresponding proteins. A possibility of random coincidence with the Caf1R is indicated against each sequence of the AraC proteins [11]. The predicted alpha-helix and beta-structure content is also indicated [12]. The hyperconserved hydrophobic and hydrophilic residues are indicated by asterisks and circles, respectively. Conserved hydrophobic residues are marked with crosses. The amino acid residues identical to those of Caf1R are indicated by capital letters. The identical and conserved residues in VirF-Y. and Matα2 sequences are indicated by colons and dots, respectively. Alpha-helical regions in the homeodomain of Matα2, revealed by the X-ray crystal analysis [6], are indicated by open boxes.

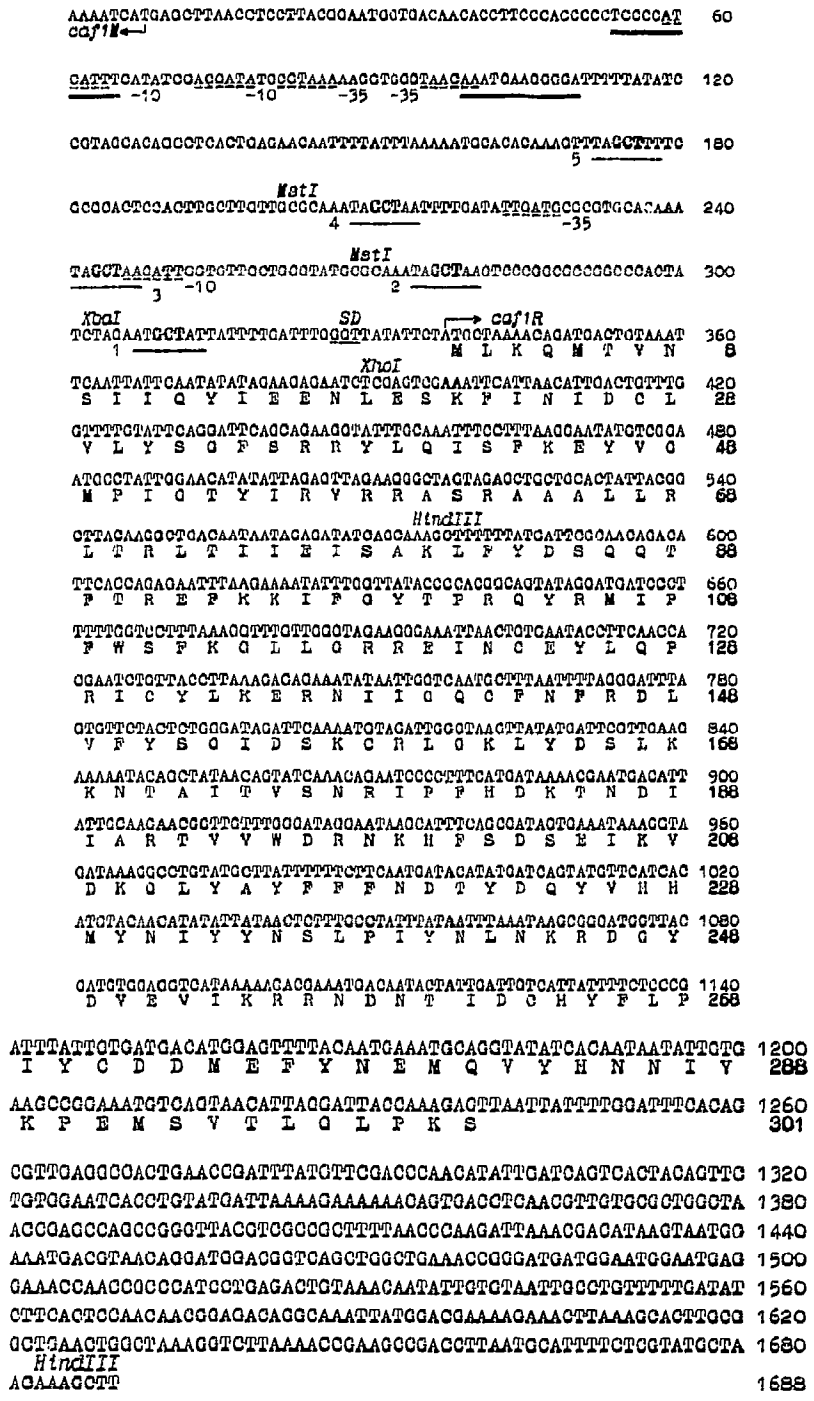

Fig. 3. Nucleotide sequence of the *caf1R* gene, and the deduced amino acid sequence. The potential upstream regulatory region (partially reported previously [2]) is also shown. Five conserved regions are marked with thin lines. Solid bars indicate the inverted repeats. Dashed lines indicate the consensus promoter regions. A part of the potential 16 S ribosomal recognition sequence is underlined and marked SD.

the Caf1R gene product should lie in the *caf1R–caf1M* intergenic region of 331 bp long. As shown in Fig. 3, this region is characterized by some interesting features. First, the conserved elements with the consensus $^{AT}_{TA}GCT^{AA}_{TT}$ are periodically repeated five times in this region at equal distances from each other. The distance between the elements was found to be 33 bp for repeats 1–2, 2–3, and 3–4, and 36 bp for repeat 4–5. This corresponds to 3 turns of a double DNA helix such that most of these sequences are to be oriented to the same DNA side. These repeated regions seem to be involved in the protein–DNA interaction. Second, there are two in-

verted repeats in the *caf1M* promoter region which might also be necessary for the regulation. We cannot exclude the possibility that proteins other than Caf1R may be involved in the regulation process. An involvement of a histone-like protein in the transcription regulation of two divergent promoters was shown in a *pap* gene cluster [8]. The *pap* regulatory region was shown to contain a pair of inverted repeats like the *caf1R–caf1M* intergenic region. Finally, it should be noted that the synthesis of the Caf1R protein is a prerequisite for F1 antigen production and hence formation of *Y. pestis* capsule. The Caf1R protein seems to induce the *caf1M* gene transcription leading to the production of Caf1M and Caf1A proteins involved in capsule formation. The Caf1R protein may be involved in signal transduction. Such signals may be temperature [9] and/or other environmental factors, such as Ca^{2+} and Mg^{2+} ions concentration [10]. Additional experiments should be carried out to elucidate the role of the C-terminal part of the Caf1R protein.

Acknowledgements: We are grateful to G.A. Zav'yalova for help in computer estimations. This work was carried out under the State Program 'Priority directions in genetics'.

REFERENCES

[1] Galyov, E.E., Smirnov, O.Yu., Karlyshev, 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.
[2] Galyov, E.E., Karlyshev, A.V., Chernovskaya, T.V., Dolgikh, D.A., Smirnov, O.Yu., Volkovoy, K.I., Abramov, V.M. and Zav'yalov, V.P. (1991) FEBS Lett. 286, 79–82.
[3] Karlyshev, A.V., Galyov, E.E., Smirnov, O.Yu., Guzayev, A.P., Abramov, V.M. and Zav'yalov, V.P. (1992) FEBS Lett. 297, 77–80.
[4] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[5] Wu, L. and Weiss, B. (1991) J. Bacteriol. 173, 2864–2871.
[6] Wolberger, C., Vershon, A.K., Lin, B., Jonson, A.D. and Pabo, C.O. (1991) Cell 67, 517–528.
[7] Menon, K.P. and Lee, N.L. (1990) Proc. Natl. Acad. Sci. USA 87, 3708–3712.
[8] Goransson, M., Sonden, B., Nilsson, P., Dagberg, B., Forsman, K., Emanuelssen, K. and Uhlin, B.E. (1990) Nature 344, 682–685.
[9] Chen, T.H. and Elsberg, S.S. (1977) Infection Immunity 15, 972–977.
[10] Chen, T.H., Elsberg, S.S., Boyles, J. and Velez, M. (1975) Infection Immunity 11, 1382–1390.
[11] Finkelstein, A.V. (1977) FEBS Lett. 82, 169–171.
[12] Zav'yalov, V.P. (1973) Biochim. Biophys. Acta 310, 70–75.