JOURNAL OF VIROLOGY, Apr. 1988, p. 1180–1185
0022-538X/88/041180-06$02.00/0

Vol. 62, No. 4

Nucleotide Sequence of the Large Double-Stranded RNA Segment of Bacteriophage φ6: Genes Specifying the Viral Replicase and Transcriptase

LEONARD MINDICH,[1]* IRIS NEMHAUSER,[1,2] PAUL GOTTLIEB,[1] MARTIN ROMANTSCHUK,[3] JACOB CARTON,[1] STEPHEN FRUCHT,[1] JEFFREY STRASSMAN,[1] DENNIS H. BAMFORD,[3] AND NISSE KALKKINEN[4]

Department of Microbiology, the Public Health Research Institute of the City of New York, Inc.,[1] and Department of Microbiology, New York University School of Medicine,[2] New York, New York 10016, and Department of Genetics[3] and Recombinant DNA Laboratory,[3] University of Helsinki, SF-00100 Helsinki, Finland

Received 19 October 1987/Accepted 22 November 1987

The genome of the lipid-containing bacteriophage φ6 contains three segments of double-stranded RNA. We determined the nucleotide sequence of cDNA derived from the largest RNA segment (L). This segment specifies the procapsid proteins necessary for transcription and replication of the φ6 genome. The coding sequences of the four proteins on this segment were identified on the basis of size and the correlation of predicted N-terminal amino acid sequences with those found through analysis of isolated proteins. This report completes the sequence analysis of φ6. This constitutes the first complete sequence of a double-stranded RNA genome virus.

Bacteriophage φ6 infects the plant pathogen *Pseudomonas phaseolicola* HB10Y. Three separate pieces of double-stranded RNA compose its genome, which is located inside a polyhedral nucleocapsid (26). The nucleocapsid has RNA polymerase activity (20) and is covered by a lipid-containing membrane (26). The assembly pathway involves the formation of a polyhedral procapsid composed of four proteins, P1, P2, P4, and P7, which is then filled with one copy each of the three pieces of double-stranded RNA per virion (4, 16). These filled procapsids are then covered with a shell of protein P8 to become nucleocapsids, which are subsequently enveloped within the lipid-containing membrane (15). This envelopment process is dependent upon the activity of the nonstructural protein P12 (17). The four proteins that constitute the procapsid are coded for by the largest of the three genomic segments, and their synthesis begins early in infection (25). The procapsid is responsible for genomic replication and transcription (24).

The study of the phage assembly process has been facilitated by the cloning of cDNA of each genomic segment (18). We recently determined the nucleotide sequence of the cDNA derived from the small RNA segment (12) and the middle segment (P. Gottlieb et al., unpublished data). In this communication, we describe the nucleotide sequence analysis of cDNA derived from the large genomic segment designated L.

MATERIALS AND METHODS

Bacterial strains, phage, and plasmids. *Escherichia coli* JM105 [Δ(*lac pro*) *thi strA endA sbcB15 hsdR4* F′ *traD36 proAB lacI*^{q} ZΔM15] was used as a host for M13 cloning. *E. coli* HB101 (*hsd20*, an r_K^- m_K^- strain) and MM294 (an r_K^- m_K^+ strain) were used as hosts for clones with plasmid pBR322 or pUC8 as a vector. Phage M13 (mp10 and mp11) replicative-form DNA (14) was used to clone cDNA fragments for the sequencing reaction. Plasmids pLMF72, pLMF306, pLMF308, and pLMF309 are pBR322 derivatives containing cDNA insert fragments of segment L (Fig. 1).

Preparation of DNA. Purified single-stranded M13 phage DNA for the sequencing reaction template was purified from 1.5-ml cultures by the procedure in the Amersham Corp. *M13 cloning and sequencing handbook*. Large plasmid preparations (200- to 500-ml overnight cultures) were prepared by the cleared-lysate method of Clewell (2), with subsequent CsCl gradient centrifugation. Restriction digests used endonucleases supplied by Boehringer Mannheim Biochemicals, and the buffer conditions were those specified by this manufacturer. Ligation reactions used T4 DNA ligase supplied by Collaborative Research, Inc. Conditions were as described in the *M13 cloning and sequencing handbook*. Procedures for 0.8% agarose and 5% polyacrylamide gel electrophoresis have been previously described (13). Transformation of *E. coli* JM105 was as described in the M13 sequencing and cloning handbook. Transformation of *E. coli* HB101 and MM294 has been previously described (18).

Protein purification and amino acid sequencing. Unlabeled and [^{3}H]leucine- and [^{3}H]alanine-labeled φ6*h*1s were grown and purified as described previously (1). Preparative and analytical protein gel electrophoresis was performed as previously described (1). Individual proteins electrophoretically eluted from gel slices were free of contaminating material as analyzed by subsequent sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

Two sequencing strategies were used to localize the amino terminus of the individual proteins upon the nucleotide sequence. First, unlabeled proteins (to which ^{14}C label had been added to facilitate protein localization within the preparative sodium dodecyl sulfate-polyacrylamide gel) were subjected to automated Edman degradation in a Beckman 890 D sequencer and a 0.1 M Quadril program. The amino acids were identified as their phenylthiohydantoin derivatives by high-pressure liquid chromatography (Varian 5020 liquid chromatograph; UV-5 detector, 269 nm).

In the second method for radiosequence analysis, the eluted tritium-labeled proteins were degraded as described above, together with 200 μg of apomyoglobin. Ninety percent of each phenylthiohydantoin-derivatized sample was counted by liquid scintillation, and the remaining 10% was analyzed by high-pressure liquid chromatography, with the

* Corresponding author.

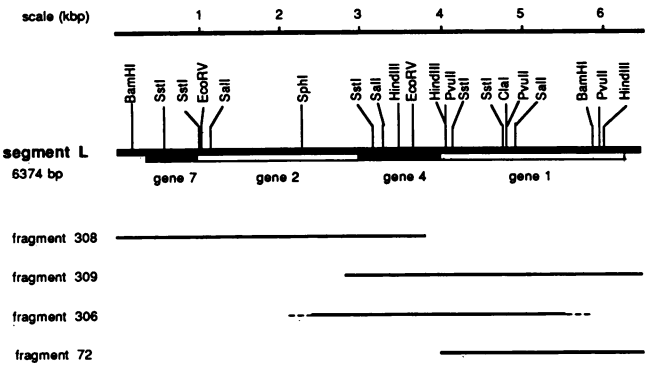


FIG. 1. Restriction map of the entire large genomic segment. Sites were obtained from restriction nuclease digestion of the cloned cDNA fragments indicated. The illustration is aligned with the direction of transcription to the right. cDNA fragments used for determination of the nucleotide sequence are shown below the map. Kbp, Kilobase pairs; bp, base pairs.

known sequence of carrier apomyoglobin used as a control for sequencer performance.

DNA sequencing. Nucleotide sequence analysis was performed by the dideoxy chain termination method of Sanger et al. (23). The sequence was determined for both strands. Fragments of cDNA for sequencing were first cloned into the unique restriction site region of the replicative form of bacteriophage M13 DNA. For this work, M13 vectors mp10 and mp11 were used (14). Our strategy involved the use of specific restriction fragments as well as the use of unselected collections of restriction fragments generated by treatment with *Taq*I. Sequences were determined on over 550 fragments. Single-stranded DNA isolated from mature M13 virus served as a template for dideoxy sequencing. The polymerase reaction was primed with a 17-nucleotide primer purchased from P-L Biochemicals, Inc., and chain extension was with the Klenow fragment of DNA polymerase I (Boehringer Mannheim Biochemicals). Labeling was performed with [α-^{35}S]dATP, >600 Ci/mmol (Amersham Corp.).

Dideoxy sequencing using dITP and reverse transcriptase. Some sections of the sequence displayed GC compressions. We resolved them by substituting dITP for dGTP in the reaction mixture (22). Chain extension reactions used avian myeloblastosis virus reverse transcriptase (Molecular Genetics Resources). In these reactions, the molar ratio of deoxy and dideoxy nucleotides was altered from that of the Klenow fragment-catalyzed reaction. Final reaction concentrations were 0.77 μM dATP to 0.027 μM ddATP, 889 μM dCTP to 819 ddCTP, 889 μM dTTP to 8.9 μM ddTTP, and 220 μM dITP to 0.35 μM ddGTP. The reaction buffer contained 33 mM Tris hydrochloride (pH 8.3), 6.7 mM $MgCl_2$, 46.7 mM KCl, and 20 mM mercaptoethanol. The 17-mer primer was present at 0.3 ng per reaction. The chain termination reactions were catalyzed by 2.5 U of avian myeloblastosis virus reverse transcriptase at 42°C for 20 min. These were subsequently chased with a mixture of four deoxynucleotide triphosphates at 222 μM each for an additional 20 min. Each chain termination reaction contained 2.5 μCi of [α-^{35}S]dATP. ^{35}S-labeled DNA fragments were denatured and analyzed by gel electrophoresis on 6 or 8% polyacrylamide gels (42.5 cm long) containing 8 M urea, 89 mM Tris-borate (pH 8.3), and 2 mM EDTA.

Computer analysis. The computer facility of the Public Health Research Institute of the City of New York was used. It consists of the Vax 11/750 computer equipped with the multiuser UNIX operating system. Nucleic acid analysis programs were provided by the sequence analysis package of the Biomathematics Computation Laboratory, Department of Biochemistry and Biophysics, University of California at San Francisco.

RESULTS

cDNA cloning of the ϕ6 genome was described by Mindich et al. (19). Plasmids were constructed that bore fragments of the L segment, and the protein reading frame positions were estimated by restriction site analysis in conjunction with in vivo complementation studies, in vitro coupled transcription translation of plasmid DNA, and reference to the cloning work of Revel et al. (21). Ultimately, the reading frame assignments were made by the identification of the N-terminal amino acid sequence of the isolated proteins and comparing these with the sequences predicted from the nucleotide sequence of the cDNA.

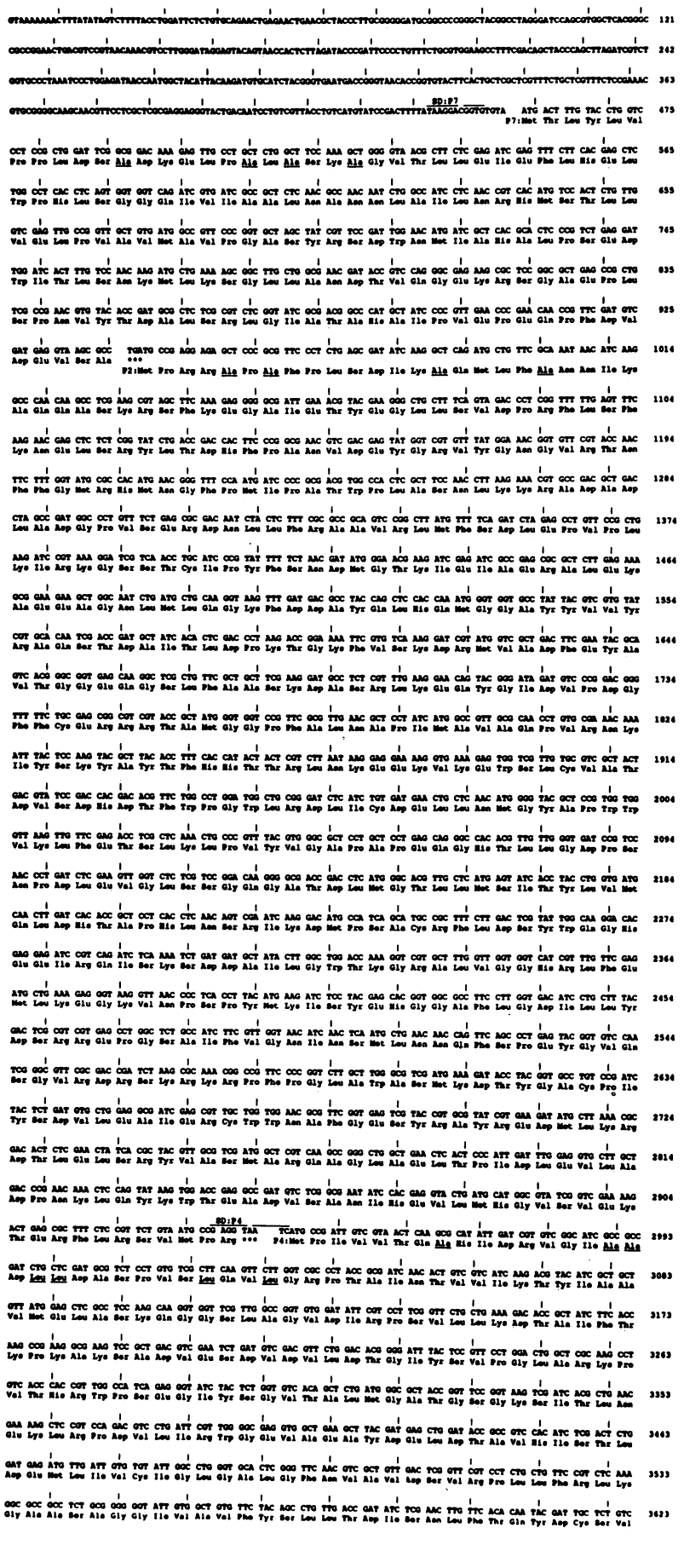
The large genomic segment of ϕ6 contains 6,374 nucleotides and it has 55.5% guanine-cytosine. Four reading frames were found that correspond to the four proteins known to be coded for by this segment (Fig. 1 and 2). As with the other two genomic segments, there were appreciable noncoding sequences at the termini. The termini of the segment agree with the sequence found by Iba et al. (9) by direct sequencing of the genomic RNA. The order of the genes is in agreement with that found by Revel et al. (21) on the basis of the protein synthesis competence of fragments on plasmids containing T7 promoters. The salient results of the sequence analysis are as follows.

Gene 7 is the first significant open reading frame in the segment. The reading frame starts with an AUG at position 458 and terminates at position 940 with a UGA. There is a Shine-Dalgarno sequence preceding the initiation codon (Fig. 2). The calculated molecular mass of P7 is 17.3 kilodaltons. The amino-terminal methionine is partially absent on the basis of amino acid sequencing. The sequence predicts an absence of cysteine, and this was confirmed by the lack of labeling of P7 in vivo when ϕ6-infected cells were exposed to radioactive cysteine (results not shown). P7 is predicted to have a high net negative charge, and this is consistent with the behavior of the protein in isoelectric focusing gels (17).

Gene 2 follows gene 7. The initiating codon begins at position 943 with an AUG that overlaps the UGA termination codon of gene 7. Gene 2 does not have a discernable Shine-Dalgarno sequence in the vicinity of the initiating codon. The reading frame ends at position 2937. The termination codon is UAA. The calculated molecular mass of P2 is 74.8 kilodaltons, and the N-terminal methionine is lost. The predicted amino acid composition indicates a basic protein, consistent with its behavior in isoelectric focusing gels (17).

Gene 4 follows gene 2. It has a Shine-Dalgarno sequence, and the initiating codon is an AUG beginning at position 2943. The reading frame terminates at position 3938. The termination codon is UAA. The calculated molecular mass of P4 is 35.0 kilodaltons, and the N-terminal methionine is lost.

Gene 1 follows gene 4. It has a Shine-Dalgarno sequence, the initiating codon is an AUG beginning at position 3951, and the reading frame terminates at position 6257 with UAA. It is notable that gene 1 ends only 117 base pairs from the end of the segment. This is much closer than those for the small



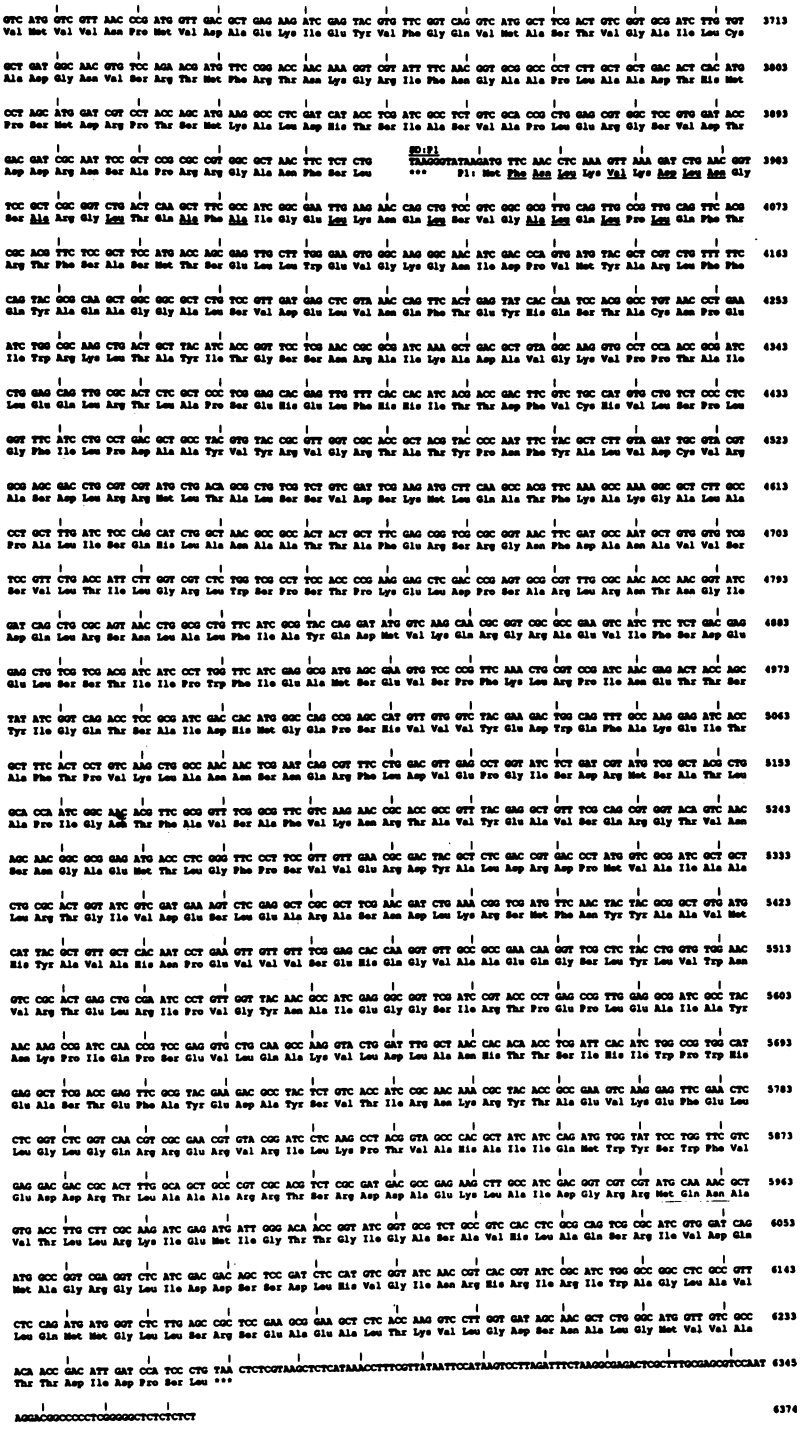


FIG. 2. Sequence of cDNA of genomic segment L of ϕ6. The sequence is that of the + strand. The translation products of the four known genes are shown. The underlined amino acids were determined by sequencing of unlabeled proteins or proteins labeled with either tritiated leucine or alanine. The Shine-Dalgarno (SD) sequences preceding the genes are overlined. The asterisks indicate stop codons.

or middle genomic segments. The calculated molecular mass of P1 is 85.0 kilodaltons. The N-terminal methionine remains on the protein. The predicted composition of P1 is consistent with the basicity found by isoelectric focusing (17).

DISCUSSION

The four proteins coded for by the large genomic segment are P1, P2, P4, and P7. These proteins are synthesized early in infection (25) and assemble to form the viral procapsid (16). The procapsid is probably both the transcriptase and the replicase of the virus. It has been shown that the procapsid structure is in the form of an open dodecahedron (10, 19, 28). An analysis of the stoichiometry of the ϕ6 virion (4) suggested that the procapsid is composed of 120 molecules each of P1 and P4, about 20 molecules of P2, and between 80 and 100 molecules of P7. Although the results of

this report change the predicted molecular weights for the proteins, the changes are not great enough to alter the expected stoichiometry of the particle significantly. Recent studies (10, 19) suggest that P1 itself can form a dodecahedral structure. It is not known whether the active site(s) for polymerase activity is shared by different proteins or whether it is located within one of the four proteins of the procapsid; however, no RNA polymerase activity has been demonstrated in any structure simpler than an RNA-filled procapsid.

A comparison of the amino acid sequences of proteins P1, P2, P4, and P7 against the Dayhoff protein sequence library as of April 1987 showed no significant similarity to other proteins. However, a search for sequences suggestive of nucleotide-binding sites revealed that P4 contains sequences highly similar to those found for a class of proteins that includes the multiple drug resistance proteins of tumor cells, transport proteins of bacteria, and UV repair proteins (Fig. 3) (7, 8, 27). The meaning of this site in P4 is not clear. Since the procapsid is involved in RNA polymerization and packaging and must itself be assembled, we could speculate that the P4 site is involved in one of these processes. This sequence is, however, not characteristic of polymerases. Recent unpublished work in our laboratory indicates that P4 has nucleotide triphosphate phosphatase activity.

Iba et al. (9) showed that there is identity among the ends of the three genomic segments for 17 bases at the 3′ end and for 18 bases at the 5′ end, with the exception of a difference in the second position. We have shown that the identity at the 3′ end is more extensive between the small and medium segments (Gottlieb et al., unpublished data). We now show that this is true for the large segment as well. Although the exact identity at the 3′ end stops at nucleotide 18, there is overall similarity that continues until 80 nucleotides from the 3′ end (Fig. 4). It appears reasonable that the regions of similarity should be involved in the regulation and mechanisms of transcription, replication, and genome packaging.

During infection, production of proteins P1, P4, and P7 is almost identical on a molar basis. Protein P2 is produced at about 10% of the rate of the others (25). Nonsense mutations in gene 7 are completely polar on the production of P2 (11). It appears that the mechanism of control of the synthesis of P2 is translational coupling. The same motif has been described for two other gene pairs in ϕ6. Production of P12 is dependent on ribosome loading on gene 8, which is immediately upstream, and production of P5 is dependent on ribosome loading on gene 9 (12). In both of these cases, the downstream product is synthesized at about 10% of the rate of the upstream product, and nonsense mutations in the upstream genes are completely polar on the production of the downstream gene product. In the three cases, the ribosome-binding site upstream from the amino acid initiating codon is either absent or spaced far away. The ϕ6 system also uses other means of control of gene expression. The production of proteins coded on the middle segment differs among the genes, although each gene has its own ribosome-binding site. It is likely that the strength of the interaction between ribosomes and some of the messages determines the level of production of individual proteins. In addition, there is a dramatic change in the pattern of single-stranded RNA synthesis during infection (3). This is certainly involved in the turn on of translation of the late proteins.

```
RWPSEGIYSGVTALMGATGSGKSITLNE          P4      111-138
          ||  |   |  |||  |
GLNLKVKSGQTVALVGNSGCGKSTTVQL          Mdr     411-438
NINLSIKQGEVIGIVGRSGSGKSTLTKL          HylB    488-505
DLNFTLRAGETLGIVGESGSGKSQSRLR          OppD     40--57
DINLDIHEGEFVVFVGPSGCGKSTLLRM          MalK     21--48
GVSLQARAGDVISIIGSSGSGKSTFLRC          HisP     24--51
NINLDIAKNQVTAFIGPSGCGKSTLLRT          PstB     28--55
NINLVIPRDKLIVVTGLSGSGKSSLAFD          UvrA-1   16--43
```

FIG. 3. Comparison of a portion of the amino acid sequence of protein P4 with those of a group of proteins with nucleotide-binding sites. The sequences are derived from the genes for the multiple drug resistance protein (Mdr), hemolysin transport protein (HylB), oligopeptide permease (OppD), and transport proteins for maltose (MalK), histidine (HisP), and phosphate (PstB) (7). A sequence from an ATP-dependent UV repair protein (UvrA-1) (5) is also shown.

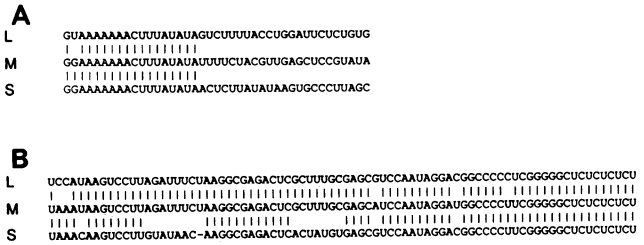


FIG. 4. Comparison of the 5′ (A) and 3′ (B) ends of the three genomic segments. In all cases, the 5′ end is at the left.

A number of viruses contain segmented genomes of double-stranded RNA. These include reovirus, rotavirus, blue tongue virus, and ϕ6. In addition, a number of particles that replicate double-stranded RNA have been described, notably, the killer particle in *Saccharomyces cerevisiae* (6). Although the replication of ϕ6 RNA differs from most of these in that transcription of ϕ6 is by strand displacement, whereas the others use a conservative mechanism, we anticipate that there is considerable homology in the proteins of these particles. In particular, we expect to find that the replicase proteins are related.

The complete nucleotide sequence of the three genomic segments has been entered into the GenBank database. The accession numbers are M17461 for segment L, M17462 for segment M, and M12921 for segment S.

ACKNOWLEDGMENTS

Parts of this work were supported by Public Health Service grant GM 34352 from the National Institutes of Health to L.M. I.N. was supported in part by National Research Service award 5 T32 AI-07180 from the National Institute of Allergy and Infectious Diseases. M.R. was supported in part by the Academy of Finland during his tenure at the Public Health Research Institute.

LITERATURE CITED

1. **Bamford, D. H., M. Romantschuk, and P. J. Somerharju.** 1987. Membrane fusion in procaryotes: bacteriophage ϕ6 membrane fuses with the *Pseudomonas syringae* outer membrane. EMBO J. **6**:1467–1473.
2. **Clewell, D. B.** 1972. Nature of Col E1 plasmid replication in *Escherichia coli* in the presence of chloramphenicol. J. Bacteriol. **110**:667–676.
3. **Coplin, D. L., J. L. Van Etten, R. K. Koski, and A. K. Vidaver.** 1975. Intermediates in the biosynthesis of double-stranded ribonucleic acids of bacteriophage ϕ6. Proc. Natl. Acad. Sci. USA **72**:849–853.
4. **Day, L. A., and L. Mindich.** 1980. The molecular weight of bacteriophage ϕ6 and its nucleocapsid. Virology **103**:376–385.
5. **Doolittle, R. F., M. S. Johnson, I. Husain, B. Van Houten, D. C. Thomas, and A. Sancar.** 1986. Domainal evolution of a prokaryotic DNA repair protein and its relationship to active-transport proteins. Nature (London) **323**:451–453.
6. **Fujimura, T., R. Esteban, and R. B. Wickner.** 1986. *In vitro* L-A

double-stranded RNA synthesis in virus-like particles from *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **83**: 4433–4437.
7. **Gros, P., J. Croop, and D. Housman.** 1986. Mammalian multidrug resistance gene: complete cDNA sequence indicates strong homology to bacterial transport proteins. Cell **47**:371–380.
8. **Higgins, C. F., I. D. Hiles, J. A. Downie, I. J. Evans, I. B. Holland, L. Gray, S. D. Buckel, A. W. Bell, and M. A. Hermodson.** 1986. A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. Nature (London) **323**:448–450.
9. **Iba, H., T. Watanabe, Y. Emori, and Y. Okada.** 1982. Three double-stranded RNA genome segments of bacteriophage φ6 have homologous terminal sequences. FEBS Lett. **141**:111–115.
10. **Ktistakis, N. T., and D. Lang.** 1987. The dodecahedral framework of the bacteriophage φ6 nucleocapsid is composed of protein P1. J. Virol. **61**:2621–2623.
11. **Lehman, J. F., and L. Mindich.** 1979. The isolation of new mutants of bacteriophage φ6. Virology **97**:164–170.
12. **McGraw, T., L. Mindich, and B. Frangione.** 1986. Nucleotide sequence of the small double-stranded RNA segment of bacteriophage φ6: novel mechanism of natural translational control. J. Virol. **58**:142–151.
13. **McGraw, T., H. Yang, and L. Mindich.** 1983. Establishment of a physical and genetic map for bacteriophage PRD1. Mol. Gen. Genet. **190**:237–244.
14. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–80.
15. **Mindich, L.** 1978. Bacteriophages that contain lipid, p. 271–335. *In* H. Fraenkel-Conrat, and R. R. Wagner (ed.), Comprehensive virology, vol. 12. Plenum Publishing Corp., New York.
16. **Mindich, L., and R. Davidoff Abelson.** 1980. The characterization of a 120S particle formed during φ6 infection. Virology **103**:386–391.
17. **Mindich, L., and J. Lehman.** 1983. Characterization of φ6 mutants that are temperature sensitive in the morphogenetic protein P12. Virology **127**:438–445.
18. **Mindich, L., G. MacKenzie, J. Strassman, T. McGraw, S. Metzger, M. Romantschuk, and D. Bamford.** 1985. cDNA cloning of portions of the bacteriophage φ6 genome. J. Bacteriol. **162**:992–999.
19. **Olkkonen, V. M., and D. H. Bamford.** 1987. The nucleocapsid of the lipid-containing double-stranded RNA bacteriophage φ6 contains a protein skeleton consisting of a single polypeptide species. J. Virology **61**:2362–2367.
20. **Partridge, J. E., J. L. Van Etten, D. E. Burbank, and A. K. Vidaver.** 1979. RNA polymerase activity associated with bacteriophage φ6 nucleocapsid. J. Gen. Virol. **43**:299–307.
21. **Revel, H. R., M. E. Ewen, J. Busslan, and N. Pagratis.** 1986. Generation of cDNA clones of the bacteriophage φ6 segmented dsRNA genome: characterization and expression of L segment clones. Virology **155**:402–417.
22. **Sanger, F., A. R. Coulson, G. F. Hong, D. F. Hill, and G. B. Petersen.** 1982. Nucleotide sequence of bacteriophage λ DNA. J. Mol. Biol. **162**:729–773.
23. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
24. **Sinclair, J. F., and L. Mindich.** 1976. RNA synthesis during infection with bacteriophage φ6. Virology **75**:209–217.
25. **Sinclair, J. F., A. Tzagoloff, D. Levine, and L. Mindich.** 1975. Proteins of bacteriophage φ6. J. Virol. **16**:685–695.
26. **Vidaver, A. K., R. K. Koski, and J. L. Van Etten.** 1973. Bacteriophage φ6: a lipid-containing virus of *Pseudomonas phaseolicola*. J. Virol. **11**:799–805.
27. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gray.** 1982. Distantly related sequences in the α and β subunits of ATP synthetase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1**:945–951.
28. **Yang, Y., and D. Lang.** 1984. Electron microscopy of bacteriophage φ6 nucleocapsid: three-dimensional image analysis. J. Virol. **51**:484–488.