Lynda Joy Simes

"THE DETERMINATION OF NUCLEOTIDE ARRANGEMENT IN OLIGONUCLEOTIDES DERIVED FROM DEOXYRIBONUCLEIC ACID"

Submitted for the degree of Master of Science in Biochemistry at Massey University, Palmerston North, New Zealand, 1969.

CONTENTS

ABBREVIATIONS AND CONVENTIONS USED IN REPRESENTING NUCLEOTIDES AND OLIGONUCLEOTIDES IN THIS THESIS

For convenience it has been found necessary to adopt a system of abbreviated representations of pyrimidine oligonucleotides, in accordance with the recommendations of the International Union of Pure and Applied Chemistry (1). All the compounds described in this thesis were deoxyribonucleotide derivatives with a regular 3'→5' internucleotide linkage and the recommended prefix "de-3', 5'-". In this thesis however the prefix has been omitted.

C and T are used to represent deoxycytidine and thymidine respectively, while p is used to represent phosphate esterified with the nucleosides. When p is written to the right of the symbol for the nucleoside this indicates the phosphate is esterified with the 3'-hydroxyl group of the nucleoside; when it is to the left, it forms the 5'-ester-.

Further abbreviations have been adopted for sequences that are expected to consist of mixtures of isomer nucleotides in unknown proportions; these are written with nucleoside symbols in parenthesis. Thus CpT represents deoxycytidylyl-(3'→5'-)-thymidine, TpC represents thymidylyl-(3'→5'-)-deoxycytidine and (CT)p represents a mixture of these two

isomers in unspecified proportions.

Several terms describing various types of isomers, used in this thesis, require definition. ISOPLITHS are oligonucleotides of identical chain length. Any group of isopliths, derived from digestion of a naturally occurring nucleic acid, will almost certainly contain sequences that differ from each other in their base composition. These are termed COMPOSITIONAL isomers. An oligonucleotide fraction representing a single compositional isomer is likely to consist of several SEQUENTIAL isomers, which differ from each other only in actual arrangement of bases.

Other abbreviations used in this thesis are:

DNA	deoxyribonucleic acid
RNA	ribonucleic acid
C	deoxycytidine
T	thymidine
Pu	purine
Py	pyrimidine
s.v.	snake venom phosphodiesterase
EDTA	ethylenediamine-tetra-acetate
Tris-HCl	tris(hydroxymethyl)amino methane-HCl

CMCD or CME-carbodiimide	N-cyclohexyl-N'-β-4-methyl-morpholinium ethyl carbodiimide
Pi	inorganic phosphate
UV	ultraviolet light
O.D.	optical density
U	an enzyme unit
E	extinction
$\frac{E_{280}}{E_{260}}$	ratio of extinction at 280 mμ and at 260 mμ

Introduction

(i) The importance of DNA sequences

The genetic material of all animals, plants bacteria, and of many animal and bacterial viruses has long been established as deoxyribonucleic acid or DNA, and recent studies (1a) have shown that its major function is to carry the genetic information required by a cell for the synthesis of species specific proteins.

This information is stored by the nucleic acid macromolecule in the form of a linear code determined by its intrinsic nucleotide sequence or primary structure. The information is carried in such a way that a specific sequence of three nucleotides has the ability to code for one of every type of amino acid found in protein. In recent years the message corresponding to each nucleotide triplet has been established (2).

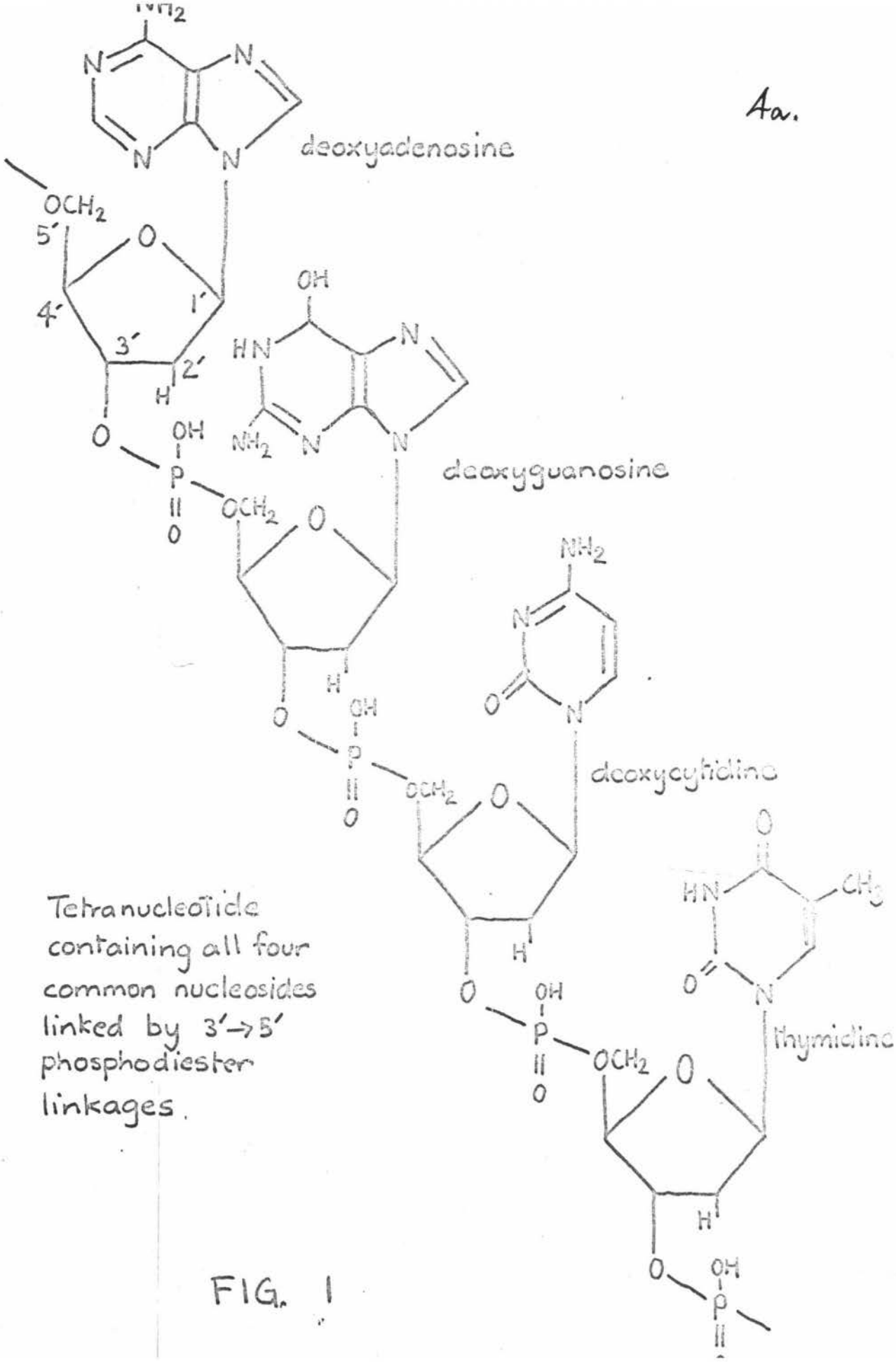
Besides coding for amino acids, nucleotide sequences exist which are known to code for ribosomal transfer RNA's. There are probably other sequences which are involved in a variety of special roles, the more important being

regulating and initiating transcription of the genetic message into functional messenger RNA, and the initiation of DNA replication. Because little information is available about the actual arrangement of bases needed to effect these functions, a knowledge of the complete nucleotide sequence of a biologically active DNA molecule may help to elucidate the nature of these extremely important processes. In addition it is hoped that new approaches can be found towards a better understanding of the actual changes to DNA caused by mutagens and carcinogenic agents, and it is also hoped that some knowledge can be obtained of the extent to which degenerate codons exist in genetic material, and the function of these codons when they do occur.

(ii) The structure of DNA

DNA molecules are linear polymers of deoxyribonucleotides linked through 3'→5' phosphodiester bridges. Four major nucleotides, two containing purine bases (adenine and guanine), and two containing pyrimidine bases (cytosine and thymine), are commonly found. A tetranucleotide containing an arrangement of these four bases is illustrated in figure 1.

In most cases the DNA molecule occurs in the form of a double right-handed helix consisting of two complementary strands of nucleotides held together by hydrogen bonding



Tetranucleotide containing all four common nucleosides linked by 3′→5′ phosphodiester linkages.

FIG. 1

(3). Stereochemical considerations restrict the formation of a stable hydrogen bonded structure to accommodate only pairing between adenine and thymine, or guanine and cytosine. Because of this restriction the two strands have complementary structures, and thus if the nucleotide sequence of one strand is specified then the sequence of the other (complementary) strand can be deduced.

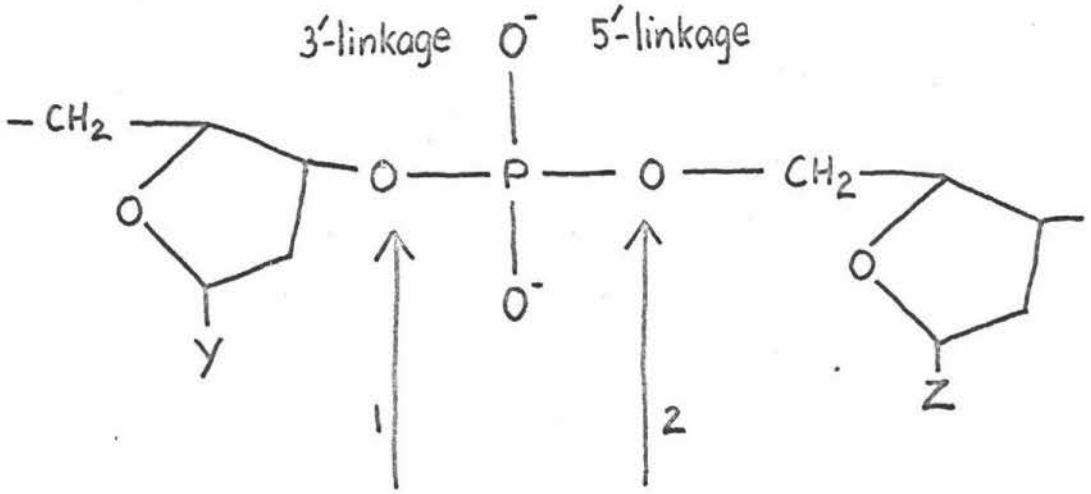
There are some exceptions to this model: the DNA of some small bacteriophages, e.g. the coliphages f1 and ØX 174, is in the form of a single covalently closed circular strand. Some other DNA's have been shown to have a circular double-stranded structure (4, 49).

Polynucleotide strands are polarised in the sense that each has a 3' and a 5' terminus, and in the Watson-Crick DNA model, the two strands of the duplex have opposite polarity.

As a consequence of the 3'→5' phosphodiester bonding of nucleosides, the internucleotide linkage can be broken in two possible ways, as is demonstrated in figure 2. This property has been found very useful in this investigation.

(iii) Determination of nucleotide sequences (15, 40, 50, 54)

Ribonucleic acid, a ribonucleotide polymer, has a similar primary structure to DNA, and is widely



Sites of hydrolysis at the internucleotide linkage

FIG. 2

distributed in cells where it has roles in the translation of the DNA genetic information into active cellular proteins. The primary structures of some small biologically active RNA molecules have been determined by the use of specific hydrolytic enzymes (endonucleases) which catalyse the cleavage of the polynucleotide chain into smaller, overlapping and identifiable sequences (5, 6).

Attempts at determination of nucleotide sequences in DNA have in the main followed along similar lines to these, but the problem is considerably more difficult than for RNA for several reasons:

a) As most cells contain many discrete chromosomes, the cellular DNA content is normally a complex mixture of double-stranded molecules; viruses, especially bacteriophage, provide the only known source of sequentially homogenous DNA. Although the DNA contained in the smallest bacteriophage is a single strand of nucleotides, it still has the high molecular weight of approximately 1.8×10^6, and contains at least 5,500 nucleotides (22).

b) Incomplete knowledge of the specificities of the few purified endonucleases that are capable of hydrolysing DNA prevents the use of these enzymes as analytical tools.

Because of this last restriction, it has been necessary to find suitable chemical aids to degrading

specific internucleotide bonds, and several of these have been described (7 - 12). The most successful of these has been the degradation induced by aromatic amines, particularly diphenylamine, in acid solution (13, 14). When DNA is incubated in 66% (v/v) formic acid containing 2% (w/v) diphenylamine, the acid conditions lead to the quantitative cleavage of the purine-deoxyribose glycosidic linkages, while the stable pyrimidine deoxyribose glycosidic bonds are completely untouched. Subsequent reaction of the resultant "apurinic acid" with the aromatic amine involving a β-elimination mechanism results in the release of inorganic phosphate from positions between adjacent purine nucleosides yielding pyrimidine nucleoside phosphates of general formula Py_nP_{n+1} (13). The reaction is shown in figure 3.

This chemical reaction gives reproducible results and it is quantitative; it produces from any DNA molecule a complex mixture of pyrimidine sequences of varying chain length and base composition. Although detailed methods have been developed for separation of these pyrimidine products into groups on the basis of their cytosine and thymine content, it has proved impossible to resolve mixtures of sequential isomers into their components. For example, the trinucleotide fraction of base composition $(CT_2)p_4$ may contain up to three sequential isomers, TTC,

Diphenylamine Reaction

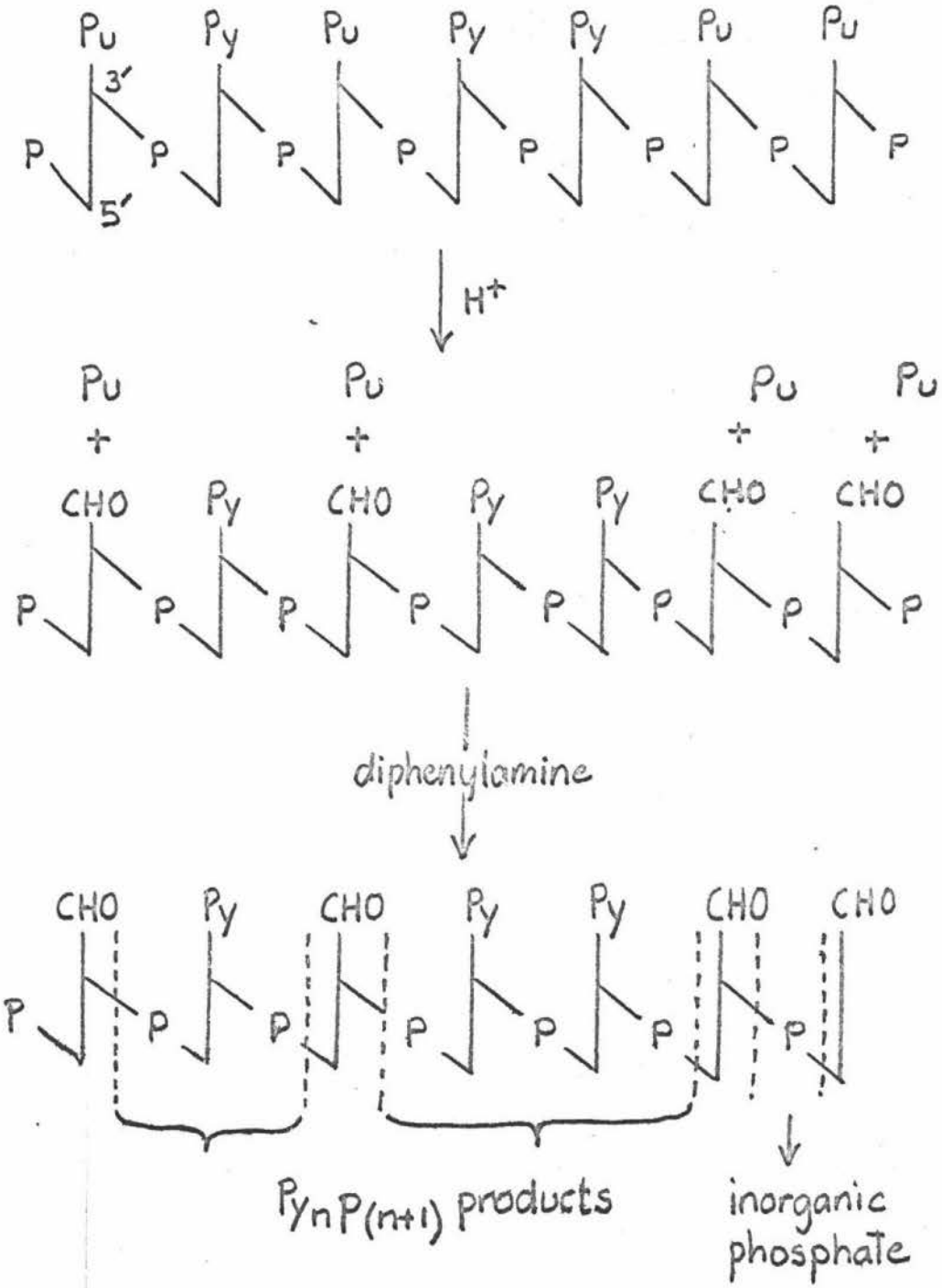


Fig 3

TCT, and CTT, which remain inseparable.

With longer nucleotides the problem becomes more complex, as larger numbers of sequential isomers can theoretically exist according to the general formula which relates numbers of possible sequential isomers to the chain length of the oligonucleotide:

For the isomer $(C_a T_b)p_{a+b+1}$ the number of sequential isomers is given by:

$$N = \frac{(a+b)!}{a! \times b!}$$

(iv) The present study

In the first part of this study, the digestion products from the action of diphenylamine on calf thymus DNA in acid solution have been separated as far as possible by physical methods, characterised, and measured quantitatively.

In addition, an investigation into possible methods of deducing the relative proportions of each sequential isomer in a mixture of isomers of identical base composition was undertaken, with special emphasis on the use of exonucleases (15).

A possible approach to this problem is found from a study of the specificities of snake venom and bovine spleen phosphodiesterases. Although neither of these enzymes will

attack a fully phosphorylated substrate of the type Py_nP_{n+1}, both will catalyse the exonucleolytic cleavage of the Py_nP_{n-1} oligonucleotides obtained by terminal dephosphorylation (16, 17):

$$\underset{(Py_nP_{n+1})}{pCpTpTp} \xrightarrow{\text{phosphomonoesterase}} \underset{(Py_nP_{n-1})}{OH\text{-}CpTpT\text{-}OH} + 2Pi$$

The specificity of snake venom phosphodiesterase is such that it will degrade terminally dephosphorylated oligonucleotides to 5'-mononucleotides, commencing its attack at the 3'-terminus, while the spleen enzyme will catalyse a similar degradation from the 5'-terminus (18).

For example, in the case of the $(CT_2)P_2$ which can exist as the following three sequential isomers (a) - (c):

(5')	(3')	
CpCpT		(a)
CpTpC		(b)
TpCpC		(c)

the two types of enzymic attack will proceed as follows:

$$C + pC + pT \xleftarrow{\text{s.v.}} CpCpT \xrightarrow{\text{spleen}} CpC + Cp + T$$

$$C + pT + pC \xleftarrow{\text{s.v.}} CpTpC \xrightarrow{\text{spleen}} Cp + Tp + C$$

$$T + pC + pC \xleftarrow{\text{s.v.}} TpCpC \xrightarrow{\text{spleen}} Tp + Cp + C$$

It will be seen that mononucleotide thymidine will arise only from the action of the spleen enzyme on isomer (a), or the snake venom enzyme on isomer (c). Measurement of the amount of this product in each type of enzyme digest should therefore allow the relative proportions of these two isomers in the mixture to be determined. Measurement of the other products will help confirm these values.

As shown above, there can be far more sequential isomers in the tetranucleotide and longer fractions than in the trinucleotide fraction. The use of these two nucleases alone therefore proves inadequate for determining the relative proportions of sequential isomers in these cases, although the result of their action does provide useful information. In other experiments therefore, attempts have been made to modify the specificities of the two nucleases, in this way deriving extra tools for the investigation of nucleotide sequences. A chemical blocking agent which attaches to the oligonucleotide was used for the modification (6, 38, 39, 41, 42, 43).

Experiments in which attempts were made to separate trinucleotide sequential isomers by ion exchange chromatography are also described.

MATERIALS

Deoxyribonucleic acid

Calf thymus DNA was prepared by the method of Kay, Simmons and Dounce (19) as modified by Chargaff (20).

Snake venom phosphodiesterase

(phosphodiesterase I E.C. 3.1.4.1.)

Phosphodiesterase isolated from Russel's viper venom was purchased from CalBiochem (Los Angeles, California, U.S.A.).

Spleen phosphodiesterase

(phosphodiesterase II E.C. 3.1.4.1.)

Bovine spleen phosphodiesterase was obtained from Worthington Biochemical Corp., (Freehold, New Jersey, U.S.A.). Each batch of enzyme (10-15 units) was dissolved in 1.0 ml water and stored frozen.

Phosphomonoesterase

Bacterial (*Escherichia coli*) alkaline phosphomonoesterase, chromatographically purified, was obtained from Worthington Biochemical Corp., (Freehold, New Jersey, U.S.A.) as a suspension in 65% $(NH_4)_2SO_4$. Each batch of enzyme (approximately 10 mg protein) was mixed with 2 ml

of 0.01 M $MgCl_2$, dialysed against 0.01 M Tris-HCl, pH 8.0 containing $MgCl_2$(0.01 M), heated at 85 - 90° for 10 minutes and cooled on ice.

Sodium p-nitrophenyl phosphate

Sodium p-nitrophenyl phosphate was obtained from the British Drug Houses Ltd., (Poole, England).

Sodium p-nitrophenyl thymidine-5'-phosphate

Sodium p-nitrophenyl thymidine-5'-phosphate was obtained from CalBiochem (Los Angeles, California, U.S.A.).

Ion Exchange Resins

i) DEAE cellulose (microgranular)

Whatman microgranular diethylamino ethyl cellulose, DE32, was obtained from H. Reeve Angel and Co. Ltd. (London, Eng.)

ii) DEAE cellulose (fibrous)

Cellex D (standard capacity) was obtained from Bio Rad (Los Angeles, California, U.S.A.). It had an exchange capacity of 0.7 meq/gm.

iii) DEAE-Sephadex A-25

DEAE Sephadex A-25 was obtained from Pharmacia (Uppsala, Sweden) and had a capacity of 3.5 meq/g and particle size 40 to 120 μ.

iv) Dowex-1, Dowex-2, and Dowex-50 x 8

Dowex-1, Dowex-2 and Dowex-50 x 8, all 200-400 mesh, were obtained from the Dow Chemical Company, (Michigan, U.S.A.).

Triethylamine

Triethylamine (reagent grade) was obtained from British Drug Houses Ltd, (Poole, Dorset, England). This was redistilled after refluxing over KOH pellets for an hour, and the fraction boiling at 89° collected.

Triethylamine bicarbonate

A stock solution of 2M triethylamine bicarbonate was prepared by slowly dropping redistilled triethylamine into ice cold distilled water through which CO_2 from dry ice was continuously bubbled. The rate of addition of triethylamine was controlled to maintain a pH of about 9 during the reaction. The solution was finally adjusted to pH 8.6 and stored at 4°.

Diphenylamine

Analar grade diphenylamine, obtained from British Drug Houses (Poole, England) was recrystallised several times from aqueous ethanol.

Formic Acid

Analar grade (98 - 100%) formic acid was obtained from British Drug Houses (Poole, England).

Urea

Reagent grade urea was obtained from May and Baker (Dagenham, England) and UV absorbing material removed by passing a solution through a column of charcoal:celite (5:1). The concentration of the stock solution was determined from its specific gravity. As strong urea solutions can extract UV absorbing materials from rubber tubing and some plastic tubing, the same rubber stoppers and plastic tubing were used throughout the work so that most of this unwanted material was soon eliminated.

N-cyclohexyl-N'-β-4-methyl-morpholinium ethylcarbodiimide (CME-carbodiimide or CMCD)

CME-carbodiimide was purchased from Aldrich Chemical Company.

METHODS

Enzyme assays

Snake Venom Phosphodiesterase

Snake venom phosphodiesterase was assayed by the method of Razzel and Khorana (21). The assay was performed in a cuvette containing 100 μmoles of Tris buffer, pH 8.9, 0.5 μmole of p-nitrophenyl thymidine-5'-phosphate, and 10 μmoles of $MgCl_2$, in a total volume of 1.0 ml. Prior to the addition of the enzyme, the cuvette was equilibriated at 37°. On addition of the enzyme the increase in optical density due to the release of p-nitrophenylate was measured at 400 mμ. An increase in optical density of 1.200 units was taken to be equivalent to the hydrolysis of 0.1 μmole of substrate.

Phosphomonoesterase activity of this enzyme was checked by measuring the release of inorganic phosphate from the same substrate.

Bacterial Alkaline Phosphatase

Alkaline phosphatase activity was measured from the rate of hydrolysis of a solution of p-nitrophenyl phosphate (0.001 M in 1.0 M Tris buffer, pH 8.0). Three mls of substrate was pipetted into each of two 1 cm cuvettes (one the blank). At zero time 0.1 ml of the enzyme solution was added to the test cuvette and the optical density was

recorded at 1 minute intervals. The rate of release of p-nitrophenol from the substrate was determined by following changes at 410 mμ, and one unit was taken as the quantity of enzyme liberating 1 μmole of p-nitrophenol under defined conditions at 25° (23).

Paper Chromatography

All separations were carried out on Whatman No.1 and 3 MM chromatography paper using solvents described in the text.

Thin Layer Chromatography

20 x 20 cm glass plates were scrubbed with detergent, washed twice in a glassware washer at 80°, air dried, and finally cleaned with acetone to remove all traces of grease.

MN cellulose 300 F_{254} (obtained from Machery and Nagel, Düren, West Germany) was mixed with a cadmium borate fluorescent indicator (24), in the proportion ten parts to one, suspended in a mixture of water and acetone (2:1), and the slurry used to coat plates to a thickness of 0.3 mm. Water was removed from the layers, either by long drying at 37°, or by drying in an oven for several hours. Randerath (25) suspended dry cellulose absorbant in acetone, resulting in a drying time of a few minutes being quite adequate, but when tried in this investigation, the method gave fluffy and very fragile layers.

Compounds were applied with a micropipette and dried immediately in a stream of warm air.

Tanks for ascending development were used. A saturated atmosphere was obtained by lining two sides of the tank with filter paper soaked in the appropriate solvent.

Paper Electrophoresis

A "Miles Hivolt" apparatus, which accommodates sheets of paper 17" x 21" was used. Whatman No.1 paper, moistened with buffer solution, was placed over the cooled flat plate between two electrode vessels filled with buffer. The paper, the ends of which dipped into these vessels, was sandwiched between two sheets of insulating polythene. The two water cooled metal plates, held together under air pressure, allowed a high voltage gradient to be employed without heating effects. The sample was applied with a micropipette to the moist paper. All separations were run in formate buffer pH 2.7.

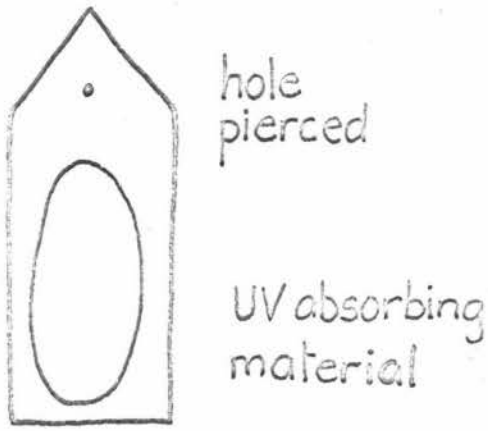
Location of Ultraviolet-Absorbing Areas on Filter Paper Sheets

These were located using a mercury filament lamp (with a gaseous chlorine filter) as UV light source, and Ilford Reflex Contact Document paper No. 50M was used to make a permanent record (26). For this purpose close

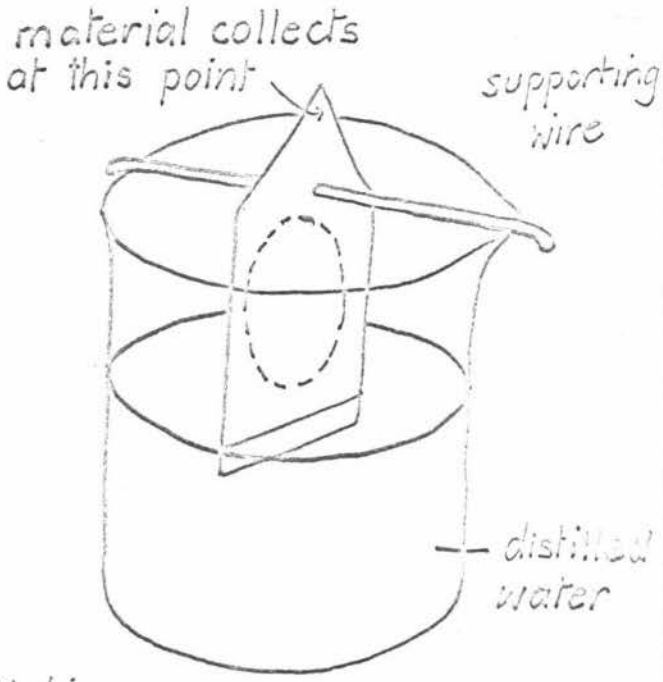
contact between the chromatogram and the photographic paper was achieved through the use of a frame consisting of a convex plywood board over which a double thickness of 0.005 inch gauge polythene sheet was tightly stretched and held in place with a pair of spring fasteners (27). The contact prints were developed with Ilford Bromophen Developer.

Elution of Nucleotides from Paper Chromatograms and Electrophoretograms

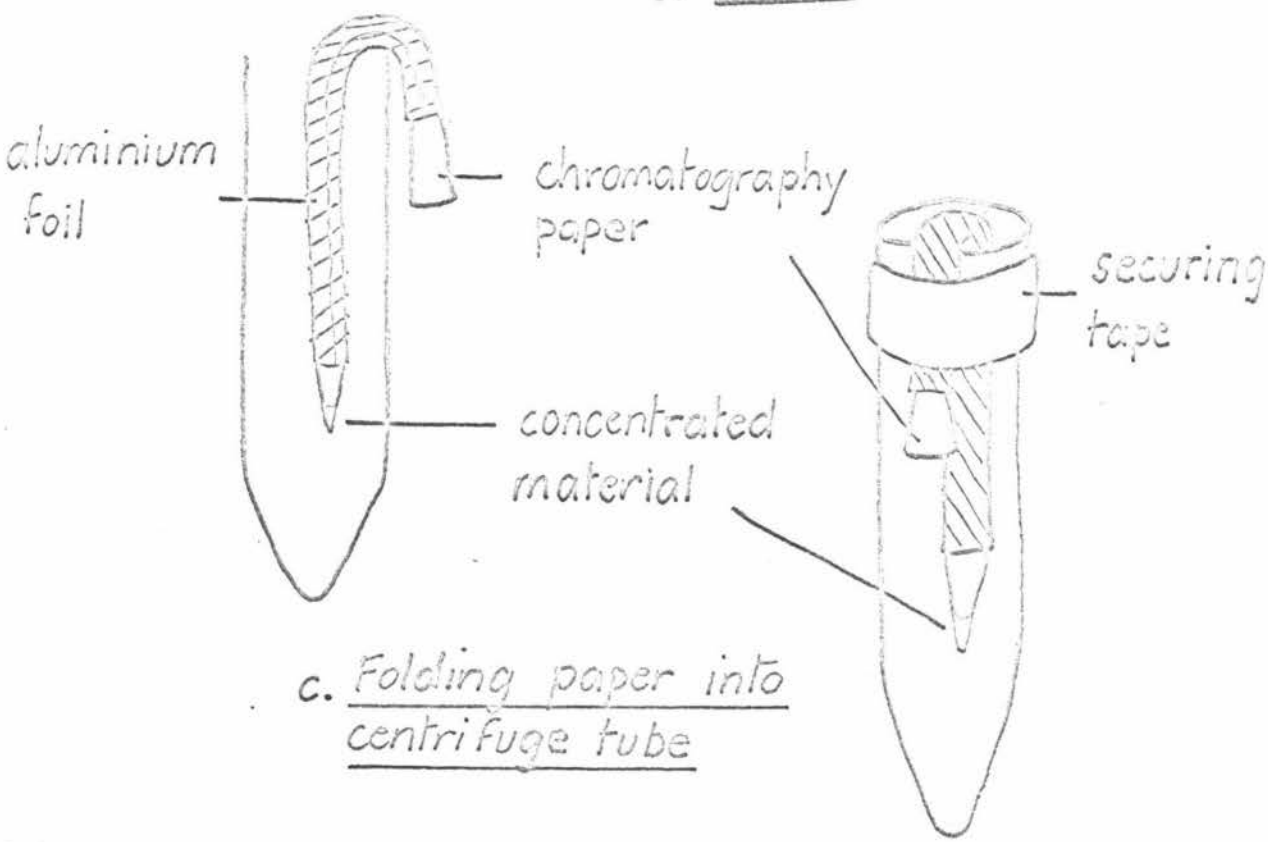
After being located, the spots were outlined and cut as shown in figure 4. A hole was pierced between spot and apex, the paper suspended by wire covered with polythene tubing and the lower edge dipped into distilled water. As water flowed up the paper by capillarity and evaporated, the nucleotide material was drawn up and concentrated in a small area at the apex after an overnight run. The wet paper was then wrapped in aluminium foil and folded several times lengthwise, forming a rigid tube. It was placed point down in a centrifuge tube, suspended about an inch from the bottom with the other end of the paper bent over the top of the tube and secured with tape. Centrifuging for a few minutes at 5000 rpm drew fluid containing the nucleotide material out from the paper to the bottom of the tube. When required, the eluates were concentrated by



a. Way spot is cut from chromatogram



b. Elution



c. Folding paper into centrifuge tube

ELUTING NUCLEOTIDE MATERIAL FROM CHROMATOGRAMS

Fig 4.

evaporation at 37° in a stream of air.

Paper Blanks

For quantitative work, equivalent areas adjacent to the nucleotide spots were cut from the chromatograms. These were eluted, the optical density (in UV light) determined, and this value used as a measure of the contribution made by background absorbance from paper, solvent and buffer. However, variable results were obtained, and an attempt was therefore made to construct a standard curve for calculating theoretical blank values for chromatograms run in one frequently used solvent. This method was not very advantageous, however, as it involved time consuming determination of paper areas and their location on the chromatogram. The procedure finally adopted was the measurement of the optical density of the eluates at 310 mμ, where extraneous material, but not nucleotide material, absorbs. This reading was then substracted from those made at the lower wavelengths (28).

Ion Exchange Chromatography on DEAE Cellulose

Preparation of DEAE Cellulose

1) DE 32

(a) Precycling. The solid exchanger was stirred into 15 volumes of 0.5M HCl and left for 30 minutes. The

supernatant liquid was decanted off and the material washed until the effluent was at pH 4, then stirred with 15 volumes of 0.5M NaOH, left for another 30 minutes, and washed in a funnel until the effluent was near neutral. The material was equilibrated by stirring in a large volume of the required buffer, allowing it to settle, and decanting, many times over. This procedure also removed all fines which are generated by physical and chemical breakdown of the cellulose material. The exchanger was stored in the wet state.

(b) Column preparat on. Occluded air was removed from a slurry of the cellulose by means of a gentle vacuum, and the slurry was poured into the glass column of required dimensions in as few passes as possible. When the bed was formed in sections the upper layers of each section were stirred slightly before the next was added. The first part of the column was allowed to settle before the tap was opened. Finally the column beds were settled by passing several bed volumes of starting buffer through them. This also removed a certain amount of unwanted U.V.-absorbing material from the column.

ii) Cellex D

The material was freed from fines by repeated settling and decantation from a large volume of water, then equilibrated with the appropriate buffer. Finally the

resin was resuspended in a small volume of buffer and the column poured in as few stages as possible.

For this material, the recommended ratio of length to diameter of column is between 3/1 and 10/1. As columns of this resin were used almost exclusively for desalting very large quantities of material, columns of size 3 x 20 cms were used in most cases.

Preparation of DEAE Sephadex

The resin was swollen in an excess of water for one hour and the fines removed by repeated decantation. The gel was then washed on a filter with 0.5 M HCl, followed by water, 0.5 M NaOH, and water again. The material was finally equilibrated with the appropriate buffer.

Preparation of Ion Exchange Resins

Cation exchangers were washed with 10% HCl, water till neutral and finally with 25% NaCl to convert the resin to the sodium form. Excess NaCl was washed out with water, and the resin equilibrated with starting buffer.

Anion exchange resins were washed with 2 M NaOH, water, and 3 M HCl until the optical density of the eluate at 260 mμ was negligible. Finally the resin was washed with water and equilibrated with starting buffer.

Application of Samples

Samples were always dissolved in the smallest possible volume of starting buffer and applied carefully to the top of the column.

Salt Gradients

A linear increase in salt concentration was obtained with a two-chambered mixing device of the type described by Parr (29).

Removal of Salt from Nucleotide Solutions

In cases where triethylamine was the only salt present, this could be readily removed by repeated evaporation on a rotary evaporator. In other cases where more effective buffering was required or where non-volatile components (e.g. urea) were added, other means of desalting the solutions were adopted.

i) Desalting on DEAE cellulose

A cellex D column was prepared as described, and the sample applied at pH 9 to 9.5. Salt and urea were washed off with a large volume of 0.01 M triethylamine bicarbonate at pH 8.6 (1500 mls needed for a column 20 cm x 3 cm). The nucleotide was then quantitatively recovered by elution with 100 - 200 mls of 2 M triethylamine bicarbonate, pH 8.6, and the eluant dried down on a rotary evaporator.

ii) Desalting on charcoal (44)

Nucleotide material was adsorbed onto a column

of charcoal:celite (5:1) when the sample was applied at pH 6-7, and urea and salt were washed from the column with distilled water. Ethanol:NH_4OH, (s.g. 0.88):water (50:2:48) was then passed through the column until all UV-absorbing material had been eluted. This solvent was removed from the nucleotide material by rotary evaporation.

Estimation of Phosphorus

Phosphorus was determined by the method of Chen, Toribara and Warner (30), as modified by Burton and Petersen (31). Phosphorus standards and the sample containing up to 8 µg of phosphorus were pipetted into tubes and brought up to 8 mls with water and 4 mls of reagent (6 N H_2SO_4:25% ammonium molybdate:water:10% ascorbate (1:1:1:2)). Each tube was capped with parafilm, and incubated at 37° for 2 hours. The optical density at 820 mµ was proportional to phosphorus concentration, 0.1 µmole of inorganic phosphate giving a reading of 0.71 approx.

EXPERIMENTAL AND RESULTS

I Preparation of Calf thymus oligonucleotides

1) DNA digestion

Approximately 1 g of DNA was dissolved in about 150 mls of distilled water and the solution stirred overnight in the cold room. A further 150 mls of distilled water was added to the viscous solution and the DNA concentration was measured by reading the O.D. at 260 mμ, where 1 mg/ml DNA ≅ O.D. of 20 (approx.)(20, 37). The final solution was found to contain 1530 mg of DNA at a concentration of 5.1 mg/ml.

Two volumes of 3% (w/v) diphenylamine in 100% formic acid were added, the mixture transferred to a glass stoppered flask and incubated at 30° in a water bath for 19 hours. The flask was shaken after about 30 minutes incubation to ensure that all the DNA was properly dissolved and the temperature equilibrated.

When the incubation was complete, an equal volume of distilled water was added to the reaction mixture, and this was transferred to a 2 litre separating funnel, and the mixture was shaken with 2 volumes of ether. The ether layer was re-extracted with ½ - 1 volume of water, and the combined aqueous layers from these extractions were re-extracted four times with fresh ether. The large volume of digest thus obtained was concentrated on the rotary

evaporator (bath temperature not higher than 40°).

ii) Removal of purines and formic acid

A column approximately 2 x 8 cms of Dowex-50 x 8 (H^+) was prepared as described in the materials section and washed with water till neutral. The digest was passed through the column and the effluent kept. The column was washed with distilled water until the O.D. at 260 mμ was below 0.05. The combined water effluents were concentrated on a rotary evaporator to 400 mls. The material was then transferred to a liquid-liquid extractor with efficient cooling so that a good head of liquid ether could be maintained in the collecting tube. This gave a fast flow of ether droplets from the sintered glass plate up through the aqueous digest solution. Removal of last traces of formic acid by the ether was assumed to be complete after 6 hours of continuous extraction.

iii) Standardisation of the digest

Sufficient digest to contain the equivalent of 500 mg of DNA was required for a convenient large scale column separation of pyrimidine oligonucleotides. This volume corresponds to 50 mg of total phosphorus or 14 mg of inorganic phosphurus, since 28% of all phosphorus contained in DNA is released by diphenylamine degradation (31). Inorganic phosphorus was estimated, as described in the methods section, as 3.4 μmole P_i/ml of

digest, and the required volume of digest was calculated to be 133 mls. The remaining digest was dried and stored at 4°.

iv) Separation of the digest products into isopliths

A column of DE 32, of dimension 3.2 x 60 cms, was prepared as described in the methods section, and equilibrated with starting buffer, 7 M urea/0.1 M acetic acid/pH 5.4, (32). Finishing buffer required to produce the desired salt gradient was identical to starting buffer but 0.2 M with respect to NaCl. Both buffers were prepared, deaerated briefly, and allowed to stand at least a day before use.

The digest, which had been evaporated to dryness, was taken up in about 20 mls of starting buffer and applied to the column through which was passed a further volume of starting buffer (several hundred mls) in order to remove all extraneous UV absorbing material before salt elution. The elution of ultraviolet-absorbing components from the column was followed with an ultraviolet flow analyser (Isco, Lincoln, Nebraska, U.S.A.). 3.5 litres of each buffer were put into the appropriate chambers of the gradient apparatus, and large fractions (20 mls) were taken. The contents of every second tube were read on the spectrophotometer at 268.5 mμ and these readings recorded in the form of a graph of the optical density of the

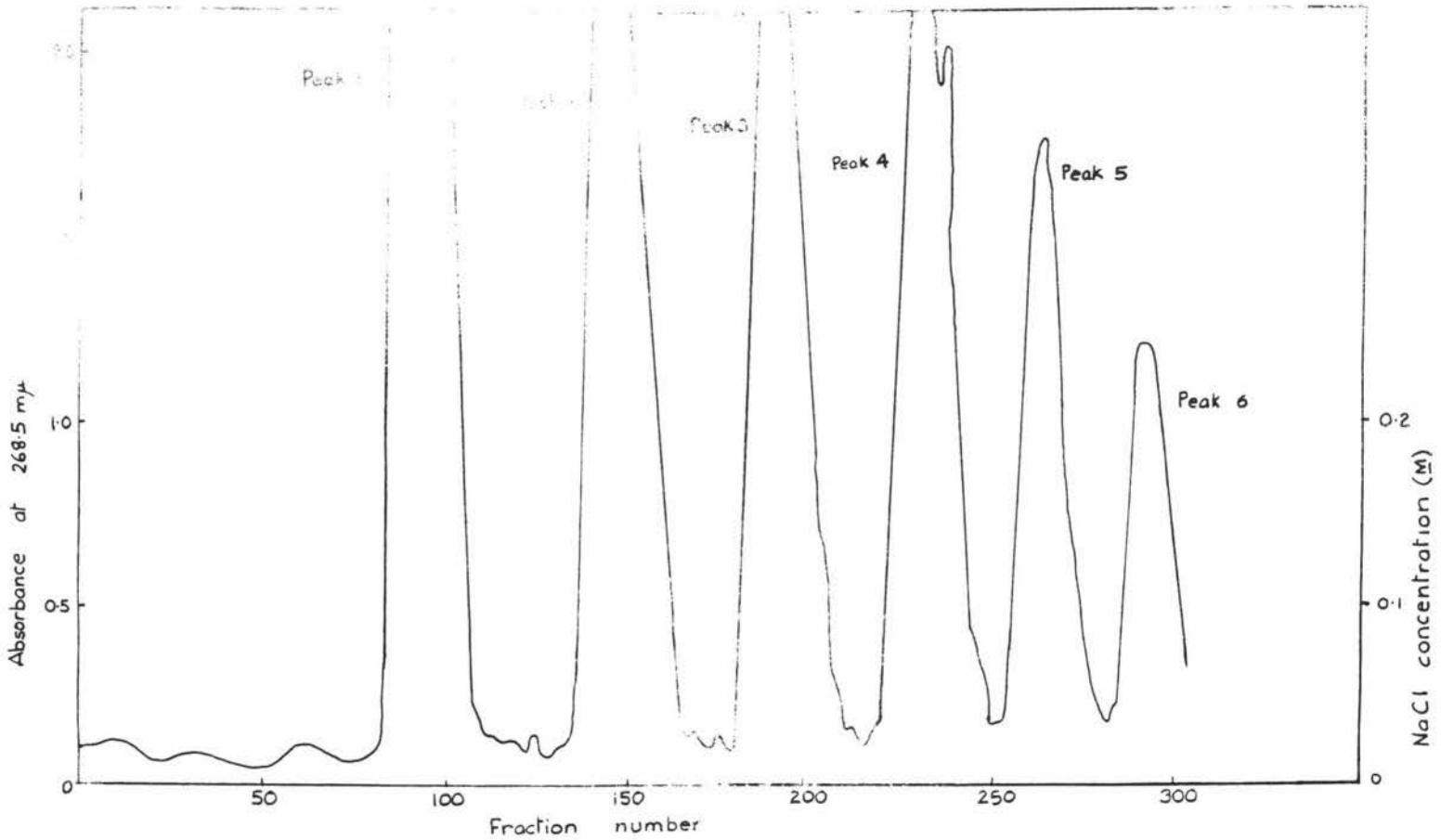


Fig 5

...ation of $Py_n\,P_{nm}$ products into groups of isopliths on a DEAE-cellulose column (60 cm. × 3 cm). The oligonucleotides were eluted with increasing NaCl concentration in a linear gradient. All eluting solutions contained 7 M urea and 0·1 M acetic acid and were adjusted to 5·4 with NaOH. Flow rate approx 55 ml/h, 20 ml fractions.

eluate as a function of fraction number and salt concentration. The separation obtained is illustrated in figure 5. The fractions constituting each peak were pooled and desalted after dilution to give a NaCl concentration below 0.05 M.

Inorganic phosphate is eluted in this system just after the peak containing free purines and extraneous non-nucleotide material.

v) Separation into compositional isomers

Mixtures of oligonucleotides of identical length but different base compositions (isopliths) may be further separated by column chromatography on DEAE cellulose in 7 M urea and 0.1 M formate, with a linear salt gradient (33). The order of emergence of the compositional isomers is consistent with their increasing negative charge at pH 3.1 to 3.5 (values measured for solutions of 7 M urea/ 0.1 M formate with and without 0.1 M NaCl respectively).

DE 32 cellulose was used for all analytical columns. The results of separating the isopliths of chain lengths 2, 3, 4, and 5 into their components are illustrated in figures 6 - 9. It will be seen that of these, only the trinucleotide fractions gave the expected number of components. In all other cases more products were observed than expected. It was concluded that extensive spontaneous dephosphorylation had occurred through a mechanism that

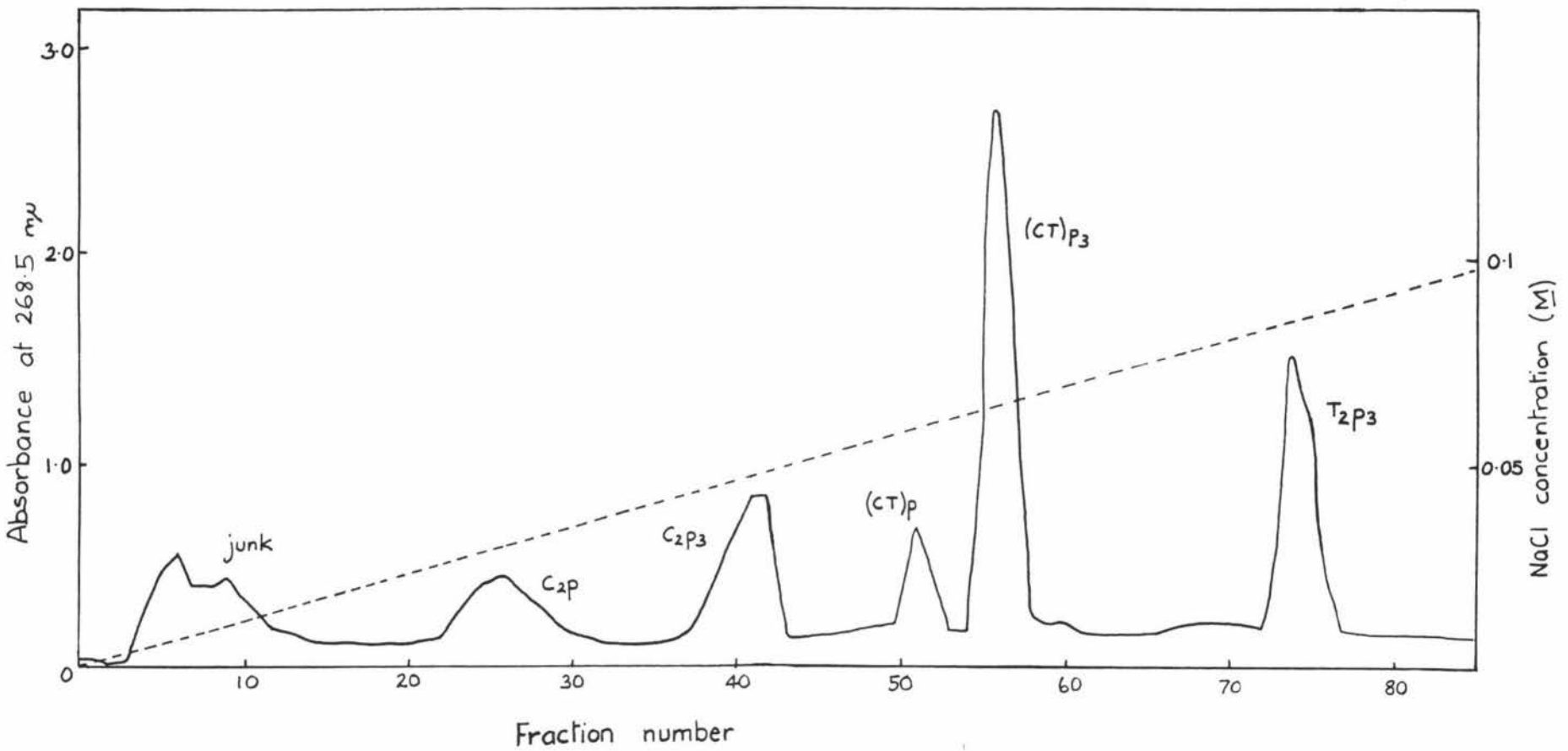


Fig 6

Fractionation of pyrimidine dimers on DEAE-cellulose in 7 M urea / 0·1 M formate by a linear salt gradient 0 - 0·1 M NaCl. Column size 14 cms x 1¾ cms. Fraction size 7 mls.

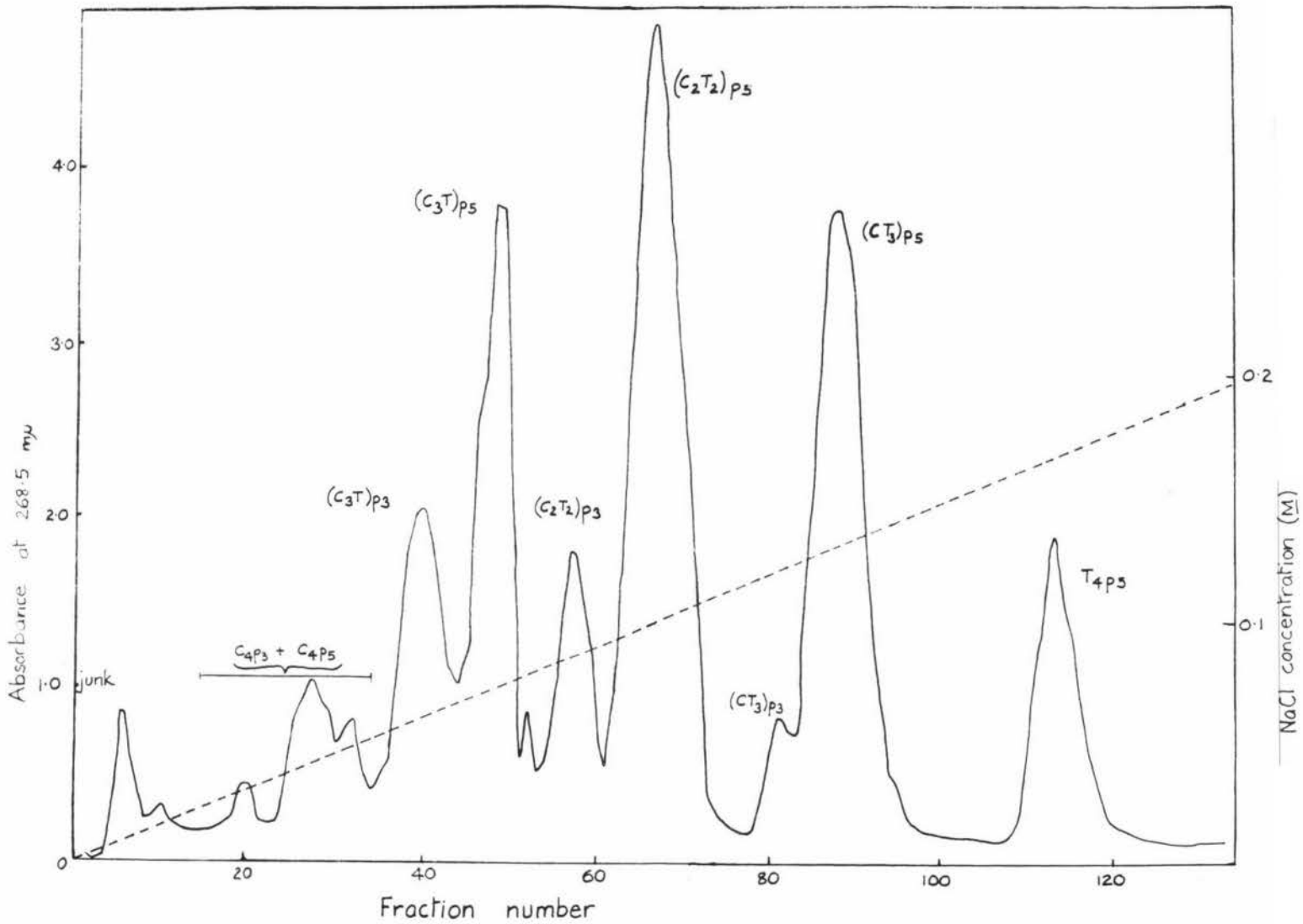


Fig 7.

Fractionation of pyrimidine tetramers on DEAE-cellulose in 7 M urea/0·1 M formate by a linear salt gradient 0 - 0·2 M NaCl. Column size 23cm x 1cm. Flow rate 20mls/hr. Fraction size 6·5 mls.

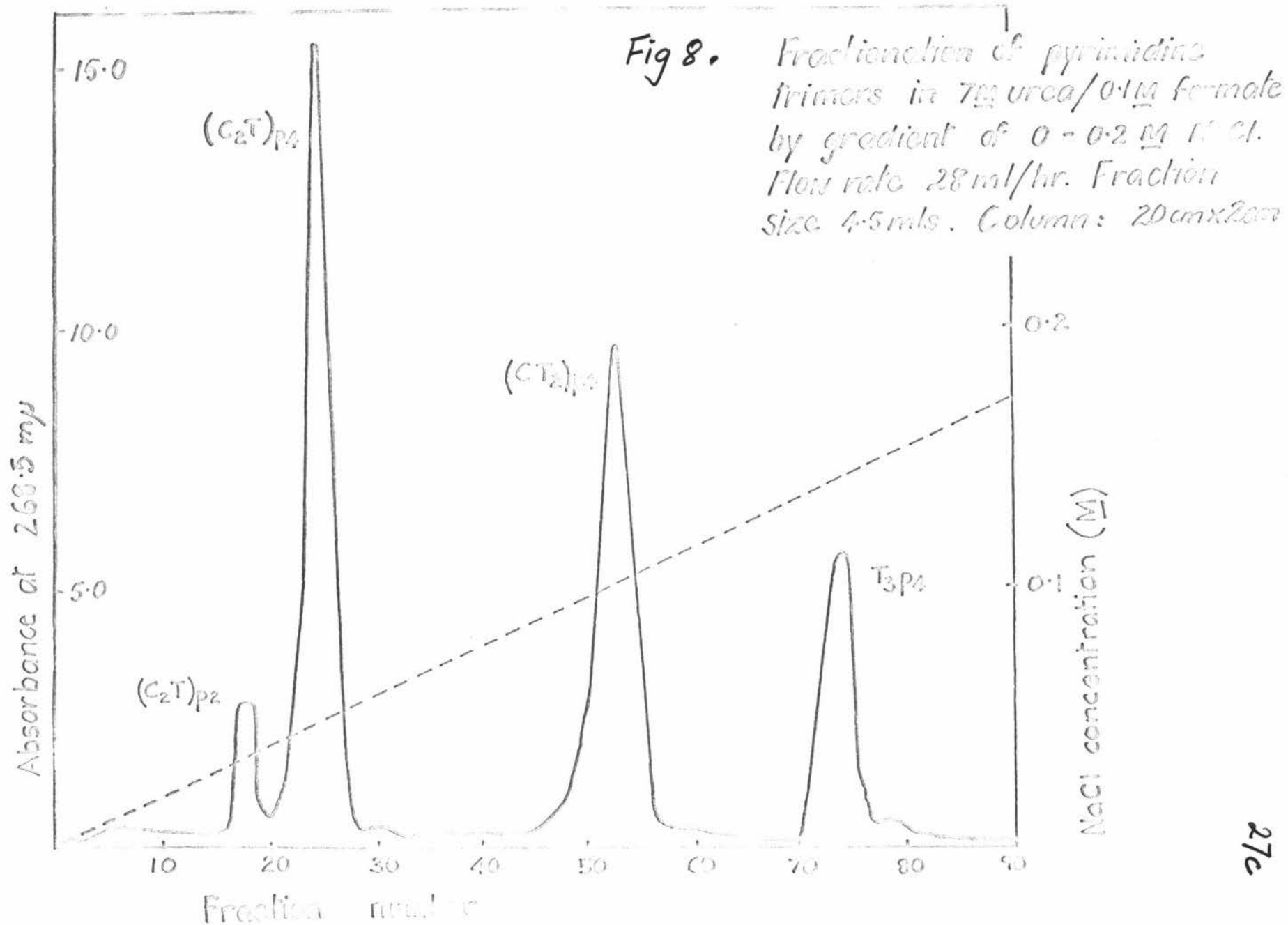


Fig 8. Fractionation of pyrimidine trimers in 7M urea/0·1M formate by gradient of 0 - 0·2 M NaCl. Flow rate 28 ml/hr. Fraction size 4·5 mls. Column: 20cm x 2cm

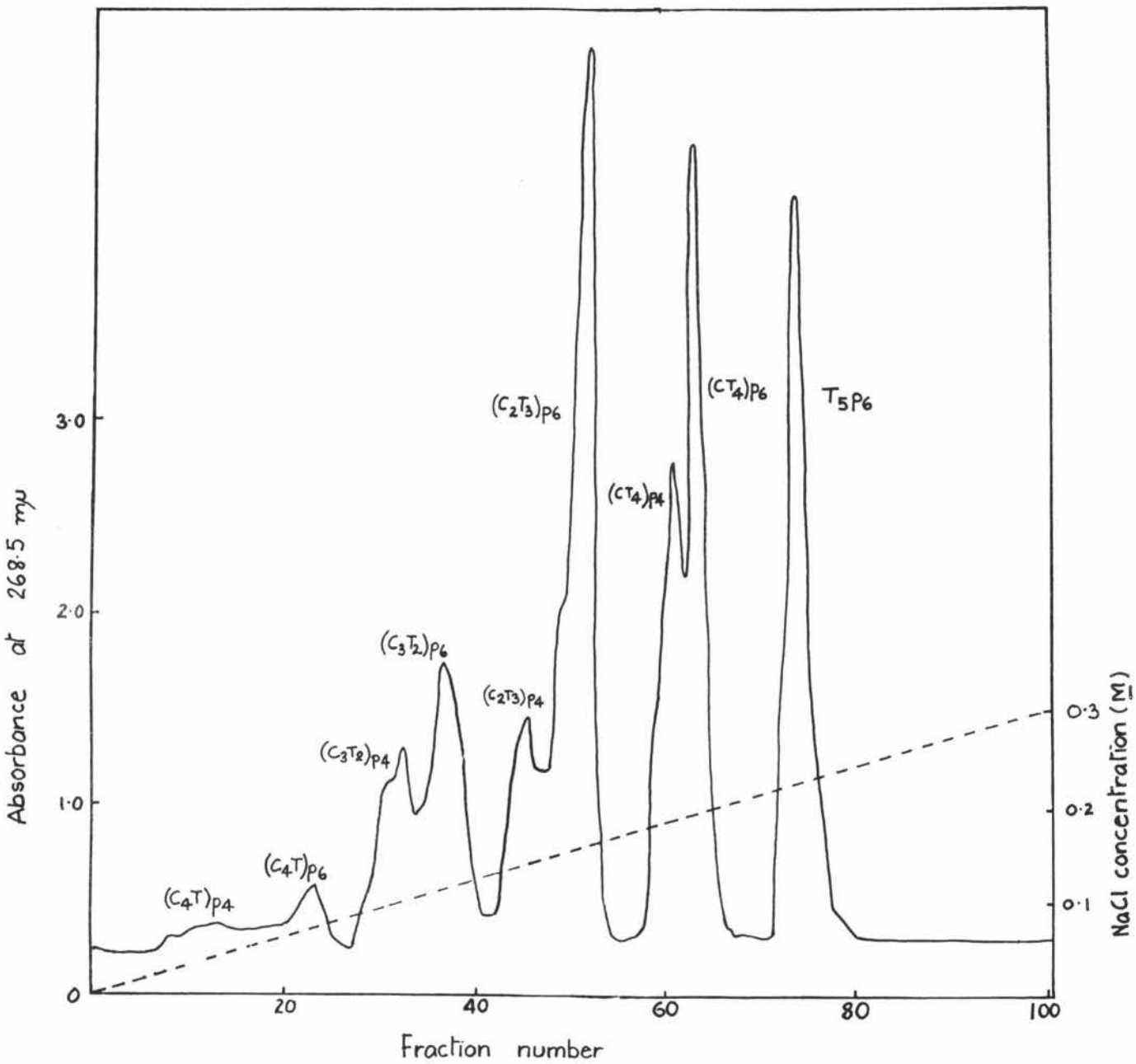


Fig 9.

Fractionation of pyrimidine oligonucleotides of length 5 on DEAE-cellulose in 7 M urea/0·1 M formate by a linear salt gradient 0 - 0·3 M NaCl. Column size 20cms by 1cm. Fraction size 10mls.

is not completely understood. It is interesting to note that sequences rich in cytosine appear to be more susceptible to this type of degradation.

vi) Separation of further material

The remaining digest corresponding to 1000 mg of DNA was separated into isopliths on a large column of DE 32 as before, except that an exponential salt gradient was used. The details of this separation are shown in figure 10. In order to try and reduce spontaneous degradation the peaks were desalted in the shortest possible time and separated in the formic acid system as before. The results are illustrated in figures 11 and 12. It will be seen that much less degradation occurred in this preparation.

An attempt was made to separate the pentanucleotide sequences on a column of DEAE Sephadex A 25 (34). The Sephadex ion exchanger was converted into the acetate form, a column 30 x 0.5 cms poured, and the column washed with 0.01 M sodium acetate buffer pH 3.0. The nucleotide material (which was made alkaline) was placed on the column without previous desalting, the column washed with 100 - 200 mls of sodium acetate 0.01 M pH 3.0 to remove urea and salt. Elution was accomplished with a linear salt gradient 0 - 0.4 M NaCl, in the acetate buffer, pH 3.0. Although

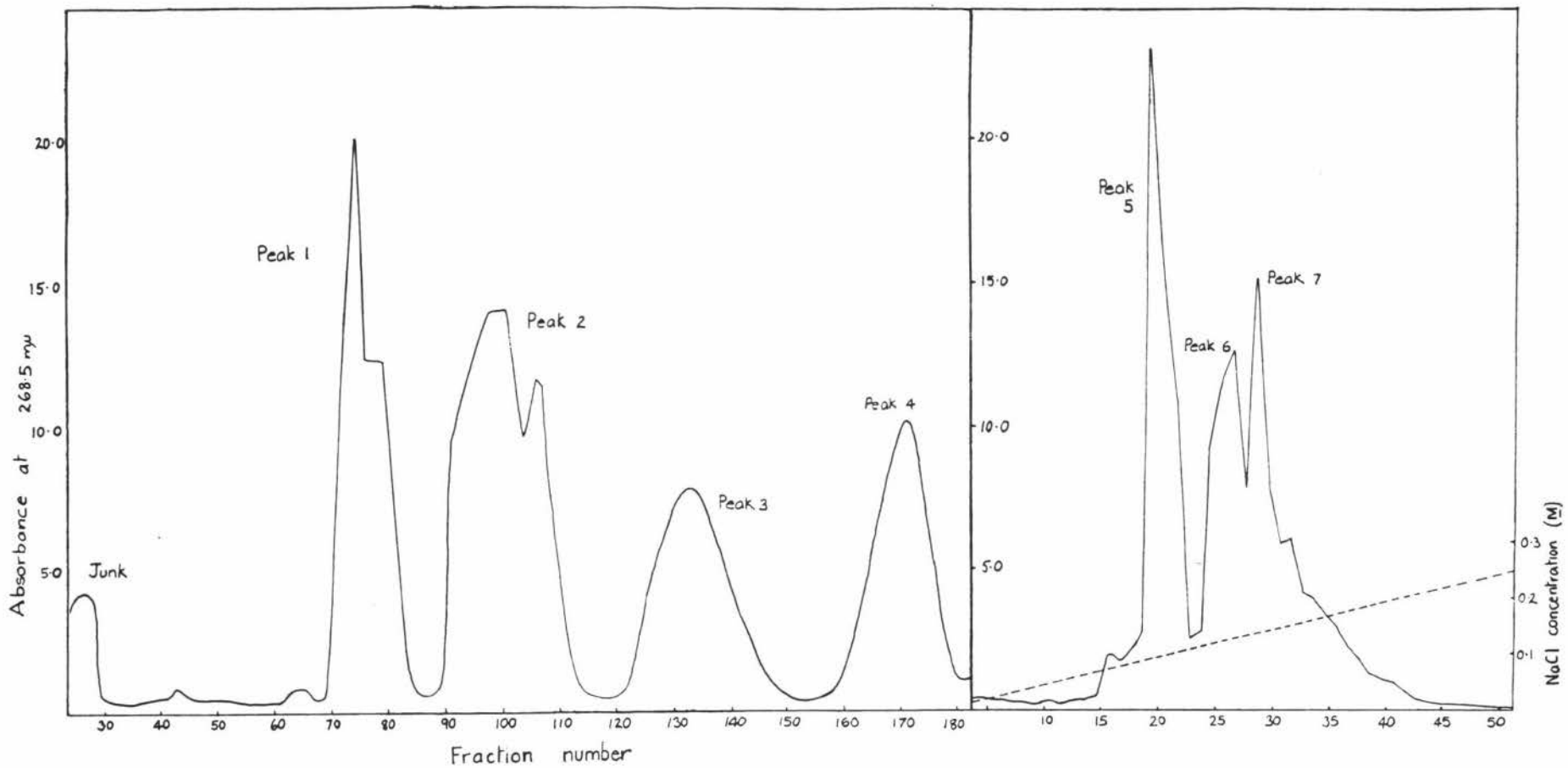


Fig 10.

Fractionation of pyrimidine isopliths from a diphenylamine digest of 1000 mg DNA on a DEAE cellulose column 50 cms x 3 cms in 7 M urea/0·1 M acetate/pH 5·4. Fraction size 20 mls. Salt gradient for first part of fractionation was exponential, 0 - 0·3 M NaCl and peaks 5 onwards were eluted as one large peak which was collected, desalted and separated again by linear salt gradient.

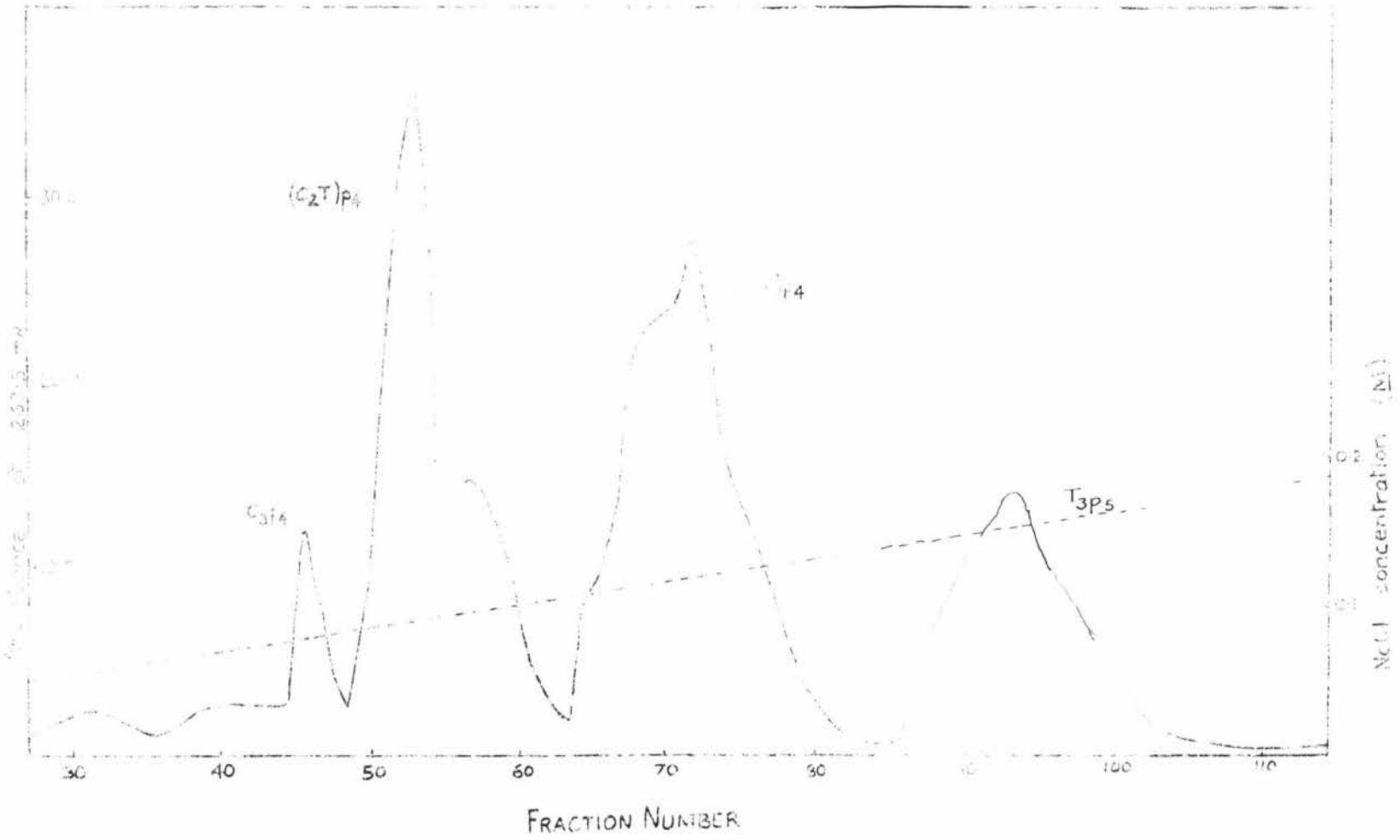


Fig 11

Fractionation of ... by gradient ... on ...

... fibers in 7 M urea/0·1 M formate
... Column size 21 cms x 0·9 cm
... size 1·7 mls.

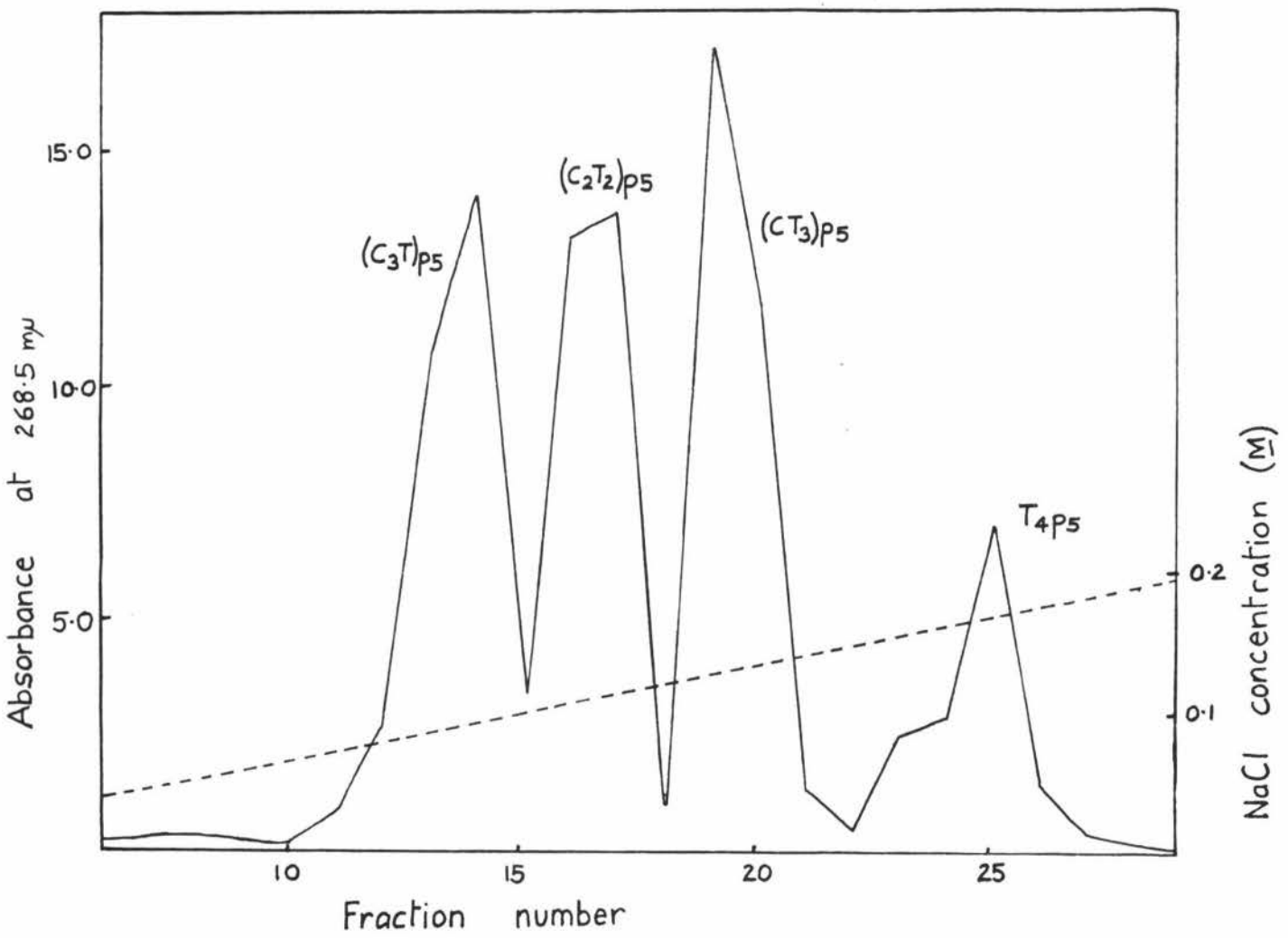


Fig 12

Fractionation of pyrimidine tetramers on DEAE cellulose in 7 M urea/0·1 M formate by a linear salt gradient 0 - 0·2 M NaCl. Column size 16 cms x 1½ cms. Fraction size 13 mls.

six peaks were expected, only two peaks were eluted, and it was obvious that some of the nucleotide material had been washed out previously with the salt and that the remainder was incompletely resolved. The remaining material was therefore desalted and fractionated on a standard DE 32 column in formate buffer, with the resulting elution pattern shown in figure 13. Both C_5p_6 and $(C_4T)p_6$ fractions were missing and yields of the remaining peaks were lower than expected.

II Measurement of recovery of pyrimidine oligonucleotides obtained from calf thymus DNA

All material that had been separated, desalted and dried down was taken up in a known amount of water and the optical density at wavelengths 260 mμ, 267 mμ and 280 mμ was measured, at pH 1. The results are shown in Table 1.

From the values for the extinction at 267 mμ, the number of millimoles of pyrimidine per 1000 mg DNA and per 100 gram atoms of DNA phosphorus were calculated for each oligonucleotide, and these values compared with those found experimentally by Petersen (27), and with the amount expected from a randomly arranged polymer of the same base composition.

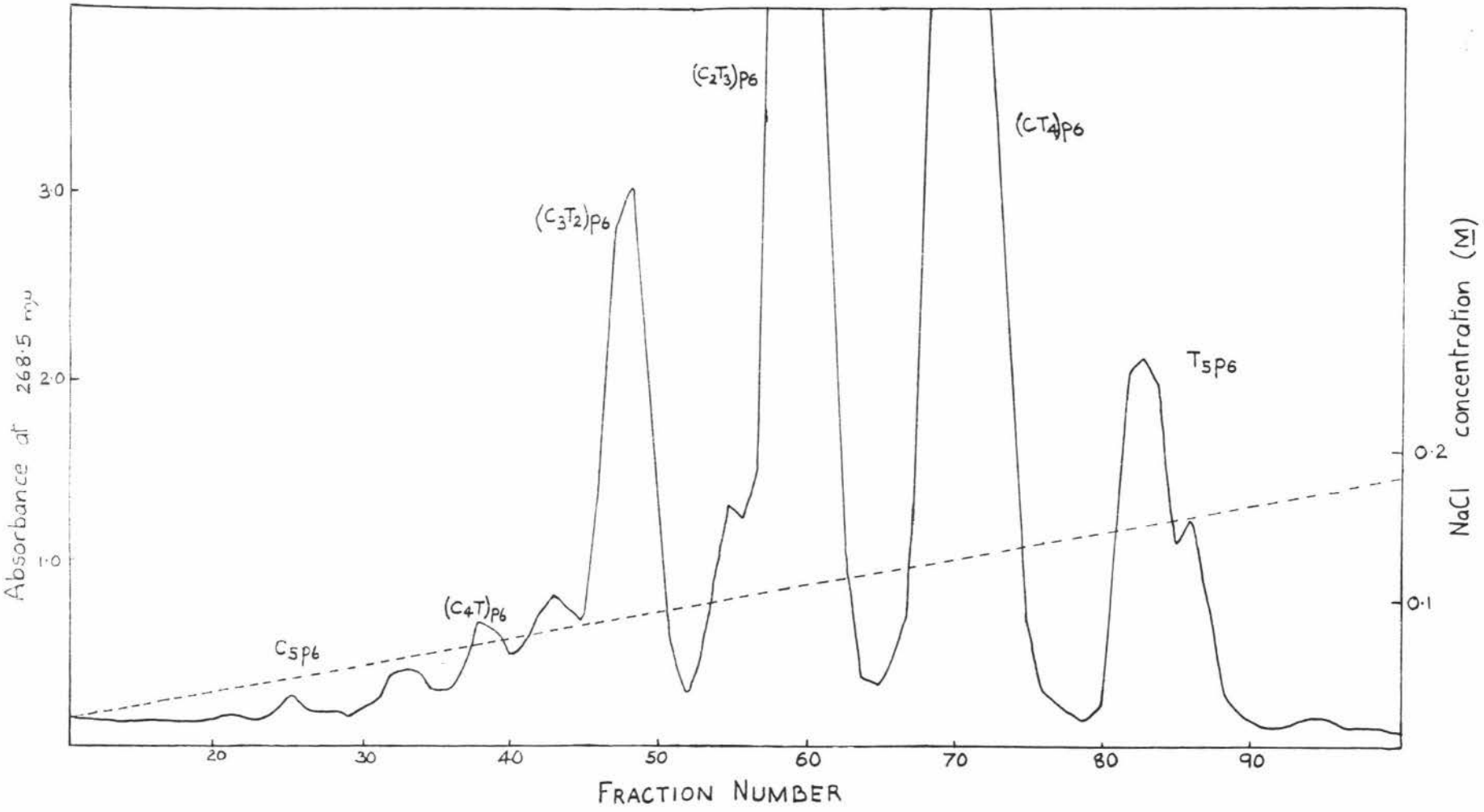


Fig 13.

Fractionation of pyrimidine oligonucleotides of length 5 on DEAE cellulose in 7 M urea/ 0·1 M formate by a linear salt gradient 0 - 0·2 M NaCl. Column size 24 cms x 1 cm. Flow rate 9 mls/hr. Fraction size 4·5 mls.

TABLE I

Oligonucleotide	E280/E260 pH 1	Calculated E280/E260
$(C_2T)p_4$	1.38	1.48
$(CT_2)p_4$	1.02	1.04
T_3p_4	0.68	0.7
$(C_3T)p_5$	1.52	1.58
$(C_2T_2)p_5$	1.18	1.24
$(CT_3)p_5$	0.93	0.95
$(C_3T_2)p_6$	1.17	1.36
$(C_2T_3)p_6$	1.07	1.11
$(CT_4)p_6$	0.92	0.96

As molar coefficient of optical density for pyrimidine at 267 mμ is 9900

atomic weight of phosphorus = 30.8

100 mg DNA contains 10 mg of phosphorus,

then, moles pyrimidine/100 gram atoms DNA phosphorus

$$= \frac{\text{O.D. units of pyrimidine/100 mg DNA} \times 1000 \times 30.8}{9900 \times 1000}$$

The results obtained are set out in Table II.

Fairly good agreement of these results was obtained with published values of the relative abundance of three pyrimidine trinucleotides and three tetranucleotides occurring in calf thymus DNA. The theoretical values shown were calculated according to Jones, Stacey and Watson (36).

III Separation of sequential isomers

Burton and Petersen (13) used a Dowex-2 chloride column and gradient elution with HCl for $(C_2T)p_4$ isomers to achieve a partial separation, ensuring that all material eluted was of the same pyrimidine composition by determining the $E\frac{280}{260}$ ratio of every fraction.

Separations were attempted in this investigation for the fractions $(CT_2)p_4$, $(C_2T)p_4$, $(CT_2)p_2$, $(C_2T)p_2$ and $(C_3T)p_3$, with an HCl gradient or in some cases with HCl/formate, HCl/urea, formate alone, and varying concentration ranges of HCl. Resins used were Dowex-1, and -2, and DE 32. About 10 to 20 optical density units of nucleotide material were applied to each column.

Oligo-nucleotide	Volume mls	E267 (at pH 1.0)	O.D. units	mmoles pyrimidine/ 1000 mg DNA (= mmoles pyrimidine/ 100 mg phosphate	moles pyrimidine/ 100 gm atoms DNA phosphorus	published value (27) (calf thymus DNA) for moles pyrimidine per 100 gm atoms DNA phosphorus	amount expected from a randomly arranged polymer of the same base composition
$(C_2T)p_4$	36	22.0	792	$\frac{792}{9900}$	$\frac{792 \times 30.8}{9900} = 2.46$	2.73	2.92
$(CT_2)p_4$	31	19.1	764	$\frac{764}{9900}$	2.38	2.64	3.86
T_3p_4	17	21.0	357	$\frac{357}{9900}$	1.11	1.28	1.70
$(C_3T)p_5$	26	17.3	454	$\frac{454}{9900}$	1.41	1.39	1.11
$(C_2T_2)p_5$	11	5.62	590	$\frac{590}{9900}$	1.84	1.99	2.20
$(CT_3)p_5$	11	4.02	440	$\frac{440}{9900}$	1.37	1.53	1.94
$(C_2T_3)p_6$	35	3.34	117	$\frac{117}{9900}$	0.37	1.27	1.30
$(CT_4)p_6$	18	8.65	158	$\frac{158}{9900}$	0.69	0.82	0.86
$(C_3T_2)p_6$	30	0.2	6	$\frac{6}{9900}$	most lost	1.06	0.98

No separation was achieved on the DE 32 columns for any eluant tried, and Dowex-2 gave poor separations which were not improved by the use of urea or formate. It is possible that a more gradual or a finely controlled HCl gradient would improve the separation that is shown in figure 14.

Determination of Sequential Isomers by Enzymic Methods

i) Assay of enzymes to be used

Snake venom phosphodiesterase. The assay procedure was as described in the methods section except that in many cases the quantities of reagents were scaled down by a factor of five. Five replicate tubes containing identical quantities of buffer, substrate and Mg^{++} ions in a volume of 0.2 ml were heated to 37°. Identical quantities of enzyme (5 µl) were added to each tube and at subsequent intervals 1 ml of NaOH was added to each tube to stop the reaction completely. The optical density of the contents of each tube was then determined at 400 mµ, and the rate of change of optical density determined graphically. An increase of 1.200 in absorbance under the defined conditions of the assay was taken to be equivalent to the hydrolysis of 0.1 µmole of substrate, and as one enzyme unit theoretically hydrolyses 0.1 µmole substrate per minutes, the units present in the enzyme sample were calculated to be 3.3 U/ml.

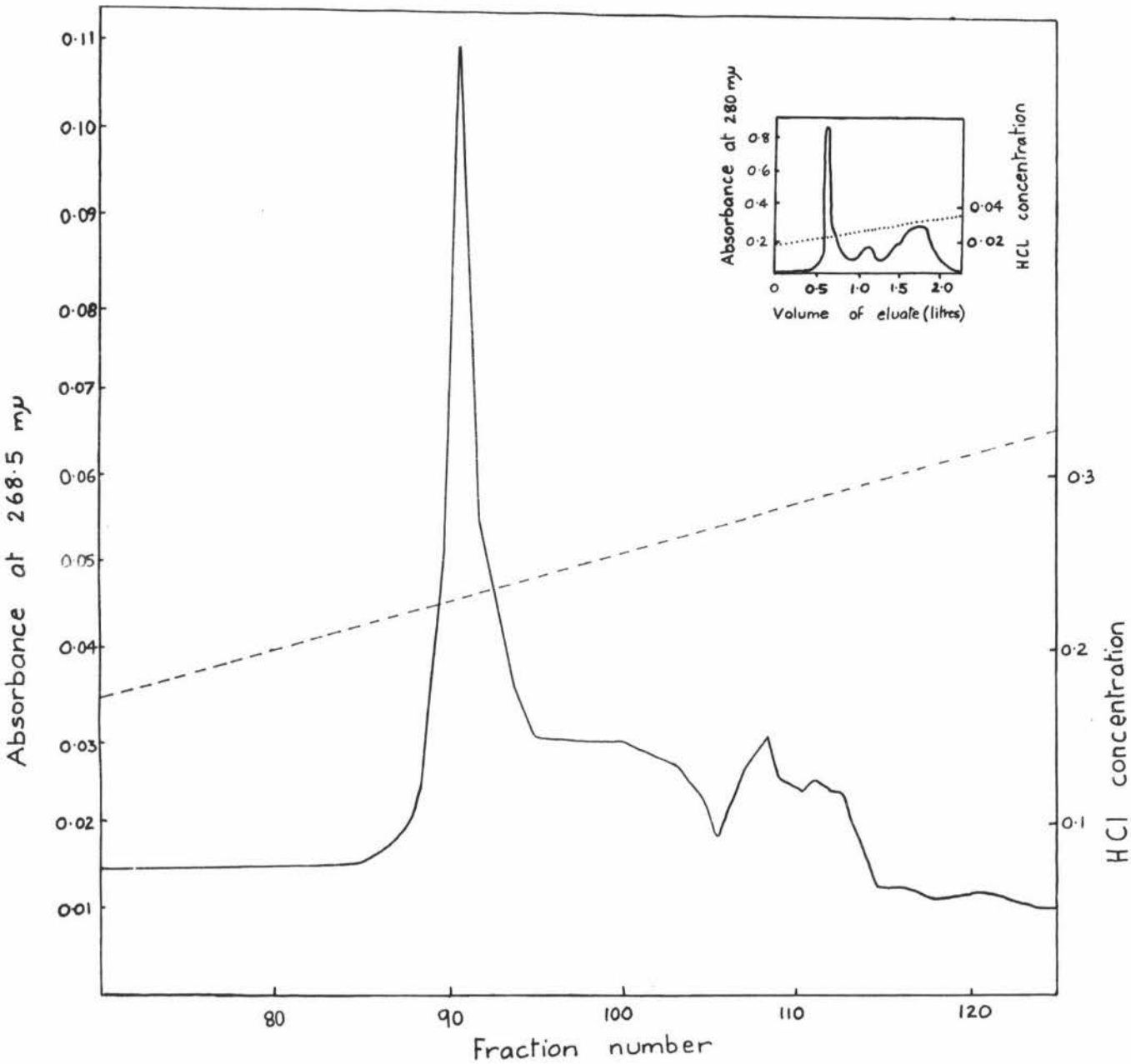


Fig 14.

Separation of $(C_2T)p_4$ isomers on Dowex-2-chloride by gradient elution with HCl. Column size 10cm x 1cm. (Inset: similar separation on Dowex-2-chloride by Burton and Petersen (20)).

For analytical purposes, approximately 0.03 units of enzyme was used to digest 10 optical density units of nucleotide material.

Test for monoesterase activity in the snake venom phosphodiesterase preparation.

This was done by measuring any release of inorganic phosphate, by contaminating monoesterase, from the substrate p-nitrophenyl thymidine-5'-phosphate. The following four systems were set up. All tubes were incubated for several hours at 37° and the inorganic phosphate content of duplicate samples from each tube was determined as described in the methods section. In addition a standard phosphate curve was produced, and the results shown in Table III were obtained.

TABLE III

Nature of system (set up in duplicate)	µmoles of inorganic phosphate per ml in tube after incubation (mean of four values)
Enzyme, substrate, buffer and Mg^{++} ions	0.032
Substrate, buffer and Mg^{++} ions	0.030
Enzyme, buffer and Mg^{++} ions	0.012
Buffer and Mg^{++} ions	0.006

It was therefore concluded that the enzyme preparation contains no monoesterase activity as no (unaccountable) inorganic phosphate was released.

Alkaline phosphatase

As supplied, the enzyme solution has a concentration of 10.5 mg/ml and 35 U/mg enzyme. The heat treatment described in the "materials" section resulted in 3 mls of solution, most of which was stored in small lots in the deep freeze.

The assay was performed as described in the methods section. The rate of increase in optical density of the incubation mixture was 0.063 units per minute. A check on the quantity of protein per ml enzyme solution was made by reading the optical density at 278 mμ, which was 2.12.

TABLE IIIa

Units/mg protein	Protein/ml enzyme solution
$= \dfrac{\Delta A/\text{min} \times 1000}{1.63 \times 10^4 \times \dfrac{\text{mg enzyme}}{\text{ml reaction mixture}}}$	$= E_{278} \times 1.43$
	$= 2.12 \times 1.43$ mg/ml
= 23U/mg protein	= 3.0 mg/ml

ii) Dephosphorylation of oligonucleotides

Contaminating enzyme action in the alkaline phosphatase preparation was apparent when nucleotide material was incubated with excessive quantities of the enzyme, or for very long incubation times, or when Mg^{++} ions were added to the incubation mixture. Therefore it was necessary to determine the minimum quantity of enzyme which was sufficient to dephosphorylate about 10 optical density units of oligonucleotide in a convenient incubation time of a few hours. Ten O.D. units represents the maximum amount of oligonucleotide from which digestion products can be easily separated by paper electrophoresis and chromatography. For this quantity (approx. 1 µmole), 10 µl of the alkaline phosphatase preparation was used (3.04 mg/ml; 23 U/mg protein), and this appeared to achieve complete dephosphorylation after about an hour of incubation at 37° and pH 8.0. When this system was left to incubate for 12 hours, quite noticeable amounts of nucleosides could be detected in the digestion mixture due to contamination in the enzyme preparation.

After incubation, the enzyme was removed from the reaction mixture by first heating at 100° for 10 minutes. After cooling, the solution was shaken with an equal volume of chloroform:amyl alcohol (20:1) which completely denatured the enzyme. Centrifuging caused the precipitated protein to lie at the interphase between chloroform and aqueous

phase; the latter could be removed with a Pasteur pipette to a clean tube.

iii) Digestion with snake venom phosphodiesterase

5 µl of the stored enzyme solution (3 U/ml) was found to be adequate for digesting 1 µmole of dephosphorylated oligonucleotide when incubated together at 37° for about 5 hours, in Tris buffer which was pH 8.9 and 0.01 M with respect to Mg^{++} ions. The Mg ions were essential for enzyme action, but were a serious handicap to subsequent paper electrophoresis, when any movement of charged mononucleotides or mononucleosides was prevented by the presence of this salt. Charged oligonucleotides were often not affected in this manner; this might reflect the nature of the Mg^{++} effect.

The enzyme was inactivated after incubation by the procedure used for alkaline phosphatase.

iv) Digestion with spleen phosphodiesterase

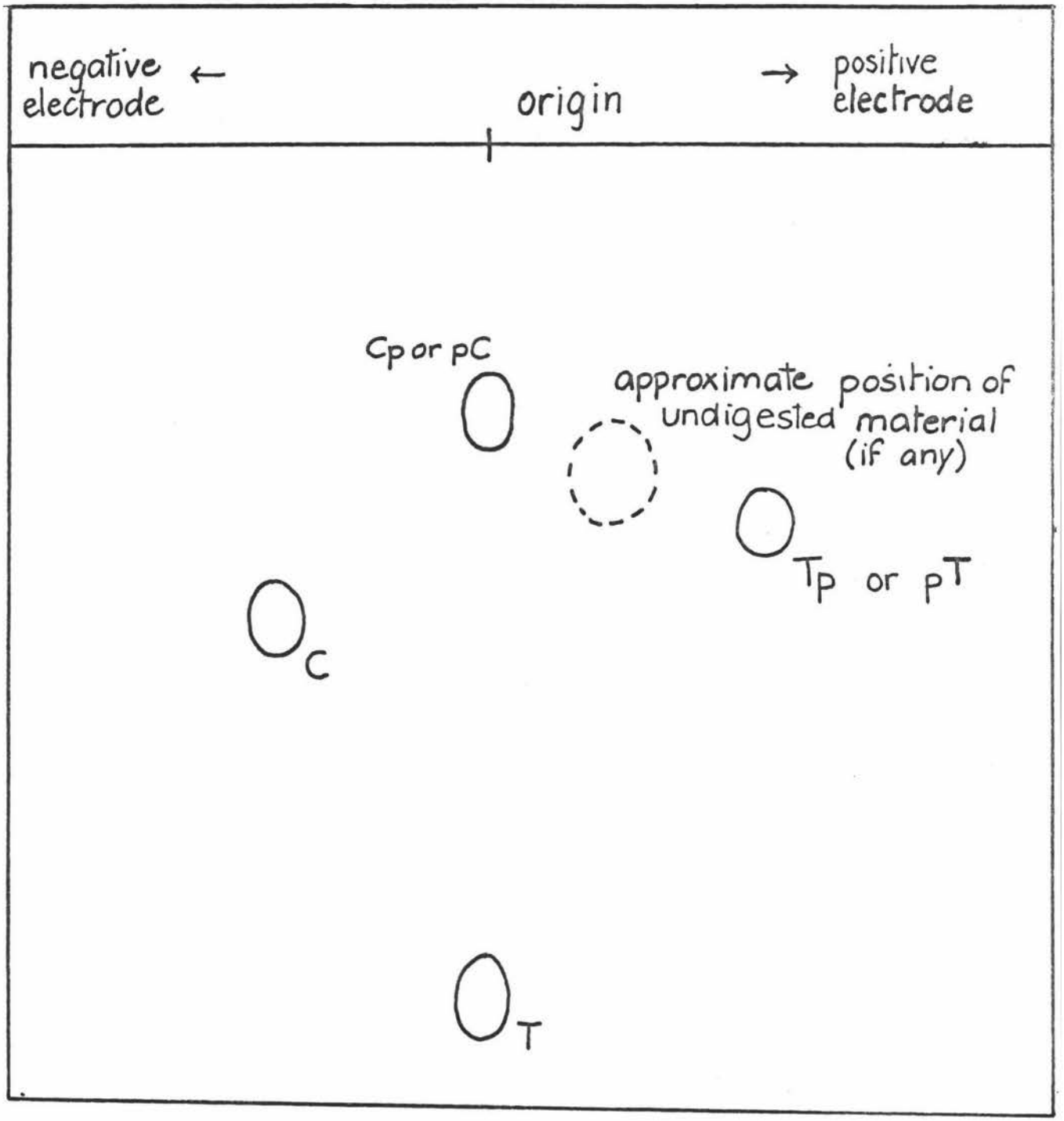
A solution of approximately 15 U/ml of this enzyme was stored in small lots in the deep freeze, and 5 µl of this was found sufficient to digest 1 µmole of nucleotide material in 2 or 3 hours. A small amount of sodium tartrate was added to all digestion mixtures to suppress possible monoesterase activity (generally 5 µl of 0.05 M

tartrate/μmole of substrate). Somewhat higher concentrations of this salt were found also to inhibit the diesterase activity to a considerable extent.

The enzyme was denatured and removed from the solution in the usual way.

v) Separation of digestion products

Oligonucleotides were treated with alkaline phosphatase, deproteinised, and digested with snake venom phosphodiesterase and spleen phosphodiesterase separately, the digests being deproteinised again and dried down in a stream of air at 37°. The spleen enzyme digest was spotted onto Whatman No.1 chromatography paper and the digest products separated by electrophoresis in formate buffer pH 2.7 at 7 Kvolts for 10 minutes. Electrophoresis separated the digested material into three components, for at this pH thymidylic acid has a nett negative charge and cytidine a nett positive charge. Thymidine and deoxycytidylic acid have no charge and remain together at the origin. The dried papers were then developed in the second dimension in n-butanol:acetone:acetic acid: 5% NH_3 solution: H_2O (9:3:2:2:4), (Solvent I), by descending chromatography for 20 to 24 hours at room temperature. This separated the digestion products from each other and from any residual undigested oligonucleotide. Figure 15 shows the relative



Result of 2-way separation of digestion products from pyrimidine oligonucleotides. (i) electrophoresis in ammonium formate buffer pH 2·7 (ii) paper chromatography (downwards) in n-butanol: acetone: acetic acid : 5% NH_3 : water (9 : 3 : 2 : 2 : 4).

positions of the pyrimidine nucleotides and nucleosides after two-dimensional separation under the system just described.

Snake venom diesterase digest mixtures could not be separated into components by electrophoresis because of the Mg^{++} ion content, even after the addition of a small quantity of EDTA into the sample. One-dimensional paper chromatography separations in Solvent I were therefore used. Although good separations in one dimension were achieved with a mixture of commercially obtained nucleosides and nucleotides (C, Cp, T and Tp), the presence of Mg and buffer ions in the actual digest mixture made resolution much less satisfactory.

vi) Determination of the quantity of sequential isomers of composition $(CT_2)p_4$.

Spleen diesterase digests of $(CT_2)p_4$ were separated by electrophoresis and snake venom diesterase digests were separated by descending paper chromatography in one dimension by Solvent I. The positions of the UV-absorbing area were identified by making contact prints, and the spots were cut out and eluted as described in the methods section.

(a) Results for Spleen diesterase

TABLE IV

Nucleotide or nucleoside	Optical Density 260 mμ	267 mμ	280 mμ	μmoles of material
C (spot that migrates to the negative terminal)	0.210			
		0.251		0.075
			0.287	0.069
				Mean = 0.072 μmoles
Cp + T (spot that remains at the origin)		0.755		0.23 μmoles
		As there should be equal quantities of Cp and T from a digestion of $(CT_2)p_2$, μmoles of Cp and T occurring in this digest will be taken as $\frac{0.23}{2}$ = 0.115 μmoles		
Tp (spot that migrates to the positive terminal)	0.540			
		0.570		0.17
			0.360	0.17
				Mean = 0.17 μmoles

Total material = 0.47 μmole

Therefore if the three sequential isomers are labelled thus:

C T T ..... (a)

T C T ..... (b)

T T C ..... (c)

the proportion of (c) in the mixture of isomers

$$= \frac{\text{μmoles C}}{\text{μmoles C + μmoles T}}$$

$$= \frac{0.072}{0.072 + 0.115}$$

$$= 0.39 \text{ (or } 39\%)$$

(b) Duplicate run

TABLE V

Nucleoside or nucleotide	Optical Density 260 mμ	267 mμ	280 mμ	μmoles of material
C	0.206			
		0.203		
			0.240	0.069 0.057 Mean = 0.063 μmole
Cp + T		0.830		0.249 Therefore there taken to be 0.125 μmoles of Cp and 0.125 μmoles of T.
Tp	0.690			
		0.730		
			0.460	0.219 0.219
				Mean = 0.219 μmole

Total material present in the digest is therefore 0.43 μmoles. The proportion of (c) in the mixture is therefore 0.34 or 34%. Therefore the mean value for the proportion of (c) in the fraction, from the two determinations is 0.36 or 36%.

From results of snake venom diesterase digestion of this trinucleotide, the ratio of the cytidine content to the cytidylic acid is all that is required to obtain the ratio

of the three isomers in the mixture.

Results obtained were:

TABLE VI

Material	E_{267}	µmoles material
cytidine (C)	0.235	0.071
cytidylic acid (Cp)	0.282	0.085

The total cytidine and cytidylic acid content is therefore

0.156 µmoles

The percentage of this material occurring in isomer (c) must be 36% (as determined previously), or

0.056 µmoles

Therefore 0.100 µmoles occurs in isomers (a) and (b).

But, 0.071 µmoles was released as cytidine (C) which can only have come from isomer (a).

Therefore 0.029 µmoles of pC came from isomer (b).

Consequently the ratio a ; b ; c ≡ CTT : TCT : TTC

must be 0.071 : 0.029 : 0.056

which is approximately **3 : 1 : 2**

vii <u>Thin layer chromatography</u>

Thin layer chromatography techniques were investigated with the aim of finding a more rapid method of separating the products of enzymic digestion than could

be obtained with paper chromatography techniques so far employed. A large variety of solvents was tried, two dimensional chromatography was attempted, and different adsorbents were investigated. The following is a brief summary of the findings:

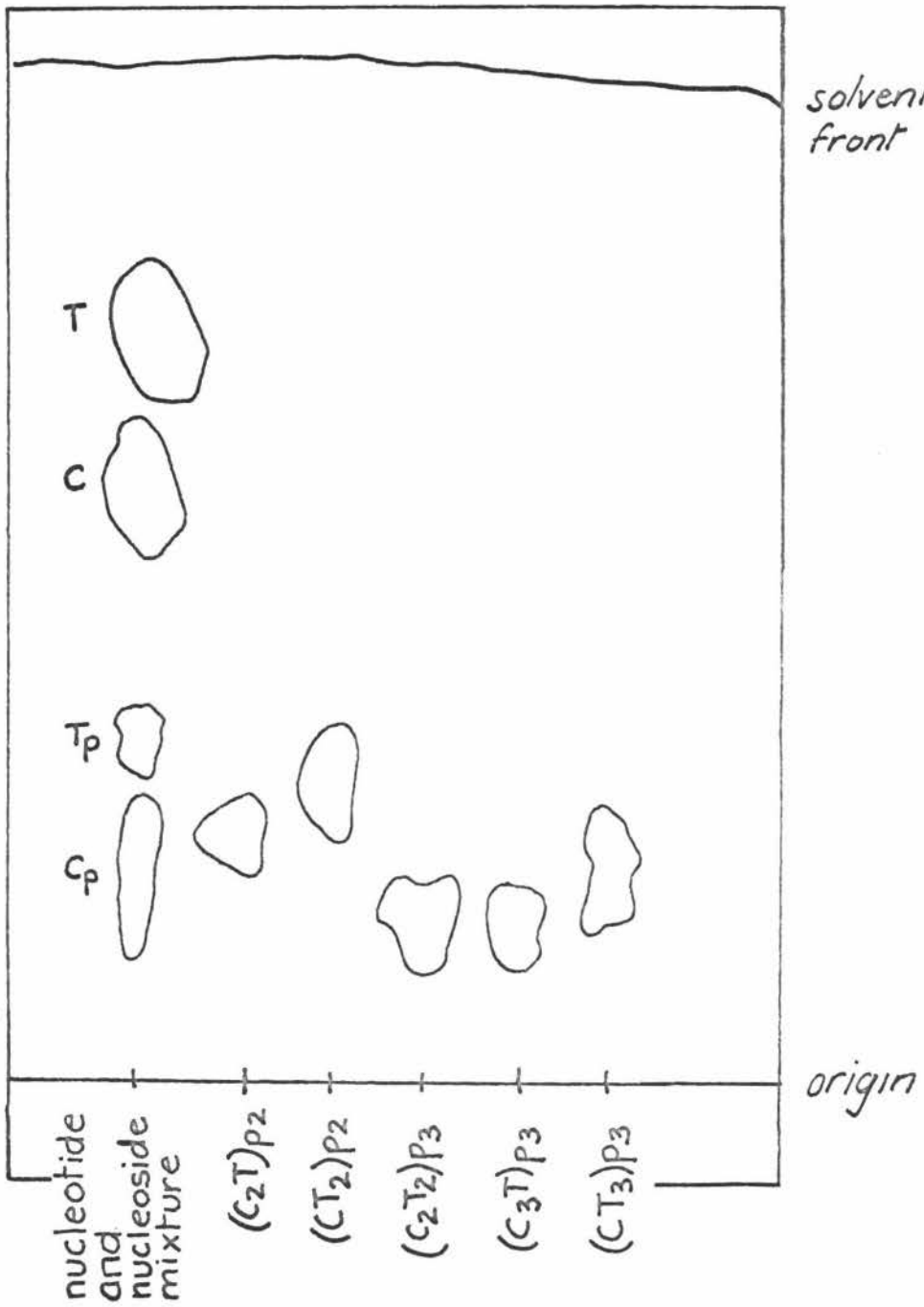
i) The use of cadmium borate in the cellulose adsorbant layer enables all nucleotide material to show up very distinctly, under UV light, as deep brown spots against a light ground, and the position of this material on the layer can easily be marked with a pencil. Layer development was found to take from half to five hours, so that the entire process of separation of digest products, from application of material to elution of spots, in fact took a fraction of the time needed for the same operations involving paper chromatography.

ii) Acid solvents such as Wyatt's (HCl:isopropanol:H_2O; 170:41:39) and solvent systems containing formic acid could not be used when the adsorbent contains cadmium borate indicator, for there was interaction between the indicator and the acid, resulting in strong UV absorption over the entire layer. Plain cellulose appeared not to react with the acid, but UV absorbing nucleotide material could not be seen at all distinctly against a background without

indicator. Acid solvents are particularly useful in the separation of nucleotides; non-acidic solvents will separate nucleosides from each other and from nucleotides. An example of a separation in a non-acidic solvent is shown in figure 16.

iii) Two dimensional thin layer development can be aided if extraneous material, carried by the solvent running in the first dimension, is scraped from the plate together with ½ inch of adsorbent material from this edge of the layer. This procedure helps to keep the second solvent front regular. However, in this study, no suitable two-dimensional system was found employing any combination of solvent, as in every case the applied material become too diffused after development to make detection easy.

The most successful non-acidic solvent used was n-butanol:acetone:acetic acid:5% NH_3:H_2O (9:3:2:2:4), saturated with EDTA. This successfully separated all four components of a snake venom or spleen diesterase digest when an artificial mixture was used, but as already mentioned actual digest mixtures were poorly separated because of buffer and Mg ions present. Another very effective solvent of note was NH_3:H_2O:isopropanol:methanol (2:3:6:7) which very adequately separated nucleosides from nucleotides after



Tracing of a thin layer chromatogram of the separation of cytidine, cytidylic acid, thymidine and thymidylic acid by the solvent n-butanol : acetone : acetic acid : 5% NH_3 : water (9:3:2:2:4) on cellulose powder. The chromatogram also shows relative rate of movement of the five dephosphorylated tri- and tetranucleotides.

only 6 to 8 cms of development (fig 17).

In nearly all cases, small oligonucleotides were found to have the same Rf values as mononucleotides and for this reason, undigested material could not be easily removed by thin layer chromatography in non-acid solvents. This is illustrated in figure 17.

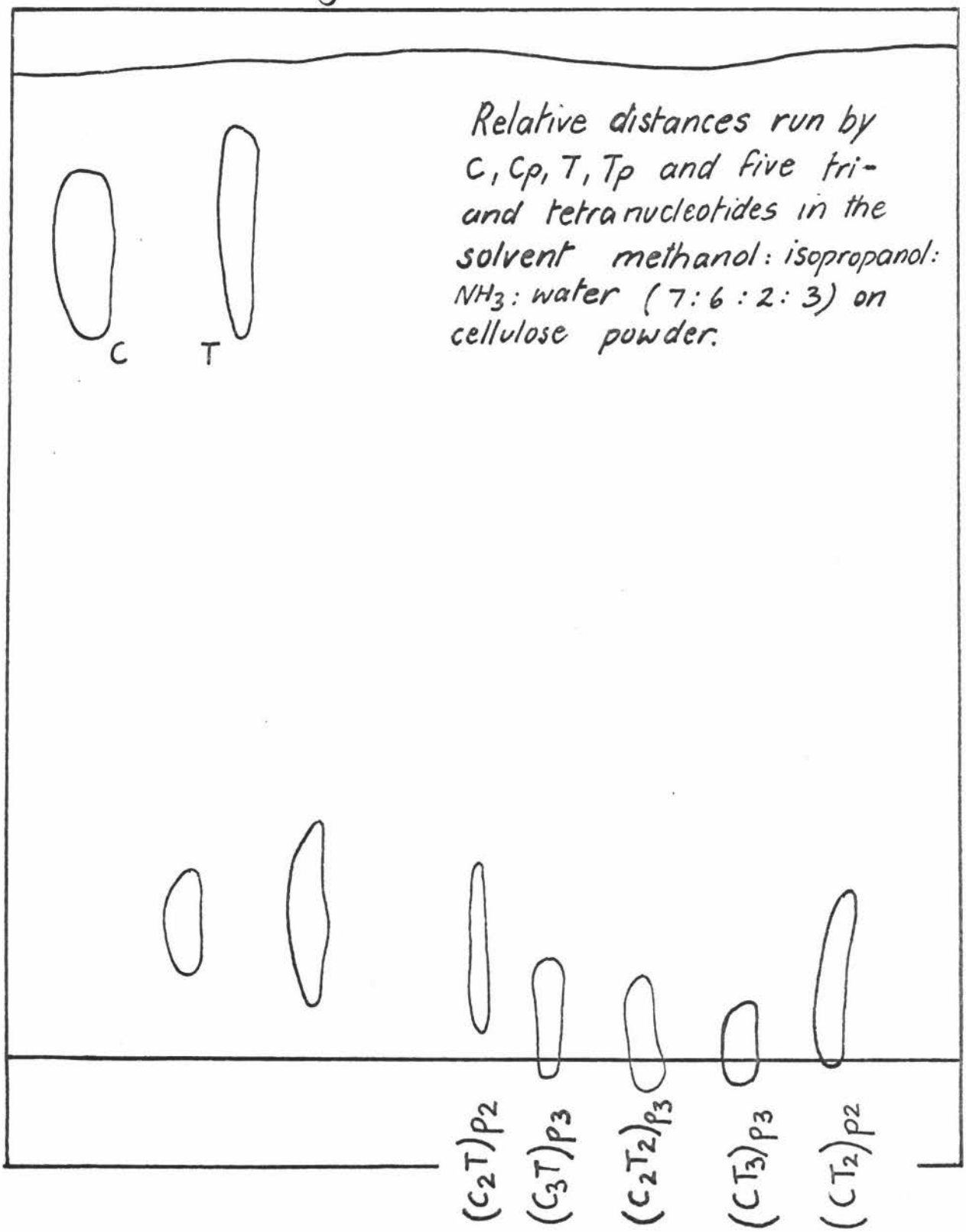
V) Experiments with CME-carbodiimide

i) Preliminary experiments with thymidylic acid

Experiments investigating the reaction of this reagent with thymidylic acid (Tp) showed that a very high concentration of reagent with long incubation times were necessary if substantial complexing or reaction between the reactants were to be achieved. The pH of the reaction mixture also appeared to be critical. Figure 18 shows the results of an experiment illustrating these observations, where the effect of pH and incubation time on the progress of the reaction can be seen clearly.

For this experiment, the following systems had been set up in triplicate, and had been incubated at 37°.

Fig 17.


Relative distances run by C, Cp, T, Tp and five tri- and tetra nucleotides in the solvent methanol: isopropanol: NH_3: water (7:6:2:3) on cellulose powder.
C
T
$(C_2T)_{P2}$
$(C_3T)_{P3}$
$(C_2T_2)_{P3}$
$(CT_3)_{P3}$
$(CT_2)_{P2}$

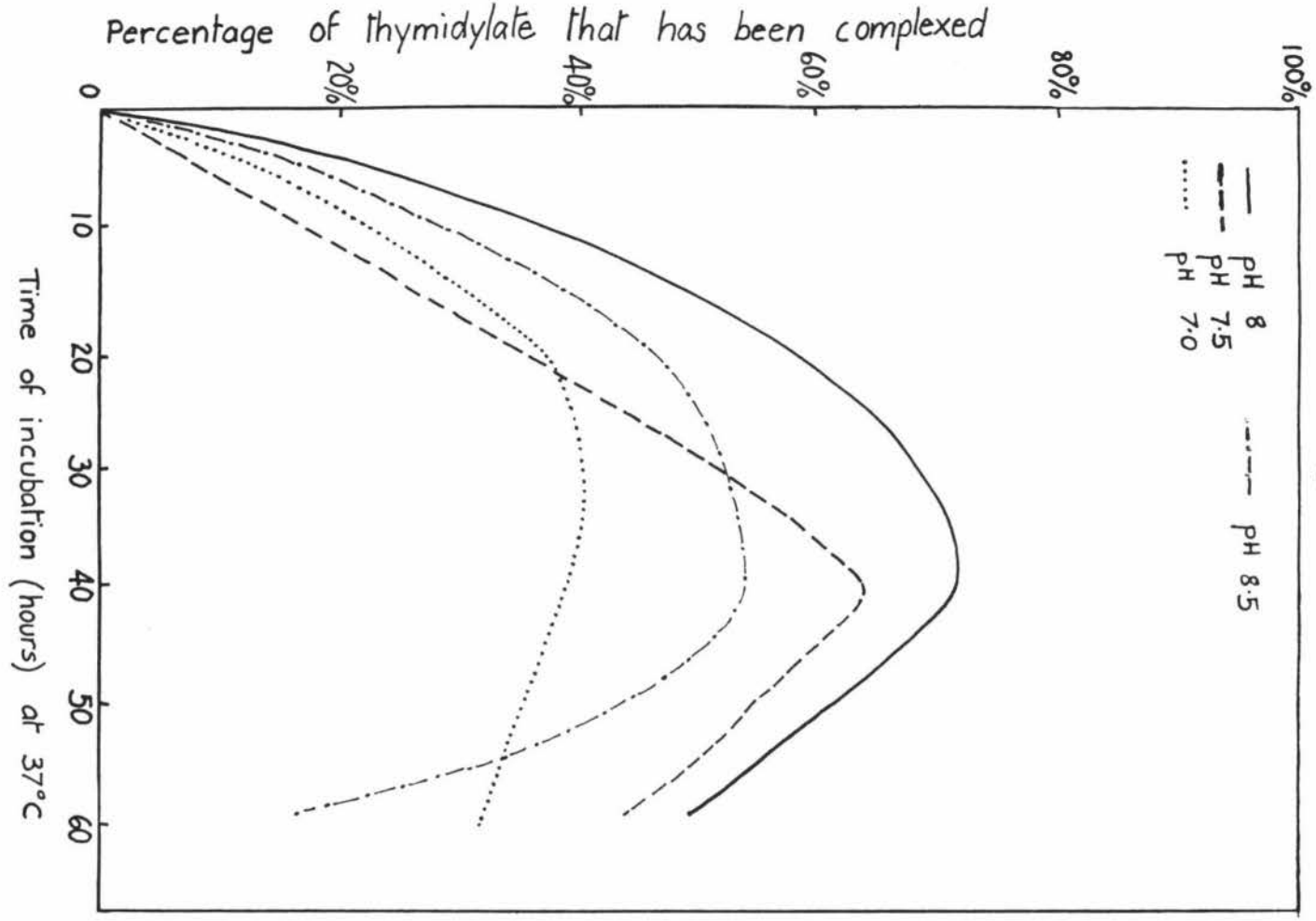


Fig 18.

Reaction of thymidylate (Tp) with N-cyclohexyl-N'-β-(4-methyl-morpholinium) ethyl carbodiimide (CMCD) as a function of time and pH. Reaction mixture contained 50 µl each of Tp solution (20 mg/ml), CMCD (100 mg/ml) and 1M tris of appropiate pH.

TABLE VII

Thymidylic acid solution 20 mg/ml	CMC-carbodiimide solution 100 mg/ml	1 M Tris buffer pH 7.0	7.5	8.0	8.5
50 µl	50 µl	50µl	-	-	-
50 µl	50 µl	-	50 µl	-	-
50 µl	50 µl	-	-	50 µl	-
50 µl	50 µl	-	-	-	50 µl

The contents of one tube at each pH were dried down and spotted onto chromatography paper after 20, 40 and 59 hours, making 12 samples in all. All papers were developed in isopropanol:water (7:3) at room temperature for 15-18 hours, and UV absorbing areas detected and eluted in the normal manner. From contact prints it was obvious that the solvent had separated the three components (unreacted CMCD and Tp, and the product CMCD-Tp). The coefficient of optical density for CMCD-Tp was calculated by summing the coefficients of CMCD and Tp (without allowance for hypochromicity). The molar optical density coefficient of CMCD, (at 267 mµ and neutral pH), was found, by measuring the optical density of a solution of the reagent of known concentration, to be approximately 150. As the molar optical density coefficient of thymidylic acid at this wavelength (and acid pH) is

9900, the extinction coefficient of the complex CMCD-Tp could be taken as being the same for this purpose. The quantities of unreacted and reacted thymidylic acid for each incubation were therefore easily calculated, and a figure obtained for the extent to which the reaction had taken place for each sample. It should be noted that CMCD was always present in excess in every incubation.

A spectrum of each of the three components was obtained on a recording spectrophotometer, and a comparison appears in figure 19. There is a noticeable shift in the absorbance curve of thymidylic acid after the attachment of a CMCD molecule, but it will be seen that this curve bears no trace of the very characteristic spectrum of the reagent molecule itself.

In this experiment, the reaction appears to have proceeded to its greatest extent after 30 to 40 hours, this activity later decreasing. Decomposition, probably of the CMCD, could be a contributing factor in the decline in formation of the complex CMCD-Tp. This decline is certainly more noticeable at the most alkaline pH observed, pH 8.5.

The concentration of CMCD in the reaction mixture in this experiment was approx. 30 mg/ml. Earlier experiments

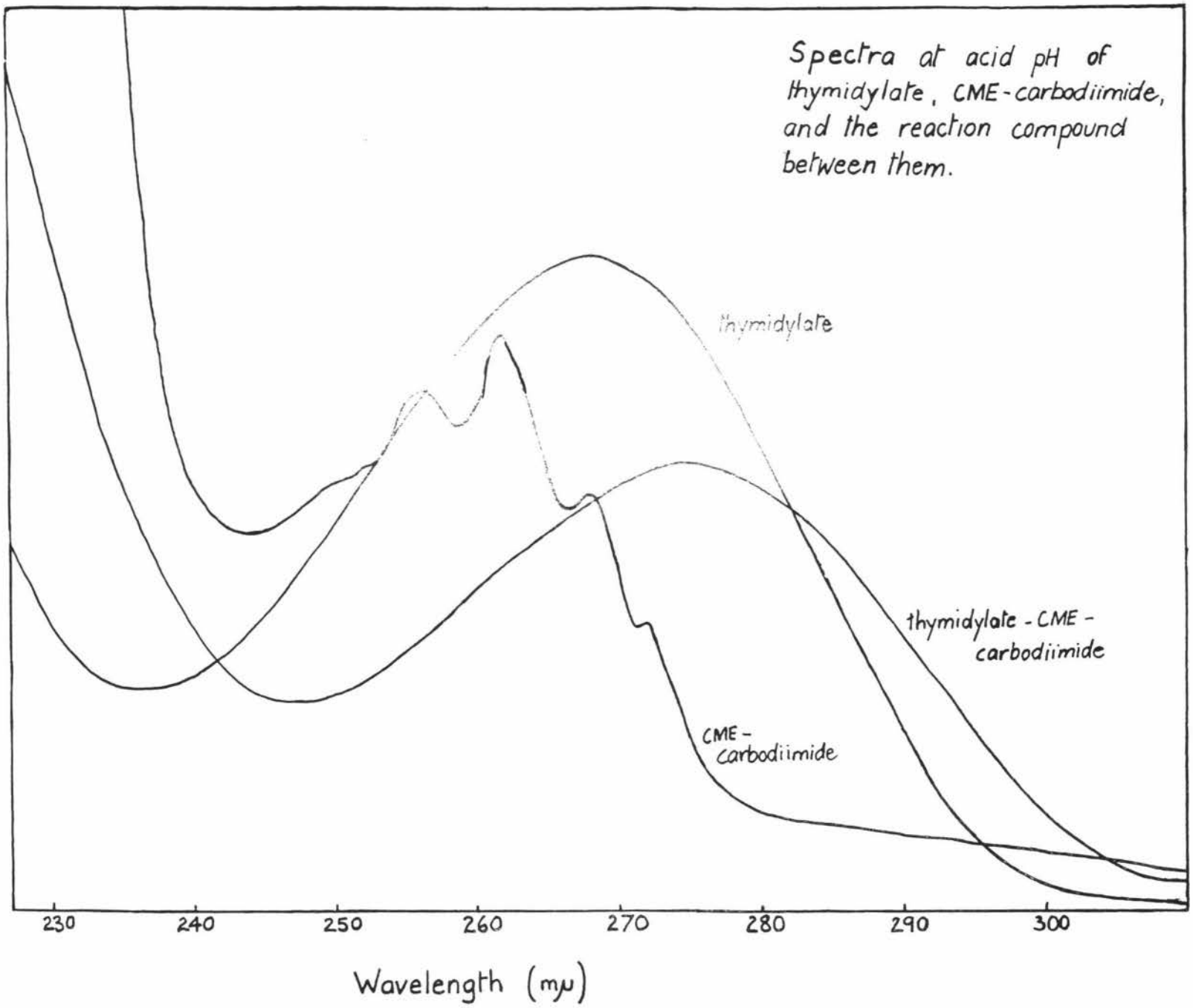

Spectra at acid pH of thymidylate, CME-carbodiimide, and the reaction compound between them.
thymidylate
thymidylate - CME - carbodiimide
CME - carbodiimide
Absorbance
230
240
250
260
270
280
290
300
Wavelength (mμ)

Fig 19.

47a.

had showed that at 100 mg/ml almost complete reaction would occur after about 70 hours incubation of the naturally viscous solution. Apart from viscocity, other undesirable effects of such high concentrations would be the susceptibility of the solution to dephosphorylation, evaporation or bacterial contamination.

The effect of Mg^{++} ions on the reaction was also briefly investigated; they were found to decrease the rate of reaction. At a concentration of 0.015 M Mg^{++}, the reaction rate that had been previously observed in the absence of magnesium was reduced by a third.

Discussion

Although this work has yielded very little in the way of original results, many useful points have been established concerning techniques chosen to investigate the problem, and many pitfalls have been recognised and noted.

The diphenylamine reaction has been chosen as the starting tool in this sequence investigation because it is possible to demonstrate quite conclusively its freedom from disrupting side reactions. During the reaction inorganic phosphate is released (as shown in figure 3), until a certain level is attained, always characteristic of the DNA species undergoing digestion, and evincing the specificity of the action of the amine. Any interfering side reactions would be expected to slowly yield P_i. Other experimental evidence available includes the fact that phosphomonoesterase will release inorganic phosphate from Py_nP_{n+1} products, again until a new stationary P_i level is reached, and this is such that the amount of residual organic orthophosphate is the same as the amount of inorganic phosphate released in the original degradation. This equivalence is to be expected since the organic phosphate remaining in the pyrimidine sequences after treatment with phosphatase should be the phosphodiester groups joining pyrimidine deoxynucleosides

and since in the double-stranded calf thymus DNA used in these experiments the number of purines equals the number of pyrimidines, the number of interpurine phosphate groups should equal the number of interpyrimidine phosphate groups. In addition, values obtained for several types of DNA for the amount of interpurine phosphate agree quite closely with the frequencies of purine-purine nearest neighbour frequencies as determined by the DNA polymerase system (51).

As a result of much recent study in techniques of column separation of nucleotide material, there have been successful methods documented for the resolution of smaller pyrimidine sequences, even though systems available for separation of oligonucleotides generally are unsatisfactory. DE 32 cellulose resin has been found to give consistently good results for the separation of lower members of the Py_nP_{n+1} series in these studies, although batches obtained from the manufacturers do differ and difficulty has been experienced in the resolution of longer sequences. Urea has been used to assist almost every separation attempted in this study. Tomlinson and Tener (32) have shown that the secondary binding forces (hydrogen bonding) between the exchangers and the nucleotide bases may be greatly reduced by including 7 M urea in the solution of the eluting salt. The separations then depend more strongly on the nett charge involved. Formamide and glycols have a similar effect and

are also widely used.

The problem of separation of longer sequences has been approached from many angles recently. One solution may lie in the use of far more complex exchangers in column work such as polynucleotide materials with which oligonucleotides could complex to varying degrees, so that a separation is effected. The longer products of the diphenylamine degradation of DNA have been successfully separated on microgranular cellulose columns. Sequences up to 16 nucleotides in length have been obtained from calf thymus DNA, and one elution peak corresponding to chain length 19 has been described for the DNA from the coliphage f1. This is the longest sequence of pyrimidines yet found in any DNA (35).

Very little separation of sequential isomers has been achieved in attempts described in this study, but the results indicate that it is certainly possible that a more exhaustive investigation of resins and eluant could produce satisfactory results for the trinucleotides $(C_2T)p_4$, $(C_2T)p_2$, $(CT_2)p_4$ and $(CT_2)p_2$. Separations of mixtures of sequential isomers of certain dinucleotides have been achieved by very long development of paper chromatograms. The theory behind the separations is not well understood, but it is possible that they occur because the base at the 5'-end of the isomers is more

accessible for interaction with the solvent or the chromatographic support.

The proportions of thymidine and deoxycytidine in the various oligonucleotides were obtained by comparing their ratio of extinction at 260 and 280 mμ with values calculated theoretically for various mixtures of integral molar proportions of thymidine and deoxycytidine. Thymidine and its nucleotides have a value of 0.7 for this ratio of extinction, at an acid pH, whereas deoxycytidine and its nucleotides have values of 2.0. As can be seen from the results, excellent agreement is obtained between calculated and experimental results for all compositional isomers separated and measured. This shows that adequate separations have been achieved.

Experiments have been described where this method of determining composition of unhydrolysed single compositional isomers can be extended, by comparing the absorption spectrum of any oligomer with calculated curves. The spectral curve is measured in 7 M urea at three pH values and digitised at 1 mμ intervals (52). Calculated curves consisting of the sums of the absorption spectra of mononucleotides in 7 M urea at the same pH's are compared with those experimentally found by means of a computer, to establish the best fit.

Quantitative measurements can be made both by UV absorption measurements or by the determination of total phosphorus, but the former method is easier and more reproducible and has therefore been used in this study. Absorbances are measured at pH 1.5 and 267 mμ, where one isosbestic point occurs in the spectra of thymidine and deoxycytidine. As spectra are pH dependent, so too are isosbestic points.

One of the most common drawbacks in the use of enzymes for specific tasks is the inclusion in the preparation of extraneous contaminating enzymes whose activities interfere with those of the enzyme under observation. Snake venom phosphodiesterase and spleen phosphodiesterase were shown to have little if any marked contamination, but alkaline phosphatase as already noted in the results was found to have a considerable Mg^{++}-inducible phosphodiesterase activity which greatly interfered with the investigation procedure. This activity was apparent on long incubation even without the presence of added Mg^{++} ions. However, if a sufficiently high concentration of the latter were included in the reaction mixture, up to 50% of the oligonucleotides tested were found to be digested to mononucleosides instead of having been simply dephosphorylated.

The two-dimensional separation system of paper electrophoresis followed by paper chromatography was found to be excellent for separation of products after spleen phosphodiesterase digestion, but quite useless when Mg^{++} ions had been introduced into the system. These, however, were necessary for digestion by snake venom phosphodiesterase. The best solution finally used for the separation of products from digestion by the latter was one-dimensional paper chromatography as described in the results section, but even this was not completely satisfactory as occasional streaking was experienced. Also, undigested material could not be adequately separated from products of digestion in a one-dimensional system.

It is now considered that two-dimensional paper chromatography would almost certainly have been more successful than any other method of separation of the products of snake venom phosphodiesterase digestion attempted in this study. However, this technique is very lengthy and there is always an increased risk of damage to the wet paper during manipulation and drying compared to other paper chormatographic methods.

The thin layer method of separation initially appeared promising, but a completely satisfactory solvent was never

found, and undigested material could not be separated from the slow moving nucleotides. The main advantages found with thin layer methods were sharpness of resolution (if the layer was not overloaded at the origin), great sensitivity, simplicity and speed. Spots on thin layers could rapidly be detected in UV light of very low intensity, and circled with a pencil. By applying the pencil very firmly, a ring of adsorbent encircling the UV absorbing material could be removed, and by applying a drop of water to the isolated spots and freezing the entire layer, the frozen samples could be lifted from the plate and placed in a test tube for elution with water or acid.

The results finally obtained for the ratio of isomers of (CT_2) are only a rough indication of the actual ratio, and this work should have been repeated many times over before results could be considered meaningful. The reliability of results obtained by ~~this~~ such exonuclease digestions will eventually be very important when other systems of sequence determination are being perfected and some means of checking the accuracy of the latter methods is required.

Results with the soluble carbodiimide, CMCD, were very disappointing and it is again obvious that work of a

more extensive nature than was attempted here is essential for this problem. Something must be learned of the mode of reaction between thymidine and CMCD before the reagent can be used under reasonable conditions for specific attachment to nucleotides, in sequence work. At present the very high reagent concentration found necessary for complete or almost complete reaction make this method very impractical.

Several other methods of varying degrees of usefulness have been proposed for the determination of sequential isomers. The estimation of isomeric components of an oligonucleotide fraction will be made easier by a method allowing the specific removal of one of the two base constituents, when dealing with fractions of the type Py_nP_{n+1} or Pu_nP_{n+1}. In addition this method may be of interest for the consideration of the potential content of a DNA molecule, in conjunction with a diphenylamine depurination, since it permits the estimation not only of these homologous pyrimidine oligonucleotide sequences that are positioned between purines, but also of those flanked by another pyrimidine.

In Chargaff's method (53) all cytosine nucleotides are removed from each isomer, leaving behind intact thymidine sequences as they were in the original isomers. Since

each isomer represents a different arrangement of thymidine residues it should be possible by examining the products to deduce the concentrations of several of the isomers. With the help of data derived from exonuclease studies on this same series of isomers, it is possible to solve many if not all of the possible sequential isomers occurring. Cytosine residues can be removed from either the original isomers, or from the latter after their terminal phosphate groups have been removed, and the difference in the two results may itself provide useful information.

This method is particularly useful for resolving the ratios of isomers containing a large proportion of thymidine.

Nucleases that cannot hydrolyse dinucleotides are potentially very useful in sequence determination, and exonucleases particularly form this category. One of the most useful is a DNA-specific phosphodiesterase from E.coli, exonuclease I. Hydrolysis begins at the 3'-hydroxyl end of the chain, but the enzyme is not able to cleave free dinucleotides or the 5'-terminal dinucleotide portion of a polydeoxynucleotide chain. Use of this enzyme is however limited at present because it has not been obtained in pure enough form for quantitative work (45).

Further chemical modification of oligonucleotides may be possible. For example, O-methylamine reacts selectively with cytosine at pH 6, forming uridine-O-methyl oxime (46), which will block the action of snake venom diesterase when this addition compound has been incorporated into oligonucleotides containing cytosine.

Chemical degradation of DNA such as that occurring on reaction with dephenylamine will never produce overlapping fragments of nucleic acid and therefore the elucidation of the exact nature of the pyrimidine sequences can only solve a small part of the main problem. The process will for example have to be repeated for the purine sequences that are characteristic of the species of DNA in question, and in addition, a totally different experimental approach attempted with a view to combining the two lots of data. One of the most ingenious approaches involves the labelling of various bases so that they can be seen at their positions along the DNA strand under the electron microscope (47, 48). Selective labelling reagents are chosen for their ability to react with or incorporate heavy atoms such as uranium. It is hoped that eventually this method will provide a means of building up fairly detailed maps of arrangements of each of the bases in DNA, which can then be compared with the results from sequence studies.

Summary

The digestion products from the action of acidic diphenylamine on calf thymus deoxyribonucleic acid have been separated as far as possible by known physical techniques, and these products characterised and measured quantitatively.

Some possible methods of deducing the relative proportions of each sequential isomer in a naturally occurring mixture of short compositionally identical isomers have been investigated, with special emphasis on the use of exonucleases. In addition attempts have been made to modify exonuclease specificity by attaching a chemical blocking agent to its substrate.

References

1 International Union of Pure and Applied Chemistry (1961) Information Bulletin No.13. Butterworth Scientific Publications, London.

1(a) Crick F.H.C., Cold Spring Harbour Symp. Quant.Biol. 31, (1966) 3.

2 Nirenberg M. et al, Cold Spring Harbour Symp. Quant. Biol. 31 (1966) 11.

3 Watson J.D. and Crick F.H.C., Cold Spring Harbour Symp. Quant. Biol. 18 (1953) 123.

4 Sinsheimer R.L., Progress in Nucleic Acid Research and Molecular Biology 8 (1968) 115.

5 Holley R.W. et al., Science 145 (1965) 1462.

6 Brownlee G.G., Sanger F. and Barrell B.G., J. Mol. Biol. 34 (1968) 379.

7 Burton K. and Petersen G.B., Biochim.Biophys. Acta 26 (1957) 667.

8 Shapiro H.S. and Chargaff E., Biochim.Biophys. Acta 26 (1957) 596 and 608.

9 Jones A.S., Ross G.W. Takemura S., Thompson T.W. and Walker R.T., J.Chem. Soc. (1964) 373.

10 Levene P.A. and Bass L.W., J.Biol. Chem 71 (1926) 167.

11 Temperli A. et al., Biochim.Biophys. Acta 91 (1964) 462.

12 Burton K. and Riley W.T., Biochem J. 98 (1966) 78.

13 Burton K. and Petersen G.B., Biochem. J. 75 (1960) 17.

14 Burton K., Lunt M.R., Petersen G.B. and Siebke, J.C., Cold Spring Harbour Symp. Quant.Biol. 28 (1961) 27.

15 Laskowski M., Advances in Enzymology 29 (1967) 165.

16 Petersen G.B. and Burton K., Biochem. J. 92 (1964) 666.

17 Razzell W.E. and Khorana H.G., J.Biol.Chem. 234 (1959)2114.

18 Boyer B.D., Lardy H. and Myrback K., Eds, "The Enzymes" Vol. 5, Academic Press, New York, 1961.

19 Kay E.R.M., Simmons N.S. and Dounse A.L., J. Am. Chem. Soc. 74 (1952) 1724.

20 Chargaff E., in "The Nucleic Acids" Vol. 1 (1955) p.307.

21 Razzell W.E. and Khorana H.G., J.Biol. Chem. 234 (1959) 2105.

22 Zinder N.D., Valentine R.C., Roger M. and Stoeckenius W., Virology 20 (1963) 638.

23 Garen A. and Levinthal C., Biochim. Biophys. Acta 38 (1960) 470.

24 Biochim.Biophys. Acta 72 (1963) 133.

25 Stahl E., Ed., "Thin Layer Chromatography", Academic Press, London (1965).

26 Markham R. and Smith J.D., Biochem. J. 45 (1949) 294.

27 Petersen G.B., Biochem. J. 87 (1963) 495.

28 Vischer E. and Chargaff E., J. Biol. Chem. 176 (1948) 703.

29 Parr C.W., Biochem. J. 56 (1954) xxvii.

30 Chen P.S., Toribara J.Y. and Warner H., Anal. Chem. 28 (1956) 1756.

31 Burton K. and Petersen G.B., Biochem. J. 75 (1960) 17.

32 Tomlinson R.V. and Tener G.M., Biochemistry 2 (1962) 703.

33 Petersen G.B. and Reeves J.M., Biochim. Biophys. Acta 129 (1966) 438.

34 Mazin A.L. and Vanyushun B.F., Biochimiya 32 (1967) 377.

35 Petersen G.B. and Reeves J.M., Biochim. Biophys. Acta 179 (1969) 510.

36 Jones A.S., Stacey M. and Watson B.E., J. Chem. Soc. (1957) 2454.

38 Naylor R., Ho N.W.Y. and Gilham P.T., J. Amer. Chem. Soc. 87 (1965) 4209.

39 Lee J.C. et al., Biochim. Biophys. Acta 95 (1965) 503.

40 RajBhandary V.L. and Stuart A., Ann. Rev. Biochem. 35 (1966) 759.

41 Drevitch V.F. et al, Biochim. Biophys. Acta 123 (1966) 207.

42 Augusti-Tocco G. and Brow G.C., Nature 206 (1965) 683.

43 Gilham P.T., J. Amer. Chem. Soc. 84 (1962) 687.

44 Crane R.K. and Lipmann F., J. Biol. Chem. 201 (1953) 235.

45 Lehman I.R. and Nusbaum A.L., J.Biol. Chem. 239 (1964) 2628.

46 Kochetkov, Budowsky and Shibaeva, Biochim. Biophys. Acta 68 (1963) 493.

47 Beer M. and Moudrianakis E.N., Proc. Nat. Acad. Sic. 48 (1962) 409.

48 Moudrianakis E.N., and Beer M., Biochim. Biophys. Acta 95 (1965) 23.

49 Dulbecco R. and Vogt M., Proc. Nat. Acad. Sci. 50 (1963) 236.

50 Rushizky and Sober, Prog. Nucl. Acid Res. 8 (1968) 171.

51 Josse J., Kaiser A.D. and Kornberg A., J. Biol. Chem. 236 (1961) 864.

52 Mandeles S. and Cantor C., Biopolymers 4 (1966) 759.

53 Shapiro H.S. and Chargaff E., Biochem. 5 (1966) 3012.

54 Holley R.W., Prog. Nucl. Acid Res. 8 (1968) 37.

List of Figures

List of Tables

I am indebted to many people for their assistance in many aspects of the experimental work and ideas behind this thesis. To my supervisor Professor G. B. Petersen for guidance and great encouragement, to Dr G. W. Butler for making available to me laboratory facilities at the Plant Chemistry Division of the Department of Scientific and Industrial Research, to Dr K. Heyes, Mrs P. Penny, Mr W. Sutton and especially Dr J. W. Lyttleton for their helpful advice, I wish to express my very grateful thanks.