A STUDY OF THE INTERACTION OF *E. COLI* RNA POLYMERASE AND BACTERIOPHAGE S13 DNA

by

Eric Rassart

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Department of Biochemistry
McGill University
Montreal

August ,1977

"Cela ne signifie nullement que la science et la pensée analytiques soient des outils vains et destructeurs, mais bien que ceux qui s'en servent doivent être plus grands que leurs instruments. Pour être un savant vraiment productif, il est nécessaire d'être plus qu'un homme de science, et un véritable philosophe est plus qu'un intellect. Car le mensuration analytique de la nature ne nous renseigne en rien sur elle si nous ne savons la considérer autrement."

"Amour et connaissance, Alan Watts"

ABSTRACT

The enzyme *E. coli* RNA-polymerase can form stable complexes with S13 and ϕX174 replicative form DNA's. The binding sites have been localized on the physical map of two restriction enzymes and by electron microscopy. On S13 RF DNA, there are three major binding sites and two weaker sites. The locations of the strong sites correspond with promoter sites at the beginning of gene A, gene B, and the region of gene D and the beginning of gene E. On ϕX174 RF DNA five sites, three classified as strong have been located. They are found at the same location as those in S13 except for one site at the beginning of gene B, which cannot be assigned at exactly the same location as in S13. Thus this area is interesting since the two phages are closely related. The S13 restriction fragment *Hin*d 8 was chosen for sequence analysis because it contains an RNA polymerase binding site and is at the beginning of gene B. T4 endonuclease IV and a new method: base specific DNA alkylation and hydrolysis were used to elucidate the sequence. The sequence was also confirmed by the rapid DNA sequencing technique of Gilbert and Maxam. The nucleotide sequence is compared to the sequence for the same region of ϕX174, and is discussed in terms of gene expression, promoter site and transcription event. The sequence also provides circumstantial evidence that the genes A and B overlap in S13.

SOMMAIRE

La RNA polymerase peut former des complexes stables avec le DNA des formes réplicatives des bacteriophages S13 et ϕX174. Les sites de couplage ont été localisés sur la carte physique produite par deux enzymes de restriction et aussi par microscopie électronique. Sur le DNA de S13 trois sites de couplage et deux beaucoup plus faibles ont été inventoriés. Les sites majeurs correspondent à des sites promoteurs et sont situés au début du gène A, du gène B et dans la région où les gènes D et E se chevauchent. Chez ϕX174, cinq sites dont trois majeurs ont été situés. On les retrouve aux mêmes positions que chez S13 exception faite d'un site, correspondant au début du gène B, qui n'a put être assigné au même emplacement que chez S13. Cette région attire particulièrement notre attention du fait que les deux phages sont étroitement reliés. Le fragment *Hin*d 8 de S13 a été choisi pour des études subséquentes de séquence, tout d'abord parce qu'il contient un site de couplage de la RNA polymerase et ensuite parce que ce fragment se retrouve dans la région du début du gène B. L'endonuclease IV induite par le bactériophage T4 ainsi qu'une nouvelle technique: l'alkylation spécifique d'une base du DNA suivie d'hydrolyse ont été utilisés pour élucider la séquence. Cette séquence a été confirmée par la rapide technique pour séquencer le DNA de Gilbert et Maxam. La séquence est comparée avec cette équivalente de ϕX174 et certains principes de base tels que, expression de gène, site promoteur et transcription seront discutés. La séquence fournit une évidence montrant que chez S13 les gènes A et B se chevauchent comme chez ϕX174.

ACKNOWLEDGEMENTS

I thank Dr. John H. Spencer for his hospitality and direction during this work in his laboratory.

I thank Drs. Klaus Harbers, Frank Grosveld, John Kaptein, Bruce Goodchild and Dean Hewish for their collaboration, friendship, criticism and suggestions.

I wish to thank Mr. Artin Kabassakalian and Mrs. Pat Douglas for technical assistance and Mrs. Evelyn Williams for the hundreds of cups of tea, her good humor and for teaching me some of the finest details of the English language.

I am very grateful to Mr. Ken Hastings and Dr. Linda Krantz for proof-reading of the introduction of this thesis.

I thank Mr. Max Zollinger and Mr. Leven Guluzian for the quality of the photographic work and Miss Odette Falardeau for typing this thesis.

I also thank the students, staff and other members of the Department of Biochemistry for their help and for providing me the necessary environment for this work.

This study was supported by the Medical Research Council of Canada for the period from 1974 to 1977 by a fund through Dr. John Spencer's grant.

TABLE OF CONTENT

	Page
DEDICATION	*ii*
ABSTRACT	*iii*
SOMMAIRE	*iv*
ACKNOWLEDGEMENT	*v*
TALBE OF CONTENTS	*vi*
LIST OF FIGURES	*xii*
LIST OF TABLES	*xvii*
LIST OF ABBREVIATIONS	*xix*
CONTRIBUTION TO ORIGINAL RESEARCH	*xx*
CHAPTER I.	
INTRODUCTION	
PREFACE	1
SECTION I.	
TRANSCRIPTION IN *E. COLI* AND COLIPHAGES	5
A. DNA-dependent RNA polymerase	6
1. Purification	6
2. Physical properties	7
3. Assembly of the subunits	8
B. Promoters and Transcription	9
1. Recognition	12
a) Interaction of RNA polymerase with the DNA template	12
i) Factors affecting binding to the template	12
ii) Measurement of template binding	16
b) The promoter concept	17
i) The "open" promoter	18
ii) Promoter site selection. -A model	19
-The sigma subunit	22

	Page
c) Mapping of RNA polymerase recognition sites	22
i) Visualization by electron microscopy	22
ii) Retention of complexes on membrane filters	23
2. Initiation	24
a) Specificity of initiation	24
b) Mapping of the RNA starting sequence sites	26
3. Elongation	28
4. Termination	28
a) Termination in the absence of *Rho* factor	29
b) Termination in the presence of *Rho* factor	29
5. A model for transcription	31
a) Positive control	33
b) Negative control i) Repressor	33
ii) Attenuator	34
SECTION II.	
DNA SEQUENCE ANALYSIS	35
A. Isolation of the starting material	35
1. Protection of specific DNA sequences	36
a) RNA polymerase	36
b) Repressors and CAP factors	37
c) Ribosomes	37
2. Nuclease digestion	37
a) Restriction enzymes	37
b) T4 induced endonuclease IV	38
	39

Page

B. Methods for DNA sequencing 39

1. Chemical degradation of DNA 39

a) Acid hydrolysis 39
b) Hydrazinolysis 40
c) DNA methylation 43
d) A rapid method for DNA sequencing 44

2. Enzymatic degradation of DNA 46

a) DNA sequencing by 3' exonuclease digestion 46
b) DNA sequencing by 5' exonuclease digestion 49

3. Other methods - primers extension 51

a) Ribosubstitution 52
b) The plus and minus method 52
i) The minus system 53
ii) The plus system 54

4. Purpose of this study 56

5. Experimental approach 57

CHAPTER II.

LOCALIZATION OF *ESCHERICHIA COLI* RNA POLYMERASE BINDING SITES ON S13 AND ϕX174 DNA: ALIGNMENT WITH RESTRICTION ENZYME MAPS

ABSTRACT 59

INTRODUCTION 60

MATERIALS AND METHODS 63

Isolation of bacteriophage S13 and ϕX174 replicative form I DNA (RF I). 63

Restriction endonuclease and restriction enzyme digestions 63

E. coli RNA polymerase 64

	Page
Carrier nucleic acids	64
Isolation of *E. coli* RNA polymerase sites from S13 RF I DNA	64
Binding of *E. coli* RNA polymerase to restriction fragments of S13 RF I DNA	65
Method A	65
Method B	66
Quantitation of *E. coli* RNA polymerase binding to S13 RF DNA restriction fragments	66
RESULTS	67
DISCUSSION	81
ACKNOWLEDGMENTS	87

CHAPTER III.

LOCALIZATION OF *ESCHERICHIA COLI* RNA POLYMERASE BINDING SITES ON S13 AND ϕX174 DNAs BY ELECTRON MICROSCOPY

ABSTRACT	88
INTRODUCTION	89
MATERIALS AND METHODS	91
Preparation of specimens for electron microscopy	91
EM measurements and calculations of results	92
RESULTS	93
DISCUSSION	98
ACKNOWLEDGEMENTS	102

Page

CHAPTER IV.

THE NUCLEOTIDE SEQUENCE OF THE B GENE PROMOTER REGION OF BACTERIOPHAGE S13

ABSTRACT 103

INTRODUCTION 104

MATERIALS AND METHODS 107

Chemicals 107

Enzymes 107

Purification of ^{32}P-labeled bacteriophage S13 RF DNA and isolation of restriction fragments 108

T4 endonuclease IV hydrolysis 109

DNA alkylation and hydrolysis 109

5' ^{32}P-labeling of alkylation and T4 endonuclease IV generated oligonucleotides 109

Fractionation of oligonucleotides by ionophoresis-homochromatography 110

Partial digestion of oligonucleotides by snake venom phosphodiesterase 110

Sequence analysis by the alkylation-hydrazinolysis - rapid gel technique 111

Isolation of an RNA polymerase binding site 111

RESULTS 112

DISCUSSION 136

CHAPTER V

DISCUSSION

1. The RNA polymerase binding sites of S13 and ϕX174 142

a) Visualization by electron microscopy 142

b) Localization using the restriction enzymes 142

Page

2. Is S13 *Hin*d 8 an RNA polymerase binding site? 147

3. Sequence analysis by T4 endonuclease IV 148

4. Sequence analysis by specific DNA alkylation and hydrolysis 149

5. The sequence of S13 *Hin*d 8 150

6. Is S13 *Hin*d 8 a promoter site 154

REFERENCES 160

APPENDIX I

ADDITIONAL RESULTS TO CHAPTER II 177

APPENDIX II

AUTORADIOGRAMS OF T4 ENDONUCLEASE IV DIGESTS AND METHYL-METHANE SULFONATE DIGESTS, OLIGONUCLEOTIDE SEQUENCE ANALYSIS OF FRAGMENT *HIN*D 8 184

LIST OF FIGURES

Page

CHAPTER I.

Section I

Figure 1. The overall process of Transcription. 10

Figure 2. Representation of Promoter Site recognition by RNA polymerase and the initiation of RNA synthesis. 14

Section II

Figure 3. Methylation mechanism for deoxyguanine in DNA and β-elimination mechanism from Gilbert and Maxam (1975). 41

Figure 4. Hydrazinolysis of deoxycytosine and cleavage of DNA in presence of piperidine. 42

Figure 5. DNA methylation and hydrazinolysis from Gilbert and Maxam (1975). 45

Figure 6. Schematic representation of sequencing DNA using site-specific restriction endonuclease and the single-stranded specific T4 endonuclease IV. 47

Figure 7. Schematic representation of sequencing a DNA fragment using specific alkylation and cleavage with methyl methane sulfonate. 48

Figure 8. Sequence chromatogram of a 5' ^{32}P-labeled oligonucleotide (Wandering spot technique). 50

Figure 9. The plus and minus systems as developed by Sanger and Coulson (1975). 55

Page

CHAPTER II.

Figure 1. Effect of KCl concentration on the binding of *E. coli* RNA polymerase to S13 RF DNA. 68

Figure 2. The effect of the molar ratio of *E. coli* RNA polymerase to S13 RF DNA on the extent of binding of the polymerase to S13 RF DNA. 69

Figure 3. Autoradiograph of *Hin*d restriction fragments of ^{32}P-labeled S13 RF DNA which bind *E. coli* RNA polymerase, separated on a polyacrylamide slab gel. 71

Figure 4. Autoradiograph of *Hae*III restriction fragments of ^{32}P-labeled S13 RF DNA which bind *E. coli* RNA polymerase, separated on a polyacrylamide slab gel. 76

Figure 5. Schematic presentation of the *E. coli* RNA polymerase binding sites of S13 and ϕX174 RF I DNAs in relation to the *Hin*d and *Hae*III restriction endonuclease cleavage maps and the genetic map. 80

CHAPTER III.

Figure 1. Electron micrographs of *E. coli* RNA polymerase molecules bound to S13 RF III DNA and ϕX174 RF III DNA. 94

		Page
Figure 2.	Histograms of the distribution of *E. coli* RNA polymerase molecules on S13 RF III and ϕX174 RF III DNA.	96
Figure 3.	Schematic presentation of the *E. coli* RNA polymerase binding sites on S13 and ϕX174 RF DNAs in relation to the *Hin*d and *Hae*III restriction endonuclease cleavage maps and the genetic map.	97
CHAPTER IV.		
Figure 1.	Schematic representation of the location of the *Hin*dII/*Hae*III 3-75 and "*Hin*d 8" fragments.	113
Figure 2.	Fingerprints of the oligonucleotides released by T4 endonuclease IV (experiment F) or by DNA alkylation and hydrolysis with methylmethane sulfonate (experiment A) of *Hin*dII 8 restriction fragments.	114-115
Figure 3.	Sequence chromatograms of T4 endonuclease IV oligonucleotides (D13 and F9) and oligonucleotides generated by methylmethane sulfonate (A15 and A34) of fragment *Hin*d 8.	116-117
Figure 4.	The sequence of *Hin*dII-*Hae*III 3-75 and "*Hin*d 8" restriction fragments.	121
Figure 5.	Partial digestion of *Hin*d 8 with *Alu*I.	123

		Page
Figure 6.	Sequence chromatograms of oligonucleotides generated by methylmethane sulfonate of fragment *Hin*d 8 labeled only at its 5' ends.	125-126
Figure 7.	Sequence autoradiography of the Gilbert-Maxam alkylation and hydrazinolysis - rapid gel method on the 75 base pairs fragment overlap of *Hin*dII-7 and *Hae*III-3.	127-128
Figure 8.	Sequence autoradiography of the Gilbert-Maxam alkylation and hydrazinolysis - rapid gel method on *Hin*d 8.	129-130
Figure 9.	*Hin*d 8 sequence representing the T4 endonuclease IV oligonucleotides obtained after RNA polymerase binding on *Hin*d 8 and DNAse digestion.	135
Figure 10.	Sequence of *Hin*dII/*Hae*III 3-75 and *Hin*d 8 with the corresponding amino acid sequence of the A protein (at the top position).	139

APPENDIX I

Figure 1.	7.5% SDS-polyacrylamide gel electrophoresis of 10, 25 and 50 µg of our RNA polymerase preparation.	178
Figure 2.	Immunological assay of our RNA polymerase preparation against anti 66,000 daltons S_1 protein (center).	178

		Page
Figure 3.	Effect of the albumin concentration to the extent of binding of the RNA polymerase to S13 RF DNA.	179
Figure 4.	Polyacrylamide gel electrophoresis of the RNA polymerase protected fragments.	180
Figure 5.	Pyrimidine oligonucleotides of A) bacteriophage S13 RF I DNA and B) the RNA polymerase protected fragments isolated from S13 RF 1 DNA.	181
Figure 6.	Polyacrylamide gel electrophoresis of *Hae*III (a) or *Hin*d (b) digested ^{3}H-labeled S13 RF I DNA before binding to the RNA polymerase and ^{32}P-labeled S13 RF I DNA after binding to the RNA polymerase.	182

LIST OF TABLES

		Page
CHAPTER I.		
Section I		
Table I.	Sequences of promoter regions.	21
CHAPTER II.		
Table I.	Binding of *E. coli* RNA polymerase of the *Hin*d restriction fragments of S13 RF I DNA.	74
Table II.	Binding of *E. coli* RNA polymerase to the *Hae*III restriction fragments of S13 RF I DNA.	75
Table III.	Binding of *E. coli* RNA polymerase to the *Hin*d II restriction fragments of ϕX174 RF I DNA.	78
Table IV.	Binding of *E. coli* RNA polymerase to the *Hae*III restriction fragments of ϕX174 RF I DNA.	79
CHAPTER IV.		
Table I.	Sequences of the oligonucleotides released by T4 endonuclease IV from fragment *Hin*d 8	118
Table II.	Sequences of the oligonucleotides released by methylmethane sulfonate alkylation and alkaline hydrolysis of fragment *Hin*d 8	119
Table III.	Sequences determined by the Maxam-Gilbert Alkylation-Hydrazinolysis-Rapid Gel.	131

		Page
Table IV.	Sequences of the oligonucleotides released by T4 endonuclease IV from that part of *Hind* 8 protected by *E. coli* RNA polymerase from DNase A digestion.	134
CHAPTER V.		
Table I.	Sequences of promoter regions.	156
APPENDIX I.		
Table I.	Binding of *E. coli* RNA Polymerase to the *Hind* and *Hae*III Restriction Fragments of S13 RF I DNA .	183

LIST OF ABBREVATIONS

Unless otherwise stated, the symbols used are the recommended SI (Système International)

A	deoxyadenosine
C	deoxycytidine
G	deoxyguanosine
T	deoxythymidine
U	uracil
N	unspecified nucleotide
R	unspecified purine nucleotide
Pu	Purine
Py	Pyrimidine
r (prefix)	ribo (eg. rN ribonucleotide)
d (prefix)	deoxy (eg. dN deoxynucleotide)

The prefix d is not used in nucleotide sequence presentations

lac	lactose
trp	tryptophan
his	histidine
ara	arabinose
gal	galactose
λ	lambda (bacteriophage)
σ	sigma protein subunit of *E. coli* RNA polymerase
SDS	Sodium Dodecyl sulfate
S	sedimentation coefficent
mRNA	messenger Ribonucleic Acid
RF	Replicative Form DNA
RFI	closed circular RF DNA

CONTRIBUTIONS TO ORIGINAL RESEARCH

1. It was shown that the *E. coli* RNA polymerase can form stable complexes with S13 RF DNA and with double-stranded fragments produced by restriction enzymes.

2. The size of the *E. coli* RNA polymerase protected fragments obtained from S13 RF DNA was shown to be between 35 and 40 base pairs.

3. The *E. coli* RNA polymerase binding sites have been localized on the S13 and ϕX174 molecules. The localization was done by measuring the binding capacity of the RNA polymerase on the different restriction fragments.

4. The Kleinschmidt's cytochrome C monolayer technique can be used successfully to visualize RNA polymerase — DNA complexes.

5. It was shown by electron microscopy that the *E. coli* RNA polymerase could form stable complexes with S13 and ϕX174 linear RF molecules and the complexes were mapped.

6. Three strong promoter sites have been found in S13 at the beginning of gene A, gene B and in the region of gene D and E. These results are in good agreement with genetic data.

7. The comparative study of the location of the RNA polymerase binding sites on S13 and ϕX174 provided additional evidence that the two phages are closely related.

8. It was shown that the T4-induced endonuclease IV can be used successfully on a denatured double-stranded DNA fragment in order to generate a specific set of deoxyoligonucleotides.

9. A site-specific cleavage method using the alkylating agent, methyl methane sulfonate, was established for degrading DNA fragments at G residues. This method was used in the sequence analysis of restriction fragment S13 *Hin*d 8.

10. The sequence of restriction fragment S13 *Hin*d 8 provided evidence suggesting that the genes A and B overlap in S13 as discovered in ϕX174.

CHAPTER I

INTRODUCTION

PREFACE

The discovery of nucleic acids was the result of the work of Friedrich Miescher in 1871. Very soon afterward, it was found that there were two kinds of nucleic acid: yeast nucleic acid (ribonucleic acid) and thymus nucleic acid (deoxyribonucleic acid). That chromosomes of higher organisms are composed of both nucleic acid and protein has been known almost since their discovery. Many early theories of heredity proposed that the genetic information must reside in the protein component of chromosomes because nucleic acids were thought to have too little structural variation to be able to specify or code for the enormous number of different proteins. The biological function of DNA remained obscure until 1944 when Avery *et al.* published their classical experiment on transformation. Direct evidence that DNA is the genetic material was derived from studies on micro-organisms: it's identification as the bacterial transforming factor established it as the probable carrier of hereditary information.

In 1953, Watson and Crick postulated a double-stranded helical structure for DNA. Their hypothesis states that each base in the template DNA specifies a different, so-called "complementary" base rather than a replica of itself. Subsequent work confirmed their postulated model and very few modifications have been required. Brachet in 1957, using very elegant histochemical studies, discovered that RNA was in some way involved in protein synthesis. A correlation was shown between the amount of RNA present in a cell and the rate of protein synthesis. From observations such as these, the idea developed that RNA provides a link between gene and protein. More specifically, it was suggested that information

encoded in the base sequence of DNA flows to RNA and then to amino-acid sequences in proteins. The first clear evidence that RNA has the capacity to carry information about protein structure came from work on tobacco mosaic virus by Gierer and Schramm in 1956. A year later, Ingram presented direct evidence that genes in DNA code for protein. In 1961, Hall and Spiegelman showed that the RNA synthesized after infection of *E. coli* by T2 virus could form hybrid molecules by specific association with the DNA of the virus. The RNA could not hybridize to other sources of DNA. Thus, the RNA seemed to serve as a "messenger" between the DNA confined to the nucleus in eukaryotic systems and the protein synthesizing system in the cytoplasm. The term "messenger" proposed by Jacob and Monod in 1961 has become the accepted name for the RNA which codes for protein.

Studies on the synthesis of the macromolecules DNA, RNA and protein and their role in the genetic pathway have revealed two levels of specificity. DNA, RNA and protein are each composed of specific sequences of monomer units; deoxyribonucleotides, ribonucleotides and amino acids respectively. The positioning of these units in the polymer chain during replication and translation is under the control of the particular DNA or RNA template involved. The mechanism by which the template dictates synthesis of the appropriate sequence involves both base pairing interactions and specific protein - nucleic acid interactions. The second level of specificity present in all three synthetic pathways is that each kind of macromolecular chain is initiated at a specific site or sites on the corresponding template. In addition, the progress of chain growth during transcription and translation is subject to termination at specific sites

on the template which define the units of transcription and translation.

The amount of RNA in a cell at any given moment is not directly related to the amount of genetic information in the DNA of the cell. The amount of prokaryotic DNA expressed in a cell when the cell is grown under optimal conditions is totally different than that expressed when the cell is grown in a poor medium. For example, there are genes that code for enzymes involved in synthesis of tryptophan, but these enzymes won't be expressed and synthesized if tryptophan is present in the medium. Thus, the cell has some mechanisms that allow some genes to be transcribed and not others. This implies that there are specific sites on the DNA where RNA polymerase can initiate and terminate the synthesis of an RNA molecule, and that these sites themselves, or the RNA polymerase are under cellular control. Thus, some transcriptional units can be very active while others are inactive. Each transcriptional unit could then be regulated by elements specific to that unit. In 1961, Jacob and Monod postulated that some transcriptional units, for example the lactose system, could be regulated by both positive and negative control elements. Considerable evidence has been accumulated to support the hypothesis of regulation of transcription, and some features will be discussed in the introduction to the thesis.

One of the first steps in understanding regulation of any process is to characterize structurally the components under regulation.

The first nucleic acid molecules sequenced were RNA molecules, for two major reasons. First, RNA molecules are of smaller size than DNA molecules. For example, tRNA, which is the smallest RNA, has an average size of only 80 nucleotides. Second, base specific nucleases were

available to degrade the RNA into smaller fragments. However, a limitation of RNA sequencing is that some DNA regions are not transcribed into RNA.

DNA sequencing has made enormous progress since 1970 due to the discovery of specific DNA nucleases such as T4 induced endonuclease IV and restriction endonucleases which degrade DNA into fragments of reasonable size. The development of powerful chromatographic techniques has also been important in the progress of DNA sequencing. In 1972, the DNA sequences of some pyrimidine oligonucleotides and the ends of some bacteriophage DNAs were reported. Five years later in 1977, Sanger's group reported the total sequence of the DNA from bacteriophage ϕX174, a DNA molecule which is over 5,000 nucleotides long.

SECTION I.

TRANSCRIPTION IN *ESCHERICIA COLI* AND COLIPHAGES

The genetic information encoded by the DNA molecule is transcribed into RNA by a DNA-dependent RNA polymerase. The enzyme catalyses the initiation, elongation and termination of the polyribonucleotide chain using ribonucleoside triphosphates as substrates (Weiss and Gladstone, 1959). The reaction requires a divalent cation (Mg^{++}) and the DNA molecule serves as the template. The reaction consists of a series of steps in which the enzyme first "recognizes" the initiation sites on the DNA template, then initiates and synthesizes RNA via the extension reaction, terminates RNA synthesis at the correct termination site and finally is released from the template. The newly synthesized RNA chain is complementary in base sequence to one strand of the DNA template. The bacterial and bacteriophage enzymes seem to recognize several genetic signals on the bacterial or bacteriophage chromosome. These signals are transformed into biochemical events involved in the processes of DNA site selection, RNA chain initiation, elongation and termination and release of the enzyme. Many regulatory processes in bacterial systems have been shown to be related to the transcriptional step. Regulation takes place either at the level of direct interaction between the RNA polymerase and a nucleotide sequence which may induce a particular conformational structure on the DNA template, or through the interaction of the enzyme or DNA template with factors that mediate regulation of transcription for example initiation or termination factors. There is increasing evidence that eukaryotic systems utilize a similar type of regulation of gene expression at the

transcriptional level.

A DNA-dependent RNA polymerase.

The *Escherichia coli* RNA polymerase enzyme is a large molecule of 400,000 to 500,000 daltons having a complex subunit structure. The enzyme has four major subunits; β', β, α and σ and a minor component ω (Burgess, 1969). There are two different active forms of this enzyme: the RNA polymerase holoenzyme and the core polymerase. The latter form does not have the σ subunit and the enzyme function is altered with respect to its capacity to recognize specific sites on the DNA template and to initiate synthesis of RNA chains. Many bacterial RNA polymerase enzymes are of this type.

The eukaryotic RNA polymerases I, II and III are also large molecules containing several polypeptide subunits. Multiple forms of RNA polymerase have been found in all eukaryotic cells including plant, insect, fungi and yeast (Chambon, 1975, Roeder, 1976).

I. Purification.

RNA polymerases, because of the complexity of their structure are difficult enzymes to purify, no universal purification method has been adopted. There are several purification procedures available for *E. coli* RNA polymerase but it is not certain that each method yields the same enzyme. Modification of one of the subunits might occur in one preparation and not in another. For example, the amount of sigma subunit associated with the enzyme varies considerably from preparation to preparation because the sigma subunit is not as tightly bound to the molecule

as the other subunits. Also there may well be unidentified components of the enzyme that are very weakly bound to the molecule and may thus be lost in the preparation. These components are not necessary for the basic steps of chain initiation and elongation since the isolated β, β', α and σ subunits can be reconstituted into an active enzyme (Sethi, 1970, Fukuda and Ishihama, 1974). However, such components could play a crucial role in site selection, termination or in regulatory mechanisms.

These considerations suggest that the properties of bacterial RNA polymerase may differ in different laboratories and that results may show a certain discrepancy due to the method of purification. The purification method of RNA polymerase in the present study is that described by Burgess and Travers (1971). The enzyme is 95% pure and contains β', β, α and σ subunits in a molar ratio of 1, 1, 2, 2.

2. Physical properties.

The holoenzyme form of *E. coli* RNA polymerase has the structure $\beta'\beta\alpha_2\sigma$ (Berg and Chamberlin, 1970). Molecular weight estimates have been made for each of the subunits by polyacrylamide gel electrophoresis in the presence of SDS. Current estimates are; β'= 160,000, β = 150,000, σ = 86,000 and α = 40,000 daltons (Burgess, 1969b, Zillig *et al.*, 1970, Berg *et al.*, 1971).

There is good evidence that β', β, α and σ are functionally important subunits of the RNA polymerase. The small component ω is missing in some preparations but is apparently not required for enzymatic activity. Mutants of *Escherichia coli* resistant to rifampicin have an RNA polymerase with an altered β subunit (Rabushay and Zillig, 1970). Temperature

sensitive mutants of *E. coli* are altered in the β' subunit (Panny *et al.*, 1974). Mutants of *E. coli* which block transcription of phage P2 show an altered α subunit (Zillig *et al.*, 1976). No mutants are known for σ but σ is essential for site recognition and initiation on many bacteriophage DNAs (Hinkle and Chamberlin, 1972a).

3. Assembly of the subunits.

E. coli RNA polymerase can be dissociated into its component subunits and reconstituted without loss of activity (Sethi, 1970). Dissociation can be achieved by treatment with detergents, organic solvents, urea, reducing agents or high salt conditions (Lill and Hartmann, 1970). None of the partially dissociated components show RNA polymerase activity but β and β' can still bind to DNA. Under carefully defined conditions, RNA polymerase activity can be restored after removal of the dissociating agent (Lill and Hartmann, 1970). The purified subunits, when mixed under appropriate conditions, regain enzymatic activity (Ishihama and Ito, 1973). It has been shown that all four different subunits must be present to get total reconstitution of the enzymatic activity. Ishihama and collaborators (1972, 1973, 1975) have carried out a step wise *in vitro* assembly of the enzyme according to the general reaction:

$$2\,\alpha \longrightarrow \alpha_2$$
$$\alpha_2 + \beta \longrightarrow \alpha_2\beta$$
$$\alpha_2\beta + \beta' \longrightarrow \alpha_2\beta\beta' \quad \text{(core enzyme)}$$
$$\alpha_2\beta\beta' + \sigma \longrightarrow \alpha_2\beta\beta'\sigma$$

The rate limiting step in the reconstitution of active polymerase is the temperature - dependent activation of the $\alpha_2\beta\beta'$ complex (core enzyme).

One of the interesting aspects of the enzyme maturation reaction lies in the great enhancement when σ subunit is added (Ishihama *et al.*, 1973). The reconstitution of RNA polymerase from separated subunits has provided a powerful tool for a better understanding of subunit function.

B. Promoters and Transcription.

The crucial determinant of the quality of the DNA transcript is clearly the accuracy of initiation. First, the RNA polymerase must recognize and bind to a DNA site before a stable complex is formed. Then, synthesis of an RNA molecule is initiated. Thus formally, we can distinguish two DNA structures: the initial binding site, and the site at which the RNA molecule is initiated. The promoter includes both the initial binding site and the startpoint but for the purpose of this study the two structures will be distinguished. DNA dependent RNA synthesis can be divided into a number of discrete steps. Each step is quite complex and has been the subject of intensive study.

For the selective transcription by *E. coli* RNA polymerase holoenzyme of transcriptional units, the commonly accepted steps are:

(1) Template binding. This involves the specific recognition of a site on the DNA by the holoenzyme, and the binding of RNA polymerase to the DNA template. The binding is followed by a conformational change and a purine ribonucleoside triphosphate is now introduced to the complex. (Maitra and Hurwitz, 1965, Travers *et al.*, 1973).

(2) RNA chain initiation: the enzyme initiation complex catalyzes formation of a phospho-diester link between rATP or rGTP and a second ribonucleoside triphosphate. There is release of inorganic pyrophosphate and

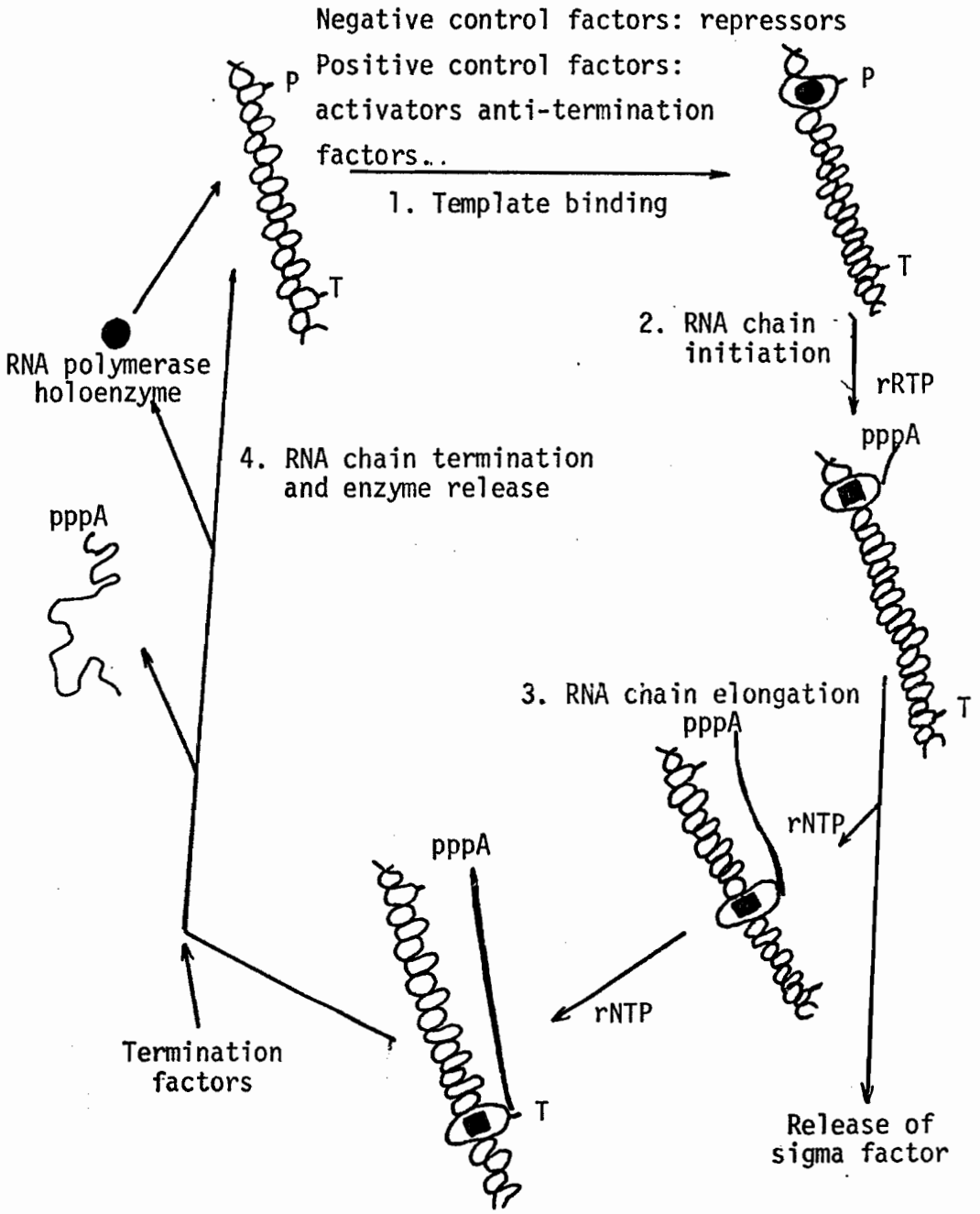


Figure 1. The overall process of Transcription

formation of a dinucleoside tetraphosphate of general formula ppprRprN.

(3) RNA chain elongation: successive ribonucleoside monophosphate residues are added to the 3' OH end of the nascent RNA chain from ribonucleoside triphosphate substrates. It is also during this step that the sigma factor is released from the complex (Wu *et al.*, 1975).

(4) RNA chain termination: the completed RNA chain and the RNA polymerase are both released from the DNA template and from each other. Two types of termination can occur during *in vitro* transcription 1) direct termination of growing RNA chain by interaction of RNA polymerase with sites on DNA and 2) *rho*-dependent termination of RNA chains at sites on DNA that require the presence of *rho* factor (Roberts, 1976).

The overall process of transcription is diagrammatically presented in Figure 1.

It is generally accepted that the fourth step (termination) rarely occurs at specific sites *in vitro* and the polymerase reaction consists of only the first three steps. In studies of transcription, because of the complexity of the complete RNA polymerase reaction, not much insight is gained on the regulatory mechanisms by the simple measurement of RNA synthesis. Even in the *in vitro* situation, the complexity of the overall reaction makes it difficult to determine the effect of various factors on RNA synthesis by measuring the rate of synthesis. To study the RNA polymerase reaction mechanism, it is necessary to follow the rate of each step and measure the sizes and properties of all of the transcriptional units on the DNA template as well. When very complex DNA templates are introduced into the system, it becomes even harder to elucidate the mechanisms involved at each step of the reaction.

1. Recognition:

a) Interaction of RNA polymerase with the DNA template.

The *in vivo* synthesis of an RNA molecule from a DNA template requires that the RNA molecule be initiated at a specific site on the DNA template. The precision of initiation depends almost uniquely upon the recognition of the specific site by the RNA polymerase followed by the binding of the enzyme to that specific sequence or DNA conformation. When this is accomplished a stable initiation complex is formed and RNA synthesis is initiated. Thus, we can distinguish two DNA structures contributing to the initiation process: the RNA polymerase recognition site and the site at which RNA synthesis is initiated (for review see (Chamberlin, 1976)).

It is known that the RNA polymerase core enzyme and holoenzyme have a general affinity for DNA or any polyanion, without any apparent specificity in site selection (Von Hippel *et al.*, 1974). The presence of the sigma subunit enables the RNA polymerase to bind at specific sites on the DNA with a greater affinity and more tightly (in terms of ease of removal), than non specific interactions. The next step in the interaction is a change in the conformational state of the DNA at the promoter site (Travers *et al.*, 1973). The two strands of the DNA at the promoter site unwind and separate and this enables the RNA polymerase to selectively transcribe one strand aligning the incoming substrate rNTPs according to the Chargaff rule of base pairing (Travers, 1974).

i) Factors affecting binding to the template

When RNA polymerase binds to the DNA, the two strands must unwind and separate producing a conformational change in the DNA molecule

(Travers, 1974, Mangel and Chamberlin, 1974). The transition between the conformational states is temperature dependent suggesting that the process has a positive enthalpy. The transition temperature thus would be a measurement of the activation energy for separating the two strands. Below that transition temperature, the promoter site is said to be closed while above that temperature it is said to be open (Chamberlin, 1974, Niyogi and Underwood, 1975, Dausse *et al.*, 1976). More direct evidence of the nature of that conformational change in the promoter region has been provided by Saucier and Wang (1972) who showed that the binding of RNA polymerase to superhelical DNA at 37°C results in an unwinding of the helix which doesn't occur if the binding is done at 0°C. They also showed that such unwinding would be consistent with either a 4-8 base pair local melting of the double helix or a conversion of DNA from the B form to the A form over a region of about 30-60 base pairs.

The first possibility is supported by the observation that the 3' end of a growing RNA chain is hybridized to the DNA strand, (Hayashi, 1965) suggesting that melting occurs over a small region of DNA.

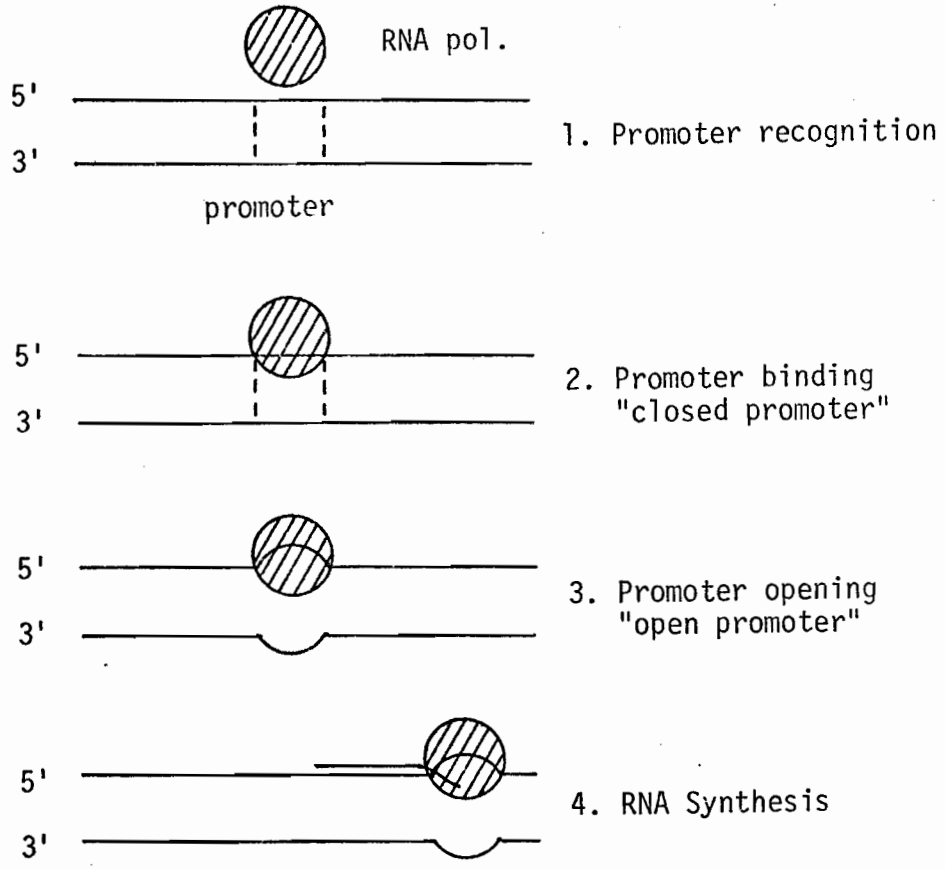


Figure 2. Representation of Promoter site recognition by RNA polymerase and the initiation of RNA synthesis. The positioning of the RNA polymerase and the promoter site is schematic and does not take into account the molecular dimensions of these entities (from Travers, 1974).

Another factor affecting the template binding of RNA polymerase and the initiation of transcription is the ionic strength. KCl has been shown to influence the different steps in transcription in different ways:

1) KCl acts by preferentially stimulating σ dependent initiation (Hoffman and Niyogi, 1973). At normal substrate levels, KCl stimulates the

RNA polymerase reaction with T4 DNA largely by promoting σ activity even when the σ concentration is very low implying that the sigma dependent initiations are promoted by KCl. At low substrate levels, KCl inhibits transcription but this may be at the level of chain initiation (Maitra *et al.*, 1970).

2) KCl prevents non specific binding of core polymerase to T4 DNA (Maitra *et al.*, 1971, Mueller, 1971).

3) KCl increases the specificity of RNA initiation and transcription. Bremer (1970) reported that in 0.2 M KCl RNA polymerase transcribes T4 DNA at a rate of 36 nucleotides/sec and T7 DNA 2.3 times faster. The chains were terminated at two preferred lengths and they were initiated at a small number of sites. At low ionic strength, the RNA chain growth rate was slower, termination was more random and the chains were initiated at additional sites on the DNA. Iida and Matsukage (1974) found that in 0.2 M KCl, initiation of RNA molecules on T7 DNA results in almost unique starting sequences but at low salt concentration other starting sequences were observed. Matsukage (1972) also observed that the specificity of *in vitro* transcription of T7 DNA was remarkably influenced by the KCl concentration. In 0.2 M KCl, the early region of the r-strand of T7 DNA is transcribed. At a low KCl concentration not only the r-strand of the early region, but also the late region and the l-strand were transcribed.

What is the effect of KCl on the template binding of RNA polymerase and the stability of the initiation complex? It is thought that KCl stabilizes the DNA like any other salt and this action results in a decrease in the efficiency of local strand separation of the promoter site. Thus the transition temperature, at which the two strands of the site can separate, is higher. The end effect is that fewer complexes are formed

on the DNA, and the specificity of binding is increased, since the KCl prevents binding to those unspecific or weak binding sites which have a higher transition temperature than the more specific ones.

This property has been used for the characterization of specific RNA polymerase binding sites which are stable even in 0.5 M KCl (Jones and Berg, 1966, Okamoto *et al.*, 1972).

ii) Measurement of template binding

Complex formation between the RNA polymerase and a DNA molecule has been demonstrated by different techniques.

1) Visualization of the complex in the electron microscope (Richardson, 1966, Bordier and Dubochet, 1974, Koller *et al.*, 1974).

2) Sedimentation of the RNA polymerase - DNA complex on gradients in the ultra-centrifuge (Richardson, 1966, Matsukage *et al.*, 1969).

3) Retention of RNA polymerase - DNA complexes on nitrocellulose membrane filters (Jones and Berg, 1966, Hinkle and Chamberlin, 1972).

4) Alteration of fluorescent properties during complex formation (Sarocchi and Darlix, 1974, Bahr *et al.*, 1976).

5) Under appropriate conditions, the number of RNA chains initiated can be correlated with the number of bound RNA polymerase molecules (Chamberlin, 1976).

6) RNA polymerase, when bound to the DNA, protects a certain region of DNA from nuclease digestion. These regions can be isolated (Le Talaer and Jeanteur, 1971, Heyden *et al.*, 1972, Okamoto *et al.*, 1972) and sequenced (Schaller *et al.*, 1975, Pribnow, 1975).

The very stable complexes between RNA polymerase and DNA define a limited

number of binding sites corresponding to initiation sites. The sequence of the protected DNA fragments from these complexes presumably contains those bases that are responsible for site recognition and complex formation. These two last techniques are more indirect techniques to measure complex formation when compared with the preceeding ones. Some of these will be explained later in more detail.

b) The Promoter concept.

In principle, the binding site for RNA polymerase can be distinguished from the promoter site. A site to which RNA polymerase binds is not necessarly a site at which RNA synthesis is initiated.

On duplex DNA, three kinds of RNA polymerase-DNA complexes have been characterized.

1) Non specific, weak, and fleeting complexes. RNA polymerase is known to bind well to ends (Vogt, 1969) and nicks in DNA (Hinkle *et al.*, 1972). The core enzyme in general exhibits non specific binding and as previously described lacks the ability to locate promoter sites on normal DNA templates (Burgess *et al.*, 1969). However, it is still able to catalize RNA synthesis. These sites are called class B sites and they have a half life of about a second (Chamberlin, 1972).

2) Sequence-specific A sites or closed promoters (Travers, 1974). These are promoters which are not used for RNA synthesis *in vivo*.

3) Activated or "open" form of the class A sites. These sites are characterized by high affinity for RNA polymerase and long half lives of about 30 to 60 hours (Hinkle and Chamberlin, 1971). At these sites, RNA polymerase can initiate RNA synthesis quickly. This open state is favored at

high temperature (Zillig *et al.*, 1971, Mangel and Chamberlin, 1974b) and at low ionic strength (Travers, 1974, Mangel and Chamberlin, 1974a).

i) The "open" promoter

The structure and properties of open promoter complexes have been studied extensively with DNAs from Bacteriophages T2 (Lill *et al.*, 1969), T3 (Bautz *et al.*, 1969), T4 (Bautz and Bautz, 1971, Rüger, 1971), T5 (Schafer *et al.*, 1973, Le Talaer *et al.*, 1973), T7 (Chamberlin and Ring, 1972, Pribnow, 1975), fd(Hyden *et al.*, 1972, Seeburg and Schaller, 1975). λ (Nakano and Sakaguchi, 1968, Le Talaer and Jeanteur, 1971) with several bacterial operons, including *Lac* (Von Hippel *et al.*, 1974, Dickson *et al.*, 1975), *gal* (Anderson *et al.*, 1974), *trp* (Squires *et al.*, 1975) and the promoter for tyrosine transfer RNA (Sekiya *et al.*, 1975). Similar promoter sites are also found in the mammalian viruses simian virus 40 and human adenovirus 2 (Allet *et al.*, 1974). The general properties of open promoter complexes can be listed as follows:

a) The RNA polymerase holoenzyme - DNA complex shows a high stability with a half life of many hours (Hinkle and Chamberlin, 1972).

b) The complex shows a dependence on temperature and ionic strength. The open promoter complex is stabilized at higher temperature and in low salt conditions (Travers, 1974, Mangel and Chamberlin, 1974a).

c) Resistance to inhibitors - RNA polymerase when bound to DNA in promoter complexes is more resistant to rifampicin (Lill *et al.*, 1969), heparin (Sogo *et al.*, 1976) and poly (rI) (Niyogi and Underwood, 1975). RNA polymerase pretreated with herapin does not bind at all to T7 DNA. After rifampicin pretreatment, the RNA polymerase still binds efficiently to

all four sites of the early region of T7 DNA, as well as unspecifically to other positions on the DNA molecule (Sogo *et al.*, 1976). However, both heparine and rifampicin inhibit RNA initiation if they are added before the RNA polymerase binds to DNA. RNA polymerase has a great affinity for poly (rI) because it is a single stranded molecule which competes with the double-stranded DNA substrate.

d) RNA polymerase bound to open promoter sites initiates RNA synthesis very rapidly when ribonucleoside triphosphates are added (Mangel and Chamberlin, 1974a).

c) The RNA polymerase in open promoter complexes, protects about 40 base pairs from DNase digestion (Le Talaer and Jeanteur, 1971, Heyden *et al.*, 1972).

ii) Promoter site selection

A Model

In the most widely accepted model for site selection by RNA polymerase the enzyme first recognizes a specific sequence on the DNA duplex and forms a closed promoter complex. The enzyme then interacts more strongly with the duplex at the same time as melting occurs in the promoter region to form an open promoter complex. This latter complex is able to initiate RNA synthesis (Travers *et al.*, 1973, Chamberlin, 1974).

It has been proposed that promoters may contain palindromic sequences which can form hairpin structures (Sobell, 1972) but a uniform duplex is thermodynamically more stable than a hairpin structure (Gralla and Crothers, 1973). However, a number of promoter regions which have been sequenced contain areas of two fold symmetry in the sequence (Lewin, 1974,

Okamoto *et al.*, 1975, Schaller *et al.*, 1975) and this type of symmetry has been implicated in enzyme-nucleic acid interaction (Lewin, 1974).

It has also been postulated that RNA polymerase will recognize A-T rich regions and in general this is the case (Le Talaer *et al.*, 1973). It is known that A-T rich regions could more easily form single-stranded loops in the DNA to which the RNA polymerase could bind than G-C rich regions. However, not all A-T rich regions are promoter sites (Gomez and Lang, 1972). Furthermore, the looped state or partial melting of DNA because of its transient nature cannot easily explain the rapid rate of promoter site selection found for example in T7 DNA (Hinkle and Chamberlin, 1972). Several groups have noted that a common sequence can be deduced from promoter sites sequenced from DNA regions protected with RNA polymerase after nuclease digestion (Dhar *et al.*, 1974, Pribnow, 1975, Schaller *et al.*, 1975, Sekiya *et al.*, 1976). The sequence

5' TAT Pu ATG 3'
3' ATA Py TAC 5'

is located 5 or 6 bases before the position of the RNA initiation site. This might be the possible melting sequence that is bound to the RNA polymerase in the open promoter complex (Pribnow, 1975). Table 1 shows a summary of all the promoter sequences with the common sequences aligned.

Actually, the promoter site structure on DNA must be more complicated and contain additional information. Schaller and co-workers (1975) have demonstrated that RNA polymerase does not bind again to the DNA fragments isolated by binding RNA polymerase and whole fd DNA followed by nuclease digestion of the unprotected DNA regions. This implies that

tyr tRNA	C G T C A T T T G A T A T G A T G C G C C C C↓G C T T C C C	Sekiya and Khorana (1974)
T_7 A_2	T A A C A T G C A G T A A G A T A C A A A T C↓G C T A G G T	Pribnow (1975a)
T_7 A_3	G T A A A C A C G G T A C G A T G T A C C A C↓A T G A A A C	Pribnow (1975b)
fd	T G C T T C T G A C T A T A A T A G A C A G↓G G T A A A G A	Schaller *et al.* (1975)
lac	T T C C G G C T C G T A T G T T G T G T G G↓A A T T G T G A	Dickson *et al.* (1975)
lac UV5	T T C C G G C T C G T A T A A T G T G T G G↓A A T T G T G A	Gralla (in preparation)
λ P_1	C A C T G G C G G T G A T A C T G A G C A C↓A T C A G C T C	Maniatis *et al.* (1975)
λ P_r	C C T C G G C G G T G A T A A T G G T T G C↓A T G T A C T A	Waltz and Pirrotta (1975)
λ P_m	C C C T T G C G G T G A T A G A T T T A A C G T↓A T G G T A	Waltz *et al.* (1976)
trp operon	T C G A A C T A G T T A A C T A G T A C G C↓A A G T T C A C	Bennet *et al.* (1976)
λ P_o	C T G T A T T T G T C A T A A T G A C T C C T↓G T T G A T A	Scherer *et al.* (1977)
gal operon	C G A T C T T T G T T A T G C T A T G G T T↓A? T T T C↓A? T A	Musso *et al.* (1977)

Table I. Sequences of promoter regions

The sequence underlined indicates the "Pribnow box", the arrows the start of transcription. (? = postulated mRNA starting point).

some DNA sequences not protected from nuclease digestion are necessary for recognition or binding to the DNA duplex.

The Sigma subunit

The sigma subunit has been shown to play a crucial role in the selection of promoter sites by RNA polymerase. In general, the sigma subunit increases the specificity of the overall reaction carried out by the RNA polymerase (Hinkle and Chamberlin, 1971). The sigma subunit activates promoter recognition, separation of the two DNA strands and initiation of RNA synthesis. It also reduces aggregation of RNA polymerase and non specific initiation at single-strand breaks in DNA. Thus, it is necessary that in any RNA polymerase binding studies, the enzyme preparation must contain sufficient amounts of sigma subunit to ensure specificity of site selection.

c) Mapping of RNA polymerase recognition sites.

i) Visualization by electron microscopy

The RNA polymerase molecule has a diameter of about 120 Å (Slayter and Hall, 1966) which is far greater than the diameter of the duplex DNA molecule and bound molecules can be visualized easily in the electron microscope. For this kind of experiment, the classical Kleinschmidt technique (1959, 1968) can be used but, in general, because of the difficulties in identifying a protein - DNA complex when using the protein monolayer technique, other modifications of the technique have been developed and used to map *E. coli* RNA polymerase binding to DNA genomes of T7 and T3 (Bordier and Dubochet, 1974, Portman *et al.*, 1974, Koller *et al.*, 1974).

ii) Retention of complexes on membrane filters

This technique was developed for rapid detection of RNA polymerase binding to DNA. Whereas native DNA and RNA polymerase, separately pass through membrane filters, a complex of the two is quantitatively retained (Jones and Berg, 1966). However, since one RNA polymerase molecule is sufficient to retain one DNA molecule (Hinkle and Chamberlin, 1971) determination of the number of RNA polymerase binding sites is difficult. It has been demonstrated by electron microscopy (Richardson, 1966) and by the membrane filter technique (Hinkle and Chamberlin, 1972a, b) that there is more than one RNA polymerase binding site per DNA molecule even in the simplest bacteriophages. As an example, Hinkle and Chamberlin (1972a) found a maximum of about 8 enzyme molecules bound to T7 DNA and four or five were found on fd RF DNA (Hyden *et al.*, 1972).

In order to map the binding sites on DNA molecules, restriction enzymes have become very useful. Restriction enzymes are endonucleases which cleave double-stranded DNA, and they are divided in two classes: class 1 and class 2. Both classes, recognize a specific sequence and their action on a DNA molecule produces a specific set of fragments. The first restriction enzyme class hydrolyzes DNA at the same site or sequence that is recognized. The second class hydrolyzes DNA at a displaced position from the recognized sequence. Restriction enzymes have been used in order to construct cleavage maps of the bacteriophage genomes, as for example, ϕX174 (Lee and Sinsheimer, 1974), fd (Takanami *et al.*, 1975) and S13 (Grosveld *et al.*, 1976); and for mapping and characterization of promoters in bacteriophage DNAs (Chen *et al.*, 1973, Allet and Solem, 1974, Humphries *et al.*, 1974, Seeburg and Schaller, 1975).

If it is possible to bind the RNA polymerase on fragments of DNA with the same efficiency and specificity as on native DNA molecules, it becomes possible to work with one unique binding site at a time in order to study the mechanism of interaction of RNA polymerase and the DNA. One can also determine which restriction fragments bind the RNA polymerase and thus study the organization of the binding sites and the control of transcription of the genome. The easiest way to identify which fragment contains an RNA polymerase binding site is to incubate the enzyme with the mixture of fragments produced from the restriction enzyme digest of a DNA and then pour the mixture on a membrane filter. Only those fragments which bind to RNA polymerase will be retained on the filter (Takaya and Takanami, 1974, Okamoto *et al.*, 1975).

2. Initiation:

a) Specificity of initiation.

For initiation of synthesis of RNA at a specific site on the DNA template, the RNA polymerase must recognize a specific DNA structure or sequence and a stable complex should be formed. The enzyme initiates RNA synthesis by binding the first ribonucleoside triphosphate and catalizing the formation of a phosphodiester link between the first ribonucleoside triphosphate which is usually a purine (Maitra and Hurwitz, 1965) and the second ribonucleoside triphosphate, involving the release of inorganic pyrophosphate. The complex now includes a DNA molecule, the RNA polymerase and a dinucleoside tetraphosphate of the type: ppprRprN. In order to obtain specific binding and initiation, the sigma factor is absolutely necessary until a dinucleoside tetraphosphate is formed

(Hinkle and Chamberlin, 1971, 1972). This dinucleoside is then elongated by sequential polymerisation of nucleoside monophosphates at the 3' -OH terminus of the growing RNA chain. The presence of rATP and/or rGTP during complex formation between RNA polymerase and DNA has been shown to stabilize the complex as soon it is formed and it has been proposed that synthesis of a dinucleoside tetraphosphate or even a short polymer (Nüsslein and Schaller, 1975) forms a very stable tertiary complex. If this is true, a dinucleoside tetraphosphate should significantly stabilize an RNA polymerase and DNA complex.

The stabilizing dinucleotide (primer) must be complementary to the DNA template strand which is transcribed and pair with the DNA strand at the initiation sequence. This behavior has been shown for the A3 promoter of T7 DNA (Pribnow, 1975), the promoter sequence is

--- A-C-C-A-C-A-T-G-A-A ---
--- T-G-G-T-G-T-A-C-T-T ---

and the RNA molecule normally synthesized is ppprAprUprGprA---. In this case transcription was stimulated by the dinucleotides rAprC, rCprA, rAprU, and rCprU (Minkley and Pribnow, 1973).

It seems then that RNA polymerase has a certain freedom of movement along the DNA template and that it can be displaced from its normal start site. This kind of displacement has been found by several groups who have studied *in vitro* messenger RNA starting sequences (Maizels, 1973, Smith *et al.*, 1974). As an example, the *lac* U.V. 5 promoter sequence is:

5' --- G-T-G-G-A-A-T-T-G-T --- 3'
3' --- C-A-C-C-T-T-A-A-C-A --- 5'

(Gralla, in prep.) and it has been reported that the initiation sequences of the *lac* mRNA transcribed from this promoter are predominantly ppprAprAprU

prU.. (85%) but also pprGprAprAprU (15%) (Maizels, 1973).

If we exclude template-directed specificity in the initiation and elongation process, the specificity of chain initiation can be narrowed to template recognition and binding to form an "open" promoter complex. This kind of specificity has been discussed previously. However, it is essential to distinguish between *in vivo* and *in vitro* initiation of transcription. *In vivo* initiation of an RNA chain is restricted to specific sites on the genome. This is true also *in vitro* if intact DNA molecules are used as templates. Thus it is potentially dangerous to work with a DNA template extensively broken or partially single-stranded. It is known that RNA polymerase can bind to a nick and eventually initiate an unspecific RNA chain (Hinkle *et al.*, 1972). The frequency of incorrect initiation by RNA polymerase when interacting with an intact DNA molecule or *in vivo* is unknown.

b) Mapping of the RNA starting sequence sites.

With several bacteriophage DNAs, it has been possible to initiate specific RNA chains *in vitro* and to isolate them. In order to localize or map the sites at which the RNA molecule is initiated, two approaches have been successful. The first, which has been followed by Takanami and co-workers (1975) on bacteriophage fd is to identify which restriction fragment can initiate RNA synthesis and correlate this with the fd restriction map (Takanami *et al.*, 1975). Takanami's group found that RNA chains were specifically initiated even after cleavage of DNA with restriction endonucleases. However, the disadvantage of this technique is that if the RNA starting site is close to a restriction enzyme cleavage

site, initiation becomes difficult. This problem can be overcome by using other restriction enzyme fragments where the same promoter is localized in the middle of the fragment.

The second approach is to test which restriction fragment will hybridize with the 5'-terminus of the isolated messenger RNA (Okamoto *et al.*, 1975, Smith and Sinsheimer, 1976). Each restriction fragment is denatured and immobilized on a nitrocellulose filter. Hybridization is performed with each immobilized restriction fragment and the amount of radioactive RNA anealed with the restriction fragment is measured by resistance to RNAse (Gillepsie and Spiegelman, 1965). The disadvantages of this technique are the same as those described above. Again, the DNA used as a template for RNA synthesis has to be intact for the same reasons as for mapping RNA polymerase binding sites. Non specific transcription can also be due to non specific initiation due to core polymerase (either originally present in the RNA polymerase preparation or formed by dissociation of sigma subunit from the holoenzyme). As it has already been described, the core polymerase is still able to catalyze the synthesis of an RNA chain. However, it is unable to localize promoter sites on DNA templates (Burgess *et al.*, 1969, Sugiura *et al.*, 1970, Bautz and Bautz, 1971). The core polymerase is probably defective in the function needed for opening the DNA strands (Saucier and Wang, 1972). This is supported by the finding that the greater affinity of core polymerase is for single-strand breaks in DNA (Vogt, 1969, Hinkle *et al.*, 1972).

Non specific transcription might be due also to initiation of RNA synthesis by RNA polymerase holoenzyme molecules bound to the DNA duplex at sites that occasionally become competent (Chamberlin, 1976).

3. Elongation:

The formation of the first phosphodiester bond during initiation of synthesis of RNA converts the binary RNA polymerase - DNA complex into a tertiary complex which includes the immature RNA chain. The tertiary complex is very stable to salt dissociation (Naito and Ishihama, 1975) and insensitive to the effects of inhibitors such as rifampicin (Mangel and Chamberlin, 1974b) heparin (Scäfer *et al.*, 1973) and high salt which normally interact with the free RNA polymerase or the binary complex. The chain elongation reaction involves the repetitive addition of ribonucleoside monophosphate residues to the 3' OH end of the growing RNA chain.

During the elongation mechanism the binding of sigma subunit to the RNA polymerase molecule is reduced and dissociation of the sigma subunit occurs (Wu *et al.*, 1975). The sigma subunit after its release can interact with other core polymerase molecules and initiate new RNA chains. The RNA elongation step and the termination of the RNA chain is thus performed by the core polymerase (Travers and Burgess, 1968).

The specificity of chain elongation is envisaged only at the level of the template-directed incorporation of nucleoside monophosphates onto the nascent RNA chain. This specificity has been tested (Bujard and Heidelberger, 1966) and an error frequency of less than 1 in 3000 has been found, indicating great fidelity during transcription.

4. Termination:

In vitro synthesis of RNA on most DNA templates proceeds for a distance along the DNA template depending on the incubation conditions and

does not usually terminate at unique sites, this produces a very disperse distribution of RNA molecules of different sizes. RNA polymerase recognizes some terminator sites *in vitro*, but it fails to stop at others unless a protein, the *rho* factor, is present. Thus, it is possible to distinguish between two kinds of termination: 1) direct termination of growing RNA chains by interaction of RNA polymerase with sites on DNA and 2) *rho* dependent termination of RNA chains at sites on DNA that require the presence of *rho* factor.

a) Termination in the absence of *Rho* factor.

In T7, termination at specific sites can take place *in vitro* in the absence of *rho* factor, either in the presence of high potassium chloride or low magnesium chloride concentrations (Minkley and Pribnow, 1973). The lack of a termination component has been suggested by experiments showing that the RNA transcripts *in vitro* were much larger than those isolated *in vivo* (Travers, 1971) (Schaefer and Zillig, 1973).

b) Termination in the presence of *Rho* factor.

Rho factor was first purified by Roberts (1968) from *E. coli* and shown to be a protein. It enables RNA polymerase to complete transcription at certain sites on bacteriophage DNAs . It has been shown that the RNA transcribed from DNA remains associated with its template, unless *rho* factor is present to cause its release (Roberts, 1969). In the absence of *rho*, the RNA and DNA sediment together as a heavy complex of more than 50S. If *rho* has been added to the system, the RNA sediments at about 32S.

The addition of *rho* causes a large decrease in the size of the RNA.

Furthermore, it has been shown that *rho* has no effect on the RNA if the latter has been synthesized in the absence of *rho* (Goldberg and Hurwitz, 1972). Terminators that function *in vitro* when *rho* factor is absent and those that function only in its presence seem to form two separate classes. However, there is evidence that *rho* factor is involved in *in vitro* termination (Richardson *et al.*, 1975). The best evidence for this is that two RNA products that are isolated from bacteriophage lambda infected cells can be made *in vitro* only in the presence of *rho* factor (Kourilsky *et al.*, 1971, Heinemann and Spiegelman, 1971). Thus, the terminators that produce these RNA species appear to coincide with those sites on the genome at which transcription ends *in vivo* in the absence of phage-specific regulators. A feature of these two RNA species from bacteriophage λ is a common sequence at the 3' end of the RNA. The sequence includes a G-C rich region followed by a stretch of 6 uracil residues terminated by an A residue (Lebowitz *et al.*, 1971, Blattner and Dahlberg, 1971). This kind of sequence has been found in the tryptophan operon leader RNA sequence (Bertrand *et al.*, 1975). The terminator that yields the *trp* operon leader RNA is effective not only *in vitro* but also *in vivo* since the majority of transcripts initiated at the *trp* promoter rarely represent the whole operon (Bertrand *et al.*, 1975). Actually, it is thought that this terminator in the leader RNA sequence is involved in *trp* operon regulation and serves as an attenuator of RNA transcription (Bertrand *et al.*, 1975).

However, it should be pointed out that neither the mechanism of action of *rho* factor, nor the control of *in vivo* termination, are well understood.

5. A Model for Transcription:

As discussed previously, transcription is a complex process which can be divided into several steps as follows:

1) A search phase in which RNA polymerase holoenzyme associates and disassociates with the DNA template until a promoter region is found.

2) Initial recognition step in which the RNA polymerase forms a specific but relatively loose complex with the closed DNA helix at the promoter (closed promoter complex).

3) The formation of an open promoter complex in which four to six base pairs of the DNA template duplex are opened. Binding of RNA polymerase to many promoters is very tight at this stage. Presumably, the RNA polymerase also selects the appropriate DNA strand for use as a template.

4) Migration of the RNA polymerase along the DNA strand until it reaches the starting point for initiation of RNA synthesis.

5) Rapid initiation of RNA synthesis in the presence of the four ribonucleoside triphosphates.

6) Elongation of the nascent RNA chain by addition of ribonucleoside monophosphate to the 3' OH end of the growing RNA chain derived from ribonucleoside triphosphate substrates.

7) Termination; when both the newly formed RNA chain and the RNA polymerase are released from the DNA template and from each other.

This is how transcription generally functions for most of the known systems but this is not necessarily the only way. Also, it should be pointed out that for many promoters step 4 does not occur, the RNA starting point coincides with the RNA polymerase binding site (Gilbert, 1976). Some results seem to show that the promoter may be a more complicated structure than was first thought, consisting of a longer DNA region than would be necessary for binding just one RNA polymerase molecule. It has been demonstrated that on T4 and T5 DNA, each promoter can bind more than one RNA polymerase molecule (Mueller, 1971, Schäfer *et al.*, 1973b). This has also been found on *E. coli* DNA by Travers (1976) who proposed that rRNA cistrons are served by multiple promoter sites called subpromoters, each of which would bind one RNA polymerase molecule. Similarly, on phage λ, Willmund and Kneser (1973) showed some promoters which could bind up to six RNA polymerase in a stable way which makes them resistant to heparin.

Step 7 does not always occur *in vitro*. For most DNAs, the termination of transcription is *rho* dependent *in vivo* and if the *rho* factor is absent, RNA synthesis may continue through the termination signal.

Now, which are the mechanisms that control the expression of transcription and at what level are they acting? In many systems and for many cistrons, there is no apparent control of transcription per se except a general control at the level of ribonucleoside triphosphate concentration or in some cases control of ribonuclease level in the cell. However, there are some examples of genuine control of transcription.

The control of RNA synthesis can be mediated by highly specific

regulatory proteins. These proteins either restrict or potentiate the initiation of RNA synthesis by binding to a DNA site in close proximity to the initial binding site for RNA polymerase.

The classic and best known example of this type of regulation is the *lac* operon in *E. coli*. The synthesis of *lac* mRNA is determined by the activity of two control elements, the *lac* repressor and the catabolite gene activator protein (CAP). So, a positive and negative control of transcription operates in this system (Dickson *et al.*, 1975).

a) Positive control

The CAP protein is a dimer which is thought to stimulate mRNA transcription by facilitating formation of the open promoter complex. The CAP interaction site generally exhibits two fold rotational symmetry and is located just before the RNA polymerase interaction site. The CAP facilitates RNA polymerase recognition and binding at the site and thus increases the mRNA expression. Furthermore, for CAP to be functional it has to be first activated by cyclic AMP.

This kind of regulation is also found in the *his* operon (Lewin, 1974), the *gal* operon (Musso *et al.*, 1977), and the *ara* operon (Greenblatt and Schleif, 1971).

b) Negative control

i) Repressor.

The repressor is a protein which recognizes a specific sequence of DNA, the operator, and prevents RNA polymerase from initiating transcription. In general, repressor activity is determined by the concentration of a low molecular weight effector which in most cases is either a

substrate or a metabolic product of the enzymes encoded by the operon. It was observed that when RNA polymerase is prebound to the promoter, the repressor cannot bind and prevent transcription. This suggests the existence of a functional overlap between the promoter and the operator regions as found in the *lac* operon (Dickson *et al.*) and the *trp* operon (Squires *et al.*, 1975).

The repressor need not necessarly be a protein. For example, in the *his* operon the repressor is the *his* t-RNA (Lewin, 1974).

ii) Attenuator

This type of regulatory element was first suggested for the *his* and isoleucine-valine operon (Wasmuth and Umbergen, 1973, Kasai, 1974) but the best example is the *trp* operon. The *trp* operon mRNA has a leader region of about 160 bases before the start of the sequences coding for the first protein, the *trp* E gene product (anthranilate synthetase component I). Genetic findings substantiate the proposal that a region between the operator and the first gene (*trp* E) normally acts to reduce operon expression (Jackson and Yanofsky, 1973). Bacterial strains in which this region, called the attenuator region, is genetically deleted exhibit elevated levels of both *trp* mRNA and *trp* operon coded enzymes (Jackson and Yanofsky, 1973). Also *in vitro* studies have provided direct evidence that *trp* operon transcription generally terminates near position 145 in the leader region, about 20 nucleotides before *trp* E (Lee *et al.*, 1976). *Trp* tRNA and *rho* factor are probably involved in this early termination. It is thought that the *rho* factor prevents the RNA polymerase molecule from entering the structural genes of the operon (Yanofsky, 1976).

SECTION II.

DNA SEQUENCE ANALYSIS

A. Isolation of the starting material.

A major problem in DNA sequence determination is the size of DNA molecules compared with RNA molecules. The smallest single-stranded viral DNA molecules have molecular weights of approximately 5,000 nucleotides and double-stranded DNA molecules start at 5,000 base pairs as for example ϕX174 DNA (Sanger *et al.*, 1977) and S13 DNA (Spencer *et al.*, 1972). Because of the availability of RNA molecules of small size and also several specific RNAses, sequencing techniques have been most extensively developed for RNA molecules. Attempts to sequence large RNA molecules such as phage MS2 or 16S ribosomal RNA have depended in part upon the success of subdivision of the molecules into smaller fragments by partial digestion with RNAse T1 (Fellner *et al.*, 1970). Some long DNA sequences have been determined indirectly by *in vitro* transcription of the DNA with RNA polymerase and sequence determination of the RNA transcripts by RNA sequencing techniques (Gilbert and Maxam, 1970, Blattner and Dahlberg, 1972).

However, the use of direct DNA sequencing techniques is essential for the investigation of sequences not transcribed *in vivo* although these can be transcribed under particular conditions. Non-transcribed sequences are of particular biological interest and a knowledge of the nucleotide sequences of the regions of DNA molecules which bind various repressors, positive control elements and promoters should help our understanding of the molecular details of their action or mechanism of recognition.

As the first step of sequence analysis of a small region of DNA, it is necessary to use specific techniques of DNA degradation to provide short DNA fragments. The following methods are used to generate DNA fragments of suitable length for sequencing:

1) Depurination. This chemical degradation destroys the N-glycoside bond of all the purine nucleotides in DNA molecules (Burton and Peterson, 1960, Burton, 1967). Formic acid treatment of the depurinated DNA then leaves the pyrimidine oligonucleotides which can be sequenced (Ling, 1972a, b).

2) Restriction enzymes are sequence specific endonucleases which cleave double-strand DNA and permit a specific division of DNA molecules into appropriate size fragments for sequence determination.

3) T4 induced endonuclease IV is another highly specific endonuclease. It cleaves single-stranded DNA at C residues (Sadowski and Bakyta, 1972).

4) Protection of some site on DNA against nuclease digestion. Some proteins eg. repressor molecules, RNA polymerase can bind strongly to specific regions of DNA and thus protect these regions from hydrolysis by DNAses. The protected DNA is then available for sequence analysis.

1. Protection of specific DNA sequences.

a) by RNA polymerase.

This approach, pioneered by Schaller and co-workers (Schaller *et al.*, 1975) permits the isolation of a DNA fragment protected by RNA polymerase against DNAse digestion, the isolated fragment can then be subjected to sequence analysis. The open promoter complexes of RNA polymerase and DNA

are very stable and define a limited number of binding sites.

Several groups have used this technique for isolation (Ruger, 1971, Le Talaer and Jeanteur, 1971, Heyden *et al.*, 1972) and sequence analysis (Schaller *et al.*, 1975, Pribnow, 1975a, b, Sugimoto *et al.*, 1975).

b) by repressors and CAP factors.

The repressor protein as well as the CAP factor are both capable of binding to different specific sites on DNA in such a way that these sites become "protected". The λ repressor binding site has been sequenced (Maniatis *et al.*, 1974) as well as the *lac* repressor binding site (Gilbert and Maxam, 1973) and more recently the *trp* repressor binding site (Bennett *et al.*, 1976). The CAP factor binding site of the *lac* operon has been sequenced by Dickson *et al.*, (1975).

c) by ribosomes.

Ribosomes were used by Steitz (1969) to protect ribosome binding sites on mRNAs and the technique has been successfully extended to the viral strands of some bacteriophage DNAs which have the same orientation and sequence as the mRNA.

Similar to RNA polymerase, ribosomes protect certain DNA regions against nuclease digestion. A ribosome binding site has been isolated from φX174 single-stranded DNA (Robertson *et al.*, 1973) and from the *trp* operon (Platt *et al.*, 1976).

2. Nuclease digestion.

a) Restriction enzymes.

DNA sequencing has become easier since the discovery of

site-specific restriction endonucleases (Kelly and Smith, 1970). Numerous restriction enzymes are now available and many recognition sequences are known (Roberts, 1976). It is therefore possible to achieve considerable specific fragmentation of DNA macromolecules. Cleavage maps have been constructed for many bacteriophage DNAs (eg. Lee and Sinsheimer, 1974, Takanami *et al.*, 1975, Grosveld *et al.*, 1976).

The use of several different restriction endonucleases at the same time provides very small restriction fragments. A restriction fragment choosen for sequence analysis can be further degraded with T4 endonuclease IV.

b) T4 induced endonuclease IV.

The majority of endonucleases acting on DNA do not show any specificity and degrade DNA into very complex mixtures of small oligonucleotides which are not suitable for sequencing.

The discovery of endonuclease IV (Sadowski and Hurwitz, 1969) provided an exceptional tool for degradation of DNA into specific oligonucleotides. This enzyme is specific for single-stranded DNA and hydrolyzes DNA at C residues, producing oligonucleotides having a pC at their 5' end (Sadowski and Bakyta, 1972, Grosveld and Spencer, 1977). In general, the average size of fragments obtained after T4 endo IV digestion varies between 5 and 20 nucleotides in length. These oligonucleotides can then be sequenced directly by partial exonuclease digestion as will be described later.

The only difficulty is that the purification of the enzyme is not easy and that the enzyme co-purifies with an endonuclease activity

(Sadowski and Bakita, 1972).

Partial digestion of phage fd DNA with the enzyme gave a specific and reproducible pattern on polyacrylamide gel electrophoresis (Ling, 1971). More recently Ziff *et al.* (1973) published conditions for partial digestion and further digestion with endonuclease IV of ϕX174 single-stranded DNA. When the digestion was carried out at low temperature and high salt, a limited digestion occured and a fragment of about 50 nucleotides was isolated and sequenced by further digestion with the enzyme and also by depurination. A sequence of 48 nucleotides has been obtained and the results show that endonuclease IV is without doubt an excellent tool for DNA sequence analysis. It can provide the basis for a detailed finger printing system for DNA when used to digest large DNA fragments such as those obtained by digestion with restriction endonucleases and fractionated gel electrophoresis (Grosveld *et al.*, 1977).

B. Methods for DNA sequencing.

1. Chemical degradation of DNA.

The discovery of site specific DNA endonucleases such as restriction enzymes and T4 endonuclease IV dates from 1970 (Smith and Wilcox, 1970, Sadowski and Hurwitz, 1969). Before then, chemical degradation was the only technique available for specific cleavage of DNA.

a) Acid Hydrolysis.

The formic acid hydrolysis of DNA in the presence of diphenylamine (Burton and Peterson, 1960) allows the selective cleavage of purine nucleotides from DNA. Long pyrimidine oligonucleotides have been obtained from different DNAs. The first experiment of depurination involving

fractionation according to chain length of the various pyrimidine oligonucleotides was on calf thymus DNA (Spencer and Chargaff, 1960, 1963). These experiments were then carried out on bacteriophage ϕX174. The pyrimidine oligonucleotides were fractionated on DEAE Sephadex by Hall and Sinsheimer (1963), who used ^{14}C-labeled phage DNA with depurinated calf thymus as a carrier. Peterson and Reeves (1966) ameliorated the technique, by separating the pyrimidine oligonucleotides of calf thymus DNA and salmon sperm DNA according to base composition. The largest pyrimidine oligonucleotide from DNA had eleven bases (Hall and Sinsheimer, 1963) and the same was found for the closely related bacteriophage S13 DNA (Cerny *et al.*, 1969). Studies of the same type have been performed on different DNAs such as fd, f1 (Ling, 1972), T7 and λ DNA (Mushynski and Spencer, 1970a, b).

There are two limitations of the depurination method. It is not possible to perform a partial digest and the technique does not yield overlapping pyrimidine oligonucleotide sequences.

However, pyrimidine sequences have provided information for comparative studies on different DNAs (Ling, 1972, Harbers *et al.*, 1976) permitted the sequence of an RNA polymerase binding site of bacteriophage fd DNA (Schaller *et al.*, 1975) and have been essential in the sequence analysis of ϕX174 DNA (Sanger *et al.*, 1977).

b) Hydrazinolysis.

As a counterpart to pyrimidine oligonucleotides analysis, hydrazinolysis is the study of the purine oligonucleotides released after treatment of the DNA with hydrazine and alkali as described by Harbermann (1963).

Figure 3 : Methylation mechanism for deoxyguanosine in DNA and β-elimination mechanism from Gilbert and Maxam (1975). Methylation for deoxyadenosine is the same except that it is the N-3 position which is alkylated.

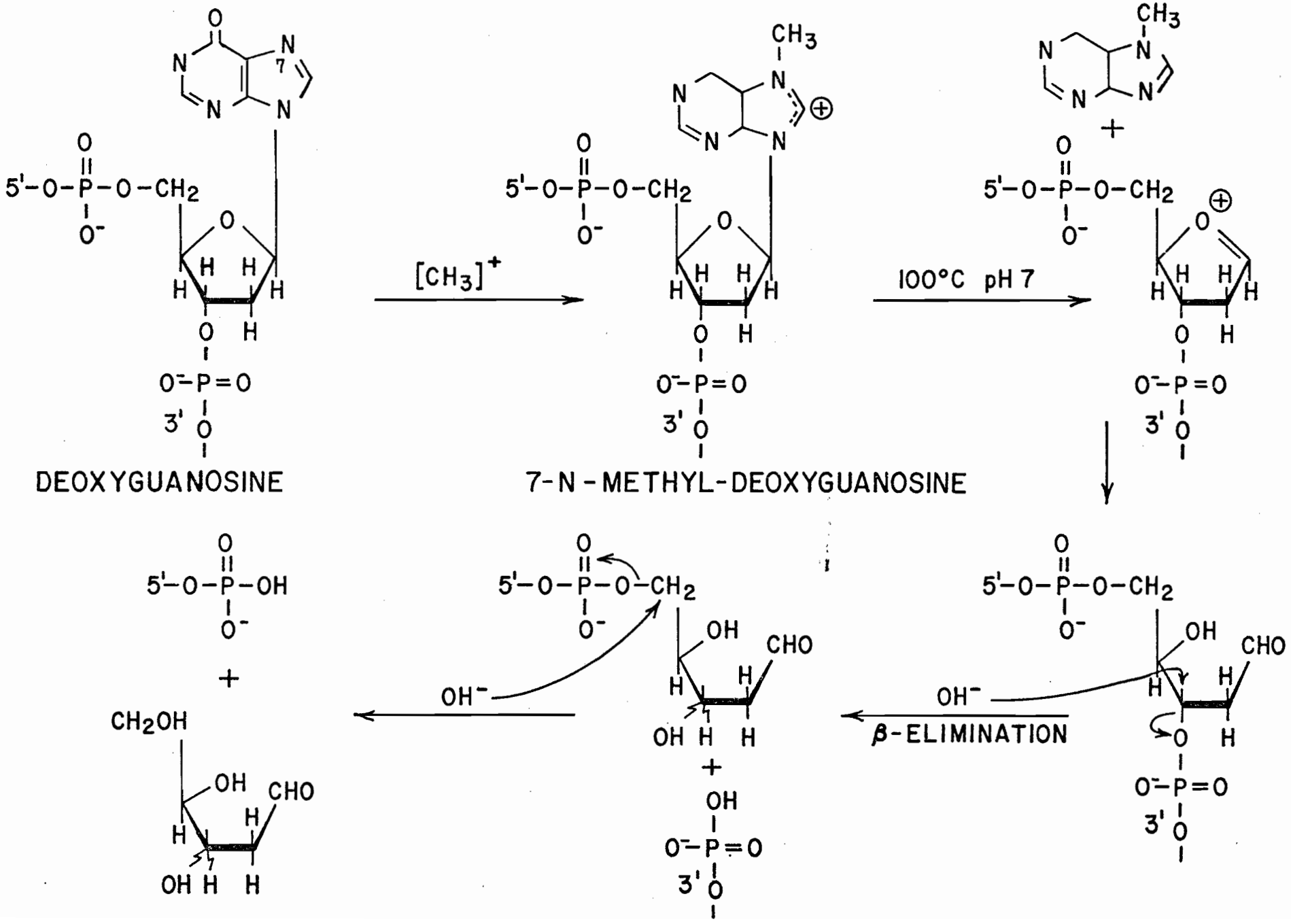
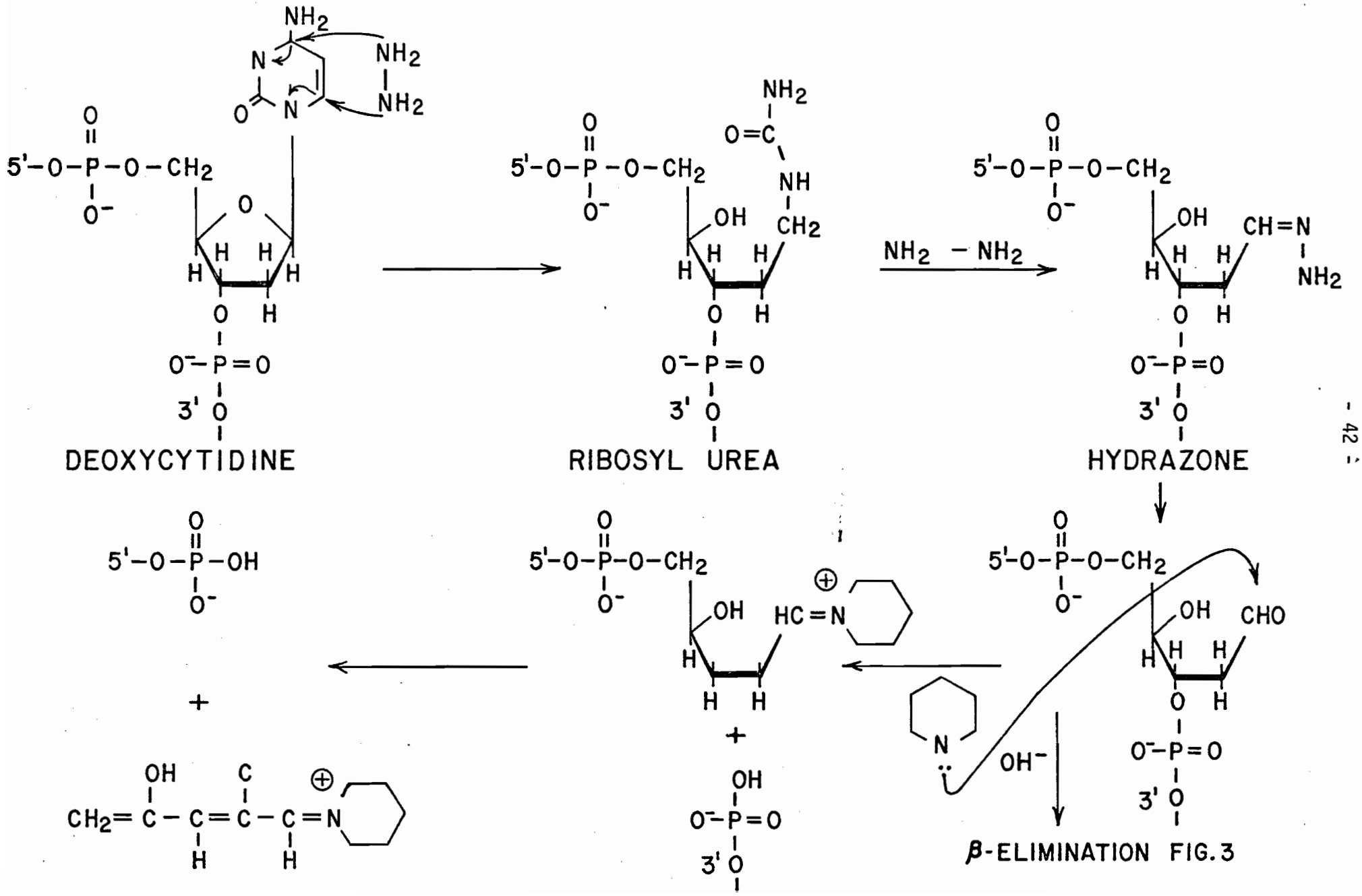

DEOXYGUANOSINE
[CH3]+
7-N-METHYL-DEOXYGUANOSINE
100°C pH 7
β-ELIMINATION
OH-
OH-

Figure 4 : Hydrazinolysis of deoxycytidine and cleavage of DNA in presence of piperidine (from Gilbert and Maxam , 1975). Hydrazinolysis of deoxythymidine is the same as deoxycytidine.


DEOXYCYTIDINE
RIBOSYL UREA
HYDRAZONE
β-ELIMINATION FIG.3

Hydrazine reacts with thymine and cytosine residues cleaving the base and leaving ribosylurea (Fig. 4). Hydrazine then reacts further to produce a hydrazone which, when incubated in alkali, allows cleavage of the DNA chain by β elimination (Fig. 3).

This analysis was first carried out by Sedat and Sinsheimer (1964) on ϕX174 DNA containing labeled purines. They showed the presence of purine oligonucleotides longer than any of the pyrimidine sequences isolated by depurination. They also showed that the proportions of several of the oligonucleotides of the two types differed. When DNA containing labeled pyrimidines was used, they showed complete destruction of the pyrimidines by hydrazinolysis. A further experiment on calf thymus DNA showed that is was possible to obtain equal proportions of purines and pyrimidine oligonucleotides of the same length, as required by the complementary structure of the DNA, although the method is not of great precision especially for large DNA molecules (Cape and Spencer, 1968, Cashmore and Peterson, 1969).

c) DNA methylation.

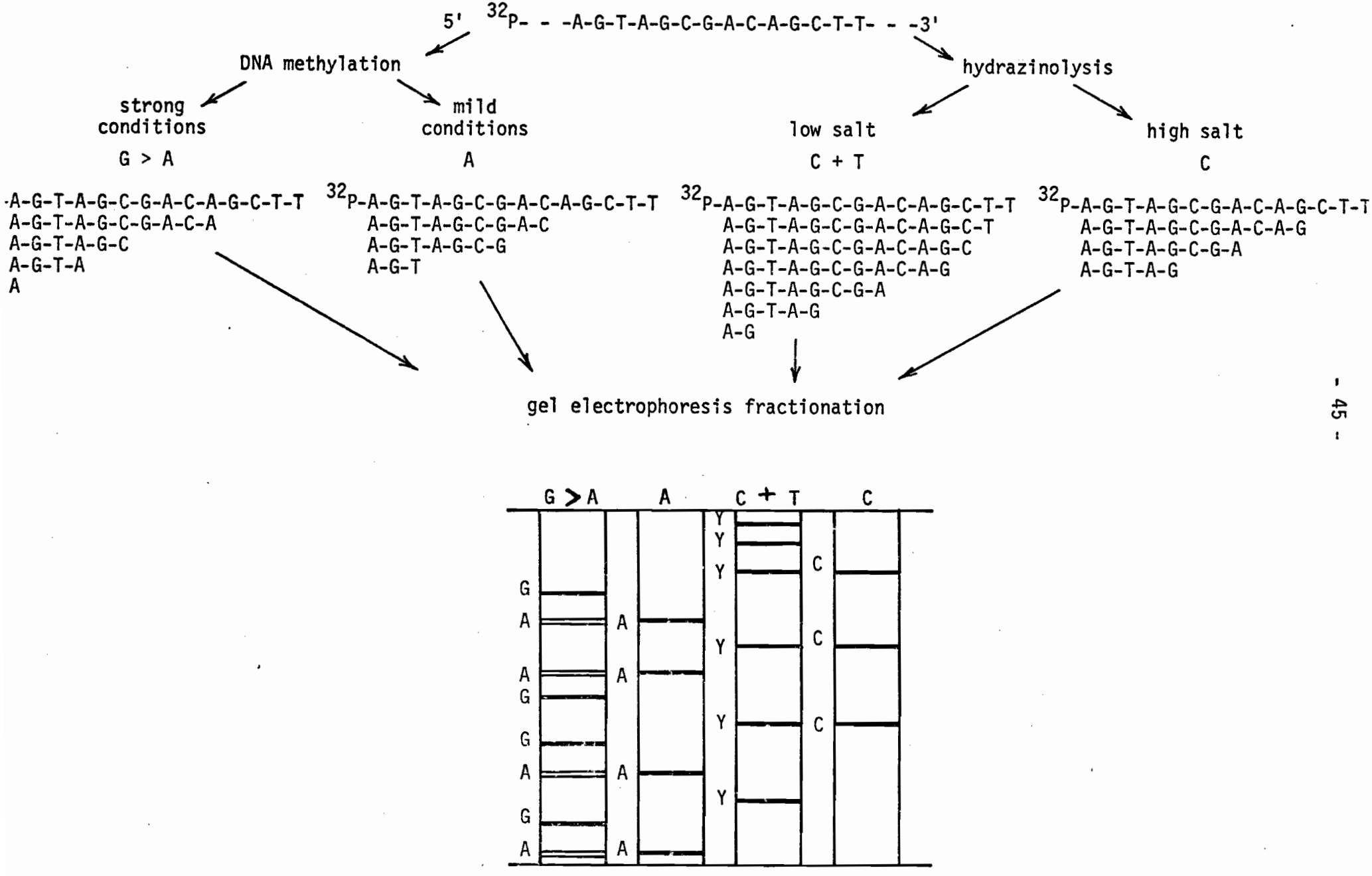
Alkylating agents have been used for a long time to induce mutations in DNA molecules (Lawley, 1966). During the past 12 years, Lawley, Brookes and their collaborators have systematically treated various DNAs with a number of methylating agents and have developed techniques for quantitation of 9 different alkylated bases. The first alkylated bases to be isolated from DNA were those present in relatively large quantities, namely 7-methyl-Guanine (7-MeG) and 3-methyl-Adenine (3-MeA). The nitrogen 7 of guanine is the most reactive nucleophilic site, both in the free base itself and in the guanosine-cytydine base pair in DNA (Pochon and Michelson, 1967, Sun and Singer, 1974).

Some methylating agents display a great specificity and will alkylate G residues almost exclusively, others alkylate the purines in varying amounts (Singer, 1975). These methylating agents are very useful for both RNA and DNA sequencing. In DNA, the methylated purines can be removed and the DNA chain cleaved in alkali by the β elimination mechanism (Fig. 3). This provides a new technique for specific degradation of DNA and is compatible with subsequent sequence analysis by partial exonuclease digestion (Fig. 8).

d) A rapid method for DNA sequencing.

A very recent DNA sequencing method developed by Gilbert and Maxam (1975) is based on partial methylation (Fig. 3) and hydrazinolysis (Fig. 4) of DNA fragments. The DNA fragment is labeled at its 5' end, with polynucleotide kinase and [γ-^{32}P]ATP. Dimethyl sulfate is used to methylate the purines in the DNA in the ratio 5 guanines to 1 adenine. The N-glycosidic bond of a methylated purine is unstable and breaks easily on heating at neutral pH, leaving an aldehyde at the C_1' position of the deoxyribose (a purinic site). Treatment with alkali cleaves the sugar from the neighboring phosphate groups (β elimination, Fig. 3). The resulting ^{32}P end labeled fragments are resolved on a polyacrylamide gel, the autoradiograph contains a pattern of dark and light bands. The dark bands arise from breakage at guanines, which methylates 5 fold faster than adenines. This strong guanine (dark bands) compared to weak adenine (light bands) pattern contains almost all the information for sequencing. However, ambiguities can arise in the interpretation of this pattern because the intensity of isolated bands is not easy to assess.

Figure 5 : DNA methylation and hydrazinolysis from Gilbert and Maxam (1975).


5' 32P- - -A-G-T-A-G-C-G-A-C-A-G-C-T-T- - -3'
DNA methylation
hydrazinolysis
strong conditions
G > A
mild conditions
A
low salt
C + T
high salt
C
-A-G-T-A-G-C-G-A-C-A-G-C-T-T
A-G-T-A-G-C-G-A-C-A
A-G-T-A-G-C
A-G-T-A
A
32P-A-G-T-A-G-C-G-A-C-A-G-C-T-T
A-G-T-A-G-C-G-A-C
A-G-T-A-G-C-G
A-G-T
32P-A-G-T-A-G-C-G-A-C-A-G-C-T-T
A-G-T-A-G-C-G-A-C-A-G-C-T
A-G-T-A-G-C-G-A-C-A-G-C
A-G-T-A-G-C-G-A-C-A-G
A-G-T-A-G-C-G-A
A-G-T-A-G
A-G
32P-A-G-T-A-G-C-G-A-C-A-G-C-T-T
A-G-T-A-G-C-G-A-C-A-G
A-G-T-A-G-C-G-A
A-G-T-A-G
gel electrophoresis fractionation
G > A
A
C + T
C
G
A
A
A
G
G
A
A
G
A
A
Y
Y
Y
Y
Y
Y
C
C
C

In order to overcome this problem, the information contained in this column of the gel is compared with another parallel column in which the breakage at guanine residues is repressed, leaving the adenine bands which now are the dark bands. This can be achieved because the N-glycosidic bond of methylated adenosine is less stable than that of the methylated guanosine.

For the alkylation of pyrimidine residues, Gilbert and Maxam (1975) used hydrazine as already described, except that they replaced the alkali treatment of the hydrazone, produced after hydrazinolysis, by piperidine. The specificity of alkylation, which is equal for both T and C residues in these conditions, is strongly enhanced for C residues in high salt so that the bands resulting only from cleavage at cytosine are observed.

So in fact, for sequencing a DNA fragment (previously ^{32}P-labeled at its 5' end), it is necessary to have four cleavage reactions: 1) strong G over weak A, 2) strong A, 3) C + T, and 4) strong C (see Fig. 5).

2. Enzymatic degradation of DNA.

a) DNA sequencing by 3' exonuclease digestion.

DNA fragments, about 20 nucleotides in length, obtained either by enzyme degradation of longer DNA fragments using endonuclease IV (Fig. 6) or by chemical degradation involving DNA methylation and hydrolysis (Fig. 7) and labeled *in vitro* with polynucleotide kinase and [γ-^{32}P]ATP, can be sequenced by partial exonuclease digestion and the products separated by ionophoresis-homochromatography (Brownlee and Sanger, 1969). The ionophoresis separates the oligonucleotides according to base composition and homochromatography separates them according to size. An

Figure 6 : Schematic representation of sequencing DNA using site-specific restriction endonuclease and the single-stranded specific T_4 endonuclease IV.

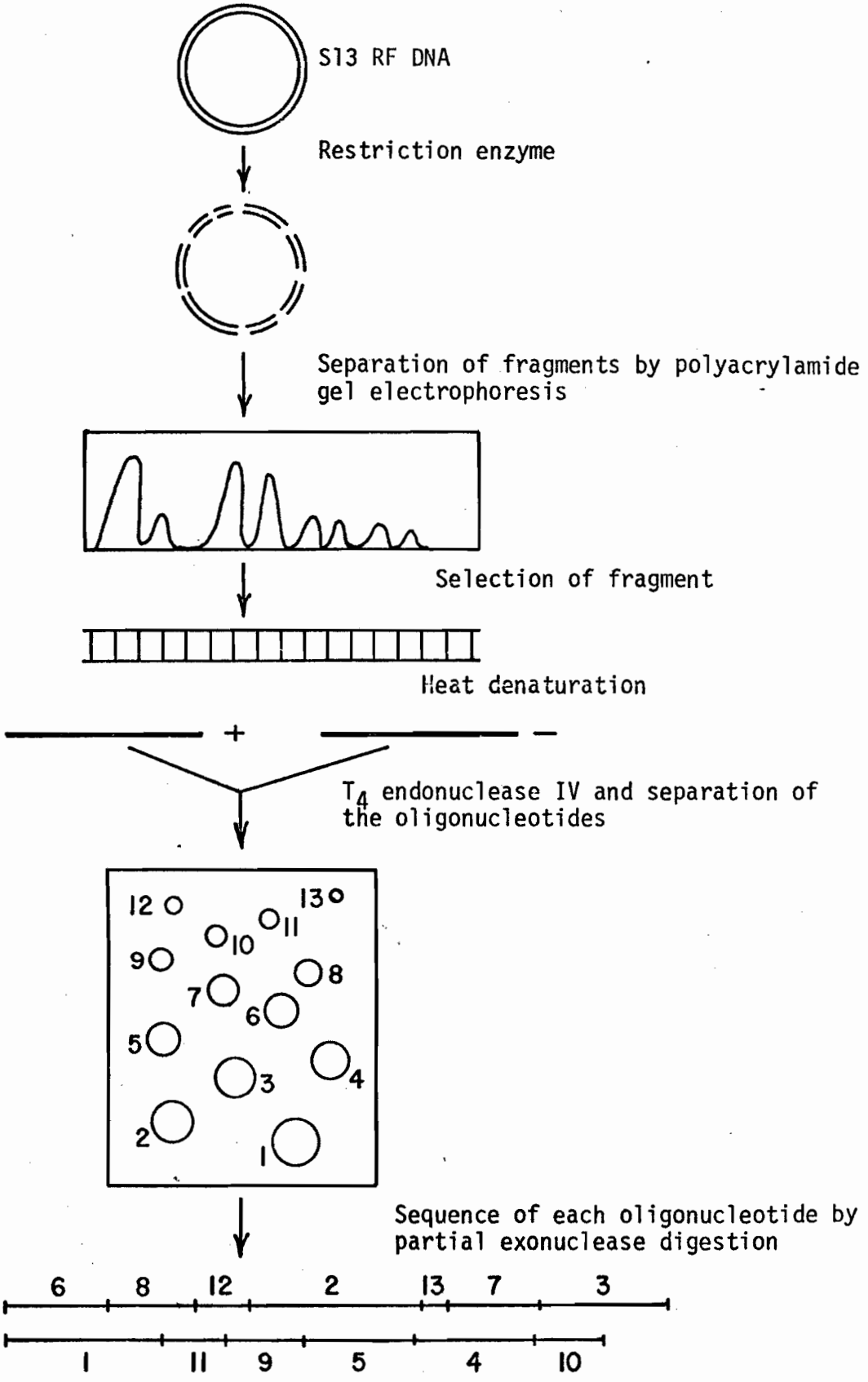
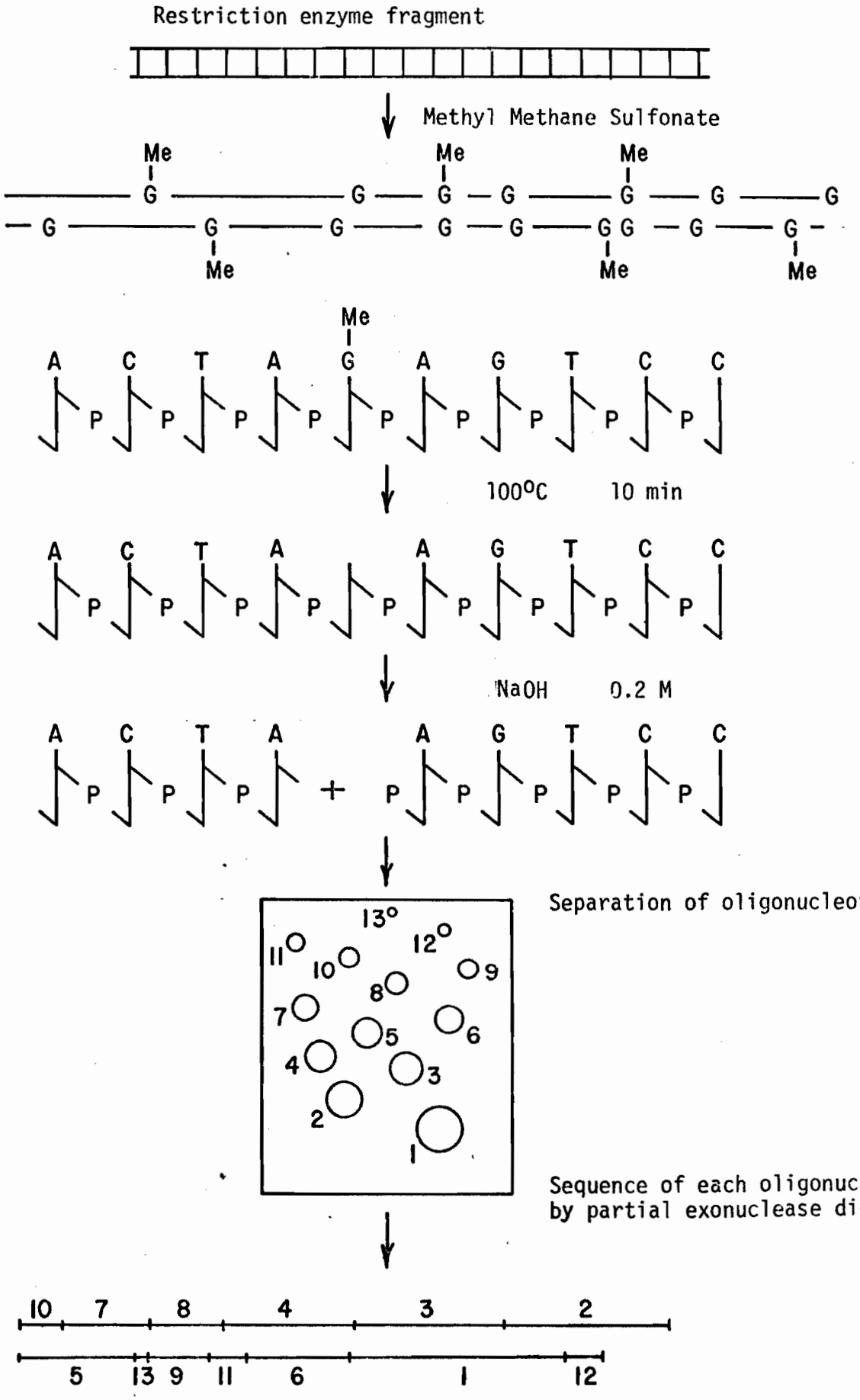

S13 RF DNA
Restriction enzyme
Separation of fragments by polyacrylamide gel electrophoresis
Selection of fragment
Heat denaturation
+
−
T4 endonuclease IV and separation of the oligonucleotides
12
13
11
10
9
8
7
6
5
4
3
2
1
Sequence of each oligonucleotide by partial exonuclease digestion
6 8 12 2 13 7 3
1 11 9 5 4 10

Figure 7 : Schematic representation of sequencing a DNA fragment using specific alkylation and cleavage with methylmethanesulfonate.


Restriction enzyme fragment
Methyl Methane Sulfonate
Me
A C T A G A G T C C
P
100°C 10 min
NaOH 0.2 M
Separation of oligonucleotide
13° 12° 11° 10° 9 8 7 6 5 4 3 2 1
Sequence of each oligonucleotide by partial exonuclease digestion
10 7 8 4 3 2
5 13 9 11 6 1 12

oligonucleotide, obtained in a pure form is partially degraded with a 3' exonuclease in such a way that the starting material and all the intermediate products including the final mononucleotide are present after the incubation (Fig. 8). The partial digest on the final chromatogram is visualized by autoradiography and the sequence can be determined from the characteristic shifts between successive spots, each corresponding to the loss of a nucleotide during the partial digestion. The loss of dA and dC or dG and dT result in a shift respectively to the left or to the right in the first dimension (ionophoresis). Also the loss of a purine versus a pyrimidine corresponds to a larger shift in the second dimension (homochromatography). Thus every nucleotide, when lost, can be characterized by the angle and the size of the shift, Fig. 8 (Galibert *et al.*, 1974).

However, the disadvantage of this technique is that it is sometimes difficult to interpret certain shifts particularly dA from dC. Certain parts of the chromatogram, especially near the 5' end of the sequence, are ambiguous to interpret and require additional sequence data.

b) DNA sequencing by 5' exonuclease digestion.

Another approach developed by Delaney and Spencer (1973) is the use of spleen phosphodiesterase, a 5' specific exonuclease.
Partial digestion with this enzyme in combination with a phosphatase, followed by phosphorylation with polynucleotide kinase introduces a 5' ^{32}P-label only at the 5' ends of the reaction product. After fractionation of the partial digest according to chain length, the sequence of the initial oligonucleotide is built up simply from the identity of

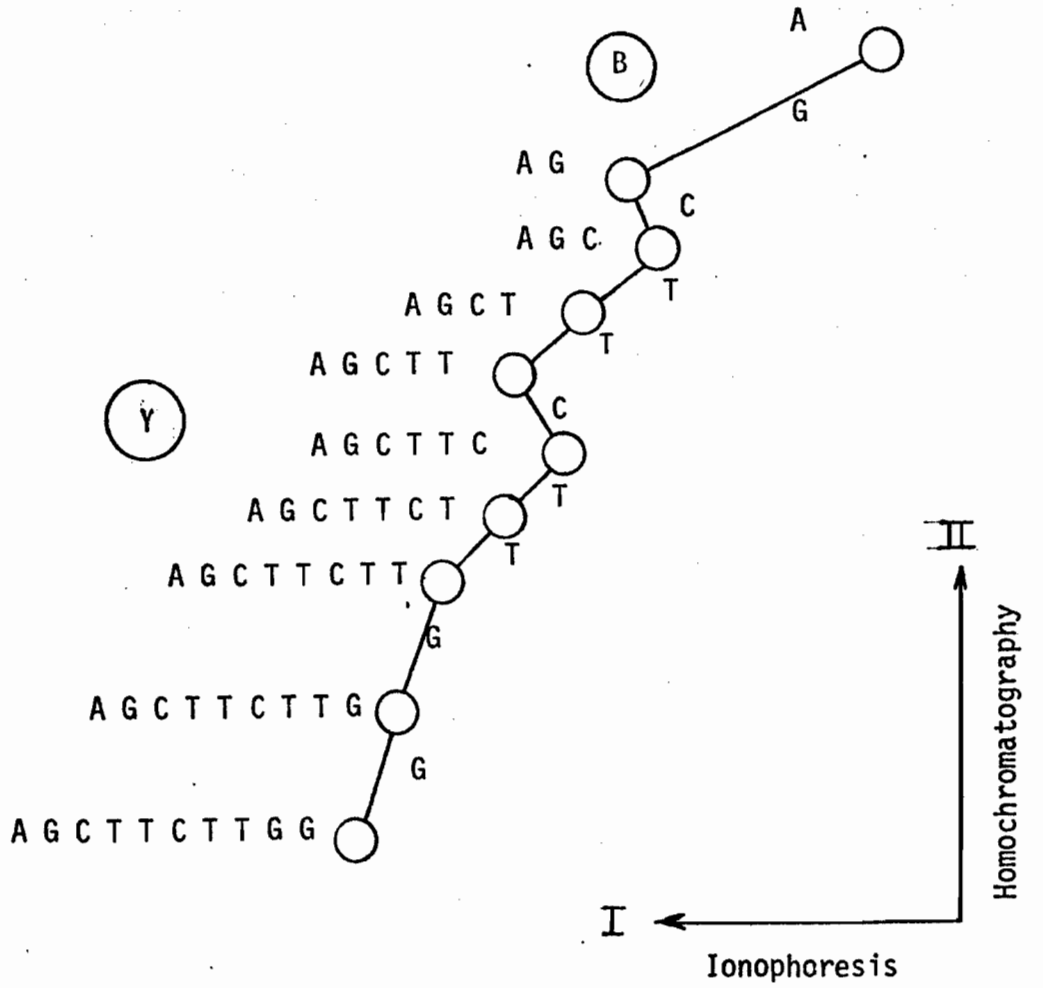
Figure 8 : Sequence chromatogram of a 5' ^{32}P labelled oligonucleotide (Wandering spot technique). B and Y are blue and yellow dye markers.

5' ^{32}P A-G-C-T-T-C-T-T-G-G

↓ partial 3' specific exonuclease

5' ^{32}P A-G-C-T-T-C-T-T-G-G
^{32}P A-G-C-T-T-C-T-T-G
^{32}P A-G-C-T-T-C-T-T
^{32}P A-G-C-T-T-C-T
^{32}P A-G-C-T-T-C
^{32}P A-G-C-T-T
^{32}P A-G-C-T
^{32}P A-G-C
^{32}P A-G
^{32}P A

↓ ionophoresis homochromatography



the 5' terminus of each oligonucleotide of the partial digest, which is easily determined after a total snake venom phosphodiesterase digestion (Delaney, 1974, Murray and Old, 1974). This method can be applied to sequence determination of oligonucleotides of about 20 nucleotides and has been successfully used to determine the nucleotide sequence of some of the longer pyrimidine oligonucleotides of S13 su N15 DNA (Delaney, 1974, Delaney and Spencer, 1976).

3. OTHER Methods - Primer extension.

The specific priming approach takes advantage of the fact that repair synthesis of DNA by DNA polymerase I can be controlled by the dependence of the reaction upon the presence of a primer complementary to the template and having a terminal 3' OH. By providing a specific primer sequence which binds to the DNA template at a unique position, it is possible to ensure that *in vitro* DNA synthesis will start at a unique location. It is possible to use a variety of different primers to obtain synchronous initiation at practically any point desired. The primers can be synthetic or natural. Schaller and co-workers (1972) have used the pyrimidine oligonucleotide C_9T_{11} from bacteriophage fd plus strand as primer and they obtained sequence specific elongation of the primer. Sanger's group (1973) have used a synthetic oligonucleotide to determine a nucleotide sequence in phage f1 DNA. The extension of a chemically synthesized primer resulted in a sequence of 50 nucleotides. More recently, restriction enzyme fragments have been used also as primers (Sanger and Coulson, 1975).

a) Ribosubstitution.

In 1963, Berg *et al.*, showed that a ribonucleotide residue can be incorporated into DNA by the *E. coli* DNA polymerase I if the extension or the repair reaction is carried out in the presence of manganese ions instead of magnesium ions. The polynucleotide synthesized in the presence of [α-^{32}P]rCTP has a ribose linkage at every C residues and can therefore be cleaved specifically at these sites by alkaline hydrolysis or with pancreatic ribonuclease, which has a pyrimidine residue base specificity.

Ribosubstitution appeared to provide a promising method for specific cleavage of a mixed ribo and deoxyribonucleotide for sequence determination. Overlaps can be obtained by the use of different ribonucleotide substituents in separate reactions.

However, a disadvantage of this method is that manganese ions are known, in general, to promote misincorporation of nucleotides with nucleic acid-polymerizing enzymes (Hall and Lehman, 1968). However, Van de Sande *et al.* (1972) reported perfect fidelity of incorporation of rC if the extension is carried out at 10^{o}C. They also found that rG was misincorporated even at low temperature. Also, the incorporation was very slow especially within consecutive G residues. Incorporation of rA was very slow and rU was not incorporated at all. Thus, these results suggest that complete fidelity can be obtained only with rC at low temperatures.

b) The plus and minus method.

This very rapid method is based on the observation that in the

repair reaction of DNA polymerase I, if one of the four deoxyribonucleoside triphosphates is present at very low concentration, the extension reaction will stop when the enzyme has to incorporate the nucleoside triphosphate which is partly missing. Using this technique, Sanger's group recently published the complete sequence of bacteriophage ϕX174 DNA (Sanger *et al.*, 1977). As primers, they mostly used restriction enzyme fragments and synthetic oligonucleotides. Ideally the synthesis should be non-synchronous and as random as possible, so that the maximum number of oligonucleotides of different length, all starting from the primer, is formed. This is achieved by the balance of NTPs and sequential sampling. The mixture is then purified on an agarose column to remove the excess triphosphates and the samples are retreated in the minus system, or in the plus system.

The minus system

Wu and Kaiser (1968), in their work on the "sticky" ends of phage λ DNA, showed that if only three of the four deoxyribonucleoside triphosphate are present with the DNA polymerase, the polymerase will incorporate nucleotides accurately until it reaches the point when the missing nucleotide should be incorporated and at that point the reaction stops. The minus system is based on the same principle. The random mixture of oligonucleotides (cleaned on agarose), which is still hybridized to the template DNA is reincubated with DNA polymerase I in the presence of three deoxyribonucleoside triphosphates. The synthesis then proceeds as far on the template as the missing triphosphate starting from the point where incubation stopped in the first reaction. Thus, if dATP is the

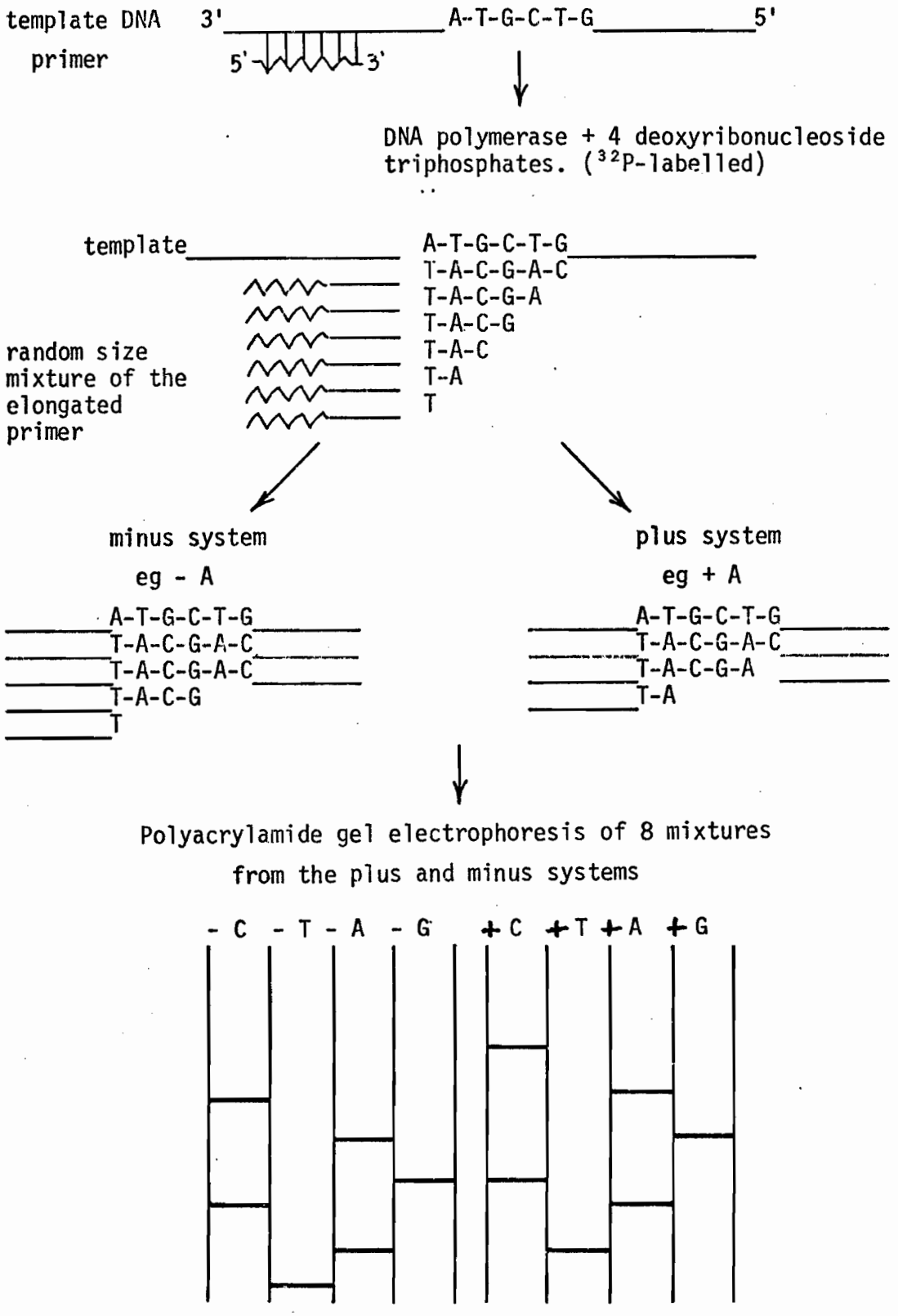
missing triphosphate (-A), each chain will terminate elongation at a position before an A residue. Separate samples are incubated, with a different one of the four triphosphates missing (Fig. 9).

The plus system

Englund (1972) showed that in the presence of only one of the four deoxyribonucleoside triphosphates, the 3' → 5' exonuclease activity of T4-DNA polymerase will degrade the randomly extended oligonucleotide until it reaches the nucleotide analogous to the one which is present in the incubation mixture. At that point, the polymerase activity is predominant over the exonuclease activity, and the one nucleotide will be replaced more quickly than it is removed, resulting in products which will all terminate at their 3' ends with the one nucleotide present in the incubation. Thus, if dATP is the triphosphate present (+A), each chain will terminate with an A residue (Fig. 9).

All the incubation mixtures from the plus and the minus system are then denatured to separate the extended primer from the template, subjected to electrophoresis on polyacrylamide gel in the presence of 8 M urea, and an autoradiograph prepared. In this fractionation system, mobility is essentially proportional to size, so that the various synthesized oligonucleotides will be arranged according to size. Ideally each oligonucleotide should be separated from its neighbour by one residue. The autoradiograph from the -A system will contain bands corresponding to positions before the A residues in the elongated chain. Thus, the positions of A residues are located. Moreover, the autoradiograph from the +A system will contain bands corresponding to positions

Figure 9 : The plus and minus systems as developed by Sanger and Coulson (1975).


template DNA 3'
A-T-G-C-T-G
5'
primer
5'
3'
DNA polymerase + 4 deoxyribonucleoside triphosphates. (^{32}P-labelled)
template
A-T-G-C-T-G
T-A-C-G-A-C
T-A-C-G-A
T-A-C-G
T-A-C
T-A
T
random size mixture of the elongated primer
minus system
eg - A
A-T-G-C-T-G
T-A-C-G-A-C
T-A-C-G-A-C
T-A-C-G
T
plus system
eg + A
A-T-G-C-T-G
T-A-C-G-A-C
T-A-C-G-A
T-A
Polyacrylamide gel electrophoresis of 8 mixtures from the plus and minus systems
- C
- T
- A
- G
+ C
+ T
+ A
+ G

of an A residue. Usually these will be in products one residue larger than the corresponding bands in the -A system, except if there is more than one consecutive A residue. In that case, the space between a -A system band and the corresponding +A system band will indicate how many consecutive A residues there are.

4. Purpose of this study.

The main objective of this study was to investigate the transcription of S13 DNA by determination of the RNA polymerase binding sites in bacteriophage S13 RF I DNA. By determining the localization and then mapping the RNA polymerase binding sites on the S13 genome, it was hoped that information on the extent of transcription into polycistronic mRNAs and how many different mRNAs were synthesized could be deduced. This information could indicate possible control mechanisms operating in S13 for gene expression. The final part of the study was sequence analysis of an RNA polymerase binding site to determine any sequence characteristics in the protein-DNA association.

At the start of the study, there was little information on the RNA polymerase binding sites. No sequence studies and very few characterization studies had been carried out on any DNAs. Slayter and Hall (1966) had visualized by electron microscopy bound RNA polymerase molecules on T7 DNA. Isolation and base composition studies of RNA polymerase binding site on bacteriophage λ (Nakano and Sakaguchi, 1968, Le Talaer and Jeanteur, 1971) and fd had been described (Heyden *et al.*, 1971), and the localization of three promoter regions in bacteriophage ϕX174 DNA (Chen *et al.*, 1973) had also been studied.

5. Experimental approach.

Bacteriophage S13 has nine different genes (Baker and Tessman 1967). Considering the number of RNA polymerase binding sites in other bacteriophages, it is unlikely that there would be a single, unique RNA polymerase binding site found in S13. It has been demonstrated by different techniques that most bacteriophage DNAs have more than one RNA polymerase binding site (Richardson, 1966, Hinkle and Chamberlin, 1972a, b). After confirmation that there was more than one RNA polymerase binding site on S13 DNA, we studied the arrangement of these sites in relation to the restriction enzyme cleavage map of S13 DNA which was in the process of construction in our laboratory (Grosveld *et al.*, 1976). The arrangement of RNA polymerase binding sites has also been done by electron microscopy. The two techniques gave similar results.

The same series of experiments has also been carried on ϕX174 for several reasons: bacteriophages ϕX174 and S13 are closely related: both contain single-stranded circular DNAs (Sinsheimer, 1959, Tessman, 1959), the same number and order of genes (Baker and Tessman, 1967, Benbow *et al.*, 1971) and moreover, the two phages recombine if mixed during infection (Tessman and Schleser, 1963). Studies on the two phage DNAs showed they were almost identical in size and physical characteristics (Sinsheimer, 1959, Spencer *et al.*, 1972). The pyrimidine patterns of the two DNAs were very similar and their sequences almost identical (Ling, 1972, Harbers *et al.*, 1976). The *Hin*d and *Hae* restriction enzyme cleavage maps of the two DNAs were very similar (Grosveld *et al.*, 1976, Godson and Roberts, 1976).

One of the major RNA polymerase binding sites was then selected

for sequence analysis. The sequencing approach is illustrated in Fig. 6 and 7. S13 RF DNA was degraded with restriction endonucleases *Hin*d II + III and the resulting fragments were separated according to size by polyacrylamide gel electrophoresis. Fragment 8, the fragment selected for analysis, was eluted from the gel and subjected to further degradation either by T4 endonuclease IV (Fig. 6) or by DNA methylation (Fig. 7) using the modification specifically adapted to sequencing double-stranded DNA fragments.

CHAPTER II

LOCALIZATION OF *ESCHERICHIA COLI* RNA POLYMERASE BINDING SITES ON S13 AND ϕX174 DNA: ALIGNEMENT WITH RESTRICTION ENZYME MAPS

ABSTRACT

E. coli RNA polymerase has been shown to bind to a limited number of *Hin*d and *Hae*III restriction enzyme fragments. On S13 RF DNA there are three major binding sites and the locations correlate with promoter sites at the beginning of genes A and B and a site overlapping gene D and the beginning of gene E. Two less definite binding sites have been localized, one in gene F and one at the gene G-H junction. In ϕX174 RF DNA five sites, each with apparently similar binding properties, have been located, four of which correspond exactly to binding sites in S13. One site, at the beginning of the B gene could not be assigned to exactly the same location found in S13. This was due in part to differences in the restriction cleavage maps in this area of the DNA and possibly to the higher background of non-specific binding in the ϕX174 experiments. The location of two of the ϕX174 sites at the beginning of gene A and D/E corresponds very well with transcription data but the site at the start of the B gene indicates the promoter site should be closer to the initiation sequence of the B protein than was previously suggested on the basis of transcription data.

INTRODUCTION

The single-stranded DNA genome of bacteriophage S13 is composed of 9 genes (Denhardt, 1975). Transcription occurs from the complementary strand of the replicative form of the DNA (Shleser *et al.*, 1969; Vanderbilt and Tessman, 1970). The RNA products synthesized *in vivo* vary in size from less than 59,000 daltons to more than 1.5×10^6 daltons (Puga and Tessman, 1973). The mRNA of 1.5×10^6 daltons represents nearly an entire genome equivalent of mRNA and from this it was concluded that the genome is transcribed in a poly-cistronic manner (Puga and Tessman, 1973). In a study of the functional stability of the mRNAs of seven of the nine genes of S13, Pollock and Tessman (1976) showed the half lives of the gene transcripts fell into five groups, B-C, D, F-G-H, A* and A. Polarity was demonstrated for the group F-G-H and indicated for B-C and the conclusion drawn that genes F, G and H have a polycistronic mRNA and B and C transcripts are polycistronic also. From these results one can predict promoters at the start of genes A, B, D and F. This prediction is supported in part by an earlier study of polarity gradients in S13 by Vanderbilt *et al.* (1972) which placed a promoter before gene A but not before G or H.

The initiation of transcription of the closely related phage ϕX174 DNA has been investigated by a number of methods. Hayashi's group (1970) isolated *in vivo* and *in vitro* mRNAs, observed a wide but reproducible distribution of sizes of molecules and concluded that the DNA was transcribed in a polycistronic manner with multiple initiation sites. Chen *et al.* (1973) studied the binding of *E. coli* RNA polymerase to restriction fragments and located potential promoter sites

on fragments which included the beginning of genes A, D and H. Smith and Sinsheimer (1976a, 1976b) identified three major *in vitro* transcripts of ϕX174 DNA and by hybridization to restriction fragments localized the initiation sites near the beginnings of genes A, B and D. Although two of the three restriction enzymes used by Smith and Sinsheimer (1976b) were the same as those used by Chen *et al.* (1973) the location of one promoter site was different in each report. Axelrod (1976a, 1976b) in an independent *in vitro* study similar to that of Smith and Sinsheimer (1976a, 1976b) also described three major transcripts and localized the initiation sites close to the beginning of the A, B and D genes. In a reinvestigation of *in vivo* transcripts Hayashi *et al.* (1976) localized initiation sites before gene B, gene C or D and probably before gene A in agreement with the *in vitro* data (Axelrod, 1976a, 1976b; Smith and Sinsheimer, 1976a, 1976b). From these data, promoters have been predicted at the start of genes A, B and D. The RNA polymerase binding to a site in the G gene may be a weak association due to sequence characteristics in this region (Sanger *et al.*, 1977).

The present study was undertaken to extend information on the initiation of transcription of S13 by determination of the distribution of *E. coli* RNA polymerase binding sites on the DNA. This approach was chosen since binding of RNA polymerase to a promoter site is the first step in transcription and studies on the *lac* operon and the *trp* operon have shown that this is a focal point at which cellular regulation of transcription takes place (Chamberlin, 1976). It does not necessarily follow that an RNA polymerase binding site is a promoter but *E. coli* RNA polymerase holoenzyme has been shown to have a high degree of specificity

for interaction and initiation of transcription on the DNAs of the filamentous phages and the strong RNA polymerase binding sites have been demonstrated to be congruent with sites for initiation of transcription (Heyden *et al.*, 1972; Okamoto *et al.*, 1972; Okamoto *et al.*, 1975; Schaller *et al.*, 1975; Seeburg and Schaller, 1975; Takanami *et al.*, 1976). We have paralleled the studies on S13 by similar analyses on ϕX174 to provide a direct comparison of the two closely related phage DNAs.

MATERIALS AND METHODS

Bacteriophage S13 wild type was kindly supplied by Drs. I. and E. Tessman and bacteriophage ϕX174am3 by Dr. D.T. Denhardt. Chloramphenicol, ethidium bromide and RNA, type II, were obtained from Sigma Chemical Co. Streptomycin sulphate was from Calbiochem and optical grade cesium chloride from the Harshaw Chemical Co. Purine nucleoside triphosphates were from Calbiochem or Terochem. Alberta, and $H_3{}^{32}PO_4$ was obtained from New England Nuclear. Pancreatic DNase I was from Worthington Biochemical Corporation. Reagents for electrophoresis, acrylamide and bis (N,N'-methylene-bis (acrylamide)) were obtained from Eastman. The acrylamide was recrystallized from chloroform before use. Collodion bags, type 100, and nitrocellulose filters, type B6 and BA-85, were from Schleicher and Schuell Inc., Sephadex G-100 was purchased from Pharmacia (Canada) Ltd.

Isolation of bacteriophage S13 and ϕX174 replicative form I DNA (RF I).

Growth was performed as described by Shleser *et al.* (1968). Isolation of the RF I DNA was by the procedure of Schekman *et al.* (1971) with the modifications described by Grosveld *et al.* (1976).

Restriction endonucleases and restriction enzyme digestions.

Endonuclease *Hin*d II + III (*Hin*d) was isolated from *Hemophilus influenzae* strain Rd and endonuclease *Hae*III from *Hemophilus aegyptius* ATCC III6. Both enzymes were generous gifts of Dr. F. Grosveld, and the isolations and reaction conditions have been described elsewhere (Grosveld *et al.*, 1976). The nomenclature for the restriction fragments is that described by Grosveld *et al.* (1976).

E. coli RNA Polymerase

E. coli RNA polymerase, isolated by the method of Burgess and Travers (1971) from *E. coli* K12 cells, was stored in the storage buffer recommended (Burgess and Travers, 1971) at a concentration of 7 mg/ml. The enzyme was a generous gift of Dr. F. Grosveld. Acrylamide gel electrophoresis of the enzyme preparation at the beginning and end of the study verified that sigma factor was present in equimolar proportions to the core enzyme (appendix I, Figure 1). The enzyme preparations also contained a protein component which migrated electrophoretically at the same position as S1 protein and which cross-reacted immunologically with S1 protein antibody. This immunological assay was kindly performed by Dr. A. Wahba (appendix I, Figure 2).

Carrier Nucleic Acids

Calf Thymus DNA was isolated by the procedure of Kay *et al.* (1952) and S13 viral DNA as described by Spencer and Boshkov (1973). Yeast tRNA was isolated from commercial yeast RNA (Type II, Sigma) by the procedure of Zubay (1966) modified to include two phenol extractions. The tRNA was a gift of Dr. F. Grosveld.

Isolation of *E. coli* RNA Polymerase Binding Sites from S13 RF I DNA.

E. coli RNA polymerase and ^{32}P-labeled S13 RF I DNA were incubated together in binding buffer (50 mM Tris-HCl buffer, pH 8.0 containing 10 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1 mM EDTA, 0.1 mM dithiothreitol, 0.1 mM ATP and 0.1 mM GTP) and either various KCl concentrations or at various RNA polymerase to DNA molar ratios dependent on the experiment. The mixtures were incubated for 10 min at 37°C then 200 µg/ml of DNase A prepared from DNase I (Junowicz and Spencer, 1973) (a gift of Dr. E. Junowicz)

was added. Incubation was continued for 40 minutes at 37°C then the mixture filtered through a nitrocellulose filter. Radioactivity retained on the filter, was measured by counting in a Toluene scintillation mixture. In some experiments, following the incubation with DNase A the mixture was extracted with phenol, the aqueous phase chromatographed on Sephadex G-100 and eluted with water. The DNA, protected from DNase A digestion by the bound RNA polymerase, eluted in the void volume and the radioactivity was measured by Cerenkov counting. Pyrimidine tract analyses of isolated binding fragments were as described by Harbers *et al.* (1976).

Binding of *E. coli* RNA Polymerase to Restriction Fragments of S13 RF I DNA.

Method A: 5 pmoles of S13 RF I DNA, digested with either *Hin*d or *Hae*III restriction enzymes, were incubated with 750 pmoles of *E. coli* RNA polymerase in binding buffer containing various concentrations of KCl for 10 minutes at 37°. Then 5 µg of heat denatured calf thymus DNA or 5 µg of single-stranded S13 DNA was added and the mixture incubated for a further 10 min. The addition of carrier DNA was to bind excess *E. coli* RNA polymerase and to compete out any non-specific low affinity binding of the *E. coli* RNA polymerase. The reaction mixtures were passed through a nitrocellulose filter (13 mm diameter), and the filter washed with binding buffer. The RNA polymerase-DNA complexes retained on the filter were eluted by two sequential extractions with 100 µl of 0.2% sodium dodecyl sulphate, the two extracts combined, 10 µg of tRNA added as carrier and the nucleic acid precipitated by addition of cold ethanol. The precipitate was resuspended in 25 µl of 0.1 M Tris-HCl buffer, pH 7.2 and the DNA fragments separated by polyacrylamide slab gel electrophoresis

(Grosveld *et al.*, 1976).

Method B: Binding experiments were also performed using the procedure of Okamoto *et al.* (1975). The *E. coli* RNA polymerase and DNA, were incubated at a molar ratio of 15:1 in a binding buffer of 40 mM Tris-HCl pH 7.9, 8 mM $MgCl_2$, 0.1 mM ATP and 0.1 mM GTP containing 50 mM KCl, passed through a nitrocellulose filter which was then washed with binding buffer containing increasing concentrations of KCl. The procedure of Okamoto *et al.* (1975) was modified by the inclusion of both ATP and GTP in the binding mixture, use of less DNA in each assay, and reduction of the total volume of the reaction mixture to 250 µl.

Quantitation of *E. coli* RNA Polymerase Binding to S13 RF DNA Restriction Fragments.

The ^{32}P labeled restriction fragment bands, separated by slab gel polyacrylamide electrophoresis, were located by autoradiography. The pieces of gel corresponding to the bands were cut out, digested in 50% H_2O_2, Brays scintillation fluid added and the radioactivity measured. Equivalent areas adjacent to the ^{32}P labeled bands were cut out and treated identically to allow correction for different backgrounds in the electrophoretic channels. To compare different experiments on different gels and reduce background effects due to different restriction fragment sizes, the assays were normalized by dividing the radioactivity present in each gel band by the number of base pairs present in that restriction fragment and dividing the quotient by the mean value of all the fragments in that particular assay. The normalized values for each assay had a mean value of 1.

RESULTS

The specificity of interaction of *E. coli* RNA polymerase holoenzyme with double stranded DNA is in part dependent upon the concentration of salt in the incubation mixture (Heyden *et al.*, 1972; Mangel and Chamberlin, 1974; Okamoto *et al.*, 1975; Chamberlin, 1976). The concentration of KCl required for optimal binding of holoenzyme to S13 RF I DNA was determined by measurement of the percentage of S13 DNA protected from DNase A digestion by bound polymerase at various KCl concentrations and the results are shown in Fig. 1. Increasing the molarity of KCl up to 0.25 M decreased the percentage of DNA protected, presumably due to reduction of non-specific binding of the polymerase which occurs at low salt concentration (Hinkle and Chamberlin, 1972; Mangel and Chamberlin, 1974; Chamberlin, 1976). Above this concentration, up to 0.5 M KCl, the amount of S13 DNA protected remained at approximately 2%. 0.35 M KCl was chosen as the optimal concentration for all further experiments. This result confirms the effect observed in other systems (Jones and Berg, 1966; Richardson, 1966; Hinkle and Chamberlin, 1972; Mangel and Chamberlin, 1974; Chamberlin, 1976) particularly the filamentous phages (Okamoto *et al.*, 1975).

To determine if the molar ratio of RNA polymerase to DNA affected the extent of binding of the polymerase to the DNA, protection of the S13 DNA from DNase A digestion, by *E. coli* RNA polymerase, was measured at RNA polymerase to DNA molar ratios of from 10 to 750. There was a slight increase in protection, with increasing ratios as shown in Fig. 2 but this was determined to be a non-specific protein effect. In the control experiment addition of increasing amounts of albumin, from 5 to

Fig. 1. Effect of KCl concentration on the binding of *E. coli* RNA polymerase to S13 RF DNA.

E. coli RNA polymerase was incubated with ^{32}P-labeled S13 RF DNA at a polymerase:DNA molar ratio of 150 in binding buffer (see Methods) containing various concentrations of KCl for 10 min at 37°C. The DNA not protected by the polymerase was hydrolyzed by digestion with DNase A for 40 min at 37°C. The resultant mixture was passed through a nitrocellulose filter and the DNA-RNA polymerase complexes retained on the filter, measured for ^{32}P radioactivity in Toluene scintillation fluid in a Beckman LS-250 scintillation spectrometer. The results are expressed as a percentage of the total ^{32}P radioactivity in the initial binding mixture. The two different symbols represent the results from two separate experiments.

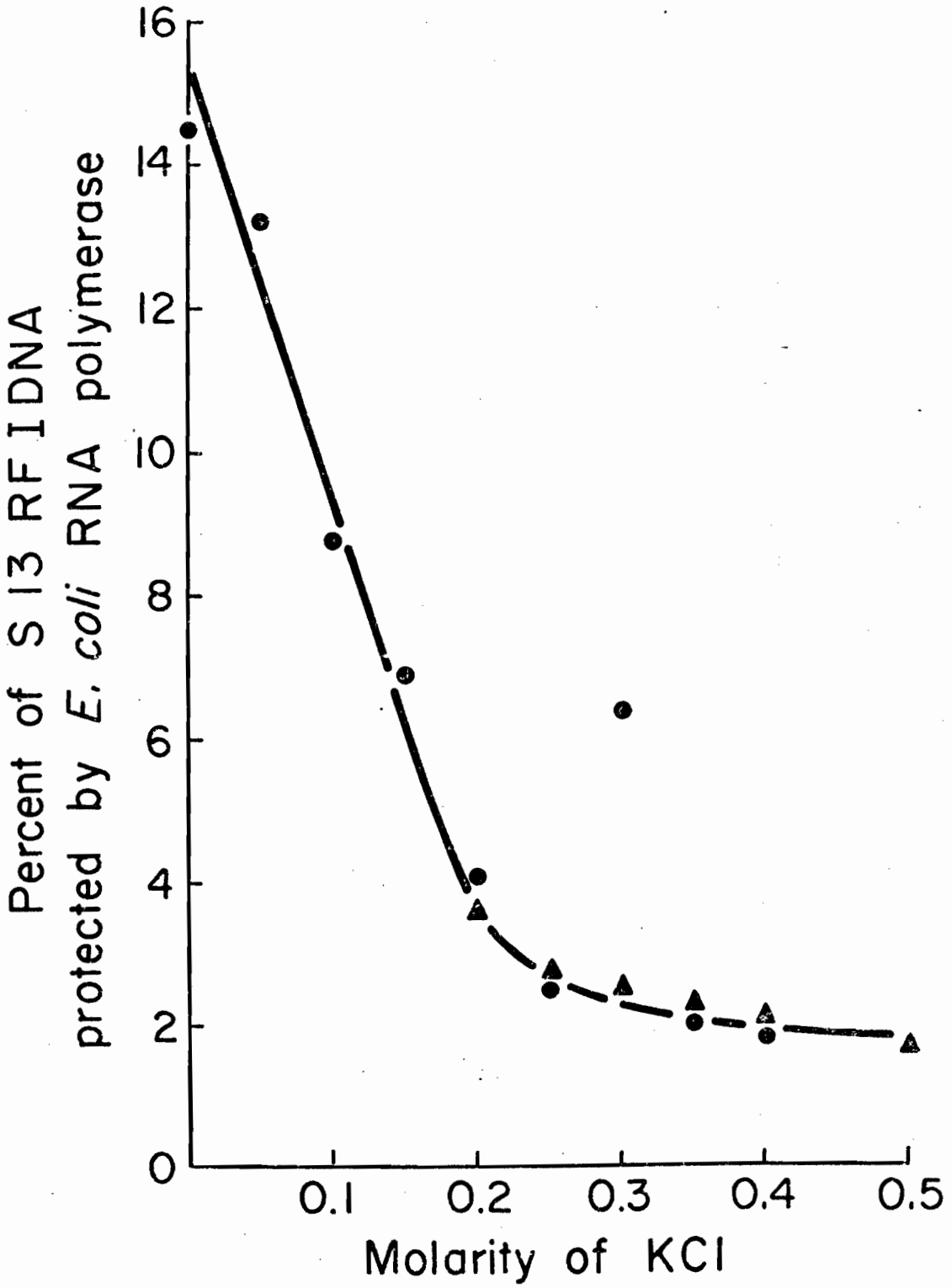
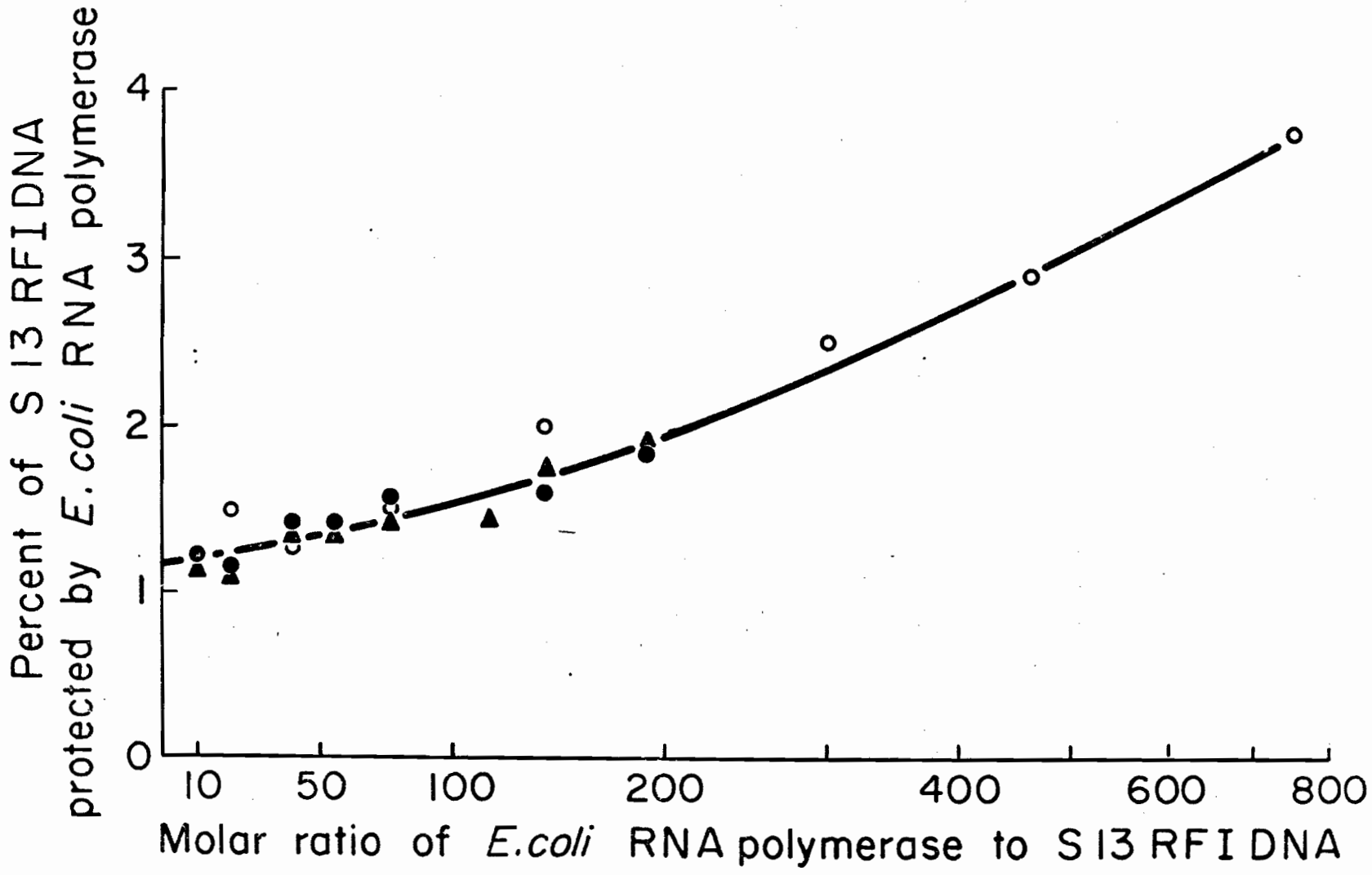

Percent of S 13 RF I DNA
protected by E. coli RNA polymerase
16
14
12
10
8
6
4
2
0
0.1
0.2
0.3
0.4
0.5
Molarity of KCl

Fig. 2. The effect of the molar ratio of *E. coli* RNA polymerase to S13 RF DNA on the extent of binding of the polymerase to S13 RF DNA.

E. coli RNA polymerase was incubated with ^{32}P-labeled S13 RF DNA at various molar ratios in binding buffer (see Methods) containing 0.35 M KCl, for 10 min at 37°C. The incubation mixture was then treated as described in the Legend to Fig. 1. The results are expressed as a percentage of the total ^{32}P radioactivity in the initial binding mixture. The three different symbols represent the results from three separate experiments.


Percent of S 13 RFI DNA protected by E. coli RNA polymerase
0
1
2
3
4
10
50
100
200
400
600
800
Molar ratio of E.coli RNA polymerase to S 13 RFI DNA

50 µg, to samples of RNA polymerase bound to S13 RF DNA at a molar ratio of 50, produced the same effect (appendix I, Figure 3). A molar ratio of 150 (polymerase:DNA) was chosen for all subsequent experiments with S13 DNA.

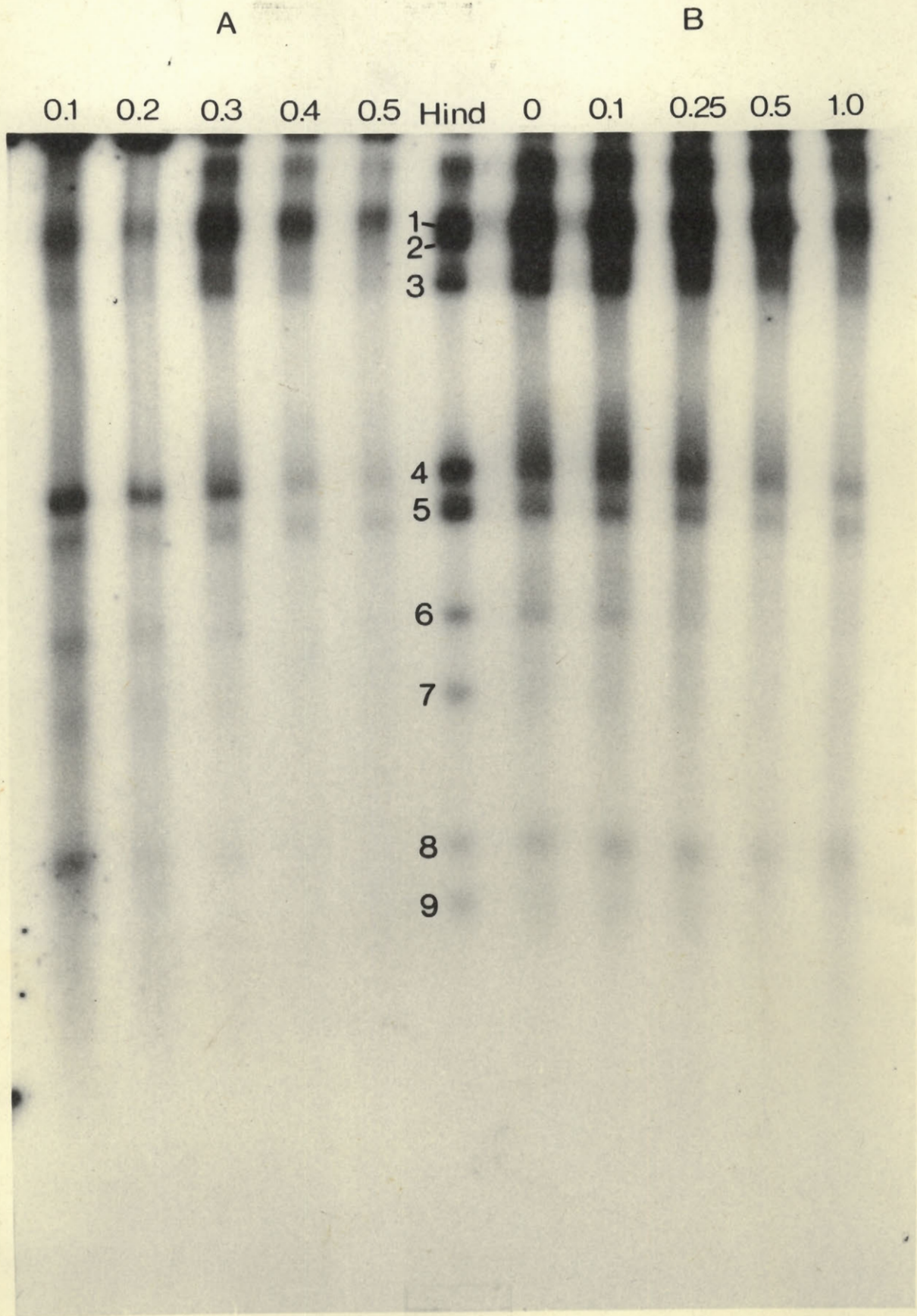
The size of the S13 RF I DNA fragments protected by *E. coli* RNA polymerase was measured by binding the polymerase and DNA at ratio of 150 in 0.35 M KCl, then incubating the mixture with DNase A. The solution was extracted with phenol and the polymerase protected fragments separated by polyacrylamide gel electrophoresis. S13 restriction fragments and oligonucleotides of known size were used as markers (Grosveld *et al.*, 1976). The polymerase protected fragments were in a narrow size range of 36 to 40 nucleotides, (appendix I, Figure 4), the same as found in other systems (Le Talaer and Jeanteur, 1971; Heyden *et al.*, 1972; Niyogi and Underwood, 1975; Pribnow, 1975a, 1975b; Schaller *et al.*, 1975; Sugimoto *et al.*, 1975; Walz and Pirrotta, 1975; Takanami *et al.*, 1976). Pyrimidine tract analysis of this fraction showed a restricted set of oligonucleotide fragments when compared to the oligonucleotide fragments from total S13 RF I DNA, indicating the DNA pieces protected by RNA polymerase were not random parts from the RF DNA (appendix I, Figure 5).

The experiments to determine the binding of polymerase to restriction fragments were designed so that the complexes formed corresponded to those that initiate transcription rapidly, (formed in low salt), and those that initiate transcription slowly (formed in high salt 0.2 M KCl) (Mangel and Chamberlin, 1974). The variations in ionic conditions also allowed discrimination between non specific binding and stable association and covered a spectrum of concentrations, in which protein aggregation was

Fig. 3. Autoradiograph of *Hin*d restriction fragments of ^{32}P labeled S13 RF DNA which bind *E. coli* RNA polymerase, separated on a polyacrylamide slab gel.

A. The *Hin*d digest of ^{32}P labeled S13 RF DNA was incubated with *E. coli* RNA polymerase at a molar ratio of 150 in binding buffer containing the concentrations of KCl indicated above each channel. The DNA fragment-RNA polymerase complexes were trapped on nitrocellulose filters and washed with binding buffer containing the appropriate concentration of KCl. The complexes were then eluted with 0.2% sodium dodecyl sulphate precipitated with cold ethanol, redissolved and separated by polyacrylamide gel electrophoresis.

B. The *Hin*d digest of ^{32}P labeled S13 RF DNA was incubated with *E. coli* RNA polymerase at a molar ratio of 15 in the binding buffer of Okamoto *et al.* (1975) containing 50 mM KCl. The DNA fragment-RNA polymerase complexes were trapped on nitrocellulose filters and then washed with binding buffer containing the KCl concentrations indicated above each channel. The complexes were then eluted and processed as in A and separated by polyacrylamide gel electrophoresis. The center channel labeled *Hin*d is a control of a total *Hin*d digest of ^{32}P labeled S13 RF DNA for use as a marker. The numbers are the *Hin*d fragment numbers (Grosveld *et al.*, 1976).


A
B
0.1
0.2
0.3
0.4
0.5
Hind
0
0.1
0.25
0.5
1.0
1
2
3
4
5
6
7
8
9

minimized (Hinkle and Chamberlin, 1972). Both purine triphosphates were present in the binding buffer and no attempt was made to differentiate ATP dependent sites from GTP dependent sites.

Localization of the binding of RNA polymerase to *Hin*d restriction fragments of S13 RF I DNA is shown in Fig. 3. In Fig. 3A, in which the restriction fragments and polymerase were bound and washed in the same KCl concentration as indicated in the Figure, stable complexes formed with fragments *Hin*d 1, *Hin*d 4, and to some extent with *Hin*d 5, *Hin*d 6 and *Hin*d 8. The conclusion is based on the presence of these fragments in experiments at all KCl concentrations. In Fig. 3B, in which the fragments and polymerase were bound in low salt and then washed with increasing concentrations of KCl (Okamoto *et al.*, 1975) as indicated, most of the restriction fragments bound polymerase at low KCl concentrations, but as the salt concentration was increased specific binding, and formation of stable complexes, could be discerned (Okamoto *et al.*, 1975). The results are essentially the same as in Fig. 3A: *Hin*d 1 and *Hin*d 4 bind polymerase although the *Hin*d 4 binding is not as clear as in A, some binding of *Hin*d 5 can be observed, there is some indication of binding of *Hin*d 6 and the binding of *Hin*d 8 is very distinct. There is also some binding of *Hin*d 2 and 3. Quantitation of a number of similar experiments with *Hin*d fragments is presented in Table 1. The data show clearly specific binding of *Hin*d 1 and *Hin*d 8. The binding of *Hin*d 4 is apparent but since there are multiple fragments of the size of *Hin*d 4 in S13 DNA it is possible that one or perhaps two only are binding. There is some binding of *Hin*d 6, *Hin*d 2 and *Hin*d 3 but it appears to be less specific. There are three *Hin*d 5 sized fragments in S13 and the binding observed in the radioautograph

(Fig. 3) is probably cumulative background which is normalized in Table 1. In the case of *Hin*d 2 the binding observed may be due in part to overlap with *Hin*d 1, and for both *Hin*d 2 and 3 the background is relatively high since they are large restriction fragments.

A similar experiment using *Hae*III fragments of S13 RF DNA for localizing RNA polymerase binding is shown in Fig. 4. In Fig. 4A, there is definite binding of *Hae*III 1, *Hae*III 2, *Hae*III 3 and *Hae*III 6 and this is confirmed by the quantitative data in Table 2.

The result is not as definite in Fig. 4B; *Hae*III 1, *Hae*III 2, *Hae*III 3 and *Hae*III 6 bind strongly and there is also an indication of binding of *Hae*III 5. However, the quantitative data from multiple experiments presented in Table 2 indicate that the *Hae*III 5 binding is probably due to cumulative background binding of the two *Hae*III 5 sized fragments present in S13. Additional experiments have been performed on S13 (appendix I, Figure 6 and Table I).

Data from analogous binding experiments using ϕX174 RF I DNA, *Hin*d II and *Hae*III restriction fragments are presented in Tables 3 and 4. The experiments were modified slightly compared to the S13 studies in that the higher concentrations of KCl were not used in the Method A experiments and the 0 M KCl wash was not used in the Method B experiments. Specific binding of *Hin*d 1, *Hin*d 2, *Hin*d 4 and *Hin*d 5 and some binding of *Hin*d 6 is apparent. Multiple fragments the size of *Hin*d 6 occur in ϕX174 thus only one or two fragments per genome may be binding. In the *Hae*III experiment (Table 4) binding of *Hae*III 1, *Hae*III 2, *Hae*III 3, *Hae*III 4 and *Hae*III 7 was found.

Although a molar ratio of *E. coli* RNA polymerase:DNA of 150 was

TABLE 1

Binding of *E. coli* RNA Polymerase to the *Hin*d Restriction Fragments of S13 RF I DNA

	Method A						Method B				
	Normalized Values for Binding of *E. coli* RNA Polymerase										
*Hin*d agment No.	0.1 M KCl	0.2 M KCl	0.3 M KCl	0.4 M KCl	0.5 M KCl	*Hin*d Control	0 M KCl	0.1 M KCl	0.25 M KCl	0.5 M KCl	1.0 M KCl
1	0.5 ± 0.1	0.9 ± 0.1	3.7 ± 0.1	4.1 ± 0.1	4.5 ± 0.5	1.1 ± 0.1	4.0 ± 0.5	4.6 ± 0.5	4.3 ± 0.5	4.3 ± 0.5	2.3 ± 0.2
2	0.2 ± 0	0.2 ± 0	1.1 ± 0	1.4 ± 0	1.5 ± 0.3	1.0 ± 0.1	1.4 ± 0.2	0.7 ± 0.2	1.6 ± 0.3	1.0 ± 0.3	1.1 ± 0
3	0.1 ± 0	0.2 ± 0	0.7 ± 0	0.7 ± 0	0.7 ± 0.2	0.9 ± 0.1	1.2 ± 0.2	1.1 ± 0.2	0.8 ± 0.2	1.0 ± 0.2	0.5 ± 0
4	0.5 ± 0	1.2 ± 0.2	0.8 ± 0	0.8 ± 0.2	1.0 ± 0.3	1.0 ± 0.1	0.7 ± 0.1	0.7 ± 0.1	0.6 ± 0.2	0.4 ± 0.1	0.3 ± 0
5	0.1 ± 0	0.2 ± 0	0.2 ± 0	0.3 ± 0	0.4 ± 0	1.0 ± 0.1	0.4 ± 0	0.4 ± 0	0.3 ± 0.1	0.3 ± 0	0.2 ± 0
6	0.8 ± 0	1.1 ± 0	0.4 ± 0	0.3 ± 0.1	0.4 ± 0.1	1.0 ± 0.1	0.5 ± 0.1	0.6 ± 0.1	0.2 ± 0.1	0	0.7 ± 0.1
7	1.1 ± 0	0.5 ± 0.2	0.3 ± 0	0	0	1.0 ± 0	0	0	0	0	0.6 ± 0
8	4.4 ± 0.4	3.5 ± 0.3	1.9 ± 0.4	1.5 ± 0.3	0.5 ± 0.1	1.0 ± 0.1	0.9 ± 0.1	0.9 ± 0.1	1.3 ± 0.1	2.1 ± 0.2	3.0 ± 0
9	1.3 ± 0.2	1.3 ± 0.2	0	0	0	1.0 ± 0	0	0	0	0	0.3 ± 0.1

Values are the mean of three experiments ± the arithmetic error of the mean

TABLE 2

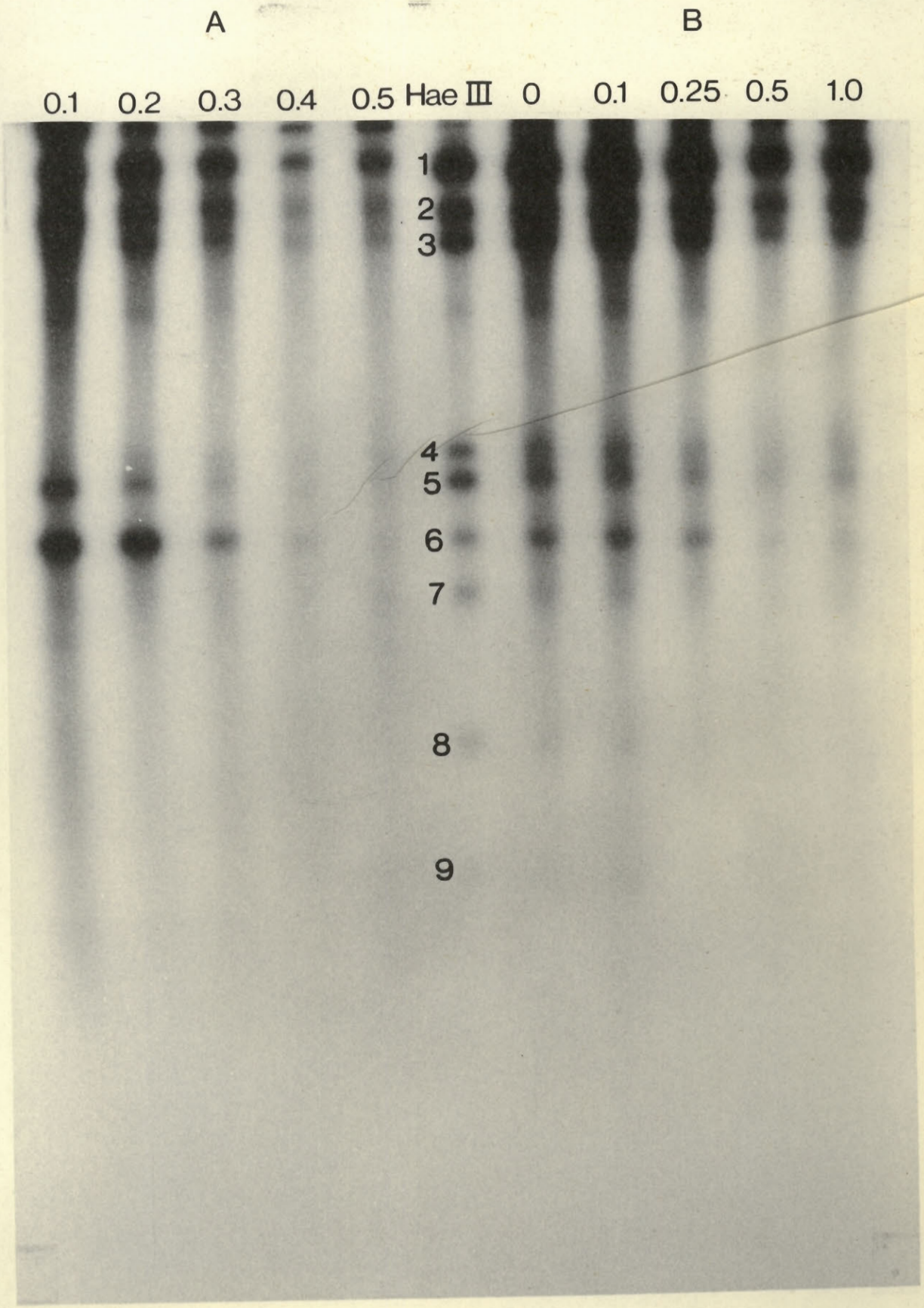
Binding of *E. coli* RNA Polymerase to the *Hae*III Restriction Fragments of S13 RF I DNA

	Method A						Method B				
	Normalized Values for Binding of *E. coli* RNA Polymerase										
*Hae*III Fragment No.	0.1 M KCl	0.2 M KCl	0.3 M KCl	0.4 M KCl	0.5 M KCl	*Hae*III Control	0 M KCl	0.1 M KCl	0.25 M KCl	0.5 M KCl	1.0 M KCl
1	1.7 ± 0	1.6 ± 0.1	2.2 ± 0.1	2.3 ± 0.4	2.2 ± 0.5	1.0 ± 0.1	3.0 ± 0.3	3.0 ± 0.2	3.5 ± 0.3	3.3 ± 0.3	3.6 ± 0.5
2	2.2 ± 0.1	1.6 ± 0.2	2.2 ± 0.1	2.3 ± 0.4	1.6 ± 0.4	1.1 ± 0.1	2.5 ± 0.3	2.5 ± 0.2	2.6 ± 0.2	1.6 ± 0.3	2.7 ± 0.2
3	1.7 ± 0.4	1.4 ± 0.3	1.1 ± 0.1	1.3 ± 0.3	1.4 ± 0.3	1.0 ± 0.1	2.3 ± 0.3	2.3 ± 0.2	2.0 ± 0.2	1.8 ± 0.2	1.7 ± 0.2
4	0	0.3 ± 0.1	0.1 ± 0	0.5 ± 0.1	0.5 ± 0.1	0.9 ± 0.1	0.3 ± 0.1	0.4 ± 0.2	0.2 ± 0.1	0.1 ± 0	0
5	0.4 ± 0	0.3 ± 0.1	0.1 ± 0	0.5 ± 0.1	0.4 ± 0.1	1.1 ± 0.1	0.2 ± 0.1	0.2 ± 0	0.1 ± 0	0.2 ± 0	0.3 ± 0
6	2.6 ± 0.1	3.7 ± 0.2	2.5 ± 0	1.5 ± 0.3	1.4 ± 0.2	1.0 ± 0.1	0.6 ± 0.1	0.6 ± 0.1	0.5 ± 0.1	0.5 ± 0.1	0.5 ± 0.1
7	0.1 ± 0	0.4 ± 0.1	0.4 ± 0	0.2 ± 0.1	0.7 ± 0.2	1.1 ± 0.1	0	0	0	0	0.2 ± 0
8	0.3 ± 0	0	0.4 ± 0	0.5 ± 0.2	0.7 ± 0.1	0.9 ± 0	0	0.1 ± 0	0.1 ± 0	0.4 ± 0	0
9	0	0	0	0.1 ± 0.1	0.3 ± 0.1	0.9 ± 0	0	0	0	0.1 ± 0	0

Values are the mean of three experiments ± the arithmetic error

Fig. 4. Autoradiograph of *Hae*III restriction fragments of ^{32}P labeled S13 RF DNA which bind *E. coli* RNA polymerase, separated on a polyacrylamide slab gel.

Conditions of incubation and analysis as in the legend to Fig. 3.


A
B
0.1
0.2
0.3
0.4
0.5
Hae III
0
0.1
0.25
0.5
1.0
1
2
3
4
5
6
7
8
9

chosen for the binding studies in Method A much lower ratios would have been acceptable. A series of experiments (results not shown) using Method A with polymerase: S13 or ϕX174 DNA at a ratio of 50 gave the same results as those at a ratio of 150. From this it is concluded that the binding observed was binary complex formation and aggregation of RNA polymerase was not a factor in the present experiments.

TABLE 3

Binding of *E. coli* RNA Polymerase to the *Hind* II Restriction Fragments of ϕX174 RF I DNA

	Method A				Method B			
	Normalized Values for Binding of *E. coli* RNA Polymerase							
Hind II Fragment No.	0.1 M KCl	0.2 M KCl	0.3 M KCl	*Hind* Control	0.1 M KCl	0.25 M KCl	0.5 M KCl	1.0 M KCl
1	2.0 ± 0.5	2.0 ± 0.5	2.2 ± 0.3	1.1 ± 0.1	2.2 ± 0.1	2.0 ± 0.2	2.1 ± 0	2.5 ± 0.3
2	1.8 ± 0.1	1.9 ± 0.3	1.8 ± 0.2	1.2 ± 0.1	1.9 ± 0	2.0 ± 0.2	1.8 ± 0	1.9 ± 0
3	0.8 ± 0.3	0.4 ± 0	0.5 ± 0.1	1.0 ± 0	0.6 ± 0.3	0.7 ± 0.3	0.5 ± 0.2	0.5 ± 0.2
4	1.8 ± 0.3	1.8 ± 0.2	1.7 ± 0.3	0.9 ± 0	1.8 ± 0.3	1.8 ± 0.3	1.7 ± 0.3	1.8 ± 0.3
5	1.2 ± 0	1.3 ± 0.1	1.2 ± 0.1	1.0 ± 0.1	1.6 ± 0.3	1.6 ± 0.3	1.5 ± 0.3	1.3 ± 0
6	0.9 ± 0.1	0.9 ± 0	0.9 ± 0.1	1.0 ± 0	0.9 ± 0.1	1.0 ± 0.1	0.9 ± 0	0.8 ± 0
7	0.4 ± 0.1	0.4 ± 0	0.4 ± 0	1.0 ± 0	0.2 ± 0.1	0.2 ± 0	0.4 ± 0.1	0.4 ± 0
8	0.4 ± 0	0.4 ± 0.1	0.6 ± 0	1.0 ± 0	0.6 ± 0.1	0.5 ± 0	0.5 ± 0.1	0.5 ± 0
9	0.3 ± 0	0.5 ± 0	0.4 ± 0	1.0 ± 0	0.3 ± 0	0.2 ± 0.1	0.2 ± 0	0.1 ± 0.1
10	0.2 ± 0	0.5 ± 0.1	0.4 ± 0	0.8 ± 0	0.2 ± 0	0.1 ± 0	0.3 ± 0	0.2 ± 0

Values are the mean of two experiments ± the arithmetic error.

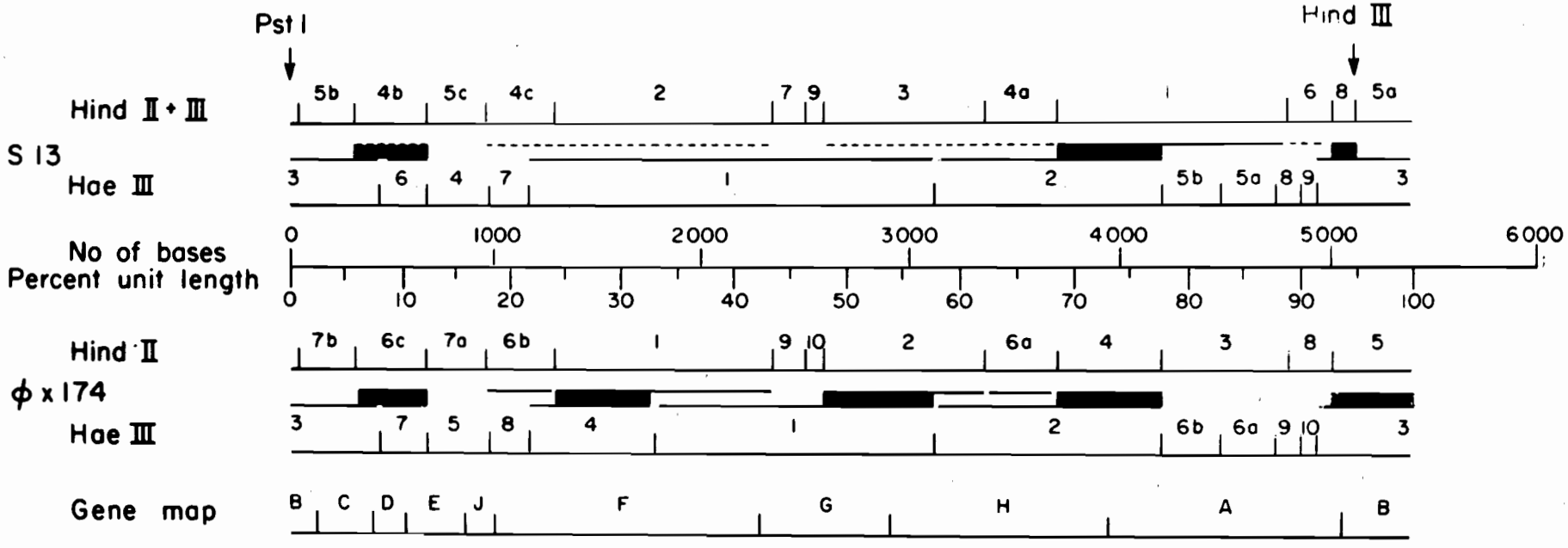
TABLE 4

Binding of *E. coli* RNA Polymerase to the *Hae*III Restriction Fragments of ϕX174 RF I DNA

	Method A				Method B			
	Normalized Values for Binding of *E. coli* RNA Polymerase							
*Hae*III Fragment No.	0.1 M KCl	0.2 M KCl	0.3 M KCl	*Hae*III Control	0.1 M KCl	0.25 M KCl	0.5 M KCl	1.0 M KCl
1	3.1 ± 0.1	2.6 ± 0.3	2.5 ± 0	1.2 ± 0	2.2 ± 0.4	2.2 ± 0.5	2.0 ± 0.5	2.2 ± 0.1
2	2.0 ± 0.4	2.0 ± 0	2.2 ± 0.3	1.1 ± 0	1.8 ± 0.3	2.1 ± 0.3	2.1 ± 0.2	2.0 ± 0.2
3	1.5 ± 0.1	1.5 ± 0.5	1.6 ± 0.3	0.8 ± 0.1	1.9 ± 0.1	1.9 ± 0.2	2.0 ± 0.2	1.6 ± 0
4	1.1 ± 0	1.5 ± 0.5	1.4 ± 0.1	0.9 ± 0	1.6 ± 0.4	1.7 ± 0.3	1.5 ± 0.1	1.2 ± 0.1
5	0.3 ± 0.2	0.4 ± 0.1	0.2 ± 0.2	1.2 ± 0.1	0.4 ± 0.1	0.3 ± 0.1	0.3 ± 0.1	0.4 ± 0
6	0.4 ± 0	0.3 ± 0	0.1 ± 0	0.9 ± 0.1	0.3 ± 0	0.3 ± 0.1	0.2 ± 0.1	0.3 ± 0.1
7	0.8 ± 0.1	0.9 ± 0	1.0 ± 0.3	1.1 ± 0.1	0.8 ± 0	0.8 ± 0.1	1.3 ± 0.2	1.2 ± 0.1
8	0.3 ± 0.1	0.5 ± 0	0.6 ± 0	1.0 ± 0.1	0.4 ± 0	0.3 ± 0	0.3 ± 0	0.4 ± 0
9	0.2 ± 0	0.2 ± 0.1	0.3 ± 0	0.9 ± 0.1	0.3 ± 0	0.3 ± 0.1	0.3 ± 0	0.4 ± 0.1
10	0.1 ± 0.1	0.2 ± 0.1	0.1 ± 0	0.9 ± 0	0.2 ± 0.1	0.1 ± 0.1	0 ± 0	0.2 ± 0.1

Values are the mean of two experiments ± the arithmetic error.

Fig. 5. Schematic presentation of the *E. coli* RNA polymerase binding sites of S13 and ϕX174 RF I DNAs in relation to the *Hin*d and *Hae*III restriction endonuclease cleavage maps and the genetic map. The S13 restriction endonuclease cleavage maps are from Grosveld *et al.* (1976) as amended by Goodchild and Spencer (1978) and those for ϕX174 from Sanger *et al.* (1977). Solid lines correspond to fragments which bound RNA polymerase clearly. Dashed lines (in the S13 frame) correspond to fragments which bound RNA polymerase less definitively. The shaded areas are the delineated binding (putative promoter) sites.


Pst I
Hind III
Hind II + III
5b 4b 5c 4c 2 7 9 3 4a 1 6 8 5a
S 13
Hae III
3 6 4 7 1 2 5b 5a 8 9 3
No of bases
0 1000 2000 3000 4000 5000 6000
Percent unit length
0 10 20 30 40 50 60 70 80 90 100
Hind II
7b 6c 7a 6b 1 9 10 2 6a 4 3 8 5
ϕ x 174
Hae III
3 7 5 8 4 1 2 6b 6a 9 10 3
Gene map
B C D E J F G H A B

DISCUSSION

The size range 35 to 40 nucleotides and the amount, approximately 2%, of the S13 DNA genome protected by *E. coli* RNA polymerase (Fig. 1, Fig. 2) from DNase hydrolysis corresponds to two to three moles of polymerase bound per S13 DNA molecule. This is the number of specific binding sites observed in the restriction fragment binding experiments.

The localization on S13 and ϕX174 RF DNAs of the *E. coli* RNA polymerase binding sites relative to the *Hin*d and *Hae*III restriction endonuclease cleavage sites and correlation with the genetic map is presented schematically in Fig. 5. The schematic was constructed by comparing the relative binding of RNA polymerase to the restriction fragments deduced from the quantitative data in Tables 1 to 4.

The binding of S13 *Hin*d 1 (Table 1) and S13 *Hae*III 2 (Table 2) places a binding site at the overlap region of these two fragments. S13 *Hae*III fragments 5b, 5a and 8 which overlap S13 *Hin*d 1, did not bind polymerase, eliminating this area from the binding site. To determine if S13 *Hin*d 4a was a possible binding area an S13 *Hin*d/*Hae*III total cross digest was incubated with *E. coli* RNA polymerase (results not shown). *Hin*d 4b and 4c are cleaved by endonuclease *Hae*III (Grosveld *et al.*, 1976) and there was no binding to S13 *Hin*d 4a. S13 *Hin*d 3 showed some evidence of polymerase binding (Table 1). The area of binding has not been delineated and may overlap with *Hae*III 2 but this would be a separate area from the *Hin*d 1-*Hae*III 2 site. The *Hin*d 1-*Hae*III 2 overlap coincides with the start of gene A.

The binding of *E. coli* RNA polymerase to S13 *Hin*d 8 (Table 1) and to S13 *Hae*III 3 (Table 2) localizes a binding site at the overlap of S13

*Hin*d 8-*Hae*III 3. The low level of binding observed with S13 *Hin*d 6 (Table 1) may be part of this binding site since it overlaps with S13 *Hae*III 3. This binding site coincides with the beginning of gene B.

Restriction fragment S13 *Hae*III 3 also overlaps with S13 *Hin*d 4b and is adjacent to S13 *Hae*III 6 which was observed to bind clearly to *E. coli* RNA polymerase (Table 2). The quantitative date in Table 1 indicates that only one *Hin*d 4 fragment is binding polymerase, thus *Hin*d 4b is the most likely fragment to be binding, placing a binding site in the area of genes D and E at the overlap of fragments *Hae*III 3 and 6-*Hin*d 4b.

The binding of S13 *Hae*III 1 is somewhat of an enigma. It could be due to a binding site which overlaps with *Hin*d 4c, but this area is well inside gene F. Some binding was observed with fragments *Hin*d 2 and 3 and the *Hae*III 1 binding may be partly a function of high background, since it is a large fragment, plus less specific binding to areas within the overlaps of *Hin*d 2 and 3 which have structural or sequence similarities to promoter regions.

The location of the three RNA polymerase binding sites, one each at the beginning of genes A and B and one which includes gene D and the beginning of E coincides with the polarity studies of the mRNA's of S13 (Vanderbilt *et al.*, 1972; Pollock and Tessman, 1976), and indicates the sites as putative promoters. This assumes that genes D, E, J, F, G and H are polycistronic.

In ϕX174 the results with both sets of restriction fragments showed a higher background, presumably due to non specific binding of polymerase. Also similar levels of binding of polymerase to a number of fragments was observed rather than a few fragments exhibiting significant levels

of distinct binding, as in S13. Consequently the identification of *E. coli* RNA polymerase binding locations is not as definitive in ϕX174 as in S13.

The binding of ϕX174-*Hin*d II 4 (Table 3) overlaps with the binding to *Hae*III 2 (Table 4) localizing a binding site at the overlap of the two fragments. As in S13 the adjacent *Hin*d II fragment (6a in this case) is one of three similar sized fragments released by *Hin*d II, only one of which binds polymerase (Table 3). Although a ϕX174 *Hin*d II/*Hae*III cross digest-binding experiment was not performed, the *Hin*d II 6a fragment is not included in the binding site since another *Hin*d II 6 fragment has been included in another binding site (see below). ϕX174 *Hin*d II 2 binds RNA polymerase and the overlap with *Hae*III 2 cannot be excluded as part of a possible binding site (see below) but we postulate that this is separate from the *Hin*d II 4-*Hae*III 2 sites. The ϕX174 *Hin*d II 4-*Hae*III 2 overlap coincides exactly with the S13 *Hin*d 1-*Hae*III 2 site at the beginning of gene A and confirms one of the sites localized by Chen *et al.* (1973). It also correlates exactly with the location derived from hybridization of ϕX174 transcripts to restriction fragments (Axelrod, 1976b; Hayashi *et al.*, 1976; Smith and Sinsheimer, 1976b).

The binding of polymerase to *Hin*d II 5 (Table 3) and *Hae*III 3 (Table 4) locates another polymerase binding site on ϕX174 DNA. However we observed no binding to *Hin*d II 8 which would extend the binding site slightly and coincide with the S13 data. The apparent larger size of this ϕX174 binding site compared to the corresponding site in S13 is due to the absence of a *Hin*d III cleavage site in ϕX174 DNA in this area. The location of this second site coincides with the start of gene B.

Chen *et al.* (1973) did not observe any significant RNA polymerase binding to *Hin*d II 5 and consequently did not predict a site at this position. However they did observe binding to *Hae*III 3. The discrepancies between our observations and those of Chen *et al.* (1973) may be due to the different RNA polymerase:DNA ratios and ionic conditions. Chen *et al.* (1973) used very high ratios and low salt (50 mM KCl) which would favour aggregation of the polymerase molecules, and accentuate non specific cooperative binding of the ligand to give a high background.

Axelrod (1976b) in one experiment examined transcription initiation under similar conditions to those of Chen *et al.* (1973) and observed suppression of initiation at the promoter site at *Hin*d II 8-*Hae*III 3.

The *in vitro* mRNA hybridization studies of ϕX174 (Axelrod, 1976b; Smith and Sinsheimer, 1976b) placed the B gene site at *Hin*d II 8-*Hae*III 3. There is no obvious explanation for this difference in site location. The high background in the present studies may have obscured binding to *Hin*d II 8 which would then include the same site area.

The location of the B gene promoter of ϕX174 is further complicated by the sequence studies (Sanger *et al.*, 1977) which placed the putative mRNA start of gene B at residue 4888 which is in *Hae*III 10 and is 176 nucleotides from the initiation of translation sequence (Sanger *et al.*, 1977) which is in fragment *Hae*III 3. The location in *Hae*III 10 was assigned on the basis of correlative sequence data of five bases at the 5' end of one group of ϕX174 transcripts (Smith and Sinsheimer, 1976c). The *Hae*III 10 location conflicts with all other data, moreover the transcripts did not hybridize to the ϕX174 *Hae*III 10 fragment (Axelrod, 1976b; Smith and Sinsheimer, 1976b). Clarification of the location of the

ϕX174 B gene promoter will require reinvestigation of the 5' termini of the B gene transcripts.

The present study indicates the B gene promoter site in S13 should be close to the initiation sequence of the B protein. Fragments S13 *Hae*III 8 and 9 have been sequenced but very few differences from the corresponding ϕX174 sequence were found (Grosveld and Spencer, 1977). The extent of the differences between the two phage DNA's in this region and whether the B start of S13 is later than in ϕX174, which has been suggested as an explanation of complex genetic data (Tessman *et al.*, 1976) must await the sequence of S13 *Hin*d 6 and 8, and correlation with additional transcription data.

The fragment ϕX174 *Hae*III 3 is adjacent to *Hae*III 7 which also bound polymerase (Table 4) and both fragments overlap with *Hin*d II 6c. As previously noted polymerase binds to one only of the three similar sized *Hin*d II 6 fragments. We have selected *Hin*d II 6c as the most likely candidate, since it is the only fragment to overlap with *Hae*III 7. This selection places a polymerase binding site at this location corresponding to the identical site in S13 covering genes D and E. Chen *et al.* (1973) located a binding site at the overlap of *Hin*d II 6c and *Hae*III 3 but did not observe any significant binding of *E. coli* RNA polymerase to *Hae*III 7. The site localized by Smith and Sinsheimer (1976b) by transcript hybridization did not include the *Hae*III 7 fragment either. However, Axelrod (1976b) did observe binding of GTP initiated, 5' proximally labeled mRNA to *Hae*III 7, correlating well with the present observations. The difference in the mRNA hybridization studies (Axelrod, 1976b; Smith and Sinsheimer, 1976b) in this region may be a function of the size of

the transcripts used for hybridization.

The remaining two binding sites in ϕX174 are located on the basis of less stringent data. The polymerase binding to *Hin*d II 1 and *Hae*III 4 delineate an area of overlap and a possible binding site. This conclusion is partially dependent on the previous selection of *Hin*d 6c as the single *Hin*d II 6 site binding polymerase since *Hin*d II 6b overlaps *Hae*III 4 also. *Hae*III 1 which is adjacent to *Hae*III 4 also binds polymerase but is not included in this overlap since *Hin*d II 2 also binds polymerase providing an overlap with *Hae*III 1. *Hin*d II 2 also overlaps *Hae*III 2, another fragment which binds polymerase, but the *Hae*III 2 binding has been assigned to the overlap with *Hin*d III 4. However, *Hae*III 2 could be involved in two binding sites overlapped with both *Hin*d II 2 and *Hin*d II 4. The *Hin*d II 1-*Hae*III 4 site is similar in location to a relatively poor binding site in S13 which may have a type of non-specific binding. The same is true of the *Hin*d II 2-*Hae*III 1 site. This latter site which is at the gene G-H junction was also observed by Chen *et al.* (1973). However they did not locate a site corresponding to the *Hin*d II 1-*Hae*III 4 site in the middle of the F gene. It has been suggested that the *Hin*d II 2-*Hae*III 1 site in ϕX174 may be the result of sequence or structural characteristics sufficiently similar to parts of promoters to result in some stable binding of RNA polymerase (Sanger *et al.*, 1977).

ACKNOWLEDGEMENTS

We thank Dr. F.G. Grosveld for generous supplies of enzymes and helpful discussions. The work was supported by grant MT 1453 from the Medical Research Council of Canada.

CHAPTER III

LOCALIZATION OF *ESCHERICHIA COLI* RNA POLYMERASE BINDING SITES ON S13 AND ϕX174 DNAs BY ELECTRON MICROSCOPY

ABSTRACT

Complexes between *E. coli* RNA polymerase and S13 and ϕX174 RFIII DNAs have been shown to form at specific locations on the phage genomes. The major locations on S13 have been mapped at 8-10% and 92-96% of the genome length starting from the unique *Pst*I cleavage site. The locations correspond to the beginning of genes D and B respectively. Four minor locations map at 18-22%, 28-32%, 50-56%, and 70-74% of the genome. The 70-74% site corresponds to the beginning of the A gene. The major locations on ϕX174 are at 8-10%, 50-54%, and 92-94% of the genome. The 50-54% site is at the start of the H gene, has an equivalent minor site on S13, but is not a promoter site. Three minor sites on ϕX174 at 20-24%, 26-32%, and 68-74% of the genome correspond to sites on S13. The data confirms the locations of sites identified by restriction fragment binding experiments (Rassart and Spencer, (1978) J. Virol., in press) and the assignment of putative promoters at the start of genes A, B, and D.

INTRODUCTION

Locations of possible promoters on coliphage DNAs have been determined by three different methods. The first, binding of *Escherichia coli* RNA polymerase to restriction fragments generated by various restriction enzymes, provides a correlation of those fragments which bind polymerase under stringent conditions with the restriction map, thus localizing the binding sites on the restriction/genetic map (Chen *et al.*, 1973; Takeya and Takanami, 1974; Okamoto *et al.*, 1975; Seeburg and Schaller, 1975; Rassart and Spencer, 1978). The second approach involves hybridization of messenger RNAs to the DNA template and to restriction fragments of the DNA. Correlation of the hybridized fragments to the restriction/genetic map locates the start sites of the various messenger RNA's, which are presumed to be the promoter sites (Axelrod, 1976; Hayashi *et al.*, 1976; Smith and Sinsheimer, 1976). The third technique, electron microscopic visualization of RNA polymerase molecules bound to the DNA template (Bordier and Dubochet, 1974; Portmann *et al.*, 1974; Giacomoni *et al.*, 1977; Zollinger *et al.*, 1977) has been made possible by recent modifications of the Kleinschmidt protein monolayer spreading technique (Zollinger *et al.*, 1977) and development of alternate spreading methods for visualization of DNA-protein complexes in the electron microscope (Dubochet *et al.*, 1971; Bordier and Dubochet, 1974; Koller *et al.*, 1974).

On bacteriophage S13 DNA we have previously located three RNA polymerase binding sites as putative promoters using the technique of binding polymerase to restriction fragments (Rassart and Spencer, 1978). The three promoters correspond to the beginning of genes A and B, and a site which includes gene D and the beginning of gene E.

A drawback to the restriction fragment localization method has been found with the filamentous DNA phages. Some restriction enzymes cleave at promoter sites, resulting in loss of promoter function, (Takanami *et al.*, 1976) thus not all RNA polymerase binding sites may be identified by this method. More recently, evidence has been presented that apart from the RNA polymerase recognition sequence on the DNA, an area 5' to the polymerase binding site may also be involved in promoter function (Okamoto *et al.*, 1977).

The present report describes the location of *E. coli* RNA polymerase binding sites on bacteriophage S13 DNA as visualized by electron microscopy using a modification of the Kleinschmidt technique in which spreading is carried out in the presence of a detergent (Zollinger *et al.*, 1977). As in the previous study, comparative experiments have been carried out with the closely related bacteriophage ϕX174 (Rassart and Spencer, 1978).

MATERIALS AND METHODS

Bacteriophage S13 wild type was kindly supplied by Drs. I. Tessman and E. Tessman and ϕX174 am3 by Dr. D.T. Denhardt. Isolation of replicative form I (RFI) DNA has been described elsewhere (Grosveld *et al.*, 1976). *E. coli* RNA polymerase was a generous gift of Dr. F. Grosveld and its purification by the method of Burgess and Travers (1971) and the details of its properties have been descibed previously (Rassart and Spencer, 1978). Restriction endonuclease *Pst*I isolated form *Providencia stuartii*, strain 164, was generously provided by Dr. B. Goodchild. Details of the isolation of the enzyme and the location of the single cleavage site in S13 RFI DNA have been described by Goodchild and Spencer (1978).

Preparation of specimens for electron microscopy

Samples of S13 or ϕX174 RFI DNA were treated with endonuclease *Pst*I to convert the circular RFI DNA molecules to linear RF III molecules (Goodchild and Spencer, 1978). Binding of *E. coli* RNA polymerase to the appropriate DNA template was performed by incubating 8 µg of polymerase with 2 µg of S13 RF III DNA or ϕX174 RF III DNA in 250 µl of binding buffer (50 mM Tris HCl, pH 8 buffer containing 50 mM KCl, 10 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1 mM EDTA, 0.1 mM dithiothreitol, 0.1 mM ATP, and 0.1 mM GTP). The mixture was incubated at 37° for 10 minutes then aliquots of 50 µl were removed, diluted 2-fold with prewarmed (37°) binding buffer to give a final DNA concentration of 4 µg/ml. Spreading on the EM grids was performed at 25° and the electron microscopy executed as described by (Zollinger *et al.*, 1977). The electron micrographs were taken at

magnifications of 29,000.

EM Measurements and Calculations of Results

For measurements of the DNA molecules the micrographs were enlarged 25 times with a modified Keuffel and Esser microfilm projector and analyzed with an electronic pen. Each DNA molecule was measured twice starting from each end and the position of the RNA polymerase molecules calculated as a percentage of the total length of the DNA molecule. From the plot of these results, a perfectly symmetrical histogram was obtained. To orient the molecules, each DNA molecule with an RNA polymerase at the same percentage length position, was plotted on another histogram and the percentage at which the other polymerases occurred noted. This was done for each RNA polymerase position on the symmetrical histogram. For each position, the other RNA polymerase molecules occur either at identical positions, which were plotted, or at mirror image positions which were discarded. When this orientation of molecules was completed for each position on the symmetrical histogram, a unique histogram was obtained showing the location of the RNA polymerase binding sites.

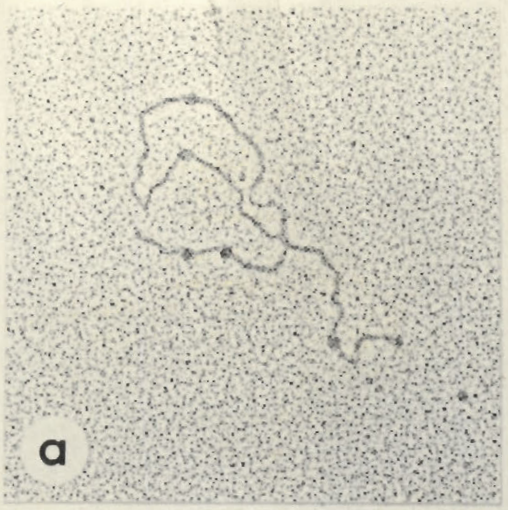
RESULTS

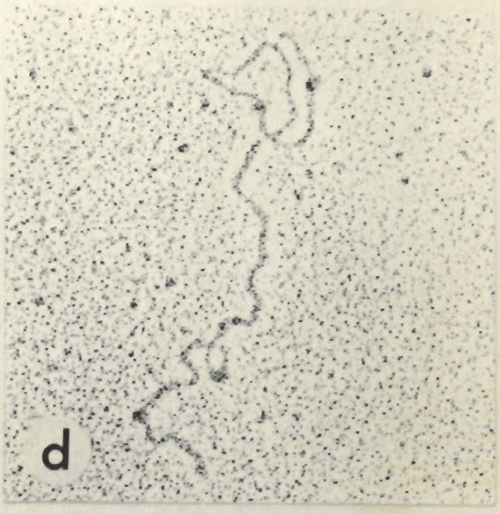
In Figure 1 are shown some typical electron micrographs of *E. coli* RNA polymerase molecules bound to S13 and ϕX174 RF III DNA molecules. The RNA polymerase molecules can be seen as dense spots bound to the DNA. Note that the ends of most DNA molecules are completely free of polymerase molecules, thus background due to non-specific binding to termini is not a problem with this methodology. These and other photomicrographs were used to construct the histograms shown in Figure 2 which delineate the position of the RNA polymerase molecules on the two phage DNAs oriented as described in the methods section. To ensure that the spreading procedure did not alter the relative size of the S13 and ϕX174 molecules the length of 54 S13 and 47 ϕX174 RF II DNA molecules were measured. Unit length T7 DNA molecules were included in the DNA samples as internal length standards. The S13 and ϕX174 DNA molecules were approximately equal in size and the lengths of 1.68± 0.02 μm and 1.69± 0.03 μm respectively, normalized to a T7 length of 12.5 μm (37 molecules measured), (MacHattie and Thomas Jr., 1975; Zollinger *et al.*, 1977) were within the range of 1.64± 0.11 μm (Kleinschmidt *et al.*, 1963) to 1.89± 0.04 μm (Chandler *et al.*, 1964) reported previously for the two DNAs (Spencer *et al.*, 1972).

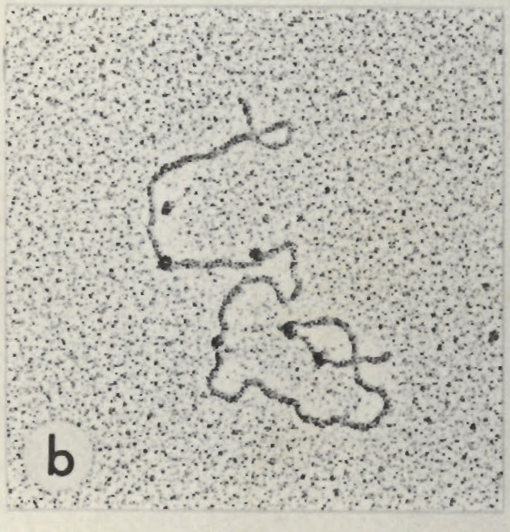
To construct the histogram in Figure 2A, 118 S13 DNA molecules were measured. There were an average of 4 RNA polymerase molecules bound per S13 DNA molecule. For the ϕX174 DNA histogram in Figure 2B, 102 DNA molecules were measured, and there were between 4 and 5 RNA polymerase molecules bound per DNA molecule. The DNA molecules measured were selected according to the following criteria: at least 2 enzyme molecules

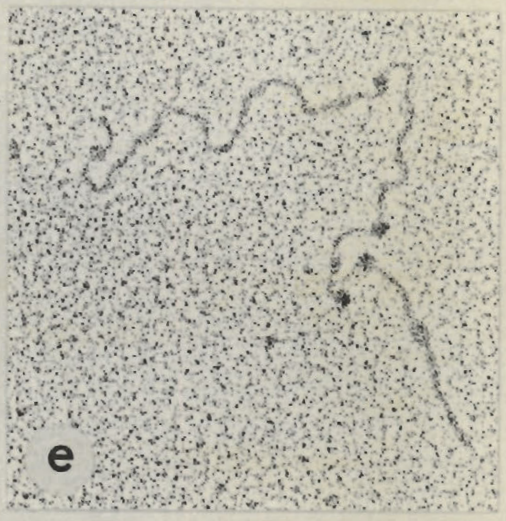
FIGURE 1

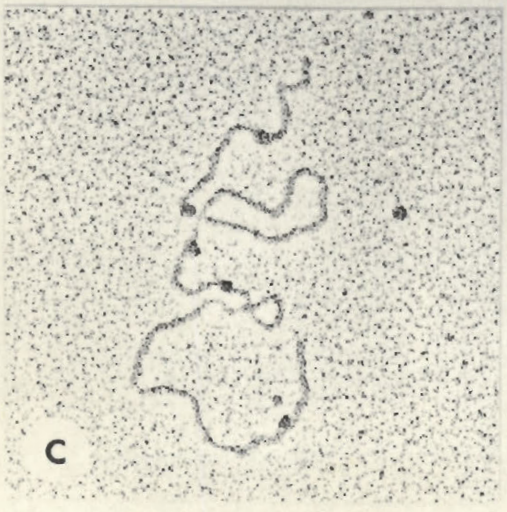
Electron micrographs of *E. coli* RNA polymerase molecules bound to S13 RF III DNA (a-c) and ϕX174 RF III DNA (d-f). Magnification x 29,000.


a


d


b


e


c


f

were bound per DNA molecule; the entire length of the molecule could be followed unambigously; the DNA molecules were intact and unit length. More than 70% of all the molecules on the EM grids fit these criteria for both the S13 and ϕX174 DNAs. The results in Figure 2 are presented in increments of 2% of the total length of the molecules. The automated tracing method gives higher resolution at the start of the measurement, and overall reproduceability is within ±1%. The 2% increment eliminates any discrepancy due to the arbitrary choice of measurement from one end of the molecule or the other.

The histograms in Figure 2 show the locations of the polymerase molecules on the two phages DNAs are almost identical. On S13 DNA (Figure 2A) there are two sites which appear to bind more strongly than others, one at 8-10%, and the second at 92-96%. There are four other sites with more than 11 molecules bound per site. These are at 18-22%, at 28-32%, at 50-56%, and at 70-74%. There is one additional area at 88-90% with more than 11 molecules bound which is probably part of the strong site at 92-96%.

On ϕX174 DNA, (Figure 2B) there are three strong sites at 8-10%, 50-54%, and 92-94%. Three other sites with more than 11 molecules bound per site are at 20-24%, 26-32%, and 68-74%.

In Figure 3, the mapping of the RNA polymerase binding sites by electron microscopy is compared to data derived from the binding of *E. coli* polymerase to *Hin*d and *Hae*III restriction fragments (Rassart and Spencer, 1978). The maps have been aligned with the genetic map of the two phages.

FIGURE 2

Histograms of the distribution of *E. coli* RNA polymerase molecules on S13 RF III DNA (A) and ϕX174 RF III DNA (B). The length increments are 2% of the genome length which corresponds to ∿100 base pairs. The orientation of the viral strand, 5' → 3' is left to right and the origin the *Pst* I cleavage site.

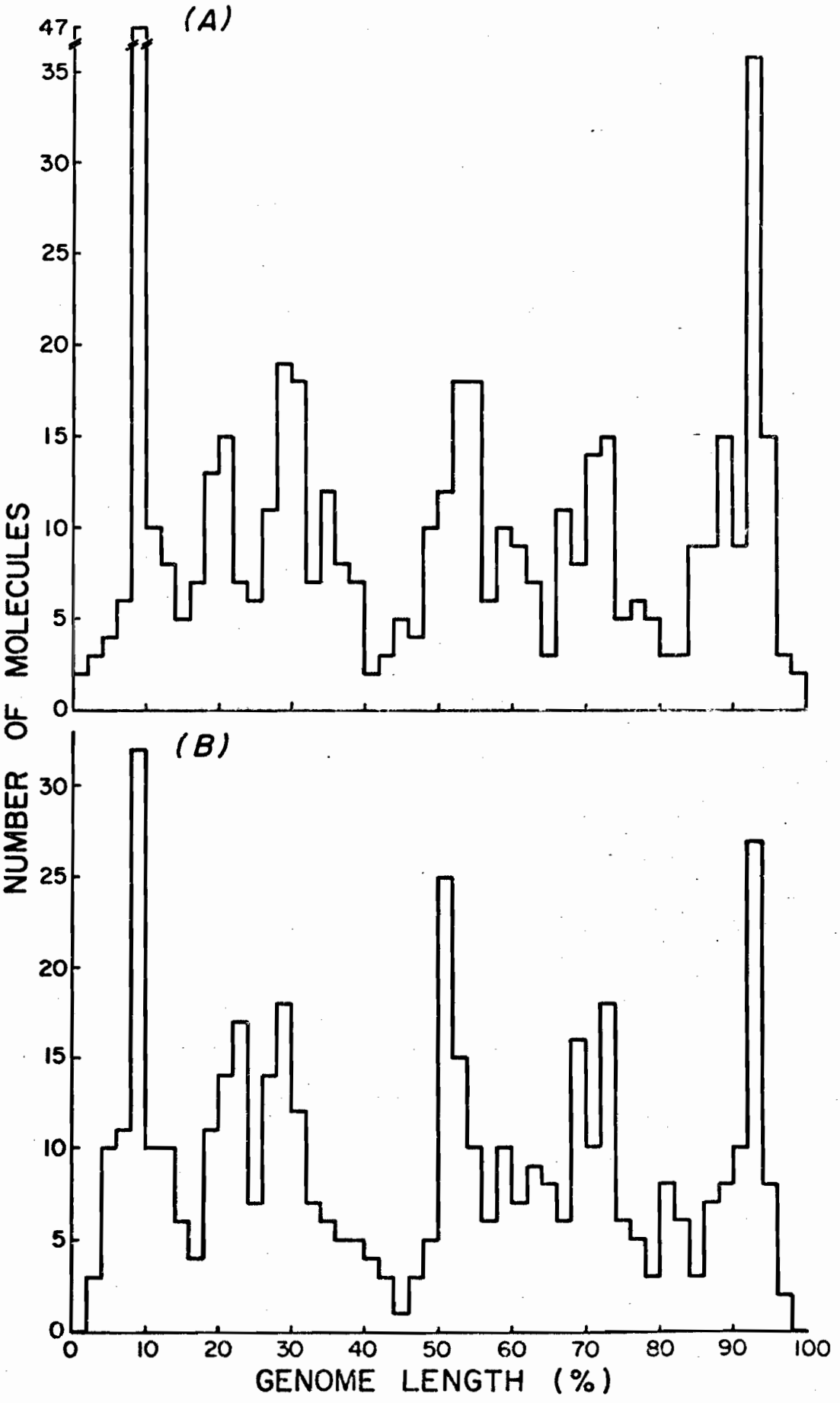
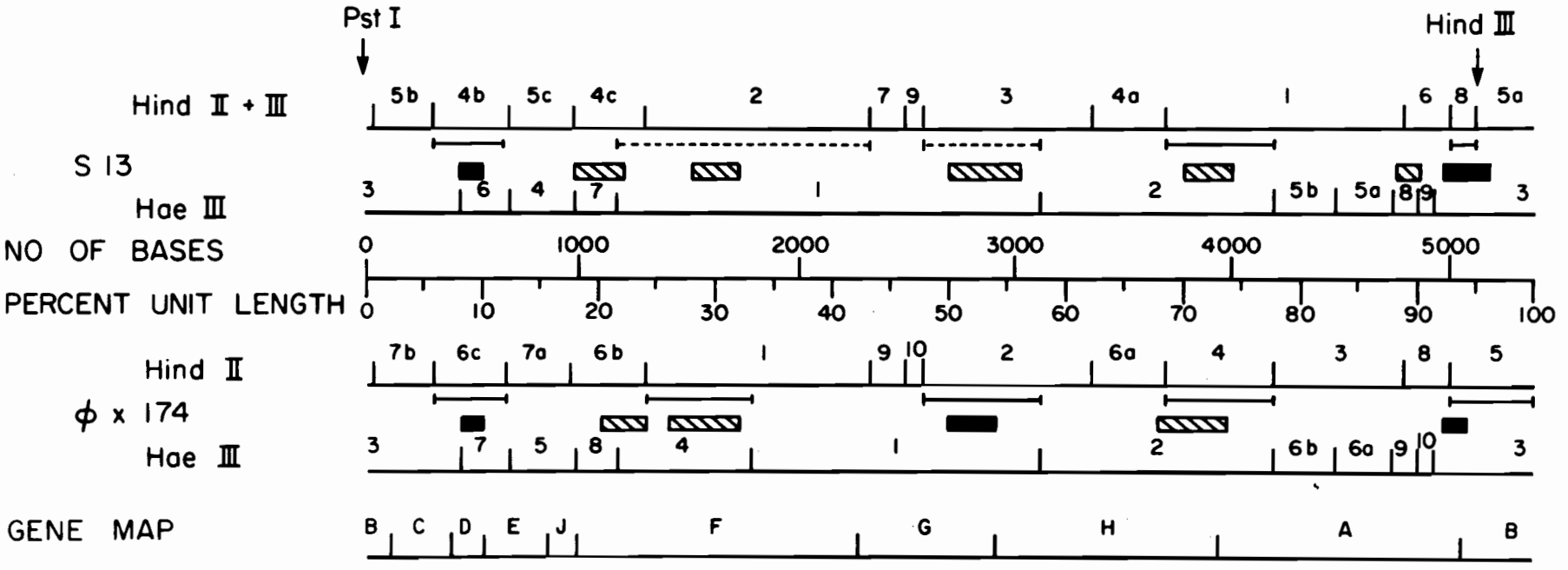

(A)
47
35
30
25
20
15
10
5
0
(B)
30
25
20
15
10
5
0
NUMBER OF MOLECULES
0
10
20
30
40
50
60
70
80
90
100
GENOME LENGTH (%)

FIGURE 3

Schematic presentation of the *E. coli* RNA polymerase binding sites on S13 and ϕX174 RF DNAs in relation to the *Hin*d and *Hae*III restriction endonuclease cleavage maps and the genetic map. Solid lines ⊢—⊣ strong binding sites, dotted lines, ⊢---⊣ , weak binding sites from restriction fragment binding data. Solid boxes ▬ , strong binding sites, hatched boxes ▨ , weak binding sites from electron microscopy data (see Fig. 2). The orientation of the viral strand, 5' → 3' is left to right and the origin the *Pst* I cleavage site.


Pst I
Hind III
Hind II + III
5b 4b 5c 4c 2 7 9 3 4a 1 6 8 5a
S 13
Hae III
3 6 4 7 1 2 5b 5a 8 9 3
NO OF BASES
0 1000 2000 3000 4000 5000
PERCENT UNIT LENGTH
0 10 20 30 40 50 60 70 80 90 100
Hind II
φ x 174
Hae III
7b 6c 7a 6b 1 9 10 2 6a 4 3 8 5
3 7 5 8 4 1 2 6b 6a 9 10 3
GENE MAP
B C D E J F G H A B

DISCUSSION

The electron microscope technique reveals more RNA polymerase binding sites on both S13 and ϕX174 DNAs then were found in the previous study of the binding of *E. coli* RNA polymerase to *Hin*d and *Hae*III restriction fragments (Rassart and Spencer, 1978). The conditions used for binding in the present study differed from those in the restriction fragment binding experiments in a number of aspects. A lower salt concentration of 50 mM KCl was used because in electron microscopy high salt concentrations of 0.2 to 0.3 M interfere in surface spreading and result in crystallization of salt on the EM grids. At lower salt concentrations, binding characteristics of the polymerase to the DNA are changed quite dramatically and binding to sites of low affinity is enhanced (Chamberlin, 1976).

The enzyme:DNA molar ratio in the present study was 30:1 whereas in the restriction fragment binding study it was 150:1. However, as previously reported, similar binding was observed to restriction fragments at a polymerase:DNA ratio of 50:1 (Rassart and Spencer, 1978). In a study of *E. coli* RNA polymerase binding to T7 DNA, Bordier and Dubochet (1974) reported 3 binding sites at a polymerase:DNA ratio of 10:1, but two more sites at a ratio of 30:1. We have measured 20 molecules of S13 from an experiment in which the polymerase:DNA ratio was 15:1 in 50 mM KCl and no differences in the number of sites were observed. Thus, the increased number of sites maybe in large part a result of the lower salt concentration used for binding, a conclusion supported by the results of the restriction binding studies undertaken at various salt concentrations (Rassart and Spencer, 1978).

The sites identified as binding sites are where more than 11 molecules of DNA were bound to any 2% increment of the DNA. The 11 molecule level was an arbitrary choice. In all cases, the width of the peaks increases closer to the base line of the histogram. This may in part be due to loss of mapping precision which is always lower, the smaller the number of molecules that are bound to any particular site. Some of the background may be due to weak binding at AT rich regions of the DNA (Le Talaer *et al.*, 1973), aggregation of the polymerase which is known to occur at low salt concentration (Berg and Chamberlin, 1970), or damaged holoenzyme in the polymerase preparation (Giacomoni *et al.*, 1977).

Figure 3 reveals that in bacteriophage S13 the two strong binding sites are located at the beginning of the D/E gene region and at the beginning of the B gene, corresponding to sites identified in the restriction fragment binding study (Rassart and Spencer, 1978). Of the four minor sites, one is located at the beginning of the A gene, and also corresponds to a site located in the restriction fragment binding study. Two of the sites, one around 30%, and one between 50-56%, are located in the *Hae*III and *Hin*d 2 and 3 areas of S13 which showed higher than background binding of polymerase but which was not sufficient from them to be identified as RNA polymerase binding sites. The site between 50-56% overlaps the beginning of the H gene, but there is strong evidence that the F, G, and H genes are transcribed polycistronically in S13 (Pollock and Tessman, 1976), thus it is doubtful that this location is a promoter site. The site around 20% does not overlap with any restriction fragment which bound RNA polymerase. These areas on S13 may have sequence characteristics which facilitate affinity for polymerase at low salt concentrations

but not at 0.2 to 0.3 M KCl.

On bacteriophage ϕX174 DNA, three strong binding sites were observed. These are located at the beginning of the D/E genes, the H gene and the B gene, and are almost in equivalent locations to sites found in S13 (Rassart and Spencer, 1978). Three other sites were observed on ϕX174 DNA, two correspond to equivalent sites on S13, at 70-74% and about 30%, the remaining site at 20-24% does overlap with an S13 site in the 20% region but is shifted in the 5' → 3' direction, and thus, is not equivalent. It is this site which does not correlate with any restriction fragment polymerase binding site in S13 nor in ϕX174. The other five sites in ϕX174 DNA all correspond exactly in location with the five polymerase binding sites observed in the restriction fragment binding study (Rassart and Spencer, 1978).

The similarities in the binding data between the two phages are closer than the previous restriction fragment binding study. However, the site on S13 DNA at 88-90% has no equivalent in ϕX174 DNA, and the extent of binding at the 50% region is greater in ϕX174 than on S13 DNA.

The site on both phage DNAs at the beginning of the A gene (∿70%) identified as a major site in the restriction fragment binding study, does not appear to be a strong binding site under the conditions used in the present study. The low salt concentration of the binding buffer in the present experiments was similar to that used by Chen *et al.* (1973) in their study of RNA polymerase binding to restriction fragments of ϕX174 DNA in which approximately the same binding site was identified. Chen *et al.* (1973) also identified sites at 50% and at ∿6%. This latter site is close to the ∿10% site but does not coincide with it. However,

the ∿10% site agrees with the transcription data of Axelrod (1976). The discrepancy at this site between the binding data of Chen *et al.* (1973) and transcript hybridization studies (Axelrod, 1976; Smith and Sinsheimer, 1976) has been discussed elsewhere Rassart and Spencer (1978). Chen *et al.* (1973) did not identify the minor sites at ∿20% and 30%.

While there are discrepancies between the data from electron microscope studies and restriction fragment binding studies (Rassart and Spencer, 1978), the three putative promoter sites identified on S13 at the beginning of genes A and B and overlapping gene D and the start of E, are partially confirmed by the present study and the D/E site is more precisely located at the start of gene D.

ACKNOWLEDGEMENTS

The authors wish to thank Drs. F.G. Grosveld and B. Goodchild for generous supplies of enzymes. The work was supported by grant MT 1453 from the Medical Research Council of Canada.

Max Zollinger has contributed to the work of this chapter as a specialist in Electron Microscopy. He did the preparation of the specimens for electron microscopy and took an active part in the measurements.

CHAPTER IV

THE NUCLEOTIDE SEQUENCE OF THE B GENE PROMOTER REGION OF BACTERIOPHAGE S13

ABSTRACT

The sequence of a 176 nucleotide fragment of S13 DNA comprising the promoter and translation initiation region of the B gene is described. Sequencing was accomplished by the use of T4 endonuclease IV on denatured double-stranded DNA restriction fragments, and the development of a cleavage procedure specific for G residues using the alkylating agent methylmethane sulfonate. The two procedures generated an overlapping series of oligonucleotides from which a sequence was constructed. Parts of the sequence were confirmed by the Maxam-Gilbert alkylation and hydrazinolysis-rapid gel technique. Comparison of the sequence with the sequence of the similar region in ϕX174 DNA locates a change from C → T in S13 at a position which then corresponds to the canonical Pribnow sequence T-A-T-G-T-T-A which is in the same area where RNA polymerase has been shown to bind in previous studies. The sequence translated into amino acid sequence data indicates that the amino terminal regions of the B protein from both ϕX174 and S13 are identical, thus differences in the electrophoretic mobility of SDS gels of the two proteins must be due to differences internal in the proteins or to post-translational modifications. The localization of the putative S13 B gene promoter in this sequence places it approximately 100 residues 3' on the viral strand to the site suggested for the B gene promoter of ϕX174 also on the basis of sequence studies. The only other sequence difference in the fragment from the ϕX174 sequence is a G → T change at position 176 which was known previously from the unique *Hin*d III cleavage site at this position.

INTRODUCTION

One of the events which leads to initiation of transcription of a gene at a promoter is interaction of the RNA polymerase with the DNA of the promoter site to form a tight binary complex (Chamberlin, 1976). Formation of this complex confers nuclease resistance on the DNA in the complex (Heyden *et al.*, 1972). This property has been used to isolate RNA polymerase binding sites from various DNAs (Heyden *et al.*, 1972), and the DNA of several of these sites, has been sequenced. A common feature of the sequence at a number of these sites is the Pribnow box; a series of seven nucleotides of general sequence T-A-T-Pu-A-T-G
A-T-A-Py-T-A-C located 5' to the point of initiation of transcription (Pribnow, 1975).

Although a site at which RNA polymerase binds is not necessarily a site at which RNA transcription occurs, the localization of polymerase binding sites is often the first step in studies designed to identify promoters on a genome. The *Escherichia coli* RNA polymerase binding sites on bacteriophage S13 DNA have recently been localized by electron microscopic visualization and mapping (Rassart *et al.*, submitted for publication) and by determination of which restriction fragments of S13 bind the polymerase (Rassart and Spencer, 1978). These locations have been correlated with the genetic map. One region which showed very strong binary complex formation was located close to the start of the B gene. The present study describes the sequence analysis of this region. Restriction enzyme *Hae*III-*Hin*dII-*Hin*dIII cleavage of S13 results in excision of this region as a restriction fragment of 98 base pairs which maps adjacent to another fragment 75 base pairs long, which in turn, is contiguous with two *Hae*III

restriction fragments, the sequences of which have been reported previously (Grosveld and Spencer, 1977). The *Hin*dIII cleavage site is a unique site on S13. *Hin*dIII does not cleave ϕX174 DNA (Goodchild and Spencer, 1978).

The B gene proteins of S13 and ϕX174 differ in their electrophoretic mobility on SDS gels although they exhibit genetic complementarity (Tessman *et al.*, 1976). The promoter site of the B gene of ϕX174 has been placed 176 nucleotides 5' to the initiation point of translation (Sanger *et al.*, 1977), but this location does not coincide with the RNA polymerase binding site identified closer to the translation initiation point (Rassart and Spencer, 1978). This is in contrast to the other ϕX promoters which coincide with polymerase binding sites (Chen *et al.*, 1973; Smith and Sinsheimer, 1976; Sanger *et al.*, 1977). The nucleotide sequence analysis of the restriction fragments S13 *Hae*III 8 and S13 *Hae*III 9, contiguous to the area of the present study, showed only minor differences from the corresponding ϕX sequences and one change in amino acid readout (Grosveld and Spencer, 1977). The unique *Hin*d III restriction cleavage site on S13 indicates at least one base difference between the two phages in this general area (Goodchild and Spencer, 1978), and it was hoped that additional sequence information would reveal other changes which might point to an explanation of the RNA polymerase binding studies. Sequence analysis could also give a first indication, based on amino acid readout, of overlap of the A and B genes in S13.

The methodological approach to the sequence analysis of this region included T4 endonuclease IV hydrolysis and the use of methylmethane sulfonate to alkylate at G residues and in conjunction with alkaline hydrolysis generate oligonucleotides which overlapped with the T4 endonuclease IV

digest as an alternative method of sequence analysis. These methods have been complemented by the rapid Maxam-Gilbert analysis method.

The choice of methylmethane sulfonate was based on the results of a systematic examination of a number of methylating agents by Lawley and Brooks (1963). They showed that the most readily alkylated bases at neutral pH are guanine and adenine, and the products are 7-methyl guanine and 3-methyl adenine respectively. Methylmethane sulfonate displayed a high preference for alkylation of the N-7 position of guanine, approximately nine times more than for the C-3 position of adenine, thus alkylation, followed by depurination and cleavage at G residues predominates sufficiently to generate a specific series of oligonucleotides for sequencing purposes.

MATERIALS AND METHODS

Chemicals

DEAE Sephadex A-25 was obtained from Pharmacia (Canada) Ltd. and $H_3{}^{32}PO_4$ from NEN (Canada) Ltd. [γ-^{32}P]ATP was prepared as described by Schendel and Wells. Dimethyl sulfate was obtained from Aldrich Chemical Co. Hydrazine, methylmethane sulfonate, acrylamide and bis (NN'-methylene-bis (acrylamide)) were Eastman products. DEAE-cellulose and cellulose for thin layer chromatography were products of Macherey, Nagel and Co. purchased from Brinkmann Instruments (Canada) Ltd. Cellulose acetate strips were from Schleicher and Schuell Inc. Keene, New Hampshire, and Cellogel strips, manufactured by Chemtron, Milan Italy were from Mandel Scientific Co.

Enzymes

Snake venom phosphodiesterase and bacterial alkaline phosphatase (electrophoretically pure) were obtained from ICN (Canada) Ltd. The alkaline phosphatase was heat treated to remove phosphodiesterases (Garen and Levinthal, 1960) and the snake venom phosphodiesterase pretreated at 37°C for 3 hrs. at pH 3.6 to remove 5' nucleotidase activity (Sulkowski and Laskowski, 1971). Polynucleotide kinase and T4 endonuclease IV were prepared from *E. coli* B infected with bacteriophage T4 *am*N 82. Polynucleotide kinase was isolated according to Richardson (1965) with the modifications described by Harbers *et al.* (1976) and was a generous gift of Dr. A. Delaney. T4 endonuclease IV was isolated by the method of Sadowski and Bakyta (1972) with the modifications described by Grosveld and Spencer (1977). Restriction endonuclease *Hae*III was isolated from *Hemophilus aegyptius* ATCC 1116 and endonuclease *Hin*dII + III (*Hin*d)

from *Hemophilus influenzae*, strain RD, by the method of Smith and Wilcox (1970), modified as described by Grosveld *et al.*, (1976). Restriction endonuclease *Hin*dIII was separated from *Hin*dII + III by the procedure of Ludwig and Summers (1975) and was a generous gift of Dr. B. Goodchild. The characteristics and properties of the separated enzymes have been described by Goodchild and Spencer. Restriction endonuclease *Alu*I was isolated from *Athrobacter luteus* ATCC 21606 as described by Roberts *et al.* DNase A purified form DNase I (Worthington Biochemical Corp) was a generous gift of Dr. E. Junowicz. *E. coli* RNA polymerase was purified by the method of Burgess and Travers (1971). The details of its properties have been described elsewhere. T4 endonuclease IV, *Hae*III, *Hin*dII + III and *E. coli* RNA polymerase were generously provided by Dr. F.G. Grosveld.

Purification of ^{32}P-labeled bacteriophage S13 RF DNA and isolation of restriction fragments.

^{32}P-labeled bacteriophage S13 RFI DNA was purified according to Schekman *et al.*, (1971) with the modifications described by Grosveld and Spencer (1977). Digestion of S13 RF DNA with restriction endonucleases *Hin*d, *Hae*III and *Alu*I were exactly as described in Grosveld *et al.* (1976) and Roberts *et al.* (1976) for *Hin*d and *Hae*III. Restriction fragments were separated by gel electrophoresis as described by Grosveld *et al.* (1976). The separated fragments were eluted from the gels by macerating the gel in 1 M NaCl and incubating the slurry for 16 hrs at 4°C. The mixture was then filtered through a fine grade sintered glass filter and the fragments purified (Grosveld and Spencer, 1977) by application to a 1 cm x 1 cm column of hydroxyapatite. The column was washed with 0.1 M potassium phosphate buffer pH 7.0 and eluted with 1.0 M potassium phosphate buffer

pH 7.0. The restriction fragments in the eluate were desalted by adsorption to DEAE cellulose and eluted with 1 M Triethylammonium bicarbonate (TEAB), pH 8, followed by evaporation to dryness.

T4 endonuclease IV hydrolysis.

The purified and desalted S13 DNA restriction fragment (5-10 µg) was dissolved in 100 µl H_2O, the solution heated at 100°C for 3 min, and cooled quickly in ice prior to addition of 10 µl of 0.2 M Tris-HCl buffer, pH 8.5, containing 0.1 M $MgCl_2$ and 50 mM 2-mercaptoethanol, followed by 20 µl of T4 endonuclease IV (20 units). The mixture was incubated at 37°C for 3 hours, then 20 µg of alkaline phosphatase added and the mixture reincubated at 60°C for 1 hr. The solution was extracted with phenol, the aqueous phase extracted with ether and applied to a 1.0 cm x 0.5 cm, DEAE Sephadex A-25 column. The oligonucleotides were eluted with 1.0 M TEAB and the solution evaporated to dryness.

DNA alkylation and Hydrolysis

The purified and desalted S13 DNA restriction fragment (5-10 µg) was dissolved in 100 µl of 0.5 M sodium phosphate buffer pH 7 and incubated with either 0.3 or 0.5 M methylmethane sulfonate for one hour at 37°C. The mixture was then extracted with ether 3 times, heated at 100°C for 10 min. NaOH added to a final concentration of 0.2 M, reincubated at 37°C for 50 min and finally neutralized with HCl. The neutralized solution was applied to DEAE Sephadex A-25, the column washed with 50 mM TEAB, pH 8.0, eluted with 1.3 M TEAB, and the eluate evaporated to dryness.

5' ^{32}P-labeling of alkylation and T4 endonuclease IV generated oligonucleotides

The oligonucleotides released from the S13 DNA restriction fragments following alkylation were incubated with 20 µg of alkaline phosphatase in 200 µl of 50 mM Tris-HCl buffer, pH 8.9, containing 10 mM $MgCl_2$, at 60°C for 1 hr. The alkaline phosphatase was inactivated by extraction with phenol and the aqueous phase extracted with ether then evaporated to dryness. The dried oligonucleotides from above or from the T4 endonuclease IV hydrolysate, were dissolved in 30 µl of H_2O and 20 µl of [γ-^{32}P]ATP, ($\sim 10^8$ dpm), 20 µl of 1% (v/v) 2-mercaptoethanol and 30 µl of polynucleotide kinase (0.015 units) added. The mixture was incubated at 37°C for 3 hrs. then applied directly to a 1 cm x 0.5 cm DEAE Sephadex A-25 column. The excess ATP was removed by elution with 0.2 M TEAB, pH 8.0, and the 5' ^{32}P-labeled oligonucleotides eluted with 1.3 M TEAB, the eluate collected and evaporated to dryness.

Fractionation of oligonucleotides by ionophoresis-homochromatography

The 5' ^{32}P-labeled oligonucleotides were separated by ionophoresis-homochromatography as described in detail elsewhere for pyrimidine oligonucleotides (Harbers *et al.*, 1976) with the minor modifications described for T4 endonuclease IV oligonucleotides (Grosveld and Spencer, 1977). The separated oligonucleotides were visualized by radioautography.

Partial digestion of oligonucleotides by snake venom phosphodiesterase

This was performed as described by Ling (1972). Spots were scraped from the thin layer plates, the DEAE cellulose washed with cold absolute ethanol, the oligonucleotides eluted with 1.3 M TEAB and the eluate evaporated to dryness. Following partial digestion with snake venom phosphodiesterase, the products were separated by ionophoresis-homochroma-

tography and the separated oligonucleotides visualized by radioautography (Brownlee and Sanger, 1969).

Sequence Analysis by the Alkylation-Hydrazinolysis - Rapid Gel technique

Restriction fragments were 5' ^{32}P-terminally labeled using [γ-^{32}P] ATP and polynucleotide kinase, cleaved by a second restriction endonuclease, the products separated, to provide fragments with one 5' terminus ^{32}P-labeled only. The 5' ^{32}P-terminally labeled fragments were sequenced by the chemical method of Maxam and Gilbert using the alternative guanine cleavage and the alternative strong adenine/weak cytosine cleavage reactions. 20% polyacrylamide sequencing gels were prepared and run as described by Maxam and Gilbert (1977).

Isolation of an RNA Polymerase Binding Site

3 µg of restriction fragment *Hin*d 8 were incubated with 35 µg of *E. coli* RNA polymerase in 150 µl of binding buffer (50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1 mM EDTA, 0.1 mM dithiothreitol, 50 mM KCl, 0.1 mM ATP and 0.1 mM GTP), for 20 min at 30°C. Then 40 µg of DNase A were added and incubation continued for a further 60 min. The incubation mixture was extracted with phenol the aqueous phase extracted with ether then applied to DEAE Sephadex A-25. The column was washed with 0.3 M TEAB to remove small oligonucleotides and the DNA retained on the column eluted with 1.3 M TEAB. The eluate was evaporated to dryness, redissolved in water, and reevaporated to dryness repeatedly. The purified DNA fragment was then hydrolysized with T4 endonuclease IV as described above. The oligonucleotides released were 5' ^{32}P-labeled, fractionated and sequenced by partial snake venom phosphodiesterase digestion and the wandering spot technique (Brownlee and Sanger, 1969).

RESULTS

Preliminary experiments to separate the strands of the 98 base pair (bp) fragment released form S13 *Hin*dII 4 by digestion with *Hin*dIII (fragment S13 *Hin*d 8, Fig. 1) by electrophoresis on cellulose acetate, a technique used for S13 *Hae*III 8 and S13 *Hae*III 9 fragments, were unsuccessful. Accordingly, a procedure was devised to denature the restriction fragment and hydrolyze the denatured strands with the single-strand specific enzyme T4 endonuclease IV. The products of a total T4 endonuclease IV hydrolysis of denatured S13 *Hin*d 8, separated by ionophoresis - homochromatography are shown in figure 2, experiment F. The radioautogram shows a distribution of oligonucleotide spots indicative of a broad range of sizes and base compositions with no unusual features. The numbered spots were scraped out and sequenced by partial snake venom phosphodiesterase hydrolysis and the wandering spot technique. Examples of the separations are shown in figure 3 (D13 and F9). The sequences derived from this and similar experiments are presented in Table 1. Overlaps of the sequences were insufficient to allow construction of a sequence for the *Hin*d 8 fragment, and a further method of analysis was sought.

Methylmethane sulfonate was selected as an alkylating agent, because of its specificity towards guanine residues over adenine residues at a ratio of 9:1. This presented the possibility of generating a congruant series of oligonucleotides from G residue cleavages to those isolated from the T4 endonuclease IV hydrolysate resulting from cleavages adjacent to C residues.

To obtain oligonucleotides generated by methylmethane sulfonate of a size suitable for partial snake venom phosphodiesterase wandering spot

Figure 1 : Schematic representation of the location of the *Hin*dII/*Hae*III 3-75 and "*Hin*d 8" fragments.
*Hin*dII/*Hae*III 3-75 can be obtained either by a *Hae*III digestion of *Hin*dII 7 or a *Hin*dII digestion of *Hae*III 3.
"*Hin*d 8" can be further degraded with *Alu*I restriction enzyme for sequencing purposes into three fragments of respectively 43, 32 and 23 base pairs (bp).

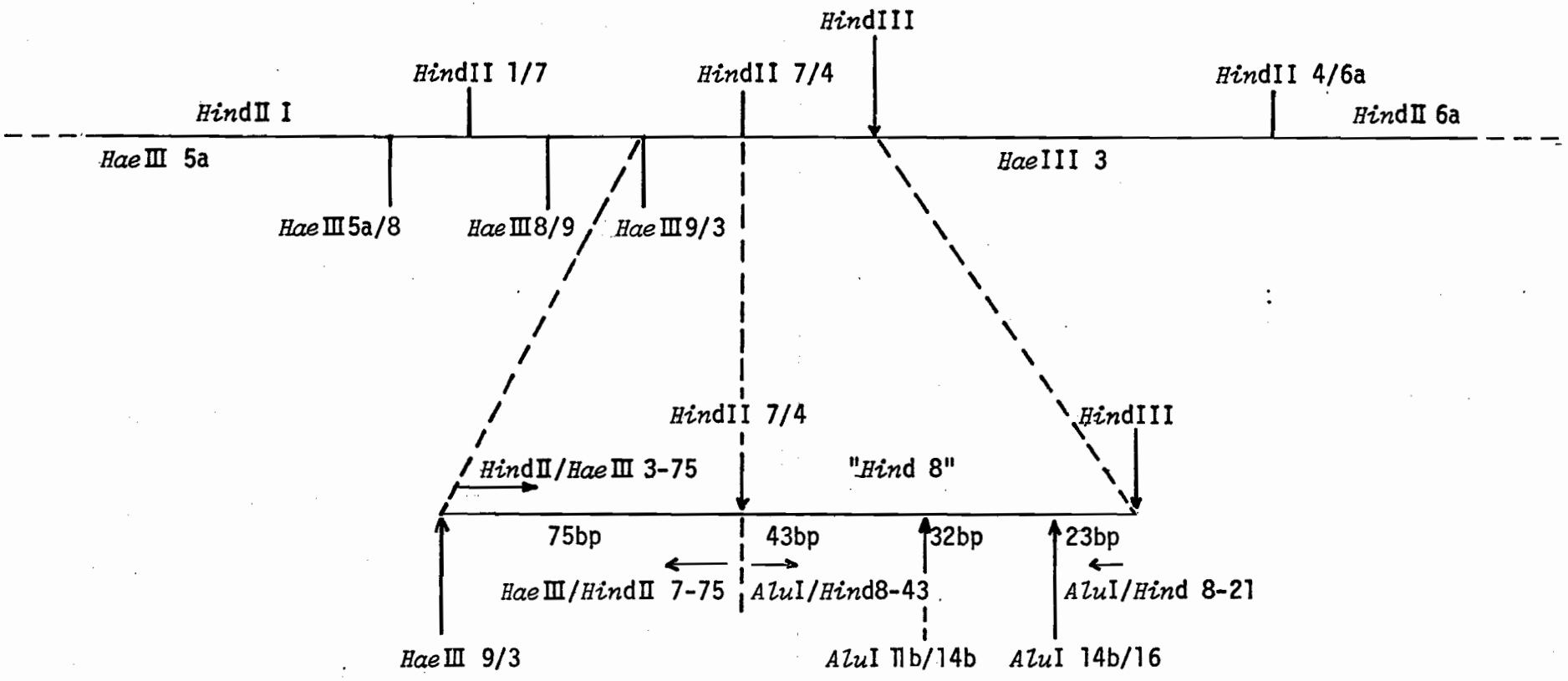
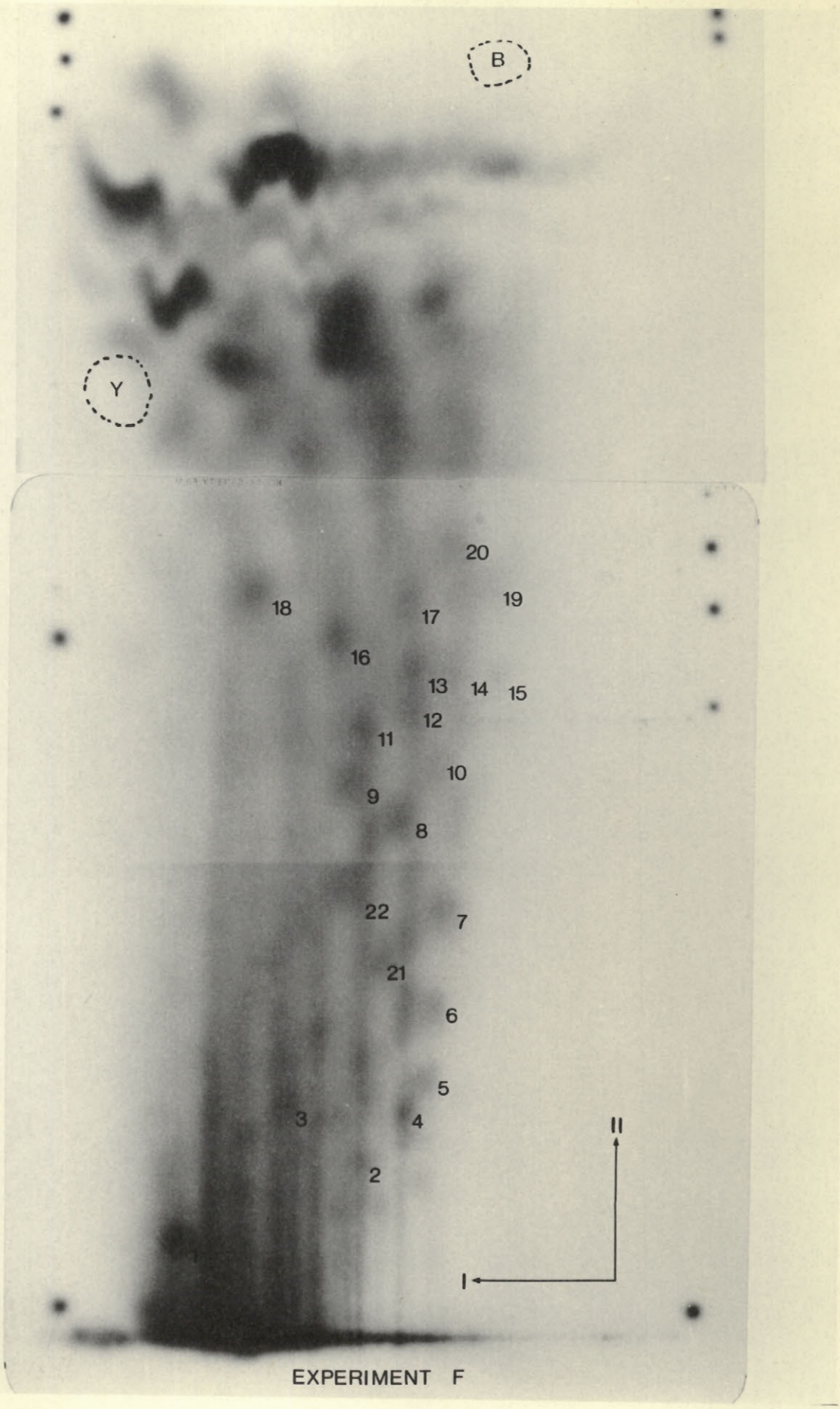

HindIII
HindII 1/7
HindII 7/4
HindII 4/6a
HindII I
HindII 6a
HaeIII 5a
HaeIII 3
HaeIII5a/8
HaeIII8/9
HaeIII9/3
HindII 7/4
HindIII
HindII/HaeIII 3-75
"Hind 8"
75bp
43bp
32bp
23bp
HaeIII/HindII 7-75
AluI/Hind8-43
AluI/Hind 8-21
HaeIII 9/3
AluI IIb/14b
AluI 14b/16

Figure 2 : Fingerprints of the oligonucleotides released by T4 endonuclease IV (experiment F) or by DNA alkylation and hydrolysis with methylmethane sulfonate (experiment A) of *Hin*dII 8 restriction fragments.

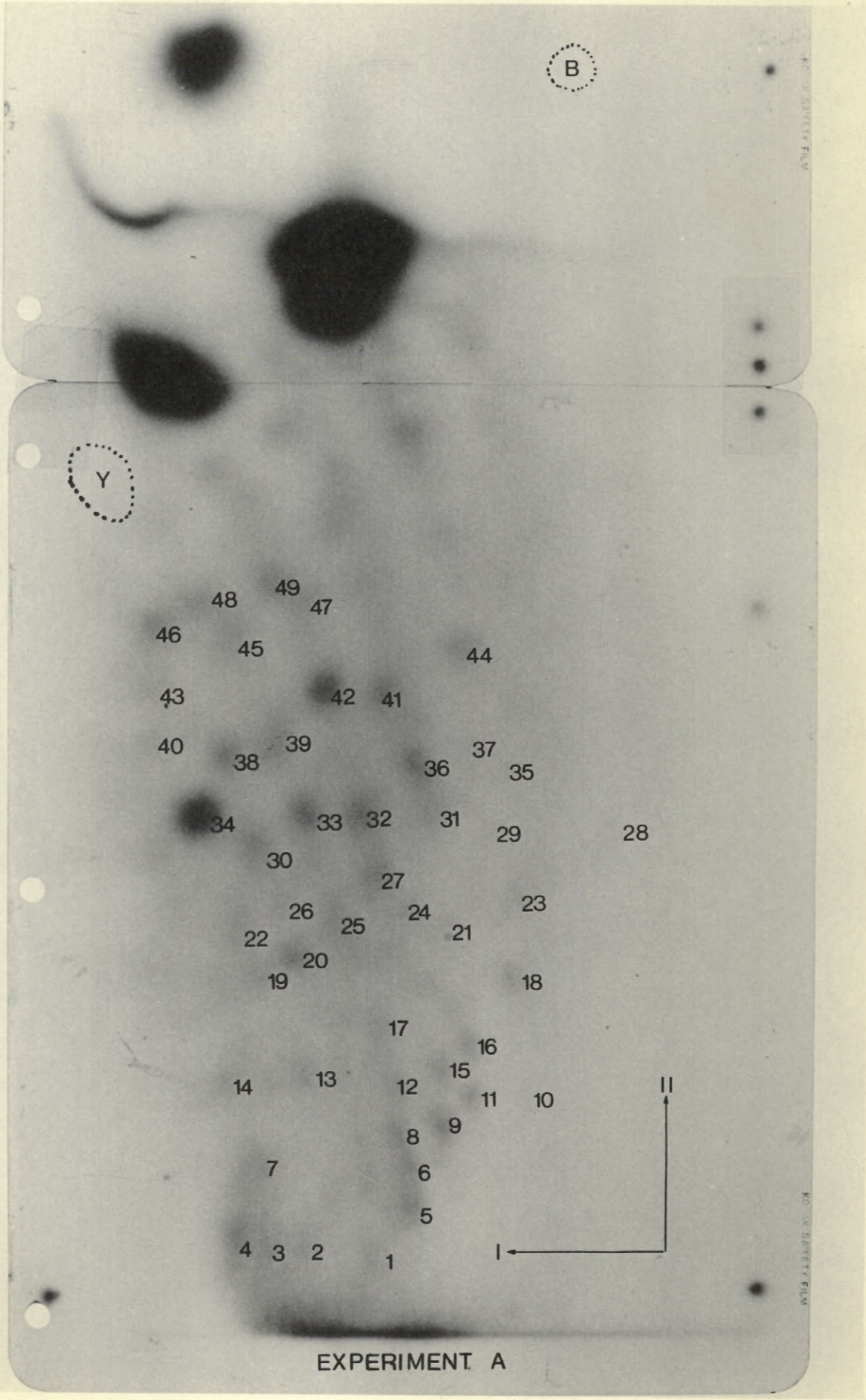
dimension I Electrophoresis on cellogel

dimension II Homochromatography

The dotted circles indicate the position of the yellow and blue dye markers.

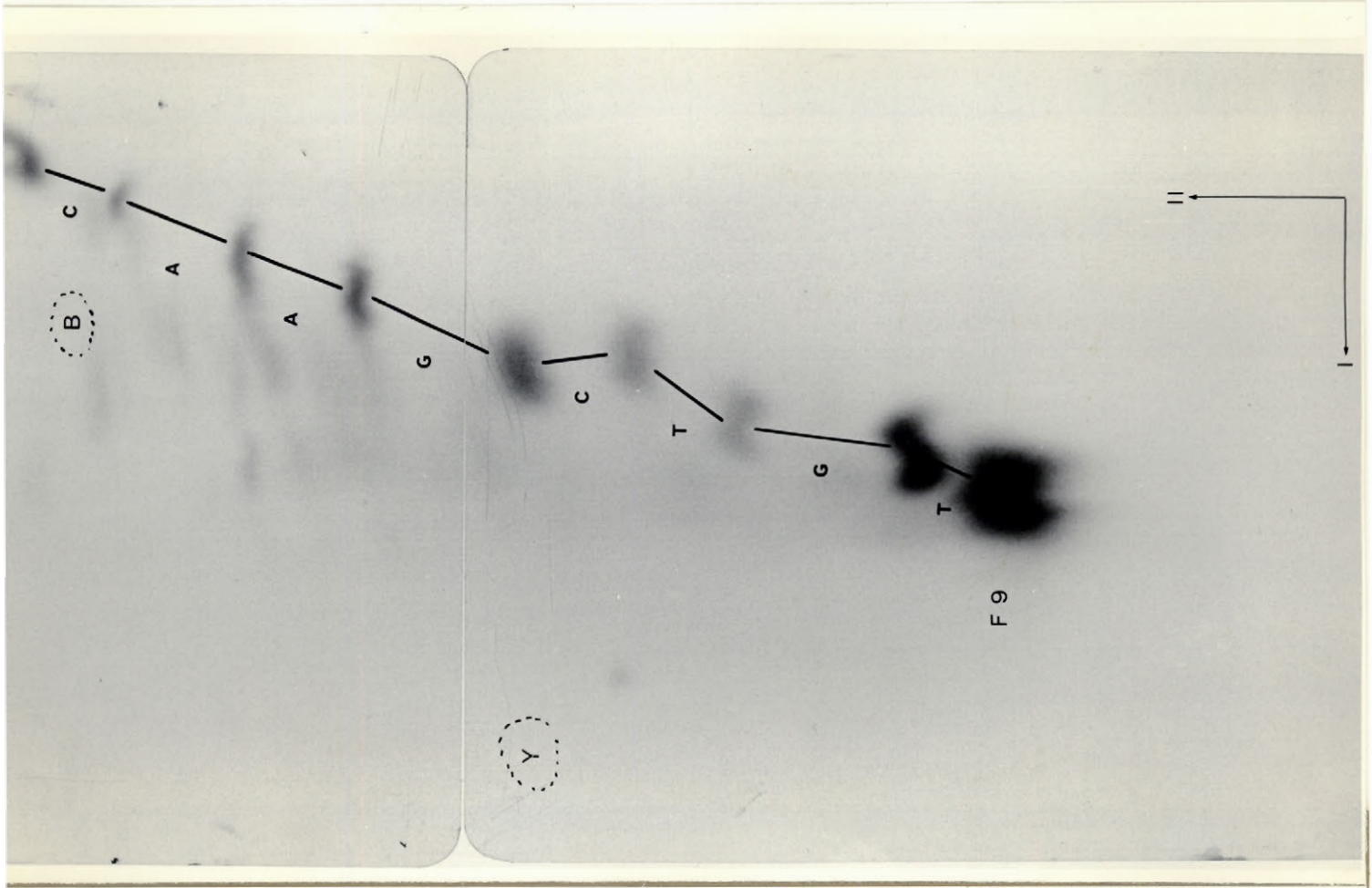


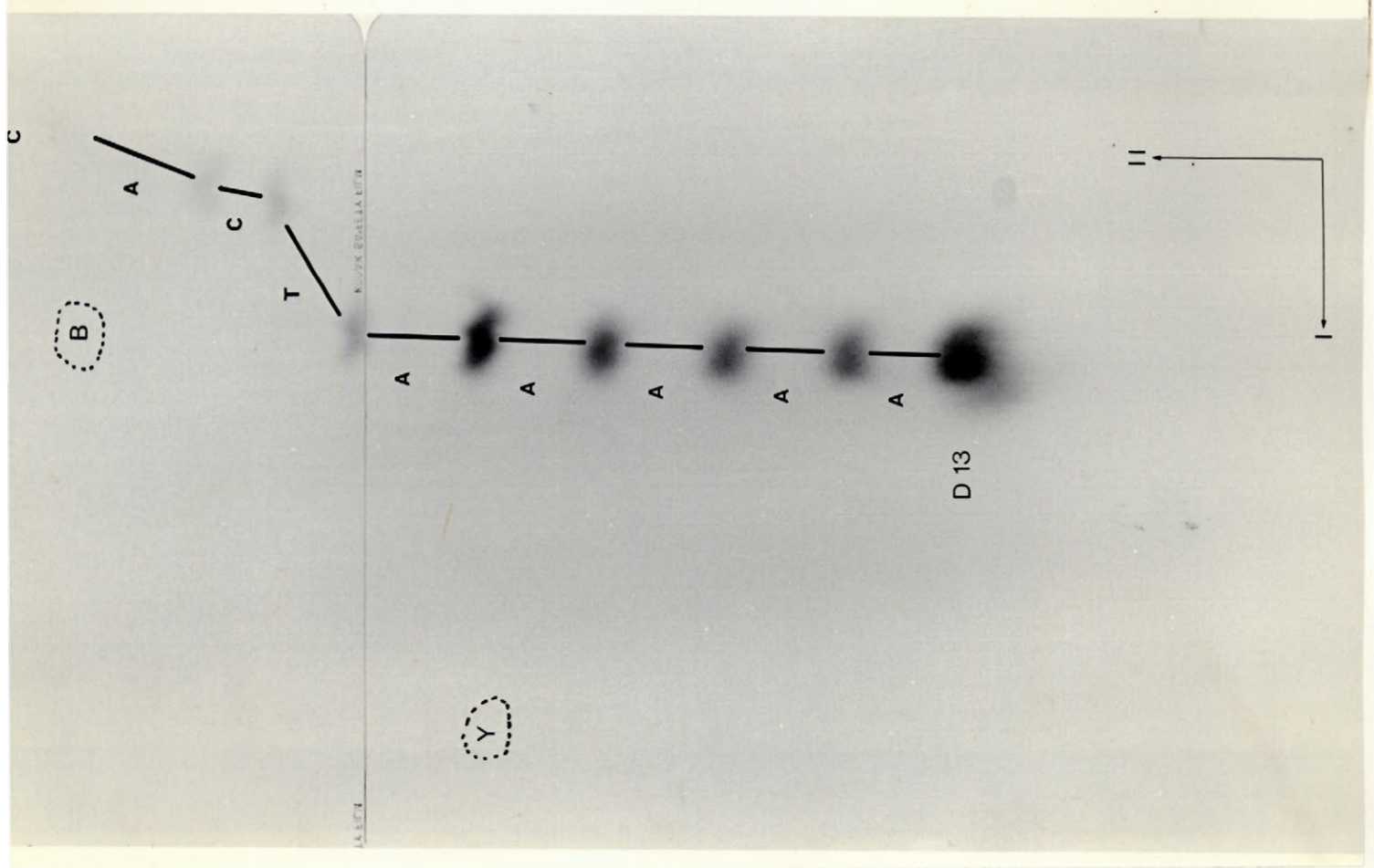
EXPERIMENT F

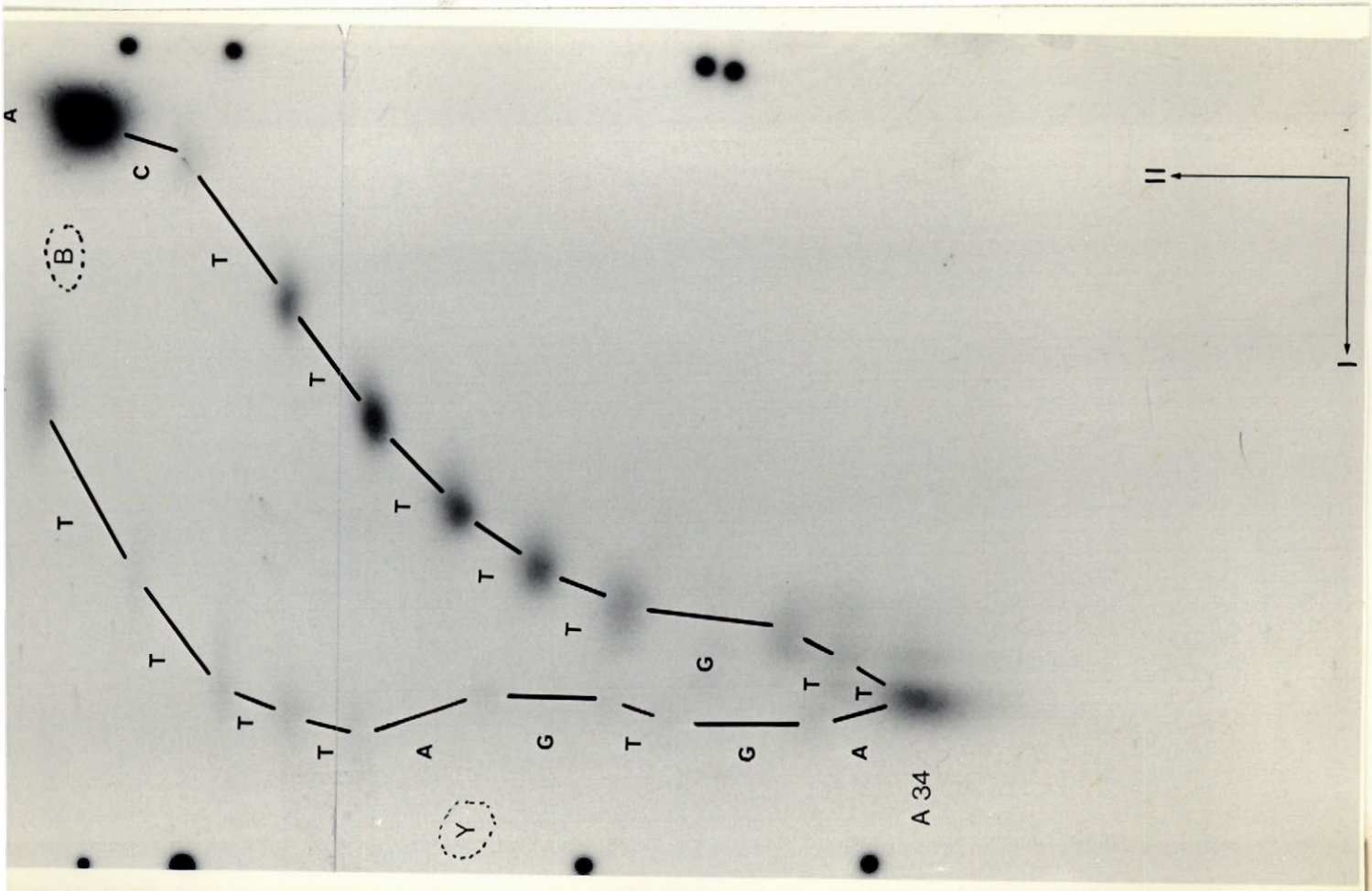


EXPERIMENT A

Figure 3 : Sequence chromatograms of T4 endonuclease IV oligonucleotides (D13 and F9) and oligonucleotides generated by methylmethane sulfonate (A15 and A34) of fragment *Hin*d 8.


B
C
A
A
G
C
T
G
T
Y
F 9
II
I


C
B
A
C
T
A
A
A
A
A
Y
D 13
II
I


A
B
C
T
T
T
T
T
T
T
T
T
T
A
G
T
G
T
A
T
Y
A 34
II
I

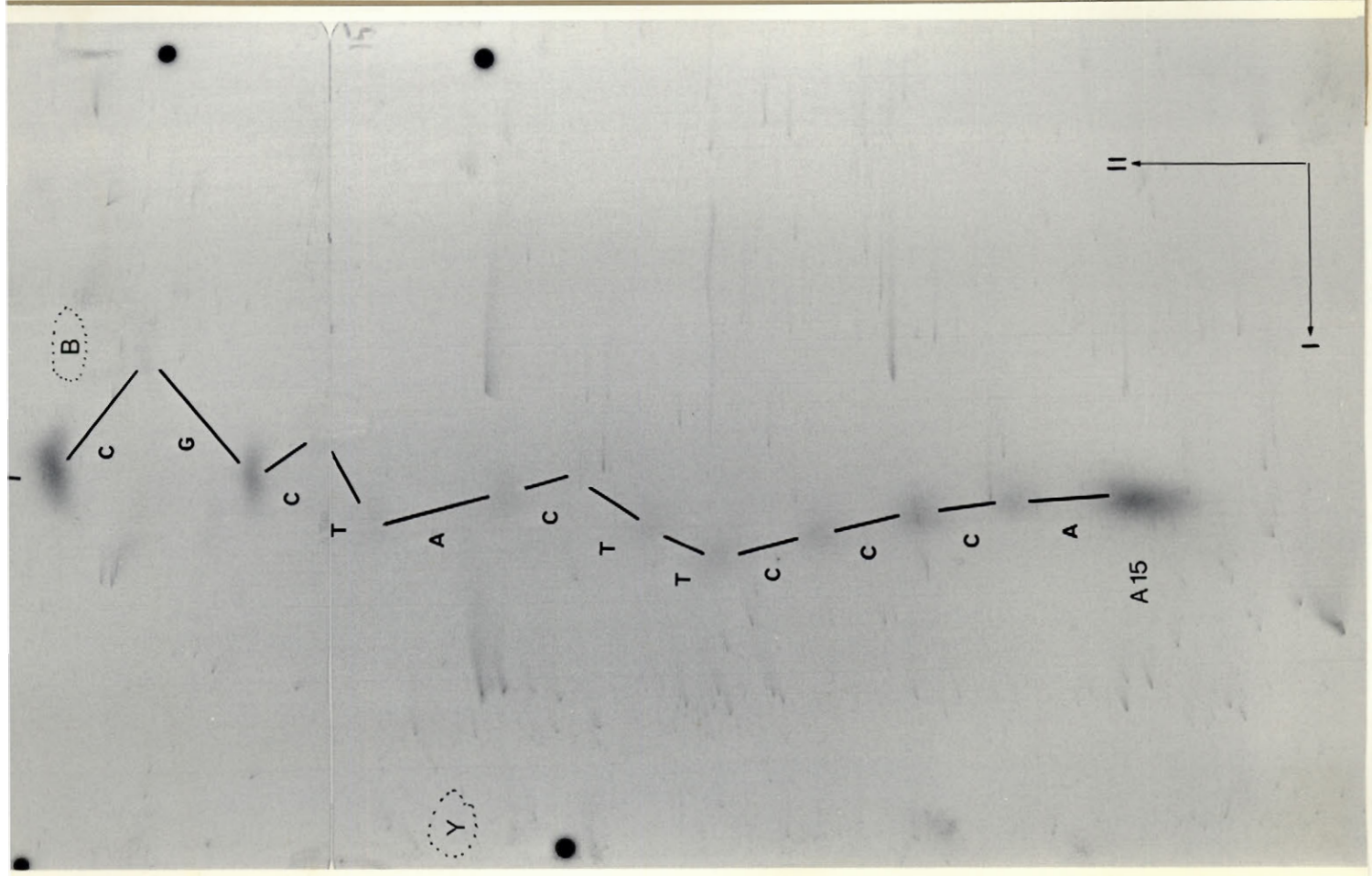

B
C
G
C
T
A
C
T
T
C
C
C
A
Y
A 15
II
I

TABLE I

Sequences of the oligonucleotides released by T4 endonuclease IV from fragment *Hin*d 8

D 8,F21	A-A-C-A-A-A-A-A-G-T
D 9	- - - -$\frac{C}{A}$-T-T-T-A-G
D11	C-C-A-T-A-T-C-T-G-A
D12	C-T-A-G-A-C-$\frac{C}{A}$-T-T-T
D13	C-A-C-T-A-A-A-A-A
D15	C-C-A-T-T-C-T-T-T-A-G
D16	C-A-T-A-T-C-T-G-A
D17,F10	C-A-A-C-T-C-A-C-T
D18,F11	C-G-C-T-A-C-T-T-C
D20	C-A-A-C-T-C-A-C
D22a	C-T-G-T-C-G-C-T-A
D22b	C-C-A-T-T-C-T-T-T
D24,F16	C-G-C-T-A-C-T-T
D25,F17	C-T-A-C-T-T-C-C
D26,D27,F18	C-T-T-T-A-G-C-T
D28	C-A-A-C-T-C-A
F 1	C-C-A-T-A-T-C-T-G-A-C-T-T-T-T- - -
F 9	C-C-A-A-G-C-T-G-T
F12	C-T-T-T-T-T-G-T-T

Letters D and F refer to different experiments

TABLE II

Sequences of the oligonucleotides released by methylmethane sulfonate alkylation and alkaline hydrolysis of fragment *Hin*d 8

A 8	C-T-C-C-T-A-G-A-C-C/A-T-T-T-A
A 9	C-T-A-C-T-T-C-C-C-A-A-G-A
A11	T-C-G-C-T-A-C-T-T-C-C-C-A-A
A13	T-C-C-A-T-A-T-C-T-G-A-C-T
A14	T-T-G-T-T-C- - - - -C-T-T-T-A
A15	T-C-G-C-T-A-C-T-T-C-C-C-A
A16	C-T-A-C-T-T-C-C-C-A-A-G
A18	C-T-A-C-T-T-C-C-C-A-A
A20	T-T-C-C-A-T-T-C-T-T-T-A
A21	A-C-C/A-T-T-T-A-G-C-A
A23	C-T-A-C-T-T-C-C-C-A
A26	T-T-C-C-A-T-T-C-T-T-T-A-G
A31	C-G-A-C-A-G-C-T-T
A32,B 4	T-C-A-G-A-T-A-T-G
A33,C5	A-G-C-T-T-C-T-T-G
A34a	T-T-T-T-T-A-G-T-G-A
A34b,B 6	A-C-T-T-T-T-T-G-T-T
A36,B14	T-C-A-G-A-T-A-T
A37,B18	G-A-G-C-T-A-A-A
A38,B12	A-C-T-T-T-T-T-G
A42,B11	A-G-C-T-T-C-T-T
A44	A-C-A-G-C-T-T
A46	T-T-T-T-T-A-G-T
A47	C-T-T-C-T-T-G
A48	G-T-T-T-T-T-A
B 8	T-C-G-C-T-A-C-T-T
B 9	A-C-T-T-T-T-T-G-T
B17	A-C-T-T-T-T-T
B13	A-C-C/A-T-T-T-A
C2	A-A-C-A-A-A-A-A-G-T-C-A

Letters A,B, and C refer to different experiments.

Sequences from experiment C are the 5' termini (see text).

sequence analysis, several concentrations of methylmethane sulfonate and the duration of alkylation were investigated. Since 3-methyladenine depurinates more rapidly than 7-methylguanine 1 hr was selected as the period of alkylation and the concentration of methylmethane sulfonate varied from 0.1 to 1.0 M. The size of the hydrolysis products was determined by homochromatographic separation. Concentrations of 0.3 to 0.5 M methylmethane sulfonate resulted in oligonucleotides of an average size of 15 nucleotides over a range of 5 to 20 nucleotides. These conditions were used in all subsequent experiments. A radioautogram of a total alkylation and alkaline hydrolysate of S13 *Hin*d 8 following 5' ^{32}P labeling, separated by ionophoresis homochromatography is shown in figure 2, experiment A. The numbered oligonucleotides were scraped from the TLC plate and partially degraded with snake venom phosphodiesterase. Examples of the separation of some of these partial hydrolysis products by ionophoresis - homochromatography is shown in figure 3 (A15 and A34). The sequences determined from this experiment and others are presented in Table II. Together with the T4 endonuclease IV generated oligonucleotides, Table I, a partial tentative sequence was constructed from sequence overlaps and this is presented as part of figure 4.

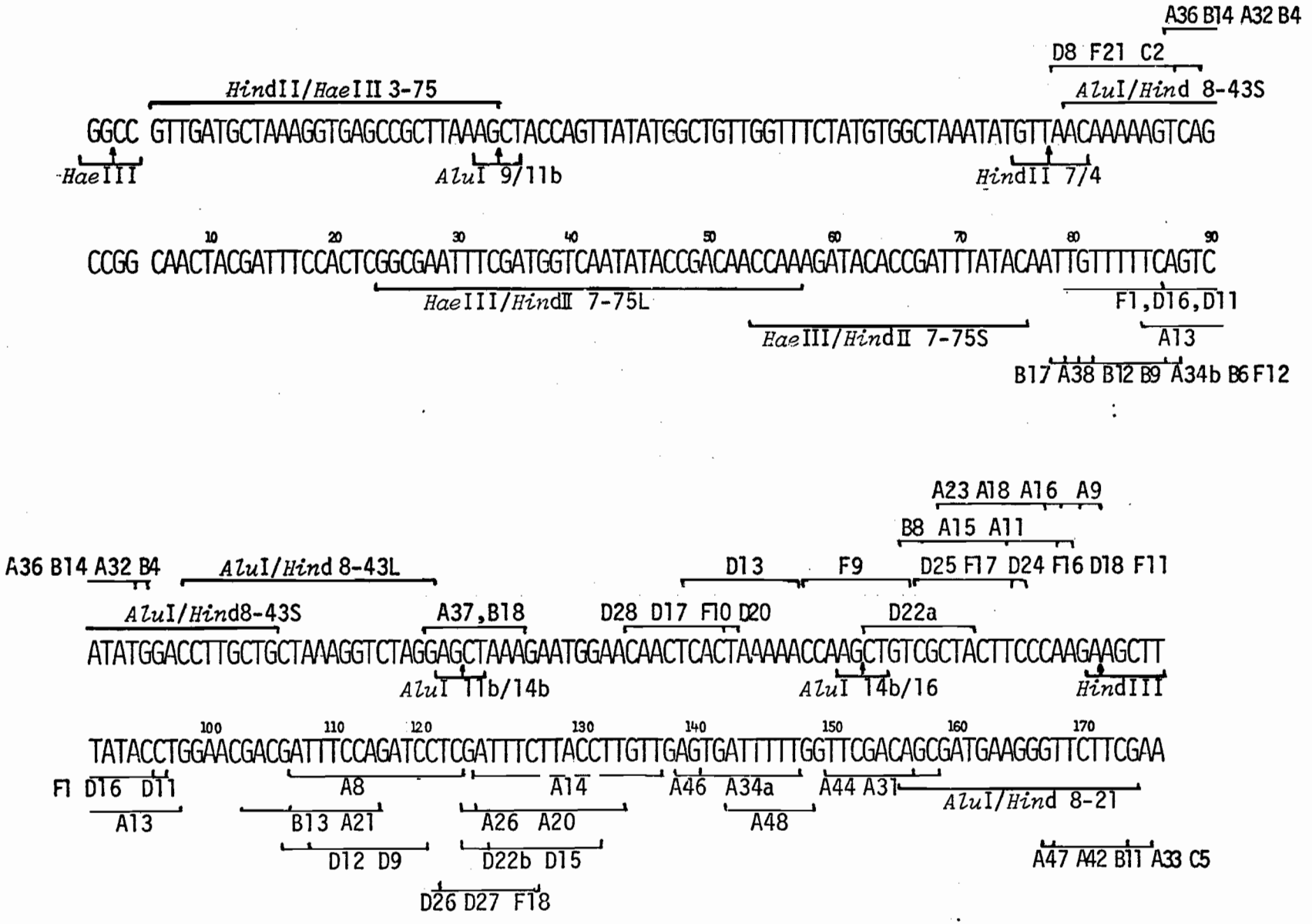
Examples of sequence alignments of the oligonucleotides shown in figure 3 are as follows, for residues 138 and 168 inclusive: on the viral strand, oligonucleotide A15 overlaps with D22a (radioautograph not shown) which in turn overlaps with F-9.

F9 A15

C-C-A-A-G-C-T-G-T-C-G-C-T-A-C-T-T-C-C-C-A

D22a

Figure 4 : The sequence of *Hin*dII-*Hae*III 3-75 and "*Hin*d 8" restriction fragments. The sequence was deduced by the Gilbert and Maxam technique (1977) for the fragment *Hin*dII-*Hae*III 3-75 and from the overlaps of the oligonucleotides (indicated by brackets) listed in tables I and II for *Hin*d 8.
The bracket accompanied by an arrow ⌊↑⌋ indicates a restriction enzyme recognition sequence and the position of hydrolysis.
The letters A,B,C,D, and F refer to different experiments.


A36 B14 A32 B4
D8 F21 C2
HindII/HaeIII 3-75
AluI/Hind 8-43S
GGCC GTTGATGCTAAAGGTGAGCCGCTTAAAGCTACCAGTTATATGGCTGTTGGTTTCTATGTGGCTAAATATGTTAACAAAAAGTCAG
HaeIII
AluI 9/11b
HindII 7/4
10 20 30 40 50 60 70 80 90
CCGG CAACTACGATTTCCACTCGGCGAATTTCGATGGTCAATATACCGACAACCAAAGATACACCGATTTATACAATTGTTTTTCAGTC
HaeIII/HindII 7-75L
F1,D16,D11
HaeIII/HindII 7-75S
A13
B17 A38 B12 B9 A34b B6 F12
A23 A18 A16 A9
B8 A15 A11
A36 B14 A32 B4
AluI/Hind 8-43L
D13
F9
D25 F17 D24 F16 D18 F11
AluI/Hind8-43S
A37,B18
D28 D17 F10 D20
D22a
ATATGGACCTTGCTGCTAAAGGTCTAGGAGCTAAAGAATGGAACAACTCACTAAAAACCAAGCTGTCGCTACTTCCCAAGAAGCTT
AluI 11b/14b
AluI 14b/16
HindIII
100 110 120 130 140 150 160 170
TATACCTGGAACGACGATTTCCAGATCCTCGATTTCTTACCTTGTTGAGTGATTTTTGGTTCGACAGCGATGAAGGGTTCTTCGAA
F1 D16 D11
A8
A14
A46 A34a
A44 A31
AluI/Hind 8-21
A13
B13 A21
A26 A20
A48
D12 D9
D22b D15
A47 A42 B11 A33 C5
D26 D27 F18

On the complementary strand A34a, with 5T's, overlaps D13 with 5A's on the viral strand. Orientation was possible since both sequences in spot A34 contained 5T runs and were easily distinguished from each other (fig. 3). Since A34a is presumably derived by alkylation of G, a G residue can be predicted at each terminus (see below). The 5' end overlaps with A48 (radioautograph not shown) which in turn must have an additional G at the 5' end providing an overlap with the C-C at the 5' end of F9 on the viral strand.

D13

G-A-C-T-A-A-A-A-A-

G-A-G-T-G-A-T-T-T-T-T-G-G

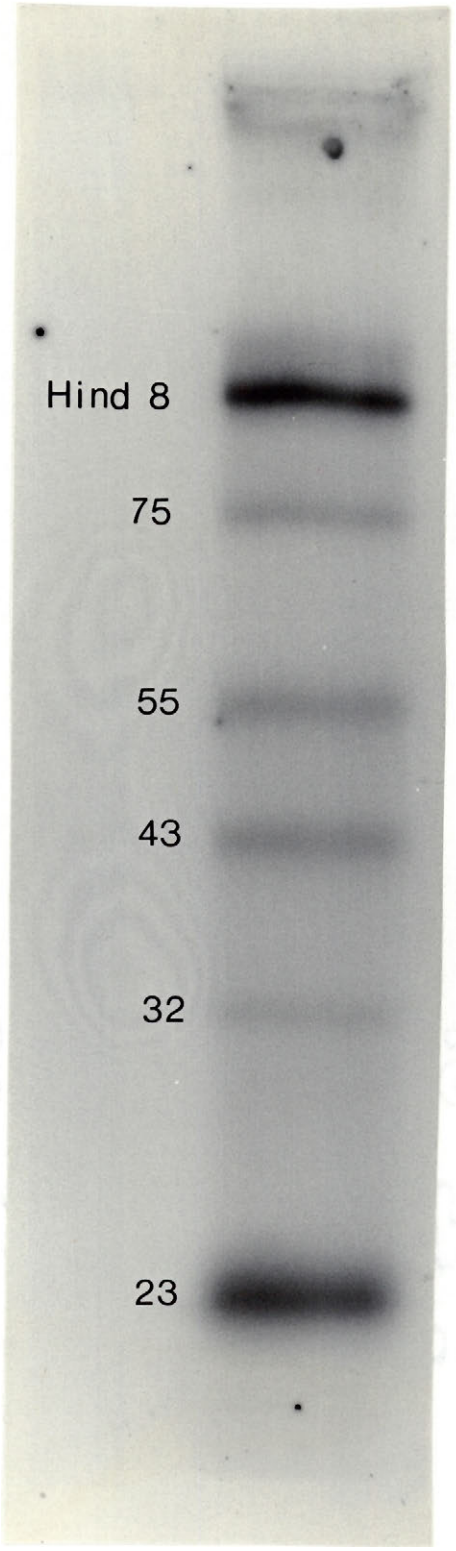
A34

A48

There are no other C-C's or G-G's in the entire *Hin*d 8 sequence which would fit with this particular overlap. F9 contains the sequence A-G-C-T which is an *Alu*I site, and this and the site at residues 119 to 123 provided the means to confirm placement of the various overlapping oligonucleotides. Referring to figure 1 which shows the location of the *Alu*I sites in the sequence (see also figure 4) *Hin*d 8 has two *Alu*I sites; *Alu*I 11b/14b, and *Alu*I 14b/16 (The fragment numbers refer to the *Alu*I map). If these sites are in the correct place, a partial hydrolysis of *Hin*d 8 by *Alu*I should result in an undigested band of *Hin*d 8 of 98 bp, a band of a 75 bp fragment (the 43 bp plus 32 bp units), a 55 bp fragment (the 32 bp plus 23 bp units) a 43 bp fragment and bands of 32 bp and 23 bp. The products of a partial hydrolysis of *Hin*d 8 with *Alu*I separated by gel electrophoresis are shown in figure 5 and are exactly those predicted.

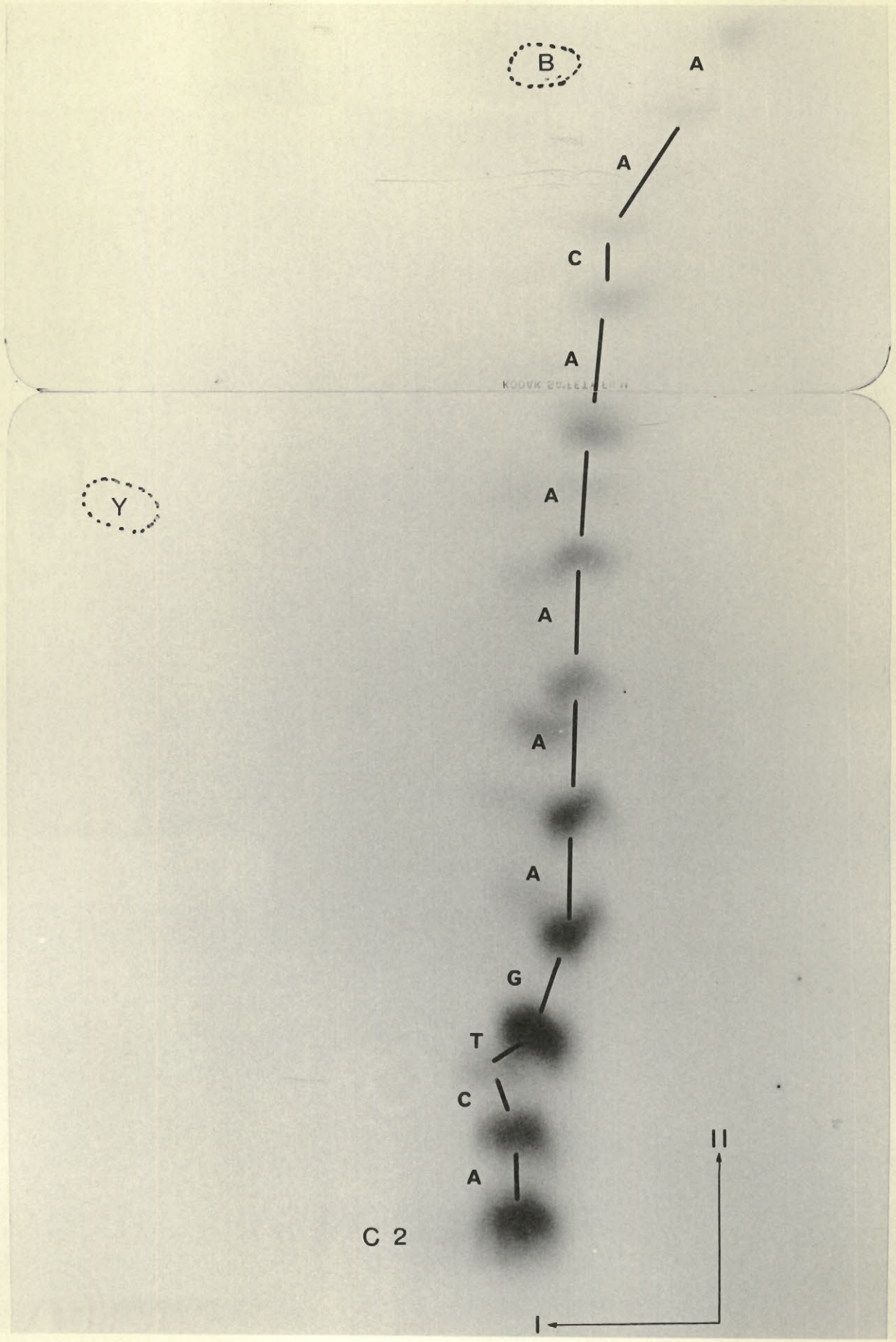
Figure 5 : Partial digestion of *Hin*d 8 with *Alu*I.

Polyacrylamide gel electrophoresis of restriction fragment *Hin*d 8 partially digested with *Alu*I.


Hind 8
75
55
43
32
23

The 5' terminal sequences of the viral and complementary strands, of *Hin*d 8 were confirmed by 5' ^{32}P terminal labeling of *Hin*d 8, alkylation and alkaline hydrolysis of the labeled fragment, separation of the hydrolisate by ionophoresis-homochromatography, scraping of the ^{32}P-labeled spots and sequence analysis by partial snake venom phosphodiesterase hydrolysis and the wandering spot technique. Examples of two analyses are shown in Fig. 6 (C5 and C2). Oligonucleotide C5 overlaps complementarily A15 (see above) and with the additional 3' G residue assumed on the basis of the method of isolation, further confirmation of the orientation, viral verus complementary strand, and sequence position is obtained. Oligonucleotide C2 overlaps complementarily A13, D16, D11, and F1, and provided similar confirmation of the 5' viral terminal sequence. Although the T4 endonuclease IV and alkylation generated oligonucleotides overlapped and provided large areas of sequence data of the *Hin*d 8 fragment, there were still areas missing, such as at positions 98 and 104, areas where the sequence was for one strand only, such as at positions 159 to 166, and problem areas such as at position 111 (Tables I and II, fragments A8, A21, B13, D9, and D12) with a possible but not confirmed difference to the ϕX174 sequence for this area. To confirm the overlap sequence data and extend the results to the adjacent fragment 5' to *Hin*d 8 (figure 1), linking up to the previously sequenced *Hae*III 8 and *Hae*III 9 fragments, the Maxam and Gilbert alkylation - hydrazinolysis sequence analysis method was applied to both *Hin*d 8 and the 75 base pair fragment generated by *Hin*dII hydrolysis of *Hae*III 3 (or alternatively *Hae*III hydrolysis of *Hin*dII 7). The results are presented in figures 7 and 8 and the sequences in Table III. The sequence overlaps with the other data are shown in figure 4.

Figure 6 : Sequence chromatograms of oligonucleotides generated by methylmethane sulfonate of fragment *Hin*d 8 labeled only at its 5' ends.


B
A
A
C
A
Y
A
A
A
A
G
T
C
A
II
I
C 2

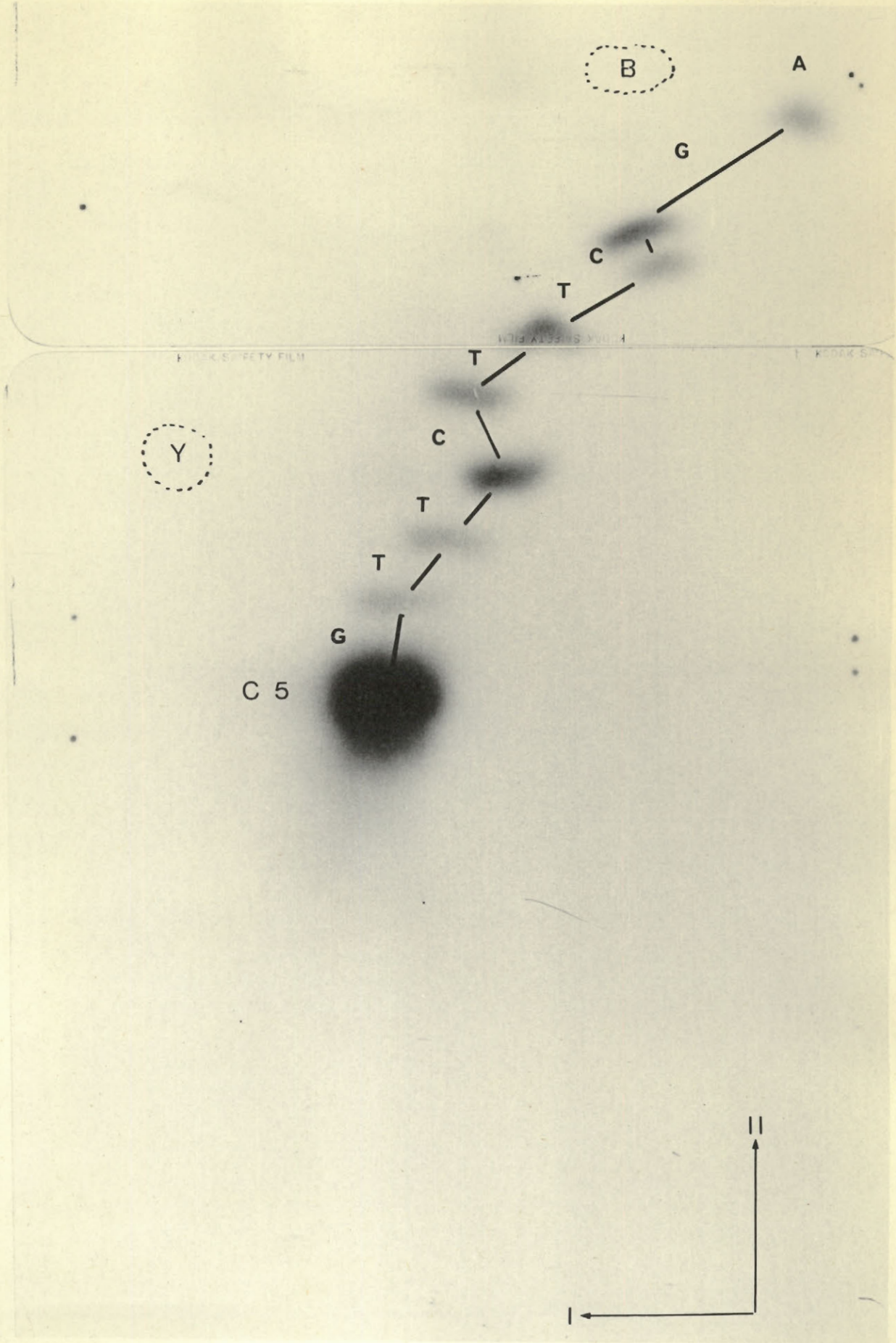
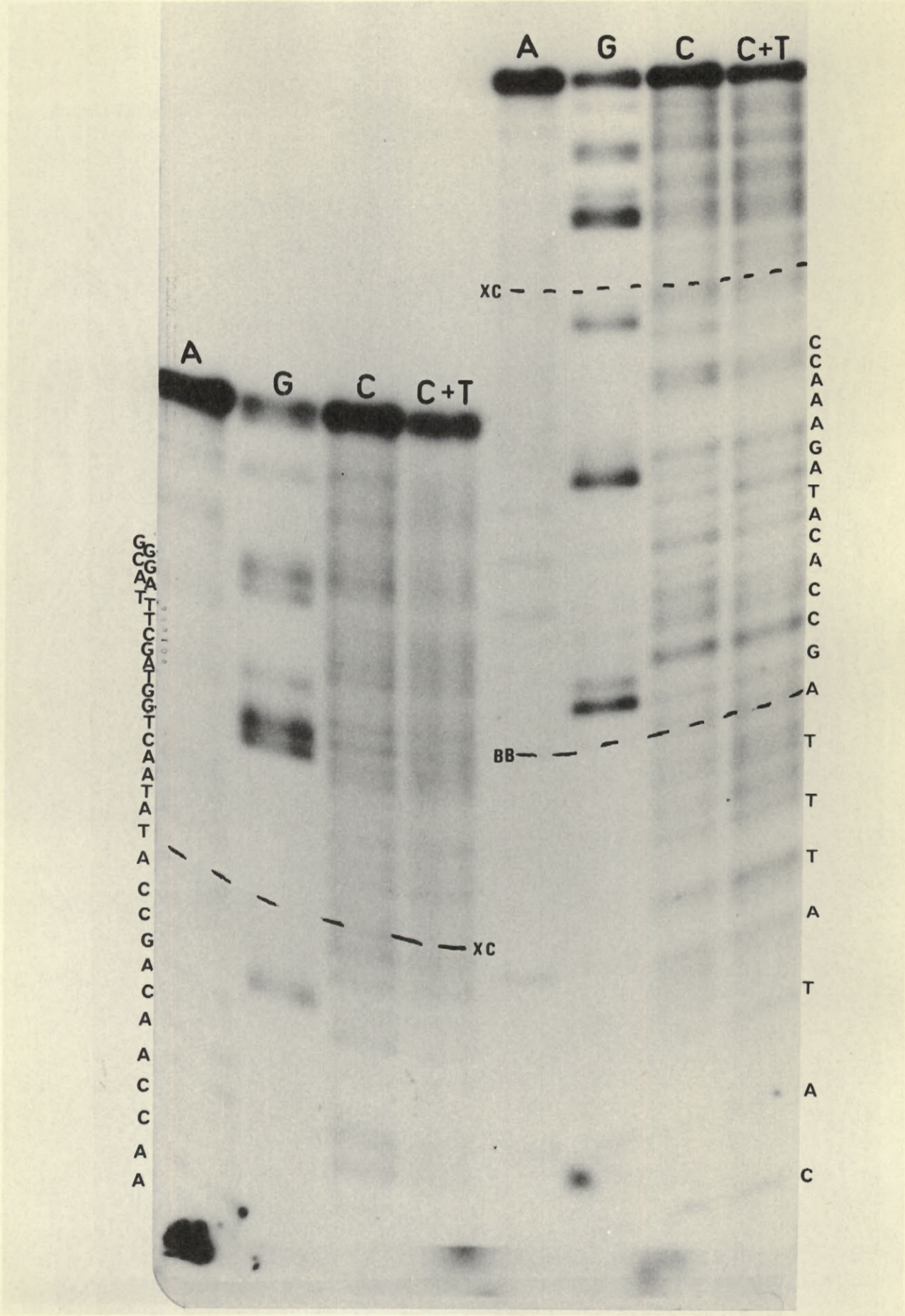

B
A
G
C
T
T
C
T
T
G
C 5
Y
II
I

Figure 7 : Sequence autoradiogram of the Gilbert-Maxam alkylation and hydrazinolysis - rapid gel method on the 75 base pairs fragment overlap of *Hin*dII-7 and *Hae*III-3.

p. 127: a long and a short electrophoresis of *Hae*III/*Hin*d 7.

p. 128: *Hin*dII/*Hae*III-3.

BB: bromophenol blue - XC cyanol FF


A
G
C
C+T
XC
BB
GGAATTCGATGGTCAATATACCGACAACCAA
CCAAAGATACACCGATTTTAATC

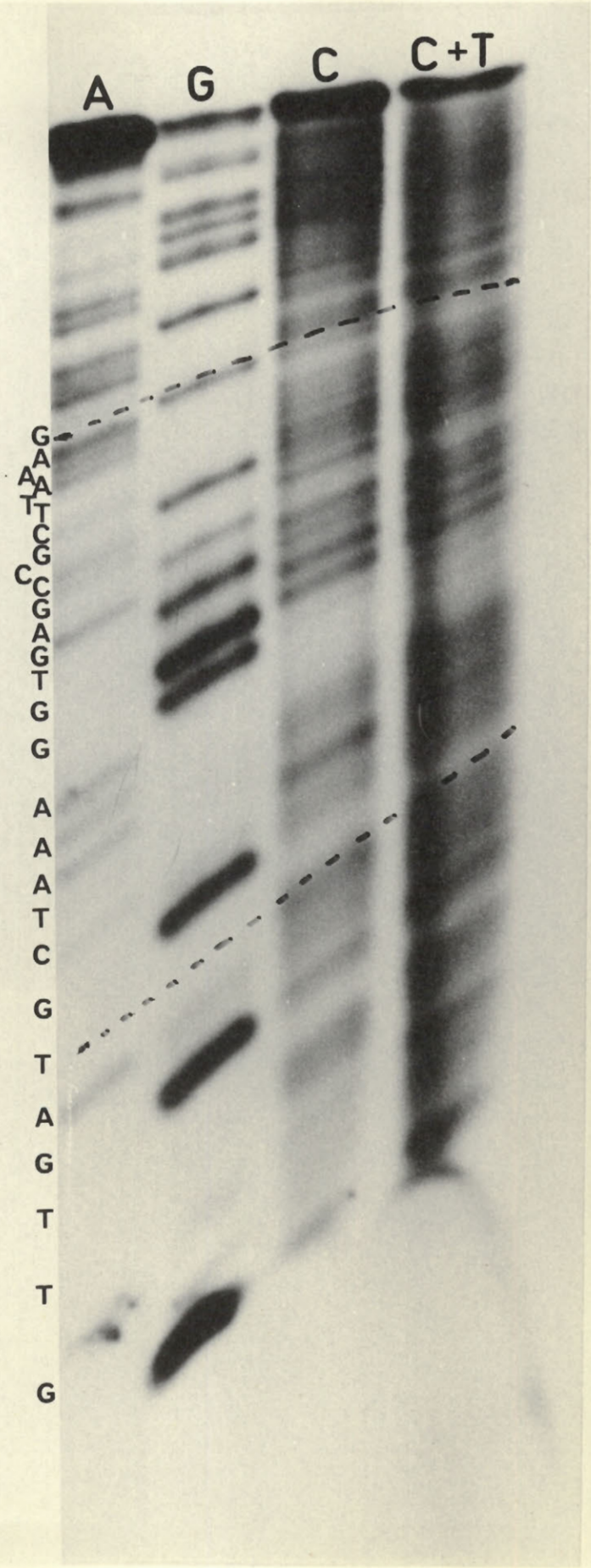
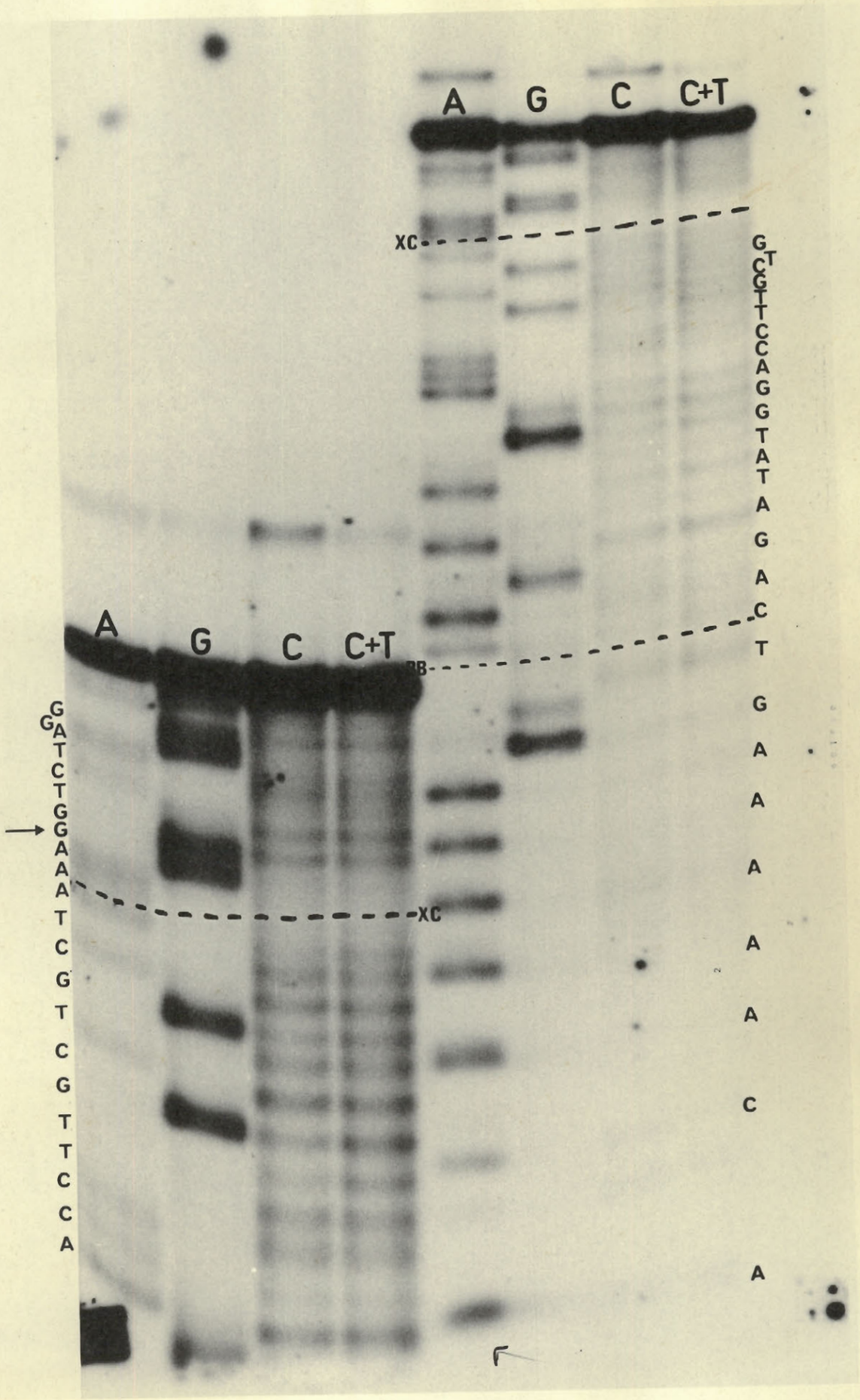

A
G
C
C+T
G
A
A
T
C
G
C
G
A
G
T
G
G
A
A
A
T
C
G
T
A
G
T
T
G

Figure 8 : Sequence autoradiogram of the Gilbert-Maxam alkylation and hydrazinolysis - Rapid gel method on *Hin*d 8.
The fragment was digested with *Alu*I to provide smaller fragments.
p. 129: a long and a short electrophoresis of *Alu*I/*Hin*d 8 (43 bp).
p. 130: *Alu*I/*Hin*d 8 (21 bp).


A
G
C
C+T
XC
GTCGTTCCAGGTATAGACTGAAAAAAAAC
A
G
C
C+T
BB
XC
GGATCTGGAAATCGTCGTTCCA

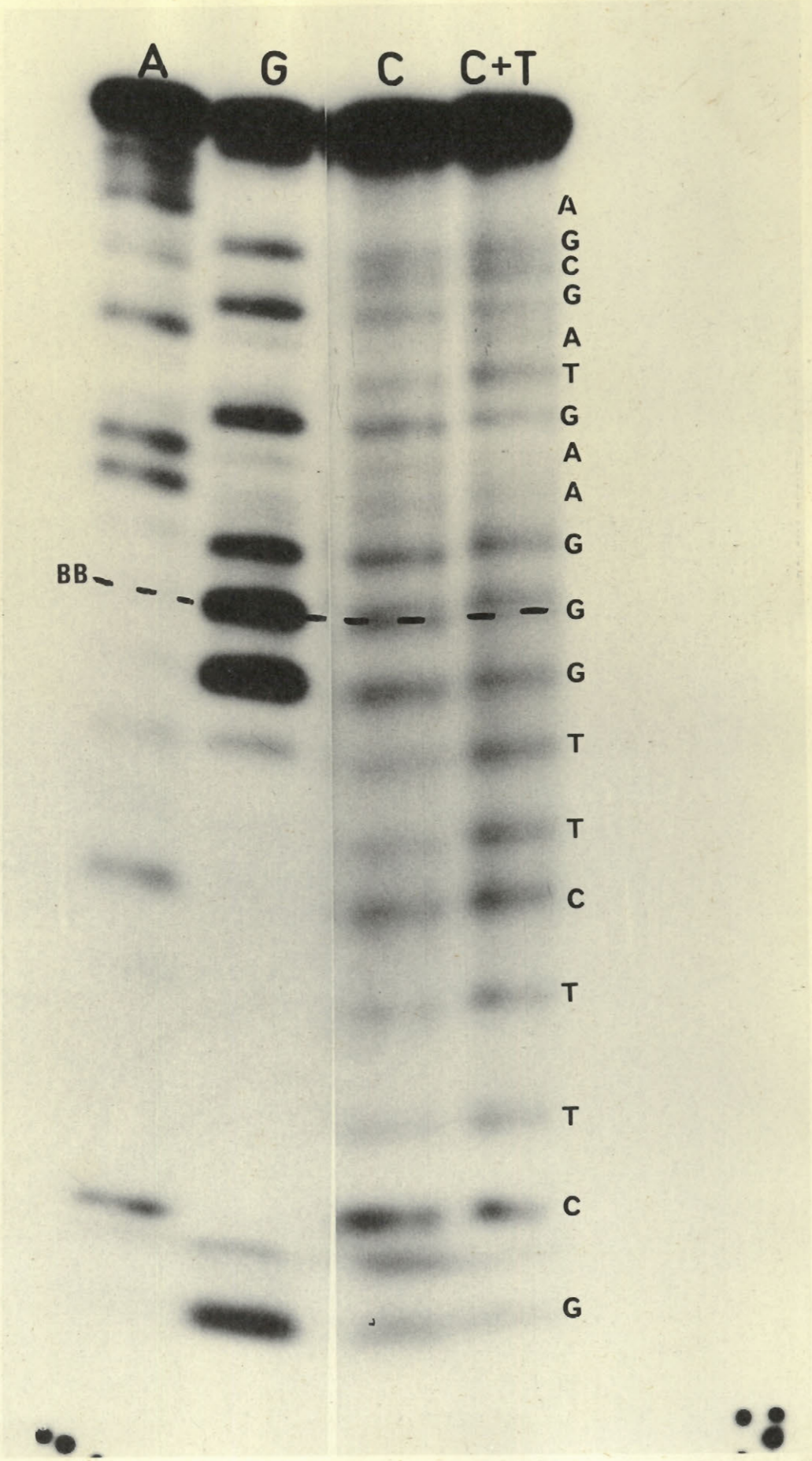

A
G
C
C+T
BB
A
G
C
G
A
T
G
A
A
G
G
G
T
T
C
T
T
C
G

TABLE III

Sequences determined by the Maxam-Gilbert Alkylation-Hydrazinolysis-Rapid Gel Method

Fragment	Sequence
*Hin*dII/*Hae*III 3-75	G-T-T-G-A-T-G-C-T- A-A-A-G-G-T-G-A-G- C-C-G-C-T-T-A-A-A-G
*Hae*III/*Hin*dII 7-75S	C-A-T-A-T-T-T-A-G- C-C-A-C-A-T-A-G-A- A-A-C-C
*Hae*III/*Hin*dII 7-75L	A-A-C-C-A-A-C-A-G- C-C-A-T-A-T-A-A-C- T-G-G-T-A-G-C-T-T- T-A-A-G-C-G-G
*Alu*I/*Hin*d 8-43S	A-C-A-A-A-A-A-G-T- C-A-G-A-T-A-T-G-G- A-C-C-T-T-G-C-T-G
*Alu*I/*Hin*d 8-43L	A-C-C-T-T-G-C-T-G- C-T-A-A-A-G-G-T-C- T-A-G-G
*Alu*I/*Hin*d 8-21	G-C-T-T-C-T-T-G-G- G-A-A-G-T-A-G-C-G-A

Sequences underlined are overlaps either ———— for short and long gel runs, or ∿∿∿∿ for complementary strand overlaps.

*Hin*dII cleavage of S13 *Hae*III 3 previously 5' ^{32}P-labeled with polynucleotide kinase and [γ-^{32}P]ATP provided the sequence at the 5' viral end of the 176 nucleotide fragment shown in figure 4 from positions 6 to 33 on the viral strand. The oligonucleotide penultimate to the *Hae*III restriction sequence G-G-C-C was not seen on the gel. The first two or three nucleotides are difficult to determine with this method of analysis. It is assumed there is only one nucleotide missing at this point, based on the sequence analysis patterns of other gels in this series of experiments. Hydrolysis of 5' ^{32}P-labeled S13 *Hin*dII 7 with *Hae*III (fig. 1), gave the sequence starting from the *Hin*dII 7/4 site in the 5' → 3' direction on the complementary strand of the DNA. The short (S) and long (L) gel separations (fig. 7) gave overlapping sequences from nucleotides 25 to 76 inclusive, which in turn, overlapped complementarily with the sequence on the viral strand, positions 6 to 33, resulting in the complete sequence of the 75 bp fragment (fig. 1). The 5' end of the complementary strand sequence overlapped into the *Hin*dII 7/4 site. Alkylation with methylmethane sulfonate and T4 endonuclease IV hydrolysis of the 75 bp fragment has also been performed (B. Goodchild, unpublished results) and these data confirm the results in figure 4.

5' ^{32}P-labeled *Hin*d 8 digested with *Alu*I (figure 1) provided two fragments for analysis by the Maxam and Gilbert technique. Fragment *Alu*I/*Hin*d 8-43 gave the sequence from positions 79-118 inclusive on the viral strand from overlaps of nucleotides 79-105 inclusive deduced from the short gel separation and 97-118 inclusive from the long gel extension (fig. 8). This sequence included nucleotide 111 referred to previously, which is marked by an arrow on figure 8. There is no difference from the

ϕX174 sequence at this position. Fragment *Alu*I/*Hin*d 8-21 gave the sequence from positions 157-175 inclusive as seen in figure 4. The *Alu*I/*Hin*d 8 sequencing gels confirmed sequences at both ends of *Hin*d 8 and provided the necessary overlaps to confirm the total sequence of this fragment.

To determine the sequence of the polymerase binding site on S13 *Hin*d 8, a sample of the fragment was incubated with *E. coli* RNA polymerase and then treated with DNase A to hydrolyse the unprotected DNA. The polymerase-DNA complex was purified, hydrolyzed with T4 endonuclease IV, the oligonucleotides 5' ^{32}P-labeled, separated and sequenced by partial snake venom phosphodiesterase and the wandering spot method. The results are presented in Table IV. The majority of the sequences correspond to the 5' viral strand end of the *Hin*d 8 fragment and the positions are shown in figure 9. In this experiment the oligonucleotides do not have C at the 5' terminus (Table IV) because some were generated by DNase A cleavages and others are from the 5' end of the fragment (fig. 9).

TABLE IV

Sequences of the oligonucleotides released by T4 endonuclease IV from that part of *Hin*d 8 protected by *E. coli* RNA polymerase from DNase A digestion

G 1	A-G-C-T-T-C-T-T-G-G-G-A-A
G 2	A-A-C-A-A-A-A-A-G-T-C-A
G 3	A-A-C-A-A-A-A-A-G-T-C
G 4	A-A-C-A-A-A-A-A-G-T
G 6	T-T-T-T-T-C-A-G-T-C-T-A
G 7	A-A-C-A-A-A-A-A-G
G 9	A-A-A-A-G-T-C-A-G-A
G10	T-T-G-G-G-A-A-G-T
G13	A-G-T-C-A-G-A-T-A-T-G
G17	T-C-C-A-T-A-T-C
G21	A-G-C-A-G-$\begin{matrix}A\\C\end{matrix}$

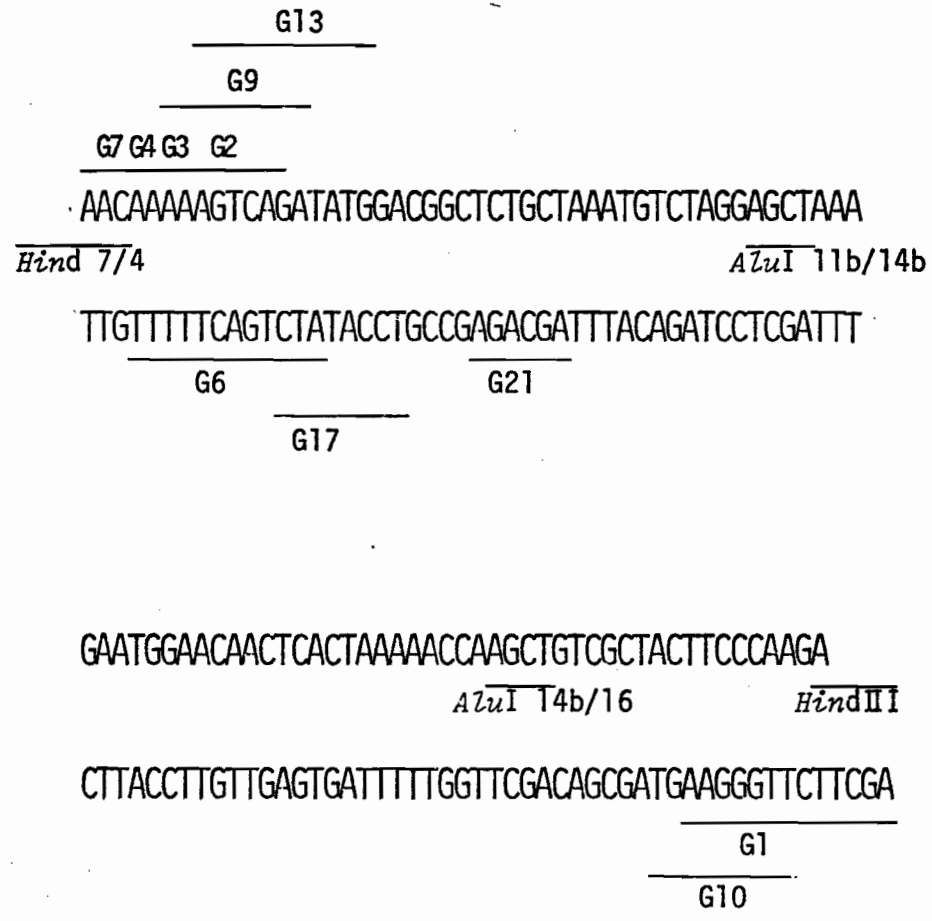


Figure 9 : *Hind* 8 sequence representing the T4 endonuclease IV oligonucleotides obtained after RNA polymerase binding on *Hind* 8 and DNAse digestion.

The oligonucleotide sequences are listed in table IV.

DISCUSSION

Use of the single strand specific enzyme T4 endonuclease IV to degrade the denatured strands of restriction fragment *Hin*d 8 demonstrated the feasability of this approach to generate oligonucleotides which can be used for sequence analysis. Certain areas of the DNA strands were not amenable to hydrolysis presumably because of more rapid renaturation in these regions or structural features which interfered with the nuclease cleavage (see below).

The specificity of methylmethane sulfonate alkylation of G residues generated alkaline cleavage sites in *Hin*d 8 and provided a different set of overlapping oligonucleotides to those from the T4 endonuclease IV hydrolysate. The different base specificity of the two reactions was confirmed further by the sequence of the nucleotides generated and the placement of overlaps within the sequence of the total fragment (fig. 4). All the oligonucleotides from the alkylation reaction were assumed to be flanked at their 5' and 3' ends by guanine residues. These residues are not given in the table of sequences (Table II) since they were not assayed directly. It should not be inferred that A residue depurination did not occur. However, the major sites for alkylation were the G residues and any A alkylations that occurred were lost in the background noise of the method. In a DNA with a high A + T content (>90%) this may not be the case (K. Harbers and J.H. Spencer, unpublished observations).

The properties of the S13 *Hin*d 8 restriction fragment such as electrophoretic mobility, base composition, and sequence indicate no unusual structural features. However, even though the methylmethane sulfonate alkylation procedure is a chemical degradative method, there were areas

of the DNA fragment which did not give rise to oligonucleotide spots of a significantly frequent occurence for identification. These included viral strand residues 96 through to 118, 127 through to 154, and complementary strand residues 99 to 113. These areas were also absent in T4 endonuclease IV digests. Thus certain areas are more prone to hydrolysis than others. Different conditions of alkylation and hydrolysis may result in these areas appearing as spots. It is possible that they were hot spots for alkylation, since each is a purine rich area and this would coincide with the actual results. If no cleavage had occurred in these regions, there should have been an enrichment of oligonucleotides from these areas resulting in spots intense enough on the radioautographs for delineating and sequencing.

The culmination of the two methods was that most of the *Hin*d 8 sequence was definable. The area around residue 100, was confirmed by the Maxam and Gilbert technique (1977). The ambiguities and problems inherent in the wandering spot technique (Wu *et al.*, 1974) were clearly apparent with the interpretation of residue 111 as an A rather than a C on the complementary strand which, by the Maxam and Gilbert technique (1977), was determined to be a G residue in the viral strand. The use of the restriction enzyme digestion with *Alu*I to confirm the orientation and organization of overlapping oligonucleotides of *Hin*d 8 was a further example of the usefulness of these enzymes in many aspects of sequence analysis.

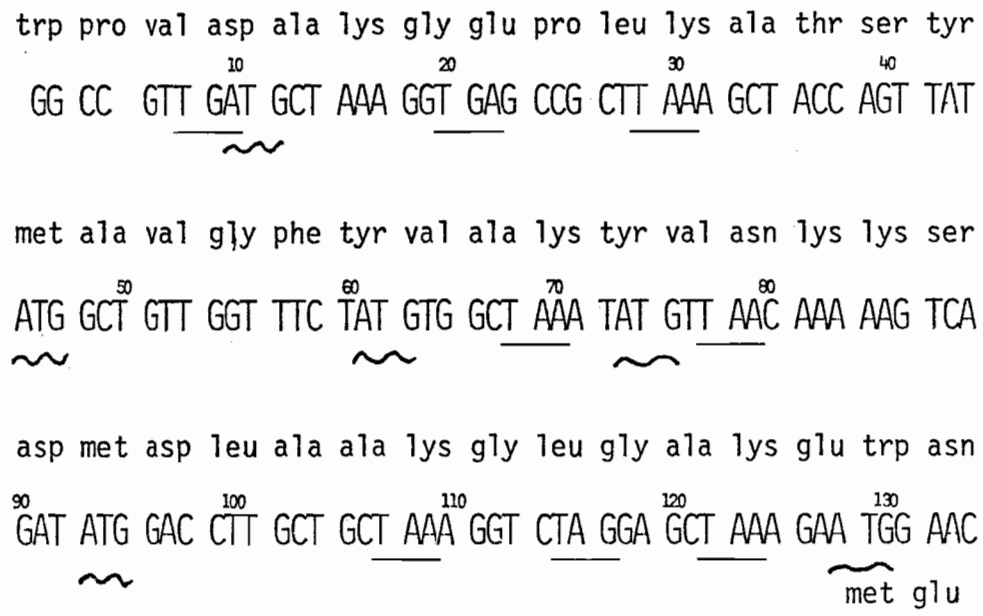
The sequence for residue 1 through to residue 75 determined by the Maxam and Gilbert technique (1977), is contiguous with the sequence of S13 *Hae*III 8 and *Hae*III 9 restriction fragments reported previously (Grosveld and Spencer, 1977). The amino acid readout of the *Hae*III 8 and

*Hae*III 9 sequence revealed the possibility of a DNA - protein binding site in that area (Grosveld and Spencer, 1977). In the same reading frame, the amino acid readout is continuous throughout the present sequence of 175 nucleotides and is given in figure 10. This provides circumstantial evidence that the sequence reported here is still within the A gene, since a continuous protein readout would be one prediction from this assumption. Comparing the A protein readout (fig. 10) with that for ϕX174, there are no amino acid changes (Sanger *et al.*, 1977). The two sequence differences from that of ϕX174 are at residues 74 and 176. In the viral strand the ϕX → S13 changes are C → T and G → T respectively. In the A protein readout, these are both third position changes and both increase the T rich predominance in the third position of the reading frame of the A protein noted elsewhere (Sanger *et al.*, 1977). Examination of the other amino acid readout frames, reveals ATG initiation sites in one reading frame at positions 10, 61, and 73. The same frame has stop codons at positions 115 and 142, eliminating this frame as a possibility for amino acid readout for the B gene in this area of S13. However, a small protein of 35 residues, maximum, would be a possibility and cannot be discounted at this time.

The third reading frame has an initiation codon at position 128 with no stop codons subsequent to that initiation codon. This could correspond to the start of the B gene protein. The site is identical to the proposed start of the B gene protein in ϕX174 (Sanger *et al.*, 1977). The amino acid sequence for this presumptive B protein start is given in figure 10 also. The sequence is identical to the equivalent sequence in ϕX174 (Sanger *et al.*, 1977) thus the difference in mobility of the S13 and ϕX174

Figure 10 : Sequence of *Hin*dII/*Hae*III 3-75 and *Hin*d 8 with the corresponding amino acid sequence of the A protein (at the top position). The B protein amino acid sequence starts at position 128 overlapping with the A protein amino acid sequence.

The sequences underlined (____) are stop codons and initiator codons (~~~).


trp pro val asp ala lys gly glu pro leu lys ala thr ser tyr
10
20
30
40
GG CC GTT GAT GCT AAA GGT GAG CCG CTT AAA GCT ACC AGT TAT
met ala val gly phe tyr val ala lys tyr val asn lys lys ser
50
60
70
80
ATG GCT GTT GGT TTC TAT GTG GCT AAA TAT GTT AAC AAA AAG TCA
asp met asp leu ala ala lys gly leu gly ala lys glu trp asn
90
100
110
120
130
GAT ATG GAC CTT GCT GCT AAA GGT CTA GGA GCT AAA GAA TGG AAC
met glu
asn ser leu lys thr lys leu ser leu leu pro lys lys leu
140
150
160
170
AAC TCA CTA AAA ACC AAG CTG TCG CTA CTT CCC AAG AAG CTT
gln leu thr lys asn gln ala val ala thr ser gln glu ala

B proteins in SDS gels is not due to changes at the NH_2 terminal end of the B protein, as has been suggested (Tessman *et al.*, 1976), but must be due to changes internal in the proteins or to post translational modifications, perhaps glycosylation. The proposal for this site for the start of the B protein provides circumstantial evidence that the A and B genes of S13 overlap, as in ϕX174 (Sanger *et al.*, 1977; Weisbeek *et al.*, 1977).

Initial interest in the sequence of *Hin*d 8 was because of the polymerase binding site localized in this area of S13 DNA. RNA polymerase binding is equated with promoter sequences which include the canonical sequence first identified by Pribnow in T7 DNA T-A-T-Pu-A-T-G (Pribnow, 1975). A sequence similar to this, T-A-T-G-T-T-A, is present at position 72 through to position 78 inclusive (Note that residue 74 is a C in ϕX174). This sequence is very similar to that found in the *lac* promoter, T-A-T-G-T-T-G (Dickson *et al.*, 1975), and the *gal* promoter (Musso *et al.*, 1977). The location of a "Pribnow" box at this position fits with the polymerase binding data to *Hin*d 8 and *Hae*III 3 and the binding of RNA polymerase to the 5' viral end of restriction fragment *Hin*d 8, as shown in the present study. This places the putative promoter site of the B gene of S13 approximately 50 nucleotides from the B protein translation start, and approximately 100 nucleotides 3' on the viral strand to the equivalent promoter site suggested for the B gene of ϕX174 (Smith and Sinsheimer, 1976b; Sanger *et al.*, 1977). The *Hin*dII cleavage site in this region may affect transcription initiation similar to the R-Hha site in the coat protein gene promoter of bacteriophage fd (Seeburg *et al.*, 1977). This may provide a means of testing for promoter function of the B gene promoter in S13.

In comparing the similarity of the sequence of ϕX174 with that of S13, the 364 nucleotides of S13 reported to date have 5 base changes from the equivalent sequence in ϕX174 (Kaptein, 1977). These results are not compatible with the heteroduplex analyses of S13 and ϕX174 from two laboratories which have ascribed a sequence similarity for the two phages at about a level of 35% mismatch (Godson, 1973; Compton and Sinsheimer, 1977). The present data give a mismatch of 1.4%. Also the regions of S13 sequences do not coincide with the regions of high homology delineated by the heteroduplex analyses. Other more extensive restriction enzyme mapping studies indicate that S13 and ϕX174 are very similar to each other in overall nucleotide sequence with an overall level of mismatch of the order of 5% (Spencer *et al.*, 1978).

Dr. Klaus Harbers has contributed to the work of this chapter by confirming the sequence of *Hin*d 8 and by sequencing *Hin*dII-7/*Hae*III 3 (75 bp) by the Maxam-Gilbert Alkylation-Hydrazinolysis-Rapid Gel Method.

CHAPTER V

DISCUSSION

1. The RNA polymerase binding sites of S13 and ϕX174.

When *E. coli* RNA polymerase is bound to S13 RF I DNA and the unprotected DNA digested with DNAse, about 2% of the starting material remains undigested. Since the RNA polymerase-protected DNase fragment is 35 to 40 nucleotides long (chapter II and appendix I, Figure 4), one can calculate that three moles of RNA polymerase are bound per S13 DNA molecule. This indicates that there are three major RNA polymerase binding sites on S13. The pyrimidine oligonucleotide composition of the RNA polymerase protected sequences was shown to be different from the pyrimidine oligonucleotide composition of S13 (appendix I, Figures 5a and b), showing that the protected sequences were specific parts of S13. However, this king of experiment was unable to give information as to the location of those sites on the S13 genome.

a) Visualization by electron microscopy.

We used the restriction enzyme *Pst* 1 (*Providentia Stuartii*) to cut both S13 and ϕX174 replicative forms at a unique site (Godson and Roberts, 1976). After hydrolysis, a linear molecule is obtained which is much more appropriate for mapping studies and measurement. After binding the RNA polymerase on S13 or ϕX174, the electron microscope reveals several binding sites. The histogram (chapter III, Figure 2) reveals 2 major peaks and 3 smaller peaks of intensity on both S13 and ϕX174.

b) Localization using the restriction enzymes.

In an attempt to localize the RNA polymerase binding sites on the S13 genome, restriction enzymes seemed to be an excellent tool to digest double stranded DNA into a specific set of fragments, which can be aligned as a physical map (Lee and Sinsheimer, 1974; Grosveld *et al.*, 1976).

Once the physical map is known, the fragments can be studied for their ability to bind the RNA polymerase. With some little differences, we obtained essentially the same results as those obtained using electron microscopy; three major binding sites for S13 with two weaker sites and five binding sites for ϕX174. The first strong binding site of S13 is located in the overlap of *Hin*d 1 and *Hae*III 2, coinciding with the start of gene A. It corresponds to the peak located at a position between 70 and 75% of the genome obtained by electron microscopy. The second site is located at the overlap of *Hin*d 8 and *Hae*III 3. This binding site coincides with the beginning of gene B, and has also been detected as a strong binding site by electron microscopy between 92 and 95% of the S13 molecule. The third strong binding site is found at the overlap of *Hin*d 4 and *Hae*III 6 which coincides with the area of gene D and E. This site, mapped by electron microscopy between 7 to 10%, is a strong binding site. We also detected binding at the region of the *Hin*d 2 - *Hae*III 1 overlap and the *Hin*d 3 - *Hae*III 1 overlap but the binding intensity was much lower and may be due to high background since they are all large fragments. These fragments may have structural or sequence similarities to promoter regions. This tends to indicate that genes D, E, J, F, G and H are polycistronic. Vanderbilt *et al.* (1972) showed that the polar effect of nonsense mutations in gene F of phage S13 (mutants *op* nF403 and *am* ngF28) extends to adjacent genes G and H, but not to genes A and B which follow the H gene in the direction of translation. Therefore genes F, G and H should be polycistronic and there should be a promoter site before gene A which is what we have found. However, the overlap between *Hin*d 3 - *Hae*III 1 which coincides with the beginning of gene H binds the RNA

polymerase to a certain extent. Electron microscopy indicates that this region is as high in terms of binding intensity as the site at the beginning of gene A for example. Also in ϕX174, the DNA region preceding gene H binds the RNA polymerase, confirmating the results of Chen *et al.* (1973). It is possible that some structural features in this area which cause a resemblence to the promoter sequence would be the cause of this binding, as discussed later.

In ϕX174, the results with the restriction fragments are not as clear, because of a higher background. This could be due to nonspecific binding of the RNA polymerase, also the restriction fragments (for the *Hae*III digest) are closer to each other in size which makes the extraction of the bands from the polyacrylamide gel more delicate and the quantitation more difficult. In addition, the observation of the histogram in the electron microscopy technique (chapter III, Figure 2) shows that the intensity of the different peaks is more uniform for ϕX174 as compared to S13. The first site is located at the overlap of *Hin*d II 4 and *Hae*III 2 fragments which coincides exactly with the site of S13 between *Hin*d I and *Hae*III 2 at the beginning of gene A. This site, mapped by electron microscopy between 69 and 74%, confirms one of the sites obtained by Chen *et al.* (1973) and correlates with the location by hybridization of ϕX174 RNA transcripts in relation to restriction fragments, (Smith and Sinsheimer 1976b; Axelrod 1976; Hayashi *et al.*, 1976). The second site in ϕX174 is found at the overlap of *Hin*d II 5 and *Hae*III 3, which coincides with the start of gene B. Chen *et al.* (1973) did not detect this site, but this may be due to the fact that their binding experiments were done at very high RNA polymerase: DNA ratios and low salt which favours nonspecific

binding. This site was detected as a strong site by electron microscopy between 93 and 95% of the ϕX174 genome. Smith and Sinsheimer (1976b) using *in vitro* mRNA hybridization placed the site for the B gene in ϕX174 within the *Hin*d II 8 fragment and the *Hae*III 3 fragment. However the *Hin*d II 8 fragment in ϕX174 does not coincide with the S13 *Hin*d 8. The location of the site is further complicated by the sequence studies of Sanger's group (1977) which placed the putative mRNA initiation site of gene B in *Hae*III 10. However the *Hae*III 10 location conflicts with the data of Smith and Sinsheimer (1976b) which did not detect any hybridization of mRNA with restriction fragment *Hae*III 10. Moreover, Axelrod (1976a and b) using mRNA transcript hybridization obtained results which do not completly agree with those studies. Reinvestigation of the location of the ϕX174 gene B promoter should be done in order to classify the discrepencies between the results of Sinsheimer's group (1976b), Sanger's group (1977), Axelrod (1976b) and our results (chapter II).

The conclusion is that in S13, the promoter site for the B gene is located at least 50 nucleotides further on the 3' side as for ϕX174. Another binding site which mapped as a strong one by electron microscopy at a position between 8 to 10% of the genome is located at the overlap of *Hin*d II 6 and *Hae*III 7. The site corresponds to the site of S13 covering genes D and E. Chen *et al.* (1973) and Smith and Sinsheimer (1976b) detected a site at the overlap of *Hin*d II 6 and *Hae*III 3 but did not observe significant binding of RNA polymerase or hybridization of RNA transcripts to *Hae*III 7. However, Axelrod (1976b) observed hybridization of GTP initiated, 5' labeled mRNA to *Hae*III 7 which correlates well with our observations. The two remaining binding sites of ϕX174

are not as strong and are located at the overlap of *Hin*d II 1 and *Hae*III 4 corresponding to the area of *Hin*d 2 - *Hae*III 1 in S13 and the peak between 25 and 30% obtained by electron microscopy. The other site located at the overlap *Hin*d II 2 - *Hae*III 1 corresponds to the weak binding site *Hin*d 3 - *Hae*III 1 in S13 and to the peak between 50 and 55% visualized by electron microscopy. This site is in the position of the gene G - H functions was also observed by Chen *et al.* (1973) and it has been suggested by Sanger *et al.* (1977) that this binding site is the result of structural characteristics similar to promoters. Examination of the sequence of ϕX174 reveals a very A + T rich composition between positions 2877 to 2894 (16 nucleotides out of 18 are A or T) 26 nucleotides before the start of the H protein (Sanger *et al.*, 1977). Therefore, it seems that this area shows a high homology with promoter sequence and could be an RNA polymerase binding site but not a promoter.

The present study, either by electron microscopy, or with the use of restriction enzymes shows a high degree of homology in the location of the RNA polymerase binding sites between ϕX174 and S13. However, this study indicates that the B gene promoter site in S13 is closer to the initiation sequence of the B protein. Thus structural changes should be selected in this area. Fragments S13 *Hae*III 8 and 9 have been sequenced (Grosveld and Spencer, 1977) but very few differences were detected after comparison with the ϕX174 sequence. The S13 *Hin*d 8 restriction fragment becomes automatically of interest to sequence, first of all, because it contains an RNA polymerase binding site (chapter II and III) and also because of the discrepancy between ϕX174 and S13 in this area. Tessman *et al.* (1976) reported a different electrophoretic mobility in polyacrylamide gels of

B proteins from ϕX174 and S13, although these can replace each other. They showed that the mobility increase of the B protein of S13 versus that of ϕX174 corresponds to a difference of about 2000 daltons. The B protein of S13 is 15 amino acids shorter than the ϕX174 protein. This could be explained by postulating a different location for the initiation site of translation for the protein. This could also imply a different location for the promoter site for the B gene in S13.

2. Is S13 *Hin*d 8 an RNA polymerase binding site?

When the RNA polymerase is bound to the restriction fragments obtained by *Hin*d on S13 RFI it can be clearly seen that restriction fragment *Hin*d 8 is retained on the filter and thus was tightly bound to the polymerase (chapter II, table 1 and Figure 3). Quantitative data in table 1 show that the normalized value for *Hin*d 8 is still 1.5 in 0.4 M KCl with method A and is 2.1 and 3.0 in 0.5 M KCl and 1 M KCl respectively with method B, showing definite binding of *Hin*d 8. The restriction fragment S13 *Hae*III 3 also binds the RNA polymerase to a considerable extent (chapter II, table 2). Also, using electron microscopy, we obtained a strong peak of binding intensity between 92 and 95% of the genome length covering the area of *Hin*d 8 in S13. There is definitively no doubt left from these results that the restriction fragment *Hin*d 8 of S13 contains an RNA polymerase binding site. It can always be argued that restriction fragments are not good substrates for the RNA polymerase and that binding of restriction fragments is not representative of an *in vivo* situation but the electron microscopy technique has the advantage that the substrate is a linear double-stranded molecule, very similar to the RF II form which

is probably the *in vivo* substrate for the RNA polymerase.

3. Sequence analysis by T4 endonuclease IV.

The sequence analysis of double-stranded DNA using the T4 endonuclease IV and 3' exonuclease digestion of the fragments has been for a long period of time the only enzymatic sequencing method for DNA. However, the action of the enzyme is restricted to single-stranded DNA. Although the strands of the *Hin*d 8 fragment could not be separated, hydrolysis of the denatured *Hin*d 8 strands was sufficient to produce oligonucleotides suitable for sequencing. T4 endonuclease IV exhibits an absolute specificity for the production of deoxyoligonucleotides ending by a 5' pC (Grosveld and Spencer, 1977). The dinucleotides T-C, G-C, A-C and C-C occur respectively 11, 12, 10 and 8 times in the *Hin*d 8 sequence. Comparison of these frequencies to those of each of the four bases occurring at the 3' ends of the oligonucleotides sequenced shows that T4 endonuclease IV exhibits a preference for the cleavage of T-C (12 times) over A-C (6 times), G-C (4 times) and C-C (2 times). These results are in good agreement with Bernardi *et al.* (1976) and Grosveld and Spencer (1977). Although, for some part of the *Hin*d 8 fragment, the enzyme displayed a poor specificity probably due to a 3' exonuclease contaminant activity (Sadowsky and Bakyta, 1972) or to the fact that we used a double-stranded DNA substrate. The total sequence of the fragment could not be obtained presumably because of more rapid renaturation of some regions of the fragment. Two areas of the fragment did not yield any aligonucleotides between positions 105 and position 133 on the viral strand and between 133 and 175 on the complementary strand. A careful examination reveals

that those two regions compose the most G rich part of the *Hin*d 8 sequence which would support the faster renaturation hypothesis we have formulated. Furthermore, there is not a single C between position 131 and position 152 on the complementary strand of *Hin*d 8. This would yield an oligonucleotide of 21 nucleotides in length and would not be detected in the high background obtained in that size range in the ionophoresis-homochromatography experiment (appendix II, Figure 2).

4. Sequence analysis by specific DNA alkylation and hydrolysis.

Certain alkylating agents such as methyl methane sulfonate display a high degree of specificity in their site of alkylation. They will alkylate G residues almost exclusively (Singer 1975). G and A residues in DNA are alkylated in a ratio of 9 to 1 (Lawley and Brookes, 1963). In contrast, in the Gilbert and Maxam technique (1977) the alkylating agent (dimethyl sulfate) is chosen so that both A and G residues are alkylated but in a ratio of 1 to 5 respectively. This allows them to distinguish on their gels between G residues (dark bands) and A residues (light bands). In DNA, the alkylated purines can be removed and the DNA chain cleaved in alkali. This results in the production of a specific set of deoxynucleotides which can be further analysed by 3' exonuclease digestion. The low alkylation by methyl methane sulfonate at A residues in our experiments yields oligonucleotides at very low concentration compared to the oligonucleotides released after alkylation at G residues so that they do not show up as strong spots on the chromatogram. If we examine all the sequences obtained by DNA alkylation with methyl methane sulfonate (chapter IV, table 2) in respect to the sequence of *Hin*d 8, we

observe that all the oligonucleotides have a G residue adjacent to each end of their sequence except for those oligonucleotides which form the 3' end of the restriction fragment. This is due to a 3' exonuclease contaminant in the *Hin*d II + III restriction enzyme preparation. Again, some areas of the *Hin*d 8 restriction fragment did not yield any oligonucleotides, in particular most of the viral strand, the two ends excepted (positions 95 to 150). A careful study of the sequence in that area reveals a part rich in A and G and another part with no G residues. The part rich in A and G would be either too degraded or produce too much background because of the alkylation of the A residues. The sequence between position 132 and position 153 does not contain any G residues but a large amount of A residues.

5. The sequence of S13 *Hin*d 8.

Using the two methods T4 endonuclease IV and DNA alkylation and hydrolysis, we were able to define most of the *Hin*d 8 sequence. In addition, the 74 base pair fragment between *Hin*d 8 and *Hae*III 8 and 9 sequenced by Grosveld and Spencer (1977) was subjected to the rapid Gilbert and Maxam (1973) sequencing method. The 74 base pair fragment was generated either by a *Hin*d II hydrolysis of *Hae*III 3 or by *Hae*III hydrolysis of *Hin*d II 7. The way the sequences were put together can be seen in chapter IV of this thesis. For the set up of the *Hin*d 8 sequence, some examples are shown in chapter IV. Briefly, all the oligonucleotide sequences were aligned and those which were complementary to each other, formed blocks of sequence. These blocks were attached to each other to form the whole *Hin*d 8 sequence. In some cases, we had sequence information

from only one strand, but it was enough to build the total sequence (chapter IV, Figure 2). It was easy to orient the fragment because of the polarity of the *Hin*d II and *Hin*d III restriction enzyme cleavage sites respectively at the 5' and 3' end of the viral strand. Moreover, the *Hin*d III restriction enzyme produces an asymmetric cleavage and leaves a single strand tail 4 nucleotides in length at the 5' end of the fragment. In addition, the sequence of the two *Alu* 1 restriction fragments forming the 5' end of the viral strand and the complementary strand was confirmed by the Gilbert and Maxam technique (1973).

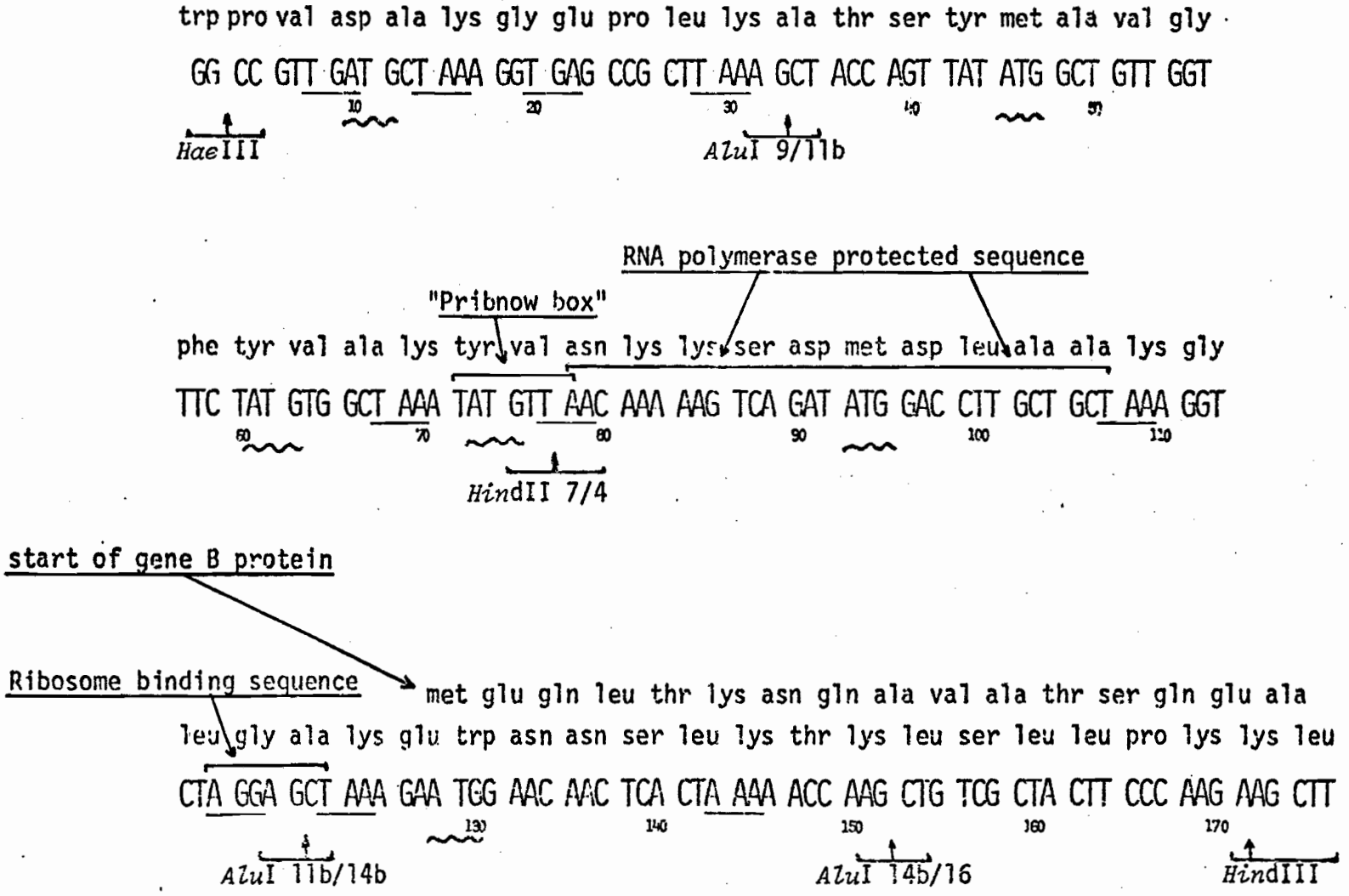
The sequence of *Hin*d 8, and that of the 75 base pair fragment (*Hae* 3 - *Hin*d 7 overlap) provide a sequence continuity with the *Hae*III 8 and 9 fragments sequenced (Grosveld and Spencer, 1977) in the gene A, B region (Figure 1). There is only one possible amino acid read out through the fragment and this sequence is in exact correlation of the gene A protein of the amino acid sequence from S13 *Hae*III 8 and 9. This provides evidence that the *Hin*d 8 sequence is still part of the A gene since there is no terminator codon in that reading frame.

Two sequence changes were found when the sequence was compared with that of ϕX174 (Sanger *et al.*, 1977). The differences are located at residues 74 and 176 of the sequence and involve changes in the viral strand from ϕX174 to S13 of C to T and G to T respectively. In the A protein frame these two changes are in the third position and thus do not affect the amino acid sequence. Also these two changes increase the predominance of T in the third position which is in agreement with the observations of Sanger *et al.* (1977) since S13 has a higher T content than ϕX174 (Spencer *et al.*, 1978). When the sequence of the *Hin*d 6 - *Hae*III 3

Figure 1 : The sequence of the viral strands of *Hin*dII-7/*Hae*III-3 and "*Hin*d 8" including the postulated amino acid sequence of gene A protein and the start and sequence of gene B protein. The lines (——) indicate nonsense codons. The lines (∼∼∼) indicate ATG codons.

The bracket accompagnied by an arrow ⌊↑⌋ indicates the restriction enzyme recognition sequence and the site of hydrolysis.

The "Pribnow box", the RNA polymerase protected sequence, and the ribosome binding sequence are indicated - (see chapter IV and V).


gene A protein
trp pro val asp ala lys gly glu pro leu lys ala thr ser tyr met ala val gly
GG CC GTT GAT GCT AAA GGT GAG CCG CTT AAA GCT ACC AGT TAT ATG GCT GTT GGT
10
20
30
40
50
HaeIII
AluI 9/11b
RNA polymerase protected sequence
"Pribnow box"
phe tyr val ala lys tyr val asn lys lys ser asp met asp leu ala ala lys gly
TTC TAT GTG GCT AAA TAT GTT AAC AAA AAG TCA GAT ATG GAC CTT GCT GCT AAA GGT
60
70
80
90
100
110
HindII 7/4
start of gene B protein
Ribosome binding sequence
met glu gln leu thr lys asn gln ala val ala thr ser gln glu ala
leu gly ala lys glu trp asn asn ser leu lys thr lys leu ser leu leu pro lys lys leu
CTA GGA GCT AAA GAA TGG AAC AAC TCA CTA AAA ACC AAG CTG TCG CTA CTT CCC AAG AAG CTT
130
140
150
160
170
AluI 11b/14b
AluI 14b/16
HindIII

overlap and that of the *Hin*d 8 fragment are compared to the corresponding fragment of ϕX174 (Sanger *et al.*, 1977), we detect a very high degree of homology. In the entire 176 base pairs, we have observed only two base changes. Also Grosveld (1977) reported 3 base changes only in the sequences of S13 *Hae*III 8 and 9 (190 bases), which corresponds to a difference of less than 2%. Kaptein (1977) reported a 5% difference between ϕX174 and S13 when looking at the pyrimidine oligonucleotide sequences (346 nucleotides). This great homology is hard to reconcile with the heteroduplex analysis of Godson (1973) which obtained 36% of base mismatch. Recently, Compton and Sinsheimer (1977) reported less homology than Godson (1973) under mild denaturing conditions and more homology under stringent conditions. The obvious conclusion that can be reached is that Godson (1973) and Compton and Sinsheimer (1977) are using an incorrect relationship between heteroduplex analysis results and the homology. It cannot be argued that the non homology is located at one unique site and that Grosveld (1977), Kaptein (1977) and ourselves have been looking at areas of high homology because first of all Godson (1973) and Compton and Sinsheimer (1977) do not report the gene A-B region as a double-stranded region in their heteroduplex analysis and also because the restriction enzyme maps of ϕX174 and S13 would not be as similar with a region of non homology covering 40% of the genome (Godson and Roberts, 1976).

One other amino acid reading frame in S13 *Hin*d 8 has stop codons along the sequence but the third reading frame has an initiator codon at position 128 with no subsequent terminator codon. This last readout could be the start of the B gene protein, and corresponds to the start of this protein in ϕX174. If the B protein starting codon in S13 is located

at the same position as in ϕX174, the smaller size of the B protein in S13 postulated by Tessman *et al.* (1976) cannot be explained by a more distal location of the translation initiation site. Two alternatives remain possible; the first one is that there is an earlier terminator codon or a deletion in the sequence, the second one is that the mobility difference of the B proteins from ϕX174 and S13 could not be due to a smaller size of the B protein of S13 but to a different level of glycosylation for example or other post translational modifications. Deletion can be ruled out because a deletion of 15 amino acids represents 45 nucleotides. This would create a length mismatch between the S13 restriction fragment and the corresponding restriction fragment in ϕX174. A difference of mobility of ϕX174 and S13 restriction fragments corresponding to 45 base pairs when compared in the range of 300 base pairs would be automatically detected. The possibility of an earlier terminator is still possible and would need additional sequence information in the area of the B terminator. The situation could also be clarified by the isolation and sequence analysis or finger printing comparison of the B proteins from ϕX174 and S13.

Finally, the fact that the starting codon for the B protein in S13 is located at the same position as in ϕX174 provides circumstantial evidence that genes A and B overlap in S13.

6. Is S13 *Hin*d 8 a promoter site.

We have already discussed the reasons why we believe that the S13 restriction fragment *Hin*d 8 contains an RNA polymerase binding site. Additional results show that if the RNA polymerase is allowed to bind to

*Hin*d 8 and the unprotected part of the fragment degraded with DNAse, a DNA fragment of about 30 base pairs remains undigested. This fragment after hydrolysis with T4 endonuclease IV, yielded oligonucleotides which were sequenced by partial snake venom phosphodiesterase (chapter IV, table 4). The majority of the oligonucleotides were part of the 5' viral end of the *Hin*d 8 fragment.

It is thought that RNA polymerase binding site equates promoter site when the canomical sequence T-A-T-R-A-T-R is present in the sequence. We find the sequence T-A-T-G-T-T-A present from position 72 to position 78 (Figure 1). This sequence is very similar to that of the *lac* promoter: T-A-T-G-T-T-G (Dickson *et al.*, 1975), the *tyr* tRNA promoter T-A-T-G-A-T-G (Sekiya and Khorana, 1974) and the *gal* promoter T-A-T-G-C-T-A (Musso *et al.*, 1977). (See table 1). It is remarkable that one base change between ϕX174 and S13 at position 74 (Figure 1) and involving a C to T change allows the formation of a promoter site although the canonical sequence is not the only feature of a promoter.

The location of the canonical sequence at this position is in good agreement with the RNA polymerase binding data .

This places the promoter site for the B gene protein approximately 50 nucleotides before the start codon of the B protein and thus about 100 nucleotides further in the 3' direction on the viral strand than the site suggested for ϕX174 (Smith and Sinsheimer, 1976b; Sanger *et al.*, 1977).

It is generally accepted that the formation of the "open" promoter complex is energy-dependent because of the necessary local unwinding or melting of base pairs during the formation of the complex (Chamberlin, 1974). It is proposed that the formation of the "open" promoter complex

tyr tRNA	C G T C A T T T G A T A T G A T G C G C C C C↓G C T T C C C	Sekiya and Khorana (1974)
T_7 A_2	T A A C A T G C A G T A A G A T A C A A A T C↓G C T A G C T	Pribnow (1975a)
T_7 A_3	G T A A A C A C G G T A C G A T G T A C C A C↓A T G A A A C	Pribnow (1975b)
fd *Hpa*-D	T G C T T C T G A C T A T A A T A G A C A G↓G G T A A A G A	Schaller *et al.* (1975)
lac	T T C C G G C T C G T A T G T T G T G T G G↓A A T T G T G A	Dickson *et al.* (1975)
lac UV5	T T C C G G C T C G T A T A A T G T G T G G↓A A T T G T G A	Gralla (in preparation)
λ P_1	C A C T G G C G G T G A T A C T G A G C A C↓A T C A G C T C	Maniatis *et al.* (1975)
λ P_r	C C T C G G C G G T G A T A A T G G T T G C↓A T G T A C T A	Waltz and Pirrotta (1975)
λ P_m	C C C T T G C G G T G A T A G A T T T A A C G T↓A T G G T A	Waltz *et al.* (1976)
trp operon	T C G A A C T A G T T A A C T A G T A C G C↓A A G T T C A C	Bennet *et al.* (1976)
λ P_o	C T G T A T T T G T C A T A A T G A C T C C T↓G T T G A T A	Scherer *et al.* (1977)
gal operon	C G A T C T T T G T T A T G C T A T G G T T↓A T T T C↓A T A	Musso *et al.* (1977)
fd *Hpa*-H	T A T T T G C T T A T A C A A T C A T C C T↓G T T T T T G G	Beck *et al.* (1977)
fd *Hpa*-C	T T G C G C T T G G T A T A A T C G C T G G↓G G T C A A A G	Takanami *et al.* (1976)
S13 *Hin*d-8	T G T G G C T A A A T A T G T T A A C A A A A↓?A A G T C A G A	

Table I. Sequences of promoter regions

The sequence underlined indicates the "Pribnow box", the arrows the start of transcription. (? = postulated mRNA starting point).

is preceeded by the formation of a "closed" promoter complex (Mangel and Chamberlin, 1974; Pribnow, 1975; chapter I, section I). Seeburg *et al.* (1977) support this view and suggest that "opening" of base pairs is essential not only for stabilizing the complex but for promoter recognition. They show that all the conditions which facilitate the opening of base pairs in DNA such as increasing the temperature or lowering the ionic strength also increase the rate of promoter selection. They suggest that there are two levels of local melting in the promoter DNA, the first one not extensive, and implicated in the recognition and the second one which covers about 10 to 15 base pairs for the stabilization of the complex. They proposed that the open region covers the 15 bases prior to the site of RNA initiation, including the "Pribnow box" and that it is A-T rich because less energy is required for opening the strands. Evidence that this region is "open" comes from the observation that dinucleotides can prime RNA chains by base pairing within that part of the promoter (Minkley and Pribnow, 1973; Dausse *et al.*, 1975).

In respect to S13 in the region of the gene B promoter, we observed that from position 68 to position 85, including the "Pribnow box" (72 to 78), 16 out of 18 nucleotides are A or T. The base content of this portion of DNA is not representative of the S13 nucleotide composition and could be considered as an additional argument in favor of a promoter site around position 75.

In addition, Shine and Delgarno (1975) noted that certain nucleotide sequences are common to the ribosome binding sites of most of the coliphages cistrons sequenced. All coliphage RNA ribosome binding sites sequenced contain all or a part of the purine rich sequence 5'rA-G-G-A-G-G-U

3' in a similar relative position on the 5' side of the initiation codon AUG (Weissmann *et al.*, 1973; Lodish, 1976). This sequence is present almost integrally in the *Hin*d 8 viral strand sequence at position 116 to 122: 5' A-G-G-A-G-C-T 3' with a single change and located 6 bases before the ATG codon of the B gene protein start at position 128 (Figure 1).

However additional experiments could be done in S13 in order to further support existing evidence that the DNA portion that we have sequenced contains the promoter site for the gene B protein. Takanami *et al.* (1976b) have shown that the G3 promoter activity of bacteriophage fd in destroyed when it is cleaved 5 nucleotides in front of the "Pribnow box" by a restriction enzyme and that the promoter activity can be recovered by replacing the DNA in front of this cleavage with any DNA fragment. Apparently DNA is required but its sequence does not play a major role. In the same field, Heyden *et al.* (1975) have shown that RNA polymerase will not form an initiation complex with short DNA fragments containing the usually recognized sequence. The experiment that can be done is to bind the RNA polymerase to restriction fragment *Hae*III 3 and after incubation with the restriction enzyme *Hin*d II, see if the site between *Hin*d II 6 and *Hin*d 8 has been "protected" by the RNA polymerase. Another experiment would be to investigate the RNA initiation on restriction fragment *Hae*III 3 before and after cleavage with *Hin*d II in order to see if the RNA initiation will be affected by the *Hin*d II cleavage. Also because of the small size of the B protein, it should be completely included in restriction fragment *Hae*III 3. It would be very interesting to isolate the mRNA made from *Hae*III 3 and see whether *in vitro* translation of the mRNA yields the B protein. This experiment would

definitively delineate the B promoter.

Although, the general features of the relation between DNA molecules and expression of genetic characteristics have been determined, there are still some points unclear in the scheme. These include the structure or the signal for DNA replication, changes of structure during recombination, repression or activation of gene transcription etc. For example, there are two A proteins, one of 62,000 daltons and one of 35,000 daltons, both in ϕX174 and in S13. Does the initiator for the smaller one occur naturally? Also, Vanderbilt *et al.* (1972) found that extracts from S13 gene B mutants sometime have the H protein missing. Is it possible that the B and H proteins are on the same mRNA? This would imply a binding site before the H gene which we found in *Hin*d 3 - *Hae*III 1 in S13 and at a higher intensity in ϕX174 at the *Hin*d II 2 - *Hae*III 1 overlap and between 50 and 55% of the genome in the electron microscope (chapter II and III). If genes B and H are on the same message this would also imply that there is RNA splicing in bacteriophage as already found in eukaryotic systems. Finally, there is evidence that gene E is completely contained within the D coding region in ϕX174 but in a different reading frame. There are great chances that the same situation occurs in S13 and this is another area where sequence investigations should be made. Some of these points, in particular those which involve interaction of specific proteins with DNA molecules, can only be elucidated by direct DNA sequence analysis. Many of these studies are presently underway but have not yet yielded enough information to enable us to understand why a particular sequence is responsible for the role it has in a specific biological function.

REFERENCES

Allet, B. and Solem, R. (1974). Separation and analysis of promoter sites in bacteriophage lambda DNA by specific endonucleases. J. Mol. Biol. 85, 475-484.

Anderson, W., Gottesman, M.E. and Pastan, I. (1974). Studies with cyclic AMP receptor and stimulation in *in vitro* transcription of the *gal* operon. J. Biol. Chem. 249, 3592-3596.

Avery, D.T., MacLeod, C.M. and McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III. J. Exp. Med. 79, 137-158.

Axelrod, N. (1976a). Transcription of bacteriophage ϕX174 *in vitro*: Selective initiation with oligonucleotides. J. Mol. Biol. 108, 753-770.

Axelrod, N. (1976b). Transcription of bacteriophage ϕX174 *in vitro*: Analysis with restriction enzyme. J. Mol. Biol. 108, 771-779.

Bahr, W., Stender, W., Scheit, K.-H. and Jovin, T.M. (1976). Binding of rifampicin to *E. coli* RNA polymerase: Thermodynamic and kinetic studies. In RNA polymerase. Losick, R. and Chamberlin, M.J. (eds.). pp. 363-396. Cold Spring Harbor Laboratory.

Baker, A. and Tessman, E. (1967). The circular genetic map of phage S13. Proc. Natl. Acad. Sci. USA 58, 1438-1445.

Bautz, E. and Bautz, F. (1970). Studies on the function of the RNA polymerase sigma factor in site selection. Cold Spring Harbor Symp. Quant. Biol. 35, 227-232.

Bautz, E., Bautz, F. and Dunn, J. (1969). *E. coli* σ factor: A positive control element in phage T4 development. Nature 223, 1022-1024.

Benbow, R.M., Hutchison, C.A., Fabricant, J.D. and Sinsheimer, R.L. (1971). Genetic map of bacteriophage ϕX174. J. Virol. 7, 549-558.

Bennett, G.N., Schweingruber, M.E., Brown, K.D., Squires, C. and Yanofsky, C. (1976). Nucleotide sequence of region preceding *trp* mRNA initiation site and its role in promoter and operator function. Proc. Natl. Acad. Sci. USA 73, 2351-2355.

Berg, D., Barrett, K. and Chamberlin, M.J. (1971). Purification of two forms of *E. coli* RNA polymerase. In Methods in enzymology. Grossman and Moldave, K. (eds.). Vol. 21 D, pp. 506-519. Academic Press, New York.

Berg, D. and Chamberlin, M.J. (1970). Physical studies on ribonucleic acid polymerase from *Escherichia coli* B. Biochemistry 9, 5055-5064.

Berg, P., Fancher, H. and Chamberlin, M.J. (1963). Synthesis of mixed polynucleotides containing ribo- and deoxyribonucleotides by purified preparations of DNA polymerase from *E. coli*. In informational macro molecules. Vogel, H.J., Bryson, V. and Lampen, J.O. (eds.). pp. 467-483. Academic Press, New York.

Bernardi, A., Matt, J., De Waard, A. and Bernardi, G. (1976). Preparation and specificity of endonuclease IV induced by bacteriophage T4. Eur. J. Biocehm. 66, 175-179.

Bertrand, K., Korn, L., Lee, F., Platt, T., Squires, C.L., Squires, C. and Yanofsky, C. (1975). New features of the regulation of the tryptophane operon. Science 189, 22-26.

Blattner, F.R. and Dahlberg, J.E. (1972). RNA synthesis startpoints in bacteriophage λ: Are the promoter and operator transcribed. Nature New Biol. 237, 227-232.

Bordier, C. and Dubochet, J. (1974). Electron microscopic localization of the binding sites of *Escherichia coli* RNA polymerase in the early promoter region of T7 DNA. Eur. J. Biochem. 44, 617-624.

Brachet, J. (1957). Biochemical Cytology, chapter VII. Academic Press, New York.

Brownlee, G.G. and Sanger, F. (1969). Chromotography of ^{32}P-labeled oligonucleotides on thin layers of DEAE-cellulose. Eur. J. Biochem. 11, 395-399.

Bujard, H. and Heidelberger, C. (1966). Fluorinated pyrimidine. XXVII. Attempts to determine transcription errors during the formation of fluorouracil containing messenger ribonucleic acid. Biochemistry 5, 3339-3345.

Burgess, R.R. (1969a). A new method for the large scale purification of *E. coli* DNA-dependent RNA polymerase. J. Biol. Chem. 244, 6160-6167.

Burgess, R.R. (1969b). Separation and characterization of the subunits of ribonucleic acid polymerase. J. Biol. Chem. 244, 6168-6176.

Burgess, R.R. and Travers, A.A. (1971). DNA-dependent RNA polymerase. In procedures in nucleic acid research. Cantoni, G.L. and Davis, D.R. (eds.). Vol. 2, pp. 851-863. Harper and Row, New York.

Burgess, R.R., Travers, A.A., Dunn, J.J. and Bautz, E. (1969). Factor stimulating transcription by RNA polymerase. Nature 221, 43-46.

Burton, K. (1967). Preparation of apurinic acid and of oligodeoxyribonucleotides with formic acid and diphenylamine. In methods in enzymology. Grossman, L. and Moldave, K. (eds.). Vol. 12A, pp. 222-224. Academic Press, New York.

Burton, K. and Peterson, G.B. (1960). The frequencies of certain sequences of nucleotides in deoxyribonucleic acid. Biochem. J. 75, 17-27.

Cape, R.E. and Spencer, J.H. (1968). Nucleotide clusters in deoxyribonucleic acids. I. Isolation of purine oligonucleotides. Can. J. of Biochem. 46, 1063-1073.

Cashmore, A.R. and Peterson, G.B. (1969). The degradation of DNA by hydrazine: A critical study of the suitability of the reaction for the quantitative determination of purine nucleotide sequences. Biochim. Biophys. Acta 174, 591-603.

Cerny, R., Mushynski, W.E. and Spencer, J.H. (1968). Nucleotide clusters in deoxyribonucleic acids. III. Separation of pyrimidine isostichs according to base composition. Biochim. Biophys. Acta 169, 439-450.

Chamberlin, M.J. (1974). The selectivity of transcription. Snell, E.E. (ed.). Vol. 43, pp. 721-775. Ann. Rev. of Biochim.

Chamberlin, M.J. (1976). Interaction of RNA polymerase with the DNA template. In RNA polymerase. Losick, R. and Chamberlin, M.J. (eds.). pp. 159-191. Cold Spring Harbor Laboratory USA.

Chambon, P. (1975). Eukaryotic nuclear RNA polymerases. Snell, E.E. (ed.). Ann. Rev. of Biochem. 44, 613-638.

Chandler, B., Hayashi, M., Hayashi, M.N. and Spiegelman, S. (1964). Circularity of the replicating form of a single-stranded DNA virus. Science 143, 47-49.

Chen, C.-Y., Hutchison, C.A. and Edgell, M.H. (1973). Isolation and genetic location of three ϕX174 promoter regions. Nature New Biol. 243, 233-236.

Compton, J.L. and Sinsheimer, R.L. (1977). Aligning the ϕX174 genetic map and the ϕX174/S13 heteroduplex denaturation map. J. Mol. Biol. 109, 217-234.

Dausse, J., Sentenac, A. and Fromageot, P. (1976). Interaction of RNA polymerase from *E. coli* with DNA. Effect of temperature and ionic strength on selection of T4 DNA early promoters. Eur. J. Biochem. 57, 569-578.

Delaney, A.D. (1974). Studies toward the primary structure of bacteriophage S13 DNA. Ph.D. thesis, McGill University.

Delaney, A.D. and Spencer, J.H. (1973). Sequence determination of oligonucleotides from S13 su N15 DNA. Fed. Proc. 32, 2536.

Delaney, A.D. and Spencer, J.H. (1976). Nucleotide clusters in deoxyribonucleic acids. XIII. Sequence analysis of the longer unique pyrimidine oligonucleotides of bacteriophage S13 DNA by a method using unlabeled starting oligonucleotides. Biochim. Biophys. Acta 435, 269-281.

Denhardt, D.T. (1975). The single-stranded DNA phages. CRC Crit. Rev. Microbiol. 4, 161-222.

Dickson, R.C., Abelson, J., Barnes, W.M. and Reznikoff, W.S. (1975). Genetic regulation: The *lac* control region. The nucleotide sequence of the *lac* control region containing the promoter and operator is presented. Science 187, 27-35.

Dubochet, J., Ducommun, M., Zollinger, M. and Kellenberger, E. (1971). A new preparation method for dark-field electron microscopy of biomacromolecules. J. Ultrastruct. Res. 35, 147-167.

Englund, P.T. (1972). The 3'-terminal nucleotide sequences of T7 DNA. J. Mol. Biol. 66, 209-224.

Fellner, P., Ehresman, C. and Ebel, J.P. (1970). Nucleotide sequence analysis of sections of the 16S RNA from *E. coli*. Cold Spring Harbor Symp. Quant. Biol. 35, 29-33.

Galibert, F., Sedat, J. and Ziff, E. (1974). Direct determination of DNA nucleotide sequence. Structure of a fragment of bacteriophage ϕX174 DNA. J. Mol. Biol. 87, 377-407.

Garen, A. and Levinthal, C. (1960). A Fine-Structure genetic and chemical study of the enzyme alkaline phosphatase of *E. coli*. I. Purification and characterization of alkaline phosphatase. Biochim. Biophys. Acta 38, 470-483.

Giacomoni, P.U., Delain, E. and Le Pecq, J.B. (1977). Electron microscopy analysis of the interaction between *Escherichia coli* DNA-dependent RNA polymerase and the replication from of phage fd DNA. Eur. J. Biochem. 78, 205-213.

Gierer, A. and Schramm, G. (1956). Infectivity of ribonucleic acid from tobacco mosaic virus. Nature 172, 702-703.

Gilbert, W. (1976). Starting and stopping sequences for the RNA polymerase. In RNA polymerase. Losick, R. and Chamberlin, M.J. (eds.). pp. 193-205. Cold Spring Harbor Laboratory.

Gilbert, W., Maizels, N. and Maxam, A. (1974). Sequence of controlling regions of the lactose operon. Cold Spring Harbor Symp. Quant. Biol. 38, 845-855.

Gilbert, W. and Maxam, A. (1973). The nucleotide sequence of the *lac* operator. Proc. Natl. Acad. Sci. USA 70, 3581-3584.

Gillepsie, D. and Spiegelman, S. (1965). A quantitative assay for DNA-RNA hybrids with DNA immobilized on a membrane. J. Mol. Biol. 12, 829-842.

Godson, G.N. (1973). DNA heteroduplex analysis of the relation between bacteriophages ϕX174 and S13. J. Mol. Biol. 77, 467-477.

Godson, G.N. and Roberts, R.J. (1976). A catalogue of cleavages of ϕX174, S13, G4 and ST-1 DNA by 26 different restriction endonucleases. Virology 73, 561-567.

Goldberg, A.R. and Hurwitz, J. (1973). Studies on termination of *in vitro* ribonucleic acid synthesis by *rho* factor. J. Biol. Chem. 247, 5637-5645.

Gomez, B. and Lang, D. (1972). Denaturation map of bacteriophage T7 DNA and direction of DNA transcription. J. Mol. Biol. 70, 239-251.

Goodchild, B. and Spencer, J.H. (1978). Mapping of the single cleavage site in S13 replicative form DNA of *Hemophilus influenzae* (*Hin*dIII) restriction enzyme. Virology 84, 536-539.

Gralla, J. and Crothers, D.M. (1973). Free energy of imperfect nucleic acid helices. II. Small hairpin loops. J. Mol. Biol. 73, 497-511.

Gralla, J. (1975 unpublished). Referenced in Sekiya, T., Takeya, T. and Khorana, H.G. 1975.

Greenblatt, J. and Schleif, R. (1971). Arabinose C protein: Regulation of the arabinose operon *in vitro*. Nature New Biol. 233, 166-170.

Grosveld, F.G. (1977). Studies on the structure of bacteriophage S13 replicative form DNA: Restriction enzyme and sequence analysis. Ph.D. thesis. McGill University.

Grosveld, F.G., Ojamaa, K.M. and Spencer, J.H. (1976). Fragmentation of bacteriophage S13 replicative form DNA by restriction endonucleases from *Hemophilus influenzae* and *Hemophilus aegyptius*. Virology 71, 312-324.

Grosveld, F.G. and Spencer, J.H. (1977). The nucleotide sequence of two restriction fragments located in the gene AB region of bacteriophage S13. Nucleic Acids Res. 4, 2235-2252.

Habermann, V. (1963). Participation of purine and pyrimidine nucleotide sequences in the molecule of calf thymus deoxyribonucleic acid. Nature 200, 782-783.

Hall, B.D. and Spiegelman, S. (1961). Sequence complementarity of T2-DNA and T2-specific RNA. Proc. Natl. Acad. Sci. 47, 137-146.

Hall, Z.W. and Lehman, I.R. (1968). An *in vitro* transversion by a mutationally altered T4-induced DNA polymerase. J. Mol. Biol. 36, 321-333.

Hall, J.B. and Sinsheimer, R.L. (1963). The structure of the DNA of bacteriophage ϕX174. IV. Pyrimidine sequences. J. Mol. Biol. 6, 115-127.

Harbers, B., Delaney, A.D., Harbers, K. and Spencer, J.H. (1976). Nucleotide clusters in DNAs. Comparison of the sequences of the large pyrimidine oligonucleotides of bacteriophage S13 and ϕX174 deoxyribonucleic acids. Biochemistry, 15, 407-414.

Hayashi, M. (1965). A DNA-RNA complex as an intermediate of *in vitro* genetic transcription. Proc. Natl. Acad. Sci. USA 54, 1736-1743.

Hayashi, M., Fujimura, F.K. and Hayashi, M. (1976). Mapping of *in vivo* messenger RNA's for bacteriophage ϕX174. Proc. Natl. Acad. Sci. USA 73, 3519-3423.

Hayashi, M., Hayashi, M.N. and Hayashi, Y. (1970). Size distribution of ϕX174 RNA synthesized *in vitro*. Cold Spring Harbor Symp. Quant. Biol. 35, 174-177.

Hayashi, M.N. and Hayashi, M. (1972). Isolation of ϕX174 specific messenger ribonucleic acids *in vivo* and identification of their 5' terminal nucleotides. J. Virol. 9, 207-215.

Hayashi, Y. and Hayashi, M. (1970). Fractionation of ϕX174 specific messenger RNA. Cold Spring Harbor Symp. Quant. Biol. 35, 171-174.

Heineman, S.F. and Spiegelman, W.G. (1971). Role of the gene N product in phage lambda. Cold Spring Harbor Symp. Quant. Biol. 35, 315-318.

Heyden, B., Nusslein, C. and Schaller, H. (1972). Single DNA polymerase binding site isolated. Nature New Biol. 96, 9-12.

Hinkle, D.C. and Chamberlin, M.J. (1970). The role of sigma subunit in template site selection by *E. coli* RNA polymerase. Cold Spring Harbor Symp. Quant. Biol. 35, 65-72.

Hinkle, D.C. and Chamberlin, M.J. (1972a). Studies of the binding of *E. coli* RNA polymerase to DNA. I. The role of sigma subunit in site selection. J. Mol. Biol. 70, 157-185.

Hinkle, D.C. and Chamberlin, M.J. (1972b). Studies of the binding of *Escherichia coli* RNA polymerase to DNA. II. The kinetics of the binding reaction. J. Mol. Biol. 70, 187-195.

Hinkle, D.C., Ring, J. and Chamberlin, M.J. (1972). Studies of the binding of *Escherichia coli* RNA polymerase to DNA. III. Tight binding of RNA polymerase holoenzyme to single-strands breaks in T7 DNA. J. Mol. Biol. 70, 197-207.

Hoffman, D.J. and Niyogi, S.K. (1973). RNA initiation with dinucleoside monophosphates during transcription of bacteriophage T4 DNA with RNA polymerase of *Escherichia coli*. Proc. Natl. Acad. Sci. 70, 574-578.

Humphries, P., McConnel, D.J. and Connolly, P. (1974). Interaction of *Escherichia coli* DNA-dependent RNA polymerase and endonuclease R *Hin*d fragments of T7 DNA. Biochim. Biophys. Acta 336, 70-78.

Iida, Y. and Matsukage, A. (1974). The effect of KCl concentration on the transcription by *E. coli* RNA polymerase. III. Starting nucleotide sequences of RNA synthesized on T7 DNA. Mol. Gen. Genetics 129, 27-35.

Ishihama, A., Fukuda, R. and Ito, K. (1973). Subunits of RNA polymerase in function and structure. IV. Enhancing role of sigma in the subunit assembly of *E. coli* RNA polymerase. J. Mol. Biol. 79, 127-136.

Ishihama, A. and Ito, K. (1972). Subunits of RNA polymerase in structure and function. II. Reconstitution of *Escherichia coli* RNA polymerase from isolated subunits. J. Mol. Biol. 72, 111-123.

Ito, K., Iwakura, Y. and Ishihama, A. (1975). Biosynthesis of RNA polymerase in *Escherichia coli*. III. Identification of intermediates in the assembly of RNA polymerase. J. Mol. Biol. 96, 257-271.

Jackson, E.N. and Yanofsky, C. (1973). The regions between the operator and first structural gene of the tryptophan operon of *E. coli* may have a regulatory function. J. Mol. Biol. 76, 89-102.

Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3, 318-356.

Jones, B.B., Chan, H., Rothstein, S., Wells, R.D. and Reznikoff, W.S. (1977). RNA polymerase binding sites in λp *lac* 5 DNA. Proc. Natl. Acad. Sci. USA 74, 4914-4918.

Jones, O.W. and Berg, P. (1966). Studies on the binding of RNA polymerase to polynucleotides. J. Mol. Biol. 22, 199-209.

Junowicz, E. and Spencer, J.H. (1973). Studies on bovine pancreatic deoxyribonuclease A. I. General properties and activation with different bivalent metals. Biochim. Biophys. Acta 312, 72-84.

Kaptein, J.S. (1977). Sequence analysis of DNA by template - Directed extension of pyrimidine oligodeoxynucleotide primers. Ph.D. Thesis. McGill University.

Kay, E.R.M., Simmons, N.S. and Dounce, A.L. (1952). An improved preparation of sodium desoxyribonucleate. J. Amer. Chem. Soc. 74, 1724-1726.

Kelly, T.J. and Smith, H.O. (1970). A restriction enzyme from *Hemophilus influenzae*. II. Base sequence of the recognition site. J. Mol. Biol. 51, 393-410.

Kleinschmidt, A.K., Burton, A. and Sinsheimer, R.L. (1963). Electron microscopy of the replicative form of the DNA of the bacteriophage ϕX174. Science 142, 961.

Koller, T.L., Sogo, J.M. and Bujard, H. (1974). An electron microscopic method for studying nucleic acid - protein complexes. Visualization of RNA polymerase bound to the DNA of bacteriophage T7 and T3. Biopolymers 13, 995-1009.

Kourilsky, P., Bourguignon, M.-F., Bouquet, M. and Gross, F. (1971). Early transcription controls after induction of prophage λ. Cold Spring Harbor Symp. Quant. Biol. 35, 305-314.

Lawley, P.D. (1966). Effects of some chemical mutagens and carcinogens on nucleic acids. Progress in nucleic acid research and molecular biology. Davidson, J.N. and Cohn, W.E. (eds.). Vol. 5, pp. 89-131. Academic Press, New York.

Lawley, P.D. and Brookes, P. (1963). Further studies on the alkylation of nucleic acids and their constituent nucleotides. J. of Biochem. 89, 127-138.

Lebowitz, P., Weissman, S.H. and Rodding, C.M. (1971). Nucleotide sequence of a ribonucleic acid transcribed *in vitro* from λ phage deoxyribonucleic acid. J. Biol. Chem. 246, 5120-5139.

Lee, A.S. and Sinsheimer, R.L. (1974). A cleavage map of bacteriophage ϕX174 genome. Proc. Natl. Acad. Sci. USA 71, 2882-2886.

Lee, F., Squires, C.L. and Yanofsky, C. (1976). Termination of transcription *in vivo* in the *E. coli* tryptophan operon leader region. J. Mol. Biol. 103, 383-394.

Le Talaer, J.-Y. and Jeanteur, P.L. (1971). Purification and base composition analysis of phage λ early promoters. Proc. Natl. Acad. Sci. 68, 3211-3215.

Le Talaer, J.-Y., Kermici, M. and Jeanteur, P.H. (1973). Isolation of *E. coli* RNA polymerase binding sites on T5 and T7 DNA: Further evidence for sigma-dependent recognition of A-T rich DNA sequences. Proc. Natl. Acad. Sci. 70, 2911-2915.

Lewin, B. (1974). Interaction of regulator proteins with recognition sequences of DNA. Cell 2, 1-7.

Lewin, B. (1974). Operon control circuits. Gene Expression. Vol. 1, pp. 310-343. Bacteriol. Genomes, John Wiley and Sons, London.

Lill, H., Lill, U., Sippel, A. and Hartmann, G. (1969). Inhibition of the RNA polymerase reaction by rifampicin. In Lepetit colloquium on RNA polymerase and transcription. Silvestri, L. (ed.). pp. 55-64, North-Holland, Amsterdam.

Lill, U. and Hartmann, R. (1970). Reactivation of denatured RNA polymerase from *E. coli*. Biochem. Biophys. Res. Commun. 39, 930-935.

Ling, V. (1971). Partial digestion of ^{32}P-fd DNA with T4 endonuclease IV. FEBS Lett. 19, 50-54.

Ling, V. (1972). Pyrimidine sequences from the DNA of bacteriophages fd f1 and ϕX174. Proc. Natl. Acad. Sci. USA 69, 742-746.

Lodish, H.F. (1976). Translational control of protein synthesis. In Annual Review of Biochemistry. Snell, E.E. (ed.). Vol. 45, pp. 39-42. Annual Reviews Inc.

Ludwig, R.A. and Summers, W.C. (1975). A restriction fragment analysis of the T7 left-early region. Virology 68, 360-373.

MacHattie, L.A. and Thomas, C.A. Jr. (1975). Viral DNA molecules. In handbook in Biochemistry and Molecular Biology, VI nucleic acids. Fasman, G.D. (ed.). p. 549, CRC Press, Cleveland, Ohio.

Maitra, U. and Hurwitz, J. (1965). The role of DNA in RNA synthesis. IX. Nucleoside triphosphate termini in RNA polymerase products. Proc. Natl. Acad. Sci. 54, 815-822.

Maitra, U., Lockwood, A.H., Dubnoff, J.S. and Guha, A. (1970). Termination, release, and reinitiation of RNA chains from DNA templates by *E. coli* RNA polymerase. Cold Spring Harbor Symp. Quant. Biol. 35, 143-156.

Maizels, N.M. (1973). The nucleotide sequence of the lactose mRNA transcribed from the UV5 promoter mutant of *E. coli*. Proc. Natl. Acad. Sci. 70, 3583-3589.

Mangel, W. and Chamberlin, M.J. (1974). Studies of RNA chain initiation by *E. coli* RNA polymerase bound to T7 DNA. II. The effect of alteration in ionic strength of chain initiation and on the conformation of binary complexes. J. Biol. Chem. 249, 3002-3006.

Mangel, W.F. and Chamberlin, M.J. (1974). Studies of ribonucleic acid chain initiation by *E. coli* RNA polymerase bound to T7 DNA. III. The effect of temperature on RNA chain initiation and on the conformation of binary complexes. J. Biol. Chem. 249, 3007-3013.

Maniatis, T., Ptashne, M., Backman, K., Kleid, D., Flashman, S., Jeffery, A. and Maurer, R. (1975). Recognition sequences of repressor and polymerase in the operators of bacteriophage lambda. Cell 5, 109-113.

Maniatis, T., Ptashne, M., Barrel, B.G. and Donelson, J. (1974). Sequence of a repressor-binding site in the DNA of bacteriophage λ. Nature 250, 394-397.

Matsukage, A. (1972). The effects of KCl concentration on the transcription by *E. coli* RNA polymerase. II. Effect on the specificity of T7 transcription. Mol. Gen. Genetics 118, 23-31.

Matsukage, A., Murakami, S. and Kameyama, T. (1969). The isolation and the characterization of DNA region bound to *Escherichia coli* RNA polymerase. Biochim. Biophys. Acta 179, 145-157.

Maxam, A.M. and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74, 560-564.

Miescher, F. (1871). On the chemical composition of pus cells. Hoppe - Seyler's Med. - Chem. Untersuch. 4, 441-460.

Minkley, E.G. and Pribnow, D. (1973). Transcription of the early region of bacteriophage T7: Selective initiation with dinucleotides. J. Mol. Biol. 77, 255-277.

Mueller, K. (1971). The function of the σ-factor of *Escherichia coli* RNA polymerase in template site selection. Mol. Gen. Genetics 111, 273-296.

Murray, K. and Old, R.W. (1974). The primary structure of DNA. Progress in Nucleic Acid Research and Molecular Biology. Cohn, W.E. (ed.). Vol. 14, pp. 117-185. Academic Press, New York.

Mushynski, W.E. and Spencer, J.H. (1970a). Nucleotide clusters in deoxyribonucleic acids. V. The pyrimidine oligonucleotides of strands *r* and *l* of bacteriophage T7 DNA. J. Mol. Biol. 52, 91-106.

Mushynski, W.E. and Spencer, J.H. (1970b). Nucleotide clusters in deoxyribonucleic acids. VI. The pyrimidine oligonucleotides of strands *r* and *l* of bacteriophage lambda DNA. J. Mol. Biol. 52, 107-120.

Musso, R., DiLauro, R., Rosenberg, M. and De Crombrugghe, B. (1977). Nucleotide sequence of the operator-promoter region of the galactose operon of *E. coli*. Proc. Natl. Acad. Sci. USA 74, 106-110.

Naito, S. and Ishihama, A. (1975). Isolation and properties of the transcription complex of *Escherichia coli* RNA polymerase. Biochim. Biophys. Acta 402, 88-104.

Nakano, E. and Sakaguchi, K. (1968). Isolation and base composition of sites on λ phage DNA which bind with RNA polymerase of *Escherichia coli*. J. Biochem. 65, 147-150.

Niyogi, S.K. and Underwood, B.H. (1975). The isolation and properties of the specific binding sites for *Escherichia coli* RNA polymerase on T4 and T7 bacteriophage DNAs. J. Mol. Biol. 94, 572-535.

Nüsslein, C. and Schaller, H. (1974). Stabilization of promoter complexes with a single ribonucleoside triphosphate. Eur. J. Biochem. 56, 563-569.

Oertel, W. and Schaller, H. (1972). A new approach to the sequence analysis of DNA. FEBS Letts. 27, 316-320.

Okamoto, T., Sugimoto, K., Sugisaki, H. and Takanami, M. (1975). Studies on bacteriophage fd DNA. II. Localization of RNA initiation sites on the cleavage map of the fd genome. J. Mol. Biol. 95, 33-44.

Okamoto, T., Sugiura, M. and Takanami, M. (1972). RNA polymerase binding sties of phage fd replicative form DNA. Nature New Biol. 237, 108-109.

Panny, S., Heil, A., Mazus, B., Palm, P., Zillig, W., Mindlin, S., Ilyina, T. and Khesin, R. (1974). A temperature sensitive mutation of the β' subunit of DNA dependent RNA polymerase from *E. coli*. FEBS Letts. 48, 241-245.

Peterson, G.B. and Reeves, J.M. (1966). An improved separation of pyrimidine oligonucleotides derived from DNA. Biochim. Biophys. Acta 129, 438-440.

Platt, T., Squires, C. and Yanofsky, C. (1976). Ribosome-protected regions in the leader-*trp*E sequence of *E. coli* tryptophan operon messenger RNA. J. Mol. Biol. 103, 411-420.

Pochon, F. and Michelson, A.M. (1967). Polynucleotides. IX. Methylation of nucleic acids, homopolynucleotides and complexes. Biochim. Biophys. Acta 149, 99-106.

Pollock, T.J. and Tessman, I. (1976). Units of functional stability in bacteriophage messenger RNA. J. Mol. Biol. 108, 651-664.

Portmann, R., Sogo, J.M., Koller, T.L. and Zillig, W. (1974). Binding sites of *E. coli* RNA polymerase on T7 DNA as determined by electron microscopy. FEBS Letts. 45, 64-67.

Pribnow, D. (1975). Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. Proc. Natl. Acad. Sci. USA 72, 784-788.

Pribnow, D. (1975). Bacteriophage T7 early promoters: Nucleotide sequences of two RNA polymerase binding sites. J. Mol. Biol. 99, 419-443.

Puga, A. and Tessman, I. (1973). Mechanism of transcription of bacteriophage S13. I. Dependence of messenger RNA synthesis on amount and configuration of DNA. J. Mol. Biol. 75, 83-97.

Rabussay, D. and Zillig, W. (1970). A rifampicin resistant RNA polymerase from *E. coli* altered in the β subunit. FEBS Letts. 5, 104-106.

Rassart, E. and Spencer, J.H. (1978). Localization of *Escherichia coli* RNA polymerase binding sites on S13 and ϕX174 DNA: Alignment with restriction enzyme maps. J. Virol. in press.

Richardson, C.C. (1965). Phosphorylation of nucleic acid by an enzyme from T4 bacteriophage-infected *Escherichia coli*. Proc. Natl. Acad. Sci. USA 54, 158-165.

Richardson, J.P. (1966). The binding of RNA polymerase to DNA. J. Mol. Biol. 121, 83-114.

Richardson, J.P., Grimley, C. and Lowery, C. (1975). Transcription termination factor *rho* activity is altered in *Escherichia coli* with *su*A gene mutations. Proc. Natl. Acad. Sci. 72, 1725-1728.

Roberts, J.W. (1969). Termination factor for RNA polymerase. Nature 224, 1168-1174.

Roberts, J.W. (1976). Transcription termination and its control in *E. coli*. In RNA polymerase. Losick, R. and Chamberlin, M. (eds.). pp. 242-272. Cold Spring Harbor Laboratory USA.

Roberts, R.J. (1976). Restriction and modification enzymes and their recognition sequences. In handbook of Biochemistry and Molecular Biology. Fasman, G.D. (ed.). Vol. II, pp. 532-534. Nucleic Acids.

Roberts, R.J., Myers, P.A., Morrison, A. and Murray, K. (1976). A specific endonuclease from *Arthrobacter luteus*. J. Mol. Biol. 102, 157-165.

Robertson, H.D., Barrell, B.G., Weith, H.L. and Donelson, J.E. (1973). Isolation and sequence analysis of a ribosome protected fragment from bacteriophage ϕX174 DNA. Nature New Biol. 241, 38-40.

Roeder, R.G. (1976). Eukaryotic nuclear RNA polymerase. In RNA polymerase. Losick, R. and Chamberlin, M.J. (eds.). pp. 285-329. Cold Spring Harbor Laboratory.

Rüger, W. (1971). Transcription of RNA polymerase binding sites isolated from T4 phage DNA. Biochim. Biophys. Acta 238, 202-211.

Sadowski, P.D. and Bakyta, I. (1972). T4 endonuclease IV. Improved purification procedures and resolution from T4 endonuclease III. J. Biol. Chem. 247, 405-412.

Sadowski, P.D. and Hurwitz, J. (1969). Enzymatic breakage of deoxyribonucleic acid. I. Purification and properties of endonuclease II from T4 phage-infected *Escherichia coli*. J. Biol. Chem. 244, 6192-6198.

Sagar Sethi, V. and Zillig, W. (1970). Dissociation and reconstitution of active DNA-dependent RNA polymerase from *E. coli*. FEBS Letts. 8, 236-239.

Saitoh, T. and Ishihama, A. (1976). Subunits of RNA polymerase in function and structure. VI. Sequence of the assembly *in vitro* of *E. coli* RNA polymerase. J. Mol. Biol. 104, 621-635.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A. III, Slocombe, P.M. and Smith, M.M. (1977). Nucleotide sequence of bacteriophage ϕX174 DNA. Nature 265, 687-695.

Sanger, F. and Coulson, A.R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J. Mol. Biol. 94, 441-448.

Sanger, F., Donelson, J.E., Coulson, A.R., Kössel, H. and Fisher, D. (1973). Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage f1 DNA. Proc. Natl. Acad. Sci. 70, 1209-1213.

Sarocchi, M.-T. and Darlix, J.-L. (1974). A spectroscopic approach to DNA transcription and protein binding. Eur. J. Biochem. 46, 481-489.

Saucier, J. and Wang, J. (1972). Angular alteration of the DNA helix by *E. coli* RNA polymerase. Nature New Biol. 239, 167-170.

Schäfer, R., Krämer, R., Zillig, W. and Cudny, H. (1973a). On the initiation of transcription by DNA-dependent RNA polymerase from *Escherichia coli*. Eur. J. Biochem. 40, 367-373.

Schäfer, R. and Zillig, W. (1973). The effects of ionic strength on termination of transcription of DNAs from bacteriophages T4, T5 and T7 by DNA-dependent RNA polymerase from *Escherichia coli* and the nature of termination of factor ρ. Eur. J. Biochem. 33, 215-226.

Schäfer, R., Zillig, W., Zechel, K., Janekovic, D. and Kruck, J. (1973b). A model for the initiation of transcription by DNA-dependent RNA polymerase from *E. coli*. Eur. J. Biochem. 33, 207-214.

Schaller, H., Gray, C. and Herrmann, K. (1975). Nucleotide sequence of an RNA polymerase binding site from the DNA bacteriophage fd. Proc. Natl. Acad. Sci. USA 72, 737-741.

Schekman, R.W., Iwaya, M., Bromstrup, K. and Denhardt, D.T. (1971). The mechanism of replication of ϕX174 single-stranded DNA. III. An enzymatic study of the structure of the replicative form II DNA. J. Mol. Biol. 57, 177-199.

Scherer, G., Hobom, G. and Kössel, H. (1977). DNA base sequence of the *Po* promoter region of phage λ. Nature 265, 117-121.

Sedat, J. and Sinsheimer, R.L. (1964). Structure of the DNA of bacteriophage ϕX174 V purine sequences. J. Mol. Biol. 9, 489-487.

Seeburg, P.H. and Schaller, H. (1975). Mapping and characterization of promoters in bacteriophage fd, f1 and M13. J. Mol. Biol. 92, 261-277.

Seeburg, P.H., Nüsslein, C. and Schaller, H. (1977). Interaction of RNA polymerase with promoters from bacteriophage fd. Eur. J. Biochem. 74, 107-113.

Seeman, N.C., Rosenberg, J.M. and Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by protein. Proc. Natl. Acad. Sci. 73, 804-808.

Sekiya, T. and Khorana, H.G. (1974). Nucleotide sequence in the promoter region of the *E. coli* tyrosine tRNA gene. Proc. Natl. Acad. Sci. USA 78, 2978-2982.

Sekiya, T., Ormondt, A. and Khorana, G.H. (1975). The nucleotide sequence in promoter region of gene for an *Escherichia coli* tyrosine transfer ribonucleic acid. J. Biol. Chem. 250, 1087-1098.

Shine, J. and Delgarno, L. (1975). Determinant of cistron specificity in bacterial ribosomes. Nature 254, 34-38.

Shleser, R., Ishiwa, H., Mannes, B. and Tessman, E.S. (1968). Early replicative form DNA of bacteriophage S13. I. Sucrose gradient analysis of replicative form made by gene IV mutants. J. Mol. Biol. 34, 121-129.

Shleser, R., Puga, A. and Tessman, E.S. (1969). Synthesis of replicative form deoxyribonucleic acid and messenger ribonucleic acid by gene IV mutants of bacteriophage S13. J. Virol. 4, 394-399.

Singer, B. (1975). The chemical effects of nucleic acids alkylation and their relation to mutagenesis and carcinogenesis. In Prog. in Nucleic Acid Research and Molecular Biology. Cohn, W.E. (ed.). Vol. 15, pp. 219-284. Academic Press, New York.

Sinsheimer, R.L. (1959). A single-stranded deoxyribonucleic acid from bacteriophage ϕX174. J. Mol. Biol. 1, 43-53.

Slayter, M.S. and Hall, C.E. (1966). Electron microscopy of RNA polymerase and RNA polymerase bound to T7 DNA. J. Mol. Biol. 21, 113-114.

Smith, D.A. and Martinez, A.M. (1974). Initiation of *in vitro* RNA synthesis with deoxyribosyl oligomers. Biochim. Biophys. Acta 353, 475-486.

Smith, H.O. and Wilcox, K.W. (1970). A restriction enzyme from *Hemophilus influenzae* I purification and general properties. J. Mol. Biol. 51, 379-392.

Smith, L.H. and Sinsheimer, R.L. (1976a). The *in vitro* transcription units of bacteriophage ϕX174. I. Characterization of synthetic parameters and measurement of transcript molecular weights. J. Mol. Biol. 103, 681-697.

Smith, L.H. and Sinsheimer, R.L. (1976b). The *in vitro* transcription units of bacteriophage ϕX174. II. *In vitro* initiation sites of ϕX174 transcription. J. Mol. Biol. 103, 699-710.

Smith, L.H. and Sinsheimer, R.L. (1976c). The *in vitro* transcription units of bacteriophage ϕX174. III. Initiation with specific 5' end oligonucleotides of *in vitro* ϕX174 RNA. J. Mol. Biol. 103, 711-735.

Spencer, J.H. and Boshkov, L. (1973). Nucleotide clusters in deoxyribonucleic acids. VIII. The pyrimidine oligonucleotides of bacteriophage S13 wild-type DNA. Can. J. Biochem. 51, 1206-1211.

Spencer, J.H., Cerny, R., Cerna, E. and Delaney, A.D. (1972). Characterization of bacteriophage S13 su N15 single strand and replicative form deoxyribonucleic acid. J. Virol. 10, 134-141.

Spencer, J.H. and Changaff, E. (1963). Studies on the nucleotide arrangement in deoxyribonucleic acids. V. Pyrimidine nucleotide clusters: Isolation and characterization. Biochim. Biophys. Acta 68, 9-17.

Spencer, J.H., Rassart, E., Kaptein, J.S., Harbers, K., Grosveld, F.G. and Goodchild, B. (1978). The S13 genome. Cold Spring Harbor Symp. Quant. Biol. (in press).

Sobell, H.M. (1972). Molecular mechanism for genetic recombination. Proc. Natl. Acad. Sci. USA 69, 2483-2487.

Sogo, J.M., Portmann, R., Koller, T.L. and Zillig, W. (1976). Effects of heparin and rifampicin on the binding of DNA-dependent RNA polymerase of *E. coli* to T7 DNA. IUB. Tenth international congress of Biochemistry Hamburg, p. 48, July 25.

Squires, C.L., Lee, F.D. and Yanofsky, C. (1975). Interaction of the *Trp* repressor and RNA polymerase with the *trp* operon. J. Mol. Biol. 92, 93.

Steitz, J.A. (1969). Polypeptide chain initiation: Nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 DNA. Nature 224, 957-964.

Sugimoto, K., Okamoto, T., Sugisaki, H. and Takanami, M. (1975). The nucleotide sequence of an RNA polymerase binding site on bacteriophage fd DNA. Nature 253, 410-414.

Suguira, M., Okamoto, T. and Takanami, M. (1970). RNA polymerase σ- factor and the selection of initiation site. Nature 225, 598-600.

Sulkowski, E. and Lashowski, M. Sr. (1971). Inactivation of 5'-nucleotidase in commercial preparation of venom exonuclease. Biochim. Biophys. Acta 240, 443-447.

Sun, L. and Singer, S. (1974). Reaction of adenosine with ethylating agents. Biochemistry 13, 1913-1920.

Takanami, M., Okamoto, T., Sugimoto, K. and Sugisaki, H. (1975). Studies on bacteriophage fd DNA. I. A cleavage map of the fd genome. J. Mol. Biol. 95, 21-31.

Takanami, M., Sugimoto, K., Sugisaki, H. and Okamoto, T. (1976). Sequence of promoter for coat protein gene of bacteriophage fd. Nature 260, 297-302.

Takaya, T. and Takanami, M. (1974). Isolation of DNA segments containing promoters from bacteriophage T3 DNA. Biochemistry 13, 5388-5394.

Tessman, E.S. and Shleser, R. (1963). Genetic recombination between phages S13 and ϕX174. Virology 19, 239-240.

Tessman, I. (1959). Some unusual properties of the nucleic acid in bacteriophage S13 and ϕX174. Virology 7, 263-275.

Tessman, I., Tessman, E.S., Pollock, T.J., Borrás, M.-T., Pjea, A. and Baker, R. (1976). Reinitiation mutants of gene B of bacteriophage S13 that mimic gene A mutants in blocking replicative form DNA synthesis. J. Mol. Biol. 103, 583-598.

Travers, A. (1970). RNA polymerase and T4 development. Cold Spring Harbor Symp. Quant. Biol. 35, 241-251.

Travers, A. (1973). Control of ribosomal RNA synthesis *in vitro*. Nature 244, 15-18.

Travers, A. (1974). RNA polymerase-promoter interactions: Some general principles. Cell 3, 97-104.

Travers, A. (1976). Multiple modes of ribosomal RNA transcription *in vitro*. Cell 8, 605-609.

Travers, A., Baillie, D.L. and Pederson, S. (1973). Effect of DNA conformation of ribosomal RNA synthesis *in vitro*. Nature New Biol. 243, 161-163.

Vanderbilt, A., Borrás, M.-T., Germeraad, S., Tessman, I. and Tessman, E.S. (1972). A promoter site and polarity gradients in phage S13. Virology 50, 171-179.

Vanderbilt, A.S. and Tessman, I. (1970). Mutagenic methods for determining which DNA strand is transcribed for individual viral genes. Nature 228, 54-56.

Van de Sande, J.H., Loewen, P.C. and Khorawa, H.G. (1972). Studies on polynucleotides. CXVIII. A further study on ribonucleotide incorporation into DNA by DNA polymerase I of *E. coli*. J. Biol. Chem. 247, 6140-6148.

Vogt, V. (1969). Breaks in DNA stimulate transcription by core DNA polymerase. Nature 233, 854.

Von Hippel, P., Revzin, A., Gross, C. and Wang, A. (1974). Nonspecific DNA binding of genome regulating proteins as a biological control mechanism. I. The *lac* operon. Equilibrium aspects 71, 4808.

Walz, A. and Pirrotta, V. (1975). Sequence of the *Pr* promoter of phage λ. Nature 254, 118-121.

Walz, A., Pirotta, V. and Ineichen, K. (1976). λ repressor regulates the switch between *Pr* and *Pr m* promoters. Nature 262, 665-669.

Watson, J.D. and Crick, F.C. (1953). Molecular structure of nucleic acids. A structure for deoxyribose nucleic acids. Nature 171, 737-738.

Weisbeek, P.J., Borrias, W.E., Langeveld, S.A., Baas, P.D. and Van Arkel, G.A. (1977). Bacteriophage ϕX174: Gene A overlaps gene B. Proc. Natl. Acad. Sci. USA 74, 2504-2508.

Weiss, S.B. and Gladstone, L. (1959). A mammalian system for the incorporation of cytidine triphosphate into ribonucleic acid. J. Amer. Chem. Soc. 81, 4118-4119.

Weissmann, C., Billeter, M.A., Goodman, H.M., Hindley, J. and Weber, H. (1973). Structure and function of phage DNA. In Annual Review of Biochemistry. Snell, E.E. (ed.). Vol. 42, pp. 303-328. Annuel Review Inc.

Willmund, R. and Kneser, H. (1973). Different binding of RNA polymerase to individual promoters. Mol. Gen. Genetics 126, 165-175.

Wu, C.W., Yarbrough, L.R. and Hillel, Z. (1975). Demonstration by fluorescence depolarisation of σ release during transcription. Federation Proceedings 34, 2441.

Wu, R., Bambara, R. and Jay, E. (1974). Recent advances in DNA sequence analysis. CRC Crit. Rev. Biochem. 2, 455-512.

Wu, R. and Kaiser, A.D. (1968). Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. J. Mol. Biol. 35, 523-537.

Yanofsky, C. (1976). Regulation of transcription initiation and termination in the control of expression of the tryptophan operon of *E. coli*. In Molecular Mechanisms in the control of gene expression. Nierlich, D.P., Rutter, W.J. and Fox, C.F. (eds.). Vol. 5, pp. 75-87. Academic Press.

Ziff, E.B., Sedat, J.W. and Galibert, F. (1973). Determination of the nucleotide sequence of a fragment of bacteriophage ϕX174 DNA. Nature New Biol. 241, 34-37.

Zillig, W., Fuchs, E., Palm, P., Rabussay, D. and Zechel, K. (1970). On the different subunits of DNA dependent RNA polymerase from *E. coli* and their role in the complex function of the enzyme. In first Lepetit colloquium on RNA polymerase. Silvestri, L. (ed.). pp. 151-157. North Holland, Amsterdam.

Zillig, W., Palm, P. and Heil, A. (1976). Function and reassembly of subunits of DNA dependent RNA polymerase. In RNA polymerase. Losick, R. and Chamberlin, M.J. (eds.). pp. 101-126. Cold Spring Harbor Laboratory.

Zillig, W., Zechel, K. and Halbwachs, H. (1970). A new method of large scale preparation of highly purified DNA-dependent RNA-polymerase from *E. coli*. Hoppe-Seylers Z. Physiol. Chem. 351, 221-224.

Zillig, W., Zechel, K., Rabussay, D., Schachner, M., Sethi, V., Palm, P., Heil, A. and Seifert, W. (1970). On the role of different subunits of DNA-dependent RNA polymerase from *E. coli* in the transcription process. Cold Spring Harbor Symp. Quant. Biol. 35, 47-58.

Zollinger, M., Guertin, M. and Mamet-Bratley, M.D. (1977). A new electron microscopic method for studying protein-nucleic acid interactions. Anal. Biochem. 82, 196-203.

Zubay, G. (1966). Isolation of transfer RNA. In Procedures in Nucleic Acid Research. Cantoni, G.L. and Davies, D.R. (eds.). Vol. 1, pp. 455-460. Harper and Row, New York.

APPENDIX I

ADDITIONAL RESULTS TO CHAPTER II

figure 1. SDS gel electrophoresis of the RNA polymerase preparation.

figure 2. Immunological assay of the RNA polymerase preparation against anti 66,000 daltons S_1 protein.

figure 3. Effect of albumin concentration to the extent of binding of the RNA polymerase on S13 RF DNA.

figure 4. Polyacrylamide gel electrophoresis of the RNA polymerase protected fragments. (Size determination).

figure 5. Pyrimidine oligonucleotides of bacteriophage S13 RF I and of the RNA polymerase protected fragments.

figure 6. Polyacrylamide gel electrophoresis of *Hae*III or *Hin*d digested ^{3}H-labeled or ^{32}P-labeled S13 RF I DNA before and after binding with the RNA polymerase.

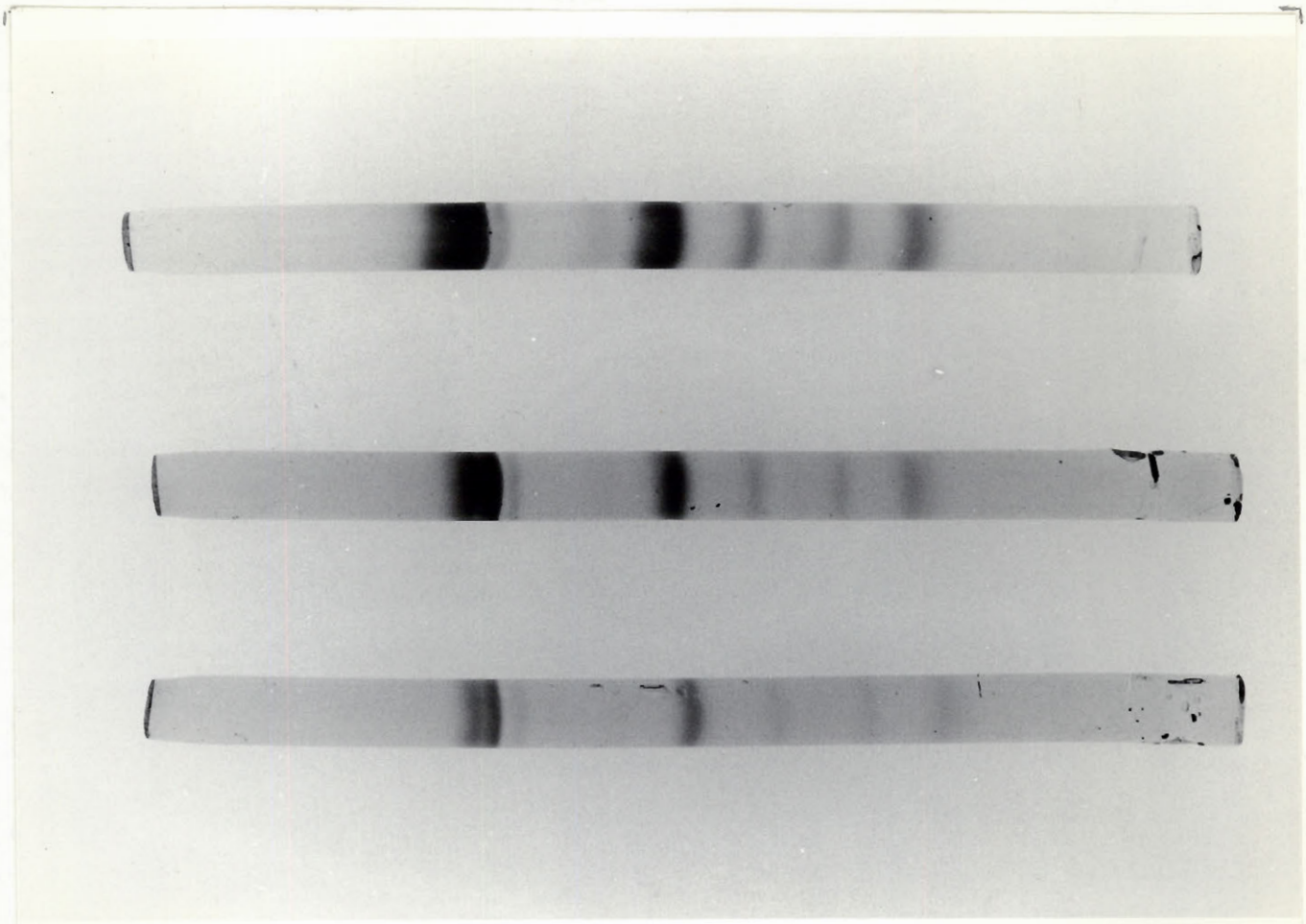
Table I. Binding of *E. coli* RNA polymerase to the *Hin*d and *Hae*III restriction fragment of S13 RF I DNA.

Figure 1 : 7.5% SDS-polyacrylamide gel electrophoresis of 10, 25 and 50 µg of the *E. coli* RNA polymerase preparation.

Figure 2 : Immunological assay of the RNA polymerase preparation against anti 66,000 daltons S_1 protein (center).

A and B: 1 and 3 µl of 66,000 daltons S_1 protein at concentration of 1.48 $O.D_{280}$/ml

C and D: 1 and 3 µl of our RNA polymerase preparation at 7 mg/ml.



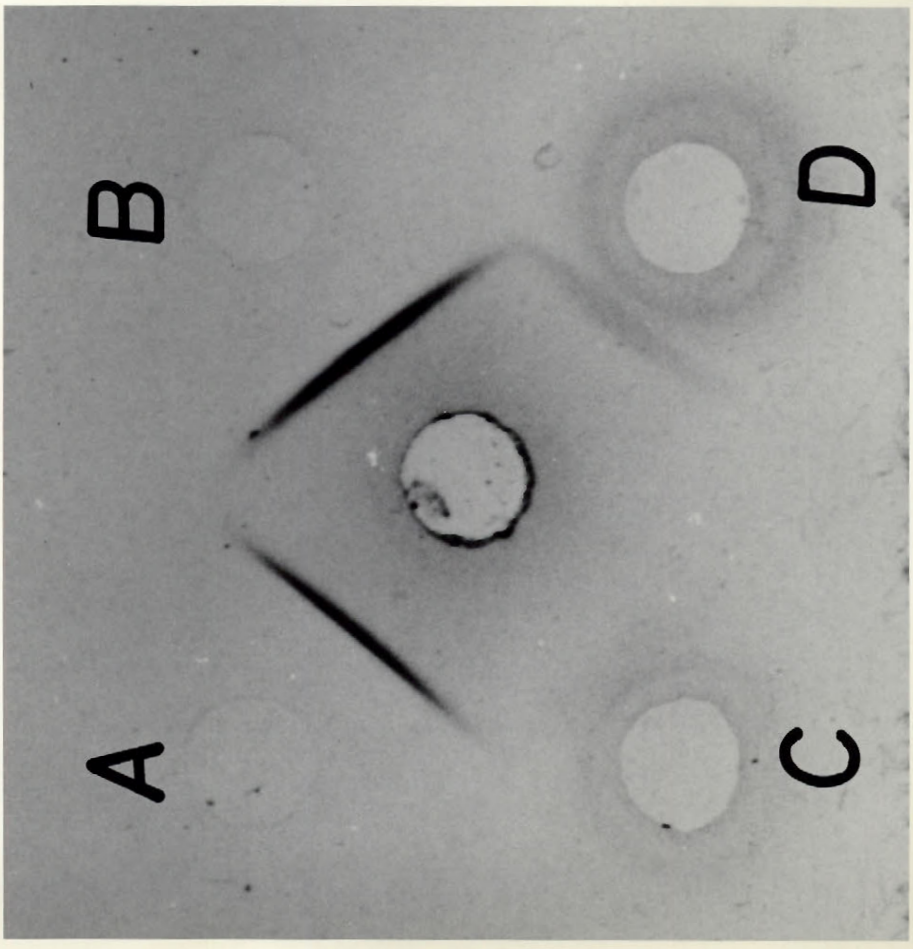

B
D
A
C

Figure 3 : Addition of increasing concentration of albumin on the extent of binding of the RNA polymerase to S13 RF DNA.
E. coli RNA polymerase was incubated with ^{32}P-labeled S13 RF DNA at various albumin concentrations in binding buffer containing 0.35 M Kcl for 10 min at 37°C. The DNA not protected by the polymerase was hydrolyzed by digestion with DNAse A for 40 min at 37°C. The resultant mixture was passed through a nitrocellulose filter and the DNA-RNA polymerase complexes retained on the filter and measured for ^{32}P radioactivity. The results are expressed as a percentage of the total ^{32}P radioactivity in the initial binding mixture.

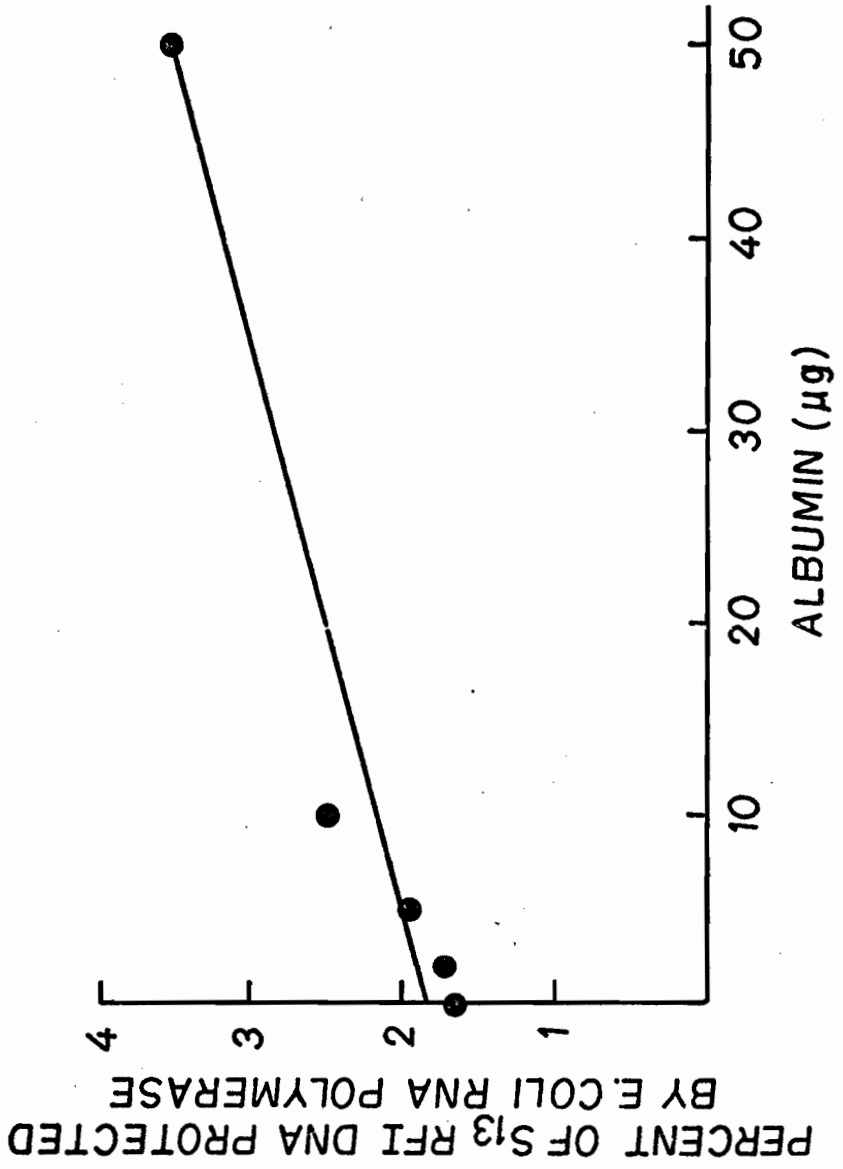

PERCENT OF S13 RFI DNA PROTECTED
BY E.COLI RNA POLYMERASE
4
3
2
1
10
20
30
40
50
ALBUMIN (µg)

Figure 4 : Polyacrylamide gel electrophoresis of the RNA polymerase protected fragments. After binding the RNA polymerase to ^{32}P-labeled S13 RF DNA followed by DNAse digestion, the protected DNA fragments were purified on Sephadex G-100, and loaded on a 12.5% polyacrylamide gel in the presence of 8 M Urea. The samples were run for 5 hours at 8 mA per gel, and then sliced into 2 mm slices, which were measured for their radioactive content. B is the bromophenol blue dye marker. The positions of denatured restriction fragment of 72, 100 and 185 markers nucleotides in length are indicated by arrows.

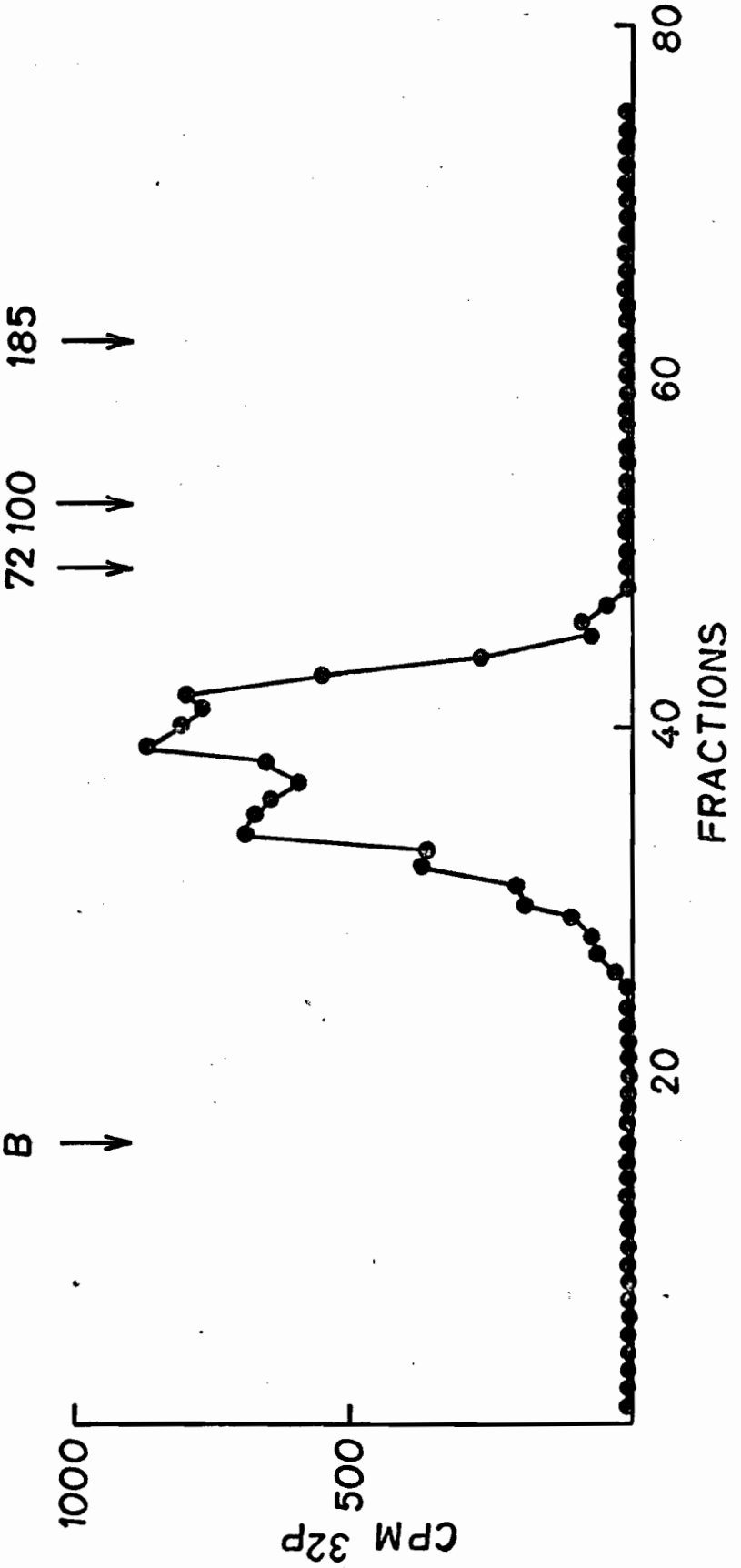

CPM 32P
1000
500
20
40
60
80
FRACTIONS
B
72 100
185

Figure 5 : Pyrimidine oligonucleotides of A) bacteriophage S13 RF I DNA and B) the RNA polymerase protected fragments isolated from S13 RF I DNA.

A) Uniformly ^{32}P-labeled RF I DNA was degraded to pyrimidine oligonucleotides which were separated by electrophoresis (horizontal dimension) and homochromatography (vertical dimension).

B) RNA polymerase was incubated with ^{32}P-labeled RF I DNA followed by a DNAse treatment. The protected DNA fragments after purification on Sephadex G100 were degraded to pyrimidine oligonucleotides and then separated as in A.

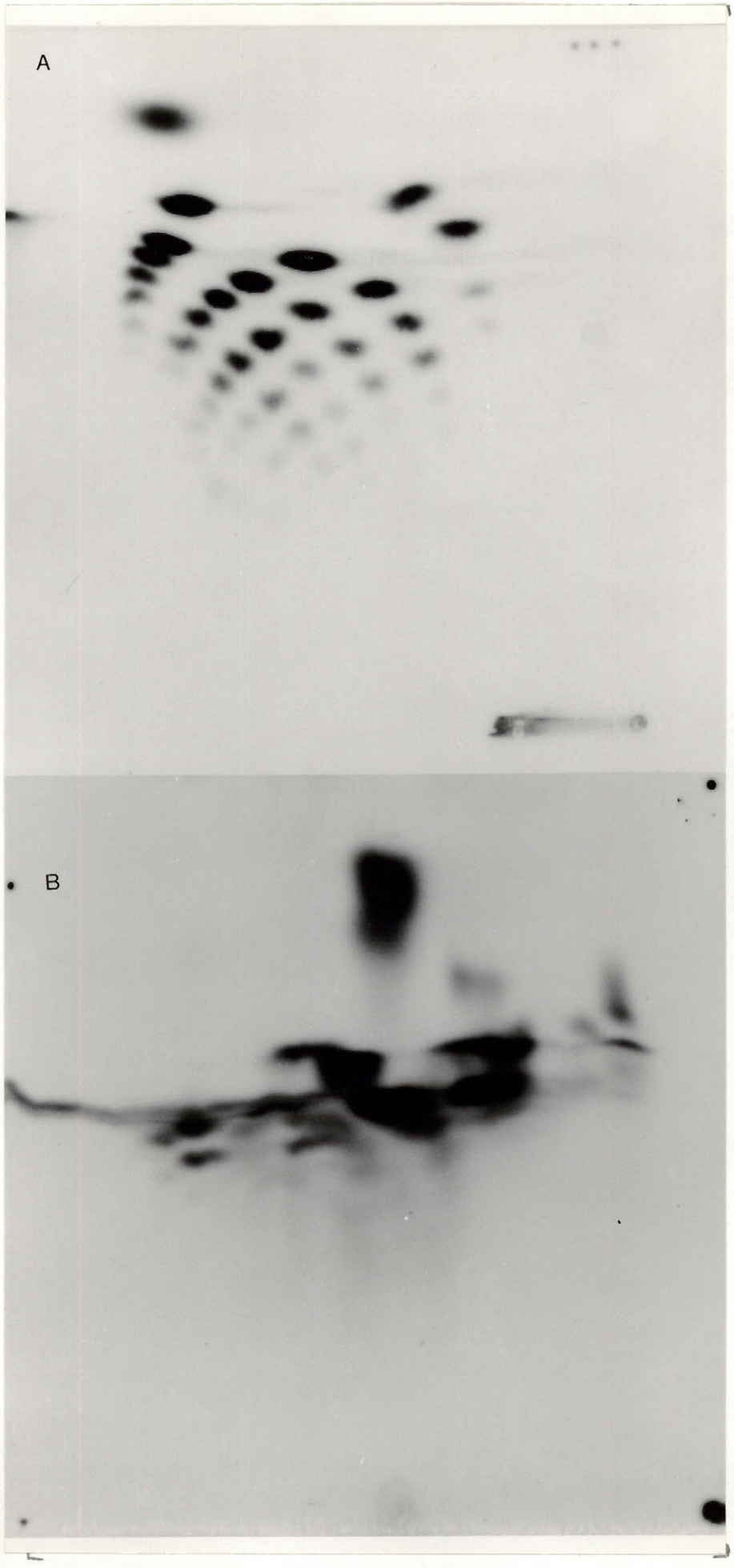

A
B

Figure 6 : Polyacrylamide gel electrophoresis of *Hae*III (a) or *Hin*d (b) digested ^{3}H-labeled S13 RF I DNA before binding to the RNA polymerase and ^{32}P-labeled S13 RF I DNA after binding to the RNA polymerase.

^{32}P-labeled S13 RF I DNA digested with *Hae*III (a) or *Hin*d (b) was incubated with the RNA polymerase for 15 min at 37°C. The mixture was passed through a nitrocellulose filter and the complexes retained on the filter. The restriction fragments involved in the complex were extracted from the filter with 0.2% SDS, ethanol precipitated and mixed with *Hae*III or *Hin*d digested ^{3}H-labeled S13 RF DNA respectively, and loaded on a 5% polyacrylamide gel. Electrophoresis was at 4 mA pertube for 10 h. Slices were 2 mm thick giving a direct measure of the distance of migration at the end of the run (mobility in mm).

Solid lines are ^{32}P-labeled S13 DNA.

Dashed lines are ^{3}H-labeled S13 DNA.

Quantitations of the results are shown in table I.

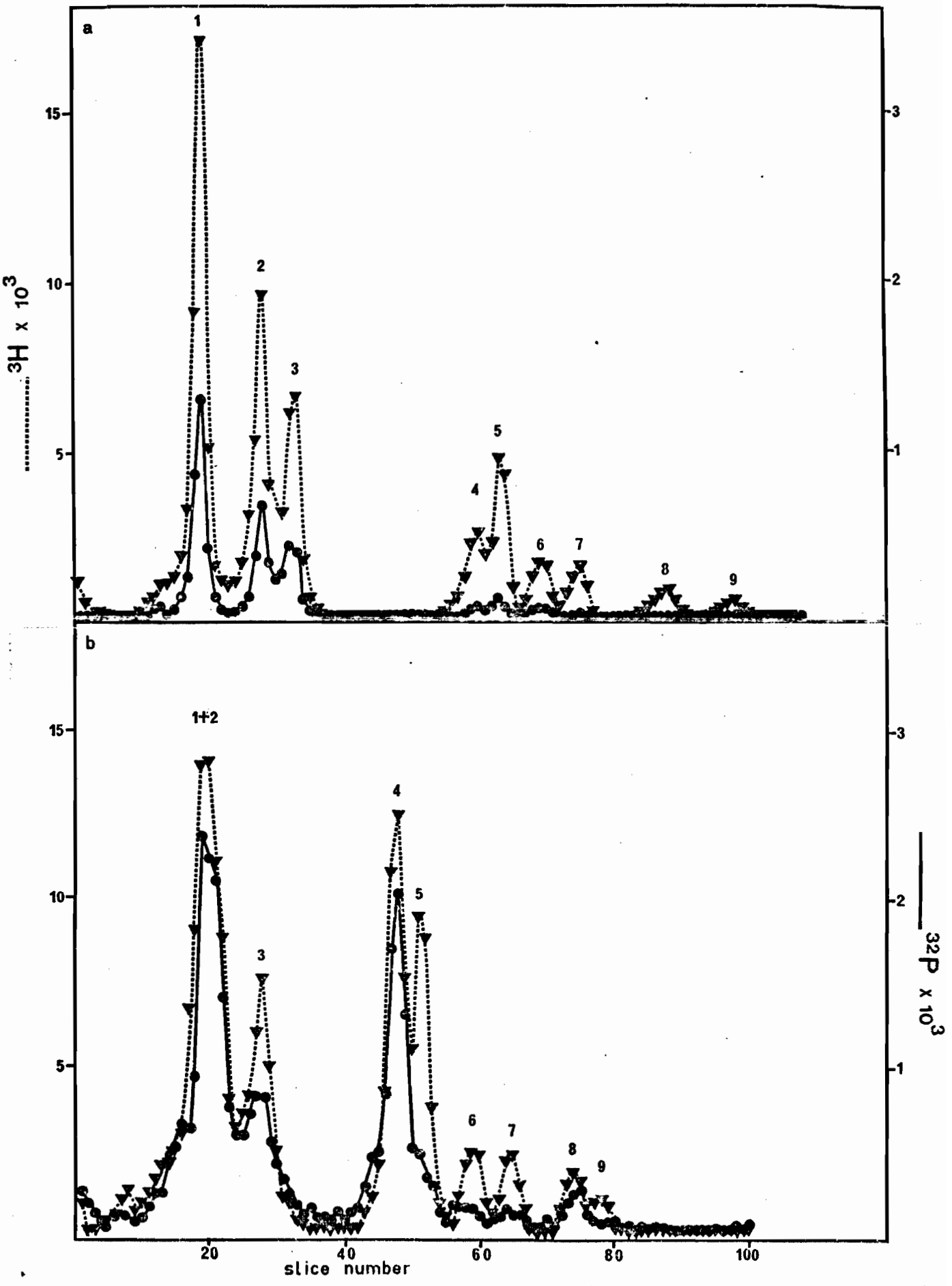

a
b
1
2
3
4
5
6
7
8
9
1+2
15
10
5
3
2
1
$^{3}H \times 10^{3}$
$^{32}P \times 10^{3}$
20
40
60
80
100
slice number

TABLE I

Binding of *E. coli* RNA Polymerase to the *Hin*d and *Hae*III Restriction Fragments of S13 RF I DNA[a]

Hae fragment number	No. of fragment	Relative intensity	*Hin*d fragment number	No. of fragment	Relative intensity
1	1	2.2	1	1	2.8
2	1	2.2	2	1	0.7
3	1	1.7	3	1	1.1
4	1	0.1	4	3	1.7-3.4-5.1
5	2	0.1-0.2	5	3	0.2-0.4-0.6
6	1	2.2	6	1	0.6
7	1	0.3	7	1	0.1
8	1	0.3	8	1	2.3
9	1	0	9	1	0.1

[a]means of 3 experiments.

APPENDIX II

AUTORADIOGRAMS OF T4 ENDONUCLEASE IV DIGESTS AND METHYLMETHANE SULFONATE DIGESTS, OLIGONUCLEOTIDE SEQUENCE ANALYSIS OF FRAGMENT *HIND* 8.

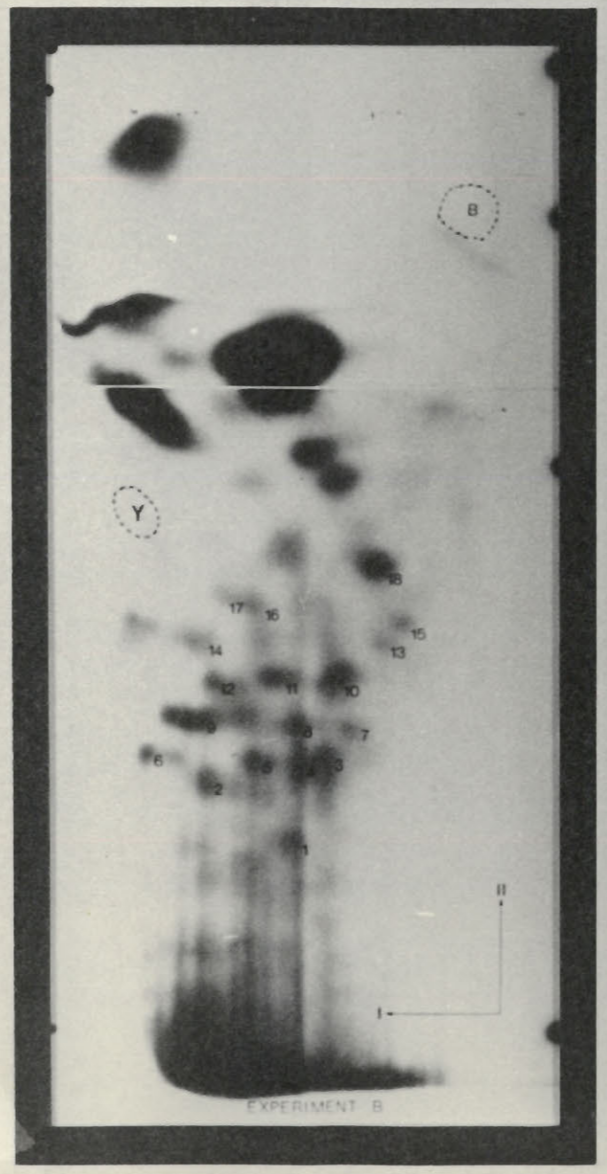
1. Autoradiograms of the oligonucleotides released by DNA alkylation and hydrolysis with methylmethane sulfonate (experiments A and B).
2. Autoradiograms of the oligonucleotides released by T4 endonuclease IV digestion (experiments D and F).

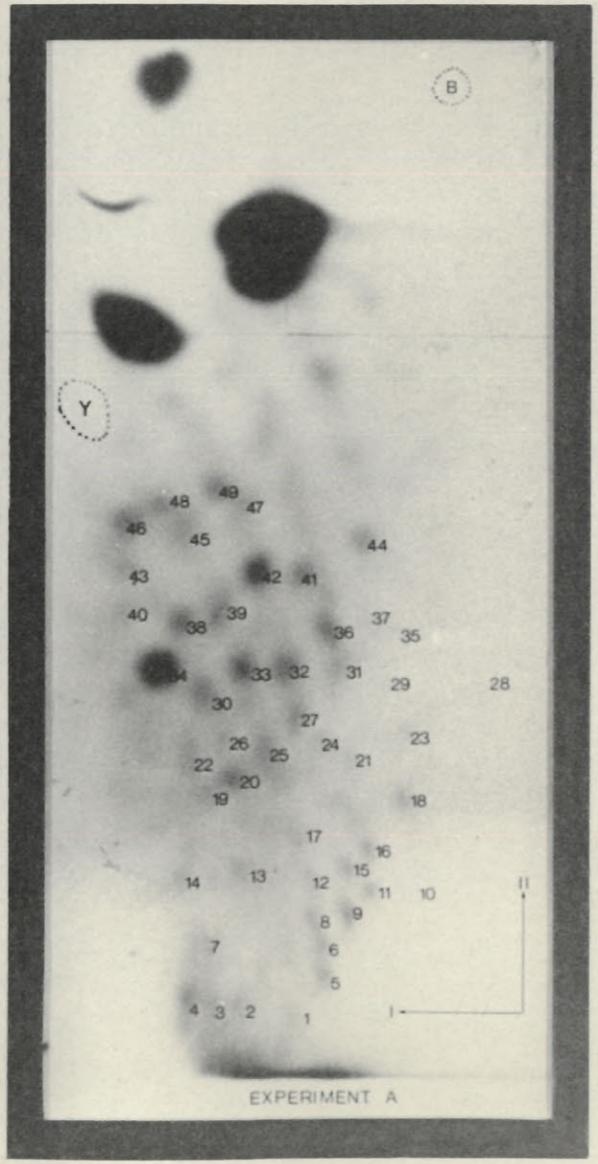
Autoradiograms of partial 3' exonuclease digests of:

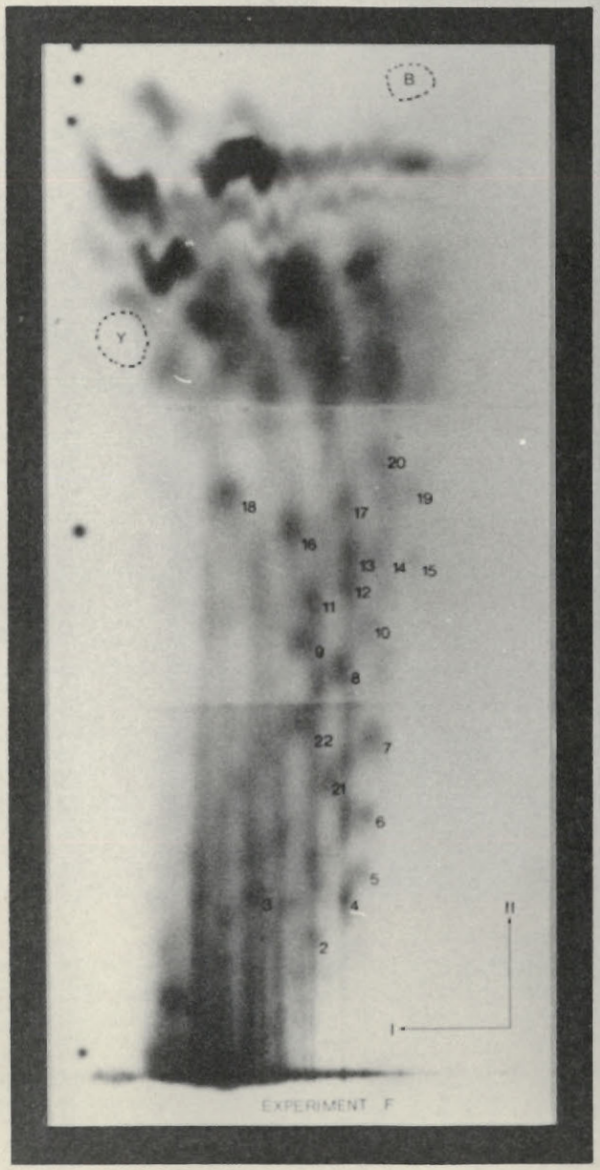
3. Experiment A, oligonucleotides 8 and 9.
4. Experiment A, oligonucleotides 11 and 13.
5. Experiment A, oligonucleotides 14 and 15.
6. Experiment A, oligonucleotides 16 and 18.
7. Experiment A, oligonucleotides 20 and 21.
8. Experiment A, oligonucleotides 23 and 26.
9. Experiment A, oligonucleotides 31 and 32.
10. Experiment A, oligonucleotides 33 and 34.
11. Experiment A, oligonucleotides 36 and 37.
12. Experiment A, oligonucleotides 38 and 42.
13. Experiment A, oligonucleotides 44 and 46.
14. Experiment A, oligonucleotides 47 and 48.
15. Experiment A, oligonucleotide A49.
    Experiment B, oligonucleotide 8.
16. Experiment B, oligonucleotides 9 and 17.
17. Experiment C, oligonucleotides 2 and 5.
18. Experiment D, oligonucleotides 8 and 9.
19. Experiment D, oligonucleotides 11 and 12.

20. Experiment D, oligonucleotides 13 and 15.
21. Experiment D, oligonucleotides 16 and 17.
22. Experiment D, oligonucleotides 18 and 22.
23. Experiment D, oligonucleotides 24 and 25.
24. Experiment D, oligonucleotides 27 and 28.
25. Experiment F, oligonucleotides 1 and 9.
26. Experiment F, oligonucleotides 12 and 21.
27. Experiment F, oligonucleotide 22.

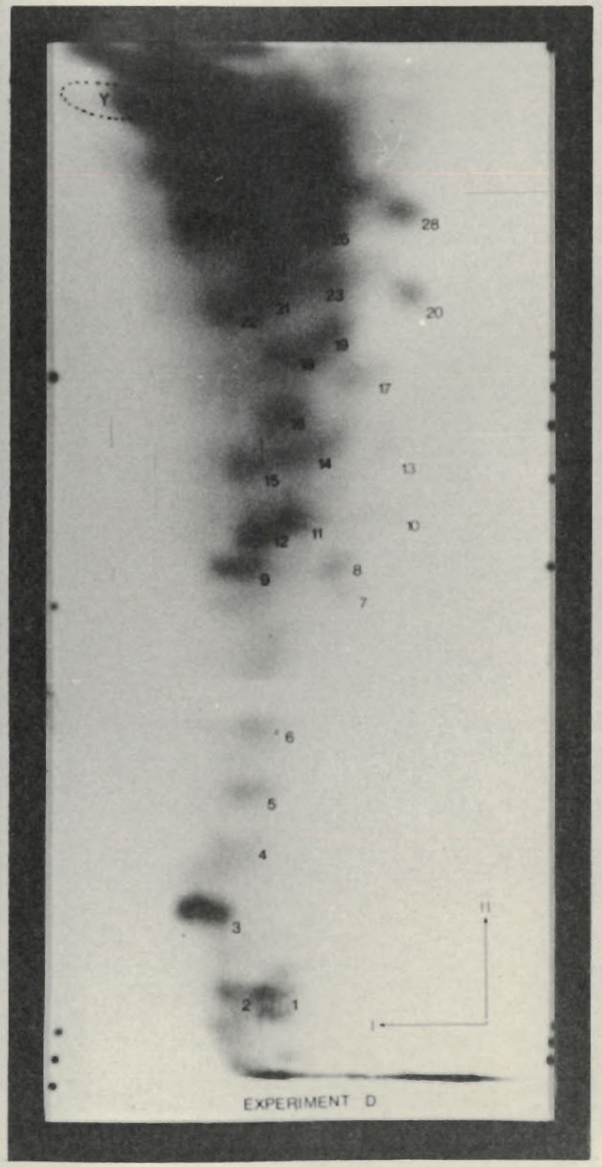
In the photographic reproduction of the original autoradiograms, some of the darker spots have blurred into each other because of overexposure in attempting to show the lighter spots. In some cases however, the lighter spots could still not be reproduced. The original autoradiograms are available.


B
Y
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
II
I
EXPERIMENT B

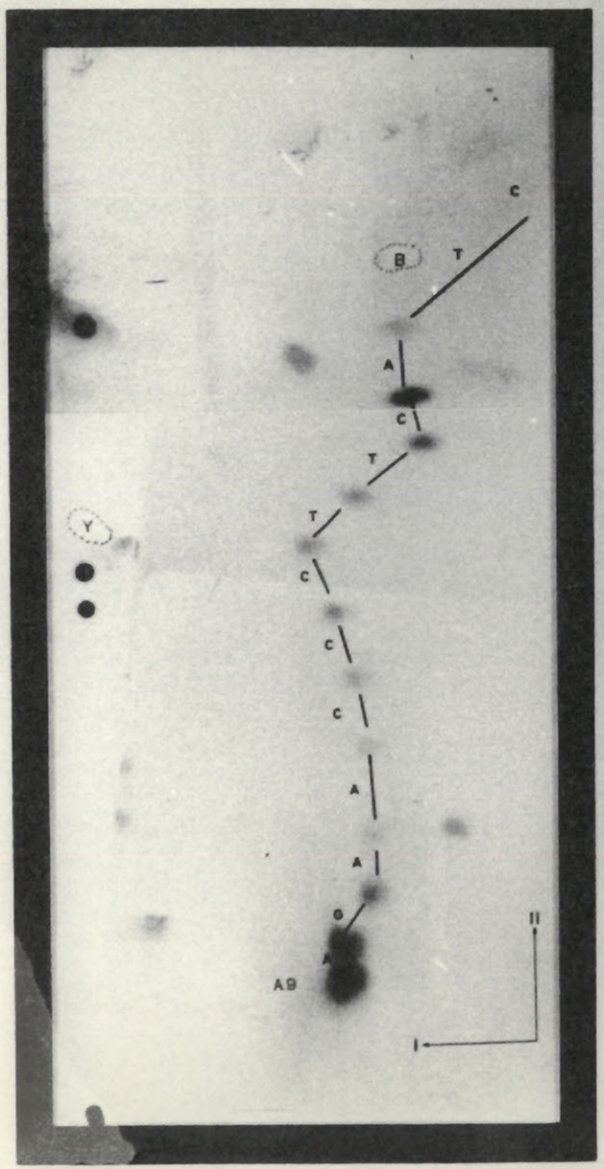

B
Y
48
49
47
46
45
44
43
42
41
40
39
38
37
36
35
34
33
32
31
29
28
30
27
26
25
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
II
I
EXPERIMENT A

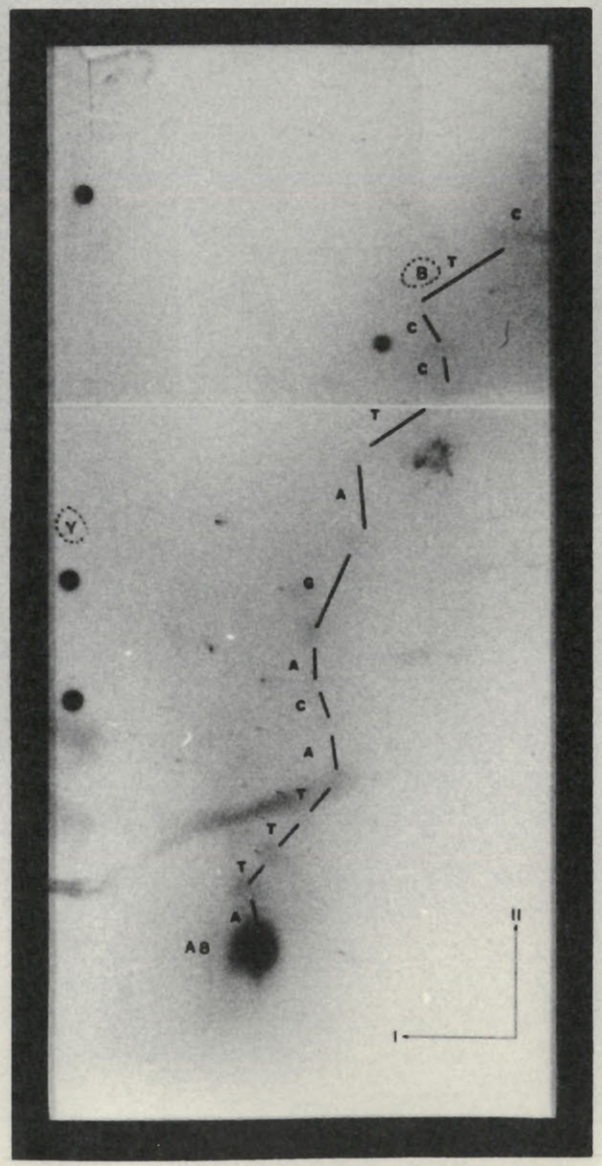


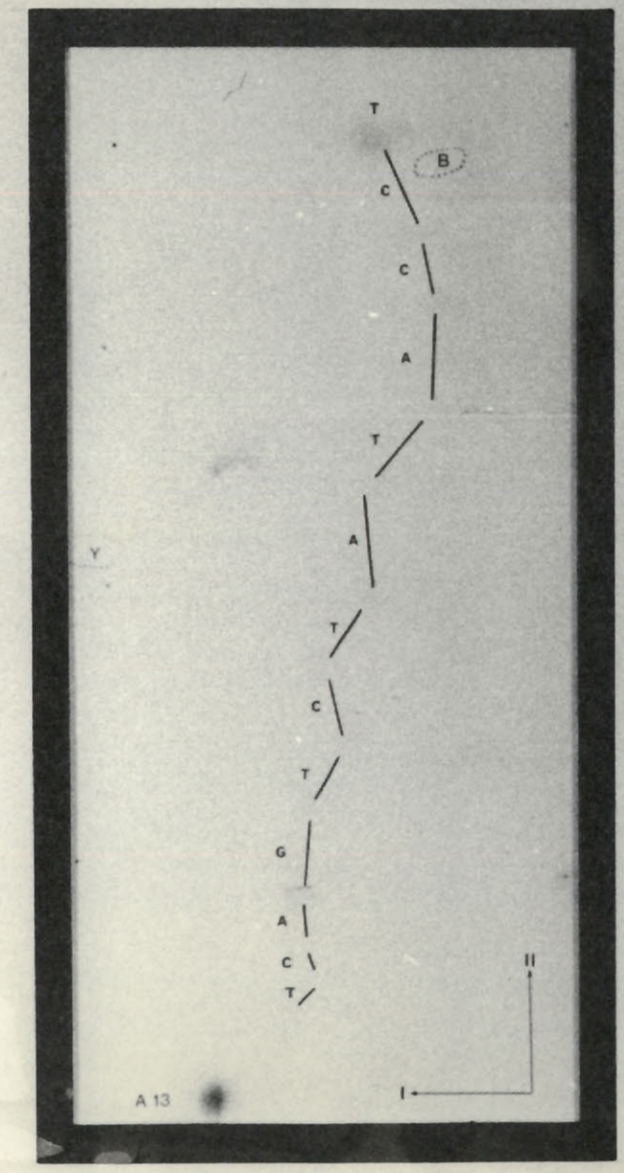
EXPERIMENT F

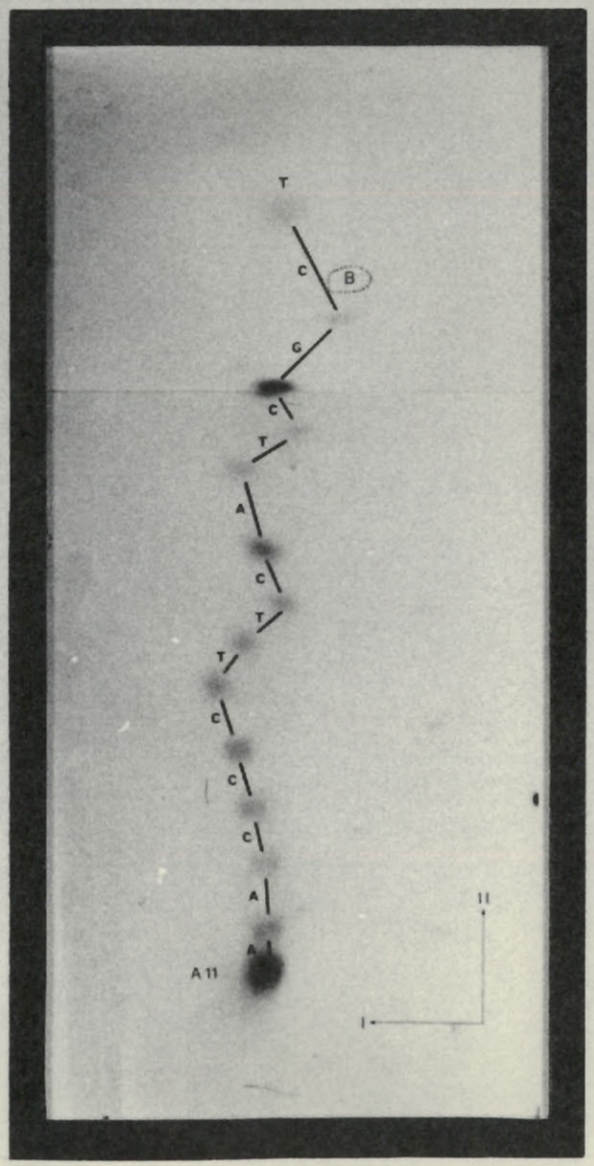


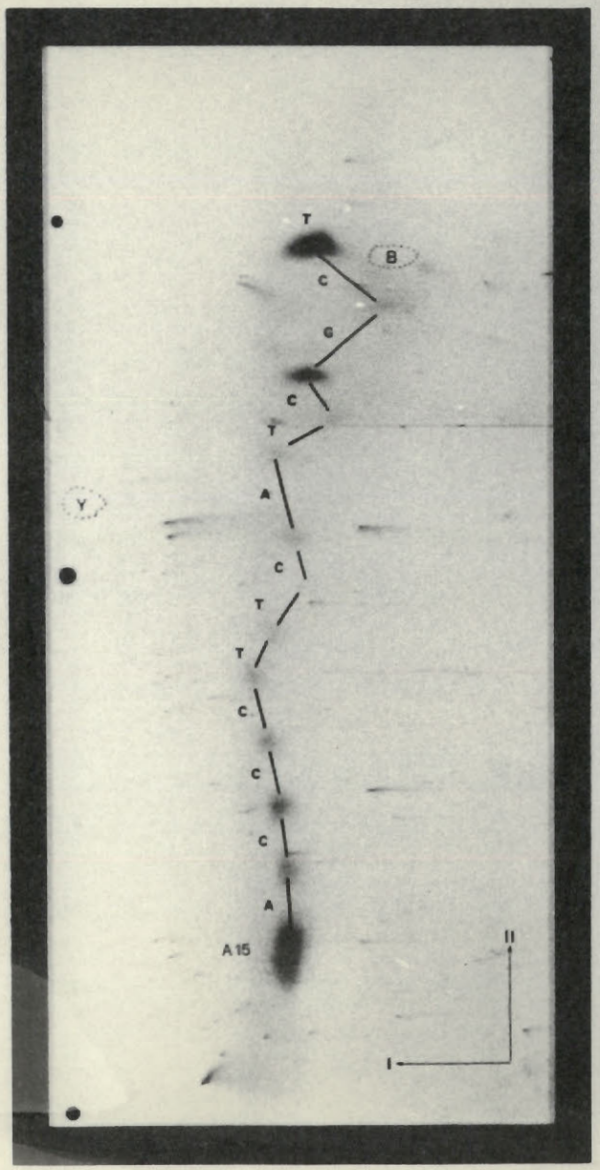
EXPERIMENT D

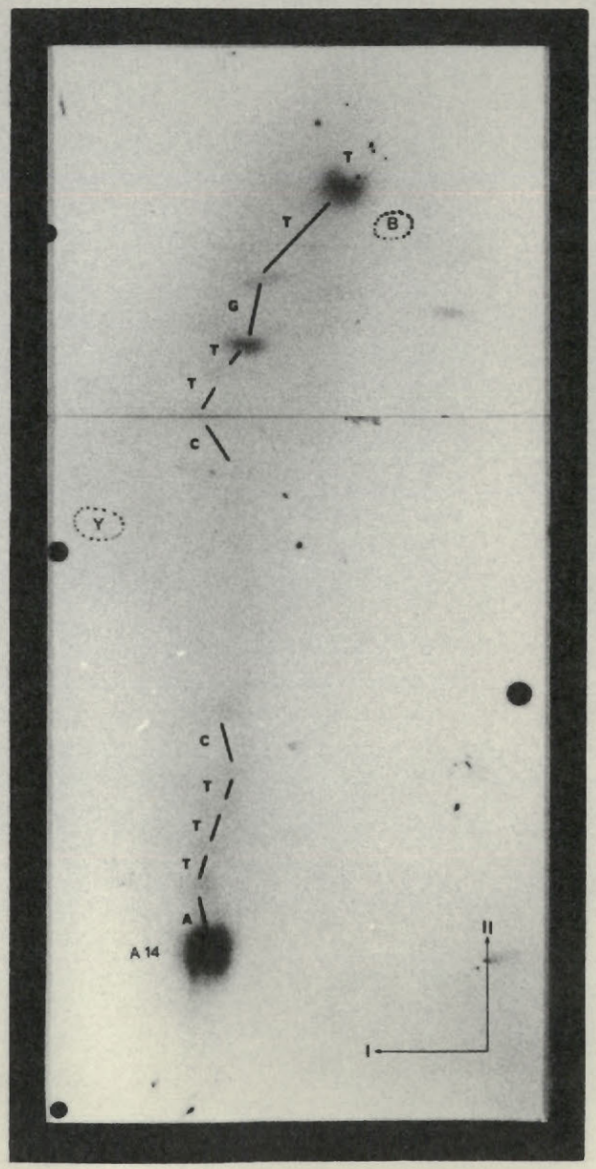

B
C
T
A
C
T
T
C
C
C
A
A
G
A
A9
Y
II
I

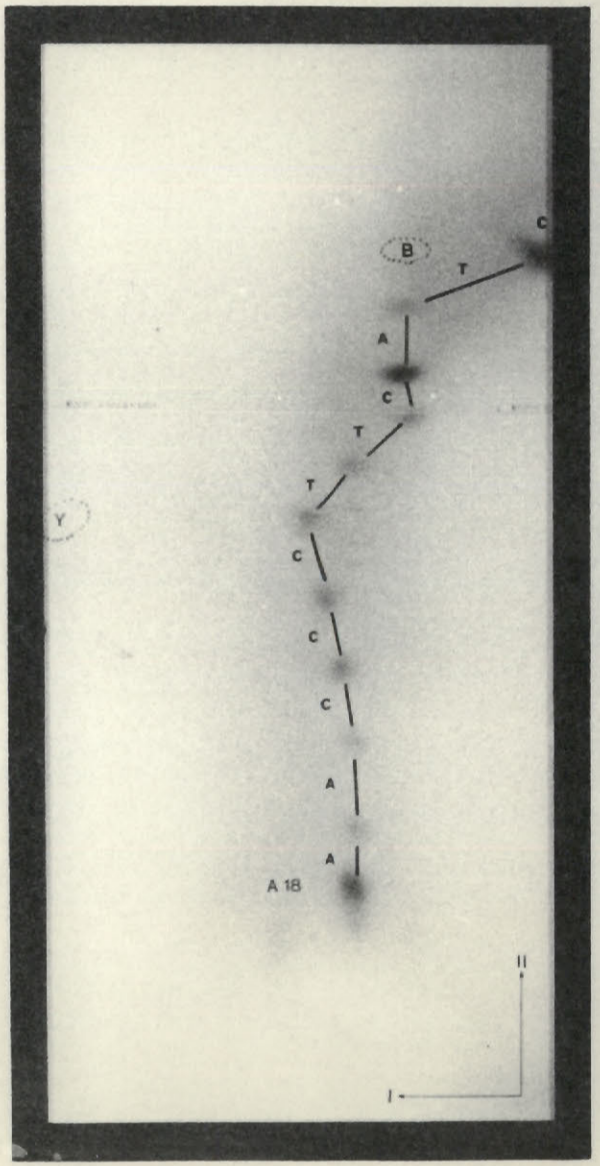

B
C
T
C
C
T
A
G
A
C
A
T
T
T
A
A8
Y
II
I

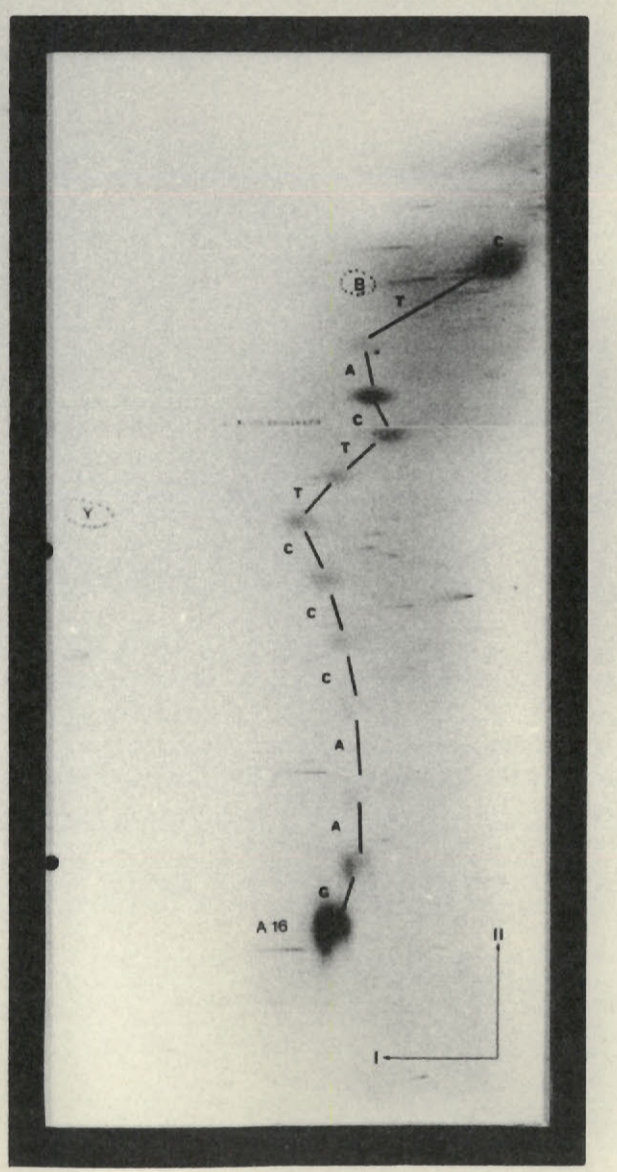

T
B
C
C
A
T
A
T
C
T
G
A
C
T
Y
II
I
A 13

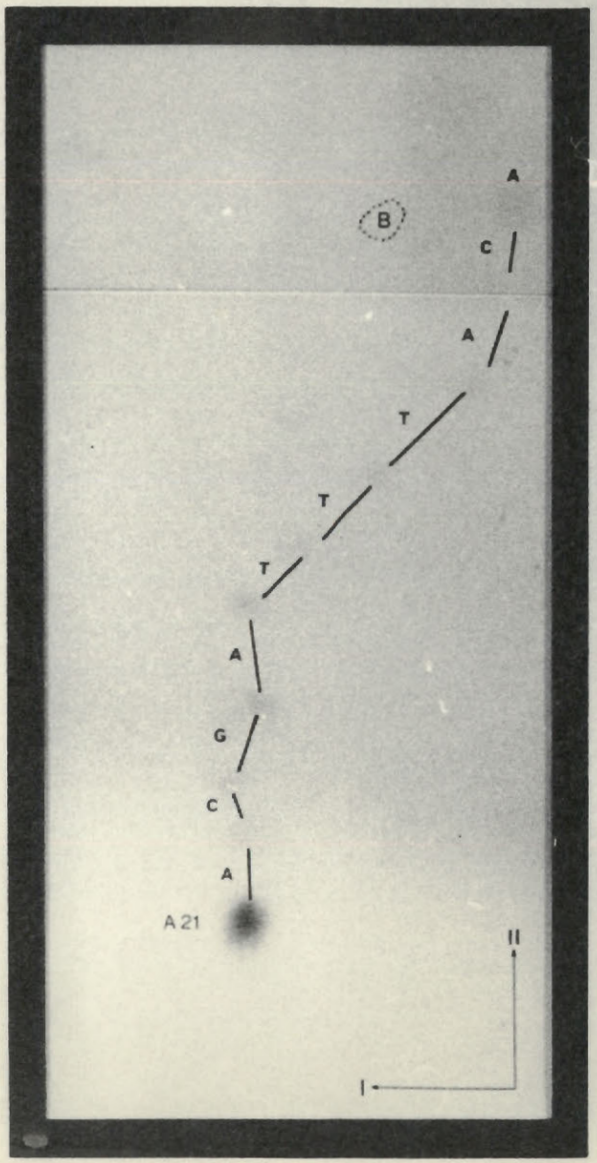

T
C
B
G
C
T
A
C
T
T
C
C
C
C
A
A
II
I
A 11

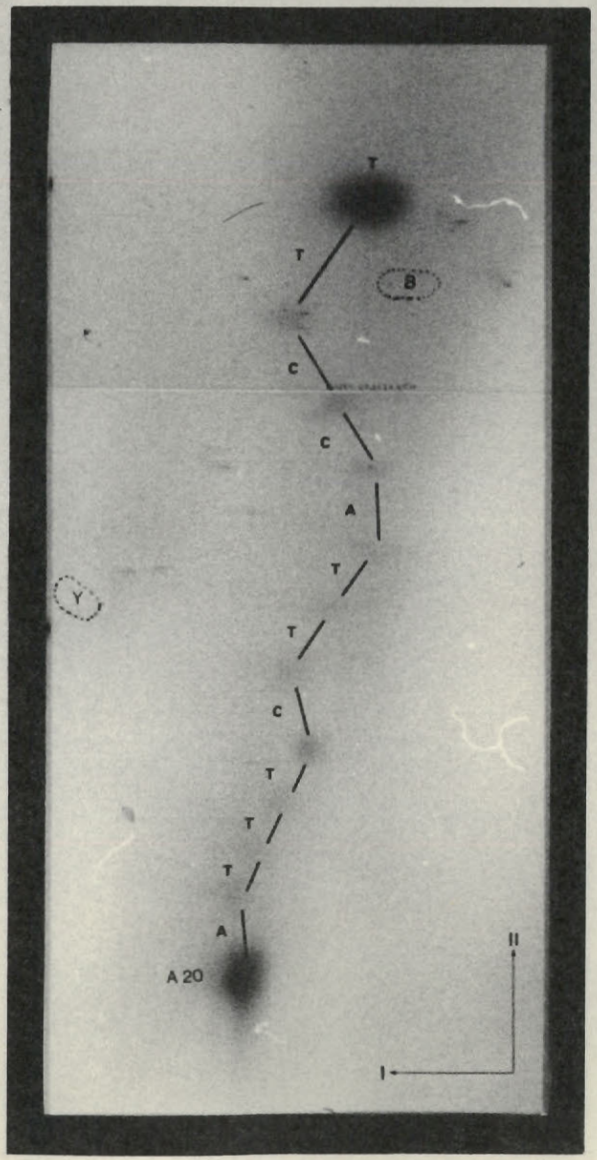

T
B
C
G
C
T
A
C
T
T
C
C
C
A
A 15
Y
II
I

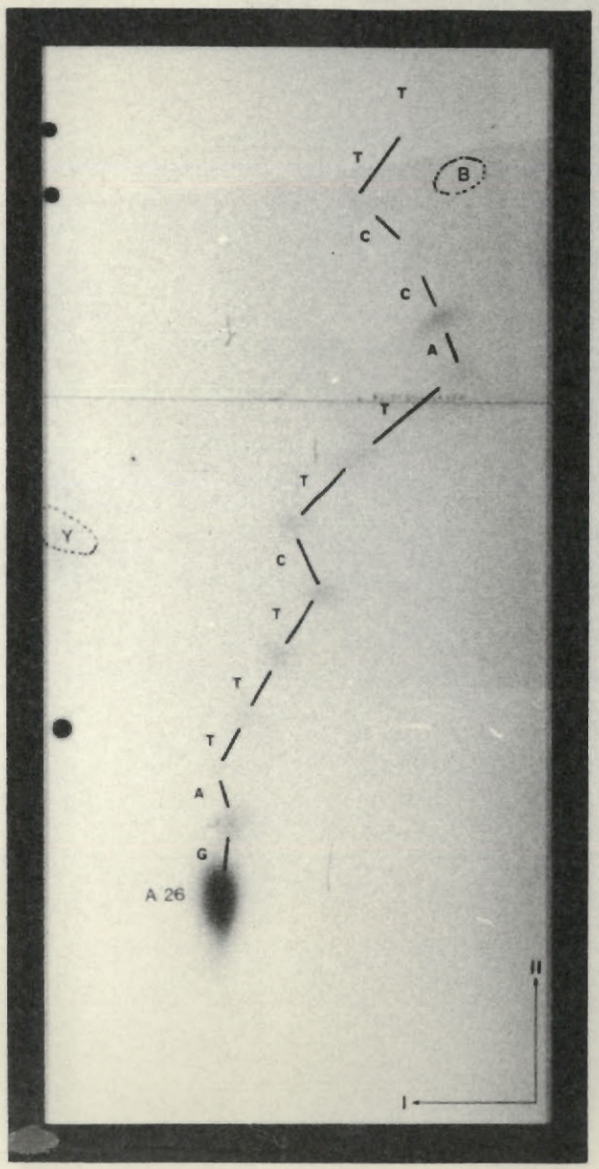

T
T
B
G
T
T
C
Y
C
T
T
T
A
A 14
II
I

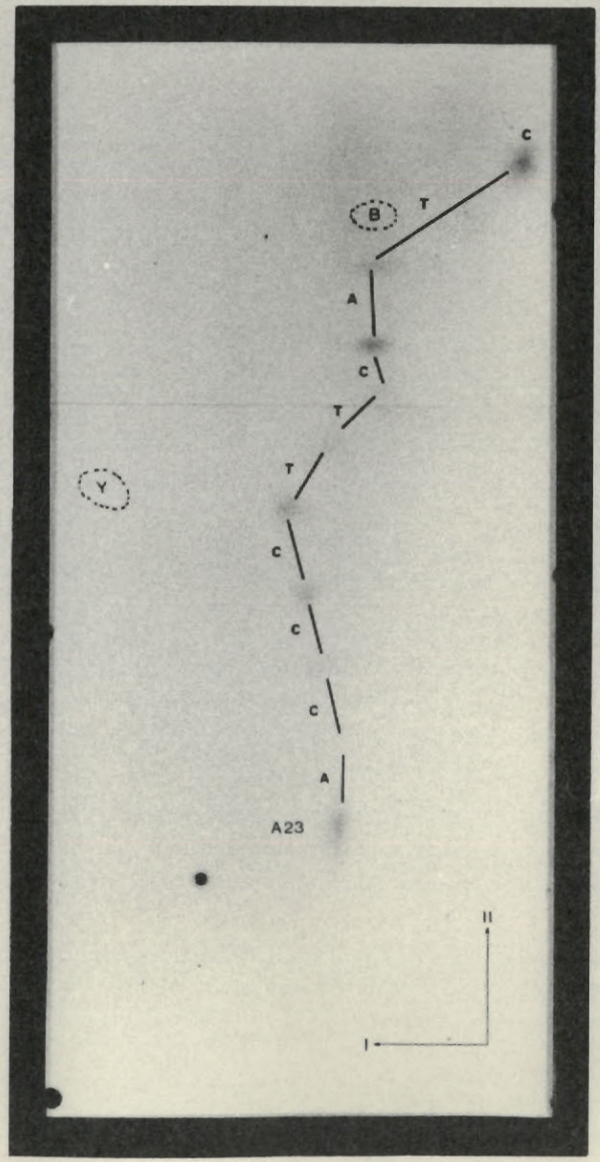

B
C
T
A
C
T
T
Y
C
C
C
C
A
A
A 18
II
I

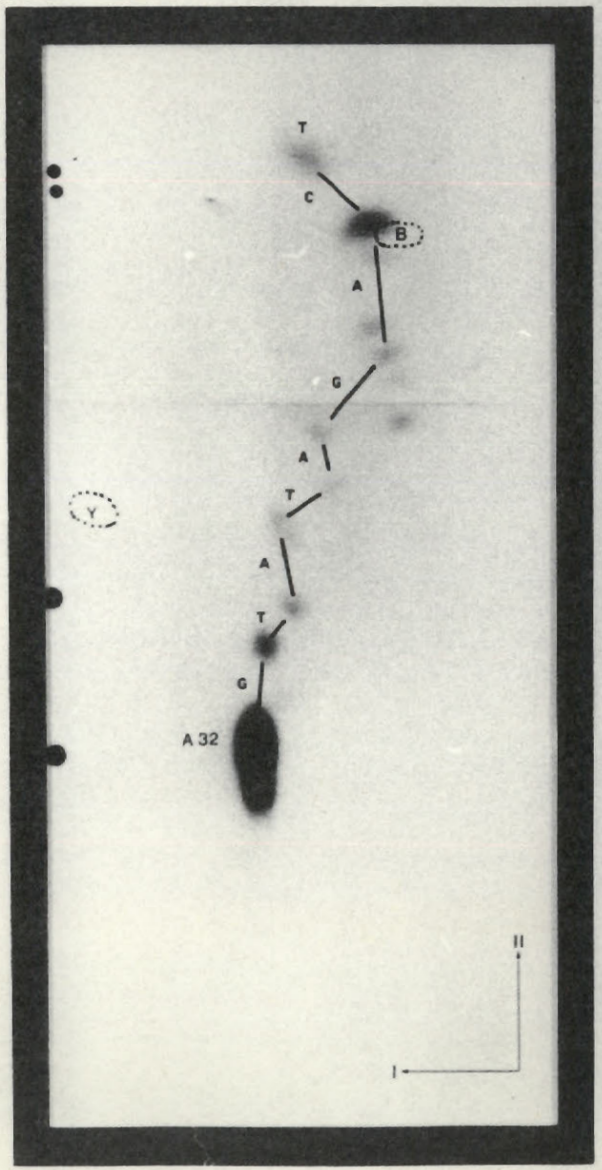

B
C
T
A
C
T
T
Y
C
C
C
C
A
A
G
A 16
II
I

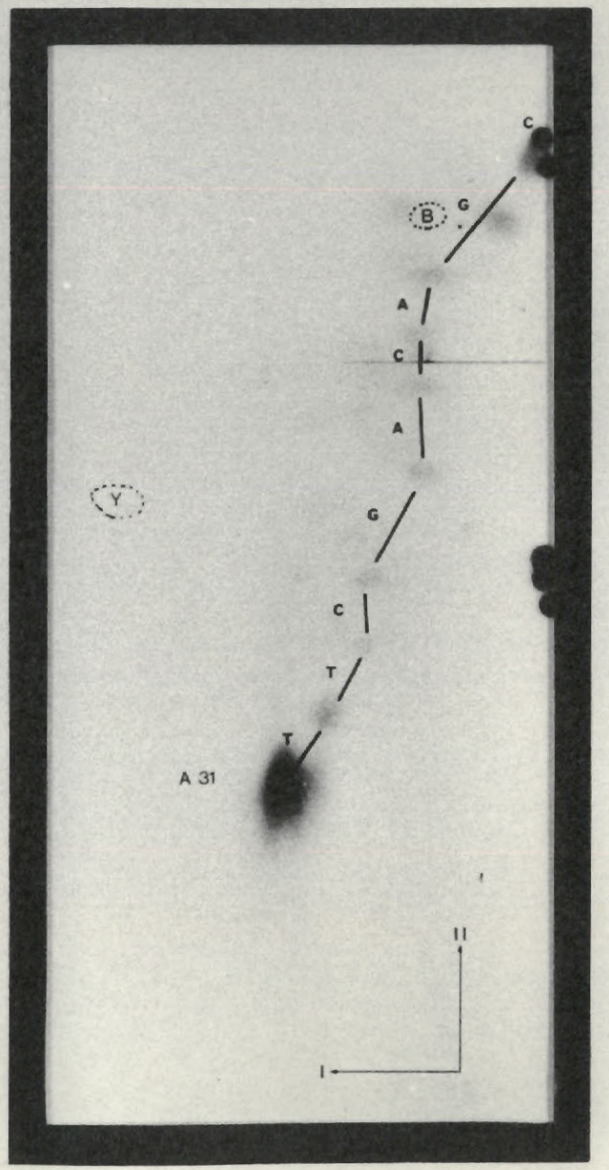

A
B
C
A
T
T
T
A
G
C
A
A 21
II
I

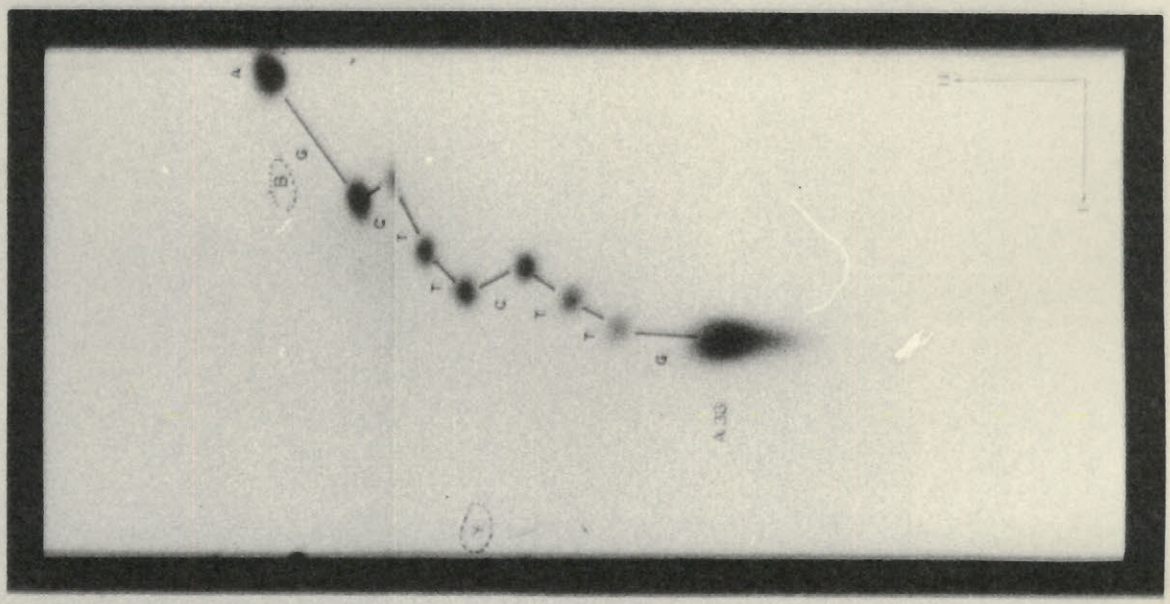

T
T
B
C
C
A
T
Y
T
C
T
T
T
A
A 20
II
I

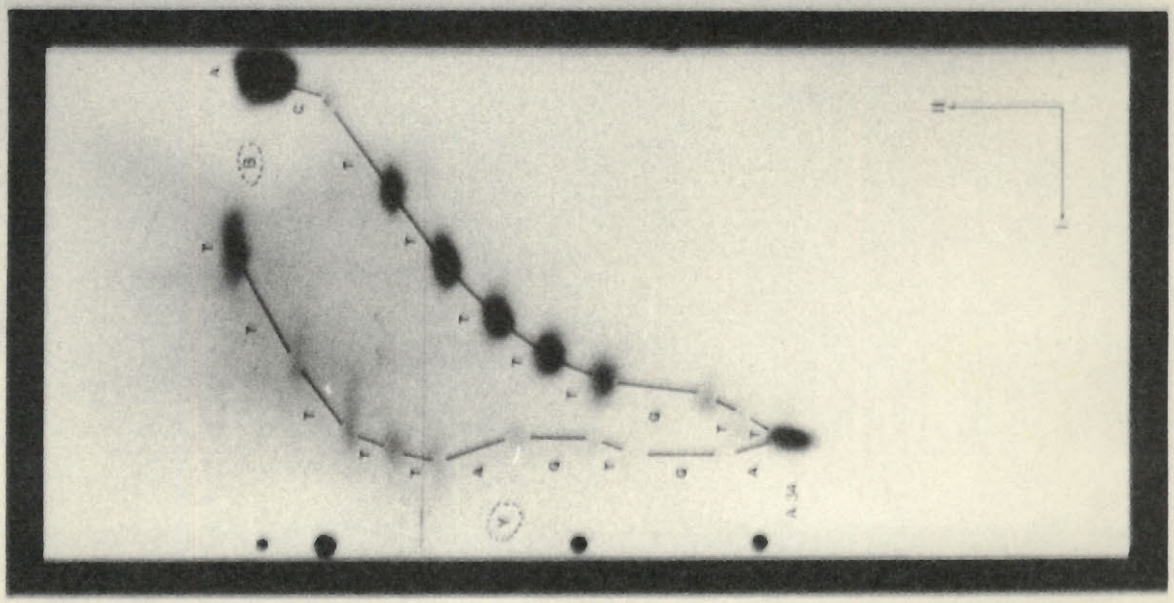

T
T
B
C
C
A
T
T
Y
C
T
T
T
T
A
G
A 26
II
I

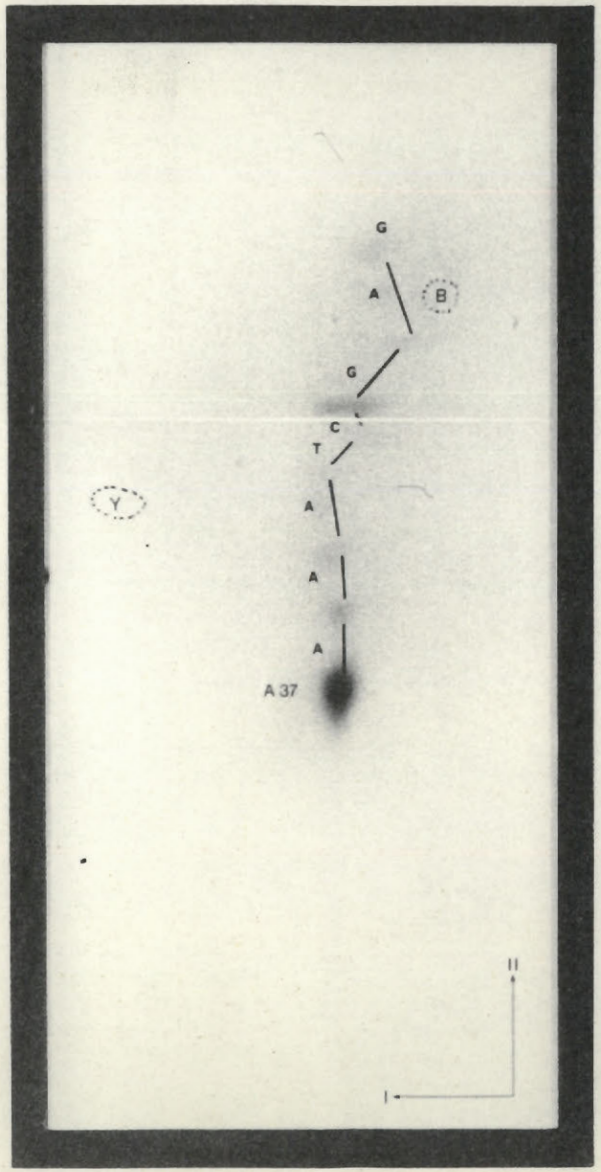

C
B
T
A
C
T
T
Y
C
C
C
A
A23
II
I

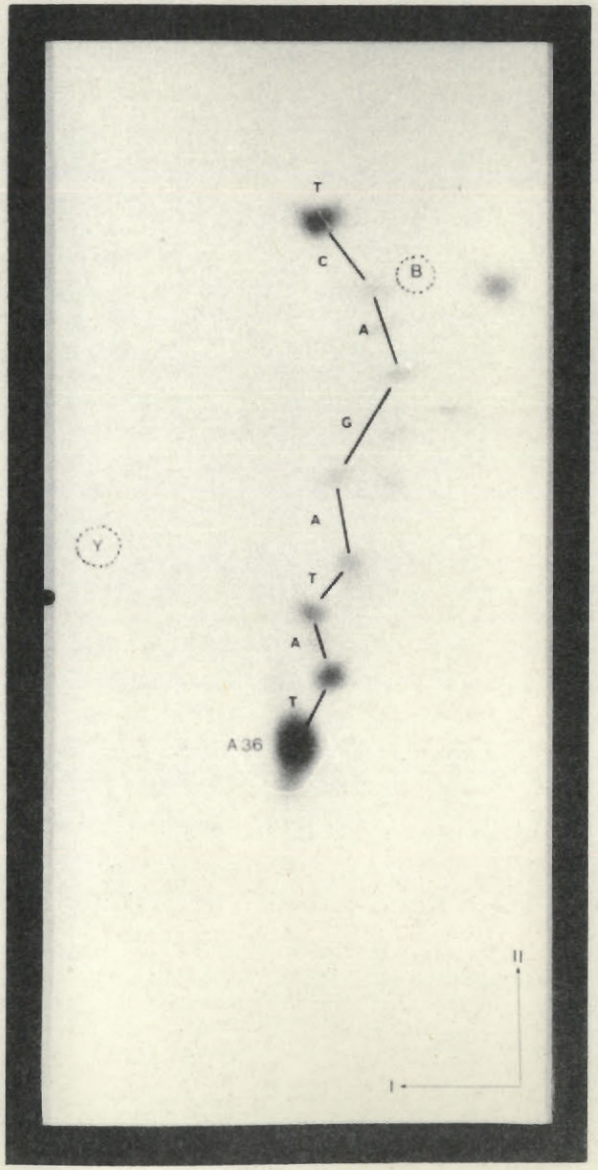

T
C
B
A
G
A
T
A
T
G
A 32
Y
II
I

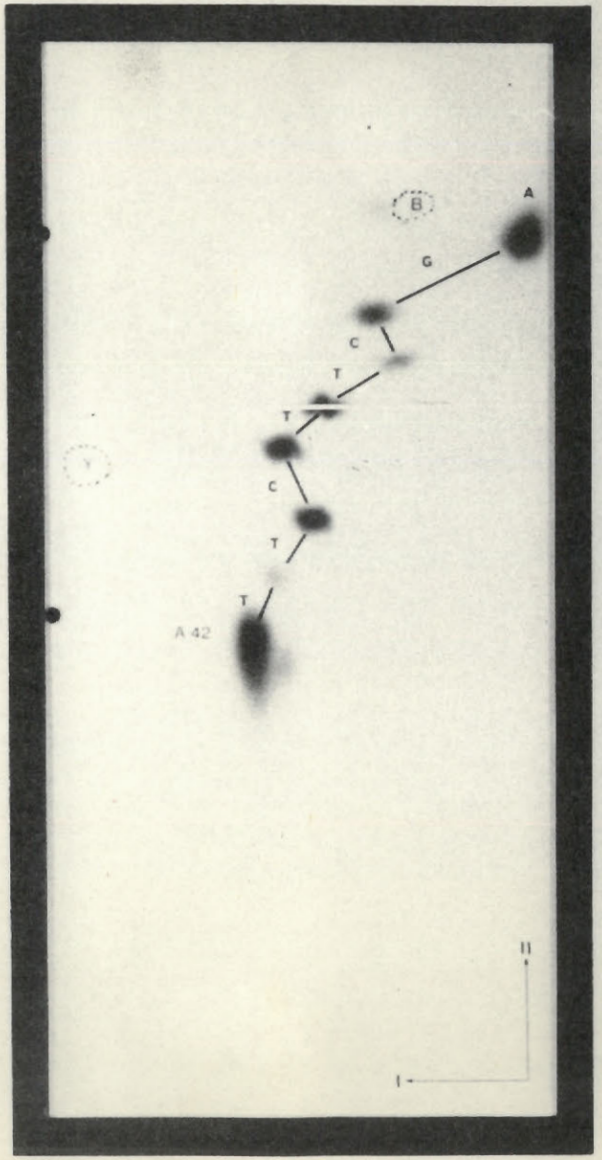

C
B
G
A
C
A
G
C
T
T
A 31
Y
II
I

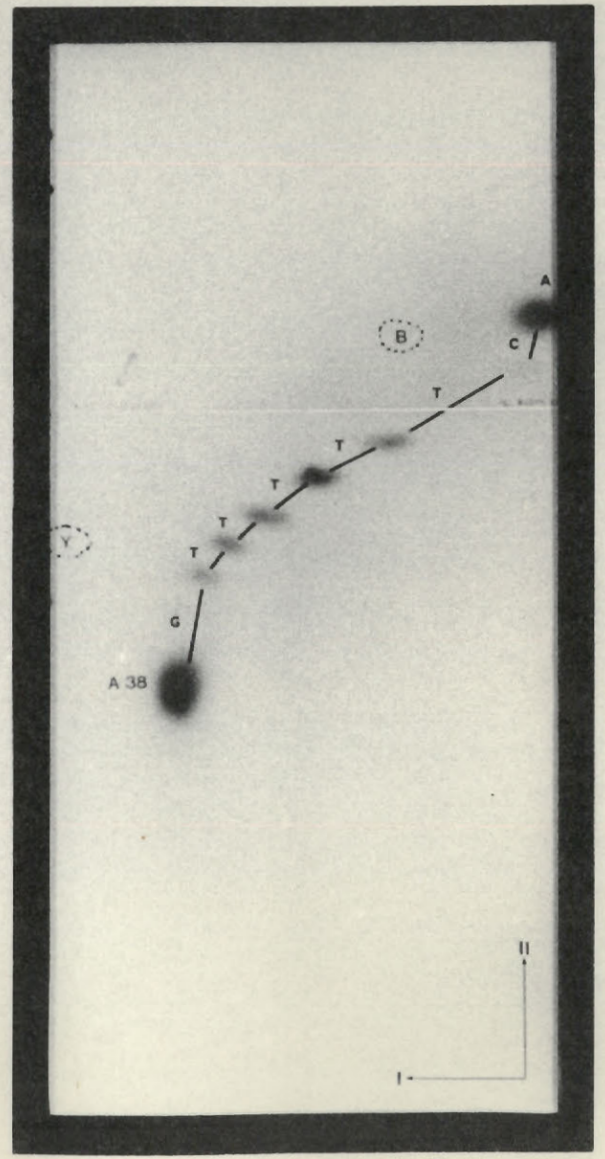


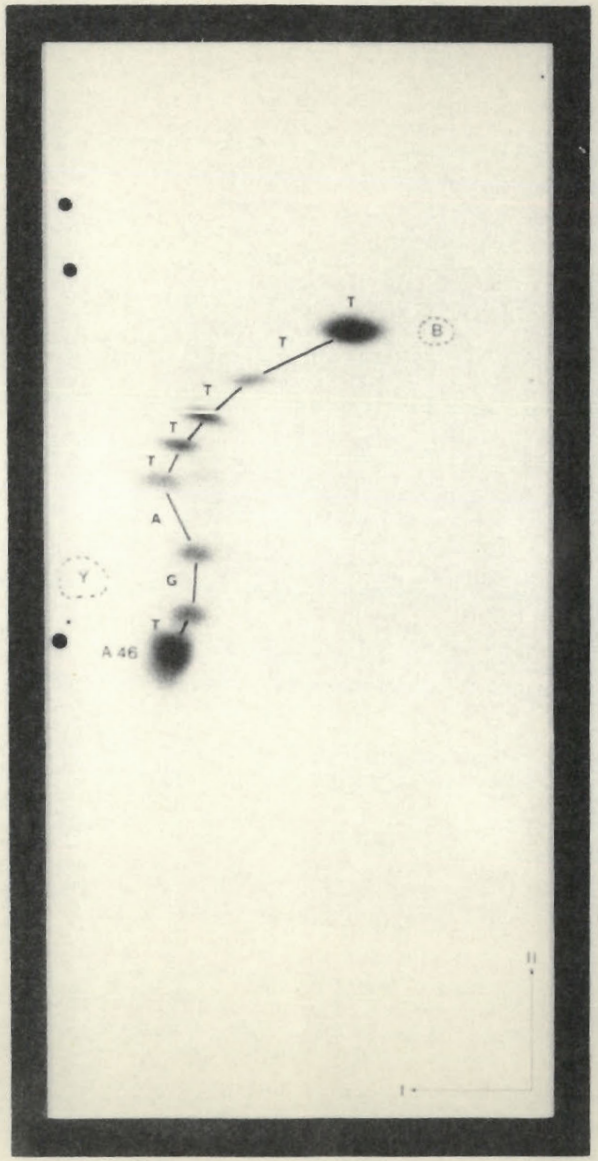


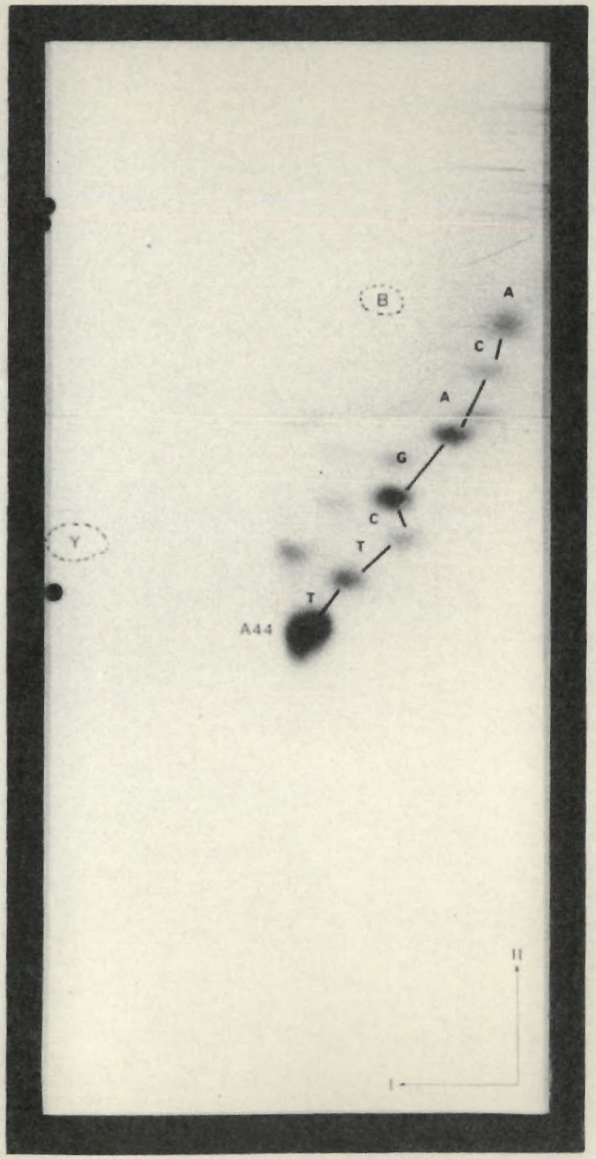

G
A
B
G
C
T
A
A
A
A 37
Y
II
I

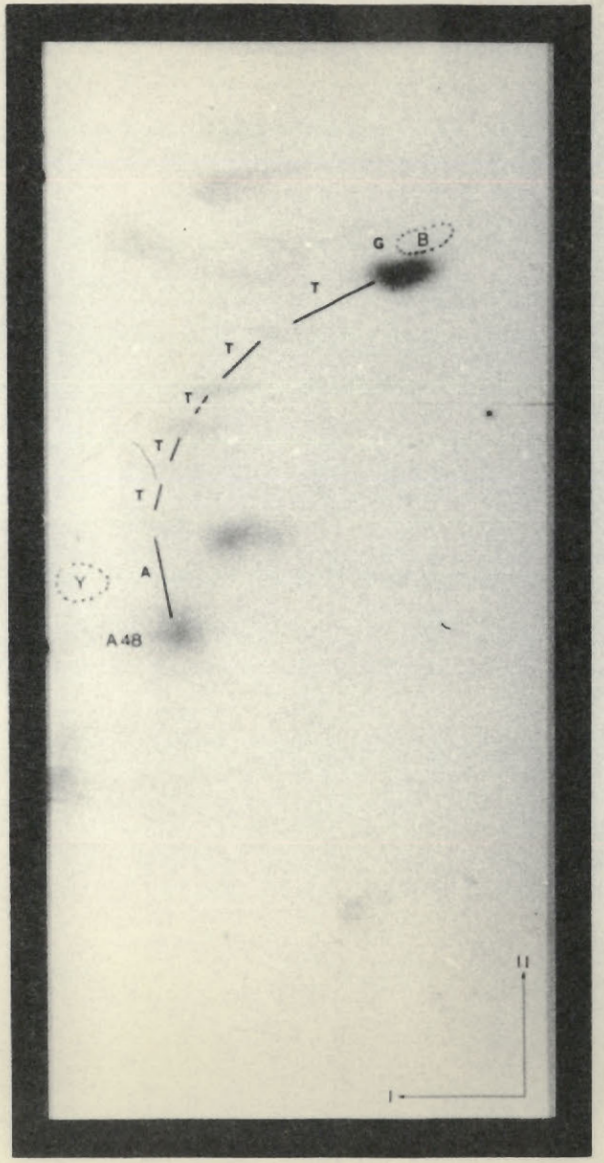

T
C
B
A
G
A
T
A
T
A 36
Y
II
I

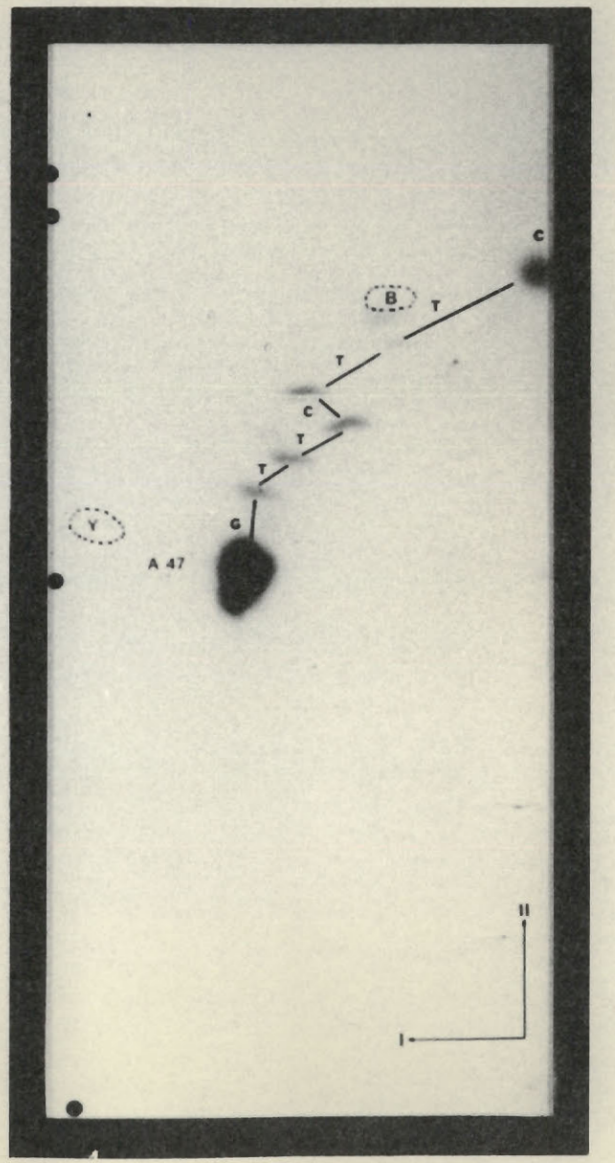

B
A
G
C
T
T
T
Y
C
T
T
A 42
II
I

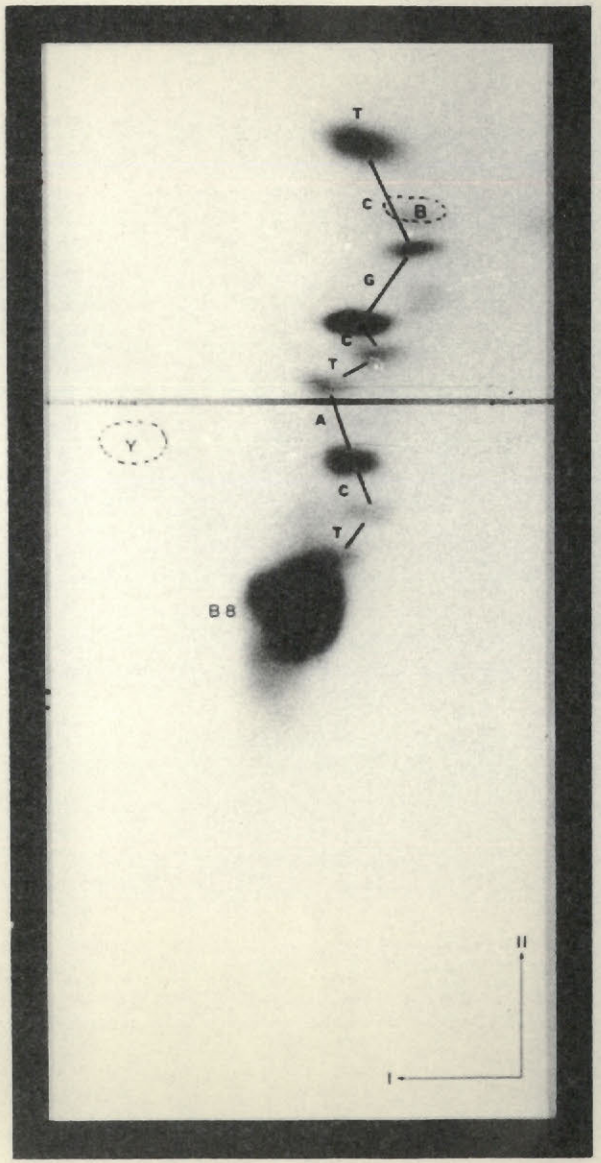

A
B
C
T
T
T
T
T
Y
T
G
A 38
II
I

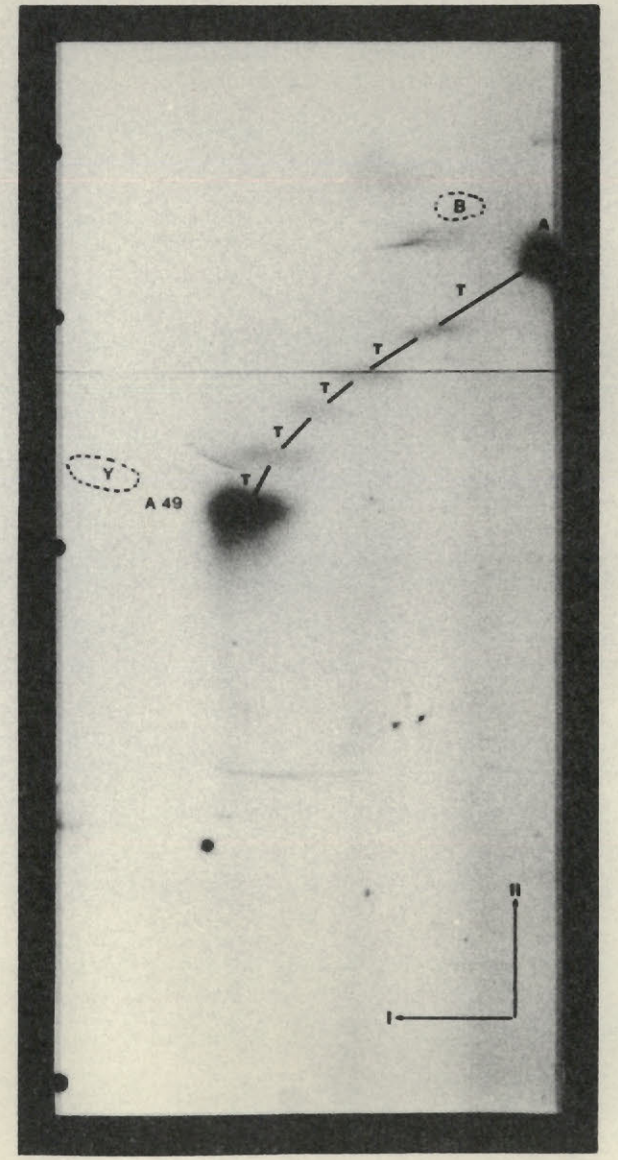

T
T
T
T
T
T
A
G
T
A 46
B
Y
II
I

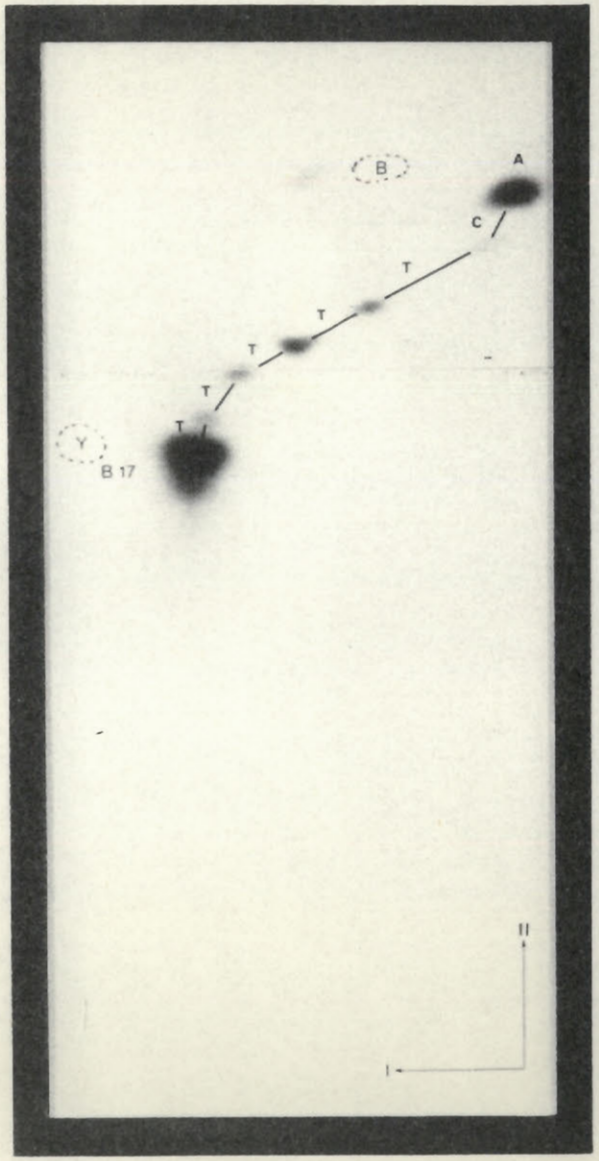

A
C
A
G
C
T
T
A44
B
Y
II
I

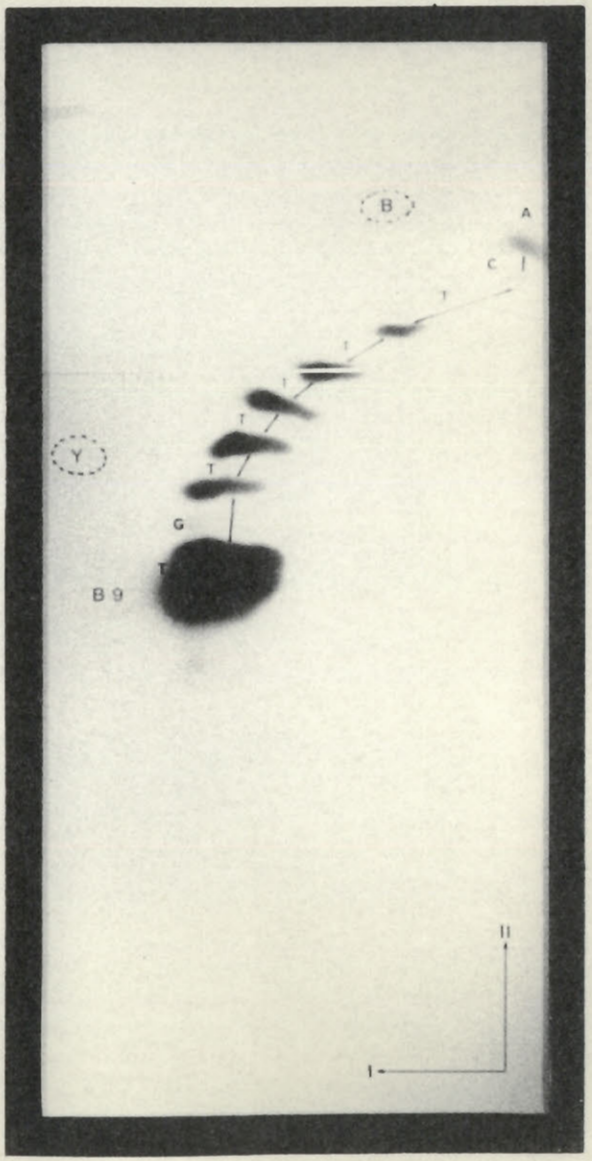

G
B
T
T
T
T
T
T
Y
A
A 48
II
I


C
B
T
T
C
T
T
Y
G
A 47
II
I

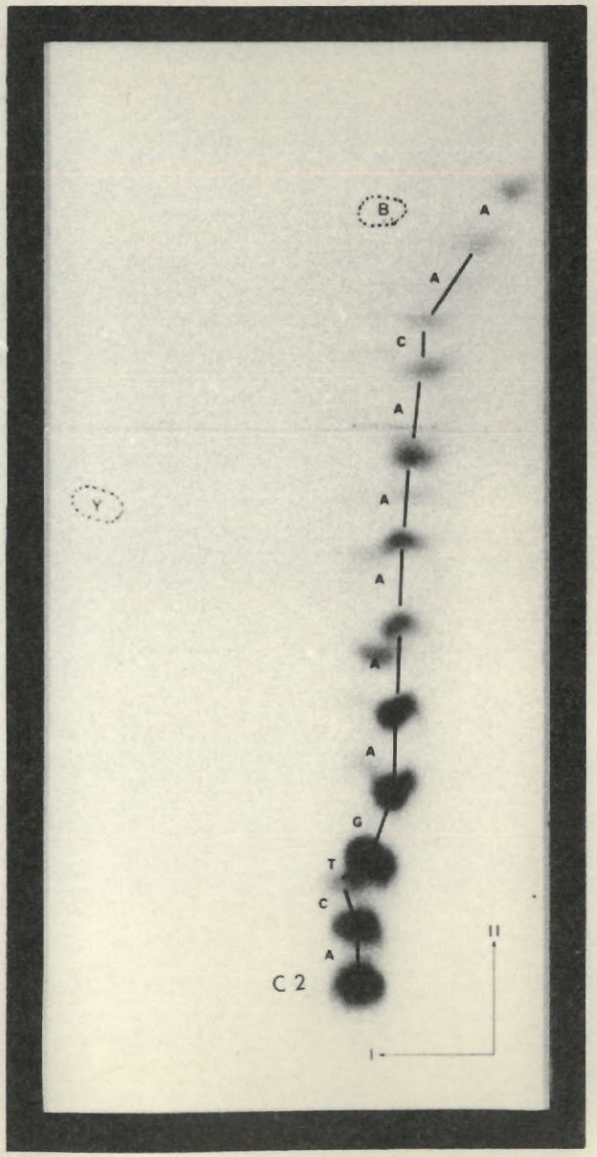

T
C
B
G
C
T
A
Y
C
T
B 8
II
I


B
A
T
T
T
T
T
Y
A 49
II
I

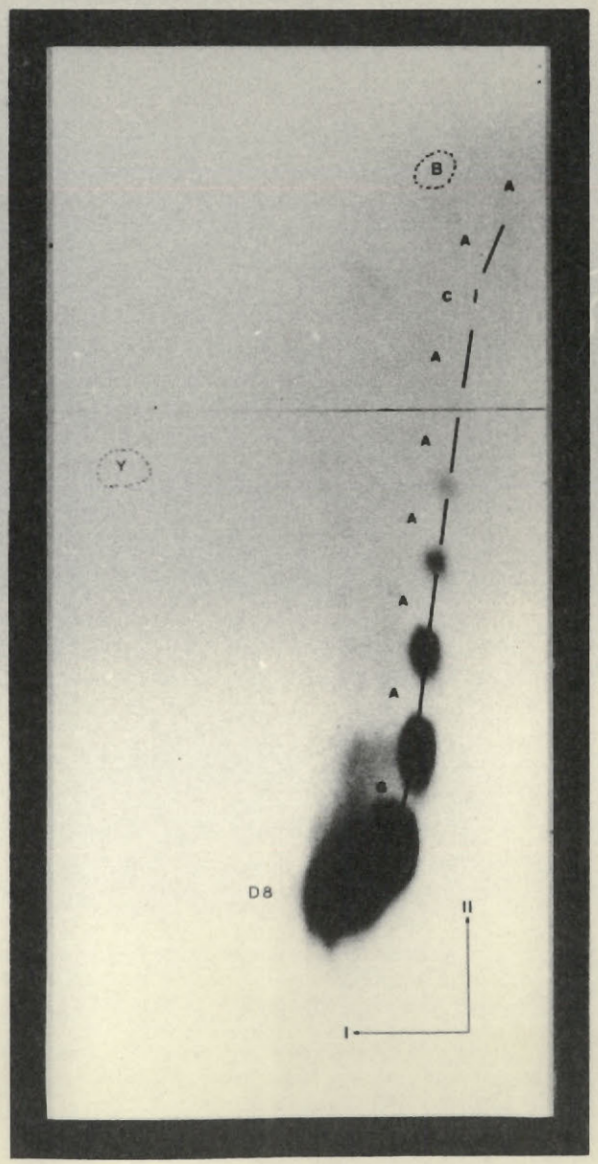

B
A
C
T
T
T
T
T
T
Y
B 17
II
I

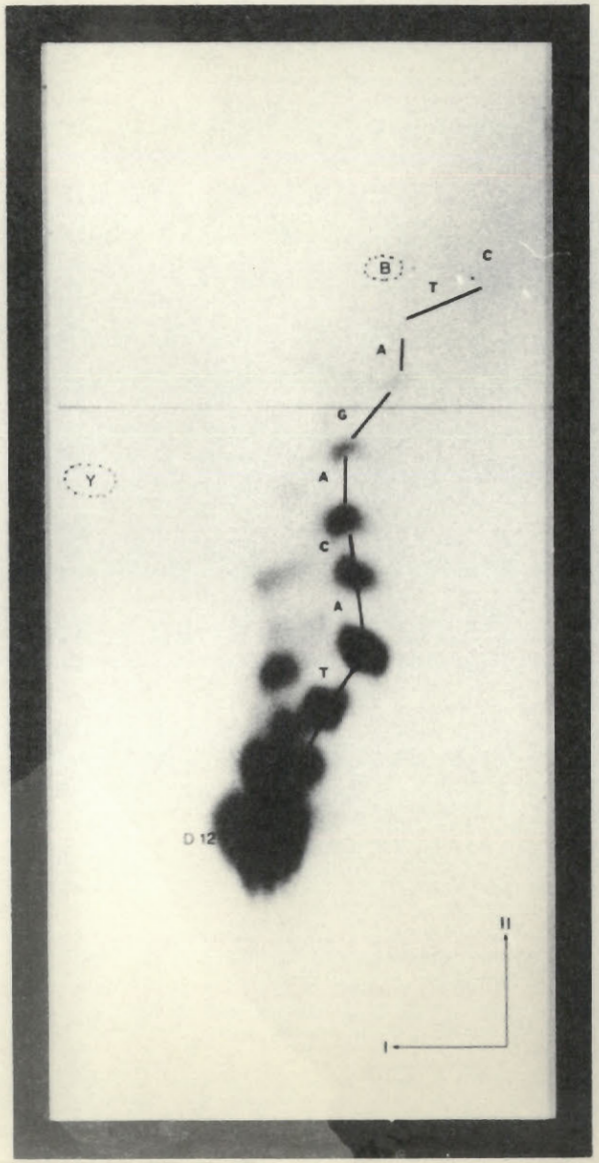

B
A
C
T
T
T
T
T
T
Y
G
T
B 9
II
I

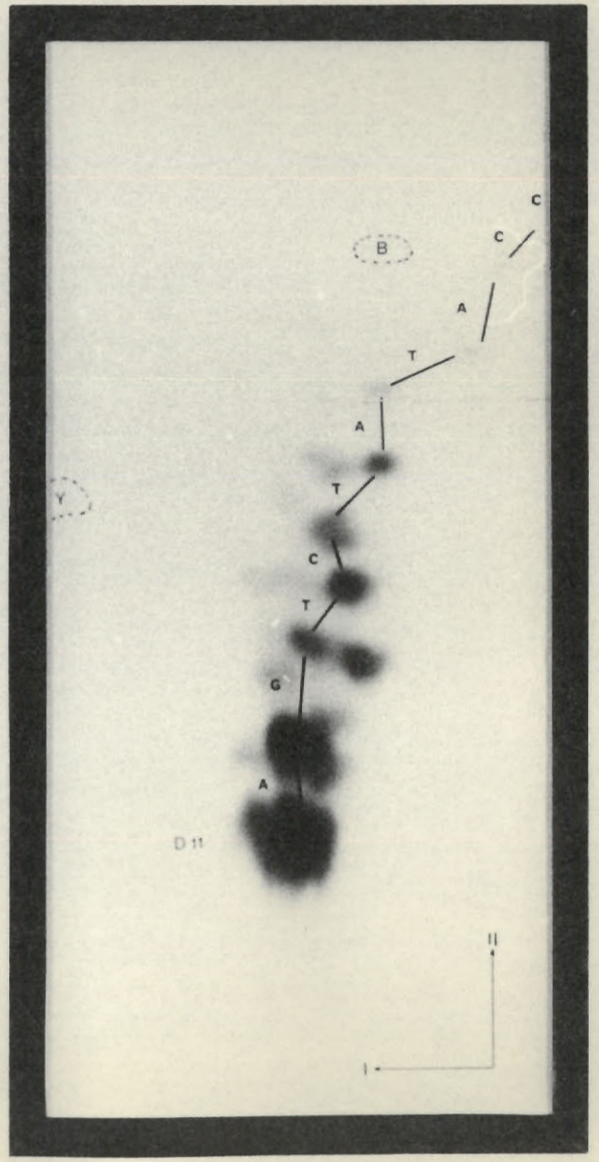

B
A
G
T
T
C
T
T
G
C 5
Y
II
I

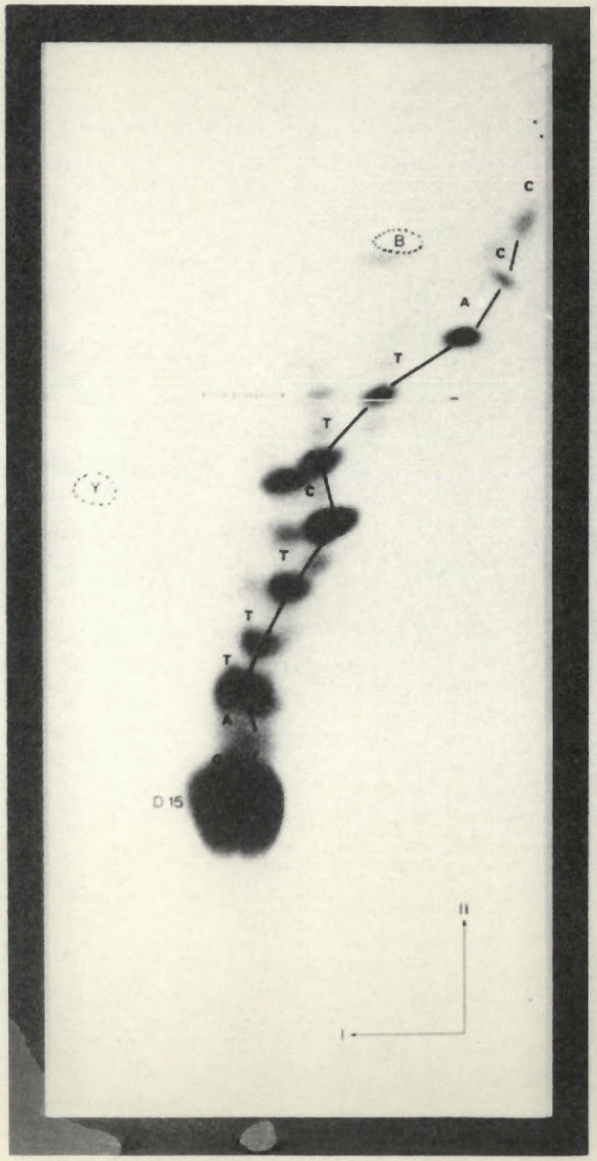

B
A
A
C
A
A
A
A
A
G
T
C
A
C 2
Y
II
I

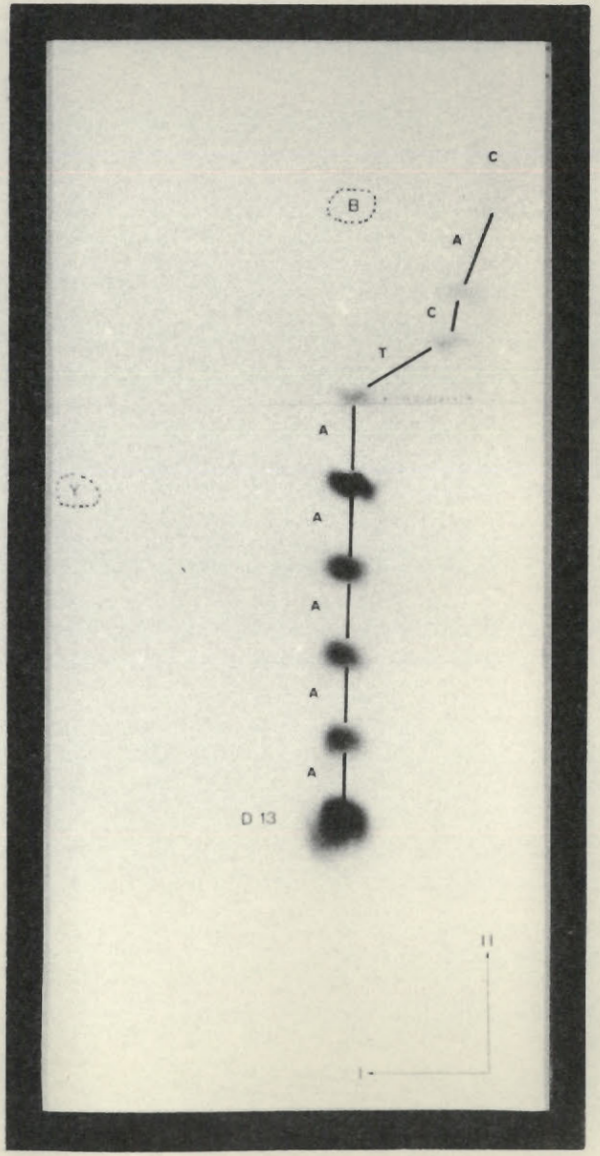

B
Y
A
T
T
T
A
G
D 9
II
I

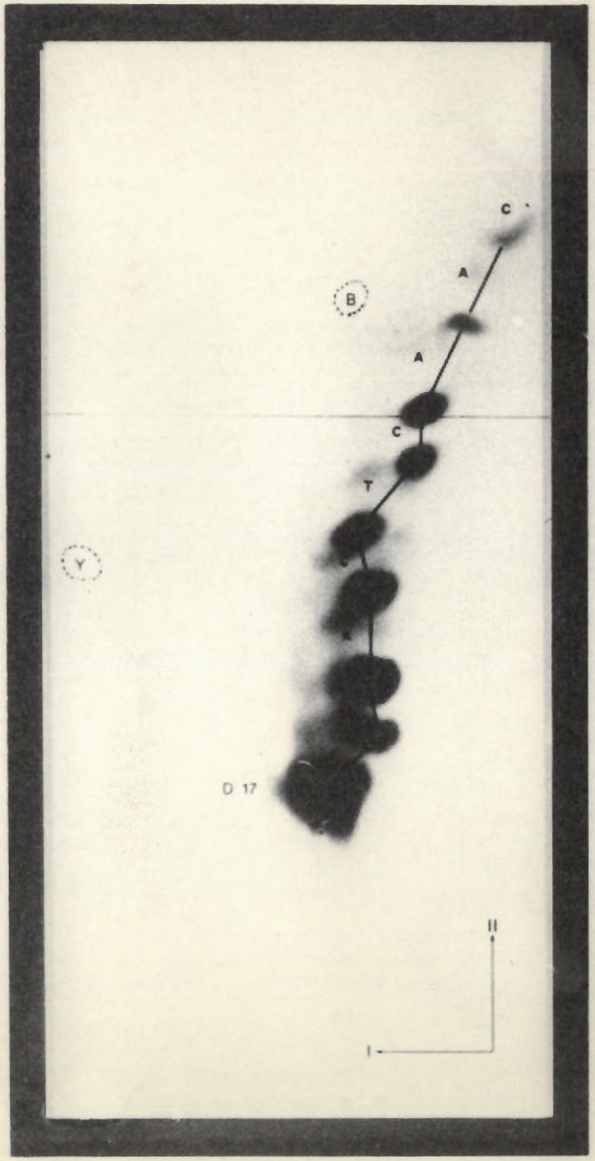

B
A
A
C
A
Y
A
A
A
A
G
D8
II
I

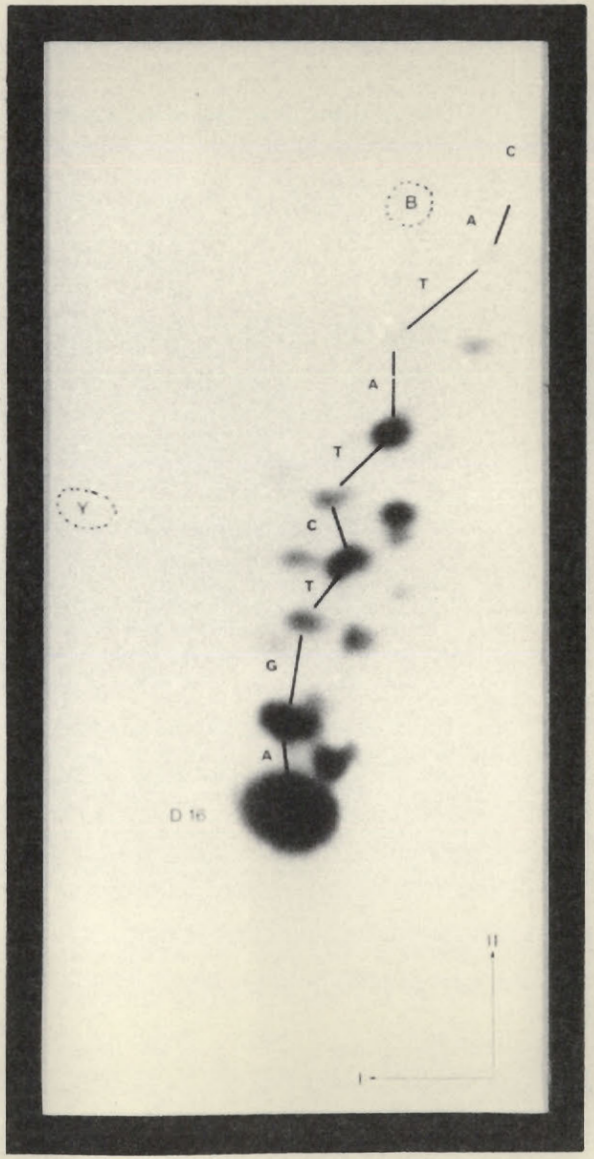

B
C
T
A
G
A
C
A
T
Y
D 12
II
I

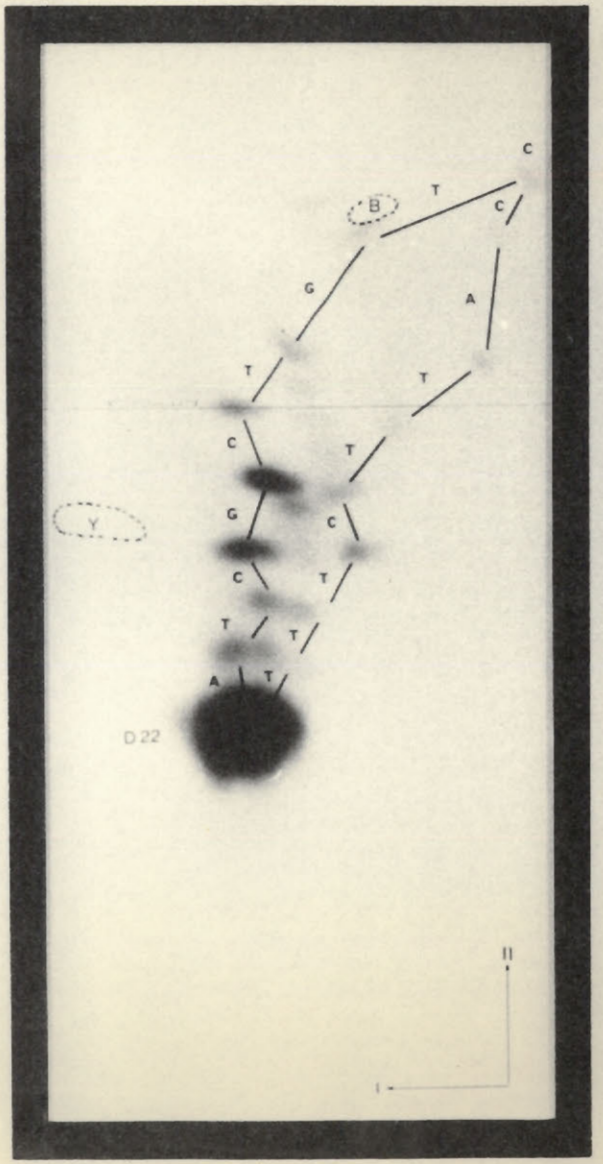

C
C
B
A
T
A
T
C
T
G
A
Y
D 11
II
I

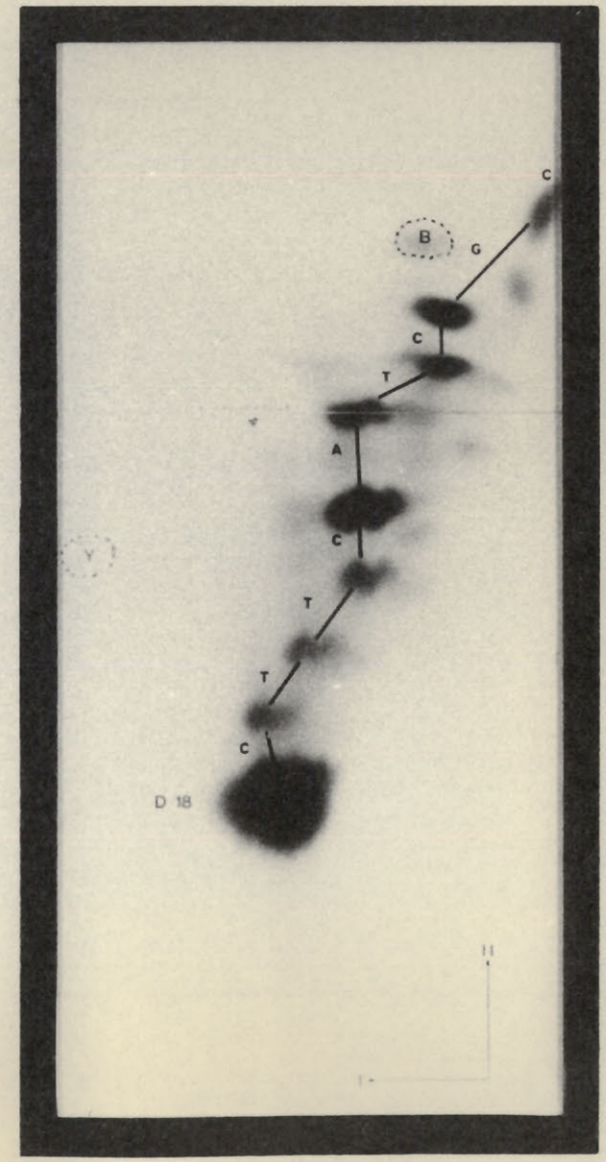

B
C
C
A
T
T
C
T
T
T
A
D 15
II
I

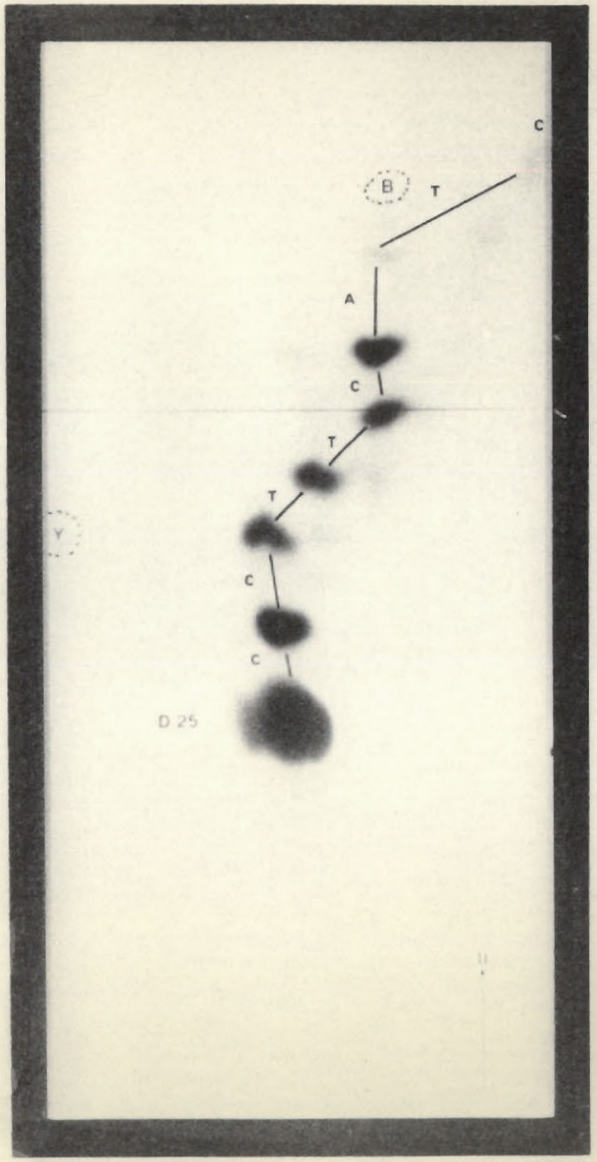

B
C
A
C
T
A
A
A
A
A
D 13
II
I

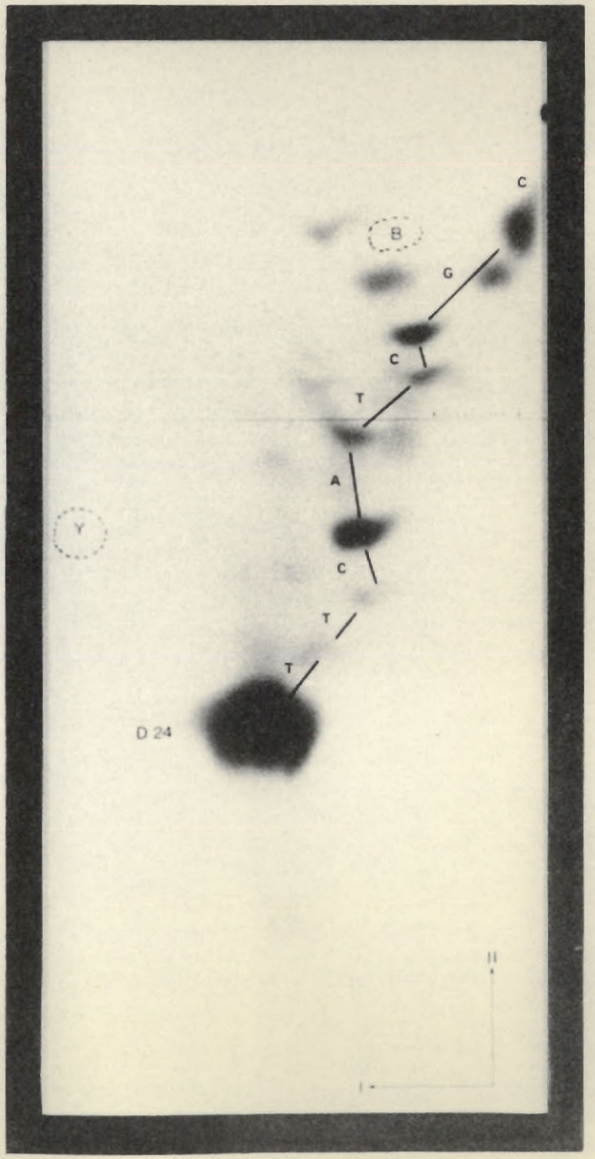

C
A
B
A
C
T
Y
D 17
II
I

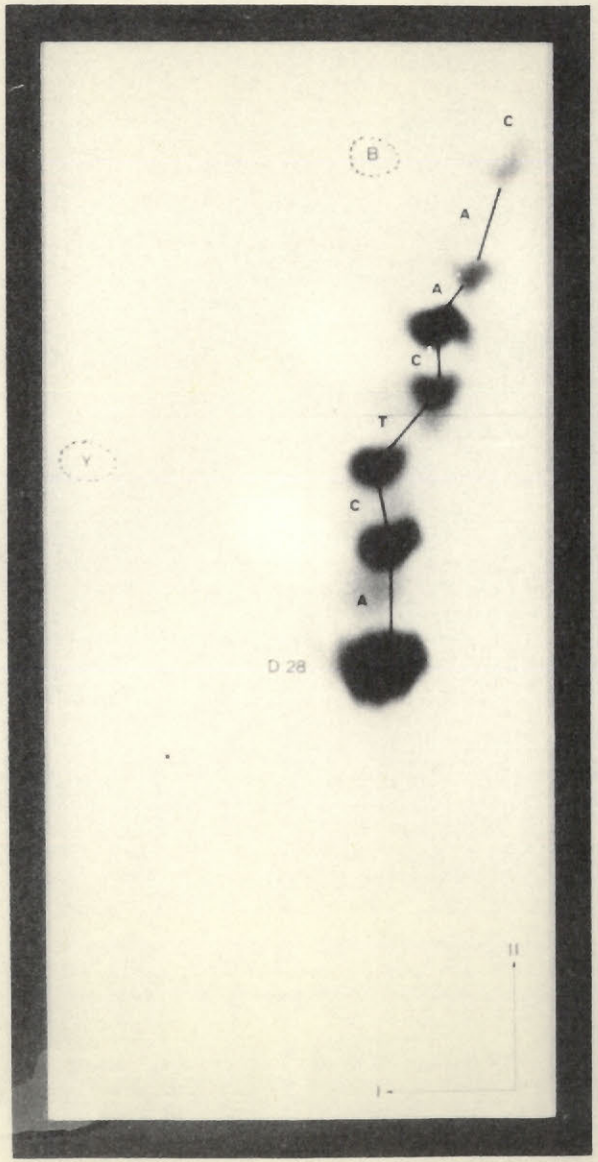

C
B
A
T
A
T
Y
C
T
G
A
D 16
II
I

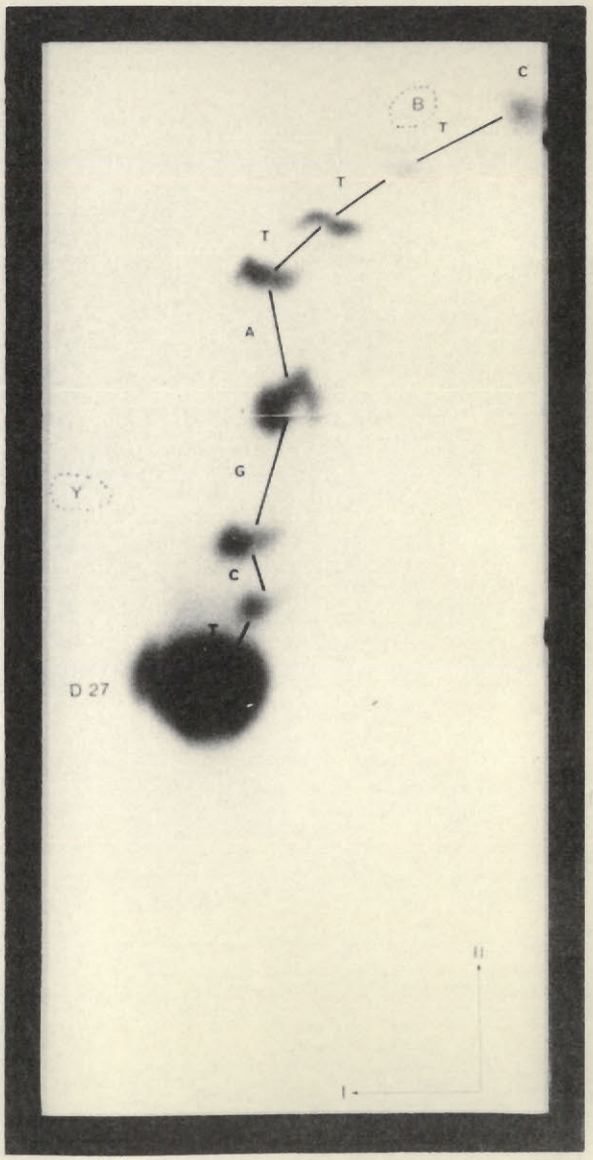

B
Y
D 22
II
I

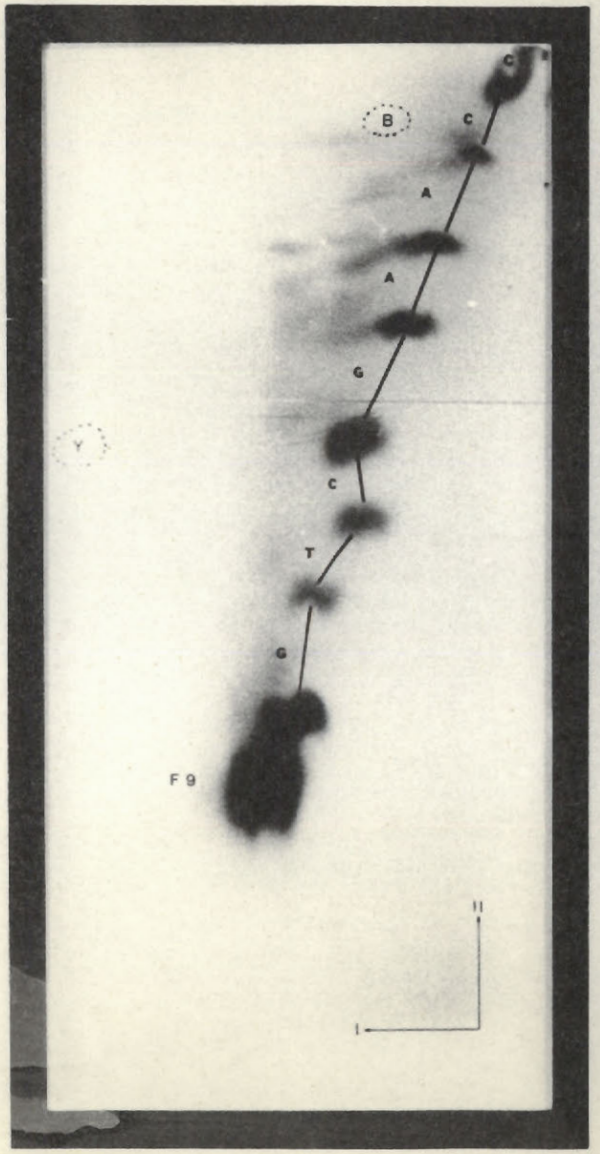

B
Y
D 18
II
I

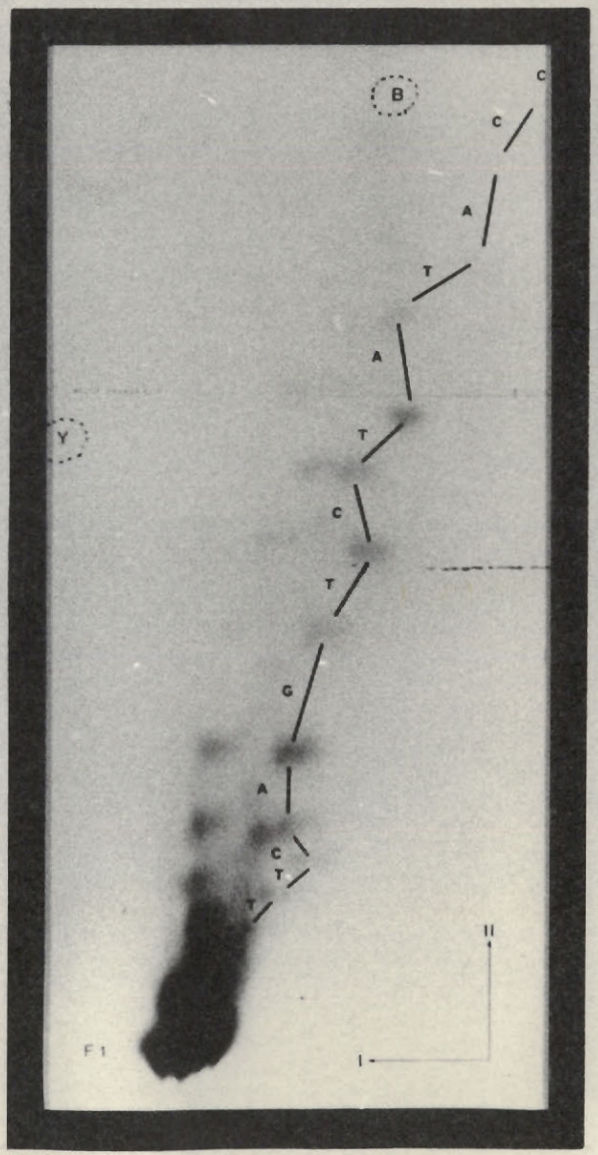

C
B
T
A
C
T
T
Y
C
C
D 25
II

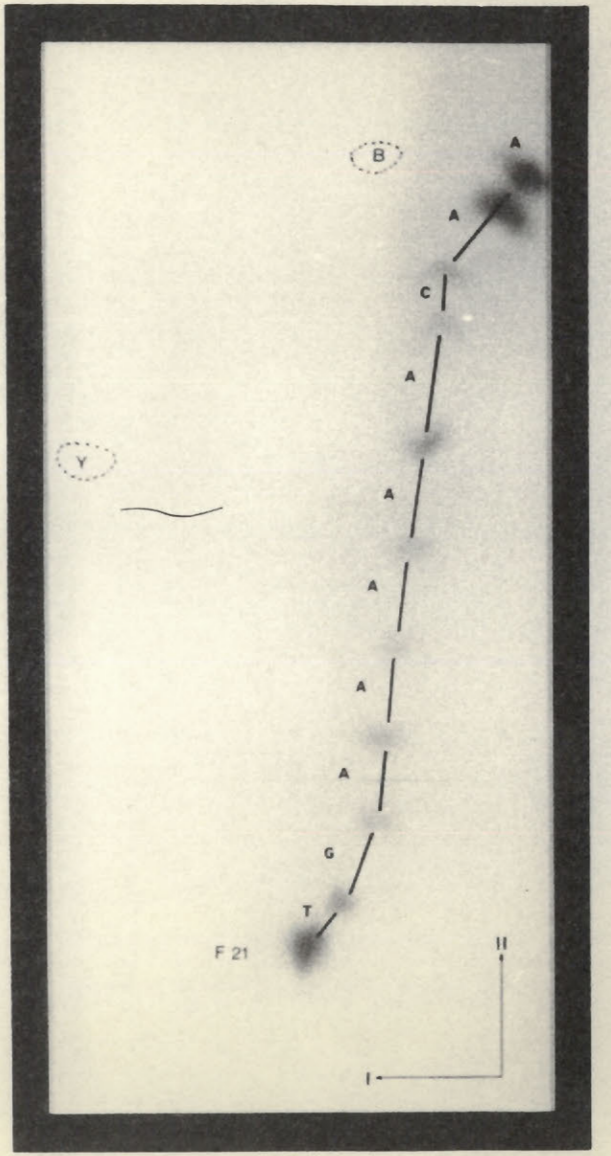

C
B
G
C
T
A
Y
C
T
T
D 24
II
I

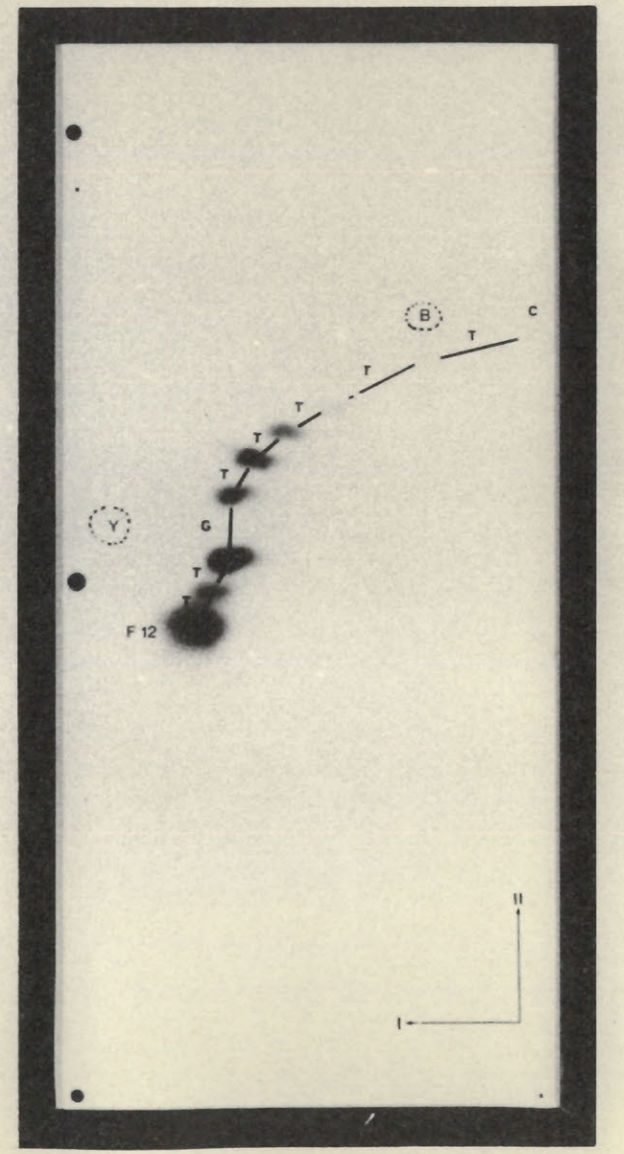

B
C
A
A
C
T
Y
C
A
D 28
II
I -

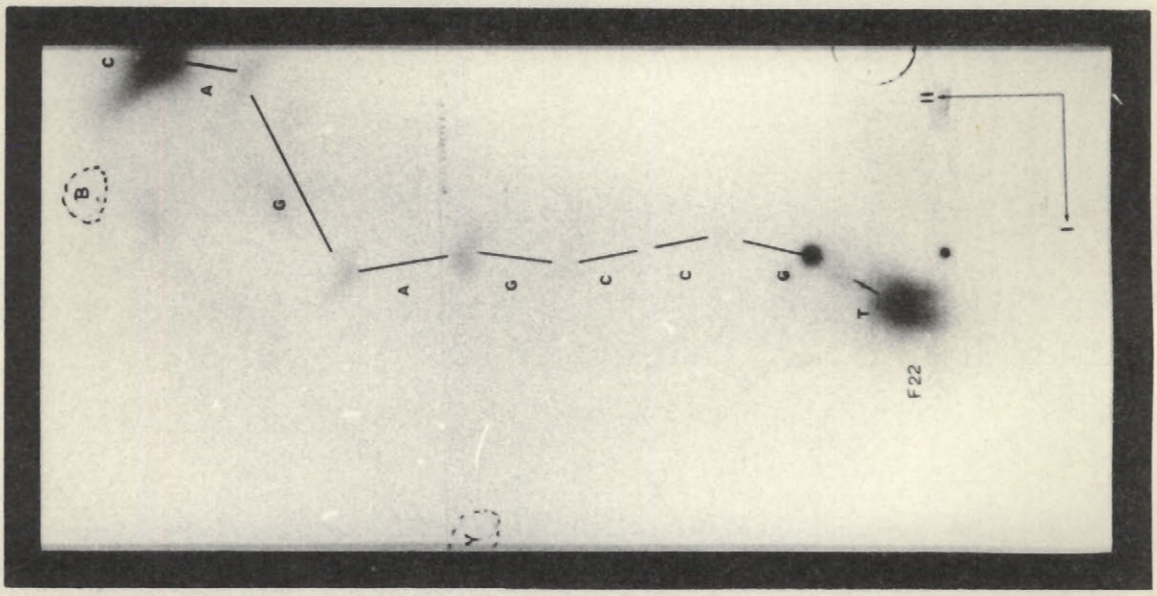

C
B
T
T
T
A
G
Y
C
T
D 27
II
I -


B
C
C
A
A
G
Y
C
T
G
F 9
II
I


B
C
C
A
T
A
Y
T
C
T
G
A
C
T
T
F 1
II
I


B
A
A
C
A
A
A
A
A
A
G
T
Y
F 21
II
I


B
C
T
T
T
T
T
T
G
T
T
Y
F 12
II
I


C
A
G
A
G
C
C
G
T
B
Y
II
I
F22