The interactions between *Eco*KI methyltransferase and specific DNA duplexes

by

Angela Her-Ser Chen

Thesis presented for the degree of Doctor of Philosophy

Department of Chemistry and

Institute of Cell and Molecular Biology

University of Edinburgh

1995

Declaration

The contents of this thesis are, unless otherwise stated, my own work. A portion of it has been accepted and published as a full paper by Nucleic Acid Research. A reprint is attached in the appendix.

Angela H. S. Chen

Department of Chemistry and Institute of Cell and Molecular Biology
University of Edinburgh

September, 1995

Acknowledgements

I wish to express my sincere gratitude to my supervisors Professor Tom Brown and Professor Noreen Murray for their guidance throughout the course of this study. Without their support, encouragement, and patience, this work could not have been successfully completed. I am also very grateful to Professor R. Ramage of Chemistry Department and Dr D. Finnegan of ICMB for allowing me to use the experimental facilities in both departments.

It was a valuable experience working with my fellow postgraduate students and colleagues both in the Department of Chemistry and Institute of Cell and Molecular Biology. The stimulating scientific discussion and technical support made my years at Edinburgh enjoyable. I am also grateful to Dr Lynn Powell and Dr David Dryden of ICMB whose enthusiastic assistance has been very important to complete this research. I would like to thank Graham and Frank for photographs and Dr Katherine McAuley-Hecht for taking the colour photographs of DNA models.

I am particularly grateful to the Oswel DNA Service for financial support and the British Council for an ORS award. My thanks also extend to Dr Dorcas Brown, a good friend who has been very generous to my family, and whose professional expertise has helped me a great deal in completing my work. A special thanks is also due to my husband Professor Hong-Yang Lu whose support made possible my study in Scotland.

Angela Chen
September, 1995

ABSTRACT

The S polypeptide of the *Eco*KI type I methyltransferase (mtase) has the function of recognising DNA. Specific contacts between *Eco*KI (mtase) and the DNA containing its recognition sequence (-AAC-$(N)_6$-GTGC-) have been investigated using various substituted oligonucleotides, namely 4-thiothymidine, 5-bromodeoxyuridine, 5-iododeoxyuridine and deoxyuridine. After ultraviolet irradiation, the DNA-S polypeptide crosslinked complexes were detected with the specific target recognition DNA, but not with the non-specific DNA, following separation by electrophoresis through a SDS-PAGE gel. The crosslinking of DNA and mtase occurs only between the bottom-strand of the DNA and the S polypeptide, as illustrated by labelling the different strands of DNA with $\gamma^{32}P$. The crosslinking efficiency between DNA-mtase increased only in one of the nine single sites of BrdU substitution in which the BrdU was substituted at the residue complementary to the first adenine in the -AAC- sequence. This demostrates a close contact between this sequence and the S subunit at the major groove.

Peptide sequencing of various trypsin digested samples of the crosslinked protein-DNA complex further illustrates that the Tyr-27 of the NH_2-terminal domain of the S polypeptide is crosslinked with DNA. This is consistent with the role of the N terminal domain of the S subunit in recognizing the -AAC- sequence.

By amino acid sequence comparison, Tyr-27 and other small sequence segments can be found at corresponding positions in other two type I enzymes, which have a similar trinucleotide recognition site. In addition the comparison reveals that motif LP-GWEW is partly retained not only in the amino recognition domain but also in the carboxyl recognition domain of the S subunit of type I enzymes.

Abbreviations

A	Adenosine
AcOH	Acetic acid
AdoMet	S-adenosyl methionine (SAM)
Ala [A]	Alanine
Amps	Ammonium persulphate
ARD	Amino recognition domain
Arg [R]	Arginine
Asp [D]	Aspartic acid
Asn [N]	Asparagine
ATP	Adenosine triphosphate
ATZ	Anilinothiozolinone
Aufs	Absorbance of UV at full scale
bp	Base pair
B.b.	Bromophenol blue
BrdU	5-Bromodeoxyuridine
C	Cytidine
CNBr	Cyanogen bromide
CRD	Carboxyl recognition domain
Cys [C]	Cysteine
DBU	1,8-Diazobicyclo[5.4.1]undec-7-ene(1,1-5)
DCM	Dichloromethane
dG	2'-Deoxyguanosine
DMF	N,N-Dimethylformamide
DMT	4,4'-Dimethoxytrityl
DPCC	Diphenyl carbamyl chloride
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
dU	Deoxyuridine
E. coli	*Escherichia coli*
ENase	Endodeoxyribonuclease
EDTA	Ethylenediamine tetraacetic acid
EtOAc	Ethyl acetate
FPLC	Fast-protein liquid chromatography
G	Guanosine
Gln [Q]	Glutamine
Glu [E]	Glutamic acid
Gly [G]	Glycine
His [H]	Histidine
HPLC	High performance liquid chromatography
hsd	(gene of) host specificity of DNA

Ile [I]	Isoleucine
IodU	5-Iododeoxyuridine
IR	Infrared spectra
kDa	kilodalton
Leu [L]	Leucine
Lys [K]	Lysine
M	Molar
Mcr	(R-M system of) modified cytosine restriction
MeOH	Methanol
Met [M]	Methionine
mtase	Methyltransferase
NH_4OAc	Ammonium acetate
NMR	Nuclear magnetic resonance spectra
nt	Nucleotides
OD	Optical density unit
PAGE	Polyacrylamide gel electrophoresis
PCA	Polymerase chain reaction
Phe [F]	Phenylalanine
PITC	Phenyl isothiocyanate
PNK	T4 polynucleotide kinase
Pro [P]	Proline
PTC	Phenyl thiocarbamyl
PTH	Phenyl thiohydantoin
PVDF	Polyvinylidene fluoride
R-M	Restriction-Modification
S	Specificity
SDS	Sodium dodecylsulfate
Ser [S]	Serine
T	Thymidine
TBE	Tris-borate-EDTA
TCA	Trichloroacetic acid
TEA	Triethylamine (Et_3N)
TEMED	N,N,N',N'-tetramethylenediamine
THF	Tetrahydrofuran
Thr [T]	Threonine
TLC	Thin layer chromatography
Trp [W]	Tryptophan
Tris	Tris (hydroxymethyl) methylamine
T^{4S}	4-Thiothymidine
Tyr [Y]	Tyrosine
Val [V]	Valine
UV	Ultraviolet

Restriction enzymes mentioned in this thesis:

R-M system	Organism
*Bam*HI	*Bacillus amyloliquefaciens*
*Cfr*AI	*Citrobacter freundii*
*Eco*57	*Escherichia coli* RFL57
*Eco*AI	*Escherichia coli* *
*Eco*BI	*Escherichia coli* *
*Eco*DI	*Escherichia coli* *
*Eco*EI	*Escherichia coli* *
*Eco*KI	*Escherichia coli* *
*Eco*prrI	*Escherichia coli* *
*Eco*DXXI	*Escherichia coli* plasmid DXXI
*Eco*R124I	*Escherichia coli* plasmid R124
*Hha*I	*Haemophilus haemolyticus*
*Pvu*II	*Proteus vulgaris*
*Sty*LTIII (*StySB*)	*Salmonella typhimurium* LT2
*StySP*I	*Salmonella potsdam*
*Taq*I	*Thermus aquaticus* YT1

*The R-M systems are from different strains of *Escherichia coli* .

By convention, R-M systems are designated by a three letter acronym derived from the name of the organism in which they occur [221]. The first letter comes from the genus, and the second and third letters from the species. The strain designation, if any, follows the acronym. Different systems in the same organism are distinguished by Roman numerals [11].

Contents

Chapter 1 Introduction

1.1 *General introduction*

The greatest advances in molecular cell biology in the recent past have been based on the analysis and manipulation of macromolecules, particularly DNA. DNA oligonucleotides (up to 100 bases long) can now be chemically synthesized entirely automatically; not only natural oligonucleotides, but any desired mutant sequence can be produced.

Endonucleases (ENase) that catalyse the sequence-specific hydrolysis of double stranded DNA have greatly facilitated the techniques of molecular cloning and DNA sequencing. The range of techniques allied to molecular cloning and nucleotide sequencing has been used to study the interactions between type I *Escherichia coli* K-12 (*Eco*KI) methyltransferase (mtase) and specific DNA duplexes.

1.2 Solid phase synthesis of oligonucleotides

The solid phase method was developed by Bruce Merrifield at the Rockefeller Institute (now Rockefeller University) in the 1950s and 1960s initially for peptide synthesis. Today almost all synthetic oligonucleotides are prepared by solid phase phosphoramidite techniques.

The solid support (usually borosilicate glass beads) to which the first nucleoside is attached at the 3′-terminus, is trapped between two filters in the synthesis column. During the synthesis the glass beads are repeatedly soaked in reagents and washed with an appropriate solvent. The excess reagents and solvents are rapidly blown out of the column to waste by a stream of inert gas. The full phosphoramidite cycle on an automatic synthesizer may consist of up to 100 separate steps. The synthesis of oligonucleotides is normally carried out in the 3′ to 5′ direction to take advantage of the high chemical reactivity of the primary 5′-hydroxyl group.

During synthesis firstly, the support-bound nucleoside is deprotected to provide a free 5′-hydroxyl group for the attachment of the second nucleotide. An excess of the second nucleotide, protected at the 5′-hydroxyl position to prevent self-polymerization and activated at the 3′-phosphate position to facilitate condensation, is then added and a support-bound dimer is formed. The unreacted support-bound monomer is capped to prevent any further monomer additions at the subsequent cycles. The phosphite of the support-bound dimer is oxidized to the phosphotriester of the support-bound dimer. The synthesis is then repeated until the desired sequence has been constructed.

1.3 The Restriction-Modification (R-M) systems

The phenomenon of restriction and modification was first discovered in *E. coli* in the early 1950s by Bertani and Weigle [1] in the course of experiments with the phages P2 and λ. Certain strains of bacteria were found to inhibit propagation of viruses grown previously on different strains. The effect was traced to sequence-specific ENases [2], whereas viruses grown in the same strain were protected by the action of mtases. The ENase involved is a restriction ENase, and the methylase is a modification methylase. The word "restriction" refers to the function of these enzymes in the bacteria of origin: a restriction ENase restricts incoming foreign DNA (for example, bacteriophage DNA or DNA taken up during transformation) by cleaving it at specific sites, called restriction sites. The primary function of the modification methylase is to protect the cell's own genome from restriction by adding a methyl group most usually to two bases, within the recognition sequence. When a methyl group is present, the restriction ENase is prevented from cutting the DNA. Together with the restriction ENase, the methylating enzyme forms a Restriction-Modification (R-M) system. R-M systems are common throughout the prokaryotic world, over two thousand members have been discovered. They have proved useful for analysing and rearranging DNA, and this has encouraged the search for more of them.

Classical R-M systems which have both a restriction and a modification ability are classified according to enzyme composition and cofactor requirements, recognition sequence symmetry and cleavage position [3]. The major features of the three classes (type I, type II and type III) of R-M systems are shown in Table 1 [4]. Type I R-M systems are hetero-oligomeric enzymes, where one enzyme complex is capable of both restriction and modification. Most of them have been found in commonly studied enteric bacteria [5]. Type II R-M systems are the most numerous, with over 2,400 members. The restriction ENase and mtase of type II systems are separate enzymes. The type II restriction ENases cut DNA within their recognition sequence. Type III R-M systems, containing 5 members, are also hetero-oligomeric enzymes, which cut DNA close to the recognition sequence. The type IIs systems [6] containing over 35 members are a subgroup of the type II systems that cleave DNA outside their recognition sequence. Some systems do not fit readily into the present classification system [7,8] and *Eco*57I has been tentatively termed a type IV system [7]. Other modification-dependent enzymes (e.g. *Mcr*BC) are not classical R-M systems, as they have no cognate methylase and only cut DNA that contains certain modifications [9].

Table 1. Characteristics of R-M systems [4]

Characteristics	System type		
	type I	type II	type III
Subunits for restriction	3	1	2
Subunits for modification	3 or 2	1	1 or 2
Cofactors for restriction@	Mg^{2+}, ATP, AdoMet	Mg^{2+}	Mg^{2+}, ATP (AdoMet)
Cofactor for modification@	AdoMet (ATP, Mg^{2+})	AdoMet	AdoMet, Mg^{2+}
Other enzymatic activities	ATPase	none	none
Distance between recognition and cleavage sites	Variable and great	Sites coincide	25-30 base pairs

@ Cofactors listed in parentheses stimulate the reaction but are not absolutely required.

1.3.1 Type I R-M systems and type IA R-M systems of *Eco*KI.

Type I R-M systems can be divided into three (IA, IB and IC) families whose members have nearly identical restriction (R) and modification (M) subunits (80-90 % identity) but different specificity (S) subunits. Representatives of all these families are shown in Table 2. The IA and IB family enzymes are chromosomally-encoded in contrast to the IC enzymes which are generally plasmid-encoded.

The type IA R-M enzyme of *Escherichia coli* K (*Eco*KI) comprising three kinds of polypeptides, R, M, and S [4, 10-12] is both an ENase and a mtase. The*Eco*KI ENase consisting of two R, two M and one S subunits ($R_2M_2S_1$) with a molecular mass of 440 kilodalton (kDa) is essential for restriction while the mtase, a trimer (M_2S_1) of molecular mass of 170 kDa consisting of one S subunit and two M subunits is required for modification [3,13]. Each R subunit contains about 1,090 amino acids, each M and S subunit contains 529 and 464 amino acids respectively.

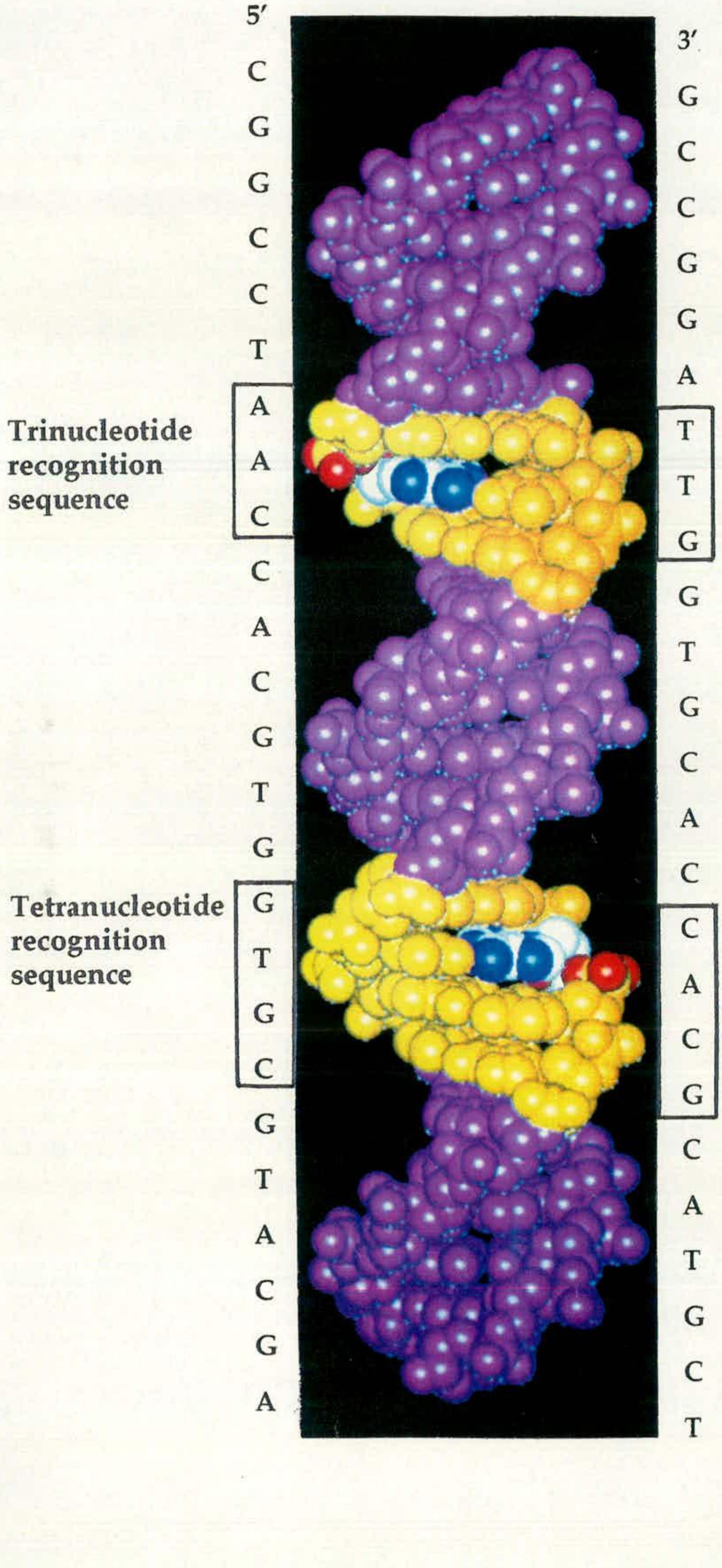
The specific recognition sites of type I enzymes are shown in Table 2. The *Eco*KI recognition sequence is -AAC-$(N)_6$-GTGC- [14,15]. It is asymmetric and comprises two components, one a trinucleotide (-AAC-), the other a tetranucleotide (-GTGC-) separated by a nonspecific spacer of 6-8 nucleotides (N) as shown in Figure 1.

Table 2. Type I target recognition sequences.

Enzyme	Target recognition sequences	Distance of target methylation adenosine (base pairs)	References
(IA) *Eco*KI	$\underline{A}$AC-N_6-GTGC TTG-N_6-C$\underline{A}$CG	10	[14]
(IA) *Eco*DI	TT$\underline{A}$-N_7-GTGY AAT-N_7-C$\underline{A}$GR	10	[116]
(IA) *Eco*BI	TG$\underline{A}$-N_8-TGCA ACT-N_8-$\underline{A}$CGT	10	[117,118]
(IA) *Sty*SPI	A$\underline{A}$C-N_6-GTRC TTC-N_6-C$\underline{A}$YC	10	[119]
(IA) *Sty*LTIII*	G$\underline{A}$G-N_6-RTAYG CTC-N_6-Y$\underline{A}$TRC	10	[119]
(IB) *Eco*AI	G$\underline{A}$G-N_7-GTCA CTC-N_7-C$\underline{A}$GT	11	[120]
(IB) *Eco*EI	G$\underline{A}$G-N_7-ATGC CTG-N_7-T$\underline{A}$CG	11	[25]
(IB) *Cfr*AI	GC$\underline{A}$-N_7-GTGG CGT-N_7-C$\underline{A}$CC	10	[27]
(IC) *Eco*R124I	GA$\underline{A}$-N_6-RTCG CTT-N_6-Y$\underline{A}$GC	9	[114]
(IC) *Eco*DXXI	TC$\underline{A}$-N_7-ATTC AGT-N_7-T$\underline{A}$AG	10	[121]
(IC) *Eco*prrI	CC$\underline{A}$-N_7-RTGC GGT-N_7-Y$\underline{A}$CG	10	[122]

**Sty*LTIII was previously known as *Sty*SB.
Adenines of methylation target are underlined.

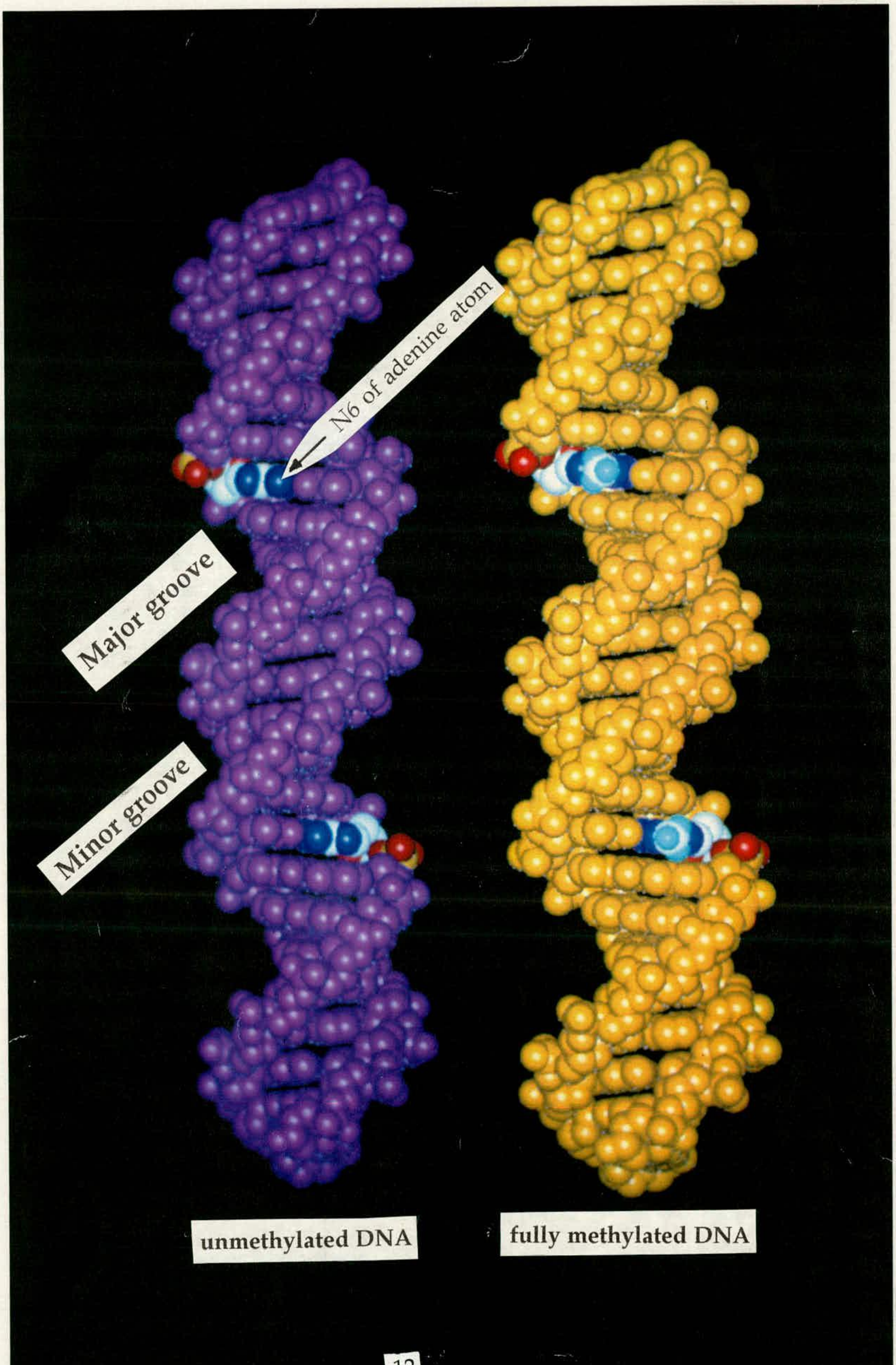
Figure 1. Model of DNA duplex (25 mer) with the *Eco*KI mtase recognition sequence, -AAC-$(N)_6$-GTGC-, (yellow balls and coloured adenines between yellow balls; carbon= white, nitrogen= navy blue, oxygen= red and phosphorus= yellow). The asymmetric and bipartite two components of the recognition sequence are separated by a nonspecific space of 6 nucleotides. The hydrogens are omitted on the model.


5′
C G G C C T A A C C A C G T G G T G C G T A C G A
3′
G C C G G A T T G G T G C A C C A C G C A T G C T
Trinucleotide recognition sequence
Tetranucleotide recognition sequence

All type I R-M enzymes methylate adenosyl residues (Table 2) in the presence of only S-Adenosyl Methionine (AdoMet, which serves as the methyl donor). Firstly the ENase (or mtase alone) binds to AdoMet and DNA containing the recognition sequence, then the target adenines are methylated by direct transfer of the methyl group from AdoMet to the N_6 [16-18] of the target adenines within the recognition sequence as illustrated in Figure 2. The methylatable adenines of type I recognition sites which have been determined are situated mostly 10 base pairs apart, about one helix turn (Figure 2). The methyl groups of the methylated *Eco*KI target adenines are in positions that do not interfere with base pairing (Figure 3) and are situated in the major groove of the B-form DNA helix (Figure 2). ATP-independent methylation of the *Eco*KI target sequence prevents restriction [19]. Methylation probably prevents cleavage by sterically hindering enzyme: substrate binding. Depending on the exact topology of the protein: DNA interface, methyl groups in some positions will interfere with binding but may not do so in others.

The *Eco*KI mtase, unlike most mtases studied to date, is relatively specific for hemimethylated targets. In fact hemimethylated DNA is the preferred substrate for methylation by *Eco*KI [3,20,21] while unmodified DNA is restricted [2].

Figure 2. Models of DNA duplex (25 mer) with the methylatable adenines of the *Eco*KI recognition sequence highlighted (specially coloured balls: carbon= white, nitrogen= navy blue, oxygen= red and phosphorus= yellow). The purple model is unmethylated DNA and the yellow model is fully methylated DNA (the methyl groups: carbon= white and hydrogen= sky blue). The methylatable adenines are in the major groove situated 10 base pairs apart about one helix turn. They are transferred from AdoMet to the N6 of adenine. The hydrogens are omitted on these models except for those in the methyl groups transferred from AdoMet.


N6 of adenine atom
Major groove
Minor groove
unmethylated DNA
fully methylated DNA

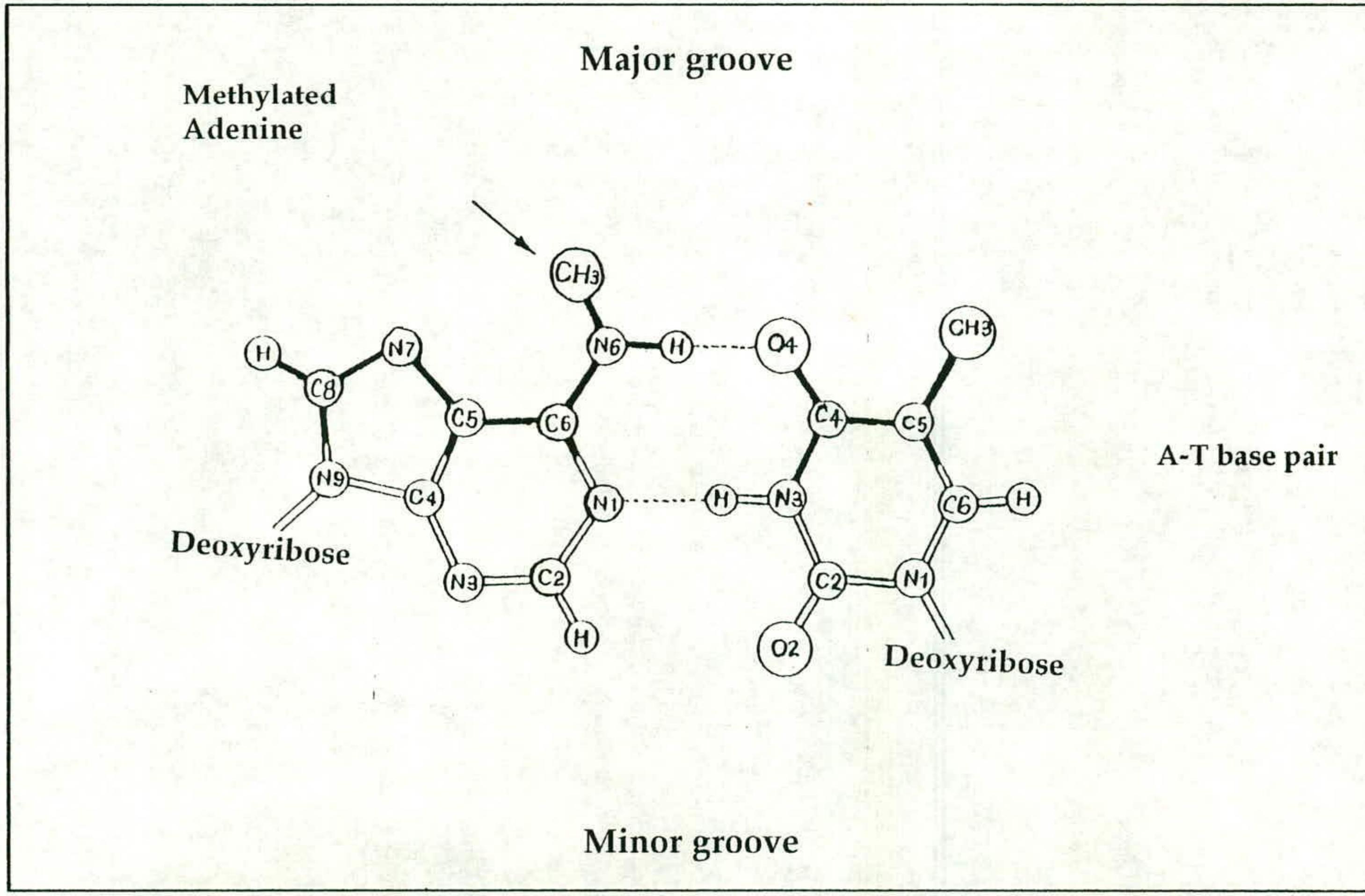


Figure 3. The methylation of DNA by mtase occurs through the direct transfer of a methyl group from AdoMet to the N6 of adenine. The methyl groups of the methylated target adenines are in positions that do not interfere with base pairing.

1.3.2 The S polypeptide of type I mtases

Many lines of evidence indicated that the S polypeptide of type I mtases dictates the recognition specificity [22-24]. The S polypeptides of type I family enzymes (IA, IB or IC) have two DNA target recognition domains (TRDs) of variable sequence, each of approximately 150 amino acids, which recognize the two components of the bipartite recognition sequence [25-28]. This organization was supported by the analysis of hybrid S genes generated by crossing-over between the central conserved regions of S genes confirming different specificities. Hybrids between *Salmonella typhimurium* LTIII (*Sty*SB) (GAGN_6RTAYG) and *Salmonella potsdam* SP (*Sty*SP) (AACN_6GTRC) displayed the predicted novel specificities: GAGN_6GTRC (SJ) and AACN_6RTAYG (SQ) [29-32]. The hybrid nature of the *hsd* S gene of *Sty*SQ, was initially confirmed by the heteroduplex studies of the *hsd*S genes of the *Sty*SP, *Sty*SB and *Sty*SQ [31]. The results showed that the *Sty*SQ specificity derived one variable region (TRD) from the *Sty*SBI gene and the other from the *Sty*SPI gene. The genetic exchange occurred within the conserved sequence of around 100 base pairs located between the two dissimilar regions each of approximately 500 base pairs [31]. The determination of DNA recognition sequence of *Sty*SPI, *Sty*SBI and *Sty*SQ [29] demonstrated that the recognition sequence of *Sty*SQ was a hybrid between the *Sty*SPI and *Sty*SBI, containing one specific recognition sequence from each parental strain. These results implied that the S polypeptide of type I mtases contains two recognition domains, each of the two domains correlate with two components of the DNA target sequence. Analysis of nucleotide sequences localized the crossover position of the *hsdS* gene of *Sty*SQ. Comparison of *hsd* S gene sequences of *Sty*SBI, *Sty*SPI and *Sty*SQ demonstrated that a polypeptide domain, recently named Amino Recognition Domain (ARD), encoded on the 5′ side of the crossover dictated recognition of the trinucleotide component of the target sequence, and a second domain, recently named Carboxyl Recognition Domain (CRD), encoded on the 3′ side of the crossover governed recognition of the tetra- or penta-nucleotide component [26]. Further the trinucleotide components of the recognition sequence of *Eco*KI, *Sty*SPI and

*Sty*SQ are identical, all recognized 5′ AAC, and the 5′ segment of their *hsd*S genes were identical or nearly so, rather than variable [26]. Minor differences within the central conserved sequence of one recombinant S polypeptide were removed so that the amino variable region derived from *Sty*SP, and the rest of the protein derived from *Sty*SB [25]. Its specificity confirmed that the ARD determined the 5′ component of the recognition sequence and, by inference, that the CRD determined the 3′ component [25].

In the case of the type IA enzymes the ARD recognizes the trinucleotide component of the recognition site while the carboxyl recognition domain (CRD) recognizes the tetranucleotide component [25,26,29]. In addition to the variable regions, the type IA S subunits have two relatively short regions which are well conserved within the family [30,31]. The arrangement of conserved and variable regions within the type IB and IC families is shown in Figure 4 for comparison with type IA. For *Eco*KI there are two M subunits per S subunit hence each M subunit may associate with each specificity domain of S subunit [13]. The repeat sequences of S polypeptides found in the conserved regions of type IA, IB and IC [27,32,33] have been implicated in interactions with the M subunits [34-36].

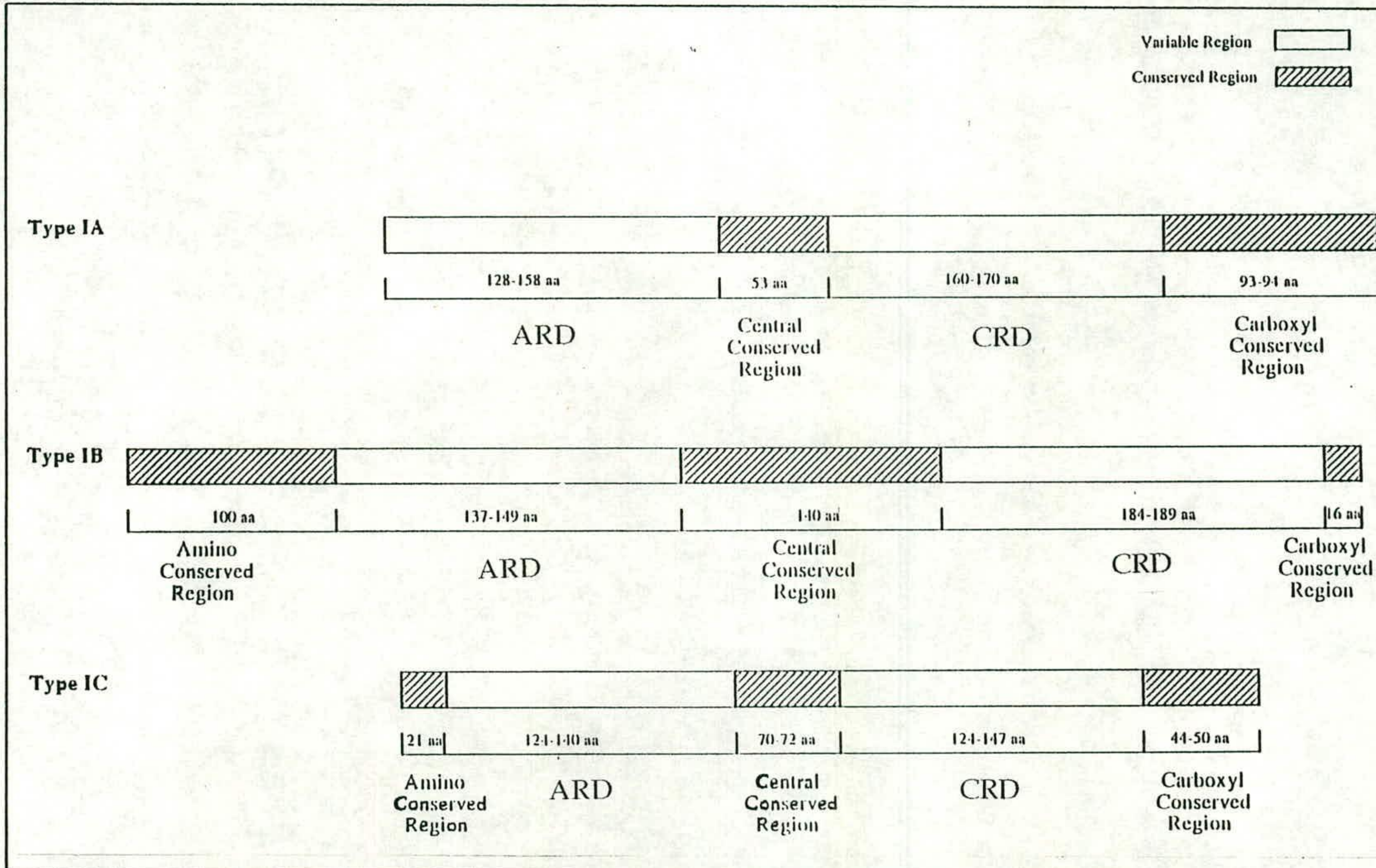


Figure 4. The organization of the S subunits of the three type I R-M system families. The regions conserved within a family are shown as shaded boxes while the ARD and CRD which recognize each part of the DNA target are shown as open boxes.

1.4 *DNA-protein interactions and oligonucleotide analogues*

The majority of the biological functions of DNA are dependent on proteins that recognise and bind to specific sequences of bases in the double helical structure. This effect is possible as the edges of each base pair in duplex DNA expose a unique array of hydrogen bonding groups (Figure 5) either in the major or minor groove [24]. Many proteins that usually bind long DNA duplexes also show similar interactions with short synthetic oligonucleotides providing the recognition sequence is present. The incorporation of modified bases into the recognition sequence is often useful for the study of the recognition process. We were particularly interested in photoreactive analogues of thymidine for ultraviolet (UV) crosslinking.

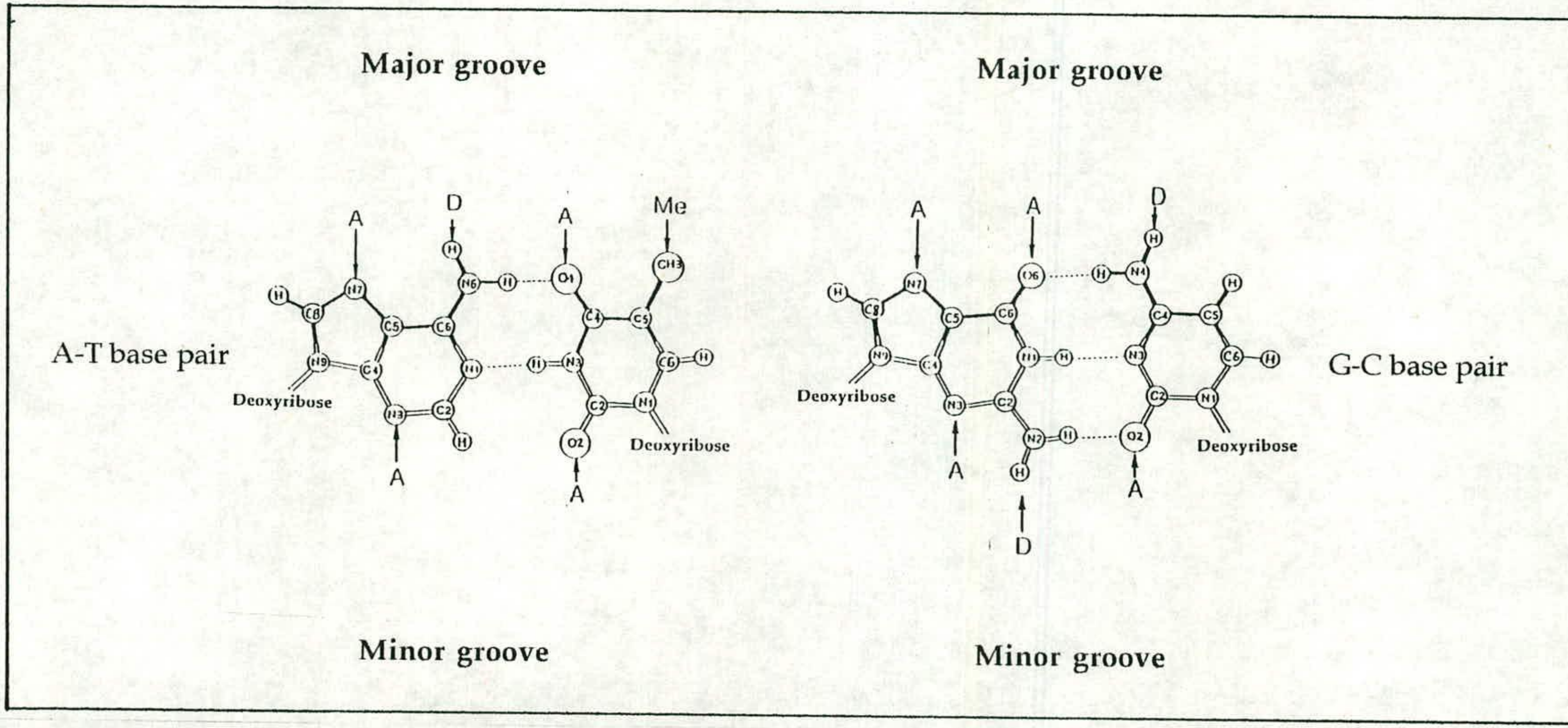


Figure 5. DNA hydrogen-bonding groups available for interactions with proteins
A = hydrogen acceptor   D = hydrogen donor

1.4.1 Photoreactive analogues of thymidine

With thymidine (T) the groups capable of interaction with proteins are show in Figure 5. Thus when T is incorporated into DNA the 4-keto oxygen atom and 5-methyl group (both in the major groove) and the 2-keto oxygen atom (in the minor groove) are capable of interaction with the protein. In 4-thiothymidine (T^{4S}, Figure 6), the thymidine analogue, the 4-keto oxygen atom of thymidine is changed to a 4-sulphur atom. In 5-Bromodeoxyuridine (BrdU, Figure 6), 5-iododeoxyuridine (IodU), and deoxyuridine (dU), the 5-methyl group of thymidine is changed to a bromine, iodine or hydrogen atom respectively. Thymidine analogues can form hydrogen bonds with appropriate amino acid side chains in proteins. Although the change of the 4-keto oxygen atom to a 4-sulphur atom in 4-thiothymidine should weaken hydrogen bonding capacity to proteins, 4-thiothymidine base pairing with adenine in DNA is still possible [37]. BrdU and thymidine are structurally very similar. In BrdU, a bromine atom substitutes the 5-methyl group of thymidine, and the 5-substituents differ little in size with van der Waals radii of the bromine atom and the methyl group at 1.95 Å and 2.00 Å respectively (Figure 7). IodU and dU provide further opportunities to probe the roles of the 5-methyl group of thymidine in the interaction with proteins. However, IodU (Figure 8) has a very much larger van der Waals radius (2.15 Å) than the methyl group of thymidine, and the hydrogen atom of dU (Figure 9) has a very much smaller van der Waals radius of hydrogen atom than methyl group of thymidine.

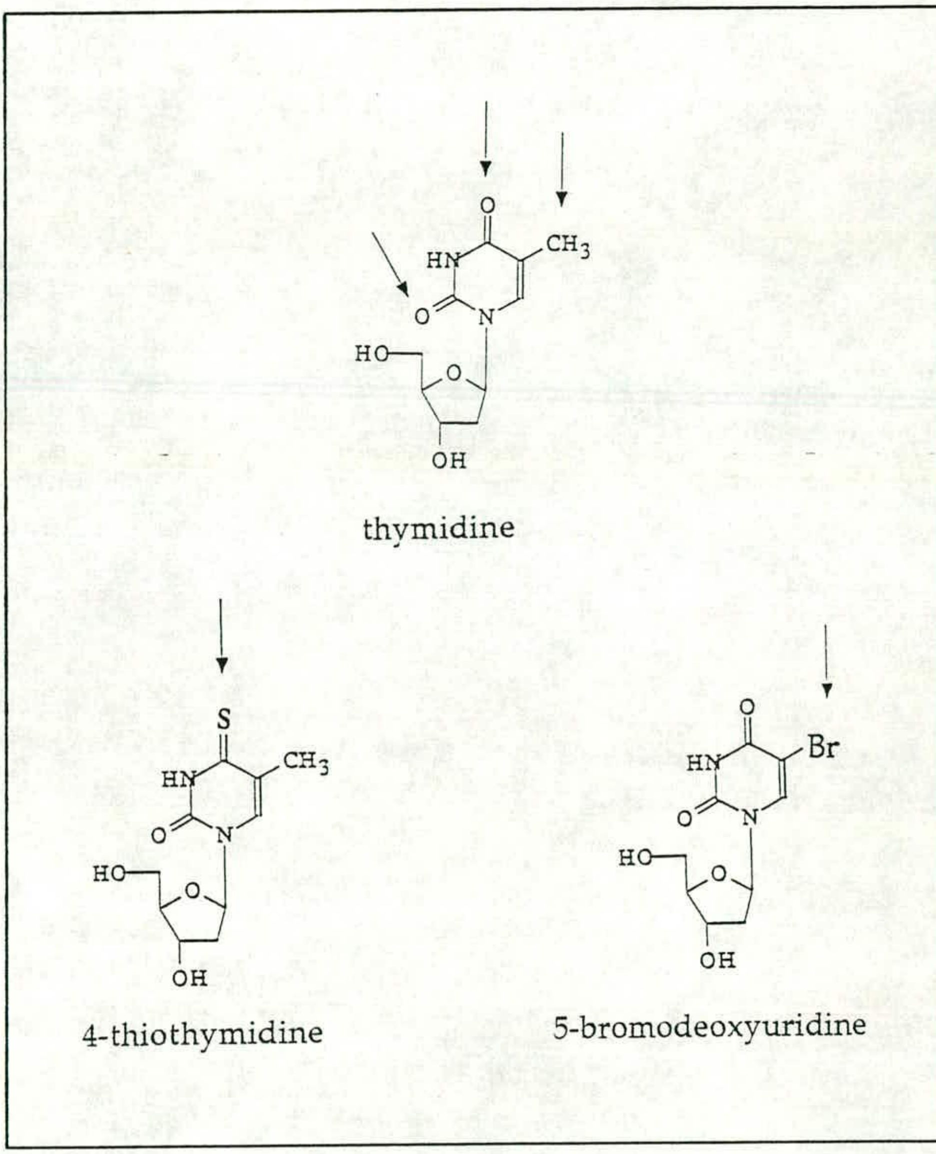


Figure 6. The structures of thymidine, 4-thiothymidine and 5-bromodeoxyuridine.
The 4-keto oxygen, 5-methyl group of major groove and the 2-keto oxygen of minor groove are capable of interacting with proteins when the base is incorporated into double stranded DNA.

Figure 7. Models of the *Eco*KI trinucleotide (-AAC-) recognition sequence duplex. The purple model is unsubstituted duplex with thymidine at the first thymidine of the -AAC- recognition sequence duplex highlighted. (specially coloured balls: carbon= white, hydrogen= sky blue, nitrogen= navy blue, oxygen= red and phosphorus= yellow). The yellow model is bromine (green ball) substitute for the 5-methyl group of thymidine at the first thymidine of -AAC- recognition sequence duplex to form 5-bromodeoxyuridine substitution. The van der Waals radius of the bromine atom is 1.95 Å compared to 2.00 Å for the methyl group.

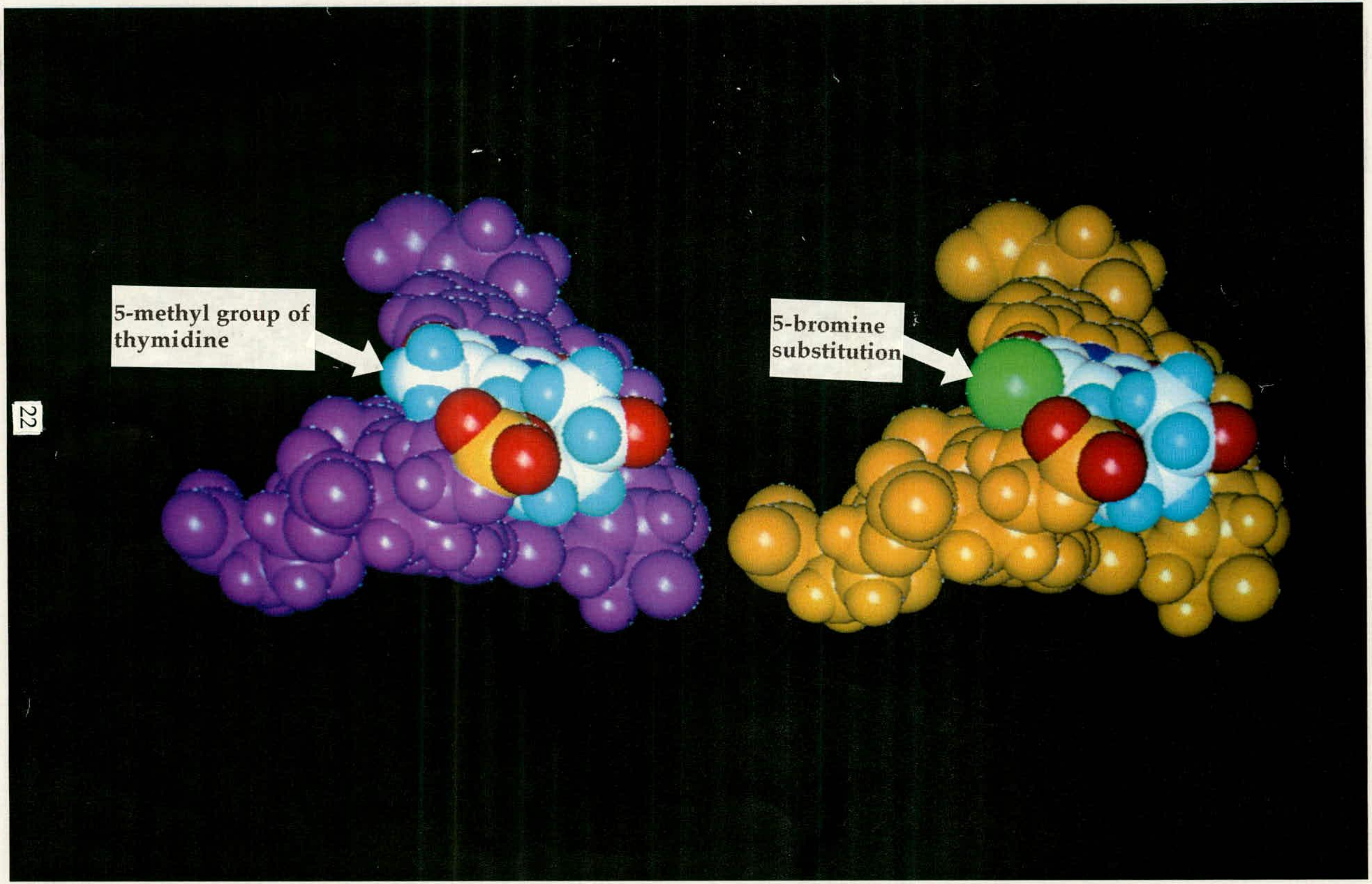

5-methyl group of thymidine
5-bromine substitution

Figure 8. Models of the *Eco*KI trinucleotide (-AAC-) recognition sequence duplex. The purple model is unsubstituted duplex with thymidine at the first thymidine of the -AAC- recognition sequence duplex highlighted. (specially coloured balls: carbon= white, hydrogen= sky blue, nitrogen= navy blue, oxygen= red and phosphorus= yellow). The yellow model is iodine (purple ball) substitutes for the 5-methyl group of thymidine at the first thymidine of -AAC- recognition sequence duplex to form 5-iododeoxyuridine substitution. The van der Waals radius of the iodine atom is 2.15 Å compared to 2.0 Å for the methyl group.

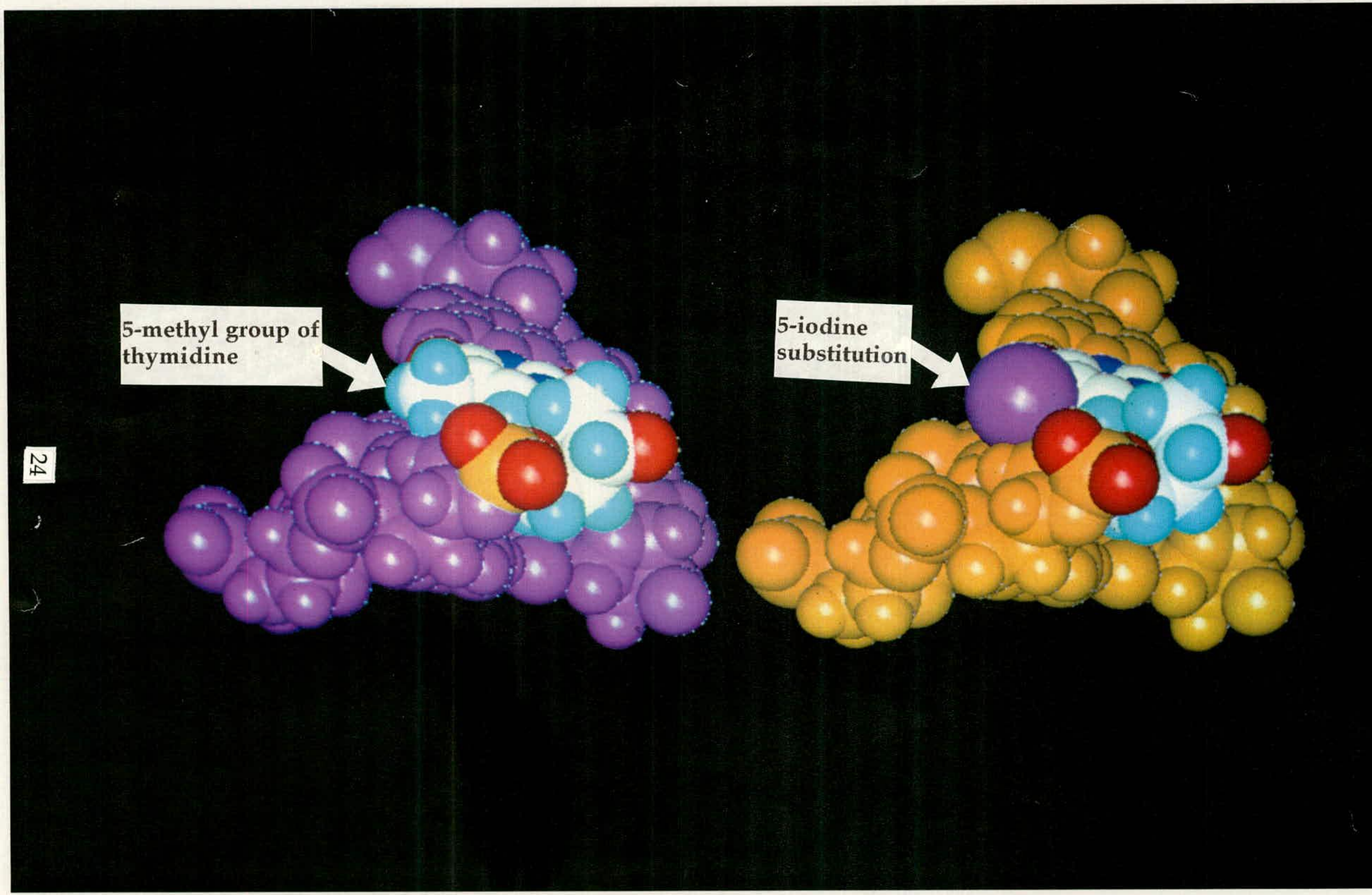
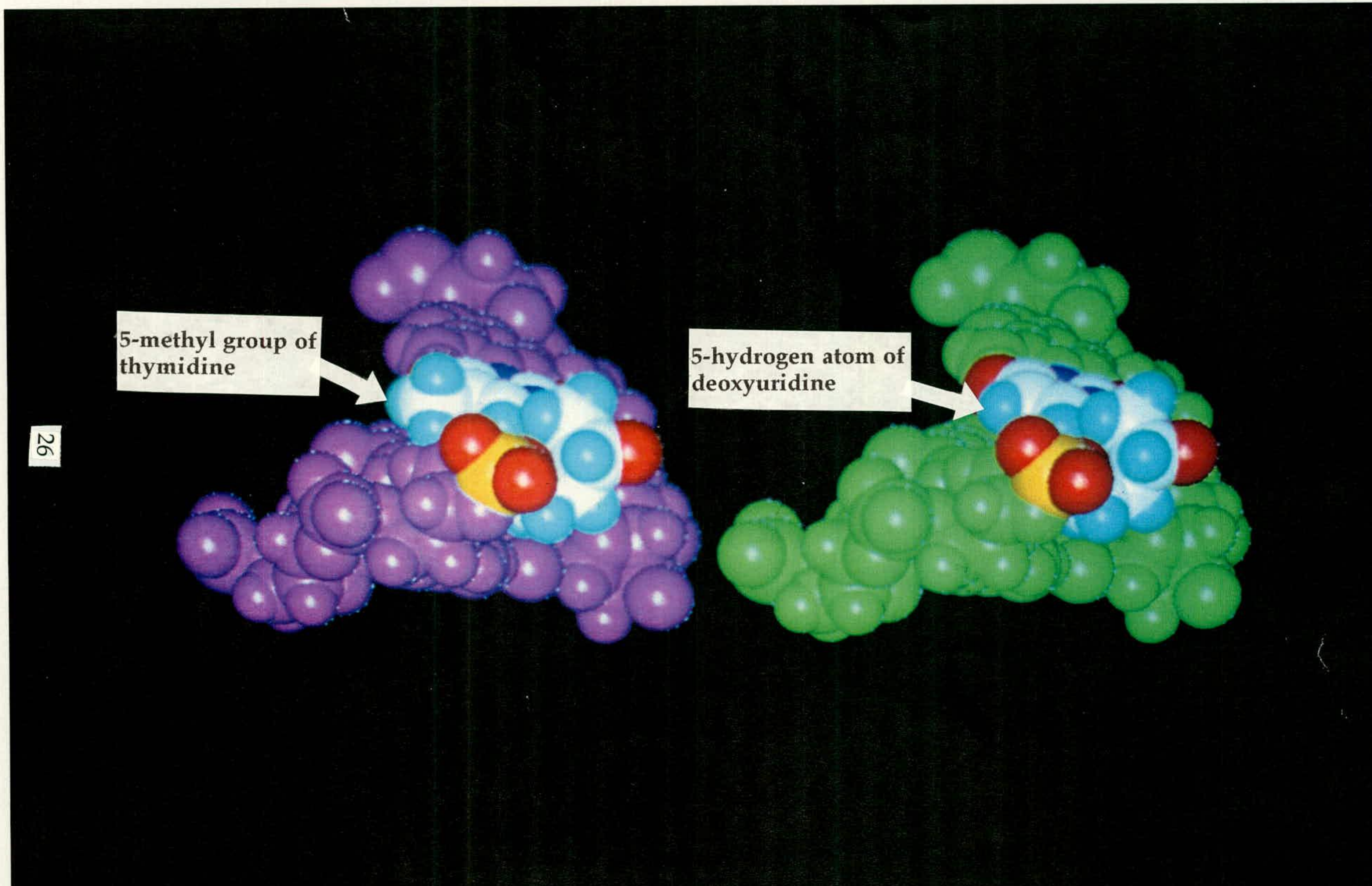

5-methyl group of thymidine
5-iodine substitution

Figure 9. Models of the *Eco*KI trinucleotide (-AAC-) recognition sequence duplex. The purple model is unsubstituted duplex with thymidine at the first thymidine of the -AAC- recognition sequence duplex highlighted. (specially coloured balls: carbon= white, hydrogen= sky blue, nitrogen= navy blue, oxygen= red and phosphorus= yellow). The green model is hydrogen (sky blue ball) substitutes for the 5-methyl group of thymidine at the first thymidine of -AAC- recognition sequence duplex to form 5-deoxyuridine substitution.


5-methyl group of thymidine
5-hydrogen atom of deoxyuridine

1.4.2 Photochemistry of BrdU and IodU

BrdU and IodU are photoreactive analogues of thymidine where the 5-methyl group is substituted by a bromine or iodine atom respectively (see ref. 38 for a review). An understanding of the photochemistry of uracils substituted in 5 position is complicated by the different excited electronic states which can be involved. The relative position of these states, the rates of crossing from one to another, possibly the state of lowest energy, will depend on the polarity of the medium, tautomeric form of the base and degree of protonation, as well as the substituent at the 5 position. Thus the nature and yield of the photochemical event will also depend on these same parameters [38]. (BrdU has an absorption maximum at a wavelength of 280nm and IodU has an absorption maximum at 287.5nm [39], which are at a longer wavelength than any base in normal DNA. It has been reported that there are no significant differences between the absorption of *E. coli* DNA and that with BrdU substituted for 22% of the thymidines [40]. The nucleoside IodU has a very similar photochemistry to BrdU. It is somewhat more difficult to incorporate IodU into DNA, and cells or viruses containing this DNA seem to behave less normally [38]. It is presumed that the problem with IodU lies with the large van der Waals radius of the iodine atom, 2.15Å, compared to 2.0Å for the methyl group and 1.95Å for the bromine atom [38].

1.4.3 UV irradiated crosslinking of BrdU substituted DNA to protein

In the dark, BrdU is unreactive but irradiation with UV light results in photodissociation of bromine with the generation of a reactive free radical at the C5 of uracil. In B-form DNA, the substituent at this position is in close proximity to the C2′ of the deoxyribose moiety of the nucleoside immediately 5′ and in the same strand. In the absence of any other reactive species, the free radical presumably abstracts a hydrogen atom from the sugar leading to decomposition of that sugar (by an unknown mechanism) and a single strand break in the DNA at the position of BrdU incorporation. A properly positioned substituent within a bound protein could donate a hydrogen to the uracil or deoxyribose radical or accept the free radical and directly crosslink to DNA, thereby avoiding strand scission [41]. Most sequence-specific DNA binding proteins cannot distinguish between BrdU and thymidine [42-46]. BrdU can be introduced at any position(s) within synthetic oligodeoxyribonucleotides using 5′-dimethoxytrityl-5-bromo-2′-deoxyuridine-β-cyanoethylphosphoramidite [45, 46].

1.4.4 Conformational variability of DNA

The conformation adopted by DNA depends both on the local sequence and on the local environment of the DNA. In general a DNA molecule is considered to possess two types of conformational flexibility. The torsional flexibility, in which the twist angle either increases or decreases corresponding respectively to a winding or unwinding of the double helix, and the axial flexibility, in which the direction of the helical axis changes [41].

The width of the minor groove in B-DNA varies depending on base composition. When the DNA is highly rich in G-C base-pairs, the stacking forces between neighbouring base pairs and the hydrogen bonds between the individual bases in a single base pair makes the twist of the DNA decrease, which leads to the minor groove becoming wider and shallower while the major groove becomes narrower and deeper resembling that found in A-DNA. While for DNA highly enriched in A-T base-pairs, the twisting increases. Overwinding (>36^{o} per base step) of DNA leads to more overlapping and less stacking of DNA. The hydration of A-T base-pair rich DNA occurs in both the minor and major grooves where it binds one or two water molecules more than G-C base-pair rich DNA. On the A-T base-pair, the hydrophobic thymine methyl group is greatly exposed, and therefore promotes water-water aggregations, resulting in a wider major groove. The degree of hydration of DNA plays a key role in its conformation. High relative humidity favours the B-form and reduced humidity or increased ionic strength leads to a transition from the B-form to C-, A-, and if sequence permits, D-, and Z- DNA. In B-DNA the major groove is wider and shallower and the minor groove is narrower and deeper compared to the A form [47].

The 3D structure of double-stranded DNA is highly polymorphic [48]. Differences in DNA sequence cause differences in the width and depth of the minor and major grooves and this variation will thus cause an external structural irregularity which can serve to distinguish different regions of DNA. The specific DNA sequences may provide a particular structure required for protein to bind at lower free energy cost than other sequences. In B-DNA the major groove is wide enough to accommodate an α –helix and the functional groups of base pairs can be contacted by proteins. In addition, the stacking interactions between the aromatic amino acid side chains such as, Tryptophan (Trp), Tyrosine (Tyr), Phenylalanine (Phe), Histidine (His), and base pairs may occur especially when the sequence is highly rich in A-T base pairs. The minor groove of B-DNA is deep and narrow (5.7Å), and thus less accessible to protein secondary structures such as α –helices. In addition it becomes even narrower (3.5Å) when the sequence is highly rich in A-T base pairs.

1.4.5 Sequence-specific nucleic acid recognition

Structural, biochemical and molecular-genetic studies of DNA-protein complexes have established two important sources of sequence specificity in DNA-protein interactions [49].

(I). Direct hydrogen bonding and van der Waals interactions between protein side chains and the exposed edges of DNA.

When DNA and protein approach closely, hydrogen bonding occurs between phosphate, sugar, bases in DNA and peptide bonds, hydrophilic, polar and neutral amino acid side chains (e.g. cysteine, serine, threonine, asparagine, glutamine and tyrosine) in protein. At the same time salt bridges can be formed between the negatively charged phosphates of DNA and the positively charged amino acid side chains (Lys and Arg) which have often been observed at the surface of protein molecules. The crystal structure studies of many sequence-specific DNA binding proteins, which have different functions, show that there does not appear to be a code for nucleic acid sequence recognition [50]. However, there are certain structural motifs which are common to certain groups of DNA binding proteins. While particular amino acid side chains do not always recognize the same base pair, there are some apparent preferences as suggested by Seeman [51]. For example, the guanidinium group of arginine often makes a bidentate interaction with the N7 and O6 of guanine for example in *Eco*RI, Trp repressor, λCro repressor and catabolite gene activator protein (CAP) [52-55]. However, it is also observed to interact with the N7s of two adjacent adenines in *Eco*RI [52]. The simultaneous interaction of amino acid side chains with two adjacent base pairs greatly enhances their suitability for sequence-specific recognition [51]. The van der Waals interactions between the protein and the 5-methyl group of thymidine appear also to contribute to specificity. In addition the hydrophobic interaction could associate thymine bases with non-polar amino acid side chains, such as Glycine (Gly), Alanine (Ala), Valine (Val), Isoleucine (Ile) and Leucine (Leu).

(II). The sequence-dependent bendability of DNA.

The direct evidence for the role of DNA distortion was obtained only when co-crystal structures of complexes, such as *Eco*RI [52,56], 434 repressor [57], Trp repressor [53] and DNase-I [58], were determined. The distortions of duplex DNA structure that have been observed in complexes include changes in twist [57], groove width [58] and kinks [52,53,57]. Recently, model building [59], the determination of the co-crystal structure as well as DNA binding experiments [55,56] established an important role for DNA kinking and bendability in the sequence-specific binding of CAP to DNA.

1.4.6- The interaction between DNA and R-M systems studied by X-ray crystallography

Crystal structures of several type II ENases have been reported, including the apo-form of *Bacillus amyloliquefaciens* HI (*Bam*HI) [60,61] and *Proteus vulgaris* II (*Pvu*II) [62], and complexes of *Eco*RI, *Eco*RV and *Pvu*II with their cognate DNAs [63-65]. Recently the structures of two type II mtases, *Haemophilus haemolyticus* I (*Hha*I) and *Thermus aquaticus* I (*Taq*I), have been reported [66,67]. The structure for *Hha*I was a co-crystal structure in the presence of protein, DNA and the AdoMet [68], while that for *Taq*I was for protein in the presence of AdoMet but not DNA [67].

*Hha*I mtase is part of a type II R-M system, which catalyses the modification of DNA by methylation of the first cytosine in its GCGC target sequence at the C5 position. Methylation of the unsubstituted DNA results in C5-methylcytosine at the target base, whereas methylation of the 5-fluoro-substituted target base results in a chemically trapped reaction intermediate [69]. The co-crystal structure of the chemically trapped reaction intermediate is a ternary complex between the *Hha*I DNA C5-mtase, S-adenosyl-L-homocysteine (the product form of the enzyme's cofactor), and a duplex 13 base pair DNA fragment containing methylated 5-fluorocytosine at its target. The DNA is covalently linked to the enzyme's catalytic cysteine residue (Cys 81) via the C6 position of the 5-fluoro-substituted target base which has been methylated at the C5 position. The structure illustrated that the single target base was completely swung out of the double helix. The covalently trapped reaction intermediate of mtase and 5-fluoro modified DNA confirmed the mechanism for cytosine methyl-transfer reactions first proposed by Wu and Santi [59] that cytosine-5-mtase functions catalytically in a manner analogous to thymidylate synthase and tRNA (uracil-5) mtase enzymes that also transfer single methyl groups from AdoMet to the C5 position of pyrimidines. This is another type of protein-induced DNA distortion, in which the target base pair is disrupted with a single base flipped-out of the double helix and located next to the enzyme's cofactor S-adenosyl-L-homocysteine [69].

The M**Taq*I DNA mtase methylates N_6 of adenine in its -TCGA-target sequence by transfer of a methyl group from the cofactor AdoMet. The crystal structure of M**Taq*I, and AdoMet shows that the three dimensional folding of M**Taq*I is similar to that of the cytosine specific *Hha*I mtase where the large β-sheet in the N-terminal domain contains all of the conserved segments typical for N6-mtase to form the cofactor binding pocket and the enzymatically functional parts, and the smaller C-terminal domain is less structured [67]. It is believed that the small domains of M* *Taq*I and *Hha*I Mtase are mainly responsible for DNA specificity recognition. The crystal structure of *Hha*I shows that two glycine-rich fragments of the small domain are directly in contact with the DNA recognition sequence in the major groove. Some loops of the M* *Taq*I in C terminal domaincontain positively charged amino acids Arg, Lys and His which are in suitable positions to interact with DNA phosphate groups[131].

The crystal structure of M* *Taq*I is the only N6 adenine specific DNA mtase which has been reported and all known type I R-M systems are N6 adenine specific DNA mtases. The similarity and the equivalence of amino acids revealed by structural alignment between M* *Taq*I and *Hha*I indicate that the AdoMet dependent mtase may share a common catalytic-domain structure. Recently, an amino acid alignment of the M polypeptide of type I mtases, M* *Taq*I and other type II N6 mtases shows the prescence of a common catalytic-domain structure [128].

1.4.7 Studies of the interaction between DNA and protein by UV-induced crosslinking.

UV-induced crosslinking has been widely used to investigate DNA-protein interactions since the 1970s. Many studies have simply exploited the intrinsic photoreactivity of the nucleotides, in particular thymine [70-73]. Other approaches have relied on nucleotide analogues with enhanced photoreactivity [74-83]. For example, crosslinking of plasmid pBR322 containing 8-azidoadenine-2'-deoxyribonucleotides to histone H1 by UV irradiation with a peak wavelength of 366-nm [75] and crosslinking of the *Eco*RV restriction ENase and modification mtase to a 4-thiothymidine substituted oligonucleotide containing the recognition site of these enzymes by UV irradiation at 340-360 nm [77] have been reported. UV irradiated crosslinking of BrdU substituted DNA to protein has been observed for many DNA-protein interactions [41,68,79-82,84]. Some examples include, the specific crosslinking of *lac* repressor protein to BrdU substituted *lac* operator at 254nm [41,68,79,81], the crosslinking between *E.coli* RNA polymerase and the BrdU substituted *lac* UV5 promoter at 254 nm [84], the bZIP DNA-binding protein-GCN4 crosslinked to BrdU substituted oligonucleotide at 254 nm [82].

1.5 Peptide sequencing

Automated protein-sequencing technology has considerably improved in recent years. The gas-phase sequenator and the use of on-line HPLC systems employing micropore columns for phenyl thiohydantoin (PTH) amino acid detection enables sequence data to be obtained with less than 10 pmol of protein or peptide [overview see 85]. In many cases, microsequencing projects are not limited by the sensitivity of the Edman degradation itself but by our ability to obtain protein/peptide samples in a form that is suitable for sequence analysis. Most protein purification protocols utilize a combination of selective precipitation, ion-exchange chromatography, gel-filtration chromatography, affinity chromatography, dialysis, and concentration steps in purifying a native and functional protein. The yields may be measured by antibody-enzyme assays, functional assays, and silver stain SDS-PAGE/immunoblots [86]. The sample preparation cartridge ProSpin™ (Applied Biosystems) has added another success to provide a form of suitable sample attachment into the hydrophobic Polyvinylidene fluoride (PVDF) membrane that can be directly submitted for sequence analysis.

1.6 *Aims*

The main difficulty of studying the interaction between a type I mtase and DNA by X-ray crystallography is to crystallise the two associated macromolecules as a complex. It is particularly problematic when the functioning mtase contains three polypeptides. In the absence of structural information on the amino acids involved in DNA recognition in type I R-M systems, the interactions between them may be investigated using 4-thiothymidine or 5-bromouracil substituted DNA [68,79-83]. UV irradiation can then be carried out to form the crosslinked DNA-protein complex and sequencing the digested crosslinked DNA-peptide complexes can be used for the identification of the crosslinked amino acid and base. The substituted DNAs are important in this investigation to allow the specific binding sites of *Eco*KI mtase with DNA to be identified without ambiguity. Therefore this method was used to study the binding of the TRDs of S polypeptides with DNA.

Chapter 2 Materials and Method

Standard laboratory reagents

BDH Chemicals Ltd, Poole, U.K.

Fisons Chemicals, Loughborough, U.K.

Sigma Chemical Co., Poole, U.K.

Aldrich Chemical Company Ltd, Dorset, U.K.

Boehringer Mannheim

Instruments

IR was carried out on a 197 infrared spectrophotometer.

^{1}H NMR spectra were recorded with a Bruker WP-200 (200 MHz). spectrometer, using $CDCl_3$ or DMSO as solvent.

F.a.b. (fast atom bombardment) mass spectra were recorded on a Kratos MS50 TC spectrometer.

UV spectra were obtained on a Lambda 15 spectrophotometer.

Applied Biosystems 380B DNA Synthesizer.

High Performance Liquid Chromatography (HPLC), Gilson.

UV handlamp set, Model UVGL-58, UVP International, Ultra-Violet Products, Ltd.

Edwards EF4 Modulo freeze dryer from Edwards High Vacuum.

2.1 General synthetic procedure on an ABI 380B DNA synthesizer

The general procedure for phosphoramidite oligonucleotide synthesis on an ABI 380B DNA synthesizer is as follows [87]:

1. Carefully check all the reagent levels on the synthesizer and the pressure of Argon gas connected to the synthesizer, and change reagent bottles and argon cylinder if necessary.
2. Screw clean collection vials onto the synthesiser and label each vial with a code number for the relevant DNA sequence.
3. Type the sequences (from 3'-terminal) to be prepared into the synthesiser.
4. Switch to the column monitor, clear the set-up for each individual synthesis column, and load in the new sequences to be synthesized.
5. Check that the choice of synthesis cycle is correct (sscel03A for 0.2 μmol; cel03A for 1 μmol).
6. Select the ending method (Trityl Off Auto will normally be the system default).
7. Select deprce03 ending method. It is also the system default.
8. Type User Name on each column monitor.
9. Print out the details of each column set-up.
10. Carefully check that all details outlined on the printout are correct.
11. Run through the Start Column procedure for each column position.
12. Check that the solvent residue (waste) bottle connected to the synthesizer is not full.
13. If the effluent from the detritylation step at each deprotection is to be collected by the fraction collector, ensure that sufficient clean dry test-tubes are present and the trityl waste line leads to the fraction collector.
14. Start the cycle, always running a "begin" procedure to purge the reagent lines and remove old reagents.

2.1.1 Determination of DNA concentration

DNA concentrations were determined spectrophotometrically using the absorbance at 254 nm. Micromolar extinction coefficients (ε) for the single stranded oligonucleotide (ss DNA) and double stranded oligonucleotides (ds DNA) are calculated from the following formula:

ε_{254} (ss DNA) = [15.4 x nA + 7.3 x nC + 11.7 x nG + 8.8 x nT] x 0.9
ε_{254} (ds DNA) = [15.4 x nA + 7.3 x nC + 11.7 x nG + 8.8 x nT] x 0.8

ε_{254} of *Eco*KI target recognition 25 mers was calculated as follows:

Top strand — 5′- C G G C C T A A C C A C G T G G T G C G T A C G A
nA = 5, nC = 8, nG = 8, nT = 4
ε_{254} = [(15.4 x 5) + (7.3 x 8) + (11.7 x 8) +(8.8 x 4)] x 0.9
= 237.78

Bottom strand — 3′- G C C G G A T T G G T G C A C C A C G C A T G C T
nA = 4, nC = 8, nG = 8, nT = 5
ε_{254} = [(15.4 x 4) + (7.3 x 8) + (11.7 x 8) + (8.8 x 5)] x 0.9
= 231.84

25 mers duplex — nA = 9, nC = 16, nG = 16, nT = 9
ε_{254} = [(15.4 x 9) + (7.3 x 16) + (11.7 x 16) + (8.8 x 9)] x 0.8
= 417.44

Molecular weight of DNA (single strand) is calculated as follows :

{(249 x nA) + (225 x nC) + (265 x nG) + (240 x nT)+ PO_2 x (n-1) + H_2O x 3n + NH_3 x (n-1) + 2}

nA = number of adenine base in DNA sequence etc.
n = number of total bases
PO_2 x (n-1) : accounts for the molecular weight of the phosphate groups
H_2O x 3n : accounts for the dyhration of DNA, approximately three water molecules per nucleotide as a rule of thumb.
NH_3 x (n-1) : accounts for the ammonium cations associated with the phosphates groups.

μM wt. of top strand 25 mers = 249 x 5 + 225 x 8 + 265 x 8 + 240 x 4 + 63 x 24 + 18 x 25 x 3 + 17 x 24 + 2 = 9.397 (mg)

Top strand 25 mers	μM wt. = 9.397 mg ε_{254} = 237.78 OD units 1.0 mg = 25.3 OD 1 OD = 39.5 μg
Bottom strand 25 mers	μM wt. = 9.388 mg ε_{254} = 240.39 OD units 1.0 mg = 25.6 OD 1 OD = 39.1 μg
25 mers duplex	μM wt. = 18.785 mg ε_{254} = 417.44 OD units 1.0 mg = 22.2 OD 1 OD = 45.0 μg

2.1.2 Reagents for phosphoramidite DNA synthesis

Base-protected 5′-dimethoxytrityl, A, G, T, C phosphoramidite monomers, tetrazole coupling catalyst, acetic anhydride and N-methyl imidazole capping reagents, trichloroacetic acid deprotection solution, iodine oxidation mixture, acetonitrile wash solvent, aqueous ammonia cleavage solution, all obtained from Applied Biosystems, Cruachem Ltd, or alternative suppliers.

Thymidine substituted analogues, 5′-(4,4′-dimethoxytrityl) 5′-bromo-2′-deoxyuridine-3′-N,N-diisopropyl cyanoethyl phosphoramidite, 5′-(4,4′-dimethoxytrityl) 5′-iodo-2′-deoxyuridine-3′-N,N-diisopropyl cyanoethyl phosphoramidite and 5′-(4,4′-dimethoxytrityl) 2′-deoxyuridine-3′-N,N-diisopropyl cyanoethyl phosphoramidite were obtained from Link Technologies (UK).

Argon gas was supplied by British Oxygen.

Gel filtration Sephadex G-25 N.A.P. (nucleic acid purification) columns were supplied by Pharmacia.

2.2 *Synthesis of 4-thiothymidine substituted oligonucleotides*

2.2.1 Reagents for synthesis of 4-nitrophenylthio-thymidine:

Analytical grade, obtained from Aldrich unless otherwise stated.
Alcohol-free anhydrous dichloromethane (DCM) was prepared by distillation from calcium hydride.
Anhydrous pyridine (C_5H_5N) was prepared by double distillation from potassium hydroxide (KOH).
Hexane ($CH_3(CH_2)_4CH_3$) was dried over sodium (Na) wire.

2.2.2 Flash-column chromatography

Flash-column chromatography was carried out on Silica Gel 60 (Fluka) in the following solvent systems:
(1) DCM/MeOH/Triethylamine (Et_3N): (95/4.5/0.5)
(2) Ethyl acetate (EtOAc)/DCM: (1/1) + Et_3N (0.1%)

2.2.3 Thin layer chromatography (TLC)

TLC on Silica Gel 60 F_{254} aluminium plates (Merck) used (3) and (4) as solvent systems.
(3) DCM/MeOH: (9/1)
(4) DCM/MeOH: (8/2)

Reagents for each step in Figure 10 are as follows:

(I) thymidine, methoxyacetic anhydride in dry pyridine

(II) and (III) 3',5'-bis-O-(methoxyacetyl)-thymidine, diphenyl phosphorochloridate, 3-nitro-1,2,4-triazole, triethylamine in acetonitrile

(IV) 3',5'-bis-O-(methoxyacetyl)-4-nitrotriazolo-thymidine, DBU [DBU:1,8-diazobicyclo[5.4.1]undec-7-ene(1,1-5)], p-nitrothiophenol, in dry dichloromethane.

(V) 3',5'-bis-O-(methoxyacetyl)-4-nitrophenylthio thymidine, ammonia (20%) in methanol

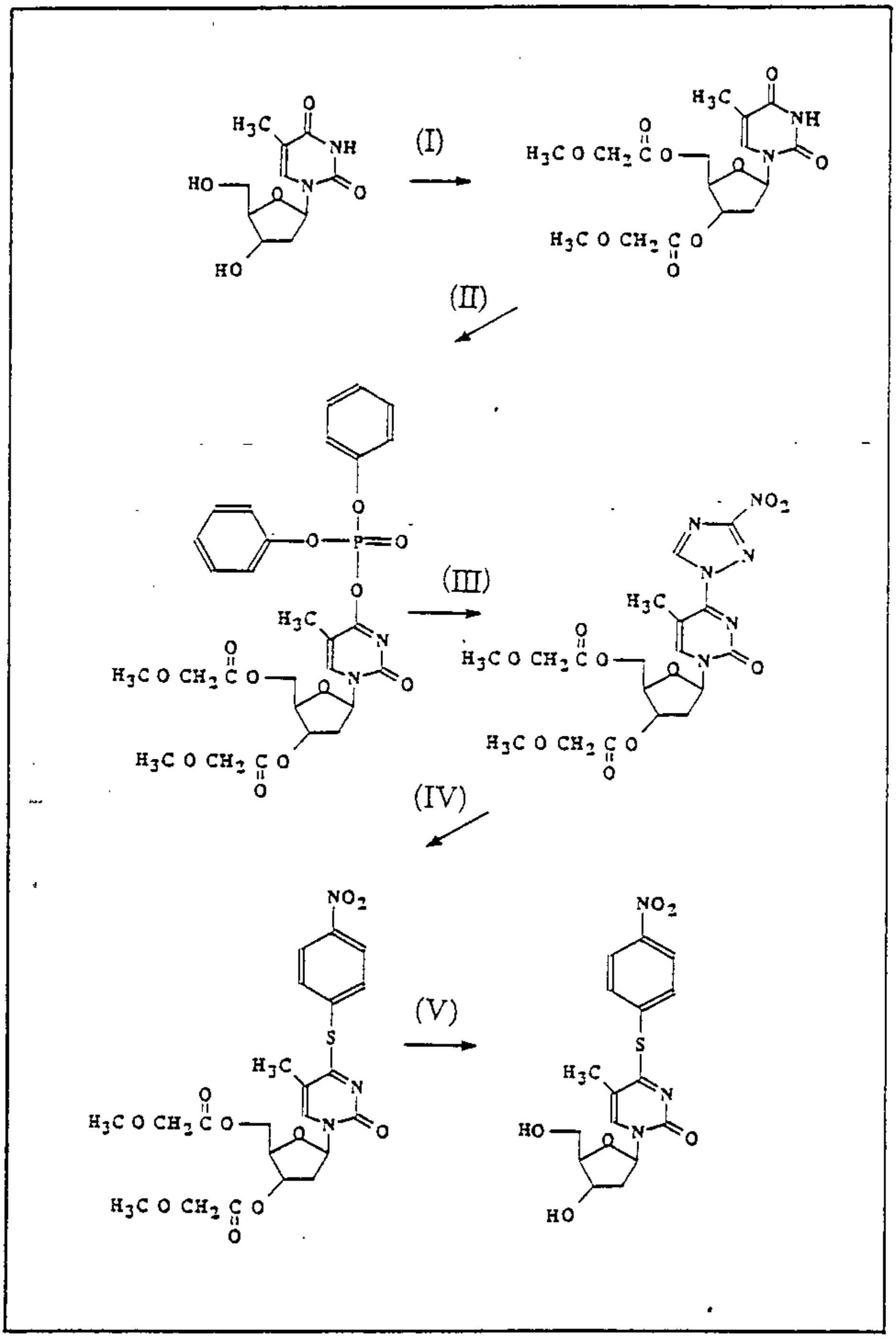


Figure 10. The synthesis of 4-nitrophenylthio thymidine

2.2.4 Synthesis of 4-nitrophenylthio-thymidine phosphoramidite

The method of synthesis of 4-nitrophenylthio-thymidine from thymidine was based on the method of Li *et al* [88] for the synthesis of 4-O-methyl thymidine. The stepwise reactions (Figure 10) were as follows:

2.2.4.1 Preparation of methoxyacetic anhydride

Methoxyacetic acid (20g, 0.22 M) and redistilled acetic anhydride (1.9ml, 0.20 M) were heated together under gentle reflux for 24 hours. The reaction mixture was then fractionally distilled through a glass helices column to give methoxyacetic anhydride. IR : C=O (1760, 1840cm-1).

2.2.4.2 Synthesis of 3', 5'-bis-O-(methoxyacetyl)-thymidine

Methoxyacetic anhydride (51.5 mM, 8.3g) was added to a stirred suspension of anhydrous thymidine (20.6 mM, 5.0g) in anhydrous pyridine (30 ml) at 20°C. Anhydrous thymidine was obtained by pre-extracting traces of water from thymidine with anhydrous pyridine (30 ml x 2) then removing the pyridine under high vacuum. This was done twice and after 30 minutes, water (2 ml) was added to the reaction mixture (which was stirred for 10 minutes) to stop the reaction. The products were then concentrated under reduced pressure and redissolved in anhydrous DCM (100 ml). The resulting solution was extracted with saturated aqueous sodium hydrogen carbonate (40 ml x 2) and washed with potassium chloride (0.1 M, 40 ml). The aqueous layers were back extracted with DCM (40 ml). The combined organic layer was dried with sodium sulphate and evaporated under reduced pressure to give 3',5'-bis-O-(methoxyacetyl) thymidine as a gum. The product was left in a vacuum desiccator overnight before extracted with diethyl ether to give a crystalline powder (6.65 g, 83%). R_f 0.45, m.p. 72°C.

2.2.4.3 Synthesis of 3',5'-bis-O-(methoxyacetyl)-4-O-nitrotriazolo-thymidine

3',5'-bis-O-(methoxyacetyl) thymidine (5.0 mM, 1.9 g) was added to a stirring solution of diphenyl phosphorochloridate (12.5 mM, 3.4 ml), 3-nitro-1,2,4-triazole (12.5 mM, 1.5 g) and TEA (25 mM, 2.5 ml) in acetonitrile (25 ml). The mixture was stirred for 90 minutes, to give the triazole derivative in a crude form [88] which was used without further purification. The reaction was carried out under anhydrous conditions. R_f 0.57.

2.2.4.4 Synthesis of 3',5'-bis-O-(methoxyacetyl)-4-nitrophenylthio-thymidine

Crude 3',5'-bis-O-(methoxyacetyl)-4-nitrotriazolo-thymidine was dissolved in dry DCM (3 ml) then added to a stirring solution of p-nitrothiophenol (1.2 mM, 0.19 g) and DBU [DBU:1,8-diazobicyclo[5.4.1]undec-7-ene(1,1-5)] (1.2 mM, 0.18 ml) in dry DCM (2.0 ml). After 10 minutes potassium phosphate buffer (0.5 M, 10 ml; pH 6.5) was added to the reaction mixture which was then extracted with dichloromethane (15 ml x 2). The dichloromethane layer was then extracted with saturated sodium bicarbonate (30 ml) and potassium chloride (0.1 M, 30 ml x 2), dried with sodium sulphate, filtered under reduced pressure and then evaporated on a rotary evaporator. The product was purified by column chromatography on a silica gel column, under gradient elution with an increasing concentration of methanol in DCM. The purified product was evaporated to a gum on a rotary evaporator. The gum crystallised after being left at room temperature for 6 hours. The reaction scheme is shown in Figure 11. R_f 0.64, 1H-NMR data ($CDCl_3$): d2.13 (m, 5H, H-2', 2"+ CH3-5), 3.15 (s, 6H, MeOAc-CH3 x 2), 4.03 (s, 4H,MeOAc-CH2 x 2), 4.20 (m, 1H, H-4'), 4.3-4.38 (m, 1H, H-3'), 4.45-4.48 (m, 2H, H-5', 5"), 6.1-6.25 (t, 1H, H-1'), 7.6-7.75 (d,3H, NTP-H-5 + H-6), 8.23 (d, 2H, NTP-H-3).

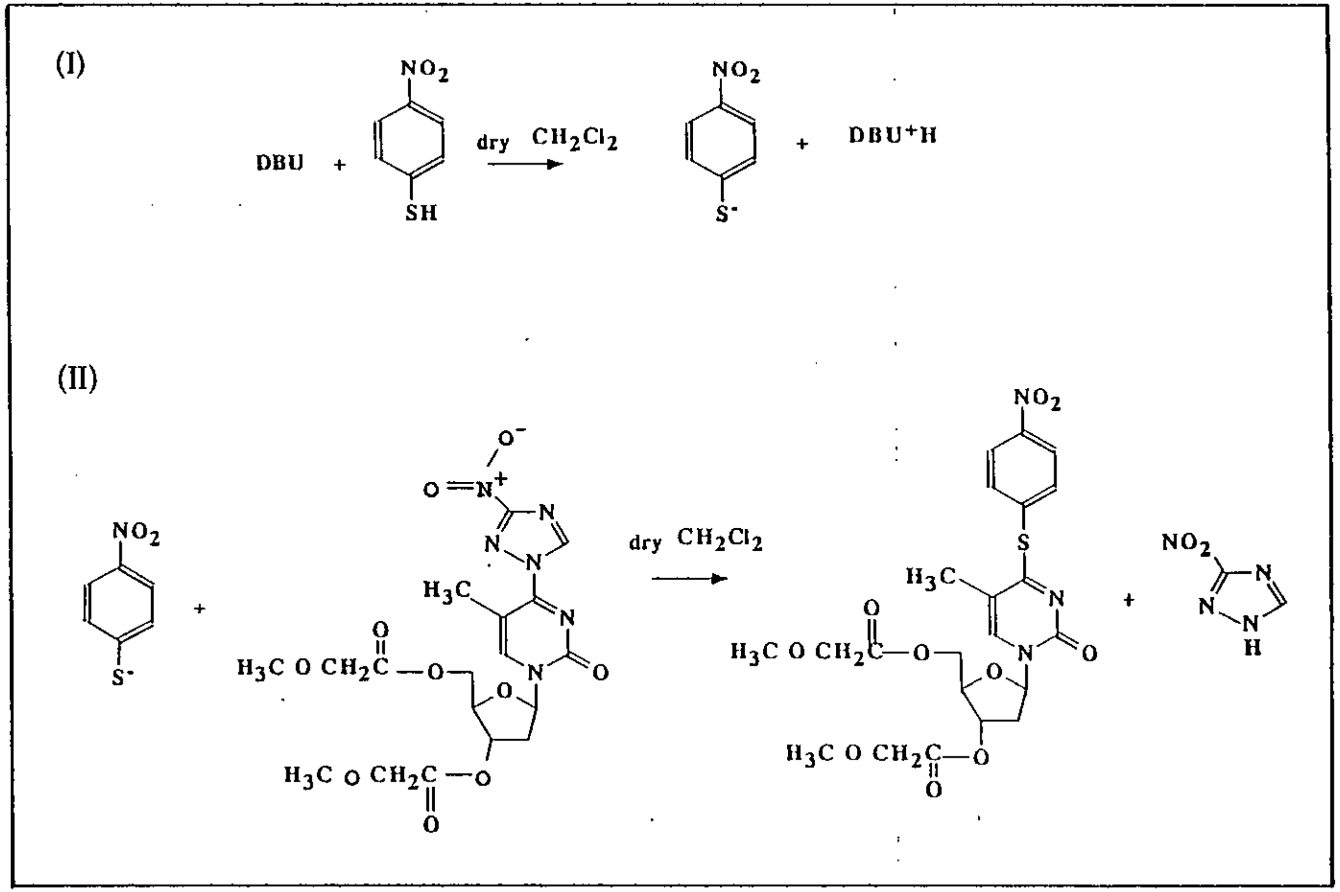


Figure 11. The synthesis of 3′,5′-bis-O-methoxyacetyl-4-nitrophenylthio thymidine.

2.2.4.5 Synthesis of 4-nitrophenylthio-thymidine

Concentrated ammonia (0.25 ml) was added to 3',5'-bis-O-(methoxyacetyl)-4-nitrophenylthio-thymidine (1.73 mM, 0.12 g) which was dissolved in methanol (1.0 ml) by gentle warming. The reaction product, 4-nitrophenylthio-thymidine, was filtered under reduced pressure and washed with diethyl ether and recrystallised from methanol. R_f 0.55, 1H-NMR data (DMSO): d 2.15 (m, 5H, H-2', 2" + CH3-5), 3.6-3.65 (m, 2H, H-5', 5"), 3.8-3.85 (m, 1H,H-4'), 4.18-4.25 (m, 1H, H-3'), 5.25-5.28 (d, 1H, OH-3'), 5.95-6.02 (t, 1H, H-1'), 7.81-7.86 (d, 2H, NTP-H-5), 8.15 (s, 1H, H-6), 8.28-8.33 (d, 2H, NTP-H-3).

2.2.4.6 Synthesis of 5'-(4.4'-dimethoxytrityl)-O-4-nitrophenylthio-thymidine

4-Nitrophenylthio-thymidine (1.73 mM, 1.0 g) was dried twice by evaporation of pyridine (10 ml x 2) under high vacuum. Dry 4-nitrophenylthio-thymidine (1.0 g) was dissolved in anhydrous pyridine (10.0 ml) and then 4,4'-dimethoxytrityl chloride (6.92 mM, 4 eq, 2.3 g) was added to the reaction mixture. The flask was sealed with a septum and flushed with nitrogen. Anhydrous Et_3N (10.38 mM, 6 eq, 1.4 ml) was transferred into the stirring mixture by a syringe. After 30 minutes the product was concentrated on high vacuum then dissolved in EtOAc (70 ml). The solid which formed was filtered off and the liquid was extracted with sodium chloride (2 x 70 ml) followed by (0.1 M) potassium bicarbonate (70 ml). The organic layer was dried with sodium sulphate and concentrated on high vacuum to an oil then purified by flash chromatography using eluant: DCM/MeOH/Et_3N : 95/4.5/0.5. The appropriate fractions pooled and evaporated *in vacuo* to give 5'-(4,4'-dimethoxytrity)-O-4-(nitrothiophenyl) thymidine. Rf 0.53, 1H-NMR data ($CDCl_3$): d2.14 (s, 5H, H-2', 2"+CH3-5), 3.78 (s, 6H, DMT-CH3 x 2), 4.5-4.6 (m, 2H, H-5', 5"), 6.01-6.05 (t, 1H,H-1'), 7.73-7.78 (d, NTP-H-5), 8.18-8.23 (d, NTP-H-3).

2.2.4.7 Synthesis of 5'-(4,4'-dimethoxytrityl)-O-(4-nitrophenylthio) thymidine-3'-(2-cyanoethyl-N, N-diisopropyl)-phosphoramidite

5'-(4,4'-dimethoxytrity)-O-(4-nitrophenylthio) thymidine (1.0 mM, 0.68 g) was dissolved twice in anhydrous DCM (10 ml x 2) and evaporated dry then dissolved in anhydrous DCM (5.0 ml). Diisopropylethylamine (4 mM, 4 eq, 1.0 ml) was added to the stirring mixture at room temperature with the exclusion of moisture, then 2-cyanoethyl N,N-diisopropylchlorophosphoramidite (1.4 mM, 4eq, 0.33 ml) was added dropwise with a syringe. After 40 minutes the product (5'-(4,4'-dimethoxytrityl)-O-4-(nitrophenylthio)-thymidine-3'-(2-cyanoethyl-N, N-diisopropyl)-phosphoramidite) was concentrated *in vacuo,* and purified by flash chromatography. Samples were dissolved in 1.0 ml of eluant (EtOAc/DCM/Et_3N : 100/100/0.1) for loading on a flash chromatography column. Following TLC monitoring, the appropriate fractions were pooled and products were concentrated by evaporating the eluent *in vacuo* . The purified product was dissolved in anhydrous DCM then precipitated into dry hexane (200 ml). After the white product precipitated, the clear supernatant was removed with a pipette and the residue was evaporated *in vacuo* . The dry product was redissolved in anhydrous acetonitrile (2.0 ml) and transferred to a weighed flask then the solvent was again removed by evaporating *in vacuo* to give the powdery product : 5'-(4,4'-dimethoxytrityl)-O-(4-nitrophenylthio) thymidine-3'-(2-cyanoethyl-N,N-diisopropyl)-phosphoramidite.

2.2.5 Synthesis of 4-thiothymidine substituted oligonucleotides

2.2.5.1 Synthesis of test sequence

After dimethoxytritylation and phosphitylation of 4-nitrophenylthio-thymidine to produce a phosphoramidite monomer for DNA synthesis, two test sequences $TTTTTTTTT^{4S}TTT$ ($T_8T^{4S}T_3$) and $CT^{4S}CT^{4S}CC$ were synthesized. Two ending methods, Trityl Off Auto and Trityl Off Manual, were compared on these two test sequences. For Trityl Off Manual ending method, two manual treatment steps were employed as follows [89]:

(1) The columns with oligonucleotide attached to the solid support were plugged between two syringes and treated with [15% v/v] *tert*-butylamine (0.75 ml) in acetonitrile (5.0 ml) for 45 minutes at room temperature then washed with acetonitrile (5.0 ml x 2).

(2) The columns were then treated with potassium thioacetate ($KSCOCH_3$, 0.3 M) in ethanol (5.0 ml) and incubated at 55^{o} C for 2 hours, washed with acetonitrile (5.0 ml x 4), then put back on the automatic synthesiser to continue the automatic deprotection/deblocking step. After the oligonucleotide was cleaved from the solid support, it was deprotected by heating in conc. aqueous ammonia then purified by reversed-phase HPLC.

2.2.5.2 Synthesis of 4-thiothymidine substituted oligonucleotides containing the specific type I *Eco*KI recognition sequence

4-Nitrophenylthio-thymidine phosphoramidites were dissolved in anhydrous acetonitrile to a concentration of 0.1 M. Results of test sequences shown that the manual treatment of *tert*-butylamine and potassium thioacetate was necessary for the incorporation of 4-thiothymidine into synthetic oligonucleotides. Trityl On Manual and Trityl Off Manual ending methods were used for the synthesis of 4-thiothymidine substituted oligonucleotide.

The manual treatment steps employed after the automatic synthesis process were the same as in section 2.2.5.1.- After the oligonucleotide containing 4-thiothymidine was cleaved from the solid support, it was deprotected as usual by heating in conc. aqueous ammonia then purified by reversed-phase HPLC.

Complementary oligonucleotides 25 mers (Figure 1), which were not substituted containing the specific type I *Eco*KI recognition sequence, were refered to as unsubstituted DNA. The complementary oligonucleotide duplex 25 mers was thus termed S25 (specific 25 mers, Figure 12). The oligonucleotide duplex which did not contain the specific type I *Eco*KI recognition sequence was named NS25 (non-specific 25 mers, Figure 12). Any single base of DNA substituted by 4-thiothymidine, BrdU, IodU or dU during synthesis, was referred to a substituted DNA (Figure 22 and Figure 13).

S25

5′ C G G C C T A A C C A C G T G G T G C G T A C G A
3′ G C C G G A T T G G T G C A C C A C G C A T G C T

NS25

5′ C G G C C T C C A C A C G T G T G T A G T A C G A
3′ G C C G G A G G T G T G C A C A C A T C A T G C T

Figure 12. The DNA duplex contains the *Eco*KI mtase recognition sequence (underlined) is referred to as S25 (specific 25 mers). The DNA duplex does not contain the *Eco*KI mtase recognition sequence (underlined) is referred to as NS25 (non-specific 25 mers).

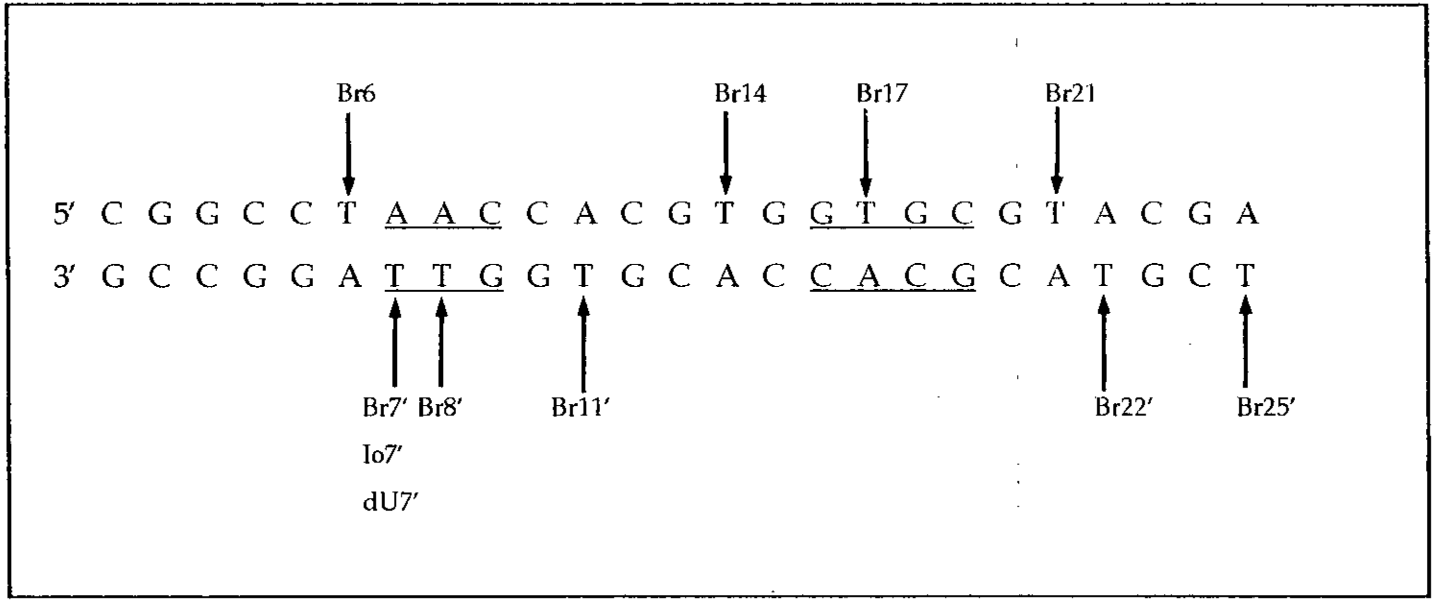


Figure 13. The DNA duplex contains the *Eco*KI target sequence (underlined). This duplex is referred to as S25. The sites of substitution of thymine by modified bases are indicated and numbered from the 5′ end of the top strand with the addition of a prime (′) indicating a substitution on the bottom strand. The duplexes containing a single substitution of BrdU, IodU or dU are designated Br6, Br14, Br17, Br21, Br7′, Br8′, Br11′,Br22′, Br25′, Io7′ and dU7′. The duplex Br500 contains BrdU at all of the indicated sites.

2.2.6 Reversed-phase HPLC purification of oligonucleotides

2.2.6.1 Trityl on oligonucleotides

The trityl on oligonucleotides were purified by reversed-phase HPLC at 296 nm and 1 x Aufs (Absorbance of UV at full scale) was selected for 1.0 μM DNA prep. A volumn of 0.7 ml sample was applied on Brownlee aquapore RP200 column (ocyl, 0.8 cm x 25 cm) and a binary solvent system was used to give a gradient of acetonitrile in aqueous ammonium acetate (NH_4OAc) buffer. The flow rate was 3.0 ml/min and the gradient as follows:

Buffer A: NH_4OAc (0.1 M), pH 7.4

Buffer B: acetonitrile (50%) in NH_4OAc (0.1 M), pH 7.4

Time	% of Buffer B
0.0 min	10.0
3.0 min	10.0
28.0 min	100.0
30.0 min	10.0
32.0 min	10.0

The second elution peak (trityl on peak) was collected and the buffer was evaporated *in vacuo*. Detritylation was performed by dissolving the oligonucleotide in 2.0 ml of water then adding 8.0 ml of acetic acid (80%) for 30 minutes. The acetic acid was then removed by evaporating with water (1.0 ml) *in vacuo*. several times. The HPLC purified and detritylated oligonucleotide (in 1.0 ml of water) was desalted by a Sephadex G-25 column (Pharmacia NAP-10 column) according to the manufacturer's instructions and the concentration of oligonucleotide was determined by spectrophotometry as described in section 2.1.1.

2.2.6.2 Trityl off oligonucleotides

The trityl off oligonucleotide was redissolved in water (0.7 ml) and purified by reversed-phase HPLC. UV absorbance = 299 nm (1.0 μM) or 295 nm (0.2 μM), Aufs = 1. A Brownlee aquapore RP200 column (ocyl, 0.8 cm x 25 cm) was used. A binary solvent system was used to give a gradient of acetonitrile in aqueous NH_4OAc buffer. The flow rate was 3.0 ml/min and the gradient as follows:

Buffer A: NH_4OAc (0.1 M) in water, pH 7.4
Buffer B: acetonitrile (25%) in NH_4OAc (0.1 M), pH 7.4

Time	% of Buffer B
0.0 min	10.0
3.0 min	10.0
28.0 min	100.0
30.0 min	10.0
32.0 min	10.0

The main peak was collected and the buffer was evaporated *in vacuo* . The redissolved oligonucleotide (in 1.0 ml of water) was desalted by a Sephadex G-25 column (Pharmacia NAP-10 column) according to the manufacturer's instructions and the concentration of oligonucleotide was determined by spectrophotometry as described in section 2.1.1.

2.2.7 Analysis of oligonucleotides substituted with 4-thiothymidine

The oligonucleotide substituted with 4-thiothymidine (2 OD) was dissolved in 0.5 ml of Tris-HCl buffer (50 mM, pH 8.8; NaCl 1M) and digested at 37°C overnight with 12.5 μl of snake venom phosphodiesterase (0.01μg/μl from SIGMA): 1.0 μg in 100 μl of H_2O and 2.5 μl of alkaline phosphatase (0.05μg/μl from SIGMA): 1.0 μl in 20 μl of H_2O. Snake venom phosphodiesterase is an exonuclease which cleaves the 3′-phosphodiester bond of an oligonucleotide from the 3′-terminal resulting in nucleotides and the final 5′ end as a nucleoside. For HPLC analysis, all nucleotides were converted to nucleosides by alkaline phosphatase which cleaves 5′-phosphates from nucleotides. The digested oligonucleotide was analyzed by HPLC at UV absorbances of 264 nm and 333 nm. A separate HPLC column was needed to purify or analyse the enzyme digested oligonucleotides, for the enzymes are very resistant to heat and detergents, therefore they can contaminate the HPLC column and cause degradation of oligonucleotides eluted from the column at a later date.

2.3 *BdU, IodU and dU substituted EcoKI specific oligonucleotides (25 mers)*

BrdU was incorporated at a single position within the complementary *Eco*KI specific oligonucleotides (25 mers) as shown in Figure 13 to give duplexes, designated Br6, Br14, Br17, Br21, Br7′, Br8′, Br11′, Br22′ and Br25′. When BrdU was incorporated at all available positions the duplex was designated duplex Br500. IodU and dU, were also substituted at a single position to give duplexes, designated Io7′ and dU7′ (Figure 13).

2.3.1 γ^{32}P 5′-end labelling of oligonucleotides using polynucleotide kinase (PNK)

The oligonucleotide was diluted with distilled water to make 15 µl of reaction solution to which PNK (obtained from S. Bruce University of Edinburgh, ICMB), (1/50, 2.0 µl), 10 x PNK buffer (2.0 µl), and [γ^{32}P] ATP (3,000Ci/mmol, Amersham), (10 µci/µl,1.0 µl) were added to make 20 µl of final reaction mixture. The reaction mixture was incubated in the Eppendorf at 37°C for 40 minutes then the enzyme was inactivated at 68°C for 10 minutes. The unincorporated radioactive isotope was removed using spun-column chromatography (SCC)[90].

Preparation of 10 x PNK buffer

Final concentration	Prepared from
Tris-HCl (500 mM, pH 7.5)	2.5 µl of Tris-HCl (2 M, pH 7.5)
$MgCl_2$ (100 mM)	1.0 µl of $MgCl_2$ (1 M)
Dithiothreitol (DTT) (50 mM)	5.0 µl of DTT (100 mM)
Distilled water	1.5 µl

Spun-column chromatography (SCC)

1) The bottom of a 1-ml disposable syringe was plugged with a small amount of sterile glass wool.
(2) The syringe was held in a 15-ml disposable plastic tube and filled with Sephadex G-25, which was presuspended in distilled water, until the resin settled down to the 1-ml level.
(3) The spun-column was centrifuged with the plastic tube in a desk top centrifuge mark 3 for 3 minutes.
(4) The spun-column was taken out and the plastic tube dried. A decapped and labelled Eppendorf (1.5 ml) was put in the plastic tube before replacing the spun-column in it. The spun-column was then ready to use.
(5) After loading the radiolabelled sample (20 μl), it was spun using a desk top centrifuge at mark 3 for 3 minutes. Eluted oligonucleotide (20 μl) was collected while the unincorporated $\gamma^{32}P$ ATP remained in the column.
(6) The radioactive syringe was carefully discarded in the radioactive waste bin and the radiolabelled sample stored at -20°C.

2.3.2 Hybridization of oligonucleotides

An equal molarity of complementary oligonucleotide was added to the radioactive labelled oligonucleotide (20 μl). To avoid the exposure to the radiolabelled mixture during boiling at 100°C, the top lid of the Eppendorf was pierced with a needle or sealed with a lid lock. The Eppendorf was placed in a plastic holder and topped by a lead top to avoid the condensed steam passing through the pierced Eppendorf top holes and changing the volume of the hybridizing mixture. After the mixture was boiled gently at 100°C for 5 minutes, it was then incubated at 68°C, 42°C and 37°C successively for 10 minutes each before being placed on ice [91].

2.4 *Gel electrophoresis*

Any charged ion or group will migrate when placed in an electric field. Since proteins carry a net charge at any pH other than their isoelectric point, they will migrate and their rate of migration will depend upon the charge density (the ratio of charge to mass) of the proteins concerned. The higher the ratio of charge to mass the faster the molecule will migrate. The application of an electric field to a protein mixture in solution will therefore result in different migration at different rates towards one of the electrodes [92].

Nucleic acids in solution generally have a negative charge because their phosphate groups are ionized; thus they migrate towards a positive electrode. However, nucleic acid molecules consisting of long chains have almost identical charge to mass ratio, whatever their length, because each residue contributes about the same charge and mass. Also, many proteins that differ in shape and mass have almost equal charge to mass ratios. Therefore, molecules are now most commonly subjected to electrophoresis in a gel, rather than a liquid solution. The size of the pores in such gels limits the rate at which molecules can move through them. Nucleic acids with identical charge to mass ratios separate according to length, with the longer ones moving more slowly. The gel electrophoretic separation according to chain length has become amazingly reliable in separation of DNA, which has made DNA sequencing possible; and it also can separate most proteins in a cell under denaturing condition.

2.4.1 Polyacrylamide Gel Electrophoresis (PAGE)

Polyacrylamide gels can be considered as porous media in which the pore size is the same order as the the size of protein molecules such that a molecular sieving effect occurs and the separation is dependent on both charge density and size. Polyacrylamide gel, as a synthetic polymer of acrylamide monomer, can always be prepared from highly purified reagents with a wide range of pore sizes in a reproducible manner provided that the polymerization conditions are standardized. In addition, polyacrylamide gel has the advantages of being chemically inert, stable over a wide range of pH, temperature, and ionic strength and is transparent. Polyacrylamide gel results from the polymerization of acrylamide monomer into long chains and the crosslinking of these by bifunctional compounds such as N,N'-methylene bisacrylamide (usually abbreviated to bisacrylamide) reacting with free funtional groups at chain termini. Polymerization of acrylamide is initiated by the addition of ammonium persulphate (Amps). In addition, N,N,N',N'-tetramethylene diamine (TEMED) is added as an accelerator, which catalyses the formation of free radicals from Amps of the polymerization process [92].

Non-denaturing PAGE was used in this study to detect the efficiencies of BrdU substituted DNA-mtase binding compared to that of unsubstituted DNA-mtase. PAGE system was Prepared as follows:

(I) 20 x TBE, pH 8.3 (1 litre) Stock solution for PAGE:

TRIS (90 mM)	218.0 g
Boric acid (90 mM)	110.0 g
EDTA (18.6 mM)	18.6 g
Distilled water	up to 1.0 liter

(II) PAGE gel mixtures (50 ml of 5% acrylamide)

Stock solution (A)	8.4 ml
Distilled water	38.9 ml
20 x TBE	2.5 ml
Amps (10%)	500 µl
TEMED	50 µl

(III) Preparation of PAGE gel:

Pouring an acrylamide gel (20 cm x 20 cm x 1.5 mm), reaction mixture was prepared by mixing acrylamide (30%, 8.4 ml), water (38.9 ml) and 20 x TBE (2.5 ml) firstly, then adding ammonium persulphate (10%, 500 µl) and TEMED (50 µl), swirling gently for oxygen inhibits polymerization. The well mixed solution was poured immediately into the gel mould to set for at least 1.5 h before use. PAGE sample buffer and the electrophoretic buffer has been described in the materials.

(IV) PAGE sample buffer (10 ml)

Glycerol (10 %, v/v)	1.0 ml
Bromophenol blue (B.b., 0.001%)	0.1 ml of (1% B.b.)
TBE (x 1)	0.5 ml of (20 x TBE)
Distilled water	8.4 ml

(V) PAGE electrophoretic buffer (1.0 liter)

1x TBE	50 ml of 20 x TBE
Distilled water	950 ml

2.4.2 SDS-PAGE system

Proteins can be fractionated in polyacrylamide gels using a buffer system designed to dissociate all proteins into their individual polypeptide subunits. The most common dissociating agent used is the ionic detergent, sodium dodecyl sulphate (SDS). The protein mixture is denatured by heating at 100°C in the presence of excess SDS and a thiol reducing reagent (e.g. β-mercaptoethanol to cleave disulphide bonds). Under these conditions, most polypeptides bind SDS [93] in a constant weight ratio (1.4 g of SDS per gram of polypeptide). The intrinsic charges of the polypeptide are insignificant compared to the negative charges provided by the bound detergent, so that the SDS-polypeptide complexes have essentially identical charge densities and migrate in polyacrylamide gels of the correct porosity strictly according to polypeptide size [94,95]. Thus, in addition to analysing the polypeptide composition of the sample, the molecular mass of the sample polypeptides can be determined by reference to the mobility of polypeptides of known molecular mass under the same electrophoretic conditions.

A discontinuous buffer system[96,97] which employed different buffer ions in the gel to those in the reservoirs, was used in this study to obtain a better gel resolution. This system has discontinuities of both buffer composition and pH. Two gel layers, stacking gel and resolving gel, were required for use with the discontinuous buffer system. The resolving gel is poured first, and allowed to polymerize under H_2O to give a flat meniscus, followed by pouring of the stacking gel in which the comb is inserted to form the sample wells. The gel must be allowed to set at least 2 hours before loading the sample. Samples are loaded onto the large-pore stacking gel polymerized on top of the small-pore resolving gel. Under the discontinuous buffer system, a relatively large volume of dilute protein sample can be applied to the gel. This is because during migration through the large -pore stacking gel, the proteins are concentrated into extremely narrow zones prior to their separation during electrophoresis in the small-pore resolving gel. An equal volume of SDS-PAGE sample buffer is mixed with the sample before the sample is loaded on the gel, which is run for 3 hours at a power of 45 mA. The sample buffer contains glycerol to increase the density of the loading sample, so the mixture will sink to the bottom of the well; β-mercaptoethanol is induced to cleave disulphide bonds of the protein and B.b is to indicate the front of buffer. After electrophoresis the fractionated polypeptides can be seen by either Coomassie Blue staining or silver staining.

SDS-PAGE system was prepared as follows:

(I) Stock solutions:

Stock solution (A) : Acrylamide (30%)

Acrylamide	58.0 g
NN'-methylenebis acrylamide	2.0 g
Distilled water	up to 200 ml

Stock solution (B) : TRIS-HCl (1.5M, pH 8.8)

Tris	18.2 g
Distilled water	up to 100 ml
pH adjusted with HCl	

Stock solution (C) : SDS (10%)

SDS (99% pure)	10.0 g
Distilled water	up to 100 ml

Stock solution (D) : Tris-HCl (0.5M, pH 6.8)

Tris	6.0 g
Distilled water	up to 100 ml
pH adjusted with HCl	

(II) SDS-PAGE gel mixtures :

(A) Resolving gel mixture (50 ml of 10% acrylamide) :

Stock solution (A)	16.7 ml
Stock solution (B)	12.5 ml
Stock solution (C)	0.5 ml
Distilled water	19.8 ml
Ammonium persulphate (10%)	500 μl
TEMED	50 μl

(B) Stack gel mixture (50 ml of 5% acrylamide) :

Stock solution (A)	2.0 ml
Stock solution (D)	2.5 ml
Stock solution (C)	100 μl
Distilled water	5.3 ml
Ammonium persulphate (10%)	100 μl
TEMED	10 μl

The preparation of each layer of the SDS-PAGE gel is similar to that of the PAGE gel as described above.

(III) SDS-PAGE sample buffer (10.0 ml) :

Final concentration	Preparation	
Tris-HCl (0.125 M, pH 6.8)	Stock solution (D)	2.5 ml
SDS (2%)	Stock solution (C)	2.0 ml
β-mercaptoethanol (1%)	β-mercaptoethanol (10%)	1.0 ml
B.b (0.001%)	B.b (1%)	0.1 ml
Glycerol (20%)	Glycerol	2.0 ml
Distilled water		2.5 ml

(IV) SDS-PAGE electrophoretic buffer (1.0 litre):

Tris (50 mM)	6.0 g
Glycine (380 mM)	28.8 g
Stock solution (C) (0.1%)	10.0 ml
pH 8.8	pH adjusted with HCl
Distilled water	up to 1 litre

2.4.3 Gel staining

2.4.3.1 Coomassie Blue Staining

Methanol-based Coomassie Blue R-250 (R = reddish hue) stain was used for routine detection of mtase on denaturing SDS-PAGE gels during its purification from the cell and was used on non-denaturing PAGE gels to detect free proteins and the DNA-mtase binding complex in this study. The minimum time required for staining depends on both the gel thickness and the gel concentration. The rate of staining is increased at higher temperatures, for example 40-50^{o}C. The excess stain on the background was removed by its diffusion into the destain solution containing 10% of acetic acid and 10% of ethanol.

Preparation of Coomassie Blue staining

(A) Staining solution (1 litre):

Coomassie Blue	1.0 g
Acetic acid	100 ml
Methanol	500 ml
Distilled water	up to 1 litre.

(B) Destain solution (1 litre):

Acetic acid (10%)	100 ml
Methanol (10%)	100 ml
Distilled water	800 ml

2.4.3.2 Silver staining

Silver staining was introduced in 1979 by Switzer *et al* [99] as a novel staining procedure which was up to 100 times more sensitive than Coomassie Blue R-250, was able to detect as little as 0.38 ng/mm^2 of bovine serum albumin. Although it is the most sensitive non-radioactive protein detection method currently available, it has a number of drawbacks. For example high backgrounds can result when the water used is not pure enough; the cost is more than other staining (e.g. Coomassie Blue staining); some protein (e.g. protein without or with very little cysteine) may not stain with some silver stains and some silver stains detect not only proteins but also DNA, lipopolysaccharides and polysaccharides.

The Silver Stain Plus Kit (Bio-Rad) was used in this study. High quality deionized distilled water and extremely clean glassware were used to obtain the best results and gloves were worn when handling gels. The commercial silver staining involves three steps: fixing, washing, and staining. The fixation step was to immobilize the proteins in the gel and to remove substances which might interfere with the staining procedures such as detergent, reducing agent or reactive buffer components like glycine. The gel was then exposed to silver nitrate solution and allowed to develop. The following preparations were for staining one mini gel (8 x 10 cm), 0.75-1.0 mm thick. After gel electrophoresis, the gel was placed in 40 ml of the fixative enhancer solution (methanol, 20 ml; acetic acid, 4 ml; deionized water, 14 ml and fixative enhancer concentrate, 2 ml) with gentle agitation for 30 minutes. After decanting the fixative enhancer solution, the gel was rinsed in (40 ml x 2) deionized distilled water for (10 minutes x 2) with gentle agitation twice. During rinsing the gel the staining and developing solution was prepared by adding (1) 1.0 ml of the silver complex solution, (2) 1.0 ml of the reduction moderator solution and (3) 1.0 ml of image development reagent stepwise. While stirring rapidly, the Development Accelerator Reagent (10 ml) was added quickly to the mixture. After rinsing, the rinsing water was decanted and the staining and developing solution (20 ml) was added immediately to stain the gel with gentle agitation. After the protein bands became visible the staining solution was decanted and 5% acetic acid was added to stop the reaction. The gel was ready to be dried or photographed. The basic mechanism of silver staining is the reduction of silver nitrate to metallic silver at the protein band, leading to the deposition of silver grains. Studies have shown the importance of basic and sulphur-containing amino acids in silver staining. However the precise mechanism of silver staining is not yet clearly known.

2.5 *Preparation and purification of EcoKI mtase*

2.5.1 Preparation of *Eco*KI mtase

2.5.1.1 Expression of the *Eco*KI mtase

Preparation and purification of the *Eco*KI mtase was based on Dryden's method [13]. In plasmid pJFMS, the *hsdM* and *hsdS* genes, encoded within a 7.8 kilobase *Sma*I-*Eco*RI fragment from λNM1048(7), were inserted into pJF118HE (expression vector) (20) in which they are transcribed from the *tac* promoter. pJF118HE encodes ampicillin resistance (AmpR).

2.5.1.2 Enzyme amplification

Bacteria (NM522(9)), freshly transformed with pJFMS (see above), were grown in 4 liters of L broth containing ampicillin (30 μM) at 37°C to an optical density of 0.5 at 650 nm before induction of transcription by the addition of isopropyl-β-D thiogalactopyranoside (IPTG, 1 mM). Cells were harvested and centrifuged after 4 hours of induction. The cell paste was stored at -20°C or -70°C until required.

2.5.1.3 Tris-MES buffer for *Eco*KI mtase purification

Tris (20 mM)
MES (20 mM)
$MgCl_2$ (10 mM)
EDTA (0.1 mM)
β-mercaptoethanol (7 mM)
pH adjusted to 6.5 or 8.0 with HCl or NaOH

2.5.1.4 Sonication of the cells

Benzamidine (20 μM) and phenylmethyl sulphonyl fluoride (PMSF), (10 μM) were added to the sonication buffer Tris-MES buffer (pH 6.5) before sonication to inhibit proteases during mtase extraction procedures. Bacteria (15 g) of cells were resuspended in Tris-MES buffer (pH 6.5) to a volume of 50 ml and sonicated in an ice bath with 30 second bursts for a total time in minutes equal to the weight of the cell paste.

2.5.2 Purification of *Eco*KI mtase

All the purification steps were carried out at 4°C. At each step of column purification, protein in the eluting fractions was monitored by UV absorbance at 280 nm. The molecular weight of proteins in the eluting fractions were determined by comparing with the marker on a SDS-PAGE gel (12% acrylamide) [93] following Coomassie Blue (0.1%) staining [13]. Fractions containing the mtase (M_2S_1) determined by SDS-PAGE were pooled. The concentration of mtase was determined by UV spectrophotometry to ascertain the concentration and quality of protein thus obtained. A peak-to-trough (A_{280}/A_{250}) ratio of 2.7 was typical as shown in Figure 14. For determining the protein concentration, the extinction coefficient at 280 nm was 8.42 for a 1% (w/v) solution in a 1 cm pathlength cell [91].

Columns for purification of the *Eco*KI mtase

(1) CM-Sepharose column (Pharmacia LKB Biotechnology Inc.)
(33 cm in length and 2.6 cm in diameter)

(2) The Heparin-Agarose column (Sigma Chemical Co.)
(15 cm in length and 1.6 cm in diameter)

(3) DEAE-Sepharose column (Pharmacia LKB Biotechnology Inc.)
(33 cm in length and 1.6 cm in diameter)

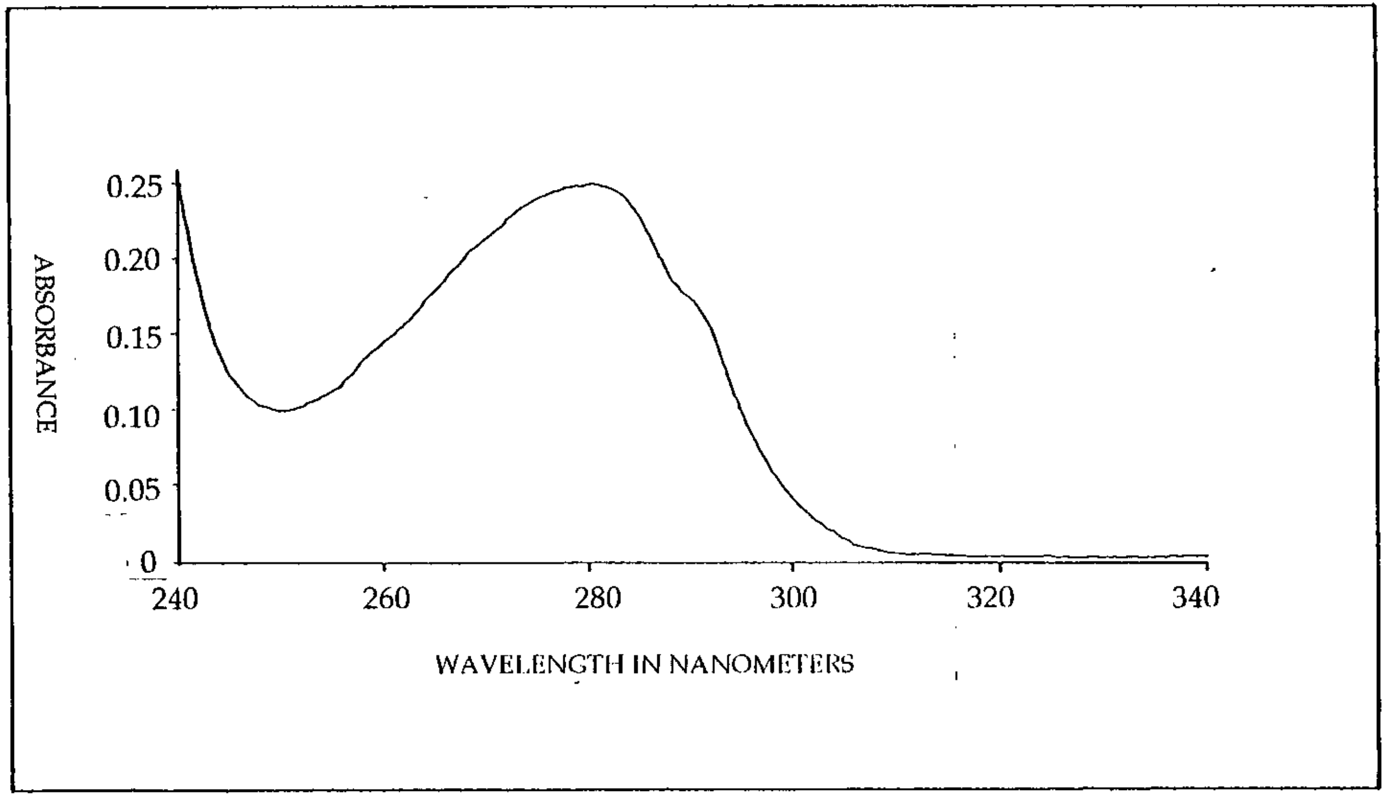


Figure 14. The UV spectrum of type I *Eco*KI methyltransferase (1.75 µm/µl)

2.5.2.1 CM-Sepharose column purification

The CM-Sepharose column was pre-equilibrated with Tris-MES buffer (pH 6.5) at a flow rate of 48 ml/h for 2 hours (100 ml of buffer is about two column volumes). The cell extract was clarified by centrifugation at 30,000 x g for 3 hours and the supernatant applied to the CM-Sepharose column which was then washed with 100 ml of Tris-MES buffer to remove unbound material. The bound protein was eluted with a gradient of 0-0.5 M NaCl in Tris-MES buffer (500 ml) at a flow rate of 24 ml/h overnight and the eluted fractions were collected every 12 minutes.

The eluted fractions were again monitored by UV at 280nm and analysed by SDS-PAGE. After Coomassie Blue staining, fractions containing the mtase were pooled. The pooled protein (58 mgs estimated from the UV spectrum) was dialyzed for 4 hours against Tris-MES buffer (pH 6.5, 2 litres) to remove NaCl. The column was regenerated with Tris-MES buffer (pH 6.5, 100 ml, NaCl 1M) at a flow rate of 48ml/h then pre-equilibrated with Tris-MES buffer (pH 6.5) before use.

2.5.2.2 Heparin-Agarose column purification

The Heparin-Agarose column was pre-equilibrated with Tris-MES buffer (50 ml, pH 6.5) at a flow rate of 48 ml/h. The dialysed mtase obtained from the CM-Sepharose column was applied to the Heparin-Agarose column at a flow rate of 48 ml/h then washed with buffer A (50 mls) at the same flow rate. The bound protein was eluted with a 500 mls gradient of 0-1.0 M NaCl-buffer A in pH 6.5 at a flow rate of 24 ml/h overnight and the eluted fractions were collected every 12 minutes. The pooled protein (36 mgs estimated from the UV spectrum) was dialysed in Tris-MES buffer (pH 8.0, 2 liters) for 4 hours. The regeneration of the Heparin-Agarose column was the same as described for the regeneration of the CM-Sepharose column.

2.5.2.3 DEAE-Sepharose column purification

The DEAE-Sepharose column was pre-equilibrated with 100 ml Tris-MES buffer (pH 8.0) at a flow rate of 48 ml/h. The dialyzed mtase obtained from the Heparin-Agarose column was applied to the DEAE-Sepharose column at a flow rate of 48ml/h then washed with 100ml Tris-MES buffer (pH 8.0) at the same flow rate. The bound protein was eluted with a 500 ml gradient of 0-0.4M NaCl inTris-MES buffer (pH 8.0) at a flow rate of 24 ml/h overnight and the eluted fractions were collected again every 12 minutes. Fractions of M_1S_1 and M_2S_1 were pooled according to the monitored UV spectrum at 280 nm and the analysis of 12% SDS-PAGE gels. The pooled protein was then precipitated by adding ammonium sulphate to 80% saturation (0.56 g/ml). The pellets obtained after centrifugation (10K for 20 minutes) were resuspended in small volumes of Tris-HCl buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl) containing 50% glycerol (usually to make a concentration of 10-20 mg/ml) and stored at -20°C. The DEAE-Sepharose column can also be regenerated with Tris-MES buffer (100 ml, pH 8.0, NaCl, 1M).

2.6 Gel retardation assays

2.6.1 Reagents for DNA-mtase binding and crosslinking

AdoMet (New England Biolabs, 32 mM)
*Eco*KI recognition sequence oligonucleotides and derivatives (Figures 12,13 and 22, synthesized in the Department of Chemistry)
*Eco*KI mtase (prepared and purified in the Institute of Cell and Molecular Biology)
Tris-HCl buffer : Tris-HCl (20 mM, pH 8.0), NaCl (100 mM)

2.6.2 Binding efficiencies of 5-BrdU substituted oligonucleotides

Gel retardation assays were used to investigate the binding efficiencies of substituted DNAs with mtase before UV crosslinking. 5-BrdU substituted DNA fragments were used. Unsubstituted DNA (S25) and non-specific DNA (NS25) were included in some studies as a control. DNAs used in these experiments were bottom strand labelled with [γ^{32}P] ATP. Mtase stock solution was desalted by using a G-25 Sephadex gel filtration column (Pharmacia, PD-10 column) and Tris-HCl buffer (20 mM Tris-HCl, pH8.0, 100 mM NaCl). Mtase (200 nM) with premixed AdoMet (100 μM) and DNA (10 nM) were added in Tris-HCl buffer (20 mM pH8.0, NaCl 100 mM) containing 5% glycerol and mixed in a microcentrifuge. The total volume of reaction mixture was 200 μl. After 15 minutes of incubation at 22°C the reaction mixture (20 μl) was loaded on a non-denaturing polyacrylamide gel (5%) and run for 1 to 1.5 h at 30 mA. Following electrophoresis, the gels were dried at 80°C for 1 h. The radioactivities of bound DNA and free DNA were monitored by autoradiography using preflashed film (Amersham) and quantitative measurements were determined by scintillation counting. Binding efficiency was represented by the fraction of radiolabelled DNA (in counts/min) in the bound DNA relative to the total amount of DNA (free+bound).

2.6.3 Binding affinities of S25, Br7′, dU7′, and Io7′

For investigation of the binding affinity of S25, Br7′, dU7′, and Io7′ with mtase, 0.1 nM of target DNA labelled in the bottom stranded, 0 to 100 nM of mtase in 10 nM intervals and 100 μM of AdoMet were used in each set of experiments. The total volume of reaction mixture was 200 μl. Other conditions for binding incubation, gel electrophoresis and quantitative measurments were the same as described above.

2.7 *UV-induced crosslinking of DNA-EcoKI mtase*

2.7.1 The optimum wavelength and time for UV irradiated crosslinking

Crosslinking of DNA-mtase resulted from irradiating the DNA-mtase binding complex with UV at 254 nm. The DNA-mtase crosslinked complexes and free DNA were estimated by SDS-PAGE followed by autoradiography (for radiolabelled DNA) of the dried gel. The DNA-mtase crosslinked complexes and the unreacted M polypeptide and S polypeptide were detected by silver staining the SDS-PAGE gel (12% polyacrylamide). The concentrations of the reaction mixture used in the following experiments were 10 nM DNA, 200 nM mtase and 100 μM AdoMet, unless otherwise stated.

Different wavelengths of UV source were investigated in order to obtain the optimum yield of crosslinked complex. 4-thiothymidine and BrdU substituted DNAs were evaluated for their ability to crosslink efficiently. The specific DNA duplex (S25) and the non-specific DNA duplex (NS25) were used as controls. DNA-mtase binding mixture was placed on a multiwell plate and kept on ice for UV-irradiation. Sample of DNA-mtase binding mixture prepared as described above (total volumn 20 μl) were irradiated at the wavelengths of 254 nm, and 365 nm by the UV handlamp set (Model UVGL-58, UVP International, Ultra-Violet Products, Ltd), 302 nm by an UV Transilluminator (Model TM-20, UVP International, Ultra-Violet Products, Ltd) and 320 nm by an UV Transilluminator (Model TM-20) with a 320 nm cut off filter for 10 minutes. The emission spectra of the UV lamps are illustrated in Figure 15.

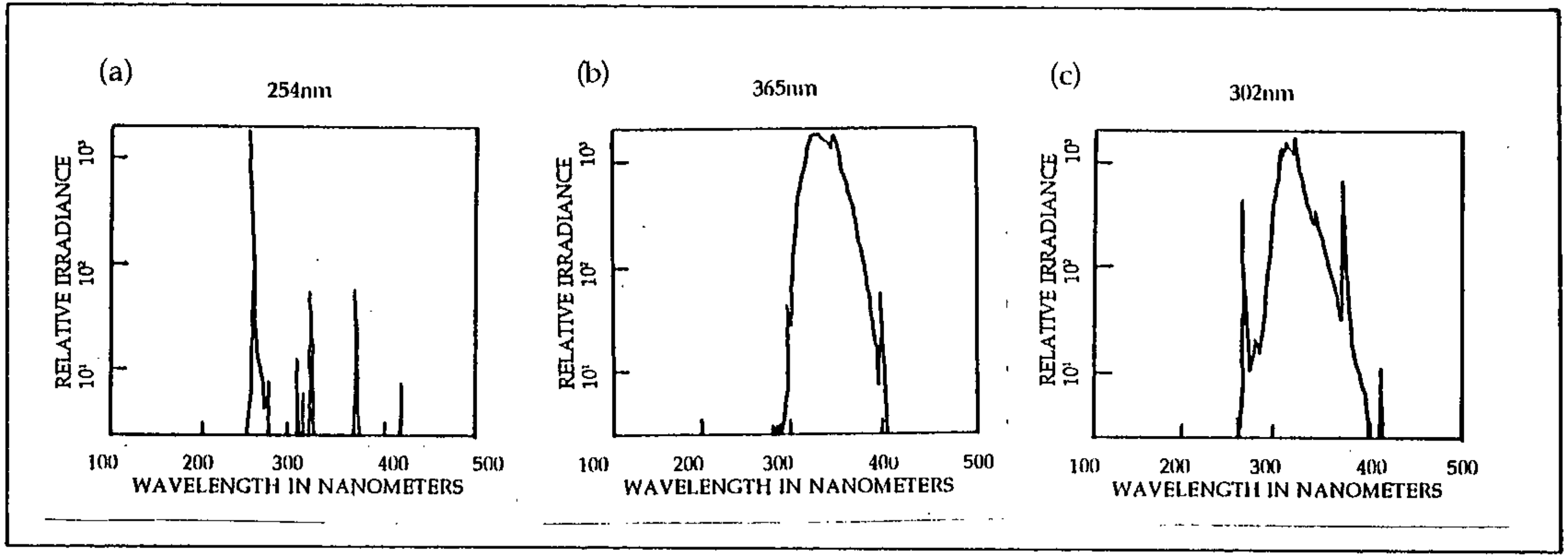


Figure 15 The emission spectra of the UV handlamp set, Model UVGL-58: (a) 254 nm and (b) 365 nm. UV Transilluminator TM-20: (c) 302 nm. (Products of UVP Internation, Ltd).

A time course of UV irradiation was carried out using the wavelength at 254 nm. BrdU substituted DNA (Br7′) was used in this experiment. The concentration of the reaction mixture was 10 nM DNA (Br7′), 200 nM mtase and 100 μM AdoMet in 20 mM Tris-HCl pH 8.0, 100 mM NaCl (total volume 200 μl).

2.7.2 Crosslinking efficiencies of S25, Br7′, dU7′ and Io7′

The crosslinking efficiencies of Br7′, Io7′ and dU7′ were compared with the unsubstituted S25 (UV irradiation at 254 nm for 10 minutes). Concentrations of the reaction mixtures were 0.68μM mtase, 2.0μM DNA and 68 μM AdoMet in 20 mM Tris-HCl pH 8.0, 100 mM NaCl (total volume 200 μl). SDS-PAGE (12% polyacrylamide) was used to separate the mtase and free DNA from the DNA-mtase crosslinking bands. Two identical gels were run in parallel and following gel electrophoresis, one of the gels was silver stained and the other was dried and recorded by autoradiography. Scintillation counting of the labelled DNA in bands identified by autoradiography provided quantitative measurements. Crosslinking efficiency was represented by the fraction of radiolabelled DNA (in counts/min) in the crosslinked DNA relative to the total amount of DNA (free+crosslinked).

2.7.3 Single stranded DNA-mtase crosslinking

To determine whether the DNA duplex-mtase crosslinked band (upper band) of the heated sample on the 12% SDS-PAGE gel was due to the rehybridization of the DNA of a single stranded DNA-mtase crosslinked complex or due to the formation of crosslinking among mtase and both strands of DNA, a 7 M urea-SDS-PAGE gel was used. DNA duplex (Br7′) was bottom-strand labelled. Samples of crosslinked complexes were either unboiled or boiled for 2 minutes before loading on gels. The unboiled samples of crosslinked complexes were loaded on both a 12% SDS-PAGE gel and a 7 M urea-SDS-PAGE gel directly after UV irradiation. The boiled samples of crosslinked complexes were quenched in liquid nitrogen immediately after boiling then loaded on both 12% SDS-PAGE gel and 7 M urea-SDS-PAGE gel for electrophoresis.

DNA duplexes (S25, Br17, Br7′ and Br500) were also labelled in different strands and the samples of crosslinked complexes were boiled before analysis using 12% SDS-PAGE and 7 M urea-SDS-PAGE gels. The top-strand, bottom-strand, and both-strands of the DNAs were labelled with [γ^{32}P] ATP respectively. The samples of crosslinked complexes were boiled for two minutes before loading for electrophoresis.

2.7.4 Identification of the nucleotide which is crosslinked by *Eco*KI mtase

To determine which nucleotide of the 25-mer was crosslinked by the *Eco*KI mtase, single 5-BrdU-substituted oligonucleotides, designated Br6, Br14, Br17, Br21, Br7′, Br8′, Br11′, Br22′, Br25′ (Figure 13), were used. Br500 with all nine 5-BrdU-substitutions was also used. Unsubstituted DNA (S25) was again included in the study for background crosslinking measurement. The oligonucleotides in this experiment were labelled with [γ^{32}P] ATP on both strands. The samples of crosslinked complexes were boiled for 2 minutes before loading on 12% SDS-PAGE gels. The quantitative measurements of the crosslinking efficiencies were as described in section 2.7.2.

2.8 *Large scale UV irradiated DNA-mtase crosslinking*

2.8 1 Experimental procedure for large scale UV irradiated DNA-mtase crosslinking

For large scale experiments, DNA duplex Br7'(20.4 μM) desalted mtase (8.8 μM) and AdoMet (70 μM) in a total volume of 4.6 ml were incubated for 15 min at 22^{o}C. The molar ratio of DNA/mtase was 2.5:1. The reaction mixture was spreaded homogeneously to form a thin layer on the inside of the lid (12.5 cm x 8.5 cm) of a multiwell plate,. which was kept on ice for UV irradiation at 254 nm. The sample was irradiated by UV at a distance of 7.0 mm for 10 minutes. The shape of the UV bulb is long and narrow (1.6 cm in diameter, 21.0 cm in length), in order to let all of the sample be irradiated directly under the UV light, the UV lamp was moved every 10 minutes along the lid of the multiwell plate where the samples were spread.

2.8.2 Purification of DNA-mtase crosslinked complexes

DNA-mtase crosslinked complexes were separated from the reaction mixtures by SDS-PAGE electrophoresis and the crosslinked complexes were eluted from the gels.

Preparation of the elution buffer :

	Final concentration
Ammonium acetate	0.5 M
EDTA	1.0 mM
SDS	0.1%
$MgCl_2$	10.0 mM
pH 7.5	adjusted with NaOH

After UV irradiation, SDS-PAGE sample buffer (1.5 ml) was mixed with the reaction mixture (4.0 ml) for loading on two denaturing SDS-PAGE gels (9% acrylamide). The crosslinked complex was separated from the free DNA by electrophoresis. The radioactive bands of crosslinked complex were monitored by autoradiography and cut out with a razor. The gel bands of crosslinked complex were crushed through a syringe (1.0 ml) into a bottle (20 ml). The crosslinked complex was eluted by shaking the crushed gels with a rotary shaker in 10 ml of elution buffer (ammonium acetate 0.5 M, EDTA 1.0 mM, SDS 0.1%, $MgCl_2$ 10.0 mM, pH 7.5 adjusted with NaOH) at room temperature overnight (at least eight hours). The eluted crosslinked complex was filtered through a syringe (5.0 ml) with a small lump of glasswool at the top of the syringe. During filtration the condensed crushed gels were occasionally stirred with a Pasteur pipette to help the elution passing through. The crushed gel in the syringe was washed with distilled water (1.0 ml) and centrifuged using a desktop centrifuge for 5 minutes for a better recovery of crosslinked complex. The crosslinked complex elution (~9 ml) was concentrated by freeze-drying for 5 hours. The freeze-dried product was dissolved in 1.6 ml of Tris-HCl (20 mM, pH 8.0) buffer at 37°C then desalted by G-25 Sephadex gel filtration columns (PD-10). The PD-10 column was prewashed with distilled water (35.0 ml) before loading the crosslinked complex elution. Resuspended freeze-dried product (2.5 ml containing 1.8 ml of sample and 0.7 ml of flask washing) was loaded on to the PD-10 column then eluted with 3.5 ml of Tris-HCl (20 mM, pH 8.0) buffer. The PD-10 column was washed with 4.0 ml of distilled water three times. The first (desalted crosslinked complex in 3.5 ml of Tris-HCl) and the second (desalted crosslinked complex in 4.0 ml of H_2O) fraction elutions from PD-10 column containing high radioactivity were then digested by trypsin.

An alternative sequence of freeze-drying and desalting was applied in later experiments to avoid the difficulty of dissolving the high salt content freeze-dried crosslinked complex in a small volume of buffer for desalting. In the alternative procedure, after the crosslinked complex was eluted from the gels, the elution buffer was desalted first then concentrated by freeze-drying for trypsin digestion.

2.8.3 Trypsin digestion of purified DNA-mtase crosslinked complexes

Bovine pancreatic trypsin with a molecular mass 24,000 kDa exhibits a highly selective specificity for the hydrolysis of peptide bonds involving the carboxyl group of the basic amino acids, arginine, and lysine [99]. It has assumed a major role in amino acid sequence determination.

Active trypsin is obtained from its inactive precursor, trypsinogen, which is stable between pH 2.0 and 4.0. In neutral and alkaline solution by limited proteolysis trypsinogen is autocatalytically activated to trypsin by small amounts of existing trypsin . Trypsin is most active in the pH range of 7-9, where it rapidly loses its activity through the continuous autolysis. However, the enzyme is protected from autolysis by the presence of 0.01 M Ca^{2+} ions [100]. The activity of trypsin can be negatively influenced by physical parameters (temperature and pH), by conformational changes of its molecule (denaturation), by chemical modifications, or by specific interactions with low molecular weight or naturally occurring polypeptide inhibitors.

The number of peptides that are produced by trypsin digestion of a polypeptide chain should, in theory be equal to the number of lysine and arginine residues in the chain plus a COOH-terminal fragment which contains neither lysine nor arginine. In fact, regions of polypeptide chain that contain adjacent lysine or arginine residues frequently prove to be resistant to trypsin cleavage, as do peptides that contain adjacent negatively charged residues, such as glutamic acid, aspartic acid, or cysteic acid. In principal, there are three kinds of substituents in the amino acid substrate which can directly influence the catalysis, i.e., the side chain, the substituent on the α-amino group, and the substituent on the carboxyl group. Additional influence can be exerted by neighbouring amino acids in the primary structure of the polypeptide chain or by the tertiary structure of a macromolecular substrate.

Both in synthetic and in polypeptide substrates, trypsin prefers the arginine side chain to the lysine side chain. This effect becomes even more pronounced at higher pH values (35:1 at pH10.0 and 52:1 at 10.7) since the ε-amino group of lysine is considerably discharged under these conditions while the guanidinium group of arginine is not affected [101]. Although the majority of positively charged groups, e.g. side chains of arginine and lysine, in native proteins is supposed to be located on the surface of the molecule, the three-dimensional structure imposes certain limitations on the action of trypsin. It has been observed in numerous studies that trypsin does not degrade native proteins prior to denaturation.

Specific activity of a trypsin preparation is expressed in units of trypsin per milligram of protein. One unit of trypsin is defined as the amount of trypsin catalysing the transformation of 1 mmole of p-toluenesulfonyl-l-arginine methyl ester per minute at 25°C and pH 8.1 in the presence of 0.01M calcium ion [102,103].

2.8.3.1 Experimental procedure for trypsin digestion

Freshly prepared trypsin (4.0 µl, 12.0 µg/µl) and $CaCl_2$ (35.0 µl, 2 M) were added to samples (the first (3.5 ml) and the second (4.0 ml) fractions from PD-10 column contained high radioactivity) of crosslinked complexes, which were eluted from the SDS-gel and desalted by PD-10 column. The mixture was incubated at 37°C overnight.

In order to obtain a very short DNA-peptide crosslinked complex for amino acid sequencing, the diphenyl carbamyl chloride (DPCC)/trypsin in ammonium acetate (200 mM, pH8.5) digestion buffer was used instead of Tris-HCl digestion buffer and redigestion of the partial digested crosslinked complexes by trypsin at the same condition was used. Possible chymotrypsin contamination of the trypsin stock solution (10 mg/ml trypsin in 100 mM Tris-HCl, 100 mM $CaCl_2$ pH 8.0) was inactivated using DPCC (Light & Co. Ltd) [104].

The purified concentrated DNA-mtase crosslinked complex was resuspended in 1 ml of ammonium acetate (200 mM, pH 8.5). DPCC/trypsin (Boehringer Mannheim) working solution (1.0 ml, 1.0 mg/ml) was added and incubated at 37°C for 12 hours. After purification by ion-exchange chromatography and analysis by 20% SDS-PAGE gel, the partially digested DNA-peptide crosslinked complex was redigested by trypsin as described previously. A satisfactory short DNA-peptide crosslinked complex was obtained for amino acid sequencing.

Preparation of DPCC/trypsin solution

(1) Buffer for trypsin digestion:
 ammonium acetate (NH_4OAc) (0.2 M, pH 8.5)

(2) Buffer for stock trypsin:

Tris, (0.1 M)/$CaCl_2$, (0.1 M) solution (pH 8.0):

Tris (0.1 M)	6.06g
$CaCl_2$ (0.1 M)	5.55g
Distilled water	up to 500 ml

pH adjusted to 8.0 with HCl

(3) Dissolving DPCC in acetone :

DPCC (Light & Co. Ltd)	10.0 mg
Analar acetone	1.7 ml

(4) DPCC/trypsin stock solution :

10.0 μg/μl trypsin in Tris, (0.1 M)/$CaCl_2$, (0.1 M), pH 8.0

Tris (0.1 M)/$CaCl_2$ (0.1 M)	4.0 ml
trypsin (Boehringer Mannheim)	40.0 mg
DPCC in acetone	20.0 μl

DPCC was added to the trypsin stock solution at room temperature to react for 90 minutes, if necessary spinning down the precipitated chymotrypsin.

(5) DPCC/trypsin working solution (1.0 μg/μl), dilute (4) with NH_4OAc (0.2 M, pH 8.5).

2.8.4 Purification of trypsin digested, DNA-peptide crosslinked complexes

2.8.4.1 Ion exchange chromatography

The separation in ion exchange chromatography is obtained by reversible adsorption. Most ion exchange experiments are performed in two main stages. The first stage is sample application and adsorption. Unbound substances can be washed out from the exchanger bed using a column volume of starting buffer (low salt). In the second stage, substances are eluted from the column by a gradient of salt separating them from each other. The separation is obtained since different substances have different affinities for the ion exchanger due to differences in their charge. These affinities can be controlled by varying conditions such as ionic strength and pH.

The Econo-Pac Q cartridge used in this study is a strongly basic anion exchanger with a charged functional group $-N^{+}(CH_3)_3$ covalently bonded to an insoluble gel. The positively charged groups are associated with mobile negatively charged counter-ions (anions). Firstly, the column is equilibrated with starting buffer counter ions. Secondly, sample substances are applied (e.g. mixture of digested peptides and DNA-peptide crosslinked complexes) by exchanging the buffer counter-ion to the higher negatively charged sample substances. Finally, the sample substances are eluted with a gradient, which desorbs the sample substances separately (e.g. peptides would be eluted at earlier stages than the more negatively charged DNA-peptides). After gradient elution the column can be regenerated and equilibrated with starting buffer for reuse or stored in 20% ethanol.

Buffer for Econo-Q ion exchange column

KH_2PO_4	20 mM
Ethanol	5%
pH 6.8	adjusted with HCl (KOH)
KCl gradient	0-0.6 M (100 ml)

2.8.4.2 Purification of DNA-peptide crosslinking complexes

The digested peptide-DNA crosslinked complex was purified from the digested mixture using an Econo-Pac Q cartridge (BIO-RAD). Trypsin digested sample mixture (2.0 ml) diluted with 20.0 ml of phosphate buffer (20 mM, 5% ethanol, pH 6.8) was centrifuged and decanted. The Econo-Pac Q cartridge was equilibrated in phosphate buffer (20 mM, 5% ethanol, pH 6.8) at 48.0 ml/h for one hour before the clear sample mixture was loaded on. After loading the sample solution, the column was washed with salt-free phosphate buffer (20 mM, 5% ethanol, pH 6.8) then eluted with a 100 ml gradient of 0-0.6 M KCl in phosphate buffer at 48.0 ml/h. The elution fractions were collected for crosslinking studies by measuring their radioactivity, and the protein and peptides in the fractions were determined using SDS-PAGE and silver staining. The radioactivity was determined using a Beckmann LS7000 scintillation counter, by adding 1.0 µl of elution fractions to 2.5 ml of Ecoscint (National Diagnostics). DNA-peptide fractions were pooled according to the monitored UV spectrum at 254 nm and the radioactivity of crosslinked DNA. The purified sample was desalted with a G-25 Sephadex gel filtration column before freeze-drying.

2.9 *Peptide sequencing-identification of amino acid crosslinked to the specific DNA*

2.9.1 The Edman reaction-phenyl isothiocyanate degradation

Proteins were sequenced by degradation from their N termini using the Edman reagent, phenyl isothiocyanate (PITC). The reaction was divided into three steps: coupling, cleavage, and conversion. During one cycle of reaction, the N-terminal residue was removed from a polypeptide and identified by HPLC. The shortened polypeptide was left with a free N terminus that could undergo another cycle of the reaction. In the coupling step PITC chemically modified the α-amino group of the N terminal residue to form a phenyl thiocarbamyl (PTC)-polypeptide. At pH 8.0-8.5 coupling was favoured at α-amino groups. In the cleavage step the PTC-N-terminal residue was rapidly cleaved from the polypeptide chain under anhydrous, acidic conditions to liberate two products, an anilinothiozolinone (ATZ)-amino acid and the n-1 polypeptide. The n-1 polypeptide had a reactive N terminus and could undergo another cycle of coupling and cleavage steps. The ATZ-amino acid was extracted from the support and delivered to a small flask. ATZ-derivatives were unstable but could be converted in the third step of the cycle into stable PTH-derivatives. Conversion was a two-step reaction and occurs in aqueous, acidic solutions. First, the ATZ-amino acid was rapidly hydrolyzed to a PTC-amino acid. In the second step, the PTC-amino acid cyclized to a stable PTH-amino acid.

2.9.2 Identification of PTH-amino acids

The PTH-amino acid generated during each sequencing cycle was identified by UV absorbance and HPLC chromatography. All 20 amino acids were easily resolved by gradient elution from a reverse-phase HPLC support. Each amino acid was detected by its UV absorbance at 269 nm and was identified by its characteristic retention time. A chromatogram of the PTH-amino acids was also shown in Figure 16. Modified amino acids could be detected in some cases. The DNA crosslinked amino acid (e.g. Tyr) could be detected by the absence of a PTH-Tyr peak [73,81,82].

The purified peptide-DNA was analyzed by SDS-PAGE (20% acrylamide) and sequenced by Prof J. Fothergill and Mr B. Dunbar of the University of Aberdeen, and Welmet of the University of Edinburgh.

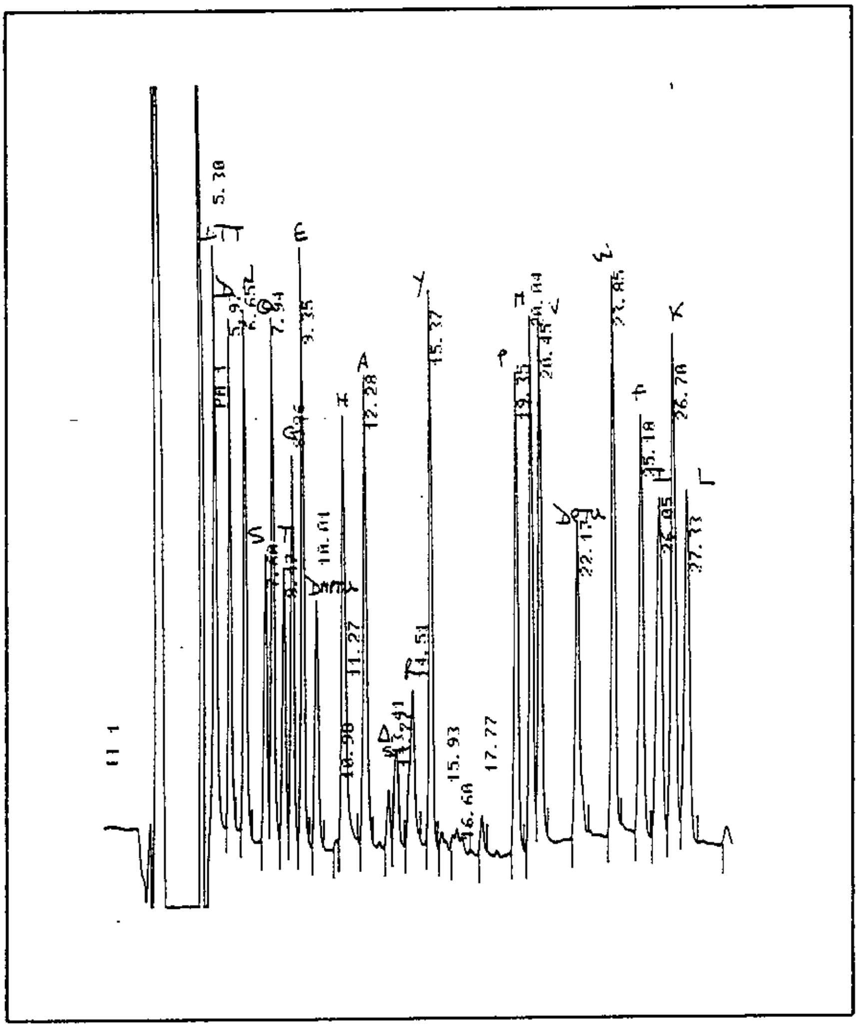


Figure 16. HPLC chromatogram of standard PTH-amino acids Amino acids running time are as follows : Asp (D) = 5.97, Asn (N) = 6.65, Ser (S) = 7.6, Gln (Q) = 7.94, Thr (T) = 8.47, Gly (G) = 8.86, Glu (E) = 9.35, His (H) = 11.27, Ala (A) = 12.28, d-Ser (S) = 13.77, Arg (R) = 14.51, Tyr (Y) = 15.37, Pro (P) = 19.35, Met (M) =

Chapter 3 Results and Discussion

3.1 *4-thiothymidine substituted oligonucleotides*

It has been observed that 4-thiouridine [105] undergoes photo oxidation to unstable intermediates which easily react with nucleophiles. Although the photoreactive mechanism of 4-thiothymidine is not yet clear, polydeoxy-4-thiothymidylic acid was used as a template analogue of *E.coli.* RNA polymerase to investigate the template binding site of the enzyme by crosslinking by UV irradiation at 330-360 nm [106]. Substitution of the target DNA sequence with 4-thiothymidine was shown to increase the yields of photoreactive product of *Eco*RV restriction ENase and modification methylase bound to DNA (340-360 nm irradiation) [77].

In this study 4-thiothymidine was synthesised and was the first base analogue incorporated into the target sequence of *Eco*KI for studying the interactions between the recognition sequence and the *Eco*KI mtase. If the yields of photoreactive products can be enhanced to a sufficient amount, the crosslinked complexes can be purified, then the specific amino acid residues at the interface of the protein-nucleic acid complexes can be identified by sequencing, following proteolysis.

3.1.1 Synthesis of 4-nitrophenylthio-thymidine

The synthesis of 4-nitrophenylthio-thymidine from thymidine involved several reactions which are illustrated in Figures 10 and 11. Firstly, the 3' and 5' hydroxyl groups of the deoxyribose sugar of thymidine were protected to form 3', 5'-bis-O-(methoxyacetyl) thymidine (Figure 10, reaction (I)). The nitrotriazolide was then prepared via the diphenyl phosphate to give 3',5'-bis-O-methoxyacetyl-4-(3-nitrotriazolo) thymidine (Figure 10, reaction (II) and (III)). This product was then treated with DBU and nitrothiophenol to form 4-nitrophenylthio thymidine (Figure 10, reaction (IV)). The reaction proceeded via the attack of the thiolate anion (Figure 11, reaction (I)), produced by addition of DBU, to displace the nitrotriazole (Figure 11, reaction (II)). DBU is a strong base, which reacts with water to produce OH^- which is potentially damaging, therefore redistilled dry DCM was used as the solvent.

The methoxyacetyl protecting groups were removed but not the nitrothiophenyl group by treatment with conc. aqueous ammonia in methanol (1 in 4 (v/v)) (Figure 10, reaction (V)). Stronger alkali, e.g. conc aqueous ammonia or 2 M sodium hydroxide, attacked the 4-nitrothiophenyl group as well as the methoxyacetyl groups resulting in the formation of 5-methyl-2'-deoxycytidine or 4-thiothymidine respectively (Figure 17).

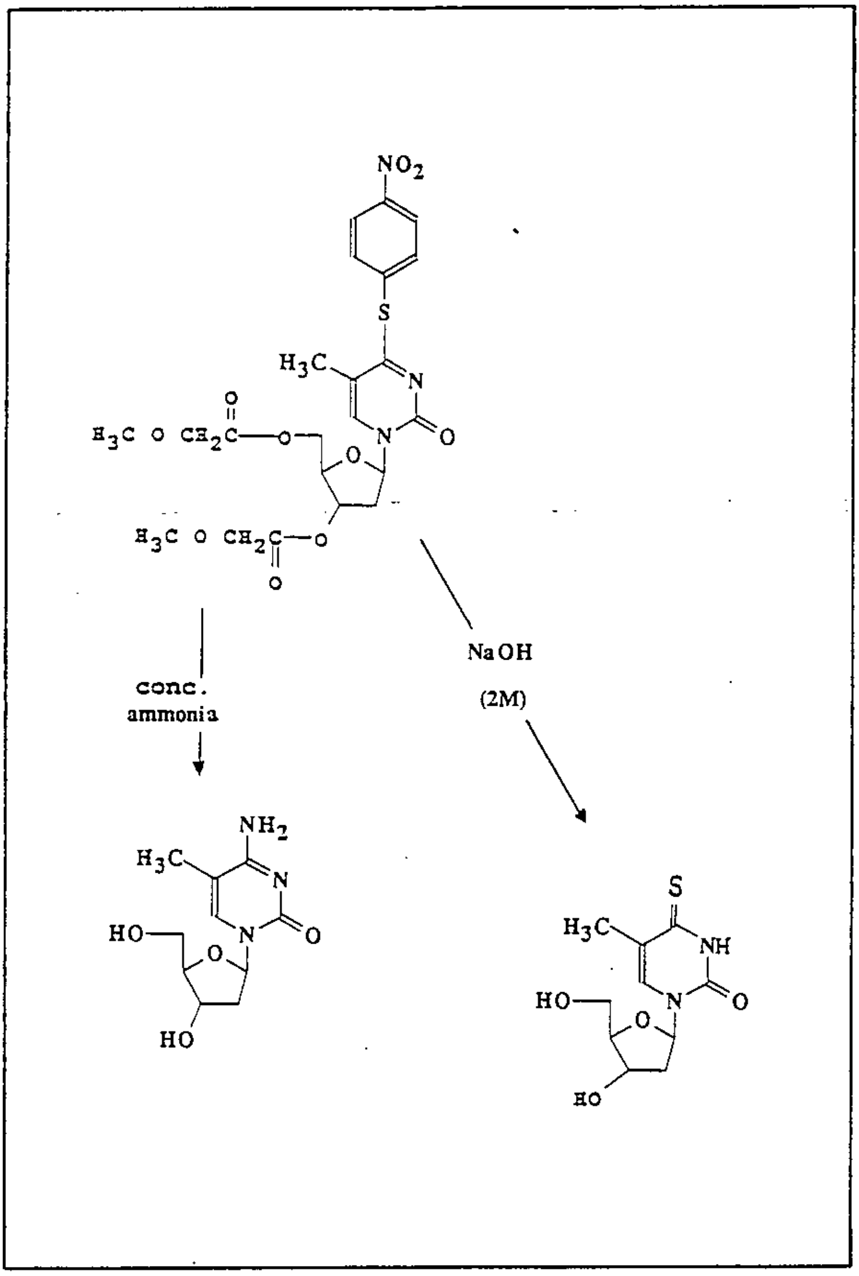


Figure 17. Reaction scheme for the formation of 5′-methyl-2′-deoxycytidine and 4-thiothymidine

3.1.2 Synthesis of oligonucleotides containing 4-thiothymidine

4-Thiothymidine and 4-nitrophenylthio thymidine have been previously synthesized by Connolly *et al* [124] and Nikiforov *et al* [89]. In this study, a different synthetic method was used as in section 2.2.4. Connolly's synthesis of 4-thiothymidine started from the preparation of 3′,5′-O-dibenzoylthymidine. 3′,5′-O-dibenzoyl 4-thiothymidine was then obtained from the reaction of 3′,5′-O-dibenzoylthymidine with Lawesson's reagent. It was then converted to 4-thiothymidine with sodium methoxide, which was in turn directly converted to 5′-dimethoxytrityl derivative [124]. The incorporation of 4-thiothymidine into synthetic oligonucleotide has been reported by Nikiforov and Connolly [89]. The phosphoramidite of 5′-O-(4,4′-dimethoxytrityl)-4-(nitrophenylthio) thymidine was used in the incorporation of 4-thiothymidine into oligonucleotides. Since the sulphur atom of 4-thiothymidine is nucleophilic and reacts with the phosphoramidites, the nitrophenyl group was used for its protection during incorporation into oligonucleotide. 5′-O-(4,4′-dimethoxytrityl)-4-(nitrophenylthio) thymidine was produced from 5′-O-(4,4′-dimethoxytrityl)-4-thiocyanatothymidine which could not be used as a protected form of 4-thiothymidine during the incorporation of 4-thiothymidine into synthetic oligonucleotide, due to the high reactivity of the thiocyanato function [89]. 5′-O-(4,4′-dimethoxytrityl)-4-thiocyanatothymidine was obtained from the reaction of 5′-O-(4,4′-dimethoxytrityl)-4-thiothymidine with cyanogen bromide (CNBr) [125]. After incorporation of 5′-O-(4,4′-dimethoxytrityl)-4-thiophenylnitro thymidine into oligonucleotides, $tBuNH_2$, $KSCOCH_3$ and NH_3 were used for the deprotection/deblocking step instead of the automatic conc. aq. NH_3 deprotection/deblocking step [89]. The UV crosslinking of 4-thiothymidine substituted synthetic dodecamers containing the recognition site d(GATATC) of the enzymes *Eco*RV ENase and *Eco*RV mtase at 340-360 nm has been reported by Nikiforov & Connolly [77]. The yields of crosslinked product were up to 35% with the mtase and up to 6% with ENase.

3.1.2.1 The coupling efficiency of test sequences

The coupling efficiency of the phosphoramidite of 4-nitrophenylthio thymidine in both test sequences $T_8T^{4S}T_3$ and $CT^{4S}CT^{4S}CC$ was more than 95% as determined by measurement of the diluted trityl colour by visible spectrometry at 495 nm. The two test sequences $T_8T^{4S}T_3$ and $CT^{4S}CT^{4S}CC$ were then used to study the deprotection/deblocking step of the synthetic oligonucleotide containing 4-thiothymidine.

3.1.2.2 Trityl on and trityl off ending methods

The trityl off ending method was used to obtain the oligonucleotide unprotected at the 5′-terminal. The oligonucleotide of trityl on ending method bore a dimethoxytrityl group on the 5′-terminal nucleotide, which gave a better separation from failure sequences during HPLC purification. Gel electrophoresis showed that trityl on and trityl off 4-thiothymidine substituted oligonucleotides were both pure, therefore the trityl on was not necessary in the synthesis of 4-thiothymidine substituted oligonucleotide.

3.1.2.3 Deprotection of test sequences

Reversed-phase HPLC showed that two products were obtained from the test sequence $T_8T^{4S}T_3$ (section 2.2.5.1), with conc. aqueous ammonia deprotection, while only one product was obtained from the $tBuNH_2$, $KSCOCH_3$ and NH_3 deprotection method.

The Trityl Off Auto and the Trityl Off Manual deprotection methods were also applied on the test sequence $CT^{4S}CT^{4S}CC$. Reversed-phase HPLC analysis showed that one oligonucleotide product was obtained from both conc. aqueous ammonia deprotection and $tBuNH_2$, $KSCOCH_3$ and NH_3 deprotection.

3.1.2.4 Analysis of the enzyme digested test sequences $CT^{4S}CT^{4S}CC$

Test sequence $CT^{4S}CT^{4S}CC$, which was deprotected with conc. aqueous NH_3, was digested by enzymes (snake venom phosphodiesterase and alkaline phosphatase) and analysed by HPLC at 254 nm. The digested $CT^{4S}CT^{4S}CC$ (Figure 18 (II)) was shown to be deoxycytidine and 5-methyl-2'-deoxycytidine in a ratio of 2:1 by comparison HPLC chromatograms with standard (Figure 18 (I) and 19). A mixed injection of standard 5-methyl-2'-deoxycytidine (1 OD) and the digested $CT^{4S}CT^{4S}CC$ (1 OD) gave two peaks at the same retention times (Figure 18 (III)) as that of the digested $CT^{4S}CT^{4S}CC$ (Figure 18 (II)). This illustrates that the digested $CT^{4S}CT^{4S}CC$ contained 5-methyl-2'-deoxycytidine and deoxycytidine instead of the desired 4-thiothymidine and deoxycytidine. Ammonolysis of the 4-position of thymine has occurred during the treatment with conc. aqueous ammonia on the automatic deprotection/deblocking step. Therefore it is concluded that the manual deprotection step of $tBuNH_2$, $KSCOCH_3$ and NH_3 treatment [89] was essential for the incorporation of 4-thiothymidine into oligonucleotides.

3.1.2.5 The stability of 4-thiothymidine to heat treatment with conc. NH_3

TLC (CH_2Cl_2:MeOH=1:1) showed that 10 minutes after treatment with conc. NH_3 at room temperature, 4-thiothymidine was still stable. (R_f of 4-thiothymidine = 0.86, R_f of 5-methyl-2'-deoxycytidine = 0.41), while 50% of 4-thiothymidine was converted to 5-methyl-2'-deoxycytidine by ammonolysis when heated at 55° C for 3 hours. Hence deprotecting the oligonucleotide containing 4-thiothymidine with $tBuNH_2$, $KSCOCH_3$ and NH_3 manually [89] was employed in this study.

3.1.2.6 Analysis of oligonucleotides containing 4-thiothymidine

The UV spectra between 200 nm and 450 nm of (I) 4-thiothymidine, (II) unsubstituted oligonucleotide (45 mer) and (III) 4-thiothymidine substituted oligonucleotide (45 mer) are shown in Figure 20. 4-Thiothymidine has a relatively high UV absorbance at 333 nm. The relative UV absorbance of 4-thiothymidine at 333 nm and at 254 nm is 10 to 1. Therefore, reversed-phase HPLC analysis of enzymes digested 4-thiothymidine substituted DNA was performed at 333 nm and results given in Figure 21. The analytical reversed-phase HPLC chromatograms showed that (I) standard 4-thiothymidine eluted at 20.5 min, (II) no nucleoside eluted at 20.5 min from unsubstituted DNA, only the digestive enzymes were detected at the beginning of elution, (III) 4-thiothymidine eluted at 20.5 min from 4-thiothymidine substituted oligonucleotide and the digestive enzymes were observed at the beginning of elution.

The size of *Eco*KI mtase binding site on DNA have been analysed by Powell *et al* [91]. It was demonstrated that the specific 25 mers gave equally good binding to the specific 45 mers with K_d values < 5 nM [91]. Therefore, instead of 45 mers, 25 mers were used in later experiments.

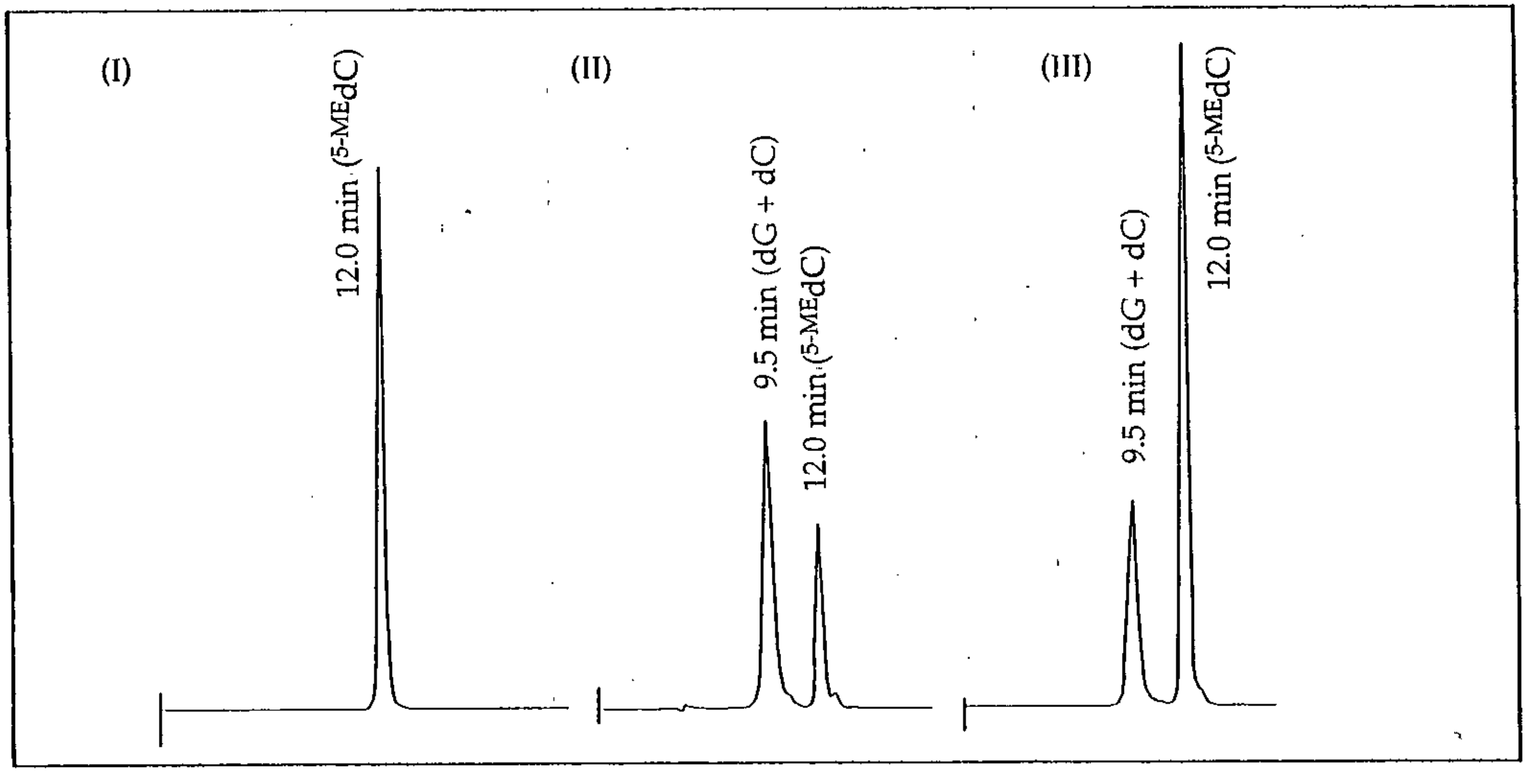


Figure 18. Analytical reversed-phase HPLC chromatograms (290 nm).
(I) 1OD of standard 5-methyl-2′-deoxycytidine ($^{5\text{-ME}}$dC) eluted at 12.0 min.
(II) 1OD of digested synthesis column 3 test sequence CT^{4S}CT^{4S}CC eluted at 9.5 min and 12.0 min.
(III) The mixed injection of (I) 1OD and (II) 1OD eluted at 9.5 min and 12.0 min.

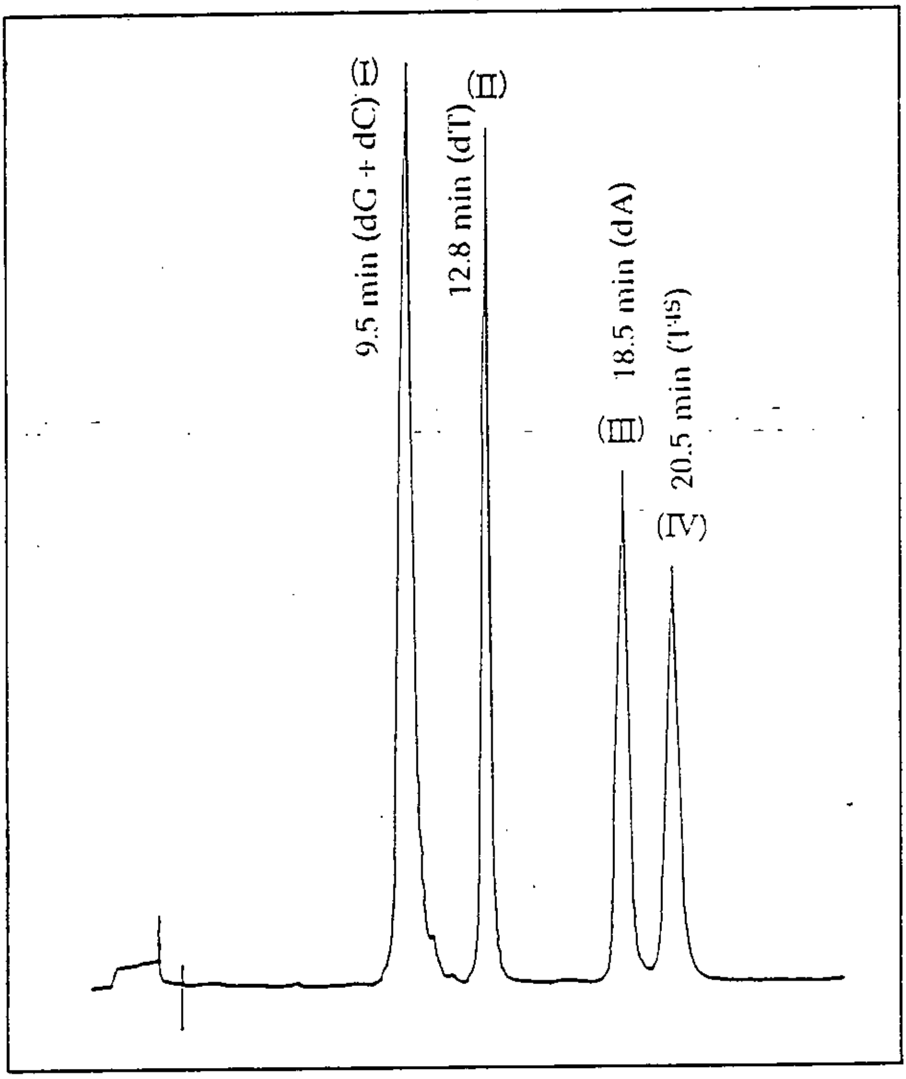


Figure 19. Analytical reversed-phase HPLC chromatogram (264 nm) of standard oligonucleotides
(I) deoxyguanosine (dG) + deoxycytidine (dC), elution time = 9.5 min,
(II) deoxythymidine (dT), elution time = 12.8 min,
(III) deoxyadenosine (dA), elution time = 18.5 min and
(IV) 4-thiothymidine (T^{4S}), elution time = 20.5 min.

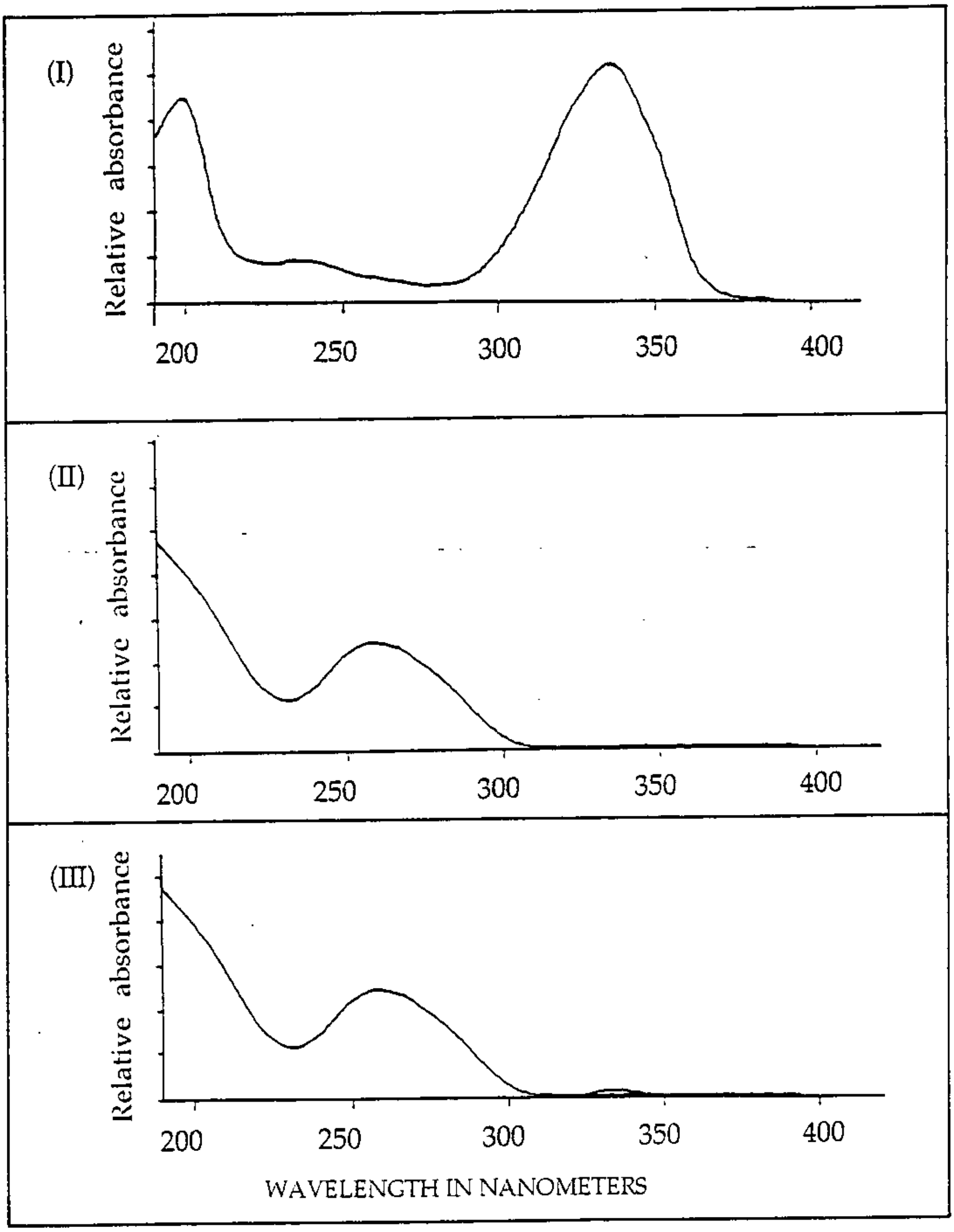


Figure 20. The UV spectra of (I) 4-thiothymidine, (II) unsubstituted oligonucleotide and (III) 4-thiothymidine substituted oligonucleotide.

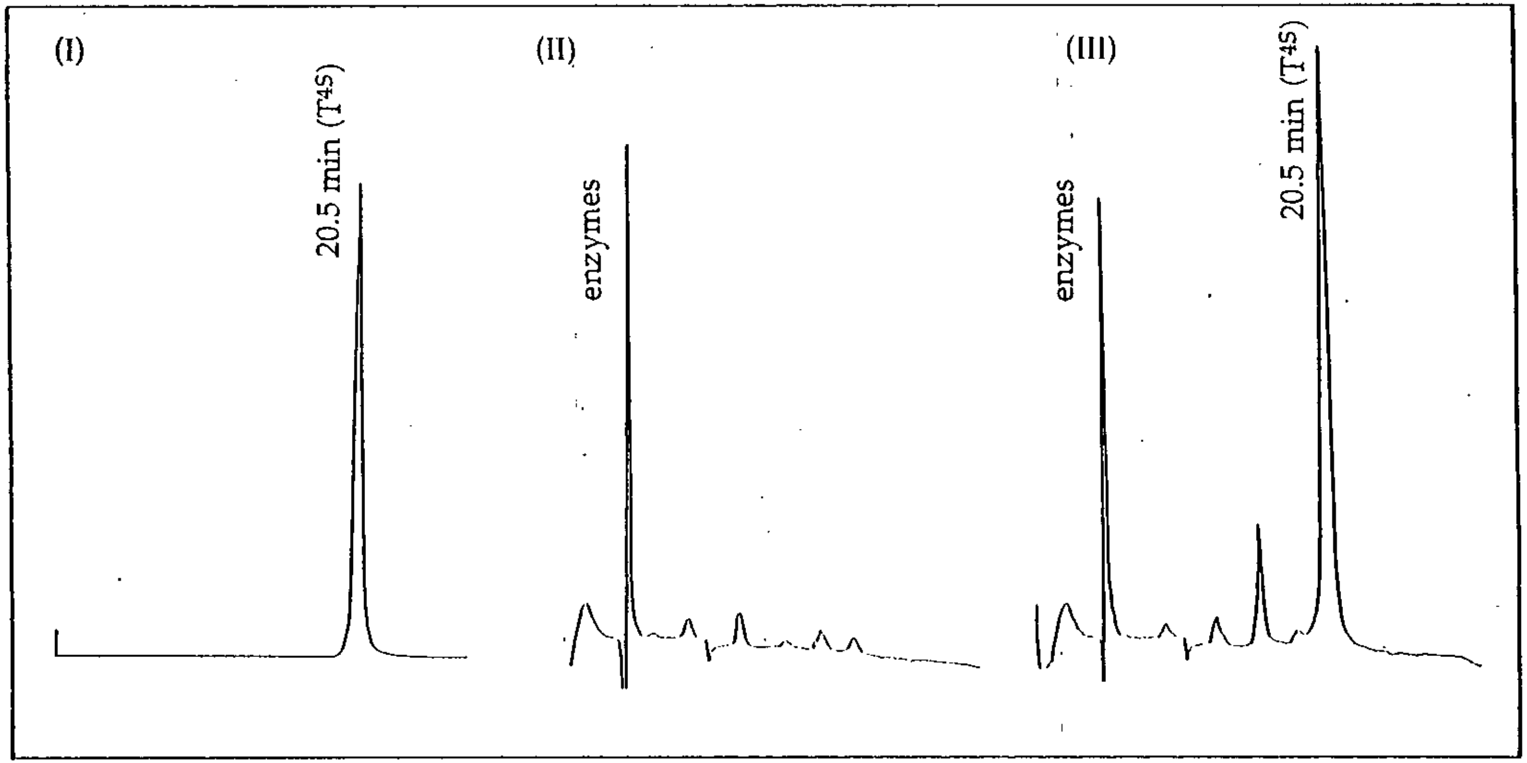


Figure 21. Analytical reversed-phase HPLC chromatograms (333nm) of
(I) standard 4-thiothymidine (elution time = 20.5 min),
(II) digested unsubstituted oligonucleotide and
(III) digested 4-thiothymidine substituted oligonucleotide (4-thiothymidine elution time = 20.5 min).
(II) and (III) were digested by phosphodiesterase and alkaline phosphatase at 37°C for 12 h.

3.1.3 UV induced crosslinking of 4-thiothymidine substituted DNA

4-Thiothymidine (UV λ_{max} pH7 335 nm. UV literature λ_{max} pH7 335 nm [126]) has a strong UV absorbance between 310 nm and 350 nm (Figure 20). For the photoreaction of 4-thiothymidine substituted oligonucleotides with *Eco*KI mtase, the wavelengths of 365 nm, 320 nm (cut off filter) and 254 nm were investigated.

Figure 22 illustrates the substitution of 4-thiothymidine at a single position (position 6, 17, 21 of the top strand of 25 mers and position 7′, 8′ and 11′ of the bottom strand of 25 mers) within the complementary specific oligonucleotides to give duplexes, designated 4-thio-6, 4-thio-17, 4-thio-21, 4-thio-7′, 4-thio-8′ and 4-thio-11′.

The efficiencies of crosslinking of 4-thiothymidine substituted oligonucleotides at different wavelength of UV irradiation were compared in Table 3. Results showed that less than 1% of crosslinked complexes were observed from S25 (unsubstituted DNA), 4-thio-17, 4-thio7′, 4-thio-8′ 4-thio-11′ and Br7′ at 320nm. The crosslinking efficiencies of S25, 4-thio-6, 4-thio-17, 4-thio-21, 4-thio7′, 4-thio-8′ 4-thio-11′ and Br7′ at 365nm were also less than 1%. No detectable crosslinking was observed from NS25 (non-specific 25 mers of the target recognition DNA for *Eco*KI). A higher percentage of crosslinking was obtained at a shorter wavelength of 254 nm for all specific oligonucleotides, NS25 still did not crosslink. The efficiency of crosslinking was 4% for S25, 4-thio-6, 4-thio-17 and 4-thio-21, 2% for 4-thio-7′, 4-thio-8′ and 4-thio-11′and 12% for Br7′. Although a wavelength of 320 nm or 365 nm should be suitable for the UV crosslinking of 4-thiothymidine, the present results showed that the crosslinking of 4-thiothymidine substituted DNA did not occur to any appreciable extent at these wavelengths.

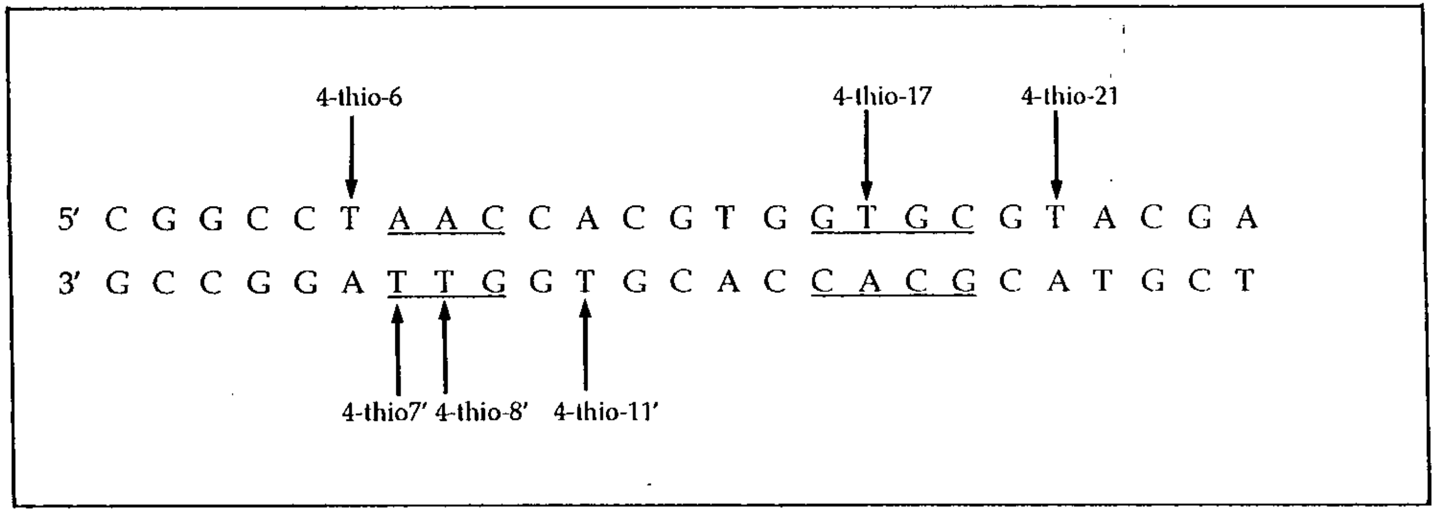


Figure 22. The DNA duplex contains the *Eco*KI target sequence (underlined). The sites of substitution of thymine by modified bases are indicated and numbered from the 5′ end of the top strand with the addition of a prime (′) indicating a substitution on the bottom strand. The duplexes containing a single substitution of 4-thiothymidine are designated 4-thio-6, 4-thio-17, 4-thio-21, 4-thio-7′, 4-thio-8′ and 4-thio-11′.

Table 3. UV irradiated crosslinking of DNA-mtase at different wavelength.

UV wavelength	254nm	302nm	320nm	365nm
Duplex	degree of crosslinking			
S25	4%	<1%	<1%	<1%
NS	0%	N	N	0%
4-thio-6	~4%	N	N	<1%
4-thio-17	~4%	N	<1%	<1%
4-thio-21	~4%	N	N	<1%
4-thio-7′	<2%	N	<1%	<1%
4-thio-8′	<2%	N	<1%	<1%
4-thio-11′	<2%	N	<1%	<1%
Br7′	12%	<2%	<1%	<1%

S25 = unsubstituted oligonucleotide
NS = non-specific oligonucleotide,
4-thio = 4-thiothymidine substituted oligonucleotides
Br7′ = 5-bromodeoxyuridine substituted oligonucleotide.
N = experiment not done.

Figure 23 illustrates the structure of unsubstituted *Eco*KI mtase target sequence (I) and 4-thio-7′ *Eco*KI mtase target sequence (II). The crosslinking efficiencies of 4-thiothymidine substitutions in the bottom strand of the DNA (4-thio-7′, 4-thio-8′ and 4-thio-11′) (approximately 2%, Table 3), were less than that of S25 (4%). This indicated that the substitution of 4-thiothymidine prevented the hydrogen bond between 4 keto oxygen of thymidine and the N6 hydrogen of adenine from forming (Figure 23 (I)). A change in DNA conformation could affect the precise contact between the DNA and the mtase, leading to a decrease in crosslinking efficiency. When 4-thiothymidine substitutions (4-thio-6, 4-thio-17 and 4-thio-21) were in the top strand of the DNA, the degree of crosslinking was about 4%, similar to that of the S25 (4%). This indicated that the substitution sites were not the key positions of the oligonucleotide which was bound by the mtase. Therefore, the crosslinking efficiency was not effected by the substitution of 4-thiothymidine at those sites and was the same as the unsubstituted DNA.

The substitution of 4-thiothymidine did not improve the UV crosslinking efficiency at any substituted position (4-thio-6, 4-thio-17 and 4-thio-21, 4-thio-7′, 4-thio-8′ and 4-thio-11′) at any wavelength (254 nm, 320 nm and 365 nm). This indicated that the UV irradiated crosslinking of DNA-*Eco*KI mtase did not occur between the substituted 4-thio atom of DNA and the mtase (Figure 23 (II)). In contrast, the 4% crosslinking efficiency of the unsubstituted oligonucleotide (S25) with mtase and the highly improved crosslinking efficiency (12%) of Br7′ with mtase at 254 nm implied that the UV induced crosslinking of *Eco*KI mtase-DNA may have occurred only between the mtase and the 5-methyl group of unsubstituted oligonucleotide (Figure 23 (II)) or the 5-bromine of the BrdU substituted DNA at 7′ position (Figure 30). Since no significant degree of crosslinking occurred between the substituted 4-thio atom of DNA and the mtase, the BrdU substituted oligonucleotide and a wavelength of 254 nm were used in subsequent experiments.

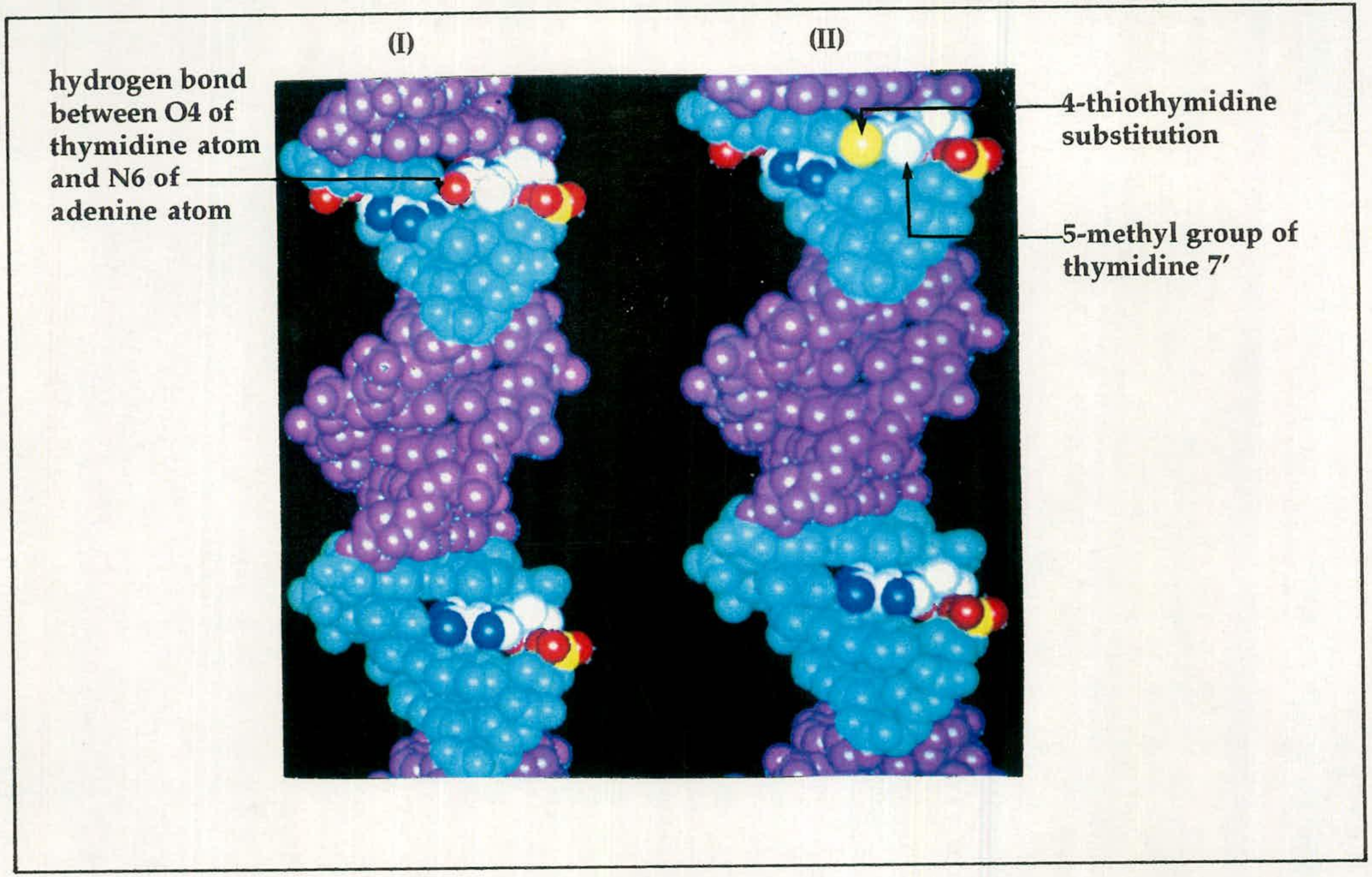


Figure 23. Models of (I) the unsubstituted *Eco*KI target sequence, (II) the 4-thiothymidine substituted *Eco*KI target sequence. A sulphur atom (yellow ball) substituted 4-oxygen (red ball) of the thymidine nucleotide at T7′ (highlighted) of the recognition sequence, AAC------GTGC, (light blue balls with the highlighted methylatable adenines, T8 and T17′). The substituted sulphur atom interferes with the formation of one hydrogen bond between the A-T base pair. It is in the major groove next to the 5-methyl group of the nucleotide, T7′, involved in DNA-mtase crosslinking .

3.2 *Preparation and purification of EcoKI mtase*

*Eco*KI mtase was over-expressed using a recombinant plasmid in *E. coli.* and isolated as described in section 2.5.1. The protein was purified through three columns. Results of each column purification are shown as follows:

3.2.1 CM-Sepharose column purification of *Eco*KI

An UV chromatogram, Figure 24, at 280 nm showed that there was a considerable amount of protein elution during sample loading and washing with NaCl-free Tris-MES buffer. Three peaks of proteins appeared on the eluted fractions of 28, 34 and 53 near the end of the NaCl gradient elution.

The eluted fractions (1, 5, 10, 15, 20, 25, 30, 35, 40, 45) of the cell extract and the NaCl-free Tris-MES buffer washing before the gradient buffer were analysed by 12% SDS-PAGE in gel I of Figure 25. The eluted fractions (65, 70, 75, 80, 85, 90, 92, 94, 96, 98, 100) of the NaCl gradient buffer elution were analysed on 12% SDS-PAGE in gel II of Figure 25. As illustrated by Coomassie Blue staining of gel II of Figure 25 fractions 75 to 90 contained polypeptides with mobility consistent with those of the mtase and were pooled for further purification. The amount of protein was estimated to be approximately 58 mg by UV spectrophotometry.

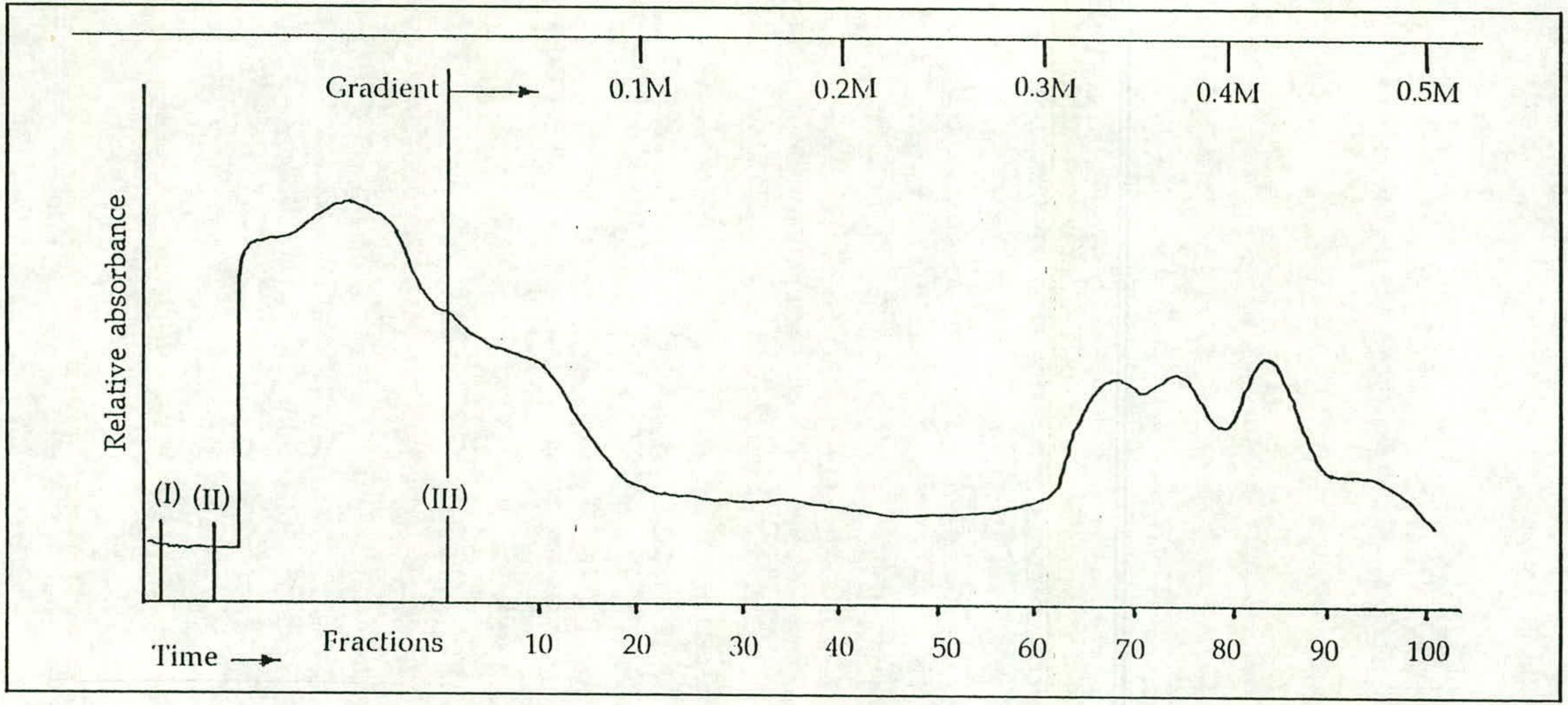


Figure 24. The UV chromatogram (280 nm) of *Eco*KI mtase purification by the CM-Sepharose column.
(I). Cell extract was applied to the column.
(II). After loading cell extract, the column was washed with 200 ml of Tris-MES buffer (Tris-MES 20 mM, $MgCl_2$ 10 mM, EDTA 0.1 mM, β-mercaptoethanol 7 mM, pH 6.5)
(III). Protein was eluted with a gradient of 0-0.5M NaCl in 500 ml of Tris-MES buffer (20 mM, pH 6.5) at a flow rate of 24.0 ml/h overnight and the eluting fractions were collected every 12 minutes.

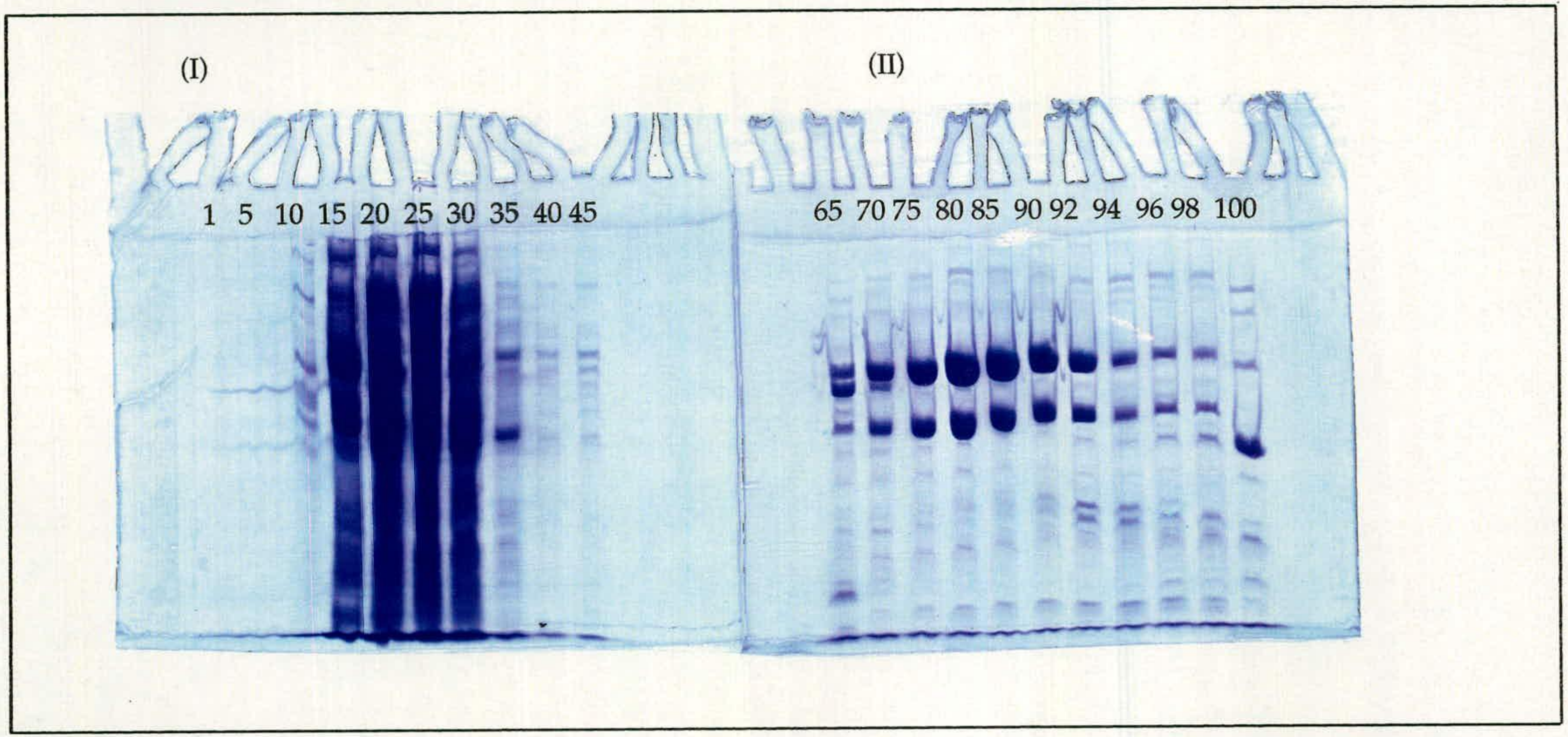


Figure 25. The SDS-PAGE (12% acrylamide) of protein eluted from the CM-Sepharose column following Coomassie Blue (0.1%) staining. Samples of eluting fractions (1, 5, 10, 15, 20, 25, 30, 35, 40 and 45) on gel (I) contained protein that did not bind to the column upon loading and was washed through NaCl free buffer. Samples on gel (II) were gradient fractions 65 to 100. Fractions 75 to 90 containing *Eco*KI mtase were pooled for further purification.

3.2.2 Heparin-Agarose column purification of *Eco*KI

Two high UV absorbance areas were observed between fractions 38-50 on the 280nm UV absorbance chromatogram (Figure 26) from the NaCl gradient elution of the Heparin-Agarose column. Mtase containing fractions 40 to 47 were pooled after the mtase purity were demonstrated by Coomassie Blue stained 12% SDS-PAGE gels (Figure 27). Estimating the amount of mtase from the UV spectrum, approximately 36 mg of mtase was recovered after purification using the Heparin-Agarose column.

3.2.3 DEAE-Sepharose column purification of *Eco*KI

Two high UV absorbance areas were obtained around fractions 47 and 57 on the UV chromatogram at 280 nm for the DEAE-Sepharose column (Figure 28). The first species is M_1S_1, the second M_2S_1 were reported previously by Dryden *et al* [13]. The separation and purity of M_1S_1 and M_2S_1 was shown on the Coomassie Blue stained 12% SDS-PAGE gels (Figure 29). M_1S_1 and M_2S_1 in fractions 46 to 49 and fractions 56 to 59 respectively were pooled. The amount of M_1S_1 and M_2S_1 recovered was 2.5 mg and 9.2 mg respectively.

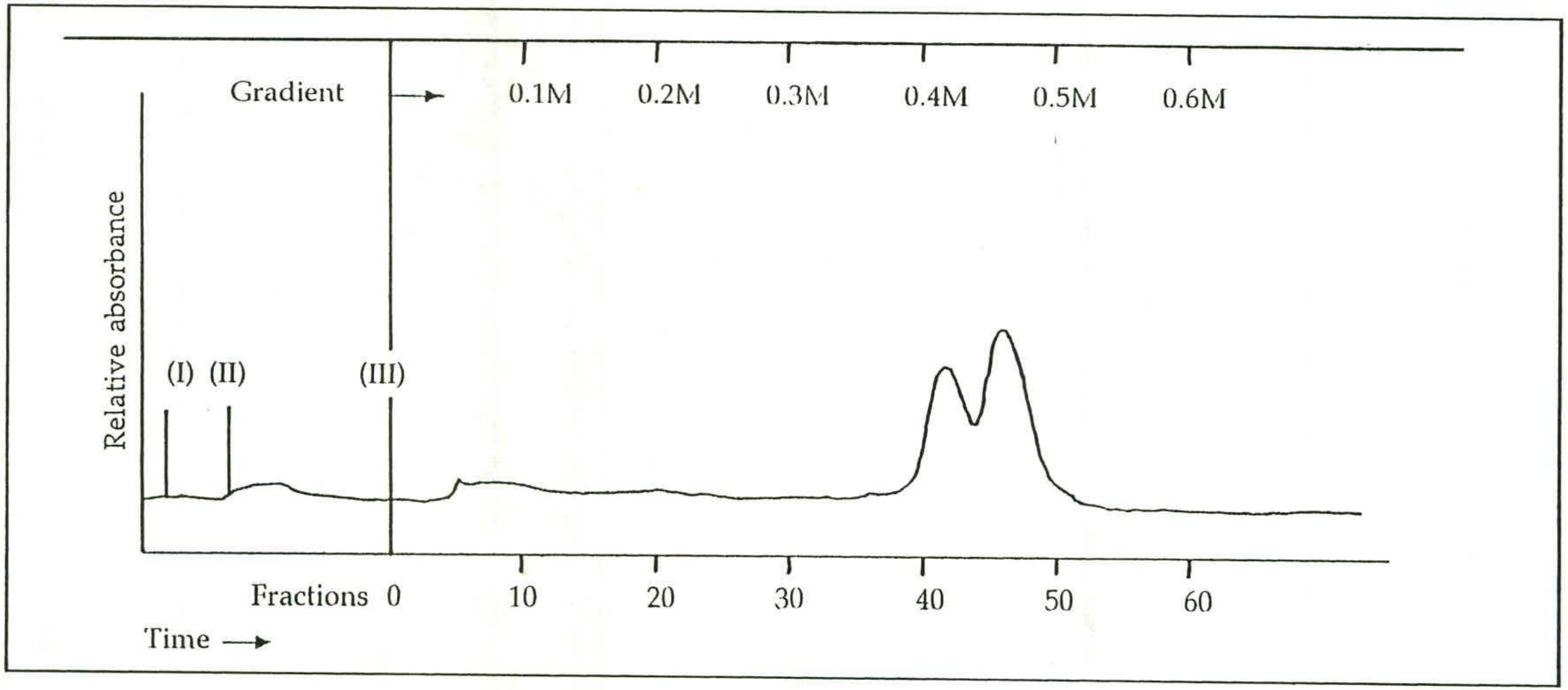


Figure 26. The UV chromatogram of *Eco*KI mtase containing fractions from the previous column applied to the Heparin-Agarose column at 280 nm.
(I). Pooled fractions of *Eco*KI mtase were applied to the column.
(II). After the sample was loaded, the column was washed with 200 ml of Tris-MES buffer (Tris-MES 20 mM, $MgCl_2$ 10 mM, EDTA 0.1 mM, β-mercaptoethanol 7 mM, pH 8.0).
(III). Protein was eluted with a gradient of 0-1.0 M NaCl in 500 ml of Tris-MES buffer (20 mM, pH 8.0) at a flow rate of 24.0 ml/h overnight and the eluting fractions were collected every 12 minutes.

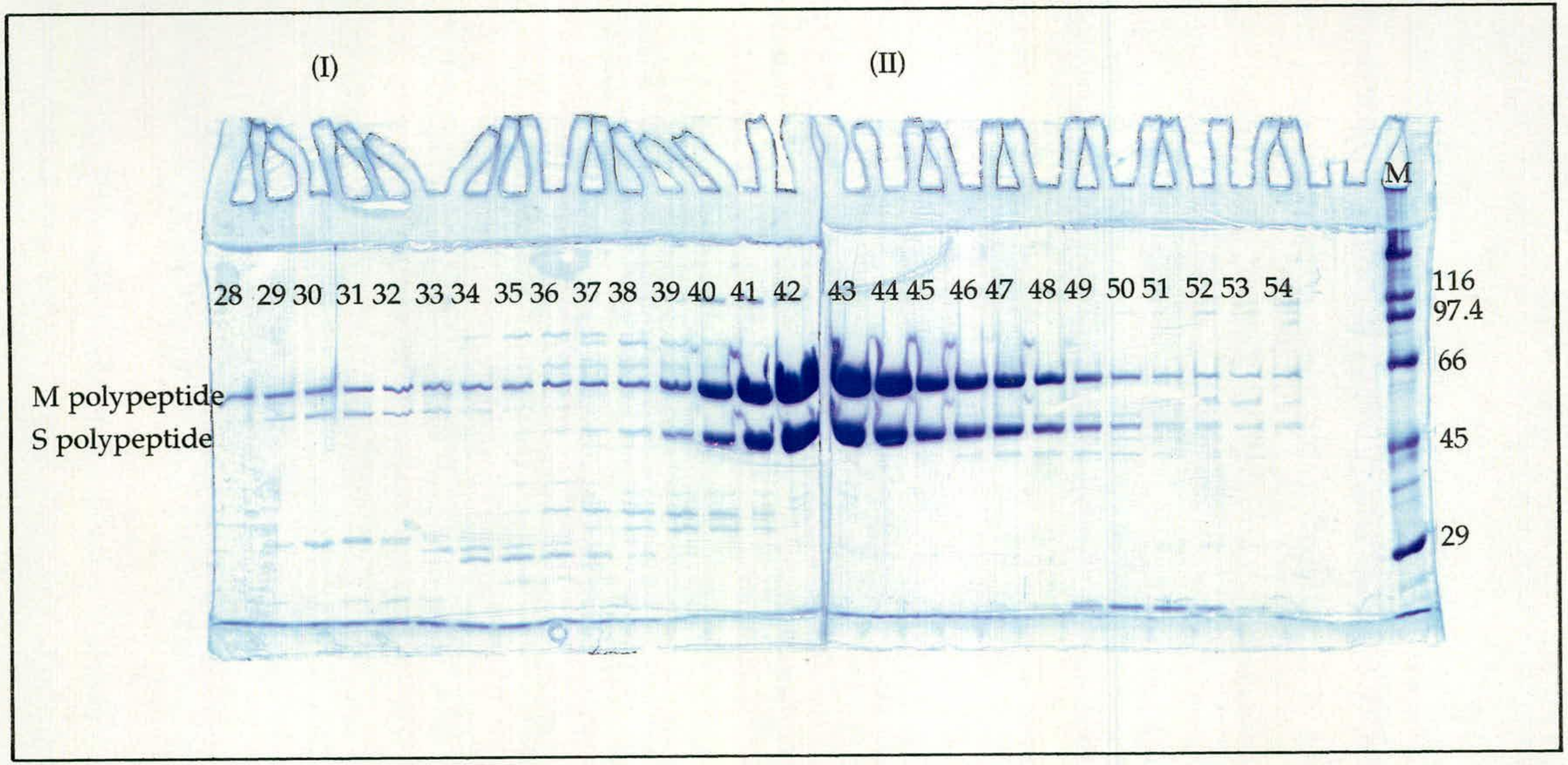


Figure 27. The SDS-PAGE (12% acrylamide) gels of fractions eluted from the Heparin-Agarose column. Protein was visualised by Coomassie Blue (0.1%) staining. Samples on gel (I) were gradient fractions 28 to 42. Samples on gel (II) were fractions 43 to 54. Fractions 40 to 47 containing *Eco*KI mtase were pooled for further purification. The molecular weights (kD) of marker (M) indicated at side.

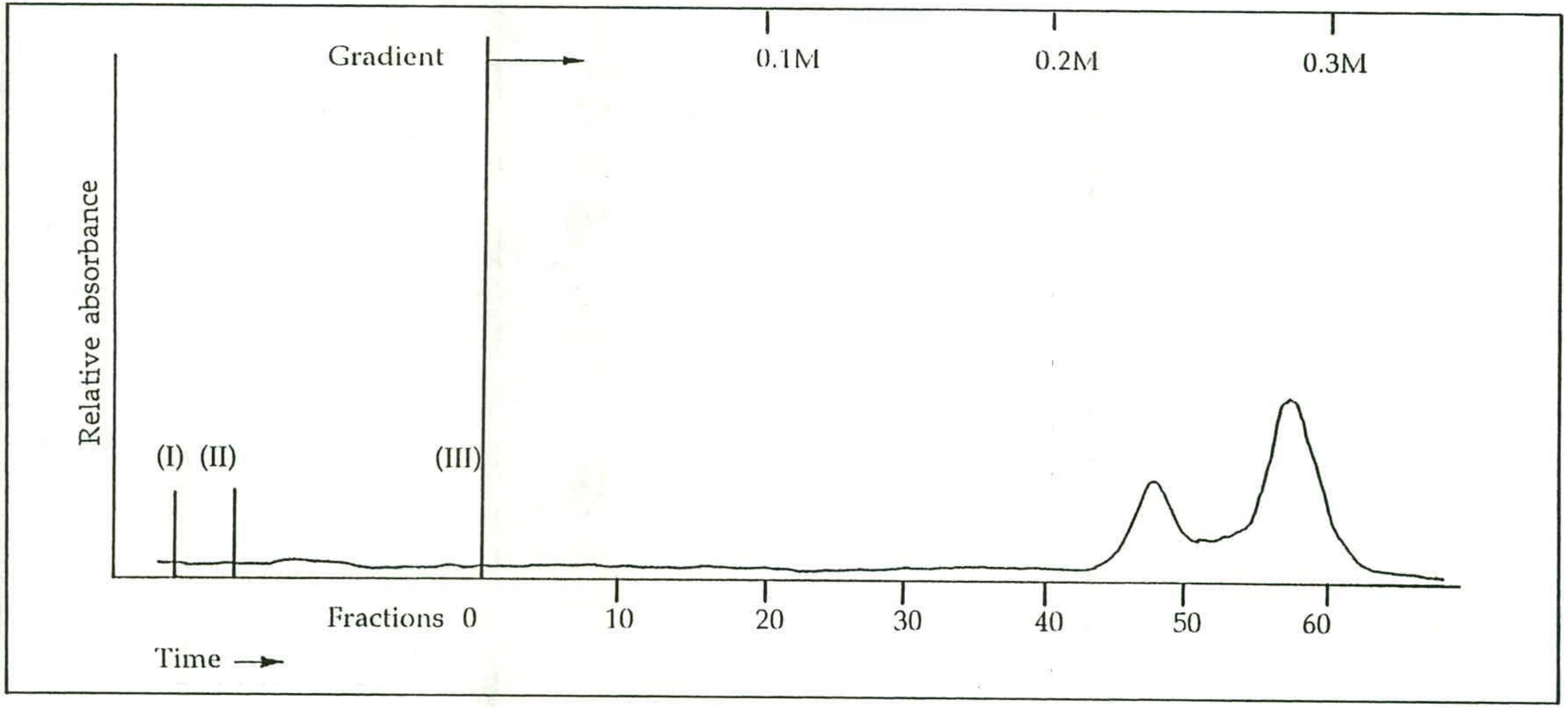


Figure 28. The UV chromatogram (280 nm) of the purification of pooled *Eco*KI mtase by the DEAE column.
(I). Pooled *Eco*KI mtase was applied to the column.
(II). After loading sample, the column was washed with 200 ml of Tris-MES buffer (Tris-MES 20 mM, $MgCl_2$ 10 mM, EDTA 0.1 mM, β-mercaptoethanol 7 mM, pH 8.0).
(III). Protein was eluted with a gradient of 0-0.4M NaCl in 500 ml of Tris-MES buffer (20 mM, pH 8.0) at a flow rate of 24.0 ml/h overnight and the eluting fractions were collected every 12 minutes.

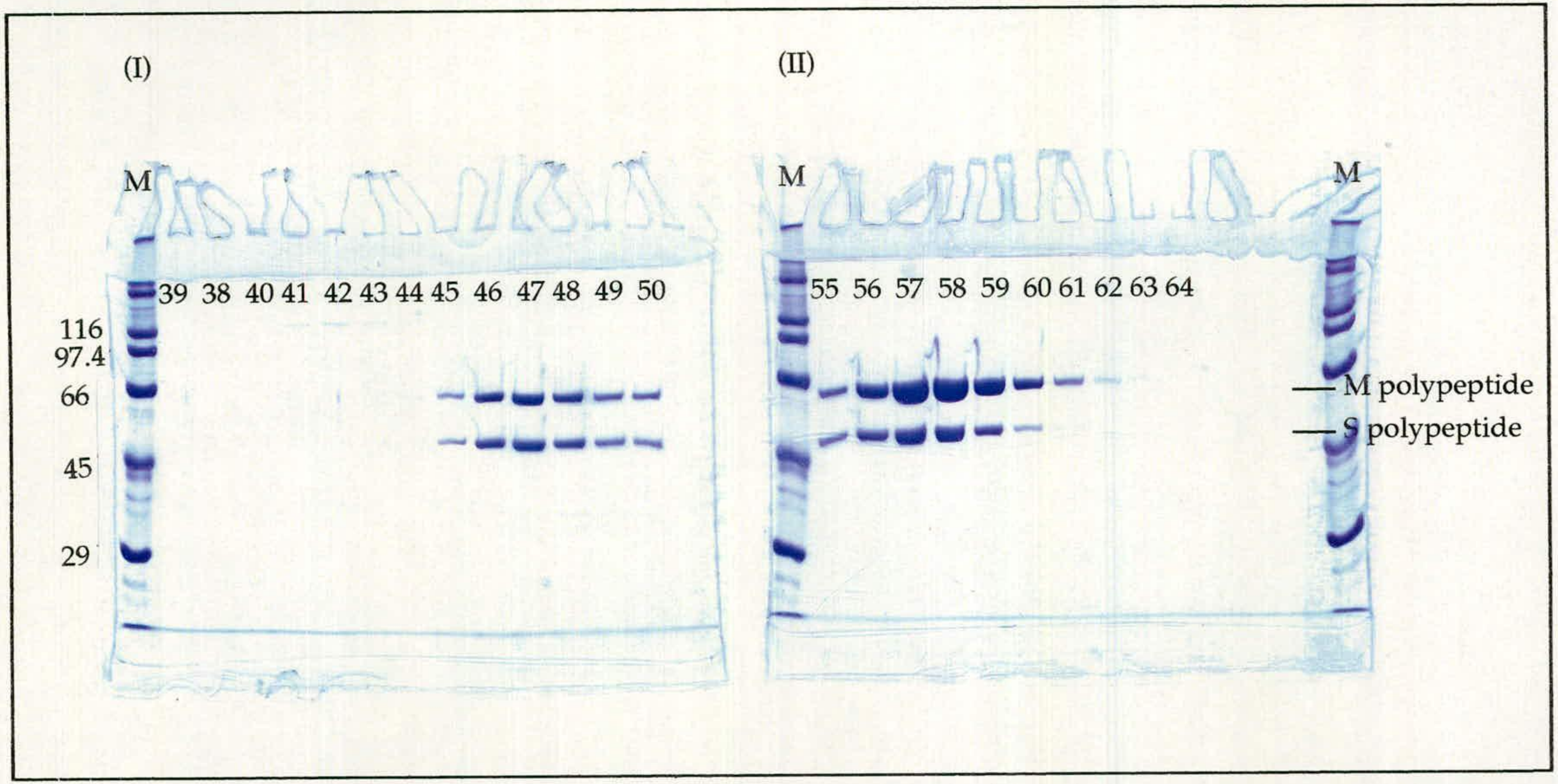


Figure 29. The SDS-PAGE (12% acrylamide) gel of fractions eluted from the DEAE-Sepharose column. Protein was visualised by Coomassie Blue (0.1%) staining. Gel (I) started with the molecular marker followed by the selective gradient fractions 38 to 50. The molecular marker were on two sides of gel (II) and fractions of 55 to 64 were in between. Fractions 46 to 49 of gel(I) and fractions 56 to 59 of gel (II) containing purified M_1S_1 and M_2S_1 were pooled respectively. The molecular weights (kD) of marker (M) are indicated at side.

3.3 *Gel retardation assays*

3.3.1 The binding efficiencies of 5-BrdU substituted oligonucleotides

Since the 5-BrdU substituted oligonucleotide was more successful in the UV irradiated DNA-mtase crosslinking experiments, the binding of 5-BrdU substituted oligonucleotides was studied. Gel retardation studies indicated similar binding efficiencies (Table 4) for Br6, Br14, Br17, Br21, Br7′, Br8′, Br11′, Br22′, Br25′ and S25 (74%-94%). Similar binding efficiencies were expected at 10 nM DNA and 200 nM mtase, because the DNA would have been saturated by mtase given the estimated binding constant (Kd) of < 10 nM. The binding affinity comparisons in section 3.3.2 show Kd < 10 nM when DNA was at a concentration of 0.1 nM.

No DNA-mtase binding complex was observed on non-denaturing PAGE for the non-specific DNA (10 nM) when incubated with mtase at a concentration lower than 200 nM, but a slower-migrating complex was observed for both specific and non-specific DNA-mtase reaction mixtures when the concentration of mtase was higher than 250 nM with 10 nM DNA in the presence of AdoMet (100 μM). This slower-migrating non-specific DNA-mtase binding complex was also observed for specific and non-specific 45 mers [91]. The formation of the non-specific binding complexes at high mtase concentration might be due to Van der Waals interactions, or salt bridges between the negatively charged phosphates of DNA and the positively charged amino acid side chains. These non-specific complexes have a higher tendency to dissociate over time than the specific complex when observed over 24 hours (results not shown). No crosslinked complex was obtained from the non-specific binding complex when 254 nm UV irradiation was applied. This implied that the conformation of the specific target sequence allowed the mtase to precisely bind to the target sequence, with some amino acid side chains in close contact with the specific nucleotides. These specific and close contacts of

DNA and mtase were responsible for the formation of crosslinked complex when UV irradiation was applied.

Table 4. Efficiency of binding and crosslinking of S25 and 5-BrdU substituted oligonucleotides.

Duplex	Percentage of duplex bound		Percentage of duplex crosslinked	
S25	84 ± 4	(5)	3.9 ± 0.2	(6)
Br500	92	(1)	12.2	(1)
Br6	83 ± 3	(9)	3.0	(1)
Br7′	79 ± 4	(5)	12.3 ± 0.2	(5)
Br8′	87 ± 7	(2)	4.3 ± 0.3	(2)
Br11′	77 ± 15	(2)	4.1 ± 0.2	(2)
Br14	74 ± 4	(5)	3.8	(1)
Br17	78 ± 4	(5)	3.0	(1)
Br21	75 ± 5	(5)	4.1	(1)
Br22′	84 ± 8	(2)	4.4 ± 0.1	(2)
Br25′	77 ± 13	(2)	4.3 ± 0.1	(2)

Mtase (200 nM) with premixed AdoMet (100 μM) and DNA (10 nM) were added in Tris-HCl buffer (20 mM pH 8.0, NaCl 100 mM) containing 5% glycerol and mixed with microcentrifuge. The total volume of reaction mixture was 200 μl. After 15 minutes of incubation at 22°C a portion of the reaction mixture (20 μl) was loaded on a non-denaturing polyacrylamide gel (5%) and run for 1 to 1.5 h at 30 mA for gel retardation assays. After incubation portions of the reaction mixture (20 μl) were UV irradiated at 254nm for 10 minutes, then was loaded on a denaturing SDS-polyacrylamide gel (12%) and run for 3 h at 45 mA to analyse the DNA-mtase crosslinked complexes.

Binding and crosslinking efficiency are represented by the fraction of radiolabelled DNA, determined by scintillation counting, in the bound or crosslinked complex relative to the total amount of DNA. The number of measurements is shown in brackets.

3.3.2 The binding affinities of S25, Br7′, dU7′ and Io7′ with mtase

After determining that the crosslinking of DNA-mtase only occurred at the residue complementary to the first adenine in the -AAC- sequence, which is thymidine 7′ of 25 mer, different substituents were used at this site. The substituents Br7′ (Figure 30), dU7′ and Io7′ at the thymidine 7′ position of S25 are structurally very similar with different functional groups at the 5-C of the pyrimidine ring (methyl group for S25, bromine for Br7′, hydrogen for dU7′ and iodine for Io7′). For the BrdU of Br7′and the thymidine 7′ of S25, the 5-substituents differ little in size, the van der Waals radii of bromine and methyl being 1.95 Å and 2.00 Å respectively [38].

Figure 31 shows the comparison of the binding efficiencies of Br7′, dU7′ and Io7′ with the unsubstituted DNA (S25). Br7′and S25 with similar van der Waals radii of the bromine and methyl groups at the 5 position of the pyrimidine ring gave very similar binding affinity, while dU7′ (Figure 31(c)) and Io7′ (Figure 31(d)) with Kd > 30 nM showed poorer binding than S25. The lower binding affinities of dU7′ and Io7′ were due to the large differences in the size of the functional groups at the 5 position of the pyrimidine ring. These results revealed that although the same specific DNA-mtase binding complex was obtained, the size of the group at the 5 position of the pyrimidine ring in position 7′ did affect the binding of DNA and mtase, and is therefore very important in the specific recognition of the target DNA by the mtase.

The gel retardation data (Figure 31) indicated that the same rate of migration on the non-denaturing PAGE gels of specific DNA-mtase binding complexes was obtained for S25 and Br7′, dU7′ and Io7′. Also from the denaturing SDS gels, the same migration of crosslinked complexes was obtained for all four complexes. This demonstrated that the binding complexes between the substituted DNA and mtase were specific.

Figure 30. Model of the *Eco*KI target sequence (25 mer) with a BrdU substituted at thymidine 7′, the residue complementary to the first adenine in the -AAC- recognition sequence (yellow balls and coloured balls between the yellow balls). The 5-bromine (green ball) of T7′ and the 6-nitrogen (N6, the blue ball as indicated in Figure 2) of A8 are in the major groove. Where photoreactive 5-bromine can be crosslinked to the S polypeptide by UV irradiation and N6 can be methylated by mtase. The N6 of the other methylation target (A17′) is about at the same plane as that of A8 following one helix turn.

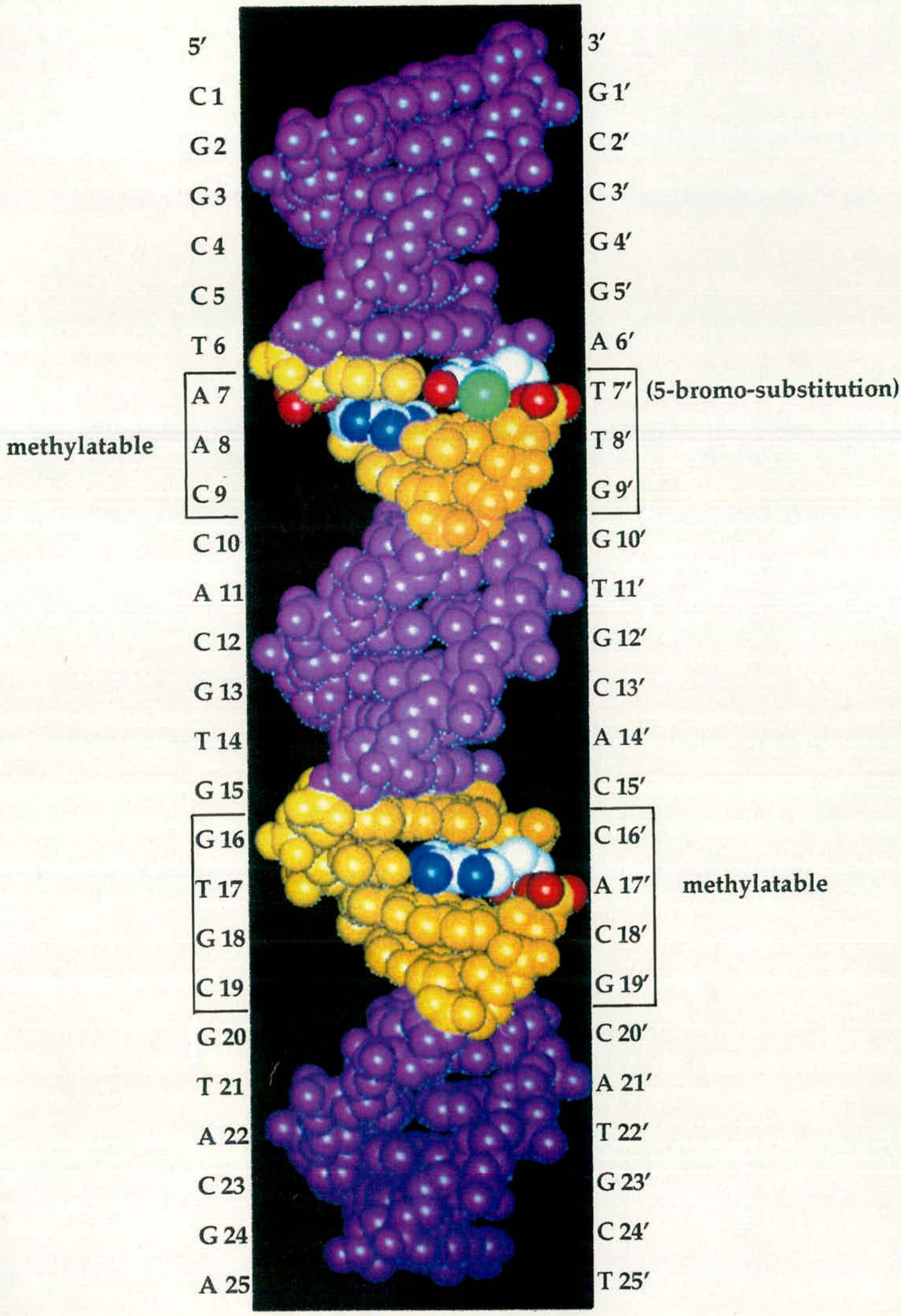
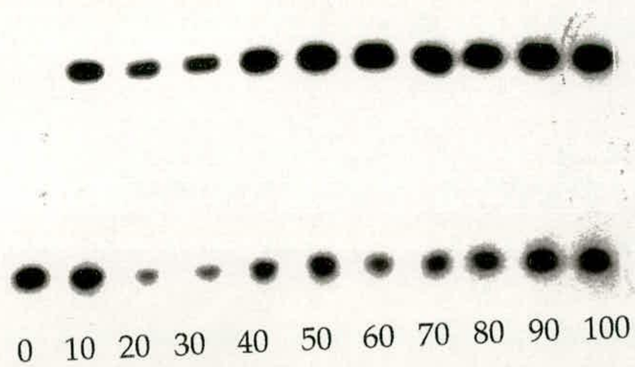

5′
3′
C 1
G 1′
G 2
C 2′
G 3
C 3′
C 4
G 4′
C 5
G 5′
T 6
A 6′
A 7
T 7′
(5-bromo-substitution)
methylatable
A 8
T 8′
C 9
G 9′
C 10
G 10′
A 11
T 11′
C 12
G 12′
G 13
C 13′
T 14
A 14′
G 15
C 15′
G 16
C 16′
T 17
A 17′
methylatable
G 18
C 18′
C 19
G 19′
G 20
C 20′
T 21
A 21′
A 22
T 22′
C 23
G 23′
G 24
C 24′
A 25
T 25′

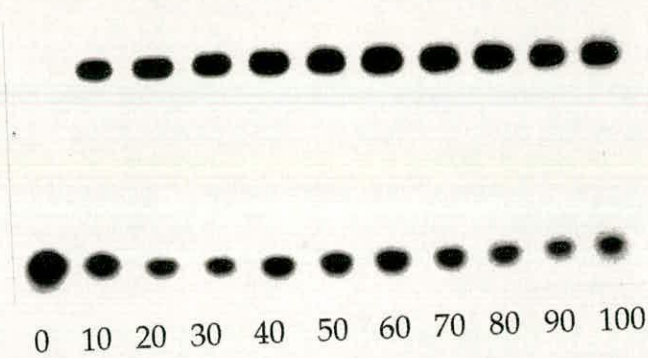
Figure 31. Comparative gel retardation experiments showing the binding of *Eco*KI mtase (0-100 nM) to DNA duplexes (0.1 nM) in the presence of AdoMet (100 μM). The top band is DNA-mtase complex and the lower band is free DNA. The mtase concentration (nM) is noted below each lane.

Panel a is unmodified DNA duplex (S25).
Panel b is 5-BrdU substituted DNA duplex (Br7').
Panel c is dU substituted DNA duplex (dU7').
Panel d is 5-IodU substituted DNA duplex (Io7').

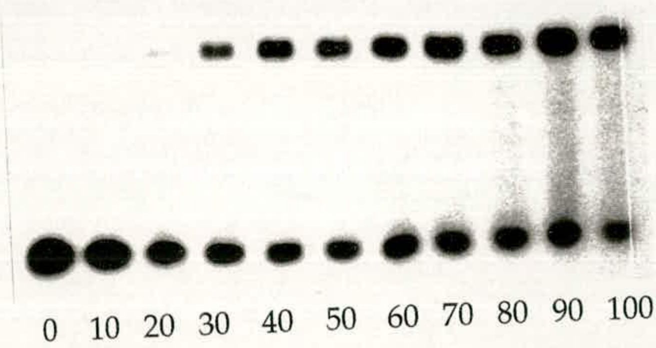
(a)


0 10 20 30 40 50 60 70 80 90 100

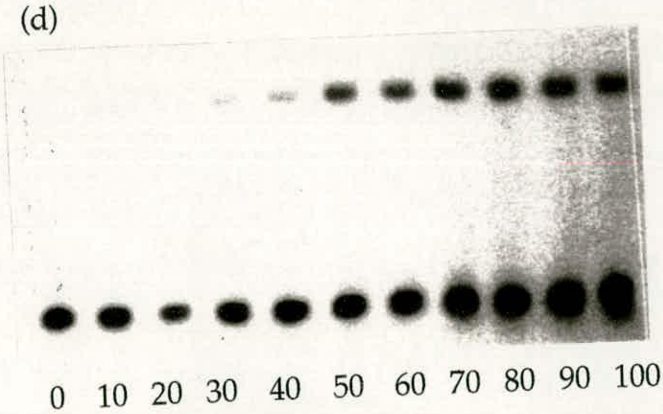
(b)


0 10 20 30 40 50 60 70 80 90 100

(c)


0 10 20 30 40 50 60 70 80 90 100

(d)


0 10 20 30 40 50 60 70 80 90 100

3.4 *UV induced photocrosslinking of DNA-mtase using BrdU substituted DNA target sequence*

3.4.1 Optimum wavelength

The efficiency of crosslinking of the BrdU single site substituted oligonucleotide (Br7') at different wavelengths of UV irradiation was compared with that of unsubstituted oligonucleotides (S25). Results in Table 3 show that less than 1% of crosslinked complex was obtained after the UV irradiation at 302 nm, 320 nm or 365 nm for S25. For BrdU substituted Br7' (Figure 31), less than 2% of crosslinked complexes were obtained at 302 nm while less than 1% of crosslinked complexes were obtained at 365 nm or 320 nm. However, a higher percentage (4% and 12%) of crosslinked complex was recovered from UV irradiation at 254 nm for S25 and Br7' respectively. Table 3 illustrates that the most efficient wavelength for UV irradiated DNA-mtase crosslinking for both Br7' and S25 is 254 nm.

Model studies with 5-bromouracil have reported that photoreactivity was very dependent upon the wavelength [107]. Monochromatic excitation at 308nm with a xenon chloride excimer laser yielded the reactive triplet state ($^3\pi,\pi*$) via a low energy excited singlet state (n, $\pi*$). However, with excitation at 254 nm with a low pressure mercury lamp two different reactions can occur. One has a higher energy singlet state ($^1\pi,\pi*$) and undergoes C-Br bond homolysis to yield the radical pair of a uracilyl radical and a bromine atom, which usually leads to BrU-DNA single strand breaks. The other, having single electron transfer from a neighbouring group (e.g. electron rich aromatic amino acid residues) in the triplet state yields the radical ion pair of $BrU^{-\cdot}$ and amino $acid^{+\cdot}$ [108], which is responsible for the crosslinking of BrU-DNA to associated proteins. If this proposed mechanism holds true, the crosslinking reaction in the present experiment should have occurred via single electron transfer from an oxidizable group in the amino acid residue, i.e. tyrosine, followed by ion radical coupling and elimination of HBr to yield a crosslink. Although Koch [109] has reported that the intersystem crossing efficiency [110] of the singlet state ($^1\pi,\pi*$), populated with irradiation at 254 nm, to the triplet state ($^3\pi,\pi*$) was only half of that from the low energy nonreactive state (n $\pi*$), excited at 308 nm, to the reactive triplet state ($^3\pi,\pi*$), the present results show that to obtain the maximum amount of mtase crosslinked complex by UV irradiation from either the unsubstituted or 5-bromo substituted DNA, 254 nm irradiation should be used rather than 302 nm irradiation. Therefore, a wavelength of 254 nm was chosen for subsequent experiments.

3.4.2 Time of UV irradiation

Crosslinking is represented by the extra double bands on the autoradiograph as indicated in Figure 32. Later it became clear that the upper band was the S polypeptide crosslinked to one strand of DNA which was hybridised to its complementary strand and the lower band was the S polypeptide crosslinked to one strand of DNA. It is evident from Figure 32 that there were no crosslinked products in lanes 1 to 3, where lane 1 was free DNA, lane 2 was mtase without DNA, and lane 3 was the binding complexes without UV irradiation. The DNA-mtase crosslinked complexes band appeared only after the reaction mixture was irradiated at 254 nm. The degree of crosslinking in lanes 5 to 10 appeared very similar, and a 10-minute period using a wavelength of 254 nm was therefore taken as the standard procedure for UV irradiation. The DNA-mtase crosslinked complex bands revealed by silver staining are shown in Figure 33. The extra bands between the M subunit and the S subunit (as indicated) are due to the crosslinked complexes.

A small quantity of some other crosslinked complexes (not shown), which migrated slower than the main bands of crosslinked complex, were also observed in some unheated samples when the concentration of mtase was very much higher than that of DNA. These extra crosslinked products bands tended to disappear when the concentration of mtase decreased (such that the ratio of protein to DNA decreased to 10 or less) or when mtase was denatured by boiling. This indicates that these extra crosslinked bands may have been due to the aggregation of protein on the initial crosslinked complex.

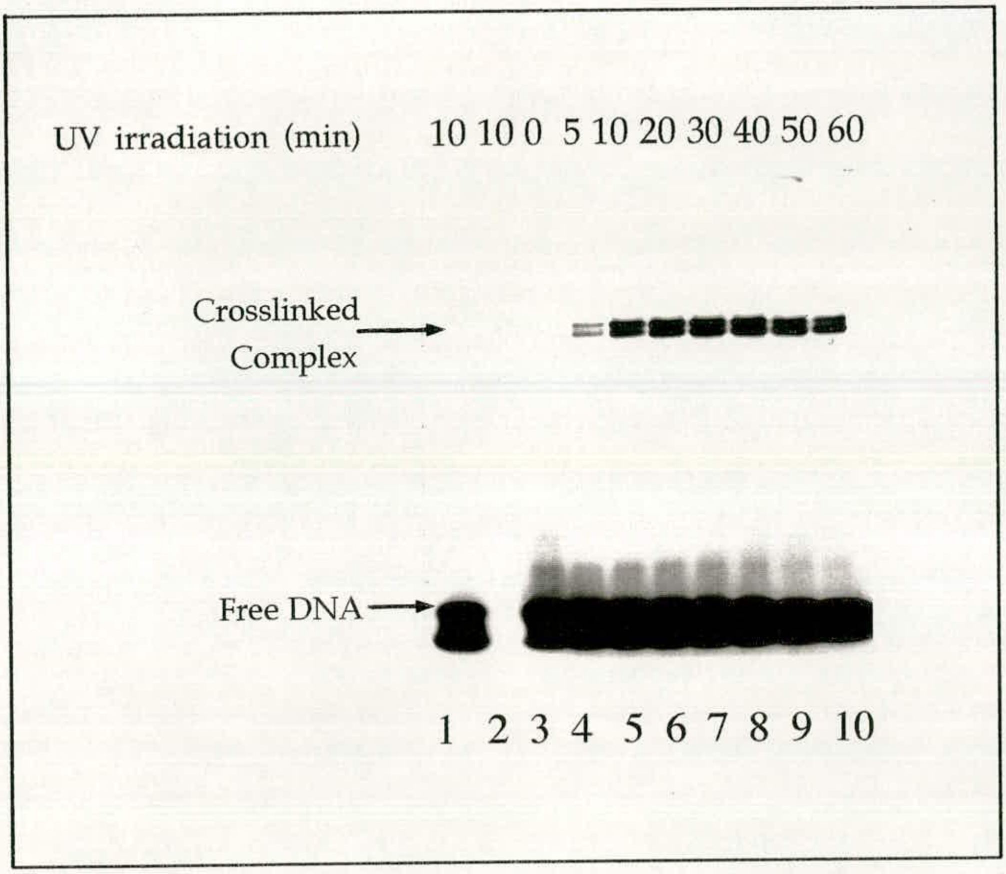


Figure 32. Time course of UV crosslinking for Br7′-mtase complex. Lane 1 contains only free DNA, lane 2 contains only mtase, lane 3 contains the binding mixture of Br7′-mtase which is not UV irradiated and lanes 4–10 contain UV irradiated Br7′-mtase crosslinked complex. The binding mixture contained Br 7′ (10 nM), mtase (200 nM) and AdoMet (100 μM). The binding mixture of lanes 3–10 were irradiated for 0 to 60 minutes respectively as indicated.

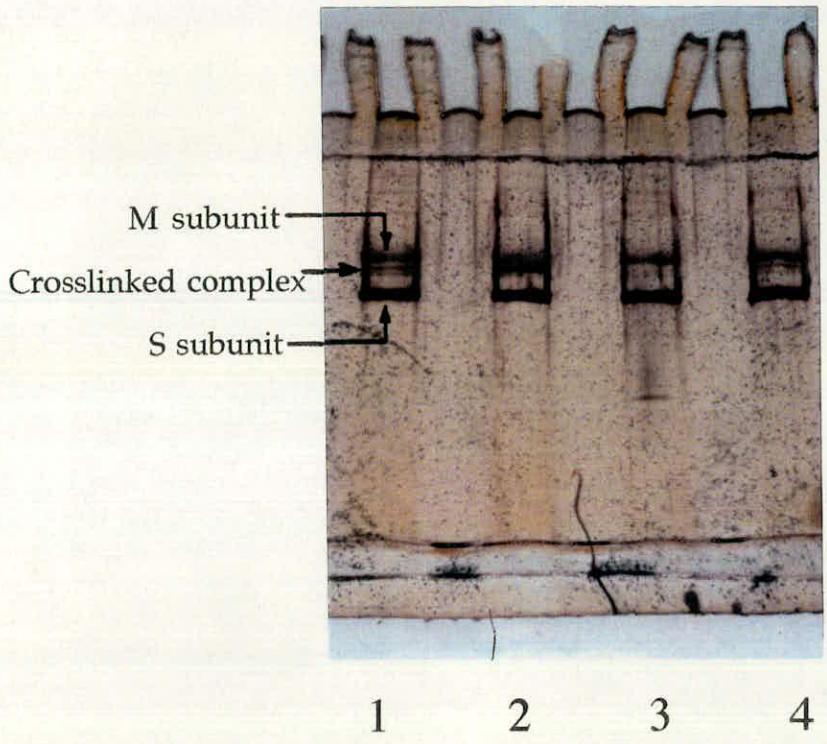


Figure 33. A silver stained 12% SDS-PAGE gel of mtase after crosslinking to duplex S25, Br7′, dU7′ and Io7′.

The S subunit crosslinked to the duplex is visible between the S subunit (lower band) and M subunit (upper band). Lanes 1-4 are S25, Br7′, dU7′ and Io7′. Crosslinking was induced by UV irradiation for 10 minutes with wavelength of 254nm.

3.4.3 Identification of the crosslinked complexes on 12% SDS-PAGE

To identify the DNA compositions of the two crosslinked bands on 12% SDS-PAGE (Figure 32), the crosslinked complexes were boiled and loaded on 7M urea-SDS-PAGE denaturing gel to obtain the single stranded DNA-mtase crosslinked complexes (as in section 2.7.3). Table 5 represents data from the autoradiography of unboiled and boiled samples of crosslinked complex on 12% SDS-PAGE and 7M urea-SDS-PAGE gels, in which the DNA duplex Br7′ was labelled on the bottom strand.

The DNA of unboiled crosslinked complex remained in a duplex form while mtase dissociated to give S and M polypeptides upon electrophoresis in a 12% SDS-PAGE gel. The unboiled samples of reaction mixture, having a heavier mass of the DNA duplex crosslinked complex which migrated slower than the single stranded DNA crosslinked complex on the SDS-PAGE gel, formed the upper band of crosslinked complex on the gel. Boiled samples of reaction mixture gave both upper and lower bands of crosslinked complex, which represented the DNA duplex-mtase crosslinked complex and the single stranded DNA-mtase crosslinked complex respectively on the SDS-PAGE gel. The formation of the lower crosslinked band was due to the procedure of boiling resulting in dissociation of the DNA duplex-mtase crosslinked complex to single stranded DNA-mtase crosslinked complex and the upper crosslinked band was formed due to the rehybridization of dissociated single stranded DNA-mtase with the complementary strand during sample loading.

Table 5. The heat treatment of crosslinked complexes on different gels.

	12% SDS-PAGE gel		7M Urea-SDS-PAGE gel	
Heat treatment	Unboiled	Boiled	Unboiled	Boiled
Crosslinked complexes				
Upper band	+	+	+	–
Lower band	–	+	+	+

Table 6. Crosslinked DNA-mtase with $\gamma^{32}P$ labelled on different strands of DNA.

	12% SDS-PAGE gel			7M Urea-SDS-PAGE gel		
Strand Labelled	Top	Bottom	Both	Top	Bottom	Both
Crosslinked complex						
Upper band	+	+	+	–	–	–
Lower band	–	+	+	–	+	+

To keep the DNA duplex-mtase completely dissociated after boiling, 7M urea-SDS-PAGE gels were used. The high concentration of urea (7M) was able to dissociate DNA duplex to single stranded DNA by destablising the hydrogen bonds between base-pairs. Table 5 shows that 7M urea had very strong ability to dissociate DNA. Both upper band and lower band were observed on the unboiled samples on the 7M urea-SDS-PAGE gel. The formation of the lower band on the unboiled sample was due to the strong dissociating ability of 7M urea during electrophoresis. For complete dissociation of the DNA duplex, the boiled samples of crosslinked complexes were quenched into liquid nitrogen before electrophoresis on the urea-SDS-PAGE gel. Table 5 shows that only the lower band was observed for the boiled sample on the urea-SDS-PAGE gel. This demonstrates that rehybridization of the single stranded DNA was prevented in the 7M urea-SDS-PAGE gel during electrophoresis once the dissociation of DNA duplex was completed before the samples were loaded on gels. The results also demonstrate that UV irradiation of DNA-mtase only produced a single stranded DNA-mtase crosslinked complex. Since the DNA duplex was bottom-strand labelled, the indication is clearly that the crosslinked complex was the bottom-strand of the DNA crosslinked to the S polypeptide.

For further confirmation that no top-strand DNA-mtase crosslinked complex is formed and that the crosslinking only occurs between the bottom-strand of the DNA and mtase, DNAs (S25, Br17, Br7' and Br500) were each labelled on the top, bottom or both strands. Samples were boiled before being loaded on gels.

Table 6 summarises the results of autoradiography of boiled crosslinked complexes with $\gamma^{32}P$ labelling on top, bottom or both strands of DNA run on 12% SDS-PAGE or 7M urea-SDS-PAGE gels.

The lower crosslinked band which was detected when the DNA was labelled only on the bottom-strand (or both-strands) became undetectable on gels (12% SDS-PAGE and 7M urea-SDS-PAGE) when the samples were labelled on the top-strand. This demonstrates that the lower crosslinked band was not due to the top-strand DNA-mtase crosslinked complex which may migrate at the same speed as the bottom-strand DNA-mtase crosslinked complex in both gels. The presence of a detectable upper crosslinked band for the top-strand labelled DNA duplexes on the 12% SDS-PAGE gel was due to the unlabelled bottom-strand DNA of the crosslinked complex rehybridizing with the labelled complementary top-strand DNA during loading and electrophoresis. However, rehybridization did not occur on the 7M urea-SDS-PAGE gel, therefore no upper crosslinked band appeared on gel for any of the strands of labelled DNA. The fact that only the lower crosslinked band appeared on 7M urea-SDS-PAGE gel for the bottom-strand (or both strands) DNA indicated that the lower crosslinked band was due to the bottom-strand DNA-mtase crosslinked complex. Therefore, the photocrosslinking of DNA-mtase only occurred between the bottom-strand DNA and mtase. The following DNA site substitutions experiments further demonstrated that only one nucleoside on the bottom-strand of DNA was crosslinked to mtase.

3.4.4 Identification of the nucleotide which is crosslinked by *Eco*KI mtase

The efficiency of crosslinking, measured by scintillation counting as the percentage of DNA present in the crosslinked complex relative to the total DNA (crosslinked + free), is illustrated in Table 4. Similar crosslinking efficiencies (~4%) were observed among the top strand substituted DNAs of Br6, Br14, Br17, Br21 and the bottom strand substituted DNAs of Br8′, Br11′, Br22′, Br25′ as well as unsubstituted DNA (S25). Br7′ was the only single site 5-bromo substituted oligonucleotide which increased the crosslinking efficiency by three fold (~12%). The crosslinking efficiency of Br500, 5-bromo-substituted at all possible thymidine sites of the DNA duplex, remained the same as Br7′ (~12%, Table 4). These results indicated that BrdU-substitution at thymidine 7′ was the only substitution which increased the efficiency of crosslinking. The autoradiograph in Figure 34 also demonstrates that Br7′ (lane 1) was the most efficient oligonucleotide for DNA-mtase crosslinking. Other DNA duplexes, Br8′, Br11′, Br22′, Br25′ (lanes 2 to 5 respectively) had similar crosslinking efficiencies to unsubstituted S25 (lane 6).

From these results it was concluded that the base at position T7′ (the thymidine of the A-T base pair adjacent to the adenine methylation targets in the trinucleotide part of the recognition sequence -AAC-) was involved in the UV induced DNA-mtase crosslinking. Therefore, when DNA-mtase crosslinking occurs between mtase and the 5-methyl group of T7′ on the unsubstituted DNA (S25) and all BrdU substituted DNA, except Br7′ and Br500, 4% crosslinking efficiency was observed and when the photoreactive bromine atom was substituted for the 5-methyl group of thymidine at T7′ on Br7′ and Br500, the crosslinking efficiency was increased up to 12%.

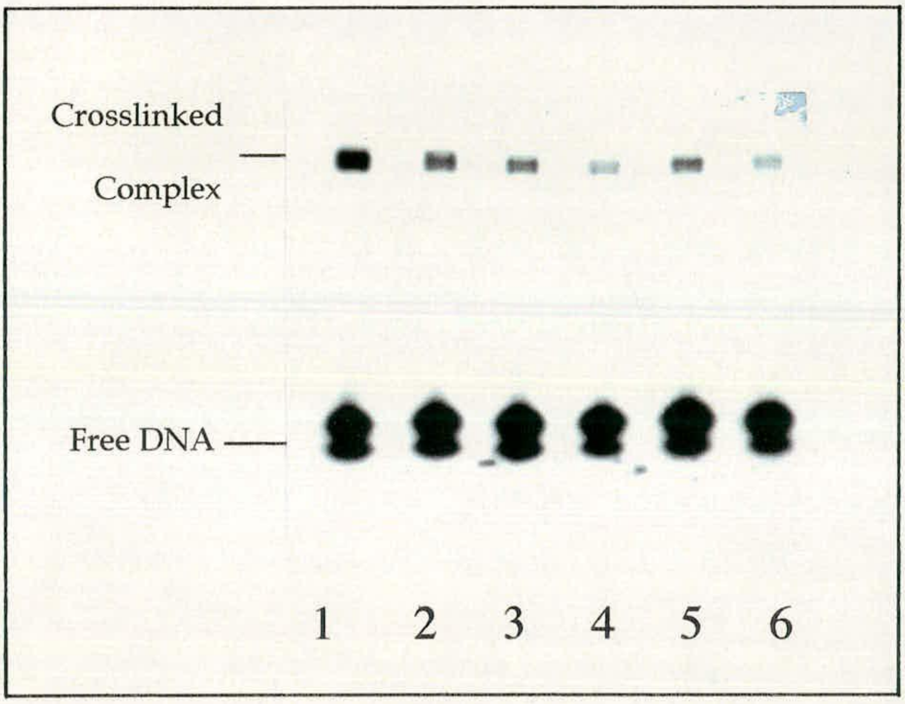


Figure 34. The autoradiograph shows the crosslinking efficiency of the different substituted DNAs. Lanes 1-5 are bottom strand substituted DNAs Br7′, Br8′, Br11′, Br22′, Br25′ and lane 6 is the controlled unsubstituted DNA, S25. Mtase (200nM) with premixed AdoMet (100µM) and DNA (10nM) were added in Tris-HCl buffer (20mM pH8.0, NaCl 100mM) containing 5% glycerol and mixed with microcentrifuge. The total volume of reaction mixture was 20µl. After 15 minutes of incubation at 22°C the reaction mixture was UV irradiated at 254nm for 10 minutes. The UV irradiated reaction mixture was mixed with same volume of sample buffer (20 µl) and run on a denaturing SDS-polyacrylamide gel (12%) for 3 h at 45 mA to analysis the DNA-mtase crosslinked complexes.

3.4.5 Different functional group substitutions on T7′

IodU and dU were substituted for T7′ for further investigation of the DNA-mtase crosslinking site and for investigating the possibility of improving crosslinking efficiency. A large propotion of crosslinked complex was necessary, so that enough crosslinked complex could be purified for sequencing and the crosslinked amino acid could be identified. Quantitative measurements showed that the degree of crosslinking, as determined by scintillation counting, was 1.0%, 2.6%, 0.6% and 0.8% for S25, Br7′, dU7′ and Io7′ respectively. It varied to a large extent when the crosslinking site was substituted by different nucleotides. Br7′ tripled the crosslinking efficiency of the unsubstituted S25, but dU7′ or Io7′ reduced the efficiency of crosslinking. This is clearly shown in the silver-stained SDS-PAGE gel of Figure 33.

Photoproducts from IodU generated in solution upon UV irradiation are clearly much greater than from BrdU, but the yield of photoproducts from IodU subtituted DNA is not always as high as that of BrdU substituted DNA [38]. 5-Iododeoxyuracil substitution has been reported to produce very efficient (70%) photocrosslinking at 308 or 312 nm for the *Oxytricha* telomere protein and its cognate DNA [111] compared to 5-bromodeoxyuracil substitution (12%) [112]. However, in the present study the 5-iododeoxyuracil substitution did not produce an efficient photocrosslinking between DNA and mtase. The higher yield of DNA-mtase crosslinked complex using Br7′ DNA compared to S-25 is because bromine is more photoreactive than the methyl group. The possible mechanism of UV irradiation for bromo-substitution at 254nm has been described above in section 3.4.1. The poor yield of crosslinked DNA-mtase complex for Io7′ (0.8%) compared to Br7′ (2.6%) and S-25 (1.0%) might be due to the effect of the iodine atom on DNA-mtase binding. Figure 30 demonstrates that the binding affinity of Io7′ is the worst among the four DNAs. Despite the poor binding affinity of Io7′, the amount of crosslinked complex (0.8%) was better than that of dU (0.6%). This is because the hydrogen atom of the dU7′ is not a photoreactive atom and cannot readily generate a photoreactive state upon UV irradiation, even though the small hydrogen atom in dU7′ has a lesser effect on binding affinity than the much bulkier iodine atom.

These results indicate that (I) the DNA-mtase crosslinking efficiency is affected by the binding affinity and (II) the efficiency of DNA-mtase crosslinking is affected not only by the size of the close-contact functional group but also by its photoreactivity.

3.5 *Large scale DNA-mtase crosslinking*

3.5.1 Purification of DNA-mtase crosslinked complex from free DNA

Using the Br7′ duplex a large quantity of DNA-mtase crosslinked complex was prepared. The yields of crosslinked complexes through different purification steps were monitored by scintillation counting of the radiolabelled DNA. The procedures for eluting the complex from the gel and desalting before trypsin digestion gave high recoveries of radiolabelled samples, 90-97% and 89-93% for each step respectively.

3.5.2 Trypsin digestion and purification of DNA-mtase crosslinked complexes

Digestion of the purified crosslinked complex with trypsin and purification of the Br7′-peptide resulted in 46-69% recovery. After trypsin digestion the Br7′-peptide was purified by ion-exchange chromatography using an Econo-Pac Q cartridge. One main peak was observed on the elution profile monitored by UV absorbance at 254nm during the purification (Figure 35a and Figure 36). The maximum absorbance of the Br7′-peptide occurred after elution for 90 minutes when the salt gradient reached to 0.5 M KCl. Its occurrence coincided with the peak from scintillation counting for the radiolabelled DNA in the fractions (Figure 35b). Silver staining analysis of the eluted fractions on 12% or 20% SDS-PAGE gels indicated that trypsin and some digested protein passed through the ion-exchange Econo-Pac Q cartridge without binding, some digested peptides eluted during the gradient before the elution of the Br7′-peptide crosslinked complex. The free DNA eluted after the elution of the Br7′-peptide crosslinked complex.

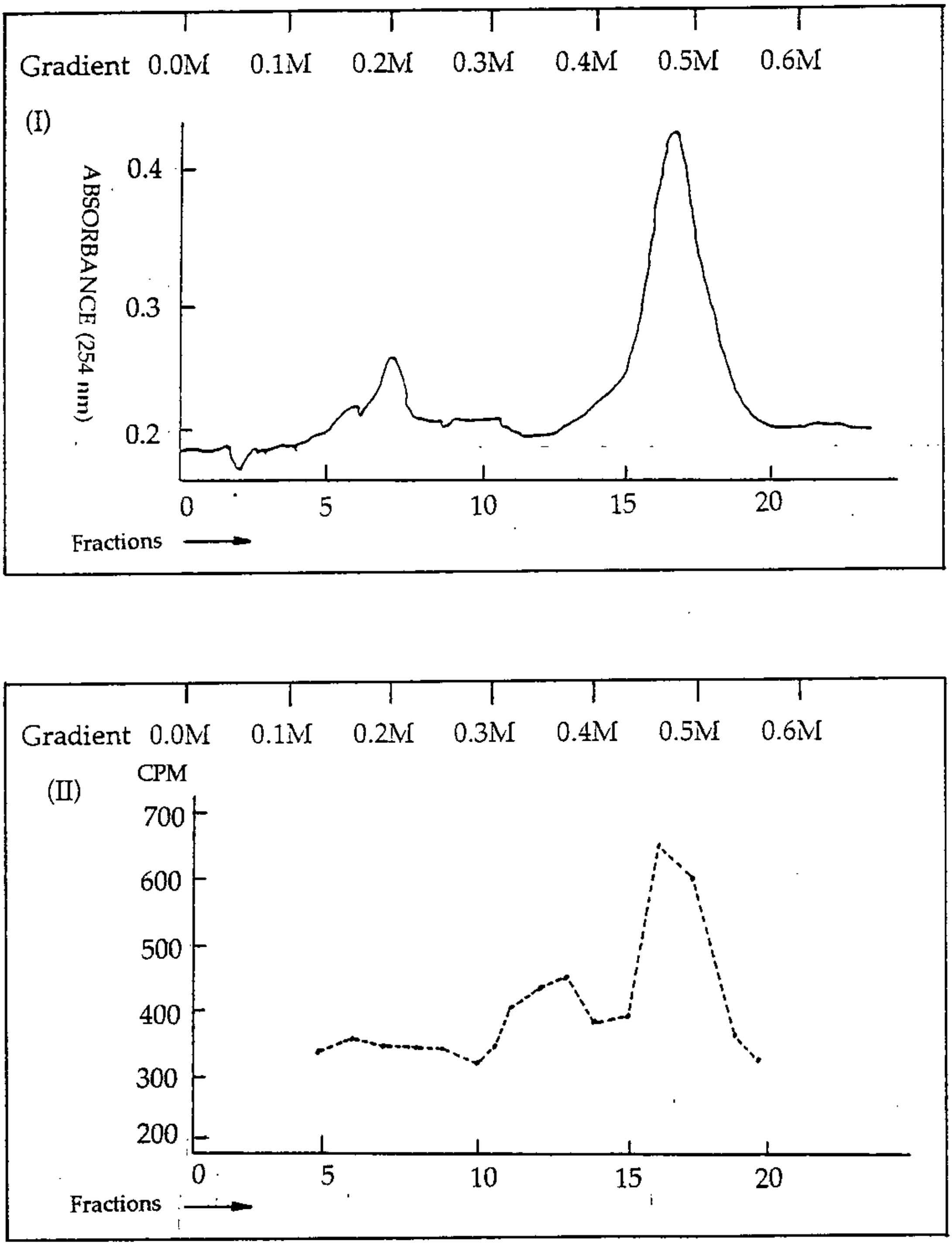


Figure 35. The UV elution profile (I) and scintillation counting (II) of the radio-labelled trypsin digested crosslinked DNA-mtase complex obtained from an anion exchange Econo-Pac Q cartridge with a 100-ml gradient of 0-0.6M KCl in phosphate buffer (20mM, 5% ethanol, pH6.8).

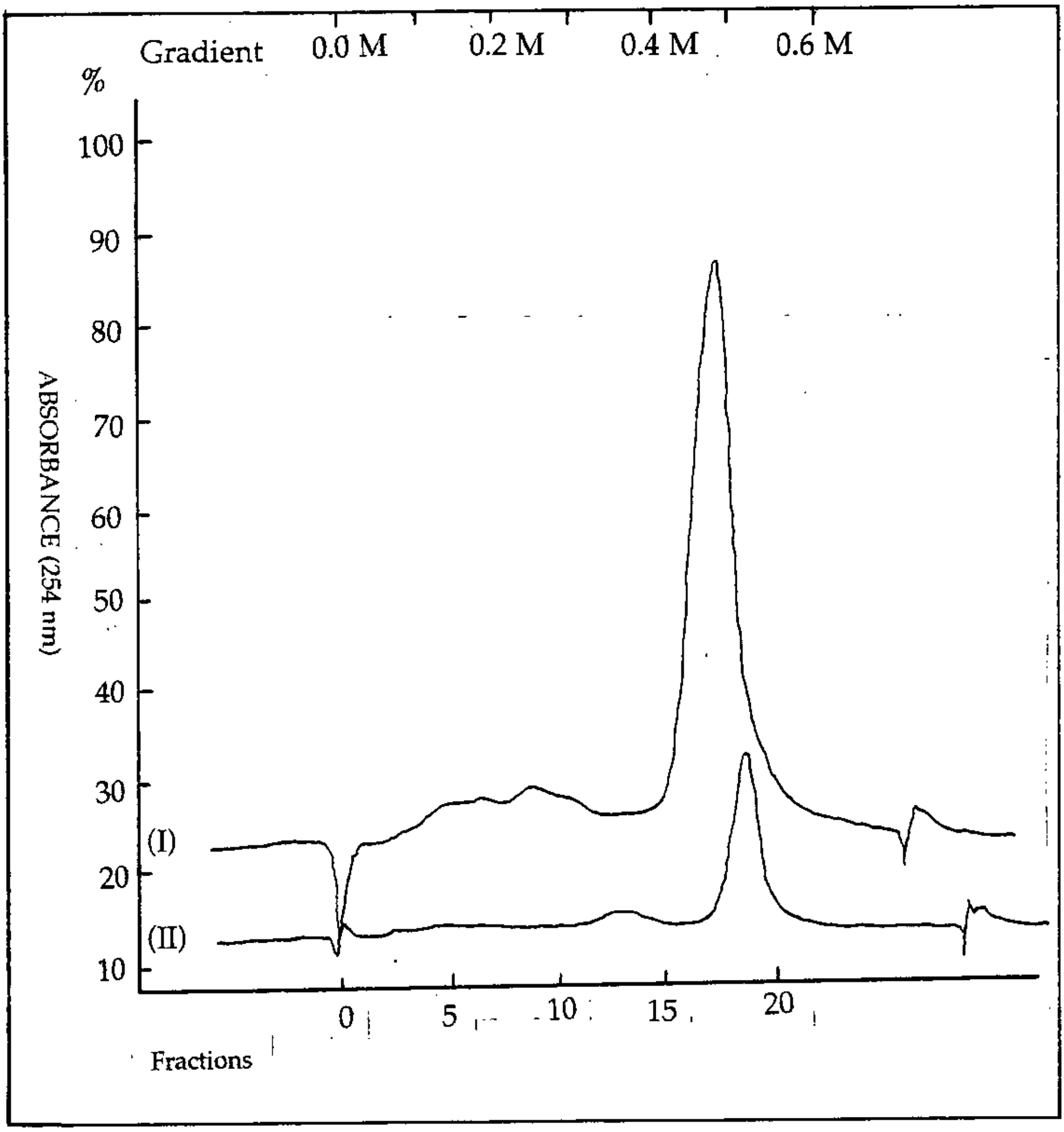


Figure 36. The UV elution profiles of (I) digested long DNA-peptide and (II) twice digested short DNA-peptide after eluting from the Econo-Pac Q cartridge with a 100-ml gradient of 0-0.6 M KCl in phosphate buffer (20 mM, 5% ethanol, pH 6.8).

Autoradiography on 20% SDS-PAGE gel showed 2 products (Figure 37). The first, intermediate DNA-peptide, produced by digestion in the presence of ammonium acetate and the second, long DNA-peptide, produced by digestion in Tris-HCl. Combining the lane 3 sample and lane 4 sample gave two separate bands (see Figure 37). This indicated that the sample in lane 4 was not a radiolabelled DNA duplex but an intermediate DNA-peptide crosslinked complex. A similar result was obtained when the DNA-peptide complex of lane 5 was combined with Br7' and run in lane 7. This confirmed that the sample in lane 5 of Figure 37 was a digested long DNA-peptide crosslinked complex, easily distinguishable from the Br7' duplex alone. The overall recovery of the long DNA-peptide crosslinked complex was approximately 2.5-2.9 % of the total DNA used.

Lane 3 in Figure 38 also shows the long DNA-peptide product. Redigestion of the long DNA-peptide, in the presence of ammonium acetate, by trypsin gave a very short DNA-peptide crosslinked product (lane 2, Figure 38). Samples in lane 4, Figure 38 were the crosslinking mixture which contained the undigested DNA-S-subunit crosslinked complex (top band), Br7' duplex (middle band) and unincorporated ATP (lower band).

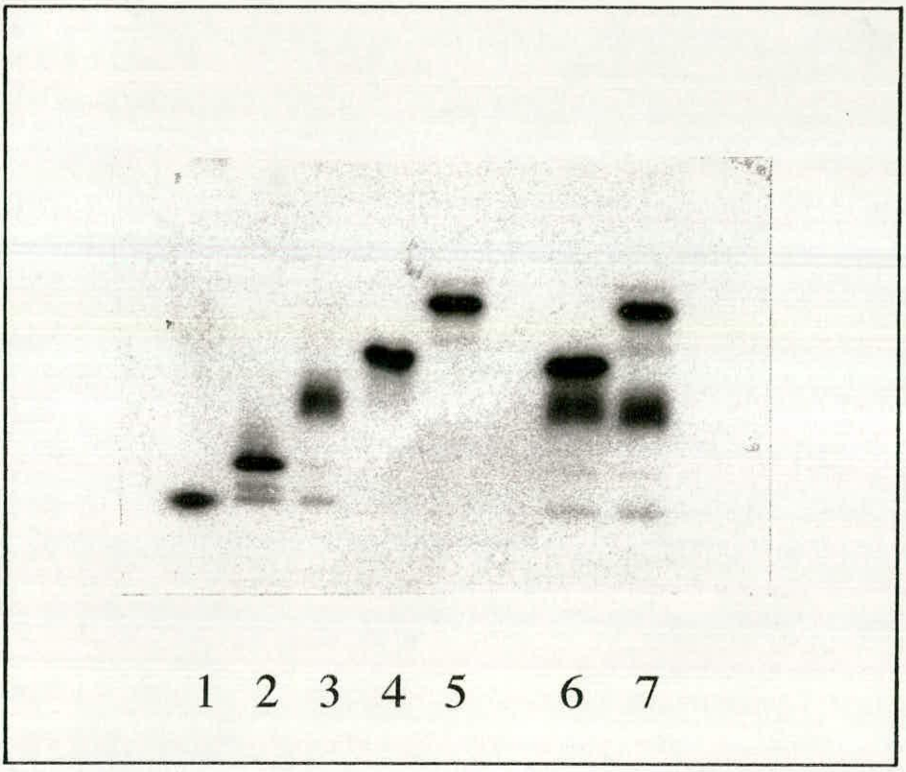


Figure 37. Digested DNA-peptide crosslinked complexes.
The autoradiogram of a 20% of polyacrylamide gel of the Br7′-peptide crosslinked complexes after trypsin digestion and purification. Lanes 1 to 3 are γ^{32}P ATP, bottom strand of Br7′ and Br7′ duplex respectively. Lane 4 shows the intermediate Br7′-peptide duplex complex obtained after trypsin treatment in the presence of ammonium acetate and lane 5 shows the long Br7′-peptide duplex complex obtained after trypsin treatment in the absence of ammonium acetate. Lane 6 shows mixed samples of lane 3 and lane 4. Lane 7 shows mixed samples of lane 3 and lane 5.

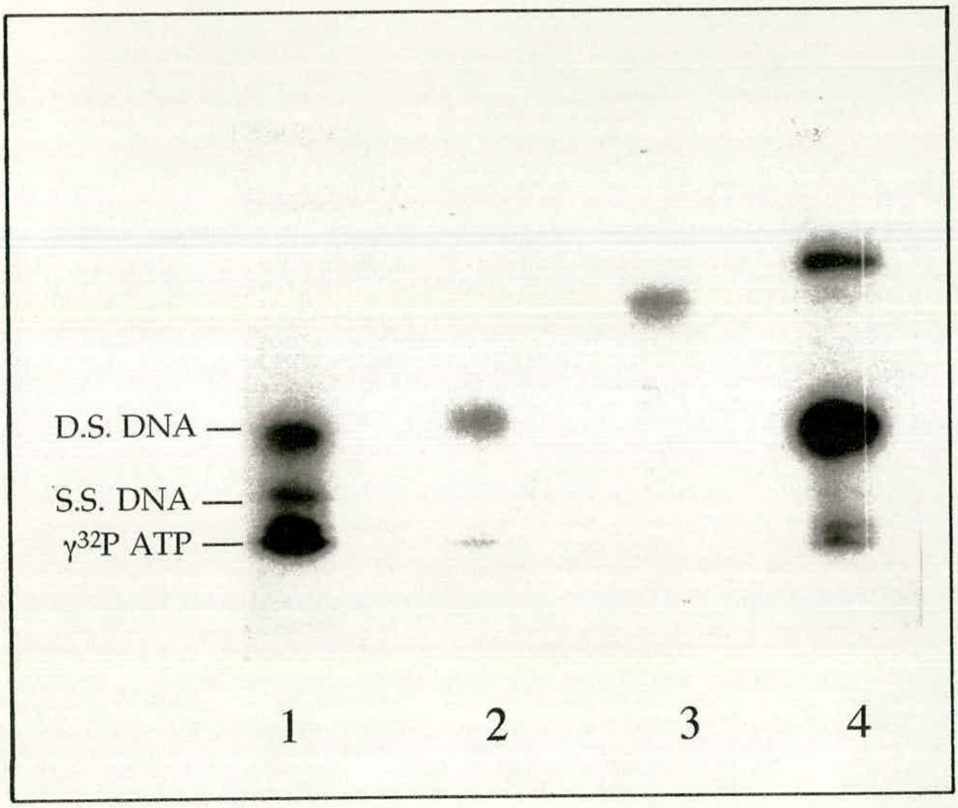


Figure 38. Digested DNA-peptide crosslinked complexes. An autoradiogram performed as in Figure 37. The upper bands in lanes 1-4 are a $\gamma^{32}P$ labelled Br7′ duplex, a double digested short DNA-peptide crosslinked complex, a long Br7′-peptide crosslinked complex and a Br7′-S subunit crosslinked complex respectively. Additional bands are visible in lanes 1 and 4. In lane 1, the two bands running ahead of ^{32}P-labelled Br7′ duplex are ^{32}P-labelled bottom strand of Br7′ and $\gamma^{32}P$ ATP as indicated. In lane 4 the two bands running ahead of the Br7′-S subunit crosslinked complex are ^{32}P-labelled Br7′ duplex DNA and $\gamma^{32}P$ ATP respectively.

3.6 *Peptide sequencing-identification of the amino acid crosslinked to DNA*

The amino acid residue that has been crosslinked can be identified by amino acid sequencing. If a solid-phase sequencer is used, the site of crosslinking can be identified by scintillation counting of an aliquot of each of the resulting PTH-amino acid derivatives. If a gas or liquid phase instrument is used, the phenylthiazolinone derivative of the $\gamma^{32}P$ labelled DNA-amino acid will not be extracted from the support disk, so a direct identification cannot be made on the basis of scintillation counting. However, there will be a "hole" at that position in the sequence that, if the primary structure of the protein is known, can be used to assign the amino acid residue involved in crosslinking [126].

These sequence analyses are usually complicated by the contamination peaks that coelute with amino acids in the first two sequencing cycles [81,113]. To eliminate this problem, a different length of peptide with variable amino terminus would allow the contamination peaks in the first two sequencing cycles to coelute with different amino acids. Therefore, the ambiguous amino acid sequence can be identified and the crosslinked DNA-amino acid sequence.

3.6.1 Sequencing of the long DNA-peptide

There were very large contamination peaks on the HPLC elution profiles of the first three cycles when sequencing the long DNA-peptide. The contamination peaks eluted where the peaks for Ser, Gly and Ala usually appeared. Therefore, although the yields of PTH-Ser of cycle 1 and the PTH-Ala of cycle 2 (Table 7) were relatively high, they were still considered as "unidentified" amino acids as they were undistinguishable from the contamination peaks. The HPLC elution profile of the PTH-Gly of cycle 3 (standard running time = 9.13 min) was hidden by the contamination peak at 9.36 of standard running time. The yields of PTH-derivatives of cycle 9, cycle 15, cycle 22 and cycle 26 were also undetectable. Results of other sequencing cycles clearly demonstrated that this long DNA-peptide was part of the ARD of the S polypeptide starting at Ser2 (The S subunit of *Eco*KI lacks the N terminal methionine and the sequence starts at Ser2). 30 cycles of sequencing were obtained from the sample of the long DNA-peptide crosslinked complex. The unidentified sequences were later determined by comparison with the Protein Sequence Database of the *Eco*KI S polypeptide.

Table 7. Quantification of the gas-phase sequencing of the long DNA-peptide from digested crosslinked complex.

Yields of the PTH-derivatives are shown in pmol, and only those amino acids present in the sequence are listed.

Data from Prof J. Fothergill and Mr B. Dunbar, Peptides Sequencing Service, the University of Aberdeen.

*Cycle	Asn	Ser	Gln	Thr	Gly	Glu	Ala	Arg	Tyr	Pro	Val	Ile	Lys	Leu	Sequence
1	----	275.6	12.83	77.83	1461	----	137.1	6.48	38.16	17.08	52.62	37.35	24.86	44.41	(Ser)@*1
2	12.51	34.45	8.30	9.21	----	----	40.01	7.42	6.06	3.90	9.52	11.60	5.24	8.25	(Ala)@*2
3	11.41	13.54	4.54	10.04	----	----	19.50	4.68	1.94	2.23	8.74	4.61	2.16	4.04	(Gly)@*3
4	7.13	4.75	5.41	6.93	----	----	8.26	3.21	1.10	6.79	6.29	8.73	11.31	3.89	Lys
5	4.73	2.02	8.37	5.75	----	9.58	4.97	2.74	2.44	3.63	7.77	3.67	6.29	25.23	Leu
6	6.00	3.29	4.21	4.90	----	----	4.51	2.02	1.42	17.94	5.69	3.48	4.13	15.60	Pro
7	3.08	2.27	3.34	6.27	----	18.62	3.26	1.46	0.84	12.34	3.87	2.56	3.40	8.50	Glu
8	2.69	1.16	2.91	4.37	10.01	13.97	2.83	1.36	1.52	6.93	4.63	1.76	3.85	5.63	Gly
9	2.00	0.67	3.04	5.25	10.07	3.09	1.11	1.04	4.07	4.03	----	1.52	2.83	6.92	(Trp)@*4
10	1.35	----	2.53	4.59	5.72	7.33	2.84	1.06	0.68	3.58	9.53	2.05	2.32	4.63	Val
11	1.51	0.51	1.29	4.43	4.01	6.21	2.96	----	0.49	2.39	7.27	8.39	1.66	5.17	Ile
12	1.78	----	2.29	4.04	3.76	5.06	8.39	----	0.60	1.98	5.17	8.53	1.37	4.45	Ala
13	1.91	----	2.18	4.00	3.45	4.59	7.29	0.97	1.13	7.38	3.75	6.64	1.11	3.77	Pro
14	1.69	----	2.66	3.06	3.15	3.91	5.02	1.11	0.97	7.44	6.74	4.56	1.15	3.33	Val
15	0.64	----	0.79	2.66	2.90	1.42	1.74	----	0.39	3.94	5.47	1.87	----	----	(Ser)@*5
16	1.55	4.04	1.057	5.08	2.53	2.58	2.85	----	0.74	3.55	5.94	2.62	0.85	2.38	Thr
17	1.45	4.32	1.74	6.09	2.51	2.36	2.41	----	0.74	2.42	7.33	2.05	0.76	2.22	Val
18	2.21	1.27	1.42	7.11	1.04	0.95	2.07	----	0.68	1.83	6.39	1.75	0.64	1.48	Thr
19	0.55	0.73	1.33	8.39	1.84	1.85	1.88	----	0.65	1.34	4.85	1.68	0.92	1.98	Thr

20	0.60	0.56	1.16	7.51	1.41	1.75	1.74	----	0.45	1.12	3.57	1.38	0.74	4.08	Leu
21	0.70	----	1.34	5.47	1.45	1.89	1.50	----	0.59	1.15	2.58	3.31	0.85	4.46	Ile
22	0.61	----	1.37	4.21	0.91	0.95	1.62	0.54	0.59	0.94	1.97	3.96	0.75	4.20	(Arg)@*[6]
23	0.59	----	1.26	3.16	2.65	3.00	1.51	1.67	0.74	0.76	1.80	3.41	0.60	3.24	Gly
24	0.80	----	1.15	2.34	2.40	0.77	1.49	2.15	----	0.82	2.38	2.63	----	2.07	Val
25	0.93	----	1:08	2.70	2.05	0.49	1.40	----	0.42	0.77	2.89	2.07	----	1.88	Thr
26	0.89	----	0.97	2.93	1.57	0.57	1.27	0.84	0.71	0.66	2.82	1.74	----	1.65	(Tyr)@*[7]
27	0.94	----	0.87	2.94	1.93	1.95	0.99	----	0.50	0.59	2.50	----	0.97	1.50	Lys
28	0.85	0.59	0.82	2.69	0.92	0.45	1.10	----	0.42	0.64	2.04	1.29	0.80	1.28	Lys
29	0.89	0.64	----	2.49	0.93	0.71	1.07	----	0.47	0.54	1.59	0.76	1.00	1.52	Glu
30	0.74	0.52	1.06	1.78	0.70	0.82	0.82	----	0.40	0.49	1.27	0.89	----	1.00	Gln

*Cycle 1 is equivalent to amino acid 2 of the S polypeptide as it lacks Met 1.

@ The amino acids of cycles 1, 2, 3, 9, 15, 22, and 26 were determined only after a comparison with the Protein Sequence Database of the *Eco*KI S polypeptide when the peptide sequence had been unambiguously identified as from the trinucleotide recognition domain of the S polypeptide.

*[1] The Ser peak of cycle 1 (HPLC standard running time = 7.81 min) was obscured by the huge contaminating peaks, which appeared at the HPLC running time of 6.18, 7.07 and 9.11 min respectively.

*[2] The Ala peak of cycle 2 (HPLC standard running time = 12.62 min) was also obscured by the huge contaminating peaks, which appeared at the HPLC running time of 6.19, 7.12 and 9.32 min respectively.

*[3] The Gly peak of cycle 3 (HPLC standard running time = 9.13 min) was masked by the contamination peak, which appeared at the HPLC running time = 9.36 min.

*[4] No Trp peak for cycle 9 (HPLC standard running = 24.15 min) was seen on the HPLC trace.

*[5] No Ser peak for cycle 15 (HPLC standard running time = 7.84 min) was seen on the HPLC trace.

*[6] The low PTH-Arg yield of cycle 22 (HPLC standard running time = 15.25 min) may be due to the attack of trypsine on this amino acid making it unidentifiable by HPLC.

*[7] The low PTH-Tyr yield of cycle 26 (HPLC standard running time = 15.61 min) is due to the crosslinking of the oligonucleotide on to this amino acid making it unidentified by HPLC.

3.6.2 Sequencing of the intermediate DNA-peptide

There were also huge contaminating peaks on the HPLC elution profiles of the intermediate DNA-peptide sequencing for the first few cycles. The contaminating peaks coeluted with Gly and Thr peaks, resulting in unidentified amino acids for cycle 1 (Figure 39) and cycle 3. Cycle 2 of the intermediate DNA-peptide sequencing was identified by the sharply increased yield of the PTH-Val (Table 8), although the contaminating peaks still remained in the Gly and Thr area of the HPLC elution profile. Cycle 4 was unidentified because of the very low yield of the recognizable PTH-derivative. This was probably due to the reaction product of this cycle being the crosslinked PTH-Tyr-DNA complex instead of PTH-Tyr (PTH-Tyr-DNA was unidentifiable by the set-up conditions for peptide sequencing). Cycle 5 to cycle 12 of the intermediate DNA-peptide sequencing (Table 8) are clearly part of the ARD of S polypeptide. Later the unidentified sequences were determined by comparison with the Protein Sequence Database of the *Eco*KI S polypeptide. The sequence of the intermediate DNA-peptide started at Gly-24 and stopped at Tyr-35.

3.6.3 Sequencing of the short DNA-peptide

The PTH-derivatives of the degradation of the short DNA-peptide (Table 9) demonstrated that it contained 6 amino acids and was a part of the ARD of the S polypeptide starting at Gly-24 and ending at Lys-29. The fact that the PTH-derivative of cycle 4 was not identifiable, once again indicated that Tyr-27 of the S subunit was crosslinked to the DNA. This complex was trapped on the filter of polyvinyl difluoride membrane and contaminants removed. Therefore, the first few cycles of sequencing can be clearly identified.

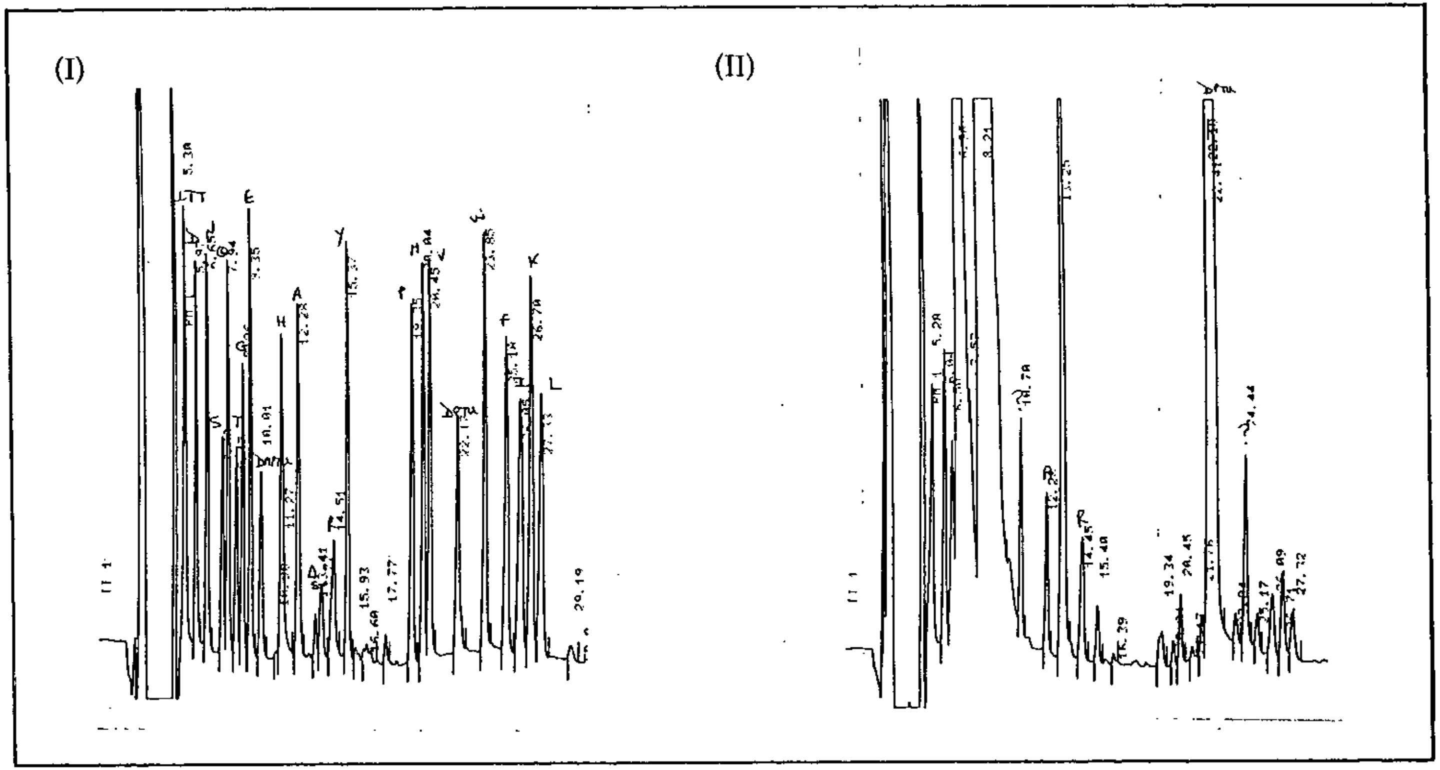


Figure 39. HPLC chromatogram of (I) standard PTH-amino acids and (II) the first cycle of intermediate DNA-peptide sequencing.
Standard amino acids running time (min) are as follows : Asp (D) = 5.97, Asn (N) = 6.65, Ser (S) = 7.6, Gln (Q) = 7.94, Thr (T) = 8.47, Gly (G) = 8.86, Glu (E) = 9.35, His (H) = 11.27, Ala (A) = 12.28, d-Ser (S) = 13.77, Arg (R) = 14.51, Tyr (Y) = 15.37, Pro (P) = 19.35, Met (M) = 20.04, Val (V) = 20.45, Trp (W) = 23.85, Phe (F) = 25.18, Ile (I) = 26.05, Lys (K) = 26.70, Leu (L) = 27.33

Table 8. Quantification of the gas-phase sequencing of the intermediate DNA-peptide from the digested crosslinked complex.

Yields of the PTH-derivatives are shown in pmol, and only those residues present in the sequence are listed.
(Data from Prof J. Fothergill and Mr B Dunbar, Peptide Sequencing Service, University of Aberdeen)

Cycle	Asn	Gln	Thr	Gly	Glu	Ala	Tyr	Val	Ile	Lys	Leu	Sequence
1	14.59	----	----	----	----	0.00	3.81	4.11	5.31	5.78	4.65	(Gly)@*1
2	12.42	----	3482	----	----	0.00	1.24	20.28	1.59	0.929	----	Val*2
3	5.86	----	1127	----	----	0.00	----	11.07	----	----	----	(Thr)@*3
4	1.44	----	----	----	----	0.82	----	3.76	----	----	----	(Tyr)@*4
5	3.05	----	----	471.1	2.39	----	----	1.74	----	10.98	----	Lys
6	2.33	1.63	----	332.7	----	0.91	----	0.66	----	16.93	----	Lys
7	1.86	1.47	----	255.0	9.11	0.71	----	----	----	11.11	----	Glu
8	1.51	8.52	----	213.8	7.83	----	----	----	----	5.25	----	Gln
9	1.00	6.77	----	151.9	4.56	7.76	----	----	----	2.25	----	Ala
10	0.77	4.25	----	127.6	3.17	6.22	----	0.65	7.94	1.13	----	Ile
11	5.30	2.68	----	109.1	----	3.24	----	----	7.51	----	----	Asn
12	3.92	0.90	----	81.36	----	1.58	2.72	----	3.07	----	----	Tyr

@ Gly, Thr and Tyr of cycle 1, 3 and 4 respectively were determined only after a comparison with the Protein Sequence Database of the S polypeptide when the intermediate peptide sequence had been unambiguously identified as part of the trinucleotide recognition domain of the S polypeptide.

*1 PTH-Gly peak of cycle 1 (HPLC standard running time = 8.86 min) was obscured by the huge contamination, which appeared at HPLC running time = 8.21 min.

*2 The peak of cycle 2 at HPLC running time = 8.42 min was due to the contamination mixed with PTH-Gly residue from cycle 1, while PTH-Val was increased from 4.11 pmol to 20.28 pmol.

*3 The PTH-Thr peak (HPLC standard running time = 8.47 min) of cycle 3 was still affected by the contamination of cycles 1 and 2.

*4 The low PTH-Tyr yield of cycle 4 (HPLC standard running time = 15.61 min) was due to the crosslinking of the oligonucleotide to this amino acid which was unidentified by HPLC.

Table 9. Quantification of the gas-phase sequencing of the short DNA-peptide from the digested crosslinked complex.

Yields of the PTH-derivatives are shown in pmol, and only those residues present in the sequence are listed.
Data from Welmet Protein Characterisation Facility, University of Edinburgh.

Cycle	Asn	Gln	Thr	Gly	Glu	Ala	Tyr	Val	Ile	Lys	Leu	Sequence
1	0.85	0.28	0.35	29.46	1.57	1.22	0.37	1.21	5.50	0.61	1.10	Gly
2	0.21	0.36	0.20	5.53	0.40	0.46	0.08	16.72	5.36	0.24	0.60	Val
3	0.20	0.19	7.11	1.61	0.25	0.30	0.16	8.86	5.56	0.03	0.17	Thr
4	0.09	0.20	4.47	0.87	0.33	0.24	0.09	2.54	5.27	0.09	0.06	(Tyr)@
5	0.08	0.14	2.09	0.80	0.23	0.33	0.08	0.94	5.33	1.76	0.21	Lys
6	0.15	0.14	1.17	0.61	0.15	0.43	0.22	0.56	5.37	2.24	0.26	Lys

@ The Tyrosine of cycle 4 was determined only after a comparison with the Protein Sequence Database of the S polypeptide when the short peptide sequence had been unambiguously identified as part of the trinucleotide recognition domain of the S polypeptide.

3.7 *The interaction between DNA and mtase*

3.7.1 Crosslinking of ARD with the trinucleotide recognition sequence (-AAC-)

Early interest in type I R-M systems has focused on the specificity of their interactions with DNA. Amino acid sequence comparisons of S polypeptide recognition domains have not identified specific amino acids responsible for DNA sequence specificity [25,26]. ARDs of the type I S polypeptides which recognize different sequences show negligible similarity, while ARDs (within a family) that recognize the same trinucleotide reveal very high similarity. Examples are the ARDs of *Eco*KI and *Sty*SPI (A family), which recognize the same 5′ trinucleotide (AAC), and show 90% identity [26], and the ARDs of *Eco*AI and *Eco*EI (B family), which recognize the same 5′ trinucleotide (GAG), and show 80% identity [25]. Comparison of ARDs between families, *Sty*SBI (A family), and *Eco*AI and *Eco*EI (B family), which recognize the same 5′ trinucleotide (GAG), reveals a much lower (44%) identity [25]. The 5′ trinucleotide recognition sequence of *Eco*R124, GAA, is the same as the half-sequence of *Eco*RI and some amino acid sequence similarity between *Eco*R124 and *Eco*RI ENase possibly responsible for sequence recognition has been noted [114]. This evidence still does not allow the pinpointing of important amino acids.

The PTH-derivatives from peptide sequencing (Table 7,8, and 9) illustrate that all three peptides of the DNA-peptide crosslinked complexes are in the ARD of the mtase S polypeptide (Table 10). All three peptides give blank sequencing cycles where Tyr-27 is expected. Therefore, Tyr-27 in the S polypeptide was crosslinked to BrdU at position 7′ through the major groove of DNA by UV irradiation at 254 nm.

Table 10. Amino acid sequences obtained from the Br7′-mtase crosslinked complexes.

*Eco*KI						↓	
S subunit	SAGKL	PEGWV	IAPVS	TVTTL	IRGVT	YKKEQ	AINY
Long Br7′-peptide	– – –KL	PEG – V	IAPV–	TVTTL	I–GVT	– – KEQ	
Intermediate Br7′-peptide					–V –	– KKEQ	AINY
Short Br7′-peptide					GVT	– KK	

Conclusion from the results of amino acid sequencing, no amino acid corresponding to Tyr-27 (indicated by arrow) was identifiable in any of the Br7′-peptide crosslinked complexes sequenced.

3.7.2 The interaction between DNA and mtase: Base-flipping out hypothesis and Bookmark hypothesis.

The recent determination of the co-crystal structures of the two type II mtase *Hha*I C5 cytosine mtase [66, 69] and *Taq*I N6 adenine mtase [67], shows that they both possess an AdoMet binding catalytic domain joined to a second, dissimilar small domain responsible for DNA recognition. Almost all of the base-specific interactions between *Hha*I and the target bases occur in the major groove through two glycine-rich loops from the small domain [69]. After or during the base recognition and binding, there are extensive conformational changes. The three essential components of the methylation reaction, the target base, the catalytic nucleophile, and the methyl donor, are brought into close proximity to each other by the flipping out of the base from the double helix [69], allowing the methylation reaction to proceed. The type I *Eco*KI mtase (M_2S_1) contains two functions, recognizing target sequence and methylating target adenosine. It was suggested by computer modelling that the M subunits contain a catalytic domain similar to that of type II mtase [128], following the proposal of a universal catalytic domain structure and base flipping mechanism for AdoMet-dependent mtases [127].

In this study, the UV crosslinking of the target DNA and S polypeptide confirms the recognition function of the S polypeptide which closely contacts the target recognition bases of DNA from the major groove. This position in the major groove coincides with the small domain of type II *Hha*I mtase. The initial interaction between DNA and S polypeptide may partly be explained by the bookmark hypothesis [115].

The bookmark interaction starts through the fast association at the outside of the DNA helix which is stablilized by the formation of salt bridges (the attraction of positively charged amino acid side chains and the negatively charged phosphates). This process is followed by slow, partial interaction of the stacking of aromatic amino acid side chains between base-pairs (bookmarking), leading to bends in the DNA [115]. The bookmarking process is accompanied by a certain selectivity for A–T base-pairs which are less strongly stacked than G–C base-pairs. Since there is no evidence for the stacking of the aromatic amino acid side chain of mtase into DNA in any type of mtase and ENase, the recognition interaction between DNA and S polypeptide could better be described as close contact rather than stacking.

Tyrosine is one of the most photoreactive amino acids [123], and BrdU is a photoreactive analogue of thymidine [38]. The close contact between the Tyr-27 of the S polypeptide and the thymidine of the specific DNA might be due to the strong electrostatic binding by the positively charged end group of Lys-28 to the negatively charged phosphate group of DNA followed by the contact of the aromatic amino acid side chain of tyrosine and A-T base pairs. Successive A-T (or T-A) base pairs (-A-T(BrdU)-T-), which are less strongly stacked than G-C base pair, around the specific DNA site of crosslinking would, if this was the case, make the contact of the aromatic side chain of Tyr-27 easier. After the specific binding of DNA and mtase (M_2S_1), the contact of tyrosine (bookmarking) could cause a conformational change in the DNA widening the major groove. This would favour the flipping out of the adjacent adenosine from the helix for methylation.

Crosslinking of Tyr-27 to Br7′ confirms much of the previous work [25-28] on the specificity of Type I R-M systems. As expected the crosslinked amino acid, Tyr-27 is within the ARD of the S polypeptide and the specific DNA crosslinking site is within the trinucleotide recognition sequence (-AAC-).

3.7.3 The absence of photocrosslinking of CRD with the tetranucleotide recognition sequence (-GTGC-)

Crosslinking between the amino acid within the CRD and the base pair within the tetranucleotide recognition sequence (-GTGC-) was also expected. Unfortunately results of UV irradiation showed that no such crosslinking was obtained from Br17 (BrdU substitution at the only thymidine site within the tetranucleotide part of the recognition sequence, which is the T of the target methylation base pair A-T). The absence of photocrosslinking between DNA and protein does not necessarily mean that there is no close contact between DNA and protein.

Many studies of DNA-protein interaction through bromodeoxyuridine substituted photocrosslinking and amino acid sequencing have shown that crosslinking often occurs at nucleotides in the DNA sequence with at least two successive A-T (or T-A) base pairs. An example is the photocrosslinking of the basic region leucine zipper (bZIP) DNA-binding motif of GCN4, a yeast transcriptional activator [82], where the Ala-238 of GCN4 crosslinks to the thymidine at position +3 of the bZIP DNA-binding site. At the bZIP DNA-binding site, position +4 is also a T-A base pair which forms the successive A-T (or T-A) base pairs with the position +3. In the studies of *lac* repressor-operator interaction, two crosslinked points are identified, the thymidine at +3 is close to His-29 of the repressor peptide 23-33 and Tyr-17 in peptide 13-22 is close to T+14 [41,81]. The sequence of the *lac* operator is AATT from +1 to +4 and ATAA from +13 to +16, which are rich in A-T (or T-A) base pairs. Proteins of the Rel family are transcription factors containing a homology region of NF-kB p50. UV-induced photocrosslinking shows the aromatic amino acid side chains tyrosine 60 and histidine 67 of NF-kB p50 crosslinked to its bromosubstituted binding DNA at thymidine +4, +8 and +45 [113]. The crosslinked bases of T+4, T+8 (for Tyr-60) and T+5 (for His-67) are in a β-interferon regulatory element. Sequence from +1 to +10 of NF-kB binding site is GGGAAATTCC. This report also shows the crosslinking of DNA and protein occurs at an A-T (or T-A) rich area and with the electron rich aromatic amino acid side chains (Tyr-60 and His-67).

Photocrosslinking of bromodeoxyuridine-substituted DNA with protein may have been easier at areas rich in A-T base-pairs, and where the closely contacted amino acid is an electron rich aromatic residue. The fact that the photocrosslinking of the tetranucleotide recognition sequence (-GTGC-) was not observed in this study may be because the area is rich in G-C base pairs, and the photoreactive ability of G or C bases are too weak and there is no electron rich amino acid in the vicinity. At present, efficient photoreactive G or C base analogues which can substitute for G-C base pairs and improve the photoreactivity are not readily available.

3.8 *Comparison of related S subunits*

As described in section 3.7.1, the ARDs of different type I systems show little similarity unless they recognise the same trinucleotide. The identification of the contact between Tyr-27 and the first A-T base pair in the sequence -AAC- suggests that this amino acid may be conserved in other ARDs that recognize the AA dinucleotide within the trinucleotide target.

Comparison of the first 60 amino acids in all type I families reveals that there is some similarity in the first part of the ARDs in some groups of type I families. The alignments (Figure 40) illustrate that Leu-21 of the S polypeptide from families IA and IC or Leu-120 of the S polypeptide from the IB family is conserved. Two different S polypeptide alignments for *Sty*SBI have been published [25,32], one contained the six residue gap after Leu-21 [25] and the other did not [32]. The former alignment was used here to allow the line up of the *Sty*SBI Leu-21 with the leucines of the other enzymes. The motif LP-GWEW is well-defined in all IB family enzymes and is retained in part in all the other systems, except *Eco*BI. The amino acid sequences shown in bold prior to Leu-21 of *Eco*KI (Figure 40) are repeated to a certain extent at the junction between the central conserved region and the CRD (Figure 41). The conservation of components of the LP-GWEW motif in S subunits from all of the families regardless of their recognition site, and the fact that the motif is repeated, suggests that these amino acids have a structural role rather than a functional DNA recognition role. The alignment of the first 60 amino acids of ARDs after Leu-21 reveals that other similarities occur (bold underlined amino acids in Figure 40) only according to common dinucleotides in the enzyme DNA recognition sequences. This suggests that these amino acids have a DNA recognition role rather than a structural role. According to the amino acid similarities after Leu-21 the enzymes are divided into four groups of common dinucleotides in the target sequence: AA*, CA*, GA* and TA* (the base pair adjacent to the methylation adenine (A*) and the methylation A*-T base pair of the trinucleotide recognition sequence). Of particular note in the AA* group are Tyr-27, Asp-39, Tyr-40, Gly-52 and Lys-53, which are conserved in two ARDs of IA family (*Eco*KI, and *Sty*SPI) that recognize -AA*C-, but also in the ARD of the different type I family (IC family, *Eco*R124I) that recognizes -GAA*-. The crosslinked Tyr-27 of *Eco*KI is one of the conserved amino acids in the AA* group. The computer Database showed that Tyr-27 of the *Eco*KI was in the region (amino acids from 25-35) of the maximum surface probabilities of ARD (amino acids from 1-157 of the S polypeptide). The region of the maximum surface probabilities in the ARDs of the other two enzymes (*Sty*SPI and *Eco*R124I) also occur between amino acids 25 to 35 of the S polypeptide, coinciding with *Eco*KI. It is possible that Tyr-27 in the S polypeptide of the *Sty*SPI and *Eco*R124I systems plays the same role in contacting the AA* dinucleotide sequence as it does in the *Eco*KI system. Therefore the DNA-mtase bromodeoxyuridine-mediated crosslinking of the Type I enzymes *Sty*SPI and *Eco*R124 by UV irradiation would be predicted to be at the same place as it occurred in *Eco*KI: the thymidine of the A-T base pair adjacent to the methylation A*-T base pair and the conserved Tyr-27 of the ARD domain within the S polypeptide. If this hypothesis could be confirmed by the crosslinking experiments with *Eco*R124, then it could suggest that the target recognition domains of the S polypeptide may have similar tertiary structure features for recognizing parts of their DNA targets.

Figure 40. Schematic Diagram and Alignment of the first 60 amino acids of the S subunit ARDs.

Type I R-M system	Trinucleotide recognition sites		Alignment of the first 60 amino acids of the S subunit ARDs. 1	21		27		39,40		52,53	
(IA) *Eco*KI	AA*C		MSAGK**LPEGW**VIAPVSTVTT	**L**	IRGVT	**Y**	KKEQAINYLKD	**DY**	LPLIRANNIQN	**GK**	FDTTDLV
(IA) *Sty*SPI	AA*C		MNRGK**LPEGW**ATAPVSTVTT	**L**	IRGVT	**Y**	KKEQALNYLQD	**DY**	LPIIRANNIQN	**GK**	FDTTDLV
(IC) *Eco*R124I	GAA*		MSEMSYLEKLLDGVEV**EWL**P	**L**	GEITK	**Y**	EQPTKYLVKAK	**DY**	HDTYTIPVLTA	**GK**	TFILGYT
(IA) *Sty*LTIII	GA*G		MSGGK**LPEGW**ATSTINEMCN	**L**		**N**	PKLKL	**D**	DDLDVGFMPMAGVPTTYLGKCNFETK		
(IB) *Eco*AI	GA*G	(95)	EKPFE**LP**D**GWEW**TT	**L**	TRIAEI	**N**	PKIDVS	**D**	DEQEISFIPMPLISTKFDGSHEFEIK		
(IB) *Eco*EI	GA*G	(95)	EFPFE**LPEGWEW**IT	**L**	SEIATI	**N**	PKIEVT	**D**	DEQEISFVPMPCISTRFDGAHDQEIK		
(IA) *Eco*BI	TGA*		MSFNSTSK**E**	**L**	IEQN I	**N**	GLLSIH	**D**	SWLPISMDSVANITNGFAFKSSEFNN		
(IB) *Cfr*AI	GCA*	(95)	DKPFE**LP**A**GWEW**VR	**L**	GEAFYIEM	**G**	QSPSSQYYNQSEEGIPFFQGKADFGKKYPTA				
(IC) *Eco*prrI	CCA*		MSELSYLEKLMDGVEV**EW**LP	**L**	SKVFNLRN	**G**	YTPSKTKKEFWANGDIPWFR				
(IC) *Eco*DXXI	TCA*		MSELSYLEKLLDGVEV**EW**VT	**L**	GSMADIGT	**G**	SSNRQDESENGIYPFYVRSK				
(IA) *Eco*DI	TTA*		MSAGK**LP**VD**W**KTVELGELIK	**L**	STGKLDANAADNDGQYPFFTCAESVSQIN						

Comparison of the first 60 amino acids of the ARDs of type I R-M systems. The family to which each belongs is shown in brackets. Each group comprises ARDs which share a common dinucleotide including the methylation site (A*). The underlined bold amino acids indicate conservations within each dinucleotide recognition group, including the Tyr-27 crosslinking site in *Eco*KI, which may be involved in recognizing the nucleotide preceding the methylation target. Other amino acids conserved within a group are underlined.

Figure 41. Alignment of amino acids at the start of the ARDs and CRDs of the S subunits.

Type I R-M system		Trinucleotide recognition sites	Amino acids at the start of the S subunit ARDs	Amino acids at the start of the S subunit CRDs
(IA)	*Eco*KI	A A*C	(5) LPEGWVIAP	(208) LTEKWRN
(IA)	*Eco*DI	T T A*	(5) LPVDWKTVE	(179) LTKEWRS
(IA)	*Eco*BI		(1) MSFNSTSKE	(209) LTKEDKD
				(230) IPEDWSE
(IA)	*Sty*SPI	A A*C	(5) LPEGWATAP	(208) LTTALRN
				(217) SLIGWHS
(IA)	*Sty*LTIII	G A*G	(5) LPEGWATST	(204) LTKELHK
				(223) IPSLWKI
(IB)	*Eco*AI	G A*G	(101) LPDGWEWTT	(389) LPEGWEW
(IB)	*Eco*EI	G A*G	(101) LPEGWEWIT	(389) LPIGWEW
(IB)	*Cfr*AI	G C A*	(101) LPAGWEWVR	(380) LPEGWEW
(IC)	*Eco*R124I	G A A*	(11) LDGVEVEWLP	(200) KEGEVEW
(IC)	*Eco*prrI	C C A*	(11) MDGVEVEWLP	(196) KEGEVEW
(IC)	*Eco*DXXI	T C A*	(11) LDGVEVEWVT	(216) KEDEVEG

Amino acids at the start of the CRDs that line up with the amino acids at the start of the ARDs before the conserved leucine (Leu21 in *Eco*KI). Amino acids in bold are the LP-GWEW motif. The conservation of components of the LP-GWEW motif in S subunits from all of the families regardless of their recognition site, and the fact that the motif is repeated, suggests that the amino acids prior to Leu 21 of *Eco*KI have a structural role.

3.9 Suggestions for future work

Site directed mutagenesis of the S polypeptide and DNA binding analysis are suggested to further clarify the interaction between the DNA and mtase. The importance of Tyr-27 of S polypeptide for DNA recognition can be tested by gel retardation assays on mutant proteins. For example, changing Tyr-27 to Phe-27, which maintains the aromatic ring of tyrosine but lacks its hydroxyl group, will demonstrate the role of the hydroxyl group in the aromatic ring of tyrosine on the mtase recognition. Changing Tyr-27 to Leu-27, which maintains a large apolar side chain but lacks the aromatic ring, will illustrate the importance of the aromatic ring in binding to DNA. The specific binding of DNA and S polypeptide will not rely only on one amino acid, the crosslinked Tyr-27, therefore the site directed mutagenesis of the S polypeptide within the region (amino acids from 25-35) of the maximum surface probabilities of ARD very likely will provide information about the DNA-mtase recognition binding. If there is any significant binding between DNA and the mutant proteins, which containing photoreactive amino acid, UV induced photocrosslinking can then be achieved and the roles of these amino acid side chains involved in the photocrosslinking reaction can be more satisfactorily understood.

*Hha*I and *Hpa*II mtase can bind DNA containing mismatches at one of the methylation target bases and even methylate uracil [129, 130]. It was proposed that the affinity of DNA binding correlates inversely with the stability of the target base pair, while the nature of the target base appears irrelevant for complex formation [129]. In this study it is confirmed that S polypeptide makes close contact with the base next to the proposed "flipping -out" methylation target base but not with the methylation target base itself. In addition, different substitutions at the 5-methyl position of thymidine changed the binding affinity (section 3.3.2). Therefore it is suggested that the mismatches on the crosslinked base pair may provide more informations on the role of the potential binding sites on the crosslinked base of the target recognition sequence. For example, if the thymidine on the crosslinked base was replaced by the cytosine, in which the 4-oxygen of thymidine is replaced by an amine group and the 5-methyl group of thymidine is replaced by a hydrogen, then the binding affinity of mtase with this mismatch will provide information on the contributions of 4-oxygen and 5-methyl of thymidine (both are in the major groove) to the sequence specific binding by mtase. Or if the adenine on the complementary strand to the crosslinked base is replaced by the guanine, in which the 6-amine of adenine is replaced by oxygen in the major groove and the 2-hydrogen of adenine is replaced by an amine group in the minor groove, then the binding affinity of this mismatch with mtase may provide information on the role of the potential binding site of adenine to the complementary strand.

UV induced crosslinking between DNA and protein has been successful in this study by using four different nucleic acid analogues which were substituted at nine different sites of oligonucleotides, and three different wavelengths of UV irradiation. The close contact point of DNA and protein, however, has only occurred between one bromo-substitutable base (i.e. thymidine) and a photoreactive amino acid (tyrosine 27). Although it was used successfully in this research, UV induced crosslinking may not be the best experimental technique to routinely determine close contacts between DNA and protein.

References

[1] Bertani, G.,Weigle, J.J. (1953). *J. Bacteriol.* 65, 113-121.
[2] Meselson, M., Yuan, R. (1968). *Nature* 217, 1110-1114.
[3] Suri, B., Nagaraja, V., Bickle, T. A. (1984). *Curr. Top. Microbiol. Immunol.* 108, 1-9.
[4] Bickle, T.A. (1987). In Neidhart, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M. and Umbarger, H.E. (eds), DNA Restriction and Modification Systems. In *Escherichia coli* and *Salmonella typhimurium*: American Society for Microbiology, Washington, DC, pp.692-696.
[5] Murray, N.E., Daniel, A.S. Cowan, G.M. & Sharp, P.M. (1993). *Mol. Microbiol.* 9, 133-143.
[6] Szybalski, W. et al (1991). *Gene.* 100, 13-26.
[7] Janulaitis, A., Petrusyte, M., Maneliene, Z., Klimasauskas, S. and Butkus, V. (1992). *Nucl. Acids Res.* 20, 6043-6049.
[8] Kong, H., Roemer, S.E., Waite-Rees, P.A., Benner, J.S., Wilson, G.G. and Nwankwo, D.O. (1994). *J. Biol. Chem.* 269, 683-690.
[9] King, G., and Murray, N.E. (1994). *Trends in Microbiol.* 2, (12) 461-501.
[10] Sain, B. and Murray, N.E. (1980). *Mol. Gen. Genet.* 180, 35-46.
[11] Wilson, G.G. and Murray, N.E. (1991). *Annu. Rev. Genet.* 25, 585-627.
[12] Bickle, T. A. and Kruger, D.H. (1993). *Microbiol. Rev.* 57, 434-450.
[13] Dryden, D.T.F., Cooper, L.P., and Murray, N.E. (1993). *J. Biol. Chem.* 268, 13228-13236.
[14] Kan, N.C., Lautenberger, J.A., Edgell, M.H. and Hutchison, C.A. (1979). *J. Mol. Biol.* 130, 191-209.
[15] Bickle, T. A., Brock, C. and Yuan, R. (1978). *Proc. Natl. Acad. Sci. USA* 75, 3099-3103.
[16] Bickle, T. A. (1982). The ATP-depent restriction endonucleases. In *Nucleases*, ed. Linn, S. M. and Roberts, R. J. pp 85-108. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
[17] Ho, D.K., Wu, J.C., Santi, D.V., Floss, H. G. (1991). *Arch. Biochem. Biophys.* 284, 264-269.
[18] Pogolotti, A.L. Jr, Ono, A., Subramaniam, R. and Santi, D.V. (1988). *J. Biol. Chem.* 263, 7461-7464.
[19] Burckhardt, J., Weisemann, J., Hamilton, D. L. and Yuan, R. (1981). *J. Mol. Biol.* 153, 425-440.
[20] Vovis, G.F., Horiuchi, K. and Zinder, N. D. (1973). *Proc. Natl. Acad. Sci. USA* 71, 3810-3813.
[21] Burckhardt, J., Weisemann, J. and Yuan, R. (1981). *J. Biol. Chem.* 256, 4024-4032.
[22] Boyer, H.W. and Roulland-Dussoix, D. (1969). *J. Mol. Biol.* 41, 459-472.
[23] Glover, S. W. and Colson, C. (1969). *Genet. Res.* 13, 227-240

[24] Fuller-Pace, F.V., Cowan, G. and Murray, N.E., (1985). *J. Mol. Biol.* 186, 65-75
[25] Cowan, G.M., Gann, A.A.F., and Murray, N.E. (1989). *Cell* 56, 103-109.
[26] Fuller-Pace, F.V. and Murray, N.E., (1986). *Proc. Natl. Acad. Sci. USA* 83, 9368-9372.
[27] Kannan, P., Cowan, G.M., Daniel, A.S., Gann, A.A.F., and Murray, N.E. (1989). *J. Mol. Biol.* 209, 335-44.
[28] Gubler, M., Braguglia, D., Meyer, J., Piekarowicz, A. and Bickle, T.A. (1992). *EMBO J.* 11, 233-240.
[29] Nagaraja, V., Shepherd, J.C.W., and Bickle, T.A. (1985). *Nature* 316, 371-372.
[30] Gough, J.A. and Murray, N.E. (1983). *J. Mol. Biol.* 166, 1-19.
[31] Fuller-Pace, F.V., Bullas, L.R., Delius, H. Murray, N.E. (1984). *Proc. Natl. Acad. Sci. USA* 81, 6095-6099.
[32] Gann, A.A.F., Campbell, A.J.B., Collins, J.F., Coulson, A.F.W., Murray, N.E. (1987). *Mol. Microbiol.* 1, 13-22.
[33] Kneale, G.G. (1994) J. *Mol. Biol.* 243, 1-5.
[34] Meister, J., MacWilliams, M., Hubner, P., Jutte, H., Skrzypek, E., Piekarowicz, A. and Bickle, T.A. (1993). *EMBO J.* 12, 4585-4591.
[35] Abadjieva, A., Webb, M., Patel, J., Zinkevich, V. and Firman, K. (1994). *J. Mol. Biol.* 241, 35-43.
[36] Cooper, L.P. and Dryden, D.T.F. (1994). *J. Mol. Biol.* 236, 1011-1021.
[37] Newman, P.C., Nwosu, V.U., Williams, D.M., Cosstick, R., Seela, F. and Connolly, B.A. (1990). *Biochem.* 29, 9891-9901.
[38] Hutchinson, F. (1973). *Q. Rev. Biophys.* 6, 201-46.
[39] Berens, K.and Shugar, D. (1963). *Acta Biochem. Polon* .10, 25-48.
[40] Boyce, R.P., and Setlow, R.B. (1963). *Biochem. Biophys. Acta* 68, 446-454.
[41] Wick, K. L., and Matthews, K.S., (1991). *J. Biol. Chem.* 266, 6106-6112.
[42] Goeddel, D.V., Yansura, D.G., Winston, C. and Caruthers, M.H. (1978). *J. Mol. Biol.* 123, 661-687.
[43] Brennan, C.A., Van Cleve, M.D. and Gumport, R.I. (1986). *J. Biol. Chem.* 261, 7270-7278.
[44] Hayawaka, T., Ono, A. and Ueca, T. (1988). *Nucl. Acids Res.* 16, 4761-4776.
[45] Fliess, A., Wolfes, H., Seela, F., Pingoud, A., (1988). *Nucl. Acids Res.* 16, 11781-11793.
[46] Risse, G., Joss, K., Neuberg, M., Brüller, H.-J., and Müller, R. (1989). *EMBO J.* 8, 3825-3832.
[47] Yoon, C., Prive, G. G., Goodsell, D.S. and Dickerson, R.E. (1988). *Proc. Natl. Acad. Sci. USA* 85, 6332-6336.
[48] Kennard, O. and Hunter, W.N., (1989). *Q. Revi. Biophys.* 22, 327-379.
[49] Steitz, T.A., (1990). *Q. Rev. Biophys.* 23, 3, 205-280.
[50] Matthews, B.W. (1988). *Nature* 335, 294-295.

[51] Seeman, N.C., Rosenberg, J.M. and Rich, A. (1976). *Proc. Natl. Acad. Sci. USA* 73, 804-808.
[52] McClarin, J.A., Frederick, C,A., Wang, B.C., Green, P., Boyer, H.W., Grable, J. and Rosenberg, J.M. (1986). *Science* 234, 1526-1541.
[53] Otwinowski, Z., Schevitz, R. W., Zhang, R.G., Lawson, C.L., Joachimiak, A., Marmostein, R.Q., Luisi, B.F. and Sigler, P.B. (1988). *Nature* 335, 321-329.
[54] Ohlendorf, D.H., Anderson, W.F., Fisher, R.G., Takeda, Y. and Matthews, B.W. (1982). *Nature* 298, 718-723.
[55] Weber, I.T. and Steitz, T.A. (1984). *Proc. Natl. Acad. Sci. USA* 81, 3973-3977.
[56] Frederick, C.A., Grable, J., M., Samudzi, C., Jen-Jacobson, L. Wang, B.C., Greene, P.J., Boyer, H.W., and Rosenberg, J.M., (1984). *Nature* 309, 327-331
[57] Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M. and Harrison, S.C. (1988). *Science* 242, 99-107.
[58] Suck, D., Lahm, A. and Oefner, C.(1988). *Nature* 332, 6163, 465-468.
[59] Wu, J.C., and Santi, D.V. (1987). *J. Biol. Chem.* 262, 4778-4786.
[60] Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1994). *Nature* 368, 660-664.
[61] Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1994). *Structure* 2, 439-452.
[62] Kim, Y., Grable, J.C., Love, R., Greene, P. and Rosenberg, J.M. (1990). *Science* 249, 1307-1309.
[63] Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K. and Wilson, K.S. (1993). *EMBO J.* 12, 1781-1795.
[64] Athanassiadis, A., Vlasi, M., Kotsifaki, D., Tucker, P.A. Wilson, K.S. and Kokkinidis, M. (1994). *Nature Struct. Biol.* 1, 469-475.
[65] Cheng, X., Balendiran, K., Schildkraut, I. and Anderson, J.E. (1994). *EMBO J.* 13, 3927-3935.
[66] Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W., Roberts, R.J. (1993). *Cell* 74, 299-307.
[67] Labahn, J., Granzin, J., Schluckebier, G., Robinson, D.P., Jack, W.E., Schildkraut, I. and Saenger, W. (1994). *Proc. Natl. Acad. Sci. USA* 91, 10957-10961.
[68] Lin, S.Y. and Riggs, A. (1974). *Proc . Natl . Acad . Sci. USA* 71, 947-951
[69] Klimasauskas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994). *Cell* 76, 357-369.
[70] Williams, K.R. and Konigsberg, W.H. (1991). *Methods Enzymol.* 208, 516-539.
[71] Paradiso, P.R., Nakashima, Y. and Konigsberg, W.H. (1979). *J. Biol. Chem.*, 254, 4739-4744.
[72] Merrill, B.M., Williams, K.R., Chase, J.W. and Konigsberg, W.H. (1984). *J. Biol. Chem.* 259, 10850-10856.

[73] Cleghon, V. and Klessig, D.F. (1992). *J.Biol. Chem.* 267, 17872-17881.
[74] Evans, R.K., Johnson, J.D. and Haley, B.E. (1986). *Proc. Natl. Acad. Sci. USA* 83, 5382-5386.
[75] Meffert, R. and Dose, K. (1988). *FEBS Lett.* 239, 190-194.
[76] Meffert, R., Rathgeber, G., Schafer, H.J., Dose, K. (1990). *Nucl. Acids Res.* 18, 6633-6636.
[77] Nikiforov, T.T. and Connolly, B.A. (1992). *Nucl. Acids Res.* 20, 1209-1214.
[78] Welsh, J. and Cantor, C.R. (1985). *Trends Biochem. Sci.* 8, 505-508.
[79] Ogata, R., and Gilbert, W. (1977). *Proc. Natl. Acad. Sci. USA* 74, 4973-4976.
[80] Wolfes, H., Fliess, A., Winkler, F. and Pingoud, A. (1986). *Eur. J. Biochem.* 159, 267-273.
[81] Allen, T.D., Wick, K.L. and Matthews, K.S. (1991). *J. Biol. Chem.*, 266, 6113-6119.
[82] Blatter, E.E., Ebright, Y.W. and Ebright R.H. (1992). *Nature*, 359, 650-652.
[83] Ahmad, I. and Rao, D.N. (1994). *J. Mol. Biol.* 242, 378-388.
[84] Simpson, R.B. (1979). *Cell* 18, 277-285
[85] Hunkapiller, M.W., Granlund-Moyer, K. and Whiteley, N.W. (1986). *Gas-phase protein/peptide sequencer in Methods of Protein microcharacterization.* (Shively, J.E. ed.) Humana Press, Clifton, New Jersey.
[86] Matsudaira, P.T. (ed.) (1989). *A practical guide to protein and peptide purification for microsequencing.* Academic press, Inc.
[87] Brown, T and Brown, D., (1991). Modern machine-aided methods of oligonucleotide synthesis in Oligonucleotide and Analogues: A Practical Approach (Eckstein, F., ed.) pp.1-24, Oxford University Press, Oxford.
[88] Li, B.F.L., Reese, C.B. and Swann, P.F. (1987). *Biochem.* 26, 1086-1093.
[89] Nikiforov, T.T. and Connolly, B.A. (1991). *Tetrahedron Letters* 32, 3851-3854.
[90] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
[91] Powell, L.M., Dryden, D.T.F., Willcock, D.F., Pain, R.H., and Murray, N.E. (1993). *J. Mol. Biol.* 234, 60-71.
[92] Hames, B.D. and Rickwood, D. (1990). *Gel electrophoresis of proteins.* (2nd ed.) IRL Press at Oxford University Press.
[93] Laemmli, U.K. (1970). *Nature* 227, 680-685.
[94] Shapiro, A.L., Vinuela, E., and Maizel, J.V. (1967). *Biochem. Biophys. Res. Commun.* 28, 815.
[95] Weber, K. and Osborn, M. (1972). *J. Biol. Chem.* 244, 4406.
[96] Ornstein, L. (1964). *Ann. NY Acad. Sci.* 121, 321.
[97] Davis, B. J. (1964). *Ann. NY Acad. Sci.* 121, 404.

[98] Switzer, R.C., Merril, C.R.and Shifrin, S. (1979). *Anal. Biochem.* 98, 231.
[99] Neurnth, H. (ed.) (1963). *The proteins-composition, structure and function.* Academic Press.
[100] Gabeloteau, C. and Desnuelle, P. (1957). *Arch. Biochem. Biophys* . 69, 475
[101] Wang, S.-S., and Carpenter, F.H. (1967). *Biochemistry* 6, 215.
[102] Hummel, B.C. (1959). *Can. J. Biochem. Physiol.* 37, 1393.
[103] Schwert, G.W., Neurath, H., Kauffman, S. and Snoke, J.E. (1948). *J. Biol. Chem.* 172, 221.
[104] Erlanger, B.F., Cooper, A.G. and Cohen, W. (1966). *Biochemistry* 5, 190-196.
[105] Favre, A., Yaniv, M., Michelson, A.M. (1969). *Biochem. Biophys. Res. Commun.* 37, 266-271.
[106] Frischauf, A.M.,Scheit, K.H., (1973). *Biochem. Biophys. Res. Commun..* 53, 1227-1233.
[107] Dietz, T.M., von Trebra, R.J., Swanson, B.J. and Koch, T.H. (1987). *J. Am. Chem. Soc.* 109, 1793.
[108] Ito, S., Saito, I., Matsuura, T., Helene, C. (1980). *J. Am. Chem. Soc.* 102, 7535.
[109] Willis, M. C., LeCuyer, K A., Meisenheimer, K.M., Uhlenbeck, O.C. and Koch, T.H. (1994). *Nucl. Acids Res.* 22, 4947-4952.
[110] Swanson, B.J., Kutzer, J.C. and Koch, T.H. (1981). *J. Am. Chem. Soc..* 103, 1274
[111] Willis, M.C., Hicke, B.J., Uhlenbeck, O.C., Cech, T.R., and Koch, T. (1993). *Science* 262, 1255-1257.
[112] Hicke, B.J., Willis, M.C., Tad H. Koch, T.H. and Cech, T.R., (1994). *Biochemistry* 33, 3364-3373.
[113] Liu, J., Sodeoka, M., Lane, W.S. and Verdine, G.L. (1994). *Proc. Natl. Acad. Sci. USA* 91, 908-912.
[114] Price, C., Lingner, J., Bickle, T.A., Firman, K. and Glover, S.W. (1989). *J. Mol. Biol.* 205, 115-125.
[115] Hélène, C. and Maurizot, J.-C. (1981). *CRC Crit. Rev. Biochem.* 10, 213-258.
[116] Nagaraja, V., Stieger, M., Nager, C., and Bickle, T.A., (1985). *Nucl. Acids Res.* 13, 389-399.
[117] Ravetch, J.V., Horiuchi, K., and Zinder, N.D., (1978). *Proc. Natl. Acad. Sci. USA* 75, 2266-2270.
[118] Lautenberger, J.A., Kan, N.C., Lackey, D., Linn, S.M., Edgell, M.H., and Hutchison III, C.A., (1978). *Proc. Natl. Acad. Sci. USA* 75, 2271-2275.
[119] Nagaraja, V., Shepherd, J.C.W., Pripfl, T., and Bickle, T.A., (1985). *J Mol. Biol.* 182, 579-585.
[120] Suri, B., Shepherd, J.C.W., Bickle, T.A., (1984). *EMBO J.* 3, 575-579.

[121] Piekarowicz, A., and Goguen, J.D. (1986). *Eur. J. Biochem.* 154, 295-298.
[122] Tyndall, C., Meister, J. and Bickle, T.A. (1994). *J. Mol. Biol.* 327, 266-274.
[123] Smith, K C. (1969). *Biochem. Biophys. Res. Commun.* 34, 3, 354-357.
[124] Connolly, B.A. and Newman, P.C. (1989). *Nucl. Acids Res.* 17, 4957-4974.
[125] Nikiforov, T.T. and Connolly, B.A. (1991). *Tetrahedron Letters* 32, 2505.
[126] Williams, K.R. and Konigsberg, W.H. (1991). *Method in Enzymology* 208, 516-539.
[127] Schluckebier, G., O'Gara, M., Saenger, W. and Cheng, X. (1995) *J. Mol. Biol.* 247, 16-20.
[128] Dryden,D.T.F., Sturrrock, S.S., and M. Winter, (1995). *Nature Stru. Bio.* 2 [8] 632-35.
[129] Klimasauakas, S. and Roberts, R.J. (1995). *Nucl. Acids Res.* 23, 1388-1395.
[130] Yang, A.S., Shen, J.C. Zingg, J. M. Mi, A and Jones P. A. (1995). *Nucl. Acids Res.* 1380-1387.
[131] Schluckebier, G., Labahn, J., Schildkraut, I. and Saenger, W. (1995). *Gene*, 157, 131-134.

 Nucleic Acids Research, 1995, Vol. 23, No. 7 1177–1183

Tyrosine 27 of the specificity polypeptide of *Eco*KI can be UV crosslinked to a bromodeoxyuridine-substituted DNA target sequence

A. Chen[1,2], L. M. Powell[1], D. T. F. Dryden[1], N. E. Murray[1,*] and T. Brown[2]

[1]Institute of Cell and Molecular Biology and [2]Department of Chemistry, University of Edinburgh, The King's Buildings, Edinburgh EH9 3JR, UK

Received December 8, 1994; Revised and Accepted February 24, 1995

ABSTRACT

The specificity (S) subunit of the restriction enzyme *Eco*KI imparts specificity for the sequence AAC(N_6)GTGC. Substitution of thymine with bromodeoxyuridine in a 25 bp DNA duplex containing this sequence stimulated UV light-induced covalent crosslinking to the S subunit. Crosslinking occurred only at the residue complementary to the first adenine in the AAC sequence, demonstrating a close contact between the major groove at this sequence and the S subunit. Peptide sequencing of a proteolytically-digested, crosslinked complex identified tyrosine 27 in the S subunit as the site of crosslinking. This is consistent with the role of the N-terminal domain of the S subunit in recognizing the AAC sequence. Tyrosine 27 is conserved in the S subunits of the three type I enzymes that share the sequence AA in the trinucleotide component of their target sequence. This suggests that tyrosine 27 may make a similar DNA contact in these other enzymes.

INTRODUCTION

The type IA restriction and modification (R–M) enzyme of *Escherichia coli* K12 (*Eco*KI) comprises three kinds of polypeptide, R (restriction), M (modification) and S (specificity) (1; for reviews see refs. 2–4). All three kinds of polypeptide are essential for restriction while only two of them, M and S, are required for modification (5,6). In the presence of *S*-adenosyl-L-methionine (AdoMet), both endonuclease and methyltransferase (mtase) recognize DNA that contains the recognition sequence A<u>A</u>C(N_6)G<u>T</u>GC (7,8). *Eco*KI methylates the N6 position of adenine at the underlined positions in the recognition sequence (9). Hemimethylated DNA is the preferred substrate for methylation (5,10,11) while unmodified DNA is restricted (12).

Many lines of evidence indicate that for type I R–M systems the S polypeptide imparts the specificity for a specific DNA sequence (13–15). Indeed it has been shown that the S polypeptide of each type I family (IA, IB or IC) has two variable DNA target recognition domains (TRDs) of negligible similarity, each of which recognizes one half of the bipartite recognition sequence (16–19). In the case of the type IA enzymes the amino recognition domain (ARD) recognizes the trinucleotide component of the recognition site while the carboxy recognition domain (CRD) recognizes the tetranucleotide component (16,17,20). In addition to the variable regions, the type IA S subunits have two relatively short regions which are well conserved within the family (21,22). The arrangement of conserved and variable regions within the type IB and IC families is shown in Figure 1 for comparison with type IA. Repeat sequences found in the conserved regions of type IA, IB and IC S polypeptides (18,23,24) have been implicated in interactions with the M subunits (25–27).

Amino acid sequence comparisons of S polypeptide recognition domains have not identified specific amino acids responsible for DNA sequence specificity (16,17). ARDs which recognize different sequences show negligible similarity, while comparison of ARDs (within a family) that recognize the same trinucleotide, reveals 80–90% identity (16,17). Comparison of ARDs recognizing the same trinucleotide between families, while showing much lower (44%) identity (16), still does not allow the pinpointing of important amino acids.

Crystal structures have been reported for several type II endonucleases, including the apo-form of *Bam*HI (28,29) and *Pvu*II (30), and complexes of *Eco*RI, *Eco*RV and *Pvu*II with their cognate DNAs (31–33). Recently the structures of two type II mtases have also been reported (34,35).

In the absence of structural information on the amino acids involved in DNA recognition in type I systems we have used a bromouracil-mediated photocrosslinking method (36–42) to identify amino acids in close proximity to the target sequence.

MATERIALS AND METHODS

Oligonucleotide synthesis and preparation of *Eco*KI methyltransferase

Complementary oligonucleotides (25mers) containing the *Eco*KI recognition sequence (AAC(N_6)GTGC) or derivatives of it were synthesized on an Applied Biosystems 380 B DNA Synthesizer utilizing β-cyanoethyl phosphoramidite chemistry (43). 5-Bromodeoxyuridine (BrdU) is an isosteric photoreactive analog of

* To whom correspondence should be addressed

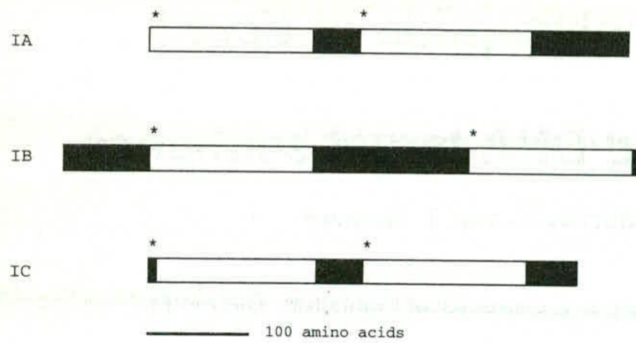


Figure 1. The organization of the S subunits of the three type I R–M system families. The regions conserved within a family are shown as shaded boxes while the ARD and CRD which recognize each part of the DNA target are shown as open boxes. Also shown is the repeated region (*) identified in the Results section in all type I S subunits.

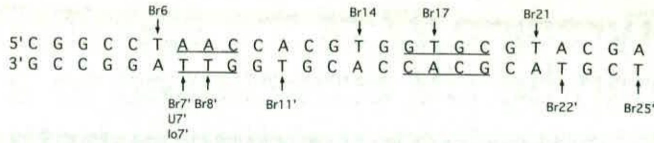


Figure 2. The 25 bp DNA duplex containing the *Eco*KI target sequence (underlined). This duplex is referred to as S25. The sites of substitution of thymine by modified bases are indicated and numbered from the 5′ end of the top strand with the addition of a prime (′) indicating a substitution on the bottom strand. The duplexes containing a single substitution of bromodeoxyuridine, deoxyuridine or iododeoxyuridine are designated Br6, Br7′, Br8′, Br11′, Br14, Br17, Br21, Br22′, Br25′, U7′ and Io7′. The duplex Br500 contains BrdU at all of the indicated sites.

thymidine (44). For 5-BrdU-substituted DNA, 5′-4,4′-dimethoxytrityl 5-bromo-2′-deoxyuridine-3′-*N,N*-diisopropyl cyanoethyl phosphoramidite was obtained from Link Technologies (UK). BrdU was incorporated either at a single position to give duplexes designated Br6, Br14, Br17, Br21, Br7′, Br8′, Br11′, Br22′ and Br25′ or at all available positions, designated duplex Br500, within the complementary oligonucleotides (Fig. 2). The unmodified duplex is called S25. The same DNA synthesis procedures were also used for the preparation of deoxyuridine and 5-iododeoxyuridine substituted DNA, duplexes U7′ and Io7′. All oligonucleotides were purified and analyzed by high performance liquid chromatography (HPLC). The oligonucleotides of the duplex were labelled respectively on the top strand, bottom strand, or both strands at the 5′-hydroxyl position using [γ-^{32}P]ATP (3000 Ci/mmol, Amersham) and T4 polynucleotide kinase (S. Bruce, ICMB). Unincorporated ATP was then removed using G25 spun column chromatography (45). Hybridization and DNA concentration determination were as described previously (46).

*Eco*KI mtase was prepared as described previously (6). Mtase stock solution was desalted by using a G25 Sephadex gel filtration column (Pharmacia, PD-10 column) and Tris–HCl buffer (20 mM Tris–HCl, pH 8.0, 100 mM NaCl). The concentration of mtase was determined by UV spectrophotometry using an extinction coefficient at 280 nm of 8.42 for a 1% (w/v) solution in a 1 cm pathlength cell (46).

Conditions for binding and UV crosslinking of substituted DNA

Gel retardation assays (46) were used to investigate the binding of the mtase with DNA oligonucleotides. To determine the optimum conditions for UV crosslinking, a solution of mtase (200 nM), AdoMet (100 μm) and DNA duplex (10 nM) in a total volume of 50 μl was put in a multiwell plate on ice and irradiated with a UV hand-held lamp. Different wavelengths of UV light and different irradiation times were investigated in order to obtain the optimum yield of crosslinked complex. In these experiments duplexes S25, Br7′, Br8′, Br11′, Br22′ and Br25′ were used. Crosslinking was examined by SDS–PAGE (47) with a 12% polyacrylamide gel of the mtase–DNA complexes, followed by silver staining (BioRad silver stain kit), or autoradiography and quantification of crosslinking efficiency by scintillation counting. To identify which one of the two DNA strands was crosslinked to protein, [γ-^{32}P]ATP was used to label one or other or both strands of the duplex. After the UV irradiation the crosslinked mixture was boiled for 2 min and cooled rapidly by quenching with liquid nitrogen before loading immediately on to a 10% polyacrylamide SDS–7 M urea gel. DNA duplexes are denatured by the boiling and quenching procedure, allowing the crosslinked DNA strands to be separated effectively from uncrosslinked DNA strands during electrophoresis on this gel.

For large scale experiments, DNA duplex Br7′ (20.4 μM), desalted mtase (8.8 μM) and AdoMet (70 μM) in a total volume of 4.6 ml were incubated for 15 min at 22°C. The reaction mixture was then spread to form a thin layer on the inside of the lid (12.5 cm × 8.5 cm) of a multiwell plate which was kept on ice for UV irradiation at 254 nm by UV handlamp set (Model UVGL-58, UVP International Ultra-Violet Products). Samples were directly irradiated by UV at a distance of 7 mm for 10 min. Crosslinking was monitored by autoradiography of SDS–PAGE gels of the irradiated samples and the efficiency of crosslinking was determined by scintillation counting of the appropriate gel slices.

Purification of large scale crosslinking complexes and trypsin digestion

The crosslinked complex was separated from the free DNA using a denaturing gel (9% polyacrylamide SDS–PAGE). The crosslinked band was cut from the gel, crushed and the DNA–protein complex eluted by shaking the crushed gel in 10 ml of elution buffer (0.5 M ammonium acetate, 1.0 mM EDTA, 0.1% SDS, 10.0 mM $MgCl_2$, pH 7.5 adjusted with NaOH) (48) at room temperature for 8 h. The crushed gel was removed by centrifugation. The solution of crosslinked complex was desalted on a G25 Sephadex gel filtration column, equilibrated with distilled water, before being freeze-dried.

The crosslinked complex was resuspended in 1 ml of ammonium acetate (200 mM, pH 8.5) or 10 mM Tris–HCl pH 8.0 and then digested by trypsin (Boehringer Mannheim). Possible chymotrypsin contamination of the trypsin stock solution (10 mg/ml trypsin in 100 mM Tris–HCl, 100 mM $CaCl_2$, pH 8.0) was inactivated using diphenyl carbamyl chloride (DPCC; Light & Co. Ltd) (49). A trypsin working solution was prepared by diluting the 10 mg/ml stock solution 10-fold with 0.2 M ammonium acetate pH 8.5 or 10 mM Tris–HCl pH 8.0. 1 ml of this 1 mg/ml trypsin working solution was added to the resuspended crosslinked complex and incubated for 12 h at 37°C. The peptide–DNA crosslinked complex was purified from digested peptides and trypsin on an Econo-Pac Q ion exchange cartridge (BioRad). After sample loading the column was washed with salt-free phosphate buffer (20 mM KH_2PO_4, 5% ethanol, pH 6.8). A 100 ml gradient from 0 to 0.6 M KCl in phosphate buffer (pH 6.8), 5% ethanol was then applied with a flow rate of 48 ml/h. The radioactive fractions were determined with a Beckmann LS7000 scintillation counter by adding 1.0 µl of elution fractions to 2.5 ml of Ecoscint (National Diagnostics). Radiolabelled peptide–DNA fractions were pooled and desalted with a G25 Sephadex gel filtration column before freeze-drying. To obtain the shortest crosslinked peptide for sequencing, the purified digested sample was treated with trypsin a second time. The purified peptide–DNA was analyzed by SDS–PAGE and sequenced by Mr B. Dunbar and Prof. J. Fothergill at Aberdeen University and the WELMET service at Edinburgh University. The WELMET samples were resuspended in deionised water, filtered through a polyvinyl difluoride membrane in a Prospin sample preparation cartridge (Applied Biosystems). The complex was trapped on the filter, washed with deionised water and dried before sequencing (50).

RESULTS

Gel retardation experiments for the unmodified DNA duplex and all of the bromodeoxyuridine substituted duplexes showed that with 10 nM DNA, 200 nM mtase was sufficient to bind 84–94% of the DNA (Table 1). Therefore, any variation in crosslinking efficiency with these duplexes can be assigned to the efficiency of the photoreaction rather than altered binding affinity.

Table 1. Efficiency of binding and crosslinking of duplexes to mtase

Duplex	Duplex bound (%)		Duplex crosslinked (%)	
S25	84 ± 4	(5)	3.9 ± 0.2	(6)
Br500	92	(1)	12.2	(1)
Br6	83 ± 3	(9)	3.0	(1)
Br7′	79 ± 4	(5)	12.3 ± 0.2	(5)
Br8′	87 ± 7	(2)	4.3 ± 0.3	(2)
Br11′	77 ± 15	(2)	4.1 ± 0.2	(2)
Br14	74 ± 4	(5)	3.8	(1)
Br17	78 ± 4	(5)	3.0	(1)
Br21	75 ± 5	(5)	4.1	(1)
Br22′	84 ± 8	(2)	4.4 ± 0.1	(2)
Br25′	77 ± 13	(2)	4.3 ± 0.1	(2)

Binding and crosslinking efficiency are represented by the fraction of radiolabelled DNA, determined by scintillation counting, in the bound or crosslinked complex relative to the total amount of DNA. The number of measurements is shown in brackets.

Preliminary experiments showed that maximum crosslinking was achieved by 10 min irradiation at 254 nm (results not shown). The percentage crosslinking is 3–4.4% for all duplexes except duplexes Br500 and Br7′ (Table 1). The Br500 and Br7′ duplexes crosslinked with 12.2–12.3% efficiency. This strongly suggests that the Br7′ substitution is the site of crosslinking to the protein. It was confirmed that crosslinking occurred only to the bottom strand by radiolabelling each DNA strand separately and ensuring that the duplex was denatured during electrophoresis on 10% polyacrylamide SDS–7 M urea gels (results not shown). This implies that the 3–4.4% level of crosslinking observed with the other duplexes is due to crosslinking by the normal thymine base at position 7′. No crosslinking was observed when bovine serum albumin was substituted for the mtase, or when a non-specific duplex lacking the recognition sequence was used (results not shown).

The substitution of the thymine at position 7′ with uracil and iodouracil reduced the crosslinking efficiency to the mtase to ~23 and 30% respectively of that obtained with Br7′ (results not shown), i.e. not significantly different from unsubstituted DNA. Gel retardation experiments (Fig. 3) showed that the Br7′ duplex bound to the mtase as well as the unmodified duplex (dissociation constant, K_d <10 nM) but that the U7′ and Io7′ duplexes bound much more poorly with K_d >30 nM. The weaker binding affinities cannot, however, account for the lower crosslinking efficiency of these duplexes as, under the conditions used for crosslinking (680 nM mtase, 2.0 µM DNA, 60 µM AdoMet in 200 µl), the amount of complex formed for U7′ and Io7′ DNA was not significantly different from that for unsubstituted DNA (results not shown). The van der Waals radii of bromine (Br7′) and the methyl group (S25) are very similar (44) while those of hydrogen (U7′) and iodine (Io7′) are smaller and larger respectively. The differences in crosslinking efficiency are probably due, therefore, to the different size of the group substituted at the 5 position of the pyrimidine ring affecting the contacts at the DNA–protein interface. Therefore, in spite of iododeoxyuridine being intrinsically more photoreactive than BrdU (44), the large size of the iodine probably disrupts the DNA amino acid contacts needed for crosslinking.

Using the Br7′ duplex a large quantity of DNA–mtase crosslinked complex was prepared. The procedures for eluting the complex from the gel and desalting gave high recoveries of radiolabelled sample (90–97% and 89–93% for each step respectively). Digestion of the complex with trypsin and purification of the peptide–Br7′ resulted in 46–69% recovery for this step. The conditions used for the ion-exchange step eluted the peptide–Br7′ complex at 0.5 M KCl. The overall recovery of the final product was 2.0–2.5% of the total DNA used.

Figure 4 shows an analysis of the crosslinked complexes used for amino acid sequencing. Figure 4A shows a silver-stained denaturing gel of the mtase crosslinked to Br7′ duplex. Between the bands representing the S and M subunits of the mtase is a new band containing the radiolabelled duplex linked to the S subunit.

Figure 4B shows an autoradiograph of a SDS–20% acrylamide gel of the intermediate peptide–Br7′ complex (lane 4), and long peptide–Br7′ complex (lane 5), produced by digestion in the presence or absence of ammonium acetate respectively. The presence of ammonium acetate increased the solubility of the S-polypeptide–Br7′ complex and resulted in a more complete trypsin digest. Figure 4C shows the short peptide–Br7′ complex (lane 2) produced by twice digesting with trypsin, the long peptide–Br7′ complex (lane 3) and the S subunit–Br7′ complex (lane 4).

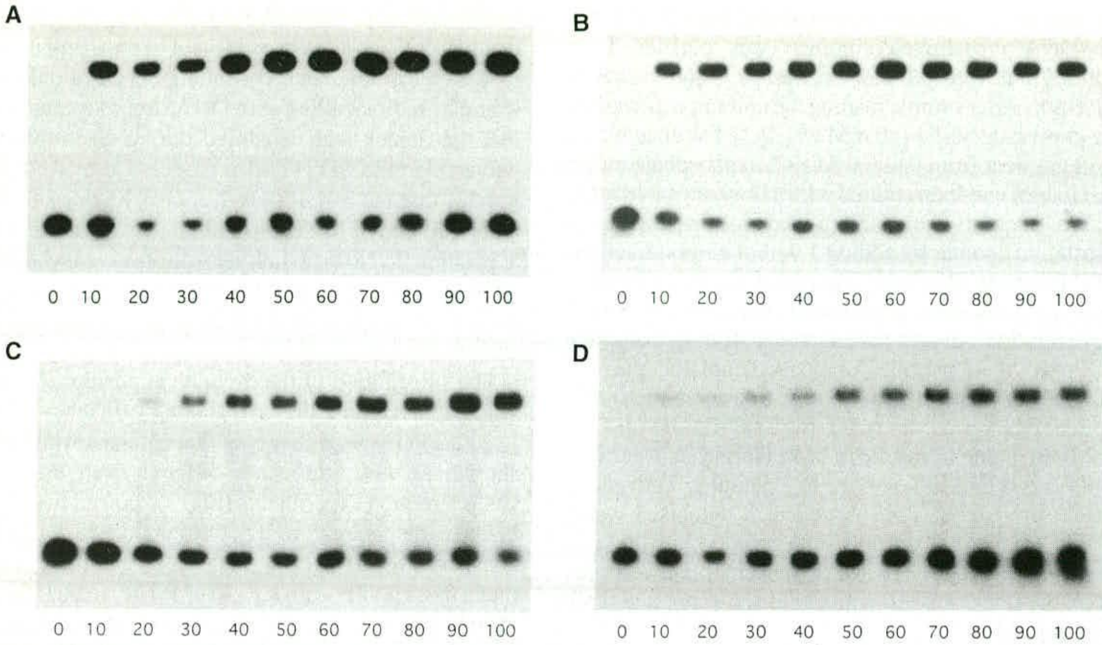


Figure 3. Binding of *Eco*KI mtase to labelled target sequences, as determined by gel retardation. Top band: DNA–mtase complex; lower band: free DNA. In all cases, 100 μM AdoMet and 0.1 nM target duplex DNA were used; the mtase concentration (nM) is noted below each lane. (**A**) Unmodified S25 duplex; (**B**) Br7′ duplex; (**C**) U7′ duplex and (**D**) Io7′ duplex.

Table 2. Amino acid sequences obtained from the crosslinked complexes containing duplex Br7′

*Eco*KI							
S subunit[a]	SAGKL	PEGWV	IAPVS	TVTTL	IRGVT	YKKEQ	AINY
Short peptide–Br7′					GVT	–KK	
Intermediate peptide–Br7′					–V–	–KKEQ	AINY
Long peptide–Br7′	– – –KL	PEG–V	IAPV–	TVTTL	I–GVT	– –KEQ	

[a]The S subunit in *Eco*KI lacks the N terminal methionine and the sequence starts at Ser2.
The first and third amino acids in the intermediate peptide–Br7′ complex and the first three amino acids of the long peptide–Br7′ complexes were not identifiable due to the presence of contaminants. The later sequencing cycles of the long peptide–Br7′ complex were difficult to interpret due to lack of material.
No amino acid corresponding to Tyr27 was identifiable in any of the peptide–Br7′ complexes sequenced.

The long peptide–Br7′ complex (Fig. 4B, lane 5; Fig. 4C, lane 3) and the intermediate peptide–Br7′ complex (Fig. 4B, lane 4) contained impurities, which made analysis of the first three amino acid sequencing cycles for the former and the first and third cycles for the latter impossible. This problem was avoided with the short peptide–Br7′ complex (Fig. 4C, lane 2) by absorbing it to a PVDF membrane and washing away the contaminants. The sequences obtained are shown in Table 2 along with the first 34 amino acids of the S subunit. The sequences obtained are uniquely found in the S subunit of *Eco*KI. Both the short and intermediate peptide–Br7′ complexes were produced by trypsin cutting after Arg23. The short peptide terminates after Lys29 but this site in the intermediate peptide has not been cut by trypsin, while in the long peptide there is no cutting after Lys5, Arg23 or Lys29. These incomplete digestions are possibly due to steric interference by the crosslinked DNA. All three peptides give blank sequencing cycles where Tyr27 is expected. Therefore, the bromodeoxyuridine at position 7′ in the DNA duplex has covalently crosslinked to Tyr27 in the S subunit.

The identification of the contact between Tyr27 and the first A–T base pair in the sequence AAC suggested that it might be conserved in other ARDs that recognize the AA dinucleotide within the trinucleotide target. Other ARDs which have a common dinucleotide sequence comprising CA or GA are also compared (Fig. 5A).

In the alignments shown in Figure 5A it is apparent that all of the ARDs have a conserved leucine (Leu21 in *Eco*KI). The two published alignments for *Sty*SBI are different (16,23), one containing the six residue gap after Leu21 (16) and the other not (23). We prefer the former alignment as this allows the line up of the *Sty*SBI Leu21 with the leucines of the other enzymes. Other alignments around Leu21 are probably equally valid and acceptable but we chose this line-up to provide us with a boundary between the LP-GWEW-containing structural regions of the S

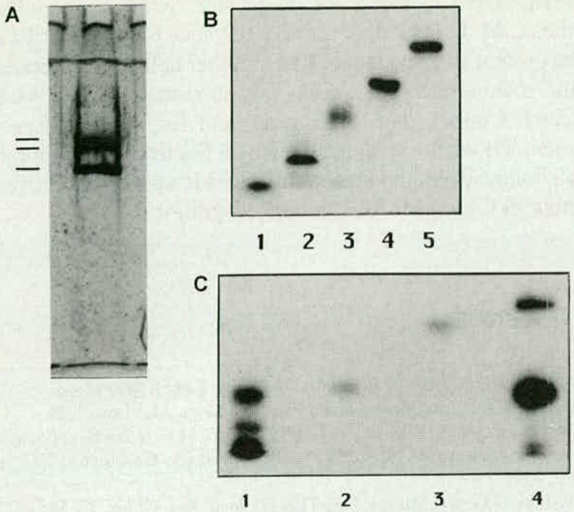


Figure 4. (**A**) A silver-stained SDS–12% polyacrylamide gel of mtase after crosslinking to duplex Br7′. The S subunit crosslinked to the duplex is visible between the S subunit (lower band) and the M subunit (upper band). (**B**) An autoradiogram of a 20% polyacrylamide SDS gel of the peptide–Br7′ crosslinked complexes after trypsin digestion and purification. Lanes 1, 2 and 3 are [γ-^{32}P]ATP, bottom strand of Br7′ and Br7′ duplex respectively. Lane 4 shows the intermediate peptide–Br7′ duplex complex obtained after trypsin treatment in the presence of ammonium acetate and lane 5 the long peptide–Br7′ duplex complex obtained after trypsin treatment in the absence of ammonium acetate. (**C**) An autoradiogram performed as in (B). The upper bands in lanes 1–4 are a ^{32}P-labelled Br7′ duplex, a double-digested short peptide–Br7′ crosslinked complex, a long peptide–Br7′ crosslinked complex and an S subunit–Br7′ crosslinked complex respectively. Additional bands are visible in lanes 1 and 4. In lane 1, the two bands running ahead of the ^{32}P-labelled Br7′ duplex are ^{32}P-labelled bottom strand of Br7′ and [γ-^{32}P]ATP. In lane 4 the two bands running ahead of the S subunit–Br7′ crosslinked complex are ^{32}P-labelled Br7′ duplex DNA and [γ-^{32}P]ATP, respectively.

subunit ARDs and the region containing conserved amino acids more likely to be important in DNA recognition, as discussed below.

The motif LP-GWEW is well-defined in all IB family enzymes and is retained in part in all of the other systems, except perhaps *Eco*BI. The small conserved region shown at the start of the IC family S polypeptides (Fig. 1) is also seen in the IA family sequences (Fig. 5A). This region in the IA family is apparent from previous alignments but was not commented on at that time (23). Sequences from the junction between the central conserved region and the CRD have also been aligned with the LP-GWEW motif (Fig. 5B). The repeat of the LP-GWEW motif in this region has been observed for members of the IB family (18), the IC family and *Eco*KI (24), but its presence in other type I systems, although apparent from earlier alignments (16,23), has also not been previously noted.

The conservation of components of the LP-GWEW motif in S subunits from all of the families regardless of their recognition site, and the fact that the motif is repeated, suggests that the amino acids shown in bold (but not underlined) prior to Leu21 of *Eco*KI have a structural role. After Leu21 other similarities occur only within the grouping of enzymes according to common dinucleotides in their recognition sequence (bold underlined amino acids in Fig. 5A), which suggests that these amino acids have a DNA recognition rather than a structural role. Of particular note are Tyr27, Asp39, Tyr40, Gly52 and Lys53, which are conserved in the two ARDs that recognise AAC, but also in the ARD of *Eco*R124I that recognises GAA, even though this system is in a different type I family from the other two. It should be noted that the second adenine in this trinucleotide is the substrate for methylation by *Eco*R124I (51). Therefore it is possible that the Tyr27 in the S subunit of *Eco*R124I also contacts the first A–T base pair in the GAA target sequence.

DISCUSSION

UV crosslinking studies have been widely used to investigate DNA–protein interactions. Many have simply exploited the intrinsic photoreactivity of the nucleotides, in particular thymine (52–55). Other approaches have relied on nucleotide analogs with enhanced photoreactivity (36–42,56–60). The most widely used of these has been BrdU in which the 5-methyl group of thymidine, which projects into the DNA major groove, is substituted by the similarly sized bromine atom (36–42). The crosslinking reaction proceeds following the loss of the bromine atom to give a free radical intermediate which can react with suitable acceptors, including amino acid side chains (60).

A

```
Type I          Trinucleotide    Alignment of the first 60 amino acids of the S subunit ARDs
R-M             recognition
system          sites
                                                           21    27              39,40           52,53
(IA) EcoKI       AA*C                  MSAGKLPEGWVIAPVSTVTT L IRGVT Y KKEQAINYLKD DY  LPLIRANNIQN GK  FDTTDLV
(IA) StySPI      AA*C                  MNRGKLPEGWATAPVSTVTT L IRGVT Y KKEQALNYLQD DY  LPIIRANNIQN GK  FDTTDLV
(IC) EcoR124I   GAA*             MSEMSYLEKLLDGVEVEWLP       L GEITK Y EQPTKYLVKAK DY  HDTYTIPVLTA GK  TFILGYT

(IA) StySBI      GA*G                  MSGGKLPEGWATSTINEMCN L         N PKLKL  D DDLDVGFMPMAGVPTTYLGKCNFETK
(IB) EcoAI       GA*G            (95)  EKPFELPDGWEWTT       L TRIAEI N PKIDVS D DEQEISFIPMPLISTKFDGSHEFEIK
(IB) EcoEI       GA*G            (95)  EKPFELPEGWEWIT       L SEIATI N PKIEVT D DEQEISFVPMPCISTRFDGAHDQEIK
(IA) EcoBI      TGA*                     MSFNSTSKE          L IEQN I N GLLSIH D SWLPISMDSVANITNGFAFKSSEFNN

(IB) CfrAI      GCA*             (95)  DKPFELPAGWEWVR       L GEAFYIEM G QSPSSQYYNQSEEGIPFFQGKADFGKKYPT
(IC) EcoprrI    CCA*             MSELSYLEKLMDGVEVEWLP       L SKVFNLRN G YTPSKTKKEFWANGDIPWFRMDDIRENGRI
(IC) EcoDXXI    TCA*             MSELSYLEKLLDGVEVEWVT       L GSMADIGT G SSNRQDESENGIYPFYVRSKNILKSDTFEF

(IA) EcoDI      TTA*                   MSAGKLPVDWKTVELGELIK L STGKLDANAADNDGQYPFFTCAESVSQINSWAFDTSAVL
```

B

```
Type I        Alignment of amino
R-M           acids at start of
system        S subunit CRDs

(IB) A           (389)LPEGWEW
(IB) E           (389)LPIGWEW
(IB) Cfr         (380)LPEGWEW
(IA) K           (208)LTEKWRN
(IA) D           (179)LTKEWRS
(IA) B           (209)LTKEDKD
                 (230)IPEDWSE
(IA) SP          (208)LTTALRN
                 (217)SLIGWHS
(IA) SB          (204)LTKELHK
                 (223)IPSLWKI
(IC) R124        (200)KEGEVEW
(IC) DXXI        (196)KEGEVEW
(IC) prr         (216)KEDEVEG
```

Figure 5. (**A**) A comparison of the first 60 amino acids of the ARDs of type I R–M systems. The family to which each belongs is shown in brackets. Each group comprises ARDs which share a common dinucleotide including the methylation site (A*). The N terminal 95 amino acids of the IB family S subunits which are not part of the ARD are omitted. The amino acids in bold type, up to and including the conserved leucine (Leu21 in *Eco*KI), are repeated at the start of the CRD, and are proposed to have a structural role. The underlined bold amino acids indicate conservations within each dinucleotide recognition group, including the Tyr27 crosslinking site in *Eco*KI, which may be involved in recognizing the nucleotide preceding the methylation target. Other amino acids conserved within a group are underlined. (**B**) This shows amino acids at the start of the CRDs that line up with the amino acids at the start of the ARD preceding the conserved leucine (Leu21 in *Eco*KI). Two alignments are shown for *Eco*BI, *Sty*SPI and *Sty*SBI, the first in each case is as previously published (23), while the second is an alternative which is in better agreement with the *Eco*AI consensus, LPEGWEW. The number of amino acids preceding the shown sequence is given in brackets.

From our results we conclude that the methyl group of thymine 7′ is in close proximity to Tyr27 in the *Eco*KI S subunit ARD. Substitution of the methyl group with the similarly sized bromine has no effect on DNA binding and increases the crosslinking yield, but the substitution with the smaller hydrogen in the U7′ duplex or the larger iodine in the Io7′ duplex reduces both binding affinity and crosslinking efficiency. In the case of the bromodeoxyuridine-substituted oligonucleotides, crosslinking only occurred between bromine at position 7′ in the duplex and Tyr27, possibly because there are no suitable reactive amino acids near to the other substitution sites.

The amino acids crosslinked to DNA via BrdU in several DNA–protein complexes have been identified (38–40) and found to correspond to important amino acids in the complex structure. For example, for *Eco*RI, methionine 137 identified at the site of crosslinking is in a region of the molecule previously implicated in DNA binding and cleavage (38) and the subsequent co-crystal structure showed that amino acid residues Met137–Gly140 make important contacts to the two outer pyrimidines of the recognition site (28,31). In the *lac* repressor protein, amino acids that crosslinked to operator DNA substituted with BrdU are consistent with NMR evidence and molecular dynamic calculations (39), while for the basic region leucine zipper (bZIP) DNA-binding motif of GCN4, a yeast transcriptional activator, the demonstration that Ala238 crosslinked to a BrdU in its binding site (40), is in agreement with the crystal structure that showed contact of Ala238 with the corresponding thymine 5-methyl group in the GCN4 bZIP element–DNA complex (61).

The demonstration that Tyr27 in the *Eco*KI S subunit ARD is in close proximity to the DNA sequence recognition site is consistent with genetic work that correlated the ARD with the recognition of the trinucleotide part of the DNA target (17). The result also gives evidence for a contact with thymine in the major groove of the DNA, a conclusion consistent with methylation interference experiments, indicating that the mtase makes major groove contacts to both halves of the bipartite recognition sequence (62).

The ARDs of different type I systems show little similarity unless they recognise the same trinucleotide (16–18). However, there is some similarity for the first part of the ARDs from all of the families, with a motif that is repeated at the beginning of the C-terminal conserved region (Fig. 5; 24) suggesting a structural role for the amino acids shown in bold prior to Leu21 of *Eco*KI (Fig. 5A). After Leu21, the sequence comparisons of the first portions of ARDs that have a common dinucleotide sequence hint that some amino acids (underlined and bold in Fig. 5A) may be conserved to allow recognition of this dinucleotide part of the DNA target. In particular, it is possible that Tyr27 in the S polypeptide of the *Eco*R124I system plays the same role in contacting the AA dinucleotide sequence as it does in the *Eco*KI system. If this hypothesis could be confirmed by, for example, crosslinking experiments with *Eco*R124I, and mutagenesis of Tyr27 in *Eco*KI and *Eco*R124I, then it might suggest that the target recognition domains of the S subunits may have similar tertiary structure features for recognizing parts of their DNA targets.

ACKNOWLEDGEMENTS

We thank Prof. J. Fothergill and Mr B. Dunbar, University of Aberdeen and Welmet Protein Characterisation Facility, University of Edinburgh, for the peptide sequencing. We thank the OSWEL DNA service for assistance with oligonucleotide synthesis, Mr L. P. Cooper for his assistance with the purification of the protein and Mrs Isobel Black for her help in the preparation of the manuscript. This work was supported by the Medical Research Council and the Science and Engineering Research Council. TB wishes to thank the Royal Society of Edinburgh for a Caledonian Research Fellowship. This is a publication from the Edinburgh Centre for Molecular Recognition.

REFERENCES

1 Sain, B. and Murray, N.E. (1980). *Mol. Gen. Genet.* **180**, 35–46.
2 Bickle, T.A. (1987). In Neidhart, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M. and Umbarger, H.E. (eds), *E. coli and S. typhimurium*. American Society for Microbiology, Washington, DC, pp. 692–697.
3 Wilson, G.G. and Murray, N.E. (1991). *Annu. Rev. Genet.* **25**, 585–627.
4 Bickle, T.A. and Krüger, D.H. (1993). *Microbiol. Rev.* **57**, 434–450.
5 Suri, B., Nagaraja, V. and Bickle, T.A. (1984). *Curr. Top. Microbiol. Immunol.* **108**, 1–9.
6 Dryden, D.T.F., Cooper, L.P. and Murray, N.E. (1993). *J. Biol. Chem.* **268**, 13228–13236.
7 Kan, N.C., Lautenberger, J.A., Edgell, M.H. and Hutchison, C.A. (1979). *J. Mol. Biol.* **130**, 191–209.
8 Bickle, T.A., Brock, C. and Yuan, R. (1978). *Proc. Natl. Acad. Sci. USA* **75**, 3099–3103.
9 Bickle, T.A. (1982). In Linn, S.M. and Roberts, R.J. (eds), *Nucleases*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 85–108.
10 Vovis, G.F., Horiuchi, K. and Zinder, N.D. (1973). *Proc. Natl. Acad. Sci. USA* **71**, 3810–3813.
11 Burckhardt, J., Weisemann, J. and Yuan, R. (1981). *J. Biol. Chem.* **256**, 4024–4032.
12 Meselson, M. and Yuan, R. (1968). *Nature* **217**, 1111–1114.
13 Boyer, H.W. and Roulland-Dussoix, D. (1969). *J. Mol. Biol.* **41**, 459–472.
14 Glover, S.W. and Colson, C. (1969). *Genet. Res.* **13**, 227–240.
15 Fuller-Pace, F.V., Cowan, G. and Murray, N.E. (1985). *J. Mol. Biol.* **186**, 65–75.
16 Cowan, G.M., Gann, A.A.F. and Murray, N.E. (1989). *Cell* **56**, 103–109.
17 Fuller-Pace, F.V. and Murray, N.E. (1986). *Proc. Natl. Acad. Sci. USA* **83**, 9368–9372.
18 Kannan, P., Cowan, G.M., Daniel, A.S., Gann, A.A.F. and Murray, N.E. (1989). *J. Mol. Biol.* **209**, 335–344.
19 Gubler, M., Braguglia, D., Meyer, J., Piekarowicz, A. and Bickle, T.A. (1992). *EMBO J.* **11**, 233–240.
20 Nagaraja, V., Shepherd, J.C.W., Pripfl, T. and Bickle, T.A. (1985). *Nature* **316**, 371–372.
21 Gough, J. A. and Murray, N.E. (1983). *J. Mol. Biol.* **166**, 1–19.
22 Fuller-Pace, F.V., Bullas, L.R., Delius, H. and Murray, N.E. (1984). *Proc. Natl. Acad. Sci. USA* **81**, 6095–6099.
23 Gann, A.A.F., Campbell. A.J.B., Collins, J.F., Coulson, A.F.W. and Murray, N.E. (1987). *Mol. Microbiol.* **1**, 13–22.
24 Kneale, G.G. (1994) *J. Mol. Biol.* **243**, 1–5.
25 Meister, J., MacWilliams, M., Hübner, P., Jutte, H., Skrzypek, E., Piekarowicz, A. and Bickle, T.A. (1993). *EMBO J.* **12**, 4585–4591.
26 Abadjieva, A., Webb, M., Patel, J., Zinkevich, V. and Firman, K. (1994). *J. Mol. Biol.* **241**, 35–43.
27 Cooper, L.P. and Dryden, D.T.F. (1994). *J. Mol. Biol.* **236**, 1011–1021.
28 Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1994). *Nature* **368**, 660–664.
29 Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1994). *Structure* **2**, 439–452.
30 Athanasiadis, A., Vlasi, M., Kotsifaki, D., Tucker, P.A., Wilson, K.S. and Kokkinidis, M. (1994). *Nature Struct. Biol.* **1**, 469–475.
31 Kim, Y., Grable, J.C., Love, R., Greene, P. and Rosenberg, J.M. (1990) *Science* **249**, 1307–1309.
32 Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K. and Wilson, K.S. (1993). *EMBO J.* **12**, 1781–1795.
33 Cheng, X., Balendiran, K., Schildkraut, I. and Anderson, J.E. (1994). *EMBO J.* **13**, 3927–3935.

34 Klimasauskas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994). *Cell* **76**, 357–369.
35 Labahn, J., Granzin, J., Schluckebier, G., Robinson, D.P., Jack, W.E., Schildkraut, I. and Saenger, W. (1994). *Proc. Natl. Acad. Sci. USA* **91**, 10957–10961.
36 Lin, S-Y. and Riggs, A.D. (1974). *Proc. Natl. Acad. Sci. USA* **71**, 947–951.
37 Ogata, R. and Gilbert, W. (1977). *Proc. Natl. Acad. Sci. USA* **74**, 4973–4976.
38 Wolfes, H., Fliess, A., Winkler, F. and Pingoud, A. (1986). *Eur. J. Biochem.* **159**, 267–273.
39 Allen, T.D., Wick, K.L. and Matthews, K.S. (1991). *J. Biol. Chem.* **266**, 6113–6119.
40 Blatter, E.E., Ebright, Y.W. and Ebright, R.H. (1992). *Nature* **359**, 650–652.
41 Ahmad, I. and Rao, D.N. (1994). *J. Mol. Biol.* **242**, 378–388.
42 Willis, M.C., LeCuyer, K.A., Meisenheimer, K.M., Uhlenbeck, O.C. and Koch, T.H. (1994). *Nucleic Acids Res.* **22**, 4947–4952.
43 Beaucage, L. and Caruthers, M.H. (1981). *Tetrahedra Lett.* **22**, 1859–1862.
44 Hutchinson, F. (1973). *Ann. Rev. Biophys.* **6**, 201–246.
45 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
46 Powell, L.M., Dryden, D.T.F., Willock, D.F., Pain, R.H. and Murray, N.E. (1993). *J. Mol. Biol.* **234**, 60–71.
47 Laemmli, U.K. (1970). *Nature* **227**, 680–685.
48 Papavassiliou, A.G. (1994). In Kneale, G.G. (ed), *DNA–Protein Interactions: Principles and Protocols. Methods in Molecular Biology*, Vol. 30, Humana Press, New Jersey, pp. 43–78.
49 Erlanger, B.F., Cooper, A.G. and Cohen, W. (1966). *Biochem. J.* **5**, 190–196.
50 Hayes, J.D., Kerr, L.A. and Cronshaw, A.D. (1989). *Biochem. J.* **264**, 437–445.
51 Taylor, I., Watts, D. and Kneale, G. (1993). *Nucleic Acids Res.* **21**, 4929–4935.
52 Williams, K.R. and Konigsberg, W.H. (1991). *Methods Enzymol.* **208**, 516–539.
53 Paradiso, P.R., Nakashima, Y. and Konigsberg, W. (1979). *J. Biol. Chem.* **254**, 4739–4744.
54 Merrill, B.M., Williams, K.R., Chase, J.W. and Konigsberg, W.H. (1984). *J. Biol. Chem.* **259**, 10850–10856.
55 Cleghon, V. and Klessig, D.F. (1992). *J. Biol. Chem.* **267**, 17872–17881.
56 Evans, R.K., Johnson, J.D. and Haley, B.E. (1986). *Proc. Natl. Acad. Sci. USA* **83**, 5382–5386.
57 Meffert, R. and Dose, K. (1988). *FEBS Lett.* **239**, 190–194.
58 Meffert, R., Rathgeber, G., Schüfer, H.J. and Dose, K. (1990). *Nucleic Acids Res.* **18**, 6633–6636.
59 Nikiforov, T.T. and Connolly, B.A. (1992). *Nucleic Acids Res.* **20**, 1209–1214.
60 Welsh, J. and Cantor, C.R. (1985). *Trends Biochem. Sci.* **8**, 505–508.
61 Ellenberger, T.E., Brandl, C.J., Struhl, K. and Harrison, S.C. (1992). *Cell* **71**, 1223–1227.
62 Powell, L.M. and Murray, N.E. (1995) *Nucleic Acids Res.* **23**, 967–974.