Volume 12 Number 10 1984 Nucleic Acids Research

Initiation and termination signals for transcription in bacteriophage M13

Mari A.Smits, Josephine Jansen, Ruud N.H.Konings and John G.G.Schoenmakers

Department of Molecular Biology, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands

Received 9 April 1984; Accepted 27 April 1984

ABSTRACT

Transcription of the infrequently expressed phage M13 genome domain, comprising genes III, VI, I and IV, has been studied in detail by hybridization and S1-nuclease mapping studies. The contiguous genes III and VI are transcribed via an 1800 nucleotide-long RNA molecule that is initiated at a promoter which overlaps with the Rho-independent termination signal between genes III and VIII. Its synthesis is terminated at a Rho-dependent terminator in the proximal part of gene I. Transcription of gene I is not mediated by an independent promoter but most probably by read-through of RNA-polymerase through this terminator. Transcription of gene IV is accomplished by synthesis of four distinct RNAs of about 1500 to 1680 nucleotides long which are initiated at a promoter located immediately in front of gene IV. Termination of these transcripts is generated at at least four different sites located in tandem within the intergenic region between genes IV and II.

INTRODUCTION

The circular genome of the F-specific filamentous coliphages M13, fd and f1(Ff) can be divided into two domains which show widely disparate transcriptional activities (1,2). The first domain, encompassing the genes II, X, V, VII, IX and VIII, is transcribed very frequently. It is separated from the second domain (genes III, VI, I and IV) by a Rho-independent transcriptional terminator T(0.25) located immediately distal to gene VIII on one side and at the other side by the intergenic region (IG) in which a packaging signal for phage morphogenesis and the replication origins for (+) and (-) strands are located (Fig. 1).

Transcription of the first domain and the regulatory principles involved have been studied in large detail (1-8, Smits et al., in preparation). Detailed information with regard to the mechanism(s) of expression of the infrequently transcribed, second domain is still lacking.

Neither in minicells containing Ff DNA nor in Ff-infected E.coli cells RNA transcripts have been found which cross the IG (1,2,6-8). Transcripts made in vitro do cross the IG although on the basis of in vitro transcription studies

in the presence of purified Rho-factor it has been hypothesized that a Rho-dependent termination signal is located in this region (9-11). Since in addition no RNA species were found which were able to direct the in vitro synthesis of both gene III- and gene IV-protein, it has further been suggested that still other transcription termination signals might be operating in this second domain (12). The occurrence of such termination signals is in agreement with our previous data (1) which indicated that the transcriptional activity of the regions immediately distal to genes VIII, VI and IV is much lower than that of the regions located upstream these sequences. Since these data only relied on the quantitation of the RNA produced by different regions of the Ff genome and a detailed analysis of these RNAs has not yet been made, we have continued our studies towards the identification and characterization of the transcripts encoded by the "silent" second domain. In particular, we have concentrated our studies on the precise location of the sites where initiation and termination of transcription occurs. The studies described have been performed with the Ff bacteriophage M13.

MATERIALS AND METHODS

Bacterial strains and bacteriophages

Phage M13 is from our own laboratory stock. RNA was prepared from M13-infected cells of E.coli K38 or from minicells produced by M13-infected cells of E.coli strain N3026 (6,13). The conditionally-lethal Rho mutant E.coli AD 1600 (*his*$^+$,*gal*3, *rho*15, *str*^{r}) was kindly provided by Dr. S. Adhya (14).

Preparation of (^{3}H)-uridine labeled RNA

E.coli cells were grown at 37°C in TPGA-medium (5) to 2.10^8 cells/ml and infected with M13 phages at a multiplicity of infection of 40. After 10 min at 37°C, (^{3}H)-uridine (100 µCi/ml) was added and cultivation was continued for another 10 minutes. The procedures for the cultivation of the minicell-producing strain E.coli N3026 and the isolation of minicells have been described earlier (6). After resuspending in TPGA medium minicells were incubated at 37°C for 20 min in the presence of (^{3}H)-uridine. Extraction of intact RNA from infected cells or minicells was performed as described by Rivera et al. (5).

Mapping of transcripts

Heat-denatured M13 RF or restriction fragments of M13 RF were covalently coupled to freshly prepared discs of diazobenzyloxymethyl (DBM) paper as described by Alwine et al. (15). After an overnight incubation under R-loop

conditions (16) with heat-denatured (^{3}H)-uridine labeled RNA in 80% (v/v) formamide, 0.4 M NaCl, 1 mM EDTA, 0.04 M Pipes-NaOH, pH 6.4, the DBM filter discs were extensively washed at room temperature in 0.01 M Tris-HCl, pH 7.6, 1 mM EDTA (TE) containing 0.1% SDS. The annealed M13-specific RNA was eluted by heating at 90°C for 1.5 min in 200 µl of TE and was recovered by precipitation with ethanol. Analysis of the RNA was performed on 2.5% polyacrylamide gels containing 0.1% SDS and 7 M urea (5). After electrophoresis, the gels were prepared for fluorography (17), dried in vacuo and contacted with X-ray film at -70°C.

S1-nuclease mapping

M13 RNA was prepared from cells grown at 37°C in TPGA-medium to 2.10^{8} cells/ml, infected at a multiplicity of infection of 40 and harvested at 20 min post-infection. Restriction fragment probes were 5'-end labeled with (γ-^{32}P)-ATP and T4-polynucleotide kinase essentially as described by Maxam and Gilbert (18). The 3'-ends were labeled with (α-^{32}P)-dCTP and T4-polymerase as described by Donelson and Wu (19). End-labeled DNA was hybridized under R-loop conditions (16) with an excess of denatured RNA in 25 µl of 80% (v/v) formamide, 0.4 M NaCl, 1 mM EDTA and 0.04 M Pipes-NaOH, pH 6.4. After a 3 hrs incubation period the solution was diluted with 100 µl of water and 125 µl of 0.06 M sodium acetate, pH 4.5, 0.02 M $ZnSO_4$, 0.5 M NaCl, 10% glycerol and subsequently treated with S1-nuclease for 1 hr at 20°C. S1-nuclease-resistent hybrid molecules were recovered by precipitation with ethanol, dissolved in 5 µl of 90% (v/v) formamide, 1 mM EDTA, 0.03% bromophenolblue and heated for 1 min at 100°C. The protected probes were electrophoresed alongside high-resolution size markers on 6% polyacrylamide/7 M urea gels in TBE buffer. The size markers were obtained either by subjecting the same DNA probes, used for the hybridization reactions, to the sequencing reactions of Maxam and Gilbert (18) or by digestion of radio-labeled M13 RF with the restriction endonuclease HinfI.

RESULTS AND DISCUSSION

After infection of E.coli cells with Ff phages no switch-off of host RNA and protein synthesis occurs. A proper identification of the phage-encoded RNAs is therefore hampered by interference of the bulk of gene products, in particular the ribosomal RNAs and their precursors, encoded by the host cell genome. Such problems can be circumvented for a large part when E.coli minicells are used. Since in minicells carrying M13 RF virtually all the RNAs synthesized de novo are M13-specific (6) and the size and number of these

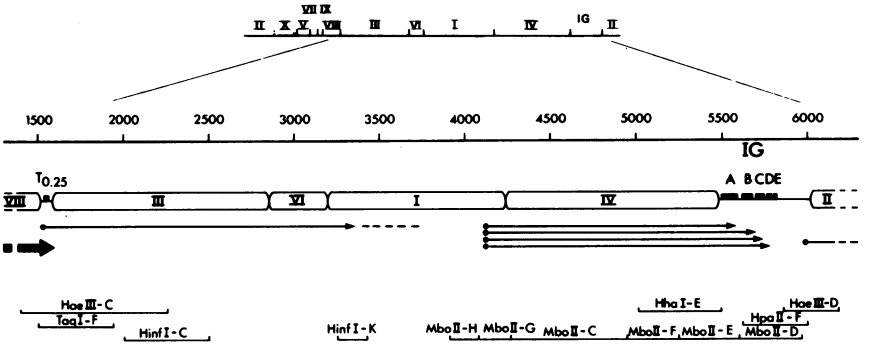

Fig. 1 Schematic drawing of the bacteriophage M13 genome. Indicated are the nucleotide positions (second line, 3), the genes (roman numerals), the Rho-independent terminator (TO.25), the intergenic region (IG) and the very stable hairpin structures (capital letters) present in the viral DNA (ΔG in kcal/mol, -64,5, -33.1, -26.0, -11.4 and -8.5, respectively). Arrows, dots and triangles denote the map positions of RNAs, promoters and termination sites, respectively. At the lower part the map positions of DNA probes, used in this study, are indicated.

RNAs are identical to the phage-specific RNAs synthesized in the infected E.coli cell, we have preferentially applied in this study (^{3}H)-uridine labeled RNA from minicells.

Hybridization of total RNA from minicells to M13 RF resulted in the retention of at least thirteen mRNA species, ranging in size from 370 (8S) nucleotides to 2000 (19S) nucleotides (Fig.2; lanes 1 and 9). The 8S, 9S, the four 11S RNAs (a-d), 13S (a-b), 14S and 19S, have already been characterized in detail (1, 5, Smits et al., in preparation). They constitute the family of primary transcripts and processed mRNAs derived from the first domain. The remainder, i.e. a 16S, one major 17S and several minor 17S RNAs and an 18S RNA are products of the second domain and have been studied in this report.

An 18S RNA of about 1800 nucleotides is selected from total RNA of minicells by DNA fragments which contain the Rho-independent termination signal T(0.25) (Fig.1) as well as by internal fragments of gene III (Fig. 2; lanes 3 and 4). This RNA is not bound to either DNA fragment TaqI-C, derived from the gene V-region (Fig. 2; lane 2) or to any other fragment containing nucleotide sequences located upstream of this terminator. More precise mapping studies indicated that the 5'-terminal end of this 18S RNA must be located somewhere in fragment TaqI-F (Fig. 1).

To map this 5'-terminal end at the nucleotide level S1-nuclease protection

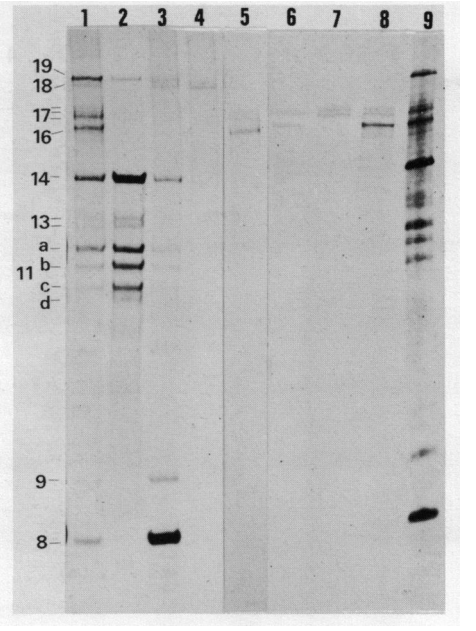

Fig. 2 Fluorograph of (^{3}H)-uridine labeled RNA species, isolated from minicells containing M13 RF, which hybridize to denatured M13 RF (lanes 1 and 9) and to the fragments TaqI-C (lane 2), HaeIII-C (lane 3), HinfI-C (lane 4), HhaI-E (lane 5), MboII-F (lane 6), MboII-D (lane 7) and MboII-E (lane 8), the origins of which are indicated in Fig. 1. The RNAs are denoted by their approximate S-values.

experiments were performed. S1-nuclease digestion of the hybrids formed upon annealing of the 5'-end labeled codogenic strand of fragment TaqI-F to RNA isolated from either infected cells or minicells resulted in the formation of one major S1-nuclease resistent molecule (Fig. 3A, lane 6). This band "a" maps the 5'terminal end of 18S RNA at nucleotide position 1544 (Fig. 3C, indicated by arrow a). Interestingly, two other hybrids became apparent upon digestion with higher S1-nuclease concentrations (Fig. 3A, lane 5). The cleavage site which generates band "b" is mapped at nucleotide position 1551 which is within the stem-loop junction of the terminator T(0.25) (Fig. 3C). Apparently, the latter structure has been formed within the hybridization probe (Fig. 3C, insert b). The cleavage site which generates band "c" coincides with the 3' terminal ends of the RNAs (8S RNA upto 19S RNA) encoded by the frequently transcribed domain of the M13 genome (1, 2, 5-8). Since 57 nucleotides of the DNA probe are complementary to the 3'-terminal end of these RNAs we assume that the protected band "c" is the result of incomplete

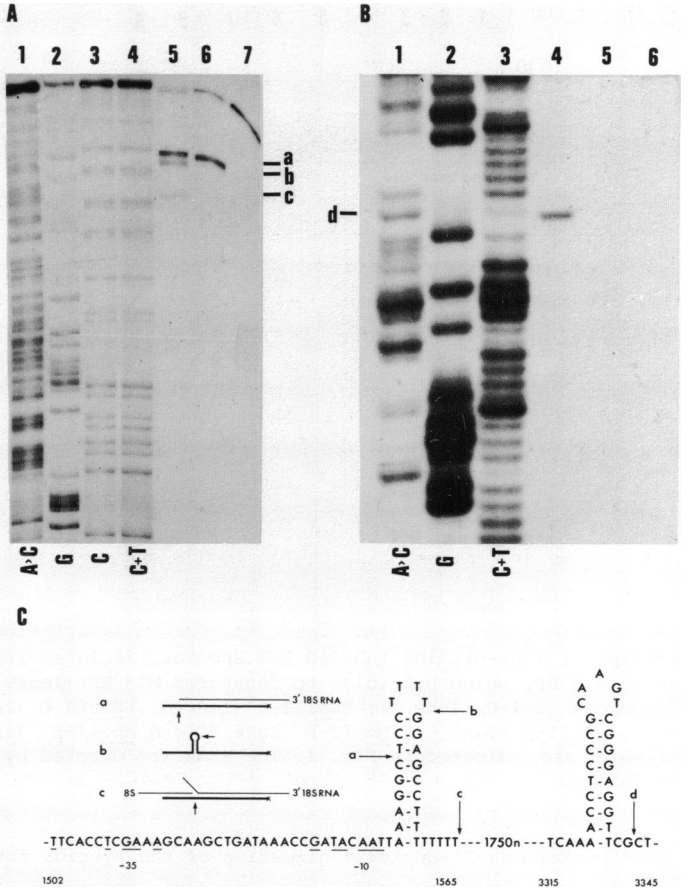

Fig. 3 S1-nuclease mapping of the 5'- and 3'-terminal end of 18S RNA.
Panel A; The codogenic strand of probe TaqI-F was labeled at its 5'-end and annealed to RNA isolated from infected cells. After hybrid formation the mixtures were treated with 1500 (lane 5), 250 (lane 6) and 0 (lane 7) units/ml of S1-nuclease. In lanes 1-4 the same probe was subjected to Maxam & Gilbert sequencing reactions (18).
Panel B; The codogenic strand of probe HinfI-K was labeled at its 3'-end and annealed to RNA isolated from infected (lanes 4 and 5) or uninfected cells (lane 6). After hybrid formation the mixtures were treated with 500 (lanes 4 and 6) or 0 (lane 5) units/ml of S1-nuclease. In lanes 1-3 the same probe was subjected to the Maxam & Gilbert sequencing reactions (18).
Panel C; Nucleotide sequence around the initiation and termination sites of 18S RNA. Homologies to the concensus E.coli promoter sequence (20, 21) are underlined and the S1-nuclease cleavage sites are indicated by arrows. Further explanations, see text.

hybridization of 18S RNA due to competition with these RNA species (Fig. 3C, insert c).

A startpoint of 18S RNA at nucleotide position 1544, as shown here, leads to the conclusion that this RNA is initiated at a promoter which partially overlaps the Rho-independent terminator T(0.25) (Fig. 3C). This conclusion is not only supported by the presence of a nucleotide sequence in front of this 5'-terminal end which is fairly homologous to the consensus of E.coli promoters (20, 21) but also by previous in vitro transcription studies (4, 22). Under the latter conditions RNA initiation at this site was only observed at high ratios of RNA polymerase to DNA (22, 23). Since the RNA initiation site is confined in the transcription termination signal T(0.25), it might well be that the activity of this promoter is influenced by the RNA polymerase molecules transcribing the DNA region immediately upstream this promoter.

From the position of its start site and the estimated length of about 1800 nucleotides we predicted that the 18S RNA is a bicistronic message covering both genes III and VI and that its 3'-terminal end is located within the HinfI-K sequence (Fig. 1). S1-nuclease protection experiments now show that its exact position is at nucleotide 3343 (Fig. 3B), which is immediately downstream a GC-rich region which possesses the intrinsic property to form a stable stem-loop structure (Fig. 3C). Since hairpin formation within a growing RNA chain causes RNA polymerase to pause (24, 25), this GC-rich region of dyad symmetry might function either as a pause site or a termination signal for transcription. We infer that the latter function is more likely. First, the transcriptional activity of the DNA region located immediately distal to this site is markedly lower that that of the region immediately upstream this sequence (1). Secondly, in the infected cell this 18S RNA can still be detected 10 min after the addition of rifampicin (unpublished data). Finally, the synthesis of 18S RNA can no longer be demonstrated under non-permissive conditions in E.coli cells carrying a conditionally-lethal mutation in the Rho gene (Fig. 4).

Genetic complementation studies have indicated that genes III, VI and I form one operon (26). As far as genes III and VI are concerned our finding of an 18S RNA transcript carrying both gene sequences is in accordance with this operon model. The detection of a larger RNA containing the genetic information of genes III, VI and I as well has been unsuccesful until now though La Farina and Model (2), studying transcription in f1-infected cells and using a slightly different approach, have recently succeeded in demon-

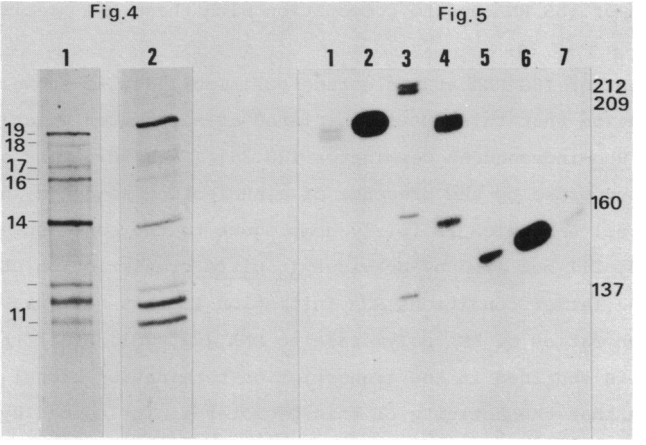

Fig. 4 Fluorograph of (^{3}H)-uridine labeled M13 specific RNAs synthesized in E.coli K38 (lane 1) and of the M13 specific RNAs synthesized under non-permissive conditions in E.coli AD 1600, bearing a conditionally lethal Rho-mutation (lane 2). Five min after infection with M13 phages at 32°C the cultures were shifted to 42°C and after another five min they were labeled with (^{3}H)-uridine for ten min.

Fig. 5 S1-nuclease mapping of the 5'-terminal end of 16S and 17S RNA. The 5'-terminal end labeled probes MboII-G (lanes 1, 2, 4) and MboII-H (lanes 5, 6, 7) were annealed to RNA isolated from infected (lanes 2, 4, 5, 6) or uninfected (lanes 1, 7) cells. After hybrid formation the mixtures were treated with 500 (lanes 1, 4, 5, 7) or 0 (lanes 2, 6) units/ml of S1-nuclease. The sizes of marker molecules (lane 3) are indicated.

strating the (weak) synthesis of such long transcripts. Their data together with ours suggest that expression of gene I is mediated by leakage of RNA polymerase through a transcription terminator which is Rho-dependent and which is located within the proximal part of gene I.

Hitherto, two protein factors (Rho and NusA) have been identified which participate with RNA polymerase in regulation of transcription termination (25). It is therefore of interest that we find a sequence CGCTCGTTA at 80 nucleotides upstream the GC-rich region of dyad symmetry and a similar sequence 330 nucleotides more upstream this region which are very identical to the sequence CGCTCTT(T)A, designated boxA, which is most probably required for NusA interaction with the transcription complex (24, 25).

Hybridization of total RNA from minicells to restriction fragments derived from the distal part of the second transcription domain resulted in the selection of one 16S RNA (1500 nucleotides), one major 17S RNA and several minor 17S RNAs varying in length from 1600 to 1680 nucleotides (Fig. 2,

lanes 5-8). These RNA species were also selected by DNA fragments (MboII-C, MboII-G) derived from the proximal part of gene IV but not by fragments containing the proximal part of gene I. These data therefore suggest that these RNA transcripts are initiated at a promoter immediately preceding gene IV. As shown in Fig 5, within the genome area covered by restriction fragments MboII-H and MboII-G only a single 5'-RNA terminus has been found by S1-nuclease protection experiments. It is located approximately 160 nucleotides upstream of the MboII-G/MboII-C junction. Precise size calibration of the S1-nuclease protected fragment showed that cleavage had occurred between nucleotide position 4105 and 4106.

The sequence in front of this site (indicated by an arrow) is:

(4059) AAATTCACTATTGACTCTTCTCAGCGTCTTAATCTAAGCTATCGCTATGTT (4111)

This sequence has many characteristics in common with E.coli promoter regions. The three bases of the Pribnow-sequence (centered around -10) which most pronouncedly affect promoter function, are conserved whereas at position -35 a sequence is found which is strikingly similar to the "recognition site" of most E.coli promoters (20, 21). From these data we infer that the 16S and 17S RNA species are initiated at the promoter indicated above which immediately preceeds the gene IV sequence. The estimate of 1500 to 1680 nucleotides for these RNA species is in agreement also with the expected capacity of these molecules to code for gene IV-protein of 426 amino acid residues in length.

The hybridization data presented in Fig. 2 and the observation that the 16S and 17S RNA species do not hybridize to fragment HaeIII-D (1) (Fig. 1) indicate that their 3'-terminal ends are located in the intergenic region (IG) between genes IV and II. This region comprises five well-defined palindromic sequences, denoted A through E (Fig. 1), which play an important role in DNA replication and phage morphogenesis (27-29). Since it has been demonstrated unambiguously that these palindromic sequences are present as very stable stem-loop structures in the single-stranded viral DNA (30), it is very likely that such structures are also formed in growing RNA chains. Consequently, they might cause RNA polymerase to slow down and/or terminate transcription. From Fig. 2 (lanes 5,6,8) it is evident that both the 16S and 17S RNAs are selected by fragments which contain the hairpin A sequence. Fragments which do not contain the latter sequence but which comprise the hairpins B through E (HpaII-F, MboII-D) do not hybridize to the 16S RNA but still bind the slightly longer 17S RNA species (Fig. 2, lane 7), suggesting that the 3'-terminal ends of 16S and 17S RNAs are located within or immedia-

tely downstream the hairpin A and B through D, respectively. Preliminary S1-nuclease protection experiments with 16S RNA and its selective DNA probes indicate that 16S RNA has a heterogeneous 3'-end. It is centered around nucleotide position 5566 which is in the 3'-interior loop of hairpin A (data not shown). Thus far we have not been able to locate the 3'-terminal ends of the various 17S RNAs. The complex banding patterns obtained has not allowed us yet to discriminate between products formed by termination events and those formed by secondary structures in both the RNA and the DNA hybridization probes.

In previous studies it has been demonstrated that transcripts made in vitro in the absence of Rho-protein cross the IG region whereas in the presence of Rho-protein transcription termination events are induced in this part of the genome (4, 9-11, 31). Since the studies presented here clearly demonstrate that very few if any transcripts cross the IG region in vivo, it is likely that the observed 3'-terminal ends of 16S and 17S RNA are generated by Rho-mediated transcription termination events. Basically similar conclusions were recently reached with respect to transcription termination in the IG region of the M13-related bacteriophage f1 (31).

REFERENCES

1. Smits, M.A., Schoenmakers, J.G.G. and Konings, R.N.H. (1980) Eur. J. Biochem. 112, 309-321
2. La Farina, M. and Model, P. (1983) J. Mol. Biol. 164, 377-393
3. Van Wezenbeek, P.M.G.F., Hulsebos, T.J.M. and Schoenmakers, J.G.G. (1980) Gene 11, 129-148
   Beck, E. and Zink, B. (1981) Gene 16, 35-58
   Hill, D.F. and Petersen, G.B. (1982) J. Virol. 44, 32-46
4. Konings, R.N.H. and Schoenmakers, J.G.G. (1978) in "Single-Stranded DNA Phages", D.T. Denhardt, D.T. Ray and D. Dressler Eds., 507-530
5. Rivera, M.J., Smits, M.A., Quint, W., Schoenmakers, J.G.G. and Konings, R.N.H. (1978) Nucl. Acids Res. 5, 2895-2912
6. Smits, M.A., Simons, G., Konings, R.N.H. and Schoenmakers, J.G.G. (1978) Biochim. Biophys. Acta 521, 27-41
7. Cashman, J.S. and Webster, R.E. (1979) Proc. Natl. Acad. Sci. USA 79, 1169-1173
8. Casman, J.S., Webster, R.E. and Steege, D.A. (1980) J. Biol. Chem. 225, 2554-2565
9. Edens, L. (1978) Ph.D. Thesis, University of Nijmegen, The Netherlands
10. Takanami, M. Okamoto, T. and Sugiuri, M. (1970) Cold Spring Harbor Symp. Quant. Biol. XXXV, 179-187
11. Takanami, M., Okamoto, T. and Sugiuri, M. (1971) J. Mol. Biol. 62, 81-88
12. La Farina, M. and Model, P. (1978) Virology 86, 368-375
13. Adler. H.I., Fisher, W.D., Cohen, A. and Hardigree, A.A. (1967) Proc. Natl. Acad. Sci. USA 57, 321-326
14. Das, A., Court, D. and Adhya, S. (1976) Proc. Natl. Acad. Sci. USA 73, 1959-1963
15. Alwine, J.C., Kemp. D.J., Parker, B.A., Reiser, J., Renart, J., Stark, G.R. and Wahl, G.M. (1980) in "Methods in Enzymology" (Grossmann, L.

and Moldave, K., eds.) 65, part I, 220-241
16. Casey, J. and Davidson, N. (1977) Nucl. Acids Res. 4, 1539-1552
17. Bonner, W.M. and Laskey, R.A. (1974) Eur. J. Biochem. 59, 83-88
18. Maxam, A.M. and Gilbert, W. (1980) in "Methods in Enzymology" (Grossmann, L. and Moldave, K. eds.) 65, part I, 499-560
19. Donelson, J. and Wu, R. (1972) J. Mol. Chem. 247, 4654-4660
20. Siebenlist, U., Simpson, R.B. and Gilbert, W. (1980) Cell 20, 269-281
21. Scherer, G.E.F., Walkinshaw, M.D. and Arnott, S. (1978) Nucl. Acids Res. 5, 3759-3773
22. Edens, L. Konings, R.N.H. and Schoenmakers, J.G.G. (1978) J. Virol. 28, 835-842
23. Schaller, H., Beck, E. and Takanami, M. (1978) in "Single-Stranded DNA Phages", D.T. Denhardt, D.T. Ray and D. Dressler, Eds., 139-163, Cold Spring Harbor Laboratory, New York
24. Platt, T. (1981) Cell 24, 10-23
25. Holmes, W.M., Platt, T. and Rosenberg, M. (1983) Cell 32, 1029-1032
26. Pratt, D., Tzagoloff, H. and Beaudoin, J. (1969) Virology 39, 42-53
27. Dotto, G.P., Enea, V.E. and Zinder, N.D. (1981) Virology 114, 463-473
28. Schaller, H., Uhlman, A. and Geider, K. (1976) Proc. Natl. Acad. Sci. USA 73, 49-53
29. Geider, K., Beck, E. and Schaller, H. (1978) Proc. Natl. Acad. Sci. USA 75, 645-649
30. Gray, C.P., Sommer, R., Polke, C., Beck, E. and Schaller, H. (1978) Proc. Natl. Acad. Sci. USA 75, 50-53
31. Moses, P.B. and Model, P. (1983) J. Mol. Biol. 172, 1-22