Volume 5 Number 12 December 1978 Nucleic Acids Research

Nucleotide sequence of gene VII and of a hypothetical gene (IX) in bacteriophage M13

T.Hulsebos and J.G.G.Schoenmakers

Laboratory of Molecular Biology, University of Nijmegen, Nijmegen, Netherlands

Received 5 September 1978

ABSTRACT

A DNA fragment containing gene VII of bacteriophage M13 has been transcribed and the nucleotide sequence of this 169-nucleotides long transcript was determined by RNA sequencing methods. Additionally, the nucleotide sequence of this gene and parts of its neighbouring genes V and VIII has been determined by the dimethylsulphate-hydrazine technique.

The reading frame of gene VII has been established by determining the nucleotide changes occurring in the transcripts of two amber mutants of this gene. From these combined data it is apparent that gene VII is only 99 nucleotides long and is immediately followed by the termination codon UGA. Its initiation codon AUG is separated from gene V by only a single nucleotide. It was noted that between the UGA termination codon of gene VII and the initiation codon of the next gene (gene VIII) there is space for another, hitherto unknown gene. This gene (IX) most probably codes for the small polypeptide ("C-protein") present in mature M13 phage particles.

INTRODUCTION

Bacteriophage M13 is a small filamentous coliphage, closely related to the phages fd, f1 and ZJ2. The genome of these phages consists of a circular single-stranded DNA which comprises only 6400 bases.

In the last few years a rapid progress is being made toward an understanding of the molecular biology of these phages (for a review see 1,2). In particular, genetic mapping, the process of viral DNA replication and the mechanism of transcription and translation has been studied in detail. The M13 genome is known to code for at least nine gene products, some of which have been well characterized regarding their biological function. In particular, the proteins encoded by gene II (nickase) and gene V (DNA-binding protein) are functional elements in the process of viral DNA replication whereas the proteins encoded by genes III and VIII are constituents of the mature phage particle. The biological function of the genes I, IV, VI and VII are still unknown although there is evidence that the products

of these genes are involved in the process of phage maturation.

By coupled transcription-translation of M13 replicative form (RF) DNA (3,4) or restriction fragments (5,6) each M13 gene product has now been identified and characterized regarding their molecular weight. The exceptions are genes VI and VII the products of which have neither been observed among the products in the infected cell (7) nor among the synthetized products in the cell-free systems applied (3-6,8). Also in minicells harbouring M13 RF, the synthesis of these two proteins could not yet be demonstrated (9).

The low level of gene VII-protein synthesis is rather intruiging. Previously we have shown that the transcription of M13 RF is initiated at nine different promotor sites and terminates at a single unique site (6,8,10,11). This central termination site has been localized immediately distal to gene VIII (11). Since gene VII is positioned on the genetic map between gene V and gene VIII (Fig. 1) and transcription proceeds in only one direction along the genetic map, gene VII is located in a region where the highest transcriptional activity of the DNA genome has been demonstrated (8). Moreover, an RNA transcript has been isolated which encompasses the coding information of genes V, VII and VIII, but upon translation of this polycistronic RNA in an *in vitro* protein synthesizing system only the abundant synthesis of the proteins of genes V and VIII is apparent (8). This strongly suggests that the expression of gene VII is controlled at the level of translation.

As a step towards the elucidation of the structural features of this gene we have undertaken to sequence the region encompassing gene VII and parts of its neighbouring genes. In the meantime, the sequence was determined of RNA transcripts derived from M13 DNA restriction fragments carrying several amber mutations in gene VII. From the results of these studies the reading frame could be determined, ultimately leading to a detailed knowledge of the primary structure of gene VII.

MATERIALS AND METHODS

Materials

Replicative form DNA, either derived from wild-type or amber mutant M13 phages, was prepared from *E.coli* C89 (su^-) or *E.coli* K37 (suI^+)-infected cells by the procedure described previously (12). The restriction enzymes R.*Hap* II and R.*Hae* III were prepared as described in a previous report (13)

EndoR.*Hha* I was purchased from New England Biolabs and EndoR.*Taq*I was from the Microbiol.Res.Establishment, Porton. *E.coli* RNA polymerase holoenzyme was a generous gift of Dr. R. Schilperoort, Leiden. The M13 nonsense mutants *am*7-H2 and *am*7-H3 were provided by Dr. D. Pratt, Davis. The primer dinucleotides GpC, CpG, ApA and ApG were purchased from Boehringer, Germany.

Restriction fragments

M13 RF (200 µg), enriched with uniformly labelled (^{32}P)-M13 RF, was digested with the appropriate endonuclease in excess using the conditions previously described (14). The digest was layered on a 3% discontinuous polyacrylamide slab gel (40 cm x 20 cm x 0.2 cm) formed in 40 mM Tris, 20 mM sodium acetate, 2 mM EDTA, pH 7.8. After electrophoresis for 16 h at 40 mA, the gels were covered with Saran wrap and autoradiographed. Gel segments corresponding to the ^{32}P-radioactive bands were excised and the DNA fragments were extracted from the gel as described by Van den Hondel *et al.* (14).

Preparation of transcripts.

RNA synthesis was carried out in a reaction mixture (0.15 ml) containing 40 mM Tris-HCl, pH 7.6, 150 mM KCl, 0.1 mM dithiothreitol, 4 mM EDTA, 15 mM $MgCl_2$, 0.1% Tween-80, 20 µM nucleoside (α-^{32}P)-triphosphate, 200 µM of each of the other ribonucleoside triphosphates, 2 pmol of "300-fragment" and about 20 pmol of *E.coli* RNA polymerase holoenzyme.

The transcription was started by the addition of $MgCl_2$. After 30 min at 37° the reaction was terminated by the addition of 150 µl of 10 mM Tris-HCl, 1 mM EDTA, pH 7.6 (buffer A) containing 150 µg of carrier tRNA per ml and 0.1% SDS. The mixture was extracted with an equal volume of freshly distilled phenol, then 0.1 vol. of 3 M sodium acetate, pH 5.6, was added to the aqueous phase and the RNA was precipitated twice with 2.5 vol. of ethanol for 1 h at -80°. The RNA was dried *in vacuo* and dissolved in 20 µl of 40 mM Tris, 20 mM sodium acetate, 2 mM EDTA, pH 7.8 buffer containing 7 M urea, 20% glycerol, 0.1% SDS and 0.1% bromophenol blue. After dissolution the RNA was heated for one min at 90°, rapidly chilled and subjected to electrophoresis on polyacrylamide gels.

The conditions for primer-dependent RNA synthesis were identical to the standard conditions except that a primer was added to a final concentration of 400 µM and the concentration of all ribonucleoside triphosphates, one of which was labelled with ^{32}P in the α-position, were 10 µM.

Gel electrophoresis and recovery of RNA from the gel

RNA products were fractionated by electrophoresis on 4% polyacrylamide slab gels (20 cm x 20 cm x 0.2 cm) which were prepared in 40 mM Tris, 20 mM sodium acetate, 2 mM EDTA, pH 7.8, containing 7 M urea and 0.1% SDS. After electrophoresis for 5 h at 30 mA (about 75 V) and autoradiography, the portions of the gel containing the RNA were cut out, crushed by piercing through a hypodermic syringe and extracted twice for 4 h with 2 ml of buffer A containing 0.1% SDS and 10 µg of carrier tRNA. The extracts were combined and the RNA was precipitated with ethanol. The precipitate was spun down, dissolved in 0.3 ml of buffer A and reprecipitated with ethanol. The RNA precipitate was dried *in vacuo* and finally dissolved in about 15µl of buffer A. The recovery of ^{32}P-labelled RNA after this isolation procedure was 70-80%.

RNA sequencing methods

Standard RNA sequencing methods were used according to Brownlee and Sanger (15) and Barrell (16). Digestion of RNA with RNase T1 (Sankyo Co.) was carried out in 10 µl of 10 mM Tris-HCl, 1 mM EDTA, pH 7.4 for 30 min at 37° using a ratio of enzyme/carrier RNA of 1:20. The resulting T1-oligonucleotides were fractionated by electrophoresis at pH 3.5 on cellulose acetate (Schleicher-Schüll) in the first dimension followed by homochromatography on DEAE-cellulose thin layer plates (Machery-Nagel, CEL 300/HR) in the second dimension. As developing medium homomixture "C" was used (15). T1-oligonucleotides eluted from fingerprints were digested with pancreatic RNase (Worthington) for 60 min at 37° with an enzyme/RNA ratio of 1:10. The pancreatic RNase products were characterized by electrophoresis on DEAE-paper (Whatman DE 81) at pH 3.5. Most of the secondary digestion products were further analysed by complete digestion with 0.5 N NaOH for 16 h at 37°. The resulting mononucleotides were fractionated by electrophoresis on Whatman 540-paper at pH 3.5 and the distribution of ^{32}P in mononucleotides was determined.

To determine the sequences of oligonucleotides for which unique sequences were not deduced by nearest neighbour analysis, partial digestion was carried out with spleen phosphodiesterase. The T1-oligonucleiotides, labelled with (α-^{32}P)-GTP and containing approximately 150 µg of carrier RNA were dissolved in 60 µl of 25 mM ammonium acetate, pH 5.7. A sample of this mixture (20 µl) was heated for 3 min at 90°. After cooling to 37°, 5 µl of spleen phosphodiesterase solution (3 mg/ml) was added and 5 µl-aliquots

were removed at 15 min intervals. The aliquots were rapidly chilled, pooled and dried *in vacuo* and subsequently dissolved in 5 µl water. The partial digestion products were fractionated by two-dimensional homochromatography using homomix C. The sequences, indicated in Table 1 by underlining, were deduced from the mobility shift pattern.

Labelling of fragments with ^{32}P at a single 5'-OH terminus

The 5'-ends of restriction fragments were dephosphorylated with bacterial alkaline phosphatase essentially as described by Maxam and Gilbert (17). Labelling of the 5'-OH ends of fragments was performed with (γ-^{32}P)-ATP and polynucleotide kinase (17). The dephosphorylated fragments (3-4 pmol) were dissolved in 45 µl of 10 mM glycine-NaOH, pH 9.5, 1 mM spermidine. 0.1 mM EDTA. The fragments were denatured by heating at 100° for 3 min, then quickly chilled and transferred to an Eppendorf tube containing 100 pmol of dried (γ-^{32}P)-ATP (spec.act. >2000 Ci/mmol). After addition of 5 µl of 0.5 M glycine-NaOH, pH 9.5, 0.1 M $MgCl_2$, 50 mM dithiothreitol, the phosphorylation was started by adding 2-3 units of polynucleotide kinase (P.L.Biochemicals). After 30 min at 37° the reaction was terminated with phenol. Carrier tRNA (10 µg) was added and after two extractions with phenol the labelled fragments were precipitated with ethanol. The precipitate was dissolved in 70 µl of buffer A, the solution was heated at 100° for 3 min and the DNA fragments were renatured by incubation at 67° for 2 h. Thereafter the appropriate restriction enzyme and buffer was added and the volume adjusted to 100 µl with buffer A. After a digestion at 37° for 2 h the 5'-labelled fragments were separated on 5% polyacrylamide gels essentially as described by Maxam and Gilbert (17).

DNA sequencing methods

Partial digestion of DNA with snake venom phosphodiesterase was carried out as described by Maniatis *et al.* (19). To 5 µl of fragment labelled at one 5'-terminal end (about 0.5 pmol) was added 5 µl of 10 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$, 10 mM mercaptoethanol, 6 mM KCl buffer and 1 µg of sonicated calf-thymus DNA, 10 µg of carrier tRNA, 7.5 ng of DNase I (Boehringer) and 2.5 ng of snake venom phosphodiesterase (Worthington). Aliquots of 2 µl, taken at 10 min intervals, were rapidly chilled, pooled, dried and finally dissolved in 5 µl ice-cold water. The partial digestion products were fractionated by two-dimensional homochromatography using homomix V (18) in the second dimension. The smaller products were eluted

from fingerprints and their sequences were determined by comparing their electrophoretic mobility on Whatman 3 MM-paper at pH 3.5 with the mobility of markers of known composition.

For DNA sequencing by chemical degradation the protocol of Maxam and Gilbert (17) was followed. Purine residues were partially methylated by dimethyl sulphate. Cleavage at Guanine was obtained by heating at neutral pH and subsequent treatment with 0.1 N NaOH at 90°. Preferential cleavage at Adenine was achieved by treatment with 0.1 N HCl followed by treatment with 0.1 N alkali at 90°. Cleavage at Cytosine and Thymine was obtained by partial hydrazinolysis followed by treatment with 0.5 M piperidine. Hydrazinolysis at Thymine was suppressed by the presence of 2 M NaCl.

Reaction mixtures were fractionated on 15% and 20% polyacrylamide slab gels (40 cm x 30 cm x 0.1 cm) which were prepared using an acrylamide/bis-acrylamide ratio of 30:1 in 50 mM Tris-borate, pH 8.3, 1 mM EDTA and 7 M urea.

RESULTS

Localization of gene VII

Previously we demonstrated that the restriction fragments *Hap*II-B_2 and *Hae*III-B (Fig. 1) contain genetic markers of gene VII. A more accurate position of these markers has recently been deduced from marker rescue experiments which showed that the M13 mutant *am*7-H2, which is an amber mutation in gene VII, is rescued not only with fragment *Taq*I-C but also with the very small *Hha*I-L fragment which is only 90 base pairs long (12) (data not shown). A second amber mutant, *am*7-H3, was rescued by both the wild-type fragments *Taq*I-H and *Hha*I-L. From this we infer that the latter fragments form parts of gene VII and that this gene is most probably located on the left-hand side of the "300-fragment" which constitutes the overlap between fragment *Hap*II-B_2 and *Hae*III-B (Fig. 1).

To substantiate this assumption, the "300-fragment" was terminally labelled with polynucleotide kinase and (γ-^{32}P)-ATP, and after subsequent cleavage with restriction enzyme *Hha*-I the fragments labelled at a single 5'-end were separated by gel electrophoresis. Each fragment was partially digested with pancreatic DNase and snake-venom phosphodiesterase as described by Maniatis *et al.* (19). The degradation products were fractionated by electrophoresis on cellulose acetate at pH 3.5 followed by homochromatography on DEAE-cellulose thin layer plates. Autoradiographs of the products generated from each 5'-terminally labelled fragment are shown in Fig. 2a

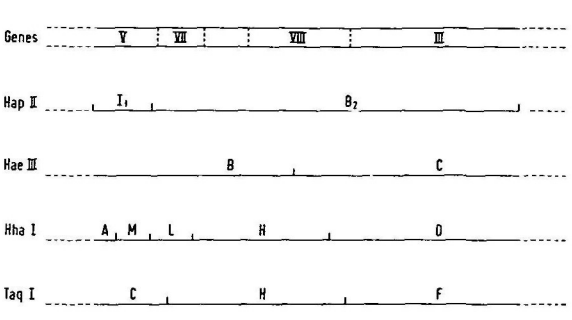

Fig. 1. Schematic diagram of a segment of the genetic map and of the restriction enzyme cleavage maps of bacteriophage M13 DNA. The Roman numerals refer to the genes. The capital letters refer to the DNA fragments which are obtained after digestion of this part of the M13 genome with the various restriction endonucleases.

and 2b. The derived sequences are summarized in Fig. 7. Interestingly, the nucleotide sequence at the left-hand terminus of the "300-fragment" corresponds to the sequence expected for the C-terminal amino acid residues *-Pro-Ala-Lys-OH* of gene V-protein (21), which is followed by the termination codon UAA whereas the nucleotide sequence at the right-hand terminus of this fragment corresponds exactly with the 5th to 9th amino acid residues *-Asp-Pro-Ala-Lys-Ala* at the N-terminal end of the major capsid protein encoded by gene VIII (22). Since the order of genes is V-VII-VIII (14) gene VII is therefore most probably confined to the "300-fragment" only.

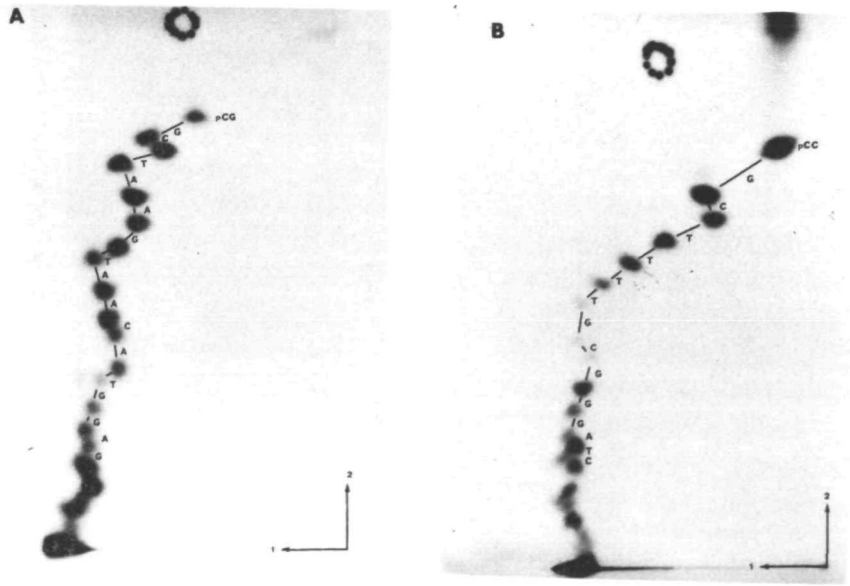

Fig. 2. Autoradiographs of two-dimensional fingerprints of oligonucleotides derived after partial digestion with snake venom phosphodiesterase of the left-hand-(A) and right-hand-(B) boundary of the "300-fragment" (*cf.* Fig. 3).

Transcription of "300-fragment"

Previously we have demonstrated that a strong promoter, designated $G_{0.18}$, is located on the "300-fragment" (10,11,20). Upon transcription of this fragment, the major product formed is an RNA species which is initiated at this promoter and which is terminated at the terminal end of the DNA fragment (11). This RNA, marked G'-RNA, is approximately 210 nucleotides long (Fig. 3 and 4). In addition, two minor RNA species are formed, readily separated from the major product on the polyacrylamide gel and which have been denoted G"-RNA and (-)RNA.

Analysis and comparison of the T1- and pancreatic oligonucleotide products obtained from the transcripts G'-RNA and G"-RNA have shown that the latter product is a prematurely terminated product consisting of the first 45 nucleotides of G'-RNA (data not shown). The (-)RNA, which is approximately 170 nucleotides long, gives rise to a completely different set of oligonucleotide products which originate from the non-codogenic viral strand (Fig. 3). If it is assumed that termination of transcription has occurred at the end of the template viral strand of the "300-fragment", the (-)RNA should cover extensive parts of fragment *Hha*I-L and *Taq*I-H (*cf.* Fig. 1) and, hence, it should be considered as a "reversed transcript" of (a large part of) gene VII. For this reason we have deduced the nucleotide sequence of (-)RNA transcribed from wild-type "300-fragment" and of (-)RNA transcribed from 300-fragments bearing various amber-7 mutations. These data enabled us to localize exactly the position of the amber mutations and allowed deduction of the reading frame of gene VII.

Under standard conditions of transcription the yield of (-)RNA is too low for nucleotide sequence analysis. To improve the yield several dinucleotide primers were tested for their capacity to stimulate the synthesis of (-)RNA. It appeared that the addition of GpC to the reaction mixture suppressed G'-RNA synthesis but did enhance the synthesis of (-)RNA several folds (Fig. 4b). Also high concentrations of rCTP had a stimulatory effect on (-)RNA synthesis (Fig. 4c) whereas no significant effects were observed with CpC (Fig. 4d), ApA, ApG and the other ribonucleoside triphosphates. Therefore, all further transcription experiments were performed with primer GpC in the reaction mixtures (final concentration 400 μM).

Analysis of T1-oligonucleotides of wild-type (-)RNA

Synthesis of (-)RNA on wild-type "300-fragment" was performed under primer-dependent transcription conditions with each of the four (α-^{32}P)

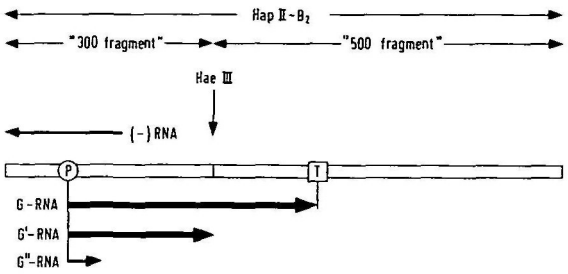

Fig. 3. Transcription map of fragment *Hap*II-B_2 and the "300-fragment". P and T refer to the promoter $G_{0.18}$ and the central terminator $T_{0.25}$, respectively.

ribonucleoside triphosphates. The transcription products were fractionated on 5% polyacrylamide slab gels and after isolation and subsequent purification, the (-)RNA was completely digested with RNase T1.

A typical fingerprint of wild-type (-)RNA, labelled with (α-^{32}P)-GTP, is shown in Fig. 5A. The distribution of ^{32}P in each spot was determined to estimate relative molar yields. All T1-oligonucleotides obtained were further characterized by digestion with pancreatic RNase and fractionation of the products by electrophoresis on DEAE-paper (16). Oligonucleotide products were further subjected to alkaline hydrolysis for nearest neighbour analysis and determination of the base composition. The results obtained are summarized in Table 1, in which nucleotide numbers correspond to the spot numbers given in Fig. 5. These analysis established the sequence of most RNase T1-oligonucleotides. The nucleotide sequences of T13, T14, T17, T22 and T23 for which unique sequences were not deduced by

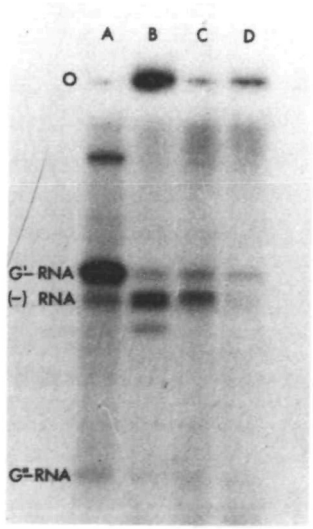

Fig. 4. Electrophoretic analysis on 4% polyacrylamide gels of the RNA products formed upon transcription of the "300-fragment" in the absence (A) and presence of the dinucleotide GpC (B) and CpC (D) or in the presence of high concentrations of CTP (C) (*cf.* Fig. 3).

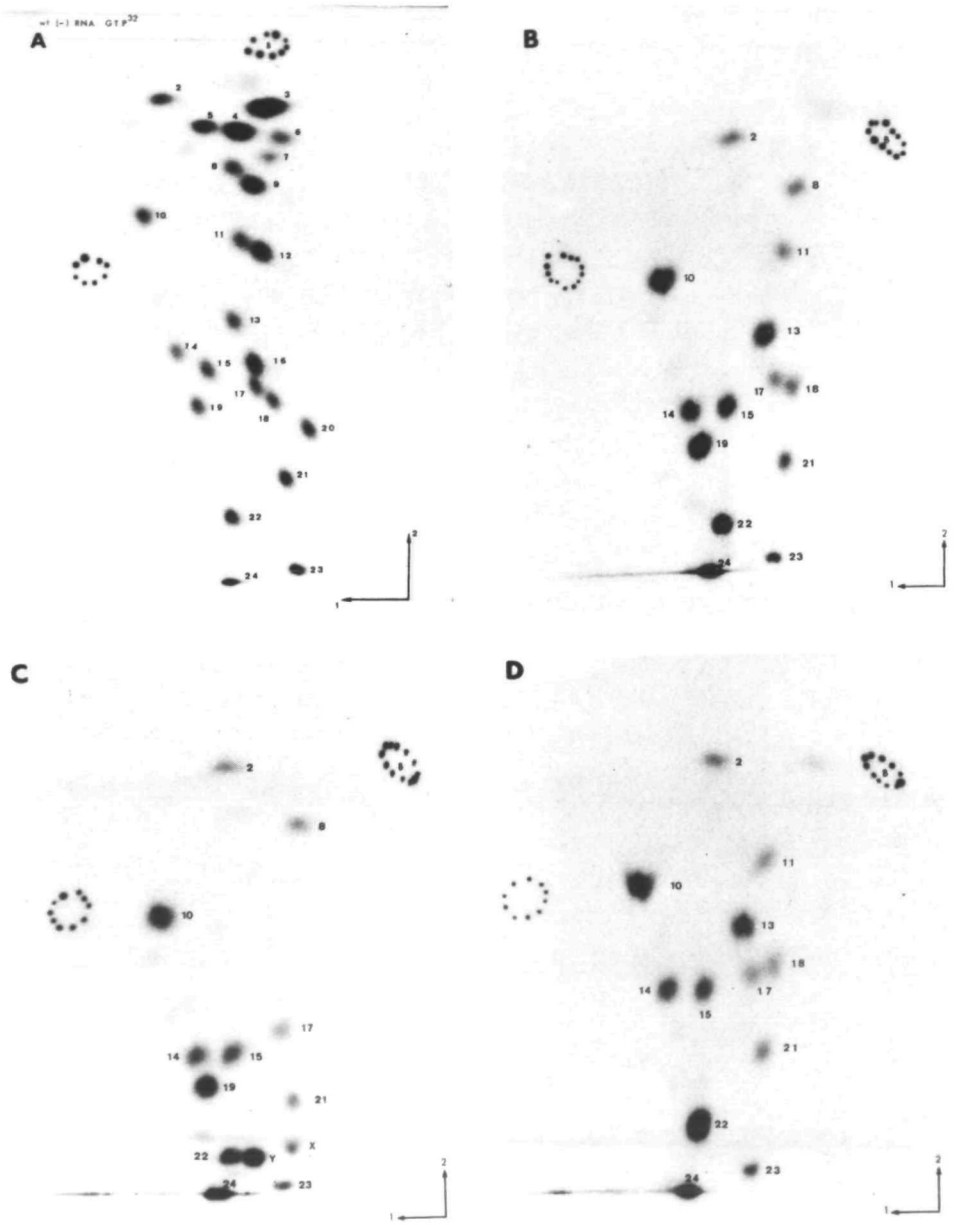

Fig. 5. Autoradiographs of the T_1-fingerprints of (-)RNA. Fingerprints of wild-type (-)RNA (A,B), *am*7-H2 (-)RNA (C) and *am*7-H3 (-)RNA (D). The RNA species have been labelled with (α-^{32}P)-GTP (A) or with (α-^{32}P)-UTP (B, C and D). Oligonucleotides 1a and 1b (Table 1) are not labelled under these labelling conditions.

Table 1. RNase A digestion products of T1 RNase resistant oligonucleotides.

Oligonucleotide[a]	(-)RNA labelled *in vitro* with (α-^{32}P)NTP[b]				Sequence[c]	Relative
	pppG	pppA	pppU	pppC		molar yields[d]
$T1_{a,b}$	-	G	-	G	G(A);G(C)	1;2
T2	U	-	G	-	UG(U)	1
$T3_{a,b}$	C	G	-	G	CG(A);CG(C)	4;1
$T4_{a,b}$	ÅĞ,ÅG	AĞ	-	-	AG(G);AG(A)	1;1
T5a	C	G	-	U	UCG(A)	1
b	C	-	-	U,G	UCG(C)	<0.5
T6	G,C	-	-	C	CCG(G)	<0.5
T7	ÅĞ	C	-	-	CAG(G)	<0.5
T8	U	G	C	C	CCUG(A)	1
T9	AÅĞ	ÅAG	-	-	AAG(G)	1
T10	U	-	G,U,ÅŬ	-	AUUUG(U)	1
T11	U	-	C	G,ÅĈ	ACCUG(C)	1
T12	AAÅG	ÅÅAĞ	-	-	AAAG(A)	1
T13	AŬ	C	G,C,ÅU	U,C	CUCCAUG(U)	1
T14	ÅG	U	2U,AĈ	AĞ,ÅC	UUACUUAG(C)	1
T15	C	U,C	ÅU	G,2AŬ	UAUCAUCG(C)	1
T16	G,AAĈ	U,AĈ,ÅAC	-	ÅC,AÅC	UACAACG(G)	1
T17	C	AAAU	AAÅU	G,C,AAAŬ	AAAUCCG(C)	1
T18	AĈ	G,U,C	AĈ	C,2ÅC	CCACUACG(A)	1
T19	U	AŬ,ÅÅAU	G,ÅU,AAÅŬ	-	AUAAAUUG(U)	1
T20	ÅG	C	-	3C,AĞ,ÅĈ	ACCCCCAG(C)	1
T21	AAÅG	ÅÅAG,ÅÅAĈ	AAAĞ	AAÅC	AAACAAAG(U)	1
T22	AÅG	U,C,AŬ,ÅAG	ÅŬ,ÅU	ÅĈ,AAĞ	AUUAUACCAAG(C)	1
T23	AAAAĈ	G,U,2C,ÅAC ÅÅÅAC	C	ÅĈ,AÅĈ,AAAÅC	C ACC,AACC UAAAACG(A)	1
T24	U	G,U,C,AĈ,ÅAŬ, ÅÅÅAĈ	2U,C,ÅU,2AĈ, AÅU	U,AŬ,3ÅC, AAAÅC	-	1
X	AAÅĞ	U,C,AĈ,ÅÅAG	AĈ	C,2ÅC	CCACUACAAAG(G)	1
Y	AŬ	U,C	G,C,ÅU,AĈ	U,C,ÅĈ,ÅC	ACCUACUCCAUG(U)	1
*am*T22	AÅG	U,C,AŬ,ÅAG	ÅŬ,ÅU	ÅĈ,AAĞ	AUUAUACCAAG(C)	1
	U	U,ÅAŬ,ÅÅAU	G,C,AÅU,AAÅŬ	C	CCUAAUAAAUUG(U)	1

a. The numbers refer to the fingerprints shown in Figure 5. X, Y and *am*T22 refer to oligonucleotides which are not present among the digestion products of (-)RNA transcribed from wild-type DNA, but which are present in the digestion products of (-)RNA transcribed from a "300-fragment" having an amber mutation in gene VII (*am*7-H2, X and Y; *am*7-H3, *am*T22).

b. pppG,pppA, etc. refer to the (α-^{32}P)-labelled ribonucleoside triphosphate precursor used to label (-)RNA.

c. The proposed nearest-neighbour bases are indicated in parenthesis. The underlined sequences were derived by partial spleen phosphodiesterase digestion analysis.

d. Relative molar yields were estimated from distribution of ^{32}P in nucleotides produced from RNAs labelled with each of four (α-^{32}P)NTP's.

nearest neighbour analysis were determined in conjunction with information obtained by partial spleen phosphodiesterase digestion of these oligonucleotides. No attempts were made to resolve completely the sequence of oligonucleotide T24.

All T1-oligonucleotides produced from (-)RNA occurred in one or more mole-equivalents, except for nucleotides UCG(C), CCG(G), and CAG(G) which were present in much lower amounts (0.2 - 0.4 moles). Since no unique T1-nucleotide containing the 3'OH-end was identified in the digest of (-)RNA and emphasizing that the 5'-end of the DNA template contains the endoR. *Hap*II recognition sequence (C.CGG) it is assumed that these extra nucleotides are incorporated to the 3'-end of the resulting transcript (-)RNA after transcription has reached to the end of the template. Evidence for such an aberrant termination of transcription at the template end is provided in the next section.

Ordering of T1-oligonucleotides of wild-type (-)RNA

Since G'-RNA is transcribed from position 90 → 300 on the "300-fragment" and (-)RNA most probably is transcribed from position 170 → 1 (in the opposite direction) a region of about 80 nucleotides in both transcripts is complementary to each other (*cf.* Fig. 3). This region represents the 5'-terminal end of the messenger RNA which codes for the precursor of the major capsid protein of phage M13, *i.e.* the product of gene VIII. The complete nucleotide sequence of this mRNA has recently been solved (23; Hulsebos and Schoenmakers, unpublished results). Initially, the sequence of this RNA was used to help specify the relative order of certain of the oligonucleotides of (-)RNA. About 50% of the T1-products, representing the 5'-terminal part of (-)RNA could be ordered in a unique sequence.

To order all RNase T1 products of wild-type (-)RNA we have deduced the DNA sequence of the region from which it is transcribed, using the chemical procedures introduced by Maxam and Gilbert (17). For this purpose, the fragments *Taq*I-H, *Hha*I-H, *Taq*I-C, *Hap*II-I$_1$ and the "300-fragment" were labelled at their 5'-hydroxy termini with (γ-^{32}P)-ATP and polynucleotide kinase. Each fragment was then cleaved with the appropriate restriction enzyme to produce fragments with a single 5'-labelled end. After electrophoretic separation, each labelled fragment was subjected to the dimethylsulphate hydrazine degradation procedure and the partial products were analysed on the sequencing gels. Representative autoradiographs of the sequencing gels are shown in Fig. 6. The DNA sequences derived from the

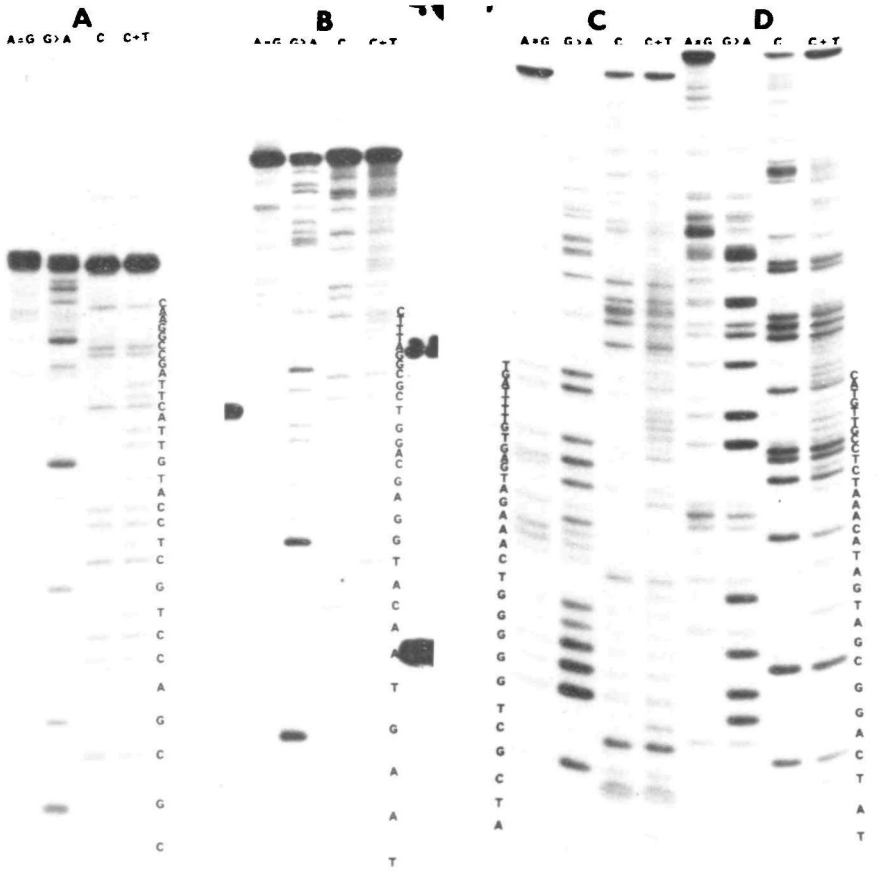

Fig. 6. Autoradiographs of DNA sequencing gels obtained after chemical degradation of: (A) the right-hand 5'-end of fragment *Taq*I-C; (B) the left-hand 5'-end of the "300-fragment"; (C) the left-hand 5'-end of the fragment *Hha*I-H and (D) the left-hand 5'-end of fragment *Taq*I-H. After labelling of the 5'-ends the first two fragments were digested with endoR. *Hha*I, while the latter two were digested with endoR. *Hae*III.

autoradiographs are presented in Fig. 7. As expected, the sequences revealed a considerable overlap with each other resulting in a complete unique sequence of about 330 base pairs. They also confirmed the sequence predicted from the known specificity of the restriction enzyme cleavage sites and the sequences at the 5'-end of the 300-fragment as deduced from partial hydrolysis with pancreatic DNase and snake venom phosphodiesterase.

Given the catalogue of all T1-RNase products obtained from the RNA sequence analysis of (-)RNA, as presented in Table 1, they can be ordered

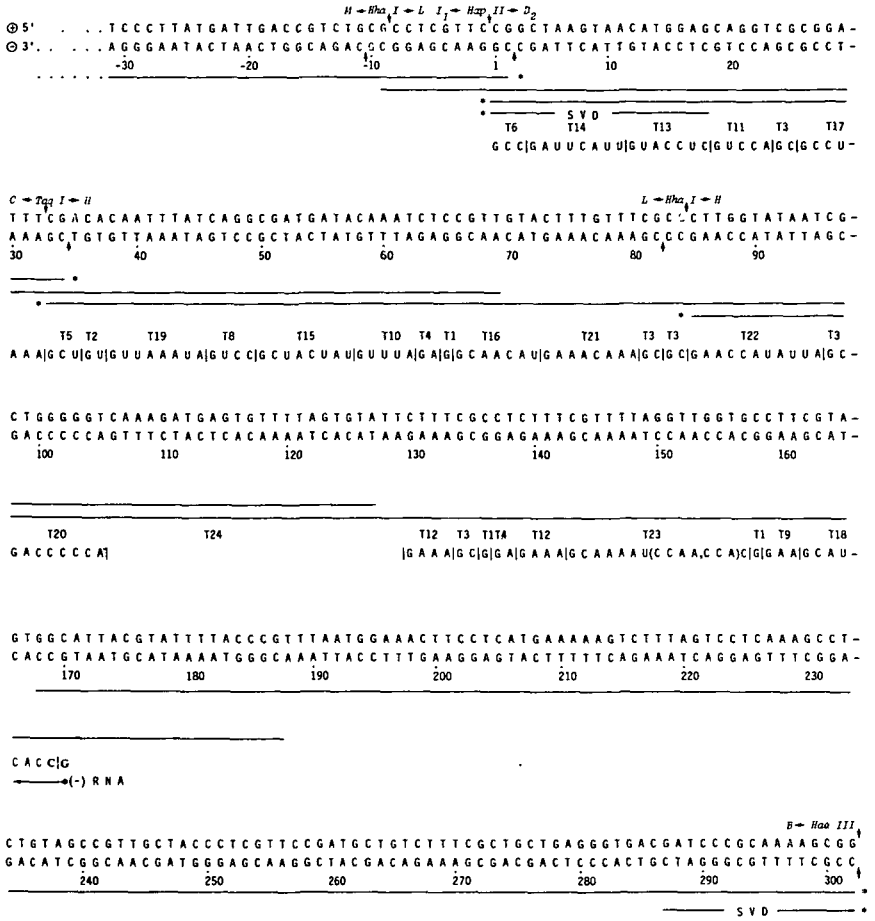

Fig. 7. Nucleotide sequence of the right-hand boundary of fragment Hap II-I_1, and of the "300-fragment" (*cf.* Fig. 1).
In this Figure the deduced nucleotide sequence of (-)RNA transcribed from the "300-fragment" is also indicated. ⊕, indicates the viral strand and ⊖, the complementary (codogenic) strand.
The underlined sequences are determined either with the chemical degradation method or by means of partial degradation with snake venom phosphodiesterase (SVD). The oligonucleotide numbers of (-)RNA correspond to those given in Table 1. The cleavage sites for the restriction endonucleases *Hha* I, *Hap* II, *Taq* I and *Hae* III are marked at the respective nucleotide positions. The capital letters above the sequence refer to the respective restriction fragments. Residue numbers indicate relative positions from the cleavage site for the restriction enzyme endo R.*Hap* II. The startpoint and direction of transcription of (-)RNA is indicated. The dinucleotide GpC used to prime the synthesis of (-)RNA is indicated with bold capital letters.

now within the sequence obtained from DNA sequence analysis. The results, summarized in Fig. 7, indicate that the RNA and DNA sequencing methods gave completely consistent data. All T1-oligonucleotides were represented by a complementary deoxy-oligonucleotide analogue except for the minor

products CCG(G), CAG(G) and UCG(C). Since the product CCG can only be derived from the *Hap*II-end of the template and we also noted that the other nucleotides CAG(G) and UCG(C) are just complementary to the sequence at position 19 → 26 of the complementary DNA strand, we assume that RNA polymerase, when reached to the end of the template, is able to switch transcription from one strand to the other. Our findings that these nucleotides are present in less than one molar yield, strongly suggest that reading back as well as termination of transcription after short complementary RNA chain growth does not occur at a fixed nucleotide position but more or less randomly. Although the mechanism involved in this reaction is quite unknown, its occurrence has already been noted by others (23). Also the pancreatic RNase digestion data of T24, the sequence of which was ambiguous, were completely consistent with those predicted from the DNA sequence deduced. We therefore feel confident to conclude that the sequence of 169 nucleotide residues representing (-)RNA is correct.

Localization of nucleotide sequence changes in amber-7 mutants

RNA sequencing techniques were used to determine exactly the positions of the single-point mutations within the nucleotide sequence of several amber mutants of gene VII. For this purpose, 300-fragments were isolated from M13 RF bearing hydroxylamine-induced amber-7 mutations and the individual 300-fragments were transcribed in the presence of the primer GpC and each of the four (α-^{32}P)-ribonucleoside triphosphates. The resulting (-)RNA species were isolated as described under Methods and subsequently digested with RNase T1. A T1-RNase fingerprint of (α-^{32}P)UTP-labelled (-)RNA transcribed from a 300-fragment bearing the *am*7-H2 mutation is shown in Fig. 5c. For comparison, a T1-RNase fingerprint of (α-^{32}P)UTP-labelled wild-type (-)RNA is also included in this figure (Fig. 5b). It is easily recognized that *am*7-H2 is a double mutant. The T1-fingerprint of *am*7-H2 (-)RNA namely contains two oligonucleotides, denoted (X) and (Y), which are not present in the fingerprint of wild-type (-)RNA. In turn, the wild-type T1-oligonucleotides T11, T13 and T18 are missing in the amber mutant (-)RNA. All T1-oligonucleotides of *am*7-H2 (-)RNA were further analysed by pancreatic RNase digestion and alkaline hydrolysis of the secondary digestion products. It turned out that all T1-oligonucleotides from the amber mutant (-)RNA were completely identical to those derived from wild-type (-)RNA (data not shown), except the oligonucleotides (X) and (Y). The nucleotide sequences of both T1-products, as deduced from the standard RNA sequencing analysis

are given in Table 1.

Given these nucleotide sequences and emphasizing that T13 is adjacent to T11 in the wild-type (-)RNA (*cf.* Fig. 7) one has to conclude that due to G → A transition in T11 a fusion product ACCUACUCCAUG is generated. A transition of G → A in (-)RNA, as a consequence of a C → T transition in the viral DNA strand is in accordance with the specificity of hydroxylamine-induced mutagenesis.

In analogy, the deduced sequence of oligonucleotide (X) suggests this sequence to be generated by a fusion of T9 and T18. The fusion is caused then as the result of a G → A transition in T18. Oligonucleotide T9, being AAG(G), will not be labelled by transcription with (α-^{32}P)-UTP and its disappearance, therefore, cannot be detected in the fingerprint shown in Fig. 5b.

Using the same approach, nucleotide sequence analyses have also been carried out with (α-^{32}P)UTP-labelled (-)RNA derived from a second amber-7 mutant, namely *am*7-H3. The T1-RNase fingerprints of this *am*7-H3 (-)RNA is shown in Fig. 5d. It is clear that in this case the wild-type oligonucleotides T8 and T19 are missing in the fingerprint of the *am*7-H3 mutant (-)RNA, whereas the relative intensity of T22 in the fingerprint of the mutant (-)RNA has markedly increased. Further analysis by pancreatic RNase digestion and alkaline hydrolysis have shown that the actual differences in both types of RNA are restricted to these oligonucleotides only. The pancreatic RNase digestion products of the T22-analogue present in the *am*7-H3 (-)RNA are shown in Table 1. This spot is actually composed of two oligonucleotides: the original oligonucleotide T22 and a new product CCUAAUAAAUUG(U). If it is emphasized that the sequence of T8 is CCUG(A) and of T19 is AUAAAUUG(U) and that both T1-products are contiguous , the new product then must be the fusion product of wild-type T8 and T19. This fusion is generated by a G → A transition in T8 which is the result of a C → T transition in the phage *am*7-H3 viral DNA strand.

Nucleotide sequence of gene VII

Since phage M13 messenger RNAs are solely transcribed from the non-viral strand of the DNA template, the DNA region analysed has been written for convenience in its codogenic RNA sequence. This is presented in Fig. 8. Also the nucleotide changes which are found in the amber mutants *am*7-H2 and *am*7-H3 are included in this figure. Both single-point mutations have changed a glutamine codon (CAG) into a nonsense codon (UAG). Both mutated

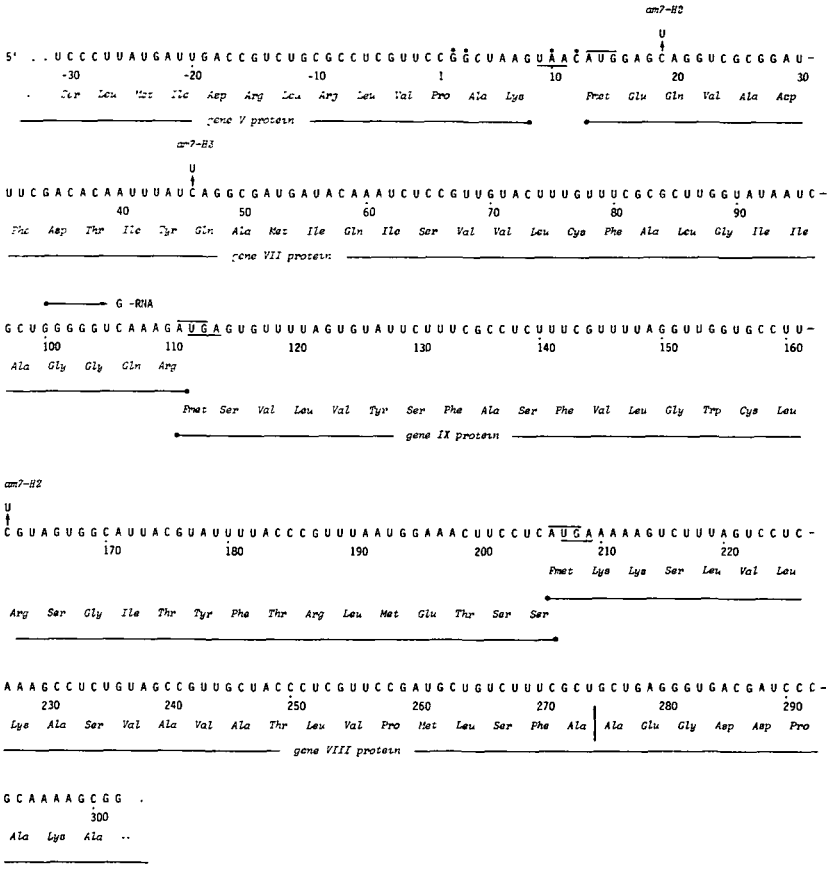

Fig. 8. Nucleotide sequence of the RNA complementary to the established nucleotide sequence of the codogenic strand presented in Fig 7. In this Figure the amino acid sequences of (a) gene VII protein, (b) the C-terminal part of gene V protein, the hypothetical gene IX protein, and of (d) the N-terminal part of the pre-coat protein encoded by gene VIII are indicated The vertical bar refers to the position where this pre-coat protein is cleaved during processing into the mature coat protein. Initiation codons for translation are overlined and termination codons are underlined. The transitions of the *am*7-H2 and *am*7-H3 mutations are marked above the positions 19,46 and 162 respectively (*cf.* Table 1).
The 5'-end of the RNA that codes for the pre-coat protein (G-RNA) is as indicated (*cf.* Fig. 3).
The nucleotides present within the ribosome binding site of gene VII and which are complementary to the 3'-terminus of 16S rRNA (29) are indicated with an asterisk.

codons are in phase to each other. Given these data, the coding region of (-)RNA and, hence of gene VII can be deduced now. As shown in Fig. 8, the initiation triplet AUG is at the 2nd triplet-frame upstream the mutated CAG codon in *am*7-H2, whereas a termination codon (UGA) is located 22 triplets downstream the mutated codon of *am*7-H3 and a second termination codon (UAG)

3 triplets more downstream the first termination codon. It is worth mentioning that in the *am*7-H2 mutant in fact two point mutations have been introduced. One C → U substitution in the CAG codon of gene VII and a second C → U substitution far more downstream and outside gene VII. Its position was found at 48 nucleotides downstream the UGA termination codon of this gene and within a region which has a hypothetical coding function (see discussion). Our inference that the AUG triplet mentioned is the initiation codon is based upon the observation that an in phase nonsense triplet UGA is present at 36 nucleotides upstream the initiation codon and that no other AUG or GUG codons are contained in this particular part of the sequence. However, a translational start of gene VII might be well further upstream if one considers the UGA codon at position -24 as a leaky terminator (*cf.* 29). Which of these two possibilities have to be favoured must await ribosomal binding studies and/or isolation and sequence determination of the gene VII-protein (see discussion).

From these sequence data it is further concluded that gene VII is separated from its next gene, *i.e.* gene VIII, by a region of 94 nucleotides. This region most probably has a coding function since it contains a purine-rich sequence with features of a ribosome binding site (24) and the appropriate codons for the initiation and termination of protein synthesis. The initiation codon (AUG) of this hypothetical gene forms an integral part of the termination signal of gene VII whereas its termination codon (UGA) overlaps with the initiation codon of gene VIII.

DISCUSSION

As a part of our studies on the regulation of transcription and translation of the small circular M13 genome we have determined the nucleotide sequence of gene VII. This has been achieved by a combination of RNA and DNA sequencing techniques. Although each approach generated certain ambiguities, in conjunction these methods gave completely consistent data and allowed deduction of an unambiguous sequence of about 330 nucleotides encompassing the complete gene VII, a still hypothetical gene, designated gene IX, and parts of the neighbouring genes V and VIII. It is shown that gene VII is only 99 nucleotides long and that this gene is immediately followed by gene IX which, in turn, partially overlaps with gene VII on one side and with gene VIII on the other.

The complete nucleotide sequence of phage fd has recently been established (25). A comparison of this sequence with the one reported in

this study for M13 shows that the sequences of the corresponding genes VII and IX are completely identical. This is quite different for several other genes of phage M13 and fd, the tentative sequences of which revealed several second-base substitutions and approximately 2-3% base-changes in the third position of the triplets. Hence, the gene VII and gene IX sequences in both filamentous phages are rather conserved.

As already pointed out in the Introduction, no clear explanation can be given yet for the undetectable level of gene VII-protein synthesis *in vivo* and *in vitro*. The low synthesis of this protein is not caused by a very low frequency of transcription of this gene. It has clearly been shown that transcription of the M13 genome occurs in a cascade-like fashion (8,20), which means that the amounts of the individual gene transcripts increase proportionally in the direction of (gene) II-V-VII-VIII with a maximal level of transcription at gene VIII. This suggests that the expression of gene VII is mainly controlled at the level of translation. A possible reason for the low synthesis of its product might be that gene VII has a low capacity of ribosome recognition. All potential ribosome binding sites of the genes of f1, fd and M13 have now been sequenced (25,26; van Wezenbeek *et al.*, unpublished results). They all show the potential to form Watson-Crick base pairs with the 3'-terminal end of 16S ribosomal RNA. From the sequence reported here it is clear that only the first, third, tenth and eleventh base upstream the initiation codon of gene VII are complementary to the 3'-end of 16S RNA (*cf.* Fig. 8). This means that gene VII has the lowest potential of Shine-Dalgarno base pairing among all filamentous phage genes. If there exists a correlation between the number of potential Shine-Dalgarno base pairs and the efficiency of ribosome binding (24,27), the low level of gene VII-protein synthesis could be explained by its low potential of base pairing with 16S RNA. That complementarity to 16S RNA, however, is not solely determining for ribosome binding has recently been evidenced by Taniguchi and Weissmann (28). They showed that also the interaction of the ribosome binding site with fMet-tRNA plays an essential role in the formation of the 70S initiation complex. Ribosome binding was substantially enhanced in case the first base following the AUG initiation triplet was mutated from a G into an A residue. From our sequence studies we know that all ribosome binding sites in M13 DNA have an A residue following their initiation triplet. Interestingly, the exceptions are gene VII (AUGG), gene VI (AUGC) and gene I (AUGG) of which it is known that protein synthesis under the direction of these genes is extremely low (3-5).

As already pointed out in the Results section, the predicted translational start of gene VII might be well further upstream if the UGA codon at position -24 is considered as a leaky terminator. In the nucleotide sequence of gene V of phage M13 and fd (25, Hulsebos and Schoenmakers, unpublished data) an AUG codon is present, which is preceded by a potential Shine-Dalgarno sequence, and which is in phase with the predicted initiation codon but not in phase with the gene V reading frame. If translation starts at this position (position -87), an overlap of the N-terminal region of gene VII distal to the leaky UGA terminator with the C-terminal end of gene V being an essential part for its function cannot be excluded. Furthermore, it seems likely that translation of gene V, the product of which is made in large amounts, would obstruct binding of ribosomes at the gene VII initiation site. This, together with a read-through at the UGA signal, might not only provide new insights in the polarity among genes V and VII but might also be a convenient control mechanism if gene VII protein is only required in small amounts.

From their nucleotide sequence studies Schaller and Takanami (25) were first to postulate that between genes VII and VIII a hitherto unidentified gene, *i.e.* gene IX, might be located. If true, this gene should code for a polypeptide which is 32 amino acids long and which does not contain the amino acid residues *His*, *Pro*, *Lys* and *Asp*. This property has enabled us to find out whether such a polypeptide exists among the products present in the M13-infected cells. Recently we have discovered that within the M13 virion a third capsid protein is present with a molecular weight of about 3300 (G. Simons, R. Konings and J. Schoenmakers, in preparation). The presence of this polypeptide, designated "C-protein", could not be detected in phage particles which were labelled with either *His*, *Pro* and *Lys*. Interestingly, a polypeptide with exactly identical properties has also been observed in minicells harbouring M13 RF (9; Simons *et al.* unpublished data). The latter polypeptide was absent, however, in minicells harbouring RF which carry a certain amber mutant in gene VII. Studies are in progress to substantiate our findings that C-protein originates from this hypothetical gene and that its synthesis is regulated by its proximal genes.

ACKNOWLEDGEMENTS

We thank Mrs. Joke de Witte for performing the SVD-digests analyses, Dr. Rob Schilperoort for providing us with RNA polymerase and Drs. Ruud Konings and Peter van Wezenbeek for critical comments and support. We also

thank Dr. Heinz Schaller for communicating the sequence of phage fd DNA. This research was supported by a grant from the Netherlands Foundation for Chemical Research (SON) with financial aid from The Netherlands Organization of Pure Research (ZWO).

REFERENCES

1. Denhardt, D.T. (1975) In: CRC Critical Reviews in Microbiology, The Chemical Rubber Company, Cleveland, Ohio, 161-223.
2. Ray, D.S. (1977) In: Comprehensive Virology, Fraenkel-Conrat, H. and Wagner, R.R., Eds. Plenum Publishing Corporation, New York, 105-178.
3. Konings, R.N.H., Hulsebos, T. and van den Hondel, C.A. (1975) Virology *15*, 570-584.
4. Model, P. and Zinder, N.D. (1974) J. Mol. Biol. *83*, 231-251.
5. Hondel van den, C.A.M., Konings, R.N.H. and Schoenmakers, J.G.G. (1975) Virology *67*, 487-497.
6. Edens, L., Konings, R.N.H. and Schoenmakers, J.G.G. (1978) J. Virol. (in press).
7. Henry, T.J. and Pratt, D. (1969) Proc. Nat. Acad. Sci. U.S.A. *62*, 800-807.
8. Edens, L., Konings, R.N.H. and Schoenmakers, J.G.G. (1978) Virology *86*, 354-368.
9. Smits, M.A., Simons, G., Konings, R.N.H. and Schoenmakers, J.G.G. (1978) Biochim. Biophys. Acta, in press.
10. Edens, L., van Wezenbeek, P., Konings, R.N.H. and Schoenmakers, J.G.G. (1975) Eur. J. Biochem. *70*, 577-587.
11. Edens, L., Konings, R.N.H. and Schoenmakers, J.G.G. (1975) Nucleic Acids Res. *2*, 1811-1820.
12. Hondel van den, C.A. and Schoenmakers, J.G.G. (1976) J. Virol. *18*, 1024-1039.
13. Hondel, van den, C.A. and Schoenmakers, J.G.G. (1976) Eur. J. Biochem. *68*, 55-70.
14. Hondel van den, C.A., Weijers, A., Konings, R.N.H. and Schoenmakers, J.G.G. (1975) Eur. J. Biochem. *53*, 559-567.
15. Brownlee, G.G. and Sanger, F. (1969) Eur. J. Biochem. *11*, 395-399.
16. Barrell, B.G. (1971) In: Procedures in Nucleic Acid Research, Cantoni, C.L. and Davies, D.R., Eds. Harper and Row, New York, Vol. 2, 751-795.
17. Maxam, A.M. and Gilbert, W. (1977) Proc. Nat. Acad. Sci. U.S.A. *74*, 560-564.
18. Jay, E., Bambara, R., Padmanabhan, R. and Wu, R. (1974) Nucleic Acids Res. *1*, 331-353
19. Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975) Proc. Nat. Acad. Sci. U.S.A. *72*, 1184-1188.
20. Konings, R.N.H. and Schoenmakers, J.G.G. (1978) In: Single-Stranded DNA Phages, Denhardt, D.T., Ray, D.T. and Dressler, D., Eds. Cold
21. Spring Harbor Laboratory, in press.
Cuypers, T., van der Ouderaa, F.J. and de Jong, W. (1974) Biochem. Biophys. Res. Commun. *59*, 557-563.
22. Asbeck, F., Beyreuther, K., Kohler, H., von Wettstein, G. and Braunitzer, G. (1969) Hoppe Seyler's Z. Physiol. Chem. *350*, 1047-1066.
23. Sugimoto, K., Sugisaki, H., Okamoto, T. and Takanami, M. (1977) J. Mol. Biol. *111*, 487-507.
24. Shine, J. and Dalgarno, L. (1974) Proc. Nat. Acad. Sci. U.S.A. *71*, 1342-1346.

25. Schaller, H. and Takanami, M. (1978) In: Single-Stranded DNA Phages, Denhardt, D.T., Ray, D.T. and Dressler, D., Eds. Cold Spring Harbor Laboratory, in press.
26. Pieczenik, G., Model, P. and Robertson, H.D. (1974) J. Mol. *90*, 191-214.
27. Steitz, J.A. and Jakes, K. (1975) Proc. Nat. Acad. Sci. U.S.A. *72*, 4734-4738.
28. Taniguchi, T. and Weissmann, C. (1978) J. Mol. Biol. *118*, 533-565.
29. Yates, J.L., Gette, W.R., Furth, M.E. and Nomura, M. (1977) Proc. Nat. Acad. Sci. U.S.A. *74*, 689-693.