Proc. Natl. Acad. Sci. USA
Vol. 78, No. 3, pp. 1336–1340, March 1981
Biochemistry

Nucleotide sequences at the termini of ϕ29 DNA

(terminal-inverted repetition/linear DNA replication/early promoter sequences)

HIROFUMI YOSHIKAWA*, THEODORE FRIEDMANN†, AND JUNETSU ITO*

*Department of Molecular and Medical Microbiology, College of Medicine, The University of Arizona, Tucson, Arizona 85724; and †The Department of Pediatrics, School of Medicine, University of California at San Diego, La Jolla, California 92093

Communicated by E. Peter Geiduschek, September 12, 1980

ABSTRACT **The nucleotide sequences of the first 422 base pairs from the left-hand end and the first 274 base pairs from the right-hand end of ϕ29 DNA were determined by using the chemical degradation method of Maxam and Gilbert. The data indicate that ϕ29 DNA has inverted terminal repetitions that are six base pairs long 5′ (-A-A-A-G-T-A-). No perfectly self-complementary sequence exists within the terminal regions of ϕ29 DNA, suggesting that DNA replication via a self-priming mechanism is improbable. The putative early promoter sequences were found in both ends of the ϕ29 DNA. The results of the sequence determination are discussed in relationship to models proposed for the mechanism of replication of linear DNA molecules.**

The genome of bacteriophage ϕ29 is a linear nonpermuted duplex DNA that is ≈18 kilobases long (1, 2) and has a protein covalently linked to each 5′ terminus (3–6). The ϕ29 DNA-terminal protein is encoded by gene 3 (*gP3*) and is essential for DNA replication (4, 7–9). Recent biochemical and electron microscopic analysis of ϕ29 replicative intermediates showed that replication is initiated at either end of the DNA and proceeds in the 5′ to 3′ direction by a strand-displacement mechanism (10–12). Similar results have been reported for replication of adenovirus DNA, which also contains terminal proteins (13–15). However, the precise mechanism by which the 5′ ends of the linear DNA replicate is not understood. The adenovirus DNA contains terminally inverted repeated sequences more than 100 nucleotides long (16–19), and this long inverted repetition is considered to play an important role in adenovirus DNA replication (14, 20). To examine whether or not some special nucleotide sequence arrangements exist at the termini of ϕ29 DNA, we have carried out base sequence analysis of its terminal fragments. Our results show that ϕ29 DNA contains a short inverted terminal repetition, six base-pairs long, rather than a long one as found in adenovirus DNA. Results complementary to our data have been obtained by Escarmis and Salas (personal communication).

MATERIALS AND METHODS

ϕ29 DNA and ϕ29 DNA-protein complex were prepared from purified phage particles as described (2, 6). Restriction endonuclease *Eco*RI was purified according to the method of Greene *et al.* (21). Restriction endonucleases *Hin*dIII, *Hha* I, *Hin*fI, *Sau*3A, and *Taq* I were purchased from Bethesda Research Laboratories, Rockville, MD. *Acc* I, *Bcl* I, *Bst*NI, *Hgi*AI, *Mbo* II, *Mnl* I, and *Rsa* I were obtained from New England BioLabs. Bacterial alkaline phosphatase was obtained from Millipore (Freehold, NJ), *Escherichia coli* DNA polymerase I (large fragment) was from New England BioLabs, T4 polynucleotide kinase was from P-L Biochemicals, and terminal deoxynucleotidyl transferase came from Bethesda Research Laboratories. [γ-^{32}P]ATP, [α-^{32}P]dNTPs, and 3′-[α-^{32}P]dATP ([α-^{32}P]cordycepin ATP) were obtained from New England Nuclear.

DNA Fragment Preparation. *Bst*NI cleaves ϕ29 DNA only once, and *Eco*RI cuts ϕ29 DNA into five fragments (see Fig. 2). For DNA sequencing of the left and right ends of the ϕ29 DNA, *Bst*NI B and *Eco*RI C fragments were isolated by sucrose gradient centrifugation as described (2). The purified *Bst*NI B or *Eco*RI C fragments (5–10 pmol) were then digested by a second restriction endonuclease: the *Bst*NI B fragment by *Bcl* I or *Hin*fI and the *Eco*RI C fragment by *Hin*dIII, *Hin*fI, or *Taq* I. The reaction mixtures for the enzymes *Bcl* I and *Taq* I contained 10 mM Tris·HCl (pH 7.4), 10 mM $MgCl_2$, 6 mM KCl, 6 mM 2-mercaptoethanol, and bovine serum albumin at 100 μg/ml. For the enzymes *Hin*dIII and *Hin*fI, the reaction mixtures contained 10 mM Tris·HCl (pH 7.4), 10 mM $MgCl_2$, and 50 mM NaCl. All reactions were for 60 min at 37°C except for *Taq* I, which was used at 60°C.

5′ End Labeling. The 5′ ends of the dephosphorylated DNA fragments were labeled by using [γ-^{32}P]ATP and T4 polynucleotide kinase according to Maxam and Gilbert (22). For labeling native 5′ ends of ϕ29 DNA, the terminal DNA fragments were treated with 0.1 M NaOH at 37°C for 90 min; the DNA sample was then neutralized with HCl, incubated at 65°C for 5 min, and cooled slowly to room temperature. Finally, the unblocked DNA was treated with alkaline phosphatase and labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase.

3′ End Labeling. The 3′ ends of the restriction enzyme fragments were labeled with ^{32}P by using [α-^{32}P]dNTP and DNA polymerase I (Klenow fragment). The reaction mixture contained 100 mM Tris·HCl (pH 7.4), 80 mM NaCl, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol, [α-^{32}P]dNTP at 500–600 Ci/mmol (1 Ci = 3.7×10^{10} becquerels), and *E. coli* DNA polymerase I. The reaction was carried out at 4°C for 15 hr and terminated by phenol extraction.

The native 3′ ends of ϕ29 DNA were labeled by using terminal deoxynucleotidyl transferase and [α-^{32}P]cordycepin triphosphate according to Tu and Cohen (23).

DNA Sequence Analysis. DNA sequence determinations were performed by the method of Maxam and Gilbert (22). Gels containing 8–25% polyacrylamide and 7 M urea were prepared as described by Sanger and Coulson (24).

RESULTS

Evidence for Blunt End of ϕ29 DNA. To examine the possibility that ϕ29 DNA contains single-stranded 5′ ends (25–27), we performed "fill-in" experiments in which four [α-^{32}P]dNTPs and *E. coli* DNA polymerase I were used. When whole ϕ29 DNA was used as a template, no incorporation of radioactivity was observed (Fig. 1); in contrast, the *Eco*RI and *Hin*dIII digests of ϕ29 DNA incorporated radioactivity. It is noticeable,

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

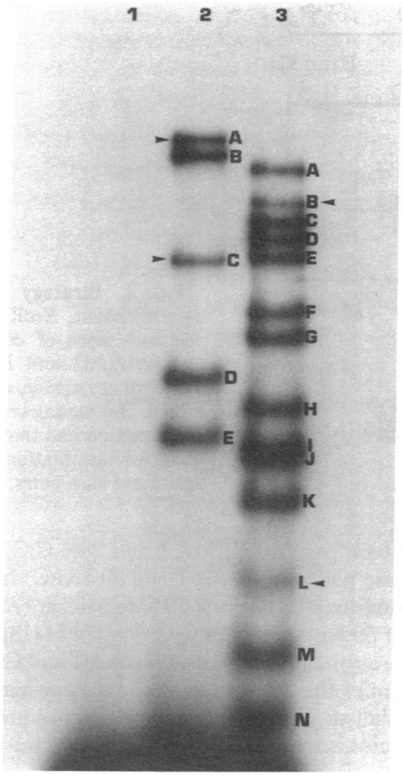

FIG. 1. Fill-in experiment of ϕ29 DNA. DNA or DNA restriction fragments ≈10 pmol were labeled with 20 μCi of [α-^{32}P]dNTPs by DNA polymerase I, and portions of these samples were subjected to electrophoresis in 1.2% agarose gel and autoradiographed. Templates are ϕ29 whole DNA (lane 1), ϕ29 DNA *Eco*RI digests (lane 2), and ϕ29 *Hin*dIII digests (lane 3). Arrows indicate terminal fragments.

however, that the terminal DNA fragments—the *Eco*RI A and C fragments and the *Hin*dIII B and L fragments—incorporated considerably less radioactivity than the internal fragments. Each of the DNA bands was then cut out from the agarose gel and assayed for radioactivity; the radioactivity of the terminal DNA fragments was about 50% that of the internal fragments (Table 1). Because ϕ29 DNA contains unblocked 3′-OH at the termini (6), these results suggest that ϕ29 DNA does not contain recessed 3′ ends.

DNA Sequence Analysis at Termini of ϕ29 DNA. Because the 5′ ends of the ϕ29 DNA are blocked by covalent linkage of a terminal protein, conventional 5′ end labeling cannot be performed. Therefore, we attempted to isolate small terminal DNA fragments suitable for DNA sequence analysis. The partial restriction enzyme cleavage maps and the strategies for determining the DNA sequences are summarized in Fig. 2.

Table 1. Incorporation of radioactivity into ϕ29 DNA *Eco*RI fragments

Fragment	Radioactivity, cpm	Activity,* %
A	2783	54
B	5134	100
C	2615	51
D	5833	114
E	5733	112

Each band of the labeled ϕ29 *Eco*RI digests was cut out from the gel. The gel slices were dissolved in 200 μl of 30% hydrogen peroxide and assayed in 20 ml of BBS-3 scintillation fluid.
* Relative to that of the *Eco*RI-B fragment.

Left-Hand Terminus. *Bcl* I cleaved the *Bst* NI B fragment once and generated two fragments (see Fig. 2). The *Bst*NI B/*Bcl* I subfragments were then labeled at their 5′ ends with kinase. The smaller *Bst*NI B/*Bcl* I fragment derived from the left-hand end of the ϕ29 DNA can be labeled at the 5′ ends of the *Bcl* I cleavage site only, because the native 5′ ends of ϕ29 DNA are inaccessible to phosphorylation. The larger *Bst*NI B/*Bcl* I fragment can be labeled at both 5′ ends. This fragment was cleaved with *Hgi*AI, and the labeled fragments were separated by electrophoresis and each was isolated from the gels and used for sequence analysis.

The smaller *Bst*NI B/*Bcl* I fragment was also labeled at the 3′ end by using DNA polymerase I and [α-^{32}P]dGTP and used for sequence analysis. To confirm the left-hand-end sequence, the *Bst*NI B/*Hin*fI digests were treated first with alkali (0.1 M NaOH) to unblock the native 5′ end and then with alkaline phosphatase and labeled with kinase. The native 3′ end of ϕ29 DNA was also labeled by using terminal transferase and [α-^{32}P]cordycepin ATP. The end-labeled DNA was cleaved with *Bcl* I, and the DNA sequence was determined from the native 5′ and 3′ ends, (Figs. 3*A* and 4*A*).

Right-Hand Terminus. The *Eco*RI C fragment was used to determine the sequence of the right-hand terminus of ϕ29 DNA (Fig. 2). *Hin*fI cuts the *Eco*RI C fragment at least four times (*Eco*RI C/*Hin*fI A to *Eco*RI C/*Hin*fI E). The terminal fragment, *Eco*RI C/*Hin*fI D, was labeled at both the 5′ and the 3′ ends by using kinase and by the fill-in method, respectively. To confirm the terminal sequence, the *Eco*RI C fragment was also cleaved with *Taq* I, and the sequence of the resulting terminal fragment was determined from the 3′ end of its restriction site. We also determined the sequence of the *Eco*RI C/*Hin*dIII fragment, which is the terminal fragment and is 274 base pairs long. The native 5′ and 3′ ends of the *Eco*RI C/*Hin*dIII fragment were labeled by unblocking, and treatment with kinase and terminal transferase. The resulting labeled fragments were then digested with *Hin*fI, and their sequences were determined from the native 5′ and 3′ ends (Figs. 3*B* and 4*B*).

The 422 base pairs of the left-hand terminus and the 274 base pairs of the right-hand terminus are shown in Fig. 5. The cleavage sites of the restriction enzymes were confirmed by digesting the end-labeled DNA fragments—*Mnl* I, *Sau*3A, and *Taq* I cut the left-hand DNA fragment, and *Acc* I, *Mnl* I, *Rsa* I, and *Taq* I cut the right-hand DNA fragment (see Fig. 5).

DISCUSSION

We have determined the nucleotide sequences of the termini of ϕ29 DNA for both strands by using 5′ end labeling by kinase and 3′ end labeling by terminal transferase or DNA polymerase I. The novel feature of the structure is that ϕ29 DNA has an inverted terminal repetition that is six base pairs long. Inverted terminal repetitions have previously been found in the chromosomes of animal viruses such as adenoviruses (16), adenoassociated virus (28), and vaccinia virus (29). This, however, is an example of an inverted terminal repeat in a bacteriophage genome. Recently, we have determined the sequences of the termini of other small *Bacillus* phage genomes and have found that all contain inverted terminal repetitions six to eight base pairs long (unpublished data). Known inverted repeated sequences range in size from 103 base pairs in adeno 5 DNA (18) to more than 10,000 base pairs in vaccinia virus DNA (29).

It is noticeable that the distribution of AT and GC base pairs in the terminal regions are not symmetric. ϕ29 DNA contains 64% AT and 36% GC (30). Our sequence data show that, at the

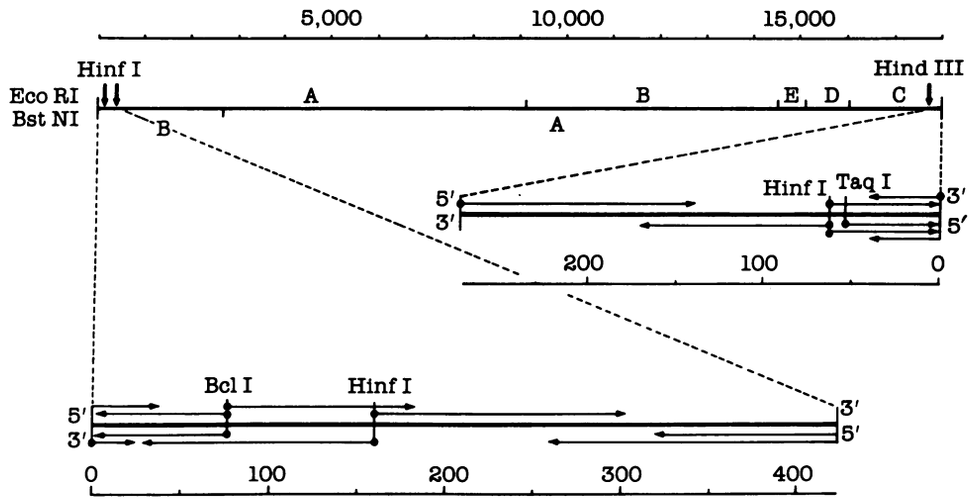

FIG. 2. Strategy for sequence determination. *Eco*RI and *Bst*NI cleavage maps of ϕ29 DNA, including *Hinf*I and *Hind*III fragments whose sequences were determined. Horizontal arrows indicate the direction and the extent of sequence determination. Numbers on scales are base pairs.

left-hand terminus, the first 50 base pairs are 58% A·T and 42% G·C, the next 50 base pairs are 52% A·T and 48% G·C, and the following 50 base pairs are 72% A·T and 28% G·C. At the right-hand terminus, the first 50 base pairs are 54% A·T and 46% G·C, the next 50 base pairs are 72% A·T and 28% G·C, and the following 50 base pairs are 70% A·T and 30% G·C. In the case of human adenovirus ad5, the first 50 base pairs are A·T rich (72%) and the next 50 base pairs are G·C rich (74%) (18). The significance of the asymmetric distribution of A·T and G·C base pairs at the termini of these viral DNAs is not known at present.

The mechanism of replication of linear nonredundant duplex DNA molecules, such as ϕ29 DNA and adenovirus DNA, is not

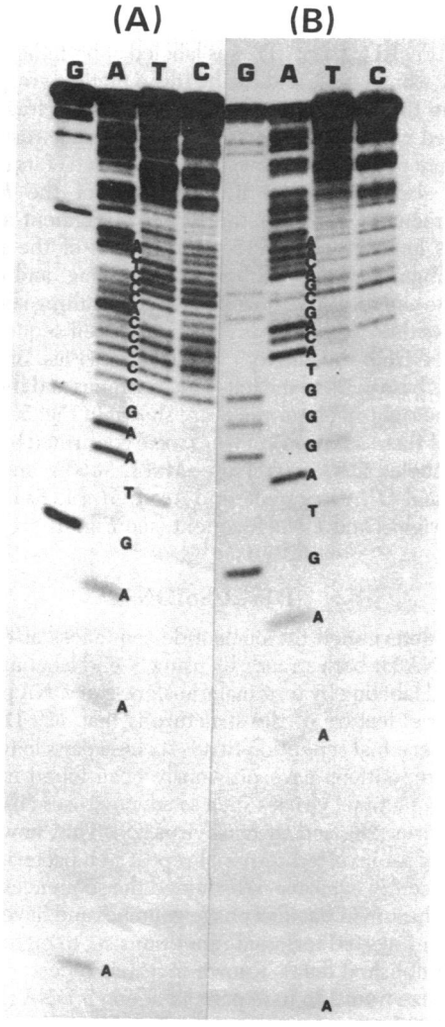

FIG. 3. Sequence determination gels of 5′-end-labeled left-hand (*A*) and right-hand (*B*) terminal ends of ϕ29 DNA. Residues 1 to 20 are indicated (20% gels).

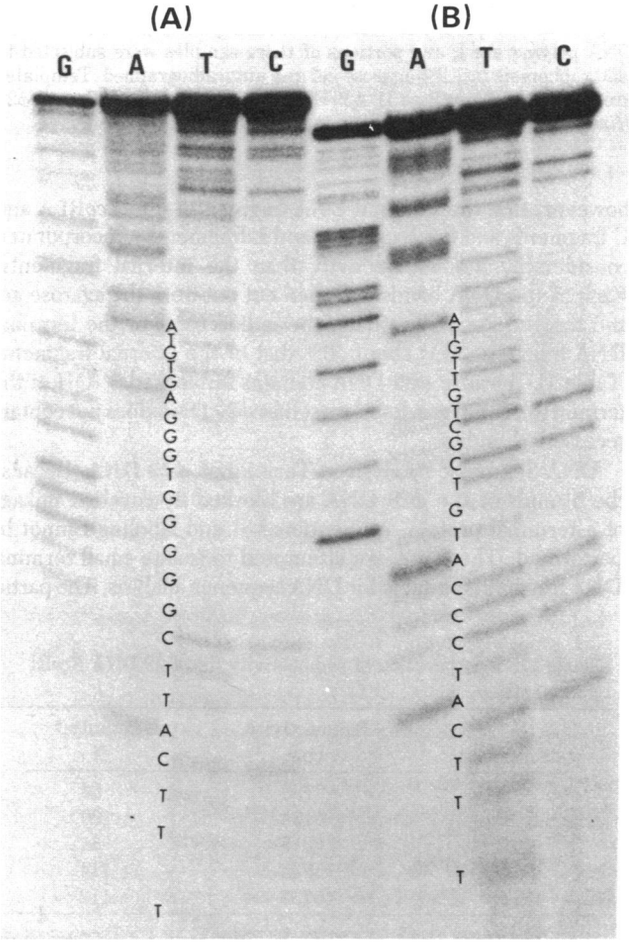

FIG. 4. Sequence-determination gels of 3′-end-labeled left-hand (*A*) and right-hand (*B*) terminal ends of ϕ29 DNA (20% gels).

FIG. 5. Terminal sequences of ϕ29 DNA. The inverted terminal repetition of six base pairs is common to both ends. Arrows indicate the direction of transcription.

currently understood but has recently been the subject of extensive investigations (14). Because all known DNA polymerases require a primer as well as a template and act in the 5′ to 3′ direction, it is not clear how the 5′ ends of linear DNA molecules are synthesized. The DNA sequence suggests that the two termini of ϕ29 DNA can neither cohere to each other to form circular intermediates in infected cells as λ phage (31) nor form concatemeric intermediates such as T7 phage (32). Indeed, no evidence for circular and concatemeric intermediates in ϕ29 DNA replication has been obtained (10–12).

Cavalier-Smith (33) has proposed a mechanism for the replication of the ends of linear chromosomal DNA molecules. According to his model, the base sequence at the termini of a linear DNA molecule is palindromic and can therefore fold back on itself to form a short duplex region so that the 3′ ends of one or both strands can provide a primer for synthesis of the complementary strand and thus generate a hairpin. Subsequently, the parental strand is nicked at the 5′ end of the palindromic sequence (which is now part of the progeny strand), and the 3′ end at the nick serves as a primer for DNA polymerase to fill out the sequences by using a daughter-strand template. This model is consistent with the structure of adenovirus-associated DNA (34, 35) and the autonomous parvovirus DNA (36). However, although ϕ29 DNA has a variety of repeat sequences having the same or opposite direction, there are no extensive self-complementary sequences that could make perfect hairpin loops. Therefore, the sequence data for ϕ29 DNA do not substantiate a mechanism for initiation of synthesis that involves the formation of hairpin loops. The nucleotide sequence analysis of adenovirus DNA also excludes the involvement of a hairpin loop structure for the initiation of replication (17–19). Lechner and Kelly (20) have pointed out that, during replication of adenovirus DNA by a strand-displacement mechanism, daughter duplex DNA and parental single-strand molecules are formed. The parental single-stranded DNA could hybridize the self-complementary terminal sequences to make a "panhandle" shaped intermediate. Then, the initiation of daughter-strand synthesis could occur at either end of the duplex DNA or the panhandle-shaped single-strand DNA by the same mechanism. The replicative intermediates of ϕ29 DNA are very similar to those of adenovirus DNA (10–12). However, because ϕ29 DNA does not contain a long inverted terminal repetition, it seems less likely that the "panhandle" is formed by parental single-strand DNA. It is noteworthy that 48% of the nucleotide sequence of the first 60 base pairs is identical between the two ends of ϕ29 DNA (Fig. 6). Whether these sequences can make single-stranded circles under physiological conditions remains to be determined.

Rekosh *et al.* (37) have proposed a model in which the terminal protein linked covalently to the 5′ end of the DNA serves as the primer. This model is supported by the fact that the newly synthesized DNA strand of adenovirus is associated with terminal protein (38, 39). It has been postulated (11, 12) that a DNA-terminal protein serves as the primer for the initiation of ϕ29 DNA replication; however, definite proof has not been obtained.

It has been shown that the early mRNAs are transcribed from both terminal regions of the ϕ29 genome (40–42). The nucleotide sequence determined in this study includes the probable promoters for ϕ29 early transcription. A sequence similar to the Pribnow box sequence, T-A-T-A-A-T-A, is located at positions 154–160 in the right-hand terminus. There is also a Shine-Dalgarno sequence, G-G-A-G-G, between the Pribnow box sequence and the initiation codon A-T-G. Thus, it is very likely

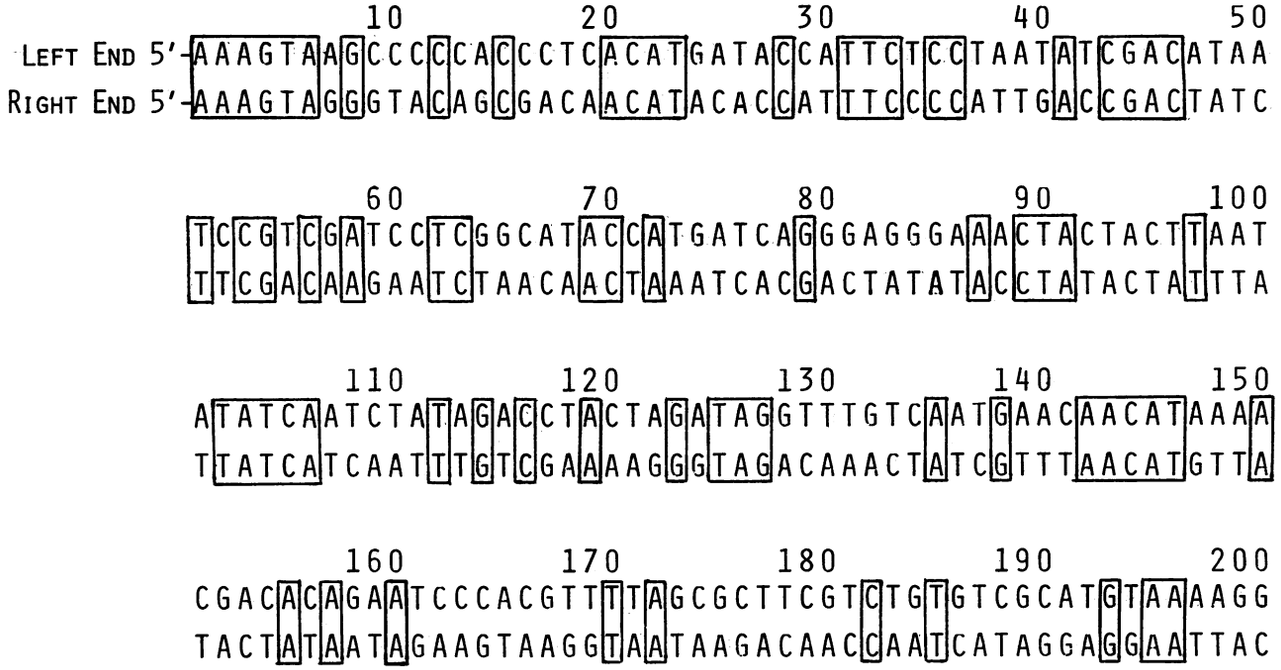

FIG. 6. Comparison of nucleotide sequences of left-hand and right-hand ends of ϕ29 DNA. Common nucleotide sequences are indicated in boxes. L and R mean left and right hand, respectively.

that this is the region transcribed and translated very early in ϕ29 infected cells. Similarly, there is a typical Pribnow box sequence, T-A-T-A-A-T-G, in the left-hand terminus at positions 333–327. Also, a typical −35 sequence, T-T-G-A-C-A, is present at positions 356–351. Recently, Davison *et al.* (43, 44) have shown that small mRNAs are transcribed from this region. It would be of interest to see whether such small RNAs are translated into polypeptides.

This research was supported by National Institutes of Health Grants GM28013 and GM28362 to J.I., who is the recipient of a faculty research award from the American Cancer Society (FRA-165).

1. Anderson, D. L., Hickman, D. D. & Reilly, B. E. (1966) *J. Bacteriol.* **91**, 2081–2089.
2. Ito, J., Kawamura, F. & Yanofsky, S. (1976) *Virology* **70**, 37–51.
3. Ortin, J., Vinuela, E., Salas, M. & Vasquez, C. (1971) *Nature (London) New Biol.* **234**, 275–277.
4. Salas, M., Mellado, R. P., Vinuela, E. & Sogo, J. M. (1978) *J. Mol. Biol.* **119**, 269–291.
5. Yehle, C. D. (1978) *J. Virol.* **27**, 776–783.
6. Ito, J. (1978) *J. Virol.* **28**, 895–904.
7. Harding, N. E., Ito, J. & David, G. S. (1978) *Virology* **84**, 279–292.
8. Yanofsky, S., Kawamura, F. & Ito, J. (1976) *Nature (London)* **259**, 60–63.
9. Mellado, R. P., Penalva, M. A., Inciarte, M. R. & Salas, M. (1980) *Virology* **104**, 84–96.
10. Ito, J., Harding, N. E. & Saigo, K. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **34**, 525–536.
11. Inciarte, M. R., Salas, M. & Sogo, J. M. (1980) *J. Virol.* **34**, 187–199.
12. Harding, N. E. & Ito, J. (1980) *Virology* **104**, 323–338.
13. Levine, A. J., Van der Vliet, P. C. & Sussenbach, J. S. (1976) *Curr. Top. Microbiol. Immunol.* **73**, 67–124.
14. Winnacker, E. L. (1978) *Cell* **14**, 761–773.
15. Robinson, A. J., Younghusband, H. B. & Bellett, A. J. D. (1973) *Virology* **56**, 54–69.
16. Wolfson, J. & Dressler, D. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 3054–3057.
17. Arrand, J. R. & Roberts, R. J. (1979) *J. Mol. Biol.* **128**, 577–594.
18. Steenbergh, P. H., Maat, J., Van Ormondt, H. & Sussenbach, J. S. (1977) *Nucleic Acids Res.* (1977) **4**, 4371–4389.
19. Shinagama, M. & Padmanabhan, R. (1979) *Biochem. Biophys. Res. Commun.* **87**, 671–678.
20. Lechner, R. L. & Kelly, T., Jr. (1977) *Cell* **12**, 1007–1020.
21. Greene, P. J., Betlach, M. C., Goodman, J. M. & Boyer, H. W. (1974) in *Methods in Molecular Biology*, ed. Weckner, R. (Dekker, New York), Vol. 9, pp. 87–111.
22. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–650.
23. Tu, C. & Cohen, S. N. (1980) *Gene* **10**, 177–183.
24. Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87**, 107–110.
25. Anderson, D. L. & Mosharrafa, E. T. (1968) *J. Virol.* **2**, 1185–1190.
26. Hirokawa, H. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1555–1559.
27. Thomas, C. A., Jr., Kelly, T. J., Jr. & Rhoades, M. (1968) *Cold Spring Harbor Symp. Quant. Biol.* **33**, 417–424.
28. Berns, K. I. & Kelly, T. J., Jr. (1974) *J. Mol. Biol.* **82**, 267–271.
29. Garon, C. F., Barbosa, E. & Moss, B. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4863–4867.
30. Rubio, V., Salas, M., Vinuela, E., Usobiaga, P., Saiz, J. L. & Llopis, J. F. (1974) *Virology* **57**, 112–121.
31. Young, E. & Sinsheimer, R. (1964) *J. Mol. Biol.* **10**, 562–564.
32. Kelly, T. J., Jr. & Thomas, C. A., Jr. (1969) *J. Mol. Biol.* **44**, 459–475.
33. Cavalier-Smith, T. (1974) *Nature (London)* **250**, 467–470.
34. Straus, S. E., Sebring, E. D. & Rose, J. A. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 742–746.
35. Lusby, E., Fife, K. H. & Berns, K. I. (1980) *J. Virol.* **34**, 402–409.
36. Bourguignon, G. J., Tattersall, P. J. & Ward, D. C. (1976) *J. Virol.* **20**, 290–306.
37. Rekosh, D. M. K., Russell, W. C., Bellett, A. J. D. & Robinson, A. J. (1977) *Cell* **11**, 283–295.
38. Stillman, B. W. & Bellett, A. J. D. (1979) *Virology* **93**, 69–79.
39. Coombs, D. H., Robinson, A. J., Bodnar, J. W., Jones, C. J. & Pearson, G. D. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **34**, 741–753.
40. Inciarte, M. R., Vinuela, E. & Salas, M. (1976) *Eur. J. Biochem.* **71**, 77–83.
41. Kawamura, F. & Ito, J. (1977) *J. Virol.* **23**, 562–577.
42. Sogo, J. M., Inciarte, M. R., Corral, J., Vinuela, E. & Salas, M. (1979) *J. Mol. Biol.* **127**, 411–436.
43. Davison, B. L., Leighton, T. & Rabinowitz, J. C. (1979) *J. Biol. Chem.* **254**, 9220–9226.
44. Davison, B. L., Murray, C. L. & Rabinowitz, J. C. (1980) *J. Biol. Chem.* **255**, 8819–8830.