JOURNAL OF VIROLOGY, Oct. 1982, p. 32–46
0022-538X/82/100032-15$02.00/0

Vol. 44, No. 1

Nucleotide Sequence of Bacteriophage f1 DNA

DIANA F. HILL AND GEORGE B. PETERSEN*
Department of Biochemistry, University of Otago, Dunedin, New Zealand

Received 22 February 1982/Accepted 17 June 1982

The nucleotide sequence of the DNA of the filamentous coliphage f1 has been determined. In agreement with earlier conclusions, the genome was found to comprise 6,407 nucleotides, 1 less than that of the related phage fd. Phage f1 DNA differs from that of phage M13 by 52 nucleotide changes, which lead to 5 amino acid substitutions in the corresponding proteins of the two phages, and from phage fd DNA by 186 nucleotide changes (including the single-nucleotide deletion), which lead to 12 amino acid differences between the proteins of phages f1 and fd. More than one-half of the nucleotide changes in each case are found in the sequence of 1,786 nucleotides comprising gene IV and the major intergenic region between gene IV and gene II. The sequence of this intergenic region (nucleotides 5501 to 6005) of phage f1 differs from the sequence reported by others through the inclusion of additional single nucleotides in eight positions and of a run of 13 nucleotides between positions 5885 and 5897, a point of uncertainty in the earlier published sequence. The differences between the sequence of bacteriophage f1 DNA now presented and a complete sequence for the DNA previously published by others are discussed, and the f1 DNA sequence is compared with those of bacteriophages M13 and fd.

Bacteriophage f1 (15, 48) belongs to a group of viruses of filamentous morphology that are specific for the F^+ and Hfr strains of *Escherichia coli*. In common with other filamentous bacteriophages (6), these viruses exhibit unusual mechanisms of replication and maturation and in many respects resemble endosymbiotic animal viruses rather than typical lytic bacteriophages. Bacterial host cell metabolism is not seriously impaired upon infection with bacteriophage f1, and the release of mature virus particles from the cell involves a continuous extrusion through the host cell membrane without concomitant lysis of the bacterium. The DNA component of the mature virion is a single-stranded, covalently closed molecule with a molecular weight of around 2×10^6. A double-stranded, covalently closed replicative form of the DNA (RFI DNA) may readily be isolated from bacterial cells infected in the presence of chloramphenicol.

The three most commonly studied members of the group of filamentous coliphages, bacteriophages f1, fd, and M13, are closely related (6, 29) and are presumably derived from a common ancestor. Although the amino acid sequence of the major coat protein has been shown to be the same for all three viruses (1, 5, 21), these viruses can be distinguished on serological grounds (31), and early studies revealed significant differences in the distribution of sequences of consecutive pyrimidine nucleotides in their DNA molecules (13, 43), in the products of cleavage of the RFI DNA by restriction endonucleases (45), and in their susceptibility to bacterial restriction (29). The complete sequences of nucleotides in the DNAs of bacteriophages fd (2) and M13 (46) and the structures of several fragments of phage f1 DNA (9, 25, 27, 28, 35) have been published. While this paper was in preparation, Beck and Zink (3) published a nucleotide sequence for the phage f1 genome and compared it with the sequence of phage fd DNA. There are, however, differences between their sequence for f1 DNA and the sequence described in this paper.

The untranslated intergenic region (nucleotides 5501 to 6005) of the filamentous coliphage genome, first noted by Vovis et al. (47), has recently assumed practical importance with the discovery that nucleotides can be incorporated in vitro into this part of the sequence without affecting the maturation and assembly of the virus particles. This observation has led to the development of derivatives of bacteriophage M13 as cloning vectors of particular value for rapid nucleotide sequencing (18, 34, 36).

Our laboratory has for many years been engaged in a study of the chemistry and primary structure of bacteriophage f1 DNA (1, 9, 23, 42–44). We report here the sequence that we have derived for the complete phage f1 genome and compare it with the data already available for this DNA (3) and for the DNAs of bacterio-

phages fd (2, 3) and M13 (46). In an earlier paper (9), we gave the full experimental evidence for the structure of the region of this DNA encompassing genes V, VII, and VIII and our rationale for the resolution of some areas of ambiguity. In this present paper, we comment briefly on the experimental evidence for those regions of the molecule (notably within gene IV) that we have found to differ markedly from the corresponding regions in fd DNA (2, 3) and on those parts of the molecule where special difficulty was experienced in interpreting the experimental data. We give experimental evidence only for those regions where our results are significantly at variance with partial sequence data previously published by others. The determination of the remainder of the sequence posed no unusual problems. We will, however, be happy to provide our detailed evidence for the structure of any specific region upon request.

Conventions. Restriction endonucleases are designated according to the abbreviated nomenclature suggested by Roberts (30). With the exception of *MnlI* fragments, which were all derived from the isolated *HinfI* A fragment and are labeled numerically in order of decreasing size, DNA fragments resulting from the complete digestion of RFI DNA with specific restriction endonucleases are identified alphabetically in order of decreasing size (as determined from the final nucleotide sequence). Restriction endonuclease products that were used, either as primers or as substrates for chemical degradation, after secondary cleavage with another restriction endonuclease are identified with the secondary enzyme indicated second (e.g., *HhaI*-L/*TaqI*).

In accordance with the convention established for the nucleotide sequences of the DNAs of bacteriophages fd (2) and M13 (46), the nucleotide sequence of the circular phage f1 DNA molecule presented in this paper is numbered in the 5′ → 3′ direction from the unique *HindII* cleavage site that is located in gene II. The phage DNA molecule represents the plus (sense) strand of the RFI DNA. Nucleotides derived from the minus strand of the RFI DNA are identified by a prime. Thus, T′20 is the nucleotide complementary to A20 of the plus strand. (Note that the nucleotides of the minus strand are numbered in the 3′ → 5′ direction.)

MATERIALS AND METHODS

Bacteriophage f1, its host, *E. coli* K38, a gene VI amber mutant of phage f1 (R7), a gene IV amber mutant (R12), and the permissive host, *E. coli* K38 (SuI), were generously given by N. D. Zinder, Rockefeller University, New York.

All chemicals, reagents, buffers, and purified enzymes were prepared or obtained as previously described (9). Restriction endonucleases were purified from the appropriate bacterial strains by the general method of Takanami (41) or were purchased from New England BioLabs, Beverly, Mass.

Bacteriophage f1 DNA. Bacteriophage f1 was grown on *E. coli* K38 as its host bacterium. The phage was purified, and the DNA was isolated as previously described (42). The conditions for the degradation of RFI DNA with restriction endonucleases, the precipitation of fragments from solution, the purification of DNA fragments by gel electrophoresis, the dephosphorylation of fragments, and their preparation for chemical degradation have already been described in detail (9).

Nucleic acid sequencing methods. Sequencing by chemical degradation was performed as described by Maxam and Gilbert (17), except that the total purine display used in sequencing many of the fragments was generated by incubating the labeled DNA fragment plus calf thymus carrier DNA (1 μg) in a total volume of 30 μl of 66% formic acid at 20°C for 3 to 5 min. The reaction was terminated by the addition of 0.3 M sodium acetate (200 μl) containing tRNA carrier (1 μg) and precipitated twice with ethanol. The degraded DNA was taken up in freshly prepared piperidine (1 M, 20 μl), incubated at 90°C for 30 min, and lyophilized. The residue was dissolved in water (20 μl), lyophilized again, and taken up in gel-loading solution (10 μl); a portion was loaded onto the gel. Sequencing by the indirect method of extension with *E. coli* DNA polymerase (Pol I) with premature chain termination was carried out as described by Sanger et al. (36), and the products of elongation were recovered as previously described (9). Priming fragments longer than 100 base pairs were removed before electrophoresis of the products. This was conveniently done by incubating the reaction mixture containing the newly synthesized products with the enzyme used to generate the primer (datum endonuclease). After the dATP chase, 1 to 2 U of the datum endonuclease were added, and the reaction mixture was incubated at 37°C for 30 min (63°C for endonuclease *TaqI*) before the products were precipitated. For removal of DNA primers generated by *AluI* and *HphI*, the dATP chase was replaced by a mixture of all four deoxynucleoside triphosphates, each at a 0.5 mM concentration, and the mixture was incubated at 37°C for 30 min before the addition of the datum endonuclease.

Gel electrophoresis for sequencing. The products, both from the chemical degradation and the chain extension experiments, were separated by electrophoresis on thin polyacrylamide gels (33). All sequencing gels were 40 by 20 by 0.02 cm, and most had an acrylamide concentration of 8%. Gels of 6 or 20% concentration were occasionally used for the separation of products with size ranges outside the normal limits. Electrophoresis was routinely performed at 1,000 V (20 to 25 mA), constant voltage, until the marker dyes had migrated the desired distance. Samples (5 μl) of each series of reaction mixtures were run on individual gels for 3, 8, 12, and 18 h to give a nucleotide sequence range from 20 to 200 nucleotides with a good overlap of sequence data between experiments. For the resolution of products containing a strong secondary structure and exhibiting compressional effects on electrophoresis, it was occasionally necessary to perform the electrophoresis at 50 W

(1,800 V), constant power, with 8% acrylamide gels that had been pre-electrophoresed for 1 h before loading the samples. Under these conditions, the gel temperature rose as high as 80°C, and many of the electrophoresis runs failed because the glass plates cracked. The nucleotide products of the reactions were detected on the sequencing gels by radioautography with preflashed film (11).

RESULTS

Sequencing strategy. The final strategy in determining the sequence of the bacteriophage f1 genome is summarized in Fig. 1. Our initial approach was to test a suitably large parent fragment of the RFI DNA covering the region of interest in each case against all restriction endonucleases available to us and to use the isolated smaller products as primers for Pol I-catalyzed extension (36) with single-stranded phage f1 DNA as the template. From the sequence data thus obtained, together with an estimate of the primer length, sufficient information was obtained to permit the provisional ordering of the restriction endonuclease products and to establish an outline sequence of the parent DNA fragment. Specific fragments of RFI DNA were then selected for direct sequencing by the chemical degradation method (17) or for use as primers for the Pol I-catalyzed extension method (36), and the provisional sequence was verified or adjusted as necessary. Since all data from the Pol I experiments referred to minus-strand sequences only, sequences derived from the chemical degradation of terminally labeled fragments could be assigned to the plus or minus strand by comparison with these data. Once a tentative sequence for the complete DNA molecule had been obtained, specific experiments were designed to confirm the structure in areas of uncertainty. The complete nucleotide sequence of the genome of bacteriophage f1 is given in Fig. 2. For ease of comparison, the initiating codons of genes I to VIII proposed by Beck et al. (2) for phage fd have been marked in, together with those of the more recently located gene IX (12, 37) and the putative gene X, which shares the same reading frame as gene II (10, 19). The reading frame of gene VI was confirmed by locating a G → A transition at position 3066′ in the sequence of the gene VI amber mutant R7. That of gene IV was confirmed by demonstrating a similar transition at position 4515′ in the sequence of the gene IV amber mutant R12.

Resolution of major ambiguities. The methods used to resolve ambiguities in the sequence determined for the region of phage f1 DNA containing genes V, VII, and VIII have already been described (9). Regions of the remainder of the genome that posed special problems of interpretation are described below, together with the arguments used to distinguish between the various possibilities.

(i) Gene II. Two regions that gave nucleotide products exhibiting strong secondary structure effects in the gel separation step were encountered at positions 38′ to 42′ and 396′ to 399′. The nucleotide arrangement in each case was resolved by sequencing the complementary strands by the chemical degradation method. Those regions of the sequence lying between nucleotides 6035 to 6075 and 6340 to 6353 have been determined solely by the Pol I-catalyzed extension method and have not been confirmed by the chemical degradation of the other strand.

(ii) Gene III. Several regions within gene III proved difficult to interpret.

(a) The DNA sequence lying between nucleotides 2284 and 2391 was found to contain an unusual eightfold repetition of the sequence GGPyGGPyGGPyTCPy (where Py = C or T). Data obtained from this region by Pol I-catalyzed extension experiments did not always show clear C bands, and the overlap between the repeating units was difficult to establish. Furthermore, extension of the priming fragment *Hin*fI G (with primer removal) always gave ambiguous gel patterns, and several experiments were required to confirm the sequence. Gel bands from the confirmatory chemical degradation of *Hin*fI-D/*Hap*II, although not well resolved, showed seven repeating clusters of pyrimidine nucleotides ending at the *Hap*II cleavage site at position 2379, which is in agreement with the sequence finally deduced.

(b) A similar fourfold repetition of groups of nucleotides was found between positions 1836 and 1892, and the results of Pol I-catalyzed extension experiments failed to demonstrate all of the C residues in this region. Confirmatory evidence was obtained from the chemical degradation of the isolated plus strand of *Hin*fI-C.

(c) Gels from both the chemical degradation experiments and the Pol I-catalyzed extension experiments relating to the sequence between nucleotides 1540 and 1550 consistently showed strong intrastrand base-pairing effects. However, these effects occurred at slightly different places in different experiments, and so the sequence in this region could be resolved.

(d) Results from Pol I-catalyzed extension experiments clearly showed residue 1967′ to be G. The corresponding nucleotide in the plus strand of *Hin*fI-D, however, failed to react with hydrazine. Nucleotides 1966 to 1970 (CCTGG) form the recognition sequence for restriction endonuclease *Eco*RII, to which the replicative form of phage f1 grown on *E. coli* K38 is resistant (45). Since 5-methylcytosine is resistant to degradation by hydrazine (22), it can be

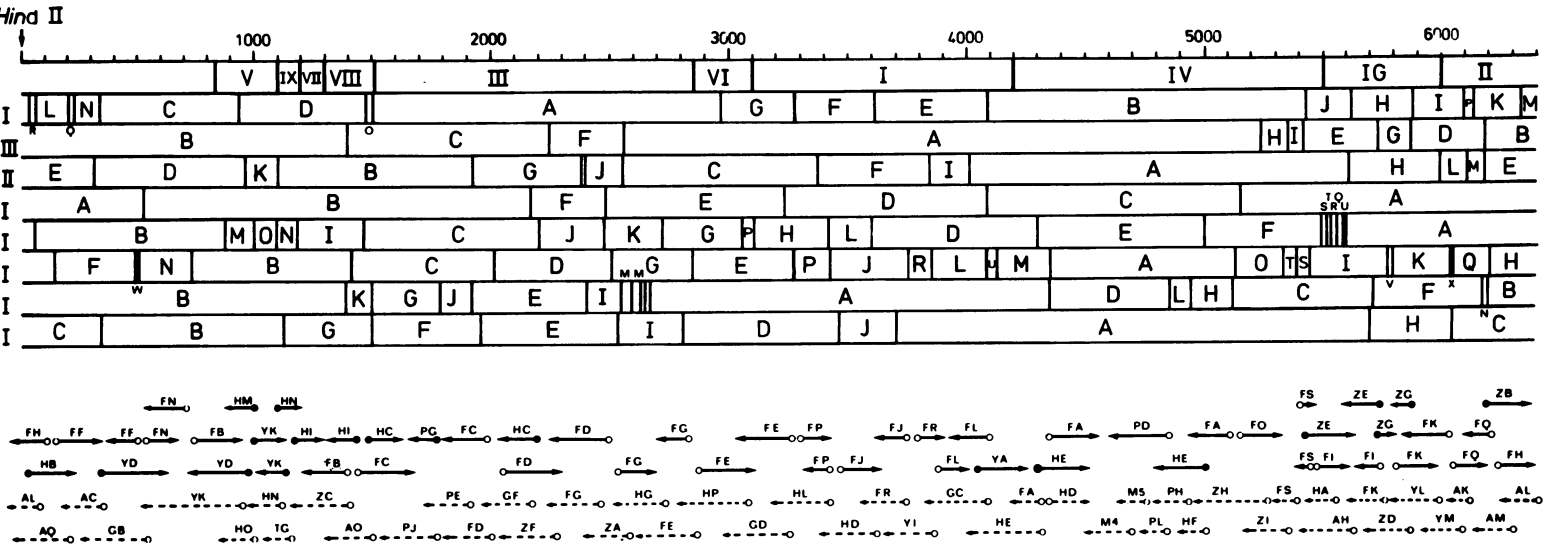


FIG. 1. Restriction fragments and summary of strategy used in sequencing the DNA of bacteriophage f1. The upper part of the figure summarizes the positions of the cleavage sites and the products of the various restriction endonucleases used in this study relative to the nucleotide number (above top line) and the gene map (below top line). Subfragments derived through the degradation of *Hin*fI-A with *Mnl*I are not shown. The lower part of the figure summarizes the strategy employed. Solid arrows denote restriction fragments that were sequenced by chemical degradation (17). Arrows terminating in solid circles were radioactively labeled at the 5′ terminus by reaction with polynucleotide kinase; those terminating in open circles were labeled at the 3′ terminus through the Pol I-catalyzed incorporation of radioactive nucleotides. Broken arrows represent sequences obtained by the Pol I-catalyzed extension method (36) with the restriction fragment shown as a primer on a single-stranded f1 DNA template. The lengths of the arrows show the lengths of the derived sequences, and the arrowheads indicate the strands to which the data refer (→ = plus strand; ← = minus strand). Sources of restriction fragments: A, *Alu*I; F, *Hin*fI; G, *Hga*I; H, *Hha*I; M, *Mnl*I; P, *Hph*I; T, *Taq*I; Y, *Hap*I; Z, *Hae*III.

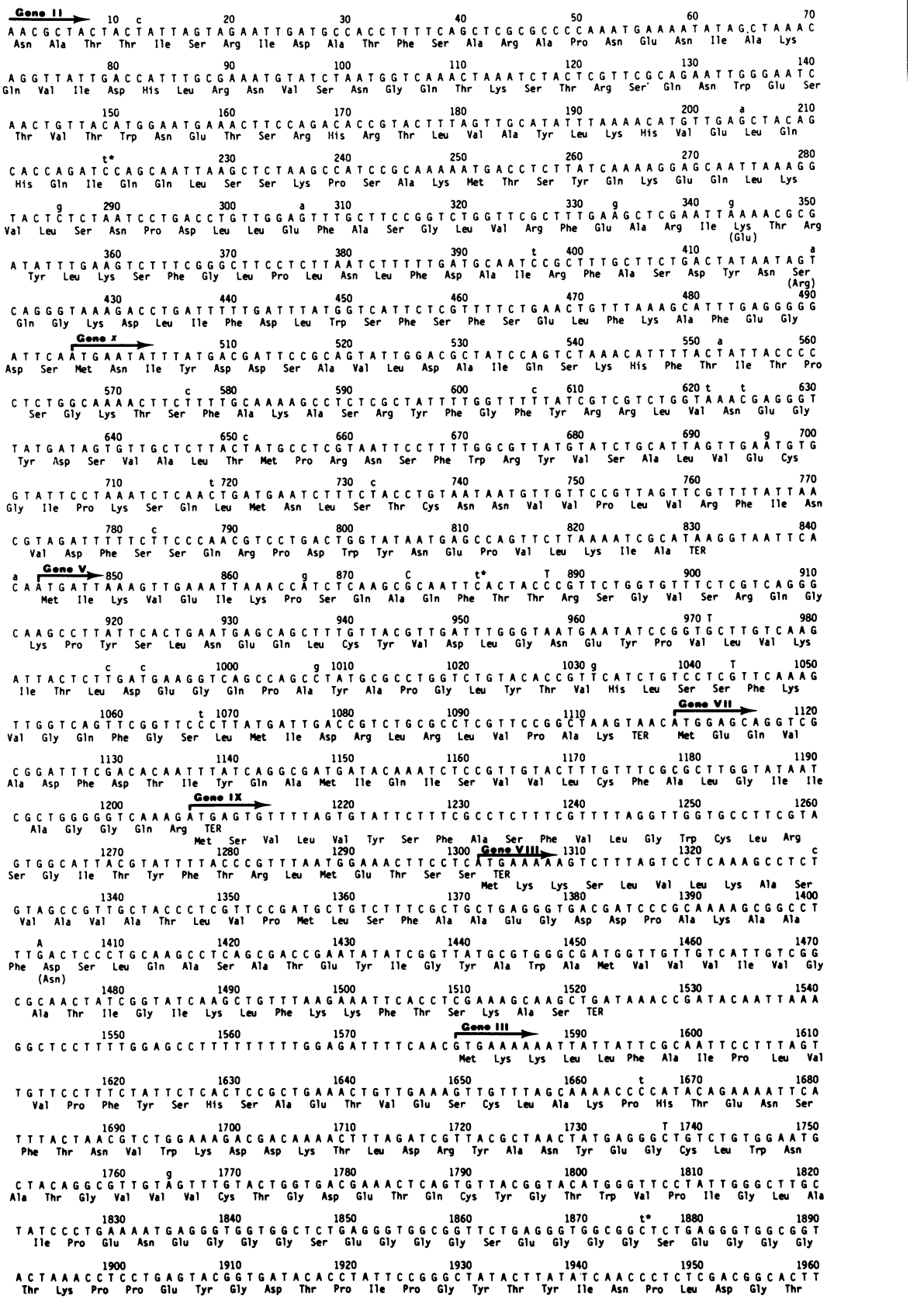
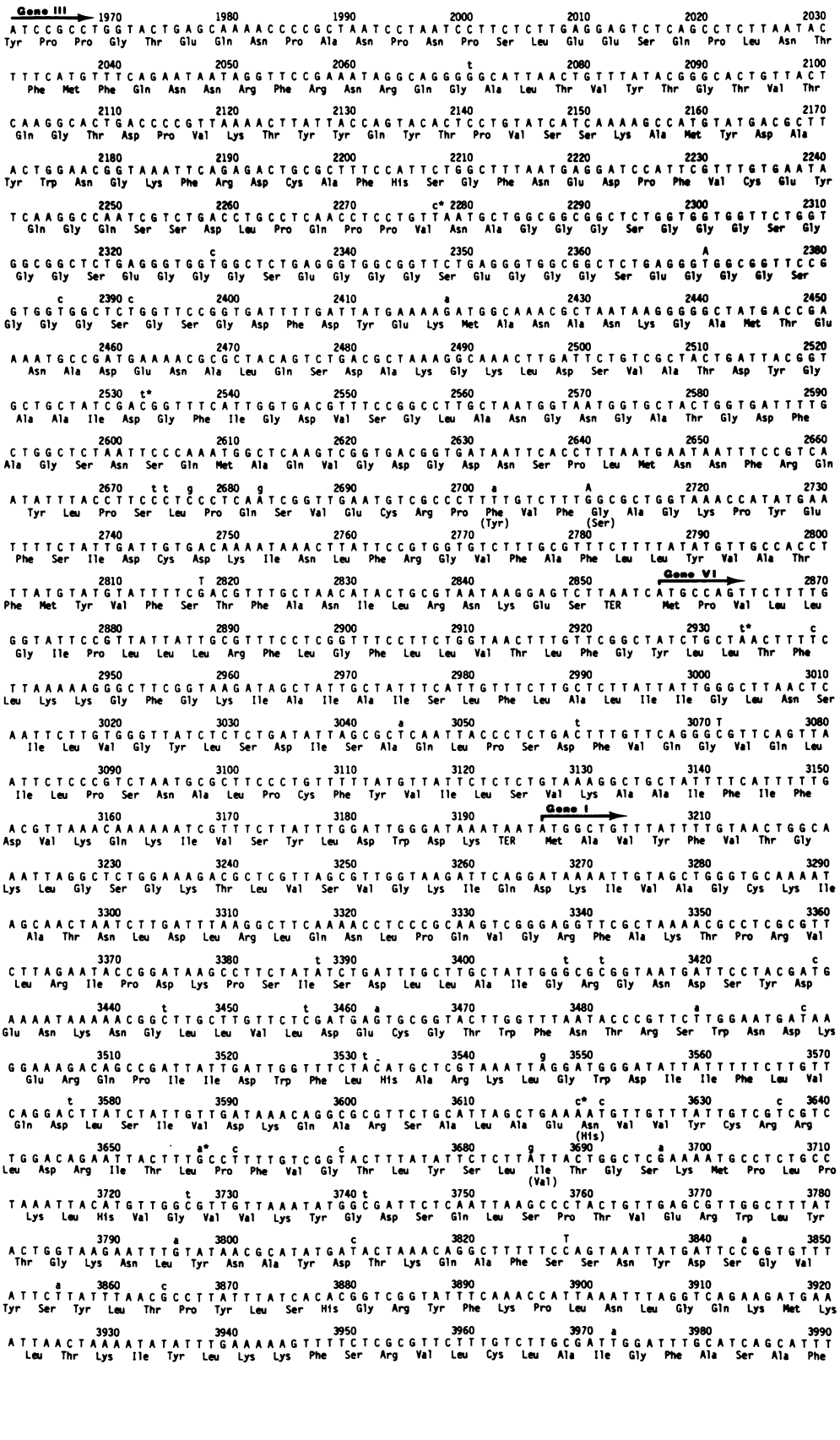


FIG. 2. Nucleotide sequence of bacteriophage f1 DNA. The sequence shown is of the viral plus strand. The initiation codons and the protein translation products of the various genes, including the putative gene X, are indicated. Positions of differences from the nucleotide sequences of phage fd and M13 DNAs are shown above the main sequence: lowercase letters, change in fd DNA only; uppercase letters, change in M13 DNA only; lowercase letters plus asterisk, same change in both fd and M13 DNAs. Amino acid changes in the corresponding translated protein products resulting from these nucleotide substitutions are indicated in parentheses below the f1 translation line.


Gene III
1970 1980 1990 2000 2010 2020 2030
A T C C G C C T G G T A C T G A G C A A A A C C C C G C T A A T C C T A A T C C T T C T C T T G A G G A G T C T C A G C C T C T T A A T A C
Tyr Pro Pro Gly Thr Glu Gln Asn Pro Ala Asn Pro Asn Pro Ser Leu Glu Glu Ser Gln Pro Leu Asn Thr
2040 2050 2060 t 2080 2090 2100
T T T C A T G T T T C A G A A T A A T A G G T T C C G A A A T A G G C A G G G G G C A T T A A C T G T T T A T A C G G G C A C T G T T A C T
Phe Met Phe Gln Asn Asn Arg Phe Arg Asn Arg Gln Gly Ala Leu Thr Val Tyr Thr Gly Thr Val Thr
2110 2120 2130 2140 2150 2160 2170
C A A G G C A C T G A C C C C G T T A A A A C T T A T T A C C A G T A C A C T C C T G T A T C A T C A A A A G C C A T G T A T G A C G C T T
Gln Gly Thr Asp Pro Val Lys Thr Tyr Tyr Gln Tyr Thr Pro Val Ser Ser Lys Ala Met Tyr Asp Ala
2180 2190 2200 2210 2220 2230 2240
A C T G G A A C G G T A A A T T C A G A G A C T G C G C T T T C C A T T C T G G C T T T A A T G A G G A T C C A T T C G T T T G T G A A T A
Tyr Trp Asn Gly Lys Phe Arg Asp Cys Ala Phe His Ser Gly Phe Asn Glu Asp Pro Phe Val Cys Glu Tyr
2250 2260 2270 c* 2280 2290 2300 2310
T C A A G G C C A A T C G T C T G A C C T G C C T C A A C C T C C T G T T A A T G C T G G C G G C G G C T C T G G T G G T G G T T C T G G T
Gln Gly Gln Ser Ser Asp Leu Pro Gln Pro Pro Val Asn Ala Gly Gly Gly Ser Gly Gly Gly Ser Gly
2320 c 2340 2350 2360 A 2380
G G C G G C T C T G A G G G T G G T G G C T C T G A G G G T G G C G G C T C T G A G G G T G G C G G T T C C G
Gly Gly Ser Glu Gly Gly Gly Ser Glu Gly Gly Gly Ser Glu Gly Gly Gly Ser Glu Gly Gly Gly Ser
c 2390 c 2400 2410 a 2430 2440 2450
G T G G T G G C T C T G G T T C C G G T G A T T T T G A T T A T G A A A A G A T G G C A A A C G C T A A T A A G G G G G C T A T G A C C G A
Gly Gly Gly Ser Gly Ser Gly Asp Phe Asp Tyr Glu Lys Met Ala Asn Ala Asn Lys Gly Ala Met Thr Glu
2460 2470 2480 2490 2500 2510 2520
A A A T G C C G A T G A A A A C G C G C T A C A G T C T G A C G C T A A A G G C A A A C T T G A T T C T G T C G C T A C T G A T T A C G G T
Asn Ala Asp Glu Asn Ala Leu Gln Ser Asp Ala Lys Gly Lys Leu Asp Ser Val Ala Thr Asp Tyr Gly
2530 t* 2540 2550 2560 2570 2580 2590
G C T G C T A T C G A C G G T T T C A T T G G T G A C G T T T C C G G C C T T G C T A A T G G T A A T G G T G C T A C T G G T G A T T T T G
Ala Ala Ile Asp Gly Phe Ile Gly Asp Val Ser Gly Leu Ala Asn Gly Asn Gly Ala Thr Gly Asp Phe
2600 2610 2620 2630 2640 2650 2660
C T G G C T C T A A T T C C C A A A T G G C T C A A G T C G G T G A C G G T G A T A A T T C A C C T T T A A T G A A T A A T T T C C G T C A
Ala Gly Ser Asn Ser Gln Met Ala Gln Val Gly Asp Gly Asp Asn Ser Pro Leu Met Asn Asn Phe Arg Gln
2670 t t g 2680 g 2690 2700 a A 2720 2730
A T A T T T A C C T T C C C T C C C T C A A T C G G T T G A A T G T C G C C C T T T G T C T T T G G C G C T G G T A A A C C A T A T G A A
Tyr Leu Pro Ser Leu Pro Gln Ser Val Glu Cys Arg Pro Phe Val Phe Gly Ala Gly Lys Pro Tyr Glu
(Tyr) (Ser)
2740 2750 2760 2770 2780 2790 2800
T T T T C T A T T G A T T G T G A C A A A A T A A A C T T A T T C C G T G G T G T C T T T G C G T T T C T T T T A T A T G T T G C C A C C T
Phe Ser Ile Asp Cys Asp Lys Ile Asn Leu Phe Arg Gly Val Phe Ala Phe Leu Leu Tyr Val Ala Thr
2810 T 2820 2830 2840 2850 Gene VI 2870
T T A T G T A T G T A T T T T C G A C G T T T G C T A A C A T A C T G C G T A A T A A G G A G T C T T A A T C A T G C C A G T T C T T T T G
Phe Met Tyr Val Phe Ser Thr Phe Ala Asn Ile Leu Arg Asn Lys Glu Ser TER Met Pro Val Leu Leu
2880 2890 2900 2910 2920 2930 t* c
G G T A T T C C G T T A T T A T T G C G T T T C C T C G G T T T C C T T C T G G T A A C T T T G T T C G G C T A T C T G C T A A C T T T T C
Gly Ile Pro Leu Leu Leu Arg Phe Leu Gly Phe Leu Leu Val Thr Leu Phe Gly Tyr Leu Leu Thr Phe
2950 2960 2970 2980 2990 3000 3010
T T A A A A A G G G C T T C G G T A A G A T A G C T A T T G C T A T T T C A T T G T T T C T T G C T C T T A T T A T T G G G C T T A A C T C
Leu Lys Lys Gly Phe Gly Lys Ile Ala Ile Ala Ile Ser Leu Phe Leu Ala Leu Ile Ile Gly Leu Asn Ser
3020 3030 3040 a 3050 t 3070 T 3080
A A T T C T T G T G G G T T A T C T C T C T G A T A T T A G C G C T C A A T T A C C C T C T G A C T T T G T T C A G G G C G T T C A G T T A
Ile Leu Val Gly Tyr Leu Ser Asp Ile Ser Ala Gln Leu Pro Ser Asp Phe Val Gln Gly Val Gln Leu
3090 3100 3110 3120 3130 3140 3150
A T T C T C C C G T C T A A T G C G C T T C C C T G T T T T T A T G T T A T T C T C T C T G T A A A G G C T G C T A T T T T C A T T T T T G
Ile Leu Pro Ser Asn Ala Leu Pro Cys Phe Tyr Val Ile Leu Ser Val Lys Ala Ala Ile Phe Ile Phe
3160 3170 3180 3190 Gene I 3210
A C G T T A A A C A A A A A A T C G T T T C T T A T T T G G A T T G G G A T A A A T A A T A T G G C T G T T T A T T T T G T A A C T G G C A
Asp Val Lys Gln Lys Ile Val Ser Tyr Leu Asp Trp Asp Lys TER Met Ala Val Tyr Phe Val Thr Gly
3230 3240 3250 3260 3270 3280 3290
A A T T A G G C T C T G G A A A G A C G C T C G T T A G C G T T G G T A A G A T T C A G G A T A A A A T T G T A G C T G G G T G C A A A A T
Lys Leu Gly Ser Gly Lys Thr Leu Val Ser Val Gly Lys Ile Gln Asp Lys Ile Val Ala Gly Cys Lys Ile
3300 3310 3320 3330 3340 3350 3360
A G C A A C T A A T C T T G A T T T A A G G C T T C A A A A C C T C C C G C A A G T C G G G A G G T T C G C T A A A A C G C C T C G C G T T
Ala Thr Asn Leu Asp Leu Arg Leu Gln Asn Leu Pro Gln Val Gly Arg Phe Ala Lys Thr Pro Arg Val
3370 3380 t 3390 3400 t t 3420 c
C T T A G A A T A C C G G A T A A G C C T T C T A T A T C T G A T T T G C T T G C T A T T G G G C G C G G T A A T G A T T C C T A C G A T G
Leu Arg Ile Pro Asp Lys Pro Ser Ile Ser Asp Leu Leu Ala Ile Gly Arg Gly Asn Asp Ser Tyr Asp
3440 t 3450 t 3460 a 3470 3480 a c
A A A A T A A A A A C G G C T T G C T T G T T C T C G A T G A G T G C G G T A C T T G G T T T A A T A C C C G T T C T T G G A A T G A T A A
Glu Asn Lys Asn Gly Leu Leu Val Leu Asp Glu Cys Gly Thr Trp Phe Asn Thr Arg Ser Trp Asn Asp Lys
3510 3520 3530 t 3540 g 3550 3560 3570
G G A A A G A C A G C C G A T T A T T G A T T G G T T T C T A C A T G C T C G T A A A T T A G G A T G G G A T A T T A T T T T T C T T G T T
Glu Arg Gln Pro Ile Ile Asp Trp Phe Leu His Ala Arg Lys Leu Gly Trp Asp Ile Ile Phe Leu Val
t 3580 3590 3600 3610 c* c 3630 c 3640
C A G G A C T T A T C T A T T G T T G A T A A A C A G G C G C G T T C T G C A T T A G C T G A A A A T G T T G T T T A T T G T C G T C G T C
Gln Asp Leu Ser Ile Val Asp Lys Gln Ala Arg Ser Ala Leu Ala Glu Asn Val Val Tyr Cys Arg Arg
(His)
3650 a* c c 3680 g 3690 a 3700 3710
T G G A C A G A A T T A C T T T G C C T T T T G T C G G T A C T T T A T A T T C T C T T A T T A C T G G C T C G A A A A T G C C T C T G C C
Leu Asp Arg Ile Thr Leu Pro Phe Val Gly Thr Leu Tyr Ser Leu Ile Thr Gly Ser Lys Met Pro Leu Pro
(Val)
3720 t 3730 3740 t 3750 3760 3770 3780
T A A A T T A C A T G T T G G C G T T G T T A A A T A T G G C G A T T C T C A A T T A A G C C C T A C T G T T G A G C G T T G G C T T T A T
Lys Leu His Val Gly Val Val Lys Tyr Gly Asp Ser Gln Leu Ser Pro Thr Val Glu Arg Trp Leu Tyr
3790 a 3800 c 3820 T 3840 a 3850
A C T G G T A A G A A T T T G T A T A A C G C A T A T G A T A C T A A A C A G G C T T T T T C C A G T A A T T A T G A T T C C G G T G T T T
Thr Gly Lys Asn Leu Tyr Asn Ala Tyr Asp Thr Lys Gln Ala Phe Ser Ser Asn Tyr Asp Ser Gly Val
a 3860 c 3870 3880 3890 3900 3910 3920
A T T C T T A T T T A A C G C C T T A T T T A T C A C A C G G T C G G T A T T T C A A A C C A T T A A A T T T A G G T C A G A A G A T G A A
Tyr Ser Tyr Leu Thr Pro Tyr Leu Ser His Gly Arg Tyr Phe Lys Pro Leu Asn Leu Gly Gln Lys Met Lys
3930 3940 3950 3960 3970 a 3980 3990
A T T A A C T A A A A T A T A T T T G A A A A A G T T T T C T C G C G T T C T T T G T C T T G C G A T T G G A T T T G C A T C A G C A T T T
Leu Thr Lys Ile Tyr Leu Lys Lys Phe Ser Arg Val Leu Cys Leu Ala Ile Gly Phe Ala Ser Ala Phe

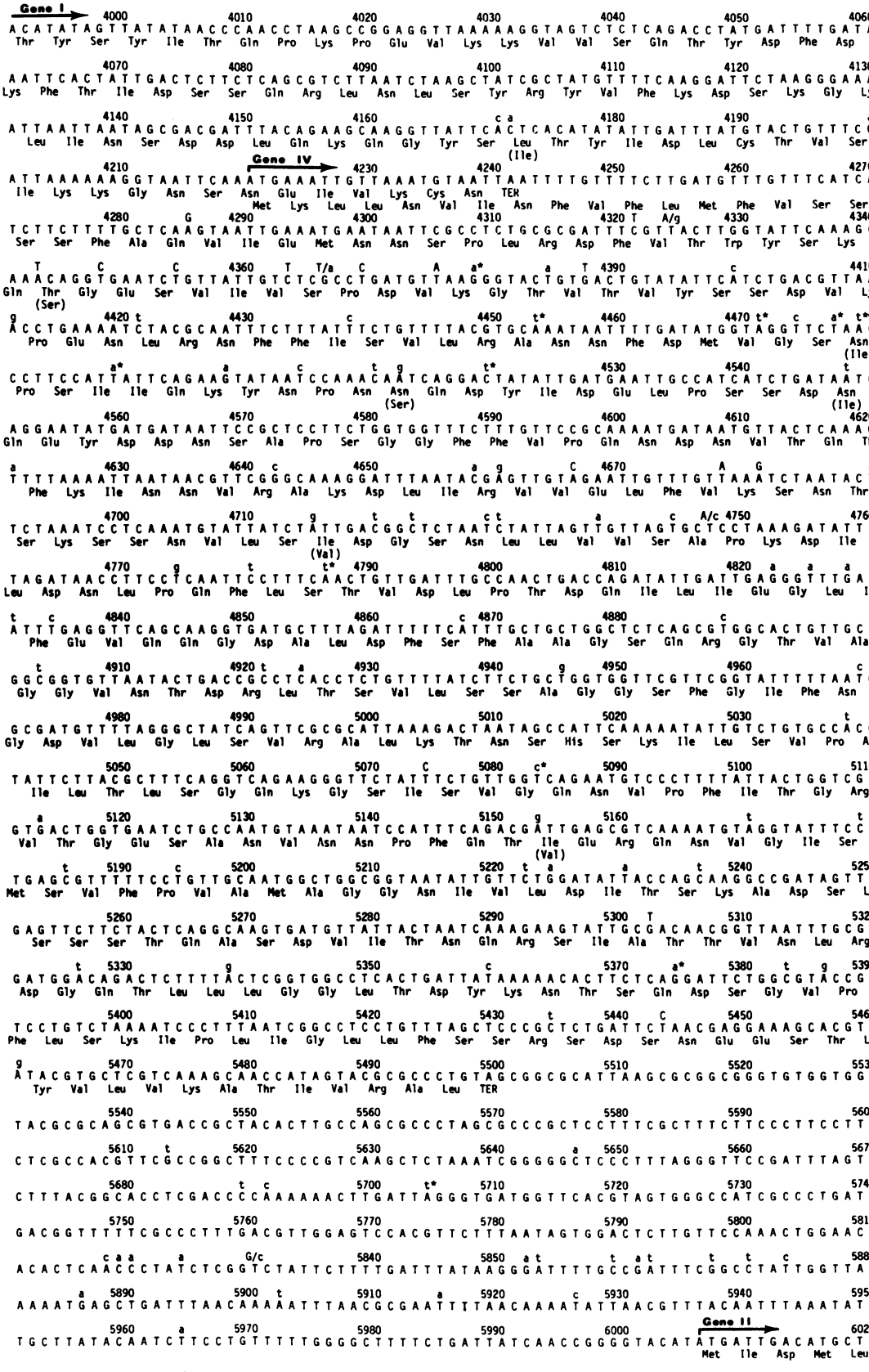


FIG. 2. Continued.

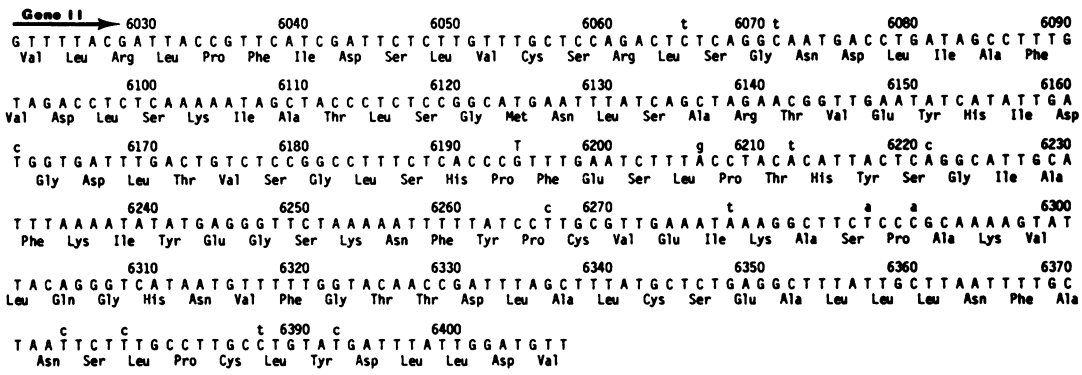


FIG. 2. Continued.

concluded that nucleotide 1967 is methylated. This is consistent with the known (4) specificity of the modification system for this enzyme. No data were obtained by chemical degradation of the corresponding C residue at position 1969′ of the RFI DNA, but it is probable that this nucleotide is also methylated.

(iii) Gene I. Although the gel patterns of products derived through both sequencing methods from the region of the DNA lying between nucleotides 3319 and 3366 consistently exhibited intrastrand base pairing in both strands, the location of these compressional effects at slightly different positions in each case permitted the resolution of the sequence in this region.

(iv) Gene IV. The portion of the nucleotide sequence lying between nucleotides 4540 and 4620 was determined solely from the Pol I-catalyzed extension of *Mnl*I-4. Attempts to use *Hha*I-D/*Alu*I as the primer on the isolated minus strand of *Hga*I-D (used as the template) gave gels exhibiting compressional effects. Comparison of the data from these experiments with the confirmed sequence showed that these banding patterns occurred immediately before two or more purine residues. It has been demonstrated that pyrimidine dimer formation in the template blocks extension by Pol I (20), and the results shown in Fig. 3a suggested that some pyrimidine dimer formation may have been induced in the template as a result of the exposure of the DNA to UV light during isolation.

The only part of the gene IV sequence to exhibit secondary structure effects was the sequence at positions 4304 to 4306. The number and order of these nucleotides were resolved by separating the products of Pol I-catalyzed extension at higher electrophoresis temperatures.

(v) Intergenic region. The final sequence that we derived for this region (Fig. 2, nucleotides 5501 to 6005) differs from that previously published by Ravetch et al. (28) through the inclusion of single nucleotides at positions 5532 (A), 5538 (A), 5559 (A), 5618 (C), 5619 (T), 5770 (C), 5794 (T), and 5915 (T). We have, moreover, determined the sequence of an additional 13 nucleotides (TGAGCTGATTTAA) at positions 5885 to 5897 to fill a gap in the partial sequence proposed by Ravetch et al. (28). Representative gel patterns supporting our assignments for each of these positions are illustrated in Fig. 3b to e.

The severe compressional effects observed on gel patterns derived from nucleotides 5480 to 5580 led to difficulties in interpreting the data derived from this region of the molecule. These effects were observed on gels obtained by both sequencing methods, and they occurred at approximately the same positions in sequences derived from either strand. Clear evidence for the nucleotide sequence in this region is restricted to the data obtained through the chemical degradation of the minus strand of *Hin*fI-I after fractionation of the products under extreme electrophoresis conditions. Other parts of the sequence of the intergenic region that failed to give clear results when first examined by the two methods used in this study were as follows.

(a) Nucleotides 5506′ and 5507′. Nucleotides 5506′ and 5507′ consistently ran with the same mobility on gels derived from the chemical degradation experiments. Although the quality of the gel patterns from confirmatory experiments was poor, the results of Pol I-catalyzed extension experiments showed T at position 5507′. Since nucleotide 5506′ gave an ambiguous G band, it was concluded that the correct assignments for nucleotides 5507′ and 5506′ were T and G, respectively.

(b) Nucleotide 5533′. The enhanced intensity of the G band corresponding to position 5533′ on gels obtained from the chemical degradation of the minus strand of *Hin*fI-I/*Hph*I (Fig. 3b) suggested the comigration of two nucleotide bands at this point. Pol I-catalyzed extension experiments and chemical degradation of the plus strand of *Hae*III-E/*Hph*I, however, both showed only one G band at this position.

(c) Nucleotide 5540′. The results of Pol I-

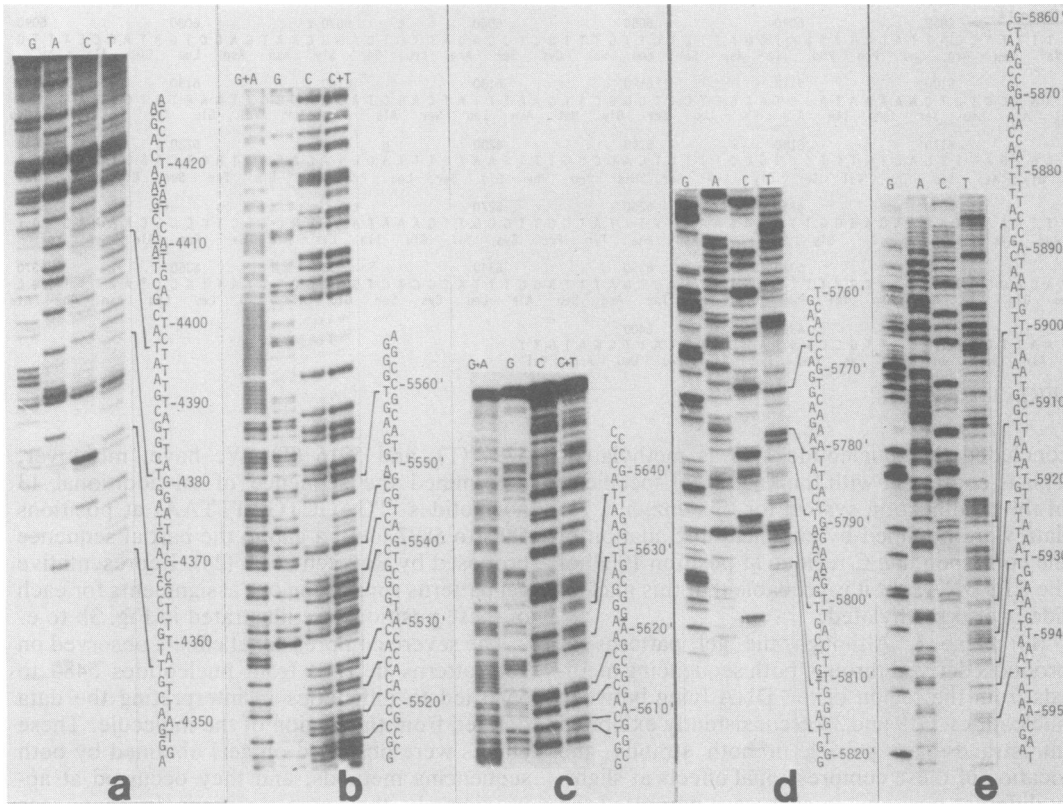


FIG. 3. Gel patterns illustrating regions of ambiguity discussed in the text. (a) Plus-strand sequence obtained through the Pol I-catalyzed extension of *Alu*I-B/*Hha* on the isolated minus strand of *Hga*I-C as the template; (b and c) minus-strand sequence obtained through the chemical degradation of the duplex *Hinf*I-I/*Hph*I labeled at the 3′ terminus through the Pol I-catalyzed incorporation of [α-^{32}P]dATP; (d and e) minus-strand sequence obtained through the Pol I-catalyzed extension of *Hap*II-L on isolated phage f1 DNA as the template. The priming fragment was removed with the datum endonuclease before the products were separated by electrophoresis.

catalyzed extension experiments suggested that two G residues may be present at position 5540′. Chemical degradation studies (Fig. 3b) did not, however, support this conclusion, and only one G residue has been assigned to this position.

(d) Nucleotides 5548′ and 5549′. Only one purine band could be seen at positions 5548′ and 5549′ on gels from chemical degradation of the minus strand of *Hin*fI-I/*Hph*I (Fig. 3b), but the intensity of this band in the G+A reaction suggested that an A band may be comigrating at this position. The gel patterns from Pol I-catalyzed extension experiments with *Hae*III-G as the primer and the chemical degradation of the plus strand of *Hae*III-E/*Hph*I, however, confirmed the presence of two purine nucleotides, G and A, in positions 5548′ and 5549′, respectively.

(e) Nucleotides 5571′ to 5573′. The intensity of the G band at position 5571′ on gels obtained from the products of chemical degradation of *Hin*fI-I/*Hph*I suggested that an A band may be comigrating at this position. Although the band spacing was not uniform on gels from Pol I-catalyzed extension experiments, three G bands could be clearly seen when *Hae*III-G was used as the primer.

(f) Nucleotides 5793′ to 5798′. Gels from Pol I-catalyzed extension of *Hae*III-D/*Hap*II suggested the sequence of the minus strand at this point to be 3′-GGACAGAG. Data from the chemical degradation of the plus strand of *Hae*III-G showed clearly, however, that the complementary sequence in this region contained two additional nucleotides. The correct sequence of the plus strand at this point is thus CTTGTTC.

DISCUSSION

Accuracy of the sequence. The nucleotide sequence presented in Fig. 2 represents the plus-strand sequence that we have established from studies of the single-stranded viral DNA and of the double-stranded RFI DNA of bacteriophage

f1. It has been assumed throughout this study that both strands of the RFI DNA contain the same number of nucleotides and that the viral strand is fully complementary to the minus strand of the RFI DNA. Approximately 85% of the nucleotide sequence has been confirmed by the two independent sequencing methods (17, 36), but since 50% of the data obtained from the chemical degradation experiments gave minus-strand data confirming the results of Pol I-catalyzed expansion studies, approximately 50% of the presented plus-strand sequence has been deduced solely from the complementarity of base pairing with the nucleotides determined to be in the minus strand.

For those regions of the molecule where no convenient restriction endonuclease sites were available (e.g., nucleotides 4450 to 4460 in gene IV) or where features of the primary sequence made interpretation difficult (e.g., nucleotides 5500 to 5600 in the intergenic region), it was not always possible to determine the sequence by both independent methods. Although each of these regions has been sequenced at least twice by the same method, this may not be valid grounds for assuming accuracy. For example, the sequence comprising nucleotides 5793 to 5798, sequenced many times by the Pol I-catalyzed extension method, always gave the same gel pattern. In the absence of additional data, two A residues in this sequence would have been omitted.

In general, the major sources of error encountered in the sequence determination were the misreading of gel patterns and the incorrect recording and copying of sequence data. The sequence described in this paper was compiled for the most part without the use of the computer programs (38) that subsequently became available to us and that largely eliminate the last two sources of error. It should be noted that a copying error remained in the sequence of genes V, VII, and VIII that we have already published (9), where nucleotide 82, C (nucleotide 924 in the complete sequence now presented), was incorrectly recorded as a G, although the amino acid translation of the corresponding codon was correctly shown to be leucine. This error has been corrected in the complete sequence given in Fig. 2.

The sequence shows the single-stranded viral DNA to contain 6,407 nucleotides in the following proportions: A, 24.56%; T, 34.69%; G, 20.69%; and C, 20.24%. The molecular weight of the free acid is 1.98×10^6. The number of nucleotides in the total genome is the same as that found by van Wezenbeek et al. (46) in the DNA of bacteriophage M13, and it confirms the observation by Beck et al. (2) and Beck and Zink (3) that the genome of bacteriophage f1 is one nucleotide shorter than that of bacteriophage fd. If any nucleotides have been inadvertently omitted from the f1 DNA sequence, they are most likely to lie in the region between nucleotides 5500 and 5600 or in the gene III region between nucleotides 2250 and 2260, for which the interpretation of data was most difficult.

The first studies of nucleotide sequences in phage f1 DNA were confined to pyrimidine tract analysis (23). Long unique tracts of consecutive pyrimidine nucleotides were observed (23, 42) and characterized (13). The pyrimidine tract characterized by Ling (13) as C_3T_8 from the DNAs of both phages fd and f1 has not been located within the DNAs of either phage (2, 3; this study), but a tract of the same length and of base composition C_2T_9, previously identified in both f1 and fd DNAs (43), was found in the complete sequences of each of these genomes. The region of the DNA sequence in which this pyrimidine tract is located forms part of a potential hairpin structure that has been implicated in the termination of viral transcription (7, 40), and gels derived from this region displayed the characteristic compressional effects arising from intrastrand base pairing. The sequence of the corresponding minus-strand region was, however, clearly established as A_9G_2. Inspection of the chromatographic data presented by Ling (14) shows that the graticule position of this pyrimidine oligonucleotide would, indeed, predict a product with a base composition of C_3T_8. It is possible that the run of nine consecutive T residues influenced the electrophoretic mobility of this tract in his experiments and led to its incorrect identification.

A DNA sequence determined by Pol I-catalyzed extension of a synthetic octanucleotide primer (35) is located between nucleotides 6350 and 6398, within the coding region for gene II. The assignment of the last eight residues at the 3′ terminus of this plus-strand sequence was based on the assumption that the primer was exactly complementary to the template sequence. The completed f1 DNA sequence shows this assumption to have been in error, since the penultimate nucleotide (G) at the 5′ terminus of the primer was not, in fact, complementary to the T residue at position 6406 of the template DNA. A decanucleotide primer (AGAAATAAAA), specifically synthesized for use on a φX174 DNA template (32), was found in similar experiments to prime uniquely on f1 viral DNA (G. B. Petersen, unpublished data). The sequence of 46 nucleotides obtained from those experiments has provided us with useful confirmation of the sequence of nucleotides 4396 to 4441 in the coding region of gene IV, and it is interesting to note that the Klenow Pol I used in those earlier experiments successfully edited the

primer sequence by removing the mismatched A residue at the 3′ terminus before catalyzing the extension of the primer.

The sequences of several separate regions of the phage f1 genome have been reported earlier by ourselves and others (9, 24–28) and, except for the intergenic region (28) discussed above, are in good agreement with the sequence for the total genome that we now present. The complete nucleotide sequence of phage f1 DNA published by Beck and Zink (3), however, differs from our sequence at a number of positions. These can be summarized (with the Beck and Zink assignments in parentheses) as follows (note, however, that nucleotide numbers greater than 3194 must be increased by one to allow our sequence to be aligned with that presented by Beck and Zink, which was based on the sequence of fd DNA): 343, A (G); 1689, C (T); 1929, C (T); 2385, T (C); 2391, T (C); 2676, C (T); 2710, G (C); 4307, C (T); 4427, A (G); 4471, A (T); 4477, T (A); 4479, A (T); 4715, A (G); 5153, A (G); 5194, T (C); 5538, A (G).

Although these nucleotide differences are all in coding positions rather than in regulatory regions, some of them are significant in that they lead to different assessments of amino acid changes when the translated gene products of phage f1 are compared with those of phages M13 and fd. There are, furthermore, significant differences in the restriction map that can be deduced from our sequence and that deduced by Beck and Zink (3). Perhaps the most important of these is the introduction of a second endonuclease *Hin*dII recognition site in the Beck and Zink sequence as a result of the substitution of G for A at position 4715. The strain of f1 that we studied, like phages fd (2, 3) and M13 (46), has a single site only for this enzyme, at position 6405. Digestion of the RFI DNA with *Hin*dII converts it to the linear RFIII form, but no further digestion occurs. Minor differences in the restriction maps that can be deduced from the two sequences arise from other substitutions. A summary of restriction endonuclease cleavage sites found by Staden's program SEARCH (38) in the sequence presented in Fig. 2 is given in Table 1.

It is possible that the 16 differences between the sequences proposed by Beck and Zink (3) and by us reflect differences between the strains of bacteriophage f1 that we examined, but it is surprising that there should be so many differences, considering that the two phage cultures were obtained from the same source. The RF DNA studied by Beck and Zink (3) was enzymically prepared in vitro, and the possibility should be considered that some of the differences between their sequence and ours, which was derived from a study of the in vivo RF DNA, may represent polymerase copying errors. We have taken the opportunity to review our own evidence for the sequence presented in Fig. 2, but we have found no grounds for changing our original assignments at each of the positions at which the two sequences differ. In nearly every case, our data are supported by both methods of analysis (17, 36) and come from studies of both strands of the DNA.

Comparison of the DNA sequence of phage f1 with those of phages M13 and fd. The sequence of the DNA of bacteriophage f1 here presented confirms that the numbers of nucleotides in the DNAs of phages f1 and M13 are identical to each other and one less than the number of nucleotides in phage fd DNA, through the deletion of one nucleotide (nucleotide 3195, A) of the fd sequence from the small intergenic region lying between genes VI and I. This deletion has no effect on any structural gene, and the genomic arrangement and coding capacities of the three viral genomes are otherwise identical.

The nucleotide sequence of phage f1 DNA (Fig. 2) differs from that of M13 DNA (46) by 52 nucleotide changes and from that of phage fd (2, 3) by 186 changes (including the nucleotide deletion mentioned above). The nucleotide differences between f1 DNA and the DNAs of the other two phages are indicated in Fig. 2. Amino acid substitutions that would result from the nucleotide changes when the corresponding translated sequences in the other phages are compared with phage f1 DNA are also indicated in Fig. 2. The positions at which these changes occur appear to be specific for each bacteriophage DNA and presumably have arisen through independent mutations of a common ancestor. Of the nucleotide substitutions observed between the DNA of phage f1 and the DNAs of phages fd and M13, 25 involve the same nucleotide position, and 21 of these coincident changes give rise to the same replacement base in the last two genomes. The distribution of the nucleotide differences among the three viral genomes is, however, not random. The coding regions for genes VII and IX are fully conserved in each case, but a significant number of the nucleotide differences are clustered in the N-terminal coding region of gene IV. The putative promoter and ribosomal binding regions are fully conserved between phages f1 and M13, but the phage fd genome contains seven nucleotide changes within these regulatory regions. There are 23 nucleotide alterations in the major intergenic region of phage f1 DNA when it is compared with the other two viral genomes (Fig. 2), but none of these changes appears to have a significant effect on the structure or stability of the proposed double-stranded structures that have been implicated in viral DNA replication (39, 46).

TABLE 1. Restriction enzyme recognition sites in bacteriophage f1 RF DNA[a]

Enzyme	Recognition sequence	Position[b]
*Acc*I	GTAGAC	6090
*Alu*I	AGCT	39, 63, 203, 229, 333, 934, 1488, 1517, 2963, 3276, 3612, 4096, 5426, 5630, 5887, 6107, 6134, 6335
*Asu*I	GGGCC	5724
*Bam*HI	GGATCC	2220
*Bbv*I	GCAGC	932, 5536
	GCTGC	1367, 2521, 3132, 4871
*Cla*I	ATCGAT	6039
*Dde*I	CTAAG	233, 1099, 4013, 4093, 4121
	CTCAG	1417, 1784, 2015, 4040, 4079, 4881, 5262, 5370, 6065, 6218
	CTGAG	1371, 1847, 1862, 1877, 1901, 1973, 2318, 2333, 2348, 2363, 6346
	CTTAG	3361
*Eco*B	TGAATATCCGGTGCT	959
	TGAGGCTTTATTGCT	6347
*Eco*PI	GGTCT	316, 1017, 5828
	AGACC	429, 4043, 6092
*Eco*RII[c]	CCTGG	1014, 1966
*Hae*I	AGGCCA	2244
	TGGCCT	5344
*Hae*II	GGCGCT	2710
	AGCGCC	5559, 5567
	AGCGCT	3039
*Hae*III	GGCC	1396, 2245, 2554, 5239, 5345, 5414, 5725, 5867, 6180
*Hga*I	GACGC	526, 2164, 2479, 3237, 4083, 5158
*Hgi*AI	GTGCTC	4743, 5465
*Hha*I	GCGC	44, 873, 1011, 1085, 1177, 1470, 2195, 2467, 2711, 3040, 3096, 3408, 3598, 4312, 4995, 5490, 5503, 5512, 5534, 5560, 5568
*Hin*dII	GTTAAC	6405
*Hin*fI	GAATC	136, 723, 4349, 5120, 6198
	GACTC	1403, 4072, 5329, 5788, 6061
	GATTC	490, 511, 2497, 3258, 3418, 3742, 3838, 4117, 5375, 5438, 6042
	GAGTC	2011, 2845, 5766
*Hpa*II (*Hap*II)	CCGG	314, 966, 1095, 1924, 2378, 2396, 2552, 3370, 3842, 4018, 5614, 5995, 6118, 6178
*Hph*I	GGTGA	1376, 1774, 1909, 2398, 2542, 2581, 2620, 2626, 4346, 4847, 5117, 5706, 6162
	TCACC	1503, 2635, 4923, 6188
*Mbo*I	GATC	216, 1382, 1714, 2221
*Mbo*II	TCTTC	781, 4075, 4271, 4937, 5255, 5587, 5962
	GAAGA	3912

TABLE 1—*Continued*

Enzyme	Recognition sequence	Position[b]
*Mnl*I	CCTC	254, 373, 560, 587, 655, 1039, 1088, 1231, 1297, 1318, 1326, 1345, 1416, 1506, 1897, 1945, 2020, 2263, 2269, 2673, 2677, 2894, 3052, 3321, 3352, 3703, 4307, 4698, 4772, 4921, 4926, 5347, 5416, 5681, 6095, 6113
	GAGG	484, 625, 1373, 1732, 1834, 1849, 1864, 1879, 2008, 2218, 2320, 2335, 2350, 2365, 3336, 4021, 4820, 4835, 5447, 6243, 6348
*Rsa*I	GTAC	173, 280, 1022, 1165, 1769, 1796, 1889, 1905, 1970, 2133, 3467, 3668, 4190, 4380, 5384, 5486, 6000, 6322
*Sfa*NI	GATGC	25, 388, 1354, 4850
	GCATC	3979
*Taq*I	TCGA	336, 1127, 1508, 1949, 2528, 2815, 3455, 3694, 5683, 6040
*Tha*I	CGCG	43, 347, 1119, 1176, 2466, 3355, 3409, 3599, 3952, 4313, 4994, 5489, 5513, 5533, 5909
*Xho*II	AGATCC	215
	GGATCC	2220

[a] No site for: *Acy*I, *Ava*I, *Ava*II, *Ava*III, *Avr*II, *Bal*I, *Bcl*I, *Bgl*I, *Eca*I, *Eco*K, *Eco*RI, *Hinc*II, *Hin*dIII, *Hpa*I, *Kpn*I, *Mst*I, *Pst*I, *Pvu*I, *Pvu*II, *Sal*I, *Sma*I, *Sst*I, *Sst*II, *Xba*I, *Xho*I.

[b] The position of the first nucleotide of the recognition site is given in each case.

[c] Phage f1 grown on *E. coli* K38 is modified and therefore resistant to *Eco*RII (45).

Codon usage. The usage of codons in the structural genes of bacteriophage f1 (with the exception of X) is shown in Table 2 and compared with the corresponding figures that we have derived by computer analysis of the sequences of bacteriophages fd (2, 3) and M13 (46). The high frequency of T residues in the third codon position observed for phages φX174 and G4 (8) is also characteristic of the three filamentous coliphages f1, fd, and M13. Of the total number of codons derived from the phage f1 DNA sequence, 49% bear a T residue in the third position, compared with 43% in phage φX174 and 33.8% in phage G4. Most of the

TABLE 2. Comparison of codon usage in the structural genes (excluding X) of bacteriophages f1, fd and M13

Codon	Phage			Codon	Phage			Codon	Phage			Codon	Phage		
	f1	M13	fd		f1	M13	fd		f1	M13	fd		f1	M13	fd
UUU	69	71	67	UCU	97	99	92	UAU	66	66	65	UGU	16	16	16
UUC	38	36	39	UCC	29	30	33	UAC	12	12	14	UGC	8	8	8
UUA	63	64	65	UCA	32	33	35	UAA	5	5	5	UGA	3	3	3
UUG	29	28	32	UCG	11	8	9	UAG	1	1	1	UGG	18	18	18
CUU	46	47	49	CCU	48	48	46	CAU	12	13	12	CGU	31	31	32
CUC	22	22	17	CCC	8	9	9	CAC	5	5	6	CGC	16	16	16
CUA	9	8	6	CCA	14	14	13	CAA	35	35	35	CGA	6	6	5
CUG	27	27	26	CCG	16	15	18	CAG	43	43	43	CGG	2	2	1
AUU	72	72	72	ACU	65	66	60	AAU	87	86	82	AGU	13	13	14
AUC	16	16	16	ACC	20	19	23	AAC	22	22	23	AGC	12	13	11
AUA	20	21	20	ACA	12	11	15	AAA	72	72	73	AGA	10	10	11
AUG	33	33	33	ACG	12	12	11	AAG	36	36	34	AGG	4	4	5
GUU	96	95	98	GCU	60	61	58	GAU	72	74	72	GGU	92	93	98
GUC	17	18	18	GCC	16	17	16	GAC	38	35	38	GGC	54	52	51
GUA	28	29	25	GCA	29	29	28	GAA	37	37	40	GGA	5	6	5
GUG	8[a]	6[a]	11[a]	GCG	15	13	17	GAG	33	33	31	GGG	11	11	9

[a] Includes initiation codon of gene III.

nucleotide changes observed between the three filamentous phages occur in this third codon position, the predominant change being a C → T or T → C transition. Only 16 of the total nucleotide substitutions result in a change in the corresponding amino acid coded for in phages M13 and fd (Fig. 2), 13 of these changes being the result of changes in either the first or second codon positions.

The relative number of nucleotide changes among the genomes of the three filamentous bacteriophages confirms earlier predictions based on pyrimidine tract distribution that, contrary to the evidence from serological (31) and bacterial restriction (16) studies, phages f1 and M13 are more closely related than are f1 and fd or M13 and fd.

ACKNOWLEDGMENTS

We are grateful to Gillian McCrorie for her skilled and patient technical assistance throughout this study, to P. A. Stockwell for his advice and assistance with the computer analyses employed in the later stages of the work, and to R. Staden for providing us with tapes of his programs.

This study was supported through project and program grants from the Medical Research Council of New Zealand and grants for equipment from the Medical Research Council of New Zealand, the Medical Distribution Committee of the New Zealand Lottery Board, and the New Zealand University Grants Committee.

LITERATURE CITED

1. **Bailey, G. S., D. Gillett, D. F. Hill, and G. B. Petersen.** 1977. Automated sequencing of insoluble peptides using detergent. J. Biol. Chem. **252**:2218–2225.
2. **Beck, E., R. Sommer, E. A. Auerswald, C. Kurz, B. Zink, G. Osterburg, H. Schaller, K. Sugimoto, H. Sugisaki, T. Okamoto, and M. Takanami.** 1978. Nucleotide sequence of bacteriophage fd DNA. Nucleic Acids Res. **5**:4495–4503.
3. **Beck, E., and B. Zink.** 1981. Nucleotide sequence and genome organisation of filamentous bacteriophages f1 and fd. Gene (Amst.) **16**:35–58.
4. **Boyer, H. W., L. T. Chow, A. Dugaiczyk, J. Hedgpeth, and H. M. Goodman.** 1973. DNA substrate site for the *Eco*RII restriction endonuclease and modification methylase. Nature (London) New Biol. **244**:40–43.
5. **Cuypers, T., F. J. van der Ouderaa, and W. W. de Jong.** 1974. The amino acid sequence of gene 5 protein of bacteriophage M13. Biochem. Biophys. Res. Commun. **59**:557–563.
6. **Denhardt, D. T.** 1975. The single-stranded DNA phages. Crit. Rev. Microbiol. **4**:161–223.
7. **Edens, L., R. N. H. Konings, and J. G. G. Schoenmakers.** 1975. Physical mapping of the central terminator for transcription on the bacteriophage M13 genome. Nucleic Acids Res. **2**:1811–1820.
8. **Godson, G. N., J. C. Fiddes, B. G. Barrell, and F. Sanger.** 1978. Comparative DNA sequence analysis of the G4 and ϕX174 genomes, p. 51–86. *In* D. Dressler and D. S. Ray (ed.), The single-stranded DNA phages. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
9. **Hill, D. F., and G. B. Petersen.** 1980. Nucleotide sequences in bacteriophage f1 DNA: nucleotide sequence of genes V, VII, and VIII. J. Virol. **34**:40–50.
10. **Konings, R. N. H., T. Hulsebos, and C. A. van den Hondel.** 1975. Identification and characterization of the in vitro synthesized gene products of bacteriophage M13. J. Virol. **15**:570–584.
11. **Laskey, R. A., and A. D. Mills.** 1977. Enhanced autoradiographic detection of ^{32}P and ^{125}I using intensifying screens and hypersensitized film. FEBS Lett. **82**:314–316.
12. **Lin, T.-C., R. E. Webster, and W. Konigsberg.** 1980. Isolation and characterization of the C and D proteins coded by gene IX and gene VI in the filamentous bacteriophage f1 and fd. J. Biol. Chem. **255**:10331–10337.
13. **Ling, V.** 1972. Pyrimidine sequences from the DNA of bacteriophages fd, f1 and ϕX174. Proc. Natl. Acad. Sci. U.S.A. **69**:742–746.
14. **Ling, V.** 1972. Fractionation and sequences of the large pyrimidine oligonucleotides from bacteriophage fd DNA. J. Mol. Biol. **64**:87–102.
15. **Loeb, T.** 1960. Isolation of a bacteriophage specific for the F^+ and Hfr mating types of Escherichia coli K-12. Science **131**:932–933.
16. **Marvin, D. A., and B. Hohn.** 1969. Filamentous bacterial viruses. Bacteriol Rev. **33**:172–209.
17. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
18. **Messing, J., B. Gronenborn, B. Müller-Hill, and P. H. Hofschneider.** 1977. Filamentous coliphage M13 as a cloning vehicle. Insertion of a *Hin*dII fragment of the *lac* regulatory region in M13 replicative form *in vitro*. Proc. Natl. Acad. Sci. U.S.A. **74**:3642–3646.
19. **Model, P., and N. D. Zinder.** 1974. *In vitro* synthesis of bacteriophage f1 proteins. J. Mol. Biol. **83**:231–251.
20. **Moore, P., and B. S. Strauss.** 1979. Sites of inhibition of *in vitro* DNA synthesis in carcinogen- and UV-treated ϕX174 DNA. Nature (London) **278**:664–666.
21. **Nakashima, Y., and W. Konigsberg.** 1974. Reinvestigation of a region of the fd bacteriophage coat protein sequence. J. Mol. Biol. **88**:598–600.
22. **Ohmori, H., J. Tomizawa, and A. M. Maxam.** 1978. Detection of 5-methylcytosine in DNA sequences. Nucleic Acids Res. **5**:1479–1485.
23. **Petersen, G. B., and J. M. Reeves.** 1969. The occurrence of long sequences of consecutive pyrimidine deoxyribonucleotides in the DNA of bacteriophage f1. Biochim. Biophys. Acta **179**:510–512.
24. **Pieczenik, G.** 1980. Predicting codon function from nucleotide sequence or survival of "fitness" of tRNA. Proc. Natl. Acad. Sci. U.S.A. **77**:3539–3543.
25. **Pieczenik, G., P. Model, and H. D. Robertson.** 1974. Sequence and symmetry in ribosome binding sites of bacteriophage f1 RNA. J. Mol. Biol. **90**:191–214.
26. **Ravetch, J., K. Horiuchi, and P. Model.** 1977. Mapping of bacteriophage f1 ribosome binding sites to their cognate genes. Virology **81**:341–351.
27. **Ravetch, J. V., K. Horiuchi, and N. D. Zinder.** 1978. Nucleotide sequence of the recognition site for the restriction-modification enzyme of *Escherichia coli* B. Proc. Natl. Acad. Sci. U.S.A. **75**:2266–2270.
28. **Ravetch, J. V., K. Horiuchi, and N. D. Zinder.** 1979. DNA sequence analysis of the defective interfering particles of bacteriophage fl. J. Mol. Biol. **128**:305–318.
29. **Ray, D. S.** 1977. Replication of filamentous bacteriophages, p. 105–178. *In* H. Fraenkel-Conrat and R. R. Wagner (ed.), Comprehensive virology, vol. 7. Plenum Publishing Corp., New York.
30. **Roberts, R. J.** 1980. Restriction and modification enzymes and their recognition sequences. Nucleic Acids Res. **8**:r63–r80.
31. **Salivar, W. O., H. Tzagoloff, and D. Pratt.** 1964. Some physical-chemical and biological properties of the rod-shaped coliphage M13. Virology **24**:359–371.
32. **Sanger, F., and A. R. Coulson.** 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J. Mol. Biol. **94**:441–448.
33. **Sanger, F., and A. R. Coulson.** 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. **87**:107–110.
34. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol.

143:161–178.
35. **Sanger, F., J. E. Donelson, A. R. Coulson, H. Kössel, and D. Fischer.** 1974. Determination of a nucleotide sequence in bacteriophage f1 DNA by primed synthesis with DNA polymerase. J. Mol. Biol. **90**:315–333.
36. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
37. **Simons, G. F. M., R. N. H. Konings, and J. G. G. Schoenmakers.** 1979. Identification of two new capsid proteins in bacteriophage M13. FEBS Lett. **106**:8–12.
38. **Staden, R.** 1977. Sequence data handling by computer. Nucleic Acids Res. **4**:4037–4051.
39. **Suggs, S. V., and D. S. Ray.** 1979. Nucleotide sequence of the origin for bacteriophage M13 DNA replication. Cold Spring Harbor Symp. Quant. Biol. **43**:379–388.
40. **Sugimoto, K., H. Sugisaki, T. Okamoto, and M. Takanami.** 1977. Studies on bacteriophage fd DNA. IV. The sequence of messenger RNA for the major coat protein gene. J. Mol. Biol. **110**:487–507.
41. **Takanami, M.** 1974. Restriction endonucleases AP, GA and H-I from three Haemophilus strains. Methods Mol. Biol. **7**:113–133.
42. **Tate, W. P., and G. B. Petersen.** 1974. A catalogue of the pyrimidine deoxyribooligonucleotides found in bacteriophage f1 DNA. Virology **57**:64–76.
43. **Tate, W. P., and G. B. Petersen.** 1974. The pyrimidine oligodeoxyribonucleotides from the DNA molecules of bacteriophages f1, fd and M13. Virology **57**:77–84.
44. **Tate, W. P., and G. B. Petersen.** 1974. Structure of the filamentous bacteriophages: orientation of the DNA molecule within the phage particle. Virology **62**:17–25.
45. **van den Hondel, C. A., and J. G. G. Schoenmakers.** 1976. Cleavage maps of the filamentous bacteriophages M13, fd, f1, and ZJ/2. J. Virol. **18**:1024–1039.
46. **van Wezenbeek, P. M. G. F., T. J. M. Hulsebos, and J. G. G. Schoenmakers.** 1980. Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: comparison with phage fd. Gene (Amst.) **11**:129–148.
47. **Vovis, G. F., K. Horiuchi, and N. D. Zinder.** 1975. Endonuclease R · *Eco*RII restriction of bacteriophage f1 DNA in vitro: ordering of genes V and VII, location of an RNA promoter for gene VIII. J. Virol. **16**:674–684.
48. **Zinder, N. D., R. C. Valentine, M. Roger, and W. Stoeckenius.** 1963. f1, a rod-shaped male-specific bacteriophage that contains DNA. Virology **20**:638–640.