JOURNAL OF BACTERIOLOGY, Apr. 1990, p. 1889–1898
0021-9193/90/041889-10$02.00/0

Vol. 172, No. 4

Nucleotide Sequence and Transcription of the Right Early Region of Bacteriophage PRD1

DAN GERENDASY AND JUNETSU ITO*

Department of Microbiology and Immunology, The University of Arizona Health Sciences Center, Tucson, Arizona 85724

Received 26 July 1989/Accepted 31 December 1989

We have sequenced the rightmost 1,700 base pairs of bacteriophage PRD1. This region encompasses the right early region and completes the sequence of all PRD1 early functions. We have also mapped the 5′ initiation site of right early transcripts in vivo and in vitro. This has allowed us to assign gene XII to an open reading frame and suggests that another open reading frame may also be expressed. Gene XII, which has been implicated in the replication process and the regulation of gene expression, is predicted to encode a protein with a molecular mass of 16.7 kilodaltons. Data base searches have revealed no significant homology between the product of this gene and other proteins. Transcription mapping studies have revealed that right early transcripts elongate from right to left and have enabled us to identify the right early promoter. This promoter behaves identically in vivo and in vitro. We also demonstrate that this promoter directs the transcription of two RNAs of different sizes in vitro.

PRD1 is a pilus-specific bacteriophage which infects a wide variety of gram-negative bacteria ranging from *Escherichia coli* and *Salmonella typhimurium* to *Pseudomonas aeruginosa* (3, 35). Its genome consists of one linear, duplex DNA which is approximately 14.7 kilobases (kb) long and which is covalently bound to a terminal protein at each of its 5′ ends (for a review, see reference 34). It also contains a perfect inverted terminal repeat of 110 to 111 base pairs bp (16, 41). Replication of the PRD1 genome occurs by a protein-priming mechanism which is similar to that used by members of the family *Adenoviridae* and the *Bacillus* bacteriophage ϕ29 (for reviews, see references 27, 34, and 40). Bacteriophage PRD1 and related strains have elicited the interest of a number of researchers because their coats contain phospholipid as well as protein (7, 29). Furthermore, PRD1 is a convenient model system with which to study protein-primed linear DNA replication, because its genome is smaller than that of any of the other known phages which replicate by a protein-priming mechanism. Also, large quantities of this phage can be isolated, and numerous conditionally lethal mutant phage and host strains exist. Because it infects *E. coli* and *S. typhimurium*, it is highly amenable to genetic and biochemical manipulation.

PRD1 expresses three genes immediately upon infection; they have been designated as early genes (genes I, VIII, and XII). All other genes, which have been mapped, have been categorized into groups consisting of middle early (genes XV and XIX) and late genes (33). Genes I and VIII, which specify DNA polymerase and terminal protein, respectively (21, 24, 42), reside on what is conventionally called the left end of the genome (see Fig. 2), while gene XII has been mapped to the far right (31). Gene XII is highly expressed and encodes a protein with an apparent molecular mass of 20 kilodaltons (kDa) as determined by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis. This protein has been implicated in the process of DNA replication, although its exact function has not been determined. It also appears to play a role in the regulation of early-gene expression. Mutations within this gene lead to decreased synthesis of DNA and failure to down regulate early-gene expression late in infection (33).

PRD1 DNA replication is initiated when phage-specific DNA polymerase catalyzes the deoxyguanylylation of a tyrosine residue within p8 (the product of gene VIII). This reaction is dependent on the presence of the phage DNA-protein complex (2). The consequent dGMP-p8 complex is then used to initiate chain elongation (45). Full-length DNA can be synthesized in vitro in the absence of p12, but the process is approximately 100 times slower than DNA synthesis in vivo, and synthesis of the second daughter strand has not yet been demonstrated (45). Since host extracts are required for in vitro DNA synthesis, it is possible that p12 is replaceable by host functions.

Previously, the left early and middle early regions of PRD1 were sequenced and the open reading frames (ORFs) encoding two early proteins (genes I and VIII) were identified (21, 24, 34). Middle early gene XV, on the left, was also tentatively assigned to an open reading frame (36). In addition, the 5′ end of the left early transcript (that of genes I, VIII, and, possibly, XV) was mapped and a promoter was identified (42). A putative promoter preceding an ORF which is believed to be gene XV has been identified; however, it has not been mapped (31).

These analyses have suggested that the transcriptional organization of PRD1 differs from that of ϕ29. Although the general genomic organization of these two phages is similar in that their late and early genes are located in the middle and on the ends, respectively, the left early genes are oriented in opposite directions. In ϕ29 all early genes are transcribed from right to left while late genes are transcribed from left to right (8, 25, 32, 39). PRD1, left early transcription occurs from left to right. The direction in which late PRD1 genes are transcribed has been inferred to be left to right based on the polarity of certain mutations within the late region (31). Although ϕ29 and PRD1 left early transcription occurs in opposite directions, the order in which their left early genes are transcribed has been maintained because the relative

* Corresponding author.

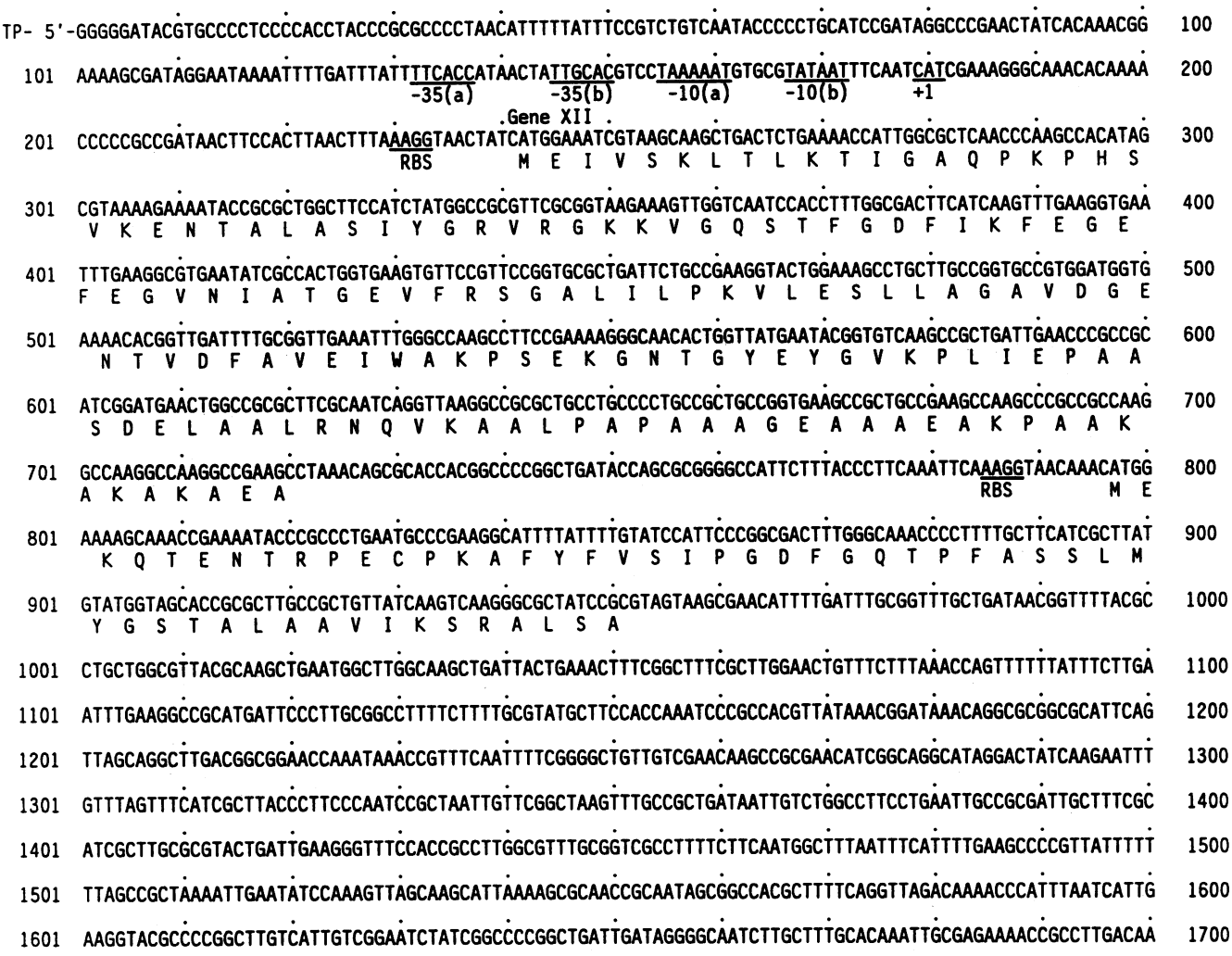

FIG. 1. Nucleotide sequence of the right 1,700 bp of PRD1. The predicted primary structure of p12 and possibly p19 is shown (see text). The putative −35 and −10 regions of the overlapping right early promoters and possible RBSs are underlined. The start site of right early transcription is also indicated. This sequence has been submitted to GenBank and assigned the accession number M30146.

positions of the genes specifying DNA polymerase and terminal protein are reversed.

To better understand the role of gene XII with respect to viral replication and transcription and to locate and possibly identify additional early genes, we have sequenced the entire right early region of PRD1. In addition, we have mapped the initiation sites of right early transcripts synthesized in vitro and in vivo, demonstrating that in vivo initiation can be faithfully reproduced in vitro. These results have allowed us to assign of gene XII and possibly gene XIX to specific ORFs (Fig. 1). They have also permitted us to identify the right early promoter and two possible terminators, as well as the direction of right early transcription. Furthermore, in vitro transcription experiments indicate that two transcripts of different sizes are generated from the right early promoter.

MATERIALS AND METHODS

Nucleotides and enzymes. [γ-^{32}P]ATP (ca. 6,000 Ci/mmol), [α-^{35}S]dATP (ca. 600 Ci/mmol), and [α-^{32}P]UTP (ca. 800 Ci/mmol) were obtained from Du Pont, NEN Research Products. Restriction endonucleases were purchased from Bethesda Research Laboratories, Inc.; New England Biolabs, Inc.; or Boehringer Mannheim Biochemicals. T4 polynucleotide kinase, calf intestinal alkaline phosphatase, and T4 DNA ligase were obtained from Promega Biotec. Sequenase and dideoxynucleotide mixtures came from U.S. Biochemical Corp. *E. coli* RNA polymerase was obtained from Boehringer Mannheim or Promega Biotec. T7 and SP6 RNA polymerases, RNase-free DNase, RNasin RNase inhibitor, and avian myeloblastosis virus reverse transcriptase were purchased from Promega Biotec.

Bacterial strains, plasmids, and phage. Bacteriophage PRD1 was grown in *S. typhimurium* LT2 carrying the drug resistance plasmid pLM2. This strain was cultivated in the presence of kanamycin (50 μg/ml) to select for the presence of pLM2, which is required for phage absorption. Strains and phage were all generously provided by Leonard Mindich, Public Health Research Institute of the City of New York. M13mp19 clones were propagated in *E. coli* JM101.

Preparation of phage DNA. Host cells were grown to a density of approximately 5×10^8 in LB broth containing kanamycin (50 μg/ml) and infected at a multiplicity of infection of about 10. After lysis, phage particles were precipitated overnight in 12% polyethylene glycol–0.4 M

NaCl and purified through a 5 to 20% (wt/vol) sucrose gradient containing 10 mM Trizma hydrochloride (pH 7.0), 5% NaCl, 1 mM $CaCl_2$, and 0.5% Tryptone. Deproteinized phage DNA was isolated by suspending phage particles in 10% SDS–200 μg of proteinase K per ml and heating for 15 min at 60°C. The DNA was then extracted twice with phenol and precipitated in ethanol.

Nucleotide sequencing. The two *Taq*I fragments located second and third from the right side of the genome (Fig. 1) were electroeluted from an agarose gel, restricted with either *Hpa*II or *Hin*pI, and cloned into M13mp19. The right-most *Bst*XI fragment was treated similarly, except that it was purified from the remaining DNA through a sucrose gradient. Dideoxy sequencing was performed with Sequenase, by using the protocol supplied with the Sequenase kit by U.S. Biochemical Corp. In some cases, when clones were lacking or confirmation was required, sequencing was performed by the method of Maxam and Gilbert (30). Every nucleotide was sequenced at least once on both strands.

In vitro transcription. Reaction mixtures contained between 0.5 and 2.5 μg of DNA. The quantity used was adjusted empirically with each new lot of *E. coli* RNA polymerase. All in vitro transcription reactions which did not entail radiolabeling contained the following in a final volume of 40 μl: 40 mM Tris (pH 8.0), 125 mM KCl, 10 mM $MgCl_2$, 0.05% bovine serum albumin, 0.1 mM EDTA, 50 μM dithiothreitol, 0.25 mM each nucleoside triphosphate, 40 U of RNasin (1 μl), and 1 U of *E. coli* RNA polymerase (1 μl). The reaction was incubated at 37°C for 30 min, at which time 3 μl of rifampin (400 μg/ml) was added along with 0.5 μl of each nucleoside triphosphate (10 mM each). After 5 min at 37°C 10 U of DNase was added, and the reaction was allowed to incubate for an additional 15 min. The resultant RNA was then extracted with 1 volume of phenol and then with 1 volume of chloroform, brought up to 0.3 M sodium acetate, and precipitated in 2.5 volumes of ethanol.

Southern blots. Southern blotting was carried out as described previously (43). Phage DNA was restricted with *Bst*XI, which cuts the PRD1 genome once approximately 1,500 bp from the right terminus. The two fragments were resolved on a 0.5% agarose gel, and the fragment originating from the right side was electroeluted. This was, in turn, digested with either *Fnu*4HI, *Hae*III, *Hin*pI, *Taq*I, *Hin*fI, *Sfa*NI, or *Hae*II. DNA which had been restricted with the first three enzymes listed was resolved on a 5% polyacrylamide gel, while the remainder was resolved on a 0.7% agarose gel. At the same time, the three single-stranded M13 clones were also electrophoresed into the agarose gel. One clone contained the second *Taq*I fragment from the right of the PRD1 genome (Fig. 1), and the other two contained the third *Taq*I fragment from the right, in opposite orientations. The agarose gel was blotted onto GeneScreen Plus (Du Pont, NEN) overnight through the capillary action of a transfer buffer containing 0.4 M NaOH and 0.6 M NaCl (6). The polyacrylamide gel was soaked in the same transfer buffer for 25 min, washed with three changes of double-distilled H_2O, and equilibrated with TBE (0.1 M Tris, 0.1 M boric acid, 2 mM EDTA). DNA was then electrophoretically transferred onto GeneScreen Plus by using the LKB semidry Multiphore Nova Blot system. Transfer was carried out for 8 h at 0.8 mA/cm^2. Blots were allowed to air dry overnight.

Hybridization of PRD1 right early in vitro transcripts to Southern blots. In vitro transcription was performed as described above, except that in the first 30-min incubation, cold UTP was replaced with [α-^{32}P]UTP (50 μCi in 5 μl). After 30 min this was chased with cold UTP for an additional 30 min. For the template, a *Hin*fI digest of the right *Bst*XI fragment was used. Before the resultant RNA was used for hybridization, it was passed over a Nu-clean R50 disposable spun column (International Biotechnologies, Inc.).

Southern blots were prehybridized for 3 h at 65°C in 10 to 20 ml of hybridization buffer (1% SDS, 10% dextran sulfate, 1.0 M NaCl). Labeled RNA probe was suspended in 0.5 ml of the same buffer, heated to 90°C for 4 min, and added to the Southern blot. Hybridization was carried out overnight at 65°C. The Southern blots were then washed six times for 5 min each in 2× SSC–0.1% SDS at 65°C (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and once in 0.1× SSC–0.1% SDS for 30 min at 45°C (23). They were then autoradiographed overnight without intensifying screens.

Isolation of RNA from phage-infected cells. The hot-phenol method used here for RNA isolation is almost identical to that described by Aiba et al. (1). A 100-ml culture of *S. typhimurium* LT2 in LB broth, with a density of approximately 5 × 10^8 cells per ml, was infected with PRD1 at a multiplicity of infection of approximately 20 phage per cell. After 15 min the cells were harvested and suspended in 3 ml of chilled 0.3 M sodium acetate–0.5% SDS–10 mM EDTA. Hot (60°C) phenol (3 ml) was then added. The phenol contained 0.1% 8-hydroxyquinoline and had been previously equilibrated against 0.1 M Tris (pH 8.0)–0.2% β-mercaptoethanol. This mixture was incubated at 60°C for 5 min with constant agitation. After centrifugation, the aqueous layer was withdrawn and extracted three more times with phenol under similar conditions. The RNA was then precipitated by the addition of 9 ml of ethanol, reprecipitated twice more, washed with 80% ethanol, and suspended in 0.5 ml of diethylpyrocarbonate-treated H_2O.

Primer extension. Primer extension was performed on 2 μl of RNA isolated from phage-infected cells as describe above and on 2 μl of RNA synthesized in vitro, also as described above. Two primers were used. One was a 26-mer, 5′-ACGC TATGTGGCTTGGGTTGAGCGCC-3′, which would be expected to hybridize approximately 100 nucleotides from the 5′ end of any RNAs that might arise from ORF B4 (Fig. 1). The other was a 30-mer, 5′-AACTGGCCGCGCTTCGCAAT CAGGTTAAGG-3′ which would be predicted to hybridize approximately 100 nucleotides away from the 5′ end of any RNAs transcribed from ORF A6 (Fig. 1).

The synthetic primer (0.5 ng) was hybridized to potential RNA transcripts in a total volume of 10 μl containing 80 mM Tris (pH 7.5), 40 mM $MgCl_2$, and 100 mM NaCl. This mixture was heated to 65°C and allowed to cool to room temperature over a period of 2 to 4 h. The following were then added directly to the hybridization buffer: 1 μl each of dGTP, dCTP, and dTTP from a 50 mM stock; 1 μl of 100 mM dithiothreitol; 1 μl of RNasin (40 U); 1 μl of [α-^{35}S]ATP; and 3 μl of avian myeloblastosis virus reverse transcriptase (33 U). The mixture was then kept at 37°C for 15 min, after which 1 μl of 50 mM dATP was added and the mixture was incubated at 42°C for 45 min. Stop buffer was then added (12 μl of 95% formamide, 20 mM EDTA, 0.05% bromphenol blue, 0.05% xylene cyanol FF), the sample was brought to 90°C for 5 min, and 2 μl was applied to a sequencing gel.

Two different sequence ladders were used as molecular mass markers. In both instances, the same oligonucleotides used for primer extension were used to prime the sequencing reaction. When primer extension was performed on transcripts arising from ORF B4, genomic PRD1 was sequenced directly after partial exonuclease III digestion. The sequences used during our attempts to perform primer extension on transcripts arising from ORF A6 were created from

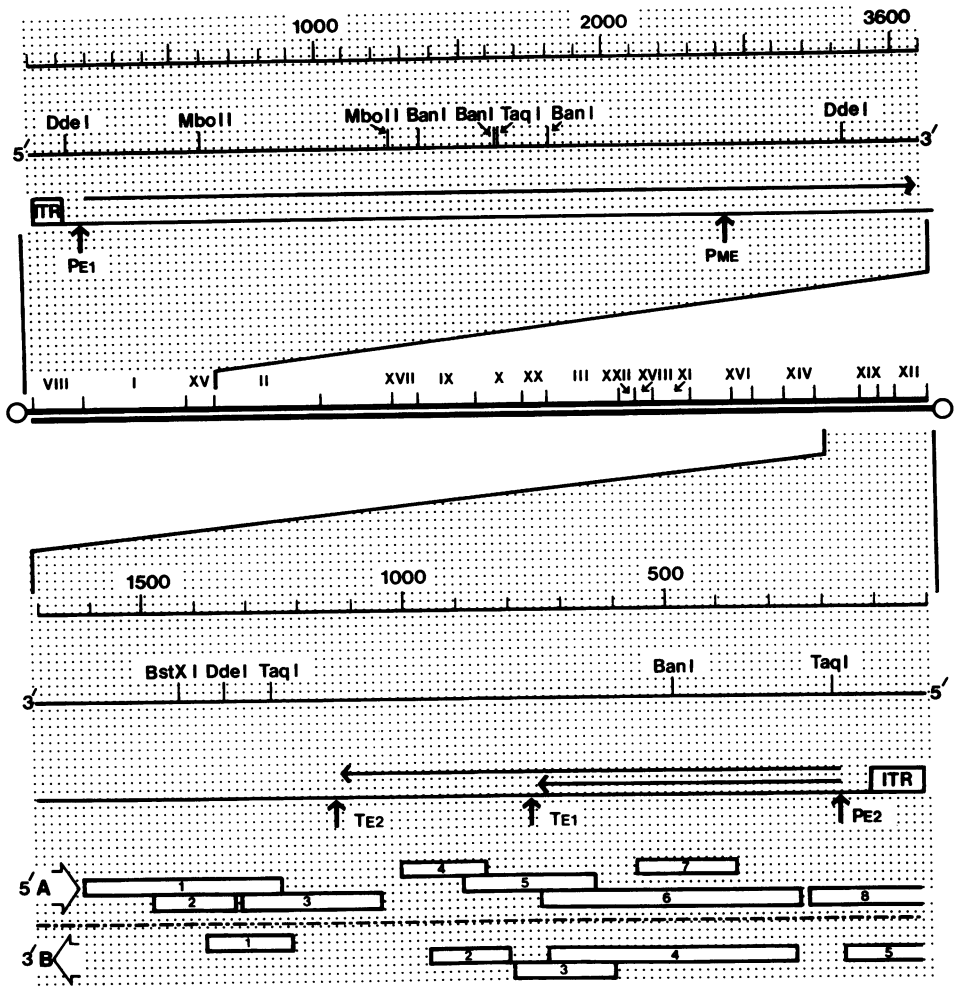

FIG. 2. Genetic, physical, transcription, and ORF map of left and right very early and medium early regions of PRD1. The genetic map was adapted from McGraw et al. (31). Genes I, VIII, and XII are expressed very early after infection, whereas genes XV and XIX are expressed slightly later. The remainder are expressed late in infection. Early and middle early regions of both sides of the genome have been sequenced, and the left early promoter (PE1) was mapped previously (13, 16, 24, 27; see the text). Promoter PME is a putative middle early promoter which precedes gene XV (24). Right early promoter (PE2), putative terminators (TE1 and TE2), and right transcripts are deduced from data presented here. ORFs which contain 50 or more residues are graphically depicted at the bottom.

an M13 clone containing the second *Taq*I fragment from the right.

Electrophoreses of RNA transcribed in vitro. In vitro transcription was performed as described above with either whole genomic phage DNA for the template or phage DNA which had been restricted with *Ban*I, *Bst*XI, *Bst*NI, *Dde*I, *Taq*I, or *Mbo*II. Both denaturing polyacrylamide and agarose gels were used. Polyacrylamide gels were prepared in exactly the same way as a sequencing gel, and the RNA was treated accordingly. After precipitation the RNA was suspended in loading buffer (95% formamide, 20 mM EDTA, 0.05% bromphenol blue, 0.05% xylene cyanol FF), heated at 90°C for 4 min, and loaded onto a 6% denaturing polyacrylamide sequencing gel (7 M urea; ratio of bisacrylamide to polyacrylamide, 1:20). To resolve higher-molecular-weight RNAs, we used 1% agarose–formaldehyde gels. These contained 2.2 M formaldehyde, 40 mM morpholinepropane sulfonic acid (MOPS), 6 mM sodium acetate, and 1 mM EDTA. RNA was suspended in the same gel buffer plus 50% formamide. Running buffer, which was not circulated, consisted of 400 mM MOPS, 60 mM sodium acetate, and 10 mM EDTA. All solutions were at pH 7.0. Gels were washed in two changes of water to remove excess urea or formaldehyde, stained with ethidium bromide, and then destained in three or four changes of water over 4 to 12 h.

Marker RNAs were synthesized with SP6 and T7 RNA polymerase from Riboprobe Gemini positive control templates provided by Promega Biotec. The conditions recommended by the manufacturer were used. This template consists of a mixture of DNA fragments containing SP6 and T7 promoters. Transcription with SP6 RNA polymerase generates RNAs of 11, 172 and 1,386 bases, whereas transcription with T7 RNA polymerase generates transcripts of 43, 679, and 1,418 bases.

RESULTS

Nucleotide sequence of the right terminus of the PRD1 genome. The DNA sequence shown in Fig. 1 covers 1,700 bp of the right-hand side of the PRD1 genome. Nucleotides 1 through 150 were published previously and are included here for clarity (16, 41). Based on physical and genetic mapping data, this sequence would be expected to cover the right early region as well as part of the right, middle early and late

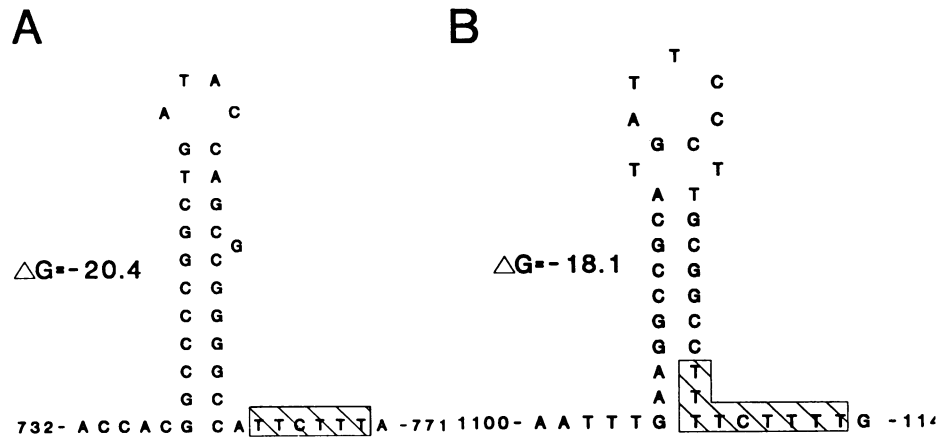

FIG. 3. Putative rho-independent transcription terminators. Free energies were calculated by using the values published by Freier et al. (14). Strings of thymidines which are boxed are typical of rho-independent terminators. These terminators would be expected to terminate right early transcripts which elongate from right to left.

regions (31). Others have had difficulty cloning genes XII and XIX and have suggested that they may be lethal to *E. coli* (31, 34). Our M13 clones contained inserts originating from all areas of the early region except that which contained the right early promoter indicated in Fig. 1 and 2 (nucleotides 120 to 180). This region was therefore sequenced by the method of Maxam and Gilbert (30). Evidence that this region contains the right early promoter is presented below. Our success in cloning fragments containing genes XII and XIX may be due to our fortuitous choice of restriction enzyme (*Taq*I cleaves the right end of the genome immediately downstream from the promoter region). This leaves open the question of whether genes XII and XIX are actually lethal when under control of their own promoter or whether the promoter itself causes plasmid instability or interferes with the ability of M13 to replicate.

One interesting anomaly which came to light during sequencing of the right side of the PRD1 genome is that it contains a site which is resistant to digestion by *Hin*fI. According to our sequence data, *Hin*fI should cleave the rightmost *Bst*XI fragment four times. The *Hin*fI site at position 262, however, was completely insensitive to digestion, even when digestion was carried out overnight in the presence of a 100-fold excess of enzyme (data not shown). We have sequenced both strands of this region by the dideoxynucleotide method as well as the method of Maxam and Gilbert to confirm that the sequence data are correct. The reason for the insensitivity of this *Hin*fI site is not understood at present.

In previous studies, 150 bp of the left and right termini of the PRD1 genome were compared. This revealed a perfect inverted terminal repeat between 110 and 111 bp in length (16, 41). The sequence presented here allows us to extend this comparison. If one allows the insertion of a 2-bp gap in the right-hand terminal sequence between nucleotides 149 and 150, the 63 bp following the 111-bp inverted repeat on the right have 51% of their nucleotides in common with those on the left. This region of homology extends up to and includes the −10 region of both the left and right early promoters. Both of these promoters are located almost exactly the same distance from the termini of the genome (data not shown).

ORF analysis. All of the ORFs in the region sequenced which have the potential to encode a polypeptide of greater than 50 amino acids are depicted at the bottom on Fig. 2. The quality of the potential ribosome-binding sites which precede ORFs B2 and B4 is striking (Fig. 1). Analysis by the method of Stormo et al. (44) suggests that they are very likely to be functional in vivo. Fickett's analysis (13) and codon bias plots (17) indicate, very convincingly, that ORFs B4 and B2 encode proteins (data not shown). Codon preference analysis suggests that ORF B4 is very highly expressed, more so than any other PRD1 gene or potential gene so far sequenced. ORF B4 is also preceded by two overlapping putative *E. coli* promoters (Fig. 1). Evidence is presented below which proves that the second of these is used in vivo and in vitro and that ORFs B4 and B2 are transcribed.

We have identified a number of regions within the rightmost 1,700 bp which conform to the general structure of *E. coli* rho-independent terminators (37). Two of these (Fig. 3) could serve to terminate transcripts which originate from the far right-hand side of the genome and which elongate toward the left. These terminators would terminate transcripts of ORF B4 and ORF B2.

The physical location of ORF B4 and the analysis presented above strongly suggest that it is, in fact, gene XII. There are only two approximately positioned ORFs (A6 and B4) that could encode a protein with a molecular mass approaching the 20 kDa determined for gene XII by SDS-polyacrylamide gel electrophoresis. ORFs A6 and B4 are capable of encoding 19.0- and 16.7-kDa proteins, respectively. ORF B4 is preceded by a very plausible promoter and a ribosome-binding site (RBS) which is very probably functional. ORF A6 is preceded by neither. Translation of ORF A6 would, necessarily, initiate with a codon specifying either a valine or a leucine, which is very rare in *E. coli*, whereas translation of ORF B4 is predicted to initiate with a codon specifying methionine. The codon preference displayed by ORF B4 indicates that it is very highly expressed, which is consistent with the high level of expression of gene XII observed in other studies (33). Finally, ORF B4 is followed by a very reasonable terminator, which is probably not coincidental.

A rough map of right early transcription. Initial mapping experiments were performed by probing Southern blots with labeled RNA synthesized in vitro from phage DNA and *E. coli* RNA polymerase. The right *Bst*XI fragment of the PRD1 genome (Fig. 2) was isolated and secondarily digested with the enzymes indicated in Fig. 4A, B, and D. These digests were then electrophoresed, blotted onto GeneScreen Plus,

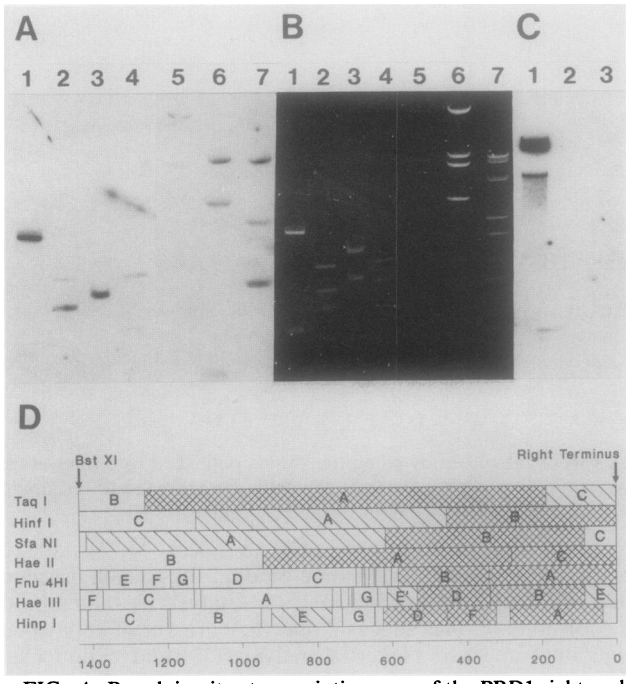

FIG. 4. Rough in vitro transcription map of the PRD1 right early region. (A) Southern blot of the gel shown in panel B. Lanes 1 to 7 correspond to lanes 1 to 7 in panel B. The right *Bst*XI fragment was digested with the following restriction enzymes prior to electrophoretic resolution: *Taq*I (lane 1), *Hin*fI (lane 2), *Sfa*NI (lane 3), *Hae*II (lane 4), *Fnu*4HI (lane 5), *Hae*III (lane 6), and *Hin*pI (lane 7). The Southern blot was probed with labeled RNA synthesized in vitro with *E. coli* polymerase and PRD1 DNA which had been restricted with *Hin*fI. (B) Gel corresponding to Southern blot in panel A. (C) Southern blot of single-stranded M13 clones of the second and third *Taq*I fragments from the right terminus. The same probe as described above was used. Lane 1 shows the second *Taq*I fragment from the right. The orientation of the insert is such that it would be expected to hybridize to transcripts which elongate from right to left with respect to the PRD1 genome. Lanes 2 and 3 show the third *Taq*I fragment from the right in opposite orientations. (D) Physical map of the right *Bst*XI fragment and summary of the results obtained from the Southern blot displayed in panel A. Fragments which hybridized to the RNA probe are cross-hatched. When the degree of hybridization varied, the fragment(s) which displayed the strongest signal are cross-hatched (▩) and those that displayed the weakest signal are hatched (▨). The nucleotide sequence of this region indicates that an additional *Hin*fI site should be located 262 bp from the right terminus; however, this site is completely insensitive to *Hin*fI digestion. We have therefore treated that fragment, which is labeled B, as one fragment.

and probed with radiolabeled RNA synthesized from *Hin*fI-digested PRD1 DNA. The results of this experiment are displayed in Fig. 4A and B. Figure 4D contains the relevant restriction map and a summary of the fragments which hybridized to the RNA probe.

The efficiencies with which labeled RNA hybridized to different restriction fragments varied. We have subjectively categorized the bands which appeared on film into dense and light bands. The high-intensity bands consist of restriction fragments that overlap the *Hin*fIB segment, which suggests that at least one major transcript originates within this fragment. The low-intensity bands may have a number of different origins. They could be derived from nonspecific transcription, specific but low-level transcription, or small quantities of the major transcript which originated from incompletely cut template. The lack of consistency concerning the positions of the low-intensity restriction fragments may be explained by differing efficiencies of transfer during the blotting procedures used (the *Fnu*4HI, *Hae*III, and *Hin*pI digests were electroblotted out of a 5% polyacrylamide gel, whereas the others were capillary blotted out of a 0.7% agarose gel). Restriction fragments *Taq*I-B and -C and *Hae*III-E and -E′ could not be resolved. In these cases, the decision on how to depict these restriction fragments in Fig. 4D was based on maintaining consistency with the rest of the data. In any case, this does not affect the conclusion that the major right transcript originates between *Taq*I-C and the left end of *Hin*fI-B. The following experiment provided information concerning the direction of transcription in this region.

We had at our disposal three single-stranded clones which were generated during the sequencing phase of these experiments. Two of these clones contained the 1.7-kb *Taq*I fragment (third from right) in opposite orientations. The other clone contained the 1.0-kb *Taq*I fragment (second from right) in an orientation that would be complementary to any RNAs transcribed from right to left in that region of the genome (Fig. 2). It would therefore be expected to hybridize with a transcript of ORF B4. The other two clones would be expected to hybridize only to right late and possibly middle early transcripts. Unfortunately, a clone containing the 1.0-kb *Taq*I fragment in the proper orientation to hybridize to transcripts of ORF A6 was not available. These three clones were electrophoresed into an agarose gel, blotted onto GeneScreen Plus, and probed with labeled RNA synthesized in vitro from the same template as described above. The result of this experiment (Fig. 4C) indicates that RNA in this region is transcribed from right to left in vitro and that under these conditions no *Hin*fI fragments within the 1.7-kb *Taq*I fragment contain functional promoters.

Determination of the 5′ start site of right early transcription. Primer extension was performed with RNA isolated from infected and uninfected *E. coli*, as well as with RNA transcribed in vitro. Two synthetic oligonucleotides were used; one was complementary to a region predicted to be about 100 nucleotides from the 5′ end of any transcripts originating from ORF B4, and the other was complementary to any transcripts that might originate from ORF A6. These were hybridized to RNA isolated from infected host cells or with RNA synthesized in vitro and were elongated with avian myeloblastosis virus reverse transcriptase in the presence of [α-^{35}S]dATP.

Figure 5 shows the results of these experiments. Three bands which differ in size by one nucleotide can be resolved in both lanes 7 and 8. These are the extension products synthesized from transcription of ORF B4 in vivo and in vitro, respectively. Attempts to resolve an elongation product from any RNAs arising from ORF A6 were unsuccessful (Fig. 5, lanes 1 and 2). No bands can be seen in lanes 3 through 6. These are controls in which RNA from uninfected cells was used in the primer extension reaction or in which RNA polymerase was left out of the in vitro transcription reaction prior to primer extension. Transcription of ORF B4 starts with equal efficiency at nucleotides 179 and 180; it occurs with lower efficiency at nucleotide 181. Nucleotides 179 through 181 are appropriately positioned downstream from a −10 and −35 promoter consensus sequence. Interestingly, this promoter region is overlapped by a second putative promoter, which is not used under these conditions (Fig. 1).

Identification of two transcripts of different sizes originating from the right early promoter. Figure 6 is a photograph of ethidium bromide-stained denaturing gels on which RNA

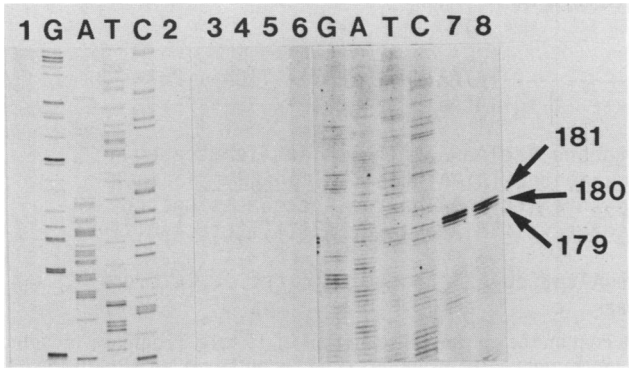

FIG. 5. Primer extension of PRD1 right early transcripts synthesized in vitro and in vivo. Lane 1 and 2 show unsuccessful attempts to primer extend any transcripts that might originate in vivo or in vitro from ORF A6 (see Fig. 2). Lanes 3 and 4 are the same as lanes 1 and 2, except that RNA polymerase was left out of the in vitro transcription reaction or RNA was isolated from uninfected host. Lanes 5 and 6 contain similar controls for primer extension of transcripts that originate from ORF B4. Lane 7 contains primer extension products of RNA isolated from an infected host originating from ORF B4. Lane 8 contains primer extension products of gene XII transcripts synthesized in vitro.

synthesized in vitro was electrophoresed. All of these bands were susceptible to degradation by RNase A and unaffected by DNase I (data not shown). RNA synthesis was performed with total genomic PRD1 DNA, which had been restricted with the indicated enzyme, as the template. A restriction map of both the left and right early regions is shown in Fig. 2. Digestion of the template DNA with *Bst*XI or *Bst*NI would not be expected to affect the sizes of any of the transcripts synthesized here, since both enzymes cleave PRD1 DNA only within regions containing late genes (*Bst*NI cleaves PRD 1 DNA only once approximately 4.8 kb from the left side). *Dde*I and *Mbo*II would be expected to result in the transcription of truncated left early RNA and full-length right early RNA. Template digested with *Taq*I would also be expected to generate truncated left early transcripts, but the effect on right early transcription would be expected to be much more dramatic because *Taq*I cuts the right early region

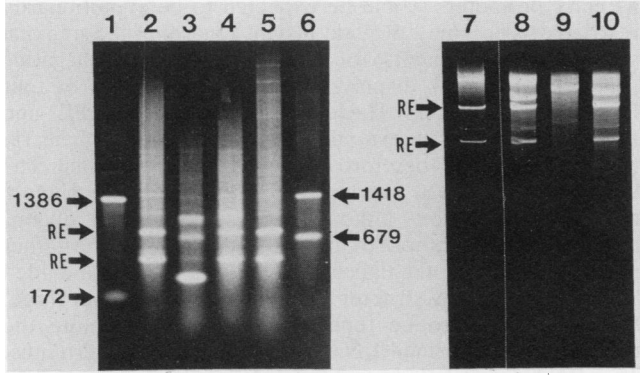

FIG. 6. In vitro PRD1 transcripts. Transcripts are resolved on an agarose-formaldehyde gel (lanes 1 to 6) or on a polyacrylamide-urea gel (lanes 7 to 10). Lanes: 1 and 6, RNA molecular mass standards; 2 and 7, transcripts generated from uncut PRD1 genomic DNA. The remaining lanes contain transcripts synthesized from PRD1 DNA which had been restricted with the following enzymes: 3, *Ban*I; 4, *Bst*XI; 5, *Bst*NI; 8, *Dde*I; 9, *Taq*I; 10, *Mbo*II. Right early transcripts are indicated (RE).

immediately downstream from the putative early promoter, between nucleotides 183 and 184 (Fig. 1 and 2).

Two bands in Fig. 6 were unaffected when the template DNA was restricted with *Dde*I or *Mbo*II. They disappeared completely, however, when the template was digested with *Taq*I. Template which is digested with *Ban*I would also be expected to produce truncated left early RNAs, but these would be fairly large and would not migrate very far into the gel. Transcripts synthesized from the right early promoter would be expected to be truncated and to be approximately 300 nucleotides in length. The same two bands which were eliminated by digestion of the template with *Taq*I also disappeared after digestion of the template with *Ban*I (Fig. 6, lane 3). A new, intense, lower-molecular-weight band appeared in this lane, and this leads us to believe that two RNAs are transcribed from the same promoter and terminate at different points downstream from the *Ban*I site, possibly at the two putative terminators identified previously (Fig. 2 and 3). The estimated lengths of the two *Taq*I-sensitive bands in Fig. 6 are approximately 800 and 500 nucleotides. The estimated length of the transcript generated from the *Ban*I-digested template is approximately 280 nucleotides.

DISCUSSION

We have sequenced the entire right early region of the PRD1 genome and, in so doing, have completed the sequence of all early regions. The data presented here demonstrate that right early transcription initiates at nucleotides 179 and 180 with equal efficiency and elongates from right to left in vivo and in vitro. Initiation also occurs with lower efficiency at nucleotide 181. This, in turn, has enabled us to assign gene XII to ORF B4 and to predict its primary structure (Fig. 1). It also indicates that ORF B2 (possibly gene XIX) is expressed. In vitro transcription experiments indicate that the right early promoter directs the synthesis of two transcripts which terminate at different points downstream from the rightmost *Ban*I site. This suggests that the two putative terminators, identified here, may be used in vitro. In addition, we now have the sequences of two proven and two putative early PRD1 promoters to analyze. This is of interest because of the wide host range of this phage.

A number of other ORFs within the right 1,700 bp of the PRD1 genome are likely to encode proteins. ORF B2, which could encode a peptide of 5,616 Da, is preceded by a notably plausible RBS and followed by a very credible terminator sequence. The molecular mass of its product, as well as its physical position on the PRD1 genome, makes it a likely candidate for gene XIX, a middle early gene of unknown function. Its apparent (determined by SDS-polyacrylamide gel electrophoresis) molecular mass has been reported to be less than 10 kDa. Sequence data obtained from the left side of the genome suggest that gene XV, also a middle early gene, may be transcribed in the same direction as genes VIII and I (36), so it would not be surprising if gene XIX was likewise transcribed with gene XII. If ORF B2 is, in fact, gene XIX, the mode by which it is down regulated, late in infection, by p12 may prove interesting. Mindich et al. (33) have suggested that the temporal regulation of PRD1 gene expression may be achieved by an antitermination mechanism.

Codon bias plots and Fickett's analysis indicate that ORF B3 also encodes a protein (data not shown), although it lacks a plausible RBS. Since it overlaps ORF B4, its expression could be achieved by ribosomal slippage or frame shifting. The size of any protein arising from this ORF (6,987 Da from

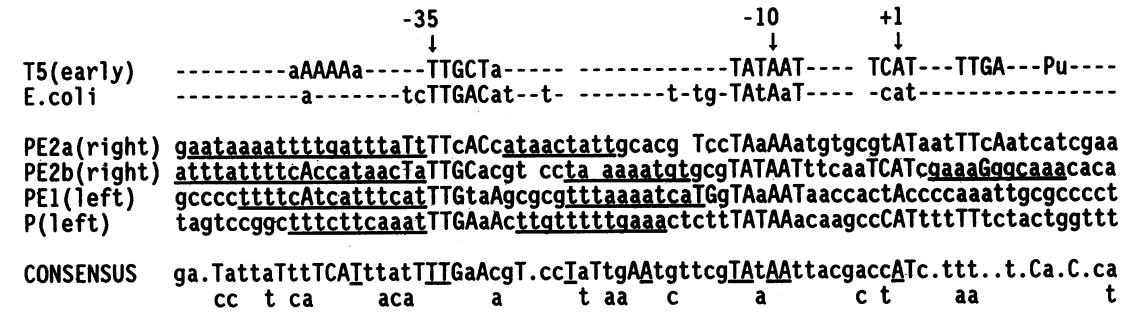

FIG. 7. Multiple alignment of PRD1 early promoters. The top two lines contain the consensus sequences of T5 early promoters (adapted from reference 4) and *E. coli* promoters (20). PE2a and PE2b are the overlapping right early promoters (PE2a is putative), and PE1 is the left early promoter. P(left) is a putative promoter which is predicted to generate a transcript which initiates 263 bp from the left end and elongates from right to left toward the left terminus. Letters are capitalized when they match one of the two consensus sequences at the top. The A+T-rich segments preceding the −35 regions are underlined, as are the A+T-rich patches located between the −35 and −10 regions. A purine-rich segment that follows the +1 region of PE2b is also underlined. A consensus sequence for the four PRD1 promoters is located at the bottom of the figure. In cases when all four nucleotides match, the nucleotide within the consensus sequence is capitalized and underlined. When three of four match the nucleotide is capitalized but not underlined. When two of four match it is displayed in lower case, and in instances when there are two possible matches, both nucleotides are displayed in lower case.

the first Met) would also be consistent with the size of gene XIX.

ORFs A1, A2, and/or A3 may encode proteins that are expressed late in infection. Should any of these ORFs encode real proteins, those arising from ORFs A1 and A3 are predicted by the method of Rao and Argos (38) to contain a membrane-bounded helix near their N terminus. The method of Eisenberg et al. (9, 10) predicts that a protein originating from ORF A3 would contain an N-terminal membrane-associated helix that could be tentatively classified as multimeric and membrane spanning, and both ORFs A1 and A2 are predicted to be integral membrane proteins by the method of Klein et al. (28) (these predictions were aided by programs supplied with the PC gene [IntelliGenetics]). Since PRD1 is a membrane-bound phage, these predictions imply that the hypothetical products of ORFs A1, A2, and A3 are structural and hence expressed late in infection. If so, this would suggest that late transcription, in this region occurs from left to right. In support of this belief, McGraw et al. (31) have observed that a certain mutant of gene III results in the loss of synthesis of proteins arising from genes XXII, XVIII, XI, and XVI (Fig. 2) and have suggested that a large transcript starting at gene III and extending to the right may be synthesized.

Extensive searches of the National Biomedical Research Foundation (NBRF) data base and GenBank have revealed no significant similarities between gene XII and any other protein. However, when the NBRF data base is searched by using the University of Wisconsin Genetics Computer Group WORDSEARCH program, a large number of histones are pulled out. The COOH terminus of p12 is rich in alanine and lysine, and it shares this characteristic with regions found in many histones. Like the histones, p12 might be a DNA-binding protein and the COOH terminus might play a role in this activity. There is also another patch of basic amino acids, Arg-Val-Arg-Gly-Lys-Lys, located at residues 33 through 38, which have the potential to interact electrostatically with DNA.

Another notable characteristic of gene XII is that a liberal number of aromatic amino acids are situated throughout the first 110 residues. This could be meaningful in that aromatic residues in conjunction with basic amino acids have been implicated in the interactions between DNA and single-stranded DNA-binding proteins (5). The possibility that gene XII encodes a single-stranded DNA-binding protein is attractive because of analogies which can be drawn between PRD1 and adenovirus. They have inverted terminal repeats of similar length, and in both cases only three virus-specific replication proteins have been discerned (for reviews, see references 22 and 26). In adenovirus these genes encode DNA polymerase, preterminal protein, and a single-stranded DNA-binding protein. Adenovirus single-stranded DNA-binding protein is similar to p12 in that it is also involved in the regulation of gene expression.

The direction in which gene XII is transcribed may have evolutionary connotations. The right early genes of both PRD1 and ϕ29 are transcribed in the same direction. Therefore, the only major organizational difference between these two phage genomes lies in the relative positions of genes specifying DNA polymerase and terminal protein and the direction in which they are transcribed. If one conjectures that the genomes of both ϕ29 and PRD1 evolved from genetic elements that had similar organizations, one might further speculate that the left side of one of their genomes underwent an inversion at some point in time.

Because of the wide host range of PRD1, the structure of its early promoters would be expected to be evolutionarily restricted. They may provide insight regarding the minimal structural requirements for a "universal" gram-negative promoter. Figure 7 displays a multiple alignment of four PRD1 early promoters. The left and right promoters PE1 and PE2b precede the known start sites of left and right early transcripts and are therefore presumed to be functional. The right promoter PE2a overlaps PE2b and probably plays a role in transcription, although DNase protection experiments would be required to prove this conclusively. Another putative left promoter [labeled P(left) in Fig. 7] is included because it has many structural similarities to the other three. Should it prove to be functional, it would promote the transcription of a small RNA out toward the left terminus. This RNA is predicted to initiate at an adenosine which is 263 bp away from the left end. Such an RNA would not be surprising, since ϕ29 transcribes an analogous RNA which has been shown to be required for packaging phage DNA into viral particles (18, 19). The bacteriophage Cp-1 and related phages also contain putative promoters at each end that direct transcription toward the termini which have been suggested to aid replication by helping to melt the DNA at

the termini (11, 12). By analogy to ϕ29, however, they are probably involved in the process of DNA packaging as well.

The fact that PE2a and P(left) bear such structural similarity to PE1 and PE2b as well as to each other leaves us with the conviction that they are functional. All four promoters contain plausible −35 and −10 regions which are separated by 16 to 18 bp. The nucleotides which are most highly conserved in the −35 and −10 regions of *E. coli* are also the most highly conserved in these promoters (the first three nucleotides of the −35 region, TTGACA, and all but the third nucleotide of the −10 region, TATAAT). All contain a +1 region which conforms to the *E. coli* consensus sequence. Most frequently the +1 region consists of CAT, where A is the initiating nucleotide. The sequence TAT is also fairly common, although it occurs less frequently than CAT. The third T is the least highly conserved (20). The demonstrated and predicted initiation sites are between 6 and 7 bp away from the −10 region. Like many *E. coli* promoters, all four of the promoters contain an A+T-rich segment upstream from the −35 region. These patches are 95, 80, 80, and 82 A+T rich (Fig. 7, top to bottom). The region between their −35 and −10 regions also contain patches which are 89, 89, 91, and 83% A+T rich. An interesting consensus can be found in the −47 to −40 regions of PRD1 early promoters, which is not immediately apparent in surveys of *E. coli* an *S. typhimurium* promoters. Either this consensus, TnnTCAT, is important for promoter function in hosts other than *E. coli* and *S. typhimurium*, or it is involved in the down regulation, late in infection, of early-gene expression by the product of gene XII.

The sequence of the promoter PE2b and its interaction with PE2a is of particular interest because it precedes gene XII, which is very highly expressed. Promoter PE2b exhibits some similarities to T5 early promoters (4, 15). Its −35 region, TTGCAC, is closer to the T5 consensus sequence, TTGCTa, than it is to that of *E. coli*, and its +1 region matches the T5 consensus exactly. Although it lacks the TTGA commonly found at positions +6 to +9, its initiation site, like that of T5, is followed by a purine-rich segment (Fig. 7). Primer extension of the right early transcript indicates that promoter PE2a is not used in the normal sense of the word; however, the fact that its predicted site of initiation falls within the −10 region of PE2b suggests that it might increase the efficiency of this promoter by funneling RNA polymerase into it.

The sequence data presented here augment previous studies by extending the known nucleotide sequence of the PRD1 genome to include all early and, possibly, middle early regions. This should facilitate the analysis of all PRD1-specific replication proteins as well as studies concerning the transcriptional organization of this phage. Although the predicted primary structure of p12 has not enabled us to infer a specific function, it is expected to aid future studies toward this end. Since transcription performed in vitro appears to faithfully reproduce in vivo events, we expect that in vitro transcription systems will provide a convenient and powerful method for dissecting the modes by which the temporal regulation of PRD1 gene expression is achieved. This is particularly so in light of the commercial availability of *E. coli* RNA polymerase which is sufficiently saturated with sigma factor. Knowledge of the transcriptional organization of the PRD1 genome may well lead to a better understanding of the evolution of linear DNAs which replicate via a protein-priming mechanism. Because of the wide host range of PRD1, analysis of its promoters may also lead to a better understanding of the essential requirements for interactions between RNA polymerase and DNA and, possibly, the evolutionary origins of RNA polymerases in general.

ACKNOWLEDGMENTS

We are extremely grateful to Leonard Mindich, Public Health Research Institute of the City of New York, for supplying phage and strains.

This investigation was supported by Public Health Service grant GM28013 from the National Institutes of Health.

LITERATURE CITED

1. **Aiba, H., S. Adhya, and B. de Crombrugghe.** 1981. Evidence for two functional gal promotors in intact *Escherichia coli* cells. J. Biol. Chem. **256**:11905–11910.
2. **Bamford, D. H., and L. Mindich.** 1984. Characterization of the DNA-protein complex at the termini of the bacteriophage PRD1 genome. J. Virol. **50**:309–315.
3. **Bradley, D. E., and E. L. Rutherford.** 1975. Basic characterization of a lipid containing bacteriophage specific for plasmids of the P, N and W compatibility groups. Can. J. Microbiol. **21**:152–163.
4. **Bujard, H., M. Brunner, U. Deuschle, W. Kammerer, and R. Knaus.** 1987. Structure-function relationship of *Escherichia coli* promoters, p. 95–103. *In* W. S. Reznikoff, R. R. Burgess, J. E. Dahlberg, C. A. Gross, M. Record, Jr., and M. P. Wickens (ed.), RNA polymerase and the regulation of transcription. Elsevier Science Publishing, Inc., New York.
5. **Chase, J. W., and K. R. Williams.** 1986. Single-stranded DNA binding proteins required for DNA replication. Annu. Rev. Biochem. **55**:103–136.
6. **Chomczynski, P., and P. K. Qasba.** 1984. Alkaline transfer of DNA to plastic membrane. Biochem. Biophys. Res. Commun. **98**:340–344.
7. **Davis, T. N., E. D. Muller, and J. E. Cronan, Jr.** 1982. The virion of the lipid-containing bacteriophage PR4. Virology **120**: 287–306.
8. **Davison, B. L., T. Leighton, and J. C. Rabinowitz.** 1980. Specificity of promotor site utilization *in vitro* by bacterial RNA polymerases on Bacillus phage ϕ29. J. Biol. Chem. **255**:8319–8330.
9. **Eisenberg, D., R. M. Weiss, and T. C. Terwilliger.** 1984. The helical hydrophobic moment: a measure of the amphiphilicity of a helix. Nature (London) **299**:371–374.
10. **Eisenberg, D., E. Schwarz, M. Komaromy, and R. Wall.** 1982. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. **179**:125–142.
11. **Escarmís, C., P. García, E. Méndez, R. López, M. Salas, and E. García.** 1985. Inverted terminal repeats and terminal proteins of the genomes of pneumococcal phages. Gene **36**:341–348.
12. **Escarmís, C., A. Gómez, E. García, C. Ronda, R. López, and M. Salas.** 1984. Nucleotide sequence at the termini of the DNA of *Streptococcus pneumoniae* phage Cp-1. Virology **133**:166–177.
13. **Fickett, J. W.** 1982. Recognition of protein coding regions in DNA sequences. Nucleic Acids Res. **10**:5303–5318.
14. **Freier, S. M., R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner.** 1986. Improved free-energy parameters for prediction of RNA duplex stability. Proc. Natl. Acad. Sci. USA **83**:9373–9377.
15. **Gentz, R., and J. Bujard.** 1985. Promoters recognized by *Escherichia coli* RNA polymerase selected by function: highly efficient promoters from bacteriophage T5. J. Bacteriol. **164**: 70–77.
16. **Gerendasy, D., and J. Ito.** 1987. The genome of the lipid-containing bacteriophage, which infects gram-negative bacteria, contains long, inverted terminal repeats. J. Virol. **61**:594–596.
17. **Gribskov, M., J. Devereux, and R. R. Burgess.** 1984. The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucleic Acids Res. **12**: 539–549.
18. **Guo, P., S. Bailey, J. W. Bodley, and D. Anderson.** 1987. Characterization of the small RNA of the bacteriophage ϕ29 DNA packaging machine. Nucleic Acids Res. **15**:7081–7090.

19. **Guo, P., S. Erickson, and D. Anderson.** 1987. A small viral RNA is required for *in vitro* packaging of bacteriophage ϕ29 DNA. Science **236**:690–694.
20. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
21. **Hsieh, J., G. Jung, M. C. Leavitt, and J. Ito.** 1987. Primary structure of the DNA terminal protein of bacteriophage PRD1. Nucleic Acids Res. **15**:8999–9009.
22. **Hurwitz, J., S. Adha, J. Field, R. Gronostajski, R. A. Guggenheimer, M. Kenny, J. Lindenbaum, and K. Nagata.** 1985. Synthesis of adenoviral DNA with purified proteins, p. 15–24. *In* Paul A. Marks (ed.), Genetics, cell differentiation, and cancer. Academic Press, Inc., Orlando, Fla.
23. **Jeffreys, A. J., and R. A. Flavell.** 1977. A physical map of the DNA regions flanking the rabbit β-globin gene. Cell **12**:429–439.
24. **Jung, G., M. C. Leavitt, J. C. Hsieh, and J. Ito.** 1987. Bacteriophage PRD1 DNA polymerase: evolution of DNA polymerases. Proc. Natl. Acad. Sci. USA **84**:8287–8291.
25. **Kawamura, F., and J. Ito.** 1977. Transcription of the genome of bacteriophage ϕ29: isolation and mapping of the major early mRNA synthesized in vivo and in vitro. J. Virol. **23**:562–577.
26. **Kelly, T. J.** 1984. Adenovirus DNA replication, p. 271–398. *In* H. S. Ginsberg (ed.), The adenoviruses. Plenum Publishing Corp., New York.
27. **Keppel, F., O. Fayet, and C. Georgopoulos.** 1988. Strategies of bacteriophage DNA replication, p. 145–262. *In* R. Calander (ed.), The bacteriophages, vol. 2. Plenum Publishing Corp., New York.
28. **Klein, P., M. Kanehisa, and C. DeLisi.** 1985. The detection and classification of membrane-spanning proteins. Biochim. Biophys. Acta **815**:468–476.
29. **Lundström, K. H., D. H. Bamford, E. T. Palva, and K. Lounatmaa.** 1979. Lipid-containing bacteriophage PR4: structure and life cycle. J. Gen. Virol. **43**:583–592.
30. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
31. **McGraw, T., H. L. Yang, and L. Mindich.** 1983. Establishment of a physical and genetic map for bacteriophage PRD1. Mol. Gen. Genet. **190**:237–244.
32. **Mellado, R. P., F. Moreno, E. Viñuela, M. Salas, B. E. Reilly, and D. L. Anderson.** 1976. Genetic analysis of bacteriophage ϕ29 of *Bacillus subtilis*: integration and mapping of reference mutants of two collections. J. Virol. **19**:495–500.
33. **Mindich, L., D. Bamford, C. Goldthwaite, M. Laverty, and G. MacKenzie.** 1982. Isolation of nonsense mutants of lipid-containing bacteriophage PRD1. J. Virol. **44**:1013–1020.
34. **Mindich, L., and D. H. Bamford.** 1988. Lipid-containing bacteriophages, p. 145–262. *In* R. Calander (ed.), The bacteriophages, vol. 2. Plenum Publishing Corp., New York.
35. **Olsen, R. H., J. S. Siak, and R. H. Gray.** 1974. Characteristics of PRD1, a plasmid-dependent broad-host-range DNA bacteriophage. J. Virol. **14**:689–699.
36. **Pakula, T. M., H. Savilahti, and D. H. Bamford.** 1989. Comparison of the amino acid sequence of the lytic enzyme from broad-host-range bacteriophage PRD1 with sequences of other cell-wall-peptidoglycan lytic enzymes. Eur. J. Biochem. **180**: 149–152.
37. **Platt, T.** 1986. Transcription termination and regulation of gene expression. Annu. Rev. Biochem. **55**:339–372.
38. **Rao, M. J. K., and P. Argos.** 1986. A conformational preference parameter to predict helices in integral membrane proteins. Biochim. Biophys. Acta **869**:197–214.
39. **Reilly, B. E., R. A. Nelson, and D. L. Anderson.** 1977. Morphogenesis of bacteriophage ϕ29 of *Bacillus subtilis*: mapping and functional analysis of the head fiber gene. J. Virol. **29**:367–377.
40. **Salas, M.** 1988. Phages with protein attached to the DNA ends, p. 169–191. *In* R. Calander (ed.), The bacteriophages, vol. 1. Plenum Publishing Corp., New York.
41. **Savilahti, H., and D. H. Bamford.** 1986. Linear DNA replication: inverted terminal repeats of five closely related *Escherichia coli* bacteriophages. Gene **49**:199–205.
42. **Savilahti, H., and D. H. Bamford.** 1987. The complete nucleotide sequence of the left early region of *Escherichia coli* bacteriophage PRD1 coding for terminal protein and DNA polymerase. Gene **57**:121–130.
43. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
44. **Stormo, G. D., T. D. Schneider, L. Gold, and A. Ehrenfeucht.** 1982. Use of the 'perceptron' algorithm to distinguish translational initiation sites in *Escherichia coli*. Nucleic Acids Res. **10**:2997–3011.
45. **Yoo, S., and J. Ito.** 1989. Protein-primed replication of bacteriophage PRD1 genome *in vitro*. Virology **170**:442–449.