4:2.8$_b$

STUDIES TOWARD THE PRIMARY STRUCTURE OF S13 DNA

STUDIES TOWARD THE PRIMARY STRUCTURE OF BACTERIOPHAGE ϕX13 DNA

by

Allen Daniel Delaney

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

July, 1974

ABSTRACT

Spleen exonuclease and bacteriophage T4 induced polynucleotide kinase have been characterized with regard to their use in DNA sequence analysis and a model describing the action of polynucleotide kinase proposed. Interfering oligodeoxyribonucleotide-like contaminants in the polynucleotide kinase preparations were characterized and removed. A general method for the determination of the sequence of oligodeoxyribonucleotides of any base composition was developed. This involved partial spleen exonuclease degradation, 5' terminal labelling with polynucleotide kinase and [γ-^{32}P] ATP, and chain length fractionation of the labelled digest products. Several fractionation systems were compared with regard to their use in this method, in particular, and in DNA sequence analysis, in general. The sequences of long pyrimidine oligonucleotides from S13^{+} DNA were determined using this and two other methods; these sequences were compared with those from the DNA of bacteriophage ØX174, and their significance regarding the evolutionary divergence of these closely related bacteriophages examined.

ABREGE

L'exonucléase de la rate et la polynucléotide kinase induite par le bacteriophage T4 ont été caractérisées a cause de leur utilité dans l'analyse séquentielle de l'ADN et un modèle décrivant l'action de la polynucléotide kinase est proposé.

Des contaminants ressemblant a des oligodéoxyribonucléotides et interférant dans la préparation de la polynucléotide kinase ont été isolés et caractérisés. Une méthode générale est developée pour la détermination de la séquence des oligodéoxyribonucléotides composés de n'importe quelle base. Celle-ci consiste en la dégradation partielle de l'exonucléase de la rate, le marquage du bout 5' de la chaine a l'aide de la polynucléotide kinase et de ATP [γ -^{32}P], et du fractionnement des produits marqués et digerés. Plusieurs séquences de fractionnement ont été comparés en raison de leur utilité dans cette methode en particulier, et dans l'analyse séquentielle de l'ADN en général.

Les séquences de longs oligonucléotides pyrimidine de l'ADN du bactériophage S13^{+} ont été déterminées en utilisant cette methode de même que deux autres ; ces séquences ont été comparées à celles de l'ADN du bactériophage ØX174 et leur importance en regard de la divergence d'évolution de ces bactériophages étroitement reliés y est examinée.

To Tricia

ACKNOWLEDGEMENTS

I wish to thank Dr. John H. Spencer for his encouragement, direction, and friendship during the last five years.

The participation of Dr. Radim Cerny and Dr. E. Cerna in the experiments described in Section 3.1 is acknowledged with thanks.

I am indebted to Dr. B. Harbers for her participation in the experiments involving sequence analysis of uniformly labelled oligonucleotides and the analyses of 5' terminally labelled oligonucleotides.

I am very grateful to Dr. Walter Mushynski and to Dr. Enrique Junowicz, who provided much of my initial training and inspiration.

I was happy to work with colleagues Dr. Jan Janda, Drs. Frank Grosveld, Dr. Klaus Harbers, Ms. Lynn Boshkov, Dr. Vivian Millington, Dr. William Thornburg, Mr. John Kaptein, and Dr. Dean Hewish, who provided aid, criticism, and enthusiasm for my work over the past few years.

I wish to thank Mr. Artin Kabassakalian for his skillful technical assistance. I am grateful to a Dutchman, Drs. Frank Grosveld, for aid in English phraseology.

I thank Ms. Lynn Boshkov and Ms. Tricia Hunt for proof-reading the thesis. Photography was performed by Mr. Levon Guluzian.

I would especially like to thank Miss Linda Pallett for her care and patience in typing this thesis.

I wish to express my appreciation to the Medical Research Council of Canada for the award of a studentship for the period 1969 to 1973, and to the Medical Research Council of Quebec for the award of a studentship for the period 1973 to 1974.

This work was carried out with the support of the Medical Research Council of Canada through their Grant MT 1453 to Dr. John H. Spencer.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Unless otherwise stated, all abbreviations follow the guidelines in Biochem. J. (1973) 131, 1 - 20.

T	thymidine
G	deoxyguanosine
C	deoxycytidine
A	deoxyadenosine
rU	uridine
rG	guanosine
rC	cytidine
rA	adenosine
X	unspecified nucleoside
C_xT_z	Pyrimidine oligonucleotide containing x deoxycytidine moieties, and z deoxythymidine moieties. The number of phosphate moieties in the oligonucleotide will not be included except where it is not apparent from the text.
A_{274}	absorbance at 274 nm
p.f.u.	plaque forming units
TLC	thin layer chromatography
CLF	chain length fractionation
SVD	snake venom exonuclease digestion
SPD	spleen exonuclease digestion

PNK	polynucleotide kinase
PME	phosphomonoesterase
PEI	polyethylenimine cellulose
TEAB	triethylammonium bicarbonate

CHAPTER 1

INTRODUCTION

1.1 Analysis of the Primary Structure of DNA

Until recently, the exact analysis of the polymeric properties of DNA molecules has been limited to base composition analyses (39, 40) and various chemical studies (41-47, 71). The most important property of DNA, the sequence of nucleotides, could not be determined with these techniques.

In contrast, the sequence of RNA molecules, since 1965 (48), has progressed to the stage where complete genomes containing 3300 nucleotides are being examined. During this same time period DNA sequence analysis reached only the trinucleotide level (51, 133). We can best discuss the difficulties of DNA sequence analysis by comparing it with RNA sequence analysis.

There are no deoxyribonucleases with specificity towards DNA similar to that shown by RNAase T1 and by pancreatic RNAase towards RNA. Pancreatic RNAase cleaves an RNA polymer at the 3' phosphate of a pyrimidine, while RNAase T1 cleaves at the 3' phosphate of a guanosine residue.

Recent studies on DNAase I (22, 52) and DNAase II (53) from hog spleen have shown limited specificity in that certain nucleotide bonds are resistant to hydrolysis. This type of degradation, however, is not useful for specific fragmentation of DNA.

In the search for a specific degradation method, many investigators have examined chemical hydrolysis of DNA molecules. Terminal stepwise oxidation of polydeoxynucleotides (44, 45, 46, 47) and depyrimidination of DNA (41, 42, 43, 139)

would be very useful if not for troublesome side reactions. Depurination of DNA, as developed by Burton and Petersen (71), has been the only truly specific chemical degradation method. The hydrolysis of DNA with formic acid in the presence of diphenylamine is analagous to the specific degradation of DNA with an endonuclease. There are, however, two limitations to the depurination method: there is no possibility of a partial digest and there can be no overlap of the pyrimidine clusters obtained.

The advent of specific DNAases such as T4 endonuclease IV (38), and restriction enzymes (54, 55), has made the DNA sequence problem more feasible, but these enzymes are not completely analagous to the RNAases mentioned above. The restriction endonucleases and T4 endonuclease IV have much more stringent specificity requirements than pancreatic RNAase or RNAase T1.

Restriction enzymes recognize and cleave DNA sequences 4 to 6 nucleotide base pairs in length, producing from a hypothetical random DNA molecule a population of fragments distributed in length around a mean of 300 to 4000 base pairs, depending on the endonuclease.

T4 endonuclease IV recognizes and cleaves the sequence --pTpCp-- in single stranded DNA, yielding from a random DNA molecule a population of fragments around a mean of 15 bases in length (67, 141, 154).

A series of degradations with restriction enzymes and finally with T4 endonuclease IV will degrade DNA into discrete fragments of relatively small size. However, the complexity of the oligonucleotide products requires a large number of separation steps and then sequence analyses of the purified homogeneous oligonucleotides to obtain complete sequence data.

RNAase T1, in contrast, will degrade a random RNA molecule to a population of fragments around a mean of four bases in length, while the complete digest products produced by pancreatic RNAase are distributed around a mean length of 2 nucleotides.

A combination of digestions with the two ribonucleases will usually yield products small enough to be sequenced directly by exonuclease partial digestions and base composition determinations. The development of fractionation procedures used in RNA sequence analysis has been aided by the fact that the enzymes used for degradation result in small fragments with specific termini. The components of a complete pancreatic RNAase digest have a pyrimidine at the 3' terminus and purines only in the rest of the molecule. Similarly, components of a complete T1 RNAase digest have a guanylate residue at the 3' terminus and only 3 nucleotides in the rest of the sequence. Some of the separation systems developed for these small oligonucleotides are specific enough to separate sequence isomers by electrophoretic or chromatographic mobility.

Except for some bacteriophage DNA's and eukaryotic DNA's from cells grown in tissue culture it is very difficult to obtain a DNA sample by in vivo labelling techniques of the high specific activity required for sequence analysis. Recent developments in in vitro labelling techniques have alleviated this problem to a considerable extent.

A qualitative difference is that the majority of DNA molecules are double stranded, with a uniform secondary structure. This reduces the possibility of specifically fragmenting DNA using partial digestion with non-specific nucleases which are hindered by secondary structure. Procedures like this are used routinely in RNA sequence studies (49, 125) and in one case have been successful in an analysis of a single stranded DNA (60).

1.2 Approaches to DNA Sequence Analysis

There are a number of approaches to DNA sequence analysis which are technologically feasible today. Which one is applicable to a given problem will depend largely on the molecule being examined. I have divided these approaches

into two categories: the direct approach consists of the fragmentation of large polymers into short pieces, analysing the sequence of each of these, and ordering them to give the total sequence; the indirect approaches are those where a biological or physical property is used in order to examine a specific small portion of the DNA molecule.

1.2.1 The Direct Approach: Fragmentation and Analysis

1.2.1.1 The Selection of a DNA Substrate

Before we can discuss a method for sequence analysis, we must select a substrate DNA. A human cell is estimated to contain 4×10^{12} daltons (134) of deoxyribonucleic acid. It will probably never be practical to investigate this genome directly. The chromosome of a bacterium such as E. coli is smaller by several orders of magnitude, with a molecular weight of 3×10^{9} daltons (135). It will be many years, however, before even bacterial DNA will be amenable to the direct approach to sequence analysis.

Bacteriophage genomes are the most popular targets for sequence analysis because of their relatively small size. Also, bacteriophage DNA can generally be prepared with a reasonably high specific radioactivity. Much is known of the genetics and organization of bacteriophages λ and T7 (62, 63), thus a considerable amount of sequence work is being done on their genomes (57, 59, 96, 106) even though they are 50,000 and 40,000 base pairs long, respectively.

In the case of the smallest coliphages, fd, f1, S13 and ØX174 with a genome only 5500 nucleotides in length, direct sequence analysis can be and is being attempted (7, 64, 65, 67, 111, 117) with the expectation of a complete genome

sequence within several years. DNA isolated from bacteriophage S13$^+$ was the source of oligonucleotides investigated in this thesis.

1.2.1.2 Fragmentation of DNA with Restriction Enzymes

The sequence of a molecule of DNA even as short as 5500 nucleotides cannot be deduced in one step ; the polymer must be broken into a number of manageable fragments which can be sequenced and ordered within the parent molecule. This thesis presents sequence methods which allow us to define a manageable fragment as one containing less than fifty nucleotides. The availability of restriction enzymes and T4 endonuclease IV makes it possible to degrade DNA specifically down to this size.

Hutchison and his colleagues (55, 66, 126) have used genetic and hybridization techniques to determine the order of the fragments produced by several restriction enzymes from the genome of bacteriophage ØX174. In future sequence studies, methods similar to these will be useful for the arrangement of DNA fragments and will eliminate the necessity for overlapping sequence data.

It is very unlikely that a single restriction enzyme would degrade a complete DNA molecule into fragments small enough for endonuclease IV degradation (Section 1.2.1.3). For example, ØX174 and S13 replicative form DNA are cleaved into fourteen and thirteen fragments (55, 141, 142) respectively, by the restriction endonuclease isolated from Haemophilus influenzae. In both cases the shortest fragments are 62 and 80 nucleotides long, while the longest fragments contain more than 1000 nucleotides. It is apparent that the longer fragments must be further digested with one or more additional restriction enzymes until the whole genome is

reduced to a set of ordered fragments which can be denatured and reduced to manageable size using T4 endonuclease IV.

It is theoretically possible to degrade DNA down to oligonucleotides less than 50 bases long using only restriction enzymes, but their level of specificity makes this a practical impossibility. The number of enzymes required would be prohibitive.

1.2.1.3 Fragmentation of DNA with T4 Endonuclease IV

In the course of the fragmentation of a large DNA molecule by any of a number of methods, it is likely that single stranded fragments will be produced which are too long for detailed sequence analysis by partial exonuclease digestion methods. Such polydeoxyribonucleotides can probably be degraded with T4 endonuclease IV, yielding a group of oligonucleotides which can each be sequenced. The upper size limit of the DNA fragments which can be examined in this manner is a function of the resolution possible with available oligonucleotide fractionation techniques.

As I have already mentioned, T4 endonuclease IV digests of a random DNA molecule will contain a population of oligonucleotides varying in size about a mean of 14.5 nucleotides. The probability of occurrence of different oligonucleotide products in such a digest is represented in Figure 1.1. The numbers above the absolute probability curve indicate the length of DNA molecule which, on complete digestion with a TC-specific T4 endonuclease IV, would probably yield one oligonucleotide of the chain length indicated by the arrow. Shorter oligonucleotides would probably occur more than once in the digest, while longer products would occur, on the average, less than once.

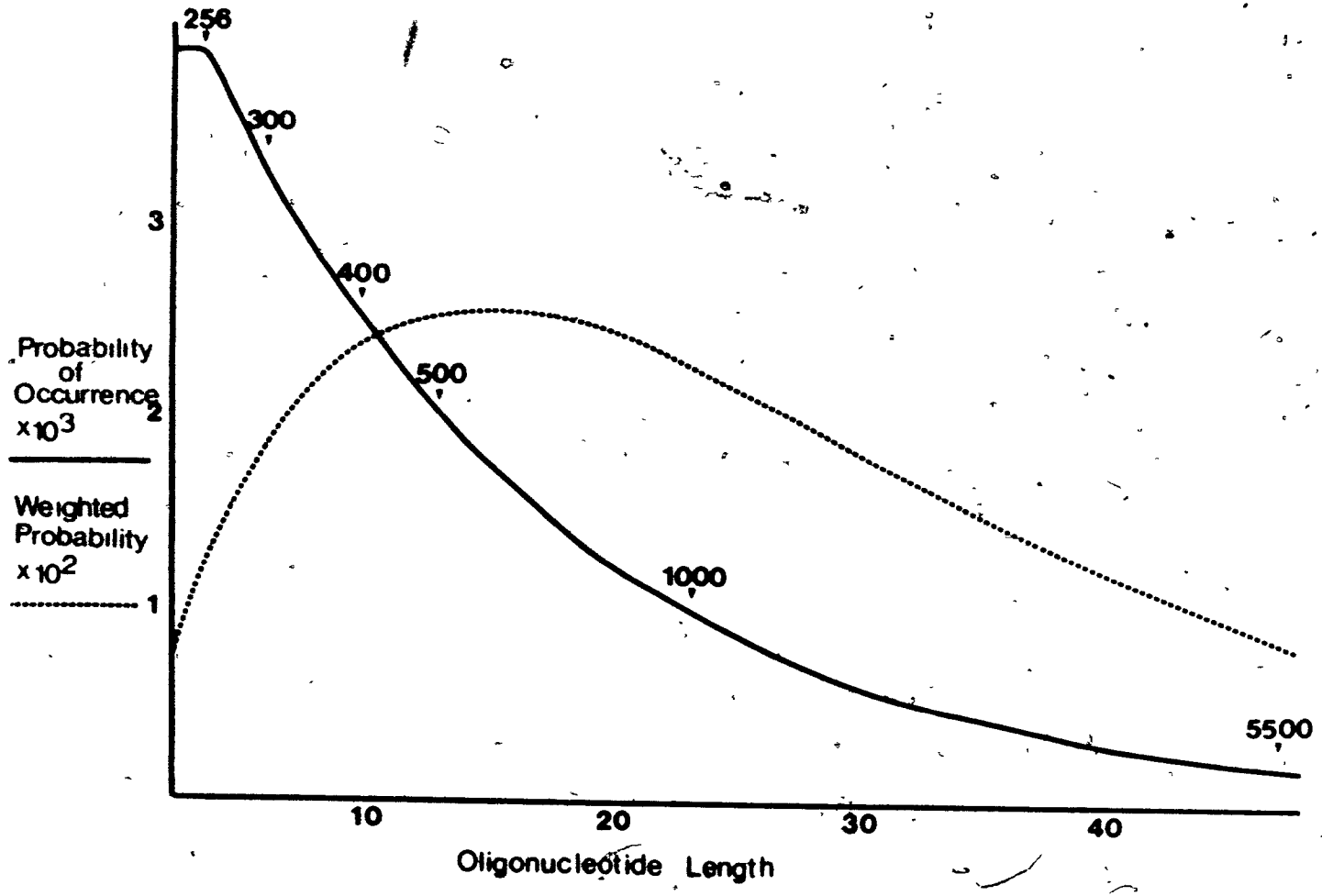


Figure 1..1 : Statistical occurrence of oligonucleotide isostichs in a T4 endonuclease IV digest of DNA of random sequence

The weighted probability of occurrence is here defined as the product of the absolute probability of occurence and the chain length of the oligonucleotide isostich. The numbers above the absolute probability curve are the reciprocals of the probabilities at that chain length, these represent the number of nucleotides in a DNA molecule which, when digested with T4 endonuclease IV, would probably yield 1 0 oligonucleotide of that chain length.

The weighted probability, defined as the product of the absolute probability and the chain length of the oligonucleotide digest product, represents the proportion of material, as moles of nucleotide, occurring in each chain length fraction in a digest of a random DNA molecule. Thus, for a radioactive digest, a maximum of radioactivity should elute with the tetradeca- and pentadeca-nucleotides.

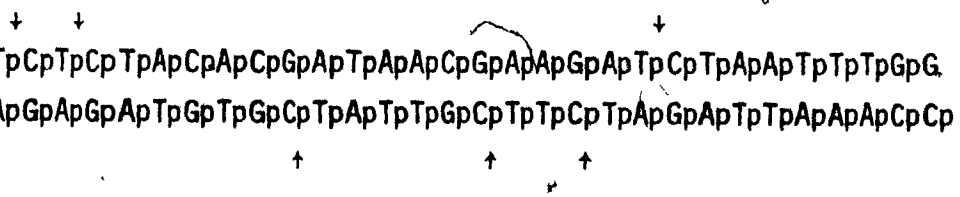
These probabilities were calculated using permutational analysis assuming random DNA sequence. From Figure I.1 it can be estimated that degradation of a double stranded DNA molecule more than 200 nucleotides in length will probably yield a mixture of fragments which will be difficult to fractionate into all of its components. Fragments smaller than this are more likely to be amenable to further analysis after degradation by T4 endonuclease IV.

I have used statistical arguments to define the boundary between "long" and "short" DNA fragments, which means the 200 nucleotide figure is not an absolute one. There are certainly polydeoxyribonucleotides containing more than 200 bases which can be analysed using T4 endonuclease IV; there will also be molecules shorter than 200 nucleotides which, because of an unusual base sequence, cannot be analysed after T4 endonuclease IV digestion.

Once a double stranded DNA fragment less than 200 base pairs in length is obtained, we encounter the problem that T4 endonuclease IV is single strand specific. F. Grosveld (68), in this laboratory, has demonstrated that double stranded DNA can be completely digested with T4 endonuclease IV if the substrate is denatured before the addition of the enzyme and the reaction performed at 37°C. An advantage of analysing a double stranded substrate is that the products will usually have overlaps necessary to determine their order. Figure 1.2 is a diagrammatic representation of a sequence experiment on a hypothetical double stranded polynucleotide thirty base pairs long illustrating this point. The bonds susceptible to hydrolysis in the presence

Figure 1.2

T4 Endonuclease IV Cleavage of a Hypothetical Fragment



of T4 endonuclease IV are indicated by the arrows. The digest products would be separated by column chromatography and their sequence determined. A summary of this hypothetical experimental data is given in Table 1.1.

We deduce that fragment 1 must be one of the 5' termini since it is the only digest product without a C at its 5' end. The absence of another such product indicates that there is a C at the other 5' terminus. These are verified by the 3' termini of fragments 5 and 6, which must be the 3' terminal sequences, since they do not end in T. Fragment 2 fits adjacent to the 5' T complementary to fragment 6. Fragment 3 must fit opposite the GAAGA in fragment 8. Fragment 4 must be adjacent to fragment 3 complementary to fragment 8. Fragment 5 is complementary to much of fragment 7, indicating that the latter is the other 5' terminal sequence. Fragment 6 overlaps the 5' end of fragment 8 and the 5' terminal sequence TCT. Finally, the overlap between fragments 7 and 8 defines the entire sequence.

Any double stranded molecule which fulfills two conditions can be sequenced in this manner: first, its T4 endonuclease IV digestion products must be separable, and second, these products must be amenable to sequence analysis.

This thesis describes sequence techniques which provide the answer to the second condition, leaving the fractionation procedures as the limiting factor.

1.2.1.4 - Oligonucleotide Sequence Analysis

Since single base specificity is not available in DNAases, it has been necessary to develop sequence methods for oligodeoxyribonucleotides using non-specific nucleases. All oligonucleotide sequence techniques described to date as well as those in this thesis depend on partial digestions with nucleases, followed

Table 1.1

Hypothetical Sequence Data

Fragment Number	Length	
1	1	pT
2	2	pCpT
3	3	TpTpCp
4	6	TpApTpTpGpCp
5	9	pCpTpApApTpTpTpGpG
6	10	ApGpApGpApTpGpTpGpCp
7	11	TpApGpApTpTpApApApCpCp
8	18	pCpTpApCpApCpGpApTpApApCpGpApApGpApT

Note: underlined sequences are written in the 3' to 5' direction.

by analysis of the progressively smaller products. Identification of all such products yields the total sequence.

In general, oligonucleotide sequence analyses fall into three major categories, uniform label methods, and two methods utilizing enzymes for labelling oligonucleotide termini.

a) Uniformly labelled oligonucleotides

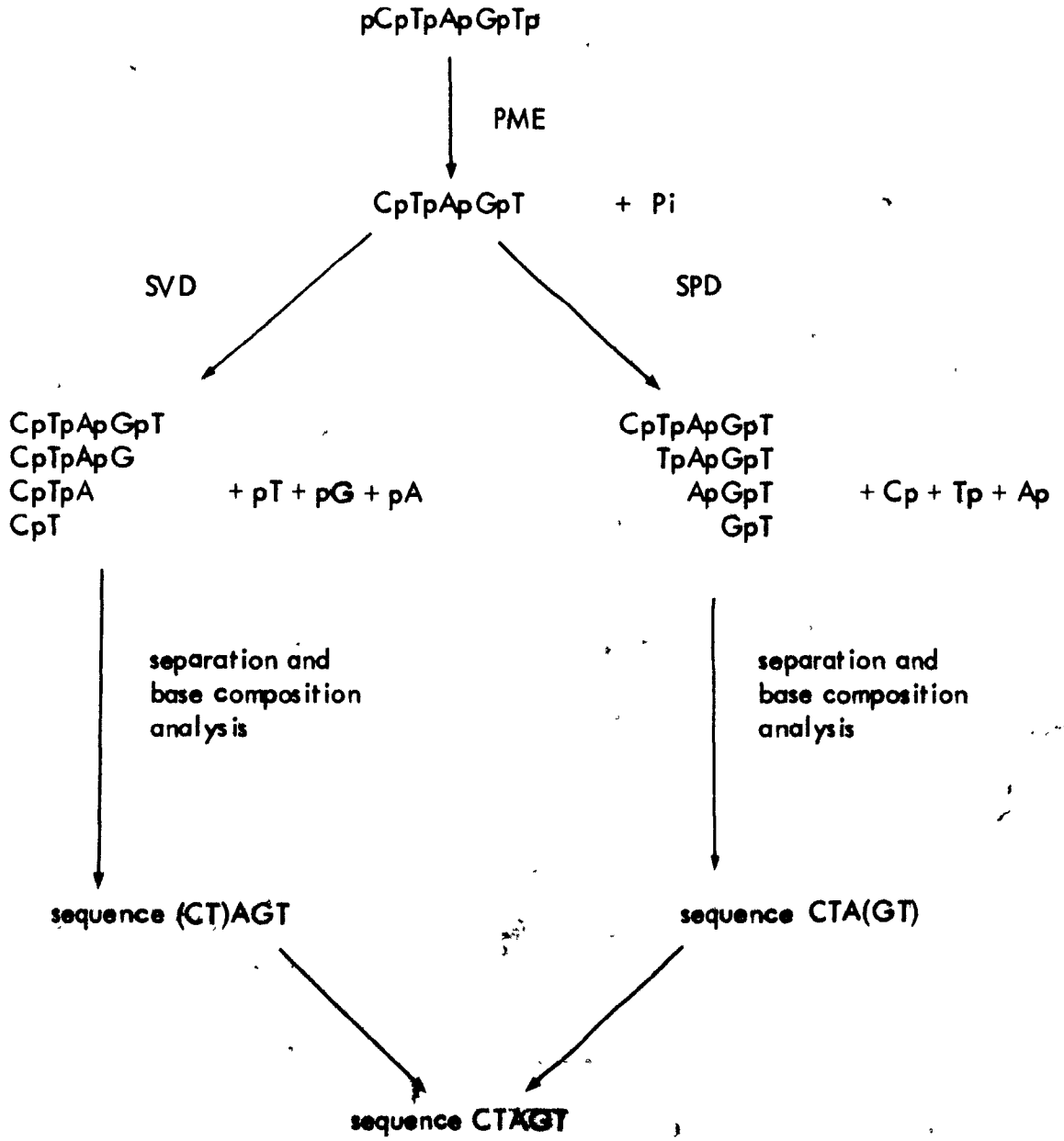
Oligonucleotides with sufficient uniform radioactive label can be sequenced using exonuclease digestion and base composition analysis. Figure 1.3 is a representation of a hypothetical sequence experiment illustrating this method. A portion of a dephosphorylated pentanucleotide is partially digested with venom exonuclease, yielding a series of digest products with nucleotides removed from the 3' end; another portion is treated with spleen exonuclease, yielding a series of products with monomers removed from the 5' end. For pyrimidine oligonucleotides the base composition of the digest products can be deduced from their migration in an electrophoresis-homochromatography system, yielding the complete sequence. If the substrate oligonucleotide contains all four bases, a base composition analysis must be performed on most digest fragments because relative migration does not distinguish between a T and a G.

b) Terminal labelling of an oligonucleotide followed by exonuclease treatment

An oligonucleotide can be labelled <u>in vitro</u> using enzymes such as polynucleotide kinase (12) or deoxynucleotidyl transferase (60). The first enzyme catalyses the transfer of the γ-phosphate of ATP to the 5' hydroxyl group of a polynucleotide. The latter esterifies several adenylate mononucleotide units to the

Figure 1.3

Sequence of a Uniformly Labelled Oligonucleotide



3' terminal hydroxyl group of an oligodeoxyribonucleotide. Excess adenylate residues can be removed from the 3' terminal region by alkaline digestion leaving a single labelled adenylate residue esterified to each 3' hydroxyl group.

An oligonucleotide labelled by incubation with polynucleotide kinase and [γ-^{32}P]ATP is purified and then partially digested with a 3' exonuclease such as venom exonuclease. Conversely, if deoxynucleotidyl transferase and [α-^{32}P]ATP are used in the labelling step the degradation enzyme is a 5' exonuclease such as spleen exonuclease. Only one exonuclease partial digest is required after terminal labelling because each sequence position, including the termini, are represented by radioactive species of different chain length.

Figure 1.4 represents a sequence determination of a hypothetical pentanucleotide after the labelling of the 5' terminus with [γ-^{32}P]ATP and polynucleotide kinase. In this case the base composition of each of the digest products must be determined by chromatographic procedures alone. This is readily accomplished for pyrimidine (Section 1.2.1.4a) oligonucleotides, but is difficult when all four bases are present (69, 76, 111).

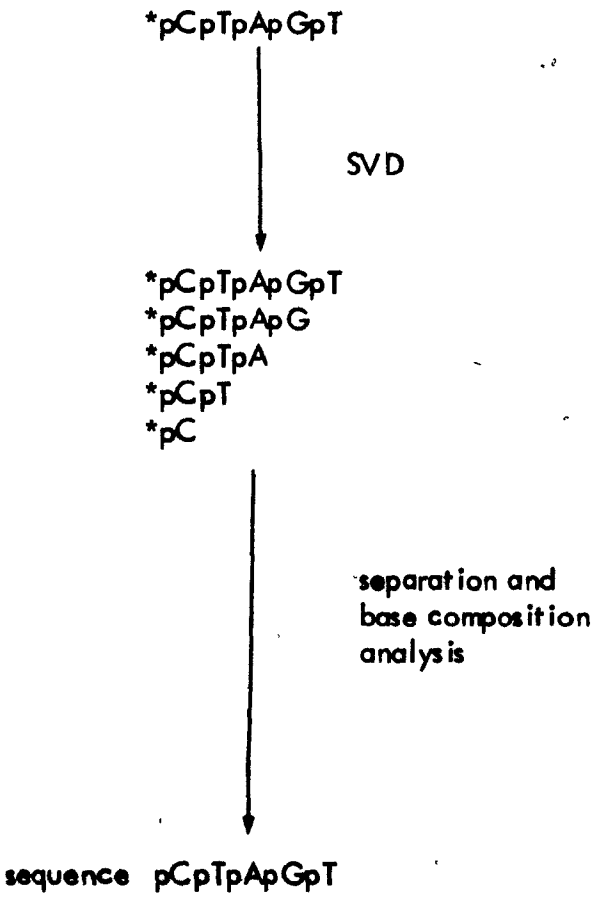
c) Exonuclease treatment of an oligonucleotide followed by terminal labelling

A method for sequence analysis which solves the problems of specific radioactivity and of base composition analysis is proposed : the oligonucleotide under study is first partially digested with spleen exonuclease; the newly generated 5' termini are labelled with [γ-^{32}P]ATP by polynucleotide kinase followed by separation of the labelled products.

Since only the 5' termini of the products are labelled, radioactive base composition analysis of each product yields the complete sequence except for the 3' terminus.

Figure 1.4

Sequence of a 5' Terminally Labelled Oligonucleotide



*p represents a radioactive phosphate ester.

The use of deoxynucleotidyl transferase and [α-^{32}P] ATP to label 3' termini after venom exonuclease partial digestion, gives an exactly complementary sequence method which will yield a complete sequence except for the 5' terminus.

When either of these methods is used alone, the unknown terminus can be identified chromatographically (51, 64, 65) or by determination of the sequence of one of the intermediate labelled digest products as described in Section 1.2.1.4 b.

Figure 1.5 represents an experiment using this approach with polynucleotide kinase to sequence a hypothetical pentanucleotide. Here the dinucleotide at the 3' terminus is analysed chromatographically (51).

Figure 1.6 represents the converse experiment using the 3' terminal labelling enzyme. The 5' dinucleotide sequence is determined using its 3' terminal label and spleen exonuclease partial digestion (1.2.1.4 b).

d) Fractionation of oligonucleotides

The resolution of the available fractionation technique determines the maximum length of oligonucleotide which can be sequenced by the methods discussed above. The development of electrophoresis homochromatography (28) and high-resolution DEAE sephadex column chromatography (22, 51) allow the separation of mixtures of oligonucleotides up to the pentaicosanucleotide level.

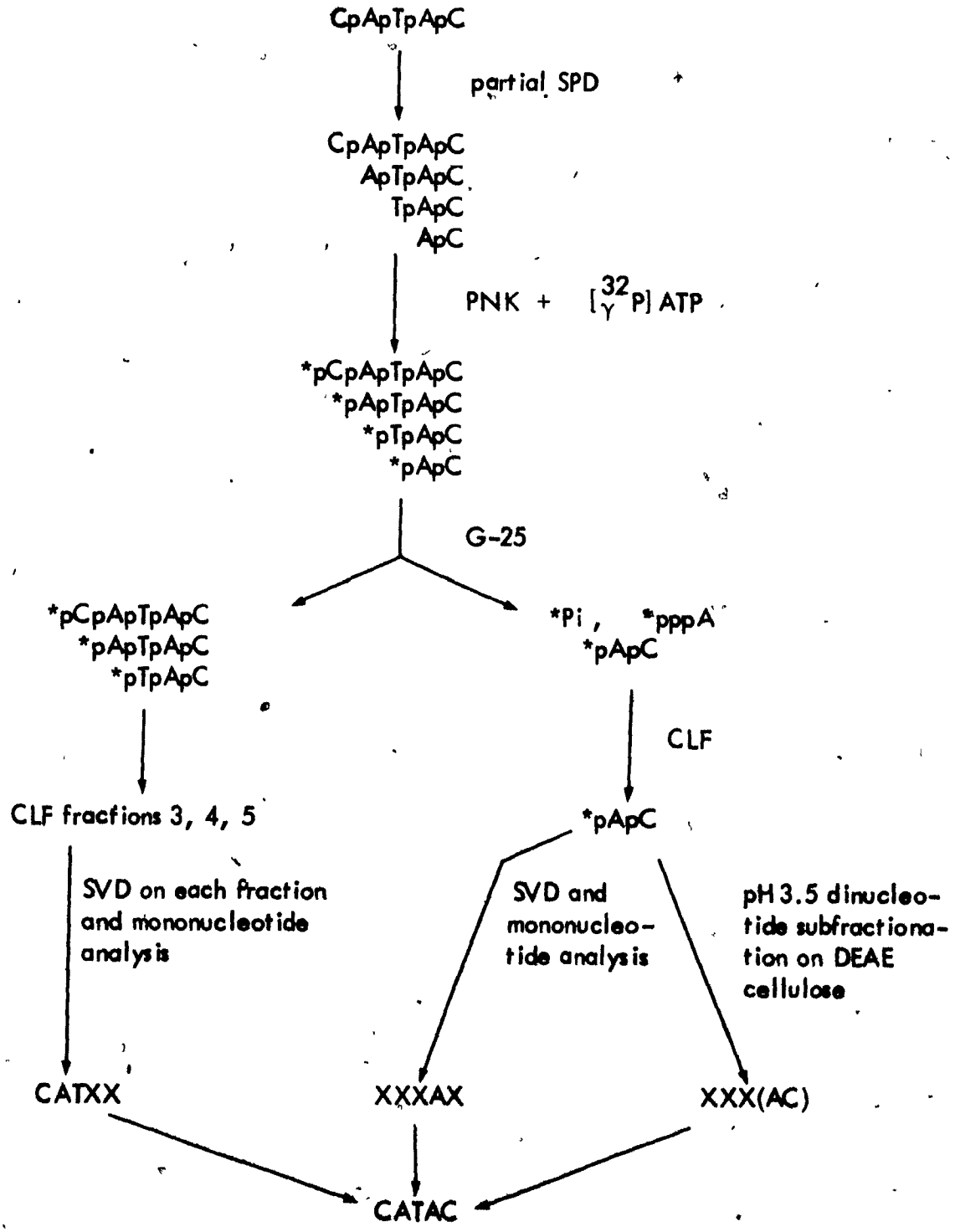
The complementary sequence methods described in Section 1.2.1.4 c allow sequence analysis on molecules up to 50 nucleotides in length.

1.2.1.5 Progress on Direct Sequence Analysis

Ling (7, 26) has sequenced pyrimidine oligonucleotides up to length 20

Figure 1.5

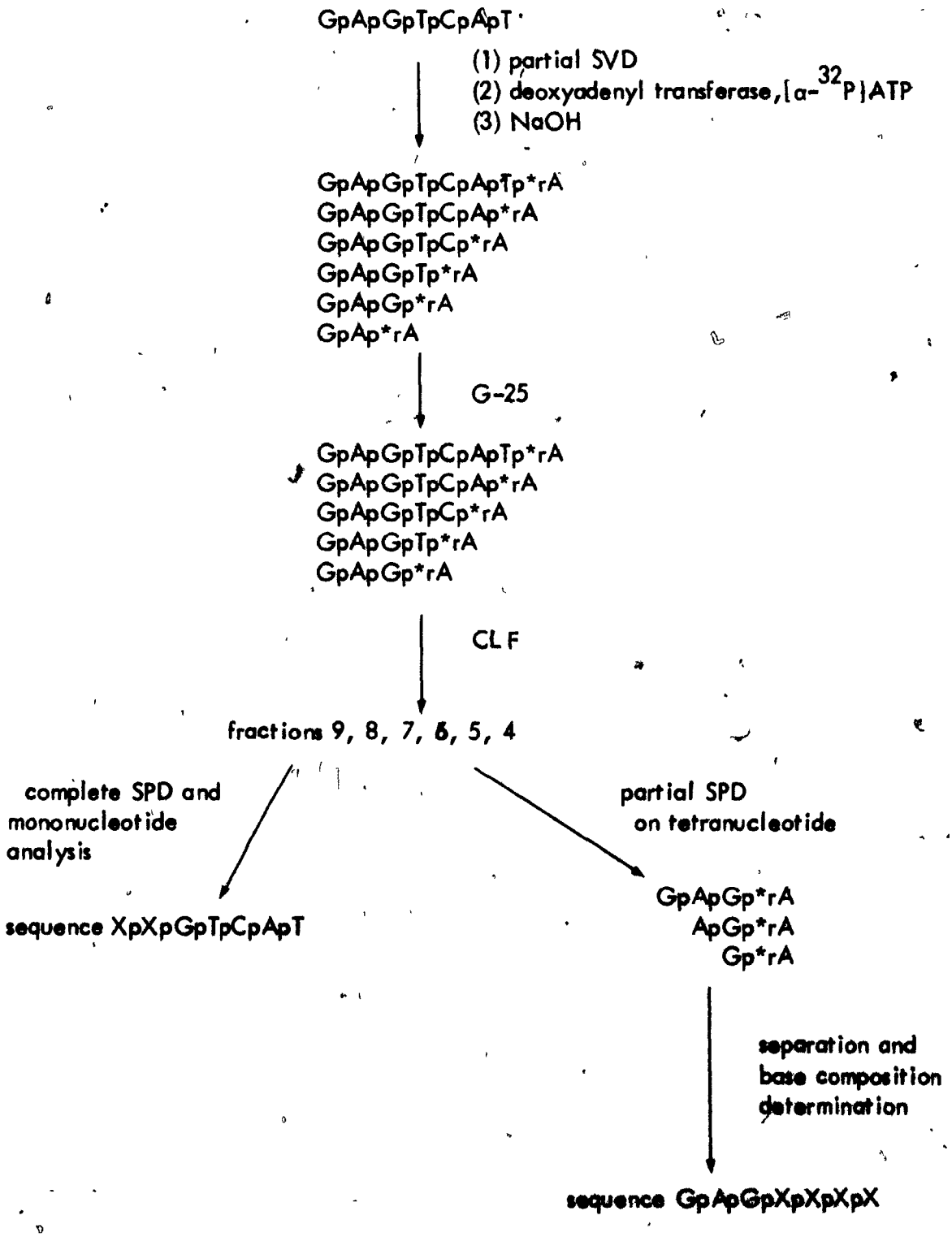
Sequence of an Oligonucleotide Using Spleen Exonuclease and Polynucleotide Kinase



*p represents a radioactive phosphate ester.

Figure 1.6

Sequence of an Oligonucleotide Using Venom Exonuclease and Deoxynucleotidyl Transferase



p* represents a radioactive phosphate ester.

in the small coliphages fd, ØX174 and f1 using the uniformly labelled sequence method described in Sections 1.2.1.4 and 2.1.10.1. Recent results from our own laboratory indicate that some of the base composition assignments for the pyrimidine clusters of ØX174 were in error. Confirmatory work on pyrimidine cluster sequences from ØX174 DNA is in progress here (27, 117).

Kossell and Roychoudhury (60) and Wu (69, 70) have presented sequences of synthetic oligonucleotides determined by methods similar to those presented by this investigator (64); they had the advantage of knowing the sequence beforehand and of having large amounts of material; they have yet to apply their methods to a biological molecule of unknown sequence.

A group of workers from Sanger's laboratory (67) have proposed a preliminary sequence of a fragment of ØX174 DNA 48 nucleotides in length isolated by partial digestion with T4 endonuclease IV. This fragment was sequenced by analysing smaller overlapping oligonucleotides produced by depurination and by complete endonuclease IV digestion. Uniform radioactive label was used throughout the procedure. The long-awaited comprehensive publication describing this sequence study has not yet appeared.

Attempts to repeat the specific fragmentation of ØX174 and S13 DNA with T4 endonuclease IV have not been successful in this laboratory (68).

1.2.2 The Indirect Approaches to DNA Sequence Analysis

The preceding pages described sequence determination as a complete characterization of a DNA molecule regardless of size or function. Many investigators, however, are unwilling to wait for complete sequence data on the small phages;

others wish to examine specific sequences within genomes much too large for direct sequence analysis. In these cases a biological or physical "handle" is used to isolate a fragment of interest from the genome.

1.2.2.1 Terminal Sequences

The most obvious regions for investigation in a linear DNA molecule are the termini. The cohesive ends, or single stranded regions, at the 5' termini of bacteriophages λ and 186 have been sequenced by Wu and his colleagues (72, 73, 74). They incorporated [α-^{32}P] nucleoside triphosphates into the terminal cohesive end regions of the DNA using the repair reaction of E. coli DNA polymerase I. Nearest neighbor analysis was then used to deduce the terminal sequences.

Englund (77, 78) used the reverse reaction of T4 DNA polymerase to examine the termini of T7 DNA, which does not have cohesive ends. He incubated the DNA in the presence of polymerase and only one nucleoside triphosphate. Under these conditions the polymerase removes nucleotides from the 3' terminus until it encounters a nucleotide corresponding to the triphosphate present in the incubation medium. Nearest neighbor analysis of the products of a series of these reactions yielded the trinucleotide at each 3' terminus of the T7 DNA. The same group (78) extended the sequence for the l-strand terminus to seven bases by isolating (71) and sequencing pyrimidine pentanucleotides labelled with T4 DNA polymerase in the presence of [α-^{32}P] TTP.

Weigel et al., (75) used Englund's (77) technique to label the 3' termini of λ DNA and then used pancreatic DNAase to generate a series of 3' terminal oligonucleotides. Identification of these oligonucleotides gave the sequence of a

hexanucleotide for the l-strand and a heptanucleotide for the r-strand. Wu has recently added one more nucleotide to the l-strand sequence (129) verifying Weigel's prediction for that nucleotide. This brings the known sequence around the λ DNA cohesive ends to 26 base pairs.

Murray (14) used polynucleotide kinase and [γ -^{32}P] ATP to label the 5' termini of λ DNA. He then digested the DNA with DNAase I to yield a series of terminal oligonucleotides. Identification of these oligonucleotides verified part of Wu's sequence and showed that coliphage 424 DNA has termini identical to those of λ DNA at least to the hexanucleotide level.

It is possible to generate from a completely double stranded linear DNA molecule single stranded tails by digestion from the 3' ends with exonuclease III, yielding 5' termini similar to the cohesive ends of λ DNA. It has been suggested that such a molecule be examined as discussed above (79).

1.2.2.2 Single Strand Specific Nuclease Digestion of DNA

Hydrolysis of partially annealed DNA with a single strand specific nuclease such as Neurospora crassa endonuclease (81, 84) or Aspergillus nuclease S1 (137) has been used in many studies of DNA and RNA function (82, 83, 84).

Single stranded DNA molecules can have significant secondary structure. From phage ØX174 DNA one can isolate fragments corresponding to about 1.5% of the molecule after self-annealing, N. crassa nuclease digestion, and gel electrophoresis (85). Such fragments are of particular interest since secondary structure has been implicated in functional sites in DNA molecules (87, 88, 89).

Another avenue opened by single strand specific enzymes is the investigation

of evolutionarily stable sequences in related phages. An intriguing study of heteroduplexes between the replicative forms of S13 and ØX174 bacteriophages by Godson (90, 145) demonstrated that only 5% of the length of any one heteroduplex molecule remained as a stable double helix in 75% formamide. A single strand specific nuclease can digest the single stranded regions of the heteroduplex leaving the duplex regions intact (81, 83, 84, 137). Godson's work (90) suggests that S13 - ØX174 heteroduplexes digested in this way would yield a double stranded fragment approximately 200 nucleotides in length.

It is likely that an evolutionarily stable DNA sequence such as this would represent functional DNA, rather than DNA coding for m-RNA. As Godson (90) has discussed, proteins and the DNA sequences coding for them can evolve considerably with little changes in function.

1.2.2.3 Sites of Interaction of Proteins and DNA

Considerable progress has been made in the isolation of segments of DNA which can be specifically protected from nuclease digestion by proteins. Ribosome binding sites (91), RNA polymerase binding sites (92, 93), and lac repressor bound DNA have been isolated (146, 147). Investigation of the binding of the CAP protein as well as that of the termination factor ρ is still at the stage (98, 99) where quantity of protein is the limiting factor.

A sequence has been proposed for a ribosome binding site isolated from ØX174 (91). This was determined using depurination and T4 endonuclease IV as described previously (Section 1.2.1.5) (67). This work was also presented as a preliminary study and has yet to be followed by a detailed article.

The isolation of DNA protected from nucleases by RNA polymerases has been performed for several years (92, 102, 103, 104). It has recently been shown that a single RNA polymerase binding site can be isolated from bacteriophage fd RF DNA (93).

It has been demonstrated that phage ØX174 RF DNA has three RNA polymerase binding sites (105). After restriction enzyme treatment and gel electrophoresis each binding site was isolated separately. This approach should be very useful in the work in progress on DNA from bacteriophages λ and T7, which has been hindered by the large number of binding sites in these genomes.

To date investigators have been unable to produce enough material for a sequence study of an RNA polymerase binding site; such a project is under way in this laboratory.

Lac repressor has been produced in large quantities (100, 101) from E. coli infected with i^q or i^{sq} strains of λ and has been used to isolate the stretch of DNA responsible for its binding (96). A sequence for most of this repressor binding site has been proposed (96) on the basis of pyrimidine cluster sequence data and RNA transcript sequence data (Section 1.2.2.4).

1.2.2.4 The Sequence of RNA Transcripts of DNA

RNA sequence techniques are so well established that it is quite plausible to determine a DNA sequence by analysis of its RNA transcript. Gilbert and Maxam (96) used RNA polymerase as a non-specific transcription enzyme in the elucidation of the sequence of the lac operator. Their study points out a disadvantage in the use of RNA polymerase on relatively small segments of DNA; transcription

is not efficient near the termini of the template.

The study by Maizels (106) of the nucleotide sequence of the lactose messenger RNA is a demonstration of the RNA transcript technique at its strongest. This presentation is of an RNA sequence, but Watson-Crick base pairing yields the DNA sequence. The sequence she presents overlaps with the lac operator sequence and with the sequence of the B-galactosidase cistron. This is the first description of a link between such functional units of DNA.

The strength of this technique lies in the use of a primer to initiate transcription of a specific site.

The disadvantages of course, are that RNA polymerase transcripts will not yield information on inter-operon regions such as their own binding site, and that any molecule with more than one RNA polymerase binding site may not give specific initiation.

1.2.2.5 Elongation of an Oligodeoxyribonucleotide Primer with DNA Polymerases

DNA polymerase has already been mentioned as a tool in 3' end labelling and terminal sequence analysis and here it will be discussed regarding its use as a tool for sequence determination within a large DNA molecule.

If a polydeoxyribonucleotide can be isolated or synthesized, which is complementary to a known region of a genome, it can be used as a primer and elongated by E. coli DNA polymerase I or T4 DNA polymerase. Here again specific initiation is the key to a meaningful result. Wu (107) has investigated the binding of a nonanucleotide to the cohesive end of λ DNA; Oertel and Schaller (108) have investigated the binding of the pyrimidine tract C_9T_{11} to fd minus strand DNA and

have elongated it using the RF complementary strand as a template. Although they did not elucidate any sequences, their data suggests that the approach is feasible.

Berg et al., (109) have shown that E. coli DNA polymerase I will incorporate ribonucleotides into a DNA polymer if manganese is the divalent ion in the incubation instead of magnesium. Since ribonucleotides are susceptible to ribonuclease or alkali digestion, this procedure (138) has been proposed as a replacement for the specific RNAases. Salser et al., (115) have shown that fingerprints can be produced from M13 DNA using ribosubstitution and alkaline hydrolysis.

Khorana s group (110) has examined ribonucleotide incorporation onto synthetic primers bound to synthetic templates. In this way they showed that, at 37°C, misincorporation of ribonucleotides is a serious problem in the ribosubstitution system. At 10°C they have demonstrated that only rCTP can be incorporated reliably. Another problem is that if two consecutive ribonucleotides must be incorporated elongation will effectively come to a halt. These facts reduce the flexibility of the procedure, but it remains a powerful technique.

Sanger et al., (111) starting with a primer only 8 nucleotides in length annealed to f1 DNA, added 50 residues to this octanucleotide and deduced their sequence. The study used timed nucleoside triphosphate incorporations, ribocytidine triphosphate incorporation , and small oligonucleotide fractionation methods.

The octanucleotide was thought to be complementary to an unambiguous sequence deduced from the amino acid sequence of the major coat protein. Instead, the sequence deduced seems to be an interesting intercistronic region with several terminator and initiator sequences.

Work in this laboratory is in progress leading to the use of unique pyrimidine nonanucleotides and undecanucleotides from S13 RF DNA as primers annealed to $S13^+$ viral DNA (148). The leading group, as regards sequence work using DNA polymerases, is Khorana and his colleagues at the Massachusetts Institute of Technology. Having synthesized the Ø80psuIII tyrosine t-RNA gene (112), they had primers which could be bound to any segment of the t-RNA cistron. Using a terminal oligonucleotide from the gene and elongating beyond the terminus with the r strand as template allowed them to examine the post-gene sequence (113). The sequence of the twenty-three nucleotides immediately following the gene was determined using a variety of techniques. The major approach was to vary the deoxynucleoside triphosphates present in the incubation mixture. Incorporation of only two or three triphosphates followed by nearest neighbor analyses allowed sequence determination in short jumps of one to four nucleotides at a time.

1.3 Mammalian DNA Sequence Analysis

The large size of mammalian genomes makes it impractical to study any unique sequences, therefore the targets in mammalian systems have been highly amplified sequences such as those occurring in many satellite DNA's .

Southern (114), using pyrimidine cluster sequence data alone, deduced the probable hexanucleotide repeating unit for the guinea pig a satellite DNA. The sequence he presented was the most frequently occurring one; the many other less frequent sequences were attributed to evolutionary divergence.

A study of the sequence properties of the kangaroo rat HS-B satellite DNA indicates that its repeating unit contains 10 nucleotides (116) and that its

evolutionary divergence is considerably less than that of guinea pig a satellite. These investigators used DNA polymerase ribosubstitution techniques and transcription into RNA as parts of their analysis. The ribosubstitution experiments were performed using r GTP as the one ribonucleoside triphosphate; this is contradictory to the demonstration by van de Sande et al., (37) that misincorporation of this ribonucleotide is a serious problem. This might explain the observation that most of the evolutionary alternative sequences involved the addition of one or more guanylic acid residues to the basic repeating unit.

In this laboratory (117), all the major pyrimidine oligonucleotides from mouse satellite DNA have been sequenced, but the repetitive unit, if there is one, is more complex than those already published, and more data are being sought before a firm hypothesis is presented.

1.4 Summary of DNA Sequences

The DNA sequence field is still very young; the DNA sequences which have been proposed form a set which is small enough to be presented in compact form. Table 1.2 summarizes the published DNA sequences to date.

1.5 Purpose of This Study

The most successful techniques to date have been those involving DNA polymerases and RNA polymerases. This is because these techniques can produce some sequence data without requiring the sequence of an oligonucleotide containing more than four bases. Manipulation of the deoxyribonucleoside triphosphates present in the polymerization reaction and the use of ribocytidine triphosphates partially fill

Table 1.2

DNA Sequences

Source	Description	Sequence	References and Notes
Ø80suIII	untranscribed strand after the suIII tyrt-RNA gene	TCACTTTCAAAAGTCCCTGAACT	(113)
ØX174 viral DNA	T4 endonuclease IV produced fragment	CCCATCTTGGCTTCCTTGCTGGGTCAGAT TGGTCGTCTTATTACCATTT	(67)
ØX174 viral DNA	ribosome binding site	AGGTTTTCTG *CTTAGGA *TTTAATC ATGTTTCAGACTTTTATTTCTCGCCAC	* indicates the position of a possible additional purine. the underlined portion codes for the N terminal sequence of the ØX174 spike protein.(91).
f1 viral DNA	sequence following an octanucleotide hybridized to the viral DNA	GGCTTTATTGCTTAATTTTGCTAATTCTTT GCCTTGCCTGTATGATTTATTGGATGGT	the doubly underlined sequence represents the complement of the bound oligonucleotide. the dotted underlines represent initiator sequences. the singly underlined sequences represent in phase termination codons (111).

Table 1.2 .....cont'd.

Source	Description	Sequence	References and Notes
coliphage λ DNA	double stranded sequence of the cohesive ends of λ DNA	5' GTTACGGGG * l strand CGGCGACCTCGCGGGT 3' 3' CAATGCCCCGCCGCTGGAGCGCCCA 5' r strand *	* indicates the position of the nicks producing linear λ DNA. (73, 74, 75).
coliphage 186 DNA	5' terminal cohesive end of the left half of 186 DNA	GGCGTGGCGGGGAAAGCAT-----	the 3' terminal T is the first nucleotide which is base paired in linear 186 DNA. (72, 74).
coliphage T7	l strand 3' terminus	-----TCCCTGT	(77, 78).
coliphage T7	r strand 3' terminus	-----AGA	(77).
coliphage λ h80C 857S7 dlac UV5	sequence of untranscribed strand of the lac repressor binding site and the beginning of the B-galactosidase cistron.	TGGAATTGTGAGCGGATAACAATTTCACACAG GAAACAGCTATGACCATGATTACGGATTCA CTGG	dotted line represents the lac operator sequence (96) single underline indicates the sequence coding for the N-terminus of B-galactosidase. (106) double underline indicates an initiator sequence.
Guinea pig	one strand of the repeating unit of α-satellite.	-----CCCTAA-----	(114).

Table 1.2 .....cont'd.

Source	Description	Sequence		References and Notes
Kangaroo rat	one strand of the repeating unit of HS-B satellite DNA.	-----ACACAGCGGG-----		(116).
Mouse	pyrimidine clusters from the heavy strand of satellite DNA.		Occurrence per 1500 base pairs	
		TTCC TCTC TCCT	230	
		TTTC	100	
		TTCT	45	
		TTTTC	85	(174)
		TTTCT	85	
		TTTTCC	145	
		TTTCTC	145	
		TTTTTC	140	
		TTTTCTC	30	
		CTTTTTC	20	
		TTTTTCT	50	
		TTTTCCTC	40	
		CCTTTTTC	40	

Table 1.2 .....cont'd.

Source	Description	Sequence	References and Notes
Mouse	pyrimidine clusters from the light strand of the satellite DNA.	Occurreace per 1500 base pairs	
		TCC 80	
		CCT 40	
		CTT 120	
		CTTT 60	

a role analagous to that of RNAase T1 and pancreatic RNAase.

In order to obtain extensive sequence information, such as Sanger et al., (111) have proposed, an oligonucleotide sequence method becomes necessary even when using DNA polymerase techniques. They proposed a sequence of 50 residues from f1 DNA and showed that it was necessary to sequence intermediate oligonucleotides containing up to 15 bases. Unfortunately, they did not describe how this was accomplished.

The only sequences, other than those of pyrimidine clusters, which have been determined by direct methods are the two large fragments isolated from ØX174 DNA (67, 91). This work has yet to be presented in a detailed form.

The advances described in this thesis provide a sequence method which is essential for direct sequence analysis of DNA and for extensive analysis by indirect methods. The specific enzymes are available which will degrade high molecular weight DNA to fragments less than 50 nucleotides long. Fractionation techniques for the analysis of oligonucleotides of this size, as well as terminal labelling methods for the transfer of high specific activity radioactive label to these molecules, are described. These procedures complete the development of a general sequence method for DNA.

CHAPTER 2

MATERIALS AND METHODS

2.1 Materials

Nucleosides, deoxynucleosides, and their mono-, di-, and triphosphates were purchased from Raylo Chemicals Inc., Edmonton, Alberta, as were thymidine 5'-monophospho-p-nitrophenyl ester and thymidine 3'-monophospho-p-nitrophenyl ester.

Chloramphenicol, p-nitrophenyl phosphate, pancreatic RNAase (E.C.2.7.7.16) (5x crystallized, grade A), lysozyme (E.C.3.2.1.17) (3x crystallized), 3-phosphoglyceric acid, 3-phosphoglyceraldehyde, nicotinic acid adenine dinucleotide, oxidized form (NAD), and reduced form (NADH), phenazine methosulfate, thiazolyl blue, 3-phosphoglyceraldehyde dehydrogenase (E.C.1.2.1.12) (yeast), and 3-phosphoglycerate phosphokinase (E.C.2.7.2.3.) (yeast), were obtained from Sigma Chemical Co.

DNAase I (E.C.3.1.4.5, grade DP) was from Worthington Biochemical Corporation.

Optical grade and radio-tracer grade cesium chloride was obtained from the Harshaw Chemical Co. Ultrapure sucrose was from Schwartz Bioresearch Co. Polyethylene glycol (Carbowax 6000) was from Union Carbide Co.

Carrier free $H_3{}^{32}PO_4$ was obtained from New England Nuclear Corporation.

All other chemicals were "Reagent" grade. Reagent grade phenol (BDH) was distilled before use.

Diphenylamine (Fisher) was recrystallized twice from ethanol (149).

Reagent grade triethylamine was distilled before use.

Solutions of reagent grade urea (BDH) (7 M) were slurried with 0.25 g activated charcoal/l. urea (150). The charcoal was removed by filtration through Whatman 3MM paper and then through a 2.5 x 5 cm DEAE cellulose column (51).

Charcoal, Norit A (Fisher), was activated by refluxing with 2 N HCl followed by washing with ethanol : ammonia : water (50:15:35) and subsequently with water to neutrality (150).

Dowex 50 was obtained from Bio-Rad Corp. as AG50 Wx4 , minus 400 mesh, and was cleaned according to Blattner and Erickson (1).

DEAE cellulose (Whatman DE-11), was sieved prior to use, the portion passing through a 200 U.S. standard sieve but retained by a 325 sieve being used in all fractionation experiments, and the portion held up by the 200 mesh being used in all desalting operations (51). The DEAE cellulose was washed with cycles of 0.5 N NaOH and 0.5 N HCl and finally left in the hydroxyl form (150). The formate form was prepared by suspending the hydroxyl form in 1 M formic acid followed by washing with water (150).

DEAE Sephadex A-25 was washed with cycles of 0.2 N NaOH and 0.2 N HCl each time washing with water until the eluting filtrate was neutral.

Sephadex G-10 , G-15 , G-25 , G-75, Sepharose 6B, and CM-Sephadex were prepared according to the instructions of the manufacturer.

Phosphocellulose (Whatman P-11) was washed with cycles of 0.2 N HCl and 0.2 N NaOH (51) each time washing with water until the eluting filtrate was neutral.

Hydroxyapatite was prepared by alkali treatment of brushite ($CaHPO_4 \cdot 2H_2O$)(4).

DNAase A, a fraction of DNAase I purified using phosphocellulose column chromatography (51,52) as described by Salnikow, Stein and Moore (2), was kindly supplied by Dr. E.Junowicz, Calf thymus DNA was prepared from fresh tissue by the method of Kay et al., (6).

TEAB buffer (3) was prepared by the addition of 1 mole triethylamine per l. solution in a vessel kept at 0°C and bubbling CO_2 through the solution until the pH was 8.0 (51).

A mixture of oligonucleotide markers was prepared by digesting 500 mg of calf thymus DNA with DNAase I as described by Junowicz and Spencer (22). This was used as marker for dinucleotide isolation (2.2.10.3c). Dinucleotide markers were prepared by fractionation of this digest on a column of DEAE Sephadex (22).

Original cultures of E. coli C and Shigella dysenteriae Y6R as well as bacteriophage SI3$^+$ and SI3suN15 were gifts from Drs. E. and I.Tessman, Purdue University, E. coli B Cr63 and bacteriophage T4D$^+$ were gifts of Dr. T.W. Conway, University of Iowa. Bacteriophage T4n82 (151) was a gift from Dr. Imo Scheffler, Harvard University. E. coli B was obtained from the Department of Microbiology, McGill University.

2.2 Methods

2.2.1 Preparation and Use of Venom Exonuclease

Venom exonuclease (E.C. 3.1.4.1), grade VPH, obtained from Worthington Biochemical Corporation, was dissolved in water and acetic acid added to pH 3.6. The solution was incubated for 3 hrs at 37°C to destroy 5' nucleotidase activity (5).

Venom exonuclease was assayed by incubating an aliquot of the enzyme

in a solution containing 10 mM $MgCl_2$, 50 mM Tris-HCl buffer, pH 9.0 and 1 mM thymidine 5' monophospho-p-nitrophenyl ester. One unit was defined as the amount of enzyme required to cause a change in absorbance at 410 nm of 1 absorbance unit per minute per ml. The average 5 mg vial from Worthington was found to contain 2000 units (24 μmoles ml^{-1} min^{-1}).

Degradation of oligonucleotides to 5' mononucleotides was accomplished as follows: the oligonucleotide solution, of volume between 0.1 ml and 2 ml, was made 50 mM in Tris-HCl buffer pH 9.0, 10 mM in $MgCl_2$, 1 mM in each of the deoxynucleoside monophosphates, and 24 units/ml in venom exonuclease; this solution was incubated 4 hours at 37°C. In control experiments less than 3% of the ^{32}P radioactivity was not absorbed by charcoal and more than 90% of the ^{32}P radioactivity eluted from a DEAE cellulose column with the mononucleotide fraction. In order to be certain that there were no products of incomplete venom exonuclease digestion, such as dinucleotides, present, 1/5th volume of a DEAE cellulose suspension (200 mesh) was added, mixed with the completed digest and then removed by centrifugation in a clinical centrifuge. The DEAE cellulose binds any oligonucleotides, leaving the mononucleotides in the supernatant.

Partial venom exonuclease digestions were performed in two ways: If the oligonucleotide under study was carrier free (less than 10 nmol) it was diluted into a final volume of 0.2 ml of 50 mM Tris-HCl buffer, pH 9.0, 10 mM $MgCl_2$, a 20 μl aliquot removed, and 0.4 units of venom exonuclease added. 20 μl aliquots were removed at one minute intervals and pipetted into 1 ml triethylamine to stop the reaction. After the entire digest had been transferred to the triethylamine, the

mixture was heated in a 100°C water bath for 10 minutes to complete the destruction of the exonuclease activity. It was then evaporated to dryness in a glass tube in a Buchler Evapo-mix operating at a bath temperatureof 45°C. The sample was dissolved in a volume of water or a buffer appropriate to the analysis being performed.

If the sample to be digested contained substantial amounts of RNA or DNA carrier, as for example the 100 µg RNA in a spot eluted from a DEAE cellulose plate after homochromatography, the digestion was done as follows (7). The oligonucleotide and the carrier were dissolved to a concentration of 10 mg/ml in a buffer containing 100 mM Tris-HCl buffer pH 8.9, 10 mM $MgCl_2$ and 400 units/ml venom exonuclease. The digest was performed at 37°C for 30 minutes and the reaction stopped by spotting the solution onto a cellulose acetate strip wet with a pH 3.5, 7 M urea buffer prior to electrophoresis (7) (2.2.9.4).

2.2.2 Preparation and Use of Spleen Exonuclease

Spleen exonuclease (E.C.3.1.4.1.), grade SPH, obtained from Worthington Biochemical Corporation, was used in some reactions containing large amounts of RNA and requiring glycerol free exonuclease.

The spleen exonuclease used in most experiments was prepared by the method of Bernardi and Bernardi (8).

This enzyme was assayed by incubating an aliquot of the enzyme in a solution containing 50 mM sodium succinate buffer pH 5.5 and 1 mM thymidine-3'-monophospho-p-nitrophenyl ester. One unit was defined as described in the venom exonuclease assay (2.2.1). The average 10 mg vial of commercial enzyme was found to contain 20-30 units (0.3-0.5 µmole ml^{-1} min^{-1}).

The spleen exonuclease prepared in this laboratory was stored at -20°C after dialysis against 50 mM Tris- HCl buffer pH 7.5, 50% glycerol at a concentration of 75 units/ml.

Partial digests were again done in two ways. A carrier free oligonucleotide was diluted into a final volume of 0.5 ml of 50 mM sodium succinate buffer, pH 5.5, a 50 µl. aliquot removed, and 0.075 units of spleen exonuclease added. Similar aliquots were removed at one minute intervals and were combined with the initial aliquot in a test tube standing in a 90°C water bath. The resulting digest was usually then treated with more enzymes, or was desalted for application to electrophoresis.

In cases where a spot recovered from a homochromatography plate was to be partially digested commercial enzyme was used and the digest was performed in a manner similar to the venom exonuclease digestion of such material (2.2.1.) (7).

2.2.3 Preparation and Use of Alkaline Phosphomonoesterase

E. coli alkaline phosphomonoesterase (E.C.3.1.3.1.) grade BAPC, obtained from Worthington Biochemical Corp., was heat treated to remove diesterases (9), and was assayed as described by Garen and Levinthal (9). The assay used p-nitrophenyl phosphate as substrate and micro-molar units were determined. The average 10 mg vial was found to contain 150 units.

Complete dephosphorylation of oligonucleotides was accomplished by diluting the sample into 50 mM Tris-HCl buffer, pH 8.0, adding 1.5 units/ml phosphomonoesterase, and incubating 45 min at 55°C.

Under these conditions there was no detectable phosphodiesterase activity on nanomolar quantities of pyrimidine oligonucleotides.

When removal of phosphomonoesterase activity was necessary, destruction of the enzyme by alkali was found to be the best method. The solution was made 0.5 M in NaOH and incubated at 37°C for 15 mins.(132). Since some enzyme can survive by clinging to the walls of a glass tube (132), the solution was then transferred to another test tube and neutralized for the next step in the procedure.

2.2.4 Preparation of [γ -^{32}P] ATP

Preliminary experiments in this thesis used [γ -^{32}P] ATP prepared by a method using phosphoglycerate kinase and phosphoglyceraldehyde dehydrogenase described by Glynn and Chappell (10). The method was modified in that yeast enzymes were used instead of those from rabbit, 5 μM NADH was included in the reaction, unlabelled phosphate was not added, the volume of the reaction mixture was kept to 0.2 ml, and the amount of ATP initially added was varied between 0.1 and 1 μmole in different experiments. These modifications decreased the yield but increased the specific activity of the product. Yields of 5% to 25% were normal, giving specific activities between 5 mCi/μmole and 45 mCi/μmole.

For the most recent experiments [γ-^{32}P] ATP was prepared by the method of Schendel and Wells (11) using the forward reaction with ADP and phosphoglyceraldehyde instead of the exchange reaction using ATP and phosphoglycerate. The method was modified by using yeast enzymes and including 0.1 mM NAD in the reaction. Yields in this reaction were between 30% and 70% and specific activities were estimated to be between 100 mCi/μmole and 1000 mCi/μmole.

Specific activities of [γ-^{32}P] ATP samples were determined using the polynucleotide kinase reaction and a known concentration of oligonucleotides with a

free 5' hydroxyl group, followed by fractionation using a DEAE cellulose column (Section 2.2.9) or a column of Sephadex G-25.

All measurements of radioactivity were performed in aqueous media (23) in a Beckman LS250 scintillation spectrometer.

2.2.5 Polynucleotide Kinase

2.2.5.1 Preparation of Polynucleotide kinase

Polynucleotide kinase was purified from T4 infected cells by the method of Richardson (12) with a few modifications. Bacteriophage T4n82 was used in our preparations of polynucleotide kinase instead of the $T4^+$ used by Richardson. T4n82 is an amber mutant deficient in DNA synthesis (15). When grown in a normal host, synthesis of early enzymes such as polynucleotide kinase is not shut off and can continue beyond the first 20 minutes (153). In our preparations infection was allowed to proceed two hours and quick cooling was not necessary before collecting the cells.

All buffers in the preparation, until the enzyme was loaded to the first DEAE cellulose column, contained 1 mM ATP. Wu (13) has recently used 0.1 mM ATP in the initial purification steps to stabilize the enzyme. Instead of a stepwise elution from the DEAE cellulose column a linear 2 l. gradient of 10 mM potassium phosphate buffer, pH 7.5, 10 mM mercaptoethanol, to 50 mM potassium phosphate buffer, pH 7.5, 10 mM mercaptoethanol was used. The activity eluted in a broad peak at 0.04 M potassium phosphate. The majority of this peak was loaded to a phosphocellulose column as described by Richardson and eluted with a linear 1 l. gradient of 0 to 0.5 M KCl in 50 mM potassium phosphate buffer, pH 7.5, 10 mM mercaptoethanol. A fairly sharp peak of kinase activity eluted at 0.18 M KCl.

This fraction was found to have traces of endonuclease and exonuclease activity and therefore it was dialysed against 10 mM potassium phosphate buffer pH 7.5, 10 mM mercaptoethanol, and loaded to another DEAE cellulose column. This column was washed extensively (600 ml) with 10 mM potassium phosphate buffer, pH 7.5, 10 mM mercaptoethanol and then eluted with 50 mM potassium phosphate buffer, pH 7.5, 10 mM mercaptoethanol. This additional DEAE column has since been suggested by Wu and Kaiser (13). This column split the polynucleotide kinase into two fractions, the first eluting with the 10 mM wash, and the second eluting with the 50 mM wash. The polynucleotide kinase fractions were then dialysed against 50 mM Tris-HCl buffer, pH 7.5, 10 mM mercaptoethanol, 50% glycerol, and stored at -20°C. Both the 10 mM and the 50 mM fractions of the polynucleotide kinase were tested under standard incubation conditions using uniformly labelled and 5' terminally labelled oligonucleotides as substrates. Fractionations on DEAE cellulose columns showed no detectable phosphomonoesterase, exonuclease, or deaminase activities in the 10 mM fraction but some exonuclease and endonuclease activity remained in the 50 mM fraction. The 10 mM fraction was used in all experiments presented in this thesis.

2.2.5.2 Polynucleotide Kinase Assay

The assay for polynucleotide kinase used as substrate a mixture of dephosphorylated oligonucleotides longer than hexanucleotides. These were prepared by degradation of calf thymus DNA by the Burton depurination procedure (71) (Section 2.2.8) followed by dephosphorylation (Section 2.2.3). They were then loaded to a DEAE cellulose column and oligonucleotides containing less than 7 bases were eluted with 0.18 M sodium chloride. The column was then washed with 100 ml

50 mM TEAB buffer to remove the salt and the oligonucleotides eluted with 1 M TEAB buffer. The buffer was evaporated, the oligonucleotides were dissolved in water and stored at -20°C.

The assay was performed in a volume of 0.1 ml of 10 mM Tris-HCl buffer, pH 8.1, 10 mM $MgCl_2$, 20 mM mercaptoethanol, containing 2 nmoles of oligonucleotides and 0.5 nmoles of [γ-^{32}P]ATP. Incubation was for 30 min at 37°C. A unit was defined as the amount of enzyme which will catalyze the transfer of phosphate at a rate of 1 nmole/min under these assay conditions.

The 50 mM fraction contained a total of 24 units at 2 units/ml; the 10 mM fraction contained a total of 18 units at 0.5 units/ml.

2.2.5.3 Phosphorylation of Oligonucleotides.

Labelling of the 5' termini of oligonucleotides was performed in volumes ranging from 0.05 ml to 1 ml. In all cases the reaction mixture was adjusted to pH 8.1 with 1 M Tris or 1 M HCl and to concentrations of 10 mM in $MgCl_2$, 20 mM in mercaptoethanol, and 0.05 units/ml in polynucleotide kinase. [γ -^{32}P]ATP was usually added in 2 to 10 fold excess. It was observed that polynucleotide kinase is not inhibited by salt concentrations up to 1 M NaCl; thus salt accumulating from other steps, such as the destruction of phosphomonoesterase, was not removed prior to the polynucleotide kinase reaction. Incubation was for 6 to 10 hours at room temperature.

2.2.6 Preparation of Bacteriophage S13$^+$ DNA

2.2.6.1 Preparation of Bacteriophage High Titre Suspensions

Three litres of E. coli C were grown in the Tris-glycerol (TG) medium of

Shleser et al., (16) to a density of 5×10^8 cells/ml (optical density at 600 nm $\simeq$ 0.45). Magnesium chloride was added to a final concentration of 2.5 mM and $S13^+$ was added to a multiplicity of infection of 0.1. When lysis was complete (3 hrs) NaCl was added to 0.5 M and polyethylene glycol to 10% W/V (25) and the culture shaken until the additions were dissolved. The culture was then left overnight at 4°C before centrifuging at 10,000 g for 10 minutes. The pellets were resuspended in 100 ml of 50 mM Tris-HCl buffer, pH 8, 10 mM EDTA. This suspension served as high-titre stock and was stored at 4°C. The titre was normally between 5×10^{11} and 10^{12} pfu/ml.

2.2.6.2 Preparation of ^{32}P-labelled $S13^+$ Replicative Form DNA

Infection of E. coli with $S13^+$ and inhibition of single-strand synthesis with chloramphenicol to accumulate RF DNA was as described by Shleser et al., (16). After collection by centrifugation, the cells were resuspended in 50 mM Tris-HCl buffer, 2 mM EDTA, lysed by treatment first with lysozyme, then with pronase and SDS at 37°C, and the E. coli DNA precipitated with 1 M NaCl as described by Hirt (17). After a phenol extraction the S13 RF DNA was precipitated overnight at -20°C by addition of one-fifth volume of 3 M sodium acetate pH 5.5, and 2 volumes of isopropanol (18). The DNA was further purified by sucrose gradient centrifugation and centrifugation in cesium chloride containing ethidium bromide as described by Schekman et al., (18). Before dialysis ethidium bromide was removed from the sample by isopropanol extraction (111). One 250 ml radioactive culture using 10 mCi of $H_3{}^{32}PO_4$ normally yielded 100 µg of DNA of specific activity 6×10^4 cpm/µg.

2.2.6.3 Preparation of Unlabelled $S13^+$ Replicative Form DNA

Unlabelled replicative form DNA was prepared as described for ^{32}P labelled RF DNA except that the cells were not collected and resuspended in phosphate-free

medium before infection. One litre of culture normally yielded 500 µg of RF DNA.

2.2.6.4 Preparation of ^{32}P-labelled $S13^+$ DNA

E. coli C were grown in Shleser's (16) Tris-glycerol (TG) medium to a cell density of 1×10^9 /ml (optical density 0.8 at 600 mµ). Magnesium chloride was added to a final concentration of 2.5×10^{-3} M and bacteriophage added to a multiplicity of infection of 5. Radioactive phosphate was added five minutes after infection and growth was allowed to continue until lysis (3 hours). The culture was then made 0.5 M in NaCl and 10% W/V in polyethylene glycol and left overnight at 4°C (25). The precipitate was collected by centrifugation at 16,000 g for 10 minutes and dissolved in 0.1 M Tris-HCl buffer, pH 8.1, 50 mM sodium borate, 10 mM EDTA. Lysozyme was added to a concentration of 1 mg/ml and the suspension stirred 10 min at room temperature and then centrifuged 15 min at 16,000 g. The purification from this stage consists of two cycles of high and low speed centrifugation, followed by cesium chloride density gradient centrifugation (152).

Phage bands were collected and dialysed 3 hours against 0.1 M Tris-HCl buffer pH 8, 0.01 M EDTA before release of the DNA by the hot phenol treatment described by Sinsheimer (19). The DNA was then dialysed extensively against water to remove phenol and salts if formic acid digestion was the next step. DNA used for other purposes was concentrated by vacuum dialysis or isopropanol precipitation, dialysed against the appropriate buffer and stored at -20°C.

One litre of culture and 10 mCi of ^{32}P normally yielded 1 mg of DNA of specific activity $\frac{30{,}000 \text{ cpm}}{\mu g}$.

2.2.6.5 Preparation of Unlabeled $S13^+$ DNA

Unlabelled viral DNA was prepared in the same manner as radioactively labelled DNA, except that radioactive label was not added.

2.2.7 Preparation and Investigation of Bacteriophage S13suN15 DNA

The preparation and the determination of the properties of S13suN15 DNA were as described by Spencer et al., (140).

Preparation of the bacteriophage DNA was similar to that of $S13^+$ DNA except that, since S13suN15 is a lysis defective strain, bacteriophage was harvested by collection of the intact cells 3 hours after infection; also, the DNA was purified further by equilibrium cesium chloride density gradient centrifugation in the presence of 0.1% sarkosyl NL97.

Temperature-absorbance profiles, ultraviolet spectra, velocity sedimentation coefficients, buoyant densities in cesium chloride, and base compositions for both S13suN15 and S13suN15 RF DNA were measured as described by Spencer et al., (140). This publication is included in this thesis as Appendix I.

2.2.8 Depurination of DNA

Formic-acid diphenylamine hydrolysis of DNA was by the method of Burton and Petersen (20). Diphenylamine and formic acid were removed as described by Spencer et al., (21).

2.2.9 Fractionation of Oligonucleotides

2.2.9.1 Desalting

Desalting of oligonucleotide solutions was performed by diluting to a salt concentration below 50 mM, absorbing the oligonucleotides to DEAE cellulose, washing the DEAE cellulose with 50 mM TEAB, and eluting the oligonucleotides with 1 M TEAB (51). Small volumes were desalted in a 12 ml conical centrifuge tube; the washings and elution were performed by vortexing the solutions with 0.5 ml of DEAE suspension and centrifuging to separate them. Large volumes were desalted by running the solutions through a 1 cm x 4 cm column of DEAE cellulose.

2.2.9.2 Gel Filtration of Oligonucleotides

Gel filtration on Sephadex G-25 columns was used to separate oligonucleotides of chain length greater than two nucleotides from small molecules such as ATP and inorganic phosphate. These columns were of dimensions 1 cm x 50 cm and were eluted with 50 mM NaCl at a flow rate of 20 ml/hr. The sample size was between 0.5 ml and 2 ml.

In some experiments dinucleotides were separated from ATP by means of gel filtration on a column of Sephadex G-15. Conditions were the same as for Sephadex G-25.

2.2.9.3 Ion Exchange Column Fractionations of Oligonucleotides

Pyrimidine oligodeoxyribonucleotides were fractionated according to chain length on a column of DEAE Sephadex A-25 with dimensions 1 cm x 25 cm. The chloride

form of the ion exchanger was washed with water in the column, the sample was loaded and the column washed with 50 to 100 ml more water. A linear gradient of 0 to 0.4 M NaCl in a total volume of 2 l. of 50 mM sodium acetate buffer, pH 5.5, 1 mM phosphate, 7 M urea was used to elute the oligonucleotides. All operations were conducted at a flow rate of 45 ml/hr. and 12 ml fractions were collected.

It has been found that the inclusion of 1 mM phosphate in the gradient buffers improves the resolution in fractionations of small amounts of material (less than 5 mg nucleotide). Chain length fractionation of oligonucleotides containing all four bases was carried out on a 1 cm x 90 cm DEAE sephadex column at 65°C as described by Junowicz and Spencer (22).

Pyrimidine oligonucleotide samples were fractionated according to base composition on columns of DEAE sephadex of dimensions 1 cm x 25 cm. Loading and washing were as described for the chain length fractionation, but a linear gradient of 0 to 1 M ammonium formate, pH 3.2, was used to elute the fractions with different base compositions. Unlike DEAE cellulose columns (152), these columns were found to be capable of resolving even picomolar amounts of nucleotide.

5' phosphorylated dinucleoside diphosphates such as those generated by DNAase I were fractionated at pH 3.4 on DEAE cellulose as described by Junowicz and Spencer (22). All the possible dinucleotides are well separated on this column and the sequence isomers pApC and pCpA are resolved.

Longer oligonucleotides containing all four bases were fractionated according to base composition on a 1 cm x 25 cm column of DEAE sephadex eluted with a gradient from 0 to 1.0 M ammonium formate, pH 3.5, total volume 2 l. Only species with very divergent base compositions can be separated in this manner.

2.2.9.4 Electrophoresis - Homochromatography

Fractionation of oligonucleotide samples according to chain length and base composition was performed using a two-dimensional system. The first dimension was electrophoresis on a cellulose acetate strip at pH 3.5. The second dimension was chromatography on DEAE TLC plates using an alkali digest of RNA as chromatographic buffer. This is the electrophoresis-homochromatography system of Brownlee and Sanger (28), described in full by Harbers et al., (174). The migration of the radioactive oligonucleotides on the plate was observed by autoradiography of the DEAE-cellulose thin layer plate.

The electrophoresis at pH 3.5 fractionates the oligonucleotide sample according to base composition; C-rich oligonucleotides migrate closer to the origin than do T-rich oligonucleotides of the same length. The thin layer chromatography fractionates the oligonucleotides according to chain length; longer oligonucleotides migrate closer to the origin than shorter ones.

In the electrophoresis-homochromatography system pyrimidine oligonucleotides migrate in such a way that a reproducible grid predicting the base composition of the components of the sample can be constructed. Figure 2.1 shows two such grids, the first constructed from the migration pattern of a mixture of dephosphorylated oligonucleotides, the second a mixture of 5' phosphorylated pyrimidine oligonucleotides.

From these diagrams it can be seen that the relative base composition of a series of progressively shorter pyrimidine oligonucleotides such as those resulting from a partial exonuclease digestion can easily be determined using the electrophoresis-homochromatography system. An oligonucleotide containing one less cytidylate residue than its

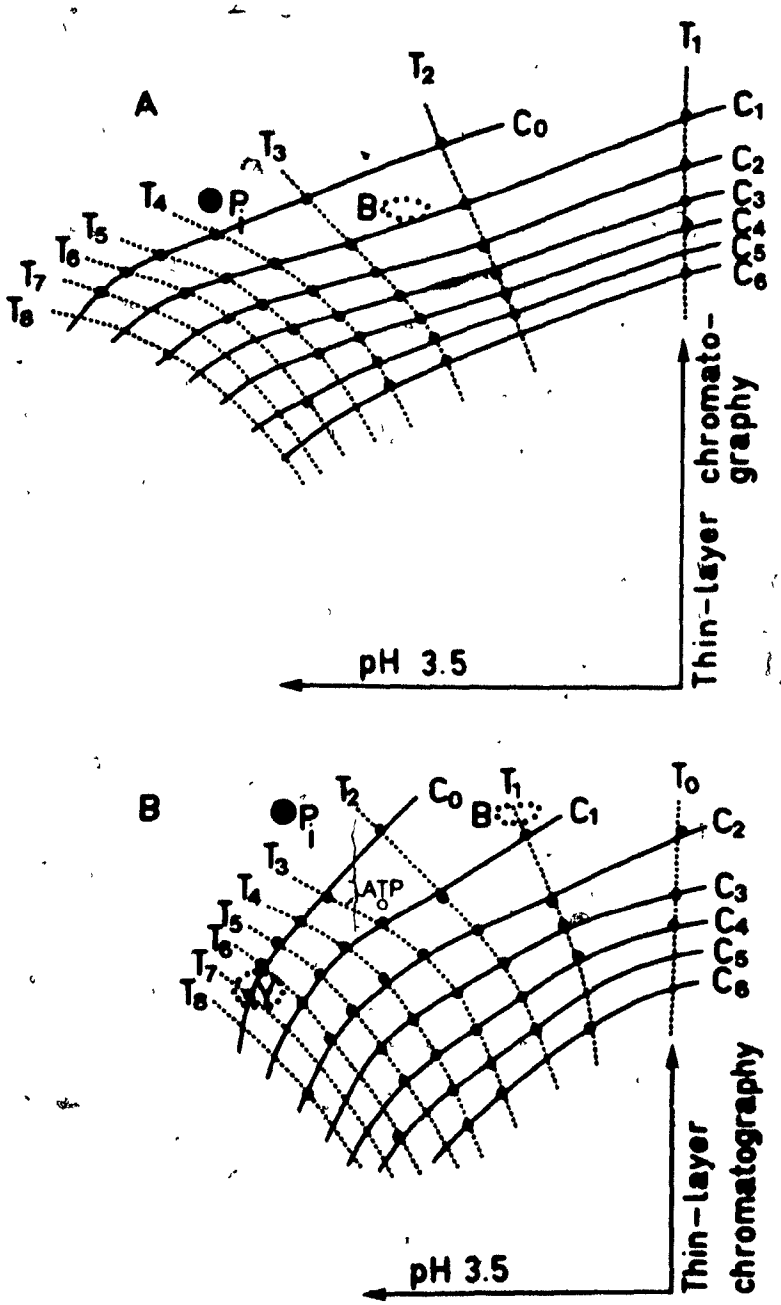


Figure 2.1: Migration of pyrimidine oligonucleotides in the electrophoresis homochromatography system

The first dimension (right to left) consisted of electrophoresis on cellulose acetate in pyridine acetate buffer, pH 3.5, 7 M urea. The contents of the strip were transferred to the base of a 20 cm x 20 cm DEAE cellulose TLC plate by elution with water, the plate dried, and thin layer chromatography carried out in the vertical direction using an RNA alkali hydrolysate in 7 M urea as eluent. Autoradiography was performed by clamping a sheet of X-ray film between the TLC plate and a board of the same dimensions and storing in a light-proof cupboard.

The intersection of the solid lines and the dashed lines indicates the migration positions of the oligonucleotide of composition shown at the top of each line; B indicates the migration position of the blue marker, Y indicates the migration position of the yellow marker.

(A) Migration of dephosphorylated pyrimidine oligonucleotides (Py_nP_{n-1})

(B) Migration of 5'-phosphorylated pyrimidine oligonucleotides (Py_nP_n)

parent oligonucleotide will migrate farther from the origin in both dimensions. One containing one less thymidylate residue will migrate closer to the origin in the electrophoresis step than its longer parent. The "zig-zag" pattern of the fractionation of a partial digest of a pyrimidine oligonucleotide yields sequence information about the parent oligonucleotide.

Oligonucleotide material was eluted from a particular region of the TLC plate by scraping the DEAE cellulose from that area, collecting it above a cotton filter plug in a pasteur pipette by suction, and washing the filter with 1 M TEAB buffer. The eluate was then evaporated to dryness to remove the TEAB.

2.2.9.5 Mononucleotide Analysis

The four 5' deoxyribonucleoside monophosphates were separated from each other on a 1 cm x 20 cm column of Dowex 50 Wx4 eluted with 0.15 M ammonium formate buffer, pH 3.2 (I) at a flow rate of 6 ml/min. Under these conditions eight base composition analyses could be performed per column per hour.

2.2.10 Sequence of Oligodeoxyribonucleotides

2.2.10.1 Uniformly Labelled Oligonucleotides

Pyrimidine oligonucleotides labelled uniformly with ^{32}P were dephosphorylated as described in Section 2.2.3. The sequence was then determined as described by Ling (26). This involved dividing the sample into two parts, digesting one with spleen exonuclease, the other with venom exonuclease, and fractionating both digests using electrophoresis-homochromatography.

The length and base composition of the digest products are determined from their position on the autoradiograph of the TLC plate; the two digests give enough information for elucidation of the complete sequence.

A diagrammatic representation of the sequence of a hypothetical pentanucleotide is given in Figure 1.3.

2.2.10.2 5' Terminal Labelling of Oligonucleotides Followed by Exonuclease Treatment

The sequence of a pyrimidine oligonucleotide which had been phosphorylated at the 5' terminus using polynucleotide kinase and [γ -^{32}P] ATP (2.2.5.3) was determined by partial digestion with venom exonuclease (2.2.1) followed by fractionation using electrophoresis-homochromatography (2.2.9.4). The sequence of the oligonucleotide was deduced from the position of each of the exonuclease derived fragments on the TLC plates, as described by Ling (26) (2.2.9.4). A diagrammatic representation of the sequence of a hypothetical pentanucleotide is given in Figure 1.4.

2.2.10.3 Exonuclease Treatment of an Oligonucleotide Followed by 5' Terminal Labelling

Most of the oligonucleotide sequences determined in this study were the result of the development of a method involving partial digestion with spleen exonuclease, labelling of the 5' termini with polynucleotide kinase and [γ -^{32}P] ATP, fractionating the resulting labelled mixture, and identifying the separated products. Experimental procedure was as follows: the oligonucleotide was dephosphorylated (2.2.3) in a volume of 50 µl and subjected to spleen exonuclease partial digestion (2.2.2). The pH of the solution was adjusted to approximately 8 (Merck Dormstadt non-bleeding pH sticks) and the phosphomonoesterase digestion repeated (2.2.3). The phosphomonoesterase was inactivated by treatment with alkali, the pH readjusted to 8.1 with HCl (2.2.3), and the oligonucleotides in the digest phosphorylated with polynucleotide kinase and [γ-^{32}P] ATP (2.2.5). After the kinase incubation 1 µmole of ATP was added as marker and the mixture fractionated on Sephadex G-25 (2.2.9.2).

a) 5' sequence analysis on columns

The major portion of the sequence of most unique oligonucleotides was determined by analysis of the larger oligonucleotide digest products eluting before ATP from the G-25 column (2.2.9.2). This fraction contained all oligonucleotides of chain length 4 or greater as well as much of the trinucleotide .

These were loaded directly to a DEAE Sephadex column and fractionated according to chain length (2.2.9.3). Individual chain length fractions were desalted (2.2.9.1), dissolved in 1 ml , and an aliquot (0.1 ml to 0.5 ml) subjected to complete venom exonuclease digestion (2.2.1). The mononucleotide composition of these digests was determined using Dowex 50 chromatography at pH 3.2 (2.2.9.5).

The sequence of an oligonucleotide from the 5' terminus down to the third position from the 3' terminus was determined in this way.

b) 5' sequence analysis using electrophoresis-homochromatography

The procedure here was the same as that described in the previous section except that the chain length fractionation was performed using two-dimensional electrophoresis-homochromatography (2.2.9.4). This system has the advantage that partial sequence information is obtained from the migration of the digest products in this system before the mononucleotide analyses are completed.

c) 3' terminal dinucleotide analysis

The fractions containing the 3' terminal dinucleotides, eluting with the first half of the ATP peak from the G-25 column, were pooled when it was necessary that these dinucleotides be identified.

Fifty mg. of oligonucleotide marker (2.1) was added to this fraction before

it was fractionated on a DEAE sephadex chain length column eluted with a buffer at pH 5.5 (2.2.9.3). This column separates the dinucleotides from inorganic phosphate [γ -^{32}P] ATP, and from some polynucleotide kinase reaction byproducts. A diagrammatic representation of the entire sequence scheme is shown in Figure 1.6.

d) 3' terminal sequence analysis

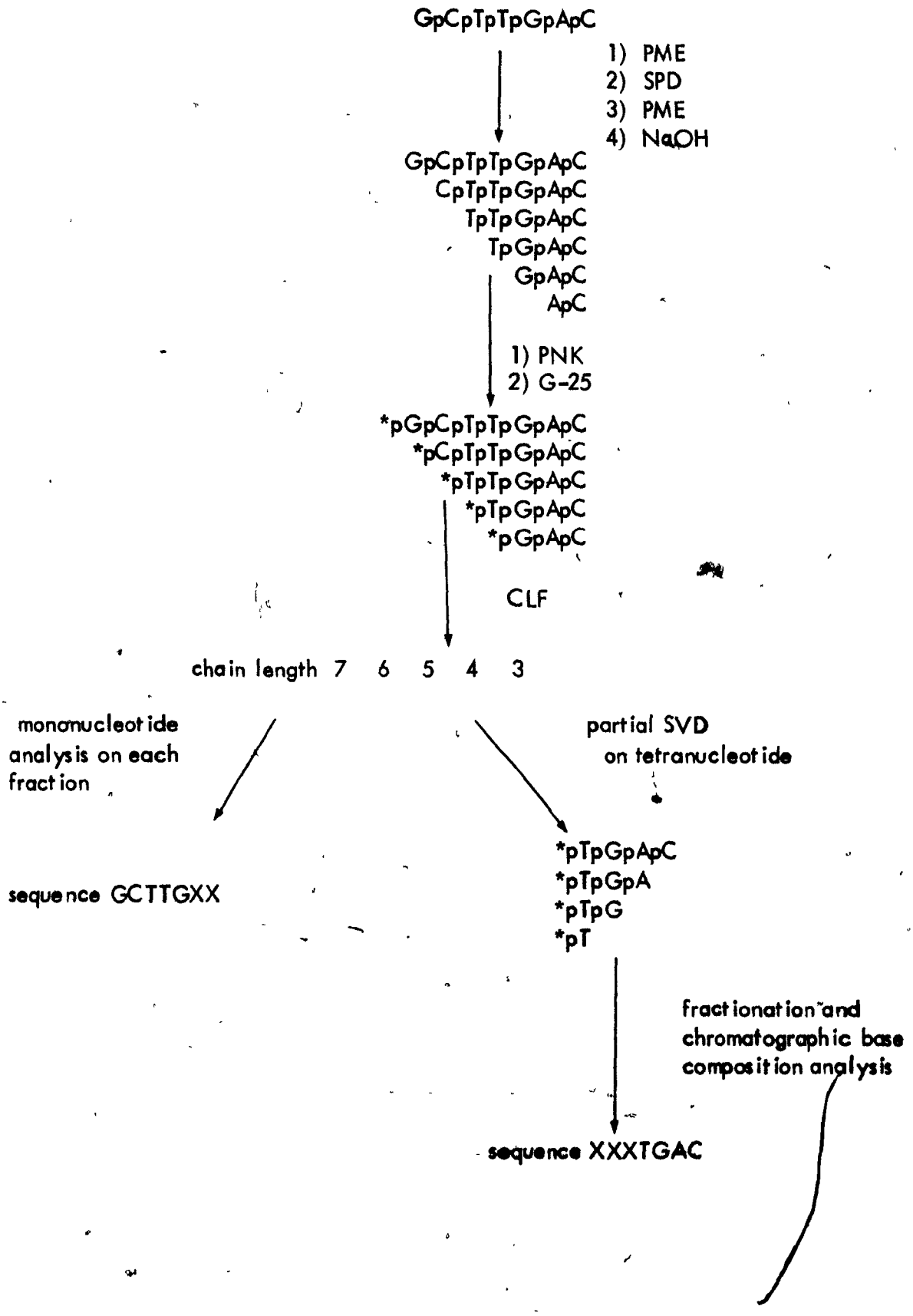
An alternative method for determining the 3' terminal nucleotide sequence of an oligonucleotide was used in some experiments. The dinucleotide region from the G-25 column was not pooled in these cases. Instead, one of the chain length fractions from the larger oligonucleotide digest products (2.2.10.3c) was sequenced using partial degradation with venom exonuclease (2.2.10.2). For pyrimidine oligonucleotides any of the labelled 5' exonuclease digest products can be used as the target for 3' sequence analysis. For oligonucleotides containing all four bases the 3' sequence analysis must be performed on the trinucleotide or tetranucleotide digest products.

A diagrammatic representation of the entire sequence scheme using this alternative for analysis of the 3' end is shown in Figure 2.2.

This approach was particularly useful for pyrimidine oligonucleotides containing isomers of similar base composition but different sequence.

Figure 2.2

Sequence of a Hypothetical Oligonucleotide



CHAPTER 3

RESULTS

3.1 Properties of Bacteriophage S13 suN15 DNA

Bacteriophages S13 and ØX174 have been reported to be genetically and immunologically closely related (121, 122) but, until 1972, the physical properties of the two bacteriophages had not been shown to be identical. Before investigating the primary structure of the S13 DNA genome, a study of the characteristics of the viral and replicative forms of S13suN15 DNA was performed in co-operation with Dr. R .Cerny and Dr. E. Cerna (140).

Results of determinations of base compositions, the relation between temperature and absorbance, and reactivity of the DNA with formaldehyde suggested that the viral DNA was single stranded. Electron micrography (performed by M. Fiandt and W. Szybalski, University of Wisconsin) demonstrated that the DNA was a circular genome; length measurements from these electron micrographs gave a molecular weight of 1.8×10^6 daltons.

Similar studies on S13suN15 replicative form DNA suggested that it was a double stranded closed circular molecule with a molecular weight of 3.5×10^6 daltons.

Velocity sedimentation analysis indicated that the S°_{20} in SSC was 24.6 for the viral DNA, and 20.7 for the closed circular replicative form. Equilibrium cesium chloride centrifugation indicated that the buoyant density of the viral DNA was 1.726 g/cc, and that of the RF DNA 1.710 g/cc.

A detailed description of this study is included in this thesis as Appendix I (140).

3.2 Properties of Spleen Exonuclease

3.2.1 Sequence Preferences of Spleen Exonuclease

In the course of this study, the susceptibility of different pyrimidine sequences to spleen exonuclease has been examined. When the concentration of oligonucleotide was very low (1 μM), spleen exonuclease did not show any significant differences in the rate of hydrolysis of individual oligonucleotides. This contrasts with the report (8) that spleen exonuclease degrades polyrC at a much lower rate than other synthetic ribopolymers .

When this exonuclease was used to degrade a radioactive oligonucleotide in the presence of unlabelled RNA oligonucleotides (30 mM) the apparent rate of degradation decreased to a different degree for individual oligonucleotides. Oligonucleotides C_6T_3 and C_5T_6, with 5' terminal sequences of TCTT ..... and CTTCC ..... were the only pyrimidine oligonucleotides from $S13^+$ DNA which were refractory to spleen exonuclease digestion under normal digest conditions (2.2.2).

3.2.2 Enzymatic Purity of Spleen Exonuclease

The sequence studies presented in this thesis are based on partial digestions using spleen and venom exonucleases. Commercially available spleen exonuclease was found to contain serious cytosine deaminase and phosphomonoesterase activities and this enzyme was therefore prepared from hog spleen by the method of Bernardi and Bernardi (8). This preparation was thoroughly examined for contaminating activities.

Except for spleen exonuclease the enzymes used in this study for the determination of DNA sequence have optimum activity between pH 8 and 9. It would have been advantageous to perform the spleen exonuclease partial digests at alkaline pH instead of at pH 5.5 in sodium succinate buffer (8). In Tris-HCl buffer, pH 8, twenty times the concentration of exonuclease had to be used for longer time periods to obtain the same level of digestion of a dephosphorylated oligonucleotide as at acid pH. A chain length fractionation of [5'-^{32}P] C_6T_4 treated with spleen exonuclease (Figure 3.1) at alkaline pH was performed to check for endonuclease or 3' exonuclease activities. It was observed that this spleen exonuclease had an endonuclease activity at pH 8.0 which degraded 45% of the starting material. The relatively low yield of octa- and nonanucleotides indicated that the extraneous activity was not an exonuclease.

Incubation of [5'-^{32}P] C_6T_4 with a twenty fold excess of spleen exonuclease in sodium succinate buffer at pH 5.5 followed by fractionation of an aliquot on DEAE cellulose at pH 3.0 did not reveal the production of any new decanucleotides by deamination of [5'-^{32}P] C_6T_4, demonstrating that oligonucleotide cytosine deaminase activity was not significant. Fractionation of another aliquot on DEAE cellulose at pH 5.5 did not reveal any products which would have resulted from endonuclease or 3' exonuclease activities.

A chain length fractionation of pyrimidine heptanucleotides (Py_7p_8) after treatment with spleen exonuclease at pH 5.5 (Figure 3.2) showed no degradation products between the mononucleotide and the heptanucleotide positions. This verified that endonuclease and nonspecific phosphomonoesterase activities were absent. On the basis of these results, all spleen exonuclease digests were performed at pH 5.5 in sodium succinate buffer.

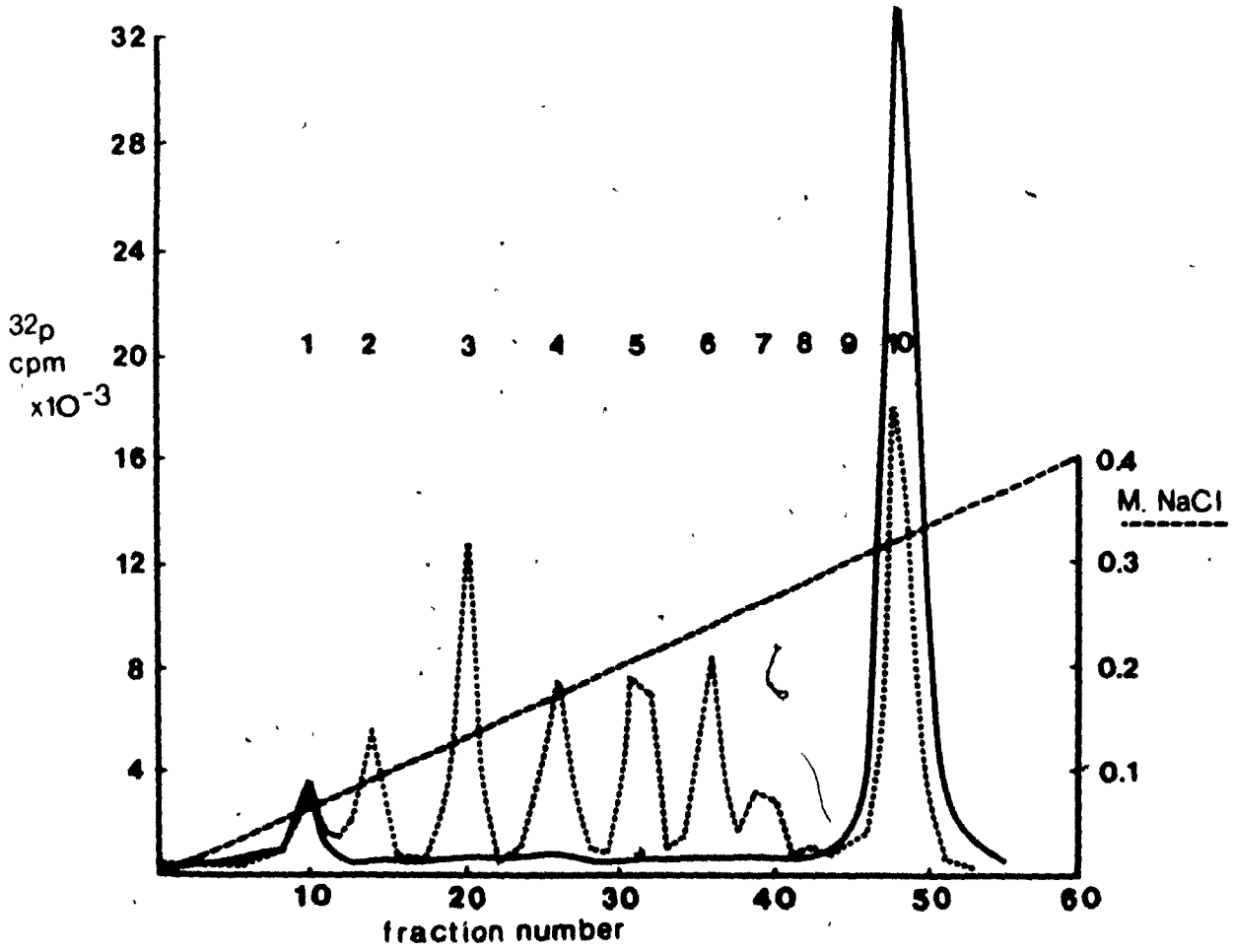


Figure 3.1 : Chain length fractionations of $(5'-^{32}P)C_6T_4$ after treatment with spleen exonuclease

Incubation was for 3 hours at 37°C in a total volume of 0.2 ml containing 100 pmoles of oligonucleotide and 0.3 units of spleen exonuclease. The solid line represents the fractionation of an incubation kept at pH 5.5 with 0.05 M Tris-HCl buffer ; the dotted line represents the fractionation of an incubation kept at pH 8.0 with 0.05 M sodium succinate buffer. The mixture was diluted to 5 ml with water before loading to a 1 cm x 25 cm DEAE cellulose column ; the wash solution was water, the eluate was 1 l of 0.05 M sodium acetate buffer, pH 5.5, 1 mM KH_2PO_4, 7 M urea, with a gradient of 0 to 0.4 M NaCl ; flow rate was 0.5 ml/min, and 16 ml fractions were collected.

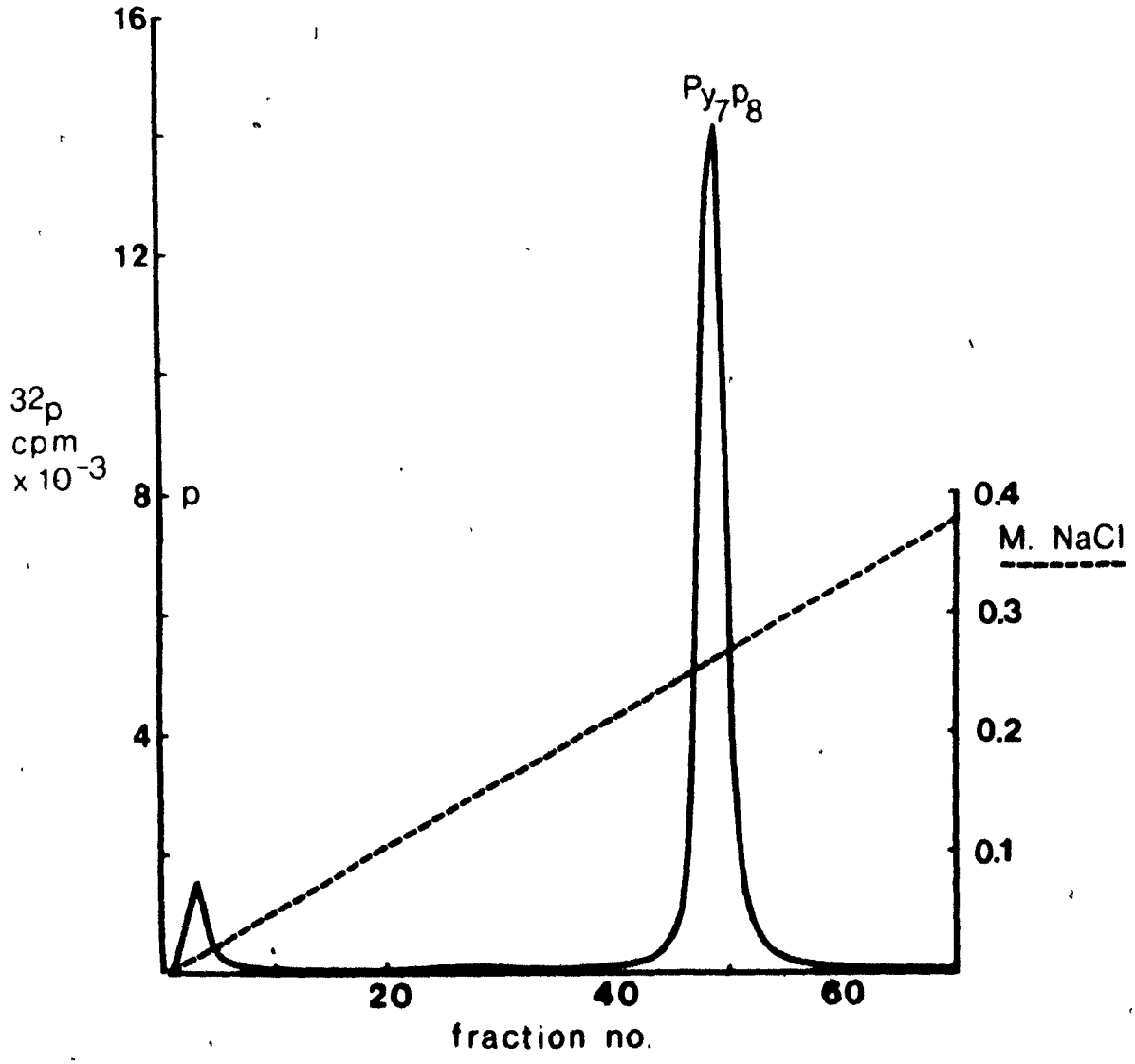


Figure 3.2 : Chain length fractionation of pyrimidine heptanucleotide $(pPy)_7p$ after treatment with spleen exonuclease.

The incubation contained 300 pmoles of heptanucleotide and was performed at pH 5.5 as described for Figure 3.1. Loading and elution were performed as described for Figure 3.1, except that 12 ml fractions were collected.

In both of the control experiments discussed above some mononucleotide and inorganic phosphate were produced ; in more extensive digests, 5' phosphorylated material could be completely degraded to monophosphates. It was calculated that this spleen exonuclease degraded 5' phosphorylated oligonucleotides at a rate 2000 times lower than 5' hydroxyl terminated oligonucleotides. It has been reported (34) that dinucleotide pTpT is degraded by spleen exonuclease at a rate 1000 times less than is TpT. That intermediate length oligonucleotides were absent in fractionations of such digests suggested that removal of the 5' terminal phosphate was the rate limiting step in these reactions.

3.3 Properties of Venom Exonuclease

Commercial venom exonuclease, after treatment (5) at pH 3.6 at 37°C for 3 hours (2.1), was found to be substantially free of contaminating activities. After complete digestion of [5'-^{32}P] C_6T_4 by venom exonuclease, 97% of the radioactivity eluted with cytidylic acid from a Dowex 50 WX4 column (Figure 3.3) indicating that this preparation is free of nucleotidase and phosphomonoesterase. Partial digestions of uniformly labelled oligonucleotides (3.4) did not contain extraneous digestion products, indicating that endonuclease and 5' exonuclease were not significant.

3.4 Sequence Analysis of Uniformly ^{32}P-Labelled Oligonucleotides Isolated from S13$^+$ DNA

Pyrimidine oligonucleotides were isolated (2.8) from 500 μg of ^{32}P-labelled S13$^+$ DNA of specific activity 5×10^4 cpm/μg (2.6.4). These oligonucleo-

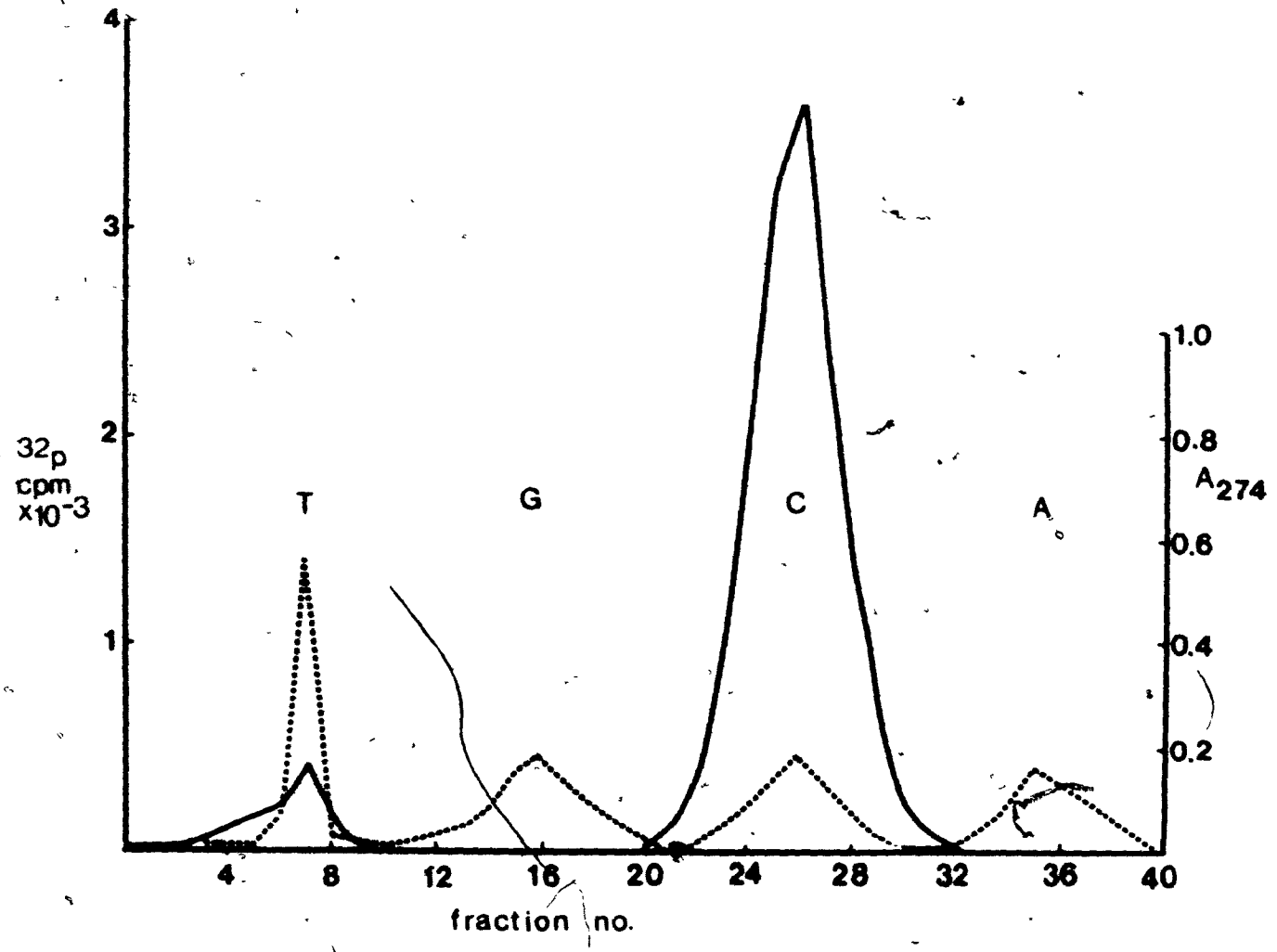


Figure 3.3 : Mononucleotide analysis of a complete venom exonuclease digest of [5'-^{32}P]C_6T_4.

Incubation was for 4 hours at 37°C in a total volume of 0.2 ml of 50 mM Tris-HCl buffer pH 9.0, 10 mM $MgCl_2$, containing 0.2 μmoles of each deoxynucleoside monophosphate, 20 pmoles (20,000 cpm) of [5' ^{32}P]C_6T_4, and 4.8 units of venom exonuclease. The digest was layered at the top of a 1 cm x 25 cm Dowex 50Wx4 column equilibrated with 0.15 M ammonium formate buffer pH 3.2 and was eluted with the same buffer. The solid line represents the elution of ^{32}P radioactivity, the dotted line the elution of the optical density of the mononucleotide markers.

tides were dephosphorylated (2.2.3) and most of the molecules containing less than five phosphates were removed by loading the mixture to a desalting column (1 x 3 cm, DEAE cellulose, 200 mesh) and eluting with 100 ml of 0.3 M TEAB. The remaining oligonucleotides were eluted with 1.0 M TEAB, evaporated to dryness, and fractionated using electrophoresis-homochromatography (2.2.9.4). The autoradiograph of this fractionation (Figure 3.4) showed the positions of migration of the individual components, identified according to the grid (Figure 2.1(A)) predicting the migration of dephosphorylated pyrimidine oligonucleotides in this system.

The oligonucleotides C_6T_4, C_2T_8, and C_5T_6 were eluted from the TLC plate (2.2.9.4), evaporated to dryness, dissolved in water, and counted. A similar experiment using 1.7 mg of $S13^+$ DNA of specific activity 3×10^4 cpm/μg yielded oligonucleotides C_6T, C_5T_2, CT_6, C_5T_3, C_4T_4, C_2T_6, and C_6T_3. Overall recovery of oligonucleotides purified in this manner is indicated in Table 3.1.

Each of these oligonucleotides was analyzed for sequence as was described in Section 2.2.10.1. Each fraction was divided into two equal portions; one of these was partially digested with spleen exonuclease (2.2.2), and the other with venom exonuclease (2.2.1). Each digest was then fractionated using electrophoresis - homochromatography (2.2.9.4) and the migration pattern of the digest products observed by autoradiography. The two autoradiographs obtained for each sample yielded complete sequence information for most of these oligonucleotides.

3.4.1 Oligonucleotide C_6T

The autoradiograph from the spleen exonuclease partial digest of C_6T

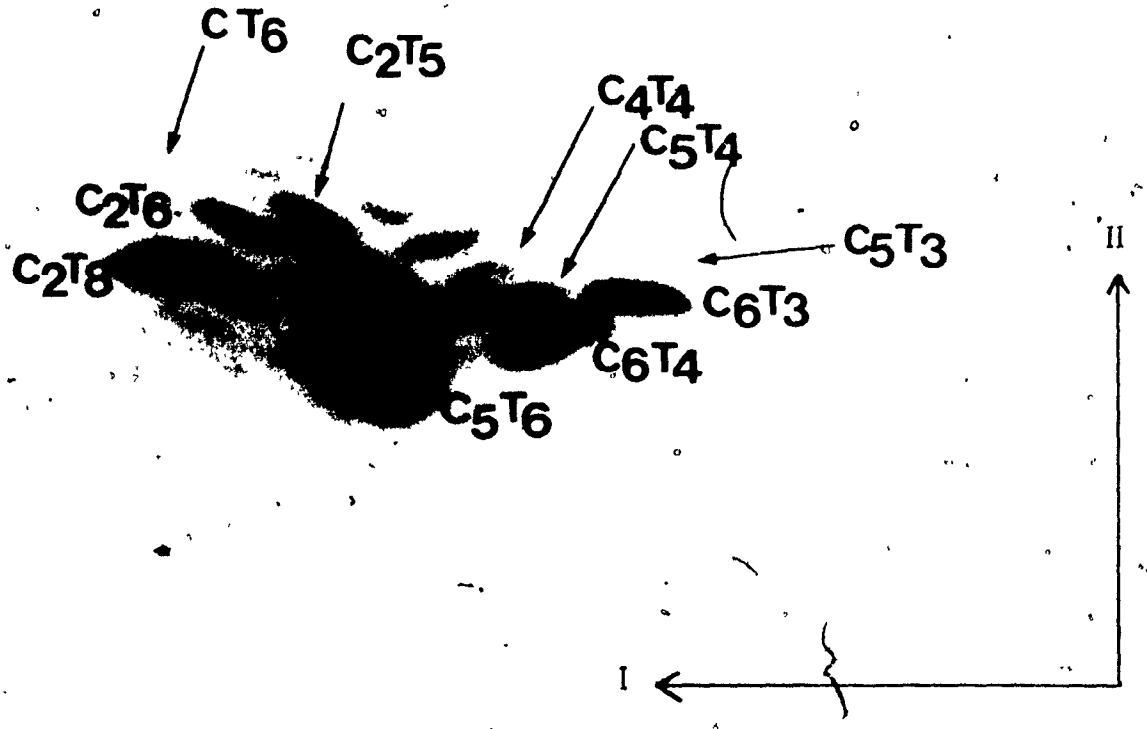


Figure 3.4 : Autoradiograph of the fractionation by electrophoresis-homochromatography of dephosphorylated pyrimidine clusters longer than tetranucleotides from $S13^+$ DNA.

The origin of this fractionation (28) was in the lower right corner. The first dimension (I), from right to left, consisted of electrophoresis on cellulose acetate in a pH 3.5 buffer containing 1% pyridine (v/v), 4.5% acetic acid (v/v), 0.5% formic acid (v/v), and 7 M urea ; electrophoresis was for 20 minutes at 4,500 volts. The cellulose acetate strip was then placed face down at the base of a 20 cm x 20 cm DEAE cellulose TLC plate and the sample transferred to the plate by placing wet strips of filter paper above the strip for 30 minutes and then the plates were dried. The second dimension (II) , in the vertical direction, consisted of thin layer chromatography for 3 hours at 60°C , using an RNA alkali digest dissolved in 7 M urea as chromatographic eluent.

Autoradiography was performed by clamping a sheet of X-ray film between the TLC plate and a wooden board the same size and storing in a light-proof cupboard.

7×10^5 cpm were used in this fractionation ; autoradiography exposure time was 4 hours.

Table 3.1

Elution of oligonucleotides from DEAE cellulose TLC plates

Isostich	Component	Recovery (cpm) Expected	Found	Yield
7	C_6T	55,600	12,000	22 %
	C_5T_2	55,600	16,000	34 %
	CT_6	55,600	20,000	36 %
8	C_5T_3	64,800	20,000	31 %
	C_4T_4	64,800	25,000	39 %
	C_2T_6	64,800	33,000	51 %
9	C_6T_3	74,000	20,000	27 %
10	C_6T_4	40,700	3,900	9.6%
	C_2T_8	40,700	5,700	14 %
11	C_5T_6	45,200	7,000	15.5%

(Figure 3.5 A) contained spots representing undigested C_6T and the first two digest products. The first migrated above and slightly to the left of the parent oligonucleotide indicating that it contained one less cytidylate than C_6T. The second migrated above and to the right of the C_5T, indicating that it was composed of one less T. Thus the digest from the 5' terminus proceeded as follows :

$$C_6T \rightarrow C_5T \rightarrow C_5$$

indicating that the sequence was CTCCCCC.

The autoradiograph from the venom exonuclease partial digest of C_6T (Figure 3.5(B)) contained spots representing undigested C_6T and five partial digest products. Each of these products had a base composition with one less cytidylate than its parent ; thus the digest from the 3' terminus proceeded :

$$C_6T \rightarrow C_5T \rightarrow C_4T \rightarrow C_3T \rightarrow C_2T \rightarrow CT$$

verifying that the 3' partial sequence was (CT)CCCCC.

3.4.2 Oligonucleotide C_5T_2

The autoradiograph from the spleen digest of C_5T_2 (Figure 3.6(A)) contained spots representing C_5T_2 and four of its digest products. The minor spots to the left of C_5T_2 and C_4T_2 could not have arisen from C_5T_2 and are probably due to contamination from adjacent oligonucleotides in the separation of the pyrimidine clusters. The digest from the 5' end proceeded :

$$C_5T_2 \rightarrow C_4T_2 \rightarrow C_4T \rightarrow C_3T \rightarrow C_3$$

giving the sequence CTCTCCC.

The autoradiograph from the venom exonuclease partial digest of this oligonucleotide (Figure 3.6(B)) contained spots representing C_5T_2 and four digest

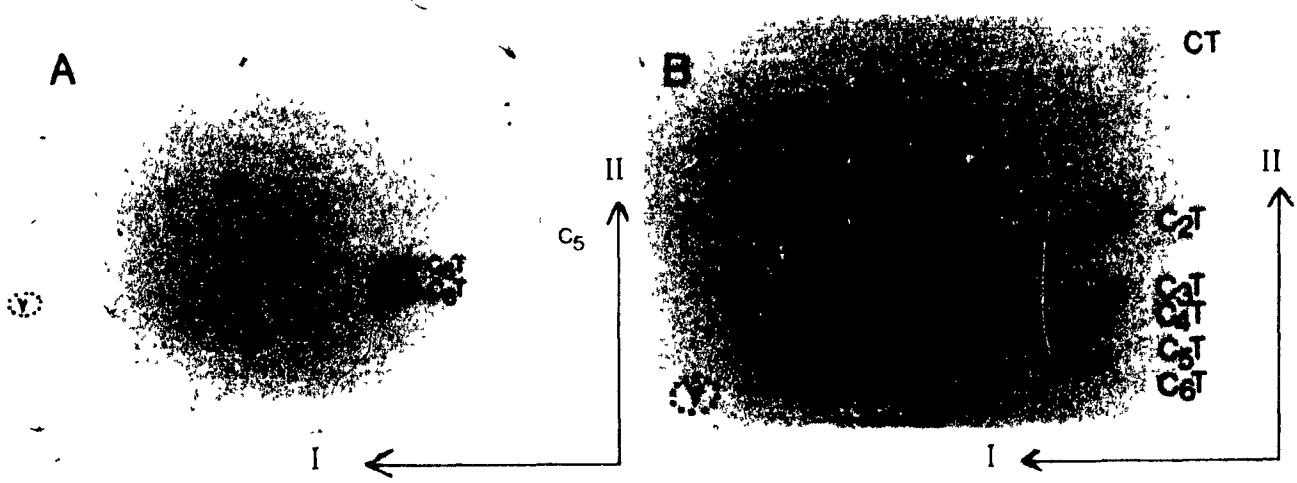


Figure 3.5 ; Autoradiographs of the fractionations by electrophoresis-homochromatography of exonuclease partial digestions of uniformly labelled oligonucleotide C_6T. (A) spleen exonuclease ; (B) venom exonuclease. Blue and yellow markers are represented by "B" and "Y".
Legend as for Figure 3.4 . 5,400 cpm were used for each fractionation ; autoradiography exposure time was 16 hours.

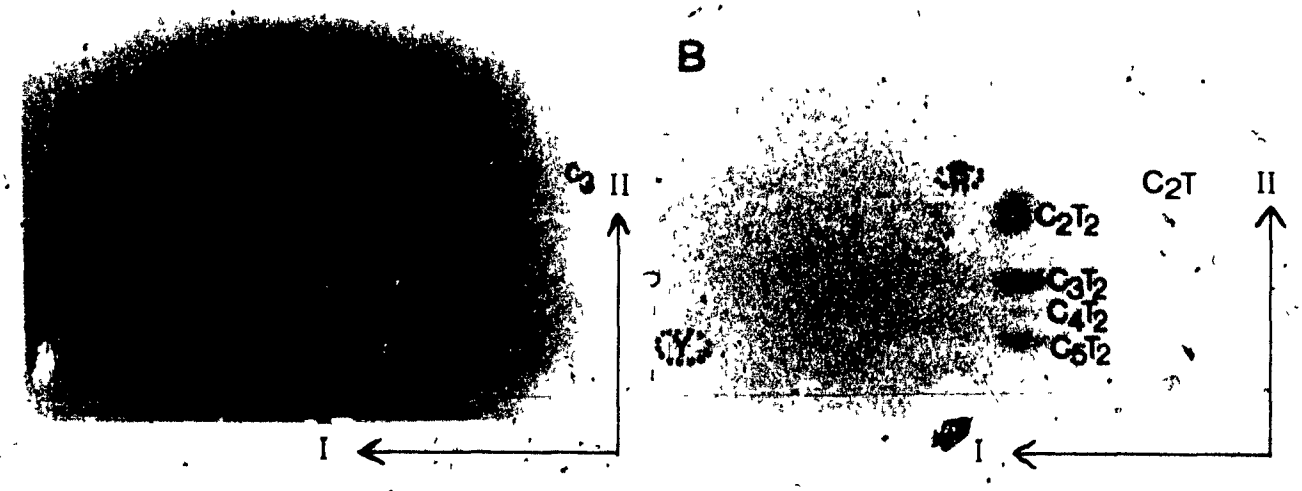


Figure 3.6 : Autoradiographs of the fractionations by electrophoresis-homochromatography of exonuclease partial digestions of uniformly labelled oligonucleotide C_5T_2. (A) spleen exonuclease ; (B) venom exonuclease. Blue and yellow markers are represented by "B" and "Y".
Legend as for Figure 3.4 . 6,700 cpm were used for each fractionation ; autoradiography exposure time was 16 hours.

products, indicating that the digest from the 3' end proceeded :

$$C_5T_2 \rightarrow C_4T_2 \rightarrow C_3T_2 \rightarrow C_2T_2 \rightarrow C_2T$$

verifying the 3' partial sequence $(C_2T)TCCC$.

3.4.3 Oligonucleotide CT_6

The autoradiograph from the spleen exonuclease digest of oligonucleotide CT_6 (Figure 3.7(A)) contained spots representing CT_6 and four digest products. These showed that the digest from the 5' end proceeded :

$$CT_6 \rightarrow T_6 \rightarrow T_5 \rightarrow T_4 \rightarrow T_3$$

and that the sequence was CTTTTTT.

The autoradiograph from the venom exonuclease digest of CT_6 (Figure 3.7(B)) contained spots representing CT_6 and four digest products. Thus the digest from the 3' end proceeded :

$$CT_6 \rightarrow CT_5 \rightarrow CT_4 \rightarrow CT_3 \rightarrow CT_2$$

verifying the 3' terminal sequence $(CT_2)TTTT$

3.4.4 Oligonucleotide C_5T_3

The autoradiograph from the spleen exonuclease digest of C_5T_3 (Figure 3.8(A)) contained spots representing C_5T_3 and three digest products. These showed that the digestion from the 5' end proceeded :

$$C_5T_3 \rightarrow C_4T_3 \rightarrow C_3T_3 \rightarrow C_3T_2$$

giving a 5' partial sequence CCT (C_3T_2).

The autoradiograph from the venom exonuclease partial digest of C_5T_3 (Figure 3.8(B)) contained spots representing C_5T_3 and five digest products. These

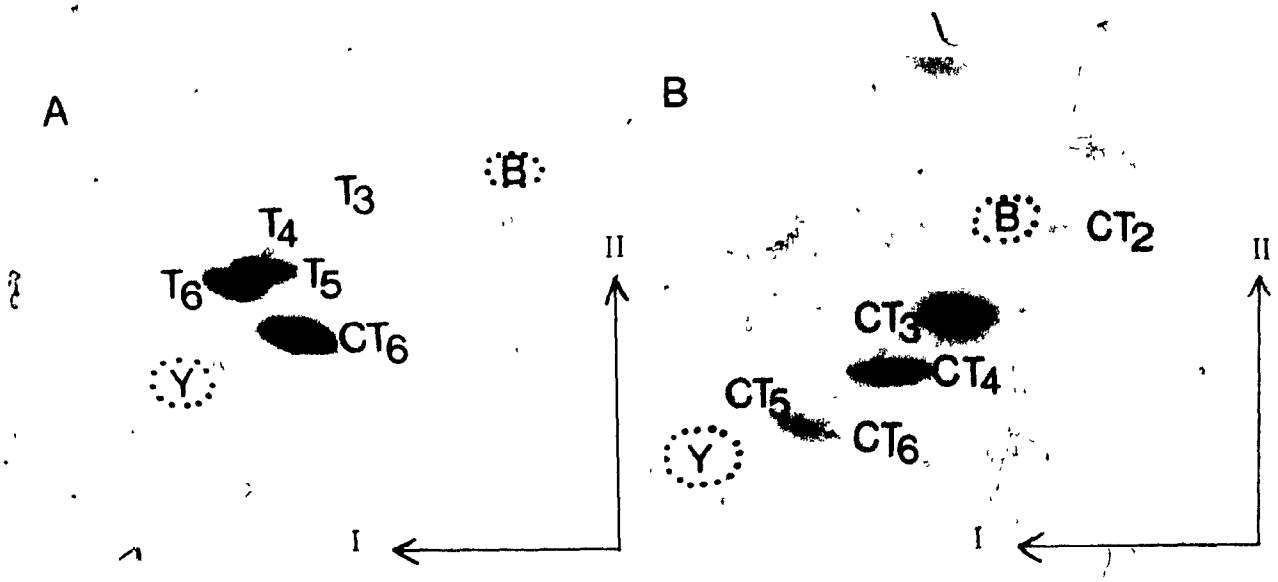


Figure 3.7 : Autoradiographs of the fractionations by electrophoresis-homochromatography of exonuclease partial digestions of uniformly labelled oligonucleotide CT_6. (A) spleen exonuclease ; (B) venom exonuclease. Blue and yellow markers are represented by "B" and "Y".
Legend as for Figure 3.4. 7,200 cpm were used for each fractionation ; autoradiography exposure time was 16 hours.

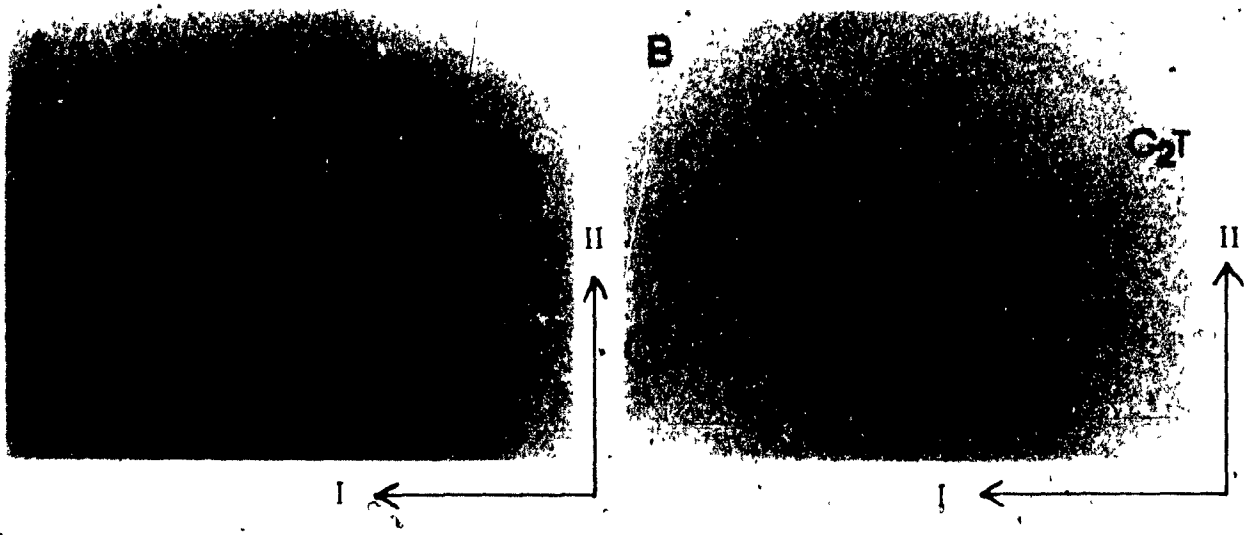


Figure 3.8 : Autoradiographs of the fractionations by electrophoresis-homochromatography of exonuclease partial digestions of uniformly labelled oligonucleotide C_5T_3. (A) spleen exonuclease ; (B) venom exonuclease. Blue and yellow markers are represented by "B" and "Y".
Legend as for Figure 3.4. 7,000 cpm were used for each fractionation ; autoradiography exposure time was 16 hours.

showed that the digest from the 3' end proceeded :

$$C_5T_3 \rightarrow C_4T_3 \rightarrow C_3T_3 \rightarrow C_2T_3 \rightarrow C_2T_2 \rightarrow C_2T$$

yielding the 3' partial sequence (C_2T) TTCCC.

The combination of these two partial sequences gave the total sequence CCTTTCCC.

3.4.5 Oligonucleotide C_4T_4

The autoradiograph from the spleen exonuclease digest of C_4T_4 (Figure 3.9(A)) contained spots representing C_4T_4 and four digest products. These indicated that the digests from the 5' end proceeded :

$$C_4T_4 \rightarrow C_3T_4 \rightarrow C_3T_3 \rightarrow C_3T_2 \rightarrow C_2T_2$$

giving the 5' partial sequence CTTC (C_2T_2).

The autoradiograph from the venom exonuclease digest of C_4T_4 (Figure 3.9(B)) contained spots representing C_4T_4 and three digest products. These showed that the digest from the 3' end proceeded :

$$C_4T_4 \rightarrow C_3T_4 \rightarrow C_3T_3 \rightarrow C_3T_2$$

giving a 3' partial sequence (C_3T_2) TTC. The combination of these two partial sequences, gave the total sequence CTTCCTTC. Since the two digests did not yield overlapping sequences, exact knowledge of the base composition of this oligonucleotide was necessary for the deduction of the complete sequence.

3.4.6 Oligonucleotide C_2T_6

The autoradiograph from the spleen exonuclease digest of C_2T_6 (Figure 3.10(A)) contained spots representing C_2T_6 and only two digest products,

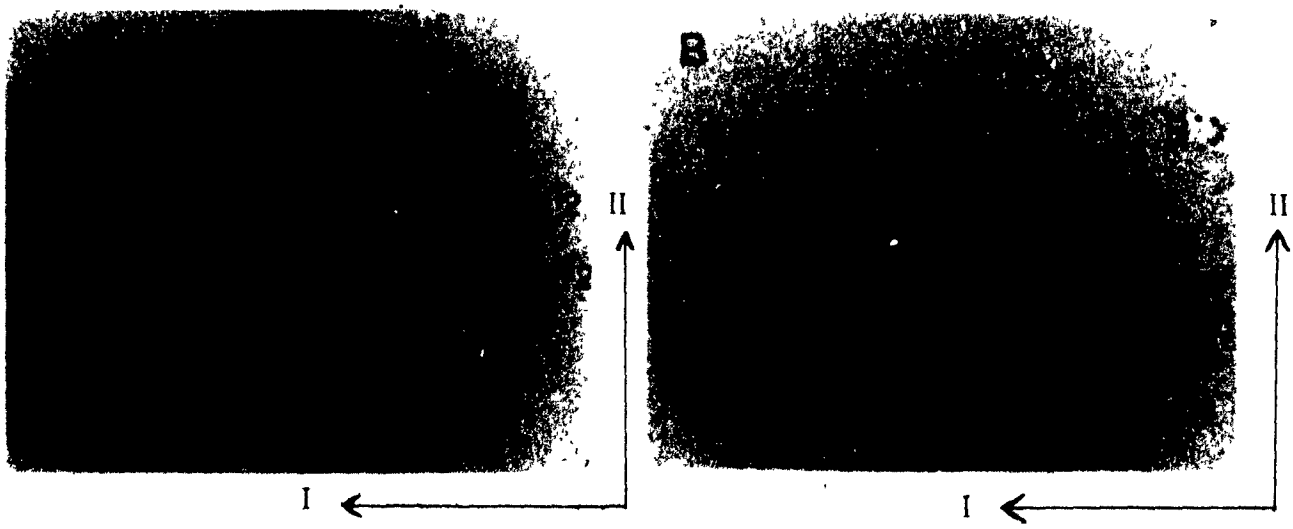


Figure 3.9 : Autoradiographs of the fractionations by electrophoresis-homochromatography of exonuclease partial digestions of uniformly labelled oligonucleotide C_4T_4. (A) spleen exonuclease ; (B) venom exonuclease. Blue and yellow markers are represented by "B" and "Y".
Legend as for Figure 3.4 . 8,200 cpm were used for each fractionation ; autoradiography exposure time was 16 hours.

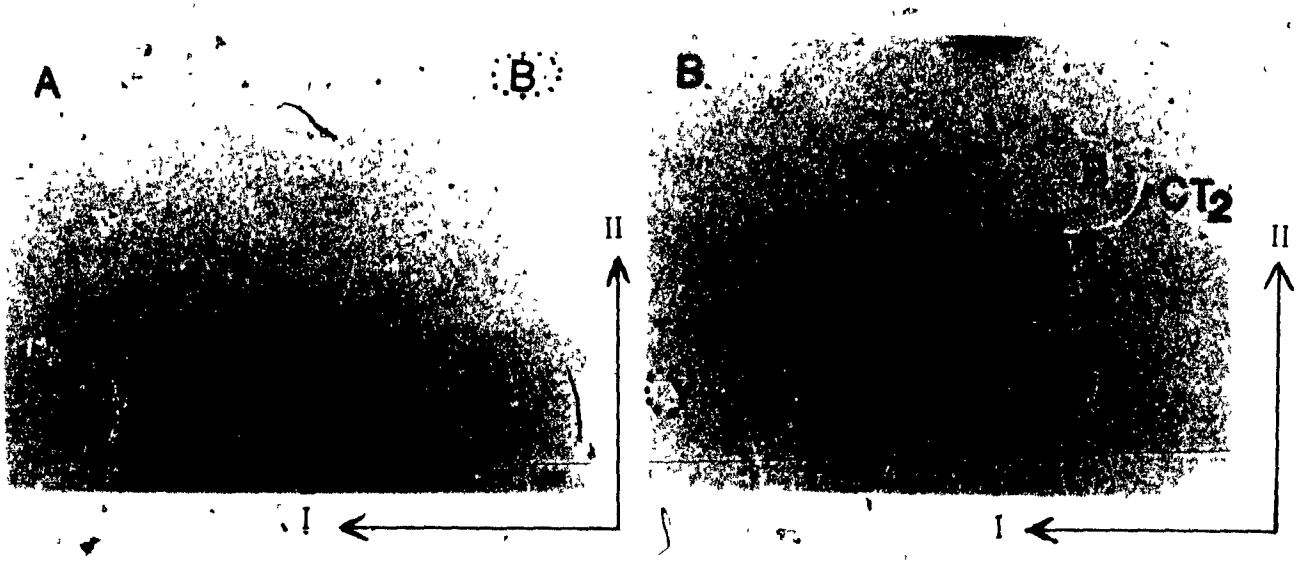


Figure 3.10 : Autoradiographs of the fractionations by electrophoresis-homochromatography of exonuclease partial digestions of uniformly labelled oligonucleotide C_2T_6. (A) spleen exonuclease ; (B) venom exonuclease. Blue and yellow markers are represented by "B" and "Y".
Legend as for Figure 3.4 . 8,500 cpm were used for each fractionation ; autoradiography exposure time was 16 hours.

CT_6 and C_2T_5. This showed that the two isomers in this fraction had different 5' termini.

The autoradiograph from the venom exonuclease digest of C_2T_6 (Figure 3.10(B)) contained spots representing C_2T_6 and nine digest products. These showed that the digest from the 3' termini of these oligonucleotides proceeded :

$$C_2T_6 \rightarrow \begin{cases} CT_6 \rightarrow CT_5 \\ C_2T_5 \rightarrow C_2T_4 \end{cases} \rightarrow CT_4 \rightarrow \begin{cases} T_4 \rightarrow T_3 \\ CT_3 \rightarrow CT_2 \end{cases}$$

Since one of the parent oligonucleotides had a C at the 5' terminus, the CT_3 in the above diagram must have the sequence CTTT. At this point the sequence information can be written as a combination of 5' and 3' terminal sequences :

TTTTC	TTC
CTTTT	CTT

Oligonucleotide CT_6 was eluted from the TLC plate of the spleen exonuclease digest fractionation, (Figure 3.10(A)) desalted, and partially digested with venom exonuclease. The autoradiograph of the fractionation of this digest (Figure 3.11) contained spots representing CT_6 and four digest products. These showed that the digest from the 3' terminus proceeded :

$$CT_6 \rightarrow T_6 \rightarrow T_5 \rightarrow T_4 \rightarrow T_3$$

giving a 3' terminal sequence of TTTTTTC.

Thus the two sequences for the oligonucleotides C_2T_6 were CTTTTTTC and TTTTCCTT.

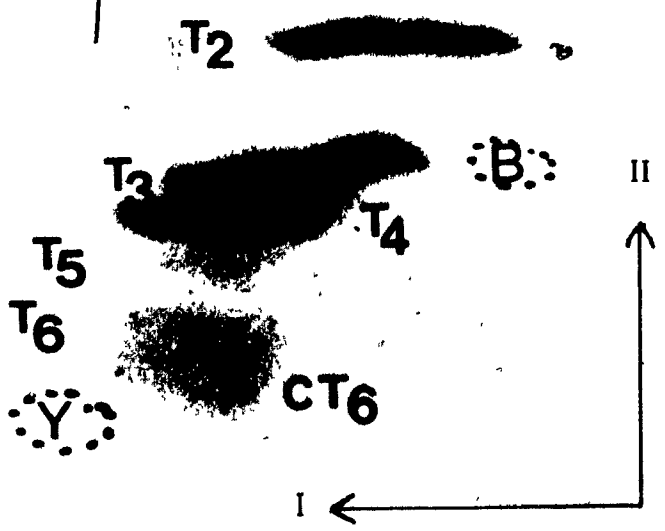


Figure 3.11 : Autoradiograph of the fractionation by electrophoresis–homochromatography of the venom exonuclease digest of oligonucleotide CT_6 isolated from C_2T_6. (Figure 3.10 (A)). Blue and yellow markers are represented by "B" and "Y".

Legend as for Figure 3.4. 960 cpm were used for this fractionation ; autoradiography exposure time was 14 days.

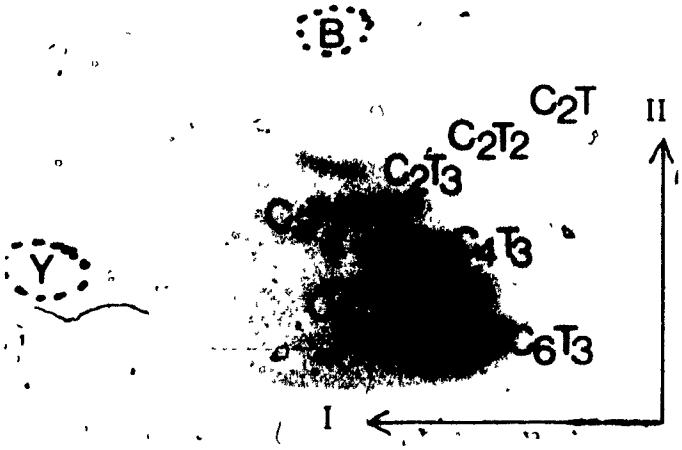


Figure 3.12 : Autoradiograph of the fractionation by electrophoresis–homochromatography of the venom exonuclease digest of oligonucleotide C_6T_3. Blue and yellow markers are represented by "B" and "Y".
Legend as for Figure 3.4 . 4,500 cpm were used for this fractionation ; autoradiography exposure time was 16 hours.

3.4.7 Oligonucleotide C_6T_3

The autoradiograph from the spleen exonuclease digest contained only one spot representing C_6T_3 showing only that this oligonucleotide was refractory to digestion.

The autoradiograph from the venom exonuclease digestion of C_6T_3 (Figure 3.12), however, contained spots representing C_6T_3 and six digest products. These indicated that the digest from the 3' end proceeded :

$$C_6T_3 \rightarrow C_5T_3 \rightarrow C_4T_3 \rightarrow C_3T_3 \rightarrow C_2T_3 \rightarrow C_2T_2 \rightarrow C_2T$$

giving a 3' terminal sequence (C_2T) TTCCCC.

This autoradiograph also contained spots which may represent CT_3 and T_3 yielding the alternate digest pattern :

$$C_6T_3 \rightarrow C_5T_3 \rightarrow C_4T_3 \rightarrow C_3T_3 \rightarrow C_2T_3 \rightarrow CT_3 \rightarrow T_3$$

and the sequence TTTCCCCCC. These spots, however, probably resulted from the minor contamination observed to the left of C_6T_3 on the autoradiograph.

3.4.8 Oligonucleotide C_6T_4

The autoradiograph from the spleen exonuclease digest of C_6T_4 (Figure 3.13(A)) contained spots representing C_6T_4 and three digest products. These showed that the digest from the 5' terminus proceeded :

$$C_6T_4 \rightarrow C_5T_4 \rightarrow C_5T_3 \rightarrow C_4T_3$$

giving the 5' partial sequence CTC (C_4T_3).

The autoradiograph from the venom exonuclease digest of C_6T_4 (Figure 3.13(B)) contained spots representing C_6T_4 and five digest products.

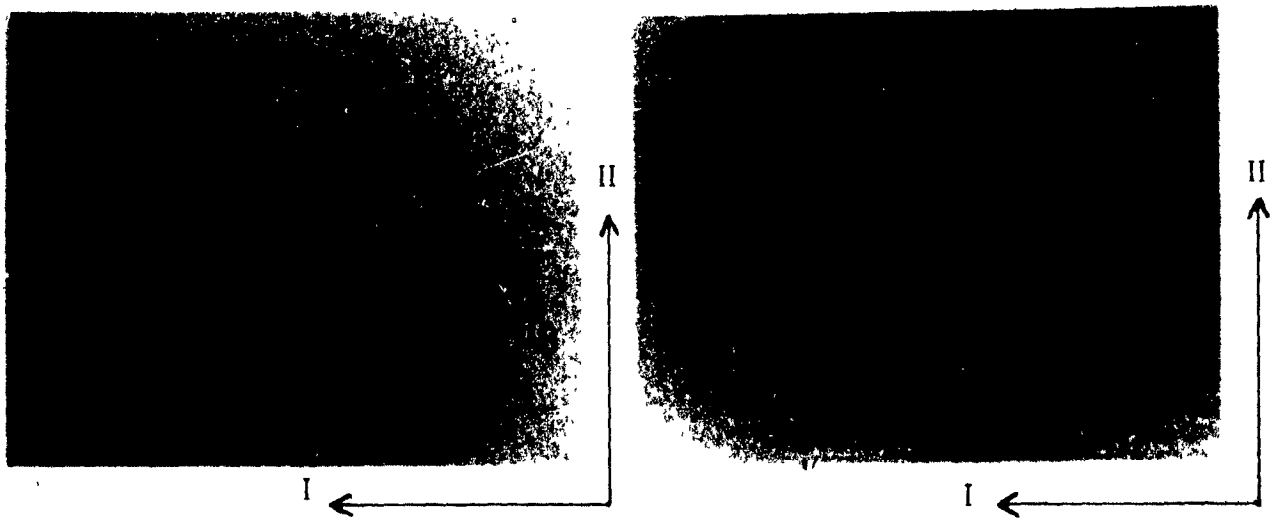


Figure 3.13 : Autoradiographs of the fractionations by electrophoresis-homochromatography of exonuclease partial digestions of uniformly labelled oligonucleotide C_6T_4. (A) spleen exonuclease ; (B) venom exonuclease. Blue and yellow markers are represented by "B" and "Y".
Legend as for Figure 3.4. 1,400 cpm were used for each fractionation ; autoradiography exposure time was 16 hours.

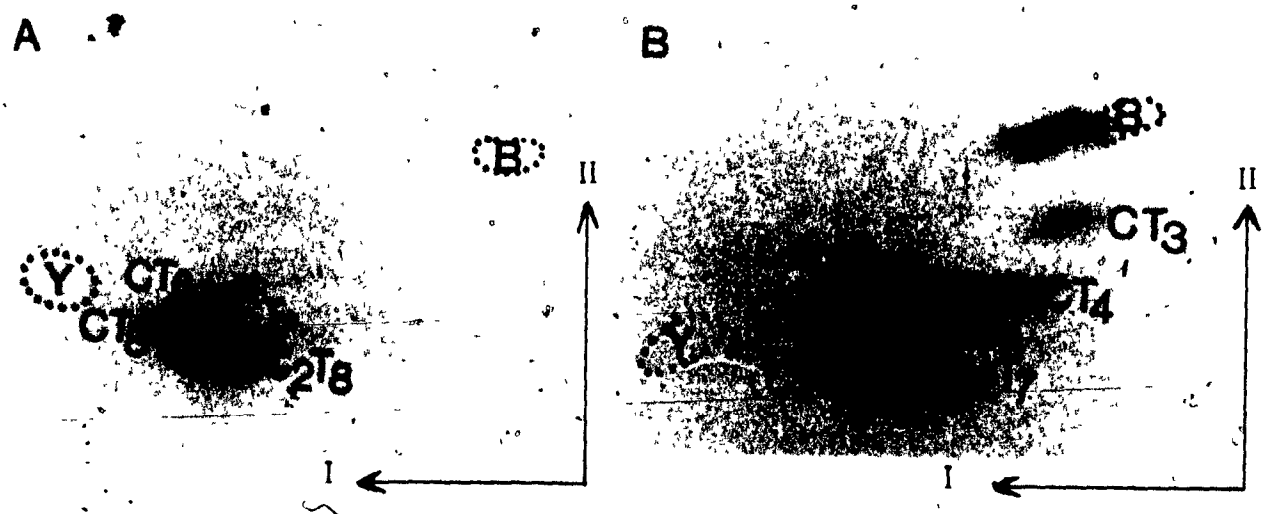


Figure 3.14 : Autoradiographs of the fractionations by electrophoresis-homochromatography of exonuclease partial digestions of uniformly labelled oligonucleotide C_2T_8. (A) spleen exonuclease ; (B) venom exonuclease. Blue and yellow markers are represented by "B" and "Y".
Legend as for Figure 3.4. 1,900 cpm were used for each fractionation ; autoradiography exposure time was 16 hours.

These showed that the digest from the 3' end proceeded :

$$C_6T_4 \rightarrow C_5T_4 \rightarrow C_4T_4 \rightarrow C_4T_3 \rightarrow C_4T_2 \rightarrow C_3T_2$$

giving the 3' partial sequence (C_3T_2)CTTCC.

The combination of these two partial sequences gave the incomplete sequence CTC(TC)CTTCC.

3.4.9 Oligonucleotide C_2T_8

The autoradiograph from the spleen exonuclease digest of C_2T_8 (Figure 3.14(A)) contained spots representing C_2T_8 and three digest products. These showed that the digest from the 5' end proceeded :

$$C_2T_8 \rightarrow CT_8 \rightarrow CT_7 \rightarrow CT_6$$

giving the 5' partial sequence CTT(CT_6).

The autoradiograph from the venom exonuclease digest of C_2T_8 (Figure 3.14(B)) contains spots representing C_2T_8 and six digest products. These showed that the digest from the 3' end proceeded :

$$C_2T_8 \rightarrow C_2T_7 \rightarrow CT_7 \rightarrow CT_6 \rightarrow CT_5 \rightarrow CT_4 \rightarrow CT_3$$

yielding the 3' partial sequence (CT_3)TTTTCT.

The combination of these two partial sequences gives the total sequence CTTTTTTTGT. Again, knowledge of the base composition was necessary for the deduction of the complete sequence.

3.4.10 Oligonucleotide C_5T_6

The autoradiograph from the spleen exonuclease partial digest of C_5T_6 (Figure 3.15(A)) contained only one spot, and thus gave no sequence information.

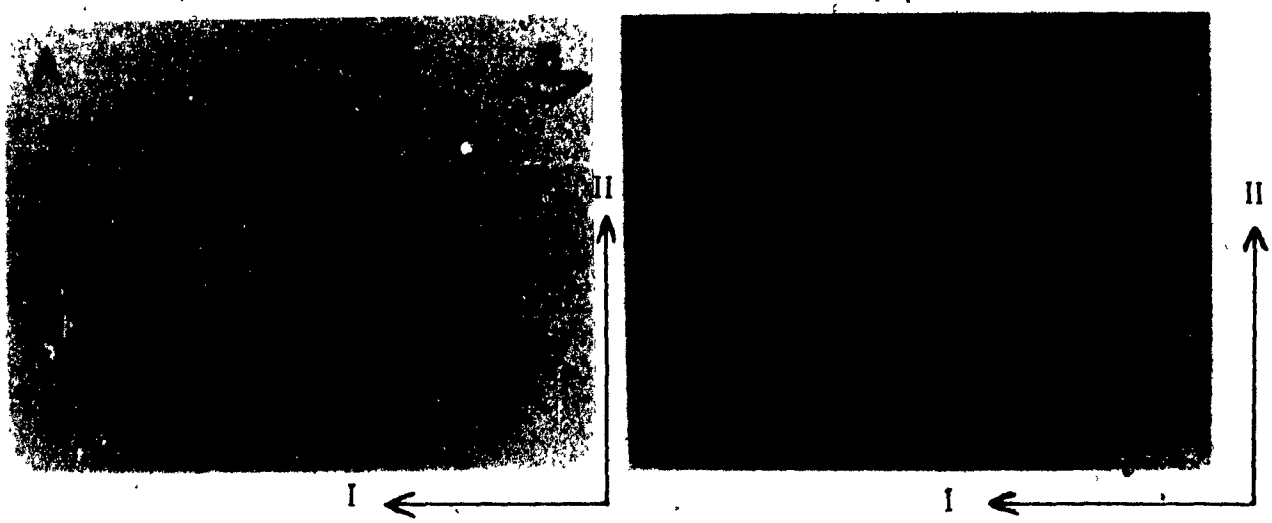


Figure 3.15 : Autoradiographs of the fractionations by electrophoresis-homochromatography of exonuclease partial digestions of uniformly labelled oligonucleotide C_5T_6. (A) spleen exonuclease ; (B) venom exonuclease. Blue and yellow markers are represented by "B" and "Y".

Legend as for Figure 3.4. 2,700 cpm were used for each fractionation ; autoradiography exposure time was 16 hours.

The autoradiograph from the venom exonuclease partial digest of C_5T_6 (Figure 3.15 (B)) contained spots representing C_5T_6 and six digest products. These indicated that the digest from the 3' end proceeded :

$$C_5T_6 \rightarrow C_5T_5 \rightarrow C_4T_5 \rightarrow C_4T_4 \rightarrow C_4T_3 \rightarrow C_3T_3 \rightarrow C_3T_2$$

yielding the 3' partial sequence (C_3T_2)TCTTCT.

3.4.11 Summary

The sequences of pyrimidine oligonucleotides from $S13^+$ DNA which have been determined using uniformly labelled material as substrate are summarized in Table 3.2.

In the cases of oligonucleotides C_6T_3 and C_5T_6 further experiments with more severe spleen exonuclease digestion conditions would certainly have yielded more sequence data ; however, the sequence of these oligonucleotides has been presented by two other techniques, and it was decided that further sequence studies on these uniformly labelled oligonucleotides were of no practical value.

3.5 Properties of Polynucleotide Kinase

3.5.1 Oligonucleotide-like Contaminants in Polynucleotide Kinase

3.5.1.1 Properties of the Contaminants

It has been found that polynucleotide kinase prepared according to Richardson (12) contained contaminants which interfered with sequence studies. When the kinase and [γ -^{32}P] ATP were incubated in the absence of oligonucleotide substrate and the mixture fractionated on a column of Sephadex G-15, considerable radioactivity eluted before the ATP fraction (Figure 3.16). A chain length

Table 3.2

Summary of the sequences deduced from uniformly labelled oligonucleotides isolated from $SI3^+$ DNA

Isostich	Components	Sequence
11	C_5T_6	(C_3T_2) TCTTCT
10	C_6T_4	CTC (CT) CTTCC
	C_2T_8	CTTTTTTCT
9	C_6T_3	(C_2T) TTCCCC*
8	C_5T_3	CCTTTCCC
	C_4T_4	CTTCCTTC
	C_2T_6	CTTTTTTC
		TTTTCCTT
7	C_6T	CTCCCCC
	C_5T_2	CTCTCCC
	CT_6	CTTTTTT

* the less likely sequence TTTCCCCCC cannot be discounted from uniform label sequence data alone.

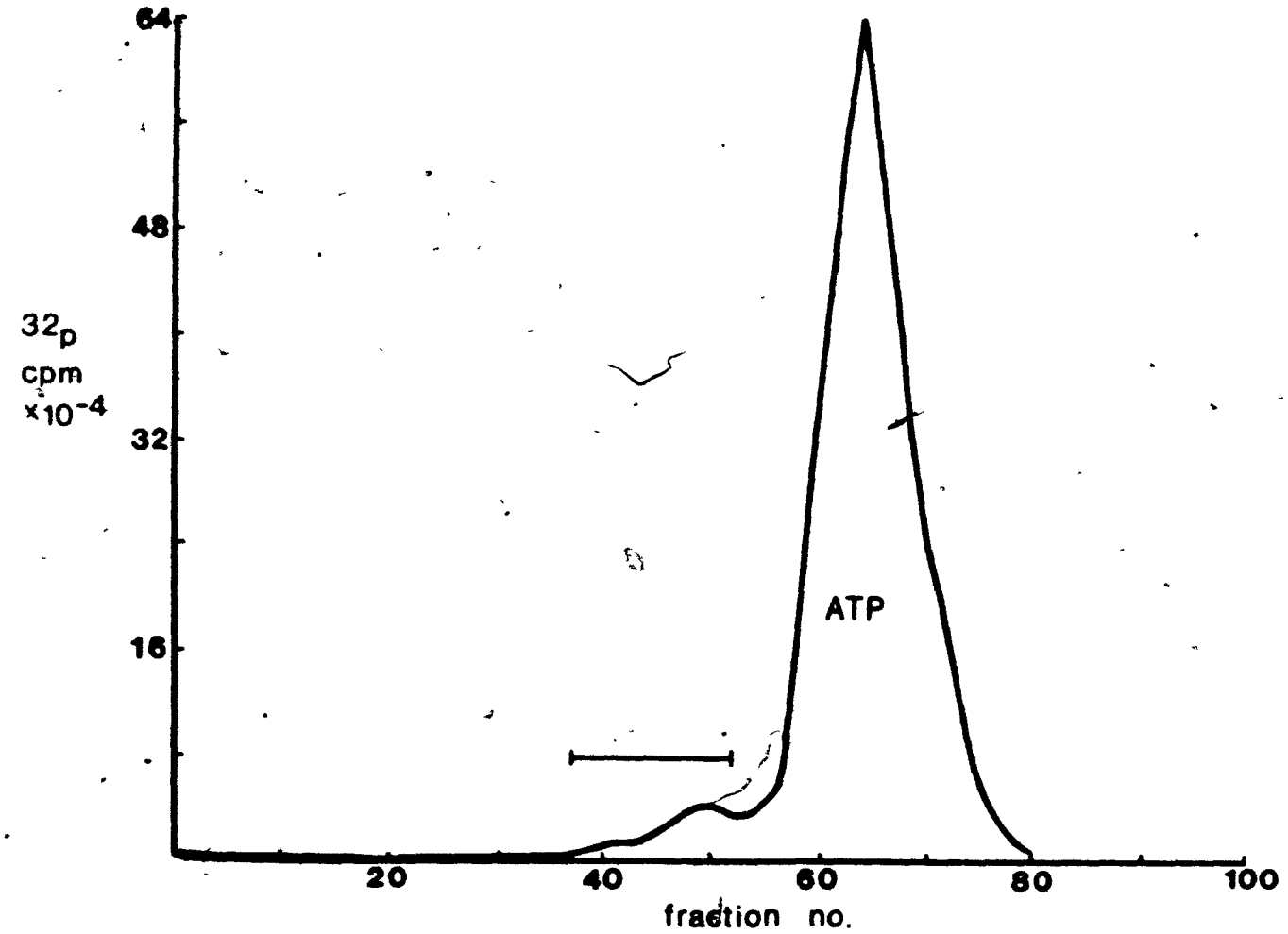


Figure 3.16 : Elution profile from a Sephadex G-15 column of an incubation of polynucleotide kinase and [γ-^{32}P] ATP.

Incubation was for 12 hours at 22°C in a total volume of 200 µl of 10 mM Tris-HCl buffer pH 8.1, 10 mM $MgCl_2$, 20 mM mercaptoethanol containing 2 nm of [γ-^{32}P]ATP and no added oligonucleotide. Column dimensions were 1 cm x 50 cm, the eluate was 0.05 M NaCl, flow rate was 0.5 ml/min and 0.4 ml fractions were collected. The bars represent the fractions pooled for further analysis.

fractionation of this material on DEAE cellulose indicated that 78% of it was similar in charge to trinucleotides, while 22% was divided among the di-, tetra-, and pentanucleotide fractions (Figure 3.17).

Extensive dialysis of this enzyme preparation against 50 mM Tris-HCl buffer, pH 7.5, 10 mM mercaptoethanol, 0.5 M KCl did not remove a significant amount of the contaminants ; nor did gel filtration on a Sephadex G-25 column eluted with the same buffer.

A spectrum of this preparation of polynucleotide kinase (Figure 3.18) showed a high A_{260}/A_{280} ratio, indicating the presence of nucleotide material .

The elution profile of ^{32}P radioactivity suggested that these contaminants had a varying number of phosphate groups ; the spectrum of the polynucleotide kinase suggested that they were nucleotide in nature ; and the fact that they were stable in alkali indicated that they might be oligodeoxyribonucleotides.

The contamination was calculated to be approximately 12 nmoles nucleotide per ml of polynucleotide kinase stock solution. This corresponds to 120 pmoles of trinucleotide contaminant in a normal labelling reaction using 0.015 units of polynucleotide kinase.

Incorporation of radioactivity into the contaminating oligonucleotides in control experiments (Figure 3.16) indicated that their concentration was from 5 to 20 nmoles /ml of stock solution. This level of interfering oligonucleotide (8 nmoles/unit) is much too high for sequence studies.

3.5.1.2 Preparation of Contaminant-free Polynucleotide Kinase

Polynucleotide kinase was prepared according to Richardson (12) except

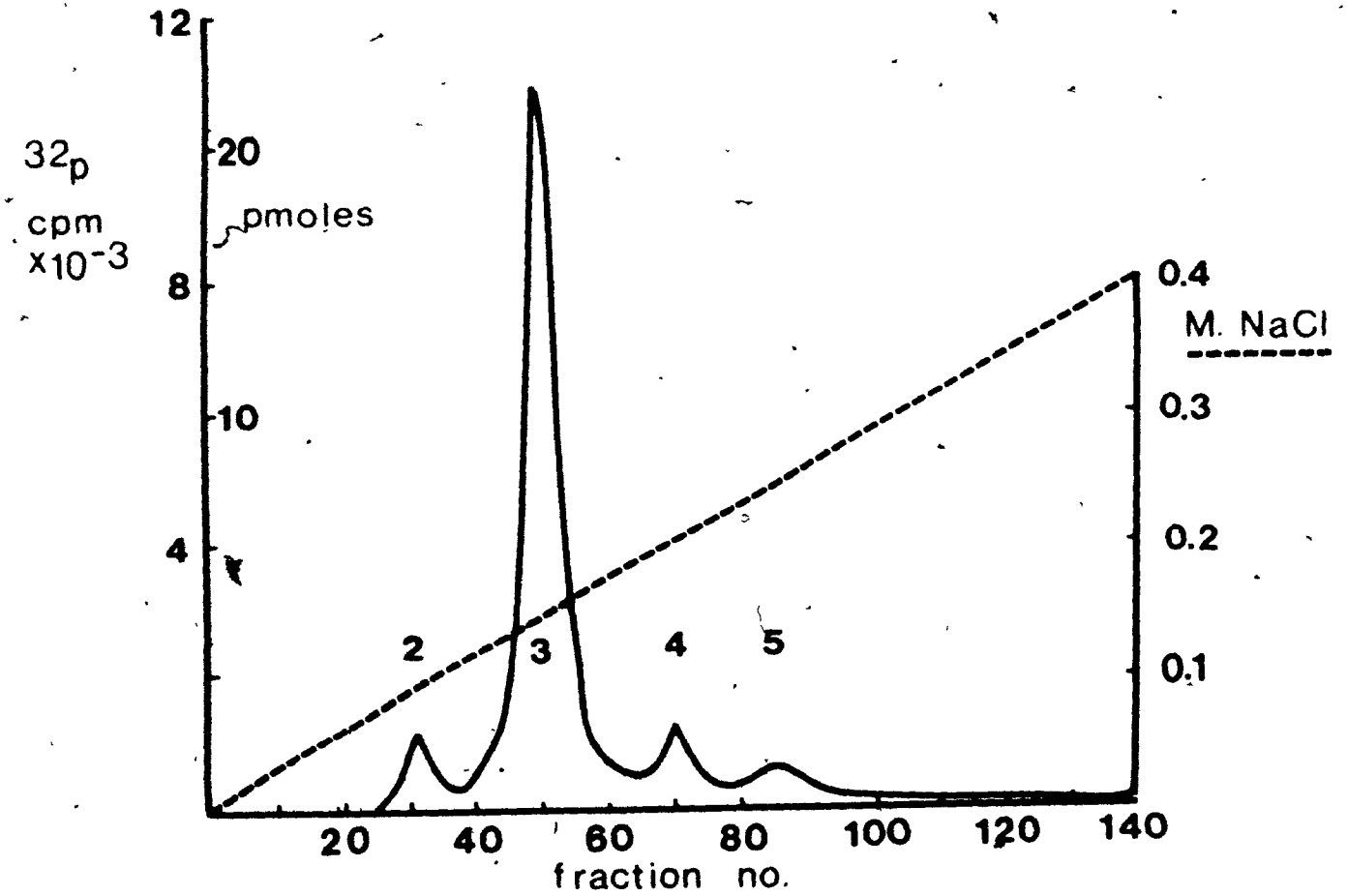


Figure 3.17 : Chain length fractionation of labelled contaminants isolated from polynucleotide kinase.

The pooled fractions from the Sephadex G-15 eluate (Figure 3.16) were loaded directly to a 1 cm x 25 cm DEAE cellulose column ; the wash solution was water, eluate was 2 l of 0.05 M sodium acetate buffer pH 5.5, 1 mM KH_2PO_4 , 7 M urea, with a gradient of 0 to 0.4 M NaCl ; flow rate was 0.6 ml/min, 14 ml fractions were collected.

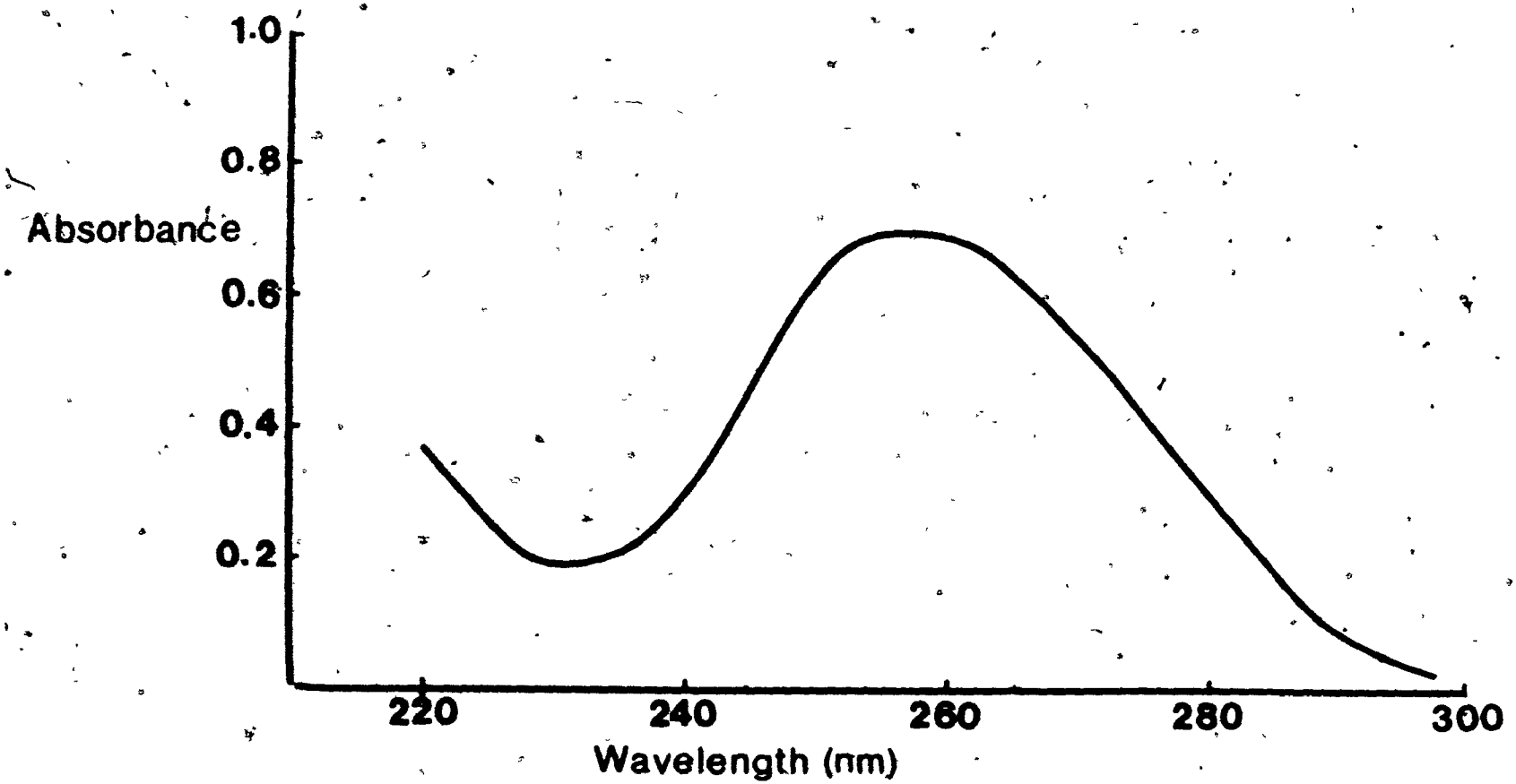


Figure 3.18 : Spectrum of polynucleotide kinase.

Polynucleotide kinase (0.5 units/ml) dissolved in 0.05 M Tris-HCl buffer, pH 7.5, 10 mM $MgCl_2$, 20 mM mercaptoethanol, 50% glycerol, was placed in the sample cuvette (1 cm path length). Buffer without enzyme was used as blank. The spectrum was measured in a Gilson 2000 spectrophotometer from 220 to 300 nm at 2 nm intervals.

1 mM ATP was added to all buffers up to and including the wash buffer of the DEAE cellulose column. Contaminants were labelled in an incubation of this preparation of enzyme with [γ-^{32}P] ATP and were isolated by Sephadex G-15 chromatography. A chain length fractionation of this material (Figure 3.19) showed that the level of interfering oligonucleotide had been reduced by a factor of 24 to 330 pmoles/unit.

Even this level of contamination would be too high for many sequence experiments and therefore this preparation of polynucleotide kinase was further purified on another DEAE cellulose column (13).

Unlike the initial DEAE cellulose column, this fractionation separated the preparation into two portions ; 40% of the enzyme activity eluted during the loading and the wash with 10 mM potassium phosphate buffer ; 60% eluted with 50 mM buffer (Figure 3.20). The latter fraction was not used since it contained a small amount of 3' exonuclease activity.

Using very high specific activity [γ -^{32}P] ATP it was shown that the 10 mM fraction contained 4 pmoles nucleotide per unit of polynucleotide kinase 80 times less than the enzyme loaded to the column. This 10 mM fraction was used in all experiments described in this thesis.

3.5.2 Enzymatic Purity of Polynucleotide Kinase

The properties of the polynucleotide kinase isolated in this laboratory were shown to be similar to those reported by others (12, 35, 36, 37).

The preparation was tested for endonuclease and exonuclease activity by incubation with pyrimidine nonanucleotide (Py_9P_{10}) in the presence of alkaline

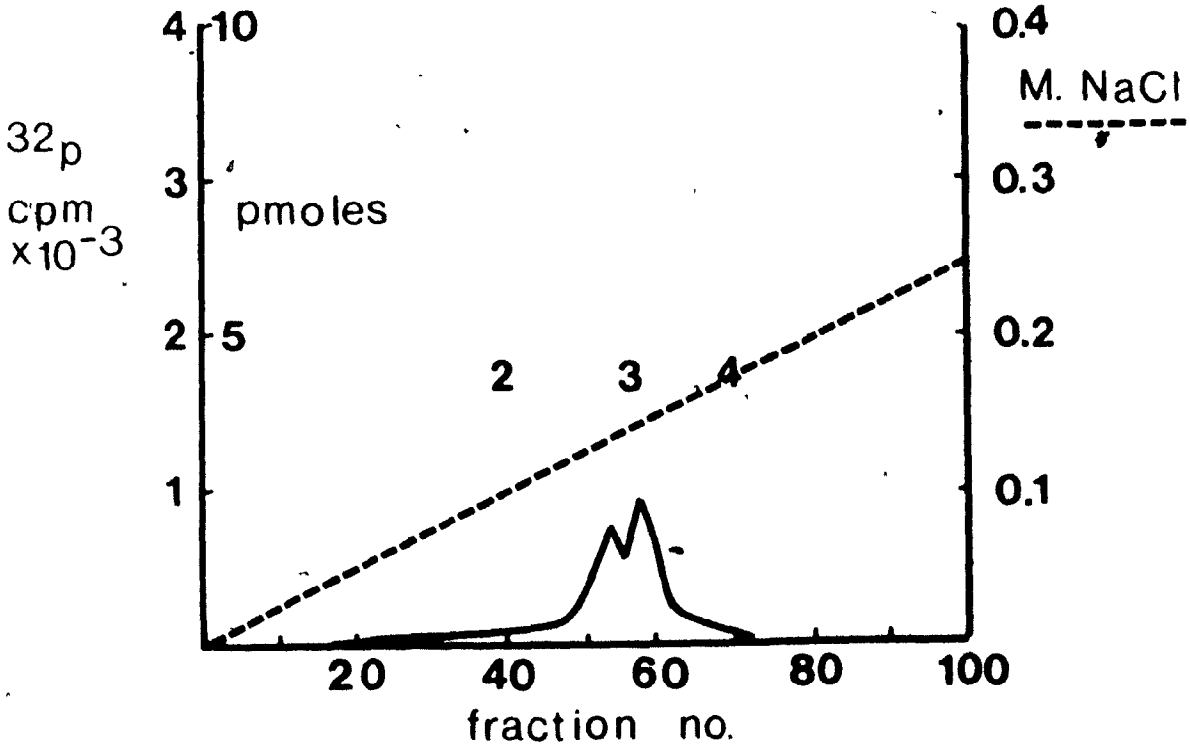


Figure 3.19: Chain length fractionation of labelled contaminants isolated from polynucleotide kinase prepared with 1 mM ATP in the initial purification buffers.

Incubation was for 12 hours at 22°C in a total volume of 0.2 ml of 10 mM Tris-HCl buffer, pH 8.1, 10 mM $MgCl_2$, 20 mM mercaptoethanol containing 2 nmoles of [γ-^{32}P]ATP, 0.01 units of polynucleotide kinase, and no added oligonucleotide. The mixture was fractionated on a Sephadex G-15 column and the contaminants loaded directly to DEAE cellulose. Column dimensions were 1 cm x 25 cm; the wash solution was water, eluate was 2 l of 50 mM sodium acetate buffer pH 5.5, 1 mM KH_2PO_4, 7 M urea, with a gradient of 0 to 0.4 M NaCl; flow rate was 0.6 ml/min, 12 ml fractions were collected.

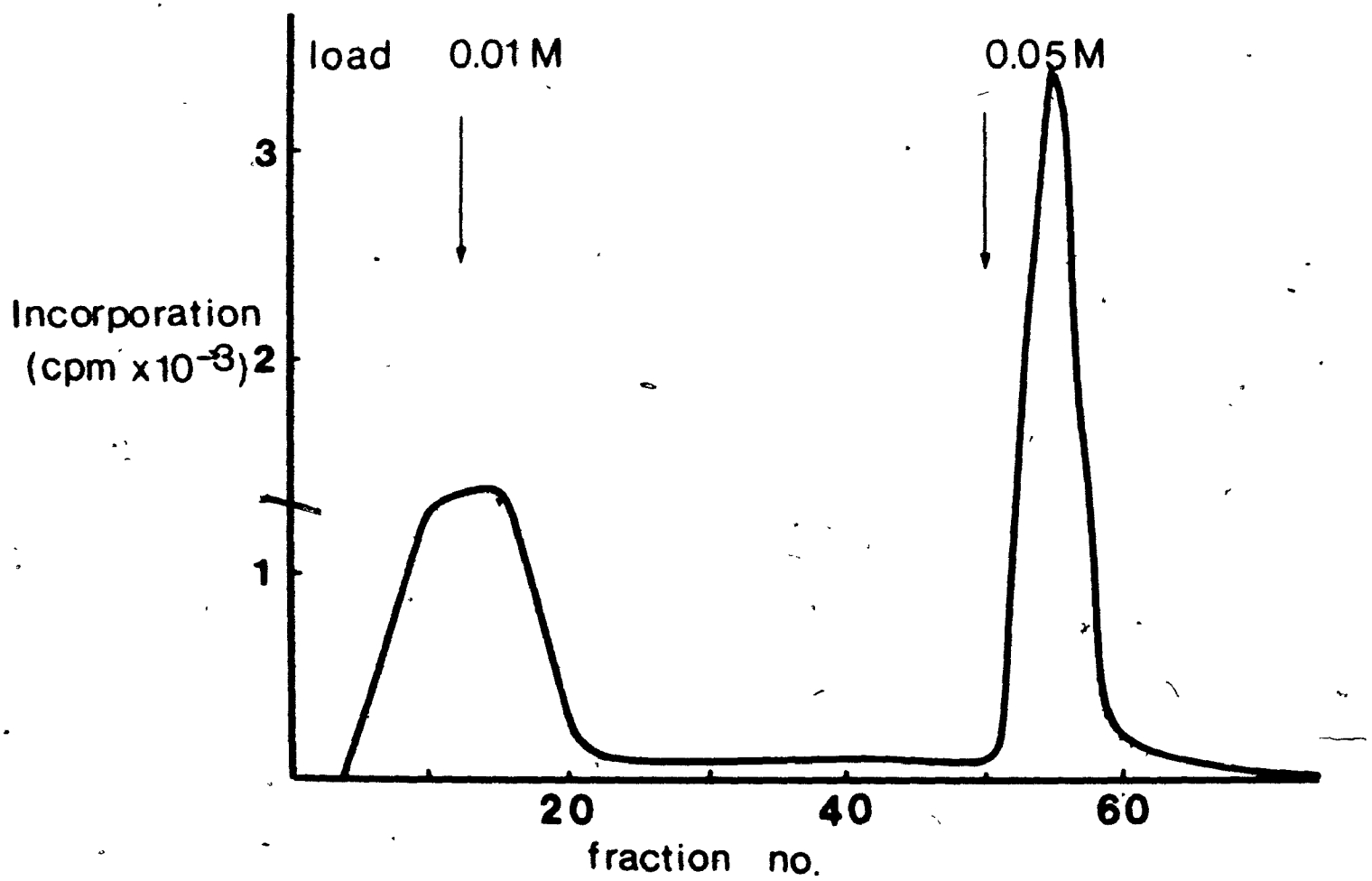


Figure 3.20 : Chromatography of polynucleotide kinase on DEAE cellulose.

Polynucleotide kinase (see text) was dialyzed against 10 mM potassium phosphate buffer pH 7.6, 20 mM mercaptoethanol, and loaded to a 1 cm by 25 cm column of DEAE cellulose washed with the same buffer. The column was eluted with 200 ml of the wash buffer followed by 100 ml of 50 mM potassium phosphate buffer, pH 7.6, 20 mM mercaptoethanol ; 20 ml fractions were collected and 30 µl of each fraction was assayed (2.2.5.2.).

phosphomonoesterase. A chain length fractionation of this incubation (Figure 3.21) showed that no oligonucleotides smaller than the nonanucleotide were present. Pyrimidine nonanucleotide incubated without phosphomonoesterase or polynucleotide kinase eluted at a volume corresponding to a molecule containing ten negative phosphate groups. The dephosphorylated nonanucleotide (Py_9P_8) from the test reaction eluted at a position corresponding to a molecule containing 8 phosphate groups. The absence of degradation products indicated that this preparation of polynucleotide kinase was free of nucleases.

Polynucleotide kinase was tested for extraneous activities under normal labelling conditions by incubation with dephosphorylated undecanucleotide C_5T_6 in the presence of [γ -^{32}P] ATP (2.4.3). The absence of labelled products other than the undecanucleotide in the chain length fractionation of this incubation (Figure 3.22) indicated that ATP dependent nucleolytic or polymerizing activities were not significant.

The endonuclease activity present in this preparation of polynucleotide kinase was assayed by the very sensitive assay described by Sadowski (38) for T4 endonuclease IV. This involved the digestion of ^{32}P-labelled S13$^+$ single stranded circular DNA with the polynucleotide kinase, followed by digestion of the nicked circles with exonuclease I and the measurement of acid precipitable radioactivity. This test indicated that in a normal labelling reaction one endonucleolytic cision would be made per 12,000 nucleotide bands.

3.5.3 Oligonucleotide 3' Phosphomonoesterase Activity of Polynucleotide Kinase

In all labelling experiments using polynucleotide kinase and [γ -^{32}P] ATP

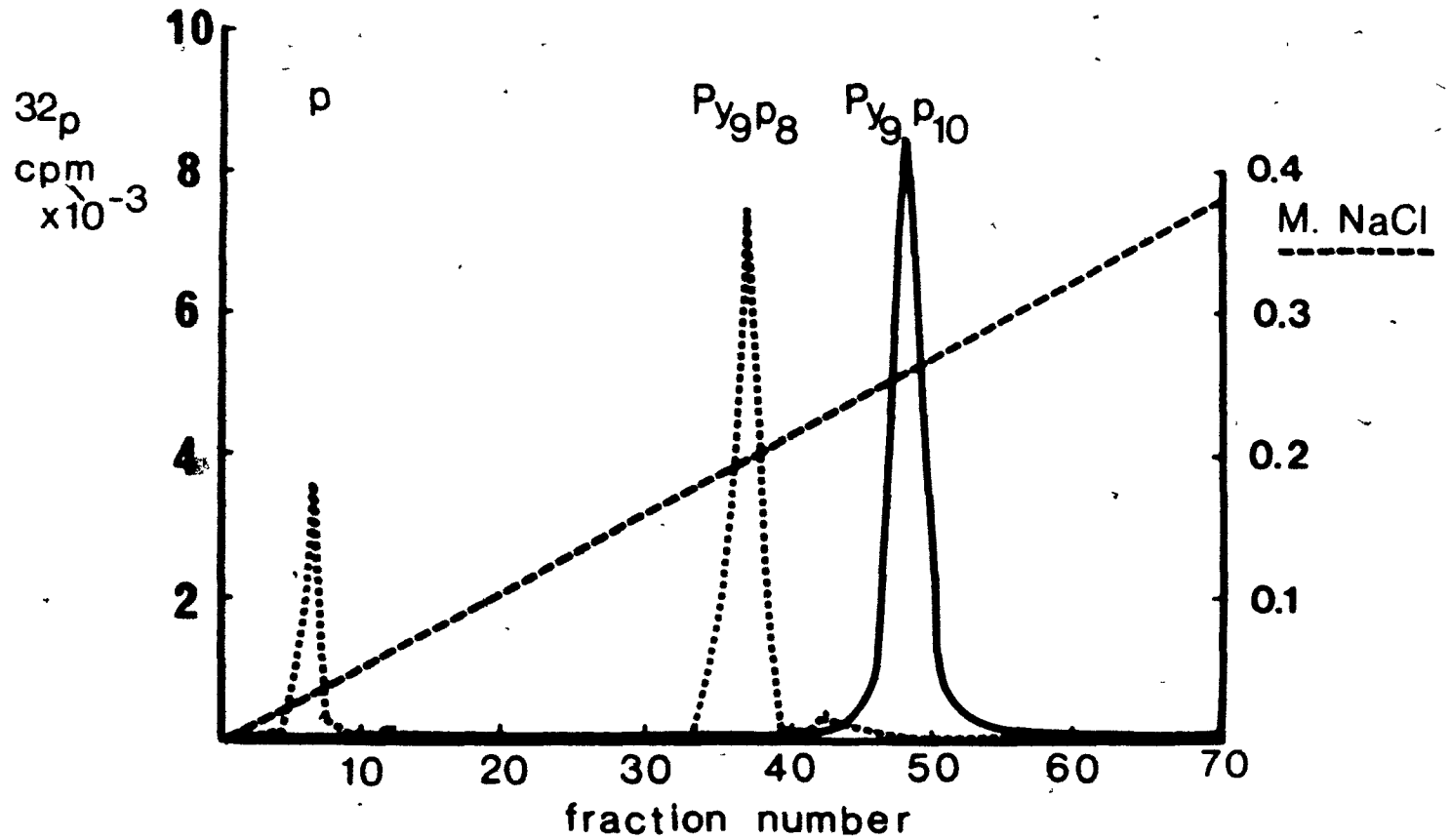


Figure 3.21 : Chain length fractionation of a pyrimidine nonanucleotide (Py_9P_{10}) after treatment with phosphomonoesterase and polynucleotide kinase.

Incubation was for 12 hours at 22°C in a total volume of 0.5 ml of 10 mM Tris-HCl buffer, pH 8.1, 10 mM $MgCl_2$, 20 mM mercaptoethanol containing 500 pmoles of ^{32}P-labelled nonanucleotide and 0.025 units of polynucleotide kinase. The mixture was diluted to 5 ml with water and loaded to a 1 cm x 25 cm DEAE cellulose column; wash solution was water, eluate was 1 l of 50 mM sodium acetate buffer, pH 5.5, 1 mM KH_2PO_4, 7 M urea, with a gradient of 0 to 0.4 M NaCl; flow rate was 0.5 ml/min ; 8 ml fractions were collected. The dotted line represents the fractionation of a similar incubation containing no enzymes.

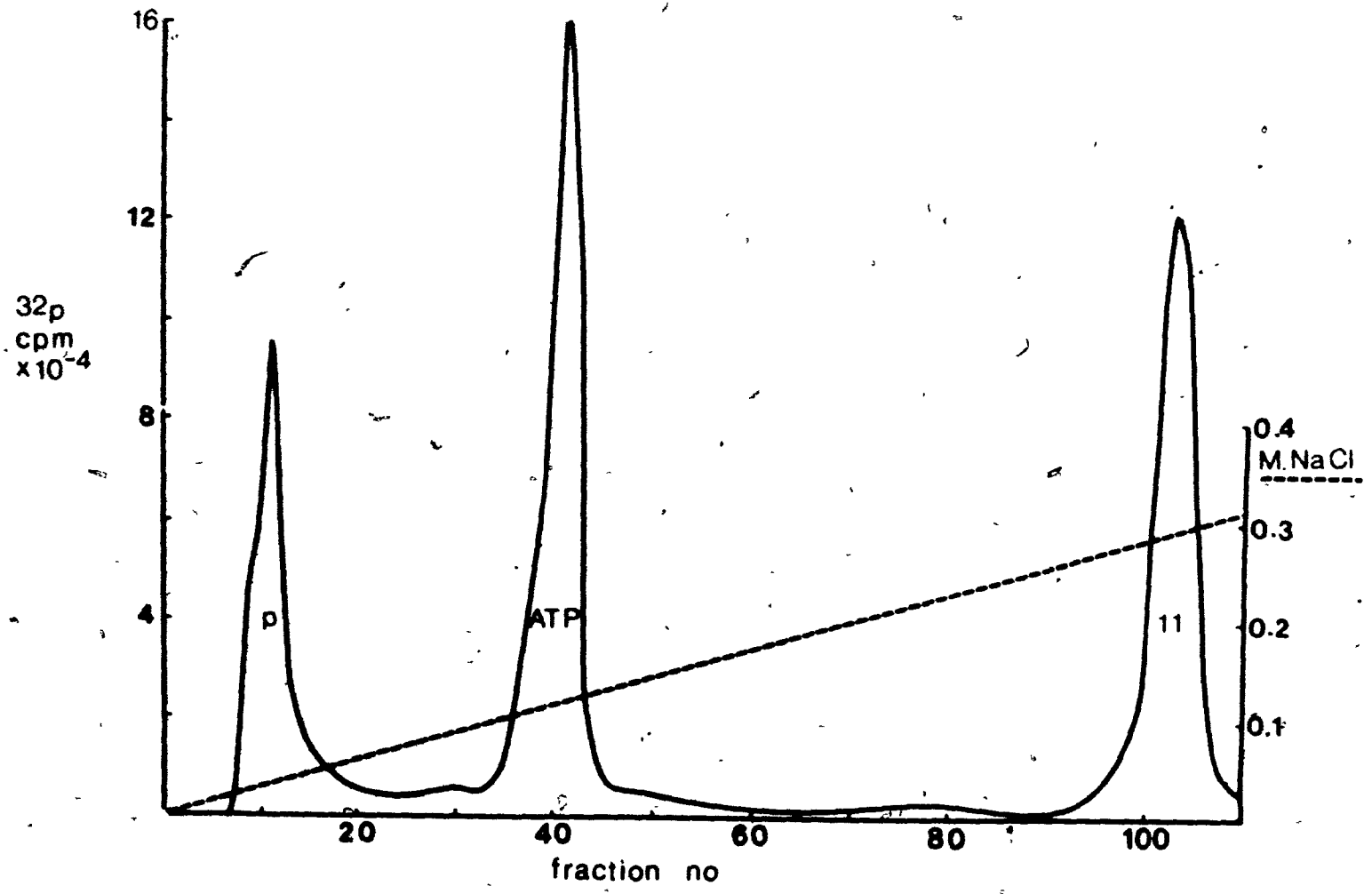


Figure 3.22 : Chain length fractionation of undeconucleotide C_5T_6 labelled using polynucleotide kinase and [γ-^{32}P]ATP

Incubation was for 16 hours at 22°C in a total volume of 0.3 ml 10 mM Tris-HCl buffer, pH 8.1, 10 mM $MgCl_2$, 20 mM mercaptoethanol, containing 250 pmoles of [γ-^{32}P]ATP, and 0.015 units of polynucleotide kinase. The mixture was diluted to 5 ml with water and loaded to a 1 cm x 25 cm DEAE cellulose column, wash solution was water, eluate was 2 l of 50 mM sodium acetate buffer, pH 5 5, 1 mM KH_2PO_4 7 M urea, with a gradient of 0 to 0.4 M NaCl ; flow rate was 0.6 ml/min , 14 ml fractions were collected

considerable amounts of inorganic phosphate were produced (Figure 3.21). Control experiments in this laboratory and elsewhere (37) have shown that ATP was not dephosphorylated by polynucleotide kinase in the absence of oligonucleotide substrates suggesting that the phosphate was hydrolyzed from the oligonucleotide.

The action of polynucleotide kinase on a pyrimidine tetranucleotide (Py_4P_5) was tested by incubation in the absence of ATP. A chain length fractionation of such an incubation (Figure 3.23(A)) showed that 45% of the radioactivity eluted in a position corresponding to a molecule containing four phosphate groups. In the control experiment (Figure 3.23(B)) all the radioactivity eluted in a volume corresponding to a molecule containing 5 phosphate groups. Fractionation of incubations which included phosphomonoesterase (Figures 3.23(C) and 3.23(D)) indicated that two phosphate groups had been removed from the tetranucleotide.

Incubation of a pyrimidine heptanucleotide (Py_7P_8) with a higher polynucleotide kinase concentration for a longer time period resulted in the quantitative removal of one phosphate from the starting material. The chain length fractionation of this incubation and of a control incubation without enzyme (Figure 3.24) showed that the kinase treated oligonucleotide contained one less phosphate than the starting material relative to the dephosphorylated decanucleotide marker.

These experiments indicated that polynucleotide kinase removed one phosphate from an oligonucleotide, but did not distinguish between the 5'- or the 3'-terminal phosphates. The observation that inorganic phosphate was produced during incubation of polynucleotide kinase and [γ-^{32}P] ATP with dephosphorylated oligonucleotides (Figure 3.22) strongly suggests that the 5' terminal ester is the one

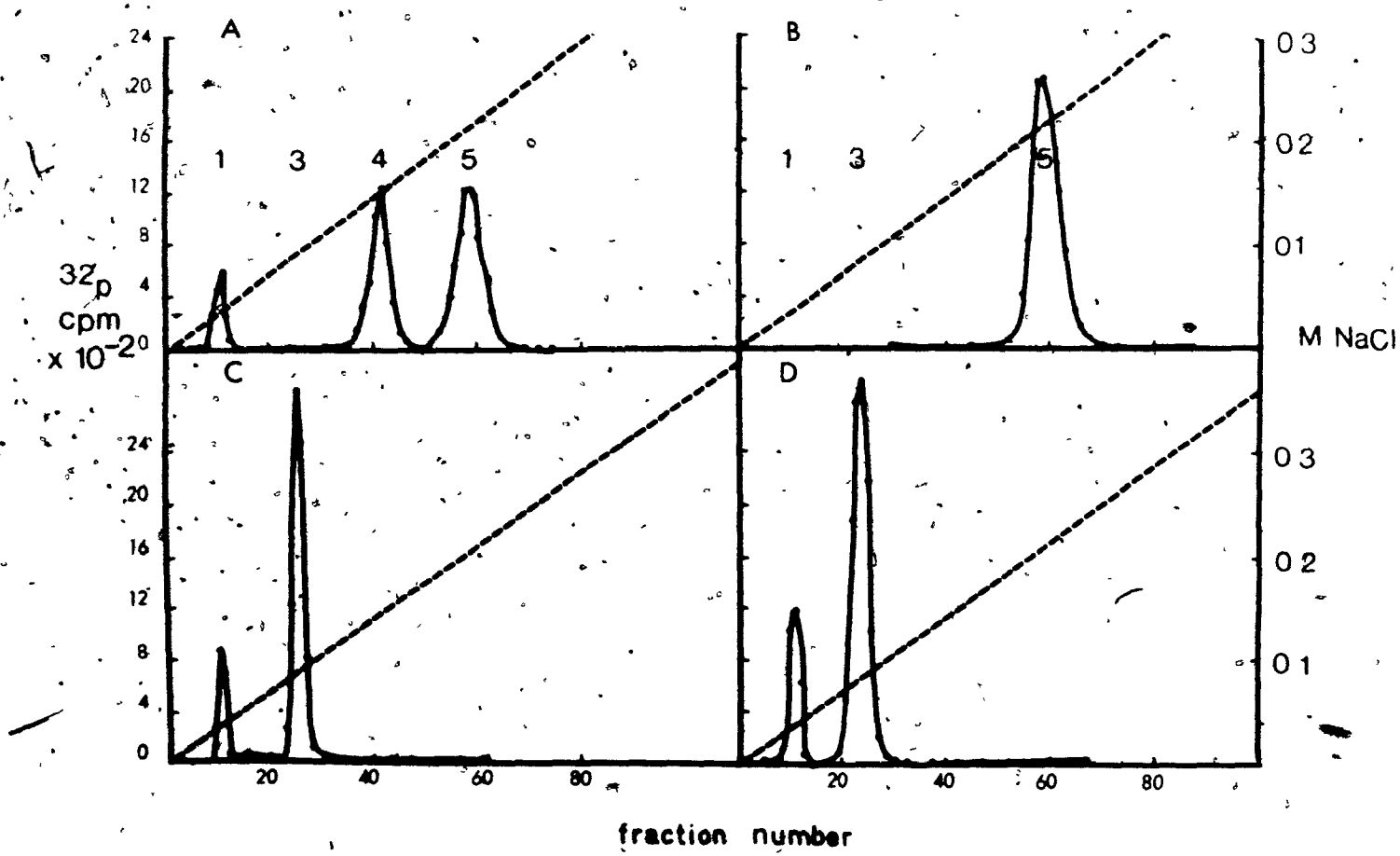


Figure 3.23 : Chain length fractionations of pyrimidine tetranucleotides after treatment with polynucleotide kinase and phosphomonoesterase.

Incubations were for 10 hours at 22°C in a total volume of 0.5 ml of 10 mM Tris-HCl buffer, pH 8.1, 10 mM $MgCl_2$, 20 mM mercaptoethanol containing 200 pmoles of tetranucleotide and the following: (A) 0.05 units of polynucleotide kinase, (B) no additions, (C) 0.75 units of phosphomonoesterase, (D) 0.75 units of phosphomonoesterase and 0.05 units of polynucleotide kinase. The mixtures were diluted to 5 ml with water and loaded to a 1 cm x 25 cm DEAE cellulose column, the wash solution was water, eluate was 1 l of 50 mM sodium acetate buffer, pH 5.5, 1 mM KH_2PO_4, 7 M urea, with a gradient of 0 to 0.4 M NaCl; flow rate was 0.6 ml/min, 10 ml fractions were collected.

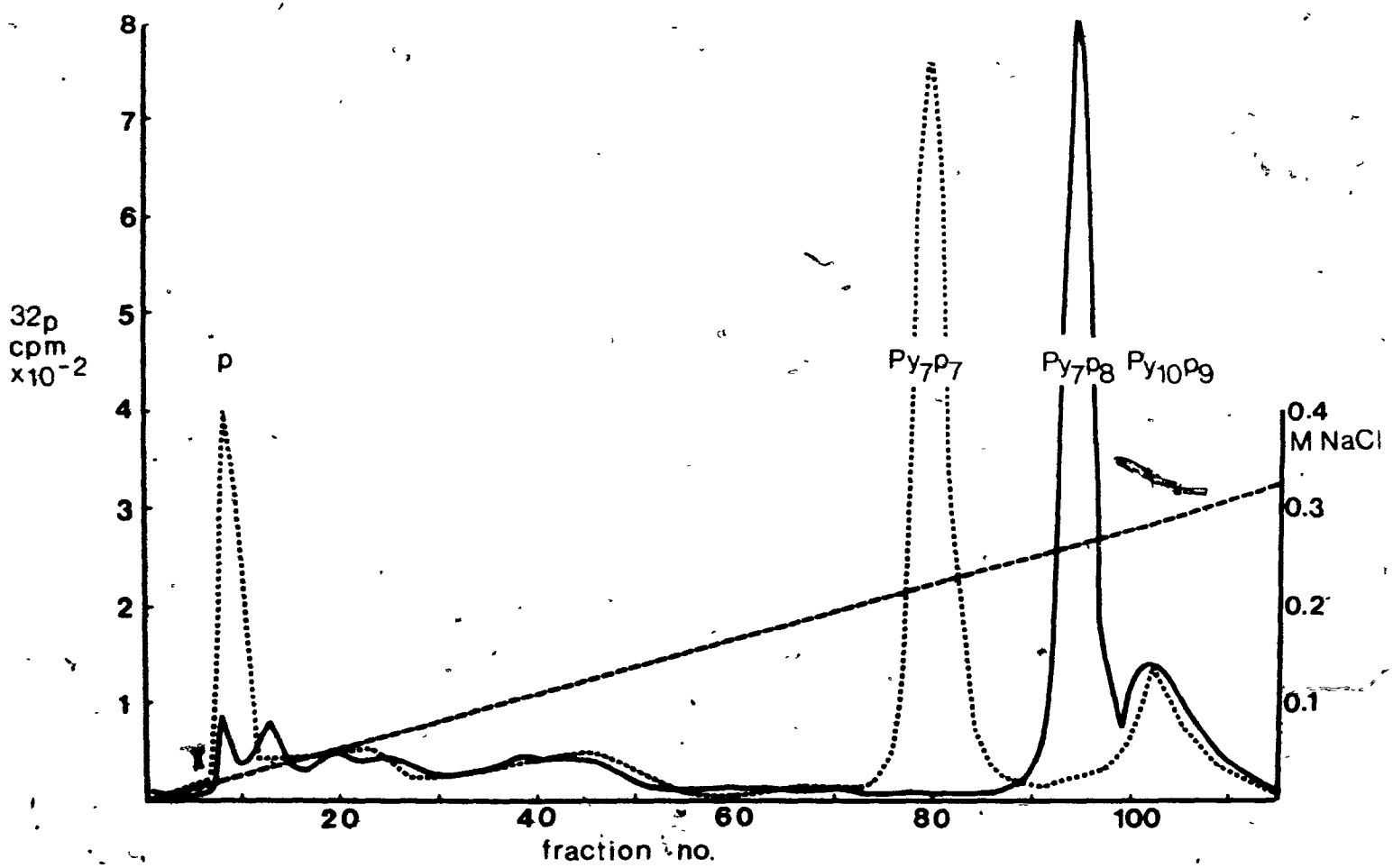


Figure 3.24 : Effect of polynucleotide kinase on a pyrimidine heptanucleotide (Py_7P_8).

The incubations were for 24 hours in a total volume of 0.2 ml of 10 mM Tris-HCl buffer, pH 8.1, 10 mM $MgCl_2$, 20 mM mercaptoethanol, containing 60 pmoles of heptanucleotide (Py_7P_8) and 20 pmoles of dephosphorylated decanucleotide ($Py_{10}P_9$). The solid line represents the chain length fractionation of an incubation containing no enzyme, the dotted line represents the fractionation of an incubation containing 0.25 units/ml polynucleotide kinase. The mixtures were diluted to 5 ml with water and loaded to a 1 cm x 25 cm DEAE cellulose column ; the wash solution was water, elyate was 2 l of 50 mM sodium acetate buffer, pH 5.5, 1 mM KH_2PO_4, 7 M urea, with a gradient of 0 to 0.4 M NaCl ; flow rate was 0.6 ml/min ; 13 ml fractions were collected.

hydrolyzed. A time course of the transfer of ^{32}P radioactivity from [γ-^{32}P]ATP to an excess of dephosphoyrlated oligonucleotides mediated by polynucleotide kinase (Figure 3.25) showed that once the [γ -^{32}P]ATP had been exhausted, the radioactivity in the oligonucleotide fraction decreased indicating that an oligonucleotide 5' phosphomonoesterase activity was present.

3.6 Sequences Deduced from 5' Terminal Labelling Followed by Exonuclease Treatment

In general, it is difficult to label deoxyribonucleic acid in vivo, to a sufficiently high specific activity for oligonucleotide sequence analysis. This quantitative problem was overcome by using polynucleotide kinase to transfer high specific activity radioactive phosphate from [γ -^{32}P]ATP to the 5' hydroxyl terminus of the oligonucleotide being investigated. In this way preparation of the DNA was simplified because low specific activity or unlabelled material could be used. The specific activity of the [γ -^{32}P]ATP used in this laboratory was sufficiently high that as little as one pmole of oligonucleotide could be used for a single sequence experiment.

Pyrimidine oligonucleotides were isolated from 47 mg of S13^{+} DNA (2.2.8) and were separated according to chain length on a DEAE Sephadex column (2.2.9.3). Individual components of the chain length fractions were isolated by chromatography at pH 3.2 on DEAE Sephadex columns, (2.2.9.3). Four to twelve nmoles of each of the three nonanucleotides, the two decanucleotides, and the undecanucleotide were obtained in this way.

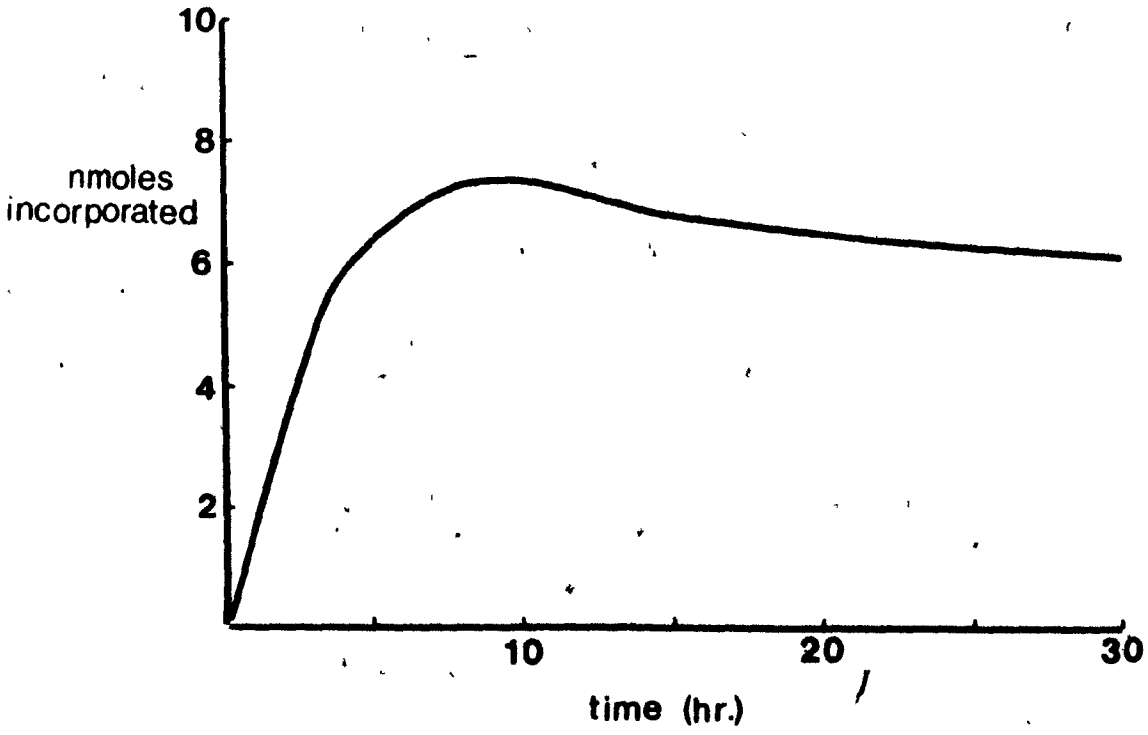


Figure 3.25: Incorporation of radioactive label from [γ-^{32}P]ATP into an excess of oligonucleotides mediated by polynucleotide kinase.

Incubation was at 22°C in a total volume of 0.5 ml of 10 mM Tris-HCl buffer, pH 8.1, 10 mM $MgCl_2$, 20 mM mercaptoethanol containing 10 nmoles [γ-^{32}P]ATP, 15 nmoles of mixed dephosphorylated pyrimidine oligonucleotides longer than hexanucleotides, and 0.025 units of polynucleotide kinase. Aliquots were removed at selected time intervals and incorporation measured by binding the labelled oligonucleotides to a DEAE cellulose suspension, removing ATP by washing with 0.18 M NaCl on a glass fiber filter, and counting the DEAE suspension.

Each of these oligonucleotide fractions was desalted (2.2.9.1), dephosphorylated (2.2.3), and 100 pmoles labelled at the 5' terminus using polynucleotide kinase and [γ-^{32}P]ATP (2.2.5.3.). Each of the labelled oligonucleotides was separated from excess ATP by chain length chromatography on DEAE Sephadex (2.2.9.3), desalted again, and concentrated. This was followed by partial digestion with venom exonuclease (2.2.1) and fractionation using electrophoresis-homochromatography (2.2.9.4). Complete sequence information was obtained for all the above nucleotides from the autoradiographs of the TLC plates from these fractionations and from 5' terminal mononucleotide analyses.

3.6.1 Oligonucleotide C_6T_3

The autoradiograph from the venom exonuclease partial digest of [5'-^{32}P] C_6T_3 (Figure 3.26) contained spots representing C_6T_3 and all 8 of its labelled digest products. These show that the digest from the 3' end proceeded :

$$C_6T_3 \rightarrow C_5T_3 \rightarrow C_4T_3 \rightarrow C_3T_3 \rightarrow C_2T_3 \rightarrow C_2T_2 \rightarrow C_2T \rightarrow CT \rightarrow T$$

giving the total sequence TCCTTCCCC.

3.6.2 Oligonucleotide C_5T_4

The autoradiograph from the venom exonuclease partial digest of [5'-^{32}P] C_5T_4 (Figure 3.27) contained spots representing C_5T_4 and all 6 of its labelled digest products. These showed that the digest from the 3' end proceeded :

$$C_5T_4 \rightarrow C_5T_3 \rightarrow C_4T_3 \rightarrow C_3T_3 \rightarrow C_3T_2 \rightarrow C_2T_2 \rightarrow CT_2$$

giving the 3' partial sequence (T_2C)CCTCCT.

A sample of [5'-^{32}P] C_5T_4 was completely digested with venom exonuclease

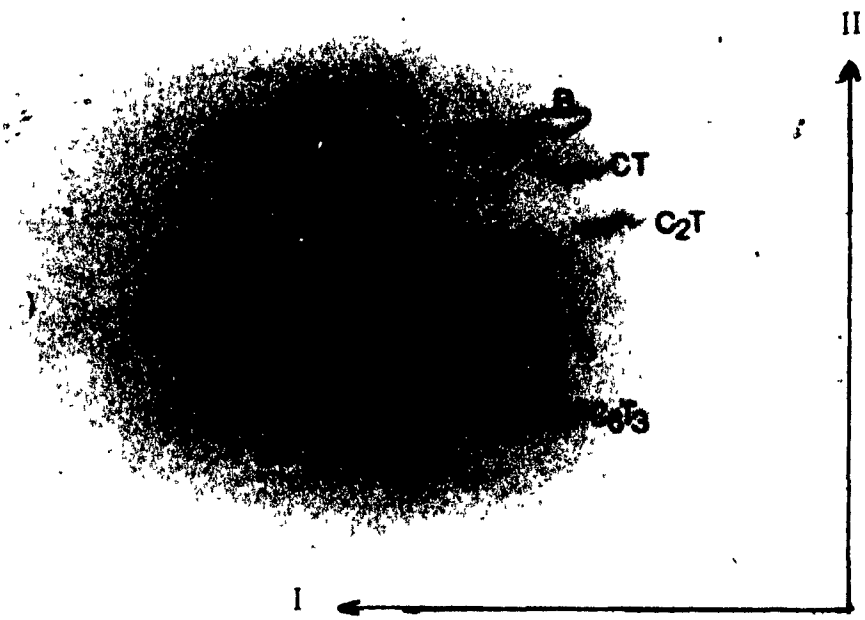


Figure 3.26 : Autoradiograph of the fractionation by electrophoresis-homochromatography of a venom exonuclease partial digest of [5'-^{32}P] C_6T_3.

100 pmoles of oligonucleotide was labelled using polynucleotide kinase and [γ-^{32}P] ATP (2.2.5.3.) ; excess ATP and phosphate was removed by chain length fractionation on DEAE Sephadex (2.2.9.3.) ; the oligonucleotide was desalted (2.2.9.1.) and a portion partially digested with venom exonuclease (2.2.1.) and fractionated by electrophoresis homochromatography. Blue and yellow markers are represented by "B" and "Y".
Legend as for Figure 3.4. 15,000 cpm were used for this fractionation ; autoradiography exposure time was 18 hours.

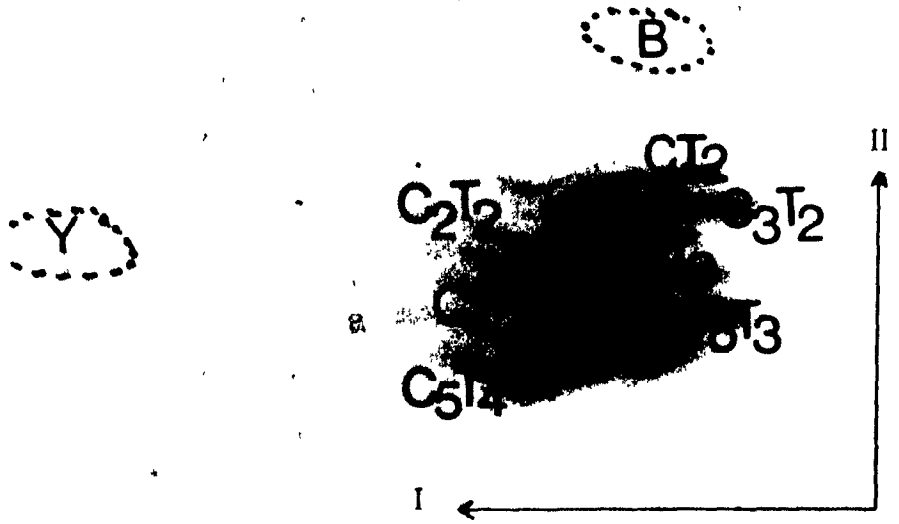


Figure 3.27 : Autoradiograph of the fractionation by electrophoresis-homochromatography of a venom exonuclease partial digest of [5'-^{32}P] C_5T_4.

Legend as for Figure 3.26 . 23,000 cpm were used for this fractionation ; autoradiography exposure was 12 hours. Blue and yellow markers are represented by "B" and "Y".

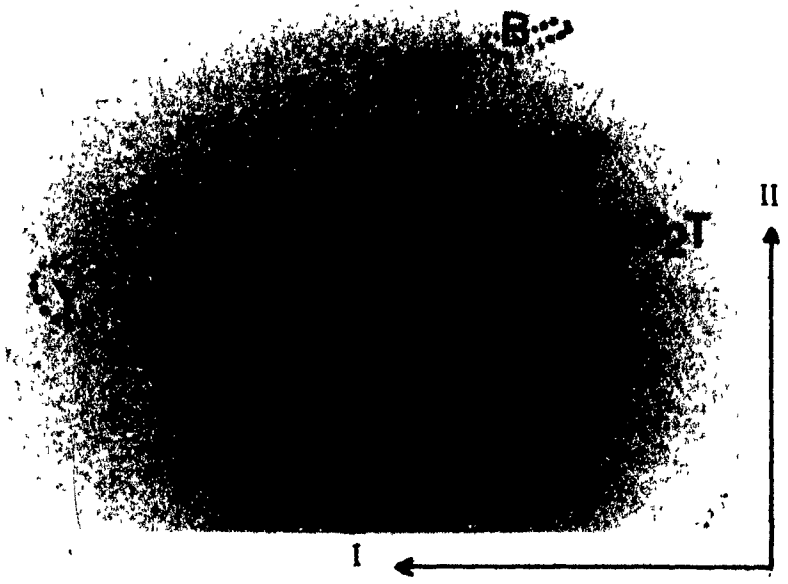


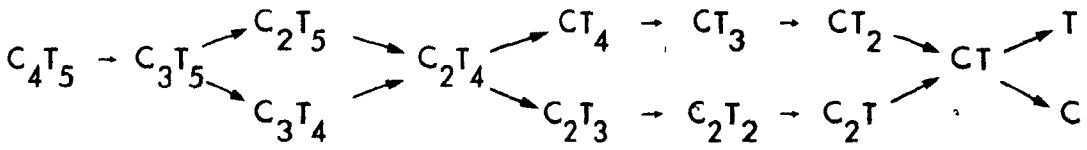
Figure 3.28 : Autoradiograph of the fractionation by electrophoresis-homochromatography of a venom exonuclease partial digest of [5'-^{32}P] C_4T_5.

Legend as for Figure 3.26 . 200,000 cpm were used for this fractionation ; autoradiography exposure was 2 hours. Blue and yellow markers are represented by "B" and "Y".

(2.2.1) followed by mononucleotide analysis . This indicated that the 5'-terminus of C_5T_4 was a cytidylate residue. This additional data yielded the total sequence CTTCCTCCT.

3.6.3 Oligonucleotides C_4T_5

The autoradiograph from the venom exonuclease digest of [5'-^{32}P] C_4T_5 (Figure 3.28) contained one spot representing the two oligonucleotides in this fraction and thirteen spots representing digest products. These showed that the digest from the 3' end proceeded :



giving the partial sequences :

CT	TTTC	TCC
TC	CTTT	CTC

More information was required to determine which partial sequences should be coupled to give two independent complete sequences.

Oligonucleotides C_3T_4 and C_2T_5 were eluted from the TLC plate of the above fractionation ; partial venom exonuclease digestion of C_3T_4 yielded no information but the autoradiograph from the digest of C_2T_5 (Figure 3.29) contained spots representing C_2T_5 and four digest products. These showed that the digest from the 3' end proceeded :

$$C_2T_5 \rightarrow C_2T_4 \rightarrow CT_4 \rightarrow CT_3 \rightarrow CT_2$$

giving the 3' partial sequence (CT_2)TTCT.

Trinucleotides CT_2 and C_2T were eluted from the original TLC plate and

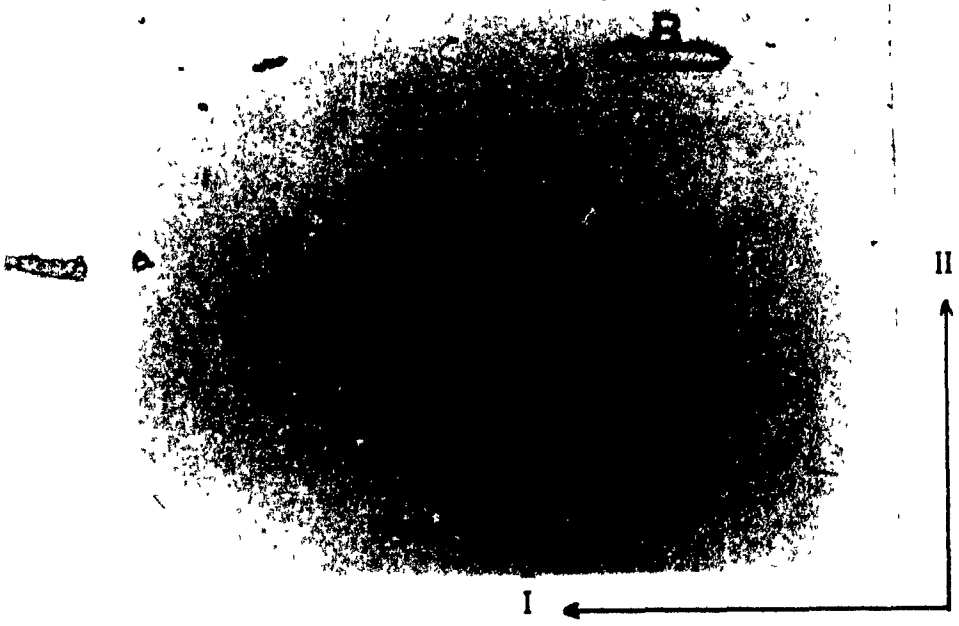


Figure 3.29 : Autoradiograph of the fractionation by electrophoresis-homochromatography of a venom exonuclease partial digest of [5'-^{32}P] C_2T_5 isolated from [5'-^{32}P] C_4T_5.

[5'-^{32}P] C_2T_5 was eluted from the TLC plate (Figure 3.28), evaporated to dryness, partially digested with venom exonuclease (2.2.1.) and fractionated using electrophoresis-homochromatography. Blue and yellow markers are represented by "B" and "Y".
Legend as for Figure 3.4. 2,000 cpm were used for this fractionation ; autoradiography exposure time was 6 days.

their 5' terminal nucleotides identified by complete digestion with venom exonuclease (2.2.1) and mononucleotide analysis (2.2.9.5). CT_2 was found to have a thymidylate residue at its 5' terminus, and C_2T a cytidylate residue.

This additional information showed that the digest of C_4T_5 from the 3' end proceeded :

$$C_4T_5 \rightarrow \begin{cases} C_2T_5 \rightarrow C_2T_4 \rightarrow CT_4 \rightarrow CT_3 \rightarrow CT_2 \rightarrow CT \rightarrow T \\ C_3T_4 \rightarrow C_2T_4 \rightarrow C_2T_3 \rightarrow C_2T_2 \rightarrow C_2T \rightarrow CT \rightarrow C \end{cases}$$

giving the two sequences CTCTTTCTC and TCTTTCTCC.

3.6.4 Oligonucleotide C_6T_4

The autoradiograph from the venom exonuclease digest of [5'-^{32}P] C_6T_4 (Figure 3.30) contained spots representing C_6T_4 and all nine labelled digest products. These showed that digest from the 3' terminus proceeded :

$$C_6T_4 \rightarrow C_5T_4 \rightarrow C_4T_4 \rightarrow C_4T_3 \rightarrow C_4T_2 \rightarrow C_3T_2 \rightarrow C_3T \rightarrow C_2T \rightarrow CT \rightarrow C$$

giving the sequence CTCCTCTTCC.

3.6.5 Oligonucleotide C_2T_8

The autoradiograph from the venom exonuclease digest of [5'-^{32}P] C_2T_8 (Figure 3.31) contained spots representing C_2T_8 and six labelled digest products. These showed that the digest from the 3' end proceeded :

$$C_2T_8 \rightarrow C_2T_7 \rightarrow CT_7 \rightarrow CT_6 \rightarrow CT_5 \rightarrow CT_4 \rightarrow CT_3$$

giving the 3' partial sequence (CT_3) TTTTCT.

A sample of [5' - ^{32}P] C_2T_8 was completely digested with venom exonuclease (2.2.1) followed by mononucleotide analysis. This indicated that the 5' terminus of

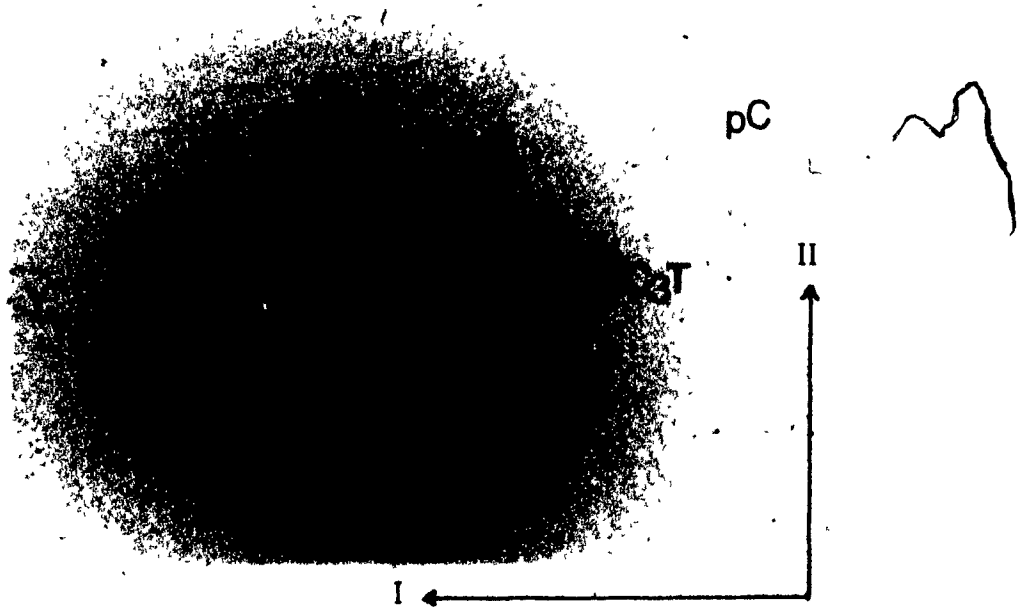


Figure 3.30 : Autoradiograph of the fractionation by electrophoresis-homochromatography of a venom exonuclease partial digest of [5'-^{32}P] C_6T_4.

Legend as for Figure 3.26. 15,000 cpm were used for this fractionation ; autoradiography exposure was 18 hours. Blue and yellow markers are represented by "B" and "Y".

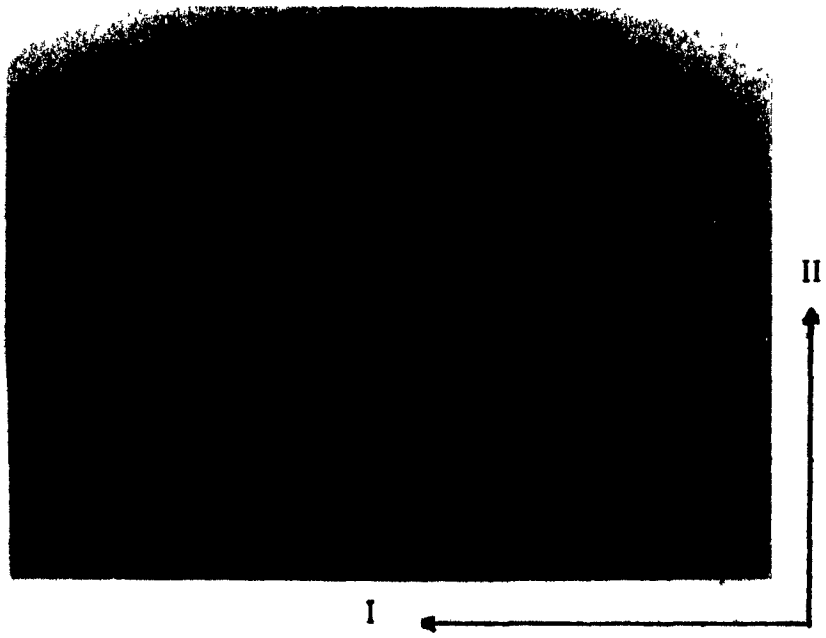


Figure 3.31 : Autoradiograph of the fractionation by electrophoresis-homochromatography of a venom exonuclease partial digest of [5'-^{32}P] C_2T_8.

Legend as for Figure 3.26. 11,000 cpm were used for this fractionation ; autoradiography exposure was 18 hours. Blue and yellow markers are represented by "B" and "Y".

C_2T_8 was a cytidylate residue. This information, along with the base composition gave the total sequence : CTTTTTTTCT.

3.5.2.6 Oligonucleotide C_5T_6

The autoradiograph from the venom exonuclease digest of [5' -^{32}P] C_5T_6 (Figure 3.32) contained spots representing C_5T_6 and all ten labelled digest products. These showed that the digest from the 3' end proceeded :

$C_5T_6 \rightarrow C_5T_5 \rightarrow C_4T_5 \rightarrow C_4T_4 \rightarrow C_4T_3 \rightarrow C_3T_3 \rightarrow C_3T_2 \rightarrow C_2T_2 \rightarrow CT_2 \rightarrow CT \rightarrow C$

giving the total sequence CTTCCTCTTCT.

3.6.7 Summary

The sequences determined using 5' terminal labelling followed by venom exonuclease digestion are summarized in Table 3.3.

Table 3.3

Summary of the sequences determined via 5' labelling followed by exonuclease digestion

Isostich	Component	Sequence
11	C_5T_6	CTTCCTCTTCT
10	C_6T_4	CTCCTCTTCC
	C_2T_8	CTTTTTTTCT
9	C_6T_3	TCCTTCCCC
	C_5T_4	CTTCCTCCT
	C_4T_5	TCTTTCTCC
		CTCTTTCTC

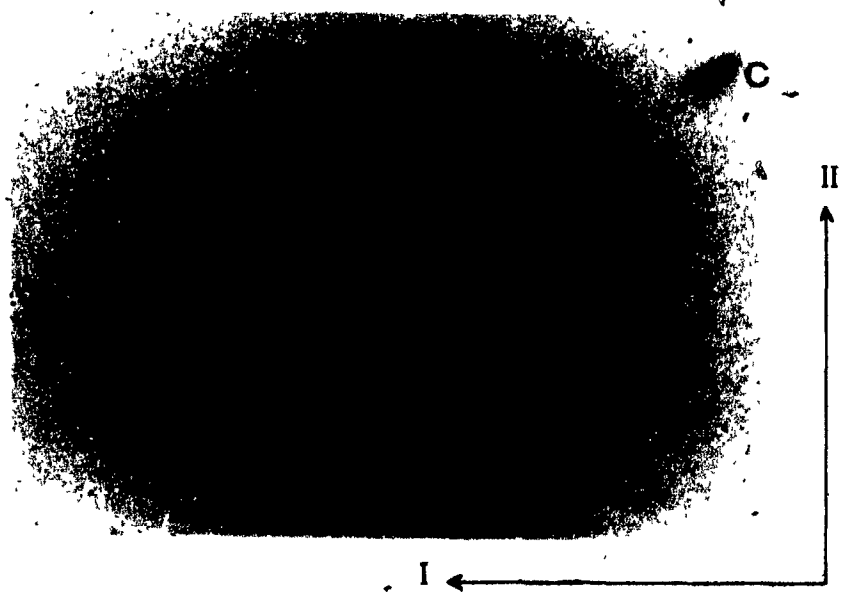


Figure 3.32 : Autoradiograph of the fractionation by electrophoresis-homochromatography of a venom exonuclease partial digest of [5'-^{32}P] C_5T_6.

Legend as for Figure 3.26. 30,000 cpm were used for this fractionation ; autoradiography exposure was 18 hours. Blue and yellow markers are represented by "B" and "Y".

3.7 Sequences Determined via 5' Exonuclease Degradation Followed by 5' Terminal Labelling

The methods discussed above involved the determination of base composition from relative migration in the electrophoresis-homochromatography system; this was simple for pyrimidine oligonucleotides and is theoretically possible for short oligonucleotides containing all four bases (69). The determination of sequence oligonucleotides containing all four bases longer than pentanucleotides requires more resolution than is available in this system.

A method for sequence analysis of oligonucleotides which circumvents this problem has been developed. The oligonucleotide under study was dephosphorylated and partially digested with spleen exonuclease and the resulting digest labelled using polynucleotide kinase and [γ-^{32}P] ATP. Fractionation of the labelled digest on a Sephadex G-25 column followed by DEAE Sephadex chromatography or by electrophoresis-homochromatography separated the labelled digest products from each other and from ATP (2.2.10.3). Identification of the 5' terminus of each of the digest products by complete venom exonuclease digestion and mononucleotide analysis yielded the sequence of the oligonucleotide except for the 3' terminus. This terminal sequence was determined by chromatographic base composition analysis of the dinucleotide digest product or by sequence analysis (2.2.10.2) of the tri- or tetranucleotide digest products.

The nona-, deca-, and undecanucleotides isolated as described in Section 3.6 from S13$^+$ DNA were examined in this manner. Except for C_4T_5, 100 pmoles of oligonucleotide were used in each experiment.

3.7.1 Oligonucleotide C_6T_4

The eluate from the Sephadex G-25 column (Figure 3.33(A)) was divided into the long oligonucleotide fraction and the dinucleotide containing fraction, as shown by the bars in the diagram. The dinucleotide fraction was combined with marker oligonucleotides and further purified by DEAE Sephadex chain length chromatography (Figure 3.33 B) . The dinucleotides from this fractionation were chromatographed on DEAE cellulose at pH 3.4 (Figure 3.33 C) and the radioactivity was observed to elute with marker pCpC , giving a 3' partial sequence of $(C_4T_4)CC$.

The long oligonucleotide fraction from the Sephadex G-25 column was fractionated according to chain length on a DEAE Sephadex column (Figure 3.33 D) and peaks containing [5' - ^{32}P] C_6T_4 and seven digest products were separated. These were individually desalted and a portion subjected to complete venom exonuclease digestion and mononucleotide analysis. The 5' termini deduced from these analyses (Table 3.4) coupled with the 3' terminal analysis above, gave the sequence CTCCTCTTCC.

3.7.2 Oligonucleotide C_2T_8

The labelled spleen exonuclease partial digest of C_2T_8 was fractionated (Figure 3.34) as was described for C_6T_4 (3.7.1). In this case the dinucleotide eluted from the DEAE cellulose column (Figure 3.34 C) with marker pCpT. The chain length fractionation (Figure 3.34 (D)) separated [5'-^{32}P] C_2T_8 and seven digest products. The analyses of the 5' termini of these fractions, as well as that of the dinucleotide (Table 3.5) gave the sequence CTTTTTTTCT.

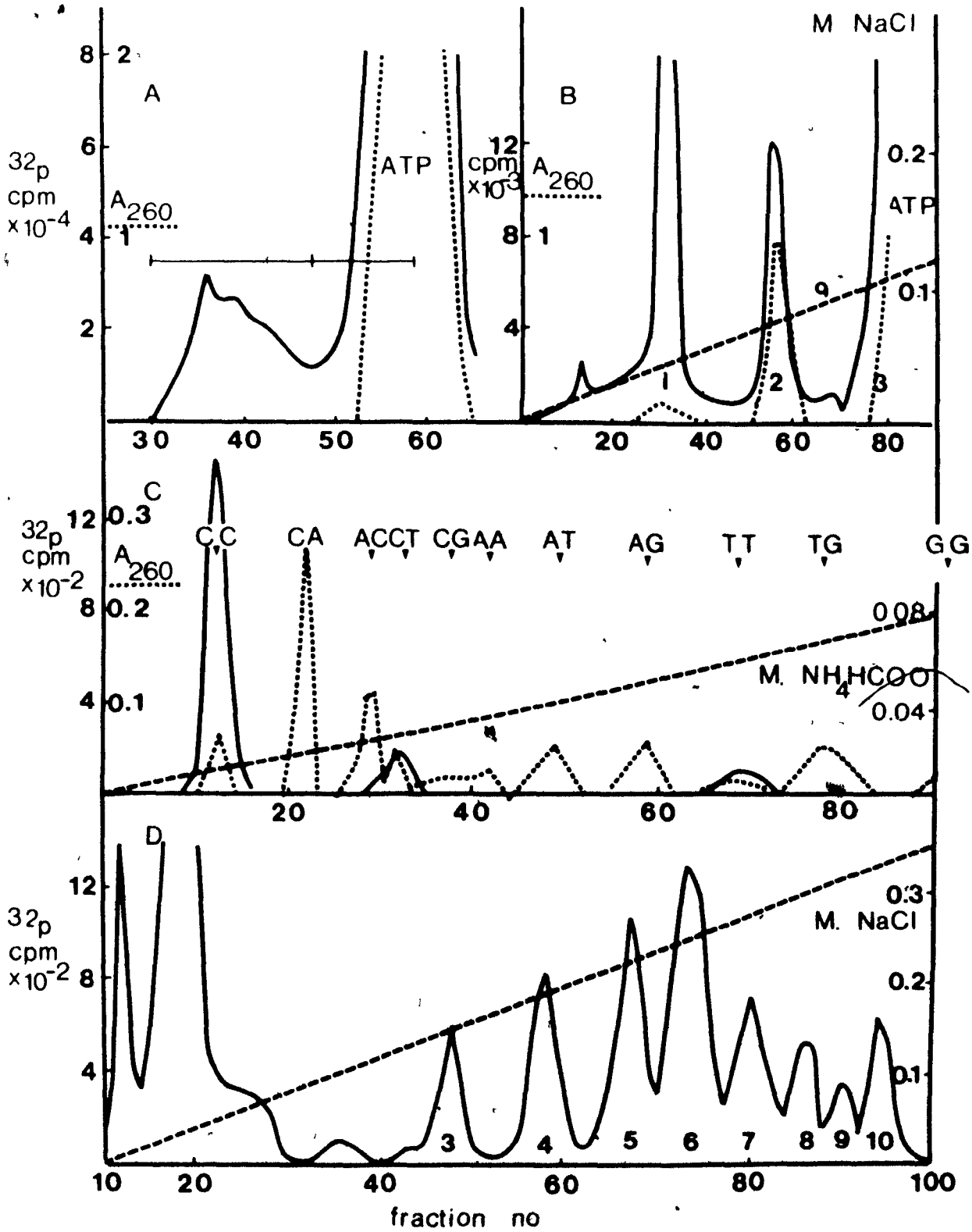


Figure 3 33 Fractionations in the analysis of the sequence of oligonucleotide C_6T_4

Legend on the following page.

Figure 3.33 :, Fractionations in the analysis of the sequence of oligonucleotide C_6T_4.

100 pmoles of oligonucleotide was partially digested with spleen exonuclease in a volume of 0.3 ml, the digest dephosphorylated, phosphomonoesterase destroyed with alkali, the digest neutralized and labelled with polynucleotide kinase and [γ -^{32}P] ATP (2.2.10.3.).

(A) Elution profile from Sephadex G-25 of the labelled spleen exonuclease digest of the oligonucleotide (2.2.10.3.) after incubation with polynucleotide kinase and [γ-^{32}P]ATP. The bars represent the fractions pooled, the long oligonucleotide fraction eluted first, followed by the dinucleotide containing fraction ; column dimensions were 1 cm x 50 cm ; eluate was 0.05 M NaCl ; flow rate was 0.3 ml/min, and 0.6 ml fractions were collected. The solid line represents the elution of radioactivity, the dotted line the elution of optical density of marker ATP.

(B) Elution profile of the chain length fractionation on DEAE Sephadex of the dinucleotide containing fraction from (A). Column dimensions were 1 cm x 25 cm, load and wash solution was water ; eluate was 2 l of 50 mM sodium acetate buffer, pH 5.5, 1 mM NaH_2PO_4 , 7 M urea, with a gradient from 0 to 0.2 M NaCl ; flow rate was 0.6 ml/min, and 14 ml fractions were collected. The solid line represents the elution of radioactivity, the dotted line the elution of oligonucleotide marker.

(C) Elution profile of a fractionation at pH 3.2 on DEAE cellulose of the dinucleotide peak from (B). Column dimensions were 1 cm x 25 cm, load and wash solution was water, eluate was a 2 l gradient from 0 to 0.1 M ammonium formate buffer, pH 3.4 ; flow rate was 1 ml/min and 17 ml fractions were collected. The solid line represents the elution of radioactivity, the dotted line the elution of marker dinucleotide optical density. The arrows indicate the elution positions of individual dinucleotides.

(D) Elution profile of the chain length fractionation on DEAE Sephadex of the long oligonucleotide fraction from (A). Column dimensions were 1 cm x 25 cm, load and wash solution was water, eluate was 2 l of 50 mM sodium acetate buffer, pH 5.5, 1 mM NaH_2PO_4, 7 M urea, containing a gradient from 0 to 0.4 M NaCl ; flow rate was 0.6 ml/min and 17 ml fractions were collected. The numbers indicate the phosphate length of the individual fractions.

Table 3.4

5' Terminal Mononucleotide Analyses: Sequence of C_6T_4

Position	Chain length fraction	cpm for analysis	Percent* T	G	C	A	Assignment
8	3	1100	99	-	1	-	T
7	4	1900	96	-	4	-	T
6	5	2000	11	-	87	-	C
5	6	3000	87	4	8	-	T
4	7	1400	9	-	91	-	C
3	8	620	16	-	84	-	C
2	9	460	75	-	25	-	T
1 (5' end)	10	900	12	-	86	-	C

* a dash indicates that less than 2% of the radioactivity eluted with that mononucleotide.

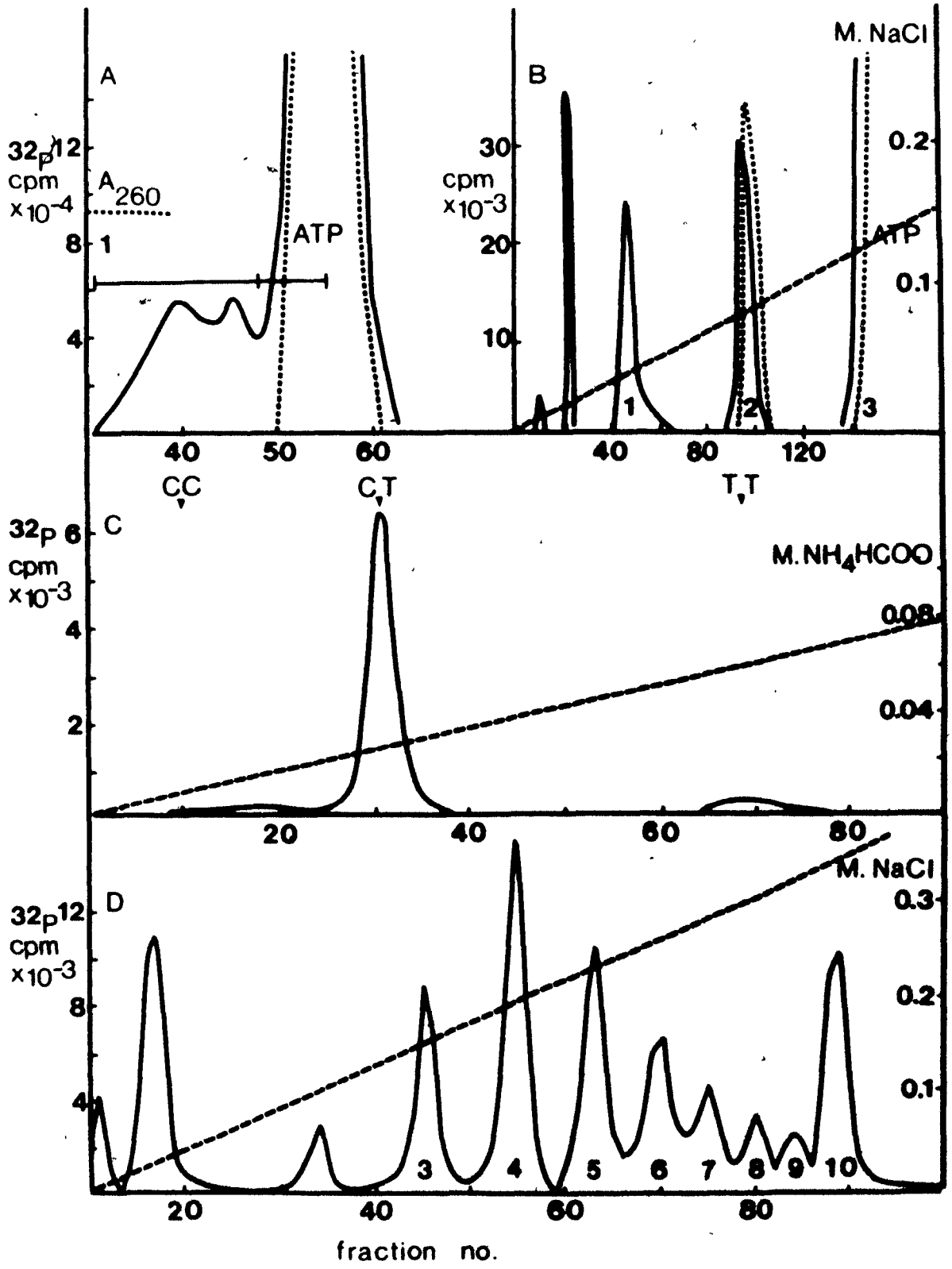


Figure 3 34 · Fractionations in the analysis of the sequence of oligonucleotide C_2T_8.

Legend as for Figure 3.33.

Table 3.5

5' Terminal Mononucleotide Analyses: Sequence of C_2T_8

Position	Chain length fraction	cpm for analysis	Percent * T	G	C	A	Assignment
9	2	39,500	2	-	97	-	C
8	3	32,000	95	-	5	-	T
7	4	41,000	96	-	4	-	T
6	5	34,000	96	-	3	-	T
5	6	18,000	97	-	3	-	T
4	7	10,300	93	-	5	-	T
3	8	9,900	92	2	5	-	T
2	9	5,100	89	-	11	-	T
1 (5' end)	10	32,000	5	-	95	-	C

* a dash indicates that less than 2% of the radioactivity eluted with that mononucleotide.

3.7.3 Oligonucleotide C_5T_6

The labelled spleen exonuclease digest of C_5T_6 was fractionated (Figure 3.35) as was described for C_6T_4 (3.7.1). In this case the dinucleotide eluted from the DEAE cellulose column (Figure 3.35(C)) with marker pCpT. The chain length fractionation (Figure 3.35(D)) separated [5' -^{32}P] C_5T_6 and eight labelled digest products. The analyses of the 5' termini of each of these fractions, as well as that of the dinucleotide, (Table 3.6) gave the sequence CTTCCTCTTCT.

3.7.4 Oligonucleotide C_6T_3

The labelled spleen exonuclease digest of C_6T_3 was fractionated (Figure 3.36) as was described for C_6T_4 (3.7.1). In this case the dinucleotide eluted from the DEAE cellulose column (Figure 3.36(C))with marker pCpC. The chain length fractionation (Figure 3.36 (D))separated [5' -^{32}P] C_6T_3 and six labelled digest products. The analyses of the 5' termini of each of these fractions, (Table 3.7) coupled with the dinucleotide analysis above, yielded the complete sequence TCCTTCCCC.

3.7.5 Oligonucleotide C_4T_5

The analysis of the oligonucleotide fraction C_4T_5 from S13$^+$ DNA by 5' exonuclease treatment followed by 5' terminal labelling was quite complex because two isomers of different sequences occurred in this fraction.

400 pmoles of dephosphorylated C_4T_5 were partially digested with spleen exonuclease for only 2 minutes instead of the usual 10 minutes (2.2.2) to obtain a high yield of the hepta- and octanucleotide digest products. This digest was

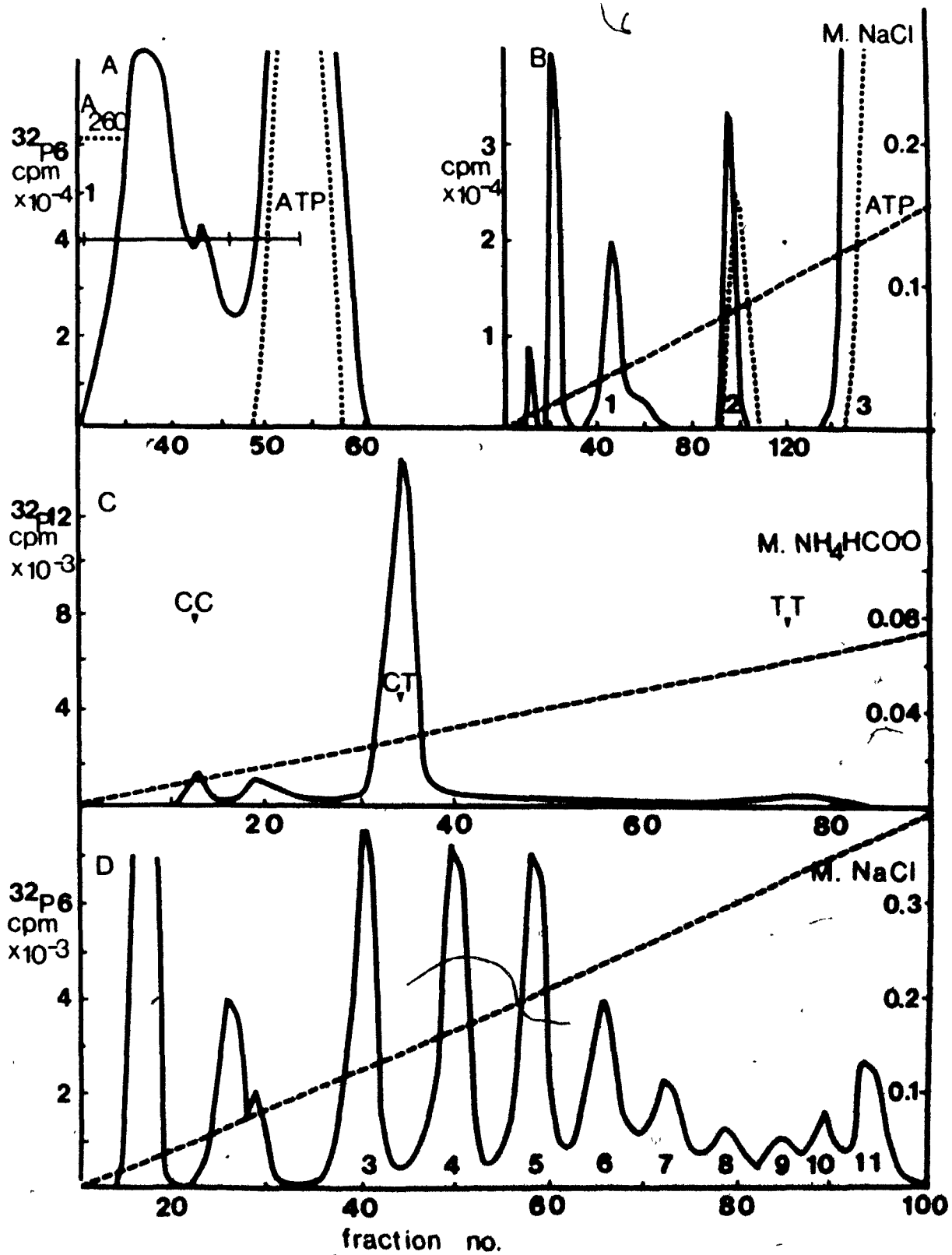


Figure 3.35 : Fractionations in the analysis of the sequence of oligonucleotide C_5T_6.

Legend as for Figure 3.33.

Table 3.6

5' Terminal Mononucleotide Analyses: Sequence of C_5T_6

Position	Chain length fraction	cpm for analysis	Percent *				Assignment
			T	G	C	A	
10	2	36,000	-	-	98	-	C
9	3	8,200	95	2	2	-	T
8	4	6,700	93	2	3	-	T
7	5	5,900	8	-	92	-	C
6	6	3,600	83	3	18	-	T
5	7	1,150	21	-	79	-	C
4	8	1,130	11	-	87	-	C
3	9	500	85	-	15	-	T
2	10	850	89	-	11	-	T
1 (5' end)	11	1,900	-	-	99	-	C

* a dash indicates that less than 2% of the radioactivity eluted with that mononucleotide.

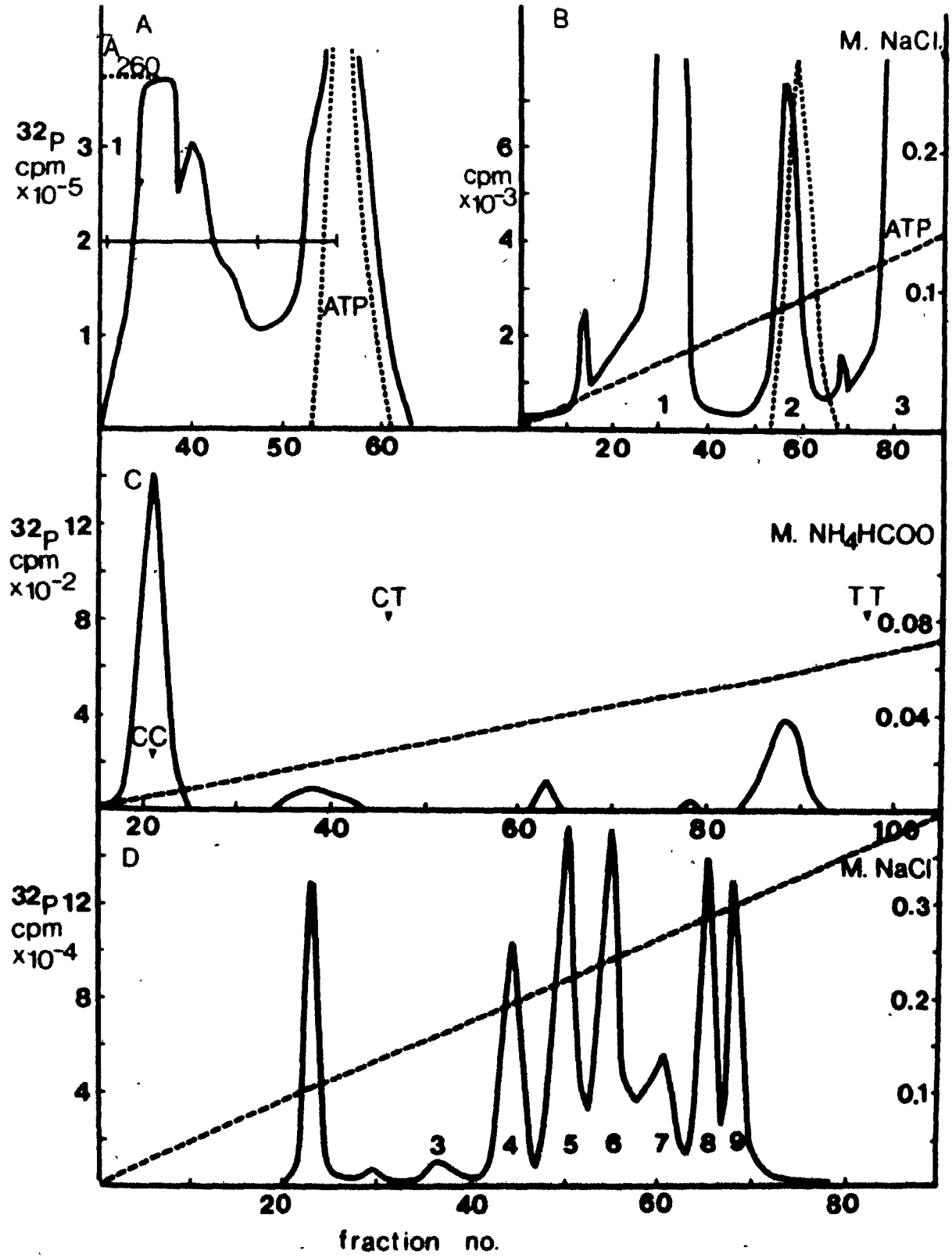


Figure 3.36 : Fractionations in the analysis of the sequence of oligonucleotide C_6T_3.

Legend as for Figure 3.33.

Table 3.7

5' Terminal Mononucleotide Analyses: Sequence of C_6T_3 *

Position	Chain length fraction	cpm for analysis	Percent* T	G	C	A	Assignment
7	3	15,000	30	-	69	-	C
6	4	48,000	13	3	84	-	C
5	5	75,000	99	-	-	-	T
4	6	106,000	98	-	2	-	T
3	7	35,000	25	2	73	-	C
2	8	76,000	10	-	90	-	C
1 (5' end)	9	89,000	99	-	-	-	T

* a dash indicates that less than 2% of the radioactivity eluted with that mononucleotide.

labelled using polynucleotide kinase and [γ - ^{32}P] ATP (2.2.5.3) and the excess ATP removed by gel filtration on a Sephadex G-25 column (Figure 3.37(A)). These labelled digest products were fractionated according to chain length on a DEAE Sephadex column (Figure 3.37(B)) and fractions corresponding to C_4T_5 and 4 digest products were separated. Complete venom exonuclease digestion and mononucleotide analysis of the nonanucleotide fraction showed that the 5' termini were heterogeneous, indicating that the octanucleotide spleen digest product must contain the non-isomeric species C_4T_4 and C_3T_5. Thus the octanucleotide peak was pooled and fractionated at pH 3.2 on a DEAE Sephadex column (Figure 3.37(C)) and the two components separated.

Each component was desalted, dephosphorylated, partially digested with spleen exonuclease, labelled again with polynucleotide kinase and [γ - ^{32}P] ATP, and then fractionated (Figures 3.38 and 3.39) as was described for oligonucleotide C_6T_4. In these cases the dinucleotide analyses did not yield any useful information. The chain length fractionation of the labelled digest of C_4T_4 (Figure 3.38(B)) separated [5' - ^{32}P] C_4T_4 and five digest product fractions. 5' terminal analyses of the tri- and tetranucleotide fractions yielded equivalent amounts of C and T, while analysis of the other products (Table 3.8) gave the tentative 5' partial sequence CTTT (C_3T).

The chain length fractionation of the labelled digest of [5'-^{32}P] C_3T_5 (Figure 3.39(B)) separated [5'-^{32}P] C_3T_5 and four digest products. Analysis of the 5' termini of these fractions (Table 3.9) gave the 5' partial sequence TCTT (C_2T_2).

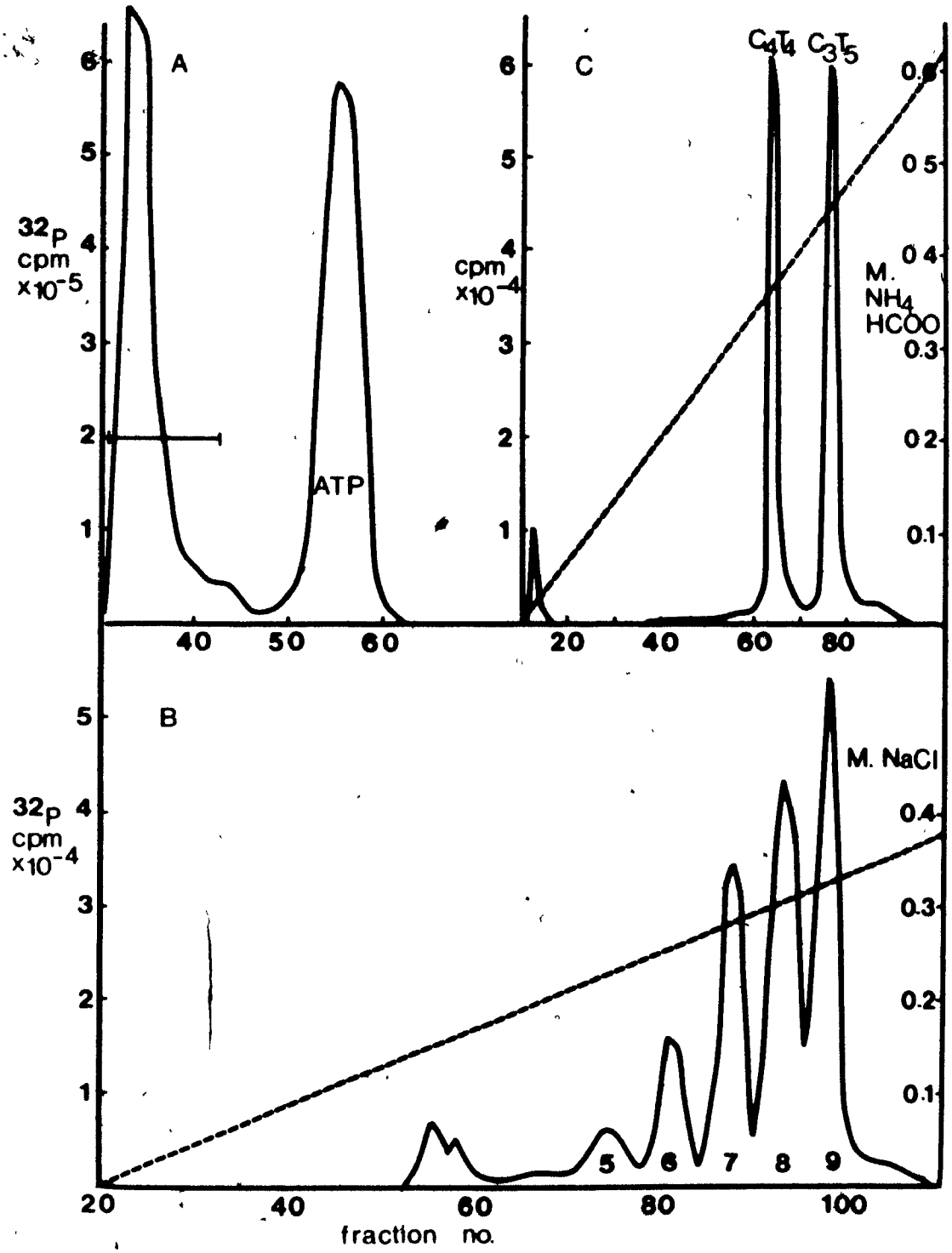


Figure 3.37 Separation of the octanucleotides in a labelled spleen exonuclease digest of nonanucleotides C_4T_5

(A) Separation of the labelled digest products from excess ATP on a Sephadex G-25 column as described for Figure 3.33 (A).

(B) Chain length fractionation of the oligonucleotides isolated in (A) as described for Figure 3.33 (D).

(C) Fractionation at pH 3.2 on DEAE Sephadex of the octanucleotides isolated in (B). Column dimensions were 1 cm x 25 cm, load and wash solutions were water, eluate was a 2 l gradient from 0 to 1.0 M ammonium formate buffer, pH 3.2, flow rate was 0.8 ml/min and 12 ml fractions were collected.

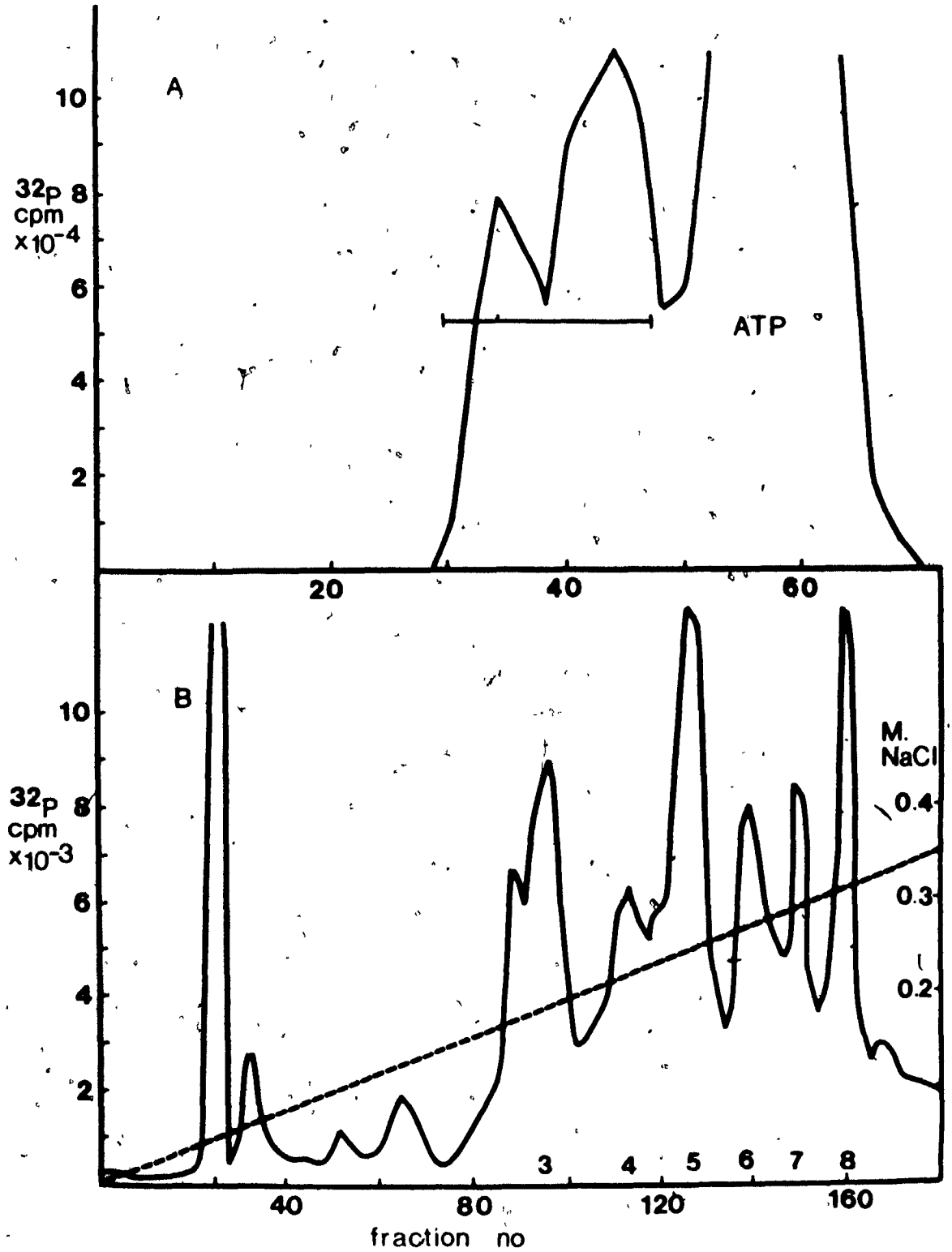


Figure 3 38 · Fractionations in the analysis of the sequence of oligonucleotide C_4T_4 from C_4T_5

(A) Separation of the long oligonucleotide fraction from ATP on a Sephadex G-25 column as described for Figure 3.33 (A).

(B) Chain length fractionation of the oligonucleotides isolated in (A) as described for Figure 3.33 (D).

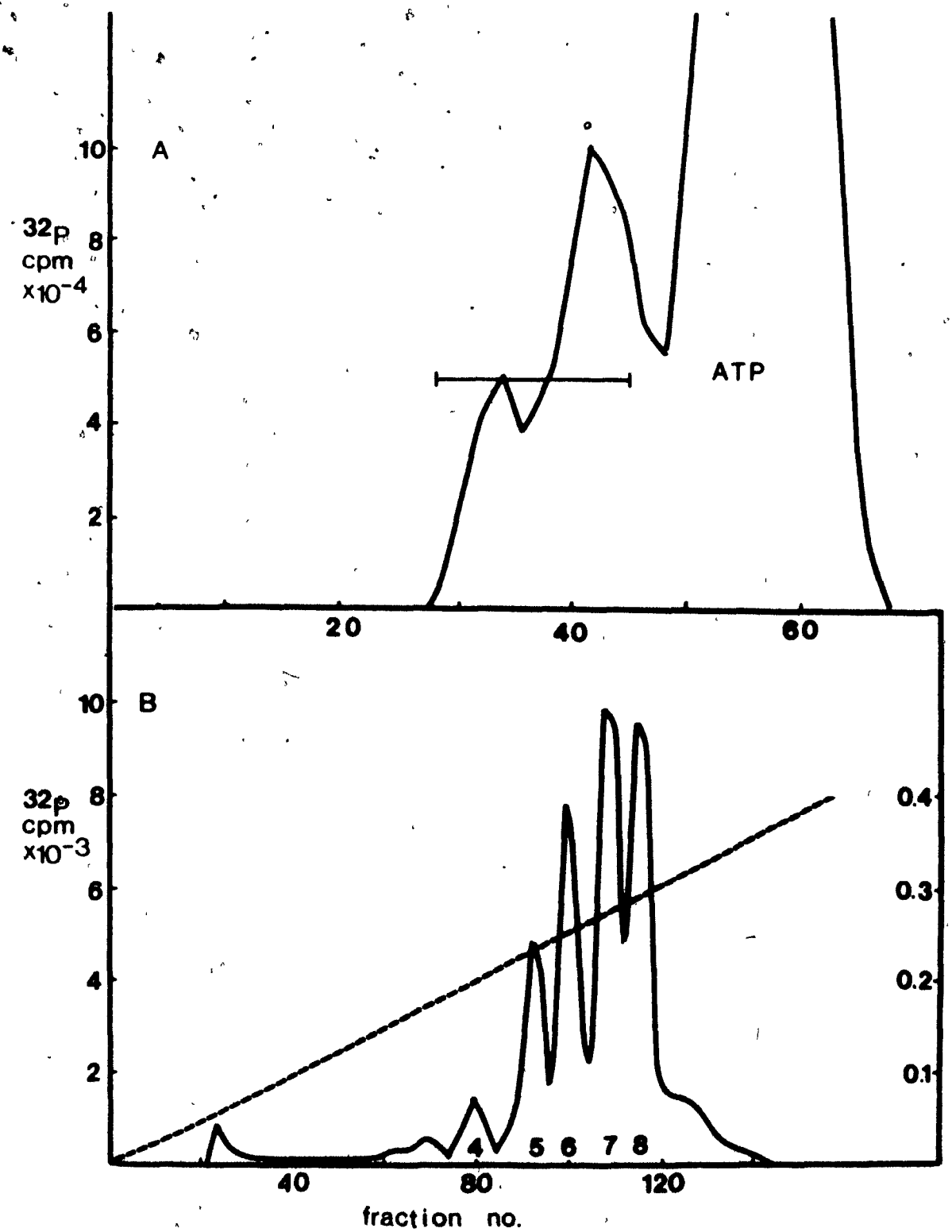


Figure 3.39 : Fractionations in the analysis of the sequence of oligonucleotide C_3T_5 from C_4T_5.

(A) Separation of the long oligonucleotide fraction from ATP on a Sephadex G-25 column as described for Figure 3.33 (A).

(B) Chain length fractionation of the oligonucleotides isolated in (A) as described for Figure 3.33 (D).

Table 3.8

5' Terminal Mononucleotide Analyses: Sequence of C_4T_4 from C_4T_5

Position	Chain length fraction	cpm for analysis	Percent * T	G	C	A	Assignment
4	5	67,000	66	-	33	-	T
3	6	29,000	93	-	6	-	T
2	7	23,000	75	-	25	-	T
1 (5' end)	8	38,000	38	-	62	-	C

Table 3.9

5' Terminal Mononucleotide Analyses: Sequence of C_3T_5 from C_4T_5

Position	Chain length fraction	cpm for analysis	Percent * T	G	C	A	Assignment
4	5	9,000	92	2	5	-	T
3	6	15,000	95	-	4	-	T
2	7	12,000	27	-	73	-	C
1 (5' end)	8	20,000	98	-	2	-	T

* a dash indicates that less than 2% of the radioactivity eluted with that mononucleotide.

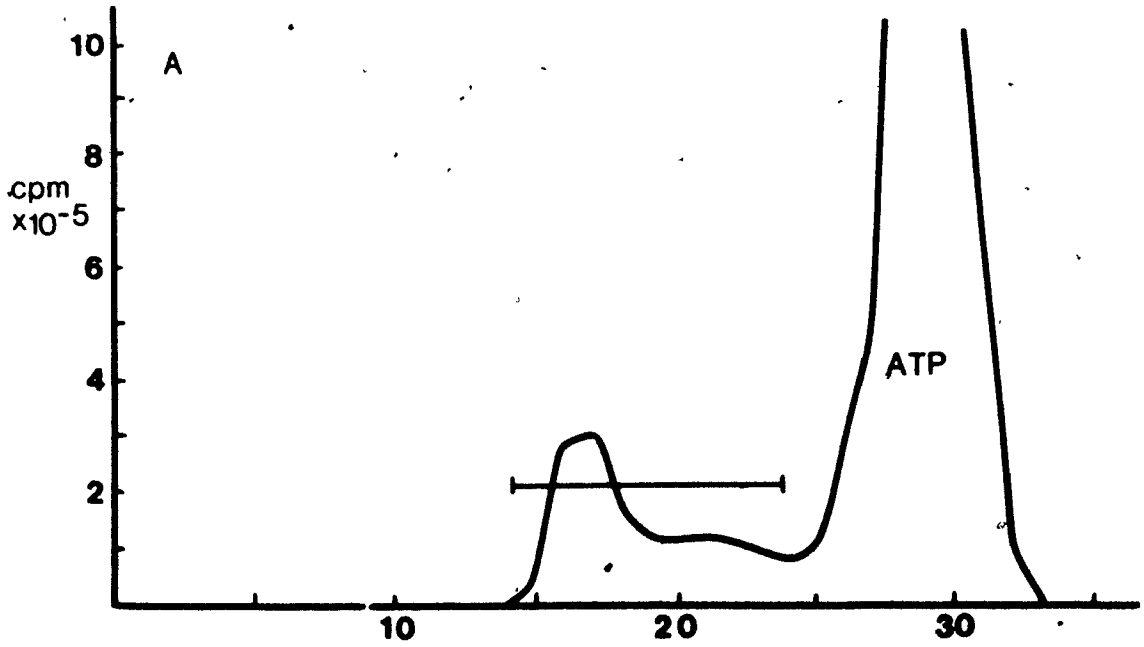
These two partial sequences yield 5' partial sequences of CTCTT (C_2T_2) and TCTTT (C_3T) for the parent oligonucleotides.

The mononucleotide analyses in the sequence determination of C_4T_4 indicated that the C_4T_4 was not homogeneous. This showed that the phosphorylation of the first partial digest of C_4T_5 was not complete and that, as expected, unphosphorylated C_4T_5 eluted from the base composition column with 5' phosphorylated C_4T_4. The level of interference in the 7 and 8 position of the C_4T_4 analysis indicate that the initial phosphorylation had only proceeded to about 50% of maximum. This limitation of the system occurs only when two successive labelling reactions must be done, and if a great excess of ATP could be used in the first step, the limitation would not apply.

As could be predicted, the C_3T_5 analysis was much clearer.

3.7.6 Oligonucleotide C_5T_4

100 pmoles of oligonucleotide C_5T_4 were dephosphorylated, partially digested with spleen exonuclease, labelled with polynucleotide kinase and [γ-^{32}P] ATP, and fractionated by gel filtration on a Sephadex G-25 column (Figure 3.40(A)). The radioactive fractions eluting before the excess ATP were pooled, desalted, and fractionated using electrophoresis-homochromatography. The autoradiograph of this fractionation (Figure 3.40 (B)) contained spots representing C_5T_4 and five digest products. Complete venom exonuclease digestion of a portion of each of these products, followed by mononucleotide analyses (Table 3.10) gave the 5' partial sequence CTTCCT (C_2T), in agreement with the 5' partial sequence which was observed by inspection of the autoradiograph : CTTCC (C_2T_2).



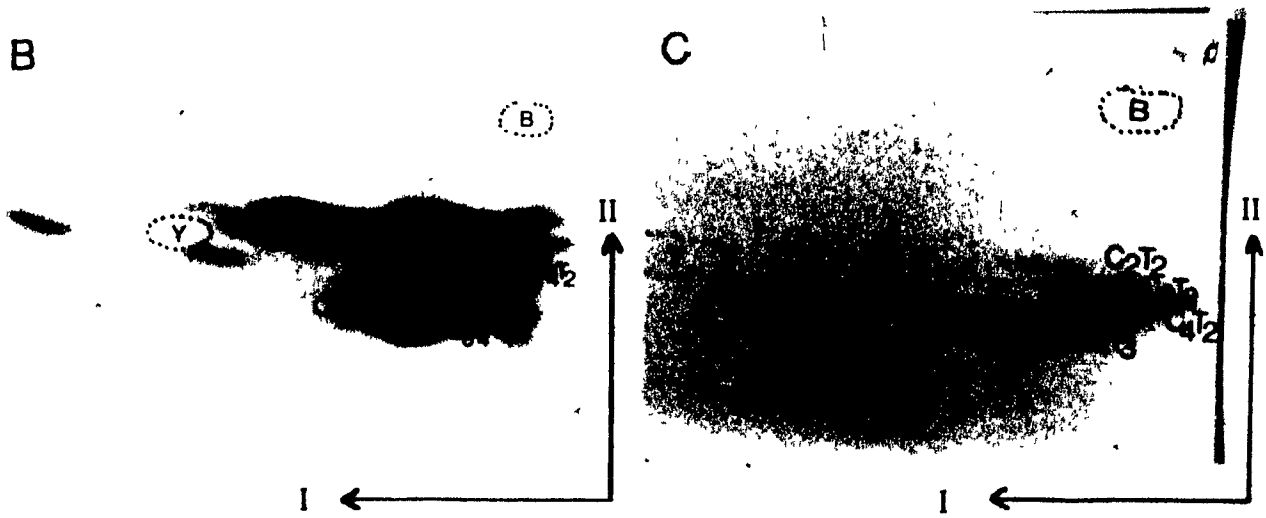


Figure 3.40 : Fractionations in the analysis of the sequence of oligonucleotide C_5T_4.

(A) Separation of the long oligonucleotides in the labelled spleen exonuclease digest from excess ATP on a Sephadex G-25 column as described for Figure 3.33 (A).

(B) Autoradiograph of the fractionation by electrophoresis homochromatography of the long oligonucleotide fraction isolated in (A). Legend as for Figure 3.4.

300,000 cpm were used in this fractionation ; autoradiography exposure time was 4 hours.

(C) Autoradiograph of the fractionation by electrophoresis homochromatography of the venom exonuclease digest of [5'-^{32}P]C_4T_3. Legend as for Figure 3.4.

3,000 cpm were used in this fractionation ; autoradiography exposure time was 3 days.

Table 3.10

5' Terminal Mononucleotide Analyses: Sequence of C_5T_4

Position	Chain length fraction	cpm for analysis	Percent* T	G	C	A	Assignment
6	4	1200	99	-	-	-	T
5	5	800	7	-	92	-	C
4	6	7000	5	-	95	-	C
3	7	8000	95	-	5	-	T
2	8	2600	99	-	-	-	T
1 (5' end)	9	2100	4	-	96	-	C

* a dash indicates that less than 2% of the radioactivity eluted with that mononucleotide.

Another portion of the heptanucleotide C_4T_3 from this fractionation was partially digested with venom exonuclease and fractionated using electrophoresis-homochromatography. The autoradiograph of this fractionation (Figure 3.40 (C)) contained spots representing C_4T_3 and three digest products. These indicated that the digest proceeded :

$$C_4T_3 \rightarrow C_4T_2 \rightarrow C_3T_2 \rightarrow C_2T_2$$

and that the 3' partial sequence is $(C_2T_2)CCT$.

The combination of these partial sequences gave the total sequence CTTCCTCCT.

3.5.3.7 Summary

The sequences determined using spleen exonuclease digestion followed by 5' terminal labelling are summarized in Table 3.11.

3.8 Sequences of Pyrimidine Oligonucleotides from $S13^+$ DNA

Table 3.12 lists all the sequences which have been determined for oligonucleotides from bacteriophage $S13^+$ DNA. The right column gives the methods which have been used for each sequence determination.

Table 3.11

Summary of the sequences determined via exonuclease digestion followed by 5' terminal labelling

Isostich	Component	Sequence
11	C_5T_6	CTTCCTCTTCT
10	C_6T_4	CTCCTCTTCC
	C_2T_8	CTTTTTTTCT
9	C_6T_3	TCCTTCCCC
	C_5T_4	CTTCCTCCT
	C_4T_5	TCTTT(C_3T)
		CTCTT(C_2T_2)

Table 3.12

Pyrimidine Oligonucleotides from Bacteriophage SI3$^+$ DNA

Isostich	Component	Sequence	Methods *
11	C_5T_6	CTTCCTCTTCT	1, 2, 3, 5
10	C_6T_4	CTCCTCTTCC	1, 2, 3, 5
	C_2T_8	CTTTTTTTCT	1, 2, 3, 5
9	C_6T_3	TCCTTCCCC	1, 2, 3, 5
	C_5T_4	CTTCCTCCT	2, 4, 6
	C_4T_5	TCTTTCTCC	1, 2, 3, 5
		CTCTTTCTC	1, 2, 3, 5
8	C_5T_3	CCTTTCCC	1
	C_4T_4	CTTCCTTC	1
	C_2T_6	CTTTTTTC	1
		TTTTCCTT	1
7	C_6T	CTCCCCC	1
	C_5T_2	CTCTCCC	1
	CT_6	CTTTTTT	1

* 1) partial digestion of uniformly labelled material (3.4).
2) partial digestion of 5'-terminally labelled material (3.6).
3) 5' sequence analysis via partial exonuclease digestion followed by terminal labelling and electrophoresis-homochromatography (3.7).
4) 5' sequence analysis via partial exonuclease digestion followed by terminal labelling and electrophoresis-homochromatography (3.7).
5) 3' terminal analysis by dinucleotide identification on columns (3.7).
6) 3' terminal analysis by sequence analysis of a 3' terminal fragment using electrophoresis-homochromatography.

CHAPTER 4

DISCUSSION

4.1 Base Preferences During Spleen Exonuclease Hydrolysis

It has been reported (8) that poly rC was degraded at a negligible rate relative to the degradation of poly rA, poly rI, or poly rU. If such specificity exists for deoxyribonucleotide sequences, it could be very useful in sequence work.

During the course of this study, spleen exonuclease was used to degrade many pyrimidine oligodeoxyribonucleotides. The availability of sequence data for these oligonucleotides allowed the examination of sequence preferences exhibited in these reactions.

When the partial digests were performed at a very low oligonucleotide concentration (1 μM) all of the oligonucleotides tested were hydrolysed in the presence of spleen exonuclease. When the oligonucleotides to be studied were isolated from a preparative electrophoresis-homochromatography fractionation (3, 4), approximately 30 mM ribonucleotide was present in the digest in the form of an alkaline hydrolysate of RNA. Under these conditions it was observed that oligonucleotides C_5T_6 (CTTCCTCTTCT) and C_6T_3 (TCCTTCCCC) were resistant to hydrolysis by spleen exonuclease.

To check whether these 5' terminal sequences were indeed resistant to spleen exonuclease hydrolysis, the sequences which were known to be hydrolysed were examined. Oligonucleotide C_4T_4 was hydrolysed by spleen exonuclease although it

had a 5' terminal hexanucleotide sequence (CTTCCT..) identical to that of C_5T_6. Oligonucleotide C_6T_4, containing the sequence TCCT...., identical to the 5' terminal sequence of C_6T_3, was observed to be hydrolysed by spleen exonuclease.

These observations show that factors besides sequence were important in determining the rate of oligonucleotide hydrolysis by spleen exonuclease. The fact that high concentrations of RNA oligonucleotides were necessary for resistance to hydrolysis to become apparent suggests that the quantity and composition of this RNA may have been a determining factor. If the partial digest incubation contained RNA oligonucleotides with much higher affinity for the spleen exonuclease active site than the radioactive oligodeoxyribonucleotide, the competition may have made the oligodeoxyribonucleotide seem resistant.

The K_m of spleen exonuclease has been shown to be 5 mM for p-nitrophenyl thymidine-3'-phosphate (34); thus the concentration of RNA in these partial digestions (30 mM) was probably enough for effective competition. The differences in the resistances of different oligonucleotides may have been due in part to the quantity and composition of the RNA eluted with the oligonucleotides from the preparative TLC plate.

These observations suggest that under the correct competitive conditions, spleen exonuclease may have specificity which may be useful in future sequence studies.

4.2 A Model for Polynucleotide Kinase Action

We have shown that polynucleotide kinase, in the absence of ATP, acts

as an oligonucleotide 5' phosphomonoesterase (3.5.3). In early experiments, when a great excess of ATP was used, interfering radioactivity was observed which was similar in charge to tetranucleotides. Van de Sande, Kleppe, and Khorana (37) have shown that this interfering material was adenosine tetraphosphate generated by the transfer of a phosphate group from the 5' end of an oligonucleotide to ATP. They also observed and characterized the reverse reaction catalysed by polynucleotide kinase, showing that the 5' phosphate of an oligonucleotide could be transferred to ADP. The phosphomonoesterase activity of the polynucleotide kinase was observed by these workers, but was ignored in the construction of their model.

A model which explains these observations is presented in Figure 4.1.

Other observations made by Van de Sande et al. (37) were that the pH-activity curve for the reverse reaction had a fairly sharp maximum at pH 6.2 and that the rate of phosphorylation in the presence of ATP increased smoothly from pH 4 to a maximum at pH 9.5. The former observation suggests that binding of the oligonucleotide 5' phosphate is facilitated by the ionization of a histidine on the enzyme and of the terminal phosphate on the oligonucleotide.

The smooth pH - activity profile of the normal phosphorylation reaction does not have an inflection near pH 7, indicating that a reaction involving a terminal phosphate group is probably not rate limiting; this suggests that the oligonucleotide binding step might be the rate limiting step in the normal phosphorylation reaction.

4.3 Fractionation Techniques Useful for Sequence Analysis

4.3.1 Sequence Analysis by Mapping Techniques

Most of the sequence work on oligoribonucleotides and much on oligodeoxy-

Figure 4.1

Mechanism of Polynucleotide Kinase Action

Normal Course of Reaction :

$$PNK + O \rightleftarrows O - PNK$$

$$O - PNK + *pppA \rightleftarrows O - PNK - *pppA$$

$$O - PNK - *pppA \rightarrow O - PNK - *p + ppA$$

$$O - PNK - *p \rightleftarrows *p - O + PNK$$

Side Reactions :

$$H_2O + O - PNK - *p \rightarrow O - PNK + *p$$

$$*pppA + O - PNK - *p \rightleftarrows O - PNK + *p*pppA$$

Abbreviations :

*p	-	^{32}p
*pppA	-	[γ - ^{32}P] ATP
O	-	dephosphorylated oligonucleotide
p-O	-	5' phosphorylated oligonucleotide
PNK	-	polynucleotide kinase

ribonucleotides have used mapping techniques for determination of relative base composition and thus sequence. A partial digest of an oligonucleotide contains a series of products each containing one nucleotide less than its parent. The identity of the terminal nucleotide of the parent, which is not present in the daughter, can often be identified by the difference in migration between the two in a two-dimensional system.

Electrophoresis of oligonucleotide mixtures on cellulose acetate at pH 3.5 followed by electrophoresis on DEAE paper at pH 1.9 was developed by Sanger (149) for sequence analysis of oligoribonucleotides. Murray (14) and Wu (129, 76) have investigated this system concerning the possibility of identifying the base composition of sequential digest products from their relative migrations. Murray achieved some success by performing two fractionations on each sample, the other having electrophoresis on aminoethyl cellulose at pH 3.5 as its first dimension. Only in this way could the addition of a pT or a pG be differentiated and the system used in a general sequence scheme.

Polyethylenimine cellulose TLC has been used in many laboratories for oligodeoxyribonucleotide fractionation.

A two-dimensional system comprised of electrophoresis on a cellulose acetate strip at pH 3.5, followed by PEI cellulose TLC using 0.85 M LiCl as chromatographic eluent, has been investigated by Wu (69). This study indicated that this system would be applicable to the sequence of short oligodeoxyribonucleotides but the relative migrations due to pG and pT are similar, as are those due to pA and pC.

A fractionation system comprised of electrophoresis on cellulose acetate at pH 3.5 followed by TLC on DEAE cellulose using a neutralized alkaline hydrolysate

of RNA in 7 M urea as chromatographic eluent (28) has been used for oligodeoxyribonucleotide sequence analysis (7, 26, 111). Higher chain length digest products can be resolved in this system than in those discussed above, but again, the major problem is to differentiate between a pT and a pG difference by relative migration.

The effectiveness of these fractionation methods for sequence analysis by mapping techniques can be summarized as follows : If it is known that the oligonucleotide being studied does not contain both G and T, the two-dimensional electrophoresis technique should be applicable to the analysis of oligonucleotides 5 to 10 nucleotides long ; the cellulose acetate electrophoresis PEI TLC system should be applicable to the analysis of oligonucleotides up to 10 to 13 nucleotides long ; the electrophoresis-homochromatography system should be applicable to oligonucleotides 15 to 20 nucleotides long. If, as is often the case, exact base composition data is not available for the oligonucleotide in question, the two-dimensional electrophoresis system would be applicable to oligonucleotides up to 5 to 10 nucleotides long ; but two analyses would have to be performed on the oligonucleotide partial digest; the cellulose acetate electrophoresis-PEI TLC system would be applicable to oligonucleotides 5 to 10 long ; the electrophoresis-homochromatography system would be applicable to oligonucleotides 5 to 9 nucleotides long. These estimates were made from an examination of published autoradiographs from the various fractionation techniques (7, 14, 69, 120).

The elucidation of relative base composition by measuring relative mobility in a two-dimensional fractionation system is a very convenient and rapid tool for sequence analysis. These methods, however, are rather sensitive to changes in the ion exchange media and in the buffers. As can be seen from the many autoradiographs in this thesis, resolution in the electrophoresis-homochromatography system was

quite variable. It was observed that the resolution was quite dependent on the pH and metal ion content of the electrophoresis buffers, the amount of material loaded to the cellulose acetate strip, and any small imperfections in the TLC plate itself. For analysis of pyrimidine oligonucleotides, where migration differences are large, the variable resolution was not a handicap, as was demonstrated by our and by Ling's work (7, 26). However, in the analysis of oligonucleotides containing all four bases, the migration differences between pT and pG, and, in the case of electrophoresis PEI TLC, between pA and pC, are small and the variable resolution would make definite sequence assignment difficult.

In conclusion, it is unlikely that mapping techniques could be used alone in a general oligonucleotide sequence scheme.

Two-dimensional fractionations of these types can be used as the chain length step in sequence methods involving exonuclease degradation followed by terminal labelling as described in Section 3.6. In this case, the determination of the base composition of each spot by mapping is unnecessary and only chain length resolution is important. The two-dimensional electrophoresis system would still only be applicable to oligonucleotides 5 to 10 nucleotides in length; the cellulose acetate electrophoresis PEI TLC system would be applicable to oligonucleotides up to about thirteen nucleotides long; the electrophoresis-homochromatography system has been demonstrated to be applicable to oligonucleotides as long as fourteen nucleotides, and it is estimated that it would be useful (69) for oligonucleotides up to twenty nucleotides long.

A disadvantage of these two-dimensional systems is that recoveries of material are usually fairly low, and the transfer of oligonucleotides low in T to the

second dimension is usually not very efficient (167). These drawbacks are often superseded by the convenience and rapidity of the electrophoresis TLC systems.

4.3.2 Column Fractionation Techniques

In general columns give better one-dimensional resolution of oligonucleotides than do electrophoresis or TLC systems. It is very cumbersome, however, to produce a two-dimensional system using columns, and, for the present, the sequence determination of long oligonucleotides by mapping procedures is limited to the electrophoresis and TLC systems. Developments in high pressure liquid chromatography (157) may make this approach possible in the near future.

The oligonucleotide sequence schemes involving exonuclease degradation followed by terminal labelling (1.2.1.4c) require only chain length fractionation; thus columns provide the resolution required for longer oligonucleotides than can be analysed using the two-dimensional systems.

4.3.2.1 DEAE Cellulose Column Fractionations

Columns of DEAE cellulose have been used to separate pyrimidine oligonucleotides isostichs up to 14 nucleotides long (33, 164), and oligonucleotides generated by DNAase digestion up to 4 nucleotides long (165, 166). This ion exchange medium seems to have a great proportion of secondary binding sites which reduce resolution, especially when purines are present in the oligonucleotides. This was additionally manifested by the observation that if less than 5 mg of nucleotide material was loaded to a 1 cm x 25 cm DEAE cellulose column, the resolution of oligonucleotide isostichs decreased considerably.

At pH 3.0, the additional binding of the purines has been put to good use (22, 51) in the subfractionation of dinucleotides and trinucleotides according to base composition. Here again, though, several milligrams of carrier dinucleotides were included in the fractionation.

The addition of carrier oligonucleotides is not a disadvantage in the chain length fractionation of a partial digest in a sequence experiment since the next step is radioactive mononucleotide analysis. It becomes a distinct problem when column fractionations are used to isolate oligonucleotides for sequence analysis, because carrier would interfere with terminal labelling performed during the sequence analysis.

4.3.2.2 Phosphate as "Pseudo" Carrier

The addition of 1 mM KH_2PO_4 to the gradient buffers used to elute DEAE cellulose columns allowed us to load less nucleotide material to the column without a serious loss of resolution. In these cases as little as 0.5 mg nucleotide could be loaded to the column. This suggests that the ion exchanger contains secondary binding sites not only for the heterocyclic rings of the oligonucleotides but also for the phosphate groups themselves.

4.3.2.3 DEAE Sephadex Column Fractionations at 25°C

Columns of DEAE Sephadex have been used to separate pyrimidine isostichs up to ten nucleotides long (160), using triethylammonium bicarbonate buffer, pH 8.7, as eluent. Isostichs generated by DNAase I up to the hexanucleotide level were separated on a DEAE Sephadex column eluted at 25°C with a 7 M urea, Tris-HCl buffer, pH 7.6 (51).

In this study DEAE Sephadex columns used for pyrimidine oligonucleotide fractionations were eluted at 25°C with 7 M urea buffered with either sodium acetate, pH 5.5, or Tris-HCl, pH 7.6. It was observed that resolution was very much better at pH 5.5 for both dephosphorylated and phosphorylated oligonucleotides. Thus DEAE Sephadex columns eluted at 25°C were routinely run at pH 5.5.

The fractionation of pyrimidine oligonucleotides according to base composition at pH 3.0 showed a further increase in resolution over the fractionation by chain length at pH 5.5. This progression indicated that the DEAE Sephadex exhibits some secondary binding for oligonucleotides at alkaline pH but much less at acid pH. The manufacturers give a titration curve for DEAE Sephadex (161) which indicates that a considerable portion of the diethylaminoethyl groups are in the form of "double DEAE" groups with a pK of 5.7. It seems likely, therefore, that the unionized form of this DEAE group is the cause of the secondary binding exhibited at pH 7.6. This suggests that the binding probably involves the heterocyclic rings of the oligonucleotides.

Another advantage of DEAE Sephadex over DEAE cellulose is that it is generally not necessary to add carrier to the sample before running the fractionation. As little as 10 picomoles (3.5 ng nucleotide) of oligonucleotide have been successfully fractionated at pH 5.5 on DEAE Sephadex. This is an additional byproduct of the low level of secondary binding of this ion exchange medium.

4.3.2.4 DEAE Sephadex Chain Length Fractionations at 65°C

Columns of DEAE Sephadex eluted with Tris-HCl buffer, pH 7.6, 7 M urea, at 65°C have been very successful in the fractionation of oligonucleotides generated

by DNAase digestion of DNA (22, 51) and fractionation of these mixtures up to the hexadecanucleotide level has been accomplished. These digests were very heterogeneous, each chain length fraction consisting of numerous oligonucleotides of different base composition. This certainly contributed to the broadening of the chain length peaks in these fractionations.

In the fractionation of a partial exonuclease digest during the sequence analysis of an oligonucleotide only one species occurs in each chain length fraction. These homogeneous isostichs would be resolved more clearly than the heterogeneous mixtures resulting from DNAase digestion.

Investigation of chromatography on DEAE Sephadex at pH 5.5 at 65°C is presently under study to see if the increased resolution is apparent at high temperatures as well as at 25°C.

It is safe to predict that ion exchange columns will be useful for sequence analysis of oligonucleotides at least 25 nucleotides long.

4.4 Pyrimidine Catalogues of $S13^+$ and S13suN15 DNA

The catalogues of the pyrimidine clusters of $S13^+$ DNA (152) and S13suN15 replicative form DNA (33) were obtained using DEAE cellulose chromatography in the presence of large amounts of unlabelled carrier pyrimidine oligonucleotides. Resolution and recoveries were excellent in these quantitative analyses.

The data of Cerny et al. (33) indicated that the recovery of octa-, nona-, deca-, and undecanucleotides was approximately 90%, relative to the total recovery from the column. These specific losses in the long oligonucleotides were probably

due to trailing of the individual isostich fractions; this was not evident in the total recoveries because these fractions represented only 2.5% of the total radioactivity.

Isolation of pyrimidine clusters from $S13^+$ RF DNA and fractionation on DEAE Sephadex has led to the conclusion that oligonucleotide C_5T_4 occurs twice in the RF DNA instead of once as reported previously for S13suN15 DNA (33). The previous assignment was a borderline case in the previous study, since its calculated occurrence was 1.46 tracts per genome, which was rounded to unity.

It was reported that oligonucleotide C_6T did not occur in $S13^+$ DNA (33, 152) and that it occurred once in the S13 replicative form DNA (33). Fractionations using electrophoresis-homochromatography have indicated that the reverse situation is true, that C_6T occurs once in $S13^+$ DNA. This observation was borne out by sequence analysis of the C_6T isolated from $S13^+$ DNA.

These minor errors were due to specific losses of long C-rich oligonucleotides on the DEAE cellulose columns, even though large amounts of carrier were present in the fractionations.

4.5 Pyrimidine Oligonucleotide Sequences from $S13^+$ DNA

4.5.1 Comparison with Sequences from Other Small Single Stranded DNA Bacteriophages

The small icosahedral bacteriophages ØX174 and S13 are closely related by genetic and immunological standards (121, 122); the same can be said of the small filamentous bacteriophages fd and f1 (123, 124). Ling has presented selected

pyrimidine sequences from the DNA of three of these : ØX174, fd, and f1. Sequences of pyrimidine oligonucleotides from S13$^+$ DNA have been presented in this thesis, and many more sequences of pyrimidine oligonucleotides from ØX174 DNA are known from work in this laboratory (117). This work has shown that Ling incorrectly assigned the base composition of two of the ØX174 oligonucleotides and thus presented incorrect sequences for them.

All of Ling's sequences, as well as all the ØX174 sequences determined in this laboratory, were determined using uniformly labelled oligonucleotides, partial exonuclease digestions, and electrophoresis-homochromatography (2.2.10.1). The sequences of the oligonucleotides from S13$^+$ DNA were determined by several methods, as has been described in this thesis. A summary of these pyrimidine oligonucleotide sequences is presented in Table 4.1.

Before a comparison of sequence homologies in these pyrimidine oligonucleotides can be made, it is necessary to decide what type of similarities are significant. Assuming a random DNA sequence, the probability of occurrence of any unique oligonucleotide in a genome will be :

$$(a)^{nA}\ (c)^{nC}\ (g)^{nG}\ (t)^{nT}$$

where a, c, g, and t represent the proportion of A, C, G, and T in the genome being observed, and nA, nC, nG, and nT represent the number of times each deoxynucleotide occurs in the unique oligonucleotide. For S13$^+$ DNA, this formula becomes (140) :

$$P_{AnCnGnTn} = (0.235)^{nA}\ (0.201)^{nC}\ (0.231)^{nG}\ (0.333)^{nT}$$

Table 4.1

Pyrimidine Oligonucleotide Sequences from the DNA of Bacteriophages S13, ØX174, fd and f1

Component	Source S13	ØX174	fd	f1
C_6T	CTCCCCC	CCTCCCC		
C_5T_2	CTCTCCC	NP		
CT_6	CTTTTTT	(CT)TTTTT		
C_5T_3	CCTTTCCC	CCTTTCCC		
C_4T_4	CTTCCTTC	CTTCCTTC		
		CTTTCTCC		
C_3T_5	NP	TCTTCTTC		
		TTTTCCTC		
C_2T_6	TTTTCCTT	TTTTCTCT		
	CTTTTTTC	CTTTTTTC		
C_6T_3	TCCTTCCCC	TCCTTCCCC		
C_5T_4	CTTCCTCCT	CTTCCTCCT		
C_4T_5	TCTTTCTCC	TCTTTCTCC	CTTCCTCTT	CTTCCTCTT
	CTCTTTCTC	CTCTTTCTC		
C_3T_6	NP	NP	TTCCTTTCT	TTCCTTTCT
			TTTCCTTCT	TTTCCTTCT
C_2T_7	NP	NP	TCTTCTTTT	TCTTCTTTT
			TTTCTTTCT	
C_7T_3	NP	CTCCTCTCCC	NP	NP
C_6T_4	CTCCTCTTCC	NP	NP	NP
C_4T_6	NP	NP	TCCTTCTCTT	TCCTTCTCTT
C_2T_8	CTTTTTTTCT	CTTTTTTTCT	TTTTTCCTTT	NP
C_5T_6	CTTCCTCTTCT	NP	TTTTTCCTCCC	
C_4T_7	NP	NP	NP	TTTTTCTTCCC
C_3T_8	NP	NP	CCTTTTTTTTC	CCTTTTTTTTC
C_9T_4	NP	NP	NP	CCTTCCCTCCCTC
C_9T_{11}	NP	NP	CTTTCTTCCCTTCC TTTCTC	CTTTCTTCCCTTCC TTTCTC

NP – not present

For example, the hexanucleotide ACAGTT has the probability of occurrence :

$$(0.235)^2\ (0.201)\ (0.231)\ (0.333)^2 = 2.84 \times 10^{-4}$$

The most probable number of times a given sequence occurs in a molecule 5500 nucleotides long would be the product of the probability of occurrence and 5500.

The most probable occurrences of some pyrimidine oligonucleotide sequences in a genome of 5500 nucleotides long are listed in Table 4.2.

Table 4.2

Probable Occurrence of Short Pyrimidine Oligonucleotides in a Random DNA Molecule 5500 Nucleotides Long

Length	Probable Occurrence in 5500 Nucleotides		
1	C : 1105		T : 1831
2	C_2 : 222	CT : 368	T_2 : 610
3	C_3 : 45		T_3 : 203
4	C_4 : 9	C_2T_2 : 25	T_4 : 68
5	C_5 : 2		T_5 : 23
6	C_6 : 0.4	C_3T_3 : 1.6	T_6 : 7
7	C_7 : 0.07		T_7 : 2.5

It can be seen from the values in the table that analysis of recurring sequences less than six nucleotides long is not statistically valid, since they probably occur more than once by random selection. If homologous sequences above six

nucleotides in length are observed, we can conclude that there may be a cause other than random selection. There are three types of sequence homology which I shall discuss. The first consists of sequence similarities within one genome.

Ling (26) has reported that the octanucleotide TTTTTCCT in fd DNA occurs in both C_2T_8 and in C_5T_6; the nonanucleotides TTTCCTTCT and TTTCTTTCT, also in fd DNA, differ in only one position. These duplications, however, do not seem to be repeated in f 1 DNA, which observation reduces the probability that they are irreplaceable sequences.

In the $S13^+$genome, we find many internal sequence homologies : the heptanucleotide CTTTTTT occurs in CT_6, C_2T_6, and C_2T_8; the heptanucleotide CTTCCTC occurs in both C_5T_4 and C_5T_6; the octanucleotide TCCTCTTC occurs in C_6T_4 and C_5T_6; the hexanucleotide CTTCCT occurs in C_4T_4, C_5T_4 and C_5T_6.

In ØX174 DNA the CT_6 duplication is repeated as is the CTTCCT redundancy in C_4T_4 and C_5T_4, but C_5T_6 is not present, thus the other duplications observed in $S13^+$DNA are not present. Another redundancy, however, is present : octanucleotide CTTTCTCC is present in C_4T_4 and C_4T_5.

Thus there are redundancies in the pyrimidine clusters of each of the small coliphages analysed. The observation that these redundancies are not always carried over from genome to genome suggests that they are not sequences important to a common function such as replication. A safer hypothesis would be that these differences arose since the phages separated evolutionarily, and that these genomes evolved in part, by deletion and duplication of existing DNA. For further interpretation it will be necessary to see if these similar sequences are located near each

other in the genome. Studies using restriction enzymes in this laboratory may shed more light on this topic.

The second set of homologous sequences I shall discuss are the similarities between ØX174 and S13 and between fd and f1.

Seven of the eleven known sequences from fd DNA have identical counterparts in f1 DNA; two others, C_2T_8 and C_5T_6, are closely related to C_4T_7 from f1 DNA; the oligonucleotide TTTCTTTCT from fd DNA is very closely related to TTTCCTTCT which occurs in both fd and f1 DNA.

Only one mutational event, such as the addition of a purine, is necessary to remove a long pyrimidine tract from the catalogue. If we make the assumption that this is the case for the three oligonucleotides in fd and f1 which do not have counterparts in the other genome, the minimum mutational divergence between the two genomes can be calculated. Thus there may be only four differences in the 120 nucleotides known for the two genomes, a 3.3% mutational difference.

Nine of the fifteen known sequences from ØX174 DNA have identical counterparts in S13 DNA and three others differ from their counterparts by only one base. In two cases the difference between oligonucleotides of the same base composition is the reversal of a CT or TC sequence; C_6T in S13 DNA has the sequence CTCCCCC, while in ØX174 DNA the sequence is CCTCCCC; also, in S13 DNA one of the C_2T_6 oligonucleotides has the sequence TTTTCCTT, while in ØX174 DNA it is TTTTCTCT. Whether this indicates that two base substitutions have taken place or that the reversal is a single mutational event is unknown.

In the same way as was done for fd DNA, and assuming that a CT to TC transition is a single mutational event, the mutational divergence between S13 and ØX174

can be calculated to be 8 mutational events in 144 nucleotides ; 5.5% divergence.

Heteroduplex studies (90) of ØX174/S13 DNA in solutions containing various concentrations of formamide have led to the prediction that S13 and ØX174 have 36% mismatched base pairs. The comparison with the estimate of 5.5% mutational divergence indicates that either long pyrimidine sequences are conserved relative to the majority of the DNA, or formamide denaturation of DNA is much more sensitive to mismatching of base pairs than has been predicted in the past (90, 161, 162). The latter seems more likely. It is possible that a small amount of base mismatching might cause a proportionally larger amount of denaturation in formamide, due to cooperative phenomena which are probably sequence dependent.

Perhaps the most meaningful homologies are those which cross the boundary between the spherical and the filamentous coliphage groups. The octanucleotide CTTTTTTT occurs in all four phage genomes ; in C_2T_8 of S13 and ØX174, and in C_3T_8 of fd and f1. This repeated occurrence suggests that this sequence may play an essential role common to the life cycles of all four bacteriophages.

This role could be in the initiation of replication, in binding to the E. coli membrane, or as a sequence specifically recognized by a bacterial endonuclease or ligase involved in the mechanism of circular DNA replication.

Several models for the replication of ØX174 DNA in E. coli have been proposed ; (168, 169) all have in common the requirement of an endonuclease to cleave RF I molecules or to release single stranded progeny molecules. The recent demonstration that the gene A protein from ØX174 is an endonuclease that cleaves ØX174 RF I or ØX174 single stranded DNA only once (170) suggests that there

might be a unique recognition sequence in the viral DNA for this enzyme. Examination of the cleavage specificity of this enzyme will determine whether the recognition sequence is unique, and perhaps whether it is evolutionarily conserved.

The sequence CTTTTTTT may not be the site discussed above, but the fact that it occurs in these four phages makes it likely that it has some specific function besides coding for protein.

4.5.2. Symmetrical features of the pyrimidine oligonucleotides from $S13^+$ and ØX174 DNA

Since the discovery of restriction enzymes (54) and the finding that their sites of cleavage are double stranded palindromic sequences, interest in such symmetrical sequence properties has risen. It has recently been shown that longer double stranded palindromes are present in the lac repressor binding site of λ DNA (96). Whether the palindromic parts of this binding site are essential for specificity of binding is under investigation by sequence analysis of binding sites from operator mutants of λ (171). Evidence for very large palindromic sequences occurring in eukaryotic chromosomal DNA has recently been reported (172).

The polarity of the phosphodiester linkage in nucleic acids precludes the possibility of palindromic symmetry occurring in single stranded DNA sequences such as pyrimidine oligonucleotides. Examination of the pyrimidine oligonucleotide sequences from $S13^+$ DNA (Table 4.1) showed that all but two heptanucleotides contained symmetrical ordering of the bases in their sequences, involving at least six in each sequence. For ØX174 DNA, all the pyrimidine oligonucleotides, except for one heptanucleotide and two octanucleotides, exhibited the same type of symmetrical ordering. These symmetrical sequences have the same statistical probability of occurrence as double stranded palindromic sequences of the same chain length in a DNA of the same size. For example, the single stranded sequence CTTTTTTC would occur as often as the double stranded palindromic sequence $\frac{\text{CTTTAAAG}}{\text{GAAATTTC}}$

in a random DNA molecule. The widespread occurrence of these sequence symmetries is pyrimidine oligonucleotides from S13 and ØX174 DNA makes it difficult to predict whether they may have biological significance or if the phenomenon is a coincidence.

Admittedly, symmetrical sequences are more likely to occur when only two nucleotides are present, as in these pyrimidine clusters, but even in a random DNA we should expect that a palindromic hexanucleotide sequence would have an approximate probability of 1/64.

Our conclusion, therefore, is that only quite long palindromic sequences are likely to be mechanistically important.

4.6 Progress in Sequence Analysis

As was discussed in the introduction, two approaches to DNA sequence analysis are possible. The direct approach became feasible only in the past few years, since the advent of the highly specific restriction endonucleases (155) and the relatively low specificity endonuclease T4 endonuclease IV (38). The indirect approach, in particular the use of DNA polymerase to elongate a primer oligonucleotide on a relatively high molecular weight template, has been possible for quite a while, but was hindered by the lack of a suitable primer. The preparation of oligonucleotide primer requires either direct fragmentation techniques unavailable until recently, or some sequence information according to which the primer can be synthesized.

These approaches have solved the problem of specificity in the analysis of DNA sequence. The problem of incorporation of sufficient radioactivity into DNA for sequence analysis has been mitigated to a great extent by the use of in vitro labelling reactions. The advent of labelling of oligonucleotides using polynucleotide kinase and [γ-^{32}P]ATP has allowed analysis on less than a pmole of oligonucleotide. The [α-^{32}P] ribo-and deoxyribonucleoside triphosphates are not usually of as high specific activity as [γ-^{32}P], but they easily allow analysis of

as little as 10 pmoles of oligonucleotide. These triphosphates are used with deoxynucleotidyl transferase for 3' terminal labelling, and with DNA polymerase during primer elongation sequence studies.

Thus the two technical problems which held DNA sequencing back relative to RNA sequencing have been overcome over the past few years, and several laboratories are now actively studying DNA sequence (14, 26, 27, 56, 57, 59, 60, 64, 65, 67, 69, 70, 72-80, 91, 96, 106-108, 110-120, 129, 155).

Another indirect method we discussed involved the use of RNA polymerase to transcribe DNA. When RNA polymerase is used as a non-specific transcription agent (96), it must be used on a relatively short molecule so that the resulting RNA mixture will not be too complex. This technique has a disadvantage in that the termini of the fragment being examined are not well transcribed.

When RNA polymerase is used in its role as a specific initiator of transcription, adding nucleotides to a primer, the template must be fragmented and purified to contain only one transcription initiation site. Maizels (106) did this elegantly using sonication and purification of repressor protein bound DNA; direct fragmentation with restriction enzymes to isolate correctly sized fragments will probably be used in the future.

For these RNA polymerase techniques, the use of [α-^{32}P] triphosphates solves the problem of radioactive incorporation.

The isolation of specific fragments of DNA protected from non-specific nucleases by bound proteins can be classified as a part of the direct fragmentation methods. These fragments, however, are usually not of uniform length, making terminal sequence analysis very difficult.

An important point is that all of the methods here discussed, with the exception of the RNA transcription methods, require, as a final step, the determination of the sequence of oligodeoxyribonucleotides.

Recent studies (111, 113) have demonstrated that DNA polymerase elongation proceeds in spurts, producing reproducible bands separable by gel electrophoresis. This property, combined with the use of ribocytidylic acid incorporation into the newly synthesized DNA chain, allows production of oligonucleotides three to twenty nucleotides long.

The direct fragmentation methods, as discussed in Section 1.2.1, yield oligonucleotides from dinucleotides to 40 nucleotides in length which must be sequenced.

Several procedures for the sequence of oligodeoxyribonucleotides have been proposed. Roychoudhury and Kossell (119) have proposed venom exonuclease partial digestion followed by labelling with [α-^{32}P] rATP and deoxynucleotidyl transferase ; identification of the 3' termini of the digest products yields the sequence. They have applied this method only to a synthetic pentanucleotide of known sequence.

We have proposed (64) spleen exonuclease digestion, labelling with polynucleotide kinase and [γ-^{32}P] ATP, and 5' terminal identification. Wu (70) used this method as a part of his analysis of some oligonucleotide sequence methods, using as substrate a synthetic oligonucleotide of the same sequence as a part of one of the λ cohesive ends.

It has been proposed (69) that the electrophoresis-homochromatography system developed by Sanger and his coworkers (28) could be used to determine the sequence

of any oligonucleotide by mapping techniques only. The proposal was based on a fractionation of a partial digest of a tetradecanucleotide of known sequence containing only one G in the thirteenth position, where it can easily be identified.

Our experience (3.4), supported by work from Murray's group (120), has been that the resolution of individual digest products in the electrophoresis-homochromatography system is not sufficient for a general sequence method. Many oligonucleotides could be sequenced in this way, particularly if the base composition is known, but many others would at best give a tentative sequence which would need verification by different sequence methods.

The method for oligonucleotide sequence analysis developed in this thesis (3.7) does not involve mapping techniques or the determination of relative base composition. Each radioactive mononucleotide analysis in this method represents a discrete position in the sequence. The limitation as to the length of oligonucleotide which can be analysed does not depend on quantitation of base composition analyses of digest products, but solely on the range of the chain length fractionation system.

Oligonucleotides less than fifteen nucleotides long can most conveniently be analysed using electrophoresis-homochromatography (4.3.1) as the chain length fractionation step after spleen exonuclease digestion and 5' terminal labelling. Oligonucleotides fifteen to twenty-five nucleotides long should be analysed using DEAE Sephadex column chromatography at 65°C. It may be possible to extend this upper limit by adjusting chromatographic conditions (4.3.2). The use of [γ-^{32}P] ATP in the labelling reaction means that less oligonucleotide substrate is required than in the methods using [α-^{32}P] triphosphates. Most important, the

use of exonuclease digestion followed by 5' terminal labelling removes the ambiguity inherent in sequence methods based on base composition analysis or on mapping.

CHAPTER 5

CLAIMS TO ORIGINAL RESEARCH

1. A completely general method for the determination of the sequence of oligodeoxyribonucleotides containing all four bases was developed.
2. The sequences of fourteen of the longest pyrimidine oligonucleotides from bacteriophage $S13^+$ DNA were determined. Comparison of these sequences with those from the DNA of closely related bacteriophage ØX174 indicated that the evolutionary divergence of the two genomes was less than predicted from heteroduplex studies. Comparison with sequences from the DNA of the small filamentous coliphages fd and f1 demonstrated that all four genomes contained the heptanucleotide CT_6.
3. Polynucleotide kinase was found to exhibit an oligonucleotide-5'-phosphomonoesterase activity and a general model for the mechanism of polynucleotide kinase action was presented.
4. Polynucleotide kinase purified from T4-infected cells was found to contain oligodeoxyribonucleotide-like contaminants. These were characterized and a method for their removal from the enzyme preparation developed.
5. Two corrections were made to the pyrimidine catalogues of $S13^+$ and S13suN15 DNA.
6. Resolution of oligonucleotides on DEAE Sephadex at 25°C at pH 5.5 was demonstrated to be superior to that at pH 7.6 or pH 8.7. In fractionations at pH 3.2, peak widths were narrower than at pH 5.5, indicating that the secondary ion exchange group, with a pK of 5.7, on DEAE Sephadex may be

involved in secondary binding of oligonucleotides at alkaline pH.

7. Two methods were developed for ion exchange chromatography of small quantities of oligonucleotide without the addition of carrier. Addition of 1 mM inorganic phosphate to the eluting buffers of DEAE cellulose columns allowed fractionation of as little as 0.5 mg of nucleotide material. The use of DEAE Sephadex in column fractionations allowed the fractionation of as little as 3.5ng of nucleotide material in the absence of carrier.

BIBLIOGRAPHY

1. Blattner, F.R., and Erickson, H.P. (1967) Anal. Biochem, 18, 220-227.

2. Salnikow, J., Stein, W.H., and Moore, S. (1969) Fed. Proc. 28, 344.

3. Bachelet, M., Guibe, M. (1951) Bull. Soc. Chim. France 18, 554.

4. Bernardi, G. (1969) Biochim. Biophys. Acta 174, 423-434.

5. Sulkowski, E., and Laskowski, M., Sr. (1971) Biochim. Biophys. Acta 240, 443-447.

6. Kay, E.R.M., Simmons, N.S., and Dounce, A.L. (1952) J. Amer. Chem. Soc. 74, 1724-1726.

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

8. Bernardi, A., and Bernardi, G. (1968) Biochim. Biophys. Acta 155, 360-370.

9. Garen, A., and Levinthal, C. (1960) Biochim. Biophys. Acta 38, 470-483.

10. Glynn, I.M., and Chappell, J.B. (1964) Biochem. J. 90, 147-149.

11. Schendel, P.F., and Wells, R.D. (1973) J. Biol. Chem. 248, 8319-8321.

12. Richardson, C.C. (1965) Proc. Nat. Acad. Sci. U.S.A. 54, 158-165.

13. Wu, R., and Kaiser, A.D. (1967) Proc. Nat. Acad. Sci. U.S.A. 57, 170-177.

14. Murray, K. (1973) Biochem. J. 131, 569-583.

15. Warner, H.R., and Lewis, N. (1966) Virology 29, 172-175.

16. Shleser, R., Ishiwa, H., Mannes, B., and Tessman, E.S. (1968) J. Mol. Biol. 34, 121-129.

17. Hirt, B. (1967) J. Mol. Biol. 26, 365-369.

18. Schekman, R.W., Iwaya, M., Bromstrup, K., and Denhardt, D.T. (1971) J. Mol. Biol. 57, 177-199.

19. Sinsheimer, R.L. (1966) In Procedures in Nucleic Acid Research, p. 572. eds. G.L. Cantoni and D.R. Davies. Harper and Row, New York.

20. Burton, K., and Petersen, G.B. (1960) Biochem. J. 75, 17-27.

21. Spencer, J.H., Cape, R.E., Marks, A., and Mushynski, W.E. (1969) Canad. J. Biochem. 17, 329-337.

22. Junowicz, E., and Spencer, J.H. (1970) Biochemistry 9, 3640-3648.

23. Clausen, T. (1968) Anal. Biochem. 22, 70-73.

24. Hanggi, U.J., Streeck, R.E., Voigt, H.P., and Zachau, H.G. (1970) Biochim. Biophys. Acta 217, 278-293.

25. Yamamoto, K.R., and Alberts, B.M. (1970) Virology 40, 734-744.

26. Ling, V. (1972) Proc. Nat. Acad. Sci. U.S.A. 69, 742-746.

27. Harbers, B., Delaney, A.D., Harbers, K., and Spencer, J.H. (1974) in preparation.

28. Brownlee, G.G., and Sanger, F. (1969) Europ. J. Biochem. 11, 395-399.

29. Jeng, Y.D., Gelfand, M., Hayashi, R., and Tessman, E.S. (1970) J. Mol. Biol. 49, 521-526.

30. Poljak, R.J., and Seruda, A.J. (1969) Virology 39, 145-146.

31. Tessman, E.S., and Shleser, R. (1963) Virology 19, 239-240.

32. Zahler, S.A. (1958) J. Bacteriol. 75, 310-315.

33. Cerny, R., Cerna, E., and Spencer, J.H. (1969) J. Mol. Biol. 46, 145-156.

34. Razzell, W.E., and Khorana, H.G. (1961) J. Biol. Chem. 236, 1144-1149.

35. Novogrodsky, A., Tal, M., Traub, A., and Hurwitz, J. (1966) J. Biol. Chem. 241, 2933-2943.

36. Panet, A., Van de Sande, J.H., Loewen, P.C., Khorana, H.G., Raae, A.J., Lillehaug, J.R., and Kleppe, K. (1973) Biochemistry 12, 5045-5049.

37. van de Sande, J.H., Kleppe, K., and Khorana, H.G. (1973) Biochemistry 12, 5050-5055.

38. Sadowski, P.D., and Hurwitz, J. (1969) J. Biol. Chem. 244, 6182-6191.

39. Belozersky, A.N., and Spirin, A.S. (1960) The Nucleic Acids, Chemistry and Biology; Vol. 3, p 147. Academic Press, New York.

40. Chargaff, E. (1955) The Nucleic Acids, Chemistry and Biology; Vol. 1, p 307. Academic Press, New York.

41. Takemura, S. (1959) Bull. Chem. Soc. Japan 32, 920-926.

42. Haberman, V. (1962) Biochim. Biophys. Acta 55, 999-1001.

43. Cape, R.E., and Spencer, J.H. (1968) Canad. J. Biochem. 46, 1063-1073.

44. Moss, G.P., Reese, C.B., Schofield, K., Shapiro, R., and Todd, L. (1963) J. Chem. Soc. 1, 1149-1154.

45. Vizsolyi, J.P., and Tener, G.M. (1962) Chem. Industr. 263-264.

46. Pfitzner, K.E., and Moffat, J.G. (1963) J. Amer. Chem. Soc. 85, 3027-3028.

47. Gabriel, T., Chen, W.T., and Nussbaum, A.L. (1968) J. Amer. Chem. Soc. 90, 6833-6837.

48. Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H., Penswick, J.R., and Zamir, A. (1965) Science 147, 1462-1465.

49. Adams, J.M., Jeppeson, P.C.N., Sanger, F., and Barrell, B.G. (1969) Nature 223, 1009-1014.

50. Min Jou, W., Haegeman, G., Yselbaert, M., Fiers, W. (1972) Nature 237, 82-88.

51. Junowicz, E. (1970) Ph.D. Thesis, McGill University.

52. Junowicz, E., and Spencer, J.H. (1973) Biochim. Biophys. Acta 312, 85-102.

53. Bernardi, G., Ehrlich, S.D., and Thiery, J.-P. (1973) Nature New Biol. 246, 36-40.

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

55. Edgell, M.H., Hutchison, C.A., III., and Sclair, M. (1972) J. Virol. 9, 574-582.

56. Padmanabhan, R., and Wu, R. (1972) J. Mol. Biol. 65, 447-467.

57. Wu, R. (1970) J. Mol. Biol. 51, 501-521.

58. Goulian, M., Lucas, Z.J., and Kornberg, A. (1968) J. Biol. Chem. 243, 627-638.

59. Englund, P.T. (1971) J. Biol. Chem. 246, 5684-5687.

60. Kossel, H., and Roychoudhury, R. (1971) Europ. J. Biochem. 22, 271-276.

61. Szekely, M., and Sanger, F. (1969) J. Mol. Biol. 43, 607-617.

62. Szybalski, W., Bovre, K., Fiandt, M., Guha, A., Hradecna, Z., Kumar, S., Lozeron, H.A., Maher, V.M., Nijkamp, H.J.J., Summers, W.C., and Taylor, K. (1969) J. Cell. Physiol. 74, Sup 1, 33-70.

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

64. Delaney, A.D., and Spencer, J.H. (1973) Fed. Proc. 32, 664 Abs.

65. Delaney, A.D., and Spencer, J.H. (1974) in preparation.

66. Benbow, R.M., Hutchison, C.A., III, Fabricant, J.D., Sinsheimer, R.L. (1971) J. Virol. 7, 549-558.

67. Ziff, E.B., Sedat, J.W., and Galibert, F. (1973) Nature New Biology 241, 34-37.

68. Grosveld, F. (1974) personal communication.

69. Jay, E., Bombara, R., Padmanabhan, R., and Wu, R. (1974) Nucleic Acids Research 1, 331-353.

70. Wu, R., Chen-pei D. Tu, and Padmanabhan, R. (1973) Biochim. Biophys. Res. Comm. 55, 1092-1099.

71. Burton, K., and Petersen, G.B. (1960) Biochem. J. 75, 17-27.

72. Padmanabhan, R., and Wu, R. (1972) J. Mol. Biol. 65, 447-467.

73. Wu, R., and Taylor, E. (1971) J. Mol. Biol. 57, 491-511.

74. Wu, R. (1970) J. Mol. Biol. 51, 501-521.

75. Weigel, P.H., Englund, P.T., Murray, K., and Old, R.W. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 1151-1155.

76. Wu, R., Tu, C.P., and Padmanabhan, R. (1973) Biochem. Biophys. Res. Comm. 55, 1092-1099.

77. Englund, P.T. (1972) J. Mol. Biol. 66, 209-224.

78. Price, S.S., Schwing, J.M., and Englund, P.T. (1973) J. Biol. Chem. 248, 7001-7006.

79. Donelson, J.E., and Wu, R. (1972) J. Biol. Chem. 247, 4661-4668.

80. Donelson, J.E., Wu, R. (1972) J. Biol. Chem. 247, 4654-4660.

81. Linn, S., and Lehman, I.R. (1965) J. Biol. Chem. 240, 1287-1293.

82. Marks, A., and Spencer, J.H. (1970) J. Mol. Biol. 51, 115-130.

83. Kohne, D.E., (1968) Biophys. J. 8, 1104-1118.

84. Rabin, E.Z., Mustard, M. and Fraser, M.J. (1968) Canad. J. Biochem. 46, 1285-1291.

85. Bartok, K. (1974) Proc. Can. Fed. Biol. Soc. 17, 116.

86. Harbers, B. (1974) personal communication.

87. Kolata, G.B. (1973) Science 184, 52-53.

88. Sobell, H.M. (1972) Proc. Nat. Acad. Sci. U.S.A. 69, 2483-2487.

89. Gierer, A. (1966) Nature 212, 1480-1481.

90. Godson, G.N. (1973) J. Mol. Biol. 77, 467-477.

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

92. Dausse, J.P., Sentenac, A., and Fromageot, P. (1972) Eur. J. Biochem. 26, 43-49.

93. Heyden, B., Musslein, C., and Schaller, H. (1972) Nature New Biol. 240, 9-12.

94. Pirotta, V. (1973) Nature New Biology 244, 13-16.

95. Maniatis, T., and Ptashne, M. (1973) Nature 246, 133-136.

96. Gilbert, W., and Maxam, A. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 3581-3584.

97. Reznikoff, W.S., and Abelson, D. (1972) Ann. Rev.Gen. 6, 133.

98. Roberts, J.W. (1969) Nature 224, 1168-1174.

99. Richardson, J.P. (1970) Cold Spring Harb. Symp. Quant. Biol. 35, 127-133.

100. Muller-Hill, B., Crapo, L., and Gilbert, W. (1968) Proc. Nat. Acad. Sci. U.S.A. 59, 1259-1264.

101. Muller-Hill, B., Beyreuther, K., and Gilbert, W. (1971) Methods in Enzymology, Vol. 21, part D, pp 483-487.

102. Ruger, W. (1971) Biochim. Biophys. Acta 238, 202-211.

103. Le Talaer, J., and Jeanteur, Ph. (1971) Proc. Nat. Acad. Sci. U.S.A. 68, 3211-3215.

104. Okamoto, T., Sugiura, M., and Takanami, M. (1972) Nature New Biology 237, 108-109.

105. Chen, Cheng-Yien, Hutchison, C.A., and Edgell, M.H. (1973) Nature New Biology 243, 232-236.

106. Maizels, N.M. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 3585-3589.

107. Wu, R. (1972) Nature New Biology 236, 198-200.

108. Oertel, W., and Schaller, H. (1972) FEBS Letters 27, 316-320.

109. Berg, P., Fancher, H., and Chamberlin, M. (1963) in Symposium on Informational Macromolecules, pp 467-483. eds. Vogel, H., Bryson, V., Lampen, J.O. Academic Press, New York and London.

110. van de Sande, J.H., Loewen, P.C., and Khorana, H.G. (1972) J. Biol. Chem. 247, 6140-6148.

111. Sanger, F., Donelson, J.E., Coulsen, A.R., Kossel, H., and Fischer, D. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 1209-1213.

112. Khorana, H.G., Agarwal, K.L., Buchi, H., Caruthers, M.H., Gupta, N.K., Kleppe, K., Kumar, A., Ohtsuka, E., Rajbhandary, U.L., van de Sande, J.H., Sgaramella, V., Teras, T., Weber, H., and Yamada, T. (1972) J. Mol. Biol. 72, 209-218.

113. Loewen, P.C., Sekiya, T., and Khorana, H.G. (1974) J. Biol. Chem. 249, 217-216.

114. Southern, E.M. (1970) Nature 227, 794-798.

115. Salser, W., Fry, K., Brunk, C., and Poon, R. (1972) Proc. Nat. Acad. Sci. U.S.A. 69, 238-242.

116. Fry, K., Poon, R., Whitcome, P., Idriss, J., Salser, W., Mazrimas, J., and Hatch, F. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 2642-2646.

117. Harbers, K., Harbers, B., and Spencer, J.H. (1974) in preparation.

118. Ling, V. (1972) Proc. Nat. Acad. Sci. 69, 742-746.

119. Roychoudhury, R., and Kossel, H. (1971) Eur. J. Biochem. 22, 310-320.

120. Murray, K. (1974) Progress in Nucleic Acid Research, Vol. 14. eds. Davidson, J.N., and Cohn, W.E. . in press.

121. Tessman, E.S., and Shleser, R. (1963) Virology 19, 239-240.

122. Zahler, S.A. (1958) J. Bact. 75, 310-315.

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

124. Denhardt, D.T., and Marvin, D.A. (1969) Nature 221, 769-770.

125. Fuke, M. (1974) Proc. Nat. Acad. Sci. U.S.A. 71, 742-745.

126. Middleton, J.H., Edgell, M.H., and Hutchison, C.A., III. (1972) J. Virol. 10, 42-50.

127. van de Sande, J.H., and Bilsker, M. (1973) Biochemistry 12, 5056-5062.

128. Szeto, K.S., and Soll, D. (1974) Nucleic Acids Res. 1, 171-181.

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

130. Erlich, D.S., Bertazzoni, U., and Bernardi, G. (1973) Eur. J. Biochem. 40, 143-147.

131. Laval, J., Thiery, J.-P., Erlich, D.S., Paoletti, C., and Bernardi, G. (1973) Eur. J. Biochem. 40, 133-137.

132. Ho, N.W.Y., and Gilham, P.T. (1973) Biochim. Biophys. Acta 308, 53-58.

133. Shapiro, H.S., and Chargaff, E. (1957) Biochim. Biophys. Acta 26, 608-623.

134. Leuchtenberger, C., Leuchtenberger, R., and Davis, A.M. (1954) Amer. J. Pathol. 30, 65-85.

135. Cairns, J. (1962) J. Mol. Biol. 4, 407-409.

136. Grosveld, F. (1972) Masters Thesis, University of Amsterdam.

137. Ando, T. (1966) Biochim. Biophys. Acta 114, 158-168.

138. Whitcome, P., Fry, K., and Salser, W. (1973) Methods in Enzymology, Vol. 29, part E, pp 295-321. eds. Grossman, L., and Moldave, K. Academic Press, New York and London.

139. Cashmore, A.R., and Petersen, G.B. (1969) Biochim. Biophys. Acta 174, 591-603.

140. Spencer, J.H., Cerny, R., Cerna, E., and Delaney, A.D. (1972) J. Virol. 10, 134-141.

141. Grosveld, F.G., and Spencer, J.H. (1974) Nucleic Acids Res. in press.

142. Maniatis, T., and van de Sande, J.H. in preparation.

143. Elson, E., and Jovin, J.M. (1969) Anal. Biochem. 27, 193-204.

144. Birnboim, H.C., Glickman, J. (1969) J. Chromatog. 44, 581-593.

145. Godson, N. (1974) Virology 58, 272-289.

146. Ptashne, M. (1967) Proc. Nat. Acad. Sci. U.S.A. 57, 306-313.

147. Ptashne, M. (1967) Nature 214, 232-234.

148. Kaptein, J. (1974) Ph.D. Research Proposal, McGill University.

149. Sanger, F., Brownlee, G.G., and Barrell, B.G. (1965) J. Mol. Biol. 13, 373-398.

150. Mushynski, W.E. (1970) Ph.D. Thesis, McGill University.

151. Warner, H.P., and Hobbs, M.D. (1967) Virology 33, 376-384.

152. Spencer, J.H., and Boshkov, L.K. (1973) Can. J. Biochem. 51, 1206-1211.

153. Hughes, S.G., and Brown, P.R. (1973) Biochem. J. 131, 583.

154. Mol, J.N.M., Borst, P., Grosveld, F.G., and Spencer, J.H. (1974) in preparation.

155. Abstracts of the EMBO Workshop on Restriction Enzymes and DNA Sequencing (1974) Ghent, Belgium.

156. Randerath, K., and Randerath, E. (1967) Methods in Enzymology, Vol. 12A, p 323, eds. Grossman, L., and Moldave, K. Academic Press, New York and London.

157. Snyder, L.R. (1972) J. Chromatog. Sci. 10, 369-379.

158. Mushynski, W.E., and Spencer, J.H. (1970) J. Mol. Biol. 51, 107-120.

159. Mushynski, W.E., and Spencer, J.H. (1970) J. Mol. Biol. 52, 91-106.

160. Hall, J.B., and Sinsheimer, R.L. (1963) J. Mol. Biol. 6, 115-127.

161. Sephadex Ion Exchangers Pamphlet (1968) Pharmacia Fine Chemicals, Uppsala, Sweden.

162. Laird, C.D., McConaughy, B.C., and McCarthy, B.J. (1969) Nature 224, 149-154.

163. McConaughy, B.L., Laird, C.D., and McCarthy, B.J. (1969) Biochemistry 8, 3289-3295.

164. Harbers, K., and Spencer, J.H. (1974) Biochemistry 13, 1094-1101.

165. Tomlinson, R.V., and Tener, G.M. (1963) Biochemistry 2, 697-702.

166. Carrara, M., and Bernardi, G. (1968) Biochemistry 7, 1121-1131.

167. Van Kreijl, C.F., Borst, P., Flavell, R.A., and Hollenberg, C.P. (1972) Biochim. Biophys. Acta 277, 61-70.

168. Denhardt, D.T. (1972) J. Theoret. Biol. 34, 487-508.

169. Dressler, D.H., (1970) Proc. Nat. Acad. Sci. U.S.A. 67, 1934-1942.

170. Henry, T.J., and Knippers, R. (1974) Proc. Nat. Acad. Sci. U S.A. 71, 1549-1553.

171. Gilbert, W., Maizels, N., and Maxam, A. (1973) Cold Spring Harb. Symp. Quant. Biol. 38. in press.

172. Wilson, D.A., and Thomas, C.A., Jr. (1974) J. Mol. Biol. 84, 115-144.

173. Wu, R., Padmanabhan, R., and Bambara, R. (1974) in Methods in Enzymology, Vol 29, part E; eds. Grossman, L., and Moldave, K. Academic Press, New York and London.

174. Harbers, K., Harbers, B., and Spencer, J.H. (1974) Biochem. Biophys. Res. Comm. 58, 814-821.

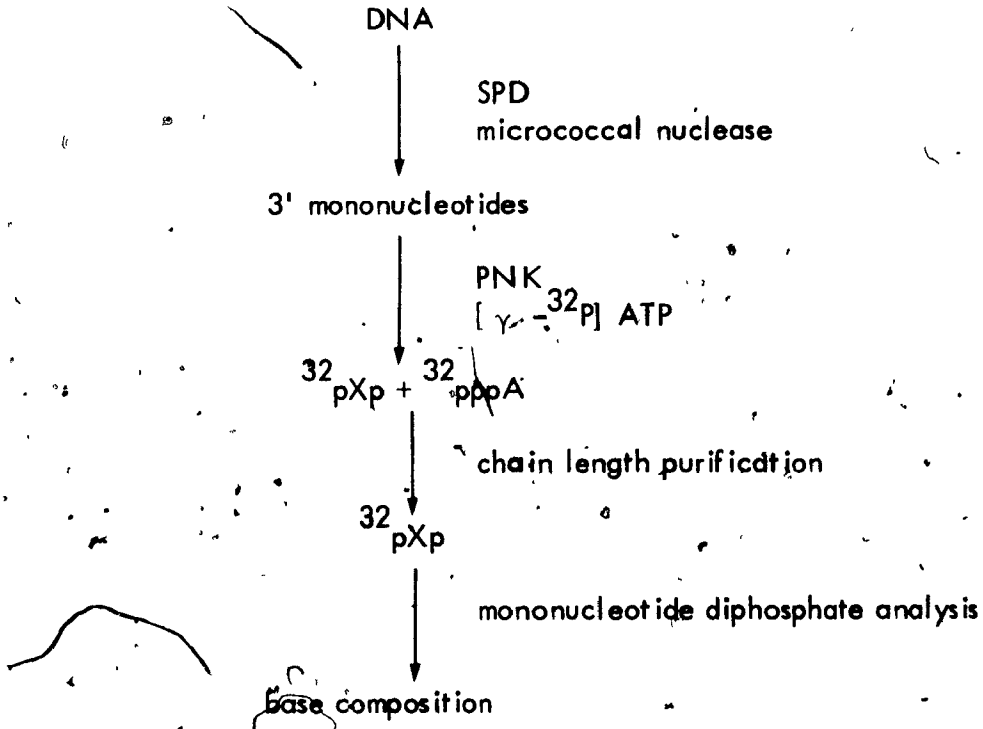
APPENDIX I

Properties of Bacteriophage S13suN15 DNA

APPENDIX II

Proposal : The Use of Polynucleotide Kinase in the Ultramicrodetermination of Base Composition.

Since it is possible to prepare [γ-^{32}P] ATP of very high specific activity, a method using polynucleotide kinase to label mononucleotides for base composition analysis would be extremely sensitive. It has been observed that polynucleotide kinase will transfer phosphate to the 5' hydroxyl of a 3' mononucleotide (12). If this reaction can be made quantitative, the following scheme becomes feasible:



To completely purify the mononucleoside diphosphate fraction it might be necessary to include two chain length fractionation steps, one at pH 5.5, where pXp has a charge of -2, and one at pH 8 , where pXp has a charge of -4.

Base composition analysis on high molecular weight DNA or on 3' phosphorylated oligonucleotide, the 3'-terminal nucleoside will not be labelled. When the substrate is a 5' phosphorylated oligonucleotide, the 5' terminal nucleoside will not be labelled to the same extent as the rest of the nucleosides.

This method may be of use in the determination of sequence on picomolar quantities of oligonucleotide using venom exonuclease digestion from the 3' end as a first step. This could extend the range of sequence methods based on polynucleotide kinase by fifteen nucleotides, to oligonucleotides containing 40 bases.