FEBS Letters 366 (1995) 46–48 FEBS 15554

The nucleotide sequence of the bacteriophage T5 *ltf* gene

Alexander V. Kaliman*, Vladimir E. Kulshin, Michael G. Shlyapnikov, Vladimir N. Ksenzenko, Valentine M. Kryukov

Institute of Biochemistry and Physiology of Microorganisms, Russian Academy of Sciences, Pushchino, Moscow Region 142292, Russian Federation

Received 10 April 1995

Abstract The nucleotide sequence of the bacteriophage T5 *Bgl*II–*Bam*HI fragment (4,835 bp in length) known to carry a gene encoding the LTF protein which forms the phage L-shaped tail fibers was determined. It was shown to contain an open reading frame for 1,396 amino acid residues that corresponds to a protein of 147.8 kDa. The coding region of *ltf* gene is preceded by a typical Shine–Dalgarno sequence. Downstream from the *ltf* gene there is a strong transcription terminator. Data bank analysis of the LTF protein sequence reveals 55.1% identity to the hypothetical protein ORF 401 of bacteriophage λ in a segment of 118 amino acids overlap.

Key words: DNA sequencing; L-shaped tail fiber; Phage adsorption; Bacteriophage T5

1. Introduction

The adsorption of bacteriophage T5 to the surface of *Escherichia coli* F is characterized by two specific steps: (i) rapid and reversible binding to the polymannose O antigen [1] and (ii) irreversible binding to the FhuA receptor protein in the outer membrane [2]. Binding to the O antigen is mediated by the L-shaped tail fibers (LTF) and accelerates adsorption by a factor of 15 [3]. However, it has been proven nonessential for infection since phages lacking the LTF are viable [4], and *E. coli* strains lacking an appropriate O antigen are infected by bacteriophage T5 [1,3]. The *ltf* gene was mapped within the *Bam*HI-D fragment of T5 DNA [5]. It is situated at the left end of the late T5 gene region near the boundary of the early and late genes, and their transcription is accomplished from different DNA strands in the opposite directions [6]. The strong transcription terminator from the boundary region (67.2% of the T5 DNA length) has been sequenced and characterized in our previous papers [7,8].

This paper deals with sequencing and analysis of the phage T5*Bgl*II–*Bam*HI fragment containing the *ltf* gene.

2. Materials and methods

Phage and plasmid DNA were isolated and purified by standard procedures [9]. Construction and analysis of recombinant plasmids pBR322-T5 and pUC-T5 were performed as in [7,9]. The nucleotide sequence was determined according to Maxam–Gilbert [10] and Sanger [11] methods. The deduced amino acid sequence of the *ltf* gene was compared with protein sequences in the Swiss Prot Database.

3. Results and discussion

To obtain the material for determining the nucleotide sequence, we decided to clone the *Bam*HI-D fragment carrying the *ltf* gene (Fig. 1). Phage DNA was digested with restriction endonuclease *Bam*HI. The restriction fragments were ligated into the *Bam*HI site of plasmid pBR322 and transformed into *E. coli* cells. The *Bam*HI-D fragment (5 kb)-containing plasmid was characterized using restriction enzymes and Southern blot analyses (data not shown). To sequence the LTF-encoding region, corresponding subclones pUC-T5 of smaller restriction fragments from the *Bam*HI-D fragment were obtained and both DNA strands were sequenced.

The nucleotide sequence of the *Bgl*II–*Bam*HI fragment of T5

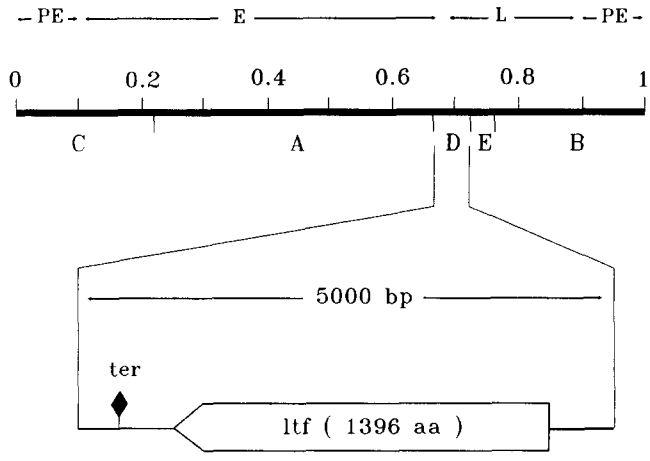


Fig. 1. Location of the *ltf* gene on the bacteriophage T5 genome. The upper line shows pre-early (PE), early (E) and late (L) regions and the *Bam*HI restriction endonuclease map of the T5 genome [6]. The horizontal arrow indicates the position and direction of *ltf* gene transcription. The transcription terminator situated at the boundary between the phage T5 early and late genes is indicated as 'ter'.

*Corresponding author. Fax: (7) (095) 9233602.
Telex: 412609 MICRO SU.
E-mail: ksenz@ibpm.serpukhov.su

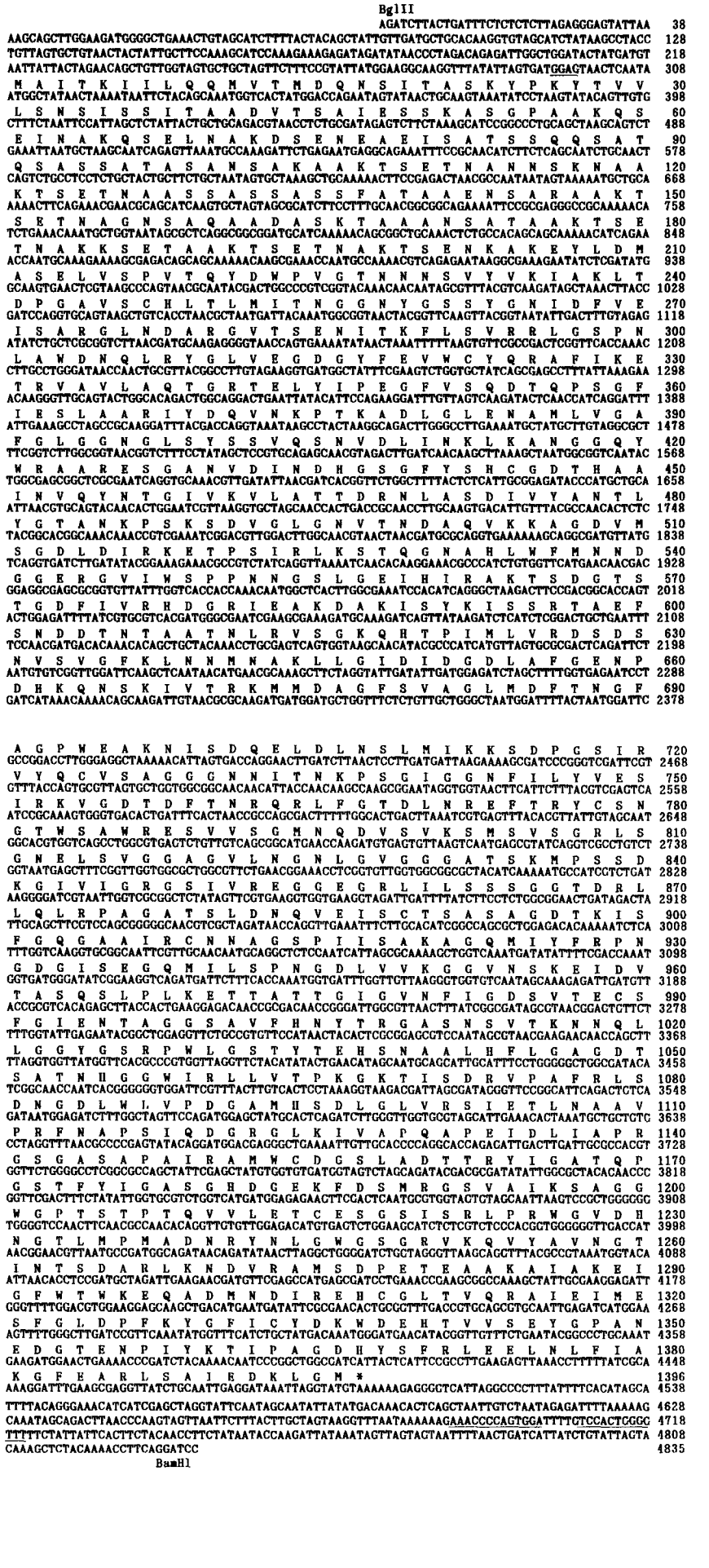
Abbreviations: LTF, L-shaped tail fibers; bp, base pair(s); kb, 1,000 bp; Da, dalton(s); kDa, 1000 Da; aa, amino acid.

The nucleotide sequence data reported in this paper have been submitted to the EMBL Data Library under the Accession Number X69460.

→

Fig. 2. Nucleotide sequence of the *ltf* gene region. The deduced amino acid sequence of the *ltf* gene is presented in a single-letter code above the nucleotide sequence. The stop codon is indicated by asterisk. The ribosome binding site and transcription terminator are underlined.

SSDI 0014-5793(95)00482-3


BglII
BamHI

```
401   (179) KATEAEKSAAAAESSKNAAATSAGAAKTSETNAAASQQSAATSASTAATKASEAATSAR
                ** .. **. . **  *.* ** *********  * ..* ** . **. ** *.. *
LTF   (81)  SATSSQQSATQSASSATASANSAKAAKTSETNANNSKNAAKTSETNAASSASSASSFAT

401   DAVASKEAAKSSETNASSSAGRAASSATAAENSARAAKTSETNARSSETAAERSASAAA (296)
       *  *  ***.*****. **  * .* *** *** *********. *****  * . *
LTF   AAENSARAAKTSETNAGNSAQAADASKTAAANSATAAKTSETNAKKSETAAKTSETNAK (198)
```

Fig. 3. Homology of a part of the phage T5 LTF protein (amino acids 81 to 198) and phage λ hypothetical protein ORF 401 (amino acids 179 to 296). Identical and similar amino acids are indicated by asterisks and dots, respectively. Numbers in parentheses indicate the residues shown.

DNA, 4,835 bp in length, is shown in Fig. 2. A search for open reading frames showed that the major one is found only in one of the DNA strands and corresponds to a protein consisting of 1,396 amino acid residues with a molecular mass of 147.8 kDa. This agrees with 150 kDa reported by Heller and Krauel [5] for a precursor of the LTF protein. The direction of *ltf* gene transcription is from the right to the left (Fig. 1) which is also in accordance with the general transcription map of bacteriophage T5 DNA [12]. The Shine–Dalgarno sequence GGAG (at nucleotides 295–298 in Fig. 2) is located 10 nucleotides from the initiation codon ATG. Upstream from the *ltf* gene there are several promoter-like sequences which show relative homology to the *E. coli* promoters [13]. However, the sequences typical of phage T5 promoters [14,15] have not been found. Downstream from the stop codon TAA there is a strong bidirectional rho-independent terminator of transcription (at nucleotides 4,691–4,721 in Fig. 2) [7,8]. This terminator is situated at the boundary between the early and late phage T5 genes [6].

An analysis of the LTF protein sequence revealed high content of Ser (10.7%), Gly (10.2%), Ala (10.2%) and Thr (7.1%). These are similar to those of the protein gp37 of bacteriophage T4 forming a distal portion of long tail fibers which contact the host receptors [16].

The study of codon usage in a number of genes [17] revealed a non-random pattern of their distribution and the sequence of the *ltf* gene of phage T5 being no exception in this respect. There is an obvious preference for codons ending with T and A residues, which might reflect a high A + T content of T5 DNA. On the whole, 64.1% codons of the *ltf* gene end in T and A residues with the greatest preference for codons ending in T residues (38%). Similar distribution was observed for other T5 genes: the D9 gene encoding DNA polymerase [18], the D15 gene encoding 5′-exonuclease [19], the *oad* gene encoding the receptor-binding protein [20] and the D10-D14 early genes encoding proteins whose functions have not yet been identified precisely [21].

The comparison of the deduced amino acid sequence of the *ltf* gene with protein sequences in the Swiss Prot Database showed that the polypeptide encoded by the *ltf* gene revealed a strong local homology to the hypothetical protein ORF 401 of bacteriophage λ. ORF 401 belongs to the phage λ fiber gene group localized in a nonessential b-region at the distal end of the $P_{R'}$ operon [22]. A particularly high homology exists between amino acids 179 to 296 of protein ORF 401 and amino acids 81 to 198 of the *ltf* gene prodct, 55.1% identity and 68.6% similarity in 118 amino acids overlap (Fig. 3).

The sequence of 1,396 amino acids encoded by the *ltf* gene determines a protein of 147.8 kDa. Heller and Krauel have earlier reported data on the cloning and expression of the phage T5 *ltf* gene [5]. Using different approaches to study polypeptides, they identified the proteins encoded by a recombinant plasmid with the *Bam*HI-D fragment containing the *ltf* gene. Only two proteins of 150 kDa and 125 kDa were identified. The 125 kDa protein corresponded to a polypeptide forming the LTF and could arise from a precursor with a molecular mass of 150 kDa, though it has not yet been detected in vivo [23]. This is in good accord with 147.8 kDa for a polypeptide encoded by the *ltf* gene sequenced in the present work. Thus, the knowledge of the *ltf* gene sequence could help to construct plasmids providing protein overproduction and study LTF protein processing.

References

[1] Heller, K.J. and Braun, V. (1982) J. Virol. 41, 222–227.
[2] Braun, V., Schaller, K. and Wolff, H. (1973) Biochim. Biophys. Acta 323, 87–97.
[3] Heller, K.J. and Braun, V. (1979) J. Bacteriol. 139, 32–38.
[4] Saigo, K. (1978) Virology 85, 422–433.
[5] Heller, K.J. and Krauel, V. (1986) J. Bacteriol. 167, 1071–1073.
[6] McCorquodale, D.J. and Warner, H.R. (1988) in: Viruses, Vol. 1, Bacteriophages (Calendar, R. Ed.) pp. 439–475, Plenum, New York.
[7] Kaliman, A.V., Krutilina, A.I. and Kryukov, V.M. (1987) Mol. Genet. Microbiol. Virol. (Russian) 10, 14–19.
[8] Kaliman, A.V., Kryukov, V.M. and Bayev, A.A. (1988) Nucleic Acids Res. 16, 6230.
[9] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
[10] Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–564.
[11] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[12] von Gabain, A. and Bujard, H. (1977) Mol. Gen. Genet. 157, 301–311.
[13] Hawley, D.K. and McClure, W.R. (1983) Nucleic Acids Res. 11, 2237–2255.
[14] Gentz, R. and Bujard, H. (1985) J. Bacteriol. 164, 70–77.
[15] Decker, K., Krauel, V., Meesmann, A. and Heller, K.J. (1994) Mol. Microbiol. 12, 321–332.
[16] Oliver, D.B. and Crowther, R.A. (1981) J. Mol. Biol. 153, 545–568.
[17] Crantham, R., Gantier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) Nucleic Acids Res. 9, 43–74.
[18] Leavitt, M.C. and Ito, J. (1989) Proc. Natl. Acad. Sci. USA 86, 4465–4469.
[19] Kaliman, A.V., Krutilina, A.I., Kryukov, V.M. and Bayev, A.A. (1986) FEBS Lett. 195, 61–64.
[20] Krauel, V. and Heller, K.J. (1991) J. Bacteriol. 173, 1287–1297.
[21] Kaliman, A.V., Kryukov, V.M. and Bayev, A.A. (1988) Nucleic Acids Res. 16, 10353–10354.
[22] Sanger, F., Coulsson, A.R., Hong, G.F., Hill, D.F. and Petersen, G.B. (1982) J. Mol. Biol. 162, 729–773.
[23] Heller, K.J. (1984) Virology 139, 11–21.