UNIVERSITY OF SOUTHAMPTON

STUDIES ON THE ECO RV RESTRICTION ENDONUCLEASE USING OLIGODEOXYNUCLEOTIDES CONTAINING MODIFIED BASES

by

Timothy Richard Waters

A thesis submitted for the degree of Doctor of Philosophy

FACULTY OF SCIENCE
DEPARTMENT OF BIOCHEMISTRY

March 1992

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SCIENCE

BIOCHEMISTRY

Doctor of Philosophy

STUDIES ON THE ECO RV ENDONUCLEASE USING OLIGODEOXYNUCLEOTIDES CONTAINING MODIFIED BASES

by Timothy Richard Waters

The Eco RV restriction endonuclease is an enzyme from E. Coli that cleaves double stranded DNA between the two central T.A base pairs of the sequence d(GATATC) with extremely high specificity.

The self complementary dodecadeoxynucleotide d(GACGATATCGTC) is a good substrate for the endonuclease and was used in this project as a model for studying the specific protein/DNA interactions of the Eco RV endonuclease. A series of oligodeoxynucleotides of this general sequence were prepared containing modifications to the dG and dC residues at the recognition sequence. Each modified base contains a conservative alteration to one of the potential protein contact points. This was achieved by the incorporation of the appropriate dC or dG analogue in place of the normal base in the recognition site of the model sequence using automated phosphoramidite based DNA synthesis techniques. The dG analogues used were 7-deazadeoxyguanosine, d[^{7C}G], 2-aminopurine deoxynucleoside, d[^{6H}G], 6-thiodeoxyguanosine, d[^{6S}G], deoxyinosine, d[I], and 3-deazadeoxyguanosine, d[^{3C}G]. The dC analogues used were 5-methyldeoxycytidine, d[^{5Me}C], and 2-pyrimidinone deoxynucleoside, d[^{4H}C]. The chemical synthesis of the analogues d[^{6H}G] and d[^{6S}G] in a form suitable for DNA synthesis are described in detail.

All of the resultant dodecamers were characterised by base composition analysis, hyperchromicity determination, melting temperature determination and circular dichroism spectroscopy. All of these oligodeoxynucleotides, with the exception of the d[^{6H}G] containing dodecamer gave data consistent with double stranded, B-form DNA. d(GAC[^{6H}G]ATATCGTC) was found to be single stranded and to overcome this problem the octadecadeoxynucleotide d(GTCGAC[^{6H}G]ATATCGTCGAC) was synthesised. This oligodeoxynucleotide was found to be double stranded and was used in the subsequent kinetic analysis.

Five of these analogue oligodeoxynucleotides were found to be endonuclease substrates using an HPLC assay described previously (Newman et al., 1990a). The d[^{7C}G] containing dodecamer and the d[^{6H}G] containing octadecamer were not cleaved at all. The k_{cat} and K_M parameters of those oligodeoxynucleotides that were substrates were determined using a UV absorbance based assay that was developed during the project. This assay is a continuous, real time technique that has the advantages of speed and accuracy over other methods of determining endonuclease activity. The results obtained have been interpretated in terms of contacts between the protein and its DNA substrate with reference to published crystal structures.

This thesis also describes techniques for obtaining resonance Raman spectra of d[^{4S}T] and d[^{4S}T] containing oligodeoxynucleotides. Initial vibrational assignments of the Raman peaks were made. From the wavenumber shift of the C=S peak, the strength of the base pairing hydrogen bond to the 4C=S of d[^{4S}T] in d(GACGA[^{4S}T]ATCGTC) was calculated to be about 7.5kJ/mol. Spectra were also obtained of d(GACGA[^{4S}T]ATCGTC) bound to the Eco RV endonuclease and some definite changes in the spectrum of the d[^{4S}T] base were observed on endonuclease binding. These were a shift in the C=S peak to higher wavenumbers, which was assigned to a reduction in the strength of the hydrogen bond to this group in the duplex, and a shift in the 5-methyl stretching vibration indicating some enzyme interaction at this point.

CONTENTS

CHAPTER 1: INTRODUCTION

CHAPTER 2: MATERIALS AND METHODS

CHAPTER 3C: KINETICS OF THE ECO RV ENDONUCLEASE

CHAPTER 3D: RESONANCE RAMAN SPECTROSCOPY

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Bernard Connolly, for his tuition and encouragement during the course of this project and I wish him every success in his new position in Newcastle. I would also like to thank Andrew Worrall for many useful discussions and for very generously allowing me to use his computers throughout the last three years. Many thanks go to Tony Parker for his patience and valuable guidance regarding the resonance Raman work.

I would like to thank all members of the 'Connolly group' and other people in the department for making 'Bolditz' an interesting place in which to work. Good luck to all 'Connolly' group members in their future careers.

I am indebted to Steve Halford for unpublished results and for valuable discussions concerning the endonuclease. I would also like to thank Fritz Winkler for unpublished endonuclease crystal data.

Thanks are also due to Geoff Kneale for assistance with the melting point determinations, Alex Drake for performing the CD spectra, Joan Street for running ^{1}H NMR spectra, and Dave Corina for providing the FAB mass spectra.

I am particularly grateful to Aidan, Birgit, Caroline, Gill and Richard for generously putting me up (or putting up with me?) for the last three months. Without their patient hospitality the writing of this thesis would have been far more difficult.

Last and by no means least I wish to thank my parents for their wonderful support (moral and financial). Without this support it would have been much harder (if not impossible) to have completed this course.

ABBREVIATIONS

Ar	aryl group
bp	base pairs
bpt.	boiling point
BSA	bovine serum albumin
Bz	benzoyl group
CD	circular dichroism
CDP	2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite
d[^{3C}A]	3-deazadeoxyadenosine
d[^{7C}A]	7-deazadeoxyadenosine
d[^{6Me}A]	*N*6-methyl-deoxyadenosine
d[^{5Br}C]	5-bromodeoxycytidine
d[^{4H}C]	2-pyrimidinone-1-β-D-(2'-deoxyriboside)
d[^{5Me}C]	5-methydeoxycytidine
d[^{3C}G]	3-deazadeoxyguanosine
d[^{7C}G]	7-deazadeoxyguanosine
d[^{6H}G]	2-aminopurine deoxynucleoside
d[^{6S}G]	6-thiodeoxyguanosine
d[I]	deoxyinosine
DMAP	4-(*N,N*-dimethylamino)pyridine
DMSO	dimethylsulphoxide
DMT	4'-, 4"-dimethoxytrityl
d[P]	purine deoxynucleoside
d[^{4H}T]	5-methyl-2-pyrimidinone-1-β-D-(2'-deoxyriboside)
d[^{2S}T]	2-thiothymidine
d[^{4S}T]	4-thiothymidine
DTE	dithioerythritol
d[U]	deoxyuridine
EDTA	ethylenediaminetetraacetic acid

FAB-MS	fast atom bombardment mass spectrometry
HEPES	*N*-(2-hydroxyethyl)piperazine-*N*'-(2-ethanesulphonic acid)
HPLC	high performance liquid chromatography
NMR	nuclear magnetic resonance
SDS	sodium dodecyl sulphate
THF	tetrahydrofuran
TLC	thin layer chromatography
T_M	melting temperature
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet

CHAPTER 1

INTRODUCTION

1.1 IMPORTANCE OF PROTEIN/DNA INTERACTIONS

Proteins that interact with nucleic acids play a central role in a great number of biological mechanisms. These include such fundamental processes as transcription, translation and replication which are carried out and controlled by proteins that bind to nucleic acids. Proteins that interact with DNA vary in their degree of sequence specificity, from those that have no sequence specificity, such as DNA polymerase I (Joyce & Steitz, 1987), to those that have a extremely high sequence specificity such as type II restriction endonucleases (Modrich, 1979). In between these two extremes come proteins such as DNase I that do not have a clearly defined sequence preference but nevertheless have enhanced reactivity at some sequences (Drew & Travers, 1984), and repressor proteins that recognise a specific sequence but can tolerate a few base changes within that sequence (Jordan & Pabo, 1988; Aggarwal *et al.* 1988; Otwinowski *et al.* 1988). Information that improves the understanding of protein/DNA interactions at the molecular level and how they give rise to sequence specificity is therefore rudimentary in understanding many fundamentally important biological mechanisms.

1.1.1 Type II Restriction/Modification Systems

Type II restriction endonucleases are a group of enzymes found in bacteria that bind to and cleave specific sites in DNA (Modrich, 1979; Roberts & Macelis, 1991). It is thought that they provide bacteria with a defence system against invading foreign DNA, such as viral DNA. Typically they recognise a specific DNA sequence of four or six base pairs. They cleave only that sequence and only when present in double stranded form, with hydrolysis occurring between two of the bases within or at a fixed position outside that site. Therefore any foreign DNA invading the bacteria will be cleaved at the cleavage site of the restriction enzyme. To protect its own DNA from its endonucleases, the bacteria have a set of enzymes that modify the native DNA to prevent hydrolysis. These are specific methylases that recognise the same sequence as the corresponding restriction enzyme and methylate one of the bases within that site. This prevents cleavage of the native DNA by the endonuclease (Modrich, 1979; Terry *et al.*, 1983; Fliess *et al.*, 1988).

Restriction enzymes are essential tools in genetic engineering and because of this a great many of them have been characterised (Roberts & Macelis, 1991). However, very little structural or mechanistic work has been carried out on them. Their incredible specificity for small DNA sequences makes them good, convenient models for studying specific DNA/protein interactions.

1.2 RESEARCH OUTLINE

Eco RV endonuclease is a restriction enzyme that binds to DNA and cleaves the phosphate bond between the central thymidine and deoxyadenosine nucleosides within the sequence d(GATATC). The aim of this project was to try and define some of the molecular interactions between the Eco RV endonuclease and DNA that enable it to be so sequence specific.

X-ray crystallography is the definitive technique for investigating protein/DNA interactions as demonstrated by the solution of several protein/DNA complexes [see section 1.4]. Indeed, the solution of various crystal structures of Eco RV endonuclease (bound to cognate DNA, bound to non-cognate DNA and the free enzyme) has led to the definition of some of the specific molecular interactions between the protein and DNA (Winkler, 1991). However, there are three points that mean that these three dimensional structures do not wholly define the interactions between the endonuclease and DNA. Firstly, the crystals were produced in the absence of the catalytic cofactor magnesium, in order to prevent DNA cleavage in the complex. It is possible that some structural rearrangement of the protein/DNA complex could occur on binding of the metal ion. Secondly, the crystal structure represents a somewhat static view in that it reveals little about what conformational changes occur before, during or after binding and hydrolysis. Thirdly, although the crystal structure can infer the presence of hydrogen bonds, it cannot be used to define their strength or their importance in the cleavage reaction.

Solution studies of the enzyme are complementary to the X-ray crystallographic work as they can give information about what occurs during the catalysis and unlike crystallography they can be carried out under (or close to) physiological conditions.

The approach used in this project was to measure the Michaelis-Menten parameters (Fersht, 1985) for the reaction of Eco RV endonuclease with various modified substrate analogues. This involved the chemical synthesis of a series of deoxyguanosine and deoxycytidine analogues, each with one of the chemical groups on the base that could be a potential protein contact point altered. This work is a continuation of the experiments previously carried out in this laboratory on a set of deoxyadenosine and thymidine analogues (Newman *et al.*, 1990a; Newman *et al.* 1990b).

The nucleoside analogues were each, in turn, incorporated in place of the normal base in the Eco RV recognition site of the self complementary dodecamer d(GACGATATCGTC). The unmodified dodecamer is a good substrate for the enzyme and therefore provides a convenient model for studying the reaction.

The base analogues were chosen on the basis of the chemical groups postulated by Seeman *et al.* (1976) to be available for contact by an incoming protein as shown in Figure 1.1. The deoxycytidine and deoxyguanosine analogues all fall, with one exception, into one of three categories:- (i) Those where a potential hydrogen bonding acceptor or donor group has been replaced by a hydrogen, (ii) those were a ring nitrogen has been replaced by a carbon, and

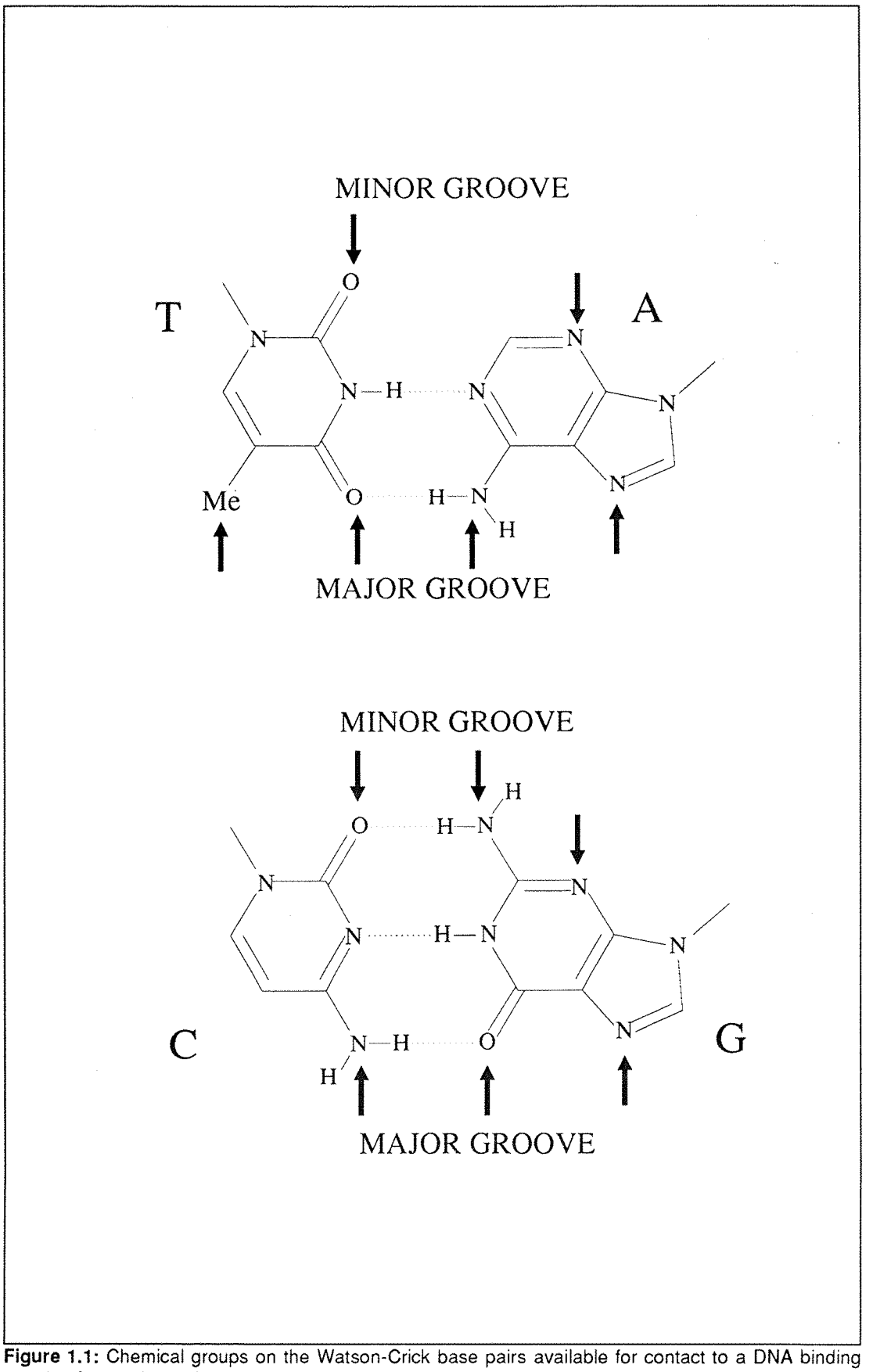


Figure 1.1: Chemical groups on the Watson-Crick base pairs available for contact to a DNA binding protein. Access is made either through the major or the minor groove (adapted from Seeman *et al.* 1977).

(iii) those where a carbonyl oxygen has been replaced by a sulphur atom. The exception is the 5-methyldeoxycytidine analogue in which an extra van der Waals group (the methyl group in the 5-position) has been introduced. The analogues are illustrated in Figures 1.2a and b and are discussed below:-

(i) deoxyinosine (dI); where the potential hydrogen bond donor exocyclic 2-amino group of deoxyguanosine has been replaced by a hydrogen. This also deletes one of the base pairing hydrogen bonds.

(ii) 2-aminopurine deoxynucleoside (d[^{6H}G]); where the potential hydrogen bond acceptor 6-carbonyl oxygen of deoxyguanosine has been replaced by a hydrogen. This causes rehybridisation of the ring resulting in the loss of the hydrogen at position 1 and consequently two base pairing hydrogen bonds are lost.

(iii) 6-thiodeoxyguanosine (d[^{6S}G]); where the potential hydrogen bond acceptor 6-carbonyl oxygen of deoxyguanosine has been replaced by a sulphur. Unlike d[^{6H}G], this does not cause rehybridisation. Sulphur forms hydrogen bonds of about half the strength of those formed by oxygen (Weiner *et al.*, 1984) and so only about half a Watson-Crick hydrogen bond is lost. However, sulphur hydrogen bonds tend to be longer than those of oxygen (Thewalt & Bugg, 1972). Also the sulphur atom (radius 1.85Å) introduces more bulk into the base than an oxygen atom (radius 1.40Å) (Weast *et al.* 1988).

(iv) 3-deazadeoxyguanosine (d[^{3C}G]); where the potential hydrogen bond acceptor ring 3-nitrogen of deoxyguanosine has been replaced by a carbon.

(v) 7-deazadeoxyguanosine (d[^{7C}G]); where the potential hydrogen bond acceptor ring 7-nitrogen of deoxyguanosine has been replaced by a carbon.

(vi) 2-pyrimidinone deoxynucleoside (d[^{4H}C]); where the potential hydrogen bond donor exocyclic 4-amino group of deoxycytidine has been replaced by a hydrogen. This deletes one of the base pairing hydrogen bonds.

(vii) 5-methyldeoxycytidine (d[^{5Me}C]); where the 5-hydrogen of deoxycytidine has been replaced with a potential van der Waal contacting methyl group. This mimics the 5-methyl group of thymidine and should show whether the extra bulk of a thymidine methyl group provides discrimination between it and deoxycytidine. This substitution is also interesting since some Type II methylases methylate the 5-position of deoxycytidine to prevent cleavage by their corresponding endonuclease (Nelson & McClelland, 1991). The use of this analogue should therefore show whether such methylation would prevent cleavage by Eco RV endonuclease.

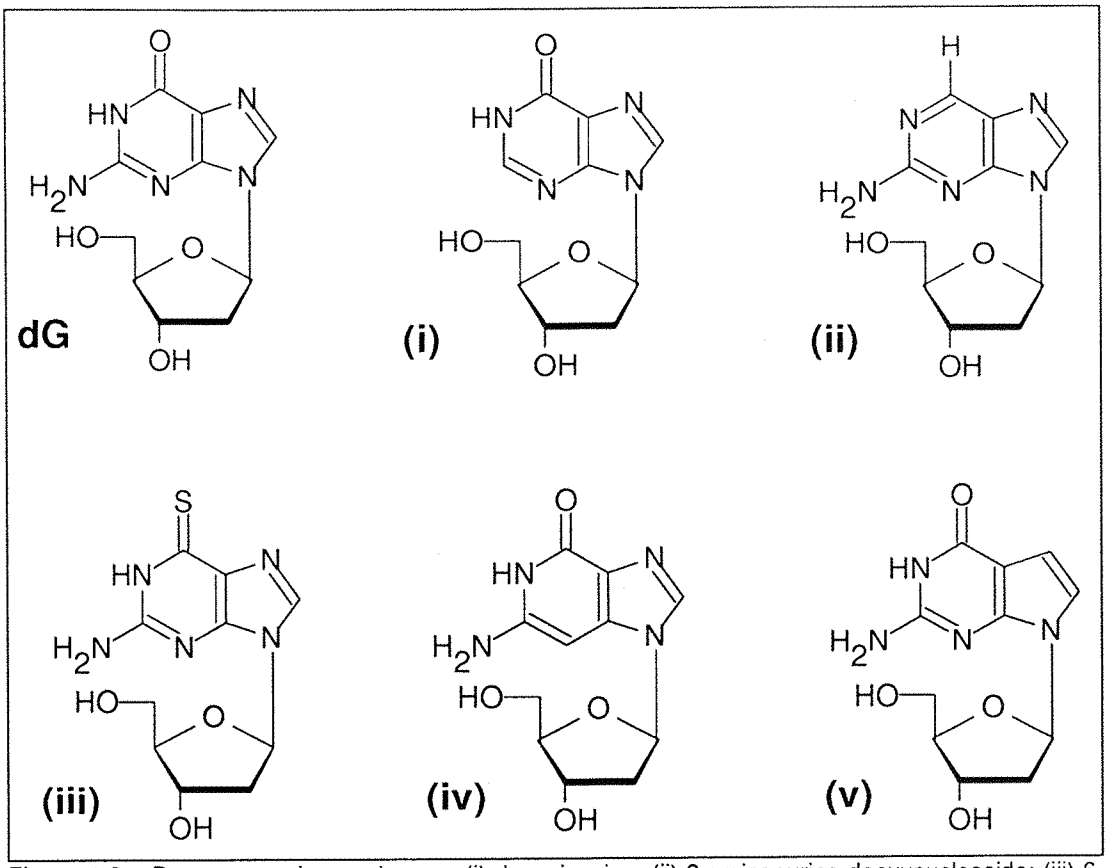


Figure 1.2a: Deoxyguanosine analogues. (i) deoxyinosine; (ii) 2-aminopurine deoxynucleoside; (iii) 6-thiodeoxyguanosine; (iv) 3-deazadeoxyguanosine; (v) 7-deazadeoxyguanosine.

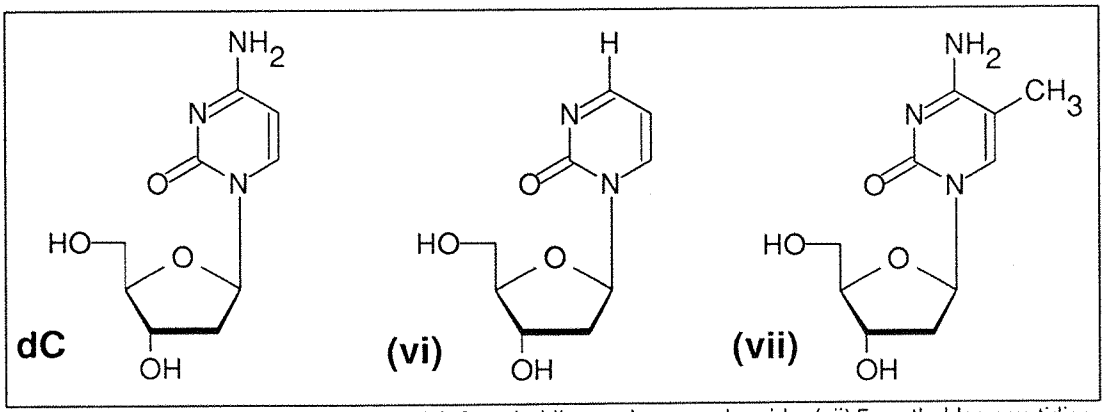


Figure 1.2b: Deoxycytidine analogues. (vi) 2-pyrimidinone deoxynucleoside; (vii) 5-methyldeoxycytidine.

1.2.1 Kinetic Analysis of Enzyme Reactions with Modified Substrates

The following analysis was developed to explore the effects of site directed mutagenesis experiments on enzyme reactions but is equally applicable to modified substrate experiments as discussed below (Fersht, 1988).

We start by considering the interactions a hydrogen bond donor group, –XH, on an enzyme, E, makes with its hydrogen bond acceptor, –Y, on a substrate, S. In water, the reaction involves the formation of the hydrogen bond between E–XH and S–Y together with

the reformation of the hydrogen bonds between bulk water and the water molecules that were solvating –XH and –Y :-

$$E\text{-}XH\cdots OH_2 + HOH\cdots Y\text{-}S \rightleftharpoons [E\text{-}XH\cdots Y\text{-}S] + HOH\cdots OH_2$$

The binding energy, ΔG_{bind}, is related to the hydrogen bonds broken and formed during the process as follows:-

$$\Delta G_{bind} = G_{EXH.YS} + G_W - G_{EXH.W} - G_{SY.W}$$

where $G_{EXH.YS}$, $G_{EXH.W}$ and $G_{SY.W}$ are the hydrogen bond dissociation energies of [E–XH···Y–S], E–XH···OH_2 and HOH···Y–S respectively. G_W is the free energy change accompanying the release of bound water molecules to bulk solvent. The value of ΔG_{bind} is extremely difficult to measure directly but in some circumstances it can be roughly determined by the use of substrate analogues. If a substrate analogue is used where –Y is replaced by a hydrogen then the binding process becomes:-

$$E\text{-}XH\cdots OH_2 + H_2O/H\text{-}S \rightleftharpoons [E\text{-}XH/H\text{-}S] + HOH\cdots OH_2$$

The apparent binding energy, ΔG_{app}, of the –Y group on the substrate is defined as the difference in the free energy between the two complexes E–XH···Y–S and E–XH/H–S. It can be defined in terms of their two dissociation constants, K_{SY} and K_{SH}, as follows (Wells & Fersht, 1986):-

$$\Delta G_{app} = RT\ln K_{SY} - RT\ln K_{SH} = RT\ln(K_{SY}/K_{SH})$$

When the reactions of substrates with an enzyme are being studied then it is more reliable to use the specificity constant k_{cat}/K_M, determined from Michaelis-Menten kinetics, than the dissociation constant (Fersht, 1985). This avoids complications that may arise from non-specific binding. The equation now becomes:-

$$\Delta G_{app} = RT\ln[(k_{cat}/K_M)_{SH} / (k_{cat}/K_M)_{SY}]$$

ΔG_{app} is a measure of the specificity an enzyme exhibits between the two substrates, S–H and S–Y, and is generally not equal to ΔG_{bind}. This is because ΔG_{bind} relates the interaction of enzyme and substrate with each other compared with their interaction with water, whereas ΔG_{app} compares the interaction between enzyme and S–Y with the interaction between enzyme and S–H. ΔG_{app} is a rough approximation of ΔG_{bind} if the E–XH/H–S complex allows free access of water to the –XH group since in this case the interaction between E–XH and Y–S has been

removed but the –XH group is still solvated. Another situation where ΔG_{app} has been found to be about equal to ΔG_{bind} is when the hydrogen bond deleted involves two groups that are not charged and where the modification does not introduce unfavourable steric factors (Fersht, 1987). Several such hydrogen bond deletion experiments, including mutagenesis of tyrosyl-tRNA synthetase (Wells & Fersht, 1986) and base mismatch studies of nucleotide-nucleotide complexes (Freier *et al.*, 1986) have given values for both ΔG_{app} and ΔG_{bind} in the range of 2.1 to 7.5kJ/mol.

Most of the base analogues used in the present study come into the category of potential hydrogen bonding groups being replaced by hydrogen. Additionally, the Eco RV endonuclease complex shows no charged amino acids involved in specific hydrogen bond formation (Winkler, 1991) and so ΔG_{app} is expected to approximate ΔG_{bind}. The analogues involving the replacement of C=S for C=O obviously do not come into this category since the extra bulk of sulphur may cause steric hindrance. Nevertheless, ΔG_{app} should still give a measure of the importance of the carbonyl oxygen replaced.

Any such analysis relies on the modification not causing any disruption to the conformation of the substrate or the enzyme. It is therefore important to ensure that the modification is as conservative as possible and that the base analogues do not disrupt the structure of the oligodeoxynucleotide. In this project two techniques have been employed to ensure that the overall DNA structure of the analogue containing oligodeoxynucleotides have not been seriously altered by the introduction of the base analogue. The first, melting temperature (T_M) looks at the ultraviolet absorbance of the oligodeoxynucleotide as a function of temperature. This technique relies on the hyperchromic effect that is seen in double stranded DNA where the UV absorbance of the DNA bases is significantly reduced by base stacking on the formation of duplex DNA. At low temperatures, an oligodeoxynucleotide would be predominantly double stranded but on increasing the temperature, the oligodeoxynucleotide duplex starts to unpair resulting in the loss of base stacking and a concomitant increase in the UV absorbance until finally at higher temperatures, the DNA is in a random single stranded conformation with minimum base stacking and maximum UV absorbance. The change from double stranded to single stranded DNA usually occurs over only a few degrees resulting in the kind of melting curve illustrated in Figure 1.3. The quantity obtained from such curves is the melting temperature, T_M, which corresponds to the temperature at which half of the DNA is in double stranded form and half is in single stranded form. The T_M values for the analogue oligodeoxynucleotides should thus show up any base analogues which cause disruption of the Watson-Crick base pairing in the oligodeoxynucleotide.

The other technique used to look at oligodeoxynucleotide structure is circular dichroism (CD) spectroscopy. This looks at the difference in absorbance of two sources of circularly polarised light, one clockwise and one anticlockwise, over a range of UV wavelengths and can

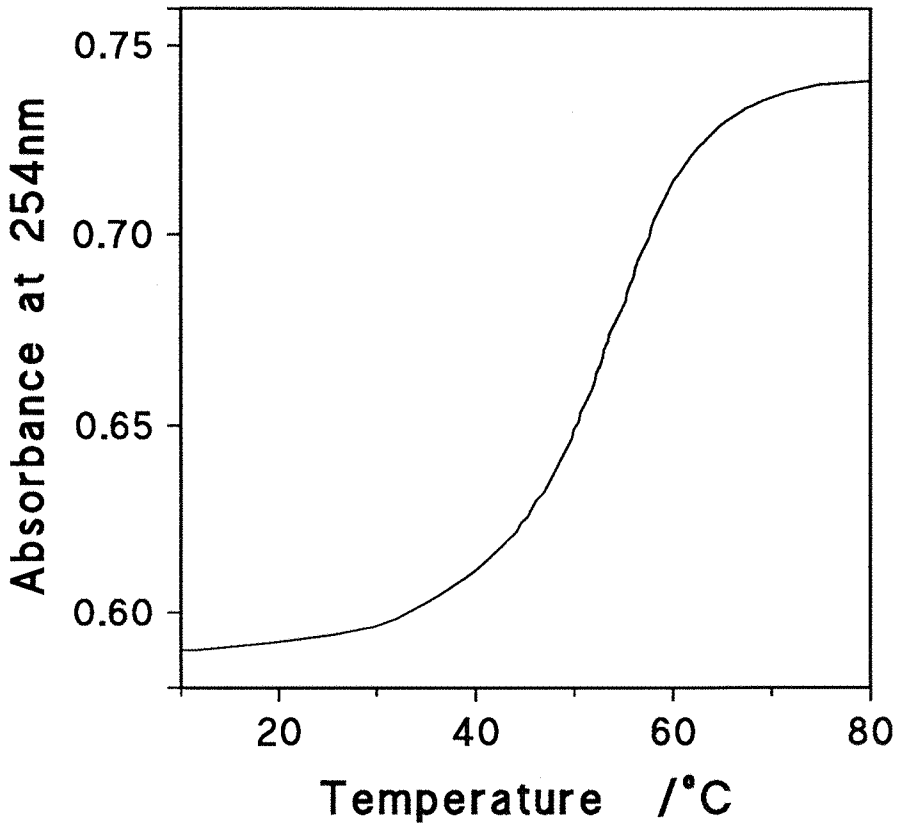


Figure 1.3: Typical melting curve of an oligodeoxynucleotide, which in this case is 12 base pairs long.

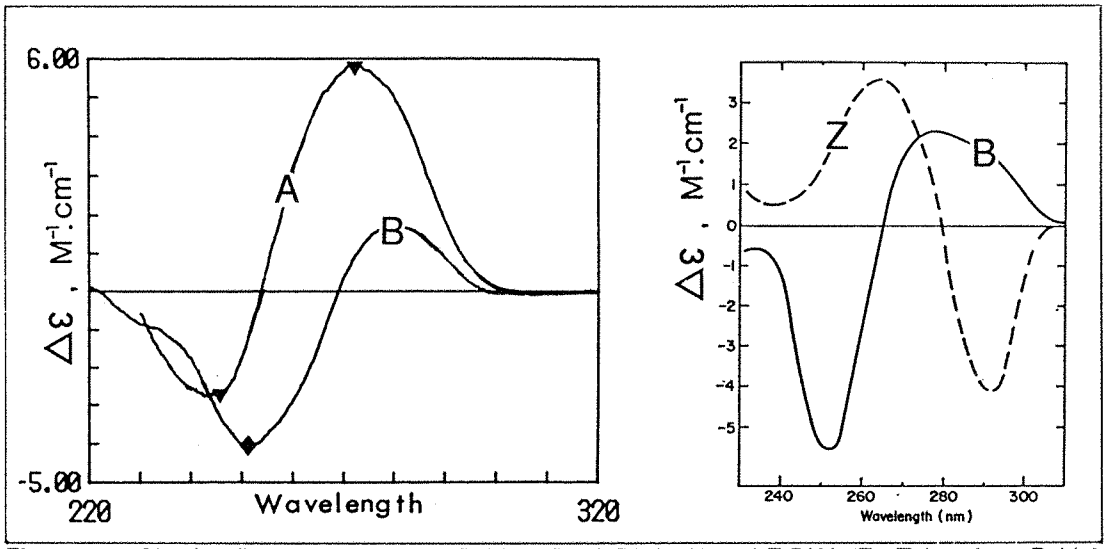


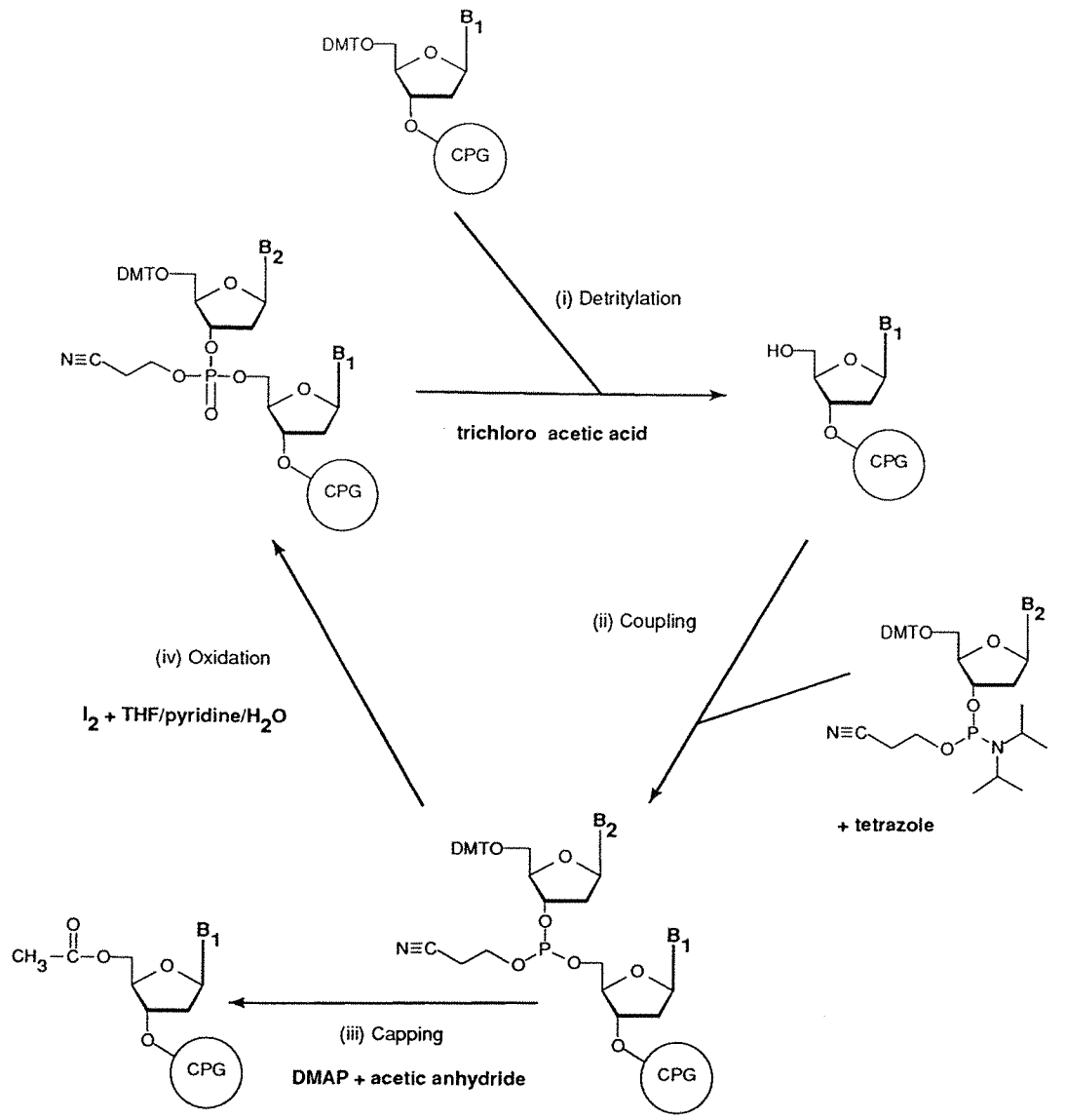
Figure 1.4: Circular dichroism spectra of B-DNA (B), A-DNA (A) and Z-DNA (Z). Taken from Pohl & Jovin, 1972 and Fairall *et al.*, 1989.

be used to determine the helical nature of DNA (Pohl & Jovin, 1972; Fairall *et al.*, 1989). At the present time, it is somewhat of a qualitative method and is usually used to distinguish between overall DNA structure of B-, Z- or A-form DNA as illustrated in Figure 1.4.

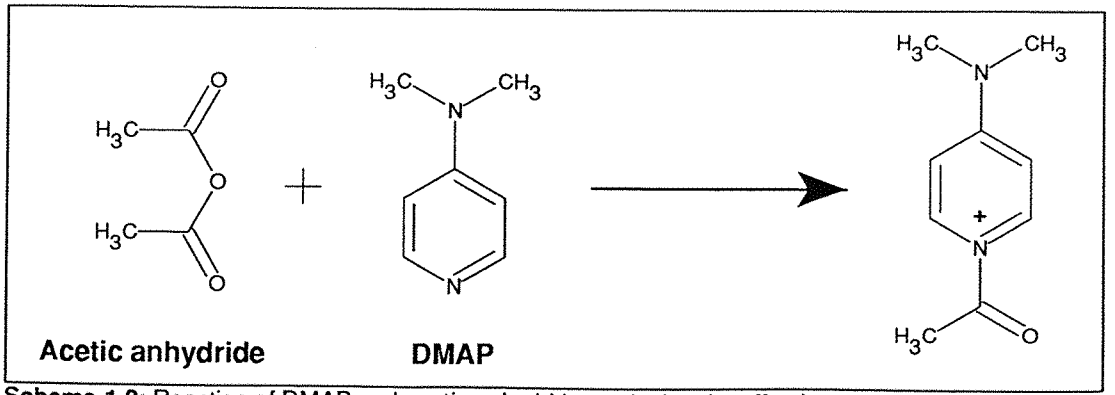
Other techniques that can be used for studying DNA structures including X-ray crystallography, proton NMR and Raman spectroscopy are discussed in later sections.

1.2.2 Current Method of Oligodeoxynucleotide Synthesis

At the present time almost all oligodeoxynucleotide synthesis involves the synthesis cycle illustrated in Scheme 1.1 which utilizes phosphite triester chemistry (Brown & Brown, 1991). The first deoxynucleoside is attached through its 3'-hydroxyl function to an inert solid support. By immobilizing the DNA in this way, excess reagents can be washed away from the propagating oligodeoxynucleotide chain, greatly simplifying purification and wash steps between each chemical reaction step. The first step of the cycle, (i), is the removal of the acid labile 4',4"-dimethoxytrityl (DMT) group from the 5'-hydroxyl by treatment with trichloroacetic acid. The following step, (ii), is the addition of the next nucleoside in the form of a protected phosphoramidite nucleoside. Phosphoramidites are nucleosides that have a suitably protected base, a DMT protected 5'-hydroxyl (Smith *et al.*, 1962) and a 3' trivalent phosphorous moiety and are used as the building blocks of DNA synthesis. The phosphorous moiety attached to the 3'-oxygen has a β-cyanoethyl protecting group and an N,N'-diisopropylamine group which acts as the leaving group in the coupling step (Sinha *et al.*, 1984). The phosphoramidite is activated by protonation of its amino function which makes it a better leaving group. By using an excess of phosphoramidite, coupling is achieved almost quantitatively by nucleophilic attack of the 5'-hydroxyl of the support bound oligodeoxynucleotide onto the phosphorous of the phosphoramidite with displacement of the protonated N,N'-diisopropylamine group. The 5'-DMT protecting group prevents self-polymerisation of the activated phosphoramidite. After coupling, some unreacted 5'-hydroxyl groups may remain and to prevent the undesirable chain extension of these groups which would result in oligodeoxynucleotides with one base missing, they must be suitably capped. This is achieved in the next step, (iii), where acetic anhydride is added together with dimethylaminopyridine (DMAP). DMAP and acetic anhydride, when mixed, form the species illustrated in Scheme 1.2 which is a powerful acetylating agent. It acetylates all the unprotected 5'-hydroxyls, preventing them from reacting further in the DNA synthesis and thus serves to truncate all oligodeoxynucleotide chains that have failed to add a nucleoside. The next step, (iv), is oxidation of the phosphite to the phosphate valency normally seen in DNA. The oxidation reagent consists of iodine in a water/pyridine/THF mix with the oxidising agent being iodine and the oxygens for the phosphate being provided by the water (Pon, 1987). The synthesis cycle is now complete and further nucleotides can be added to the bound oligodeoxynucleotide by repeating the cycle of (i) detritylation of the terminal 5'-hydroxyl, (ii) phosphoramidite coupling, (iii) capping of unreacted 5'-hydroxyls, and (iv)



Scheme 1.1: Synthesis cycle for automated DNA synthesis.



Scheme 1.2: Reaction of DMAP and acetic anhydride producing the effective acetylating agent shown.

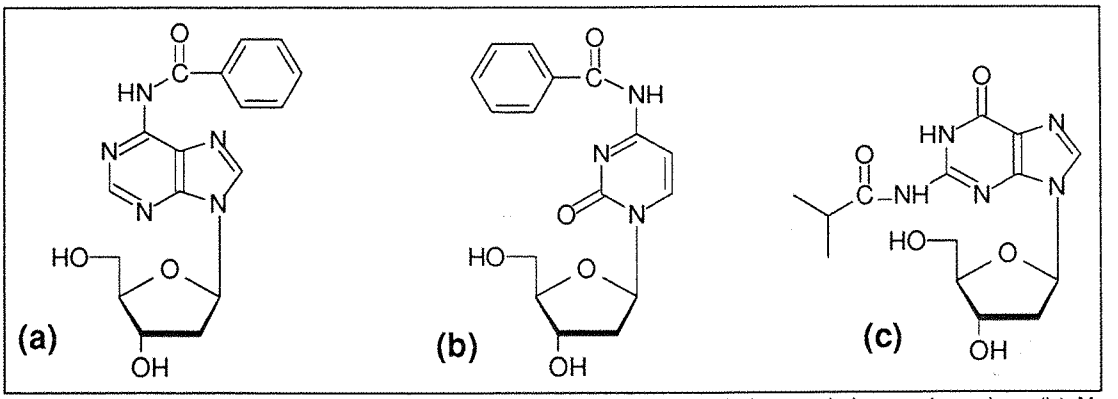


Figure 1.5: Amine protecting groups used in DNA synthesis. (a) N-benzoyl deoxyadenosine; (b) N-benzoyl deoxycytidine; (c) N-isobutyryl deoxyguanosine.

oxidation of the phosphite to phosphate.

After synthesis is complete, the oligodeoxynucleotide is cleaved from the solid support, the phosphorous β-cyanoethyl groups are removed and the base protecting groups removed by the treatment with concentrated aqueous ammonia. The oligodeoxynucleotide is then in the form of natural DNA and if necessary can now be purified. Purification is generally achieved by one of the following three methods (Gait, 1984; Zon & Thompson, 1986). (i) Ion exchange hplc. This separates oligodeoxynucleotides on the basis of the number of phosphate groups (and hence size). (ii) Gel electrophoresis which separates on the basis of size. (iii) Reverse phase hplc. This last technique is extremely useful and requires that the 5'-DMT group of the last nucleoside be left on. Because of its great hydrophobicity, the DMT group enables the correct sequence to be readily separated by reverse phase chromatography from all of the failure sequences which due to the capping step will lack the DMT group. The DMT group can subsequently be removed by treatment with acetic acid at room temperature for half an hour.

The release of the DMT cation (Scheme 1.1, step (i)) provides a useful method for quantifying the yield of each coupling step. After coupling, capping and oxidation, only those chains correctly synthesised will release the DMT group on acid treatment . By virtue of its distinct orange colour in acid, the amount of DMT and hence the amount of correctly coupled oligodeoxynucleotides can be determined from the absorbance of the released DMT.

To prevent the exocyclic amine functions of the three bases, deoxyguanosine, deoxyadenosine and deoxycytidine, from reacting during the synthesis steps they need to be protected. Since thymidine has no reactive exocyclic groups it needs no protection. The most commonly used protecting groups are the acetyl groups shown in Figure 1.5 with deoxyadenosine and deoxycytidine protected by benzoylation and deoxyguanosine protected by isobutyrylation (Ti *et al.*, 1982). The removal of these groups is the slowest part of the final ammonia deprotection reaction (5 hours at 55°C). As well as this long time being a possible inconvenience, the prolonged exposure to ammonia can be a problem in some applications where modified oligodeoxynucleotides are synthesised (Newman *et al.*, 1990a; Waters & Connolly, 1991a). This problem can now be overcome by the use of the base labile protecting

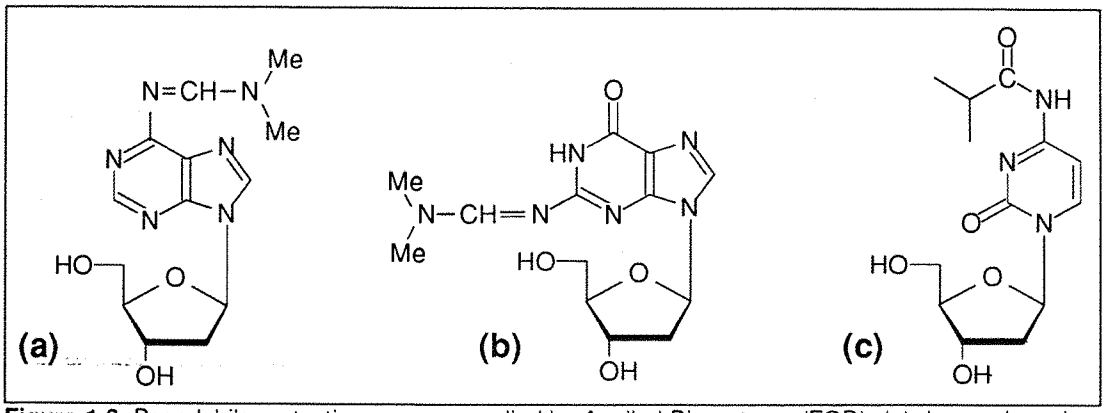


Figure 1.6: Base labile protecting groups supplied by Applied Biosystems (FOD). (a) deoxyadenosine; (b) deoxyguanosine; (c) deoxycytidine.

groups shown in Figure 1.6 (McBride *et al.*, 1986) which are now available commercially from Applied Biosystems (known as FOD phosphoramidites). These protecting groups can be removed by a one hour ammonia treatment at 50°C or at room temperature overnight.

1.3 THE STRUCTURE OF DNA

The structure of uniform B-DNA, first proposed by Watson and Crick (1953) based on fibre diffraction patterns, is shown in Figure 1.7a. It was thought to be a uniform right-handed helix formed by two single DNA strands running in opposite directions and joined by hydrogen bonding between bases from opposing strands. The bases are on the inside of the helix and the hydrophilic phosphates on the outside of the helix. From the diagram, two grooves in the helix are clearly visible and are indicated as the major and minor grooves. With this structural model available it was possible to propose systems for recognition of specific sequences of DNA by proteins. Seeman *et al* (1976) suggested that recognition of a specific DNA sequence by a DNA binding protein could be by direct contact to the bases themselves. An incoming protein would have access to the DNA bases either through the major or the minor groove of B-DNA. Each base pair presents a different array of possible contacts for a protein in these grooves as illustrated in Figure 1.1. The groups are all hydrogen bond acceptors or donors except for the 5-methyl group on thymidine which could provide van der Waals contact. A DNA binding protein could therefore select for a particular DNA sequence by having a complementary array of hydrogen bond donors and acceptors spatially arranged so that only when presented with its correct recognition sequence would all its hydrogen bonds be able to be formed, as illustrated in Figure 1.8. This type of recognition is known as 'direct' readout.

When X-ray structures of *crystals* of short stretches of DNA were solved it became clear that the structure of DNA was far more variable than had first been thought (Dickerson & Drew, 1981). An extreme example was the discovery of a completely new form of DNA that was a left-handed helix (Wang *et al.*, 1979; Wang *et al.*, 1981) now known as Z-DNA (Figure 1.7b).

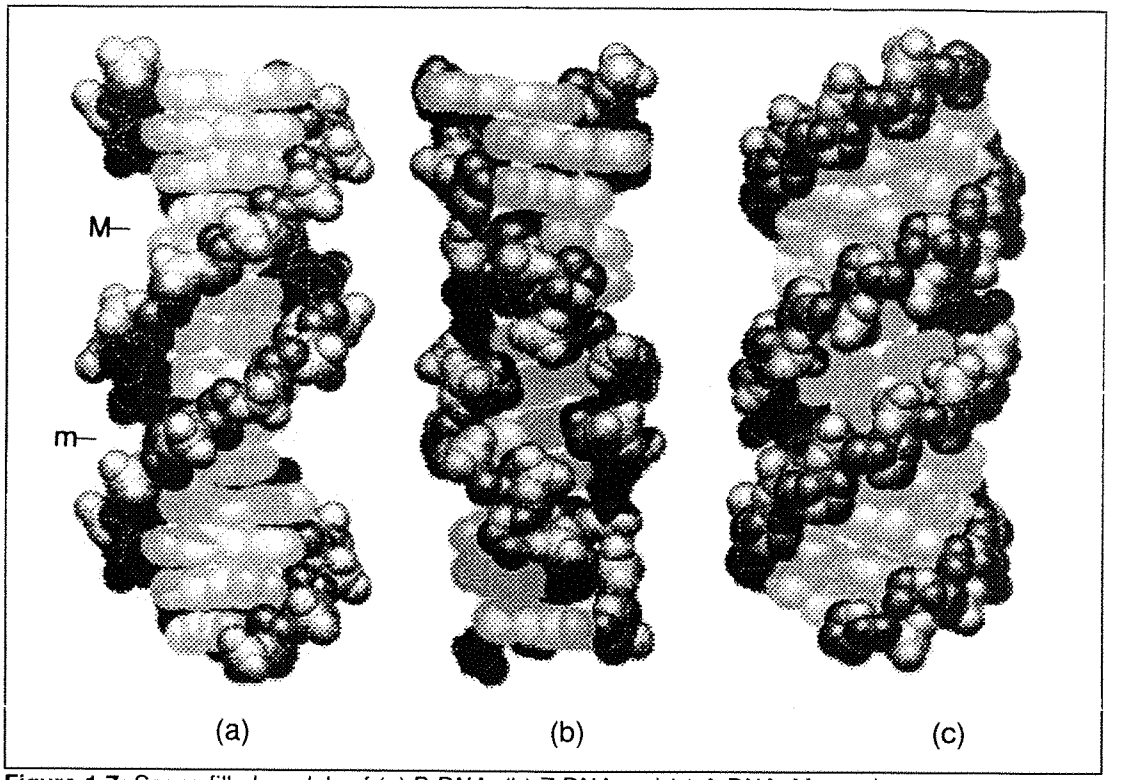


Figure 1.7: Space filled models of (a) B-DNA, (b) Z-DNA and (c) A-DNA. M = major groove; m = minor groove. Taken from Saenger (1984).

More importantly in the present context, crystal structures of oligodeoxynucleotides that had an overall B-DNA structure showed considerable deviation from classical uniform B-DNA at the local level (Dickerson & Drew, 1981; Nelson *et al.*, 1987; Drew *et al*, 1988).

Some of the structural parameters used to describe these variations are illustrated in Figure 1.9. In some cases these variations have been shown to be dependent upon DNA sequence. For example, propeller twist is normally greater in T.A base pairs than G.C pairs (Dickerson & Drew, 1981; Saenger, 1984). This, in part, is due to the extra resistance to twist of three hydrogen bonds in G.C base pairs as opposed to the two in A.T base pairs. However, a much larger flattening of propeller twist is seen when the next base pair to a G.C base pair contains a deoxyguanosine in the opposite strand due to steric clash between the two deoxyguanosines. Propeller twist of the two base pairs brings the 2-NH_2 groups on each deoxyguanosine into close contact thus limiting the amount of twist possible. If, however a deoxyadenosine is in the opposite strand then large twist is possible since the deoxyadenosine has no 2-NH_2 group that would clash with that of the deoxyguanosine.

Such sequence dependent local structural variations could provide a different mechanism for the recognition of a particular sequence by a protein that does not involve direct contact of the protein to the DNA bases but rather recognition of the DNA conformation caused by the particular sequence. This type of recognition is known as 'indirect' readout (Otwinowski *et al.*, 1988).

Another deviation from 'normal' B-DNA now known to be a common feature of DNA is curvature of the helical rod. The ability of a stretch of DNA to undergo a bend and the

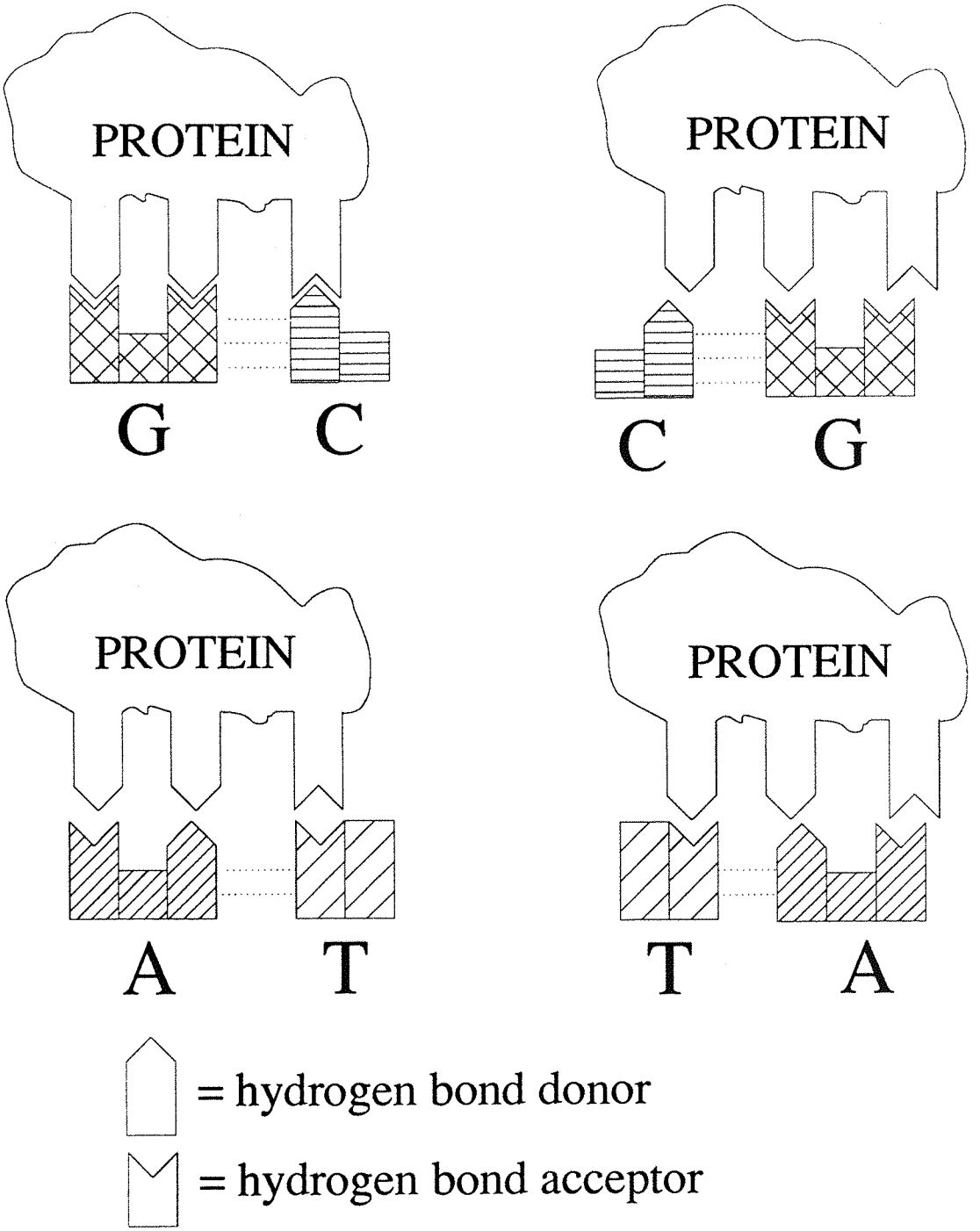


Figure 1.8: Example of protein/DNA direct readout via the major groove. In the case shown here, only a G.C base pair can form all three hydrogen bonds.

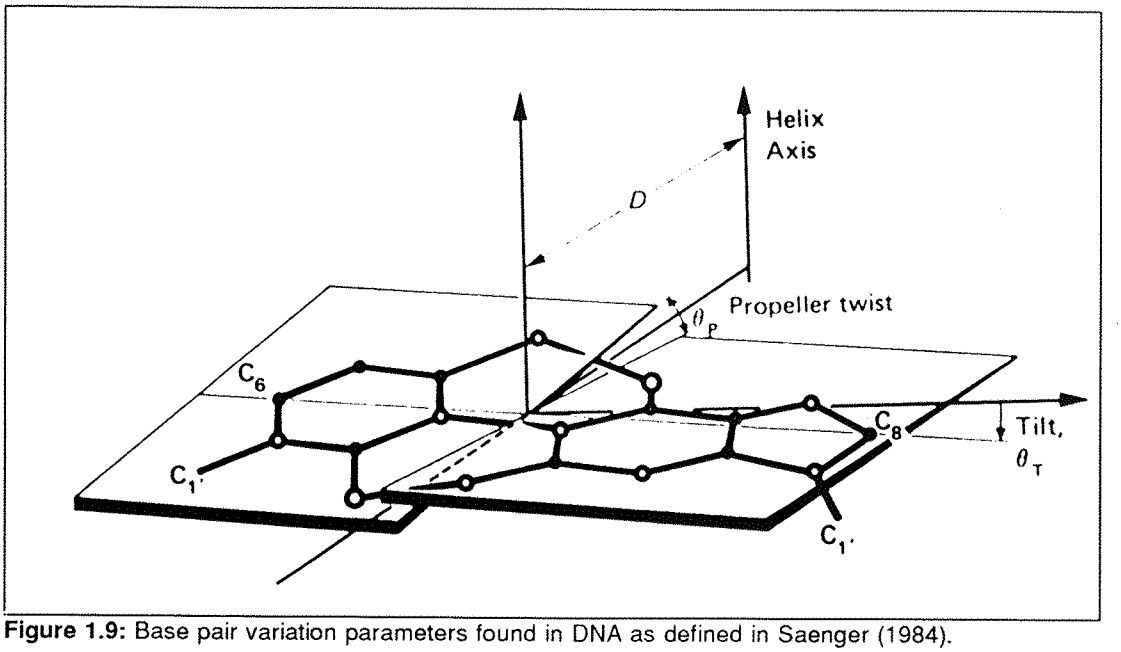


Figure 1.9: Base pair variation parameters found in DNA as defined in Saenger (1984).

direction in which it is most easily bent are considered to be strongly dependent upon the sequence of the DNA (Calladine & Drew, 1986). Several protein/DNA complexes have been shown to have protein bound to bent DNA (Aggarwal *et al.*, 1988; Otwinowski *et al.*, 1988) and since flexibility of DNA is to some extent sequence dependent, the need to be able to bend the DNA in order to form the correct complex may confer a certain degree of sequence discrimination in the protein (Travers, 1989).

1.4 GENERAL FEATURES OF PROTEIN/DNA COMPLEXES

The structure of several DNA binding proteins bound to their recognition sequence have been solved and so far four classes of DNA binding motifs have emerged (Harrison & Aggarwal, 1990; Freemont *et al*, 1991; Harrison, 1991).

1.4.1 Helix-Turn-Helix Proteins

The most common of these DNA binding motifs is the helix-turn-helix motif first proposed on the basis of the structure of the λ Cro and CAP protein structures (Steitz *et al.*, 1982) and has since then been found in a number of prokaryotic repressors and activators. The motif consists of about twenty amino acids that form two helices linked by a tight β-turn so that the helices are at an angle of around 90 degrees to each other. One of the helices, normally termed the recognition helix, sits in the major groove of DNA in the protein/DNA complex and provides the majority of the contacts to the bases. The N-termini of both helices point towards the phosphate backbone on either side of the major groove and help to anchor the recognition helix position through hydrogen bonds and through the electrostatic interaction

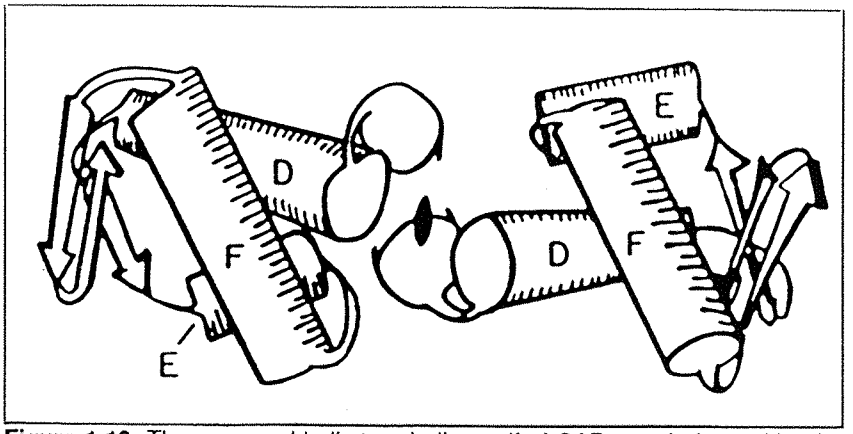


Figure 1.10: The proposed helix-turn-helix motif of CAP protein formed by the two α-helices, E and F, with the F helices fitting into the DNA major groove (Steitz *et al.*, 1982).

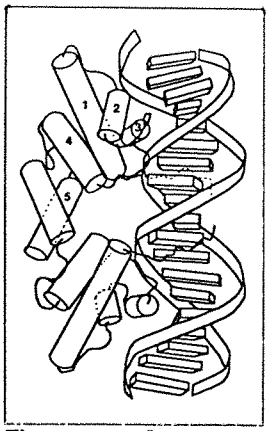


Figure 1.11: Structure of λ-repressor bound to DNA (Jordan & Pabo, 1988).

of the positive helix dipole with the negatively charged phosphates (Figure 1.10). The majority of helix-turn-helix proteins are dimers and bind to two successive major grooves as shown in Figure 1.11.

Although they employ the same recognition motif, helix-turn-helix proteins interact with their target DNA in significantly different ways. This is reflected in the variation of the exact positioning of the recognition helix in the major groove, the structure of the DNA itself and in the specific contacts that the protein makes to the DNA bases. The latter two points enable helix-turn-helix proteins to utilise different mechanisms for sequence specific recognition of DNA. In the λ repressor complex with its operator DNA sequence, O_L1 (Jordan & Pabo, 1988), sequence recognition is achieved by specific contacts between the DNA bases and amino acids in the recognition helix and also from its N-terminal arm that wraps around the DNA in the major groove (Figure 1.11). The specificity of the 434 repressor for its consensus sequence ACAAXXXXXXTTGT is also largely achieved by direct contacts between the protein and the outer four DNA bases (Aggarwal *et al.*, 1988). However, the specific contacts do not wholly define the sequence specificity of 434 repressor since they do not involve the inner base pairs which, to a lesser extent, do affect the binding of the repressor. The DNA in the complex is somewhat bent and is considerably overwound in the centre and it is thought that the different ability that different central DNA sequences would have to adopt this conformation gives rise to the different binding affinities (Figure 1.12). In the case of the *E. Coli* TrpR repressor only a few direct contacts are made to the base pairs and these are not enough to explain the specificity observed (Otwinowski *et al.*, 1988). Two other mechanisms are suggested to be involved. In the crystal structure, three tightly bound water molecules are each hydrogen bonded both by the protein and the DNA and it is thought that this provides specific recognition of the DNA bases involved by water mediated hydrogen bonds to the protein. For example, one of the water molecules is hydrogen bonded to the N7 of a deoxyadenosine, the O6 of a deoxyguanosine and the main chain NH of residue 80 in the repressor thus linking the two bases to the protein (Figure 1.13). As for 434 repressor, indirect

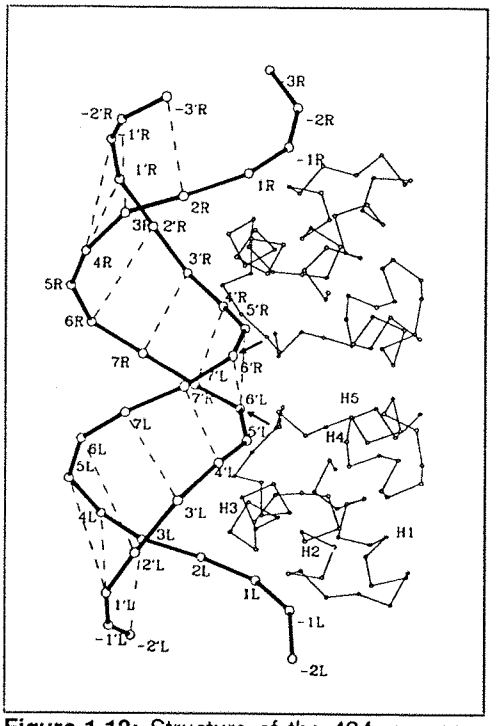


Figure 1.12: Structure of the 434 repressor bound to DNA (Aggarwal *et al.*, 1988).

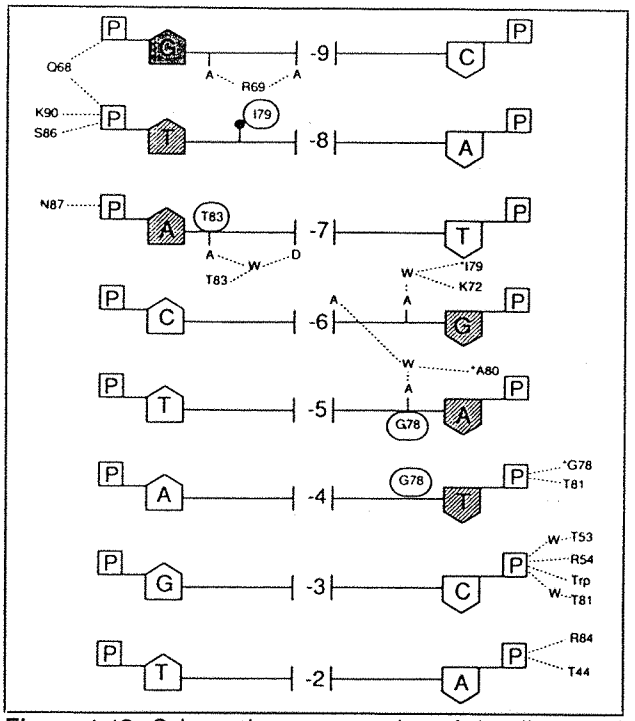


Figure 1.13: Schematic representation of the direct and water mediated hydrogen bonds in the trp repressor/ DNA complex (taken from Freemont *et al.*, 1991).

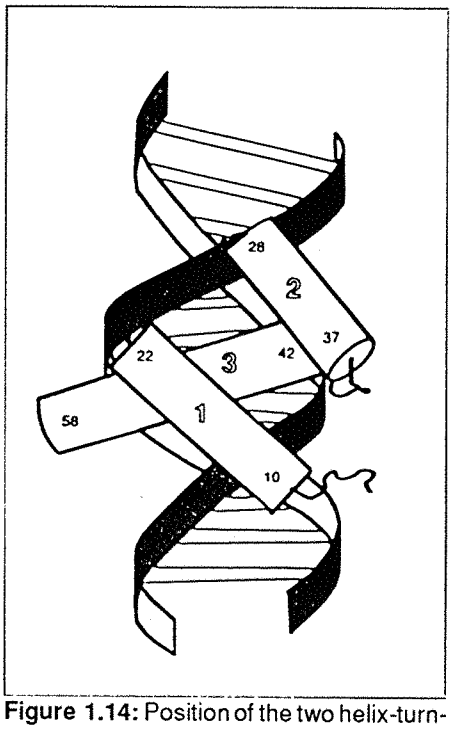


Figure 1.14: Position of the two helix-turn-helix helices (2 & 3) in the DNA of the engrailed homeodomain/DNA complex (Kissenger *et al.*, 1990).

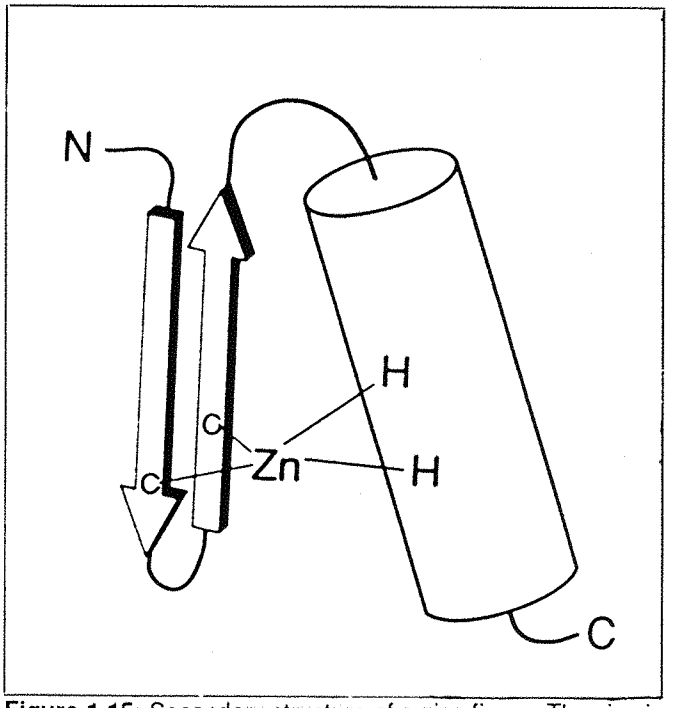


Figure 1.15: Secondary structure of a zinc finger. The zinc is coordinated to two cysteine residues in a β-sheet and two histidines in an α-helix (taken from Freemont *et al.*, 1991).

readout is also thought to be involved since the DNA in the complex is distorted (although to a lesser extent than 434 repressor) to fit the surface of the trp repressor.

Helix-turn-helix proteins have been found not only in prokaryotic repressors but in various types of DNA binding proteins that even includes a class of eukaryotic proteins known as homeodomains (Qian *et al.*, 1989; Kissenger *et al.*, 1990) and so appears to be a universal DNA recognition motif (Figure 1.14).

1.4.2 Zinc Finger Proteins

Another protein structure that now appears to be quite common in DNA binding proteins, is the zinc finger, which involves the coordination of zinc to conserved cysteine and histidine residues which although far apart in the amino acid sequence, are brought together in the zinc finger domain. The structure of a zinc finger was first suggested by Berg (1988) to consist of an antiparallel two stranded β-sheet containing the two conserved cysteine residues linked through coordination to the zinc atom to the two conserved histidine residues in an α-helix as shown schematically in Figure 1.15. A single zinc finger on its own, although able to fold correctly, binds poorly to DNA and DNA binding proteins in this class have from two to nine zinc fingers. Berg also addressed the question of how several zinc fingers in a single protein could be arranged in order to bind to DNA. Two models for DNA binding had previously been proposed (Fairall *et al.*, 1986) and in each case the α-helix of each finger lies in the major groove of the DNA. Model I involves the protein wrapping around the DNA with the zinc fingers following the path of the helix in the major groove, as shown in Figure 1.16. In model II, the protein is bound on one face of the DNA with the zinc fingers sitting in the major groove on alternate sides of the DNA (Figure 1.16, model II). Bergs' proposed secondary structure for the zinc finger favoured the first model and the first crystal structure of a zinc finger/DNA complex solved was found to have essentially the same secondary structure as that proposed by Berg and to bind to DNA in a manner consistent with model I (Lee *et al.*, 1989; Pavletich & Pabo, 1991). The crystal structure was of the three zinc fingers of Zif268 (a mouse immediate early protein) complexed with an 11 base pair oligodeoxynucleotide and was found to have the protein wrapped around the DNA, following the helical path, with the zinc fingers contacting the DNA in the major groove (Figure 1.17). Each finger has essentially the same secondary structure and is orientated in the same way with respect to the DNA with each finger contacting a three base pair sub-site. Contacts to the DNA bases are made by residues in the amino terminal part of the α-helix of each zinc finger with the sequence recognised by each finger appearing to depend upon the identity of the side chains of these residues. For example, fingers 1 and 3 recognise the same sub-site GCG and contact the first G through a charged arginine residue (the sixth residue in the α-helix) that hydrogen bonds with both the N7 and the O6 of the deoxyguanosine (Figure 1.18). In finger 2, which recognises the sub-site TGG, the sixth residue in the α-helix is not arginine but threonine which does not appear to be involved in DNA binding. However, the third residue in finger 2, histidine, does contact the

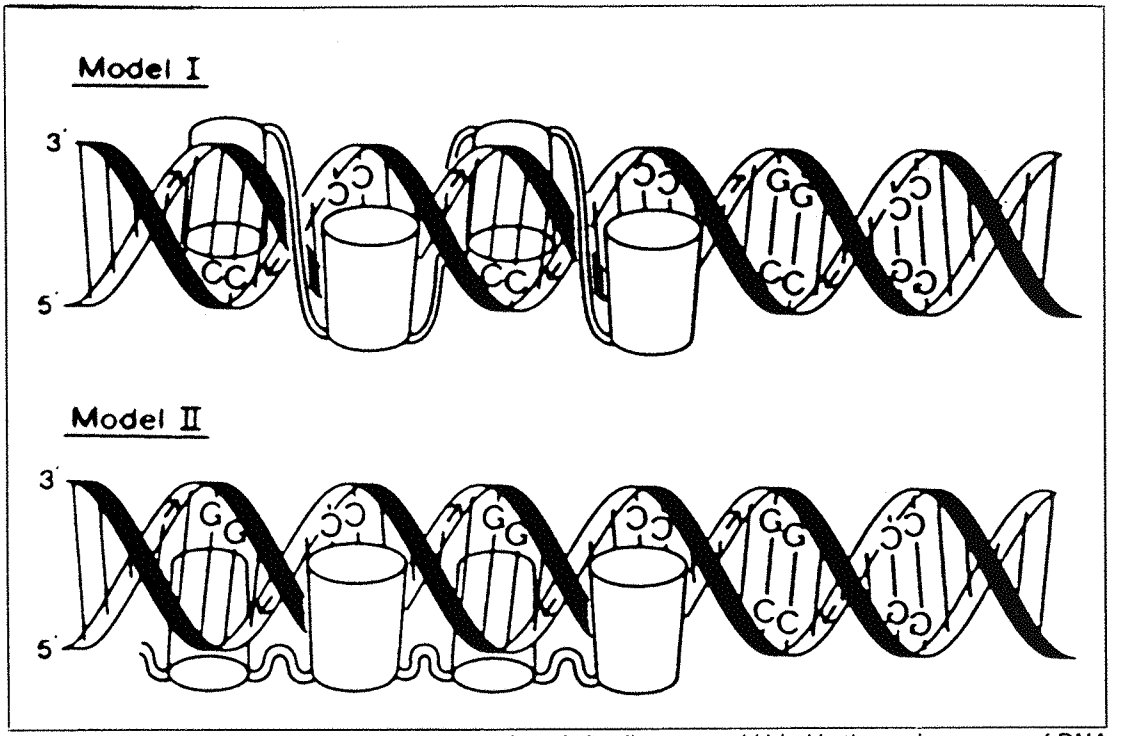


Figure 1.16: Two different models of how a series of zinc fingers could bind in the major groove of DNA (Fairall *et al.*, 1986).

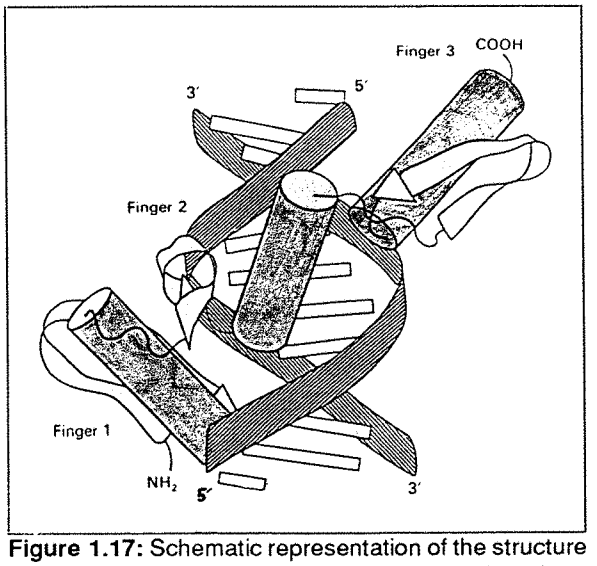


Figure 1.17: Schematic representation of the structure of the Zif269 three zinc finger complex (taken from Freemont *et al.*, 1991).

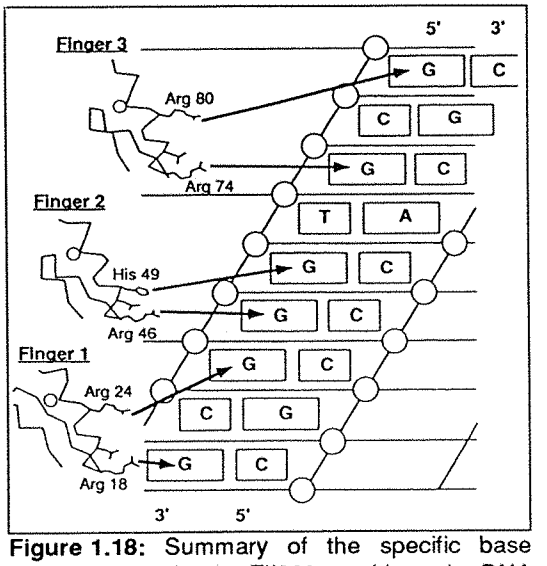


Figure 1.18: Summary of the specific base contacts made by the Zif268 peptide to the DNA (Pavletich & Pabo, 1991).

DNA making a hydrogen bond with the N7 or O6 (the X-ray data cannot define which) of the central deoxyguanosine of its sub-site (Figure 1.18). Fingers 1 and 3 have glutamic acid in this position which does not contact the DNA. From this structure it appears that zinc fingers could interact with DNA in a more conserved and uniform manner than helix-turn-helix proteins although it remains to be seen whether other zinc finger proteins will use their fingers in different ways to bind DNA.

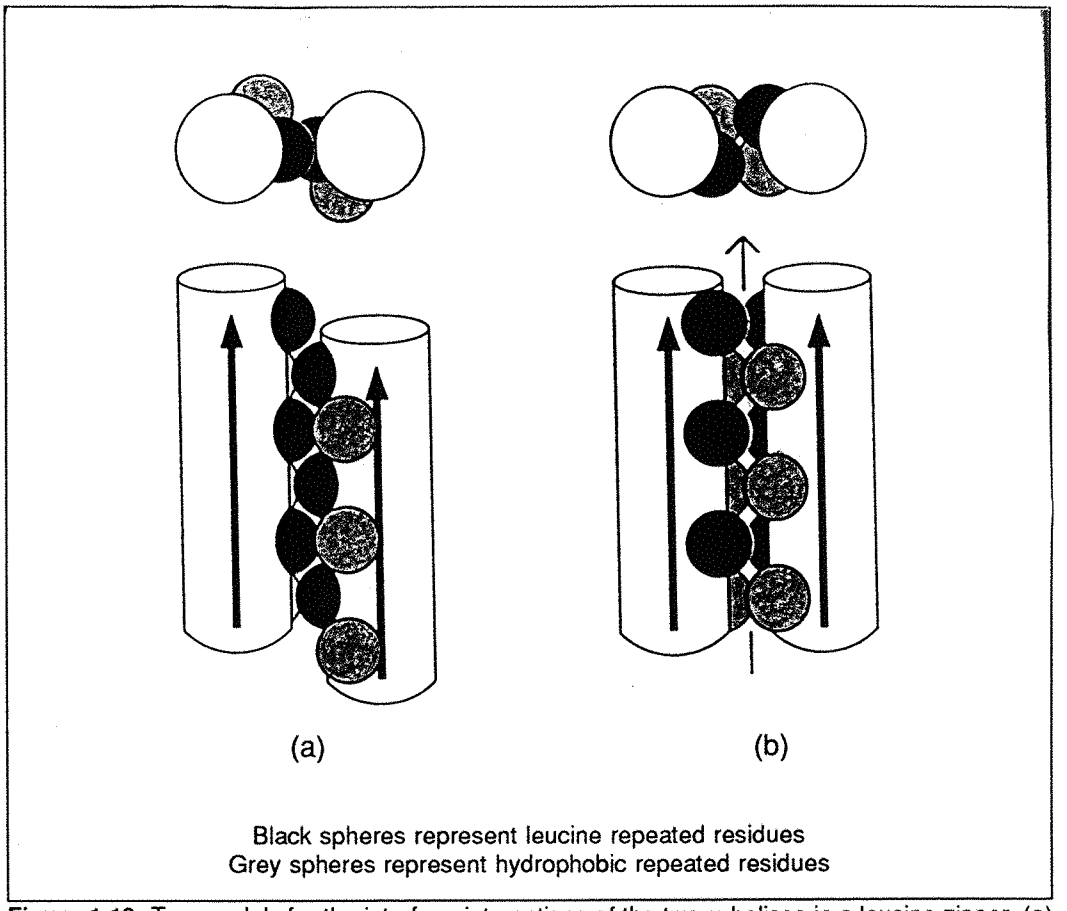


Figure 1.19: Two models for the interface interactions of the two α-helices in a leucine zipper. (a) The leucines are interdigitated which for parallel helices gives an asymmetric structure. (b) Coiled-coil arrangement which is symmetric.

1.4.3 Leucine Zipper Proteins

Another structural motif that is found in DNA binding proteins is the so called leucine zipper. Leucine zipper sequences are characterised by a periodic repeat of leucines at every seventh position which, if the zipper regions are α-helical, would place the leucines on one side of the helix. There are also hydrophobic residues repeated at every seventh position. The zipper region itself is not directly involved in the binding of protein to DNA but forms the dimer interface which was originally thought to involve the two α-helices on separate subunits coming together antiparallel with the leucine residues interdigitating forming the 'zipper'. Cross linking studies however have indicated that the α-helices are parallel (O'Shea *et al*, 1989). 2D-NMR studies did indeed show that a synthetic 33 residue zipper dimer region of the yeast transcriptional activator GCN4 was helical but that only one set of subunit proton signals was present showing that the dimer must be symmetrical (Oas *et al.*, 1990). This argues against the kind of arrangement depicted in Figure 1.19a where the leucines are interdigitated since in this model the subunits are asymmetrical. A coiled coil conformation as shown in Figure 1.19b is symmetrical and would therefore give rise to a single set of NMR signals.

An X-ray structure of the zipper region of GCN4 recently published is in agreement

with the dimer adopting a parallel coiled coil conformation (O'Shea *et al.*, 1991). The DNA binding region consists of a stretch of about 30 residues that are to the amino terminal side of the zipper α-helix and consists of a large proportion of basic residues (Nekabeppu & Nathans, 1989). The X-ray studies also showed that a significant amount of the basic region is α-helical although the amount of helix appears to vary considerably with temperature indicating that it is poorly defined.

1.4.4 β-Proteins

In contrast to all the proteins described above which make contacts with the DNA mainly through α-helices, the beta proteins use a two stranded β-sheet structure that binds in the grooves of DNA. In the recently soloved crystal structure of the *E. Coli* MetJ repressor the recognition motif is formed by a β-strand from each subunit coming together to form an antiparallel β-ribbon that sits in the major groove of the operator DNA (Rafferty *et al.*, 1989; Figure 1.20).

The protein makes eight hydrogen bonds to the bases in the major groove through this β-sheet as well as some non-specific contacts to the phosphate backbone by amino acids outside the β-sheet region. The DNA in the complex is bent towards the protein by about 25 degrees resulting in compression of the major groove. The complex also includes the corepressor S-adenosylmethionine needed for increased binding to DNA although its mode of action is not clear from the crystal structure.

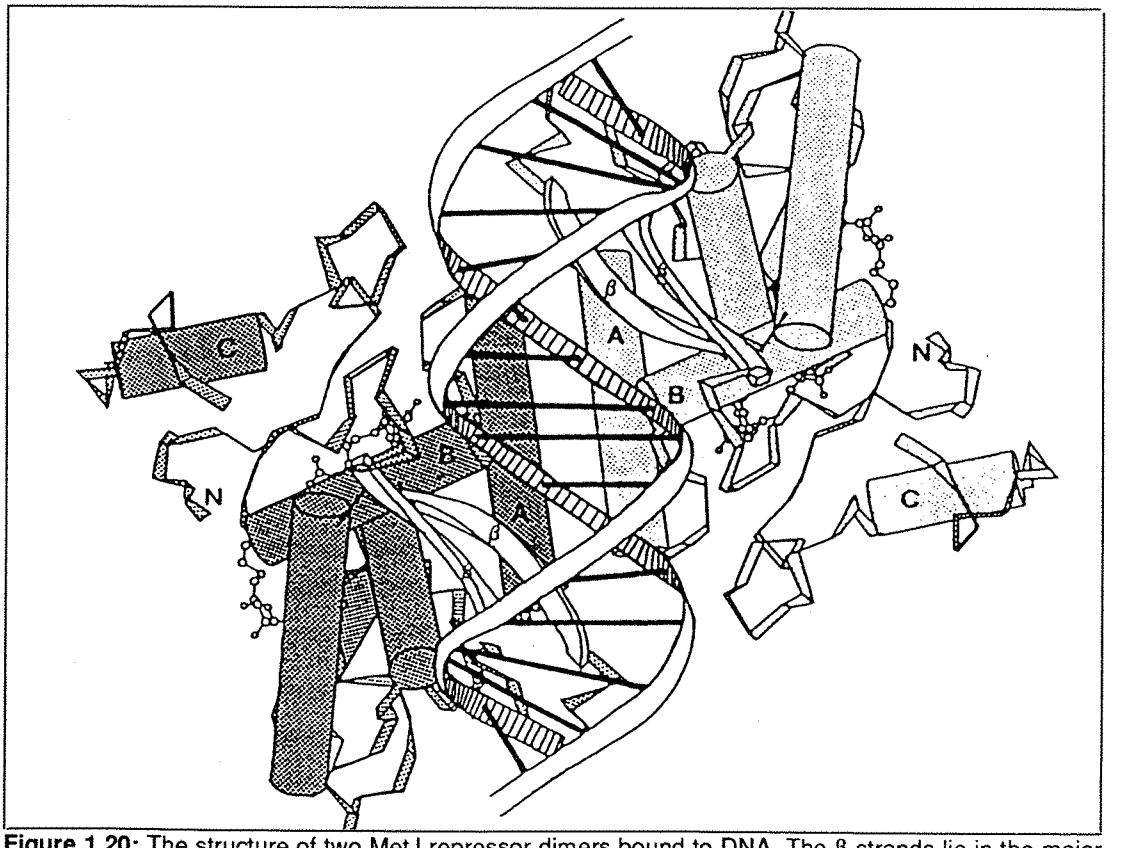


Figure 1.20: The structure of two MetJ repressor dimers bound to DNA. The β-strands lie in the major groove of the DNA (taken from Perutz, 1990).

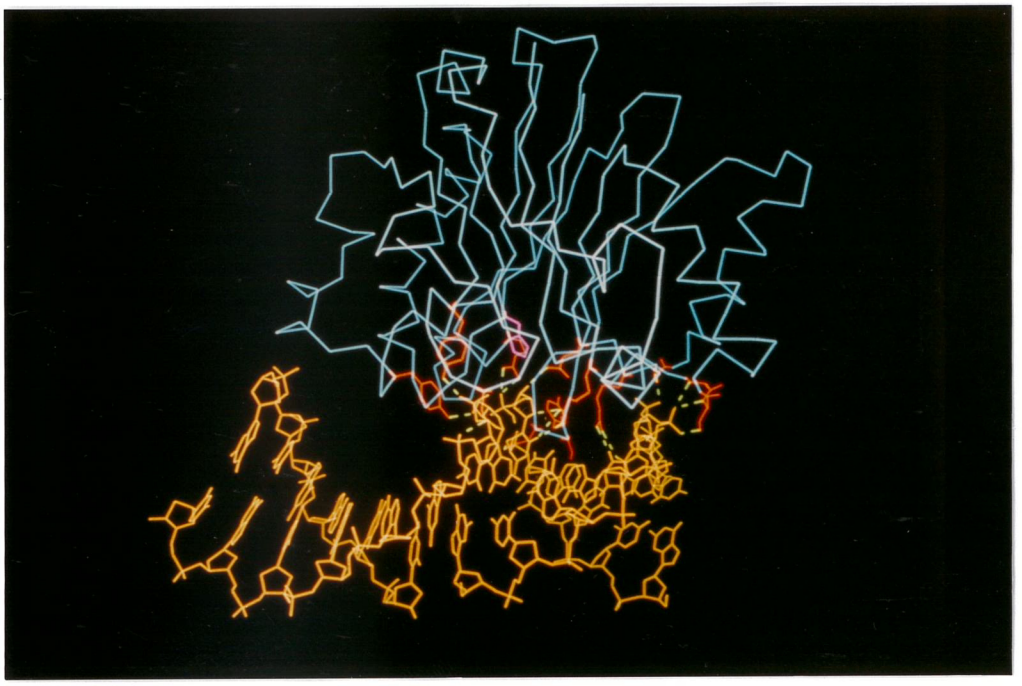


Figure 1.21: Structure of DNase I (blue) bound to an oligodeoxynucleotide (yellow). Residues of the enzyme contacting the DNA are coloured red (photograph courtesy of D.Suck, EMBL, Heidelberg).

1.4.5 Other DNA Binding Proteins

The three dimensional structures of several DNA modifying enzymes have been solved and have been found to have modes of DNA binding that do not fall into any of the structural classes described above. These include DNase I, which is an endonuclease that cleaves DNA at all sites but with different DNA sequences being cleaved at different rates (Drew & Travers, 1984). This variation in the rate of cleavage can be divided into two trends; global and local. The global trend is the rate of cleavage averaged over about four base pairs and therefore ignores local variations in rate. The global rate changes gradually over a stretch of DNA and was found to have some dependence upon DNA sequence; stretches of A.T or G.C base pairs were found to be cleaved very poorly. This was interpreted as being due to the width of the minor groove of the DNA where the DNase I molecule was thought to bind. It was thought that the protein preferred a 'normal' minor groove width of about 12Å and that runs of A.T would cause the groove to become narrower than the optimum whilst runs of G.C would cause the groove to become wider. Both would result in DNase I having a lowered affinity at such sites and hence a reduced rate of cleavage. The local trend in cleavage rates was defined as the differences in rates of cleavage that was observed between adjacent bases. In some cases there was a stark contrast in the rate of cleavage between adjacent bases but there was no discernable sequence dependence. It was suggested that the local trend could be due to variations in the orientation of phosphates.

The structure of DNase I co-crystallised with an octanucleotide that forms a pseudo-continuous 14mer in the complex has been solved to a resolution of 2Å (Suck *et al.*, 1988). The structure (Figure 1.21) shows that the protein does indeed bind in the minor groove of the DNA

but that the DNA is somewhat distorted in the complex. The DNase I contacts the DNA through an exposed protein loop and this causes the minor groove to widen somewhat. Also the DNA helix is bent away from the complex towards the major groove by 21.5 degrees. The fact that the minor groove is widened in the complex should mean that runs of G.C that have a wider than normal minor groove should be cleaved at a faster rate than random sequences of DNA where in fact the opposite is observed. The minor groove width cannot therefore be the sole determinant of the global cleavage trend. The bending of the DNA in the complex probably plays a part in the global variation in cleavage rates since sequences that are amenable to being bent towards the major groove would more readily form a complex than sequences that are inherently rigid. The co-crystal provided no insight into the DNA structural parameters that affect local cleavage trends.

1.5 ECO RI ENDONUCLEASE

The Eco RI type II restriction/modification system is the most thoroughly studied type II system. The sequence recognised by the system is duplex d(GAATTC) with the endonuclease cleaving (|) between the G and the A bases and the methylase adding a $-CH_3$ group (m) to the 6-amino nitrogen of the inner A base (Dugaiczyk *et al.*, 1974):-

G | A A^m T T C
C T T A^m A | G

Both the enzymes have been prepared from *E. Coli* cells and purified to homogeneity (Modrich & Zabel, 1976; Rubin & Modrich, 1977) and both their gene sequences elucidated (Newman *et al.*, 1981; Greene *et al.*, 1981). Despite the fact that they recognise the same DNA site there is no significant homology between the two enzymes nucleotide sequences or their amino acid sequences, as determined from their nucleotide sequences. However, the endonuclease has been found to have a 50% nucleotide sequence identity with that of Rsr I endonuclease (Stephenson *et al.*, 1989), a type II endonuclease from *Rhodobacter Sphaeroides* and an isoschizomer of Eco RI. The similarity at the amino acid level was found to be even greater and resided in regions of the Eco RI endonuclease known from crystal structure and mutagenesis data to be important in the functioning of the enzyme.

The two Eco RI enzymes show little similarity at the functional level. The endonuclease acts as a dimer and under physiological conditions can cleave both strands of DNA in one binding event (Modrich & Zabel, 1976). In contrast the methylase functions as a monomer and can only transfer one methyl per DNA binding event (Rubin & Modrich, 1977).

1.5.1 Physical and Catalytic Properties of the Eco RI Endonuclease

The gene sequence of the Eco RI endonuclease gave a polypeptide of 31,000Da molecular weight (Newman *et al.*, 1981) which agrees well with the molecular weight of 29,000Da for the endonuclease determined by gel electrophoresis under denaturing conditions (Modrich & Zabel, 1976). In solution however, the active form of the endonuclease was found to be a dimer of this subunit (Modrich & Zabel, 1976; Greene *et al.*, 1975).

Cleavage of DNA by the Eco RI endonuclease occurs on the 3' side of the scissile phosphate and results in inversion of the phosphate (Connolly *et al.*, 1984a). The simplest explanation of the inversion of the phosphate is that the reaction proceeds by direct attack on the phosphate and not via the formation of a covalent enzyme/DNA intermediate.

The cleavage of double stranded DNA by the endonuclease is a two step mechanism which involves sequential cleavage of the two strands. With Col E1 plasmid DNA as substrate, an enzyme bound intermediate which is cleaved in one DNA strand was implicated. Therefore, the enzyme cleaves one strand of the Col E1 substrate and then the other strand before dissociating (Modrich & Zabel, 1976). From single turnover experiments, the cleavage of the second DNA strand was found to be only slightly slower than cleavage of the first strand and the rate determining step was deduced to be product release by the enzyme.

With plasmid pMB9 as substrate however, an intermediate product containing a single strand break was found to accumulate to concentrations greater than that of the enzyme (Halford, 1983). The intermediate in this case must therefore dissociate from the enzyme before the second strand scission occurs. Other DNA substrates have been found to fall into one or other of these categories but the factors determining which path they follow are not known. It may be the nature of the DNA sequences flanking the recognition site that are contacted directly by the protein or it may be that the secondary structures of the DNA substrates around the recognition site differ.

Flanking sequences are known to affect the rate of cleavage by the Eco RI endonuclease. For example, under certain conditions the rate of cleavage of the five Eco RI sites on bacteriophage λ vary markedly from site to site (Halford *et al.*, 1980). The simplest explanation of this observation is that the flanking sequences vary from site to site and have differing effects on the cleavage reaction. Evidence of direct involvement of outside sequences came from the cleavage of a DNA substrate consisting of a duplex of two synthetic decadeoxynucleotides:- d(GGGAATTCTT).d(AAGAATTCCC) (Alves *et al.*, 1984). It was found that the rate of cleavage of the former strand was slower than that of the latter strand. The enzyme would appear, therefore, to prefer A.T base pairs on the 5' side of the recognition sequence to G.C base pairs, although the results could not distinguish between the cause being direct contact to the flanking bases by the enzyme or to the indirect effect the flanking bases could have on the DNA secondary structure.

1.5.1.i Facilitated Diffusion Mechanism

The endonuclease binds to non-specific sites as well as its recognition sequence and it has been suggested that it can then diffuse along the DNA to a specific Eco RI site (Jack *et al.*, 1982). This facilitated diffusion mechanism is envisioned to involve the random collision of endonuclease with DNA producing (predominantly) non-specific complexes and the enzyme then moving randomly along the DNA until a recognition site is encountered. Such a mechanism would result in the rate of association of the enzyme with its target sequence being faster than that expected for normal diffusion limited collision. If such a mechanism was involved in Eco RI then the difference in cleavage rates observed for the five Eco RI sites on λ DNA (Halford *et al.*, 1980) could be due to the flanking sequences affecting the efficiency of the one-dimensional diffusion along the DNA in different ways. Such a model, however, does not explain the different rates observed for different flanking bases in a synthetic oligodeoxynucleotide (Alves *et al.*, 1984) since such a substrate is too small for non-specific binding outside the recognition site. If one-dimensional diffusion along DNA is an important factor in the formation of specific complexes by Eco RI endonuclease, then the size of the substrate should affect the rate of association between the enzyme and its recognition sequence. With short substrates only direct collision between enzyme and target sequence can result in a specific complex. As the length of the substrate increases the involvement of non-specific complex formation followed by diffusion along the DNA to the target site becomes increasingly more important resulting in an increase in the rate of formation of specific complexes as the size of the DNA increases. This was in fact found to be the case for the Eco RI endonuclease.

It had been shown previously from filter binding studies that Eco RI endonuclease binds preferentially to its recognition site, without cleavage, in the absence of magnesium ions (Halford & Johnson, 1980; Terry *et al.*, 1983). Such complexes were formed between Eco RI endonuclease and a number of DNA substrates of between 34bp and 6,200bp in length (Jack *et al.*, 1982). Each of these substrates contained a single Eco RI site roughly in the middle of the polynucleotide. As the DNA length was increased the rate of dissociation of these preformed complexes was found to increase until a plateau was reached at about 4,000bp when increasing chain length no longer affected the dissociation rate (Figure 1.22). If however, solutions of DNA and endonuclease were allowed to reach equilibrium, then the concentration of specific complexes was found to be independent of DNA chain length. That is, the equilibrium constant for the enzyme binding its target sequence was the same for DNA substrates of all lengths. Since the equilibrium constant is the ratio of dissociation rate to association rate, these two results imply, as predicted from the facilitated diffusion mechanism, that the rate of specific complex formation increases with increasing DNA chain length. This was confirmed by the direct measurement of rates of association which were determined by a preferential cleavage assay. Radiolabeled DNA and Eco RI endonuclease were mixed together in the absence of magnesium and samples removed at certain times after the addition of the

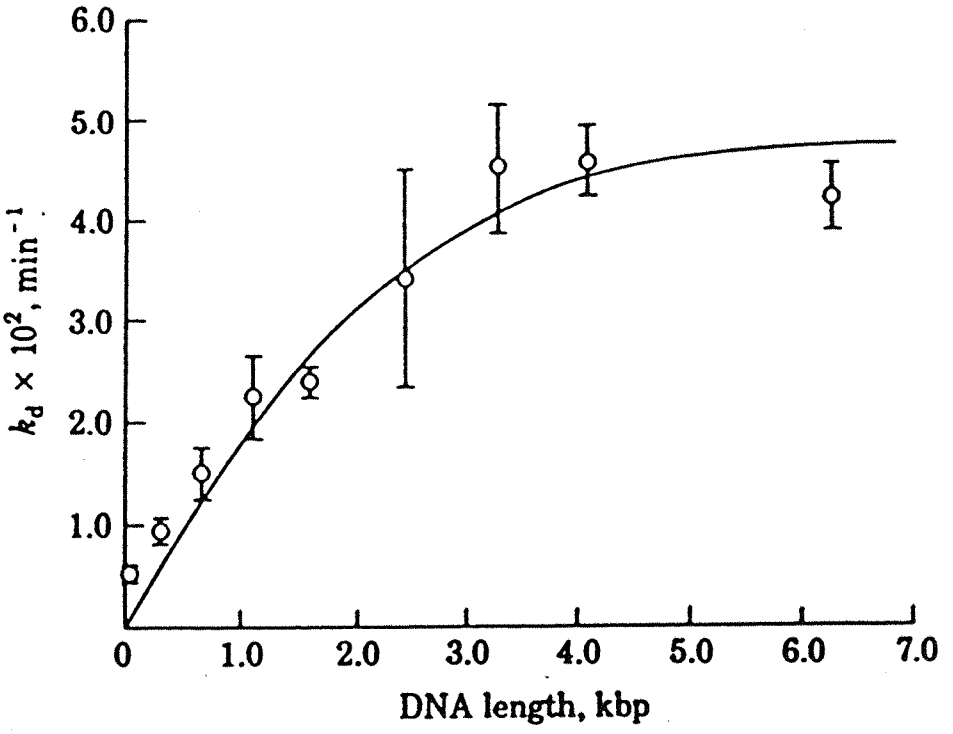


Figure 1.22: Dependence of dissociation rate constant on DNA chain length (Jack *et al.*, 1982). Error bars are for 95% confidence limits.

endonuclease. Excess unlabelled plasmid DNA was added together with magnesium ions and the reaction allowed to continue for 30 seconds. Under these conditions only preformed specific complexes between the endonuclease and the labelled substrate should result in cleavage of labelled substrate. Quantification of the labelled cleavage products should therefore give the amount of specific complexes formed at the time the sample was removed. Association rates determined using this assay were faster for longer DNA substrates than for shorter substrates in agreement with the equilibrium and dissociation data.

1.5.2 Solution Studies Investigating the Specific Protein/DNA Interactions of the Eco RI Endonuclease

Alkylation interference and protection studies have been used to probe potential Eco RI endonuclease/DNA contacts (Lu *et al.*, 1981). A 34bp fragment of DNA containing a single Eco RI site was used as the binding substrate. In interference studies of the purines involved in the binding of the enzyme, the 34bp DNA was randomly methylated with dimethylsulphate (a reagent that methylates the N3 of deoxyadenosines and the N7 of deoxyguanosines) so that each DNA duplex had on average one methylated base. The modified DNA was then mixed with the endonuclease, allowed to reach equilibrium, and then enzyme-bound DNA was removed from free DNA by nitrocellulose filtering. Those purines which when methylated reduced the affinity of the enzyme for its cognate site could thus be identified. Protection studies of the purines were also carried out in which the endonuclease/DNA complexes were

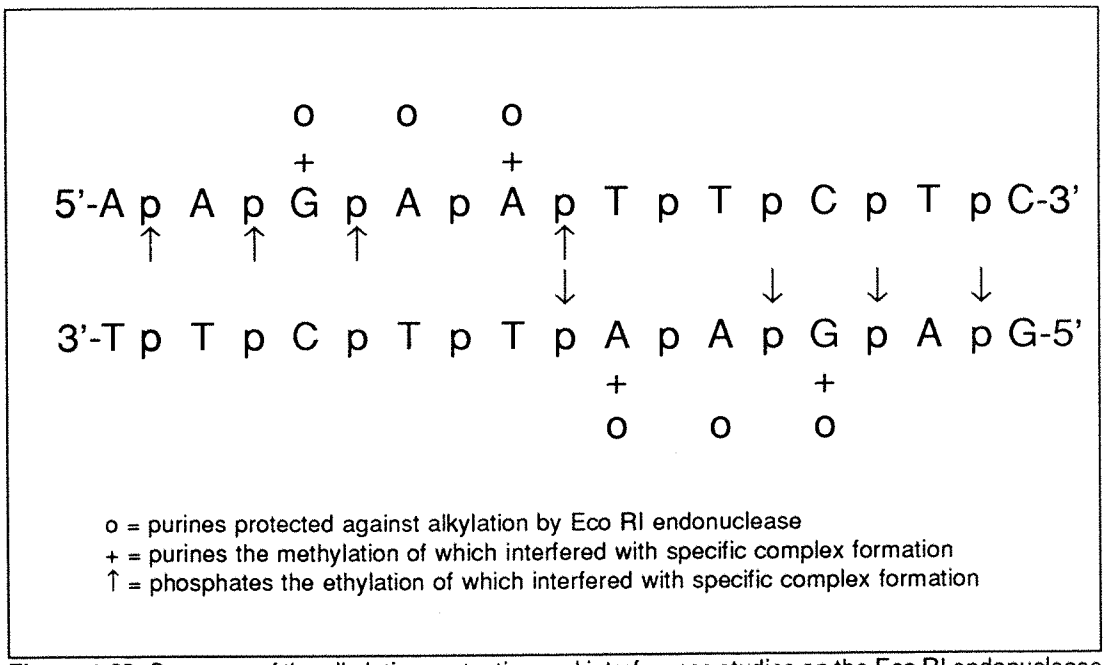


Figure 1.23: Summary of the alkylation protection and interference studies on the Eco RI endonuclease carried out by Lu *et al.* (1981).

formed first. These complexes were then exposed briefly to the methylating reagent and the purines that were protected by the endonuclease against methylation identified. The results are summarised in Figure 1.23 and show that only bases within the recognition site are involved. Methylation of all the purines within the recognition site was prevented by enzyme/DNA complex formation and the methylation of the deoxyguanosines and the inner (but not the outer) deoxyadenosines significantly interfered with binding. An important feature of the results is the symmetry between the two DNA strands implying that the enzyme dimer forms a symmetric complex. When the methylation protection and interference experiments were carried out with a DNA substrate that had been methylated (on the N6 of the inner deoxyadenosine) by the Eco RI methylase, no specific protection or interference was observed. This is consistent with the view that modification of DNA by the Eco RI methylase prevents formation of specific Eco RI endonuclease/DNA complexes.

Similar interference experiments were performed by the same group on the DNA phosphates. Phosphates were randomly ethylated and then those ethylphosphates that interfered with endonuclease binding identified. These results are also shown in Figure 1.23 and again show symmetry between the two DNA strands. Another important feature of these results is the fact that phosphates outside the recognition sequence are implicated in binding. This therefore provides evidence that base sequence outside the recognition site could have a direct effect on the endonuclease reaction. It should be borne in mind that drawbacks of this method may have a significant effect on the results. Firstly the experiments were carried out in the absence of magnesium ions, the presence of which may alter the way the endonuclease binds to DNA. Secondly, it is possible that the presence of the alkyl groups on the DNA causes the DNA to alter its conformation from that of the native form making its interactions with

In both cases, oligodeoxynucleotides containing a base analogue that had one of the potential contact points removed were tested for cleavage by the endonuclease. Since these oligodeoxynucleotides are self complementary, the modified base appears in both strands and so two potential hydrogen bonds are removed. Oligodeoxynucleotides that were cleavable then had their Michaelis-Menten parameters, k_{cat} and K_M, determined in order to give a measure of the effect of the modification. The specificity constant (k_{cat}/K_M) of a modified substrate relative to that of the unmodified substrate gives an indication of the importance of the functional group deleted (Fersht, 1985; and see section 1.2.1). As a rough guide, the loss of two hydrogen bonds in water results in a typical energy change of −13kJ/mol (Fersht, 1987) which would give rise to a relative specificity of 0.005 at 25°C.

Lesser *et al.* (1990) also looked at the effect of modified base substitutions in one strand of the Eco RI site of a synthetic oligodeoxynucleotide substrate. Instead of measuring k_{cat} and K_M however, they measured the rate of cleavage of one strand in a preformed enzyme/DNA complex (k_1) and the equilibrium constant (K_A) for the association of enzyme and DNA in the absence of cleavage. In this case, relative values of $k_1 \times K_A$ give a measure of the energy lost on deletion of a contact point analogous to the k_{cat}/K_M values.

The results obtained for all these experiments discussed above that involve deletion of a potential contact point are summarised in Table 1.1 and are discussed below.

G.C base pair: The relative specificities of the d[^{7C}G] and the d[^{6H}G] containing oligodeoxynucleotides indicate that a hydrogen bond is lost in both cases implying that the ring 7-nitrogen and the 6-carbonyl groups are both contacted by the protein. The three results for the d[I] containing substrates taken together indicate that there is no direct contact to the 2-amino group of deoxyguanosine in the minor groove. The small decrease (approximately ten fold) in relative specificity observed for these substrates is presumably due to some indirect effect on the structure of the oligodeoxynucleotide caused by the presence of the deoxyinosine base. Indeed, it has been shown that the replacement of deoxyguanosine by deoxyinosine in the Eco RI site of d(CTGAATTCAG) increases the curvature of the oligodeoxynucleotide by about 10% (Diekmann & McLaughlin, 1988). It was proposed that the absence of a substituent on deoxyinosine in the minor groove was responsible for the increase in the curving of the DNA. The details of how the increased curvature of the DNA affects the endonuclease reaction are not known but it appears in this case that the replacement of deoxyguanosine by deoxyinosine involves more than the simple deletion of a potential hydrogen bonding group.

The only result for deoxycytidine analogues is the one for the 5-methyldeoxycytidine containing octamer which was found not to be cleaved by the endonuclease. The situation in this case is somewhat different to the other analogues since a potential hydrogen bonding group has not been deleted. However, the result shows that the presence of the methyl group on 5-methyldeoxycytidine (and by implication the methyl group on thymidine in a non-cognate site) cannot be tolerated by the enzyme.

Table 1.1: Summary of the reaction of Eco RI endonuclease with substrates (either oligodeoxynucleotide or plasmid) containing modified bases. [a]Values given as relative k_{cat}/K_M; [b]values given as relative $k_1 x K_A$; [c]values given as relative rates; [d]not double stranded; [e]did not cleave; [f]see text.

ANALOGUE	McLaughlin *et al.*, 1987 (k_{cat}/K_M)[a]	Brennan *et al.*, 1986 (k_{cat}/K_M[a]	Lesser *et al.*, 1990 ($k_1 \times K_A$)[b]	Berkner & Folk, 1977[c]	Modrich & Rubin, 1977[c]	Contact Implied?
	CTGAATTCAG	GGAATTCC	17mer	PBS2	Col E1	
-GAATTC-	1.00	1.00	1.00	1.00	1.00	—
[I]AATTC	0.17	0.10	—	—	~1	no
[^{7C}G]AATTC	—	—	0.017	—	—	yes
[^{6H}G]AATTC	not d/s[d]	—	0.015	—	—	yes
G[P]ATTC	0.06	—	—	—	—	yes
GA[P]TTC	0.52	—	—	—	—	no
GAA[U]TC	0.34	0.28	—	—	—	no
GAAT[U]C	0.04	d/c[e]	—	—	—	yes(?)[f]
GAA[U][U]C	—	—	—	~1	—	(?)[f]
GAATT[^{5Me}C]	—	d/c[e]	—	—	—	yes

Outer A.T base pair: The result for a purine in the outer deoxyadenosine position of the Eco RI site shows that the exocyclic 6-amino group of that deoxyadenosine is contacted by the protein. The result of replacing the outer thymidine with deoxyuridine using oligodeoxynucleotide substrates would seem to indicate that the enzyme does make contact with the 5-methyl group. However, the cleavage of plasmid DNA with thymidines replaced by deoxyuridine (Berkner & Folk, 1977) would seem to suggest that the 5-methyl groups on both thymidines within the recognition site are unimportant. This result should be treated with caution though as the rate of cleavage was not compared to that of an adequate control, as discussed above.

Inner A.T base pair: The virtually unchanged specificity constant for purine replacing the inner deoxyadenosine in the decamer, shows that the 6-amino group of that deoxyadenosine is not important for the endonuclease reaction. Also, the results for replacement of the inner thymidine with deoxyuridine show that the loss of the methyl group of this thymidine has little effect on the cleavage.

1.5.3 Eco R1* activity of the Endonuclease

Under conditions of low ionic strength and high pH, the Eco RI endonuclease also cleaves at sites other than its recognition sequence (Polisky *et al.*, 1975; Woodbury *et al.*, 1980a). The majority of these non-canonical sites (termed Eco RI* sites) were found to differ from the recognition site by just one base pair. Under Eco RI* conditions the canonical site was still the best substrate and a definite hierarchy of the likelihood of cleavage could be seen for

the Eco RI* sites. The replacement of magnesium ions by manganese ions in 'normal' Eco RI buffers also produced Eco RI* activity (Hsu & Berg, 1978). The cleavage pattern of phage ϕX174 DNA in low ionic strength and high pH buffers was found to be essentially the same as the pattern produced in manganese buffers and so the two different conditions cause cleavage at the same sites on the DNA (Woodbury *et al.*, 1980a). A sequence found to be common to many of the Eco RI* sites was NAATTC (where N is any nucleotide) although no real correlation could be made between base sequence and reactivity.

In two recent studies the cleavage of oligodeoxynucleotides containing Eco RI* sites has been investigated in a more systematic way (Lesser *et al.*, 1990; Thielking *et al.*, 1990). In both cases they looked at the reactivity of all nine possible base pair substitutions (see Table 1.2) but under normal buffer conditions. The rates of cleavage for each of the two different DNA strands (which in the Eco RI* substrates are no longer equivalent) were determined for preformed endonuclease/DNA complexes. Also in both studies they measured the equilibrium constant of association for the Eco RI* substrates and the endonuclease in the absence of magnesium. The results obtained agree fairly well between the two separate sets of data. For example, they both found that cleavage in the half site that did not contain the non-canonical base pair was generally faster than in the other modified half site. One noticeable exception to this observation is the GACTTC sequence which was cleaved faster in the modified half site than in the unmodified half site for both studies. The rate of cleavage multiplied by the association constant (k_1 x K_A) gives a value which is analogous to the specificity constant obtained from Michaelis-Menten kinetics and gives a measure of the relative specificity of each substrate. Values of k_1 x K_A relative to the control sequence are given for the unmodified half sites in Table 1.2 for both sets of results. Comparison of the data in this table shows that there is reasonable, although not exact, agreement between the two sets of results. Two trends common to both sets of data can be seen. (i) Base pair substitutions in the first position of the recognition site are generally better substrates than substitutions in the third position which in turn are generally better than substitutions in the second position. This trend is more pronounced in the results of Lesser *et al.* with a prominent exception in the results of Thielking *et al.* being the GACTTC sequence which was their best Eco RI* substrate. The trend is in agreement with the earlier experiments which identified NAATTC substrates as amongst the best Eco RI* sites. (ii) Within sites there is reasonable agreement of the hierarchy of specificity seen in the two sets of results. A particularly noticeable feature is that base pair substitutions that involve base pair reversal (for example C.G for G.C in the first position) give rise to oligodeoxynucleotides that are the poorest substrates within that site.

Lesser *et al.* looked at the results obtained in more detail together with data obtained for base analogue substitution [see section 1.5.2] and converted the specificity values into losses of energy involved on going from the normal sequence to the modified substrate. They looked at the changes in the structure of the major groove from that of the cognate site for each of the base pair and base analogue substitutions and attempted to correlate these changes with the

Table 1.2: Summary of the cleavage reaction of Eco RI endonuclease at sites other than its recognition sequence. [a]Values given as rate of cleavage of unmodified half site (bottom strand as shown in column 1), k_1, multiplied by the rate of association, K_A, relative to the cognate sequence. ↑ = site of cleavage in lower strand.

CLEAVAGE SITE	Lesser *et al.*, 1990 relative $k_1 \times K_A$[a]	Thielking *et al.*, 1990 relative $k_1 \times K_A$[a]
-GAATT C- -CTTAA-G-	1.00	1.00
-**T**AATTC- -**A**TTAAG-	1.5×10^{-5}	6.3×10^{-7}
-**A**AATTC- -**T**TTAAG-	1.4×10^{-6}	3.5×10^{-7}
-**C**AATTC- -**G**TTAAG-	1.0×10^{-7}	6.2×10^{-8}
-G**C**ATTC- -C**G**TAAG-	2.7×10^{-8}	1.3×10^{-9}
-G**G**ATTC- -C**C**TAAG-	1.5×10^{-9}	1.3×10^{-9}
-G**T**ATTC- -C**A**TAAG-	2.9×10^{-10}	6.9×10^{-10}
-GA**C**TTC- -CT**G**AAG-	2.5×10^{-7}	7.1×10^{-7}
-GA**G**TTC- -CT**C**AAG-	8.9×10^{-8}	1.3×10^{-6}
-GA**T**TTC- -CT**A**AAG-	6.2×10^{-10}	$<1.4 \times 10^{-10}$

observed reaction energy changes. On the basis of their results, they suggested that the simple loss of a specific hydrogen bond between the protein and a base pair unaccompanied by any other changes, gives rise to a loss in energy of only 4kJ/mol to 8kJ/mol, in agreement with other studies (Fersht, 1987). Similarly, the apposition of two hydrogen bond donors or two hydrogen bond acceptors also only contribute about 4kJ/mol to 8kJ/mol. These specific contributions however, are not enough to explain the large reaction energy changes observed for the base pair substitutions of 25kJ/mol to 54kJ/mol. It was suggested that the rest of the unfavourable energy came from the loss of other non-specific protein/DNA contacts and an increase in the energy required to form the transition state complex.

Although the rate of cleavage of Eco RI* sites is much slower than at the cognate site, it is still fast enough to lead to cleavage of *E. Coli* genomic DNA under the conditions of the cell. Such DNA cleavage would result in the cell being inviable and so *in vivo* the cell must have a method of preventing double stranded cleavage. One such mechanism could involve the Eco RI methylase. It has been shown that the Eco RI methylase methylates sites other than

its cognate site under reaction conditions similar to those that enhance Eco RI endonuclease star activity (Woodbury *et al.*, 1980b). Plasmid DNA thus methylated was found to be resistant to cleavage by Eco RI endonuclease under star conditions whereas unmethylated DNA of the same plasmid is cleaved into several fragments under the same conditions (Woodbury *et al.*, 1980a). It was therefore suggested that endonuclease cleavage of host DNA at Eco RI* sites was prevented by methylation of these sites *in vivo* by Eco RI methylase. Other suggestions of what prevents hydrolysis of the genomic DNA centre on the fact that reaction at star sites would normally result in cleavage in only one DNA strand. In the cell, this nicked DNA could be repaired by DNA ligase before reassociation of the Eco RI endonuclease and cleavage of the other strand could occur, thus preventing double strand cleavage of the hosts DNA (Lesser *et al.*, 1990; Thielking *et al.*, 1990).

1.5.4 Crystal Structure of the Eco RI Endonuclease Bound to Cognate DNA

Type II restriction enzymes require magnesium ions to cleave the substrate DNA but they are not necessary for the enzyme to bind to the recognition site (Jack *et al.*, 1982). Utilising this, the endonuclease has been co-crystallised with oligodeoxynucleotide substrates in the absence of magnesium in order to determine the structure of the complexes by X-ray crystallography. A three-dimensional structure of the Eco RI endonuclease bound to an oligodeoxynucleotide substrate, d(TCGCGAATTCGCG), has been presented (McClaren *et al.*, 1986). Recently however the original connectivity between elements of secondary structure has been found to be incorrect and a revised protein chain tracing has been proposed (Kim *et al.*, 1990). Although few details were presented, some comparison of the features of the new model with those of the old model were given and these are discussed below (Figure 1.24). The architecture of the protein is essentially the same, with a five stranded parallel β-sheet surrounded by α-helices, but the connectivity through loops and helices at the surface of the complex is different. As in the old model, a parallel bundle of four α-helices fit into the major groove of the DNA and so Eco RI endonuclease has a DNA binding motif not found in any other DNA binding protein structure so far determined. Also the same as in the original model, specific contacts to DNA bases are made through this motif, although the exact residues involved in the new model were not discussed. Other specific contacts are made through an extended chain motif (Met137 to Ala142) that runs through the major groove of the DNA and was not present in the original model. It was originally proposed that no specific interactions were made with the pyrimidines. However, the revised model reveals contacts between the extended chain and pyrimidines, although again the exact groups involved were not discussed. This is consistent with biochemical evidence obtained using pyrimidine analogues (Brennan *et al.*, 1986; McLaughlin *et al.*, 1987; see section 1.5.2). In both enzyme models no protein contacts are made in the minor groove which is again consistent with base analogue experiments (Modrich & Rubin, 1977; Brennan *et al.*, 1986; McLaughlin *et al.*, 1987; see section 1.5.2).

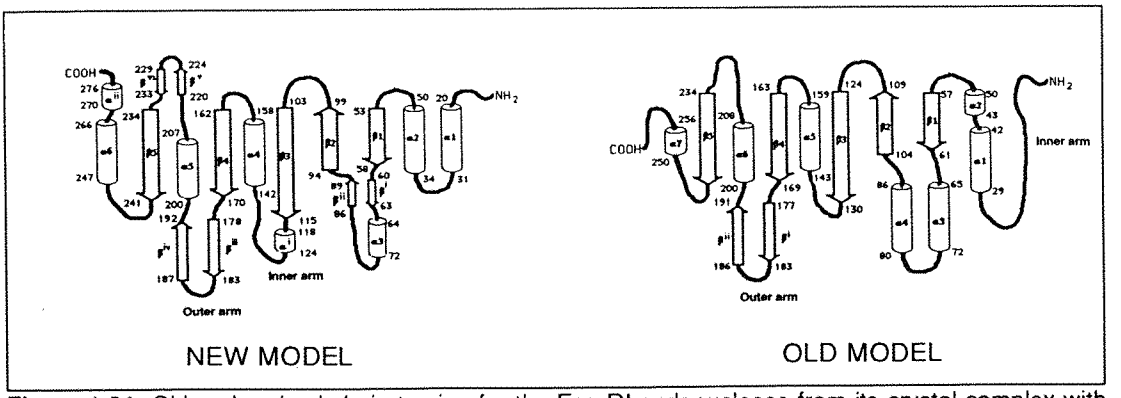


Figure 1.24: Old and revised chain tracing for the Eco RI endonuclease from its crystal complex with an oligodeoxynucleotide (Kim *et al.*, 1990).

The new assignment of specific amino acid residues helps to explain some of the results observed for mutation experiments which were hard to reconcile with the old model. The most prominent example of this involves the mutation of Glu111. In the old model it was assigned to a portion of the enzyme well removed from the protein/DNA interface. However, mutation of this amino acid to glutamine resulted in an enzyme that had a decrease in cleavage rate of approximately 10,000 fold relative to the wild type enzyme (Wright *et al.*, 1989). This is consistent with the revised model which puts Glu111 near the scissile bond of the DNA. It is interesting to note that the Glu to Gln mutation results in a lowering of the cleavage rate and not of the DNA binding. Also the mutation of aspartic acid for Glu111, which retains a negative charge on the side chain, gives rise to a mutant that has a similar specific activity to that of the wild type enzyme. It therefore seems that residue Glu111 is involved in the catalytic step and helps speed up the cleavage through its negative charge.

The original model identified hydrogen bonds between the purines of the Eco RI site and residues Glu144, Arg145 and Arg200 as the specific protein/DNA contacts (McClaren *et al.*, 1986). Conservative mutations of any one of these amino acids did indeed reduce the rate of cleavage by several orders of magnitude (Wolfes *et al.*, 1986; Needels *et al.*, 1989; Alves *et al.*, 1989). However, none of the mutants showed altered sequence specificity from GAATTC despite losing what in the old model were proposed to be specific hydrogen bond contacts. In the new model, these three amino acids are still in the vicinity of the protein/DNA interface but as yet they have not been identified as making direct contacts to the DNA bases and so the data from these mutagenesis experiments may need to be reinterpreted.

The structure of the DNA, which does not depend upon how the protein chain is traced, is the same as originally proposed (Frederick *et al.*, 1984; McClaren *et al.*, 1986). That is, it is not a straight uniform stretch of B-form DNA but contains three distinct kinks (Figure 1.25). One of them (the neo-1 kink) is in the centre of the recognition sequence and the other two (neo-2 kinks) are symmetrically placed about the dyad axis between the recognition sequence d(GAATTC) and the rest of the oligodeoxynucleotide. The kinks are termed *neo* because it appears that they are caused by the binding of the DNA to the enzyme. The crystal structure of a virtually identical oligodeoxynucleotide, d(CGCGAATTCGCG), in the absence of any

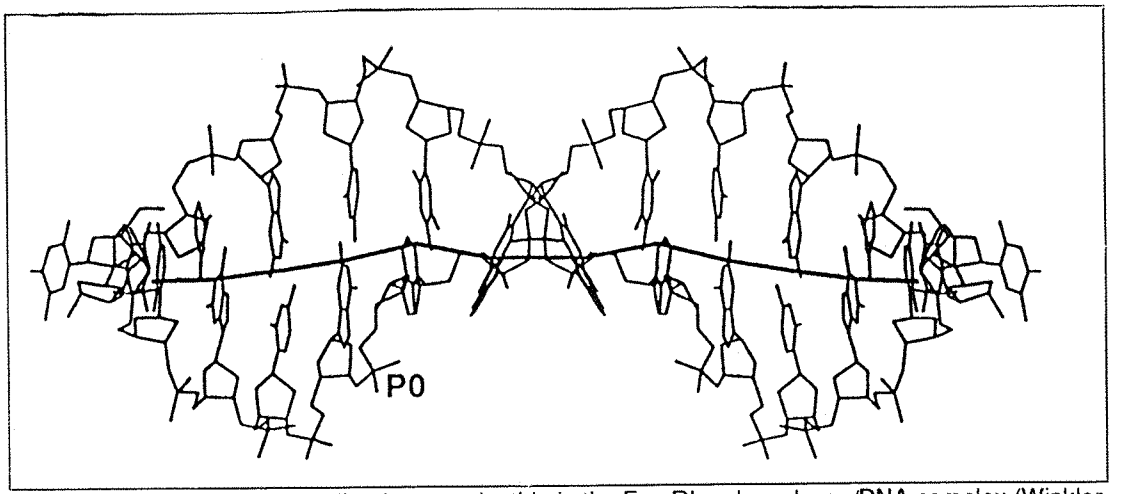


Figure 1.25: Structure of the oligodeoxynucleotide in the Eco RI endonuclease/DNA complex (Winkler, 1991). The line running through the DNA shows the helical axis and the scissile bond is indicated by PO.

protein has no such similar large structural deviations from B-DNA (Dickerson & Drew, 1981). The central neo-1 kink unwinds the DNA by approximately 25 degrees which widens the major groove by about 3.5Å greater than that seen in normal B-DNA. This serves to accommodate the four α-helix bundle that would otherwise not fit into the major groove.

1.6 ECO RV ENDONUCLEASE

Eco RV is a type II restriction/modification system and recognises the sequence d(GATATC) in duplex form (Schildkraut *et al.*, 1984). The endonuclease cleaves (|) between the two central residues (Schildkraut *et al.*, 1984; D'Arcy *et al.*, 1985) whereas the methylase transfers a methyl group (m) to the 6-amino group of the outer deoxyadenosine (Nwosu *et al.*, 1988):-

G A^{m} T|A T C
C T A|T A^{m} G

The amino acid sequences of both the endonuclease and the methylase have been determined from their gene sequences (Bougueleret *et al.*, 1984). There is no apparent homology between the Eco RI enzymes and the Eco RV enzymes contrary to what might be expected from the similarity of their recognition sequences, where only the two central bases are different (GA**TA**TC and GA**AT**TC). Also, as for other type II systems, there is no detectable homology between the Eco RV endonuclease and the Eco RV methylase despite the fact that they recognise the same sequence. Investigations into the way the two Eco RV enzymes interact with their target DNA sequence should therefore give an insight into two different approaches proteins can use to recognise the same particular DNA sequence.

The Eco RV system has several advantages over other restriction/modification systems which make it a more suitable all round choice for molecular studies. Both the methylase and the endonuclease have been cloned and overexpressed successfully (Bougueleret *et al.*, 1985; Nwosu *et al.*, 1988) and can be obtained in the relatively large quantities needed for kinetic studies and other techniques such as NMR and crystallography.

Another advantage of Eco RV endonuclease is its high solubility which means that techniques that require high concentrations of enzyme, such as NMR spectroscopy, can be used. This obviously widens the scope of investigation possible giving a more general understanding of how the enzyme works.

1.6.1 Crystal Structure of Various Eco RV Endonuclease Complexes

As mentioned above, the endonuclease can be obtained in large amounts from a genetically engineered *E.Coli* strain (Bougueleret *et al.*, 1985). A two column purification procedure enabled sufficient pure endonuclease to be prepared from this overproducer for it to be crystallised (D'Arcy *et al.*, 1985). The crystals thus produced diffracted X-rays to a resolution of about 2Å. Crystals of the endonuclease bound to three different oligodeoxynucleotides, one containing the recognition sequence and two lacking it, have also been grown (Winkler *et al.*, 1991). The structure of the endonuclease in the absence of bound oligodeoxynucleotide (the apo-enzyme) has been solved to a resolution of 2.8Å, and is shown in Figure 1.26 (Winkler, F. & Halford S.E.; personal communication), and recently this has been refined to 2.5Å (Winkler, 1991). The dimeric structure is clearly visible and a cleft can be seen between the two subunits. The three dimensional structure of the enzyme bound to a non-cognate oligodeoxynucleotide and of it bound to the cognate decadeoxynucleotide d(GGGATATCCC) have also both been solved to a resolution of 3Å. In both cases, the DNA is bound at the bottom of the cleft (Winkler, 1991). This is demonstrated in Figure 1.27 which shows the endonuclease bound to the cognate decamer, looking down the axis of the DNA with the protein in a similar orientation as in Figure 1.26. In order for the DNA to fit into the cleft, the enzyme has to open up by about 15Å, with the subunit interface acting as a kind of hinge.

There are considerable quaternary structural differences between the three types of enzyme structure (apo-enzyme, non-cognate complex and cognate complex) which can be related to the functional properties of the enzyme. Both the enzyme complexes with DNA contact the DNA bases through two short loops from each subunit, as shown in Figure 1.27. In the apo-enzyme structure these loops are poorly ordered and presumably this flexibility enables them to probe the base sequence.

In the non-cognate complex two octamers of sequence d(CGAGCTCG) are bound per enzyme molecule with the oligodeoxynucleotides stacked end to end and the junction in the middle of the complex. The DNA can therefore be considered as a bound hexadecamer with the central phosphate missing. The complex deviates considerably from two fold symmetry

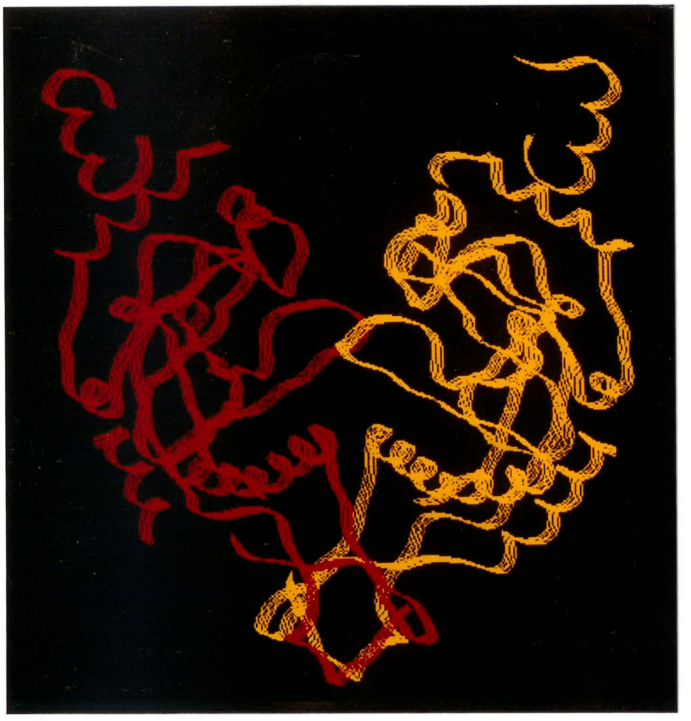


Figure 1.26: Structure of the Eco RV endonuclease apo-enzyme showing the dimeric structure and the DNA binding cleft (photograph courtesy of S.Halford, University of Bristol).

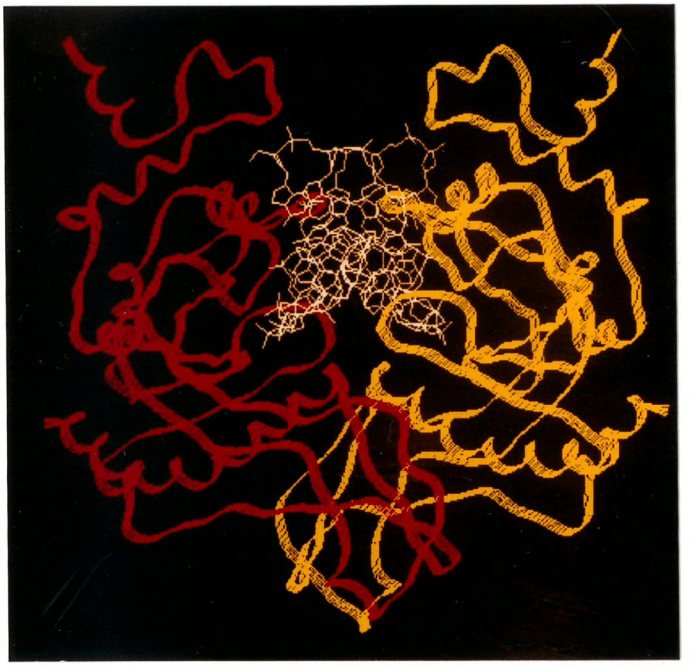


Figure 1.27: Structure of the Eco RV endonuclease bound to the decamer d(GGGATATCCC) looking down the axis of the DNA. The two loops interacting with the DNA grooves are clearly visible (photograph courtesy of S.Halford, University of Bristol).

presumably due to the lack of two fold symmetry in the bound 'hexadecamer'. Only one direct protein contact to the DNA bases is made, with most of the contacts being made to the DNA phosphates. These non-specific contacts are mostly from protein atoms that are also involved in binding phosphates in the cognate complex.

The most striking difference between the cognate complex and the non-cognate complex is that the DNA in the former deviates severely from B-form DNA whereas the non-cognate DNA is essentially B-form. The oligodeoxynucleotide in the cognate complex is very distorted along its whole length with unwinding of the helix of about 50 degrees between the four central base pairs. As in the case of the Eco RI endonuclease complex with DNA, the central two base pairs are unstacked. There is an almost 90 degree bend in the central axis of the DNA bound to the Eco RV endonuclease from end to end with a pronounced kink in the centre (Figure 1.28). This contrasts with the structure of the DNA in the Eco RI endonuclease complex which although distorted, is so only at specific points in the DNA, namely the neo-1 and neo-2 kinks (Frederick *et al.*, 1984). Another difference between the structure of the cognate DNA's in the Eco RI and Eco RV endonuclease complexes is that the major groove in the RI complex is widened whereas in the RV complex it has become more narrow and deeper.

It should be pointed out that both these structures were obtained from crystals that were grown in the absence of magnesium, the necessary catalytic co-factor, and so further changes in the DNA structure on binding magnesium ions cannot be ruled out.

Closer inspection of the endonuclease/DNA interface in the Eco RV cognate complex shows that contact to the DNA through the two loops from each enzyme sub-unit is made into both the major and minor grooves. Again, this differs from the case for the Eco RI endonuclease which contacts only the major groove with the minor groove left exposed to solvent. The loop shown at the bottom of the cleft in Figure 1.27 is centred around amino acid 70 and contacts the DNA in the minor groove. The loop at the top of the cleft is centred around amino acid 184 and contacts the major groove. Several specific contacts are made exclusively to the DNA bases within the recognition sequence through these loops and are shown schematically in Figure 1.29. It can be seen that all the potential contact points in the major groove of the G.C and the outer A.T base pairs in the recognition site are hydrogen bonded directly to amino acids of the 184 loop. The 184 loop is well ordered in the cognate complex and is stabilised by several intraloop hydrogen bonds. Conversely, not all of the potential contact points in the minor groove of these two base pairs are bonded. The crystal structure indicates that the side chain amide nitrogen of Asn70 is at a distance suitable for making hydrogen bonds with both the deoxycytidine 2C=O of the G.C base pair and the deoxyadenosine 3N of the outer A.T base pair. However, the hydrogen bonding angles for contact to the two bases indicated that a hydrogen bond would only be likely to be strong between Asn70 and the 2C=O of the deoxycytidine. The most remarkable feature of the specific contacts is that none are made to the two central base pairs. This is surprising because it results in the loss of six

Figure 1.28: Side view of the DNA bound in the Eco RV endonuclease cognate complex demonstrating the distortion of the oligodeoxynucleotide (photograph courtesy of S.Halford, University of Bristol).

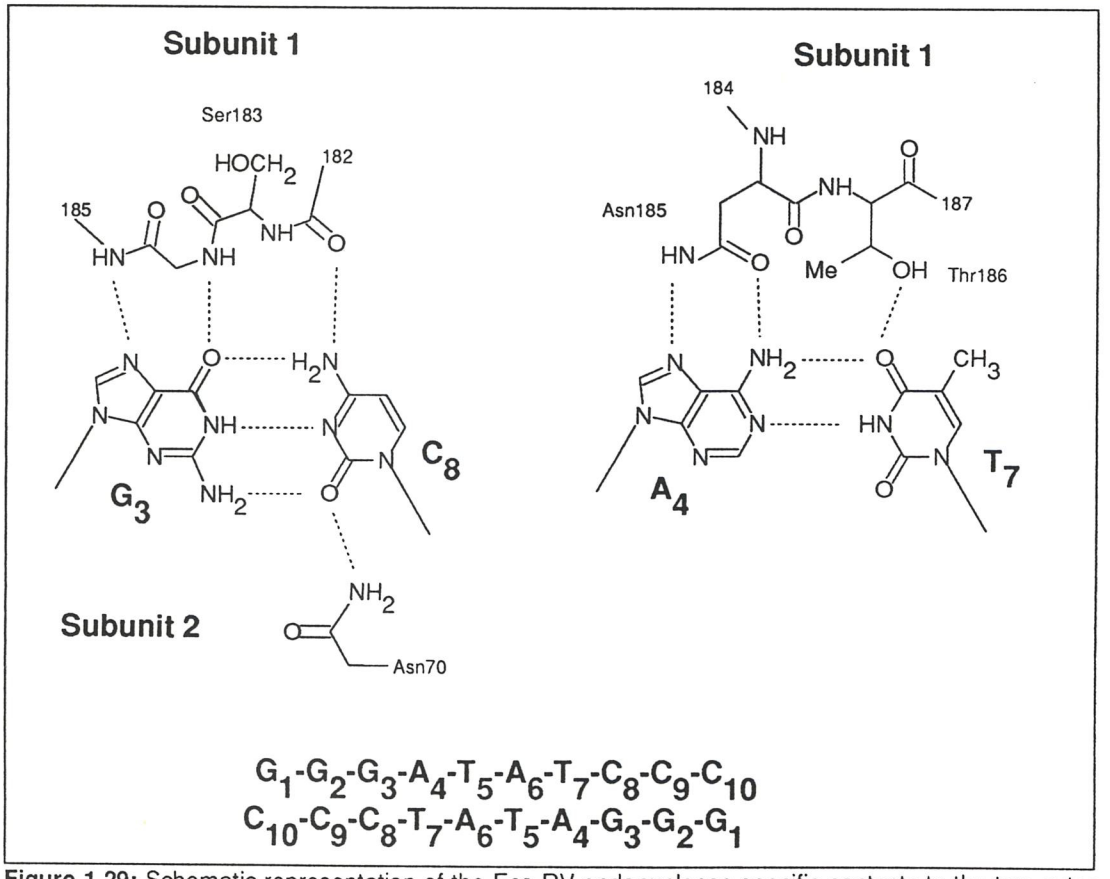


Figure 1.29: Schematic representation of the Eco RV endonuclease specific contacts to the two outer DNA base pairs in the recognition sequence d(GATATC).

potential hydrogen bonds in an apparently high energy complex. Also, alteration of either of these two bases to any other base is known to reduce cleavage by the endonuclease by more than six orders of magnitude (Taylor & Halford, 1989). The lack of specific contacts to these bases may be due to the crystal lacking magnesium which could cause a change in the structure of the complex and bring side chains into contact with the central A.T base pairs. Another alternative is that there are indeed no specific contacts to these bases in the cleavage reaction and that specificity is conferred by the indirect readout of the central d(AT) sequence in a manner discussed by Travers (1989). This would involve only the d(GATATC) sequence being able to adopt the distorted structure seen in the specific complex. The magnitude, however, of the decrease in cleavage observed when either central base pair is altered makes it seem likely that several factors are involved in the specific recognition of the central bases.

1.6.2 Site Directed Mutagenesis of the Eco RV Endonuclease

Using the three dimensional structure as a guide, some site directed mutagenesis work has been carried out to try and confirm contact points implied in the crystal structure as well as to identify amino acids not directly involved in the complex as shown by the crystal but that are necessary for catalysis, such as those that bind the magnesium cofactor (Thielking *et al.*, 1991a). Those mutants that fall into the first category involve alteration of residues Asn70, Asn185 and Thr186. Asn70 is shown in the crystal to make contact to the minor groove of the DNA through its side chain and this is supported by the fact that the replacement of Asn70 by glutamine, which has the same functional group but is larger, gives a mutant that is less active than the wild type endonuclease by a factor of 100. Similarly, mutations of the amino acids Asn185 and Thr186 which are shown to contact the DNA in the major groove, also lead to less enzyme activity. Removal of the $-CONH_2$ function (which is directly involved in specific DNA binding) of Asn185 by replacing it with alanine completely eliminates activity, as does retaining the functionality but introducing more bulk by mutating it to glutamine. Removing its hydrogen bond donor NH_2 group but keeping the =O hydrogen acceptor group by replacing it with aspartic acid gives rise to a mutant with a very low activity. Substitution of serine for Thr186, which conserves the -OH function but reduces the size of the side chain, totally inactivates the enzyme as does replacing it with asparagine.

Other amino acids that were found to be sensitive to mutation include Ser183 and Asn188. The removal of the $-NH_2$ function of Asn188 by mutation to aspartic acid leads to no activity, and the retention of the $-CONH_2$ functionality by mutating it to glutamine greatly reduces the activity. Asn188 is close to the phosphate backbone and its $-NH_2$ group is within hydrogen bonding distance of one of the phosphates. It appears however, that the introduction of extra bulk by the N188Q mutant is more important than the loss of hydrogen bonding since the abolition of the $-CONH_2$ function achieved by replacing Asn188 with alanine gives rise to a mutant whose activity is only about a tenth down on that of the wild type. The complete lack of activity in the N188D mutant is perhaps not surprising since it puts a negatively charged

side chain next to the negatively charged phosphate.

Ser183 does not make direct contact to the DNA but it does make an intrastrand hydrogen bond in the 184 loop through its side chain -OH. Abolishing the -OH function by mutating Ser183 to alanine presumably disrupts the structure of the 184 loop since it leads to an enzyme with greatly reduced activity. It also appears that the loop is disrupted by the introduction of extra bulk since the replacement of the serine by threonine gives rise to a very poor activity.

The crystal structure of the Eco RV cognate complex reveals two acidic amino acids (Asp74 and Asp90) and one basic residue (Lys92) close to the scissile phosphate. Another acidic side chain, Glu45, although not as close, is also in the vicinity (Winkler, 1991). Mutations of these amino acids do not have any significant affect on the binding of the enzyme to DNA in the absence of magnesium, but do however show that they are important for the cleavage reaction (Selent *et al.*, 1991). The replacement of either of the two acidic aspartic residues by the neutral alanine gives mutants with essentially no cleavage activity. Similarly, replacing either of them with asparagine abolishes activity. If however, either of them are replaced by glutamic acid, which retains the carboxylate function but introduces a larger side chain, little change in activity compared with the wild type endonuclease is observed. Similar effects are observed for Glu45 although the results are not as clear cut. Namely, retention of the carboxylate function by mutation to aspartic acid gives an enzyme of similar reduced activity to that of a mutant where the carboxylate function is lost by replacing Glu45 with glutamine. Substitution of Lys92 by either glutamine or glutamic acid virtually abolishes activity.

Because mutations of these residues affects the cleavage reaction but not the DNA binding and because of their vicinity to the cleaved phosphate, they have been proposed to constitute the catalytic and magnesium binding centre. The arrangement of two acidic residues and a basic residue in the sequence (Asp74....Asp90 X Lys92) is also found close to the scissile phosphate in the Eco RI endonuclease complex in a similar sequence (Asp91....Glu111 X Lys113) (Figure 1.30). The secondary structural arrangement of the two sequence motifs is also very similar. There is however no obvious counterpart of Glu45 in the Eco RI complex. Selent *et al.* propose that the Eco RI and Eco RV endonucleases follow a similar reaction mechanism with the attack of a hydroxyl ion (activated by a metal ion and one of the acidic amino acids) on the phosphate to produce a pentavalent phosphate intermediate which is stabilised by the metal ion or the basic amino acid. The phosphorous-oxygen bond is then cleaved with inversion of configuration at the phosphate as observed for both Eco RI (Connolly *et al.*, 1984a) and Eco RV (Grasby, 1991). The structure of the Eco RV cognate complex in the absence of magnesium however, shows that none of the acidic residues are correctly positioned for in-line attack of the phosphate by an activated water molecule (Winkler, 1991). Only if considerable structural rearrangement of the site occurs on magnesium binding could this situation be changed.

The active site is preformed in the Eco RV apo-enzyme and so allosteric activation of

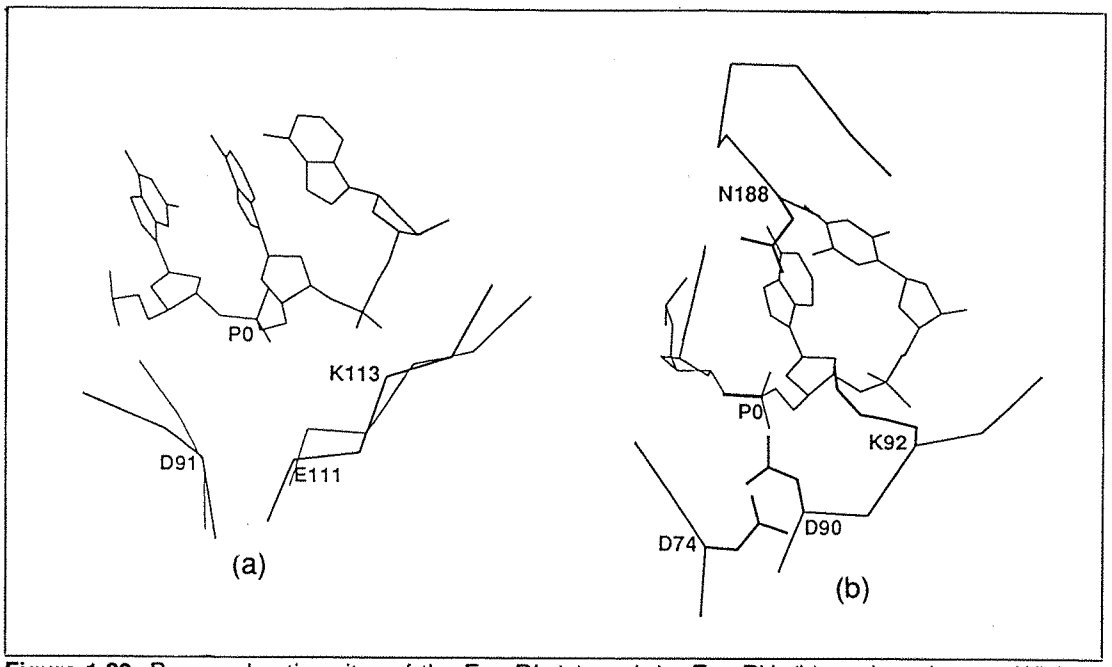


Figure 1.30: Proposed active sites of the Eco RI, (a), and the Eco RV, (b), endonucleases (Winkler, 1991). PO indicates the scissile phosphate bond.

the enzyme by assembly of the enzyme active site only occurring when a cognate DNA complex is formed seems unlikely. However, only in the cognate complex is the scissile DNA phosphate positioned in the enzyme active site, enabling cleavage to occur. This couples the recognition of the Eco RV sequence with the catalytic step. Only the Eco RV sequence can form the energetically unfavourable cognate conformation (which is stabilised by formation of the specific protein/DNA hydrogen bonds) needed to correctly position the phosphate in the catalytic site. Therefore only the Eco RV sequence can be cleaved.

1.6.3 Kinetic Mechanism of the Eco RV Endonuclease

Some detailed kinetic studies of the Eco RV endonuclease have been carried out with the plasmid substrate pAT153 (Halford & Goodall, 1988). These involved measuring the rate of cleavage of the single Eco RV site on the plasmid at various pH's and at various magnesium ion concentrations. The pH optimum was found to be around 7.5 and at this pH closed circle DNA (plasmid with no cleavage) was converted directly into linear DNA (cleaved in both strands) without any accumulation of open circle DNA (cleaved in one strand). At pH6, significant quantities of open circle DNA were produced during the reaction (Figure 1.31). Greater quantities of open circle DNA were produced than endonuclease present indicating that the reaction involved endonuclease binding at the recognition site, cleaving one of the DNA strands and then dissociating from the DNA. Similar results were obtained when the magnesium concentration was varied. Saturating magnesium ions (>2mM) gave reactions where both DNA strands were cleaved in a concerted reaction whereas low concentrations of magnesium resulted in single stranded cleavage being detected. These results are thought to be a consequence of the fact that the enzyme acts as a dimer. At low magnesium

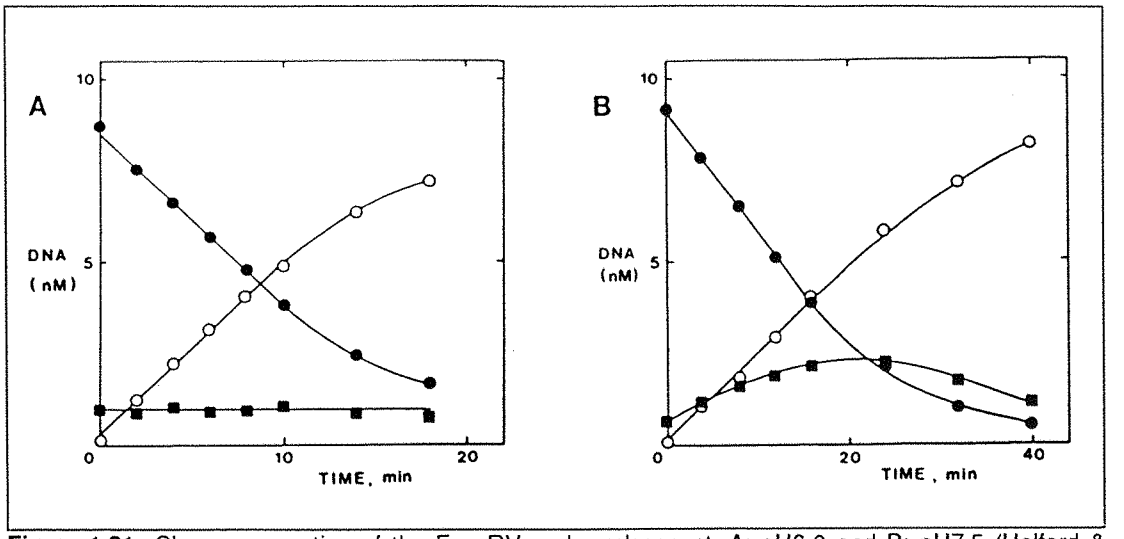


Figure 1.31: Cleavage reaction of the Eco RV endonuclease at, A: pH6.0 and B: pH7.5 (Halford & Goodall, 1988). ●=closed circle DNA, ■=open circle DNA, ○=linear DNA.

concentrations there would not be enough magnesium for all enzyme/DNA complexes formed to have a magnesium ion bound to both protein subunits and the bound enzyme would therefore not be able to cleave both strands. At low pH, higher concentrations of magnesium (>>20mM) are needed to saturate the enzyme and so the effect observed at low pH is due to a lower affinity of the enzyme/DNA complex for magnesium. This lower affinity could be due to protonation of the $-COO^-$ groups of the two aspartic acid residues proposed to be involved in magnesium binding [see section 1.6.2].

To try and determine the rate limiting step, this study also looked at transient kinetics and single turnover experiments. Experiments that looked at cleavage of the plasmid before a steady state was reached did give a transient phase that was dependent upon enzyme concentration. This indicated that the rate limiting step was after cleavage of both strands of the plasmid (Fersht, 1985) and is most probably release of product by the endonuclease. The conditions of the reaction were the same as those that produced no open circle DNA under steady state conditions and also no open circle DNA was observed during the transient phase. Single turnover experiments, in which a large excess of enzyme over substrate is used so that all the substrate is bound and only the rate of cleavage of bound DNA is observed, were also carried out. When performed at pH6, the production of nicked DNA is seen as for steady state conditions, but in this case DNA nicked in one strand is not released from the enzyme and so the effect observed is due only to the rate of cleavage of the phosphate bonds in each strand. These results gave values of $1.7min^{-1}$ for cleavage of the first strand and $0.9min^{-1}$ for cleavage of the second strand. If, in an enzyme complexed with nicked DNA, the affinity of a subunit next to the uncleaved strand for magnesium is the same as that of a subunit in an enzyme complexed with closed circle DNA, then the fact that the rate of cleavage of the second strand is virtually half that of the first strand is probably due to statistical factors. This is because the cleavage of the first strand can happen in either strand and so for reaction to occur, magnesium need only bind to one or other of the subunits. Whereas for cleavage of the second

strand to occur, magnesium must bind to the subunit next to the uncleaved strand and so has half as many sites available that will produce reaction.

1.6.4 Eco RV* Activity of the Endonuclease

It has been found that like Eco RI endonuclease, Eco RV endonuclease cleaves not only its recognition sequence d(GATATC) but also other plasmid sites that differ in only one base pair, although at vastly reduced rates (Halford *et al.*, 1986). In the presence of dimethyl sulphoxide and at high pH, the endonuclease cleaves all of the eighteen possible sequences that differ from the recognition site GATATC by one base pair except for TATATC, CATATC, GATATA and GATATG. The best non-cognate site was found to be GTTATC, but even with this site the specificity (k_{cat}/K_M) is about a thousand times lower than that for the cognate site under these conditions (Taylor & Halford, 1989). Under the more usual conditions of pH7.5 in the absence of DMSO, the discrimination between cognate and non-cognate sites is greater, with the next best site having a k_{cat}/K_M a million times less than that of the cognate site. Also the mechanism of cleavage at the non-cognate site differs from that at the normal site where hydrolysis occurs in one DNA binding event. Non-cognate sites are cleaved in only one strand per enzyme/DNA complex:-

G-A-T-A-T-C end G-A-T A-T-C
C-T-A-T-A-G -------> C-T-A T-A-G

Cognate Site Cleavage

G-T-T-A-T-C end_1 G-T-T-A-T-C end_2 G-T-T A-T-C
C-A-A-T-A-G -----> C-A-A T-A-G -----> C-A-A T-A-G

Non-Cognate Site Cleavage

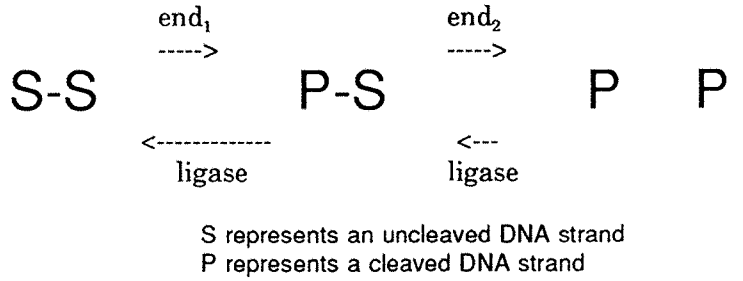
End_1 and end_2 represent two different endonucleases molecules.

In a similar manner to the cognate mechanism at low pH, the difference in mechanism is thought to be due to the enzyme/non-cognate complex having a lower affinity for magnesium ions. The enzyme/cognate complex would be expected to always contain a magnesium ion in both endonuclease subunits and therefore able to cleave both strands. With non-cognate sites, the enzyme complexes' lower affinity for magnesium would result, on average, in only one endonuclease subunit per complex containing an ion and therefore only one strand would be cleaved. This is consistent with the crystal structures of the Eco RV complex with cognate and with non-cognate DNA where the presumed metal binding site is fully assembled in the former

but not in the latter (Winkler, 1991).

Even with this large discrimination between cognate and non-cognate sites, it was calculated that the amount of cleavage at the next best site during a cell cycle would be significant enough (around a hundred strand breaks) to disrupt the cell. Three mechanisms were suggested that might be involved in making the cells viable. (i) The endonuclease is compartmentalized away from the chromosome, (ii) The non-cognate sites are protected from cleavage in some way, and (iii) The strand breaks are repaired by some kind of proof-reading mechanism. The Eco RV methylase would be the most likely candidate to provide protection for the non-cognate sites and it was indeed found to protect sequences other than d(GATATC) (Taylor *et al.*, 1990). However, these sites were found to be different from the non-cognate sites cleaved by the endonuclease and so , unlike the case for Eco RI (Woodbury *et al.*, 1980a & b), the methylase does not protect a hosts DNA from cleavage at non-cognate sites.

The protection of chromosomal DNA is thought to come from a proof-reading mechanism involving DNA ligase. DNA ligase is an enzyme which joins breaks in DNA but it repairs DNA which has been cleaved in one strand much faster than DNA cleaved in both strands (Lehman, 1974). It could therefore, if the relative activities of the two enzymes allow, repair endonuclease cleavage at non-cognate sites (on host DNA) since they would involve single strand breaks whilst being unable to repair cleavage at cognate sites (in unmethylated foreign DNA) since they involve double stranded breaks:-



S represents an uncleaved DNA strand
P represents a cleaved DNA strand

This mechanism has been shown to work *in vitro* where no non-cognate cleavage was observed when DNA ligase was present in the endonuclease reaction mixture (Taylor & Halford, 1989). It has also been demonstrated to work *in vivo* where Eco RV endonuclease and DNA ligase producing *E. Coli* strains were viable whereas those producing Eco RV endonuclease but lacking the normal ligase gene were not (Taylor *et al.*, 1990)

1.6.5 Binding of the Eco RV Endonuclease to DNA

Although the Eco RV endonuclease was found to cleave its cognate site at least a million times faster than non-cognate DNA, it was found that in the absence of magnesium it bound with equal affinity to cognate and non-cognate DNA sites (Taylor *et al.*, 1991).

Protein/DNA complexes were detected by a gel shift assay which showed that at least as many as fifteen enzyme molecules could bind to one 381bp DNA fragment with roughly equal affinity regardless of whether it contained an Eco RV site or not. Similar results were obtained for 235, 100 and 55bp DNA fragments but with the maximum number of bound protein molecules per fragment decreasing with DNA size. The trend of maximum number of bound enzyme molecules was compatible with the endonuclease occupying about 15bp, as expected from the protein/DNA crystal structure (Winkler, 1991). Analysis of the gel shift data gave a value of about 10^6M for the equilibrium association constant for the first molecule of endonuclease binding to any DNA site, cognate or non-cognate. This is in contrast to the results observed for Eco RI endonuclease which binds specific DNA preferentially to non-specific DNA by several orders of magnitude (Terry *et al.*, 1983; Lesser *et al.*, 1990). These results imply that specificity for the Eco RV reaction occurs at a stage or stages after DNA binding.

Preferential cleavage assays of pre-formed Eco RV endonuclease complexes with DNA containing a single Eco RV site showed cleavage occurring at the Eco RV site. However, the fraction cleaved was greater than that expected from the fraction of Eco RV endonucleases estimated to be occupying cognate sites and so considerable translocation without dissociation of the enzyme must have occurred. This is compatible with the linear diffusion mechanism proposed for the Eco RI endonuclease (Jack *et al.*, 1982).

The D90A mutant of Eco RV endonuclease carries a mutation of one of the putative catalytic amino acids and although it does not cleave DNA it can bind to it (Selent *et al.*, 1991). In the absence of magnesium, D90A produces essentially the same multiple binding pattern on a 377bp DNA fragment as does the wild type endonuclease (Thielking *et al.*, 1991b). In the presence of magnesium, the multiple bands formed by binding of D90A can be reduced to one specific band. Quantitatively however, the D90A mutant binds a specific 20bp oligodeoxynucleotide about a hundred times more strongly than a non-specific 20bp oligodeoxynucleotide in the absence of magnesium. In the presence of magnesium, this preference was enhanced with the specific binding being increased and the non-specific binding decreased so that the preference in the presence of magnesium is more than four thousand fold. These results suggest that some specificity in the Eco RV endonuclease reaction is brought about by preferential binding in the presence of magnesium. However, the magnitude of the preferential binding is not great enough to account for all the specificity and so further discrimination must occur during events after DNA and magnesium binding.

1.6.6 Reaction of the Eco RV Endonuclease with Modified Oligodeoxynucleotide Substrates

As mentioned before the use of base analogues has been used successfully to investigate Eco RI endonuclease binding (Berkner & Folk, 1977; Modrich & Rubin, 1977; Brennan *et al.*, 1986; McLaughlin *et al.*, 1987; Lesser *et al.*, 1990). More recently, three cases have been reported that involve the use of modified bases to probe the protein/DNA contacts of Eco RV endonuclease (Fliess *et al.*, 1988; Mazzarelli *et al.*, 1989; Newman *et al.*, 1990b). All three Eco RV cases involved incorporation of a modified base in place of the normal base within the recognition sequence GATATC of a short oligodeoxynucleotide and then testing to see if it was cleaved by the endonuclease. Analogues that were substrates then had their Michaelis-Menten parameters (K_M and k_{cat}) determined in order to give a measure of the importance of the contact point altered by the introduction of the base analogue.

One of the most important aspects that an analogue should have is that the alteration introduced should be subtle enough so that it only affects the contacting of the enzyme at that particular point and does not cause any major change in the structure of the DNA. This basically means that alterations that introduce large groups into the DNA or those that severely disrupt the Watson-Crick base pairing should be avoided. The analogues used in the three Eco RV papers are shown in Figure 1.32 indicating which potential contact point has been altered in each case. These analogues all fall into one of four categories; (i) Those where a ring nitrogen has been replaced by a carbon atom, (ii) Those where an extra potential van der Waals contacting group has been introduced (as in 5-bromodeoxyuridine), (iii) Those where an exocyclic group (either $-NH_2$, -Me or =O) has been replaced by a hydrogen atom, and (iv) Those where an exocyclic =O group has been replaced by a =S group. The final group of analogues are important since the replacement of a carbonyl group by a hydrogen atom causes rehybridisation of the ring which results in the loss of a hydrogen atom that is involved in base pairing. This, as was mentioned above, is undesirable because it could alter the overall structure of the DNA. Replacement of the carbonyl oxygen by a sulphur atom reduces the hydrogen bonding potential at that point to about half that of oxygen (Weiner *et al.*, 1984) and although it introduces slightly more bulk into the base, maintains the hybridisation of the ring and therefore hopefully maintains the base pairing of the analogue.

The kinetic results obtained for the analogue substrates are summarised in Table 1.3. There is reasonably good agreement between the different sets of results, particularly between those of Newman *et al.* and those of Fliess *et al.*. For the unmodified recognition site, the former obtain a value of 3.8μM for K_M and a value of 6.9min^{-1} for k_{cat} whereas the latter have a K_M of 4μM and a k_{cat} of 7min^{-1}. There is also good agreement between their results for substrates containing the base analogues 7-deazadeoxyadenosine and deoxyuridine. Fliess *et al.* get no cleavage for deoxyuridine in place of either thymidine or for 7-deazadeoxyadenosine in place of the outer deoxyadenosine, and Newman *et al.* get measurable but very low rates of cleavage for these analogues. With 7-deazadeoxyadenosine in place of the inner deoxyadenosine

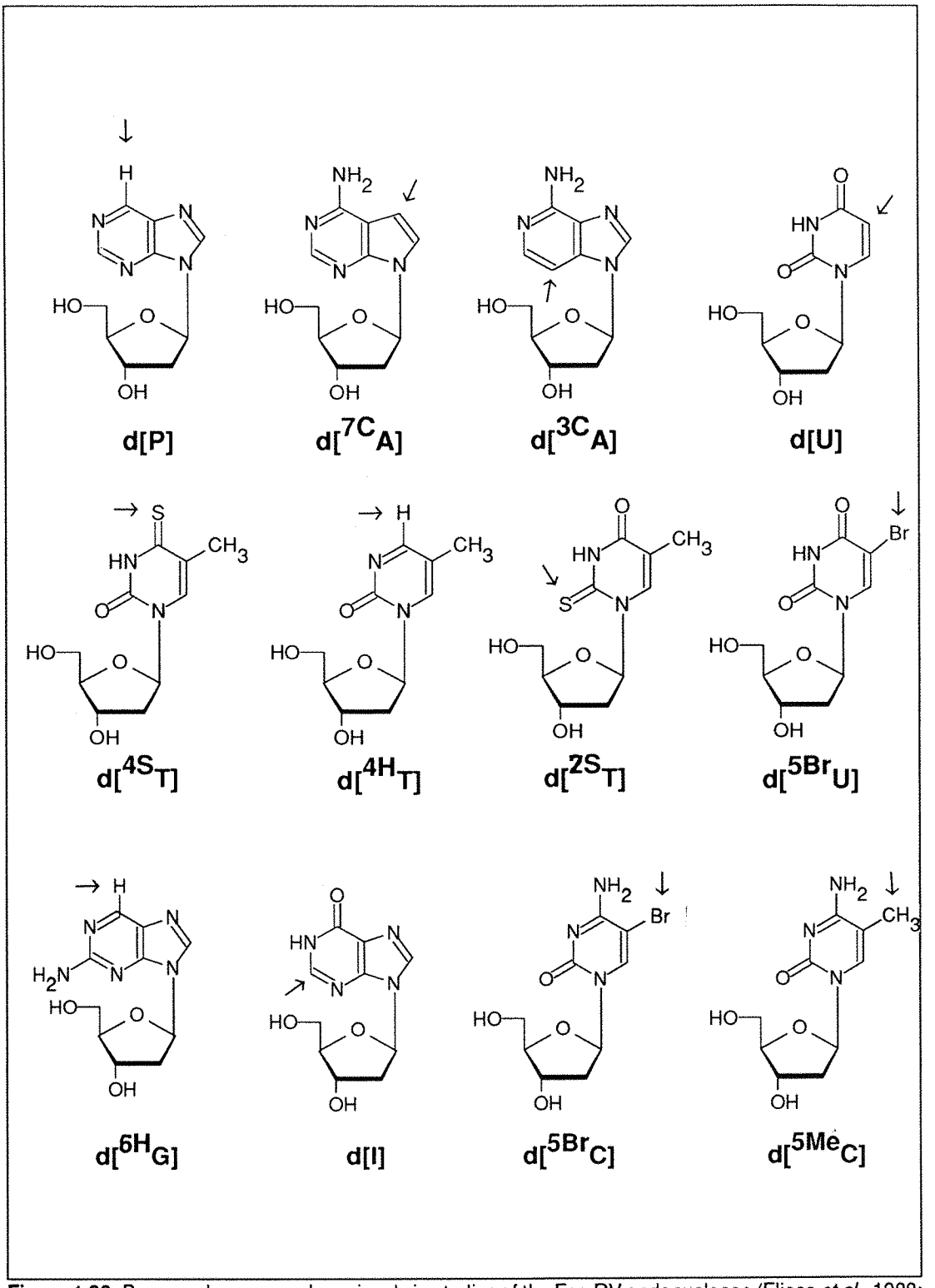


Figure 1.32: Base analogues used previously in studies of the Eco RV endonuclease (Fliess *et al.*, 1988; Mazzarelli *et al.*, 1989; Newman *et al.*, 1990b):-

d[P]	= purine deoxynucleoside	d[^{7C}A]	= 7-deazadeoxyadenosine
d[^{3C}A]	= 3-deazadeoxyadenosine	d[U]	= deoxyuridine
d[^{4S}T]	= 4-thiothymidine	d[^{4H}T]	= 5-methyl-2-pyrimidinone deoxynucleoside
d[^{2S}T]	= 2-thiothymidine	d[^{5Br}U]	= 5-bromodeoxyuridine
d[^{6H}G]	= 2-aminopurine deoxynucleoside	d[I]	= deoxyinosine
d[^{5Br}C]	= 5-bromodeoxycytidine	d[^{5Me}C]	= 5-methyldeoxycytidine

they both obtain similar values for the Michaelis-Menten parameters.

The value obtained by Mazzarelli *et al.* for the K_M of the unmodified sequence is down by about a tenth and that for the k_{cat} is about a fifth of the values for the other two sets of results. The reasons for this could be the different bases flanking the recognition site which do not have a direct effect on the specific contacts of the endonuclease but could alter the overall binding (K_M) and the rate of cleavage (k_{cat}) of the substrate. Flanking sequences are known to be able to influence the cleavage of plasmid DNA by Eco RI endonuclease (Halford, 1983). Even with the differences in the Michaelis-Menten constants for the unmodified site, the results obtained for analogue substrates containing deoxyuridine in place of either thymidine and purine deoxynucleoside in place of either deoxyadenosine agree favourably, albeit only qualitatively, with those of Newman *et al.*. In both cases all four analogues give very low rates of cleavage.

The Michaelis-Menten parameters for the plasmid pAT153 have been determined as being < 0.5nM for K_M and $0.9min^{-1}$ for k_{cat} (Halford & Goodall, 1988). The slower rate of cleavage in the plasmid could be due to the rate limiting step being product release for plasmids but not for short oligodeoxynucleotides. The release of oligodeoxynucleotide products would be faster than for plasmid DNA and so an increased k_{cat} is observed. The much smaller K_M seen in the plasmid probably reflects the fact that the facilitated diffusion mechanism, where the endonuclease bound to DNA at a non-specific site can move along the DNA until it finds a recognition site, is not available to the endonuclease when reacting with short oligodeoxynucleotides. This kind of mechanism has been proposed for both the Eco RI endonuclease (Jack *et al.*, 1982) and the Eco RV endonuclease (Taylor *et al.*, 1991; Thielking *et al.*, 1991b).

The contact points predicted from the modified base results agree well with those seen in the cognate crystal structure [discussed in section 1.6.1] for the outer G.C and A.T base pairs (Winkler, 1991). The analogues d(-G[^{3C}A]TATC-) and d(-GATA[U]C-) give changes in the specificity constant k_{cat}/K_M expected for the loss of a hydrogen bond and a van der Waals contact respectively (Fersht, 1987). The crystal structure does indeed show that the 3-nitrogen of the outer deoxyadenosine is in contact distance of an amino acid side chain although the angles in the structure are not ideal for a hydrogen bond. Also, there is protein density in van der Waals contact distance (from the side chain of Asn188) of the methyl group of the outer thymidine. Four other substrate analogues whose specificity constants could not be determined because they were either cleaved too slowly [d(-G[P]TATC-)] or because they were not cleaved at all [d(-G[^{7C}A]TATC-), d(-GATA[^{4H}T]C-) and d(-GATA[^{4S}T]C-)] obviously alter atoms that are important in the cleavage reaction. This is backed up by the crystal structure which shows that all these points are contacted by the enzyme. However, since no specificity constants were obtainable in these cases due to the cleavage reactions being too slow, it must be assumed that more than the simple deletion of one hydrogen bonding point results from the introduction of

Table 1.3: Published results of the kinetic parameters obtained for the reaction of the Eco RV endonuclease with base analogue containing oligodeoxynucleotides (sequence of the oligodeoxynucleotide given in parentheses). [a]Relative specificity constant given as a percentage of the control oligodeoxynucleotide; [b]Actual specificity constant of control oligodeoxynucleotide given in $s^{-1}M^{-1}$; [c]Rate as a percentage of control only given; t = trace of cleavage; n/d = not determined; s/s = found to be single stranded.

ANALOGUE	Newman *et al.*, 1990b. (GACGATATCGTC)			Fliess *et al.*, 1988. (AAAGATATCTT)			Mazzarelli *et al.*, 1989. (CTGATATCAG)		
	k_{cat}	K_M	k_{cat}/K_M[a]	k_{cat}	K_M	k_{cat}/K_M[a]	k_{cat}	K_M	k_{cat}/K_M[a]
GATATC	6.9	3.8	30,200[b]	7	4	29,200[b]	0.35	0.46	12,600[b]
G[P]TATC	t	n/d	n/d				t	n/d	n/d
G[^{7C}A]TATC	0	-	-	0	-	-			
G[^{3C}A]TATC	0.057	4.0	0.8%						
GA[U]ATC	0.013	3.8	0.2%	0	-	-	t	n/d	n/d
GA[^{4S}T]ATC	0.016	0.5	1.8%						
GA[^{4H}T]ATC	0	-	-						
GA[^{2S}T]ATC	1.87	1.0	108%						
GA[^{5Br}U]ATC				2	2	57%			
GAT[P]TC	0.067	8.0	0.5%				t	n/d	n/d
GAT[^{7C}A]TC	1.9	0.8	131%	2.5	2	71%			
GAT[^{3C}A]TC	0.1	1.4	4%						
GATA[U]C	0.006	6.5	0.05%	0	-	-			
GATA[^{4S}T]C	0	-	-						
GATA[^{4H}T]C	0	-	-						
GATA[^{2S}T]C	0.1	4.0	1.4%						
GATA[^{5Br}U]C				[c]37%	n/d	n/d			
[I]ATATC				0.2	1	11%	0.16	0.2	102%
[^{6H}G]ATATC							s/s	-	-
GATAT[^{5Me}C]				3	1	171%			
GATAT[^{5Br}C]				[c]123%	n/d	n/d			
GATAT[U]				0	-	-			

one of these analogues.

When the deoxyguanosine analogue deoxyinosine is introduced in the recognition site, a substrate is produced that has a specificity constant relative to the normal substrate that is reduced in one case by about a tenth (Fliess *et al.*, 1988) and not at all in another case (Mazzarelli *et al.*, 1989) compared with the normal substrate. This difference is far too small to suggest that the introduction of deoxyinosine has resulted in the loss of a protein/DNA hydrogen bond and this is again supported by the crystal structure which shows no contact to the exocyclic amino group of that deoxyguanosine.

The one result for the two outer base pairs that shows a discrepancy between the base modification studies and the three-dimensional structure is that obtained for d(-GATA[^{2S}T]C-) which shows a large decrease in the specificity constant but is not contacted in the crystal. The reason for this is probably that the analogue substrate is not B-form DNA as shown by its CD spectrum (Connolly & Newman, 1989) and the alteration in the specificity constant seen is due to the change in overall structure and not the specific alteration of the 2=O to 2=S.

The crystal structure shows no protein contacts to the inner A.T base pairs although the analogue results predict that several of the contact points of these bases are important to the cleavage reaction. This may be due to either the bases playing a role in the binding of the magnesium cofactor (which is not present in the crystal) or a conformational change in the enzyme that is induced by magnesium binding creating further enzyme contacts to these bases. Another possibility is that the base analogues in the central A.T base pairs affect the ability of the oligodeoxynucleotide to form the distorted DNA conformation seen in the cognate complex.

1.7 RAMAN SPECTROSCOPY

The Raman effect was first discovered by C.V.Raman in 1928 (Raman & Krishnan, 1928) and results from the interaction of an incident beam of light or radiation with the molecules of a sample. The scattering of the incident light caused by collision with the molecule, gives rise to radiation of a different frequency to that of the incident beam due to the loss or gain of energy of the photons acquired during the collision (Krishnan, 1971). If the molecule exists in a vibrational excited state before the light collision then the interaction of the light with the molecule can cause the molecule to drop to a lower energy state, giving up energy, resulting in the scattered light having a higher energy than the incident beam. This scattered light of higher frequency is known as anti-Stokes radiation. Alternatively, the collision of the light may cause a molecule to rise to a higher energy state resulting in the scattered radiation having a lower energy than the incident light. Scattered light of lower frequency than the incident light is known as Stokes radiation. Normally there are more molecules in an unexcited (ground) state than in an excited state and so Stokes scattering is

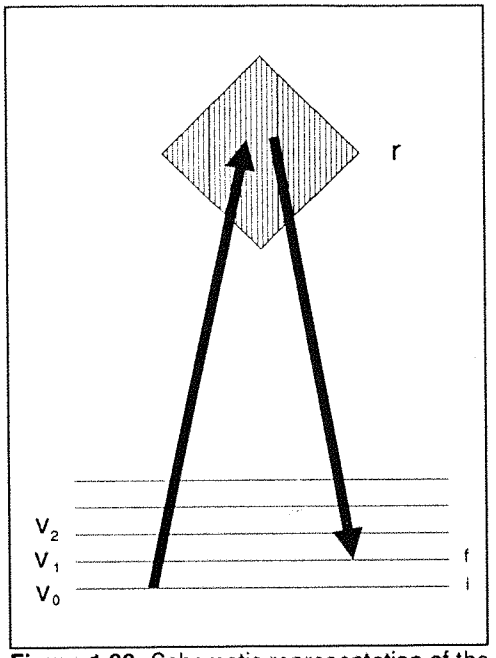


Figure 1.33: Schematic representation of the Raman effect. On collision with an incident photon, the molecule undergoes a transition from the ground state level, i, to an excited level, f, via a virtual intermediate state r.

generally more intense than anti-Stokes scattering and is hence more frequently observed in Raman spectroscopy.

Raman scattering can be visualised in terms of the quantum mechanical view of the effect (Nishimura *et al.*, 1978). Here the effect is considered as a process involving a transition from an initial energy state, i, to a final state, f, via a virtual intermediate level, r. Figure 1.33 illustrates such a scheme for a Stokes line which emerges as a photon with a decrease in energy from that of the incident light of $E_f - E_i$.

In most applications of Raman spectroscopy, the initial state, i, is a ground vibrational level (v_0) and the final state, f, is the next excited vibrational level (v_1) although overtones corresponding to the final state being two vibrational levels above the ground state (v_2) can also be observed. A Raman spectrum therefore results from transitions within the different vibrational modes of a molecule. Another spectroscopic technique, infra-red, is also due to the vibrational transitions of a molecule. However, in this case the spectrum of radiation directly absorbed by the molecule is observed. For example in Figure 1.33, an infra-red absorption band for the vibrational transition would be seen at a frequency corresponding to the energy $E_f - E_i$. Raman and infra-red spectra are not however, identical since the two phenomena have different selection rules stating which vibrational transitions are allowed. In general, vibrations that alter the *dipole moment* of a molecule (for example the asymmetric stretch of carbon dioxide shown in Figure 1.34 but not its symmetric stretch) are infra-red active. Vibrations are Raman

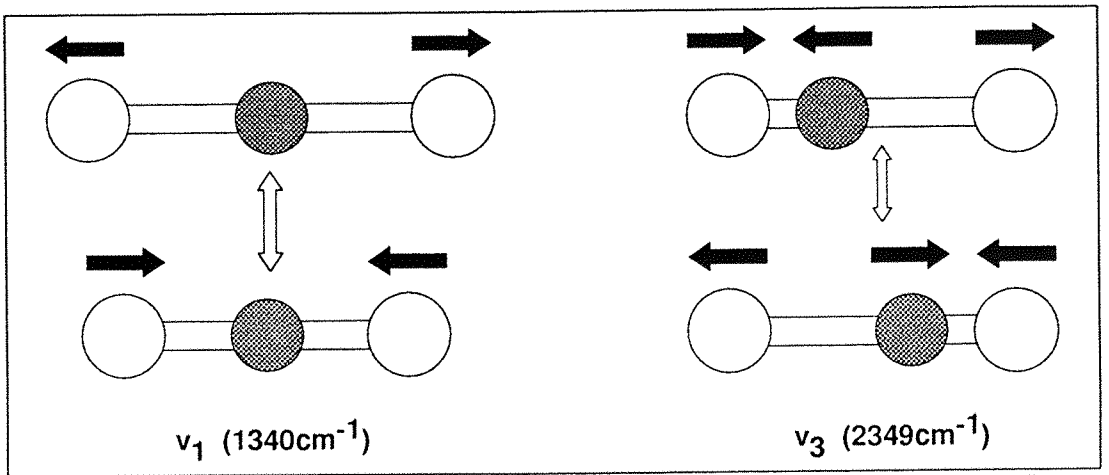


Figure 1.34: Two of the vibrations of the carbon dioxide molecule.The symmetric stretch (v_1) is Raman active whereas the antisymmetric stretch (v_3) is infrared active.

active if they alter the *polarizability* of a molecule as in the symmetric stretch of carbon dioxide (but not its asymmetric stretch). However, only for symmetric molecules are the two mutually exclusive.

1.7.1 Normal Mode Raman Studies of DNA

With a molecule as complex as a nucleic acid it is not presently possible to assign all the Raman bands to particular vibrational modes of the molecule. However, since Raman spectroscopy can be carried out on solids as well as liquids, X-ray diffraction data obtained from many crystals of nucleic acid model compounds (such as mononucleosides, mononucleotides and oligodeoxynucleotides) has considerably aided the interpretation of the Raman spectra of such compounds. This has enabled many useful 'marker' Raman bands that correspond to a particular structural conformation of a nucleic acid to be established. These markers, (reviewed in Thomas & Wang 1988 and Peticolas *et al.*, 1987) include bands that can be assigned to vibrations of various backbone groups , as given in Table 1.4, and are diagnostic of A-, B- or Z-forms of DNA.

Table 1.4: Backbone conformation markers found in DNA Raman spectra. Frequencies given in cm^{-1}.

GROUP	A-DNA	Z-DNA	B-DNA
OPO	706±5		790±3
	807±3	745±3	828±2 (G.C)
			835±2 (G.C/A.T)
			839±2 (A.T)
PO_2^-	1099±1	1095±1	1092±1
CH_2	1418±2	1425±2	1422±2

Table 1.5: Nucleoside conformational markers found in DNA Raman spectra. Frequencies are in cm^{-1}.

BASE	C3'-endo/anti	C2'-endo/anti	C3'-endo/syn
G	664±2	682±2	625±3
	1318±2	1333±3	1316±2
A	1335±2	1339±2	1310±5
C	780±2	782±2	784±2
	1252±2	1255±5	1265±2
T	745±2	748±2	770±2
	777±2	790±3	
	1239±2	1208±2	

Other bands can be attributed to particular bases and are diagnostic of different conformations of the corresponding sugar rings (Table 1.5). Using these markers, the Raman spectrum of a DNA polymer can yield a great deal of information about the structure of the DNA. For example, Raman spectral comparisons of crystalline d(CGCGCG), which has been shown by X-ray crystallography to be Z-form (Wang *et al.*, 1981), with aqueous poly-d(GC).poly-(GC) have shown the salt induced B- to Z-DNA transition of the polydeoxynucleotide (Thamann *et al.*, 1981). In high salt the polymer has essentially the same Raman spectrum as the Z-form d(CGCGCG) crystals whereas in low salt, the polymer exhibits a Raman spectrum typical of B-DNA. This salt induced B to Z transition has also been observed by circular dichroism spectroscopy (Pohl & Jovin, 1972).

Another salt induced transition of DNA has also been observed by Raman spectroscopy. Aqueous poly-dG.poly-dC has been seen to have a Raman spectrum typical of B-DNA in 0.03M sodium chloride but in 4M sodium chloride the spectrum becomes indicative of A-DNA (Nishimura *et al.*, 1986). In 1M salt, the transition was found to be temperature dependent with an A-DNA structure at 10°C and B-DNA at 60°C. They also found that deoxycytidine 5-methylation stabilised the A-form.

Since Raman spectra can be obtained from crystals as well as from solution, a comparison of DNA crystal X-ray structures can be made with its structure in aqueous solution. For example, Benevides *et al.* (1988) have found that d(CGCAAATTTGCG) exhibits a greater heterogeneity in sugar ring conformation and phosphate backbone angles in the crystal than in aqueous solution.

Raman spectroscopy has also been used to study protein/DNA interactions, despite the presence of the protein Raman lines complicating the interpretation of the DNA spectrum. Benevides *et al.* (1991) investigated the interaction of phage lambda repressor with two of its operator sites, O_L1 and O_L3. They found that the repressor induced subtle changes in the spectrum of the two operator DNAs. These changes were identified as due to:- (i) alteration to the phosphate backbone geometry, (ii) hydrophobic interaction between protein and some of the thymidine methyl groups, (iii) hydrogen bonding to some of the deoxyguanosine 7N and 6C=O groups, and (iv) alteration to some sugar conformations. One of the most important results of this study was that these spectral changes were different for the two different operator sites and could therefore reflect how the protein distinguishes between O_L1 and O_L3.

Changes in the Raman spectrum of a 176bp DNA Cro binding site have also been observed to occur upon binding of the Cro protein (Evertsz *et al.*, 1991). These include destacking of the G.C base pairs and also an increase in the A-like character of the DNA.

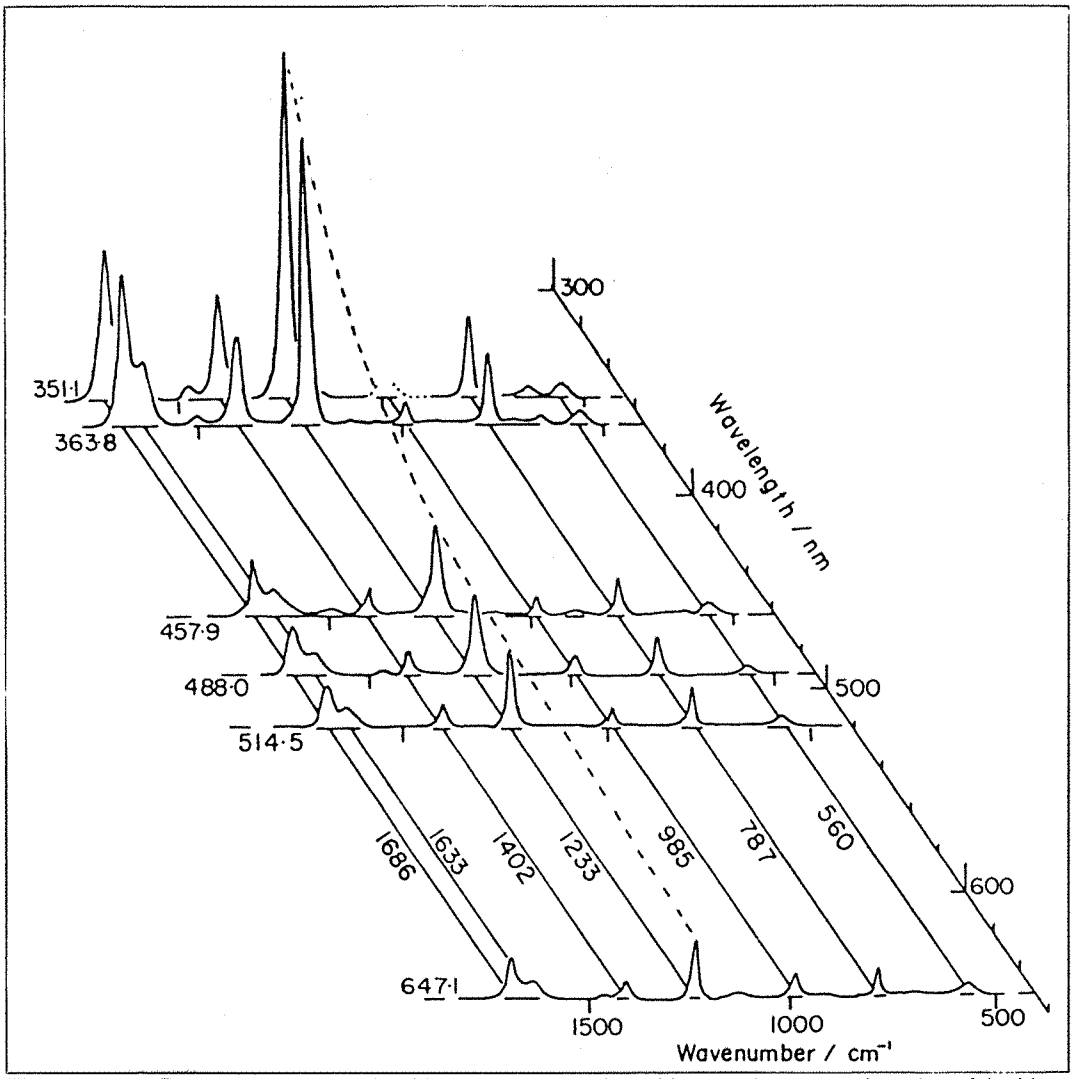


Figure 1.35: Raman spectra of β-uridine 5'-phosphoric acid at various wavelengths of incident radiation. Some Raman lines are enhanced as the absorption band at 260nm is approached (Nishimura *et al.*, 1978).

1.7.2 Resonance Raman Spectroscopy

When the incident light frequency approaches that of an absorption band of a molecule some bands in its Raman spectrum become considerably more intense. This is known as the resonance Raman effect and is illustrated by the Raman spectra of β-uridine 5'-phosphoric acid at various incident wavelengths in Figure 1.35 (Nishimura *et al.*, 1978). Uridine has an electronic absorbance band at around 260nm and as the exciting radiation approaches this band, some of the Raman lines, such as the 1233cm^{-1} line, are greatly enhanced. Others, such as the 985cm^{-1} line, experience no resonance effect. A general rule that governs whether a particular Raman band undergoes resonance enhancement has been proposed (Nishimura & Tsuboi, 1980). It states that if the electronic excited state responsible for the absorption band causes a change in the molecular geometry that mirrors a normal vibrational mode of the ground state molecule, then that vibrational mode will be resonance Raman enhanced. For this reason, resonance Raman spectroscopy can provide information about the electronic structure of molecules.

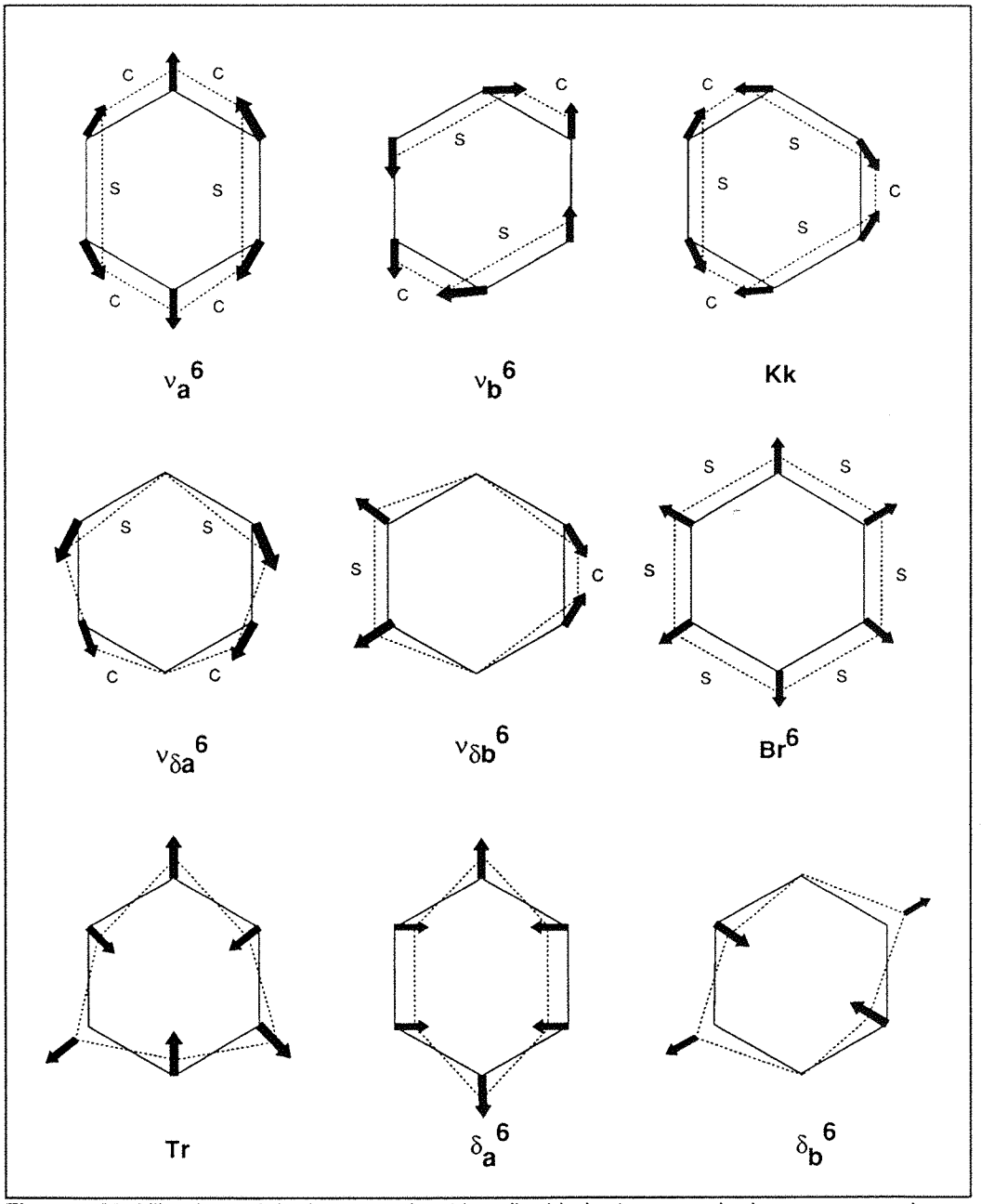


Figure 1.36: Vibrations of the benzene ring (described below). s=stretched, c=compressed.

Table 1.6: Descriptions of the ring vibrations of the benzene ring.

NAME	DESCRIPTION
ν_a^6 ν_b^6	1595cm^{-1} (e_{2g}) of benzene
$\nu_{\delta a}^6$ $\nu_{\delta b}^6$	1484cm^{-1} (e_{1u}) of benzene
Kk	Kekulé vibration, 1308cm^{-1} (b_{2u}) of benzene
Br^6	Six-membered ring breathing vibration, 992cm^{-1} (a_{1g}) of benzene
Tr	Triangle vibration, 1010cm^{-1} (b_{1u}) of benzene
δ_a^6 δ_b^6	605cm^{-1} (e_{2g}) of benzene

More information can be obtained from a Raman spectrum if the actual vibrational modes corresponding to each Raman line are known. This is extremely difficult to achieve for a molecule as complex as a nucleic acid. However, a great deal of work has been carried out in assigning the Raman bands of the uracil molecule and of some of its derivatives. In complex molecules, vibrational modes are not caused by the simple vibration of one chemical bond in isolation but usually involve contributions from several bond vibrations. Vibrations of the uracil molecule can be divided into two types:- (i) those that involve mainly vibrations of exocyclic groups, such as the carbonyl functions, and (ii) those that involve vibration and deformation of the six membered ring. The ring vibrations can be usefully characterised by the vibrations seen in the benzene ring as shown in Figure 1.36 and described in Table 1.6. Although the actual vibrations involved in uracil would deviate significantly from these idealised vibrations, they provide a useful nomenclature for describing the uracil modes of vibration.

The observed normal mode Raman bands for uracil in the range 1800-700cm^{-1} are given in Table 1.7 (Nishimura *et al.*, 1981). A few of these bands are characteristic of certain chemical groups and as such are readily assignable. These include the 1716cm^{-1} band which is typical of a carbonyl stretching vibration. For a more detailed analysis of the vibrational modes, an *ab initio* molecular orbital calculation of the uracil molecule was carried out (Nishimura *et al.*, 1981; Tsuboi *et al.*, 1987). This gave a set of calculated force constants for the molecular bonds. These in turn could be used to predict the wavenumber and the molecular vibrations involved in each mode. This enabled the Raman bands to be assigned without much difficulty. However, adjustment of the *ab initio* values by using conversion factors obtained from the simpler molecules urea, acrolein and formamide (considered to be fragments of the uracil molecule) gave better predictions of the wavenumbers. These predicted wavenumbers and a brief description of each vibrational mode are given in Table 1.7. This method also predicted with reasonable accuracy the effects of deuterium replacement of some of the exchangeable protons.

Six Raman bands of aqueous dUMP are resonance enhanced by excitation at 260nm. These lines are labelled as UrI to VI and from the vibration assignments of uracil, they can easily be assigned to particular vibrations as shown in Table 1.7. As previously stated, a general rule for the resonance enhancement of a Raman band is whether the electronic transition brings about a change in the molecular geometry that mirrors that vibrational mode. This can be demonstrated by the resonance Raman spectrum for uracil excited at 260nm, which lacks the 2C=O vibrational band. This indicates that the 260nm UV band involves an electronic transition ($\pi\pi^*$) which does not involve the 2C=O double bond. In fact, the electrons involved in that electronic transition are associated mainly with the 5C=6C-4C=O section with the 2C=O double bond not being in conjugation with that section (Nishimura & Tsuboi, 1980).

Table 1.7: Calculated and observed frequencies of the normal vibrational modes of uracil together with observed resonance Raman lines of uridine monophosphate and 4-thiouridine.

Calculated Frequency	Vibrational Mode	Observed Frequency in solid uracil	Frequency of Resonance Raman Line	
			dUMP	4-thiouridine
1720	C_2=O stretch	1716	-	-
1674	C_4=O stretch	1662	1686 **UrI**	-
1631	C_5=C_6 stretch, ν_b^6	1611	1633 **UrII**	1619
1495	N_1H bend	1507	-	-
1474	N_3H bend	1462	-	-
1410	ring stretch, ν_a^6	1422	1475 **UrIII**	1482
-	ring stretch, $\nu_{\delta a}^6$ + N_1-C stretch	-	1402 **UrIV**	1372
1386	C_5H.C_6H in phase bend	1398	-	-
1239	Kekulé, Kk	1236	1233 **UrV**	1243
1208	C_5H.C_6H out of phase bend	1217	-	-
1114	ring stretch, $\nu_{\delta b}^6$	1104	-	-
1047	ring stretch, $\nu_{\delta a}^6$	1010	-	-
984	ring deformation, Tr	988	-	-
792	ring breath, Br	792	787 **UrVI**	(1160)
-	C_4=S stretch coupled with ring breath at 1160cm^{-1}	-	-	(709)

4-Thiouridine is a derivative of uridine in which the 4C=O is replaced by 4C=S and it occurs naturally in some tRNAs (Figure 1.37). It has a strong UV absorption band at around 330nm (identical to that of 4-thiothymidine shown in Figure 1.38) and exhibits a resonance Raman spectrum, when excited by 363.8nm argon ion laser radiation, similar to that of uridine (Nishimura *et al.*, 1978). Some of the Raman bands (Table 1.7) can be assigned to the four uridine bands UrII, III, IV and V (since 4C=O has been replaced by 4C=S, UrI is obviously absent). The 709cm^{-1} band is caused by the C=S stretching vibration and is considered to be strongly coupled to the ring breathing vibration at 1160cm^{-1}. Again there appears to be no resonance enhancement of the 2C=O vibration for this electronic transition.

The relatively high wavelength absorbance band of 4-thiouridine means that it can be resonance Raman excited at around 330nm without inducing a resonance Raman effect in any other biochemical molecule, such as other nucleotides or protein amino acids. This enables the 4-thiouridine residue to be used as a resonance Raman probe of biochemical mechanisms. It has been used as such in studies of tRNAs where resonance Raman spectra of 4-thiouridine present in some tRNAs could be recorded (Nishimura *et al.*, 1976). Some of these tRNAs gave the same spectrum as free 4-thiouridine. Others gave a completely different spectrum and it

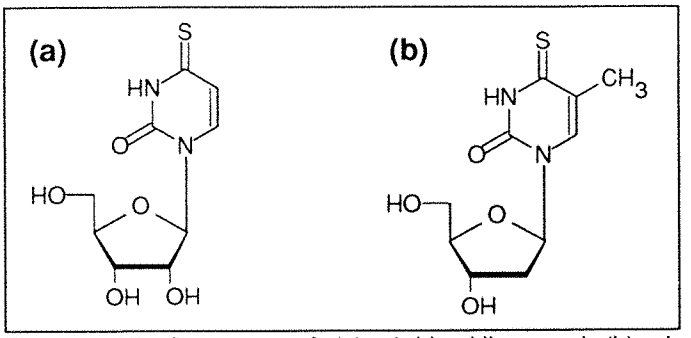


Figure 1.37: Structures of (a), 4-thiouridine, and (b), 4-thiothymidine.

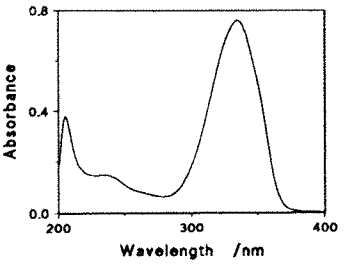


Figure 1.38: UV spectrum of 4-thiothymidine at pH7.2.

was proposed that this was due to modification of the 4-thiouridine residue in the tRNA by crosslinking to another nucleotide base.

4-Thiothymidine has a very similar structure (Figure 1.37) and UV spectrum (Figure 1.38) to that of 4-thiouridine and is therefore expected to have a similar resonance Raman spectrum. In this project, it was hoped that with detailed assignments of its resonance Raman bands, 4-thiothymidine could be used as a resonance Raman probe for studying protein/DNA interactions.

CHAPTER 2

MATERIALS AND METHODS

2.1 SYNTHESIS OF BASE ANALOGUES

2.1.1 Materials

All solvents were obtained from BDH (Poole, Dorset) except for anhydrous *N,N*-dimethylformamide which was obtained from Aldrich (Gillingham, Dorset) and petroleum spirit (40-60) which was from Interchem (Ipswich, Suffolk). Deoxyguanosine was obtained from Genofit (Geneva, Switzerland). All other organic reagents were from Aldrich (Gillingham, Dorset).

Silica gel for flash columns was from BDH (Poole, Dorset) and reverse phase LiChroprep RP-18 was from Aldrich (Gillingham, Dorset).

Pyridinium form Dowex resin was made by washing 100g of Dowex 50W-X8(Na) (BDH, Poole, Dorset) with the following:-

(i) 1l of water
(ii) 1l of 2M hydrochloric acid
(iii) 1l of water
(iv) 1l of 1:1 pyridine in water
(v) 1l of water

2.1.2 Drying of Solvents

Pyridine free from primary and secondary amines was obtained by refluxing with ninhydrin for 5 hours followed by distillation. This was then dried by refluxing with potassium hydroxide for 2.5 hours and then distilling (b.pt. 115°C).

Dry tetrahydrofuran (**THF**) was obtained when needed by refluxing with sodium metal until dry (benzophenone as indicator) followed by distillation (b.pt. 67°C).

Dry **methanol** was obtained when needed by refluxing with magnesium turnings and iodine for 30 minutes followed by distillation (b.pt. 65°C).

Dry **triethylamine** (b.pt. 89°C) and ***N,N*-diisopropylethylamine** (b.pt. 127°C) were obtained by refluxing with calcium hydride for 6 hours followed by distillation.

2.1.3 Experimental and Analytical

Chemical reactions were routinely monitored by thin layer chromatography (TLC) using aluminium backed silica gel plates coated with a fluorescent indicator (silica gel 60 F_{254}, Merck 5549). Plates were developed with one of the following solvent systems and then viewed with 254nm UV light:-

(percentages are v/v)

A: 3% methanol in dichloromethane
B: 4% methanol in ethylacetate
C: 10% methanol in dichloromethane
D: 10% triethylamine + 45% ethylacetate + 45% dichloromethane

Reverse phase TLC was carried out using RP-18 F_{254} glass backed plates (Merck 15423).

Flash chromatography (Still *et al.*, 1978) was performed using Merck silica gel 60 (40 to 60μm particle size) and a positive nitrogen pressure of 0.5 atmospheres. One of three column sizes was used depending on the scale of the reaction; 20 x 3cm, 25 x 4cm or 35 x 5cm. Reverse phase flash chromatography was performed using LiChroprep RP-18. Fractions collected from flash columns were monitored by TLC.

Rotary evaporation of solvents was carried out using a Buchi rotary evaporator under water pump vacuum or, where stated, under oil pump vacuum.

Modified deoxynucleosides were characterised by proton nuclear magnetic resonance (^{1}H NMR) spectrometry, fast atom bombardment mass spectrometry (FAB-MS) and in some cases, UV spectrometry. ^{1}H NMR spectroscopy was performed on a Brucker AM360 360MHz spectrometer, with the sample dissolved in the solvent stated, and using tetramethylsilane as internal standard or, for D_2O solutions, 3-(trimethylsilyl)propionic-2,2,3,3,-*d*4 acid sodium salt. Mass spectra were recorded using a V.G. TS250 spectrometer with the sample analyzed in a matrix of glycerol or nitrobenzyl alcohol. The samples were bombarded with a beam of neutral Xenon atoms at 9-10kV. UV spectra were recorded on a Uvikon 930 UV/visible spectrometer (Kontron Instruments). Fluorescence spectra were recorded using a Hitachi F-2000 fluorescence spectrophotometer with a xenon lamp excitation source.

2.1.4 Synthesis of the 2-Aminopurine Analogue, d[^{6H}G]

The synthesis of the 2-aminopurine analogue was based on the method of McLaughlin *et al* (1988) and uses 2'-deoxyguanosine (I) as the starting material.

2.1.4.i Synthesis of 3',5'-*O*, 2-*N*-tribenzoyldeoxyguanosine (II)

2.850g (10mmol) of 2'-deoxyguanosine monohydrate (I) was dried by coevaporation in a rotary evaporator under oil pump vacuum two times with 25ml dried pyridine. The solid was then suspended in 35ml of dried pyridine. 0.37g (3mmol) of dimethylaminopyridine and 9.05g (40mmol) of benzoic anhydride were added and the mixture gently refluxed overnight under anhydrous conditions. After about 18 hours, TLC (system C) showed that all the starting

material had been converted to product. Excess benzoic anhydride was quenched after cooling by the careful addition of 10ml of saturated aqueous sodium bicarbonate solution. After removing the solvents on a rotary evaporator the off-white solid obtained was worked up by dissolving in 200ml of dichloromethane and washing twice with an equal volume of 5% (w/v) aqueous sodium bicarbonate. The organic layer was dried over anhydrous sodium sulphate and then evaporated to a yellow solid. The crude product was purified by flash chromatography on a silica gel column eluting with 0-3% methanol in dichloromethane. Evaporation of the solvent from all fractions containing product yielded 4.92g (85%) of a white solid that was pure by TLC.

^{1}H NMR data ($CDCl_3$ solvent): δ2.66-2.74ppm (1H, m, H2'); δ3.24-3.33ppm (1H, m, H2''); δ4.65-4.72ppm (2H, m, H4'+ H5'); δ5.03-5.10ppm (1H, m, H5''); δ5.95-5.99ppm (1H, m, H3'); δ6.28-6.35ppm (1H, t, J=6.79Hz, H1'); [δ7.26ppm (s, $CHCl_3$)]; δ7.3-8.2ppm (16H, m, ArH from benzoyls + H8); δ9.70ppm (1H, s, 2-NHBz); δ12.20ppm (1H, s, N1H).

FAB-MS data: 580 ($M + H^+$, 30%), 256 (M - sugar + H^+, 100%).

2.1.4.ii Synthesis of 3',5'-*O*, 2-*N*-tribenzoyl-6-(2,4,6-triisopropyl)benzenesulphonyl-deoxyguanosine (III)

4.05g (7mmol) of dry 3',5'-*O*, 2-*N*-tribenzoyldeoxyguanosine (II) was dissolved in 100ml of anhydrous grade dichloromethane together with 4.75ml (28mmol) of dry diisopropylethylamine and 0.61g (5mmol) of dimethylaminopyridine. 4.24g (14mmol) of 2,4,6-triisopropylbenzenesulphonyl chloride was added and the solution stirred with the exclusion of moisture at room temperature. After 2 hours, TLC (system A) showed that all the starting material had been converted cleanly to product. The reaction mixture was transferred to a 1l flask and evaporated to an orange solid on a rotary evaporator. The large flask is needed since severe frothing of the product occurs on evaporation. The crude product was purified by flash chromatography eluting with 0-1.5% methanol in dichloromethane. 4.00g (68%) of product was obtained which was pure by TLC.

^{1}H NMR data (in $CDCl_3$): δ1.23-1.31ppm (18H, m, CH_3 on the benzenesulphonyl group); δ2.86-2.96ppm (2H, m, H2' + 4-CH$(CH_3)_2$ on the benzenesulphonyl group); δ3.35-3.27ppm (1H, m, H2''); δ4.25-4.33ppm (2H, m, 2- & 6-CH$(CH_3)_2$ on the benzenesulphonyl group); δ4.66-4.70ppm (1H, m, H4'); δ4.72-4.77ppm (1H, m, H5'); δ4.83-4.87ppm (1H, m, H5''); δ5.94-5.97ppm (1H, m, H3'); δ6.55-6.59ppm (1H, m, H1'); δ7.23ppm (2H, s, 2 ArH from the benzenesulphonyl group); [δ7.27ppm ($CHCl_3$)]; δ7.35-8.15ppm (16H, m, ArH from benzoyls + H8); δ8.65ppm (1H, s, 2-NHBz).

FAB-MS data: 846 (M + H^+, 15%), 579 (M - the benzenesulphonyl group + H^+, 18%), 521 (M - sugar + H^+, 30%), 255 (M - the benzenesulphonyl group - sugar + H^+, 100%)

2.1.4.iii Synthesis of 3',5'-*O*, 2-*N*-tribenzoyl-6-hydrazinodeoxyguanosine (IV)

3.14g (3.7mmol) of dry 3',5'-*O*, 2-*N*-tribenzoyl-6-(2,4,6-triisopropyl)benzenesulphonyl-deoxyguanosine (III) was dissolved in 70ml of dried THF. The reaction vessel was sealed with a rubber septum and cooled to 0°C in an ice bath. 0.25ml (12mmol) of anhydrous hydrazine was added gradually over 10 minutes through the septum. The mixture was then allowed to reach room temperature. After 2 hours all the starting material had reacted (TLC system A) and so the reaction was stopped by drying on a rotary evaporator to a solid, which was then taken up in 150ml ethylacetate. The solution was washed twice with an equal volume of 5% aqueous sodium bicarbonate and once with an equal volume of saturated aqueous sodium chloride. The yellow organic solution was dried over anhydrous sodium sulphate and then reduced to a yellow solid on a rotary evaporator. 3.10g of crude product was obtained in this way and was used without further purification for the next synthesis step. A small amount of the crude product (~100mg) was purified by recrystallisation from methanol for NMR and mass spectroscopic analyses.

^{1}H NMR data (in $CDCl_3$): δ2.77-2.83ppm (1H, m, H2'); δ3.25-3.33ppm (1H, m, H2''); δ4.63-4.66ppm (1H, m, H4'); δ4.71-4.75ppm (1H, m, H5'); δ4.86-4.91ppm (1H, m, H5''); δ5.89-5.92ppm (1H, m, H3'); δ6.48-6.51ppm (1H, m, H1'); [δ7.27ppm (s, $CHCl_3$)]; δ7.3-8.1ppm (16H, m, ArH from benzoyls + H8)

FAB-MS data: 594 (M + H^+, 32%) 270 (M - sugar + H + H^+, 100%)

2.1.4.iv Synthesis of 3',5'-*O*-dibenzoyl-2-(*N*-benzoylamino)purine deoxynucleoside (V)

2.97g of crude 3',5'-*O*, 2-*N*-tribenzoyl-6-hydrazinodeoxyguanosine (IV) was dissolved in 150ml of 5% water in THF. 2g (8.6mmol) of silver I oxide was added and the mixture heated under reflux. After 2.5 hours reaction appeared complete by TLC (system B) and so the mixture was allowed to cool. Solvents were removed on a rotary evaporator and the residue taken up in 100ml of ethyl acetate. The majority of solids were removed by filtration through celite before washing twice with an equal volume of saturated aqueous potassium iodide, twice with 10% aqueous sodium thiosulphate and twice with water. After drying over anhydrous sodium sulphate, the organic layer was reduced to a black solid using a rotary evaporator. The crude product was purified by flash chromatography eluting with 0-2.5% methanol in dichloromethane to give 1.37g (69% for steps 2.1.4.iii and 2.1.4.iv) of an off-white solid that was pure by TLC.

^{1}H NMR data (in $CDCl_3$): δ2.84-2.90ppm (1H, m, H2'); δ3.35-3.43ppm (1H, m, H2"); δ4.72-4.80ppm (2H, m, H4' + H5'); δ4.90-4.95ppm (1H, m, H5"); δ5.97-6.00ppm (1H, m, H3'); δ6.55-6.59ppm (1H, m, H1'); [δ7.27ppm (s, $CHCl_3$)]; δ7.3-8.5ppm (16H, m, ArH from benzoyls + H8); δ8.95ppm (1H, s, 2-NHBz); δ9.05ppm (1H, s, H6).

FAB-MS data: 564 (M + H^+, 16%), 240 (M - sugar + H + H^+, 100%)

2.1.4.v Synthesis of 2-(*N*-benzoylamino)purine deoxynucleoside (VI)

1.00g (1.78mmol) of 3',5'-*O*-dibenzoyl-2-(*N*-benzoylamino)purine deoxynucleoside (V) was dissolved in 20ml of pyridine and cooled to -15°C. 2.7ml of chilled methanol was added followed by 2.5ml of 2M aqueous sodium hydroxide (added slowly over 5 minutes). Reaction was complete as shown by TLC (system B) after 40 minutes and was stopped by the addition of pyridinium form Dowex, until the solution was neutral. The solution was allowed to reach room temperature and the Dowex was filtered off and then washed with a 1:1 mixture of pyridine and water. The filtrate was reduced on a rotary evaporator to a browny-white solid which was rendered anhydrous by coevaporation three times with dried pyridine. The crude product was used directly for the next stage without purification or characterisation.

2.1.4.vi Synthesis of 5'-*O*-dimethoxytrityl-2-(*N*-benzoylamino)purine deoxynucleoside (VII)

The crude product (VI) from the above step was suspended in 20ml of dried pyridine. 0.90g (2.66mmol) of dimethoxytrityl chloride was added and the solution stirred overnight at room temperature, with the exclusion of moisture. After 20 hours the reaction was complete (as shown by TLC, system C) and was stopped by the addition of 6ml of methanol. Solvents were removed on a rotary evaporator to give an orange oil which was redissolved in 50ml dichloromethane. This was worked up by washing twice with 5% aqueous sodium bicarbonate and then drying the organic layer over anhydrous sodium sulphate. Solvents were evaporated off to give a yellow oil which was purified by flash chromatography eluting with 2% methanol in dichloromethane containing 0.5% triethylamine. The purified product was evaporated to give 0.64g (55% for steps 2.1.4.v and 2.1.4.vi) of a solid that was pure by TLC.

^{1}H NMR data (in $CDCl_3$): δ2.60-2.70ppm (1H, m, H2'); δ2.95-3.00ppm (1H, m, H2"); δ3.36-3.40ppm (1H, m, H5'); δ3.47-3.51ppm (1H, m, H5"); δ3.74-3.80ppm (6H, s, -OCH_3 from DMT); δ4.20-4.25ppm (1H, m, H3'); δ6.58ppm (1H, t(J=6.46Hz), H1'); δ6.7-7.9ppm (18H, m, ArH from benzoyl and DMT groups [also includes $CHCl_3$ s at 7.27ppm]); δ8.2ppm (1H, s, H8); δ8.65ppm (1H, s, 2-NHBz); δ9.0ppm (1H, s, H6).

FAB-MS data: 658 (M + H^+, 1%), 354 (M - DMT + H^+, 2%), 303 (DMT^+, 100%), 240 (M - sugar + H + H^+, 26%)

2.1.4.vii Synthesis of 3'-*O*-[(*N*,*N*-diisopropylamino)-2-cyanoethylphosphoramidite]-5'-*O*-dimethoxytrityl-2-(*N*-benzoylamino)purine deoxynucleoside (VIII)

0.68ml (3.75mmol) of diisopropylethylamine was added through a rubber septum to 0.50g (0.76mmol) of dried 5'-*O*-dimethoxytrityl-2-(*N*-benzoylamino)purine deoxynucleoside (VII). 10ml of dried THF was added through the septum and the solution cooled to 0°C. 0.25ml (1.1mmol) of 2-cyanoethyl, *N*,*N*-diisopropylamino-chlorophosphoramidite was added carefully to the cooled stirring solution through the septum. After 2 hours, TLC (system D) showed the reaction to be complete and so excess reagent was quenched by the addition of 3ml of methanol. The solution was reduced on a rotary evaporator to an orange tar which was redissolved in 50ml of 10% triethylamine in ethylacetate. This was washed twice with an equal volume of 5% aqueous sodium bicarbonate and twice with saturated aqueous sodium chloride. The organic layer was dried over anhydrous sodium sulphate, reduced to a thick dark yellow oil on a rotary evaporator and then redissolved up in a small amount of 10% triethylamine in a 1:1 mix of ethylacetate and dichloromethane. This was purified by flash chromatography, eluting with the same solvent system, using a silica gel column that had been pre-equilibrated with the solvent system (to reduce decomposition of the product on the silica). Pooled fractions containing product were reduced to give a sticky yellow solid. In order to facilitate handling of the phosphoramidite, it is preferable to obtain it as a free flowing solid. This was achieved by dissolving the sticky solid in a minimum volume of toluene and then precipitating out the phosphoramidite by adding the solution to about 400ml of vigorously stirring petroleum spirit (40-60). After about 2 minutes stirring, the solid was allowed to settle and most of the solvent decanted off. The solid was separated from the remainder of the solvent by filtration and dried overnight in a desiccator under oil pump vacuum to give 0.33g (51%) of white solid that was pure by TLC.

2.1.4.viii Synthesis of deblocked 2-aminopurine deoxynucleoside (IX)

Some deblocked 2-aminopurine deoxynucleoside was prepared for characterisation and for use as a standard for base composition analysis. 200mg (0.35mmol) of 3',5'-*O*-dibenzoyl-2-(*N*-benzoylamino)purine deoxynucleoside (V) was dissolved in 2ml of a 1:1 mixture of concentrated aqueous ammonia and ethanol and sealed in a bijou bottle. The solution was heated at 55°C in a water bath overnight. After cooling on ice the bottle was opened and the solution dried down to an off white solid using a vacuum centrifuge. The solid was suspended in 0.5ml of water, washed twice with diethyl ether and then the water evaporated off to give 82mg (93%) of product which was not quite pure by TLC (20% methanol in ethylacetate). 50mg of this was purified for characterisation by recrystallising from methanol to give 32mg (65%) of pure product.

^{1}H NMR data (in 1:1 D_2O and CD_3OD): δ2.4-2.5ppm (1H, m, H2'); δ2.75-2.87 (1H, m, H2''); [δ3.25-3.35ppm (m, CHD_2OD)]; δ3.70-3.85ppm (2H, m, H5' + H5''); δ4.08ppm (1H, q(J=3.5Hz), H4'); δ4.57-4.64ppm (1H, m, H3'); [δ4.75ppm (s, HOD); δ5.0ppm (s, CD_3OH)]; δ6.35-6.45ppm (1H, m, H1'); δ8.27ppm (1H, s, H8); δ8.61ppm (1H, s, H6).

FAB-MS data: 252 (M + H^+, 3%), 251 (M^+, 5%), 136 (M - sugar + H + H^+, 53%)

UV data: pH7 λ_{max} at 298nm

2.1.5 Synthesis of the 6-Thiodeoxyguanosine Analogue d[^{6S}G]

The synthesis of the thiodeoxyguanosine analogue uses 2'-deoxyguanosine nucleoside monohydrate as the starting material and goes through the same first two steps as in the 2-aminopurine analogue synthesis [2.1.4.i and 2.1.4.ii]. After that the syntheses diverge and the thio-function is introduced by a novel method involving substitution by sulphide ions.

2.1.5.i Synthesis of 3',5'-*O*, 2-*N*-tribenzoyl-6-thiodeoxyguanosine (X)

3.96g (4.68mmol) of 3',5'-*O*, 2-*N*-tribenzoyl-6-(2,4,6-triisopropyl)benzenesulphonyl-deoxyguanosine (III) was dissolved in 60ml of dried THF. 0.25g (5.38mmol) of dilithium sulphide was added and the solution stirred at room temperature. After 2.5 hours, TLC (system A) showed that all the starting material had reacted. The solvent was removed on a rotary evaporator leaving a yellow solid. This was redissolved in 200ml dichloromethane and washed twice with 5% aqueous sodium bicarbonate and once with saturated sodium chloride. The organic layer was dried over anhydrous sodium sulphate and then dried down on a rotary evaporator to a yellow solid. The crude product was purified on a silica gel column, eluting with 1% methanol in trichloromethane, to give 2.45g (88%) of yellow solid that was pure by TLC.

^{1}H NMR data (in $CDCl_3$): δ2.68-2.78ppm (1H, m, H2'); δ3.25-3.34ppm (1H, m, H2''); δ4.51-4.83ppm (2H, m, H4' + H5'); δ5.03-5.09ppm (1H, m, H5''); δ6.00-6.05ppm (1H, m, H3'); δ6.31-6.36ppm (1H, t(J=6.7Hz), H1'); [δ7.27ppm (s, $CHCl_3$)]; δ7.34-8.23ppm (16H, m, ArH from benzoyls + H8); δ9.6ppm (1H, s, 2-NHBz); δ13.4ppm (1H, s, N1H).

FAB-MS data: 596 (M + H^+, 27%), 272 (M - sugar + H^+, 100%)

UV data (CH_2Cl_2): λ_{max} at 333nm

2.1.5.ii Synthesis of 2-*N*-benzoyl-6-thiodeoxyguanosine (XI)

The method used here is essentially the same as step 2.1.4.v in the synthesis of the 2-aminopurine analogue. 2.38g (4.00mmol) of 3',5'-*O*, 2-*N*-tribenzoyl 6-thiodeoxyguanosine (X) produced 1.59g of crude product which was >95% pure as judged by TLC (system B). The majority of this was used without purification for the next step but a small amount (0.35g) was purified by recrystallisation from methanol to be used for characterisation by FAB-MS and ^{1}H NMR.

^{1}H NMR data (in d_6-DMSO): δ2.35-2.45ppm (1H, m, H2'); [δ2.60ppm (q(J=-1.9Hz, d_5-DMSO)]; δ2.65-2.77ppm (1H, m, H2"); [δ3.42ppm (s, absorbed water in DMSO)]; δ3.60-3.75ppm (2H, m, H5' + H5"); δ3.93ppm-3.99ppm (1H, q(J=4.0Hz), H4'); 4.48-4.53ppm (1H, m, H3'); δ5.05ppm§ (1H, t(J=5.39Hz), 5'OH); δ5.42ppm§ (1H, d(J=3.5Hz), 3'OH); δ6.40ppm (1H, t(J=6.7Hz), H1'); δ7.65-8.22ppm (5H, m, ArH from benzoyl); δ8.55ppm (1H, s, H8); δ12.25ppm§ (1H, s, 2-NHBz); δ13.75ppm§ (1H, s, N1H)

§These peaks disappear on shaking the solution with a little D_2O.

FAB-MS data: 388 (M + H$^+$, 70%), 272 (M - sugar + H + H$^+$, 100%).

2.1.5.iii Synthesis of 2-*N*-benzoyl-S-(2-cyanoethyl)-6-thiodeoxyguanosine (XII)

6.3g of anhydrous potassium carbonate and 4.8ml of 3-bromopropionitrile were added to 100ml of dry dimethylformamide and the mixture stirred vigorously. 1.18g of the crude 2-*N*-benzoyl-6-thiodeoxyguanosine (XI) product obtained above was dissolved in 40ml of dry dimethylformamide and added dropwise to the stirring mixture over a period of about 5 minutes. The reaction was left at room temperature overnight after which time TLC (system C) showed that all the starting material had been converted cleanly into product. The reaction mixture was filtered to remove solid material and the solids washed with 100ml of 15% methanol in dichloromethane. The filtrate was then evaporated to a yellow oil which was redissolved in 150ml of 20% methanol in dichloromethane and insoluble material was removed by filtration. Removal of the solvents from the filtrate resulted in 1.09g of solid that was >95% pure by TLC (system C). For use in compound characterisation, 0.36g of the crude product was purified by flash chromatography eluting with 8% methanol in dichloromethane yielding 0.23g (64%) of solid that was pure by TLC. 50mg of this was further purified by recrystallisation from ethylacetate to give 34mg of pure product.

^{1}H NMR data (in d_6-DMSO): δ2.35-2.45ppm (1H, m, H2'); [δ2.60ppm (q, d_5-DMSO)]; δ2.80-2.90ppm (1H, m, H2"); δ3.28ppm (2H, t(J=6.8Hz), SCH_2CH_2CN); [δ3.43ppm (s, water in DMSO)]; δ3.58-3.76ppm (4H, m, H5' + H5" + SCH_2CH_2CN); δ3.94-4.00ppm (1H, m, H4'); δ4.50-4.58ppm (1H, m, H3'); δ5.00 (1H, t(J=5.51Hz), 5'OH); δ5.41ppm (1H, d(J=4.05Hz), 3'OH);

[continued over page]

δ6.47ppm (1H, t(J=6.75Hz), H1'); δ7.60-7.75ppm (3H, m, ArH from benzoyl); δ8.02-8.08ppm (2H, m, ArH from benzoyl); δ8.7ppm (1H, s, H8); δ11.1ppm (1H, s, 2-NHBz).

FAB-MS data: 441 (M + H^+, 90%), 325 (M - sugar + H + H^+, 100%).

UV data: pH6.5 λ_{max} 293nm

2.1.5.iv Synthesis of 5'-*O*-dimethoxytrityl-2-*N*-benzoyl-S-(2-cyanoethyl)-6-thiodeoxyguanosine (XIII)

0.70g of the crude 2-*N*-benzoyl-S-(2-cyanoethyl)-6-thiodeoxyguanosine (XII) produced above was 5'-dimethoxytritylated as in step 2.1.4.vi. After the work up procedure, the crude product was purified by flash chromatography using 0.5% pyridine in ethylacetate as eluent. 0.67g (47% for the three steps 2.1.5.ii, 2.1.5.iii and 2.1.5.iv) of product was obtained which was pure by TLC (system A).

^{1}H NMR data (in $CDCl_3$): δ2.50-2.60ppm (1H, m, H2'); δ2.90-3.00ppm (1H, m, H2''); δ3.10ppm (2H, t(J=6.95Hz), SCH_2CH_2CN); δ3.35-3.45ppm (2H, m, SCH_2CH_2CN); δ3.50-3.60ppm (2H, m, H5' + H5''); δ3.73 (3H, s, OCH_3 from DMT); δ3.74ppm (3H, s, OCH_3 from DMT); δ4.10-4.25ppm (1H, m, H4'); δ4.80-4.85ppm (1H, m, H3'); δ6.46ppm (1H, t(J=6.57Hz), H1'); δ6.68-7.75ppm (18H, m, ArH from benzoyl and DMT); δ8.05ppm (1H, s, H8); δ8.45ppm (1H, s, 2-NHBz).

FAB-MS data: 743 (M + H^+, 1%), 303 (DMT^+, 100%).

2.1.5.v Synthesis of 3'-*O*-[(*N*,*N*-diisopropylamino)-2-cyanoethylphosphoramidite]-5'-*O*-dimethoxytrityl-2-*N*-benzoyl-S-(2-cyanoethyl)-6-thiodeoxyguanosine (XIV)

0.61g (0.82mmol) of pure 5'-*O*-dimethoxytrityl-2-*N*-benzoyl-S-(2-cyanoethyl)-6-thiodeoxyguanosine (XIII) was converted to its phosphoramidite derivative in the usual manner as described in 2.1.4.vii. The crude product obtained was purified by flash chromatography eluting with 2% triethylamine in a 1:1 mixture of ethylacetate and dichloromethane to yield 0.54g (70%) of product that was pure by TLC (same solvent system as for flash chromatography).

2.1.5.vi Synthesis of deblocked 6-thiodeoxyguanosine (XV)

0.40g (0.67mmol) of 3',5'-*O*, 2-*N*-tribenzoyl-6-thiodeoxyguanosine (X) was suspended in 20ml of dry methanol under a nitrogen atmosphere. 0.76ml (3.3mmol) of a 25% wt/wt solution of sodium methoxide in methanol was added and the solution stirred at room temperature overnight. After 20 hours, all the starting material had reacted as seen by reverse phase TLC

(using RP-18F_{254} plates developed in a 2:3 solvent mix of water in methanol). The reaction was therefore stopped by the addition, in small batches, of pyridinium form Dowex resin until the solution was at neutral pH. The resin was filtered off, washed with a 2:3 mixture of water in methanol and the filtrate reduced on a rotary evaporator to give 0.16g (86%) of crude product. 0.060g of the crude solid was purified by flash chromatography using LiChroprep RP-18 as the reverse stationary phase and a 2:3 mixture of water in methanol as the eluent. This yielded 0.027g (45%) of pale yellow product that was pure by reverse phase TLC and virtually pure by HPLC with deoxyguanosine (<1%) as the only contaminant.

^{1}H NMR data (in d_6-DMSO): δ2.27-2.37ppm (1H, m, H2'); δ2.56-2.68ppm (1H, m, H2''); [δ3.44ppm (s, absorbed water in DMSO)]; δ3.57-3.63ppm (1H, m, H5'); δ3.64-3.69ppm (1H, m, H5''); δ3.89ppm-3.93ppm (1H, q(J=4.0Hz), H4'); 4.42-4.46ppm (1H, m, H3'); δ5.02ppm (1H, broad s, 5'OH); δ5.37ppm (1H, broad s, 3'OH); δ6.21ppm (1H, t(J=6.8Hz), H1'); δ6.9ppm (2H, s, 2-NH_2); δ8.21ppm (1H, s, H8); δ12.05ppm (1H, s, N1H).

FAB-MS data: 284 (M + H^+, 50%), 168 (M - sugar + H + H^+, 100%).

UV data: pH=4.0; λ_{max} at 258 and 341nm.
pH=6.9; λ_{max} at 255 and 341nm.
pH=12.0; λ_{max} at 251 and 319nm.

2.1.5.vii Kinetics of the Ammonia Deblock of 2-*N*-benzoyl-S-(2-cyanoethyl)-6-thiodeoxyguanosine (XII)

To determine how long it would take to remove the cyanoethyl and benzoyl protecting groups from the 6-thiodeoxyguanosine analogue under the usual oligodeoxynucleotide deblocking conditions [see section 2.2.4] and to see whether the deblocked deoxynucleoside would be stable to these conditions, a sample of 2-*N*-benzoyl-S-(2-cyanoethyl)-6-thiodeoxyguanosine (XII) was heated in 35% aqueous ammonia at 55°C. 100mg of XII was dissolved in 5ml of aqueous ammonia in a sealed flask and heated to 55°C. 200μl samples were removed at various times, dried down and then redissolved in water. The samples were analyzed by reverse phase HPLC using an octadecyl silyl column at room temperature and a gradient of 0 to 15% buffer B [see section 2.2.2 for buffer composition] over 8 minutes followed by an isocratic gradient of 32% B.

2.1.6 Synthesis of the 3-deazadeoxyguanosine, d[^{3C}G], 7-deazadeoxyguanosine, d[^{7C}G], and 2-pyrimidinone-1-β-D-(2'-deoxyriboside), d[^{4H}C], Analogues

The 3-deazadeoxyguanosine and 2-pyrimidinone-1-β-D-(2'-deoxyriboside) analogues were provided as the 5'-DMT protected phosphoramidite derivatives by B. Connolly (University of Southampton). The synthesis of the 3-deazadeoxyguanosine derivative was based on a method published by Revankar *et al.* (1984) with modification of a few of the steps.

The synthesis of the 2-pyrimidinone-1-β-D-(2'-deoxyriboside) involved essentially the same pathway as that used for the equivalent thymidine analogue (Connolly & Newman, 1989) except that deoxyuridine rather than thymidine was used as the starting material.

The 7-deazadeoxyguanosine was a kind gift of Phillip Gottlieb (University of Delaware, USA) and was supplied as the 5'-dimethoxytritylated deoxynucleoside derivative. It was phosphitylated as in 2.1.4.vii.

2.2 SYNTHESIS AND PURIFICATION OF OLIGODEOXYNUCLEOTIDES CONTAINING BASE ANALOGUES

2.2.1 Materials

Snake venom phosphodiesterase (2mg/ml), from *Crotalus durissus*, and nuclease P1 (1mg of solid) were purchased from Böhringer Mannheim (Lewes, East Sussex). Calf intestinal alkaline phosphatase (1U/μl) and T4 polynucleotide kinase (8U/μl) were from Promega (Southampton, Hampshire). NAP columns were from Pharmacia (Uppsala, Sweden).

All DNA synthesis reagents, including deoxyguanosine, deoxyadenosine, thymidine, deoxycytidine, deoxyinosine and 5-methyldeoxycytidine phosphoramidites, were obtained from Applied Biosystems (Warrington, Cheshire). All other phosphoramidites were prepared as described in section 2.1.

2.2.2 Experimental and Analytical

Reverse phase high performance liquid chromatography (HPLC) was carried out on a Varian 5000 liquid chromatograph fitted with an Apex octadecyl silyl column of dimensions 25cm x 0.45cm (Jones Chromatography, Mid Glamorgan). Detection was routinely at 254nm by means of a Varian UV-50 variable detector and chart recorder. Peak areas were determined by computer integration using an Apple IIe personal computer connected to the UV detector. Buffer systems were constructed from the following buffers:-

Buffer A: 100mM aqueous triethylammonium acetate (pH6.5)
+ 5% (v/v) acetonitrile.

Buffer B: 100mM aqueous triethylammonium acetate (pH6.5)
+ 65% (v/v) acetonitrile.

Buffer C: 100mM aqueous triethylammonium bicarbonate (pH6.5)
+ 5% (v/v) acetonitrile.

Buffer D: 100mM aqueous triethylammonium bicarbonate (pH6.5)
+ 65% (v/v) acetonitrile.

Buffer E: 100mM aqueous triethylammonium acetate (pH6.5).

Triethylammonium acetate buffers were prepared using HPLC grade acetonitrile (Rathburn Chemicals Ltd., Walkerburn, Scotland), HPLC grade triethylamine and acetic acid (BDH, Poole, Dorset). These buffers were degassed under water pump vacuum. Unless otherwise stated, buffer A was used as reservoir A and columns were run at 55°C.

Triethylammonium bicarbonate buffers were freshly prepared by bubbling carbon dioxide through 100mM aqueous triethylamine containing the appropriate amount of acetonitrile.

Other buffers were as follows:-

Buffer α10:-

10mM magnesium chloride
100mM sodium chloride
50mM HEPES (pH 7.5)

Buffer β:-

10mM magnesium chloride
5mM dithioerythritol
50mM Tris (pH 7.6)
0.1mM spermidine
0.1mM EDTA

Buffer γ:-

10mM magnesium chloride
100mM sodium chloride
25mM KH_2PO_4 (pH 7.2)

Buffer δ:-

5mM sodium chloride
5mM HEPES (pH 7.5)
5% (v/v) acetonitrile

UV spectroscopy was performed using a Uvikon 930 spectrometer (Kontron Instruments) fitted with a thermostatted cell holder. Fluorescence spectra were recorded using a Hitachi F-2000 fluorescence spectrophotometer with a xenon lamp excitation source.

Melting temperature (T_M) determinations were carried out by G. Kneale (Portsmouth Polytechnic) using a Perkin-Elmer Coleman 572 spectrometer. The cuvette temperature was raised from 10°C to 90°C at a rate of between 0.5 and 1.0°C per minute with the actual temperature being measured in an adjacent dummy cuvette.

Circular dichroism spectra were recorded by A. Drake (Birkbeck College, University of London) using a Jasco J40-CS spectrometer.

2.2.3 Synthesis of Oligodeoxynucleotides

All oligodeoxynucleotides were synthesised on a 1μmol scale using a fully automated Applied Biosystems 381A DNA synthesizer and standard phosphoramidite chemistry [see section 1.2.2]. All modified base phosphoramidites were dried overnight *in vacuo* and then dissolved in the appropriate amount of anhydrous acetonitrile to give the usual concentration of 0.1M. These solutions were filtered through 0.5μ teflon filters before use and were then used in the 'X' position of the machine. Normal amide protected phosphoramidites were used in the other positions on the machine except in the cases of the d[^{4H}C] analogues where the more easily deprotected FOD phosphoramidites were used[1]. The normal 1μmol synthesis cycle was used without alteration for the base analogues and all oligodeoxynucleotides were synthesised with the final 5'-*O*-dimethoxytrityl group left on.

2.2.4 Deblocking of Oligodeoxynucleotides

All oligodeoxynucleotides were cleaved from the glass support and the base protecting groups removed by treatment with 35% aqueous ammonia in a sealed bottle. 55°C for three to five hours was used for those produced using normal phosphoramidites, with care being taken to ensure that no more than four hours was used in the case of d(GAC[^{6S}G]ATATCGTC). Those produced from FOD phosphoramidites were deprotected by treatment with 35% aqueous

[1] FOD (fast oligodeoxynucleotide deblock) is the trademark of Applied Biosystems (Warrington, Cheshire) for their more easily deprotected phosphoramidites based on the work of McBride *et al.* (1986).

ammonia in a sealed bottle for one hour at 55°C or eight hours at ambient temperature.

2.2.5 Purification of Oligodeoxynucleotides

After drying down from aqueous ammonia, the deprotected oligodeoxynucleotides were redissolved in about 300µl of water and washed twice with butan-1-ol to remove the now free base protecting groups (this step was omitted for oligodeoxynucleotides produced from FOD phosphoramidites). These 'trityl-on' oligodeoxynucleotides were then purified by reverse phase HPLC using a column heated to 55°C and generally a gradient of 25 to 70% HPLC buffer B over 20 minutes, injecting 25µl samples. In some cases the gradient needed to be adjusted to improve peak separation. The eluent was viewed by ultraviolet absorbance at 254nm and the desired 'trityl-on' oligodeoxynucleotides eluted after about twelve minutes with this system. Fractions containing the required oligodeoxynucleotide were pooled and evaporated to dryness. The dimethoxytrityl group was removed by dissolving the solid up in about 1ml of 70% acetic acid and then leaving to react for 1.5 hours at room temperature. The acetic acid was then removed either by evaporation and then coevaporation three times from water or by elution through a NAP-25 Sephadex gel filtration column. NAP columns were pre-equilibrated and eluted with buffer δ. 0.5ml fractions were collected and those containing oligodeoxynucleotide as judged by UV absorbance were pooled and evaporated to dryness.

After the final evaporation, the solid residue was dissolved up in ~300µl of water and viewed by HPLC (5-15% buffer B over 15 minutes). At this stage most oligodeoxynucleotides were >95% pure and could be further purified simply by HPLC using the same gradient. However, because more than one peak was visible in their HPLC traces, the d[^{6S}G] and the d[^{4H}C] analogues required slight alterations to this procedure as discussed below.

2.2.5.i Purification of 'DMT-off' d(GAC[^{6S}G]ATATCGTC)

Preparations of the detritylated d[^{6S}G] analogue oligodeoxynucleotide were found to contain disulphides. Therefore to maximise the yield the oligodeoxynucleotide was treated with dithioerythritol (DTE) prior to purification by HPLC. Also, to improve separation from a contaminant peak and to give better resolution of the unusually broad product peak, a gradient of 0 to 40% buffer B over 20 minutes was used.

2.2.5.ii Purification of 'DMT-off' d(GACGATAT[^{4H}C]GTC)

The HPLC of the detritylated d[^{4H}C] analogue oligodeoxynucleotide showed three main peaks. In order to separate these peaks a shallower HPLC gradient was needed and the three products were purified using an isocratic gradient of 7% buffer B. The last peak to elute was found to be the desired product from its base composition analysis and was used in the following purification and characterisation steps.

2.2.6 Desalting of Oligodeoxynucleotides

Oligodeoxynucleotides were desalted by injecting down an octadecyl silyl HPLC column at room temperature eluting with volatile triethylammonium bicarbonate buffers. A gradient of 0 to 40% buffer D (with buffer C in reservoir A) was used in all cases except that of the d[^{6S}G] oligodeoxynucleotide were a shallower gradient of 0 to 20% D was used in order to separate product from a small contaminating peak. Collected fractions containing product were evaporated to dryness and the residual bicarbonate salts removed by coevaporation three times with analar grade methanol. The final oligodeoxynucleotides were redissolved in about 200µl water and stored at -20°C.

2.2.7 Characterisation of Analogue Oligodeoxynucleotides

All oligodeoxynucleotides were analyzed to check for correct base composition, stability of DNA duplex and deviation from B-form structure as outlined below.

2.2.7.i Base Composition Analysis of the Oligodeoxynucleotides

Base composition of oligodeoxynucleotides was determined by digestion with snake venom phosphodiesterase which cleaves DNA completely to its constituent 5'-nucleotides (Laskowski, 1971). Alkaline phosphatase, which removes 5'-phosphates from nucleotides (Sambrook *et al.*, 1989), was also included in the reaction mixture as removal of the phosphate group facilitates subsequent separation by HPLC. The following constituents were dissolved in buffer α10:-

1. 5-10nmol oligodeoxynucleotide
2. 10µg Snake Venom Phosphodiesterase
3. 10U Alkaline Phosphatase
4. 10µg Nuclease P1)§

§This also cleaves DNA to its constituent nucleotides (Kihara *et al.*, 1976) and needed to be added to oligodeoxynucleotides containing the analogues d[^{6S}G], d[^{6H}G], d[^{7C}G] and d[I] since these were resistant to cleavage by snake venom phosphodiesterase alone.

The deoxynucleosides were separated by HPLC using a gradient of 3 to 18% buffer B (with buffer E as reservoir A) over 15 minutes and the majority of them viewed at 254nm. The relative amount of each deoxynucleoside was calculated from the area of the peak divided by the extinction coefficient of the deoxynucleoside. Because the absorbances at 254nm of d[^{6S}G] (ε_{254} 9,000M^{-1}cm^{-1}; Fox *et al.*, 1958) and of d[^{4H}C] (ε_{254} 2,150M^{-1}cm^{-1}; Gidea & McLaughlin, 1989) are so small, UV detection of these deoxynucleosides was switched from 254nm to their respective λ_{max}'s of 340nm for d[^{6S}G] (ε_{340} 20,000M^{-1}cm^{-1}) and 305nm for d[^{4H}C] (ε_{305} 6,470M^{-1} cm^{-1}). Also because of the tendency of d[^{6S}G] to form disulphides, samples from this analogue

were treated with 5mM DTE at pH8 prior to HPLC injection. All peaks were identified by co-injection with solutions of the standard deoxynucleosides.

2.2.7.ii Hyperchromicity Measurement of the Oligodeoxynucleotides

In order to determine the extinction coefficients of the various oligodeoxynucleotides, the change in absorbance at 254nm on digestion of the oligodeoxynucleotides was observed. Oligodeoxynucleotides at a concentration of about 5μM in buffer α10 were digested with 20μg of snake venom phosphodiesterase [and 20μg of Nuclease P1 when needed; see 2.2.7.i] at 25°C. The absorbance at 254nm was followed with time and the initial and final (i.e. when the absorbance no longer increased but stayed constant) absorbances were recorded. To ensure complete digestion, the final samples were viewed by HPLC using a gradient of 3 to 23% buffer B (with buffer E as reservoir A) over 20 minutes. The increase in absorbance observed together with the sum of extinction coefficients at 254nm of the constituent deoxynucleosides (given in Table 2.1) can then be used to calculate the extinction coefficient of the oligodeoxynucleotide:-

$$\varepsilon_{254} = (\text{sum of deoxynucleoside } \varepsilon_{254})/\text{hyperchromicity}$$

Table 2.1: Extinction coefficients of deoxynucleosides used in this study. [a]From Fasman (1975), [b]from that given for the riboside in Fox *et al.* (1958), [c]from Winkeler & Seela (1983), [d]calculated from its' UV spectrum and the ε_{max} value given in Revankar *et al.* (1984), [e]from Gildea & McLaughlin (1989).

DEOXYNUCLEOSIDE	x 10^3 EXTINCTION COEFFICIENT /$M^{-1}cm^{-1}$
dA	14.3[a]
dG	13.5[a]
dT	7.0[a]
dC	6.0[a]
dI	10.0[a]
d[^{5Me}C]	5.1[a]
d[^{6S}G]	9.0[b]
d[^{6H}G]	3.0[b]
d[^{7C}G]	12.9[c]
d[^{3C}G]	7.3[d]
d[^{4H}C]	2.1[e]

2.2.7.iii Determination of the Melting Temperature (T_M) of the Oligodeoxynucleotides

In order to check that the analogue oligodeoxynucleotides were double stranded under the conditions used for the enzyme kinetics, and to assess the stability of the analogue duplices, their melting temperatures (T_M) were measured. This was done by measuring the absorbance at 254nm of 4-6μM oligodeoxynucleotide in buffer α10 as a function of temperature. From a graph of absorbance against temperature the T_M was determined as the point of greatest rate of change of absorbance.

2.2.7.iv Circular Dichroism Spectroscopy of the Oligodeoxynucleotides

The circular dichroism (CD) spectra of all the oligodeoxynucleotides were recorded in the UV region at a concentration of about 5μM in buffer γ to check for any major deviation from B-form DNA.

2.2.7.v Ultraviolet Spectroscopy of d(GAC[^{6S}G]ATATCGTC), d(GAC[^{6H}G]ATATCGTC), d(GACGATAT[^{4H}C]GTC) and d(GAC[^{3C}G]ATATCGTC)

Because the λ_{max}'s of d[^{6S}G], d[^{6H}G] and d[^{4H}C] are well removed from those of the normal bases (254-271nm), the UV spectra of their corresponding oligodeoxynucleotides were recorded to see whether the presence of the modified base could be confirmed by the appearance of an extra absorbance band not seen in normal DNA. 3-deazadeoxyguanosine also has a shoulder in its' UV spectrum that is above the range of normal bases and so the UV spectrum of d(GAC[^{3C}G]ATATCGTC) was also recorded.

2.2.7.vi Fluorescence Spectroscopy of d(GACGATAT[^{4H}C]GTC) and d(GAC[^{6H}G]ATATCGTC)

The excitation and emission spectra of the d[^{6H}G] and the d[^{4H}C] containing dodecamers were recorded at 25°C. Amounts of oligodeoxynucleotide that gave satisfactory fluorescence were added to buffer γ and their concentrations determined after fluorescence spectra were recorded by measuring the UV absorbance of the solutions. The excitation and emission spectra of the free d[^{6H}G] (0.25μM) and d[^{4H}C] (22.3μM) nucleosides were also recorded in buffer γ. Relative quantum yields were determined for the oligodeoxynucleotides and the deoxynucleosides from the areas of the emission bands and the relative absorbances at the corresponding excitation maxima.

The fluorescence of d(GAC[^{6H}G]ATATCGTC), with the excitation wavelength at 300nm and the emitted light at 360nm collected, was recorded as the temperature of the solution was raised from 25°C to 70°C.

2.2.8 Phosphorylation of Oligodeoxynucleotides

Oligodeoxynucleotides were phosphorylated after 'DMT-off' purification but before performing the desalting step. Reactions were carried out at 37°C in 500µl of buffer β as follows:-

1. 100nmol (200µM) of double stranded oligodeoxynucleotide.
2. 50µl of 0.1M ATP (gives final concentration of 10mM).
3. 80U of T4 Polynucleotide Kinase.

Reactions were followed by HPLC (0 to 15% buffer B over 15 minutes) with the phosphorylated oligodeoxynucleotide appearing just before the starting material. Generally after a few hours the reaction had virtually gone to completion but to ensure 100% conversion, the reactions were left at 37°C for 2 to 3 days.

Phosphorylated oligodeoxynucleotides were purified and desalted in the same step by injecting down an HPLC column at room temperature and eluting with volatile buffers as described in section 2.2.6. The identity of the phosphorylated oligodeoxynucleotides were confirmed by treatment of a small sample with alkaline phosphatase in buffer α10 followed by HPLC co-injection with the corresponding non-phosphorylated oligodeoxynucleotide (gradient of 0 to 15% buffer B over 15 minutes).

2.2.9 Stability of d(GACGATAT[^{4H}C]GTC) to Deblocking Conditions

As the thymidine equivalent of d[^{4H}C] is known to be sensitive to the ammonia deblock procedure (Connolly & Newman, 1989), the stability of d(GACGATAT[^{4H}C]GTC) to 35% aqueous ammonia at 55°C was investigated. About 100nmol of the oligodeoxynucleotide was dissolved in 0.5ml of 35% aqueous ammonia and then heated at 55°C in a sealed tube. After 2 hours, the solution was cooled, evaporated to dryness and then redissolved in water. The solution was analyzed by HPLC (0 to 15% buffer B over 20 minutes) and found to consist of two main peaks. The later eluting peak co-eluted with a standard sample of d(GACGATAT[^{4H}C]GTC). The earlier eluting peak, designated as X, was purified by HPLC for further analysis.

The stability of X to ammonia was also investigated. About 30nmol of X in 35% aqueous ammonia was heated to 55°C in a sealed tube. After 22 hours, the solution was cooled and evaporated to dryness. After redissolving in water HPLC of the mixture using the same gradient as above, revealed that most of X had been converted into two different peaks, Y and Z. The earlier eluting peak, Y, was found to be converted into a different, later eluting peak on treatment with alkaline phosphatase. This peak and peak Z were purified by HPLC for further analysis. All three compounds, X, Y and Z had there base composition determined as described in 2.2.7.i.

2.3 ECO RV ENDONUCLEASE PREPARATION AND PURIFICATION

The Eco RV endonuclease was prepared from a sample of frozen cell paste of the overproducing *E. Coli* strain (Bougueleret *et al.*, 1985) that had previously been produced (Newman, 1989). Purification was based on the method of D'Arcy *et al* (1985).

2.3.1 Materials

Benzamidine, dithioerythritol (DTE) and phenylmethylsulphonylfluoride were from Sigma (Poole, Dorset). All other chemicals were from BDH (Poole, Dorset) and of analar grade.

Sephadex G-75-50 was obtained from Sigma (Poole, Dorset) and phosphocellulose P11 was obtained from Whatman (Maidstone, Kent). Centriprep-10 concentration tubes were obtained from Amicon (Beverly, USA) and used as described by the manufacturers.

2.3.2 Experimental and Analytical

Buffers were as follows:-

Buffer ε:-

20mM KH_2PO_4 (pH 7.4)
2mM DTE
1mM EDTA
10% glycerol
0.1mM phenylmethylsulphonylfluoride
0.1mM benzamidine

Buffer ζ:-

20mM KH_2PO_4 (pH 7.4)
2mM DTE
1mM EDTA
0.2M sodium chloride

Sample Disruption Buffer:-

60mM Tris (pH 6.8)
2% (w/v) SDS
10% (v/v) glycerol
5% (v/v) 2-mercaptoethanol
0.1% bromophenol blue

SDS Running Buffer:-

25mM Tris (pH 8.3)

250mM glycine

0.1% SDS

SDS polyacrylamide (12%) gels (Sambrook *et al.*, 1989) of 14cm x 12cm were run at a constant current of between 20 and 40mA. 50µl of protein samples were added to an equal volume of sample disruption buffer and then heated at 100°C for 5 minutes. After briefly spinning down, 5µl of each of these protein samples were loaded onto the gel. Molecular weight markers of 66, 45, 36, 29, 21.1 and 14.2 kDaltons were used as standards. Protein bands were stained with 0.25% (w/v) Brilliant Blue G in a 10:45:45 mix of acetic acid:methanol:water for approximately one hour at 37°C. Gels were destained in 5% methanol and 7.5% acetic acid in water until the gel background was no longer blue.

2.3.3 Extraction of the Endonuclease

20g of frozen cell paste was thawed out gently in 100ml of buffer ε containing 0.8M sodium chloride chilled on ice. The suspension was homogenised using a motor driven homogeniser fitted with a teflon pestle. The cells were then sonicated at 0°C with an MSE Soniprep 150. Sonication was for 12 minutes using a cycle of 15 seconds sonication followed by 15 seconds off.

The suspension was then centrifuged at -2°C for 3 hours at 20,000rpm in a Beckman J2-21 centrifuge fitted with an 8 x 50ml rotor. The yellow supernatents were pooled and diluted with the appropriate volume of buffer ε to give a final sodium chloride concentration of 0.2M.

2.3.4 Phosphocellulose Column Purification of the Endonuclease

About 30ml of dry phosphocellulose was made into a slurry in buffer ε containing 0.2M sodium chloride. The buffer was removed by filtration and the solid resuspended in buffer ε containing 0.2M sodium chloride. The pH of the suspension was adjusted with 0.2M aqueous potassium hydroxide to pH7.4. The solid was removed by filtration, resuspended in buffer ε containing 0.2M sodium chloride and the pH checked to ensure it was at 7.4. The suspension was degassed using a water pump and then used to make up a 10cm x 2cm column by the slurry method.

After the phosphocellulose column had been pre-equilibrated with buffer ε containing 0.2M sodium chloride, the solution of soluble cell protein extracted from the cell paste was loaded onto the column. The column was washed through with buffer ε containing 0.2M sodium

chloride overnight, at a flow rate of about 60ml/h, to remove protein that did not bind to the column. Protein bound to the column (including endonuclease) was eluted by running a sodium chloride gradient of 0.2M to 0.8M in 500ml of buffer ε through the column. The gradient was run at 60ml/h and 10ml fractions were collected using an LKB 7000 Ultrarac fraction collector.

All collected fractions had their absorbance at 280nm measured on a Uvikon 930 spectrometer. Using this as a guide, some of the fractions expected to contain protein were run on an SDS polyacrylamide gel in order to discern which fractions contained endonuclease.

Fractions containing endonuclease were pooled and the protein precipitated by the addition to the gently stirring solution on ice, of ammonium sulphate to give a final concentration of 80% (w/v). The solid was isolated by centrifugation at -2°C for 30 minutes at 20,000rpm followed by careful decanting off of the supernatant.

2.3.5 Gel Filtration Column Purification of the Endonuclease

33.5g of Sephadex G-75-50 (bead size 20-50μm) were soaked in water for 24 hours to give approximately 400ml of swollen gel. This was made up to give a 75% (v/v) swollen gel suspension in water (*i.e.* made up to 540ml with water) and then degassed under water pump vacuum. A 90cm x 2.5cm column was prepared from this by pouring the slurry into the glass column, letting the liquid run out of the bottom under gravity and continuously topping up the column with slurry until the desired height of gel was reached.

After pre-equilibrating the column with buffer ζ, the precipitate obtained from the phosphocellulose column purification was redissolved in 5ml of buffer ζ and added to the top of the gel filtration column. The column was eluted with 300ml of the same buffer at a rate of 15ml/h and 10ml fractions were collected. The absorbance at 280nm of the collected fractions was measured on a Uvikon 930 spectrometer. Using this as a guide, samples of some of the fractions expected to contain protein were run on an SDS polyacrylamide gel. On the basis of this, fractions containing pure endonuclease were pooled and then concentrated to 0.5ml using Centriprep tubes. 0.5ml of glycerol was added with careful mixing to the final endonuclease solution. 1ml of an 8 times dilution of this solution was made in a solution of 30% (v/v) glycerol in buffer ζ. Both enzyme solutions were stored at -20°C.

2.3.6 Endonuclease Concentration Determination

The enzyme concentrations of the two solutions were determined using both the absorbance of the solutions at 280nm (ε_{280} 104,400$M^{-1}cm^{-1}$; D'Arcy *et al.*, 1985) and a Bio-Rad Protein Assay system based on the colour change observed when a dye binds to a protein (Bradford, 1976). A standard plot was made for the protein assay method using 0-0.8mg ml^{-1} BSA solutions.

2.4 ENZYME KINETICS

2.4.1 Materials

Eco RV restriction endonuclease was prepared and purified as described in section 2.3. Phosphorylated and non-phosphorylated dodecamers of the control sequence d(GACGATATCGTC) and those containing base analogues were prepared and characterised as described in section 2.2 as was the tridecamer d(TGACGATATCGTC). BSA (solid), snake venom phosphodiesterase from *Crotalus durissus* (2mg/ml), and nuclease P1 (1mg of solid) were from Böhringer Mannheim (Lewes, East Sussex). Calf intestinal alkaline phosphatase (1U/μl) was from Promega (Southampton, Hampshire)

Reaction Buffers (all of the same ionic strength):

Buffer α2:-

2mM magnesium chloride
124mM sodium chloride
50mM HEPES (pH 7.5)

Buffer α5:-

5mM magnesium chloride
115mM sodium chloride
50mM HEPES (pH 7.5)

Buffer α10:-

10mM magnesium chloride
100mM sodium chloride
50mM HEPES (pH 7.5)

Buffer α20:-

20mM magnesium chloride
70mM sodium chloride
50mM HEPES (pH 7.5)

Buffer α25:-

25mM magnesium chloride
55mM sodium chloride
50mM HEPES (pH 7.5)

Buffer α40:-

40mM magnesium chloride
10mM sodium chloride
50mM HEPES (pH 7.5)

Reverse phase HPLC was performed as described in section 2.2 and UV spectra recorded on a Uvikon 930 spectrometer.

2.4.2 HPLC Assay of the Endonuclease Reaction with Oligodeoxynucleotides

To determine whether they were substrates or not for the Eco RV endonuclease, samples of the analogue oligodeoxynucleotides were incubated at 25°C with the enzyme in 200μl of buffer α10 as follows:-

1. 4nmol (20μM) double stranded oligodeoxynucleotide.
2. 10U Alkaline Phosphatase.
3. 340pmol (1.7μM) endonuclease dimer.

(1mM DTE was included in the cleavage of the d[^{6S}G] oligodeoxynucleotide)

Any reaction that occurred was observed by HPLC using a column heated to 55°C and a gradient of 0 to 15% buffer B over 20 minutes. The two product peaks are well resolved from each other and the dodecamer (the use of alkaline phosphatase in the reaction mixture helps this by dephosphorylating one of the products) and elute before the dodecamer. If no reaction was observed after one hour then a further 680pmol of enzyme was added to the reaction. If this failed to produce cleavage after another hour then another 680pmol of endonuclease was added. Oligodeoxynucleotides that were not cleaved after 24 hours under these conditions were specified as non-substrates.

Cleavable oligodeoxynucleotides were allowed to be hydrolysed completely and then the two product peaks produced were purified by HPLC. The base composition of the products was then established using the method described in section 2.2.7.i to ensure that cleavage by the endonuclease had occurred in the correct position.

2.4.3 UV Assay of the Endonuclease Reaction with Oligodeoxynucleotides

The cleavage of double stranded dodecamer by the endonuclease produces hexameric DNA which is single stranded at 25°C. This causes a hyperchromic increase in the UV absorbance of the reaction mixture that is proportional to the extent of cleavage. This property provides the basis of a convenient assay for studying Eco RV endonuclease reactions (Waters & Connolly, 1991b).

2.4.3.i Determination of the Values of the Hyperchromicity Changes Induced by the Endonuclease Cleavage of Oligodeoxynucleotides

A prerequisite of the UV assay is that the increase in absorbance at the assay temperature on going from double stranded dodecamer substrate to single stranded hexamer products is known accurately. This was found for all of the endonuclease substrate dodecamers and the control tridecamer in a way analogous to that used for 'normal' hyperchromicity measurement as described in 2.2.7.ii. Reactions were carried out in 1ml of buffer α25 contained in 1cm path length quartz cuvettes thermostatted at 25°C. The change in absorbance was recorded at 254nm on cleaving approximately 4nmol (4μM) of oligodeoxynucleotide substrate with enough endonuclease to completely cleave the oligodeoxynucleotide in one to five hours. With the d(GAC[^{6S}G]ATATCGTC) oligodeoxynucleotide, 10μmol (10mM) of DTE was included in the reaction mixture. The reactions were followed by UV absorbance at 254nm and the initial and final absorbance values at 254nm were recorded. The final absorbance was taken when the UV absorbance stopped increasing and this was confirmed as being the end of the reaction by HPLC analysis of the final solution (0 to 15% buffer B over 20 minutes). The initial and final wavelength scans were also recorded over the range 200nm-400nm so that endonuclease hyperchromicities at wavelengths other than 254nm could be determined.

2.4.3.ii Comparison of the UV and HPLC Methods of Assaying Eco RV Endonuclease Cleavage of d(GACGATATCGTC)

The initial rate of cleavage of the non-phosphorylated control oligodeoxynucleotide was determined at 25°C using the HPLC and the UV assays. 6.8nmol (6.8μM) of non-phosphorylated d(GACGATATCGTC) was reacted with 10pmol (10nM) of endonuclease in 1ml of buffer α10 together with 5U of alkaline phosphatase. The reaction was carried out in a 1cm path length quartz cuvette and was followed by the increase in 254nm absorbance with time. Samples of 15μl were removed from the reaction at various times and quenched by mixing with 5μl of 0.5M sodium hydroxide. The rate of cleavage was determined using the HPLC assay by injecting the samples onto an HPLC column which separated out the two products and the remaining starting material (0 to 15% buffer B over 20 minutes). The relative amounts of substrate and product hexamers were calculated from the areas of their respective peaks found by computer integration. The computer integrator had been calibrated by injection of known quantities of both hexameric products (d(GACGAT) and d(ATCGTC)) and of the control dodecamer substrate and so the peak areas could be converted into amounts of product and substrate. The rate of cleavage was also determined from the rate of increase in UV absorbance of the reaction.

2.4.3.iii Stability of Dilute Endonuclease

To check to see whether dilute solutions of the endonuclease needed to be stabilised by the inclusion of BSA in the dilution buffer, the UV assay was used to assess the activity of two different solutions of 100nM endonuclease. One 100nM enzyme solution was made by diluting stock endonuclease with buffer α25 containing 1mg/ml BSA and another 100nM solution by diluting stock endonuclease with buffer α25 alone. The solutions were kept on ice and their activity determined at zero time, after 1 hour, 2.5 hours and 6 hours. Samples of 5pmol (5nM) of enzyme were added to 2nmol (2μM) of control dodecamer in 1ml of buffer α25 in a 1cm path length cell at 25°C and the rate of cleavage measured using the change in 254nm absorbance.

In order to check that the BSA had no contaminating DNA cleaving activity, a blank was run by adding the same volume of buffer α25 containing 1mg/ml BSA as the endonuclease was added in (*i.e.* 50μl), to 2nmol (2μM) of control oligodeoxynucleotide in 1ml of buffer α10 and observing any change in the 254nm absorbance.

2.4.3.iv Dependence of the Rate of Cleavage on Endonuclease Concentration

To check that the rate of oligodeoxynucleotide cleavage was linear with respect to the endonuclease concentration, a sample of 1μM endonuclease and one of 100nM endonuclease were made up by diluting the stock solution with buffer α25 containing 1mg/ml BSA. The rate of cleavage of non-phosphorylated control oligodeoxynucleotide was determined at various enzyme concentrations by adding different amounts of endonuclease (2 to 10pmol from the 100nM solution and 20 to 100pmol from the 1μM solution) to 3nmol (3μM) of control oligodeoxynucleotide in 1ml of buffer α25 at 25°C and monitoring the change in UV absorbance 254nm.

The UV assay was also used to determine the amount of endonuclease present in samples of unknown enzyme concentration obtained from a phosphocellulose column purification of the protein (similar to that described in 2.3.4) in order to ascertain which fractions contained Eco RV endonuclease. 5μl of each fraction was added to 2.25nmol (4.5μM) of non-phosphorylated control dodecamer in 0.5ml of buffer α10 and their activity determined. As a comparison, the absorbance of some of the fractions were measured at 280nm.

2.4.3.v Initial Rate Determination of Control and Base Analogue Containing Oligodeoxynucleotides by the UV Assay

The rates of cleavage of non-phosphorylated analogue oligodeoxynucleotides that were Eco RV endonuclease substrates were measured using the UV assay. Approximately 5nmol (5μM) of double stranded oligodeoxynucleotide were dissolved in 1ml of buffer α25 in a 1cm path length quartz cuvette. Reactions were started by the addition of endonuclease (the amount added depending on the ease with which the particular substrate was cleaved). The cuvettes were kept at 25°C and the rate of change of the absorbance at 254nm followed with

time. To prevent binding of DNA to the quartz cuvettes, they were siliconised. This was done by rinsing out the cleaned cuvettes with a 2% dimethyldichlorosilane solution in 1,1,1-trichloroethane, allowing them to dry and then thoroughly rinsing them out with water.

2.4.3.vi Michaelis-Menten Kinetics of Oligodeoxynucleotides

The Michaelis-Menten kinetic parameters, k_{cat} and K_M, of phosphorylated control dodecamer and phosphorylated analogue dodecamers that were endonuclease substrates were determined at 25mM magnesium ion concentration. Each determination consisted of measuring the initial rates of cleavage (first few percent of substrate cleaved) at six different substrate concentrations using the UV assay. An endonuclease concentration of between 2nM and 300nM was used, depending on the reactivity of the substrate, and substrate concentrations were varied in the range of 0.2μM to 8.3 μM double stranded oligodeoxynucleotide.

In order to get accurate k_{cat} and K_M values, the substrate concentration should be varied around the K_M value. Using the UV assay, results could be plotted as the experiment progressed. If from these initial plots the substrate concentrations were found not to be near the K_M value, then initial rates could be determined at appropriate substrate concentrations without the need to redo the whole experiment. Reactions at substrate concentrations below 1.2μM were performed in 2ml of buffer α25 in 2cm path length quartz cuvettes and those above 1.2μM were performed in 1ml of buffer α25 in 1cm path length cuvettes. The cuvettes were siliconised as described in section 2.4.3.v. A fresh dilute endonuclease solution was prepared for each determination. Stock endonuclease was diluted with buffer α25 containing 1mg/ml BSA and kept on ice for the duration of the experiment.

The Michaelis-Menten kinetic parameters of non-phosphorylated control dodecamer, d(GACGATATCGTC), and of non-phosphorylated d(TGACGATATCGTC) were also determined.

Values of k_{cat} and K_M were determined from Hanes plots (Fersht, 1985) of the substrate concentration divided by the initial rate ($[S_0]/V$) against substrate concentration ($[S_0]$). The thermodynamic constants, ΔG_{app}, were calculated from the k_{cat} and K_M values as described in section 1.2.1.

2.5 RESONANCE RAMAN SPECTROSCOPY

Resonance Raman spectral studies were performed on the 4-thiothymidine deoxynucleoside and 4-thiothymidine containing oligodeoxynucleotides.

2.5.1 Materials

Deuterium oxide was from Aldrich (Gillingham, Dorset). Organic solvents were from BDH (Poole, Dorset) and were of spectroscopic or similar grade. 4-thiothymidine containing deoxynucleoside was provided by B.Connolly (University of Southampton) and was prepared

as described in Connolly and Newman (1989). 5'-*O*-(4,4'-dimethoxytrityl)-4-(S-*p*-nitrophenyl)-thiothymidine was supplied by T.Nikiforov (University of Southampton) and prepared as described in Nikiforov and Connolly (1991). DNA synthesis reagents were from Applied Biosystems (Warrington, Cheshire). All other chemicals were from BDH (Poole, Dorset).

Eco RV endonuclease was prepared as described in section 2.3.

2.5.2 Experimental and Analytical

UV Buffer:

Buffer γ:-

10mM magnesium chloride
100mM sodium chloride
25mM KH_2PO_4 (pH 7.2)

Raman Buffers:

Buffer α0:-

130mM sodium chloride
50mM HEPES (pH 7.5)
1mM EDTA

Buffer η:-

10mM KH_2PO_4 (pH 7.5)
100mM sodium chloride

A Coherent Innova 90K argon ion laser tuned to 363.8nm and at a power of 10 to 50mW was used as the excitation source. Laser plasma lines were removed from the laser beam by passing it through a Pellin-Broca dispersing prism. Spectra were obtained from 100µl samples in a rotating quartz cuvette (diameter 10mm x 3mm high). Sample concentrations were between 2 and 10 absorbance units (corresponds to 90-450µM for the 4-thiothymidine deoxynucleoside) at their λ_{max}'s and were deoxygenated by bubbling argon through them prior to running spectra. Raman scattered light was collected in the backscattering mode (*i.e.* at 180 degrees to the incident beam) with a Spex Triplemate spectrometer fitted with a 2400g/mm UV blazed holographic grating in the final dispersion stage. Detection of radiation was with a Spectroscopic Instruments IRY700 UV-intensified diode array which had the reticon cooled to -35°C and the photocathode cooled to -20°C. Spectra were calibrated to an accuracy of $\pm 2cm^{-1}$ with ethanol, carbon tetrachloride, toluene and acetone Raman peaks (wavenumbers taken from Schrader, 1989) each time the grating was moved. At least two runs, of approximately three minutes each, were performed for each spectrum (four runs for the majority) to reduce signal-to-noise.

2.5.3 Synthesis of Oligodeoxynucleotides Containing 4-Thiothymidine

The 4-thiothymidine containing oligodeoxynucleotides d(AG[^{4S}T]TC), d(GACGA[^{4S}T]ATCGTC) and d(GACGATA[^{4S}T]CGTC) were prepared using standard automated phosphoramidite techniques [see section 2.2] on a one μmol scale with 5'-*O*-(4,4'-dimethoxytrityl)-4-(S-*p*-nitrophenyl)thiothymidine in the X position. These oligodeoxynucleotides were deprotected using a published procedure (Nikiforov & Connolly, 1991). Oligodeoxynucleotides were then dried down *in vacuo* and dissolved up in ~200μl of water. Oligodeoxynucleotides were HPLC purified, detritylated and desalted as described in sections 2.2.5 and 2.2.6 and then stored at -20°C.

The UV spectra of all three 4-thiothymidine containing oligodeoxynucleotides and of the 4-thiothymidine deoxynucleoside were recorded in buffer γ.

2.5.4 Resonance Raman Spectroscopy of the 4-Thiothymidine Deoxynucleoside, [^{4S}T], in H_2O and in D_2O

Spectra of two separate 100μl samples of 450μM (A_{335}=10) 4-thiothymidine in water were recorded. One was over the approximate range of 600-1450cm^{-1} and the other over the range 1000-1800cm^{-1}. Spectra were calibrated and peak positions determined. The two spectra were then joined and the background fluorescence removed by subtraction of a polynomial fitted curve.

This was repeated for two samples of 4-thiothymidine in deuterium oxide. Spectra were also recorded in buffer α0 and in buffer η.

The laser power dependence of the resonance Raman spectrum of 4-thiothymidine was investigated using a 100μl sample of 450μM (A_{335}=10) in water. Spectra in the region 600-1450cm^{-1} region were recorded at various laser powers (3 to 25mW). The areas of two peaks (709cm^{-1} and 1158cm^{-1}) were determined for all the spectra and plotted against laser power.

2.5.5 Resonance Raman Spectroscopy of d(AG[^{4S}T]TC), d(GACGA[^{4S}T]ATCGTC) and d(GACGATA[^{4S}T]CGTC)

Spectra over the range 1000-1800cm^{-1} were recorded from three separate 100μl samples of ~540μM (A_{337}=10) single stranded d(AG[^{4S}T]TC) either in buffer α0 or in buffer η. Since buffer did not affect the peaks, these three spectra were added together and peak positions determined after spectrum calibration. Spectra over the range 600-1450cm^{-1} were also recorded of two separate 100μl samples of d(AG[^{4S}T]TC) (same concentration) in buffer α0. These two spectra were calibrated, added together and peak positions determined.

The high and low wavenumber spectra were merged and fluorescence background

subtracted.

For both d(GACGA[^{4S}T]ATCGTC) and d(GACGATA[^{4S}T]CGTC) , a spectrum was recorded of one 100µl sample of 400-500µM (A_{337}=10) double stranded oligodeoxynucleotide in buffer α0 over the range 600-1450cm^{-1}. The same samples were used to record spectra over the range 1000-1800cm^{-1}. After the spectra were calibrated, peak positions were determined. Corresponding high and low wavenumber spectra were merged and fluorescence background subtracted.

2.5.6 Resonance Raman Spectroscopy of d(GACGA[^{4S}T]ATCGTC) Bound to Eco RV Endonuclease

Spectra of two samples of 100µl volume of 110µM endonuclease plus 100µM (A_{337}=2) d(GACGA[^{4S}T]ATCGTC) in buffer η were recorded in the range 600-1450cm^{-1}. To one of the samples 4µl of 0.5M magnesium chloride (to give 20mM [Mg^{2+}]) was added and another spectrum recorded in the same region. The spectrum of a sample of 100µl of 110µM endonuclease in buffer η was also recorded.

Spectra recorded in the absence of magnesium were added together. All spectra had background endonuclease, endonuclease storage buffer and fluorescence spectra subtracted.

One of the samples exposed to endonuclease in buffer η without magnesium and the one exposed in the presence of magnesium were examined by reverse phase HPLC. An octadecyl column heated to 55°C was used with a gradient of 0 to 20% buffer B over 20 minutes [buffers and other HPLC details given in section 2.2.2].

CHAPTER 3A

SYNTHESIS OF BASE ANALOGUE CONTAINING OLIGODEOXYNUCLEOTIDES

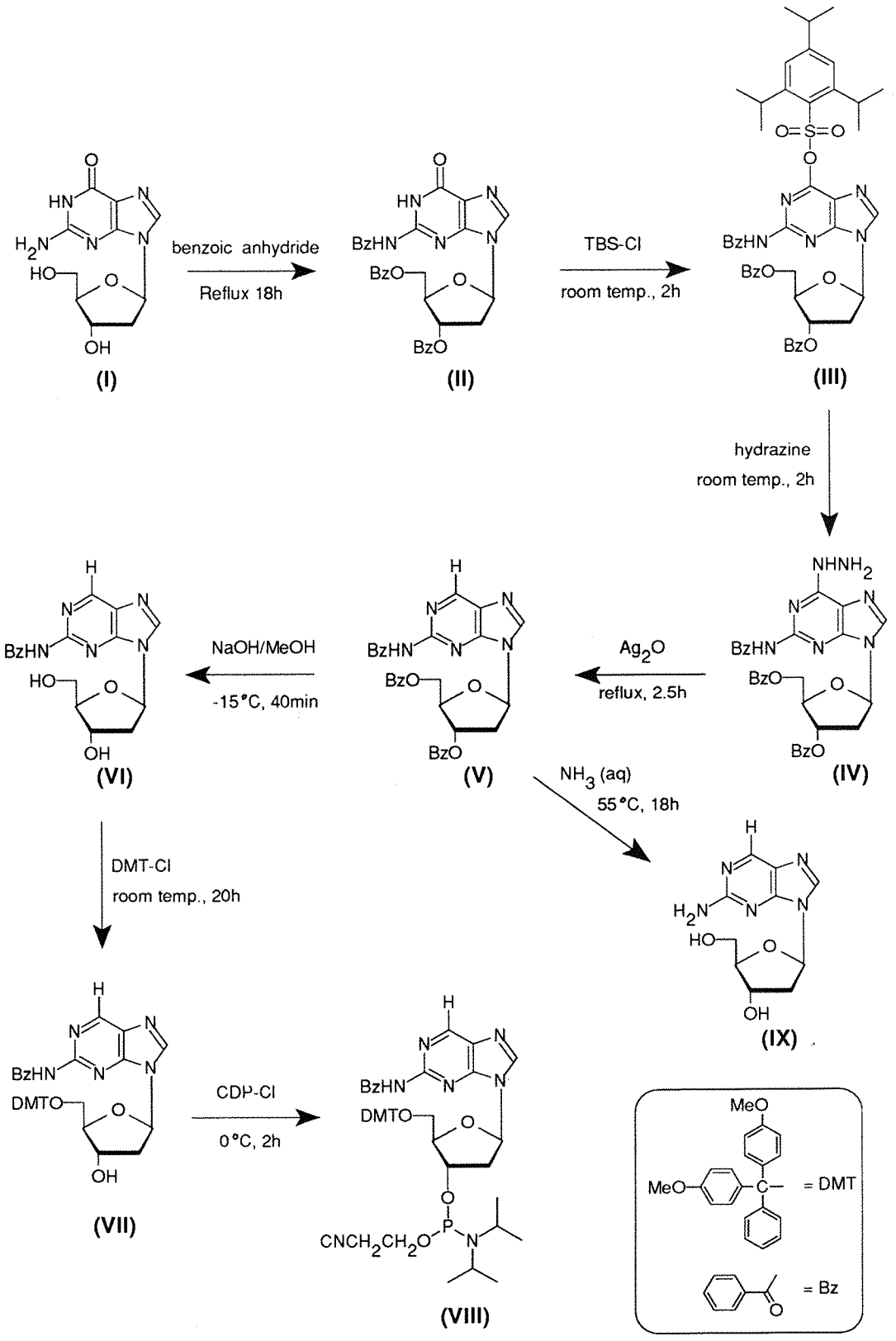
3a.1 SYNTHESIS OF BASE ANALOGUES

3a.1.1 Synthesis of the d[^{6H}G] Base Analogue

The synthesis of the protected phosphoramidite derivative of d[^{6H}G] is shown in Scheme 3a.1. The method used is based on that published by McLaughlin *et al.* (1988) and only a few minor alterations were employed (Connolly, 1991). The first step involved the protection of the exocyclic amino group and the sugar hydroxyls. This was done by reacting dry 2'-deoxyguanosine (I) with benzoic anhydride under anhydrous reflux conditions. Following the reaction by TLC revealed that the starting material was converted to another compound within about two hours and that this compound was converted to the desired tribenzoyl derivative (II) after about a further sixteen hours. The intermediate product was presumably 3',5'-*O*-dibenzoyldeoxyguanosine with the less reactive amino group taking longer to benzoylate. In the next step, the 6-oxygen was efficiently converted to the 6-(2,4,6-triisopropylbenzene-sulphonyl) derivative (III) by reaction at room temperature with the appropriate benzenesulphonyl chloride. The presence of the benzenesulphonyl group is known to labilise the 6-position of deoxyguanosine to substitution reactions (Gaffney & Jones, 1982a & b; Gaffney *et al.*, 1984) and in the next step of the reaction scheme it is displaced by hydrazine to give the hydrazino derivative (IV). This product was used in the next step after the work-up procedure but without further purification. By not isolating the hydrazino compound (IV) the yield for the two steps III → IV → V was improved compared with the published yield. This was probably because hydrolysis of the unstable hydrazino compound occurs during the work-up and purification procedures resulting in conversion back to tribenzoyl deoxyguanosine in a similar manner to that reported for some 4-hydrazino pyrimidine compounds (Gildea & McLaughlin, 1989). The oxidation of the hydrazino function with silver I oxide in the presence of water produced the required 2-aminopurine base in compound V. The oxidation of hydrazino derivatives of thymidine, cytidine and deoxycytidine with silver I oxide has also been reported (Cech & Holy, 1977; Gildea & McLaughlin, 1989) and the general mechanism follows the scheme given below, with the formation of nitrogen being the driving force:-

$$\text{R-NH-NH}_2 + \text{Ag}^{\text{I}}_2\text{O} \rightarrow \text{R-N=NH} + 2\text{Ag}^0 + \text{H}_2\text{O} \rightarrow \text{R-H} + \text{N}_2(\uparrow)$$

The next three steps produced the analogue in a form suitable for phosphoramidite based DNA synthesis. Firstly the two hydroxy benzoyl groups were removed by mild alkaline hydrolysis. Performing this reaction with aqueous sodium hydroxide at -15°C limited the reaction to the removal of just the 3'- and 5'-*O*-benzoyls leaving the *N*-benzoyl group intact. A 5'-dimethoxytrityl group was then introduced in the usual manner by reacting VI with dimethoxytrityl chloride at room temperature (Gait, 1984). Finally VII was reacted with 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite to introduce the 3'-phosphoramidite functionality (VIII) required for the DNA synthesis coupling reaction. Purification of the final



Scheme 3a.1: Reaction scheme for the synthesis of the 2-aminopurine deoxynucleoside phosphoramidite. TBS-Cl = 2,4,6-triisopropylbenzenesulphonyl chloride; CDP-Cl = 2-cyanoethyl, *N,N*-diisopropylaminochlorophosphoramidite.

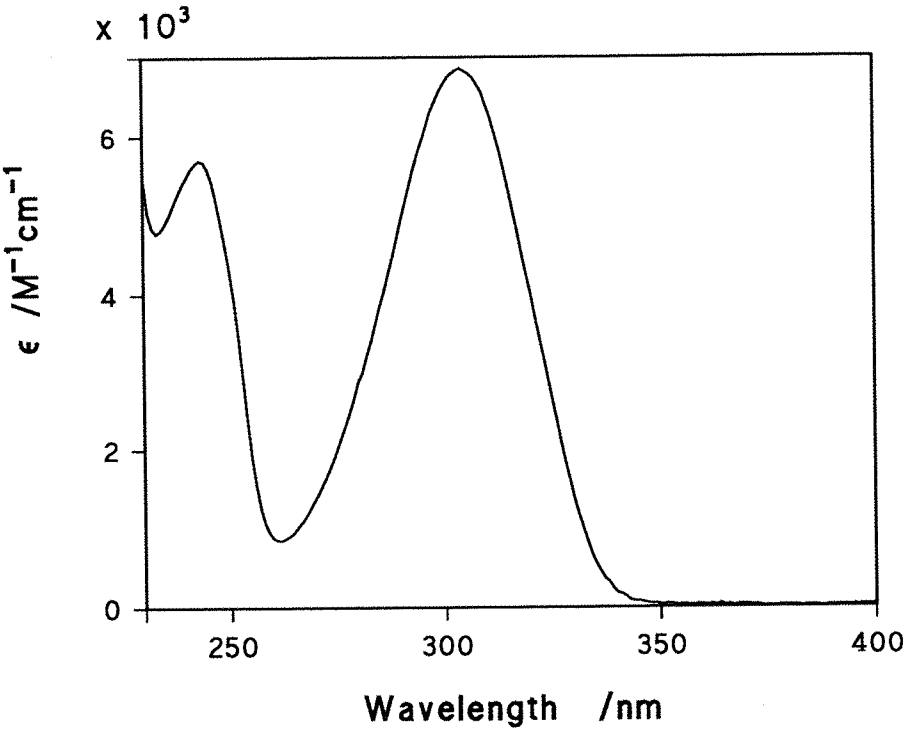


Figure 3a.1: UV spectrum of 2-aminopurine deoxynucleoside at pH7.2.

phosphoramidite derivative was by flash chromatography on silica gel, collecting both of the diastereoisomers which were resolvable on the flash column. In this and all the work-up steps of VIII it was necessary to keep the solutions basic in order to prevent acid catalysed activation of the P–N bond and subsequent hydrolysis. The column purified phosphoramidite evaporated down to a sticky yellow solid that was pure by TLC. To make the weighing out and the handling of the phosphoramidite easier, the sticky residue was precipitated out in petroleum spirit resulting in a white, free flowing solid.

For use as a standard and for characterisation purposes, some of the deprotected 2-aminopurine deoxynucleoside (IX) was prepared from the tribenzoyl protected derivative (V) by reaction with ammoniacal ethanol at 55°C (the ethanol helping to dissolve the tribenzoyl derivative). After overnight treatment, the deoxynucleoside was dried down and then washed with diethyl ether to remove benzamide. Recrystallisation of this product from methanol gave pure 2-aminopurine deoxynucleoside as judged by TLC. The proton NMR and FAB mass spectroscopic data of this compound were in agreement with the expected structure and also with the literature results. Particularly characteristic of the 2-aminopurine deoxynucleoside was the peak at δ8.61ppm in the ^{1}H NMR due to the C6 proton and the molecular ion peak at mass 251 in the FAB-MS spectrum. Also characteristic of the 2-aminopurine base was its UV spectrum which showed a peak at 304nm (Figure 3a.1). This is in agreement with a published spectrum of the 2-aminopurine riboside (Fox *et al.*, 1958). The 2-aminopurine deoxynucleoside was also characterised by its fluorescent spectral properties. Its excitation and emission spectra are given in Figure 3a.2. The excitation maximum of 302nm and the emission maximum of

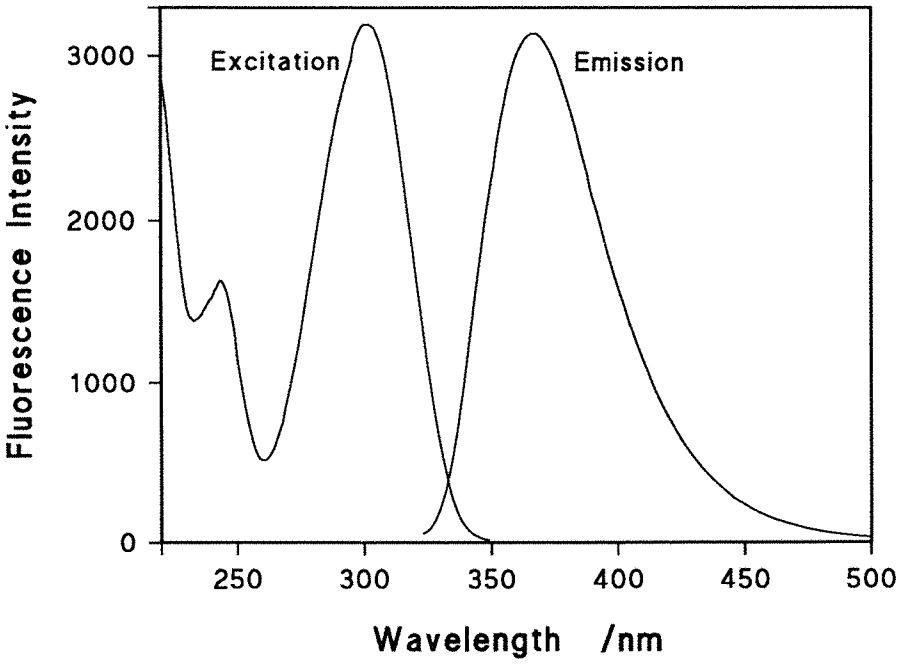


Figure 3a.2: Fluorescence spectrum of 2-aminopurine deoxynucleoside at pH7.2.

367nm are in close agreement with the published values of 303nm and 370nm respectively for the 2-aminopurine riboside (Ward *et al.*, 1969).

The 2-aminopurine base has been prepared previously by other methods including hydrogenation of the 6-chloro-deoxyguanosine compound (Milne & Townsend, 1974) and by reduction of the 6-thioguanosine riboside with Raney nickel (Fox *et al.*, 1958). The method used here (McLaughlin *et al.*, 1988) is a more convenient route to the 2-aminopurine deoxynucleoside and the overall yield of about 15% from compound II to the phosphoramidite (VIII) is satisfactory.

All of the intermediates except for 2-(*N*-benzoylamino)purine deoxynucleoside (VI) were characterised by ^{1}H NMR and FAB mass spectroscopy and gave data that was in agreement with published data (McLaughlin *et al.*, 1988). The phosphoramidite (VIII) was not characterised but was found to couple efficiently in the DNA synthesis of d(GAC[^{6H}G]ATATCGTC), as judged by the release of the dimethoxytrityl cation. The 2-*N*-benzoyl group provided adequate protection during the DNA synthesis and was removed along with the protecting groups of the other bases by overnight treatment with 35% ammonia at 55°C.

3a.1.2 Synthesis of the d[^{6S}G] Base Analogue

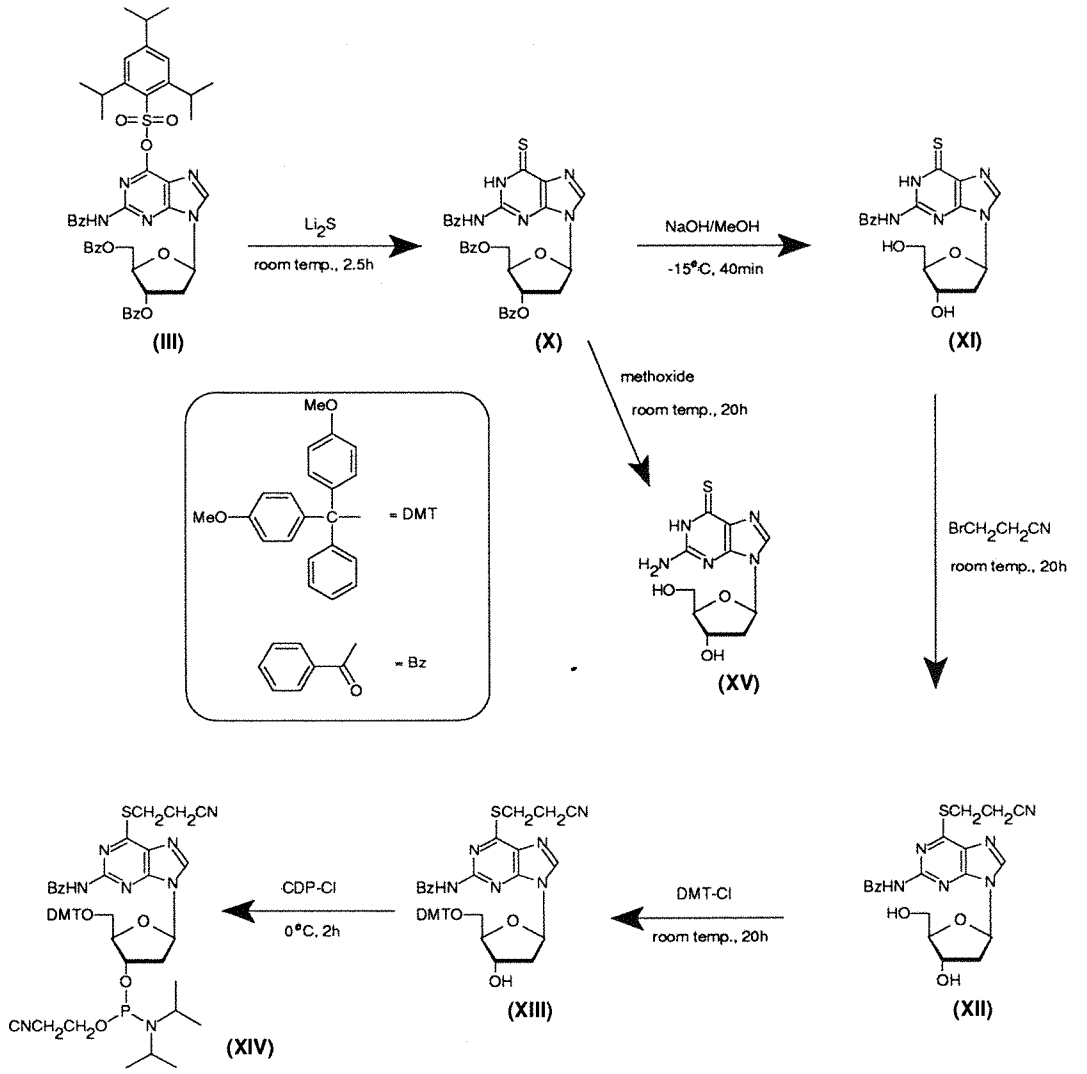
The synthesis of the 6-thiodeoxyguanosine analogue was achieved using a novel procedure (Waters & Connolly, 1991a; Connolly, 1991) which is outlined in Scheme 3a.2. The starting material was 2'-deoxyguanosine which was converted to the sulphonyl derivative (III) as described for the 2-aminopurine deoxynucleoside synthesis. After this the two syntheses diverged with the next step for the synthesis of 6-thiodeoxyguanosine being the introduction of the thio-function. As in the 2-aminopurine deoxynucleoside synthesis, the lability of the 6-(2,4,6-triisopropyl)benzenesulphonyl group to nucleophilic displacement was utilised, with direct displacement of the sulphonyl group by sulphide ion. This reaction is a novel method of introducing the 6-thio functionality and it was found to proceed cleanly and efficiently under mild conditions, yielding around 90% of pure product. This method has the advantages of being quicker and simpler than other previously published methods of producing 6-thiodeoxyguanosine which involve a glycosylation step (Ueda *et al.*, 1978; Hanna *et al.*, 1988) and also gives higher yields.

Since developing this reaction, two other methods of synthesising 6-thiodeoxyguanosine from deoxyguanosine have been reported (Kung & Jones, 1991; Xu *et al.*, 1991). They are both similar to the method described here in that they both involve displacement of a labile 6-functionality on deoxyguanosine by a sulphur nucleophile. Kung and Jones displaced a 6-pyridyl group with hydrosulphide ion whilst Xu *et al.* displaced a 6-mesitylene sulphonyl group with thioacetate ion.

In production of the required 6-thiodeoxyguanosine phosphoramidite the sugar benzoyl groups were removed in a limited alkaline hydrolysis reaction, and then the 6-thio group was protected with a 2-cyanoethyl group to give compound XII. This 2-cyanoethyl introduction was attempted using a method based on that of Johnston *et al.* (1958) for producing S-substituted 6-thiopurine derivatives. Various solvents, bases, reactants (*i.e.* acrylonitrile and 1-bromo-propionitrile) and stoichiometries were tried but the method described in section 2.1.5.iii was the only one that gave complete conversion of the starting material. All other reaction attempts stopped after about 50% conversion, but it must be pointed out that the conditions for this step have not been optimised.

The 5'-dimethoxytrityl group and the 3'-[(*N*,*N*-diisopropylamino)-2-cyanoethylphosphoramidite] functions were then introduced using established procedures (Gait, 1984) to give the desired phosphoramidite (XIV) in an overall yield of 17% from deoxyguanosine.

Some of the free 6-thiodeoxyguanosine deoxynucleoside was prepared for use as a standard and for characterisation purposes. This was achieved by removing the benzoyl groups from tribenzoyl 6-thiodeoxyguanosine (X) with methoxide treatment at room temperature. The 6-thiodeoxyguanosine produced in this way was of high purity but was purified further by reverse phase flash chromatography to give >99% pure 6-thiodeoxyguanosine as judged by reverse phase HPLC of the compound. Attempts to recrystallise the compound were not



Scheme 3a.2: Reaction scheme for the synthesis of the 6-thiodeoxyguanosine phosphoramidite. CDP-Cl = 2-cyanoethyl, *N,N*-diisopropylaminochlorophosphoramidite.

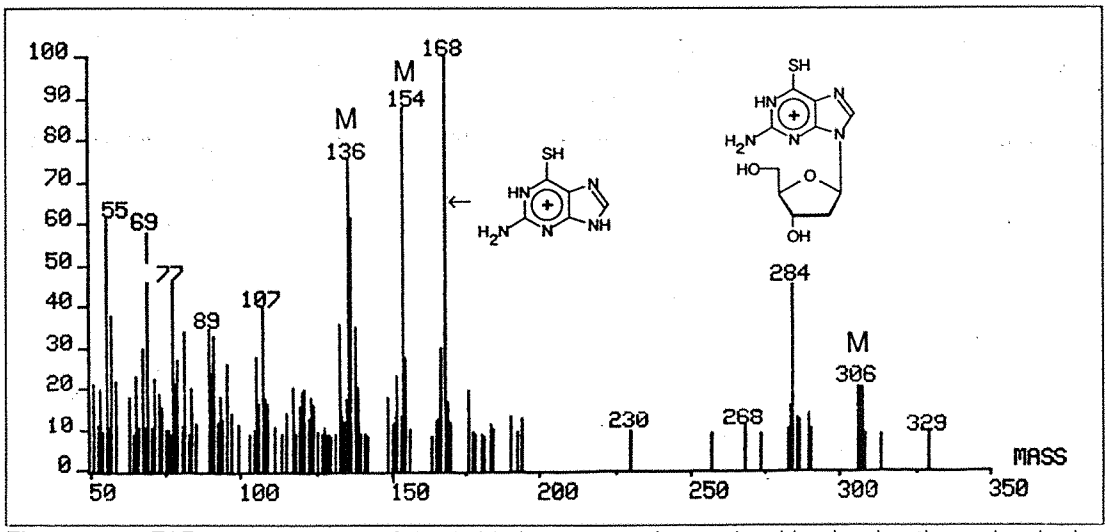


Figure 3a.3: FAB-mass spectrum of the 6-thiodeoxyguanosine nucleoside showing the molecular ion peak and one other identifiable fragment ion peaks. Peaks labelled M are derived from the matrix.

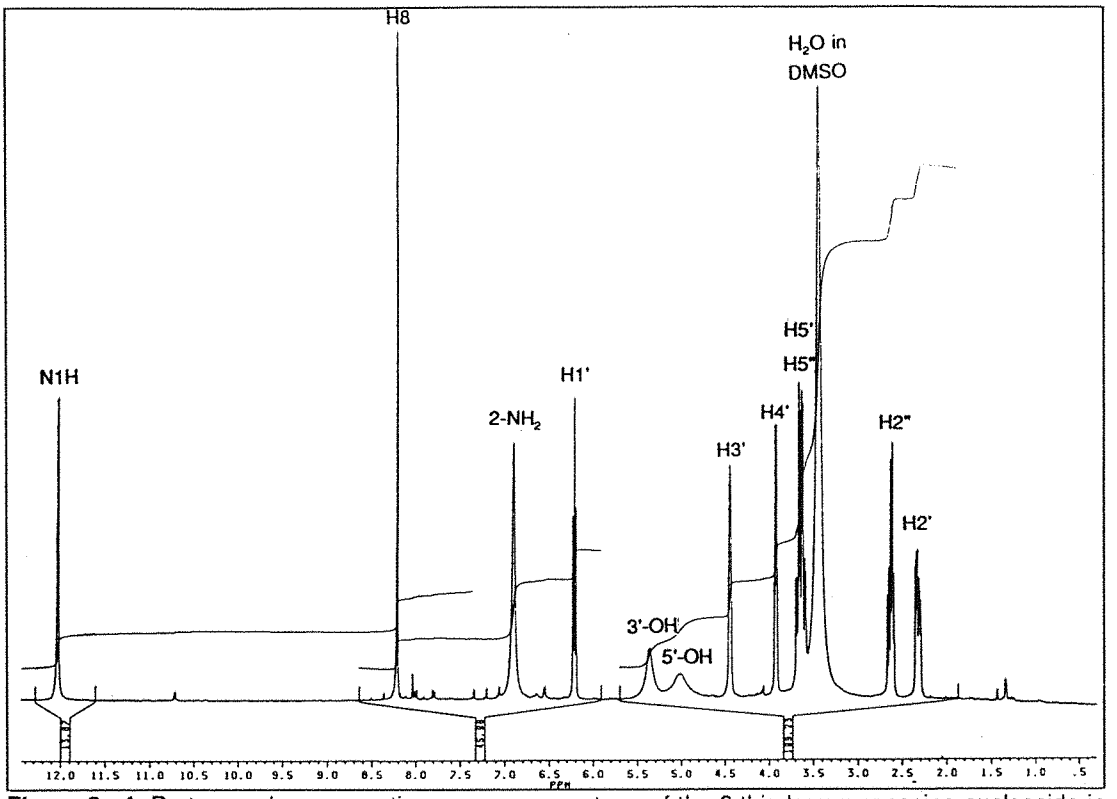


Figure 3a.4: Proton nuclear magnetic resonance spectrum of the 6-thiodeoxyguanosine nucleoside in d_6-DMSO. Peak assignments are as indicated.

successful because of the compounds poor crystallinity and also because of the susceptibility of the 6-sulphur to oxidation. The FAB-MS and ^{1}H NMR for the purified deoxynucleoside are given in Figures 3a.3 and 3a.4 respectively. The peaks in the ^{1}H NMR spectrum and the assignments given are in agreement with those published for 6-thiodeoxyguanosine synthesised by other routes (Hanna *et al.*, 1988; Kung & Jones, 1991). The broadness of the peaks assigned to the 2-amino, 3'-hydroxy and 5'-hydroxy protons is due to their constant exchange reaction with protons from a small amount of water present in the solvent. The peaks assigned to the 3'- and 5'-hydroxy protons were confirmed as exchangeable by their disappearance in the ^{1}H NMR spectrum of 2-*N*-benzoyl-6-thiodeoxyguanosine on shaking the sample with D_2O (exchangeable protons are replaced with deuterium from the D_2O and so do not show up in the ^{1}H NMR spectrum). The FAB mass spectrum shows the expected molecular ion peak at mass 284 and also a peak due to the heterocycle ion at mass 168 as shown in Figure 3a.3. The deoxynucleoside was further characterised by its UV spectrum. The UV spectra of 6-thiodeoxyguanosine at pH 4.0, 6.9 and 12.0 are shown in Figure 3a.5. The spectra exhibit four isobestic points and are virtually identical to the spectra of the 6-thioguanosine riboside (Fox *et al.*, 1958).

All other intermediates were characterised by ^{1}H NMR and FAB-MS and gave satisfactory data. The peak at ~340nm in the UV spectrum of 6-thiodeoxyguanosine at pH7 (Figure 3a.5) is characteristic of the C=S function and provides a means of checking for its presence after the sulphide substitution [step 2.1.5.i]. The UV spectrum of tribenzoyl 6-thiodeoxyguanosine (X) confirmed the presence of the 6C=S group by its λ_{max} at 333nm. When

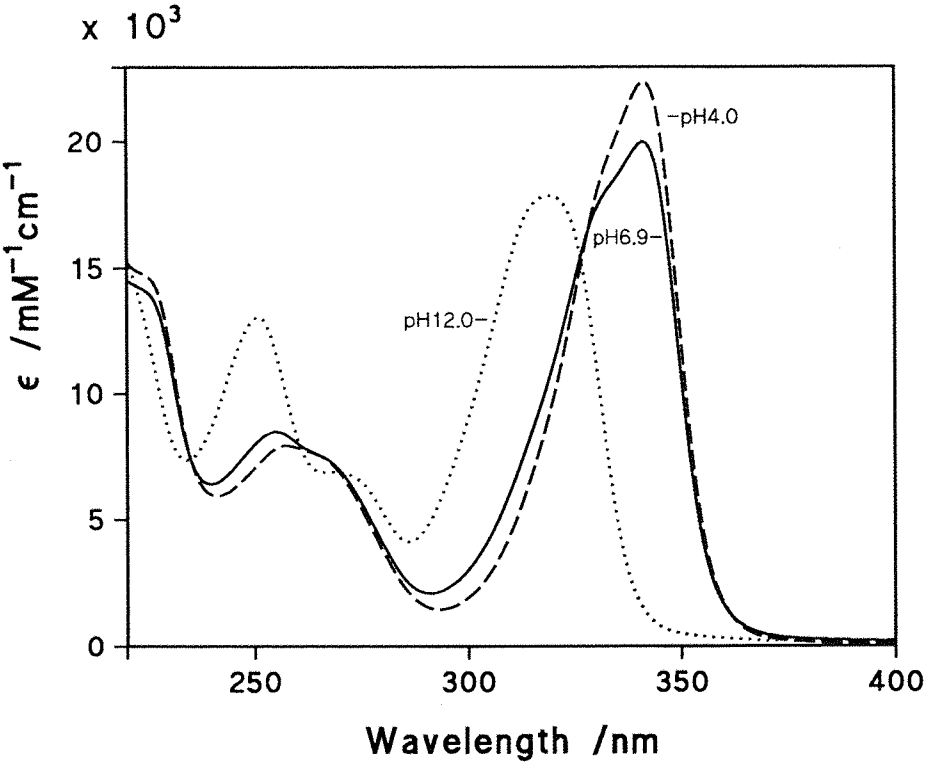


Figure 3a.5: UV spectrum of 6-thiodeoxyguanosine nucleoside at pH4.0 (---), pH6.9 (——) and pH12 (...).

the sulphur atom is not double bonded but is bound to both the ring 6-carbon and another group, such as when it is alkylated or as in acid solution when it is protonated (Figure 3a.5), this absorbance band undergoes a blue shift of about 30nm. The presence of the UV band at 293nm in compound XII therefore confirms the correct introduction of the cyanoethyl group at the sulphur.

The 6-thiodeoxyguanosine nucleoside was found in solution to have a tendency to be oxidised by air. This normally resulted in the formation of dimers linked by a disulphide bond which could be reduced back to the monomer by treatment with DTE at pH8. Formation of the disulphide could also be brought about by oxidation with iodine. On formation of the disulphide bond the sulphur is no longer double bonded and has a λ_{max} shift from 341nm for the monosulphide, to 320nm for the disulphide. The oxidation process could thus be observed by UV spectroscopy as demonstrated in Figure 3a.6. Further oxidation of the 6-thiodeoxyguanosine nucleoside was also inferred from the slow conversion to deoxyguanosine (probably via the 6-sulphonate) which occurred in 6-thiodeoxyguanosine solutions kept at room temperature over a number of days. This desulphurisation process also been observed to occur in 6-thiodeoxyguanosine mono- and tri-phosphates by Cowart & Benkovic (1991) who quote a t½ of four days at room temperature. For the 6-thiodeoxyguanosine produced here, no detectable deoxyguanosine was observed overnight at room temperature in aqueous solution although at 55°C approximately 25% conversion to deoxyguanosine was observed over four days.

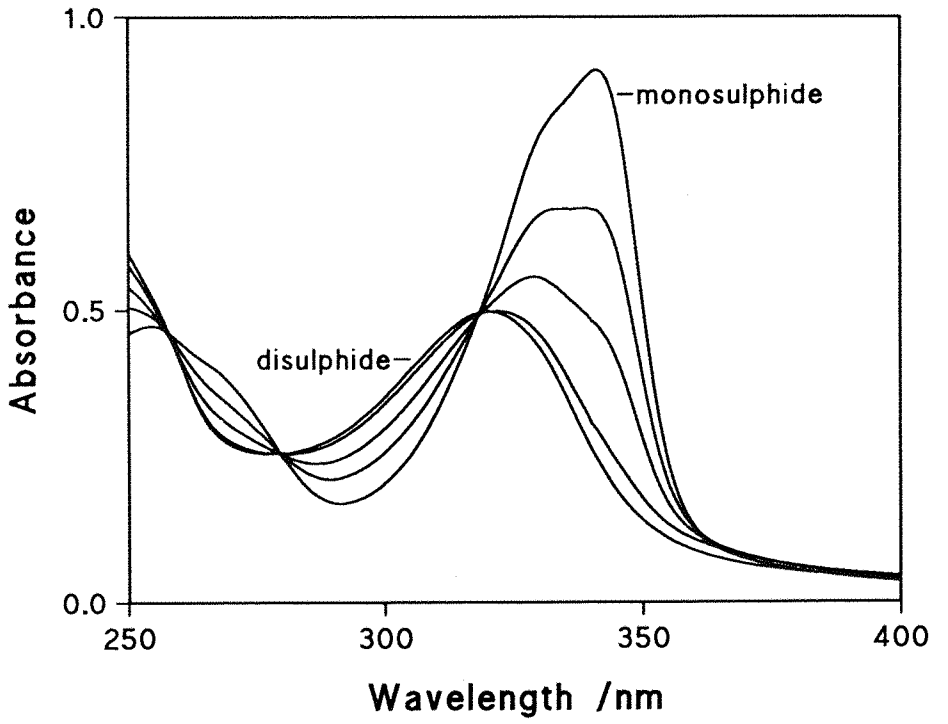
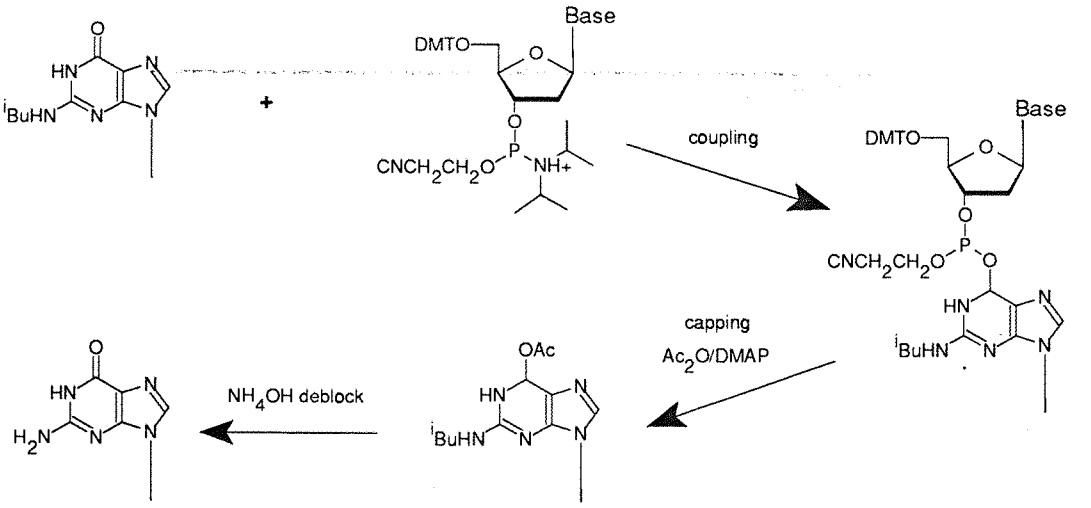


Figure 3a.6: UV spectra of the iodine oxidation of 6-thiodeoxyguanosine to its disulphide.

In order to incorporate the 6-thiodeoxyguanosine into an oligodeoxynucleotide by phosphite triester DNA synthesis, the 6-thio group needed to be protected. This was to prevent oxidation of the sulphur during the repeated iodine steps in the synthesis cycle and also to prevent addition of tetrazole activated phosphoramidites to the sulphur during the coupling step [see section 1.2.2]. This latter reaction is known to occur to a small extent with deoxyguanosine as outlined in Scheme 3a.3. The addition product of deoxyguanosine is not, however, a serious problem since it is converted to the 6-acetylated derivative during the capping step which is subsequently converted back to the 6-carboxyl function during the ammonia deblock procedure. Since the sulphur is a better nucleophile it was expected that phosphoramidite addition to the 6-thio group of 6-thiodeoxyguanosine would present a problem and so it was decided to protect it. In the only other previously reported method of synthesising 6-thiodeoxyguanosine containing oligodeoxynucleotide, the 6-thio group was protected as the 6-mesitylenesulphonyl derivative (Rappaport, 1988). Initial attempts were made at DNA synthesis with a 6-S-sulphenylmethyl (6-S-SCH_3) protected phosphoramidite which has been found to be suitable for the sulphur protection of 4-thiothymidine during DNA synthesis (Connolly & Newman, 1989). The 6S-SCH_3 protected phosphoramidite of 6-thiodeoxyguanosine was, however, found not to couple in oligodeoxynucleotide synthesis. Attempts were then made to protect the 6-thio function with the 2-cyanoethyl protecting group which has been suggested as a suitable protecting group for the 6-carbonyl function of deoxyguanosine for DNA synthesis



Scheme 3a.3: Reaction of the 6-carbonyl of deoxyguanosine during DNA synthesis.

(Gaffney & Jones, 1982b). The 2-cyanoethyl protected 6-thiodeoxyguanosine phosphoramidite (XIV) coupled efficiently in the DNA synthesis of d(GAC[^{6S}G]ATATCGTC) as judged by the released dimethoxytrityl cation. The 2-cyanoethyl group could be easily removed and was found to have successfully preserved the thio-function during the DNA synthesis. In a recent publication, cyanoethyl protected 6-thiodeoxyguanosine has also been used successfully in DNA synthesis (Christopherson & Broom, 1991). The cyanoethyl group is removed by treatment with alkali and in order to optimise the deblock procedure, the kinetics of the removal of the benzoyl and cyanoethyl protecting groups in 35% aqueous ammonia at 55°C were investigated, as described in 2.1.5.vii. The results are shown graphically in Figure 3a.7. The reaction appeared to follow an A → B → C mechanism with the cyanoethyl group of the starting material (A in the mechanism) being removed first to give 2-*N*-benzoyl-6-thiodeoxyguanosine (B) followed by removal of the benzoyl group to give 6-thiodeoxyguanosine (C). No other compounds were detected and so the formation of S-(2-cyanoethyl)-6-thiodeoxyguanosine by removal of the benzoyl group from the starting material does not occur. Although the data are somewhat crude, the points agree reasonably well with a best-fit curve based on the kinetics of an A → B → C reaction (Frost & Pearson, 1961) as shown in the figure. The curves were obtained by varying the values k_1 and k_2 (the rate constants for the first and second reactions respectively) until the total squared error between the curves and the data points was minimised. The values for the two constants were identical at about $2s^{-1}$. More importantly for the present application, the graph shows that 97% of the deoxynucleoside is completely deblocked after three hours without any other detectable products. However, a fourth peak started to appear after four hours and increased with time. This compound coeluted with 2,6-diaminopurine deoxynucleoside and was found to have an identical UV spectrum (λ_{max} at 255nm and 279nm; Ueda *et al.*, 1978). It therefore appears that the 6-thiodeoxyguanosine analogue undergoes a similar substitution of sulphur by ammonia as that seen for 4-thiothymidine (Connolly & Newman, 1989). This is in contrast to the results of Christopherson and Broom (1991) who detected approximately 25% conversion of 6-thiodeoxyguanosine to deoxyguanosine after four

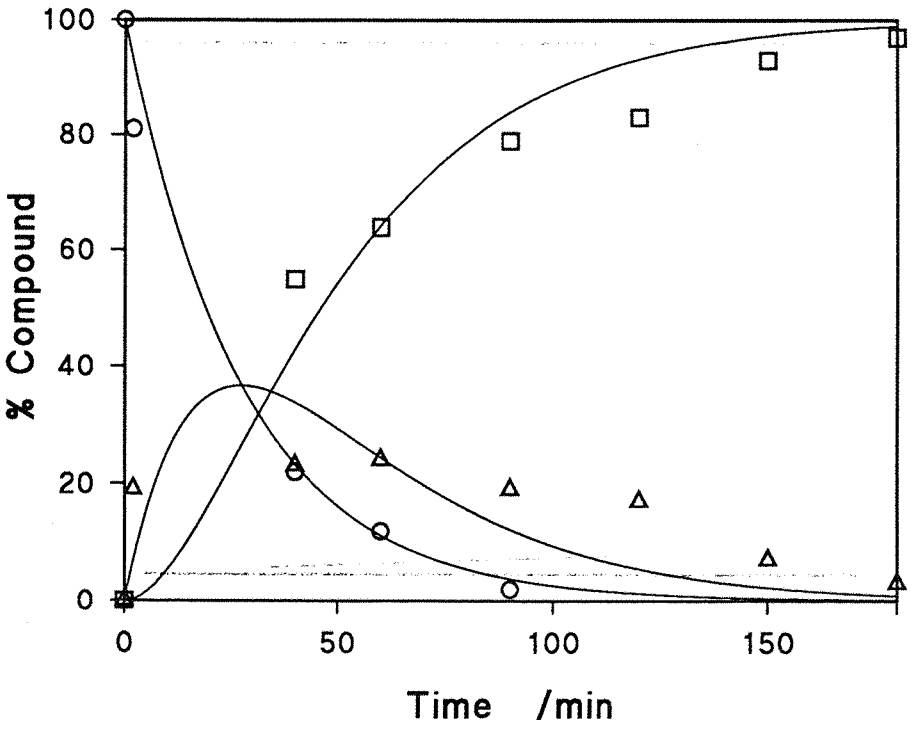


Figure 3a.7: Ammonia deblock kinetics of 2-*N*-benzoyl-S-(2-cyanoethyl)-6-thiodeoxyguanosine (○) going to 6-thiodeoxyguanosine (□) via 2-*N*-benzoyl-6-thiodeoxyguanosine (△).

hours treatment of a protected d([^{6S}G]T) dinucleotide derivative with concentrated ammonium hydroxide at room temperature. To maximise the yield of deblocked 6-thiodeoxyguanosine containing oligodeoxynucleotide, with minimum formation of 2,6-diaminopurine deoxynucleoside containing oligodeoxynucleotide, a time of four hours was chosen for the ammonia treatment of the oligodeoxynucleotide. At this time in the experiment no unblocked 6-thiodeoxyguanosine was detected and there was only about 1% contamination with 2,6-diaminopurine deoxynucleoside.

3a.2 SYNTHESIS AND PURIFICATION OF OLIGODEOXYNUCLEOTIDES CONTAINING BASE ANALOGUES

The oligodeoxynucleotides listed below were synthesised on a one μmol scale using an automated DNA synthesizer:-

d(GACGATATCGTC) d(TGACGATATCGTC) d(GAC[^{6S}G]ATATCGTC)
d(GAC[^{6H}G]ATATCGTC) d(GAC[^{3C}G]ATATCGTC) d(GAC[^{7C}G]ATATCGTC)
d(GAC[I]ATATCGTC) d(GACGATAT[^{5Me}C]GTC) d(GACGATAT[^{4H}C]GTC)
d(GTCGACGATATCGTCGAC) d(GTCGAC[^{6H}G]ATATCGTCGAC)

All base analogue phosphoramidites coupled successfully in the DNA synthesis in yields of >98%, as judged by the concentration of the released orange dimethoxytrityl cation, without the need to alter the normal cycle program of the automated DNA synthesizer.

3a.2.1 Deblocking of Oligodeoxynucleotides

The deblocking of all of the oligodeoxynucleotides, except for the d[^{6S}G] and d[^{4H}C] containing oligodeoxynucleotides, was by the usual overnight treatment at 55°C with 35% ammonia. With the d[^{6S}G] containing oligodeoxynucleotide the deblock was limited to four hours for the reasons discussed above in section 3a.1.2.

The d(GACGATAT[^{4H}C]GTC) dodecamer was found to be unstable to the conditions of ammonia deblock, as shown in Figure 3a.8, being converted into a compound designated X. The base composition of X by digestion with snake venom phosphodiesterase and nuclease P1 [see section 2.2.7.i] was found to be C_2, G_3, T_3, and A_3 with no detectable d[^{4H}C]. It therefore appeared that X was produced from d(GACGATAT[^{4H}C]GTC) by the modification of the d[^{4H}C] base by ammonia to give d(GACGATAT[?]GTC). The fact that no fifth base corresponding to the altered d[^{4H}C] base was detected in the digest of X could be due to the UV chromophore of d[^{4H}C] being destroyed or it could be that the altered base has been lost altogether by depyrimidation. The X oligodeoxynucleotide was found to react further at 55°C in aqueous ammonia to give two more peaks labelled Y and Z in Figure 3a.9 (presumably by chain cleavage). Treatment of Y with alkaline phosphatase converted it into a different oligodeoxynucleotide, showing that Y had a 5'-phosphate. The base composition of Y was found to be C_1, G_1 and T_1. Digestion of Z with snake venom alone gave the four normal deoxynucleoside HPLC peaks plus a fifth peak, Q, in the ratio C_1, G_2, T_1, A_3 and Q_1. Treatment of the digestion mixture with nuclease P1 converted Q into thymidine giving two equivalents of thymidine. All the data can be explained by the mechanism shown in Scheme 3a.4, which is modified from that given for a d[^{4H}T] containing oligodeoxynucleotide by Connolly and Newman (1989). The d[^{4H}C] base is converted to a saturated pyrimidine by addition of ammonia across the 5,6-double bond giving X. Pyrimidine bases are known to be saturated by strong nucleophiles such as hydrazine, but do not normally react with ammonia (Kochetkov & Budovski, 1972). The substitution of the 4-NH_2 of deoxycytidine (or the 4C=O group of thymidine) by a hydrogen labilises the 5,6-double bond to such addition reactions. This is because the 4-NH_2 group releases electrons into the ring system of the pyrimidine. This helps to counteract the π-electron deficiency of the pyrimidine ring, caused by the electron withdrawing effect of the ring nitrogens, and makes the pyrimidine more aromatic. Removal of the 4-NH_2 group thus makes the system less aromatic and so the 5,6-double bond moves towards isolated non-aromatic character and is therefore more easily attacked by nucleophiles. The rate of the reaction for d[^{4H}C] was approximately fifteen times faster than that for d[^{4H}T]. This is probably because of the steric hindrance that the 5-methyl group of d[^{4H}T] presents to

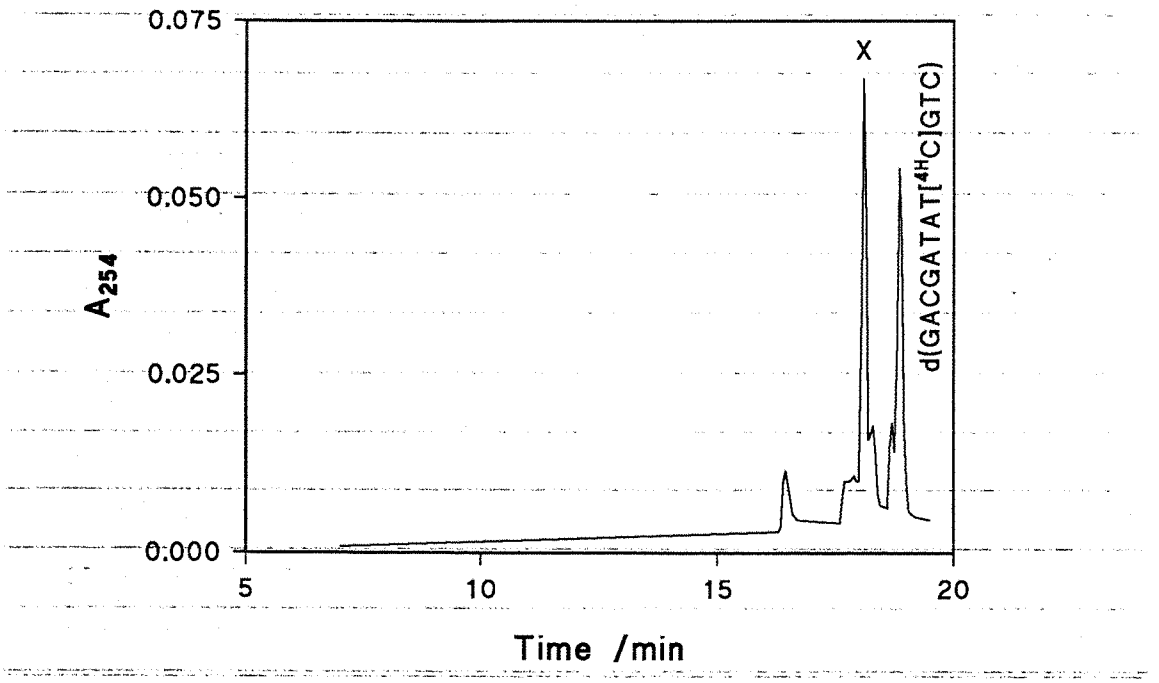


Figure 3a.8: HPLC trace of the reaction of d(GACGATAT[^{4H}C]GTC) with ammonia, producing compound X (gradient is 0-12% buffer B over 16 minutes).

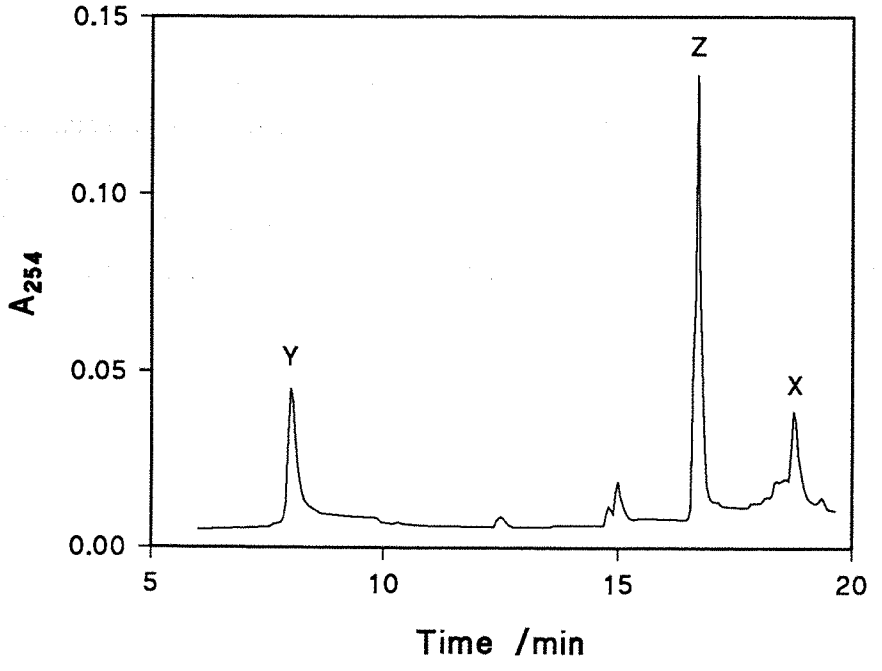
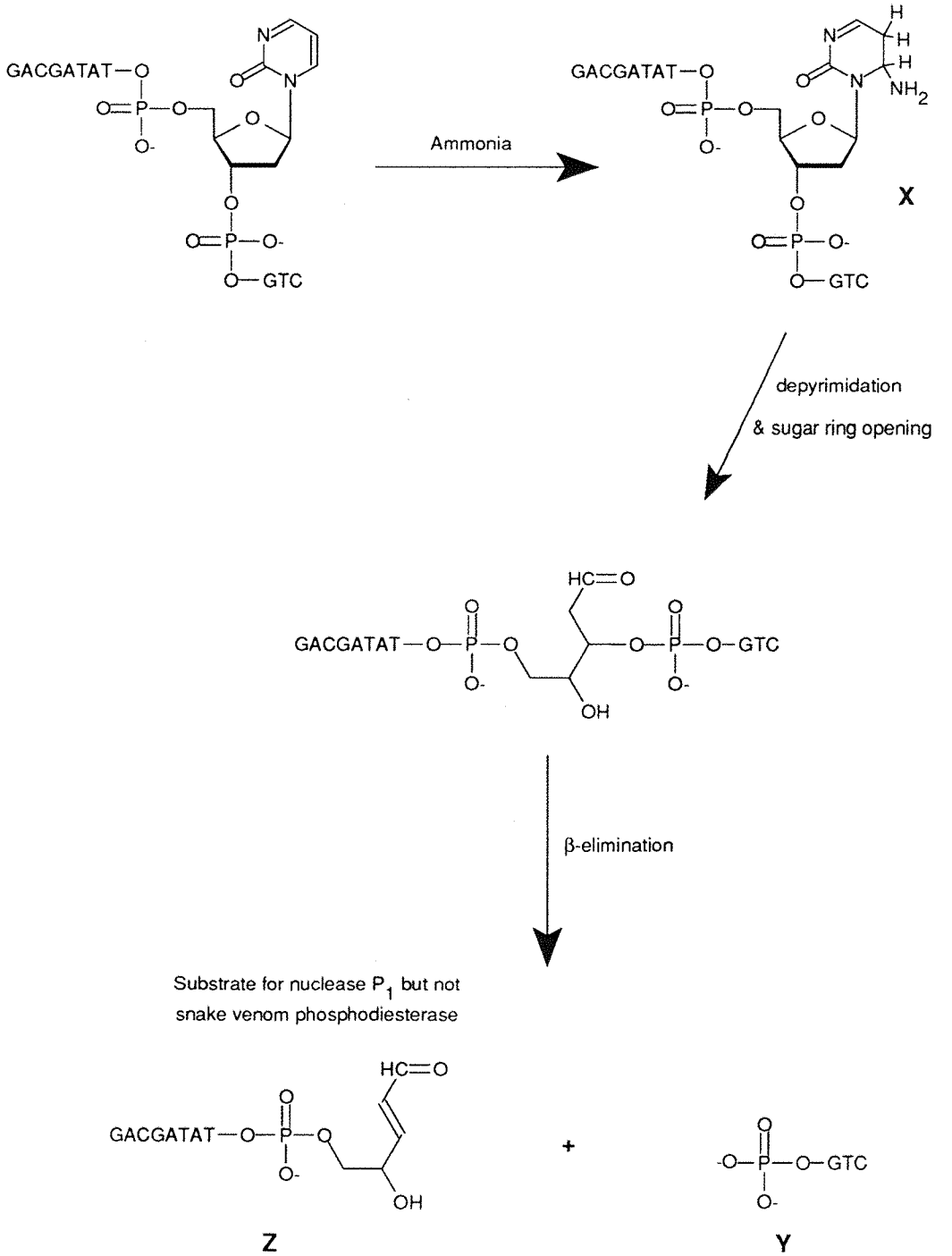


Figure 3a.9: HPLC trace of the reaction of compound X with ammonia, producing Y and Z (gradient is 0-12% buffer B over 16 minutes).

the incoming NH_3 nucleophile. The saturated base would not be detected by UV absorbance of the HPLC eluent in the digest of X since its chromophore has been destroyed. The oligodeoxynucleotides Y and Z are produced by cleavage of X following depyrimidation of the altered d[^{4H}C] base. Y has a 5'-phosphate which can be removed by alkaline phosphatase and Z has a 3'-phosphate derivative on its terminal thymidine. When Z was digested with snake venom phosphodiesterase the 3'-phosphate derivative on the terminal thymidine was not a substrate for this enzyme. The terminal thymidine with the 3'-phosphate moiety attached thus



Scheme 3a.4: Reaction scheme for the reaction of d(GACGATAT[^{4H}C]GTC) with concentrated aqueous ammonia during the deblock step.

gives rise to the fifth peak in the digestion. Nuclease P1 does however cleave this thymidine derivative and so was able to cleave the 3'-phosphate group from the terminal thymidine.

To try and minimise the extent of reaction during the ammonia deblock of d(GACGATAT[^{Ac}C]GTC), the dodecamer was synthesised using FOD phosphoramidites. This enabled the ammonia deblock step to be limited to eight hours at room temperature.

3a.2.2 Purification of Oligodeoxynucleotides

All of the oligodeoxynucleotides were purified with their 5'-dimethoxytrityl groups on, by reverse phase HPLC. Retaining the 5'-DMT group provides a useful route for the separation of the required DNA sequence from all the failure sequences. Since only the correct sequence has the hydrophobic 5'-dimethoxytrityl group, only it binds strongly to a reverse phase HPLC column, whereas all the failure sequences are much more easily eluted because they lack the dimethoxytrityl group. A typical analytical trace of a 'DMT-on' oligodeoxynucleotide is shown in Figure 3a.10. Preparative HPLC was performed by injecting samples of about 100nmol and then collecting fractions containing the pure oligodeoxynucleotide. After drying down the samples *in vacuo* and removing the 5'-terminal DMT group by acid treatment, the oligodeoxynucleotides were then viewed again by reverse phase HPLC.
Normally oligodeoxynucleotides were >95% pure at this stage as the example in Figure 3a.11 shows and were further HPLC purified without any problems as described in section 2.2.5.

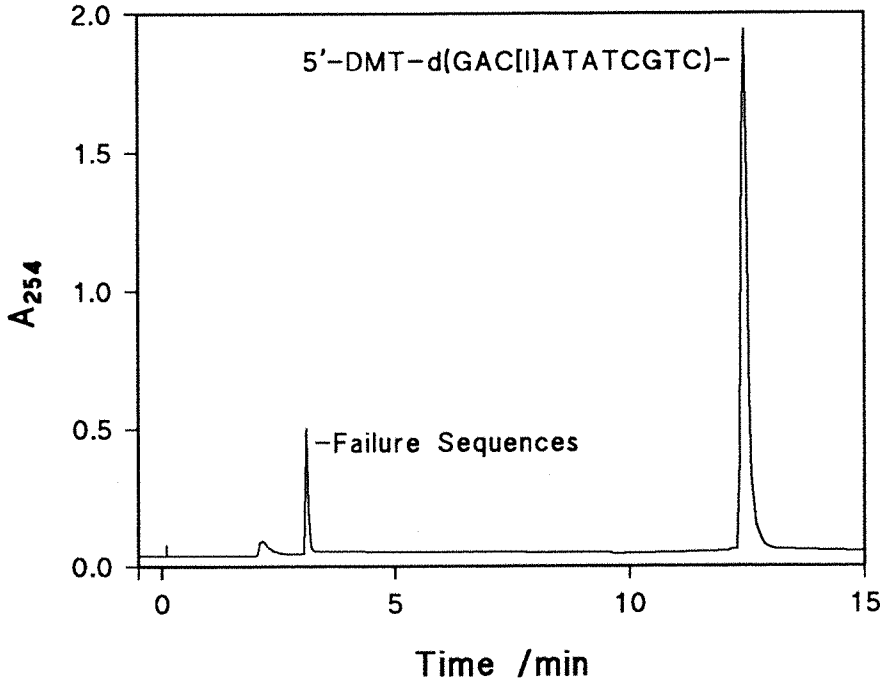


Figure 3a.10: HPLC trace of crude 5'-DMT-d(GAC[I]ATATCGTC) with a gradient of 25-60% buffer B over 15 minutes.

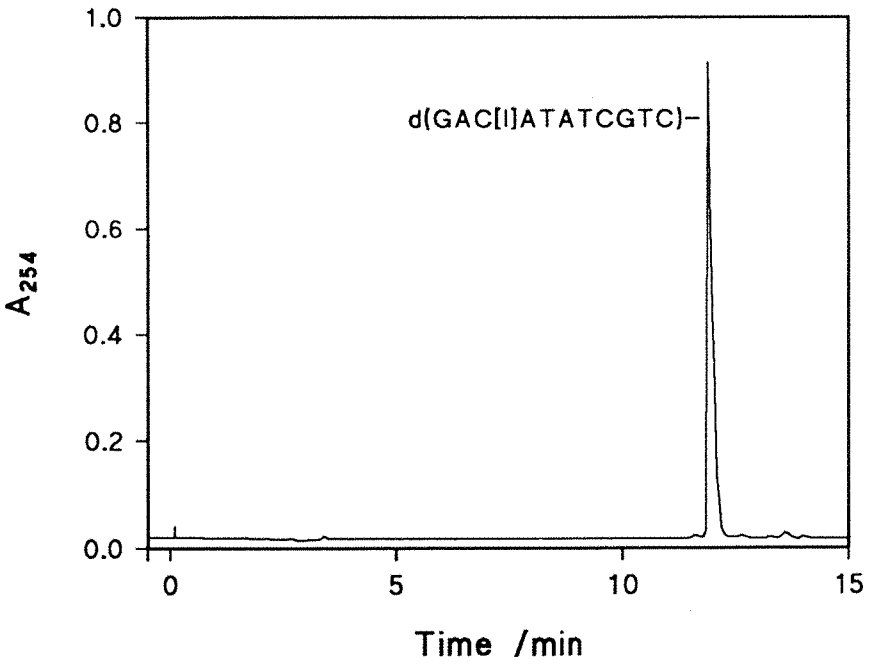


Figure 3a.11: HPLC trace of 'DMT-on' purified d(GAC[I]ATATCGTC), after detritylation, using a gradient of 5-15% buffer B over 15 minutes.

However, there were contaminant peaks in the traces of d(GACGATAT[^{4H}C]GTC) and d(GAC[^{6S}G]ATATCGTC). The trace of DMT-off d(GACGATAT[^{4H}C]GTC) has two contaminating peaks as shown in Figure 3a.12. These correspond to the two oligodeoxynucleotides designated X and Z and discussed in section 3a.2.1. The parent dodecamer d(GACGATAT[^{4H}C]GTC) and the two oligodeoxynucleotides X and Z would all have borne a 5'-dimethoxytrityl group and so were collected together during the previous HPLC purification. The other ammonia product of d(GACGATAT[^{4H}C]GTC), oligodeoxynucleotide Y, would have lacked the dimethoxytrityl group and would have eluted with the HPLC breakthrough peak. These ammonia deblock products of the d[^{4H}C] oligodeoxynucleotide were produced despite the use of FOD phosphoramidites and the milder ammonia conditions they allowed. This again shows that the d[^{4H}C] base is more reactive than the d[^{4H}T] base. Using the isocratic HPLC gradient described in the experimental section, d(GACGATAT[^{4H}C]GTC) was purified from the other contaminants.

The HPLC trace of d(GAC[^{6S}G]ATATCGTC) shown in Figure 3a.13 has three peaks. The main peak was found to be the desired dodecamer and the last peak to elute was found to be its disulphide. The smaller peak eluting immediately before the main peak was found, by co-elution with a standard, to be the control sequence, d(GACGATATCGTC). This peak was often found to vary in size even from injections of the same sample. This appears to be an effect caused by some of the d(GAC[^{6S}G]ATATCGTC) oligodeoxynucleotide co-eluting with the small amount of contaminant d(GACGATATCGTC). An isocratic gradient of 8% buffer B was found to successfully separate these two oligodeoxynucleotides without this misleading column

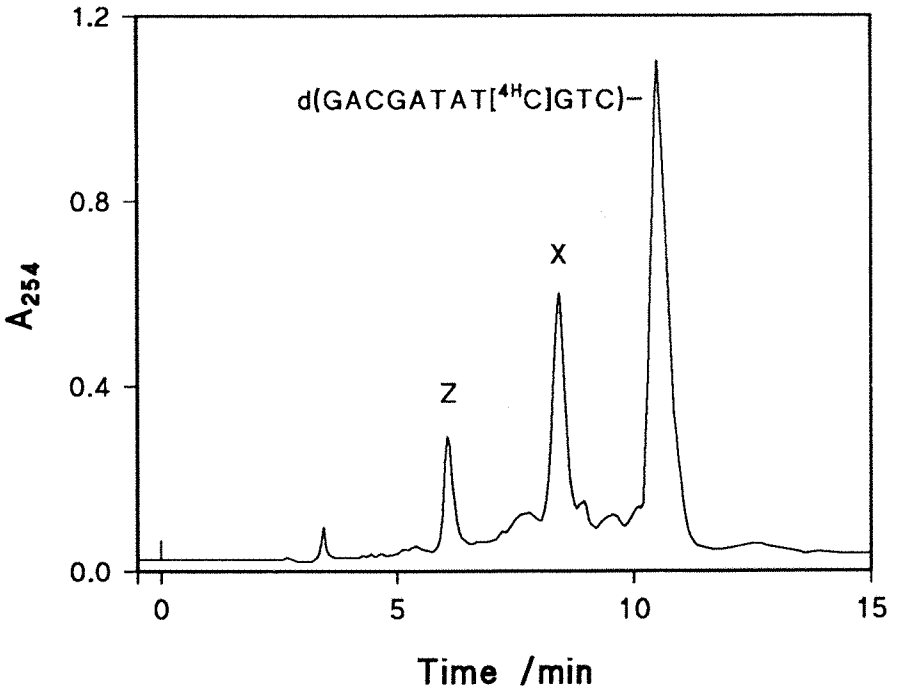


Figure 3a.12: HPLC trace of 'DMT-on' purified d(GACGATAT[^{4H}C]GTC), after detritylation, using an isocratic gradient of 7% buffer B.

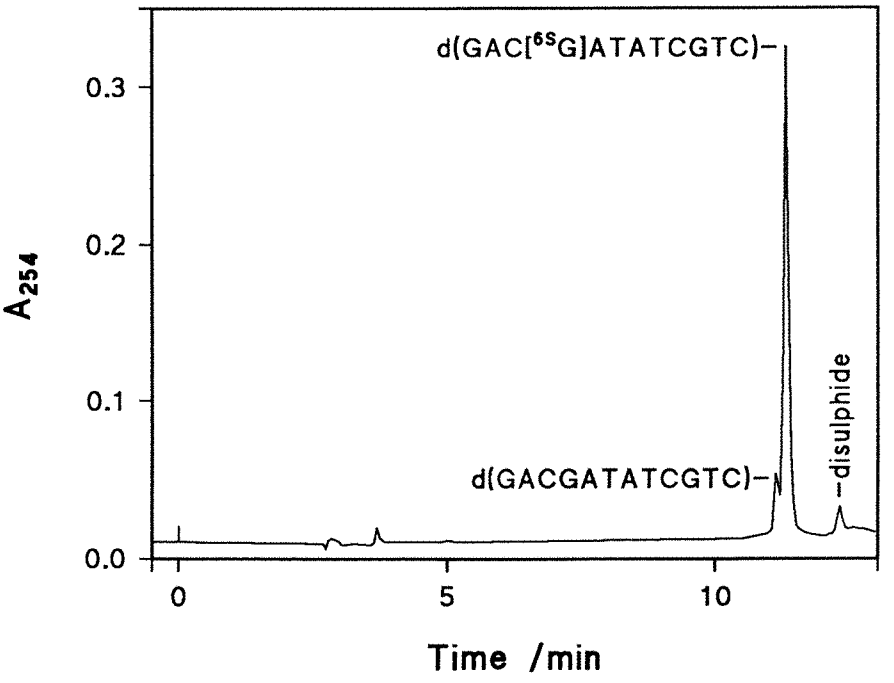


Figure 3a.13: HPLC trace of 'DMT-on' purified d(GAC[^{6S}G]ATATCGTC), after removal of the DMT group, using a gradient of 0-40% buffer B over 20 minutes.

effect and showed that the control oligodeoxynucleotide was generally present as less than 5% of the total UV absorbance.

There was no detectable 2,6-diaminopurine containing oligodeoxynucleotide in the HPLC trace of the detritylated d[^{6S}G] oligodeoxynucleotide, and so the deblock time of four hours seemed to have worked well. The oligodeoxynucleotides tendency to form disulphide dimers was found to be greater than that of the free deoxynucleoside but this could be minimised by leaving solutions of the oligodeoxynucleotide at room temperature for as little time as possible. When necessary the disulphide could be converted back to the monomer by treatment with DTE. Solutions of d(GAC[^{6S}G]ATATCGTC) were not kept permanently in the presence of DTE but were treated prior to HPLC purification. Also, to improve peak definition, an HPLC gradient that was different from usual was used. In this way, d(GAC[^{6S}G]ATATCGTC) was purified successfully in high yield and separated from the majority of the control oligodeoxynucleotide.

Using these protocols all the DMT-off oligodeoxynucleotides were purified by HPLC to give oligodeoxynucleotides that were all >98% pure as judged by HPLC (although the d[^{6S}G] oligodeoxynucleotide again showed the presence of a disulphide peak).

Finally all the oligodeoxynucleotides were desalted by reverse phase HPLC using volatile triethylammonium bicarbonate buffers as described in section 2.2.6. In the case of the d[^{6S}G] oligodeoxynucleotide, it was again treated with DTE just prior to HPLC injection and

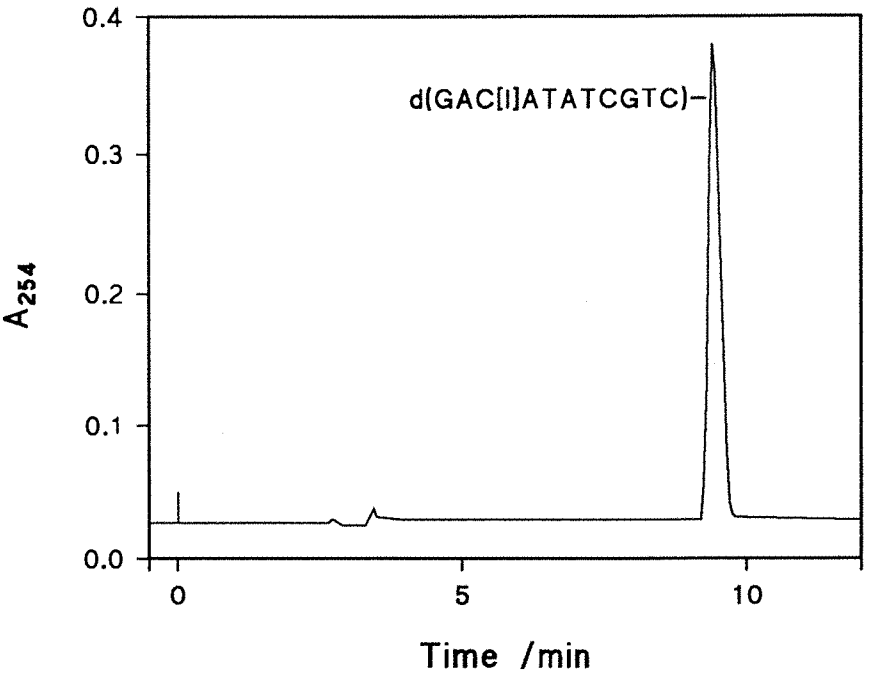


Figure 3a.14: HPLC trace of pure, desalted d(GAC[I]ATATCGTC) using a gradient of 5-15% buffer B over 15 minutes.

a shallower HPLC gradient was used in order to separate the product from contaminant control oligodeoxynucleotide. Pure desalted oligodeoxynucleotides were produced in this manner and all, with the exception of the d[^{6S}G] oligodeoxynucleotide, had no contamination detectable by HPLC as shown by the example HPLC trace in Figure 3a.14. The d[^{6S}G] oligodeoxynucleotide did contain some control oligodeoxynucleotide but this was less than 1% of the total 254nm absorbing material. The control oligodeoxynucleotide in this case was probably produced by a small amount of desulphurisation of d(GAC[^{6S}G]ATATCGTC), but the level of contamination was so low it did not present a problem.

3a.2.3 Phosphorylation of Oligodeoxynucleotides

All oligodeoxynucleotides that were substrates for the endonuclease, with the exception of d(TGACGATATCGTC) and the control octadecamer, were phosphorylated using T4 polynucleotide kinase as described in section 2.2.8. The method used was that of Sambrook *et al.* (1989) and was found to be most conveniently performed using oligodeoxynucleotides that had been HPLC purified DMT-off but were not desalted. The ability to carry out the reaction before the desalting step meant that less oligodeoxynucleotide was wasted since an extra HPLC purification step, which would inevitably result in loss of some of the oligodeoxynucleotide, was avoided. T4 polynucleotide kinase is known to be strongly inhibited by ammonium ions (Sambrook *et al.*, 1989) but the high concentrations of triethylammonium ions expected to be

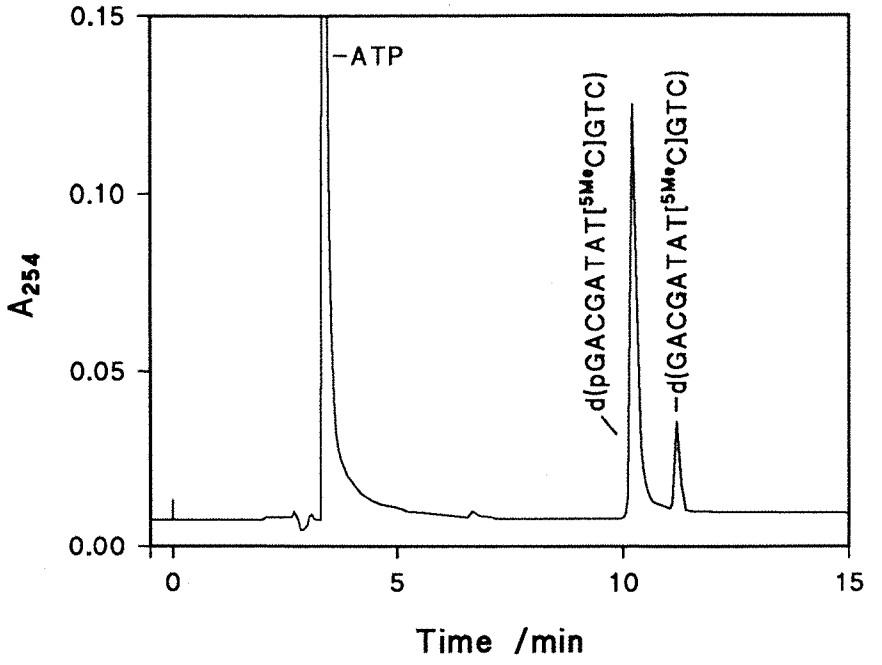


Figure 3a.15: HPLC trace of the phosphorylation of d(GACGATAT[^{5Me}C]GTC) after 21 hours. A gradient of 5-15% buffer B over 15 minutes was used.

present in the oligodeoxynucleotides used here which were dried down from triethylammonium acetate buffers did not seem to have any significant effect on the enzyme. Reactions were followed by reverse phase HPLC as described in materials and methods, with the phosphorylated oligodeoxynucleotide eluting just before the non-phosphorylated oligodeoxynucleotide (see example in Figure 3a.15). Generally phosphorylations were complete within a few hours but because separation of phosphorylated dodecamers from non-phosphorylated dodecamers was difficult by reverse phase HPLC, reactions were allowed to continue for two to three days to ensure maximum conversion. When reactions were complete, the phosphorylated oligodeoxynucleotides were purified and desalted by reverse phase HPLC using volatile triethylammonium bicarbonate buffers. In this way, desalted phosphorylated dodecamers were prepared with less than 1% contamination by non-phosphorylated DNA. The phosphorylated dodecamers were characterised by treatment of a sample with alkaline phosphatase to remove the 5'-phosphate, followed by HPLC co-elution with the corresponding standard non-phosphorylated dodecamer.

3a.3 CHARACTERISATION OF OLIGODEOXYNUCLEOTIDES

3a.3.1 Base Composition Analysis of Oligodeoxynucleotides

Because of the high reliability of automated DNA synthesis, DNA sequences are rarely determined. Problems that normally arise in DNA synthesis generally lead to total failure of the synthesis rather than incorrect sequence synthesis. With modified deoxynucleosides, some of which could potentially react during synthesis, it is important to ensure that the modified deoxynucleoside has been incorporated into the sequence without alteration. In order to be able to do this, it is not necessary to determine the absolute DNA sequence since the base composition alone provides adequate evidence that the deoxynucleoside analogue is present in the oligodeoxynucleotide and that it has not been modified during the DNA synthesis.

Base compositions were determined by totally digesting the oligodeoxynucleotides with snake venom phosphodiesterase (and nuclease P1 for those oligodeoxynucleotides that were resistant to the phosphodiesterase) together with alkaline phosphatase and then calculating the relative amounts of each deoxynucleoside thus produced. Digestion mixtures were analyzed by reverse phase HPLC. Using the gradient described in section 2.2.7.i the four normal deoxynucleosides were well resolved from each other and from the modified deoxynucleosides as illustrated in Figure 3a.16. The positions of the modified deoxynucleotides are given relative to the normal deoxynucleotides. Their positions relative to each other are not known and are given as shown in Figure 3a.16 based on their relative retention times. The deoxynucleoside peak areas were determined by computer integration and then converted to relative amounts

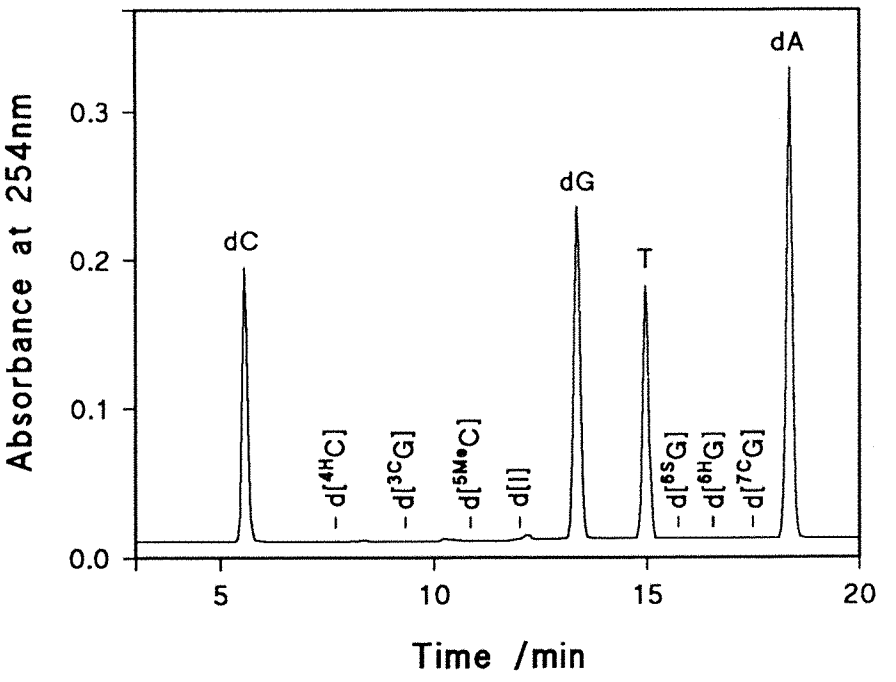


Figure 3a.16: HPLC trace showing the relative elution positions of the four 'normal' and the seven analogue deoxynucleosides (gradient is 3-23% buffer B over 20 minutes with buffer E as reservoir A).

of deoxynucleoside by correcting for their extinction coefficients. The majority of the deoxynucleosides were detected at 254nm but because of their low absorbances at this wavelength, d[^{6S}G] and d[^{4H}C] were detected at their λ_{max}'s of 341nm and 303nm respectively. Separation of these deoxynucleosides from the four normal deoxynucleosides was good enough to enable the UV detector to be switched to their λ_{max} wavelengths to detect their elution without the need to perform a separate HPLC run. Despite the low absorbance of d[^{6H}G] at 254nm (ε_{254} = 3.0 x $10^3 M^{-1}cm^{-1}$) it was not found necessary to detect it at its λ_{max} of 304nm since it gave a satisfactory peak area at 254nm. In some other applications, though, such as when it is present in a longer oligodeoxynucleotide and therefore constitutes less of the total deoxynucleoside composition, it may be necessary to detect it at 304nm.

All of the analogue containing oligodeoxynucleotides gave the expected base compositions as shown in Table 3a.1.

Table 3a.1: Base compositions of the oligodeoxynucleotides determined as described in section 2.2.7.i.

ANALOGUE	Expected Base Composition	Determined Base Composition
d(GACGATATCGTC)	A_3, C_3, G_3, T_3	$A_{3.0}$, $C_{3.0}$, $G_{3.2}$, $T_{2.9}$
d(GAC[^{6S}G]ATATCGTC)	A_3, C_3, G_2, T_3, [^{6S}G]$_1$	$A_{2.7}$, $C_{3.1}$, $G_{2.0}$, $T_{2.9}$, [^{6S}G]$_{1.2}$
d(GAC[^{6H}G]ATATCGTC)	A_3, C_3, G_2, T_3, [^{6H}G]$_1$	$A_{2.8}$, $C_{3.3}$, $G_{2.1}$, $T_{3.0}$, [^{6H}G]$_{0.9}$
d(GAC[^{7C}G]ATATCGTC)	A_3, C_3, G_2, T_3, [^{7C}G]$_1$	$A_{3.0}$, $C_{3.0}$, $G_{1.9}$, $T_{3.1}$, [^{7C}G]$_{0.9}$
d(GAC[^{3C}G]ATATCGTC)	A_3, C_3, G_2, T_3, [^{3C}G]$_1$	$A_{2.9}$, $C_{3.0}$, $G_{2.0}$, $T_{3.1}$, [^{3C}G]$_{1.0}$
d(GAC[I]ATATCGTC)	A_3, C_3, G_2, T_3, [I]$_1$	$A_{2.9}$, $C_{3.1}$, $G_{2.0}$, $T_{3.0}$, [I]$_{1.0}$
d(GACGATAT[^{4H}C]GTC)	A_3, C_2, G_3, T_3, [^{4H}C]$_1$	$A_{2.9}$, $C_{2.2}$, $G_{3.2}$, $T_{2.9}$, [^{4H}C]$_{0.7}$
d(GACGATAT[^{5Me}C]GTC)	A_3, C_2, G_3, T_3, [^{5Me}C]$_1$	$A_{2.8}$, $C_{2.2}$, $G_{3.0}$, $T_{3.0}$, [^{5Me}C]$_{0.9}$
d(TGACGATATCGTC)	A_3, C_3, G_3, T_4	$A_{2.9}$, $C_{3.2}$, $G_{3.0}$, $T_{4.0}$
d(GTCGACGATATCGTCGAC)	A_4, C_5, G_5, T_4	$A_{4.2}$, $C_{4.8}$, $G_{5.0}$, $T_{4.0}$
d(GTCGAC[^{6H}G]ATATCGTCGAC)	A_4, C_5, G_5, T_4, [^{6H}G]$_1$	$A_{3.9}$, $C_{4.9}$, $G_{4.2}$, $T_{4.1}$, [^{6H}G]$_{1.0}$

3a.3.2 Hyperchromicity and Extinction Coefficients of the Oligodeoxynucleotides

The base stacking interactions in double stranded DNA are known to reduce the UV absorbances of the DNA bases (Saenger, 1984). This hyperchromic effect results in the molar extinction coefficient of an oligodeoxynucleotide being less than the sum of the molar extinction coefficients of its constituent deoxynucleosides. The actual extinction coefficient of an oligodeoxynucleotide can be determined in an indirect manner by completely digesting it into its deoxynucleotides. When the oligodeoxynucleotide is digested, the individual deoxynucleotides are no longer base stacked and so the UV absorbance is then due to the sum of the extinction coefficients of the constituent deoxynucleotides. The ratio of the UV absorbance of the digested oligodeoxynucleotide to that of the intact oligodeoxynucleotide is therefore equal to the ratio of the sum of the deoxynucleotides extinction coefficients to the extinction coefficient of the intact oligodeoxynucleotide. That is:-

$$\text{Hyperchromicity} = \frac{\text{absorbance of digested oligo}}{\text{absorbance of intact oligo}} = \frac{\sum\varepsilon_{\text{nucleotides}}}{\varepsilon_{\text{intact oligo}}}$$

$$\therefore \quad \varepsilon_{\text{intact oligo}} = (\sum\varepsilon_{\text{nucleotides}})/(\text{hyperchromicity})$$

For meaningful kinetic experiments to be carried out on the oligodeoxynucleotides synthesised here, an accurate method of concentration determination is required. This can be

accurately and conveniently done by measuring the UV absorbance of an oligodeoxynucleotide provided its extinction coefficient is known. The hyperchromicity values of all of the oligodeoxynucleotides were therefore determined by digesting about half an OD unit of each oligodeoxynucleotide with snake venom phosphodiesterase, or, for the phosphodiesterase resistant d(GAC[^{6S}G]ATATCGTC), d(GAC[^{6H}G]ATATCGTC), d(GAC[^{7C}G]ATATCGTC) and d(GAC[I]ATATCGTC) dodecamers, snake venom phosphodiesterase and nuclease P1. Reactions were carried out in quartz cuvettes and the digestions were followed by the change in 254nm absorbance with time. When the UV absorbance plateaued out, the digestions were checked by HPLC to ensure that the cleavage into deoxynucleotides was complete. Hyperchromicity values were then calculated from the final (*i.e.* completely digested oligodeoxynucleotide) 254nm absorbance divided by the initial (*i.e.* intact oligodeoxynucleotide) 254nm absorbance. Molar extinction coefficients for each oligodeoxynucleotide were then calculated from these hyperchromicities and the sum of the constituent deoxynucleoside extinction coefficients as described above. The resulting hyperchromicities and extinction coefficients are given in Table 3a.2. All of the hyperchromicity values except for that of d(GAC[^{6H}G]ATATCGTC) are very similar and fall into the narrow range of 1.34 to 1.55. This gives a very crude indication that the base stacking interactions and hence the structures of these oligodeoxynucleotides are similar. The unusually low hyperchromicity of d(GAC[^{6H}G]ATATCGTC) shows that there is considerably less base stacking in this dodecamer than in the other oligodeoxynucleotides.

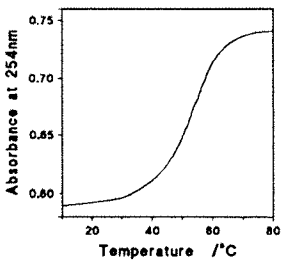
3a.3.3 Melting Temperatures of the Oligodeoxynucleotides

Since the Eco RV endonuclease, as for other type II restriction endonucleases, only cleaves double stranded DNA (Modrich, 1979) it was important to check that all of the oligodeoxynucleotides used formed stable duplexes under the endonuclease assay conditions. This was done by measuring the UV absorbance of approximately 3μM oligodeoxynucleotide as a function of temperature. At low temperatures a DNA duplex has a relatively low UV absorbance due to the hyperchromic effect of base stacking (Saenger, 1984). Raising the temperature of the DNA beyond a certain point disrupts the double stranded helix as the increased thermal motion overcomes the weak hydrogen bonding and base stacking interactions that hold the duplex together. Because disruption of the double helix disrupts the base stacking that gives rise to the hyperchromic effect, the 'melting' of an oligodeoxynucleotide is accompanied by a rise in its UV absorbance. This results in the type of melting curve shown for the control dodecamer, d(GACGATATCGTC), in Figure 3a.17a. On the lower plateau (below ~30°C), the dodecamer is essentially all double helical. On the upper plateau (above ~60°C), the dodecamer is essentially in a random coil configuration. In between these values the double helix is partially disrupted. The point of inflexion in the curve (when the rate of change of absorbance with temperature is at its greatest) is designated the melting temperature, T_M. At this temperature in a typical melting curve, the oligodeoxynucleotide is half double stranded

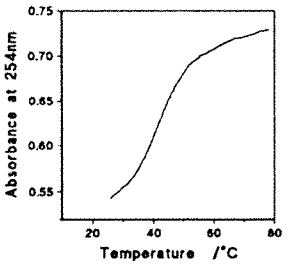
Table 3a.2: Melting temperatures, hyperchromicities and extinction coefficients (254nm) of the oligodeoxynucleotides. [a]Units of $M^{-1}cm^{-1}$ (for double stranded DNA which is the true endonuclease substrate). [b]No helix-to-coil transition was observed. [c]Not determined. [d]These gave biphasic melting curves and therefore two melting temperatures.

ANALOGUE	Melting Temperature, T_M	Hyperchromicity	Extinction Coefficient at 254nm[a]
d(GACGATATCGTC)	53°C	1.48	166 x 10^3
d(GAC[^{6S}G]ATATCGTC)	43°C	1.39	170 x 10^3
d(GAC[^{6H}G]ATATCGTC)	-[b]	1.19	188 x 10^3
d(GAC[^{7C}G]ATATCGTC)	52°C	1.41	173 x 10^3
d(GAC[^{3C}G]ATATCGTC)	47°C	1.42	164 x 10^3
d(GAC[I]ATATCGTC)	46°C	1.45	164 x 10^3
d(GACGATAT[^{4H}C]GTC)	33°C	1.34	175 x 10^3
d(GACGATAT[^{5Me}C]GTC)	54°C	1.42	171 x 10^3
d(TGACGATATCGTC)	-[c]	1.41	183 x 10^3
d(GTCGACGATATCGTCGAC)	56°C & 72°C[d]	1.55	236 x 10^3
d(GTCGAC[^{6H}G]ATATCGTCGAC)	44°C & 70°C[d]	1.40	246 x 10^3

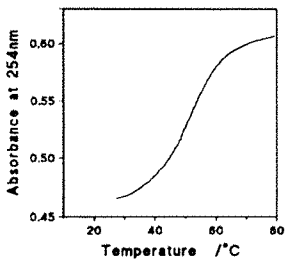
and half single stranded. The melting curves of all the oligodeoxynucleotides used are given in Figure 3a.17 and the melting temperatures determined from these curves are given in Table 3a.2. The control dodecamer has a T_M of 53°C and would therefore be double stranded at the endonuclease reaction temperature used of 25°C. The majority of the T_M's for the analogue containing dodecamers are greater than 40°C and thus they also are expected to be virtually 100% double stranded at 25°C. The d[^{5Me}C] and d[^{7C}G] containing oligodeoxynucleotides show no significant difference in their T_M's compared with the control dodecamer and so the presence of these deoxynucleoside analogues in the dodecamers cause no alteration to the duplex stability. A slight decrease in the T_M of a d[^{7C}G] containing octadeoxynucleotide of 4°C has also been reported (Seela & Driller, 1986). Other reports for d[^{5Me}C] containing oligodeoxynucleotides have noted increases in their T_M's compared with dC containing oligodeoxynucleotides (Butkus *et al.*, 1987; Brennan *et al.*, 1986). The increase of 1°C for d(GACGATAT[^{5Me}C]GTC) seen here is in line with this increased stability but is too small to attach any real significance to it. The replacement of dG with d[^{3C}G] does not disrupt the Watson-Crick base pairing hydrogen bonds and so the slightly reduced T_M of 47°C for the d[^{3C}G] containing dodecamer is probably caused by a reduction in the base stacking. The d[I] containing dodecamer has a slightly lower T_M (46°C) than the control dodecamer. This is not surprising as a base pairing hydrogen bond (two per duplex) has been lost by the substitution of d[I] for dG which loses the 2-amino group normally involved in the Watson-Crick base pairing. A similar drop of 10°C in the T_M of a d[I] containing octadeoxynucleotide has also been observed (Ono & Ueda, 1987a). The T_M of the d[^{6S}G] analogue is also similarly lower due presumably to reduction in the strength of hydrogen bonding at the deoxyguanosine 6-position



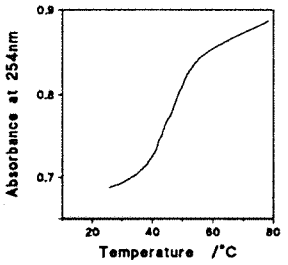
(a) d(GACGATATCGTC)



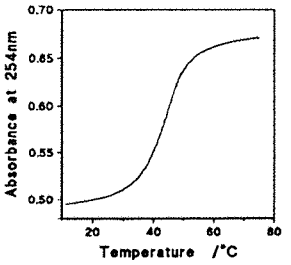
(b) d(GAC[^{6S}G]ATATCGTC)



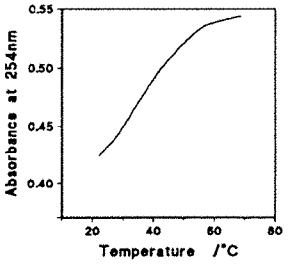
(c) d(GAC[^{7C}G]ATATCGTC)



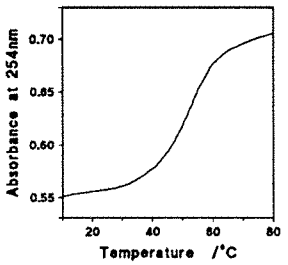
(d) d(GAC[^{3C}G]ATATCGTC)



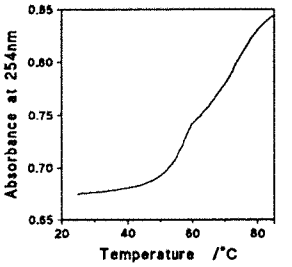
(e) d(GAC[I]ATATCGTC)



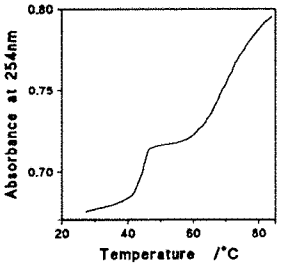
(f) d(GACGATAT[^{4H}C]GTC)



(g)
d(GACGATAT[^{5Me}C]GTC)



(h)
d(GTCGACGATATCGTCGAC)



(i)
d(GTCGAC[^{6H}G]ATATCGTCGAC)

Figure 3a.17: Melting curves of the oligodeoxynucleotides.

when a sulphur rather than an oxygen is at this position. The extra bulk of the sulphur atom may also reduce the base stacking in the duplex thus destabilising the helix.

The somewhat low T_M of 33°C for the d[^{4H}C] dodecamer is surprising since although a base pairing hydrogen bond has been lost as in the deoxyinosine analogue, its T_M is a further 13°C lower. A similar T_M reduction of 20°C has also been recorded for the replacement of dC by d[^{4H}C] in a decamer (Gildea & McLaughlin, 1989). Again base stacking interactions may play a part in the reduced duplex stability.

All of the above values of T_M mean that those analogue oligodeoxynucleotides discussed would be predominantly double stranded at 25°C. This is not the case for the d[^{6H}G] dodecamer which exhibited no helix-to-coil transition and must therefore be assumed to be single stranded

at room temperature. This is not entirely surprising since the d[^{6H}G] base makes two less base pairing hydrogen bonds than does deoxyguanosine and since each dodecamer duplex would contain two d[^{6H}G] bases, the double stranded oligodeoxynucleotide is destabilised by the loss of four hydrogen bonds. Similar results for other d[^{6H}G] containing oligodeoxynucleotides have been observed although the actual destabilising effect varied with the position of d[^{6H}G] within the oligodeoxynucleotide and the length of the oligodeoxynucleotide (McLaughlin *et al.*, 1988). To overcome the problem of d(GAC[^{6H}G]ATATCGTC) being single stranded at the endonuclease assay temperature of 25°C and thus not representing a conservative deletion of the 6-carbonyl oxygen of deoxyguanosine, the longer d(GTCGAC[^{6H}G]ATATCGTCGAC) oligodeoxynucleotide was synthesised. It was hoped that the extra base pairs would stabilise the two [^{6H}G].C base pairs enough to keep the oligodeoxynucleotide double stranded at 25°C. The melting curve of this oligodeoxynucleotide is given in Figure 3a.17i and exhibits a clear biphasic nature. This arises from the oligodeoxynucleotide not melting entirely in one process but from melting occurring locally in one part of the oligodeoxynucleotide first and then at higher temperatures the rest of the duplex melting. The first local melting process most likely occurs in the central region around the two d[^{6H}G] residues as these form fewer base pairing hydrogen bonds than normal deoxynucleosides and are therefore expected to be where the duplex is weakest. Despite the biphasic nature of the curve it can be seen that the oligodeoxynucleotide would be double stranded at 25°C. The two melting events are well enough separated to enable the evaluation of the melting temperatures for each process and were found to be 44°C and 70°C. The melting curve of d(GTCGACGATATCGTCGAC), the control of the d[^{6H}G] containing octadecamer, is shown in Figure 3a.17h. It too shows a slight biphasic nature but much less pronounced than in the d[^{6H}G] oligodeoxynucleotide. This may be caused by the local melting of the four central A.T base pairs at a lower temperature than the rest of the oligodeoxynucleotide. The lower melting temperature of A.T rich regions is a well documented phenomenum and is due to A.T base pairs having only two hydrogen bonds as opposed to the three of G.C base pairs. The two melting events are less well resolved than for the d[^{6H}G] oligodeoxynucleotide but nevertheless reasonable T_M values can still be obtained. The first T_M of 56°C is 12°C higher than that of the d[^{6H}G] octadecamer whereas the second T_M of 72°C is roughly the same in both cases. This is in agreement with the hypothesis that the first transition is due to the melting of the central region since it is this transition affected by the presence of the d[^{6H}G] residues which are expected to cause local destabilisation.

3a.3.4 Circular Dichroism Spectroscopy of the Oligodeoxynucleotides

CD spectroscopy is a technique that looks at the differential absorbance of clockwise and anti-clockwise circularly polarised light by a compound. For any differential absorbance to occur, the compound must be chiral and for DNA, the nature of the helix chirality is the dominant factor in its CD spectra. A right handed B-form DNA structure, as typified by the Dickerson dodecamer d(CGCGAATTCGCG), has a large negative peak at around 250nm and a smaller positive peak (about half the intensity of the negative peak) at around 280nm (Fairall *et al.*, 1989). The x-axis crossover point between the positive and negative peaks is at around 270nm. Other B-form oligodeoxynucleotides have similar spectra. The drastic differences in the helix structure of Z-DNA compared with B-DNA cause drastic changes in the CD spectrum of Z-DNA. The reversal of the handedness of the helix brings about a reversal in the CD spectrum. The positive maximum is moved to about 260nm and the negative peak, which is now only slightly larger than the positive peak, is at about 290nm (Pohl & Jovin, 1972; Butkus *et al.*, 1987). The crossover point also changes and is moved up to around 280nm. Probably more relevant to the needs of this study, CD spectroscopy can also distinguish A-form from B-form DNA, although the differences are less clearly defined. The spectrum of calf thymus DNA in 80% ethanol has been suggested as a model of A-DNA (Fairall *et al.*, 1989). In this A-DNA spectrum, the positive peak is about twice the intensity of the negative peak, as opposed to half in the B-DNA spectrum. The maxima are both shifted to lower wavelengths, relative to B-DNA, with the positive peak at about 270nm and the negative peak at about 240nm. Also the crossover point is shifted down to about 250nm. The CD spectra of DNA samples do not necessarily exhibit either A- or B-DNA characteristics but can show features that are intermediate between the two forms.

CD spectra of all the oligodeoxynucleotides, except for d(TGACGATATCGTC), were recorded in a phosphate buffer very similar in composition to the endonuclease reaction buffer. The CD spectrum of d(GACGATATCGTC) is shown in Figure 3a.18a The majority of the other oligodeoxynucleotide CD spectra were virtually indistinguishable from this spectrum and all of them had positive maxima at around 280nm and negative maxima at around 250nm. The ratios of the positive peak intensities to the negative intensities all, with the exception of d(GAC[^{6H}G]ATATCGTC), fell into the range 0.72 to 0.90. The oligodeoxynucleotides used here therefore exhibit CD spectra typical of B-form DNA. The slightly higher ratio of peak intensities for d(GAC[^{6H}G]ATATCGTC) of ~1 is suggestive of its single stranded nature since an increase in this ratio has been observed when the temperature of a dodecamer has been raised above its melting temperature (data not shown). This ratio difference between d(GAC[^{6H}G]ATATCGTC) and the other oligodeoxynucleotides is probably too small to be taken as any real proof of single stranded nature and it is also not uncommon for double stranded DNA to have peak ratios of 1 or higher.

Another feature of the CD spectra supportive of the B-DNA nature of the oligodeoxynucleotides is the fact that all, except for d(GAC[I]ATATCGTC), have their crossover

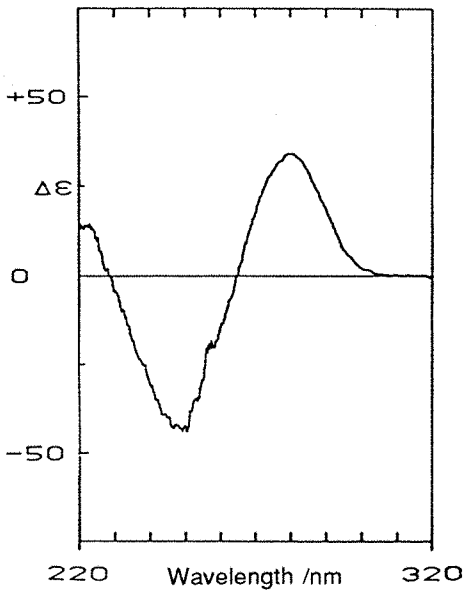


Figure 3a.18a: CD spectrum of d(GACGATATCGTC).

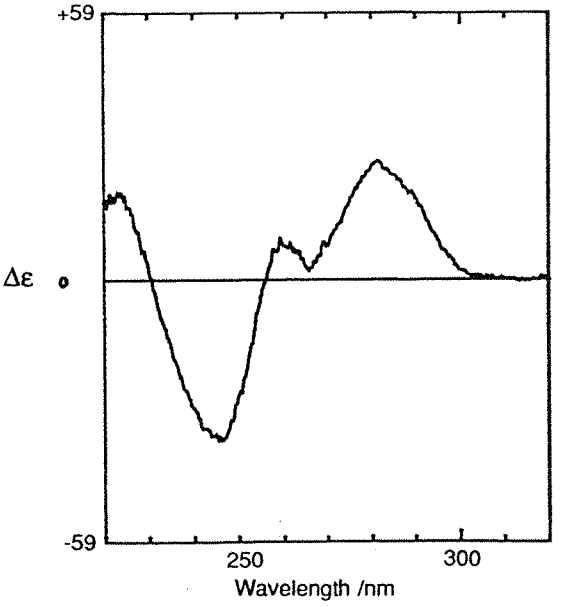


Figure 3a.18b: CD spectrum of d(GAC[I]ATATCGTC).

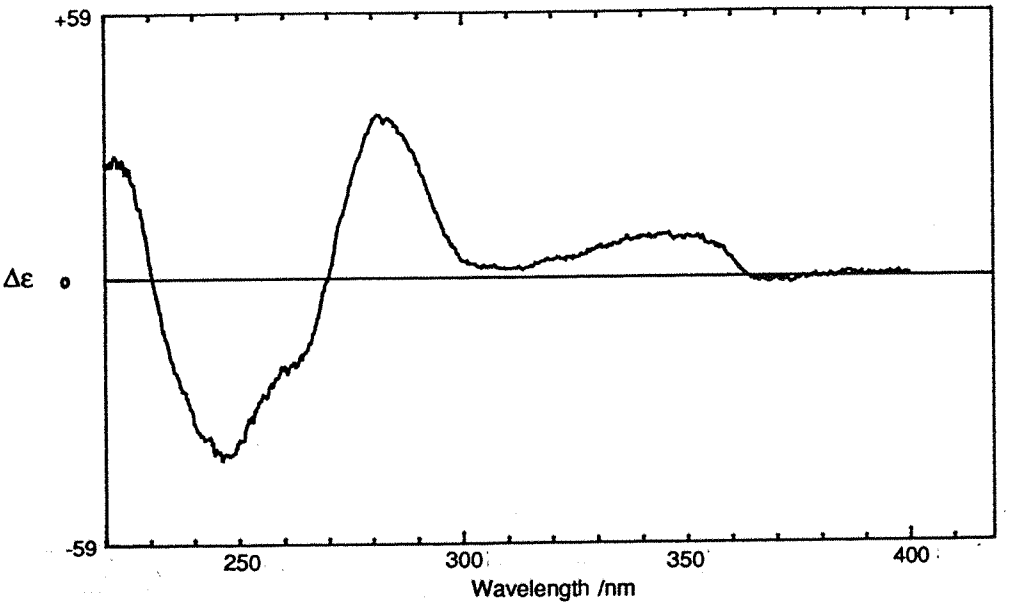


Figure 3a.18c: CD spectrum of d(GAC[^{6S}G]ATATCGTC).

points at ~265nm. The d[I] dodecamer has an interesting feature in the extra positive peak centred around 260nm that appears as a shoulder on the main peak (Figure 3a.18b). The presence of this extra peak causes the shift of the crossover point down to ~255nm. A very similar 'blip' has also been reported for another d[I] containing oligodeoxynucleotide, d(GGA[I]ATCTCC) (Ono & Ueda, 1987a). This could arise because the λ_{max} of d[I] at 248nm is at a lower wavelength than the other deoxynucleosides causing it to show up as a separate peak in the CD spectrum. The exceptionally high λ_{max} of d[^{6S}G] at 341nm is well enough resolved from the rest of the deoxynucleoside absorbances to show up as an entirely separate peak in the CD spectrum of dGAC[^{6S}G]ATATCGTC), with a λ_{max} at around 345nm (Figure

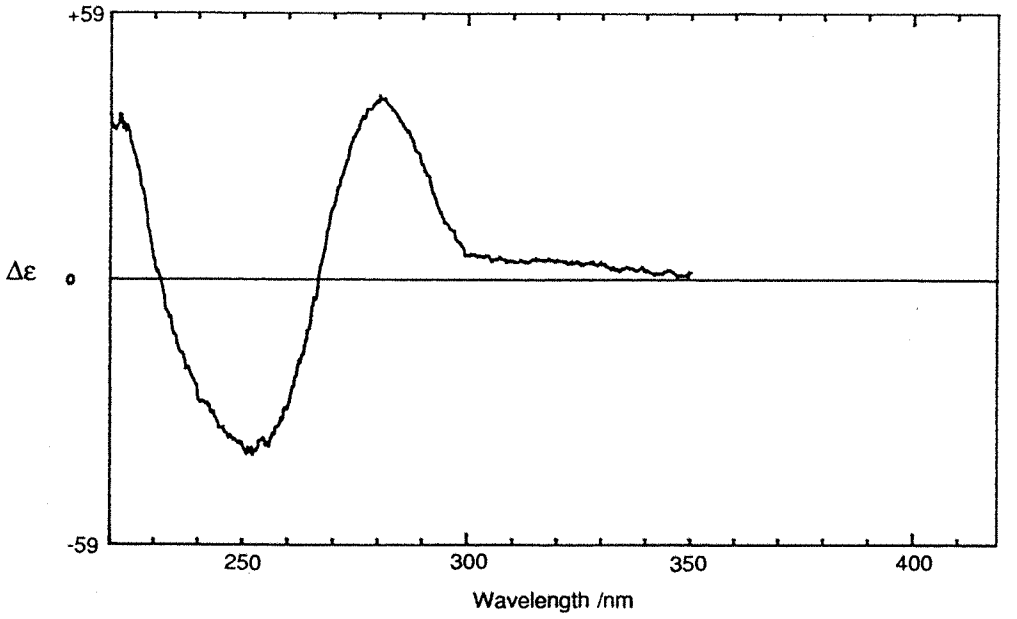


Figure 3a.18d: CD spectrum of d(GAC[^{6H}G]ATATCGTC).

3a.18c). The λ_{max} of d[^{6H}G] (303nm) is also higher than the normal oligodeoxynucleotide absorbance and can be seen as a shoulder on the main positive peak in the CD spectrum of d(GAC[^{6H}G]ATATCGTC) (Figure 3a.18d)

3a.3.5 Ultraviolet Spectroscopy of d(GAC[^{6S}G]ATATCGTC), d(GAC[^{6H}G]ATATCGTC), d(GACGATAT[^{4H}C]GTC) and d(GAC[^{3C}G]ATATCGTC)

The UV spectrum of the d[^{6S}G] containing oligodeoxynucleotide is shown in Figure 3a.19 together with that of the control dodecamer and the free d[^{6S}G] deoxynucleoside. The extra band due to the presence of the modified base is clearly visible and is well removed from the main DNA band peaking at 260nm. The λ_{max} has been shifted from 341nm for the free base to 348nm in the DNA. An explanation of this could be that the Watson-Crick base pairing and/or the base stacking interactions in double stranded DNA cause a decrease in the energy needed for the electronic transfer that is responsible for the band, and hence shift it to higher wavelengths. The hyperchromic effect on the d[^{6S}G] residue is slightly less than the average effect on the rest of the deoxynucleosides since the ratio of the ε_{max} for the free deoxynucleoside to the ε_{max} of d[^{6S}G] in the dodecamer is approximately 1.25 whereas the hyperchromicity of the oligodeoxynucleotide at 254nm is 1.39 [Table 3a.2]. Of course the hyperchromicity value at 254nm is not exactly equivalent since the shifts in λ_{max}'s of other deoxynucleosides, that may occur on incorporation into the oligodeoxynucleotide, are obscured by the main UV peak. Therefore, unlike for d[^{6S}G], hyperchromicity values for a particular normal deoxynucleoside cannot be individually determined.

The d[^{6H}G], d[^{3C}G] and the d[^{4H}C] containing oligodeoxynucleotides also show an extra band (Figures 3a.20, 21 and 22) but because the λ_{max}'s of the three bases are relatively small

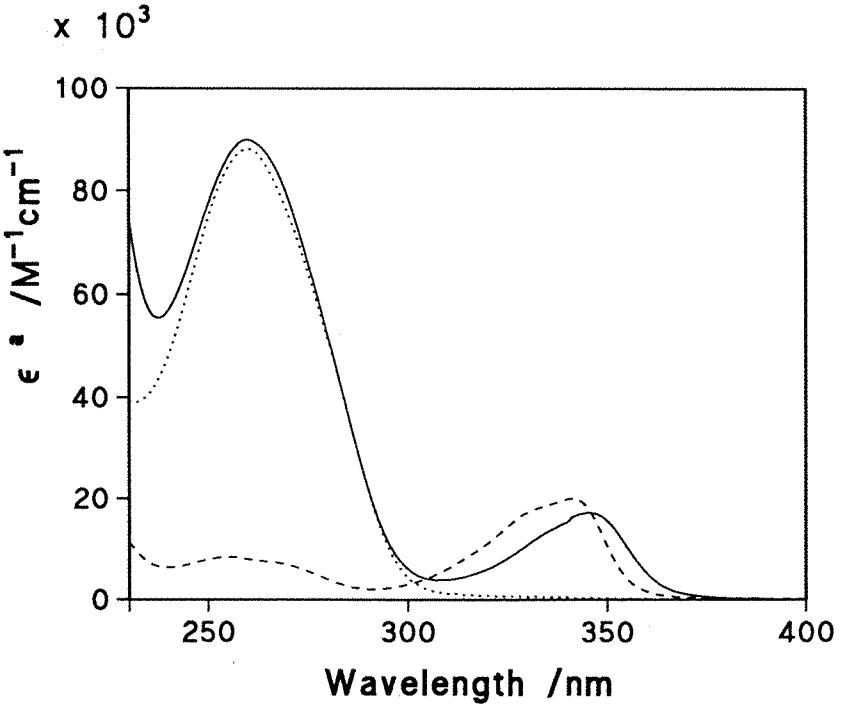


Figure 3a.19: UV spectrum of d(GAC[^{6S}G]ATATCGTC) (——), d(GACGATATCGTC) (···) and d[^{6S}G] (---). [a]Extinction coefficient per mole of single stranded DNA.

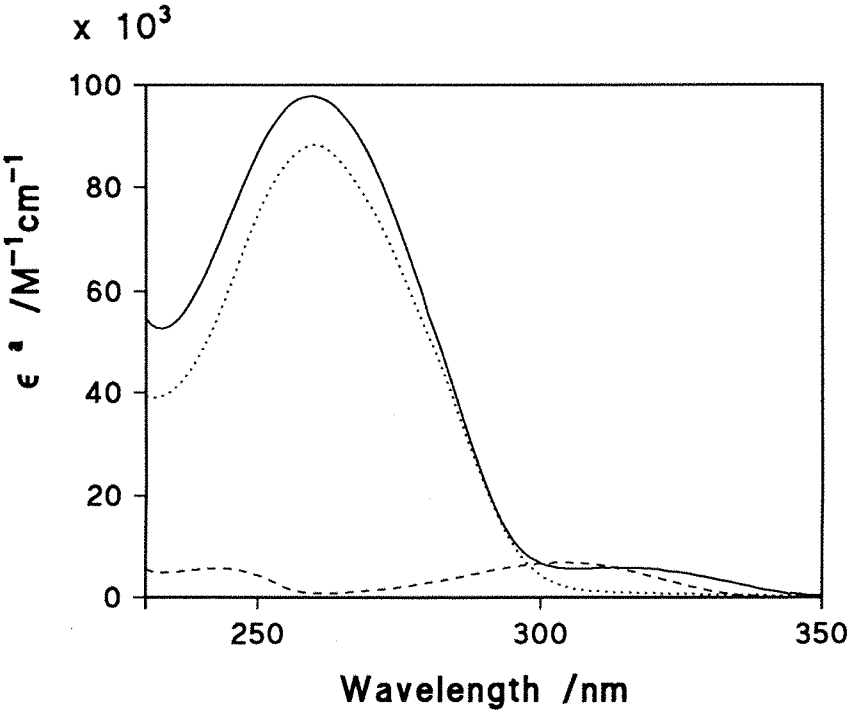


Figure 3a.20: UV spectrum of d(GAC[^{6H}G]ATATCGTC) (——), d(GACGATATCGTC) (···) and d[^{6H}G] (---). [a]Extinction coefficient per mole of single stranded DNA.

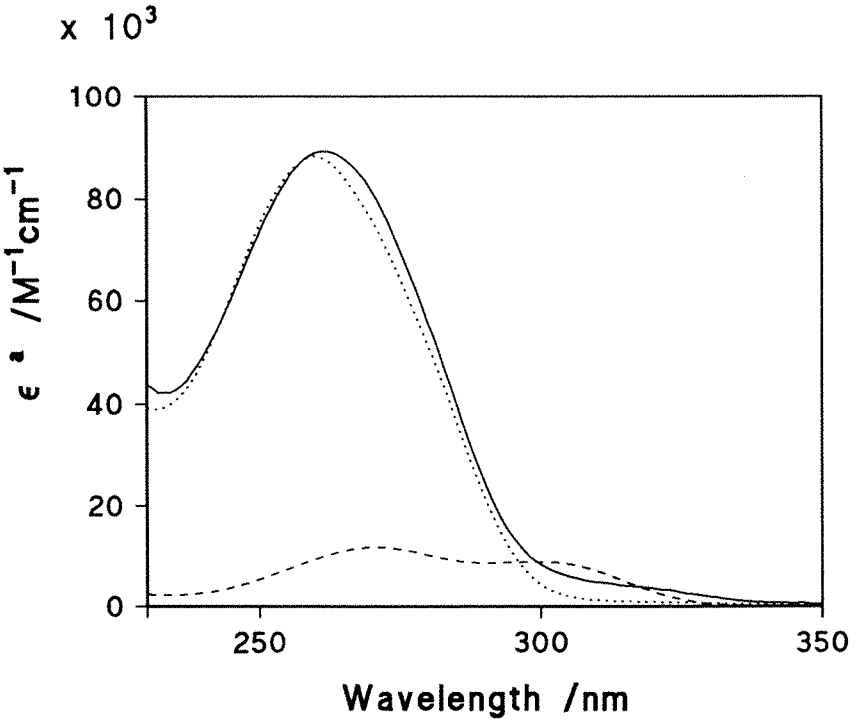


Figure 3a.21: UV spectrum of d(GAC[^{3C}G]ATATCGTC) (——), d(GACGATATCGTC) (···) and d[^{3C}G] (---). [a]Extinction coefficient per mole of single stranded DNA.

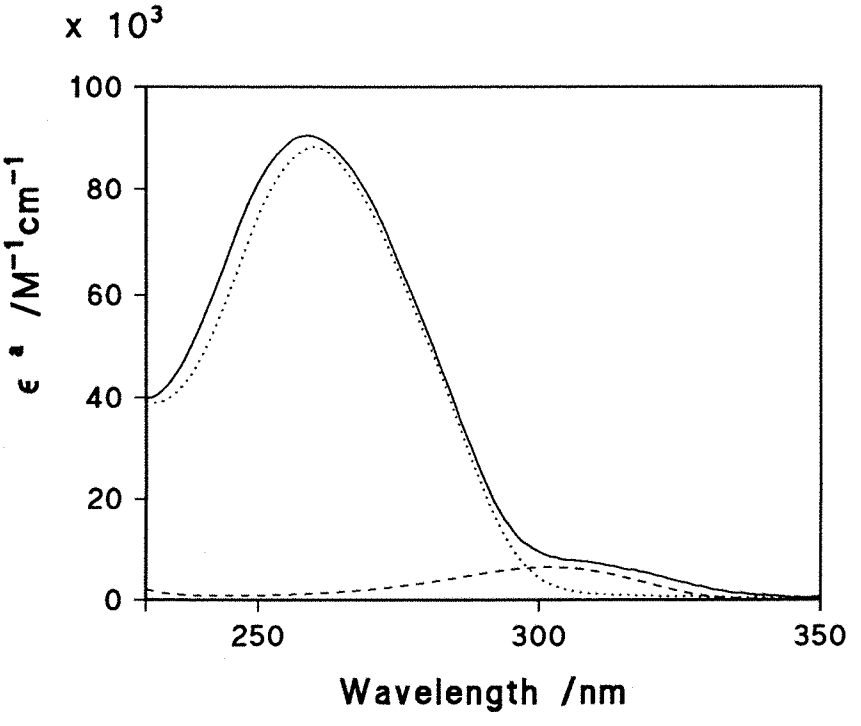


Figure 3a.22: UV spectrum of d(GAC[^{4H}C]ATATCGTC) (——), d(GACGATATCGTC) (···) and d[^{4H}C] (---). [a]Extinction coefficient per mole of single stranded DNA.

and closer to that of the rest of the DNA, they only appear as shoulders to the main DNA band. They are shown together with the UV spectrum of the control dodecamer and the corresponding free deoxynucleoside for comparison. The shoulders due to the modified bases are clearly present confirming their incorporation into the DNA. Because of the closeness of these shoulders to the main peak they probably include some absorbance from the four normal deoxynucleosides. The hyperchromic effect on the individual modified deoxynucleosides cannot therefore be determined.

3a.3.6 Fluorescence Spectroscopy of d(GAC[^{6H}G]ATATCGTC) and d(GACGATAT[^{4H}C]GTC)

The fluorescence emission and excitation spectra of the free d[^{6H}G] deoxynucleoside and of d(GAC[^{6H}G]ATATCGTC) are given in Figures 3a.23a and b respectively. The general shape of the two emission bands are the same and both have maxima at 367nm. The two excitation bands are slightly different with the maximum of the deoxynucleoside at 302nm being shifted up to 305nm in the dodecamer. There is also an extra peak in the excitation spectrum of the oligodeoxynucleotide at around 260nm and is presumably due to some effect caused by the other deoxynucleosides in the oligodeoxynucleotide. The excitation and emission maxima of d[^{6H}G] are in close agreement with those published for the 2-aminopurine riboside of 303nm and 370nm respectively (Ward *et al.*, 1969). The relative quantum yields of the oligodeoxynucleotide and the deoxynucleoside were calculated from the areas of their emission bands and taking into account the UV absorbances of the solutions at the excitation wavelength. The quantum yield for the oligodeoxynucleotide was found to be approximately a twentieth of that for the free deoxynucleoside. Ward *et al.* (1969) reported a quantum yield reduction of greater than 250 fold for polyribonucleotides containing alternating 2-aminopurine residues but that this reduction was only about 50 fold when the polymer was thermally denatured. Although direct comparisons are not entirely valid (because, for example, the ^{6H}G riboside was base paired with U or T derivatives in the polyribonucleotides whereas in the dodecamer used here d[^{6H}G] is paired with dC), the result here of only a 20 fold reduction in the quantum yield is in line with d(GAC[^{6H}G]ATATCGTC) being single stranded at 25°C.

Because the structurally similar d[^{4H}T] deoxynucleoside has been reported to be fluorescent (Connolly & Newman, 1989), it was expected that the d[^{4H}C] analogue would also be fluorescent. This was indeed found to be the case as shown by its excitation and emission spectra in Figure 3a.24a. Although the exact details were not given, it appears that the intensity of fluorescence for d[^{4H}C] is of the same order of magnitude as that of d[^{4H}T] (Newman, 1989). The fluorescence intensity of d[^{4H}C] however, is approximately 200 times less than that of d[^{6H}G]. The excitation maximum for d[^{4H}C] was found to be 299nm, and its emission maximum was at 372nm. These values are both about 15nm lower than those seen for the d[^{4H}T] molecule. The fluorescence spectra of d(GACGATAT[^{4H}C]GTC) were also recorded

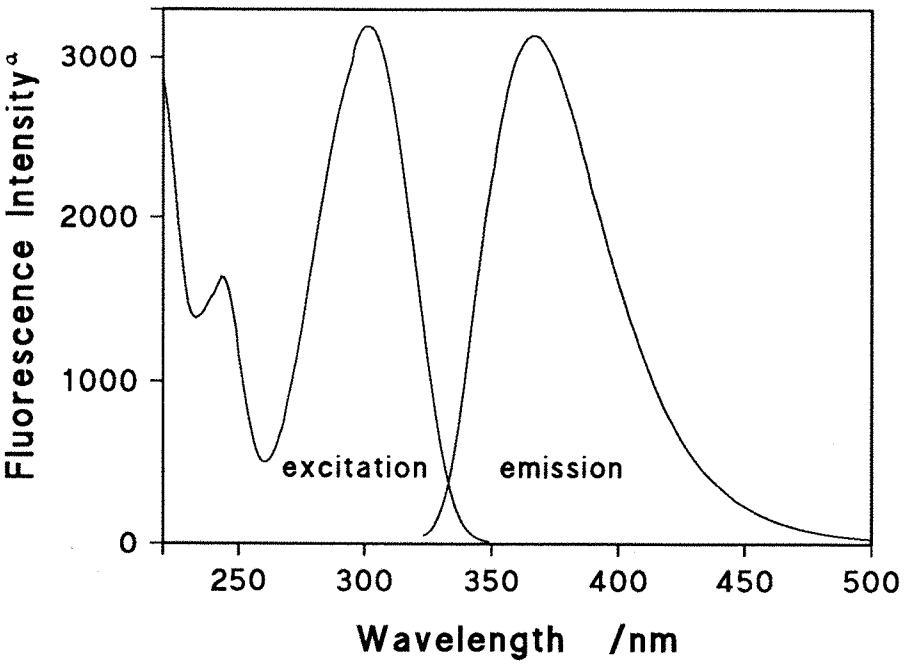


Figure 3a.23a: Fluorescence spectrum of 0.25μM d[^{6H}G]. [a]Arbitrary units.

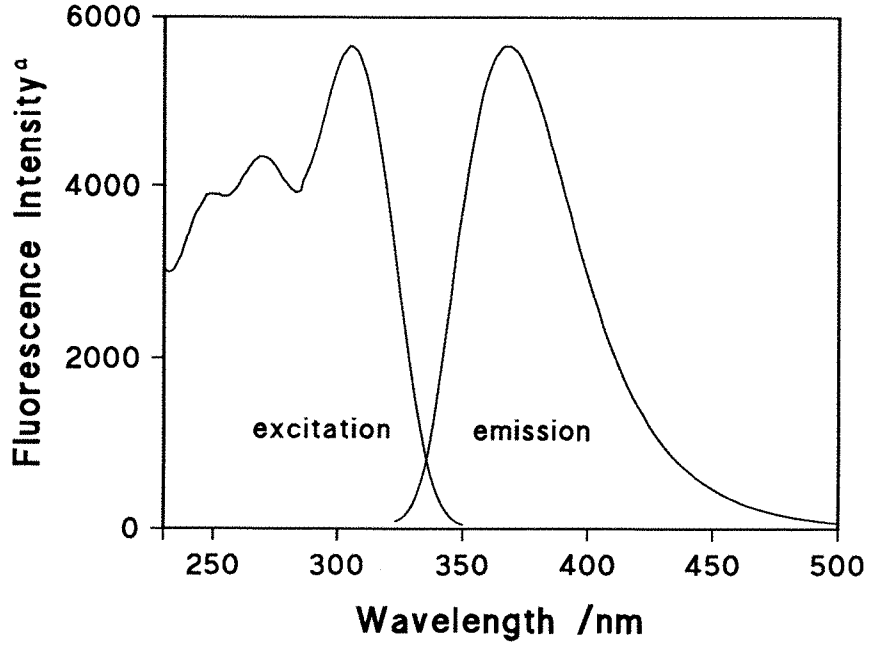


Figure 3a.23b: Fluorescence spectrum of d(GAC[^{6H}G]ATATCGTC) at 5.9μM double stranded concentration. [a]Arbitrary units.

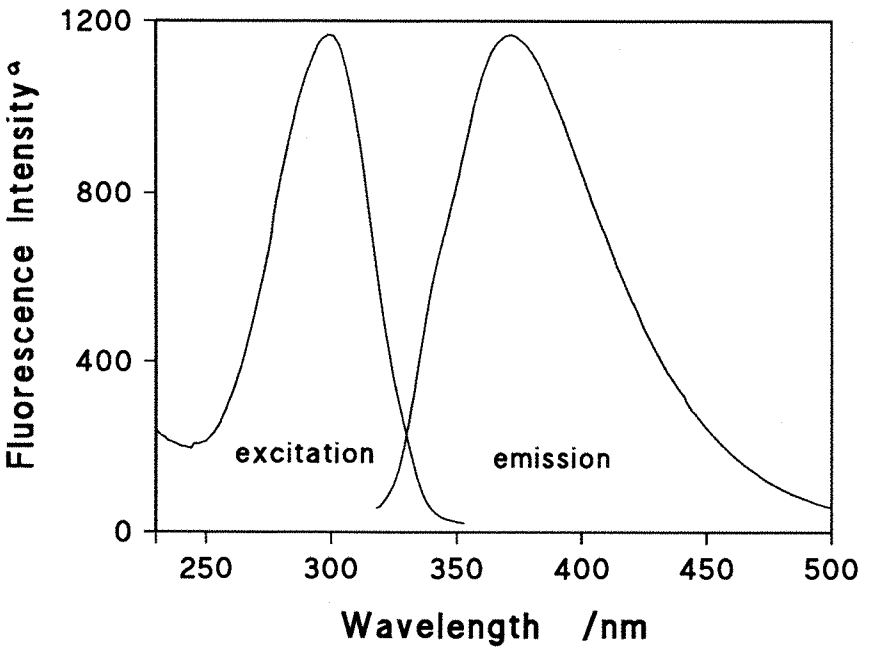


Figure 3a.24a: Fluorescence spectrum of 22.3μM d[^{4H}C]. [a]Arbitrary units

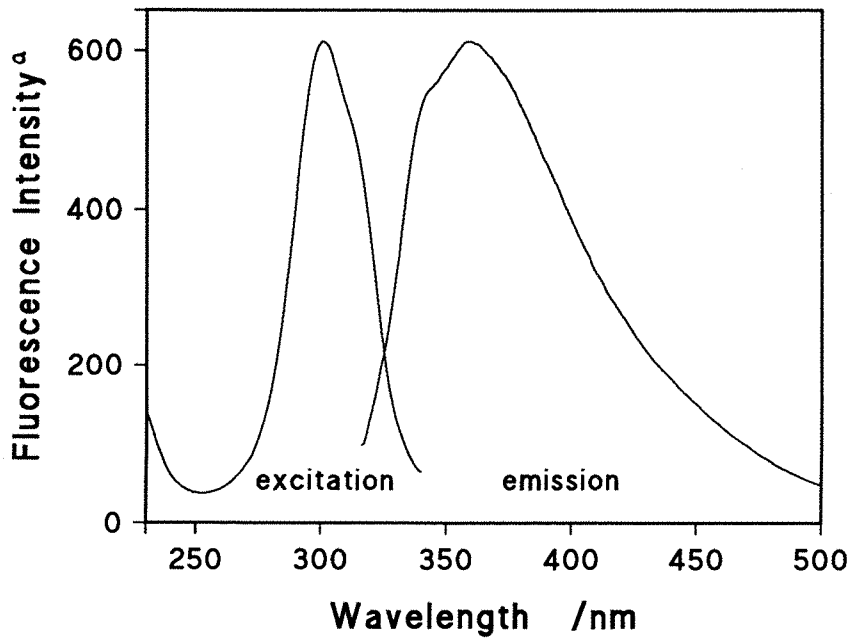


Figure 3a.24b: Fluorescence spectrum of d(GACGATAT[^{4H}C]GTC) at 59μM double stranded concentration. [a]Arbitrary units.

and are shown in Figure 3a.24b. The excitation maximum of the d[^{4H}C] residue in the dodecamer is at 301nm, which is at essentially the same wavelength as for the free deoxynucleoside. The emission maximum in the oligodeoxynucleotide is, however, shifted 13nm to 359nm. This in contrast to the effect seen for the incorporation of d[^{4H}T] into two different dodecamers which both resulted in the emission maximum being shifted to *higher* wavelengths. The quantum yield for d(GACGATAT[^{4H}C]GTC) was approximately a twentieth of that for the free deoxynucleoside. This also contrasts with the results of the two d[^{4H}T] containing dodecamers (Connolly & Newman, 1989) with the quenching seen for d[^{4H}C] being about ten times that seen for d[^{4H}T]. The quenching of d[^{4H}C] however, is about the same as that seen for the incorporation of d[^{6H}G] into its dodecamer. Direct comparisons of the quenching seen with the different base analogues in oligodeoxynucleotides is difficult since there are many variable factors involved. These include different surrounding deoxynucleosides and also different base pair hydrogen bonding. On top of this the d(GAC[^{6H}G]ATATCGTC) oligodeoxynucleotide is single stranded whereas the others are double stranded.

To determine whether the ordered structure of a double stranded DNA helix was an important factor in the quenching of the fluorescence of d[^{4H}C], the d(GACGATAT[^{4H}C]GTC) dodecamer was melted by heating the oligodeoxynucleotide solution from 25°C to 70°C. The melting process was monitored by the fluorescence emitted at 360nm with 300nm exciting light. The results are given in Figure 3a.25 and although the exact temperature at any particular time between the initial temperature of 25°C and final temperature of 70°C is not

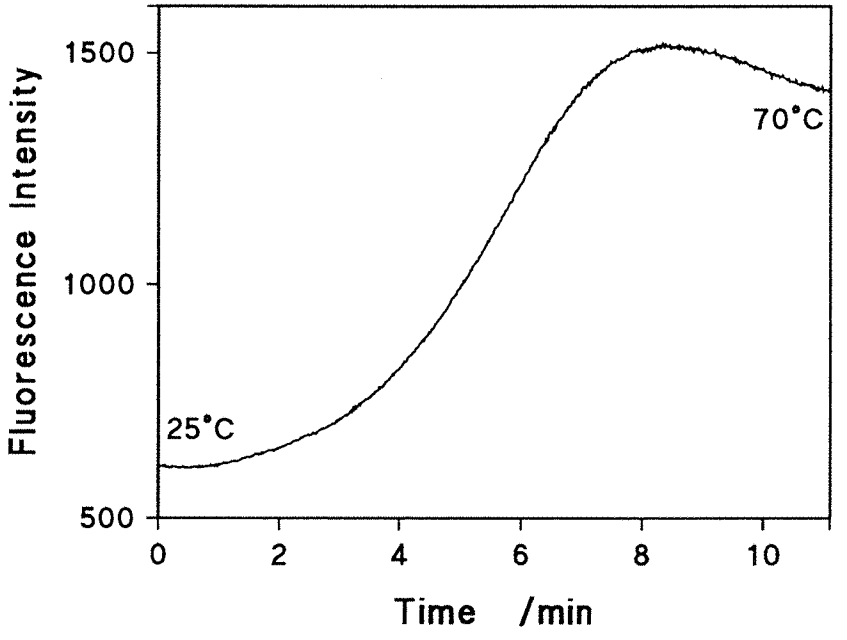


Figure 3a.25: Melting of 59μM (double stranded) d(GACGATAT[^{4H}C]GTC) followed by fluorescence (excitation at 300nm, emission at 360nm).

known, two important features can be seen in the figure. These are, (i) that the fluorescence increases up to a maximum of about 2½ times the original intensity, and (ii) that the fluorescence gradually decreases after reaching the maximum until the final temperature is reached. The increase in the fluorescence is due to the oligodeoxynucleotide duplex melting out as seen for other fluorescent nucleoside containing polynucleotides (Ward *et al.*, 1969). On top of this increase is a more gradual decrease in the fluorescence. This is caused by the rate of fluorescence quenching by solvent molecules becoming greater as the temperature is increased, due to their greater thermal motion increasing their collision rate with the oligodeoxynucleotide. This decrease in fluorescence only becomes apparent after the duplex has melted since it is far outweighed by the increase in fluorescence caused by the oligodeoxynucleotide melting. Strictly speaking, to get a true value of the decrease in fluorescence associated with duplex formation, this temperature dependent quenching by the solvent molecules should be taken into account. Also, to give a true measure of the quenching associated with duplex formation, the change of absorbance at 300nm with temperature should be taken into account. The absorbance at 254nm is increased by about 20% on going from 25°C to 70°C and the absorbance at 300nm is not expected to be much different. The fluorescence quenching caused by the duplex formation of d(GACGATAT[^{4H}C]GTC) can therefore be estimated to be roughly a decrease of 50%. This means that the fluorescent yield of melted dodecamer is about a tenth of that of the free deoxynucleoside and so the incorporation into a DNA chain, rather than duplex formation, is the dominant factor in the reduction of the fluorescence of d[^{4H}C] in the dodecamer d(GACGATAT[^{4H}C]GTC).

3a.4 CONCLUSIONS OF THE SYNTHESIS AND CHARACTERISATION OF THE BASE ANALOGUE CONTAINING OLIGODEOXYNUCLEOTIDES

In conclusion, all of the required deoxyguanosine and deoxycytidine analogues were synthesised and incorporated successfully into oligodeoxynucleotides in the required positions. The presence of all of the modified deoxynucleosides in their respective oligodeoxynucleotides was confirmed by base composition analysis. In addition to this, the presence of d[^{6H}G], d[^{6S}G], d[^{4H}C] and of d[^{3C}G] in their respective dodecamers was confirmed by the presence of their UV bands in the UV spectra of these oligodeoxynucleotides. Fluorescence spectroscopy of d(GAC[^{6H}G]ATATCGTC) and of d(GACGATAT[^{4H}C]GTC) also confirmed that they contained the modified bases d[^{6H}G] and d[^{4H}C].

Only the DNA synthesis of the d[^{4H}C] containing oligodeoxynucleotide presented any real problems. The high reactivity of the deoxynucleoside in the conditions of the ammonia deblock step after DNA synthesis reduced the yield by about 50% and also made the subsequent HPLC purification steps more complicated.

The d[^{6S}G] containing oligodeoxynucleotide was found to have a greater tendency to

form disulphides than the free deoxynucleoside. This should not cause a problem in most applications since this oxidation can easily be prevented by the inclusion of DTE in the oligodeoxynucleotide solution.

All of the dodecamers, with the exception of d(GAC[^{6H}G]ATATCGTC), were shown by their melting curves to be double stranded at 25°C and by their CD spectra to have essentially B-form DNA structures. The dodecamer d(GAC[^{6H}G]ATATCGTC) was concluded to be single stranded at 25°C on the basis of the following lines of evidence:- (i) Its melting curve showed no helix-to-coil transition in the temperature range 10° to 70°C. (ii) Its hyperchromicity of 1.19 is particularly low (the other dodecamers fall into the range 1.34 to 1.48) indicating less base stacking as would be expected for a random coil, single stranded oligodeoxynucleotide. (iii) Although not significant in itself, the CD spectrum of d(GAC[^{6H}G]ATATCGTC) has a slightly higher positive to negative peak ratio than the other dodecamers. An increase in this ratio has also been observed when other dodecamers have been heated past their melting temperatures. (iv) The reduction in the fluorescence of d[^{6H}G] on incorporation into the dodecamer is more in line with the fluorescence reduction seen for 2-aminopurine riboside present in thermally denatured ribocopolymers than the greater reduction seen when it is present in the double stranded ribocopolymers (Ward *et al.*, 1969). Without any further investigations, points (iii) and (iv) cannot be said to have any great significance in themselves but they do add to the other evidence that d(GAC[^{6H}G]ATATCGTC) is single stranded. By far the most important evidence comes from the melting curve of the oligodeoxynucleotide.

Because of the single stranded nature of d(GAC[^{6H}G]ATATCGTC) and the problems this would cause for any endonuclease studies, a longer d[^{6H}G] containing oligodeoxynucleotide was synthesised. The octadecamer d(GTCGAC[^{6H}G]ATATCGTCGAC) was synthesised and was shown by its melting curve to be double stranded at 25°C. Its CD spectrum was also typical of B-DNA. By way of a control for this d[^{6H}G] octadecamer, d(GTCGACGATATCGTCGAC) was synthesised. It too gave satisfactory T_M and CD data.

The tridecamer d(TGACGATATCGTC) was not characterised by CD spectroscopy or by its melting curve, but was expected to be essentially the same as the control dodecamer.

3b.1 ECO RV ENDONUCLEASE PREPARATION

3b.1.1 Purification of the Eco RV Endonuclease

The purification protocol of the Eco RV endonuclease was based on a method previously published (D'Arcy *et al.*, 1985), with a few minor changes. 20g of frozen cell paste, obtained as described previously (Newman, 1989) from an Eco RV overproducing *E. Coli* strain (Bougueleret *et al.*, 1985), was gently thawed out in a buffer containing 0.8M sodium chloride. Lower concentrations of sodium chloride in this and the subsequent sonication step have been found to cause pelleting of some of the Eco RV endonuclease with the cell debris when the cell suspension is spun down (Luke *et al.*, 1987). The protease inhibitors phenylmethylsulphonyl fluoride and benzamidine were included to prevent degradation of the endonuclease by native *E. Coli* proteases liberated during the sonication step. After the cells were disrupted by sonication, insoluble cell debris was removed by centrifugation at 20,000rpm. The yellow supernatant contains all the endonuclease which can be seen in the gel electrophoresis of the supernatant as a band at 29kDa amongst the many other protein bands (Figure 3b.1, lane 4). The supernatant was diluted with buffer to reduce the sodium chloride concentration to 0.2M before loading onto a pre-equilibrated phosphocellulose column. At this salt concentration, the

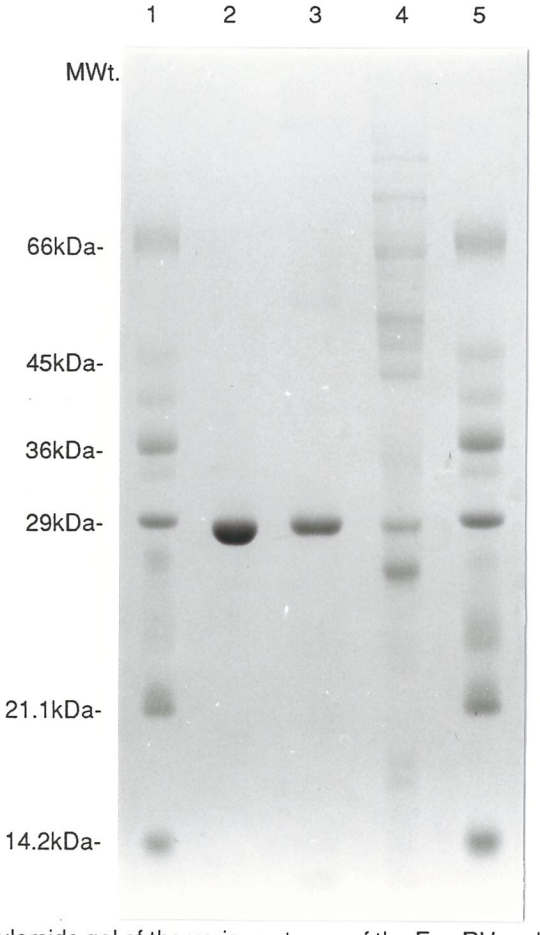


Figure 3b.1: Polyacrylamide gel of the various stages of the Eco RV endonuclease purification. Lanes 1 and 5, molecular weight markers; lane 2, after the gel filtration column; lane 3, after the phosphocellulose column; lane 4, cell supernatant.

CHAPTER 3B

ECO RV ENDONUCLEASE PURIFICATION AND ASSAYS OF ITS ACTIVITY

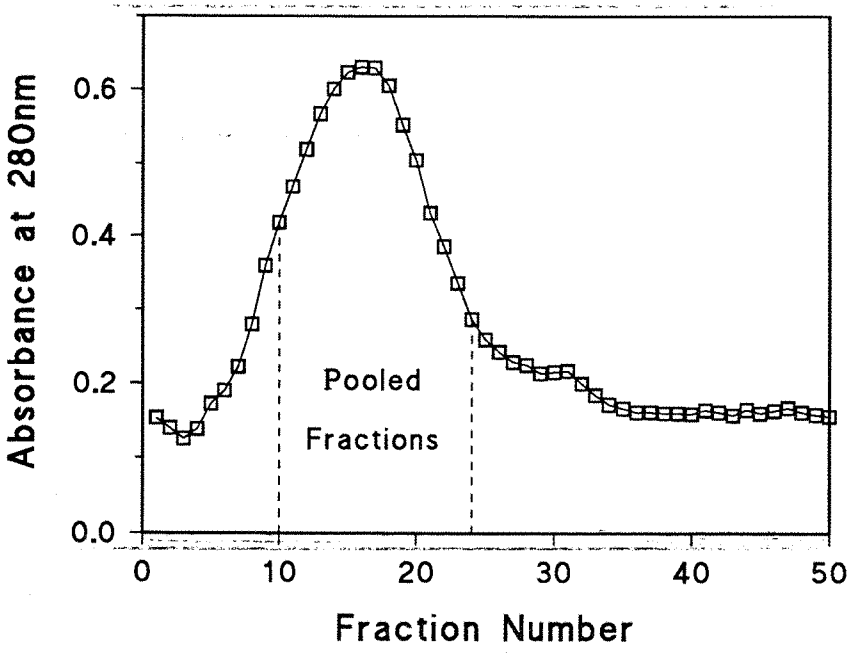


Figure 3b.2: Absorbance at 280nm of the 10ml fractions collected from the phosphocellulose column purification step of the endonuclease.

negatively charged phosphocellulose matrix does not bind any of the cellular nucleic acids or the majority of the cellular proteins and so these are washed through the column with the wash buffer. Proteins that bind to negatively charged compounds, including the Eco RV endonuclease, which binds DNA, were retained on the column during the washing process. These column bound proteins were then eluted off the column by increasing the sodium chloride concentration linearly from 0.2M to 0.8M over 500ml of buffer. Collected fractions of 10ml volume had their UV absorbance at 280nm measured to ascertain which of them contained protein. The results given in Figure 3b.2 show the presence of one main peak that eluted between about 0.3M and 0.6M sodium chloride. Further analysis of the fractions in this peak by SDS gel electrophoresis showed that the peak was virtually all endonuclease together with some minor contaminants. On the basis of the SDS gel, fractions 10 to 24 were pooled to give the best possible yield with minimum contamination. The endonuclease was precipitated from the solution with 80% (w/v) ammonium sulphate to give a preparation that was judged by SDS gel electrophoresis to be >95% pure (Figure 3b.1, lane 3). The enzyme at this stage contained just one main contaminating protein that ran on the SDS gel with a molecular mass of greater than 14kDa. This is presumably the 12kDa protein observed after phosphocellulose column purification of the Eco RV endonuclease by others (Luke *et al.*, 1987; Newman, 1989).

The endonuclease was further purified by gel filtration. A matrix of Sephadex G-75-50 (Sigma), which has a recommended fractionation range of 3kDa to 70kDa, was used as the

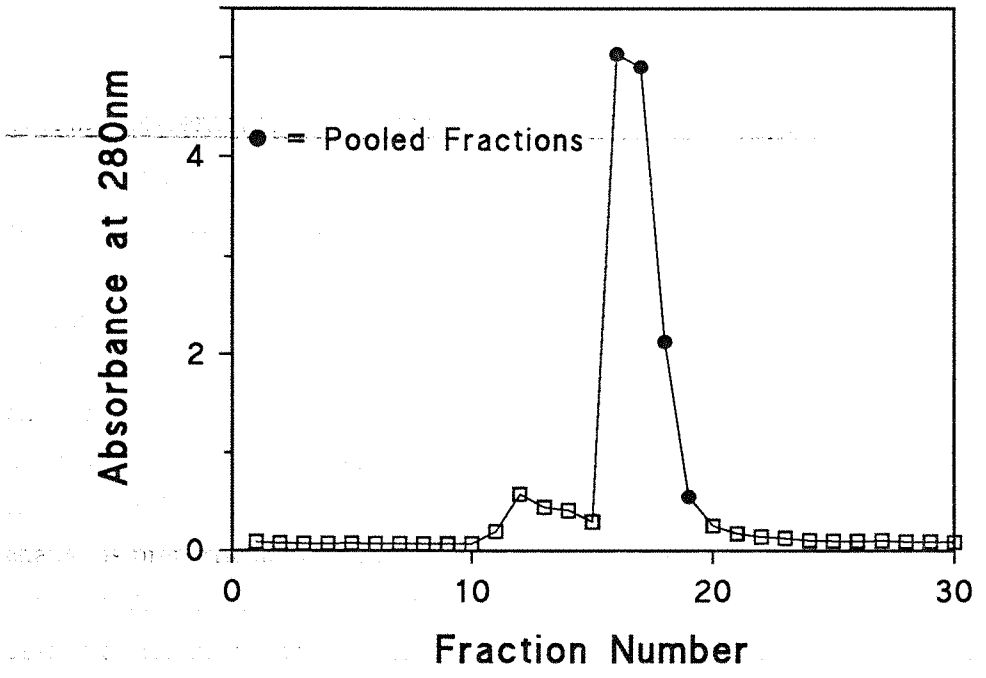


Figure 3b.3: Absorbance at 280nm of the 10ml fractions collected from the gel filtration column purification of the endonuclease.

stationary phase. 10ml fractions were collected from the column and were initially analyzed for protein content by measuring their absorbance at 280nm. The results are shown in Figure 3b.3 which shows one main sharp peak proceeded by one much smaller peak. Analysis of the fractions by SDS polyacrylamide gel electrophoresis revealed that the main peak was the endonuclease. The small, earlier eluting peak consisted mostly of a protein with a molecular mass of >66kDa together with some of the endonuclease. The gel showed that fractions 16 to 19, which contained the vast majority of the endonuclease, were uncontaminated and so these were pooled. The volume of the pooled fractions was reduced to about 0.5ml using centriprep tubes. 0.5ml of cold glycerol was added to the 0.5ml of concentrated endonuclease. 1ml of an 8 times dilution of this enzyme solution was made up in buffer ξ containing 30% glycerol. The two enzyme solutions were stored at -20°C.

SDS polyacrylamide gel electrophoresis of the final endonuclease solution showed the enzyme to be >98% pure as shown in Figure 3b.1 (lane 2). Luke *et al.* (1987) suggested replacing the gel filtration step with a blue agarose gel purification since they found it the best matrix for removing the 12kDa protein contamination present after the phosphocellulose column purification. However, the gel filtration column proved adequate here since no detectable band was seen at 12kDa in the SDS gel of the final enzyme preparation. Also important from the point of view of the enzyme kinetic studies is that the endonuclease preparation should contain no nuclease contamination. This was found to be the case here since incubation of the d(GACGATATCGTC) dodecamer produced only a single cut at the

expected Eco RV site. Also, incubation of oligodeoxynucleotides that are not Eco RV substrates with high levels of the enzyme (~8μM) gave no detectable hydrolysis of the DNA.

3b.1.2 Concentration Determination of the Eco RV Endonuclease

In this, and all other following sections, endonuclease concentrations are given for the enzyme *dimer* since this is the active form of the enzyme.

Two different methods were used to determine the concentrations of the two stock endonuclease solutions. One of these methods was to measure the UV absorbance of the enzyme solutions. This relies on the accuracy of the molar extinction coefficient of the Eco RV endonuclease which has been calculated as $104{,}400 M^{-1}cm^{-1}$ (for the dimeric enzyme) at 280nm (D'Arcy *et al.*, 1985). The second method, a Bio-Rad protein assay, is based on the colour change of a dye that occurs on binding to a protein (Bradford, 1976). The magnitude of the change is proportional to the amount of protein present and so, when the colour change has been standardised against samples of known protein concentration, the colour change can be used to determine the protein concentration of an unknown. Ideally the standardising protein solutions of known concentration should be of the same protein as that whose concentration is being determined. When this is not possible, as in the case here for the Eco RV endonuclease, other proteins such as BSA can be used successfully as the standard.

The concentrations of the two stock solutions, made as described in section 2.3.5, were

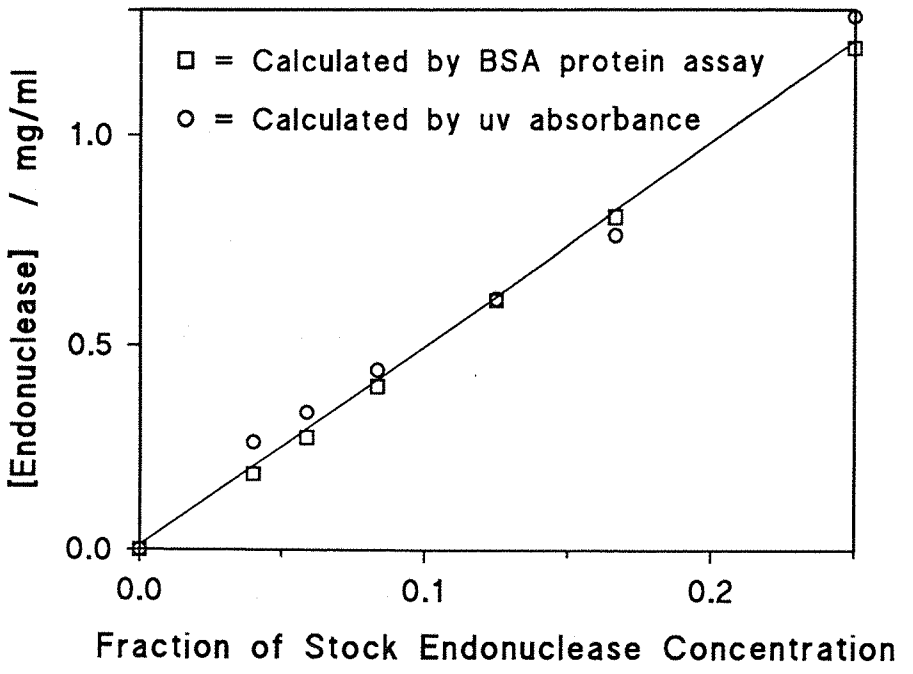


Figure 3b.4: Concentration determination of the endonuclease by its absorbance at 280nm (○) and by the Bio-Rad assay (□).

determined by their UV absorbance and with the Bio-Rad protein assay kit. There was reasonably good agreement between the two methods for the more concentrated stock solution with the UV absorbance giving a concentration of 672μM and the protein assay giving a concentration of 695μM. Taking an average of these two values gives a concentration of 683μM. 1ml of this solution was purified from 20g of the cell paste and so the yield was 34nmol, or 2mg, per gram of cell paste. This is of similar yield to the 2-4mg/g reported by D'Arcy *et al.* (1985) and the 1.5mg/g reported by Luke *et al.* (1987).

The more dilute stock solution was required for the kinetic studies and so the concentration needed to be determined more accurately. This was done by determining the concentration of six different dilutions of the enzyme by both methods. There was excellent agreement between the two methods in this case, as shown in Figure 3b.4 with the UV absorbance giving a concentration of 83.63μM and the Bio-Rad method giving 83.64μM. Combination of all the sets of data gave an enzyme concentration of 83.6 (±0.8)μM. To avoid variations caused by using different enzyme preparations, only this endonuclease solution was used for the enzyme kinetics.

3b.2 ASSAYS OF ECO RV ENDONUCLEASE ACTIVITY

Previously two main assays have been used for measuring endonuclease activity. One of these is based on the separation of products and substrate by reverse phase HPLC and the other is based on the separation of radiolabelled products and substrate by polyacrylamide gel electrophoresis or homochromatography.

3b.2.1 HPLC Assay

The dodecamer substrate, d(GACGATATCGTC) can be easily separated from its two hexameric Eco RV cleavage products, d(GACGAT) and d(pATCGTC), by reverse phase HPLC. A representative trace of such a cleavage reaction of d(GACGATATCGTC) is shown in Figure 3b.5. The endonuclease produces 5'-phosphorylated d(pATCGTC) from the cleavage reaction but the removal of the 5'-phosphate with alkaline phosphatase was found to greatly improve the resolution of this peak. After removal of this phosphate the oligodeoxynucleotides elute in the order d(GACGAT), d(ATCGTC) and then d(GACGATATCGTC).

This type of HPLC set-up can be used for determining the rate of the endonuclease reaction (Fliess *et al.*, 1988; Newman *et al.*, 1990a). This involves removing samples of an endonuclease reaction, quenching them by heat denaturation or with concentrated alkali, and then viewing them by HPLC. The extent of the cleavage reaction is determined for each sample from the relative areas of one of the product peaks (the other product is produced in equal amounts) and that of the substrate peak. To allow for the different extinction coefficients of the

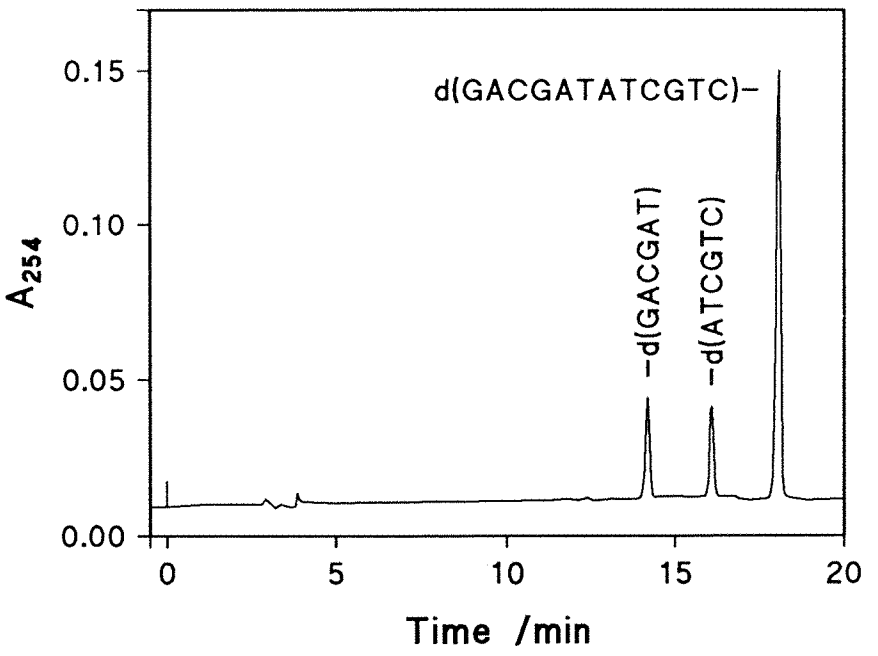


Figure 3b.5: HPLC trace of the partial cleavage of d(GACGATATCGTC) by the Eco RV endonuclease (after alkaline phosphatase treatment). HPLC gradient is 0 to 15% buffer B over 20 minutes.

substrate and the products that gives rise to different responses of the HPLC detector, the detector is calibrated for each of the peaks by injecting samples of known amounts of each oligodeoxynucleotide. If the total amount of *double* stranded substrate present initially was T and the amount of *single* stranded product formed after a certain amount of time, t, is P then the amount of *double* stranded substrate left at time t is (T-P/2). Thus, the ratio of relative amount of single stranded product formed (P_R) to relative amount of double stranded substrate left (S_R) after time t, found by HPLC integration is:-

$$P_R/S_R = P/(T-P/2)$$

Rearrangement of this equation gives the amount of single stranded product, P, formed after time t in terms of the known quantities T, S_R and P_R:-

$$P = 2TP_R/(2S_R+P_R)$$

The HPLC assay is particularly useful for qualitative examination of endonuclease cleavages of oligodeoxynucleotides since the products and the substrate are easily separable and can be viewed directly. Therefore any non-specific cleavage of the substrate that may occur would be clearly evident. This method is also useful for the isolation of the products from a cleavage reaction for characterisation purposes. As well as this, complete digestion of a

substrate would reveal any residual non-cleavable oligodeoxynucleotide contaminant that may normally co-elute with the substrate. Although the HPLC assay has been used successfully for determining the rates of endonuclease reactions (Fliess *et al.*, 1988; Newman *et al.*, 1990a), it has several disadvantages that limit its usefulness as a quantitative method. These include the limited sensitivity of the UV detection system and the poor accuracy of the technique of peak area determination. Another major disadvantage of the this assay is that it is very time consuming. A single rate determination requiring say, ten time points which each have to be individually injected on to the HPLC column, would take about five hours.

3b.2.2 Polyacrylamide Gel Electrophoresis Assay

Another method of separating the products from the dodecamer substrate is polyacrylamide gel electrophoresis. A polyacrylamide gel run in the presence of urea separates DNA on the basis of DNA length and by using a suitable percentage of polyacrylamide (~17%), the hexameric products can be well resolved from the dodecamer starting material (Newman *et al.*, 1990b). By using an oligodeoxynucleotide substrate that is radiolabelled with a 5'-^{32}P phosphate group, the substrate and the 5'-end cleavage product that would also be labelled, can be detected by autoradiography. A typical autoradiogram for the hydrolysis of [^{32}P]d(pGACGATATCGTC) by the Eco RV endonuclease is shown in Figure 3b.6. After chromatographic separation of reaction samples removed at various times, the relative amount of single stranded product (P) can then be determined by scintillation counting using the formula:-

$$\text{single stranded product, } P = 2TP_C/(S_C+P_C)$$

Where P_C and S_C are the number of counts for product and substrate respectively and T is the total amount of *double* stranded substrate initially present.

The polyacrylamide gel electrophoresis method has been used for studying the Eco RV endonuclease (Newman *et al.*, 1990b) and the Hind II, Sal I and Taq I endonucleases (Jiricny *et al.*, 1986). Another analogous method for separating and determining the relative amounts of radioactive substrate and products is chromatography on DEAE-cellulose TLC plates. This separation technique also separates on the basis of DNA length and has been used successfully for studying the oligodeoxynucleotide cleavage reactions of the Eco RI (Brennan *et al.*, 1986) and the Bgl II, Sau 3AI and Mbo I (Ono *et al.*, 1984; Ono & Ueda, 1987a) restriction endonucleases.

The gel assay has several advantages over the HPLC method for determining hydrolysis rates. Most importantly its accuracy and sensitivity are far superior to those of the HPLC assay. Another important advantage of the gel assay is that several samples can be run simultaneously thus cutting down the time needed for determining a rate. Its major drawback

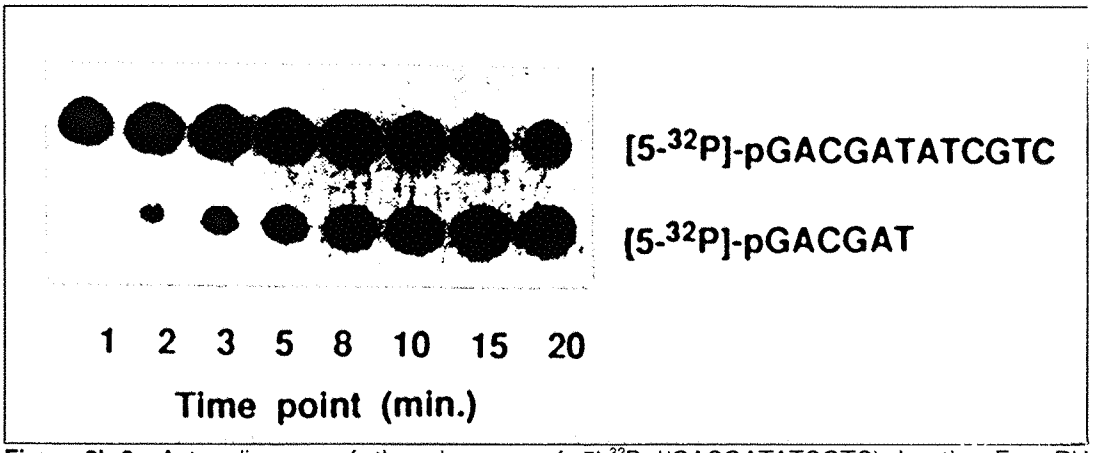


Figure 3b.6: Autoradiogram of the cleavage of 5'-^{32}P-d(GACGATATCGTC) by the Eco RV endonuclease (from Newman *et al.*, 1990b).

however, is the need for a ^{32}P label and the hazards associated with the phosphorylation, handling and purification of the radiolabelled oligodeoxynucleotide. Again, as for the HPLC assay, a major disadvantage is the fact that it is a discontinuous assay.

3b.2.3 Continuous Ultraviolet Absorbance Assay

As discussed in the section on the melting curves of DNA [3a.3.3], double stranded DNA has a lower UV absorbance than it does when single stranded. This hyperchromic effect is associated with the greater base stacking, which reduces the UV absorbance, that occurs in double stranded DNA (Saenger, 1984). The cleavage of an oligodeoxynucleotide by a restriction endonuclease leads to the formation of shorter oligodeoxynucleotide products. If the starting oligodeoxynucleotide substrate is from about 10 to 16 base pairs long, then the products formed will often be short enough (<8 bases) to be single stranded at 25°C. The cleavage of such an oligodeoxynucleotide therefore results in the production of single stranded products from a double stranded substrate and because of the hyperchromic effect, the formation of the single stranded products is accompanied by an increase in the UV absorbance. A reaction of this type can therefore be followed by the increase in UV absorbance with time. The viability if this type of UV assay for measuring the rate of hydrolysis of d(GACGATATCGTC) by the Eco RV endonuclease was investigated. The UV assay was then developed as a technique for assaying enzyme activity and for carrying out Michaelis-Menten kinetics on d(GACGATATCGTC) and its base analogue containing equivalents (Waters & Connolly, 1991b).

3b.3.3.i Initial Rate of Cleavage of d(GACGATATCGTC) by the Eco RV Endonuclease

The hydrolysis of d(GACGATATCGTC) by Eco RV endonuclease results in the two hexameric products d(GACGAT) and d(pATCGTC). The dodecamer has a T_M of 53°C [section 3a.3.3] and so at 25°C it is double stranded. The two products are only six bases long and as such are not expected to be double stranded at 25°C. The cleavage of the dodecamer by the endonuclease should therefore be accompanied by an increase in the UV absorbance of the

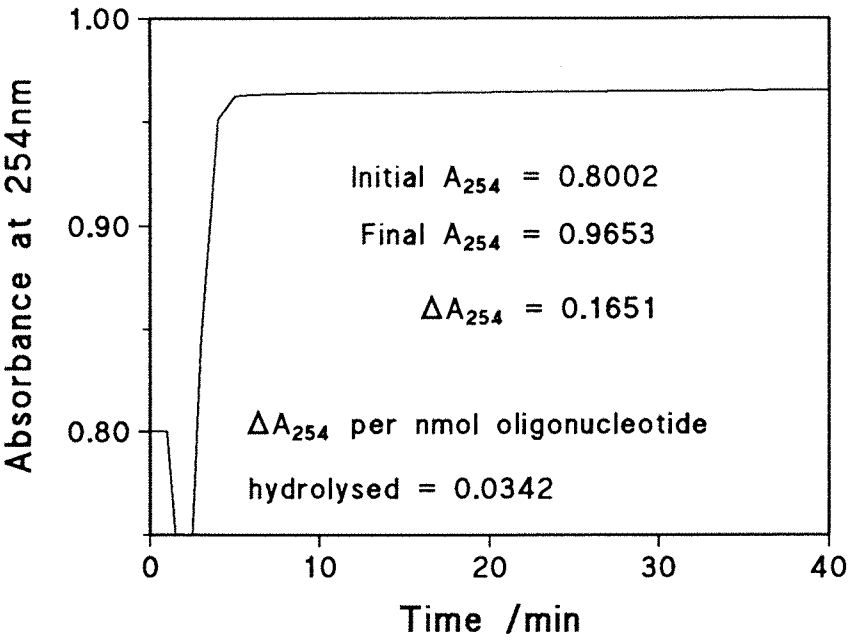


Figure 3b.7: Change in the absorbance at 254nm for the complete cleavage of d(GACGATATCGTC) by the Eco RV endonuclease.

reaction mixture because of the hyperchromic effect. This was indeed found to be the case. The complete cleavage of d(GACGATATCGTC) by approximately 85pmol (85nM) of endonuclease resulted in an increase in the UV absorbance of the reaction solution, in a 1cm path length cell, of about 20% as shown in Figure 3b.7. The maximum increase in the UV absorbance of the reaction is associated with the total cleavage of the dodecamer as proved by HPLC which showed just the two product peaks [HPLC as described in 3b.2.1]. Since the increase in UV is directly proportional to the amount of dodecamer cleaved, the value of this increase can be related to the change in absorbance at 254nm that occurs per nmol of substrate cleaved. In the example shown in Figure 3b.7, the initial absorbance was 0.8002 and the final absorbance was 0.9653. The extinction coefficient of d(GACGATATCGTC) at 254nm is 166 x $10^3 M^{-1}cm^{-1}$ [section 3a.3.2]. Therefore:-

$$\text{The initial concentration of double stranded d(GACGATATCGTC)} = A/\varepsilon_{254}$$
$$= 0.8002/(166 \times 10^3) = 4.82\mu M$$

The reaction volume was 1ml and so the initial amount of d(GACGATATCGTC) present was 4.82nmol of double strands. The total cleavage of the dodecamer therefore produced a total of 9.64nmol of single strands for each product. Therefore the change in the absorbance associated with each nmol of single stranded product formed:-

$\Delta A_{254} = (A_{initial} - A_{final})$/nmol of product formed = (0.9653-0.8002)/9.64

= **17.1 x 10^{-3}** OD units per nmol of single stranded product formed

This value was termed the endonuclease hyperchromicity factor and it can be used to convert absorbance changes into actual amounts of product formed.

By using less endonuclease than that used in Figure 3b.7, the initial rate of hydrolysis of d(GACGATATCGTC) by the endonuclease could be determined. 6.8nmol (6.8μM) of control dodecamer was incubated at 25°C with 10pmol (10nM) endonuclease in 1ml of buffer α10 (10mM $MgCl_2$, 100mM NaCl, 50mM HEPES, pH7.5). The reaction was performed in a 1cm path length cell and was followed by the UV absorbance at 254nm as a function of time. The UV trace obtained in this way was converted into nmol of single stranded product formed against time by dividing the change in UV absorbance ($A_{initial}$ - $A_{at\ time\ t}$) by the factor given above (17.1 x 10^{-3} OD units per nmol). To express the reaction in terms of nmol of double stranded substrate consumed, this is simply divided by two. By way of a comparison, samples of the same reaction were removed at various times and then quenched for analysis by the HPLC assay [section 3b.2.1]. The HPLC peak areas were converted into amounts of product formed as described. The results of the two different assay methods of the same reaction were superimposed and are shown together in Figure 3b.8. The graph shows very good agreement between the two techniques. The UV assay is obviously superior since it has far more data points than the HPLC method (about 500 compared with 11!) and it is also far less erratic than the HPLC data. The variability of the HPLC points could be due to inconsistent efficiency in the quenching of the samples or it could be due to the errors inherent in the determination of the HPLC peak areas. These results show that the UV assay can be used successfully to determine rates of oligodeoxynucleotide cleavage by endonucleases. It also highlights some of the advantages of this assay over other methods. Firstly, it provides a way of continuously monitoring a reaction and it therefore eliminates the inaccuracies associated with a discontinuous method of removing samples and quenching them. These inaccuracies include determination of the exact time of removal and the effectiveness of the enzyme quenching step. Secondly, it is much quicker to perform, taking no longer than the time the reaction itself takes. In contrast, the HPLC assay for the reaction shown in Figure 3b.8 took over five hours to perform. Assaying samples by polyacrylamide gel electrophoresis is quicker than the HPLC method but still takes considerably longer than the UV assay. Thirdly, the reaction is monitored in real time. This can be a great advantage especially if something goes wrong with the reaction. With both the HPLC and polyacrylamide gel electrophoresis assays this would not be discovered until after the work and the time needed to perform the assays had been put in. Finally, the UV assay can monitor much smaller time intervals of a reaction than either of the other methods. The limitation at the moment in this respect is the time it takes to mix the reactants properly. With a suitable mixing set-up, the UV assay should find applicability as a means of monitoring transient kinetic reactions.

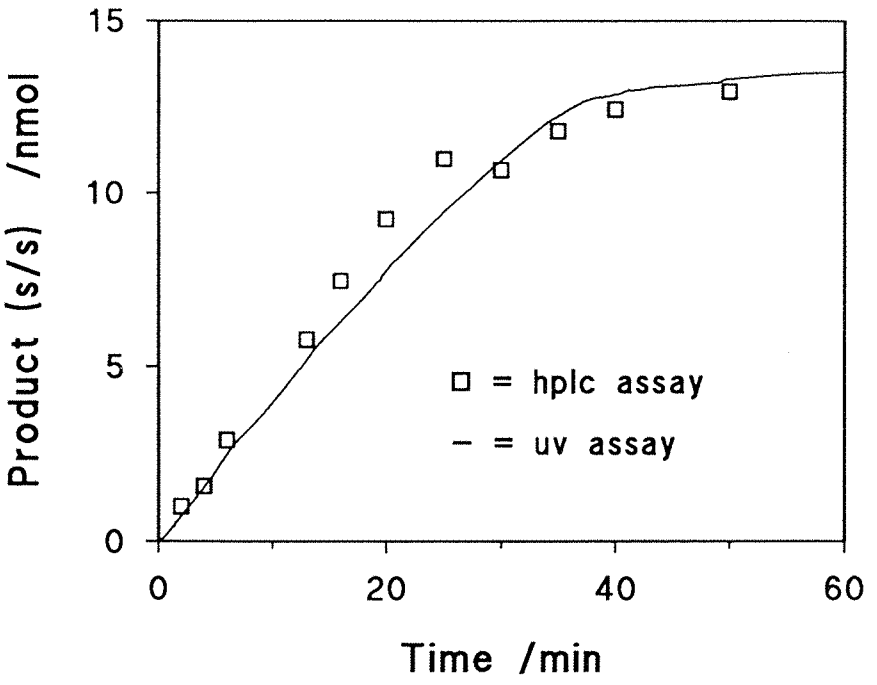


Figure 3b.8: Comparison of the Eco RV endonuclease hydrolysis of d(GACGATATCGTC) assayed by HPLC (□) and by the UV absorbance change (—). s/s = single stranded.

The UV assay was used to study various aspects of the Eco RV endonuclease reaction including the Michaelis-Menten kinetics of the analogue containing dodecamers. These experiments are discussed below and in the following chapter. During the development of this assay, two precautions were found to be important for obtaining accurate and true rates. Most important of these, especially for measuring slow rates, was to ensure that the temperature of the DNA solution in the cuvette was equilibrated with that of the cell housing before initialising the reaction by the addition of enzyme. Small temperature changes were found to cause small but noticeable changes in the UV absorbance which could have a relatively large effect on the evaluation of slow reaction rates. To ensure that the temperature had reached equilibrium, the UV absorbance was monitored until it no longer changed before adding the endonuclease. The second precaution was to siliconise the cuvettes as described in section 2.4.3.iv. This was to prevent binding of the DNA to the glass and is particularly important for low concentrations of oligodeoxynucleotide.

3b.2.3.ii Stability of the Dilute Eco RV Endonuclease

The stability of the dilute enzyme (100nM) at 0°C was investigated. Two separate dilute solutions were made up; one by diluting stock endonuclease with buffer α25 (25mM $MgCl_2$, 55mM NaCl, 50mM HEPES, pH7.5) and the other by diluting with buffer α25 containing 1mg/ml BSA. In this way, the possible stabilising effects of BSA could also be investigated. The dilute solutions were kept on ice and their activities were assayed by

measuring their rates of cleavage of 2μM d(GACGATATCGTC) at 25°C in buffer α25 using the UV assay. Enzyme activities were determined in this way at zero time, after 1 hour, 2.5 hours and 6 hours on ice.

The results are shown graphically in Figure 3b.9. It can clearly be seen from this that the presence of the BSA in the dilution buffer does indeed stabilise the dilute enzyme with only about 6% activity of the enzyme being lost over 6 hours at 0°C whereas over 28% is lost when the enzyme is diluted with buffer containing no BSA. The 6% decrease in activity for the enzyme in BSA is acceptable from the point of view of experiments such as Michaelis-Menten kinetic determinations since these experiments should take no more than an hour, corresponding to only 1% loss in activity of the enzyme. Perhaps more importantly, the results show a considerably lower (~80%) initial activity for the enzyme diluted with non-BSA buffer compared with that diluted with BSA containing buffer. This indicates that the BSA helps to stabilise the protein during the dilution process. No cleavage of the dodecamer due to possible contaminating nucleases in the BSA was observed when the oligodeoxynucleotide was incubated with the BSA containing buffer in the absence of endonuclease.

Halford and Goodall (1988) have used 2-mercaptoethanol in the endonuclease dilution buffer as a precautionary stabilising measure. The theory behind using 2-mercaptoethanol to help stabilise the endonuclease was that it prevented the single cysteine known to be present in the enzyme (Cys20), from being oxidised. Chemical modification of this cysteine has been shown to totally inactivate the enzyme (Luke *et al.*, 1987). However, under the conditions used

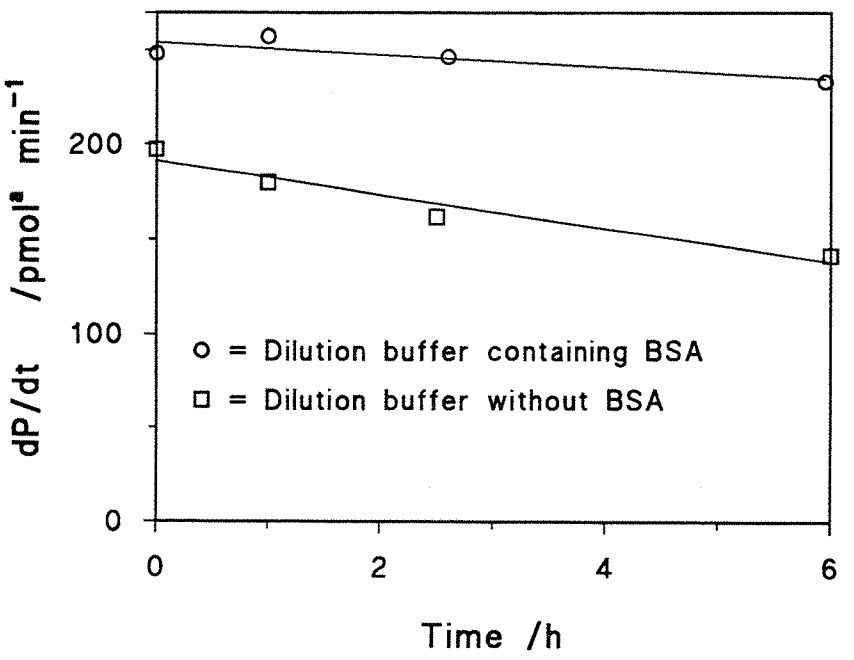


Figure 3b.9: Stability of dilute Eco RV endonuclease (100nM) at 0°C in buffer α25 containing BSA (○) and buffer α25 lacking BSA (□). [a]Single stranded product.

here it has been found that the use of 2-mercaptoethanol causes no noticeable change in the stability of the endonuclease (Newman, 1989). Also, it has since been found that Cys20 can be mutated to a serine without loss of activity (Halford, S.E.; personal communication).

3b.2.3.iii Dependence of the Rate of d(GACGATATCGTC) Cleavage on the Endonuclease Concentration

The dependence of the rate of the endonuclease cleavage of 3μM d(GACGATATCGTC) was investigated as a function of enzyme concentration over the range 2nM to 100nM. Rates of each reaction were determined at 25°C using the UV assay. The rate lines obtained are shown in Figure 3b.10 and the rates of the initial linear portions of each reaction were determined by linear regression. These rates were plotted against enzyme concentration and the graph obtained is shown in Figure 3b.11. This graph clearly shows an excellent linear dependence of the reaction rate upon the endonuclease concentration. The gradient of this graph provides a good value for the apparent turnover number at this substrate concentration (the real turnover number is determined at saturating substrate concentration). The value obtained was $26min^{-1}$, or 448nmol of double stranded dodecamer cleaved per minute per mg of endonuclease dimer. This is almost four times the value of 115nmol/min/mg obtained by Newman (1990a) for the same oligodeoxynucleotide. Even taking into account the different conditions used (25mM Mg^{2+} here and 10mM Mg^{2+} used by Newman *et al.*) the turnover

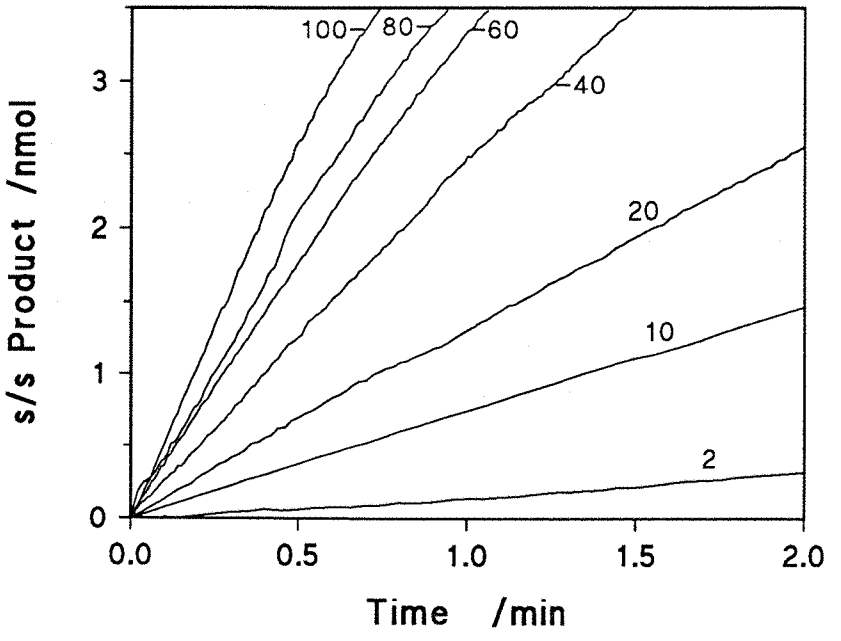


Figure 3b.10: Rate lines of the Eco RV endonuclease cleavage of d(GACGATATCGTC) at various enzyme concentrations determined using the UV assay. Concentrations of enzyme indicated for each line are in nM. s/s = single stranded.

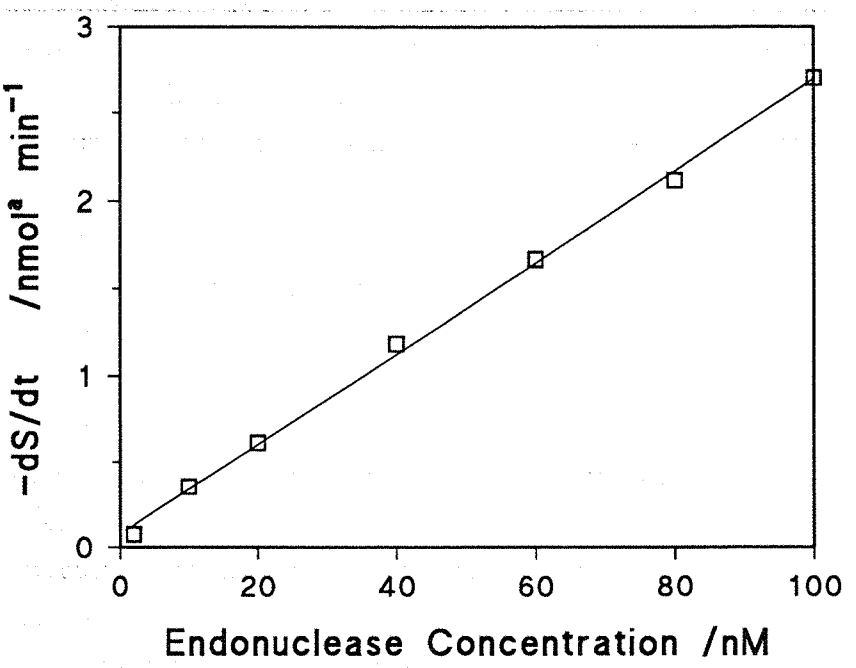


Figure 3b.11: Dependence of the rate of cleavage of d(GACGATATCGTC) by the Eco RV endonuclease on the enzyme concentration. [a]Double stranded substrate.

number is still considerably larger. This would suggest a more active preparation of endonuclease used in these studies or perhaps reflects the use of BSA in the dilution buffer used here. This increased activity is discussed in a later section in terms of the real turnover number.

Since the rate of cleavage was found to be linearly related to enzyme concentration it became obvious that the UV assay could be used to measure the concentration of endonuclease in an unknown sample, such as the fractions collected from a column purification of the enzyme. To test this, fractions collected from a phosphocellulose column purification of the endonuclease [as in section 2.3.4] were assayed using the control dodecamer and measuring the rate of oligodeoxynucleotide cleavage by UV. 5µl of each fraction was reacted with 2.25nmol (4.5µM) of d(GACGATATCGTC) in 0.5ml of buffer α10 and the rates of cleavage determined by the UV method. As a comparison, the UV absorbance at 280nm of some of the fractions was also measured. The results for both the rate determinations and the 280nm absorbances are given in Figure 3b.12. The UV assay should in theory provide a better way of assessing which fractions contain endonuclease than the 280nm absorbance since only endonuclease should show up in the former whereas all UV absorbing material would show up in the latter. This is born out by the graph which shows a much sharper peak and a greatly reduced background level for the cleavage assay than for the 280nm absorbance. The UV assay also has the advantage of speed over the other method normally used for assessing which fractions contain endonuclease, SDS polyacrylamide gel electrophoresis, taking less than an hour to perform.

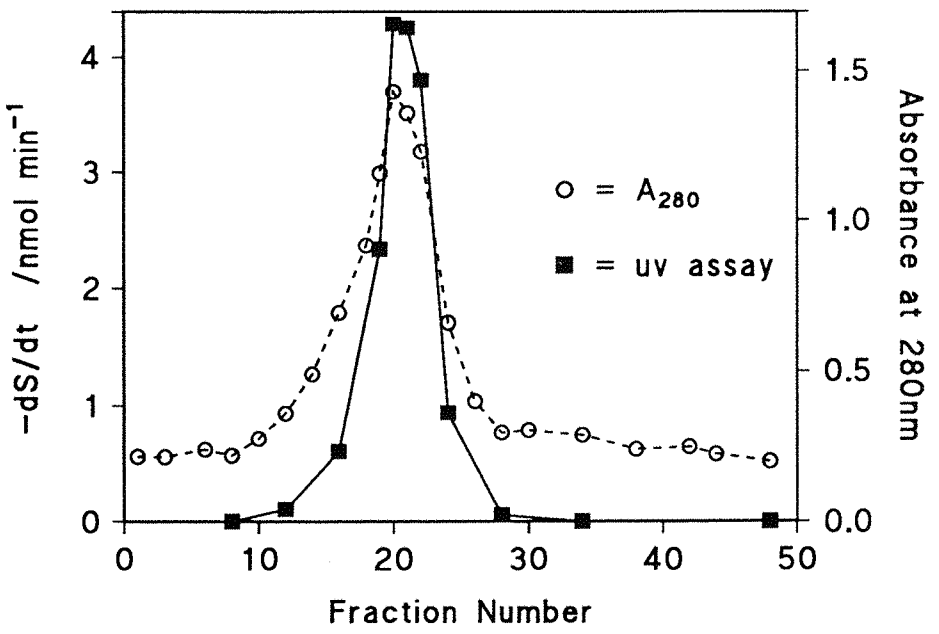


Figure 3b.12: UV assay of the Eco RV endonuclease activity of fractions collected from a phosphocellulose column purification of the enzyme measured by the rate of cleavage of d(GACGATATCGTC) (■). Their absorbance at 280nm is also given (O).

CHAPTER 3C

KINETICS OF THE ECO RV ENDONUCLEASE

3c.1 DETERMINATION OF WHICH BASE ANALOGUE CONTAINING OLIGODEOXYNUCLEOTIDES ARE ENDONUCLEASE SUBSTRATES

All analogue containing oligodeoxynucleotides were tested to see whether or not they were substrates for the Eco RV endonuclease. 20μM solutions of each oligodeoxynucleotide were incubated with up to 1.7nmol (8.5μM) endonuclease at 25°C. At various times after the enzyme addition, samples were removed from the mixture and viewed by HPLC [see section 3b.2.1]. The HPLC separated the two products from each other and from the starting material for all oligodeoxynucleotide substrates. Alkaline phosphatase, which removes the 5'-phosphate from one of the products, (d(pATCGTC)), was included in the reaction mixture to improve peak resolution. In all of the dodecamer cleavages, the hexameric products eluted in the order d(GACGAT) first and d(ATCGTC) second (or the analogue containing equivalents) with the substrate eluting considerably later [for example see Figure 3b.5, section 3b.2.1]. In the case of d(TGACGATATCGTC), the heptameric product d(TGACGAT) eluted just after the other hexameric product and for the control octadecamer the 5' nonameric product eluted after the 3' product. The results for the cleavage tests are given in Table 3c.1. Those analogues that are designated as non-substrates are oligodeoxynucleotides that were not cleaved under the conditions described in 2.4.2. The HPLC method can detect cleavage of as little as 1% of an oligodeoxynucleotide substrate. Oligodeoxynucleotides that have no detectable product after 20 hours are therefore cleaved at a rate of less than 33fmol/min. Non-substrates were exposed to 1.7nmol of endonuclease for this time and so this corresponds to a turnover of less than $2x10^{-5}min^{-1}$. This is more than a million fold less than that of the control oligodeoxynucleotide and is the kind of rate reduction seen with non-cognate Eco RV endonuclease substrates (Taylor & Halford, 1989). Importantly, even under these conditions of high endonuclease concentration, no non-specific cleavage of the oligodeoxynucleotides that were not Eco RV endonuclease substrates was detected as typified by the HPLC trace in Figure 3c.1. This shows that no non-specific nucleases are present in the endonuclease preparation.

Of all of the oligodeoxynucleotides tested only three were non-cleavable; the d[^{6H}G] containing dodecamer and octadecamer, and the d[^{7C}G] containing dodecamer. The lack of cleavage of the d[^{6H}G] containing dodecamer is not surprising since it has been shown to be single stranded at the assay temperature and is therefore not expected to be a substrate for the Eco RV endonuclease which requires double stranded DNA. However, the d[^{6H}G] containing octadecamer was also not cleaved, even though its melting curve showed it to be double stranded at 25°C. Also, its CD spectrum showed it to be B-form DNA. Its resistance to endonuclease cleavage must therefore be a direct result of the presence of the modified base. The same must also be true for the d[^{7C}G] containing dodecamer since it too has CD and T_M data consistent with a double stranded, B-form oligodeoxynucleotide. These results show that both the 6-carbonyl oxygen and the ring 7-nitrogen of the deoxyguanosine in the Eco RV recognition site are very important in the endonuclease cleavage reaction. The crystal structure of a cognate complex does indeed show hydrogen bonds from the endonuclease to the 6-

Table 3c.1: Table showing which of the oligodeoxynucleotides were Eco RV substrates. The base compositions and the inferred cleavage products from these compositions are also given for the substrates. [a]Not double stranded at 25°C. [b]Not determined.

Oligodeoxynucleotide	Substrate ?	Base Composition of Products	Inferred Products
d(GACGATATCGTC)	yes	$C_1\ G_2\ T_1\ A_2$	d(GACGAT)
		$C_2\ G_1\ T_2\ A_1$	d(pATCGTC)
d(GAC[^{6H}G]ATATCGTC)	no[a]	—	—
d(GAC[^{6S}G]ATATCGTC)	yes	$C_{1.1}\ G_{1.1}\ T_{1.1}\ A_{2.0}\ [^{6S}G]_{0.9}$	d(GAC[^{6S}G]AT)
		$C_{1.9}\ G_{1.1}\ T_{2.1}\ A_{1.0}$	d(pATCGTC)
d(GAC[^{7C}G]ATATCGTC)	no	—	—
d(GAC[I]ATATCGTC)	yes	$C_{1.1}\ G_{1.0}\ T_{1.0}\ A_{1.7}\ [I]_{1.3}$	d(GAC[I]AT)
		$C_{2.1}\ G_{1.0}\ T_{2.0}\ A_{0.9}$	d(pATCGTC)
d(GAC[^{3C}G]ATATCGTC)	yes	$C_{1.1}\ G_{1.1}\ T_{1.0}\ A_{2.0}\ [^{3C}G]_{0.8}$	d(GAC[^{3C}G]AT)
		$C_{2.1}\ G_{1.0}\ T_{1.9}\ A_{0.9}$	d(pATCGTC)
d(GACGATAT[^{5Me}C]GTC)	yes	$C_{1.1}\ G_{2.0}\ T_{1.0}\ A_{1.9}$	d(GACGAT)
		$C_{1.0}\ G_{1.1}\ T_{2.0}\ A_{0.9}\ [^{5Me}C]_{1.0}$	d(pAT[^{5Me}C]GTC)
d(GACGATAT[^{4H}C]GTC)	yes	$C_{1.1}\ G_{2.1}\ T_{1.0}\ A_{1.9}$	d(GACGAT)
		$C_{1.1}\ G_{1.1}\ T_{2.0}\ A_{0.9}\ [^{4H}C]_{1.0}$	d(pAT[^{4H}C]GTC)
d(TGACGATATCGTC)	yes	n/d[b]	—
d(GTCGACGATATCGTCGAC)	yes	$C_{2.0}\ G_{3.0}\ T_{2.0}\ A_{1.9}$	d(GTCGACGAT)
		$C_{3.1}\ G_{2.0}\ T_{2.0}\ A_{1.9}$	d(pATCGTCGAC)
d(GTCGAC[^{6H}G]ATATCGTCGAC)	no	—	—

carbonyl oxygen and the 7-nitrogen of the deoxyguanosine in the Eco RV recognition site (Figure 3c.2). However, more than the simple deletion of a hydrogen bond by the substitution of these base analogues must be involved since such a deletion should not decrease the rate of cleavage so dramatically that no cleavage is detectable. Previous results for hydrogen bond deletions have given relative k_{cat}/K_M values in the range 4% to 40% per deletion (Fersht, 1987). Therefore, assuming there is no drastic alteration in K_M, the rate drop expected should still be in the detection range of the HPLC assay. One possible explanation of why there is no cleavage is that the presence of the modified deoxynucleosides may cause a local perturbation in the DNA structure that prevents the endonuclease from reacting with the oligodeoxynucleotide. Another possibility is that the hydrogen bonds to the enzyme that are deleted by the incorporation of the analogues are part of a group of hydrogen bonds that form cooperatively. If this is so then the deletion of one of these hydrogen bonds could prevent the formation of the

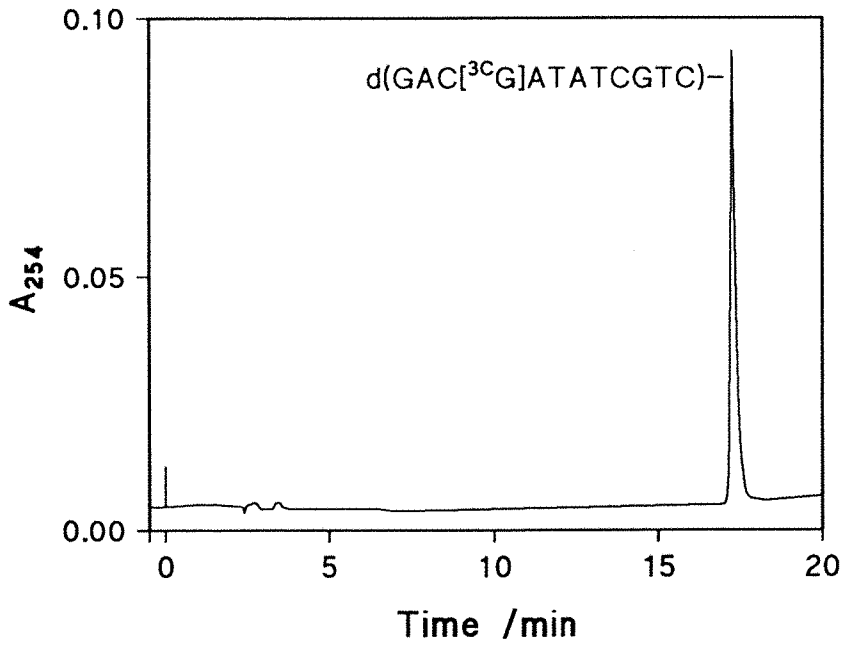


Figure 3c.1: HPLC trace of d(GAC[^{7C}G]ATATCGTC) after treatment with 8.5μM endonuclease for 24 hours. No nuclease cleavage of the oligodeoxynucleotide is visible.

other hydrogen bonds and thus greatly lower the ability of the endonuclease to form the transition state complex. The protein contacts to these two deoxyguanosine groups shown in the crystal structure (Figure 3c.2) are from main chain protein atoms that are part of a loop that extends into the major groove [see section 1.6.1]. In the cognate complex this loop is also stabilized by intraloop hydrogen bonds but in the absence of cognate DNA, this loop is poorly ordered (Winkler, 1991). This suggests that the hydrogen bonds of this loop do indeed form with a degree of cooperativity.

The cleavage of the control dodecamer by the Eco RV restriction endonuclease occurs between the central T and dA deoxynucleosides and leads to the formation of two hexameric products:-

$$\begin{matrix} \text{d(5'-GACGATATCGTC-3')} \\ \text{d(3'-CTGCTATAGCAG-5')} \end{matrix} \quad \rightarrow \quad \text{2d(GACGAT) + 2d(pATCGTC)}$$

To check that the cleavage reactions observed for the analogue containing dodecamers was also occurring at the expected position for the Eco RV endonuclease, cleavage reactions of each of the substrates were allowed to go to completion for product analysis. The two product peaks were purified by HPLC and their base compositions determined as described in section 2.2.7i. All of the dodecamer substrates gave the correct deoxynucleoside compositions for the two hexameric products expected from cleavage between the central T and dA deoxynucleosides as shown in Table 3c.1. The base composition analysis of the cleavage products of the control

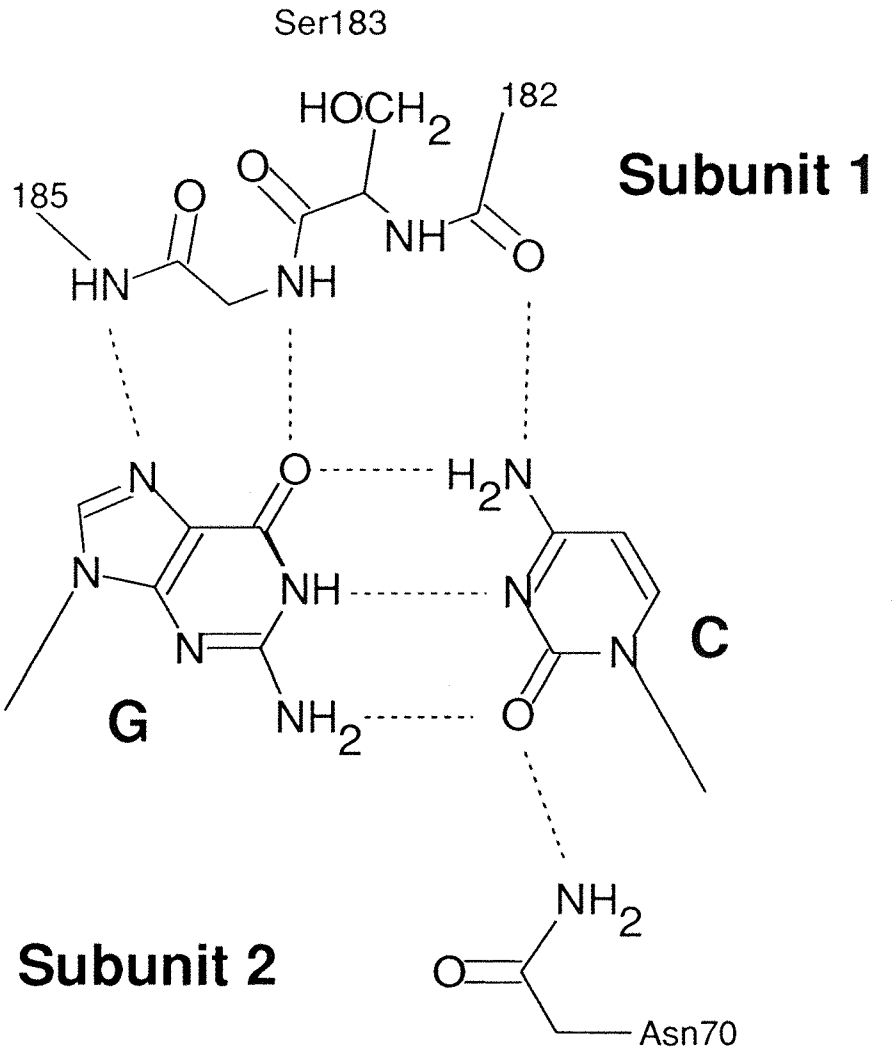


Figure 3c.2: Contacts to the G.C base pair in the Eco RV site of the Eco RV endonuclease cognate crystal complex (Winkler, 1991),

octadecamer also gave the expected ratios for cleavage at the Eco RV site. The cleavage products of d(TGACGATATCGTC) were not characterised by their base compositions but one of its product peaks, after alkaline phosphatase treatment, co-eluted with standard d(ATCGTC) which is one of the expected products from the Eco RV endonuclease cleavage of the tridecamer:-

$$\begin{matrix} \text{d(5'-TGACGATATCGTC -3')} \\ \text{d(3'- CTGCTATAGCAGT-5')} \end{matrix} \rightarrow \text{2d(TGACGAT) + 2d(pATCGTC)}$$

Although this is not conclusive proof, it is extremely unlikely that cleavage would occur at any other position.

3c.2 DETERMINATION OF THE RATES OF ECO RV ENDONUCLEASE HYDROLYSIS OF THE OLIGODEOXYNUCLEOTIDES

In order to be able to determine the rates of cleavage using the UV assay, the change in UV absorbance per nmol of oligodeoxynucleotide cleaved for each substrate needs to be determined. This was done for all substrates, with the exception of the control octadecamer, using the method described in section 2.4.3.i and discussed for the control dodecamer in section 3b.2.3.i. Basically, a known amount of each substrate was totally cleaved by the Eco RV endonuclease and the change in absorbance at 254nm associated with this recorded. Cleavage reactions were confirmed as being finished by HPLC which revealed the presence of only the two expected product peaks and no detectable substrate peak. Since the amount of oligodeoxynucleotide was known and the increase in absorbance associated with its cleavage was known, the change in absorbance per nmol of substrate cleaved could be calculated. Wavelength scans over the range 400nm to 200nm were also recorded for each substrate before and after the total digestion by the endonuclease. This was so that the change in absorbance at wavelengths other than 254nm resulting from each nmol of substrate cleaved could also be calculated. Situations when cleavage reactions may need to be monitored at wavelengths other than 254nm are discussed later in section 3c.3.

The endonuclease hyperchromicity factors for the oligodeoxynucleotides at 254nm are listed in Table 3c.2. Complete cleavage of the substrates resulted in percentage increases in the 254nm absorbance of between 20% and 32%. Comparison of these values with those for the 254nm absorbance increases associated with the complete digestion of the oligodeoxynucleotides to their constituent deoxynucleosides by nucleases [see Table 3a.2; section 3a.3.2] reveals that the latter values are only slightly higher. Therefore most of the hyperchromicity of the oligodeoxynucleotides must be associated with the formation of double stranded DNA rather than the formation of the DNA chain *per se*. This is as expected, since most of the hyperchromicity of DNA comes from the base stacking in duplex DNA (Saenger, 1984).

All of these oligodeoxynucleotide substrates, had their rates of cleavage at 5μM substrate concentration evaluated using the UV method as described in section 2.4.3.v. The more easily cleaved oligodeoxynucleotides, d(GACGATATCGTC), d(TGACGATATCGTC), d(GACGATAT[^{4H}C]GTC) and d(GACGATAT[^{5Me}C]GTC), were reacted with 5pmol (5nM) endonuclease. The more resistant d[^{3C}G] and d[I] containing dodecamers were reacted with 50pmol (50nM) endonuclease and the particularly resistant d(GAC[^{6S}G]ATATCGTC) oligodeoxynucleotide was incubated with 400pmol (400nM) endonuclease. The rate lines obtained are shown in Figure 3c.3. Initial rates were calculated by linear regression of the initial linear portions of the graphs as shown by the dotted lines. These rate values are given in Table 3c.2. The results obtained can be split into three categories:- (i) Dodecamers whose rates are essentially the same as the control dodecamer. That is, d(GACGATAT[^{4H}C]GTC) and d(GACGATAT[^{5Me}C]GTC). (ii) Dodecamers whose rates are reduced by a factor of about ten;

Table 3c.2: Table showing the percentage increases in the absorbance at 254nm associated with the total cleavage of the oligodeoxynucleotides by the Eco RV endonuclease, and the initial turnover rates of the substrates determined at 5μM concentration using the UV assay. [a]Not determined. [b]Given as per nmol of double strands cleaved in 1ml volume and using 1cm path length cells. [c]Determined from Figure 3c.3. [d]Determined at 5.6μM substrate concentration. [e]Determined using the HPLC assay.

OLIGODEOXYNUCLEOTIDE	% Increase in A_{254} after Total Eco RV cleavage	Increase in A_{254} per nmol of substrate cleaved[b]	Turnover[c] /min^{-1} (% relative to d(GACGATATCGTC))
d(GACGATATCGTC)	21%	0.0342	30.1 (100%)
d(GAC[^{6S}G]ATATCGTC)	31%	0.0524	0.014 (0.05%)
d(GAC[I]ATATCGTC)	32%	0.052	2.22 (7%)
d(GAC[^{3C}G]ATATCGTC)	29%	0.0498	1.80 (6%)
d(GACGATAT[^{5Me}C]GTC)	21%	0.0362	22.8 (76%)
d(GACGATAT[^{4H}C]GTC)	22%	0.0384	25.3 (84%)
d(TGACGATATCGTC)	15%	0.0266	32.1[d] (107%)
d(GTCGACGATATCGTCGAC)	n/d[a]	—	132[e] (440%)

d(GAC[^{3C}G]ATATCGTC) and d(GAC[I]ATATCGTC). (iii) d(GAC[^{6S}G]ATATCGTC) whose rate of cleavage is greatly reduced (>2,000 fold) compared with the control dodecamer. These results are difficult to interpret without obtaining both the K_M and k_{cat} values but it is clear that the presence of the methyl group in the d[^{5Me}C] analogue, or the deletion of the 4-amino group, as in the d[^{4H}C] analogue, has little effect on the reaction. In contrast, the alteration of 4=S for 4=O in the d[^{6S}G] analogue has a very large effect on the reaction. The importance of the 6=O group of the deoxyguanosine has already been demonstrated by the complete lack of cleavage of the d[^{6H}G] containing octadecamer. The results for the d[I] and d[^{5Me}C] containing dodecamers are in very close agreement with published Eco RV endonuclease cleavage rates for d[I] and d[^{5Me}C] containing undecamers (Fliess *et al.*, 1988). d(AAA[I]ATATCTT) was cleaved at 7%, and d(AAAGATAT[^{5Me}C]TT) was cleaved at 60% of the rate of cleavage for the normal undecamer.

The rate of cleavage of the octadecamer, d(GTCGACGATATCGTCGAC), has been determined using the HPLC method at a substrate concentration of 5μM (Connolly, B.A., personal communication) and is included in Table 3c.2 for comparison. Its rate is 4.4 times that of the control dodecamer. This is probably the result of the dodecamers not entirely filling the binding cleft of the endonuclease and not being able to make the additional contacts to the enzyme that the octadecamer can. The crystal structure of the endonuclease bound to a non-cognate hexadecamer does indeed show extra contacts to phosphates that beyond the length of a dodecamer (Winkler, 1991). These contacts are to phosphates at positions -6 and -7 (the central scissile phosphate defines the 0 position and the negative sign refers to phosphates on the 5' side of it). Although no crystal structure is available of a cognate complex involving an

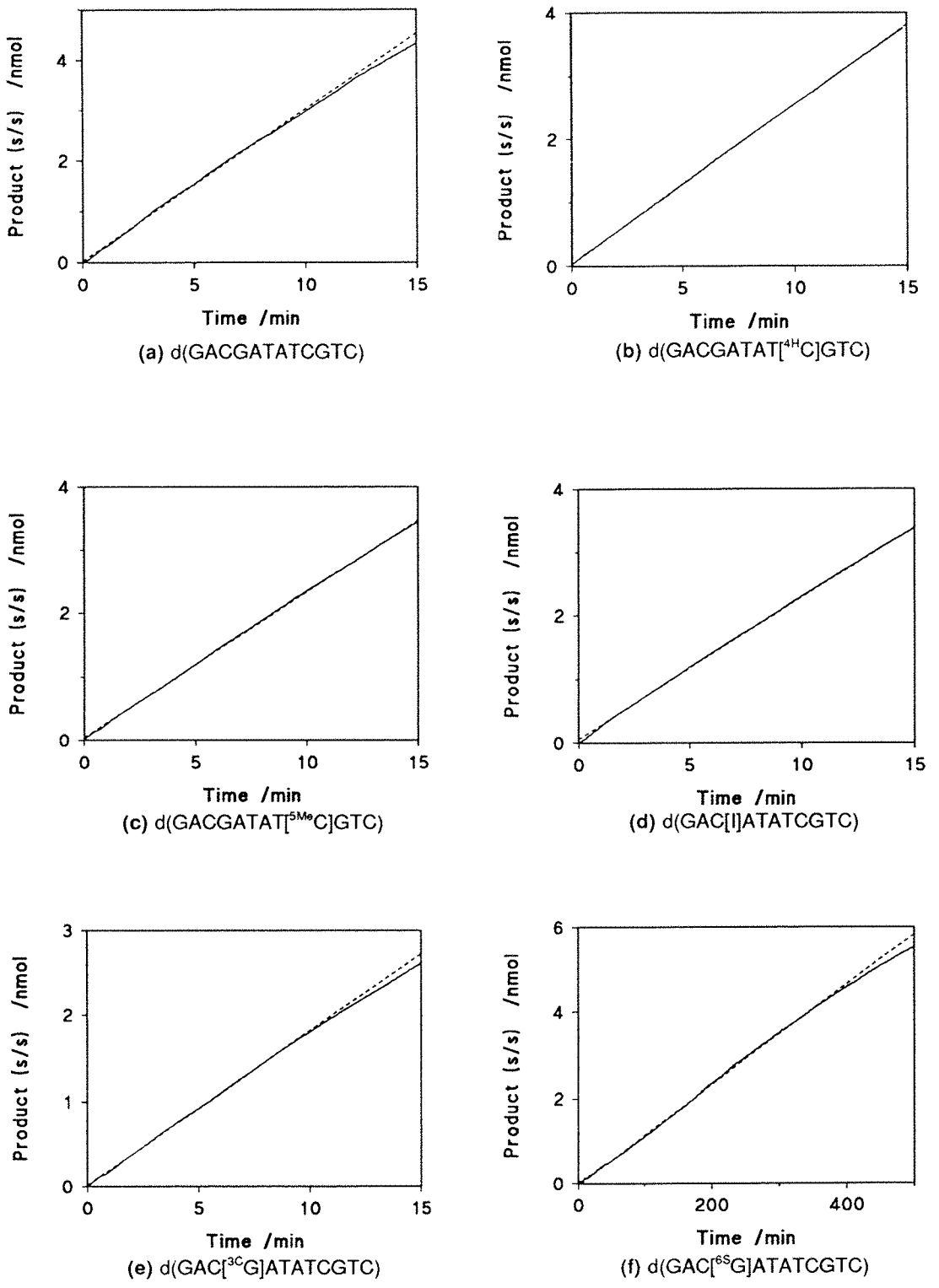


Figure 3c.3: Initial rates of cleavage of the non-phosphorylated dodecamer Eco RV substrates at 5μM substrate concentration. (a) to (c) were cleaved with 5nM endonuclease, (d) and (e) with 50nM, and (f) with 400nM endonuclease.

oligodeoxynucleotide of this length, it appears that these contacts are possible to equivalent phosphates in hexadecamer or longer cognate oligodeoxynucleotides.

3c.3 MICHAELIS-MENTEN KINETIC ANALYSES OF THE OLIGODEOXYNUCLEOTIDES

The Michaelis-Menten parameters k_{cat} and K_M can be determined by measuring the initial rate of a reaction over a range of substrate concentrations (Fersht, 1985). Complementary studies using deoxyadenosine and thymidine analogues determined k_{cat} and K_M values at 25mM magnesium ion concentration (Newman *et al.*, 1990b). At this magnesium concentration, the Eco RV endonuclease reaction with d(pGACGATATCGTC) was found to be saturated with magnesium and so the Michaelis-Menten studies of Newman *et al.* and the ones performed here, were carried out in buffer α25 (25mM $MgCl_2$, 55mM NaCl, 50mM HEPES, pH7.5). In the previous study, the initial rates for the k_{cat} and K_M determinations were measured using the polyacrylamide gel electrophoresis assay [section 3b.2.2]. In this study, initial rates were determined using the UV assay. As mentioned above, in order to calculate k_{cat} and K_M values, initial rates over a range of substrate concentrations need to be measured. It was found that measuring the endonuclease cleavage reaction at 254nm in 1cm path length cells, meant that the practical substrate concentration range for measuring cleavage rates of d(GACGATATCGTC) was about 0.4μM to 7.5μM. Going below the lower limit produced absorbance changes too small to measure whilst going above the upper limit resulted in absorbances at 254nm that were outside the linear range of the spectrometer. These limits however, could be extended by the use of different path length cells. Cells with path lengths of between 0.2cm and 5cm are widely available and so in theory the range of concentrations of d(GACGATATCGTC) that can be used could be extended to between 0.08μM and 37μM. The practicalities of using the complete range of path length cells available was not investigated and there are potential problems associated with their use. Using longer path lengths increases the absorbance changes due to cleavage but also increases the background absorbance resulting in a greater noise level. Another potential inconvenience is that longer path length cells require larger reaction volumes and hence larger quantities of reactants. On the other hand, using shorter path lengths decreases the noise but also decreases the absorbance change per nmol of oligodeoxynucleotide cleaved. This may be a problem for slow reaction rates. For the Michaelis-Menten kinetic analyses of the dodecamers, it was found that extending the substrate range by the use of 2cm path length cells, as well as the usual 1cm cells, was adequate. The 2cm path length cells doubled the UV response to oligodeoxynucleotide cleavage. They also increased the noise level but the increase in noise was found to be low enough not to be a problem.

Another way of extending the upper limit of substrate concentration is by monitoring

the reaction at a different wavelength. Wavelength scans (200nm to 400nm) of the oligodeoxynucleotides before and after total endonuclease cleavage showed that the absorbance increases at other wavelengths were large enough to enable the reaction to be followed not only at 254nm but also at other wavelengths. For example, the absorbance at 240nm of all the dodecamers is approximately half that at 254nm but the absorbance increase at 240nm on endonuclease hydrolysis is almost as large as that at 254nm. Thus, by monitoring the reaction at 240nm, twice as much substrate could be used with very little loss in the sensitivity of the UV response to cleavage. This approach was used for the Michaelis-Menten kinetic determination of one of the analogue containing dodecamers, d(GAC[^{3C}G]ATATCGTC), and is discussed in more detail in the relevant section [3c.3.1.iv].

3c.3.1 Michaelis-Menten Kinetic Analysis of the Phosphorylated Control and Analogue Containing Dodecamers

An attempt was made to find the Michaelis-Menten constants for the non-phosphorylated control oligodeoxynucleotide by using the UV method for determining the rate of cleavage at various dodecamer concentrations. This was unsuccessful because there was no change observed in the rate of cleavage on going from 6.4μM substrate to 1.3μM. This indicated that the enzyme must be saturated with substrate at the substrate concentrations used, which meant that K_M must be <<1μM. The value calculated before for the *phosphorylated* control, d(pGACGATATCGTC), was 3.8μM (Newman *et al.* 1990b) which would be expected to give a large difference in the rates of cleavage of between 1.3μM and 6.4μM substrate. The fact that no change was observed could be explained by the difference in the two methods of rate determination where the earlier study monitored the formation of ^{32}P labelled product. However, the UV assay has been shown to give a genuine measure of the rate of cleavage. The difference in K_M is due to the presence of the 5'-phosphate (^{32}P label) in the previous study and the rate of cleavage of phosphorylated control oligodeoxynucleotide was found to vary considerably over the same substrate concentration range used for the non-phosphorylated control.

Partly because of the difficulties described above in determining the Michaelis-Menten parameters of the non-phosphorylated analogue dodecamers but mainly to be consistent with the previous complementary work carried out on deoxyadenosine and thymidine analogues (Newman *et al.* 1990b), it was decided that all the K_M and k_{cat} determinations for the control and analogue containing dodecamers should be done using 5'-phosphorylated oligodeoxynucleotides.

To determine K_M and k_{cat} values for the phosphorylated dodecamers their initial rates of cleavage were each determined at six different substrate concentrations using the UV assay. The UV lines obtained, were converted to nmol of product formed as a function of time. Initial rates were determined by linear regression of the initial linear portions of these rate lines.

Table 3c.3: Kinetic data for the reaction of the Eco RV endonuclease with oligodeoxynucleotides containing deoxyguanosine and deoxycytidine analogues. [a]Eco RV site contained within the sequence d(GACGATATCGTC). [b]Given as rate of cleavage relative to the control. [c]Given as kJ per mol of modified base.

BASE	ECO RV SITE[a]	k_{cat} /min^{-1}	K_M /μM	k_{cat}/K_M /$s^{-1}M^{-1}$	ΔG_{app}[c] /$kJmol^{-1}$
–	GACGAT	33	2.9	186000	0
G_1	[^{6S}G]ACATC	(0.05%)[b]	n/d	n/d	n/d
	[I]ACATC	0.63	1.76	5950	-4.3
	[^{3C}G]ACATC	0.28	3.51	1340	-6.1
C_6	GATAT[^{5Me}C]	21	0.97	362000	+0.8
	GATAT[^{4H}C]	38	2.0	317000	+0.7

Plots of initial reaction rate (V) against initial substrate concentration (S_0) were made to check that an adequate range of substrate concentrations had been covered for each substrate. The initial rates were also plotted as a Hanes plot of initial substrate concentration, [S_0], against [S_0] divided by the initial rate of cleavage, V (Fersht, 1985). The value of K_M is given by the x-axis intercept of such a plot and the value of k_{cat} can be calculated from the gradient which is equal to the reciprocal of V_{max}. From the specificity values of k_{cat}/K_M relative to the control dodecamer, values for the apparent binding energy, ΔG_{app}, for each analogue containing substrate could be calculated [as described in section 1.2.1], using the formula below:-

$$\Delta G_{app} = RT\ln[(k_{cat}/K_M)_{analogue}/(k_{cat}/K_M)_{control}]$$

The results obtained are given in Table 3c.3 and are discussed in more detail below.

3c.3.1.i Kinetics of the Phosphorylated Control Dodecamer, d(pGACGATATCGTC)

The initial rates of cleavage of d(pGACGATATCGTC) over a range of substrate concentrations using 5nM endonuclease are shown on Figures 3c.4a and b. One of the advantages of using the UV assay is that the initial absorbance of the reaction solution gives a very accurate measure of the initial substrate concentration. A Hanes plot of the results obtained is given in Figure 3c.4c and gave a good straight line indicating that under these conditions the endonuclease followed Michaelis-Menten kinetics. From the line, a value of 2.9μM (double stranded concentration) for K_M was calculated. This is in good agreement with the K_M of 3.8μM obtained previously for the Eco RV endonuclease reaction with the same substrate, as determined using ^{32}P labelled substrate in a polyacrylamide gel assay (Newman *et al.*, 1990b). It also agrees well with the value of 4μM obtained for the non-phosphorylated

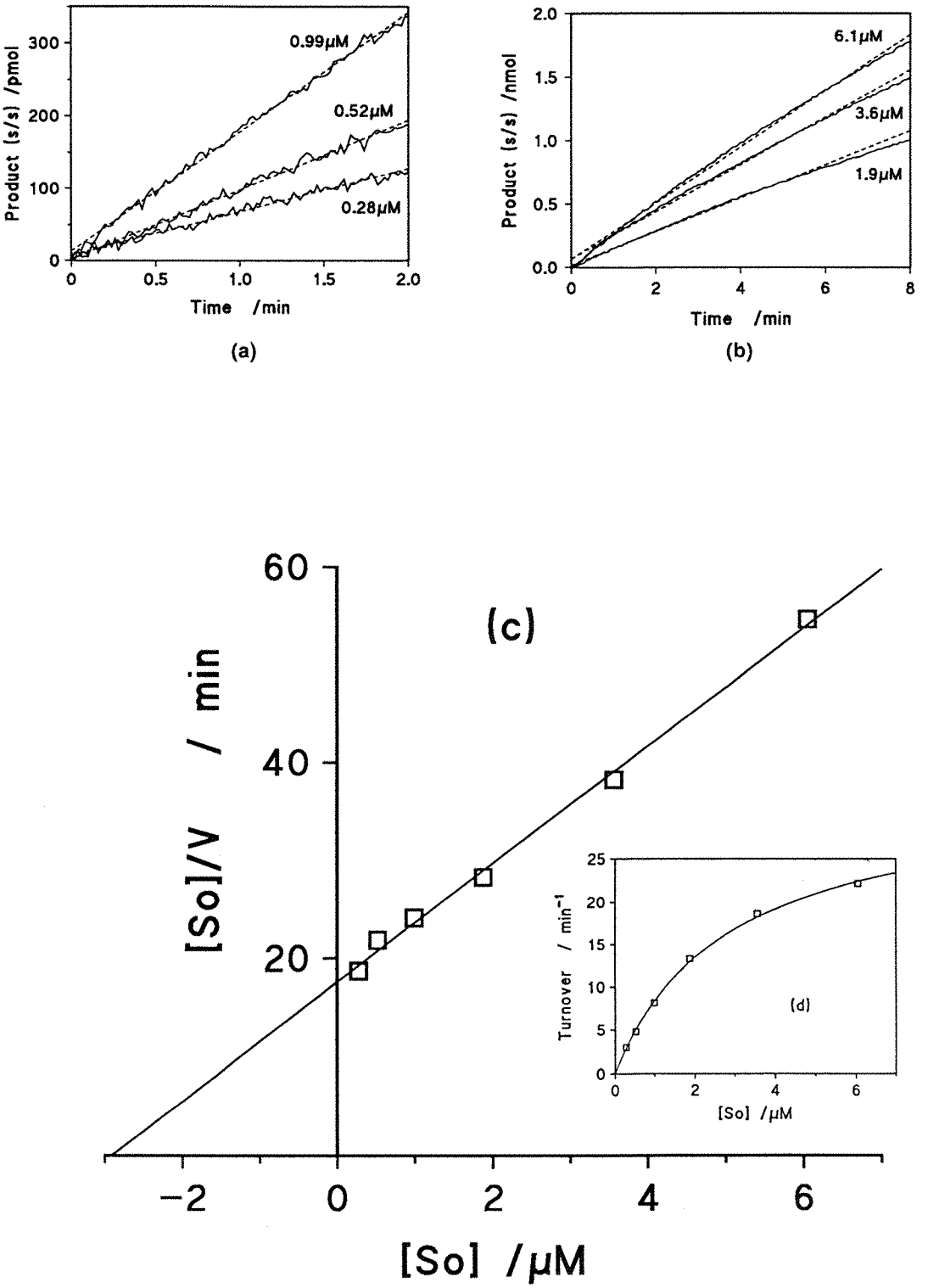


Figure 3c.4: Eco RV endonuclease cleavage reaction of d(pGACGATATCGTC) at various substrate concentrations. **(a)** Rate lines obtained in 2ml volumes using 2cm path length cells. **(b)** Rate lines obtained in 1ml volumes using 1cm path length cells. **(c)** Hanes plot. **(d)** Plot of initial substrate concentration, $[S_0]$, versus initial turnover rate.

oligodeoxynucleotide, d(AAAGATATCTT), by Fliess *et al.* (1988) although the conditions used by them were different (11°C temperature and 10mM Mg^{2+}). A somewhat lower K_M of 0.46μM was obtained for d(pCTGATATCAG) by Mazzarelli *et al.* (1989) and may reflect the different bases flanking the Eco RV site. Flanking sequences are known to affect the reaction of Eco RI endonuclease with its recognition site. For example, the reaction of Eco RI endonuclease with the different Eco RI sites on a particular plasmid has been shown to vary from site to site (Halford *et al.*, 1980). This difference in rate was proposed to be caused by different sequences flanking the different sites. Recently, more direct evidence for the interaction of flanking sequences with the Eco RI endonuclease came from the different rates of cleavage that were observed for oligodeoxynucleotides that varied in length and in flanking sequences (Van Cleve & Gumport, 1992). In the case of the Eco RV endonuclease it is known that the enzyme makes contact to phosphate groups outside the Eco RV site in its complex with d(GGGATATCCC) (Winkler, 1991).

K_M values of the order of 0.1μM to 7μM have also been observed for the reaction of the Eco RI endonuclease with oligodeoxynucleotide substrates (Greene *et al.*, 1975; Brennan *et al.*, 1986; McLaughlin *et al.*, 1987), although a recent study has found that K_M varied considerably with oligodeoxynucleotide length (Van Cleve & Gumport, 1992). K_M's of 3.3μM and 0.2μM have also been observed for the reaction of the Bgl II and Mbo I endonucleases respectively, with d(pGGAGATCTCC) (Ono & Ueda, 1987a).

The gradient of the graph in Figure 3c.4c yielded a k_{cat} value of $33min^{-1}$ (given as the number of mols of *double* stranded substrate cleaved per minute per mol of *dimeric* enzyme). This is almost five times the value of $6.9min^{-1}$ obtained previously for the same substrate (Newman *et al.*, 1990b). This probably reflects the different approaches used to concentrate the enzyme after the final purification step. In the earlier study, the endonuclease was isolated by precipitation with ammonium sulphate followed by redissolving the protein precipitate in buffer, whereas here the enzyme was concentrated by reducing the solution volume using a system based on membrane filtration [section 3b.1]. The former method may result in the denaturation of some of the endonuclease during the precipitation and redissolving processes and thus produce an endonuclease preparation that is not fully active. In the method used here, these problems are avoided. The inclusion of BSA in the dilution buffer used here may also be a cause of the increase in k_{cat}.

The k_{cat} obtained in this study is also higher than the value obtained by Fliess *et al.* (1988) of $7min^{-1}$ for d(AAAGATATCTT) and the much smaller value of $0.7min^{-1}$ determined by Mazzarelli *et al.* (1989) for d(pCTGATATCAG). These differences may again be the result of different enzyme preparations or they may arise from different base flanking sequences.

The k_{cat} and K_M values calculated here give a specificity constant (k_{cat}/K_M) of $1.86 \times 10^5 s^{-1}M^{-1}$ which is approximately five times that obtained for the same oligodeoxynucleotide by Newman *et al.* (1990b). This is mainly due to the increased k_{cat} value discussed above. A much greater specificity constant of $>3 \times 10^7 s^{-1}M^{-1}$ has been reported for the Eco RV

endonuclease reaction with a single plasmid Eco RV site (Halford & Goodall, 1988; Taylor & Halford, 1989). This greater specificity was despite a smaller k_{cat} value of $0.9min^{-1}$ and is due to a much smaller K_M value. No reduction in rate was observed on varying the plasmid concentration from 30nM to 2nM and so K_M must be <<2nM. A value of 0.5nM was quoted and this much smaller value of K_M observed for the plasmid is probably a result of the facilitated diffusion mechanism playing an important part in the plasmid reaction. With the large plasmid substrate the endonuclease can bind non-specifically and then diffuse along the DNA to its target site, thus speeding up the rate of encounter between the enzyme and its recognition site. With short oligodeoxynucleotides this process is not possible. A similar effect has also been noted with the Eco RI endonuclease as discussed in section 1.5.1 and it too has been attributed to facilitated diffusion. Another possible cause for the higher K_M values with oligodeoxynucleotide substrates is that because they are short they lack phosphates that are contacted by the endonuclease. From the crystal structures of a cognate and of a non-cognate DNA complex contacts are expected between the enzyme and DNA phosphates outside the Eco RV site. With the central scissile phosphate defining the zero position, all phosphates from -2 (5' side) through to +4 (3' side) are contacted by the endonuclease in the cognate complex. In addition, the enzyme contacts phosphates -6 and -7 in the non-cognate complex. Although these two phosphates are not present in the decamer of the cognate complex, equivalent contacts do seem possible for longer DNA substrates.

3c.3.1.ii Kinetics of d(GAC[^{6S}G]ATATCGTC)

Since d(GAC[^{6S}G]ATATCGTC) was so poorly cleaved [>2,000 fold reduction in the cleavage rate compared with d(GACGATATCGTC); see Table 3c.2] and because the results with other oligodeoxynucleotides showed that 5'-phosphorylation reduced the rate still further, no attempt was made to determine k_{cat} and K_M for the d[^{6S}G] containing dodecamer. The concentration of endonuclease that was expected to be required to obtain measurable initial rates would have been greater than the concentration of the substrate needed for Michaelis-Menten type kinetics and so would not give meaningful k_{cat} and K_M values. A similar problem was encountered for the poorly cleaved oligodeoxynucleotide, d(GACG[P]TATCGTC) (Newman *et al.*, 1990b). The cleavage of d(GACG[P]TATCGTC) was found to be approximately a thousand fold down on that of d(GACGATATCGTC) at 20μM substrate concentration and it was found to be impossible to get measurable rates over a wide enough range of substrate concentrations to enable the determination of k_{cat} and K_M for d(pGACG[P]TATCGTC).

Despite the lack of k_{cat} and K_M data for d(GAC[^{6S}G]ATATCGTC) it can still be concluded that the 6C=O group of the deoxyguanosine in the Eco RV site is a very important determinant in the endonuclease reaction. This is further supported by the lack of cleavage in the d[^{6H}G] containing octadecamer. This result is in agreement with the cognate crystal structure which shows a hydrogen bond between this 6-carbonyl oxygen and a protein main chain nitrogen, as shown in Figure 3c.2.

3c.3.1.iii Kinetics of d(pGAC[I]ATATCGTC)

The results for the kinetics of d(pGAC[I]ATATCGTC) using 100nM endonuclease are shown in Figure 3c.5. From the Hanes plot, K_M was found to be 1.8μM and k_{cat} was 0.63min^{-1}. Relative to the control dodecamer d(pGACGATATCGTC), d(pGACGATAT[I]GTC) therefore has essentially the same K_M value but a considerably reduced k_{cat}. The specificity constant of 5.95 x 10^3s^{-1}M^{-1} for d(pGAC[I]ATATCGTC) represents an approximate 30 fold drop in specificity compared with d(pGACGATATCGTC). This gives a ΔG_{app} of –8.5kJ/mol for the d[I] containing oligodeoxynucleotide. Since no structural deviations from double stranded B-DNA were apparent from the CD and T_M data, and because no charged groups are involved, ΔG_{app} is expected to approximate ΔG_{bind}. Therefore each 2-amino group of the two deoxyguanosines in the Eco RV site are expected to contribute –4.3kJ/mol to the binding energy of d(pGACGATATCGTC) to the endonuclease. This value is within the range of values previously reported for hydrogen bond deletions (2.1kJ/mol to 7.5kJ/mol; Fersht, 1987). Fliess *et al.* (1988) report a specificity of 11% for d(AAA[I]ATATCTT) relative to the native sequence, with the reduction, as for d(pGAC[I]ATATCGTC), arising from a reduction in k_{cat}. Their relative specificity of 11% gives a value of -2.7kJ/mol for the ΔG_{bind} of each deoxyguanosine 2-amino groups. This value, although about half that obtained here for d(pGAC[I]ATATCGTC), is still within the range expected for a hydrogen bond deletion. In contrast, Mazzarelli *et al.* (1989) see little change in k_{cat} or K_M for d(pCT[I]ATATCAG) compared with its control decamer, and a virtually identical specificity constant. Also, the crystal structure of a cognate DNA complex with Eco RV endonuclease shows no contact to the 2-amino group of the deoxyguanosine in the Eco RV site (Winkler, 1991). Taking all the presently available information into account, the 2-amino group of the deoxyguanosine is probably not involved in hydrogen bond formation to the endonuclease during the cleavage reaction. The reduced specificity seen here and by Fliess *et al.* for the substitution of deoxyinosine for deoxyguanosine is probably the result of some structural effect caused by the presence of the modified base. The CD and T_M characterisations of the deoxyinosine containing oligodeoxynucleotides revealed no apparent deviation from an overall double stranded B-DNA structure under the assay conditions. However, local changes in the DNA structure that could affect the endonuclease reaction cannot be ruled out. Also, the 2-amino group is in the minor groove of the DNA which in the crystal structure faces into the enzyme. Thus, although the crystal structure shows no direct hydrogen bond to the 2-amino group, the endonuclease reaction may be sensitive to changes in this area of the DNA. Another possibility is that the substitution of deoxyinosine for deoxyguanosine affects the flexibility of the DNA and by this means affects the ability of the DNA to form the transition state complex. The cognate crystal complex shows that the bound DNA is extremely distorted and it may be that the replacement of deoxyinosine for deoxyguanosine in the Eco RV site makes it more difficult for the DNA to adopt the cognate complex structure.

Similarly contradicting results have been observed for the Eco RI endonuclease reaction with substrates that have deoxyguanosine replaced by deoxyinosine. Ten fold drops in

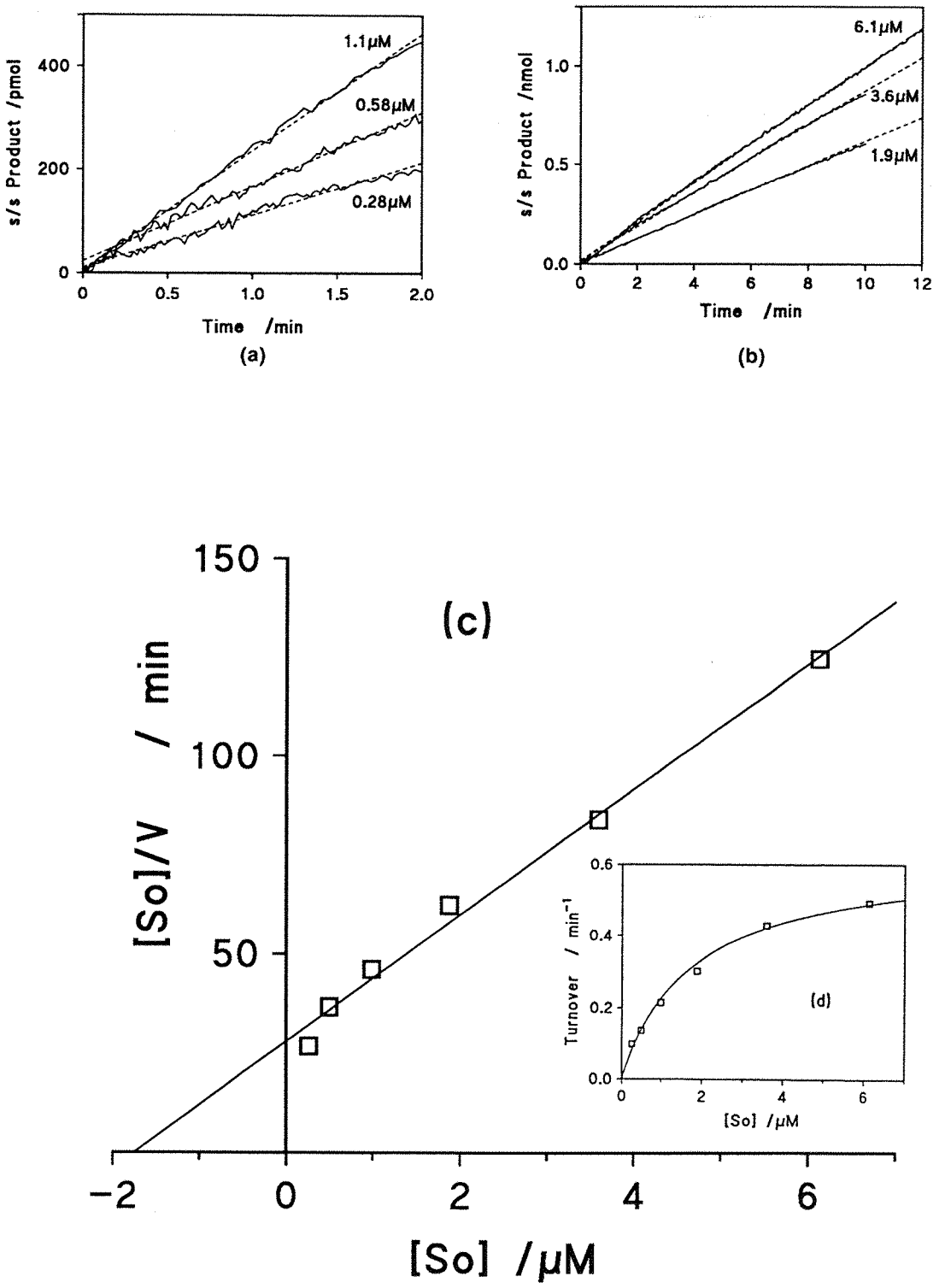


Figure 3c.5: Eco RV endonuclease cleavage reaction of d(pGAC[I]ATATCGTC) at various substrate concentrations. **(a)** Rate lines obtained in 2ml volumes using 2cm path length cells. **(b)** Rate lines obtained in 1ml volumes using 1cm path length cells. **(c)** Hanes plot. **(d)** Plot of initial substrate concentration, $[S_0]$, versus initial turnover rate.

specificity have been determined for the substitution of deoxyinosine for deoxyguanosine in the Eco RI site of two separate oligodeoxynucleotides (Brennan *et al.*, 1986; McLaughlin *et al.*, 1987). Whereas, the same substitution in the single Eco RI site of a plasmid had no apparent effect on the endonuclease reaction (Modrich & Rubin, 1977). Also, the crystal structure of an Eco RI endonuclease/DNA complex shows no protein contacts to the DNA in the minor groove (McClaren *et al.*, 1986; revised in Kim *et al.*, 1991). The apparent anomalous effect of deoxyinosine in the Eco RI endonuclease reaction has been proposed as being a consequence of the different curvature of deoxyinosine containing oligodeoxynucleotides compared with the native sequence (Diekmann & McLaughlin, 1988) and a similar effect may well be involved in the Eco RV endonuclease reaction.

3c.3.1.iv Kinetics of d(pGAC[^{3C}G]ATATCGTC)

This oligodeoxynucleotide was found to be particularly resistant to cleavage and an enzyme concentration of 300nM was needed to get measurable rates. This falls outside the linear dependence of enzyme activity on enzyme concentration investigated in this study [section 3b.2.3.iii] but it does fall within the linear range reported previously (0 to 1.2μM; Newman, 1989).

For the lower five concentrations the rates were determined as usual by observing the change in absorbance at 254nm. From the rates of these lower substrate concentrations it became obvious that the value of K_M for d(pGAC[^{3C}G]ATATCGTC) was such that, in order to get reliable values for k_{cat} and K_M, the cleavage rate of a higher substrate concentration was

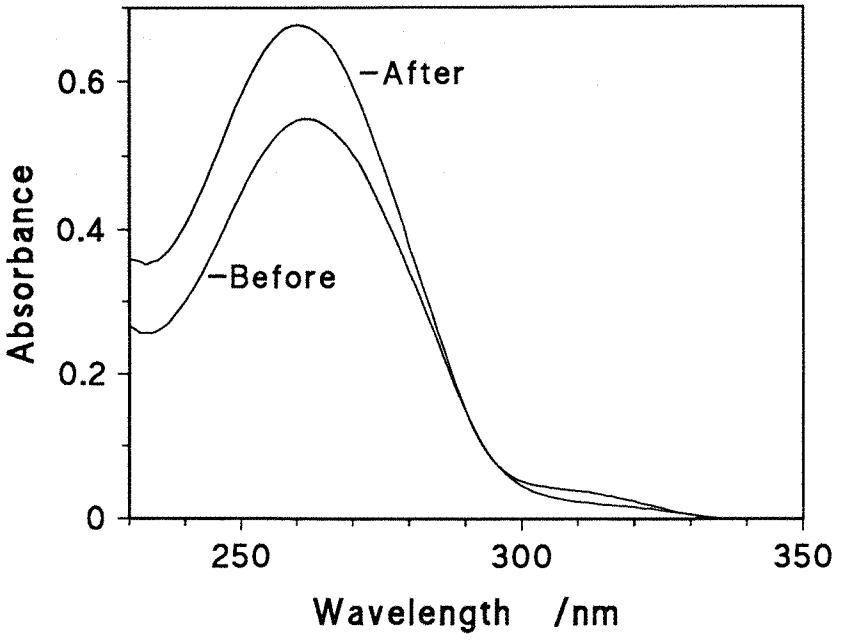


Figure 3c.6: Wavelength scan of d(GAC[^{3C}G]ATATCGTC) before and after total hydrolysis by the Eco RV endonuclease.

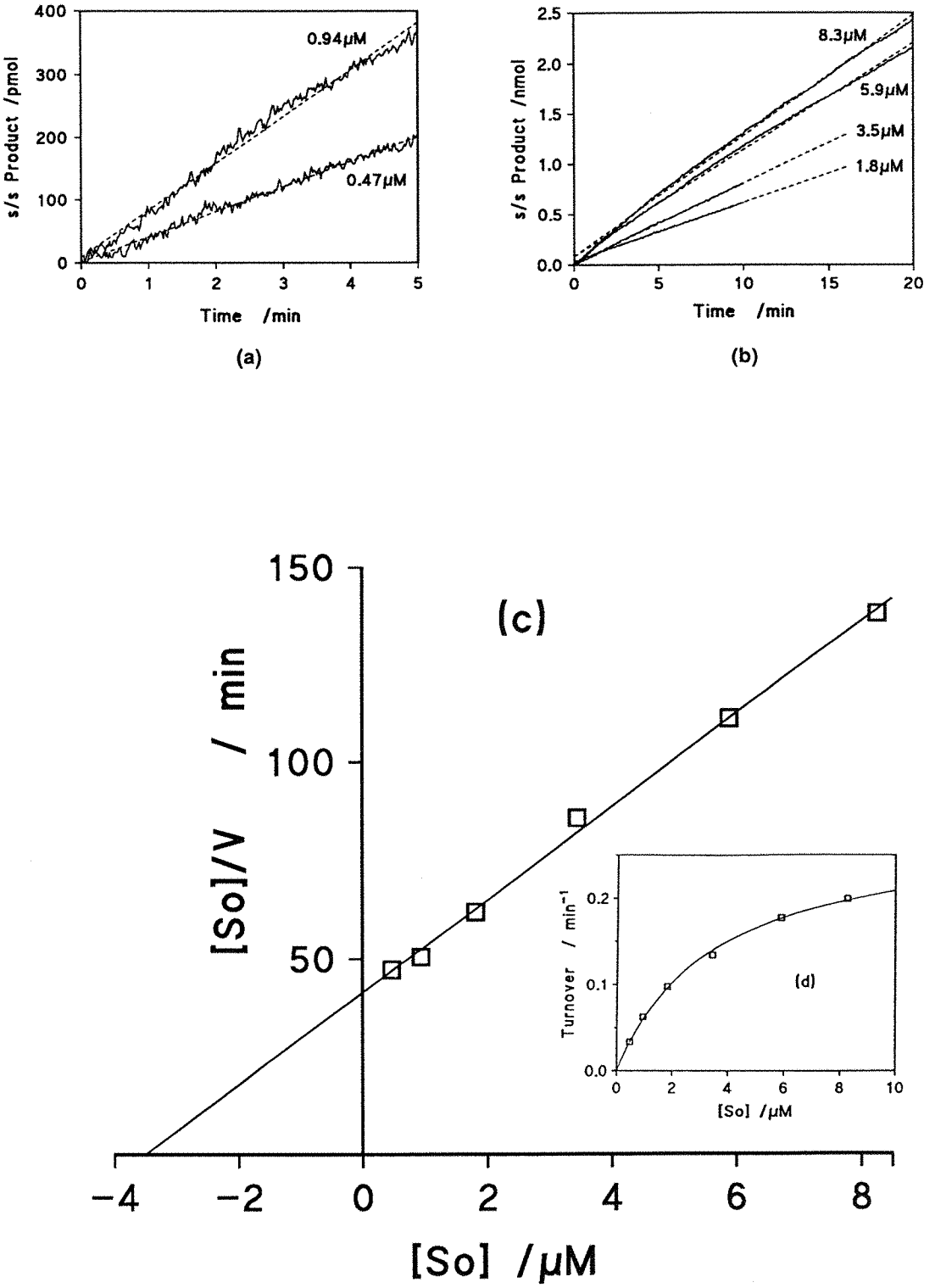


Figure 3c.7: Eco RV endonuclease cleavage reaction of d(pGAC[^{3C}G]ATATCGTC) at various substrate concentrations. **(a)** Rate lines obtained in 2ml volumes using 2cm path length cells. **(b)** Rate lines obtained in 1ml volumes using 1cm path length cells. The sample at 8.3μM substrate concentration was measured at 245nm. **(c)** Hanes plot. **(d)** Plot of initial substrate concentration, $[S_0]$, versus initial turnover rate.

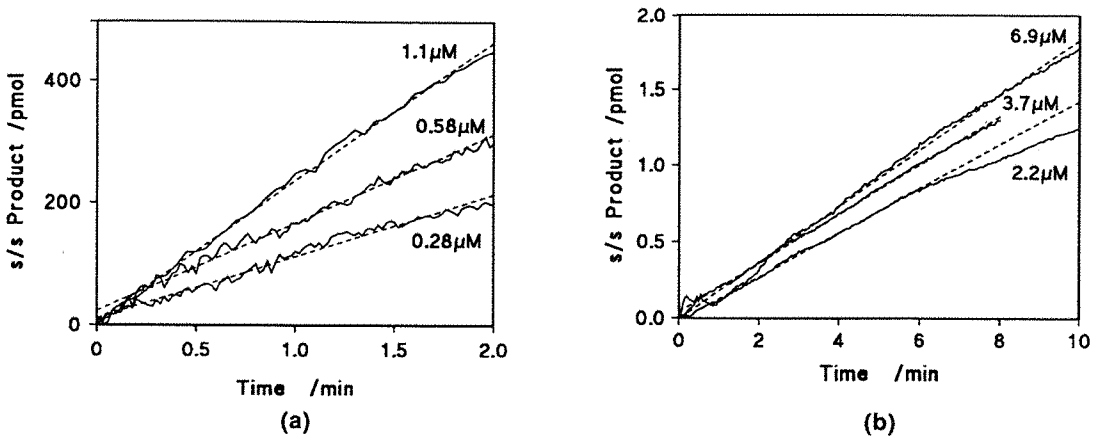
needed. A higher concentration in 1cm path length cells, however, would have been outside the linear absorbance range of the spectrometer at 254nm. In order to overcome this, the rate of hydrolysis of 8.3μM d(pGAC[^{3C}G]ATATCGTC) was measured at *245*nm where the oligodeoxynucleotide has a lower absorbance. To be able to convert the change in absorbance at 245nm to amount of substrate cleaved, the absorbance change at 245nm per nmol of oligodeoxynucleotide cleaved needs to be known. This was calculated as being 0.043 OD units per nmol of d(pGAC[^{3C}G]ATATCGTC) cleaved, from the wavelength scan of a known amount of d(pGAC[^{3C}G]ATATCGTC) before and after total hydrolysis by the Eco RV endonuclease [see section 3c.2] as shown in Figure 3c.6. The UV trace obtained at 245nm for the cleavage of 8.3μM substrate could thus be converted into nmol of product formed as a function of time as shown in Figure 3c.7 together with the rate lines for the other substrate concentrations. The initial rates obtained were plotted as a Hanes graph (Figure 3c.7c).

Ideally, the substrate concentration should be considerably greater (preferably about ten fold) than that of the enzyme. However for the cleavage of the lowest concentration of d(pGAC[^{3C}G]ATATCGTC), the substrate concentration is only slightly higher than that of the endonuclease. Despite this, a good straight line was obtained for the Hanes plot and from this graph, a k_{cat} of 0.28min^{-1} was determined. The specificity constant of d(pGAC[^{3C}G]ATATCGTC) is thus 1.34 x 10^{3}s^{-1}M^{-1} which is over a hundred fold lower than that of the control dodecamer. A ΔG_{app} of -12kJ/mol was calculated from the relative specificity and this corresponds to -6.1kJ/mol for each d[^{3C}G] nucleoside. This value is of the order expected for the conservative deletion of a hydrogen bonding group (Fersht, 1987) and from this result it would be expected that the Eco RV endonuclease makes a hydrogen bond contact to the ring 3-nitrogen of the deoxyguanosine in the recognition site. However, no hydrogen bond contact to the 3-nitrogen of the deoxyguanosine is implicated in the crystal structure of an Eco RV endonuclease/oligodeoxynucleotide complex (Winkler, 1991). Nevertheless, it cannot be ruled out that a water mediated hydrogen bond may occur between the protein and the 3-nitrogen nor that during the reaction (*i.e.* in the transition state complex in the presence of magnesium), the enzyme may make a hydrogen bond to the 3-nitrogen. Also, as for d(pGAC[I]ATATCGTC), the group altered is in the minor groove which, in the crystal complex, is buried in the endonuclease. Therefore, even if no direct hydrogen bonds to the 3-nitrogen are involved, the enzyme is probably sensitive to changes in this vicinity. Another possible cause of the reduction in specificity of d(pGAC[^{3C}G]ATATCGTC) is that the presence of the deoxynucleoside analogue either alters the structure of the oligodeoxynucleotide in a way that affects the endonuclease reaction or that it alters the ability of the oligodeoxynucleotide to adopt the distorted DNA structure seen in the cognate crystal structure.

3c.3.1.v Kinetics of d(pGACGATAT[^{5Me}C]GTC)

The initial rates of cleavage of d(pGACGATAT[^{5Me}C]GTC) with 5nM endonuclease, determined as a function of substrate concentration are given in Figure 3c.8a and b. From the Hanes plot of these results (Figure 3c.8c), the Michaelis-Menten parameters were calculated to be 0.97μM for K_M and 21min^{-1} for k_{cat}. Thus, both k_{cat} and K_M are slightly reduced compared with the control dodecamer. The parameters give a specificity constant of 3.51 x 10^5s^{-1}M^{-1} for d(pGACGATAT[^{5Me}C]GTC) which is approximately twice that of the control. Almost identical results were obtained by Fliess *et al.* (1988) for d(AAAGATAT[^{5Me}C]TT). This oligodeoxynucleotide was also found to have a drop in both k_{cat} and K_M, and its specificity constant was 1.7 times that of the corresponding control undecamer. The slightly higher specificity constants for the d[^{5Me}C] containing oligodeoxynucleotides suggest a slightly higher preference by the Eco RV endonuclease for d[^{5Me}C] over dC. The apparent binding energy, ΔG_{app}, calculated for d(pGACGATAT[^{5Me}C]GTC) from its specificity constant was found to be +1.6kJ/mol. Thus, each of the extra 5-methyl groups on the d[^{5Me}C] residue contributes about +0.8kJ/mol to ΔG_{app}. It is therefore apparent that there is no steric hinderance between the methyl group on d[^{5Me}C] and the enzyme compared with deoxycytosine. The 5-methyl group of a thymidine nucleoside in place of deoxycytidine in the Eco RV site would thus not enable Eco RV endonuclease to distinguish it from deoxycytidine. The Eco RV endonuclease is known from a crystallized complex to make contact to the DNA in the vicinity of the deoxycytidine in the recognition site (Winkler, 1991). However from the van der Waals radii of the groups involved it appears possible that a 5-methyl group on the deoxycytidine could be accommodated in this complex.

5-methyldeoxycytidine has been used in place of deoxycytidine in a number of studies involving the cleavage of oligodeoxynucleotides by various endonucleases. With the Mva I endonuclease (Kubareva *et al.*, 1988) no effect on the cleavage rate was observed when 5-methyldeoxycytidine was substituted for the deoxycytidine in its recognition site (CC[T/A]GG). Conversely, if 5-methyldeoxycytidine replaced deoxycytidine in the recognition site of Eco RI no cleavage was observed for the Eco RI endonuclease (Brennan *et al.*, 1986) nor for its isoschizomer, Rsr I endonuclease (Aiken *et al.*, 1991). Replacement of the first deoxycytidine of the recognition site of Eco RII (CC[T/A]GG) reduced the Eco RII endonuclease cleavage rate whereas replacing the inner deoxycytidine with 5-methyldeoxycytidine completely abolished activity (Yolov *et al.*, 1985). The latter result for Eco RII is not surprising since the Eco RII methylase, like several other type II modification methylases, protects DNA from cleavage by the corresponding endonuclease by methylation of the 5-position of a deoxycytidine (Nelson & McClelland, 1991) which in the case of Eco RII is the inner deoxycytidine (Boyer *et al.*, 1973).



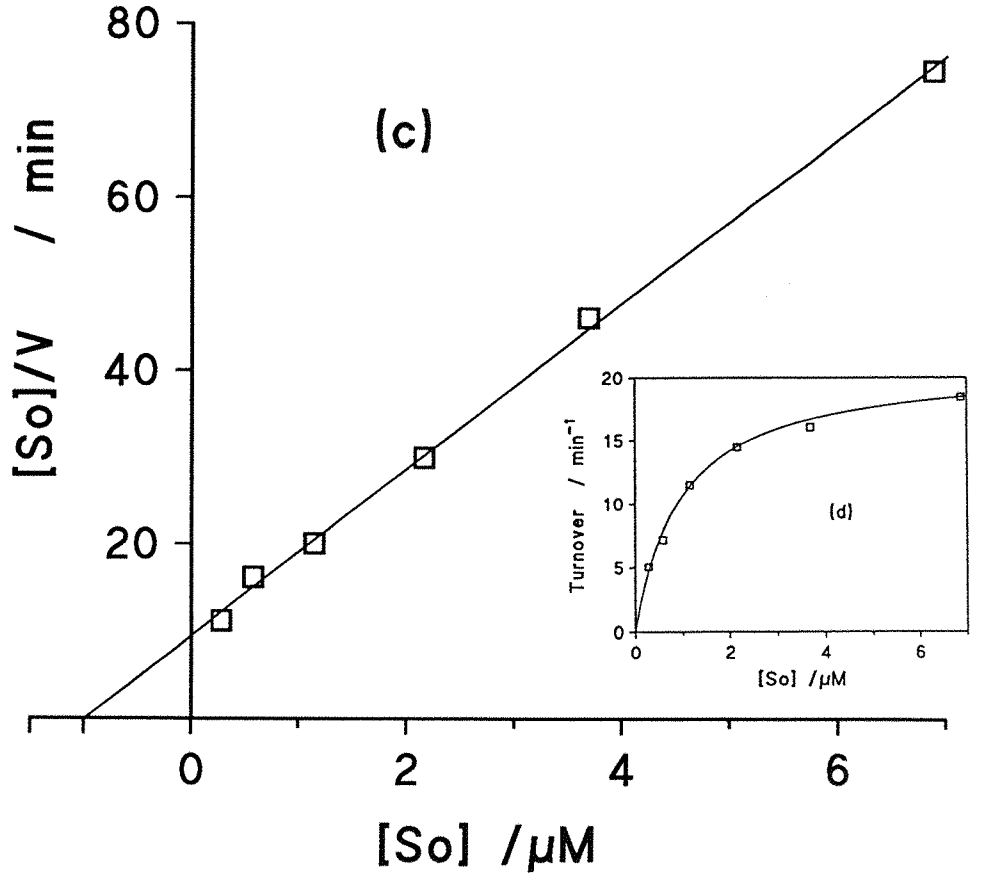


Figure 3c.8: Eco RV endonuclease cleavage reaction of d(pGACGATAT[^{5Me}C]GTC) at various substrate concentrations. **(a)** Rate lines obtained in 2ml volumes using 2cm path length cells. **(b)** Rate lines obtained in 1ml volumes using 1cm path length cells. **(c)** Hanes plot. **(d)** Plot of initial substrate concentration, [S_0], versus initial turnover rate.

3c.3.1.vi Kinetics of d(pGACGATAT[^{4H}C]GTC)

Using the UV assay, six good rate lines were obtained for the cleavage of d(pGACGATAT[^{4H}C]GTC) by 5nM endonuclease and are shown in Figures 3c.9a and b. The Michaelis-Menten parameters obtained from the straight line in the Hanes graph (Figure 3c.9c) were 38min^{-1} for k_{cat} and 2.0μM for K_M. These values are very similar to those of the phosphorylated control dodecamer and give a specificity constant of 3.17 x 10^5s^{-1}M^{-1}. This is 1.7 times the specificity constant of d(pGACGATATCGTC) and gives a ΔG_{app} for d(pGACGATAT[^{4H}C]GTC) of +1.3kJ/mol (or +0.66kJ/mol for each d[^{4H}C] residue). Since the conservative deletion of the 4-amino group on deoxycytidine by replacing it with a hydrogen is expected to fulfil the necessary criteria of not causing steric hindrance and not involving charged groups [see section 1.2.1] the assumption can be made that ΔG_{app} approximately equals ΔG_{bind} (Fersht, 1987). The very small value for ΔG_{bind} of +0.66kJ/mol for each 4-amino group deleted can be taken as zero within experimental error. This result therefore suggests that no contact is made between the Eco RV endonuclease and the 4-amino group of the deoxycytidine in the recognition site. This contradicts the prediction from the crystal structure of d(GGGATATCCC) bound to the Eco RV endonuclease that a hydrogen bond exists between the 4-amino group of the deoxycytidine and the main chain carbonyl oxygen of amino acid 182 (Winkler, 1991 and see Figure 3c.2). However, this contact shows considerable discrepancy between the two strands of the DNA in the complex. In one strand, the interatomic distance between the protein carbonyl oxygen and the deoxycytidine 4-amino group is 2.8Å which is in the range expected for a NH···O type hydrogen bond (2.7Å to 3.1Å; Saenger, 1984). In the other strand, though, this distance is 3.2Å which is outside the expected hydrogen bonding range and so in reality, the hydrogen bond to the deoxycytidine 4-amino group may not exist. Unfortunately, since a protein main chain atom is involved, it cannot be mutated to confirm (or disprove) the result obtained here with d(pGACGATAT[^{4H}C]GTC).

Interestingly, the result for d(pGACGATAT[^{4H}C]GTC) and that for d(pGACGATAT[^{5Me}C]GTC) suggest that the Eco RV endonuclease may not distinguish directly between deoxycytidine and thymidine in the sixth position of the Eco RV site. The 4-amino group of deoxycytidine does not appear to be important and one might predict that the 4C=O group of thymidine could be rather well tolerated. The enzyme also tolerates the presence of the 5-methyl group of thymidine since the substitution by d[^{5Me}C] has no real effect on the reaction. Of course, the presence of deoxycytidine can be distinguished indirectly by contacting its base paired partner, deoxyguanosine, but the results of this study predict that a T.G mismatch base pair at this position should be a reasonably good Eco RV endonuclease substrate.

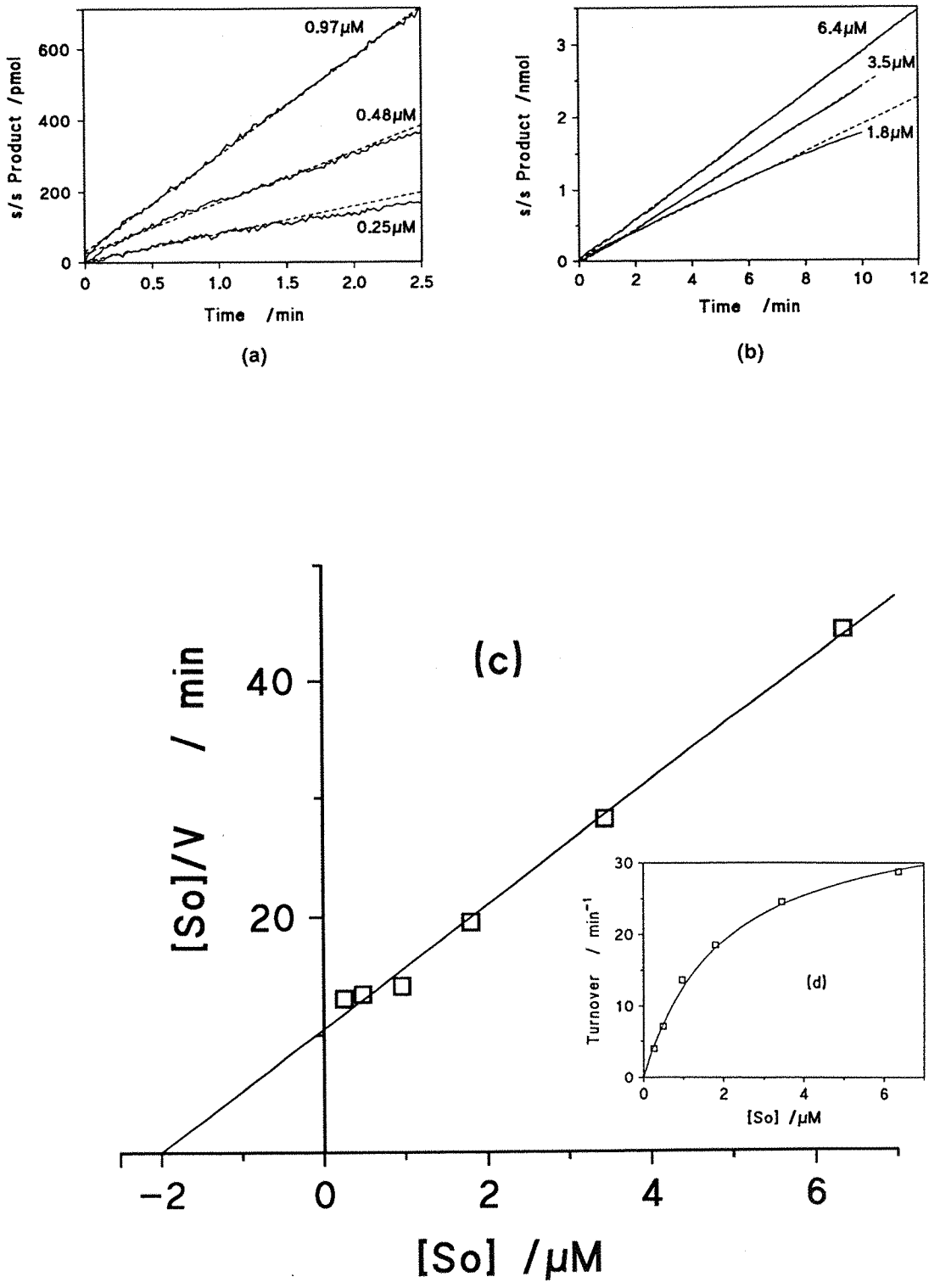


Figure 3c.9: Eco RV endonuclease cleavage reaction of d(pGACGATAT[^{4H}C]GTC) at various substrate concentrations. **(a)** Rate lines obtained in 2ml volumes using 2cm path length cells. **(b)** Rate lines obtained in 1ml volumes using 1cm path length cells. **(c)** Hanes plot. **(d)** Plot of initial substrate concentration, $[S_0]$, versus initial turnover rate.

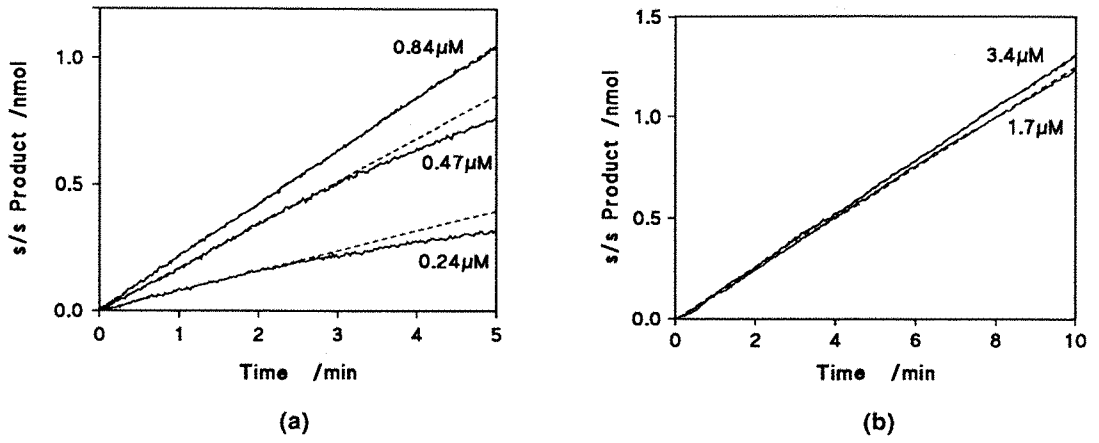
3c.3.2 Michaelis-Menten Kinetic Analysis of the Non-Phosphorylated Control Dodecamer, d(GACGATATCGTC) and Tridecamer, d(TGACGATATCGTC)

As mentioned in section 3c.3.1, no significant change in the rate of cleavage of d(GACGATATCGTC) was observed using substrate concentrations from 6.4μM to 1.3μM, indicating a K_M of <<1μM. In an attempt to obtain values for k_{cat} and K_M for d(GACGATATCGTC), the rates of cleavage of substrate concentrations lower than 1μM were determined. The results obtained using 2nM endonuclease are shown in Figure 3c.10 and over this substrate range, changes in rate were indeed observed. Although ideally rates at lower substrate concentrations should be obtained to cover a wider range relative to K_M, a reasonable straight line was obtained for the Hanes plot. From this line, a K_M of 0.47μM and a k_{cat} of 38min^{-1} were obtained. The k_{cat} of 38min^{-1} is essentially the same as the 33min-1 obtained in this study for the phosphorylated control dodecamer, d(pGACGATATCGTC). Therefore the maximum rate of oligodeoxynucleotide hydrolysis is the same for the phosphorylated and the non-phosphorylated dodecamers. The K_M value for the non-phosphorylated control, however, is less than a sixth of that for the phosphorylated oligodeoxynucleotide (K_M=2.9μM).

The decreased K_M with no change in k_{cat} for d(GACGATATCGTC) means that the presence of the 5'-phosphate affects binding events prior to the actual cleavage reaction. One of these steps that could plausibly be responsible for the different K_M value is the formation of the endonuclease/substrate complex. To attempt to distinguish whether this reduced binding was due to steric or charge repulsion of the terminal phosphate, the Michaelis-Menten parameters of the tridecamer d(TGACGATATCGTC) were determined.

The tridecamer, d(TGACGATATCGTC) is of the same sequence as the control dodecamer except that it has an extra overhanging thymidine on the 5'-end. d(TGACGATATCGTC) contains a phosphate in the equivalent position to the 5'-phosphate of d(pGACGATATCGTC), but unlike the terminal phosphate which carries a double negative charge at the assay pH, it is a phosphodiester and therefore carries only a single negative charge. If the effect of the 5'-phosphate of d(pGACGATATCGTC) is due to steric hindrance then the K_M of d(TGACGATATCGTC) would be expected to be the same as that of the phosphorylated dodecamer since the phosphate is still present. If however it is charge repulsion of the terminal phosphate of d(pGACGATATCGTC) that is responsible for its higher K_M, then the K_M of d(TGACGATATCGTC) would be expected to be different to that of d(pGACGATATCGTC).

The initial rate lines for the cleavage of d(TGACGATATCGTC) with 5nM endonuclease determined at six different substrate concentrations are shown in Figure 3c.11. Initial rates were also plotted in a Hanes graph (Figure 3c.11c) to give a good straight line showing that the tridecamer follows Michaelis-Menten kinetics. The line yielded a k_{cat} of 35min^{-1} and a K_M of 0.37μM. d(TGACGATATCGTC) therefore has a K_M approximately an eighth of that for d(pGACGATATCGTC) (2.96μM) but virtually identical to that of d(GACGATATCGTC)



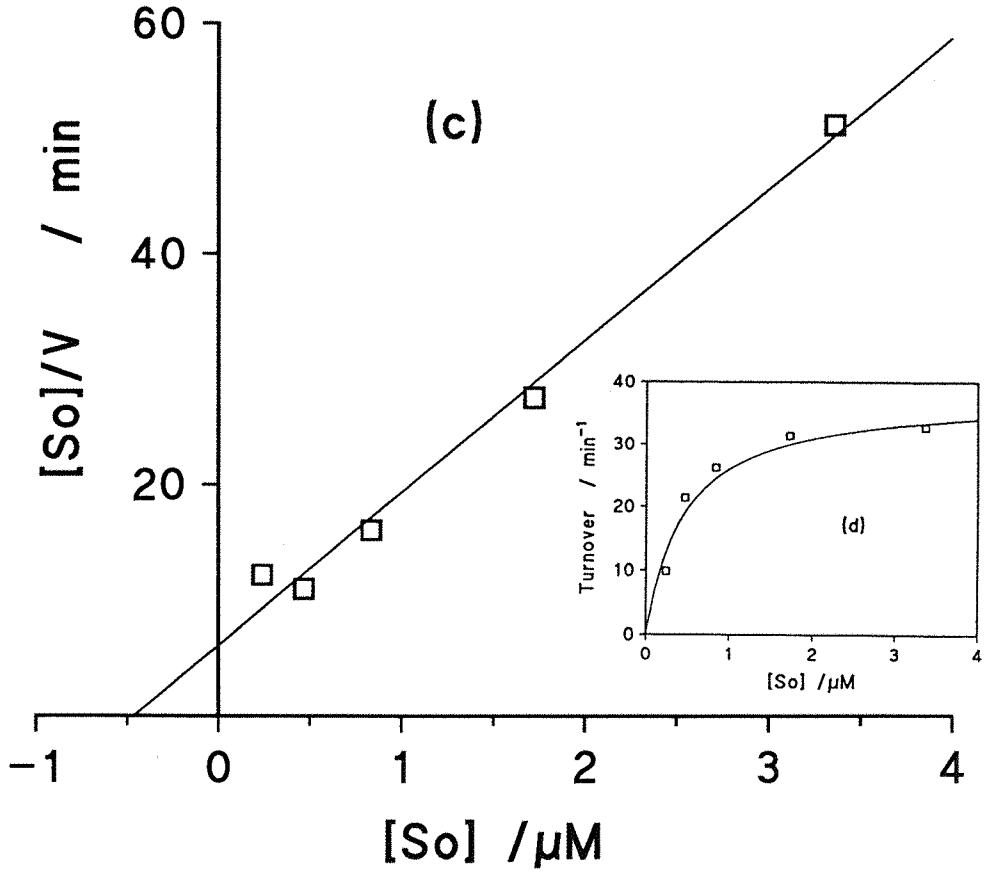


Figure 3c.10: Eco RV endonuclease cleavage reaction of d(GACGATATCGTC) at various substrate concentrations. **(a)** Rate lines obtained in 2ml volumes using 2cm path length cells. **(b)** Rate lines obtained in 1ml volumes using 1cm path length cells. **(c)** Hanes plot. **(d)** Plot of initial substrate concentration, $[S_0]$, versus initial turnover rate.

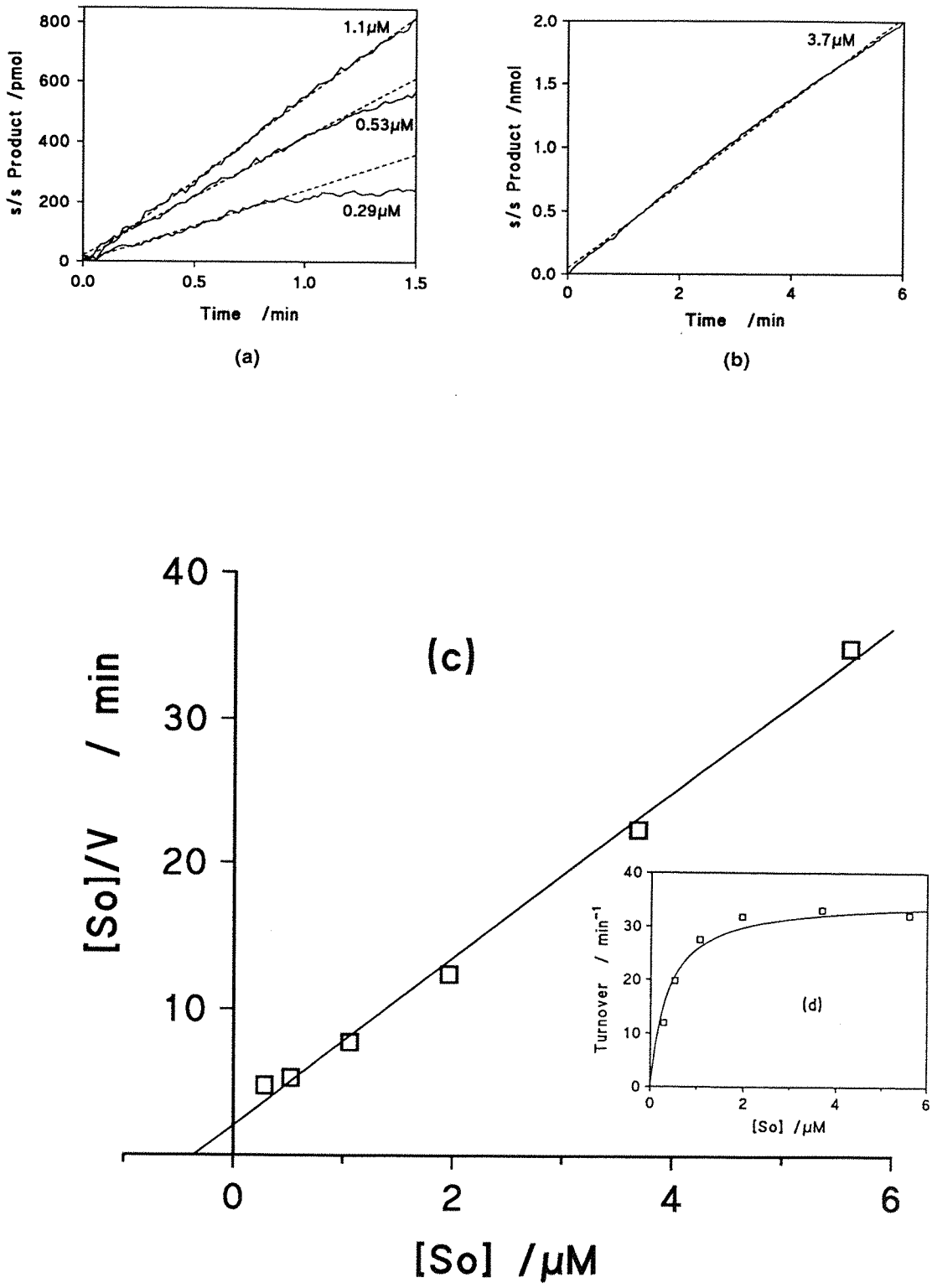


Figure 3c.11: Eco RV endonuclease cleavage reaction of d(TGACGATATCGTC) at various substrate concentrations. **(a)** Rate lines obtained in 2ml volumes using 2cm path length cells. **(b)** Rate line obtained in 1ml volumes using 1cm path length cells; Cleavage reactions of 2.0μM and 5.6μM substrate were virtually identical to the one shown. **(c)** Hanes plot. **(d)** Plot of initial substrate concentration, $[S_0]$, versus initial turnover rate.

(0.47μM). The k_{cat} values for all three oligodeoxynucleotides are approximately equal (results summarised in Table 3c.4). It can therefore be concluded that the 5'-phosphate group of d(pGACGATATCGTC) does not cause an increase in K_M due to steric effects, but rather because of the negative charges associated with it. When the equivalent phosphate is a phosphodiester, as in d(TGACGATATCGTC), the increase in K_M is abolished. This result suggests that the 5'-phosphate of d(pGACGATATCGTC) is brought near to a negative charge on the protein on binding to the endonuclease. However, since the crystal structure of the cognate complex contains an oligodeoxynucleotide of only ten base pairs long, it is difficult to speculate whether any negatively charged protein residues expected to be in the vicinity of this phosphate.

Table 3c.4: Michaelis-Menten parameters for the phosphorylated and non-phosphorylated control dodecamers and for the control tridecamer.

Oligodeoxynucleotide	k_{cat} /min^{-1}	K_M /μM
d(GACGATATCGTC)	38	0.47
d(pGACGATATCGTC)	33	2.9
d(TGACGATATCGTC)	35	0.37

CHAPTER 3D

RESONANCE RAMAN SPECTROSCOPY

3d.1 RESONANCE RAMAN SPECTROSCOPY

The UV spectrum of 4-thiothymidine (Figure 3d.1) shows a λ_{max} at 335nm. This is well removed from the λ_{max}'s of other chromophores present in proteins and nucleic acids and so it was thought that this would make 4-thiothymidine a useful resonance Raman probe for studying protein/DNA systems.

To investigate this, three oligodeoxynucleotides were synthesised using an automated DNA synthesizer. The method employed to introduce the 4-thiothymidine nucleoside into oligodeoxynucleotides has been reported previously (Nikiforov & Connolly, 1991) and involves the incorporation of a 4-S-(*p*-nitrophenyl) protected 4-thiothymidine moiety into the required position.

Using this synthesis protocol the two dodecamers, d(GACGA[^{4S}T]ATCGTC) and d(GACGATA[^{4S}T]CGTC), and the pentamer, d(AG[^{4S}T]TC), were prepared. These were purified and desalted as described in sections 2.2.5 and 2.2.6 to give oligodeoxynucleotides that were pure as judged by reverse phase HPLC. The two dodecamers have previously been characterised by CD spectroscopy, base composition analysis and melting point determination (Connolly & Newman, 1989). On the basis of the CD spectra and the melting temperatures they are both expected to be double stranded and predominantly B-form DNA at room temperature. The extinction coefficients at 254nm of the two dodecamers have also been previously determined (77.5 x $10^3 M^{-1}cm^{-1}$ for single stranded d(GACGA[^{4S}T]ATCGTC) and 78.0 x $10^3 M^{-1}cm^{-1}$ for single stranded d(GACGATA[^{4S}T]CGTC)). That for d(AG[^{4S}T]TC) was determined using the method described in section 2.2.7.ii and was found to be 40.1 x $10^3 M^{-1}cm^{-1}$ for the single stranded oligodeoxynucleotide. The UV spectra of these oligodeoxynucleotides are shown in Figure 3d.1 together with that of 4-thiothymidine. The UV peak due to the presence of the 4-thiothymidine residue in the oligodeoxynucleotides is clearly visible and is well resolved from the absorbance due to the other bases. The intensity of this peak is much less in the dodecamers than that of the free deoxynucleoside at equal molar concentrations due to the hyperchromic effect. As expected for a single stranded oligodeoxynucleotide, the 4-thio peak exhibits a much smaller hyperchromic effect in d(AG[^{4S}T]TC). Another point of interest is the different λ_{max} values for the 4-thio chromophore for the three oligodeoxynucleotides. In d(GACGATA[^{4S}T]CGTC) and in d(AG[^{4S}T]TC) it is at 337nm which is essentially the same wavelength as for the free nucleoside. In d(GACGA[^{4S}T]ATCGTC) however, it is shifted up by 12nm to 349nm. This indicates that the 4-thiothymidine base is in different environments in the three different oligodeoxynucleotides.

The sequences of the two dodecamers d(GACGA[^{4S}T]ATCGTC) and d(GACGATA[^{4S}T]CGTC) were chosen in order to ultimately study their interaction with the Eco RV endonuclease by resonance Raman spectroscopy. Their enzymatic reactions with the endonuclease have already been investigated (Newman *et al.*, 1990b). It was found that d(GACGA[^{4S}T]ATCGTC) was cleaved at a reduced rate but that d(GACGATA[^{4S}T]CGTC) was not cleaved at all. The specificity constant of d(GACGA[^{4S}T]ATCGTC) relative to

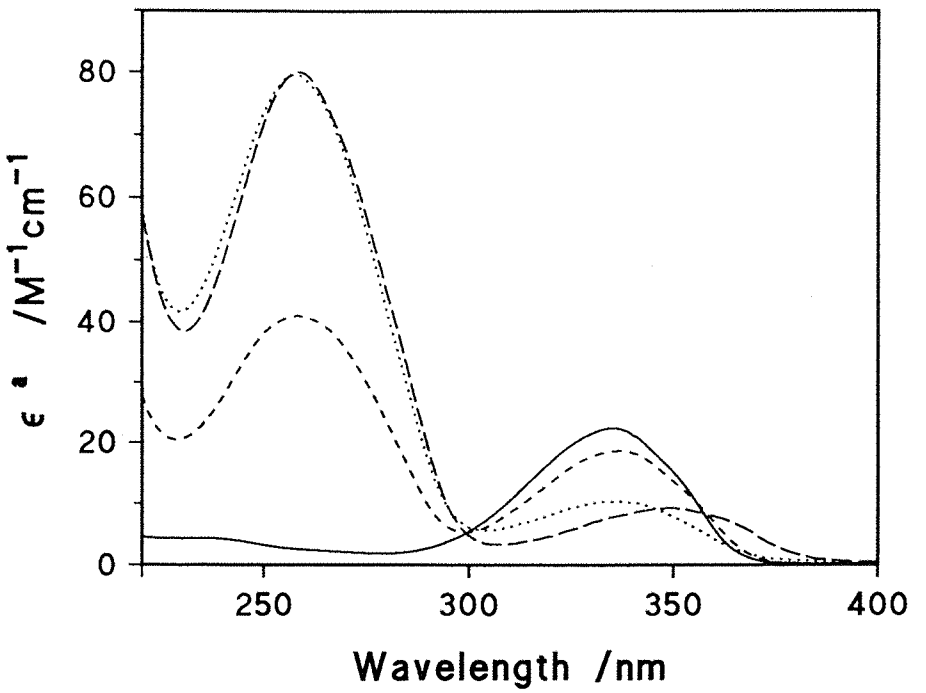


Figure 3d.1: UV spectra at pH 7.2 of 4-thiothymidine (——), d(AG[^{4S}T]TC) (---), d(GACGA[^{4S}T]ATCGTC) (— — —) and d(GACGATA[^{4S}T]CGTC) (···). ^{a}For single stranded DNA.

d(GACGATATCGTC) was found to be 0.02 with most of this reduction resulting from a reduction in k_{cat} and not by an increase in K_M.

The pentamer d(AG[^{4S}T]TC) was chosen since it was expected to model the behaviour of 4-thiothymidine present in single stranded DNA. The pentamer was not characterised by CD spectroscopy nor was its melting temperature determined but because of its short length and the lack of self-complementarity, it is expected to be a random coil at 25°C. A comparison of the Raman spectrum of d(AG[^{4S}T]TC) and the two dodecamers will therefore highlight differences in the properties of 4-thiothymidine in double and single stranded DNA.

3d.2 COLLECTION OF RAMAN SPECTRAL DATA

Initially an excimer laser was used to excite ('pump') a dye laser to produce a pulsed (10ns pulse of energy 6mJ/pulse at 10Hz) laser beam of 340nm wavelength. This pulsed laser beam was used as the Raman excitation source in attempts to record resonance Raman spectra of 4-thiothymidine. The scattered Raman radiation was collected at 90 degrees to the incident beam and was electronically gated so that only the first 30ns of radiation scattered after the laser pulse was collected. This was to reduce the amount of background scattered light collected. Using this set-up, reasonable resonance Raman spectra could be obtained from 100ml samples of 450μM (A_{335}=10) 4-thiothymidine that were circulated through the laser beam by

an electronic pump. Even with the gating system, the background was high and the Raman signal was found to decay significantly with the length of exposure to the laser.

Much better spectra were obtained using a continuous beam argon ion laser source. Although the average power of the two lasers is approximately the same, the pulsed dye laser produces its power all in one pulse whereas the argon ion laser is continuous. The peak power of the dye laser (6MW) is so large that it is far more likely to cause problems such as heating of the sample, pushing the molecule into an excited state, and inducing photochemical reactions. With the lower continuous power (<100mW) of the argon ion laser, these problems were greatly reduced.

Using the argon ion excitation source and placing the sample in a rotating quartz cell (10mm diameter x 3mm high) the amount of sample needed was greatly reduced. With the set-up shown in Figure 3d.2 good spectra could be obtained using as little as 100µl of 90µM (A_{335}=2; a total of 2.3µg) 4-thiothymidine. The argon ion laser line at 363.8nm was used and was passed through a Pellin-Broca prism before being directed onto the sample in the rotating cell. The beam was focused onto the sample and scattered radiation collected at 180 degrees to the incident beam. This backscattered radiation was focused into the slit of the spectrograph using the collecting lens. Radiation of different wavelengths was then separated by a dispersion grating and then detected with a diode array. A grating of 2400g/mm was used as it gave the highest resolution of wavenumbers and with this grating a range of about 800cm^{-1} could be viewed at any one time.

Despite not gating the Raman signal, the background with the argon ion laser was much less as compared with the pulsed laser. However it was still found (although to a lesser extent than with the pulsed laser) to increase with laser exposure. This background was thought to be due to the fluorescence of a compound in the sample. Since it increased with exposure it was apparent that the fluorescence was not due to the 4-thiothymidine itself but rather a photo-induced reaction product of it. It was found that deoxygenation of the sample and keeping it under argon reduced the fluorescence background and greatly reduced the rate of further increase in the fluorescence. It therefore seemed likely that the fluorescent product resulted from a photo-induced reaction between oxygen and 4-thiothymidine. Oxygen would not be photochemically activated directly by 340nm radiation but by collision with an excited 4-thiothymidine molecule. This possible scheme is outlined in Scheme 3d.1. Firstly a 4-thiothymidine molecule absorbs a photon of 340nm wavelength causing an electronic transition to the next allowed excited state, a singlet state (^{1}T). The singlet state may then spontaneously emit a photon (fluorescence) and return to the ground state. Alternatively, the singlet molecule, ^{1}T, may transfer to the lower excited triplet state (^{3}T; *i.e.* has two unpaired electrons) by a process known as intersystem crossing. This triplet state may again spontaneously emit a photon (phosphorescence) and return to the ground state or alternatively it may be quenched by collision with another molecule. Oxygen, which is triplet in its ground state, is a known

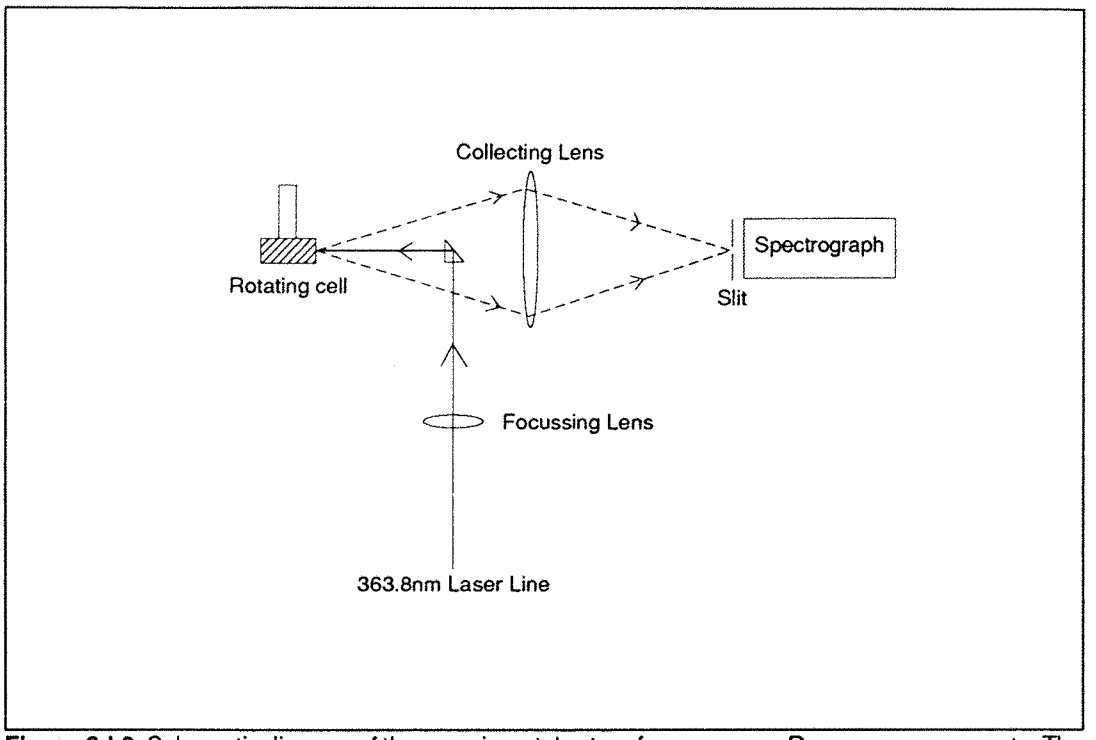
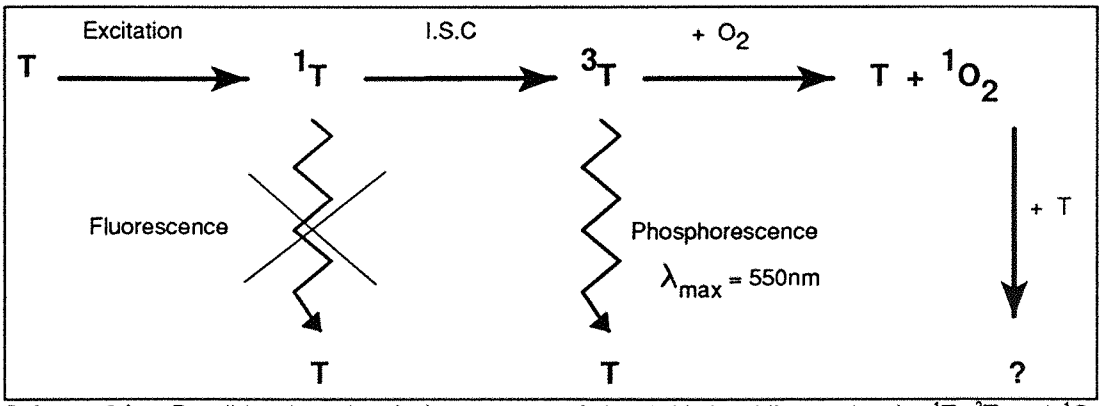


Figure 3d.2: Schematic diagram of the experimental set-up for resonance Raman measurements. The laser is focused onto the sample in the rotating cell and backscattered radiation is directed into the spectrograph slit.



Scheme 3d.1: Possible photochemical processes of the 4-thiothymidine molecule. ^1T, ^3T and $^1O_2$ represent singlet ^{4S}T, triplet ^{4S}T and singlet oxygen respectively. I.S.C. = intersystem crossing.

quencher of excited triplet state molecules. This quenching results in the formation of singlet oxygen, $^1O_2$, which is a very reactive species. Evidence for this kind of process comes from work carried out on excited state 4-thiouridine (Salet *et al.*, 1983); a nucleoside very similar in structure to 4-thiothymidine. It is known that 4-thiouridine forms a triplet state on irradiation with 353nm light (must be via the singlet state since a direct transition to the triplet is forbidden) and that this can be quenched with oxygen. It has also been found that 4-thiouridine does not fluoresce, but that it does exhibit a weak phosphorescence with a λ_{max} at 550nm. (Salet *et al.*, 1983; Pochon *et al.*, 1971). Therefore the singlet state does not decay by fluorescence. It thus seems that a reaction of singlet oxygen with 4-thiothymidine produces a species that fluoresces in the 363.8nm laser beam and that removing oxygen from the sample

greatly reduces this process. Examination by reverse phase HPLC of samples of 4-thiothymidine exposed to the laser showed that several minor products were being produced. The most abundant of these was found by co-elution with a standard to be thymidine. The amount of the photo-products was found to be greater in samples with oxygen present than in deoxygenated samples. As thymidine does not absorb at 363nm it cannot fluoresce when irradiated with this wavelength light and so the fluorescence must be due to one of the other photo-products. These products were produced only in quantities of about 1 to 2% but since fluorescence is much more intense than Raman (approximately a million fold), only a small amount of the compound needs to be present to produce a problematic effect.

3d.3 RESONANCE RAMAN SPECTROSCOPY OF FREE 4-THIOTHYMIDINE

Using the set-up shown in Figure 3d.2, good quality resonance Raman spectra of 100µl of 450µM (A_{335}=10) 4-thiothymidine in water were recorded as described in section 2.5.4. Two separate ranges of wavenumbers were observed (600-1450cm^{-1} and 1000-1800cm^{-1}) and the raw data obtained in this way are shown in Figures 3d.3a and b. The spectra were calibrated by Raman peaks of ethanol, carbon tetrachloride, toluene and acetone, and then the peak positions of 4-thiothymidine determined. The two spectra were merged to give Figure 3d.3c which then had the background fluorescence removed by subtraction of the polynomial fitted curve shown in the figure. This resulted, after smoothing of the data, in the final spectrum shown in Figure 3d.3d. All other spectra given in this and later sections [except section 3d.5] were produced in this manner.

Spectra of 4-thiothymidine in buffer α0 and in buffer η showed no noticeable differences to those recorded in water.

In order to be able to interpret any changes that were observed in the Raman spectrum of 4-thiothymidine when it was present in oligodeoxynucleotides, and any changes observed on

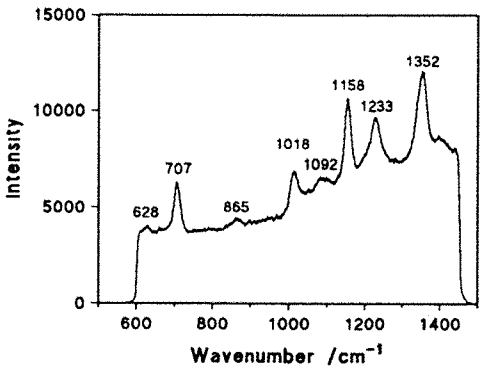


Figure 3d.3a: Raw resonance Raman spectrum of 450µM 4-thiothymidine in water (low wavenumber region).

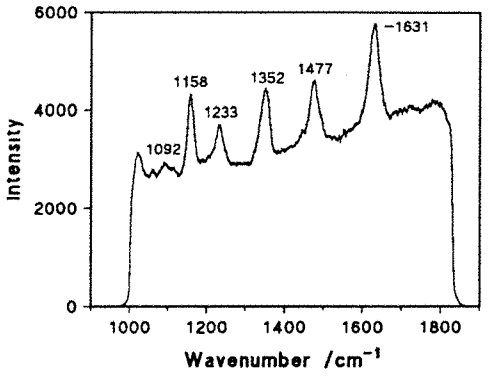


Figure 3d.3b: Raw resonance Raman spectrum of 450µM 4-thiothymidine in water (high wavenumber region).

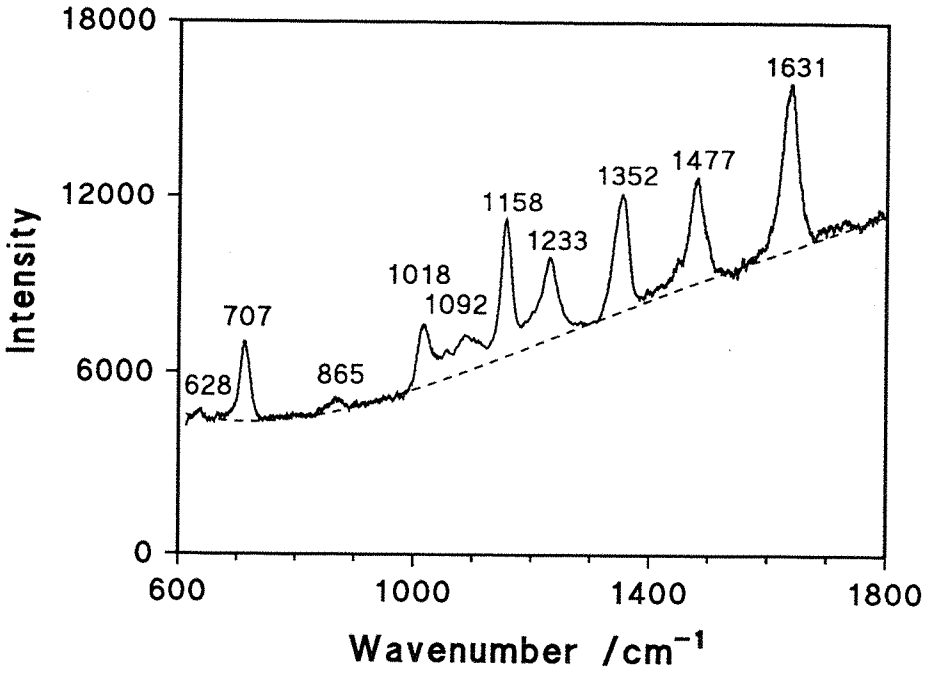


Figure 3d.3c: Merged resonance Raman spectra from high and low wavenumber regions of 4-thiothymidine. The dashed line shows the polynomial fitted curve used to subtract out background fluorescence.

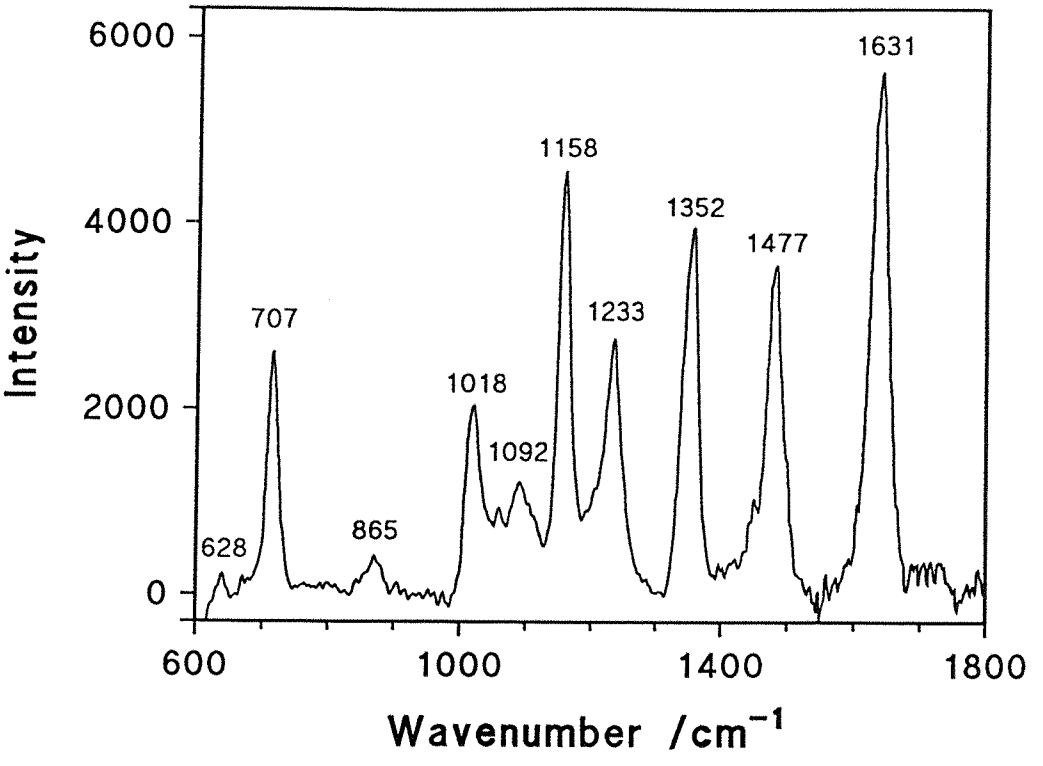


Figure 3d.3d: Final fluorescence subtracted and smoothed resonance Raman spectrum of 4-thiothymidine.

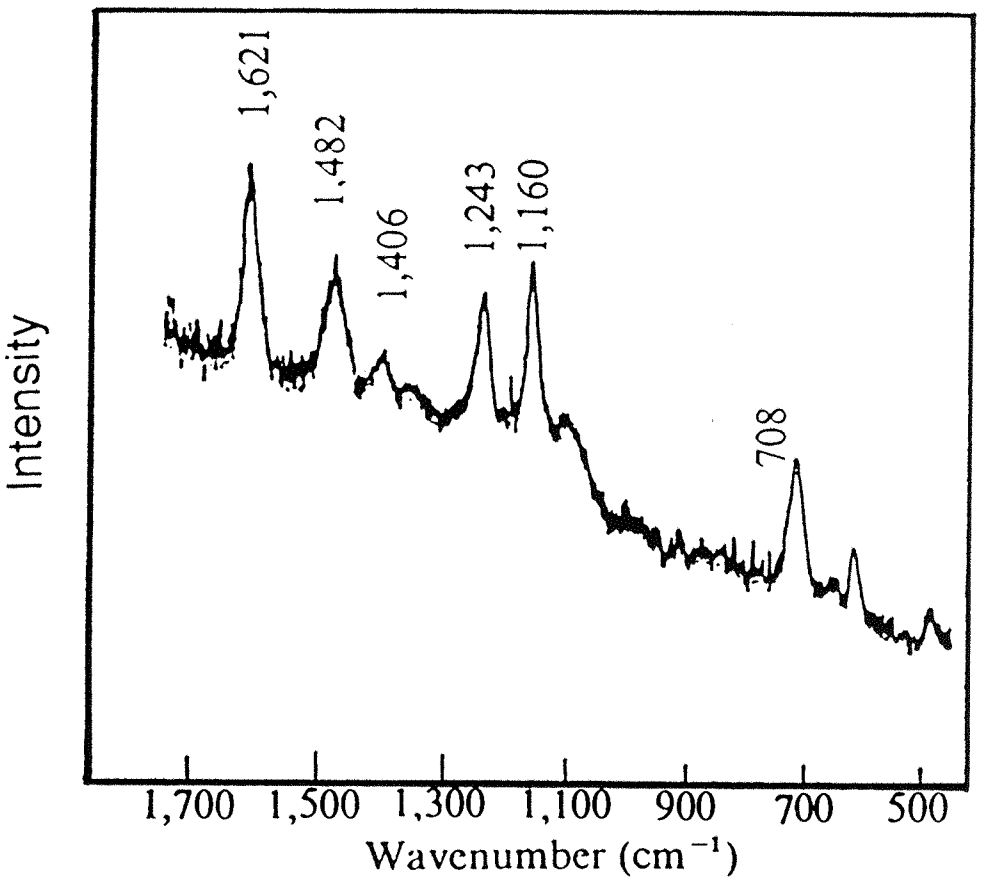


Figure 3d.4: Resonance Raman spectrum of 4-thiouridine taken from Nishimura *et al.*, 1976.

protein binding of these oligodeoxynucleotides, the peaks need to be assigned to particular vibrations of the molecule. This was greatly aided by the work that has been done previously on uracil, uridine and 4-thiouridine (for a review see Tsuboi *et al.*, 1987). The resonance Raman spectrum of 4-thiothymidine is very similar to that of 4-thiouridine (Nishimura *et al.*, 1976; Nishimura *et al.*, 1978) as shown in Figure 3d.4. Several of the 4-thiothymidine peaks can thus be immediately assigned to the same vibrational modes as for 4-thiouridine (see Table 3d.1). The assignments of the 4-thiouridine peaks and the nomenclature used to describe the vibrations was discussed in section 1.7.2. These are based mostly on studies done on uracil and uracil derivatives including uridine monophosphate (Tsuboi *et al.*, 1987). Thus the 1631cm^{-1} peak is assigned to the ν_b^6 ring stretch (which includes the C5=C6 stretch), the 1477cm^{-1} peak to the ν_a^6 ring stretch and the 1243cm^{-1} peak to the Kekulé (Kk) ring stretch. As for the 4-thiouridine molecule, 4-thiothymidine also has two peaks at 707cm^{-1} and 1158cm^{-1} which are considered to be due to the C=S stretch being strongly coupled to the ring breathing vibration (Br^6; Nishimura *et al.*, 1978). The main difference between the 4-thiouridine and the 4-thiothymidine spectra is the appearance of an intense peak at 1352cm^{-1} in the latter. In the 4-thiouridine spectrum, a much weaker peak is seen at 1372cm^{-1} and this has been assigned to a blend of N1-C1' stretch and the ring stretch, $\nu_{\delta a}^6$ (Tsuboi *et al.*, 1987). The strong peak in the 4-thiothymidine spectrum at 1352cm^{-1} can be assigned to a combination of the N1-C1' stretch coupled with the $\nu_{\delta a}^6$ ring stretch together with the C5-Me stretch on the basis of the

Table 3d.1: Resonance Raman peaks and their vibrational assignments for 4-thiouridine [see section 1.7.2], 4-thiothymidine in water and 4-thiothymidine in deuterium oxide (N3-d_1). §May be the other way around (see text).

VIBRATIONAL MODE	4-Thiouridine	4-Thiothymidine	N3-d_1 4-Thiothymidine
ring stretch, ν_b^6 (C5=C6)	1619	1631	1632
ring stretch, ν_a^6	1482	1477	1477
unknown	—	—	1404
N1-C1' stretch coupled with ring stretch, $\nu_{\delta a}^6$	1372	1352	1366
C5-methyl stretch	—		1330
Kekulé, Kk	1243	1233	1218
ring breath, Br^6, coupled with C=S stretch	1160	1158	1163
§ ring stretch, $\nu_{\delta b}^6$	—	1092	1090
§ ring deformation, Tr	—	1018	1017
unknown	—	865	—
C=S stretch coupled with ring breath, Br^6	709	707	699
ring deformation, δ_a^6	—	628	—

following:- (i) The 5-methyl group is the only structural difference between 4-thiouridine and 4-thiothymidine, (ii) an extra, strong peak at 1379cm^{-1} is also a prominent difference between the resonance Raman spectra of deoxythymidine monophosphate and that of deoxyuridine monophosphate (Fodor *et al.*, 1985), and (iii) a peak at 1376cm^{-1} in the normal mode Raman spectra of oligodeoxynucleotides has been assigned to the thymidine C5 methyl group (Benevides *et al.*, 1991).

This just leaves the minor peaks at 1092cm^{-1}, 1018cm^{-1} and 628cm^{-1}. On the basis of studies of uracil and uracil derivatives, these three have been assigned to the ring stretch, $\nu_{\delta b}^6$, the ring deformation, Tr, and the ring deformation, δ_a^6, respectively. This also fits in with the fact that symmetric modes are generally more intense in Raman spectroscopy than asymmetric modes, since the more intense band at 1018cm^{-1} is assigned to the symmetric Tr deformation and the less intense 1092cm^{-1} band is assigned to the more asymmetric mode, $\nu_{\delta b}^6$.

It is interesting to compare the resonance Raman spectrum of 4-thiothymidine with its infrared (IR) spectrum shown in Figure 3d.5. The positions of the Raman peaks are indicated in the spectrum. Several of these peaks can immediately be seen to correspond to peaks in the IR spectrum. Namely the Raman peaks at 1631cm^{-1}, 1477cm^{-1} and 1092cm^{-1}. The peak at 1155cm^{-1} in the IR spectrum also probably corresponds to the Raman peak at 1158cm^{-1}. A noticeable absence in the IR spectrum is the lack of a C=S peak at 707cm^{-1}. Also

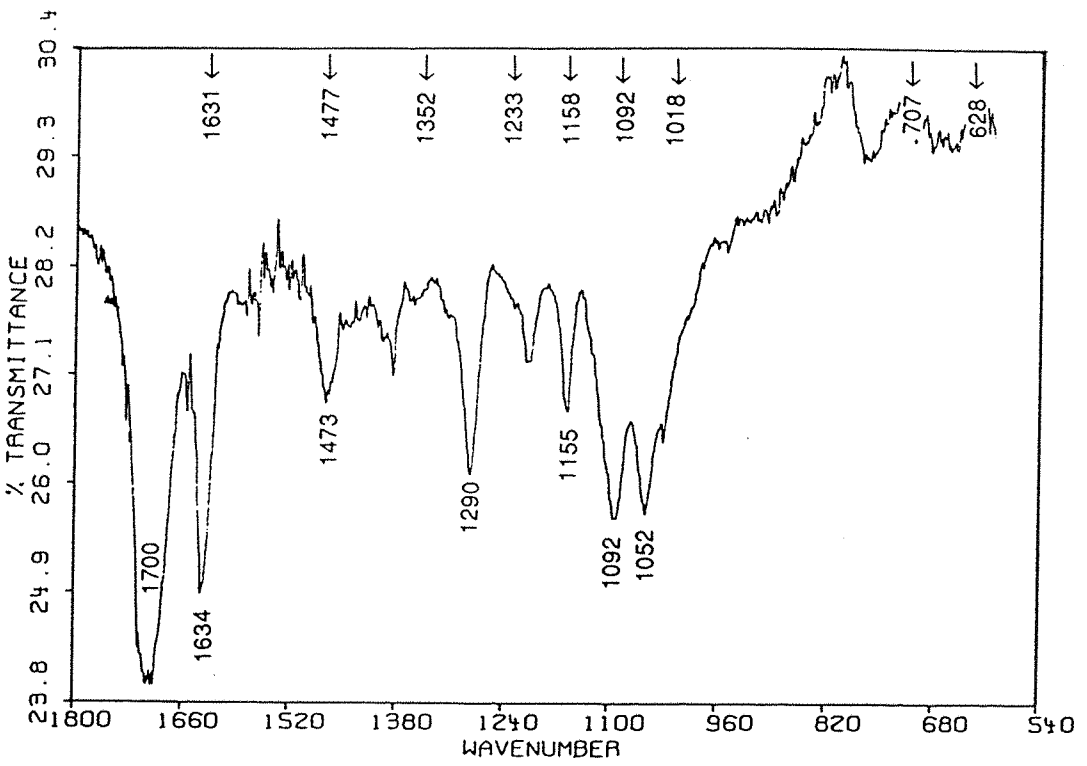


Figure 3d.5: Infrared spectrum of 4-thiothymidine (provided by C.Wharton, University of Birmingham). The positions of peaks seen in the resonance Raman spectrum are indicated by arrows.

important is the presence of a strong peak in the IR spectrum at around 1700cm^{-1}. This can be assigned to the C2=O stretching vibration which is absent from the resonance Raman spectrum for the reasons stated in section 1.7.2.

To assist in the assignment of resonance Raman bands, spectra of 4-thiothymidine in deuterium oxide were run. The replacement of the exchangeable proton on N3 in 4-thiothymidine with deuterium was expected to affect some of the bands. This was indeed found to be the case as shown in Figure 3d.6. The main changes compared with the water spectra was the appearance of two extra peaks in the 1200-1450cm^{-1} region and also wavenumber shifts of a few of the other peaks. The origin of the extra peak at 1401cm^{-1} in D_2O is not known. The peaks at 1330cm^{-1} and 1366cm^{-1} in D_2O were attributed to the splitting up of the 1352cm^{-1} peak seen in water. The 1352cm^{-1} peak seen in the water spectrum was assigned to the two modes, $\nu_{\delta a}{}^{6}$ plus N1-C1' stretch and the C5-Me stretch, being degenerate and therefore producing a single band. Presumably deuteriation affects these two modes in different ways causing the 1352cm^{-1} peak to be split up into the two components. This is supported by the fact that the total relative areas of the two peaks in deuterium oxide is approximately equal to the relative area of the 1352cm^{-1} peak in water. These and other peak assignments are given in Table 3d.1. Other changes observed include the shift of the 707cm^{-1} C=S peak to 699cm^{-1} and this seems to be due to the difference between the C=S group being solvated in D_2O as opposed to H_2O. A corresponding shift of similar magnitude to higher wavenumbers is seen for the peak from the coupled ring breath and C=S stretch mode, to 1163cm^{-1}. The Kekulé ring vibration

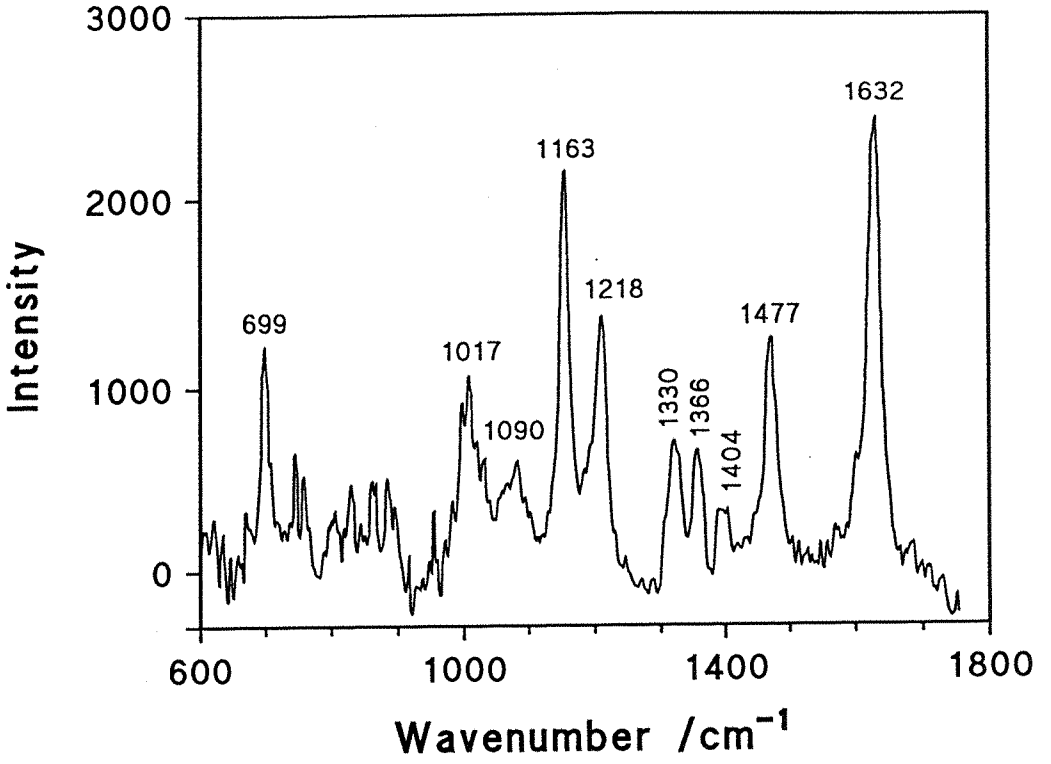


Figure 3d.6: Resonance Raman spectrum of 450μM 4-thiothymidine in deuterium oxide.

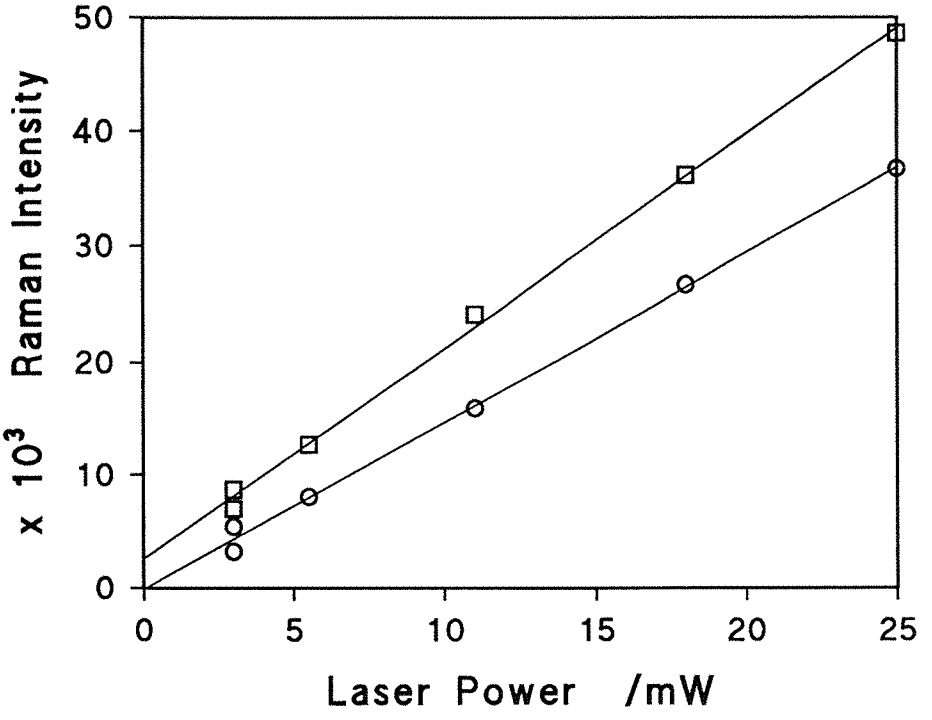


Figure 3d.7: Graph of the power dependence of the resonance Raman intensities of the 707cm^{-1} (○) and 1158cm^{-1} (□) peaks of 4-thiothymidine.

is shifted down by $15cm^{-1}$ in D_2O relative to the spectrum in H_2O. None of the other peaks are significantly affected by the change in solvent.

In order to ensure that the Raman spectrum of 4-thiothymidine in its *ground* state was being observed and not that of an excited state, the area of two of the Raman peaks ($707cm^{-1}$ and $1158cm^{-1}$) was investigated as a function of the laser power. If an excited state is being viewed then a two photon process is involved; one being absorbed by the 4-thiothymidine molecule to push it into the excited state, and the other photon being scattered by the excited molecule to produce the Raman spectrum. Such a situation would result in the Raman intensity being proportional to the *square* of the laser power. If however the ground state Raman spectrum is being observed then only one photon, which is scattered by the molecule, is involved. In this case the Raman intensity is *directly* proportional to the laser power. The results plotted for the two peaks in Figure 3d.7 clearly show that the ground state and not an excited state is being observed.

3d.4 RESONANCE RAMAN SPECTROSCOPY OF OLIGODEOXYNUCLEOTIDES

3d.4.1 Resonance Raman Spectroscopy of d(GACGA[^{4S}T]ATCGTC)

The resonance Raman spectrum of d(GACGA[^{4S}T]ATCGTC) (produced as described for 4-thiothymidine) is shown in Figure 3d.8. The Raman scattering of this and the other dodecamer were found not to decay noticeably on exposure to the laser. This is in contrast to the case for the 4-thiothymidine nucleoside where the Raman signal was found to be gradually decreased and the background fluorescence increased. Deoxygenation of the sample greatly reduced this problem seen with the deoxynucleoside but even in these samples the signal gradually altered, presumably due to oxygen leaking into the cell. With the oligodeoxynucleotides however, the background fluorescence was much less than for the deoxynucleoside and did not increase with time. Also, reverse phase HPLC of oligodeoxynucleotide samples that had been exposed to the laser showed no detectable decomposition. The 4-thiothymidine residue therefore seems to be protected from reaction in some way in the oligodeoxynucleotides. This is probably due to excited state 4-thiothymidine (either ^1T or ^3T in Scheme 3d.1) being able to lose its energy harmlessly through an alternative vibrational relaxation pathway. It may also be that the site of oxygen attack is shielded to some degree in the oligodeoxynucleotide and so limiting the rate of reaction. If this was the case then the vibrational relaxation pathway, which may have been slower than the oxygen reaction, would now be faster. This would return excited state 4-thiothymidine to the ground state before it had a chance to react either chemically or with oxygen to produce

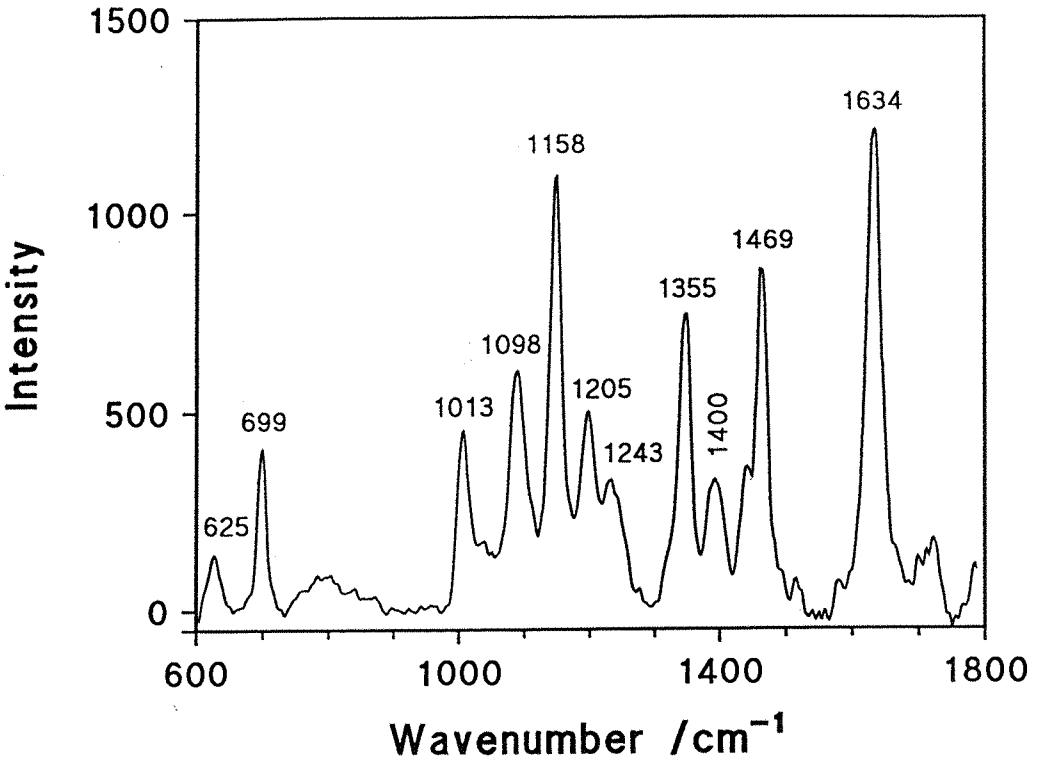


Figure 3d.8: Resonance Raman spectrum of 540μM double stranded d(GACGA[^{4S}T]ATCGTC) in buffer α0.

singlet oxygen.

Comparison of Figure 3d.8 with the spectrum of the nucleoside shows that although some of the features are the same in both cases, some significant differences are apparent. The peaks and their assignments are listed in Table 3d.2 together with those for 4-thiothymidine. The 1634cm^{-1} (ν_b^6 including C5=C6 stretch), 1158cm^{-1} (Br^6 coupled with C=S stretch), 1013cm^{-1} (Tr deformation) and 625cm^{-1} (δ_a^6 deformation) peaks are all essentially the same as in the nucleoside. The 1469cm^{-1} (ν_a^6 ring stretch) is 8cm^{-1} lower than that in the nucleoside. This may be significant but without knowing the exact bond vibrations involved in this mode it is difficult to interpret. The considerable increase in the intensity of the peak at 1098cm^{-1} can be attributed not to an increase in the ring deformation ($\nu_{\delta a}^6$) intensity but to the PO_2^- symmetrical stretch of the oligodeoxynucleotide phosphates (Thomas & Wang, 1988). This is not a resonant enhanced band but because there are eleven times as many phosphates as 4-thiothymidine residues, and because it is normally a relatively intense peak, it shows up in the oligodeoxynucleotide spectra. The other changes in the spectrum are analogous to those seen for 4-thiothymidine in D_2O. That is, the C=S peak is shifted by 8cm^{-1} to 699cm^{-1} and the 1352cm^{-1} peak is split. The effect of hydrogen bonding to the 4-thio function and the N3 proton in the oligodeoxynucleotide may well have a similar effect to the hydrogen bonding of the former to the D_2O solvent and the exchange of the N3 atom for a deuterium atom. The 1352cm^{-1} peak seen in the nucleoside spectrum appears to be split into two peaks in the oligodeoxynucleotides at 1355cm^{-1} and at 1400cm^{-1}. The former is assigned to the C5—Me

Table 3d.2: Resonance Raman peaks (cm^{-1}) and their vibrational assignments for 4-thiothymidine in water and the oligodeoxynucleotides d(GACGA[^{4S}T]ATCGTC), d(GACGATA[^{4S}T]CGTC) and d(AG[^{4S}T]TC) in buffer α0.

VIBRATIONAL MODE	4-Thiothymidine	d(GACGA[^{4S}T]ATCGTC). [d(GACGATA[^{4S}T]CGTC) in brackets]	d(AG[^{4S}T]TC)
ring stretch, ν_b^6 (C5=C6)	1631	1634 (1633)	1632
ring stretch, ν_a^6	1477	1469 (1470)	1474
ring stretch, $\nu_{\delta a}^6$ coupled with N1-C1' stretch	1352	1400 (1395)	—
C5-methyl stretch		1355 (1357)	1353
unknown	—	1243 (1247)	1254
Kekulé, Kk	1233	1205 (1211)	1230
ring breath, Br^6, coupled with C=S stretch	1158	1158 (1159)	1157
ring stretch, $\nu_{\delta b}^6$	1092	—	—
ring stretch, $\nu_{\delta b}^6$ plus PO_2^- stretch	—	1098 (1096)	1100
ring deformation, Tr	1018	1013 (1014)	1018
unknown	865	—	—
C=S stretch coupled with ring breath, Br^6	707	699 (703)	706
ring deformation, δ_a^6	628	625 (632)	—

stretch whilst the latter is assigned to the N1-C1' plus the $\nu_{\delta a}^6$ ring stretch. In the double helix of the oligodeoxynucleotides, the constraints imposed on the N1-C1' bond (such as *syn* orientation of the base) is probably the cause of the shifting of its Raman peak (by about $40cm^{-1}$). Another major difference between the nucleoside and the oligodeoxynucleotide spectra is the appearance of an extra peak at $1243cm^{-1}$. This peak was assigned as being due to the deoxyribose-phosphate backbone but its exact origins are not known.

3d.4.1.i Hydrogen Bonding of the 4C=S Group in the DNA Helix

The hydrogen bond of the 4-thio function in the d(GACGA[^{4S}T]ATCGTC) duplex is responsible for the $8cm^{-1}$ shift of its Raman band to $699cm^{-1}$ (see Figure 3d.8). This is supported by the fact that the $699cm^{-1}$ peak can be 'melted out' to a certain extent by heating the oligodeoxynucleotide (Figure 3d.9). This $8cm^{-1}$ shift shows that the 4-thio group is participating in maintaining the stability of the helix. The energy of the hydrogen bond formed between the 4-thio group on 4-thiothymidine and the 6-amino group on the base pair partner

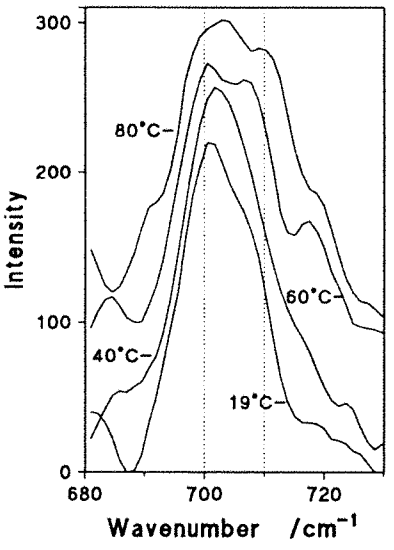


Figure 3d.9: Melting of the C=S peak of 4-thiothymidine in d(GACGA[^{4S}T]ATCGTC) [540μM double stranded in buffer η].

deoxyadenosine relative to the one formed between the 4-thio group and water can be calculated by two methods. The first relies on an estimate that there is a linear relationship between bond energy and wavelength shift as long as the wavenumber shift is small in relation to the overall bond energy (Badger & Bauer, 1937). Thus for an $8cm^{-1}$ shift and assuming a value of 600kJ/mol for the C=S bond (Gans, 1971), the change in ΔH is:-

$$\Delta\Delta H = (8/707) \times 600 = 6.7kJ/mol$$

Another method of calculation uses the empirical correlation between the change in the bond force constant, F, and ΔH (Gans, 1971). In this case the vibration is assumed to be simple harmonic and the change in force constant is calculated from the equation for simple harmonic motion [(frequency)^2=force constant/reduced mass]:-

$$\text{Ratio of force constants} = (707)^2/(699)^2 = 1.023$$

Therefore there is a 2.3% decrease in the force constant. The linear relationship between F and ΔH is:-

$$\Delta H = 0.61F$$

The decrease in ΔH is therefore, 0.61 x 2.3% = 1.4%

$$\therefore \Delta\Delta H = 1.4\% \times 600 = 8.4kJ/mol$$

Although the two methods do not necessarily give the exact value the fact that they are very similar is encouraging. Also they are of similar magnitude to the strength of Watson-Crick hydrogen bonds determined for DNA duplexes in water (3.4kJ/mol to 6.7kJ/mol; Freier *et al.*, 1986). The strength of sulphur hydrogen bonds *in vacuo* are about half the strength of those of oxygen (Weiner *et al.*, 1984). For example:-

$HOH\cdots OH_2$	-27kJ/mol
$HOH\cdots S(H)CH_3$	-13kJ/mol

However, in water we are looking at the exchange reaction of the group hydrogen bonded to water with the group being hydrogen bonded in the duplex:-

$$4C{=}S\cdots HOH + H_2O\cdots H_2N\text{-}6C \longrightarrow 4C{=}S\cdots H_2N\text{-}6C + H_2O\cdots HOH$$

We are therefore measuring the difference in the hydrogen bonding of the sulphur atom to water and of it hydrogen bonded in DNA. In tyrosyl-tRNA site directed mutagenesis experiments, sulphur was found to form effective hydrogen bonds in water (Wilkinson *et al.*, 1983) as appears be the case in the d(GACGA[^{4S}T]ATCGTC) oligodeoxynucleotide. The strength of the C=S hydrogen bond in d(GACGA[^{4S}T]ATCGTC) is further supported by the fact that its melting temperature, T_M, is virtually identical to that of d(GACGATATCGTC) (52°C and 53°C respectively; Connolly & Newman, 1989).

3d.4.2 Resonance Raman Spectroscopy of d(GACGATA[^{4S}T]CGTC)

The resonance Raman of d(GACGATA[^{4S}T]CGTC) was virtually identical to that of d(GACGA[^{4S}T]ATCGTC). The peak wavenumbers are given in Table 3d.2 and it can be seen that, with the exception of a few peaks, they are in the same positions as in the other dodecamer. The small variations however, may be significant. For instance the smaller 4cm^{-1} shift of the C=S peak down to 703cm^{-1} suggests a less strong hydrogen bond (calculated strength: 3.4kJ/mol, method 1 and 4.2kJ/mol, method 2) than in d(GACGA[^{4S}T]ATCGTC). The UV spectra of the two oligodeoxynucleotides shown in Figure 3d.1 also imply different environments for the C=S group by the differences in their λ_{max}'s (337nm and 349nm). The differences in the Raman peaks at 1211cm^{-1} and 1247cm^{-1} in d(GACGATA[^{4S}T]CGTC) compared with 1205cm^{-1} and 1243cm^{-1} in d(GACGA[^{4S}T]ATCGTC) probably also reflect different local environments for the 4-thiothymidine residue in the two oligodeoxynucleotides.

3d.4.3 Resonance Raman Spectroscopy of d(AG[^{4S}T]TC)

The resonance Raman spectrum of d(AG[^{4S}T]TC) is shown in Figure 3d.10 with the peak positions and assignments in Table 3d.2. This oligodeoxynucleotide was chosen so that it could provide a model for single stranded DNA and give support to some of the interpretation of the Raman spectra of the double stranded oligodeoxynucleotides. Some of the assignments made to the double stranded dodecamers are indeed supported by the spectrum of d(AG[^{4S}T]TC). Firstly the C=S peak is not significantly shifted in d(AG[^{4S}T]TC) and so the shifts seen in the dodecamers do seem to be caused by the hydrogen bonding of the sulphur in double stranded DNA. Also no peak at ~1400cm^{-1} is visible for d(AG[^{4S}T]TC). This peak was assigned to the N1-C1' plus $\nu_{\delta a}{}^{6}$ stretch in the double helix which was said to be separated from the C5-Me stretch by the constraints imposed on it in the duplex. In d(AG[^{4S}T]TC) it appears that the single stranded DNA chain does not cause the 1353cm^{-1} peak to split probably because there is more freedom of conformation for the 4-thiothymidine base than in the duplex. As for the dodecamers however, the peak seen at 1233cm^{-1} in the nucleoside spectrum is split into two

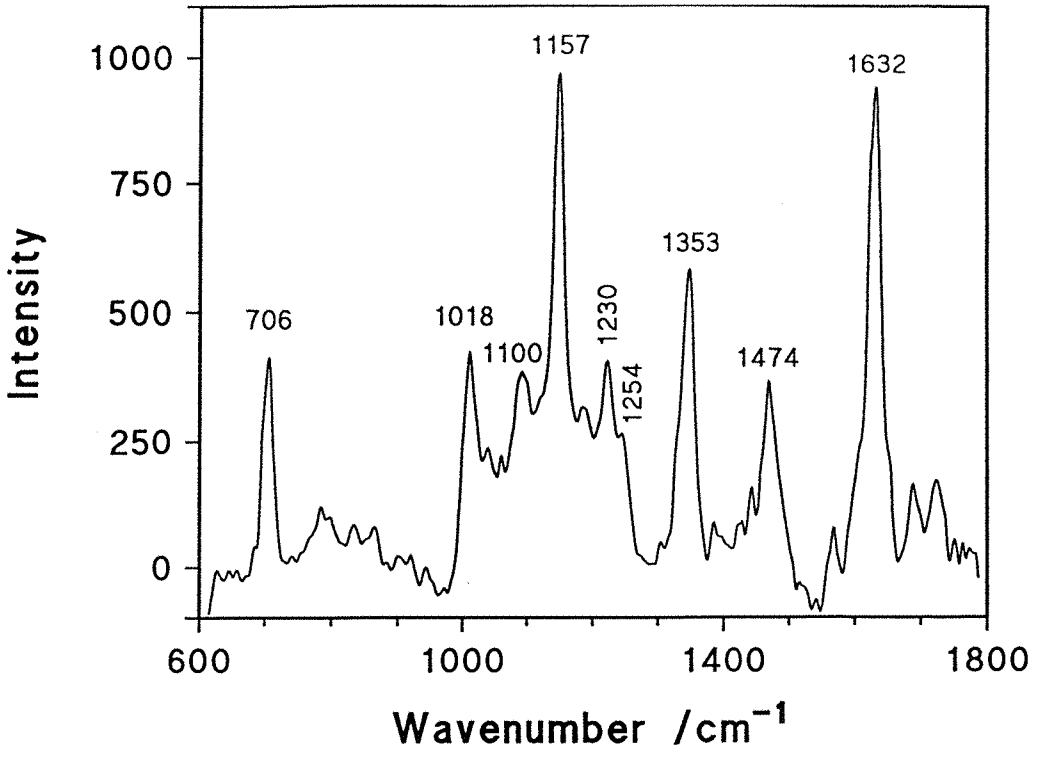


Figure 3d.10: Resonance Raman spectrum of 540μM single stranded d(AG[^{4S}T]TC).

peaks in d(AG[^{4S}T]TC) (1254cm^{-1} and 1230cm^{-1}). The reason for this splitting has not been determined but the fact that it occurs in both single and double stranded DNA limits the possible causes. For example it cannot be due to Watson-Crick hydrogen bonding. Again as for the dodecamers, an increase in the intensity of the peak at 1100cm^{-1} is seen but to a much lesser extent. This is assigned to the PO_2^- stretch and the lower intensity for d(AG[^{4S}T]TC) than that seen in the dodecamers is due to the fewer phosphate groups in the pentamer (four as opposed to eleven). All in all, the Raman spectrum of d(AG[^{4S}T]TC) shows features that are intermediate between those of the free nucleoside and those of the double stranded dodecamers. Unlike the dodecamers however, the C=S group in d(AG[^{4S}T]TC) does not seem to be protected from photochemical reaction since its Raman intensity decreases on laser exposure. Also reverse phase HPLC revealed that several products were formed on laser exposure of d(AG[^{4S}T]TC).

3d.5 RESONANCE RAMAN SPECTROSCOPY OF d(GACGA[^{4S}T]ATCGTC) BOUND TO THE ECO RV ENDONUCLEASE

The resonance Raman spectrum of 100μM d(GACGA[^{4S}T]ATCGTC) in the presence of 110μM Eco RV endonuclease in buffer η (no magnesium) is shown in Figure 3d.11a. Higher concentrations of the chromophore proved impractical because of the problem of precipitation of higher concentrations of the endonuclease on addition of the oligodeoxynucleotide. Despite the high background fluorescence, Raman peaks due to the 4-thiothymidine residue are visible (labelled in the figure). Other peaks are also visible and these are attributed to the Raman spectrum of glycerol (not resonance enhanced). This was present in the enzyme storage buffer and was added with the enzyme resulting in a final concentration of 8% (v/v). These peaks can also be seen in the Raman spectrum of a 110μM solution of endonuclease in the absence of DNA shown in Figure 3d.11b. The glycerol peaks and a significant amount of the background fluorescence was removed from the spectrum of endonuclease bound d(GACGA[^{4S}T]ATCGTC) by subtraction of an endonuclease curve (Figure 3d.11b) which had been normalised to the 848cm^{-1} and 921cm^{-1} glycerol peaks. This gave the spectrum shown in Figure 3d.11c which then had the remaining fluorescence removed by subtraction of a polynomial fitted curve. After the peak positions were determined, this spectrum was then smoothed to give that shown in Figure 3d.11d.

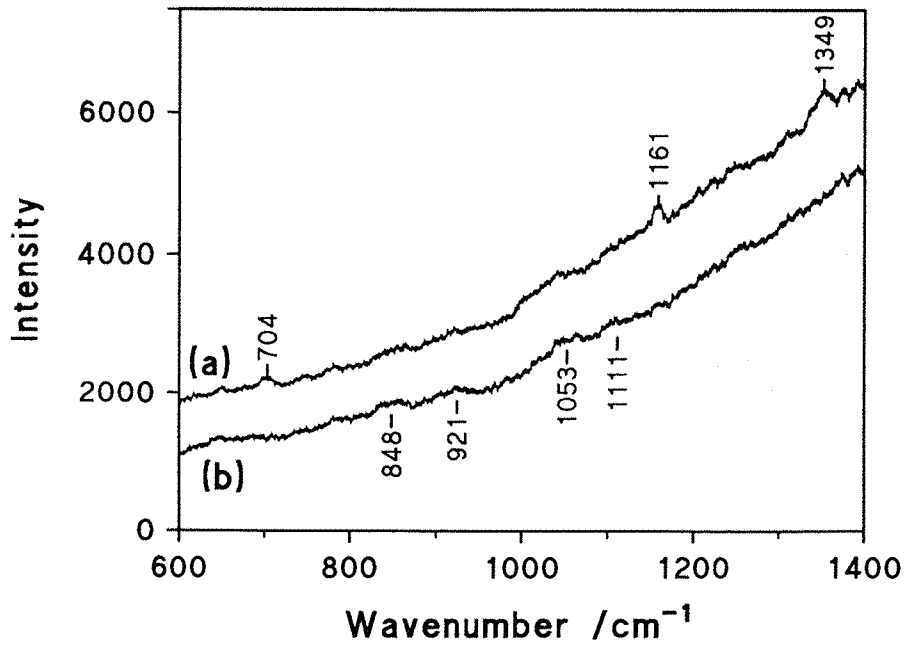


Figure 3d.11: Resonance Raman spectra of **(a)** d(GACGA[^{4S}T]ATCGTC) bound to the endonuclease; some of the Raman peaks are indicated, and **(b)** the endonuclease alone; glycerol peaks are indicated.

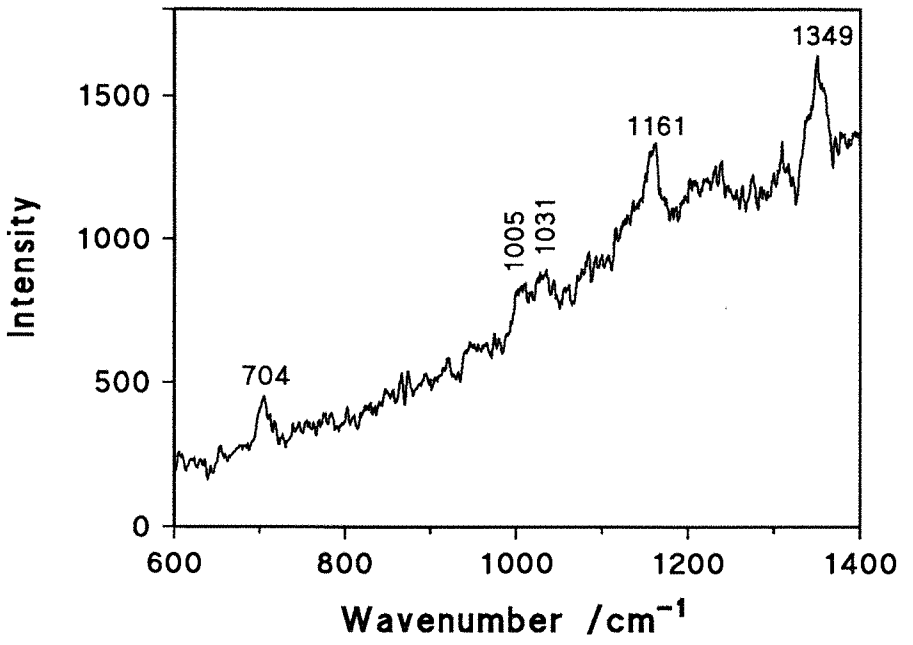


Figure 3d.11c: Resonance Raman spectrum of d(GACGA[^{4S}T]ATCGTC) bound to the Eco RV endonuclease with the glycerol/endonuclease background spectrum subtracted.

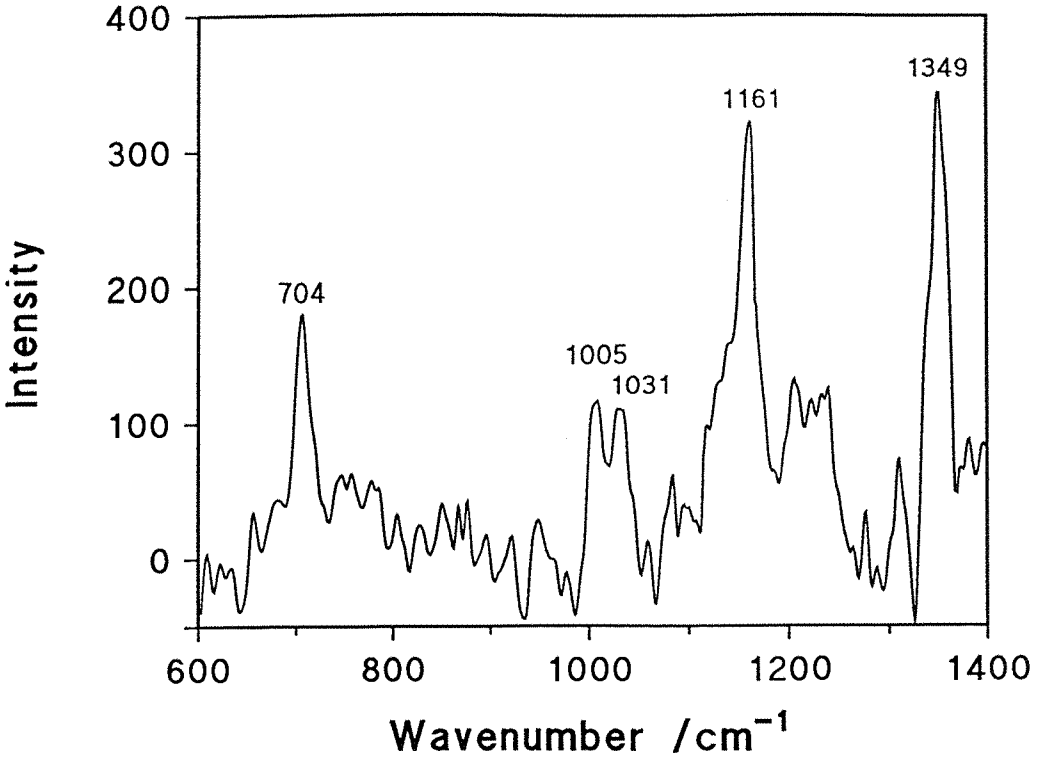


Figure 3d.11d: Final background subtracted and smoothed resonance Raman spectrum of endonuclease bound d(GACGA[^{4S}T]ATCGTC) [in the absence of magnesium].

Some definite changes in the spectrum of the oligodeoxynucleotide can be seen when it is in the presence of the endonuclease as opposed to when it is free in solution. From the association constant for DNA and the endonuclease (~10^{-6}M; Taylor & Halford, 1991) the vast majority of the oligodeoxynucleotide is expected to be enzyme bound at the concentrations used here. It is therefore the interaction of the endonuclease with the oligodeoxynucleotide that has caused the peak shifts seen in Figure 3d.11d. The most striking difference is the 5 cm^{-1} shift of the C=S peak up to 704cm^{-1} which is accompanied by a 3cm^{-1} shift in the coupled ring breathing mode up to 1161cm^{-1}. This can be interpreted as disruption of the double helix hydrogen bonding to the 4-thio function. Indeed the crystal structure of a cognate oligodeoxynucleotide bound to the endonuclease shows that the two central A.T base pairs are unstacked, distorted and unable to form either their Watson-Crick hydrogen bonds or hydrogen bonds to the protein (Winkler, 1991). The C5-Me stretching vibration has also been shifted on endonuclease binding. This peak is known to be sensitive to its local environment (Benevides *et al.*, 1991) and the 6cm^{-1} shift down to 1349cm^{-1} observed here indicates a definite change in the 5-methyl environment. This could be caused by van der Waals contact between the methyl and the protein as implicated by base analogue studies (Newman *et al.*, 1990b) or by some less direct effect such as the distortion of the DNA. Interpretation of the significance of the 1031cm^{-1} and 1005cm^{-1} peaks is difficult because of the close proximity to a glycerol peak (1053cm^{-1}). A relatively small error in the subtraction of this glycerol peak could result in considerable alteration in the appearance of these two peaks and so interpretation of them has not been attempted.

Addition of magnesium to the endonuclease and oligodeoxynucleotide solutions had no noticeable affect on the Raman spectrum indicating that no major structural changes occur in the vicinity of the 4-thiothymidine residue on binding of magnesium. Analysis of the enzyme bound oligodeoxynucleotides by HPLC after the Raman spectra were recorded showed that no detectable cleavage of the oligodeoxynucleotide occurred in the absence of magnesium. In the presence of magnesium, less than 5% cleavage had occurred during the recording of the resonance Raman spectra.

In summary, although the quality of the enzyme bound oligodeoxynucleotide spectra are not very good, definite interpretable changes are observed. To add support to these results and to ensure that a specific rather than a non-specific complex is being observed, studies of a 4-thiothymidine containing oligodeoxynucleotide that lacks an Eco RV site need to be performed.

CHAPTER 4

CONCLUSION

4.1 CONTINUOUS ULTRAVIOLET ABSORBANCE ASSAY OF ENDONUCLEASE ACTIVITY

During the course of this project a new method of continuously monitoring the Eco RV endonuclease reaction was developed. The assay is based on the increase in UV absorbance that occurs when cleavage of a double stranded DNA substrate results in single stranded products. For this reason, the length of the oligodeoxynucleotide substrate and the position of the cleavage site should be such that the products formed are single stranded at the assay temperature. It should be relatively straight forward to design substrates for other endonucleases that fulfil these criteria and so this assay should find general applicability to other endonuclease reactions. Previously used methods for following endonuclease reactions are based on the removal of samples from a reaction mixture, quenching them, and then separating the products from the substrate. Separation of the products and substrate can be achieved by reverse phase HPLC, homochromatography on DEAE-cellulose TLC plates or polyacrylamide gel electrophoresis. The UV assay has several advantages over these other methods as summarised below:-

(i) It is much quicker than the other methods and is relatively simple to perform.

(ii) Reactions can be monitored in real time and so results can be interpreted immediately.

(iii) Unlike the other methods, the UV assay is continuous and thus eliminates potential problems associated with the removal and quenching of samples. Because of this, it gives more accurate measurements of reaction rate.

(iv) Unlike the gel electrophoresis and homochromatography techniques, it does not require the use of ^{32}P labelled substrates. The hazards and inconveniences with the preparation and use of radiolabelled oligodeoxynucleotides are therefore avoided.

A possible disadvantage of the UV assay is the limited range of substrate concentrations that can be used. Even the use of different path length cells, which give a potential range of 0.08μM to 37μM for d(GACGATATCGTC), may not cover the required concentrations in some applications. The upper limit is dependent upon the loss of the spectrometer's linear response to the concentration of chromophore, at high absorbance values. The lower limit is really dependent upon the sensitivity of the technique. With low concentrations of substrate the cleavage reaction may produce such a small change in the absorbance that it cannot be reliably detected. In this study however, the UV assay was found to be a very convenient and accurate method of determining the rates of Eco RV endonuclease cleavage of oligodeoxynucleotides. It was possible to use it as a way of assessing the Eco RV endonuclease activity of unknown samples. It also enabled the successful determination of the kinetic parameters, k_{cat} and K_M, of a number of oligodeoxynucleotides and base analogue containing oligodeoxynucleotides.

4.2 ENZYMOLOGY OF THE ECO RV ENDONUCLEASE WITH BASE ANALOGUE CONTAINING OLIGODEOXYNUCLEOTIDES

In this study, the reaction of the Eco RV endonuclease with oligodeoxynucleotides containing deoxycytidine and deoxyguanosine analogue substitutes in the Eco RV recognition site have been examined. Firstly, the oligodeoxynucleotides were tested to see whether or not they were Eco RV endonuclease substrates. Secondly, those that were substrates had their k_{cat} and K_M values determined. From the specificity constants (k_{cat}/K_M), the apparent binding energies (ΔG_{app}) could be calculated (Fersht, 1985). The results obtained here, and in the complementary study on deoxyadenosine and thymidine analogues (Newman *et al.*, 1990b) are given in Table 4.1. These results fall into three categories:-

(i) Analogues that are cleaved at the same rate as the control. These oligodeoxynucleotides involve the substitution of the 2C=O of the first thymidine by 2C=S, the substitution of the 7N of the second deoxyadenosine by 7CH and the substitution of the 4-NH_2 of the deoxycytidine by a hydrogen. The analogue involving the addition of a methyl group at the 5-position of the deoxycytidine also falls into this category. These results can be taken to mean that the Eco RV endonuclease does not contact these sites and that it is not in close enough contact to the deoxycytidine 5-position to disallow a methyl group there.

(ii) Analogues that are either not cleaved at all or produce only a trace of cleavage. The replacement of the 6C=O of the deoxyguanosine by 6C=S or the replacement of the 6-NH_2 of the first deoxyadenosine both result in extremely slow cleavage by the endonuclease (<0.1% of the control). These oligodeoxynucleotides were too slowly cleaved to enable the determination of their k_{cat} and K_M parameters. Oligodeoxynucleotides that were not cleaved at all are those that involved the replacement of the 6C=O of the deoxyguanosine by a hydrogen, the replacement of the 7N of the first deoxyadenosine by 7CH, replacement of the 4C=O of the first thymidine by a hydrogen, and the replacement of the 4C=O of the second thymidine by either 4C=S or a hydrogen. The potential contacts altered in these oligodeoxynucleotides are obviously very important to the Eco RV endonuclease reaction.

(iii) Analogues that cause a reduction in specificity resulting in a ΔG_{app} of between -4.3kJ/mol and -9.4kJ/mol. In all these cases the reduction in specificity arises from a reduction in k_{cat}. K_M values are virtually the same for all of the oligodeoxynucleotides. If ΔG_{app} approximates ΔG_{bind}, then all but two of the values obtained for these oligodeoxynucleotides are of the expected magnitude for the deletion of hydrogen bonds (2.1kJ/mol to 7.5kJ/mol) and would therefore imply that for these oligodeoxynucleotides, groups that bind to the endonuclease during the reaction are being deleted. The two ΔG_{app} values that are above this range are from oligodeoxynucleotides that involve the deletion of a 5-methyl group from a thymidine and so concern the deletion of a potential van der Waals contact, not a hydrogen bonding group.

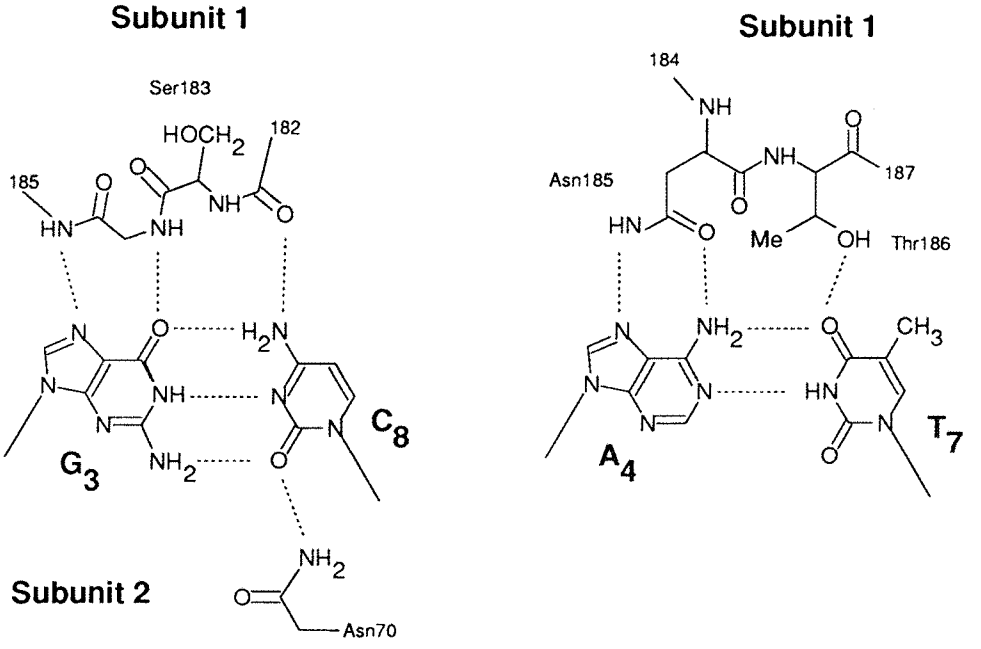
Recently the specific contacts between the Eco RV endonuclease and d(GGGATATCCC) in a crystal complex, have been reported (Winkler, 1991; see Figure 4.1). These contacts are given in Table 4.2 and comparison with the kinetic results show some interesting features.

Table 4.1: Kinetic data for the reaction of the Eco RV endonuclease with oligodeoxynucleotides containing base analogues. [a]Eco RV site contained within the sequence d(GACGATATCGTC). [b]From this study. [c]From Newman *et al.* (1990b). [d]Given as initial rates relative to the control sequence. [e]Given as kJ per mol of modified base. n/d=not determined. n/c=not cleaved. [m]minor groove contact (other groups are in the major groove).

BASE	ECO RV SITE[a]	k_{cat} /min^{-1}	K_M /μM	k_{cat}/K_M /$s^{-1}M^{-1}$	ΔG_{app}^{e} /$kJmol^{-1}$
–	GACGAT	33[b] (6.9)[c]	2.9[b] (3.8)[c]	186000[b] (30200)[c]	0
G_1	[^{6S}G]ACATC	(0.05%)[b,d]	n/d	n/d	n/d
	[^{6H}G]ACATC	n/c	–	–	–
	[^{7C}G]ACATC	n/c	–	–	–
	[I]ACATC[m]	0.63[b]	1.76[b]	5950[b]	-4.3
	[^{3C}G]ACATC[m]	0.28[b]	3.51[b]	1340[b]	-6.1
A_2	G[P]TATC	(?)[c,d]	n/d	n/d	n/d
	G[^{7C}A]TATC	n/c	–	–	–
	G[^{3C}A]TATC[m]	0.057[c]	4.0[c]	238[c]	-6.0
T_3	GA[U]ATC	0.013[c]	3.8[c]	55.8[c]	-7.8
	GA[^{4S}T]ATC	0.016[c]	0.5[c]	533[c]	-5.0
	GA[^{4H}T]ATC	n/c	–	–	–
	GA[^{2S}T]ATC[m]	1.9[c]	0.95[c]	32900[c]	+0.1
A_4	GAT[P]TC	0.067[c]	8.0[c]	139[c]	-6.7
	GAT[^{7C}A]TC	1.9[c]	0.8[c]	39600[c]	+0.3
	GAT[^{3C}A]TC[m]	0.1[c]	1.4[c]	1190[c]	-4.8
T_5	GATA[U]C	0.006[c]	6.5[c]	15.4[c]	-9.4
	GATA[^{4S}T]C	n/c	–	–	–
	GATA[^{4H}T]C	n/c	–	–	–
	GATA[^{2S}T]C[m]	0.1[c]	4.0[c]	417[c]	-5.3
C_6	GATAT[^{5Me}C]	21[b]	0.97[b]	362000[b]	+0.8
	GATAT[^{4H}C]	38[b]	2.0[b]	317000[b]	+0.7

Firstly, the deletion of any of the protein/DNA contacts seen in the crystal, except for the contact to the 4-amino group of the deoxycytidine, either totally abolish cleavage or reduce it to an extremely slow rate. Despite the crystal structure implicating a hydrogen bond to the 4-amino group of the deoxycytidine, replacing the amino group with a hydrogen results in an oligodeoxynucleotide that is at least as good a substrate as the control dodecamer. Secondly, substrates that give ΔG_{app} values of the order expected for the deletion of a hydrogen bond, involve deletion of groups not contacted by the protein in the crystal structure.

There are two possible explanations for why some of the analogue oligodeoxynucleotides are not cleaved (or cleaved extremely slowly). The first is that the introduction of the base analogue has not only deleted the hydrogen bond contact but has also caused some kind of change in the structure of the DNA which renders it uncleavable by the endonuclease. The CD and T_M data for all of these oligodeoxynucleotides show them to be double stranded at the assay temperature of 25°C and of an overall B-form DNA structure. It is still possible however, that significant local deviations in DNA structure may occur in some of the oligodeoxynucleotides. The second possible explanation of the lack of cleavage is that the hydrogen bond that the protein makes to the deleted point is part of a group of hydrogen bonds that work cooperatively. The DNA is known to be very distorted in the enzyme complex and it could be that the formation of one hydrogen bond could, by slight distortion of the complex, bring other hydrogen bond donors and acceptors into contact enabling further hydrogen bonds



G_1-G_2-G_3-A_4-T_5-A_6-T_7-C_8-C_9-C_{10}
C_{10}-C_9-C_8-T_7-A_6-T_5-A_4-G_3-G_2-G_1

Figure 4.1: Contacts made by the Eco RV endonuclease to the G.C and the outer A.T base pairs in the crystal complex (Winkler, 1991).

Table 4.2: Kinetic data for the reaction of the Eco RV endonuclease with oligodeoxynucleotides containing base analogues compared with the contacts implicated in the Eco RV crystal complex (Winkler, 1991). [a]Eco RV site contained within the sequence d(GACGATATCGTC). [b]Given as kJ per mol of modified base. [c]This analogue involves the addition of an *extra* group, 5-methyl. n/d=not determined but cleaved at a greatly reduced rate. n/c=not cleaved. [m]Minor groove contact (other groups are in the major groove).

BASE	ECO RV SITE[a]	ΔG_{app}[b] /kJmol^{-1}	CONTACT GROUP	CONTACT IN CRYSTAL?	CONTACT DURING REACTION?
–	GACGAT	0	–	–	–
G_1	[^{6S}G]ACATC	n/d	6=O	Asn185 NH	yes
	[^{6H}G]ACATC	n/c			
	[^{7C}G]ACATC	n/c	7N	Gly184 NH	yes
	[I]ACATC	-4.3	2-NH_2[m]	no	no
	[^{3C}G]ACATC	-6.1	3N[m]	no	no
A_2	G[P]TATC	n/d	6-NH_2	Asn185 Oδ1	yes
	GAT[^{7C}A]TC	n/c	7N	Gly184 NH	yes
	G[^{3C}A]TATC	-6.0	3N[m]	no	no
T_3	GA[U]ATC	-7.8	5-Me	no	no
	GA[^{4S}T]ATC	-5.0	4=O	no	no
	GA[^{4H}T]ATC	n/c			
	GA[^{2S}T]ATC	+0.1	2=O[m]	no	no
A_4	GAT[P]TC	-6.7	6-NH_2	no	no
	GAT[^{7C}A]TC	+0.3	7N	no	no
	GAT[^{3C}A]TC	-4.8	3N[m]	no	no
T_5	GATA[U]C	-9.4	5-Me	Asn188	yes
	GATA[^{4S}T]C	n/c	4=O	Thr186 Oγ1	yes
	GATA[^{4H}T]C	n/c			
	GATA[^{2S}T]C	-5.3	2=O[m]	no	no
C_6	GATAT[^{5Me}C]	+0.8	5-Me[c]	–	no
	GATAT[^{4H}C]	+0.7	4-NH_2	Gly182 =O	no

to be formed. Deletion of such bonds would therefore prevent formation of the other hydrogen bonds thus greatly lowering the ability of the enzyme to form the cognate complex.

The importance of the contacts to the 6-NH_2 and 7N groups of the first deoxyadenosine has also been demonstrated by site directed mutagenesis (Thielking *et al.*, 1991a). Mutation of Asn185 to alanine or glutamine abolished activity and mutating it to asparagine produced a mutant of extremely low activity. The contact between Thr186 and the 4=O of the second thymidine was also shown to be important. Mutating Thr186 to either a serine or an asparagine produced mutants that were totally inactive. These results suggest a degree of cooperativity in the hydrogen bonds of the enzyme/substrate complex. Unfortunately, since the other specific contacts in the crystal structure are from main chain atoms they cannot be mutated to investigate the importance of these contacts.

There are a number of possible causes for the reduction in specificity seen for some of the oligodeoxynucleotides that contain modified bases that do not form contacts to the protein as determined by the crystal structure. It could be that indirect, water mediated contacts are made to the groups on the DNA deleted in the analogue oligodeoxynucleotides. Another possibility is that protein contacts to these points are not made in the crystal complex but are realised in later stages of the endonuclease reaction. The crystal structure was determined in the absence of the catalytic cofactor magnesium and conformational changes may occur on binding of the metal ion. The fact that these crystals crack when magnesium is added do suggest that magnesium binding causes a conformational change in the complex. The most likely explanation, however, is that the presence of the base analogues perturb the structure of the DNA in a way that lowers the rate of reaction with the endonuclease. With the exception of the 2-thiothymidine containing dodecamers which had CD spectra more typical of A-DNA than B-DNA (Connolly & Newman, 1989), none of the oligodeoxynucleotides exhibited CD or melting temperature data suggestive of any gross deviation from B-DNA. Smaller, local deviations in the DNA structure would probably not be detected by CD spectroscopy and so the possibility exists that the base analogues cause local perturbations in the DNA that reduce the ability of the endonuclease to cleave the oligodeoxynucleotide. The greatly distorted DNA structure in the enzyme cognate oligodeoxynucleotide structure suggests another possible cause of reduced specificity for some of the oligodeoxynucleotides. This high energy distorted structure appears to be an integral part of the Eco RV endonuclease reaction and so the ability of a stretch of DNA to adopt this structure would have a significant affect in the reaction. The presence of the modified bases in some of the analogue oligodeoxynucleotides may reduce the ability of the DNA to form the bound DNA structure characteristic of the true cognate sequence. This could be brought about by a reduction in the flexibility of the analogue oligodeoxynucleotide or by a decrease in the stabilising forces, such as stacking interactions, that may play a part in holding the DNA in the distorted structure seen in the cognate complex. A better understanding of the influence that these base analogues have on DNA structure is really needed for a detailed interpretation of their affect on the cleavage reaction.

Techniques such as NMR, X-ray crystallography and Raman spectroscopy could be used to study the analogue oligodeoxynucleotide structures in more detail.

With the benefit of both the kinetic results and the crystal structure, it can be concluded that the contacts indicated in the last column of Table 4.2 are involved in the Eco RV endonuclease reaction. The question now arises as to whether these contacts are sufficient to enable the endonuclease to distinguish d(GATATC) from any other sequence. The answer is yes for the first two base pairs in the sequence. This is illustrated in Figures 4.2 and 4.3. In the first base pair of the recognition sequence, G.C, there are two hydrogen bonds between the protein and the bases. One occurs between the main chain NH of Asn185 and the 7N of the deoxyguanosine, and the other contact is between the main chain NH of Gly184 and the 6C=O of the deoxyguanosine. No other base pairs can form both hydrogen bonds to the protein. In addition, in some cases incorrect base pairs result in the apposition of one of the protein hydrogen bond donors and a hydrogen bond donor of one of the bases. This type of apposition has been shown to be very unfavourable in the Eco RI endonuclease reaction and is probably a very important part of the discrimination mechanism of the Eco RI endonuclease (Lesser *et al.*, 1990). As well as the contacts shown here, the crystal structure also shows a contact in the

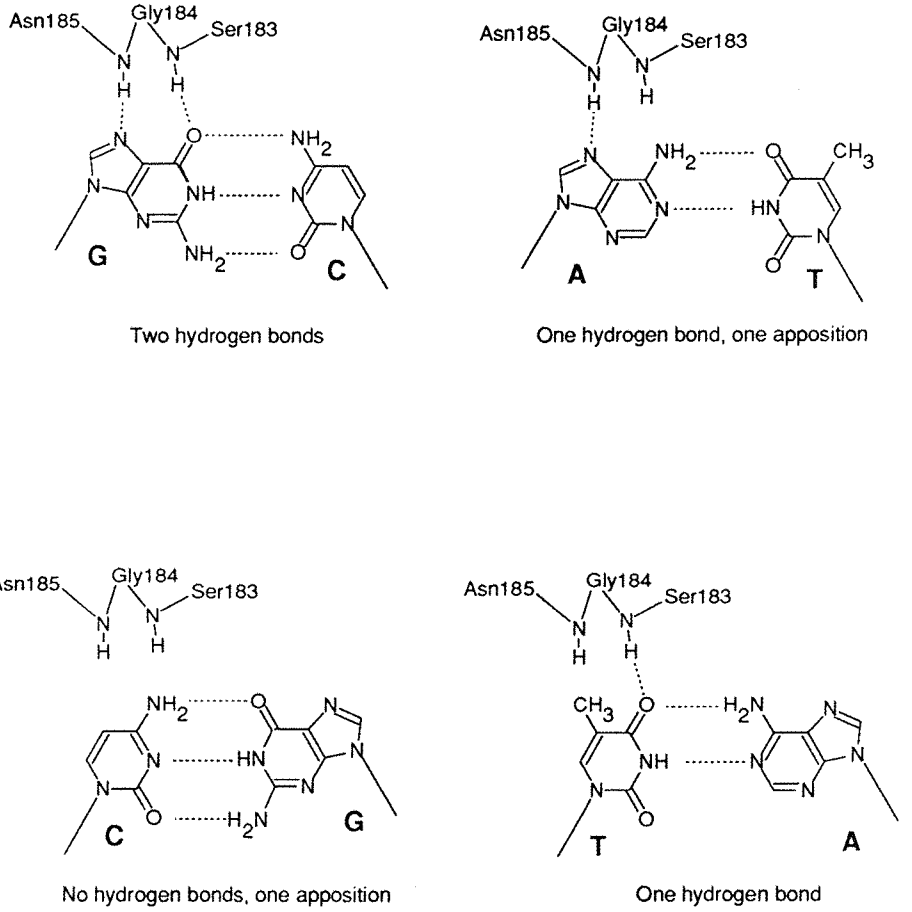


Figure 4.2: The hydrogen bonds made between the Eco RV endonuclease and the first base pair in the Eco RV site, G.C, on formation of the specific complex. The consequences of replacing G.C with any other base pair are also shown.

minor groove between the side chain nitrogen of Asn70 and the 2=O group of the deoxycytidine (Winkler, 1991). Unfortunately, no base analogue was used here to investigate the importance of this possible contact. If it does exist, then it may provide extra discrimination between different base pairs, although all four bases would present a hydrogen bond acceptor in roughly the same position in the minor groove (2=O of thymidine and 3N of deoxyadenosine and deoxyguanosine). The contacts from the protein to the second base pair, A.T, involve two bidentate hydrogen bonds between the side chain of Asn185 and the 7N and 2-NH_2 of the deoxyadenosine, a hydrogen bond between the side chain of Thr186 and the 4=O of the thymidine, and a van der Waals contact between the side chain of Asn188 and the 5-methyl of this thymidine. This arrangement of contacts discriminates well between the cognate A.T base pair and other base pairs as shown in Figure 4.3. Other base pairs would result in the loss of one or more hydrogen bonds, the van der Waals contact and in some cases the apposition of H-bond donors or acceptors. At first glance the T.A base pair would appear to be a particularly bad fit, forming no hydrogen bonds or van der Waals contact, and having the apposition of both a pair of hydrogen bond donors and a pair of hydrogen bond acceptors. The sequence d(GTTATC) is however the best Eco RV* site for the endonuclease (Taylor & Halford, 1989). It may be that the protein recognition motif is able to move relative to the DNA base pair enabling it to form three hydrogen bonds to the T.A base pair as shown in Figure 4.3. This arrangement, although probably introducing more strain into the complex for the cognate

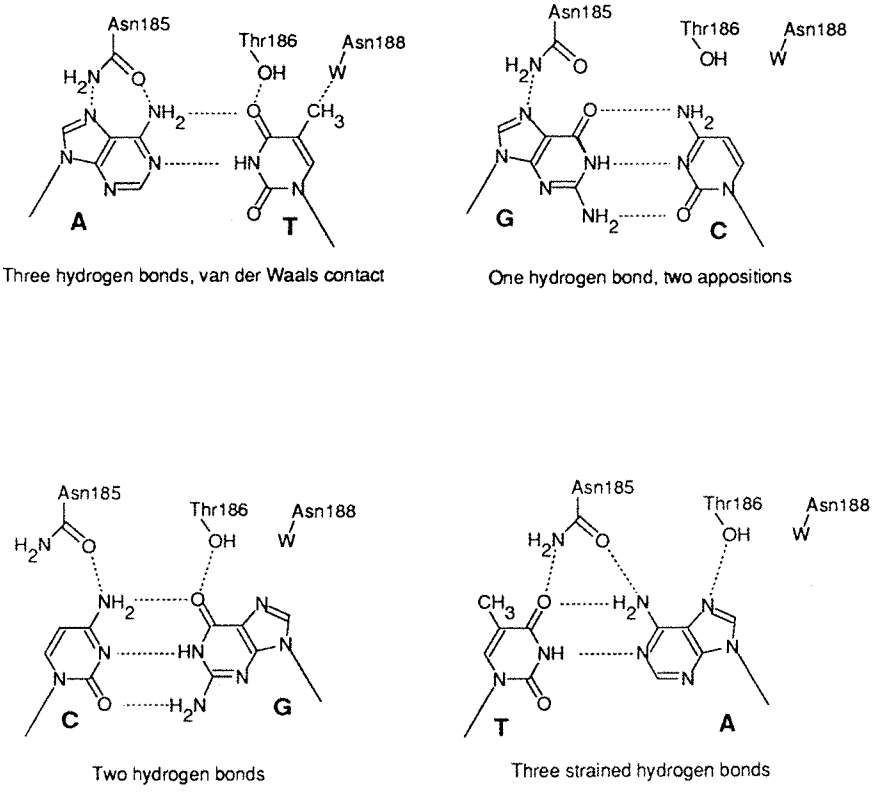
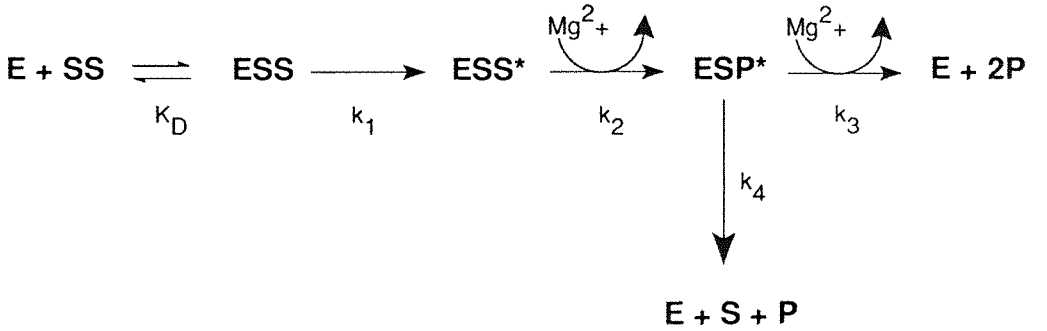


Figure 4.3: The three hydrogen bonds and the van der Waals contact (W) made between the endonuclease and the second base pair in the Eco RV site, A.T, on formation of the specific complex. The situation for other base pairs is also shown.

sequence, may be enough to explain the Eco RV* activity of d(GTTATC).

No contacts to the central two base pairs are seen in the crystal structure. How then does the endonuclease differentiate between these two base pairs and any other possible base pair arrangement? One possibility is that during the reaction the complex undergoes a conformational change, perhaps on magnesium binding, so that the endonuclease makes contact to these bases. Another explanation is based on the observation that the central base pair step (T.A/A.T) in the cognate crystal structure is essentially unstacked [see Figure 1.28, section 1.61]. This base pair step has been calculated to have the lowest stacking energy of any of the possible base pair steps (Saenger, 1984) and it could be that the extra energy needed to unstack any other base pair step provides a level of discrimination between the T.A/A.T step and any other step. Some of the modified bases used in the central two base pairs by Newman *et al.* (1990b) may result in different stacking energies for this step and this could explain the reduced specificities of these oligodeoxynucleotides. In the complex of the MetJ repressor with an oligodeoxynucleotide, no contacts are made to the central T.A and A.T base pairs even though they have been shown to be important for the binding of the protein (Phillips, S; personal communication). The specificity in this case was attributed to the low stacking energy of the T.A/A.T step allowing the overtwisting of the DNA seen at this position in the complex. In a recent study, however, experimentally derived stacking energies suggest that differences in the stacking energies of the various base pair steps are much less than previously calculated (Delcourt & Blake, 1991) and may not be enough to discriminate between different steps.

The mechanism proposed in Scheme 4.1 explains the evidence available to date. It involves the initial equilibrium between free enzyme (E) plus free substrate (S) and an enzyme/substrate complex (ES). This complex probably resembles that seen in the non-cognate crystal structure of Winkler (1991) and is basically non-specific. The fact that it is non-specific is supported by the results of Taylor and Halford (1991) who show that the Eco RV endonuclease binds all DNA sites with equal affinity. The next step is the formation of the specific complex (ES*). The distorted DNA seen in the cognate complex with the endonuclease brings the scissile phosphate of the DNA into the catalytic site of the enzyme allowing magnesium binding and catalysis to occur. The formation of this high energy complex is probably the rate determining step with the subsequent magnesium binding and phosphate hydrolysis reactions (k_2 and k_3) occurring at a much faster rate. With normal, cognate sites, magnesium binding and cleavage of the second strand (k_3) is much faster than the rate of release of DNA cleaved only in one strand (k_4). Essentially, cleavage of the two strands is simultaneous and so no single stranded cleavage is observed. With Eco RV* sites, the affinity for magnesium is much less and so k_3 is reduced and becomes comparable to k_4 resulting in a significant amount of single stranded cleavage being detected for Eco RV* reactions (Taylor & Halford, 1989). In contrast, it is not thought that the base analogues used in the series of dodecamers shown in Table 4.2 affect magnesium binding and so no single stranded cleavage is expected for these oligodeoxynucleotides.



Scheme 4.1: Proposed scheme for the Eco RV endonuclease reaction with DNA. E = endonuclease; SS = double stranded DNA; ESS = non-specific complex; ESS* = specific complex; SP = DNA cleaved in one strand only; P = cleaved DNA product.

The results for the base analogue containing dodecamers show that significant differences from the native sequence occur in the k_{cat} values but that the K_M values are much the same. If k_1 is by far the slowest step, then $K_M \approx K_D$ (and $k_{cat} \approx k_1$; Fersht, 1985). This is supported by the fact that the K_M values obtained for the dodecamers are of the same order of magnitude as the dissociation equilibrium constant determined for the binding of the endonuclease to DNA (Taylor & Halford, 1991). Larger, plasmid substrates have lower K_M values due to the facilitated diffusion mechanism. Thus K_M would represent the ability of the dodecamer substrates to form the non-specific type complex, ES. This would not be expected to be altered much by the presence of the modified bases since no direct contacts are made to the DNA bases and so this fits with the kinetic data. On the other hand, formation of the ES* complex requires deformation of the DNA and the formation of specific hydrogen bonds to some of the DNA bases to stabilise the structure. The rate of formation of ES* (k_1) would therefore be expected to be altered by the presence of modified bases that either eliminate one of these hydrogen bonds or reduce the ability of the DNA to distort into the specific complex structure.

REFERENCES

Aggarwal, A.K.; Rodgers, D.W.; Ptashne, M. & Harrison, S.C. (1988) *Science* 242 899-907

Aiken, C.R.; McLaughlin, L.W. & Gumport, R.I. (1991) *J. Biol. Chem.* 266 19070-19078

Alves, J.; Pingoud, A.; Haupt, W.; Langowski, J.; Peters, F.; Maass, G. & Wolf, C. (1984) *Eur. J. Biochem.* 140 83-92

Alves, J.; Rüter, T.; Geiger, R.; Fliess, A.; Maass, G. & Pingoud, A. (1989) *Biochemistry* 28 2678-2684

Badger, R.M. & Bauer, S.H. (1937) *J. Chem. Phys.* 5 839-851

Benevides, J.M.; Wang, A.H-J.; van der Marel, G.A.; van Boom, H. & Thomas, G.J. (1988) *Biochemistry* 27 931-938

Benevides, J.M.; Weiss, M.A. & Thomas, G.J. (1991) *Biochemistry* 30 5955-5963

Berg, J.M. (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85 99-102

Berkner, K.L. & Folk, W.R. (1977) *J. Biol. Chem.* 252 3185-3193

Bodnar, J.W.; Zempsky, W.; Warder, D.; Bergson, C. & Ward, D.C. (1983) *J. Biol. Chem.* 258 15206-15213

Bougueleret, L.; Schwartzstein, M.; Tsugitu, A. & Zabeau, M. (1984) *Nucleic Acids Res.* 12 3659-3676

Bougueleret, L.; Tenchini, M.L.; Botterman, J. & Zabeau, M. (1985) *Nucleic Acids Res.* 13 3823-3839

Boyer, H.W.; Chow, L-T.; Dugaiczyk, A.; Hedgpeth, J. & Goodman, H.M. (1973) *Nature (New Biology)* 244 40-43

Bradford, M.M. (1976) *Anal. Biochem.* 72 248-254

Brennan, C.A.; van Cleve, M.D. & Gumport, R.I. (1986) *J. Biol. Chem.* 261 7270-7278

Brown, T. & Brown, D.J.S. (1991) *Oligonucleotides and Analogues - a practical approach* (Ed.: Eckstein, F.) IRL Press, Oxford

Butkus, V.; Klimašauskas, S.; Petrauskienė, L.; Manelienė, Z.; Janulaitis, A. Minchenkova, L.E. & Schyolkina, A.K. (1987) *Nucleic Acids Res.* 15 8467-8478

Calladine, C.R. & Drew, H.R. (1986) *J. Mol. Biol.* 192 907-918

Cech, D. & Holy, A. (1977) *Colect. Czech. Chem. Commun.* 42 2246-2260

Christopherson, M.S. & Broom, A.D. (1991) *Nucleic Acids Res.* 19 5719-5724

Connolly, B.A. (1991) *Oligonucleotides and Analogues - a practical approach* (Ed. Eckstein, F) IRL Press, Oxford

Connolly, B.A. & Newman, P.C. (1989) *Nucleic Acids Res.* 17 4957-4974

Connolly, B.A.; Eckstein, F. & Pingoud, A. (1984a) *J. Biol. Chem.* 259 10760-10763

Connolly, B.A.; Potter, B.V.L.; Eckstein, F.; Pingoud, A. & Grotjahn, L. (1984b) *Biochemistry* 23 3443-3453

Cowart, M. & Benkovic, S.J. (1991) *Biochemistry* 30 788-796

D'Arcy, A.; Brown, R.S.; Zabeau, M.; van Resandt, R.W. & Winkler, F.K. (1985) *J. Biol. Chem.* 260 1987-1990

Delcourt, S.C. & Blake, R.D. (1991) *J. Biol. Chem.* 266 15160-15169

Dickerson, R.E. & Drew, H.R. (1981) *J. Mol. Biol.* 149 761-786

Diekmann, S. & McLaughlin, L.W. (1988) *J. Mol. Biol.* 202 823-834

Drew, H.R. & Travers, A.A. (1984) *Cell* 37 491-502

Drew, H.R.; McCall, M.J. & Calladine, C.R. (1988) *Ann. Rev. Cell Biol.* 4 1-20

Dugaiczyk, A.; Hedgpeth, J.; Boyer, H.W. & Goodman, H.W. (1974) *Biochemistry* 13 503-512

Dwyer-Hallquist, P.; Kézdy, F.J. & Agarwal, K.L. (1982) *Biochemistry* 21 4693-4700

Evertsz, E.M.; Thomas, G.A. & Peticolas, W.L. (1991) *Biochemistry* 30 1149-1155

Fairall, L.; Rhodes, D. & Klug, A. (1986) *J. Mol. Biol.* 192 577-591

Fairall, L.; Martin, S. & Rhodes, D. (1989) *EMBO J.* 8 1809-1817

Fasman, G. (Ed.) (1975) *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* Vol. 1, 3rd ed., CRC Press, Ohio

Fersht, A.R. (1985) *Enzyme Structure and Mechanism*, 2nd ed., Freeman, New York

Fersht, A.R. (1987) *Trends Biochem. Sci.* 12 301-304

Fersht, A.R. (1988) *Biochemistry* 27 1577-1580

Fisher, E.F. & Caruthers, M.H. (1979) *Nucleic Acids Res.* 7 401-416

Fliess, A.; Wolfes, H.; Seela, F. & Pingoud, A. (1988) *Nucleic Acids Res.* 16 11781-11793

Fodor, S.P.A.; Rava, R.P.; Hays, T.R. & Spiro, T.G. (1985) *J. Am. Chem. Soc.* 107 1520-1529

Fox, J.J.; Wempen, I.; Hampton, A. & Doerr, I.L. (1958) *J. Am. Chem. Soc.* 80 1669-1675

Frederick, C.A.; Grable, J.; Melia, M.; Samudzi, C.; Jen-Jacobson, L.; Wang, B-C.; Greene, P.; Boyer, H.W. & Rosenberg, J.M. (1984) *Nature* 309 327-330

Freemont, P.S.; Lane, A.N. & Sanderson, M.R. (1991) *Biochem. J.* 278 1-23

Freier, S.M.; Sugimoto, N.; Sinclair, A.; Alkema, D.; Neilson, T.; Kierzek, R.; Caruthers, M.H. & Turner, D.H. (1986) *Biochemistry* 25 3214-3219

Frost, A. & Pearson, R. (1961) *Kinetics and Mechanism* John Wiley & Sons, New York

Gaffney, B.L. & Jones, R.A. (1982a) *Tetrahedron Lett.* 23 2253-2256

Gaffney, B.L. & Jones, R.A. (1982b) *Tetrahedron Lett.* 23 2257-2260

Gaffney, B.L.; Marky, L.A. & Jones, R.A. (1984) *Tetrahedron* 40 3-13

Gait, M.J. (Ed.) (1984) *Oligonucleotides Synthesis - a practical approach*, IRL Press, Oxford

Gans, P. (1971) *Vibrating Molecules* Chapman & Hall, London

Gildea, B. & McLaughlin, L.W. (1989) *Nucleic Acids Res.* 17 2261-2281

Grasby, J. (1991) *PhD Thesis*, University of Southampton, UK

Greene, P.J.; Poonian, M.S.; Nussbaum, A.L.; Tobias, L.; Garfin, D.E.; Boyer, H.W. & Goodman, H.M. (1975) *J. Mol. Biol.* 99 237-261

Greene, P.J.; Gupta, M.; Boyer, H.W.; Brown, W.E. & Rosenberg, J.M. (1981) *J. Biol. Chem.* 256 2143-2153

Halford, S.E. (1983) *Trends Biochem. Sci.* 8 455-460

Halford, S.E. & Goodall, A.J. (1988) *Biochemistry* 27 1771-1777

Halford, S.E. & Johnson, N.P. (1980) *Biochem. J.* 191 593-604

Halford, S.E.; Johnson, N.P. & Grinstead, J. (1980) *Biochem. J.* 191 581-592

Halford, S.E.; Lovelady, B.M. & McCallum, S.A. (1986) *Gene* 41 173-181

Hanna, N.B.; Ramasamy, K.; Robins, R.K. & Revankar, G.R. (1988) *J. Heterocyclic Chem.* 25 1899-1903

Harrison, S.C. (1991) *Nature* 353 715-719

Harrison, S.C. & Aggarwal, A.K. (1990) *Annu. Rev. Biochem.* 59 933-969

Hirakawa, A.Y. & Tsuboi, M. (1975) *Science* 186 790-797

Hsu, M-T. & Berg, P. (1978) *Biochemistry* 17 131-138

Jack, W.E.; Terry, B.J. & Modrich, P. (1982) *Proc. Natl. Acad. Sci. (USA)* 79 4010-4014

Jiricny, J.; Wood, S.G.; Martin, D. & Ubasawa, A. (1986) *Nucleic Acids Res.* 14 6579-6590

Johnston, T.P.; Holum, L.B. & Montgomery, J.A. (1958) *J. Am. Chem. Soc.* 80 6265-6271

Jordan, S.R. & Pabo, C.O. (1988) *Science* 242 893-899

Joyce, M.C. & Steitz, T.A. (1987) *Trends Biochem. Sci.* 12 288-292

Kihara, K.; Nomiyama, H.; Yukuhiro, M. & Mukai, J-I (1976) *Anal. Biochem.* 75 672-673

Kim, Y.; Grable, J.C.; Love, R.; Greene, P.J. & Rosenberg, J.M. (1990) *Science* 249 1307-1309

Kissenger, C.R.; Liu, B.; Martin-Blanco, E.; Kornberg, T.B. & Pabo, C.O. (1990) *Cell* 63 579-590

Kochetkov, N.K. & Budovski, E.I. (1972) *Organic Chemistry of Nucleic Acids* part B, Plenum Press, London

Krishnan, R.S. (1971) *The Raman Effect*, Vol. 1 (Ed.: Anderson, A) Marcel Dekker, New York

Kubareva, E.A.; Pein, C-D.; Gromova, E.S.; Kuznezova, S.A.; Tashlitzki, V.N.; Cech. D. & Shabarova, Z.A. (1988) *Eur. J. Biochem.* 175 615-618

Kung, P-P. & Jones, R.A. (1991) *Tetrahedron Lett.* 32 3919-3922

Laskowski, M. (1971) *The Enzymes* Vol. IV, 3rd ed., (Ed. Boyer, P.D.) Academic Press, London

Lee, M. S.; Gippert, G.P.; Soman, K.V.; Case, D.A. & Wright, P.E. (1989) *Science* 245 635-637

Lehman, I.R. (1974) *Science* 186 790-797

Lesser, D.R.; Kurpiewski, M.R. & Jen-Jacobson, L. (1990) *Science* 250 776-786

Lu, A-L.; Jack, W.E. & Modrich, P. (1981) *J. Biol. Chem.* 256 13200-13206

Luke, P.A.; McCallum, S.A. & Halford, S.E. (1987) *Gene Amplication and Analysis*, Vol. V: Restriction Endonucleases and Methylases (Ed.: Chirikian, J.G.) Elsevier, New York

Mazzarelli, J.; Scholtissek, S. & McLaughlin, L.W. (1989) *Biochemistry* 28 4616-4622

McBride, L.J.; Kierzek, R.; Beaucage, S.L. & Caruthers, M.H. (1986) *J. Am. Chem. Soc.* 108 2040-2048

McClaren, J.A.; Frederick, C.A.; Wang, B-C.; Greene, P.; Boyer, H.W.; Grable, J. & Rosenberg, J.M. (1986) *Science* 234 1526-1541

McLaughlin, L.W.; Benseler, F.; Graeser, E.; Piel, N. & Scholtissek, S. (1987) *Biochemistry* 26 7238-7245

McLaughlin, L.W.; Leong, T.; Benseler, F. & Piel, N. (1988) *Nucleic Acids Res.* 16 5631-5644

Milne, G.H. & Townsend, L.B. (1974) *J. Med. Chem.* 17 263-268

Modrich, P. (1979) *Q. Rev. Biophys.* 12 315-369

Modrich, P. & Rubin, R.A. (1977) *J. Biol. Chem.* 252 7273-7278

Modrich, P. & Zabel, D. (1976) *J. Biol. Chem.* 251 5866-5874

Nakabeppu, Y. & Nathans, D. (1989) *EMBO J.* 8 3833-3841

Needels, M.C.; Fried, S.R.; Love, R.; Rosenberg, J.M.; Boyer, H.W. & Greene, P.J. (1989) *Proc. Natl. Acad. Sci. (USA)* 86 3579-3583

Nelson, M. & McLelland, M. (1991) *Nucleic Acids Res.* 19 2045-2071

Nelson, H.C.M.; Finch, J.T.; Luisi, B.F. & Klug, A. (1987) *Nature* 330 221-226

Newman, P.C. (1989) *PhD Thesis*, University of Southampton, UK

Newman, A.K.; Rubin, R.A.; Kim, S-H. & Modrich, P. (1981) *J. Biol. Chem.* 256 2131-2139

Newman, P.C.; Nwosu, V.U.; Williams, D.M.; Cosstick, R.; Seela, F. & Connolly, B.A. (1990a) *Biochemistry* 29 9891-9901

Newman, P.C.; Williams, D.M.; Cosstick, R.; Seela, F. & Connolly, B.A. (1990b) *Biochemistry* 29 9902-9910

Nikiforov, T.T. & Connolly, B.A. (1991) *Tetrahedron Lett.* 32 3851-3854

Nishimura, Y. & Tsuboi, M. (1980) *Science* 210 1358-1360

Nishimura, Y.; Hirakawa, A.Y.; Tsuboi, M. & Nishimura, S. (1976) *Nature* 260 173-174

Nishimura, Y.; Hirakawa, A.Y. & Tsuboi, M. (1978) *Advances in Infrared and Raman Spectroscopy*, Vol. 5 (Eds.: Clark, R.J.H. & Hester, R.E.)

Nishimura, Y.; Tsuboi, M.; Kato, S. & Morokuma, K. (1981) *J. Am. Chem. Soc.* 103 1354-1358

Nishimura, Y.; Torigoe, C. & Tsuboi, M. (1986) *Nucleic Acids Res.* 14 2737-2748

Nwosu, V.U.; Connolly, B.A.; Halford, S.E. & Garnet, J. (1988) *Nucleic Acids Res.* 16 3705-3720

Oas, T.G.; McIntosh, L.P.; O'Shea, E.K.; Dahlquist, F.W. & Kim, P.S. (1990) *Biochemistry* 29 2891-2894

Ono, A. & Ueda, T. (1987a) *Nucleic Acids Res.* 15 3059-3072

Ono, A. & Ueda, T. (1987b) *Nucleic Acids Res.* 15 219-232

Ono, A.; Sato, M.; Ohtani, Y. & Ueda, T. (1984) *Nucleic Acids Res.* 12 8939-8949

O'Shea, E.K.; Rutkowski, R. & Kim, A. (1989) *Science* 243 538-542

O'Shea, E.K.; Klemm, J.D.; Kim. P.S. & Alber, T. (1991) *Science* 254 539-544

Otwinowski, Z.; Schevitz, R.W.; Zhang, R-G.; Lawson, C.L.; Joachimiak, A.; Marmorstein, R.Q.; Luisi, B.F. & Sigler, P.B. (1988) *Nature* 335 321-329

Pavletich, N.P. & Pabo, C.O. (1991) *Science* 252 809-817

Perutz, M.F. (1990) *Mechanisms of cooperative and allosteric regulation of proteins*, Cambridge University Press, Cambridge

Peticolas, W.L.; Kubasek, W.L.; Thomas, G.A. & Tsuboi, M. (1987) *Biological Applications of Raman Spectroscopy*, Vol. 1 (Ed.: Spiro, T.G.) Wiley, New York

Phillips, S.E.V. (1991) *Curr. Opinions Struct. Biol.* 1 89-98

Pochon, F,; Balny, C.; Scheit, H. & Michelso, A.M. (1971) *Biochem. Biophys. Acta* 228 49-56

Pohl, F.M. & Jovin, T.M. (1972) *J. Mol. Biol.* 67 375-396

Polisky, B.; Greene, P.; Garfin, D.E.; McCarthy, B.J.; Goodman, H.M. & Boyer, H.W. (1975) *Proc. Natl. Acad. Sci. (USA)* 72 3310-3314

Pon, R.T. (1987) *Nucleic Acids Res.* 15 7203

Qian, Y.Q.; Billeter, M.; Otting, G.; Müller, M.; Gehring, W.J. & Wüthrich, K. (1989) *Cell* 59 573-580

Rafferty, J.B.; Somers, W.S.; Saint-Girons, I. & Phillips, S.E.V. *Nature* 341 705-710

Raman, C.V. & Krishnan, K.S. (1928) *Nature* 121 501-502

Rappaport, H.P. (1988) *Nucleic Acids Res.* 16 7253-7267

Revankar, G.R.; Gupta, P.K.; Adams, A.D.; Dalley, N.K.; McKernan, P.A.; Cook, P.D.; Canonico, P.G. & Robins, R.K. (1984) *J. Med. Chem.* 27 1389-1396

Roberts, R.J. & Macelis, D. (1991) *Nucleic Acids Res.* 2077-2109

Rubin, R.A. & Modrich, P. (1977) *J. Biol. Chem.* 252 809-817

Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York

Salet, C.; Bensasson, R.V. & Favre, A. (1983) *Photochemistry and Photobiology* 38 521-525

Sambrook, J.; Fritsch, E.F. & Maniatis, T. (1989) *Molecular Cloning* Vol. 3, 2nd ed., Cold Spring Harbour Laboratory Press, New York

Schildkraut, I.; Banner, C.; Rhodes. C. & Parekh, S. (1984) *Gene* 27 327-329

Schrader, B. (1989) *Raman/Infrared Atlas of Organic Compounds* 2nd ed., VCH, New York

Seela, F. & Driller, H. (1986) *Nucleic Acids Res.* 14 2319-2332

Seeman, N.C.; Rosenberg, J.M. & Rich, A. (1976) *Proc. Natl. Acad. Sci. (USA)* 73 804-808

Selent, U.; Rüter, T.; Köhler, E.; Liedtke, M.; Thielking, V.; Alves, J.; Oelgeschläger, T.; Wolfes, H.; Peters, F. & Pingoud, A. (1991) *Biochemistry* (submitted)

Sinha, N.D.; Bienat, T.; McManus, J. & Köster, H. (1984) *Nucleic Acids Res.* 12 4539-4557

Smith, M.; Rammler, D.H.; Goldberg, I.H. & Khorana, H.G. (1962) *J. Am. Chem. Soc.* 84 430-440

Steitz, T.A.; Ohlendorf, D.H.; McKay, D.B.; Anderson, W.F. & Matthews, B.W. (1982) *Proc. Natl. Acad. Sci. (USA)* 79 3097-3100

Stephenson, F.H.; Ballard, B.T.; Boyer, H.W.; Rosenberg, J.M. & Greene, P.J. (1989) *Gene* 85 1-13

Suck, D.; Lahm, A. & Oefer, C. (1988) *Nature* 332 464-468

Taylor, J.D. & Halford, S.E. (1989) *Biochemistry* 28 6198-6207

Taylor, J.D.; Goodall, A.J.; Vermote, C.L. & Halford, S.E. (1990) *Biochemistry* 29 10727-10733

Taylor, J.D.; Badcoe, I.G.; Clarke, A.R. & Halford, S.E. (1991) *Biochemistry* 30 8743-8753

Terry, B.J.; Jack, W.E.; Rubin, R.A. & Modrich, P. (1983) *J. Biol. Chem.* 258 9820-9825

Thamann, T.J.; Lord, R.C.; Wang, A.H-J. & Rich, A. (1981) *Nucleic Acids Res.* 9 5443-5457

Thewalt, U. & Bugg, C.E. (1972) *J. Am. Chem. Soc.* 94 8892-8898

Thielking, V.; Alves, J.; Fliess, A.; Maass, G. & Pingoud, A. (1990) *Biochemistry* 29 4682-4691

Thielking, V.; Selent, U.; Köhler, E.; Wolfes, H.; Pieper, U.; Geiger, R.; Urbanke, C.; Winkler, F. & Pingoud, A. (1991a) *Biochemistry* 30 6416-6422

Thielking, V.; Selent, U.; Köhler, E.; Landgraf, A.; Wolfes, H.; Alves, J. & Pingoud, A. (1991b) *Biochemistry* (submitted)

Thomas, G.J. & Wang, A.H-J. (1988) *Nucleic Acids and Molecular Biology*, Vol. 2 (Eds.: Eckstein, F. & Lilley, D.M.J.) Springer-Verlag, Berlin

Ti, G.S.; Gaffney, B.L. & Jones, R.A. (1982) *J. Am. Chem. Soc.* 104 1316-1319

Travers, A.A. (1989) *Annu. Rev. Biochem.* 58 427-452

Tsuboi, M.; Nishimura, Y.; Hirakawa, A.Y. & Peticolas, W.L. (1987) *Biological Applications of Raman Spectroscopy*, Vol. 2 (Ed.: Spiro, T.G.) Wiley, New York

Ueda, T.; Miura, K. & Kasai, T. (1978) *Chem. Pharm. Bull.* 26 2122-2127

Van Cleve, M.D. & Gumport, R.I. (1992) *Biochemistry* 31 334-339

Wang, A.H-J.; Quigley, G.J.; Kolpak, F.J.; Crawford, J.L.; van Boom, J.H.; van der Marel, G. & Rich, A. (1979) *Nature* 282 680-686

Wang, A.H-J.; Quigley, G.J.; Kolpak, F.J.; van der Marel, G. & van Boom, J.H. (1981) *Science* 211 171-176

Ward, D.C.; Reich, E. & Stryer, L. (1969) *J. Biol. Chem.* 244 1228-1237

Waters, T.R. & Connolly, B.A. (1991a) *Nucleosides Nucleotides* (in press)

Waters, T.R. & Connolly, B.A. (1991b) *Anal. Biochem.* (submitted)

Watson, J.D. & Crick, F.H.C. (1953) *Nature* 171 737-738

Weast, R.C.; Astle, M.J. & Beyer, W.H. (Eds.) (1988) *CRC Handbook of Chemistry and Physics*, 69th ed., CRC Press, Florida

Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S. & Weiner, P. (1984) *J. Am. Chem. Soc.* 108 765-784

Wells, T.N.C. & Fersht, A.R. (1986) *Biochemistry* 25 1881-1886

Wilkinson, A.J.; Fersht, A.R.; Blow, D.M.; & Winter, G. (1983) *Biochemistry* 22 3581-3586

Winkeler, H-D. & Seela, F. (1983) *J. Org. Chem.* 48 3119-3122

Winkler, F.K. (1991) *Curr. Opinions Struct. Biol.* (in press)

Winkler, F.K.; D'Arcy, A.; Blöcker, H.; Frank, R. & van Boom, J.H. (1991) *J. Mol. Biol.* 217 235-238

Wolfes, H.; Alves, J.; Fliess, A.; Geiger, R. & Pingoud, A. (1986) *Nucleic Acids Res.* 14 9063-9080

Woodbury, C.P.; Hagenbüchle, O. & von Hippel, P.H. (1980a) *J. Biol. Chem.* 255 11534-11546

Woodbury, C.P.; Downey, R.L. & von Hippel, P.H. (1980b) *J. Biol. Chem.* 255 11534-11546

Wright, D.J.; King, K. & Modrich, P. (1989) *J. Biol. Chem.* 264 11816-11821

Xu, Y-Z,; Zheng, Q. & Swann, P.F. (1991) *Tetrahedron Lett.* 32 2817-2820

Yolov, A.A.; Vinogradova, M.N.; Gromova, E.S.; Rosenthal, A.; Cech, D.; Veiko, V.P.; Metelev, V.G.; Kosykh, V.G.; Buryanov, Y.A.; Bayev, A.A. & Shabarova, Z.A. (1985) *Nucleic Acids Res.* 13 8983-8998

Zon, G. & Thompson, J.A. (1986) *Biochromatography* 1 22-32