Eur. J. Biochem. 70, 577–587 (1976)

# Mapping of Promoter Sites on the Genome of Bacteriophage M13

Luppo EDENS, Peter VAN WEZENBEEK, Ruud N. H. KONINGS, and John G. G. SCHOENMAKERS

Laboratorium voor Moleculaire Biologie, Universiteit van Nijmegen

(Received June 4/September 13, 1976)

With the aid of transcription studies on restriction fragments of bacteriophage M13 DNA it has been demonstrated that at least eight promoter sites are located on the M13 genome. Five of these promoters initiate the synthesis of RNA chains which contain at their 5′-terminal end pppG (G promoters), while the other three promoters initiate RNA chains which start exclusively with pppA (A promoters). The positions of these promoter sites on the physical map are: 0.82 ($G_{0.82}$), 0.88 ($G_{0.88}$), 0.94 ($G_{0.94}$), 0.01 ($G_{0.01}$), 0.08 ($G_{0.08}$), 0.36 ($A_{0.36}$), 0.51 ($A_{0.51}$) and 0.56 ($A_{0.56}$). The G promoters were found to be clustered within a distance of one-third of the genome length from the central termination site for transcription (map position 0.77). The A promoters, however, were found at greater distances from this termination signal. Based upon the incorporation of [$\gamma$-$^{32}$P]ATP or [$\gamma$-$^{32}$P]GTP, the capacity of these promoters to initiate the synthesis of RNA chains varies. The strongest G promoters are $G_{0.82}$, $G_{0.94}$ and $G_{0.08}$ and the strongest A promoter is $A_{0.36}$.

As judged from their position on the genetic map, it is postulated that two promoters, namely $G_{0.94}$ and $G_{0.01}$, are located within gene II. The other promoters are most probably located immediately in front of the gene VIII/VII boundary ($G_{0.82}$), and immediately in front of gene V ($G_{0.88}$), gene II ($G_{0.08}$), gene IV ($A_{0.36}$), gene I ($A_{0.51}$) and gene VI ($A_{0.56}$). No evidence has been obtained so far for the existence of a promoter immediately in front of gene III.

The genome of the small filamentous bacteriophage M13 consists of a circular, single-stranded DNA ($M_r \approx 2 \times 10^6$) which, upon infection, is converted into a double-stranded, circular replicative form (RF) molecule (for a recent review, see [1]). During M13 growth only the nonviral strand of the RF molecule is transcribed by its host RNA polymerase [2–4]. By convention transcription proceeds counterclockwise on the genetic map and Roman numerals are used to designate the eight M13 genes [5–7] (Fig. 1).

In a previous communication [8] we have presented evidence that transcription *in vitro* of M13 replicative form I results in the formation of at least seven discrete RNA species ranging in size from about 8 S ($M_r \approx 1.1 \times 10^5$) up to 26 S ($M_r \approx 17 \times 10^5$). The two largest RNA species (23 S and 26 S) are initiated with pppA (A-start RNA) while the other five (8 S, 11 S, 14 S, 17 S and 19 S) are initiated with pppG (G-start RNA) [8]. Translation of these individual RNA species in a protein-synthesizing system revealed that they all share the coding information for the major capsid protein encoded by gene VIII (our unpublished results). From these results it was inferred that the synthesis of all these RNA species is terminated at a unique termination signal which is located immediately distal to the 3′ end of gene VIII. The latter conclusion has recently been substantiated by transcription studies *in vitro* on a DNA restriction fragment (fragment *Hap*II-$B_2$; Fig. 1) which previously has been shown to direct the synthesis of gene-VIII protein in a DNA-dependent protein-synthesizing system [8, 10].

The synthesis *in vitro* of a discrete number of RNA species implies that an equivalent number of RNA initiation sites (promoter sites) are located on the double-stranded replicative form molecule. The approximate positions of three of these promoters have already been mapped with the aid of coupled transcription and translation studies *in vitro* [7, 10].

*Abbreviations.* RF, double-stranded replicative form DNA; RF I, RF in which both strands are covalently closed; RF III, linear RF of full genome-length; *Hap*II, *Hae*III and *Hind*II refer to the DNA fragments produced by the restriction endonucleases from *H. aphirophilus* (endoR · *Hap*II), *H. aegyptius* (endoR · *Hae*III) and *H. influenzae* serotype d (endoR · *Hind*II).

*Enzymes.* RNA polymerase or nucleoside triphosphate RNA nucleotidyltransferase (EC 2.7.7.6); restriction endonucleases (EC 3.1.4.30).

*Note.* The direction of transcription proceeds counterclockwise on the genetic map. This is also the 5′ to 3′ polarity of the restriction fragments.

However, in order to map these and the other promoters more precisely we have studied RNA synthesis under the direction of various restriction fragments of M13 RF molecules. From the results of such experiments it became evident that only those fragments gave rise to the synthesis of a discrete RNA product which were presumed to contain a promoter site on the basis of their location on the physical map. The length of the RNA chain formed then was determining for the position of the promoter on this fragment.

It could be demonstrated that the number of promoter sites detected on restriction fragments and the ribonucleoside triphosphates used for initiation are consistent with the number and starting nucleotide sequences of the RNA chains synthesized on the intact M13 genome.

## MATERIALS AND METHODS

### *Enzymes*

*Escherichia coli* RNA polymerase holoenzyme was a generous gift from Dr R. Schilperoort (University of Leiden). The restriction endonucleases endoR · *Hind*II, endoR · *Hap*II and endoR · *Hae*III have been isolated as described previously [14–16].

### *Replicative Form I DNA and Restriction Fragments*

The method for the preparative isolation and purification of M13 replicative form I DNA has been described [14, 16]. Digestion of it with the restriction endonucleases was carried out under the conditions previously described [5]. After digestion, the fragments were separated by electrophoresis on a 3% discontinuous polyacrylamide slab gel [14] and further purified as described by van den Hondel *et al.* [5].

Full-length linear RF III molecules, produced by cleavage of M13 RF I with endoR · *Hind*II, was isolated by sucrose density gradient centrifugation [11].

### *RNA Synthesis* in vitro

RNA synthesis *in vitro* was performed in a standard reaction mixture (0.1 ml) which contained: 4 μmol Tris-HCl (pH 7.9), 15 μmol KCl, 0.8 μmol $MgCl_2$, 0.1 μmol dithiothreitol, 0.01 μmol EDTA, 0.1% Tween-80, 25 μg bovine serum albumin, 0.2 pmol M13 DNA restriction fragment and 8.0 pmol *E. coli* RNA polymerase holoenzyme. After a preincubation period of 5 min at 37 °C, ribonucleoside triphosphates were added to a final concentration of 80 μM, except for the ($\gamma$-$^{32}$P)-labeled triphosphate whose concentration was 6 μM. Initiation was allowed to proceed for 5 min at 37 °C after which time the ATP or GTP concentration was adjusted to 80 μM and rifampicin was added (final concentration 25 μg/ml). RNA synthesis was then continued for 5 min at 37 °C.

M13 RF I or full-length RF III molecules were transcribed under similar reaction conditions with the exception that the molar ratio of RNA polymerase to DNA template was 10.

After RNA synthesis, 10 μg of carrier tRNA was added and the reaction mixtures were extracted with phenol. The aqueous layer was passed through a Sephadex G-50 column (10 × 1 cm) and the RNA was precipitated twice with 2.5 vol. of ethanol at −20 °C. The RNA was collected by high-speed centrifugation (15 min at 150000 × *g*) and dissolved in 15 μl of deionized formamide.

### *Polyacrylamide Gel Electrophoresis*

The RNA species synthesized *in vitro* were analysed on vertical slab gels (16 × 14 × 0.2 cm), containing 3.8% polyacrylamide in 98% formamide [13]. After electrophoresis for 4 h at 30 mA the wet gel slab was exposed to X-ray film (Kodak RP/R54). The approximate lengths of the RNA species synthesized *in vitro* were estimated from their relative electrophoretic mobilities. As electrophoretic markers *E. coli* ribosomal RNAs and denatured restriction fragments of M13 RF were used. The chain lengths for the various RNAs was assumed to be 3000 nucleotides for 23-S, 1500 nucleotides for 16-S and 120 nucleotides for 5-S ribosomal RNA [17]. For the lengths of the restriction fragments the data given by van den Hondel and Schoenmakers [14] were used. On the gel system used no detectable differences were observed between the relative electrophoretic mobilities of the ribosomal RNAs and the restriction fragments of identical length.

### *RNA-DNA Hybridization*

The method for RNA-DNA hybridization was similar to the method of Gillespie and Spiegelman [18] as modified by Petterson *et al.* [19].

### *Nomenclature of Promoter Sites*

Promoters which initiate the synthesis of RNA chains which start with pppG are denoted by G, while those which initiate the synthesis of RNA which starts with pppA are denoted by A. The position of each promoter is given in map units and is indicated by a suffix corresponding to the position of the promoter on the physical map (*i.e.* $G_{0.82}$ means a promoter which is located at 0.82 map units on the physical map and the RNA of which is started with pppG).

## RESULTS

We reported previously [8] that transcription of circular M13 replicative form I by purified *E. coli* RNA polymerase holoenzyme results in the formation of at least two large RNA species (23 S and 26 S) which are initiated with pppA and five species (8 S, 11 S, 14 S, 17 S and 19 S) which are initiated with pppG. From their relative electrophoretic mobilities on the polyacrylamide gel one can calculate that these RNA species range in size from approximately 360 nucleotides for the smallest (8 S) up to about 5000 nucleotides for the largest (26 S) RNA product. Also the frequency of synthesis varies among the individual RNA species (Table 1). If it is taken into account that all these RNA species are terminated at the same site and that this central termination site is located immediately distal to gene VIII [8], then the approximate positions of initiation regions on the genetic and physical map can be determined. The positions estimated from transcription data on intact M13 RF molecules are presented in Table 1. Due to the inaccuracy of estimating sizes of RNA from their relative electrophoretic mobilities, especially in the high-molecular-weight region, the positions given are only fairly accurate for sites which initiate the relatively small RNA chains. In order to verify the existence of such promoter regions and to map these promoter sites more accurately we studied RNA synthesis under the direction of restriction fragments of M13 RF. The fragments used were the endoR · *Hae*III, the endoR · *Hap*II and the endoR · *Hind*II fragments, whose physical orders are illustrated in Fig. 1.

As shown in Table 2, RNA synthesized on restriction fragments with RNA polymerase holoenzyme did not hybridize to single-stranded phage DNA but did find complements with denatured M13 RF. This indicates that under the transcription conditions used the RNA is predominantly transcribed from the minus strand of the restriction fragments, *i.e.* the strand which is also used as a template during the tran-

Table 2. *Hybridisation of bacteriophage M13 RNA synthesized* in vitro

$^{32}$P-labeled RNA was synthesized *in vitro* either on M13 RF I or on RF which has been digested with the appropriate restriction endonucleases. About 0.04 μg RNA ($5.5 \times 10^4$ counts · min$^{-1}$) was incubated for 20 h at 37 °C with either heat-denatured M13 RF (2 μg), M13 ssDNA (1 μg), or denatured T7 DNA (2 μg) which were immobilized on nitrocellulose filters. The hybridisation buffer consisted of 50% formamide, 0.6 M NaCl, 0.05 M Tris-HCl, pH 7.4, 10 mM EDTA and 0.1% sodium dodecylsulfate. Non-hybridized RNA was removed from the filters by digestion with pancreatic ribonuclease and ribonuclease T1. Subsequently, the filters were washed several times with 0.3 M NaCl, 0.03 M sodium citrate buffer, dried, and counted in PPO/POPOP/toluene scintillation fluid

| Replicative form used for transcription | Amounts of [$^{32}$P]RNA hybridized to | | |
|---|---|---|---|
| | M13 RF | M13 viral DNA | T7 DNA |
| | % | | |
| Untreated M13 RF | 65 | 0.7 | 0.4 |
| EndoR · *Hap*II-treated RF | 51 | 2.2 | 0.6 |
| EndoR · *HaeIII*-treated RF | 53 | 1.3 | 0.4 |

Table 1. *Relative strengths and positions of the promoters on the M13 genome*

The RNA species synthesized *in vitro* under the direction of M13 RF have been designated 8-S, 11-S, 14-S *etc.* [8]. The sizes of these RNA chains were approximated from their relative electrophoretic mobilities on formamide slab gels with *E. coli* ribosomal RNA as markers. The lengths of the RNA species synthesized on restriction fragments of M13 RF were estimated similarly using *E. coli* ribosomal RNA and denatured restriction fragments as markers. The position of a promoter on a restriction fragment was calculated from the length of the RNA chain synthesized on this fragment, on the assumption that synthesis has ceased at its 3′ end. The length of the RNA chains synthesized on M13 RF finally has been calculated from the distance between the central termination site (map position 0.77 [8]) and the promoter site as determined by transcription of the various restriction fragments.. The strength of the various promoters was estimated from the $^{32}$P content of the RNA chains initiated at the respective promoter sites, using either [$\gamma$-$^{32}$P]ATP or [$\gamma$-$^{32}$P]GTP as the sole radioactive precursor during transcription. The relative strength of the G promoters and A promoters were evaluated and expressed as a fraction of the strength of promoter $G_{0.94}$ and $A_{0.36}$, respectively. n.d. = not determined

| RNA species | Promoter | Position of promoter as deduced from the length of RNA transcribed from | | | Estimated length of RNA chain (nucleotides) | Relative strength of promoter |
|---|---|---|---|---|---|---|
| | | RF I | RF III | restriction fragments | | |
| 8-S | $G_{0.82}$ | 0.80 | 0.80 | 0.82 | 360 | 0.50–0.80 |
| 11-S | $G_{0.88}$ | 0.90 | 0.90 | 0.88 | 700 | 0.05–0.15 |
| 14-S | $G_{0.94}$ | 0.95 | 0.95 | 0.94 | 1200 | 1.00 |
| 17-S | $G_{0.01}$ | 0.00 | 0.00 | 0.01 | 1600 | 0.10–0.15 |
| 19-S | $G_{0.08}$ | 0.10 | 0.10 | 0.08 | 2000 | 0.60–0.90 |
| 23-S | $A_{0.36}$ | 0.30 | 0.35 | 0.36 | 3900 | 1.00 |
| 26-S | $A_{0.51}$ | n.d. | 0.50 | 0.51 | 4800 | 0.15–0.30 |
| 26-S | $A_{0.56}$ | n.d. | 0.50 | 0.56 | 5100 | |

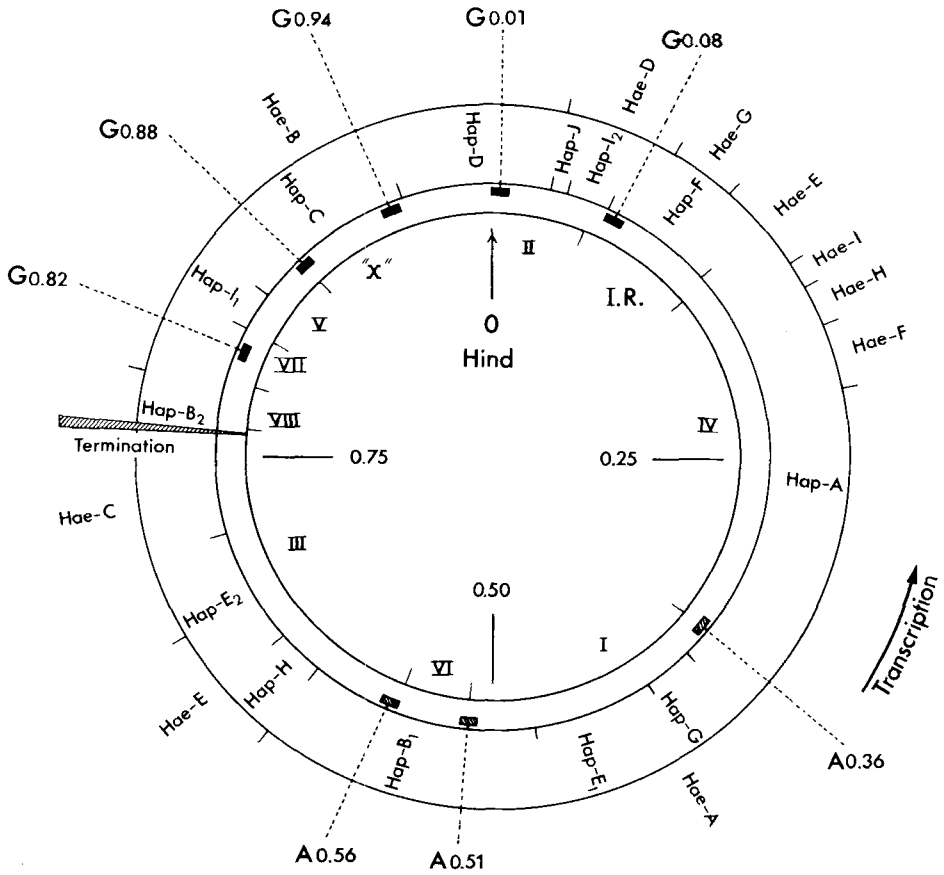


Fig. 1. *Localization of promoter sites on the genetic map and the physical maps of bacteriophage M13 DNA.* The inner circle represents the genetic map [5]. The middle and outer circles show the locations of the endoR · *Hap*II and the endoR · *Hae*III fragments [14]. The arrow indicates the cleavage site of M13 RF by endoR · *Hind*II [14]. The direction of transcription is counterclockwise around the genetic map [7]. The positions of the promoters discussed in this paper are indicated: G promoters with black bars, and A promoters with hatched bars. The position of the central termination site for transcription [8] is indicated. "X" refers to the C-terminal part of gene II in which an internal start of RNA synthesis is located. This RNA directs the synthesis of protein X [10, 15] (and Edens, L., Konings, R. N. H. & Schoenmakers, J. G. G., unpublished results). I.R. refers to the intergenic region between gene II and gene IV [15, 32] which contains the replication origin for parental M13 RF [23]

scription *in vivo* and *in vitro* of the intact M13 genome [2, 3]. Only a limited number of restriction fragments of M13 RF were able to direct the synthesis of a unique RNA product, as will be shown in the following paragraphs: namely those fragments which were presumed to contain a promoter on the basis of their location on the physical map (Table 1). These combined results indicate that proper strand selection and the specificity of RNA initiation is still retained after cleavage of M13 RF molecules into restriction fragments. The efficiency of initiation, however, appeared to be influenced by restriction endonuclease digestion.

## LOCATION OF A PROMOTERS

From their length in nucleotides it is to be expected that the synthesis of the ATP-initiated RNA chains starts at the map positions 0.3 (23-S) and 0.5 (26-S), respectively (Table 1 and Fig. 1). Fragments which encompass these sites and, hence, are presumed to contain these A-promoter regions are *Hap*II-A, *Hap*II-$B_1$, *Hae*III-A and RF III. The transcriptional results obtained with these fragments are described in the following sections.

### *Fragment* Hap*II-A*

Genetic analysis and coupled transcription and translation studies *in vitro* have demonstrated that the complete gene IV is located on fragment *Hap*II-A ($\approx$ 1530 base pairs long [5, 7]). Since the size of gene IV is almost equivalent to the size of fragment *Hap*II-A it was concluded that a promoter is located proximal to the 5′ end of this restriction fragment.

As shown in Fig. 2h, transcription of fragment *Hap*II-A in the presence of ($\gamma$-$^{32}$P)-labeled ATP and subsequent analysis of the products on formamide polyacrylamide slab gels revealed that this fragment directs the synthesis of a single RNA species which is approximately 1400 nucleotides long. Similar to fragment *Hap*II-A, this RNA species directs in a

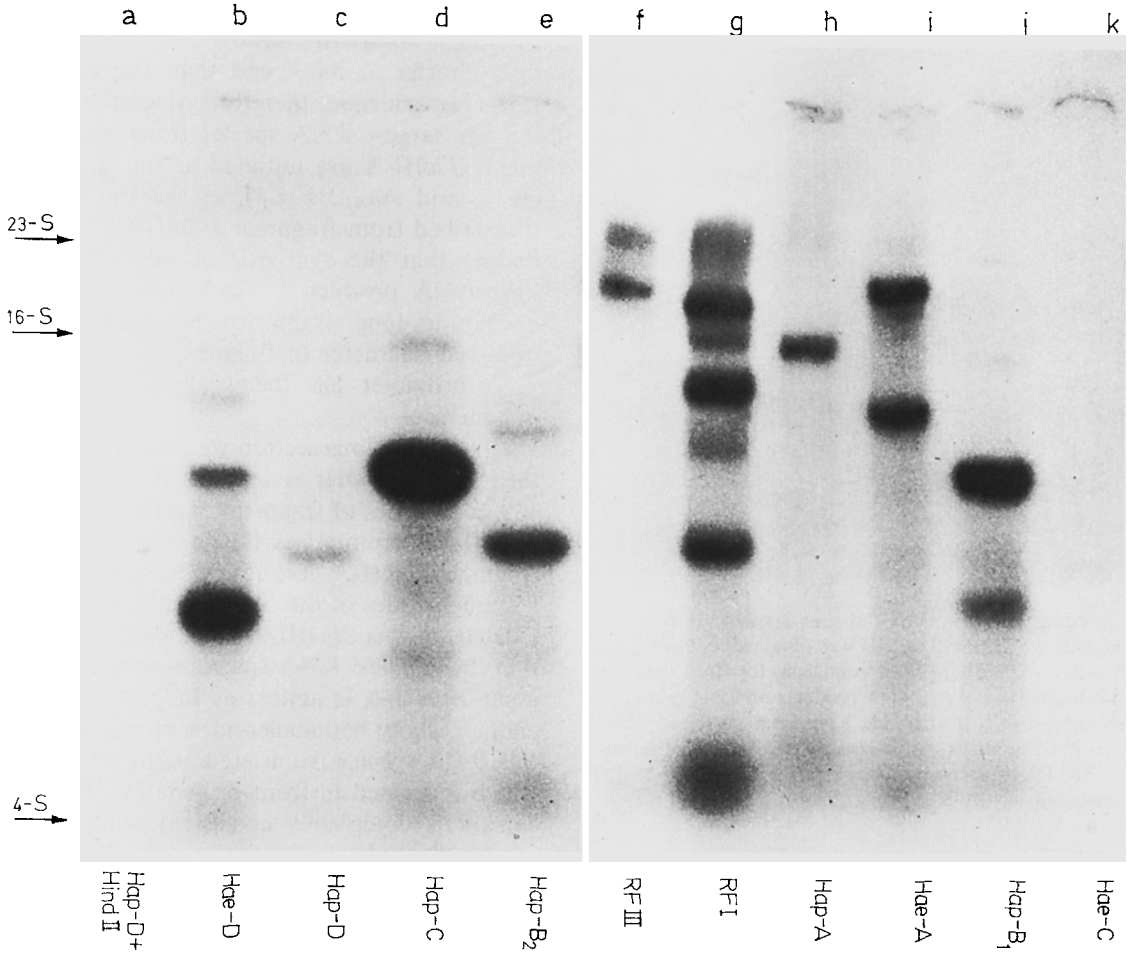


Fig. 2. *Autoradiograph of the RNA species synthesized under the direction of M13 RF I or under the direction of various restriction fragments of M13 RF.* RNA synthesis was carried out in the presence of [$\gamma$-$^{32}$P]ATP or [$\gamma$-$^{32}$P]GTP and was performed as described under Methods. The RNA products were analysed on vertical slab gels (16 × 14 × 0.2 cm) containing 3.8% polyacrylamide in 98% formamide [13]. The arrows indicate the positions of migration of the ribosomal RNA species isolated from *E. coli* cells. The RNA products were synthesized under the direction of: (a) *Hap*II-D which was split with endoR · *Hind*II, (b) *Hae*III-D, (c) *Hap*II-D, (d) *Hap*II-C, (e) *Hap*II-$B_2$, (f) RF III, *i.e.* RF which was split with endoR · *Hind*II, (g) RF I, (h) *Hap*II-A, (i) *Hae*III-A, (j) *Hap*II-$B_1$, (k) *Hae*III-C. The synthesis of the RNA products (a – e, g) were carried out in the presence of [$\gamma$-$^{32}$P]GTP, whereas the others (h – k, f) were carried out in the presence of [$\gamma$-$^{32}$P]ATP

protein-synthesizing system the synthesis of the polypeptide encoded by gene IV [7] (and unpublished results). These observations suggest, therefore, that this RNA species is initiated at the same promoter on fragment *Hap*II-A as is the RNA transcribed from this fragment in the DNA-dependent protein-synthesizing system [7, 10]. This A promotor, designated $A_{0.36}$, therefore must be located at or within a distance of approximately 150 nucleotides from the 5′ end of this restriction fragment (Fig. 1).

### *Fragment* Hap*II-B*$_1$

With the aid of marker rescue experiments we have demonstrated that restriction fragment *Hap*II-$B_1$ ($\approx$800 base pairs long) contains genetic markers for both gene VI and I and very likely also for gene III [5, 15]. The order of these genes on the genetic map is 5′-III-VI-I-3′ (Fig. 1).

Contrary to expections, transcription of fragment *Hap*II-$B_1$ with RNA polymerase holoenzyme resulted in the formation of two, instead of one, RNA species, the sizes of which are approximately 500 and 200 nucleotides long (Fig. 2j). Comparison of the bands on the autoradiograph indicated that they differ in intensity. This suggests that these RNA species are transcribed from promoters which differ in their affinity for RNA polymerase holoenzyme.

Genetic analysis as well as protein synthesis studies *in vitro* [5, 7, 15] have indicated that the size of gene VI must be small and, furthermore, that the N-terminal ends of both genes VI and I are located proximal rather

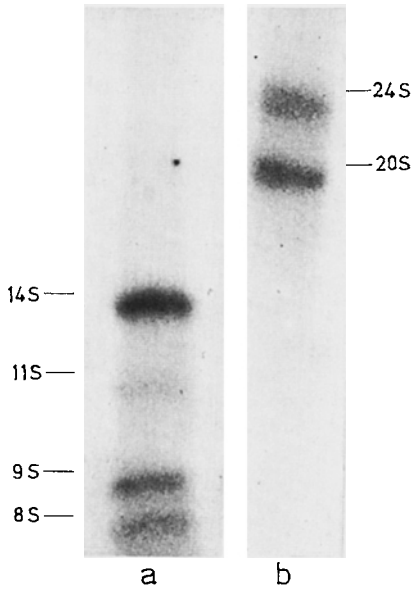


Fig. 3. *Autoradiograph of the RNA species synthesized under the direction of M13 RF III.* M13 RF III was obtained by cleavage of RF I with endoR · *Hind*II [10]. The conditions for RNA synthesis *in vitro* and the analysis of the RNA products on 3.8% polyacrylamide slab gels in 98% formamide have been described under Methods. (a) RNA products synthesized on M13 RF III in the presence of [$\gamma$-$^{32}$P]GTP. (b) RNA products synthesized on M13 RF III in the presence of [$\gamma$-$^{32}$P]ATP

than distal to the 3′ end of fragment *Hap*II-$B_1$. Since gene VI and gene I are adjacent to each other on the genetic map it is attractive to postulate that the promoters of the RNA species transcribed from fragment *Hap*II-$B_1$ are located immediately in front of the N-terminal ends of genes VI ($A_{0.56}$) and I ($A_{0.51}$), *i.e.* at 500 and 200 nucleotides from the 3′ end of fragment *Hap*II-$B_1$ (Fig. 1).

### *Fragment* Hae*III-A*

In earlier reports [5, 7, 10] evidence was given that fragment *Hae*III-A encompasses the C-terminal end of gene III, the complete genes VI and I and about 1000 nucleotides of the N-terminal end of gene IV (Fig. 1). As shown on the cleavage maps of M13 RF the complete nucleotide sequences of fragments *Hap*II-$B_1$, *Hap*II-$E_1$, and *Hap*II-G, and about 1000 nucleotides of the 5′ end of fragment *Hap*II-A are contained on this large restriction fragment (Fig. 1). Transcription of fragment *Hae*III-A in the presence of [$\gamma$-$^{32}$P]-ATP results in the formation of two major RNA species which are about 2000 and 850 nucleotides long (Fig. 2i), and a minor one, the size of which is approximately 1750 nucleotides and which is recognized as a faint band just beneath the largest RNA product on the formamide gel (Fig. 2i).

In the previous sections we have demonstrated that the RNA species transcribed from fragment *Hap*II-$B_1$ are approximately 500 and 200 nucleotides long. Fragment *Hap*II-$B_1$, however, is about 1650 nucleotides shorter at its 3′ end than fragment *Hae*III-A [5]. This evidence, therefore, strongly suggests that the two largest RNA species transcribed from fragment *Hae*III-A are initiated at the same promoters ($A_{0.56}$ and $A_{0.51}$, Fig. 1) as are the RNA species transcribed from fragment *Hap*II-$B_1$. Moreover, our finding that the synthesis of the 1750-nucleotides-long RNA product is much lower than the 2000-nucleotides-long transcript is consistent with the observed character of these promoters of which the $A_{0.51}$ promoter has the weakest affinity for RNA polymerase.

In the previous section we also demonstrated that the $A_{0.36}$ promoter is located at or within the first 150 nucleotides of fragment *Hap*II-A. Since fragment *Hae*III-A encompasses the first 1000 nucleotides of fragment *Hap*II-A, it is concluded that approximately 850 nucleotides of the N-terminal end of gene IV are located on this *Hae*III-A fragment. The observation that the smallest RNA species transcribed from fragment *Hae*III-A is almost as long as this N-terminal end, *i.e.* about 850 nucleotides, strongly suggests that this RNA species is initiated at the promoter ($A_{0.36}$) which is located in front of gene IV. The latter conclusion is completely consistent with our previous results of coupled transcription and translation studies which showed that fragment *Hae*III-A is only capable of synthesizing the N-terminal part of gene-IV protein [7, 10].

### *RF III*

Cleavage of circular M13 RF I with the restriction endonuclease R · *Hind*II results in the formation of a linear double-stranded molecule (RF III) of genome length [11]. The single cleavage site for this restriction enzyme has been defined on the physical map at zero point position (Fig. 1) [14, 16].

As is shown in Fig. 2f and 3b, transcription of RF III in the presence of [$\gamma$-$^{32}$P]ATP, followed by electrophoretic analysis on formamide slab gels, gives rise to the formation of two ATP-initiated RNA chains, the sizes of which are approximately 2100 and 3200 nucleotides. If it is assumed that the synthesis of these RNA chains has ceased at the 3′-terminal end of the linear RF molecule (*cf.* [20]), their starting positions should be located at about 0.3 and 0.5 map units (Table 1 and Fig. 1). These positions are almost identical to the positions of the $A_{0.36}$ and $A_{0.56}$ (or $A_{0.51}$) promoters as deduced from the transcription data on the fragments *Hae*III-A *Hap*II-A and *Hap*II-$B_1$ (see above).

It cannot be ascertained whether the largest RNA species (3200 nucleotides long) is a homogeneous transcriptional product of the strong promoter $A_{0.56}$

or consists of a mixture of chains, which deviate only slightly in length and which originate from both the $A_{0.56}$ promoter and the (weak) $A_{0.51}$ promoter.

## LOCATION OF G PROMOTERS

Previously we reported that five promoters are present on the M13 genome which give rise to the synthesis of pppG-initiated RNA chains [8]. The relative strengths of these G-promoters, as deduced from the incorporation values of [$\gamma$-$^{32}$P]GTP into RNA chains, and their approximate map positions, as deduced from the length of RNA chains synthesized on intact RF I, are given in Table 1.

In order to map these G promoters more accurately the same strategy as used above for the mapping of the A promoters was followed.

### *Fragment* Hae*III-D*

Restriction fragment *Hae*III-D, which is approximately 310 base pairs long [14], was recently shown to encompass a part of the intergenic space between gene II and gene IV and a part of the N-terminal end of gene II [15].

Transcription of this fragment results in the formation of one major RNA product of about 250 nucleotides long and several minor RNA species (Fig. 2b). Interestingly, the minor species, the sizes of which are about 600 and 900 nucleotides, are one and more fragment-length longer than the major RNA transcript. These observations strongly suggest that termination of transcription does not always occur at the 3′-terminal end of this particular fragment (*cf.* [20–22]) but that the RNA polymerase infrequently is able to switch transcription at this end from one strand to the other. The observation that the decrease in intensity of the RNA bands parallels their increase in size is consistent with the latter hypothesis. In this respect it has to be emphasized that fragment *Hae*III-D is the sole fragment we have isolated so far which demonstrates this template switching so clearly. Since the replication origin for complementary strand synthesis of progeny M13 RF is located near this fragment [23], it is possible that this fragment has a particular secondary and tertiary structure. This in turn may explain why transcription is not always terminated at the 3′ end of this restriction fragment.

Our observation that the N-terminal end of gene II is located on fragment *Hae*III-D [15] together with the observation that the RNA initiation site is located at a distance of approximately 250 nucleotides from its 3′ end, strongly suggests that this $G_{0.08}$ promoter is located in the intergenic region between gene II and gene IV at a position which is immediately proximal to the N-terminal end of gene II (Fig. 1).

### *Fragment* Hap*II-D*

As is shown in Fig. 1, restriction fragment *Hap*II-D ($\approx$560 base pairs long) encompasses almost completely the first half of gene II [5, 7, 15]. It does not contain, however, the N-terminal end of this gene. Transcription of fragment *Hap*II-D results in the weak synthesis of an RNA species which is about 350 nucleotides long (Fig. 2c). This observation suggests that a promoter, namely promoter $G_{0.01}$ (Fig. 1), is located on this fragment at approximately 200 base pairs distal to the 5′ end.

Cleavage of restriction fragment *Hap*II-D with the restriction endonuclease R · *Hind*II results in the formation of two new DNA fragments which are 240 and 300 base pairs long [14]. The 5′ end of fragment *Hap*II-D is located on the smallest of these two fragments (Fig. 1). Transcription of either fragment did not result in the formation of detectable amounts of RNA (Fig. 2a). These observations may be explained by the following alternatives: either (a) the biological activity of the promoter site on fragment *Hap*II-D is completely destroyed due to cleavage of this fragment by the restriction endonuclease R · *Hind*II or (b) the promoter-containing DNA fragment which remains after cleavage of *Hap*II-D with endoR · *Hind*II gives rise to the synthesis of such a small RNA species that its presence on the polyacrylamide gel is difficult to detect. Independent of the correct explanation, the results are in any case in full agreement with the conclusion that the $G_{0.01}$ promoter on *Hap*II-D is located at about 200 nucleotides from the 5′ end of this fragment.

### *Fragment* Hap*II-C*

Genetic studies, as well as protein synthesis studies *in vitro*, have recently indicated that this fragment, which is about 650 base pairs long, constitutes the C-terminal part of gene II and the N-terminal part of gene V [5, 10, 15]. The order of these genes on the genetic map is 5′-II-V-3′ (Fig. 1). Transcription of fragment *Hap*II-C results in the formation of one major and one minor RNA product, the sizes of which are 570 and 180 nucleotides, respectively (Fig. 2d). The promoters which initiate these pppG-containing RNA chains, therefore, have to be positioned at 0.94 and 0.88 map units (Fig. 1). As judged from the intensity of the bands on the autoradiograph, it is evident that the promoters $G_{0.94}$ and $G_{0.88}$ differ significantly in their affinity for RNA polymerase. This difference is not only noted with fragment *Hap*II-C but is also apparent when intact RF I is used as a template for transcription (*cf.* Table 1).

Infrequently, the weak synthesis of an RNA species ($\approx$1300 nucleotides), which was about twice as long as the fragment template, could be observed. It

is probable that the latter product is initiated at the strong $G_{0.94}$ promoter, which is located proximal to the 5′ end of fragment *Hap*II-C. Elongation then proceeds beyond the 3′ end of this fragment, most probably by switching transcription from one strand to the other.

DNA-dependent protein synthesis studies *in vitro* have demonstrated that fragment *Hap*II-C directs the synthesis of protein X, a polypeptide which is about 130 amino acid residues long [5, 7] and whose complete genetic information is located within the C-terminal end of gene II [15, 33]. From our present observation that the major RNA product of fragment *Hap*II-C, which is initiated at the strong $G_{0.94}$ promoter, is large enough to code for this polypeptide, we infer that in front of this 'gene' the $G_{0.94}$ promoter is positioned.

### *Fragment* Hap*II-B*$_2$

Restriction fragment *Hap*II-$B_2$ is approximately 800 base pairs long (Fig. 1). In a DNA-dependent cell-free system it directs the synthesis of the polypeptide encoded by gene VIII [10]. Genetic analysis and transcription and translation studies *in vitro* have indicated that besides gene VIII also gene VII and the N-terminal part of gene III are located on this restriction fragment [5, 7, 8, 10, 15]. The order of these genes on the genetic map is 5′-VII-VIII-III-3′ (Fig. 1).

Previously, we have demonstrated that the promoter $G_{0.82}$ as well as the central termination site for transcription are located on fragment *Hap*II-$B_2$ [8]. These regulatory sites, which mark the ends of 8-S RNA (360 nucleotides), are located at 90 and 450 nucleotides, respectively, from the 5′ end of this restriction fragment (Fig. 1).

As shown in Fig. 2e, fragment *Hap*II-$B_2$ directs, besides the synthesis of the 8-S RNA, also the weak synthesis of an RNA species which is about 720 nucleotides long. The 8-S RNA product has previously been shown to code for the synthesis of gene VIII protein only [8, 33]. The length of the larger RNA product corresponds to the distance between the $G_{0.82}$ promoter and the 3′ end of fragment *Hap*II-$B_2$. This observation, together with the result that this large RNA is also able to direct the synthesis of gene VIII protein (unpublished results) suggests that this 720-nucleotides-long RNA is the product of a transcription process which is initiated at promoter $G_{0.82}$ but which is not terminated until the RNA polymerase has reached the 3′ end of this restriction fragment. From these results and our unpublished data we infer that the RNA polymerase holoenzyme leaks infrequently, for reasons still unknown, through this rho-independent termination signal which is located close to the center of fragment *Hap*II-$B_2$.

### *RF III*

Previously we demonstrated that transcription of RF I molecules in the presence of [$\gamma$-$^{32}$P]GTP results in the formation of five GTP-initiated RNA chains, the sizes of which correspond to 8 S, 11 S, 14 S, 17 S and 19 S [8]. This is demonstrated again in Fig. 2g. In contrast, transcription of RF III gives rise to only four GTP-initiated RNA chains of 360, 450, 700 and 1200 nucleotides (Fig. 3a). The former is identical to 8-S RNA, the largest products correspond to 11-S and 14-S RNA, respectively. As compared to RF I, two RNA products, *i.e.* 17-S and 19-S RNA are missing but a shorter one, namely 9-S RNA, is present (Fig. 3a). These observations are consistent with our findings (above) that the endoR · *Hind*II cleavage site on the M13 genome is located in or at least in the vicinity of the $G_{0.01}$ promoter. Transcription from the $G_{0.08}$ promoter then results in the formation of a prematurely terminated RNA chain of about 450 nucleotides (9-S RNA).

## DISCUSSION

After infection of an *Escherichia coli* cell with bacteriophage M13 not all phage-specific proteins are synthesized in equimolar amounts [24]. Some proteins, such as the 'DNA-unwinding protein' encoded by gene V and the major capsid protein encoded by gene VIII, are synthesized in much larger quantities than the other phage-specific proteins. The latter proteins are also the major products synthesized in M13-DNA-directed reactions *in vitro* [6, 7, 25]. To explain these noticeable differences of gene expression several regulatory mechanisms have been proposed [25]. The available experimental data [3, 8, 10, 26, 27, 30] favour the model according to which the replicative form molecule is transcribed into several distinct polycistronic mRNAs which are initiated at different promoter sites (RNA initiation sites) but all are terminated at a unique termination signal which is located immediately after gene VIII [8, 27].

There are several indications that such a regulatory mechanism of gene expression is not solely restricted to the filamentous coliphages but that it is also operative during the expression of the late genes of the phages T7 [28], T4 and $\lambda$ (E. Young, B. Studier and M. Pearson, personal communications). Whether this mechanism of gene expression possesses a more universal character however, awaits, further investigation.

The results presented in this paper provide strong evidence that at least eight promoter sites are located on the M13 genome. The map positions of these promoter sites are shown in Fig. 1. The RNA transcribed from three of these promoters ($A_{0.56}$, $A_{0.51}$ and $A_{0.36}$) are initiated with pppA (A-start RNA) while the RNAs transcribed from the other five promoters

($G_{0.08}$ through $G_{0.82}$) are initiated with pppG (G-start RNA). All of these RNA species share the coding information for the major capsid protein (our unpublished results) which support the observation that on the M13 genome a rho-independent termination of transcription occurs at a unique site which is located immediately distal to gene VIII [8].

From the results of the experiments in which replicative form I molecules are transcribed in the presence of either [$\gamma$-$^{32}$P]ATP or [$\gamma$-$^{32}$P]GTP it can be concluded that the RNA chains are initiated with different frequencies (Table 1). Hence, both strong and weak promoters are located on the M13 genome. The strongest promoters are $G_{0.08}$, $G_{0.94}$, $G_{0.82}$ and $A_{0.36}$ (Fig. 1). Based upon the $^{32}$P radioactivity incorporated into each RNA chain, the remaining promoters ($G_{0.01}$, $G_{0.88}$, $A_{0.51}$ and $A_{0.56}$) revealed about 0.25–0.1-fold weaker affinity for RNA polymerase holoenzyme. Whether these differences in affinity reflect differences in nucleotide sequences between these promoters has yet to be established. From nucleotide sequence analysis of the strong promoters $G_{0.82}$ and $G_{0.94}$ of phage fd it is known that the longest sequence common to both promoters is only T-A-T-A-A-T [21,22,29].

Recently, Seeburg and Schaller [30] have also studied promoter locations on the genome of M13 and its relatives fd and f1. Their method consisted of binding RNA polymerase to restriction fragments in the presence or absence of various triphosphate mixtures. The results obtained by this group cover to a large extent the results we have obtained since they also found promoter sites on the restriction fragments *Hap*II-A, *Hap*II-$B_1$, *Hap*II-$B_2$, *Hap*II-C, *Hap*II-D and *Hap*II-F. Unlike our observations they did not find two promoter sites on fragments *Hap*II-$B_1$ and *Hap*II-C. These discrepancies, however, are almost certainly due to the different techniques used for the location of promoter sites. Okamoto *et al.* [31] have studied the location of promoter sites on the fd genome with the aid of the filter-binding assay and, in addition, with the same technique used here. However, they detected only four promoter sites on the fd genome. Three of these promoters (G promoters) were localized at the same map positions as we have found for the strong promoters $G_{0.82}$ (*Hap*II-$B_2$), $G_{0.94}$ (*Hap*II-C) and $G_{0.08}$ (*Hae*III-D) on the M13 genome. The fourth promoter (A promoter), however, was localized at map position 0.46 (Fig. 1). This position is completely different from the positions of the A promoters found on the fd genome by Seeburg and Schaller [30] and on the M13 genome by our group. At present we have no reasonable explanation for this disagreement. An explanation might be, though not very likely, that both groups are working with variants of the original fd strain.

Given the positions of the various M13 phage promoters on the physical map, it is very attractive to correlate their positions with the genetic map (Fig. 1). In previous communications [7,10] we have provided evidence for the existence of a promoter in front of gene IV, a promoter in front of gene VIII, and an RNA initiation site which was positioned somewhere in the middle of gene II. The promoters which accord with such positions are the $A_{0.36}$, the $G_{0.82}$ and the $G_{0.94}$ promoter, respectively. Several studies have demonstrated now that the deduced position of $G_{0.82}$ and $A_{0.36}$ in front of these genes is correct [8,32] (and our unpublished results). In the meantime, it has been shown that $G_{0.94}$ represents an internal start of RNA synthesis which is located within the C-terminal region of gene II and which leads to the synthesis of protein X [15]. Moreover, transcription of several new restriction fragments of M13 RF and subsequent translation of the synthesized RNA products now have demonstrated unambiguously that promoter $G_{0.88}$ is positioned in front of gene V, that promoter $G_{0.08}$ is located in front of gene II and that $G_{0.01}$ represents a second but weak intragenic initiation site (our unpublished results).

The biological function of these intragenic initiation sites within gene II is not known. As far as protein X is concerned, we have no evidence at present whether the amino acid sequence of this protein is identical to the amino acid sequence of the C-terminal part of gene II protein. In other words, it is not known whether the mRNA coding for protein X is translated in the same reading frame as the mRNA coding for gene II protein. If not, one is left with the unique situation that one of the M13 genes, namely gene X, is enclosed within another gene, *i.e.* gene II.

Since we do not know the exact positions of the 5' ends of genes VI and I on the genetic map, it is impossible to define unambiguously which promoters are located in front of these genes. The experimental data presented in this study, however, favour the hypothesis that the promoters $A_{0.56}$ and $A_{0.51}$ are positioned in front of these genes. Genetic studies have indicated that *in vivo* a polarity exists among genes III, VI and I in that the expression of genes VI and I is dependent on the expression of gene III. Our findings that promoters are located most probably in front of the N-terminal ends of genes I and VI suggest that both genes can be expressed independently of gene III. The biological implications of these discrepancies can not yet be understood.

Previously we have presented evidence that on the M13 genome termination of transcription occurs at a unique site which is located between the genes III and VIII [8]. Furthermore, we have demonstrated (unpublished results) that the mixture of RNAs transcribed from M13 RF directs in a protein-synthesizing system the synthesis of gene III protein. These results

are in favour of a promoter in front of gene III. However, up to now, we have failed to demonstrate the presence of such a promoter. Three hypothesis could account for these observations.

a) A promoter is located in front of gene III but its capacity to bind RNA polymerase holoenzyme is destroyed by cleavage of the circular replicative form I molecules into linear DNA fragments.

b) A promoter is located in front of gene III but its affinity for RNA polymerase is very weak resulting in the synthesis of very small amounts of high-molecular-weight RNA ($M_r \approx 2 \times 10^6$). Such small amounts of RNA might have escaped detection on the polyacrylamide gel system used.

c) No promoter is located in front of gene III, but termination of transcription is not stringent. Hence, transcription of gene III is the result of a transcription process whereby RNA polymerase leaks infrequently through the central termination site located between the C-terminal end of gene VIII and the N-terminal end of gene III (Fig. 1). As a consequence, small amounts of high-molecular-weight transcripts of gene III are synthesized which are, by definition, heterogeneous in nature and therefore difficult to detect by electrophoretic analysis. Our observation that fragment *Hap*II-$B_2$ (see Results) directs, besides 8-S RNA, also the synthesis of a 720-nucleotides-long RNA which also carries the coding information of for gene VIII protein, can only be explained by such a 'read-through' transcription process and is, therefore, in agreement with the latter hypothesis.

The results obtained thus far are consistent with the multi-promoter single-terminator model suggested by Sugiura *et al.* [36]. That the different size classes of phage M13 mRNA synthesized *in vitro* are indeed the result of such a 'cascade' mechanism of transcription has recently been demonstrated unambiguously by transcription and subsequent translation studies of various M13 DNA restriction fragments (unpublished results).

The proposed cascade mechanism, giving rise to RNA molecules with overlapping nucleotide sequences, may prove to be an important mechanism for regulating the amount of the various gene products of this phage. Particularly, the genes which are located proximal to the central termination site, *i.e.* gene VIII and gene V, will be transcribed more frequently than the other genes resulting in the formation of larger amount of their encoded proteins. Our previous observations that the DNA-binding protein, encoded by gene V, and the major coat protein, encoded by gene VIII, are synthesized *in vitro* in much larger amounts than the other phage proteins [7, 25] (*cf.* [6]) are in any case consistent with the proposed cascade mechanism.

Gene V protein and gene VIII protein are also the products which are most abundantly present in the infected cell [24]. However, no direct evidence is available yet which demonstrates that the proposed model of transcription is also relevant for the control of M13 gene expression *in vivo*. Preliminary investigations indicate that in the infected cell several RNA species are present which range in size from 8 S up to 30 S and which share the coding information for the polypeptide encoded by gene VIII (unpublished results). Whether these RNA species are initiated at the same promoters and whether they are formed by cascade transcription of the M13 genome, is currently under investigation.

## REFERENCES


1. Denhardt, D. T. (1975) *CRC Crit. Rev. Microbiol. 4*, 161–223.
2. Jacob, E. & Hofschneider, P. H. (1969) *J. Mol. Biol. 46*, 359–363.
3. Okamoto, T., Sugiura, M. & Takanami, M. (1969) *J. Mol. Biol. 45*, 101–111.
4. Jacob, E., Jaenisch, R. & Hofschneider, P. H. (1970) *Nature (Lond.) 277*, 59–60.
5. Van den Hondel, C. A., Weyers, A., Konings, R. N. H. & Schoenmakers, J. G. G. (1975) *Eur. J. Biochem. 53*, 559–567.
6. Model, P. & Zinder, N. D. (1974) *J. Mol. Biol. 83*, 231–251.
7. Konings, R. N. H., Hulsebos, T. & van den Hondel, C. A. (1975) *J. Virol. 15*, 570–584.
8. Edens, L., Konings, R. N. H. & Schoenmakers, J. G. G. (1975) *Nucleic Acid Res. 2*, 1811–1820.
9. Reference deleted.
10. Van den Hondel, C. A., Konings, R. N. H. & Schoenmakers, J. G. G. (1975) *Virology, 67*, 487–497.
11. Van den Hondel, C. A. & Schoenmakers, J. G. G. (1973) *Mol. Biol. Rep. 1*, 41–45.
12. Konings, R. N. H., Jansen, J., Cuypers, T. & Schoenmakers, J. G. G. (1973) *J. Virol. 12*, 1466–1472.
13. Staynov, D. Z., Pinder, J. C. & Gratzer, W. B. (1972) *Nat. New Biol. 235*, 108–110.
14. Van den Hondel, C. A. & Schoenmakers, J. G. G. (1975) *Eur. J. Biochem. 53*, 547–558.
15. Van den Hondel, C. A. Pennings, L. & Schoenmakers, J. G. G. (1976) *Eur. J. Biochem. 68*, 55–70.
16. Van den Hondel, C. A. & Schoenmakers, J. G. G. (1976) *J. Virol. 18*, 1024–1039.
17. Fellner, P. (1974) in *Ribosomes* (M. Nomara, A. Tissières & P. Lengyel, eds) pp. 169–191, Cold Spring Harbor Press, New York.
18. Gillespie, D. & Spiegelman, S. (1965) *J. Mol. Biol. 12*, 829–842.
19. Petterson, U., Sambrook, J., Delius, H. & Tibbetts, C. (1974) *Virology, 59*, 153–167.
20. Heyden, B., Nüsslein, C. & Schaller, H. (1975) *Eur. J. Biochem. 55*, 147–155.
21. Schaller, H., Gray, C. P. & Herrmann, K. (1975) *Proc. Natl Acad. Sci. U.S.A. 72*, 737–741.
22. Sugimoto, K., Okamoto, T., Sugisaki, H. & Takanami, M. (1975) *Nature (Lond.) 253*, 410–414.
23. Horiuchi, K. & Zinder, N. D. (1976) *Proc. Natl Acad. Sci. U.S.A. 73*, 2341–2345.
24. Henry, T. J. & Pratt, D. (1969) *Proc. Natl Acad. Sci. U.S.A. 62*, 800–807.
25. Konings, R. N. H. (1973) *FEBS Lett. 35*, 155–160.
26. Takanami, M., Okamoto, T. & Sugiura, M. (1970) *Cold Spring Harbor Symp. Quant. Biol. 34*, 179–187.
27. Chan, T., Model, P. & Zinder, N. D. (1975) *J. Mol. Biol. 99*, 369–382.

28. Pachl, C. A. & Young, E. T. (1976) *Proc. Natl Acad. Sci. U.S.A. 73*, 312–316.
29. Takanami, M., Sugimoto, K., Sugisaki, H. & Okamoto, T. (1976) *Nature (Lond.) 260*, 297–302.
30. Seeburg, P. H. & Schaller, H. (1975) *J. Mol. Biol. 92*, 261–277.
31. Okamoto, T., Sugimoto, K., Sugisaki, H. & Takanami, M. (1975) *J. Mol. Biol. 95*, 33–44.
32. Vovis, G. F., Horiuchi, K. & Zinder, N. D. (1975) *J. Virol. 16*, 674–684.
33. Pratt, D., Tzagoloff, H., Erdahl, W. S. & Henry, T. J. (1967) in *Molecular Biology of Viruses* (J. S. Colter & W. Parachych, eds) pp. 219–238, Academic Press, New York.
34. Lyons, L. B. & Zinder, N. D. (1972) *Virology, 49*, 45–60.
35. Sugiura, M., Okamoto, T. & Takanami, M. (1969) *J. Mol. Biol. 43*, 299–315.

L. Edens, P. van Wezenbeek, R. N. H. Konings, and J. G. G. Schoenmakers
Laboratorium voor Moleculaire Biologie, Katholieke Universiteit Nijmegen, Toernooiveld, Nijmegen, The Netherlands