JOURNAL OF VIROLOGY, Apr. 1988, p. 1186–1193
0022-538X/88/041186-08$02.00/0

Vol. 62, No. 4

Nucleotide Sequence of the Tail Sheath Gene of Bacteriophage T4 and Amino Acid Sequence of Its Product

FUMIO ARISAKA,[1]* TATSUSHI NAKAKO,[1] HIDEO TAKAHASHI,[2] AND SHIN-ICHI ISHII[1]

Department of Biochemistry, Faculty of Pharmaceutical Sciences, Hokkaido University, Sapporo,[1] and Institute of Applied Microbiology, University of Tokyo, Bunkyo-ku, Tokyo 113,[2] Japan

Received 18 May 1987/Accepted 1 January 1988

The nucleotide sequence of gene 18 of bacteriophage T4 was determined by the Maxam-Gilbert method, partially aided by the dideoxy method. To confirm the deduced amino acid sequence of the tail sheath protein (gp18) that is encoded by gene 18, gp18 was extensively digested by trypsin or lysyl endopeptidase and subjected to reverse-phase high-performance liquid chromatography. Approximately 40 peptides, which cover 88% of the primary structure, were fractionated, the amino acid compositions were determined, and the corresponding sequences in DNA were identified. Furthermore, the amino acid sequences of 10 of the 40 peptides were determined by a gas phase protein sequencer, including N- and C-terminal sequences. Thus, the complete amino acid sequence of gp18, which consists of 658 amino acids with a molecular weight of 71,160, was determined.

The most striking morphological change in bacteriophage T4 upon infection is the contraction of its tail sheath. The tail sheath is 95 nm long and 20 nm in diameter and consists of 144 protomers of the sheath protein, gp18, arranged in 24 annuli (5, 13, 15). The structure of the tail sheath has been studied in detail by three-dimensional image reconstruction from electron micrographs of negatively stained specimens and hydrated, nonstained specimens by cryoelectron microscopy (1, 20, 30). Contraction of the sheath is triggered by conformational change in the baseplate upon infection, and the change in arrangement of the sheath protomers is transmitted like a wave toward the head-proximal end of the tail (24). The head and the inner tube rotate during contraction with respect to the baseplate fixed on the outer membrane of *Escherichia coli* (6). The resultant contracted sheath is 38 nm long and 25 nm in diameter. The contracted sheath is very stable; it does not dissociate with 8 M urea or 6 M guanidine hydrochloride, but it does dissociate when boiled in sodium dodecyl sulfate (33).

The morphological pathway of the tail structure has been worked out by King and co-workers (14, 15). The association of monomeric sheath protein gp18 with the tube-baseplate initiates as soon as tube protein monomer gp19 starts to polymerize on the baseplate. The length of the tail sheath is determined by the tube, whose length is determined by an elongated protein, possibly gp48 (9). There is a variety of missense, as well as nonsense, mutations in gene 18. They should be useful in studying the structure-function relationship of the protein. Furthermore, gp18 can be readily isolated in sizable quantity as a monomeric form which can reassemble onto the tube-baseplate to form the extended sheath (2, 34). Taking advantage of this system, we wish to elucidate the mechanism of sheath contraction at the molecular level. As a first step toward this goal, we determined the primary structure of gp18.

MATERIALS AND METHODS

Enzymes and reagents. DNase I and trypsin (bovine pancreas) were purchased from Sigma Chemical Co. Lysyl endopeptidase, a lysine-specific serine protease from "*Achromobacter lyticus*" M467-1, was purchased from Wako Junyaku Co. Polynucleotide kinase and all of the restriction enzymes used were purchased from Takara Shuzo Co. Methanesulfonic acid (4 M) was purchased from Pierce Chemical Co., and all other reagents were analytical grade, from either Nakarai Chemicals or Wako Junyaku Co. Nutrient broth, tryptone, yeast extract, Bacto-Agar, and Casamino Acids were purchased from Difco Laboratories.

Medium and buffer. M9A medium was used to grow *E. coli*. It contained, per liter of water, 6 g of Na_2HPO_4, 3 g of KH_2PO_4, 0.2 g of $MgSO_4 \cdot 7H_2O$, 0.5 g of NaCl, 1 g of NH_4Cl, 4 g of glucose, and 10 g of Casamino Acids. Phages were stored in B^+ buffer, which contained, per liter of water, 7 g of Na_2HPO_4, 4 g of NaCl, and 3 g of KH_2PO_4, supplemented with 1 mM $MgSO_4$ for use.

Phage, plasmids, and bacterial strains. T4D.23*am*H11 is from our collection. T4D.18*am* mutants (Fig. 1) were kindly supplied by W. B. Wood and J. King. *E. coli* BE was used as a nonpermissive host for amber mutant phages, and *E. coli* CR63 was the permissive host. *E. coli* JM103 was used to propagate M13 phages and to prepare replicative-form DNA of M13 phage. Three *E. coli* BE strains, which carry plasmids p662, p655, and p664, respectively, were kind gifts from Tom Mattson. Each of the three plasmids contains an *Eco*RI fragment which forms approximately one-third of gene 18 (Fig. 1).

Marker rescue test. Marker rescue of T4D.18*am* mutants by plasmids was performed as described by Mattson et al. (21). The results are also shown in Fig. 1.

DNA preparations for sequencing. DNA preparations for the Maxam-Gilbert method have been described by Christensen and Young (8). DNA was 5′ end labeled with polynucleotide kinase, and the strands were then separated by polyacrylamide gel electrophoresis, or alternatively, the double-stranded DNA was cut with a restriction enzyme and the desired fragment was purified by polyacrylamide gel electrophoresis.

DNA sequencing procedures. The Maxam-Gilbert method was done essentially as described by Maxam and Gilbert (22). The dideoxy method was carried out as described by Messing (23).

Preparation of gp18. Gp18 was prepared as described by

* Corresponding author.

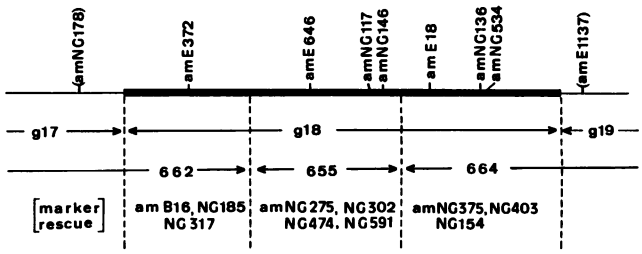

FIG. 1. Marker rescue of 18*am* mutations. Marker rescue experiments were carried out as described by Mattson et al. (21). The relative positions of some of the mutations were measured by standard crossing experiments with 17*am*NG178 or 19*am*E1137 and are shown above the horizontal line. The extents of T4 DNA carried by plasmids p662, p655, and p664 are indicated above the marker rescue results.

Tschopp et al. (34). Monomeric gp18 thus prepared was used within a week. Tails, which are the immediate source of monomeric gp18, could be stored for several months as a solution with 0.02% sodium azide at 4°C without loss of activity as judged by the reassociation ability of isolated gp18. Freezing tails for storage causes some decrease in the yield of monomeric gp18.

For Edman degradation of whole protein, gp18 was further purified by sodium dodecyl sulfate-gel electrophoresis as described by Hunkapiller et al. (11).

Proteolytic digestion of gp18. Monomeric gp18 in 1 mM phosphate buffer, pH 7.0, was heat denatured at 90°C for 10 min, and trypsin, treated with TPCK (L-1-tosylamide-2-phenylmethyl chloromethyl ketone) was added at a weight ratio of 1/100. After incubation at 37°C for 2 h, trypsin was further added at a weight ratio of 1/100 and incubated at 37°C for 12 h. The reaction was stopped by heating in boiling water for 3 min, and after cooling, trichloroacetic acid was added to a final concentration of 5%. After removal of precipitates by centrifugation at 2,300 × *g* for 15 min, the trichlororoacetic acid was removed by ether extraction and water was removed by evaporation. The gp18 digest was dissolved in water at about 4 mg/ml, and insoluble materials were removed with a 0.45-μm (pore size) membrane filter (Millipore Corp.) before application to high-performance liquid chromatography (HPLC).

HPLC of gp18 digests. Peptides were analyzed and separated on a reverse-phase HPLC system (Hitachi 655 or Waters 600) equipped with a gradient former, a degasser (Erma ERC-3510), and UV and fluorescence monitors (Hitachi 638 and JASCO FP110, respectively). A column of Nucleosil $5C_{18}$ (0.4 by 25 cm) was used. Water was deionized and doubly glass distilled before use. Twenty microliters of the sample was first analytically examined, and then 200 μl was applied and peaks were manually fractionated.

Amino acid analysis. The amino acid compositions of peptides, as well as of the whole protein, were determined, after hydrolysis with 4 M methanesulfonic acid at 115°C for 24 h as described by Simpson et al. (29), on an amino acid analyzer (Hitachi 835).

Sequence determination of peptides and protein. The amino acid sequences of peptides and the protein were determined on a gas phase protein sequencer (Applied BioSystems 470A). Usually, 0.5 to 1 nmol of sample was applied and analyzed. Phenylthiohydantoin derivatives of amino acids were analyzed and identified by HPLC.

RESULTS

DNA sequence determination of gene 18. Three plasmids, p662, p655, and p664, were used to sequence gene 18. The sequencing strategy is shown in Fig. 2b. The orientation of the 0.6-kilobase *Eco*RI fragment from p655 was originally decided on the basis of the open reading frame and then later confirmed by peptide analysis of gp18 as described in the next section. Because of the difficulty encountered in finding appropriate restriction sites in the 5'-end region of gene 18 by the Maxam-Gilbert method, the dideoxy method using M13 phage was also used (Fig. 2b). As a result, a sequence of approximately 2,000 bases that cover all of gene 18, the 3' end of gene 17, and the 5' end of gene 19 was determined

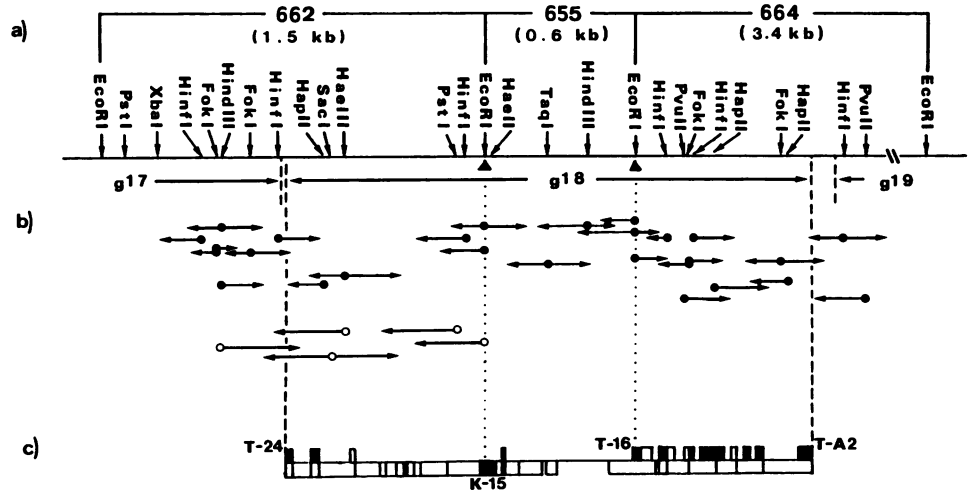

FIG. 2. DNA sequencing of gene 18 and peptide analysis of its product. (a) Restriction map of the gene 18 region. (b) Orientations and lengths of nucleotide sequences obtained (arrows). Arrows with closed circles represent sequence information from the Maxam-Gilbert method, and arrows with open circles are from dideoxy sequencing using M13 clones. (c) Assignment of tryptic and lysyl endopeptidase fragments of gp18. Locations of the peptides identified are drawn in comparison with the DNA sequence. Boxes above the continuous line are for tryptic peptides, and those under the line are for lysyl endopeptidase-digested peptides. Open boxes represent peptides identified by amino acid composition only, and closed boxes represent those identified by amino acid sequence as well.

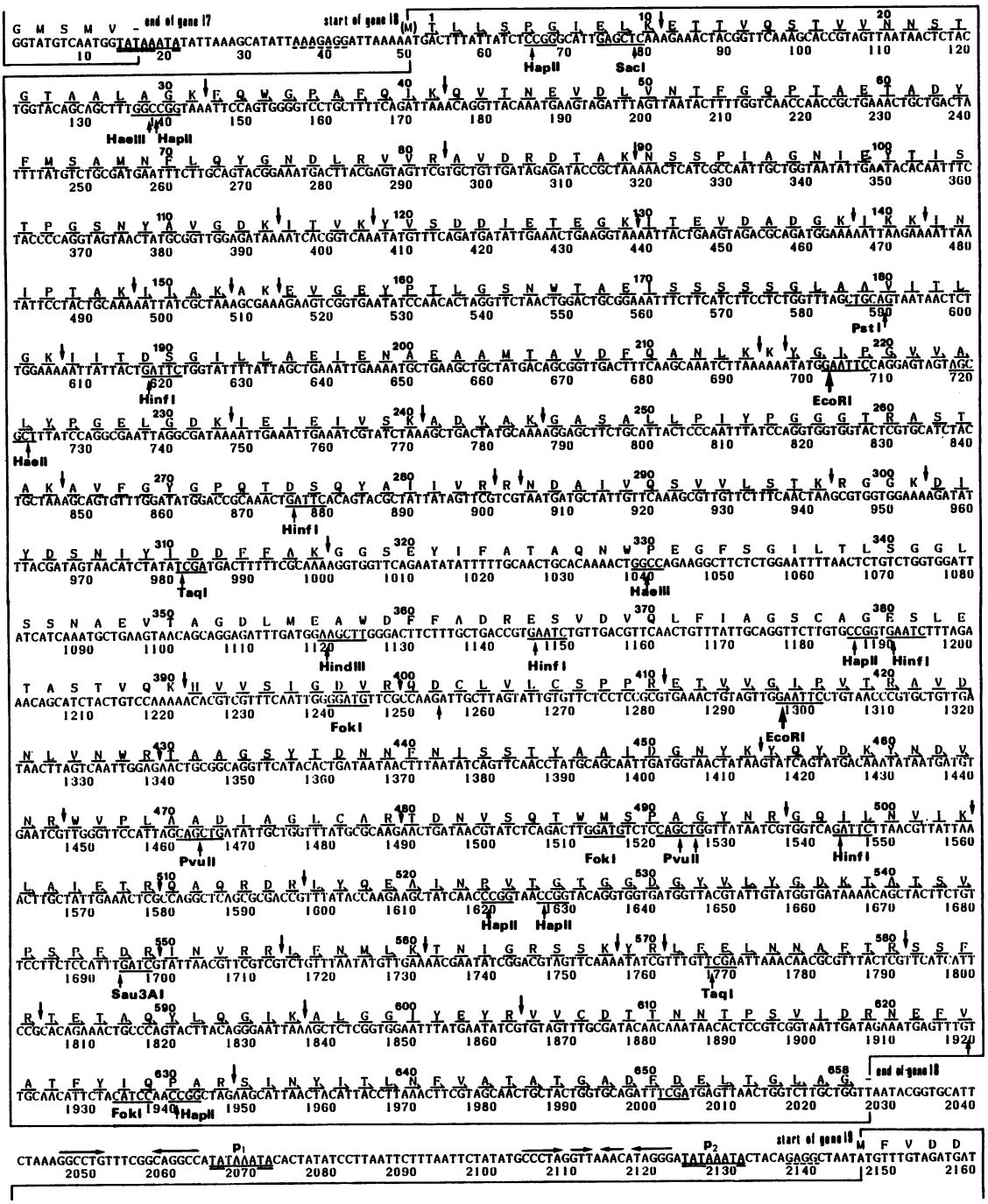

FIG. 3. Sequence of gene 18 and its product. T4 late promoter consensus sequences TATAAATA for genes 18 and 19 are doubly underlined by solid lines, and the putative Shine-Dalgarno sequences for the two genes are underlined by broken lines. Arrows pointed downward along the amino acid sequence indicate the cleavage sites by trypsin or lysyl endopeptidase, and those pointed upward along the nucleotide sequence indicate the restriction sites. The amino acids identified by analysis of the peptides are underlined by single bars, and those identified by amino acid sequence analysis as well are underlined by arrows. The N-terminal sequence for the whole protein was determined from Thr 1 to Pro 5. Two predicted stem-and-loop structures near the 5′ end of the two promotors between genes 18 and 19 are marked by arrows. Their stabilization energies, calculated by the method of Ooi and Takanami (25), are −19.4 kcal (1 cal = 4.184 J) (near p_1) and −15.4 kcal (near p_2), respectively.

(Fig. 3). The 3′ end of gene 17 has been sequenced by D. Powell, J. Franklin, F. Arisaka, and G. Mosig (manuscript in preparation), and the 5′ end of gene 19 has also been determined (3).

Description of the DNA sequence. Figure 3 shows the DNA sequence determined as described above. As indicated in the figure, the T4 phage late promoter consensus sequence, TATAAATA (8), and a putative Shine-Dalgarno sequence,

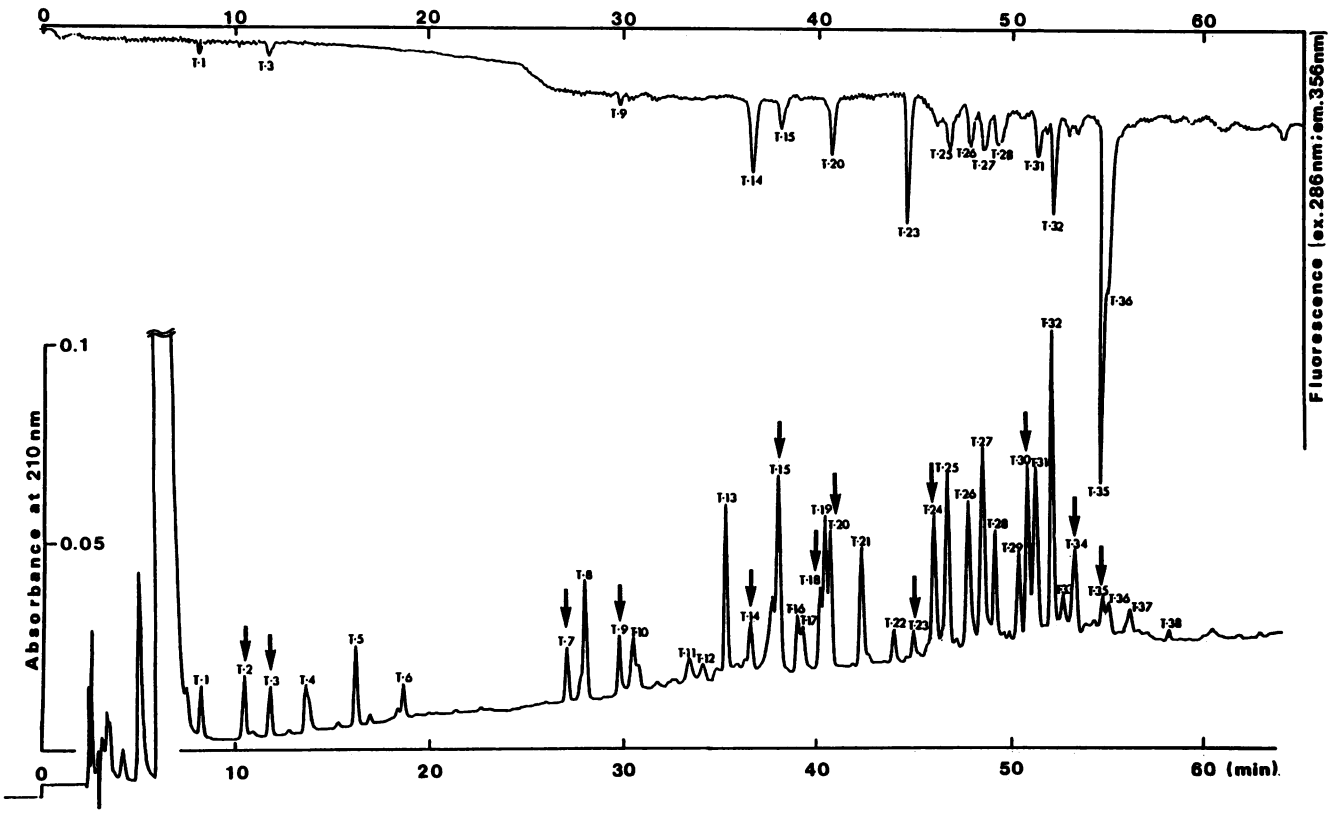

FIG. 4. Isolation of tryptic peptides of gp18 by reverse-phase HPLC. A Nucleosil $5C_{18}$ column (0.4 by 25 cm) was used, with elution with a 4 to 36% acetonitrile linear gradient in 0.1% trifluoroacetic acid at 1 ml/min. The lower and upper elution profiles represent A_{210} and fluorescence emission (excitation at 286 nm, emission at 356 nm) from tryptophan residues, respectively. The peaks indicated by arrows were fractionated and rechromatographed under the same conditions except for 10 mM ammonium formic acid, pH 6.7, instead of 0.1% trifluoroacetic acid.

TAAaGAGG, are upstream from the initiation codon for gene 18; the capital letters in the latter sequence are complementary to the 3′ end of the 16S rRNA. The significance of two late promoters and the predicted two stem-and-loop structures between genes 18 and 19 are addressed in the Discussion.

Analysis of tryptic fragments of gp18. Purified gp18 was heat denatured, digested with trypsin, and then subjected to reverse-phase HPLC (Fig. 4). A number of well-separated peaks were fractionated and further purified by rechromatography on reverse-phase HPLC. Their amino acid sequences were determined by a gas phase protein sequencer; the results are indicated by arrows in the sequence (Fig. 3). All of the sequences were located on the DNA sequence. Peptide T-24 is the N-terminal one; the N-terminal threonine was further confirmed by Edman degradation of the whole polypeptide chain. Peptide T-16 corresponds to the junction sequence between DNA fragments 655 and 664 (Fig. 2b). Peptide K-15, corresponding to the junction between 662 and 655, was also found in the lysyl endopeptidase digest of gp18, as described below.

When a protein sequence is deduced from the DNA sequence, it is essential to determine at least the N-terminal and C-terminal sequences of its gene product to obtain the complete primary structure of the protein unambiguously. Although the N-terminal peptide was found among the tryptic fragments of gp18 separated by HPLC (Fig. 4), the C-terminal peptide was not. We therefore tried to isolate the C-terminal peptide from the tryptic digest of gp18 by affinity chromatography, in which the ability of immobilized anhydrotrypsin to bind peptides specifically with arginine or lysine at their C termini was used (12, 17). We thus identified T-A2 as the C-terminal peptide (Fig. 2c). The complete primary structure of gp18, consisting of 658 amino acid residues with a molecular weight of 71,160, was determined (Fig. 3). Excellent agreement of the amino acid sequence with the amino acid composition of the protein was obtained (Table 1). A predicted secondary structure of gp18 by the method of Chou and Fasman (7) is shown in Fig. 5.

Peptide mapping of gp18 by lysyl endopeptidase. One-dimensional peptide mapping of gp18 was carried out on reverse-phase HPLC. The objective of the peptide mapping was twofold. (i) We intended to confirm further the primary structure of gp18, and (ii) for future study of the higher-order structure of gp18, it is desirable to establish a system by which one can readily identify a peptide whose chemical nature has been altered by chemical modification or amino acid substitution by mutation. Lysyl endopeptidase was used because of its high specificity for lysine residues and because it produces fewer peptides than does trypsin (approximately 30 instead of more than 50). gp18 was S-carboxymethylated with monoiodoacetic acid under denaturing and reducing conditions and digested by lysyl endopeptidase as described in Materials and Methods. The resultant digest was subjected to reverse-phase HPLC, and all of the major peaks were fractionated. After further purification by rechromatography, the amino acid composition of each peptide was determined and the position of the

TABLE 1. Amino acid composition of gp18[a]

Amino acid	Expected no. of residues from nucleotide sequence (amino acid analysis)
Asx	80 (83.9)
Thr	55 (54.0)
Ser	47 (44.8)
Glx	55 (59.9)
Pro	23 (22.8)
Gly	56 (57.1)
Ala	67 (70.8)
Cys	5 (5.0)
Val	50 (39.5)
Met	6 (5.8)
Ile	50 (41.1)
Leu	42 (43.0)
Tyr	31 (30.7)
Phe	24 (24.0)
His	1 (1.0)
Lys	31 (31.0)
Trp	7 (5.5)
Arg	28 (28.6)

[a] gp18 had a total of 658 residues and a molecular weight of 71,160.

TABLE 2. Assignment of lysyl endopeptidase fragments in HPLC[a]

Peptide	Residue no.
K-1	1–10
K-2	11–31
K-3	32–41
K-4	42–89
K-5	90–114
K-6	115–118
K-7	119–129
K-8	130–138
K-9	139–141
K-10	142–148
K-11	149–152
K-12	Not recovered
K-13	155–185
K-14	186–214
K-15	216–232
K-16	233–240
K-17	241–245
K-18	246–26
K-19	266–297
K-20	Not recovered
K-21	302–316
K-22	Not recovered
K-23	391–454
K-24	455–459
K-25	460–503
K-26	504–538
K-27	539–560
K-28	561–568
K-29	569–595
K-30	596–658

[a] These data are taken from Fig. 6.

peptides along the corresponding DNA sequence was identified. The results are shown in Fig. 6 and Table 2. The identified peaks cover 88% of the total primary structure of gp18 (Fig. 2c).

DISCUSSION

The complete nucleotide sequence of gene 18 and the deduced amino acid sequence, most of which was confirmed by peptide analysis, have been presented. Concerning the nucleotide sequence, a T4 phage late promoter consensus sequence was found 29 bases upstream from the initiation codon. Since Stahl et al. (31) have demonstrated that genes 17 and 18 have no polarity effect, it is very likely that the consensus sequence is the true late promoter for gene 18. The stop codon for gene 17 was in the late promoter sequence for gene 18 and no terminatorlike structure was found near the 3′ end of gene 17. On the other hand, two late promoter consensus sequences were found between genes 18 and 19. Originally, we expected to find no promoter sequence between genes 18 and 19 because of the significant polarity effect between these two genes (31). An S1 mapping experiment for that region of the DNA was carried out to elucidate the roles of the two late promoter consensus sequences and the terminatorlike structure at the 3′ end of

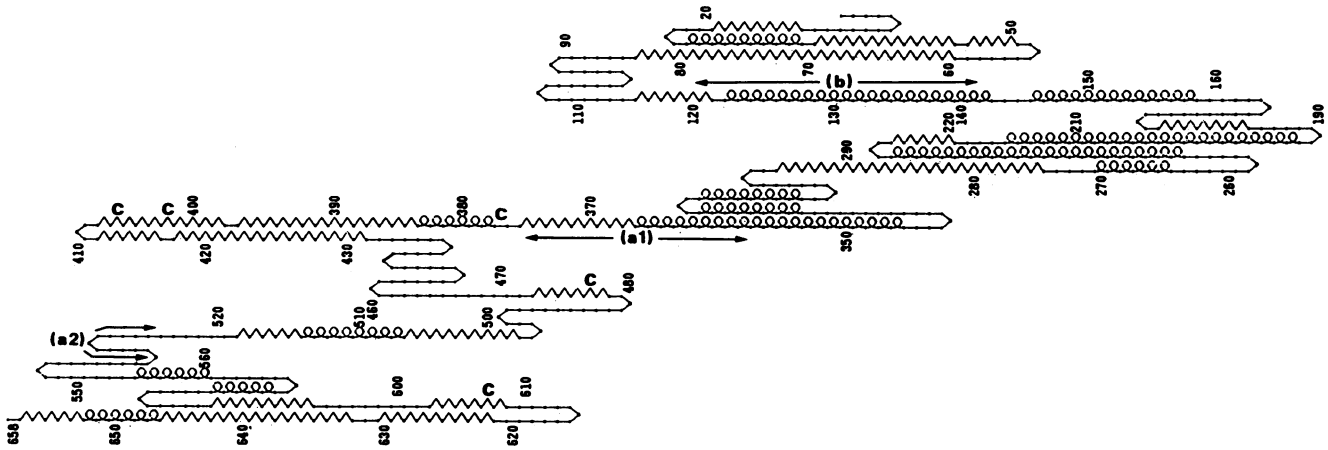

FIG. 5. Predicted secondary structure of gp18. The secondary structure of gp18 was estimated by the method of Chou and Fasman (7). Symbols: ⍺, α helix; ∧, β sheet; ⊃, β turn; —, remainder. The locations of five cysteine residues are indicated (C). Also, sequences homologous to the GTP- or ATP-binding domains of some nucleotide-binding proteins (a1 and a2) and the EF hand of the Ca^{2+}-binding domain (b; see Discussion), respectively, are indicated by arrows.

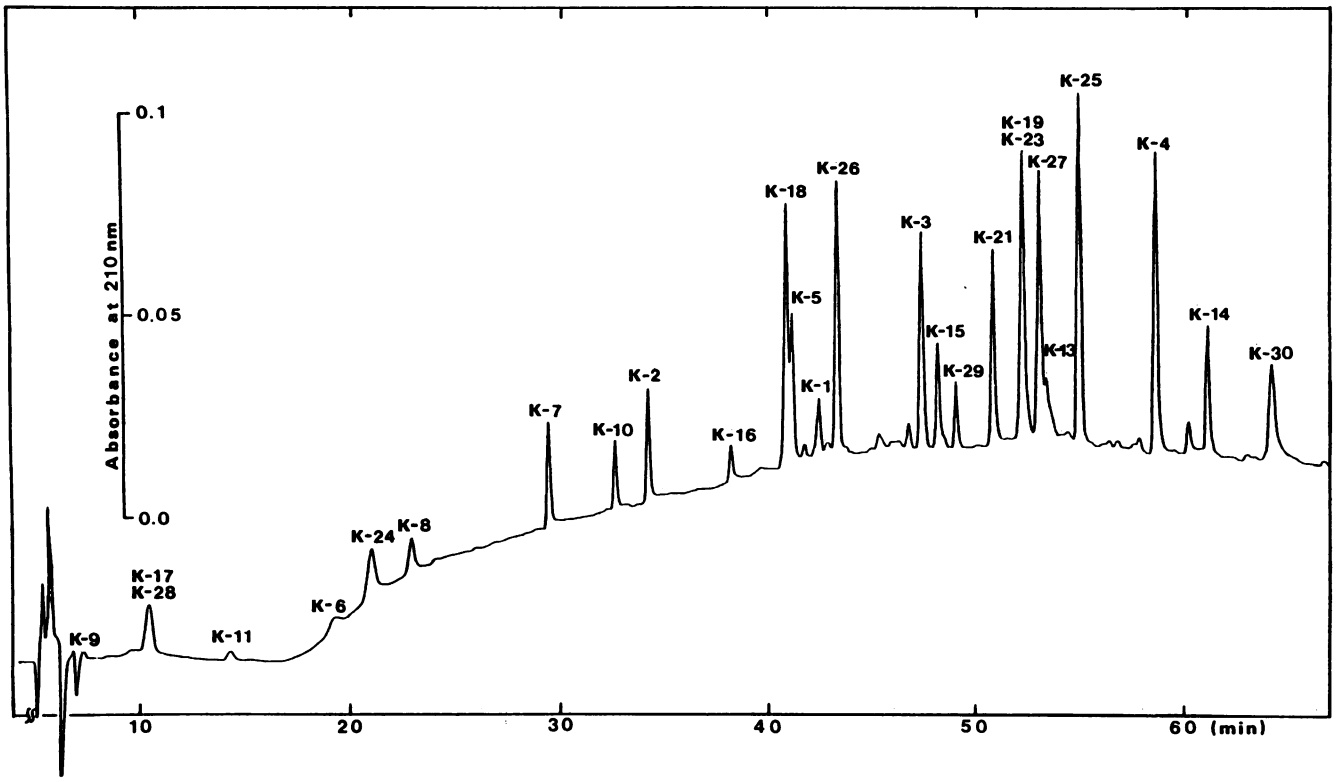

FIG. 6. Peptide mapping of gp18 by lysyl endopeptidase. An elution profile of reverse-phase HPLC as measured by absorption at 210 nm is shown. Conditions were as described in the legend to Fig. 4, except that the peptides were eluted at a speed of 0.5 ml/min. All of the major peaks were fractionated, and the amino acid compositions were determined. The identified peptides in the primary structure are listed in Table 2.

each late promoter sequence. It revealed that both of the consensus sequences serve as weak promoters but most of the transcripts arise from further upstream of the two promoters, possibly from the promoter for gene 18 (3).

When we started the peptide analysis of gp18, we assumed that the N-terminal amino acid residue was blocked, because Brenner et al. (5) could not detect the N-terminal amino acid by the 2,4-dinitrophenol method. Since we identified the N-terminal peptide with which Edman degradation was feasible, we reexamined the 2,4-dinitrophenol method on the whole protein and found 2,4-dinitrophenol-threonine after hydrolysis (data not shown). We used monomeric gp18 for N-terminal analysis instead of the contracted sheath which they used. This might explain why they did not detect the N-terminal residue. Contracted sheath is known to be extremely stable (33). Recently, Ben-Bassat and Bauer (4) have presented an empirical rule for the removal of the N-terminal methionine based on the second amino acid residue. The removal of N-terminal methionine in gp18 agrees with their rule that it is removed when the second residue is threonine. The proposed rule also agrees with the fact that the N-terminal methionine of gp19 is retained when the second residue is phenylalanine (3).

Among the three peptides not recovered (Table 3; Fig. 6), peptide 317-390 contains most of the sequence not confirmed by peptide analysis. This peptide is considered to be hydrophobic and appears to aggregate and precipitate during protease digestion. Attempts to recover this peptide with reasonable yield have been unsuccessful. However, it should be noted that where confirmation of the nucleotide sequence by peptide analysis was not carried out, the nucleotide sequence was determined in both directions (Fig. 2b).

Determination of the primary structure is the first step toward elucidation of the structure-function relationship of the tail sheath protein. Twenty-nine mutant strains of T4 with distinct phenotypes are present in gene 18. They include *ts* (growth at 30°C but not at 42°C; two strains), *cs* (growth at 37°C but not at 25°C; three strains), *hs* (heat labile; four strains), *CBW* (can infect in the presence of high concentration of polyethylene glycol in the medium; three strains), and *amber* mutants (17 strains). Identification of the mutation sites in gene 18 or gp18 will give us information concerning roles of the amino acid residues at the mutation sites. For example, it is likely that *CBW* mutants have an altered amino acid residue located on the surface of the tail sheath, where it must interact with the tail fibers. Also, the peptide map produced by lysyl endopeptidase will help us to identify peptides which contain a substituted amino acid or those which have a chemically modified amino acid residue.

TABLE 3. Molecular weights of sheath and tube proteins

Phage	Mol wt (10^3) of:		Reference
	Sheath protein	Tube protein	
T4	71	19	This study
SPO1	60	12	26
Mu	55	13	10
P2	46	20	19
PS17	36	18	28

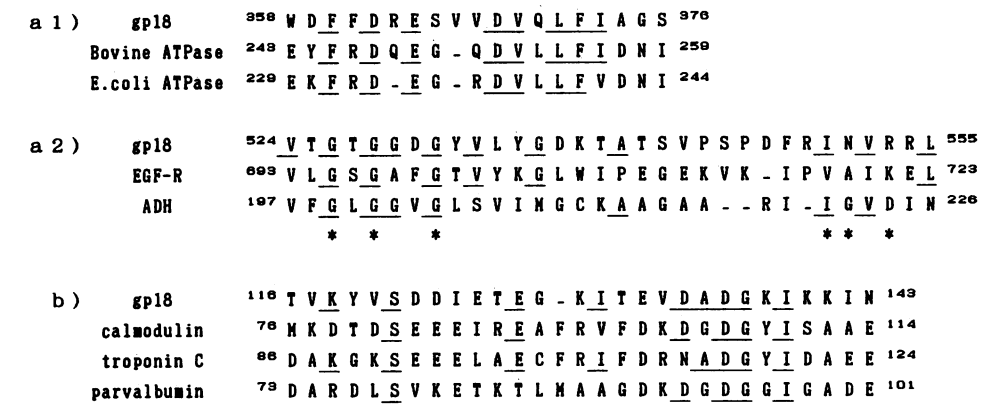

FIG. 7. Sequences in gp18 homologous to nucleotide-binding sites (a1, reference 18; a2, reference 32) and the EF hand (b, reference 16). Abbreviations: EGR-R, epidermal growth factor; ADH, alcohol dehydrogenase.

Some characteristic features of gp18 drawn from the primary structure may be pointed out. There are five Cys residues, and all of them are in the C-terminal half of the molecule (Fig. 5). Preliminary experiments have indicated that two of them form a disulfide bond (data not shown). There is a sequence in the N-terminal region with five consecutive serine residues (Fig. 3). This region is considered to assume an undefined conformation, possibly located on the surface of the protein. Also, there is only one histidine residue in the total of 658 residues.

Recently, it was reported that gp18 is a GTP-binding protein (27). We therefore searched for homology to nucleotide-binding proteins. Some homology was found (Fig. 7a1 and a2). Further study is necessary to assess the significance of the homology and the role of GTP in the sheath protein. There is also a region homologous to Ca^{2+}-binding proteins, the so-called EF hand (16; Fig. 7b). The significance of this homology is not clear, but Ca^{2+} might be important to stabilize the native structure of gp18.

It has been reported that gp18 has 2 to 3% hexosamine (5, 34). Amino acid analysis of gp18 has shown that it has some amino sugar whose elution time in the analyzer does not coincide with that of any commonly encountered amino sugar such as glucosamine, galactosamine, or mannosamine (R. Muramatsu, F. Arisaka, and S. Ishii, unpublished data). However, we did not detect this amino sugar in the amino acid analyses of peptides in the present study, which covered 88% of the total primary structure. In particular, it was apparent that there are no covalently attached amino sugars in those peptides whose sequences were determined by Edman degradation in a protein sequencer. The structure of the amino sugar and its mode of linkage to the protein remain to be elucidated.

Another approach to elucidation of the structure-function relationship of the sheath protein would be to compare the primary structure of T4 with those of other phages. Molecular weights of contractile tail sheath proteins from a number of bacteriophages are listed in Table 3. Although tail tube proteins have relatively constant molecular weights, the molecular weights of the sheath proteins vary considerably. We postulate that these sheath proteins share a homologous domain which plays an important role in sheath contraction.

Recently, tail sheath and tube genes from *Pseudomonas aeruginosa* phage PS17 have been cloned (T. Shinomiya, personal communication). Also, the sequence of the genome of phage 186, closely related to phage P2, will be determined (J. B. Egan, personal communication). We therefore expect that the primary structures of at least two phages other than T4 will soon be available. The sequence data will be very useful in studying the structure-function relationship of the sheath protein and may provide insights into the evolutionary relationship among these phages.

ACKNOWLEDGMENTS

We thank T. Mattson for providing plasmids 662, 655, and 664, G. Mosig for sending data before publication, T. Senba and co-workers, Nikkaki Co., Tokyo, and N. Ochi and co-workers, Chugai Pharmaceutical Co., Tokyo, for determination of the amino acid sequences of peptides. Thanks are also due Frederick A. Eiserling and Robert L. Duda for valuable comments on the manuscript.

This work was supported in part by grants-in-aid for scientific research from the Ministry of Education, Science, and Culture of Japan.

LITERATURE CITED

1. **Amos, L. A., and A. Klug.** 1975. Three dimensional image reconstructions of the contractile tail of the T4 bacteriophage. J. Mol. Biol. **99**:51–73.
2. **Arisaka, F., J. Tschopp, R. van Driel, and J. Engel.** 1979. Reassembly of the bacteriophage T4 tail from the core-baseplate and the monomeric sheath protein P18: a cooperative association process. J. Mol. Biol. **132**:369–386.
3. **Arisaka, F., L. Ishimoto, G. Kassavetis, T. Kumazaki, and S. Ishii.** 1988. Nucleotide sequence of the tail tube structural gene of bacteriophage T4. J. Virol. **62**:882–886.
4. **Ben-Bassat, A., and K. Bauer.** 1987. Amino-terminal processing of proteins. Nature (London) **326**:315.
5. **Brenner, S., G. Streisinger, R. W. Horne, S. P. Champe, L. Barnett, S. Benzer, and M. W. Rees.** 1959. Structural components of bacteriophage. J. Mol. Biol. **1**:281–292.
6. **Caspar, D. L. D.** 1980. Movement and self-control in protein assemblies. Quasi-equivalence revisited. Biophys. J. **32**:103–135.
7. **Chou, P. Y., and G. D. Fasman.** 1978. Empirical predictions of protein conformation. Annu. Rev. Biochem. **47**:251–276.
8. **Christensen, A. C., and E. T. Young.** 1982. T4 late transcripts are initiated near a conserved DNA sequence. Nature (London) **299**:369–371.
9. **Duda, R. L., and F. A. Eiserling.** 1982. Evidence for an internal component of the bacteriophage T4D tail core: a possible length-determining template. J. Virol. **43**:714–720.
10. **Giphart-Gassler, M., C. Wijffelman, and J. Reeve.** 1981. Structural polypeptides and products of late genes of bacteriophage Mu. J. Mol. Biol. **145**:139–163.
11. **Hunkapiller, M. W., E. Lujan, F. Ostrander, and L. E. Hood.** 1983. Isolation of microgram quantities of proteins from poly-

acrylamide gels for amino acid sequence analysis. Methods Enzymol. **91**:227–236.
12. **Ishii, S., H. Yokosawa, T. Kumazaki, and I. Nakamura.** 1983. Immobilized anhydrotrypsin as a specific affinity adsorbent for tryptic peptides. Methods Enzymol. **91**:378–383.
13. **Kellenberger, E., and E. Boy de la Tour.** 1964. On the fine structure of normal and "polymerized" tail sheath of phage T4. J. Ultrastruct. Res. **11**:545–563.
14. **Kikuchi, Y., and J. King.** 1975. Genetic control of bacteriophage T4 baseplate morphogenesis. III. Formation of the central plug and overall assembly pathway. J. Mol. Biol. **99**:695–716.
15. **King, J., and N. Mykolajewycz.** 1973. Bacteriophage T4 tail assembly: proteins of sheath, core and baseplate. J. Mol. Biol. **75**:339–358.
16. **Kretsinger, R. H., and C. E. Nockolds.** 1973. Carp muscle calcium-binding protein. II. Structure determination and general description. J. Biol. Chem. **248**:3313–3326.
17. **Kumazaki, T., T. Nakako, F. Arisaka, and S. Ishii.** 1986. A Novel method for selective isolation of C-terminal peptides from tryptic digests of proteins by immobilized anhydrotrypsin: application to structural analyses of the tail sheath and tube proteins from bacteriophage T4. Proteins Struct. Funct. Genet. **1**:100–107.
18. **Leberman, R., and U. Egner.** 1984. Homologies in the primary structure of GTP-binding proteins: the nucleotide-binding site of EF-Tu and p21. EMBO J. **3**:339–341.
19. **Lengyel, J. A., R. N. Goldstein, M. Marsh, and R. Calendar.** 1974. Structure of bacteriophage P2 tail. Virology **62**:161–174.
20. **Lepault, J., and K. Lenard.** 1985. Three-dimensional structure of unstained, frozen-hydrated extended tails of bacteriophage T4. J. Mol. Biol. **182**:431–443.
21. **Mattson, T., G. van Howe, A. Bolle, G. Selzer, and R. Epstein.** 1977. Genetic identification of cloned fragments of bacteriophage T4 DNA and complementation by some clones containing early T4 genes. Mol. Gen. Genet. **154**:319–326.
22. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
23. **Messing, J.** 1983. New M13 cloning vectors. Methods Enzymol. **101**:20–78.
24. **Moody, M. F.** 1973. Sheath of bacteriophage T4. III. Contraction mechanism deduced from partially contracted sheaths. J. Mol. Biol. **60**:613–635.
25. **Ooi, T., and M. Takanami.** 1981. A computer method for construction of secondary structure from polynucleotide sequence: possible structure of the bacterial replication origin. Biochim. Biophys. Acta **655**:221–229.
26. **Parker, M. L., and F. A. Eiserling.** 1983. Bacteriophage SPO1 structure and morphogenesis. I. Tail structure and length regulation. J. Virol. **46**:239–249.
27. **Serysheva, I. I., A. I. Tourkin, S. Y. Venyaminov, and B. F. Poglazov.** 1984. On the presence of guanosine phosphate in the tail of bacteriophage T4. J. Mol. Biol. **179**:565–569.
28. **Shinomiya, T., and S. Shiga.** 1979. Bactericidal activity of the tail of *Pseudomonas aeruginosa* bacteriophage PS17. J. Virol. **32**:958–967.
29. **Simpson, R. J., M. R. Neuberger, and T.-Y. Liu.** 1976. Complete amino acid analysis of proteins from a single hydrolysate. J. Biol. Chem. **251**:1936–1940.
30. **Smith, P. R., U. Aebi, R. Josephs, and M. Kessel.** 1976. Studies of the structure of the T4 bacteriophage tail sheath. J. Mol. Biol. **106**:243–275.
31. **Stahl, F. W., J. M. Crasemann, C. Yegian, M. M. Stahl, and A. Nakata.** 1970. The linkage map of bacteriophage T4. Genetics **64**:157–170.
32. **Sternberg, M. J. E., and W. R. Taylor.** 1984. Modelling the ATP-binding site of oncogene products, the epidermal growth factor receptor and related proteins. FEBS Lett. **175**:387–392.
33. **To, C. M., E. Kellenberger, and A. Eisenstark.** 1969. Disassembly of T-even bacteriophage into structural parts and subunits. J. Mol. Biol. **46**:493–511.
34. **Tschopp, J., F. Arisaka, R. van Driel, and J. Engel.** 1979. Purification, characterization and reassembly of the bacteriophage T4D tail sheath protein P18. J. Mol. Biol. **128**:281–286.