Proc. Nat. Acad. Sci. USA
Vol. 70, No. 4, pp. 1151–1155, April 1973

The 3′-Terminal Nucleotide Sequences of Bacteriophage λ DNA

(*ter* function/recognition site/rotational symmetry/coliphage)

PAUL H. WEIGEL, PAUL T. ENGLUND, KENNETH MURRAY*, AND ROBERT W. OLD*

Department of Physiological Chemistry, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205; and
*Department of Molecular Biology, University of Edinburgh, Edinburgh, Scotland

Communicated by Albert L. Lehninger, February 8, 1973

ABSTRACT **Analyses of radioactive oligonucleotides in endonuclease digests of 3′-terminally labeled λ DNA revealed the 3′ terminal sequence -GTTACG for the *l* strand and -ACCCGCG for the *r* strand. These sequences, together with those previously known for the 5′ cohesive ends, provide a total of 25 known base-pairs in the vicinity of the termini. When the cohesive ends are paired, the sequence between the nicks can be bisected by a 2-fold rotational axis of symmetry. Five of the first eight base-pairs, on either side of the axis, are rotationally symmetric. This symmetry may be involved in the recognition of the site by an enzyme responsible for formation of the cohesive ends.**

The DNA of bacteriophage λ is a duplex molecule of about 46,500 base-pairs, with single-stranded projections of 12 nucleotides at each 5′-terminus. These projections, or cohesive ends, are of complementary sequence and, by their specific base-pairing, the DNA may form either a ring or a linear aggregate (1, 2).

The mechanism by which the cohesive ends are formed is not completely understood. The mature viral DNA molecule may be produced from a replicative intermediate, possibly a concatemer, by endonucleolytic attack at the site of the cohesive ends (for a review, see ref. 3). This endonuclease is the *ter* function, whose existence was inferred from genetic studies (4) and which has been detected recently in extracts of *Escherichia coli* infected with phage λ (5). Presumably, this endonuclease binds to a unique sequence of nucleotides in the region of the cohesive ends and introduces two nicks, 12 base-pairs apart, on opposite strands. The mature λ DNA molecules may then be released by disruption of the base pairs between the nicks.

Our objective was to identify the nucleotide sequence recognized by the *ter* enzyme. The sequence of bases between the two nicks is known (2). To determine sequences beyond these nicks, we labeled λ DNA at specific sites at or near its 3′-termini (6) and determined the sequence of oligonucleotides produced by degradation of the DNA with a nonspecific endonuclease (7). The 3′-terminus of the *l* strand is -GTTACG, and that of the *r* strand is -ACCCGCG. Brezinski and Wang have recently informed us that they have independently obtained sequences in agreement with these results (8).

MATERIALS AND METHODS

DNA. Phage were produced from the λ-lysogen *E. coli* N1383, (λcI857 *Sam7*), a gift of Dr. John Little (9, 10), and DNA was isolated (6). Analysis of the DNA on an alkaline sucrose gradient indicated that more than 90% of the strands were intact.

3′-Terminal Labeling. DNA termini were labeled with [^{32}P]nucleotides in a reaction catalyzed by the T4 DNA polymerase (6). The DNA [3.7 μmol of nucleotide in 1.95 ml containing 46 mM NaCl–86 mM Tris·HCl (pH 8.0)] was heated to melt cohered ends (74°, 10 min, followed by quick cooling to 0°). A solution (0.55 ml) containing polymerase (6), a single [α-^{32}P]deoxynucleoside triphosphate (6), $MgCl_2$, and 2-mercaptoethanol was added immediately after cooling. Final concentrations were: 1.5 mM DNA, 67 mM Tris·HCl (pH 8.0), 36 mM NaCl, 6 mM $MgCl_2$, 6 mM 2-mercaptoethanol, 0.1 mM triphosphate (2×10^{10} cpm/μmol), and 600 units/ml of polymerase. Incubation was at 11°, and when acid-insoluble radioactivity in an aliquot reached a limiting value (usually 60 min), the reaction was terminated by addition of EDTA (to 20 mM) and by extraction with phenol. Triphosphates were removed by exhaustive dialysis against 1.0 M NaCl–10 mM Tris·HCl (pH 8.0)–1.0 mM EDTA. The DNA was then purified by zone sedimentation in an alkaline sucrose gradient (5–20% sucrose in 0.2 M NaOH–0.8 M NaCl–1 mM EDTA; 25,000 RPM, 15 hr, 5°, Spinco SW25 rotor). This purification removed small labeled DNA fragments, which contained about 30% of the radioactivity, but which were not detectable by A_{260} measurements. These fragments were probably not a consequence of endonuclease activity in the incubation mixture, because similar fragments were not observed in identical experiments with T7 DNA (6); they were probably present in the DNA used for labeling. The purified intact [^{32}P]DNA strands were then dialyzed against 50 mM NaCl–10 mM Tris·HCl (pH 8.0)–1.0 mM EDTA. A portion of the DNA was used for strand separation (11), and a sample of each isolated strand was used for nearest-neighbor analysis (12). The remainder of the purified [^{32}P]DNA was used for sequence analysis.

Endonuclease Digestion of the Labeled DNA. Labeled DNA was transported from Baltimore to Edinburgh in evacuated, sealed tubes. The solutions were concentrated to 1.0 ml by rotary evaporation, dialyzed against 0.3 mM NaCl, and further concentrated to 0.1 ml. The labeled DNA was then digested with pancreatic DNase (Worthington Biochemical Corp.). Conditions were chosen to achieve an equable distribution of radioactive oligonucleotides from trial digests analyzed by electrophoresis on Whatman AE81 paper (pH 3.5). Typical conditions were 50 μg/ml of DNase in 20 mM Tris·HCl (pH 7.5)–10 mM $MgCl_2$ at 37° for 3–12 hr. The

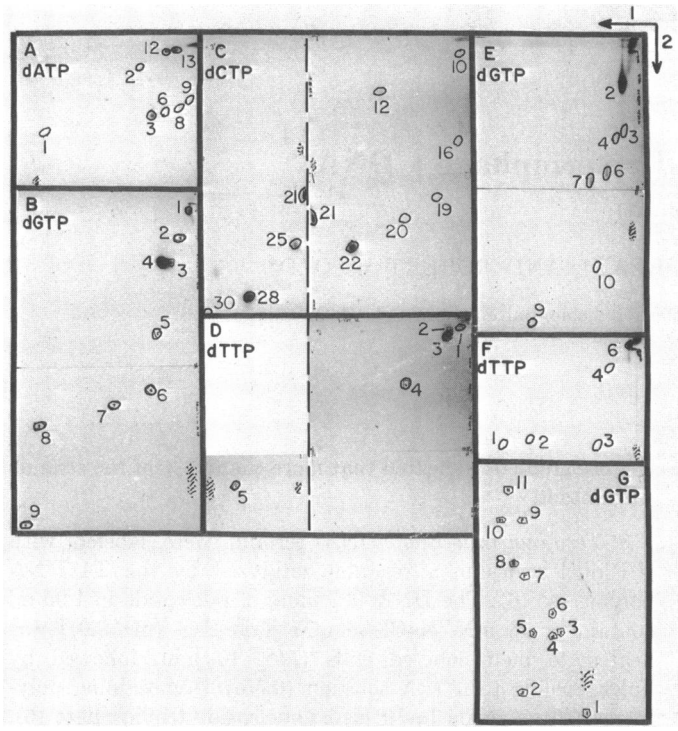

FIG. 1. Radioautograms of two-dimensional oligonucleotide maps of DNase digests of 3′-terminally labeled λ DNA. Electrophoretic method *iii* was used for maps *A*, *B*, *C*, and *D*, method *iv* for maps *E* and *F*, and method *v* for map *G*. The *labeling* triphosphate is shown on each panel. *Numbers* by oligonucleotide spots refer to code numbers in Table 1. The *hatched* zones enclosed by *dots* indicate the position of the blue dye, xylene cyanol FF. The *broken line* through panels *C* and *D* denotes the site where the chromatogram was cut after the first dimension. The *numbered arrows* denote the first and second dimensions in the two-dimensional separations.

digests were fractionated by electrophoresis by the following methods: *i*. AE81 paper at pH 3.5; *ii*. Whatman DE81 paper at pH 2; *iii*. two-dimensional on DE81 paper at pH 9.7, followed by pH 2 (13); *iv*. two-dimensional on AE81 paper at pH 3.5 followed by DE81 paper at pH 2; *v*. two-dimensional on cellulose acetate at pH 3.5 in 7 M urea, followed by transfer to thin layers of polyethyleneimine–cellulose and chromatographic development with 1.6 M formic acid (adjusted to pH 3.6 with pyridine) (14). Labeled oligonucleotides were located by radioautography and eluted (15).

Analysis of Oligonucleotides. The electrophoretic mobility of each [^{32}P]oligonucleotide was determined by methods *i* and *ii*. Oligonucleotides isolated by method *i* were examined by method *ii*, and *vice versa*. The 3′-penultimate base of each oligonucleotide was determined, after dephosphorylation with *E. coli* alkaline phosphatase (Worthington), by digestion with spleen phosphodiesterase (a generous gift of G. Bernardi) and analysis of the 3′-mononucleotides by electrophoresis by method *i* or by chromatography on polyethyleneimine–cellulose with 5% formic acid. (phosphatase digestion: 2 μg of enzyme in 20 μl of 60 mM NH_4HCO_3 at 37° for 45 min, followed by removal of phosphatase on phosphocellulose (16); phosphodiesterase digestion: 0.5 unit/ml (17) in 10 μl of 50 mM ammonium acetate–1 mM EDTA (pH 5.6) at 37° for 45 min).

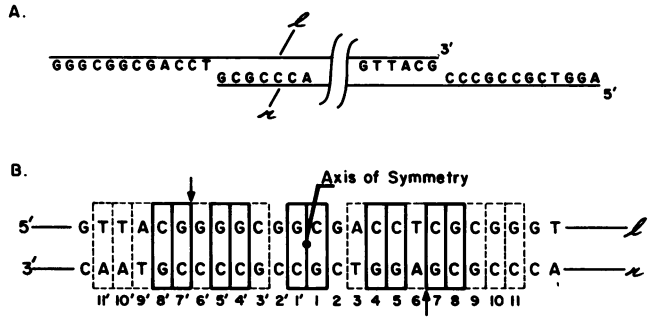

FIG. 2. (*A*) Nucleotide sequences at the termini of λ DNA. The sequences of the cohesive ends, as well as the 3′-terminal dGMP residues, were identified by Wu and Taylor (2) (see also ref. 7). (*B*) Symmetry in the region of the cohesive ends. The ends are paired, and nicks are indicated by *vertical arrows*. The 2-fold axis of symmetry is perpendicular to the page. Symmetrical base pairs are enclosed in *solid boxes* and pairs symmetrical in purine–pyrimidine orientation are enclosed in *dashed boxes*. Base pairs are *numbered* on either side of the axis.

RESULTS

Incubation of λ DNA with the T4 DNA polymerase and a single [α-^{32}P]deoxynucleoside triphosphate results in the addition of a [^{32}P]nucleotide to a natural 3′-terminus provided that the triphosphate is complementary to the first base in the adjacent cohesive end. However, if no nucleotide can be added to the natural terminus, then the enzyme will sequentially remove nucleotides from this terminus until a nucleotide is reached that can be exchanged for a labeled one (6). For example, during incubation with [α-^{32}P]dGTP, three [^{32}P]nucleotides would be added to the 3′ terminus of the *l* strand and the 3′-terminal dGMP residue of the *r* strand would be exchanged for a labeled one (see Fig. 2*A* for structure of cohesive ends). Measurements of the extent of labeling with each triphosphate† and of the distribution of radioactivity between the two separated strands supported this general scheme. With [α-^{32}P]dGTP, about four [^{32}P]-nucleotides were incorporated per molecule, three of which were in the *l* strand. With [α-^{32}P]dATP or [α-^{32}P]dCTP, about two nucleotides were incorporated per molecule, with one on each strand. With [α-^{32}P]dTTP, only about one nucleotide was incorporated per molecule, even after 165 min of incubation, and about 80% of the radioactivity was in the *l* strand. Presumably, no dTMP residue exists close enough to the 3′-terminus of the *r* strand to allow efficient exchange for a radioactive nucleotide.

The strategy for 3′-terminal sequence analysis of the labeled DNA is based upon that developed for 5′-terminally labeled DNA (7). If 3′-terminally labeled DNA is degraded with a nonspecific endonuclease, all the labeled oligonucleotides in the digest must belong to one of two families. Each family arises from one of the strands and consists of a homologous series of oligonucleotides whose sequences overlap progressively from the 3′-terminus. Although all labeled oligonucleotides in the digest must have the same 3′-terminal base, each may be assigned to its family by identification of its 3′-penultimate base, which generally differ in the two families. Once all oligonucleotides have been thus assigned,

† Measurements were corrected for the fact that about 30% of the radioactivity was present in small fragments.

their sequences can be determined from the difference in mobility (methods *i* and *ii*) between adjacent members of each homologous series. This difference in mobility between adjacent homologs is diagnostic of the single base by which they differ (15, 18). Sequences can be confirmed by electrophoresis by methods *iii*, *iv*, or *v*. Finally, nearest-neighbor analysis of the separated strands is used to assign each family to one of the two strands of the DNA.

Application of this strategy to λ DNA labeled at the 3'-termini with [^{32}P]deoxyadenylate revealed the terminal sequence -GTTA on the *l* strand and -ACCCGCGA on the *r* strand (Table 1, Fig. 1). Inspection of the cohesive end sequences (see Fig. 2*A*) indicates that labeling of the *r* strand must have occurred by addition of a [^{32}P]dAMP residue to the 3'-terminal dGMP residue of that strand. Therefore, the terminal sequence of the *r* strand must be -ACCCGCG. Labeling of the *l* strand must have been preceded by the hydrolytic removal of nucleotides from the terminus.

Interpretation of the experiments with [α-^{32}P]dGTP was aided by Wu and Kaiser's finding that the 3'-terminal residue of each strand is dGMP (19). Labeling of λ DNA with [α-^{32}P]dGTP should result in exchange of the terminal dGMP residue of the *r* strand for [^{32}P]dGMP, and the addition of three [^{32}P]dGMP residues to the terminal dGMP

TABLE 1. *Oligonucleotides derived from 3'-terminally labeled λ DNA*

	l strand				*r* strand			
Labeling triphosphate	Oligonucleotides*	Mobility¶ *i*	Mobility¶ *ii*	Code numbers§	Oligonucleotides*	Mobility¶ *i*	Mobility¶ *ii*	Code numbers§
dATP	TA	0.68	1.73		GA	0.43	1.49	
	TTA	0.29	0.60	A1	CGA	0.28	1.42	
	GTTA	0.07	0.17	A2	GCGA	0.07	0.51	A3
	GGTTA**	0.02	0.06	A12	CGCGA	0.04	0.50	A6
					CCGCGA	0.03	0.46	A8
					CCCGCGA	0.02	0.41	A9
					ACCCGCGA	0.01	0.30	
dGTP	GGG‖	0.07	0.25	E2 G8	CG	0.59	1.87	E9 G1
	GG	0.27	0.79	B8 E7 G5	GCG	0.15	0.68	B7 E6
	GGGG	0.02	0.09		CGCG	0.10	0.70	G3
	CGGGG	0.01	0.09					
	CGGG	0.04	0.03					
	CGG	0.16	0.73	B7 E6 G4				
	CG	0.59	1.87	E9 G1				
	ACG	0.21	1.13	B9 E10				
	ACGG	0.05	0.50	B5 E3 G6				
	TACGG			G7				
	TACG	0.10	0.52	E4				
	TTACGG	0.01	0.09	G10				
dCTP	AC	0.71	2.28	C28	GC	0.60	1.84	C25
	TAC	0.40	1.44	C21	CGC	0.37	1.71	C22
	TTAC	0.17	0.49	C12	CCGC	0.26	1.41	C20
	GTTAC	0.03	0.15		CCCGC	0.10	1.32	C19
					ACCCGC	0.04	0.94	C16
					CC†	0.93	2.41	C30
					CCC†	0.62	2.08	
dTTP	TT	0.75	0.87	F1				
	GT†	0.60	0.86	F2	AT‡	0.59	1.65	
	GTT	0.17	0.25	D4 F4	AAT	0.24	0.98	D5 F3
	GGTT**	0.05	0.06	D3 F6				

* Assignment of an oligonucleotide to a family was by identification of the 3'-penultimate base. Assignment of a family to a strand of λ DNA was by nearest-neighbor analysis of separated strands (in each case about 85% of the radioactivity was in a single 3'-mononucleotide). Dinucleotide sequences determined by nearest-neighbor analysis on separated strands are underlined.

† These oligonucleotides were probably labeled because the polymerase occasionally failed to halt degradation after it reached the first nucleotide that could be exchanged for a labeled one.

‡ These oligonucleotides may have arisen from the *r* strand, although little radioactivity was incorporated into that strand.

§ Letters refer to panels and numbers refer to oligonucleotides in Fig. 1. Oligonucleotides not listed in these columns were identified only by methods *i* and *ii*.

¶ Mobility during electrophoresis by methods *i* and *ii* relative to the dye xylene cyanol FF.

‖ Structure confirmed by partial digestion with venom phosphodiesterase (13).

** The 5' base is G or T. G is more probable.

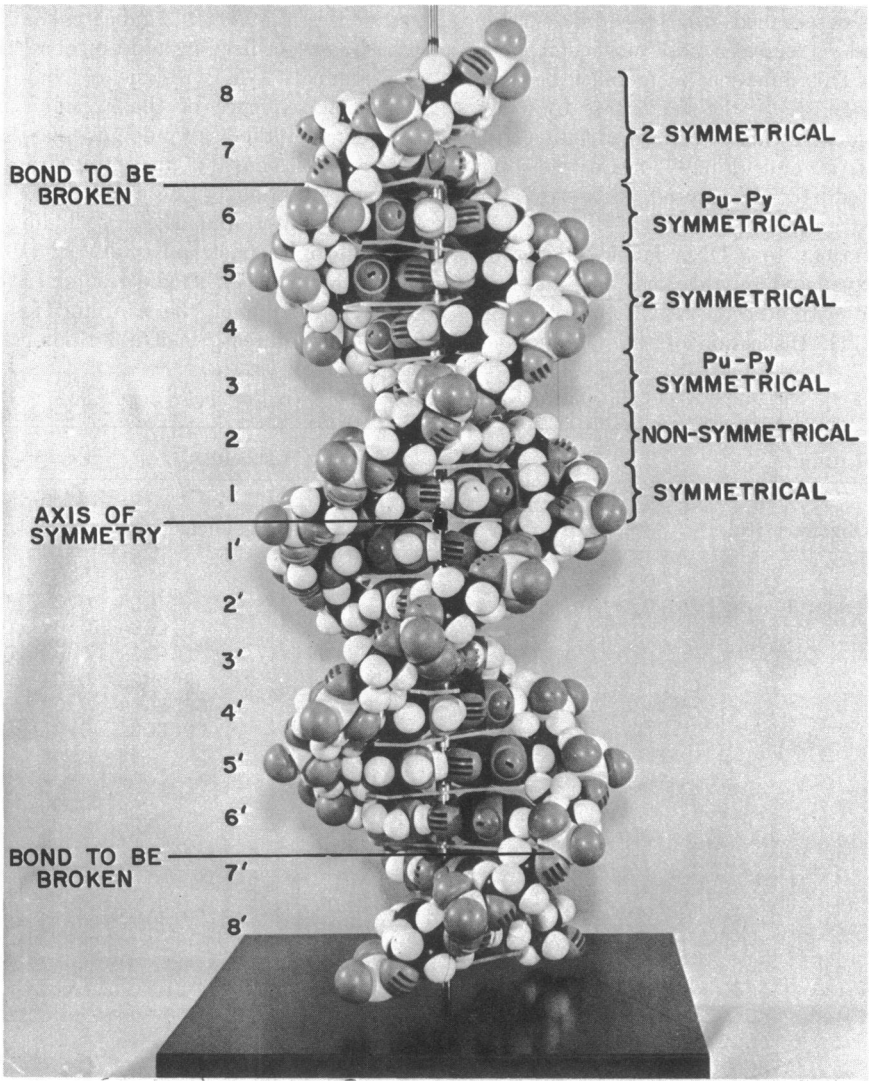

FIG. 3. Model of the "cohesive ends" region of the precursor of mature λ DNA (constructed of Corey–Pauling–Koltun models from the Ealing Corp.). Numbering of base-pairs is the same as in Fig. 2. The 2-fold axis of symmetry is perpendicular to the page.

of the *l* strand. The sequences derived from the *r* strand confirm sequences obtained from the DNA labeled with [α-^{32}P]-dATP. The sequences from the *l* strand define its 3'-terminus, and overlap sequences determined from the DNA labeled with [α-^{32}P]dATP. The terminal sequence of the *l* strand is, therefore, -GTTACG.

Additional oligonucleotides were isolated from DNA labeled with [α-^{32}P]dCTP and [α-^{32}P]dTTP. Their sequences confirmed those already deduced.

DISCUSSION

We can now write a sequence of 25 base-pairs, within which lie the nicks produced by the *ter* function. The most striking property of this sequence is rotational symmetry (Fig. 2*B*). When the two cohesive ends are paired, the sequence between the two nicks is bisected by a 2-fold rotational axis. Not only are the nicks symmetrical about this axis, but five of the first eight base-pairs on either side of the axis are symmetrical. Also, an additional five of the first 11 pairs on either side of the axis are symmetrical in their purine–pyrimidine orientation. The term "hyphenated symmetry" has been used to describe sequences of this type in which the symmetry is interrupted by some nonsymmetrical base-pairs (20).

The abundance of G-C base-pairs in this sequence greatly increases the probability of a fortuitous symmetrical arrangement of base-pairs about a central axis. However, the 2-fold symmetry does suggest mechanisms by which the *ter* enzyme might recognize this sequence. For example, a dimeric enzyme, containing identical subunits that are symmetrical about a 2-fold axis, could bind to the DNA so that its axis would align with that of the symmetrical sequence. Symmetrical catalytic sites on the dimeric enzyme could then simultaneously introduce the two nicks. A hyphenated sequence of 10 symmetrical base-pairs would occur by chance only once in about a million base-pairs. Since λ DNA contains only about 46,500 base-pairs, it is reasonable to assume that the sequence would occur only once. However, the shorter sequence of five base-pairs with which each subunit is associated would occur by chance about 50 times in a random molecule the size of λ DNA, making it unlikely that the enzyme would be active other than as a dimer.

There are several ways in which a dimeric enzyme could

bind to this symmetrical sequence. One is that the enzyme binds to the native helix and interacts with functional groups within the grooves of the DNA. Inspection of a space-filling model shows that the phosphodiester bonds to be broken, as well as some possible specificity-conferring functional groups on the symmetrical base-pairs, occur on the same face of the helix (Fig. 3). One subunit could interact with base-pair **1** in the minor groove and base-pairs **4**, **5**, **7**, and **8** in the major groove; the other subunit could interact in a corresponding way with the prime-numbered pairs. It is possible that base pairs symmetrical only in purine–pyrimidine orientation are also involved in specific interaction with the enzyme. It is also possible that this sequence may dictate local stereochemical characteristics of the helix that differ from those of the B form (21), and that these characteristics may be recognized by the enzyme.

Other recognition mechanisms might require more drastic changes in the local structure of the DNA, such as the formation of loops or cruciforms as suggested by Gierer (22), or perhaps even the generation of more extensive single-stranded regions (23). Cruciforms may be formed from any rotationally symmetrical sequence and, although the present example is complicated by hyphenation, it may be possible to derive these structures if one invokes G-T and A-C base-pairs, for which there is some precedent (24–26).

Recognition of the symmetrical sequence by the λ *ter* function appears possible by these or other mechanisms. However, until rigorous physical evidence is available, it may not be possible to distinguish among them.

Symmetrical recognition sequences occur in other systems. Spleen acid DNase (27), actinomycin D (28), and restriction and modification enzymes from *Hemophilus influenzae* (ref. 29, and unpublished work of Old, Murray, and Roizes) and *E. coli* plasmids (ref. 30, and unpublished results of Bigger, Murray, and Murray) recognize rotationally symmetrical sequences. Available evidence suggests that the spleen and *E. coli* plasmid RI enzymes may be dimers of identical subunits (27, 30). Extensive symmetry similar to that surrounding the cohesive ends of λ DNA is not apparent within the sequence of 19 base-pairs at the cohesive ends of DNA from phages 186, P2, and 299 (31, 7), although the identification of the sequences adjacent to the cohesive ends could reveal the existence of some rotational symmetry. The rotational symmetry in the terminal sequences of λ DNA, however, is striking and provides an attractive model for the interaction of the *ter* function with its substrate.

We are grateful to Patricia Clinkenbeard for expert technica assistance. We thank Drs. James Wang for a preprint of an unpublished manuscript, and John Little and Michael Hershfield for helpful advice. P.H.W. is supported by NIH Training Grant 5T01 GM00184-13 and P.T.E. by NIH Grant CA 13602-04 and by a Faculty Research Award from the American Cancer Society. K.M. is supported by the Science Research Council (U.K.) and R.W.O. by a research training grant from the Medical Research Council (U.K.).

1. Hershey, A. D., Burgi, E. & Ingraham, L. (1963) *Proc. Nat. Acad. Sci. USA* **49**, 748–755.
2. Wu, R. & Taylor, E. (1971) *J. Mol. Biol.* **57**, 491–511.
3. Yarmolinsky, M. B. (1971) in *The Bacteriophage Lambda*, ed. Hershey, A. D. (Cold Spring Harbor Laboratory, New York), pp. 97–111.
4. Mousset, S. & Thomas, R. (1969) *Nature* **221**, 242–244.
5. Wang, J. C. & Kaiser, A. D. (1973) *Nature New Biol.* **241**, 16–17.
6. Englund, P. T. (1972) *J. Mol. Biol.* **66**, 209–224.
7. Murray, K. & Murray, N. E. (1973) *Nature*, in press.
8. Brezinski, D. & Wang, J. C. (1973) *Biochem. Biophys. Res. Commun.* **50**, 398–404.
9. Parks, J. S., Gottesman, M., Shimada, K., Weisberg, R. A., Perlman, R. L. & Pastan, I. (1971) *Proc. Nat. Acad. Sci. USA*, **68**, 1891–1895.
10. Varmus, H. E., Perlman, R. L. & Pastan, I. (1970) *J. Biol. Chem.* **245**, 2259–2267.
11. Szybalski, W., Kubinski, H., Hradecna, Z. & Summers, W. C. (1971) in *Methods in Enzymology*, eds. Grossman, L. & Moldave, K. (Academic Press, New York), Vol. XXI D, pp. 383–413.
12. Josse, J., Kaiser, A. D. & Kornberg, A. (1961) *J. Biol. Chem.* **236**, 864–875.
13. Murray, K. (1970) *Biochem. J.* **118**, 831–841.
14. Southern, E. M. & Mitchell, A. R. (1971) *Biochem. J.* **123**, 613–617.
15. Sanger, F., Brownlee, G. G. & Barrell, B. G. (1965) *J. Mol. Biol.* **13**, 373–398.
16. Székely, M. & Sanger, F. (1969) *J. Mol. Biol.* **43**, 607–617.
17. Carrara, M. & Bernardi, G. (1968) *Biochemistry* **7**, 1121–1132.
18. Murray, K. (1973) *Biochem. J.* **131**, 569–583.
19. Wu, R. & Kaiser, A. D. (1968) *J. Mol. Biol.* **35**, 523–537.
20. Arber, W. & Linn, S. (1969) *Annu. Rev. Biochem.* **38**, 467–500.
21. Bram, S. & Tougard, P. (1972) *Nature, New Biol.* **239**, 128–131.
22. Gierer, A. (1966) *Nature* **212**, 1480–1481.
23. Crick, F. H. C. (1971) *Nature* **234**, 25–27.
24. Crick, F. H. C. (1966) *J. Mol. Biol.* **19**, 548–555.
25. Paetkau, V., Coulter, M. B., Flintoff, W. F. & Morgan, A. R. (1972) *J. Mol. Biol.* **71**, 293–306.
26. Hirsch, D. (1971) *J. Mol. Biol.* **58**, 439–458.
27. Bernardi, G. (1968) *Advan. Enzymol.* **31**, 1–49.
28. Sobell, H. M. & Jain, S. C. (1972) *J. Mol. Biol.* **68**, 21–34.
29. Kelly, T. J., Jr. & Smith, H. O. (1970) *J. Mol. Biol.* **51**, 393–409.
30. Hedgpeth, J., Goodman, H. M. & Boyer, H. W. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 3448–3452.
31. Padmanabhan, R. & Wu, R. (1972) *J. Mol. Biol.* **65**, 447–467.