

NUCLEOTIDE SEQUENCE OF A
TRANSCRIPTION TERMINATION
REGION IN COLIPHAGE T7

By John C. Boothroyd

A Thesis presented for the degree of
Doctor of Philosophy

Department of Molecular Biology
University of Edinburgh
Scotland

June 1979

TO YVONNE

Abstract

A detailed physical map of restriction endonuclease cutting sites has been determined for the region surrounding the transcription-termination site at the end of the early operon of coliphage T7. Using this information and the chain-terminating method for sequencing DNA, a nucleotide sequence 370 base-pairs in length has been determined corresponding to the DNA between 18.88 and 19.81 on the T7 physical map.

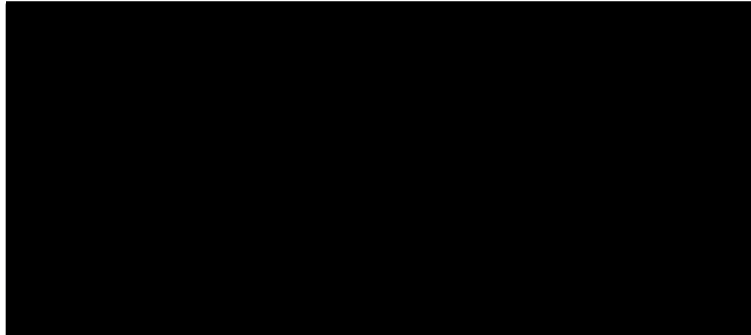
By comparison with the data of others, including nucleotide sequences for regions of known physiological importance, several interesting functional sites have been tentatively identified within the sequence presented. These include: 1) the termination site (map position 18.96) for transcription of the early operon by E.coli RNA polymerase; 2) two sites (19.35 and 19.65) likely to act as promoters for the initiation of class II transcripts by T7 RNA polymerase; 3) two regions (18.89-19.31 and 19.34-19.61) each of which includes a plausible ribosome-binding site and which probably code for a pair of previously unreported class II genes tentatively designated genes 1.4 and 1.5, respectively; 4) the ribosome-binding site (position 19.65) and initial 18 codons of the class II gene 1.7.

A particularly striking feature of this sequence is the frugal arrangement of several of the proposed functional sites so that two perfect, direct repeats of 21 base-pairs contain: the stop-codons for the putative genes 1.4 and 1.5, the two class II promoters, and the ribosome-binding sites for genes 1.5 and 1.7, respectively.

Declaration

I hereby declare that I alone have composed this thesis and that, except as noted, all the work presented within it is my own.

June 1979



Acknowledgements

I would like to take this opportunity to thank all those who have played a significant part in the production of this thesis:

Richard Hayward for his friendship and outstanding supervision; Kevin O'Hare and Andy Newman for their stimulating companionship in and out of the lab. I could not have asked for three finer people with whom to work;

all the other past and present members of the Department for their friendship and innumerable helpful discussions;

my family upon whose love and advice I shall always rely; the E.U. Renaissance Singers who collectively befriended me and tried so hard to satisfy my yearning for mountains and music;

my peripatetic friends, especially Julia and Bruce for all they have done for me;

Anna for the efficient typing of this thesis;

Professors Murray and Pollock for allowing me to work within the Department;

my parents, The Royal Commission for the Exhibition of 1851, The Royal Society of Canada and McGill University for financial support.

Abbreviations:

A ₂₆₀	absorbance at 260 nm
A:T	adenosine:thymine base-pair
A + T	adenosine plus thymine content
bp	base-pair
Bpb	bromophenol blue
dH ₂ O	distilled water
ddNTP	2',3'-dideoxyribonucleoside 5'-triphosphate
DpolI	<u>E.coli</u> DNA polymerase I (Klenow A fragment)
ds	double stranded
DTT	dithiothreitol
EndoR.	restriction endonuclease
<u>Kg</u>	1000 x the acceleration due to gravity
KPi	KH ₂ PO ₄ + K ₂ HPO ₄
Krpm	1000 x rpm
2-me	2-mercaptoethanol
pfu	plaque-forming units
R	purine nucleotide
rb	ribosome binding (site)
rif	rifampicin
Rpol	RNA polymerase
rt	room temperature
ss	single stranded

CONTENTS

	Page
CHAPTER 1. Introduction	1
1. <u>E.coli</u> RNA polymerase	1
2. Transcription	4
a. Initiation	
i. Promoter	5
ii. Operator	7
b. Elongation	10
3. Termination	11
a. Mechanism of termination <u>in vitro</u>	12
i. Rho-independent terminators	12
ii. Rho-dependent terminators	20
b. Regulation of termination <u>in vivo</u>	
i. Antitermination mediated by the N gene product of bacteriophage λ	24
ii. Attenuation in the leader region of the <u>trp</u> , <u>phe</u> and <u>his</u> operons	28
4. Background and approach to the present study of termination in coliphage T7	36
a. Coliphage T7	37
i. Class I gene expression	38
ii. Class II and III gene expression	43
b. Nucleotide sequencing	44
CHAPTER 2. Materials	
1. Bacteria and bacteriophage	46
2. Media	46
3. Restriction endonucleases	46
4. Other enzymes	47
5. Nucleotides	47
6. Radiochemicals	47
7. Gel electrophoresis	47
8. Gel photography	48
9. Autoradiography	48
10. Chromatography	48
11. Miscellaneous	49
12. Suppliers	50
CHAPTER 3. Methods	
1. General	52
a. Siliconisation of glassware	52
b. Sterilisation of dialysis sacs	52
c. Centrifugation	52
d. Phenol extraction	53
e. Ethanol precipitation	53

	Page
2. T7 growth and DNA purification	
a. Phage titration	54
b. Large scale lysates	54
c. Small scale lysates	56
d. Plate lysates	56
e. T7 DNA strand separation	57
3. Restriction endonuclease (EndoR.) preparation	
a. EndoR. HpaI	58
b. EndoR. HpaII	60
c. EndoR. AvaI	60
4. Restriction endonuclease reactions	. 61
5. Gel electrophoresis	61
a. Vertical agarose	62
b. Horizontal agarose	64
c. Preparative polyacrylamide	65
d. Analytical polyacrylamide	67
e. Sequencing polyacrylamide	68
6. Gel photography	70
7. Autoradiography	70
8. DNA fragment-purification from gels	71
a. Electroelution	71
b. Passive elution	72
9. DNA sequencing	
a. Dideoxy-blocking of template	73
b. Sequencing	74
c. Computer analysis	76
10. Binding of DNA fragments to nitrocellulose filters	77
 CHAPTER 4. Physical mapping of restriction endonuclease cutting sites in the T7 WT genome	 79
1. Enzymes which do not cut	83
2. Enzymes which make between 1 and 9 cuts	83
a. AvaI	83
b. BclI	84
c. KpnI	85
d. MboI	85
e. XbaI	86
3. Enzymes which make more than 9 cuts	86
a. HaeII	87
b. HaeIII	88
c. HgiAI	90
d. HhaI	92
e. HinII	93
f. HpaI	95
g. HpaII	96
4. Deletion size estimations	98
5. Mapping of deletions	98

	Page
a. T7 C5	99
b. T7 LG3	99
c. T7 LG26	100
d. T7 LG37	100
6. Summary	101
CHAPTER 5. Analysis of ternary complexes of Rpol, DNA and RNA	102
CHAPTER 6. Nucleotide sequence determination in the region of the T7 early transcription termination site, near 19.29 on the physical map	
1. Approach	107
2. Results	110
3. Structural features	113
a. Base composition	113
b. Direct repeats	113
c. Inverted repeats	113
CHAPTER 7. Discussion	115
1. Mapping of EndoR. cutting sites	115
2. Fragment binding to nitro-cellulose	116
3. Nucleotide sequencing	
a. Artefacts	116
b. Functional features	
i. Termination of transcription at T7 t ₁	118
ii. Class II transcription	124
iii. Translation	
(1) Gene 1.3 translational stop-site	126
(2) Translation of other genes	127
c. Physical features	
(1) Base composition	129
(2) Direct repeats	129
(3) Inverted repeats	129
4. Further studies	130

REFERENCES

INDEX OF FIGURES AND TABLES

FIGURE	FOLLOWS PAGE NO.	FIGURE	FOLLOWS PAGE NO.
1	6	24	101
2	13	25	101
3	24	26	101
4	28	27	101
5	34	28	102
6	38	29	110
7	57	30	110
8	75	31	110
9	75	32	114
10	80	33	118
11	81	34	125
12	84		
13	84		
14	85		
15	85		
16	87	TABLE	
17	88	1	46
18	90	2	46
19	92	3	46
20	92	4	60
21	95	5	75
22	95	6	97
23	96	7	98

CHAPTER 1 INTRODUCTION

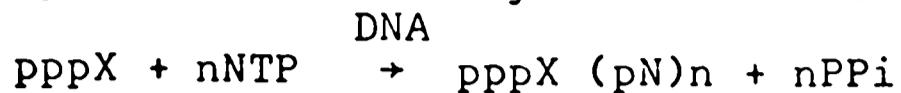
Of fundamental importance to the survival of any organism is an ability to respond rapidly and economically to changes in its local environment. For the gram-negative bacterium Escherichia coli such responses are often made through the regulation of gene expression at the level of transcription, for this is the first step in the pathway from DNA-encoded blueprint to active cellular constituent. As there is only a single major species of DNA-dependent RNA polymerase (Rpol) in E.coli, this enzyme must be able to respond both generally to changes in the overall metabolic state within the cell, and specifically to fluctuations in the levels of individual metabolites or other cellular constituents. As a result, E.coli Rpol is a necessarily complex protein, a complexity reflected in its physical and biochemical properties. This thesis will describe investigations which attempt to further our understanding of one such biochemical function, that of transcription termination.

1. E.coli RNA polymerase

The properties of this enzyme have been extensively dealt with in the book RNA Polymerase (ed. Losick and Chamberlin, 1976) and more recently reviewed by Lathe (1978).

E.coli Rpol can be isolated in two forms: core and holo enzyme (Burgess, 1976). Both appear to be active in the monomer form and both contain the polypeptide subunits

α , β and β' in the molar ratio 2:1:1. They differ in specificity of template and overall activity, and in that holoenzyme contains the additional subunit σ . They both catalyse the essentially irreversible reaction:



where X usually is a purine nucleoside and N is adenosine, guanosine, cytidine or uridine.

a. α subunit

This subunit, present at two copies per complete enzyme, has a molecular weight of 37,000 (Ovchinnikov et al, 1977).

Although no known function has yet been specifically ascribed to it, mutations in the gene coding for it (rpoA) can alter the specificity of transcription (Fujiki et al, 1976).

b. β subunit

This subunit, present at only one copy per active enzyme molecule, has a molecular weight of 145,000 (Lowe, Hager and Burgess, personal communication). Evidence has been obtained suggesting that it is responsible for substrate (NTP) binding (Armstrong et al, 1976), and template (DNA) binding in holo enzyme : promoter complexes (Hillel and Wu, 1978). Mutations conferring resistance to the drugs rifampicin and streptolydigin map in the gene (rpoB) coding for this subunit (Rabussay and Zillig, 1969; Iwakura et al, 1973).

c. β' subunit

This subunit is also present at only one copy per active enzyme molecule and has a molecular weight of 150,000 (Lowe

et al, personal communication). It is coded for by the rpoC gene and has been implicated in the primary binding of enzyme to DNA template (Fukuda and Ishihama, 1974).

d. σ subunit

This subunit, of molecular weight about 82,000 (Lowe et al, personal communication), is present only in holoenzyme molecules. It is present in the cell in less than stoichiometric amounts compared with the other Rpol subunits (Ishihama et al, 1976), but this is consistent with its more rapid turnover during transcription (Travers and Burgess, 1969). Although it cannot bind alone to DNA, it can be chemically cross-linked to DNA in holoenzyme: promoter complexes (Hillel and Wu, 1978), and it is functionally responsible for the recognition by holoenzyme of sequence-specific regions of DNA (Chamberlin, 1974). It is coded for by the rpoD gene.

The binding of the σ subunit to Rpol core enzyme induces a conformational change in the enzyme structure (Wu et al, 1976), with a concomitant change in the enzyme's functioning: i.e., 1) a decrease in the non-specific binding of Rpol to DNA, 2) an increase in the specific binding to promoter sites on DNA, 3) an increase in the rate of transcription initiation, 4) a decrease in the non-specific initiation of transcription at single-stranded breaks in DNA, and 5) the acquisition of the ability to melt specific regions of DNA to form "open" promoter complexes (Chamberlin, 1976).

e. Other subunits

One or more small (8000-12000 daltons), weakly associated

polypeptide species (ω) are sometimes found in highly purified holoenzyme preparations, but these do not appear essential to the full functioning or activity of the enzyme (Burgess, 1976).

Recently, however, Ishihama and Saitoh (1979) have reported a protein with a MW of 22500, which is stably associated with purified Rpol. This protein reduces the activity of Rpol holoenzyme on certain templates in vitro; moreover, it is apparently identical to SSP, the predominant protein synthesised under conditions of extreme starvation. Its function in vivo has not yet been determined, but it is likely to involve the regulation of transcription as part of the stringent response.

2. Transcription

In E.coli, genes which are likely to be expressed in response to a common stimulus are often grouped within a single transcriptional unit, the operon (Jacob and Monod, 1961). At the simplest level, the start of such an operon contains a regulatory region, or operator, where a variety of negative and/or positive effector molecules act to alter the frequency of initiations by Rpol at an adjacent or overlapping, region, the promoter. Transcription, once initiated, will then proceed to the end of the operon where a terminator region causes the Rpol to release the nascent RNA transcript and to dissociate from the DNA template. In fact, this picture is often considerably complicated by the existence of one or more additional operator/promoter regions at the start of or even within the operon. Furthermore, it appears that the efficiency of some terminators

is inherently low or can vary in response to a variety of positive and/or negative effectors, so that an operon need not be limited to only one terminator. Such regulatable terminators have been identified between genes and even before the first structural gene, in particular operons.

Together, these many regulatory sites enable the cell to finely tune its transcription so that each gene is expressed at the optimal rate, with an overall economy of effort. The mechanism and some examples of regulation of transcription will be discussed below.

a. Initiation

i) Promoter

The initiation of transcription by Rpol holoenzyme occurs almost exclusively at defined sites in the genome (Jacob and Monod, 1961). This specificity is wholly dependent on the association of the σ subunit with core enzyme (Chamberlin, 1974). The mechanism by which promoter-complexes form can be briefly summarised as 1) Rpol binds to DNA in a weak, non-specific reaction and then diffuses to the promoter region; 2) at the promoter, some primary, secondary or tertiary structure is recognised and a specific, "closed-complex" is formed between the Rpol and the DNA. (The stability of this "closed-complex" does not correlate with the strength (i.e. initiation frequency) of the promoter.); 3) a localised melting (10-15 bp) of the DNA occurs, resulting in the formation of a highly stable "open-complex" (Chamberlin, 1976).

This "open-complex" protects about 45 basepairs of DNA from nuclease attack (Pribnow, 1975). The transcription start site which, under artificial conditions, may be at any of 5 or 6 adjacent nucleotides (Hayden et al, 1975; Minkley and Pribnow, 1973), is roughly centred within this protected region. Once the open-complex has formed, initiation appears to be equally efficient in all promoters, suggesting that it is the transition from closed to open complex which determines the relative efficiency of a promoter (Dausse et al, 1976; Chamberlin, 1976).

In comparing the DNA sequence which is protected in open-complexes at several different promoters, Pribnow (1975) has observed a 7 bp sequence centred about 10 bp before (i.e. upstream of) the RNA start-site which is largely conserved in all known examples (Fig. 1). This sequence is A + T rich, a feature which may be important in permitting the localised melting and resulting formation of open-complexes at promoter sites. Mutations within this sequence can alter the relative efficiency of initiation at promoters (J. Gralla, cited in Gilbert, 1976; Musso et al, 1977).

A second region of partial homology has also been observed, centred about 35 bp before the RNA start-site (Fig. 1; Maniatis et al, 1975b). Although this lies outside the region protected in open-complexes, several promoter mutations have been mapped within it (Dickson et al, 1977). This region may play an important role in promoter recognition (i.e. closed-complex formation) as Rpol will

Figure 1. Nucleotide Sequences in Promoter Regions

The nucleotide sequences of DNA in 20 different promoters are presented for comparison. Only the codogenic or sense-strand is shown. The sequences are aligned to produce maximum homology in two regions, centred about 10 and 35 bp before the RNA start site. Accordingly, some sequences have been split between positions -22 and -23 or -23 and -24.

The underlined nucleotides indicate the RNA start-site(s) in each promoter.

The 'average' sequence is given at the bottom of the figure, along with the frequency of the 'average' nucleotide at each position. Only those nucleotides present in 10 or more cases are considered significant enough to include in the 'average' sequence.

This list was compiled by Siebenlist (1979). A complete list of references is included in that report.

Figure 1.

CAAACGGTTGACAACATGA	AGTAAACACGGTACGATGTACCACAT	T ₇ A3
AAAGAGTATTGACTTAAAGT	CTAACCTATAGGATACTTACAGCCAT	T ₇ A1
TGGCGGTGTTGACATAAATA	CCACTGGCGGTGATACTGAGCACATC	λp _L
CGTGCGTGTTGACTATTTTA	CCTCTGGCGGTGATAATGGTTGCATG	λp _R
TGCCGAAGTTGAGTATTTTT	GCTGTATTTGTCATAATGACTCCTGT	λp _O
TCTTTTTGATGCAATTCGCT	TTGCTTCTGACTATAATAGACAGGGT	fdX
CATTAACGTTTACAATTTAA	ATATTTGCTTATAACAATCATCCTGTT	fdII
AATTGTTGTTGTTAACTTGT	TTATTGCAGCTTATAATGGTTACAAA	SV40
CGTCAGGATTGACACCCTCC	CAATTGTATGTTTTTCATGCCTCCAAA	OXA
ATGAGCTGTTGACAATTAAT	CATCGAACTAGTTAACTAGTACGCAA	trp
ATCGAATGGCGCAAACCTT	TCGCGGTATGGCATGATAGCGCCCGG	lacI
TCTCATGTTTGACAGCTTAT	CATCGATAAGCTTTAATGCGGTAGTT	tet
AACAGGTATTGACAACATGAAGTAACATGCAGTAAGATACAAATCGC		T ₇ A2
GCACGGTGTTAGATATTTATCCCTTGCGGTGATAGATTTAACGTATG		λp _{rm}
TTCTCTTGTTGACATTTTAAAAGAGCGTGGATTACTATCTGAGTCCG		OXD
TAAATAGCTTGCAAATACGTGGCCTTATGGTTACAGTATGCCCATC		OXB
CCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGAAT		lacUV5
TTTATTCCATGTCACACTTTTCGCATCTTTGTTATGCTATGGTTATT		gal(-cap)
AATCTCCGTTGTACTTTGTT	TCGCGCTTGGTATAATCGCTGGGGT	fdVIII
CGTAACACTTTACAGCGGCG	CGTCATTTGATATGATGCGCCCGCT	Tyr tRNA

TGTTGACAATTT T

10 17 16 13 10 11 11
12 19 13 15 9 12

TG TATAATR

11 15 13 14 15
12 17 13 20

not rebind to the 45 bp protected fragment, once dissociated (Pribnow, 1975; Schaller et al, 1975). The shortest fragment so far isolated which is capable of specifically binding Rpol extends to about 60 bp before the RNA start-site (Gilbert, 1976).

There is often considerable overall homology between promoters with a similar origin or function (Gilbert, 1976; Lathe, 1978; Walz et al, 1976). This may be indicative of a common evolutionary origin and/or functional requirements.

Despite the observation that these two main regions of homology are, at best, only approximate (Fig. 1), the activity of a given promoter is acutely sensitive to its DNA sequence: a single basepair change may alter the efficiency of transcription initiation by an order of magnitude or more. The importance of this cannot be overstressed when making comparisons and predictions on the basis of sequence data alone.

ii) Operator

Adjacent to or overlapping the promoter, there can be a regulatory region wherein a variety of positive and/or negative factors can effect an alteration in the frequency of transcription initiation. This region is designated the operator (Jacob and Monod, 1961). Two examples each of negative and positive control are briefly discussed below.

(1) Negative Control

When the concentration of intracellular lactose is low, transcription of the lac operon is repressed because the

lac repressor (the product of the laci gene which lies outside the lac operon) is free to bind with very high specificity to a region overlapping the lac promoter, lacop, thus sterically preventing Rpol:promoter complex formation (Majors, 1975). At high concentrations of lactose, a repressor:lactose complex forms which has a much reduced affinity for lacop DNA, and thus binding of Rpol to the lac promoter is unobstructed.

In the case of the trp operon, and under conditions of excess intracellular L-tryptophan, the trp aporepressor (product of the trpR gene) binds L-tryptophan to form an active repressor complex (Fig. 5). This can then bind to a site overlapping the trp promoter, trpop and thus physically block the binding of Rpol to the trp promoter (Squires et al, 1975). In the absence of L-tryptophan, the affinity of this aporepressor for trpop is much reduced and thus Rpol:promoter complex formation is unhindered.

(2) Positive Control

Under conditions of a high intracellular concentration of cyclic AMP (cAMP) (e.g. in the absence of glucose), the pleiotropic positive effector, CRP (cAMP receptor protein), can bind cAMP to form an active CRP:cAMP complex. In the case of the lac operon, this complex can bind to a region centred around 60.5 bp before the RNA start site, and so stimulate by up to fifty-fold the frequency of transcription initiation (Beckwith et al, 1972). This region has a two-fold rotational symmetry; mutations which alter the efficiency of CRP-mediated promoter activation also affect this symmetry (Dickson et al, 1975; Gilbert, 1976).

When bacteriophage λ is in the prophage state, transcription of the repressor gene (λcI) is initiated at the promoter for repressor maintenance, λp_{rm} . The activity of this promoter is dependent on stimulation by repressor bound to λo_R1 (Ptashne et al, 1976), a regulatory site located about 65 basepairs before the start site for p_{rm} -initiated cI mRNA. This apparently autocatalytic system is controlled by the fact that at higher concentrations, repressor will bind to two additional regulatory regions, λo_R2 and λo_R3 , resulting in a decrease in the frequency of initiation at λp_{rm} (Walz et al, 1976). This negative control is probably effected by a steric exclusion mechanism similar to that already described for the trp and lac operons, as λo_R2 and λo_R3 overlap λp_{rm} .

The distance between λo_R1 and λp_{rm} is about the same as that between the lac CRP:cAMP binding site and lacp, suggesting that the positive control in both cases may be effected through a similar mechanism. Gilbert (1976) has noted that the distances involved are such that the core Rpol, the σ subunit and the positive effector molecule might be in physical contact with one another when all three are bound to the DNA. The activation signal therefore might be transmitted through a chain of allosteric, protein: protein interactions. Alternatively, Wells et al (1977) have discussed the possibility that binding of a positive effector at one site on the DNA molecule might markedly affect the stability of the double helix some distance away. If this effect were one of destabilisation, open-complex formation might be enhanced and thus the frequency of

transcription initiation could be increased (assuming that open-complex formation is the rate-limiting step in such initiation). Wells et al (1977) have termed this mechanism telestability".

b. Elongation

Soon after the initiation of transcription, the sigma subunit is released from the transcribing complex, and the elongation of the RNA transcript proceeds in a 5' to 3' direction, catalysed by the Rpol core alone (Krakow and von der Helm, 1971). This ternary elongation-complex protects about 26 basepairs of DNA from attack by nucleases (with a preference for the codogenic or sense strand) and about 22 nucleotides of the nascent RNA (Rohrer and Zillig, 1977). The rate of elongation varies with the template, but in vivo averages about 43 nucleotides/s (Manor et al, 1969). Rpol may, however, pause at specific sites on the DNA template in vitro (Darlix and Fromageot, 1972; Gilbert, 1976). Such pausing occurs at multiple sites in the lac operon when transcription is carried out under substrate-limiting conditions. The nucleotide sequences of several of the latter pausing sites have been determined (Maizels, 1973), as have the sequences of a number of mutations which decrease pausing at one such site (Gilbert et al, 1974). All these mutations represent G/C to A/T changes in a G + C rich region 7-10 nucleotides upstream of the pausing site. Since G + C rich RNA:DNA hybrids can be very stable (more so even than double-stranded DNA), Gilbert (1976) has suggested that pausing reflects the relatively long time needed to dissociate G + C rich nascent RNA from template

DNA. This suggests that within the elongating complex, the expected RNA:DNA hybrid region extends about 10 bp back from the site of elongation. The relevance of this to transcription termination will be discussed below.

3. Termination

The ends of operons are defined by terminator regions wherein Rpol, possibly in conjunction with one or more accessory factors, recognises an as yet ill-defined signal and terminates transcription. Termination consists essentially of three steps: the elongation process is halted, the RNA transcript is released, and the Rpol:DNA complex dissociates. The relative order of these last two steps is not yet certain but with T7 t_1 in vitro, at least, the given order appears to be correct (O'Hare, 1978).

Termination sites, however, are not confined to the ends of operons. In recent years it has become increasingly clear that regulatable terminators occur within some operons, and that these can play an important part in the regulation of gene expression (for reviews see Lathe, 1978; Adhya and Gottesman, 1978; and Roberts, 1976). Although in vivo such terminators may be positively and/or negatively regulated, the examples so far studied in vitro appear to be fully active in terminating transcription by purified Rpol combined, where necessary, with the termination factor rho, but in the absence of any operon-specific factors. This is also true of termination at the ends of operons. Hence the termination signal must be encoded within the nucleotide sequence of the terminator region. Comparative studies of different

terminators (ideally including mutant terminators) might reveal the crucial features of these nucleotide sequences. I shall discuss below what such studies have so far indicated about the mechanism of termination, and then consider two examples of its in vivo regulation.

a. Mechanism of termination in vitro

Terminators have been divided into two classes according to whether or not they are dependent on rho factor for activity in vitro. It should, however, be noted that the current terminology is somewhat misleading, as many so-called rho-independent terminators are dependent on rho for activity in vivo (e.g. trpa, Korn and Yanofsky, 1976; trpt, Wu and Platt, 1978) and even their in vitro activity is in some cases altered by rho. This is perhaps most striking in the case of the λ 4S or oop RNA, where addition of rho in vitro causes a shift in the RNA stop-site by one base-pair downstream, as well as a five-fold increase in the yield of transcript (Rosenberg et al, 1975). Adhya and Gottesman (1978) have suggested that terminator complexes are normally slow to dissociate from λ 4S DNA in vitro, and that such complexes can effectively block initiations and/or elongations: rho acts by stimulating dissociation of the terminator complex, and so permits more rapid reinitiations. Nevertheless, the distinction between terminators according to dependence on rho in vitro, however qualified, is real and hence these two classes will be treated separately.

i. Rho-Independent Terminators

There are three notable features present to some extent in all eight rho-independent terminators so far sequenced

(Figure 2a): a) termination in all cases occurs immediately following a run of 4-8 consecutive T:A base-pairs, which are usually part of a larger, A + T rich sequence; 2) immediately preceding this sequence is a G + C rich region of 4-11 bp; 3) this G + C rich region, and in some cases a portion of the A + T rich region, represents part of a two-fold rotational symmetry. There is no detectable homology downstream of the RNA stop-sites.

Gilbert (1976) has incorporated the first two of these features into a preliminary model for the mechanism of termination. As already discussed in regard to elongation, he suggests that pausing can occur shortly after transcription of a G + C rich sequence, because of the slow dissociation of the relatively stable, G + C rich nascent RNA:DNA hybrid: such dissociation is assumed to be required some ten bp upstream of the elongation site as a prerequisite for translocation. If this pausing occurs while a run of T:A base-pairs is being transcribed, then according to Gilbert's model, termination will result because of the relative instability of the A + T rich RNA:DNA hybrid within or near the elongation site. Presumably once this portion of the nascent RNA:DNA hybrid has dissociated, the G + C rich hybrid upstream is somehow destabilised to allow complete release of the transcript perhaps through some conformational change in the polymerase or through telestability effects in the polynucleotide such as proposed by Wells et al, 1977. It is important to note that this model assumes the existence of an RNA:DNA hybrid (some 10 bp in length) within the transcribing complex, an assumption for which there is no directly supportive evidence.

Figure 2. Nucleotide Sequences in the Region of Transcription Termination Sites

a. Rho-Independent

The nucleotide sequence of DNA surrounding eight rho-independent terminators are presented for comparison. Only the codogenic, or sense strand is shown.

The sequences have been aligned according to the transition point from G + C rich to A + T rich (indicated by the vertical dotted line), which lies just upstream of the RNA stop-site. Nucleotides which form part of a region of two-fold rotational symmetry are underlined. The centre of symmetry in each case is indicated by o.

Where known, the RNA stop-sites observed in vivo are indicated by Δ above the appropriate nucleotide; and the in vitro sites by ∇ below. Large triangles indicate the preferred stop-sites.

These sequences were reported by: fd coat, Sugimoto et al (1977); Ø80 M3, Pieczenik et al (1972); λ4S, Schwartz et al (1978); λ6S, Sklar et al (1975); ØX174-H, Sanger et al (1977); trpt, Wu and Platt (1978); pha, Zurawski et al (1978); trpa, Lee et al (1976); Bertrand et al (1977).

b. Rho-Dependent

The nucleotide sequences of DNA, in the regions surrounding 2 terminators, which are wholly dependent on rho for activity in vitro, are presented for comparison. Only the codogenic or sense strand is shown.

The sequences have been aligned on the CAATCAA homology which lies just upstream of the RNA stop-site.

Details are as in a.

The tRNA₁^{tyr} sequence was reported by Küpper et al (1978) and the λt_R1 sequence by Rosenberg et al (1978).

Figure 2a.

fd coat 5' . . . CAAGCTGATAAACCGATACAAATTAAAGGCTCCTTTTGGAGCCCTTTTTC^Δ
Ø80 M3 ACAGTGTGATAAAGGCCCTTGAAGTTA^ΔCCGCTTCAAGGGCTTTTAA^{ΔΔ}
λ 4S CTCCATCTGGATTGTTCAGAACGCTCGGTTGCCCGCGGCTTTTATTGGTGAGAAATCGCAGCAA^Δ
λ 6S CAGTTCGCAGGTAATAQT^ΔTAGAGCCTGCATAACGGTTTCGGGATTTTATTATATCTGC^Δ
ØX174 H CCTCCCAAATTGTATGTTTTCATGCCCTCCAAATCTTGGAGGCTTTTATTATGGTTCGTTCTATTACCC^Δ
trp t CGACGGCAGTTAATCCACAGCCGCCAGTTCCGCTGGCGGCATTTTAACTTTCTTAAATGAAGCCGG^Δ
phe a AATGCCAAGACGAACAATAAAGCCCTCCCAAATCGGGGGCCTTTTATTGATAACAATAAGGCAAC^Δ
trp a TGCGTAAAGCAATCAGATACCCAGCCGCCCTAATGAGCGGGCTTTT^{ΔΔ}TTTGAACAATAATAGAGAATA^Δ

Figure 2b.

tRNA^{tyr}₁ 5' . . . CACCATCACTTCAAAAGTCCCTGAACTCTCAAGCGAATCCGCAATCAAATA^ΔTATTCTGCCCATGCGGGG^Δ
λ t_R1 ATTAACACACCTATGGTGTATGCATTTATT^ΔTTCATACATTCAATCA^{ΔΔ}ATTGTTATCTAAGGAAATAC^Δ

McMahon and Tinoco (1978) have attempted to quantify Gilbert's model by assessing the average terminator sequence in terms of the free energy of formation of each of the nearest neighbour base-pairs. From this they have deduced that an essential feature of terminators may be a most stable nearest neighbour term (i.e. GpG or CpC) followed by two variable terms, and then a series of seven weak terms, such that at least five of the seven are of the weakest possible sort (i.e. ApA or TpT). The most efficient terminator would therefore contain $(pG)_2$ or $(pC)_2pN(pA)_8$ or $(pT)_8$.

Adhya and Gottesman (1978) have proposed that in addition to the above features (i.e. G + C rich followed by A + T rich), the region of two-fold rotational symmetry may also play an important part in termination. They suggest that a stable, base-paired stem-loop structure of DNA and/or nascent RNA forms in such regions, and that this in some way interacts with the transcribing Rpol to prevent further elongation. If this interaction is sufficiently strong, and if the region also has the two features required by Gilbert's model then termination will result.

This 3-feature model is supported by the totality of the following results:

(i) Neff and Chamberlin (1978) have demonstrated that the efficiency of termination at T7 t_1 , in vitro, is reduced when GTP is replaced by its analogue ITP in the transcription reaction. ITP is a substrate for Rpol, but I:C- or I:dC- containing double helices are less stable

than those containing G:C or G:dC. This is consistent with the model which predicts that pausing and thus termination would be less likely to occur under conditions where the putative RNA:DNA hybrid, and/or RNA:RNA base-paired stem-loop structures, are less stable.

(ii) An analysis of the complete nucleotide sequence of bacteriophage ØX174 by McMahon and Tinoco (1978) revealed only five sequences which should be active terminators according to their modification of Gilbert's hypothesis (which does not include as a requisite feature a region of two-fold rotational symmetry). The distances from promoter A to these five putative terminators, coupled with their predicted relative efficiencies (calculated from the closeness of fit to the optimal sequence, see above) correlate well with the size and relative molar proportions of the transcripts known to be initiated at pA in vivo. (Interestingly only the most efficient of these sites has a region of significant two-fold rotational symmetry preceding the RNA stop-site.)

(iii) Sugimoto et al (1977) have observed a particularly interesting phenomenon which occurs upon transcription of a restriction fragment of bacteriophage fd (RF) DNA. This fragment contains a promoter for transcription, but no terminator. Rpol can therefore initiate transcription at a specific site, and might then be expected to continue elongation until it reaches the end of the fragment. The nucleotide sequence of the resulting transcript, however, suggests that in fact the polymerase somehow "switches" DNA strands when it reaches the end of

the fragment and continues transcription until a G + C rich, 11 base-pair inverted repeat has been transcribed. Termination then occurs. This is consistent with the suggestion that a stem-loop structure in the product can trigger termination: Rpol, which could otherwise synthesise a very long inverted repeat, in fact stops when the resulting putative stem-loop structure reaches a critical size or stability. This phenomenon, although perhaps quite complex, is most intriguing, and would undoubtedly repay further study.

(iv) Transcripts initiated in vitro at the ØX174 promoter A (pA) apparently read through a termination site located 20 bp downstream, which is active for transcripts initiated upstream of pA (Sanger et al, 1977). These investigators have pointed out that whereas in the latter case a stable, 8 bp stem-loop might form in the transcript, initiation at pA occurs within the relevant region of two-fold rotational symmetry, so that no equivalent stem-loop structure could form in the short (pA-initiated) transcripts. It is particularly significant that pA-initiated transcripts do contain the G + C rich followed by A + T rich sequences suggested by Gilbert (1976) and McMahon and Tinoco (1978) to be enough for efficient termination. It seems likely that the stem-loop is an essential feature of at least this terminator.

(v) Although the efficiency of termination at trpa in vivo is regulated in response to intracellular concentrations of L-tryptophan (see below), in vitro it is about 95% efficient with purified Rpol in the absence of

rho or any other protein factors (Lee et al, 1976). Stauffer et al (1978) have isolated 5 point mutations which decrease the efficiency of termination at trpa in vitro by 9- to 15-fold (Fig. 4a). All of the changes occur in the G + C rich inverted repeat which precedes the RNA stop-site; they also reduce the predicted stability of the putative stem-loop structure by about 50%. Moreover, three of the mutations occur in the upstream portion of the inverted repeat, and thus do not alter the G + C rich sequence which immediately precedes the run of T:A base-pairs. This again points to the importance of a stable stem-loop at least for termination at trpa.

(vi) Bertrand et al (1977) have isolated three mutants containing deletions which extend to within or near the trpa region (Fig. 4a,b). One of these deletions, trpΔ147, extends from within trpC (downstream of trpa but within the same operon) to between positions 152 and 146 in trpL: that is, to within 5-11 bp downstream of the RNA stop-site. It has no apparent effect on the efficiency of termination in vitro. This suggests that the sequence 5-11 bp downstream of the stop-site is not important for terminator activity in vitro. However, the significance of this result is severely limited by the lack of information on the sequence in trpC which is fused to the trpa region in trpΔ147.

The second deletion studied, trpΔ142, extends from within trpD (located between trpa and trpC) to between positions 125 and 120, that is, to about 16-20 bp upstream of the stop-site. The deleted region therefore includes the G + C rich followed by A + T rich sequences, as well as

a part of the inverted repeat implicated in the terminator function. As predicted, transcription of $\Delta 142$ DNA in vitro is not terminated in the leader region, though again the significance of this result is somewhat uncertain.

The third of these deletions, trp Δ 1419, extends from within trpC to position 138 in trpL. This deletion is known to replace the sequence distal to 138 (i.e. including the last four of the 8 consecutive T:A^{bp} in the stop-region) with a sequence from trpC which commences with a G:C bp. Transcription of $\Delta 1419$ DNA in vitro is not terminated at trpa. This result indicates that using purified Rpol and in the absence of other factors, a run of more than four T:A bp in the RNA stop-region is essential for trpa activity. This result was not significantly altered by addition of rho to the transcription reaction.

(vii) A region within the structural gene for λ cro contains a sequence reminiscent of the average terminator described above (Rosenberg et al, 1978), namely:

5'...AAGGCCATTCATGCAGCCGAAAGATTTTTTAA...3'

In vitro transcription studies described in the same report, however, indicate that Rpol does not terminate transcription at this site. Two possibly significant differences between this sequence and that of the average terminator do, however, exist, namely: 1) the predicted free energy of formation of the putative G + C rich stem-loop is only -10 kcal/mol, compared with -19.6 for the weakest structure suggested in Fig. 2a (i.e. for trpa); and 2) the run of 7 T:A base-pairs is separated from the G + C rich sequence by 5 bp. Although four of these are A:T, the fifth is an intervening G:C, so

that only 4 out of the 7 nearest neighbour base-pairs which follow the G + C rich region are "least stable" as defined by McMahon and Tinoco (1978). This is below the minimum of 5 suggested by the latter authors as being necessary for terminator function.

Together with the mutation analyses described for trpa, this result indicates that the signal for termination is rather precise. Accordingly, any model which attempts to explain termination must have a degree of precision incorporated within it. Such exactitude is easily (albeit inelegantly) accounted for in a model which supposes that a terminator is recognised by virtue of its specific **nucleotide sequence (possibly irrespective of the particular physical or structural considerations discussed by Gilbert (1976) and Adhya and Gottesman (1978)), in a manner analogous to that** ascribed to type II restriction endonucleases (Roberts, 1978). (This possibility is further explored when discussing the results presented in this thesis; Chapter 7.) An important functional distinction between these two alternatives (which, it must be emphasised, are in no way mutually exclusive) is that the physical/energetic interpretations predict that termination at these sites should be roughly comparable using any RNA polymerase, whereas the specific sequence interpretation predicts that a totally unrelated Rpol might not recognise the termination signal. To test this distinction it might be interesting to investigate the efficiency of termination in vitro by T7 Rpol (a superficially simpler enzyme than E.coli Rpol) at one of the terminators shown in Fig. 2a.

Finally, it is relevant to note that termination, at least in vitro, is not an exactly reproducible event: the efficiency seems never to be 100%, and the stop-site in at least some cases varies within 2 or 3 base-pairs (e.g. trpa; Lee et al, 1976). There are also reports of one to five additional A residues (not encoded by the terminator DNA) being incorporated into the 3' end of transcripts at or immediately following termination in vivo and in vitro (Kramer et al, 1974; Rosenberg et al, 1975).

ii. Rho-Dependent Terminators

The protein factor rho (MW about 50000) is the product of the rho gene, previously designated suA (Richardson et al, 1975), nitA (Inoko and Imai, 1976) or psu (Korn and Yanofsky, 1976). It has two major activities in vitro: stimulation of transcription termination (Roberts, 1969), and nucleoside triphosphate phosphohydrolase (NTPase; Lowery-Goldhammer and Richardson, 1974). Only partial rho mutants have so far been isolated so that evidently some rho functioning is essential to the cell's survival.

The stimulation of transcription termination by rho is non-stoichiometric with respect to Rpol; moreover, rho can be added after the initiation of transcription and still catalyse termination, so that its activity is evidently not dependent on an interaction with Rpol holoenzyme (Goldberg and Hurwitz, 1972). It will not catalyse termination by T7 Rpol, suggesting that its interaction with E.coli Rpol is specific. Alternatively, the DNA template for T7 Rpol may lack a sequence necessary for rho action.

The nucleotide sequences of two terminators wholly

dependent on rho for activity in vivo and in vitro have so far been determined (Fig. 2b). Although neither of them possesses all three features found in rho-independent terminators, there are some similarities between the two classes.

The rho-dependent RNA stop-site for transcription of the gene coding for tRNA₁^{tyr} in vitro is at one of 5 base-pairs within an A + T rich region (Küpper et al, 1978). Although this follows a G + C rich region, only 3 out of the 7 nearest neighbour base-pairs are of the weakest type and there is no two-fold rotational symmetry present in the terminator region. Hence termination would not be predicted according to the three-feature model described above.

In the case of λt_{R1} the RNA stop-site lies within an A + T rich sequence, but this does not follow a G + C rich region, and neither is there a run of 5 out of 7 weakest nearest neighbour terms (Rosenberg et al, 1978). However, there is a region of two-fold rotational symmetry near the RNA stop-site, so that a stable stem-loop structure may be able to form within the nascent transcript and/or the DNA strands.

The partial resemblance of these two sites to rho-independent terminator sequences is consistent with the suggestion that rho stimulates termination at natural but weak termination sites (Rosenberg et al, 1978). This is further supported by: 1) the observation that in the absence of rho, purified Rpol can pause at λt_{R1} for about 60 seconds before transcribing the DNA beyond the terminator

(Rosenberg et al, 1978); and 2) the isolation of an rpoB mutant which restores termination at rho-dependent sites in vivo (e.g. at trpt) in all seven different rho mutants so far tested, suggesting that the mutant polymerase recognises these sites independently of rho (Guarente and Beckwith, 1978). It will be interesting to learn whether purified Rpol from this rpoB mutant strain is able to recognise λt_R1 or the terminator for tRNA₁^{tyr} in vitro in the absence of rho.

Interestingly, these two sites contain the heptamer sequence CAATCAA near the RNA stop-site. This may be a part of the specific signal for rho-dependent termination. In the case of tRNA₁^{tyr}, the entire sequence shown in Fig. 2b differs in only five places from a similar sequence which precedes it in the same transcript by about 180 bp (Küpper et al, 1978). Termination, however, does not occur in this upstream sequence. One of the differences is a C to T transition giving CAATTAA instead of CAATCAA, consistent with the suggestion that the latter sequence is an essential part of the termination signal.

A mutant in λt_R1 which increases the efficiency of termination (cin1) also increases the extent of the two-fold rotational symmetry in the terminator region (Rosenberg et al, 1978). Mutations which decrease the efficiency of termination at λt_R1 (cnc1 and cnc8) reduce this symmetry. Hence, although CAATCAA may represent a part of the termination signal, it is not sufficient for termination at λt_R1 at least.

Adhya and Gottesman (1978) have proposed a model for

rho action. They suggest that rho binds to an exposed region of the nascent transcript, and then migrates rapidly along the RNA (using NTP hydrolysis as a source of energy) toward an Rpol paused at a rho-dependent termination site. Rho then interacts with the transcription complex (i.e. Rpol and/or a specific nucleotide sequence), and somehow stimulates termination. This model is supported by the finding that the NTPase activity of rho is strictly dependent in vitro on the presence of RNA. This stimulatory effect is greatest with synthetic polynucleotides which are long and which contain at least 1 in 20 C-residues, but which do not contain any G (Lowery and Richardson, 1977). It is noteworthy that there are only 5 G-residues in the 50 nucleotides preceding the 3'-terminus of the λt_{R1} -terminated and the $tRNA_1^{tyr}$ transcripts (Fig. 2b). Furthermore, these regions of the transcripts would not normally /be sequestered by ribosomes in vivo: the cro gene's stop codon is about 90 nucleotides upstream of λt_{R1} (Rosenberg et al, 1978) and the $tRNA_1^{tyr}$ gene is presumably not translated.

The interaction which causes termination may be between Rpol and rho. This suggestion is supported by the finding that one rpoB mutation suppresses a given set of rho mutations, while a different rpoB mutation may suppress an overlapping or even a completely distinct set of rho mutations (Das et al, 1978; Guarente, 1979). Although some of these effects might be explained without postulating a direct interaction (e.g. the effect of a slowly migrating rho might be suppressed by a similarly retarded Rpol), it remains plausible that a direct protein:protein interaction normally occurs between Rpol and rho.

The above model for rho action appears to explain mutational polarity wherein a nonsense mutation in one gene affects the expression of other, distal genes. Adhya and Gottesman (1978) suggest that in this case the stop-codon causes the release of ribosomes, thus exposing the normally translated mRNA. This allows rho to bind to the RNA and effect termination at a normally inactive rho-dependent terminator downstream of the mutation.

Similarly this model may explain polarity due to insertion of an IS1 element as this contains stop codons in all three frames in either orientation (Ohtsubo and Ohtsubo, 1978). If a rho-dependent terminator lay between the inserted IS1 and the next translation start site, termination would occur. Polarity due to IS2 insertion is more direct, for it actually contains a rho-dependent terminator (active in only one orientation: Starlinger and Saedler, 1976).

b. Regulation of termination in vivo

i. Antitermination mediated by the N gene product of Bacteriophage λ

The N gene of bacteriophage λ is the first gene of the major leftward operon (Fig. 3: Hershey, 1971). Upon infection or prophage induction this gene is transcribed from p_L , resulting in a protein product of MW \sim 13,500 (Shaw et al, 1978). This protein (gpN) antagonises termination at specific sites on the λ genome (Salstrom and Szybalski, 1978), so that Rpol reads through into distal genes. The antitermination effect is generally phage specific:

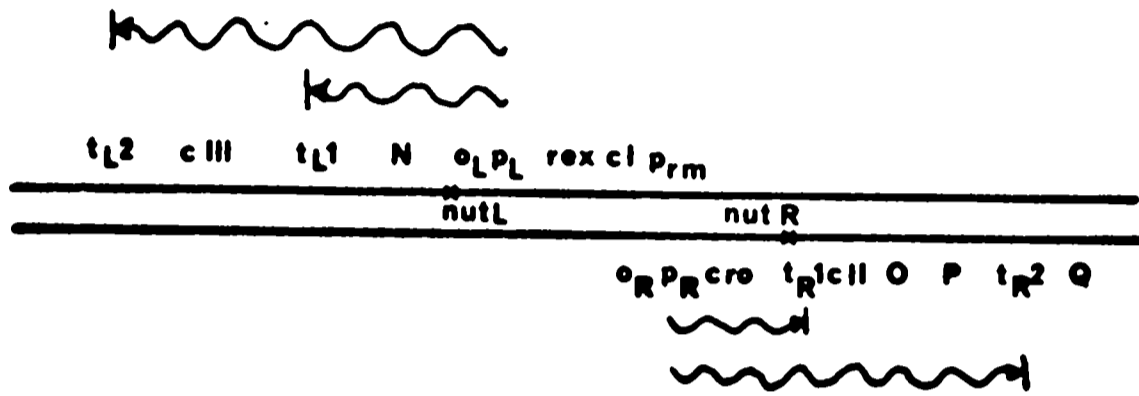
Figure 3. Aspects of the Regulation of Transcription in Bacteriophage λ Mediated by the N Gene Product (gpN).

- a. Transcription from p_L and p_R under conditions of low or no N gene expression (e.g. during establishment of lysogeny, or in a λN^- strain). In this case transcripts reaching t_L1 and t_R1 (and as a result of read-through, t_L2 and t_R2) will be terminated with a fairly high efficiency. In the case of t_L2 , this termination is apparently rho-independent.

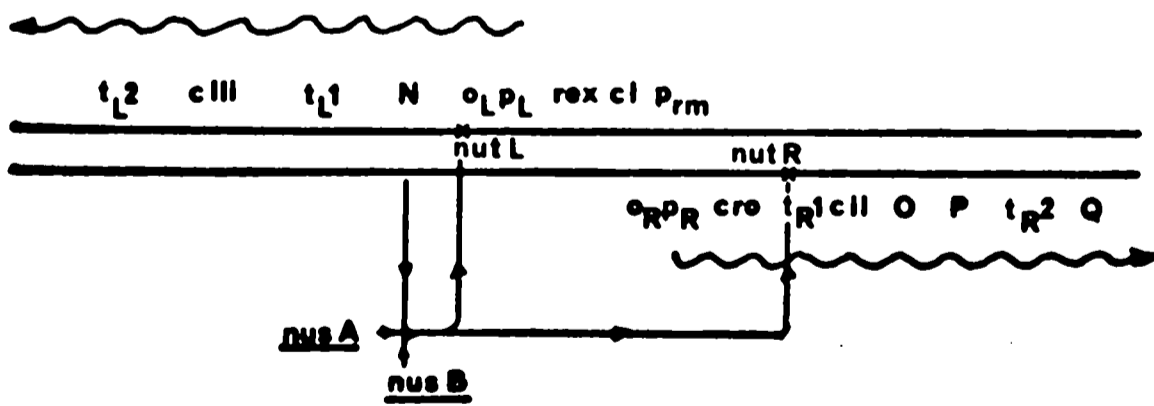
- b. Transcription from p_L and p_R under conditions of active N gene expression (e.g. during lytic growth of a λN^+ strain). In this case gpN antagonises termination at the four terminators shown (at least), giving rise to efficient readthrough and distal gene expression. This activity is affected by mutations in the host genes nusA and nusB.

Figure 3.

a.



b.



for example, the N gene of λi^{21} will not complement λN^- mutants (Friedman et al, 1973; λi^{21} contains the immunity region, including the N gene, of phage 21). This suggests that there must be some interaction with a specific site on the homologous phage genome. By studying deletions, two such N-utilisation sites have been mapped to just left of p_L (nutL) and just to the right of the cro gene (nutR) (Adhya et al, 1974; Salstrom and Szybalski, 1976). The nucleotide sequences of these two regions have recently been determined (Rosenberg et al, 1978). The results show that a 17 bp sequence in nutL is almost exactly reproduced in nutR (16 out of 17 bp are identical), and that this repeated sequence has a two-fold rotational symmetry. That this sequence is indeed a part of the functional nut site, at least in the case of nutL, is demonstrated by the finding that a nutL⁻ phenotype can apparently result from a single bp change within the 17 bp sequence (unpublished results cited by Rosenberg et al, 1978).

The likeliness of a direct functional interaction between Rpol and gpN has been demonstrated genetically (Franklin, 1974) and biochemically (Epp and Pearson, 1976). Once this interaction has occurred, at a specific nut site, the antitermination effect can be observed at nearly all terminators encountered downstream by the same Rpol. This has been demonstrated, for example, at trpt in λ /trp fusions (Segawa and Imamoto, 1976), and for a variety of polar mutations (Adhya et al, 1974), whose effects were suppressed provided that transcription initiation had occurred at either λp_L or λp_R .

The antitermination activity of purified gpN has not yet been demonstrated in vitro. This may be related to the fact that gpN action in vivo is apparently dependent on at least two host genes, nusA (for N-utilising substance) and nusB (also termed gro), which are clearly distinct from any of the known Rpol or rho structural genes (Friedman et al, 1976).

The observation that rho mutants can suppress the λN^- phenotype indicates that there are no essential genes in bacteriophage λ which are separated from their promoters by fully rho-independent terminators, and which are dependent on gpN-antagonism of termination for expression (Das et al, 1976). However, this does not mean that the antiterminating activity of gpN is effective only with rho-dependent terminators. Since the efficiency of termination at t_L2 is apparently not decreased when a rho ts 15 mutant is shifted to a non-permissive temperature, t_L2 may be a truly rho-independent terminator in vivo (Adhya and Gottesman, 1978; rho ts 15 relieves termination at all rho-dependent terminators so far tested). If this is so, and since gpN antagonises termination at t_L2 (Salstrom and Szybalski, 1978), gpN cannot be simply an anti-rho factor, but rather it must mediate antitermination in some other way.

This conclusion is further supported by the finding that there is a terminator to the left of the b2 region of λ which is rho-dependent but whose efficiency is not affected by gpN when transcription is initiated at p_L (S. Adhya, unpublished results, cited in Adhya et al, 1976).

On the basis of these results, Adhya and Gottesman (1978) have proposed three possible mechanisms for gpN action:

1) gpN interacts with Rpol at a nut site and causes a conformational change in the Rpol, so that specific terminators are not recognised. (Whether this alteration in the transcribing complex, once effected, is maintained throughout the remainder of the transcription cycle in the absence of a continuous association of Rpol with gpN, is not at present known.)

2) gpN antagonises ribosome discharge so that rho entry (or its access to Rpol) is blocked, and hence rho-dependent terminators are not recognised. This does not explain the apparent ability of gpN to antagonise termination at t_L^2 , though there may be transcription:translation coupling even at rho-independent terminators.

3) gpN could bind to nut sites in RNA and affect the structure of the nascent RNA so as to prevent termination, possibly by causing an interaction between RNA and some other cellular component(s) (nusA or B?).

The absence of a workable in vitro system for investigating the activity of gpN has meant that there is little information to help decide between these three alternatives.

The product of the λQ gene probably also has an anti-terminating activity (Roberts, 1975). As the λQ^- phenotype cannot be suppressed by rho mutants (Das et al, 1976), it may be that Q acts at terminators which are essentially rho-independent. As with all such studies, however, it must be emphasised that rho mutants seem to be necessarily

leaky (presumably rho⁻ is a lethal phenotype). Hence residual rho activity may be sufficient to effect termination at sites normally antagonised by gpQ.

Finally, it is interesting to note that assuming gpN is not excessively stable in vivo, any reduction in the frequency of initiations at p_L will have a non-linear effect on the expression of genes distal to N. This may be important in the establishment of lysogeny.

ii. Attenuation in the leader region of the trp, phe and his operons.

Between the trpop complex and the start of the first structural gene (trpE), there is a leader region (trpL) which, though transcribed, codes for no known protein product (Fig. 4a; Bertrand et al, 1976). However, mutants having deletions extending into this region (and studied in a trpR⁻ background to exclude any possible effects of the operator control system) exhibit an increased expression of the trp operon and are unresponsive to changes in the intracellular levels of L-tryptophan (Trp) whereas trpL⁺ strains do respond (Jackson and Yanofsky, 1973). This was found to reflect the deletion of a regulatable terminator of transcription, now termed the attenuator, or trpa, located at the operator-distal end of trpL. The activity of trpa apparently varies in response to the intracellular concentration of Trp or more precisely to the relative concentration of charged vs. uncharged tRNA^{trp} (Morse and Morse, 1976).

Nucleotide sequencing studies have shown that transcription termination at trpa both in vivo (Bertrand et al,

Figure 4. E.coli trp operon

a. Genetic Map and Deletions (not to scale).

The operator/promoter (op), attenuator (a), and distal terminator (t) are shown below the line; the leader region (L) and the five structural genes (E, D, C, B and A) are shown above the line.

The regions deleted in the mutants trp Δ 147, trp Δ 142 and trp Δ 1419 are indicated below the line.

b. trpL Sequence and Mutations

The nucleotide sequence of the sense strand of DNA in the trpL region is shown. Positions are given in base-pairs from the initiation point for trpL transcription.

The mutations indicated at positions 116, 117, 118 and 132 produce a decreased efficiency of transcription termination at trpA (Stauffer et al, 1978). Those with the changes indicated at positions 29 and 75 are the trpL29 and trpL75 mutants of Zurawski et al (1978a) and have an increased efficiency of termination.

The three deletion mutants are as described in the text (Bertrand et al, 1977). The dashed lines indicate uncertainty as to the exact end point of the deletions.

Arrows pointing down between positions 136 through 141 indicate the site and relative frequency (length of the arrow) of termination in vivo.

Regions of two-fold rotational symmetry are indicated by lines beneath the sequence, with the respective centres of symmetry indicated by dots.

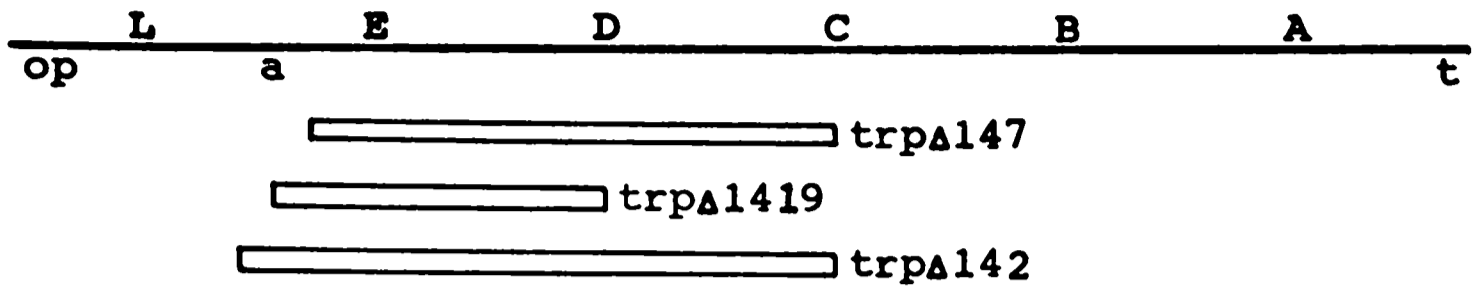
c. trpL Secondary Structure

Two possible stem/loop structures which might form in the trpL transcript are shown. The tandem Trp codons and the stop codon of the putative leader polypeptide are indicated by thin underlining. The complementarity between the Trp codons and the first loop is indicated by thick underlining.

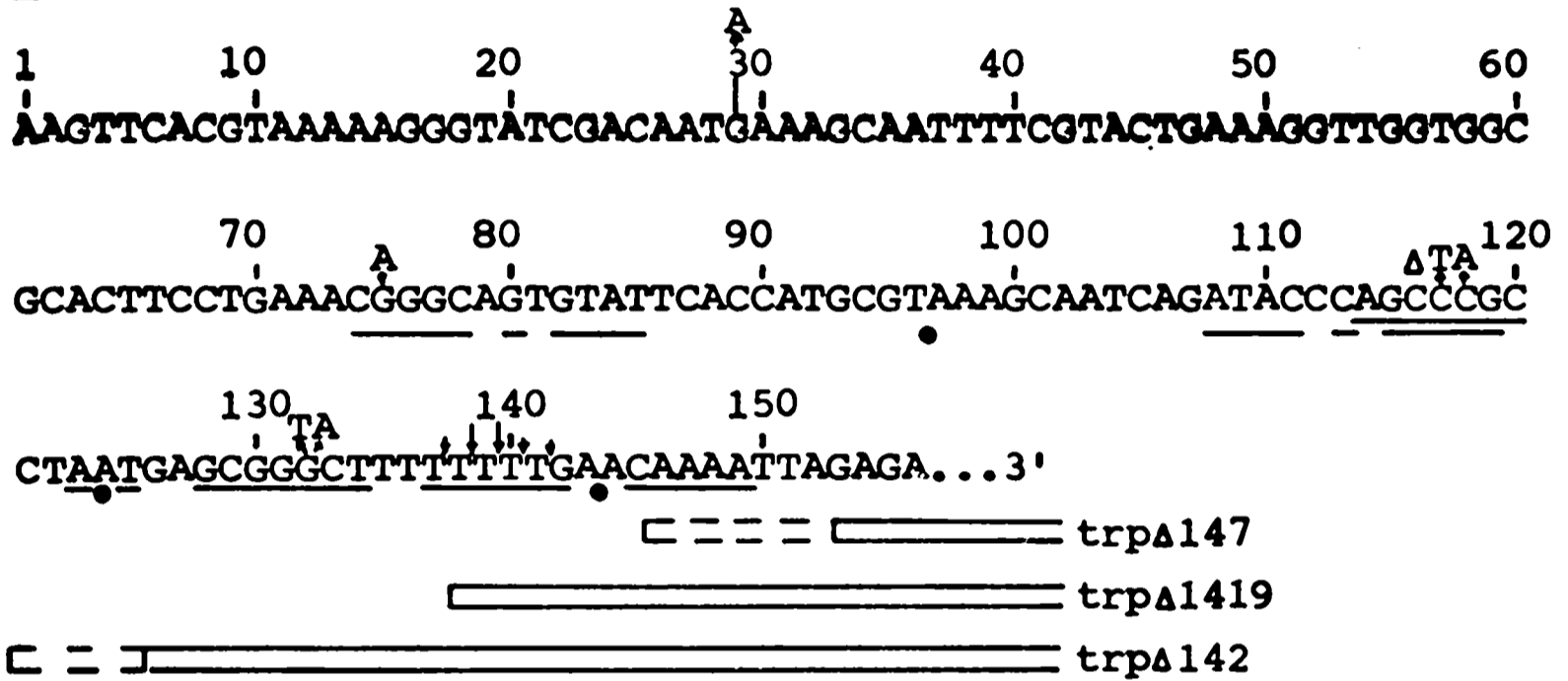
Positions are as given in b.

Figure 4.

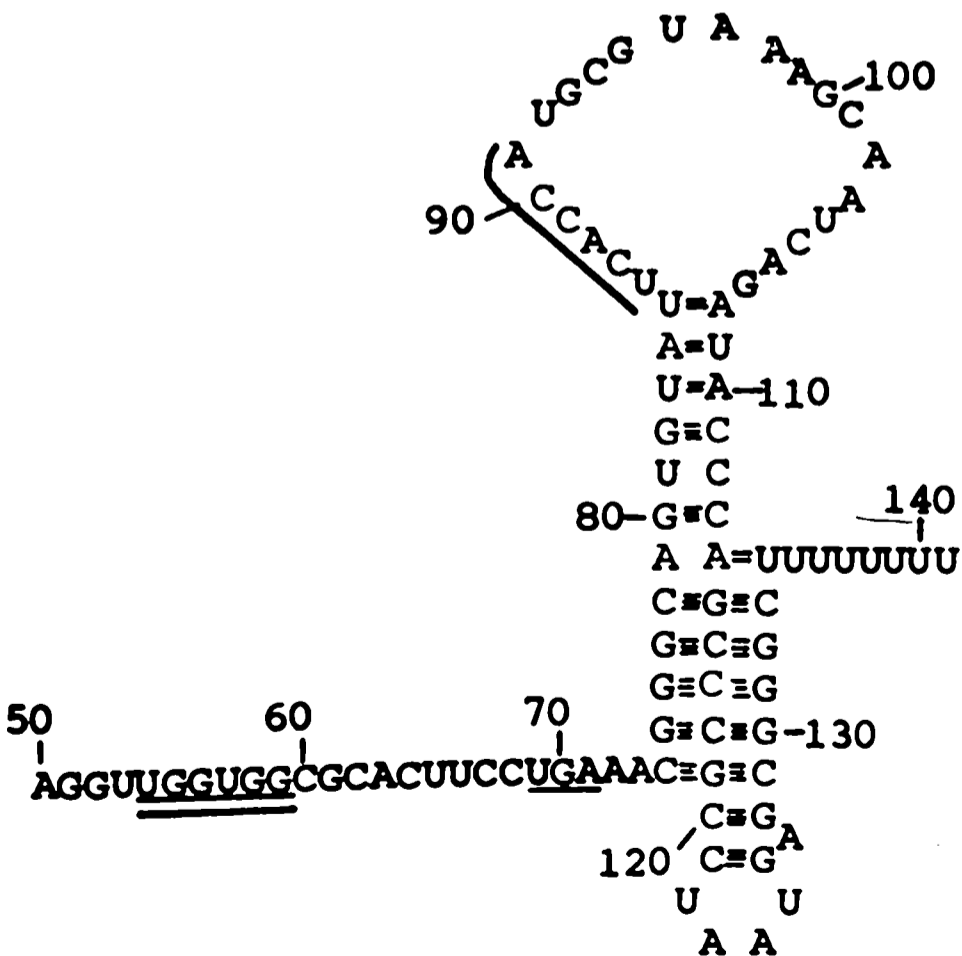
a.



b.



c.



and in vitro (Lee et al, 1976) occurs about 140 base-pairs from the initiation site for trp mRNA in a run of 8 consecutive T:A base-pairs which follow a G + C rich region (Figure 4b). This G + C rich sequence lies in a region of 2-fold rotational symmetry corresponding to the operator-distal ("second") stem-loop structure discussed below. (Although the T + A rich sequence also lies in a region of 2-fold rotational symmetry, it could not form a stable stem-loop in the trpL transcript, because only half the symmetric region would be present.) The above sequence organisation is typical of terminators which are rho-independent in vitro (Fig. 2).

Nucleotide sequencing studies have also revealed that the trpL transcript might code for a short polypeptide (Lee and Yanofsky, 1977). It has also been shown that trpL RNA can bind ribosomes in vitro (Platt et al, 1976) and that these are capable of initiating translation in vivo in trp deletions (i.e. trpLE fusions; Miozzari and Yanofsky, 1978a) and in trp/lac fusions (H. Somner, cited in Schmeissner et al, 1977). However, no leader polypeptide encoded by the trpL region of wild type strains has yet been isolated. It may be that RNA secondary structure normally prevents trpL translation, or that the protein product is unstable and/or produced in very small amounts. Such a polypeptide, if translated, would be 14 amino acids long, with tandem Trp residues at positions 10 and 11. Moreover, two overlapping regions of the trpL transcript can be drawn in base-paired stem-loop structures. The reality of such structures is indicated by the relative

resistance of these regions to digestion by T1 RNAase. The first or operator-proximal stem-loop involves base-pairing between nucleotides 74 to 85 and 108 to 119 (numbered from the start of the trpL transcript; Fig. 4c). However, as the putative stop codon for trpL translation lies at 69 to 71, ribosomes which reach this position might unfold or interfere with formation of the first stem-loop. Lee and Yanofsky (1977) therefore proposed that regulation at trpa may be effected as follows. Under conditions of adequate intracellular levels of Trp-tRNA^{trp}, ribosomes will translate the trpL transcript to the stop codon at position 70, inhibiting formation of the first stem-loop and thus favouring formation of the second (114 to 121 paired with 126 to 134). This may represent the activation or exposure of the termination signal so that attenuation is effected. As termination at trpa is rho-dependent in vivo, it may be that it is the rho-recognition site which is activated or exposed. Consistent with the hypothesis that formation of the second stem-loop is required for termination, the analagous structure in Salmonella typhimurium is less stable and attenuation less efficient than in E.coli. (The mechanism in both cases appears similar: see below.)

Now, under conditions where tRNA^{trp} is largely uncharged, ribosomes translating trpL should tend to pause at the tandem Trp codons. This would allow formation of the first stem-loop. The resulting RNA secondary or tertiary structure would then obscure or obliterate the termination signal (e.g. the second stem-loop, or a sequence

necessary for rho recognition and/or action). Readthrough at trpa would thus be favoured and trp operon expression accordingly increased.

Subsequent findings, however, suggest that amendments to this model may be necessary. If the merodiploid strain trpT⁺/trpT su7 is grown in excess glutamine, tRNA^{trp}_{su7} (which recognises the UAG stop codon instead of UGG) will be efficiently charged with glutamine (Gln). Yanofsky and Soll (1977) have shown that in such a situation trp operon expression does not increase in response to lowered levels of Trp, even though ribosomes should still stall at the tandem Trp codons. They suggest therefore that some factor must normally recognise or bind Trp-tRNA^{trp} to induce termination. They further suggest that Gln-tRNA^{trp}_{su7} is still recognised by the factor, so that termination remains efficient, whatever the Trp-tRNA^{trp} concentration, in the above experiment. They argue that this factor is unlikely to be tryptophanyl-tRNA synthetase as this has a 50-fold lower affinity for tRNA^{trp}_{su7} than for normal tRNA^{trp}.

Earlier, Pouwels et al (1975) found that crude extracts prepared from E.coli grown on glucose medium (i.e. with a high metabolic rate) contain a factor (probably a protein of MW in the range 10-100 K) which can specifically stimulate trp operon expression in vitro. This appears to act by relieving attenuation at trpa and thus is thought to be an antitermination factor (Pannekoek et al, 1975). Further studies (Pouwels and Pannekoek, 1976) have shown that this response can also be elicited in trpS⁻ strains

suggesting that this factor is not Trp-tRNA^{trp} synthetase, and that Trp-tRNA^{trp} is not involved in this aspect of the regulation of attenuation. The apparent discrepancy between these results and those of Morse and Morse (1976, see above) may be due to the use of a relA strain in the experiments of Pouwels and Pannekoek, as opposed to relA⁺ in those of Morse and Morse. The latter authors in the same report show that the Trp-tRNA^{trp}-mediated response is dependent on an active relA gene. It is therefore reasonable to suppose that the glucose-induce relief of attenuation may be via a completely different pathway, and thus may involve a different (but possibly overlapping) set of factors.

Zurawski et al (1978a) have isolated two cis-dominant point mutants which show a decreased relief of attenuation in response to limiting levels of Trp and a higher efficiency of termination in vivo. One of these (L75, Fig. 4b) would reduce the stability of the first stem-loop structure, consistent with Lee and Yanofksy's model wherein the second stem-loop is necessary for termination. However, the other mutation (L29) alters the AUG start codon for the putative trpL polypeptide to AUA. Translation of trpL is therefore unlikely to occur, so that formation of the first stem-loop should be unimpeded and thus, according to the original model, attenuation should be prevented. That this is not the case suggests that stalled ribosomes play an active part in the relief of attenuation. Zurawski et al have also noted that there is considerable complementarity between the region containing the Trp codons (54 to 59) and

a region within the loop of the first stem-loop (86 to 91: Fig. 4c). Stalled ribosomes might act by favouring formation of a base-paired structure between these latter two regions, which alone or in conjunction with the first stem-loop would be responsible for neutralising the termination signal. Alternatively, the presence of stalled ribosomes could be necessary to prevent pairing of 54-59 with 86 to 91 and so allow formation of the first stem-loop.

Stauffer et al (1978) have isolated several mutants which show a decreased efficiency of attenuation in vivo under conditions of excess Trp. These contain single base-pair changes all of which would reduce the stability of the second stem-loop structure of the model (Fig. 4b,c). Relief of termination was greatest in those mutants where the first stem-loop was not also destabilised. This supports the suggestion that formation of the second stem-loop is an important positive feature and formation of the first stem-loop an important negative feature in the operation of the termination signal.

Bertrand et al (1977) have characterised three mutants carrying deletions affecting trpL (Fig. 4b). Their results have already been discussed in section a, above.

Considering all this evidence, an expanded model of regulation at trpa can be constructed. When a low concentration of intracellular Trp is reflected in a pre-dominance of uncharged $tRNA^{trp}$ over $Trp-tRNA^{trp}$, ribosomes translating the nascent trpL transcript will stall at the tandem Trp codons at position 54 to 59, resulting in a rel dependent synthesis of ppGpp and formation of a complex

base-paired structure in the nascent RNA (probably including the first stem-loop and possibly additional pairing between positions 54 to 59 and 86 to 91). This somehow neutralises the termination signal and/or the site of rho action, so that attenuation at trpa is relieved and the trp operon is expressed. Relief of attenuation can also occur during growth in glucose medium through the action of a specific antitermination cofactor.

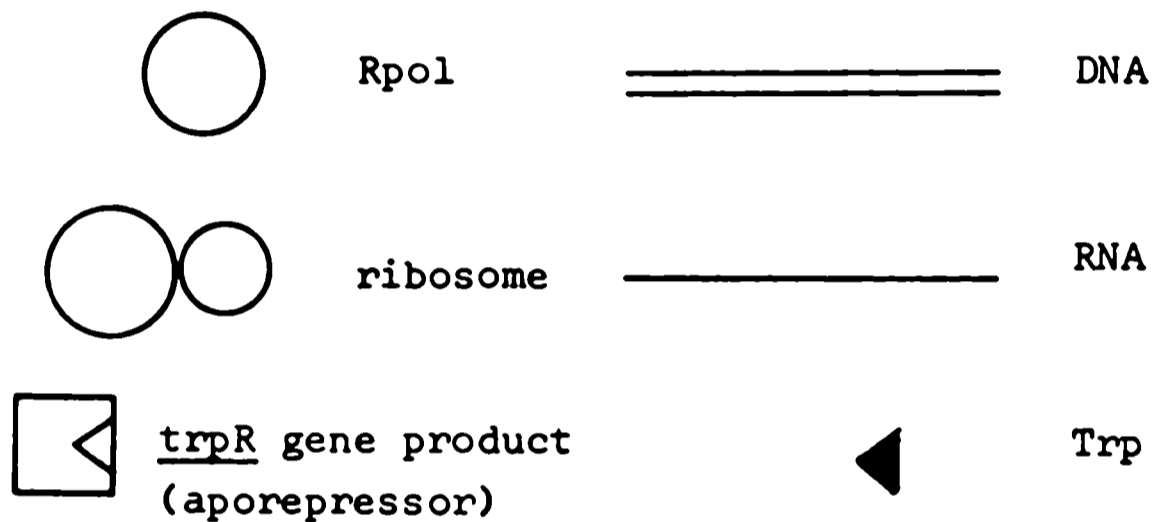
Under conditions where an adequate supply of intracellular Trp is reflected in a predominance of Trp-tRNA^{trp} over tRNA^{trp}, ribosomes translating the nascent trpL transcript will reach the stop codon at position 69 to 71. This will prevent the formation of the complex base-paired structure proposed above, and instead will favour the formation of the simple, "second" stem-loop structure (positions 114 to 134). In the presence of the general termination factor rho and a specific Trp-tRNA^{trp} activated termination factor, Rpol will then terminate transcription in the run of U's (positions 134 to 141) at trpa. These control pathways are schematically illustrated in Fig. 5.

The trp operons of Serratia marcescens and Salmonella typhimurium are also capable of responding to the intracellular levels of Trp through regulation at an attenuator, preceding the first structural gene (Lee et al, 1978; Miozzari and Yanofsky, 1978). DNA-sequencing studies have shown that two tandem Trp codons are present in a putative protein coding region of the trpL transcript, in both species (Lee et al, 1978). Two stem-loop structures

Figure 5. Regulation of trp Operon Expression in E.coli

Regulation of trp operon expression is here shown diagrammatically (details in text). Positions and genetic designations are as given in Fig. 4.

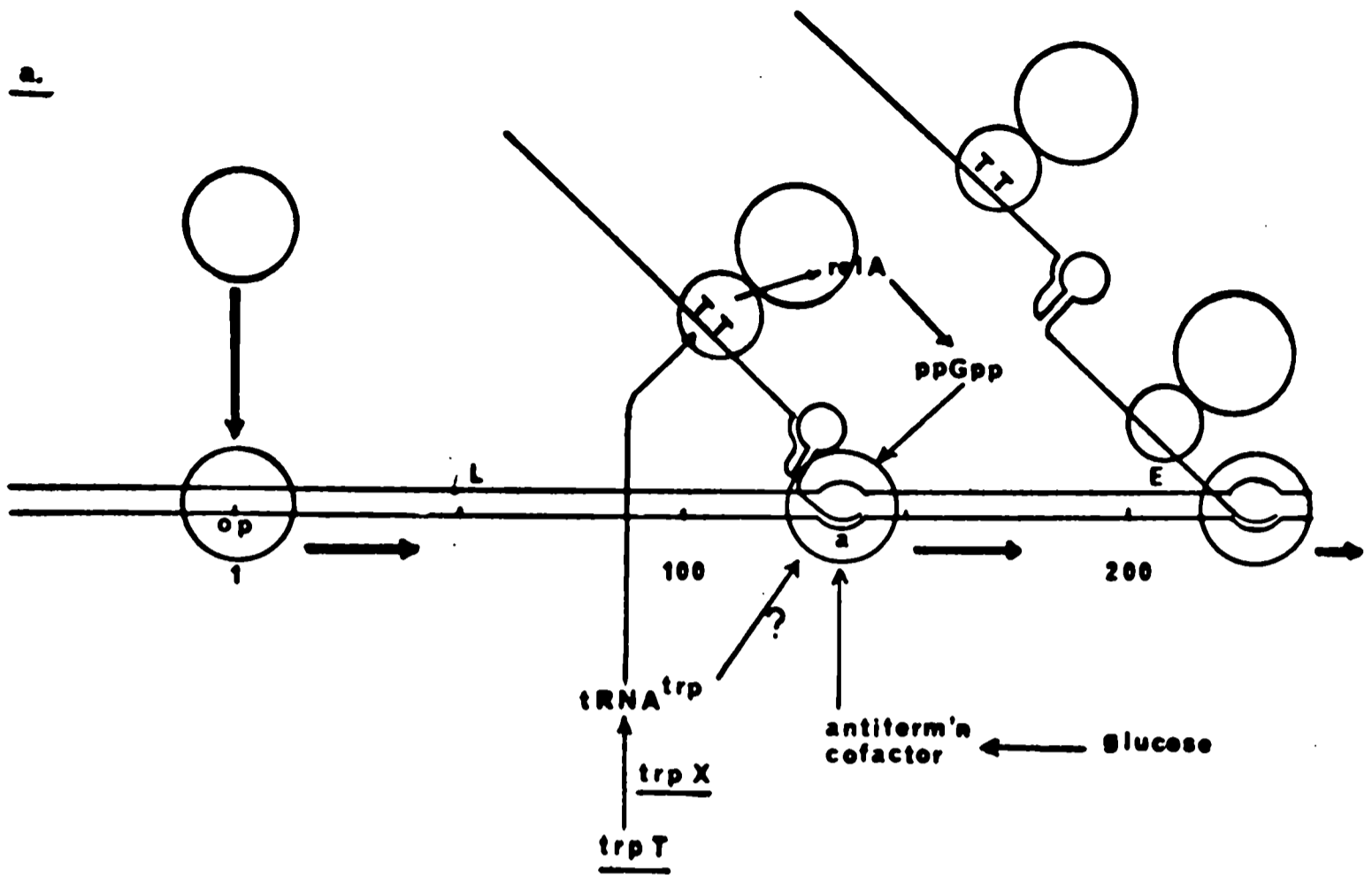
- a. Increased expression under conditions of limiting concentrations of intracellular Trp, or of high metabolic rate resulting from growth on glucose.
- b. Decreased expression under conditions of elevated concentrations of intracellular Trp, or low metabolic rate due to growth on poor carbon sources.



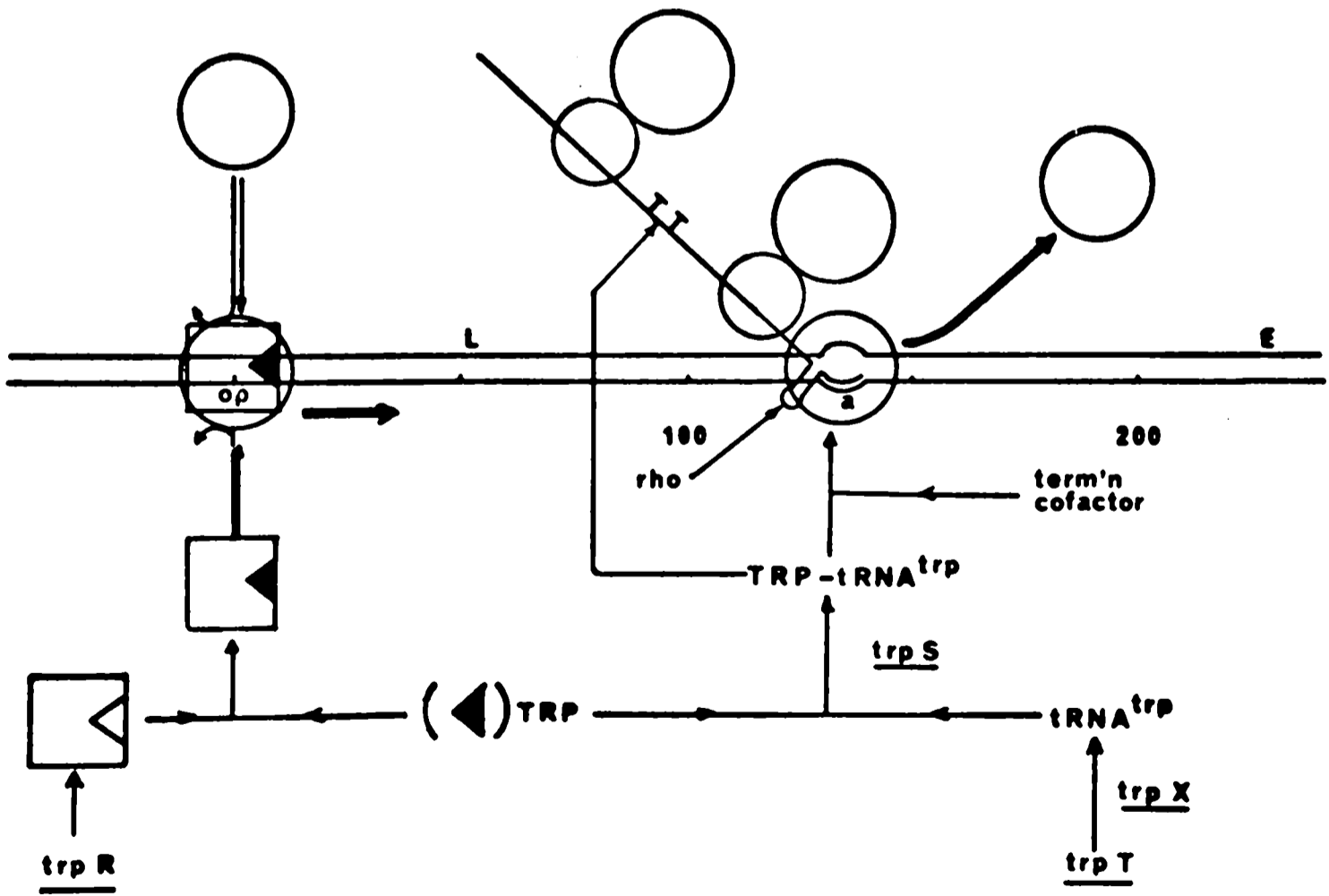
<u>trpR</u> :	<u>trp</u> aporepressor	<u>trpT</u> :	tRNA ^{trp} structural
<u>trpS</u> :	Trp-tRNA ^{trp} synthetase	<u>trpX</u> :	tRNA ^{trp} modification

Figure 3.

a.



b.



analogous to those postulated in the trpa region of E.coli, can again be drawn, in both cases, so that it seems likely that attenuation of trp operon expression is regulated in the same manner in all three bacterial species.

One of the genes involved in the de novo synthesis of L-phenylalanine (Phe) in E.coli is in an operon which is responsive to the cellular level of Phe (Zurawski et al, 1978b). There is an attenuator (phea) located operator-proximal to the only structural gene of the operon, which requires a properly modified tRNA^{phe} for activity, and is rho-dependent in vivo. DNA sequencing studies show a region of pheL which is analogous to trpL and which, if translated, would encode a short polypeptide 15 amino acids long, 7 of which would be Phe residues. Two stable stem-loop structures can be drawn in the phea region, as well as a base-paired structure between the region coding for the Phe residues and the loop of the operator-proximal stem-loop structure. These results strongly suggest that regulation of the phe operon at phea involves a mechanism similar to that operating in trp.

The genes involved in the de novo synthesis of L-Histidine (His) in E.coli and S.typhimurium are organised in a large operon whose expression is responsive to the intracellular levels of His probably mediated through His-tRNA^{his} (Smith, 1972). Kasai (1974) has proposed that a regulatable attenuator precedes the first structural gene of this operon. Although no direct evidence for such attenuation has yet been reported, DNA sequencing studies have disclosed a region of this operon preceding the first

structural gene which, if transcribed, might code for a polypeptide 16 amino acids long, 7 of which would be tandem His residues (DiNocera et al, 1978). About 30 base-pairs operator-distal to this is a region which, if transcribed, could form either of two potentially very stable stem-loop structures, followed by an A + T rich region. These similarities to the trpa region of E.coli suggest that a similar mode of regulation may be operating here. It is noteworthy that the his operator in S.typhimurium, at least, is not regulated by a His/aporepressor complex analogous to that of the trp operon. This may explain the exaggerated features of the putative his attenuator region (i.e. the 7 tandem His codons, and the high stability of the alternative stem-loop structures).

4. Background and Approach to the Present Study of Termination in Coliphage T7

Termination of transcription by E.coli Rpol at the end of the early operon of coliphage T7 occurs in a defined region (t_1) with an efficiency in vitro of up to 95% (Millette et al 1970; Peters and Hayward, 1974a,b; O'Hare, 1978). This activity is independent of the addition of accessory protein factors.

T7 has been the subject of extensive studies in vivo and in vitro (Studier, 1969, 1972; Hausmann, 1976) including the isolation and characterisation of several deletion mutants which affect termination at t_1 (Simon and Studier, 1973). Moreover, transcription of T7 DNA by purified host Rpol appears to be a faithful reflection of

the in vivo situation. T7 t_1 was therefore considered to be a suitable system in which to study termination in vitro. The particular approach chosen for these studies has been to determine the nucleotide sequence in the region of t_1 . By comparison with other known terminator sequences, and ideally with those of mutations affecting termination at t_1 , it was hoped that this work might increase our understanding of the termination event in general.

I shall discuss below the relevant background information on T7, and then briefly outline the strategy adopted.

a. Coliphage T7

The genetics and biochemistry of coliphage T7 have been extensively dealt with by Studier (1969, 1972) and, more recently, by Hausmann (1976). I shall confine myself to a discussion of gene expression within this virus.

The genome of T7 is a linear, double-stranded DNA molecular which is not circularly permuted (Studier, 1972) but which does contain a short terminal redundancy of about 70 bp at either end (Ludwig and Summers, 1975). The genome comprises some 40000 bp (McDonnell et al, 1977) and thus has a molecular weight in the order of 25×10^6 . About 30 genes have so far been detected, 20 of which are evidently essential for laboratory growth on a wild-type host (Studier, 1972).

Detailed physical maps showing the cutting sites for several type II restriction endonucleases have been determined for the leftmost portion of the genome (Ludwig and Summers, 1976; Gordon et al, 1978; Siebenlist, 1979) but only one

report has so far appeared which extends the mapping as far as t_1 (McDonnell et al, 1977).

The genes of T7 can be grouped into three (overlapping) classes according to the period of the infection cycle during which they are expressed (Studier, 1969). As might be expected each of these classes comprises functionally related genes. Each is also physically clustered on the genome.

i) Class I gene expression (Fig. 6)

These genes are expressed from about 4 through 8 min after infection, and are located in the leftmost 20% of the genome, the so-called "early region" (Studier, 1969). They are collectively involved in neutralising the host's defence system and in directing the cell's metabolic processes towards the expression of the T7 genes.

There have been several reports describing transcription of the early operon in vitro and in vivo. I shall consider the two situations individually.

Purified E.coli Rpol initiates transcription of T7 DNA preferentially at three closely spaced promoters designated A_1 , A_2 and A_3 , located near the left end of the genome: specifically, at map positions 1.2, 1.5 and 1.8, respectively (Stahl and Chamberlin, 1977; Siebenlist, 1979).

If the Rpol:DNA ratio is very high, polymerase molecules will also bind to four other sites on T7 DNA (Stahl and Chamberlin, 1977; Koller et al, 1978; ^{Delius et al, 1973}). These have been designated minor promoters B, C, D and E as transcription can be initiated at these sites under certain conditions in vitro (Stahl and Chamberlin, 1977; McConnell, 1979a).

Figure 6. Genetic and Physical Map of the Early Region of T7.

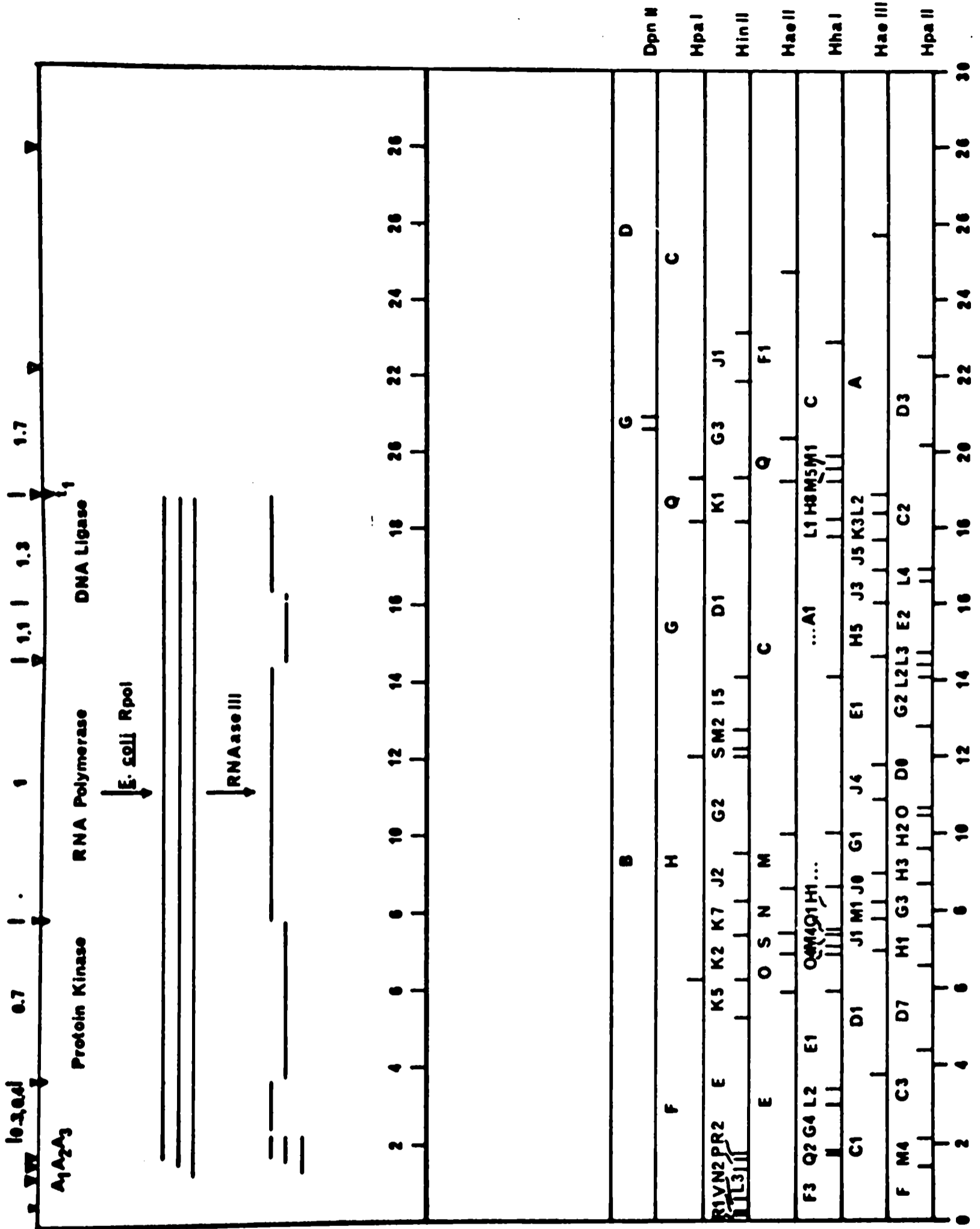
- a. Genetic map showing the structural genes and regulatory sites within the early region. Genes are designated by a number (essential genes by whole numbers, and inessential by fractional numbers) and by their activity, where known (see text). Transcripts initiated by E.coli Rpol at the three major early promoters (A_1 , A_2 and A_3 ; designated by ∇) and terminated at t_1 ($\bar{\nabla}$) are shown before and after processing by RNAase III. Initiation and termination can also occur at minor promoter ($\underline{\nabla}$) and terminator ($\bar{\nabla}$) sites, located where indicated.

Class II T7 Rpol promoters are indicated by (∇).

The scale is in T7 map units. One map unit = 1% T7 DNA = 400 bp.

- b. Physical map showing the cutting sites for 7 restriction endonucleases as determined by F.W. Studier (personal communication). The scale is as above.

Figuro 6a



A

Three of these sites, B, C and D are located in the early region, at positions 0.6, 3.7 and 7.7, while E is located near the right end of the genome, at position 92.1 (Delius *et al.*, 1973; Koller *et al.*, 1977; McConnell, 1979a; Minkley and Pribnow, 1973; Studier, personal communication).

Elongation proceeds from left to right, terminating with up to 95% efficiency at a site designated t_1 (Millette *et al.*, 1970; Peters and Hayward, 1974a,b; O'Hare, 1978). This site defines the end of the early operon and is located near map position 18.9 (Simon and Studier, 1973; see Chapter 7 for a complete discussion of the data for this mapping). The resulting polycistronic transcript can be cleaved by purified RNAase III, apparently in a way which faithfully mimics the physiological situation to yield five independent mRNAs and at least four short, apparently untranslated species (Rosenberg *et al.*, 1974; Robertson *et al.*, 1977). The five mRNAs can be efficiently translated *in vitro* to give 6 polypeptides (Dunn and Studier, 1975), now designated 0.3, 0.4, 0.7 1, 1.1 and 1.3. (The 0.3 and 0.4 polypeptides are encoded on a single processed mRNA: Studier, 1975; Steitz and Bryan, 1977.)

Translation of these genes *in vitro* can also occur from the unprocessed polycistronic transcript, in most cases with much the same efficiency as from the RNAase III-cleaved messages (Dunn and Studier, 1975; Yamada and Nakada, 1976). The main known exception to this is that translation of the 0.3-coding region is much reduced in the unprocessed transcript. This could reflect an inhibitory effect of the extra RNA in the latter transcript on the ability of

ribosomes to bind and/or initiate translation at the start of 0.3 (Steitz and Bryan, 1977).

As indicated by its whole number designation, the only class I gene product essential for growth on the wild type host is that of gene 1. This product has been identified as a T7 specific Rpol (Chamberlin et al, 1970) which is responsible for transcription of the class II and III genes.

Three of the other early genes have also been identified with a particular function. Gene 0.3 codes for a protein (MW ~14000) which is involved in overcoming the host's restriction system (Studier, 1975) and in superinfection exclusion (McAllister and Barrett, 1977b).

Gene 0.7 codes for a protein kinase (MW ~36000) which apparently phosphorylates the host Rpol (Zillig et al, 1975). Phenotypically, this gene is required for the rapid shut off of early gene transcription (McAllister and Barrett, 1977c; Hesselbach and Nakada, 1977) and for the efficient recognition of termination sites within and at the end of the early operon (Ponta et al, 1974; Pfennig-Yeh et al, 1978).

Gene 1.3 codes for a DNA ligase which is probably involved in DNA synthesis and repair (Studier, 1972). However, as gene 1.3⁻ mutants are effectively complemented by the host's DNA ligase, its expression is not normally essential to the virus. Consistent with its function, gene 1.3 is also expressed as a class II gene: it is transcribed by T7 Rpol initiating at the class II promoter near map position 15 (Oakley and Coleman, 1977).

No activity has yet been clearly ascribed to the 0.4 and 1.1 genes.

According to Darlix (1974), termination of transcription occurs at intercistronic regions within the early operon when the termination factor rho is added to the transcription mix in vitro. Dunn and Studier (1973) observed a similar effect, although they concluded that the termination so produced was not confined to specific sites. More recently, McConnell (1979b) has shown that transcripts initiated at the C promoter are frequently terminated at two or more downstream sites located just before the start of the gene 1 coding region. This result was obtained in the absence of added rho factor. Termination at these sites can also occur when high concentrations of Rpol, together with the initiation inhibitor rifampicin, are used in vitro (Darlix and Horaist, 1975). Kassavetis et al (1978) have suggested that this occurs because open promoter complexes are formed at these sites, but are unable to initiate transcription because of the presence of rifampicin. These inhibited complexes would effectively block progress by polymerases which had initiated transcription upstream, prior to addition of the drug.

Transcription in vivo is apparently very similar to that in vitro. Initiation again occurs almost exclusively at the A promoters (Studier, 1972), although Pfennig-Yeh et al (1978) reported evidence that initiation may occasionally occur within the early operon.

Elongation then proceeds from left to right until a site located between the 0.4 and 0.7 genes is reached. In

an RNAase III⁺ host, about 50-75% of the transcripts will be terminated at this site (Hercules et al, 1976; Pfennig-Yeh et al, 1978). The remaining transcription complexes continue elongation to a site between genes 0.7 and 1 where according to Hercules et al (1976) a further 50% of the transcripts are terminated. This leaves about one-quarter of the initial transcribing complexes which continue to the end of the operon, where termination at a site (designated t₁) downstream of gene 1.3 occurs (Studier, 1972). The efficiency of termination at this site is at least 75% (McAllister and Barrett, 1977a).

Termination at the two intraoperonic sites is dependent on an RNAase III⁺ phenotype in the host, and on 0.7⁺ in the phage (Pfennig-Yeh et al, 1978). Efficient termination at t₁ is also apparently dependent on active T7 gene 0.7 (Ponta et al, 1974). The product of this gene is a protein kinase, which can phosphorylate E.coli Rpol (Zillig et al, 1975). Pfennig-Yeh (1978) have suggested three possible explanations for the effect of this gene on termination:

- 1) phosphorylation of the host Rpol is necessary for recognition of these terminators;
- 2) phosphorylated polymerases bind to sites near to or overlapping these terminators (e.g. perhaps by unproductive binding at the minor promoters B and C, in the case of termination before genes 0.7 and 1), and so block elongation past these sites by active, unmodified polymerases;
- or 3) protein kinase stimulates termination through a direct protein:protein interaction with Rpol at the termination site. (The ability of purified Rpol to recognise t₁ in vitro in the absence

of the gene 0.7 product is reminiscent of the difference between in vitro and in vivo dependency on rho factor observed for the so-called rho-independent terminators: evidently there is some antitermination activity (protein or otherwise) present in vivo which is somehow suppressed by rho and/or gene 0.7.)

The dependence on RNAase III for terminator function at the intraoperonic sites might indicate a direct involvement of RNAase III in termination. Alternatively the expression of the T7 gene 0.7 in an RNAase III⁻ host may be reduced, in which case the RNAase III phenotype would affect termination indirectly. (The mechanism of this reduction could be similar to that discussed above, in the case of gene 0.3 expression.)

Shut-off of early gene expression in vivo is primarily mediated through the gene 0.7 product (McAllister and Barrett, 1977c; Hesselbach and Nakada, 1977). However, in T7 0.7⁻ mutants shut-off does occur after some delay: this effect is evidently dependent on the gene 2 product (Hesselbach and Nakada, 1977). The mechanism of shut-off is not known for certain, although the 0.7-mediated effect may be related to the phosphorylation of the E.coli Rpol, while there is also evidence that the gene 2 product binds directly to the host polymerase and inhibits it in vitro (Hesselbach and Nakada, 1977).

ii) Class II and III

The class II genes are responsible for T7 DNA replication and host DNA degradation (Studier, 1972). Transcription of these genes occurs approximately from 6 to 15 min after

infection at 30° (Studier, 1972) or from 4 to 16 min after infection at 37° (McAllister and Wu, 1977). Class II transcription is initiated at several promoter sites both in vivo (Pachl and Young, 1978) and in vitro (Golomb and Chamberlin, 1974). Termination of transcription initiated at class II promoters occurs predominantly at a site to the right of gene 10 near map position 60.

The class III genes code for structural, assembly, and DNA-maturation functions (Studier, 1972). These genes are expressed from about 8 min after infection until the time of lysis. Promoters for class III transcription are present at several sites in vivo, the leftmost of which is at about position 45 on the physical map, i.e. to the left of gene 8 (Rosa, 1979). Genes 8 through 10 therefore are expressed from both class II and class III promoters. These class III transcripts are terminated at the same site as class II (i.e. map position 60): the remainder terminate near the right end of the genome (Golomb and Chamberlin, 1974; Pachl and Young, 1978; Rosa, 1979).

During the period of overlap of expression, class II genes are transcribed in vivo about as efficiently as class III. In vitro, however, class II transcription is much less efficient than class III (McAllister and Barrett, 1977a).

The switch from class II to class III gene expression is not yet understood, although it is apparently dependent on the class II gene 3.5, "lysozyme" (McAllister and Wu, 1978).

b. Nucleotide Sequencing

There are presently two alternative methods for deter-

mining the nucleotide sequence of DNA: the chemical modification method of Maxam and Gilbert (1977) and the chain-terminating method of Sanger et al (1977).

Both methods require a detailed physical map of restriction endonuclease cutting sites in the area of interest, in order that a suitable fragment of DNA can be isolated for use in sequence analyses. Accordingly, the generation of such a map of the t_1 region of T7 DNA was the first step in these sequencing studies (see Chapters 4 and 5).

The approach chosen for the actual sequence determination was the chain-terminating method. This is potentially more rapid, more economical and simpler to use than the chemical modification method, provided that it can successfully be adapted to the system under study. Because the strands of T7/^{DNA} can easily be separated (Szybalski et al, 1971) it seemed likely that such an adaptation would indeed be possible. This was the strategy adopted; the results are described in Chapter 6.

CHAPTER 2 - MATERIALS

1. Bacteria and Bacteriophage

The strains of bacteria and bacteriophage used in this study are described in Tables 1 and 2, respectively. Bacteriophage stocks were stored at 2° in phage buffer with a drop of CHCl₃ or in CsCl solution. Stocks of bacteria were kept at -70° in nutrient broth, 15% (v/v) glycerol, 0.5% (w/v) NaCl.

2. Media

T-broth (Studier, 1969) contained 10 g Bacto Tryptone and 5 g NaCl in 1 l distilled water (d H₂O).

L-broth (Lennox, 1955) contained 10 g Bacto Tryptone, 5 g Bacto Yeast Extract and 10 g NaCl in 1 l d H₂O. pH was adjusted to 7.2 with NaOH.

BBL-broth contained 10 g BBL-trypticase and 5 g NaCl in 1 l d H₂O. pH was adjusted to 7.2 with NaOH.

For T7 phage titrations, bottom agar was 1% (w/v) Bacto agar and top agar was 0.7% (w/v) Bacto agar, both in T-broth.

For T7 plate lysates to be used as sources of DNA, bottom agar was 1.5% (w/v) agarose in L-broth, and top agar was 0.65% (w/v) agarose in BBL-broth. Bacto products were from Difco; BBL were from Becton-Dickinson.

3. Restriction Endonucleases (Roberts, 1978)

All restriction endonucleases (endo R.) used in this study were type II (Nathans and Smith, 1975) and are described in Table 3.

Table 1 Bacterial Strains

Strain	Description	Source	Ref
<u>Escherichia coli</u> B	su ⁻ , prototroph	F.W. Studier	1
<u>Escherichia coli</u> BBW/1	B, su ⁺ , prototroph	R. Hausmann	2

1. F.W. Studier. *Virology* 39, 562-574 (1969)

2. R. Hausmann and B. Gomez. *J.Virol.* 1, 779-792 (1967)

Table 2 Bacteriophage Strains^a

Strain	Deletion Limits ^b	RNA Affected	Protein Affected
T7 WT	-	-	-
T7 C5 ^c	3.2 - 8.0	0.3, 0.7, 1	0.3, 0.7, 1
T7 LG3 ^c	15.2 - 19.2	1, 1.1, 1.3	1.1, 1.3
T7 LG13	15.6 - 20.4	1, 1.1, 1.3, T	1.1, 1.3
T7 LG26	19.1 - 23.5	1.3, T	1.3, 1.7
T7 LG37	15.2 - 21.0	1, 1.1, 1.3, T	1.1, 1.3
T7 am 342a ^c	-	-	1

^a All strains were obtained from F.W. Studier (some of them via D. McConnell) and are as described in M.N. Simon and F.W. Studier, *J.Mol.Biol.* 79, 249-265 (1973)

^b As given in Simon and Studier (1973). Note, however, that data presented in the results section of this thesis have been used to more accurately map these deletions as summarised in Table 7.

^c Obtained from F.W. Studier as T7 C5 LG3 am 342a.

Table 3. Restriction Endonucleases

Endor	Bacterial Origin	Specificity ^a	Source ^b	Use ^c
AvaI	Anabaena variabilis	CYCGRG	see Methods	G
BamI	Bacillus amyloliquefaciens	GGATCC	K. Brown, K. Murray	G
BclI	Bacillus caldolyticus	TGATCA	MRE, P. Broda	G
BglI	Bacillus globigii	?	K. Murray	G
EcoRI	Escherichia coli RY13	GAATTC	K. Brown, K. Murray	G
HaeII i	Haemophilus aegyptius	RGGCGY	K. Murray	G
HaeII ii	Haemophilus aegyptius	RGGCGY	NEBL, N. Murray	S
HaeII iii	Haemophilus aegyptius	RGGCGY	NEBL	S
HaeIII	Haemophilus aegyptius	GGCC	A. Smith	S
HgiAI	Herpetosiphon giganteus HP1023	G ^T GA ^T CA ^T _A	N. Brown	G
HhaI	Haemophilus haemolyticus	GGCC	K. O'Hare	S
HincII	Haemophilus influenzae Rc	GTYRAC	NEBL	S
HindII i	Haemophilus influenzae Rd	GTYRAC	R. Thompson	G
HindII ii	Haemophilus influenzae	GTYRAC	MRE, P. Broda	S
HindIII	Haemophilus influenzae	AAGCTT	K. Brown, K. Murray	G
HpaI i	Haemophilus parainfluenzae	GTTAAC	K. Murray	G

HpaI ii	Haemophilus parainfluenzae	GTTAAC	see Methods	G
HpaI iii	Haemophilus parainfluenzae	GTTAAC	NEBL, N. Murray	G
HpaII	Haemophilus parainfluenzae	C [↓] CGG	see Methods	S
KpnI	Klebsiella pneumoniae OK8	GGTACC [↓]	MRE, P. Broda	G
MboI	Moraxella bovis	[↓] GATC	B. Smith	G
PstI	Providencia stuartii 164	CTGCAC [↓] G	MRE, P. Broda	G
SalI	Streptomyces albus G	G [↓] TCGAC	K. Brown, K. Murray	G
SmaI	Serratia marcescens S _b	CCC [↓] GGG	K. Brown, K. Murray	G
XbaI	Xanthomonas badrii	T [↓] CTAGA	K. Murray	G
XhoI	Xanthomonas holcicola	C [↓] TCGAG	K. Murray	G

^a As given in R. Roberts. Gene 4, 183-193 (1978)

^b Where known to be different, the commercial supplier is given first (MRE and NEBL, see list of suppliers, section 12), followed by the local donor. These many gifts are gratefully acknowledged.

^c G, used in gel analyses only.

S, also used in sequencing studies.

4. Other Enzymes

E.coli RNA polymerase (EC 2.7.7.6) was generously given by K. O'Hare.

E.coli DNA polymerase I (Klenow A fragment, EC 2.7.7.7) was purchased from Boehringer.

5. Nucleotides

Ribo-, 2'-deoxyribo-, and 2',3'-dideoxyribonucleoside 5'-triphosphates were purchased from PL. Solutions in d H₂O were stored at -20°.

Poly r(U,G) containing randomly polymerised guanylic and uridylic acid residues in a ratio of 1:1 to 1:1.5, was from Miles.

6. Radiochemicals

The Radiochemical Centre supplied α -³²P 2'-deoxycytidine 5'-triphosphate, ammonium salt (PB165, >350 Ci/mmole) and 2'-deoxyguanosine 5'-triphosphate, triethylammonium salt (PB166, >350 Ci/mmole).

7. Gel Electrophoresis

N,N'-methylene-bis-acrylamide, 2 x recrystallised acrylamide and N,N,N',N'-tetramethylethylenediamine (TMED) were from Serva.

Analar formamide and stock solutions of acrylamide and bis-acrylamide in d H₂O, for use in 7 M urea sequencing gels, were deionised by stirring for 30 min with ~1 g/10 ml Amberlite IR-120 (from Hopkins and Williams) ion exchange resin and filtering through a glass sinter to remove the resin.

Bromophenol blue (Bpb) was from Sigma.

Xylene cyanol FF (XCFF) was from BDH.

Ethidium bromide was grade B from Calbiochem.

Agarose was from either Sigma (type II) or Miles.

Electrophoresis tape, perspex spacers, glass plates and vertical gel electrophoresis apparatus were from Raven.

Plastikard was plain white sheet from Slaters.

8. Gel Photography

UV illumination was by either a Mineralight C-5 (overhead, 254 nm peak transmission) or a Transilluminator C-62 (365 nm peak transmission) obtained from Ultra Violet Products.

FP4 film (4 x 5 in) was from Ilford.

9. Autoradiography

The gel drier was Model 224 from Bio-rad.

X-ray film was XH-1 or BB-5 from Kodak.

Intensifying screens were either Fast Tungstate from Ilford or Cronex Xtra-Life Lightning-Plus from Du Pont.

The flash unit was an Agfatronic 150 A from Hamilton-Tait behind a Safelight filter (6B) from Ilford.

Saran wrap was from Spontex.

10. Chromatography

Bio-gel A 0.5 m was from Bio-rad.

Diethylaminoethyl cellulose (DE52) and phosphocellulose (P11) were from Whatman.

Sephadex G-100 was from Pharmacia.

Conductivity meter was from Radiometer.

11. Miscellaneous

All other chemicals were analytical grade, where possible, obtained through various general suppliers.

Nitrocellulose filter discs (0.45 μm pore size) were from Sartorius.

Streptolydigin was a gift from Dr G.B. Whitfield Jnr. of the Upjohn Co. Rifampicin was kindly donated by Dr M. Coldman of Lepetit. Stock solutions of both (100 mg/ml in dimethyl formamide) were stored at -20° .

Polyethylene glycol was Carbowax 6000 from Union Carbide. Repelcote, a 0.2% solution of dimethyl dichlorosilane in CCl_4 was from Hopkins and Williams. Sarkosyl NL35 (a 35% (w/v) solution of sodium lauryl sarcosinate) was from Geigy.

Bovine serum albumin (BSA) was step V from Armour. A 20 mg/ml solution in 50 mM Tris-Cl (pH 7.8) was frozen and thawed several times to destroy contaminating nucleases.

Visking dialysis tubing was from Scientific Instrument Centre.

Diethylpyrocarbonate (DEP), dithiothreitol (DTT) and 2-mercaptoethanol (2-me) were from Sigma.

Analytical grade phenol was redistilled over metallic zinc in a system flushed with nitrogen. It was collected under $\text{d H}_2\text{O}$ and stored in the dark at -20° . 25 ml aliquots were equilibrated against 50 mM Tris-Cl (pH 7.8) before use.

CsCl was baked at 180° for 18 h and stored dry in a desiccator.

Samples of bacteriophage λ DNA and the plasmid pBR322 (from E.coli) were generously donated by A. Newman.

12. Suppliers

Armour Pharmaceuticals Co Ltd, Eastbourne, UK.
BDH Chemicals Ltd, Poole, Dorset, UK.
Becton-Dickinson, Empire Way, Wembley, UK.
Bio-Rad Labs, Bromley, Kent, UK.
Boehringer Corporation, Lewes, Sussex, UK.
Calbiochem, Bishops Stortford, Herts, UK.
Difco, West Molesey, Surrey, UK.
Du Pont (UK) Ltd, Huntingdon, Cambs, UK.
Geigy (UK) Ltd, Manchester, UK.
Hamilton-Tait Ltd, Penicuik, Midlothian, UK.
Hopkins and Williams, Chadwell Heath, Essex, UK.
Ilford Ltd, Basildon, Essex, UK.
International Enzymes Ltd, Windsor, Berks, UK.
Kodak, Hemel Hempstead, Herts, UK.
Lepetit Pharmaceuticals Ltd, Maidenhead, Berks, UK.
Miles Biochemicals, Stoke Poges, Slough, UK.
Microbial Research Establishment (MRE), Salisbury, UK.
New England Bio-Labs (NEBL), Beverly, Mass, USA.
Pharmacia GB Ltd, Uxbridge Road, London, UK.
P-L Biochemicals Inc, c/o International Enzymes Ltd.
Radiochemical Centre, Amersham, Bucks, UK.
Radiometer, Copenhagen, Denmark.
Raven Scientific Ltd, Haverhill, Suffolk, UK.
Sartorius Membranfilter GmbH, Gottingen, W. Germany.
Scientific Instrument Centre Ltd, Leake Street, London, UK.
Serva, Uniscience Ltd, Fulham, London, UK.
Sigma, Poole, Dorset, UK.
Slaters Plastickard Ltd, Matlock Bath, Derbs, UK.

Spontex Ltd, Croydon, Kent, UK.

Ultra Violet Products Inc, Endover, UK.

Union Carbide Ltd, Hythe, Southampton, UK.

Upjohn Co, Kalamazoo, Michigan, USA.

Whatman Biochemicals Ltd, Maidstone, Kent, UK.

CHAPTER 3 - METHODS

1. General

a) Siliconisation of Glassware

Glassware to be used in nucleic acid work (or wherever recovery was important) was thoroughly cleaned and siliconised before use by soaking overnight at room temperature (r.t.) in conc. persulphuric acid (200 g $\text{Na}_2\text{S}_2\text{O}_8$ in 2.5 l conc. H_2SO_4), rinsing thoroughly in tap H_2O and then distilled H_2O (d H_2O), and finally drying at 100° . This acid-cleaned glassware was then briefly immersed in Repelcote, drained, dried at 100° , rinsed once in d H_2O and dried for storage. The silicone coating is resistant to autoclaving and temperatures as high as 120° , and survives several gentle washes in mild detergent.

b) Sterilisation of Dialysis Sacs

20-50 cm lengths of Visking dialysis tubing were boiled 15 min in 1% (w/v) Na_2CO_3 , rinsed thoroughly with d H_2O , filled with 0.01 M EDTA and dialysed vs d H_2O overnight. Sacs were then emptied, rinsed with d H_2O and stored in 0.1 mM EDTA at 4° . Sterile solutions and technique were used throughout.

c) Centrifugation

All medium and high-speed spins (i.e. above 10,000 rpm) are individually described elsewhere, and are expressed in terms of average g force (followed by the actual rotor type and speed used), temperature, and time of run. The particular machine used is indicated by the rotor type.

Low speed spins (less than 10,000 rpm) are not similarly detailed as a variety of machines and rotors were used depending on the volume to be spun and machine availability.

d) Phenol Extraction

Nucleic acid solutions were mixed with an equal volume of redistilled phenol pre-equilibrated against buffer), and gently shaken on ice for 5-10 min in siliconised glass tubes or plastic snap-cap tubes. The mixtures were spun at 5000 g (5 kg) at r.t. for 5 min to separate the phases. The phenol (lower) phase was removed and back-extracted with a small amount of TE buffer (10 mM Tris-Cl (pH 7.5), 1 mM EDTA; typically 0.3 x original volume). The aqueous phases were combined and re-extracted with an equal volume of phenol.

Residual phenol was removed by three ether extractions: an equal volume of ether was added to the solution, mixed for 5 min, spun as above and the ether (upper) phase removed and discarded. Residual ether was allowed to evaporate at r.t. for one hour.

e) Ethanol Precipitation

Nucleic acid solutions in siliconised corex tubes or plastic snap-cap tubes were mixed with 2 volumes ethanol and 0.1 volume 3 M CH₃ COOK (unless the salt concentration already exceeded 0.3 M) and left at -20° overnight or -70° for 20 min. The precipitate was pelleted at 18 Kg (12.5 Krpm, Sorvall SS34; full speed, Quickfit Microcentrifuge), 4° for 20 min, washed once in 80% ethanol, dried in vacuo and resuspended to the desired volume in an appropriate buffer, usually TE.

2. T7 Growth and DNA Purification

a) Phage Titration

T7 phage suspensions were sequentially diluted to an expected titre of 10^3 plaque forming units (pfu) per ml. 0.1 ml of such a dilution was mixed with 0.2 ml of a stationary phase E.coli culture in 3 ml molten (45°) top agar and the mixture poured onto a plastic petri plate (9 cm diam.) containing about 30 ml set bottom agar. Phage plaques were visible after 3-4 hours at 37° .

b) Large Scale Lysates

0.5 l cultures of exponentially growing E.coli B ($\sim 10^8$ cells/ml) in 2 l flasks were infected with sufficient T7 to give a multiplicity of infection of 0.1 (E.coli BBW/1 was the host for T7 C5LG3 am 342a growth). Growth with vigorous aeration at 37° was continued until lysis was complete as judged by eye. Lysates were pooled, made 0.5 M NaCl, 0.2% (v/v) chloroform, stirred for 1 h at 4° , and then left to aggregate overnight in the cold. The entire volume was spun at 10 Kg (8 Krpm, Sorvall GSA) for 15 min to remove cell debris. The supernatants were pooled, made 6% (w/v) with polyethylene glycol and left to settle for 48 hours in the cold (Yamamoto et al, 1970). The bulk of the supernatant was siphoned off and the remaining slurry spun at 10 Kg (8 Krpm, Sorvall GSA) for 10 min, 4° . The phage-containing precipitate was resuspended in 10 ml phage buffer per litre of original lysate, and cleared by spinning at 5 Kg (6.5 Krpm, Sorvall SS34) for 10 min. The supernatant was retained and the pellet washed with 5 ml phage buffer per litre original lysate.

The combined supernatants were cleared by a final spin at 10 Kg (9.2 Krpm, Sorvall SS34) for 5 min, before layering onto a preformed 25 ml CsCl step gradient: 3 ml of CsCl solution in phage buffer, density 1.78 g/cm³, 4 ml of 1.65 g/cm³, 4 ml of 1.54 g/cm³, 5 ml of 1.47 g/cm³, 4 ml of 1.40 g/cm³ and 5 ml of 1.24 g/cm³, in 70 ml polyallomer tubes. The gradients were overlaid with 34 ml of phage suspension followed by paraffin oil, and spun at 63 Kg (22.5 Krpm, MSE 59591), 10⁰ for 18 h. Fractions were collected from the bottom of the gradient and titrated for phage. Peak fractions were pooled and brought to 17 ml in 1 x 3 in polyallomer tubes with a CsCl solution, density 1.50 g/cm³ in phage buffer, and spun at 51 Kg (22.5 Krpm, Bechmann SW25.1), 10⁰ for 66 h. The blue-grey phage band was drawn into a 5 ml syringe by piercing the side of the tube. This was titrated and either stored at 2⁰ or used directly for purification of DNA. Suspensions stored for greater than one month were rebanded by repeating the last spin before attempting to purify DNA.

To extract DNA, such suspensions were dialysed exhaustively against 0.1 M Tris-Cl (pH 8.0), 0.1 M NaCl, 0.01 M EDTA to remove CsCl, and then treated with phenol to remove protein. Final dialysis to remove residual phenol, ether and salt was against TE buffer (four changes). The purified DNA solutions thus obtained were stored at 2⁰ in sterile snap-cap tubes. Approximately 0.6 mg DNA/1 T7 WT or T7 C5LG3 am 342a lysate and 0.05 mg DNA/1 T7 LG37 lysate, was obtained.

c) Small Scale Lysates (Studier, 1969)

5 ml of a fresh overnight culture of E.coli were mixed with 15 ml prewarmed (37°) T-broth and about 10^7 phage (or a single plaque) and shaken at 37° until lysis was complete. 1 g NaCl and 0.4 ml chloroform were added, shaking continued for 10 min at 37° , and then the lysates were cleared by a 10 min spin at 5 Kg. Phage were pelleted by spinning the cleared supernatant at 55 Kg (29 Krpm, Beckman type 30), 4° for 75 min. The pellet was resuspended in 1 ml phage buffer and given a final clearing spin at 7 Kg for 10 min. The resulting supernatant was titrated for phage and stored at 2° .

For DNA extraction, 5×10^{11} pfu in 0.2-0.8 ml were laid over a 1.5 ml CsCl step gradient (0.5 ml of 1.6 g/cm³, 0.5 ml of 1.5 g/cm³ and 0.5 ml of 1.4 g/cm³) in polyallomer tubes and spun at 102 Kg (33 Krpm, Beckman SW50.1), 4° for 2 h. The visible phage band was removed by syringe, dialysed against TE buffer to remove CsCl, phenol-extracted, ether-extracted, and then ethanol-precipitated. Pellets were resuspended in 50 μ l TE buffer and stored at 2° . 5×10^{11} pfu yielded approximately 25 μ g DNA.

d) Plate Lysates (Cameron and Davis, 1977)

10^3 - 10^4 phage derived from a single plaque were mixed with 0.2 ml stationary phase E.coli ($1-2 \times 10^9$ /ml) in 3 ml molten (45°) BBL-agarose (0.65%) and overlaid on L-agarose (1.5%) plates. Incubation was at 37° until lysis was confluent (~ 4 hours). 5 ml of 10 mM Tris-Cl (pH 7.5), 10 mM EDTA were added to the plates

which were then left overnight at 4°. To 4 ml of supernatant on ice, 0.4 ml of 0.5 M EDTA (pH 8.5), 0.2 ml of 2 M Tris base, 0.2 ml of 10% (w/v) sodium dodecyl sulphate, and 10 µl diethylpyrocarbonate were added and the mixtures heated in open tubes in a fume cupboard at 65° for 30 min. 1 ml 5 M CH₃ COOK (pH 8-9) was then added and the mixtures left on ice for 1 h. The solutions were spun in polyallomer tubes at 27 Kg (15 Krpm, Sorval SS34), 4° for 10 min. The supernatant containing DNA was ethanol-precipitated, resuspended in 0.4 ml TE buffer and then extracted with phenol and ether as usual. The yield was 10-20 µg T7 DNA per plate.

e) T7 DNA Strand Separation (Szybalski et al, 1971)

CsCl-purified bacteriophage were dialysed against 1 mM EDTA (pH 8.0) and adjusted to an absorbance at 260 nm (A_{260}) of 10. To 0.5 ml aliquots in sterile polyallomer tubes, 40 µl of 5% (v/v) sarkosyl and 60 µl of 1 M NaOH were added and the mixture left for 9' at r.t. 200 µg poly r(U,G) in 55 µl was added, the solution gently mixed, left one min at r.t. and then neutralised with 60 µl of 2 M NaH₂PO₄ checking the final pH to be 7-8. The total volume was brought to 1.2 ml with 1 mM EDTA (pH 8.0) and then mixed with 4.8 ml saturated CsCl solution in 1 mM EDTA (pH 8.0), checking the final density to be 1.73-1.74 g/cm³. After overlaying with paraffin, the tubes were capped and spun at 95 Kg (38 Krpm, Beckman 50 Ti), 10° for 66 h. 0.2 ml fractions were collected from the bottom of the tube, diluted to 0.4 ml with 1 mM EDTA (pH 8.0) and the A_{260}

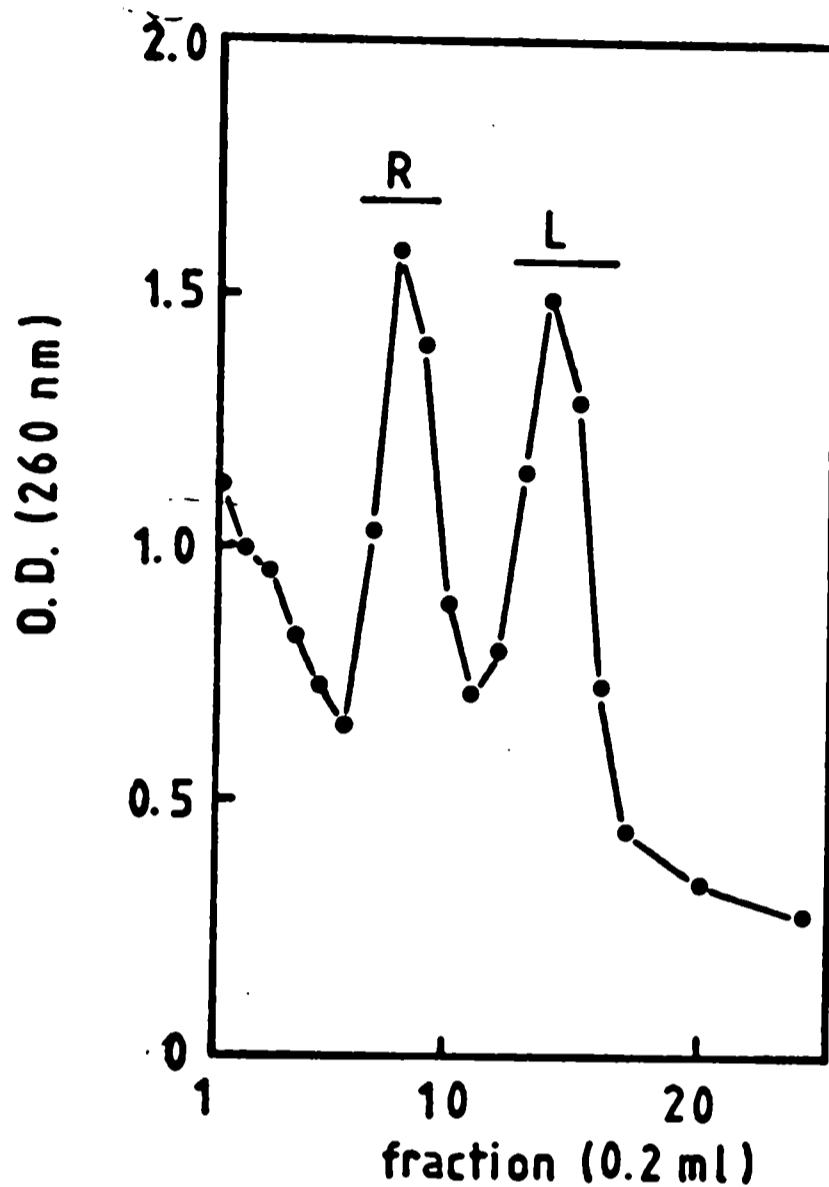


Figure 7. T7 WT DNA Strand-Separation

An absorbance profile (260 nm) of a CsCl fractionation of T7 WT DNA single-strands in the presence of poly r(U,g). The faster sedimenting peak (fractions 8-11) contained r-strand hybridised to poly r(U,G), while the slower sedimenting peak (fractions 12-16) contained free l-strand. The outer fractions indicated by the bar above each peak were pooled, and stored after treatment as described in the text.

determined and plotted. The density of three fractions (bottom, middle and top of gradient) was determined by precisely weighing 25 μ l samples. Two discrete A_{260} peaks were obtained, (Fig. 7), the denser of which was the poly r(U,G)/r-strand complex and the lighter, the l-strand alone.

Outer fractions from each peak were pooled, made 0.1 M NaOH, and incubated at 37° for 5 h to hydrolyse the poly r(U,G). The solutions were neutralised with HCl and dialysed against 10 mM Tris-Cl (pH 8.0), 10 mM NaCl, and 1 mM EDTA at 2°. The final A_{260} was determined and 1 ml aliquots stored in sterile siliconised vials at -35°.

3. Restriction Endonuclease (EndoR.) Preparation

a) Endo R. HpaI (Sharp et al, 1973)

All materials and buffers were autoclaved before use. All steps were performed at as near to 0° as practical. One unit of active enzyme digests 1 μ g T7 WT DNA to completion in 1 h under standard conditions.

9.86 g frozen concentrate of Haemophilus parainfluenzae grown on 3.7% Brain-heart infusion broth, were resuspended in 20 ml of 0.01 M Tris-Cl (pH 7.4), 0.001 M 2-mercaptoethanol (2-me) and lysed by sonication (15 x 30 sec at maximum output with an MSE 100 Watt Ultrasonic Disintegrator). The lysates were spun in polycarbonate tubes at 95 Kg (39 Krpm, Beckman type 65), 2° for 75 min. The supernatant (23 ml) was made 1 M NaCl and divided into three aliquots. Each aliquot was fractionated separately

by passing down an agarose column (A 0.5_m, Bio-rad; 66 x 2 cm diam., in 1 M NaCl, 0.02 M Tris-Cl, pH 7.4, 0.01 M 2-me) at 15 ml/h. 3 ml fractions were collected and assayed for activity by incubating 2 μ l samples from every third fraction in a total volume of 50 μ l, containing 10 mM Tris-Cl, pH 7.4, 10 mM MgCl₂, 10 mM 2-me, and 20 μ g/ml T7 WT DNA, for 2 h at 37^o. Digests were analysed by gel electrophoresis. Peak fractions (eluting at about 0.8 V_T) from one such loading were stored at 2^o (as a precaution) while those from the other two fractionations were pooled, dialysed against buffer C (10 mM KPi (pH 7.6), 10 mM 2-me, 0.1 mM EDTA, 10% glycerol) and loaded at 1 ml/h onto a phosphocellulose column (P-11, Whatman; 15 x 1.2 cm diam. in buffer C). After washing with 4 column volumes of buffer C at 5 ml/h, the HpaI activity was eluted at the same rate with a 90 ml linear gradient of 0-0.6 M KCl in buffer C, collecting 1 ml fractions. The conductivity and thus the KCl concentration of every third fraction was determined. Alternate fractions were assayed for HpaI activity. The peak fractions, eluting at about 0.35 M KCl, were stored individually at 2^o. Further purification of some of these fractions on a Sephadex G-100 column (49 x 1.5 cm diam, in buffer C) did not significantly improve the purity but did result in a very low recovery of active enzyme.

About 500 units active HpaI were obtained, with no appreciable loss of activity after several months storage. Incubation of 1 unit at 37^o for 1 hour in reaction buffer

(see Table 3) resulted in the loss of all detectable HpaI activity.

The method of De Fillipes (1974) yielded only a small amount of HpaI activity which was also heavily contaminated with HpaII.

b) EndoR. HpaII (Sharp et al, 1973)

The method used was as described above for HpaI purification, but with the following modifications.

A sonicated lysate from 11.8 g Haemophilus parainfluenzae was cleared, fractionated on an agarose column and assayed as described for HpaI except that the entire cleared supernatant was fractionated in a single loading and 6 ml fractions were collected. Four peak fractions (eluting at about $0.6 V_T$) were pooled, dialyzed, fractionated on a phosphocellulose column and assayed as described, except that loading was at 2 ml/h, elution was at 4 ml/h, and 2 ml fractions were collected. Six peak fractions (molarity of KCl not determined) were pooled, dialysed against buffer (10 mM Tris-Cl, pH 7.4, 10 mM 2-me, 0.3 mM EDTA and 50% (v/v) glycerol) and stored as three 2 ml aliquots at -20° . About 5-6000 units were obtained, with no appreciable loss of activity over 2.5 years storage.

c) Preparation of EndoR. AvaI (Murray et al, 1976)

AvaI was prepared from Anabaena variabilis as described in Murray et al (1976) except as noted below.

A freeze-thaw lysate of 20 g of Anabaena variabilis was cleared, ammonium sulphate fractionated, and chromatographed on a DEAE-cellulose column (DE-52, Whatman) as

Table 4. EndoR Reaction Conditions

EndoR	Tris-Cl		2-me mM	DTT mM	MgCl ₂ mM	BSA μg/ml	NaCl mM	Triton % (v/v)
	mM	pH						
AvaI	10	7.5	10		10			
BamI	10	7.5	10		10		50	0.005
BclI	6.6	7.4		1	6.6			
BglI	10	7.5	10		10			
DpnI	10	7.5	10		10			
EcoRI	10	7.5	10		10		100	0.005
HaeII i,ii	6.6	7.4		1	6.6			
HaeII iii	6	7.4	6		6	100		
HaeIII	6.6	7.4		1	6.6			
HgiAI	7	7.5	7		7		100	
HhaI	6	7.9	6		6		100	
HincII	10	7.9		1	7	100	60	
HindII i,ii	6.6	7.4		1	6.6			
HindIII	10	7.5	10		10		50	0.005
HpaI i	10	7.5	10		10		100	
HpaI ii	10	7.5	10		10		80	
HpaI iii	6.6	7.4		1	6.6		50	
HpaII	10	7.5	10		10			
KpnI	6.6	7.4		1	6.6			
PstI	6.6	7.4		1	6.6			
SalI	6	7.4	6		6	100	150	
SmaI	30	9.0	10		3		150*	
XbaI	10	7.5	10		10		100	
XhoI	10	7.5	10		10			

* KCl

described, except that the fraction precipitating between 30% and 40% (w/v) NH_4SO_4 was used, the column was 16.5 x 4 cm diam., and 9 ml fractions were collected. Phospho-cellulose-fractionation and concentration on DEAE-cellulose were as described. The peak fraction from the final DEAE-cellulose step was mixed with an equal volume of glycerol and split into five 1.5 ml aliquots for storage at -20° . A total of about 1000 units was obtained, with no appreciable loss of activity during 15 months storage.

4. Restriction Endonuclease Reactions

Reactions in a total volume of less than 50 μl were carried out in heat-sealed, drawn-out glass capillaries (100 x 1 mm diam.): those of more than 50 μl , in sterile snap-cap tubes. Volumes were greater than 10 μl for ease of handling, but otherwise minimised to maintain high enzyme and DNA concentrations. Reactions were generally initiated by mixing the appropriate amounts of DNA (in solution), 10 x reaction buffer, d H_2O and lastly enzyme. Incubation was at 37° for typically 1-3 h. The reactions were stopped by adding excess EDTA, usually in the form of gel electrophoresis loading buffer (see below). Table 4 lists the conditions used for each EndoR. These were as recommended by their respective donors (see Table 3).

5. Gel Electrophoresis

Five basic types of gel were used to resolve EndoR digests of DNA according to size.

a) Vertical Agarose (Sharp et al, 1973)

These gels were used where the expected digestion pattern contained relatively few and mostly large (greater than 400 bp) fragments of double-stranded DNA, or where high resolution was unnecessary. They were superseded by the horizontal agarose gel system, midway through these studies.

Two sizes of gel were used, according to the resolution desired: 13.5 cm wide x 0.3 cm thick and either 15 or 33 cm long. Similarly, low (1% (w/v)) or high (2% (w/v)) agarose solutions were used in the resolving portion of the gel according to the resolution required.

To make the gels, two molten agarose solutions were prepared by mixing:

- i) 2 g dry agarose with 100 ml TAE buffer (40 mM Tris-acetate pH 8.3, 33 mM NaCH₃COO, 1 mM EDTA), giving a 2% (w/v) solution;
- ii) sufficient dry agarose and TAE buffer to give 75 or 150 ml (for short (15 cm) or long (33 cm) gels, respectively) of a 1 or 2% (w/v) agarose solution as required. The agarose was dissolved by vigorous stirring on an electric hot plate/magnetic stirrer, allowed to boil for 1 min, and then cooled in a 50^o water bath. A glass sandwich was made, consisting of one notched plate (21 or 37 x 16.7 x 0.25 cm with a 2 x 14 cm centred notch at one end) separated by a pair of lightly greased perspex spacers (19 or 40 x 1.3 x 0.3 cm) from an unnotched plate (21 or 37 x 16.7 x 0.25 cm). The latter plate was first coated on the inner side with a solution

of 0.1% (w/v) agarose in water, and baked at 100° for 20 min. A U-shaped ridge of grease was placed around the notch at the top of a Raven gel electrophoresis apparatus, and the glass sandwich clamped in place with several 2 inch butterfly clips so that the notches in the glass plate and the apparatus were aligned. The assembled apparatus was tilted back to a 45° angle and the 100 ml molten 2% agarose solution poured into the bottom reservoir, immersing the bottom of the glass plates and running up about 2 cm between them. When this had set, the apparatus was returned to the upright position.

With a perspex comb (13 x 2 x 0.15 cm with 1-13 cuts 1.5 cm deep, centred in a perspex mount, 13 x 2 x 0.3 cm) at the ready, the second molten agarose solution was poured into the top of the sandwich, filling the cavity to about 2 mm below the notch. The comb was quickly inserted so that the teeth extended their full length below the notch. When set (15-30 min), the comb was carefully removed and the upper and lower reservoirs filled with TAE buffer. The samples to be analysed were mixed with 5 µl loading buffer (L-TAE: 0.5 x TAE, 0.125 M EDTA, 0.1% bromophenol blue (Bpb), 50% v/v glycerol) desiccated if necessary, and loaded by under-laying with a 50 µl Hamilton syringe so that the samples filled the well to no more than 8 mm high. A constant voltage gradient of no more than 10 v/cm was applied for several hours, as individually described.

To locate the DNA, the gel apparatus was disassembled and the gel stained in 1 µg/ml ethidium bromide for

20 min with gentle shaking. The gel was then washed 60 min in tap water before being photographed as described below (6).

b) Horizontal Agarose (McDonnell et al, 1977)

These gels superceded vertical agarose gels and had the same applications.

A 200 ml solution of molten agarose (0.5-1.5% (w/v)) in TAE buffer was prepared as described above. A perspex platform (30 x 16 x 0.5 cm) with fixed side walls (30 x 1 x 0.5 cm) and clamped-on end walls (14 x 1 x 0.5 cm) was positioned horizontally. Using U-shaped perspex mounts, a perspex comb (13 x 2 x 0.15 cm) with extended ends and typically twelve evenly spaced notches (1 x 0.3 cm) was positioned across the platform, 7 cm from one end with the downward pointing teeth separated by about 1 mm from the platform. The molten agarose solution was poured into the mold and allowed to set. A little buffer was then squirted onto the gel around the comb, and the comb gently removed. The clips were removed from the end pieces and the gel platform positioned over the buffer chambers (nearly full with TAE buffer), with the cathode chamber nearest the comb end. Buffer saturated paper wicks (2 layers of 3 MM paper, Whatman) were then hung from the gel so that efficient electrical contact between gel and buffer chambers was made.

Samples were loaded in a volume sufficient to just fill the wells (35 μ l for 13 well-gels) using a variable micro-pipetter. 60 V was applied across the gel (typically registering 150 V on the power pack voltmeter, due to the

resistance of the paper wicks) for 30 min, to run the dye and nucleic acids into the gel matrix. The wells were then topped up with buffer and the gel covered with Saran-wrap. A syphon between the two buffer chambers, in combination with a pumping of buffer from one chamber to the other, prevented the formation of a severe pH gradient which would otherwise distort the resolution of the samples. Up to 60 V (across the gel) were then applied for several hours (typically overnight) so that optimal resolution was obtained. DNA was detected as described above.

c) Preparative Polyacrylamide (Maniatis et al, 1975)

These gels, at various concentrations, were primarily used to resolve relatively large amounts of DNA for subsequent elution and purification.

A glass sandwich was assembled as described in 5a above, using uncoated glass plates and three spacers (two side and one bottom) of either 0.3 or 0.15 cm thickness ("thick" and "thin" respectively). The sandwich was clamped together and its edges were sealed with 2% (w/v) molten agar, and then it was placed in a large tank (42 or 24 x 20 x 15 cm for long or short gels respectively), below an ample water supply.

The acrylamide mixture was prepared in a side-arm flask by mixing the appropriate volumes of stock solutions of acrylamide/bis-acrylamide (29% and 1% (w/v), respectively), 10% (w/v) $(\text{NH}_4)_2 \text{S}_2\text{O}_8$ (prepared fresh every seven days) and 10 x TBE buffer (0.9 M Tris base, 0.89 M H_3BO_3 , 0.025 M EDTA, pH 8.3) and brought to 40 ml (short, thin gels),

80 ml (short, thick and long, thin gels) or 160 ml (long, thick gels) with d H₂O. The final concentrations are individually given in figure legends. The flask was stoppered and the solution degassed under reduced pressure. TMED was added and the solution poured into the glass sandwich, the tank being simultaneously filled with water, so that the level of the acrylamide solution just exceeded that of the water. The sandwich was filled to 1 cm below the notch. The solution was overlaid with a small volume of 2-butanol, and the gel allowed to set (1-2 h). The 2-butanol was then decanted and the top of the gel rinsed well with d H₂O. Individual perspex teeth (1.5 x 0.7 x 0.15 or 0.3 cm) were inserted until just touching the top of the gel, and molten agarose (2% w/v) in TBE buffer poured in between the teeth and/or side spacers, to give the desired number and width of wells. The perspex teeth, bottom spacer and clamps were removed and the sandwich positioned on the Raven apparatus as described in 5a. The chambers were filled with TBE buffer and any air in the wells or beneath the gel rinsed away. Samples were mixed with 5 µl L-TBE buffer (0.5 x TBE, 0.125 M EDTA, 0.1% Bpb, 50% v/v glycerol) for every 0.15 cm² of well base and desiccated so that when loaded they filled the well to a height of no more than 2 mm. A constant voltage gradient of up to 6 V/cm was then applied until the necessary resolution was obtained. Ethidium bromide staining was as described in 5d below.

d) Analytical Polyacrylamide

Gradient polyacrylamide gels were used for the analysis of restriction endonuclease digests of T7 DNA containing fragments between 40 and 5000 bp in size. TBE buffer was used throughout.

A long thin glass sandwich was assembled and placed in the empty tank, as described in 5c. A two-chamber linear gradient marker was suspended above a magnetic stirrer and positioned a few cm above the glass sandwich. Two 30 ml solutions were then prepared as described in 5c: a 3% (w/v) acrylamide mixture (3 ml acrylamide/bis-acrylamide (29% and 1% w/v respectively), 3 ml 10 x TBE buffer, 0.25 ml $(\text{NH}_4)_2\text{S}_2\text{O}_8$ and 23.75 ml d H_2O) and a 10% (w/v) acrylamide mixture (10 ml acrylamide/bis-acrylamide, 3 ml 10 x TBE buffer, 0.25 ml $(\text{NH}_4)_2\text{S}_2\text{O}_8$ and 16.75 ml d H_2O). After degassing (as in 5c), 6 μl TMED was added to each, and the high and low concentration solutions poured into the first (outlet) and second chamber, respectively. A 12 mm magnetic bar was added to the first chamber, and mixing commenced. The passage between the two chambers was opened, and a slow, steady flow of solution was then allowed to pass into the sandwich so that about 10 min were required to complete the pouring. The sandwich was filled to about 7 cm below the notch and a small amount of 2-butanol overlaid. When the gel had set, the 2-butanol was decanted and the upper surface rinsed well with d H_2O . 7 μl TMED was then added to a degassed, 20 ml stacking (2.5%) acrylamide mixture (1.67 ml acrylamide/bis-acrylamide (29 and 1% w/v

respectively), 2 ml 10 x TBE buffer, 0.25 ml 10% (w/v) $(\text{NH}_4)_2 \text{S}_2\text{O}_8$ and 16 ml d H_2O), and the mixture poured into the sandwich, filling it to 2 mm below the notch. A perspex comb (13 x 2.5 x 0.15 cm, with twelve evenly spaced notches, 0.6 x 0.15 cm) was then inserted (avoiding trapped air). When set, the clamps, bottom spacer and comb were removed and the apparatus assembled as in 5c. The gel was left for 24 hr before use and then prerun for 6 h at 150 V. Samples were prepared, loaded and run as in 5c. The run was stopped when the Bpb was 1 cm from the bottom of the gel. Staining was in 3 $\mu\text{g}/\text{ml}$ ethidium bromide for 10 min followed by washing for 8 min in H_2O . Photography was as described in 6 below.

e) Sequencing Polyacrylamide (7 M urea) (Sanger and Coulson, 1978)

These gels were used for the analysis of small ^{32}P -labelled single-stranded DNA fragments produced in DNA sequencing experiments. TBE buffer was used throughout.

A glass sandwich consisting of one long notched plate (ethanol washed, and Repelcoted on the inner surface) and one long unnotched plate, separated by two Plastikard spacers (37 x 1 x 0.04 cm) was assembled and the sides and bottom sealed with Raven electrophoresis tape. A 30 ml 8% (w/v) acrylamide solution was mixed and gently heated to 40° (6 ml of deionised acrylamide/bis-acrylamide (38% and 2% (w/v) respectively), 3 ml of 10 x TBE buffer, 0.15 ml of 10% $(\text{NH}_4)_2 \text{S}_2\text{O}_8$, 12.6 g urea and 8.25 ml d H_2O). 30 μl TMED was added and the sandwich, held at about 20° from the horizontal, completely filled. A Plastikard

comb (14 x 3.5 x 0.04 cm, with thirteen evenly spaced notches, 0.6 x 0.15 cm), prewashed for 5 sec in acetone and rinsed with ethanol and d H₂O, was quickly inserted so that the polyacrylamide teeth, when formed, would just protrude above the notched plate. When the gel had set, all the tape was removed, except for a strip running down each side to hold the plates together. Excess polyacrylamide was scraped away from around the comb, which was then gently removed and the apparatus assembled as usual. The gel was prerun, under cover, for 1-4 h at a constant current of 20 mA, to evenly preheat the gel.

The samples were mixed with 2-3 volumes L-form buffer (98% v/v deionised formamide, 0.3% (w/v) Bpb, 0.3% (w/v) Xylene cyanol ff (XCFF), 10 mM EDTA), heated to 100° for 3 min, cooled on ice and then 5-10 µl loaded per well (diffused urea was flushed out of the wells before loading). The gel was run under cover at 20-25 mA, giving an even heating to about 60°. Subsequent loadings on separate tracks were sometimes made when the XCFF marker dye from the first loading had migrated 15-20 cm from the origin (i.e. after 2-3 h). When the marker dyes indicated optimal resolution had been obtained (typically when the Bpb of the second loading had reached the bottom), the power was turned off and the gel allowed to cool for 1-2 min. The apparatus was then disassembled and the gel fixed in 10% (v/v) acetic acid for 10 min and washed in H₂O for 2'. Gel drying and autoradiography were as described in 7, below.

6. Gel Photography

Gels stained with ethidium bromide were photographed with comparable results in one of two manners.

a) Overhead Short Wave UV Illumination

Gels were laid on a piece of black polythene and illuminated from above in a Mineralight cabinet containing 4 short wave (254 nm) UV lamps. A 5 min exposure through a red filter onto FP4 film gave optimal resolution for all but analytical polyacrylamide gels, which required 8 min. Film processing was as recommended by Ilford for FP4, rated at ASA 125.

b) Long Wave UV Transillumination

Gels were laid directly onto a Chromatovue trans-illuminator and illuminated from below with six long wave UV (365 nm) lamps (15 Watts each). A 13 s exposure was used for all gels. Camera, film and processing were as in 6a, above.

7. Autoradiography

Wet gels on a glass plate were covered with Saran wrap and placed in contact with a piece of X-ray film. A second glass plate was placed over this and the sandwich clamped together for exposure in the dark at r.t.

For greater efficiency and resolution, gels were dried with heating onto 3 MM paper (Whatman) using a Bio-rad gel drier (20 min for sequencing gels, 0.04 cm thick, and 60 min for analytical polyacrylamide gels, 0.15 cm thick). The dried gel on paper was taped to a glass plate and overlaid with a piece of light-activated X-ray film (pre-

exposed to a single flash from 50 cm by an Agfa flash unit through an Ilford 6B safelight filter). A piece of intensifying screen (Ilford, Fast Tungstate or Du Pont, Cronex) was then laid face down over the film and the sandwich completed with another glass plate. This was all wrapped in 2 layers of aluminium foil, placed in a thick, black polythene bag, clamped, and left to expose in the dark at -70° . To avoid condensation, it was brought to r.t. before disassembling for developing.

Processing was as recommended by the film manufacturer.

8. DNA Fragment-Purification from Gels

Two basic protocols were used, varying in their ease of use and the purity of the resulting products.

a) Electroelution (McDonnell et al, 1977)

This method was simple, and rapid (less than a day) and gave a high recovery (50-100%). The DNA thus prepared, however was contaminated by a mild inhibitor of restriction endonucleases, and gave less clean results in sequencing studies than those produced by passive elution (8b).

Regions of preparative gels containing fragments of DNA to be eluted for sequencing were first identified by cutting longitudinal slices from the gel (usually a marker track and a small portion of the preparative track itself), staining in ethidium bromide, and photographing as described above. Using the photograph as a guide, the regions were then excised from the unstained remainder of the gel and passed through a sterile plastic syringe into a dialysis sac. 2 to 5 ml EE buffer (5 mM Tris base, 2.5 mM acetic

acid, pH about 6) was added, the sac sealed and placed in a shallow tray with a platinum electrode at each end so that the sac lay at right angles to the current flow. Sufficient EE buffer was added to completely cover the sacs and 100 V applied across the tray. After 1.5 h the polarity of the current was reversed for 5 min.

The contents of each sac were filtered under minimal negative pressure (1-2 cm mercury) through a GF/C filter, collecting the filtrate in a siliconised glass vessel. The sac and contents were washed with 2 ml EE buffer. The combined filtrates were ethanol precipitated. If the product was to be used in sequencing experiments, single-stranded (ss)-DNA (template) was added with the ethanol. The precipitated DNA was resuspended in the buffer appropriate for its ultimate use.

The same procedure was followed to electroelute fragments for purposes other than sequencing, except that the fragments were visualised directly by ethidium bromide staining and long wave UV fluorescence. To remove ethidium bromide, the filtrates were extracted four times with 1-butanol before ethanol precipitation (ethidium bromide but not DNA enters the 1-butanol phase).

b) Passive Elution

Appropriate regions of preparative polyacrylamide gels were localised and excised as described for sequencing experiments in 8a. The gel pieces were passed through a syringe into a siliconised corex tube containing 5 ml PE buffer (50 mM Tris-Cl (pH 8.0), 100 mM NaCl, 10 mM EDTA), parafilm sealed and gently swirled at 37^o overnight. The

gel bits were spun down by centrifuging at 5 Kg for 5 min, the supernatant removed and the loose pellet washed with 3 ml PE buffer. The supernatants were combined and passed through a 1 ml DEAE-cellulose column (DE-52, Whatman, in PE buffer) in a siliconised, glass-wool-stoppered, pasteur pipette. The flow-through was recycled and the column then washed with about 3 ml PE buffer. The DNA was eluted in 1-1.5 ml PEN buffer (PE buffer, 1 M NaCl) and precipitated by the addition of two volumes of ethanol. Again, if the product was to be used for sequencing, ss-DNA was added just prior to the ethanol.

This method, though more complex and less rapid than electroelution, gave a cleaner preparation both in terms of recutting with other restriction endonucleases and in sequencing experiments. The recovery was similar.

9. DNA Sequencing

a) Dideoxy-Blocking of Template

Purified T7 WT single-stranded DNA ('l' or 'r') at 30 µg/ml was denatured in Hin NaCl buffer (6.6 mM Tris-Cl (pH 7.4), 6.6 mM MgCl₂, 1 mM DTT and 50 mM NaCl), by heating at 100° for 3 min. Intrastrand annealing was then encouraged by incubating initially at 50° and then slowly lowering the temperature to 20° over a 2 h period. The mixture was then made 0.03 mM dATP, dGTP and dCTP, 0.25 mM 2',3'-dideoxy TTP (ddTTP), and incubated for 60 min at r.t. with 3 units E.coli DNA polymerase I (DpolI; Klenow A fragment). The reaction was stopped by making it 20 mM EDTA and heating to 70° for 10 min.

Polynucleotide was separated from mononucleotide by passing over a 6 ml column (13 x 0.8 cm diam.) of Sephadex G-100 in 5 mM Tris-acetate (pH 8), 0.25 mM EDTA at r.t. The flow rate was approximately 10 ml/h and 0.5 ml fractions were collected. The first A_{260} peak-fractions (containing only polynucleotide) were pooled and stored at 2° until coprecipitated with purified restriction fragment.

This should result in the presence of 2',3'-dideoxy-ribothymidine at the 3' ends of most polynucleotides which might otherwise act as primers in the DNA sequencing experiments described below.

b) Sequencing (Sanger et al, 1977)

In a typical experiment the template (dideoxy-blocked, single-stranded T7 WT DNA, l- or r-strand) was coprecipitated with the primer (purified restriction fragment) as described in section 8, above. The precipitate was redissolved in Hin NaCl buffer, denatured by heating at 100° for 3 min in glass capillaries, and then incubated for 60-100 min at 60° to promote interstrand pairing. This yielded a mixture containing the primed template necessary for sequencing (being full length single-strand annealed to the complementary strand of the restriction fragment) as well as reannealed double-stranded restriction fragment and the various unannealed single-stranded DNA molecules.

This mixture was divided into five aliquots designated G, A, T, C and H, and treated as described below. A schematic illustration of the theory underlying these experiments is given in Figure 8. Details of the specific

reaction conditions and solutions used in one experiment are shown in Tables 5a and b and in Figure 9.

A solution of three unlabelled dNTPs in buffer (termed $X^{\circ}N$, where X is the labelled nucleotide to be added and N is the nucleotide present in only very low concentrations; (e.g. N = dATP in the A reaction, see Table 5a) was mixed with a small volume of the appropriate ddNTP in aqueous solution (e.g. ddATP in the A reaction, see Fig. 9) and the mixture used to redissolve the dried-down $[\alpha\text{-}^{32}\text{P}]\text{-dXTP}$ (except in the case of the H reaction where the ddNTP was omitted). This mixture was added to an aliquot of the primed template and then DpolI was added and the reactions incubated at r.t. for 15 min (except for the H reaction which was left for only 2 min). Excess unlabelled dXTP corresponding to the ^{32}P - nucleotide was added and the incubation continued so as to chase any extension complexes which might have stalled because of the relatively low concentration of the labelled nucleotide (Fig. 8).

The restriction endonuclease used to generate the 3' end of the primer (along with a second EndoR in the case of the H-reaction) was then added and the reactions incubated at 37° for 15 min. The reactions were stopped by the addition of loading buffer (containing EDTA to chelate the essential Mg^{++} ions) and denatured as described in section 5e, above. This gave for each of the G, A, T and C reactions, a family of ^{32}P -labelled extensions with identical 5'-ends, varying in total length, but all ending (3') with the ddN used in the reaction

Figure 8. Schematic Illustration of One Sequencing
Experiment

A typical sequencing experiment is here presented schematically to show the theory behind the chain-terminating method.

The 900 bp HpaII C2/HaeII digestion product was hybridised to T7 WT r-strand and used as a primed template for the DpolI-catalysed extension reaction.

The autoradiograph of the gel produced by this experiment is shown in Fig. 30f.

Figure 8.

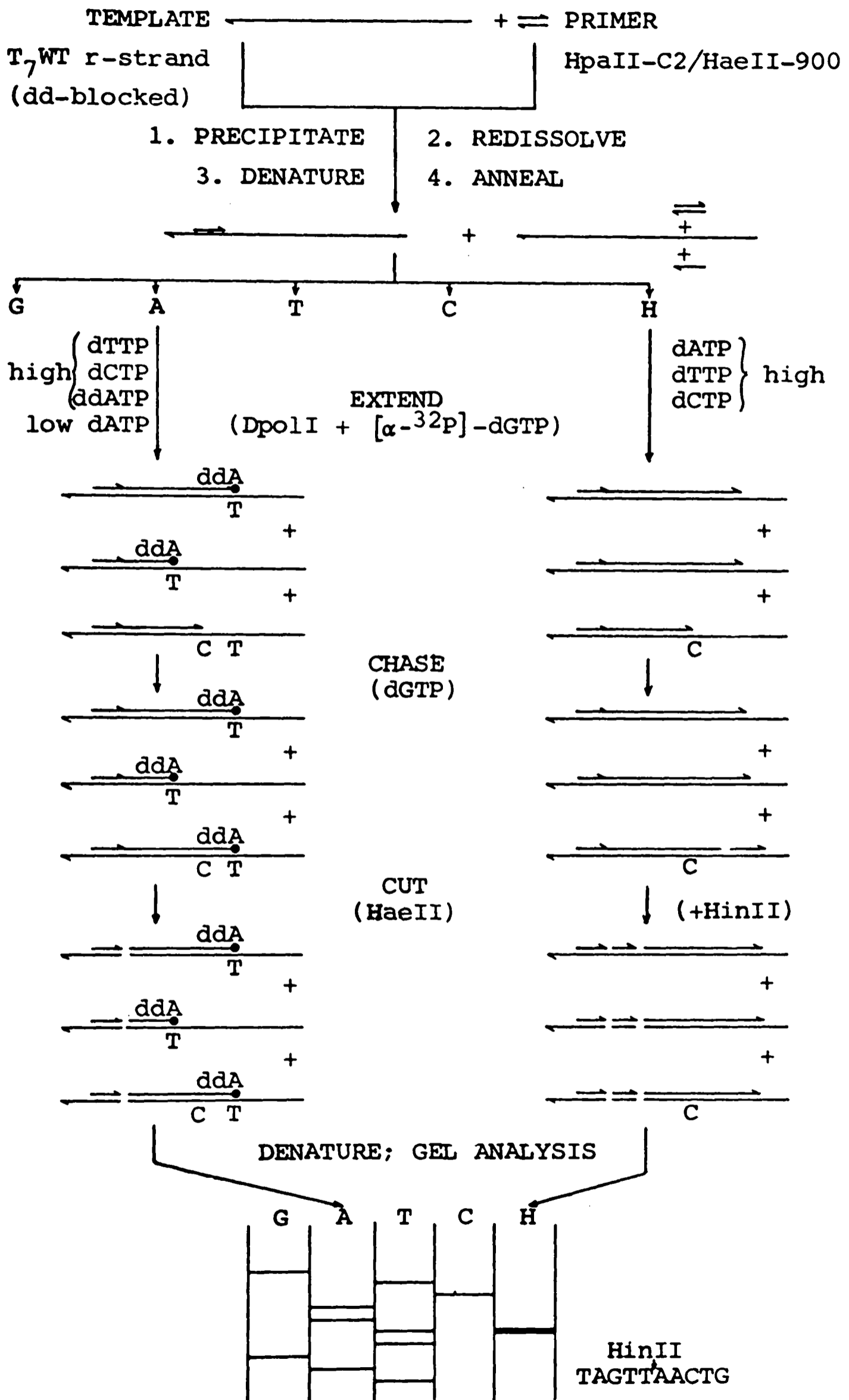


Figure 9. Detailed Protocol of One sequencing Experiment

The detailed protocol is shown for the sequencing experiment which was schematically illustrated in Fig. 8.

Unless otherwise indicated all numbers indicated volume added in μl .

The autoradiograph of the gel produced by this experiment is shown in Fig. 30f.

Specific variations used in other experiments are individually given in the appropriate figure legends or text of Chapter 6.

Figure 9.

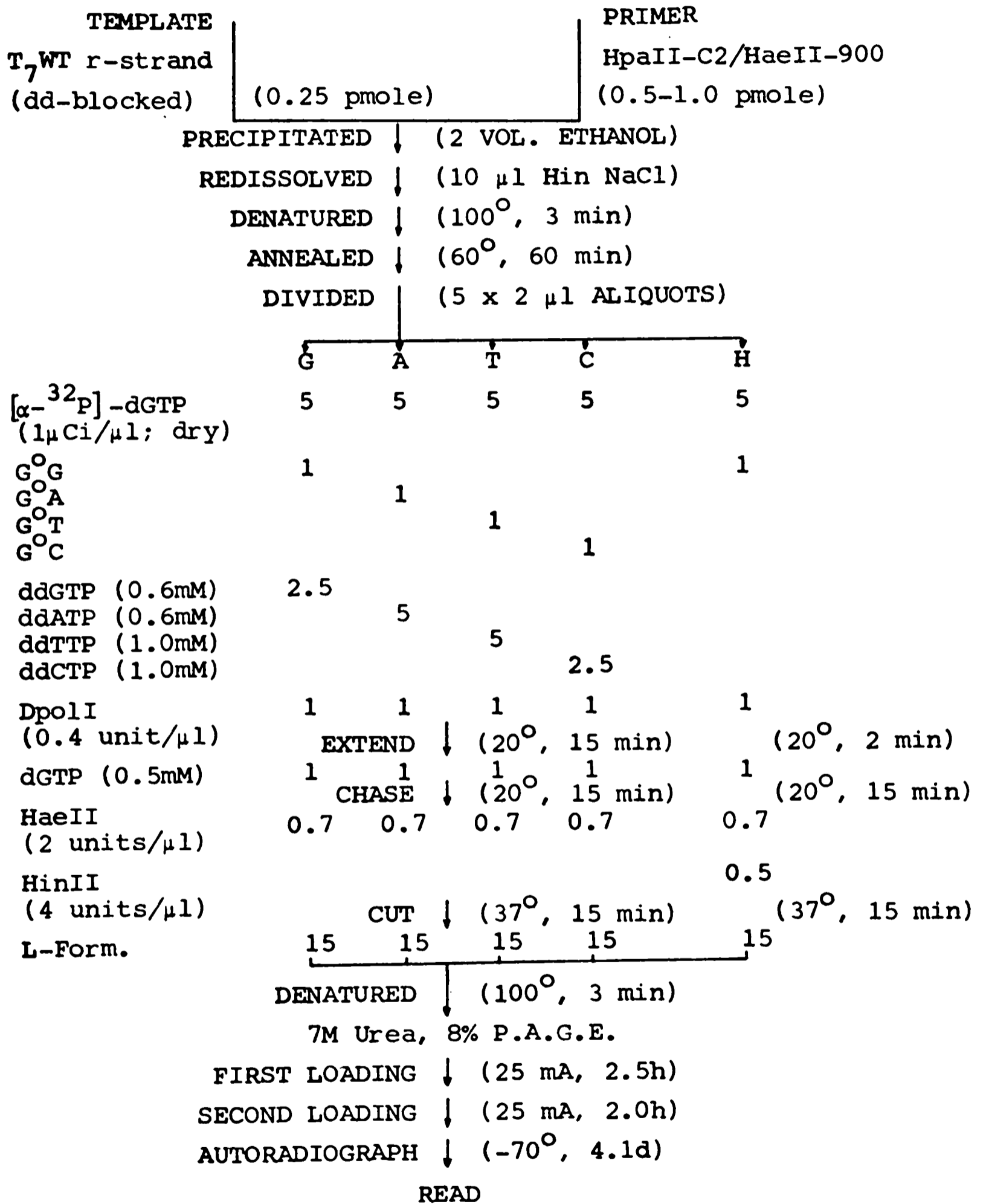


Table 5a. Nucleotide/Buffer Mixtures for Sequencing Experiments (G^oN)*

	dATP (mM)	dTTP (mM)	dCTP (mM)	Buffer
G ^o G	0.125	0.125	0.125	2.5 x Hin
G ^o A	0.011	0.17	0.17	3.3 x Hin
G ^o T	0.17	0.011	0.17	3.3 x Hin
G ^o C	0.17	0.17	0.011	3.3 x Hin

* Using [α -³²P]-dGTP as the labelled nucleotide.

Table 5b. Final Nucleotide Concentrations (μ M) in Extension Reactions in One Sequencing Experiment

Reaction	[α - ³² P]-dGTP	dATP	dTTP	dCTP	ddNTP
G	2	25	25	25	120 (ddGTP)
A	1.3	1	25	25	300 (ddATP)
T	1.3	25	1	25	500 (ddTTP)
C	2	35	30	1.3	200 (ddCTP)
H	3	40	40	40	nil

(e.g. ddA in the A-reaction). The H-reaction should contain labelled extensions which indicate the distance between the 3' end of the primer and the first downstream cutting site for the other EndoR used; depending on the length of the extensions, labelled restriction fragments, completely internal to these extensions, may be produced.

The products of each of these reactions were analysed by electrophoresis on adjacent tracks of a 7 M urea polyacrylamide gel, as described in sections 5e and 7, above. This gel system resolves single-stranded oligonucleotides, according to size (15 to 200 nucleotides long), with sufficient resolution to distinguish between bands containing DNA fragments which differ in length by only one nucleotide (in these experiments, the dd-nucleotide at the 3' end of the molecule). From an analysis of which track contained the darkest band at any given position (i.e. of defined length), a specific sequence was derived for the region in which the 3' termini of the extensions lay.

c) Computer Analysis

Computer analysis of the nucleotide sequence was kindly ^{out} carried/by Dr J.F. Collins. His programme searched for and listed all regions at which intrastrand basepairing would result in a negative free energy change, based on the rules and parameters defined by Tinoco et al (1973). In order to allow detection of potential, complex structures containing many such regions, the size of the loop of unpaired nucleotides linking the base-paired regions was not taken into account in the initial free energy estimations.

10. Binding of DNA Fragments to Nitrocellulose Filters

Reactions were set up by mixing T7 DNA, transcription buffer and E.coli RNA polymerase, incubating for 5 min at 37° and then initiating transcription by the addition of the four NTPs, so that the final mixture (in a volume of 60 µl) contained 60 mM Tris-Cl (pH 7.8), 1 mM EDTA, 180 mM KCl, 16 mM MgCl₂, 1 mM DTT, 1 mg BSA/ml, 20% (v/v) glycerol, 0.4 mM GTP, ATP, UTP and CTP, 150 µg DNA/ml, and sufficient E.coli RNA polymerase to give an active enzyme:DNA molar ratio of 2:1. After 1 min at 37°, further initiations were inhibited by addition of rifampicin to 25 µg/ml and incubation continued for a further 5 min at 37°. Streptolydigin was then added to 20 µg/ml. One min later 16 units of EndoR. HaeIII were added and the mixture left at 37° for 40 min. It was then split into two aliquots and filtered at 2° in either high or low salt conditions to isolate ternary, or both binary and ternary complexes, respectively. Of the filtrates, only those in low salt were normally analysed because those from high salt filtration caused severe distortion of neighbouring tracks on polyacrylamide gel analysis.

High salt filtration was performed by diluting the digest in 0.2 ml high salt washing buffer (10 mM Tris-Cl pH 7.8, 1 mM EDTA, 0.5 M NaCl), and then passing it through a 2.5 cm nitrocellulose filter (Sartorius) with mild suction (2 cm mercury). The filter was washed four times with 1 ml high salt washing buffer and then transferred to a siliconised glass scintillation vial for elution with 0.4 ml low salt washing buffer (10 mM Tris-Cl

(pH 7.8), 1 mM EDTA, 0.05 M NaCl) containing 1% (v/v) sarcosyl NL-35. After swirling for 30 min at 37^o, eluates were ethanol-precipitated, resuspended in 20 μ l TE buffer, mixed with 5 μ l TBE-loading buffer, desiccated to \sim 10 μ l, and analysed on gradient polyacrylamide gels (see section III 5d).

Low salt filtration was performed by diluting each sample in 0.2 ml low salt washing buffer at 2^o, and then passing it through a nitrocellulose filter disc as above. The filtrate was collected in a small siliconised vial (3 x 1 cm diam). Four 0.25 ml low salt washes of the filter were collected in the same vial, which was then set aside on ice while the filter was given four 1 ml low salt washes (filtrate discarded). The filter-bound material was eluted, and analysed as described above. The filtrate was ethanol-precipitated and analysed in parallel on polyacrylamide gels.

CHAPTER 4 - PHYSICAL MAPPING OF RESTRICTION ENDONUCLEASE

CUTTING SITES IN THE T7 WT GENOME

Two general approaches to mapping were used. The first was to compare the digestion patterns on gel electrophoresis of T7 WT DNA with those of deletion mutants affecting regions of DNA near the early transcription termination site. This site was originally mapped by Simon and Studier (1973) at between 20.2 and 20.3, later revised to 19.3 (McDonnell et al, 1977) and most recently to 18.9 by F.W. Studier (personal communication). Two deletions were used to obtain the bulk of the information: Δ LG37, which inactivates the transcription terminator, and Δ LG3 (in the double deletion mutant T7 C5LG3) which does not. The phenotypes of the mutants are detailed in Table 1, and a summary of the mapping information is given in Table 7. The use of the double deletion mutant did not generally confuse the results but rather gave some extra information.

The second approach was to purify particular restriction fragments from gels, recut them with another EndoR., and compare the resulting fragments with those in total DNA digests. This method provided most of the detailed mapping information around the transcription termination site. For this purpose electroelution proved to be an effective and rapid method for purifying fragments.

The following points apply generally to the mapping work and should be kept in mind throughout.

- i) The T7 map is divided into 100 units with the end

nearest the early region arbitrarily designated '0'. This end is also labelled 'left', so that transcription of the early operon proceeds from left to right. Each unit represents 1% T7 DNA, or 400 bp. Map positions will be expressed in absolute terms (e.g. 16.6) to avoid confusion with fragment sizes which will be expressed in base pairs (e.g. 450 bp) and deletion sizes which will be expressed in base pairs or % T7 (e.g. 1470 or 3.68%).

ii) Based on heteroduplex mapping (Simon and Studier, 1973), Δ LG3 and Δ LG37 are assumed to have identical left limits to their respective deletions. This assumption proved to be completely consistent with the results presented below.

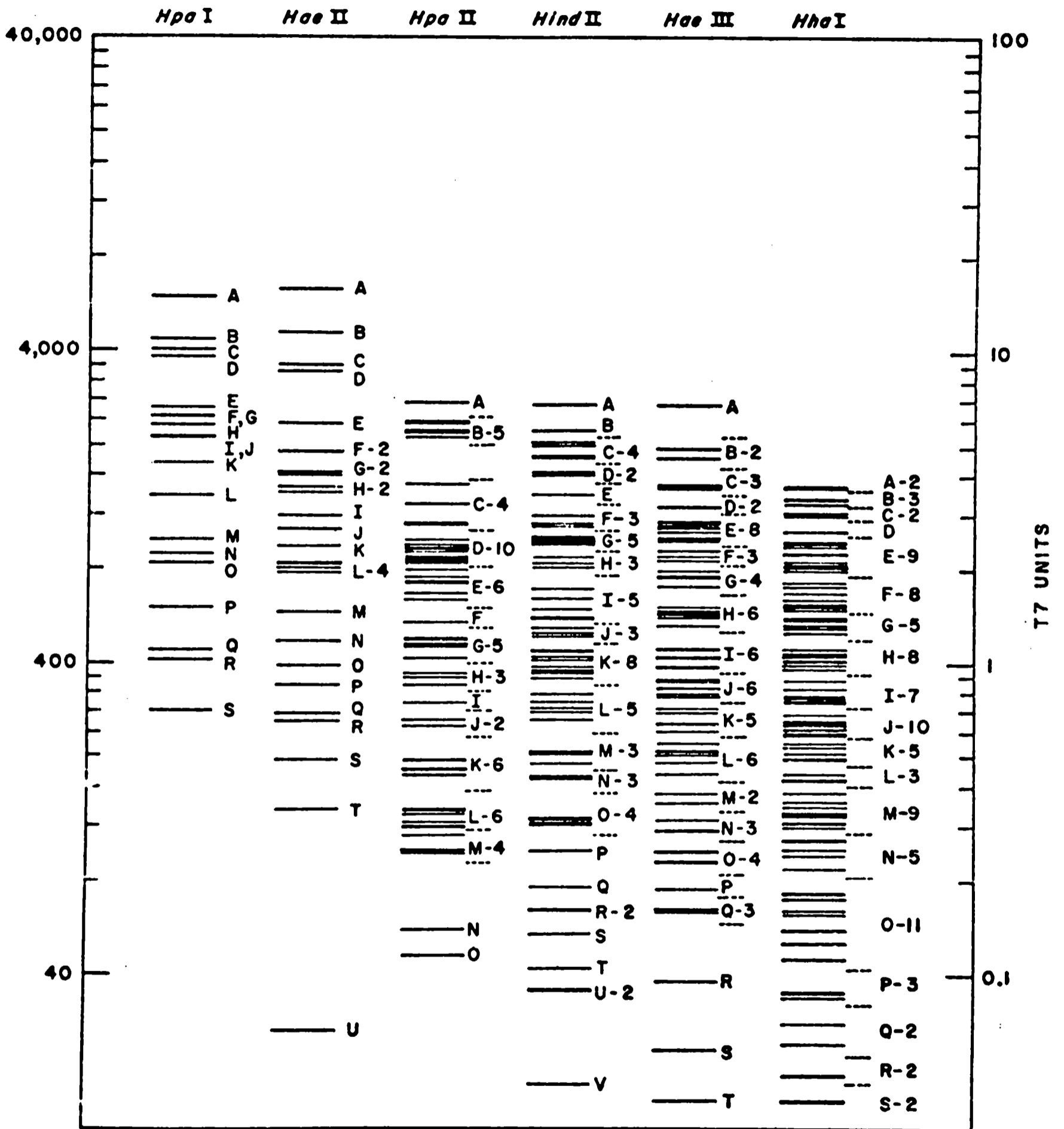
iii) For ease in comparing results. Studier's system of nomenclature for restriction fragments has been adopted (F.W. Studier, personal communication). This is shown for several different endonucleases in Fig. 10. Each size class is assigned a capital letter from the alphabet (using A for the largest, B for the next largest, etc.) and an arabic number for each fragment within the size class (1 for the largest, 2 for the next largest, etc.). Relative mobilities can vary with the gel system and running temperature (McDonnell et al, 1977; Shinnick et al, 1975), but, except as noted, this phenomenon has not significantly affected the use of the above nomenclature.

iv) Restriction endonuclease digests of DNA will be abbreviated DNA/EndoR. (e.g. T7 WT/HpaII). The resulting restriction fragments will be termed EndoR.A, EndoR.B, etc. (e.g. HpaII A, HpaII B, etc.). Digestion of one of

Figure 10. T7/EndoR. Fragment Sizes and Nomenclature

The band-patterns obtained upon polyacrylamide gel analysis of restriction endonuclease digests of T7 WT DNA are shown for the enzymes HpaI, HaeII, HpaII, HindII, HaeIII and HhaI. This diagram indicates the sizes of the resulting fragments in base-pairs and in % T7 (1% T7 = 400 bp), as well as a systematic nomenclature assigning each fragment a name as detailed in the text. This information was communicated by F.W. Studier, prior to publication.

Figure 10.



these fragments by a second EndoR. will be abbreviated EndoR₁A/EndoR₂ (e.g. HpaII A/HaeII).

v) A new band or fragment present in the digest pattern of a deletion mutant, but not that of T7 WT, will be designated Y^o when it migrates faster than the smallest size class (Y), or X' when it migrates with or just above size class X in the T7 Wt digest. Where two new fragments are observed, using the same EndoR., but different deletions, which both migrate faster than the smallest size class, the smaller of the two will be designated Y^{oo} and the larger, Y^o. Similarly, X' and X'' will be used for the faster and slower migrating fragments respectively, when two deletions yield two different new fragments migrating with or just above size class X.

vi) Recutting of a purified fragment with a second EndoR. (e.g. HpaII.A/HaeII), normally yields two fragments not present in the limit digest of total DNA produced by this second enzyme (i.e. T7 WT/HaeII). Such fragments will be designated 'heterospecific' to indicate that they were produced by the action of two endonucleases with different recognition specificities.

vii) Size estimates for digests containing fragments generally smaller than 5000 bp were based on information presented in Fig. 10 (McDonnell et al, 1977; F.W. Studier, unpublished results). These estimates were closely confirmed when HaeII digests of the plasmid pBR322 (recently sequenced by G. Sutcliffe, unpublished results) were used as standards (Figure 11). For those digests containing fragments generally larger than 300 bp, HindIII or EcoRI

Figure 11. Comparison of a T7 WT/HaeIII digest with
pBR322/HaeII

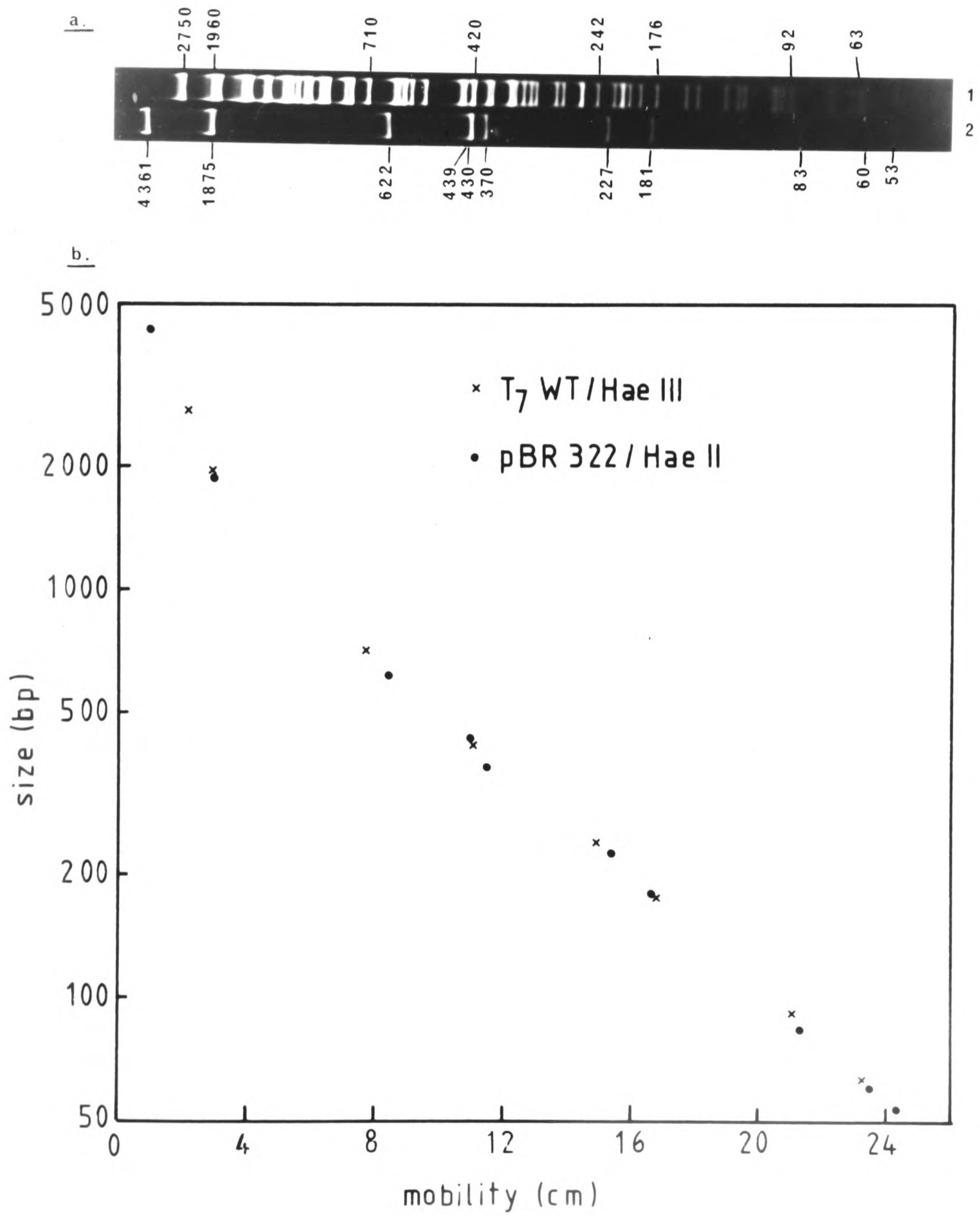
- a. Analysis of T7 WT/HaeIII and pBR322/HaeII digests by electrophoresis on a gradient polyacrylamide gel. Voltage was 4.7 V/cm for 16 h.

Track 1 3 μ g T7 WT/HaeIII

2 0.5 μ g pBR322/HaeII; 0.5 μ g pBR322/EcoRI

- b. A plot of mobility (in a.) vs. size for the pBR322/HaeII fragments and a selection of T7 WT/HaeIII fragments of comparable size. pBR322/HaeII sizes were derived from the complete sequence of pBR322 (G. Sutcliffe, unpublished results). T7 WT/HaeIII size estimates were as given by F.W. Studier (Fig. 10, personal communication).

Figure 11.



digests of various strains of bacteriophage λ were normally used as standards (Newman et al, 1979; Gottesman and Adhya, 1977). The fragment sizes so derived were normalised to agree with the present estimate of 40,000 bp as the size of T7 WT DNA (McDonnell et al, 1977).

viii) Densitometric scanning was used to determine the number of fragment species in any band which appeared darker than expected on gel photographs (except as noted in section c, below).

ix) The amounts of DNA used in fragment-recutting analyses are not given in figure legends as the concentration of DNA in purified fragment preparations was not normally determined. The same is true of plate-lysate DNA preparations.

x) The smallest fragments detected by ethidium bromide staining of agarose gels were about 200 bp, and for polyacrylamide, about 40 bp; smaller fragments may have gone undetected. It is however, highly unlikely that this would have significantly affected the accuracy of mapping. Note that the 'H' reaction in sequencing experiments did not reveal any previously undetected EndoR. cutting sites (chapter 6).

xi) To aid in interpreting the results presented in this chapter, Figure 27 should be folded out and used as a ready reference. This figure combines the mapping conclusions of Studier (Figure 6) with those which have been independently derived from the results which follow (cf. Fig. 25 and 26).

These mapping results will be dealt with in three groups: those enzymes which make no cuts in T7 WT DNA, those which make 1 to 9 and those which make more than 9. Within each group, enzymes will be treated alphabetically.

1. Enzymes which do not cut

T7 WT DNA incubated with the following EndoR's under optimal conditions for digestion (see Table 4), comigrated with undigested T7 WT DNA on agarose gel electrophoresis, with no other detectable bands: BamI, BglI, EcoRI, HindIII, PstI, Sall, SmaI and XhoI.

Parallel positive control experiments using phage λ DNA which is known to be cut by these enzymes (Roberts, 1978) gave limit digests in the case of BamI, EcoRI, HindIII, PstI and Sall. Parallel controls were not run in the case of BglI, SmaI and XhoI, but these enzymes were concurrently active in their donors' experiments.

2. Enzymes which make between 1 and 9 cuts

T7 WT DNA digested with the following EndoR.'s under optimal conditions migrates as a specific pattern of bands on gel electrophoresis, representing 2-10 fragment species: AvaI, BclI, DpnI, KpnI and XbaI.

a. AvaI (Figure 12)

AvaI digestion of T7 WT DNA yields 5 equimolar bands, A, B, C, D and E (Fig. 12a). Comparison with standards indicate that these correspond to DNA-fragments of 10540, 10270, 10000, 5320 and 3870 bp, respectively. AvaIA alone is missing from T7 LG37 and T7 C5LG3 digests,

being replaced by D" (8660 bp) and D' (7125 bp), respectively. Since even AvaI E is too large to fit between the C5 deletion and the left end of T7, AvaI A must be the leftmost AvaI fragment, and thus the leftmost AvaI site maps at about 26.3.

AvaI D/HpaII digests yield several fragments, one of which comigrates with HpaII C1 (Fig. 12b). This latter fragment was identified by Ludwig and Summers (1975) as containing the right end of T7 WT DNA. This suggests that AvaI D is the rightmost AvaI fragment. Alternatively, one of the AvaI D/HpaII heterospecific fragments might fortuitously comigrate with HpaII C1. F.W. Studier has mapped the AvaI sites to 26.2, 39.8, 65.4 and 75.3, making the order of the fragments A-D-B-E-C (personal communication). However my AvaI B+C/HpaII digests did not contain a band comigrating with HpaII C1 (Fig. 12b). This suggests that neither AvaI B nor AvaI C is the rightmost AvaI fragment. The results presented here therefore indicate the order A-(B, C, E)-D. In the absence of Studier's detailed evidence, no further evaluation of this discrepancy can be made.

b. BclI (Fig. 13)

BclI digestion of T7 WT DNA yields two equimolar bands on gel electrophoresis, A and B. Comparison with standards indicates the smaller band B corresponds to DNA fragments of about 8280 bp, making A, therefore, about 31700 bp. T7 LG37 and T7 C5LG3 digests both lack B, and have in its stead B^o (6220 bp) and B^{oo} (4940 bp), respectively. BclI B, therefore, is the lefthand BclI

Figure 12. Mapping of AvaI cutting sites

- a. Analysis of T7/AvaI digests by electrophoresis on a horizontal agarose (0.5%) gel. Voltage was 1 V/cm for 26 h and then 0.5 V/cm for 14 h. (McDonnell et al (1977) and Shinnick et al (1975) report that low voltage gradients such as these give better resolution of large fragments of DNA.)

Track 1. 0.2 μ g λ cI857 + λ rif^d18/HindIII
2. 0.15 μ g T7 LG37/AvaI
3. 0.15 μ g T7 C5LG3/AvaI
4. 0.15 μ g T7 WT/AvaI

Size estimates of T7/AvaI fragments were derived by comparison with the λ /HindIII digest in track 1. (Gottesman and Adhya, 1977; Newman et al, 1979).

- b. Analysis of HpaII digests of purified T7 C5LG3/AvaI fragments by electrophoresis on a gradient polyacrylamide gel. Voltage was 5 V/cm for 18 h. Bpb migrated off the gel shortly before the run was stopped.

Track 1. 2 μ g T7 C5LG3/HpaII
2. AvaI D/HpaII
3. AvaI D'/HpaII
4. AvaI B + C/HpaII

Figure 13. Mapping of BclI cutting sites

Analysis of T7/BclI digests by electrophoresis on a horizontal agarose (0.5%) gel (details as in Fig. 12).

Track 1. 0.2 μ g λ cI857 + λ rif^d18/HindIII
2. 0.15 μ g T7 WT/BclI
3. 0.15 μ g T7 C5LG3/BclI
4. 0.15 μ g T7 LG37/BclI
5. 0.15 μ g T7 LG26/BclI

Size estimates of the T7/BclI fragments were derived by comparison with the λ /HindIII digest in track 1 (Gottesman and Adhya, 1977; Newman et al, 1979).

Figure 12.

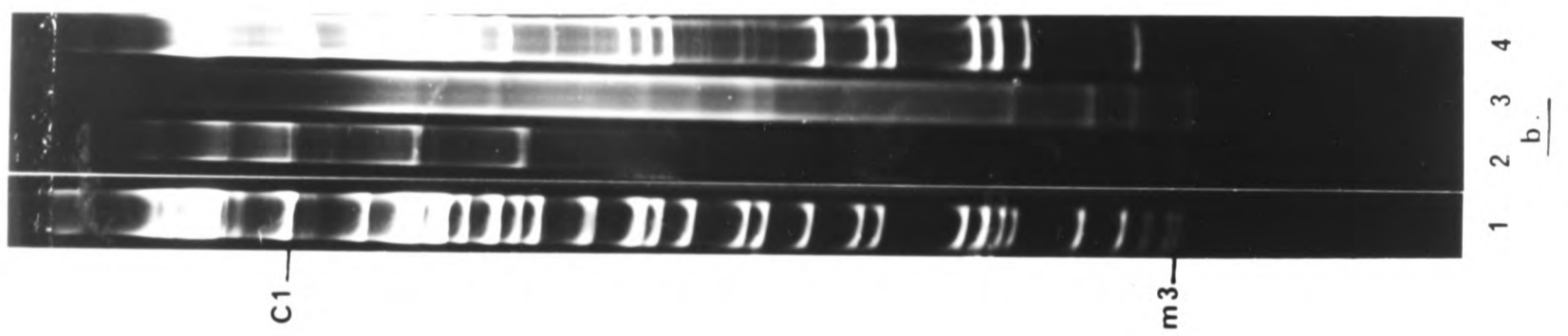
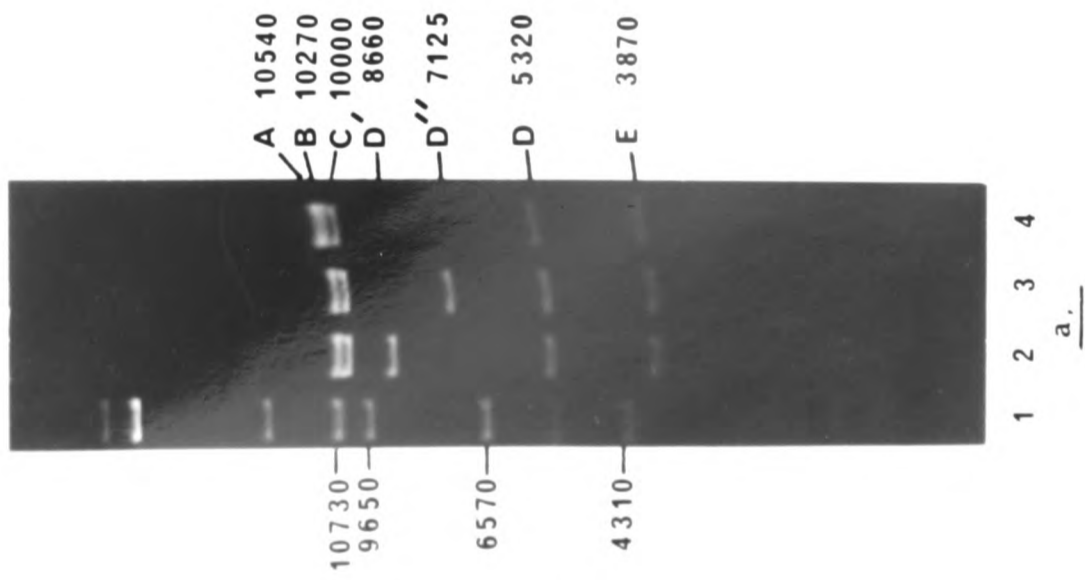
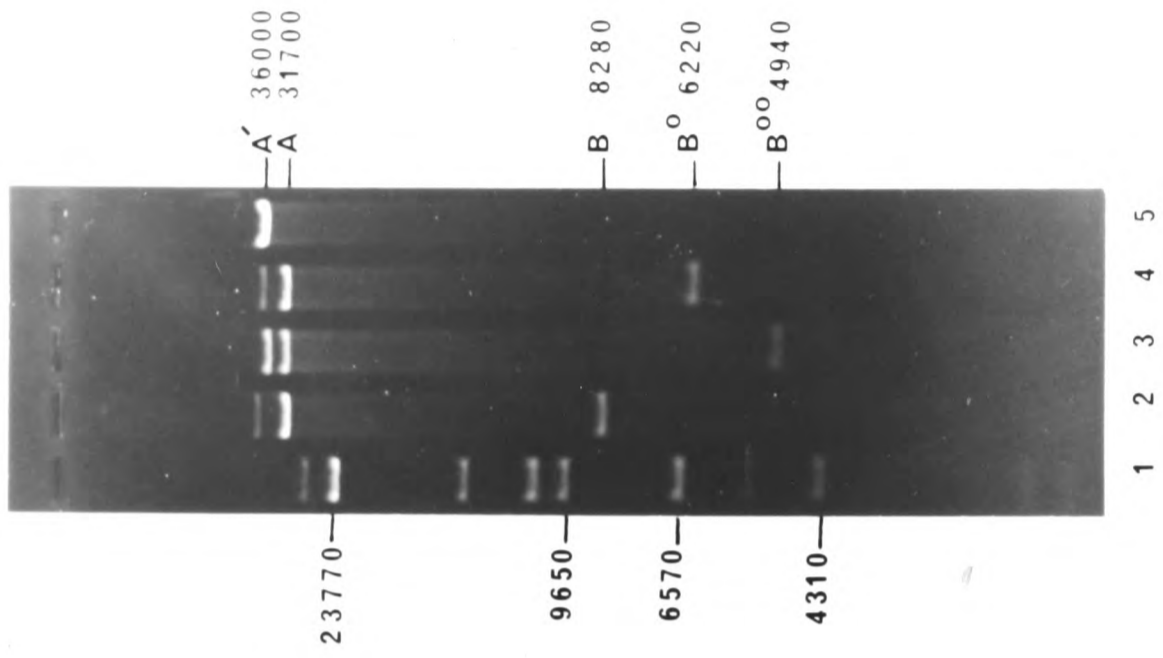


Figure 13.



fragment, placing the single BclI cutting site at about 20.7.

T7 LG26/BclI digests lack both A and B and have in their stead a single new band A' (about 36000 bp). The Δ LG26 mutation therefore deletes the sole BclI site, consistent with their respective mapped positions (Δ LG26 deletes from 17.75-18.2 to 21.43-21.88, see Table 7).

c. KpnI (Fig. 14)

KpnI digestion of T7 WT DNA yields 4 equimolar bands on gel electrophoresis, A, B, C and D. Size standards indicate that these correspond to DNA fragments of 16100, 14650, 5650 and 3590 bp, respectively. T7 LG37/KpnI digests lack only fragment D and have in its place a new band D^{oo}, of about 1750 bp. The region deleted by Δ LG37 is therefore wholly contained within fragment D. T7 C5LG3/KpnI digests lack both KpnI C and D, and have instead two new fragments, D' (3800 bp) and D^o (2250 bp). Each deletion is therefore wholly contained within one of the missing fragments. By analogy with T7 LG37, it must be KpnI D which wholly contains Δ LG3, and thus KpnI C wholly contains Δ C5. The order of the fragments is, therefore, C-D-(A,B), placing the three KpnI cutting sites at 14.1, 23.1 and 59.7 or 63.3

d. MboI (Fig. 15a)

MboI digestion of T7 WT DNA, analysed in such a way that only fragments larger than about 1000 bp would be detected, yielded five bands, A, B, C, D and E. Comparison with the T7 C5LG3/AvaI digest in the adjacent track indicated sizes of >15000, 8000, 3800, 3150 and 2850 bp,

Figure 14. Mapping of KpnI cutting sites

- a. Analysis of T7/KpnI digests by electrophoresis on a horizontal agarose (1%) gel. Voltage was 2 V/cm for 20 h.

Track 1. 1 μ g AN λ 17/EcoRI
2. 1 μ g T7 WT/KpnI
3. 0.8 μ g T7 LG37/KpnI
4. 1 μ g T7 C5LG3/KpnI

- b. Analysis of T7/KpnI digests by electrophoresis on a horizontal agarose (0.5%) gel (details as in Figure 12).

Track 1. 0.15 μ g T7 LG37/KpnI
2. 0.15 μ g T7 C5LG3/KpnI
3. 0.15 μ g T7 WT/KpnI
4. 0.2 μ g λ cI857 + λ rif^d18/HindIII

Size estimates for T7/KpnI fragments C and D were obtained by comparison with the λ /EcoRI digest (track a 1, Newman et al 1979); and for A and B, by comparison with the λ /HindIII digest (track b 4, Gottesman and Adhya, 1977; Newman et al, 1979).

Figure 15. Mapping of MboI and XbaI cutting sites

- a. Analysis of a T7 WT/MboI digest by electrophoresis on a short, vertical agarose (1.2%) gel. Voltage was 10 V/cm for 2.5 h.

Track 1. 0.3 μ g T7 C5LG3/AvaI
2. 0.15 μ g T7 WT/MboI

Size estimates of the T7 WT/MboI fragments were derived by comparison with the T7 C5LG3/AvaI digest in track 1, (cf. Fig. 12).

- b. Analysis of T7/XbaI digests by electrophoresis on a long, vertical agarose (1%) gel. Voltage was 7 V/cm for 17 h.

Track 1. 0.5 μ g T7 WT/AvaI
2. 0.4 μ g T7LG37/XbaI
3. 0.5 μ g T7 WT/XbaI

Figure 14

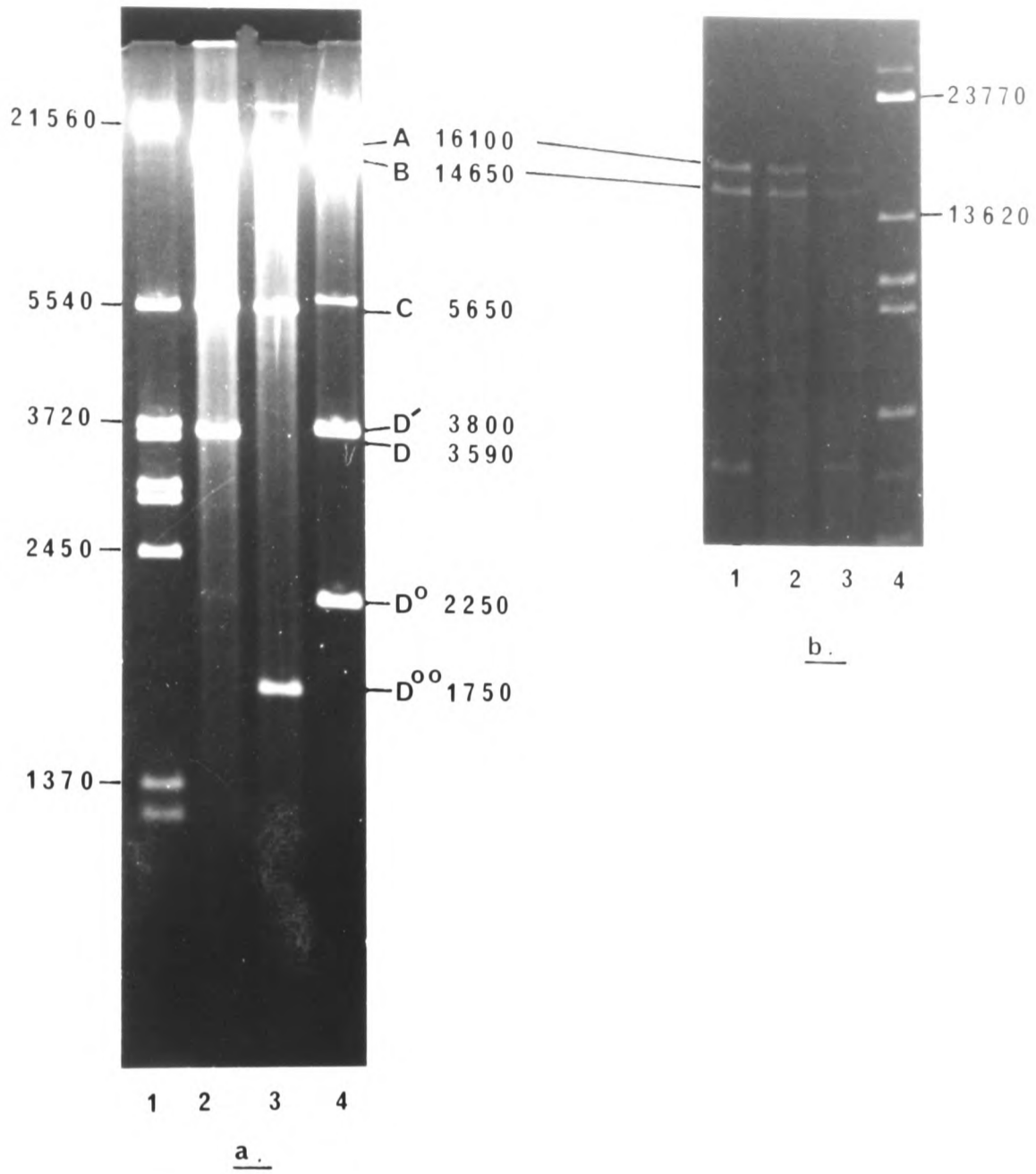
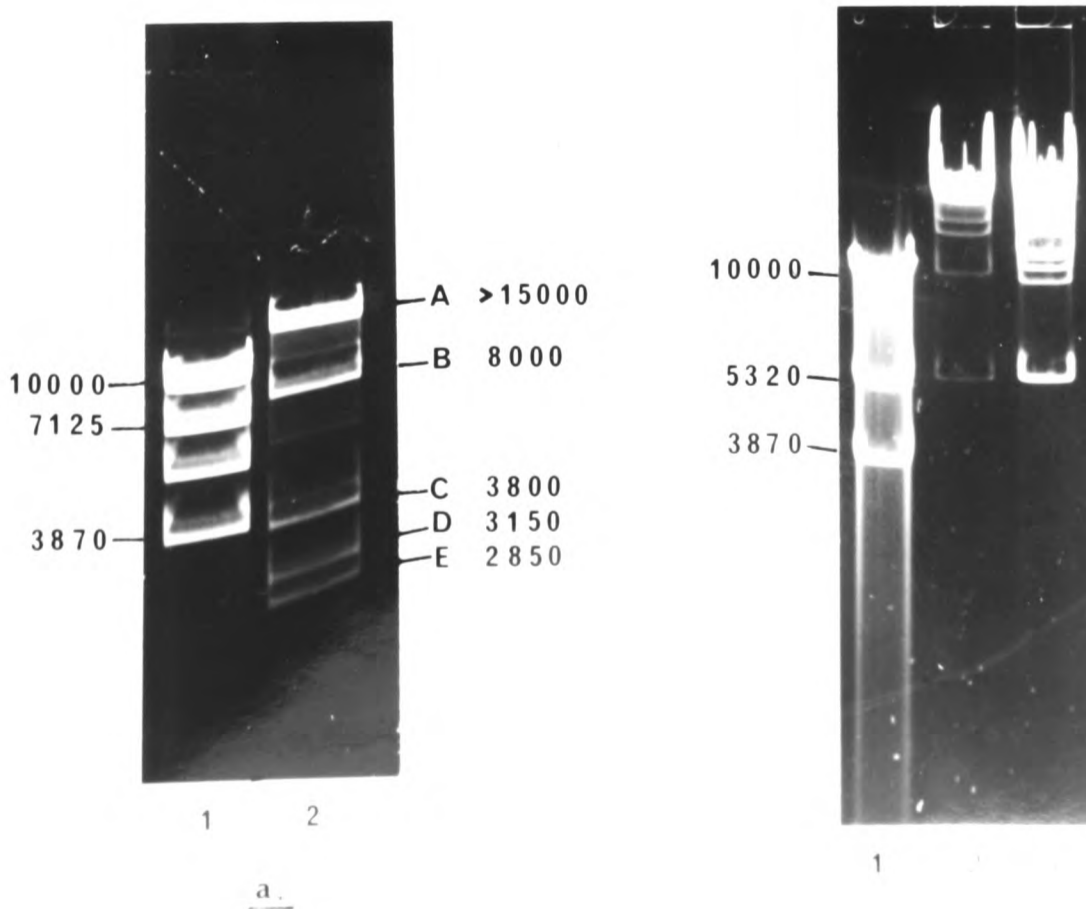


Figure 15.



which are consistent with the results of McDonnell et al, (1977): 21428, 8268, 3820, 3116 and 2840 bp, respectively. The two smaller fragments, F (408 bp) and G (120 bp), reported by these authors, would have run off the gel shown. Scarcity of enzyme prevented a more complete analysis.

e. XbaI (Fig. 15b)

A single attempt to investigate XbaI cutting of T7 WT DNA yielded a partial digest containing only a few, slowly migrating bands on agarose gel electrophoresis. The fastest of these represented DNA fragments of about 5000 bp. The number of bands observed in these digests is consistent with only a few (less than 10) cutting sites in T7 WT DNA. Scarcity of enzyme prevented a more complete analysis.

3. Enzymes which make more than 9 cuts

T7 WT DNA, digested with the following enzymes, migrated as a specific pattern of bands on gel electrophoresis, representing more than 10 species of fragment: HaeII, HaeIII, HgiAI, HhaI, HinII, HpaI and HpaII. The results obtained with these enzymes are presented below. Wherever relevant, a comparison is made with the results obtained by F.W. Studier (personal communication, Fig. 6). Where regions of the gel contained 3 or more comigrating fragments (i.e. in a single band), densitometric scanning was not felt to be sufficiently reliable to determine the actual number of fragments present. Studier's conclusions were then assumed correct.

a. HaeII (Fig. 16)

The pattern obtained on HaeII digestion of T7 WT DNA (Fig. 16a) was indistinguishable from that of Studier, except that the smallest of his 27 fragments, U, was not detected on any gel. T7 C5LG3/HaeII digests lack fragments C (3660 bp), E (2360 bp), N (475 bp), O (395 bp) and S (195 bp) and in their stead have two new bands, F'' (2150 bp) and G' (1600 bp). T7 LG37/HaeII digests similarly lack fragment C (3660 bp) as well as Q (285 bp), and show a single new fragment F' (2000 bp). Since Δ LG37 extends beyond Δ LG3 only on the right, Q must be adjacent to and rightward of C, while C must wholly contain the region deleted by Δ LG3. The difference between the sizes of C and F'' (3660 - 2150 = 1510 bp) is reasonably close to the estimated size of Δ LG3 (1426 bp, Table 7) whereas the difference between C and G' greatly exceeds it (2060 bp). Δ LG3, therefore is responsible for the new fragment F'', and thus Δ C5 eliminates fragments E, N, O and S and creates G'.

AvaI D'/HaeII digests (AvaI.D' being the new fragment in T7 C5LG3/AvaI digests) yield F'', F1 (1920 bp), G', M (585 bp), P (335 bp), Q and S' (250 bp) (Fig. 16b, c). This is completely consistent with Studier's map (Fig. 6) and further predicts that adjacent to and rightward of HaeII F1 lies HaeII P, followed 250 bp further to the right by the leftmost AvaI site at 26.2

HpaII C2/HaeII digests yield three fragments of 900, 285, and 100 bp (Fig. 16d). The 285 bp fragment comigrates with and is presumed identical to HaeII Q. The other two

Figure 16. Mapping of HaeII cutting sites

- a. Analysis of T7/HaeII digests by electrophoresis on a gradient polyacrylamide gel. Voltage was 5 V/cm for 18 h.

Track 1. 1 μ g T7 WT/HaeII
2. 1 μ g T7 LG37/HaeII
3. 1 μ g T7 C5LG3/HaeII

- b. Analysis of a T7 C5LG3/AvaI D'/HaeII digest by electrophoresis on a horizontal agarose (1%) gel. Voltage was 1.5 V/cm for 18 h.

Track 1. 1 μ g T7 WT/HaeII
2. 2 μ g T7 C5LG3/HaeII
3. AvaI D'/HaeII

- c. Analysis of a T7 C5LG3/AvaI D'/HaeII digest by electrophoresis on a short, thin polyacrylamide (4%) gel. Voltage was 2.5 V/cm for 15 h.

Track 1. 1 μ g T7 WT/HaeII
2. AvaI D'/HaeII
3. 2 μ g T7 C5LG3/HaeII

- d. Analysis of HpaII C2/HaeII digests by electrophoresis on a long, thin, polyacrylamide (8%) gel. Voltage was 3.5 V/cm for 16 h and then 4.5 V/cm for 5 h.

Track 1. 1.5 μ g T7 WT/HaeII
2. HpaII C2/HaeII

Track 2 represents the extreme edges of a much wider track as this gel was used to prepare HpaII C2/HaeII fragments for sequencing.

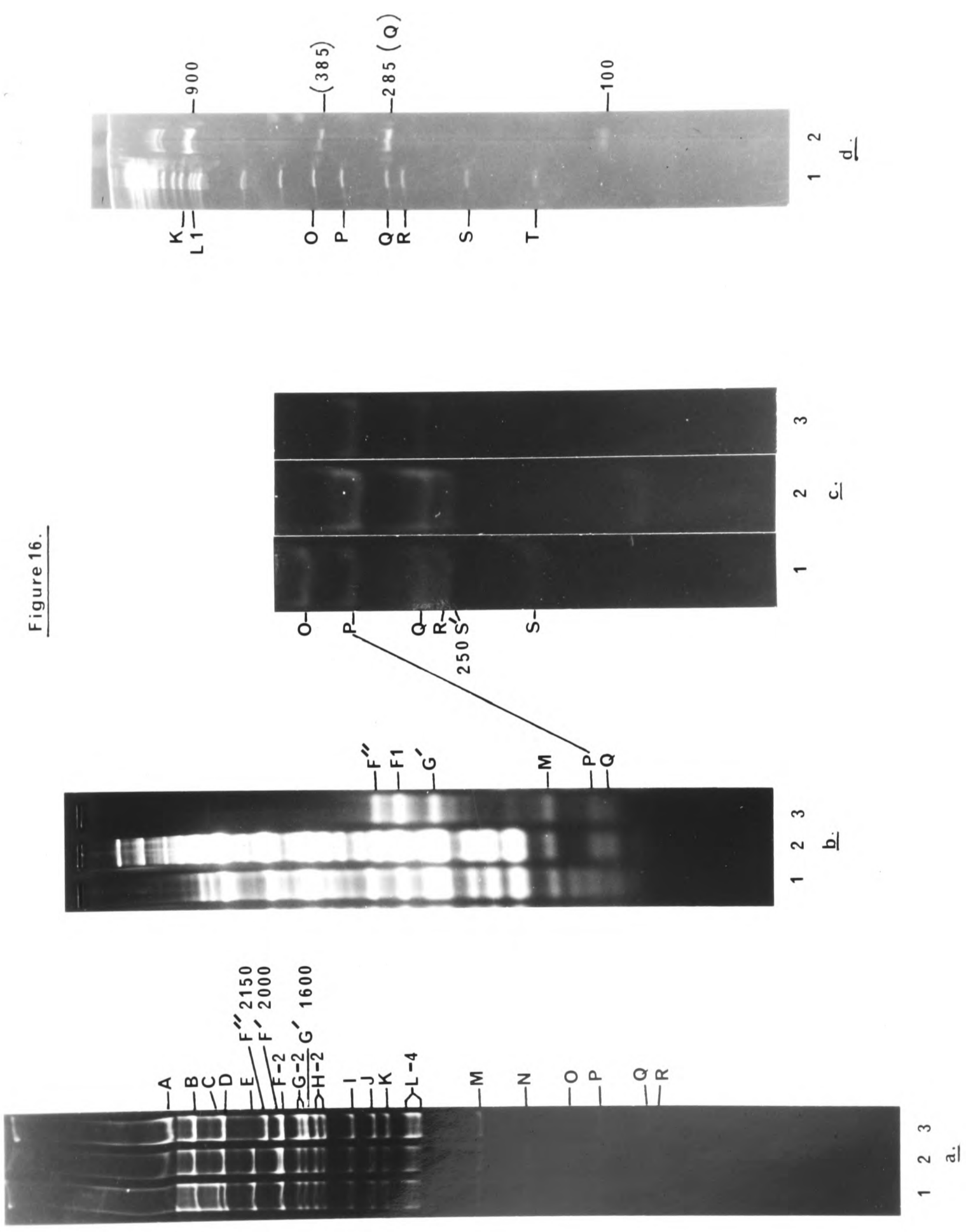


Figure 16.

fragments are therefore heterospecific. HpaII C2 can extend no further than 400 bp beyond the right limit of Δ LG37, since 400 bp is the size of the new fragment in T7 LG37/HpaII digests (see below). Moreover, this limit of Δ LG37 lies within HaeII Q. The size of the right heterospecific fragment in HpaII C2/HaeII must therefore be less than 400 bp. The order of the HpaII C2/HaeII fragments is, therefore, 900 -Q(285)-100. (Consistent with this interpretation there is a 385 bp fragment detectable in subequimolar amounts in Fig. 16d; this is the size predicted for a partial digestion product comprising the 285 and 100 bp fragments.)

b. HaeIII (Fig. 17)

The band pattern obtained on HaeIII digestion of T7 WT DNA (Fig. 17a) was indistinguishable from that of Studier except that the three smallest of his 68 fragments (R, S and T) were not detectable on any gel. T7 C5LG3/HaeIII digests lack fragments C1 (1500 bp), D1 (1270 bp), E1 (1140 bp), H5 (550 bp), J1 (350 bp), J3 (345 bp), J5 (330 bp), K3 (250 bp) and M1 (150 bp), and in their place have two new bands D' (1430 bp) and E' (1220 bp). T7 LG37/HaeIII digests lack fragments A (2750 bp), E1, H5, J3, J5, K3 and L2 (210 bp), and have instead a new band, A' (3800 bp). L2 and A, therefore lie outside and to the right of Δ LG3 in the order L2-A because L2 and not A is smaller than the size of Δ LG37 (1945 bp, see Table 7). On the basis of the estimated size of Δ C5 (1844 bp) and Δ LG3 (1426 bp), and given the fragments affected by these deletions, D' must be produced by Δ C5

Figure 17. Mapping of HaeIII cutting sites

- a. Analysis of T7/HaeIII digests by electrophoresis on a gradient polyacrylamide gel. Voltage was 5 V/cm for 18 h.

Track 1. 2.5 μ g T7 WT/HaeIII
2. 2.5 μ g T7 LG37/HaeIII
3. 2.5 μ g T7 C5LG3/HaeIII

- b. Analysis of a T7 C5LG3/AvaI D'/HaeIII digest by electrophoresis on a gradient polyacrylamide gel. Voltage was 5 V/cm for 16 h. Mobility of HaeIII D' is also more clearly shown than in a. (above).

Track 1. 2 μ g T7 C5LG3/HaeIII
2. AvaI D'/HaeIII

- c. Analysis of T7 C5LG3/AvaI purified fragments recut with HaeIII by electrophoresis on a gradient polyacrylamide gel. Voltage was 4.5 V/cm for 35 h. (stacking gel 2.4%) buffer was TPE, 36 mM Tris base, 30 mM NaH_2PO_4 , 1 mM EDTA, pH 7.7.

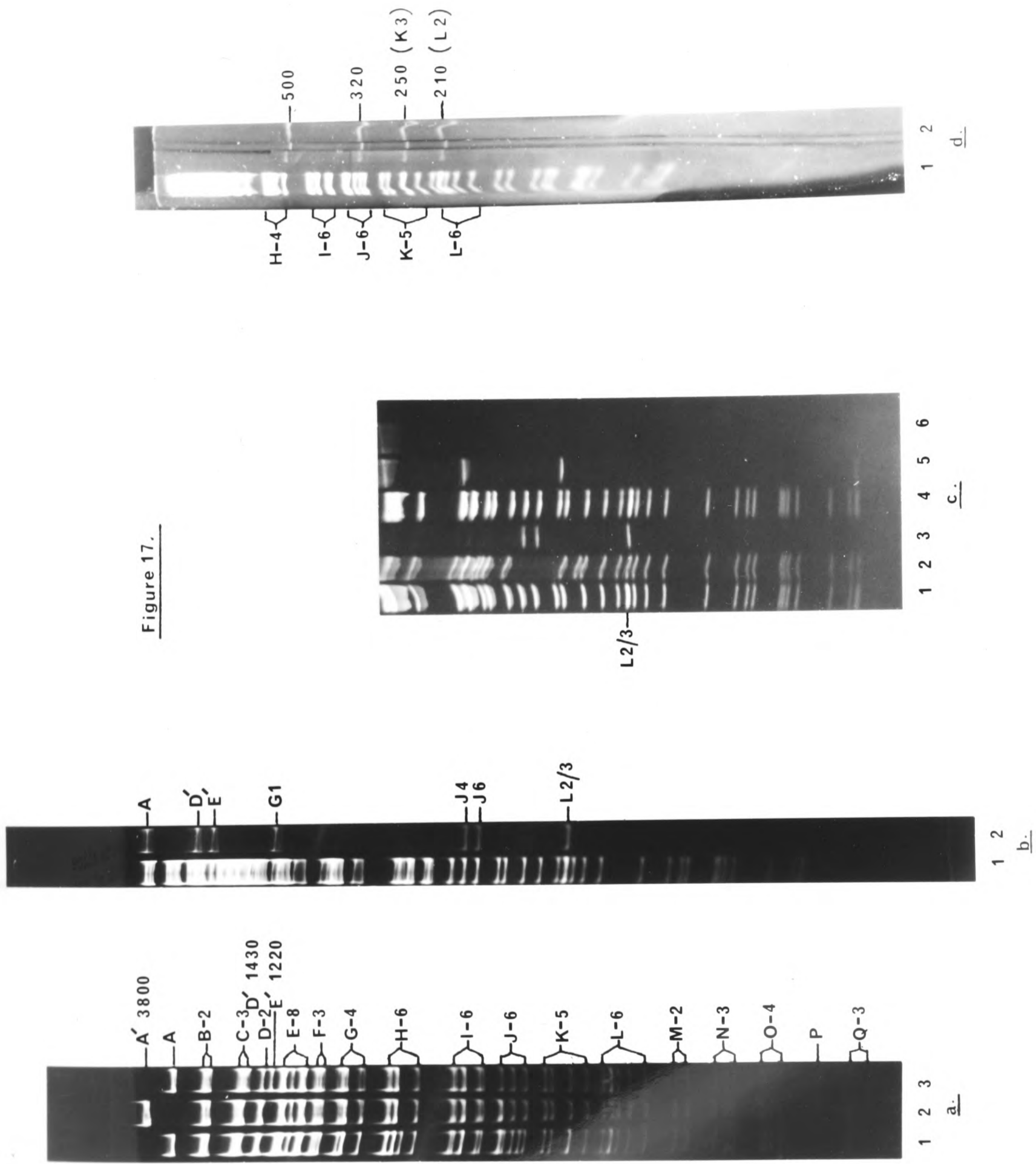
Track 1. 2.1 μ g T7 C5LG3/HaeIII
2. AvaI B + C/HaeIII
3. AvaI D'/HaeIII
4. 2.1 μ g T7 C5LG3/HaeIII
5. AvaI D/HaeIII
6. AvaI E/HaeIII

- d. Analysis of HpaII C2/HaeIII digest by electrophoresis on a long, thin polyacrylamide (8%) gel. Voltage was 3 V/cm for 18 h.

Track 1. 1.5 μ g T7 WT/HaeIII
2. HpaII C2/HaeIII

Track 2 represents the extreme edges plus a thin slice from the middle of a much wider track, as this gel was used to prepared HpaII C2/HaeIII fragments for sequencing.

Figure 17.



(because HaeIII C1 + D1 + J1 + M1 - D' (1430 bp) = 1840 bp) and E' by Δ LG3 (HaeIII E1 + H5 + J3 + J5 + K3 - E' (1220 bp) = 1395 bp).

T7 C5LG3/AvaI D'/HaeIII digests yield A, D', E', G1, J4, J6 and a double band at the L2/3 position (Fig. 17b). Alone among these (excluding D' and E'), the second fragment at L2/3 is not given by Studier among the HaeIII fragments lysing to the left of 25.7 (Fig. 6), where he considers his mapping to be complete. Therefore, either HaeIII L3 lies within the region between HaeIII A (endpoint, 25.7) and the AvaI site at 26.2, with a tiny, undetectable heterospecific fragment being produced, or else the second fragment in the L2/3 position is itself the heterospecific fragment. AvaI B + C/HaeIII digests contain a fragment in the L2/3 position, supporting the second alternative above (Fig. 17c).

HpaII C2/HaeIII digests yield four fragments, 500, 320, 250 and 210 bp (Fig. 17d). HaeIII K3 and L2 comigrate with and are presumed identical to the 250 and 210 bp fragments, respectively. The other two must be, therefore, the flanking heterospecific fragments. Since Δ LG37 affects both HaeIII K3 and L2, whereas Δ LG3 only affects K3, and since HpaII C2 contains the right limit of both deletions (see below), HaeIII L2 lies adjacent to and rightward of HaeIII K3, giving the order K3-L2-A.

The new fragment in T7 LG37/HaeIII digests, HaeIII A', is 1050 bp larger than HaeIII A, so that the left limit of Δ LG37 must lie in HaeIII E1 (at 1140 bp, E1 is the only other fragment affected by LG37 which is larger

than 1050 bp), within 90 (1140 minus 1050) bp of its right end. Since HaeIII E' is 1220 bp, the right limit of the relevant deletion, Δ LG3, must lie within 170 (1220 minus 1050) bp of the right end of HaeIII K3 (assuming that Δ LG3 and Δ LG37 have identical left limits). Accordingly, if the 320 bp band in HpaII C2/HaeIII were the right heterospecific fragment, the right limit of Δ LG3 would have to lie within 700 (170 + 210 (L2) + 320) bp of the right end of HpaII C2. However, the HinII results presented below imply that the right limit of Δ LG3 lies more than 770 bp from this end of HpaII C2. The right heterospecific fragment must therefore be represented by the 500 bp band. (170 + 210 (L2) + 500 = 880 bp would then be the maximum distance between the right limit of Δ LG3 and the right end of HpaII C2, consistent with the HinII data.) The order of the HpaII C2/HaeIII fragments is therefore 320 - 250(K3) - 210 (L2) - 500 bp.

c. HgiAI (Figs. 18 and 19)

HgiAI digests of T7 WT DNA, analysed on gel electrophoresis, contain 23 fragments, ranging in size from 60 to 7150 bp (Fig. 18a). T7 C5LG3/HgiAI digests lack fragments A (7150 bp) and B (6250 bp) and in their place have a new fragment, A' (\sim 9000 bp). T7 LG37/HgiAI digests also lack A and B and have instead a new fragment, A'' (\sim 11000 bp). The single site affected therefore lies within the region deleted by Δ LG3, implying that HgiAI A and B are adjacent and contain most if not all of the T7 genome to the left of position 33.5.

T7 LG26/HgiAI digests (Fig. 18b) lack only fragment B,

Figure 18. Mapping of HgiAI cutting sites

- a. Analysis of T7/HgiAI digests by electrophoresis on a horizontal agarose (1.5%) gel. Voltage was 1 V/cm for 16 h, and then 2 V/cm for 3h.

Track 1. 1.5 μ g T7 WT/HpaI
2. 1.5 μ g T7 WT/HgiAI
3. 1.5 μ g T7 LG37/HgiAI
4. 2.5 μ g T7 C5LG3/HgiAI

- b. Analysis of T7/HgiAI digests by electrophoresis on a horizontal agarose (1%) gel. Voltage was 2 V/cm for 19 h.

Track 1. 1 μ g T7 LG26/HgiAI
2. 1.8 μ g T7 WT/HgiAI

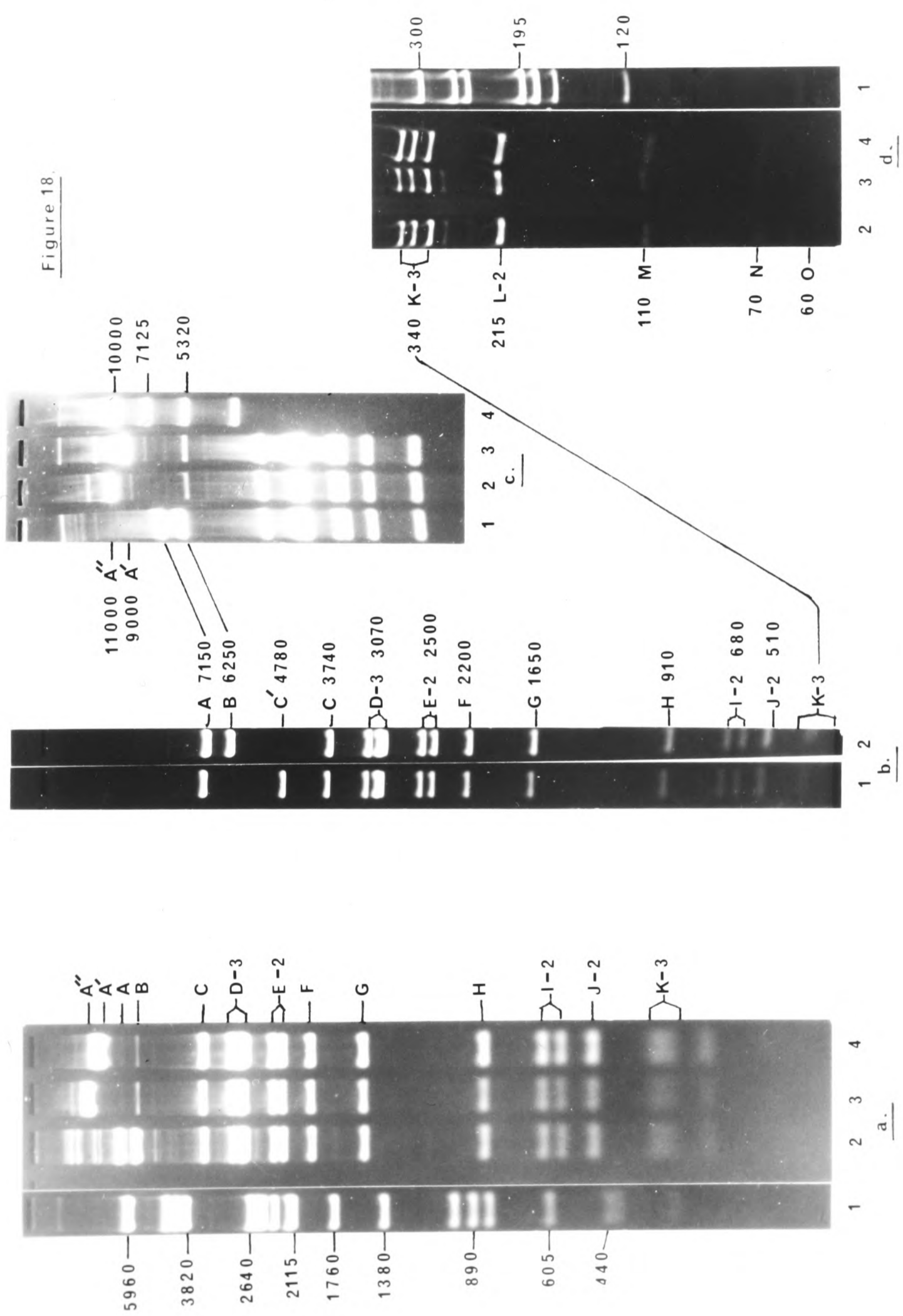
The sizes of T7/HgiAI fragments A'' and A' were estimated by comparison with the T7 C5LG3/AvaI digest in track 4, part c; those of fragments A through J2 by comparison with the T7 WT/HpaI digest in track 1, part a; and those of fragments M1 through 0 by comparison with the AvaI B + C/HpaII digest in track 4, part d.

- c. Analysis of T7/HgiAI digests by electrophoresis on a horizontal agarose (1%) gel. Voltage was 1 V/cm for 21 h.

Track 1. 1.5 μ g T7 WT/HgiAI
2. 1.5 μ g T7 LG37/HgiAI
3. 1.5 μ g T7 C5LG3/HgiAI
4. 0.6 μ g T7 C5LG3/AvaI

- d. Analysis of T7/HgiAI digests by electrophoresis on a gradient polyacrylamide gel. Voltage was 5 V/cm for 18 h.

Track 1. AvaI B + C/HpaII
2. 2 μ g T7 C5LG3/HgiAI
3. 2 μ g T7 LG37/HgiAI
4. 2 μ g T7 WT/HgiAI



and have in its stead a new fragment C' (4780 bp). B, therefore, lies to the right of A.

HgiAI A/HpaII digests (Fig. 19) contain a fragment comigrating with HpaII F which, according to Ludwig and Summers (1976) and Gordon et al (1978), is the leftmost HpaII fragment. HgiAI A, therefore, is the leftmost HgiAI fragment.

Two bands present in the HgiAI A/HpaII digest (300 and 620 bp) are not accounted for by Studier's HpaII map (Fig. 6). One is presumably the HgiAI/HpaII heterospecific fragment, and the other a partial digestion product. The smaller fragment is the preferred heterospecific candidate because: (i) no two adjacent HpaII fragments in this region add to 300 bp (L2 plus L3 come closest (275 bp), but they do not appear to be present in subequimolar amounts); (ii) HpaII F is present in subequimolar amounts, suggesting incomplete digestion, and combined with the adjacent HpaII M3 adds to 630 bp; (iii) HpaII L4 (16.6-16.9 on Studier's HpaII map, Fig. 6) is present in the HgiAI A/HpaII digest, but HpaII C2 (16.9-20.15) is not. The HgiAI site therefore maps at $16.9 + x/400$ where x is the size (bp) of the heterospecific fragment present in the digest. If $x = 300$, then the HgiAI site would map at about 17.7; if $x = 620$, then it would map at about 18.5. However, Δ LG3 which deletes this HgiAI site, extends no further to the right than 18.2 (Table 7); (iv) 17.7 is in closer agreement with the estimated size of HgiAI A (17.9% T7) than 18.5%. The single HgiAI site in the early region, therefore is near 17.7, with the next site near 33.3.

d. HhaI (Fig. 20)

The band pattern obtained on HhaI digestion of T7 WT DNA was indistinguishable from that reported by Studier, except that the six smallest of the 97 fragments (Q1,2, R1,2 and S1,2) were not visible on any gel.

T7 C5LG3/HhaI digests lack fragments A1 (1480 bp), E1 (990 bp), H1 (440 bp), I6 (295 bp), L1 (170 bp), L2 (165 bp), M4 (135 bp), O1 (75 bp) and O4 (65 bp). Only one new fragment, I' (345 bp), has been detected. T7 LG37/HhaI digests lack fragments A1, H8 (380 bp), L1 and M5 (130 bp) and instead have a new fragment J' (240 bp). Evidently Δ LG3 affects A1 and L1, while Δ LG37 extends to affect H8 and M5 on the right. Δ C5 therefore affects E1, H1, I6, L2, M4, O1 and O4. On the basis of size estimates for Δ LG3 and Δ C5 (Table 7), they should create new fragments of 225 and 320 bp respectively. Δ C5 is thus most likely to be responsible for the new fragment I'. The Δ LG3 created new fragment may lie concealed in a multifragment band.

T7 LG26/HhaI digests lack fragments C1 (1225 bp), H8, L1, M1 (150 bp) and M5. No new fragment has been detected. Since HhaI L1 is missing in both T7 LG26 and T7 LG37 digests, whereas A1 is missing only in T7 LG37, the fragments in this region can be further ordered as A1 - L1 - (M8, M5) - (M1, C1).

HpaII C2/HhaI digests contain five bands, 380, 170, 150, 130 and 110 bp. These add to only 940 bp suggesting that the 380 bp band (smeared on the gel shown in Fig. 20) is a doublet, thus bringing the sum of the sizes to about

Figure 19. Mapping of HgiAI (contd.)

Analysis of HgiAI A/HpaII digest by electrophoresis on a gradient polyacrylamide gel. Voltage was 4 V/cm for 24 h.

- Track 1. 2.5 μ g T7 WT/HpaII
- 2. HgiAI A/HpaII

Figure 20. Mapping of HhaI cutting sites

Analysis of T7/HhaI and HpaII C2/HhaI digests. Details as in Fig. 19.

- Track 1. HpaII C2/HhaI
- 2. 2.5 μ g T7 WT/HhaI
- 3. 2.5 μ g T7 WT/HhaI
- 4. 2.5 μ g T7 LG37/HhaI
- 5. 2.5 μ g T7 C5LG3/HhaI
- 6. 2.5 μ g T7 LG26/HhaI

Figure 19.

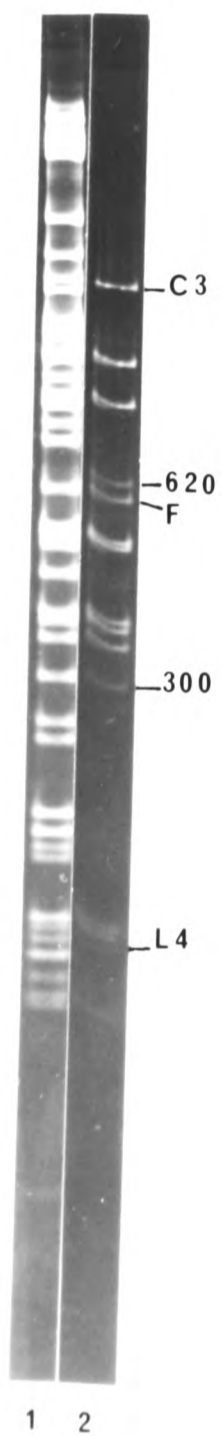
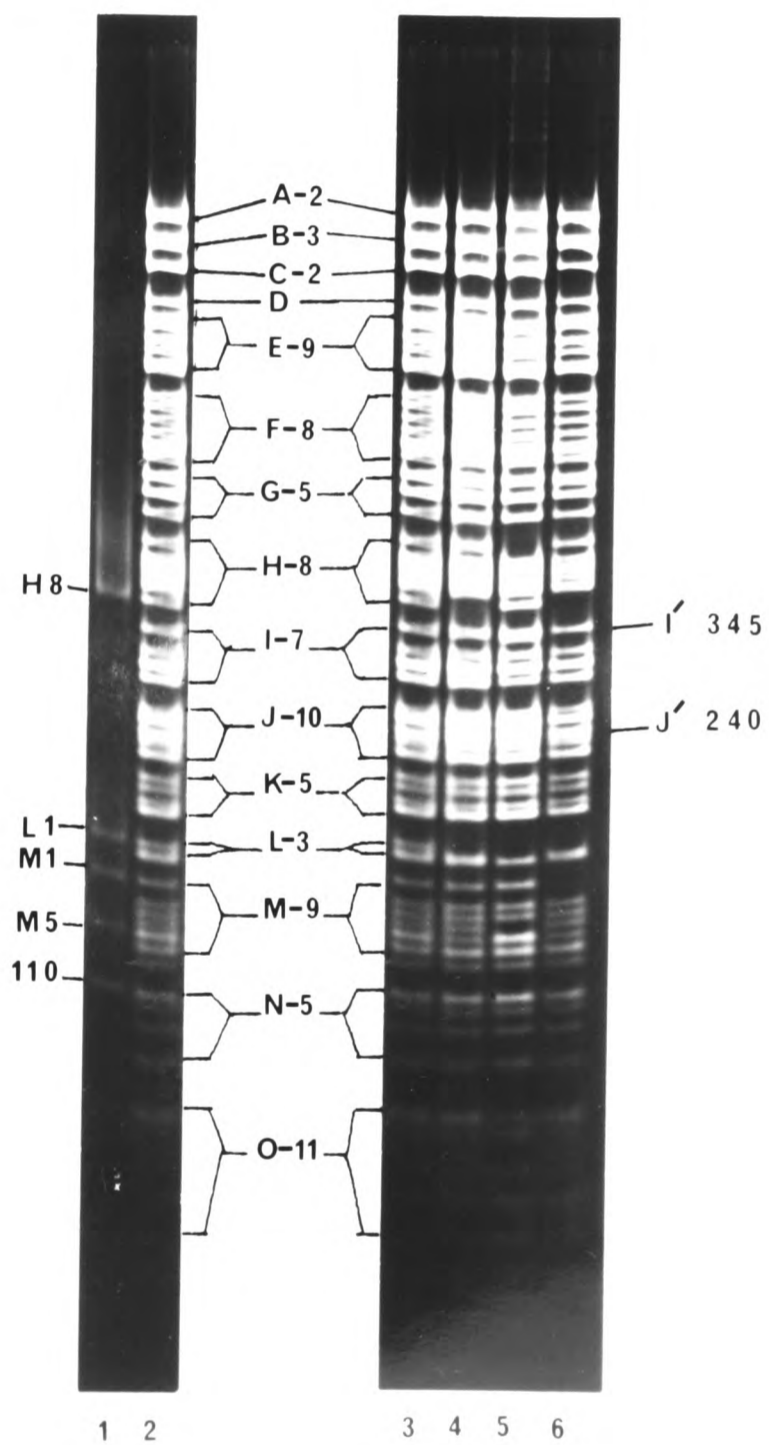


Figure 20.



1320 bp, close to the estimated size of HpaII C2 (1290 bp). The four largest bands comigrate with and are presumed to contain fragments identical to H8, L1, M1 and M5, respectively, making the other 380 bp and the 110 bp species the flanking heterospecific fragments. As HaeII recognition sites form a subset of HhaI sites (Table 3) the HpaII C2/HaeII fragments should be present as such, or subdivided in the HpaII C2/HhaI digest. The 110 bp HpaII/HhaI heterospecific fragment is therefore presumed to be identical to the 100 bp HpaII C2/HaeII heterospecific fragment (Fig. 16d). No 285 bp fragment (HaeII Q) is present in the HpaII C2/HhaI digest, but HhaI M1 and M5 add to 280 bp and are therefore probably generated by a HhaI site within the HaeII Q. The remaining HpaII C2/HhaI fragments, H8 and L1, and the 380 bp heterospecific fragment add to 930 bp, about the size of the left heterospecific fragment in the HpaII C2/HaeII digest (900 bp). Together with the deletion results given above, this suggests that the order of the HhaI fragments in the region of the transcription termination site is A1-L1-H8-M5-M1-C1.

e. HinII (Fig. 21)

HincII and HindII are isoschizomers differing only in their bacterial source, H.influenzae strains c and d, respectively. Strain c lacks the HindIII activity found in strain d: thus HincII preparations are generally preferable to HindII. However, as HindIII does not cut T7 WT DNA, HindII and HincII could be used interchangeably in the present work and will be considered collectively as HinII.

The gel pattern obtained on HinII digestion of T7 WT DNA was indistinguishable from that of Studier, except that the four smallest of his 60 fragments T, U1, U2 and V) were not detected (Fig.21). T7 C5LG3/HinII digests lack D1 (1660 bp), E (1400 bp), K2 (440 bp), K5 (385 bp) and K7 (365 bp) but contain instead two new fragments, I' (690 bp) and M' (215 bp). T7 LG37/HinII digests lack D1, G3 (1000 bp) and K1 (440 bp) and have in their stead a new fragment F' (1120 bp). HinII K1, missing from T7 LG37/HinII digests, is distinct from HinII K2, missing from T7 C5LG3/HinII digests, because HinII K1 is the arbitrary designation of the HinII fragment identical to HpaI Q, which is also missing from T7 LG37/HpaI digests but not those of T7 C5LG3/HpaI. Accordingly HinII D1 wholly contains the region deleted by Δ LG3. Moreover, M' is in the expected size range for a new fragment created from HinII D1 by Δ LG3 (i.e. $1660 - 1425 = 235$ bp). (Δ C5 must therefore be responsible for I'.) HinII G3 is clearly larger and HinII K1 smaller than the difference in size between Δ LG3 and Δ LG37 ($1946 - 1426 = 520$ bp, see below), so that the order of the fragments affected by Δ LG37 must be D1-K1-G3.

HpaII C2/HinII digests yield three fragments, 510, 440 and 330 bp (Fig. 21c). HinII K1 comigrates with and, on the basis of the above mapping information, is presumed identical to the 440 bp fragment, so that the other two species are the flanking heterospecific fragments. As the new HpaII fragment created by Δ LG3 is either 900 or 925 bp (see below), HinII K1 plus the right heterospecific HpaII C2/HinII) fragment must add to less

than 925 bp. The smaller (330 bp) heterospecific fragment must therefore lie to the right of *Hin*II K1, and the larger (510 bp) to the left, making the order 510-440 (K1)-330 bp.

f. HpaI (Fig. 22)

*Hpa*I digestion of T7 WT DNA yields 19 fragments with relative mobilities as described by McDonnell et al (1977), who also report the complete *Hpa*I digestion map.

T7 C5LG3/*Hpa*I digests lack fragments F (2500 bp), G (2464 bp) and H (2312 bp) but contain instead two new fragments, E' (3000 bp) and M' (1100 bp), implying either Δ C5 or Δ LG3 must be wholly contained within one of the missing fragments. T7 LG37/*Hpa*I digests lack fragments C (4072 bp), G and Q (440 bp) and in their stead have a new fragment B' (5000 bp). The region deleted by Δ LG3 is therefore wholly contained within *Hpa*I G, implying the new fragment produced by Δ LG3 should be about 1040 bp (2464 minus 1426; Table 7), close to the observed size of M'. Evidently *Hpa*I E' is created by the C5 deletion.

Since *Hpa*I C is larger than the region deleted by Δ LG37, *Hpa*I Q must lie between *Hpa*I G and C. (Note that *Hpa*I Q and C must both be to the right of *Hpa*I G, since they are not affected by the LG3 deletion.). Therefore, the order of these fragments is G-Q-C. Not even the smallest *Hpa*I fragment, S (284 bp) could lie between or to the left of *Hpa*I G, F and H, as this would place the left end of *Hpa*I C at 20.0, well outside the region deleted by Δ LG37. Therefore, the five leftmost T7 WT/*Hpa*I fragments are (F,H)-G-Q-C, in that order. This agrees with the order F-H-G-Q-C reported by McDonnell et al (1977) and with Gordon et al (1978).

Figure 21. Mapping of HinII cutting sites

- a. Analysis of T7/HinII digests by electrophoresis on a gradient (4-10%) polyacrylamide gel. The buffer used in the stacking (2.5% polyacrylamide) gel and cathode buffer chamber was TBE, as usual. However, the gradient gel and anode chamber contained E buffer, 40 mM Tris acetate (pH 7.2), 20 mM NaCH₃COO, 2 mM EDTA. Voltage was 6 V/cm for 9 h.

Track 1. 1.5 µg T7 C5LG3/HinII
2. 1.7 µg T7 LG37/HinII
3. 1.5 µg T7 WT/HinII

- b. Analysis of a T7 WT/HinII digest by electrophoresis on a short, thick polyacrylamide (3.5%) gel. The stacking gel was agarose (0.7%) in TPE buffer (see Fig. 11c). TBE buffer was used in the polyacrylamide gel and both electrode chambers. Voltage was 4.5 V/cm for 9 h, then 2 V/cm for 10 h.

Track 1. 0.8 µg T7 WT/HinII

- c. Analysis of a HpaII C2/HinII digest by electrophoresis on a short, thin polyacrylamide (8%) gel. Voltage was 4.5 V/cm for 19 h.

Track 1. 1.5 µg T7 WT/HinII
2. HpaII C2/HinII

Figure 22. Mapping of HpaI cutting sites

Analysis of T7/HpaI digests by electrophoresis on a vertical, long, thick agarose (1%) gel. Voltage was 3 V/cm for 16.5 h.

Track 1. 2 µg T7 WT/HpaI
2. 1.2 µg T7 LG37/HpaI
3. 2 µg T7 C5LG3/HpaI

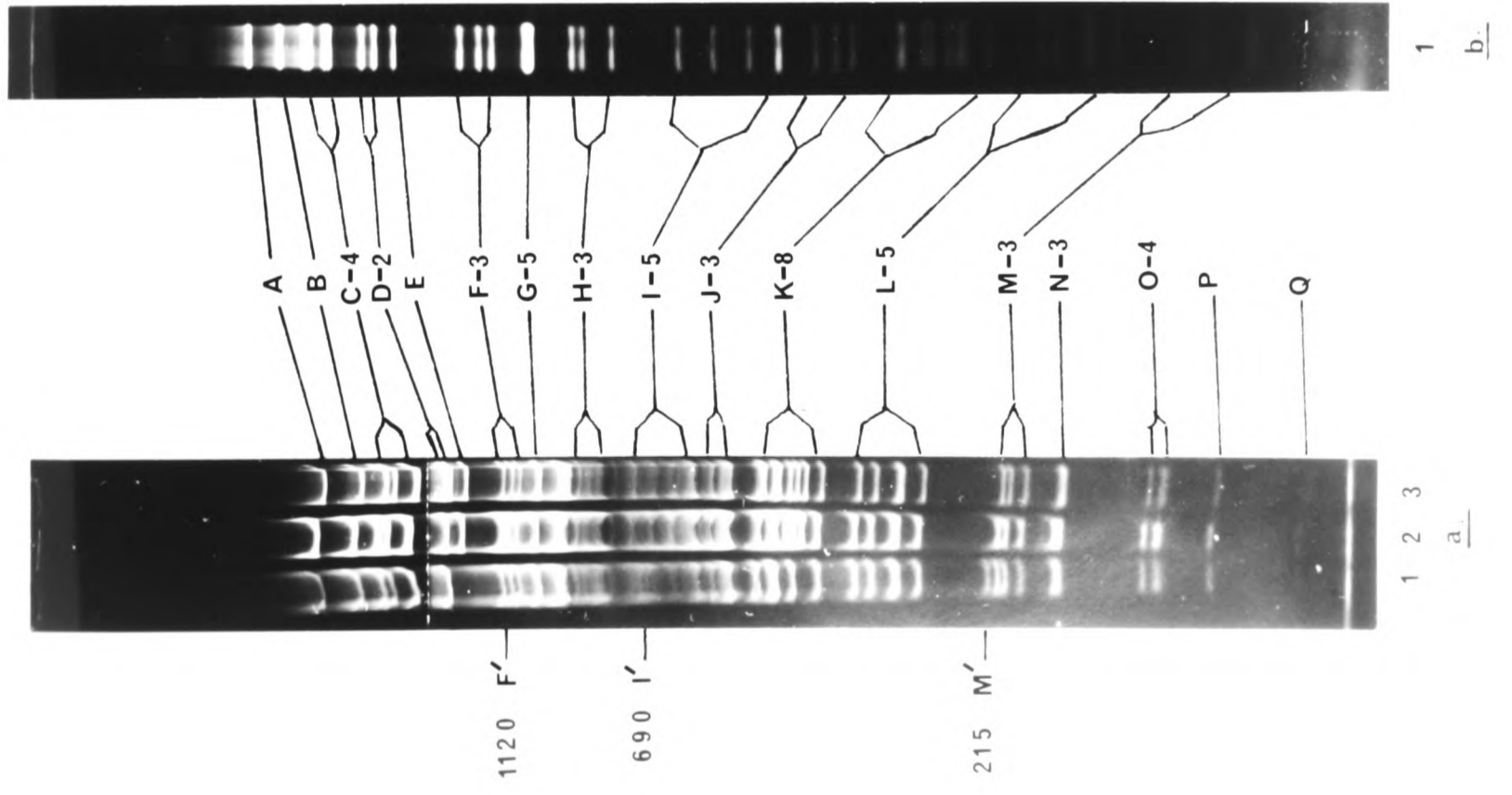


Figure 21

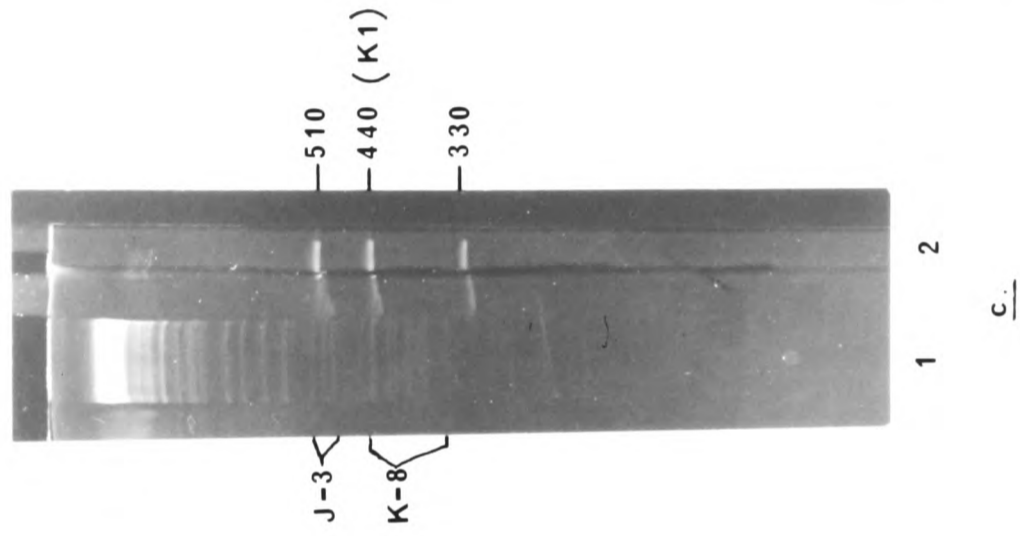
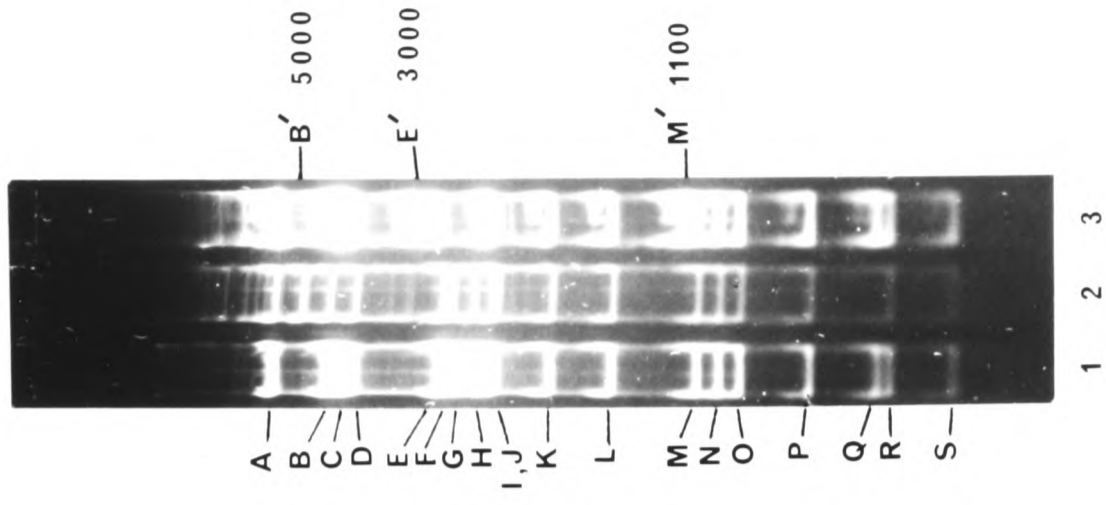


Figure 22



g. HpaII (Fig. 23)

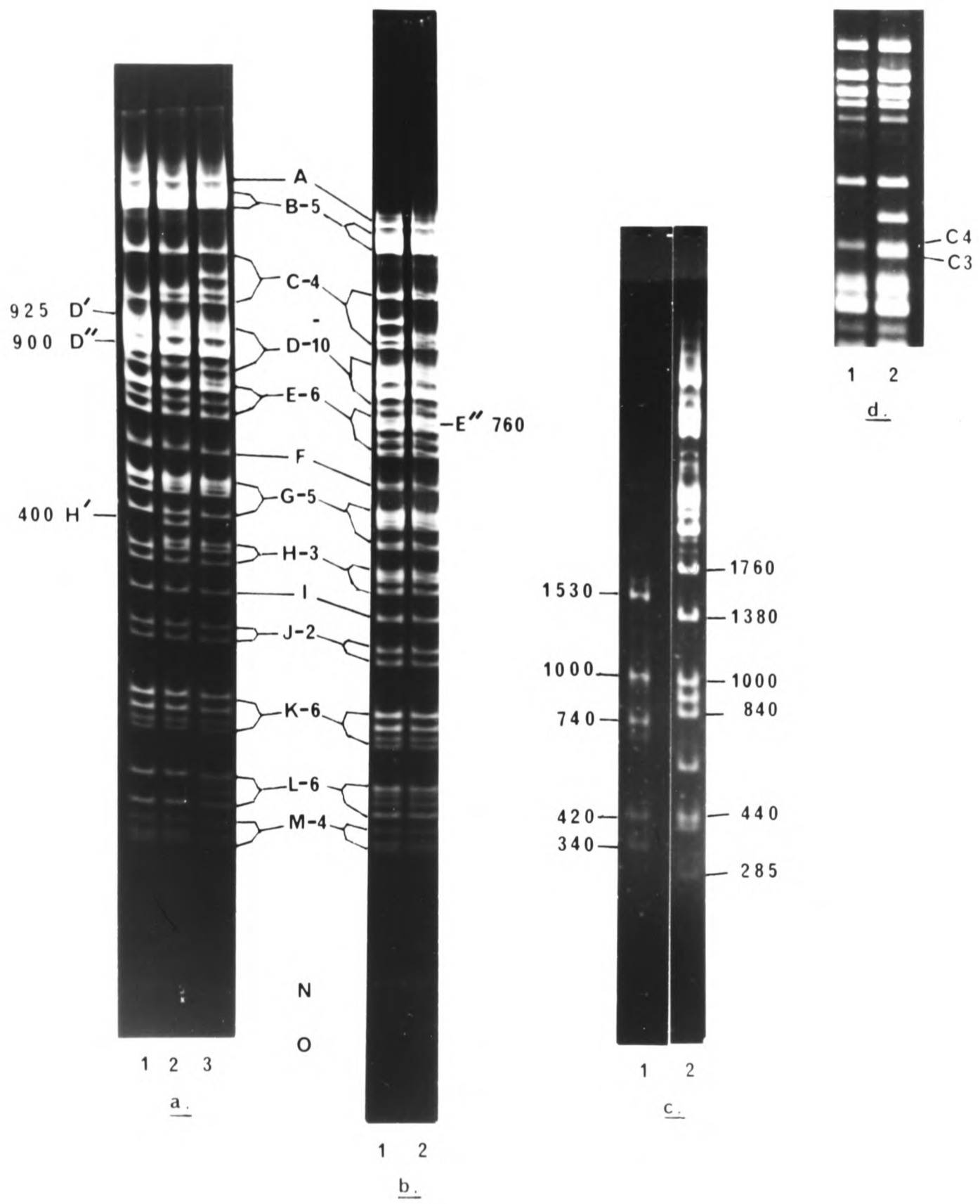
The gel pattern obtained on HpaII digestion of T7 WT DNA was indistinguishable from that communicated by Studier, except that only three fragments were present in the M group, as against four in Studier's report (Fig. 23, vs Fig. 6). No other difference was detected.

T7 C5LG3/HpaII digests lack fragments C2 (1290 bp), C3 (1150 bp), D7 (840 bp), E2 (760 bp), G3 (460 bp), H1 (370 bp), L3 (135 bp) and L4 (125 bp), but have instead two new fragments, D'' (925 bp) and D' (900 bp) (Fig. 23a). T7 LG37/HpaII digests lack fragments C2, E2, L3 and L4 and in their stead have a new fragment H' (400 bp). Evidently Δ LG37 must end on the right within the same HpaII fragment as Δ LG3, since no extra fragments are affected by Δ LG37. Since the right limit of Δ LG37 lies within HaeII Q (see 'a', above) and since HaeII Q is wholly contained within HpaII C2 (Fig. 16d), the right limits of Δ LG3 and Δ LG37 must lie within HpaII C2. The fragments E2, L3 and L4 must lie immediately to the left of C2, but their relative order cannot be deduced.

T7 LG26/HpaII digests (Fig. 23b) lack fragments C2 and D3 (940 bp), and in their stead have a new fragment E'' (760 bp). HpaII D3 therefore lies adjacent to and rightward of HpaII C2.

HpaI C/HpaII digests yield five fragments: 1530, 1000, 740, 420 and 340 bp (Fig. 23c). Though a direct comparison was not made, the 340 bp fragment is probably identical to the 330 bp right heterospecific fragment of HpaII C2/HinII digests and would therefore be the left

Figure 23



heterospecific fragment in the HpaI C/HpaII digest. (HpaI cuts at a subset of the sites recognised by HinII, including, on the basis of similarity in size of HpaI Q and HinII K1 and on mapping information, the site defining the left end of HpaI C.) Given that the left portion of HpaI C and the right portion of HpaII C2 overlap (e and f, above) and since HpaII D3 lies immediately to the right of HpaII C2, the 1000 bp fragment is probably identical to HpaII D3. Though HpaII C1 is about the right size to be the 1530 bp fragment, it maps to the extreme right of the T7 WT genome (Ludwig and Summers, 1975) and thus is clearly not this fragment.

Similarly HpaII C2, which on the basis of size is the only other T7 WT/HpaII fragment possibly identical to the 1530 bp fragment, maps elsewhere (see above). The right heterospecific fragment in the HpaI C/HpaII digest is therefore the 1530 bp species. The relative order of the 740 and 420 bp species cannot be deduced but they must lie between the 1000 bp (D3) and 1530 bp fragments. Therefore, the HpaI C/hpaII fragments are in the order 320 - 740 (D3) - (740, 420) - 1530 bp.

T7 C5LG3/AvaI D'/HpaII digests include a fragment comigrating with 'm3' (the third band in my M group; Fig. 12b). This is probably identical to Studier's 'M4', which he maps as the penultimate leftmost HpaII fragment in the T7 WT genome. A fragment comigrating with 'm3' has also been detected as an RNA polymerase-bound complex which is stable at low ionic strength (unpublished results with R. Hayward). Since Studier's 'M4' is thought to

Table 6. Estimations of Deletion Sizes

DELETION	EndoR.	(a) SUM OF AFFECTED FRAGMENTS	(b) NEW FRAGMENT CREATED	(a-b) DIFFERENCE	AVERAGE
ΔC5	HaeII	3425	1600	1825	1844 (4.61%)
	HaeIII	3270	1430	1840	
	HinII	2590	690	1900	
	HpaI	4812	3000	1812	
ΔLG3	HaeII	3660	2150	1510	1426 (3.57%)
	HaeIII	2615	1220	1395	
	HinII	1660	215	1445	
	HpaI	2464	1110	1354	
ΔLG26	HhaI	2230	760	1470	1470 (3.68%)
	HpaII	6250	4780	1470	
ΔLG37	HaeII	3945	2000	1945	1946 (4.87%)
	HhaI	3100	1120	1980	
	HinII	6976	5000	1976	
	HpaI	2310	400	1910	
	HpaII	2160	240	1920	

All sizes are given in base pairs; the average difference is also given in % T7.

Only those restriction endonucleases which affected and produced accurately measurable fragments are included.

contain the major early promoter A2 (Siebenlist, 1979), I conclude that my 'm3' is identical to Studier's M4, and that the discrepancy between our results involves some other M fragment.

Agarose gel electrophoresis inverts the relative mobilities of HpaII C3 and C4: the fourth band is absent and the third band present in the C-group of T7 C5LG3/HpaII digests when analysed on agarose gels (Fig. 23d), but vice-versa when examined on polyacrylamide gels (Fig. 23a). I have followed Studier in defining HpaII C3 and C4 by their mobilities on polyacrylamide. This change in relative mobility may be due to the different buffers used in the two systems (Tris-acetate in agarose gels and Tris-borate in polyacrylamide).

4. Deletion Size Estimations

An estimate of the size of a deletion can be made by comparing restriction endonuclease digestion patterns of DNA with and without this deletion, and subtracting the length of the new fragment present from the total length of those fragments missing from the deletion DNA digests. By averaging the results obtained using a variety of enzymes (for which accurate estimates of fragment sizes were known) the following figures were derived (Table 6): $\Delta C5$, 1844 bp (4.61% T7); $\Delta LG3$, 1426 bp (3.57% T7); $\Delta LG26$, 1470 bp (3.68% T7); $\Delta LG37$, 1946 bp (4.87% T7).

5. Mapping of Deletions

The results of the previous sections (concerning the sizes of deletions and their effects on EndoR. digestion

Table 7. Mapping of Deletions

DELETION	METHOD	LEFT LIMIT	RIGHT LIMIT	DELETION SIZE	
ΔC5	a	1.75-3.75	7.8-8.2		
	b	3.19-3.59	7.8-8.2	4.61	h
	c	2.9	8.1	5.2	i
	d	2.76	7.70		
	e	2.76-3.02	7.44-7.70	4.68	j
	f	3.3	8.1	4.8	k
ΔLG3	a	14.35-14.65	17.7-18.2		
	b	14.35-14.63	17.92-18.2	3.57	h
	c	15.2	19.2	4.0	i
	d	14.44	18.25		
	e	14.44-14.65	18.04-18.25	3.6	j
ΔLG13	c	15.6	20.4	4.8	i
	d	14.82	19.39		
	e	14.82-15.07	19.14-19.39	4.32	j
	c,g	15.2	20.4	5.2	i,g
	d,g	14.44	19.39		
	e,g	14.44-14.71	19.12-19.39	4.68	j,g
ΔLG26	a	17.75-18.2	20.7 -22.45		
	b	17.75-18.2	21.43-21.88	3.68	h
	c	19.1	23.5	4.4	i
	d	18.15	22.34		
	e	18.15-18.38	22.11-22.34	3.96	j
ΔLG37	a	14.35-14.65	19.3 -19.5		
	b	14.43-14.63	19.3 -19.5	4.87	h
	c	15.2	21.0	5.8	i
	d	14.44	19.96		
	e	14.44-14.74	19.66-19.96	5.2	j

- a. given which recognition sites for each relevant EndoR. are eliminated and which made adjacent by a deletion, constraints are placed on the left and right limits of that deletion.
- b. the deletion size (see Table 6) has been added to the left limit figures and subtracted from the right limit figures in 'a', to give new right and left limits, respectively. The overlaps between these new limits and those in 'a' provide the figures on line 'b'.
- c. as given by Simon and Studier (1973)
- d. figures in 'c' were adjusted by a factor of 19.2/20.2 as the T7 t₁ site, originally mapped at 20.2 by Simon and Studier (1973), has been remapped to 19.2 (McDonnell et al, 1977).
- e. figures in 'd' were further constrained by the deletion sizes (see 'k', below) by the procedure already described for 'b'.
- f. as given by Gordon et al (1978)
- g. reconciling the T7 LG13 and T7 D2 limits (see text).
- h. from table 6.
- i. as given by Simon and Studier (1973)
- j. figures in 'i' multiplied by 0.9 as recommended in McDonnell et al (1977)
- k. as given by Gordon et al (1978)

patterns) can be used to position these deletions within certain limits on Studier's map (Fig. 6). Table 7 summarises my conclusions and compares them with those of Simon and Studier (1973), McDonnell et al (1977) and in the case of T7 C5, Gordon et al (1978). For each deletion, my figures given on the second line ('b') are considered to be the most accurate and are assumed correct throughout this study.

a. T7 C5

The region deleted by $\Delta C5$ includes the HaeIII sites at 3.70 and 7.80 but not the HinII site at 1.75 and the HaeIII site at 8.20. However as the gel analyses described above indicate the size of $\Delta C5$ to be 1844 bp or 4.61% T7 and if the right limit is between 7.80 and 8.20, then the left limit must lie between 3.19 and 3.59. These estimates are slightly to the right of those derived from McDonnell et al (1977) and Simon and Studier (1973), 2.76-3.02 and 7.44-7.70 (Table 7e) but agree well with the estimate of Gordon et al, 3.3 to 8.1 (Table 7f).

b. T7 LG3

The region deleted by $\Delta LG3$ includes the HaeIII sites at 14.65 and 17.7, but not the HpaII site at 14.35 and the HinII site at 18.2. However, as the gel analyses described above indicate the size of $\Delta LG3$ to be 1426 bp or 3.57% T7, and since the left limit lies between 14.35 and 14.63, the right limit must lie between 17.92 and 18.2. This agrees well with the estimates from Simon and Studier (1973) as modified by McDonnell et al (1977): 14.44-14.65 and 18.04-18.25 (Table 7e).

c. T7 LG26

The region deleted by Δ LG26 includes the HhaI site at 18.2 and the BclI site at 20.7, but not the HhaI site at 17.75 nor the HpaII site at 22.45. However, as the gel analyses described above indicate the size of Δ LG26 to be 1470 bp or 3.68% T7, and since the left limit is between 17.75 and 18.2, the right limit must lie between 21.43 and 21.88. This left limit is in good agreement with the estimate from Simon and Studier (1973) as modified by McDonnell et al (1977), 18.15-18.49 (Table 7e) but my right limit lies slightly to the left of theirs (22.11-22.45).

d. T7 LG37

The region deleted by Δ LG37 includes the HaeIII site at 14.65 and the HinII site at 19.3, but not the HpaII site at 14.35 and the HhaI site at 19.5. However, as the gel analyses described above indicate the size of Δ LG37 to be 1946 bp or 4.87% T7 (Table 6), and since the right limit is between 19.3 and 19.5, the left limit must lie between 14.43 and 14.63. This left limit is in good agreement with the estimate from Simon and Studier (1973) as modified by McDonnell et al (1977), 14.52-14.90 (Table 7e); but my estimate of the right limit lies well to the left of theirs (19.66-19.96).

A comparison of HpaII, HaeIII and HinII digests of T7 LG37 and T7 LG13 DNA showed no detectable difference between the two mutants, strongly suggesting that the size and limits of these two deletions are identical. Simon and Studier (1973) listed them as having identical phenotypes but different limits. However, the left limit

which they ascribed to Δ LG13 was in doubt because another deletion, D2 (which unlike Δ LG3, Δ LG13 and Δ LG37 does not affect the gene 1 RNA) was assigned the limits 15.3 and 20.4 compared with 15.6 and 20.4 for LG13. If the Δ D2 figures are correct, and the phenotypic difference is real, then the left limit of Δ LG13 should map near 15.2 (using Simon and Studier's mapping). Its right limit could not be much altered without affecting its t^- phenotype, so that Δ LG13 might actually lie between 15.2 and 20.4 (Table 7c, g) or on the revised scale of McDonnell et al (1977), between 14.44-14.71 and 19.12-19.39 (Table 7e, g). This would put it in very close agreement with Δ LG37 as positioned by me. To check that these strains had not become confused at Edinburgh, DNA was prepared from plate-lysates of the original samples received from F.W. Studier. Again, the EndoR. digests were identical (Fig. 24). Since the phenotypes of the original T7 LG13 and T7 LG37 strains were not distinguishable, it seems reasonable to assume that the t^- strain used in the present studies is indeed T7 LG37.

6. Summary

The mapping information derived from experiments described above, is summarised in Figures 25 and 26. A compendium of these results and those of F.W. Studier (personal communication, Fig. 6) is presented in Fig. 27.

Figure 24. Comparison of T7 LG13 and T7 LG37

Analysis of HaeIII and HinII digests of T7 LG13 and T7 LG37 by electrophoresis on a horizontal agarose (1.5%) gel. Voltage was 2 V/cm for 2.5 h, and then 1.3 V/cm for 13 h.

- Track 1. 1.5 μ g T7 WT/HaeIII
2. T7 LG37/HaeIII
3. T7 LG13/HaeIII
4. 1.5 μ g T7 WT/HinII
5. T7 LG37/HinII
6. T7 LG13/HinII

The concentrations of T7 LG37 and T7 LG13 DNA used in each case were not determined.

Figure 24.

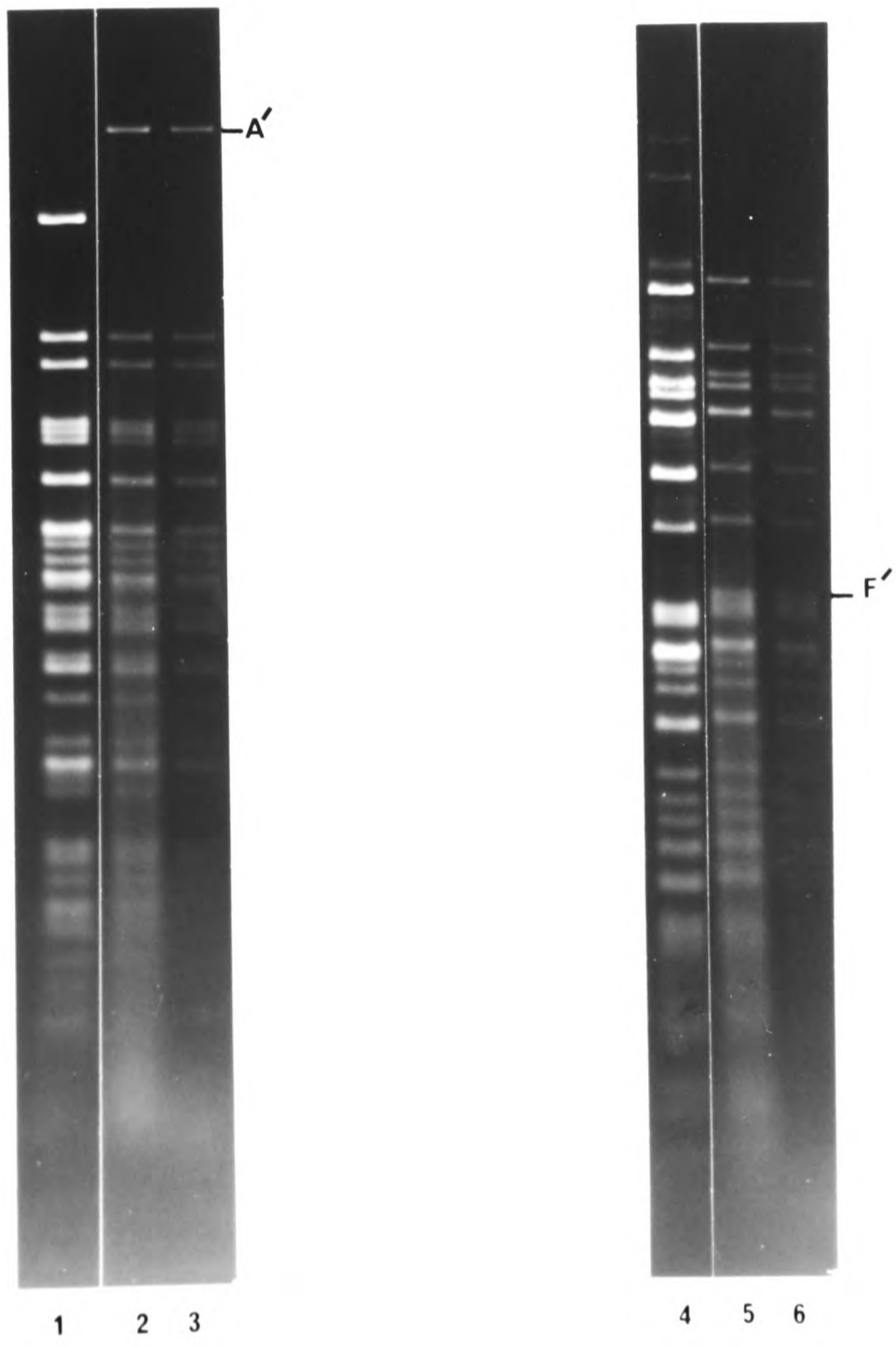


Figure 25. Large Scale Map of EndoR. Cutting Sites

A map of the entire T7 genome showing the cutting sites determined for the enzymes *Ava*I, *Bcl*I, *Kpn*I and *Hgi*AI. All positions were derived from estimates of the sizes of fragments and are relative to 0 or 100 on the T7 map. The results presented here are completely independent of the work of Studier.

Figure 26. EndoR. cutting sites between positions 10 and 30 on the T7 map

A summary of the cutting sites mapped between 10 and 30 on the T7 map for ten restriction endonucleases. All positions were determined by experiments described in the text, and then expressed relative to the *Hpa*I site mapped by McDonnell et al (1977) at 19.29. (The latter site is here considered as 19.3, because the size estimates used to map the cutting sites for other enzymes were not generally considered accurate to $\pm 0.01\%$ T7 (4 bp).) Although for consistency and simplicity, Studier's nomenclature has been adopted (Figs. 6 and 10), the results presented here are completely independent of his unpublished work.

Some of the sites for these Endo R.'s have not yet been accurately mapped and hence the data presented in this figure is incomplete in the regions designated '?'.
Some of the sites for these Endo R.'s have not yet been accurately mapped and hence the data presented in this figure is incomplete in the regions designated '?'.

Brackets indicate all of the relevant fragments in a particular region have been identified, but their relative order, and hence the exact position of the cutting sites have not.
Brackets indicate all of the relevant fragments in a particular region have been identified, but their relative order, and hence the exact position of the cutting sites have not.

Figure 25.

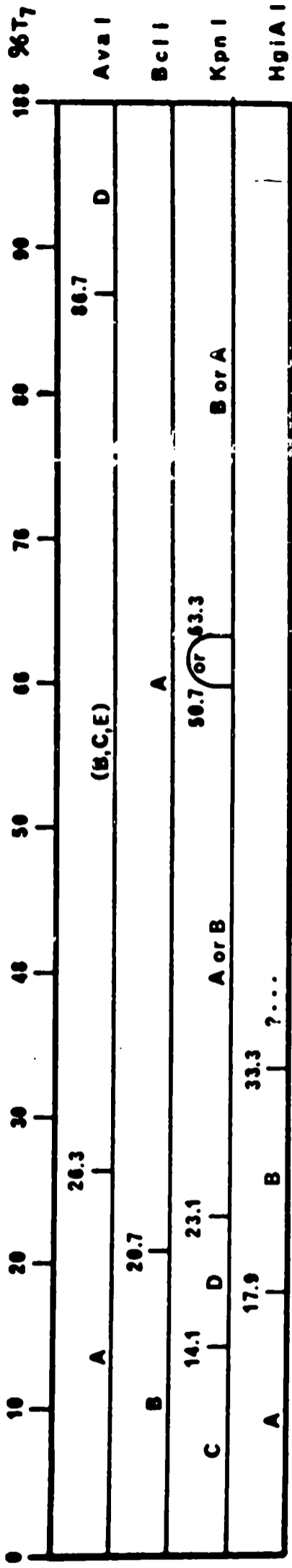


Figure 26.

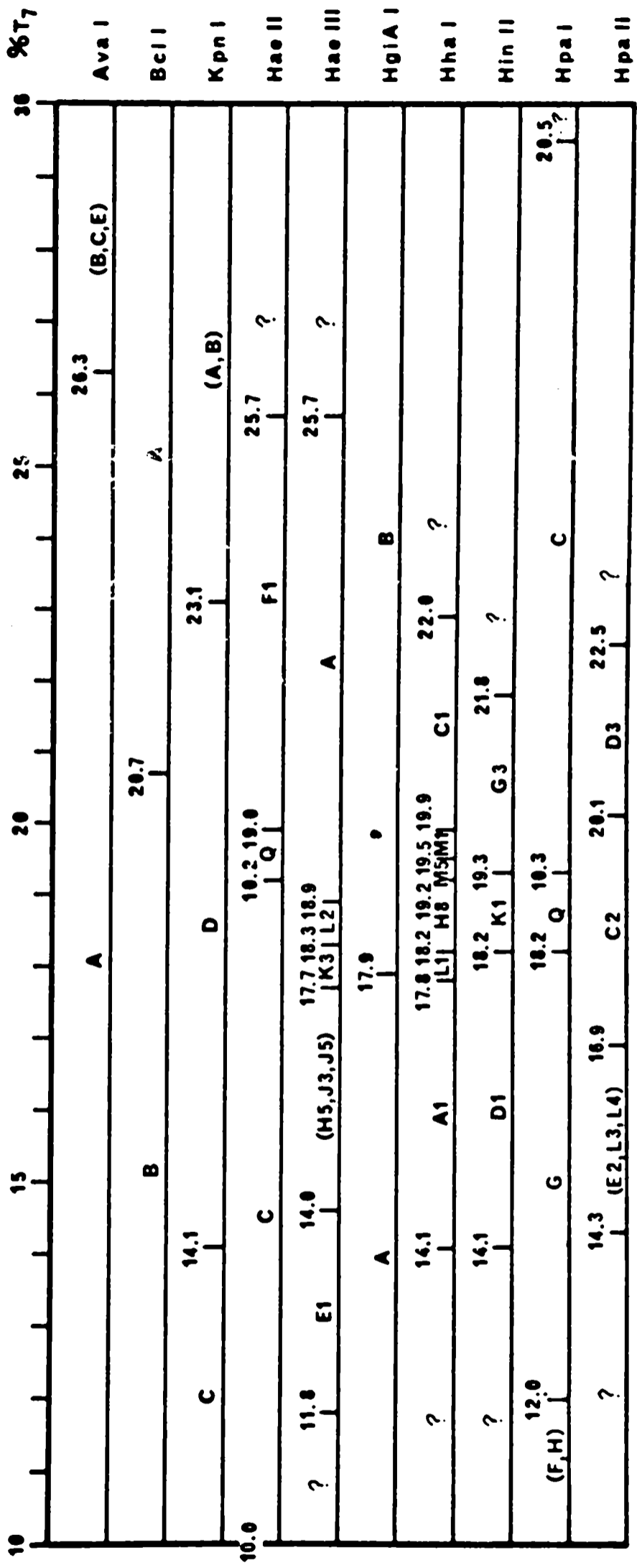
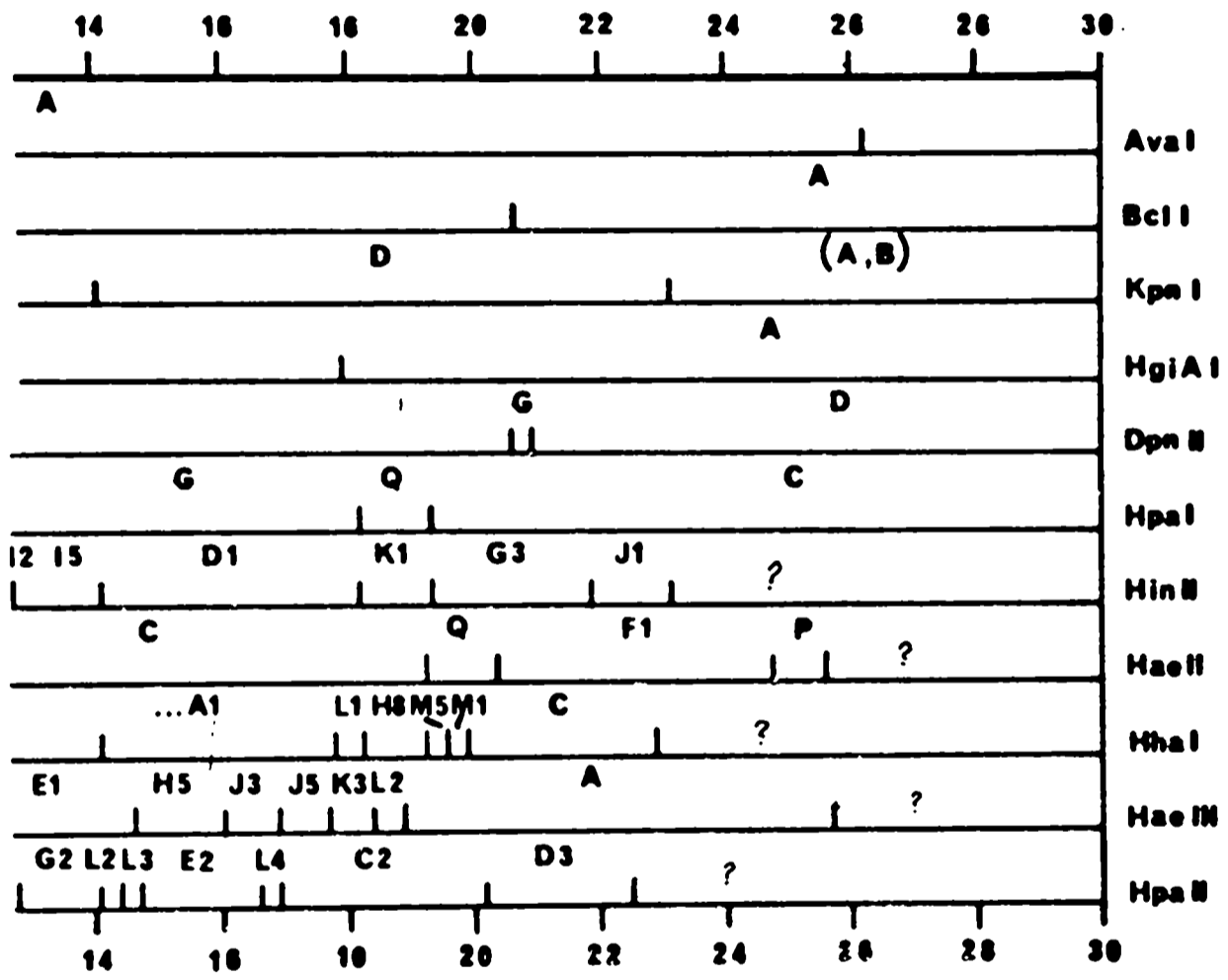


Figure 27. Map of EndoR. cutting sites and deletion limits
between positions 0 and 30 on the T7 map

A compendium is here presented, combining my independent results (Figs. 25 and 26) with the unpublished results of F.W. Studier (Fig. 6), as well as some of my conclusions which are dependent on Studier's mapping work.

The use of brackets and '?' is as described in fig. 26.



CHAPTER 5

ANALYSIS OF TERNARY COMPLEXES OF Rpol, DNA AND RNA

When the molar ratio of active Rpol to T7 DNA is 2:1, many DNA molecules will be simultaneously transcribed by 2 polymerase molecules. According to the results of O'Hare (1978), when the first Rpol reaches the transcription terminator it releases the nascent RNA relatively quickly ($t_{\frac{1}{2}} < 3$ min) and so becomes a binary complex (and thus presumably unstable in high salt filtration conditions; Chamberlin, 1976). However, the RNA is released much more slowly, so that the second polymerase is sterically prevented from reaching the termination site. This second enzyme therefore remains for a considerable time as a ternary complex (and thus stable to high salt filtration). If the transcription elongation inhibitor streptolydigin (stl; Cassani et al, 1971) is then added, the stalled ternary complex should be stabilised thus allowing time for the complete digestion of the DNA with a restriction endonuclease. Filtering under high salt conditions should then lead to the selective retention of fragments in ternary transcription complexes as well as in open promoter complexes (Chamberlin, 1976) on nitrocellulose filters. The retained fragments of DNA can be eluted with 1% detergent, recovered by ethanol precipitation and analysed on polyacrylamide gels to determine which fragments of DNA are specifically involved. The results of one such experiment using endoR.HaeIII are shown in Figure 28a.

With one notable exception, the low-salt filtrate (Track 1) is indistinguishable from the usual T7 WT/HaeIII

Figure 28. Analysis of Rpol:DNA Complexes Filtered on Nitrocellulose

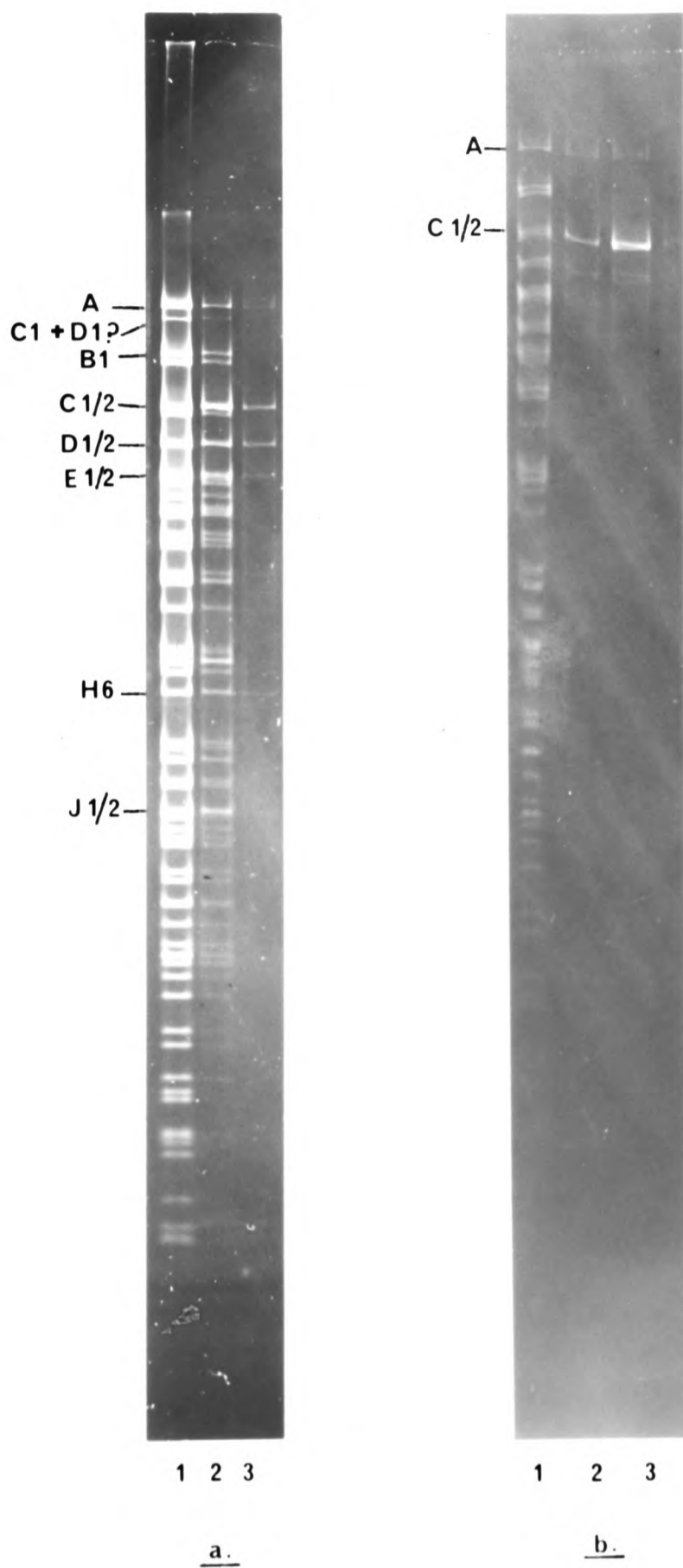
- a. A streptolydigin-inhibited transcription reaction was incubated with HaeIII, and filtered through nitrocellulose under low (0.05 M) or high (0.5 M) salt conditions, as described in Methods. Analysis was by electrophoresis on a gradient polyacrylamide gel; voltage was 4.5 V/cm for 15.6 h.

Track 1. Low salt:filtrate
2. Low salt:filter-bound
3. High salt:filter-bound

- b. A transcription reaction was incubated with HaeIII, filtered and analysed as described above except that:
i) the active enzyme to DNA molar ration was 4:1,
ii) streptolydigin was not used,
iii) HaeIII was added 9 min after the transcription reaction started,
iv) filtering was under high salt conditions only and was at both 2° and 20°, and
v) electrophoresis was at 4.7 V/cm for 3 h, then 4 V/cm for 14 h.

Track 1. 2°:filtrate
2. 20°:filter-bound
3. 2°:filter-bound

Figure 28.



digest pattern (cf. Figure 17) and so can be used as a marker to identify the bands in other tracks. The exception to this identity is a single, extra fragment present in subequimolar amounts which migrates between HaeIII A and BI on the gel shown in Fig. 28a. Now, Studier (personal communication) and Gordon et al (1978) have mapped the B promoter to about the same position on the T7 map as the HaeIII C1/D1 junction. Thus Rpol bound to the B promoter might protect the DNA from cutting by HaeIII at this site. Therefore, although the extra fragment observed in this experiment could result from incomplete digestion by HaeIII for trivial reasons, its uniqueness and its size (2650 bp) are consistent with its being an Rpol-protected partial digestion product comprising HaeIII C1 (1500 bp) and HaeIII D1 (1270 bp). Its presence in the filtrate is easily explained by a low efficiency of retention on filtering and/or by a weak association between the polymerase and DNA.

The fraction of material retained on the filter in this case was evidently not enough to be detectable by examination of the filtrate.

The low-salt, filter-bound material (Track 2) appears to contain disproportionately dense bands at the C1/2, H6 and J1/2 positions superimposed on a low background of all fragments. HaeIII C1 contains the promoters A₁, A₂, A₃, D and possibly B (F.W. Studier, personal communication; Gordon et al, 1978). HaeIII J1 contains the C promoter (McConnell, 1979a). The region contained within HaeIII H6 is unknown, but it may include the E promoter which maps

at position 92 on the T7 physical map (Koller et al, 1978).¹ Binary complexes formed between Rpol and these promoters (stable at low ionic strength) are probably responsible for the extra material retained. The putative HaeIII C1 + D1 partial digestion product (see above) is present in this track to a greater relative extent (compare with HaeIII B1) than in Track 1. This is consistent with its postulated derivation. The non-specific background may reflect DNA binding by Rpol (e.g. to the ends of fragments) and/or by HaeIII. The latter alternative is supported by the finding (O'Hare, Boothroyd and Hayward, unpublished results) that the background varies with the endonuclease used (HinII gave the highest and HpaII the lowest background of those studied).

High salt filtration effectively eliminates this non-specific background in the filter-bound material (Track 3). However, the following HaeIII fragments were retained (the relative densities of the bands are indicated in brackets): A (+), C1/2 (+++), D1/2 (++) , E1/2 (+) and H6 (+). This suggests that they were present as high salt-stable ternary Rpol complexes when filtered. Such complexes may have arisen in one or more of the following ways:

- 1) Rifampicin was used in these experiments to prevent reinitiation of transcription after the first cycle was completed. However, it does not prevent high salt-stable complex formation, but rather acts by preventing translocation after the formation of the first phosphodiester bond during initiation (McClure and Cech, 1978). This probably explains the retention of the promoter-containing

1. see footnote 1 pg. 106.

fragment HaeIII C1, and possibly HaeIII D1 (promoter B (?), see Fig. 6) and HaeIII H6 (promoter E (?)). The small size of HaeIII J1, which contains the C promoter, may have prevented its detection in this experiment.

2) Some polymerase molecules may have either stalled, paused, or progressed at an unusually slow rate so that stl added 6 min after initiation resulted in elongation complexes being 'frozen' at various places in the early operon. This probably explains the retention of HaeIII fragments from the early operon (i.e. HaeIII C1, D1 and E1, at least).

3) Finally, as detailed above, slow release of DNA from binary terminator complexes could result in the freezing by streptolydigin of ternary complexes lying just short of the termination site.

Since no Rpol binding site has been mapped within the region contained in HaeIII A (18.9-25.7 on the T7 map; see Chapters 1 and 4), the first mechanism described above seems a most unlikely explanation for the specific retention of HaeIII A. One or both of the remaining possibilities, therefore, most probably explains this result. However, it is conceivable that HaeIII A lies just outside the early operon and that Rpol reads through the terminator and then stalls within HaeIII A due to the action of streptolydigin.

Now, the retention of HaeIII A under high salt filtration was about 21% of that of the promoter-containing fragment HaeIII C1 (determined by densitometer scanning of Figure 28a, Track 3). Moreover, the efficiency of termination at the

end of the early operon has been determined to be approximately 95% (K. O'Hare, 1978; calculated from experiments performed in the absence of streptolydigin, but with the same enzyme preparation, under identical conditions, and within a few months before these studies). Together, these two results argue against the possibility that ternary complexes involving HaeIII A are mostly a result of *stl*-induced stalling of polymerases which have read-through the terminator.

To check that *stl* was not in some other way responsible for these results, the experiments were repeated but without the addition of *stl*. Instead, the slow release of DNA from binary terminator complexes was relied on to stall ternary, and thus high salt stable, complexes just before the terminator. Preliminary results show that HaeIII A is again specifically retained on Nitro-cellulose upon high salt filtration (Fig. 28b, Track 3). This strongly suggests that the transcription terminator at the end of the early operon lies within HaeIII A.

Similar studies using endoR. HpaII indicate that the terminator lies within HpaII C2, consistent with the deletion studies described in Chapter 4, and tentatively within HaeII C (O'Hare, Boothroyd and Hayward, unpublished).

1. It should also be noted that the fragment migrating to the H6 position (520 bp) in tracks a2 and a3 (fig. 28) may in fact represent a partial digestion product comprising the adjacent fragments J1 (350 bp) and M1 (150 bp): the HaeIII site between these two fragments has been shown to lie about 20 bp downstream of the RNA start-site at promoter C and about 10 bp upstream of the in vitro termination sites located in this region (McConnell, 1979b). Hence this partial digestion may be a result of specific protection by Rpol bound in a binary complex at promoter C or of *stl*-induced pausing of Rpol at the nearby termination site.

CHAPTER 6

NUCLEOTIDE SEQUENCE DETERMINATION IN THE REGION OF
THE T7 EARLY TRANSCRIPTION TERMINATION SITE, NEAR
19.29 ON THE PHYSICAL MAP

1. Approach

As outlined in the Introduction and detailed in Methods, the dideoxynucleotide chain-terminating method of sequencing DNA requires a primed-template to act as substrate for the extension reaction catalysed by DpolI. The simplest way to produce such a primed-template, in a system where the strands of DNA can be easily and cleanly separated, is to determine a detailed physical map of restriction sites in the region of interest, and then to hybridise purified fragments of DNA to the full-length, single-stranded DNA template. This was the approach adopted in these studies.

In the original description of this method (Sanger et al, 1978), bacteriophage ϕ X174 (+)-strand was used as the template, so that there was no possibility of intrastrand hybridisation (loopback) of the template leading to spurious, self-primed substrates for DpolI. However, in the case of the linear, single-stranded T7 DNA, such self-annealing might occur. This could generate a family of dideoxy-terminated extension which, if they recreated an appropriate restriction site, would result in a second set of specific bands being superimposed on the main sequence result. Self-annealing might occur at various positions so that several such minor sets of bands might be produced. In order to minimise this possibility, the template

(single-stranded T7 DNA, purified by the poly (U,G)-binding method) was incubated under conditions favouring the formation of loopbacks. DpolI, 3 dNTPs and 1 dd NTP were then added so that any 3'-termini able to act as primers should become extended, and ultimately terminated with a dd NMP. To test the effectiveness of this treatment in reducing background 'noise', mock sequencing experiments were performed in the absence of any added primer fragment. As hoped, the dideoxy-blocked template gave a lower background than unblocked template in the resolving part of the gel. The significance of this is uncertain because the amount of sample loaded onto the gel was not strictly controlled. It was later learned, however, that the original developers of the method had independently adopted this modification when using linear template (A. Smith, personal communication).

In order to isolate pure restriction fragments for use as primers from digests as complex as those shown for T7/HaeII, -/HaeIII and -/HinII, it became apparent that a two-stage purification would be necessary, as the fragments most suitable for use as primers lie in very congested regions of the gels. HpaII C2, however, was found to contain all the relevant restriction sites, and to migrate relatively far from other HpaII fragments. Accordingly, large quantities of HpaII C2 were purified by electroelution from polyacrylamide gels; electroelution was found to be the optimal method for this purpose as it was quick, easy and gave a high yield of fragment in a form sensitive to redigestion by other endonucleases. HpaII C2 was then recut

with HaeII, HaeIII or HinII. The resulting products were resolved on polyacrylamide gels and eluted (electroelution was at first used for this purpose but the sequence results obtained were found to be generally cleaner when passive elution, a slightly more involved procedure, was used).

Primers so purified were coprecipitated with template at a molar ratio of about 3:1, resuspended in low ionic-strength buffer, heat-denatured, and left to anneal at 60° for 60 min. These conditions were chosen to favour the formation of the required primed-template over loopbacks.

The extension reaction was carried out in conditions designed to increase the radioactivity and specificity of the family of polynucleotides corresponding to the main sequence. To this end each reaction contained more [α -³²P]-dNTP and DpolI than originally recommended (i.e. 5 vs 1 μ ci and 0.4 vs 0.2 units, respectively). As well, a higher than recommended molar ratio of ddNTP to dNTP was found to improve the results (Table 5b; i.e. between 60 and 500:1 instead of 50-250:1). Finally, it was found that the urea could be efficiently removed by gentle washing of the gels in 10% acetic acid for 10 min. The gels could then be dried and autoradiographed using a tungstate screen intensifier, flash-activated film, and exposure at -70°. These modifications allowed the chain-terminating method to be successfully applied to the sequencing of T7 DNA. It should, however, be noted that these changes were only gradually adopted during these studies (see individual Figure legends), so that only in the latter stages was the full potential of the technique realised.

In order to determine an extensive and largely unambiguous sequence in the region of interest many different primers were used, and extensions were made in both directions. Figure 29 shows schematically the primer used, the direction of extension, and the region in which useful sequence information was obtained, for every productive sequencing experiment.

2. Results

The individual results are shown in Figure 30. A complete interpretation of all these results is given in Figure 31. There are nine possible designations for any one position in a given experiments. These are listed and described in the legend to Figure 31. Uncertainty or ambiguity in individual results could arise in one or more of the following ways:

1) The strength of a band might vary according to the relative affinity of DpolI for ddNTP versus dNTP at any given position, depending on the surrounding sequence.

2) Local sequence might also affect the absolute affinity of DpolI for dNTPs so that it might pause at a given position and so generate bands of varying intensity in all (or several) tracks. This might or might not obscure the "true" band for this position.

3) The cold dNTP chase following the labelled extension reaction should reduce any effects of pausing by DpolI due to the relatively low concentration of the ^{32}P -labelled dNTP. However, contaminants in the $[\alpha\text{-}^{32}\text{P}]$ -dNTP preparations might cause sequence specific chain

Figure 29. Schematic Summary of Sequencing Experiments

The figure includes a detailed map (see Results, Chapter 4) of the cutting sites for the enzymes HaeII (AII), HaeIII (AIII), HhaI (HI), and HpaI (PI) within the HpaII (PII) fragment C2 (note: all HaeII sites are also recognised by HhaI, and all HpaI sites by HinII).

Map units (% T7) are given at the top, as are the distances in base-pairs from the HpaI site at 19.29% on the T7 map.

The thin lines beneath the map represent the various primers used, with the thicker, arrowed lines showing the direction of extension and the region for which useful sequence information was obtained.

Also shown are the relevant limits of three T7 deletions (see Results, Chapter 4).

<u>KEY</u>	<u>Primer</u>	<u>Template</u>
a,b,c,d	HaeIII L2	T7 WT r-strand
e,f	HpaII C2/HaeII 900	T7 WT r-strand
g	HinII K1	T7 WT r-strand
h,i	HpaII C2/HaeII 100	T7 WT l-strand
j	HpaII C2/HinII 330	T7 WT l-strand
k	HaeII Q	T7 WT l-strand

Figure 26.

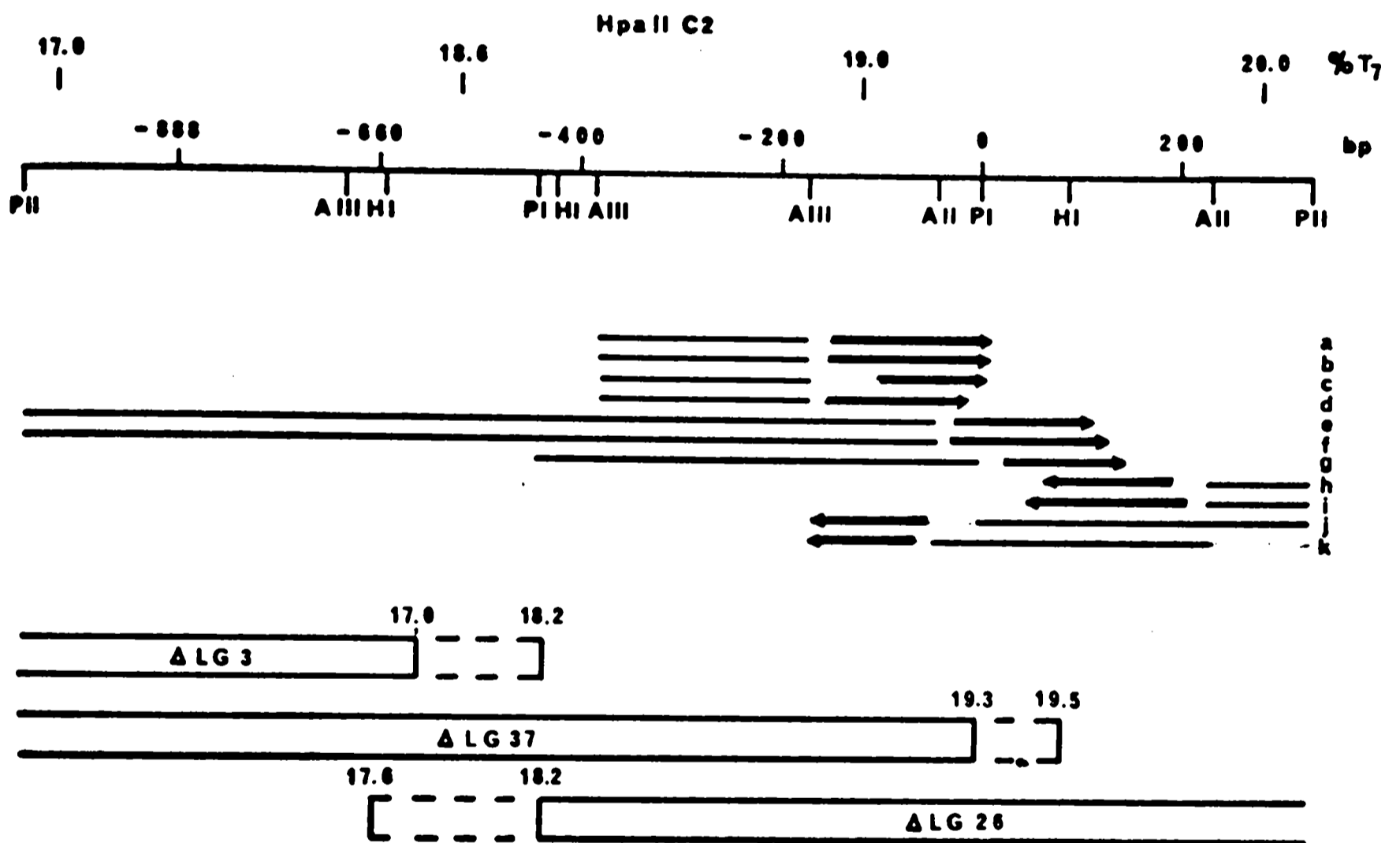


Figure 30. Nucleotide Sequencing Results

The autoradiograms produced in 11 different sequencing experiments (Fig. 29a-k) are reproduced. Details for each of these experiments are as described in Figures 8 and 9 and in Table 5a and b, except as noted in individual legends. The numbering of band positions corresponds to that used in Figure 31, wherein a complete interpretation of these results is compiled.

The tops of the autoradiograms in all cases are not shown, as these provided no useful information. The number 4 beneath a track indicates that a small sample from each of the 4 reactions (i.e. G, A, T and C) was loaded onto this track to provide a reference ladder of bands. Subscripts 1 and 2 beneath track designations indicate first and second loading, respectively. Unless otherwise indicated amounts added to reactions are expressed as volumes in μl .

Unusual bands are discussed in the light of the final sequence given in Fig. 31.

a. T7/HaeIII L2 primer on T7 WT/r-strand template.

C* indicates that a portion of the C reaction was not incubated with HaeIII before analysis. The N band in all tracks at $\sim +30$ is about the right size ($177 + 30 = 207$ bp) to represent labelled primer, i.e. HaeIII L2. The smaller N-band at -111 is not present in C*, suggesting that it results from recutting by HaeIII. It may be indicative of a HaeIII half-site, reminiscent of the HgiAI situation discussed in detail in Chapter 7. Its detection in this experiment alone may reflect unintended variations in the ionic strength in this experiment.

Variations in protocol:

	primer (electroeluted) + template			
	4 x 5 μ l aliquots			
	G	A	T	C
[α - 32 P] dGTP				
G ^N				
ddNTP (mM)	1.5(0.25)	1.5(0.6)	1(1)	1.5(0.25)
DpolI (0.3 u/ μ l)	1	1	1	1
dGTP				
HaeIII (2 u/ μ l)	1	1	1	1
L-form	20	20	20	20

1st loading 2.6 h, 23 mA
 2nd loading 1.7 h, 20 mA
 autoradiographed wet, 16 d exposure

b. T7/HaeIII L2 primer on T7 WT/r-strand template.

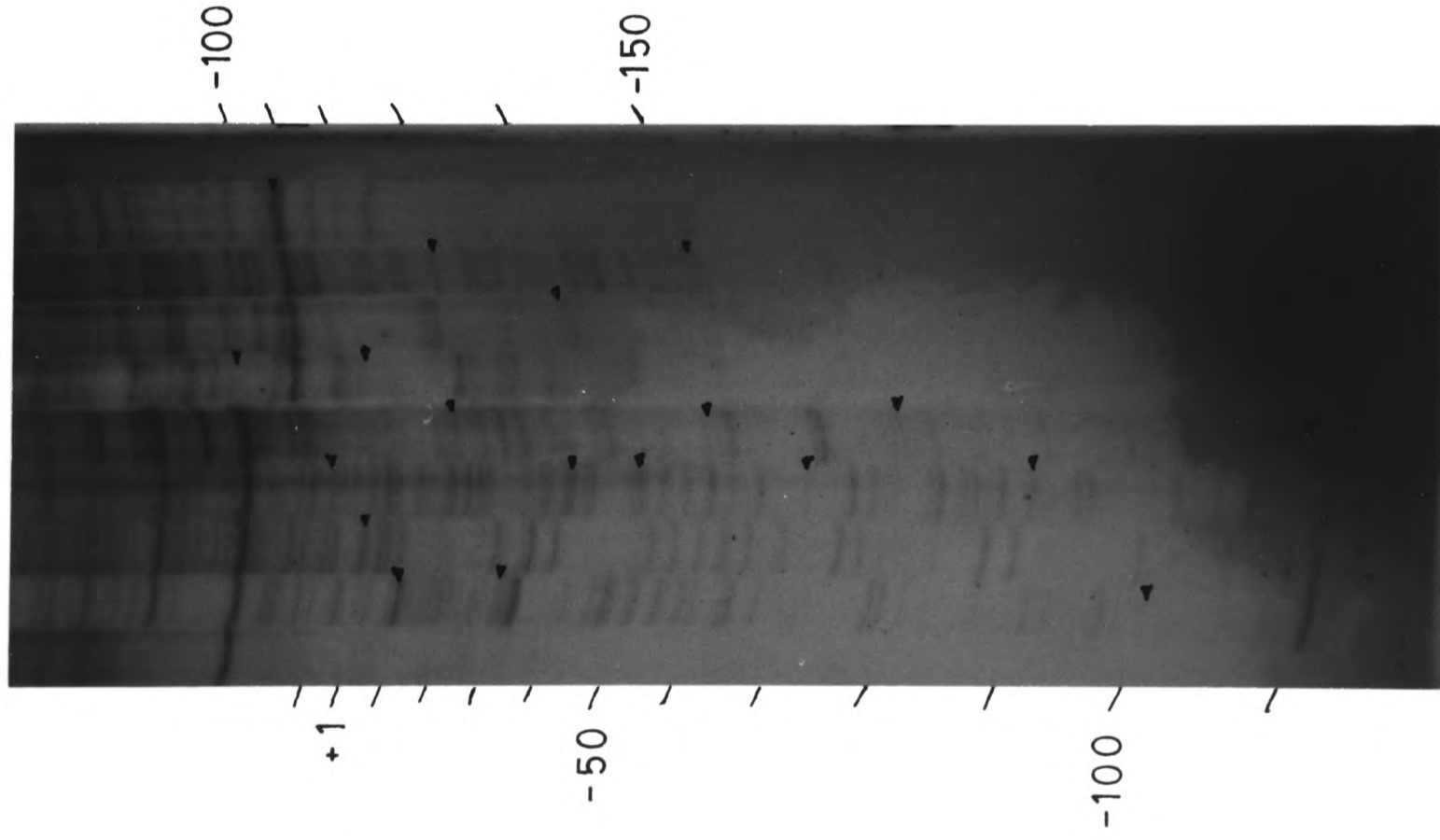
The band at position -62 in the Hg track probably indicates that there is a "half" HgiAI-recognition site at this position (see Chapter 7). The N-band at $+30$ is about the right size ($177 + 30 = 207$ bp) to represent labelled primer, i.e. HaeIII L2.

Variations in protocol:

	primer (electroeluted) + template					
	(4) G	(4) A	(4) T	(4) C	(2) Hg	(2) Hi
[α - 32 P] -dGTP						
G ^N						
ddNTP (mM)	2(0.25)	2(0.6)	1.6(1)	2.6(0.25)		
DpolI						
dGTP	1.5	1.5	1.5	1.5	1.5	1.5
HaeIII (2 u/ μ l)	1	1	1	1	1	1
HgiAI (0.5 u/ μ l)					5	
HinII (0.5 u/ μ l)						5
L-form	25	25	25	25	25	25

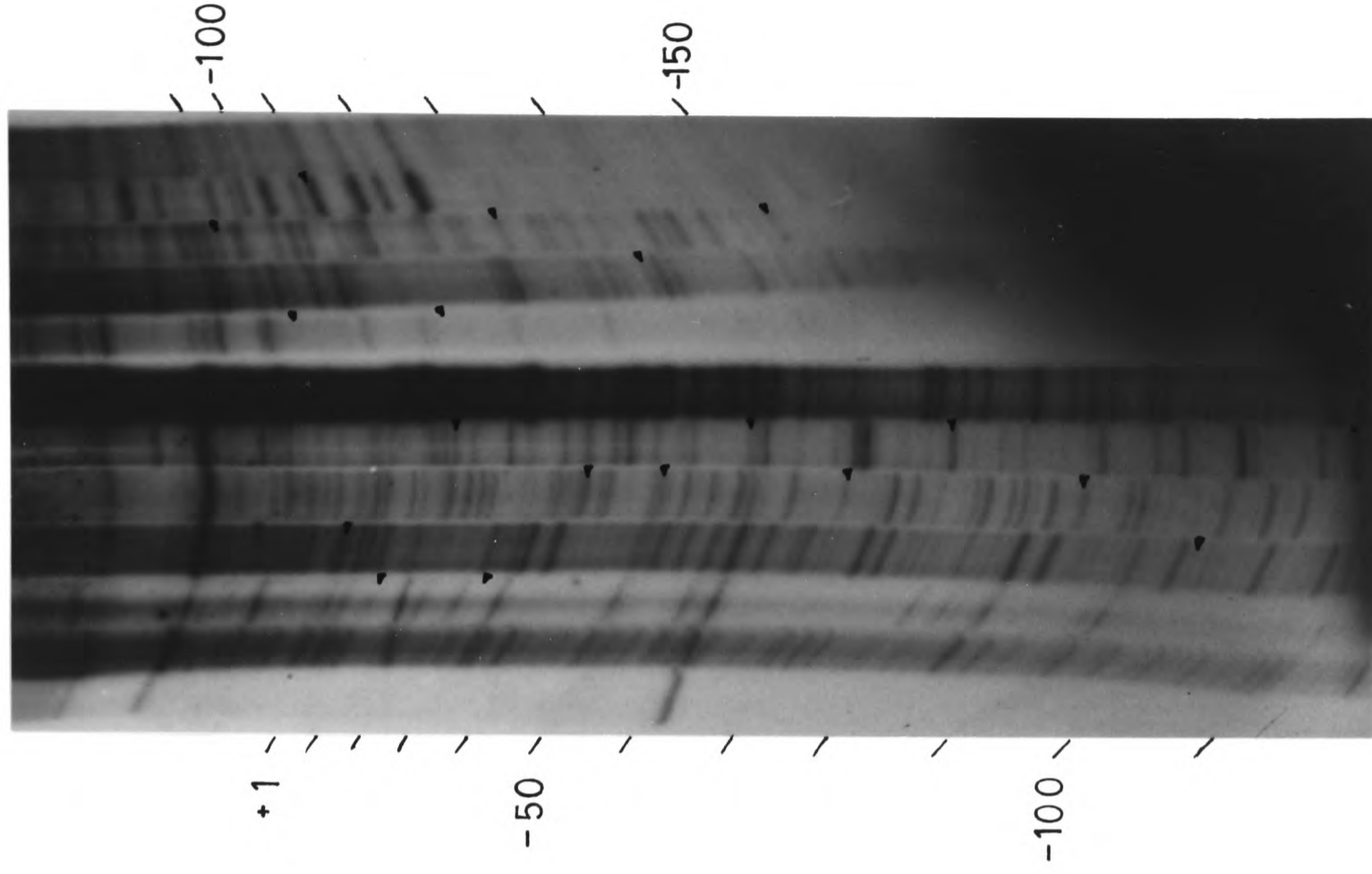
1st loading 2.3 h, 23 mA
 2nd loading 1.6 h, 20 mA
 autoradiographed wet, 3.5 d exposure

C₁ G₁ A₁ T₁ C₁ G₂ A₂ T₂ C₂



a.

Hg₁ G₁ A₁ T₁ C₁ Hi₁ G₂ A₂ T₂ C₂ 4₂



b.

c. T7/HaeIII L2 primer on T7 WT/r-strand template.

The band at -43 in the Ha track probably represents the HaeII site mapped to this position (Chapter 4) and predicted from the sequence (Fig. 31). The band at -49 in the Ha track may be due to half-cutting by HaeII at the HhaI site (+86) as this would yield a fragment 129 nucleotides long, i.e. from -43 to +86), about the size of fragments migrating to this position (177 - 49 = 128 bp). This is analagous to the HgiAI half-site discussed in Chapter 7.

The band at about -1 probably represents the HinII site mapped to this position (Chapter 4) and predicted from the sequence (Fig. 31).

Variations in protocol:

primer (electroeluted) + template

G (4)	A (4)	T (4)	C (4)	Hi (2)	Ha(2)
-------	-------	-------	-------	--------	-------

[α - ³² P]-dGTP						
G ^N						
ddNTP (mM)	1(0.6)	2(0.6)	1.6(1)	1(1)		
Dpoll						
dGTP	1.5	1.5	1.5	1.5	1.5	1.5
HaeIII (2 u/ μ l)	1	1	1	1	1	1
HaeII (0.5 u/ μ l)						5
HinII (0.15 u/ μ l)					5	
L-form	20	20	20	20	20	20

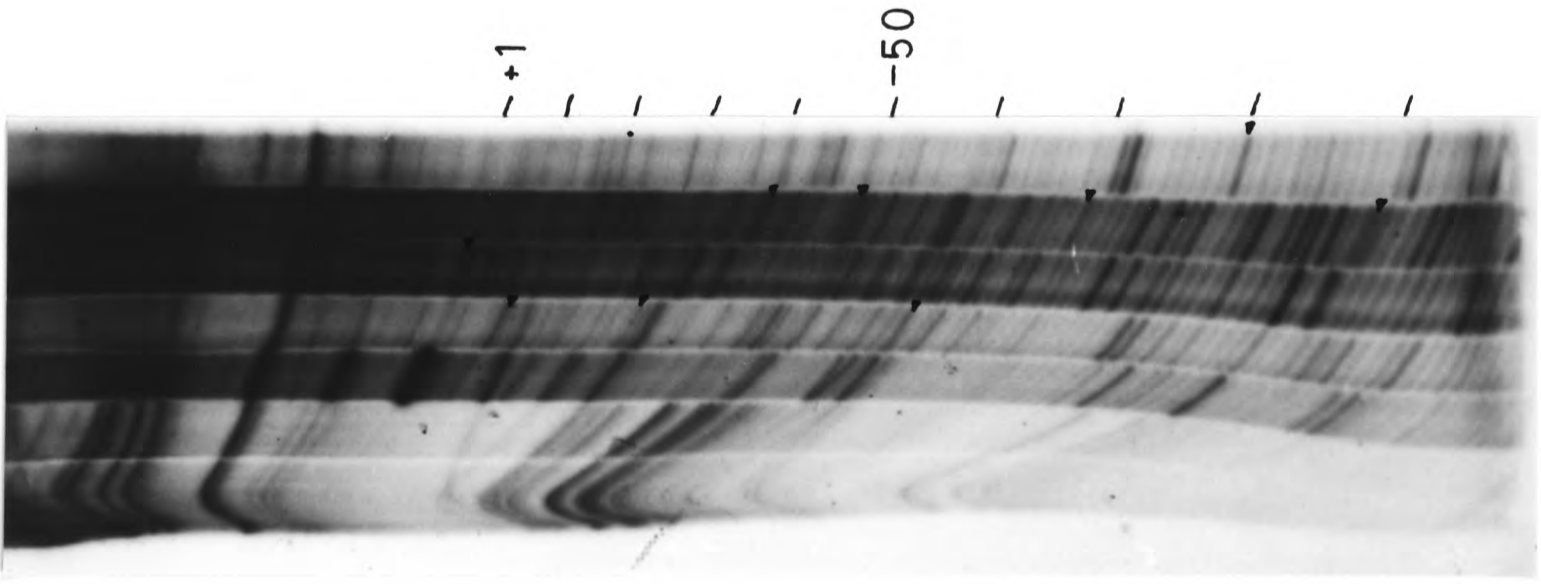
4.5 h, 24 mA
 autoradiographed wet, 12 d exposure

d. T7/HaeIII L2 primer on T7 WT/r-strand template.

The samples analysed on this gel were those initially prepared and analysed as described in c, above. They were stored (in L-form) at -20^o for 2.5 d before being loaded onto this gel for reanalysis. An interpretation of the unusual bands is given in c.

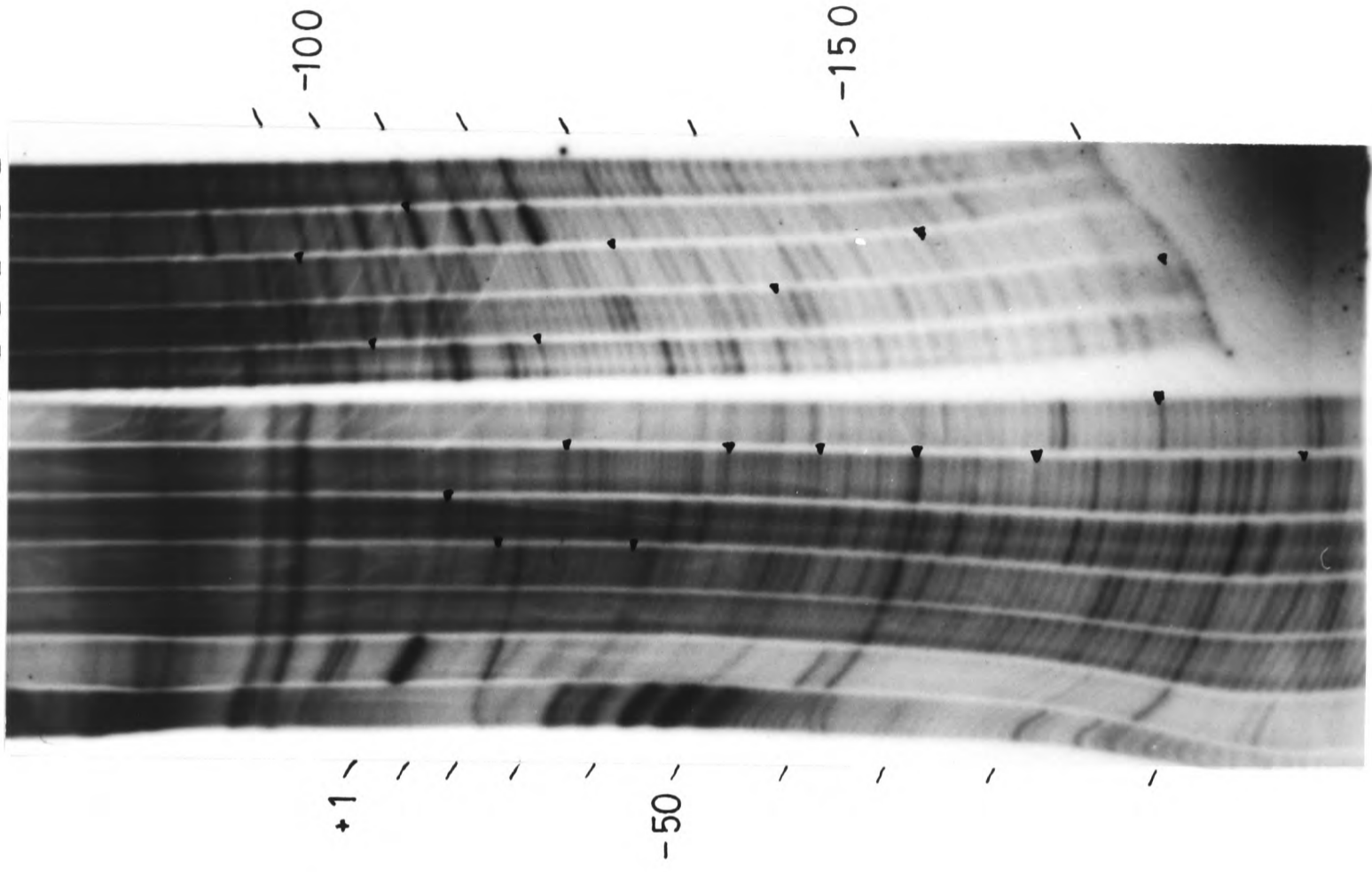
1st loading: 3.2 h, 25 mA
 2nd loading: 1.9 h, 23 mA
 autoradiographed wet, 16 d exposure

Ha HaHi G A T C



c.

Ha₁Hi₁ G₁ A₁ T₁ C₁ G₂A₂ T₂ C₂⁴₂



d.

e. T7/HpaII C2/HaeII - 900 primer on T7 WT/r-strand template.

The band at position -1 in the H track probably represents the HinII site mapped to this position (Chapter 4) and predicted from the sequence (Fig. 31).

Variations in protocol:

primer (electroeluted) + template

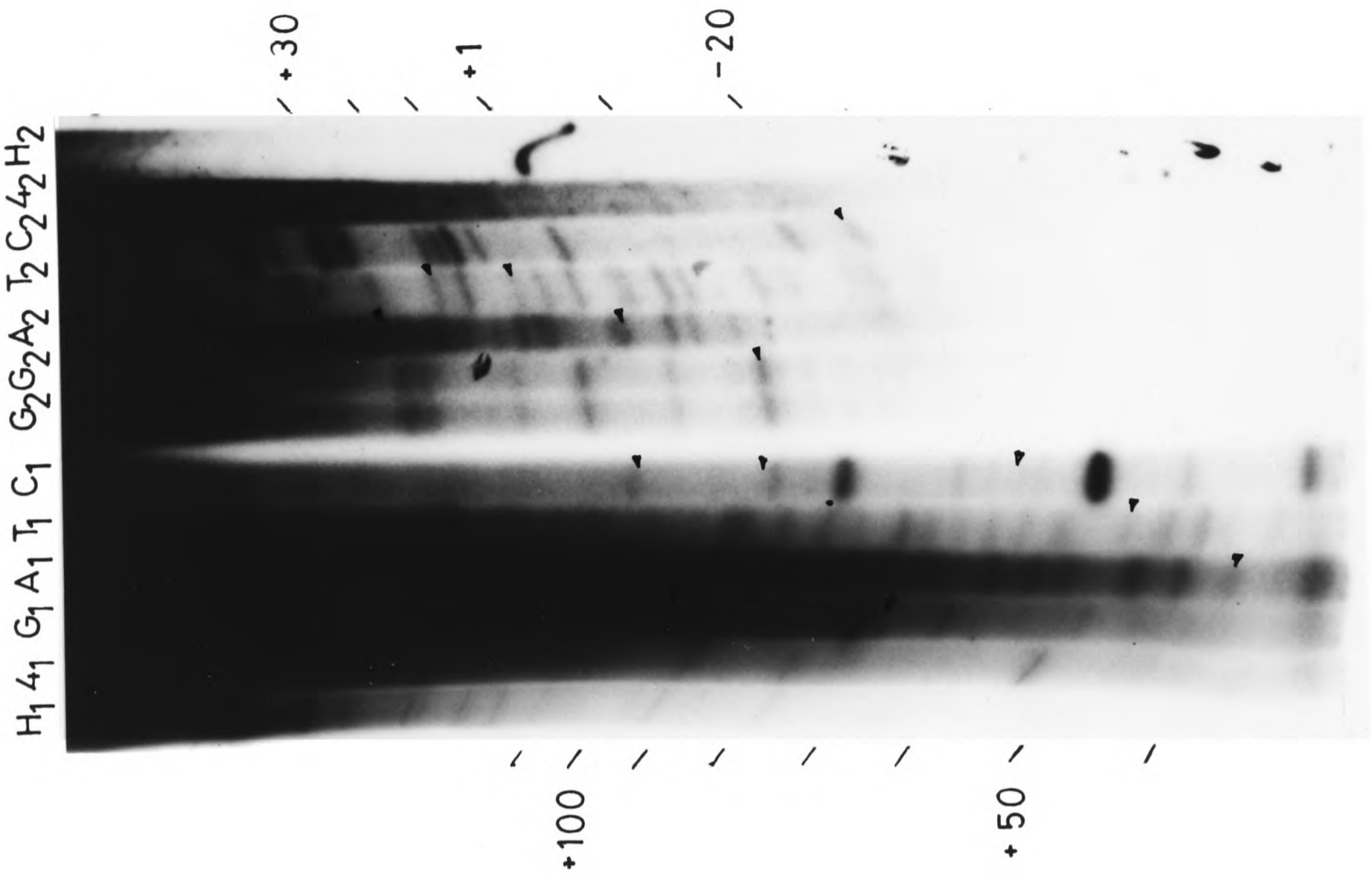
(5) (5) (5) (5) (2.5)
G A T C H

	(5) G	(5) A	(5) T	(5) C	(2.5) H
[α - ³² P]-dGTP					
G ^{ON}					
ddNTP (mM)	1 (0.6)	2 (0.6)	1.6 (1)	1 (1)	
DpolI					
dGTP	1.5	1.5	1.5	1.5	1
HaeII (2 u/ μ l)	1	1	1	1	1
HinII (0.5 u/ μ l)					5
L-form					

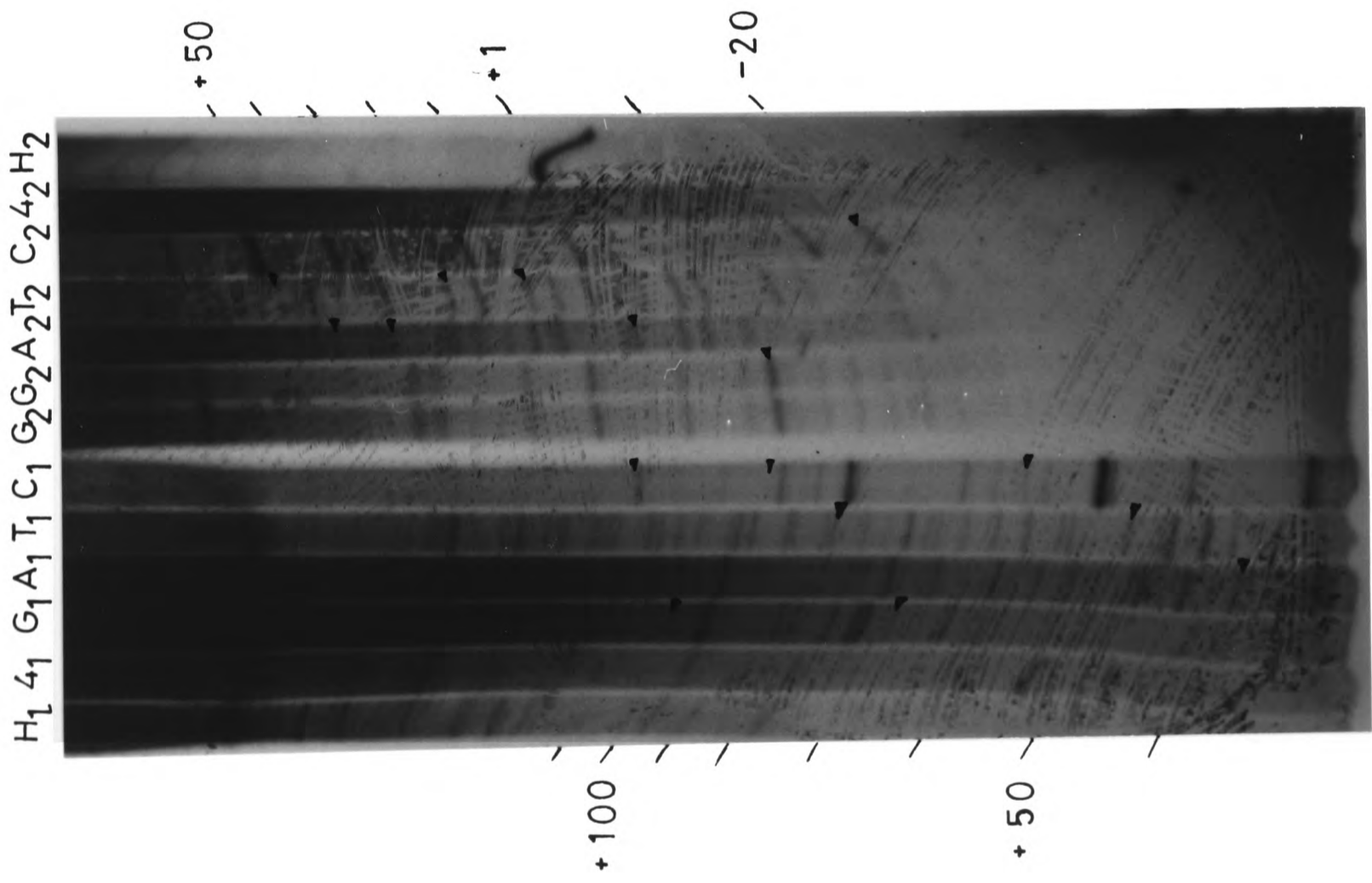
1st loading: 3.3 h, 23 mA

2nd loading: 2.1 h, 21 mA

- i. autoradiographed wet, 13 d exposure
- ii. dried and reautoradiographed with intensification, 2 d exposure.



e. ii.



e. i.

f. T7/HpaII C2/HaeII - 900 primer on T7/r-strand template.

The band at position -1 in the H track probably represents the HinII site mapped to this position (Chapter 4) and predicted from the sequence (Fig. 31).

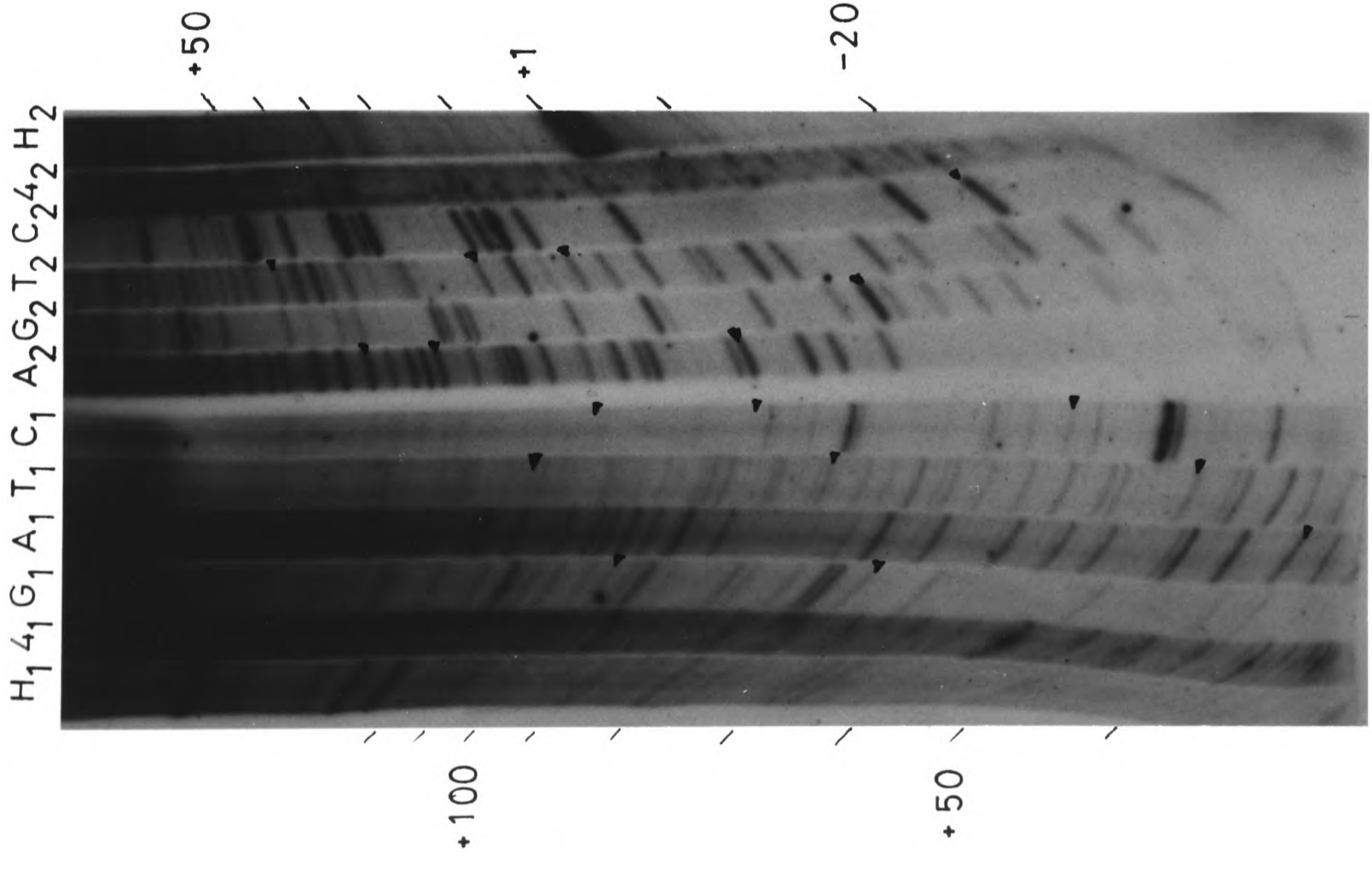
The protocol for this experiment is described in detail in Table 5a and b, and Figs. 8 and 9.

g. T7/HinII K1 primer on T7/r-strand template.

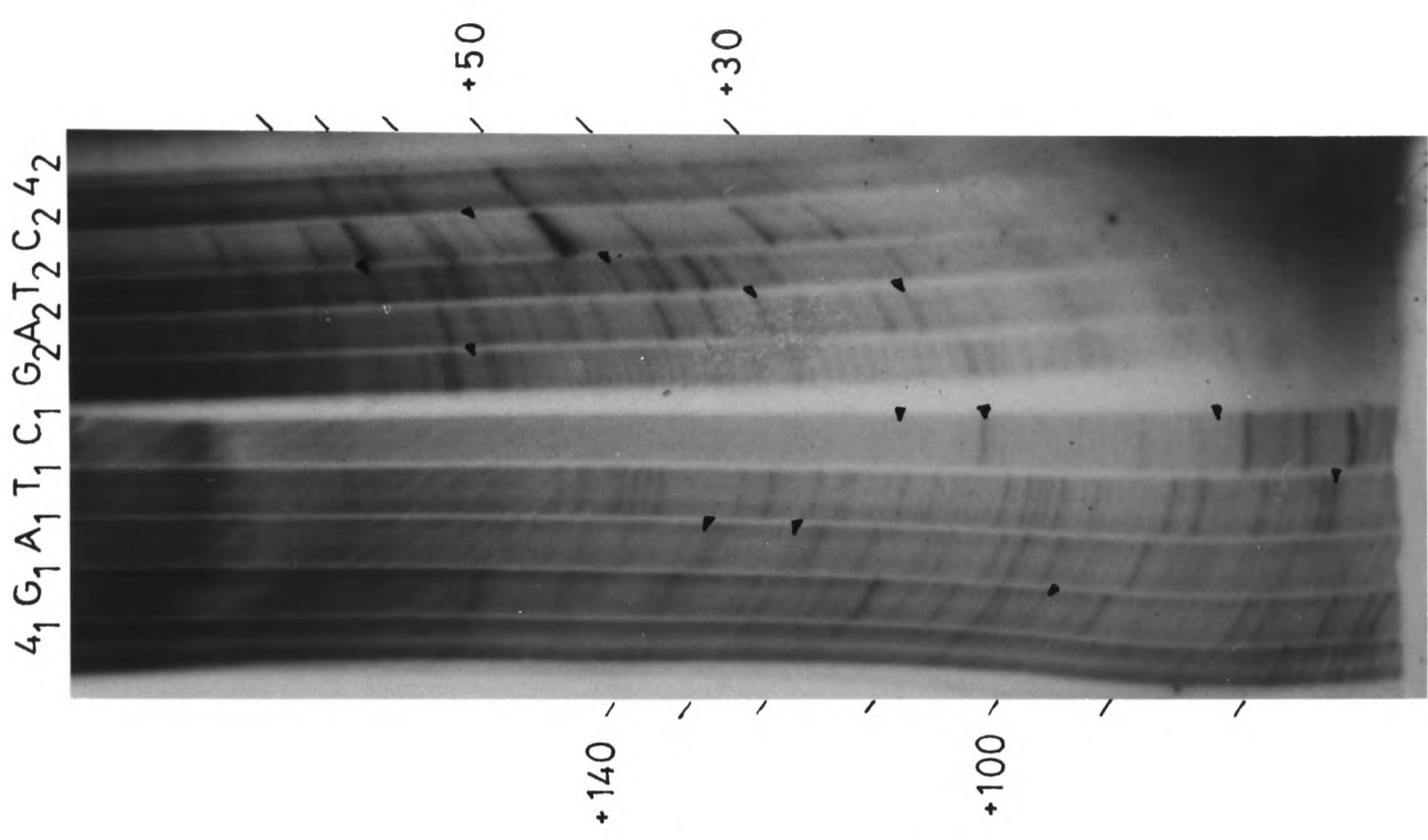
Variations from protocol:

	primer (electroeluted) + template			
	4 x 5 μ l aliquots			
	G	A	T	C
$[\alpha\text{-}^{32}\text{P}]\text{-dGTP}$				
G^{ON}				
ddNTP (mM)	2 (0.6)	4 (0.6)	3 (1)	2 (1)
DpolI				
dGTP	2	2	2	2
HinII (4 u/ μ l)	0.7	0.7	0.7	0.7
L-form				

1st loading: 3h, 25 mA
 2nd loading: 1.5 h, 25 mA
 autoradiographed wet, 15 d exposure



f.



g.

h. T7/HpaII C2/HaeII - 100 primer on T7/l-strand template.

The N band at position +127 is about the right size to be labelled primer (228 - 127 = 101 bp; these extensions were to the left from the HaeII site at +228 so that a band at +127 represents a fragment about 101 nucleotides in length).

The origin of the N band at position +102 is unknown.

Variations in protocol:

	primer (electroeluted) + template				
[α - ³² P]-dGTP					
G ^{ON}					
ddNTP (mM)	1 (0.6)	1 (0.6)	2 (1)	1 (1)	
DpolI					
dGTP					
HaeII					
HinII					
	1st loading: 3.3h, 20 mA				
	2nd loading: 2.5h, 20 mA				
	exposure, 22 h				

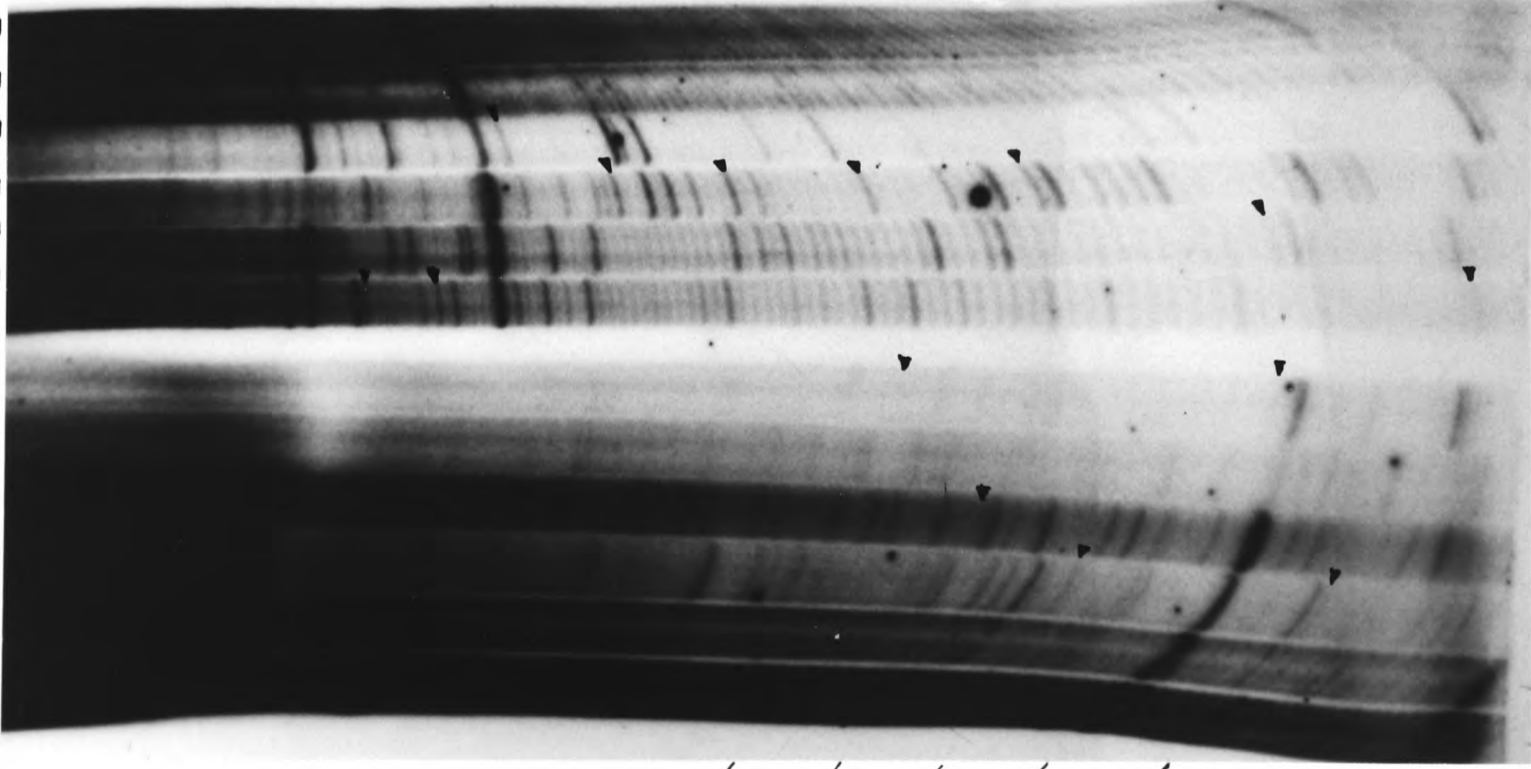
i. T7/HpaII C2/HaeII - 100 primer on T7/l-strand template.

The origin of the N band at position +127 is probably as described in h, above. The N band at position +101 is of unknown origin.

Variations in protocol:

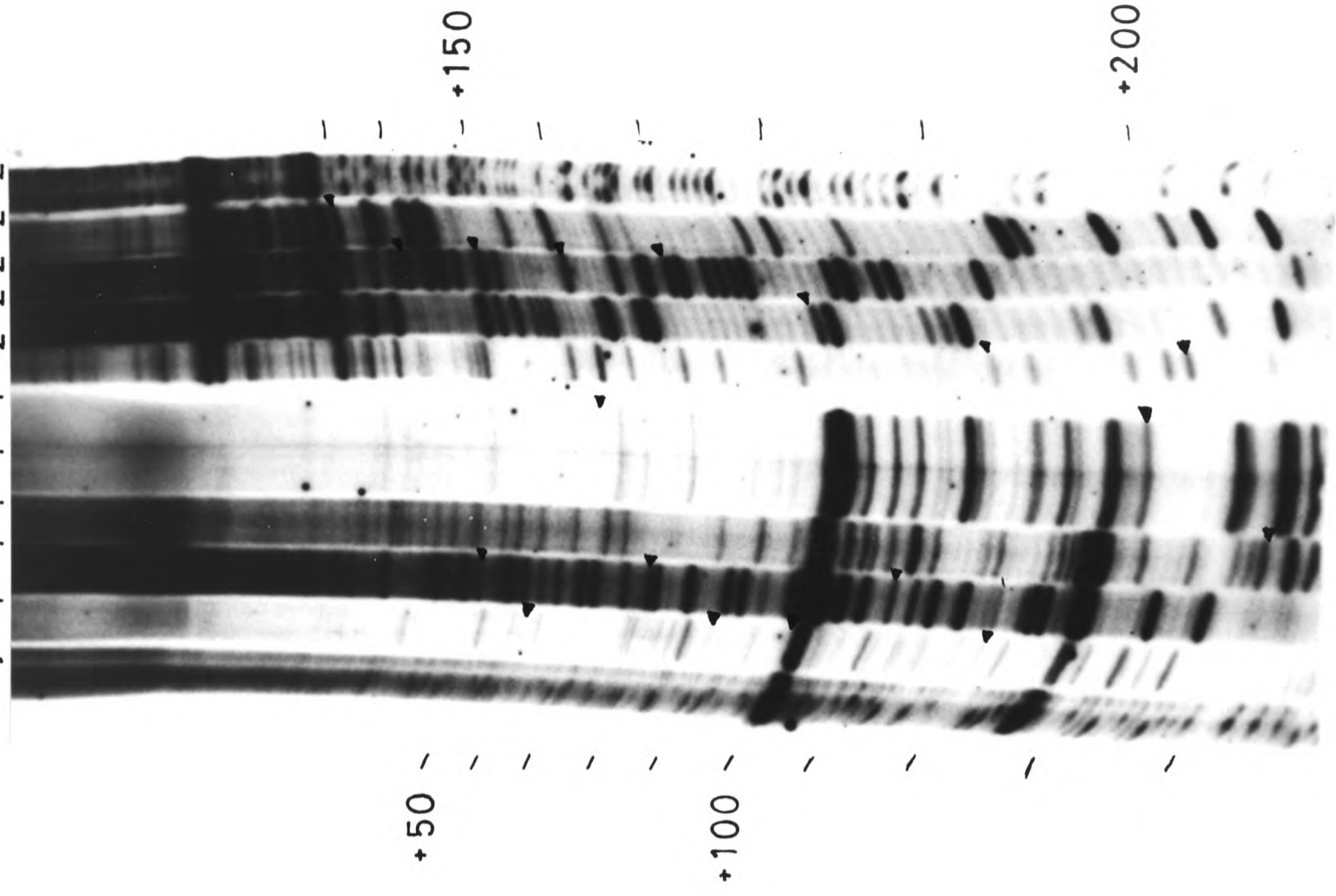
	primer + template			
[α - ³² P]-dGTP				
G ^{ON}				
ddNTP (mM)	1.7 (0.6)	3.5 (0.6)	3.5 (1)	1.7 (1)
DpolI				
dGTP				
HaeII				
L-form				
	1st loading: 2.5 h, 25 mA			
	2nd loading: 2 h, 23 mA			
	exposure, 15 h			

H₁ 4₁ G₁ A₁ T₁ C₁ G₂A₂T₂C₂⁴H₂



h.

4₁ G₁ A₁ T₁ C₁ G₂A₂T₂C₂⁴



i.

Figure 31. Compendium of Sequence Results

Sequence information derived from the autoradiograms shown in Figure 30a-k is here presented in a manner allowing a comparison of the different results. The polarity in all cases is 5' to 3' and the sequence is of the l- or sense-strand. Distances are given in nucleotide residues from the HpaI site at 19.29% on the T7 map (McDonnell et al, 1977). The bottom line in each 100 nucleotide section shows the final designation for each position.

KEY

<u>Symbol</u>	<u>Description</u>
X	No band in any track, yet one expected.
A O	Weak band in one track.
A *	Moderate or strong band in the indicated track; weaker bands in one or more of the other tracks.
A n	Moderate bands in three or more tracks, but the given one predominates.
A G	Equally strong bands in two tracks.
N N	Equally strong bands in three or more tracks.
AA N	Additional strong band in all tracks, apparently interpolated at the indicated position in the specific sequence ladder.
A *	Strong band in one track, but with anomalous mobility relative to the nonspecific background ladder.
A	Clear band in one track only.

20 30 40 50 60 70 80 90 100 110
! ! ! ! ! ! ! ! ! ! !

e. TAAAGGAGGTACACACCATTXATXACTXAATGCCATTACTCATCGTCATTGTAGGATGCCCTTGGTTCCAXT^GT^GCGGATGATGATATGCCAGA
* * * * *
f. TAAAGGAGGTACACACCATTGATGTACTTAATGCCATTACTCATCGTCATTGTAGGATGCCCTT^TGTTCCACTGTAGCGATGATGATGCCAGXGGTTA
* * * * *
g. GGTACACAXCXITGATGTACTTAXT^CCCATT^GCT^AATCGTCC^GTXT^TGGATGCCCTTGGCTXCTXTXTAGCGATGATGNTATGXCCAGATTTGTXA
* * * * *
h. TAAXTGTAGAAATGCCXT^GCGCTCCACTGTANCGATGANGA^CATGCXA^TATGGTCA
* * * * *
i. TGCC^ATTTACTCATCGTCATTGTAGGATGCCCTT^TGCTCCACTGTAX^TC^TGATXN^GGA^AATATATGGT^CA
* * * * *

! TAAAGGAGGTACACACCATTGATGTACTTAAATGCCATTACTCATCGTCATTGTAGGATGCCCTTGGCTCCACTGTAGCGATGATGATGCCAGATGGTCA
! ! ! ! !

120 130 140 150 160 170 180 190 200 210
! ! ! ! ! ! ! ! ! ! !

f. TGCTTAATAC
O*OOO***O
g. XGXTTAAATAXGXXTXXXTAAAGGXG
* * * * *
h. CGCTTAANACCGACTCATAAAGGAGACACTATAATG^CTTTCGACTTTCATTACACA^AAGCGTTAAGAXTTCA^XGG
* * * * *
i. T^GC^TT^TA^NA^CC^GA^CT^CA^C^{GA}T^GA^AG^GA^GA^CA^CT^AT^AT^GT^TC^GA^CT^TC^AT^TA^CA^CA^AAGCGTTAAGAAATTTCA^CGGTTCCCGGTGCTGA
C^TY^NO['] N
! CGCTTAATACCGACTAAGGAGACACTATAATGTTTCGACTTTCATTACACA^AAGCGTTAAGAAATTTCA^CGGTTCCCGGTGCTGA
! ! ! ! !

terminations or pausing, resulting in bands of varying intensity at a given position in one or more tracks.

The further phenomena listed below could generate minor families of polynucleotides which would be completely unrelated to the main sequence. If of sufficient intensity, these might confuse or even obliterate the require result. As such contaminants would probably be of substantially different base composition, they might have altered mobilities relative to the main sequence bands of corresponding length. This could result in extra bands appearing between the main sequence bands, or, alternatively, these contaminant bands could form the background ladder between whose "rungs" the main sequence bands might migrate.

4) One or more minor sequences might be generated by extensions of specific loopback(s) of either the template strand (which must also generate a recognition site for the relevant EndoR.) or one or other of the primer-fragment strands. Extensions arising from looped-back primer-fragment are likely to yield not only a minor sequence, but also a very strong band in all tracks since many such extensions would not be chain-terminated before reaching the 5'-terminus of the molecule.

5) Renatured, double-stranded primer fragments with 5'-single-stranded projections (produced either by the action of the endonuclease used to generate the fragment, or by 3'-degradation during preparation) will be extended to the ends of the molecules ('filled-in'). This will produce a large number of labelled molecules of identical length and thus a very strong band in all tracks (or a

double band, if the strands are of markedly different base composition).

6) A primed-template, whose primer 3'-terminus has been degraded, will, when extended, recut, and denatured, release a labelled fragment of primer length and thus generate a band in all tracks.

7) Complementary strand contamination in the single-stranded T7 DNA (prepared as described in Methods) is estimated to be not more than 5% (O'Hare, 1978; Peters, 1974). However, this may be sufficient to generate a minor sequence when hybridised to the corresponding strand of the primer fragment.

8) Contaminating site-specific endonucleases in the EndoR. preparations (which were used to produce a unique 5'-terminus following the extension reaction) might generate a family of extensions with a different 5'-terminus.

Virtually all the individual results compiled in Figure 31, are clearly discernible from the autoradiograms shown in Fig. 30a-k. A few, however, deserve special mention and are commented upon in the appropriate figure legends.

The bottom line for each 100 nucleotide section in Figure 31 gives the final conclusion for each position compiled from all the individual results. Only one such designation is not derived solely from the sequencing experiments. Position +85 must be C rather than T, because +82 to +85 is the only possible candidate for the HhaI site (GCGC) mapped in this region (see Chapter 4).

3. Structural Features

The sequence data leftward of position -163 (i.e. -164 through -177) are not considered firm enough to be included in the following analyses. These will deal, therefore, only with the region from -163 to +207 (including 9 uncertainties).

a) Base Composition

Excluding the 9 uncertainties this region is A + T rich, being 57% A + T and 43% G + C. Because the label used throughout was [α -³²P]-dGTP, there may have been a tendency for G residues to appear uncertain. If this is allowed for by assuming G wherever possible (i.e. in 8 of the 9 uncertainties), the region is at least 56% A + T.

b) Direct Repeats

There are three major direct repeats in this region. The most striking of these is the 21 bp-long sequence at +7 through +27 (DR1a), perfectly repeated at +124 through +144 (DR1b).

The second is the sequence from -160 through -142 (DR2a), which is nearly identical (17, 18 or 19 out of 19, depending on the uncertainties at -158 and -149) to that at + 139 to +157 (DR2b).

Finally, there is a nearly tandem duplication at -96 through -87 (DR3a) and -84 through -74 (DR3b) which is exact if the G at -78 is ignored. A sequence closely related to these occurs at -20 through -11 (DR3c): it differs by two A to G transitions at -13 and -11.

c) Inverted Repeats

Inverted repeats in a DNA sequence indicate the possibility

that the corresponding RNA transcript might be able to form a stem-loop structure through intrastrand basepairing. The stability of such potential structures can be estimated according to the parameters defined by Tinoco et al (1973). A computer was therefore programmed to search for any inverted repeats in the region sequenced, and to calculate the free energy change if a transcript were to assume a stem-loop structure involving these repeats.

Of the many theoretical structures produced by this programme, only two were found to have a predicted free energy change (ΔG) of greater than -10 kcal/mol. The less stable of these is a branched basepaired structure (i.e. containing two "hairpin" loops) involving nucleotides +59 through +116 (Fig. 32a). It has a predicted ΔG of formation of -13 kcal/mols: however, the effect of the branching on its stability cannot be predicted from the data of Tinoco et al, so that -13 kcal/mol may be an over- or underestimate of the true value.

An extremely large, but imperfect, inverted repeat exists between positions -133 and +14. An RNA transcript of this region can be drawn in a complex structure containing many unpaired nucleotides but only one "hairpin" loop (Fig. 32b). The predicted ΔG of formation for this structure is -49.2 kcal/mol, assuming that the G/C ambiguity at position -132 is in fact C, and that nucleotides -30 and -31 are both G residues. Different nucleotides at these three positions could alter the total stability of the structure by as much as +18.4 kcal/mol, i.e. to a ΔG of formation of -30.8 kcal/mol.

Other factors which might affect the formation of these structures are discussed in Chapter 7.

Figure 32. Most Stable Predicted RNA Secondary Structures

- a. Possible base-pairing in an RNA transcript containing the region from +59 through +116 (numbering of nucleotides is as in Fig. 30). The predicted free energy change of formation for this structure (Tinoco et al, 1973) is -13 kcal/mol. The contribution of each segment to this overall value is given alongside the drawing.

- b. Possible base-pairing in an RNA transcript containing the region from -133 through +14. The predicted free energy change of formation for this structure is -49.2 kcal/mol. Other details are as for a.

Figure 32a.

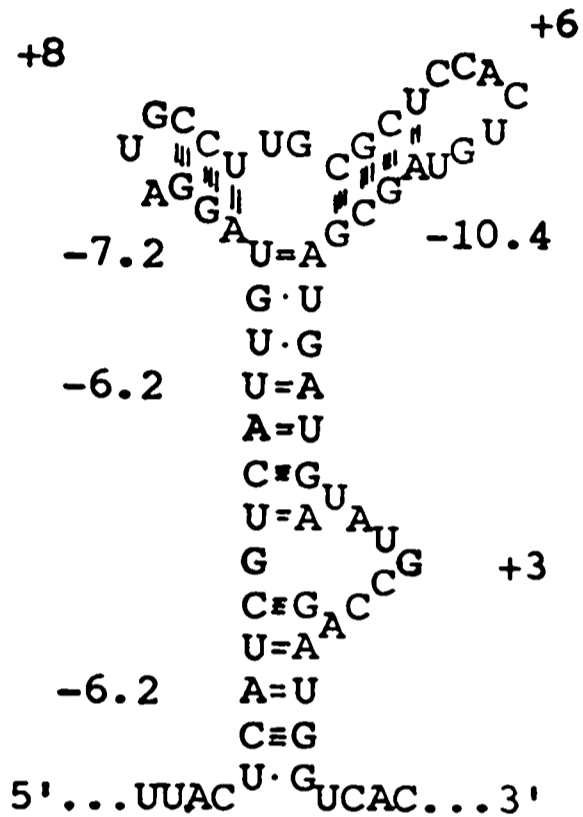
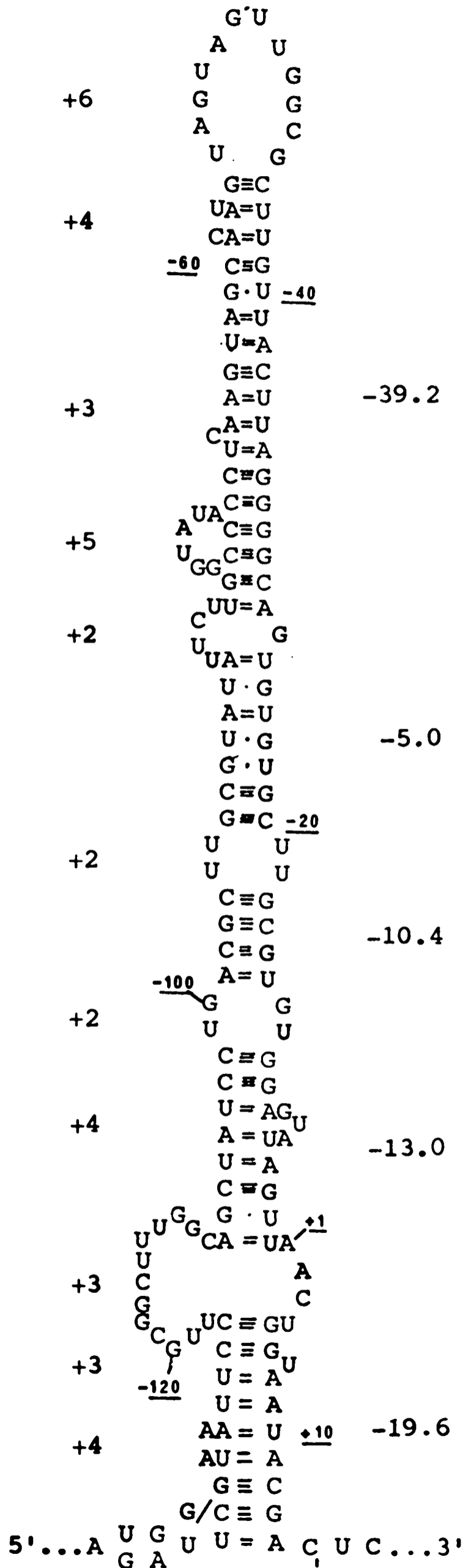


Figure 32b.

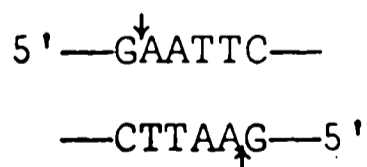


CHAPTER 7 - DISCUSSION

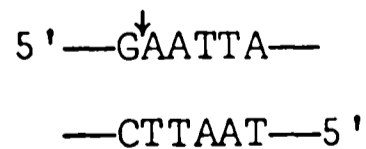
1. Mapping of EndoR. Cutting Sites

The results presented in Chapter 4 are completely self consistent and are also in full agreement with the published work of other groups (Ludwig and Summers, 1976; McDonnell et al, 1977; Gordon et al, 1978). Moreover, with the two exceptions already discussed (i.e. HpaII M4 and AvaI D), my results are confirmed by the unpublished work of F.W. Studier (Fig. 6; personal communication).

Nucleotide sequencing in the region 18.85 to 19.8 on the physical map has confirmed the mapping data and indicated only one unexpected site: Fig. 30b shows that a faint but significant band is present in the H track, at position -61. This track contained the products of a double digestion (HaeIII and HgiAI) of labelled extensions from a HaeIII L2 primer hybridised to the r-strand of T7. This result predicts that -65 to -60 is recognised by HgiAI, whereas the mapping results described in Chapter 4 indicated no such site in this region. Recently, however, Bishop (1979) has identified what he terms EcoRI "half-sites", i.e. sites at which EcoRI will apparently recognise and cut only one of the DNA strands:

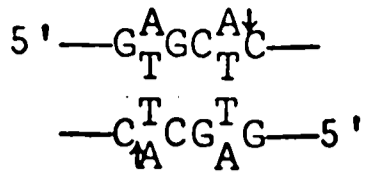


EcoRI site

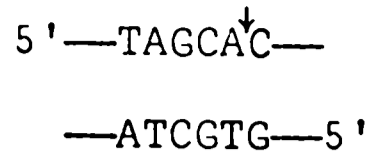


EcoRI "half-site"

Such a phenomenon might explain the apparent conflict in my HgiAI results, since by analogy to EcoRI, -65 through -60 may be a HgiAI half-site:



HgiAI site



HgiAI "half-site"

Such half-sites would not be detected under non-denaturing conditions such as those used in the mapping studies of Chapter 4, whereas they would be (if the cut strand were labelled) under the denaturing conditions of a sequencing gel.

2. Fragment Binding to Nitro-Cellulose

The results described in Chapter 5 are completely consistent with the predicted positions of the T7 early promoters in relation to the HaeIII cutting sites (Studier, personal communication, Fig. 6; Gordon et al, 1978). Furthermore, my results suggest that termination of transcription at the end of the early operon occurs within HaeIII A: that is, to the right of 18.85 on the T7 physical map. This technique may prove of general usefulness in studies of this sort.

3. Nucleotide Sequencing

a. Artefacts

The adaptation of the chain-terminating method for sequencing DNA to the T7 system required several modifications before a satisfactory result was routinely obtained. Nevertheless several artefacts do persist, as described in Chapter 6, and thus multiple attempts using different primers and extensions in both directions were necessary to generate a reasonably clear and unambiguous sequence. Uncertainties which remain even after this might be resolved

using a different $[\alpha\text{-}^{32}\text{P}]\text{-dNTP}$ (dGTP was the labelled nucleotide used in all these experiments) and/or by employing the single-site ribosubstitution method of Brown (1978). This is a modification of the chain-terminating method wherein instead of releasing extensions from primers by cutting with a restriction endonuclease, a single rNMP is incorporated into the 3'-terminus of the primer before the extension reaction, so that the extensions can be released by alkali digestion (alkali specifically cleaves the phosphodiester based on the 3' side of the ribose sugar group). This modification overcomes the problem of generating two sequence results when two recognition sites for the same endonuclease are in close proximity (one of which represents the 3'-end of the primer).

Ideally, however, uncertainties in the sequence result might best be dealt with by employing a completely independent method for sequencing DNA, such as that developed by Maxam and Gilbert (1977).

It is interesting to note that the tentative assignments reached in individual experiments are rarely contrary to the final conclusion. This suggests that the four tentative G assignments in my sequence are likely to prove correct. Moreover, Sanger (1977) has noted that in a run of identical nucleotides, the first position may give a heavy band in gel analyses whereas subsequent ones may yield substantially weaker signals. This probably explains the uncertainty of the G designations as positions -30 and -31 as these are the distal two nucleotides in a run of four tandem G residues, and this region has been sequenced in one direction only.

b. Functional Features

i. Termination of Transcription at T7 t₁:

On the basis of filter binding studies described in Chapter 5 (i.e. HaeIII A contains at least a part of the early operon, and deletion studies in Chapter 4 (i.e. ΔLG37, but not ΔLG3, deletes the terminator function), T7 t₁ should lie within the region -177 to +132. The latter limit is based on the assumption that ΔLG37, which extends no further to the right than +82, could not affect termination at a site greater than 50 bp downstream. Hence, I conclude that T7 t₁ lies within the sequence shown in Figure 33.

Since this terminator is rho-independent in vitro (Millette et al, 1970; Peters and Hayward, 1974a,b), one might expect it to possess, a priori, the three features found in common to such terminators, as described in Chapter 1. If so, termination should occur in a run of 4-8 T:A or A:T base-pairs, immediately following a G + C rich sequence within a region of two-fold rotational symmetry. However, nowhere in the sequence shown in Figure 33 are all three of these features observed. Moreover, there is no region which displays to any significant extent even two such features.

Rosenberg et al (1978) and Klüpper et al (1978) have suggested that CAATCAA may represent a part of the termination signal at sites which are rho-dependent in vitro (Chapter 1). This sequence might be, however, of some significance even at rho-independent sites and thus Figure 33 was scanned for similar sequences. The closest

Figure 33. Nucleotide Sequence and Proposed Functional Sites between Map Positions 18.85 and 19.81

The nucleotide sequence shown in Figure 31 is presented here with the functional sites as proposed in the text.

Only the 1- or sense-strand of DNA is shown: nucleotides are numbered from the HpaI site at 19.29 (McDonnell et al, 1977).

Nucleotide designations:

A alternative possibilities
G
g preliminary or possible
· tentative or probable
G certain

Translational Features:

TGA
*** stop-codon and phase
3

ATG
... start-codon and phase
1

AGG region of complementarity to 3'-terminus of 16S rRNA

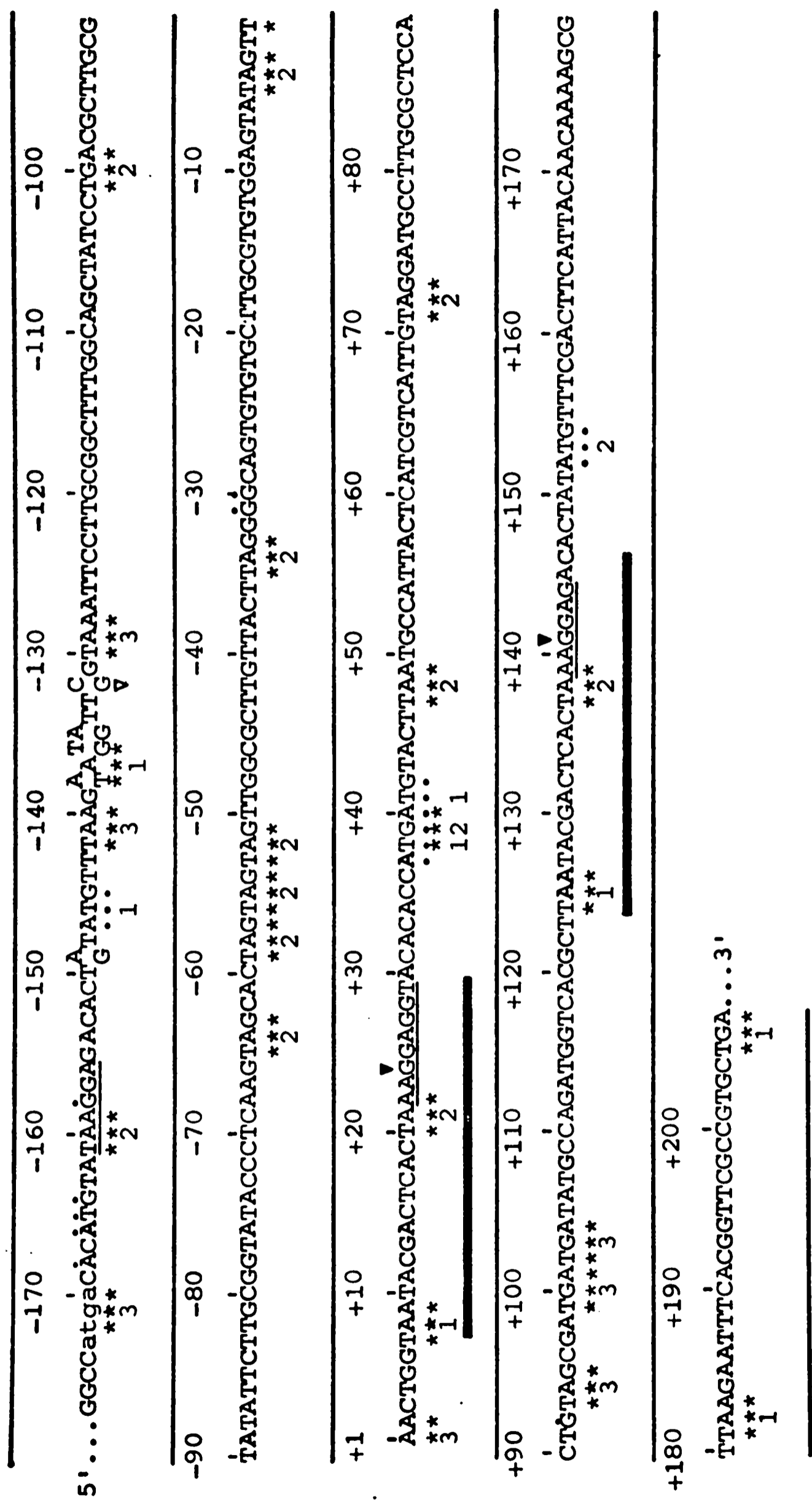
Transcriptional Features:

C
G class I RNA stop-site
▽

ATA class II promoter

▽
G class II RNA start-site.

Figure 33.



homology observed was 5/7 at positions +17 to +23, +134 to +140 and +169 to +175. It would be surprising if such a low degree of homology were to prove significant.

Previous work from this laboratory (Peters and Hayward, 1974a,b) clearly indicated $Y\overset{\downarrow}{C}G\overset{A}{T}$ as the most likely sequence surrounding the RNA stop site at T7 t_1 (the arrow indicates the RNA 3'-terminus). To obtain this result, RNA synthesised on T7 DNA (using purified E.coli MRE600 Rpol) was hybridised to its complementary strand and extended with DpolI and $[\alpha\text{-}^{32}\text{P}]\text{-dGTP}$. Specifically rCM ^{32}P (2', 3') could then be released by alkali or pancreatic RNAase, indicating the sequence $Y\overset{\downarrow}{C}G$ at the RNA stop site. Other $[\alpha\text{-}^{32}\text{P}]\text{-dNTPs}$ yielded no significant rNMP under the same conditions. Similar, unpublished experiments using RNAase U_2 indicated that the RNA 3'-terminal sequence was probably $A Y_{1-3} C_{OH}$, while RNAase T_1 released a considerably longer terminal oligonucleotide (Peters and Hayward, personal communication).

Using a completely different approach Peters and Hayward (1974c) suggested that the RNA terminates with three successive C residues. (This approach attempted to sequence directly the 3'-end of the RNA, using the NaB^3H_4 method of end-labelling RNA.)

Nowhere in Figure 33 is the sequence CCCG observed. Thus the latter conclusion probably must be discounted. However, allowing for ambiguities, $Y\overset{\downarrow}{C}G\overset{A}{T}$ is present at three places: -133 to -130, +62 to +65 and +157 to +160. Since, ΔLG37 extends no further to the right than +82 (Chapter 4), but deletes the terminator function, it seems highly unlikely that +159 represents the RNA stop site. The other

two candidates cannot be distinguished on these grounds. To decide between them I shall therefore consider the relevant results so far obtained in other laboratories.

Millette et al (1970), also using the NaB^3H_4 method of 3'-end labelling RNA, found that T7 early RNA synthesised in vitro is terminated predominantly with U (75%) and that the 3'-terminal T_1 -released oligonucleotide is a single species approximately 10 nucleotides long. Maitra et al (1970) using alkali digestion also found predominantly U at the 3'-terminus. The apparent discrepancy of this result from that of Peters and Hayward (1974a,b) may be due to strain differences in the respective sources of Rpol, or conceivably in the T7, itself. Alternatively, as noted in Chapter 1, the termination factor rho can shift the RNA stop-site one base-pair downstream at otherwise "rho-independent" terminators in vitro. In both candidate stop-sites (i.e. -132 and +63), U is the predicted penultimate RNA nucleotide, so that rho or some other contaminating factor in the different Rpol preparations might be altering the actual stop-site by one nucleotide within the same terminator region.

Kramer et al (1974) were unsuccessful in their attempts to determine the sequence at the 3'-terminus of gene 1.3 mRNA synthesised in vivo. However, they did note that the 3'-terminal T_1 oligonucleotide was heterogeneous in length, being about 5-7 nucleotides long, and that it was pyrimidine-rich (but not exclusively U; M. Rosenberg, personal communication). The difference between this result and that of Millette et al (1970) may also be attributable to

strain differences or more likely to a difference between the in vivo and in vitro termination events. Such a difference has been observed in the case of trpa (Fig. 2).

In its present form, candidate -132 is in no way inconsistent with the T1 digestion results of Kramer et al (1974). However, because of ambiguities at position -138, -136 and -135, comparisons at this stage are not very significant. The results of Millette et al (1970), on the other hand, are somewhat at variance with the -132 candidate as the latter predicts a terminal T1-digestion product of at most 7 nucleotides (the G residue at -139 is definite). The +63 candidate predicts a product 12 nucleotides in length. This is in only slightly better agreement with Millette et al (1970) and quite different from the results of Kramer et al (1974).

Using T7 early mRNA synthesised in vivo, McAllister and Barrett (1977a) showed substantial hybridisation with T7/HpaI Q, but not with T7/HpaI C. They therefore suggested that the RNA-terminus lies near to and probably to the left of the HpaI C/Q boundary (+1/-1 in Fig. 33). This result favours the -132 candidate.

Recent, unpublished results of Pries and Delius (personal communication) indicate that the RNA stop-site is within HpaI Q, about 147 ± 20 base-pairs from its right end (i.e. -147 ± 20 in Fig. 33). In their experiments, T7 early mRNA synthesised in vitro was hybridised to T7/HpaI fragments under conditions favouring the formation of RNA:DNA hybrids and the length of the hybrid estimated by electron microscopy.

Although no details have yet been obtained, Studier (personal communication) has also very recently revised his estimate of the RNA terminus to near position 18.9, or about -156 in Figure 33.

Both these results clearly favour -132 as the more likely of the two candidate sites.

Finally, it should be noted that, although the sequence preceding the -132 candidate (as shown in fig.33.) does not contain any two-fold rotational symmetry, it may represent part of an unusually large inverted repeat, the centre of symmetry for which lies near to or beyond the left limit of my reliable sequence data (i.e. -163). If this is the case, a stable stem-loop structure might, in fact, be able to form in a transcript terminated at -132. Such a structure might, then, be involved in t_1 functioning as generally suggested by Adhya and Gottesman (1978).

Taken together, all the above information suggests that transcription of the early operon in T7 terminates at position -132 or about 18.96 on the physical map of T7.¹ This conclusion raises two important questions.

First, why is T7 t_1 apparently so different in its sequence composition from all other "rho-independent" terminators so far sequenced? As already mentioned in Chapter 1, the terminators shown in Figure 2, though independent of rho in vitro, are not necessarily so in vivo. Indeed the activity of several of these is severely decreased in mutant strains containing only partially active rho. (The rho⁻ phenotype appears to be lethal so that definitive testing in vivo is difficult to achieve.) Hence, it may be that T7 t_1 will prove to be the only truly rho-independent terminator in vivo. (No rho mutants have

1. This site, although A + T rich, exhibits none of the features commonly found in rho-independent terminators (i.e. it apparently has no two-fold rotational symmetry, and no run of 4-8 T:A or A:T bp preceded by 2 or more G:C or C:G bp; see chapter 1).

yet been found which detectably alter the activity of T7 t₁ in vivo; Kieffer et al, 1977.) At this point it is difficult to say what difference complete rho-independence might make in the optimal terminator sequence.

A possibly more important distinction of T7 t₁ is the fact that it must not be recognised by T7 Rpol: this enzyme freely transcribes the region between positions 14.6 and about 60 on the T7 physical map, both in vitro (Golomb and Chamberlin, 1974) and in vivo (Pachl and Young, 1978). Attempts to explain the significance of the average rho-independent terminator sequence (see Chapter 1) have so far **concentrated on physical or energetic considerations, which do not include a specific interaction with E. coli Rpol.** If this is indeed the correct interpretation of their mode of action (which is by no means established), then such terminators should presumably be just as active in terminating transcription by T7 Rpol as by E.coli Rpol. However, this is clearly not the case at T7 t₁. Thus, a different sort of terminator might be expected at this site: that is, one which is specifically recognised by E.coli Rpol by virtue, for example, of its primary nucleotide sequence rather than by any higher order structural or energetic considerations. This I consider to be a quite likely explanation for the so far exceptional terminator sequence at T7 t₁. In this regard, it is interesting to note that the Rpol's from seven different genera of bacteria are all capable of recognising T7 t₁ in vitro, albeit to varying extents (Wiggs et al, 1979). Although on first consideration this might again suggest an energetic or physical block,

the ability to accurately recognise specific transcriptional signals on the DNA of a different genus may be of substantial survival value in allowing the efficient and faithful expression of exchanged genetic information.¹

The second main question raised by the proposed position of the terminator is whether the putative ribosome binding site at -161 through -145 (see below) plays a role in the regulation of termination at T7 t_1 . A model for such coupling has been formulated by Zurawaki et al (1978) to explain the regulation of attenuation of trpA (see also Chapter 1). However, there is, as yet, no evidence for the suggestion, nor any reason to expect, that T7 t_1 is a regulatable terminator.

ii. Class II Transcription

The class II genes of T7 (located in the region 15 to 45 on the physical map, and expressed from about 4 to 16 min after infection; McAllister and Wu, 1978) are transcribed by the early (class I) gene 1 product, T7 Rpol (Chamberlin et al, 1970). Oakley et al (1975) found that purified T7 Rpol could bind to T7 /HpaII C2, suggesting that a T7 Rpol promoter is located on this fragment

1. A third, alternative explanation of this unusual terminator sequence might be that by the time a class I transcription complex reaches t_1 in vivo, the polymerase may have already been phosphorylated by the T7 0.7 protein kinase activity. Such modified polymerases might require a completely different signal for termination.

(i.e. between 16.9 and 20.1 on the T7 map). However, the specificity of this binding is unclear, because only the largest of the T7/HpaII fragments were retained on the Nitrocellulose filters.

More recently, Pachl and Young (1978) have isolated an in vivo class II transcript which extends from about the end of gene 1.3 (i.e. near map position 19) to the right, including genes 3.5 and dup. These authors argue that this is not a product of RNAase III digestion of a larger transcript, though they cannot exclude the possibility that it is a result of some other processing event.

Finally, Kassavetis and Chamberlin (1979) have mapped a class II promoter active in vitro to about position 18.9. Hence, the sequence shown in Fig. 33 may include a T7 Rpol class II promoter. To investigate this possibility, the sequence was scanned for homology to the class II promoter (map position 14.6) sequenced by Oakley and Coleman (1977). This revealed two sequences (wholly contained within the perfect direct repeats DR1a and DR1b, i.e. at +7 through +27 and +127 through +144) which are almost exactly homologous to the Oakley and Coleman sequence. Moreover, four class III promoters (also transcribed by T7 Rpol but expressed from 8 min after infection until lysis) have recently been sequenced (Rosa, 1979; Oakley et al, 1979), and these also show this homology. All these sequences are presented for comparison in Fig. 34a.

By analogy with the transcripts initiated at the other T7 Rpol promoters, the 5'-sequence of transcripts initiated at both these promoters would be pppGGAG-. Consistent with

Figure 34. Nucleotide Sequence Homologies within T7

a. Late Promoters

The nucleotide sequences of one class II (Oakley and Coleman, 1977) and four class III (Rosa, 1979; Oakley et al, 1979) T7 Rpol promoter regions are compared with those of the two putative class II promoters identified in Figure 33. (The numbering of the nucleotides in the latter two cases is as in Fig. 33.) Only the 1- or sense-strand sequence is shown.

Approximate locations on the physical map of T7 and the class of the promoter are given on the right.

The RNA start-site where known is indicated by * above the appropriate nucleotide.

The sequences are aligned by homology with the 23 base-pairs (enclosed within the box) which are identical in the four class III promoters. Non-homologous nucleotides within this box are underlined.

b. Ribosome-Binding Sites

The nucleotide sequence of the gene 0.3 ribosome binding site (rb-0.3; Steitz and Bryan, 1977) and of the proposed gene 1 site (rb-1; McConnell, 1979b) are compared with the three putative sites suggested for genes 1.4 (rb-1.4), 1.5 (rb-1.5) and 1.7 (rb-1.7). (The numbering of nucleotides in the latter three cases is as in Fig. 33.) The sequences are aligned by the AUG codon. This codon and the region complementary to the 3'-end of 16S rRNA are indicated by dots beneath the sequence.

Other homologous regions are underlined; these include an almost perfectly conserved tetranucleotide sequence ACAC, occurring 4-6 nucleotides before the AUG codon, and a somewhat less complete homology with the heptanucleotide sequence ACUCACU, centred 17-20 nucleotides before the AUG codon.

Figure 34a.

			T ₇ Map	
			Position	Class
5'...GAAAT	TAATACGACTCACTATAG [*] GGAGA	CCACA...3'	57	III
GGAAA	TAATACGACTCACTATAG [*] GGAGA	GGCGA	87	III
GAATT	TAATACGACTCACTATAG [*] GGAGA	CCTCA	55	III
TAAAT	TAATACGACTCACTATAG [*] GGAGA	TAGGG	46.5	III
CAAAT	<u>CA</u> ATACGACTCACTATAG [*] <u>AG</u> GGGA	CAAAC	14.6	II
+2		+34		
ACTGG	TAATACGACTCACTAA <u>AG</u> GGAG <u>GT</u>	ACACA	19.35	II ?
+119		+151		
ACGCT	TAATACGACTCACTAA <u>AG</u> GGAG <u>AC</u>	ACTAT	19.65	II ?

Figure 34b.

5'..... <u>AACUGCACGAGGU</u> <u>AACACAAGAUG</u> ...3'	rb-0.3
GAUU <u>UACUAACUGGAAGAGGCACUAAAUG</u>	rb-1
-173 aug <u>aCaCaug</u> UAUAA <u>gGAGACACU</u> ^A <u>UAUG</u>	rb-1.4
+11 ACG <u>ACUCACUAAAGGAGGUACACACCAUG</u>	rb-1.5
+126 AUACG <u>ACUCACUAAAGGAGACACU</u> UAUG	rb-1.7

this, Bishayee et al (1976) found predominately pppGG(A) at the 5' end of T7 late transcripts.

Therefore, these two neighbouring sequences, which are located at about 19.35 and 19.6 on the T7 physical map, almost certainly represent the promoter mapped by Kassavetis and Chamberlin (1978) to 18.9. The method used by these authors (i.e. RNA sizing) was approximate and hence, as they point out, two species of RNA of roughly similar size would not necessarily have been resolved by their analyses. Furthermore, they have mapped the leftmost T7 Rpol promoter to about 13.5, whereas Oakley and Coleman (1977) using more accurate means found that this promoter is about 60 bp from the right end of HpaII L3, i.e. at 14.6 on Studier's map (Fig. 6). The position of my pair of sites is therefore well within the limits of the apparent experimental error of Kassavetis and Chamberlin's mapping.

iii. Translation

(1) Gene 1.3 Translational Stop-Site

Gene 1.3 is translated from early mRNA (Simon and Studier, 1973) and thus must lie to the left of T7 t_1 , i.e. probably to the left of -132. The relevant stop codon for this gene therefore is at or to the left of -137 (codons are numbered according to their middle nucleotide), depending on its phase of translation. It is interesting to note that if -137 is indeed the terminus for gene 1.3 translation it would be only the fourth active UAG out of 54 stop codons so far known for E.coli and its phages (H. Grosjean, compiled for tRNA Book II, Cold Spring Harbor

Laboratory, 1979: personal communication). Preliminary evidence (B. Will and R. Hayward, personal communication using an amber-suppressing host, suggests that the gene 1.3 coding region could not end with UAG. Finally, if the ambiguities at positions -136 and -138 (Fig. 33) are resolved so that the -137 codon is not UAG, then -141 is the rightmost possible stop codon for gene 1.3.

(2) Translation of Other Genes

A class II transcript (Pachl and Young, 1978) of the sequence shown in Fig. 33 would contain three regions which might act as ribosome binding (rb) site. These are tentatively designated rb-1.4 (-161 to -145), rb-1.5 (+22 to +39) and rb-1.7 (+139 to +154). This prediction is based on the results of Shine and Dalgarno (1974) and Steitz and Jakes (1975) which indicate that rb-sites typically contain a sequence of 3-9 nucleotides which is complementary to a region near the 3'-terminus of 16 S rRNA, and which is followed at a suitable distance (about 7 nucleotides) by an AUG (or rarely a GUG) start codon. The nucleotide sequences of rb-1.4, -1.5 and -1.7 show additional homology to each other as well as to the rb sites of T7 gene 0.3, rb-0.3 (Steitz and Bryan, 1977), and T7 gene 1, rb-1 (McConnell, 1979b). This further supports the possibility of a physiological significance for these sequences. They are all presented in Fig. 34b for comparison.

rb-1.4 may be the initiation codon for translation of a previously unreported gene, 1.4. This gene would be translated in phase 1 and, assuming that the -137 codon is not UAG, would yield a polypeptide 51 amino acids long

(about 6000 daltons), terminating at the UAA codon at +8. If gene 1.3 extends to within this sequence, its vigorous translation might prevent gene 1.4 activity. However, as demonstrated in the case of ØX174 (Barrell et al, 1976; Smith et al, 1977), the expression of such overlapping genes need not be mutually exclusive in vivo.

rb-1.5 may represent the start of another previously unreported gene, 1.5. This gene must lie outside gene 1.3, as it is preceded by stop codons in all three phases. Translation (in phase 1) could initiate at either of the tandem AUG codons (+38 or +41) and terminate at the UAA at +125, yielding a polypeptide comprising 29 or 28 amino acids (about 3000 daltons). This putative gene could not be gene 1.7, as the latter's product has an estimated length of at least 135 amino acids (Simon and Studier, 1973).

If real, the class II genes 1.4 and 1.5 would both be deleted from the T7 LG37 mutant (which can grow on E.coli WT) and must therefore be inessential. It is interesting to note that Studier (personal communication) has found a number of small, previously undetected, T7 specific proteins within the class I (early) region. It is also relevant to note that recent in vitro studies with λDNA have indicated rb-sites where no gene has yet been mapped genetically (Calame and Ihler, 1977).

rb-1.7 may represent the start of gene 1.7, as they both lie outside the region affected by ΔLG37, and, if active, rb-1.7 is the initiation site (phase 2) for a polypeptide at least 17 amino acids long. (No phase 2 stop codons are present to the right of +153 in the sequence so far determined; Fig. 33).

c. Physical Features

(1) Base Composition

The significance of the base composition of the region sequenced (56-57% A + T) is unclear at present. It may reflect no more than random variation from the overall T7 DNA composition of 52% A + T (Studier, 1969).

(2) Direct Repeats

The largest direct repeats (DR1a and b) observed in this sequence contain: 1) the stop codons for the two putative genes 1.4 and 1.5; 2) the 16 S rRNA complementarity implicated in rb-1.5 and rb-1.7; and 3) the two putative class II T7 Rpol promoters at 19.35 and 19.65 on the T7 physical map.

The second most striking direct repeats (DR2a and b) contain rb-1.4 and rb-1.7.

The proposed similarities in function, therefore, may explain the otherwise surprising degree of homology within these two pairs of repeats. DR1a and b may be of particular interest as highly economical intercistronic regions.

No likely function has yet been specifically ascribed in the case of DR3a and b.

(3) Inverted Repeats

It may be significant that the very large region of two-fold rotational symmetry represented in Fig. 32b, lies between but not including rb-1.4 and rb-1.5. This could reflect a requirement that active rb-sites be free of any secondary structure.

Because of their location, ^{neither} / of the structures represented in Fig. 32 is likely to be involved in termination at T7 t_1 .

4. Further Studies

These results suggest several interesting experiments which might now be carried out. Such further studies would test and extend the conclusions reached in this thesis.

a) The filter binding studies should be repeated using HaeII and HpaI to further delineate the terminator region, and to further explore the usefulness of this technique (which was developed in collaboration with Kevin O'Hare).

b) The ambiguities and uncertainties in the nucleotide sequence might be eradicated through an extension of the sequencing work, preferably using the method of Maxam and Gilbert (1977) to minimise the chance of repetition of some of the artefacts found in the present studies. This should help in defining the RNA stop-site by the criteria already discussed.

c) The region sequenced should be extended to the left and right to check the predictions regarding genes 1.3 and 1.7, and to investigate the possibility of a large stem-loop structure upstream of T7 t_1 .

d) in vitro ribosome-binding and -protection studies using mRNA and/or DNA containing rb-1.4, -1.5 and -1.7, could help determine whether these rb-sites are physiologically significant (Pieczenik et al, 1974; Calame and Ihler, 1977).

e) The activity of the proposed class II promoters at 19.35 and 19.65 on the physical map should be investigated in vitro using purified fragments of T7 DNA and T7 Rpol in experiments analogous to those of Oakley and Coleman (1977) and Rosa (1979).

f) The construction and molecular cloning of a recombinant DNA molecule carrying T7 t_1 closely linked to T7 p_A would enable an RNA of manageable size to be isolated and sequenced (Donnis-Keller et al, 1977; Stanley and Vassilenko, 1978; Kramer and Mills, 1978). This could be used to identify the exact RNA stop-site in vitro, and possibly in vivo.

g) The insertion of T7 t_1 between an active promoter (e.g. T7 p_A) and a "selectable" gene (e.g. E.coli lacZ) could enable terminator mutants to be isolated and sequenced to better define the functionally important features of T7 t_1 . The recently developed techniques of in vitro mutagenesis might prove especially useful in this regard (Hutchison et al, 1978; Muller et al, 1978; Shortle and Nathans, 1978).

References

- Adhya, S. and M. Gottesman. 1978. *Ann.Rev.Biochem.* 47, 967-996
- Adhya, S., M. Gottesman and B. De Crombrughe. 1974. *Proc.Nat.Acad.Sci.USA* 71, 2534-2538
- Adhya, S., M. Gottesman, B. De Crombrughe and D. Court. 1976. In RNA Polymerase, ed. R. Losick and M. Chamberlin. New York, Cold Spring Harbor Laboratory
- Allet, B.R., R. Roberts, R. Gesteland and R. Solem. 1974. *Nature* 249, 217-221
- Armstrong, V.W., H. Sternbach and F. Eckstein. 1976. *Biochem.* 15, 2086-2091
- Barrell, B.G., G.M. Air and C.A. Hutchison III. 1976. *Nature* 264, 34-41
- Beckwith, J., T. Grodzicker and R. Arditti. 1972. *J.Mol.Biol.* 69, 155-160
- Bertrand, K. and C. Yanofsky. 1976. *J.Mol.Biol.* 103, 339-349
- Bertrand, K., C. Squires and C. Yanofsky. 1976. *J.Mol.Biol.* 103, 319-337
- Bertrand, K., L.J. Korn, F. Lee and C. Yanofsky. 1977. *J.Mol.Biol.* 117, 227-247
- Bishayee, S., A. Yudelevich and U. Maitra. 1976. *Biochem. Biophys.Res.Comm.* 68, 626-633
- Bishop, J.O. 1979. *J.Mol.Biol.* 128, 545-559
- Brown, N.L. 1978. *FEBS Lett.* 93, 10-15
- Burgess, R.R. 1969. *J.Biol.Chem.* 244, 6168-6176
- Burgess, R.R. 1976. In RNA Polymerase, ed. R. Losick and M. Chamberlin, pp. 69-100. New York, Cold Spring Harbor Laboratory.
- Calame, K. and G. Ihler. 1977. *J.Mol.Biol.* 116, 841-853
- Cameron, J.R. and R. Davis. 1977. *Nucleic Acids Res.* 4, 1429-1448
- Cassani, G., R.R. Burgess, H.M. Goodman and L. Gold. 1971. *Nature New Biol.* 230, 197-200
- Chamberlin, M.J. 1976. In RNA Polymerase, ed. R. Losick and M. Chamberlin, pp. 17-67. New York, Cold Spring Harbor Laboratory.

Chamberlin, M.J., J. McGrath and L. Waskell. 1970.
Nature 228, 227-231

Darlix, J.L. 1974. Biochimie 56, 693-701

Darlix, J.L. and P. Fromageot. 1972. Biochimie 54, 47-54

Darlix, J.L. and M. Horaist. 1975. Nature 256, 288-292

Das, A., D. Court and S. Adhya. 1976. Proc.Nat.Acad.Sci.
USA 73, 1959-1963

Das, A., C. Merril and S. Adhya. 1978. Proc.Nat.Acad.Sci.
USA 75, 4828-4832

Dausse, J.-P., A. Sentenac and P. Fromageot. 1976.
Eur.J.Biochem. 65, 387-393

De Fillipes, F.M. 1974. Biochem.Biophys.Res.Comm. 58,
586-596

Dickson, R., J. Abelson, W. Barnes and W. Reznikoff.
1975. Science 187, 23-35

Dickson, R.C., J. Abelson, P. Johnson, W.S. Reznikoff and
W.M. Barnes. 1977. J.Mol.Biol. 111, 65-75

Di Nocera, P.P. F. Blasi, R. Dilauro, R. Frunzio and
C.R. Bruni. 1978. Proc.Nat.Acad.Sci.USA 75, 4276-4280

Donis-Keller, H., A. Maxam and W. Gilbert, 1977. Nucleic
Acids Res. 4, 2527-2538

Dunn, J.J. and F.W. Studier. 1973. ProcNat.Acad.Sci.USA
70, 1559-1563

Dunn, J.J. and F.W. Studier, 1975. J.Mol.Biol. 99, 487-499

Dunn, J.J., E. Buzash-Pollert and F.W. Studier. 1978.
Proc.Nat.Acad.Sci.USA 75, 2741-2745

Engback, F., C. Gross and R. Burgess. 1976. Mol.gen.Genet.
143, 291-295

Epp, C. and M.L. Pearson. 1976. In RNA Polymerase, ed.
R. Losick and M. Chamberlin, pp. 667-691. New York, Cold
Spring Harbor Laboratory

Franklin, N. 1974. J.Mol.Biol. 89, 33-48

Friedman, D.I., G.S. Wilgus and R.J. Mural. 1973. J.Mol.Biol.
81, 505-516

Friedman, D.I., M. Baumann and L.S. Baron. 1976. Virol.
73, 119-127

Fujiki, H., P. Palm, W. Zillig, R. Calendar and M. Sunshine.
1976. Mol.gen.Genet. 145, 19-22

- Fukuda, R. and A. Ishihama. 1974. *J.Mol.Biol.* 87, 523-540
- Gilbert, W. 1976. In RNA Polymerase, ed. R. Losick and M. Chamberlin, pp. 193-205. New York, Cold Spring Harbor Laboratory
- Gilbert, W.N., N. Maizels and A. Maxam. 1974. *Cold Spring Harbor Symp.Quant.Biol.* 38, 845-855
- Godson, G.N., B.G. Barrell, R. Staden and J.C. Fiddes. 1978. *Nature* 276, 236-247
- Goldberg, A.R. and J. Hurwitz. 1972. *J.Biol.Chem.* 247, 5637-5645
- Golomb, M. and M. Chamberlin. 1974. *Proc.Nat.Acad.Sci.USA* 71, 760-764
- Gordon, R.L., P. Humphries and D.J. McConnell. 1978. *Mol.gen.Genet.* 162, 329-339
- Gottesman, S. and S. Adhya. 1977. In DNA Insertion Elements, Plasmids and Episomes, ed. A.I. Bukhari, J.A. Shapiro and S.L. Adhya, pp. 712-718. New York, Cold Spring Harbor Laboratory.
- Guarente, L. 1979. *J.Mol.Biol.* 129, 295-304
- Guarente, L.P. and J. Beckwith. 1978. *Proc.Nat.Acad.Sci. USA* 75, 294-297
- Guarente, L.P., D.H. Mitchell and J. Beckwith. 1977. *J.Mol.Biol.* 112, 423-436
- Hausmann, R. 1976. *Curr.Topics in Microbiol. and Immunol.* 75, 77-110
- Hercules, K., S. Jovanovich and W. Sauerbier. 1976. *J.Virol.* 17, 642-658
- Hershey, A.D. (ed.) 1971. The Bacteriophage λ . Cold Spring Harbor Laboratory, New York.
- Hesselbach, B.A. and D. Nakada. 1977. *J.Virol.* 24, 736-745
- Heyden, B., C. Nüsslein and H. Schaller. 1975. *Eur.J. Biochem.* 55, 147-155
- Hutchison III, C.A., S. Phillips, M.H. Edgell, S. Gillam, P. Jahnke and M. Smith. 1978. *J.Biol.Chem.* 253, 6551-6560
- Hillel, Z. and C.-W. Wu. 1978. *Biochem.* 17, 2954-2961
- Imoko, H. and M. Imai. 1976. *Mol.gen.Genet.* 143, 211-221
- Ishihama, A., M. Taketo, T. Saitoh and R. Fukuda. 1976. In RNA Polymerase, ed. R. Losick and M. Chamberlin, pp. 485-502. New York, Cold Spring Harbor Laboratory

- Ishihama, I. and T. Saitoh. 1979. J.Mol.Biol. 129, 517-530
- Iwakura, Y., A. Ishihama and T. Yura. 1973. Mol.gen.Genet. 121, 181-196
- Jackson, E.N. and C. Yanofsky. 1973. J.Mol.Biol. 76, 89-101
- Jacob, F. and J. Monod. 1961. J.Mol.Biol. 3, 318-356
- Kasai, T. 1974. Nature 249, 523-527
- Kassavetis, G.A. and M.J. Chamberlin. 1979. J.Virol. 29, 196-208
- Kassavetis, G.A., K.M. Kaya and M.J. Chamberlin. 1978. Biochem. 17, 5798-5804
- Kieffer, M., N. Neff and M.J. Chamberlin. 1977. J.Virol. 22, 548-552
- Koller, T., O. Kübler, R. Portman and J.M. Sogo. 1978. J.Mol.Biol. 120, 121-131
- Korn, L.J. and C. Yanofsky. 1976. J.Mol.Biol. 103, 395-409
- Krakow, J.S. and K. von der Helm. 1971. Cold Spring Harbor Symp.Quant.Biol. 35, 73-83
- Kramer, F.R. and D.R. Mills. 1978. Proc.Nat.Acad.Sci.USA 75, 5334-5338
- Kramer, R.A., M. Rosenberg and J.A. Steitz. 1974. J.Mol. Biol. 89, 767-776
- Küpper, H., T. Sekiya, M. Rosenberg, J. Egan and A. Landy. 1978. Nature 272, 423-428
- Lathe, R. 1978. Curr.Topics in Microbiol. and Immunol. 83, 37-91
- Lee, F. and C. Yanofsky. 1977. Proc.Nat.Acad.Sci.USA 74, 4365-4369
- Lee, F., C.L. Squires, C. Squires and C. Yanofsky. 1976. J.Mol.Biol. 103, 383-393
- Lee, F., K. Bertrand, G. Bennett and C. Yanofsky. 1978. J.Mol.Biol. 121, 193-217
- Lennox, E.S. 1955. Virology 1, 190-206
- Losick, R. and M. Chamberlin (eds.) 1976. RNA Polymerase, Cold Spring Harbor Laboratory, New York.
- Lowery-Goldhammer, C. and J.P. Richardson. 1974. Proc.Nat. Acad.Sci.USA 71, 2003-2007
- Lowery, C. and J.P. Richardson. 1977. J.Biol.Chem. 252, 1381-1385

- Ludwig, R.A. and W.C. Summers. 1975. *Virology* 68, 360-373
- Ludwig, R.A. and W.C. Summers. 1976. *Virology* 71, 278-290
- Maitra, U., A.H. Lockwood, J.S. Dubnoff and A. Guha. 1970. *Cold Spring Harbor Symp.Quant.Biol.* 35, 143-156
- Maizels, N.M. 1973. *Proc.Nat.Acad.Sci.USA* 70, 3585-3589
- Majors, J. 1975. *Proc.Nat.Acad.Sci.USA* 72, 4394-4398
- Maniatis, T., A. Jeffrey and H. van de Sande. 1975a. *Biochem.* 14, 3787-3794
- Maniatis, T., M. Ptashne, K. Backman, D. Kleid, S. Flashman, A. Jeffrey and R. Maurer. 1975b. *Cell* 5, 109-113
- Manor, H., D. Goodman and G.S. Stent. 1969. *J.Mol.Biol.* 39, 1-29
- Maxam, A. and W. Gilbert. 1977. *Proc.Nat.Acad.Sci.USA* 74, 560-564
- McAllister, W.T. and C.L. Barrett. 1977a. *Virol.* 82, 270-287
- McAllister, W.T. and C.L. Barrett. 1977b. *J.Virol.* 24, 709-711
- McAllister, W.T. and C.L. Barrett. 1977c. *J.Virol.* 23, 543-553
- McAllister, W.T. and R.J. McCarron. 1977. *Virol.* 82, 288-298
- McAllister, W.T. and H.-L. Wu. 1978. *Proc.Nat.Acad.Sci.USA* 75, 804-808
- McClure, W.R. and C.L. Cech. 1978. *J.Biol.Chem.* 253, 8949-8956
- McConnell, D. 1979a. *Nucleic Acids Res.* 6, 525-544
- McConnell, D. 1979b. Personal communication.
- McDonell, M.W., M.N. Simon and F.W. Studier. 1977. *J.Mol.Biol.* 110, 119-146
- McMahon, J.E. and I. Tinoco, Jr. 1978. *Nature* 271, 275-277
- Millette, R.L., C.D. Trotter, P. Herrlich and M. Schweiger. 1970. *Cold Spring Harbor Symp.Quant.Biol.* 35, 135-142
- Minkley, E. and D. Pribnow. 1973. *J.Mol.Biol.* 77, 255-277
- Miozzari, G.F. and C. Yanofsky. 1978a. *J.Bact.* 133, 1457-1466
- Miozzari, G.F. and C. Yanofsky. 1978b. *Nature* 276, 684-689

- Morse, D.E. and A.N.C. Morse. 1976. J.Mol.Biol. 103, 209-226
- Muller, W., H. Weber, F. Meyer and C. Weissman. 1978. J.Mol.Biol. 124, 343-358
- Murray, K., S.G. Hughes, J.S. Brown and S.A. Bruce. 1976. Biochem.J. 159, 317-322
- Musso, R.E., R. Dilauro, S. Adhya and B. de Crombrughe. 1977. Cell 12, 847-854
- Nathans, D. and H.O. Smith. 1975. Ann.Rev.Biochem. 44, 273-293
- Neff, N.F. and M.J. Chamberlin. 1978. J.Biol.Chem. 253, 2455-2460
- Newman, A.J., T.G. Linn and R.S. Hayward, 1979. Mol.gen. Genet. 169, 195-204
- Oakley, J.L. and J.E. Coleman. 1977. Proc.Nat.Acad.Sci. USA 74, 4266-4270
- Oakley, J.L., J.A. Pascale and J.E. Coleman. 1975. Biochem. 14, 4684-4691
- Oakley, J.L., R.E. Strothkamp, A.H. Sarris and J.E. Coleman. 1979. Biochem. 18, 528-537
- O'Hare, K. 1978. Ph.D. Thesis, University of Edinburgh
- Ohtsubo, H. and E. Ohtsubo. 1978. Proc.Nat.Acad.Sci.USA 75, 615-619
- Ovchinnikov, Y.A., V.M. Lipkin, N.N. Modyanov, O.Y. Chertov and Y.V. Smirnov. 1977. FEBS Lett. 76, 108-111
- Pachl, C.A. and E.T. Young, 1978. J.Mol.Biol. 122, 69-101
- Pannekoek, H., W.J. Brammar and P.H. Pouwels. 1975. Mol.gen.Genet. 136, 199-214
- Peters, G.G. 1974. Ph.D. Thesis, University of Edinburgh
- Peters, G.G. and R.S. Hayward. 1974a. Eur.J.Biochem. 48, 199-208
- Peters, G.G. and R.S. Hayward. 1974b. FEBS Proc. 33, 361
- Peters, G.G. and R.S. Hayward. 1974c. Biochem.Biophys.Res. Comm. 61, 809-816
- Pfennig-Yeh, M., H. Ponta, M. Hirsch-Kauffmann, H.-J. Rahmsdorf, P. Herrlich and M. Schweiger. 1978. Mol.gen.Genet. 166, 127-140
- Peiczenik, G., B.G. Barrell and M.L. Gefter. 1972. Arch.Bioch.Bioph. 152, 152-165

- Piecznik, G., P. Model and H.D. Robertson. 1974. J.Mol.Biol. 90, 191-214
- Platt, T., C. Squires and C. Yanofsky. 1976. J.Mol.Biol. 103, 411-420
- Ponta, H., H.-J. Rahmsdorf, S.H. Pai, M. Hirsch-Kauffmann, P. Herrlich and M. Schweiger. 1974. Mol.gen.Genet. 134, 281-297
- Pouwels, P.H. and H. Pannekoek. 1976. Mol.gen.Genet. 149, 255-265
- Pouwels, P.H. and J. van Rotterdam. 1975. Mol.gen.Genet. 136, 215-226
- Pribnow, D. 1975. Proc.Nat.Acad.Sci.USA 72, 784-788
- Ptashne, M., K. Backman, M.Z. Humayun, A. Jeffrey, R. Maurer, B. Meyer and R.T. Sauer. 1976. Science 194, 156-161
- Rabussay, T. and W. Zillig. 1969. FEBS Lett. 5, 104-107
- Richardson, J.P., C. Grimley and C. Lowery. 1975. Proc. Nat.Acad.Sci.USA 72, 1725-1728
- Roberts, J. 1969. Nature 224, 1168-1174
- Roberts, J.W. 1975. Proc.Nat.Acad.Sci.USA 72, 3300-3304
- Roberts, J.W. 1976. In RNA Polymerase, ed. R. Losick and M. Chamberlin, pp. 247-271. New York, Cold Spring Harbor Laboratory.
- Roberts, R.J. 1978. Gene 4, 183-193
- Roberts, T.M., H. Shimatake, C. Brady and M. Rosenberg. 1977. Nature 270, 274-275
- Rohrer, H. and W. Zillig. 1977. Eur.J.Biochem. 79, 401-409
- Rosa, M.D. 1979. Cell 16, 815-825
- Rosenberg, M., R.A. Kramer and J.A. Steitz. 1974. J.Mol.Biol. 89, 777-782
- Rosenberg, M., S. Weissman and B. De Crombrughe. 1975. J.Biol.Chem. 250, 4755-4764
- Rosenberg, M., B. De Crombrughe and R. Musso. 1976. Proc.Nat.Acad.Sci.USA 73, 717-721
- Rosenberg, M., D. Court, H. Shimatake, C. Brady and D. Wulff. 1978. Nature 272, 414-423
- Salstrom, J.S. and W. Szybalski. 1978. Virol. 88, 252-262
- Sanger, F. and A.R. Coulson. 1978. FEBS Lett. 87, 107-110

- Sanger, F., G.M. Air, B.G. Barrell, N.L. Brown, A.R. Coulson, J.C. Fiddes, G.A. Hutchison III, P.M. Slocombe and M. Smith. 1977a. *Nature* 265, 687-695
- Sanger, F., S. Nicklen and A.R. Coulson. 1977b. *Proc.Nat. Acad.Sci.USA* 74, 5463-5467
- Sanger, F., A.R. Coulson, T. Friedman, G.M. Air, B.G. Barrell, N.L. Brown, J.C. Fiddes, C.A. Hutchison III, P.M. Slocombe and M. Smith. 1978. *J.Mol.Biol.* 125, 225-246
- Schaller, H., C. Gray and K. Hermann. 1975. *Proc.Nat.Acad. Sci.USA* 72, 737-741
- Schmeissner, U., D. Ganem and J.H. Miller. 1977. *J.Mol.Biol.* 109, 303-326
- Segawa, T. and F. Imamoto. 1976. *Virology* 70, 181-184
- Sharp, P.A., B. Sugden and J. Sambrook. 1973. *Biochem.* 12, 3055-3063
- Shaw, J.E., B.B. Jones and M.L. Pearson. 1978. *Proc.Nat. Acad.Sci.USA* 75, 2225-2229
- Shine, J. and L. Dalgarno. 1974. *Proc.Nat.Acad.Sci.USA* 71, 1342-1346
- Shinnick, T.M., E. Lund, O. Smithies and F.R. Blattner. 1975. *Nucleic Acids Res.* 2, 1911-1929
- Shortle, D. and D. Nathans. 1978. *Proc.Nat.Acad.Sci.USA* 75, 2170-2174
- Siebenlist, U. 1979. *Nucleic Acids Res.* 6, 1895-1907
- Simon, M.N. and F.W. Studier. 1973. *J.Mol.Biol.* 79, 249-265
- Sklar, J., P. Yot and S.M. Weissman. 1975. *Proc.Nat.Acad. Sci.USA* 72, 1817-1821
- Smith, M., N.L. Brown, G.M. Air, B.G. Barrell, A.R. Coulson, C.A. Hutchison III and F. Sanger. 1977. *Nature* 265, 702-705
- Squires, C.L., F.D. Lee and C. Yanofsky. 1975. *J.Mol.Biol.* 92, 93-111
- Stahl, S.J. and M.J. Chamberlin. 1977. *J.Mol.Biol.* 112, 577-601
- Stanley, J. and S. Vassilenko. 1978. *Nature* 274 (5666), 87-89
- Stauffer, G.V., G. Zurawaki and C. Yanofsky. 1978. *Proc. Nat.Acad.Sci.USA* 75, 4833-4837
- Steitz, J.A. and R.A. Bryan. 1977. *J.Mol.Biol.* 114, 527-543

- Steitz, J.A. and K. Jakes. 1975. Proc.Nat.Acad.Sci.USA 72, 4734-4738
- Strome, S. and E.T. Young. 1978. J.Mol.Biol. 125, 75-93
- Studier, F.W. 1969. Virology 39, 562-574
- Studier, F.W. 1972. Science 176, 367-376
- Studier, F.W. 1975. J.Mol.Biol. 94, 283-295
- Starlinger, P. and H. Saedler. 1976. Curr.Topics in Microbiol. and Immunol. 75, 111-152
- Sugimoto, K., H. Sugisaki, T. Okamoto and M. Takanami. 1977. J.Mol.Biol. 111, 487-507
- Szybalski, W., H. Kubinski, A. Hradecna and W.C. Summers. 1971. In Methods in Enzymology, ed. L. Grossman and K. Moldave. Vol. XXI, 383-413. Academic Press, New York
- Tinoco, I., P.N. Borer, B. Dengler, M.D. Levine, O.C. Uhlenbeck, D.M. Crothers and J. Gralla. 1973. Nature New Biol. 246, 40-41
- Travers, A.A. and R.R. Burgess. 1969. Nature 222, 537-540
- Walz, A., V. Pirotta and K. Ineichen. 1976. Nature 262, 665-669
- Wells, R.D., R.W. Blakesley, S.C. Hardies, G.T. Horn, J.E. Larson, E. Selsing, J.F. Burd, H.W. Chan, J.B. Dodgson, K.F. Jensen, I.F. Nes and R.M. Wartell. 1977. C.R.C. Crit.Rev. in Biochem. 4, 305-340
- Wiggs, J.L., J.W. Bush and M.J. Chamberlin. 1979. Cell 16, 97-109
- Williams, R.C. 1977. Proc.Nat.Acad.Sci.USA 74, 2311-2315
- Wu, A.M. and T. Platt. 1978. Proc.Nat.Acad.Sci.USA 75, 5442-5446
- Wu, F.Y.-H., L.R. Yarbrough and C.-W. Wu. 1976. Biochem. 15, 3254-3258
- Yamada, Y. and D. Nakada. 1976. J.Virol. 18, 1155-1159
- Yamamoto, K.R., B.N. Alberts, R. Benzinger, L. Lawhorne and G. Treiber. 1970. Virology 40, 734-744
- Yanofsky, C. and L. Soll. 1977. J.Mol.Biol. 113, 663-677
- Zillig, W., H. Fujiki, W. Blum, D. Janekovic, M. Schweiger, H.-J. Rahmsdorf, H. Ponta and M. Hirsch-Kauffmann. 1975. Proc.Nat.Acad.Sci.USA 72, 2506-2510

Zurawski, G., D. Elseviers, G.V. Stauffer and C. Yanofsky.
1978a. Proc.Nat.Acad.Sci.USA 75, 5988-5992

Zurawski, G., K. Brown, D. Killingly and C. Yanofsky.
1978b. Proc.Nat.Acad.Sci.USA 75, 4271-4275

Addenda:

Robertson, H.D., E. Dickson and J.J. Dunn. 1977. Proc.Nat.
Acad.Sci.USA 74, 822-826

Salstrom, J.S. and W. Szybalski. 1976. Fed.Proc. 35, 1538

Smith, J.D. 1972. Ann.Rev.Genet. 5, 245-256

Delius, H., H. Westphal, and N. Axelrod. 1973. J. Mol. Biol. 74: 677.