Proc. Natl. Acad. Sci. USA
Vol. 75, No. 1, pp. 50–53, January 1978
Biochemistry

Structure of the origin of DNA replication of bacteriophage fd

(initiation of DNA replication/RNA polymerase binding site/nucleotide sequence/hairpin structures in single-stranded DNA)

CHRISTOPHER P. GRAY*, REINHOLD SOMMER, CHRISTINE POLKE†, EWALD BECK, AND HEINZ SCHALLER

Mikrobiologie der Universität Heidelberg, Im Neuenheimer Feld 230, 6900 Heidelberg, West Germany

Communicated by Norton D. Zinder, September 22, 1977

ABSTRACT **An RNA-polymerase-protected DNA fragment of 125 nucleotides from the origin of single-strand to double-strand replication of bacteriophage fd (ori-DNA) was located on the physical map of the phage genome. A stretch of 187 base pairs of DNA including the ori-DNA was sequenced. This DNA segment contains regions with a highly asymmetric pyrimidine/purine distribution next to regions with 2-fold symmetry that form stable hairpin structures in the viral DNA strand.**

The filamentous bacteriophages fd, M13, and f1 contain a small, circular, single-stranded DNA of about 2×10^6 molecular weight (6400 nucleotides). Conversion of the viral DNA into the double-stranded replicative form as well as the replication of both strands of this DNA molecule are initiated and terminated at the same site on the phage genome (1–3), the origin of DNA replication. Detailed *in vitro* studies of the initiation reaction with the viral DNA as a template have indicated that the origin serves on this DNA strand as a signal for *Escherichia coli* RNA polymerase to synthesize a short RNA primer (4, 5). A single-strand specific *E. coli* DNA binding protein (6) has been implicated as an essential auxiliary protein for specific recognition of the origin structure (7, 8). In a previous study (8) we isolated and characterized a 120-nucleotide fragment of the origin region from fd (ori-DNA) as an RNA polymerase binding site that is protected by the enzyme against nuclease attack (8). We now report the nucleotide sequence of the ori-DNA fragment.

MATERIALS AND METHODS

Isolation of fd [^{32}P]DNA (8), pyrimidine tract analysis (8), separation of DNA fragments (8), and oligonucleotide $T_{11}C_9$-primed DNA synthesis (9) have been described. Restriction nucleases used were *Hpa* II, *Hae* III, *Hin*fI, *Alu* I, and *Hha* I prepared as described (10). DNA was cleaved in 10 mM Tris·HCl, pH 8/5 mM $MgCl_2$/ 0.5 mM dithiothreitol/50 mM NaCl/5% glycerol. DNA fragments were separated in slab gels of 6% polyacrylamide with 45 mM Tris·borate, ph 8.3/2.5 mM EDTA as buffer.

Synthesis of fd Replicative Form (RF) DNA *In Vitro*. fd DNA (2 nmol) and oligonucleotides from fd RF DNA (chain-length, 10–30; about 10 nmol) were annealed at 45° for 10 min in 10 ml of 0.1 M Tris·HCl/200 mM KCl (11). The mixture was cooled to 20° and the volume was increased to 20 ml with the addition of the four dNTPs (one radiolabeled) to 300 μM each, ATP to 50 μM, $MgCl_2$ to 7 mM, *E. coli* DNA polymerase I to 40 units/ml, and T4 DNA ligase to 1 unit/ml. DNA synthesis at 20° was followed by determining acid-insoluble radioactivity in small aliquots. Formation of covalently closed fd RF DNA was assessed by the fraction of labeled DNA molecules that was not denatured at pH 12.2 and not retained on nitrocellulose filters (see below). When the latter assay indicated 70–80% of closed fd RF DNA (2–4 hr), the reaction was terminated by addition of EDTA to 50 mM, Tris base to 0.1 M, and 10 ml of phenol. DNA was precipitated from the aqueous phase by addition of sodium acetate to 0.1 M and isopropanol (2.5 vol); after 2 hr at −20°, it was collected by centrifugation.

The sediment was dissolved in 4 ml of 10 μM Tris/0.1 mM EDTA/0.5 M NaCl; 0.5-ml portions were diluted at 0° with 1 ml of denaturation buffer (25 mM KPO_4/10 mM Tris/0.1 mM EDTA/0.9 M NaCl, titrated to pH 12.2 with KOH), after 1 min neutralized by the addition of 1 ml of 70 mM KH_2PO_4, and immediately filtered through two nitrocellulose filters (Sartorius MF 15; diameter, 50 mm). The filters were washed with 1 ml of 0.5 M NaCl. The combined filtrates were pooled, concentrated on a flash evaporator to about 2 ml, and desalted on a column of Sephadex G-50 (70 ml, 38 cm) in 10 mM Tris, pH 8.0/0.1 mM EDTA. The DNA concentration was determined spectrophotometrically with a conversion factor of 12.5 pmol of fd RF per A_{260} unit. Yields of isolated RF DNA varied from 50 to 80% of the input fd DNA; the fraction of the product consisting of covalently closed molecules was >98%.

Oligonucleotides from fd RF DNA. fd RF DNA (about 1 mg/ml) was degraded to about 50% acid solubility by incubation with pancreatic DNase (5 μg/ml) at 37° in 50 mM Tris·HCl, pH 8.0/5 mM $MgCl_2$/0.5 mM $CaCl_2$. EDTA was added to 20 mM and the nuclease was inactivated at 100° for 5 min.

DNA Sequencing. Synthesis of [γ-^{32}P]ATP and labeling of DNA fragments were carried out essentially as described (12). The reagents used were dimethyl sulfate (G), and hydrazine (C + T, C) (11), and 66% formic acid/2% diphenylamine (G + A) (13). Incubation with the latter reagent was for 10 min at 20°.

RESULTS AND DISCUSSION

Fine Mapping of the ori-DNA. Hybridization of the isolated ori-DNA with restriction fragments had shown that this DNA fragment was contained within *Hpa*-H, a restriction fragment of about 380 base pairs (8). For more precise mapping, shorter subfragments of the *Hpa*-H fragment and of adjacent DNA were compared to the ori-DNA by fingerprint analysis of the respective oligopyrimidine tracts.

fd viral strand [^{32}P]DNA was converted into its double-stranded form by *in vitro* repair synthesis using *E. coli* DNA polymerase I and then cleaved with *Hpa* II plus either *Hae* III or *Alu* I. The DNA fragments were separated on polyacrylamide gels, and those that map within or close to *Hpa*-H (Fig. 1) were analyzed by standard fingerprinting techniques. The results obtained with the *Hpa*/*Hae* restriction fragments

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

Abbreviations: ori-DNA, DNA fragment from the origin of DNA replication; RF, replicative form DNA.

* Present address: Department of Biology and Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139.
† Present address: Institut für Immunologie, Deutsches Krebsforschungszentrum, 6900 Heidelberg, West Germany.

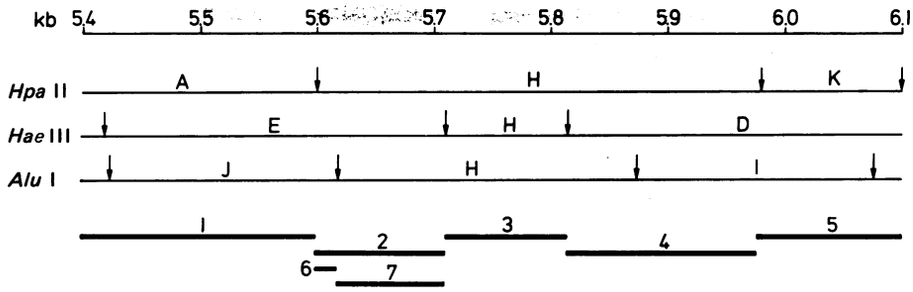

FIG. 1. Cleavage map of fd DNA at the origin of DNA replication for three restriction nucleases. Map distances are given in kilobases (kb) in 5′ to 3′ polarity from the single *Hin*dII cleavage site and assuming a total chainlength of 6400 nucleotides. Fragments 1–7 were isolated from uniformly labeled fd [^{32}P]DNA and used for the pyrimidine tract analysis described in Table 1.

(fragments 1–5) are listed in Table 1. They demonstrate that virtually all pyrimidine tracts typical for the ori-DNA (7) are contained in a 110-nucleotide subfragment *Hpa*-H/*Hae*-E (fragment 2) except for the additional nucleotide T_3C_5 and the missing nucleotide T_5C. The latter is present in the 3′-proximal fragment 3 (*Hae*-H). The nucleotide T_3C_5 was detected in the 17-nucleotide fragment 6 (*Hpa*-H/*Alu*-J) and therefore is located at the 5′ end of the *Hpa*-H fragment. These data place the ori-DNA in the left third of the *Hpa*-H fragment, starting from the *Alu* J/H junction (position 5620) and extending into the *Hae*-H fragment to about position 5750. The same conclusion was also obtained from analysis of restriction fragment *Hae*-E/*Alu*-H (fragment 7 in Fig. 1).

Nucleotide Sequences. The nucleotide sequence of the ori-DNA and of adjacent regions was determined by using the 5′-terminally labeled polynucleotides shown in Fig. 2. Most of these (a–g) were obtained by secondary cleavage of 5′-labeled restriction fragments from fd RF; polynucleotide f was obtained by using the oligonucleotide $T_{11}C_9$ (14) as a primer for site-specific repair synthesis (9). This oligonucleotide maps in the fd viral DNA just ahead of the 5′ end of fragment *Hpa*-H.

Table 1. Pyrimidine tract analysis of DNA fragments

Pyrimidine tract	Relative molar yields						
	1	2	3	4	5	o	s
$T_{11}C_9$	0.95	—	—	—	—	—	—
T_6C	—	—	—	1.2	—	—	—
T_5C_2	—	—	—	0.97	0.87	—	—
T_5C	—	—	1.6	0.77	—	0.8	1
T_5	—	—	—	1.8	—	—	—
T_4C_3	—	0.6	—	—	—	0.9	1
T_4C_4	0.9	—	—	—	—	—	—
T_4	—	—	—	—	0.88	—	—
T_3C_5	—	0.5	—	—	—	<0.3	—
T_3C_3	—	—	1.13	—	—	<0.3	—
T_3C_2	—	—	0.9	—	1.67	—	—
T_3C	2.0	1.25	—	—	—	0.95	1
T_3	0.65	2.3	—	9.1	0.7	1.95	2
T_2C_4	0.6	—	—	—	0.7	—	—
T_2C_2	—	2.17	1.15	—	—	1.9	2
T_2C	0.7	2.6	—	1.88	1.0	2.15	2
T_2	3.4	—	1.0	3.9	2.0	0.2	—
TC_4	—	0.7	—	—	—	0.65	1
TC_3	1.27	1.2	1.08	—	0.74	1.4	2
TC_2	0.62	—	4.2	—	1.1	0.1	—
TC	1.57	2.2	3.2	4.87	2.75	2.6	3
T	9.6	5.1	4.4	5.82	8.6	6.0	6
C_2	2.9	—	0.8	—	1	—	1
C_1	23.6	5.0	6.9	8.23	3.4	3.8	4
P_i	22.1	20.6	13.6	37	14.9	34	34
Chain length	190	112	103	166	125	~120	125

The distribution of pyrimidine tracts (8) in the viral strand of fd RF DNA fragments was determined as described in the *text* and normalized to the size of each restriction fragment. Numbering and map positions of the fragments are as given in Fig. 1. Pyrimidine tract analysis of the ori-DNA fragment (8) is included for comparison (o), as well as the theoretical yields calculated from the nucleotide sequence given in Fig. 3 (s).

DNA chains a–f were partially degraded by base-specific chemical reactions, and the resulting mixtures of fragments were resolved by gel electrophoresis on 20% acrylamide gels as described by Maxam and Gilbert (12). Sequences were confirmed by analysis of the complementary DNA strand or by the use of the plus–minus method of Sanger and Coulson (15) (data not shown). The results obtained are summarized in Fig. 3, which shows the sequence of the whole region consisting of 187 base pairs. This sequence is consistent with the pyrimidine tract distribution in the ^{32}P-labeled restriction fragments of the viral DNA strand (Table 1) and of the complementary strand (not shown) and also with the order of pyrimidine tracts next to the 3′ end of the $T_{11}C_9$ primer: $T_{11}C_9$-C_2-C_1-T_3C_3-T_3C_5-TC-T_2C_2-TC-T_4C_3-T_2C_2-(T-T_3C)-(C-TC_3)-TC_4-T_2C- (ref. 9; unpublished data). Furthermore, all recognition sequences for restriction nucleases have been confirmed by cleavage with the appropriate enzyme.

The ori-DNA fragment is located within the sequenced DNA segment by the position of its pyrimidine tracts (Table 1) which coincide exactly with those of a stretch of about 125 nucleotides of the viral DNA strand. The 5′ terminus of the ori-DNA is marked by the absence of oligonucleotide T_3C_5 and by the presence in twice molar amounts of T_2C_2 and TC. The 3′ end is located between nucleotides T_5C and T_3C_3, the latter occurring only as a weak spot in fingerprints of the 120-nucleotide ori-DNA fragment (8). It should be noted that both outside oligopyrimidines, T_3C_5 and T_3C_3, were present in larger precursor fragments of the ori-DNA and that the spot for 3′-terminal T_5C was missing in fingerprints from somewhat shortened ori-DNA fragments that were isolated after extended nuclease digestion (ref. 8; unpublished data).

Hairpin Structures in the ori-DNA. A rather compact structure by base-pairing of self-complementary nucleotide sequences in the ori-DNA had been indicated by its enhanced mobility during gel electrophoresis and by its high resistance to the action of single-strand specific *E. coli* DNA binding protein or nucleases (8). Furthermore, a stable hairpin structure has been located in fd DNA at the map position of the origin by electron microscopy after photochemical crosslinking (16). Fig. 4 shows the most stable of several possible secondary structures of the ori-DNA viral strand. It consists of two big hairpins of 55 and 60 nucleotides, respectively, that comprise most of the ori-DNA sequence. A short spacer of 13 nucleotides between these structures matches also fairly well with a nucleotide sequence beyond the 3′ end of the ori-DNA segment.

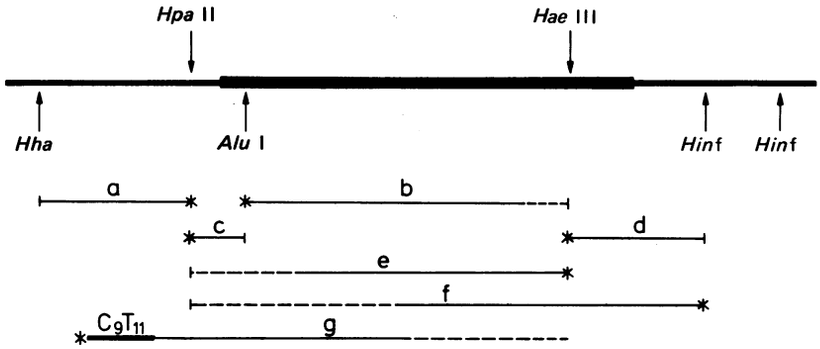

FIG. 2. Labeled DNA strands used for sequence analysis. The position of ori-DNA is indicated by the bold line in the cleavage map. Pertinent sites for cleavage by restriction enzymes are shown. Polynucleotides a–g were obtained by secondary cleavage of 5′-labeled restriction fragments (12) and fragment f, by elongation of 5′-^{32}P-labeled C_9T_{11} primer (9). An asterisk indicates the 5′-^{32}P terminus; broken lines indicate the non-sequenced parts of DNA strands.

However, this outside sequence appears not to be stably base-paired to the ori-DNA·RNA polymerase complex because it does not resist exonuclease I digestion during isolation of the ori-DNA as a protected fragment. This does not exclude contacts of the enzyme with this "outside region," because single-stranded DNA is not protected against exonuclease I degradation when complexed with RNA polymerase (8).

The proposed secondary structure is supported by the observation that the isolated ori-DNA fragment is quite resistant to nuclease S1 and is cleaved into defined substructures on prolonged incubation (8). Furthermore, a close relationship exists between the ori-DNA and fd DNA core, a mixture of several hairpin fragments that can be isolated from fd DNA by digestion with single-strand specific nucleases (17). Two major fd DNA core fragments map at the positions of the two hairpin structures in the ori-DNA and show oligopyrimidine fingerprints that fit the nucleotide sequence of the ori-DNA hairpins (C. P. Gray and E. Beck, unpublished data).

Recently, two highly S1-resistant fragments have been isolated from the fd DNA core and have been identified by DNA sequencing as subfragments of the 5′-proximal ori-DNA hairpin structure (C. P. Gray, unpublished data). Cleavage by nuclease S1 occurred at the two positions of this hairpin where a mismatched base pair interrupt the double-stranded stem (Fig. 4). This result provides direct evidence for the presence of a dominant hairpin structure in the origin region of fd DNA.

Further Discussion. The present structural analysis of the origin of DNA replication has centered on the ori-DNA, a DNA fragment of 125 nucleotides isolated previously from fd DNA as a specific complex with RNA polymerase (and possibly *E. coli* DNA binding protein). A DNA segment of 187 base pairs from fd RF DNA was sequenced comprising the ori-DNA and the region where synthesis of the RNA primer and the elongation of the DNA chain start and terminate.

The nucleotide sequence that we have determined reveals two unusual features: (*i*) extended self-complementary sequences that result in stable hairpin structures in the DNA single-strand, and (*ii*) long runs of pyrimidine nucleotides, located mainly outside of the self-complementary nucleotide sequences. These features are also present in the nucleotide sequences from the origin of DNA replication of plasmid Col E1 (18) and of bacteriophage λ (19) as well as in promoters (20) and terminators of transcription. This suggests a potential mechanism for activation of the origin in double-stranded DNA by destabilization of the double helix and by conformational changes—e.g., into hairpin structures (21). At the origin of fd these may be triggered by the nick introduced into the viral DNA strand by gene II protein, a viral initiation factor for phage DNA replication. Evidence for the presence of hairpin structures next to chain interruptions is indicated by the unusual electrophoretic mobilities of double-stranded DNA fragments from the fd origin (7), which sometimes can be resolved into two species of identical nucleotide sequence that migrate differently on polyacrylamide gels (unpublished data).

An alternative mechanism in which RNA polymerase recognizes the origin in its double-stranded form as a promoter seems unlikely because none of the six major (22) or the minor RNA polymerase binding sites from fd RF DNA maps at the origin. Furthermore, the origin does not contain the nucleotide sequences that have been implicated in promoter recognition and binding (23, 24).

In single-stranded fd DNA, recognition of the origin seems to be mainly determined by the high secondary structure of the origin region, which prevents it from being covered by the *E.*

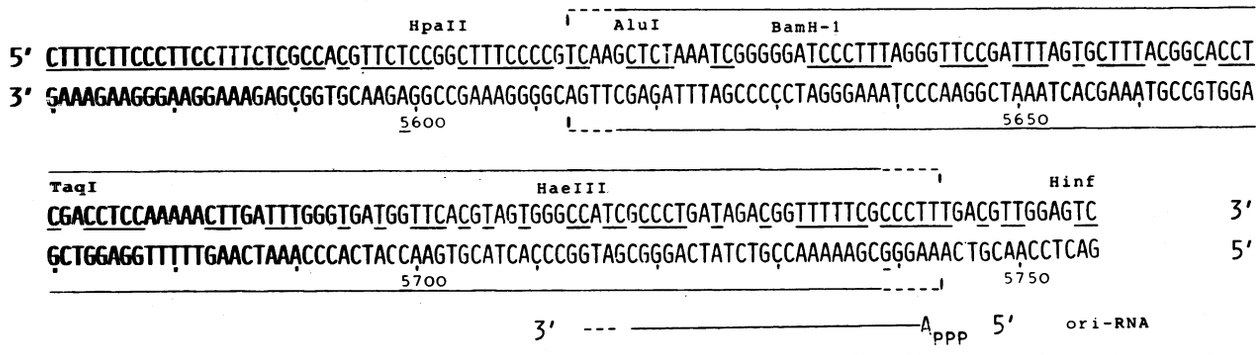

FIG. 3. Nucleotide sequence of the origin region. A, G, T, and C represent the dNMP residues. Cleavage sites by restriction nucleases are indicated. The upper line gives the sequence from the viral DNA strand; the pyrimidine tracts are underlined. Numbers represent map positions relative to the single *Hin*dII cleavage site in fd RF DNA. The ori-DNA sequence is boxed, and the position and polarity of the primer RNA are indicated.

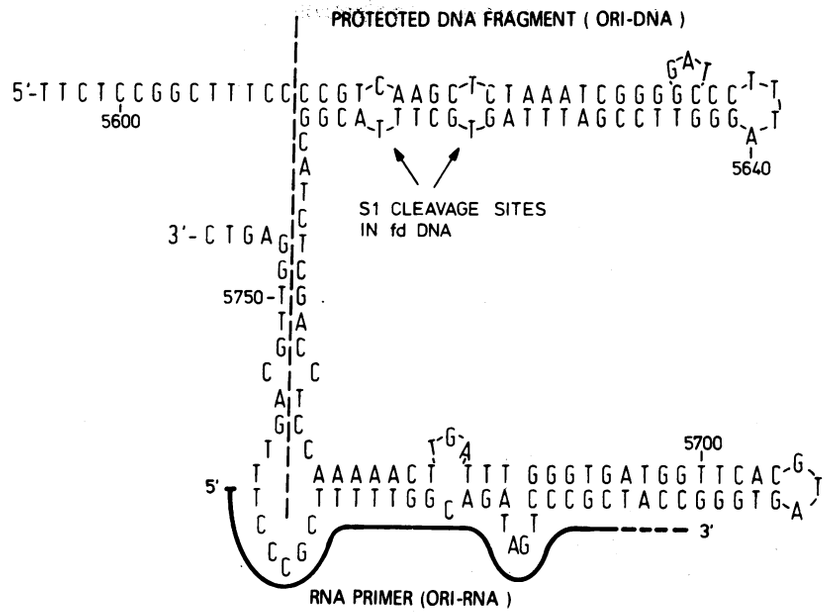

FIG. 4. Secondary structure of fd viral DNA at the origin of DNA replication. Nucleotides from map positions 5596–5756 are folded to give the highest degree of base-pairing, and the structure is kinked arbitrarily at the ends of the two main hairpin structures. The ends of two hairpin structures that have been isolated from fd DNA after extensive digestion with the single-strand specific nuclease S1 are indicated by the arrows. The region that is protected against nuclease digestion in the preinitiation complex (ori-DNA) is bordered by the dashed line. The sequence of the RNA primer (ori-RNA) is indicated by the bold line.

coli DNA binding protein. Discrimination against other hairpin structures in the viral DNA—e.g., which map at positions 500 and 5500 (C. P. Gray, unpublished data)—may be due to the particular high stability of the hairpins at the origin or to a simultaneous interaction of the RNA polymerase with both hairpins. This could explain why the ori-DNA is much larger than the 40 base pairs protected by the RNA polymerase in promoter complexes in double-stranded DNA (25).

The data presented in this paper do not allow any final conclusions as to the functional role of all parts of the ori-DNA. The 3′-terminal segment codes for the primer RNA (ori-RNA), which initiates at and terminates close to the single-stranded loops at the ends of the base-paired stem of the 60-nucleotide hairpin. This has been inferred from the fingerprint analysis of the ori-DNA and from the restriction mapping of DNA chains primed by the ori-RNA on the fd viral DNA strand (5).

The 5′-proximal hairpin seems to be of minor importance for the initiation reaction. This is also indicated by experiments that show that integration of nonviral DNA into the *Bam*HI site in map position 5630 does not abolish viral DNA replication (R. Herrmann, personal communication). However, the facts that the 5′-terminal ori-DNA segment forms the most stable hairpin in fd DNA, that its secondary structure is conserved at the origin of the related bacteriophage f1 (26), and that it is protected against nuclease digestion suggests that it is part of the preinitiation complex.

We are grateful to A. Maxam and W. Gilbert for providing detailed protocols of their DNA sequencing methods prior to publication. We also thank the Institut für Virusforschung, Deutsches Krebsforschungszentrum Heidelberg, for hospitality and the Deutsche Forschungsgemeinschaft for support.

1. Tabak, H. F., Griffith, S., Geider, K., Schaller, H. & Kornberg, A. (1974) *J. Biol. Chem.* **249,** 3049–3054.
2. Horiuchi, K. & Zinder, N. D. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 2341–2345.
3. Suggs, S. V. & Ray, D. S. (1977) *J. Mol. Biol.* **110,** 147–163.
4. Wickner, W., Brutlag, D., Schekman, R. & Kornberg, A. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 965–969.
5. Geider, K., Beck, E. & Schaller, H. (1978) *Proc. Natl. Acad. Sci. USA* **75,** in press.
6. Sigal, N., Delius, H., Kornberg, T., Gefter, M. L. & Alberts, B. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 3537–3541.
7. Geider, K. & Kornberg, A. (1974) *J. Biol. Chem.* **249,** 3999–4005.
8. Schaller, H., Uhlmann, A. & Geider, K. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 49–53.
9. Oertel, W. & Schaller, H. (1972) *FEBS Lett.* **27,** 316–319.
10. Roberts, R. J., Breitmeyer, J. B., Tabachnik, N. F. & Meyers, P. A. (1975) *J. Mol. Biol.* **91,** 121–123.
11. Heyden, B. (1970) Diplomarbeit, Universität Tübingen.
12. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
13. Burton, K. (1967) in *Methods in Enzymology,* eds. Grossman, L. & Moldave, K. (Academic Press, New York), Vol. 12A, pp. 222–224.
14. Ling, V. (1972) *J. Mol. Biol.* **64,** 87–102.
15. Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* **94,** 441–448.
16. Shen, C-K. J. & Hearst, J. E. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 2649–2653.
17. Schaller, H., Voss, H. & Gucker, S. (1969) *J. Mol. Biol.* **44,** 445–458.
18. Tomizawa, J., Ohmori, H. & Bird, R. E. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 1865–1869.
19. Denniston-Thompson, K., Moore, D. D., Kruger, K. E., Furth, M. E. & Blattner, F. R. (1977) *Science,* in press.
20. Schaller, H., Gray, C. & Herrmann, K. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 737–741.
21. Gierer, A. (1966) *Nature* **212,** 1480.
22. Seeburg, P. H. & Schaller, H. (1975) *J. Mol. Biol.* **92,** 261–277.
23. Pribnow, D. (1975) *J. Mol. Biol.* **99,** 419–443.
24. Seeburg, P. H., Nüsslein, C. & Schaller, H. (1977) *Eur. J. Biochem.* **74,** 107–113.
25. Heyden, B., Nüsslein, C. & Schaller, H. (1972) *Nature New Biol.* **240,** 9–12.
26. Ravetch, J. V., Horiuchi, K. & Zinder, N. D. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4219–4222.