2543

THE NUCLEOTIDE SEQUENCE OF SEVERAL GENES AND REGULATORY ELEMENTS ON THE BACTERIOPHAGE M 13 GENOME

T.J.M. Hulsebos

THE NUCLEOTIDE SEQUENCE OF SEVERAL GENES AND REGULATORY ELEMENTS ON THE BACTERIOPHAGE M13 GENOME

Promotor: Prof.Dr. J.G.G. Schoenmakers

THE NUCLEOTIDE SEQUENCE OF SEVERAL GENES AND REGULATORY ELEMENTS ON THE BACTERIOPHAGE M 13 GENOME

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN
DOCTOR IN DE WISKUNDE EN NATUURWETENSCHAPPEN
AAN DE KATHOLIEKE UNIVERSITEIT TE NIJMEGEN
OP GEZAG VAN DE RECTOR MAGNIFICUS
PROF DR P G A B WIJDEVELD
VOLGENS HET BESLUIT VAN HET COLLEGE VAN DECANEN
IN HET OPENBAAR TE VERDEDIGEN OP
VRIJDAG 12 DECEMBER 1980
DES NAMIDDAGS OM 2 00 UUR PRECIES

door

THEODORUS JOHANNES MARIA HULSEBOS

geboren te Raalte

1980

Druk Sneldruk Boulevard Enschede

Graag wil ik mijn erkentelijkheid betuigen aan allen die hebben bijgedragen aan de totstandkoming van dit proefschrift.

Ingevolge het promotiereglement van de Katholieke Universiteit is het helaas niet toegestaan leden van het wetenschappelijk corps te bedanken. De volgende personen mag ik echter wel met name bedanken:

Riet Matze en Gonda van Duren, die mij vooral in de initiële fase van het onderzoek geholpen hebben met het RNA sequentie werk.

Joke de Witte en Ben Peters, die in het kader van hun doktoraalstudie een wezenlijke bijdrage hebben geleverd aan het onderzoek, beschreven in de hoofdstukken III en VII.

Angélique Brackel voor het nauwgezet typen van de tekst van dit proefschrift.

De heer A. Derks van de afdeling Ontwerp voor het op de valreep tekenen van de figuren.

De heer G. Dekkers van de afdeling Fotografie voor de vakkundige verzorging van de vele foto's.

Bert Puyk voor het corrigeren van de Engelse tekst en in het bijzonder Els voor haar ondersteuning en aanmoediging gedurende de voorbereidingstijd van dit proefschrift.

This investigation was carried out in part under the auspices of the Netherlands Foundation for Chemical Research (SON) and with financial aid from the Netherlands Organization for the advancement of Science (ZWO).

Het in dit proefschrift beschreven onderzoek werd verricht in het Laboratorium voor Moleculaire Biologie van de Faculteit der Wiskunde en Natuurwetenschappen van de Katholieke Universiteit te Nijmegen onder leiding van Prof.dr. J.G.G. Schoenmakers.

CONTENTS

CHAPTER I

INTRODUCTION

CHAPTER I

INTRODUCTION

I.1. THE SINGLE-STRANDED DNA PHAGES

In bacteria and all higher organisms the genetic information is stored in the form of double-stranded DNA. In viruses, however, a whole scala of nucleic acids is used as genetic information carrier. The virus particle may contain single-stranded DNA, double-stranded DNA, single-stranded RNA or, even, double-stranded RNA.

Bacteriophage M13, the subject of this thesis, belongs to the single-stranded DNA containing phages. Based on their shape, these phages are subdivided into two groups: the icosahedral phages and the filamentous phages. The best characterized members of the first group are φX174, G4 and S13[1-3]. The second group can be subvided into two classes. The class I viruses infect only F- or I-pili containing *E.coli* strains. Representatives of the F-pili specific bacteriophages are M13, f1, fd and ZJ/2[4-7]. The bacteriophages If1 and If2[7,8] belong to the I-pili specific phages. The class II viruses differ from the class I viruses in their molecular architecture, as measured by X-ray diffraction studies.[9-12] Members of class II are Pf1 and Xf[13-14], which grow on *Pseudomonas aeruginosa* and *Xanthomonas oryzae*, respectively.

Another major difference between the icosahedral and filamentous phages lies in the process by which the progeny phages are 'released' into the surrounding medium. The icosahedral phages lyse their host, while the filamentous phages are extruded continuously without killing the host. On the other hand, both phage groups resemble each other in many aspects, for instance in the small size of their single-stranded DNA (mol. weight 2×10^6), in the number of genes contained within their

genome (about 9-10) and in the way their DNA is replicated.

In this Introduction, the main properties of the filamentous single-stranded DNA phages (abbreviated Ff) will be discussed. Special attention will be drawn to bacteriophage M13 and the closely related phages f1 and fd. Sometimes, reference will be made to analogous phenomena observed in the icosahedral phage group.

I.2. STRUCTURE OF THE Ff PHAGE

The wild-type Ff phage has a length of 900-1000 nm. Its circular, single stranded DNA (6400 bases, mol. weight ± 2×10^6) is contained within a central core, 2.5 nm in diameter, and surrounded by a protein cilinder with an 4 nm inner diameter and a 6 nm outer diameter. About 2700 B-protein molecules (mol. weight 5200, encoded by gene VIII) are present in the phage coat. They constitute 99% of the total protein content of the coat. The B-protein molecules form a left-handed helix around the central core with 4.5 molecules per turn[9-12]. The molecules overlap each other like the scales on a fish.

About four molecules of a minor coat protein, the A-protein (mol. weight 42700, encoded by gene III) are located at one end of the phage particle[15-18]. Recently, two additional minor proteins were detected in the coat of M13[19]. One of these proteins, C-protein (mol. weight 3300) is encoded by the newly discovered gene IX (Chapter III of this thesis). It cannot be excluded at this moment, that C-protein is in fact a doublet protein, not only encoded by gene IX but also by gene VII (G. Simons *et al*, personal communication). The second, additional minor coat protein, D-protein (mol. weight 11500) is encoded by gene VI, as could be demonstrated by amino acid analysis (R.E. Webster *et al*, personal communication) and Edman degradation (G. Simons *et al*, personal commincation) of the purified protein. Evidence is accumulating, that the

gene III- and gene VI-protein molecules are located at one end of the phage coat and the gene IX- (and possibly also gene VII-)protein molecules at the other hand (R.E. Webster *et al* , personal communication).

Although the normal length of the Ff virion is 900-1000 nm, this figure can vary considerably. A few percent of the phages in a wild-type stock are of double length. They contain predominantly two unit-length circles[20-22] but about 10% of these double-length phages contain double-length molecules[23]. Also miniphages can be found in a wild-type phage stock. These phages vary in size from 20 to 50% of the normal unit-length and contain correspondingly smaller DNA[24-26]. A third abnormal form, the midiphage, is also present although in low amount in a wild-type stock. Midiphages vary in size from 1.2 to 1.8x the normal phage-length and their single-stranded DNA is correspondingly larger[27]. The great flexibility in the size of the Ff phages, largely determined by the amount of DNA to be packaged, has made them very suitable as cloning vehicles. Large pieces of foreign DNA can be inserted into the duplex replicative form DNA of the Ff phages by recombinant DNA techniques[28-36] (P. van Wezenbeek, personal communication).

The progeny phages, which can be obtained quite easily in large amounts after transfection of *E.coli*, contain the inserted DNA in the single-stranded form. This is of great advantage for a number of biochemical and genetical purposes, for instance DNA sequence determination of the inserted DNA by the chain-termination method[37-40] and introduction of various mutations in the inserted DNA by the site-directed mutagenesis approaches[41-44].

1.3. GENES AND GENE-PRODUCTS OF THE Ff PHAGE

The genetic content of the Ff phages was initially analyzed by Pratt *et*

al [45,46].

They isolated amber and temperature-sensitive mutants of bacteriophage M13 and used them in standard complementation assays. These studies revealed the presence of eight complementation groups (genes) on the M13 genome. The relative positions of the corresponding complementation groups on the circular genetic map of f1 was determined by recombination experiments[47,48]. The precise alignment of the individual genes could be obtained by a combination of two very powerful technical approaches. At first, *in vitro* DNA-dependent protein synthesizing systems were developed (see below). With these systems it became possible to identify various gene-products of the filamentous phages and, consequently, to estimate the approximate lengths of several genes. Secondly, the discovery of base-sequence specific restriction enzymes[49,50] has allowed the production of well-defined fragments of the phage genome. Several physical maps were constructed by a variety of methods[51-62].

The genetic map was correlated with the physical maps by application of the marker-rescue technique[51,52,54,57,63] and by *in vitro* transcription-translation studies of individual restriction fragments[52,64,65]. The positions of the *M13 genes* and the physical maps of the restriction enzymes *Hap II* and *Hae III* are shown in Fig 1.

The detection and identification of the gene-products of the Ff phages *in vivo* is seriously hampered by the fact that host-protein synthesis continues after infection at nearly normal rate, which results in a very high background of protein synthesis. Initially, it was tried to circumvent this difficulty by U.V.-irradiation of the cells before infection[15,66,67]. However, the development of *in vitro* DNA-dependent protein synthesizing systems has considerably facilitated the identification of the various gene-products[68-71].

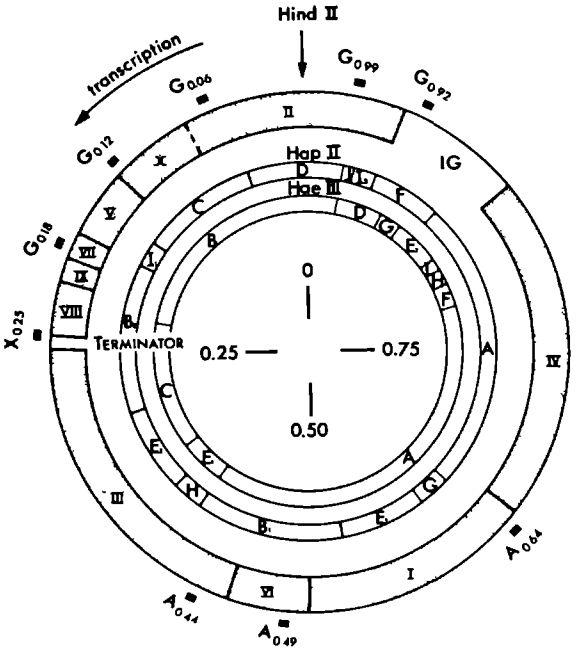


Fig 1 Circular genetic map and physical maps of the bacteriophage M13 genome. The Roman numerals in the dotted circular area represent the M13 genes. The direction of transcription is as indicated. The approximate positions of the G-promoters, the A-promoters and the promoter in front of gene III ($X_{0.25}$) are indicated by black bars. The latter promoter forms an integral part of the rho-independent central terminator. IG refers to the intergenic region in which the replication origins for both the viral and complementary strand are located. The single site at which M13 RF is cleaved with restriction enzyme *Hind II* is indicated. The inner circles represent the physical maps obtained after cleavage of M13 RF with restriction enzymes *Hap II* and *Hae III*.

By comparison of the proteins synthesized by coupled transcription-translation *in vitro* of M13 replicative form DNA, derived from wild-type and amber mutant phages, Konings *et al* [71] could unambiguously identify the proteins encoded by six of the eight M13 genes (*i.e.* genes I, II, III, IV, V and VIII). In parallel studies by Model and Zinder[72] the corresponding gene-products of bacteriophage were identified. A seventh

protein, the so-called X-protein, was detected in both of these studies, but could not be assigned to any of the known genes. The products of the genes VI and VII could not be identified by the *in vitro* protein synthesizing systems.

The molecular weights of the six identified M13 gene-products, as determined from their mobility in SDS-polyacrylamide gels are listed in Table I (column 4). The molecular weights given for the gene V- and gene VIII-protein were calculated from their estimated amino acid sequences[73-79].

The introduction of new DNA sequencing techniques[80,81] made it possible to resolve the sequence of large pieces of DNA in a relatively short time. By means of these new techniques the complete nucleotide sequence of M13 DNA has been established in our lab[82] (this thesis). The coding regions of all M13 genes were determined by establishing the nucleotide changes, caused by amber mutation(s) in the respective genes (this thesis[82-84]). In this way it was possible to locate the positions of all M13 genes at the nucleotide level and to predict also the sizes of the proteins, encoded by the genes VI and VII, which could not be identified with the aid of *in vitro* protein synthesis studies. The exact location of gene X could be deduced (Chapter V of this thesis). In addition, the sequencing studies revealed the presence of a new M13 gene, gene IX (Chapter III of this thesis). The latter gene, located between gene VII and gene VIII (Fig 1) encodes C-protein, one of the recently discovered minor capsid proteins[19].

DNA sequence analysis has also allowed the precise localization of the Intergenic Region (I.G. in Fig 1). This region, of which the existence was already demonstrated by van den Hondel *et al* [62] in M13 and by Zinder and co-workers[52,53] in Fig 1 is the only part of the Ff phage genome, that does not encode proteins. It contains the origins of replic-

Table 1 GENE AND GENE-PRODUCTS OF BACTERIOPHAGE M13

gene	nucleotides	protein molecular weight from DNA	protein molecular weight from protein	physiological role
I	1044	39,500	36,000[71]	morphogenesis?
II	1230	46,117	46,000[71]	DNA replication single-strand DNA synth.
III	1272 (1213)	44,748 (42,675)	59,000[71] (56,000)[71]	minor capsid protein
IV	1278	45,791	48,000[71]	morphogenesis?
V	261	9,666	9,688[73-75]	single-strand DNA synth.
VI	336	12,264	-	morphogenesis? minor capsid protein?
VII	99	3,587	-	morphogenesis? minor capsid protein?
VIII	219 (150)	7,622 (5,234)	5,800[71] (5,196)[76-79]	major capsid protein
IX	96	3,654	3,300[19]	minor capsid protein
X	333	12,670	12,000[71]	unknown

ation of the viral and complementary DNA strand[85-87].

As already demonstrated by Konings *et al* [71], the gene III- and gene VIII-protein made *in vitro*, are larger than their *in vivo* counterparts, as isolated from the phage coat, suggesting that these proteins are formed as precursors. This could be confirmed by nucleotide sequence analysis of the regions encoding the precursors (Chapter IV of this thesis, see ref 79,83,84,88) and by partial amino acid sequence determinations of the N-terminal extensions, present in both precursors[17,89]. The 18 extra N-terminal amino acid residues of the gene III-precursor and the 23 extra N-terminal amino acid residues of the gene VIII-precursor are split off by (a) signal peptidase(s)[89-93], thereby generat-

ing the mature form of both proteins.

The sizes of the M13 genes and the molecular weights of the gene-products, both deduced from the nucleotide sequences of these genes, are listed in Table 1, column 2 and 3, respectively. In general, there is reasonable agreement between the estimated and calculated molecular weights of the M13-specific proteins, except for the protein encoded by gene III. The molecular weight of the latter protein, deduced from the nucleotide sequence is considerable lower than the molecular weight, estimated from its electrophoretic mobility in SDS-polyacrylamide gels[71,72]. This discrepancy is most probably due to the unusual structural features of the gene III-protein, resulting in an anomalous electrophoretic behaviour[84].

I.4. THE REPLICATION CYCLE OF THE Ff PHAGE

From electronmicroscopic observations[94,95] it has been concluded that infection of male *E.coli* cells by the Ff phages starts with the attachment of one end of the phage to the tip of an F-pilus. After attachment, retraction of the F-pilus occurs and the virus penetrates into the cell. Gene III-protein, of which only a few copies are present in the virion, is needed for this adsorption process[96]. However, alternative schemes have been proposed for the initial stage of infection.

Especially, the important role of the F-pilus in the adsorption of the phage is seriously doubted[97]. Anyway, during penetration the viral DNA is decapsidated and enters the replication cycle. The capsid protein molecules, encoded by gene VIII, are deposited in the bacterial inner membrane and can be re-used during the assembly of the progeny viruses[98]. Based on the observation that gene III-protein remains attached to the viral DNA, even after its conversion to the replicative

form, it has been suggested that this protein serves as a 'pilot-protein', guiding the DNA to a place on the inner membrane where DNA replication is initiated[96]. The replication of the viral DNA can be divided into three stages (schematically drawn in Fig 2).

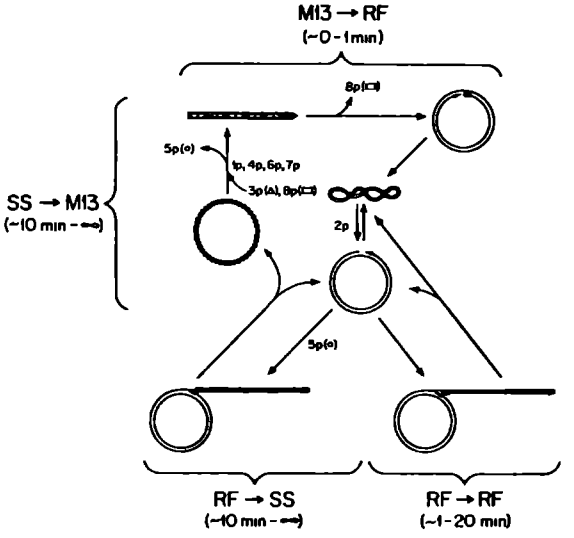


Fig 2 The replication cycle of filamentous phage. The three stages of replication of filamentous phage are indicated schematically along with an indication of the approximate duration of each stage and the requirements for viral gene products. (Outer circle) viral strand; (inner circle) complementary strand; (arrow head) 3'-OH terminus; (⟿) RNA; (□) gene-VIII protein, 8p; (o) gene-V protein, 5p; (Δ) gene-III protein, 3p. (From Ray, ref 185)

i Conversion of the viral, single-stranded DNA (ss) into the parental replicative form (ss→RF).

ii Replication of the RF molecule, resulting in a pool of daughter RF molecules (RF→RF).

iii Synthesis of progeny single-strands by an asymmetric replication process, in which the complementary strands of the RF molecules serve as templates for the repeated displacements of new viral strands (RF→ss).

Only host-functions are involved in the first stage of the replication cycle. Based upon *in vitro* studies[99] and the observation that the formation of parental RF is inhibited by rifampicine[97,100], it is generally accepted now, that synthesis of the complementary strand starts with the synthesis of a short primer RNA. This RNA, about 30 nucleotides long[101], is synthesized in the Intergenic Region (Fig 1) by the rifampicine-sensitive *E.coli* RNA polymerase. Complementary strand synthesis by DNA polymerase III-holoenzyme is initiated at the 3'-end of the primer RNA. After a full round of replication the primer RNA is removed and replaced by DNA through the combined action of the 5'→3' exonuclease and polymerase activity of DNA polymerase I. The complementary strand is closed by *E.coli* DNA ligase. The resulting relaxed replicative form (RF IV) is converted into the supertwisted form (RFI) by the host DNA-gyrase[102,103].

The thus formed parental RFI is replicated in the second stage of the replication cycle. Up to 200 progeny RF molecules can be produced per cell during the first 10-20 min of infection[104]. Synthesis of the phage encoded gene II-protein is required in this stage, because infection in the presence of chloramphenicol, an inhibitor of protein synthesis, or infection of a non-permessive host with amber mutants in gene II does not proceed beyond the parental RF formation[105]. In addition to the phage-encoded gene II-protein, a number of host-functions are involved in the RF replication[100,105-113]. The RF replication is clearly associated with the host membrane system[114-116] and occurs most probably by a Rolling Circle mechanism[117]. It starts with the introduction of a nick in the viral strand of the parental RF[118,119]. Gene II-protein, which acts as a nicking-closing enzyme is responsible for this step.

The position of the nick, *i.e.* in the Intergenic Region, has been de-

termined recently[120]. The supertwisted parental replicative form (RFI) is by the action of gene II-protein converted in an open circular form (RFII). In a subsequent step, the generated 3'-OH terminus in the viral strand serves, most probably, as primer for the synthesis of a new viral strand in a direction opposite to the direction of the complementary strand synthesis. The old viral strand is displaced during synthesis of the new one. After a full round of replication, the completely displaced viral strand is split off and circularized. Gene II-protein is also involved in the latter processes[118,119], analogous to the multifunctional action of the gene A-protein of ϕx174[121,122]. On the displaced circular viral strand a new complementary strand is synthesized. The resulting RFII-molecules are converted, *via* the relaxed closed intermediate form (RFIV), into RFI molecules by the action of the host ligase and gyrase.

The switch-over from RF replication to the third stage of the replication cycle, the synthesis of progeny viral strands, is entirely determined by the amount of gene V-protein in the cell[123-125]. Besides gene V-protein, the gene II-protein and several host-functions[108,112,113,126-128] are indispensable in this stage of the replication cycle.

As replication proceeds, more RF molecules become available for transcription and translation. This results in an enormous increase in the amount of gene V-protein in the infected cell. Gene V-protein binds, cooperatively, to the viral strands which are displaced on the replicating RF molecules and prevents synthesis of the complementary strands[129,130]. Gene II-protein is, most probably, involved in the split-off and circularization of the displaced viral strands.

At 60-75 minutes after infection, the host cell contains about 200 progeny viral strands[131], each of which are covered with approximately

1300 gene V-protein molecules[66,132]. The viral DNA - gene V-protein complexes pass through the inner membrane, whereby the gene V-protein molecules are replaced by the capsid protein molecules which have accumulated during infection. The 1300 gene V-protein molecules are replaced by about 2700 gene VIII-protein molecules (and a few molecules of the minor capsid proteins). The gene V-protein molecules are released into the cytoplasm and can be re-used in the binding of newly displaced viral strands[132].

It has been shown genetically that assembly of the phages requires the products of the genes I, III, IV, VI, VII and VIII[7,45]. The products of genes III, VI, VII, VIII (and also IX) are structural components of the phage coat. Whether they exert additional functions in the assembly process is not known at this moment. Also, the function of the gene I- and gene IV-product remains to be clearified.

I.5. EXPRESSION OF THE Fϕ PHAGE GENOME

The RF molecules, produced during infection, are implicated in the various stages of the replication cycle or serve as template for the synthesis of the phage-specific mRNA's by the host-RNA polymerase. Only the complementary strands of the RF molecules are transcribed[7,133-140]. Transcription proceeds counterclockwise on the genetic map, as indicated in Fig 1[65,71,72].

Until now, the transcription process has best been studied *in vitro*. Transcription of RF *in vitro*, in the absence of the transcriptional termination factor *rho*, results in the synthesis of a discrete set of RNA's, ranging in size from 8S (360 nucleotides) to 26S (5000 nucleotides). In addition, a heterogenous mixture of RNA molecules with a size equivalent to or larger than one genome

length is formed[136-138,141]. The 8S to 26S RNA species are transcribed by a cascade-like mechanism: the various messengers initiate at different promoters on the genome, but they all terminate at a single site, the *rho*-independent central terminator. As was demonstrated by Edens *et al* [137], the latter regulatory element is located immediately distal to gene VIII. Edens *et al* [138] located nine different promoters on the M13 genome. These *in vitro* promoters, as well as the promoters on the phage f1 and fd DNA have been mapped by several independent methods[57,65,71,72,136-138,142-147]. Although the initially deduced positions of the promoters on the respective Ff phage genomes differed markedly, there is now general agreement about the positions shown in Fig 1.

With the exception of genes VII (and IX), all M13 genes are preceded by a promoter. In addition, gene II contains two internal promoters, of which one preceeds the presumptive gene X. *In vitro* transcription studies with γ-^{32}P-labelled ribonucleoside triphosphates have shown that the smaller RNA's, initiating at positions 0.18, 0.12, 0.06, 0.99 and 0.92 on the physical map, all start with pppG. The starting nucleotide of the large RNA's, initiating at map positions 0.64, 0.49 and 0.44, proved to be pppA[137,138]. The starting nucleotide of the RNA, initiating at position 0.25, has not been determined unambiguously, but is most probably pppU[148].

The strength of the various promoters has been approximated by several independent methods[137,138,149]. On the average the G-promoters are stronger than the A-promoters. The strength, as determined from the amount of γ-^{32}P-labelled ribonucleoside triphosphate incorporated in the respective RNA's, decreases in the order $G_{0.06} > G_{0.18} \geq G_{0.92} > G_{0.99} \geq G_{0.12}$ for the G-promoters and $A_{0.64} > A_{0.49} \geq A_{0.44}$ for the A-promoters[137,138]. The strength of the $X_{0.25}$ promoter has not been determined yet.

As a result of the cascade mode of transcription, the region preceding the central terminator, which contains the coding information for the gene V-, VII-, IX- and VIII-protein, is most frequently transcribed *in vitro*. Consequently, an abundant synthesis of gene V- and gene VIII-protein has been found in DNA-dependent transcription-translation systems to which RF-DNA was added[71,72]. Genes VII and IX, however, which are positioned between genes V and VIII, are not expressed to a detectable level, despite the great number of transcripts, containing the coding information for these genes. Also, gene VI-protein has never been detected in these *in vitro* systems, although the products of the neighbouring genes, gene III and gene I, are synthesized. These data suggest, that the expression of the Ff phage genome is, at least *in vitro*, not only regulated at the transcriptional level, but also at the level of translation.

Much present research about the Ff phage gene expression is concerned with the question, how far the above discussed model, based mainly on *in vitro* studies, reflects the *in vivo* situation. Earliest studies by Jacob *et al* [133] have indicated that the maximum size of the M13-specific mRNA's, isolated from the infected cell, is smaller than the maximum size of the RNA species, synthesized *in vitro*. Similar observations have been made for the M13-mRNA's synthesized in RF containing minicells[134,135].

Also the observation of Edens[150] that the transcriptional termination factor *rho* changes the mode of Ff phage transcription *in vitro*, does already indicate that the transcription process *in vivo* is more complex.

A first successful attempt to elucidate the *in vivo* transcription pattern of the filamentous phage genome was performed by Rivera *et al* [151]. By sequence analysis of the 5'- and 3'-terminal end of 8S RNA, isolat-

ed from the infected cell, they could demonstrate that at least this RNA species is identical to its *in vitro* counterpart, *i.e.* initiates at promoter $G_{0.18}$ and terminates at the central terminator. Detailed hybridization studies of Smits *et al* [152] have indicated that six of the nine promoters detected *in vitro* are also operative in the infected cell. The hybridization studies furthermore indicated that the region in front of the central terminator is, analogous to the *in vitro* situation, most frequently transcribed *in vivo*. The majority of the phage-specific RNA species, present in the infected cell, direct the synthesis of at least the gene VIII-protein in an RNA dependent translation system[152-155].

All major *in vivo* RNA's share a 3'-OH terminal end which is identical to the 3'-OH terminal end of the *in vivo* 8S RNA species, as expected for a transcriptional termination event at the central terminator[151,152,155]. By analysis of the 5'-ends of the *in vivo* transcripts, it was demonstrated that several phage-specific mRNA's are not primary transcripts, but are the products of some type of RNA processing[152,154,155].

RNA processing might explain the absence of very long phage-specific RNA species in the infected cell. An additional explanation might be the presence of *rho*-dependent transcriptional termination signals on the Ff phage genome. Although direct proof for the existence of *rho*-dependent termination sites on the M13 genome is still lacking, the hybridization data of Smits *et al* [152] suggest that the regions immediately distal to gene VI and gene IV encompass such sites. These signals would cause a considerable shortening of the long RNA's. Interestingly, the region immediately distal to gene IV coincides with the region where Edens[150], studying the influence of the termination factor *rho* on the *in vitro* transcription of M13 RF, has located a *rho*-dependent term-

ination signal. Whether processing and additional *rho*-dependent termination of transcription can sufficiently explain the whole scala of phage-specific mRNA's in the infected cell requires, however, further investigation.

I.6. NUCLEOTIDE SEQUENCE ANALYSIS OF DNA

For many years the sequence analysis of nucleic acids has primarily been focussed on RNA. The main reason for this preference was the availability of small RNA molecules with defined lengths which could easily be degraded to smaller pieces by base-specific ribonucleases. The virtual absence of base-specific deoxyribonucleases and the fact that even the smallest DNA molecules are very large (more than 5000 nucleotides) have greatly hampered the development of DNA sequencing techniques.

The first direct DNA sequence determinations were perfomed by pyrimidine tract analysis of unfragmented DNA and by repair DNA synthesis on the 5'-protruding cohesive ends of various temperate bacteriophages (reviewed in ref 156). Specific, small DNA fragments were initially obtained by the binding of macromolecules, such as RNA polymerase[142,143,157], the λ and lac repressor[158,159] and ribosomes[160,161] to DNA and subsequent digestion of the unprotected DNA by DNases. Another method, applied in the earlier sequencing studies of φx174 DNA, was limited digestion of the viral DNA with T4 endonuclease IV[162-165]. The sequences of the small DNA fragments were then deduced by pyrimidine tract analysis, partial digestion with exonucleases or analysis of transcripts.

The discovery of a new class of deoxyribonucleases, the so-called restriction enzymes, marks a turning point in the sequence analysis of DNA. Especially, the type II restriction enzymes proved to be very useful. These enzymes generally recognize a stretch of 4-6 nucleotides and

split double-stranded DNA within or close to the recognition sequence[166]. With the aid of these restriction enzymes it became possible to cleave almost every DNA region into well-defined fragments. Since then, a number of DNA sequencing procedures have been developed which make use of restriction fragments.

I.6.1. DNA sequence analysis by the transcription method

Initially no general procedures were available for the direct sequencing of restriction fragments. To circumvent this problem, restriction fragments were transcribed *in vitro* by *E.coli* RNA polymerase in the presence of four ribonucleoside triphosphates, of which one was ^{32}P-labelled in the α-position. The nucleotide sequence of the transcripts were then determined by standard RNA sequencing procedures[167, 168]. The ^{32}P-labelled transcripts were digested with RNase T1, which cleaves specifically after G residues. The resulting oligonucleotides were separated by two-dimensional homochromatography: electrophoresis on a cellulose-acetate strip in the first dimension, followed by homochromatography on a DEAE-cellulose thin layer in the second dimension. Each nucleotide was then further analyzed by digestion with pancreatic RNase (or RNase A) which cleaves specifically after C and U residues. The pancreatic RNase products were hydrolyzed by alkali treatment. The latter procedure gives information about the transfer of label to the 'nearest neighbour'. The sequencing procedure was then repeated with each of the other ribonucleoside triphosphates as labelled precursor. In a similar way the pancreatic RNase products of ^{32}P-labelled transcripts were analyzed by RNase T1 digestion and subsequent alkali hydrolysis of the RNase T1 products. The sequence of the majority of oligonucleotides could be determined with these methods. Additional methods, such as partial digestion with exonucleases or digestion with other

base-specific ribonucleases, were often necessary to deduce the sequence of large oligonucleotides or oligonucleotides with a peculiar sequence arrangement. The complete digestion products were ordered by partial digestion of the transcript and by using the overlapping sequence information of the RNase T1- and pancreatic RNase-oligonucleotides.

Native and denatured fragments have been transcribed (see for example ref 143-145, 169-174). In the latter method the ordering of the RNase T1 (and pancreatic RNase) oligonucleotides is facilitated by the fact that these oligonucleotides originate from both strands and, consequently, must be complementary. Controlled asymmetrical transcription of native fragments can often be obtained by the addition of a short primer to the transcription mixture[143,165,169,171,175]. Primer dependent transcription and analysis of oligonucleotides with the standard RNA sequencing procedures have been performed in Chapter III and IV of this thesis.

I.6.2. The mobility shift method

In this method a DNA (or RNA) chain, terminally labelled with ^{32}P at the 5'- or 3'-end, is partially degraded with an exonuclease under such conditions that a complete series of degradation products is generated[176-181]. Partial digestion can be obtained by incubation with snake venom phosphodiesterase, which removes sequentially 5'-(d) NMP residues in the 3'→5' direction, or by incubation with spleen phosphodiesterase, which starts at the 5'-end and removes 3'-(d)NMP residues. The partial degradation products are fractionated in the two-dimensional homochromatography system, described in I.6.1. In this fractionation system the removal of a single nucleotide from an oligonucleotide of N residues causes a shift of in the position of the N-1 product which is characteristic for that nucleotide (see Fig 3). The sequence of a nucleic acid

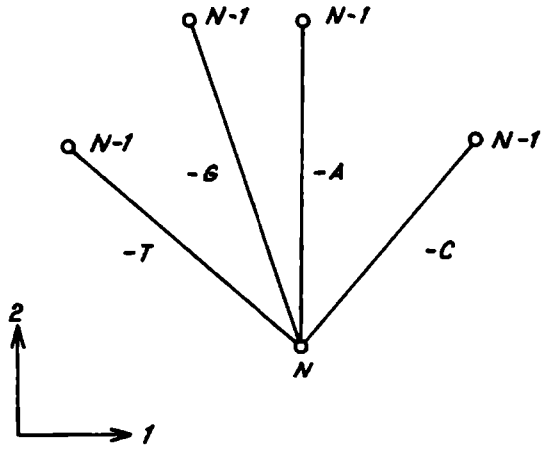


Fig 3 The mobility shift method.
Expected effect of the removal of a single nucleotide on the position of an oligonucleotide.
1. Electrophoresis on cellulose acetate at pH 3.5
2. Homochromatography on DEAE-cellulose thin layer

chain can thus directly be read from the mobility shifts of its degradation products in the two-dimensional homochromatography system.

The mobility shift method was originally used in the analysis of products obtained by 'primer extension'[156,181]. In the latter method DNA polymerase is allowed to produce small labelled fragments by extension of the primer in a DNA primer/template complex. The method is still useful for the sequence determination of the 5'- or 3'-terminal ends of restriction fragments. However, due to the low resolving power of the two-dimensional fractionation system no more than 25 nucleotides (starting at the labelled end) can be determined.

This technique was applied in the Chapters III and IV for the sequence determination of RNase T1 oligonucleotides and in the Chapters IV and V for the sequence determination of the 5'-ends of restriction fragments.

I.6.3. The 'plus and minus' method

DNA sequencing by analysis of transcripts (and by partial degradation) is very laborious, especially in case longer molecules are studied. The repeated fractionation and analysis is very time-consuming and one often ends up with insufficient amounts of material for the final analysis. To overcome this problem, several new DNA sequencing techniques were developed during the last five years. With these methods it became possible to deduce a DNA sequence directly from the fractionation system, thereby avoiding further analytical steps. As these methods heavily rely on the availibility of small defined DNA fragments, their development become only possible after the introduction of the restriction enzymes.

The three major fast DNA sequencing methods, *i.e.* the 'plus and minus' method, the chain-termination method and the Maxam-Gilbert method will be described here.

The general principle of the 'plus and minus' method[182] is shown in Fig 4. A primer-template is made by annealing a restriction fragment (or a synthetic oligodeoxynucleotide) to single-stranded DNA. DNA polymerase I is allowed to extend the primer for a limited time in the presence of all four deoxyribonucleoside triphosphates, of which one is labelled with ^{32}P. This will result in a mixture of extension products, all with the same 5'-end but differing at their 3'-ends. Ideally this mixture should be as complex as possible, with all extension products present. The mixture is passed through a gel filtration column to remove the excess deoxyribonucleoside triphosphates and then incubated under various conditions in the 'plus' or 'minus' system.

In the 'minus' system re-incubation with DNA polymerase I is carried out in the presence of only three deoxyribonucleoside triphosphates. For instance, in the '-A system', dATP is omitted. Each chain in the

heterogenous mixture will be extended now until a position is reached where a dAMP residue should be incorporated. After fractionation of the products and autoradiography, a banding pattern will appear, in which each band represents an oligodeoxynucleotide which would be followed by a dAMP residue. Similar incubations are carried out in the absence of one of the other deoxyribonucleoside triphosphates (-G, -C and -T system). The four incubation mixtures are denaturated and fractionated in parallel by electrophoresis on a denaturating polyacrylamide slab gel.

Under these conditions the fractionation of the oligonucleotides is according to size. The sequence of the DNA can directly be read from the autoradiograph. As the 'minus' system is in practice not always completely reliable, Sanger and Coulson[182] also developed the 'plus' system to confirm and supplement the sequence deduced in the 'minus' system. In the 'plus' system, the heterogenous mixture of extension products is incubated with T4 DNA-polymerase and only one of the four deoxyribonucleoside triphosphates. In the absence of deoxyribonucleoside triphosphates the enzyme acts as an exonuclease, which degrades the DNA, starting at the 3'-end. In the presence of deoxyribonucleoside triphosphates its polymerizing activity prevails. Incubation of the heterogenous mixture in the presence of only dATP (in the + A system) will result in a degradation of each chain until the position of an A residue is reached. At this position a dAMP residue is incorporated faster than it is removed. The products of the +A reaction are thus oligodeoxynucleotides with a dAMP residue at their 3'-ends. Analogous incubations are performed in the presence of each of the other deoxyribonucleoside triphosphates. The four 'plus' mixtures are run in parallel with the 'minus' mixtures. In addition, the heterogenous mixture of extension products without enzyme treatment is applied to the gel (-O- lane in Fig 4).

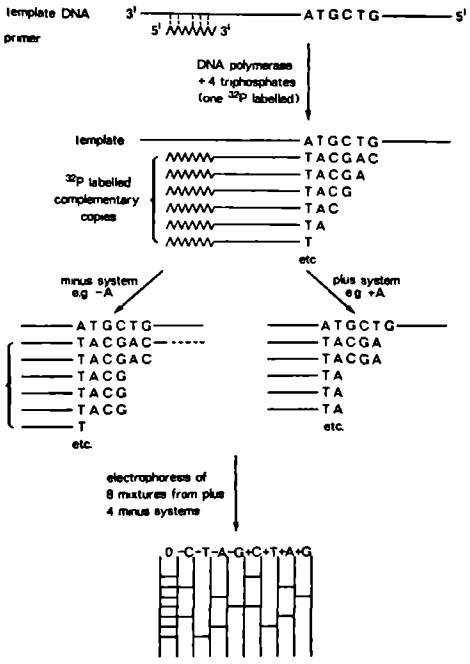


Fig 4 The principle of the 'plus and minus' method (from Sanger and Coulson, see ref 182).

As is shown in Fig 4, each oligodeoxynucleotide is represented by two bands, one in the 'plus' system, which defines its 3'-terminus and one in the 'minus' system, which defines the next residue. For instance, the lowest -T band indicates that the oligodeoxynucleotide is followed by a T residue at its 3'-end. This oligonucleotide is represented by a band in the +T lane. The latter band co-migrates with a band in the -A lane, indicating that dAMP is the next nucleotide, which is confirmed by the presence of the one nucleotide longer band in the +A system, and so on.

The resolving power of the polyacrylamide gel system is limited by

the lengths of the extension products which are separated. In case a short synthetic oligodeoxynucleotide is used as primer, there is no need to remove it. However, the most suitable primers are restriction fragments. These are split off by incubation with a restriction enzyme before application of the 'plus' and 'minus' mixtures to the gel. The main disadvantage of the method is that tracts of the same residue within the DNA sequence are represented by a gap in the reading pattern. It is often difficult to establish the exact number of residues which have caused the gap, although in theory each residue should be represented by a band in the untreated mixture (the -O- lane).

1.6.4. The chain-termination method

Two years after the introduction of the 'plus and minus' method, Sanger *et al* [37] published a new DNA sequencing technique, which is even more rapid and more accurate. Similar to the 'plus and minus' method, a primer is annealed to a single-stranded template DNA and extended with DNA polymerase I (Klenow fragment). Base-specific extension products are obtained by inclusion of the 2',3'-dideoxy- or arabinonucleoside analogues of the deoxyribonucleoside triphosphates (see Fig 5).

For instance, primer extension in the presence of deoxythymidine triphosphate (dTTP) and 2',3'-dideoxythymidine triphosphate (ddTTP) as well as the other three deoxyribonucleoside triphosphates (of which one is labelled in the α-position with ^{32}P), will occasionally result in the incorporation of the dideoxy analogue. As the latter contains no free 3'-hydroxyl group, the chain cannot be extended further and chain-termination will occur. Under appropriate conditions (depending on the ddTTP/dTTP ratio) a heterogenous mixture of extension products is obtained, all with the same 5'-end but with ddTMP residues at their 3'-ends.

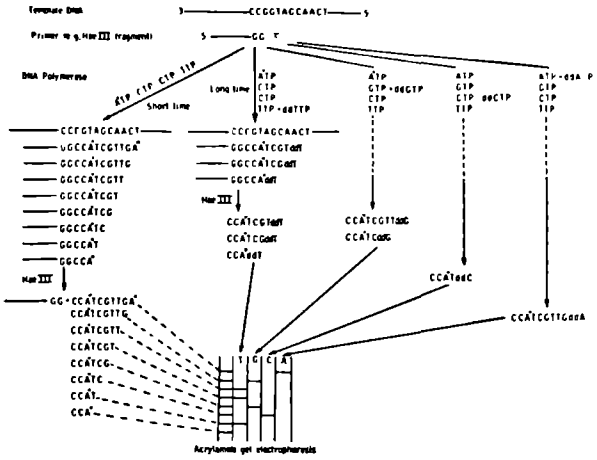


Fig 5 The principle of the chain-termination method (from Sanger *et al* , see ref 37)

In an analogous way, chain extensions are performed in the presence of one of the other dideoxy derivatives and the four dNTP's. After removal of the primer fragment, the four incubation mixtures are run in parallel on a denaturating polyacrylamide slab gel. A mixture including all extension products is also applied to the gel to serve as a reference. After autoradiography, the nucleotide sequence can directly be read from the gel.

A main advantage of this method, as compared to the 'plus and minus' method is the fact that intermediate nucleotides in a stretch of the same residue show up as individual bands. Moreover, the method is more

rapid and simple to perform and requires less material, as better incorporation of the ^{32}P-label is obtained. The chain-termination method is only applicable to single-stranded DNA. However, two variations on this method, the exonuclease method of Smith[183] and the nick translation method of Maat and Smith[184], have been devised which enable the sequence termination of double-stranded DNA templates. The (original) chain-termination method was used in Chapter VI to deduce the positions of several amber mutations in gene V of the bacteriophages M13, f1 and fd at the nucleotide level.

I.6.5. The Maxam and Gilbert method

In the same year as the chain-termination method was published (1977), Maxam and Gilbert[80] introduced a chemical method for sequencing DNA. The starting material for this method is single- or double-stranded RNA, labelled at only one end with ^{32}P. For this purpose, a restriction fragment is first labelled at its 5'-ends with (γ-^{32}P)-ATP and T4 polynucleotide kinase. To obtain DNA, labelled at only one end, the restriction fragment is subsequently cleaved by another restriction enzyme or the strands are separated.

The method, essentially, consists of three consecutive chemical reactions: a specific modification of one of the four bases, removal of the modified base from its sugar and DNA strand scission at that sugar. For instance, the G residues are specifically methylated with dimethyl sulphate (Fig 6a). In a second reaction, the methylated G residues are removed. Finally, a break in the DNA strand is introduced by an elimination reaction in which the sugars without base are cleaved from the neighbouring phosphates. Partial chemical cleavage is obtained by choosing conditions under which only a few percent of the bases are specifically modified. After removal of the modified bases and strand scis-

sion a whole series of overlapping molecules is obtained, all with the same ^{32}P-labelled end, but with different 3'-ends (exemplified for the G residues in Fig 6B).

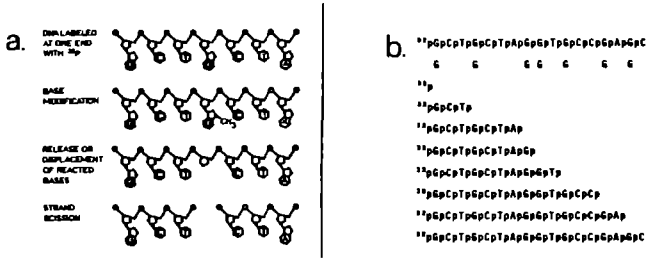


Fig 6 The principle of the Maxam-Gilbert method (from Maxam and Gilbert, see ref 186).

A ^{32}P-labelled DNA fragment is subjected to four, separate, base-specific chemical cleavage reactions. The four different sets of products are electrophoresed in parallel on a denaturating polyacrylamide gel. Only the degradation products which contain the labelled end are visualized by subsequent autoradiography. The sequence can directly be read from the banding pattern that emerges. Base-specific cleavages can be obtained in as much as twelve different ways[186]. Of these, six have been applied in the nucleotide sequencing studies described in the Chapters III and V of this thesis and only these will be explained in more detail.

The strong guanine weak/adenine cleavage

Dimethyl sulphate methylates the guanines at the N7 position and the adenines at the N3 position. The glycosidic bond of a methylated purine is unstable and breaks on heating at neutral pH, leaving a free sugar moiety. Treatment with 0.1 M alkali at 90°C will cleave the sugar from the neighbouring phosphate groups. After fractionation and autoradiography of the end-labelled products, a pattern of dark and light

bands will appear. The dark bands originate from the breakage at G residues, as these methylate 4-10 times faster than the adenines.

The strong adenine/weak guanine cleavage

The glycosidic bond of a methylated adenosine is less stable than that of a methylated guanosine. Gentle treatment with 0.1 N HCl releases the adenines preferentially. Cleavage at the sugar moieties with 0.1 N alkali and fractionation of the products results in dark adenine and weak guanine bands on the autoradiograph.

Strong adenine/weak cytosine cleavage

Treatment of the DNA with strong alkali opens the adenine and (to a less extent) the cytosine rings. The ring-opened products are displaced and the phosphate groups eliminated with piperidine. After fractionation, the adenine cleavage products appear as dark bands on the autoradiograph, while the cytosine cleavage products are weak.

Cleavage at guanine

Cleavage, exclusively at guanine residues can be obtained by heating the methylated DNA in 1.0 M piperidine. This reaction opens the 7-methyl guanine ring, displaces it from the sugar and eliminates the phosphates from the free sugar moiety. Only guanine cleavage products appear on the autoradiograph.

Cleavage at cytosine and thymine

Hydrazine reacts with cytosine and thymine, cleaving the base and leaving ribosylurea. Hydrazine may react further to produce a hydrazone. After partial hydrazinolysis in 15-18 M aqueous hydrazine at 20°C, the DNA is cleaved with 0.5 M piperidine. This cyclic secondary amine, as the free base, displaces all the products of the hydrazine reaction from the sugars and catalyzes the β-elimination of the phosphates. Cytosine and thymine cleavage products appear as evenly intensive bands on the autoradiograph.

Cleavage at cytosine

The presence of 2 M NaCl preferentially suppresses the reaction of thymines with hydrazine. Then, the piperidine breakage produces bands only from cytosine.

The Maxam-Gilbert method is somewhat more laborious than the chain-termination method. However, an apparent advantage of the first method is that it requires no enzymatic copying *in vitro*. The latter method may sometimes lead to erroneous results, due to the problem in faithfull copying with DNA polymerase I of certain template DNA's, such as DNA's with strong secondary structures. A second advantage of the Maxam-Gilbert method is that conformation of a deduced sequence can, in general, easily be obtained by nucleotide sequencing of the opposite strand.

1.7. The aim of the present investigation

The region on the bacteriophage M13 genome in front of the *rho*-independent central terminator of transcription is most frequently transcribed. In accordance with this frequent transcription, the products of genes V and VIII are synthesized in much larger amounts than the other phage-specific proteins. Gene VII which is located between genes V and VIII, is also frequently transcribed. Nevertheless, its product has not been detected among the M13-specific gene-products either in the infected cell or among the products synthesized in *in vitro* cell-free systems. The reasons for the absence of a detectable level of gene VII-protein synthesis are not understood.

Another puzzling aspect of the expression of the M13 region, which comprises the genes II to VIII, is that under the direction of this region a protein is synthesized *in vitro*, designated X-protein, which is

not synthesized at a detectable level in the infected cell.

The undetectable level of gene VII-protein synthesis on one hand and the differential expression of a DNA region which encodes X-protein on the other might be attributed to either a regulation at the transcriptional or at the translational level.

In order to get more insight into the genetic organization of the gene X-gene VIII region and in the structural features of the regulatory elements, which are operative in this region, we have determined its DNA sequence.

To do this, it first proved to be necessary to extend our collection of restriction enzyme cleavage maps. The construction of two additional maps, the *Hha* I and *Mbo* II restriction enzyme cleavage maps is described in Chapter II.

The precise location and the nucleotide sequence of gene VII is determined in Chapter III. In this Chapter standard RNA sequencing methods are used to determine the positions of two amber mutations in gene VII, with which the reading frame of gene VII can be deduced.

Classical RNA sequencing techniques are also used in Chapter IV to sequence a small RNA, which arises from the gene VII-VIII region, but the genetic information of which is too small to encode one of the characterized M13 gene-products.

In Chapter V, the entire nucleotide sequence of the region comprising the genes X, V, VII, IX and VIII has been elucidated with the Maxam-Gilbert method of DNA sequencing. From the deduced DNA sequence the positions and sequences of genes and regulatory elements in this region could be deduced at the nucleotide level.

In Chapter VI the amber mutations in several amber5-mutants are localized by DNA sequencing of the respective phage DNA's using the chain-termination method. The two types of amber5-mutants which emerge from

the sequencing studies are applied in two subsequent studies: at first, they are used to investigate the regulatory mechanism which underlies the gene V-gene VII polarity phenomenon (Chapter VI). Secondly, the amber mutants are applied to generate various specifically modified gene V-proteins, which probably can be used as a tool for studying the mechanism of DNA-gene V-protein interaction.

1. Sertic, V. and Boulgakov, N. (1935) C.R. Soc. Biol. *119*, 1270-1272.
2. Godson, G.N. (1974) Virology *58*, 272-289.
3. Burnet, F.M. (1927) Br. J. Exp. Pathol. *8*, 121-129.
4. Hofschneider, P.H. (1962) Z. Naturforsch. *18b*, 203-205.
5. Loeb, T. (1960) Science *131*, 932-933.
6. Hoffmann-Berling, H., Marvin, D.A. and Durwald, H. (1963) Z. Naturforsch. *18b*, 876-883.
7. Marvin, D.A. and Hohn, B. (1969) Bacteriol. Rev. *33*, 172-209.
8. Meynell, G.G. and Lawn, A.M. (1968) Nature *217*, 1184-1186.
9. Marvin, D.A., Wiseman, R.L. and Wachtel, E.J. (1974) J. Mol. Biol. *82*, 121-138.
10. Marvin, D.A., Pegram, W.J., Wiseman, R.L., Wachtel, E.J. and Marvin, F.J. (1974) J. Mol. Biol. *88*, 581-600.
11. Berkowitz, S.A. and Day, L.A. (1976) J. Mol. Biol. *102*, 531-547.
12. Day, L.A., Wiseman, R.L. and Marzec, C.J. (1980) Nucleic Acids Res. *7*, 1393-1403.
13. Minamishima, Y., Takeya, K., Ohnishi, Y. and Amako, K. (1968) J. Virol. *2*, 208-213.
14. Kuo, T.-T., Haung, T.-C. and Chow, T.-Y. (1969) Virol. *39*, 548-554.
15. Pratt, D., Tzagoloff, H. and Beaudoin, J. (1969) Virol. *39*, 42-53.
16. Marco, R. (1975) Virol. *68*, 280-283.
17. Goldsmith, M.E. and Konigsberg, W.H. (1977) Biochemistry *16*, 2686-2694.
18. Woolford, J.L., Jr, Steinman, H.M. and Webster R.E. (1977) Biochemistry *16*, 2694-2700.
19. Simons, G.F.M., Konings, R.N.H. and Schoenmakers, J.G.G. (1979) F.E.B.S. Letters *106*, 8-12.

20. Scott, J.R. and Zinder, N.D. (1967). In: The molecular biology of viruses, Colter, J.S. and Paranchych, W., Eds. Academic Press, New York, pp 211-218.

21. Salivar, W.O., Henry, T. and Pratt, D. (1967) Virol. *32*, 41-51.

22. Beaudoin, J. and Pratt, D. (1974) J. Virol. *13*, 466-469.

23. Wheeler, F.C., Benzinger, R.H. and Bujard, H. (1974) J. Virol. *14*, 620-630.

24. Griffith, J. and Kornberg, A. (1974) Virol. *59*, 139-152.

25. Hewitt, J.A. (1975) J. Gen. Virol. *26*, 87-94.

26. Enea, V. and Zinder, N. (1975) Virol. *68*, 105-114.

27. Wheeler, F.C. and Benzinger, R.H. (1978) In: the single-stranded DNA phages, Denhardt, D.T., Dressler, D. and Ray, D.S., Eds. Cold Spring Harbor Laboratory, 571-581.

28. Messing, J. and Gronenborn, B. (1977) Hoppel-Seyler's Z. Physiol. Chem. *358*, 278.

29. Messing, J., Gronenborn, B., Müller-Hill, B. and Hofschneider, P.H. (1977) Proc. Natl. Acad. Sci. *74*, 3642-3646.

30. Ray, D.S., and Kook, K. (1978) Gene *4*, 109-119.

31. Herrmann, R., Neugebauer, K., Zentgraf, H. and Schaller, H. (1978) Mol. Gen. Genet. *159*, 171-178.

32. Winter, G. and Fields, S. (1980) Nucleic Acids Res. *8*, 1965-1974.

33. Herrmann, R., Neugebauer, K., Pirkl, E., Zentgraf, H. and Schaller, H. (1980) Mol. Gen. Genet. *177*, 231-242.

34. Zacher III, A.N., Stock, C.A., Golden II, J.W. and Smith, G.P. (1980) Gene *9*, 127-140.

35. Kaguni, J., LaVerne, L.S. and Ray, D.S. (1979) Proc. Natl. Acad. Sci. U.S.A. *76*, 6250-6254.

36. Kaguni, J. and Ray, D.S. (1979) J. Mol. Biol. *135*, 863-878.

37. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. *74*, 5463-5467.

38. Barnes, W.M. (1979) Gene *9*, 127-139.

39. Anderson, S., Gait, M.J., Mayol, L. and Young, I.G. (1980) Nucleic Acids Res. *8*, 1731-1745.

40. Heidecker, G., Messing, J. and Gronenborn, B. (1980) Gene *10*, 69-73.

41. Razin, A., Hirose, T., Itakura, K. and Riggs, A.D. (1978) Proc. Natl. Acad. Sci. U.S.A. *75*, 4268-4270.

42. Hutchison III, C.A., Phillips, S., Edgell, M.H., Gillam, S., Jahnke, P. and Smith, M. (1978) J. Biol. Chem. *253*, 6551-6560.

43. Gillam, S. and Smith, M. (1979) Gene *8*, 81-97.

44. Gillam, S. and Smith, M. (1979) Gene *8*, 99-106.

45. Pratt, D., Tzagoloff, H. and Erdahl, W.S. (1966) Virology *30*, 397-410.

46. Pratt, D. and Erdahl, W.S. (1968) J. Mol. Biol. *37*, 181-200.

47. Lyons, L.B. (1971) Ph.D. thesis, The Rockefeller University, New York.

48. Lyons, L.B. and Zinder, N.D. (1972) Virology *49*, 45-60.

49. Smith, H.O. and Wilcox, K.W. (1970) J. Mol. Biol. *51*, 379-391.

50. Kelly, T.J., Jr. and Smith, H.O. (1970) J. Mol. Biol. *51*, 393-409.

51. Horiuchi, K., Vovis, G.F., Enea, V. and Zinder, N.D. (1975) J. Mol. Biol. *95*, 147-165.

52. Vovis, G.F., Horiuchi, K. and Zinder, N.D. (1975) J. Virol. *16*, 674-684.

53. Enea, V., Horiuchi, K., Turgeon, B.G. and Zinder, N.D. (1977) J. Mol. Biol. *111*, 395-414.

54. Ravetch, J., Horiuchi, K. and Model, P. (1977) Virology *81*, 341-351.

55. Ravetch, J.V., Horiuchi, K. and Zinder, N.D. (1977) Proc. Natl. Acad. Sci. U.S.A. *74*, 4219-4222.

56. Vovis, G.F. and Lacks, S. (1977) J. Mol. Biol. *115*, 525-538.

57. Seeburg, P.H. and Schaller, H. (1975) J. Mol. Biol. *92*, 261-279.

58. Takanami, M., Okamoto, T., Sugimoto, K. and Sugisaki, H. (1975) J. Mol. Biol. *95*, 21-31.

59. Hondel van den, C.A. and Schoenmakers, J.G.G. (1973) Mol. Biol. Rep. *1*, 41-45.

60. Hondel van den, C.A. and Schoenmakers, J.G.G. (1975) Eur. J. Biochem. *53*, 547-558.

61. Hondel van den, C.A. and Schoenmakers, J.G.G. (1976) J. Virol. *18*, 1024-2039.

62. Hondel van den, C.A., Pennings, L. and Schoenmakers, J.G.G. (1976) Eur. J. Biochem. *68*, 55-70.

63. Hondel van den, C.A., Weyers, A., Konings, R.N.H. and Schoenmakers J.G.G. (1975) Eur. J. Biochem. *53*, 559-567.

64. Model, P., Horiuchi, K., McGill, C. and Zinder, N.D. (1975) Nature *253*, 132-134.

65. Hondel van den, C.A., Konings, R.N.H. and Schoenmakers, J.G.G. (1975) Virology *67*, 487-497.

66. Webster, R.E. and Cashman, J.S. (1973) Virology *55*, 20-38.

67. Lin, N.S.-C. and Pratt, D. (1974) Virology *61*, 334-342.

68. Konings, R.N.H. (1973) F.E.B.S. Lett., *35*, 155-160.

69. Konings, R.N.H., Jansen, J., Cuypers, T. and Schoenmakers, J.G.G. (1973) J. Virol. *12*, 1466-1472.

70. Konings, R.N.H. and Schoenmakers, J.G.G. (1974) Mol. Biol. Rep. *1*, 251-256.

71. Konings, R.N.H., Hulsebos, T. and Hondel van den, C.A. (1975) J. Virol. *15*, 570-584.

72. Model, P. and Zinder, N.D. (1974) J. Mol. Biol. *83*, 231-251.

73. Cuypers, T., Ouderaa van der, F.J. and Jong de, W.W. (1974) Biochem. Biophys. Res. Commun. *59*, 557-563.

74. Nakashima, Y., Dunker, A.K., Marvin, D.A. and Konigsberg, W. (1974) F.E.B.S. Lett. *40*, 290-292.

75. Nakashima, Y., Dunker, A.K., Marvin, D.A. and Konigsberg, W. (1974) F.E.B.S. Lett. *43*, 125.

76. Asbeck, F., Beyreuther, K., Köhler, H., Wettstein von, G. and Braunitzer, G. (1969) Hoppe-Seyler's Z. Physiol. Chem. *350*, 1047-1066.

77. Snell, D.T. and Offord, R.E. (1972) Biochem. J. *127*, 167-178.

78. Nakashima, Y. and Konigsberg, W. (1974) J. Mol. Biol. *88*, 598-599.

79. Bailey, G.S., Gillett, D., Hill, D.F. and Petersen, G.B. (1977) J. Biol. Chem. *252*, 2218-2225.

80. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. *74*, 560-564.

81. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. *74*, 5463-5467.

82. Wezenbeek van, P.M.G.F., Hulsebos, T.J.M. and Schoenmakers, J.G.G. (1980) Gene *11*, 129-148.

83. Hulsebos, T. and Schoenmakers, J.G.G. (1978) Nucleic Acids Res. *5*, 467-4698.

84. Wezenbeek van, P. and Schoenmakers, J.G.G. (1979) Nucleic Acids Res. *6*, 2799-2818.

85. Tabak, H.F., Griffith, J., Geider, K., Schaller, H. and Kornberg, A. (1974) J. Biol. Chem. *249*, 3049-3054.

86. Horiuchi, K. and Zinder, N.D. (1976) Proc. Natl. Acad. Sci. U.S.A. *73*, 2341-2345.

87. Suggs, S. and Ray, D.S. (1977) J. Mol. Biol. *110*, 147-164.

88. Sugimoto, K., Sugisaki, H., Okamoto, T. and Takanami, M. (1977) J. Mol. Biol. *111*, 487-507.

89. Chang, C.N., Blobel, G. and Model, P. (1978) Proc. Natl. Acad. Sci. U.S.A. *75*, 361-365.

90. Chang, C.N., Model, P. and Blobel, G. (1979) Proc. Natl. Acad. Sci. U.S.A. *76*, 1251-1255.

91. Mandel, G. and Wickner, W. (1979) Proc. Natl. Acad. Sci. U.S.A. *76*, 236-240.

92. Ito, K., Mandel, G. and Wickner, W. (1979) Proc. Natl. Acad. Sci. U.S.A. *76*, 1199-1203.

93. Ito, K., Date, T. and Wickner, W. (1979) J. Biol. Chem. *255*, 2123-2130.

94. Caro, L.G. and Schnös , M. (1966) Proc. Natl. Acad. Sci. U.S.A. *56*, 126-132.

95. Jacobson, A. (1972) J. Virol. *10*, 835-843.

96. Jazwinski, S.M., Marco, R. and Kornberg, A. (1973) Proc. Natl. Acad. Sci. U.S.A *70*, 205-209.

97. Marco, R., Jazwinski, S.M. and Kornberg, A. (1974) Virology *62*, 209-223.

98. Smilowitz, H. (1974) J. Virol. *13*, 94-99.

99. Geider, K. and Kornberg, A. (1974) J. Biol. Chem. *249*, 3999-4005.

100. Brutlag, D., Schekman, R. and Kornberg, A. (1971) Proc. Natl. Acad. Sci. U.S.A. *68*, 2826-2829.

101. Geider, K., Beck, E. and Schaller (1978) Proc. Natl. Acad. Sci. U.S.A. *75*, 645-649.

102. Marians, J.K., Ikeda, J., Schlagman, S. and Hurwitz, J. (1977) Proc. Natl. Acad. Sci. U.S.A. *74*, 1965-1968.

103. Yasumoto, C.S. and Hurwitz, J. (1977) Proc. Natl. Acad. Sci. U.S.A. *74*, 4195-4199.

104. Hohn, B., Lechner, H. and Marvin , D.A. (1971) J. Mol. Biol. *56*, 143-154.

105. Fidanián, H.M. and Ray, D.S. (1972) J. Mol. Biol. *72*, 51-63.

106. Mitra, S. and Stallions, D.R. (1976) Eur. J. Biochem. *67*, 37-45.

107. Olsen, W.L., Staudenbauer, W.L. and Hofschneider, P.H. (1972) Proc. Natl. Acad. Sci. U.S.A. *69*, 2570-2573.

108. Staudenbauer, W.L. (1974) Eur. J. Biochem. *49*, 249-256.

109. Staudenbauer, W.L., Olsen, W.L. and Hofschneider, P.H. (1973) Eur. J. Biochem. *32*, 247-253.

110. Dasgupta, S. and Mitra, S. (1976) Eur. J. Biochem. *67*, 47-51.

111. Ray, D.S., Dueber, J. and Suggs, S. (1975) J. Virol. *16*, 348-355.

112. Haldenwang, W.G. and Walker, J.R. (1977) J. Virol. *22*, 23-30.

113. Fidanián, H.M. and Ray, D.S. (1974) J. Mol. Biol. *83*, 63-82.

114. Forsheit, A.B. and Ray, D.S. (1971) Virology *43*, 647-664.

115. Staudenbauer, W.L. and Hofschneider, P.H. (1971) Biochem. Biophys. Res. Commun. *42*, 1035-1041.

116. Kluge, F., Staudenbauer, W.L. and Hofschneider, P.H. (1971) Eur. J. Biochem. *22*, 350-354.

117. Gilbert, W. and Dressler, D. (1969) Cold Spring Harbor Symp. Quant. Biol. *33*, 473-477.

118. Meyer, T.F. and Geider, K. (1979) J. Biol. Chem. *254*, 12636-12641.

119. Meyer, T.F. and Geider, K. (1979) J. Biol. Chem. *254*, 12642-12646.

120. Meyer, T.F., Geider, K., Kurz, C. and Schaller, H. (1979) Nature *278*, 365-367.

121. Eisenberg, S., Griffith, J. and Kornberg, A. (1977) Proc. Natl. Acad. Sci. U.S.A. *74*, 3198-3202.

122. Langeveld, S.A., Mansfeld van, A.D.M., Baas, P.D., Jansz, H.S., Arkel van, G.A. and Weisbeek, P.J. (1978) Nature *271*, 417-420.

123. Salstrom, J.S. and Pratt, D. (1971) J. Mol. Biol. *61*, 489-501.

124. Mazur, B.J. and Model, P. (1973) J. Mol. Biol. *78*, 285-300.

125. Mazur, B.J. and Zinder, N.D. (1975) Virology *68*, 490-502.

126. Bouvier, F. and Zinder, N.D. (1974) Virology *60*, 139-150.

127. Staudenbauer, W.L. and Hofschneider, P.H. (1972) Proc. Natl. Acad. Sci. *69*, 1634-1637.

128. Chen, T.-C. and Ray, D.S. (1976) J. Mol. Biol. *106*, 589-604.

129. Oey, J.L. and Knippers, R. (1972) J. Mol. Biol. *68*, 125-138.

130. Alberts, B.M., Frey, L. and Delius, H. (1972) J. Mol. Biol. *68*, 139-152.

131. Ray, D.S., Bscheider, H.P. and Hofschneider, P.H. (1966) J. Mol. Biol. *21*, 473-483.

132. Pratt, D., Laws, P. and Griffith, J. (1974) J. Mol. Biol. *82*, 425-439.

133. Jacob, E. and Hofschneider, P.H. (1969) J. Mol. Biol. *46*, 359-363.

134. Smits, M.A. Simons, G., Konings, R.N.H. and Schoenmakers, J.G.G. (1978) Biochim. Biophys. Acta *521*, 27-44.

135. Sugiura, M., Okamoto, T. and Takanami, M. (1969) J. Mol. Biol. *43*, 299-315.

136. Chan, T.-S., Model, P. and Zinder, N.D. (1975) J. Mol. Biol. *99*, 369-382.

137. Edens, L., Konings, R.N.H. and Schoenmakers, J.G.G. (1975) Nucleic Acids Res. *2*, 1811-1820.

138. Edens, L., Wezenbeek van, P., Konings, R.N.H. and Schoenmakers, J.G.G. (1976) Eur. J. Biochem. *70*, 577-587.

139. Okamoto, T., Sugimoto, K., Sugisaki, H. and Takanami, M. (1975) J. Mol. Biol. *95*, 33-44.

140. Pieczenik, G., Horiuchi, K., Model, P., McGill, C., Mazur, B.J., Vovis, G.F. and Zinder N.D. (1975) Nature *253*, 131-132.

141. Okamoto, T., Sugiura, M. and Takanami, M. (1969) J. Mol. Biol. *45*, 101-111.

142. Schaller, H., Gray, C. and Herrmann, K. (1975) Proc. Natl. Acad. Sci. U.S.A. *72*, 737-741.

143. Sugimoto, K., Okamoto, T., Sugisaki, H. and Takanami, M. (1975) Nature *253*, 410-414.

144. Sugimoto, K., Sugisaki, H., Okamoto, T. and Takanami, M. (1975) Nucleic Acids Res. *2*, 2091-2100.

145. Takanami, M., Sugimoto, K., Sugisaki, H. and Okamoto, T. (1976) Nature *260*, 297-302.

146. Edens, L., Konings, R.N.H. and Schoenmakers, J.G.G. (1978) Virology *86*, 354-367.

147. Edens, L., Konings, R.N.H. and Schoenmakers, J.G.G. (1978) J. Virol. *28*, 835-842.

148. Schaller, H., Beck, E. and Takanami, M. (1978) Sequence and Regulatory Signals of the Filamentous Phage Genome. In: Single-stranded DNA phages, Denhardt, D.T., Ray, D.S. and Dressler, D., Eds. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp 139-163.

149. Seeburg, P.H., Nüsslein, C. and Schaller, H. (1977) Eur. J. Biochem. *74*, 107-113.

150. Edens, L. (1978) Ph.D. thesis, University of Nijmegen, The Netherlands.

151. Rivera, M.J., Smits, M.A., Quint, W., Schoenmakers, J.G.G. and Konings, R.N.H. (1978) Nucleic Acids Res. 5, 2895-2912.

152. Smits, M.A., Schoenmakers, J.G.G. and Konings, R.N.H. (1980) Eur. J. Biochem., in press.

153. La Farina, M., and Model, P. (1978) Virology *86*, 368-375.

154. Cashman, J.S. and Webster, R.E. (1979) Proc. Natl. Acad. Sci. U.S.A. *76*, 1169-1173.

155. Cashman, J.S., Webster, R.E. and Steege, D.A. (1980) J. Biol. Chem. *255*, 2554-2562.

156. Wu, R., Bambara, R. and Jay, E. (1974) CRC Crit. Rev. Biochem. *2*, 455-512.

157. Pribnow, D. (1975) Proc. Natl. Acad. Sci. U.S.A. *72*, 784-788.

158. Maniatis, T. and Ptashne, M. (1973) Proc. Natl. Acad. Sci. U.S.A. *70*, 1531-1535.

159. Gilbert, W. and Maxam, A.M. (1973) Proc. Natl. Acad. Sci. U.S.A. *70*, 3581-3584.

160. Robertson, H.D., Barrell, B.G., Weith, H.L. and Donelson, J.E. (1973) Nature New Biol. *241*, 38-40.

161. Robertson, H.D. (1975) J. Mol. Biol. *92*, 363-375.

162. Ziff, E.B., Sedat, J.W. and Galibert, F. (1973) Nature New Biology *241*, 34-38.

163. Sedat, J.W., Ziff, E.B. and Galibert, F. (1976) J. Mol. Biol. *107*, 391-416.

164. Blackburn, E.H. (1975) J. Mol. Biol. *93*, 367-374.

165. Blackburn, E.H. (1976) J. Mol. Biol. *107*, 417-432.

166, Roberts, R.J. (1978) Gene *4*, 183-193.

167. Barrell, B.G. (1971) In: Procedures in Nucleic Acid Research, Cantoni, G.L. and Davies, D.R. Eds., Harper and Row, New York, Vol 2, pp 751-779.

168. Brownlee, G.G. (1973) Determination of sequences in RNA, North Holland - Elsevier, Amsterdam.

169. Maizels, N.M. (1973) Proc. Natl. Acad. Sci. U.S.A. *70*, 3585-3589.

170. Zain, S., Weissman, S.M., Dhar, R. and Pan, J. (1974) Nucleic Acids Res. *1*, 577-594.

171. Volckaert, G., Contreras, R., Soeda, E., Voorde van de, A. and Fiers, W. (1977) J. Mol. Biol. *110*, 467-510.

172. Subramanian, K.N., Bhar, R. and Weissman, S.M. (1977) J. Biol. Chem. *252*, 333-339.

173. Mansfeld van, A.D.M., Vereijken, J.M. and Jansz, H.S. (1976) Nucleic Acids Res. *3*, 2827-2843.

174. Mansfeld van, A.D.M. (1979) Ph.D. Thesis, State University of Utrecht.

175. Axelrod, N. (1976) J. Mol. Biol. *108*, 753-770

176. Ling, V. (1972) J. Mol. Biol. *64*, 87-102.

177. Ghangas, G.S., Jay, E., Bambara, R. and Wu, R. (1973) Biochem. Biophys. Res. Comm. *54*, 998-1007.

178. Wu, R., Tu, C.D., Padmanabhan, R. (1973) Biochem. Biophys. Res. Comm. *55*, 1092-1099.

179. Rensing, R.F.E. and Schoenmakers, J.G.G. (1973) Eur. J. Biochem. *33*, 8-18.

180. Jay, E., Bambara, R., Padmanabhan, R. and Wu, R. (1974) Nucleic Acids Res. *1*, 331-353.

181. Bambara, R., Jay, E. and Wu, R. (1974) Nucleic Acids Res. *1*, 1503-1520.

182. Sanger, F. and Coulson, A.R. (1975) J. Mol. Biol. *94*, 441-448.

183. Smith, A.J.H. (1979) Nucleic Acids Res. *6*, 831-848.

184. Maat, J. and Smith, A.J.H. (1978) Nucleic Acids Res. *5*, 4537-4545.

185. Ray, D.S. (1978) *In vivo* Replication of Filamentous Phage DNA. In: The single-stranded DNA Phages, Denhardt, D.T., Ray, D.S. and Dressler, D., Eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. pp 325-339.

186. Maxam, A.M. and Gilbert, W. (1980) Sequencing End-labelled DNA with Base-Specific Chemical Cleavages. In: Methods in Enzymology, Grossman, L. and Moldave, K., Eds., Vol. 65, pp 499-560.

CHAPTER II

RESTRICTION FRAGMENT MAPS OF BACTERIOPHAGE M13

CHAPTER II

RESTRICTION FRAGMENT MAPS OF BACTERIOPHAGE M13

ABSTRACT

Two new physical maps of the bacteriophage M13 genome have been constructed. The fifteen fragments, generated by cleavage of the M13 replicative form DNA (RF) with the restriction enzyme *Hha* I, were ordered by the primed DNA-synthesis method, in which each fragment was used as a primer. The order of the eight *Mbo* II DNA fragments was determined by partial digestion with *Mbo* II of large M13 fragments, which were labelled at only one end. With these and previously constructed physical maps, the M13 genome can be dissected now into more than 100 fragments for further DNA sequence analysis.

INTRODUCTION

The genome of the small bacteriophage M13 consists of a circular, single-stranded DNA molecule of about 6400 bases which, upon infection, is converted into a double-stranded replicative form (RF) DNA molecule.

The use of restriction endonucleases to generate small, specific fragments from replicative form DNA, and the construction of restriction fragment maps potentially allows the precise locations of genetic functions within the genome. Restriction fragment maps have been used to order the filamentous phage genes[1-3] and to trace the origin of replication of the viral and complementary strand[4-7]. Restriction fragment maps have also been used as templates for *in vitro* transcription and coupled transcription-translation studies[8-13]. From the results of these analyses the promoter sites operating on the M13 genome could be mapped and the location of the termination site for transcription could be identi-

fied.

In the course of our sequencing studies it became evident that it was necessary to extend our collection of restriction fragment maps of bacteriophage M13 DNA. The previously published maps were constructed by analysis of partial restriction enzyme digestion products[14] and the analysis of overlapping sets of fragments produced by reciprocal digestion with two restriction enzymes[5,14,15]. In a third method use was made of fragment-primed DNA synthesis[14].

In this paper we report the construction of two additional restriction fragment maps, namely the *Hha* I- and the *Mbo* II-map. The former restriction fragment map was derived by using the fragment-primed DNA synthesis method. In this method the minus strand of M13 restriction fragments serves as primer for DNA synthesis on a viral template strand. If DNA synthesis is carried out after a short pulse with ^{32}P-labelled deoxyribonucleoside triphosphates and synthesis is allowed to continue at least past the next restriction enzyme site, subsequent cleavage of the extended DNA with the same restriction enzyme then results in a specific labelling of that fragment which is adjacent to the 3'-end of the primer. The *Mbo* II map was constructed by using an approach originally developed by Smith and Birnstiel[16,17]. The method is based on partial digestion of a large fragment, which is labelled at only one end, with the restriction enzyme of which the map has to be established, *i.e.* *Mbo* II. A spectrum of partial digestion products is generated, in which the labelled fragments form a simple overlapping series, all with a common labelled terminus. The partial digestion products are electrophoretically fractionated according to size and detected by autoradiography. The size of a partial digestion product, and consequently the distance from an enzyme cleavage site to the labelled terminus, can be calculated by comparison of its mobility with that of molecular weight

markers. The order of the labelled fragments and their lengths correspond to the order of the cleavage sites and their positions on the DNA molecule.

The 15 fragments produced by the restriction enzyme *Hha* I and the 8 *Mbo* II fragments can be ordered unambiguously. With the aid of these new maps and those currently available[5,14,15,18] (P. van Wezenbeek, unpublished results) the M13 DNA can be cleaved now at more than 100 positions. This will enable us to generate fragments, the sizes of which are potentially useful for DNA sequence analysis.

MATERIALS AND METHODS

Materials

M13 viral DNA and M13 replicative form DNA were isolated according to previously published procedures[1]. The restriction enzymes *Hha* I, *Mbo* II and *Bam* H1 were obtained from New England Biolabs. *Hind* II was from Boehringer, Germany. The isolation of *Hap* II and *Hae* III have been described[5]. Polynucleotide kinase from T4-infected cells was obtained from P.L.Biochemicals and *E.coli* DNA polymerase (Klenow fragment) from Boehringer, Germany.

Gel electrophoresis of restriction fragments

Fractionation of restriction fragments was carried out on a discontinuous slab gel (40 cm x 20 cm x 0,2 cm) consisting of a 3% polyacrylamide gel (26 cm) on top of a 10% polyacrylamide layer (10 cm). Very large restriction fragments were fractionated on a 1.5% agarose slab gel (26 cm) on top of a 4% polyacrylamide layer (10 cm). The gels were prepared in electrophoresis buffer containing 40 mM Tris, 20 mM sodium acetate, 2 mM EDTA, pH 7.8. Electrophoresis was performed at room temperature for 20 h with a constant current of 40 mA after which time the bromo-

phenol blue marker had moved to the end of the 10% polyacrylamide gel.

Isolation of Hha I-fragments

M13 RF (approx. 100 μg) supplemented with ^{32}P-labelled RF was incubated with 100 units of *Hha* I for 2 h at 37°C in a reaction volume of 0.5 ml, containing 10 mM Tris-HCl, pH 7.6, 6 mM 2-mercaptoethanol and 50 mM NaCl. The digestion mixture was extracted twice with phenol and the DNA-fragments were precipitated at -20°C for 24 h after the addition of 0.1 vol of 3 M sodium acetate, pH 5.6 and 2.5 volumes of ethanol. The precipitate was collected by centrifugation, washed once with 96% ethanol and finally dissolved in 0.1 ml electrophoresis buffer. After electrophoresis of the DNA-fragments on a 3% discontinuous slab gel and subsequent autoratiography the *Hha* I fragments were cut out and extracted as described previously[1]. The purified fragments were dissolved in 10 mM Tris-HCl, pH 7.6 and 1 mM EDTA and used in the fragment-primed DNA synthesis.

Primed DNA synthesis with Hha I-restriction fragments

Prior to DNA synthesis, primer/template heteroduplexes were prepared by hybridization of the appropriate fragments to M13 viral DNA strands. The hybridization was carried out in a reaction mixture (0.05 ml) containing 20 mM Tris-HCl, pH 7.6, 0.2 mM EDTA, 100 mM NaCl, 1 pmole of purified DNA-fragment and 3 pmoles of circular, single-stranded M13 viral DNA. The mixture was heated to 100°C for 5 min in sealed polypropylene tubes and incubated for 2 h at 56°C to allow hybridization to occur.

Fragment-primed DNA synthesis was carried out in 50 mM Tris-HCl, pH 7.6, 100 mM NaCl, 10 mM $MgCl_2$ and 0.5 mM EDTA. The standard reaction mixture contained in addition: 80 pmoles of (α-^{32}P)-dATP (spec act

300-400 Ci/mmol, from Amersham, England), 3 mmoles each of unlabelled dCTP, dGTP and TTP, 0.2 pmole of primer/template DNA and 0.5 unit of DNA polymerase I (Klenow fragment). Label incorporation was allowed to proceed at 20°C for 20 min. Thereafter, 3 mmoles of unlabelled dATP (5 µl) was added to the mixture and the incubation continued for 20 min at 33°C. The reaction was terminated by phenolextraction and the DNA was precipitated from the aqueous phase at -20°C by the addition of 0.1 volume 3 M sodium acetate, pH 5.6, and 2.5 volumes of ethanol. The extended DNA was dissolved in 50 µl enzyme buffer and digested to completion with *Hha* I. The resulting digestion products were fractionated on a 3% discontinuous polyacrylamide slab gel and the location of the radioactive bands was determined by autoradiography. The radioactive bands were identified with the aid of a *Hha* I digest of nick-translated M13 RF. Nicktranslation was carried out as described by Jeppesen *et al*[19].

Isolation of DNA fragments labelled at a single 5'-OH end

M13 RF (8 µg) was digested with restriction enzyme *Bam* H1 in excess at 37°C in a reaction volume of 0.1 ml containing 6 mM Tris-HCl, pH 7.6, 6 mM $MgCl_2$, 6 mM 2-mercaptoethanol and 50 mM NaCl. The whole digestion mixture was transferred to a polypropylene tube containing 100 pmole of dried (γ-^{32}P)-ATP (spec act 1500-2000 Ci/mmol). Thereafter, 1-2 units of polynucleotide kinase were added and both 5'-ends of the linear RF molecule were labelled by the kinase-mediated exchange reaction[20]. After 30 min at 37°C the reaction was terminated with phenol. After two extractions the DNA was precipitated twice with ethanol to remove most of the unincorporated label. The precipitate was dissolved in 0.1 ml buffer containing 6 mM Tris-HCl, pH 7.6, 6 mM $MgCl_2$, 6 mM 2-mercaptoethanol and 50 mM NaCl and digested to completion with *Hind* II. The resulting two fragments, labelled only at their *Bam* H1 ends, were separat-

ed by electrophoresis on a 3% polyacrylamide slab gel. The fragments were located by autoradiography, extracted from the gel and precipitated with ethanol.

In case of labelling the *Hind* II ends the order of digestion with the restriction enzymes was reversed.

Partial digestion of DNA fragments labelled at a single 5'-OH end

Fragments, labelled at their *Bam* H1- or *Hind* II-ends were dissolved in 50 µl of 10 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 1 mM DTT and 6 mM KCl. Unlabelled RF (1 µg) was added as a carrier and the digestion was started by the addition of 2 units of *Mbo* II enzyme, At 0, 1, 5 and 20 min a 10 µl sample was taken and added to 3 µl of dye mixture containing 50% glycerol, 0.1% bromophenol blue and 0.5% sodium dodecyl sulfate. The samples were heated for 5 min at 60°C and directly loaded on a vertical 1.5% agarose slab gel on top of a 4% polyacrylamide layer. ^{32}P-labelled *Hap* II- and *Hae* III-fragments were applied to adjacent slots to serve as molecular weight markers. After electrophoresis, the polyacrylamide gel layer was removed and the agarose gel dried. The radioactive bands in the gel were localized by autoradiography.

RESULTS

Hha I restriction cleavage map

M13 RF was digested to completion with the restriction enzymes *Hap* II, *Hha* I, *Taq* I and *Mbo* II. The resulting fragments were fractionated in a 4% polyacrylamide slab gel and made visible by ethidiumbromide staining (Fig 1). In lane 2 11 *Hha* I fragments (A-K) can readily be detected. However, 3 additional, small fragments (L, M and N), which could not be made visible by the staining procedure, proved to be present when uniformly ^{32}P-labelled M13 RF was used as substrate for the *Hha* I restict-

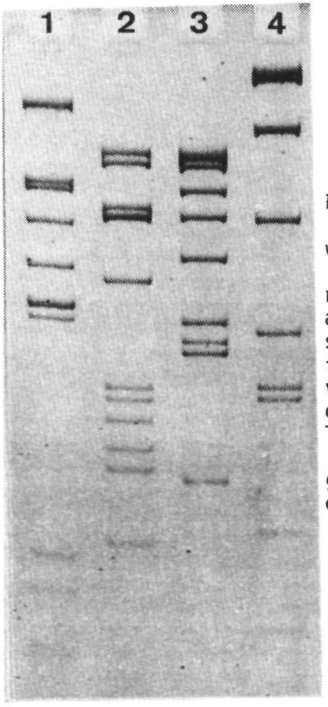


Fig 1 Gel electrophoresis of fragments produced by digestion of M13 RF with various restriction enzymes

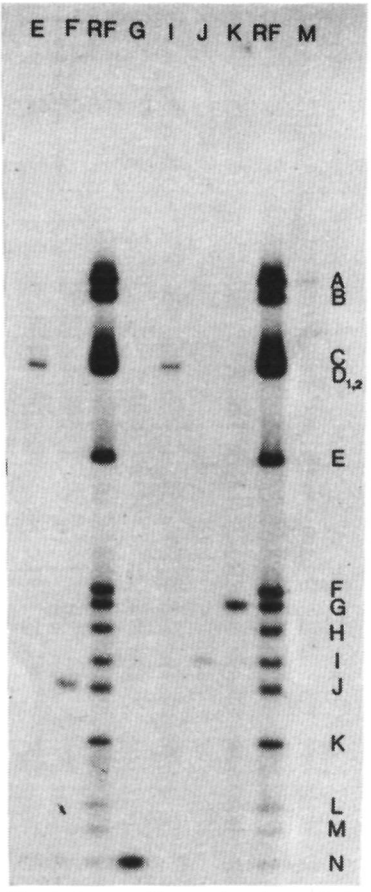
M13 was digested to completion with the restriction enzymes *Hap* II, *Hha* I, *Taq* I and *Mbo* II. The resulting fragments were separated on a 4% polyacrylamide slab gel for 16 h at 40 mA. The fragments were visualized by staining with a solution containing 1 μg/ml ethidium bromide, 10 mM Tris-HCl, pH 7.6 and 1 mM EDTA.

Lane 1, *Hap* II digest; *lane* 2, *Hha* I digest; *lane* 3, *Taq* I digest; *lane* 4, *Mbo* II digest.

ion enzyme (cf. Fig 2, *lane* 3 and 8).

The size of the DNA fragments was calculated on the basis of the relative ^{32}P-yield of each fragment from uniformly ^{32}P-labelled M13 RF (Table 1). Taken into account a total of 6400 basepairs per M13 RF molecule, a range of 990 basepairs for the largest fragment *Hha* I-A upto about 45 basepairs for the smallest fragment *Hha* I-N is obtained.

To determine the neighbour fragment of each *Hha* I fragment, a template-primer DNA hybrid was constructed by denaturating and reannealing each purified fragment in the presence of circular M13 DNA viral strands. By offering ^{32}P-labelled deoxyribonucleoside triphosphates for short pulse periods, followed by a chase of unlabelled deoxyribonucleoside triphosphates, DNA polymerase was made to synthesize specifically-started DNA chains which were radioactively labelled only in the regions vi-



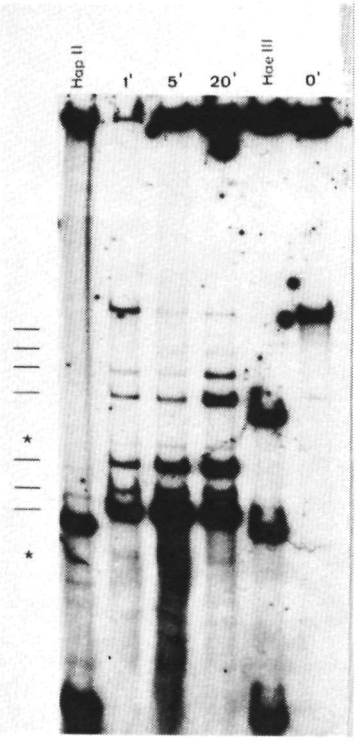


Fig 2 Autoradiograph of a 3% discontinuous slab gel showing the fragments labelled by primed DNA synthesis using M13 viral DNA as template and the *Hha* I fragments E (*lane* 1), F (*lane* 2), G (*lane* 4), I (*lane* 5), J (*lane* 6), K (*lane* 7) and M (*lane* 9) as primer.
Marker fragments (*lane* 3 and 8) were obtained by digestion of ^{32}P-labelled M13 RF with *Hha* I).

Fig 3 Autoradiograph of a 1.5% agarose slab gel, showing the products obtained by partial digestion of the large Bam H1-HII fragment, labelled at the Bam H1 end, with Mbo II.

The large *Bam* H1-*Hind* II fragment, labelled at its *Bam* H1 end, was partially digested with *Mbo* II. Samples were taken immediately before addition of the enzyme (0') and at 1, 5 and 20 min and electrophorized in a 1.5% agarose slab gel. *Lane* 1, ^{32}P-labelled *Hap* II digest of M13 RF; *lane* 2, sample taken at one minute; *lane* 3, at five minutes and *lane* 4, at twenty minutes after the addition of the enzyme; *lane* 5, ^{32}P-labelled *Hae* III digest of M13 RF; *lane* 6, sample taken immediately before addition of the enzyme. Only the upper part of the gel is shown.
The two faint bands, denoted by asteriks, most probably represent the partial *Mbo* II products of the small *Bam* H1-*Hind* II fragment, which slightly contaminates the large *Bam* H1-*Hind* II fragment preparation.

Table 1 SIZES OF FRAGMENTS PRODUCED BY DIGESTION WITH THE RESTRICTION ENZYMES *Hha I* AND *Mbo II*

Hha I fragment	Size(basepairs)[a]	*Mbo* II fragment	Size(basepairs)[c]
A	990	A	3100
B	930	B	1250
C	740	C	650
$D^{b}_{1,2}$	720	D	400
E	515	E	350
F	325	F	300
G	310	G	200
H	280	H	150
I	255		
J	245		
K	205		
L	95		
M	70		
N	45		

a. Sizes were calculated on the basis of the relative ^{32}P-yield of each fragment from uniformly labelled ^{32}P-RF and assuming a total of 6400 basepairs per M13 RF molecule
b. By plotting the logarithm of the ^{32}P-content of each fragment versus its electrophoretic mobility fragment *Hha* I-D proved to be a doublet.
c. Sizes were calculated from the distance between two adjacent *Mbo* II cleavage sites, of the positions which were determined by partial digestion of large M13 fragments labelled at one 5'-terminus.

cinal to the primer. After subsequent cleavage of the extended DNA by *Hha* I, the digestion products were subjected to electrophoresis on 3% discontinuous polyacrylamide slab gels and analysed by autoradiography.

Several examples of autoradiographic analysis of the newly synthesized neighbour fragments are presented in Fig 2. *Lane* 2 shows the product obtained when fragment *Hha* I-F is extended. It is clear that frag-

ment *Hha* I-J is the fragment which is preferentially labelled by the *Hha* I-F primed DNA synthesis and, hence, argues for a contiguous position of both fragments on the M13 genome in the directions 5'...*Hha* I-F-*Hha* I-J...3'.

In an analogous way we conclude from this autoradiograph that *Hha* I-E is followed by *Hha* I-D, *Hha* I-G by *Hha* I-N, *Hha* I-I by *Hha* I-D, *Hha* I-J by *Hha* I-I, *Hha* I-K by *Hha* I-G and *Hha* I-M by *Hha* I-A. In each case the primer fragment itself was not labelled. This indicates that the fragments were intact and apparantly did not contain single-stranded nicks which might cause incorporation of ^{32}P-label in the primer fragment region.

The results of the priming studies with all *Hha* I fragments are summarized in Table 2. An apparent ambiguity in this Table arises when

Table 2 NEIGHBOUR FRAGMENT(S) IDENTIFIED BY Hha I FRAGMENT PRIMED DNA SYNTHESIS USING M13 VIRAL DNA AS TEMPLATE

Hha I fragment	Neighbour fragment(s)[a]
A	B
B	E
C	K
$D_{1,2}$	C,H
E	D,(N)
F	J,(N)
G	N
H	L,(A)
I	D
J	I,(D)
K	G,(D)
L	M,(A)
M	A
N	F

a. Fragments in parenthesis were weakly labelled.

fragment *Hha* I-D is considered. When this fragment is used as a primer both fragments *Hha* I-C and *Hha* I-H are labelled. On the other hand, fragment *Hha* I-D is labelled by primed DNA synthesis with *Hha* I-E but also with *Hha* I-I. These results can easily be explained by the fact that fragment *Hha* I-D is a double fragment.

The order of fragments deduced from the results presented in Table 2 is: A→B→E→D_1→C→K→G→N→F→J→I→D_1→H→L→M.

The *Hha* I restriction fragment map was aligned with the previously constructed *Hap* II, *Hae* II, *Hae* III, *Taq* I and *Alu* I maps by double digestion of M13 RF DNA with *Hha* I and either restriction enzyme *Hind* II or *Bam* H1. The latter two enzymes cleave M13 RF only once. Double digestion with *Hha* I and *Hind* II results in the disappearance of fragment *Hha* I-B (930 bp) and the formation of two new fragments, 880 and 50 bp long. *Bam* H1 cleaved fragment *Hha* I-I (255 bp). Two new fragments were generated with a length of 230 and 25 bp (data not shown). The results of both double-digestion experiments are only compatible with an alignment of the *Hha* I restriction fragment map as shown in Fig 5.

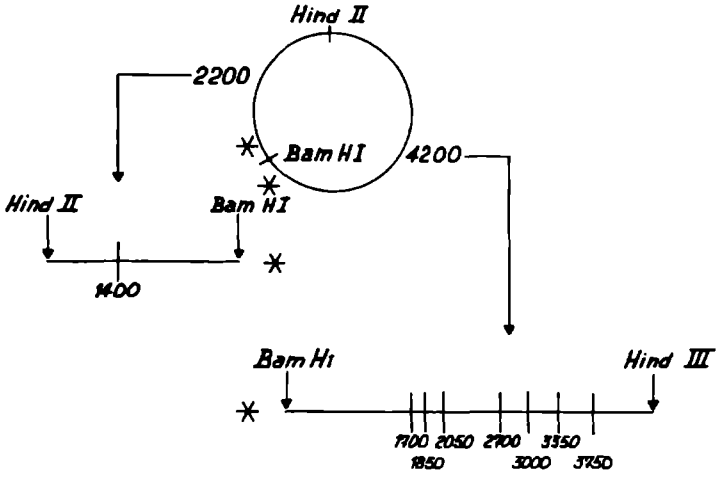
Mbo II restriction cleavage maps

Complete digestion of M13 RF with the restriction enzyme *Mbo* II results in the formation of 8 discrete fragments (Fig 1, *lane* 4). The approximate lengths of the *Mbo* II fragments can be deduced from their mobility relative to that of the *Hap* II- and *Taq* I- marker fragments (Fig 1, *lanes* 1 and 3). They range in size from about 3000 basepairs for the largest fragment *Mbo* II-A to 150 basepairs for the smallest fragment *Mbo* II-H. The *Mbo* II fragments were ordered by partial digestion of large M13 RF fragments which were labelled at a single 5'-terminus. To generate these fragments, M13 RF was cleaved with *Bam* H1. The linearized RF molecule was labelled at its 5'- ends with (γ-^{32}P)-ATP and polynucleotide kinase and subsequently digested with *Hind* II. Two fragments

were formed, 4200 and 2200 basepairs long which were labelled only at their *Bam* H1 ends. Both *Bam* H1-*Hind* II fragments were separated by electrophoresis and, after purification, subjected to partial digestion with *Mbo* II. Samples were taken immediately before addition of the enzyme (0') and after 1, 5 and 20 min and then applied to a 1.5% agarose slab gel. ^{32}P-labelled *Hap* II- and *Hae* III-fragments of known molecular weight were run in parallel as marker fragments. Fig 3 shows the fragments that were formed in case the large *Bam* H1-*Hind* II fragment was partially digested with *Mbo* II. The sizes of the partial digestion products formed were 3750, 3350, 3000, 2700, 2050, 1850 and 1700 basepairs. In a similar way, the small *Bam* H1-*Hind* II fragment was partially digested with *Mbo* II. In this case only one ^{32}P-labelled degradation product was formed with a length of 1400 basepairs (data not shown, but see legend to Fig 3).

The lengths of the partial digestion products determine the distances between the *Mbo* II cleavage sites and the *Bam* H1 site. These distances are indicated in Fig 4. The lengths of the individual *Mbo* II-fragments can be calculated now from this Figure. These fragments were arranged according to size in Table 1. From this Table and the data presented in Fig 4 it is evident that the order of *Mbo* II fragments is A→H→G→C→F→E→D→B. Fragment *Mbo* II-A is cleaved by *Bam* H1 into two fragments, 1400 and 1700 basepairs long. Restriction enzyme *Hind* II cleaves fragment *Mbo* II-B generating two fragments with a length of 450 and 800 basepairs, respectively (data not shown). The *Mbo* II cleavage map, aligned with the previously constructed maps through the single *Bam* H1 and *Hind* II cleavage sites is presented in Fig 5.

In an analogous way partial digestions have also been carried out with the small and large fragments labelled at their respective *Hind* II ends (data not shown). The positions of the *Mbo* II cleavage sites



*denotes ^{32}P-label

Fig 4 The sizes of products (basepairs), obtained by partial digestion of the small and large *Bam* H1-*Hind* II fragment with *Mbo* II.

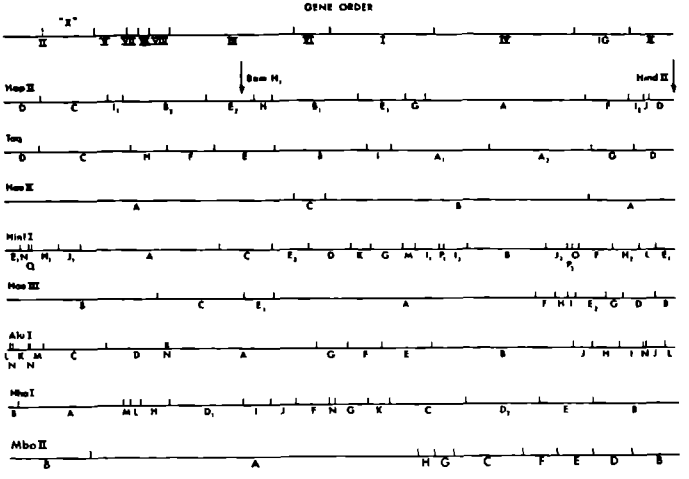


Fig 5 Genetic map and restriction fragment maps of M13 RF cleaved with *Hap* II, *Taq* I, *Hae* II, *Hae* III, *Hinf* I, *Alu* I, *Bam* H1, *Hind* II, *Hha* I and *Mbo* II.

read from the *Hind* II site were in complete agreement with the positions read from the *Bam* H1 site.

DISCUSSION

In this chapter, two techniques were applied to construct additional restriction fragment maps of bacteriophage M13 RF.

The eight *Mbo* II fragments were ordered by partial digestion of the two large fragments, generated by cleavage of M13 RF with *Bam* H1 and *Hind* II. This method has the advantage of being rapid and simple to perform. A disadvantage, inherent to this method, is the rather inaccurate size-determination of large partial fragments in agarose gels, resulting in less accurate cleavage site locations. Moreover, cleavage sites very near to each other would probably escape detection because of the very small size increment of a large partial fragment, caused by the addition of the small fragment separating both cleavage sites. Because digestion of M13 RF with *Mbo* II yields a limited number of rather large fragments (Fig 1, *Lane* 4), the partial digestion method could successfully be applied to construct the *Mbo* II cleavage map.

Very small fragments are present in a complete digest of M13 RF with *Hha* I. For this reason we did not try to construct a *Hha* I map with the aid of the partial digestion method. The fragment primed DNA synthesis method was used in this case since in this method the DNA synthesis conditions can easily be modified, so as to facilitate the detection of small fragments. In addition, the small fragment itself can be used as primer fragment in the extension reaction, thereby directly determining its position with respect to the neighbour fragment. The method has the disadvantage of being more laborious because of the need to isolate the individual restriction fragments in a pure form. As shown in

this chapter all 15 *Hha* I-fragments could unambiguously be ordered in this way.

In Fig 5 a summary is given for all M13 restriction fragment maps presently known. Also, the approximate positions of the M13 genes are indicated. It is clear that with the aid of these maps the whole M13 genome can be dissected now in fragments small enough for DNA sequence analysis. This will enable us to unravel the structure of the individual M13 genes and the regulatory elements involved in the process of transcription and replication of the filamentous M13 genome.

1. Hondel van den, C.A., Weijers, A., Konings, R.N.H. and Schoenmakers, J.G.G. (1975) Eur. J. Biochem. *53*, 559-567.
2. Seeburg, P.H. and Schaller, H. (1975) J. Mol. Biol. *92*, 261-277.
3. Horiuchi, K., Vovis, G.F., Enea, V. and Zinder, N.D. (1975) J. Mol. Biol. *95*, 147-165.
4. Tabak, H.F., Griffith, J., Geider, K., Schaller, H. and Kornberg, A. (1974) J. Biol. Chem. *249*, 3049-3054.
5. Hondel van den, C.A., Pennings, L. and Schoenmakers, J.G.G. (1976) Eur. J. Biochem. *68*, 55-70.
6. Horiuchi, K. and Zinder, N.D. (1976) Proc. Natl. Acad. Sci. U.S.A. *73*, 2341-2345.
7. Suggs, S. and Ray, D.S. (1977) J. Mol. Biol. *110*, 147-163.
8. Vovis, G.F., Horiuchi, K. and Zinder, N.D. (1975) J. Virol. *16*, 674-684.
9. Hondel van den, C.A., Konings, R.N.H. and Schoenmakers, J.G.G. 1975) Virology *67*, 487-497.
10. Edens, L., Konings, R.N.H. and Schoenmakers, J.G.G. (1975) Nucleic Acids Res. *2*, 1811-1820.
11. Edens, L., Wezenbeek van, P., Konings, R.N.H. and Schoenmakers, J.G.G. (1976) Eur. J. Biochem. *70*, 577-587.
12. Edens, L., Konings, R.N.H. and Schoenmakers, J.G.G. (1978) J. Virol. *28*, 835-842.
13. Edens, L., Konings, R.N.H. and Schoenmakers, J.G.G. (1978) Virology *86*, 354-367.
14. Hondel van den, C.A. and Schoenmakers, J.G.G. (1975) Eur. J. Biochem. *53*, 547-558.
15. Hondel van den, C.A. and Schoenmakers, J.G.G. (1976) J. Virol. *18*, 1024-1037.

16. Smith, H.O. and Birnstiel, M.L. (1976) Nucleic Acids Res. 3, 2387-2398.

17. Boseley, P.G., Tuyns, A. and Birnstiel, M.L. (1978) Nucleic Acids Res. 5, 1121-1138.

18. Hondel van den, C.A. and Schoenmakers, J.G.G. (1973) Mol. Biol. Rep. 1, 41-45.

19. Jeppesen, P.G.N., Sanders, L. and Slocombe, P.M. (1976) Nucleic Acid Res. 3, 1323-1339.

20. Berkner, K.L. and Folk, W.R. (1977) J. Biol. Chem. 252, 3176-3184.

CHAPTER III

NUCLEOTIDE SEQUENCE OF GENE VII AND OF A HYPOTHETICAL GENE (IX) IN BACTERIOPHAGE M13

Part of the results described in this Chapter has been published:

Hulsebos, T. and Schoenmakers, J.G.G. (1978) Nucleic Acids Res. 5, 4677-4698.

CHAPTER III

NUCLEOTIDE SEQUENCE OF GENE VII AND OF A HYPOTHETICAL GENE (IX) IN BACTERIOPHAGE M13

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 has been 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 ($Su\ I^+$)-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 µq), 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].

Marker rescue

Marker rescue experiments were performed by the method of Taketo[30] as described by van den Hondel *et al* [14]. Wild-type M13 restriction fragments (0.1 pmol), dissolved in 50 µl 50 mM Tris-HCl, pH 7.6 and 1 mM EDTA, were denaturated by heating for 5 min at 100°C and then rapidly chilled.

50 µl of amber mutant viral DNA's (1 pmol in 10 mM Tris-HCl, pH 7.6, 0.1 M NaCl and 1 mM EDTA) were added and the mixtures incubated for 3 h at 67°C. For transfection 50 µl of the hybrid containing solutions were mixed with 0.1 ml $CaCl_2$-treated *E.coli* K38 (Su^-) cells at 0°C. The mixtures were incubated for 10 min at 0°C, then for 2.5 min at 37°C and chilled. The transfected cells were mixed with 5 ml agar and plated for infective centers together with 0.2 ml *E.coli* K38 (Su^-) as indicator strain.

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 15 µl 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 (3mg/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 fractionat-

ed 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^{o}. Preferential cleavage at Adenine was achieved by treatment with 0.1 N HCl followed by treatment with 0.1 N alkali at 90^{o}. 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/bisacrylamide 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 (van den Hondel *et al* , 1975). In order to determine the nucleotide sequence of gene VII,

a more accurate position of these markers within the M13 restriction fragment map was needed. As several new restriction fragment maps of M13 DNA have recently been constituted (see Chapter II) a more precise alignment of gene VII could be achieved by marker rescue experiments. The results of these studies are summarized in Table 1.

Table 1 RESCUE OF AMBER 7 MUTATIONS BY RESTRICTION FRAGMENTS

viral DNA	M13 fragment	Infective centers per plate
am7-H2	-	2
am7-H2	*Taq* I-C	96
am7-H2	*Taq* I-H	3
am7-H2	*Hha* I-L	44
am7-H3	-	3
am7-H3	*Taq* I-C	6
am7-H3	*Taq* I-H	416
am7-H3	*Hha* I-L	23

$CaCl_2$-treated *E.coli* K38 (Su^-) cells were transfected with M13 amber 7 viral DNA's to which denaturated M13 wild-type restriction fragments were hybridized. The transfected cells were plated for infective centers together with *E.Coli* K38 (Su^-) as indicator strain. Approximately $3x10^{11}$ viral DNA molecules and $3x10^{10}$ molecules restriction fragment were used per plate.

M13 mutant *am7*-H2, which contains an amber mutation in gene VII, is rescued not only by restriction fragment *Taq* I-C but also by the very small restriction fragment *Hha* I-L. A second amber mutant, M13 *am7*-H3 is rescued by both restriction fragments *Taq* I-H *Hha* I-L. From this we infer that the latter fragment forms part 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).

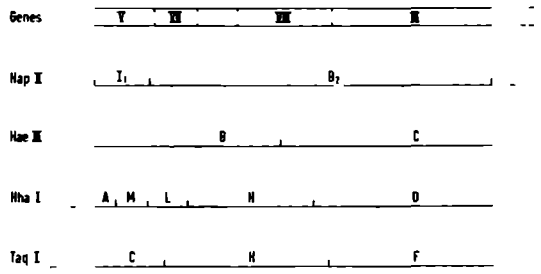


Fig 1 Schematic diagram of a segment of the genetic map and of the restriction enzyme cleavage maps of bacteriophage M13 DNA. The Roman numerals refer to the genes. The capital letters refer to the DNA fragments which are obtained after digestion of this part of the M13 genome with the various restriction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

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 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

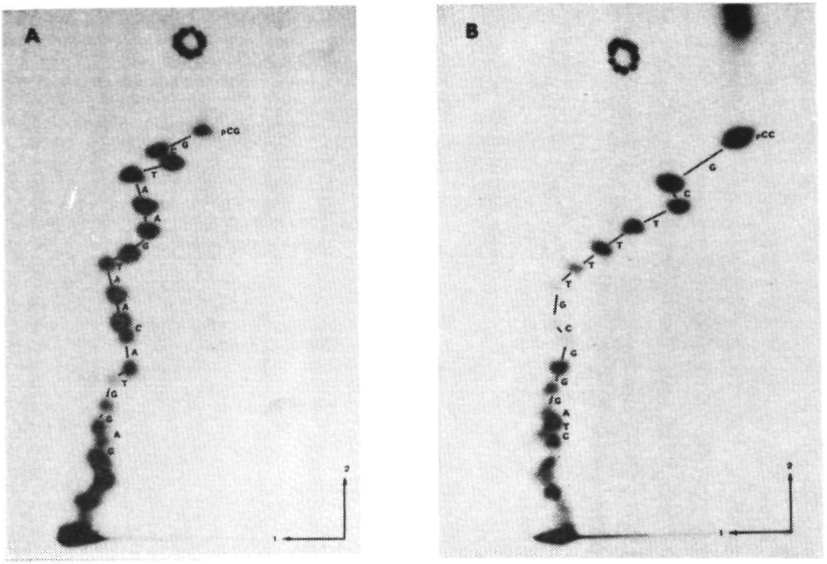


Fig 2 Autoradiographs of two-dimensional fingerprints of oligonucleotides derived after partial digestion with snake-venom phosphodiesterase of the left-hand-(A) and right-hand-(B) boundary of the "300-fragment" (*cf.* Fig 3).

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.

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.

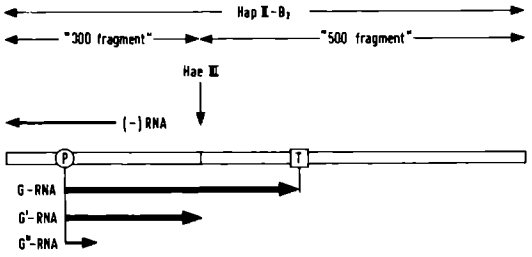


Fig 3 Transcription map of fragment *Hap* II-B_2 and the "300-fragment". P and T refer to the promoter $G_{0.18}$ and the central terminator $T_{0.25}$, respectively.

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 (see Chapter IV). 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

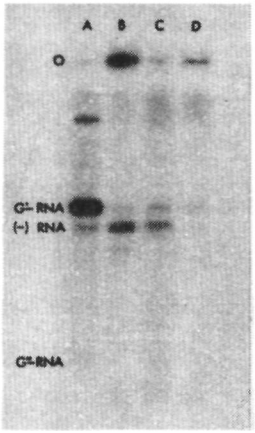


Fig 4 Electrophoretic analysis on a 5% polyacrylamide gel of the RNA products formed upon transcription of the "300-fragment" in the absence (A) and presence of the dinucleotide GpC (B) and CpC (D) or in the presence of high concentrations of CTP (C) (*cf.* Fig 3).

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) 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 2, in which nucleotide numbers correspond to the spot numbers given in Fig 5. These analyses 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 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 geneVIII. The nucleotide sequence of this mRNA encoded by the closely related bacteriophage fd has recently been solved[23]. It appeared to us that the 5' terminal regions of the fd-encoded major coat protein message and that of G'-RNA, transcribed from M13 DNA, are exactly identical. The only difference noted was a pppGGGGG start of the M13-transcript as compared to a pppGGGG start in the fd-mRNA. Therefore, the sequence of this RNA was initially 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 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 restrict-

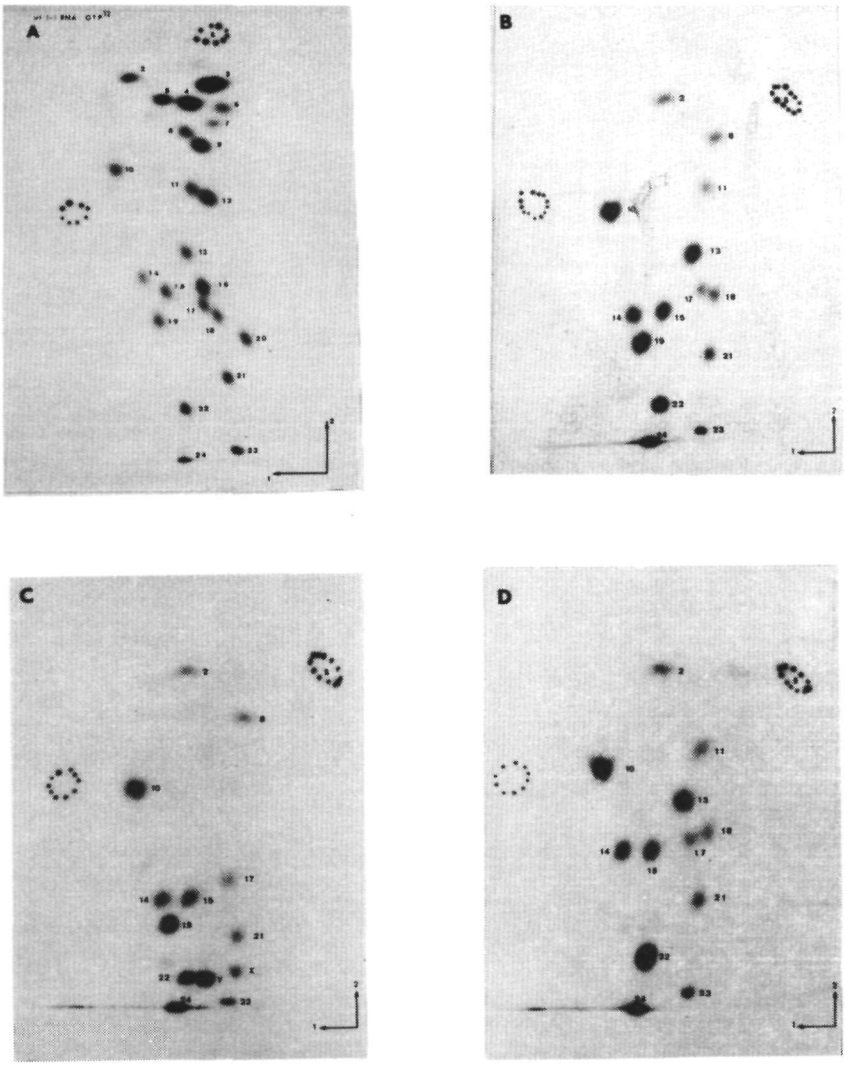


Fig 5 Autoradiographs of the T1-fingerprints of (-)RNA. Fingerprints of wild-type (-)RNA (A,B), *am*7-H2 (-)RNA (C) and *am*7-H3 (-)RNA (D). The RNA species have been labelled with (α-^{32}P)-GTP (A) or with (α-^{32}P)-UTP (B,C and D). Oligonucleotides 1a and 1b (Table 1) are not labelled under these labelling conditions.

Table 2. 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
$T4_{a,b}$	[illegible],[illegible]	[illegible]	-	-	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	[illegible]	C	-	-	CAG(G)	<0.5
T8	U	G	C	C	CCUG(A)	1
T9	A[illegible]	[illegible]AG	-	-	AAG(G)	1
T10	U	-	G,U,[illegible]	-	AUUUG(U)	1
T11	U	-	C	G,[illegible]	ACCUG(C)	1
T12	AA[illegible]G	[illegible]	-	-	AAAG(A)	1
T13	A[illegible]	C	G,C,[illegible]U	U,C	CUCCAUG(U)	1
T14	[illegible]G	U	2U,A[illegible]	A[illegible],[illegible]C	UUACUUAG(C)	1
T15	C	U,C	[illegible]U	G,2A[illegible]	UAUCAUCG(C)	1
T16	G,AA[illegible]	U,A[illegible],[illegible]AC	-	[illegible]C,A[illegible]C	UACAACG(G)	1
T17	C	AAAU	AA[illegible]U	G,C,AAA[illegible]	AAAUCCG(C)	1
T18	A[illegible]	G,U,C	A[illegible]	C,2[illegible]C	CCACUACG(A)	1
T19	U	A[illegible],[illegible]AU	G,[illegible]U,AA[illegible]	-	AUAAAUUG(U)	1
T20	[illegible]G	C	-	3C,A[illegible],[illegible]	ACCCCCAG(C)	1
T21	AA[illegible]G	[illegible]AG,[illegible]A[illegible]	AAA[illegible]	AA[illegible]C	AAACAAAG(U)	1
T22	A[illegible]G	U,C,A[illegible],[illegible]AG	[illegible],[illegible]U	[illegible],AA[illegible]	AUUAUACCAAG(C)	1
T23	AAAA[illegible]	G,U,2C,[illegible]AC [illegible]AC	C	[illegible],A[illegible],AAA[illegible]C	C ACC,AACC UAAAACG(A)	1
T24	U	G,U,C,A[illegible],[illegible]A[illegible], [illegible]A[illegible]	2U,C,[illegible]U,2A[illegible], A[illegible]U	U,A[illegible],3[illegible]C, AAA[illegible]C	-	1
X	AA[illegible]G	U,C,A[illegible],[illegible]AG	A[illegible]	C,2[illegible]C	CCACUACAAAG(G)	1
Y	A[illegible]	U,C	G,C,[illegible]U,A[illegible]	U,C,[illegible],[illegible]C	ACCUACUCCAUG(U)	1
cmT22	A[illegible]G	U,C,A[illegible],[illegible]AG	[illegible],[illegible]U	[illegible],AA[illegible]	AUUAUACCAAG(C)	1
	U	U,[illegible]A[illegible],[illegible]AU	G,C,A[illegible]U,AA[illegible]	C	CCUAAUAAAUUG(U)	1

a. The numbers refer to the fingerprints shown in Figure 5. X, Y and cmT22 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 (cm7-H2, X and Y; cm7-H3, cmT22).

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.

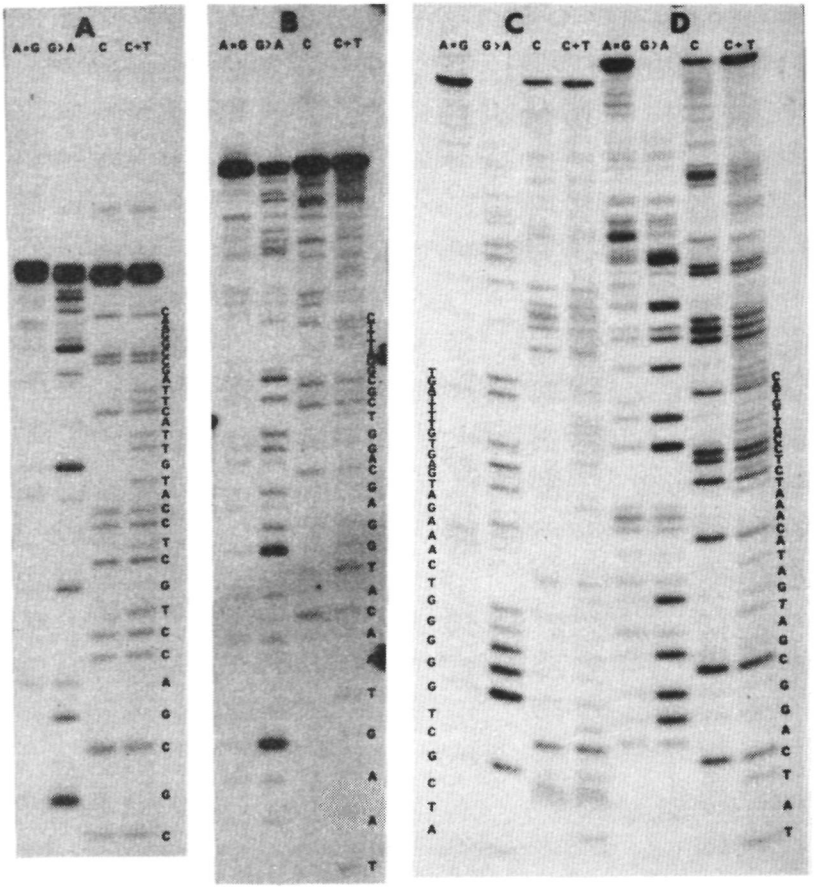


Fig 6 Autoradiographs of DNA sequencing gels obtained after chemical degradation of: (A) the right-hand 5'-end of fragment *Taq* I-C; (B) the left-hand 5'-end of the "300-fragment "; (C) the left-hand 5'-end of the fragment *Hha* I-H and (D) the left-hand 5'-end of fragment *Taq* I-H. After labelling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

ion 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 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

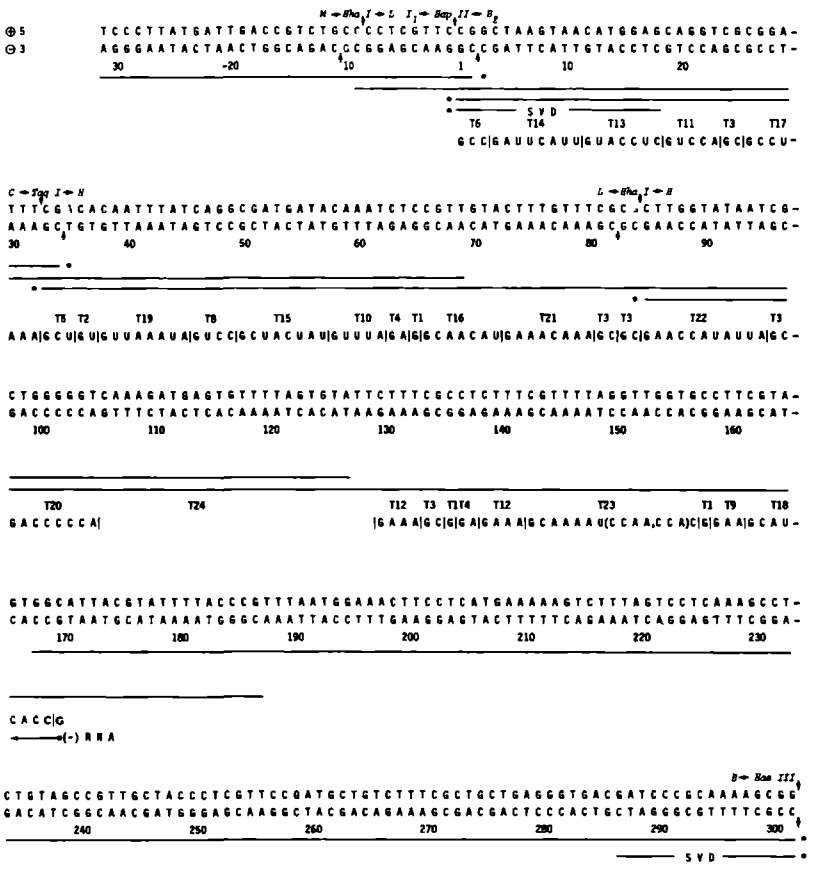


Fig 7 Nucleotide sequence of the right hand boundary of fragment Hap II I_1 and of the "300-fragment" (*cf* Fig 1)
In this Figure the deduced nucleotide sequence of (-)RNA transcribed from the "300-fragment" is also indicated ⊕, indicates the viral strand and ○ , the complementary (codogenic) strand
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 *Bha* 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

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 2.

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 2. 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 RNA's 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

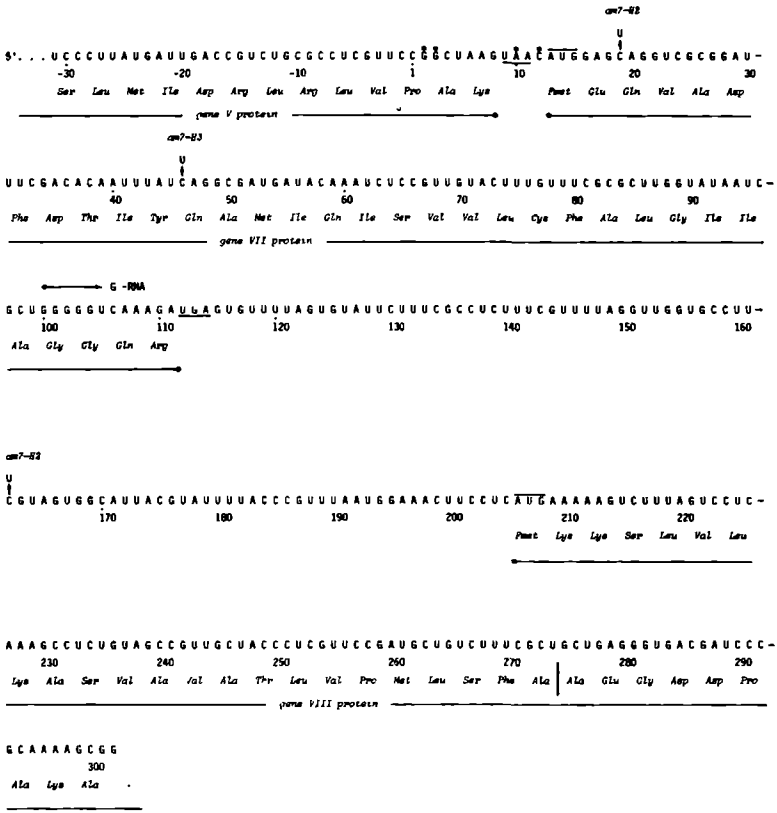


Fig. 8. Nucleotide sequence of the RNA complementary to the established nucleotide sequence of the codogenic strand presented in Fig. 7. In this Figure the amino acid sequences of (a) gene VII protein, (b) the C-terminal part of gene V protein, the hypothetical gene IX protein, and of (d) the N-terminal 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 *cm7*-H2 and *cm7*-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.

*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 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] (see Chapter V) 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 (C. 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).

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 Hondel, van den, 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. *28*, 835-842.
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 *521*, 27-44.
10. Edens, L., Wezenbeek van, 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., Pennings, L. 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.S. and Dressler, D., Eds. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. pp 507-530.

21. Cuypers, T., Ouderaa van der, F.J. and Jong de, W. (1974) Biochem. Biophys. Res. Commun. *59*, 557-563.

22. Asbeck, F., Beyreuther, K., Kohler, H., Wettstein von, 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., Beck, E. and Takanami, M. (1978) In: Single-Stranded DNA Phages, Denhardt, D.T. and Dressler, D., Eds. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp 139-163.

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.

30. Taketo, A. (1972) J. Biochem. (Tokyo) *72*, 973-979.

CHAPTER IV

FURTHER STUDIES ON THE IN VITRO TRANSCRIPTION OF THE GENE VII BOUNDARY OF BACTERIOPHAGE M13

CHAPTER IV

FURTHER STUDIES ON THE IN VITRO TRANSCRIPTION OF THE GENE VII BOUNDARY OF BACTERIOPHAGE M13

ABSTRACT

In vitro transcription of the '300-fragment', containing gene VII of bacteriophage M13, results in the synthesis of G'-RNA and two minor RNA's, *i.e.* (-)RNA and G"-RNA (*cf.* Chapter III).

The nucleotide sequence of G"-RNA has been determined by RNA sequence analysis of fragments and ordering of these fragments on the DNA sequence of the '300-fragment'. From our data it is concluded that G"-RNA is a prematurely terminated product, consisting of the first 45-50 nucleotides of G'-RNA and that it is transcribed from the complementary strand. It is inferred that G"-RNA is generated by blockage of the elongation of G'-RNA by RNA polymerase molecules, attached to the (-)RNA promoter.

INTRODUCTION

Hybridization studies on RNA's, isolated from M13-infected cells, have indicated that only the complementary strand of the M13 double-stranded replicative form (RF) is transcribed[1,2]. The results of *in vitro* transcription studies[3-8] are in accordance with the strand specificity, observed in the infected cell.

However, an exception seems to be the transcription of the '300-fragment' (Chapter III). Transcription of this fragment results in the synthesis of G'-RNA and two minor RNA's, *i.e.* (-)RNA and G"-RNA (Fig 1). Our sequence analysis has demonstrated that (-)RNA is not the product of the complementary strand but is transcribed from the viral strand.

Its starting position and its localization within the nucleotide sequence of the '300-fragment' are exactly known. Also, the starting position and the sequence of the 5'-terminal part of G'-RNA have been determined[9] (Hulsebos and Schoenmakers, unpublished results).

In this chapter we have deduced the nucleotide sequence of G"-RNA and determined in this way its starting position and the strand from which it is transcribed. Taking these data together, it is now possible to give a very accurate transcription map of the '300-fragment'. A model, which might explain the peculiar transcription pattern of this fragment will be discussed.

MATERIALS AND METHODS

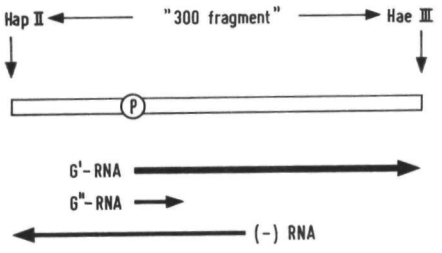
All materials and methods, including the preparation of ^{32}P-labelled transcripts, gel electrophoresis, recovery of RNA from the gel and most of the RNA sequencing methods have been described in detail in a previous communication[10] (cf. Chapter III). Two additional RNA sequencing techniques were used. Digestion of oligonucleotides with RNase-U_2 was carried out in 10 µl of 0.05 M sodium acetate, pH 4.5, 2 mM EDTA and 0.1 mg of bovine serum albumin at 1 unit/ml for 4 hours of 37°C. The digestion products were separated by electrophoresis on DEAE-paper (Whatman DE 81) in 7% formic acid[11] and analyzed by alkaline digestion.

For the analysis of the 5'-terminal end, G"-RNA labelled with (α-^{32}P)-GTP was digested with pancreatic RNase. The digest was fractionated by two-dimensional chromatography on polyethyleneimine (PEI)-cellulose thin-layer plates (Macherey-Nagel Co, 20 cm x 20 cm)[12]. The plate was developed in the first dimension with 1.3 M lithium formate-7 M urea (pH 3.5) for 2 hours and then with 1.8 M lithium formate-7 M urea (pH 3.5) for 6 hours. The plate was washed with methanol and developed

in the second dimension with 0.6 M lithium chloride-7 M urea-0.02 M Tris (pH 8.0) for 6 hours. The 5'-terminal oligonucleotide, containing the triphosphate end, stayed at the origin under these conditions. It was eluted from the plate and digested to completion with RNase-T1. The resulting products were fractionated by electrophoresis on DEAE-paper and their radioactive content determined.

RESULTS AND DISCUSSION

In vitro transcription of the '300-fragment' results in the synthesis of three RNA's (Fig 2A). The regions from which two of these RNA's, G'-RNA and (-)RNA, originate are already known from sequencing studies[9,10] (Hulsebos and Schoenmakers, unpublished results).



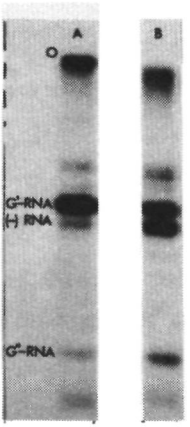


Fig 1 Transcription of the '300-fragment'.
P refers to the promoter $G_{0.18}$·G'-RNA and G"-RNA are transcribed from the complementary stand, whereas (-)RNA is transcribed from the viral strand.

Fig 2 Electrophoretic analysis on a 4% polyacrylamide gel of the (α-^{32}P)-GTP labelled products, formed upon transcription of the '300-fragment' in the presence (A) or absence (B) of the dinucleotide GpC.

As the strand specificity and localization of the third RNA band, G"-RNA, are still unknown, we have determined the nucleotide sequence of this small RNA. For this purpose, wild-type '300-fragment' was transcribed under unprimed conditions with each of the four (α-^{32}P)-ribonucleoside triphosphates. After fractionation of the various transcription products on a 4% polyacrylamide gel, G"-RNA was extracted and, after purification, digested to completion with RNase-T1. The resulting products were fractionated by electrophoresis at pH 3.5 on cellulose acetate in the first dimension, followed by homochromatography with homomixture C[13] on a DEAE-cellulose plate in the second dimension. A typical T1-fingerprint of G"-RNA, labelled with (α-^{32}P)-GTP only, is shown in Fig 3.

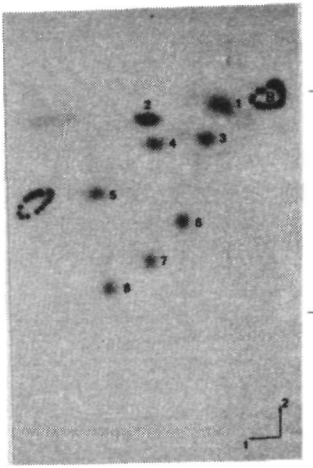


Table 1 RNase A - DIGESTION PRODUCTS OF T1 RNase RESISTANT OLIGONUCLEOTIDES

Oligonucleotide[a]	G"-RNA labelled *in vitro* with (α-^{32}P)-labelled				sequence[b]	relative molar yields[c]
	GTP	ATP	UTP	CTP		
$T1_{a,b}$	G	-	G	-	G(U); G(G)	1; 3
T2	U	-	G	-	UG(U)	2
T3	$\overset{*}{A}G$	-	$\overset{*}{A}\overset{*}{G}$	-	AG(U)	1
T4	$\overset{*}{A}\overset{*}{U}$	G	$\overset{*}{A}\overset{*}{U}$	-	AUG(A)	1
T5	$\overset{*}{A}\overset{*}{G}$	U	3U; $\overset{*}{A}\overset{*}{G}$	-	UUUUAG(U)	1
T6	$\overset{**}{A}A\overset{*}{G}$	$C,\overset{**}{A}A\overset{*}{G}$	-	U	UCAAAG(A)	1
T7	C	-	2U; 2C; G	2U; C	CCUCUUUCG(U)	1
T8	C	U	2U; C; $\overset{**}{A}\overset{*}{U}$	2U; G	UAUU(CU,UU)CG(C)[d]	1

a. The numbers refer to the fingerprint shown in Fig 3.

b. The proposed nearest-neighbour bases are indicated in parenthesis. The underlined sequence was derived by partial spleen phosphodiesterase digestion analysis.

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

d. RNase U2 digestion yielded two products which were fractionated by electrophoresis on DEAE-paper (Whatman DE 81) in 7% formic acid. The fast moving product (mobility 1.94 with respect to the xylene cyanol marker) proved to be UA.

e. Asterisks in column 2 denote transfer of label after digestion of the pancreatic RNase products with alkali.

Fig 3 Autoradiograph of the T1-fingerprint of G"-RNA, labelled with (α-^{32}P)-GTP.
The numbers on the fingerprint correspond to the oligonucleotide listed in Table 1.

The individual T1-oligonucleotides were characterized by digestion with pancreatic RNase and fractionation of the products by electrophoresis on DEAE-paper at pH 3.5. Subsequently, the pancreatic RNase

products were further subjected to alkaline hydrolysis and the resulting mononucleotides were separated by electrophoresis on Whatman 540-paper at pH 3.5 for nearest neighbour analysis and determination of base composition. The sequence of oligonucleotides T1, T2, T3, T4, T5 and T6 could unambiguously be deduced from these analyses. Additional information, however, was necessary to establish the sequence of oligonucleotides T7 and T8. The sequence of T7 was deduced in conjunction with the information obtained by partial spleen phosphodiesterase analysis of the oligonucleotide, labelled with (α-^{32}P)-GTP (Table 1). The nucleotide sequence of T8 could be established in part by applying the purine-specific enzyme RNase-U2. Digestion of T8 with the latter enzyme yielded only two products, of which one was 5'-UA-3' and the other the remaining part of T8. Combination of this information with the pancreatic RNase digestion data results in the tentative sequence UAUU(CU,UU)CG(C) for this oligonucleotide.

By comparison of the deduced nucleotide sequences of the T1-oligonucleotides with the DNA sequence of the '300-fragment' the region from which G"-RNA is transcribed could be determined. G"-RNA covers at least the region between nucleotide positions 101 and 144, as the oligonucleotides T1-T8 originate from this region (Fig 4). It is furthermore concluded that this small RNA is transcribed from the complementary strand.

Comparison with the nucleotide sequence of G'-RNA reveals that G"-RNA constitutes the 5'-terminal region of this RNA. G'-RNA starts with pppGp at nucleotide position 100[9] (Hulsebos and Schoenmakers, unpublished results). Although it is very probable that G"-RNA also starts with pppGp, such a conclusion could not be inferred from the fingerprinting results. Oligonucleotides with triphosphate ends have a strong tendency to smear during the fractionation procedure and, hence, are not recognized as a discrete spot on the autoradiographs[15]. On the other hand,

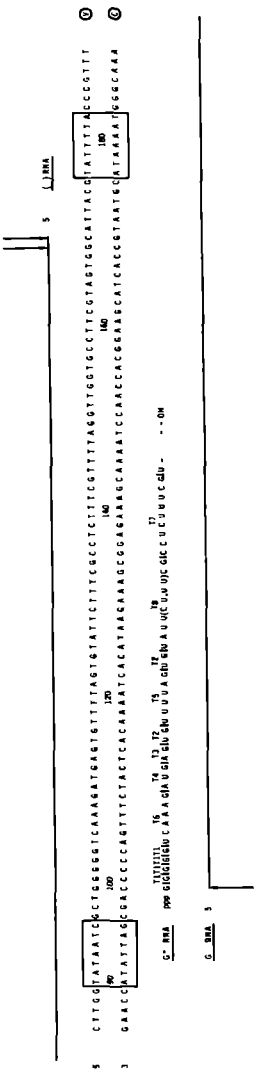


Fig 4 Detailed transcription of the region between nucleotide positions 85 and 190 on the '300-fragment'. V and C denotes viral and complementary strand, respectively. The RNA polymerase binding site sequences ('Pribnow Boxes') are indicated.

a starting position at nucleotide 100 is very probable, since RNase-T1 digestion of (α-^{32}P)-GTP labelled G"-RNA yielded 3 moles of Gp in addition to the other RNase-T1 oligonucleotides, as expected if G"-RNA synthesis starts at this position (Table 1). In addition, an oligonucleotide with the sequence CUG(G), derived from positions 98 to 100, has never been observed.

To confirm that G"-RNA starts with pppGp at nucleotide position 100, a 5'-terminal analysis has been carried out with G"-RNA, labelled with (α-^{32}P)-GTP. The latter RNA was digested with pancreatic RNase and the digestion products were fractionated by two-dimensional chromatography on a PEI-cellulose plate[12] using conditions under which triphosphate containing oligonucleotides do not migrate. The radioactive spot at the origin was eluted from the plate and digested with RNase-T1. The products formed were then fractionated by electrophoresis on DEAE-paper at pH 3.5 and their radioactivity determined. Two spots were seen, one at the origin, the other with a mobility of 1.80 relative to the xylene cyanol marker and a radioactive content of 860±20 cpm and 1230±30 cpm, respectively. These results can only be explained if one assumes that upon digestion of G"-RNA with pancreatic RNase, a 5'-terminal product pppGpGpGpGpGpUp, is formed which, after labelling with (α-^{32}P)-GTP and digestion with RNase-T1 yields pppGp and Gp in a radioactive ratio of 2:3. From these results and the foregoing data we conclude that G"-RNA starts at nucleotide 100.

The 3'-terminal end of G"-RNA could not be determined very accurately. In each T1-digest of G"-RNA, oligonucleotide CCUCUUUCG(U) (T7) was always present in a relative molar yield of one (Table 1). On the other hand, the oligonucleotide UUUUAG(G) (positions 145-150) has never been found. From these observations we infer that G"-RNA terminates between nucleotide positions 146 and 150.

The starting nucleotide of primed (-)RNA has been determined previously[10] (cf. Chapter III). Whether (-)RNA, synthesized in the absence of primer, starts at the same position is not known. However, we assume that this RNA starts at or very near the starting position of primed (-)RNA because of the following reason: the rate-limiting step in the initiation of RNA chains is the formation of the first phosphodiester bond. Dinucleotides can overcome this restriction and stimulate the synthesis of RNA chains in case they are complementary to the normal transcriptional initiation site or to a limited region around this site[14-19]. It is assumed that these dinucleotides act by basepairing with the DNA at or near the initiation site, providing 3'-OH ends which can then be extended by RNA bases. The dinucleotide GpC stimulates in this way the synthesis of (-)RNA (see Fig 2B). As was reported previously, other dinucleotides had no significant stimulatory effect[10]. So, most probably, the transcriptional initiation site of unprimed (-)RNA is located in the region in which GpC hybridizes (indicated by the arrows in Fig 4).

G"-RNA has no sequence characteristics which could explain that the termination of its synthesis occurs between nucleotides 146 and 150. Termination of transcription, either *in vivo* or *in vitro*, occurs at transcriptional terminators. RNA's ending at such termination sites have in general a stretch of six to eight uridinylate residues at their 3'-OH terminal end, which is preceeded by a G-C rich region, capable of intramolecular basepairing (for a review, see ref 20). However, these characteristic features are absent in the 3'-terminal end of G"-RNA. The mechanism of termination of G"-RNA synthesis is not known, but we assume that this RNA is formed by a blockage of transcriptional elongation caused by RNA polymerase molecules, which are attached to the promoter of (-)RNA. The promoters for *E.coli* RNA polymerase, sequenced so

far, all have two regions of at least partial homology. One region is located about 35 basepairs in front of the transcriptional initiation site and is involved in the initial recognition of the promoter by RNA polymerase[21].

The second region is the so-called Pribnow Box, $\begin{matrix} 5' \text{ TATPuATG } 3' \\ 3' \text{ ATAPyTAC } 5' \end{matrix}$ (Pu is purine base, Py is pyrimidine base) five to seven nucleotides upstream from the transcriptional start site[22]. The latter region provides a binding site for the RNA polymerase molecules. Such a binding site sequence is present in front of the (-)RNA initiation site (indicated by a box around nucleotide 180 in Fig 4). A second Pribnow Box, with opposite orientation, can be found in front of the start of G'-RNA. It constitutes the binding site sequence of the already extensively characterized $G_{0.18}$ promoter[9] (Hulsebos and Schoenmakers, unpublished results).

RNA polymerase molecules, attached to binding sites, cover in general 40-50 basepairs of which 15-20 code for the RNA that is initiated[22]. This would imply that the RNA polymerase binding site of (-)RNA extends to nucleotides 150-155 on the '300-fragment'.

Given the very low synthesis of (-)RNA under unprimed conditions we suppose that RNA polymerase molecules do bind at the (-)RNA promoter but initiate RNA synthesis only rarely. On the other hand, RNA polymerase molecules bind and initiate very efficiently at the strong $G_{0.18}$ promoter resulting in the synthesis of significant amounts of G'-RNA. This synthesis is most probably hindered by RNA polymerase molecules which are firmly attached to the (-)RNA promoter, ultimately leading to the formation of G"-RNA. A similar blocking phenomenon resulting in premature termination of transcription at promoters has been observed during the *in vitro* transcription of ϕx174 DNA[19] and T7 DNA[23].

In this study the transcription of the gene VII boundary was performed under *in vitro* conditions. Whether this aberrant transcription is

also operative in the infected cell is not clear at this moment, because we do not know exactly how far the transcription conditions used resemble the *in vivo* situation. It seems possible that RNA polymerase binding to the (-)RNA promoter is a mechanism for regulating the amount of gene VIII-protein, formed by reducing downstream transcription. A similar regulatory mechanism has been proposed by Kassavetis *et al* [23] in case of phage T7 DNA transcription. Further studies are needed to investigate the possible significance of this regulatory phenomenon for the *in vivo* expression of the filamentous phages.

1. Jacob, E. and Hofschneider, P.H. (1969) J. Mol. Biol. *46*, 359-363.
2. Smits, M.A., Simons, G., Konings, R.N.H. and Schoenmakers, J.G.G. (1978) Biochim. Biophys. Acta *521*, 27-44.
3. Sugiura, M., Okamoto, T. and Takanami, M. (1969) J. Mol. Biol. *43*, 299-315.
4. Chan, T.-S., Model, P. and Zinder, N.D. (1975) J. Mol. Biol. *99*, 369-382.
5. Edens, L., Konings, R.N.H. and Schoenmakers, J.G.G. (1975) Nucleic Acids Res. *2*, 1811-1820.
6. Okamoto, T., Sugimoto, K., Sugisaki, H. and Takanami, M. (1975) J. Mol. Biol. *95*, 33-44.
7. Pieczenik, G., Horiuchi, K., Model, P., McGill, C., Mazur, B.J., Vovis, G.F. and Zinder, N.D. (1975) Nature *253*, 131-132.
8. Edens, L., Wezenbeek van, P., Konings, R.N.H. and Schoenmakers, J.G.G. (1976) Eur. J. Biochem. *70*, 577-587.
9. Sugimoto, K., Sugisaki, H., Okamoto, T. and Takanami, M. (1977) J. Mol. Biol. *111*, 487-507.
10. Hulsebos, T. and Schoenmakers, J.G.G. (1978) Nucleic Acids Res. *5*, 4677-4698.
11. 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, pp 751-795.
12. Mirzabekov, A.D. and Griffin, B.E. (1972) J. Mol. Biol. *72*, 633-643.
13. Brownlee, G.G. and Sanger, F. (1969) Eur. J. Biochem. *11*, 395-399.
14. Downey, K.M., Jurmark, B.S. and So, A.G. (1971) Biochemistry *10*, 4970-4975.
15. Maizels, N.M. (1973) Proc. Natl. Acad. Sci. U.S.A. *70*, 3585-3589.

16. Hoffman, D.J. and Niyogi, S.K. (1973) Proc. Natl. Acad. Sci. U.S.A. *70*, 574-578.

17. Minkley, E.G. and Pribnow, D. (1973) J. Mol. Biol. *77*, 255-277.

18. Küpper, H., Contreras, R., Landy, A. and Khorana, H.G. (1975) Proc. Natl. Acad. Sci. U.S.A. *72*, 4754-4758.

19. Axelrod, N. (1976) J. Mol. Biol. *108*, 753-770.

20. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. *13*, 319-353.

21. Gilbert, W. In: RNA polymerase, Losick, R. and Chamberlin, M., Eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1976) pp 193-205.

22. Pribnow, D. (1975) Proc. Natl. Acad. Sci. U.S.A. *72*, 784-788.

23. Kassavetis, G.A., Kaya, K.M. and Chamberlin, M.J. (1978) Biochemistry *17*, 5798-5804.

CHAPTER V

NUCLEOTIDE SEQUENCE OF A REGION IN BACTERIOPHAGE M13 DNA ENCOMPASSING THE GENES X, V, VII, IX AND VIII

Part of the results described in this Chapter has been published: Wezenbeek van, P.M.G.F., Hulsebos, Th.J.M. and Schoenmakers, J.G.G. (1980) Gene *11*, 129-148.

CHAPTER V

NUCLEOTIDE SEQUENCE OF A REGION IN BACTERIOPHAGE M13 DNA ENCOMPASSING THE GENES X, V, VII, IX AND VIII

ABSTRACT

The nucleotide sequence of the region in bacteriophage M13 DNA encompassing the genes X, V, VII, IX and VIII has been deduced with the aid of the Maxam-Gilbert method. The sequence analysis revealed that gene X constitutes the C-terminal part of gene II and that both genes are using the same reading frame. The positions of the genes V and VIII in the DNA sequence were traced with the help of their known amino acid sequences. Together with the previously determined positions of the genes VII and IX a detailed picture of the genetic organization of this part of the M13 genome can now be given. Furthermore, the primary structures of the three promoters $G_{0.06}$, $G_{0.12}$ and $G_{0.18}$, operating in this part of the genome, and of the ribosome binding sites preceeding the M13 genes have been elucidated. In addition, the precise location of the rho-independent central terminator at which all *in vitro* transcripts terminate has been deduced. The presence of secondary structures in these transcripts and their possible involvement in the differential expression of this part of the genome have been discussed.

INTRODUCTION

The small filamentous bacteriophage M13 is composed of a circular single-stranded DNA molecule (6400 bases, mol.wt. 2×10^6) encapsulated in a protein coat consisting of approximately 2700 B-protein monomers (mol.wt. 5200), encoded by gene VIII, and 5-6 molecules of A-protein (mol.wt. 42,600), encoded by gene III[1]. Recently, two additional minor

components were discovered in the M13 phage coat: C-protein (mol.wt. 3650), encoded by gene IX and D-protein (mol.wt. 11,500) of which the genetic origin is still unknown[2].

After infection, the circular single-stranded DNA molecule is converted into the double-stranded replicative form (RF) by synthesis of a complementary strand. *In vitro* studies have indicated that the RF template is transcribed according to a cascade mechanism. According to this mechanism transcription starts at nine promoters, but all transcripts terminate at one site on the M13 DNA genome (for a recent review, see ref 3). This *rho*-independent termination site, the so-called central terminator, is located immediately after gene VIII[4]. As a consequence of this mode of transcription, genes located in front of the central terminator are transcribed more frequently than those at greater distances (Fig 1).

In the region most frequently transcribed three promoters, designated $G_{0.06}$, $G_{0.12}$ and $G_{0.18}$ have been found by *in vitro* transcription studies[4-6]. In order to locate the positions of these promoters and of the central terminator at the nucleotide level, we have determined the nucleotide sequence of the region which encompasses these regulatory elements.

In addition, elucidation of the nucleotide sequence of this particular part of the M13 genome might enable us to give an explanation for the following phenomena:

i *In vitro* transcription and translation studies have indicated that the C-terminal part of gene II contains the coding information of a protein of about 12,000 dalton. The genetic origin of this protein, designated X-protein, is still unclear[7-10]. It is not yet known whether gene X, which encodes the X-protein, uses the same reading frame as gene II, starting at an internal initiation codon and ending at the termination

codon of gene II, or that it uses a different reading frame with its own initiation and termination codon.

ii *In vitro* transcription-translation of M13 replicative form DNA (RF) results in the synthesis of a gene VIII-protein which is considerably longer than the gene VIII-encoded product, present in M13 virions[8]. As the amino acid sequence of the latter protein, the major coat protein, has been determined[11-14], the DNA region which encodes this mature form can be traced at the nucleotide level. From examination of the nucleotide sequences of the flanking regions we might be able to deduce the amino acid sequence of the precursor and the N-terminal or C-terminal location of the precursor peptide.

iii By sequence analysis and mapping of amber mutants we have shown previously that gene VII and the hypothetical gene IX are located between gene V and gene VIII[15] (see Chapter III). In accordance with the cascade mode of transcription, the products of the latter genes are made in abundance both *in vivo*[16,17] and *in vitro*[7,8]. Although genes VII and IX are also very frequently transcribed, their translational products have never been detected among the products synthesized with the aid of various *in vitro* transcription-translation systems[7-10]. The mechanisms which regulate this differential expression are incompletely understood. It is quite clear, however, that next to a regulation at the level of transcription, a regulation at the translational level must be operative. A detailed knowledge of the nucleotide sequence of these genes and their neighbouring genes enables us to consider the involvement of factors which influence their translation, such as the polycistronic or monocistronic nature of the various phage-messages, the strength of ribosome binding sites and the presence of secondary structures in the RNA's.

In this study we have determined the nucleotide sequence of the region encompassing the M13-genes X, V, VII, IX and VIII. The deduced DNA sequence has a length of about 1300 nucleotides. The consequences of the derived nucleotide sequence for the regulatory phenomena, described above, will be discussed.

MATERIALS AND METHODS

Materials

The restriction enzymes *Hap* II and *Hae* III were isolated as described previously[18]. *Hga* I, *Hha* I and *Mbo* II were purchased from New England Biolabs. *Taq* I was obtained from the Microbiol. Res. Establishment, Porton. *Hinf* I was a generous gift of Dr. P. van Wezenbeek. Bacterial alkaline phosphatase and snake venom phosphodiesterase were from Wortington B.C. and polynucleotide kinase from P.L. Biochemicals. Dimethyl sulphate was obtained from Aldrich Co, hydrazine from Eastman Kodak Co. and piperidine from Fisher Scientific. These reagents were used without further purification. (γ-^{32}P)-ATP was prepared as described by Maxam and Gilbert[19] using HCl-free, carrier-free $^{32}PO_4^{3-}$, supplied by the Radiochemical Centre, Amersham, England.

M13 replicative form DNA and restriction fragments

M13 replicative form DNA (RF) was isolated from phage-infected cells as described by van den Hondel *et al* [20]. Restriction enzyme digestions of M13 RF were done according to the instructions of the enzyme manufacturers. The generated fragments were separated by electrophoresis on 4% or 5% polyacrylamide slab gels and recovered from the gels by previously published procedures[20].

5'-Terminal labelling of restriction fragments

Initially, DNA-fragments were dephosphorylated with Bacterial alkaline phosphatase and subsequently the denaturated fragments were labelled with T_4-polynucleotide kinase and (γ-^{32}P)-ATP, as described previously[15]. As this labelling procedure - at least in our hands - frequently resulted in severe losses of radioactive material, we preferred in the later stages of our sequencing studies a direct labelling procedure by making use of the exchange activity of polynucleotide kinase at neutral pH[21]. In these cases, restriction fragments (3-4 pmol) were dissolved in 100 µl of 10 mM Tris-HCl, pH 7.6, 1 mM EDTA, 7 mM $MgCl_2$ and 7 mM 2-mercaptoethanol or 1 mM dithiothreitol and transferred to an Eppendorf tube containing 100 pmol of dried (γ-^{32}P)-ATP (spec. act. 1000-2000 Ci/mmol). The exchange reaction was started by the addition of 2-3 units of polynucleotide kinase. After 30 min at 37°C, the reaction was terminated by the addition of an equal volume of phenol. The ^{32}P-labelled restriction fragments were phenol-extracted twice and precipitated from the waterphase with ethanol.

Separation of the 5'-labelled ends of restriction fragments

Restriction fragments, labelled at both 5'-ends, were cut into two or more smaller fragments with a second restriction enzyme. The digestion was carried out in a reaction volume of 100 µl using conditions in accordance with the restriction enzyme applied. The fragments, bearing a single ^{32}P-labelled 5'-end, were separated on 5% polyacrylamide slab gels (acrylamide/bisacrylamide ratio of 20:1) in 40 mM Tris, 20 mM sodium acetate, 2 mM EDTA, pH 7.8.

In those cases, in which the separation of the labelled ends could not be achieved by restriction enzyme digestion, the strands of fragments were separated as described by Maxam and Gilbert[19]. Successful

separation could routinely be achieved with fragments ranging in size from about 50-300 basepairs. Restriction fragments were first denaturated by heating for 2 min at 100°C in 80 µl of 0.38 N NaOH. Then, 20 µl of dye mixture (50% glycerol, 5 mM EDTA, 0.25% xylene cyanol FF, 0.25% bromophenol blue was added and the denatured fragments were applied to a 5% polyacrylamide slab gel (acrylamide/bisacryl amide ratio of 30:1) in 50 mM Tris-borate, pH 8.3 and 1 mM EDTA. The dimensions of the slab gels were 40 cm x 20 cm x 0.4 cm. The slots had a width of 3 cm and no more than 2 pmol denaturated restriction fragments were applied to each slot.

Electrophoresis was for 16-20 h at 10 mA. After autoradiography the separated strands of fragments were eluted from the gel, as described by van den Hondel *et al* [20].

DNA sequencing methods

To deduce the nucleotide sequence of the 5'-labelled ends of restriction fragments, these fragments were subjected to partial digestion with snake venom phosphodiesterase, as described in a previous report[15].

For DNA sequencing by chemical degradation the protocol of Maxam and Gilbert[19] was followed. Purine residues were partially methylated by dimethyl sulphate. Strong adenine/weak guanine cleavage was obtained by treatment of the partially methylated DNA with 0.1 N HCl followed by treatment with 0.1 N NaOH at 90°C. Strong guanine/weak adenine cleavage was achieved by heating at neutral pH and subsequent treatment with 0.1 N NaOH at 90°C. Cleavage at cytosine and thymine residues was obtained by partial hydrazinolysis followed by treatment with 0.5 M piperidine. Hydrazinolysis at thymine residues was suppressed by the presence of 2 M NaCl.

In a later stage of our sequencing studies two alternative cleavage

procedures were used, one specific for guanine residues, the other with a strong preference for adenine residues. A cleavage, exclusively at guanine residues was obtained by heating the DNA, partially methylated by dimethyl sulphate in 1 M piperidine at 90°C. The alternative strong adenine/weak cytosine cleavage reaction was carried out by heating the non-methylated DNA in 1.5 M NaOH at 90°C, followed by treatment with 1 M piperidine.

After chemical degradation, the products of the individual cleavage reactions were divided into three parts and fractionated on 10%, 20% and 25% polyacrylamide slab gels (acrylamide/bisacrylamide ratio 20:1, 30:1 and 30:1, respectively) containing 50 mM Tris-borate, pH 8.3, 1 mM EDTA and 7 M urea. The dimensions of the slab gels were 40 cm x 25 cm x 0.1 cm. All gels were pre-electrophoresed for 2 h at 1000 V before application of the four cleavage reactions. Electrophoresis was for various times and at various (constant) voltages. In general, 10% slab gels were run for at least 7 h at 1250 V, 20% slab gels for 16-20 h at 1000 V and 25% slab gels for 4 h at 1250 V or 16 h at 650 V. After electrophoresis, the gels were covered with Saran Wrap and autoradiographed for 1-4 weeks at -20°C.

The nucleotide sequences were stored and studied using the computer programmes devised by Staden[22].

RESULTS

The chemical degradation method of Maxam and Gilbert[19] has been used for the sequence determination of the region covering the genes X, V, VII, IX and VIII. Fig 2 shows the approximate location of these genes. It also shows the cleavage sites of the restriction enzymes *Hap* II, *Hae* III, *Taq* I, *Hinf* I, *Hha* I, *Mbo* II and *Hga* I in this region.

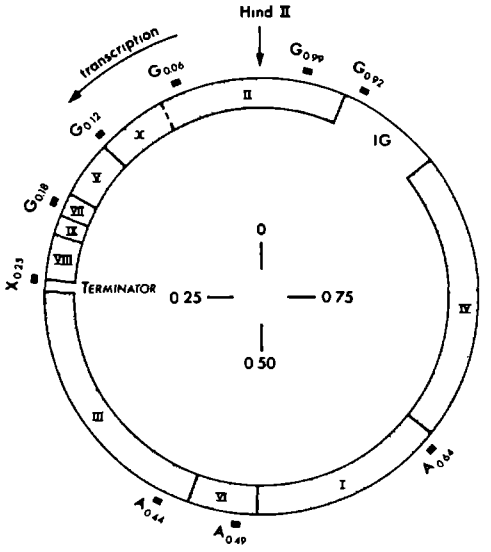


Fig 1 Circular genetic map of the bacteriophage M13 genome. The Roman numerals refer to the genes. The direction of transcription is as indicated. The approximate positions of the G-promoters, the A-promoters and the promoter in front of gene III ($X_{0.25}$) are indicated with black bars. The latter promoter forms an integral part of the central terminator. IG refers to the intergenic region in which the replication origins for both the viral and complementary strand are located. The single site at which RF DNA is cleaved with restriction enzyme *Hind* II is indicated.

All cleavage sites have been mapped before[23] (Chapter III, P. van Wezenbeek, unpublished results) except for the single *Hga* I cleavage site, which emerged from the deduced DNA sequence. Also, the presence of two, instead of one, cleavage sites for the restriction enzyme *Hin*fI, generating the very small fragment *Hin*fI-Q_2, became clear during the course of this sequence work. From the distribution of the various cleavage sites, it is evident that fragments can be generated now by (a combination of) these enzymes which are small enough for an unambiguous sequence determination.

To deduce the nucleotide sequence of this region, the individual fragments were labelled at their 5'-ends with (γ-^{32}P)-ATP and poly-

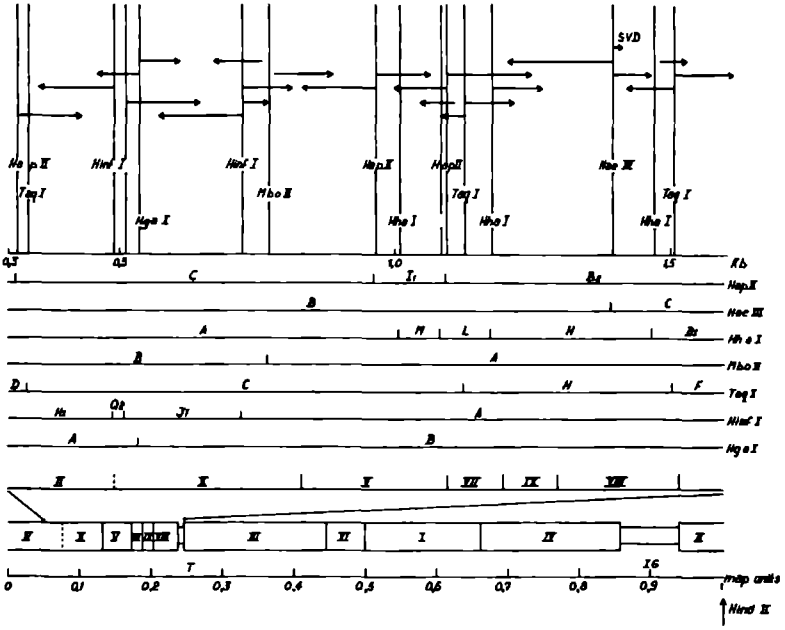


Fig 2 Linear genetic map of bacteriophage M13 and the restriction enzyme cleavage maps of the DNA region encompassing the genes X, V, VII, IX and VIII. The symbol T stands for the *rho*-independent termination site of transcription. IG refers to the intergenic region in which the replication origin for the viral and complementary strand are located. The arrows represent extent and direction of the individual sequence determinations by means of the chemical degradation method of Maxam and Gilbert[19]. The sequence designated SVD has been deduced by partial snake venom phosphodiesterase digestion.

nucleotide kinase. Preparation of fragments carrying a single 5'-labelled end was achieved by cleavage with a second restriction enzyme, followed by separation of the resulting fragments on polyacrylamide slab gels. In some cases, a successful separation of terminally labelled ends could not be achieved by this procedure due to a peculiar distribution of restriction enzyme cleavage sites. Single 5'-end labelled molecules were then obtained by strand separation. An example is shown in Fig 3D, in which the individual strands of fragment *Hap* II-I_1, were

separated by alkaline denaturation and electrophoresis on a 5% polyacrylamide gel, as described in Materials and Methods. Also, the fragments 1, 2 and 3, generated by cleavage of fragment *Hap* II-C with *Hinf* I (Fig 3A) were subjected to this procedure. As is shown in Fig 3B and 3C, the strands of fragment 2 and of fragment 3 were separated after electrophoresis, although some renaturation occurred. Due to its length, the strands of fragment 1 could not be separated under the conditions used (data not shown).

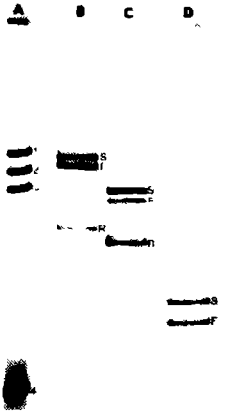


Fig 3 Strand separation of restriction fragments
Fragment *Hap* II-C was digested with restriction enzyme *Hinf* I. The resulting fragments were labelled at their 5'-ends with polynucleotide kinase and (γ-^{32}P)-ATP and fractionated on a 5% polyacrylamide gel (acrylamide/bisacrylamide ratio of 20:1) (A). Fragments 2 and 3 were eluted from the gel and their strands were separated by alkaline denaturation and electrophoresis on a 5% polyacrylamide gel (acrylamide/bisacrylamide ratio of 30:1) as described in Materials and Methods (B and C, respectively). D shows the separation of the individual strands of fragment *Hap* II-I_1, ^{32}P-labelled at both ends, on a 5% polyacrylamide gel (acrylamide/bisacrylamide ratio 30:1). S and F refer to the slow and fast strand, respectively. R indicates renatured fragment.

Restriction fragments which were labelled at one end, or labelled separated strands were subjected to chemical degradation by the method of Maxam and Gilbert[19]. The reaction products were divided into three parts and applied to 25%, 20% and 10% DNA sequencing gels. The 25% slab gels were used to read nucleotides 2-30 from the labelled terminus, 20% gels for nucleotides 20-70, whereas on 10% gels, after a large number of nucleotides were allowed to run off, a reading could be made until nucleotides 150-170.

Some examples of DNA sequencing gels are shown in Fig 4A-4E. The extent and direction of the sequence determinations are indicated by arrows in Fig 2. Several measures were taken to ensure the accuracy of the final sequence. For instance, each sequence was established by at least two independent sequence determinations. Moreover, as can be seen in Fig 2, a large part of the sequence was determined on both the viral and complementary strand. Because of the presence of a rather large number of restriction sites, extensive parts of the DNA were sequenced in the same direction starting from two different restriction sites. Also, it was possible to confirm, in this way, the positions of all previously established restriction sites in this part of the genome.

Moreover, additional cleavage sites, which had escaped detection by standard restriction enzyme analysis, could now be traced (like one of the *Hin*6 I cleavage sites at nucleotide positions 490 or 511, indicated in Fig 4B). Furthermore, the exact location of cleavage sites of several other restriction enzymes, of which no maps have been constructed, could be predicted from the DNA sequence (like the *Hga* I cleavage site at nucleotide position 535).

The recognition sequence of the latter enzyme, which is separated from the cleavage site (*cf.* 5'-GACGCNNNNN↓ -3' / 3'-CTGCGNNNNNNNN↑ -5') is indicated in Fig 4A.

There are only two regions in the DNA sequence where we could not sequence across a restriction site, *i.e.* the *Hap* II-C/I_1 junction and the *Hae* III-B/C junction. In both instances we can exclude the possibility that there are two (or more) cleavage sites next to each other. Fig 4C shows the left-hand side sequence and Fig 4D the right-hand side sequence determined at the *Hap* II-C/I_1 junction. Both sequences start with -G-G-, the second and third nucleotide in the *Hap* II recognition sequence, 5'-C↓CGG-3' / 3'-GGC↑C-5', so we feel confident to conclude that we actually have sequenced the 5'-ends of fragment *Hap* II-C and *Hap* II-I_1. Based

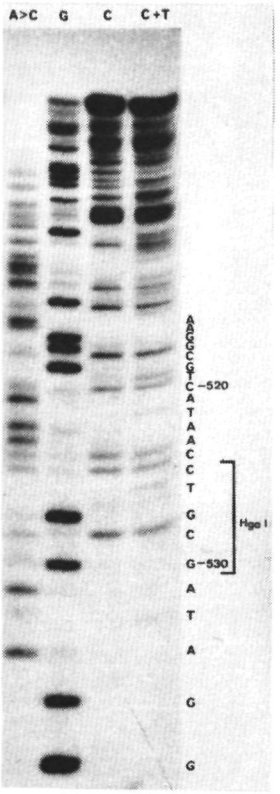


Fig 4A

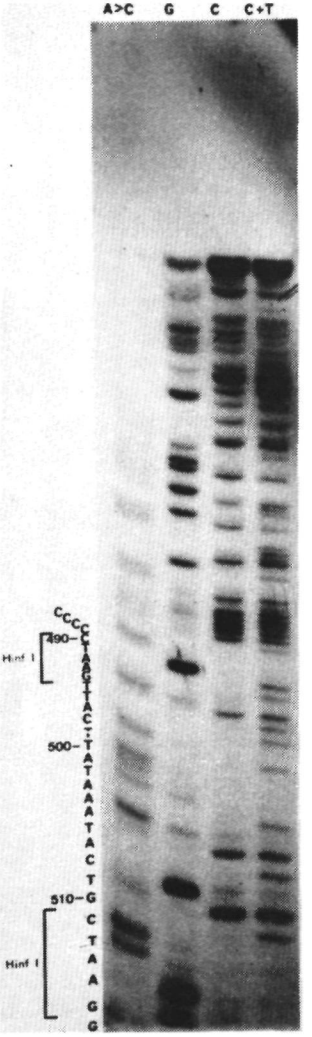


Fig 4B (right)

Fig 4 Autoradiographs of sequencing gels obtained after chemical degradadation of: (A,B) the right-hand 5'-end of fragment *Hap* II-*Hga* I; (C) the right-hand 5'-end of fragment *Hinf* I-*Hap* II (fragment 1 in Fig 3A); (D) the 5'-end of the slow strand of fragment *Hap* II-I_1 (cf. Fig 3D) and (E) the left-hand 5'-end of fragment *Taq* I-F. After labelling the 5'-ends, fragment *Hap* II-*Hga* I was digested with *Taq* I, fragment *Hinf* I-*Hap* II with *Mbo* II and *Taq* F with *Hap* II. The chemical degradation products were fractionated on 25% (A,C,D) or 20% (B,E) polyacrylamide slab gels.

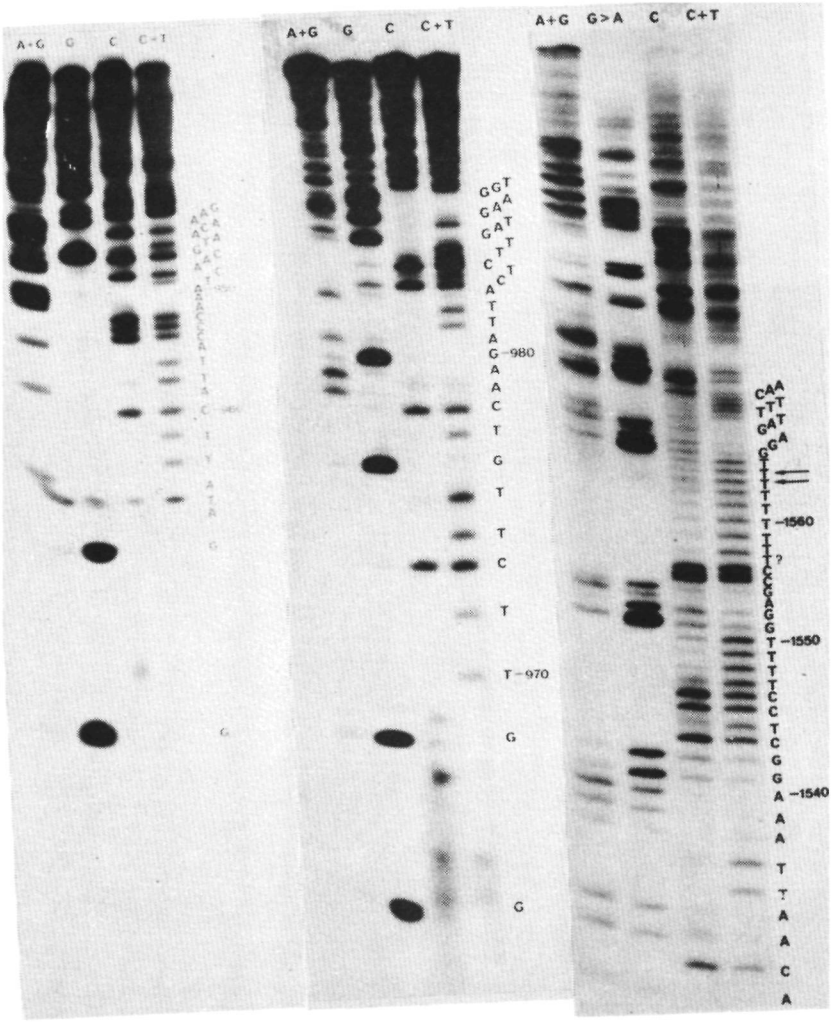


Fig 4C

Fig 4D

Fig 4E

(Legend, see page 119)

upon the amino acid sequence of the M13 gene V-protein[24], the left-hand side sequence appears to code for amino acid residues 32-42 (nucleotides 936-967) and the right-hand side sequence for amino acid residues 43-52 (nucleotides 968-998) of this protein. From this observation we conclude that both sequences originate from the same *Hap* II cleavage site and that the presence of a repetition of *Hap* II cleavage sites can be excluded.

The same conclusion can be drawn for the *Hae* III-B/C junction. The sequence of the left-hand side appears to code for the 5th to 9th amino acid residue of the mature gene VIII-protein, as was established in our previous studies[15]. The sequence of the right-hand side was determined both by the chemical degradation method (data not shown) and by partial digestion with snake venom phosphodiesterase, essentially as described previously[15]. The sequence, which can be deduced from the mobility shift pattern, shown in Fig 5, specifies the amino acid residues 10 to 15 (nucleotides 1398-1413) of the mature gene VIII-protein. The possibility that there are two or more *Hae* III cleavage sites next to each other in this part of the genome is therefore excluded.

In Fig 6 the elucidated nucleotide sequence of the region covering the genes X, V, VII, IX, VIII is presented. As the viral DNA strand has the same polarity as the viral mRNA's, the sequence depicted in Fig 6 represents the viral strand sequence in a 5' to 3' direction. The derived viral DNA sequence has a length of about 1300 nucleotides and is numbered using the single *Hind* II cleavage site indicated in Fig 2 as zero (reference) point.

Localization of genes and regulatory elements

Knowledge of the nucleotide sequence makes it now possible to localize the positions of the genes X, V, VII, IX and VIII with their respective

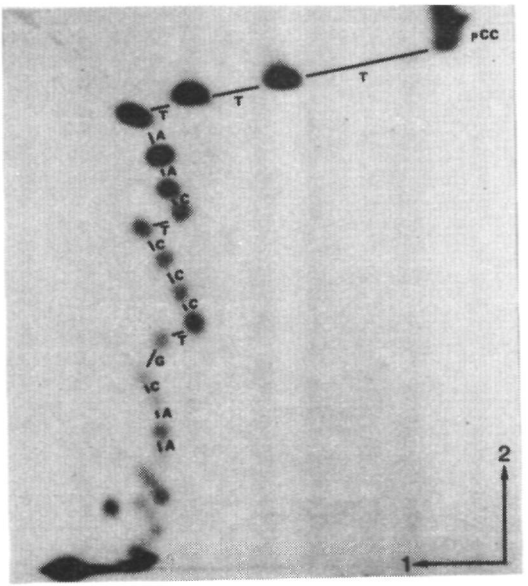


Fig 5 Autoradiograph of a two-dimensional fingerprint of oligonucleotides derived after partial digestion with snake venom phosphodiesterase of the left-hand 5'-end, labelled with ^{32}p, of fragment *Hae* III-C.

ribosome binding sites at the nucleotide level. Also, the exact positions of the various promoters and the *rho*-independent central terminator, which are operative in this region, can be deduced now.

Localization of genes

In vitro transcription-translation studies of M13 RF revealed the presence of a polypeptide, X-protein (mol.wt. 12,000), for which no gene-product relationship could be established[8]. As X-protein is also synthesized in a DNA-dependent cell-free system under the direction of fragment *Hap* II-C[9], it can be concluded that the coding region for X-protein is contained within this fragment, *i.e.* between nucleotide positions 316 and 966 (see Fig 6).

We have looked for open reading frames between these nucleotide positions, long enough to code for a polypeptide with a molecular weight of approximately 12,000. There appears to be only one such reading frame

which satisfies this condition. It starts with the codon ATG at position 497 and terminates with TAA at position 829. The protein encoded by this region contains 111 amino acid residues and has a molecular weight of 12,670. That the ATG codon at position 497 is most likely the initiation codon of gene X is strongly supported by the presence of a potential ribosome binding site in front of gene X. This site shows very good complementary with the 3'-OH terminal end of 16S-ribosomal RNA[25]. The nine nucleotides, capable of basepairing are underlined in Table 1.

Table 1 DNA SEQUENCES OF RIBOSOME BINDING SITES IN M13 DNA

16S rRNA	3'-A U U C C U C C A C U A G-5'	
gene X	T T T G A G G G G G A T T C A A T G A A T A T T T A T	
		Met Asn Ile Tyr
gene V	A T A A G G T A A T T C A C A A T G A T T A A A G T T	
		Met Ile Lys Val
gene VII	T T C C G G C T A A G T A A C A T G G A G C A G G T C	
		Met Glu Gln Val
gene IX	C G C T G G G G G T C A A A G A T G A G T G T T T T A	
		Met Ser Val Leu
gene VIII	A A T G G A A A C T T C C T C A T G A A A A A G T C T	
		Met Lys Lys Ser

The nucleotides which are complementary to the 3'-OH terminal end of 16S-ribosomal RNA are underlined.

Several investigations have indicated that gene X is contained within gene II[7-9,18]. However, it remained unclear whether both genes use the same reading frame or different ones. The derived nucleotide sequence now clearly demonstrates that the two other possible reading frames in the gene X region each give rise to a premature termination of gene II-protein synthesis and, consequently, have to be rejected. So, we conclude that gene II uses the same reading frame as gene X and terminates likewise with the TAA codon at position 829. That gene X in

fact represents the C-terminal part of gene II is supported by the observation of Model and McGill (quoted in ref 26), who demonstrated that the synthesis of X-protein of the closely related bacteriophage f1 is affected by a late amber mutant in gene II.

The biological significance of a gene, which is capable of directing the synthesis of two proteins, is not clear at present. It has to be emphasized that gene A of bacteriophage ϕx174 also gives rise to the synthesis of two proteins, A- and $\overset{*}{A}$-protein, of which the latter protein constitutes the C-terminal part of A-protein[27,28]. Similar functions have been attributed to the gene II-protein of M13 and the gene A-protein of ϕx174[29-33]. On the other hand, the mechanism of synthesis of X-protein and its ϕx174-analogue $\overset{*}{A}$-protein are completely different. $\overset{*}{A}$-protein is formed by a translational re-initiation event on the A-protein message, whereas gene X is equipped with its own promoter (*i.e.* $G_{0.06}$) and, consequently, can be translated from a separate messenger.

The DNA sequence starting with the ATG codon at position 843 and terminating with TAA at position 1106 corresponds exactly with the amino acid sequence of gene V-protein, as determined by Cuypers *et al* [24]. Gene V is separated from gene X (II) by a small intergenic region consisting of 14 basepairs. This region is part of the ribosome binding site of gene V (Table 1) which has exactly the same nucleotide sequence as the experimentally determined ribosome binding site sequence of gene V in bacteriophage f1[34]. The only difference is the presence of a cytosine residue at position 841 which is adenine in case of f1. However, this difference does not influence the Skine-Dalgarno basepairing with the 3'-OH terminal end of 16S-ribosomal RNA.

In a previous report we have localized two amber mutations, at positions 1114 and 1141, in gene VII[15]. It was concluded that gene VII covers the region from nucleotide 1108 to 1206 and codes for a protein

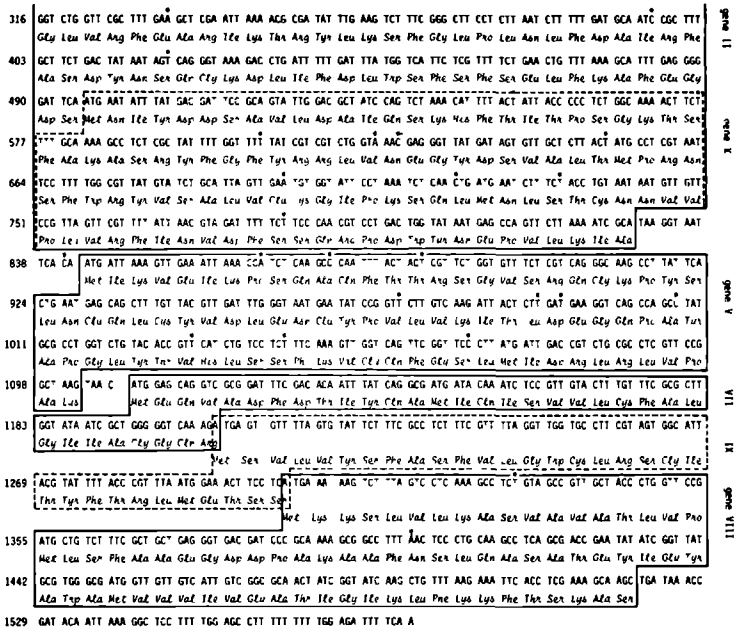


Fig 6 Nucleotide sequence of the genes X, V, VII, IX and VIII and the amino acid sequences of their encoded protein. Nucleotides are numbered in the 5'→3'-direction with the *Hind* II site (see Fig 2) as zero (reference) point.

with a length of 33 amino acid residues. Its potential ribosome binding site has a very low capacity for Shine-Dalgarno basepairing with the 3'-OH terminal end of 16S-ribosomal RNA (Table 1).

The existence of gene IX was suggested by the presence of an open reading frame between gene VII and gene VIII from nucleotide positions 1206 to 1301[15]. However, definitive proof for its existence could not be given because no amber mutations are presently known which are located in this gene. Evidence for the existence of gene IX has recently been provided by Simons *et al* [2], who demonstrated that one of the minor coat proteins, (*i.e.* C-protein), is encoded by gene IX. The ribosome binding site of gene IX

shows good complementarity with the 3'-OH terminal end of 16S-ribosomal RNA (Table 1).

The amino acid sequences of the major coat protein (encoded by gene VIII) of the related bacteriophages f1 and fd have been determined[11,13,14] and proved to be identical. Although the sequence of the M13 coat protein has not been solved yet, the amino acid compositions of these proteins from M13, fd and f1 show no differences[11], strongly suggesting that the major coat protein sequences of these filamentous phages are identical. This is confirmed by our sequence analysis of gene VIII. The DNA region from nucleotide position 1370 to 1519 in Fig 6 fits completely with the amino acid sequence of the mature major coat protein of fd and f1. The only difference is the twelfth amino acid residue in the mature coat protein, which is aspartic acid in fd/f1 and asparagin, encoded by AAC at position 1403, in M13. The latter difference might explain the differences in electrophoretic mobility of M13 and f1 virions in agarose gels, as noted by Moses *et al* [35].

The observations of Konings *et al* [8] suggest that the *in vitro* synthesized gene VIII-protein is larger than the mature gene VIII-protein, isolated from the filamentous phage coat. The mature gene VIII-protein starts with alanine instead of the usual methionine residue. This already indicates that gene VIII-protein is formed as precursor with the precursor peptide at the N-terminal end. Two potential initiation codons for this precursor can be found in front of the region encoding the mature form, at positions 1306 and 1355. The precursor most probably starts with the initiation codon at position 1306 because of the following reasons:

i The region in front of this initiation codon shows good complementarity with the 3'-OH terminal end of 16S-ribosomal RNA (Table 1) in contrast

to the region in front of the alternative initiation codon at position 1355.

ii The nucleotide sequence of the ribosome binding site in front of the initiation codon at position 1306 corresponds exactly to the experimentally determined ribosome binding site of gene VIII in bacteriophage f1[34].

iii Chang *et al* have partially sequenced the N-terminal part of the f1-precursor[36]. Their amino acid sequence data are in agreement with the theoretical amino acid sequence which can be predicted from the DNA sequence, starting at the initiation codon at position 1306. The N-terminal extension of the precursor has a length of 23 amino acid residues.

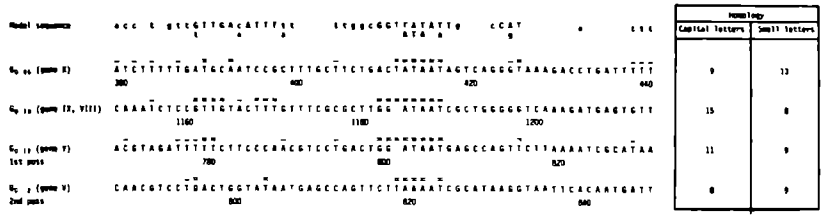
Localization of promoters and central terminator

To start transcription *E.coli* RNA polymerase recognizes and binds to a specific region of the DNA, called the promoter. By comparison of 54 procaryotic promoter sequences, Siebenlist *et al* [37] have demonstrated that promoters contain two regions of homology.

At first the 'binding sequence' or Pribnow box[38] with probable sequence TATAAT, at about 10 basepairs ahead of the RNA initiation site. Secondly, the 'recognition sequence'[39], with probable sequence TTGACA, about 35 basepairs upstream the RNA initiation site. However, these structural requirements are not absolute. Scherer *et al* [40] concluded from earlier studies that there is at most a preference for a particular base at a specific position in a promoter. Their model sequence, which includes the ideal 'recognition' and 'binding sequences' is shown in Table 2.

By *in vitro* transcription of M13 RF and/or M13 RF restriction fragments and by length measurements of the resulting RNA's, Edens *et al* [4-6,41]

Table 2 *DNA SEQUENCES OF PROMOTERS IN BACTERIOPHAGE M13*



An enlargement of this table is shown on page 204

The upper part of the Table shows the model promoter sequence of Scherer *et al* [40]. Capital letters indicate high preference, small letters low preference (as defined by Scherer *et al* [40]) for particular bases at specific positions. Both indicate no preference. The sequences of the promoters $G_{0.06}$ and $G_{0.18}$ and the two possible $G_{0.12}$ promoter sequences have been aligned for maximum homology with the model sequence. An exact base to base correspondence between model sequence and promoter sequence is indicated by a single or double bar, depending on the homology with a small or capital letter, respectively.

identified nine M13 promoters and deduced their (approximate) positions. In this way they located the promoters $G_{0.06}$, $G_{0.12}$ and $G_{0.18}$ in front of the genes X, V and VIII, respectively. We have used the results of these length measurements and the above mentioned model sequence to locate the latter three promoters more precisely within the deduced nucleotide sequence.

Promoter $G_{0.06}$ gives rise to the synthesis of an RNA species on fragment *Hap* II-C which starts with pppG and which has a length of about 570 nucleotides[5]. As the synthesis of this RNA species is most probably terminated at the 5'-terminal end of the non-viral strand of fragment *Hap* II-C, at position 966, this implies that the initiation site

of the RNA can be found around position 396 (Fig 6). Comparison of the nucleotide sequence around this position with the model sequence reveals that the DNA sequence from nucleotide positions 380 to 441 shows the best fit with the model sequence (Table 2). As stated before, the initiation site of an RNA species, starting at a particular promoter, can generally be found about 10 nucleotides distal to (the middle of) the TATAAT sequence. This would mean that the G-start RNA initiates at positions 423, 424 or 425.

This conclusion is strongly supported by the work of Sugimoto *et al* [42] and Heyden *et al* [43] who determined the sequence of the 5'-terminal region of this RNA derived from the closely related bacteriophage fd. The latter RNA also starts at one of the three indicated positions.

The weak promoter $G_{0.12}$ is also located on fragment *Hap* II-C. *In vitro* transcription of this fragment results in the synthesis of the RNA described above, and of a minor RNA which has a length of about 160 nucleotides. The 5'-terminus of this minor RNA is also pppG[5]. These combined data suggest that the latter RNA is initiated around nucleotide position 806. There are two regions in this part of the sequence which show partial homology with the model sequence. These are located between positions 770 and 831 and between positions 787 and 848 in Fig 6. Both promoter sequences are shown in Table 2. Because the RNA species which initiates at this promoter starts with pppG, this would mean that the most likely starting positions are 831 or 832 for the first possible promoter location and 814 for the second one.

Promoter $G_{0.18}$ has been located on fragment *Hap* II-B[24,5]. The 8S RNA species, initiating at this promoter, terminates in the central terminator. The latter regulatory element is also located on this fragment[4]. G"-RNA, of which the RNA sequence has been determined (Chapter IV), constitutes the 5'-terminal part of 8S RNA (or G-RNA). G"-RNA and 8S RNA

start with pppGGGGGG and are initiated at position 1195. This is in contrast to the situation in case of bacteriophage fd, where the corresponding RNA starts with four G residues at position 1196[44]. The M13 promoter $G_{0.18}$ which initiates the synthesis of 8S RNA, and which is located between nucleotide positions 1152 and 1199, shows very good homology with the model sequence (Table 2).

Several groups have demonstrated that *in vitro* initiation of transcription occurs at different sites on the filamentous phage genome, but that all transcripts terminate at a unique site, the so-called central terminator (reviewed in ref 3). As a consequence of this multipromoter-single terminator model all M13 transcripts should have identical 3'-terminal sequences. With the aid of standard RNA sequencing procedures Edens[45] has elucidated the 3'-OH terminal end of one of these transcripts, the 8S RNA. It contains a stretch of seven to eight uridenylate residues, preceeded by a G-C rich region. The latter region can be drawn in the form of a hairpin (shown in Fig 7e). This sequence characteristics are common to other 3'-OH terminal ends of procaryotic RNA's which end at *rho*-independent termination signals[46].

In Fig 4E the DNA sequencing gel is shown from which the DNA sequence of the central terminator region can be deduced. In this region the synthesis of 8S RNA is terminated at nucleotide positions 1563 of 1564. Non-regular spacing between the bands can be seen around position 1555. The aberrant mobility of the bands is most probably caused by the presence of the secondary structure (shown in Fig 7e) in the DNA, even under the strong denaturating conditions under which this gel was run. There is some uncertainty about the nucleotide at position 1557. Although this sequencing gel shows a C-residue at this position, in other sequencing gels the band in the C-track was less prominent or even missing, indicating that a T should be read. Also from the work of Edens[45],

who has found C_2U_7(U)-OH at the 3'-OH terminal end of 8S RNA, it is evident that a T residue is present at position 1557.

Interestingly, the DNA template permits the synthesis of nine U residues at the 3'-OH terminal end of 8S RNA. However, the sequence C_2U_9-OH has never been detected. 8S RNA terminates *in vitro*[44,45] and *in vivo*[47] predominantly with C_2U_8-OH and only rarely (less than 20%) with C_2U_7-OH.

The DNA sequence distal to the central terminator shows no extensive sequence homology or a relationship in secondary structure with comparable regions of other procaryotic *rho*-independent terminator[46]. This would imply that the region immediately distal to the termination signal is not important for the termination process. However, the involvement of more distantly located regions in the termination process, as was recently suggested for the termination of transcription at the end of the tryptophan operon in *E.coli*[48], cannot be excluded at this moment.

DISCUSSION

The base composition of bacteriophage M13 DNA shows a rather high T-content (35.8%, *cf.* ref 51). This frequent occurrence of T residues is also apparent in the DNA sequence deduced in this study (35.1%), although there is a considerable variation in the T-content of each gene (Table 3, column 3). Van Wezenbeek *et al* [49] noted a high frequency of codons with T at the third positon in the M13-genes III, VI and I (about 50%) analoguous to the situation found in ϕX 174 DNA[28]. The same preference for T in the third position can be noticed for the M13 genes X, V and IX (Table 3, column 4).

During the course of our sequencing studies, the groups of Schaller and Takanami[50] have determined the complete nucleotide sequence of the

Table 3 *ANALYSIS OF T-CONTENT OF THE GENES X, V, VII, IX AND VIII OF BACTERIOPHAGE M13*

gene	nucleotides	amino acids	total T-content[a]	codons with T at third position[b]
X	333	111	129 (39%)	64 (58%)
V	261	87	88 (34%)	44 (51%)
VII	99	33	29 (29%)	10 (30%)
IX	96	32	42 (44%)	15 (47%)
VIII	219	73	59 (27%)	20 (27%)

a. Values in brackets indicate percentage of nucleotides in a gene which are Thymine;
b. Values in brackets indicate percentage of codons in a gene with Thymine at the third position.

closely related bacteriophage fd. The nucleotide sequence, deduced in this study, differs at 26 positions from the corresponding sequence on the fd genome. The differences are indicated by asterisks in Fig 6 and are listed in Table 4.

Table 4 *NUCLEOTIDE INTERCHANGES IN THE GENE X-GENE VIII REGION OF BACTERIOPHAGE M13 AND FD*

M13 fd	positions
A↔ G	333, 696, 866, 1403 (1st base)
A↔ T	621
C↔ A	841 (intergenic)
C↔ G	875, 1007
C↔ T	396, 624, 718 (1st base), 1067
T↔ A	420, 552
T↔ G	971, 1031, 1043
T↔ C	576, 606, 651, 732, 783, 887 989, 992, 1330

Two of the base changes are found at the first base of triplets. The CTG codon at position 718 in M13 is replaced by TTG in fd. However, both triplets encode leucine. The adenine residue at position 1403 in M13 is changed into a guanine residue in fd. Only in this case, the nu-

cleotide substitution results in an amino acid replacement: AAC in M13 codes for asparagine, whereas GAC in fd codes for aspartic acid. The latter base change is also reflected in the *Hinf* I digestion pattern of both phage DNA's[52]. As the base change in M13 alters the recognition sequence GACTC of *Hinf* I into AACTC, the analogue of M13 fragment *Hinf* I-A is absent in fd. Another nucleotide substitution is situated in a non-coding region, namely in the small intergenic region between the genes X and V. Cytosine at position 841 in M13 is replaced by adenine in fd.

The remainder of the base differences (90%) concerns base changes at the third position of triplets. These nucleotide replacements are such, that the coding capacity of the respective triplets has not been changed. These 'silent' nucleotide substitutions support the established reading frames for the genes in this region.

The sequence differences are not equally distributed among the genes. By far the most changes are found in the genes, X and V, whereas no changes at all are present in the genes VII and IX. Another striking feature is the involvement of thymine in the nucleotide changes between M13 and fd. 19 Out of the 26 replacements (73%) are of the kind X↔T. We do not know the biological significance of this preference of base changes in which thymine is involved.

Taken together, it is evident that the sequence differences between M13 and fd are small in this region. The same conclusion can be drawn for other regions of the M13 and fd genome[49,50,53].

Edens *et al* [4-6,41] have demonstrated that *in vitro* transcription of M13 RF initiates at nine different promoters. All transcripts terminate at the *rho*-independent central terminator which is located immediately distal to gene VIII. Three of these promoters, $G_{0.06}$, $G_{0.12}$ and $G_{0.18}$ are operative in the gene X-gene VIII region. In this study these pro-

moters could be located very precisely by the search for nucleotide sequences, homologous to a model promoter sequence[40]. The search for the positions of the promoters was facilitated not only by the knowledge of the approximate lengths of the transcripts but also by our knowledge of the 5'-terminal nucleotide sequence of two of the three transcripts, initiated at these promoters[15,42-44] (see Chapter IV).

Two promoter sequences were found for $G_{0,12}$. Which of these two promoter sequences is actually involved in the *in vitro* synthesis of the 11S RNA species, which initiates at this promoter cannot be answered as the 5'-terminal sequence of this RNA is still unknown. On the basis of its better sequence homology with several other procaryotic promoters the one located between positions 770 and 831 (Table 2) is the more likely candidate, although we realize that good homology is only an indication but does not prove that such a sequence actually functions as a promoter.

Shine and Dalgarno[25] noticed that a purine-rich sequence preceeding the initiation codon of many genes shows complementarity with the 3'-OH terminal sequence of 16S-ribosomal RNA. They suggested that this sequence may participate directly, by basepairing with the 3'-OH terminal end of 16S-ribosomal RNA, in the binding of the 30S ribosomal subunit to the beginning of a gene. The potential ribosome binding sites listed in Table 1 show, with the exception of the ribosome binding site in front of gene VII, rather good complementarity with the 3'-OH terminal end of 16S-ribosomal RNA. Only the sequences in front of gene V and gene VIII have been isolated by binding of ribosomes to *in vitro* synthesized RNA's[34].

Our knowledge of the nucleotide sequence of recognition and binding sites of M13 promoters and the detailed information on the structure

of ribosome binding sites on itself does not provide us a deeper insight in the regulation of the expression of the M13 genes. Expression of the various phage genes is not only determined by the complexity of promoter selection by *E.coli* RNA polymerase, but is also determined by the efficiency of transcription and subsequent translation of the message. Furthermore, in most instances, the processes of transcription and translation are coupled, and therefore, the frequency of initiation of transcription at a particular promoter but also the efficiency of ribosome binding, the rate of movement of ribosomes along the growing mRNA chain, the distances between translational termination and initiation signals, etc, are all factors which might influence, though to a varying extent the expression of a certain gene.

As very recently also RNA processing events have been observed to occur with several M13 RNA transcripts[54-56] and the biological half-lives of individual M13 messages vary considerable[57-59,68] then it is clear that a variety of unknown structural factors have also to be considered in the regulation of gene expression. For this reason we have looked for the occurence of secondary structures within genes and their flanking regions, in order to find out whether such structures, if present, might contribute to our understanding of the regulation of M13 gene expression.

The potential ribosome binding site in front of gene X shows very strong complementarity with the 3'-OH terminal end of 16S-rRNA (Table 1). Notwithstanding this excellent complementarity, Pieczenick *et al* [34], upon incubation of the *in vitro* synthesized RNA's with ribosomes, could not isolate the ribosome binding site of this gene, despite the fact that the 14S RNA which contains the gene X message is one of the predominant species synthesized[5]. Moreover, in an RNA-dependent translation system the synthesis of gene X-protein is markedly depressed as com-

pared to its synthesis in a coupled transcription-translation system[6,8,9]. To our opinion these phenomena can be explained by assuming that the ribosome binding site of gene X is implicated in the very stable secondary structure, which is drawn in Fig 7a. With the aid of a computer program, designed to search for potential secondary structures, Huang and Hearst[60] classified the corresponding structure in the viral DNA as one of the most stable secondary structures in the filamentous phage genome. This structure may hinder the binding of ribosomes in front of gene X when completed RNA is used as template. In a coupled transcription-translation system, the ribosome binding site of gene X will already be loaded by ribosomes before the complementary part of the hairpin has been synthesized. In this case, initiation of gene X-protein synthesis will not be hindered because the secondary structure in the RNA is not formed. Similar hindrance of ribosome loading by intramolecular base-pairing has also been suggested in RNA phage translation[61], in λ *cro*-gene expression[62] and in the expression of gene 0.3 of phage T7[63].

In this respect it is worth mentioning that the synthesis of X-protein has not yet been observed either in *E.coli* minicells bearing M13 RF as a plasmid[64] or in M13 infected *E.coli* cells[59,65], despite the fact that in both cases the presence of a 14S RNA has clearly been demonstrated. If ribosome binding is hindered by the secondary structure mentioned, then a protection against RNA breakdown will not occur in the region between this structure and the next ribosome binding site. It might give a plausable explanation for the undetectable level of gene X-protein synthesis and the rapid processing of 14S RNA species in the infected cell[54-56].

In Fig 7b the two possible starting positions of 11S RNA which initiates at the weak promoter $G_{0.12}$, are indicated. Translation of 11S RNA results in the synthesis of gene V- and gene VIII-protein[6].

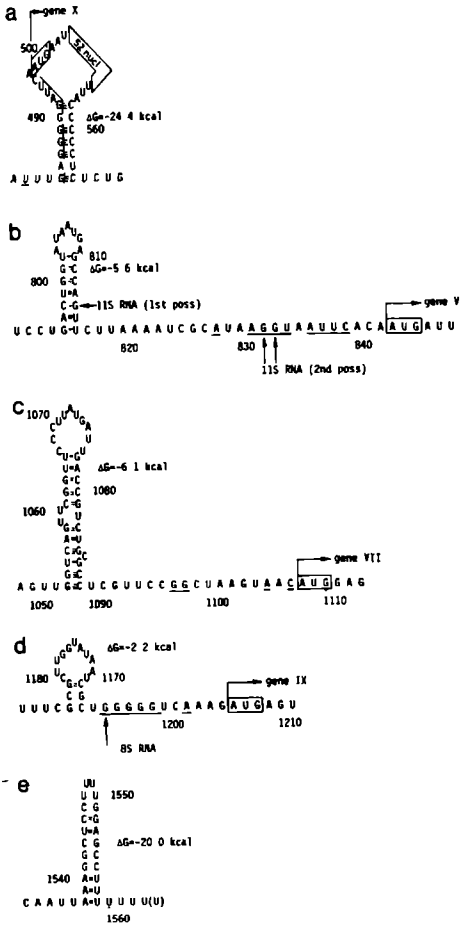


Fig 7 Possible secondary structures in the transcripts of various parts of the gene X-gene VIII region.

The regions in the RNA's preceeding the initiation codons of the genes X, V, VII and IX are shown in a, b, c and d, respectively. In also the sequence about 60 nucleotides distal to the initiation codon of gene X, which shows very strong complementarity with the ribosome binding site of gene X, is indicated. Figure 7e shows the 3'-OH terminal end of M13 8S RNA as deduced by Edens[45].

The bases in front of the boxed initiation codons which are capable of Shine-Dalgarno basepairing are underlined. The starting nucleotide of 8S RNA and the possible starting nucleotides of 11S RNA are indicated by arrows. Free energies (ΔG) have been calculated according to the rules of Tinoco *et al* [67].

As initiation of 11S RNA synthesis at positions 831 or 832 would partially destroy the ribosome binding site of gene V, we suppose that initiation is more likely to occur at position 814. Initiation at the latter position leaves the ribosome binding site of gene V intact. The biological significance of the small hairpin structure at nucleotide positions 797 to 805 remains unclear.

The undetectable level of gene VII-protein synthesis, *in vivo*[16,17] as well as *in vitro*[7,8] might be explained by the very low potential of Shine-Dalgarno basepairing for the ribosome binding site of gene VII (Table 1 and Fig 7c). Because gene VII is not equipped with a promoter, translation only occurs on RNA's which initiate at upstream promoters. The large hairpin structure in front of gene VII will probably not interfere with the attachment of ribosomes to the weak ribosome binding site as this structure will normally be melted out by ribosomes, translating the gene V message. However, if gene V synthesis is prematurely terminated, this hairpin might be formed and might influence the attachment of ribosomes to the ribosome binding site of gene VII. Whether this hypothesis is valid for an explanation of the observed polarity among genes V and VII[65,66] will be discussed in the next chapter.

Gene IX is preceeded by a region which shows rather good complementarity with the 3'-end of 16S ribosomal RNA (Table 1 and Fig 7d). Translation of 8S RNA, which initiates at position 1195 in the $G_{0.18}$ promoter, results in the synthesis of only the gene VIII-protein[6], although the whole coding information for the gene IX-product is included in this transcript (Fig 7d). We have, at present, no explanation for the absence of gene IX-protein. Perhaps the region which preceeds the initiation codon of gene IX is too short to allow firm binding of the ribosomes. On the other hand, transcripts which initiate at upstream promoters, also fail to direct the synthesis of gene IX-protein[6]. We have

looked for stable secondary structures locate immediately in front of gene IX, which might be of influence in ribosomal attachment at the ribosome binding site of gene IX. The only loop structure found is the one shown in Fig 7d. Its stability is, in our opinion, too low for negative influence on the ribosome attachment.

Gene VIII-protein is readily synthesized in coupled as well as uncoupled *in vitro* systems[4-10]. The properties of the ribosome binding site of gene VIII are in accordance with this observation: the ribosome binding site shows considerable complementarity with the 3'-OH terminal end of 16S-ribosomal RNA (Table 1), it is not blocked by intramolecular basepairing and the attachment of ribosomes is not hindered by neighbouring secondary structures.

1. Day, L.A. and Wiseman, R.L. (1978) A Comparison of DNA Packaging in the Virions of fd, Xf and Pf1. In: Single-Stranded DNA Phages, Denhardt, D.T., Ray, D.S. and Dressler, D., Eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp 605-625.
2. Simons, G.F.M., Konings, R.N.H. and Schoenmakers, J.G.G. (1979) FEBS Lett. *106*, 8-12.
3. Konings, R.N.H. and Schoenmakers, J.G.G. (1978) Transcription of the Filamentous Phage Genome. In: Single-Stranded DNA Phages; Denhardt, D.T., Ray, D.S. and Dressler, D., Eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp 507-530.
4. Edens, L., Konings, R.N.H. and Schoenmakers, J.G.G. (1975) Nucleic Acids Res. *2*, 1811-1820.
5. Edens, L., Wezenbeek van, P., Konings, R.N.H. and Schoenmakers, J.G.G. (1976) Eur. J. Biochem. *70*, 577-587.
6. Edens, L., Konings, R.N.H. and Schoenmakers, J.G.G. (1978) Virology *86*, 354-367.
7. Model, P. and Zinder, N.D. (1974) J. Mol. Biol. *83*, 231-251.
8. Konings, R.N.H., Hulsebos, T. and Hondel van den, C.A. (1975) J. Virol. *15*, 570-584.
9. Hondel van den, C.A., Konings, R.N.H. and Schoenmakers, J.G.G. (1975) Virology *67*, 487-497.
10. Vovis, G.F., Horiuchi, K. and Zinder, N.D. (1975) J. Virol. *16*, 674-684.
11. Asbeck, F., Beyreuther, K., Köhler, H., Wettstein von, G. and Braunitzer, G. (1969) Hoppe-Seyler's Z. Physiol. Chem. *350*, 1047-1066.
12. Snell, D.T. and Offord, R.E. (1972) Biochem. J. *127*, 167-178.

13. Nakashima, Y. and Konigsberg, W. (1974) J. Mol. Biol. *88*, 598-600.

14. Bailey, G.S., Gillett, D., Hill, D.F. and Petersen, G.B. (1977) J. Biol. Chem. *252*, 2218-2225.

15. Hulsebos, T. and Schoenmakers, J.G.G. (1978) Nucleic Acids Res. *5*, 4677-4698.

16. Henry, T.J. and Pratt, D. (1969) Proc. Natl. Acad. Sci. U.S.A. *62*, 800-807.

17. Lin, N.S.-C. and Pratt, D. (1974) Virology *61*, 334-342.

18. Hondel van den, C.A., Pennings, L. and Schoenmakers, J.G.G. (1976) Eur. J. Biochem. *68*, 55-70.

19. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. *74*, 560-564.

20. Hondel van den, C.A., Weyers, A., Konings, R.N.H. and Schoenmakers, J.G.G. (1975) Eur. J. Biochem. *53*, 559-567.

21. Berkner, K.L. and Folk, W.R. (1977) J. Biol. Chem. *252*, 3176-3184.

22. Staden, R. (1977) Nucleic Acids Res. *4*, 3037-3051.

23. Hondel van den, C.A. and Schoenmakers, J.G.G. (1975) Eur. J. Biochem. *53*, 547-558.

24. Cuypers, T., Ouderaa van der, F.J. and Jong de, W.W. (1974) Biochem. Biophys. Res. Commun. *59*, 557-563.

25. Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U.S.A. *71*, 1342-1346.

26. Horiuchi, K., Vovis, G.F. and Model, P. (1978) The Filamentous Phage Genome: Genes, Physical Structure and Protein Products. In: Single-Stranded DNA Phages, Denhardt, D.T., Ray, D.S. and Dressler, D., Eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp 113-137.

27. Weisbeek, P.J., Borrias, W.E., Langeveld, S.A., Baas, P.D. and Arkel van, G.A. (1977) Proc. Natl. Acad. Sci. U.S.A. *74*, 2504-2508.

28. Sanger, F., Coulson, A.R., Friedmann, T., Air, G.M., Barrel, B.G., Brown, N.L., Fiddes, J.C., Hutchison III, C.A., Ilocombe, P.M. and Smith, M. (1978) J. Mol. Biol. *125*, 225-246.

29. Meyer, T.F., Geider, K., Kurz, C. and Schaller, H. (1979) Nature *278*, 365-367.

30. Meyer, T.F. and Geider, K. (1979) J. Biol. Chem. *254*, 12636-12641.

31. Meyer, T.F. and Geider, K. (1979) J. Biol. Chem. *254*, 12642-12646.

32. Eisenberg, S., Griffith, J. and Kornberg, A. (1977) Proc. Natl. Acad. Sci. U.S.A. *74*, 3198-3202.

33. Langeveld, S.A., Mansfeld, van, A.D.M., Baas, P.D., Jansz, H.S., Arkel van, G.A. and Weisbeek, P.J. (1978) Nature *271*, 417-420.

34. Pieczenick, G., Model, P. and Robertson, H.D. (1974) J. Mol. Biol. *90*, 191-214.

35. Moses, P.B., Boeke, J.D., Horiuchi, K. and Zinder, N.D. (1980) Virology *104*, 267-278.

36. Chang, C.N., Blobel, G. and Model, P. (1978) Proc. Natl. Acad. Sci. U.S.A. *75*, 361-365.

37. Siebenlist, U., Simpson, R.B. and Gilbert, W. (1980) Cell *20*, 269-281.

38. Pribnow, D. (1975) Proc. Natl. Acad. Sci. U.S.A. *72*, 784-788.

39. Gilbert, W. (1976) In: RNA polymerase, Losick, R. and Champerlin, M., Eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp 193-205.

40. Scherer, G.E.F., Walkinshaw, M.D. and Arnott, S. (1978) Nucleic Acids Res. *5*, 3759-3773

41. Edens, L., Konings, R.N.H. and Schoenmakers, J.G.G. (1978) J. Virol. *28*, 835-842.

42. Sugimoto, K., Okamoto, T., Sugisaki, H. and Takanami, M. (1975) Nature *253*, 410-414.

43. Heyden, B., Nüsslein, C. and Schaller, H. (1975) Eur. J. Biochem. 55, 147-155.

44. Sugimoto, K., Sugisaki, H., Okamoto, T. and Takanami, M. (1977) J. Mol. Biol. *III*, 487-507.

45. Edens, L. (1978) Ph.D. Thesis, University of Nijmegen, The Netherlands.

46. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. *13*, 319-353.

47. Rivera, M.J., Smits, M.A., Quint, W., Schoenmakers, J.G.G. and Konings, R.N.H. (1978) Nucleic Acids Res. 5, 2895-2912.

48. Wu, A.M., Chapman, A.G., Platt, T., Guarente, L.P. and Beckwith, J., Cell *19*, 829-836.

49. Wezenbeek van, P. and Schoenmakers, J.G.G. (1979) Nucleic Acids Res. *6*, 2799-2818.

50. Beck, E., Sommer, R., Auerswald, E.A., Kurz, Ch., Zink, B., Osterburg, G., Schaller, H., Sugimoto, K., Sugisaki, H., Okamoto, T. and Takanami, M. (1978) Nucleic Acids Res. 5, 4495-4503.

51. Marvin, D.A. and Hohn, B. (1969) Bacteriol. Rev. *33*, 172-209.

52. Hondel van den, C.A. and Schoenmakers, J.G.G. (1976) J. Virology *18*, 1024-1039.

53. Wezenbeek van, P.M.G.F., Hulsebos, T.J.M. and Schoenmakers, J.G.G. (1980) Gene, 129-148.

54. Smits, M.A., Schoenmakers, J.G.G. and Konings, R.N.H. (1980) Eur. J. Biochem., in press.

55. Cashman, J.S. and Webster, R.E. (1979) Proc. Natl. Acad. Sci. U.S.A. *76*, 1169-1173.

56. Cashman, J.S., Webster, R.E. and Steege, D.A. (1980) J. Biol. Chem. *255*, 2554-2562.

57. Jacob, E., Jaenisch, R. and Hofschneider, P.H. (1973) Eur. J. Biochem. *32*, 432-443.

58. Lica, L. and Ray, D.S. (1976) J. Virol. *18*, 80-84.

59. La Farina, M. and Model, P. (1978) Virology *86*, 368-375.

60. Huang, C.-C. and Hearst, J.E. (1980) Anal. Biochem. *103*, 127-139.

61. Fiers, W. (1979) Structures and Function of RNA Bacteriophage. In: Comprehensive Virology, Fraenkel-Conrat, H. and Wagner, R., Eds., Plenum Publishing Corporation, N.Y., Vol. 13, pp 69-204.

62. Iserentant, D. and Fiers, W. (1980) Gene 9, 1-12.

63. Dunn, J.J. and Studier, F.W. (1979) Processing of Bacteriophage T7 RNA's by RNase III. In: From gene to protein: information transfer in normal and abnormal cells, Russell, T.R., Brew, K., Faber, H. and Schultz, J., Eds., Academic Press, Miami Winter Symposia, Vol. 16, pp 261-269.

64. Smits, M.A., Simons, G., Konings, R.N.H. and Schoenmakers, J.G.G. (1978) Biochim. Biophys. Acta *521*, 27-44.

65. Lyons, L.B. (1971) Ph.D. Thesis, The Rockefeller University, New York.

66. Lyons, L.B. and Zinder, N.D. (1972) Virology *49*, 45-60.

67. Tinoco, I. Jr., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature *246*, 40-41.

68. Jaenisch, R., Jacob, E. and Hofschneider, P. (1970) Nature *277*, 59-60.

CHAPTER VI

GENE V OF THE FILAMENTOUS BACTERIOPHAGES M13, f1 AND fd; LOCALIZATION OF AMBER MUTATIONS IN GENE V AND THEIR USE IN THE STUDY OF GENE V-GENE VII POLARITY

CHAPTER VI

GENE V OF THE FILAMENTOUS BACTERIOPHAGES M13, f1 AND fd; LOCALIZATION OF AMBER MUTATIONS IN GENE V AND THEIR USE IN THE STUDY OF GENE V-GENE VII POLARITY

ABSTRACT

The positions of the amber mutation in several amber5-mutants of the filamentous phages M13, f1 and fd have been deduced by nucleotide sequence analysis with the aid of the chain-termination method. Of the six amber5-mutants examined, four are mutated at position 999 (*i.e.* f1 R13, f1 R99, M13 *am*5-H1 and M13 *am*5-H3) and one is mutated at position 906 (*i.e.* fd 122). From the sequence analysis of the sixth amber 5-mutant, f1 R148, it appears that this amber mutant is not mutated in gene V, but in gene VII at position 1114.

In contrast to what has been observed with conditionally lethal mutants in other operons, the gene V-mutants, when tested in a complementation assay with well-defined amber7-mutants, did not reveal any gradient in polarity. A model which might explain the latter phenomenon is discussed.

INTRODUCTION

After infection of the *E.coli* cell, the single-stranded DNA of the filamentous phage is converted into a double-stranded replicative form (RF) by synthesis of a complementary strand[1]. The parental RF is then replicated by a process in which at first the viral DNA strand of RF is nicked by the phage-encoded gene II-protein[2]. The generated 3'-OH terminus serves as a primer for the synthesis of a new strand, whereby the resident viral strand is displaced. After circularization, the displac-

ed viral strand is converted into RF by synthesis of the complementary strand, etc. The switch-over from RF synthesis to progeny viral strand DNA synthesis takes place when sufficient amounts of the phage-encoded gene V-protein have accumulated as a result of expression of the synthesized RF molecules[3,4,5]. Gene V-protein binds to the displaced viral strands[6,7] and prevents, in this way, the conversion to RF molecules. The experimental data of some authors suggest that a second function of gene V-protein might be the repression of the synthesis of gene II-protein[8,9].

Gene V, which encodes this single-stranded DNA binding protein, has been located in a region of the genome, which is very frequently expressed, *in vivo* as well as *in vitro*[10-13]. Genetic studies of Lyons and Zinder[14,15] have shown that amber mutations in gene V are polar on the contiguous gene VII. Based on these observations, it was suggested that genes V and VII form an operon. However, the mechanism which underlies the polar effect of amber mutations in gene V on gene VII expression has not been resolved, yet.

A general characteristic of operons is that amber mutations which are located near the amino terminal end of a gene tend to be more polar on distal genes than mutations near the carboxy terminal end[16]. Such positional effects of amber5-mutants on the gene V-gene VII polarity have never been studied so far. Therefore, one way to unravel the problem of this polarity is to investigate in detail whether there are differences in polar effects of gene V-amber mutants on gene VII-expression and whether these effects are dependent on the positions of the amber mutations. To do this, it is necessary to study phages with the amber mutation in gene V in as much as possible positions. For this purpose, we have determined by nucleotide sequencing the positions of the amber mutations in several M13, fd and f1 amber5-mutants.

The amber5-mutants, selected in this way, were subsequently used in genetic complementation tests to establish their effect on gene VII-expression. Based on the results of these tests a model is discussed which might explain the observed polarity between gene V and gene VII.

MATERIALS AND METHODS

Bacteria and phages

The *E.coli* strains KA 798 (Su^-) and KA 805 (Su I) were kindly provided by Dr. B. Glickman, University of Leiden. The M13 amber mutants *am*5-H1 and *am*5-H3 were originally obtained from Dr. D. Pratt. The f1 amber mutants R13, R99 and R148 were provided by Dr. N.D. Zinder, fd 122 was a gift of Dr. H. Schaller. Reversion rates of these mutants ranged from 10^{-6} to 10^{-4}.

Materials and enzymes

Restriction enzyme *Hap* II was prepared as described previously[17]. Restriction enzyme *Taq* I was obtained from the Microbiol. Res. Establishment, Porton Down and *Hha* I from New England Biolabs. *E.coli* DNA polymerase I (Klenow fragment) was purchased from Boehringer, Mannheim. (α-^{32}P)-dATP (spec.act. 300-400 Ci/mmol) was supplied by Amersham, England.

Viral DNA and restriction fragments

Viral DNA was extracted from purified mutant phage, grown on *E.coli* KA 805 (Su I), as described by van den Hondel *et al* [17,18]. Wild type M13 RF was isolated, purified and digested with restriction enzymes according to previously published procedures[17,18].

DNA sequencing procedure

The nucleotide sequence of amber mutant DNA was determined with the chain termination procedure of Sanger *et al* [19]. One pmol of wild-type M13 restriction fragment and 0.2 pmol of amber mutant viral DNA were dissolved in 10 µl of H-buffer, containing 6.6 mM Tris-HCl, pH 7.4, 6.6 mM $MgCl_2$, 50 mM NaCl and 1 mM DTT. The mixture was sealed in a capillary tube and heated for 3 minutes at 100°C, followed by 15 minutes at 67°C. Two µl samples of the hybrids formed were added to 2 µl of each of four incubation mixtures. Each mixture contained 1.5 x H-buffer, 1 µCi (α-^{32}P)-dATP (spec.act. 300-400 Ci/mmol) and the following triphosphates:

T: 0.13 mM dCTP, 0.13 mM dGTP, 0.08 mM dTTP, 1 mM ddTTP

C: 0.13 mM dTTP, 0.13 mM dGTP, 0.08 mM dCTP, 0.35 mM ddCTP

G: 0.13 mM dTTP, 0.13 mM dCTP, 0.08 mM dGTP, 0.35 mM ddGTP

A: 0.13 mM dTTP, 0.13 mM dCTP, 0.13 mM dGTP, 0.75 mM ddATP

DNA synthesis was started by the addition of 1 µl (0.25 units) of DNA polymerase I (Klenow fragment) to each reaction mixture. Incubation was at room temperature for 15 min, followed by a chase with 1 µl of dATP (0.5 mM), for another 15 min. Large primer fragments were subsequently split off by incubation (5 min at 37°C) with 1 µl (1-2 units) of restriction enzyme. In certain cases the splitting off was omitted in order to avoid a double banding pattern, generated by a second restriction site in the direct neighbourhood. Twenty µl of dye mixture (0.3% xylene cyanol, 0.3% bromophenol blue, 10 mM EDTA in deionized formamide) was added and the reaction mixtures were denatured for 3 min at 100°C. Ten µl of each mixture was applied to a 10% polyacrylamide slab gel (acrylamide/bisacrylamide ratio 20:1, prepared in 50 mM Tris-borate, pH 8.3, 1 mM EDTA and 7 M urea). Electrophoresis was carried

out at 1250 V for various times, depending on the positions of the regions to be sequenced.

Genetic complementation

Genetic complementation tests were carried out as described by Pratt *et al* [20]. Non-permissive *E.coli* cells (KA 798) were grown in Trypton medium[21] at 37°C to a density of $2x10^8$ cells/ml. A sample was taken for assay of bacterial colony formers. The culture was shifted to 32°C and infected with the appropriate pair of amber mutants (multiplicity of infection was 50 for each of the phages). After an absorption period of 10 min, the infected culture was diluted and plated for plaque formation with KA 798 as indicator bacteria. The plates were incubated overnight at 32°C. Complementation values, defined as the ratio of plaque formers to total cells plated, were calculated for the various double infections. Each value was corrected for wild-type revertants, found in single infections of both phages. The criteria for complementation, as defined by Pratt *et al* [20] have been applied. Complementation was strong in case the complementation values were greater than 0.15, weak when the values were between 0.01 and 0.15 and negative (no complementation) when the values were less than 0.01.

RESULTS

Selection of amber5-mutants for sequence determination

In order to study the positional effects of amber mutations in gene V on gene VII-expression, it is necessary to have amber mutations in as much as possible positions within gene V. Because of this reason we examined not only amber mutants in gene V of M13, but also amber mutants in gene V of the closely related phages f1 and fd. The positions of the

amber mutations were deduced by nucleotide sequence analysis, using the chain-termination method of Sanger *et al* [19]. The following phages were examined: the M13 amber5-mutants M13 *am*5-H1 and M13 *am*5-H3, the f1 amber5-mutants f1 R13, f1 R99 and f1 R148 and the fd amber5-mutants fd 122. The regions to be sequenced in the DNA of these phages were selected on the basis of the following considerations:

i The complete nucleotide sequence of the gene V region in wild-type M13, f1 and fd has been deduced now (Chapter V of this thesis[22,23]. Although there are differences in the nucleotide sequence of gene V in the three phage DNA's, these are restricted to the third base of the coding triplets which in no case lead to amino acid replacements in the resulting gene V-protein. The nucleotide sequence of gene V in wild-type M13, f1 and fd and the amino acid sequence of gene V-protein are shown in Fig 1.

ii All the amber mutations studied were induced by hydroxylamine or nitrous acid treatment of wild-type phages. Both agents induce preferentially C→T transitions[24]. The amber codons will therefore most probably be generated by the conversion of CAG to TAG. From Fig 1 it is clear that there are four potential candidates for conversion to amber mutations, namely the CAG-codons at positions 906, 933, 999 and 1056.

iii The regions where the amber mutations could be located are further restricted by the information obtained from previous marker rescue experiments[18,25-27].

This methodology revealed that the amber mutation in fd 122 was rescued by the fd-analogue of the M13 restriction fragment *Hap* II-C[25]. The location of this fragment is shown in fig 2. From the positions of gene V and fragment *Hap* II-C within the nucleotide sequence, shown in fig 1, it can be concluded that this amber mutation is most probably located

at position 906 or 933 *i.e.* in the N-terminal part of gene V.

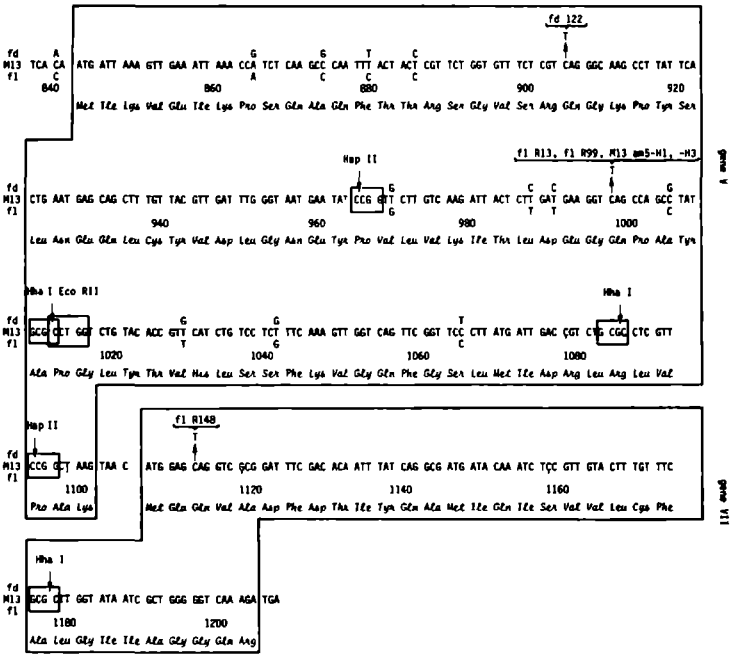


Fig 1 Positions of the amber mutation in fd 122, f1 R13, f1 R99, f1 R148, M13 am5-H1 and M13 am5-H3.
The figure shows the DNA sequence of the gene V-gene VII region in wild-type M13 DNA and the sequence differences which occur in wild-type fd and f1 DNA. The amino acid sequence of wild-type M13 gene V-protein is also shown. The wild-type f1 and fd gene V-protein are identical to the wild-type M13 gene V-protein. The recognition and cleavage sites for the restriction enzymes *Hap* II, *Hha* I and *Eco* RII are indicated by boxes and downwards pointing arrows, respectively. The C→T transitions which occur in fd 122 at position 906, in f1 R13, f1 R99, M13 *am*5-H1 and M13 *am*5-H3 at position 999 and in f1 R148 at position 1114 are indicated by upwards pointing arrows.

The M13 amber mutant *am*5-H3 was rescued by M13 fragment *Hap* II-I_1[18]. This fragment is located in the C-terminal part of gene V, between positions 966 and 1095 (see Fig 1 and 2). This would mean that the amber mutation in M13 *am*5-H3 can only be located at position 999 or 1056.

Since there is no marker rescue information available for M13 *am*5-H1,

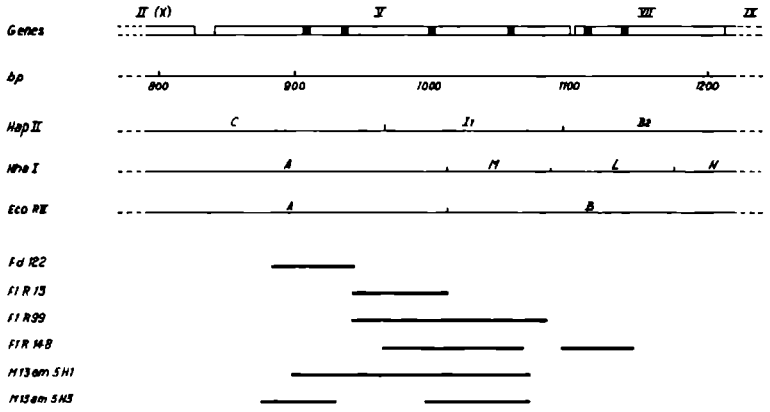


Fig 2 Genetic map and physical maps of a segment of bacteriophage M13 DNA. The Roman numerals refer to the genes. The capital numerals refer to the DNA fragments obtained after digestion of this part of the M13 genome with various restriction enzymes. The positions of the CAG codons in the genes V and VII of wild-type phages are indicated by black boxes. The horizontal bars indicate the regions in the various amber mutant DNA's of which the sequence has been deduced.

the amber mutation in this phage can occupy, in theory, each of the four probable positions in gene V.

The f1 amber5-mutants R13 and R99 were rescued by the f1-analogue of M13 fragment *Hap* II-I_1[27], but also by fragment Eco RII-A[26]. Therefore both mutations will be located in the overlapping region of these fragments, between nucleotides 966 and 1014 (see Fig 1 and 2). The most probable location for the amber mutations will be the CAG-codon at position 999.

A rather puzzling situation is noted for the amber5-mutant f1 R148. This mutant was rescued by the f1-analogue of M13 fragment *Hap* II-B_2 and by f1 fragment *Eco* RII-B[26,27]. If f1 R148 contains an amber mutation in gene V, then the mutation must be located between the *Hap* II-cleavage site at position 996 and the C-terminal end of gene V at pos-

ition 1104. However, no CAG-codon is present in this part of gene V. It is possible, therefore, that this amber mutation is created by an unexpected mutagenic event. Of course, the same holds true for the other amber phages, discussed above. Alternatively, f1 R148 is not mutated in gene V, but in another gene, contained within fragment *Hap* II-B_2.

Determination of nucleotide changes in amber5-mutants

The nucleotide sequence of the regions which most probably contain the amber mutation was deduced with the chain-termination method of Sanger *et al* [19]. For this purpose, the minus-strands of wild-type M13 restriction fragments were hybridized to the circular, single-stranded DNA of the mutant phages and used as primers for limited DNA synthesis in the presence of various dideoxy-triphosphates and (α-^{32}P)-dATP.

After DNA synthesis and an appropriate chase with unlabelled dATP, the restriction fragments were split off from the newly synthesized DNA. The reaction products were denatured and then applied to the DNA sequencing gels. The gels were run at constant voltages for various times, depending on the positions of the regions to be sequenced. The choice of these regions was determined by the localization predictions, made in the previous section. In case of M13 *am*5-H1 we have sequenced all possible amber locations because of the absence of marker rescue information for this phage.

As the most probable location of the amber mutation in f1 R99 was position 999, the restriction fragment *Hha* I-M (Fig 2) was selected as primer. After removal of the primer, the newly synthesized DNA gives rise to the sequencing gel, shown in Fig 3A. For comparison, the corresponding region in wild-type M13 DNA is shown in Fig 3B. The only difference between the two sequences in this region is a G→A change at

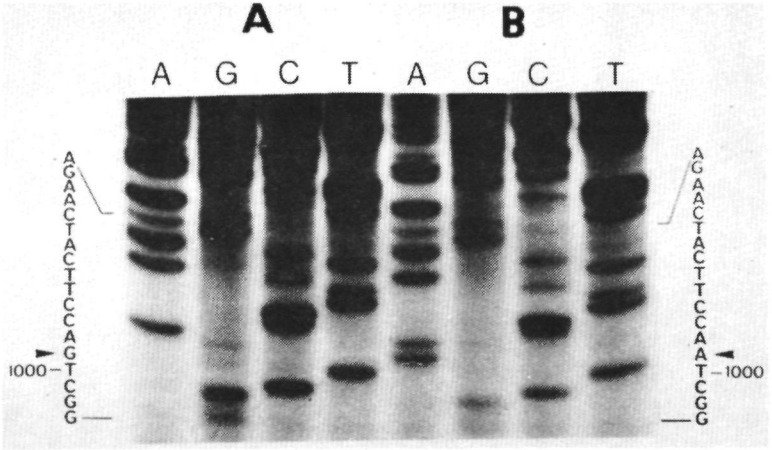


Fig 3 Autoradiograph of a DNA sequencing gel from the sequence determination using fragment *Hha* I-M as primer and f1 R99 viral DNA (A) or wild M13 viral DNA (B) as template. The top and bottom part of the gel are not shown.

position 999, which replaces the wild-type sequence 5'-TGGCTGACC-3' in 5'-TGGCTAACC-3'. As the viral strand is used as template, the sequence represents the complementary strand sequence. Consequently in the viral strand the wild-type sequence 5'-GGTCAGCCA-3' is replaced by 5'-GGTTAGCCA-3' in f1 R99. This result confirmed our prediction that, in phage f1 R99, CAG at position 999 is converted into the amber codon TAG.

To trace the amber mutation in f1 R13 we sequence the same DNA region as described above, but in this experiment the primer fragment *Hha* I-M was not split off after the DNA synthesis reaction. As is shown in Fig 4 the amber mutation in f1 R13 is also present at position 999. Because the amber mutation in M13 *am*5-H3 is most probably located at positions 999 or 1056 within fragment *Hap* II-I_1, we sequenced this region by using fragment *Hap* II-B_2 as primer (see Fig 2). For the local-

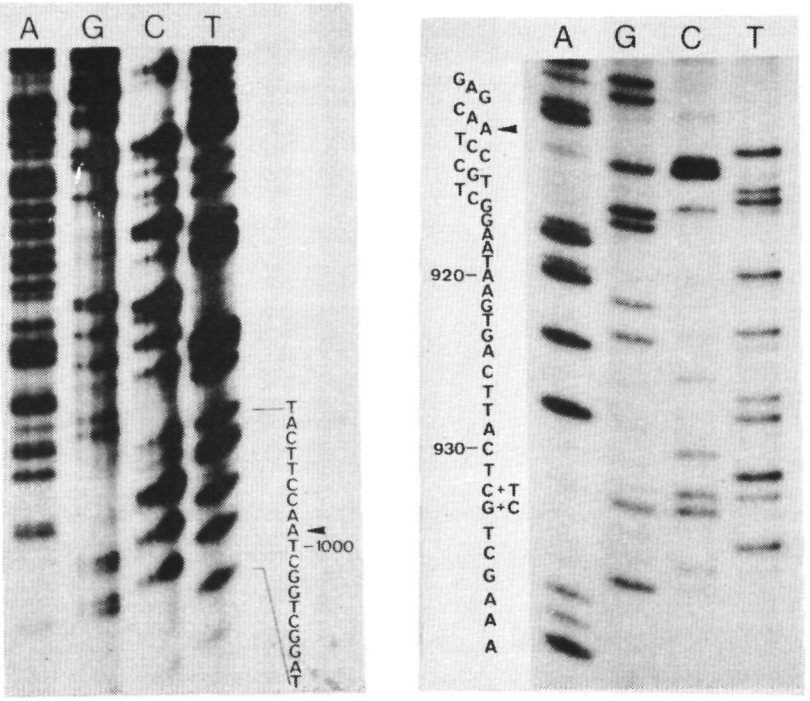


Fig 4 Autoradiograph of a DNA sequencing gel from the sequence determination using fragment *Hha* I-M as primer and f1 R13 viral DNA as template.
In this case the primer fragment was not removed after the DNA synthesis reaction.
The top and bottom part of the gel are not shown.

Fig 5 Autoradiograph of a DNA sequencing gel from the sequence determination using fragment *Hap* II-I_1 as primer and fd 122 viral DNA as template.
The top part of the gel is not shown.

ization of the amber mutation in M13 *am*5-H1, for which no marker rescue data were available, all probable amber locations were sequenced. For this purpose the restriction fragments *Hap* II-B_2, *Hha* I-M and *Hap* II-I_1 were used as primer in the limited DNA syntheses reaction. The result-

ing sequencing gels revealed that, in analogy with the situation found for f1 R13 and f1 R99, the amber mutation in M13 *am*5-H1 as well as M13 *am*5-H3 was located at position 999 (data not shown).

The amber mutation in gene V of phage fd 122 was predicted to be located at position 906 or 933. In order to sequence both possible positions, fragment *Hap* II-I_1 was chosen as primer (Fig 2). The only difference observed with respect to the wild-type fd sequence[22] is a G→A change at position 906, as is shown in Fig 5. The corresponding C→T change in the viral strand replaces the wild-type sequence 5'-CGTCAGGGC-3' by 5'-CGTTAGGGC-3' in fd 122. An anomalous read-out can be observed at positions 932 and 933. The wild-type C at position 932 can also read T. Similarly, the wild-type G at position 933 can also read C. The anomaly was also observed in the DNA of M13 *am*5-H1, the other amber mutant of which the DNA was sequenced in this region and in the DNA of a temperature sensitive M13 mutant. Further studies are needed to find out whether this observation has any biological significance or that this anomaly is merely a sequencing artefact.

In the previous section, it was suggested that the amber5-mutant f1 R148 contains an amber mutation between the *Hap* II-cleavage site at position 1095 and the C-terminal end of gene V at position 1103 (Fig 1). Alternatively, the amber mutation is located in another gene, contained within fragment *Hap* II-B_2. As, by genetic criteria, amber5- and amber7-mutants behave grossly identical, the most likely alternative is an amber mutation in gene VII. To verify both possibilities, we sequenced the C-terminal part of gene V and the most probable amber locations in gene VII. The latter are the CAG-codons at position 1114 and 1141 (mutated into amber codons in M13 *am*7-H2 and M13 *am*7-H3, respectively[28]. Fragment *Hha* I-H was used as primer. To limit the DNA chain extensions, we doubled the dideoxytriphosphate concentration in the various react-

ion mixtures. Otherwise, the subsequent digestion with restriction enzyme *Hha* I would create an additional banding pattern by nicking of the enzyme at the *Hha* I-L/*Hha* I-M junction (Fig 2). The resulting sequencing gel (Fig 6) shows quite unexpectedly that guanine at position 1114 in the wild-type sequence[23] is replaced by adenine in f1 R148. The wild-type sequence 5'-GACCTGCTC-3' is changed into 5'-GACCTACTC-3'. The complementary region in the viral DNA strand of f1 R148 thus contains the amber codon 5'-GAGTAGGTC-3' at this position. The second possible amber location in gene VII, at position 1141, remains unchanged. Also no 5'-CTA-3' sequence, complementary to an amber codon, can be detected in the upper part of the sequencing gel which represents the nucleotide sequence of the C-terminal part of gene V. From this result and the marker rescue data we infer that f1 R148 is not an amber5-mutant but a gene VII-mutant which contains the amber mutation at position 1114.

In order to detect additional nucleotide sequence deviations in the DNA of the amber mutant phages, we occasionally sequenced other parts of gene V in the respective phages. These regions are indicated by lines in Fig 2. The sequencing data of these regions revealed that, apart from the amber mutations and the sequence anomaly described above, no other sequence alterations could be detected as compared to the wild-type M13, f1 and fd sequences.

Genetic complementation

There are four well-spread CAG-codons in the gene V sequence of the filamentous phages, which could in theory be converted into amber codons by the mutagenic agents used. However, of the six amber5-mutants investigated, four were mutated at position 999 (f1 R13, f1 R99, M13 *am*5-H1 and M13 *am*5-H3), only one at position 906 (fd 122) and one proved to be located

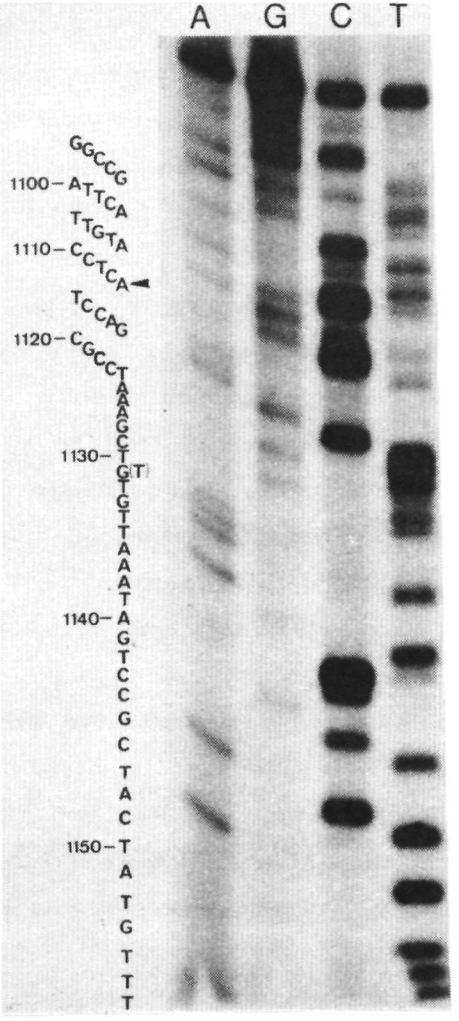


Fig 6 Autoradiograph of a DNA sequencing gel from the sequence determination using fragment *Hha* I-H and f1 R148 viral DNA as template. The top part of the gel is not shown.

outside gene V, namely at position 1114 in gene VII (f1 R148). No mutants were found with an CAG→TAG change at position 933 or 1056. The

absence of amber mutations in the latter positions limits the possibility to investigate the positional effects of amber mutations in gene V on gene VII-expression. Nevertheless, we have performed experiments with fd 122 and two representatives of the second amber location, f1 R13 and M13 *am*5-H1. The effects of the amber5-mutations on gene VII-expression cannot be measured directly, because gene VII-protein synthesis, *in vivo*[10,11] as well as *in vitro*[12,13] is too low to be detected. However, gene VII-expression *in vivo* can be detected by a complementation assay, devised by Pratt *et al* [20]. This method is based on the general principle that mixed infection of non-permissive *E.coli* cells with two conditional lethal mutants will result in progeny phage production, when such mutations are located in different genes. No progeny will be formed when both mutations are located in the same gene, because complementation of the defective gene-products will not occur. In a similar way polar mutants will not or only weakly complement mutants with a mutation in the distal genes of an operon. Gene VII-expression by a particular amber5-mutant can thus be measured by double infection of non-permissive *E.coli* cells by the amber5-mutant and an amber mutant, defective in gene VII and subsequent determination of progeny virus production. As non-permissive host *E.coli* KA 798 was used and double infection was carried out with the various amber5-mutants and one of the well-defined amber7-mutants, M13 *am*7-H2 and M13 *am*7-H3. The progeny virus production was determined and the complementation value, defined as the ratio of plaque formers to total cells plated, was calculated as described in Materials and Methods. In control experiments the complementation value for intragenic crosses proved to be 0 (data not shown). We also performed double infections with the individual amber5- and amber7-mutants (including f1 R148) and M13 *am*8-H1. The results obtained are shown in Table 1.

Table 1 COMPLEMENTATION BEHAVIOUR OF AMBER5- AND AMBER7-MUTANTS

amber mutant pair			complementation value[a]
fd 122	+	M13 *am*8-H1	0.25
fd 122	+	M13 *am*7-H2	0.001
fd 122	+	M13 *am*7-H3	0.002
M13 *am*5-H1	+	M13 *am*8-H1	0.23
M13 *am*5-H1	+	M13 *am*7-H2	0
M13 *am*5-H1	+	M13 *am*7-H3	0
f1 R13	+	M13 *am*8-H1	0.23
f1 R13	+	M13 *am*7-H2	0
f1 R13	+	M13 *am*7-H3	0
f1 R148	+	M13 *am*8-H1	0.33
f1 R148	+	M13 *am*7-H3	0
M13 *am*7-H2	+	M13 *am*8-H1	0.06
M13 *am*7-H3	+	M13 *am*8-H1	0.37

[a]Ratio of plaque formers to total cells plated, corrected for wild-type revertants, found in single infections of both phages[20]. Complementation is classified as strong when the ratio is greater than 0.15, weak when the ratio is between 0.01 and 0.15 and negative (no complementation) when the ratio is less than 0.01.

As expected, in the tests were M13 *am*8-H1 was the second amber mutant, high complementation values were obtained. These values ranged from 0.23 to 0.37. The only exception is the mixed infection with M13 *am*7-H2 and M13 *am*8-H1, where weak complementation was observed. Exactly the same results, including the weak complementation for a mixed M13 *am*8-H1-M13 *am*7-mutant infection were made by Pratt *et al* [29]. In contrast, mixed infections with the individual amber5-mutants and one of the amber7-mutants never resulted in a significant complementation value. We conclude from these tests that there is no complementarity between the amber5- and amber7-mutants and hence there is, at least for the two positions examined, no positional effect of the amber5-mutants on gene VII

expression.

It is worth noting that the complementation between f1 R148 and M13 *am*8-H1 was always found to be 'strong'. This is in marked contrast to the 'weak' complementation between M13 *am*7-H2 and M13 *am*8-H1, mentioned above, although both mutants have their amber mutation at the same position in gene VII. Also, the complementation between M13 *am*7-H3 and M13 *am*8-H1 is strong. We consider the most probable explanation for this 'weak' complementation between M13 *am*8-H1 and M13 *am*7-H2 the presence of an additional mutation in another gene. We previously showed that M13 *am*7-H2 not only contains an amber mutation in gene VII but also a missence mutation in gene IX[28].

As a result, the arginine residue at position 18 in wild-type gene IX-protein is replaced by cysteine in the gene IX-protein, encoded by M13 *am*7-H2. This second base substitution is absent in gene IX of M13 *am*7-H3[28] and probably also in gene IX of f1 R148. The altered gene IX-protein, encoded by M13 *am*7-H2 might negatively influence the phage production in a mixed infection with M13 *am*8-H.

DISCUSSION

For many years, the marker rescue assay was the most powerful method for the localization of conditional lethal mutants on the filamentous phage genome[17,18,25-27]. However, the accuracy of this method is restricted by a lower limit in the lengths of the restriction fragments that can successfully be used[30]. Moreover, generation of small DNA fragments requires the establishment of detailed restriction maps.

As rapid methods for DNA sequencing have recently been developed[19,31,32], it became possible to determine the positions of amber mutations at the nucleotide level. In case of the filamentous phages, the chain-

termination method of Sanger *et al* [19] is extremely suitable for this purpose, because it requires only small amounts of single-stranded DNA which can easily be prepared from the supernatant of a bacterial culture, infected with the filamentous phage. This is in contrast to the chemical degradation method of Maxam and Gilbert[32], which requires the laborious preparation of the individual amber mutant replicative form DNA's and subsequent preparation of restriction fragments.

We have successfully applied the chain-termination method in the localization of the positions of five amber5-mutations. The regions to be sequenced were dictated by the available marker rescue data and the known nucleotide sequences of the gene V-regions in wild-type M13-, f1- and fd-DNA. The accuracy of the method is exemplified by f1 R148 and fd 122 which were originally classified, by genetic criteria, as an amber mutant in gene V[15] and an amber mutant in gene V or gene VII (D. Marvin, as quoted in ref 25), respectively. The marker rescue information for fd 122[25] already suggested that fd 122 is a gene V rather than a gene VII-mutant. This is confirmed by the sequencing gel shown in Fig 5, which clearly demonstrates that fd 122 contains the amber mutation at position 906 in gene V. The marker rescue assays performed with f1 R148[26,27] did not allow a definite conclusion to be drawn on its precise genetic location. However, the sequence determination (Fig 6) clearly demonstrates that f1 R148 is not an amber mutant in gene V, but in gene VII at position 1114.

The sequencing studies, carried out at the five amber5-mutants, revealed a preference of the amber mutations for certain positions in the gene V-sequence. This is rather unexpected, since, in theory, the CAG-codons at positions 906, 933, 999 and 1056 have an equal chance to become mutated by the chemical agents that were applied. However, only the CAG-codon at position 906 (in fd 122) and the CAG-codon at position

999 (in f1 R13, f1 R99, M13 *am*5-H1 and M13 *am*5-H3) were actually found to be replaced by the amber codon.

What causes this strong preference is by no means clear. The nucleotide sequence of gene V does not contain regions of strong secondary structure, which could otherwise hamper the access of the mutagenic agent. On the other hand we do not know whether the amber mutants, made available to us, although isolated independently, do represent a true selection of the amber5-mutants that can really exist. Until proven otherwise, we favor the idea that the isolation of amber5-mutants, after random mutagenization of wild-type phages and subsequent growth on a serine-inserting *E.coli* suppressor strain, does not permit the isolation of phages with an amber mutation at positions 933 or 1056. Probably, replacement of the original glutamine residue by serine at the corresponding positions 51 or 72 in the gene V-protein, results in the synthesis of an inactive molecule and, hence, no viable phages will be produced or selected.

The availability of only two locations in the nucleotide sequence of gene V, where amber codons have been generated, limits a detailed investigation of the positional effects of amber5-mutations on gene VII-expression. The available amber5-mutants, however, do not show any complementation at all with amber mutants in gene VII (Table 1). Maybe, the assay is not sensitive enough to detect (differences in) very small complementation values. On the other hand, Pratt *et al* [20] using the same methodology found clear differences in complementation between various amber3-mutants and an amber mutation in gene VI.

In terms of polarity, both amber5-mutants exert a very strong polar effect on gene VII-expression, irrespective of the positions of these mutants withing gene V.

We have as yet no explanation for this peculiar phenomenon, but as a working hypothesis we favor the following model. As gene VII is not equipped with a promoter, gene VII-protein can only be synthesized on polycistronic mRNA's which also contain the gene V-message. As discussed in Chapter V these RNA's contain a region in the C-terminal part of gene V which has the potential to form a rather stable hairpin structure (see Fig 7). Under wild-type conditions ribosomes, translating the gene V-message, will prevent the formation of the hairpin. After reaching the end of gene V, the ribosomes will dissociate and diffuse into the cytoplasm. Attachment of the 30S ribosomal subunit and subsequent initiation of gene VII-protein synthesis will only infrequently occur, as the ribosome binding site sequence of gene VII is very weak[28].

Introduction of an amber mutation in gene V causes premature dissociation and release of the ribosomes at the amber codon. No ribosomes will reach the end of gene V and consequently the hairpin can be formed now. This structure will prevent by steric hindrance, the attachment of a 30S ribosomal subunit to the initiation codon of gene VII, which results in a complete stop of the already low level of gene VII-protein synthesis. A similar inhibitory function has been proposed by Cannistraro and Kennell[33] for a hairpin structure in front of the ribosome binding site of the β-galactosidase gene. The model might explain the total lack of complementation between the amber5-mutants and the amber mutants in gene VII (Table 1), as both types of amber5-mutants give rise to the formation of the hairpin structure at the end of gene V (see Fig 7). An essential feature of this model is its explanation of polarity at the level of translation. This is in contrast to the model proposed by Adhya and Gottesman[34] who explain the polarity observed in several *E.coli* operons at the transcriptional level.

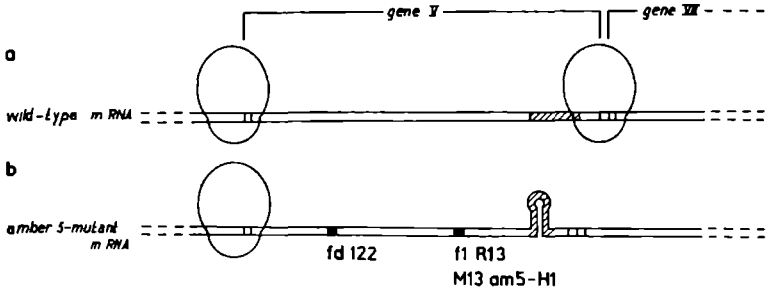


Fig 7 A model for gene V-gene VII polarity.
Ribosomes translate the gene V- and gene VII-message of polycistronic mRNA's transcribed from wild-type RF DNA (a). Initiation of translation of the gene VII-message will only infrequently occur, due to the weak ribosome binding site of this gene. Ribosomes, translating the gene V message prevent formation of a potential hairpin structure (indicated by the hatched region) in the C-terminal part of gene V. Premature termination of translation of the gene V-message at the amber codons of the amber5-mutant mRNA's (indicated by black boxes) allows the formation of the hairpin structure (b). This structure completely prevents, by steric hindrance, attachment of the 30S ribosomal subunit to the ribosome binding site of gene VII and no gene VII-protein synthesis will occur.

In the latter model polarity is caused by *rho*-induced termination of transcription at a site distal to the amber codon. *Rho* can interact with the nascent RNA at such a site and induce termination of transcription, because the RNA is no longer protected by ribosomes against '*rho*-attack'. In case of the gene V-gene VII polarity this would mean that a *rho*-dependent termination signal should be present in the C-terminal part of gene V. However, *rho*-induced termination seems to be a very unlikely explanation for gene V-gene VII polarity, as recent experiments have clearly demonstrated that the polycistronic RNA's which carry the gene V- and gene VII-message are still present in an intact form in non-permessive cells, infected with amber mutants in gene V

(Smits and Konings, personal communication). The latter observation also excludes degradation of uncovered RNA^{35} as a model for gene V-gene VII polarity.

1. Geider, K. and Kornberg, A. (1974) J. Biol. Chem. *249*, 3999-4005.
2. Fidanián, H.M. and Ray, D.S. (1972) J. Mol. Biol. *72*, 51-63.
3. Salstrom, J.S. and Pratt, D. (1971) J. Mol. Biol. *61*, 489-501.
4. Mazur, B.J. and Model, P. (1973) J. Mol. Biol. *78*, 285-300.
5. Mazur, B.J. and Zinder, N.D. (1975) Virology *69*, 490-502.
6. Oey, J.L. and Knippers, R. (1972) J. Mol. Biol. *68*, 125-138.
7. Alberts, B.M., Frey, L. and Delius, H. (1972) J. Mol. Biol. *68*, 139-152.
8. Horiuchi, K., Vovis, G.F. and Model, P. (1978) The Filamentous Phage Genome: Genes, Physical Structure and Protein Products. In: Single-Stranded DNA Phages, Denhardt, D.T., Ray, D.S. and Dressler, D., Eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp 113-137.
9. Webster, R.E. and Rementer, M. (1980) J. Mol. Biol. *139*, 393-405.
10. Henry, T.J. and Pratt, D. (1969) Proc. Natl. Acad. Sci. U.S.A. *62*, 800-807.
11. Lin, N.S.-C. and Pratt, D. (1974) Virology *61*, 334-342.
12. Model P. and Zinder, N.D. (1974) J. Mol. Biol. *83*, 231-251.
13. Konings, R.N.H., Hulsebos, T. and Hondel van den, C.A. (1975) J. Virol. *15*, 570-584.
14. Lyons, L.B. (1971) Ph.D. Thesis, The Rockefeller University, New York.
15. Lyons, L.B. and Zinder, N.D. (1972) Virology *49*, 45-60.
16. Newton, W.A., Beckwith, J.R., Zipser, D. and Brenner, S. (1965) J. Mol. Biol. *14*, 290-296.
17. Hondel van den, C.A., Pennings, L. and Schoenmakers, J.G.G. (1976) Eur. J. Biochem. *68*, 55-70.

18. Hondel van den, C.A., Weyers, A., Konings, R.N.H. and Schoenmakers, J.G.G. (1975) Eur. J. Biochem. *53*, 559-567.

19. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. *74*, 5463-5467.

20. Pratt, D., Tzagaloff, H. and Erdahl, W.S. (1966) Virology *30*, 397-410.

21. Marvin, D.A. and Hohn, B. (1969) Bacteriol. Rev. *33*, 172-209.

22. Beck, E., Sommer, R., Auerswald, E.A., Kurt, Ch., Zink, B., Osterburg, G., Schaller, H., Sugimoto, K., Sugisaki, H., Okamoto, T. and Takanami, M. (1978) Nucleic Acids Res. *5*, 4495-4503.

23. Hill, D.F. and Petersen, G.B. (1980) J. Virol. *34*, 40-50.

24. Shugar, D., Huber, C.P. and Birnbaum, G.I. (1976) Biochim. Biophys. Acta *447*, 274-284.

25. Seeburg, P.H. and Schaller, H. (1975) J. Mol. Biol. *92*, 261-277.

26. Vovis, G.F., Horiuchi, K. and Zinder, N.D. (1975) J. Virol. *16*, 674-684.

27. Ravetch, J., Horiuchi, K. and Model, P. (1977) Virology *81*, 341-351.

28. Hulsebos, T. and Schoenmakers, J.G.G. (1978) Nucleic Acids Res. *5*, 4677-4698.

29. Pratt, D., Tzagoloff, H. and Beaudoin, J. (1969) Virology *39*, 42-53.

30. Hutchison, C.A., III and Edgell, M.H. (1971) J. Virol. *8*, 181-189.

31. Sanger, F. and Coulson, A.R. (1975) J. Mol. Biol. *94*, 441-448.

32. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. *74*, 560-564.

33. Cannistraro, V.J. and Kennell, D. (1979) Nature *277*, 407-409.

34. Adhya, S. and Gottesman, M. (1978) Ann. Rev. Biochem. *47*, 967-996.

35. Contesse, G., Crepin, M. and Gros, F. (1970) Transcription of the lactose operon in *E.coli*. In: The Lactose Operon, Beckwith, J.R. and Zipser, D., Eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp 111-142.

CHAPTER VII

THE GENERATION OF MODIFIED GENE V-PROTEINS AND THEIR USE IN THE STUDY OF DNA - GENE V-PROTEIN INTERACTION

The DNA - Gene V-protein binding experiments, described in this study, were performed in collaboration with N.C.M. Alma-Zeestraten and B.J.M. Harmsen of the Department of Biophysical Chemistry, University of Nijmegen.

CHAPTER VII

THE GENERATION OF MODIFIED GENE V-PROTEINS AND THEIR USE IN THE STUDY OF DNA - GENE V-PROTEIN INTERACTION

ABSTRACT

The glutamine residues at positions 22 and 53 in the amino acid sequence of gene V-protein have been substituted by serine and tyrosine residues by infection of *E.coli* KA 805 (Su I) and *E.coli* KA 807 (Su III) cells with the amber5-mutant phages fd 122 and f1 R99, respectively. The intracellular content of each specifically modified gene V-protein has been determined by immunoprecipitation. The various modified gene V-proteins have been isolated by DNA-agarose chromatography and further purified by Sephadex G-75 gel filtration. The purified proteins are characterized by their ability to bind to single-stranded DNA. The binding studies reveal that substitution of the glutamine residue at position 22 by serine does not significantly alter the binding properties of the gene V-protein. Substitution of the glutamine at position 53 by serine results in a gene V-protein which binding is characterized by an n-value, considerably lower than the n-value of the binding of wild-type gene V-protein.

INTRODUCTION

Gene V-protein, a DNA-binding protein, plays an essential role in the synthesis of progeny viral single strands during infection of *E.coli* by the filamentous phages. When, as a result of a phage-directed protein synthesis the concentration of gene V-protein in the infected cell has reached a certain level, there is a switch-over from replicative form (RF) replication to single-stranded DNA synthesis[1-3]. This change is caused by the strong and cooperative binding of gene V-protein to

the displaced viral strands of the 'Rolling Circle' intermediates[4,5]. The resulting complexes, consisting of circular viral strands completely covered by gene V-protein, are precursors in the assembly process of the filamentous phages. The final stage of this process takes place at the cell membrane and involves the exchange of about one molecule of gene V-protein for two molecules of gene VIII-protein[6].

The binding of gene V-protein to single-stranded DNA has already been the subject of numerous investigations. Titration studies, using spectroscopic detection of complex formation, have shown that one gene V-protein covers ca four nucleotides[7-9]. This figure seems to be somewhat higher for complexes isolated from infected cells[7,9]. The protein exists predominantly as a dimer in solution[10]. The amino acid sequences of gene V-protein, encoded by the closely related bacteriophages M13, f1 and fd are identical[11-13]. The protein has a molecular weight of 9690 and contains 87 amino acid residues. Five are tyrosine of which three, at position 26, 41 and 56, are located on the surface of the molecule[8,9]. At least one tyrosine residue is involved in the interaction of gene V-protein with single-stranded DNA via stacking, as was concluded from nuclear magnetic resonance studies[14-17].

The latter studies[15-17] also suggested the involvement of at least one phenylalanine residue in this binding. Recent X-ray diffraction analysis of crystallized gene V-protein have substantiated the involvement of these amino acid residues in the binding reaction[18]. By U.V. irradiation of infected cells, Lica and Ray[19] cross-linked gene V-protein to DNA and isolated the complexes. Tryptic digestion of gene V-protein in these complexes revealed that the region encompassing amino acid residues between positions 70 and 77 are involved in the binding reaction.

In this Chapter we present an alternative approach to the study of DNA - gene V-protein interaction. The method, formerly applied by Miller *and co-workers*[20,21] in the study of lac repressor binding, is based on the idea that amino acid residues can be replaced at specific positions in a protein by suppression of amber mutations in the coding information of the protein. The effects of such specific replacements on the binding might give new insights into the mechanism of the binding reaction itself.

In the previous Chapter we have demonstrated that phage fd 122 contains the amber mutation at position 906, whereas f1 R99 is mutated at positions 999. The two CAG-codons present in the nucleotide sequence of wild-type gene V at these positions, encode the glutamine residues 22 and 53 in the amino acid sequence of gene V-protein. By infection of various *E.coli* suppressor strains with fd 122 or f1 R99, these glutamine residues can, in theory, be replaced now by many other amino acid residues. Especially, the possibility to introduce an extra tyrosine residue at the positions of both glutamines seems very attractive, because of the well-documented involvement of tyrosine residues in the DNA - gene V-protein interaction[13-17].

In this study we demonstrate that, of all the suppressor strains tested, only infection of the strains which are capable of inserting serine or tyrosine results in the synthesis of modified gene V-proteins which still possess a (limited) biological activity. The modified gene V-proteins were isolated, chromatographically purified and then subjected to DNA-binding studies. Preliminary results of these studies and the further perspectives of this approach in DNA - gene V-protein interaction research will be discussed.

Bacteria and phages

The KA strains used in this study were constructed by introduction of the F-plasmid in XA strains, which originate from Dr J.H. Miller (Miller *et al*, J. Mol. Biol. *109* (1977) 275-203). The KA strains were kindly made available to us by Dr B. Glickman, University of Leiden, and are listed in Table 1.

Table 1 *E.coli suppressor strains*

Strain	Miller ref	Type	Insertion
KA798	GMI/XAC	no suppressor	-
KA805	XA100	amber	*ser*
KA806	XA101	amber	*gln*
KA807	XA102	amber	*tyr*
KA808	XA103	ochre/amber	*lys*
KA809	XA104	amber	*leu*
KA810	XA105	ochre/amber	*gln*
KA811	XA106	ochre/amber	*tyr*

Bacteriophages f1 R99 and f1 R148 (amber mutants in gene V) were provided by Dr N.D. Zinder, New York. The amber5-mutant fd 122 was a gift of Dr H. Schaller, Heidelberg.

Immunoprecipitation

E.coli suppressor bacteria were grown in MTP medium[22] at 37°C to a density of $1x10^8$ cells/ml. The temperature was shifted to 32°C and one ml of the culture was infected with amber mutant phage with a multiplicity of infection of ca 10.

After an adsorption period of five minutes 12.5 µCi of ^{14}C-labelled amino acids (50 µCi/ml, protein hydrolysate of algal protein from Amersham, England) was added and infection continued for 90 minutes at 32°C. Then, the infected bacteria were spun down and washed once with sterile water. The bacterial pellet was resuspended in 0.5 ml of sterile water and sonified with a Branson sonifier to lyse the bacteria. The lysate was made 1x PBSTDS (PBSTDS buffer contains 10 mM Sodium phosphate, pH 7.2 and 0.15 M NaCl) and centrifuged for 30 min at 180.000xg.

Immunoprecipitation was performed by incubating 100 µl of the resulting supernatant with 50 µl of gene V-protein antiserum (a kind gift of Dr G. Simons, University of Nijmegen) in 1x PBSTDS and a total volume of 0.5 ml. After incubation for 15-18 h at 4°C, the mixture was layered on a sucrose cushion (200 µl of 10% sucrose in 1x PBSTDS) and centrifuged for 5 min in an Eppendorf minifuge at maximal speed. The supernatant was carefully removed and the precipitate washed three times with 0.5 ml of PBSTDS. After washing, the dried pellet was dissolved in 30 µl of cracking buffer, containing 0.0625 M Tris-HCl, pH 6.8, 8 M urea, 2% SDS, 5% 2-mercaptoethanol and 0.001% bromophenol blue, and heated for five minutes at 100°C. One-tenth of the mixture was assayed for radioactivity in a liquid scintillation counter. The remaining part was analysed on a 15% acrylamide - 0.4% bisacrylamide slab gel (14 cm x 12 cm x 0.1 cm) prepared in 0.05 M Tris-HCl, pH 8.9, 8 M urea and 0.1% SDS. After electrophoresis for 3 h at 30 mA the gel was fixed in 15% trichloroacetic acid (30 min) and prepared for fluorography as described by Bonner and Laskey[23].

Isolation of gene V-protein

Gene V-protein was isolated as described by Garssen *et al*[14] with modifications (N.C.M. Alma-Zeestraten *et al*, unpublished results). *E.coli*

suppressor strain was grown in 10 L M3 medium (Difco) supplemented with 0.2 mM $CaCl_2$ at 37°C until a density of $2x10^8$ cells/ml. The temperature was shifted to 32°C and amber mutant phage was added with a multiplicity of 10. After 4.5 h at 32°C the infected cells were collected by centrifugation and washed once with 500 ml of 40 mM Tris-HCl, pH 7.8, 10 mM magensium acetate and 60 mM KCl. The cells were suspended in 180 ml of 20 mM Tris-HCl, pH 8.1, 10 mM $MgCl_2$, 2 mM $CaCl_2$, 1 mM 2-mercaptoethanol and 1 mM EDTA and disrupted by sonification until the absorbance at 600 nm was 0.1 of the original value. The lysate was centrifuged at low speed to remove celldebris and clearified by high-speed centrifugation (2.5 h at 200,000xg). The resulting supernatant was made 2 M NaCl and 10% (w/v) Poly Ethylene Glycol 6000 and incubated for two hours at 4°C to precipitate the DNA. The precipitate was spun down (20 min, 19,200xg). The supernatant was dialyzed extensively against three 3 L changes of buffer A (20 mM Tris-HCl, pH 8.1, 50 mM NaCl, 5 mM EDTA and 1 mM 2-mercaptoethanol). The dialyzed extract was centrifuged at low speed to remove a light precipitate, made 10% (v/v) in glycerol and applied to a DNA-agarose column[24] (12 cm x 1 cm) which had been equilibrated in buffer A plus 10% (w/v) glycerol. Extensive washing with this buffer removed most of the applied proteins. Gene V-protein was eluted from the column with a linear 0.05 M - 2 M NaCl-gradient in buffer A plus 10% (w/v) glycerol (50 ml). Fractions containing gene V-protein were pooled, dialyzed against 1 mM Na-cacodylate, pH 6.9 and 50 mM NaCl and concentrated by ultrafiltration using an Amicon CEC1 column eluate concentrator. Gene V-protein was separated from contaminating high molecular weight material by gel filtration through a Sephadex G-75 column (30 cm x 2 cm). Equilibration and elution of the column were with 1 mM Na-cacodylate, pH 6.9 and 50 mM NaCl. Fractions containing pure gene V-protein were pooled and concentrated again with the

Amicon CEC1 column eluate concentrator. Gel electrophoresis was used throughout the isolation procedure to identify fractions containing gene V-protein. Samples to be analyzed were dialyzed against water, lyophilyzed and finally dissolved in 30 µl of cracking buffer. The samples were heated for five minutes at 100°C and applied to 15% polyacrylamide slab gels. After electrophoresis for 3 h at 30 mA, the gels were stained with Coomassie Brilliant Blue R-250.

DNA - gene V-protein binding assay

The binding of gene V-protein to single-stranded DNA was measured by titration of poly dA with purified gene V-protein following the tyrosine fluorescence at 303.5 nm (excitation at 276 nm). The method applied in this study has been described by Pretorius *et al*[9] but included several improvements. A full description of these improvements will be published elsewhere (Alma-Zeestraten *et al*, in preparation). In short, titrations were performed at 4°C in 1 mM Na-cacodylate pH 6.9, 120 mM NaCl by addition of increasing amounts of a solution containing gene V-protein (max 400 µl) to 800 µl 10-30 µMol poly dA (mean mol weight 150,000, Boehringer, Mannheim). The resulting fluorescence was corrected for the dilution of poly dA, introduced by the addition of gene V-protein. The concentrations of the gene V-protein containing solutions were determined from the absorbance at 276 nm using a molar absorption coefficient of 7100 M^{-1} cm^{-1} (ref 28). Concentrations of poly dA solutions were determined with $\epsilon(P)_{276\ nm} = 8520$. The number of bases covered by one gene V-protein molecule was (roughly) calculated from the equation $n = \frac{[\text{poly dA}]}{\Delta[\text{gene V-protein}]}$ in which [poly dA] is the concentration of poly dA and Δ[gene V-protein] the difference in concentration of gene V-protein at the reflection points in the titration curve (see Fig 4).

In the previous chapter we demonstrated that the two CAG (glutamine) codons at positions 906 and 999 in the wild-type DNA sequence of gene V are changed into amber codons in fd 122 and f1 R99, respectively. To find out whether a replacement of the original glutamines by other amino acid residues results in a biologically still active gene V-protein, these phages were tested for their ability to grow on various E. *coli* strains, which differ in their suppression specificity. As is shown in Table 2, plaque formation was only found on KA 805, KA 806 and KA 807 which insert resine, glutamine and tyrosine, respectively. Insertion of the original glutamine residue results in a lower titer than serine insertion, both at position 22 (in fd 122) and 53 (in f1 R99).

This phenomenon is in accordance with the observed differences in suppression efficiency of Su I (*ser*) and Su II (*gln*) E.*coli* strains, as measured by other methods (*cf* ref. 25). The absence of plaques, others than caused by revertants, on the KA 810 and KA 811 strains seems contradictory to the phage production on the KA 806 and KA 807 strains, which insert the same amino acid residues. A possible explanation might be that the former strains also suppress the ochre signals at the end of several filamentous phage genes (*cf.* ref 26), resulting in a greater percentage of biologically inactive phage-specific proteins. The same explanation is possibly valid for the KA 808 strain which inserts lysine at ochre and amber codons. The absence of plaques on the leucine inserting KA 809 strain might be caused by a low suppression efficiency. Alternatively, this amino acid replacement gives rise to the synthesis of a biologically inactive gene V-protein or a protein that is rapidly degraded in the infected cell (*cf* ref 27). We have not further investigated the various possible explanations for the absence of phage pro-

Table 2 PLATING OF *fd 122* AND *f1 R99* ON VARIOUS *E.COLI* SUPPRESSOR STRAINS

strain	type	insertion	efficiency of plating fd 122	f1 R99
KA798	no suppressor	-	3×10^{8}	2×10^{8}
KA805	amber	*ser*	4×10^{13}	5.6×10^{13}
KA806	amber	*gln*	2×10^{13}	1.5×10^{13}
KA807	amber	*tyr*	1.7×10^{13}	6×10^{12}
KA808	ochre/amber	*lys*	$<10^{9}$	$<10^{9}$
KA809	amber	*leu*	$<10^{9}$	$<10^{9}$
KA810	ochre/amber	*gln*	$<10^{9}$	$<10^{9}$
KA811	ochre/amber	*tyr*	$<10^{9}$	$<10^{9}$

duction in these suppressor strains but have focussed our attention on the tyrosine insertions by *E.coli* KA 807 in the gene V-protein encoded by fd 122 and f1 R99 in order to find out whether introduction of an extra tyrosine residue does or does not dramatically change the DNA-protein interaction. In the meantime we also studied whether serine insertions into gene V-protein (by the KA 805 strain) do influence its binding properties.

Determination of gene V-protein production by immunoprecipitation

To get insight into the amount of gene V-protein, produced by the amber 5-mutant infected *E.coli* KA 805 and KA 807 strains, we determined the gene V-protein content by immunoprecipitation. For this purpose the *E.coli* suppressor strains were grown to 1×10^{8} cells/ml and infected with fd 122 or f1 R99 (MOI=10) in the presence of a ^{14}C-labelled amino acid mix. In control experiments the same amount of wild-type M13 was added or phage addition was omitted. After 90 min at 32°C, the bacteria were collected by centrifugation and lysates were prepared as described under Materials and Methods. Gene V-protein was immunoprecipitated by incubating appropriate a-

mounts of the purified lysates with rabbit antiserum, directed against wild-type gene V-protein. 0.9 Part of each immunoprecipitate was analyzed on a 15% polyacrylamide slab gel (Fig 1). The remaining part was assayed for radioactivity (Table 3).

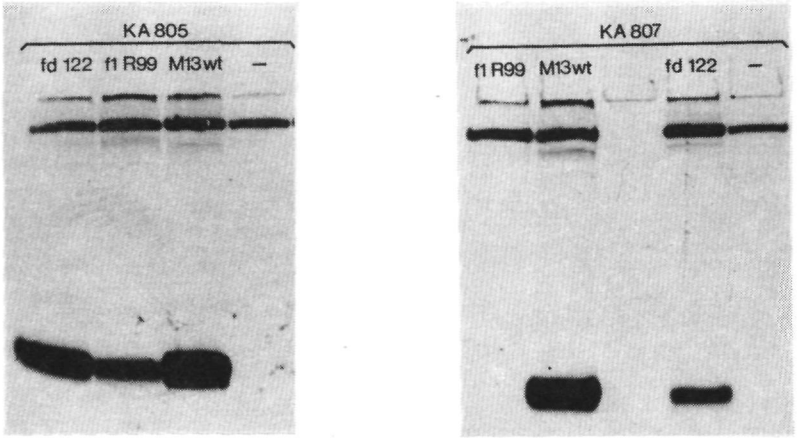


Fig 1 Electrophoretic analysis of immunoprecipitates
The Figure shows the ^{14}C-labelled products immunoprecipitated with antiserum directed against wild-type gene V-protein from *E.coli* KA 805 and KA 807 cells, infected with fd 122, f1 R99, M13 wt and not infected (-). Infection of cells and subsequent immunoprecipitation were performed as described in Materials and Methods. 0.9 Part of each immunoprecipitate was applied to the 15% polyacrylamide slab gels. The remaining part was assayed for radioactivity (Table 3).

Table 3

	phage	counts/ min	% of wild-type		phage	counts/ min	% of wild-type
KA805	fd 122	323.7	60	KA807	fd 122	234.4	30
	f1 R99	229.4	30		f1 R99	117.4	0
	M13 wt	523.3	100		M13 wt	536.4	100
	-	63.4	0		-	54.0	0

Assay of radioactivity in 0.1 part of each immunoprecipitate of which the electrophoretic analysis is shown in Fig 1.

Apart from a high molecular weight contaminant, which is also present in the non-infected bacteria, only gene V-protein is precipitated, demonstrating the specificity of the precipitation reaction. Precipitation under the applied conditions is quantitative, as was established in control experiments (data not shown). As is shown in Table 3, gene V-protein production by the amber mutant infected *E.coli* suppressor strains is low relative to the production of wild-type gene V-protein. We could not detect gene V-protein synthesis in the KA 807 strain, infected by f1 R99. However, some biologically active gene V-protein has to be present in these infected cells, as a low but distinct phage production was noted (see previous reaction). Possibly, the gene V-protein content was too low to be precipitated by the antiserum.

Isolation of modified gene V-proteins

The modified gene V-proteins were isolated from 10 L of *E.coli* KA 805 and KA 807 cultures, infected with fd 122 or f1 R99. Extracts were made as described in Materials and Methods and applied to a DNA-agarose column. Gene V-protein was eluted from the column with a linear 0.05 M-2.00 M NaCl gradient. The elution profiles of the four extracts are shown in Fig 2B-E. As a control an extract was made, in a similar way, of a 10 L culture of *E.coli* KA 807, infected with f1 R148 (Fig 2A). The latter mutant contains the amber mutation in gene VII (*cf* Chapter VI) and, consequently, should produce a wild-type gene V-protein.

The fractions containing gene V-protein were identified by electrophoretic analysis on 15% polyacrylamide slab gels. They are marked by hatched bars in the elution profiles.

The fractions in the second half of the f1 R148/KA 807 eluate and the fd 122/KA 805 (Fig 2A and B) eluate contained electrophoretically pure gene V-protein (Fig 3, *lane* 2 and 3). They were collected and, after

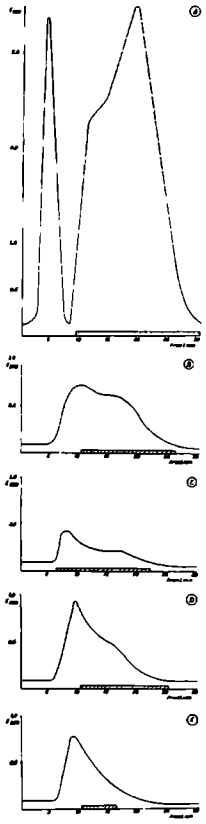


Fig 2 Elution profiles of extracts prepared from cultures of E.coli KA805 cells, infected with fd 122 (B) or f1 R99 (C) and from E.coli KA807 cells, infected with f1 R148 (A), fd 122 (D) or f1 R99 (E).
Extracts were made from 10 L of infected cultures as described in Materials and Methods and applied to an a DNA-agarose column (12 cm x 1 cm). Elution was with a linear 0.05 M-2.00 M NaCl gradient (fractions 0-30). The gene V-protein containing fractions are indicated by a hatched bar.

dialysis, directly used in the DNA binding assay.

All gene V-protein containing fractions in the fd 122/KA 807 and the f1 R99/KA 805 eluate (Fig 2C and D) were contaminated with high-molecular weight proteins. The fractions of both eluates were separately collected, concentrated by ultrafiltration and subjected to Sephadex-G75 filtration in order to separate the gene V-protein from the high-molecular weight contaminants. The gene V-protein preparations, obtained from the column were electrophoretically pure (Fig 3, *lane* 5 and 7).

Despite several efforts it was not possible to obtain enough f1 R99/

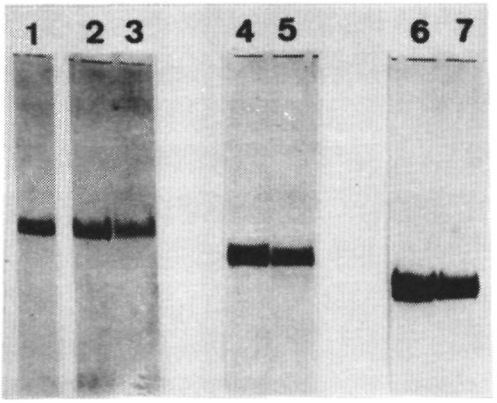


Fig 3 Electrophoretic analysis of modified gene V-proteins
The Figure shows the electrophoretic analysis of the purified gene V-proteins which are used in the DNA-binding assay. This Figure is a composite of three 15% polyacrylamide slab gels which were run for various times. Wild-type gene V-protein (*lanes* 1, 4 and 6) was included in each electrophoretic run in a parallel *lane* to serve as marker. After electrophoresis, proteins were visualized by staining with Coomassie Brilliant Blue.
Lane 2 Gene V-protein isolated from f1 R148 infected *E.coli* KA 807 cells
Lane 3 Gene V-protein isolated from fd 122 infected *E.coli* KA 805 cells
Lane 5 Gene V-protein isolated from f1 R99 infected *E.coli* KA 805 cells
Lane 7 Gene V-protein isolated from fd 122 infected *E.coli* KA 807 cells

KA 807 gene V-protein in a pure form. The very low amount of gene V-protein, in and around fraction 14 (Fig 2E), was heavily contaminated with high-molecular weight material, which could not be removed by gel filtration without severe losses of the protein.

We have not quantitated the yields of the modified gene V-proteins. However, it is clear from the elution profiles of the DNA-agarose column that the yields of these proteins are very low (not more than a few percent) as compared to the yield of wild-type gene V-protein, isolated from f1 R148 infected *E.coli* KA 807 cells.

Binding of modified gene V-protein to single-stranded DNA

The modified gene V-proteins were characterized by determination of their binding properties. The binding of gene V-protein to single-stranded DNA was studied in a titration experiment, in which the decrease in tyrosine fluorescence of the gene V-protein upon binding was measured.

Wild-type gene V-protein contains five tyrosine residues per molecule of which three, at positions 26, 41 and 56, are located at the surface of the molecule[8,9]. Binding of gene V-protein to single-stranded DNA results in a severe quencing of its tyrosine fluorescence. This is most probably caused by intercalation of one or more of the tyrosine residues, located at the surface, between the bases of the single-stranded DNA which results in a loss of their fluorescence.

In general, three phases can be recognized in a titration of single-stranded DNA with wild-type gene V-protein (see for example Fig 4A). In the first fase there is a linear increase in fluorescence, caused by the addition of gene V-protein. At the first deflection point, binding of gene V-protein to the single-stranded DNA starts, resulting in a lower total fluorescence. After the second deflection point the DNA is fully saturated with gene V-protein molecules and only the further addition of non-binding gene V-protein is measured.

The binding of the modified gene V-protein to the single-stranded polymer poly dA was studied. For each gene V-protein preparation the absorbance ratio 280 nm/260 nm (E280/E260) was calculated. The n-values, number of bases covered by one gene V-protein molecule, could (roughly) be determined from the titration curves as described in Materials and Methods).

The gene V-protein, isolated from f1 R148 infected *E.coli* KA 807 cells has a rather high E280/E260 ratio (1.9), comparable to the ratio

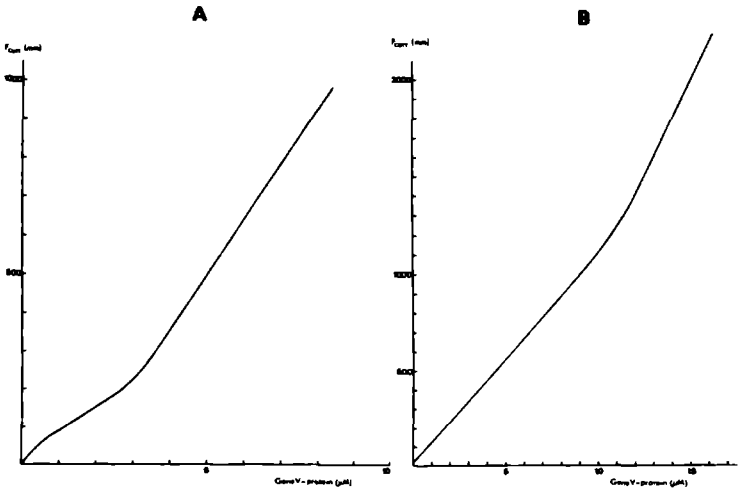


Fig 4 Titration of poly dA with modified gene V-proteins
Titrations were performed as described in Materials and Methods. The protein fluorescence was excited at 276 nm and measured at 303.5 nm.
A. 10 µMol poly dA titrated with gene V-protein isolated from fd 122 infected *E.coli* KA 805 cells
B. 26 µMol poly dA titrated with gene V-protein isolated from f1 R99 infected *E.coli* KA 805 cells
Note the difference in scale between both titration experiments.

of gene V-protein, isolated from cells infected with wild-type phage (E280/E260 = 2.1). Its binding property was identical to that of wild-type gene V-protein (n = 4-4,5). This once more indicates, as was established previously (Chapter VI), that this protein is encoded by a phage, not mutated in gene V.

The tritration of poly dA with gene V-protein, isolated from fd 122 infected *E.coli* KA 805 cells is shown in Fig 3A. The calculated n-value[4] is of the same magnitude as the n-value of wild-type gene V-protein. This indicates that the binding properties of gene V-protein drastically changed by a replacement of glutamine by serine at position 22. The E280/E260 ratio (1.7) is low, compared to wild-type gene V-protein, which suggest that some denaturation of the modified gene V-prot-

ein has occurred.

We could not detect binding between gene V-protein, isolated from fd 122 infected *E.coli* KA 807 cells and poly dA under the standard titration conditions. Also, the reverse titration of gene V-protein (10 µmol) with poly dA did not show any binding. This is rather unexpected for two reasons. At first infection of the *E.coli* KA 807 strain with fd 122 results in phage production (Table 2), indicating that biologically active gene V-protein has been formed. Secondly, modified gene V-protein binds to the DNA-agarose column as it could otherwise not be isolated. On the other hand, the E280/E260 ratio of the finally obtained protein is very low (1.3). Probably, the insertion of tyrosine at position 22 of the gene V-protein sequence still results in (partly?) biologically active gene V-protein, but the protein thus formed is rather labile and its biological activity is lost during the final steps of the purification procedure. Despite several efforts we have not succeeded in the isolation of sufficient quantities of this modified protein with a higher E280/E260 ratio.

The E280/E260 ratio of gene V-protein, isolated from f1 R99 infected *E.coli* KA 805 cells, was also very low (1.4), suggesting that the purified protein was, at least to a significant extent, in a denatured state. In analogy with the fd 122/KA 807 gene V-protein, we expected to find that this protein will not bind to single-stranded DNA. However, the *n*-value calculated from the titration curve (Fig 4B) is conderably smaller (2.2-2.6) than the *n*-value of wild-type gene V-protein (4-4.5). These data suggest that the replacement of glutamine by serine at position 53 in the amino acid sequence of gene V-protein results in an increase of the number of gene V-protein molecules which can cover the single-stranded DNA.

We do not know whether the observed *n*-value is an intrinsic property of

the modified gene V-protein, as the question remains unsolved whether one succeeds in isolating this protein in a less denatured form and, if so, whether the latter protein should reveal a higher *n*-value.

DISCUSSION

The use of suppression of amber mutations to create modified proteins offers an alternative approach in the analysis of protein structure and function. With this technique amino acid substitutions can be directed at various sites in a protein. With the set of *E.coli* suppressor strains, currently available, as many as five different amino acid substitutions can, in theory, be effected at each site in a protein for which an amber mutation is available.

However, the technique,applied in this study to specifically modify the gene V-protein of the filamentous bacteriophages, revealed some limitations. At first, it appeared that not all theoretically possible insertions at the two amber positions in gene V gave rise to a biologically still active gene V-protein. Apart form the original glutamine insertion, only serine and tyrosine insertion resulted in the production of progeny amber mutant phages (Table 2). Whatever causes the failure of the *E.coli* KA 808 (Su V) and *E.coli* KA 809 (see Chapter VI) strains to produced progeny phages, it is clear that they cannot be used directly for the introduction of lysine and leucine at specified positions in the gene V-protein structure. Secondly, to our opinion, the serine and tyrosine substitutions enhance, although to a varying extent, the labile character of the gene V-protein. This might be an explanation for the very low amounts of modified gene V-proteins which have been bound to the DNA-agarose column (Fig 3B, C, D and E) as compared to the amount of wild-type gene V-protein (Fig 3A). These low amounts of modi-

fied proteins are not primarly caused by a low level of synthesis of these proteins in the infected cells, as the quantities of modified gene V-proteins present in the cell-extracts are higher than the amounts of modified gene V-proteins recovered from the DNA-agarose columns as compared to the yield of wild-type gene V-protein in both instances (cf Fig 2 and Table 3). Possibly, many of the modified proteins have already lost (a part of) their biological activity during the first steps of the isolation procedure and, consequently, did not bind to the DNA-agarose column.

The labile character is also exemplified by the low E280/E260 ratios of the various modified proteins finally obtained. Especially the introduction of an extra tyrosine residue at position 22 in the amino acid sequence of gene V-protein seems to result in a very labile protein as this protein initially bound to a certain extent to the DNA-agarose column but failed to bind any poly dA in the binding assay.

Our failure to isolate a modified gene V-protein with a tyrosine insertion at position 53 migh also be due to a very labile character of this protein, but it has to be emphasized that the synthesis of this protein, as judged from our immunoprecipitation data, is very low (Fig 2 and Table 3).

As far as the serine insertions are concerned, we have concluded that insertion of a serine at position 22 does not alter the binding properties of gene V-protein. Serine insertion at position 53 does result in a protein which, irrespective of its low E280/E260 ratio, has a binding capacity which suggests that this protein is only capable of covering less than four nucleotides per monomer (n-value is 2.2-2.6). Similar low n-values are observed also in nuclear magnetic resonance studies of binding of wild-type gene V-protein to short polynucleotides (chain length < 20 nucleotdies), whereas binding of wild-type gene V-protein to

long polynucleotide chains is characterized by an n-value of 4-4,5[7-9]. We do not know whether the low n-value observed with the serine insertion at position 53 is caused by a partial denaturation of the protein studied or whether this protein lacks the intrinsic property of wild-type gene V-protein to shift its binding capacity from low to $n = 4$ values in contact with short and long polynucleotides. Further studies are needed to solve this point of interest.

1. Salstrom, J.S. and Pratt, D. (1971) J. Mol. Biol. *61*, 489-501.
2. Mazur, B.J. and Model, P. (1973) J. Mol. Biol. *78*, 285-300.
3. Mazur, B.J. and Zinder, N.D. (1975) Virology *68*, 490-502.
4. Oey, J.L. and Knippers, R. (1972) J. Mol. Biol. *68*, 125-138.
5. Alberts, B.M., Frey, L. and Delius, H. (1972) J. Mol. Biol. *68*, 139-152.
6. Day, L.A. and Wiseman, R.L., Comparison of DNA Packaging in the Virons of fd, Xf and Pf1. In: The single-stranded DNA Phages, Denhardt, D.T., Dressler, D. and Ray, D.S., Eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp 605-625.
7. Pratt, D., Laws, P. and Griffith, J. (1974) J. Mol. Biol. *82*, 425-439.
8. Anderson, R.A., Nakashima, Y. and Coleman, J.E. (1975) Biochemistry *14*, 907-914.
9. Pretorius, H.T., Klein, M. and Day, L. (1975) J. Biol. Chem. *250*, 9262-9269.
10. Cavalieri, S.J., Neet, K.E. and Goldthwait, D.A. (1976) J. Mol. Biol. *102*, 697-711.
11. Cuypers, T., Ouderaa van der, F.J. and Jong de, W.W. (1974) Biochem. Biophys. Res. Commun. *59*, 557-564.
12. Nakashima, Y., Dunker, A.K., Marvin, D.A. and Konigsberg, W. (1974) FEBS Lett. *40*, 290-292 and *43*, 125.
13. Hill, D.F. and Petersen, G.B. (1980) J. Virology *34*, 40-50.
14. Garssen, G.J., Hilbers, C.W., Schoenmakers, J.G.G. and Boom van, J.H. (1977) Eur. J. Biochem. *81*, 453-463.
15. Garssen, G.J., Kaptein, R., Schoenmakers, J.G.G. and Hilbers, C.W. (1978) Proc. Natl. Acad. Sci. U.S.A. *75*, 5281-5285.

16. Coleman, J.E. and Armitage, I.M. (1978) Biochemistry *17*, 5038-5045.

17. Garssen, G.J., Tesser, G.I., Schoenmakers, J.G.G. and Hilbers, C.W. (1979) Biochim. Biophys. Acta *607*, 361-371.

18. McPherson, A., Jurnak, F.A., Wang, A.H.J., Molineux, I. and Rich, Rich, A. (1979) J. Mol. Biol. *134*, 379-400.

19. Lica, L. and Ray, D.S. (1977) J. Mol. Biol. *115*, 45-59.

20. Smitz, A., Coulondre, C. and Miller, J.H. (1978) J. Mol. Biol. *123*, 431-456.

21. Miller, J.H., Coulondre, C., Hofer, M., Schmeissner, U., Sommer, H., Schmitz, A. and Lu, P. (1979) J. Mol. Biol. *131*, 191-222 (1979).

22. Marvin, D.A. and Hohn, B. (1969) Bacteriol. Rev. 33, 172-209.

23. Bonner, W.M. and Laskey, R.A. (1974) Eur. J. Biochem. *46*, 83-88.

24. Schaller, H., Nüsslein, C., Bonhoeffer, F.J., Kurz, C. and Nietzschman, I. (1972) Eur. J. Biochem. *26*, 474-481.

25. Lewin, B.M. (1974) Gene Expression, John Wiley and Sons, London, Vol. 1, pp 169-225.

26. Wezenbeek van, P.M.G.F., Hulsebos, Th.J.M. and Schoenmakers, J.G.G. (1980) Gene *11*, 129-148.

27. Schlotmann, M. and Beyreuther, K., Eur. J. Biochem. *95*, 39-49.

28. Day, L.A. (1973) Biochemistry *14*, 907-914.

In the first part of this thesis we describe the elucidation of the nucleotide sequence of the region on the bacteriophage M13 genome which is most frequently transcribed. This region which encompasses the genes II, V, VII and VIII is located in front of the *rho*-independent central terminator of transcription. Several features concering the expression of this region were tried to clearified: for example, the intruiging observation that gene VII is not expressed to a detectable level, either *in vivo* or *in vitro*. This is rather peculiar as this gene is located by genetic criteria between genes V and VIII. One would expect that also the product of gene VII, like the products of genes V and VIII, is abundantly made in the infected cell as well as *in vitro* cell-free systems.

Another unresolved problem is that of X-protein. This protein is encoded by a DNA region which is located in the C-terminal part of gene II. It is still an open question whether gene X, which is equipped with its own promoter, uses the same reading frame as gene II or whether it uses a different one. Furthermore, X-protein is synthesized *in vitro* in rather large amounts, but this protein has not yet been detected in the infected cell.

To gain more insight into the factors that regulate the differential expression of this important part of the M13 genome we have deduced its DNA sequence.

As nucleotide sequencing studies require, in general, rather small DNA fragments, we have first extended our collection of restriction enzyme cleavage maps in order to increase the number of restriction enzyme cleavage sites in the region of interest. The constructions of the *Hha* I and *Mbo* II restriction enzyme cleavage maps have been des-

cribed in Chapter II. The ordering of *Hha* I fragments was achieved by using each DNA fragment as a primer for the *in vitro* synthesis of its respective neighbour. The *Mbo* II fragments were ordered by partial digestion analysis of large M13 DNA fragments which were labelled at only one end.

In Chapter III the nucleotide sequence of gene VII has been deduced. A 'reversed' transcript of the so-called '300-fragment' could be made by performing the transcription in the presence of dinucleotide primer. The nucleotide sequence of the 'reversed' RNA transcript which comprises the entire coding information of gene VII, was elucidated by sequence analysis of T1-oligonucleotides and ordering of T1-fragments. The reading frame of gene VII was established by analysis of the positions of two amber mutations in this gene. The positions were traced by comparison of the T1-oligonucleotides of the 'reversed' transcripts of wild-type and amber7-mutant phage DNA's.

From our sequencing studies the presence of a new, hitherto unidentified gene became apparent. Its postulated length is 96 nucleotides. Firm evidence is now available that this gene, called gene IX, really exists and codes for one of the minor coat proteins.

The transcription of gene VII and its flanking regions was studied in more detail in Chapter IV. The RNA sequence of a short RNA, originating from this region, has been deduced by standard RNA sequencing techniques. It was shown that this RNA consists of the first 45-50 nucleotides of the mRNA which encodes the major coat protein. It is argued that the short RNA is most probaly generated by a premature termination of the synthesis of the major coat protein message by RNA polymerase molecules firmly bound to a promoter-like DNA sequence.

The complete DNA sequence of the gene X-VIII region, comprising about 1300 basepairs, has been deduced in Chapter V using the Maxam-

Gilbert method of DNA sequencing.

Our sequencing studies revealed that gene X constitutes the C-terminal part of gene II and uses the same reading frame as the latter gene. The positions of the genes V and VIII in the DNA sequence were determined on the basis of the known amino acid sequences of their gene-products. Together with the positions of the genes VII and IX, a detailed picture of the genetic organization of this part of the M13 genome could be given. The sequence data furthermore provided detailed information on the primary structure and the precise locations of the various regulatory elements in the gene X-VIII region such as the three promoters $G_{0.06}$, $G_{0.12}$ and $G_{0.18}$, the *rho*-independent central terminator of transcription and the ribosome binding sites. The occurence of rather stable secundary structures in the RNA transcripts have been discussed in relation to the regulation of gene expression and to RNA processing *in vivo*.

In the second part of this thesis our knowledge of the nucleotide sequence has been applied for gaining more information on the function of gene V and its product.

Chapter VI deals with the long standing problem of gene V-gene VII polarity. This polarity was investigated by studying the positional effects of amber5-mutations on gene VII expression. For this purpose we have determined the positions of the amber mutations in several amber5-mutants using the chain-termination method of DNA sequencing. Of the six amber5-mutants analyzed, the amber mutation was actually found at only two positions, namely at position 906 and at position 999 in the nucleotide sequence of gene V, whereas of one amber5-mutant (f1 R148) it could be demonstrated that the amber mutation was located in gene VII instead of gene V.

Our complementation analyses with these amber5-mutants suggest that both mutants abolish the gene VII-expression. A model, which might explain the similar effects of both gene V-mutant on gene VII-protein synthesis has been discussed.

Since gene V-protein acts as a DNA-binding protein during the life cycle of the phage, both amber5-mutants have been applied for specifically modifying gene V-protein and to study these altered proteins in their interaction with single-stranded DNA. The modified proteins were generated by infection of various *E.coli* suppressor strains with both types of amber5-mutants. It was found that, of the seven different *E.coli* suppressor strains, only the Su I (serine) and the Su III (tyrosine) were capable of producing biologically still active gene V-protein upon infection with the two amber5-mutants.

The modified proteins, *i.e.* one with a tyrosine insertion at position 22 and two proteins with a serine insertion at position 22 and 53, respectively, have been isolated and purified to homogeneity.

Our binding experiments, carried out by measuring the change in tyrosine fluorescence of the modified proteins upon binding to poly dA, have demonstrated that the serine insertion at position 22 does not change the binding characteristics of gene V-protein. Under the standard binding conditions applied, the gene V-protein with an extra tyrosine residue at position 22 failed to bind to poly dA. The data obtained with gene V-protein with a serine insertion at position 53 suggest that this protein as a monomer covers less than four nucleotides. Detailed physical-chemical studies of these modified proteins in their interaction with DNA are limited by their labile character.

In het eerste gedeelte van dit proefschrift wordt de opheldering beschreven van de nucleotiden volgorde van het gebied op het genoom van de bacteriofaag M13, wat het meest wordt getranscribeerd. Dit gebied ligt voor de *rho*-onafhankelijke centrale terminator van de trancriptie en omvat de genen II, V, VII en VIII.

We hebben getracht om een verklaring te vinden voor een aantal tot nu toe onopgeloste problemen betreffende de expressie van dit gebied, zoals bijvoorbeeld de observatie dat gen VII, zowel *in vivo* als *in vitro*, niet in die mate tot expressie komt dat zijn gen-produkt aangetoond kan worden. Dit laatste is vreemd, omdat gen VII, op genetische gronden, tussen gen V en gen VIII is gelegen. De gen-produkten van beide laatste genen worden, zowel in de geïnfecteerde cel als in celvrije *in vitro* systemen, in grote hoeveelheden aangemaakt. Men zou verwachten dat dit ook het geval is voor het gen VII-produkt.

Een ander probleem is het X-eiwit. Van dit eiwit weten wij dat het wordt gecodeerd door een gebied dat gelegen is in het C-terminale deel van gen II. Wij weten ook dat het gen, wat codeert voor dit eiwit, - gen X - een eigen promotor heeft. Onbekend is echter of gen X gebruik maakt van hetzelfde "lees-frame" als gen II of van een ander. Een tweede probleem inzake het gen X-eiwit betreft de regulatie van de synthese van dit eiwit. Het eiwit wordt *in vitro* in redelijk grote hoeveelheden gemaakt. *In vivo* heeft men het echter tot nu toe niet aan kunnen tonen.

Met het doel om meer inzicht te krijgen in de faktoren die betrokken zijn bij de regulatie van de expressie van dit belangrijke gebied op het M13 genoom hebben wij de nucleotiden volgorde van dit gebied opge-

helderd.

Omdat bij de opheldering van nucleotiden volgordes in het algemeen van kleine fragmenten wordt uitgegaan, hebben wij onze bestaande collectie van splitsingskaarten van restrictie enzymen uitgebreid om zodoende meer splitsingsplaatsen te creeëren in het gebied waarvan wij de nucleotiden volgorde wilden ophelderen. De bepaling van de splitsingskaarten voor de restrictie enzymen *Hha* I en *Mbo* II is beschreven in Hoofdstuk II.

De volgorde van de *Hha* I fragmenten werd bepaald door ieder DNA-fragment als primer te gebruiken voor de *in vitro* synthese van zijn naaste buur. De volgorde van de *Mbo* II fragmenten werd bepaald door partiële digestie van lange M13 fragmenten, die slechts aan één kant een radioaktief label dragen.

In Hoofdstuk III is de nucleotiden volgorde van Gen VII opgehelderd. Door het zogenaamde "300-fragment" te transcriberen in de aanwezigheid van een dinucleotide als primer kon de synthese van een RNA worden gestimuleerd, die getranscribeerd wordt van de virale streng. Dit RNA bevat de gehele codogene informatie van gen VII. De nucleotiden volgorde van dit RNA werd opgehelderd door de sequentie van de oligonucleotiden die ontstaan door RNAse-T1 digestie op te lossen en door de volgorde van deze T1-oligonucleotden te bepalen. Het "lees-frame" van gen VII kon afgeleid worden uit de plaats van de amber mutaties in twee amber7-mutanten. Deze plaatsen konden bepaald worden door T1-oligonucleotides van de wilde-type vorm van bovengenoemd RNA te vergelijken met die van de RNA's, getranscribeerd van het DNA van de amber7-mutanten. Door het nucleotiden volgorde onderzoek werd het bestaan van een nieuw gen - gen IX - met een lengte van 96 nucleotiden, aannemelijk gemaakt. Er zijn nu

overtuigende evidenties voorhanden, die aantonen dat gen IX echt bestaat. Gen IX codeert een eiwit dat in geringe hoeveelheden in de faagmantel voorkomt.

De transcriptie van gen VII en de gebieden aan de weerszijden van dit gen werden meer gedetailleerd bestudeerd in Hoofdstuk IV.

Met behulp van standaard RNA sequentie technieken werd de nucleotiden volgorde van een klein RNA - wat afkomstig was uit dit gebied - opgehelderd. Dit RNA blijkt 45-50 nucleotiden lang te zijn en een nucleotiden volgorde te hebben, identiek aan die van het 5'-terminale gedeelte van het RNA, dat codeert voor het gen VIII eiwit. Dit RNA wordt waarschijnlijk gevormd door een voortijdige terminering van de synthese van het RNA, dat codeert voor het gen VIII eiwit door toedoen van RNA polymerase moleculen die hechten aan een gebied op het DNA, dat promoter-achtige eigenschappen heeft.

De volledige nucleotiden volgorde van het gen X-VIII gebied is opgehelderd in Hoofdstuk V.

De volgorde van de ca. 1300 nucleotiden is bepaald met behulp van de chemische splitsingstechniek van Maxam and Gilbert. Uit de nucleotiden volgorde blijkt dat gen X het C-terminale deel van gen II omvat en dat gen X in het zelfde "frame" wordt afgelezen als gen II.

De posities van de genen V en VIII in de nucleotiden volgorde werden bepaald met behulp van de bekende aminozuur-sequenties van hun respektievelijke gen-produkten. Samen met de posities van de genen VII en IX kan nu een gedetailleerd beeld van de genetische organisatie van dit gedeelte van het M13 genoom worden gegeven. Uit de nucleotiden volgorde werd ook gedetailleerde informatie verkregen over de primaire struktuur en de preciese plaats van de verschillende regulatoire elementen, zoals

de drie promotors $G_{0.06}$, $G_{0.12}$ en $G_{0.18}$, de *rho*-onafhankelijke centrale terminator van de transcriptie en de bindingsplaatsen voor ribosomen. Tenslotte is de aanwezigheid van vrij stabiele secundaire strukturen in de RNA transcripten van dit gebied in relatie gebracht met de regulatie van de gen-expressie en met de specifieke splitsing van deze transcripten zoals dat in de geïnfecteerde cel wordt waargenomen.

In het tweede gedeelte van dit proefschrift is de kennis van de nucleotiden volgorde toegepast in het verkrijgen van meer inzicht in de funktie van gen V en het gen V-eiwit.

Hoofdstuk VI heeft het reeds lang bekende fenomeen van de gen V-gen VII polariteit tot onderwerp. Deze polariteit werd onderzocht door het effect na te gaan van de plaats van amber mutaties in gen V op de expressie van gen VII. Met dit doel voor ogen is de plaats van de amber mutatie in verschillende amber5-mutanten vastgesteld door de nucleotiden volgorde bepaling met behulp van de ketenbeëindigingsmethode van Sanger en medewerkers.

Onder de 6 amber5-mutanten die geanalyseerd zijn, werden maar twee amber mutatie-plaatsen gevonden, namelijk op plaats 906 en op plaats 999 in de nucleotiden volgorde van gen V. Van één amber5-mutant (f1 R148) kon aangetoond worden dat deze gemuteerd is in gen VII in plaats van gen V. De complementatie-experimenten,die werden uitgevoerd met deze amber5-mutanten,suggereerden dat beide mutanten de expressie van gen VII in gelijke mate (ver)hinderen. Een model waarmee het identieke effekt van beide amber5-mutanten op de gen VII-eiwitsynthese verklaard zou kunnen worden, werd in dit Hoofdstuk besproken.

Omdat gen V-eiwit gedurende de levenscyclus van de faag een enkelstrengs DNA bindende funktie heeft, werden beide amber5-mutanten gebruikt om het gen V-eiwit op specifieke plaatsen te veranderen en wer-

den de gemodificeerde eiwitten gebruikt om hun interaktie met enkelstrengs DNA te bestuderen. De gemodificeerde eiwitten werden gemaakt door verschillende *E.coli* suppressor stammen te infecteren met elk van beide amber5-mutanten. Het bleek dat van de zeven *E.coli* suppressor stammen, slechts de Su I-stam, die serine inserteert en de Su III-stam die tyrosine inserteert in staat waren om gen V-eiwitten te produceren die nog biologisch aktief zijn. De gemodificeerde eiwitten, één met tyrosine op plaats 22 en twee eiwitten met respektievelijk een serine insertie op plaats 22 en 53 werden geïsoleerd en gezuiverd tot een homogeen preparaat was verkregen.

De binding van de gemodificeerde gen V-eiwitten aan enkelstrengs DNA werd vervolgd door de verandering in de fluorescentie van de tyrosine residuen in het gen V-eiwit te meten en na binding aan poly dA. Het bleek dat de serine insertie op plaats 22 de bindingskarakteristieken van gen V-eiwit niet verandert. Onder de standaardbindingscondities bond het gen V-eiwit met een extra tyrosine op plaats 22 in het geheel niet aan poly dA. De bindingsexperimenten met het gen V-eiwit dat een serine insertie op plaats 53 heeft, suggereren dat de monomere vorm van dit eiwit minder dan vier nucleotiden verbindt.

Gedetailleerd fysisch-chemische karakterisering van de interaktie van deze gemodificeerde eiwitten met DNA wordt beperkt door het labiele karakter van deze eiwitten.

CURRICULUM VITAE

De schrijver van dit proefschrift werd op 14 oktober 1949 te Raalte geboren.

Na het behalen van het diploma HBS-B (Thomas a Kempis Lyceum, Zwolle) begon hij in september 1968 met de studie Biologie aan de Katholieke Universiteit te Nijmegen.
Het kandidaatsexamen Biologie (B4) werd in februari 1972 afgelegd.
Het doktoraalexamen Biologie, met als hoofdvak Chemische Cytologie (Hoofd: Prof.dr. Ch.M.A. Kuyper) en als bijvakken Botanie (Hoofd: Prof. dr. H.F. Linskens) en Moleculaire Biologie (Hoofd: Prof.dr. J.G.G. Schoenmakers) werd in december 1974 afgelegd.

Hij was van december 1974 tot oktober 1979 werkzaam op het Laboratorium voor Moleculaire Biologie van de Katholieke Universiteit te Nijmegen. In deze tijd werd het in dit proefschrift beschreven onderzoek verricht. Sinds 1 januari 1980 is hij als wetenschappelijk medewerker werkzaam op de Afdeling Genetica (Hoofd: Prof.dr. W.H.G. Hennig) van de Katholieke Universiteit.

Addendum: enlargement of table 2 (Chapter V)

		Homology	
		Capital letters	Small letters
Modal sequence	a c c . t . g t t G T T G A c A T T T t t . . . . t t g g c G G T T A T A T T g . . . c C A T . . . . . a . . . . . t t t t a a A T A a g		
$G_{0.06}$ (gene X)	A T C T T T T G A T G C A A T C C G C T T T G C T T C T G A C T A T A A T A G T C A G G G T A A A G A C C T G A T T T T T 380 400 420 440	9	13
$G_{0.18}$ (gene IX, VIII)	C A A A T C T C C G T T G T A C T T T G T T T C G C G C T T G G T A T A A T C G C T G G G G G T C A A A G A T G A G T G T T 1160 1180 1200	15	8
$G_{0.12}$ (gene V) 1st poss	A C G T A G A T T T T T C T T C C C A A C G T C C T G A C T G G T A T A A T G A G C C A G T T C T T A A A A T C G C A T A A 780 800 820	11	9
$G_{0.12}$ (gene V) 2nd poss	C A A C G T C C T G A C T G G T A T A A T G A G C C A G T T C T T A A A A T C G C A T A A G G T A A T T C A C A A T G A T T 800 820 840	8	9

STELLINGEN

I

Uit de waarneming dat de aminozuursamenstelling van het gen V-eiwit, dat gecodeerd wordt door een amber7-mutant van bacteriofaag f1, niet verandert ten opzichte van het wilde type gen V-eiwit, trekken Mazur en Zinder ten onrechte de conclusie dat gen VII niet het C-terminale deel van gen V omvat.

Dit proefschrift

Mazur, B.J. en Zinder, N.D. Virol. (1975) 68, 284-285.

II

Omdat Gattoni *et al* niet bepalen welke van de eiwitten uit hn RNP partikels rechtstreeks een interaktie aangaan met hn RNA, is hun conclusie dat de RNA-eiwit interaktie niet aan stringente specificiteit onderworpen is, aanvechtbaar.

Gattoni, R., Stevenin, J. en Jacob, M., Eur. J. Biochem. (1980) 108, 203-211.

III

Uit het feit dat organische kationen bij *Saccharomyces cerevisae* geen lek veroorzaken van celmateriaal dat bij 260 nm absorbeert, mogen Pena *et al* niet concluderen dat organische kationen dan ook de integriteit van de celmembraan niet aantasten.

Pena, A., Mora, M.A. en Corrasco, N., J. Membr. Biol. (1979) 47, 261-284.

IV

De "urncellen" die bij *Siponculiden* een agglutinerende rol vervullen, moeten eerder als lichaamseigen dan wel als interne parasieten worden beschouwd.

In de publikatie van de ruimtelijke struktuur van het gen V-eiwit van bacteriofaag fd zijn door McPherson *et al* de resultaten dermate inadequaat weergegeven dat wij onze platte denkbeelden over eiwit-nucleinezuur interakties helaas geen drie-dimensionaal cachet kunnen geven.

McPherson, A., Jurnak, F.A., Wang, A.H.J., Molineux, I. en Rich, A., J. Mol. Biol. (1979) 134, 379-400.

VI

Op grond van hun experimenten kunnen Ritz *et al* niet concluderen dat het monoclonale antilichaam gericht tegen het "Common Acute Lymphoblastic Leukaemia Antigen" specifiek is voor leukemische cellen.

Ritz, J., Pesando, J.M., Notis-McConarty, J., Lazarus, H. en Schlossman, S.F., Nature (1980) 283, 583-585.

VII

De conclusie van Chooi dat de "intervening sequences" in de ribosomale genen van *Drosophila melanogaster* worden getranscribeerd, is gebaseerd op onvoldoende experimentele gegevens.

Chooi, W.Y., Chromosoma (Berl.) (1979) 74, 57-74.

VIII

Voor het initiatiecodon van verschillende genen van ϕx174 worden door Bahramian in het RNA secundaire strukturen verondersteld, die betrokken zouden zijn bij de initiatie van de eiwitsynthese. Uit energetische overwegingen moet het voorkomen van deze strukturen en de veronderstelde biologische funktie als uitermate onwaarschijnlijk worden beschouwd.

Bahramian, M.B., J. theor. Biol. (1980) 84, 103-118.

IX

Gezien de korte duur van de arbeidscontracten voor "post-docs" enerzijds en de huidige malaise op de woningmarkt anderzijds, is voor hen de aanschaf van een caravan te prefereren boven de aankoop van een huis.

Nijmegen, 12 december 1980 T. Hulsebos