# NUCLEOTIDE SEQUENCES IN DEFINED REGIONS OF VIRAL DNA

by

Robert William Old

A thesis presented for the degree of
Doctor of Philosophy of the University of Edinburgh

Department of Molecular Biology
University of Edinburgh
Scotland

August, 1973

# PREFACE

This thesis describes work done in the Department of Molecular Biology during the tenure of a Medical Research Council Scholarship between October 1970 and September 1973. All the experimental work described in the following pages is my own except where indicated. I have in certain places cited collaborative experiments performed with Drs. K. Murray, N.E. Murray, P. Englund and P. Weigel.

ACKNOWLEDGEMENTS

It is a pleasure to acknowledge my indebtedness to my supervisor, Dr K. Murray for patient advice and encouragement. I am also grateful to Drs. P. Englund, P. Weigel and G. Roizes for fruitful collaboration. I thank T. Bruce and S.G. Hughes for growing batches of bacterial cells. I thank the Medical Research Council for a scholarship for training in research methods and Professors W. Hayes and M.R. Pollock for a place in their department. Finally I thank those who provided manuscripts and results in advance of publication, and colleagues and friends in the laboratory for helpful and enjoyable discussion.

## NOTATION OF NUCLEOTIDES

Since this thesis is concerned mainly with deoxynucleotides, the usual prefix 'd' has generally been omitted. In some instances, however, to avoid confusion, the prefix 'd' has been included. Ribonucleotides bear the prefix 'r', except for riboadenosine-5′-triphosphate which is abbreviated ATP.

## SUMMARY

A number of highly specific endonucleases act at a very limited number of sites in DNA, and belong to an important class of proteins that are able to recognise specific targets in a DNA molecule. The Ter endonuclease, which generates the cohesive ends of mature bacteriophage lambda DNA, and restriction endonucleases are examples of such specific endonucleases. Also, certain restriction endonucleases provide a means of degrading DNA into fragments of defined genetic content that are useful in the analysis and manipulation of the primary structure of DNA. This thesis is concerned with the application of radioactive terminal labelling of DNA molecules to the determination of nucleotide sequences in viral DNA at the sites of action of the Ter endonuclease, and the restriction endonucleases R.$Eco_{P1}$ and R.$Hin_{d}$ III.

Investigation of a nucleotide phosphotransferase from E. coli, as a possible means of labelling DNA at the 3′-terminus, showed that whereas the enzyme catalyses phosphate group transfer to the 3′-hydroxyl group of ribo- and deoxyribo-dinucleoside monophosphates, higher oligodeoxynucleotides are not phosphorylated. Native DNA and RNA are also not phosphorylated.

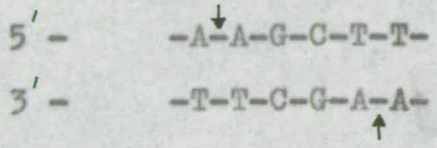
Analyses of radioactive oligonucleotides in digests of bacteriophage lambda DNA that had been labelled at or

near its 3′-termini in exonuclease-repair reactions catalysed by T4 DNA polymerase revealed the sequences -G-T-T-A in the l-stand and -A-C-C-C-G-C-G in the r-stand. These sequences, together with others known in the region of the termini provide a total of 25 known base pairs in the vicinity of the termini. When the cohesive ends are paired, the sequence between the nicks has an axis of 2-fold rotational symmetry. Five of the first eight base pairs on either side of the axis are rotationally symmetric.

The 5′-terminal labelling method has been applied to unmodified lambda DNA fragments produced by treatment with R.Eco$_{P1}$. The extent of labelling was small and no new 5′-terminal sequences were detected, suggesting that the 5′-termini created by R.Eco$_{P1}$ do not participate in the polynucleotide kinase reaction. R.Eco$_{P1}$ does not have a strict requirement for the presence of exogenous S-adenosyl methionine as cofactor. Lambda DNA fragments produced by treatment with R.Eco$_{P1}$ were resolved as bands upon gel electrophoresis.

The sequence of DNA base pairs at the site of action of R.Hin$_d$ III has been determined. Oligonucleotides present in digests of lambda DNA fragments that had been 5′-terminally labelled in a reaction catalysed by polynucleotide kinase were analysed. The following sequence of nucleotides in the vicinity of the break made in lambda DNA by R.Hin$_d$ III was deduced from these data, from

analyses of oligonucleotides present in digests of DNA fragments that had been 3′-terminally labelled in a reaction catalysed by T4 DNA polymerase and from electron microscopic observations.




The endonuclease R.Hin$_d$ III generates short cohesive termini. The most striking feature of the sequence is its symmetry.

Endonuclease R.Hin$_d$ III makes six double-strand breaks in SV40 DNA. The SV40 DNA fragments have been resolved by gel electrophoresis and their lengths have been determined by electron microscopy. The DNA of bacteriophage T7 grown on several E. coli strains is resistant to R.Hin$_d$ III.

These results are discussed in relation to the mechanism by which the endonucleases recognise their targets within DNA molecules.

# INDEX

# Chapter 1

# INTRODUCTION

## 1.1 Nucleotide sequences in DNA

Brown and Todd, in 1952, adopted a general structure for nucleic acids in which the nucleoside units are linked by 3',5'-phosphodiester bonds. Since then there has been rapid progress in methods for the determination of nucleotide sequences in nucleic acid, so that now the determination of the sequence of entire RNA chains from small coliphage is feasible (about 3300 nucleotides). Much has depended upon the development of micromethods for the fractionation and analysis of $^{32}P$-labelled oligonucleotides (Sanger et al., 1965). Many of these methods are applicable to sequence determination in DNA. In many respects, however, RNA sequence determination has advanced more rapidly than that of DNA, for a number of reasons. But several recent developments in DNA enzymology, in particular the exploitation of endonuclease IV (Ziff et al., 1973; Robertson et al., 1973) and certain restriction enzymes are providing a great stimulus to direct sequence determination in DNA, as is the development of mapping procedures enabling the sequences of quite large oligonucleotides (up to twenty residues), belonging to a single homologous set, to be 'read off' directly (Sanger et al., 1973; ~~Morrison~~ Murray and Murray, 1973).

There are a number of recent reviews of sequencing

methodology. Those of Brownlee (1971), de Wachter and Fiers (1971), Szekely (1971), Barrell (1971) and Randerath and Randerath (1967) provide an exhaustive collation of available methods. A review by Murray and Old (1973) deals with some recent developments in DNA sequence determination illustrating a variety of experimental approaches.

The ultimate interest in nucleic acid sequence analysis must lie in our understanding of biological problems. Regions of a DNA molecule involved in a specific interaction, or having a known function such as the specification of an identifiable product, or mediating control of transcription or replication of DNA, command attention. The recognition of particular targets, within a relatively much larger DNA molecule, by proteins is a component of many important biological processes. As originally suggested by Vogel (1964) (and subsequently discussed by many others - in particular see Britten and Davidson (1969) and Crick (1971)) the great majority of DNA in higher organisms does not appear to code for proteins but is used for control purposes.

Comparison of nucleotide sequences at sites interacting with a particular protein can define those features of the sequence that are essential for recognition. In some cases it is also feasible to define the recognition sequence from a study of mutants with a functionally altered nucleic acid site. The

primary structure of the site may indicate the likely secondary structure of the site, or may fall into one of several classes of sequence which share a common recognition mechanism. This unilateral structural approach to sequence recognition has been strikingly successful in the study of ribosome binding sites in the RNA of phages Qβ and R17 (Goodman et al., 1969; Cory et al., 1970), a site in T7 RNA synthesised in vitro (Arrand and Hindley, 1973), and the site of the single-stranded DNA of ØX174 (Robertson et al., 1973).

Recognition of sequences in DNA by proteins

The mechanism by which certain proteins are able to recognise specific sequences in double-stranded DNA is a central question in molecular biology. There have been many models put forward for DNA structures that might occur at recognition sites.

Sequences having an axis of rotational symmetry may form, by self-complementarity within strands, a cross-shaped structure that is recognised as a singularity in specific interaction with a protein (Gierer, 1966). The unpaired regions at the top of the loops may be particularly important. Such structural singularities have been discussed extensively by Sobell (1973a; 1973b). Crick (1971) has suggested that recognition of base sequence takes place in rather extensive single-stranded regions of DNA. His argument is a general one; for

most proteins whose tertiary structure is known, the active site lies in a shallow groove or cavity and not a protruding piece of protein structure. Thus it may not be easy for a protein of reasonable size to recognise more than a limited number of base-pairs.

The direct effects of base sequence upon the local conformation of the DNA duplex or upon the contour of electron density in the grooves of the helix are factors of which little is known, but it may be in this way that a sequence of bases is recognised. The different conformations of duplex DNA suggested by Bram and Tougard (1972), the various types of helix proposed for a range of polynucleotides giving characteristically different X-ray diffraction patterns and the effect of base sequence on the conformation of poly d (I, C) (Wells et al., 1970) all encourage the view that base sequence per se may be recognised directly in duplex DNA.

What direct evidence is there for any of these suggestions? Can sequences be recognised directly in DNA, and if so how extensive can they be? In order to answer such questions it is essential that initially we learn more about sequences in double-stranded DNA that are recognised specifically by proteins. Highly specific endonucleases, such as restriction endonucleases and the Ter function responsible for the production of the cohesive ends of phage lambda DNA, offer a favourable opportunity to investigate this problem. Also,

restriction endonucleases are becoming an invaluable tool in the study of DNA; firstly as a means of fragmenting DNA in a precise way to obtain fragments of defined genetic content or location within the genome, and secondly as a probe for the detection in DNA of the particular sequence, perhaps reiterated many times, recognised by the enzyme.

Sequence determination at the sites recognised by the Ter endonuclease and certain restriction endonucleases forms a central part of the work described in this thesis. An account of these endonucleases, and methods that may be used in studying the sites they recognise is given in the following sections.

## 1.2 Cohesive ends of bacteriophage lambda DNA

### Nature of the cohesive ends

The DNA of phage lambda is a duplex molecule of about 46,500 base pairs, with single-stranded projections of twelve nucleotides at each 5′-terminus (Wu and Taylor, 1971). The ends of the molecule are cohesive because the single-stranded regions are mutually complementary in sequence, and by their specific base-pairing the DNA may form circles or linear aggregates *in vitro* (Hershey *et al.*, 1963).

Experiments by Strack and Kaiser (1965) established the nature of the cohesive ends and showed that repair

reactions with E. coli DNA polymerase I and radioactively labelled deoxynucleoside triphosphates could be used to study the composition and sequence of this region of the molecule. Wu (Wu, 1970; Wu and Taylor, 1971) used the repair method with $^{3}H$ and $\alpha$-$^{32}P$-labelled triphosphates. The sequence complementary to each cohesive end was then deduced from analyses of oligonucleotides present in digests of the products of both complete and partial (with one or more dNTP omitted) repair reactions. Spleen and venom exonuclease digestion was used for analysis of oligonucleotides and identification of the nucleoside ($^{3}H$-labelled) released (where necessary after treatment of the oligonucleotide with phosphatase) gave the terminal nucleotides. Mononucleotides were counted for both $^{3}H$ and $^{32}P$ so that nearest neighbour (Josse et al., 1961) results were also obtained from spleen exonuclease digests. The sequences of the cohesive ends of lambda DNA (see Chapter 5, Fig. 2a) show the anticipated complementarity of bases.

The cohesive end sequences of several lambdoid phage DNAs have been examined directly by Murray and Murray (1973). Terminal labelling at the 5′-terminus of the DNA in the polynucleotide kinase reaction (Richardson, 1965; Novogrodsky et al., 1966) was followed by digestion with pancreatic DNase under very mild conditions, and analysis of the radioactive oligonucleotides by mapping methods. The cohesive ends of DNA from phages 82, 21,

424 and phi 80 have been examined in this way and have sequences identical to those of lambda.

Murray and Murray have also examined directly cohesive end sequences of the DNA of the P2-like family of phages. The cohesive end sequences of these are larger than those of lambda: nineteen nucleotides instead of twelve (Padmanabhan and Wu, 1972). Some differences were found among the three phages 186, P2 and 299 (Murray and Murray, 1973).

Bacteriophage lambda DNA replication and Ter

The mechanism by which the cohesive ends of phage lambda DNA are formed is not completely understood.

It is well known that shortly after infection, the cohesive ends join and the two single-chain interruptions in the resulting molecule are subsequently sealed by ligase (Gellert, 1967) so that within five minutes after infection most of the DNA molecules are converted into covalently closed circles (Young and Sinsheimer, 1964; Bode and Kaiser, 1965). Synthesis of lambda DNA begins about five minutes after infection, and continues until the end of the latent period. Replication of lambda DNA appears to take place in two ways.

Until about fifteen minutes after infection synthesis is confined to DNA in a closed circular form (Carter et al., 1969). Density labelling permits phage DNA in the first round of replication to be separated

from host DNA. Tomizawa and Ogawa (1968) and Schnos and Inman (1970) have isolated DNA having a buoyant density intermediate between light and hybrid and have examined the DNA in the electron microscope. Two-thirds of the structures isolated in this way are circular and most of the circles are topologically like the Greek letter theta (Figure 1.1). These structures have important properties. The length a + b equals the characteristic length of a single whole lambda DNA molecule. The length of a equals that of a′ and varies from 0 to 100% the length of monomeric lambda DNA. And not only does the length of a equal that of a′ but the base sequence in each segment is the same (Schnos and Inman, 1970). These doubly branched circles have all the properties of circles in different stages of replication, with a and a′ having replicated and b awaiting replication. That the replication proceeds bi-directionally around the circle was established through the characteristic molecular partial denaturation map.

Fifteen minutes after infection the synthesis of closed circles ends and from then until the end of the latent period a new 'fast-sedimenting' form of lambda DNA can be extracted from the infected bacteria (or induced lysogens). This DNA is distinguished from replicating circles in that it releases, upon denaturation, single strands of greater than monomer length. Radioactive pulse-chase experiments show that the fast-sedimenting DNA is a precursor of linear monomers found in phage

particles (Carter et al., 1969).

What is the precursor of fast-sedimenting DNA? The fate of radioactive thymidine presented in a two-minute pulse starting five minutes after infection has been studied by Carter et al. At 35 minutes after infection one-third of the radioactivity was found in covalently closed circles, one-third in linear monomers and one-third in fast-sedimenting DNA. Thus if fast-sedimenting DNA is the only precursor of linear molecules, two-thirds of the DNA labelled by the pulse must have become fast-sedimenting. Unless the precursors of late DNA are preferentially lost during extraction, they must have been visible in the electron microscopic studies where little other than circular structures were apparent. Covalently closed circles are not precursors to fast-sedimenting DNA (Carter et al., 1969). Therefore nicked circles synthesised early are probably the precursors of fast-sedimenting DNA.

Fast-sedimenting DNA has several interesting properties. The sedimentation velocity ranges from 1 to 1.7 times that of linear monomers (Skalka, 1971) and when it is denatured in alkali, single-stranded monomeric circles are released (Kiger and Sinsheimer, 1969). These properties could be accounted for by a monomeric circle with a linear tail of variable length; that is, by a rolling circle (Gilbert and Dressler, 1968).

Despite the obscurity of the structure of late DNA

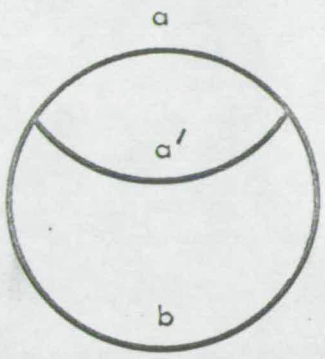



Figure 1.1. Doubly-branched circle of replicating lambda DNA.

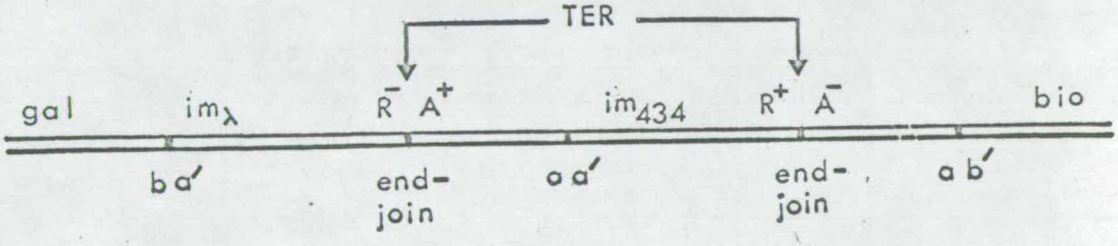



Figure 1.2. Structure of dilysogenic strain used by Mousset and Thomas. ba′, aa′ and ab′ are the Campbell recognition regions. The strain is a dilysogenic derivative of a rec $A^-$ strain that is permissive ($su^+$) for amber (sus) mutants. $R^-$ and $A^-$ are sus mutants.

maturation of the DNA must involve endonucleolytic attack at the site of the cohesive ends. Two staggered single-chain scissions must be introduced at precise points, twelve base-pairs apart on opposite strands. Mature lambda DNA molecules may then be released by disruption of the base-pairs between the nicks. The enzyme (or enzymes) responsible for generating the termini has been called Ter, a name coined by Mousset and Thomas (1968) who inferred the existence of this function from genetic studies on the excision of a lambda chromosome from a doubly lysogenic bacterium lacking the recombination functions rec, red and int.

Strains lysogenic for both $\underline{im}_{\lambda}\underline{R}^{-}$ and $\underline{im}_{434}\underline{A}^{-}$ phages were constructed - one of these is shown in Figure 1.2. The structure of such tandem dilysogens was checked by joint P1-transduction of prophage markers with bacterial gal and bio markers. On superinfection of the dilysogen shown, with a heteroimmune $\underline{red}^{-}$, $\underline{int}^{-}$ phage, the phage particles recovered with either prophage immunity were almost all $\underline{im}_{434}$, and in addition all these $\underline{im}_{434}$ phage were $\underline{R}^{+}\underline{A}^{+}$ recombinants. Thus it is clear that dilysogenic excision takes place specifically between genes R and A, as expected if mediated by the termini generating function.

The excised phage is precisely that recombinant which cuts made at the end-joins would be expected to produce. If the normal prophage excision mechanism is

not blocked, its action masks the excision due to cuts at the end-joins (Szpirer and Brachet, 1970). This is because integration-excision functions are expressed early; the Ter function is expressed late.

More recently, Wang and Kaiser (1973) have initiated the biochemical characterisation of the Ter enzyme. Fortunately, there is a means of detecting cohesive ends: transfection by lambda DNA (Kaiser and Hogness, 1960; Kaiser, 1962) requires the presence of at least one cohesive end capable of cohering with DNA of a helper phage. This assay was used to detect the increase of infectivity of covalently closed $\lambda\, \underline{b}_2\, \underline{im}_{21}$ DNA after incubation with various cell extracts. The covalently closed DNA substrate contained 20-30% (by mass) of the circular DNA as dimeric circles.

A number of observations were made:

(i) Extracts of lambda-infected cells contain an activity that generates cohesive ends from closed circular molecules.

(ii) The activity is dependent upon the presence of ATP.

(iii) The activity is dependent upon the function of gene A; either amber mutation or deletion of gene A removes the activity.

Equilibrium favours the joining of cohesive ends under the conditions used by Wang and Kaiser. However, heating at $70^{\circ}$ for five minutes did not lead to an increase in infectivity. However it may be that in the

assay, the products of the intramolecular dissociation of the cohesive ends are effectively trapped and sequestered from this reversible process.

These experiments indicate that an activity having the anticipated properties of Ter is determined at least in part by gene A of lambda. (The participation of host functions and some other phage function cannot be ruled out.) Several proteins known to act or presumed to act at specific sites in lambda DNA are determined by genes situated close to their corresponding sites. This generalisation applies to most if not all the following genes: int, xis, O, P, cI, cII, cIII, cro, N and Q (Hershey, 1971). Such an arrangement assures that these genes and their targets are not too readily separated by recombination. Perhaps the same reasons of security explain the location of the end-join next to gene A.

The ATP dependence of Ter is a property shared by a number of other nucleases including the K, B and P1 restriction enzymes (Meselson and Yuan, 1968; Linn and Arber, 1968; Roulland-Dussoix and Boyer, 1968) and the rec BC nuclease (Goldmark and Linn, 1970).

What is the molecular basis of the interaction between the Ter enzyme and the cohesive end site? Part of this thesis describes experiments to determine sequences beyond the nicks introduced by Ter. These sequences are amenable to analysis by 3′-terminal labelling lambda DNA. In conjunction with the sequence

of the twelve base pairs between the nicks, these sequences suggest mechanisms for the interaction of Ter with its target and predict properties of the enzyme.

## 1.3 Host-controlled restriction and modification of DNA

### Systematic description

Host-controlled restriction and modification is a phenomenon exhibited by many strains of bacteria. It is most readily detected as the inactivation of phage following transfer from one bacterial strain to another. Thus lambda grown on one host strain may form plaques on a second host strain with a much lower efficiency than on the first, i.e. the phage are restricted by the second host. If the surviving phage are now used to infect the second host strain they are no longer restricted; but if they are first cycled through the original host strain they are again restricted on the second host. The non-heritable change conferred on the phage by the second strain is called modification. Host-controlled restriction and modification has been reviewed very extensively recently (Arber and Linn, 1969; Arber, 1971; Boyer, 1971; Meselson et al., 1972; Arber, 1973).

The restricted phage adsorbs normally to the restrictive host and injects its DNA. When the phage are labelled with $^{32}P$ it is apparent that the phage DNA undergoes degradation to acid-soluble fragments soon

after injection into restrictive bacteria (Arber and Dussoix, 1962; Dussoix and Arber, 1962; Lederberg and Meselson, 1964). No DNA degradation occurs with appropriately modified phage. The endonuclease primarily responsible for the degradation of unmodified DNA was first isolated from E. coli strain K by Meselson and Yuan (1968).

Arber and Dussoix investigated the modification of lambda by prophage P1 and showed that modification is a property of the phage DNA, (Arber and Dussoix, 1962). They found that most of the progeny issuing from strain K after infection with λ.K, P1 had lost the capacity to grow on strain K (P1). But by labelling the parental phage DNA either with $^{32}P$ or with deuterium they could show that phage particles with at least one conserved parental strand could still grow on K (P1). This finding indicated that the P1-lysogenic host modified the DNA by adding something that strain K could neither add nor remove.

Modified DNA is protected because it is methylated at a (very limited) number of sites (Arber and Linn, 1969; Smith et al., 1972; Brookes et al., 1972). The bacterial DNA of the second host is impervious to its own restriction endonuclease because it is appropriately modified. DNA methylated in one strand only - either as a result of one round of semi-conservative replication in a non-modifying host, or as a heteroduplex constructed

*in vitro* - is resistant to restriction (Arber and Dussoix, 1962; Dussoix and Arber, 1962; Meselson and Yuan, 1968). This may be necessary for the protection of host DNA immediately after replication.

DNA restriction and modification is not confined to bacteriophage infection. DNA transferred by transduction or conjugation (Arber and Morse, 1965) or transfection (Benzinger, 1968) is also subject to the constraint of host-specificity.

Host-specificity is found in many species and strains of bacteria and more than one host-specificity system may operate in a single bacterium, in which case each system acts independently and their effects are cumulative. The responsible genes can be located on the bacterial chromosome or in plasmids. It should be noted that not all DNA-containing phages are subject to a given host-specificity system.

## DNA restriction and modification enzymes

The enzymes from *E. coli* strains K and B, and P1-lysogenic *E. coli* were the first restriction enzymes studed *in vitro* (Meselson and Yuan, 1968; Linn and Arber, 1968; Roulland-Dussoix and Boyer, 1968). Recently, a number of restriction enzymes have been identified in a wide variety of sources. Table 1.1 is a partial list of the restriction and modification systems known to date and gives a nomenclature for them

(Danna *et al.*, 1973; Nathans and Smith, 1973). Some of the recently discovered so-called restriction endonucleases listed in Table 1.1 have not formally been shown to be part of a host-controlled restriction and modification system; they are however inactive on host DNA *in vitro*.

The K, B and P1-specific restriction and modification enzymes

The enzymology of R.$Eco_K$ and R.$Eco_B$ has been studied extensively. The two enzymes are very similar in all properties except their own particular specificity, and the genes corresponding to them in strains K and B may be regarded as allelic (Glover and Colson, 1969). R.$Eco_{P1}$, from the initial report of its properties (Meselson and Yuan, 1968), appeared to be rather similar to these two endonucleases.

R.$Eco_K$ and R.$Eco_B$ require magnesium ions, ATP and S-adenosyl methionine as cofactors. ATP is hydrolysed during restriction *in vitro* (Yuan *et al.*, 1972; Eskin and Linn, 1972b). The products of hydrolysis are ADP and $P_i$. The ATPase reaction of these enzymes is curious in that the number of ATP molecules hydrolysed is vastly in excess of the number of DNA phosphodiester bonds broken; the exact ratio depends upon the reaction conditions but it is of the order of $10^5$ ATP molecules per single-strand break. In addition, there is no

Table 1.1 Restriction and Modification Systems: Nomenclature and Identification

| Host | Restriction Modification System[a] | How identified[b] | References |
|---|---|---|---|
| *Bacillus subtilis* | $Bsu_{X5}$ | G,R | Bron (1973) |
| *Haemophilus aegyptius* | Hae | R | Middleton *et al.* (1973); this thesis appendix I |
| *H. parainfluenzae* | Hpa I | G,R | Gromkova & Goodgal (1972); Sharp *et al.* (1973) |
| | Hpa II | G,R | Ditto |
| *H. influenza* | $Hin_a$ I | G | Piekarowicz & Glover (1972) |
| | $Hin_a$ II | G | Ditto |
| | $Hin_b$ | G | Ditto |
| | $Hin_c$ I | R | Landy (quoted in Nathans & Smith, 1973) |
| | $Hin_c$ II | R | Ditto |
| | $Hin_d$ I | G,M,R | Glover & Piekarowicz (1972); Roy & Smith (1973); Gromkova *et al.* (1973) |
| | $Hin_d$ II | M,R | Roy & Smith (1973); Kelly & Smith (1970); Smith & Wilcox (1970) |
| | $Hin_d$ III | M,R | Roy & Smith (1973); this thesis |
| | $Hin_e$ | G | Piekarowicz & Glover |
| | $Hin_f$ | G | Ditto |

Table 1.1 (continued)

| Host | Restriction Modification System[a] | How identified[b] | References |
|---|---|---|---|
| *Escherichia coli* | $Eco_A$ | G | Arber *et al.* (1972) |
| | $Eco_B$ | G,M,R | Arber (1965); Linn & Arber (1968); Roulland-Dussoix & Boyer (1968); Lautenberger & Linn (1972) |
| | $Eco_K$ | G,R | Arber (1965); Meselson & Yuan (1968) |
| | $Eco_{15}$ | G | Arber & Wauters-Willems (1970) |
| | $Eco_{P1}$ | G,M,R | Arber (1965); Brockes *et al.* (1972); Meselson & Yuan (1968); this thesis |
| | $Eco_{RI}$ | G,M,R | Yoshimori (1971); Hedgpeth *et al.* (1972); Boyer (quoted in Nathans & Smith, 1973) |
| | $Eco_{RII}$ | G,M,R | Yoshimori (1971); Bigger *et al.* (1973); Boyer *et al.* (1973) |
| *Salmonella typhimurium* | $Sty_{LT}$ | G | Glover & Colson (1967) |
| | $Sty_{N3}$ | G | Hattman (1971) |

Legend to Table 1.1: (a) Restriction enzymes are designated R, and modification enzymes M, followed by the three letter code describing the organism. Strain or type identification is written as a subscript. When a particular strain has more than one identified restriction and modification

system, these are indicated by Roman numerals. Thus restriction enzymes from H. influenza strain Rd are R.Hin$_d$ I, etc.

(b) R, restriction endonuclease identified.

G, system identified by efficiency of plating of phage.

M, modification methylase identified.

---

theoretical [redacted] energy requirement for the catalysis of the DNase reaction; many nucleases have no energy-providing cofactors. The requirements of the ATPase reaction are the same as those for restriction. Unmodified DNA is required. Hydrolysis of ATP continues long after DNA degradation has ceased, but no ATPase is observed when restricted DNA but no unmodified DNA is presented to the enzyme. Restricted DNA is required for the maintenance of ATPase once initiated. Eskin and Linn (1972) suggest that after an initial ATP-dependent hydrolysis of unmodified DNA, the enzyme assumes an altered form which can no longer serve as a nuclease (little or no turnover can be detected with R.Eco$_B$ as a nuclease) but which is active as an ATPase. Whether this ATPase is significant in vivo is unclear. It may be controlled by the further degradation of the DNA to acid-soluble fragments that occurs after the primary restriction.

No comparable ATPase activity is found with R.Eco$_{P1}$

but the hydrolysis of one or a few ATP molecules per polynucleotide chain scission would not have been detected (Habermann, quoted in Yuan et al., 1972).

One may speculate that SAM acts as methyl donor in enzymatic methylation of the DNA during restriction. It is known from work with monofunctional methylating agents such as methyl methane sulphonate that DNA can be broken by methylation (Lawley, 1963). Methyl methane sulphonate reacts with purine nucleophilic centres such as the N7 of guanine and the N3 of adenine. This labilises the N-deoxyriboside bond. The resulting apurinic structure is labile at non-neutral pH so that the phosphodiester bonds on either side of the deoxyribose residue may break. Linn (quoted in Hershey, 1971) and Brockes (1972) have proposed that R.Eco$_K$ and R.Eco$_B$ cause methylation of certain purine residues that would be followed by local depurination and breakage of the DNA in enzyme catalysed reactions. Radioactivity from methyl-labelled SAM does not appear in the DNA fragments produced in K-specific restriction, but it is present in DNA at early stages of the reaction, probably at the N3 position of adenine (Brockes, 1972).

R.Eco$_{P1}$ was initially reported to have a requirement for SAM similar to that of R.Eco$_K$ (Meselson and Yuan, 1968).

Direct determination of nucleotide sequences at the sites of restriction by R.Eco$_K$ with the 5'-terminal

labelling method has given consistently negative results (Murray et al., 1973). No new terminal sequences due to restriction were found (whether the fragments were first treated with alkaline phosphatase or not). Eskin and Linn (1972a) have reported that R.Eco$_B$ behaves similarly, but do not present their data.

The inability to find new 5′-terminal sequences in this type of experiment may be a result of the mechanism of DNA hydrolysis. If depurination occurs at the site of hydrolysis and phosphodiester bond cleavage occurs so as to leave an apurinic deoxyribose residue at the 5′-terminus (either with or without a 5′-monophosphate) then participation in the polynucleotide kinase reaction (Novogrodsky and Hurwitz, 1966) is not anticipated. Fragments produced from duplex closed circular replicative form (RFI) of phage fd DNA by the action of R.Eco$_B$ are susceptible to lambda exonuclease, exonuclease III, the rec BC nuclease, and after denaturation, exonuclease I (Eskin and Linn, 1972a). These results imply that the termini of the restricted DNA have 3′-hydroxyl and 5′-phosphoryl groups, although the behaviour of an apurinic site with these enzymes is unknown.

An alternative explanation for the failure to find a sequence at the sites of restriction comes from the recent experiments of Horiuchi and Zinder (1972). They showed that the sites of cleavage of the duplex circular RF of phage f1 DNA do not correspond to the genetic sites

(sB-sites) that confer susceptibility to R.Eco$_B$. The number of possible cleavage sites is larger than the number of sB-sites. This conclusion comes from the observation that linear molecules produced by the action of R.Eco$_B$ on a mutant of f1 RFI DNA that has just a single sB-site can be circularised by denaturation and renaturation.

A similar observation was made for the linear molecules produced by the action of R.Eco$_{P1}$. In this case however it has been suggested that cleavage occurs at one of a number of possible sites partly as a result of P1-specific modification by the enzyme preparation (Haberman et al., 1972). There was no single-site sP1 mutant available. Mulder and Delius (quoted in, Mulder and Delius, 1972) have shown that R.Eco$_{P1}$ makes a single break in circular SV40 DNA, and that this break is not at a unique site.

The sB-site is unaffected when f1 DNA is cleaved in restriction since an sB-site can be used repeatedly by R.Eco$_B$; the circles reformed from linear molecules after denaturation and renaturation can be cleaved by the same enzyme. These observations indicate that the sB-sites are not the sites of cleavage by R.Eco$_B$. It is possible that the enzyme recognises unmodified DNA and then travels along the DNA to some point, perhaps dictated by a random process, before breaking the DNA; alternatively the DNA may turn back on itself so as to come into contact with

the enzyme that remains at the sB-site.

The possibility that these enzymes travel along the DNA is extremely interesting. The lactose and lambda repressors react so rapidly with their operators that the rate of reaction surpasses that of the diffusion-limited process (Riggs et al., 1970). An extreme solution to this problem is to suggest that the operator sequence is located by a pseudo-one-dimensional walk along the DNA molecule (Gilbert and Muller-Hill, 1970). R.Eco$_B$ may show similar kinetics but the detailed analysis has yet to be done (Boyer et al., 1971). Thus it may be that the enzyme travels along the DNA both before and after recognition of an unmodified site.

Horiuchi and Zinder also showed that cleavage of the duplex circular RF of (wild-type) phage f1 DNA by R.Eco$_B$ gave rise to full length linear molecules, even though f1 DNA contains two well separated sB-sites. The linear molecules are not susceptible to R.Eco$_B$. This result is in direct conflict with the observation of Eskin and Linn (1972) that restriction of unmodified phage fd RF DNA having two well separated sB-sites (similar to those of the closely related f1 DNA) produces fragments smaller than the full length linear molecules that are produced from a mutant DNA with a single sB-site. The reason for the discrepancy between these two sets of observations is totally unclear and it is to be hoped that the situation will soon be resolved.

This insusceptibility of linear f1 molecules to R.Eco$_B$ is puzzling; one possible explanation is that linear DNA molecules must be larger than about $3.5 \times 10^6$ in molecular weight before there is an appreciable chance of their being broken. Circular molecules have properties similar to longer linear molecules. However there are preliminary results (Morrison, personal communication) indicating that a fragment of lambda DNA of about $4.5 \times 10^6$ in molecular weight having an sK-site located about one-third of the length from one end is susceptible to R.Eco$_K$.

Analysis by zone centrifugation in neutral sucrose gradients or by gel electrophoresis, of DNA fragments produced by the action of R.Eco$_K$ on sK-site mutant DNAs of lambda X phi 80 hybrid phages (Murray et al., 1972) shows that this endonuclease also cleaves at sites removed from the unmodified sK-site (Murray et al., 1973).

If the site of endonucleolysis is indefinite with respect to nucleotide sequence, the subsequent randomisation of terminal sequences in a terminal labelling experiment would be hard to detect and distinguish from a failure to label the termini created by restriction. This is the case even in experiments where restricted lambda DNA fragments are used in the polynucleotide kinase reaction and the known 5′-terminal sequences serve as an internal control for the labelling reaction. Experiments to determine the sequence at the site of

hydrolysis of unmodified lambda DNA by R.Eco$_{P1}$ are reported in this thesis. During the course of these experiments similar experiments have been performed with R.Eco$_K$ (Murray et al., 1973). However, it has simultaneously become apparent that these two endonucleases are significantly different in several respects; this will be considered in the discussion.

Whereas the restriction endonucleases R.Eco$_B$ and R.Eco$_K$ exhibit unusual properties, the modification methylases M.Eco$_B$ (Lautenberger and Linn, 1972) and M.Eco$_{P1}$ (Brockes et al., 1972) are less bizarre. It is probable that the sites methylated by the modification enzymes correspond to the genetically defined target sites. Measurements of the number of bases methylated in B-specific modification in vivo (Smith et al., 1972) and in vitro (Kuhnlein and Arber, 1972) of fd DNA having zero, one or two sB-sites indicated that DNA is modified by methylation of two adenine bases per site. The product of methylation is a 6-methylamino purine. More convincing evidence comes from experiments with M.Eco$_{P1}$ (Brockes, 1972).

A highly purified preparation of M.Eco$_{P1}$ was used to methylate unmodified DNA from phages lambda and 82 with $^{14}$C-[Me]-labelled SAM, and the labelled DNA was purified and digested with pancreatic DNase. Oligonucleotides were fractionated by two-dimensional ionophoresis. Although tiresomely long exposures were necessary, radioautographs were used for analysis of partial digests of $^{14}$C-labelled

oligonucleotides with venom phosphodiesterase. These results, together with the positions of the oligonucleotides on the nucleotide maps showed that the major components of the digests were pA-G-mA, pG-mA, pG-mA-T, pG-mA-T-C, pmA-T, and pmA-T-A, although several other unidentified oligonucleotides were also present. (mA indicates $^{14}$C- Me -6-methylaminopurine.) It is tempting to infer that these oligonucleotides are all derived from pA-G-mA-T-C, itself part of the rotationally symmetrical hexanucleotide pA-G-mA-T-C-T. However this conclusion requires identification of the hexanucleotide as the major methylated hexanucleotide derived from modified DNA, which has not been done. These results do show that M.Eco$_{Pl}$ methylates adenine residues within specific sequences. In all probability these are the recognition targets (sPl-sites) for the modification (and restriction) enzymes. M.Eco$_B$ and M.Eco$_{Pl}$ have no absolute requirement for cofactors other than SAM. ATP can have a slight stimulatory effect.

Restriction and modification enzymes from Haemophilus

The interesting properties of the restriction enzymes just discussed have proven somewhat frustrating in direct approaches to the study of the sequence recognised by these proteins. The apparent lack of catalysis (i.e. no turnover) of these enzymes is of practical importance making it difficult to obtain large

quantities of restricted DNA. This is not the case with several other restriction endonucleases. The first of these studied *in vitro* was an enzyme preparation, initially called R.H, obtained from *Haemophilus influenzae* strain Rd by Smith and Wilcox (1970). This enzyme made about forty breaks in T7 DNA.

The 5′-terminal sequences around these breaks were determined from an elegant experiment (Kelly and Smith, 1970) in which T7 DNA, uniformly labelled with $^{33}P$, was restricted and dephosphorylated and the 5′-termini labelled with $^{32}P$ in the polynucleotide kinase reaction. After digestion of the product to 5′-mononucleotides, only pA and pG carried the $^{32}P$-label, with 63% of the $^{32}P$ in pA and 37% in pG. (The corresponding figures obtained when P22 DNA was used as substrate were 85% pA and 15% pG.) Analysis of the 5′-di and trinucleotides in DNase digests of the labelled restriction products showed that the two sequences pA-A-C and pG-A-C were present and it was concluded that the enzyme breaks both strands of the duplex on the 5′-side of the sequence pR-A-C. The 3′-terminal sequences at the points of breakage were determined by analysis of a micrococcal nuclease digest of a limit digest of $^{32}P$-uniformly labelled T7 DNA with the restriction enzyme. Since micrococcal nuclease gives fragments with a 3′-phosphate and 5′-hydroxyl group, the 3′-terminal nucleotides from sites of breakage by R.H differ from those from other

parts of the molecule. Two dinucleoside monophosphates were obtained: TpC and TpT, the latter appearing in higher yield. These results are consistent with the enzyme making an even (as opposed to a staggered) break across the two strands through the centre of the rotationally symmetrical sequence 5′- G-T-Y-R-A-C.

The nucleotide sequences at the points of breakage of lambda DNA by R.$Eco_{RI}$ (Hedgpeth et al., 1972) and R.$Eco_{RII}$ (Bigger et al., 1973) were determined using the polynucleotide kinase reaction. Electron microscopy and repair reactions with Rous Sarcoma Virus DNA polymerase have been used to confirm the nature of the breaks produced (Boyer et al., 1973). These sequences will be considered in detail in the discussion.

Further work with R.H showed that distinctive patterns of radioactive bands could be obtained in radioautographs of acrylamide gels (or mixed acrylamide-agarose gels) after electrophoresis of $^{32}P$-labelled SV40 DNA (Danna and Nathans, 1971) and RF ØX174 DNA (Edgell et al., 1972). Danna and Nathans found with R.H that DNA from large-plaque, small-plaque and minute-plaque strains of SV40 gave fragments with specific differences in their mobility on gel electrophoresis. They were able to order the 11 fragments of the SV40 genome by analysis of partial digest products and by analysis of an overlapping set of fragments produced by a restriction endonuclease preparation from H. parainfluenzae. In addition, the

single site in SV40 DNA cleaved by R.Eco$_{RI}$ (Morrow and Berg, 1972) was located. With this site as a reference point the cleavage sites were located on a physical map of SV40 DNA (Danna et al., 1973). Danna and Nathans (1972) exploited this system in an elegant series of experiments with pulse-labelled DNA which showed that replication of SV40 DNA commenced at a specific site and proceeded bidirectionally around the circular molecule to terminate at a point about half-way round.

During the course of this work it was noticed that different preparations of R.H gave rather different fragmentation patterns with SV40 DNA. This prompted attempts to purify the R.H preparation further (Smith, personal communication). On chromatography on DEAE-cellulose a broad peak of endonuclease activity was eluted in a NaCl gradient. This peak was detected with salmon sperm DNA as substrate. However only the earlier fractions of this peak were active on T7 DNA. It was apparent that R.H was composed of more than one enzyme. Smith suggested that an endonuclease (R.Hin$_d$ II) was present that was active on T7 and salmon sperm DNA, accompanied by another endonuclease (R.Hin$_d$ III) that was inactive on T7 DNA. R.Hin$_d$ II gave 5 fragments with SV40 DNA. The activity of R.Hin$_d$ III was lost however, precluding further experiments.

The nucleotide sequence at the site of action of R.Hin$_d$ III was determined in experiments described in

this thesis.

Recently Roy and Smith have investigated the DNA methylases of H. influenzae Rd (Roy and Smith, 1973a; 1973b). Four separate activities were found, each of which required SAM as methyl group donor and each differed in its ability to methylate various DNAs in vitro.

DNA methylase I ($M.Hin_d$ I) is related to the genetically described restriction and modification system in H. influenzae Rd (Glover and Piekarowicz, 1972). This methylase is active on H. influenzae Rd DNA. This is an unexpected result because DNA methylases are normally not active on host DNA. However Glover and Piekarowicz have noted that the strain used in these studies is phenotypically unstable and consists of a mixed population of cells that is 50% $r^+m^+$ and 50% $r^-m^-$ for the restriction and modification system manifested by the efficiency of plating of phage HP1c1 on various H. influenzae strains. Therefore $M.Hin_d$ I from the $r^+m^+$ cells may be methylating unmodified DNA from $r^-m^-$ cells. Gromkova et al. (1973) have purified a restriction activity from strain Rd which appears to be specified by the genetic system described by Glover and Piekarowicz. This enzyme destroys transfecting activity of DNA from phage grown on $r^-m^-$ cells but not $r^+m^+$ cells. This restriction enzyme requires ATP and SAM as cofactors. $M.Hin_d$ I

is apparently the modification enzyme corresponding to this enzyme.

DNA methylase M.Hin$_d$ II is active on T7 DNA and protects it against the R.H preparation described by Smith and Wilcox (1970). Conversely, R.H destroys methylase II sites in T7 DNA.

DNA methylase M.Hin$_d$ III, like R.Hin$_d$ III, does not affect T7 DNA.

The biological function of DNA methylase IV remains unknown. The possibility exists that this enzyme serves some function other than modification of restriction sites.

Each of these four methylases specifically methylates adenine residues in the 6-amino position. Roy and Smith performed methylation reactions with each of these methylases and $^3$H-[Me]-SAM. The DNA was then digested to dinucleotides with the Bacillus subtilis phage nuclease, SP3 DNase (Trilling and Aposhian, 1968), followed by removal of the 5′-phosphoryl group with phosphatase to give dinucleoside monophosphates. These were analysed by chromatography on ion-exchange columns. Dinucleoside monophosphate species containing the 3′ or 5′ neighbour of the methylated base are resolved so that a trinucleotide with a centrally placed methylated adenine is determined. The relationship of these methylation sites to the site of nucleolysis by the corresponding restriction enzymes will be considered in

the discussion.

There are now known several examples of various strains of Haemophilus species with more than one restriction and modification system. These are shown in Table 1.1. It is possible that some of the systems in different strains have a common sequence specificity.

## 1.4 Terminal labelling methods

There are two general approaches to sequence determination in a defined region of a relatively much larger DNA molecule. Either the DNA in that region must be isolated and purified from the rest of 'bulk' DNA or a radioactive label is selectively introduced into the molecule at the region itself. In studying nucleotide sequences in viral DNA at the sites of action of specific endonucleases, the latter approach in the form of terminal labelling, is the appropriate one.

### 5′-terminal labelling

The polynucleotide kinase reaction is a very powerful tool in 5′-terminal sequence analysis (Murray, 1973; Szekely and Sanger, 1969) of DNA molecules or large fragments of them such as those produced by restriction enzymes. The reaction is extremely sensitive when $\gamma$-$^{32}$P-ATP of very high specific radio-

activity (about 10 mCi/umole or greater) is used. If the labelled DNA preparation is digested under mild conditions with a non-specific nuclease (such as pancreatic DNase) that breaks virtually randomly with respect to sequence, the range of radioactive products will be equivalent to a partial venom phosphodiesterase digest of the largest of the labelled oligonucleotides. If the DNA or polynucleotide is single-stranded then the labelled nucleotides will form a progressively overlapping homologous 5′-terminal series, and the sequences of members of this series can be deduced through the characteristic change in ionophoretic mobility on ion-exchange papers of an oligonucleotide brought about by the addition (or subtraction) of one nucleotide and the use of this change to identify the base by which two such nucleotides differ (Sanger et al., 1965; Murray, 1970; Murray, 1973). When two complementary ion-exchange systems are used for analysis of the partial digest (DE-81 paper at pH 2 and AE-81 paper at pH 3.5) it is possible to deduce uniquely the sequence of the oligonucleotides present in the digest. The extent of the sequence that can be deduced in this way is dependent upon the base composition to some extent, for this affects the mobility of the oligonucleotides and hence the resolution of the larger members of the series, but usually the method is applicable to nucleotides up to six residues.

The characteristic mobility shift (M) values for the addition (or subtraction) of each nucleotide are given in Table 1.2. The following points should be noted. With ionophoresis on DE-81 paper at pH 2 the effect of adding a pC residue is very small and may go unnoticed; and the effects of pG and pT residues are similar so that any sequence deduced is ambiguous where either of the bases G or T occurs. Ionophoresis of the digest on AE-81 paper at pH 3.5 solves these uncertainties; the M values for pC residues are unmistakable, and although M values for pA and pT residues are now ambiguous, pG has a value characteristically larger than that for pT (Murray, 1973).

From these two ionophoreses the 5′-terminal sequence can be deduced uniquely and the 5′-dinucleotide is identified from its mobility on the two papers. The sequence of the dinucleotide can be deduced by ionophoresis on AE-81 paper at pH 3.5 if appropriate dinucleotides are used as reference compounds, for the members of each pair of sequence isomers have slight differences in mobility (Murray, 1973). However it is more usual to deduce the sequence of the dinucleotide by identification of the 5′-terminal nucleotide produced upon hydrolysis with venom phosphodiesterase.

Thus the 5′-terminal sequence of single-stranded DNA can be deduced directly by inspection of autoradiographs of the two ionophoretograms. However, one would

Table 1.2 M Values for Deoxynucleotides

The M value is the ratio of the difference in mobility (or $R_P$) between two adjacent oligonucleotides of a homologous series to the mobility of the slower of the two. On AE-81 at pH 3.5 the values tend to increase with increasing size of the oligonucleotides.

| | Range of M values on: | |
|---|---|---|
| Base | DE-81, pH 2 | AE-81, pH 3.5 |
| C | 0.05 - 0.2 | 0.5 - 0.7 |
| A | 0.4 - 1.0 | 0.9 - 1.5 |
| G | 1.4 - 3.0 | 2.0 - 3.0 |
| T | 1.4 - 3.0 | 1.0 - 1.5 |

seek to confirm this sequence with partial digests of oligonucleotides. This ensures that no oligonucleotide in the series has been missed. The most likely error due to overlooking a nucleotide in the 5′-terminal series would be the interpretation of the difference due to the combined effect of pT and pC as a difference due to pG.

The products from similar experiments with a double-stranded molecule or fragment comprise two such families of overlapping nucleotides, but providing that these nucleotides can be assigned to their respective terminal family, the analysis can be completed in the same way

(Murray and Murray, 1973). If the two 5′-terminal nucleotides differ this assignment is trivial, but if they are the same the assignment must be based upon terminal di- and trinucleotides, etc., after partial exonuclease digestion.

3′-terminal labelling

The radioactive products present in a digest of 3′-terminally labelled DNA can be analysed in a way analogous to those from 5′-terminally labelled DNA (Weigel et al., 1973; Murray and Old, 1973).

From the start of the work described in this thesis it was apparent that a 3′-terminal labelling reaction that could be used in a similar way to, and in conjunction with, the polynucleotide kinase reaction would be extremely useful in determining the DNA structure at termini created by the action of specific endonucleases on duplex DNA. Two immediate considerations prompted the investigation of a likely 3′-terminal labelling reaction. (i) On the hypothesis that the apparent inertness of DNA termini produced by restriction with R.Eco$_K$ to terminal labelling in the polynucleotide kinase reaction (Murray et al., 1973) is due to a 5′-terminal apurinic site, the 3′-terminal labelling approach is attractive. (ii) Application of 3′-terminal labelling to phage lambda DNA provides access to the sequences in the region of the nicks produced by the Ter endonuclease.

## Nucleotide phosphotransferase from *E. coli*

In surveying the nucleoside phosphotransferase of *E. coli* (Brawerman and Chargaff, 1955) an enzyme was encountered that may be described as a nucleotide phosphotransferase: an enzyme transferring organically bound ester phosphate to the 3′- or 2′- hydroxyl of a nucleoside 5′-phosphate. Brunngraber and Chargaff (1970) described further observations of this enzyme showing that it is relatively unspecific in the nature of the nucleotides that may be phosphorylated. This fact prompted experiments described in this thesis to determine whether the enzyme will catalyse the phosphorylation of DNA or polydeoxynucleotides at their 3′-terminus.

The observations of Brunngraber and Chargaff included the following. All nucleosides and nucleoside 5′-monophosphates examined served as acceptors of phosphate groups. Those examined included the four common ribo and deoxyribo derivatives. Deoxyribonucleosides, deoxyribonucleoside 5′-monophosphates and deoxyribonucleoside 5′-triphosphates were phosphorylated at the 3′-hydroxyl (exclusively) in good yield. The phosphate group donor in all these reactions was p-nitrophenyl phosphate, which was the most efficient donor tested and surpassed adenosine 3′-monophosphate, the only nucleotide tested as donor. The physiological substrates are unknown as is the biological significance of the enzyme.

Since the enzyme was not fully purified it is not certain that all the activities described should be attributed to the same protein.

Exonuclease-repair reactions with T4 DNA polymerase

Recently two methods of 3′-terminally labelling DNA have been published. One of these makes use of terminal deoxynucleotidyl transferase (Kato et al., 1967) and has been developed by Kossel and Roychoudhury (1971). The other method is due to the work of Englund and involves exonuclease-repair reactions catalysed by the DNA polymerase of phage T4 (Englund, 1971a; 1971b). This latter method has been the basis of several experiments described in this thesis and is therefore described in detail.

The method uses the exonuclease activity associated with the T4 DNA polymerase which releases 5′-mononucleotides exclusively from 3′-hydroxyl termini. Linear duplex DNA can be degraded completely (Goulian et al., 1968; Englund, 1971b), but in the presence of any single deoxynucleoside triphosphate degradation of duplex DNA is drastically reduced (although degradation of single-stranded DNA is unaffected). The triphosphate does not actually inhibit the exonuclease activity, but after initial exonucleolysis until a complementary base is reached on the opposite strand, it serves as substrate for replacement of the nucleotide just removed. Subsequent reaction consists of the alternating incorporation

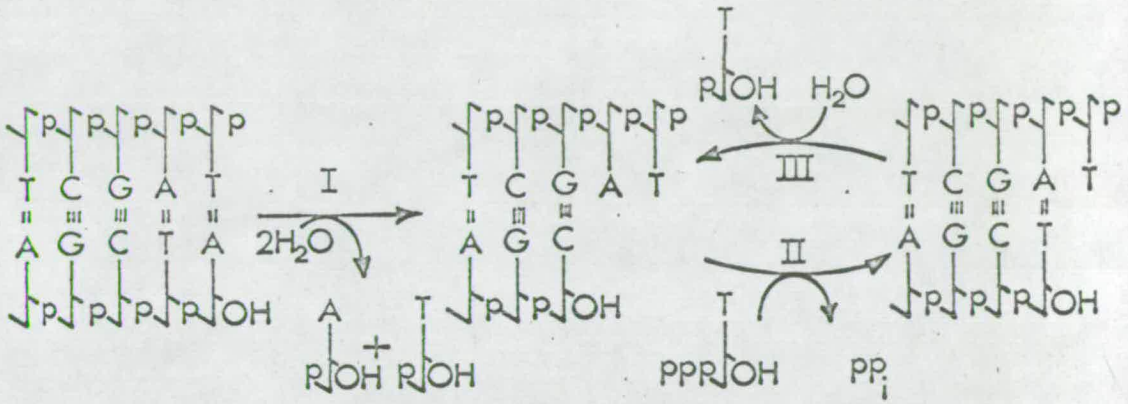



Figure 1.3. A scheme for the hydrolysis of linear duplex DNA by the T4 DNA polymerase in the presence of dTTP.

and removal of the 3′-terminal nucleotide (Englund, 1971a). These reactions are summarised in Figure 1.3.

The reaction may be used for sequence analysis in two ways: (i) with uniformly labelled DNA and an analysis of nucleotides released, (ii) as a terminal labelling method with non-labelled DNA after the incorporation of radioactivity from an $\alpha$-$^{32}$P-labelled deoxynucleoside triphosphate. In the latter mode radioactivity is incorporated at or near the original 3′-terminus. In either case four reactions can be carried out with each triphosphate in turn. In order to limit the exonucleolytic degradation the reaction must be carried out at low temperature (11°). Possibly at higher temperatures the enzyme recognises the termini as single strands because of a tendency to local denaturation and their degradation would then be unaffected by the presence of triphosphate.

The use of T4 DNA polymerase as a 3′-terminal labelling method is the basis of the experiments reported in this thesis to determine 3′-terminal nucleotide sequences of lambda DNA. These experiments were part of a collaborative series involving Drs. P. Englund, P. Weigel and K. Murray. The terminal labelling reactions were performed in Dr. Englund's laboratory, and the labelled material transported to Edinburgh. Similar labelling reactions have been performed by myself in experiments to determine 3′-terminal sequences of

lambda DNA fragments produced by the action of R.Hin$_d$ III, that are reported in this thesis.

## 1.5 Conclusion. Research Aims

There are many examples of proteins that interact specifically with particular sites in duplex DNA molecules. It is essential that we learn about the nucleotide sequences at such sites. These sequences may have properties which indicate mechanisms by which they are recognised. 'Are there any general recognition mechanisms?' is a question that can only be answered by studying a number of examples of the phenomenon.

Highly specific endonucleases such as the Ter enzyme and various DNA restriction enzymes are examples of such proteins. The sequences at their sites of action are amenable to sequence analysis by terminal labelling methods. In addition DNA restriction enzymes are potentially very important reagents: their capability of putting a limited number of specific breaks into DNA has immediate utility in sequencing studies as well as for the manipulation of the primary structure of DNA molecules.

The primary aims of research described in this thesis were as follows:

(i) To investigate whether a nucleotide phosphotransferase isolated from E. coli is able to

catalyse phosphorylation of the 3′-hydroxyl of polydeoxynucleotides; a reaction that could be the basis of a useful 3′-terminal labelling method.

(ii) To use the 3′-terminal labelling approach to determine the nucleotide sequences around the site of action of the Ter endonuclease.

(iii) To determine nucleotide sequences at the sites of action of R.$Eco_{P1}$ and R.$Hin_{d}$ III.

# Chapter 2

# MATERIALS

## 2.1 Bacterial strains

The bacterial strains used in this work are shown in Table 2.1.

## 2.2 Bacteriophage strains

Phage λcI857 S7 (Goldberg and Howe, 1969) is a thermoinducible lysis-defective phage convenient in DNA preparations. S7 is an amber mutation suppressible by su $III^+$. The phage was obtained by heat induction of lysogenic bacteria.

Phage T7 was obtained from Dr. R.S. Hayward.

Phage T4am N82 is a mutant defective in DNA synthesis. The levels of several 'early' enzymes (including the T4 DNA polymerase) that are induced upon infection of non-permissive host are higher than those induced by wild-type T4 (Wiberg et al., 1962). The phage was obtained from Dr. N. Symonds.

## 2.3 Media

The following media were used for the growth of bacteria and bacteriophages. All quantities are given in g/litre $H_2O$ unless otherwise stated.

L Broth (pH 7.2): Difco Tryptone 10, yeast extract 5, NaCl 10.

Table 2.1 Bacterial strains. *Escherichia coli* K12 derivatives unless stated otherwise.

| | Strain | Description | Reference | Source |
|---|---|---|---|---|
| 1 | 1100 | endonuclease $I^-$, $r_K^+ m_K^+$ | Durwald & Hoffman-Berling (1968) | I.R. Lehman |
| 2 | 1100$r_K$- (P1) | endo $I^-$, $r_K^- m_K^\pm$, $r_{P1}^+ m_{P1}^+$ | - | J.P. Brockes |
| 3 | 1100 PolA | endo $I^-$, polA1, $su^-$ | Brockes (1972) | J.P. Brockes |
| 4 | W1485 ( λ, P1) | λcI857 S7, $r_{P1}^+ m_{P1}^+$ | - | J.P. Brockes |
| 5 | C600 | $r_K^+ m_K^+$ | Appleyard (1954) | N.E. Murray |
| 6 | 803 | $r_K^- m_K^-$ | Wood (1966) | N.E. Murray |
| 7 | 803 ( λ) | λcI857 S7, $r_K^- m_K^-$ | - | N.E. Murray |
| 8 | CR63 | $suI^+$ | Appleyard *et al*. (1956) | N.E. Murray |
| 9 | *E. coli* B | wild-type | - | M. Peacey |
| 10 | *E. coli* B ($R_I$) | endo $I^-$, $r_B^+ m_B^+$, $r_{RI}^+ m_{RI}^+$ | Yoshimori (1971) | L.V. Crawford |
| 11 | *H. influenzae* Rd | wild-type | Glover and Piekarowicz (1972) | S.W. Glover |

Low-phosphate Medium (J. Abelson, personal communication): KCl 1.5, NaCl 5, $NH_4Cl$ 1, Tris base 12.1, adjusted to pH 7.4 with HCl. 20% Bactopeptone (Difco) was adjusted to pH 9 with $NH_4OH$, centrifuged to remove the precipitated $MgNH_4PO_4.6H_2O$, and then adjusted to pH 7.4. The product was diluted 1:100 in the salts solution, autoclaved, and made 0.4% in glucose and 1 mM in $MgSO_4$ before use.

Phage Buffer: $KH_2PO_4$ 3, $Na_2HPO_4$ 7, NaCl 5, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.001% gelatin.

2.4 Enzymes and proteins

(a) Pancreatic deoxyribonuclease, electrophoretically pure, was obtained as a solid from Worthington Biochemical Corpn., Freehold, New Jersey, U.S.A. It was dissolved in 0.1 M sodium acetate, 0.005 M $MgCl_2$ and stored at $-10^{\circ}$ before use. Pancreatic RNase A was also purchased from Worthington.

(b) Endonuclease I from *E. coli*, prepared as described by Lehman *et al.* (1962), was generously provided by Dr. P.R. Brown.

(c) Snake venom phosphodiesterase (*Crotalus adamanteus*) was obtained from Worthington, or from Whatman Biochemicals Ltd., Maidstone, Kent, U.K.

(d) Spleen and DNase and spleen phosphodiesterase were the generous gift of Dr. G. Bernardi.

(e) E. coli alkaline phosphatase was obtained from Worthington or Whatman. The suspension, in a solution of ammonium sulphate, was dissolved in 0.1 M ammonium bicarbonate to give a protein concentration of 1 mg/ml. This was heated at 100° for five minutes to inactivate any contaminating DNase. The velocity sedimentation behaviour of lambda DNA (1 μg) in neutral sucrose density gradients was unchanged by incubation (for 30 minutes at 37°) in a solution containing phosphatase (0.18 mg/ml) that had been prepared in this way.

(f) Polynucleotide kinase, purified from T4 am 122 infected E. coli B/r by modifications of the procedure of Richardson (1971), was the gift of A. Morrison or Dr. K. Murray. This enzyme was stored unfrozen.

(g) Nucleoside monophosphate kinase, nucleoside 5′-diphosphate kinase and creatine kinase were obtained from Boehringer Mannheim.

(h) Pronase (protease from Streptomyces griseus), Horse heart cytochrome c (grade 6), and Bovine serum albumin were obtained from the Sigma Chemical Company, Kingston-upon-Thames, Surrey, U.K.

## 2.5 Chemicals

All chemicals were reagent grade or better, and, unless stated otherwise, were used without further purification.

The following chemicals were obtained from the Sigma Chemical Company: p-nitrophenyl phosphate (sodium salt), agarose ('for electrophoresis'), ethidium bromide, SAM-iodide (grade 1), calf thymus DNA, salmon sperm DNA, nucleosides, nucleotides and ATP (disodium salt), dimethyl sulphoxide, trichloroacetonitrile, acetonitrile.

The following chemicals were obtained from British Drug Houses, Poole, Dorset, U.K.: N-tris-(hydroxymethyl) methyl-2-aminoethanesulphonic acid, EDTA, phenol, CsCl, 2-mercaptoethanol, bromophenol blue.

The dinucleoside monophosphate ApT was a kind gift from Dr. K. Murray.

Streptomycin sulphate was a gift from Glaxo Laboratories Ltd., Ulverston, Lancs., U.K.

The scintillant 5-(4-biphenyl)-2-(4-6-butylphenyl)-1-oxa-3,4-diazole (Butyl-PBD) was obtained from Ciba Ltd., Duxford, Cambs., U.K.

SAM was purified by ion-exchange chromatography on a column of Zeokarb 226 resin, using the procedure reported by Brookes (1972).

Dimethyl sulphoxide, trichloroacetonitrile, acetonitrile and triethylamine for use in the preparation of deoxynucleoside triphosphates were redistilled from, and

stored over, $CaH_2$.

Phenol was redistilled under nitrogen and stored under de-aerated water. It was used within a few days of redistillation.

The dyes used as markers on paper electrophoresis, Xylene Cyanol FF, Orange G and acid Fuschin were obtained from G.T. Gurr Ltd., London, U.K.

## 2.6 Radiochemicals

All radiochemicals were obtained from the Radiochemical Centre, Amersham, Bucks., U.K.

Essentially carrier-free $^{32}P$-orthophosphoric acid in dilute HCl solution pH 2-3, was purchased at 10 mCi/ml.

Solutions of $\gamma$-$^{32}P$-ATP of very high specific radioactivity (greater than 10 mCi/umole) were evaporated to dryness in a vacuum desiccator. The ATP was redissolved in $H_2O$ to give a 1 mM solution.

$^{3}H$-[Me]-thymidine was obtained as a solution. The specific radioactivity was 22 mCi/umole.

## 2.7 Chromatographic and ion-exchange media

The following papers were obtained from W. and R. Balston Ltd., Maidstone, Kent: Whatman Chromedia AE-81, DE-81, and P-81; and Whatman No. 1 and No. 52.

The following Whatman ion-exchange media were obtained from H. Reeve Angel and Co. Ltd., London, U.K.: P-11 phosphocellulose, DE 52 microgranular diethylamino-

ethyl (DEAE) cellulose.

Sephadex G-100 was obtained from Pharmacia Fine Chemicals AB, Uppsala, Sweden.

Pretreatment of ion-exchange media for column chromatography

P-11 phosphocellulose was de-fined, precycled and equilibrated as described by Burgess (1969). Special care was taken to ensure that this high-capacity exchanger was precisely equilibrated before use.

DE52 cellulose was suspended in $H_2O$, de-fined, and equilibrated by intermittent stirring with several changes of buffer over a period of several hours. Between buffer changes, the exchanger was recovered by filtration under reduced pressure.

Equilibration of ion-exchangers was checked by pH and conductivity measurements at each buffer change.

## 2.8 DNA preparations

(a) $^{32}P$-labelled λcI857 S7 DNA.

This was prepared bearing various host-controlled modifications. Unmodified lambda ( λ.0) was prepared from 803 ( λcI857 S7), λ.K was prepared from W1485 ( λcI857 S7) and λ.KP from W1485 ( λcI857 S7, P1).

100 ml L-broth was inoculated with 2 ml of an overnight culture of the appropriate lysogen and incubated at 34° in a shaking water bath until a cell density of

about 5 x $10^8$ cells/ml was reached. The temperature was then raised to 43° (to induce) and maintained for 15 minutes. The cells were then quickly pelleted by centrifugation (MSE 8 x 50 rotor, 10 K.r.p.m. 10 minutes) and resuspended in 50 ml low-phosphate medium containing 1-10 mCi $^{32}$P-orthophosphate. The buffering capacity of the medium was sufficient to readily neutralise the dilute HCl solution in which the orthophosphate was supplied. Incubation was continued at 37° for a further 3-4 hr. The bacteria were pelleted by centrifugation, resuspended in a convenient small volume of phage buffer and lysed with $CHCl_3$. After about 30 minutes, during which the suspension was shaken intermittently, the lysate was clarified by centrifugation. CsCl was added to 41.5% (w/w) and the phage banded by centrifugation (MSE 3 x 5 rotor, 33 K.r.p.m. for about 36 hrs.) The visible phage band was collected by puncturing the bottom of the centrifuge tube. The phage were then rebanded using the same conditions as before. The final bands were collected, pooled and stored at 0° - 4°.

DNA was prepared by phenol extraction. The phage suspension was diluted if necessary (so that the $A_{260}$ of the suspension was less than 10), dialysed against 10 mM Tris-HCl, 1 mM EDTA, pH 8.0, and then extracted three times, each with an equal volume of freshly redistilled phenol that had been equilibrated with 0.5 M Tris-HCl pH 8.0. The aqueous layer was dialysed

exhaustively against 10 mM Tris-HCl, 1 mM EDTA pH 8.0 and stored at 0°.

(b) Non-radioactive lambda DNA was prepared in a similar way, except that the induced lysogen was not pelleted and transferred to different medium; incubation was in L-broth throughout. In some cases lambda DNA was prepared from phage preparations that were kindly provided by Dr. K. Murray or Drs. P. Batten and J.P. Brockes.

$^{3}$H-labelled lambda DNA preparations were provided by Drs. K. and N.E. Murray.

(c) $^{32}$P-labelled T7.B DNA.

100 ml of low-phosphate medium was inoculated with 2 ml of an overnight culture of E. coli B and incubated at 37° until a cell density of about 5 x $10^{8}$ cells/ml was reached. The cells were then infected with phage T7.B at a multiplicity of 0.1. 2 mCi of $^{32}$P-ortho-phosphate was added 30 minutes before infection. Incubation was continued at 37° until lysis was visible (about thirty minutes after infection). The lysate was cleared by centrifugation (MSE 8 x 50 rotor, 10 K.r.p.m., 10 minutes) and the phage pelleted by high speed centrifugation (MSE 3 x 65 rotor, 23 K.r.p.m., 3 hrs). The pellet was resuspended in a convenient small volume of phage buffer and layered onto step gradients. These gradients consisted of three CsCl solutions at 1.7, 1.5 and 1.3 g/ml in phage buffer; made by layering lighter

solutions over dense ones in a MSE 23 ml centrifuge tube. Each layer had a volume of 4 ml. Phage bands were visible after centrifugation (MSE 3 x 23 rotor, 25 K.r.p.m., 4 hrs) and were collected by puncturing the bottom of the tube. The phage were then mixed with a solution of CsCl at 1.5 g/ml in phage buffer and re-banded by centrifugation (MSE 3 x 5 rotor, 35 K.r.p.m., for about 36 hrs). The bands were collected, pooled and stored at 0-4°.

DNA was prepared from the T7 phage by phenol extraction as for lambda.

(d) $^{3}$H-labelled T7.K and T7.803 DNA.

These DNAs were prepared from phages grown on C600 and 803 respectively. The T7 stock was propagated on *E. coli* B; therefore the C600 culture was infected with T7 phage that had been precycled through C600 so that they were appropriately modified. Similarly the phage used to infect the 803 culture had been precycled through 803.

These preparations were similar to the ones of $^{32}$P-labelled T7 DNA, except that the medium was L-broth and 2.5 mCi $^{3}$H-thymidine was added 30 minutes before infection.

(e) Preparations of SV40 DNA were the generous gift of Dr. L.V. Crawford.

## 2.9 $^{32}$P-labelled ribonucleotides from E. coli

100 ml of low phosphate medium containing 10 mCi $^{32}$P-orthophosphate was inoculated with 5 ml of an overnight culture of 1100. The culture was incubated at 37° for 6 hours. The cells were collected by centrifugation (MSE 8 x 50 rotor, 7 K.r.p.m., 15 minutes) and lysed by the addition of 210 µg sodium dodecylsulphate in 3 ml of 0.1 M NaCl, 0.01 M trisodium citrate. Ice-cold ethanol (6 ml) was added. The precipitate was collected by spooling onto a glass rod and extracted once with 1 M KCl (1 ml) and three times with 1.4 M NaCl (1 ml each). Between extractions debris was removed by centrifugation. The combined extracts were treated with ice cold ethanol (9 ml). The precipitate was collected by spooling, and dispersed in 1 mM NaCl (1.5 ml). Pronase (2 mg) was added and the suspension kept at 23° overnight before extraction (3 x) with neutralised phenol. The final aqueous phase was extracted (8 x) with diethyl ether and exposed to an air jet to evaporate residual ether.

This nucleic acid preparation (1.5 ml) was made 0.2 M in KOH, incubated at 37° overnight, neutralised with acetic acid and applied to a column (45 cm x 1 cm diameter) of Sephadex G-100 in 0.5 mM NaCl. Two well resolved peaks of radioactivity were eluted. Fractions comprising the later peak were pooled and evaporated to dryness. The ribonucleotides present in this material were separated by descending chromatography on Whatman

3MM paper with the following solvent system: isobutyric acid, 1 N ammonia, 0.02 M EDTA (100:60:0.8, v/v/v). The ribonucleoside monophosphates were located by autoradiography (exposure: 7 minutes) and identified by comparison with authentic markers (visible in u.v. light) run on the same chromatogram. The 2′- and 3′-isomers were not resolved. Ribonucleotides were eluted efficiently, from excised sections of the chromatogram, with $H_2O$. Ribonucleotides prepared in this way had an initial specific radioactivity of about 1.5 nCi/umole.

## 2.10 Preparation of $\alpha$-$^{32}$P-dATP

The preparation was based on the trichloroacetonitrile method of Symons (1968) for the synthesis of $^{32}$P-labelled nucleoside monophosphates. The unprotected deoxynucleoside was used (Symons, 1969). Organic reagents used in the synthesis were redistilled and stored over calcium hydride.

10-20 mCi carrier-free $^{32}$P-orthophosphoric acid was transferred to a 100 ml round bottom flask. A steel syringe needle was not used, since metal ions can contaminate the dilute acid solution. Non-radioactive $H_3PO_4$ (1 μmole) was added and the solution taken to dryness in a rotary evaporator to remove HCl. A suspension of deoxyadenosine (12 mg) in acetonitrile (0.5 ml) was added, and washed in with a further two 0.5 ml portions of acetonitrile. The mixture was dried

under gentle vacuum. Dimethyl sulphoxide (0.3 ml) was added and mixed to dissolve as much nucleoside as possible. 50 µl of a mixture of triethylamine and dimethyl sulphoxide (10% triethylamine, by volume) was added to the reaction flask, followed by 50 µl of a mixture of trichloroacetonitrile and dimethyl sulphoxide (10% trichloroacetonitrile, by volume). The contents of the flask were mixed and incubated at 37°. The solution turned a pale straw colour. After incubation for 2 hrs the solution was dried under vacuum in a rotary evaporator. $H_2O$ (1 ml) was added and a small sample (less than 1 µl) was taken for analysis by ionophoresis on AE-cellulose paper at pH 3.5. Nucleoside monophosphate yields of 60-80% relative to $P_i$ input were obtained.

About 50% of the monophosphate mixture was the 5'-isomer. This was enzymically converted to the triphosphate as follows. The reaction mixture contained, in addition to mononucleotide, in 1 ml: 74 µmoles Tris-HCl pH 8.0, 50 µmoles $MgCl_2$, 20 µmoles KCl, 3 µmoles rATP, 0.6 mg bovine serum albumin, 10 µg nucleoside monophosphate kinase, 10 µg nucleoside diphosphate kinase, 0.4 mg creatine phosphate and 20 µg creatine kinase. After incubation at 37° for 1 hr the mixture was heated in a boiling water bath for 4 minutes.

The boiled mixture was diluted with 10 ml 0.01 M ammonium bicarbonate and applied to a small column (4 cm x 1 cm diameter) of DEAE-cellulose that had been

equilibrated with 0.01 M ammonium bicarbonate. Triphosphate appeared as the final peak of radioactivity on elution with a linear gradient (160 ml) running from 0.01 M - 0.4 M ammonium bicarbonate. The pooled fractions comprising this peak were dried in a rotary evaporator and the remaining salt washed with $H_2O$. The triphosphate was taken up in $H_2O$ and the rATP from the kinase reaction destroyed by reaction with periodate as follows.

The reaction mixture (1 ml) contained: 100 µmole $NaIO_4$, 50 µmole triethylamine-HCl pH 8.5. After incubation at room temperature (21°) for 5 minutes, 0.1 ml of 1 M DL-lysine (pH 9.0) was added and incubation continued at 45° for 90 minutes. Excess periodate was destroyed by the addition of 0.2 ml of 2 M glycerol, and incubation at room temperature for 20 minutes.

The $\alpha$-$^{32}$P-triphosphate was purified from this mixture by chromatography on DEAE-cellulose under the conditions described above. Fractions containing $\alpha$-$^{32}$P-dATP were pooled, dried down in a rotary evaporator and taken up in a small volume of $H_2O$. The drying was repeated three times to remove ammonium bicarbonate. Finally the triphosphate solution was dried in a vacuum desiccator and dissolved in $H_2O$ to give a final concentration of 1mM. The overall yield of triphosphate relative to $P_i$ input was 10-20%.

## 2.11 Preparation of 5′-oligodeoxynucleotides

5′-oligonucleotides were prepared by digesting calf-thymus DNA with pancreatic DNase and fractionating them according to chain length (Tomlinson and Tener, 1963).

Calf-thymus DNA (1.2 mg) was digested with a solution (1 ml) of pancreatic DNase (50 µg/ml) in 10 mM Tris-HCl pH 7.4, 7 mM $MgCl_2$. After incubation at 37° for 3 hr the reaction mixture was heated at 100° for 5 minutes, allowed to cool and diluted with 10 ml of 10 mM Tris-HCl pH 7.8, 7 M urea. The mixture was applied to a column (32 x 1 cm diameter) of DEAE-cellulose that had been equilibrated with 10 mM Tris-HCl pH 7.8, 7 M urea. Oligonucleotides were eluted in a linear gradient (1 l) running from 0.0 - 0.3 M NaCl in that buffer. Fractions (8 ml) were collected, and their absorbance at 260 nm measured. Oligonucleotide peaks up to tetranucleotides were well resolved. Peaks of di-, tri- and tetra-nucleotides were pooled separately, and diluted with four times their original volume of 0.02 M ammonium carbonate (pH 8.6). The oligonucleotide solutions were desalted by application to columns (7 x 1 cm diameter) of DEAE-cellulose that had been equilibrated with 0.02 M ammonium carbonate. The adsorbed oligonucleotides were well washed with this buffer and eluted with a small volume of 2 M ammonium carbonate. This volatile buffer was removed by repeated evaporation to dryness (five times) in a rotary evaporator. Finally the oligo-nucleotides were taken up in 0.1 ml $H_2O$.

# Chapter 3

# METHODS

## 3.1 DNA methodology

(a) Sucrose gradients for zonal sedimentation analysis of DNA.

These were formed using a two-chambered device of the sort described by Britten and Roberts (1960). Density gradients running from 20% - 6% sucrose in 10 mM Tris-HCl pH 7.4, 0.04% sodium dodecyl sulphate had a total volume of 2.2 ml. A sample of up to 0.2 ml was layered on top of the gradient. Centrifugation was at 50 K.r.p.m. in an MSE 3 x 3 rotor for 2 hr at 20°. The gradient was fractionated by puncturing the bottom of the tube with a hollow needle. Three-drop fractions were collected directly onto 2.1 cm Whatman GFC discs, dried, and assayed for radioactivity in a liquid scintillation spectrometer after the addition of 2.5 ml of toluene based scintillant (4 g Butyl-PBD/litre AnalaR toluene).

(b) Gel electrophoresis of DNA.

Electrophoresis was performed in a 1% agarose gel cast in a space (0.4 cm x 17 cm wide x 40 cm long) between two glass plates. The plates were separated by narrow perspex strips at the sides and a perspex slot former giving 1 cm gel slots at the bottom. The gel was cast in this inverted position, the joints in the apparatus

being sealed by coating their surfaces lightly with paraffin wax. The whole apparatus was held together with spring clips.

The gel was prepared by refluxing agarose (2.5 g) with electrophoresis buffer (250 ml; 40 mM Tris-HCl, 20 mM sodium acetate, 1 mM EDTA, adjusted to pH 8.2 with acetic acid) until a clear solution was obtained, allowing this to cool to about 50° and then casting the gel. A layer (about 3 cm deep) of 3% agarose was poured over the 1% gel to hold the slab in place when the apparatus was turned to the running position.

Samples were prepared by mixing them with 10 µl of electrophoresis buffer containing 50% glycerol, 0.04% bromophenol blue. If the resulting volume was greater than 20 µl it was reduced by evaporation in a vacuum desiccator. Samples were then carefully layered into the gel slots from a capillary. Electrical contact between the upper buffer reservoir and the buffer over the gel slots was by six layers of Whatman 3MM paper soaked in electrophoresis buffer; the bottom of the gel apparatus dipped into the lower buffer reservoir. Gels were run at 40 mA, usually for 14 hr.

(i) $^{32}$P-labelled DNA. After electrophoresis the gel was removed from the apparatus and sheet of Whatman No. 1 paper placed over it. Water was withdrawn from the gel by suction through a sintered plastic slab, initially at room temperature, later at 100°. The gel

dried down to a thin layer that adhered to the paper and was placed in contact with X-ray film for autoradiography.

(ii) Non-radioactive DNA. Duplex DNA was made visible by running the gel in the presence of ethidium bromide. Ethidium bromide was added to the refluxed agarose solution just before the gel was cast, and also to the electrophoresis buffer in the reservoirs. The concentration was 0.5 mg/l throughout (Sharp et al., 1973). After electrophoresis the gel was removed from the apparatus and viewed under UV light (ChromatoVue, UV Products Inc., San Gabriel, California, U.S.A.). DNA bands fluoresced orange against a dark background and were photographed on panchromatic film through a red filter. As little as 0.05 μg DNA in a single band was readily detectable with this technique.

(c) Electron microscopy of DNA. The method of spreading with cytochrome c was essentially that described by Lang and Mitani (1970).

The following solution was freshly prepared by mixing stock solutions: volume 1 ml, 0.15 M ammonium acetate, 3.5 mM formaldehyde, 1.4 ng cytochrome c. The DNA (10 μl of a solution at 10-50 μg/ml) was added and samples (40 μl) were taken and placed as drops on a Teflon block. Samples were usually taken in duplicate. After 15 minutes a carbon-coated grid was touched to the surface of the drop, then placed in the surface of ethanol for 10 seconds

and finally drained with filter paper. To examine hydrogen-bonded circles, this procedure was performed in a cold room, with all solutions equilibrated at 5°.

Prepared grids were shadowed by Pt-evaporation in a Balzers vacuum evaporator. The samples were rotated rapidly during shadowing.

Grids were examined and photographed in a Siemens 1A electron microscope. The photographic plate was placed in a projector and DNA length measurements were made with a map measurer on tracings of the enlarged image.

## 3.2 Oligonucleotide methodology

### (a) Ionophoresis and radioautography

Details of these were as described by Sanger et al. (1965) and Murray (1970). The following buffer systems were used in paper ionophoresis:

pH 2, 2.5% formic acid - 8.7% acetic acid (v/v);

pH 3.5, 0.5% pyridine - 5% acetic acid (v/v);

pH 9.7, 5% triethylamine carbonate.

When triethylamine carbonate buffer was used it was necessary to add triethylamine to the white spirit used as coolant in the tanks (0.5 - 1.0% v/v).

Radioautographs were made on Kodak Blue Brand X-ray film.

(b) Analyses of oligonucleotides

Details of the micro techniques used for these analyses, and for the elution of nucleotides from the ion-exchange papers with 30% (v/v) triethylamine carbonate solution were as described by Sanger et al. (1965).

(i) Partial and complete digestions with venom phosphodiesterase.

Partial digests for the determination of the sequence of 5′-terminally labelled oligonucleotides were made by incubation of the nucleotide in a solution (10 µl) of the enzyme (0.01 mg/ml) in 0.02 M Tris-HCl pH 8.5 at room temperature (20°). Samples were taken at zero time, 15 minutes and 1 hr for ionophoresis on DEAE-cellulose paper at pH 2 and AE-cellulose paper at pH 3.5.

Complete digests were made with an enzyme concentration of 0.1 mg/ml and incubation at 37° for 2 hr. Complete digests were analysed by ionophoresis on AE-cellulose paper at pH 3.5, because in this system the four 5′-mononucleotides are quite satisfactorily resolved when the blue dye (Xylene Cyanol FF) has migrated 25 cm from the origin. Nucleotides were detected by radioautography, and, where necessary, quantitated by counting with a low-background, anticoincidence, gas-flow counter (Nuclear Chicago Ltd., High Wycombe, Bucks., U.K.). When very small quantities of radioactivity were involved,

reference markers (10 mg/ml; 10 µl) were run coincidently with the samples and located by viewing under u.v. light.

(ii) Complete digestion with spleen phosphodiesterase.

Oligonucleotides from a pancreatic DNase digest of DNA, labelled at its 3′-terminus in exonuclease-repair reactions with T4 DNA polymerase, were first dephosphorylated before treatment with spleen phosphodiesterase.

Oligonucleotides were dephosphorylated by incubation in a solution (10 µl) of alkaline phosphatase (0.5 mg/ml) in 10 mM Tris-HCl pH 8.0 at room temperature (22°) for 90 minutes. After incubation the sample was applied to a small pennant of P-81 phosphocellulose paper that had been prewashed with $H_2O$. The dephosphorylated oligonucleotide was eluted with $H_2O$. Phosphatase was not eluted.

Complete spleen phosphodiesterase digests were made in a solution (20 µl) of enzyme (1.5 units/ml) in 0.05 M ammonium acetate pH 5.6, 1 mM EDTA at 37° for 1 hr. Some complete spleen digests were analysed by ionophoresis on AE-cellulose paper at pH 3.5, however this system does not resolve adenosine 3′-monophosphate and guanosine 3′-monophosphate; subsequently ionophoresis on Whatman No. 52 paper at pH 3.5 was used since on this system all four 3′-mononucleotides are well resolved when the blue dye (Xylene Cyanol FF) has migrated 10 cm. Nucleotides were detected by radioautography and

quantitated in the same way as the 5′-mononucleotides from complete venom phosphodiesterase digests.

## 3.3 Preparation of T4 DNA polymerase

The source of this enzyme was 1100 PolA that had been infected with T4 am N82. This strain was chosen because it is endo $I^-$. This is useful in the preparation of T4 DNA polymerase for use with a defined high molecular weight primer-template because endonuclease I partially co-purifies with the polymerase (Goulian et al., 1968). The PolA mutation ensures that the polymerase activity is predominantly due to T4 infection; if infection is incomplete E. coli polymerase I could be a complication in the early stages of purification (see Table 3.1). T4 am N82 stocks were prepared on the permissive host CR63.

The purification procedure of Goulian et al. (1968) requires alteration since with the endonuclease $I^-$ strain an autolysis step is precluded. The following procedure was used. Enzyme assays were performed according to Goulian et al. (1968).

64 g of infected cells (multiplicity of infection: 4) were partially thawed, suspended in 20 ml buffer (0.05 M glycylglycine pH 7.0, 2 mM EDTA, 10 mM 2-mercaptoethanol) and blended in a Waring blender with 200 g acid-washed glass beads for 10 minutes at high speed. The cell

Table 3.1 DNA polymerase activity in extracts of 1100, 1100 PolA, and T4 am N82-infected 1100 PolA

Two 500 ml cultures of 1100 PolA and one 500 ml culture of 1100 were grown up in L-broth at 37° to $2.10^8$ cells/ml. One of the 1100 PolA cultures was infected with T4 am N82 phage (multiplicity of infection:10) and incubated for a further 75 minutes before the cells were harvested by centrifugation. The other two cultures were harvested immediately. Cell extracts were prepared by sonication in buffer (0.05 M Tris-HCl pH 7.4, 0.01 M 2-mercaptoethanol), and centrifugation to remove debris (MSE 6 x 100 rotor, 18 K.r.p.m., 30 minutes). DNA polymerase was assayed according to Goulian et al. (1968).* Protein was assayed by the method of Lowry et al. (1951).

| Source of Extract | DNA polymerase units/mg protein |
|---|---|
| 1100 | 5.0 |
| 1100 Pol A | 0.1 |
| 1100 Pol A infected with T4 am N82 | 14.6 |

* These conditions are for T4 DNA polymerase, but are similar to those for E. coli polymerase I (Jovin et al., 1969).

debris was removed from the decanted supernatant by low speed centrifugation. To the cleared extract (260 ml) 30 g solid ammonium sulphate was added slowly. The precipitate that formed after thirty minutes was removed by centrifugation (MSE 6 x 250 rotor, 10 K.r.p.m., 15 minutes). A further 50 g of ammonium sulphate was added to the supernatant. The precipitate was collected by centrifugation, resuspended in 0.05 M potassium phosphate pH 6.5, 10 mM 2-mercaptoethanol and dialysed against the same buffer before being applied to a phosphocellulose column (15 cm x 3.6 cm diameter) equilibrated with the same buffer. Protein was eluted in a linear concentration gradient of about 1 litre total volume running from 0.05 M to 0.4 M potassium phosphate buffer pH 6.5 containing 10 mM 2-mercaptoethanol.

Active fractions appeared after the passage of about 700 ml of the gradient, were pooled, dialysed against 0.02 M potassium phosphate pH 7.4, 10 mM 2-mercaptoethanol and applied to a DEAE-cellulose column equilibrated with the same buffer. This was followed by a linear gradient (500 ml total volume) having limits of 0.02 M and 0.15 M potassium phosphate pH 7.4, containing 10 mM 2-mercaptoethanol. Active fractions appeared after the passage of about 275 ml of the gradient. They were pooled and concentrated by pressure ultrafiltration (Amicon PM10 membrane). The product was dialysed against 0.1 M potassium phosphate pH 7.0, 10 mM 2-mercapto-

ethanol, 50% glycerol. This final fraction had a specific activity of 11,300 units/mg (2000 units/ml) and was stored at $-10^{o}$.

## 3.4 Preparation of nucleotide phosphotransferase from E. coli 1100

Standard assay: The assay measures the amount of uridylic acid produced from uridine with p-nitrophenyl phosphate as phosphate donor. The procedure is adapted from Brunngraber and Chargaff (1970).

The assay reaction mixture (0.4 ml) was as follows: p-nitrophenyl phosphate 100 mM, uridine 20 mM, sodium acetate (pH 5.1) 100 mM, enzyme at a concentration such that less than 25% of the donor was hydrolysed in 1 hr. After incubation at $37^{o}$ for 1 hr a portion (50 µl) was withdrawn into ethanol (20 µl) to stop the reaction and chromatographed (descending) on Whatman No. 1 paper in the following solvent system.

1-propanol, .880 ammonia, $H_2O$ (11:7:2 v/v/v). Uridylic acid ($R_F$ 0.20) was located under UV light, eluted in 0.1 N HCl and measured spectrophotometrically at 260 mu.

Elution was found to give quantitative recovery. At pH 2, uridylic acid, $a_M$ (260) = $9.9 \times 10^9$.

Thymidine 3'-monophosphate ($R_F$ 0.23) was eluted and estimated by the same procedure in assays where thymine

replaced uridine as phosphate acceptor. At pH 2, thymidylic acid $a_M$ (260) = 10 x $10^3$.

The specific activity is expressed as μmoles of uridylic acid produced in 1 hr by 1 mg protein in 1 ml of assay mixture. Protein content of enzyme solutions was determined by the method of Lowry et al. (1951) with bovine serum albumin as standard.

Enzyme purification

Nucleotide phosphotransferase was prepared from E. coli 1100 according to the procedure of Brunngraber and Chargaff (1970). Grinding of the cells with alumina was replaced by sonication for a total of 3 minutes with cooling in an ice bath. The final enzyme fraction (step III, from DEAE-cellulose step) had a specific activity of 5.4.

This preparation, like that of Brunngraber and Chargaff, was contaminated with at least one phosphatase capable of hydrolysing p-nitrophenyl phosphate. (Phosphatase was assayed by omitting uridine from the assay described above. The amount of p-nitrophenol that was produced was measured spectrophotometrically (425 mu) after the addition of $Na_2CO_3$ to develop the yellow (alkaline) colour.)

The phosphotransferase was further purified by chomatography on phosphocellulose as follows. Step III phosphotransferase preparation (36 ml) was dialysed

against 0.001 M sodium acetate pH 6.0 and applied to a phosphocellulose column (8 cm x 1 cm diameter) that had been previously equilibrated with the same buffer. Phosphotransferase activity was eluted mid-way in a linear gradient (total volume 200 ml) running from 0.001 M to 0.4 M sodium acetate pH 6.0. Pooled active fractions were dialysed against 0.001 M sodium acetate pH 6.0, and had a final specific activity of 9.6. The ratio of phosphotransferase:phosphatase was not significantly improved, however, by this additional chromatography step.

## 3.5 Preparation of $R.Eco_{P1}$

### Assay

The assay reaction mixture (0.1 ml) contained: 50 mM N-tris (hydroxymethyl)methyl-2-aminoethane-sulphonate pH 8.0, 0.25 mM EDTA, 6 mM $MgCl_2$, 13 mM 2-mercaptoethanol, about 0.1 $\mu$g $^{32}P$-labelled $\lambda$.KP DNA (several thousand cpm) and a similar amount of $^{3}H$-labelled $\lambda$.0 or $\lambda$.K DNA, 1 mM ATP, 0.03 mM SAM with the enzyme sample as final addition. (The lambda DNA was previously deconcatenated by heating at 65° for 3 minutes.) Incubation was at 30° for 20 minutes. The reaction was stopped by the addition of EDTA (20 $\mu$l; 0.2 M). Finally, the whole reaction mixture was layered onto neutral sucrose density gradients for zone sedimentation analysis.

Active enzyme fractions degraded unmodified DNA to fragments that were resolved from (intact) modified DNA.

Enzyme purification

The enzyme was prepared essentially according to the procedure of Meselson and Yuan (1968). P1-specific restriction activity was readily detectable even in crude extracts of *E. coli* 1100$r_K$-(P1).

The enzyme was eluted from the DEAE-cellulose column at a phosphate concentration of 0.09 M, and from the phosphocellulose column at 0.18 M. The glycerol gradient fraction was titrated to find the minimum amount necessary to fully restrict a given amount of DNA.

In some experiments described in this thesis a preparation of R.$Eco_{P1}$ was used that was the gift of L. Morss. It had been prepared by the procedure just described.

## 3.6 Preparation of R.$Eco_{RI}$

The enzyme was prepared from *E. coli* B (RI) according to a procedure adapted from Yoshimori (1971). All procedures were performed at low temperature (about 4°).

Frozen cells (74 g) were suspended in 200 ml of extraction buffer (10 mM potassium phosphate pH 7.0, 7 mM 2-mercaptoethanol, 1 mM EDTA) and sonicated with a Dawes sonicator at maximum setting for a total of

8 minutes with cooling in an ice-bath. A further 100 ml of buffer was added and cell debris removed by centrifugation (MSE 6 x 100 rotor, 16 K.r.p.m., 20 minutes). The supernatant volume was 335 ml, and to this was added slowly 100 ml of 5% streptomycin sulphate in extraction buffer. The precipitate which had formed after 30 minutes was removed by centrifugation (MSE 6 x 250 rotor, 10 K.r.p.m., 10 minutes). To the supernatant (400 ml) was added slowly 400 ml of a saturated ($5^{o}$) ammonium sulphate in extraction buffer. The precipitate that had formed after 30 minutes was collected by centrifugation (MSE 6 x 250 rotor, 10 K.r.p.m., 10 minutes), resuspended in extraction buffer and dialysed against extraction buffer. The volume after dialysis was 150 ml. Solid NaCl was added to a final concentration of 0.35 M.

The solution was applied to a phosphocellulose column (15 cm x 2.5 cm diameter) that had been equilibrated with a solution of 0.35 M NaCl in extraction buffer. Elution was with a linear gradient (700 ml) running from 0.35 M to 0.8 M NaCl in extraction buffer, at a flow rate of 40 ml/hr. Sixty-three fractions (approx. 12 ml) were collected.

Fractions were assayed by a gel electrophoresis. Reactions (15 $\mu$l) were performed in capillaries and contained the following: 50 mM NaCl, 40 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 2.4 ug $\lambda$.0 DNA and 5 $\mu$l of column fraction. Incubation was at $37^{o}$ for

30 minutes when the reaction was stopped with 0.2 M EDTA (5 µl). The DNA product was analysed by electrophoresis in 1% agarose gels containing ethidium bromide. Just one activity was found that gave specific fragmentation of the λ DNA. It gave the 6 fragments corresponding to $R.Eco_{RI}$ (Allet et al., 1973), and peaked at fractions 24-26. Active fractions were pooled, dialysed against extraction buffer (pH 6.5), and applied (at a flow rate of 40 ml/hr) to a column of DEAE-cellulose (5 cm x 1 cm diameter) equilibrated with extraction buffer at pH 6.5. The restriction activity was eluted with 5 ml of 0.5 M NaCl in extraction buffer at pH 7.0. The solution was dialysed against extraction buffer containing 30% glycerol and stored at $0^{o}$. The enzyme was titrated against lambda DNA to determine the amount of enzyme required for the reaction to go to completion. When a considerable excess of enzyme (twenty times that necessary) was used the gel assay showed that further fragmentation of λ.O DNA occurred. Whether this is due to another restriction activity or whether there can be attack of 'secondary' sites by enzyme in large excess is not known. The enzyme requires $Mg^{2++}$, but no other cofactors, for activity (Yoshimori, 1971).

## 3.7 Preparation of R.Hin$_d$ III

This enzyme was prepared from H. influenzae Rd obtained from Dr. S.W. Glover. The initial preparation of the extract was performed by G. Roizes, following the procedure of Smith and Wilcox (1970) for the preparation of R.H. The activity did not affect H. influenzae Rd DNA (G. Roizes, personal communication). However it was apparent that R.H and this preparation were not identical. This will be considered in the discussion. The experiments described in this thesis were performed with R.Hin$_d$ III prepared as follows.

An extract of H. influenzae Rd purified to the ammonium sulphate step was the generous gift of G. Roizes. This material was diluted ten times to 50 ml with 0.01 M potassium phosphate pH 7.4 and applied to a phosphocellulose column (10 cm x 1 cm diameter) equilibrated with the same buffer. Protein was eluted stepwise with 10 ml portions of 0.01 M potassium phosphate pH 7.4 containing increasing molarities of KCl as follows: 0.0, 0.1, 0.2, 0.3 and 0.4 M. Fractions of 2 ml were collected and assayed as follows. Reactions (20 µl) were performed in capillaries and contained 50 mM NaCl, 40 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 0.6 ug $^{32}P$ λ.C DNA and 5 µl of column fraction. Incubations were at 37° for 30 minutes. The reactions were stopped with 5 µl 0.2 M EDTA and the DNA product

analysed by electrophoresis in 1% agarose gels that finally were dried and autoradiographed.

The bulk of the endonuclease activity that gave rise to discrete DNA bands on gel electrophoresis, was eluted in the 0.4 M step. Endonuclease was present in the previous fractions but it did not give rise to DNA bands. This may have been because of inherent non-specifity of the endonuclease or contamination with exonuclease.

Active fractions at the 0.4 M step were pooled, dialysed against 0.2 M NaCl, 0.02 M Tris-HCl pH 7.4, 10 mM 2-mercaptoethanol in a solution of glycerol (50%), and stored at $-10^{o}$.

# Chapter 4

## NUCLEOTIDE PHOSPHOTRANSFERASE AND 3′-TERMINAL LABELLING OF DNA

### 4.1 Phosphate group donors

Brunngraber and Chargaff (1970) have reported details of the phosphotransferase activity including the pH dependence and activity towards deoxyribo and ribo-mononucleotides. In their studies p-nitrophenyl phosphate was phosphate group donor. Para-nitrophenyl phosphate was more efficient than 3′-riboadenylic acid. The latter substance is also a product of phosphorylation of r-adenosine by phosphotransferase with p-nitrophenyl phosphate as donor.

Table 4.1 shows the results obtained when a number of other compounds were tested as phosphate group donors. No new effective phosphate donors were found.

### 4.2 Phosphate group acceptor studies

#### (a) Time-course of thymidine phosphorylation

The time-course of phosphorylation of thymidine, giving thymidine 3′-monophosphate, was investigated by incubating thymidine with enzyme under standard assay conditions, as described in the legend to Figure 4.1. Under these conditions the enzyme was fully active after incubation for 2 hr. Subsequent experiments with the

Table 4.1 Phosphate group donors

Each potential phosphate donor was incubated, at the concentrations indicated, with 0.12 unit phosphotransferase under standard assay conditions. After incubation at 37° for 1 hr a sample (50 µl) of each reaction was withdrawn for chromatography and spectrophotometric determination of uridylic acid as described in Methods.

| Donor | Concentration of donor (mM) | m µmoles uridylic acid produced |
|---|---|---|
| p-nitrophenyl phosphate | 100 | 122 |
| " | 20 | 98 |
| phenyl phosphate | 100 | < 5 |
| " | 20 | " |
| pyrophosphate | 50 | " |
| " | 5 | " |
| rATP | 20 | " |
| " | 5 | " |
| dATP | 20 | " |
| orthophosphate + ATP | 50, 5 | " |

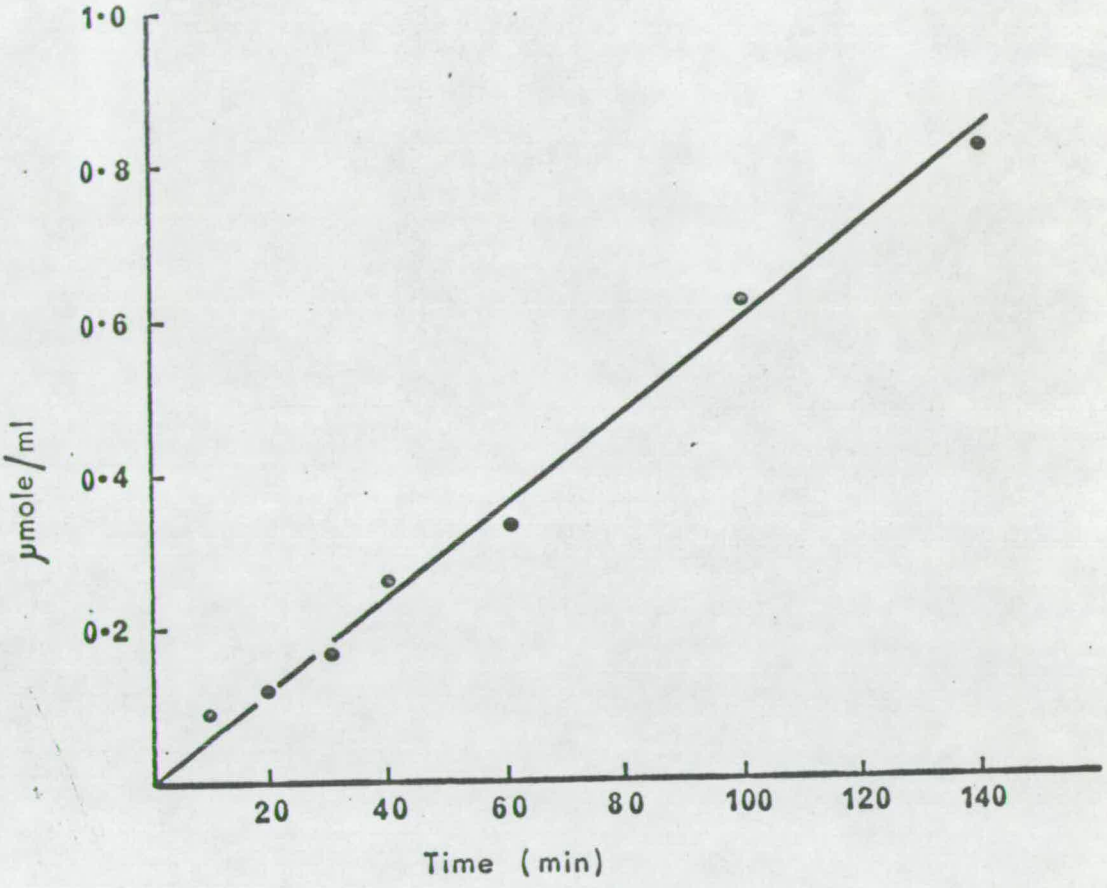



Figure 4.1. Time-course of phosphorylation of thymidine. The reaction mixture (0.8 ml) contained p-nitrophenyl phosphate 100 mM, thymidine 20 mM, sodium acetate (pH 5.1) 100 mM and nucleotide phosphotransferase. During the course of incubation at 37° samples (50 µl) were withdrawn for spectrophotometric determination of the concentration of thymidine 3′-monophosphate in the reaction mixture (Methods).

phosphotransferase were performed using incubation for 2 hr at 37°.

(b) Phosphorylation of dinucleoside monophosphates

The ability of the phosphotransferase to phosphorylate dinuceloside monophosphates was investigated with two ribodinucleoside monophosphates (r-CpU and r-UpC) and one deoxyribodinucleoside monophosphate (ApT).

Each dinucleoside monophosphate (20 µg) was incubated with 2′-(3′-)-$^{32}$P-riboadenylic acid (10,600 cpm, prepared as in Methods) in a solution (20 µl) of enzyme (0.09 units/ml) in 0.1 M sodium acetate buffer pH 5.1. After incubation at 37° for 2 hr the radioactive products were analysed by ionophoresis on AE-cellulose paper at pH 3.5 and autoradiography.

Putative r-CpUp, r-UpCp and ApTp spots (mobilities relative to Xylene Cyanol FF: 0.74, 0.78, 0.61) were eluted and partially digested with spleen phosphodiesterase. The partial digests were analysed by ionophoresis on AE-cellulose paper at pH 3.5. These analyses (Fig. 4.2) confirmed that the dinucleoside monophosphates were phosphorylated at their 2′- or 3′-hydroxyl group, and in the case of ApT this must have been at the 3′-hydroxyl group.

(c) Phosphorylation of 5′-oligonucleotides

The abilityof the phosphotransferase to phosphorylate

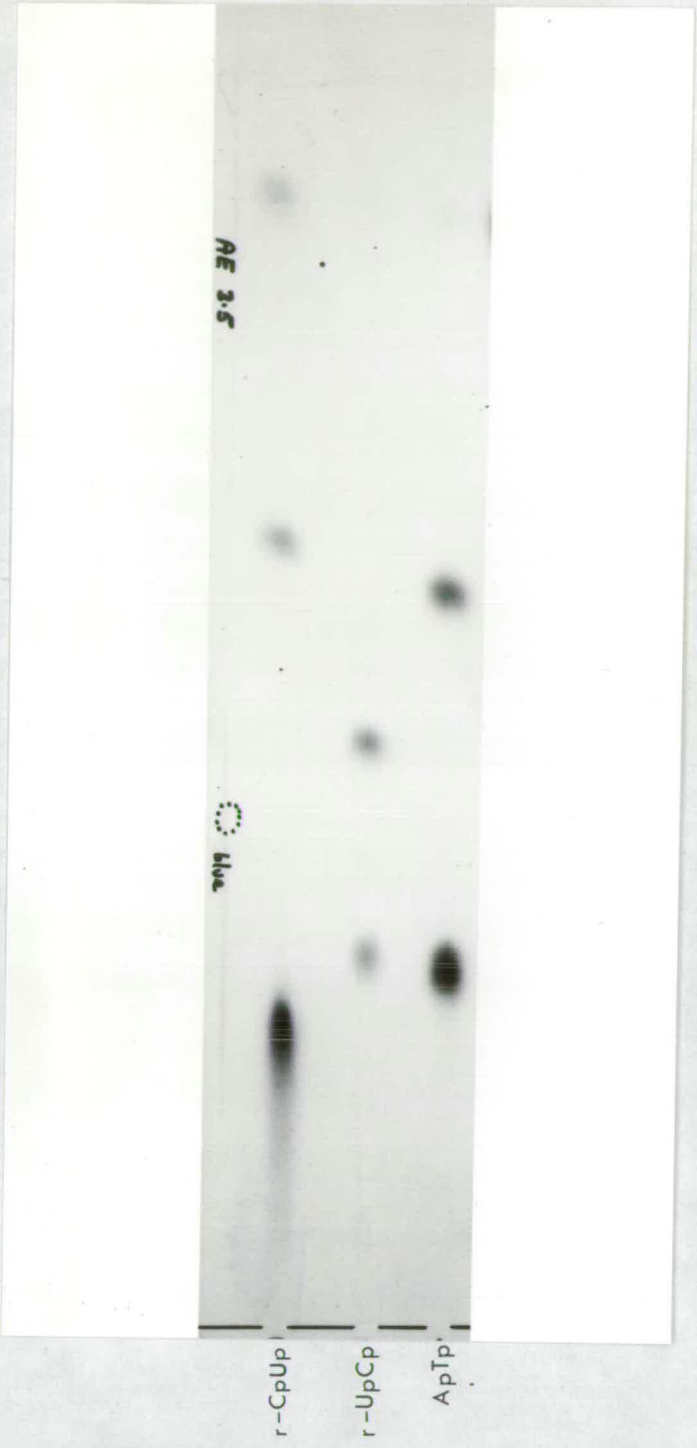



Figure 4.2. Partial digestion of r-CpUp, r-UpCp and ApTp with spleen phosphodiesterase. The nucleotides were partially digested with spleen phosphodiesterase (0.5 units/ml) in 0.1 M ammonium acetate pH 5.6, 1 mM EDTA at $37^{o}$ for 30 minutes. The partial digests were analysed by ionophoresis on AE-cellulose paper at pH 3.5. The position of the blue dye is indicated.

Figure 4.3. Oligonucleotides of defined chain-length.

Solutions of 5′-dinucleotides, etc. (20-30 nmoles) prepared as described in the text, were added to 0.1 ml of 10 mM Tris-HCl pH 8.0 containing alkaline phosphatase (50 µg/ml) and incubated at 37° for 2 hr. Alkaline phosphatase was inactivated by adding N HCl to pH 2, in the presence of 1 mM EDTA. After 10 minutes at 23° the solution was brought to about pH 7.4 with Tris base. $MgCl_2$ and 2-mercaptoethanol were added to give final concentrations of 10 mM and 15 mM, followed by $\gamma$-$^{32}$P-ATP ($4 \times 10^5$ cpm: sp. radioactivity about 1 mCi/µmole) and 10 µl of a preparation of polynucleotide kinase. After incubation at 37° for 1 hr the mixture was concentrated by evaporation and applied to AE-81 paper for analysis at pH 3.5. (a) dinucleotides (b) trinucleotides (c) tetranucleotides. The position of the blue dye is indicated.

Figure 4.4. 5′-oligonucleotides are not a substrate for phosphotransferase.

5′-dinucleotides, etc. (20-30 n moles) were incubated with 2′-(3′-)$^{32}$P-riboadenylic acid ($3.6 \times 10^4$ cpm; prepared as in Materials) in a solution (30 µl) of enzyme (0.09 units/ml) in 0.1 M sodium acetate pH 5.1, at 37° for 2 hr. Radioactive products were analysed by ionophoresis on AE-81 paper at pH 3.5.

Samples: (1) 5′-dinucleotides + thymidine (2 n moles).
(2) 5′-trinucleotides.
(3) 5′-tetranucleotides.

The position of the blue dye is indicated.


a b c


1 2 3

5′-dinucleotides, 5′-trinucleotides and 5′-tetranucleotides was investigated.

The 5′-oligonucleotides were isolated from a pancreatic DNase digest of calf-thymus DNA by fractionation according to chain-length upon chromatography on DEAE-cellulose in the presence of 7 M urea (see Materials).

The preparations of 5′-dinucleotides, etc., were checked by 5′-terminal labelling in the polynucleotide kinase reaction after dephosphorylation. Figure 4.3 shows the results of the analysis of the di-, tri- and tetranucleotide preparations. Mononucleotides would not be detected by this procedure because a nucleoside is not a substrate for polynucleotide kinase (Novogrodsky and Hurwitz, 1965).

5′-di-, tri- and tetra-nucleotides were incubated with phosphotransferase as described in the legend to Figure 4.4. The autoradiograph shows that 5′-oligonucleotides are not phosphorylated even when thymidine, present as an internal control, is phosphorylated. The oligonucleotides do not inhibit phosphorylation. Thus deoxynucleotides larger than a dinucleoside monophosphate cannot be labelled at their 3′-hydroxyl group in a reaction catalysed by the phosphotransferase.

(d) Phosphorylation of DNA and RNA

The results of the previous section show that 5′-oligonucleotides are not phosphorylated by the

Table 4.2 DNA and RNA are not substrates for phosphotransferase

0.1 ml of DNA solution ($A_{260}$ 1.3), prepared as described in the text, was incubated in a solution (0.3 ml) containing phosphotransferase (0.09 units/ml) and 2′-(3′)-$^{32}$P-riboadenylic acid ($6.2 \times 10^5$ cpm) in 0.1 M sodium acetate pH 5.1. A similar reaction was performed in which 10 µg sRNA (Sigma: from yeast) replaced DNA. Blank reactions were performed omitting enzyme. After incubation at 37° for 2 hr, 0.1 ml salmon sperm DNA (1 mg/ml) was added followed immediately by 0.4 ml of a cold solution of perchloric acid (6%). The precipitate was resuspended in 1 M Tris pH 8.0 (1 ml) and repricipitated with perchloric acid solution (1 ml). The precipitate was filtered onto Whatman GFC discs, dried, and assayed for radioactivity in a liquid scintillation spectrometer (see Methods).

| nucleic acid | phospho-transferase | cpm |
|---|---|---|
| DNA | + | 36 |
| DNA | - | 38 |
| RNA | + | 27 |
| RNA | - | 22 |

phosphotransferase. The experiments described in this section were designed to test whether sRNA and native DNA can be phosphorylated by the enzyme.

Salmon sperm DNA (0.1 mg/ml) was incubated with a low concentration of pancreatic DNase (1 μg/ml) in 10 mM Tris-HCl pH 7.4, 5 mM $MgCl_2$. Incubation was continued at room temperature (20°) until about 5% of the DNA was made acid-soluble (27 minutes) when the reaction was stopped by the addition of EDTA to a final concentration of 10 mM. After extraction with neutralised phenol the product was dialysed against 10 mM Tris-HCl pH 8.0.

DNA prepared in this way was incubated with phosphotransferase as described in the legend to Table 4.2. A reaction was also carried out with sRNA (Sigma: yeast sRNA) in place of DNA. The results (Table 3) show that DNA and RNA are not substrates in the phosphorylation reaction. The DNA preparation may have contained some small oligonucleotides but the results of the previous section show that these oligonucleotides neither compete with nor inhibit the reaction.

The enzyme does not form the basis of a 3′-terminal labelling method for DNA (or RNA) sequence analysis.

# Chapter 5

## DNA NUCLEOTIDE SEQUENCE RECOGNISED BY THE TER ENDONUCLEASE OF BACTERIOPHAGE LAMBDA

The exonuclease-repair reactions catalysed by T4 DNA polymerase that have been discussed in the Introduction can be applied to phage lambda DNA in an analysis of its 3′-terminal nucleotide sequences. These nucleotides lie beyond the nicks introduced by the Ter endonuclease. The sequence of bases between the two nicks is known (Wu and Taylor, 1971).

### 5.1 Endonuclease digestion of labelled DNA

Lambda DNA was terminally labelled by incubation with T4 DNA polymerase in the presence of $\alpha$-$^{32}$P-dATP (specific radioactivity: about 10 mCi/μmole). These reactions were performed by Dr. P. Englund (Johns Hopkins Univ., Baltimore, U.S.A.). After labelling intact single strands were purified by zone sedimentation in alkaline sucrose gradients. Labelled DNA was transported from Baltimore to Edinburgh in evacuted, sealed tubes.

DNA solutions were concentrated to 1 ml by evaporation in a rotary evaporator, dialysed against 0.3 mM NaCl and further concentrated to 0.1 ml in a vacuum desiccator. The labelled DNA was then digested with pancreatic DNase. Digestion of the denatured DNA

required high concentrations of endonuclease and incubation for long periods. Conditions were chosen from trial digests analysed by ionophoresis on AE-cellulose paper at pH 3.5. The bulk of the material was digested using 0.5 mg/ml of DNase in 20 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$ at 37° for 22 hr. This gave a satisfactory distribution of radioactive oligonucleotides. A sample of the labelled material was digested with endonuclease I from E. coli.

## 5.2 Fractionation and analysis of oligonucleotides

Oligonucleotides in the pancreatic DNase digests were fractionated by ionophoresis on AE-cellulose paper at pH 3.5 and on DE-cellulose paper at pH 2.0. Two-dimensional ionophoresis on DE-cellulose paper at pH 9.7 followed by DE-cellulose paper at pH 2.0 was used for the endonuclease I digest. Labelled oligonucleotides (Fig. 5.1) were located by autoradiography and eluted.

The mobility of each $^{32}P$-oligonucleotide was determined on both ionophoresis systems. Oligonucleotides isolated by ionophoresis on AE-cellulose paper were examined on DE-cellulose paper and vice versa. The 3′-penultimate base of each oligonucleotide was determined, after dephosphorylation with alkaline phosphatase, by digestion with spleen phosphodiesterase and ionophoretic analysis of the 3′-mononucleotides (as described in Methods).

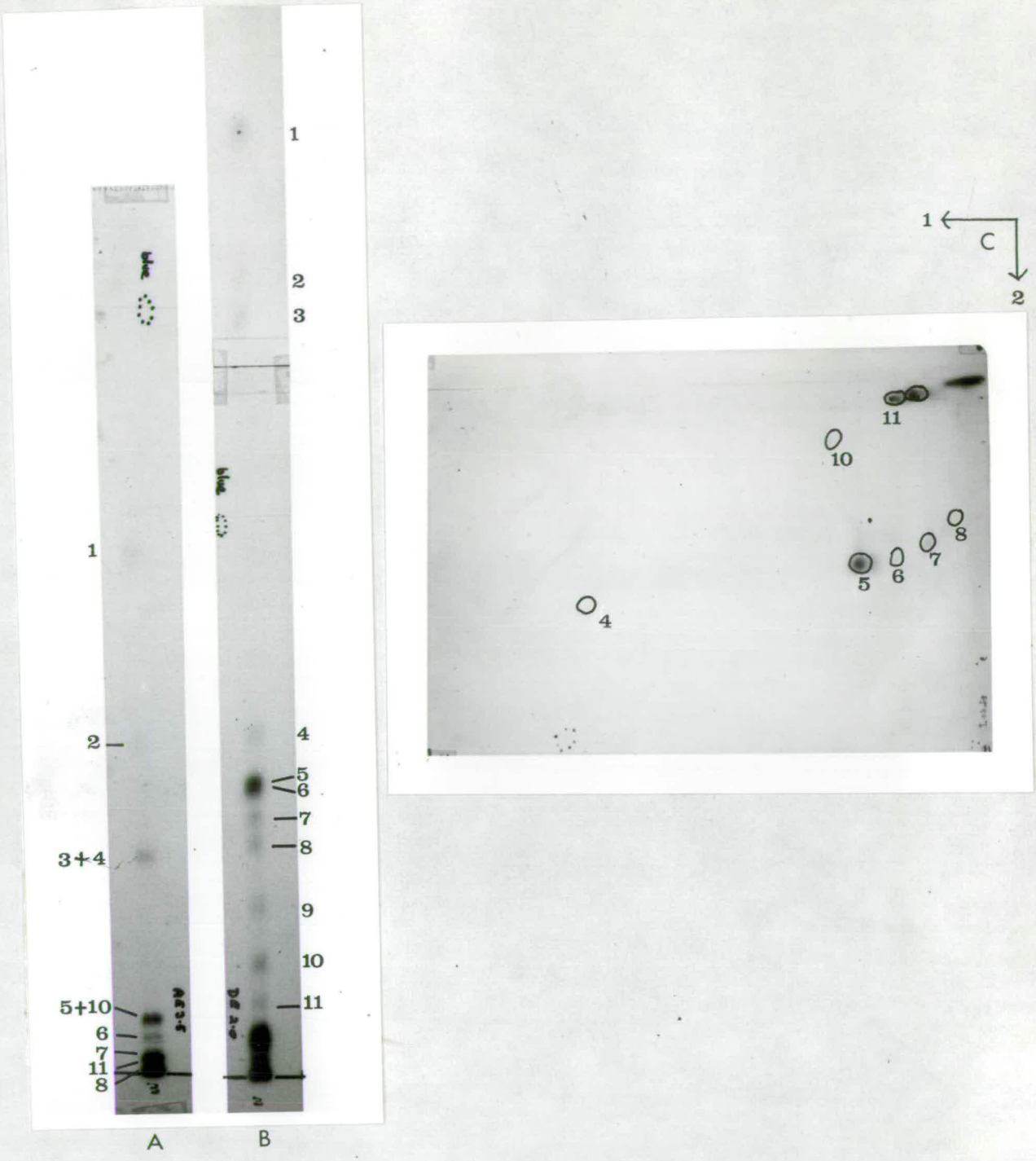



Figure 5.1. Radioactive oligonucleotides in a pancreatic DNase digest of lambda DNA terminally labelled by incubation with T4 DNA polymerase and $\alpha$-$^{32}$P-dATP. Autoradiographs of oligonucleotides after ionophoresis on (A) AE-cellulose paper at pH 3.5 (B) DE-cellulose paper at pH 2 (C) DE-cellulose paper at pH 9.7 in the first dimension, followed by DE-cellulose paper at pH 2 in the second dimension.

## 5.3 Nucleotide sequences

The strategy for 3′-terminal sequence analysis of the labelled DNA was discussed in the Introduction. We recall that if 3′-terminally labelled DNA is degraded with a non-specific endonuclease, all the labelled oligonucleotides in the digest must belong to one of two families. Each family arises from one of the strands and consists of a homologous series of oligonucleotides whose sequences overlap progressively from the 3′-terminus.

All the oligonucleotides in these digests necessarily have A as the 3′-terminal base. Each oligonucleotide may be assigned to its family by identification of its 3′-penultimate base (either T or G). Once all oligonucleotides have been thus assigned their sequences can be determined from the difference in mobility, on electrophoresis, between adjacent members of each homologous series. The identities of oligonucleotides given in Table 5.1 were deduced in this way.

These results reveal the sequence -G-T-T-A on the l-strand and -A-C-C-C-G-C-G-A on the r-strand of lambda DNA. Inspection of the cohesive end sequences (see Fig. 5.2a) indicates that labelling of the r-strand must have occurred by addition of a $^{32}$P-dAMP residue to the 3′-terminal dGMP residue of that strand. Therefore, the terminal sequence of the r-strand must be -A-C-C-C-G-C-G. Labelling of the l-strand must have been preceded by the hydrolytic removal of one or more

Table 5.1 Oligonucleotides derived from lambda DNA terminally labelled by incubation with T4 DNA polymerase and $\alpha$-$^{32}$P-dATP

(i) Oligonucleotides separated on DE-81 pH 2

| nucleotide code no. | 3′-penult. base | mobility (cm) on: DE-81 pH 2 | AE-81 pH 3.5 | identity l-strand | r-strand |
|---|---|---|---|---|---|
| | | (blue = 26.0) | (blue = 37.5) | | |
| 1 | T | 45.3 | 25.3 | TA | |
| 2 | G | 36.7 | 16.2 | | GA |
| 3 | G | 34.9 | 10.4 | | CGA |
| 4 | T | 15.8 | 10.9 | TTA | |
| 5 | G | 13.5 | 2.7 | | GCGA |
| 6 | G | 13.5 | 1.7 | | CGCGA |
| 7 | G | 12.0 | 1.1 | | CCGCGA |
| 8 | G | 10.7 | 0.6 | | CCCGCGA |
| 9 | G | 7.8 | 0.3 | | ACCCGCGA |
| 10 | T | 5.1 | 2.6 | GTTA | |
| 11 | T | 2.1 | 0.9 | GGTTA* | |

Table 5.1 (continued)

(ii) Oligonucleotides separated on AE-81 pH 3.5

| nucleotide code no. | 3′-penult. base | mobility (cm) on: AE-81 pH 3.5 | DE-81 pH 2 | identity l-strand | r-strand |
|---|---|---|---|---|---|
| | | (blue = 35.3) | (blue = 28.2) | | |
| 1 | T | 24.3 | 50.5 | TA | |
| 2 | G | 15.3 | 44.4 | | GA |
| 3 | G | 10.2 | 42.5 | | CGA |
| 4 | T | 10.2 | 16.0 | TTA | |
| 5 | G | 2.5 | 13.9 | | GCGA |
| 10 | T | 2.5 | 4.7 | GTTA | |
| 6 | G | 1.2 | 13.8 | | CGCGA |
| 7 | G | 0.8 | 12.7 | | CCGCGA |
| 11 | T | 0.7 | 1.7 | GGTA* | |
| 8 | G | 0.6 | 11.5 | | CCCGCGA |

Table 5.1 (continued)

(iii) Oligonucleotides separated on DE-81 pH 9.7 in the first dimension, and on DE-81 pH 2 in the second dimension

| nucleotide code no. | 3′-penult. base | mobility (cm) on: DE-81 pH 9.7 | DE-81 pH 2 | identity l-strand | r-strand |
|---|---|---|---|---|---|
| | | (blue = 30.5) | (blue = 26.0) | | |
| 4 | T | 29.5 | 16.5 | TTA | |
| 5 | G | 8.7 | 13.5 | | GCGA |
| 6 | G | 6.4 | 12.5 | | CGCGA |
| 7 | G | 3.8 | 11.6 | | CCGCGA |
| 8 | G | 1.7 | 10.7 | | CCCGCGA |
| 10 | T | 11.0 | 4.4 | GTTA | |
| 11 | T | 6.1 | 1.5 | GGTTA* | |

Legend to Table 5.1

Nucleotide code numbers refer to Figure 5.1. The 3′-penultimate base was identified by ionophoresis of 3′-mononucleotides on AE-81 at pH 3.5 (Ap and Gp not resolved) or by chromatography on polyethyleneimine-cellulose thin layers with 5% formic acid (Weigel et al., 1973). Assignment of a nucleotide family to a strand of lambda DNA was by nearest-neighbour analyses of separated strands. Dinucleotide sequences confirmed by these nearest-neighbour analyses are underlined (Weigel et al., 1973).

Mobilities may be converted to $R_F$ values by taking the ratio of the distance migrated by the nucleotide to that migrated by the blue dye (Xylene Cyanol FF).

Nucleotide sequences refer to 5′-oligonucleotides. The 5′-terminal phosphate has been omitted for clarity.

* The 5′-base is G or T. G is more probable.

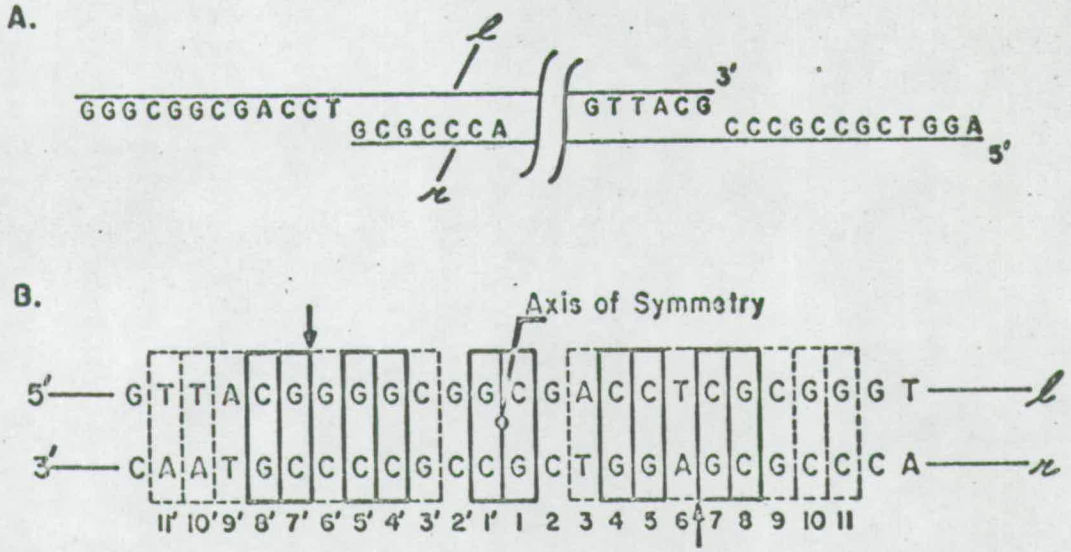



Figure 5.2. (A) Nucleotide sequences at the termini of lambda DNA. The sequences of the cohesive ends, as well as the 3′-terminal dGMP residues, were determined by Wu and Taylor (1971). (B) Symmetry in the region of the cohesive ends. The ends are paired, and nicks are indicated by vertical arrows. The 2-fold axis of symmetry is perpendicular to the page. Symmetrical base pairs are enclosed in solid boxes and pairs symmetrical in purine-pyrimidine orientation are enclosed in dashed boxes. Base pairs are numbered on either side of the axis.

nucleotides from the terminus.

Further experiments (not performed by myself) in this collaborative series (Weigel et al., 1973) using $\alpha$-$^{32}$P-dGTP showed that the -G-T-T-A sequence of the l-strand is preceded by -C-G. Additional oligonucleotides were isolated from DNA labelled with the other two deoxynucleoside triphosphates. Their sequences confirmed those already deduced, but unfortunately did not extend them.

We can now write a sequence of 25 base-pairs, within which lie the nicks produced by the Ter function. The most striking property of this sequence is rotational symmetry (Fig. 5.2b). When the cohesive ends are paired the sequence between the two nicks is bisected by a 2-fold rotational axis. Five of the first eight base-pairs on either side of the axis, and the nicks themselves, are symmetrical. Also an additional five of the first 11 pairs on either side of the axis are symmetrical in their purine-pyrimidine orientation. The term 'hyphenated symmetry' has been used to describe sequences of this type in which the symmetry is interrupted by some nonsymmetrical base pairs (Arber and Linn, 1969).

# Chapter 6

## LAMBDA DNA FRAGMENTS PRODUCED BY R.Eco$_{P1}$

### 6.1 DNA restriction by endonuclease R.Eco$_{P1}$

Bacteriophage lambda is very efficiently restricted by the P1-specific host-specificity system (e.o.p. $2 \times 10^{-5}$; Arber and Linn, 1969). This presumably indicates a relatively large number of sites at which the restriction enzyme acts.

The sedimentation profiles in neutral sucrose gradients of $^{3}H$ λ.K DNA and $^{32}P$ λ.K (P1) DNA incubated with R.Eco$_{P1}$ are shown in Figure 6.1. Appropriately modified DNA is unaffected. Unmodified DNA is degraded to fragments of high molecular weight with little or no non-sedimenting radioactivity. The number of breaks introduced into intact DNA can be estimated using the formula of Studier (1965). The position of the unmodified DNA peaks in Figures 6.1b and 6.1c are the same and therefore represent the DNA limit product. The maxima of the restricted peaks occur at a position corresponding to the production of about 8 breaks in lambda DNA.

The lambda DNA fragments produced by the action of R.Eco$_{P1}$ can be further resolved by electrophoresis in 1% agarose gels. The results of such an experiment are given in Figure 6.2, where the sedimentation profile and gel pattern of a sample of restricted $^{32}P$-labelled DNA are shown side by side.

Figure 6.1. Sedimentation profiles of a mixture of $^{32}P$ λ.KP DNA (0.12 μg) and $^{3}H$ λ.K DNA (0.15 μg) incubated for 20 minutes (A) without enzyme (B) with 2 μl $R.Eco_{Pl}$ preparation (C) with 5 μl $R.Eco_{Pl}$ preparation. Conditions for digestion with $R.Eco_{Pl}$ and details of the sedimentation analysis were as described in Methods. ——— • ——— $^{32}P$, ———o——— $^{3}H$.

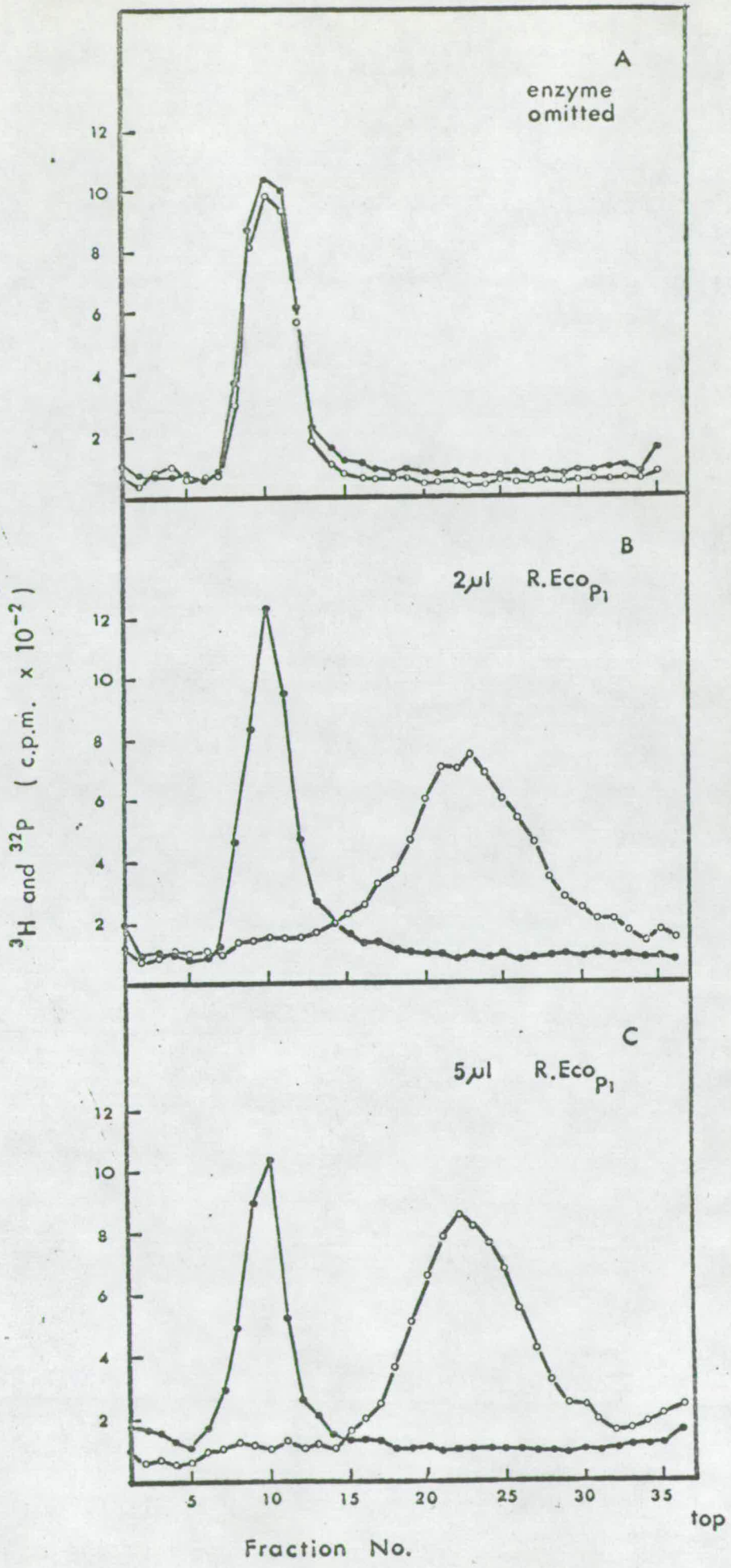
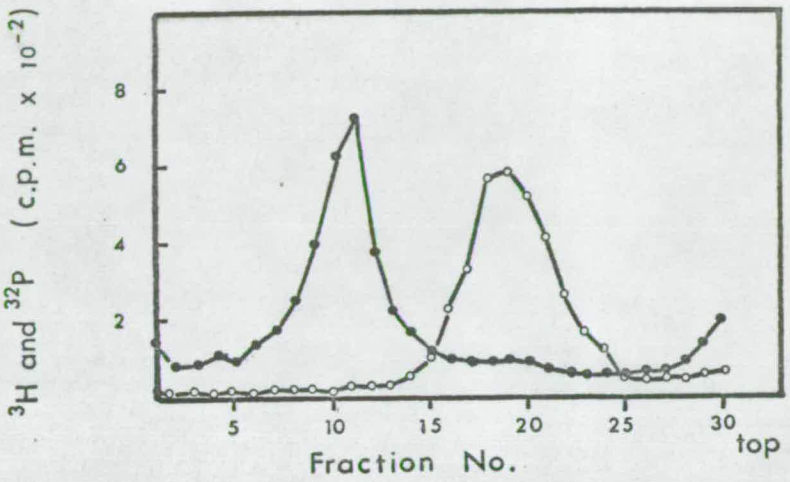

A
enzyme
omitted
B
2µl R.Eco$_{P_1}$
C
5µl R.Eco$_{P_1}$
$^3H$ and $^{32}P$ ( c.p.m. x $10^{-2}$ )
12
10
8
6
4
2
5
10
15
20
25
30
35
top
Fraction No.

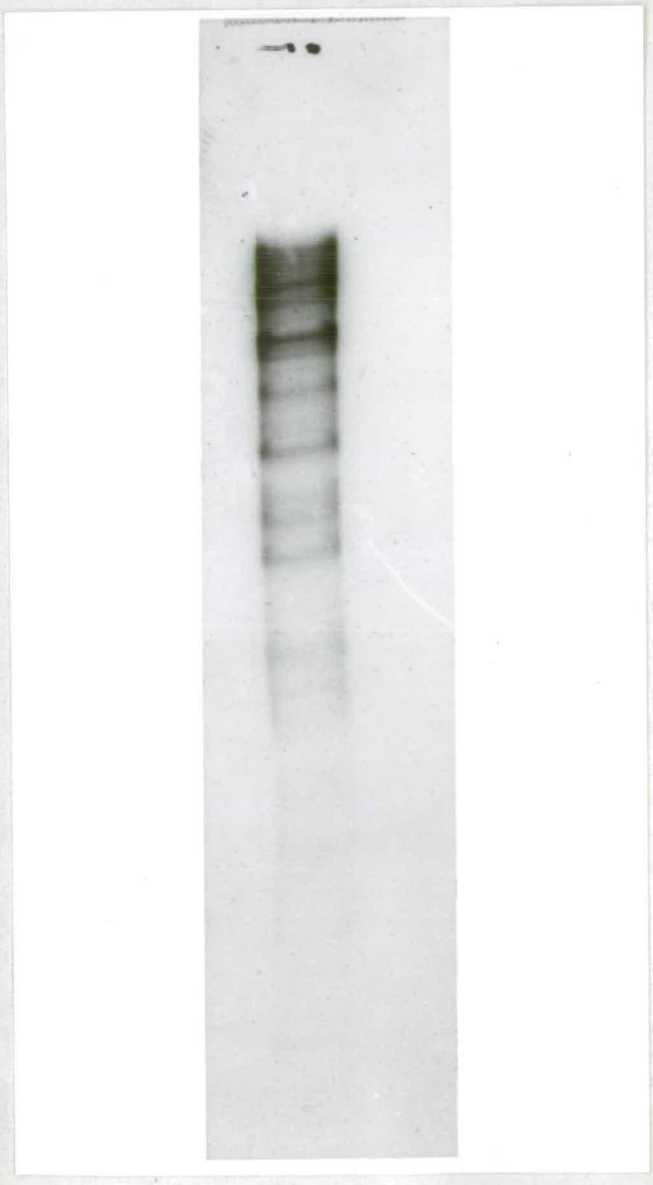
Figure 6.2. Lambda DNA fragments produced by incubation of a mixture of $^{3}H$ λ.KP DNA (0.15 μg) and $^{32}P$ λ.KP DNA (0.70 μg) with 10 μl $R.Eco_{P1}$ preparation, as described in Methods. A sample of the reaction mixture was used for analysis (B) by sedimentation in a neutral sucrose density gradient (B) by electrophoresis in a 1% agarose gel. The gel autoradiograph is reproduced actual size. Bromophenol blue migrated about 24 cm. Anode towards bottom.

——•—— $^{3}H$, ——o—— $^{32}P$.

A




B



It is quite clear that R.Eco$_{P1}$ does give rise to fragments that form bands on gel electrophoresis. This is in marked contrast to R.Eco$_K$ (Murray et al., 1973) which does not give bands at all with lambda DNA, and means that the number of possible sites of breakage is very much smaller with R.Eco$_{P1}$ than for R.Eco$_K$, or that numbers of potential sites are clustered together.

The endonuclease R.Eco$_{P1}$ requires ATP and $Mg^{++}$. Omission of ATP or $Mg^{2+}$, or addition of EDTAto sequester the divalent metal ion, leaves unmodified DNA unaffected in a restriction assay. However the results shown in Figure 6.3 show that the enzyme has no strict requirement for the addition of SAM to the reaction, contrary to the report of Meselson and Yuan (1968). This observation has been confirmed and extended by Murray (personal communication) in a series of reactions with various enzyme : unmodified DNA ratios in the presence and absence of added SAM. A low concentration of enzyme, insufficient for restriction of a certain amount of unmodified DNA, in the absence of added SAM, is sufficient for restriction when SAM is added.

It is not known whether SAM is consumed in the reaction or acts only catalytically, possibly as an allosteric effector. It is quite possible that the enzyme retains tightly bound SAM throughout the purification procedures used in its isolation, and this must be the case on the hypothesis that endonucleolysis occurs

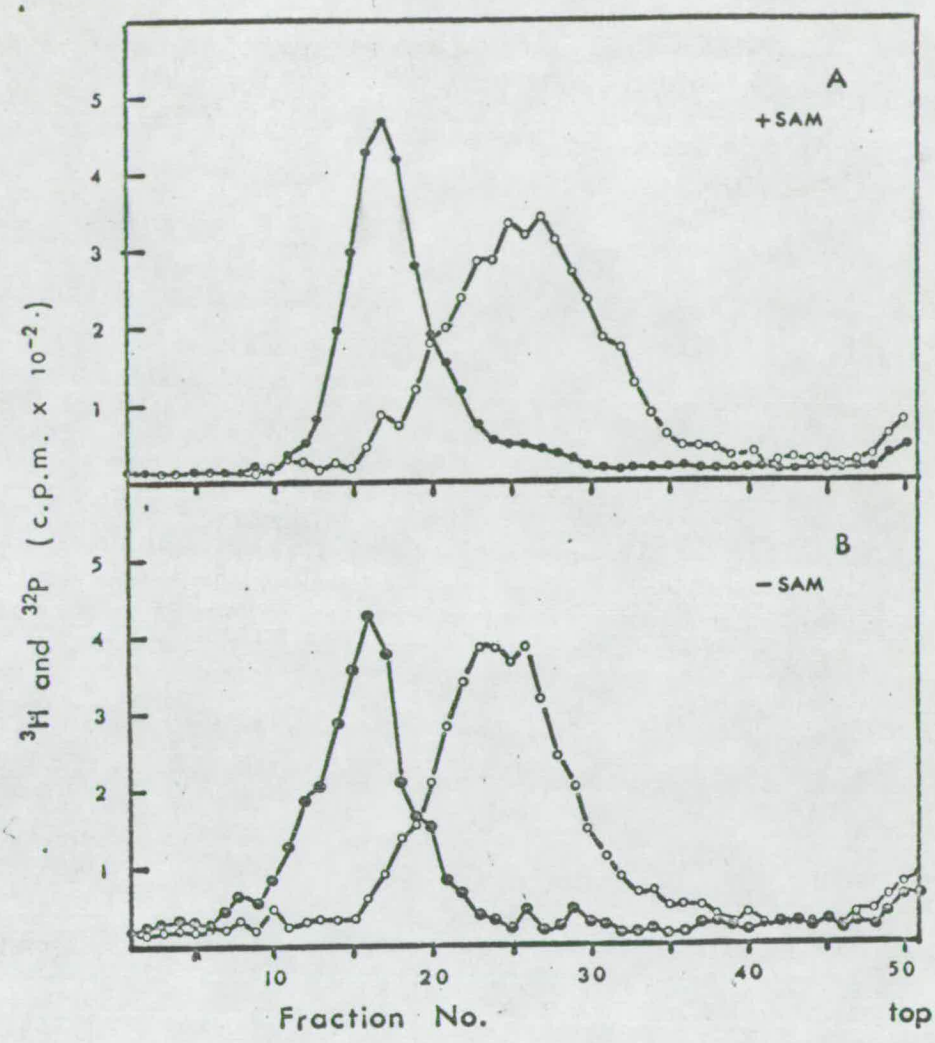



Figure 6.3. Cofactor requirement of R.Eco$_{P1}$. Sedimentation profiles of a mixture of $^{32}P$ λ.KP DNA (0.44 μg) and $^{3}H$ λ.O DNA (o.55 μg), after incubation with 5 μl of R.Eco$_{P1}$ preparation. Conditions were as described in Methods except for the following: (A) SAM concentration 3 x $10^{-5}$ M (B) SAM omitted. Sedimentation analyses were as described in Methods except that 2-drop fractions were collected.

via depurination after methylation by methyl-group transfer from SAM.

## 6.2 Absence of exonucleolytic activity

In sucrose gradient sedimentation analysis little or no non-sedimenting material is produced from restricted DNA. Therefore it seems unlikely that the enzyme itself has an exonucleolytic activity or that it is contaminated with exonuclease activity. In order to substantiate this further the digest was examined for released nucleotides. A solution of $^{32}$P-labelled λ.0 DNA was extensively digested with an excess of endonuclease. Enzyme was omitted from a control incubation. The restricted DNA product was analysed after incubation by zone sedimentation analysis with a $^{3}$H-labelled intact λ DNA marker. Restriction was efficient.

Samples of each incubation were applied to AE-cellulose paper, coincident with 5′-mononucleotide reference markers (10 mg/ml; 10 µl), for ionophoresis at pH 3.5. The mononucleotide zones were located under U.V. light and counted for radioactivity in an anti-coincidence, low background, gas-flow counter. The results (Table 6.1) show that essentially no activity was found associated with the mononucleotide spots.

The ionophoretogram was autoradiographed, so that nucleotides other than 5′-mononucleotides could be detected. None were apparent after 2 weeks of autoradiography.

Table 6.1 Release of nucleotides by R.Eco$_{P1}$

A digest of $^{32}$P- λ.C DNA with R.Eco$_{P1}$ was analysed by ionophoresis on AE-81 at pH 3.5, as described in the text. 5′-mononucleotide marker spots were cut out and assayed for radioactivity in a gas-flow counter.

| | counts/10 minutes* | |
|---|---|---|
| Nucleotide | R.Eco$_{P1}$ present | R.Eco$_{P1}$ omitted |
| pT | 275 | 248 |
| pC | 251 | 220 |
| pA | 175 | 325 |
| pG | 297 | 325 |
| origin | $3.3 \times 10^6$ | $3.1 \times 10^6$ |

* A blank value of 247 has been subtracted. This value was obtained by counting a blank region of the ionophoretogram equivalent in size to the mononucleotide spots.

## 6.3 Terminal labelling of lambda DNA fragments produced by restriction

The polynucleotide kinase reaction was used to investigate the termini of lambda DNA that had been restricted by R.$Eco_{P1}$.

Non-radioactive λ.0 DNA (4.1 μg) was incubated in a reaction with R.$Eco_{P1}$ (50 μl) under conditions similar to those given in the legend to Figure 2. A control incubation with $^{32}P$ λ.K and $^{3}H$ λ.KP DNAs and a small portion of the reaction mixture. Sedimentation in a neutral sucrose gradient confirmed that the restriction was efficient. After incubation at 37° for 30 minutes the reaction mixture was extracted with neutralised phenol, dialysed exhaustively first against 0.3 M NaCl, 0.01 M Tris-HCl pH 8.0, 0.001 M EDTA to remove ATP and then against 0.01 M Tris-HCl pH 8.0, 0.001 M EDTA. The DNA solution (2.0 ml) was treated with alkaline phosphatase (20 μg; incubation, 37° for 40 minutes), extracted with neutralised phenol, and then dialysed against 20 mM Tris-HCl pH 7.4.

The lambda DNA fragments were incubated with polynucleotide kinase (50 μl: final fraction) in a reaction mixture (2.45 ml) containing 20 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol and 20 nmole of $\gamma$-$^{32}P$-ATP. After incubation at 37° for 40 minutes the reaction was stopped by the addition of 0.2 M EDTA

(0.2 ml) and the products dialysed against 0.3 M NaCl, 0.01 M Tris-HCl pH 8.0, 0.001 M EDTA. This dialysis removed most of the unreacted ATP. The radioactive high molecular weight products of the reaction were isolated as the excluded peak of radioactivity upon chromatography on a column (45 cm x 1 cm diameter) of Sephadex G-100 eluted with 0.01 M Tris-HCl pH 8.0, 0.001 M EDTA. Fractions were pooled, dialysed against 0.5 mM Tris-HCl pH 7.5 and concentrated by evaporation in a vacuum desiccator. About 3,700 cpm were incorporated into DNA.

A sample of the DNA was digested to 5′-mononucleotides to determine the 5′-terminal nucleotides labelled in the polynucleotide kinase reaction. The results (Table 6.2) show that the majority of the radioactivity was incorporated into pA and pG.

Using the known specific radioactivity of the $\gamma$-$^{32}$P-ATP and the molecular weight of lambda DNA, a total of about 4,200 cpm is calculated as the incorporation of radioactivity into 4.1 $\mu$g of intact lambda DNA. Therefore it seems probable that only those 5′-nucleotides that terminate whole lambda DNA molecules (i.e. pA and pG) were labelled in this reaction, and that the 5′-termini created by the action of R.Eco$_{P1}$ are not substrates for polynucleotide kinase even after treatment with phosphatase. These results could only otherwise be obtained if the terminal labelling reaction was very

Table 6.2 Labelling of lambda DNA with polynucleotide kinase after incubation with R.Eco$_{P1}$

Lambda DNA fragments produced by treatment with R.Eco$_{P1}$ were incubated with polynucleotide kinase and $\gamma$-$^{32}$P-ATP as described in the text. A sample of the labelled material was digested to mononucleotides in a solution (100 µl) of pancreatic DNase (50 µg/ml) and venom phosphodiesterase (20 µg/ml) containing 30 mM Tris-HCl pH 7.5 and 10 mM $MgCl_2$. Incubation was at 37° for 2 hr. Radioactive 5′-mononucleotides were fractionated by ionophoresis on AE-cellulose paper at pH 3.5, and located under U.V. light by the presence of marker mononucleotides. Mononucleotide regions were assayed for radioactivity in a gas-flow counter.

| Nucleotide species | count/40 min* |
|---|---|
| pT | 997 |
| pC | 898 |
| pA | 1929 |
| pG | 1835 |

* A background value of 183 has been subtracted. This value was obtained by counting an appropriately sized blank region of the ionophoretogram.

inefficient and the newly created 5′-terminal bases were at least predominantly A and G in equal proportions.

A sample of the labelled material was digested with pancreatic DNase, as described in the legend to Figure 4, and analysed by ionophoresis on DE-cellulose paper at pH 2.0. There are two difficulties in this type of experiment. Firstly, the amount of DNA that can be used is limited by the fact that a large proportion of an endonuclease R.Eco$_{P1}$ preparation must be expended in restricting a few μg of lambda DNA. In this respect R.Eco$_{P1}$ resembles R.Eco$_{B}$, for which no turnover as an endonuclease can be demonstrated (Eskin and Linn, 1972a). This means that the amount of radioactivity incorporated into the natural ends of whole lambda DNA is small. Secondly, the background labelling is inevitably increased by the incubation with R.Eco$_{P1}$ and subsequent processing of the DNA. However, nucleotides are apparent in Figure 6.4 which have mobilities consistent with nucleotides derived from the 5′-termini of lambda DNA (Murray, 1973). These did not contain sufficient radioactivity for analysis by digestion with venom phosphodiesterase.

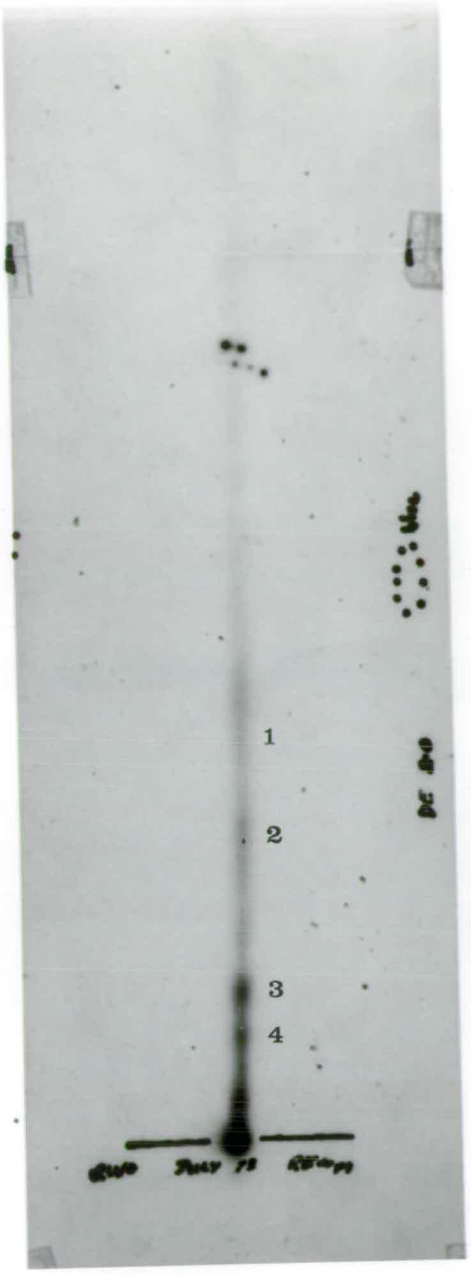



Figure 6.4. Polynucleotide kinase reaction and terminal labelling of fragments of lambda DNA after incubation with R.Eco$_{Pl}$. Lambda DNA fragments produced by treatment with R.Eco$_{Pl}$ were incubated with polynucleotide kinase and $\gamma$-$^{32}$P-ATP as described in the text. A sample of the labelled material was digested by incubation with pancreatic DNase (50 $\mu$g/ml) in a solution of Tris-HCl pH 7.5 (30 mM) and $MgCl_2$ (10 mM) at 37° for 1 hr. The digest was applied to DE-cellulose paper for ionophoresis at pH 2. Spots 1, 2, 3 and 4 have mobilities consistent with pG-G, pA-G-G, pG-G-G (+ pG-G-G-C), and pA-G-G-T (+ pA-G-G-T-C) respectively.

# Chapter 7

## DNA NUCLEOTIDE SEQUENCE RECOGNISED BY RESTRICTION ENDONUCLEASE R.Hin$_d$ III

### 7.1 Purification of restriction endonuclease from H. influenzae strain Rd

The enzyme was partially purified from wild-type H. influenzae Rd according to the procedure given in Methods. No cofactors other than $Mg^{++}$ were present in the restriction assay. Although the identity of the restriction endonuclease prepared in this way will not be discussed until section 7.7, it will be referred to as R.Hin$_d$ III after the suggestion of Smith (personal communication). All of the following experiments were performed with the final fraction.

### 7.2 Action of R.Hin$_d$ III on phage lambda DNA

Figure 7.1 shows the zone sedimentation profile, in neutral sucrose density gradients, of lambda DNA after treatment with R.Hin$_d$ III. Using the formula of Studier (1965), the average molecular weight of the $\lambda$ DNA fragments produced by R.Hin$_d$ III is calculated as about $1.8 \times 10^6$. Therefore the enzyme makes about 17 double-strand breaks in lambda DNA.

The fragments of lambda DNA produced by R.Hin$_d$ III

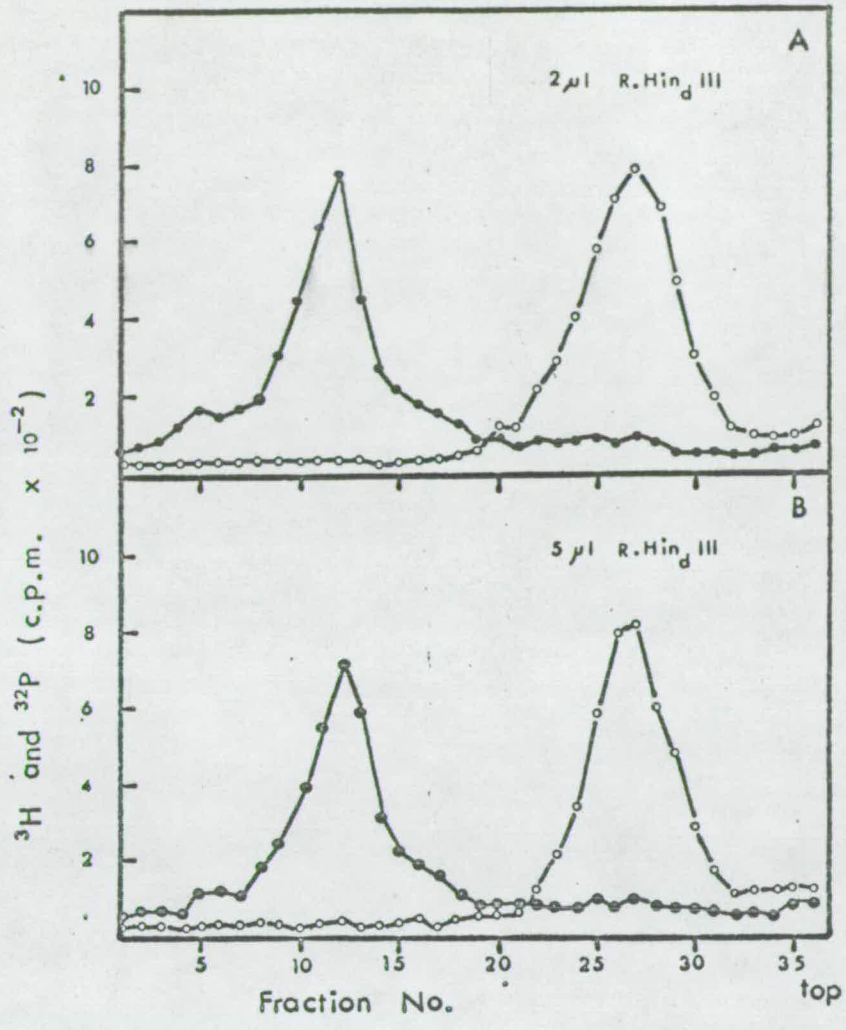



Figure 7.1. Sedimentation profile of lambda DNA after treatment with R.Hin$_d$ III. $^3$H λ.0 DNA (0.13 µg) was incubated under standard assay conditions (Methods) with R.Hin$_d$ III. After the reaction had been stopped by the addition of EDTA, $^{32}$P λ DNA was added as a sedimentation marker. Sedimentation analyses were as described in Methods. ——o—— $^3$H, ——●—— $^{32}$P.

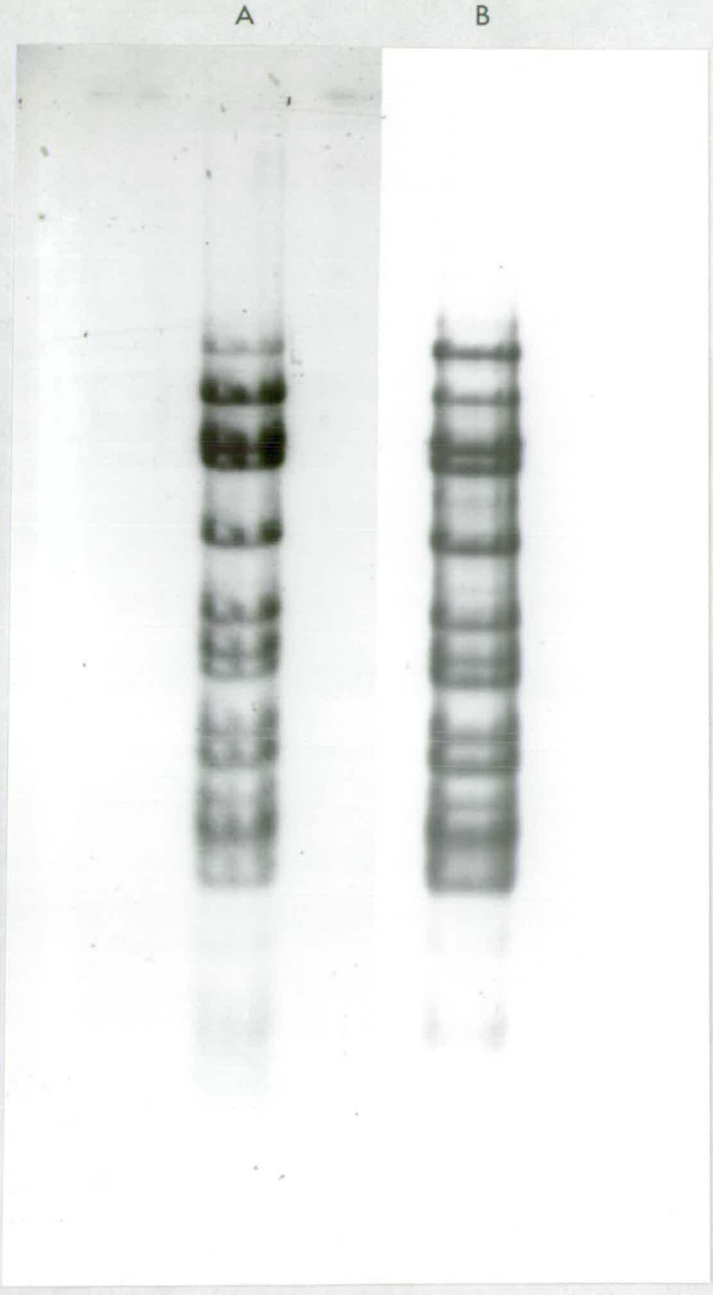



Figure 7.2. Agarose gel electrophoresis of lambda DNA after treatment with R.Hin$_d$ III. (A) $^{32}$P λ DNA (1.5 μg) and (B) $^{32}$P λ-sK$^o$ DNA (1.5 μg) were incubated with R.Hin$_d$ III preparation (5 μl) under standard assay conditions, at 37$^o$ for 2 hr. The DNA fragments were fractionated by electrophroesis on 1% agarose gels. The gel autoradiograph is reproduced actual size. Bromophenol blue migrated about 30 cm. Anode towards bottom.

give complicated band patterns upon electrophoresis in 1% agarose gels. Autoradiographs of the bands obtained with $^{32}P$-labelled DNA are shown in Figure 7.2. These digests, which can be obtained quite reproducibly, contain some partial products.

## 7.3 Action of R.Hin$_d$ III on SV40 DNA

Fragments of SV40 DNA produced by treatment with R.Hin$_d$ III are shown in Figure 7.3. Non-radioactive SV40 DNA fragments were analysed by electrophoresis in 1% agarose gels. Six fragments were apparent. Therefore SV40 DNA contains six sites susceptible to R.Hin$_d$ III. (If two sites are so close as to produce a DNA fragment with molecular weight less than about $10^5$, they would appear a single site.)

The length of each of the six fragments of SV40 DNA was determined by electron microscopy. Fragments were located by fluorescence in u.v. light after electrophoresis in a 1% agarose gel containing ethidium bromide, and eluted by macerating excised (wet) gel sections in 0.5 M ammonium acetate (0.3 ml).

The DNA was prepared for electron microscopy by spreading with cytochrome c. Electron microscopy (Fig. 7.4) and DNA length measurements were carried out using the procedures given in Methods.

Electron micrographs of the SV40 DNA preparation used in these experiments indicated that about 70% of the DNA molecules were supercoiled covalently closed circular structures. The remainder (30%) were open circular molecules (Fig. 7.4a). The preparation was

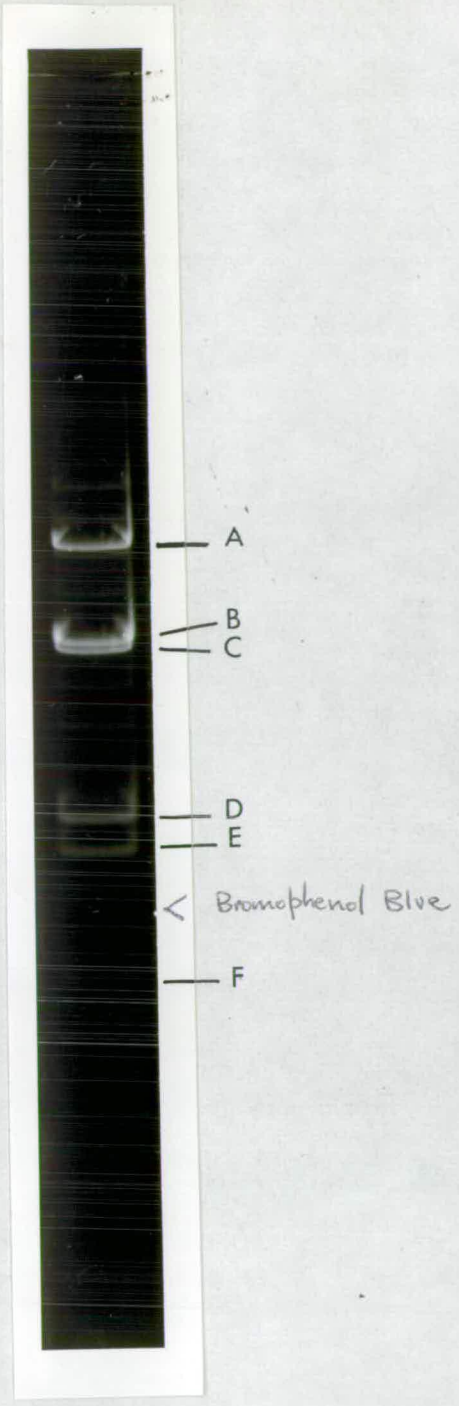



Figure 7.3. Fragments of SV40 DNA produced on treatment with R.Hin$_d$ III, analysed by electrophoresis in a 1% agarose gel containing ethidium bromide. The gel was photographed in u.v. light. Bands are lettered from cathode to anode. The fragments were prepared by treatment of SV40 DNA (3 $\mu$g) with R.Hin$_d$ III preparation (2.5 $\mu$l) at 37$^{o}$ for 45 minutes under standard assay conditions. The position to which Bromophenol Blue migrated is indicated.

Figure 7.4. Electron micrographs of SV40 DNA and fragments of SV40 DNA produced by treatment with restriction endonucleases.

(i) Supercoiled and open circular molecules of SV40 DNA.

(ii) Linear molecules of SV40 DNA. These molecules were produced by incubation of SV40 DNA (1.2 $\mu$g) with R.Eco$_{RI}$ (0.5 $\mu$l) at 37$^{o}$ for 45 minutes under the conditions for digestion with R.Eco$_{RI}$ given in Methods.

Electron micrographs A-F show the correspondingly lettered fragments of SV40 DNA eluted from a 1% agarose gel similar to that shown in Fig. 7.3.

Magnification is approx. 50,500 x in photographs i, ii, A, B and C.

Magnification is approx. 100,000 x in photographs D, E and F.

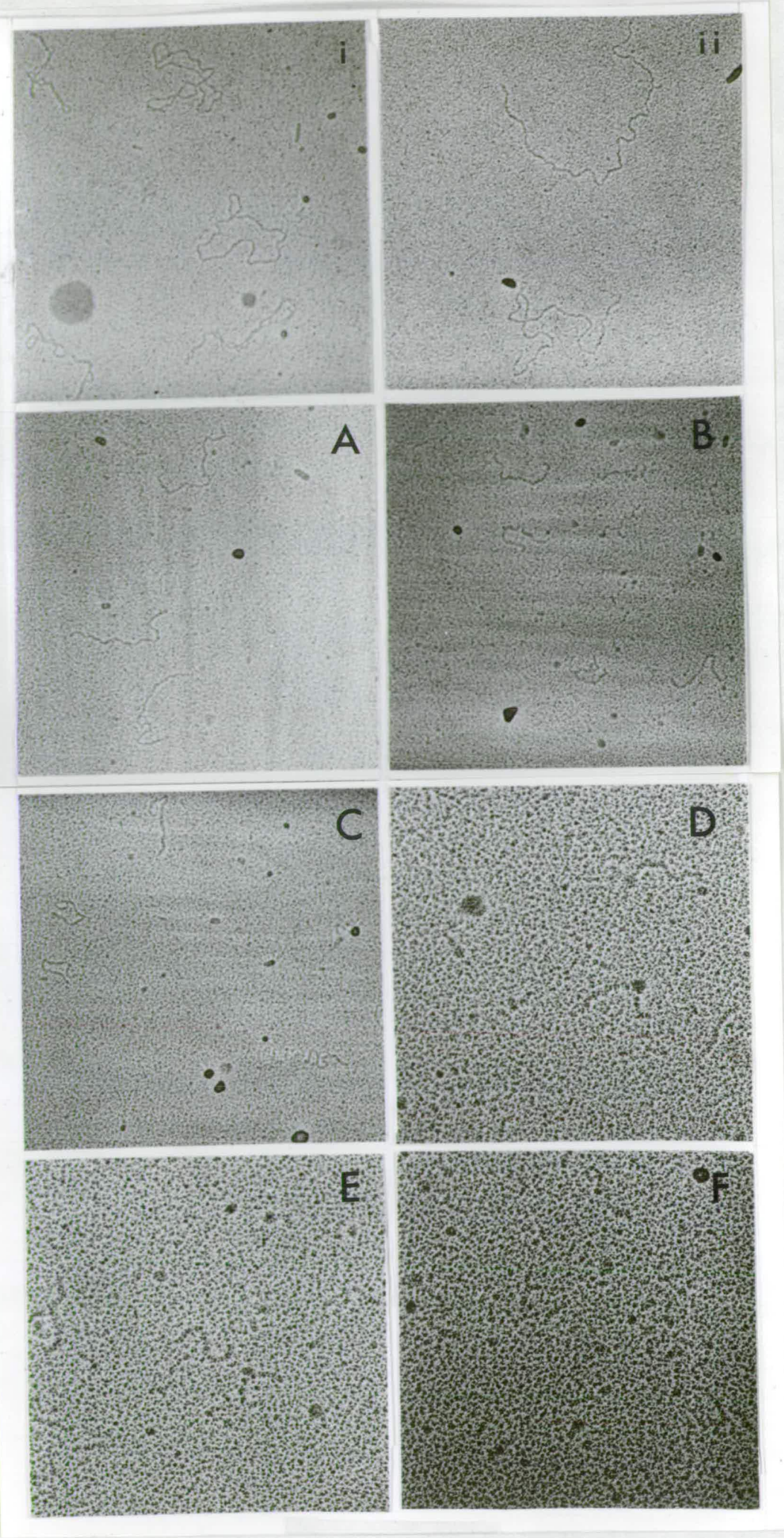

i
ii
A
B
C
D
E
F

Table 7.1 Molecular length determination of fragments of SV40 DNA produced by R.Hin$_d$ III

| | SV40 circles | RI product | R.Hin$_d$ products A | B | C | D | E | F |
|---|---|---|---|---|---|---|---|---|
| No. of molecules measured | 8 | 9 | 14 | 16 | 14 | 13 | 8 | 12 |
| Average length (μm) | 1.80 | 1.80 | 0.64 | 0.40 | 0.38 | 0.14 | 0.13 | 0.08 |
| Standard deviation (± μm) | 0.04 | 0.04 | 0.05 | 0.05 | 0.03 | 0.03 | 0.04 | 0.04 |
| Fragment as % SV40 DNA | - | 100 | 35.4 | 22.0 | 21.0 | 8.0 | 7.2 | 4.7 |
| MW* x $10^6$ | - | 3 | 1.06 | 0.66 | 0.63 | 0.24 | 0.22 | 0.15 |

* Calculated on the basis of 3 x $10^6$ as the molecular weight of SV40 DNA (Crawford and Black, 1964).

sensitive to R.Eco$_{RI}$, giving full-length linear molecules that are the result of breakage at a specific site in the SV40 DNA molecule (Morrow and Berg, 1972). Linear molecules produced by treatment with R.Eco$_{RI}$ are shown in Figure 7.4b. All of the molecules (50 out of 50) were sensitive to R.Eco$_{RI}$ and only full-length linear molecules were produced. This indicates that the SV40 DNA preparation did not contain a significant proportion of defective molecules.

The results of the length determinations for each of the six fragments of SV40 DNA produced by R.Hin$_d$ III are given in Table 7.1. The sum of the lengths of the fragments is in good agreement (98.3%) with the length of the whole molecule. In view of the standard deviations of the length measurements such good agreement is fortuitous. No attempt was made to remove the ethidium bromide that was eluted from the gel with the DNA fragments. Intact SV40 DNA molecules and linear monomers were prepared in the absence of ethidium bromide. These results show that the effect of the remaining ethidium bromide did not significantly affect the length of the DNA fragments.

These results, together with the mobilities (Fig. 7.3) of the fragments of lambda DNA produced by R.Eco$_{RI}$ which are of known molecular weight (Allet et al., 1973), calibrate the gel electrophoresis system. The molecular weight versus mobility curve is given in

Fig. 7.5. The system provides a useful method for molecular weight determinations of double-stranded DNA molecules, by interpolation within the range 0.2 - 20 million daltons, when samples are run in conjunction with the fragments of known molecular weight. Untreated SV40 DNA (forms I and II) migrates considerably faster than the linear fragments produced by treatment with $R.Eco_{RI}$.

In a recent paper Danna et al. (1973) give the cleavage map of the SV40 genome, in which the locations of the sites of cleavage by several restriction enzyme preparations are given. They have tested a subfraction of R.H which is one of two subfractions ($R.Hin_d$ II and $R.Hin_d$ III) having different specificities (H.O. Smith, personal communication). They state that one of these fractions ($R.Hin_d$ III) cleaves SV40 into six fragments. They deduce, from an analysis of the digest products, that $R.Hin_d$ III cleaves SV40 DNA at Hin sites 4, 6, 7, 9, 10 and 11 (see Figure 7.6). Given that the $R.Hin_d$ III sites comprise this sub-set of the Hin sites, and from the sizes of the 11 fragments produced by R.H, the following fragment sizes are predicted as the products of SV40 DNA produced by $R.Hin_d$ III: 34.0%, 22.5%, 20.5%, 10.5%, 8.5%, 4.0% (expressed as a percentage of total SV40 DNA). These values are in rather good agreement with, and confirm those given in Table 7.1. This is an encouraging result, for the smaller Hin fragments were

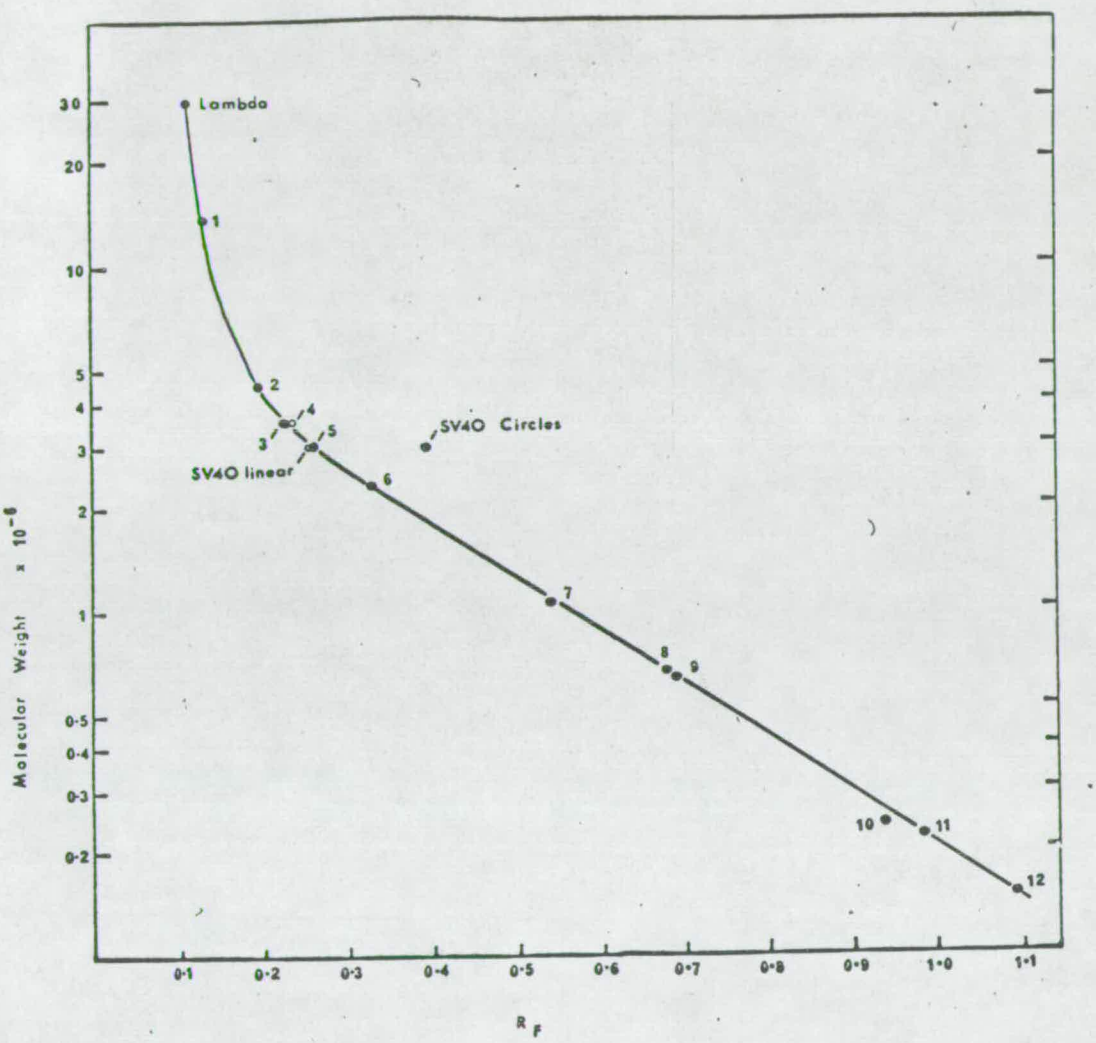



Figure 7.5. Relationship between molecular weight and mobility of duplex DNA fragments upon electrophoresis in 1% agarose gels containing ethidium bromide. The relative mobilities of DNA fragments are unaffected by the omission of ethidium bromide.

1–6 Fragments of lambda DNA produced by R.$Eco_{RI}$.

7–12 Fragments of SV40 DNA produced by R.$Hin_{d}$III.

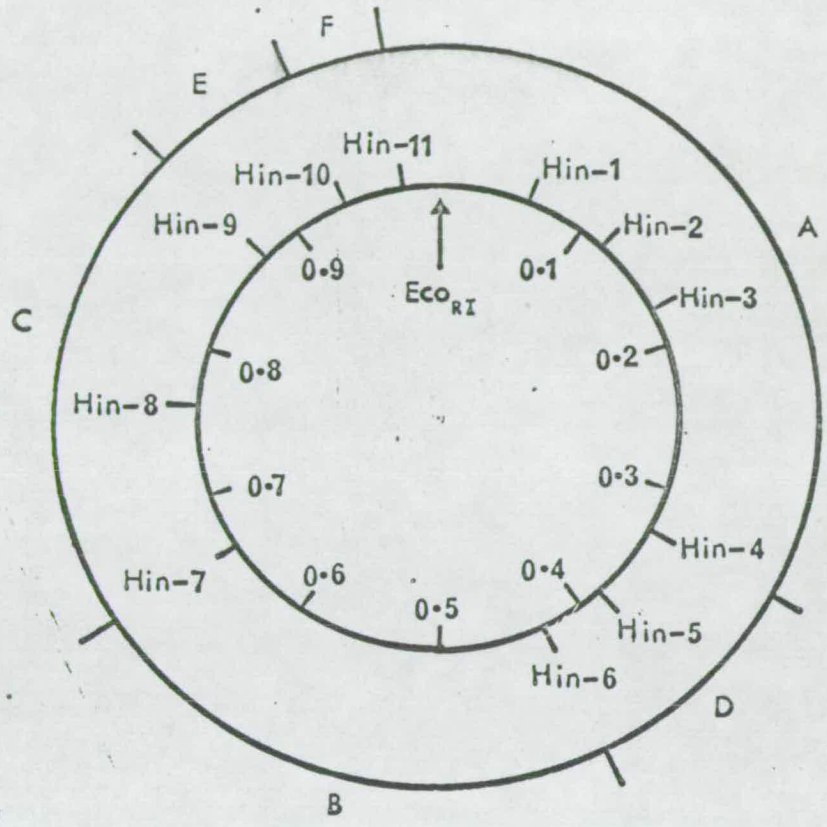



Figure 7.6. Location of the fragments of SV40 DNA produced by R.Hin$_d$ III on the physical map of SV40 DNA. This map is due to the work of Danna et al. (1973).

sized on the basis of the distribution of radiolabel in a complete R.H digest of $^{32}$P-SV40 DNA with the assumption of unimolar yield (Danna and Nathans, 1971).

This correlation enables the DNA fragments visible in Figure 7.4 to be located within the SV40 genome.

The SV40 DNA fragments produced by R.Hin$_d$ III alone, or in combination with other restriction endonucleases, could be of considerable use in studies of defined regions of the SV40 genome (see Chapter 8).

## 7.4 Nucleotide sequence at the site of action of R.Hin$_d$ III

### (a) 5′-terminal sequence of lambda DNA fragments produced by R.Hin$_d$ III

The polynucleotide kinase reaction was used to terminally label lambda DNA fragments produced by R.Hin$_d$ III.

Bacteriophage lambda DNA (300 µg) was treated with R.Hin$_d$ III (22 µl) in a 1.0 ml reaction under standard assay conditions. A sample (0.1 ml) of the reaction mixture was withdrawn at the start of the incubation and mixed with 3 µg of $^{3}$H-labelled lambda DNA to check the course of the reaction. Small portions (10 µl) of this sample were taken at intervals during incubation at 37°, and analysed by sedimentation in neutral sucrose

gradients after the addition of EDTA (to stop the reaction) and $^{32}$P-labelled lambda DNA as sedimentation marker.

After 5 hr incubation at 37° the large scale reaction was stopped by the addition of EDTA to 20 mM. Sedimentation of the labelled DNA in the test reaction showed that the DNA had been extensively degraded.

The DNA fragments were treated with bacterial alkaline phosphatase (20 µg/ml; 1 hr at 37°), extracted with neutralised redistilled phenol, dialysed exhaustively against 0.01 M Tris-HCl pH 8.0, 0.001 M EDTA and finally dialysed against 0.02 M Tris-HCl pH 7.4.

DNA fragments were terminally labelled as follows. The reaction mixture (1 ml) contained: 25 µg DNA, 20 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 10 mM 2-mercapto-ethanol, 10 nmoles $\gamma$-$^{32}$P-ATP and polynucleotide kinase (50 µl: final fraction). The solution was incubated at 37° for 45 minutes, a further 50 µl of the polynucleotide kinase preparation was added and incubation continued for another 45 minutes. The reaction was stopped by the addition of EDTA to 20 mM. The solution was dialysed against 0.3 M NaCl, 0.01 M Tris-HCl pH 8.0, 0.001 M EDTA and finally dialysed against 0.01 M ammonium bicarbonate. Labelled DNA fragments were purified by chromatography on a column (45 x 1 cm diameter) of Sephadex G-100 run in 0.01 M ammonium bicarbonate. They appeared as a peak of radioactivity at the void volume. Fractions

were combined and dialysed against 0.5 mM NaCl.

The labelled DNA was digested with pancreatic DNase (50 μg/ml) in a solution (0.1 ml) containing 10 mM $MgCl_2$, 10 mM Tris-HCl pH 7.4. After incubation at 37° for 1 hr the solution was concentrated by evaporation in a vacuum desiccator.

Nucleotides present in the pancreatic DNase digest of the labelled material were fractionated by ionophoresis on AE-cellulose paper at pH 3.5 and on DE-cellulose paper at pH 2.0. Long (blue dye: 67 cm on AE-81; 61.5 cm on DE-81) ionophoresis runs were performed in addition to shorter runs (blue dye: 37 cm on AE-81; 21 cm on DE-81) so that relatively slow moving nucleotides could be resolved. Labelled nucleotides were located by autoradiography (Fig. 7.7) and eluted. The mobility of each $^{32}P$-oligonucleotide was determined on both ionophoresis systems. The 5′-terminal mononucleotide of each oligonucleotide was determined by complete digestion with venom phosphodiesterase. Oligonucleotides were also analysed by partial digestion with venom phosphodiesterase as described in Methods.

Table 7.2 gives the properties and identities of the labelled oligonucleotides present in the digest. These oligonucleotides belong to a single 5′-terminal family with A as the 5′-terminal base. A single tetranucleotide, pA-G-C-T, was present. The only pentanucleotide identified was p-A-G-C-T-T. Figure 7.8 shows the

Figure 7.7. Fractionation of nucleotides in digests of 5′-terminally labelled fragments of lambda DNA produced by R.Hin$_d$ III.

(A) Fractionation by ionophoresis on AE-cellulose paper pH 3.5 (blue = 37 cm).

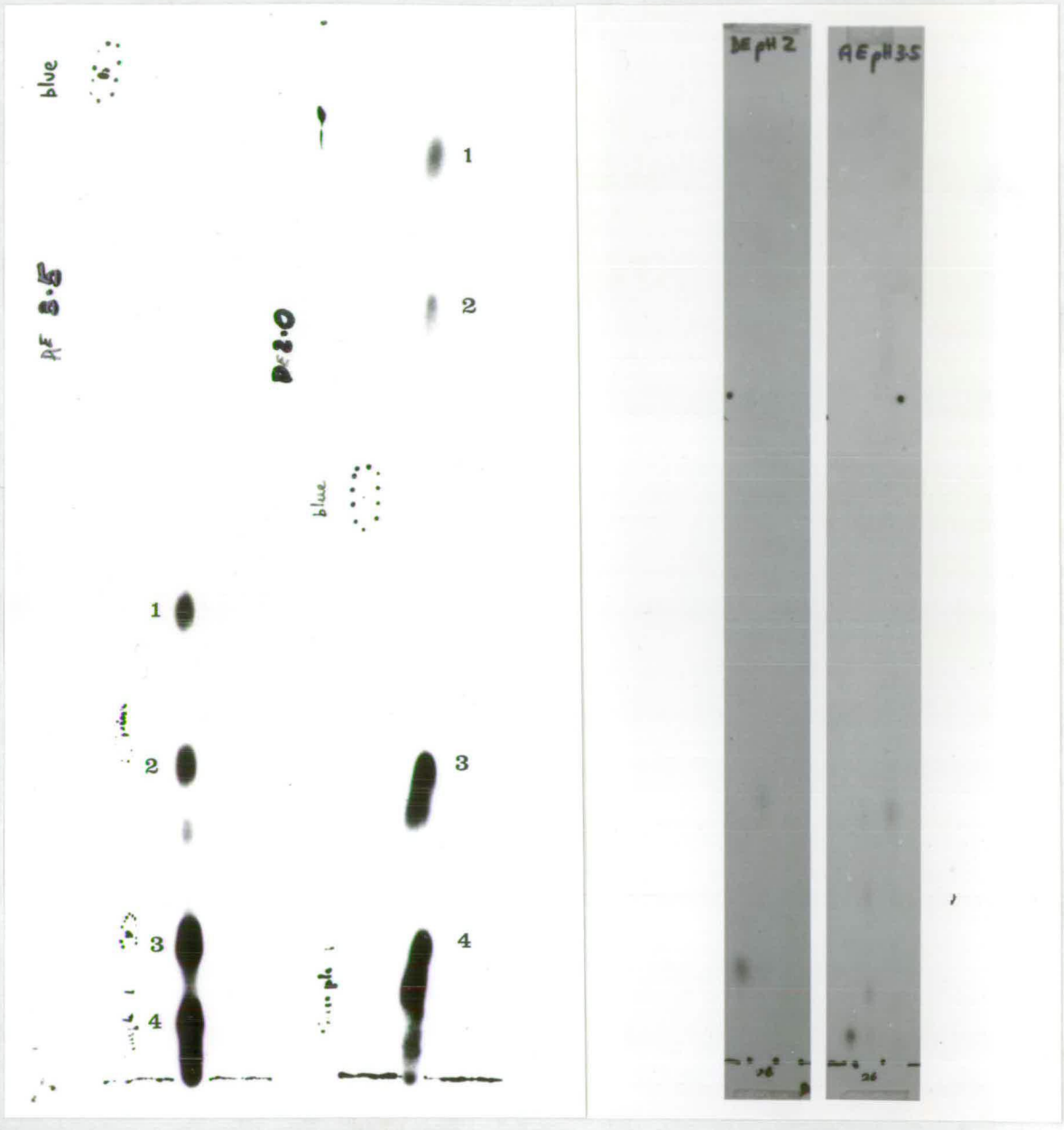
(B) Fractionation by ionophoresis on DE-cellulose paper pH 2 (blue = 21 cm).

Identities of numbered oligonucleotides are given in Table 7.2. Faint unnumbered spots represent nucleotides corresponding to the site of action of R.Hin$_d$ II. Slight contamination with R.Hin$_d$ II was apparent in this experiment in which the lambda DNA was incubated with the R.Hin$_d$ III preparation for 5 hr.

Figure 7.8. Products of partial digestion of nucleotide 5 (Fig. 7.7) with venom phosphodiesterase.

(A) Ionophoresis on AE-cellulose paper pH 3.5

(B) Ionophoresis on DE-cellulose paper pH ~~3.5~~ 2.0.




A B A B

Table 7.2 Radioactive oligonucleotides derived from labelled R.Hin$_d$ III fragments of lambda DNA

| nucleotide code no. | 5′-terminal base | $R_F$ on DE pH 2.0 | $R_F$ on AE pH 3.5 | identity |
|---|---|---|---|---|
| 1 | A | 1.60 | 0.46 | AG |
| 2 | A | 1.34 | 0.31 | AGC |
| 3 | A | 0.50 | 0.13 | AGCT |
| 4 | A | 0.18 | 0.06 | AGCTT |

Code numbers refer to Figure 7.7.

$R_F$ values are relative to Xylene Cyanol FF. The values are average values obtained from several ionophoresis runs.

Sequences refer to 5′-oligonucleotides. The 5′-terminal phosphate has been omitted.

products of partial digestion of this pentanucleotide with venom phosphodiesterase. The partial digests were analysed by ionophoresis on AE-cellulose paper at pH 3.5 and DE-cellulose paper at pH 2.0. All of the smaller nucleotides present in the original pancreatic DNase digest are produced upon partial digestion of the penta-nucleotide.

No other pentanucleotide was identified, but the presence of others in lower amounts could not be ruled out. The low mobility of such penta- (and higher) nucleotides on the ionophoresis systems makes analysis at this level less certain, for although the larger nucleotides may be positively identified it is extremely difficult to be certain that they are unique.

The sequence A-G-C-T-T may be written as part of a double-stranded sequence having an axis of rotational symmetry.

(b) DNA fragments produced by R.Hin$_d$ III have short cohesive ends

The restriction endonucleases R.Eco$_{RI}$ and R.Eco$_{RII}$ both make staggered breaks in duplex DNA, giving rise to cohesive termini (Mertz and Davis, 1972; Hedgpeth et al., 1972; Bigger et al., 1973; Boyer et al., 1973). The 5′-terminal sequence of fragments produced by R.Hin$_d$ III suggested that this enzyme may make staggered breaks also. This possibility was investigated in the following

experiments in which it was shown that short linear fragments of lambda DNA produced by treatment with R.Hin$_d$ III can be cyclised through intramolecular hydrogen bonding.

(i) Bacteriophage lambda DNA (2.5 µg) was treated with a solution (0.2 ml) of R.Hin$_d$ III (2 µl) under standard assay conditions. After incubation at 37° for 30 minutes 10 µl of 0.4 M EDTA was added to stop the reaction.

(ii) A similar incubation was performed in which the EDTA was added before the enzyme.

Both DNA solutions were then brought to a NaCl concentration of 0.1 M by addition of 1 M NaCl, and kept in a cold room at 5° for 2 days. The DNA was prepared for electron microscopy using the cytochrome c spreading procedure described in Methods. All steps up to the Pt-shadowing stage were performed at 5°. Grids were also prepared at 23° with DNA from solution (i) that had been warmed to 37° for 3 minutes just prior to spreading.

The DNA of solution (i) consisted of molecules that were much shorter than whole lambda DNA. These fragments were able to form circular molecules that were stable at 5° (Figure 7.9). These circular molecules had contour lengths in the range 0.5 - 2.0 microns. Circular molecules were not present after heating to 37° (Table 7.3). DNA in solution (ii) consisted of molecules that had been

degraded very little. About 80% of the molecules were intact. Some circular monomers of whole lambda were present. This shows that R.Hin$_d$ III is inactive when $Mg^{++}$ ions are sequestered by EDTA and that the small DNA fragments of solution (i) that could circularise were not introduced as contaminant in the enzyme preparation.

These results show that short linear fragments of lambda DNA produced by R.Hin$_d$ III can circularise at 5°, and that the circles are unstable at 37°. The proportion of circular molecules was rather low even at 5°. The investigation by Mertz and Davis (1972) on the cohesion of the termini produced by R.Eco$_{RI}$ (four A.T base pairs) under conditions similar to, but not identical with, those used here indicate a $T_m$ of about 6° for the equilibrium between linear and circular monomers. Therefore the linear DNA fragments produced by R.Hin$_d$ III have short cohesive ends about four nucleotides in length.

It should be noted that R.Hin$_d$ II breaks the two strands of duplex DNA at opposite points making an even as opposed to staggered break (Kelly and Smith, 1970). The presence of R.Hin$_d$ II as a contaminant in the R.Hin$_d$ III preparation could not account for the formation of cohesive termini.

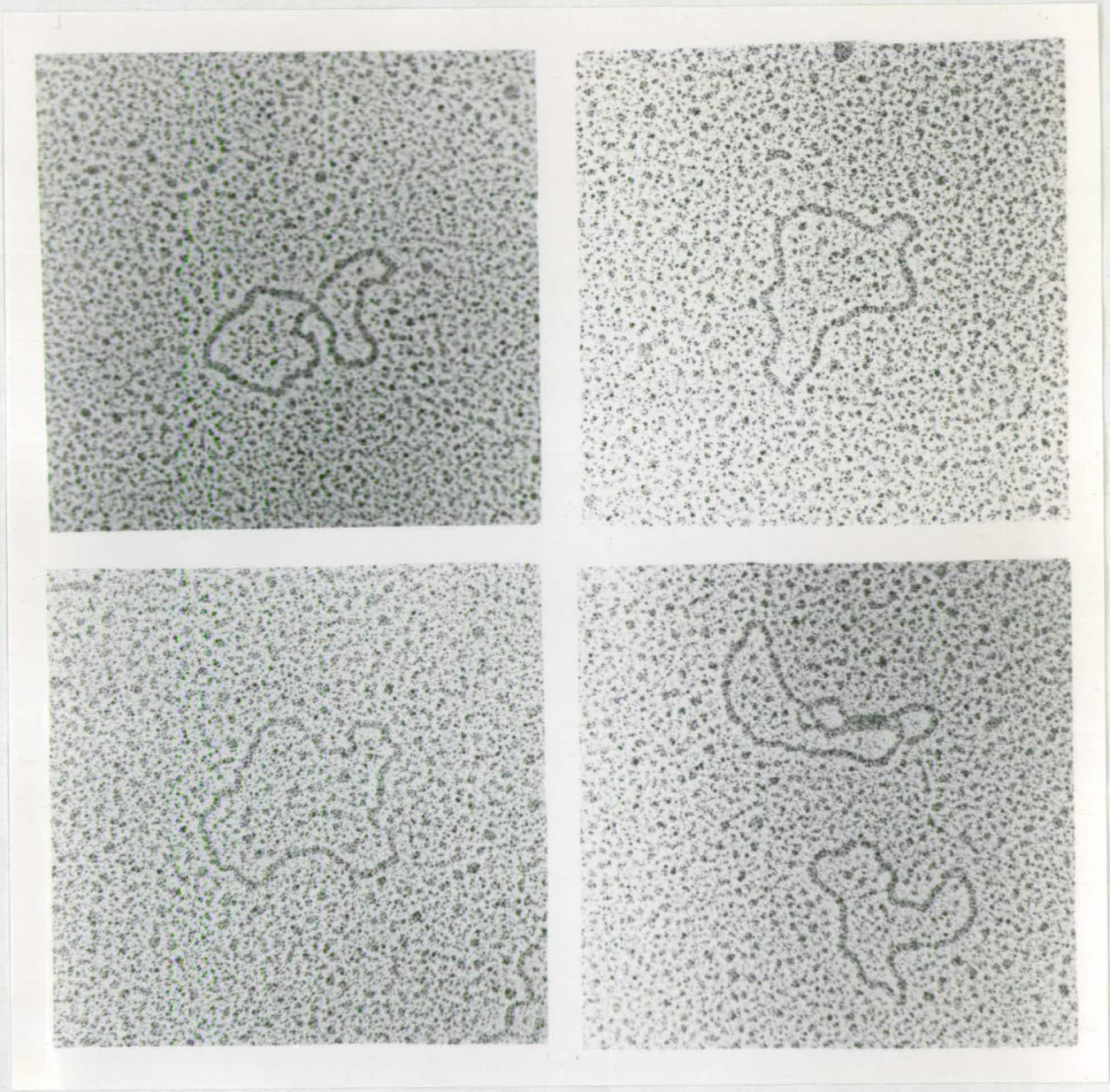


Figure 7.9. Electron micrographs of lambda DNA fragments produced by R.Hin$_d$ III. These fragments have been circularised by intramolecular hydrogen bonding. Magnification approx. 130,000 X

Table 7.3 Molecular structure of fragments produced by R.Hin$_d$ III treatment of lambda DNA

DNA solutions that had been incubated with R.Hin$_d$ III in the presence or absence of EDTA were treated as described in the text and prepared for electron microscopy. Molecules were counted in fields of view chosen at random.

| Treatment | No. of molecules observed | No. of unscorable molecules | No. of circular molecules | Circular molecules as % of total |
|---|---|---|---|---|
| solution (i) kept at 5° | 400 | 30 | 28 | 7% |
| solution (i) heated to 37° | 400 | 13 | 0 | 0% |

(c) 3′-termini of lambda DNA fragments produced by R.Hin$_d$ III

The results of the two previous sections show that the DNA fragments produced by R.Hin$_d$ III have (p) A-G-C-T-T as their 5′-terminal nucleotide sequence, and that the fragments have short cohesive ends. These cohesive ends could consist of either 3′- or 5′-single-stranded projections. Figure 7.10 shows the possible structures for the endonuclease R.Hin$_d$ III substrate region. The structures all have an axis of rotational symmetry. Structures 2 and 3 are unlikely on the basis that they involve a specific sequence of ten or more base pairs. Any particular sequence of ten base pairs would be expected to occur by chance about once every $10^6$ base-apirs in DNA with the composition of lambda DNA. Since both lambda DNA and SV40 DNA are degraded to fragments that are very much smaller than this it is extremely unlikely that structures 2 or 3 describe the endonuclease R.Hin$_d$ III substrate region.

The structure of the termini created by R.Hin$_d$ III was investigated further using reactions catalysed by the T4 DNA polymerase. The 5′-terminal labelling approach showed that the pentanucleotide p A-G-C-T-T was present at the 5′-termini of the fragments. If this is the only pentanucleotide sequence, then the structure of the termini is as follows:

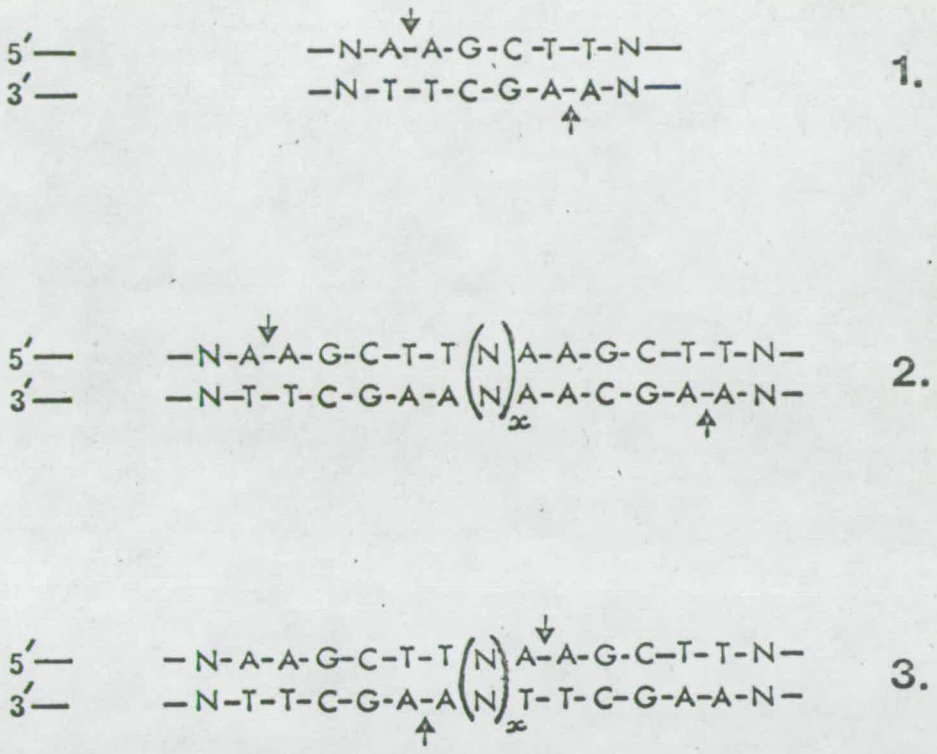



Figure 7.10. Possible structures for the endonuclease R.Hin$_d$ III substrate region. Analysis of the sequence at the 5′-ends of R.Hin$_d$ III product has shown that the enzyme breaks both strands of duplex DNA on the 5′-side of the sequence A-G-C-T-T. The product also has short cohesive ends. The figure shows structures for the endonuclease substrate region consistent with these findings. N represents any of the four possible nucleotides, and does not necessarily imply that they all occur at any particular position.

5′ (p) A-G-C-T-T——

3′ (p) A——

In a reaction with $\alpha$-$^{32}$P-dATP and T4 DNA polymerase a dAMP residue will be incorporated by repair at the 3′-hydroxyl group (after treatment of the fragments with phosphatase if necessary). This situation is similar to that encountered in the reaction with dATP and T4 DNA polymerase at the 3′-hydroxyl group of the r-strand of whole lambda DNA (Chapter 5).

Reactions were performed as follows. Lambda DNA (240 µg) was digested with R.Hin$_d$ III (10 µl) in a 1.0 ml reaction mixture under standard assay conditions. The reaction was stopped by the addition of EDTA, the solution extracted with neutralised redistilled phenol, and dialysed exhaustively against 63 mM Tris-HCl pH 8.0.

The DNA fragments (30 µg) were incubated with T4 DNA polymerase (200 units) at 11° in a solution (0.3 ml) containing 7 mM $MgCl_2$, 7 mM 2-mercaptoethanol, 60 mM Tris-HCl pH 8.0 and 0.1 mM $\alpha$-$^{32}$P-dATP (specific radioactivity about 7 mCi/umole). After incubation for 80 minutes the reaction was stopped by the addition of EDTA (50 µl, 0.4 M), the solution extracted with neutralised redistilled phenol and dialysed against 0.3 M NaCl, 0.01 M Tris-HCl pH 8.0, 0.001 M EDTA. The labelled DNA fragments were purified by chromatography on Sephadex G-100 as described for 5′-terminally labelled DNA.

Samples of labelled DNA were applied to Whatman

GF/C filters. All of the radioactivity (greater than 95%) was acid insoluble, indicating that the purification of the fragments had effectively removed unincorporated precursor.

A nearest-neighbour analysis was performed on the 3′-terminally labelled material as follows. A sample of the DNA was digested to 3′-mononucleotides in a solution (30 µl) containing spleen endonuclease (5 µl; a gift from Dr. G. Bernardi) and spleen exonuclease (0.3 units) in 0.05 M ammonium acetate pH 5.6, 0.001 M EDTA. After incubation at 37° for 1 hr the sample was concentrated by evaporation and applied to Whatman No. 52 paper for ionophoresis at pH 3.5. The 3′-mononucleotides were run in conjunction with 5′-mononucleotides added as u.v. markers (about 20 µl of a solution containing each at 10 mg/ml was necessary with No. 52 paper). The ionophoretogram was also autoradiographed to ensure that the radioactivity in 3′-mononucleotides co-migrated with the 5′-mononucleotide markers. Appropriate regions were cut out and assayed for radioactivity in a gas-flow counter. The results of this analysis are shown in Table 7.4.

The ability of the polymerase to label the fragments indicates that the termini have 3′-hydroxyl groups (Goulian et al., 1968). However, it is not certain that endonuclease R.Hin$_d$ III creates termini with a free 3′-hydroxyl group since the endonuclease preparation was

Table 7.4 3′-terminal nucleotides of DNA fragments produced by R.Hin$_d$ III

The 3′-terminal nucleotides of DNA fragments produced by R.Hin$_d$ III were determined by nearest-neighbour analysis of fragments terminally labelled by incubation with $\alpha$-$^{32}$P-dATP and T4 DNA polymerase. See text.

| 3′-mononucleotide | counts/10 min.* | % total |
|---|---|---|
| Tp | 254 | 6 |
| Gp | 110 | 3 |
| Ap | 3185 | 78 |
| Cp | 518 | 13 |

* A background of 584 counts has been subtracted. This value was obtained by counting an appropriately sized blank region of the ionophoretogram.

contaminated at a low level with phosphatase (data not shown).

A sample of the 3′-terminally labelled material was digested with pancreatic DNase under conditions similar to those used for the 5′-terminally labelled fragments. Radioactive oligonucleotides were fractionated by ionophoresis, and detected by autoradiography (Fig. 7.11).

The nearest-neighbour analysis shows that the 3′-base of the DNA fragments is A (about 80% of the radioactivity was in Ap). In addition, the radioactive oligonucleotides derived from the terminally labelled DNA by pancreatic DNase digestion suggest that the sequence becomes random at the 5′-hexanucleotide position. These results confirm the sequence shown in structure 1 of Figure 7.10. These results do not however distinguish between the possible cohesive end structures shown in Figure 7.10.

It should be noted that although sequence divergence occurs at the 5′-terminal hexanucleotide position, different combinations of nucleotides at the 6th (or higher) positions, on both sides of the axis of symmetry, may affect sequence recognition by the endonuclease. For instance, the requirement for symmetry may extend beyond the 5′-terminal hexanucleotide position. This is a general consideration that applies to all of the published restriction endonuclease substrate sequences. Such effects are most easily studied with a set of

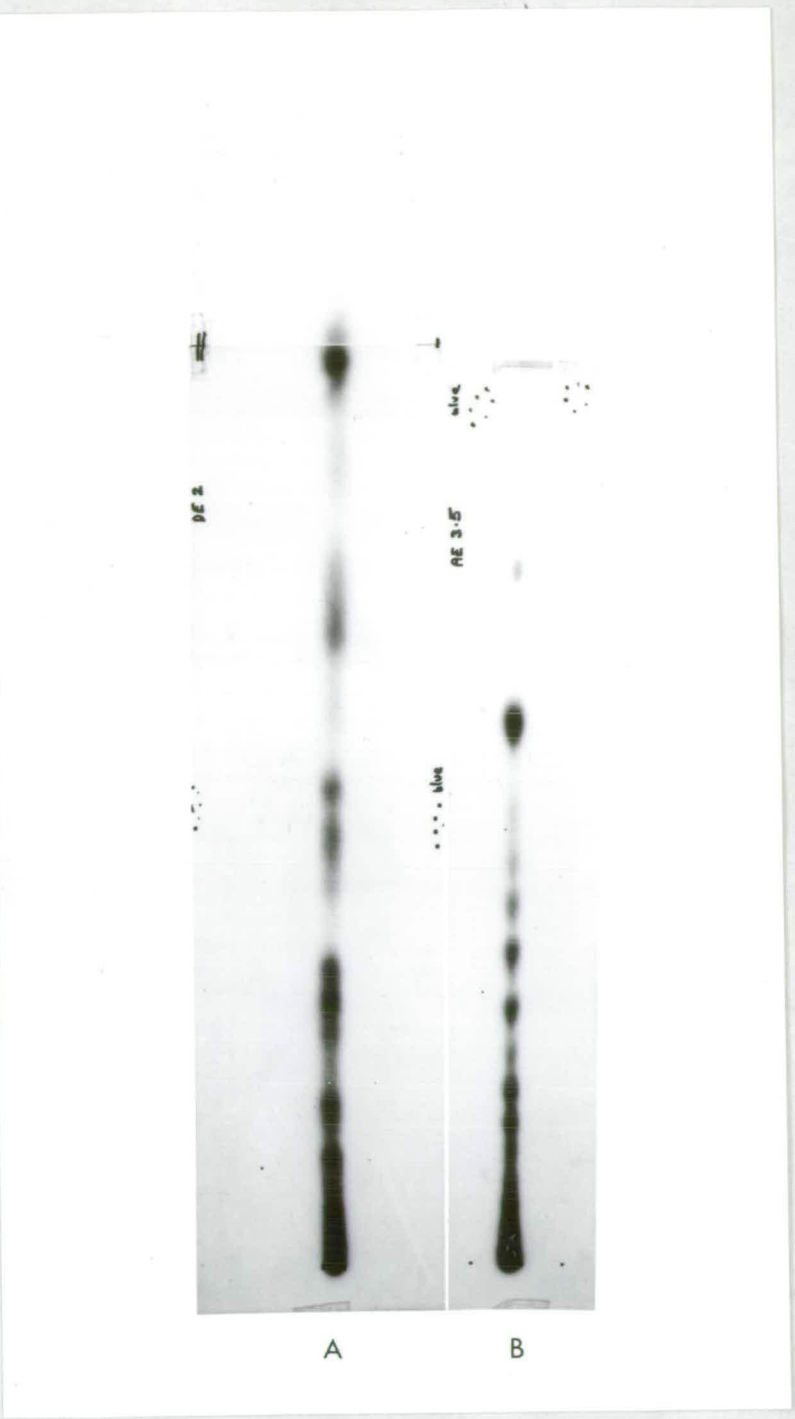



Figure 7.11. Fractionation of nucleotides in digests of lambda DNA fragments produced by R.Hin$_d$ III that have been labelled by incubation with T4 DNA polymerase and $\alpha$-$^{32}$P-dATP.
(A) DE-cellulose paper pH 2. (B) AE-cellulose paper pH 3.5.

single-site DNA substrates. It is unfortunate that the system where a set of such site mutants is available, the E. coli K system, is not amenable to sequence analysis by the polynucleotide kinase approach (Murray et al., 1973).

## 7.5 Resistance of phage T7 DNA

In their studies on the sequence recognised by the endonuclease preparation that was originally called R.H, Kelly and Smith (1970) used T7 DNA as substrate. Although R.Hin$_d$ III was present in R.H its activity was hidden because it is not active on T7 DNA. This is demonstrated in the experiments of Figure 7.12.

Zone sedimentation in neutral sucrose density gradients shows that T7.0 (T7.803), T7.K and T7.B DNAs are resistant to R.Hin$_d$ III, whereas $\lambda$.0 DNA is degraded in the same solution of enzyme. Therefore T7 DNA is neither substrate nor inhibitor. Analysis by gel electrophoresis shows that no small fragment of T7 DNA (down to a molecular weight of about 0.2 x $10^6$) is produced by cleavage at a site near to one end of the molecule. The production of such a fragment might go undetected by the sedimentation analysis.

Lambda DNA has many sites that are susceptible to endonucleolysis by R.Hin$_d$ III. Yet lambda DNA is not very much larger than T7 DNA. The fragments of lambda

Figure 7.12. Resistance of T7 DNA to R.Hin$_d$ III. Sedimentation profiles of mixtures of T7 DNA and lambda DNA incubated for 45 minutes with R.Hin$_d$ III. Conditions for digestion with R.Hin$_d$ III and details of the sedimentation analysis were as described in Methods.

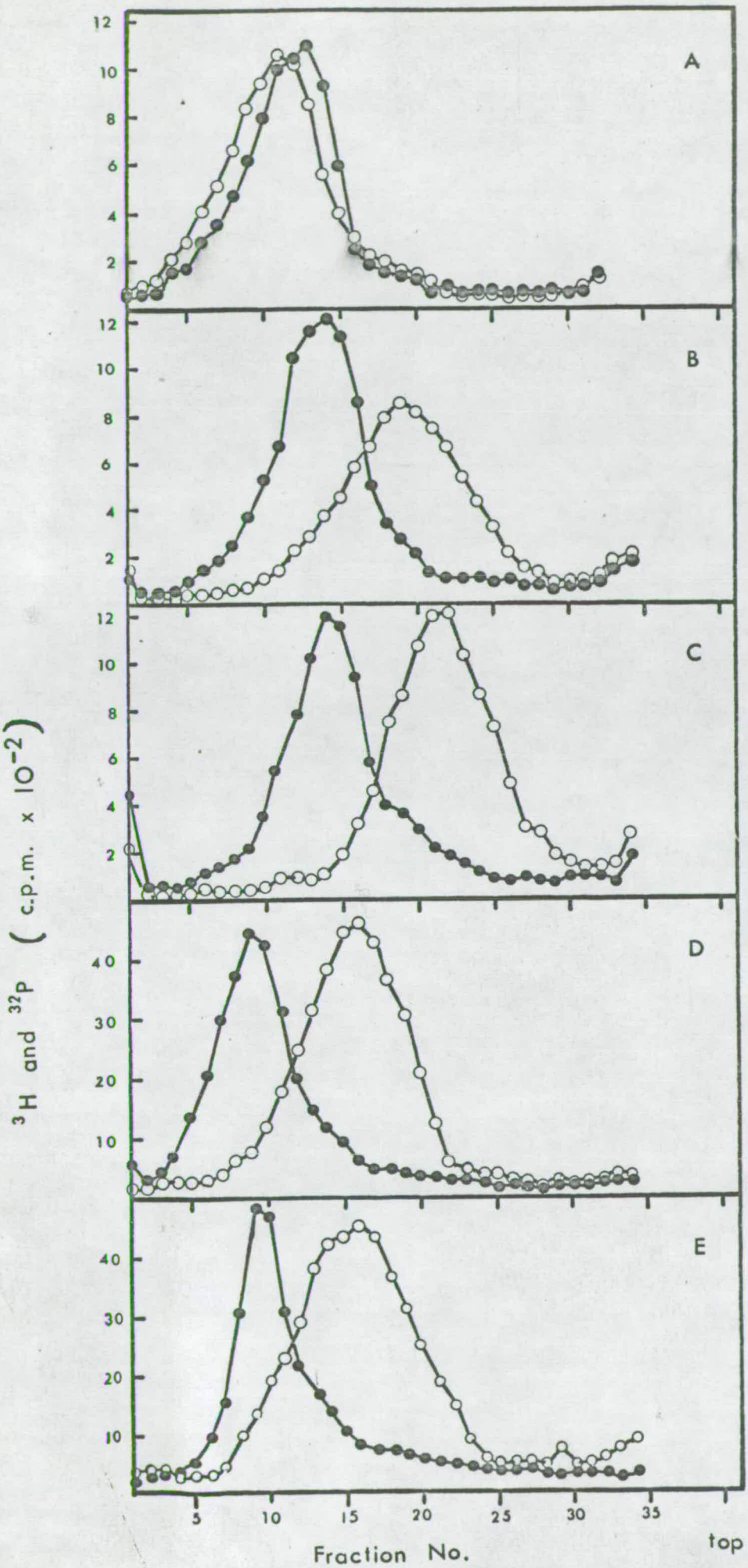
(A) $^{32}P$ λ.0 DNA (0.03 μg) and $^{3}H$ T7.0 DNA (0.12 μg). R.Hin$_d$ III omitted. ——•—— $^{3}H$, ——o—— $^{32}P$.

(B) as in A except for addition of 2 μl R.Hin$_d$ III preparation.

(C) as in A except for addition of 5 μl R.Hin$_d$ III preparation.

(D) $^{32}P$ λ.0 DNA (0.03 ug) and $^{3}H$ T7.K DNA (0.14 μg). 5 μl R.Hin$_d$ III. ——•—— $^{3}H$, ——o—— $^{32}P$.

(E) $^{3}H$ λ.0 DNA (0.1 μg) and $^{32}P$ T7.B (0.09 μg). 5 μl R.Hin$_d$ III. ——•—— $^{32}P$, ——o—— $^{3}H$.


A
B
C
D
E
$^3H$ and $^{32}P$ ( c.p.m. x $10^{-2}$ )
Fraction No.
top

DNA that are produced by the enzyme are very much smaller than whole T7 DNA (Fig. 7.12). The base composition of the DNAs are similar.

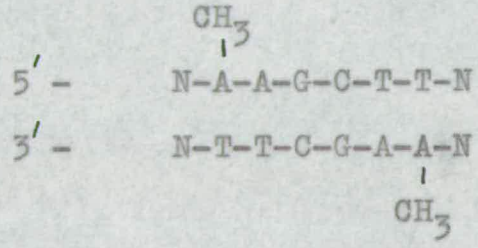
It is extremely unlikely that by chance alone T7 DNA does not contain the recognition sequence. Since DNAs from lambda and T7 phage grown on the same host strain are respectively sensitive and resistant any differential modification of the DNAs must be induced by T7 infection. Roy and Smith (1973) have shown that modification of DNA by the methylase M.Hin$_d$ III is brought about by the methylation (with SAM as methyl group donor) of certain adenine residues, and that this methylase is inactive on T7 DNA. In their survey of the induction of DNA methylases upon infection of E. coli with various phages, Hausmann and Gold (1966) did not detect any induced by T7. Although this negative result is not conclusive, it suggests that the resistance of T7 DNA is not due to modification of sites by transfer of methyl groups from SAM. Other forms of modification cannot be ruled out.

If T7 DNA does indeed contain the recognition sequence (the omission of the hexanucleotide sequence would surely be a constraint requiring an extremely large selection pressure for its maintenance) then these findings are a caution in the use of restriction enzymes as probes for the detection of particular sequences in DNA. Hitherto unrecognised modification of the DNA may lead to an apparent reduction in the frequency of occurrence of the sequence.

## 7.6 Identity of the enzyme

The endonuclease studied in experiments reported in this chapter has been named R.Hin$_d$ III on the basis of the following criteria.

The enzyme is inactive on T7 DNA. This property was inferred by Smith and Nathans (personal communication) in their studies of R.H. They suggested that the component in R.H that accompanied R.Hin$_d$ II should be called R.Hin$_d$ III. The numbers are assigned on the basis of the DNA methylases isolated from H. influenzae Rd. In their recent studies of the bilateral nearest-neighbours of 6-methyl-aminopurine residues in salmon sperm DNA that had been methylated by M.Hin$_d$ III, Roy and Smith (1973a; 1973b) concluded that the methylation occurred within the sequence N-mA-A (where N represents any nucleotide). This site at which M.Hin$_d$ III acts is consistent with the restriction site as follows:




M.Hin$_d$ III does not methylate T7 DNA (Roy and Smith, 1973a; 1973b).

It remains to be shown that methylation by purified M.Hin$_d$ III protects DNA against challenge with R.Hin$_d$ III.

Kelly and Smith (1970) found that treatment of T7 DNA with R.H gave pA and pG as the 5′-terminal nucleo-

tides in the proportions 63% and 37% respectively. However, treatment of phage P22 DNA gave 85% pA and 15% pG (they did not show that the P22 DNA fragments all bore a 5′-phosphoryl group). This higher proportion of pA is explained if both R.Hin$_d$ III and R.Hin$_d$ II are active on P22 DNA.

Treatment of SV40 DNA with R.H produces 11 fragments (Danna and Nathans, 1971), and R.Hin$_d$ II produces 5 fragments (H.O. Smith, personal communication; Danna et al., 1973). R.Hin$_d$ III produces 6 fragments, the sizes of which conform with those expected from the Hin-site map and other results of Danna et al. (1973).

It is significant that R.Hin$_d$ III was recovered in the 0.4 M KCl step in the chromatography upon phosphocellulose that was used in its isolation (Methods). The R.H preparation of Smith and Wilcox, which was a mixture of R.Hin$_d$ II and R.Hin$_d$ III, was taken from the 0.2 M step and some early fractions of the 0.3 M step (Smith and Wilcox, 1970).

# Chapter 8

# DISCUSSION

## 8.1 Nucleotide phosphotransferase

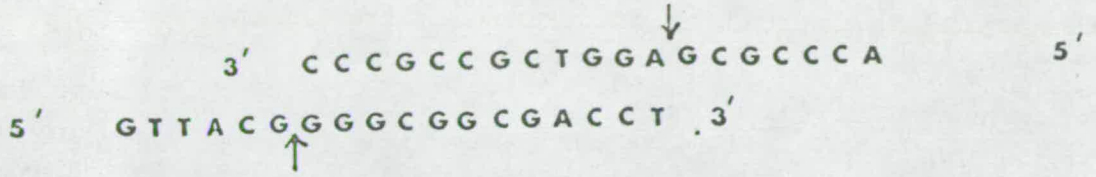
Evidence presented in Chapter 4 shows that a nucleotide phosphotransferase that catalyses phosphorylation of the 3′- (and 2′-) hydroxyl groups of nucleosides and nucleoside-5′-monophosphates also catalyses phosphorylation of dinucleoside monophosphates. However 5′-di- and higher deoxynucleotides are not phosphorylated; nor are sRNA and native DNA phosphorylated. Therefore the enzyme does not provide a method of 3′-terminal labelling in nucleotide sequence determination. However it may be of use in the synthesis of mononucleoside diphosphates of the form pNp and tetraphosphates of the form pppNp (Brunngraber and Chargaff, 1970). Para-nitrophenyl phosphate was the most efficient phosphate group donor tested.

## 8.2 Nucleotide sequence recognition by the lambda Ter endonuclease

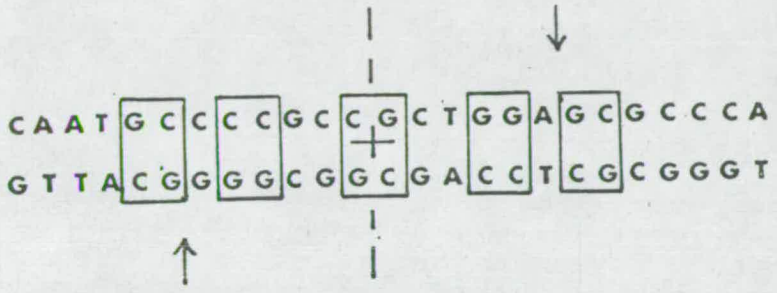
The nucleotide sequence of 25 base pairs that is now known in the region of lambda DNA at which the endonuclease Ter acts is shown in Figure 8.1. The cohesive ends are shown base-paired. The sequence of twelve base pairs between the nicks generated by reaction with the Ter enzyme is bisected by an axis of rotational

symmetry. The symmetry is interrupted by some non-symmetrical base pairs, and extends two base-pairs beyond the nicks that define the limit of the cohesive ends.

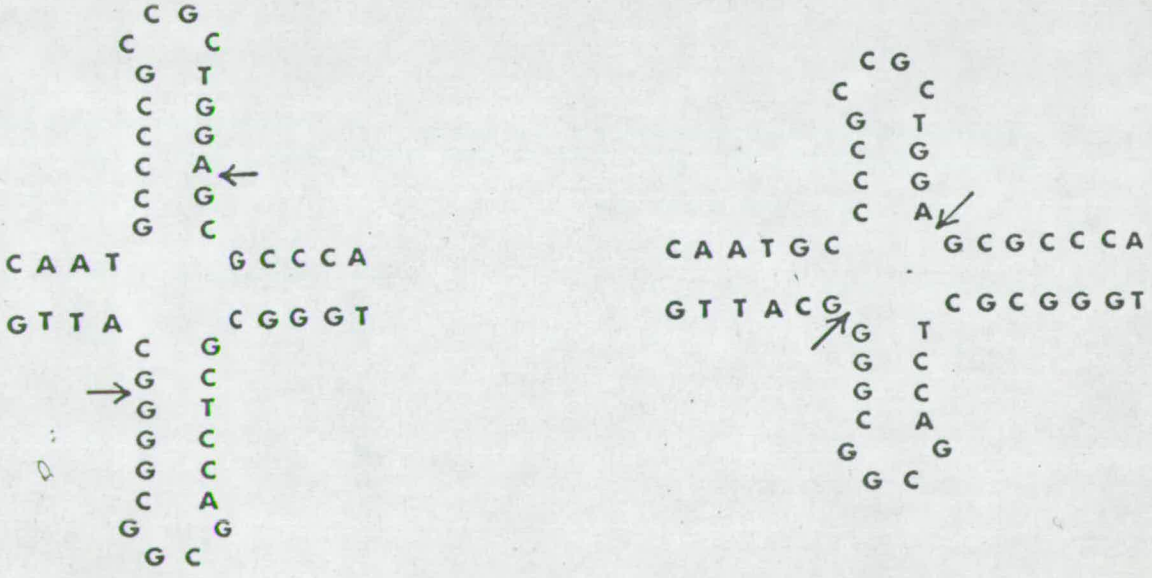
The cohesive end sequences of 4 other lambdoid DNAs that have been examined (Murray and Murray, 1973) are identical with those of lambda, but their 3′ sequences have not yet been determined. However, it has been shown that Ø80 efficiently excises recombinant lambda genomes from tandem lambda dilysogens (Szpirer and Brachet, 1970), which means its Ter mechanism can efficiently recognise the target sequence of lambda. In addition, Ø80 can complement mutants in λ gene A (Sato et al., 1968). If all the symmetrically disposed nucleotides around the cohesive ends are essential for the proper function of the Ter systems one would expect at least the 3′-terminal dinucleotides of Ø80 and lambda DNA to be identical. Preliminary results suggest that the right-hand cohesive ends of these two DNAs have identical 3′-pentanucleotide sequences (Bambara et al., 1973). A study of the homology separated DNA strands of Ø80 and lambda reveals a short region of homology of no more than 50 base pairs at the extreme right-hand end of the heteroduplex. A much longer region of homology exists at the left-hand end. It will be of interest to see the extent of 3′-terminal identity in the DNA of other lambdoid phages and also whether their Ter systems




The Terminal Nucleotide Sequences of λ DNA.




Symmetry Around the Cohesive Ends.




Possible Transient Structures of the Cohesive End Region.

Figure 8.1 .

cross react with each other, for a set of such sequences could define those nucleotides that are part of the recognition sequence.

The abundance of G.C pairs in the region of lambda DNA at which Ter acts greatly increases the probability of a fortuitous symmetrical arrangement of base pairs about a central axis. However, the two-fold symmetry does suggest mechanisms by which the Ter enzyme might recognise this sequence.

Any sequence that has such symmetry may be drawn in the form of a hydrogen-bonded structure with two hairpin loops; the size of the structure clearly depends on the extent of the symmetrical sequence. Although the sequence of the cohesive end region of lambda is not perfectly symmetrical, it is still possible to make loops from them if some non-standard base-pairs are involved. Two of several possible structures of this type are shown in Figure 8.1. They provide a prominent physical feature for recognition by the enzyme, and bring the phosphodiester bonds that are broken closer together, but would probably be too unstable to exist even transiently (Tinoco et al., 1971).

The symmetry of the nucleotide sequence itself may be the basis for the recognition by the Ter endonuclease. If, for instance, the enzyme were a dimer (or higher even-numbered oligomer) of symmetrical subunits arranged about an axis of rotational symmetry, it could bind to

the DNA with its axis aligned with that of the symmetrical sequence, and symmetrically disposed catalytic sites could introduce the two nicks simultaneously.

A hyphenated sequence of 10 base pairs (like any other decanucleotide sequence) would be expected to occur by chance about once every $10^6$ base pairs, so it is improbable that it would occur more than once in lambda DNA (about 46,500 base pairs). However the sequence of five base pairs that would be associated with each sub-unit of the enzyme would be expected to occur about fifty times in a molecule the size of lambda DNA. The enzyme would therefore be expected to be active only as a dimer. There is good reason to believe that oligomeric proteins composed of identical subunits display rotational symmetry (Monod, Wyman and Changeux, 1965). Wang and Kaiser (1973) have shown that the Ter activity can be detected in crude extracts of lambda-infected cells. It will be of interest to see if this prediction of the enzyme structure is upheld.

A space-filling model of the sequence (Fig. 8.2) shows that the phosphodiester bonds to be broken and the symmetrical base pairs (and therefore their functional groups that may confer specificity) can be 'seen' from one aspect of the helix. One sub-unit of the enzyme could interact with base paid 1 via the minor groove of the helix and base pairs 4, 5, 7 and 8 via the major

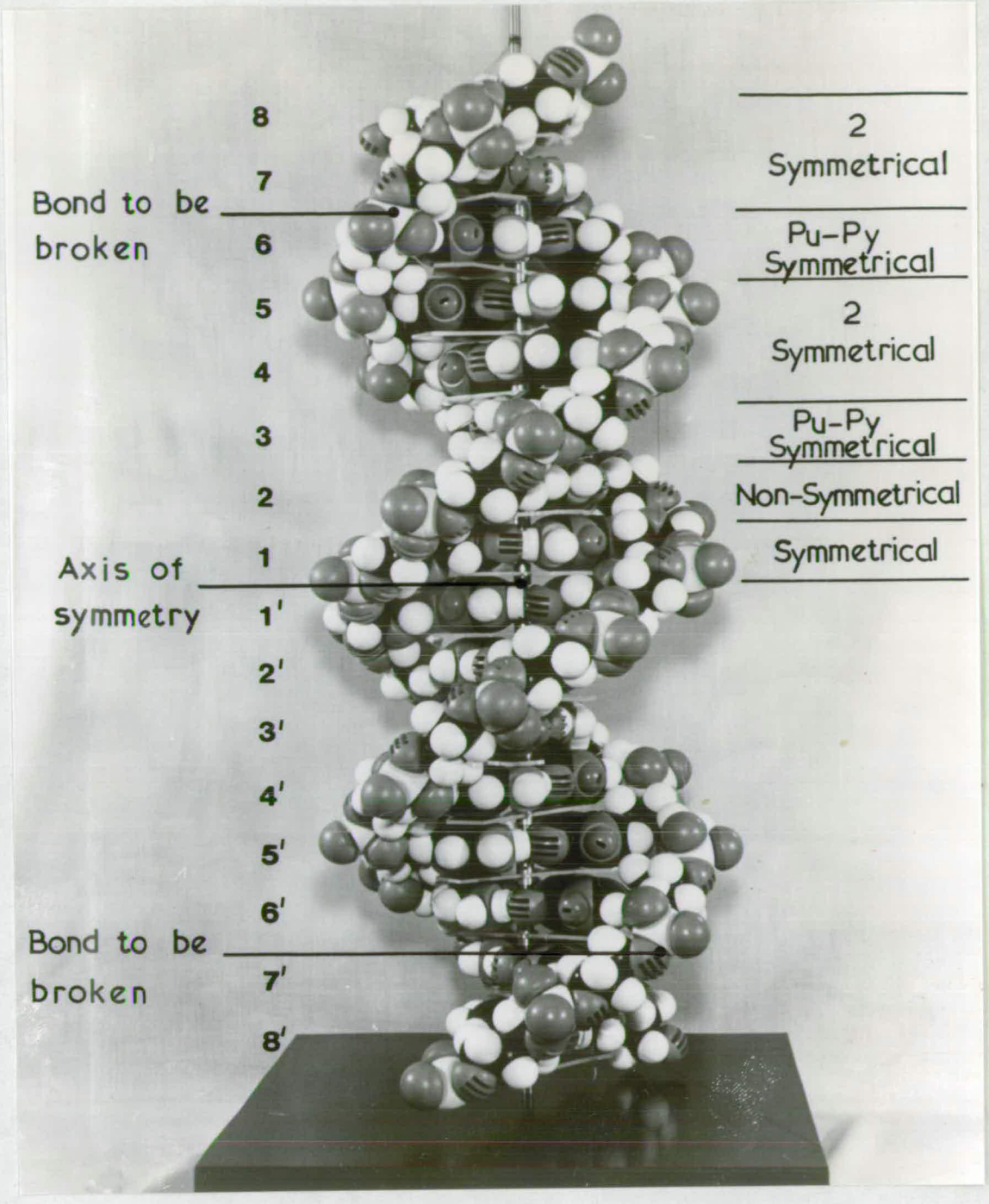



Figure 8.2. Model of the cohesive end region of lambda DNA. Numbering of the base pairs is the same as in Figure 5.2. The 2-fold axis of symmetry is perpendicular to the page. The model was constructed by Drs. P. Englund and P. Weigel, who kindly supplied the original photograph.

groove, and the other subunit with the corresponding prime-numbered bases. (Base-pairs 8 and 8′ fall within the compass of a protein comprising two identical subunits of molecular weight 50,000.) It is possible that base pairs symmetrical only in their purine-pyrimidine orientation are also involved in specific interaction with the enzyme. It is also possible that this sequence may dictate local stereochemical characteristics of the helix that differ from those of the B form (Bram and Tougard, 1972), and that these characteristics may be recognised by the enzyme.

Recently Gilbert and Maxam (1973) have isolated microgram amounts of the E. coli DNA fragment protected by the lac repressor from digestion with DNase. The sequence of this fragment, determined indirectly via its in vitro transcription product, includes the symmetrical sequence shown in Figure 8.3. The symmetry is hyphenated.

More is known of the lac operator region through sequence analysis of lac mRNA transcribed in vitro (Gilbert and Maizels, 1973), for the operator sequence can be recovered at the beginning of the lac message. Gilbert et al. have extended their studies to operator constitutive mutants. One of these has a G.C to A.T base change in the left-hand symmetry region of the operator (Fig. 8.3).

Since both the Ter endonuclease and the lac repressor recognise a sequence having hyphenated

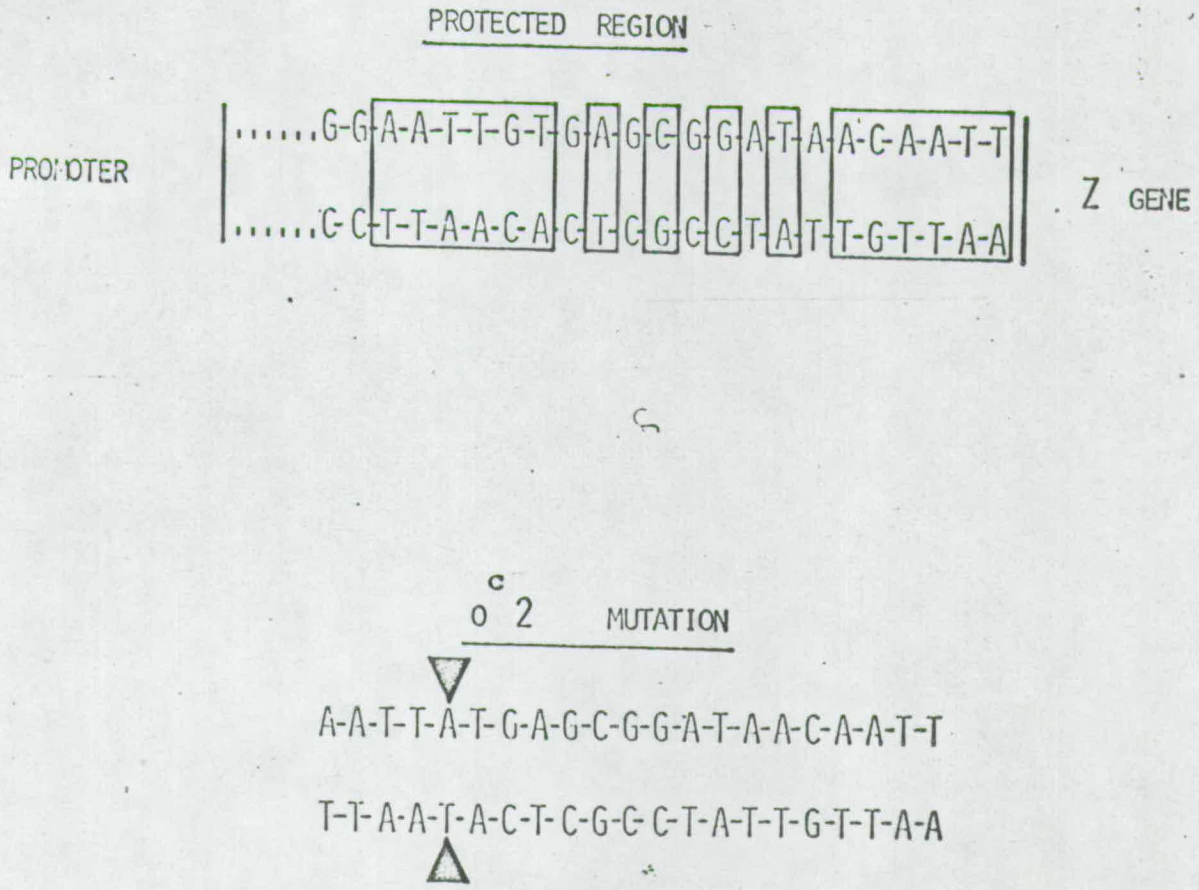



Figure 8.3. Nucleotide sequences in the region of the lac operator.

symmetry, and since both recognition regions are of about the same size, it therefore seems likely that similar recognition mechanisms are involved.

It is of interest that, in the binding form, the *lac* repressor exists as a tetramer of identical subunits with a monomer molecular weight of about 38,000; that the Arrhenius activation energy of binding is very low (about 8 Kcal/mole) making any model involving enzyme-induced looping out or denaturation of extensive regions of the duplex unlikely (Riggs *et al.*, 1970); and that a single base-change that upsets the hyphenated symmetry leads to constitutivity. These properties are consistent with models in which the symmetry of the sequence itself is the basis for recognition by a protein having an axis of rotational symmetry.

The hyphenated symmetry at the cohesive end region of lambda DNA is not a property that is shared by the terminal sequences of the non-lambdoid temperate phages (Murray and Murray, 1973; Padmanabhan and Wu, 1972; Wang and Brezinski, 1973): there is no axis of rotational symmetry when the cohesive ends are paired; in fact the near absence of symmetry in the sequences is rather surprising. This means that formation of a looped structure, in these cases, is virtually impossible, for even if the sequences beyond the 3′-termini proved to be rotationally symmetrical, an unpaired loop of about twenty base pairs would be extremely destabilising.

Direct recognition of the base sequence in the duplex appears the most favourable mode of recognition in these examples.

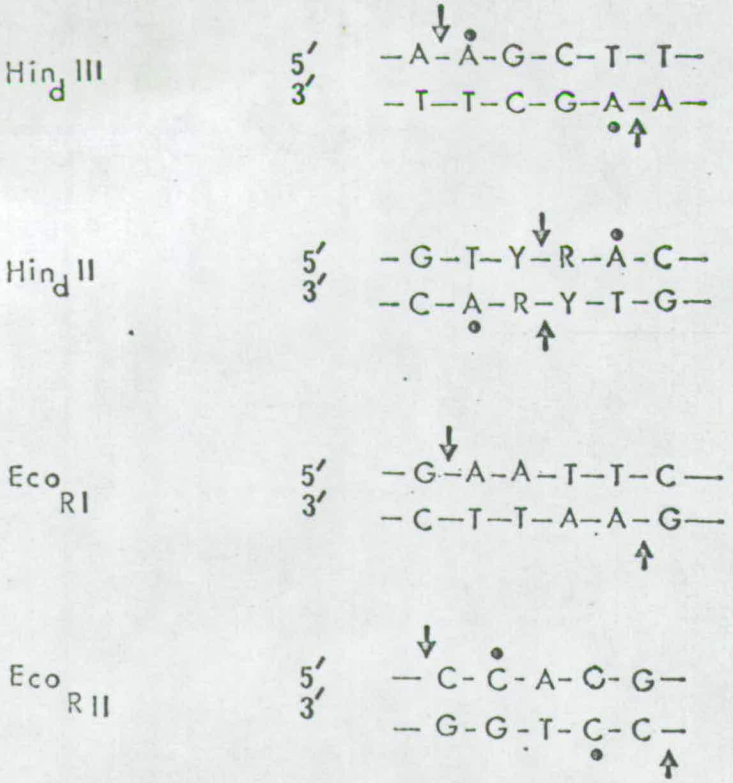
Wang and Brezinski have proposed a model involving the alignment of two copies of cohered ends (in the same or different molecules) with their helix axes anti-parallel. The duplex sequences now become rotationally symmetrical. The proper alignment would be mediated by a protein that has a DNA recognition site and is active in the dimeric state. The DNA sequence recognised by the protein is not defined by the model, any part of the sequence common to both 186 and P2 DNA in the region of the cohesive ends could be involved (P2 and 186 appear to have Ter function that cross react). However it may be noted that the sequence G-G-C-G-G occurs in the l-strand cohesive end of 186, P2 and lambda DNA, so that if the Ter systems of all three are related this pentamer may be one of the essential elements involved in recognition (by itself, it would be insufficient). This interesting speculative model is applicable to processes such as integration and excision of prophage as well as to Ter function.

## 8.3 Nucleotide sequence recognition by R.Hin$_d$ III and other restriction and modification enzymes

Nucleotide sequences around the sites of reaction of some restriction and modification enzymes are shown in Figure 8.4, together with the R.Hin$_d$ III sequence. All of these sequences either have an axis of rotational symmetry or can be included within a corresponding sequence that has rotational symmetry (e.g. N-mA-A in the case of M.Hin$_d$ III). The R.Hin$_d$ II sequence has a central ambiguity and it is not known if the symmetry of a single site is always exact.

Since all the known sequences at the sites of reaction of restriction and modification enzymes have rotational symmetry it is likely that this reflects one rather general mechanism of nucleotide sequence recognition. Whether the restriction endonucleases R.Eco$_K$, R.Eco$_B$ and R.Eco$_{P1}$ conform to this mechanism is not known because the nucleotide sequence(s) at the site of endonucleolysis of any of these three has yet to be determined. However, the results of Brockes (1972) on the sequence(s) around the base methylated by M.Eco$_{P1}$ suggest that a symmetrical sequence may be involved in modification. R.Eco$_{P1}$ must distinguish between modified and unmodified DNA and therefore must recognise and interact with the symmetrical sequence. Genetic studies indicate that the restriction enzyme has (at least) one subunit

References:




This thesis; Roy & Smith (1973b)

Kelly & Smith (1970); Roy & Smith (1973b)

Hedgpeth et al. (1973)

Bigger et al. (1973); Boyer et al. (1973)

Figure 8.4 . Nucleotide sequences around the sites of action of some restriction and modification enzymes. Arrows show the position of endonucleolysis. Dots indicate sites of modification methylation.

in common with the modification enzyme, and this may carry out the recognition or specificity function (Wood, 1966).

$R.Eco_{P1}$ appears to resemble $R.Eco_K$ and $R.Eco_B$ in making each double-strand break at one of a number of sites (Horiuchi and Zinder, 1972; Mulder and Delius, 1972). Its ability to produce lambda DNA fragments that form bands on gel electrophoresis means that either these sites are much fewer in number than those available to $R.Eco_K$, or that they are grouped in clusters (within about 1000 base pairs). On the hypothesis that $R.Eco_K$, $R.Eco_B$ and $R.Eco_{P1}$ travel along the DNA from the (unmodified) modification sequence to a site of endonucleolysis, these results would be explained if the sites available to $R.Eco_{P1}$ occur closer to its modification sequence than is the case for $R.Eco_K$. The massive ATP hydrolysis that is associated with restriction by $R.Eco_K$, and its absence in the case of $R.Eco_{P1}$ (Yuan, Heywood and Meselson, 1972) may be phenomena that are related to this difference.

Results presented in Chapter 6 show that upon restriction with $R.Eco_{P1}$ no new terminal sequences are created that can be detected with the polynucleotide kinase terminal labelling approach. Although the interpretation of the results is made difficult by the increase in 'background' labelling that occurs in this type of experiment, the results suggest that the newly

created 5′-termini are not substrates for polynucleotide kinase. The possibility that R.Eco$_{P1}$ gives rise to a virtually random set of 5′-termini that are substrates for polynucleotide kinase cannot be absolutely excluded. Murray et al. (1973) have obtained similar results with R.Eco$_K$. These results are consistent with the hypothesis, discussed in the Introduction, in which endonucleolysis occurs at a site of depurination after methylation of a purine residue by the restriction endonuclease. However, R.Eco$_{P1}$ is active without the addition of SAM to the reaction mixture, although its activity is stimulated by the addition of SAM. It is quite possible that SAM remains bound to the enzyme throughout its isolation from the cell extract. Alternatively, the depurination hypothesis may still hold, with the labilisation of the N-deoxyriboside bond being independent of methyl group transfer from SAM. An investigation of the SAM-dependent methylation reactions catalysed by R.Eco$_{P1}$ is in progress in this laboratory.

Of the sequences shown in Figure 8.4, that of R.Hin$_d$ II was the first to be determined. Kelly and Smith (1970) noted that the presence of a symmetrical recognition site allows the same protein to recognise and nick opposite strands of the DNA duplex. If the recognition process involved single strands, and if the sequence were asymmetrical, two proteins with different recognition specificities would be required. Kelly and

Smith also point out that the symmetry of the recognition sequence may also reflect underlying symmetry in the enzyme. Much of the discussion of the recognition process of the Ter endonuclease applies directly to restriction endonucleases, and indeed was stimulated by the ubiquity of symmetrical sequences. Preliminary evidence indicates that the R.Hin$_d$ II enzyme consists of four subunits, each of molecular weight 23,000 (Smith, personal communication). The R.Hin$_d$ III sequence, together with the others shown in Figure 8.4, indicate that the symmetry of the sequence recognised by R.Hin$_d$ II is not an isolated phenomenon.

It should be noted that whereas all the sequences shown in Figure 8.4 have the property of rotational symmetry in common, there are variations in the positions (with respect to the axis of symmetry) at which phosphodiester bond breakage and modification occur.

The staggered breaks made by certain restriction endonucleases, including R.Hin$_d$ III, give rise to DNA fragments with short cohesive termini. These DNA fragments can be recombined <u>in vitro</u>, to give rise to hybrid molecules that are derived from genomes that may not normally recombine in nature. The DNA fragments can be covalently joined in reactions catalysed by DNA ligase (Sgaramella, 1972). The availability of a range of restriction endonucleases that create cohesive termini not only provides a means of specifically

degrading DNA to give regions of defined genetic content; but also provides a biochemical means by which new functionally defined segments of genetic information can be introduced into the cell.

## REFERENCES


Appleyard, R.K., McGregor, J.F. & Baird, K.M. (1965) Virology 2, 565-574.

Appleyard, R.K. (1954) Genetics 39, 440-452.

Allet, B., Jeppesen, P.G.N., Katagiri, K.J. & Delius, H. (1973) Nature New Biol. 241, 120-122.

Arber, W. (1965) Annu. Rev. Microbiol. 19, 365-378.

Arber, W. (1971) in The Bacteriophage Lambda, Cold Spring Harbor Lab., New York.

Arber, W. (1973) Progress in Nucleic Acid Research and Molecular Biology 14, in press.

Arber, W. & Dussoix, D. (1962) J. Mol. Biol. 5, 18-36.

Arber, W. & Linn, S. (1969) Annu. Rev. Biochem. 467-500.

Arber, W. & Morse, M.L. (1965) Genetics 51, 137-148.

Arber, W. & Wauters-Willems, D. (1970) Molec. Gen. Genet. 108, 203-217.

Arrand, J.R. & Hindley, J. (1973) Nature New Biol. 244, 10-13.

Bambara, R., Ghangas, G.S., Jay, E. & Wu, R. (1973) Fed. Proc. 32, 2537.

Barrell, B.G. (1971) in Procedures in Nucleic Acid Research, Vol. 2, 751-779.

Benzinger, R. (1968) Proc. Nat. Acad. Sci. U.S. 59, 1294-1299.

Bigger, C.H., Murray, K. & Murray, N.E. (1973) Nature New Biol. 244, 7-10.

Bode, V.C. & Kaiser, A.D. (1965) J. Mol. Biol. 14, 399-417.

Boyer, H.W. (1971) Annu. Rev. Microbiol. 25, 153-176.

Boyer, H.W., Chow, L.T., Dugaiczyk, A., Hedgpeth, J. & Goodman, H.M. (1973) Nature New Biol. 244, 40-44.

Bram, S. & Tougard, P. (1972) Nature New Biol. 239, 128-131.

Brawerman, G. & Chargaff, E. (1955) Biochim. Biophys. Acta 16, 524-532.

Britten, R.J. & Davidson, E.H. (1969) Science 165, 349-357.

Britten, R.J. & Roberts, R.B. (1960) Science 131, 32-33.

Brockes, J.P., Brown, P.R. & Murray, K. (1972) Biochem. J. 127, 1-10.

Brockes, J.P. (1972) Ph.D. thesis, Univ. of Edinburgh.

Bron, S. (1973) Ms. in preparation.

Brown, D.M. & Todd, A.R. (1952) J. Chem. Soc. 52-58.

Brownlee, G.G. (1971) in Laboratory Techniques in Biochemistry and Molecular Biology (Work, T.S. & Work, E. (eds.) Vol. 3, pt. 1, North-Holland Publishing Co., Amsterdam.

Brunngraber, E.F. & Chargaff, E. (1970) Proc. Nat. Acad. Sci. U.S. 67, 107-112.

Burgess, R.R. (1969) J. Biol. Chem. 244, 6160-6167.

Carter, B.J., Shaw, B.D. & Smith, M.G. (1969) Biochim. Biophys. Acta 195, 494-505.

Cory, S., Spahr, P.F. & Adams, J.M. (1970) Cold Spr. Harb. Symp. Quant. Biol. 35, 1-23.

Crawford, L.V. & Black, P.H. (1964) Virology 24, 388-392.

Crick, F.H.C. (1971) Nature 234, 25-27.

Danna, K.J. & Nathans, D. (1971) Proc. Nat. Acad. Sci. U.S. 68, 2913-2917.

Danna, K.J. & Nathans, D. (1972) Proc. Nat. Acad. Sci. U.S. 69, 3097-3100.

Danna, K.J., Sack, G.H. & Nathans, D. (1973) J. Mol. Biol. 78, 363-376.

Durwald, H. & Hoffman-Berling (1968) J. Mol. Biol. 34, 331-346.

Dussoix, D. & Arber, W. (1962) J. Mol. Biol. 5, 37-49.

Edgell, M.H., Hutchison, C.A. & Sclair, M. (1972) J. Virology 9, 574-582.

Englund, P.T. (1971a) J. Biol. Chem. 246, 3269-3276.

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

Eskin, B. & Linn, S. (1972a) J. Biol. Chem. 247, 6183-6191.

Eskin, B. & Linn, S. (1972b) J. Biol. Chem. 247, 6192-6196.

Gellert, M. (1967) Proc. Nat. Acad. Sci. U.S. 57, 148-155.

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

Gilbert, W. & Dressler, D. (1968) Cold Spr. Harb. Symp. Quant. Biol. 33, 473-484.

Gilbert, W. & Maizels, N. (1973) Ms. in preparation.

Gilbert, W. & Maxam (1973) Ms. in preparation.

Gilbert, W. & Muller-Hill, B. (1970) in The Lac Operon, Cold Spring Harbor Lab., New York.

Glover, S.W. & Colson, C. (1969) Genet. Res. 13, 227-240.

Glover, S.W. & Piekarowicz, A. (1972) Biochem. Biophys. Res. Comm. 46, 1610-1617.

Goldberg, A.R. & Howe, M. (1969) Virology 38, 200-202.

Goldmark, P.J. & Linn, S. (1970) Proc. Nat. Acad. Sci. U.S. 67, 434-441.

Goodman, H.M., Billeter, M.A., Hindley, J. & Weissmann, C. (1969) Nature 224, 1083-1086.

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

Gromkova, R. & Goodgal, S.H. (1972) J. Bact. 109, 987-992.

Gromkova, R., Laporte, J. & Goodgal, S.H. (1973) Abstracts of the annual meeting of the American Society for Microbiology, 73rd annual meeting, p. 72.

Haberman, A., Heywood, J. & Meselson, M. (1972) Proc. Nat. Acad. Sci. U.S. 69, 3138-3141.

Hausmann, R. & Gold, M. (1966) J. Biol. Chem. 241, 1985-1994.

Hedgpeth, J., Goodman, H.M. & Boyer, H.W. (1972) Proc. Nat. Acad. Sci. U.S. 69, 3448-3452.

Hershey, A.D. (1971) The Bacteriophage Lambda, Cold Spring Harbor Lab., New York.

Hershey, A.D., Burgi, E. & Ingraham, L. (1963) Proc. Nat. Acad. Sci. U.S. 49, 748-755.

Horiuchi, K. & Zinder, N.D. (1972) Proc. Nat. Acad. Sci. U.S. 69, 3220-3224.

Josse, J., Kaiser, A.D. & Kornberg, A. (1969) J. Biol. Chem. 236, 864-875.

Jovin, T.M., Englund, P.T. & Bertsch, L.L. (1969) J. Biol. Chem. 244, 2996-3008.

Kaiser, A.D. (1962) J. Mol. Biol. 4, 275-287.

Kaiser, A.D. & Hogness, D.S. (1960) J. Mol. Biol. 2, 392-415.

Kato, K.I., Goncalves, J.M., Houts, G.E. & Bollum, F.J. (1967) J. Biol. Chem. 242, 2780-2789.

Kelly, T.J. & Smith, H.O. (1970) J. Mol. Biol. 51, 393-409.

Kiger, J. & Sinsheimer, R. (1969) J. Mol. Biol. 40, 467-490.

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

Kuhnlein, U. & Arber, W. (1972) J. Mol. Biol. 63, 9-19.

Lang, D. & Mitani, M. (1970) Biopolymers 9, 373-379.

Lautenberger, J.A. & Linn, S. (1972) J. Biol. Chem. 247, 6176-6182.

Lawley, P.D. (1966) Progress in Nucleic Acid Research and Molecular Biology, Vol. 5.

Lederberg, S. & Meselson, M. (1964). J. Mol. Biol. 8, 623-628.

Lehman, I.R., Roussos, G.G. & Pratt, E.A. (1962) J. Biol. Chem. 237, 819-828.

Linn, S. & Arber, W. (1968) Proc. Nat. Acad. Sci. U.S. 59, 1300-1306.

Lowry, O.H., Rosebrough, N.J., Farr, A.L. & Randall, R.J. (1951) J. Biol. Chem. 193, 265-275.

Mertz, J.E. & Davis, R.W. (1972) Proc. Nat. Acad. Sci. U.S. 69, 3370-3374.

Meselson, M. & Yuan, R. (1968) Nature 217, 1110-1114.

Meselson, M., Yuan, R. & Heywood, J. (1972) Annu. Rev. Biochem. 41, 447-466.

Middleton, J.H., Edgell, M.H. & Hutchison, C.A. (1972) J. Virology 10, 42-49.

Monod, J., Wyman, J. & Changeux, J. (1965) J. Mol. Biol. 12, 88-118.

Morrison, A. & Murray, K. (1973) Biochem. J. Ms. in preparation.

Morrow, J.F. & Berg, P. (1972) Proc. Nat. Acad. Sci. U.S. 69, 3365-3369.

Mulder, C. & Delius, H. (1972) Proc. Nat. Acad. Sci. U.S. 69, 3215-3219.

Murray, K. (1970) Biochem. J. 118, 831-841.

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

Murray, K. & Murray, N.E. (1973) Nature New Biol. 243, 134-139.

Murray, K. & Old, R.W. (1973) Progress in Nucleic Acid Research and Molecular Biology 14, in press.

Murray, N.E., Batten, P.L. & Murray, K. (1973) J. Mol. Biol., in press. (Vol. 81)

Murray, N.E., Manduca de Ritis, P., Foster, L. (1972) Mol. Gen. Genet. 120, 261-281.

Nathans, D. & Smith, H.O. (1973) Ms. in preparation.

Novogrodsky, A. & Hurwitz, J. (1966) J. Biol. Chem. 241, 2923-2932.

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

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

Piekarowicz, A. & Glover, S.W. (1972) Molec. Gen. Genet. 116, 11-25.

Randerath, K. & Randerath, E. (1967) Methods in Enzymology 12A, 323-347.

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

Richardson, C.C. (1971) in Procedures in Nucleic Acid Research, Vol. 2 (Cantoni, G.L. & Davies, D.R. (Eds.)) Harper and Row, New York.

Riggs, A.D., Bourgeois, S. & Cohn, S. (1970) J. Mol. Biol. 53, 401-417.

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

Roulland-Dussoix, D. & Boyer, H.W. (1969) Biochim. Biophys. Acta 195, 219-229.

Roy, P. & Smith, H.O. (1973a) J. Mol. Biol., in press.

Roy, P. & Smith, H.O. (1973b) J. Mol. Biol., in press.

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

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

Schnos, M. & Inman, R. (1970) J. Mol. Biol. 51, 61-73.

Sgaramella, V. (1972) Proc. Nat. Acad. Sci. U.S. 69, 3389-3393.

Sharp, P.A., Sugden, W. & Sambrook, J. (1973) Biochem., in press.

Skalka, A. (1971) in The Bacteriophage Lambda, Cold Spring Harbor Lab., New York.

Smith, H.O. & Wilcox, K.W. (1970) J. Mol. Biol. 51, 397-391.

Smith, J.D., Arber, W. & Kuhnlein, U. (1972) J. Mol. Biol. 63, 1-8.

Sobell, H.M. (1973a) Adv. in Genetics 18, in press.

Sobell, H.M. (1973b) Progress in Nucleic Acid Research and Molecular Biology 14, in press.

Strack, H.B. & Kaiser, A.D. (1965) J. Mol. Biol. 12, 36-49.

Studier, F.W. (1965) J. Mol. Biol. 11, 373-390.

Symons, R.H. (1968) Biochim. Biophys. Acta 155, 609-610.

Symons, R.H. (1969) Biochim. Biophys. Acta 190, 548-550.

Szekely, M. (1971) in Procedures in Nucleic Acid Research, Vol. 12 (Cantoni, G.L. & Davies, D.R. (Eds.)) Harper and Row, New York.

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

Szpirer, J. & Brachet, P. (1970) Molec. Gen. Genet. 108, 78-92.

Tinoco, I., Uhlenbeck, O.C. & Levine, M.D. (1971) Nature 230, 362-367.

Tomizawa, J. & Ogawa, T. (1968) Cold Spr. Harb. Symp. Quant. Biol. 33, 533-551.

Tomlinson, R.V. & Tener, G.M. (1963) Biochem. 2, 697-702.

Trilling, D.M. & Aposhian, H.V. (1968) Proc. Nat. Acad. Sci. U.S. 60, 214-221.

Vogel, F. (1964) Nature 201, 847.

Wang, J.C. & Brezinski, D.P. (1973) Proc. Nat. Acad. Sci. U.S., in press.

Wang, J.C. & Kaiser, A.D. (1973) Nature New Biol. 241, 16-17.

de Wachter, R. & Fiers, W. (1971) Methods in Enzymology 21D, 167-178.

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

Wells, R.D., Larson, J.E., Grant R.C., Shortle, B.E., Cantor, C.R. (1970) J. Mol. Biol. 54, 465 - 497.

Wiberg, J.S., Dirksen, M-L., Epstein, R.H., Luria, S.E. & Buchanan, J.M. (1962) Proc. Nat. Acad. Sci. U.S. 48, 293-301.

Wood, W.B. (1966) J. Mol. Biol. 16, 118-133.

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

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

Yoshimori, R.N. (1971) Ph.D. thesis, University of California.

Young, E.T. & Sinsheimer, R.L. (1964) J. Mol. Biol. 10, 562-564.

Yuan, R., Heywood, J. & Meselson, M. (1972) Nature New Biol. 240, 42-43.

Ziff, E.B., Sedat, J.W. & Galibert, F. (1973) Nature New Biol. 241, 34-37.

# Appendix

## 5′-TERMINAL SEQUENCE OF DNA FRAGMENTS PRODUCED BY R.Hae

Middleton et al. (1972) have described a restriction endonuclease from H. aegyptius which cleaves the double-stranded replicative form DNA of phage ØX174 into 11 specific limit fragments.

The polynucleotide kinase reaction has been used in the determination of the 5′-terminal sequence of phage lambda DNA fragments produced by treatment with R.Hae. A sample of endonuclease R.Hae preparation was the gift of Dr. D. Gould and H. Cook. The enzyme was prepared from H. aegyptius by a modification of the procedure of Middleton et al. (1972).

Lambda DNA (220 µg) was digested in a reaction mixture (1 ml) containing 7 mM Tris-HCl pH 7.4, 7 mM $MgCl_2$ and 7 mM 2-mercaptoethanol and 10 µl R.Hae preparation. After incubation at 37° for 2 hr the reaction was stopped by the addition of EDTA (10 µmole). Bacterial alkaline phosphatase (10 µg) was added and incubation at 37° continued a further 30 minutes. DNA fragments were then dialysed against 20 mM Tris-HCl pH 7.4.

The fragments were terminally labelled in a reaction (1.2 ml) containing 20 mM Tris-HCl pH 7.4, 8 mM $MgCl_2$, 8 mM 2-mercaptoethanol, 10 nmoles $\gamma$-$^{32}$P-ATP and

polynucleotide kinase (10 µl: final fraction). The solution was incubated at 37° for 45 minutes, the reaction stopped by the addition of EDTA (20 µmoles). The solution was extracted with neutralised phenol and dialysed against 0.3 M NaCl, 0.01 M Tris-HCl pH 8.0, 0.001 M EDTA and finally dialysed against 0.01 M ammonium bicarbonate. The labelled DNA fragments were purified by chromatography on a column (45 x 1 cm diameter) of Sephadex G-100 run in 0.01 M ammonium bicarbonate. They appeared as a peak of radioactivity in the void volume. Fractions were combined and dialysed against 1 mM Tris-HCl pH 7.4 and concentrated to about 50 µl by evaporation.

The terminally labelled fragments were digested with pancreatic DNase, and radioactive oligonucleotides present in the digest were fractionated by ionophoresis on DE-cellulose paper at pH 2 and AE-cellulose paper at pH 3.5, as described for terminally labelled DNA fragments produced by R.Hin$_d$ III (Chapter 7).

Figure A.1 shows radioactive oligonucleotides fractionated on AE-paper. Oligonucleotides were eluted and analysed by partial digestion with venom phosphodiesterase. These analyses were performed by A. Morrison. Table A.1 gives the properties and identities of labelled oligonucleotides shown in Figure A.1. M values in Table 1 exhibited on DE-cellulose paper at pH 2 by homologous nucleotides, are in some instances on the low

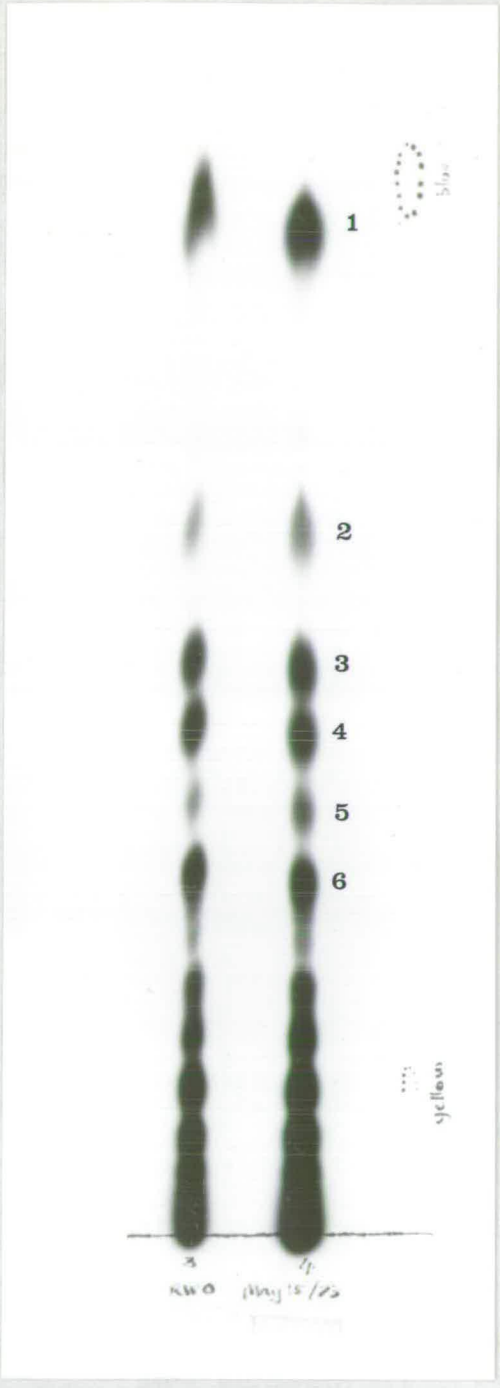



Figure A.1. Nucleotides in digests of 5'-terminally labelled fragments of lambda DNA produced by treatment with R.Hae. Nucleotides were fractionated on AE-cellulose paper at pH 3.5. Identities of numbered nucleotides are given in Table A.1 .

Table A.1 Radioactive oligonucleotides derived from 5′-terminally labelled R.Hae fragments of lambda DNA

| nucleotide code no. | 5′-terminal base | $R_F$ on DE pH 2 | $R_F$ on AE pH 3.5 | identity |
|---|---|---|---|---|
| 1 | C | 2.26 | 0.95 | C C |
| 2 | C | 2.17 | 0.68 | C C C |
| 3 | C | 1.66 | 0.54 | C C T |
| 4 | C | 2.06 | 0.45 | C C A |
| 5 | C | 2.12 | 0.41 | C C C C |
| 6 | C | 1.60 | 0.34 | C C G |

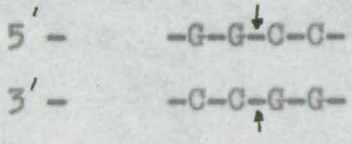
Code numbers refer to Figure 1.

$R_F$ values are relative to Xylene Cyanol FF.

Sequences refer to 5′-oligonucleotides. The 5′-terminal phosphate has been omitted.

side of the ranges given in Table 1.2 (Chapter 1). This is anticipated and characteristic of immediate homologues of so fast a migrating nucleotide as pC-C.

The identities of the nucleotides given in Table A.1 show that the 5′-terminal sequence becomes fully degenerate at the trinucleotide position. Analyses of other oligonucleotides provided evidence for degeneracy at the tetranucleotide position, so that it can be concluded that endonuclease R.Hae catalyses phosphodiester bond cleavage on the 5′- side of the dinucleotide sequence -C-C-. Since a single 5′-terminal dinucleotide sequence is created by R.Hae it is tempting to conclude that the substrate region recognised by the endonuclease has the structure



This structure could be confirmed by identification of the 3′-terminal nucleotides of DNA fragments produced by the action of R.Hae. In addition, such a tetranucleotide sequence occurs in duplex DNA molecules of defined sequence that have been synthesised using the procedures of Khorana (Kleppe et al., 1971; Kleppe and Khorana, 1972). These molecules comprise a set of test substrates.

While it is premature to draw firm conclusions on the substrate region recognised by R.Hae, the recovery

of a single 5′-terminal dinucleotide sequence in the experiments described here implies that the enzyme recognises a sequence having an axis of rotational symmetry.

## REFERENCES (APPENDIX)

Kleppe, R. & Khorana, H.G. (1972) J. Biol. Chem. 247, 6149-6156.

Kleppe, K., Ohtsuka, E., Kleppe, R., Molineux, I. & Khorana, H.G. (1971) J. Mol. Biol. 56, 341-361.

*Reprinted from*
*Proc. Nat. Acad. Sci. USA*
Vol. 70, No. 4, pp. 1151–1155, April 1973

# The 3′-Terminal Nucleotide Sequences of Bacteriophage λ DNA

**(*ter* function/recognition site/rotational symmetry/coliphage)**

PAUL H. WEIGEL, PAUL T. ENGLUND, KENNETH MURRAY*, AND ROBERT W. OLD*

Department of Physiological Chemistry, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205; and
*Department of Molecular Biology, University of Edinburgh, Edinburgh, Scotland

*Communicated by Albert L. Lehninger, February 8, 1973*

**ABSTRACT** **Analyses of radioactive oligonucleotides in endonuclease digests of 3′-terminally labeled λ DNA revealed the 3′ terminal sequence -GTTACG for the *l* strand and -ACCCGCG for the *r* strand. These sequences, together with those previously known for the 5′ cohesive ends, provide a total of 25 known base-pairs in the vicinity of the termini. When the cohesive ends are paired, the sequence between the nicks can be bisected by a 2-fold rotational axis of symmetry. Five of the first eight base-pairs, on either side of the axis, are rotationally symmetric. This symmetry may be involved in the recognition of the site by an enzyme responsible for formation of the cohesive ends.**

The DNA of bacteriophage λ is a duplex molecule of about 46,500 base-pairs, with single-stranded projections of 12 nucleotides at each 5′-terminus. These projections, or cohesive ends, are of complementary sequence and, by their specific base-pairing, the DNA may form either a ring or a linear aggregate (1, 2).

The mechanism by which the cohesive ends are formed is not completely understood. The mature viral DNA molecule may be produced from a replicative intermediate, possibly a concatemer, by endonucleolytic attack at the site of the cohesive ends (for a review, see ref. 3). This endonuclease is the *ter* function, whose existence was inferred from genetic studies (4) and which has been detected recently in extracts of *Escherichia coli* infected with phage λ (5). Presumably, this endonuclease binds to a unique sequence of nucleotides in the region of the cohesive ends and introduces two nicks, 12 base-pairs apart, on opposite strands. The mature λ DNA molecules may then be released by disruption of the base pairs between the nicks.

Our objective was to identify the nucleotide sequence recognized by the *ter* enzyme. The sequence of bases between the two nicks is known (2). To determine sequences beyond these nicks, we labeled λ DNA at specific sites at or near its 3′-termini (6) and determined the sequence of oligonucleotides produced by degradation of the DNA with a nonspecific endonuclease (7). The 3′-terminus of the *l* strand is -GTTACG, and that of the *r* strand is -ACCCGCG. Brezinski and Wang have recently informed us that they have independently obtained sequences in agreement with these results (8).

## MATERIALS AND METHODS

*DNA.* Phage were produced from the λ-lysogen *E. coli* N1383, (λcI857 *Sam*7), a gift of Dr. John Little (9, 10), and DNA was isolated (6). Analysis of the DNA on an alkaline sucrose gradient indicated that more than 90% of the strands were intact.

*3′-Terminal Labeling.* DNA termini were labeled with [$^{32}$P]nucleotides in a reaction catalyzed by the T4 DNA polymerase (6). The DNA [3.7 μmol of nucleotide in 1.95 ml containing 46 mM NaCl–86 mM Tris·HCl (pH 8.0)] was heated to melt cohered ends (74°, 10 min, followed by quick cooling to 0°). A solution (0.55 ml) containing polymerase (6), a single [α-$^{32}$P]deoxynucleoside triphosphate (6), $MgCl_2$, and 2-mercaptoethanol was added immediately after cooling. Final concentrations were: 1.5 mM DNA, 67 mM Tris·HCl (pH 8.0), 36 mM NaCl, 6 mM $MgCl_2$, 6 mM 2-mercaptoethanol, 0.1 mM triphosphate ($2 \times 10^{10}$ cpm/μmol), and 600 units/ml of polymerase. Incubation was at 11°, and when acid-insoluble radioactivity in an aliquot reached a limiting value (usually 60 min), the reaction was terminated by addition of EDTA (to 20 mM) and by extraction with phenol. Triphosphates were removed by exhaustive dialysis against 1.0 M NaCl–10 mM Tris·HCl (pH 8.0)–1.0 mM EDTA. The DNA was then purified by zone sedimentation in an alkaline sucrose gradient (5–20% sucrose in 0.2 M NaOH–0.8 M NaCl–1 mM EDTA; 25,000 RPM, 15 hr, 5°, Spinco SW25 rotor). This purification removed small labeled DNA fragments, which contained about 30% of the radioactivity, but which were not detectable by $A_{260}$ measurements. These fragments were probably not a consequence of endonuclease activity in the incubation mixture, because similar fragments were not observed in identical experiments with T7 DNA (6); they were probably present in the DNA used for labeling. The purified intact [$^{32}$P]DNA strands were then dialyzed against 50 mM NaCl–10 mM Tris·HCl (pH 8.0)–1.0 mM EDTA. A portion of the DNA was used for strand separation (11), and a sample of each isolated strand was used for nearest-neighbor analysis (12). The remainder of the purified [$^{32}$P]DNA was used for sequence analysis.

*Endonuclease Digestion of the Labeled DNA.* Labeled DNA was transported from Baltimore to Edinburgh in evacuated, sealed tubes. The solutions were concentrated to 1.0 ml by rotary evaporation, dialyzed against 0.3 mM NaCl, and further concentrated to 0.1 ml. The labeled DNA was then digested with pancreatic DNase (Worthington Biochemical Corp.). Conditions were chosen to achieve an equable distribution of radioactive oligonucleotides from trial digests analyzed by electrophoresis on Whatman AE81 paper (pH 3.5). Typical conditions were 50 μg/ml of DNase in 20 mM Tris·HCl (pH 7.5)–10 mM $MgCl_2$ at 37° for 3–12 hr. The

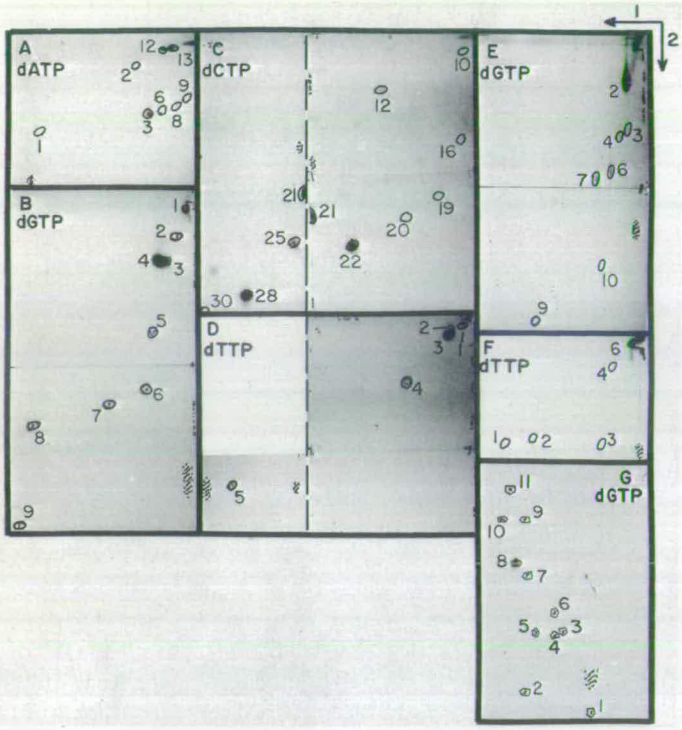



FIG. 1. Radioautograms of two-dimensional oligonucleotide maps of DNase digests of 3′-terminally labeled λ DNA. Electrophoretic method *iii* was used for maps *A*, *B*, *C*, and *D*, method *iv* for maps *E* and *F*, and method *v* for map *G*. The *labeling* triphosphate is shown on each panel. *Numbers* by oligonucleotide spots refer to code numbers in Table 1. The *hatched* zones enclosed by *dots* indicate the position of the blue dye, xylene cyanol FF. The *broken line* through panels *C* and *D* denotes the site where the chromatogram was cut after the first dimension. The *numbered arrows* denote the first and second dimensions in the two-dimensional separations.

digests were fractionated by electrophoresis by the following methods: *i*. AE81 paper at pH 3.5; *ii*. Whatman DE81 paper at pH 2; *iii*. two-dimensional on DE81 paper at pH 9.7, followed by pH 2 (13); *iv*. two-dimensional on AE81 paper at pH 3.5 followed by DE81 paper at pH 2; *v*. two-dimensional on cellulose acetate at pH 3.5 in 7 M urea, followed by transfer to thin layers of polyethyleneimine–cellulose and chromatographic development with 1.6 M formic acid (adjusted to pH 3.6 with pyridine) (14). Labeled oligonucleotides were located by radioautography and eluted (15).

*Analysis of Oligonucleotides.* The electrophoretic mobility of each [$^{32}$P]oligonucleotide was determined by methods *i* and *ii*. Oligonucleotides isolated by method *i* were examined by method *ii*, and *vice versa*. The 3′-penultimate base of each oligonucleotide was determined, after dephosphorylation with *E. coli* alkaline phosphatase (Worthington), by digestion with spleen phosphodiesterase (a generous gift of G. Bernardi) and analysis of the 3′-mononucleotides by electrophoresis by method *i* or by chromatography on polyethyleneimine–cellulose with 5% formic acid. (phosphatase digestion: 2 μg of enzyme in 20 μl of 60 mM $NH_4HCO_3$ at 37° for 45 min, followed by removal of phosphatase on phosphocellulose (16); phosphodiesterase digestion: 0.5 unit/ml (17) in 10 μl of 50 mM ammonium acetate–1 mM EDTA (pH 5.6) at 37° for 45 min).

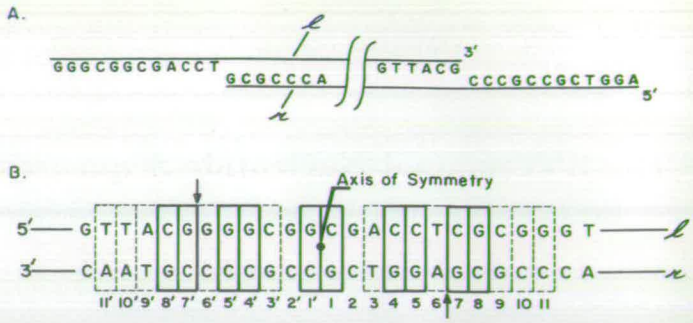



FIG. 2. (*A*) Nucleotide sequences at the termini of λ DNA. The sequences of the cohesive ends, as well as the 3′-terminal dGMP residues, were identified by Wu and Taylor (2) (see also ref. 7). (*B*) Symmetry in the region of the cohesive ends. The ends are paired, and nicks are indicated by *vertical arrows*. The 2-fold axis of symmetry is perpendicular to the page. Symmetrical base pairs are enclosed in *solid boxes* and pairs symmetrical in purine–pyrimidine orientation are enclosed in *dashed boxes*. Base pairs are *numbered* on either side of the axis.

## RESULTS

Incubation of λ DNA with the T4 DNA polymerase and a single [α-$^{32}$P]deoxynucleoside triphosphate results in the addition of a [$^{32}$P]nucleotide to a natural 3′-terminus provided that the triphosphate is complementary to the first base in the adjacent cohesive end. However, if no nucleotide can be added to the natural terminus, then the enzyme will sequentially remove nucleotides from this terminus until a nucleotide is reached that can be exchanged for a labeled one (6). For example, during incubation with [α-$^{32}$P]dGTP, three [$^{32}$P]nucleotides would be added to the 3′ terminus of the *l* strand and the 3′-terminal dGMP residue of the *r* strand would be exchanged for a labeled one (see Fig. 2*A* for structure of cohesive ends). Measurements of the extent of labeling with each triphosphate† and of the distribution of radioactivity between the two separated strands supported this general scheme. With [α-$^{32}$P]dGTP, about four [$^{32}$P]-nucleotides were incorporated per molecule, three of which were in the *l* strand. With [α-$^{32}$P]dATP or [α-$^{32}$P]dCTP, about two nucleotides were incorporated per molecule, with one on each strand. With [α-$^{32}$P]dTTP, only about one nucleotide was incorporated per molecule, even after 165 min of incubation, and about 80% of the radioactivity was in the *l* strand. Presumably, no dTMP residue exists close enough to the 3′-terminus of the *r* strand to allow efficient exchange for a radioactive nucleotide.

The strategy for 3′-terminal sequence analysis of the labeled DNA is based upon that developed for 5′-terminally labeled DNA (7). If 3′-terminally labeled DNA is degraded with a nonspecific endonuclease, all the labeled oligonucleotides in the digest must belong to one of two families. Each family arises from one of the strands and consists of a homologous series of oligonucleotides whose sequences overlap progressively from the 3′-terminus. Although all labeled oligonucleotides in the digest must have the same 3′-terminal base, each may be assigned to its family by identification of its 3′-penultimate base, which generally differ in the two families. Once all oligonucleotides have been thus assigned,

† Measurements were corrected for the fact that about 30% of the radioactivity was present in small fragments.

their sequences can be determined from the difference in mobility (methods *i* and *ii*) between adjacent members of each homologous series. This difference in mobility between adjacent homologs is diagnostic of the single base by which they differ (15, 18). Sequences can be confirmed by electrophoresis by methods *iii*, *iv*, or *v*. Finally, nearest-neighbor analysis of the separated strands is used to assign each family to one of the two strands of the DNA.

Application of this strategy to λ DNA labeled at the 3′-termini with [$^{32}$P]deoxyadenylate revealed the terminal sequence -GTTA on the *l* strand and -ACCCGCGA on the *r* strand (Table 1, Fig. 1). Inspection of the cohesive end sequences (see Fig. 2*A*) indicates that labeling of the *r* strand must have occurred by addition of a [$^{32}$P]dAMP residue to the 3′-terminal dGMP residue of that strand. Therefore, the terminal sequence of the *r* strand must be -ACCCGCG. Labeling of the *l* strand must have been preceded by the hydrolytic removal of nucleotides from the terminus.

Interpretation of the experiments with [$\alpha$-$^{32}$P]dGTP was aided by Wu and Kaiser's finding that the 3′-terminal residue of each strand is dGMP (19). Labeling of λ DNA with [$\alpha$-$^{32}$P]dGTP should result in exchange of the terminal dGMP residue of the *r* strand for [$^{32}$P]dGMP, and the addition of three [$^{32}$P]dGMP residues to the terminal dGMP

TABLE 1. *Oligonucleotides derived from 3′-terminally labeled λ DNA*

| | *l* strand | | | | *r* strand | | | |
|---|---|---|---|---|---|---|---|---|
| Labeling triphosphate | Oligonucleotides* | Mobility¶ *i* | Mobility¶ *ii* | Code numbers§ | Oligonucleotides* | Mobility¶ *i* | Mobility¶ *ii* | Code numbers§ |
| dATP | TA | 0.68 | 1.73 | | GA | 0.43 | 1.49 | |
| | TTA | 0.29 | 0.60 | A1 | CGA | 0.28 | 1.42 | |
| | GTTA | 0.07 | 0.17 | A2 | GCGA | 0.07 | 0.51 | A3 |
| | GGTTA** | 0.02 | 0.06 | A12 | CGCGA | 0.04 | 0.50 | A6 |
| | | | | | CCGCGA | 0.03 | 0.46 | A8 |
| | | | | | CCCGCGA | 0.02 | 0.41 | A9 |
| | | | | | ACCCGCGA | 0.01 | 0.30 | |
| dGTP | GGG‖ | 0.07 | 0.25 | E2 G8 | CG | 0.59 | 1.87 | E9 G1 |
| | GG | 0.27 | 0.79 | B8 E7 G5 | GCG | 0.15 | 0.68 | B7 E6 |
| | GGGG | 0.02 | 0.09 | | CGCG | 0.10 | 0.70 | G3 |
| | CGGGG | 0.01 | 0.09 | | | | | |
| | CGGG | 0.04 | 0.03 | | | | | |
| | CGG | 0.16 | 0.73 | B7 E6 G4 | | | | |
| | CG | 0.59 | 1.87 | E9 G1 | | | | |
| | ACG | 0.21 | 1.13 | B9 E10 | | | | |
| | ACGG | 0.05 | 0.50 | B5 E3 G6 | | | | |
| | TACGG | | | G7 | | | | |
| | TACG | 0.10 | 0.52 | E4 | | | | |
| | TTACGG | 0.01 | 0.09 | G10 | | | | |
| dCTP | AC | 0.71 | 2.28 | C28 | GC | 0.60 | 1.84 | C25 |
| | TAC | 0.40 | 1.44 | C21 | CGC | 0.37 | 1.71 | C22 |
| | TTAC | 0.17 | 0.49 | C12 | CCGC | 0.26 | 1.41 | C20 |
| | GTTAC | 0.03 | 0.15 | | CCCGC | 0.10 | 1.32 | C19 |
| | | | | | ACCCGC | 0.04 | 0.94 | C16 |
| | | | | | CC† | 0.93 | 2.41 | C30 |
| | | | | | CCC† | 0.62 | 2.08 | |
| dTTP | TT | 0.75 | 0.87 | F1 | | | | |
| | GT† | 0.60 | 0.86 | F2 | AT‡ | 0.59 | 1.65 | |
| | GTT | 0.17 | 0.25 | D4 F4 | AAT | 0.24 | 0.98 | D5 F3 |
| | GGTT** | 0.05 | 0.06 | D3 F6 | | | | |

* Assignment of an oligonucleotide to a family was by identification of the 3′-penultimate base. Assignment of a family to a strand of λ DNA was by nearest-neighbor analysis of separated strands (in each case about 85% of the radioactivity was in a single 3′-mononucleotide). Dinucleotide sequences determined by nearest-neighbor analysis on separated strands are underlined.

† These oligonucleotides were probably labeled because the polymerase occasionally failed to halt degradation after it reached the first nucleotide that could be exchanged for a labeled one.

‡ These oligonucleotides may have arisen from the *r* strand, although little radioactivity was incorporated into that strand.

§ Letters refer to panels and numbers refer to oligonucleotides in Fig. 1. Oligonucleotides not listed in these columns were identified only by methods *i* and *ii*.

¶ Mobility during electrophoresis by methods *i* and *ii* relative to the dye xylene cyanol FF.

‖ Structure confirmed by partial digestion with venom phosphodiesterase (13).

** The 5′ base is G or T. G is more probable.

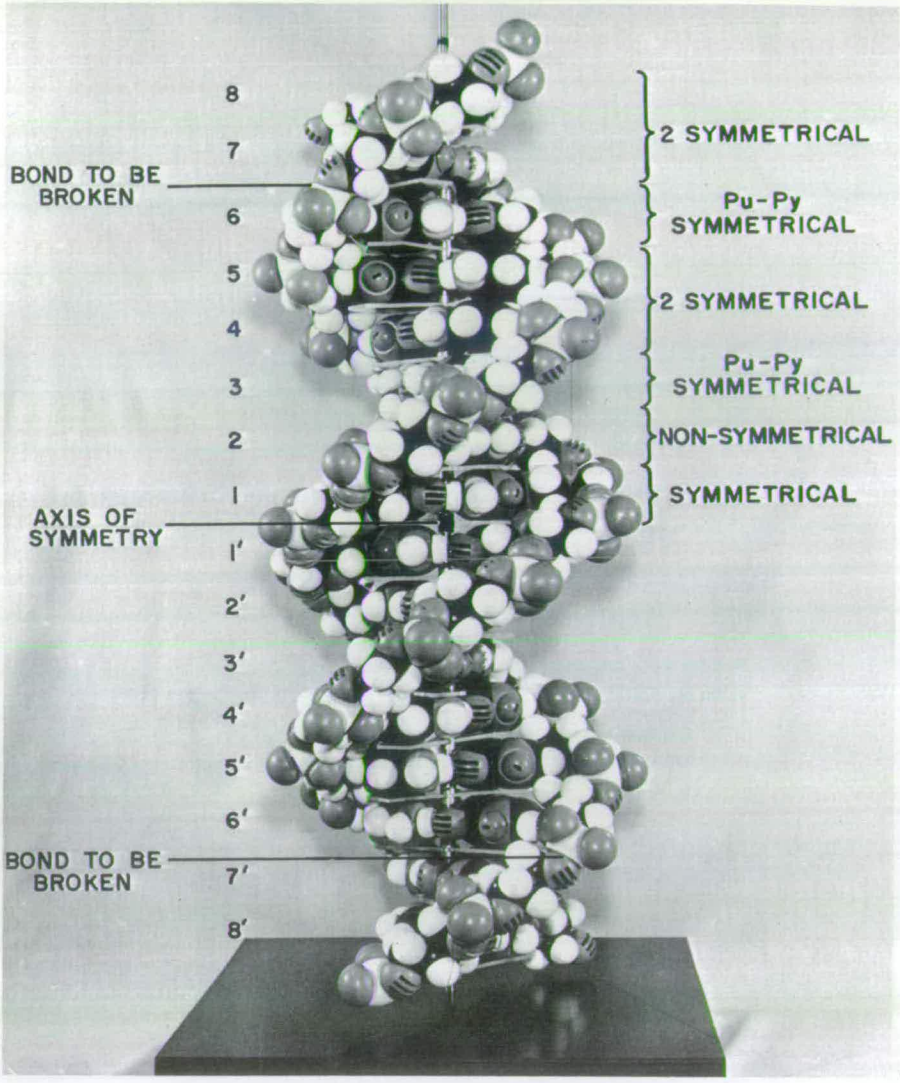



FIG. 3. Model of the "cohesive ends" region of the precursor of mature λ DNA (constructed of Corey–Pauling–Koltun models from the Ealing Corp.). Numbering of base-pairs is the same as in Fig. 2. The 2-fold axis of symmetry is perpendicular to the page.

of the *l* strand. The sequences derived from the *r* strand confirm sequences obtained from the DNA labeled with [$\alpha$-$^{32}$P]-dATP. The sequences from the *l* strand define its 3′-terminus, and overlap sequences determined from the DNA labeled with [$\alpha$-$^{32}$P]dATP. The terminal sequence of the *l* strand is, therefore, -GTTACG.

Additional oligonucleotides were isolated from DNA labeled with [$\alpha$-$^{32}$P]dCTP and [$\alpha$-$^{32}$P]dTTP. Their sequences confirmed those already deduced.

## DISCUSSION

We can now write a sequence of 25 base-pairs, within which lie the nicks produced by the *ter* function. The most striking property of this sequence is rotational symmetry (Fig. 2*B*). When the two cohesive ends are paired, the sequence between the two nicks is bisected by a 2-fold rotational axis. Not only are the nicks symmetrical about this axis, but five of the first eight base-pairs on either side of the axis are symmetrical. Also, an additional five of the first 11 pairs on either side of the axis are symmetrical in their purine–pyrimidine orientation. The term "hyphenated symmetry" has been used to describe sequences of this type in which the symmetry is interrupted by some nonsymmetrical base-pairs (20).

The abundance of G-C base-pairs in this sequence greatly increases the probability of a fortuitous symmetrical arrangement of base-pairs about a central axis. However, the 2-fold symmetry does suggest mechanisms by which the *ter* enzyme might recognize this sequence. For example, a dimeric enzyme, containing identical subunits that are symmetrical about a 2-fold axis, could bind to the DNA so that its axis would align with that of the symmetrical sequence. Symmetrical catalytic sites on the dimeric enzyme could then simultaneously introduce the two nicks. A hyphenated sequence of 10 symmetrical base-pairs would occur by chance only once in about a million base-pairs. Since λ DNA contains only about 46,500 base-pairs, it is reasonable to assume that the sequence would occur only once. However, the shorter sequence of five base-pairs with which each subunit is associated would occur by chance about 50 times in a random molecule the size of λ DNA, making it unlikely that the enzyme would be active other than as a dimer.

There are several ways in which a dimeric enzyme could

bind to this symmetrical sequence. One is that the enzyme binds to the native helix and interacts with functional groups within the grooves of the DNA. Inspection of a space-filling model shows that the phosphodiester bonds to be broken, as well as some possible specificity-conferring functional groups on the symmetrical base-pairs, occur on the same face of the helix (Fig. 3). One subunit could interact with base-pair **1** in the minor groove and base-pairs **4**, **5**, **7**, and **8** in the major groove; the other subunit could interact in a corresponding way with the prime-numbered pairs. It is possible that base pairs symmetrical only in purine–pyrimidine orientation are also involved in specific interaction with the enzyme. It is also possible that this sequence may dictate local stereochemical characteristics of the helix that differ from those of the B form (21), and that these characteristics may be recognized by the enzyme.

Other recognition mechanisms might require more drastic changes in the local structure of the DNA, such as the formation of loops or cruciforms as suggested by Gierer (22), or perhaps even the generation of more extensive single-stranded regions (23). Cruciforms may be formed from any rotationally symmetrical sequence and, although the present example is complicated by hyphenation, it may be possible to derive these structures if one invokes G-T and A-C base-pairs, for which there is some precedent (24–26).

Recognition of the symmetrical sequence by the λ *ter* function appears possible by these or other mechanisms. However, until rigorous physical evidence is available, it may not be possible to distinguish among them.

Symmetrical recognition sequences occur in other systems. Spleen acid DNase (27), actinomycin D (28), and restriction and modification enzymes from *Hemophilus influenzae* (ref. 29, and unpublished work of Old, Murray, and Roizes) and *E. coli* plasmids (ref. 30, and unpublished results of Bigger, Murray, and Murray) recognize rotationally symmetrical sequences. Available evidence suggests that the spleen and *E. coli* plasmid RI enzymes may be dimers of identical subunits (27, 30). Extensive symmetry similar to that surrounding the cohesive ends of λ DNA is not apparent within the sequence of 19 base-pairs at the cohesive ends of DNA from phages 186, P2, and 299 (31, 7), although the identification of the sequences adjacent to the cohesive ends could reveal the existence of some rotational symmetry. The rotational symmetry in the terminal sequences of λ DNA, however, is striking and provides an attractive model for the interaction of the *ter* function with its substrate.

We are grateful to Patricia Clinkenbeard for expert technica assistance. We thank Drs. James Wang for a preprint of an unpublished manuscript, and John Little and Michael Hershfield for helpful advice. P.H.W. is supported by NIH Training Grant 5T01 GM00184-13 and P.T.E. by NIH Grant CA 13602-04 and by a Faculty Research Award from the American Cancer Society. K.M. is supported by the Science Research Council (U.K.) and R.W.O. by a research training grant from the Medical Research Council (U.K.).

1. Hershey, A. D., Burgi, E. & Ingraham, L. (1963) *Proc. Nat. Acad. Sci. USA* **49**, 748–755.
2. Wu, R. & Taylor, E. (1971) *J. Mol. Biol.* **57**, 491–511.
3. Yarmolinsky, M. B. (1971) in *The Bacteriophage Lambda*, ed. Hershey, A. D. (Cold Spring Harbor Laboratory, New York), pp. 97–111.
4. Mousset, S. & Thomas, R. (1969) *Nature* **221**, 242–244.
5. Wang, J. C. & Kaiser, A. D. (1973) *Nature New Biol.* **241**, 16–17.
6. Englund, P. T. (1972) *J. Mol. Biol.* **66**, 209–224.
7. Murray, K. & Murray, N. E. (1973) *Nature*, in press.
8. Brezinski, D. & Wang, J. C. (1973) *Biochem. Biophys. Res. Commun.* **50**, 398–404.
9. Parks, J. S., Gottesman, M., Shimada, K., Weisberg, R. A., Perlman, R. L. & Pastan, I. (1971) *Proc. Nat. Acad. Sci. USA*, **68**, 1891–1895.
10. Varmus, H. E., Perlman, R. L. & Pastan, I. (1970) *J. Biol. Chem.* **245**, 2259–2267.
11. Szybalski, W., Kubinski, H., Hradecna, Z. & Summers, W. C. (1971) in *Methods in Enzymology*, eds. Grossman, L. & Moldave, K. (Academic Press, New York), Vol. XXI D, pp. 383–413.
12. Josse, J., Kaiser, A. D. & Kornberg, A. (1961) *J. Biol. Chem.* **236**, 864–875.
13. Murray, K. (1970) *Biochem. J.* **118**, 831–841.
14. Southern, E. M. & Mitchell, A. R. (1971) *Biochem. J.* **123**, 613–617.
15. Sanger, F., Brownlee, G. G. & Barrell, B. G. (1965) *J. Mol. Biol.* **13**, 373–398.
16. Székely, M. & Sanger, F. (1969) *J. Mol. Biol.* **43**, 607–617.
17. Carrara, M. & Bernardi, G. (1968) *Biochemistry* **7**, 1121–1132.
18. Murray, K. (1973) *Biochem. J.* **131**, 569–583.
19. Wu, R. & Kaiser, A. D. (1968) *J. Mol. Biol.* **35**, 523–537.
20. Arber, W. & Linn, S. (1969) *Annu. Rev. Biochem.* **38**, 467–500.
21. Bram, S. & Tougard, P. (1972) *Nature, New Biol.* **239**, 128–131.
22. Gierer, A. (1966) *Nature* **212**, 1480–1481.
23. Crick, F. H. C. (1971) *Nature* **234**, 25–27.
24. Crick, F. H. C. (1966) *J. Mol. Biol.* **19**, 548–555.
25. Paetkau, V., Coulter, M. B., Flintoff, W. F. & Morgan, A. R. (1972) *J. Mol. Biol.* **71**, 293–306.
26. Hirsch, D. (1971) *J. Mol. Biol.* **58**, 439–458.
27. Bernardi, G. (1968) *Advan. Enzymol.* **31**, 1–49.
28. Sobell, H. M. & Jain, S. C. (1972) *J. Mol. Biol.* **68**, 21–34.
29. Kelly, T. J., Jr. & Smith, H. O. (1970) *J. Mol. Biol.* **51**, 393–409.
30. Hedgpeth, J., Goodman, H. M. & Boyer, H. W. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 3448–3452.
31. Padmanabhan, R. & Wu, R. (1972) *J. Mol. Biol.* **65**, 447–467.