Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 699–703, February 1984
Biochemistry

Major coat protein and single-stranded DNA-binding protein of filamentous virus Pf3

(bacteriophage of *Pseudomonas aeruginosa* PAO1(RP1)/DNA and protein sequences/signal sequence/DNA and virus structure/molecular evolution)

DEBRA GLUCK PUTTERMAN*, ARTURO CASADEVALL*, PAULA D. BOYLE*, HUEY-LANG YANG*, BLAS FRANGIONE†, AND LOREN A. DAY*

*The Public Health Research Institute of the City of New York, 455 First Avenue, New York, NY 10016; and, †Department of Pathology, New York University School of Medicine, New York, NY 10016

Communicated by Sarah Ratner, October 12, 1983

ABSTRACT **The region of the Pf3 virus genome encoding its major coat protein and its single-stranded DNA-binding protein is organized somewhat like the corresponding region of the fd (M13, f1) genome. Nevertheless, the major coat protein is unique among the major coat proteins of fd and the other filamentous phages studied in that it lacks a signal sequence and appears to be a direct translation product and in that it has fewer basic amino acid residues than its equivalent of DNA phosphates in the virion. These features are relevant to considerations of both protein insertion into membranes and DNA structure in filamentous viruses. The single-stranded DNA-binding protein also has a sequence that is different from the sequences of single-stranded DNA-binding proteins from other filamentous viruses.**

Bacteriophage Pf3 infects *Pseudomonas aeruginosa* PAO1 bearing the RP1 plasmid (1). It is one of a variety of filamentous viruses of Gram-negative bacteria, the best known of which are the closely related fd, f1, and M13 phages, all of which infect male strains of *Escherichia coli*. Other filamentous phages are Pf1 (*P. aeruginosa*, strain K), Xf (*Xanthomonas oryzae*), and IKe and If1, both of which grow on *E. coli* strains bearing N and I plasmids, respectively. These viruses each contain a single-stranded circular DNA molecule encased in a slender protein sheath composed of thousands of identical protein subunits of molecular weight about 5,000 (the major coat protein) and a few copies each of a few coat proteins located at the ends of the virions (the minor coat proteins). Various comparative physical studies have shown that the packing of the DNA inside the different viruses is not the same, and also that intracellular complexes between the DNAs and their single-stranded DNA (ss DNA)-binding proteins are different (2-13).

The best-characterized of these systems is that of fd (f1, M13), and it serves as the basis for comparisons (see ref. 12 for reviews). Of particular relevance to the present study is that the major coat protein subunits of fd span the membrane of the infected cell before they form a coat around the viral DNA. Their insertion into the membrane involves the processing of a signal sequence from a precursor protein. During virus assembly, which occurs at the membrane, the viral DNA leaves its complex with its single-stranded DNA-binding protein as it becomes covered by major coat protein subunits and is extruded into the medium. The assembly and extrusion process does not cause cell lysis.

In this paper, we present the results of an investigation of a region on the genome of Pf3 that encodes the DNA-binding protein and the coat protein of this virus. Comparisons of this region of DNA in Pf3 with its fd counterpart indicate close similarities in overall organization yet quite different amino acid sequences for the structural proteins themselves and the absence of a signal sequence for the major coat protein.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Bacteria and Bacteriophage. Pf3 bacteriophage and its host *P. aeruginosa* (PAO1) bearing the RP1 plasmid (1) were obtained originally from D. E. Bradley. Phage Pf3 was propagated and purified from single-plaque isolates (3, 4). A standard medium containing 10 g of tryptone, 1 g of yeast extract, 8 g of NaCl, 1 g of glucose, and 220 mg of $CaCl_2$ per liter was used.

DNA. Covalently closed circular duplex Pf3 DNA [replicative form (RF) DNA] was isolated from Pf3-infected *P. aeruginosa*, and DNA restriction fragments of it were prepared by digestions with the appropriate enzymes (14, 15). When mixtures of fragments were used, the restriction digest was simply precipitated with ethanol and then suspended in 10 mM Tris acetate, pH 7.5. Individual fragments were separated by electrophoresis on low-melting-point agarose (Bethesda Research Laboratories; Marine Colloids, Rockland, ME), followed by passage over DEAE-cellulose (Whatman DE-52) with elution by 2 M NaCl/1 mM EDTA/10 mM Tris·HCl, pH 7.5. The DNA used for sequence analysis was a recombinant plasmid consisting of the largest EcoRI fragment (3.8 kilobases) of Pf3 DNA inserted into the cloning vector pBD214 as described (14, 15).

Isolation of DNA-Binding Protein. The ss DNA-binding protein of Pf3 was isolated on DNA-agarose columns similar to columns used to isolate the gene *5* protein of fd (16) or by gel filtration of precursor complexes disrupted in 2 M NaCl buffer (13).

DNA-Directed Protein Synthesis *In Vitro*. Cell-free extracts derived from a *recB* mutant strain of *E. coli* were used for *in vitro* protein syntheses as described (17), except that Triton X-100 at a final concentration of 0.2% was included in the incubation mixtures. Tritiated leucine was the radioactive label.

$NaDodSO_4$/Polyacrylamide/Urea Gels. Proteins were analyzed by electrophoresis according to a modified protocol from Bethesda Research Laboratories. A resolving gel of 15% acrylamide (0.4% bisacrylamide), in 6 M urea/0.1% SDS $NaDodSO_4$ buffered by 0.1 M sodium phosphate, pH 7.2, and a stacking gel of 3.5% acrylamide in the same buffer were used. The running buffer was 0.1 M sodium phosphate, pH 7.2, and 0.1% $NaDodSO_4$. Samples were heated in and loaded onto the gels in 6 M urea/1% $NaDodSO_4$/1% mercaptoethanol. Gels were stained in 0.1% Coomassie blue. Pro-

Abbreviations: ss DNA, single-stranded DNA; RF, replicative form.

teins synthesized *in vitro* were detected by autoradiography performed as described (17). Labeled and unlabeled molecular weight standards were protein mixtures obtained from Bethesda Research Laboratories.

Determination of Primary Structure. The amino acid sequence of the Pf3 major coat protein was determined by a combination of (*i*) automated Edman degradation and (*ii*) determination of the amino acid sequence of peptides obtained from trypsin and subtilisin digestion (18, 19). The intact virus was used for both procedures. Amino acid analysis was carried out on a Durrum D 500 analyzer. Sequence analysis was by automated Edman degradation with a Beckman 890C sequencer (20) with 0.1 M Quadrol. Phenylthiohydantoins were identified by high-pressure liquid chromatography (Waters Associates) with a μBondapak C_{18} column and by amino acid analysis after hydrolysis of the phenylthiohydantoin derivatives (20). Tryptic and subtilisin peptides were isolated by high-voltage electrophoresis (21). Manual sequence analysis was performed as described (22).

DNA Sequence Analysis. The method of Maxam and Gilbert (23) was used, with one modification used by several laboratories: limited DNA cleavage at guanines and adenines was accomplished by incubation of labeled DNA in 67% (wt/wt) formic acid at 20°C for 5 min, with subsequent reaction steps identical to those described by Maxam and Gilbert for limited DNA cleavage at pyrimidines.

RESULTS

Protein Chemistry. The major coat protein of Pf3 has an apparent molecular weight of only ≈4,000 (Fig. 1). Under the conditions used, fd protein has a mobility corresponding to its known molecular weight of 5,240. Overloading of the gel produced aggregates of the major coat protein subunits and retarded migration rates. A faint band at ≈50,000 M_r could be clearly discerned in all five lanes of the original gels of Pf3 virus, with decreased intensity corresponding to decreased input concentrations. This band was assigned as a minor coat protein component of the virion. From the relative intensities it was surmised that the major coat protein constitutes more than 95% of the total protein in the virus, and, according to molecular weight studies, there are approximately 2,500 subunits per virion (3, 4, 14). The ss DNA-binding protein–DNA complexes, isolated from Pf3-infected cells, gave a band at ≈6,000 M_r when disrupted and electrophoresed under the same conditions (13).

The amino acid sequence of the Pf3 major coat protein is

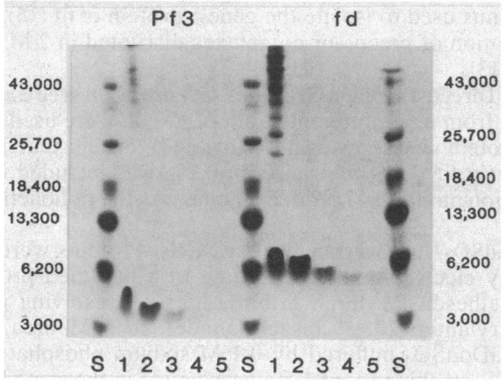

FIG. 1. Electrophoresis of disrupted Pf3 and fd viruses under conditions for low molecular weight proteins. The lanes labeled S contain molecular weight standards. Lanes 1–5 for each virus are for input concentrations, in μg/ml, of 1250, 400, 125, 40, and 12.5, respectively. Samples loaded in lanes 3, 4, and 5 contained 6 M urea, but samples loaded in lanes 1 and 2 contained only 1.5 M and 4.5 M urea for Pf3 and 3M and 5.5 M urea for fd, respectively.

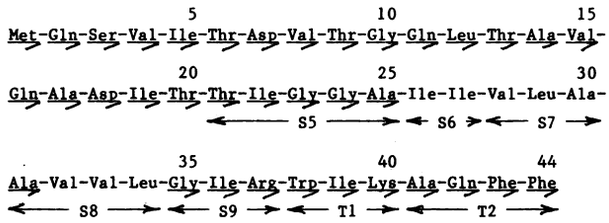

FIG. 2. Amino acid sequence of the Pf3 major coat protein. →, Positions obtained by sequence determination; T, tryptic peptides; S, subtilisin peptides.

shown in Fig. 2. The proposed sequence was deduced as follows. Samples of intact virus were first submitted to automated Edman degradation with a liquid-phase sequencer, which gave the sequence of 20 residues from the NH_2 terminus. Two peptides were obtained after trypsin digestion (T_1 and T_2), and the insoluble material was redigested with subtilisin and relevant peptides were purified and partially sequenced. Their positions from residue 26 were ordered by using the DNA sequence.

The NH_2 terminus of the DNA-binding protein was subjected to automated Edman degradation. The sequence obtained was: Met-Asn-(Leu or Ile)-Gln-(Leu or Ile)-Thr-Phe-Thr-Asp-Ser-(Val or Met)-Arg-Gln-Gly-.

Location of the Genes for the Coat Protein and the ss DNA-Binding Protein. An *in vitro* system for coupled transcription and translation of DNA templates produced several proteins when programmed with double-stranded Pf3 DNA (5.8 kilobases). Three proteins were synthesized in large amounts and their molecular weights estimated from gel migration were nominally 4,000, 5,000, and 6,000 (Fig. 3, lane 3). The smallest of these (the M_r 4,000 band) migrates with the major coat protein from the virus itself (Fig. 1), so the gene producing it became a candidate for the major coat protein gene. The genes producing the bands at M_r ≈5,000 and ≈6,000 were also candidates because the analogous protein in fd is post-translationally processed. Part or all of the band of M_r ≈6,000 was expected to be the ss DNA-binding protein (13). The intensity of the M_r 4,000 band was dramatically reduced, virtually to zero, by using either Pf3 RF DNA cleaved with *Hin*dIII or the largest fragment of Pf3 DNA produced by *Eco*RI after further cleavage by *Hin*dIII (Fig. 3,

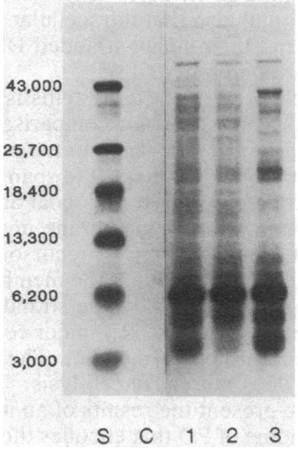

FIG. 3. Autoradiograph of the products of an *in vitro* translation reaction. Lane S shows molecular weight standards. Lane C shows the control with no DNA added. The DNAs used to program the system were lane 1, large (3.8-kilobase) *Eco*RI fragment from Pf3; lane 2, large *Eco*RI fragment cleaved by *Hin*dIII; lane 3, intact supercoiled Pf3 RF DNA.

lane 2). *Hin*dIII cleaves Pf3 DNA at only one site (15).

Nucleotide sequence determinations extending in both directions from the *Hin*dIII site revealed a sequence starting 120 bases away that encoded a protein having an NH_2-terminal region of 25 amino acids and a COOH-terminal region of 10 amino acids that were identical to those obtained from direct amino acid sequence determinations on the protein from the virus. In addition, the sequence of the nine intervening amino acids given by the DNA sequence could be reconciled with peptides from this region. We therefore concluded the DNA sequence was that of the gene for the major coat protein. The protein is composed of 44 amino acids and has a molecular weight of 4,630. The search for the Pf3 ss DNA-binding protein gene was focused on the region between the major coat protein gene and the *Eco*RI recognition site about 600 bases upstream from it, on the basis of the gene order in the fd genome. Sequence determinations in this region revealed an open reading frame for a protein having an NH_2-terminal sequence in agreement with the results above and an amino acid composition in agreement with direct analyses (13). The protein consists of 78 amino acids and has a molecular weight of 8,907, somewhat higher than expected on the basis of its electrophoretic mobility. The sequence of 900 bases spanning the genes for the two proteins is presented in Fig. 4.

```
                                                                    60
AATTCTTATGATTGCTCACAATATCTGTCATCCGCCTCGCCCTCACTGTCAGCTCAAGCT

                                                                   120
CTTATGCTTGCTCAATGCCAAAGCTGTGTCTTACTTTATGGTGGGCACTGACGATAAAAT

                                                                   180
CGAAGCTGAGGTTTGGGTTTTTGCTGATGAATCCGTTGCTGTTGCTCGCCCAGGTCAGCC

                                                                   240
TGTAAAGGCTTACCATACCATTCAATCTGTTTGCTTCCATTGAGGTGCATTTTATGAACA
                                                          MetAsnIle

                                                                   300
TTCAAATTACCTTTACCGATTCCGTCCGCCAGGGTACTTCCGCTAAAGGGAATCCTTACA
  GlnIleThrPheThrAspSerValArgGlnGlyThrSerAlaLysGlyAsnProTyrThr

                                                                   360
CGTTCCAAGAAGGGTTCTTGCATTTGGAAGATAAGCCCTTTCCTCTCCAGTGCCAGTTCT
  PheGlnGluGlyPheLeuHisLeuGluAspLysProPheProLeuGlnCysGlnPhePhe

                                                                   420
TTGTTGAGTCGGTCATTCCTGCCGGTTCTTATGAGGTGCCGTATCGCATTAACGTGAACA
  ValGluSerValIleProAlaGlySerTyrGlnValProTyrArgIleAsnValAsnAsn

                                                                   480
ATGGGCGTCCTGAATTGGCTTTTGATTTTAAAGCCATGAAGCGTGGCGTAATGAGTTATTA
  GlyArgProGluLeuAlaPheAspPheLysAlaMetLysArgAla******

                                                                   540
TGTTTGTGTTCAGCTTGTTAATGATGTTTGTCATGAATGGGCTGAACGTTCAGATTGCTA

                                                                   600
AGCTTGCCGGAAGGTTCAGGCTTGCAAATTGGCGGGATGTTATTGCTACTTTCCGCCACC
                                        MetLeuLeuLeuLeuSerAlaThr

                                                                   660
GCTTGGGGTATTCAACAAATAGCCCGTTTACTTTTAAATCGTTGATGAGGTGTCTTTTAT
AlaTrpGlyIleGlnGlnIleAlaArgLeuLeuLeuAsnArg******
                    ***              ***IleValAspGluValSerPheMet

                                                                   720
GCAATCCGTGATTACTGATGTGACAGGCCAACTGACAGCCGGTGCAAGCTGATATCACTAC
 GlnSerValIleThrAspValThrGlyGlnLeuThrAlaValGlnAlaAspIleThrThr

                                                                   780
CATTGGTGGTGCTATTATTGTTCTGGCCGCTGTTGTGCTGGGTATTCGCTGGATCAAAGC
 IleGlyGlyAlaIleIleValLeuAlaAlaValValLeuGlyIleArgTrpIleLysAla

                                                                   840
GCAATTCTTTTGATCCGTCCTTGGGCTTTTGGCCTCAATCGTTATAAGGGGGCTTCGGCT
 GlnPhePhe***                                        Transcription

                                                                   900
CCCTTATTCGTTTAGCGGCTAAAATTTTTCAATTCACGGGGCTTTTATGGAGATTATGGA
Terminator
```

FIG. 4. Region of the Pf3 genome encoding the ss DNA-binding protein (nucleotides 234–467, underlined) and the major coat protein (nucleotides 659–790, underlined), with nucleotide numbering beginning at an *Eco*RI cleavage site (nucleotides 1–5 plus a preceding G). Boxes indicate the −35 and −10 regions of assigned promoters and the terminator at nucleotides 824–853. The parentheses around nucleotides 475 and 476 indicate that there may be more than two bases, but this uncertainty does not affect the conclusions drawn. Asterisks indicate termination codons.

DISCUSSION

Coat Protein Synthesis Without a Signal Sequence. The NH_2-terminal amino acid of the coat protein in the virion is methionine, encoded by ATG at nucleotide 659. AUG is the usual codon for translational initiation. The nearest ATG codon upstream from the protein and in the same reading frame occurs at nucleotide 470, but intervening translational stop codons occur at nucleotides 539, 620, and 635. A good Shine–Dalgarno ribosome recognition sequence, U-G-A-U-G-A-G-G-U-G, is encoded seven bases upstream from the ATG codon at nucleotide 659. It shows extensive complementarity with the 3′ termini of 16S ribosomal RNAs in *P. aeruginosa*, A-U-U-C-C-U-C-U-C-Y (Y = pyrimidine nucleoside) and in *E. coli* A-U-U-C-C-U-C-C-A-C-Y (24). A good Shine–Dalgarno sequence would be expected for a protein made in large amounts such as a major structural protein. The codon GUG, sometimes used to initiate translation, occurs at nucleotide 650. Translational initiation at this GUG would have to take place with a poor Shine–Dalgarno sequence and involve subsequent cleavage of three amino acids to leave an NH_2-terminal methionine. We do not consider this a likely possibility, and three amino acids would, in any case, not constitute a signal sequence of the types involved in membrane insertion mechanisms (25, 26). On the basis of the data and these arguments, we conclude that the protein is synthesized without a signal sequence and propose that the mature protein is a direct production of translation.

All the major coat proteins of the other filamentous viruses whose sequences have been determined (Fig. 5) start with an amino acid other than methionine and hence appear to be products of post-translational processing. All have hydrophobic central domains, a characteristic of intrinsic membrane proteins. All of the viruses are extruded into the medium without cell lysis, implying similar assembly mechanisms. It seems probable that all the coat proteins are, as in fd, intrinsic membrane proteins prior to virus assembly. Because of the important role of the membrane and because of the important role of post-translational processing of a signal sequence in membrane insertion in the fd (f1, M13) system (35–38), we find it interesting that the Pf3 coat protein is not synthesized with a signal sequence and does not appear to be post-translationally processed.

Coat Protein Sequence and DNA Structure in the Virion. On average there are 2.4 nucleotides present in the virion per subunit (4, 39), yet there are only two basic amino acids in the subunit. This situation is unique to Pf3 virus because the major coat proteins of the other viruses have more basic residues than their equivalent number of nucleotides in the virion. The way the phosphate charges are neutralized in Pf3 virions is therefore expected to be different from the way they are neutralized in the other systems (Fig. 5). Another unique feature is the presence of two phenylalanines at the COOH terminus. Circular dichroism and absorbance measurements indicate the absence of base–base stacking in the virion (6, 7), and Raman spectral analysis suggests an unusual DNA backbone structure (8). The amino acid sequence combined with the spectral data and with theoretical considerations embodied in "the pitch connection" (39) lead us to propose that the DNA structure in the Pf3 virion is an inside-out or inverted helix (I-DNA) with phosphates in and bases out. The bases could interact with the two phenylalanines and the nearby tryptophan.

In spite of the large differences in sequence, especially in the basic COOH-terminal portions, significant amino acid sequence homology occurs among the coat proteins listed in Fig. 5. The partial homologies between Pf3, Xf, and Pf1 are of particular interest. As examples, sequence matches between Pf3 and Xf can be seen simply by aligning methionine-1 of Pf3 and glycine-2 of Xf; among the first 25 positions

VIRUS	AMINO ACID SEQUENCE	n/s
	- - - + - + ++ +	
IKe	AEPNAATNYATEAMDSLKTQAIDLISQTWPVVTTVVVAGLVIRLFKKFSSKAV	(?)
	-- + - - + ++ +	
If1	ADDATSQAKAAFDSLTAQATEMSGYAWALVVLVVGATVGIKLFKKFVSRAS	(?)
	- -- + - - + ++ +	
fd, f1 (ZJ/2) [M13]	AEGDDPAKAAFDSLQASATEYIGYAWAMVVVIVGATIGIKLFKKFTSKAS [N] (A)	(2.4)
	- - + ++	
Pf1	GVIDTSAVQSAITDGQGDMKAIGGYIVGALVILAVAGLIYSMLRKA	(1.0)
	- - - + + ++	
Xf	SGVGDGVDVVSAIEGAAGPIAAIGGAVLTVMVGIKVYKWVRRAM	(2.0)
	- - + +	
Pf3	MQSVITDVTGQLTAVQADITTIGGAIIVLAAVVLGIRWIKAQFF	(2.4)

FIG. 5. Comparison of Pf3 coat protein with the coat proteins of eight other filamentous bacterial viruses, as drawn from references 19 and 27–34. The one-letter convention is used, with the positions of aromatic side chains emphasized by underlining and those of basic and acidic groups by + and − signs above the letters even though the states of protonation and charge are not known. The number of DNA phosphate groups present per subunit in each virion is given as nucleotides per subunit (n/s) ratios (2–4).

there are 8 with matches and 8 others with residues related by a single base change. For Pf3 and Pf1, alignment at methionine-1 of Pf3 and valine-8 of Pf1 yields, among the first 7 positions, 5 positions with matches and 2 others related by one base change. The amino acid sequences of Pf1 and Xf have direct homology of about 42% (40). Limited homologies are evident in simple pairwise comparisons between any of Pf1, Xf, and Pf3 with any fd, IKe, and If1, the latter three all being closely homologous.

In addition to the findings of this study, other data and theories support the existence of widely varying DNA structures in filamentous bacteriophages. The virion of Pf1 has an unusual DNA structure that is probably an I type, but quite different from the I-DNA structure expected for Pf3, and the DNA structures in both of these viruses must be vastly different from more normal, right-handed DNA helices with phosphates out and bases in proposed for fd, If1, IKe, and Xf (5–9, 39). In view of the structural variety, we find it intriguing that there are the extensive, yet partial, sequence homologies in the proteins that determine the DNA structures.

ss DNA-Binding Protein. The amino acid sequences are now available for the ss DNA-binding proteins of Pf3 (this study), fd and M13 (41, 42), IKE (43), and Pf1 (44). The proteins range in size from 78 amino acids (Pf3) to 144 amino acids (Pf1). As a group, they show far less homology than the major coat proteins (13, 43). All of the sequences begin with methionine. In the case of Pf3 the start codon is preceded six bases upstream by a good Shine–Dalgarno sequence, C-A-U-U-G-A-G-G-U-G, and a similar arrangement occurs for fd. Presumably none of the DNA-binding proteins is post-translationally processed, as would be consistent with the fact that they are not membrane proteins. Comparative data on the complexes between these proteins and their DNAs show a variety of DNA structures, with as much or more diversity as in the virions (10, 11, 13).

Genome Organization. The genes for the ss DNA-binding protein and the major coat protein are separated by about 185 nucleotides. Our data indicate an open reading frame for an amphipathic peptide of 22 amino acids which, because of the hydrophobic nature of its NH_2-terminal 16 residues, might interact with the membrane in some way. The peptide would have a reasonable Shine–Dalgarno sequence, but whether or not it is expressed is unknown. Thirty-one bases downstream from the major coat protein gene is the sequence A-T-A-A-G-G-G-G-G-C-T-T-C-G-G-C-T-C-C-C-T-T-A-T-T-C-G-T-T-T (nucleotides 824–853); with one G·U pair, the transcript from this sequence could form a stem and loop structure with a 10-base-pair stem and a 4-base loop that would be followed by a U-rich 3′ terminus, like those of ρ-independent terminators (45, 46). This overall organization is similar to that for fd, and block diagrams of these corresponding regions of the fd and Pf3 genomes are presented in Fig. 6. The promoters indicated for Pf3 in Figs. 4 and 6 have been assigned on the basis of the canonical RNA polymerase recognition sequence having a −35 region (T-G-T-T-G-A-C-A-A-T-T-T) and a −10 region (T-A-T-R-A-T-R; R = purine nucleoside) as established for *E. coli* and coliphages (48–50). The assigned promoters have reasonable positions with respect to the two known genes and the proposed terminator sequence. The separation of the Pf3 major coat protein gene from the three assigned promoters (through *Hin*dIII cleavage at nucleotide 540) virtually eliminates the *in vitro* production of coat protein (Fig. 3). Also, supercoiled Pf3 RF DNA yields a larger relative amount of the major coat protein, but about the same amount of the DNA-binding protein (data not shown), suggesting at least two transcriptional control units.

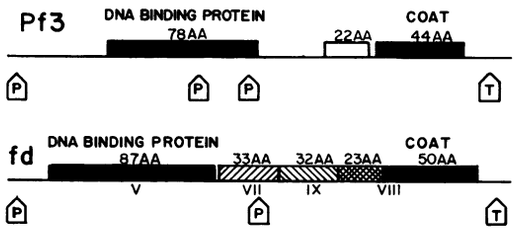

FIG. 6. Block diagram for a 781-base-pair portion of the Pf3 genome, as compared to the well-characterized analogous region of 790 base pairs on the fd genome (29, 31, 47). Transcription is from left to right, P indicates promoter site, T indicates transcription termination site, and AA indicates amino acids. The filled bars represent known proteins in both systems, the products of genes *V* and *VIII* in the case of fd. The hatched bars represent two of the known minor coat proteins of fd, the products of genes *VII* and *IX*. The open bar represents an open reading frame given by the Pf3 sequence. The cross-hatching indicates the signal sequence for fd, but there is no corresponding signal sequence for Pf3.

In fd, transcripts of different sizes all contain the coat protein sequence (50–53).

With respect to translational control, we find it interesting that the last seven bases of the two Shine–Dalgarno sequences proposed for the major coat protein and the ss DNA-binding protein are identical. Both sequences show extensive complementarity with the 3′ termini of *E. coli* and *P. aeruginosa* 16S RNA, consistent with strong expression both *in vivo* and in the *in vitro* system.

Molecular Evolution. All known filamentous phages have the same overall morphology and appear to have similar life cycles. The rather widely variant fd and Pf3 have similar organization over part, if not all, of their genomes (this study and ref. 14). These facts and the partial homologies in the sequences of the major coat proteins suggest that the variety of actual sequences observed is the result of divergent evolution from an ancestral host–virus system. This appears more likely than convergent evolution to partially homologous proteins. Divergent evolution of these systems implies the occurrence of evolutionary steps that have altered the structure of DNA within the virions, and that have introduced, or have eliminated, signal sequences on the major coat proteins.

Note Added in Proof. After this manuscript was submitted, a publication of the sequence of 415 bases spanning the major coat protein gene of Pf3 appeared (54). The T at our nucleotide 676 was adopted from the published results of Luiten *et al.* (54) because our data, in spite of sequence determinations on both strands, indicated an ambiguous C at this third position in a codon for threonine.

The authors are very grateful to E. Rosenwasser, F. Prelli, E. Lusskin, R. Smith, T. J. Gryczan and C. J. Lusty for their help in various aspects of this study. Financial support was through National Institutes of Health Grants AI09049 (to L.A.D.), 5-T32-CA09161 (to D.G.P.), and 5-T32-GM07308 (to A.C.) and other National Institutes of Health grants (to B.F.).

1. Stanisich, V. A. (1974) *J. Gen. Microbiol.* **84,** 332–342.
2. Day, L. A. & Wiseman, R. L. (1978) in *The Single-Stranded DNA Phages*, eds. Denhardt, D., Dressler, D. & Ray, D., (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) pp. 605–625.
3. Berkowitz, S. A. & Day, L. A. (1980) *Biochemistry* **19,** 2696–2702.
4. Newman, J., Day, L. A., Dalack, A. & Eden, D. (1982) *Biochemistry* **21,** 3352–3358.
5. Day, L. A., Wiseman, R. L. & Marzec, C. J. (1979) *Nucleic Acids Res.* **7,** 1393–1403.
6. Casadevall, A. & Day, L. A. (1982) *Nucleic Acids Res.* **10,** 2467–2481.
7. Casadevall, A. & Day, L. A. (1983) *Biochemistry* **22,** 4831–4842.
8. Thomas, G. J., Jr., Prescott, B. & Day, L. A. (1983) *J. Mol. Biol.* **165,** 321–356.
9. Cross, T. A., Tsang, P. & Opella, S. J. (1983) *Biochemistry* **22,** 721–726.
10. Pratt, D., Laws, P. & Griffith, J. (1974) *J. Mol. Biol.* **82,** 425–435.
11. Gray, C. W., Kneale, G. G., Leonard, K. R., Siegrist, H. & Marvin, D. A. (1982) *Virology* **116,** 40–52.
12. Denhardt, D. T., Dressler, D. & Ray, D. S. eds. (1978) in *The Single-Stranded DNA Phages* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
13. Casadevall, A. (1983) Dissertation (New York University, New York, NY).
14. Putterman, D. G. (1983) Dissertation (New York University, New York, NY).
15. Putterman, D. G., Gryczan, T. J., Dubnau, D. & Day, L. A. (1983) *J. Virol.* **47,** 221–223.
16. Alberts, B. M., Frey, L. & Delius, H. (1972) *J. Mol. Biol.* **68,** 139–152.
17. Yang, H.-L., Ivashkiv, L., Chen, H.-Z., Zubay, G. & Cashel, M. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 7029–7033.
18. Michaelson, T. E., Frangione, B. & Franklin, E. C. (1977) *J. Biol. Chem.* **252,** 883–889.
19. Frangione, B., Nakashima, Y., Konigsberg, W. & Wiseman, R. L. (1978) *FEBS Lett.* **96,** 381–389.
20. Frangione, B., Rosenwasser, E., Prelli, F. & Franklin, E. C. (1980) *Biochemistry* **19,** 4304–4308.
21. Frangione, B. & Milstein, C. (1968) *J. Mol. Biol.* **33,** 893–906.
22. Gray, W. R. (1972) *Methods Enzymol.* **25,** 333–343.
23. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
24. Shine, J. & Dalgarno, L. (1975) *Nature (London)* **254,** 34–38.
25. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67,** 835–851.
26. Emr, S. D., Hall, M. N. & Silhavy, T. J. (1980) *J. Cell Biol.* **86,** 701–711.
27. Snell, D. T. & Offord, R. E. (1972) *Biochem. J.* 127, 167–178.
28. Nakashima, Y. & Konigsberg, W. (1974) *J. Mol. Biol.* **88,** 598–600.
29. Bailey, G. S., Gillet, D., Hill, D. F. & Petersen, G. B. (1977) *J. Biol. Chem.* **252,** 2218–2225.
30. Beck, E., Sommer, R., Auerswald, E. A., Kurz, Ch., Zink, B., Osterburg, G., Schaller, H., Sugimoto, K., Sugisaki, H., Okamoto, T. & Takanami, M. (1978) *Nucleic Acids Res.* **5,** 4495–4503.
31. van Wezenbeek, P. M. G. R., Hulsebos, T. J. M. & Schoenmakers, J. G. G. (1980) *Gene* **11,** 129–148.
32. Hill, D. F. & Petersen, G. B. (1982) *J. Virol.* **44,** 32–46.
33. Nakashima, Y., Wiseman, R. L., Konigsberg, W. & Marvin, D. A. (1975) *Nature (London)* **253,** 68–71.
34. Nakashima, Y., Frangione, B., Wiseman, R. L. & Konigsberg, W. (1981) *J. Biol. Chem.* **256,** 5792–5797.
35. Chang, C. N., Blobel, G. & Model, P. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 361–365.
36. Ito, K., Mandel, G. & Wickner, W. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 1199–1203.
37. Wickner, W. (1980) *Science* **210,** 861–868.
38. Russel, M. & Model, P. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 1717–1721.
39. Marzec, C. J. & Day, L. A. (1983) *Biophys. J.* **42,** 171–180.
40. Doolittle, R. (1981) *Science* **214,** 149–159.
41. Cuypers, T., Van der Ouderaa, F. J. & de Jong, W. W. (1974) *Biochem. Biophys. Res. Commun.* **59,** 557–563.
42. Nakashima, Y., Dunker, A. K., Marvin, D. A. & Konigsberg, W. (1974) *FEBS Lett.* **43,** 125.
43. Peeters, B. P. H., Konings, R. N. H. & Schoenmakers, J. G. G. (1983) *J. Mol. Biol.* **169,** 197–215.
44. Maeda, K., Kneale, G. G., Tsugita, A., Short, N. J., Perham, R. N., Hill, D. F. & Petersen, G. B. (1982) *EMBO J.* **1,** 255–261.
45. Sugimoto, K., Sugisaki, H., Okamoto, T. & Takanami, M. (1977) *J. Mol. Biol.* **111,** 487–507.
46. Pribnow, D. (1979) in *Biological Regulation and Development, Gene Expression*, ed. Goldberger, R. F., (Plenum, New York), Vol. 1, pp. 219–277.
47. Beck, E. & Zink, B. (1981) *Gene* **16,** 35–58.
48. Pribnow, D. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 784–788.
49. Gilbert, W. (1976) in *RNA Polymerase*, eds. Losick, R. & Chamberlin, M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 193–205.
50. Takanami, M. K., Sugimoto, K., Sugisaki, H. & Okamoto, T. (1976) *Nature (London)* **260,** 297–302.
51. Okamoto, T., Sugiura, M. & Takanami, M. (1969) *J. Mol. Biol.* **45,** 101–111.
52. Seeburg, P. H. & Schaller, H. (1975) *J. Mol. Biol.* **92,** 261–277.
53. Edens, L., Konings, R. N. H. & Schoenmakers, J. G. G. (1978) *Virology* **86,** 354–367.
54. Luiten, R. G. M., Schoenmakers, J. G. G. & Konings, R. N. H. (1983) *Nucleic Acids Res.* **11,** 8073–8085.