JOURNAL OF VIROLOGY, May 1978, p. 429-434
0022-538X/78/0026-0429$02.00/0

Vol. 26, No. 2
Printed in U.S.A.

178-Nucleotide Sequence Surrounding the *cos* Site of Bacteriophage Lambda DNA

BRIAN P. NICHOLS AND JOHN E. DONELSON*
Department of Biochemistry, University of Iowa, Iowa City, Iowa 52242

Received for publication 18 October 1977

A nucleotide sequence of 61 nucleotides at the left end and 117 nucleotides at the right end of DNA from bacteriophage λ*c*I857*Sam*7 was determined by the Maxam and Gilbert method. A perfect inverted repeat sequence of 10 nucleotides is near the left end, and one of 15 nucleotides is near the right end. DNA from another closely related λ strain, λ*c*I857*prm*116*Sam*7, has about 10% divergence in the sequence of the first 110 nucleotides at the right end and has a 17-member perfect inverted repeat sequence.

The first DNA sequence of biological significance to be determined was the following 12-long cohesive end sequence of bacteriophage λ DNA (17):

p G G G C G G C G A C C T C—
$_{OH}$G—
left end
—G$_{OH}$
—C C C C G C C G C T G G A p
right end

This sequence was subsequently confirmed by others (10, 16), and, when extended to the first few internal nucleotides, was found to possess a hyphenated twofold axis of symmetry which is thought to be the recognition site of the putative *ter* (or *cos*) endonuclease (2, 5, 11).

Experiments bearing on the mechanism of packaging of the λ chromosome into preformed heads suggest that the *cos* site alone is not sufficient for this process, but that there may be other unique sequences both to the left and right of the *cos* sequence (7, 13) which are recognized by specific proteins. We therefore decided to investigate these DNA sequences using the Maxam and Gilbert DNA sequencing technique (9).

The right end of λ DNA can be labeled by incubating a mixture of λ DNA and [α-^{32}P]dGTP with *Escherichia coli* DNA polymerase I. The DNA polymerase will incorporate three [^{32}P]-dGMP residues onto the right 3′ terminus and then stop. Further synthesis will not occur because the triphosphate of the next residue to be incorporated, dCMP, is not provided in the reaction mixture. No synthesis will occur at the left end since dATP, which is required to supply the first nucleotide, dAMP, is not present during the incubation. Hence the DNA is ready for sequencing after gel filtration to separate the labeled DNA from unreacted triphosphate. Similarly, the left end can be labeled by incubating the DNA and polymerase with [α-^{32}P]dATP. In this case, a single [^{32}P]dAMP is incorporated onto the left end, whereas no synthesis occurs at the right end.

In this report we present sequences of 61 nucleotides at the left end and 117 nucleotides at the right end (as measured from the symmetrical axis of the *cos* site) that were determined using DNA extracted from bacteriophage λ*c*I857*Sam*7. A perfect inverted repeat sequence of 10 nucleotides was present near the left end, and one of 15 nucleotides was present near the right end. The sequence of the right end of DNA extracted from another λ bacteriophage, *c*I857*prm*116*Sam*7, was also determined and found to differ by about 10%. One of the base pair changes extended the symmetrical sequence near the right end to a 17-member inverted repeat. When the left and right ends were aligned together as exists in the circular DNA molecule or the lysogenic state, terminator codons were found in all three phases of both strands, indicating that a structural gene does not encompass the entire region.

MATERIALS AND METHODS

Isolation and labeling of λ DNA. Bacteriophage λ*c*I857*Sam*7 (obtained from M. Feiss) was grown and harvested, and the DNA was released from the phage particle by dialysis against 50% formamide, all as described in reference 15. DNA from bacteriophage λ*c*I857*prm*116*Sam*7 (the gift of E. Rosen and G. Gussin) was obtained by the same procedures. The right end was labeled in an incubation mixture (250 μl) containing 50 mM potassium phosphate, pH 7.3, 8 mM $MgCl_2$, 100 to 200 μg of λ DNA, and 0.04 mCi of [α-^{32}P]dGTP (specific activity = 250 Ci/mmol; from Amersham). The mixture was heated to 65°C for 10

min to denature those cohesive ends that were hybridized together and then held at 0°C for 10 min. The reaction was started by adding 50 U of *E. coli* DNA polymerase I (Boehringer-Mannheim), and the incubation was conducted at 6°C for 2 h. The lower temperature minimizes incorporation into random internal nicks (17). Portions were removed at various times to follow the incorporation of label into trichloroacetic acid-precipitable material. The reaction was essentially complete at 30 min, so the reaction was frequently stopped at that time by adding Na_2EDTA to 25 mM and NaOH to 150 mM, and the entire reaction mixture was passed over a Sephadex G-100 column (0.5 by 10 cm). The column was eluted with 50 mM triethylammonium bicarbonate, pH 7.8. The void volume fractions containing the labeled DNA were combined and distributed among four to eight plastic conical tubes (5-ml capacity), and the salt was removed by repeated lyophilization.

The left end was labeled in an identical reaction, except that [α-^{32}P]dATP replaced [α-^{32}P]dGTP.

Partial base modification and cleavage. The procedures of Maxam and Gilbert (9) were followed exactly, using the alternate G and alternate strong A/weak C procedures. Times of reaction were adjusted according to the length of sequence desired. Reaction times of 25 min gave fragments of approximately 75 nucleotides and shorter, whereas reaction times of 10 min gave fragments of 150 nucleotides and shorter.

RESULTS AND DISCUSSION

Individual time courses of the incorporation of [^{32}P]dAMP into the left end and [^{32}P]dGMP into the right end of λ*c*I857*Sam*7 DNA showed that the reactions were almost complete at 30 min, although only about two of the expected three dGMP residues and 85% of the expected one dAMP residue were incorporated per molecule. This somewhat unexpected result was initially of concern in the case of the [^{32}P]dGMP-labeled material, because it suggested that the mixture of DNA molecules had a variable number of [^{32}P]dGMP residues at the right end which would confuse the interpretation of the Maxam-Gilbert sequencing gels. However, as Fig. 1C and D and Fig. 2 show, this turned out not to be a problem, i.e., doublet or triplet bands suggesting a variable number of [^{32}P]dGMP residues on the end were not seen. Thus each right end appeared to have either three labeled dGMP residues or none at all. Since the incubation conditions were selected to minimize incorporation at internal nicks, it seems likely that the incorporation was lower at the ends than expected because (i) not all cohesive ends denatured during the 65°C heat step or (ii) some of the cohesive ends annealed during the quick-cool and incubation steps. No attempt was made to distinguish between these two and/or other possibilities.

After separating the labeled λ*c*I857*Sam*7 DNA from the unreacted triphosphates, the DNA was subjected to the base modification reactions followed by the cleavage and gel electrophoresis steps. Figure 1 shows autoradiograms of several representative sequencing gels, and Fig. 3 shows the sequence determined. From the autoradiograms of the cleavage products of λ DNA-A* (λ DNA labeled with [^{32}P]-dAMP) in Fig. 1A and B, the sequence of 61 nucleotides at the left end can be deduced. The sequence of the first eight internal nucleotides (positions 7 through 14 of the lower or r-strand in Fig. 3) is the same as that published earlier (2, 10, 16).

Figures 1C and D and Fig. 2A show autoradiograms of the cleavage products of λ DNA-3G* (λ DNA labeled with three [^{32}P]dGMP's). By analyzing these and other autoradiograms, a 117-nucleotide sequence was deduced. The first seven internal residues (positions −7 through −13 of the l-strand in Fig. 3) correspond to the sequence published earlier (2, 5, 11, 16). Since this sequence includes all four bases, it provides strong evidence that the modification and cleavage reactions are proceeding reliably.

Figure 3 presents the deduced sequence information with the two cohesive ends annealed together as exists in the circular λ DNA molecule or the lysogenic state.

The sequence from positions −5 through +5 was taken from references 11 and 17. The rest of the minus (−) positions were obtained using λ DNA-3G*, and the rest of the plus (+) positions were obtained from λ DNA-A*. The cohesive ends extend from positions −6 through +6. A guanine plus cysotine-rich region which includes the *cos* site (18 out of 20 base pairs are dG:dC) is followed immediately on the left end by an adenine-thymine-rich region (19 out of 23 base pairs are dA:T). All three phases of both strands possess at least one termination codon, as shown in Fig. 3. The r-strand has one potential initiator codon which is in phase with a terminator codon that is 18 triplets away. The l-strand contains four potential initiator codons, only one of which is in phase with a terminator codon downstream within the region sequenced. Two of the others, if actual initiation sites, would begin a structural gene which spans the *cos* site.

Two regions possess perfect twofold rotational symmetry, i.e., a 10-nucleotide region near the left end and a 15-nucleotide region near the right end, as shown in Fig. 3. In fact, a variety of potential regulatory signals are present in the region between positions −38 and −68 near the right end. In this region the r-strand has four terminator codons, three of which are contiguous, and the l-strand also has four terminators,

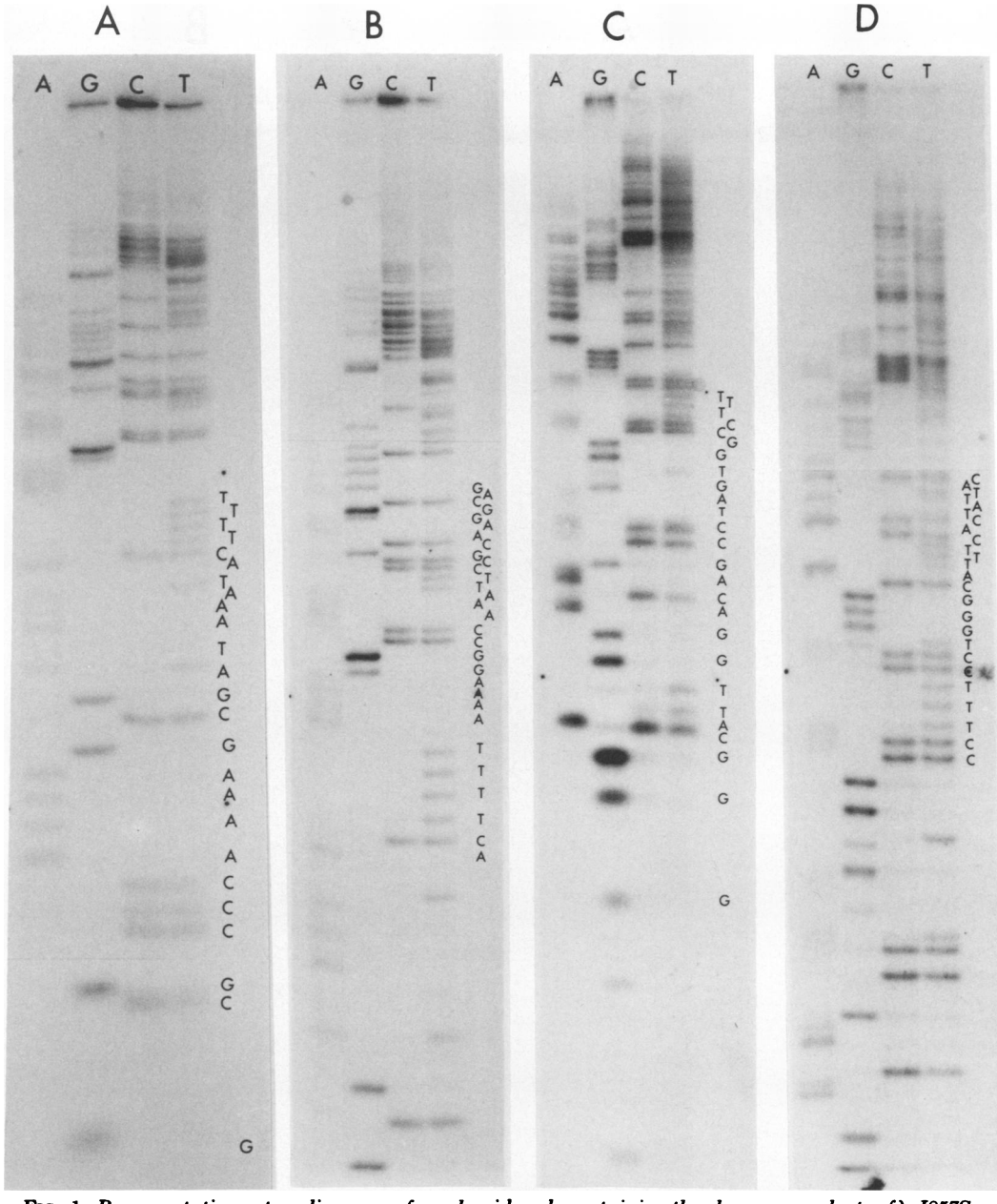

FIG. 1. *Representative autoradiograms of acrylamide gels containing the cleavage products of λcI857Sam7 DNA, (A) and (B), labeled with [^{32}P]dAMP, and (C) and (D), labeled with [^{32}P]dGMP. The gels shown in (A) and (C) ran for a shorter length of time than those shown in (B) and (D).*

two of which are contiguous. In addition to the axis of sequence symmetry encompassing 15 residues, there is a symmetrical axis of terminator codons encompassing four codons, two on each strand. This region also contains three potential initiator codons, i.e., two in the l-strand, for which there are no in-phase terminator codons downstream within the region sequenced, and one in the r-strand. Of the structural genes known in λ, the R gene maps the nearest to the right end, although it is thought to be about 2,300 base pairs away (4). It codes for an endolysin whose amino acid sequence is known (8). As expected, none of the nucleotide sequence presented here matches any region of the amino acid sequence of this protein.

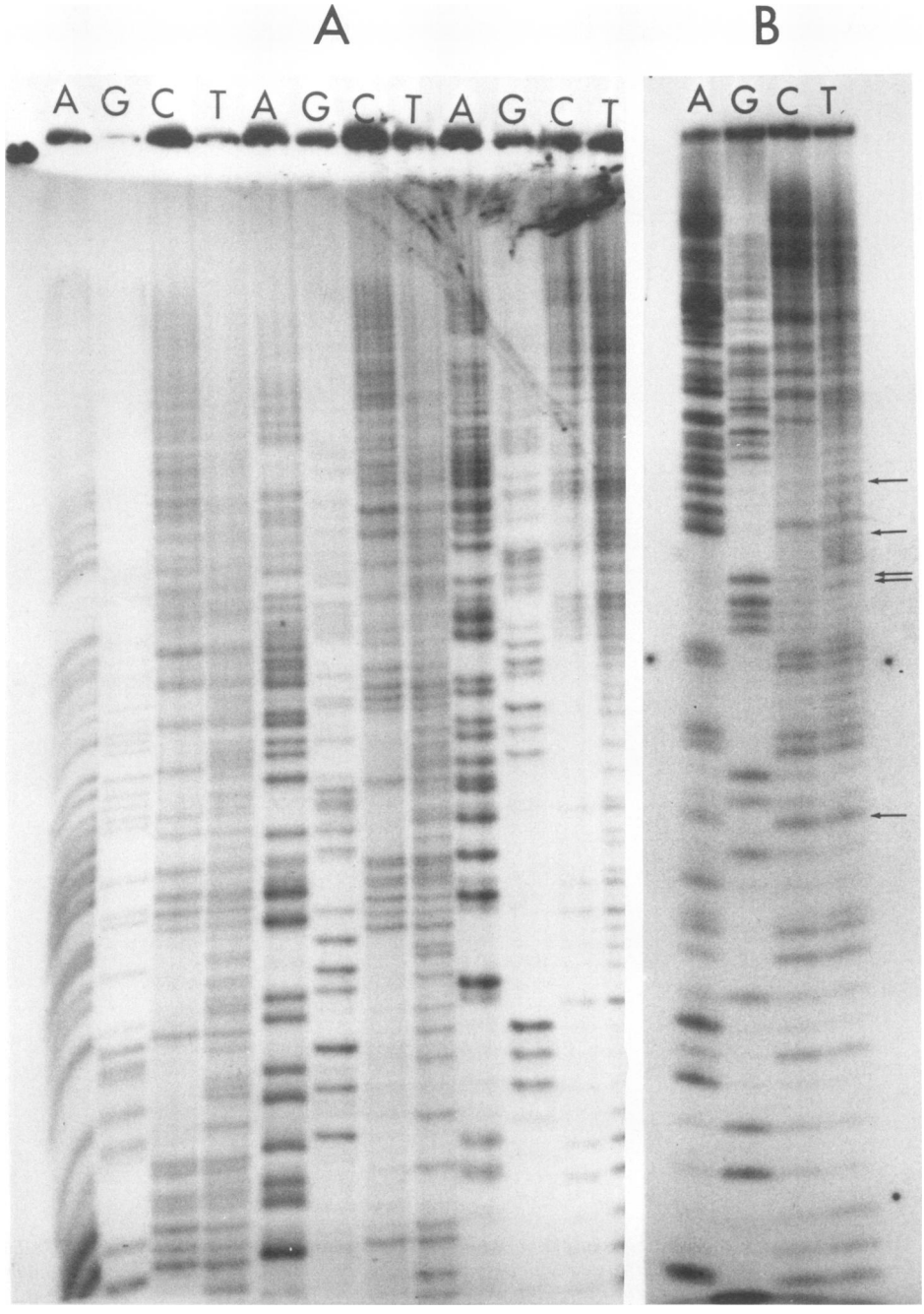

FIG. 2. *(A) A representative autoradiogram of three groups of the four cleavage products of λcI857Sam7 DNA labeled with [^{32}P]dGMP from which the sequence from positions −10 to −117 (Fig. 3) can be determined. The group of four on the left have been electrophoresed for the longest time, the group on the right for the shortest time. (B) A representative autoradiogram of an acrylamide gel of the cleavage products of λprm116 DNA labeled with [^{32}P]dGMP. The arrows indicate places at which the sequence varies from that of λcI857Sam7 DNA.*

The sequence adjacent to the right end was also determined, using another strain of bacteriophage λ, *c*I857*prm*116S*am*7, or abbreviated as λ*prm*116. Figure 2B shows a representative autoradiogram of the cleavage products from the DNA of this strain labeled with [^{32}P]dGMP.

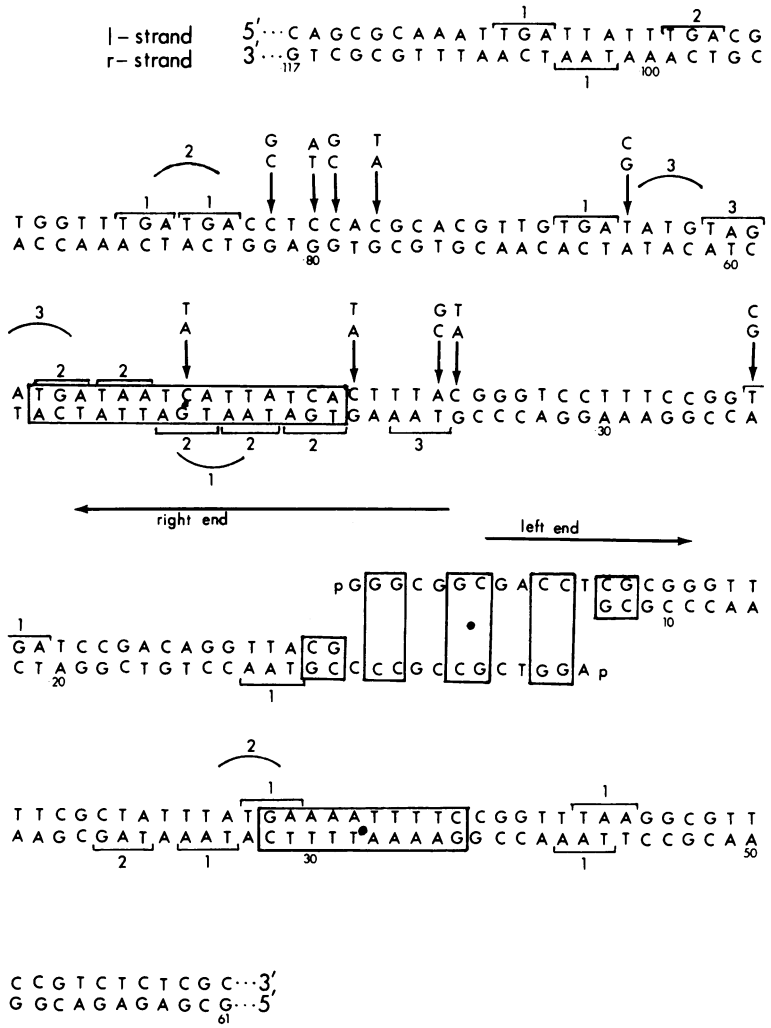

FIG. 3. *The sequence of λcI857Sam7 deduced from the combined information of autoradiograms similar to those shown in Fig. 1 and 2. The sequences enclosed within boxes are symmetrical about an axis indicated by a dot. The symbol ⌐ indicates a termination codon, and the symbol ⌒ indicates a potential initiation codon. The numbers adjacent to these symbols indicate the three relative phases in which the triplet is found. The base pair differences of λprm116 DNA are shown above the vertical arrows.*

The five arrows indicate sites at which the sequence differs from the λcI857*Sam7* sequence shown in Fig. 1A and B. When the gel was electrophoresed for a much longer time (similar to the one shown in Fig. 2A), another five differences were observed for a total of ten differences between these two λ DNAs in the first 110 nucleotides from the right end. These changes are shown in Fig. 3. They include both transitions and transversions and eliminate three of the above-mentioned terminator codons. The change at position −50 occurs at the center of the 15-member inverted repeat sequence (but does not disturb its symmetry) and the change at position −42 extends it to a 17-member inverted repeat. According to the equation of Brezinski (1), the probability that a sequence of this symmetry will occur, i.e, that eight positions are followed by eight symmetrical positions, is:

$$(½)^8 \sum_{i=8}^{8} \sum_{j=8}^{i} \frac{8!(½)^i}{i!(i-j)!(8-i)!} = (½)^8(½)^8$$

$$= (¼)^8 = 1.5 \times 10^{-5}$$

If the same probability equation is applied to the hyphenated symmetrical sequence about the *cos* site, a sequence of eight positions (beginning at the first perfect symmetrical position) with

five perfectly symmetric positions, two purine/purine or pyrimidine/pyrimidine symmetric positions, and one nonsymmetric position, the probability is:

$$(\tfrac{1}{2})^8 \sum_{i=7}^{8} \sum_{j=5}^{i} \frac{8!(\tfrac{1}{2})^i}{(i)!(i-j)!(8-i)!} = 8.5 \times 10^{-3}$$

Therefore, the symmetrical sequence between positions −42 and −58 in λ*prm*116 DNA is $8.5 \times 10^{-3}/1.5 \times 10^{-5} = 560$ times less likely to occur randomly than the symmetry of the *cos* region—a sequence thought to be recognized by a specific (*cos*) nuclease during the production of monomer length λ DNA from concatenated DNA. Similar calculations for the smaller seven-position symmetrical sequence near the right end of λ*c*I857*Sam*7 DNA show that it is $8.5 \times 10^{-3}/(\tfrac{1}{4})^7 = 140$ times less likely to occur randomly than the *cos* symmetry.

It is not possible to deduce the biological significance of a given DNA region from its sequence alone. However, proteins that bind to or cleave specific DNA regions, such as repressors and restriction enzymes, frequently do so by recognizing sequences with a twofold axis of symmetry. Thus, the occurrence of such sequences near the right and left ends of λ DNA suggests, but doesn't prove, that they may also be recognized by a specific protein(s).

Since packaging of the chromosome begins at the left end (13), and the right end is subsequently positioned so that it extends into the tail (3, 12, 14), good candidates for the proteins that may recognize these sequences are those which are involved in DNA packaging and/or tail assembly.

The complete genetic histories of the two λ strains, λ*c*I857*Sam*7 and λ*prm*116, are not known. The right end of both was obtained by crossing in the *Sam*7 marker from a λimm^{434} strain. This mutation prevents lysis of the host, but does not interfere with phage multiplication (6). Therefore, this region of the DNA seems to have undergone 10% divergence since the original strain, *Sam*7, was constructed in 1967, perhaps due, in part, to laboratory mutagenesis. Since bacteriophage λ can accommodate many base changes in this region, it suggests that the sequence of other regions may also be unique to the given strain. Hence λ strains must be carefully characterized before DNA sequence analysis is begun and biological significance is assigned to a deduced sequence.

ACKNOWLEDGMENTS

This research was supported by Public Health Service grant GM-21696 from the National Institute of General Medical Sciences and National Science Foundation grant PCM 76-13461.

LITERATURE CITED

1. **Brezinski, D. P.** 1975. Statistical significance of DNA sequence symmetries. Nature (London) **253**:128–130.
2. **Brezinski, D. P., and J. C. Wang.** 1973. The 3′-terminal nucleotide sequence of λ DNA. Biochem. Biophys. Res. Commun. **50**:398–404.
3. **Chattoraj, D. K., and R. B. Inman.** 1974. Location of DNA ends in P2, 186, P4 and lambda bacteriophage heads. J. Mol. Biol. **87**:11–22.
4. **Davidson, N., and W. Szybalski.** 1971. Physical and chemical characteristics of lambda DNA, p. 13–44. *In* A. D. Hershey (ed.), The bacteriophage lambda. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
5. **Ghangas, G. S., E. Jay, R. Bambara, and R. Wu.** 1973. Nucleotide sequence analysis of DNA. XI. The 3′-terminal sequences of bacteriophage λ and φ80. Biochem. Biophys. Res. Commun. **54**:998–1007.
6. **Harris, A. W., D. W. A. Mount, C. R. Fuerst, and L. Siminovitch.** 1967. Mutations in bacteriophage λ affecting host cell lysis. Virology **32**:553–568.
7. **Hohn, B.** 1975. DNA as substrate for packaging into bacteriophage lambda, *in vitro*. J. Mol. Biol. **98**:93–106.
8. **Imada, M., and A. Tsugita.** 1971. Amino acid sequence of λ phage endolysin. Nature (London) New Biol. **233**:230–231.
9. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
10. **Murray, K., and N. E. Murray.** 1973. Terminal nucleotide sequences of DNA from temperate coliphages. Nature (London) New Biol. **243**:134–139.
11. **Murray, K., and R. W. Old.** 1974. The primary structure of DNA, p. 117–185. *In* Waldo E. Cohn (ed.), Progress in nucleic acid research and molecular biology, vol. 14. Academic Press Inc., N.Y.
12. **Padmanabhan, R., R. Wu, and V. C. Bode.** 1972. Arrangement of DNA in lambda bacteriophage heads. III. Location and number of nucleotides cleaved from λ DNA by micrococcal nuclease attack on heads. J. Mol. Biol. **69**:201–207.
13. **Syvanen, M.** 1975. Processing of bacteriophage lambda DNA during its assembly into heads. J. Mol. Biol. **91**:165–174.
14. **Thomas, J. O.** 1974. Chemical linkage of the tail to the right-hand end of bacteriophage lambda DNA. J. Mol. Biol. **87**:1–9.
15. **Thomas, M., and R. W. Davis.** 1975. Studies on the cleavage of bacteriophage lambda DNA with *Eco*RI restriction endonuclease. J. Mol. Biol. **91**:315–328.
16. **Weigel, P. H., P. T. Englund, K. Murray, and R. W. Old.** 1973. The 3′-terminal nucleotide sequences of bacteriophage λ DNA. Proc. Natl. Acad. Sci. U.S.A. **70**:1151–1155.
17. **Wu, R., and E. Taylor.** 1971. Nucleotide sequence analysis of DNA. II. Complete nucleotide sequence of the cohesive ends of bacteriophage λ DNA. J. Mol. Biol. **57**:491–511.