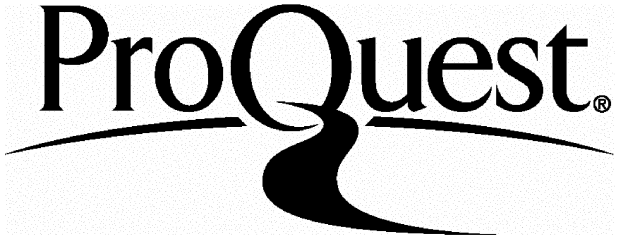
A Novel Thermostable Restriction Modification System

Sejal Patel

Submitted for the degree of Doctor of Philosophy (PhD)

University College London



ABSTRACT

A thermostable restriction modification system has been isolated from a thermophilic bacterium. The isolate designated Rot34A1 was thought to be the extreme thermophile *T. filiformis*. The strain grows optimally at 70°C and was isolated from a thermal site in Rotorua, New Zealand. The endonuclease named *Tfi*I, in accordance to the rules proposed by Smith and Nathans (1973), recognises the sequence $GA^{A}/_{T}TC$ which is a hitherto unknown specificity and has not been reported from either mesophilic or thermophilic sources.

The endonuclease was purified and characterised. The molecular weight (M_r) of *Tfi*I endonuclease, estimated under denaturing conditions was 37 000. The M_r of the native form of *Tfi*I endonuclease, as estimated by gel filtration was approximately 75 000 i.e., *Tfi*I endonuclease is a dimer in its active form.

The optimal pH for *Tfi*I endonuclease activity was determined to be pH 8.0 although *Tfi*I endonuclease exhibits activity over a broad pH range. The enzyme is remarkably thermostable, surviving at room temperature for several weeks and having a half life of greater than one hour at 65°C. Other characteristics of *Tfi*I endonuclease have been determined such as ability to cleave single stranded DNA and it's salt requirement.

Two kinds of "star activity" were observed for *Tfi*I endonuclease; the indiscriminate endonucleolytic activity exhibited in certain buffers and the relaxed specificity exhibited by buffers containing Mn^{2+}.

The *Tfi*I methylase was partially purified and a single step method was developed to separate the methylase from the endonuclease. The site at which the *Tfi*I methylase incorporates methyl groups into DNA was determined.

Genomic libraries were created in both plasmid and phage vectors and strategies for screening the libraries are discussed.

ACKNOWLEDGEMENTS

I am extremely grateful to John Ward and Don Cowan for their supervision throughout this project. Their guidance and support have been invaluable.

Thanks to Dr Hsuan for proving the internal sequencing data.

To Siân Shore (my mentor) and Carol French thanks for all the advice over the past few years and thanks also Paul Shadbolt for all his help.

Thanks to Neil Broadway and Mike Orford for making FPLC so much more fun than System Gold and for introducing me to Huntley Street.

Apologies to all on the Ground Floor (past and present) for not mentioning you all individually -thanks for having to put up with the mood swings and the mess.

A big thanks to Majib ur Rehman for all his support and special thanks to Ian Collins (with sincerity!) and Richard Braley (for the proof reading and encouragement).

Finally thank you to all my family and friends whose constant support and food kept me going!

TABLE OF CONTENTS

TABLE OF CONTENTS

List of Figures

TABLE OF CONTENTS

TABLE OF CONTENTS

List of Tables

Abbrevations

A	adenosine
AdoMet	S-adenosyl-L- methionine
Amp	ampicillin
ATP	adenosine triphosphate
βgal	β-galactosidase
bp	base pair
BSA	bovine serum albumin
C	cytosine
C·	controlling element
cpm	counts per minutes
ds	double stranded
Dam	DNA-adenine methyltransferase
Dcm	DNA-cytosine methyltransferase
DTT	dithiothreitol
EDTA	ethylene diamine tetraacetic acid
ENase	restriction endonuclease
EtBr	ethidium bromide
G	guanine
kb	kilobase
M· (MTase)	methylase
NAD+	nicotinamide adenine dinucleotide (oxidised form)
NADH	nicotinamide adenine dinucleotide (reduced form)
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis

PCR	polymerase chain reaction
PGA	3-phosphoglyceric acid
PGK	3-phosphoglyceric phosphokinase
R·	restriction endonuclease
RM	restriction modification system
SAM	S-adenosyl-L- [methyl-^{3}H] methionine
SDS	sodium dodecyl sulphate
ss	single stranded
T	thymine
Tc	tetracyline
TCA	Trichloroacetic acid
T_{opt}	Temperature optimum
TVR	terminal variable region
U	uracil
v/v	volume per volume
w/v	weight per volume
w/w	weight per weight
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
ve	elution volume
vo	void volume

Abbreviation of amino acids

A	(Ala)	Alanine
C	(Cys)	Cysteine
D	(Asp)	Aspartate
E	(Glu)	Glutamic acid
F	(Phe)	Phenylalanine
G	(Gly)	Glycine
H	(His)	Histidine
I	(Ile)	Isoleucine
K	(Lys)	Lysine
L	(Leu)	Leucine
M	(Met)	Methionine
N	(Asn)	Asparagine
P	(Pro)	Proline
Q	(Gln)	Glutamine
R	(Arg)	Arginine
S	(Ser)	Serine
T	(Thr)	Threonine
V	(Val)	Valine
W	(Trp)	Tryptophan
Y	(Tyr)	Tyrosine

1 Introduction

1.1 RESTRICTION AND MODIFICATION

Present day DNA technology is largely dependent upon the ability to cut DNA at specific sites with restriction endonucleases. Prior to 1970 there was no method available for cutting DNA into discrete fragments. A solution to this problem eventually grew from research into the phenomenon of host controlled restriction and modification. Arber and Linn (1968) demonstrated host controlled restriction and modification. The first endonuclease was isolated by Messelson and Yuan (1968).

Restriction describes the ability to reduce the biological activity of infecting DNA by enzymatic cleavage. It is a hydrolysis reaction in which phosphodiester bonds within the DNA backbone are cleaved by the restriction endonuclease. The restricting host also possesses the ability to protect it's own DNA from restriction, by modification. Modification is the result of DNA methylation. In bacteria, which have a restriction modification system, the modification enzyme methylates bases within the restriction endonuclease target site. Methylation of certain bases in the host DNA alters the endonuclease recognition sites and serves to protect the organism's own DNA from being digested; i.e., the organism is immune. Hence a restriction system always consists of two enzymatic elements, namely a nuclease and a methylase.

1.2 NOMENCLATURE

Smith and Nathans (1973) proposed a system of nomenclature for restriction endonucleases and modification methyltransferases which has been generally accepted. Under these rules the first three letters reflect the genus and the species from which the restriction enzyme has been isolated, followed by the Roman numerals which show the

number of different enzymes isolated from the same strain (For example *Hind* III means the endonuclease was isolated from the strain *Haemophilus influenzae* strain d and was the third endonuclease to be isolated from this organism). To differentiate between the restriction endonuclease and the modification methyltransferase it is generally accepted that the methyltransferase is distinguished by the capital letter M· with a raised dot in front of the enzyme name. Recently the endonucleases and methylases are referred to collectively as restriction modification systems (RM). The endonucleases are referred to as the --- ENases and the methyltransferases as the ---MTases (for example the *Eco*RI system would be referred to as the *Eco*RI ENase and the *Eco*RI MTase (Kessler and Manta, 1990).

1.3 CLASSIFICATION OF RESTRICTION MODIFICATION SYSTEMS

Based on their subunit structures, reaction mechanisms and genetic organisation, restriction enzymes fall into three general classes, namely I, II and III (Table 1.1 shows the different characteristics of the classes).

Class I enzymes exhibit both restriction and DNA modification activities located on different subunits of multifunctional enzyme complexes. They require magnesium ions, ATP and S-adenosyl methionine (AdoMet) as cofactors. These enzymes cleave DNA at unspecified sites usually 100 to 1000 base pairs downstream of the specific recognition sequences.

Class III enzymes like Class I also combine restriction and DNA modification activities on single enzymatic complexes composed of different subunits. However these enzymes lack both the ATPase activity of the class I enzymes and their absolute requirement of S-adenosyl methionine. Class III enzymes recognise specific sequences and cleave DNA at fixed sites 25 to 27 base pairs distal to their recognition sequence.

Table 1.1 Characteristics of classes of restriction modification systems.

[The different properties of the classes of restriction modification systems are shown below. N= any nucleotide, bp = base pair]

	CLASS I	CLASS II	CLASS III
Restriction modification activities	single multifunctional enzyme	separate endonuclease & methylase	separate enzymes with a subunit in common
Protein structure of endonuclease	3 different subunits	Homodimers	2 different subunits
Requirement for restriction	ATP, Mg^{2+}, AdoMet	Mg^{2+}	ATP, Mg^{2+}, AdoMet
Sequence of host specificity sites	non rotational symmetry	rotational symmetry	non rotational symmetry
	e.g. *Eco*B	e.g. *Eco*RI	e.g. *Eco*PI
	T G A N_2 T C G C	**G A A T T C**	**A G A C C**
Cleavage Site	Possibly random at least 1000 bp from host specificity site	At or near host specificity site	24-26 bp 3' of host specificity site

All class II systems are comprised of two different proteins, i.e., the restriction enzymes are separate from the methylases, which are relatively small compared to their Class I and III counterparts and their cofactor requirements are simple. *In vitro* restriction requires only magnesium ions and unmodified DNA while the methylase is dependent only on the presence of S-adenosyl methionine and unmodified or partially modified DNA.

It is the class II enzymes which have received the most intense attention over the years as they are capable of reproducibly cutting DNA in a fixed position with respect to the sequence they recognise.

1.4 CLASS II RESTRICTION ENDONUCLEASES

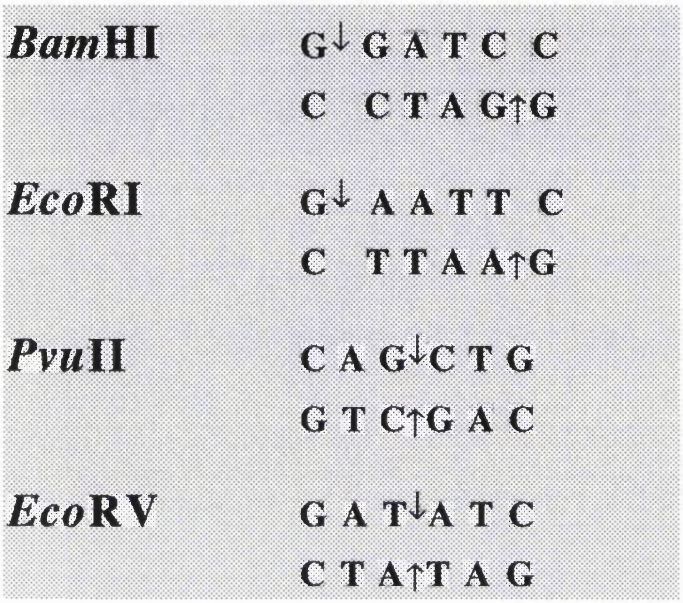
Endonucleases cleave phosphodiester bonds at defined positions in or adjacent to the recognition site, resulting in 5' or 3' self-complementary terminal extensions (e.g., *Bam*HI and *Eco*RI) or fully base paired termini [e.g., *Eco*RV and *Pvu*II, (Figure 1.1) see Kessler and Manta (1990) for review].

The recognition sequences usually exhibit two fold symmetry with the cleavage sites in each strand arranged around the axis of symmetry. Some sequences are degenerate (e.g., *Hha*II where the recognition site is GANTC where N is variable).

Isoschizomers are restriction endonucleases which have been isolated from different microorganisms, but are characterised by identical sequence specificity and identical cut positions. Examples of isoschizomers are *Nde*II and *Sau*3AI both cutting at identical cut positions within the identical recognition sequence 5'-↓GATC-3' (arrow indicates cut site).

Figure 1.1 DNA recognition sequences of *Eco*RI, *Eco*RV, *Bam*HI and *Pvu*II

[The site of cleavage within the recognition sequence is indicated by the arrows.]



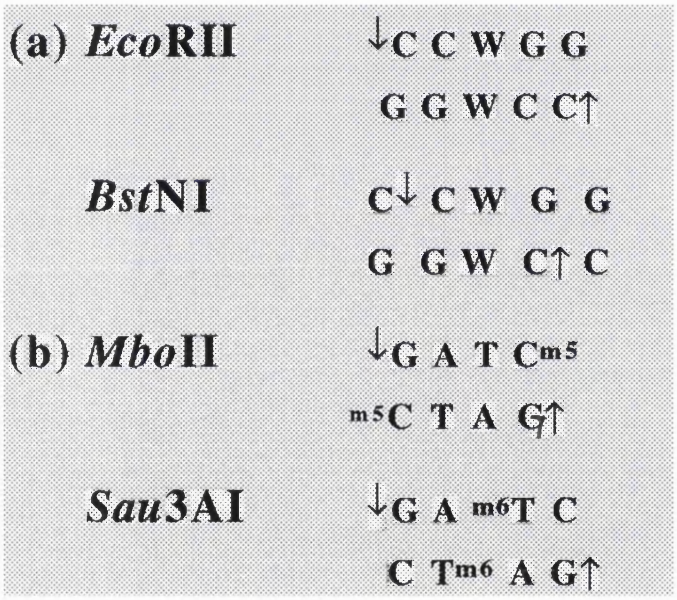
Class II endonucleases usually act as homodimers; i.e., interact with their substrate as a dimer (Modrich, 1982). The endonucleases can be a range of sizes from 157 amino acids (*Pvu*II) to 576 amino acids (*Bsu*RI) [Wilson and Murray, 1991]. Some isoschizomers can cleave DNA at different positions within the same recognition sequence. Where the cleavage sites are identical, isoschizomers can have different sensitivity to methylation (Figure 1.2).

Canonical site specific methylation refers to the methylation exhibited by the methylase of a specific restriction modification system. This methylation always inhibits DNA cleavage by it's corresponding (or cognate) restriction endonuclease. For example M·*Bam*HI methylase modifies GGAT^{m4}CC, and *Bam*HI endonuclease cannot cut this methylated sequence. However non canonical site methylation will inhibit the rate of DNA cleavage. In about half of the cases tested, (McClelland *et al.*, 1994) methylation at non canonical sites caused at least ten-fold inhibition. In other cases non canonical methylation has no

effect on restriction cleavage. For example *Bam*HI cuts DNA which has been modified at GGATC^{m4}C or GGATC^{m5}C but cannot cut DNA methylated at GGAT^{m5}CC.

Figure 1.2 Class II isoschizomers

[Two examples of isoschizomers, (a) isoschizomers with the same recognition sequence but different cleavage sites, (b) isoschizomers with identical cleavage sites but different methylation sites (W is either A or T).]



Endonucleases can show preferential cleavage; i.e., within a given DNA molecule some recognition sites are cleaved more easily than other sites. This phenomenon is partly explained by the influence of adjacent sequences. *Pvu*II is inhibited by methylation outside the *Pvu*II recognition site. Experiments showed that with methylated DNA containing two cleavage sites for *Pvu*II, one site being digested four to eight fold more slowly than the other. With unmethylated DNA, the two sites were cleaved at the same rate in this system (Liu *et al.*, 1992).

As more and more restriction enzymes are being isolated, the classification system definitions have been found to be insufficient. Class IIS enzymes (ENases-IIS) are a subclass of Class II restriction modification systems. The enzymes interact with two discrete sites: the recognition site, which is 4-7 bps long, and the cleavage site usually 1-20 bp away from the recognition site. A total of 35 ENase-IIS have been isolated (80 if isoschizomers are included).The recognition sequences are totally asymmetric and all of the characterised ENase-IIS are monomeric (Bellemare and Potvin, 1990; Szybaski *et al*., 1988; Degtyarev *et al.,* 1990).The Class IIS endonucleases are normally twice the size of Class II endonucleases and appear to act as monomers (Wilson and Murray, 1991).

The *Bcg*I endonuclease isolated from *Bacillus coagulans* may be a member of yet another subgroup of restriction enzymes. This restriction endonuclease differs from the other groups as summarised in Table 1.2 (Kong *et al.,* 1993).

Class II restriction endonucleases require Mg^{2+} as a cofactor for cleavage activity. Other divalent metals replace Mg^{2+} but with reduced specificity (Hsu and Berg, 1978). This relaxation of activity is termed as "star activity" (Polisky *et al.*, 1975). *Bam*HI is among the endonucleases that shows a similar relaxation of specificity in buffers containing Mn^{2+} and also in high concentrations of compounds such as glycerol, ethylene glycol, ethanol or dioxane. It is assumed that this is due to altered DNA-protein interactions (Jack *et al.*, 1991).

Although some restriction endonucleases cleave single stranded DNA specifically, most are unable to or do so with low activity and specificity.

Class II endonucleases which are not accompanied by a class II methylase have alternative methods of protection their DNA from restriction, e.g. no endonuclease sites within the host DNA (Herrero *et al.*, 1984).

Table 1.2 Comparison between *Bcg*I and other classes of restriction endonuclease [Table reproduced from Kong *et al.,* (1993) ds - double stranded].

Restriction endonuclease	Cleavage factors	Recognition sequence	Cleavage characteristics
Class I e.g., *Eco*K	Mg^{2+},ATP, AdoMet	Asymmetric; interrupted; bipartite e.g., CTGAAG	Non specific; ds, on both sides of recognition sequence
Class II e.g., *Bam*HI	Mg^{2+}	Symmetric; continuous or interrupted e.g., GGATCC	Specific; ds, symmetric within recognition sequence
Class IIs e.g., *Fok*I	Mg^{2+}	Asymmetric; e.g., GGATG	Specific; ds, asymmetric on one side of recognition sequence
Class III e.g., *Ecop*I	Mg^{2+}, ATP (stimulated by AdoMet)	Asymmetric; continuous e.g., CTGAAG	Semi specific; ds, asymmetric on one side of recognition sequence
_Bcg_I	**Mg^{2+}, AdoMet**	**Asymmetric interrupted e.g., $CGAN_6TGC$**	**Specific; ds; symmetric, on both sides of recognition sequence**
*Eco*571, *Gsu*I, *Bsg*I	Mg^{2+} (stimulated by AdoMet)	Asymmetric; continuous e.g., CTGAAG	Specific; ds, asymmetric on one side of recognition sequence

*Dpn*I is the best characterised of unaccompanied restriction endonucleases, cleaves the sequence G^{m6}ATC. However this sequence does not occur in the hosts genome as this bacterium does not have specific cytosine methylation (Lacks and Greenberg, 1977).

1.5 CLASS II MODIFICATION METHYLASES

DNA methylases transfer methyl groups from the donor S-adenosylmethioine onto the adenine or cytosine residues within a specific DNA sequence. Most class II methylases recognise double stranded DNA (which exhibits two fold rotational symmetry) and methylate a single residue on both strands of the recognition sequence. They produce one of three products, N^6-methyladenine, modifying the exocyclic amino group of adenine, 5-methylcytosine and N^4 methylcytosine, modifying the exocyclic group of cytosine. The latter type of modification was only discovered in the early 1980s (Janulaitis *et al.*, 1983; Ehrlich *et al.*, 1985, see Section 1.6.3).

No class II methylase has been reported to be less specific than its endonuclease counterpart (Wilson and Murray, 1991).

A number of methyltransferases occur separately and are involved in repair mechanisms, unaccompanied by endonucleases. *Escherichia coli* possesses two such unaccompanied methylases, DNA-adenine methyltransferase (*dam*) associated with methyl directed mismatch repair and DNA-cytosine methyltransferase (*dcm*) associated with short patch repair (Marinus, 1987).

Methylases which have been characterised to date appear to function as monomers: single polypeptide chains between 228 amino acids (M·*Alu*I) to 587 amino acids (M·*Hgi*DI) in length (Wilson and Murray, 1991). The exception is M·*Bst*1503I, where the active form is composed of four identical subunits, each of 105 000 kDa (Levy and Welker, 1981).

1.6 OCCURRENCE OF RESTRICTION MODIFICATION SYSTEMS

There were 1284 endonucleases and 130 methylases of known specificities isolated from 1117 different organisms by January 1990 (Kessler and Manta, 1990). One year later another 800 endonucleases had been characterised, including 179 different class-II specificities (Roberts and Macelis, 1992). Some specificities are common: isoschizomers of *Hae*III, *Eco*RII and *Pst*I, have been found 50 times over, while others are rare [*Xba*I has only been found once (Wilson and Murray, 1991)].

1.6.1 RESTRICTION MODIFICATION SYSTEMS IN NON BACTERIAL SYSTEMS

The vast majority of restriction enzymes come from bacteria, although a few non bacterial systems have been reported. Virulent viruses of the unicellular alga, *Chlorella* , carry restriction modification systems (Xia *et al.*, 1986; Zhang *et al.,* 1992). Other non bacterial systems which have site specific endonucleases include self splicing introns from yeast. Though not strictly "true" restriction enzymes, they are used *in vitro* for cutting DNA (Watabe *et al.,* 1981; Shibata *et al.,* 1984). *Hsa*I from man and an endonuclease from *Chlamydomonas reinhardtii* (Sklar *et al.*, 1986) have also been identified (see Kessler and Höltke, 1986, for review and primary sources). There are endonucleases encoded by group I introns in both eukaryotes and bacteriophages which cleave intronless alleles.

1.6.2 RESTRICTION MODIFICATION SYSTEMS IN BACTERIAL SYSTEMS

Over 10 000 bacterial strains have been examined and sequence specific endonucleases have been found in roughly one quarter (Wilson and Murray, 1991); the remaining three quarters may lack restriction modification systems, or possess them in forms that have

eluded detection. Among the bacteria that do possess restriction modification systems, approximately half have multiple systems: usually two or three, but sometimes more (Laue *et al.*, 1991).

1.6.2.1 RESTRICTION MODIFICATION SYSTEMS IN THERMOPHILES

Restriction modification systems have been isolated from many thermophilic sources (e.g., Shaw *et al.*, 1991; Zebela *et al.* 1990; Slatko *et al.*, 1987; Kiss *et al.*, 1985). The thermophilic restriction modification systems are mostly thermostable isoschizomers of mesophilic counterparts such as the *Bst*LVI restriction endonuclease isolated from *Bacillus stearothermophilus* LV, an isoschizomer of *Cla*I (Lobos and Vásquez, 1993) although others cleave at novel sites; e.g., *Tfi*I (this study). Within the Archaea, various plasmids have been found to encode restriction modification systems including the *Methanobacterium thermoformicicum* strain (Nölling *et al.*, 1992). The hyperthermophile *Desulfurococcus mobiis* (Dalgaard *et al.*, 1993) has been found to have a site specific endonuclease encoded by an archaeal intron. This enzyme, I-*Dmo* I is unusual among the intron endonucleases in that it is thermostable and is expressed only from linear and cyclised intron species and not from its precursor RNA.

^{m5}C has been identified as a "hot spot" for mutation (Cowlondre *et al.*, 1978). At elevated temperatures ^{m5}C is deaminated to thymine and this will result in an A-T pairing instead of a G-C pairing. When Ehrlich *et al.* (1985) identified N^4 methylcytosine in half the thermophiles screened, they suggested an evolutionary origin of m^4 cytosine residues in thermophilic bacteria to prevent this mismatch (N^4 methylcytosine is deaminated to uracil which is replaced by cytosine via the uridine specific DNA repair pathway, hence preventing mutation). However on improved screening the same research group have shown N^4 methylcytosine in mesophiles.

1.7 PROTEIN STRUCTURE - AMINO ACID HOMOLOGY

1.7.1 AMINO ACID HOMOLOGY BETWEEN RESTRICTION ENDONUCLEASES

With the exception of some isoschizomers, class II endonucleases exhibit no common primary sequence homology. Isoschizomers that do show amino acid sequence homology include the *Eco*RI and *Rsr*I endonucleases (Stephenson *et al.,* 1989), *Bsu*BI and *Pst*I endonucleases (Xu *et al.,* 1992), *Bsu*FI and *Msp*I (Kapfer *et al.*, 1991), *Tth*HB81 and *Taq*I (Barany *et al.,* 1992a), *Fnu*DI and *Ngo*PII, *Cfr*91 and *Xma*I (Wilson, 1991). Comparison of the isoschizomers *Eco*RI and *Rsr*I showed their amino acid sequence 50% identity within a 266 amino acid overlap between the deduced amino acid sequences of *Rsr*I and *Eco*RI. The sequence alignment showed six regions 8-17 amino acid in length with 75-100% sequence identity (Stephenson *et al.*, 1989). Recently the *Mun*I restriction modification system (Siksnys *et al.*, 1994) has been cloned. The central part of the recognition sequence, including the position of the scissile bond ($^{\downarrow}$AATT), are common for *Mun*I, *Eco*RI and *Rsr*I (Figure 1.3).

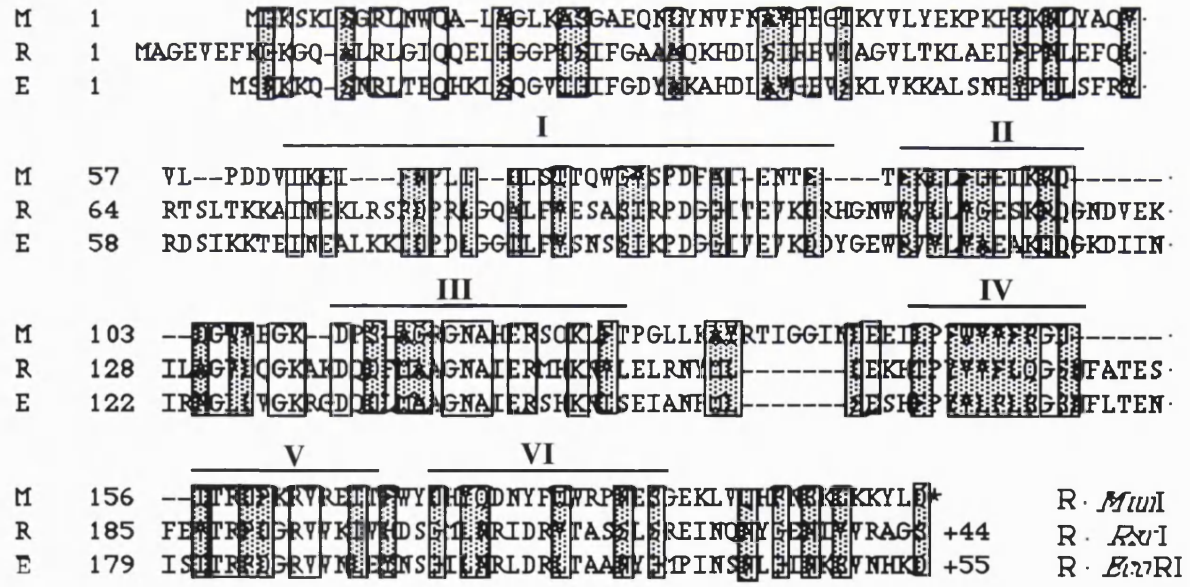
Comparisons between the R·*Mun*I deduced amino acid sequence, *Eco*RI and *Rsr*I revealed five of the six conserved regions were present in R·*Mun*I (Figure 1.4). These structural and functional significances of these conserved regions are discussed in greater detail in Section 1.8.2.

Figure 1.3 Recognition Sites of *Eco*RI, *Rsr*I and *Mun*I

***Eco*RI**	C$^{\downarrow}$A A T T C
(*Rsr*I)	G T T A A$^{\uparrow}$G
***Mun*I**	C$^{\downarrow}$A A T T G
	G T T A A $^{\uparrow}$C

Figure 1.4 Alignment of *Eco*RI, *Rsr*I and *Mun*I

[Columns with open boxes indicate amino acid residues that are identical between aligned proteins. Columns with grey background indicate similar amino acids (Taylor, 1986). Roman numerals I-IV above aligned sequence correspond to functionally and structurally important elements of the *Eco*RI and *Rsr*I enzymes, according to Stephenson *et al.* (1989). Numbers to the left of the alignment refer to the positions in the amino acid sequences of individual proteins. Reproduced from Siksnys *et al.* (1994).]



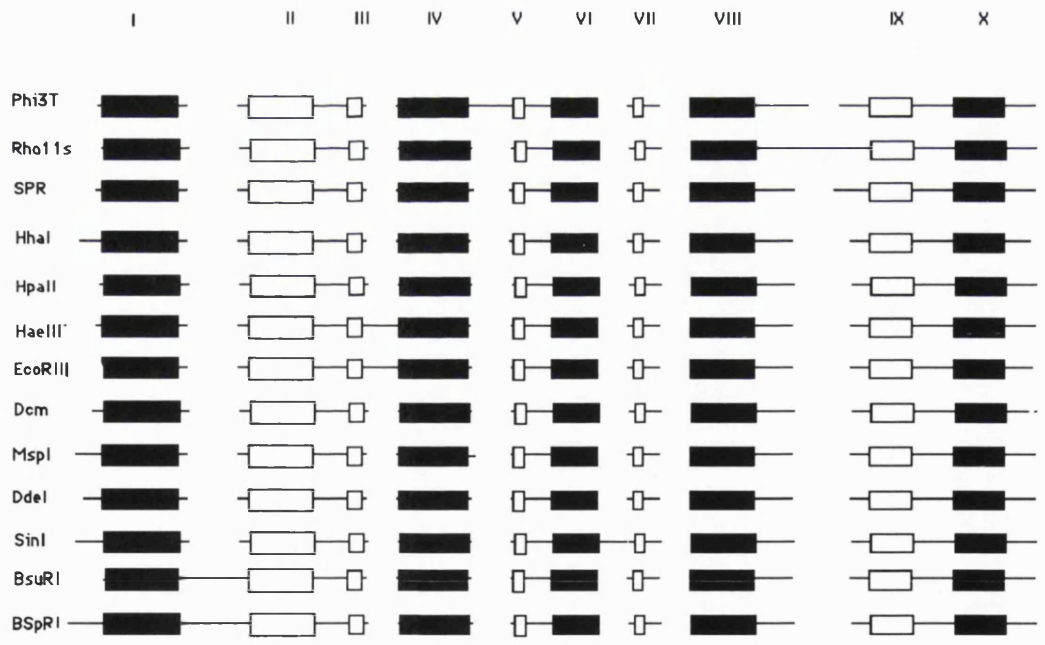
1.7.2 AMINO ACID HOMOLOGY BETWEEN METHYLASES

In contrast to the endonucleases there are extensive similarities among the primary amino acid sequence of methylases. Approximately ninety methylase genes have been sequenced and approximately seven architectural classes have been distinguished (Klimasauskas *et al.*, 1989).

One class comprises enzymes that form 5-methylcytosine in DNA (m^5C-MTase). Members of this group share at least ten common amino acid motifs (Figure 1.5). Towards the -COOH terminus of these enzymes is a "terminal variable region" (TVR) between conserved motifs VIII and IX. This is thought to form the sequence specific recognition domain (Pósfai *et al.*, 1989). The M·*Alu*I is the largest known methylase (predicted M_r 59 kDa) and also has the largest known variable region of any known monospecific DNA methylase, larger than even that of most multispecific methylases (Zhang *et al.*, 1993).

Figure 1.5 Schematic diagram of the alignment of MTase sequences

[Each line represents one sequence. Gaps were introduced in the alignment where the lines are interrupted. Boxes indicate where the ten blocks of conserved residues occur. Filled boxes indicate the five highly conserved blocks; the open boxes represents the less conserved blocks. The variable region lies between blocks VIII and IX. Reproduced from Pósfai *et al.* (1989).]



The remaining classes have the enzymes that form N4-methylcytosine (m^4C-MTases) and N6-methyladenine (m^6A-MTases). The m^4C-MTases and the m^6A-MTases share some similarity at primary amino acid level. The methylation of the exocyclic amino group of adenine and cytosine may be similar (Tao *et al.*, 1989), although the mechanism has yet to be elucidated. There are two common motifs shared by these two types of methylases, DPF-GSGT and TSPPY, which were found to be similar to those of the m^5C-MTases.

1.7.3 AMINO ACID HOMOLOGY BETWEEN RESTRICTION ENDONUCLEASES AND METHYLASES

Endonucleases and their cognate methylases share DNA sequence specificity but not primary amino acid sequence similarity. Comparisons of primary amino acid sequences of methylase specificity (target recognition) domains and their cognate endonucleases for nine different restriction modification systems showed no similarity (Chandrasegaran and Smith, 1988). One exception is the *Eco*RV restriction modification system in which a short motif similar to one found in several heterospecific endonucleases and methylases as well as in the cognate M·*Eco*RV has been identified in R·*Eco*RV (Thielking *et al.*, 1991). The lack of similarity between endonucleases and their cognate methylases suggests that restriction and modification enzymes are unrelated and that they recognise their targets by different strategies. It is hypothesised that the MTase and ENase genes may have evolved separately, only becoming adjacent to one another on the chromosome when their combined activity provided a selective advantage (Kaszubska *et al.*, 1989; Wilson and Murray, 1991).

Comparisons between primary sequences of R·*Eco*RII and the proposed sequence domains of three DNA-[cytosine-C5] methyltransferases identified a region containing

sequence motifs that were partially conserved (Kossykh *et al.*, 1993). The authors suggested that the motifs were important for sequence recognition by R·*Eco*RII and the isospecific methylases. Whether this evidence suggests that endonucleases and methylases arose from a common ancestor is yet to be determined. However *Eco*RII has been shown to require at least two recognition sites for endonucleolytic cleavage (Kroger and Hoborn, 1984). The enzyme must bind to an activator site before it will cut a cleavage site. *Eco*RII also has been found to have similarity between a 29 amino acid sequence in the carboxyl end of *Eco*RII and the motif defining the integrase family of recombinases (Topal and Conrad, 1993).

1.8 PROTEIN-NUCLEIC ACID INTERACTIONS

1.8.1 PROTEIN-NUCLEIC ACID INTERACTIONS OF ENDONUCLEASES

Structural studies of DNA binding proteins and their complexes with DNA have proceeded at an accelerating pace in recent years, due to technical advances in molecular genetics, DNA synthesis, protein crystallography and nuclear magnetic resonance.

Restriction endonucleases differ from other gene regulatory proteins in several ways. The DNA sequence recognised by the restriction endonuclease is usually centred at the dyad axis of the DNA, tends to be 4-8 base pairs in length and shows a strict dyad symmetry. In comparison the DNA sequence recognised by repressors extends over 14-20 base pairs in length and shows only partial dyad symmetry. In particular the sequence specificity exhibited by restriction endonucleases is far greater than exhibited by repressors (see Steitz, 1990 for review). A slight change in the methylation of a base within a restriction enzyme target site can result in complete inhibition of the activity of the restriction enzyme.

The elucidation of the crystal structures of a number of endonucleases (*Eco*RI, *Eco*RV, *Bam*HI and *Pvu*II) has been a significant advance in the understanding of the structural aspects of protein-nucleic acid interactions (Kim *et al.,* 1990; Winkler *et al.,* 1993; Newman *et al.,* 1994; Cheng *et al.,* 1994).

Though restriction endonucleases show no primary amino acid homology (except in certain examples of isoschizomers) they have been shown to have certain structural homologies. *Eco*RI, *Eco*RV, *Bam*HI and *Pvu*II have all been found to display an αβ architecture since all contain a 5-stranded β sheet surrounded on either side by α helices (Figures 1.6, 1.7, 1.8 and 1.9). The central core of the endonucleases consists of a mixed β-sheet, which carries the active site at one end, facing the reactive phosphate group. In their functional dimeric form, the endonucleases are U-shaped, with a prominent cleft for DNA binding (Aggarwal, 1995). *Eco*RI and *Bam*HI are structurally related and bind DNA from the major groove side whereas *Eco*RV and *Pvu*II, which are structurally homologous (Figures 1.7 and 1.8), approach DNA from the minor groove side.

The first structure to be determined with the cognate DNA was *Eco*RI endonuclease. McClarin *et al.* (1986) determined the crystal structure of the *Eco*RI endonuclease complexed with a 12 base pair DNA duplex including the enzyme recognition sequence.

In the absence of Mg^{2+} the enzyme will not hydrolyse the phosphodiester bond but binds the nucleotide. The crystal structure was determined at 3Å resolution using a single platinum heavy atom derivative, with no Mg^{2+} present. The structure of the complexed protein -metal ion -specific DNA has not been elucidated as this structure is unstable and the DNA is cleaved. Vipond and Halford (1995) have shown that Ca^{2+} ions appear to mimic the role of Mg^{2+}. Ca^{2+} gives a stable ternary complex in which the DNA-bound nuclease can not cleave the DNA.

Figure 1.6 Structures of *Bam*HI and *Eco*RI

[(a) *Bam*HI dimer (originally published by Newman *et al.*, 1994), (b) *Eco*RI- DNA complex. The proteins are represented as ribbons and the DNA as a stick model, using the program MOLSCRIPT. The twofold axis runs vertically in the plane of the paper. One of the subunits is shown shaded and with labels for the secondary structural elements. N and C mark the N and C termini of the subunits. Reproduced from Aggarwal (1995).]

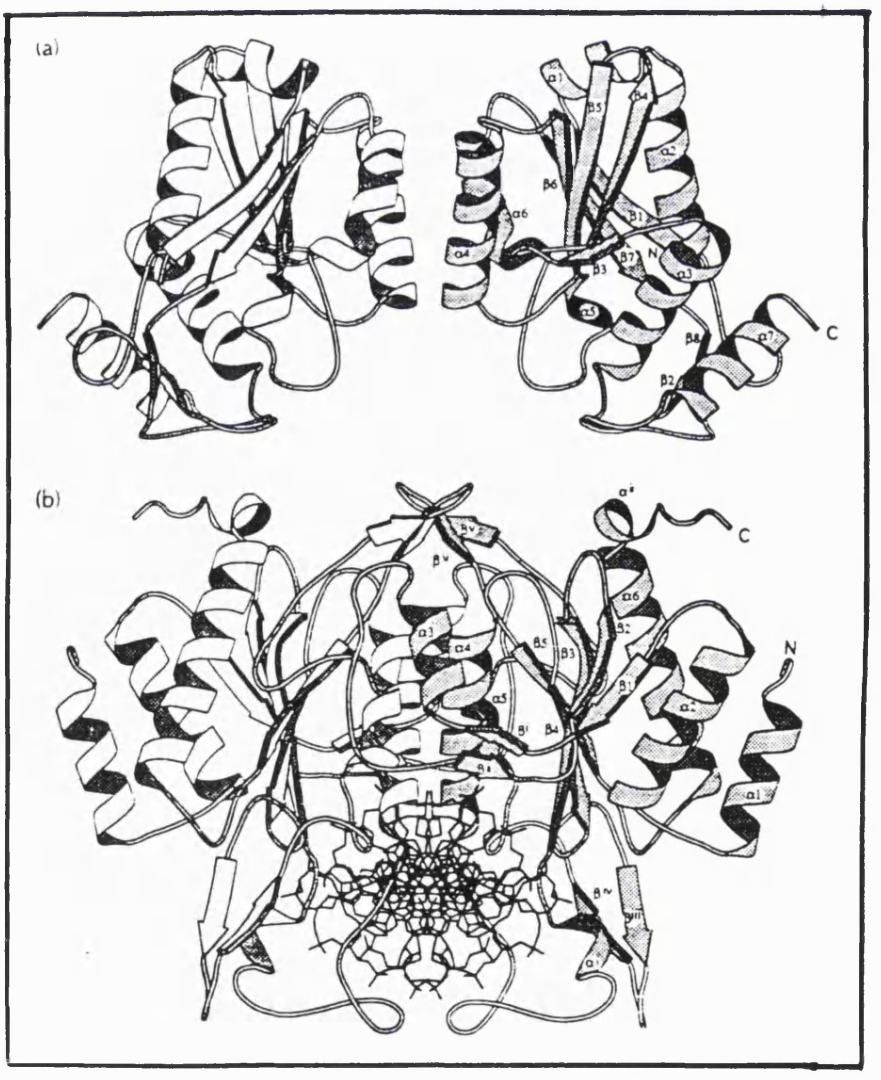
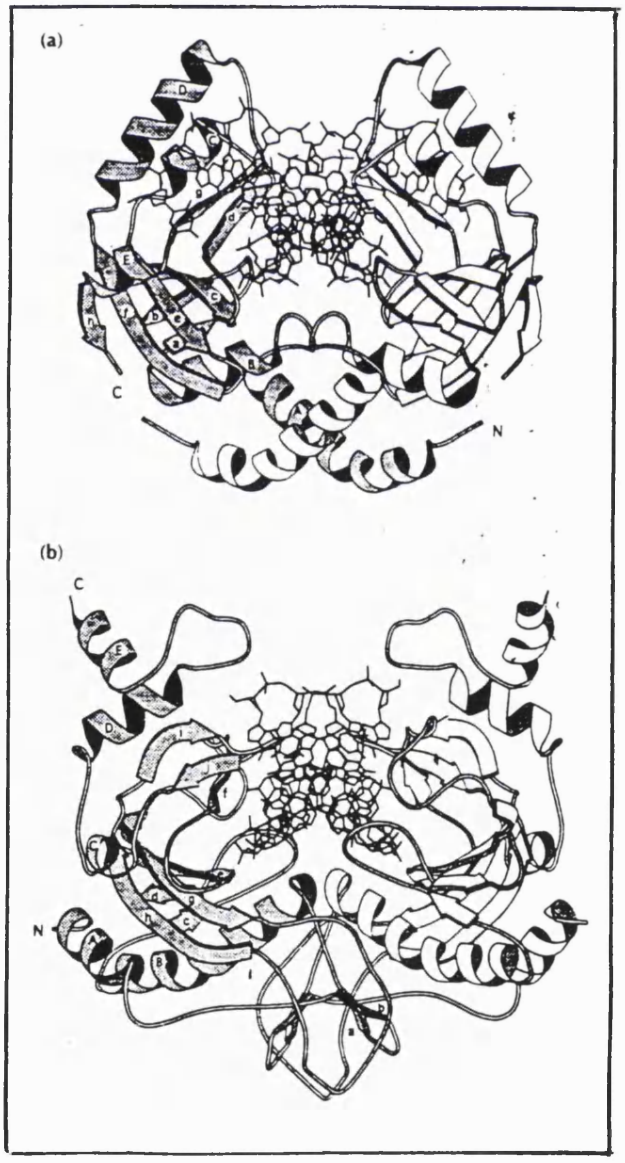


Figure 1.7 Structures of *Pvu*II and *Eco*RV

[Structures of (a) *Pvu*II-DNA (originally published by Cheng *et al.*, 1994) and (b) *Eco*RV-DNA (originally published by Winkler *et al.*, 1993) complexes. The proteins are represented as ribbons and the DNA as a stick models. One of the subunits is shown shaded and with labels for the secondary structural elements Reproduced from Aggarwal (1995).]



*Eco*RI endonuclease induces DNA kinking (McClarin *et al.*, 1986; Fredrick *et al.*, 1984). Two types of DNA "kinks" are produced by the binding of the DNA to *Eco*RI. The first is termed a "neoI" kink and occur at the central dyad axis, the second (the "neoII" kink) is observed 3 base pairs from the dyad axis on both sides of the dyad giving rise to a bend between the DNA axis on either side of the kink. This allows the DNA sequence to be in the correct orientation for cleavage to occur.

Twelve hydrogen bond donors and acceptors from protein side chains provide the sequence specificity exhibited by this enzyme. The side chains are complementary to the donors and acceptors presented by the exposed edge of the base pair in the recognition sequences. Two arginines (Arg^{145}, Arg^{200}) and one glutamic acid (Glu^{144}) each make two hydrogen bonds to the DNA bases in both subunits of the dimer. These three side chains are able to penetrate the deep major groove by emanating from the ends of the α helices that protrude from the protein into the DNA major groove. Four parallel β sheets (two from each dimer) are "pushed" into the groove for the purpose of direct DNA-base sequence recognition. Each side chain makes two hydrogen bonds with bases Arg^{200} to the N_7 and O_6 of guanine; the other two side chain interactions serve to cross link the adenine of the two adjacent base pairs. It has been suggested that a catalytically active form of the enzyme is formed once the requisite sequence specific DNA protein interactions have formed (McClarin *et al.*, 1986; Kim *et al.*, 1990).

Comparison of the amino acid sequence of the *Rsr*I, (an isoschizomer of *Eco*RI) with *Eco*RI showed six regions of homology which correspond to key structural and functional regions of *Eco*RI (Figure 1.4; Stephenson *et al.*, 1989). There are six conserved amino acid regions in the sequence alignment of *Eco*RI and *Rsr*I. Comparisons between *Mun*I and *Eco*RI (*Rsr*I) showed five of the conserved regions (Figure 1.4; Siksnys *et al.*, 1994). According to the X-ray structure of *Eco*RI, (Kim *et*

al., 1990), the residues from conserved regions I and II form a β-meander that contains the active centre. The sequence similarities between *Mun*I and *Eco*RI (*Rsr*I) suggest the endonucleases share common structural organisation of the catalytic site (Siksnys *et al.*, 1994).

Region III (Figure 1.4) forms an extended polypeptide chain motif and a part of the inner recognition helix (Kim *et al.*, 1990). It is at this motif that the set of hydrogen bonds and Van der Waal contacts occur with the inner tetranucleotide sequence (AATT). The same amino acids are found within the *Mun*I endonuclease, namely Gly^{140}, Asn^{141}, Ala^{142}, Arg^{145} from region III and Gln^{115} from region II (Siksnys *et al.*, 1994).

Regions IV and V correspond to parts of structural motifs of *Eco*RI ensuring correct spatial orientation of the main sequence recognition elements (McClarin *et al.*, 1986; Kim *et al.*, 1990). The regions are identical in *Eco*RI and *Mun*I suggesting a similar spacial orientation between *Mun*I and DNA as for the *Eco*RI-DNA interaction (Siksnys *et al.*, 1994).

Region VI forms a part of a loop and outer recognition helix in *Eco*RI providing a structural basis for discriminating the external nucleotide pairs of the hexanucleotide sequence (Kim *et al.*, 1990). There is no similarity in this region (Siksnys *et al.*, 1994), probably due to the different external nucleotide pairs recognised by *Eco*RI and *Mun*I (CG for *Mun*I and GC for *Eco*RI). Kim *et al.* (1990) showed that *Eco*RI discrimination for GC is by both hydrogen bond and Van der Waal contacts between Met^{137} and Ala^{138} and the C base. These amino acids are not conserved in *Mun*I, which instead has a Arg residue located in the region suggesting this residue has interactions with the G base. *Sma*I and *Xma*I endonucleases are imperfect isoschizomers recognising the sequence CCCGGG. *Sma*I cleaves between the internal CG to produce blunt end scissions whereas *Xma*I cleaves between the external cytosines to produce a four base, 5'

overhang. Withers and Dunbar (1993) revealed that each of the endonucleases induces bending of the DNA. The bending of the helix axis appears to be in opposite orientations. The orientation of the *Sma*I induced bend appeared to be towards the major groove and is similar to the bend induced by *Eco*RV (also produces blunt end scissions). The *Xma*I appeared to bend the DNA towards the minor groove. These studies provide further examples of endonuclease induced DNA bending to facilitate restriction by enhancing local protein-DNA interactions. *Eco*RV has been crystallised both with it's specific DNA sequence and also with non specific DNA (Winkler *et al.*, 1993). The conformation of the two DNAs are different. The non specific DNA adopts a β DNA conformation and does not exhibit the bending and unwinding observed with the specific DNA. The major consequence of this is a high affinity Mg^{+2} binding site formed only with the specific DNA. Recently two groups have shown that DNA distortion is not universal for DNA binding to endonucleases (see Aggarwal, 1995 for review). The *Pvu*II complex reveals specific DNA that retains a β DNA like conformation without the kinking and untwisting seen in the *Eco*RI and *Eco*RV complexes (Cheng *et al.*, 1994). This has also been observed with the *Bam*HI complex.

The endonucleases seem to show structural homology which is not detectable at primary level. The structural homology appears to be related to the cleavage patterns of the enzymes (Figure 1.1). *Eco*RI and *Bam*HI share cleavage recognition similarity and also structural similarity; i.e., cleave DNA to produce a 5' overhang. *Pvu*II and *Eco*RV differ from the *Eco*RI and *Bam*HI both structurally and with respect to the sites they recognise; i.e., cleave DNA to produce blunt ends and are structurally similar to one another.

The active sites for the endonucleases are similar in all four endonucleases. However the DNA recognition structural elements are diverse: *Eco*RV uses a loop, *Pvu*II, a antiparallel β sheet and both *Bam*HI and *Eco*RI recognise by the same mechanism, though *Bam*HI lacks the extended chain motif of *Eco*RI (Figures 1.8 and 1.9).

Figure 1.8 Secondary structures of *Bam*HI and *Eco*RI

[Secondary structures of (a) *Bam*HI, (b) *Eco*RI (originally published by Newman *et al.*, 1994). The shaded regions correspond to the common core motif. Reproduced from Aggarwal (1995).]

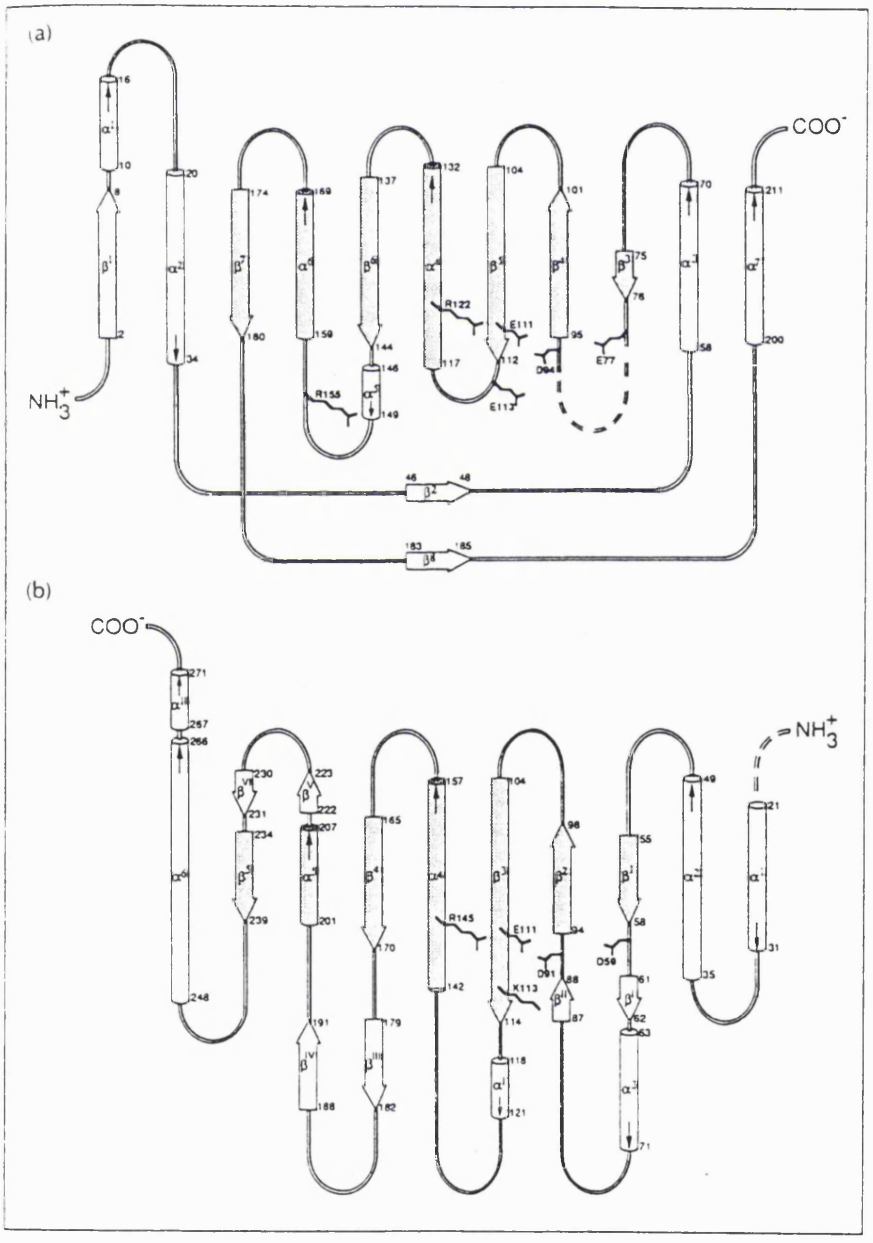
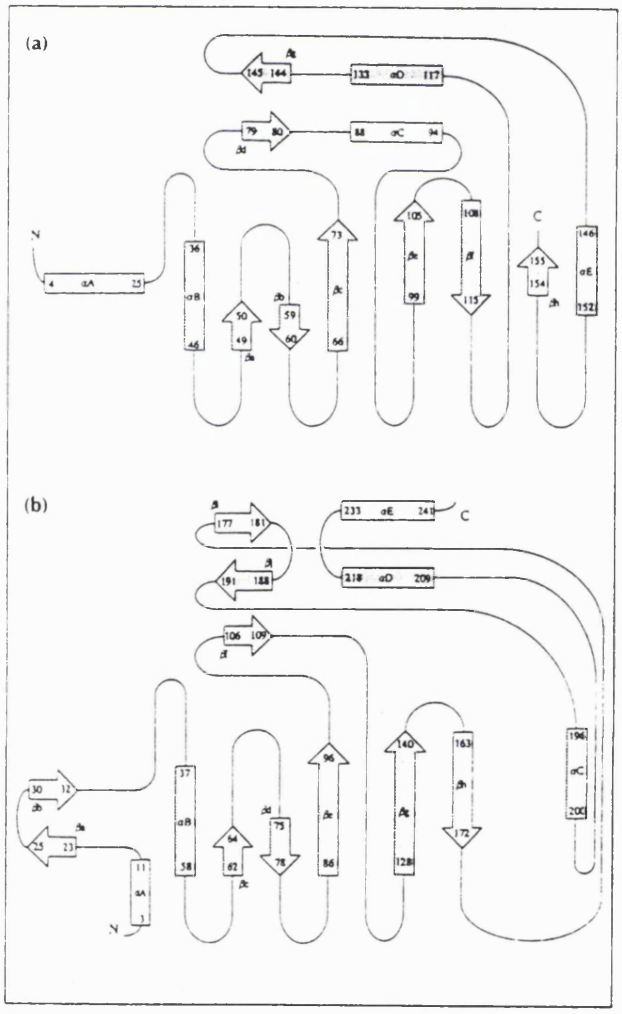


Figure 1.9 Secondary structures of *Pvu*II and *Eco*RV

[Secondary structures of (a) *Pvu*II and (b) *Eco*RV (originally published by Athanasiadis *et al.*, (1994). The shaded regions correspond to topologically equivalent regions. Reproduced from Aggarwal (1995).]



1.8.2 PROTEIN-NUCLEIC ACID INTERACTIONS OF METHYLASES

The *Hha*I MTase from *Haemophilus haemolyticus* is a m^5C-MTase which recognises the GCGC duplex sequence and catalytically transfers a methyl group from AdoMet to the first cytosine of the sequence. The catalytic transfer from AdoMet involves a covalent intermediate. This is the result of nucleophilic attack at the cytosine C6 position by the conserved Cys^{81} residue on the *Hha*I MTase (Figure 1.10; Cheng *et al.*, 1993). Replacement of the proton by a fluorine at the cytosine C5 position chemically traps the covalent reaction intermediate. The crystal structure for the chemically-trapped *Hha*I methyltransferase, S-adenosylmethionine and a duplex 13-mer DNA oligonucleotide containing methylated 5-fluorocytosine at it's target, has been determined at 2.8 Å resolution (Klimasaukas *et al.*, 1994).

This crystal structure has shown that the enzyme remodels DNA so as to present a particular base to the active site (Klimasaukas *et al.*, 1994). *Hha*I MTase binds DNA as a monomer (a binary complex), the cleft providing the binding site for DNA (Cheng *et al.*, 1993). The active site lies in a 20 residue loop (amino acids 80-99) in the large domain. This loop has six of the residues that are conserved (Figure 1.5) including the catalytically active Cys^{81}. On binding, the loop undergoes a conformational change bringing the sulphydryl group of Cys^{81} into close proximity with the target cytosine, allowing the formation of the covalent link (Klimasaukas *et al.*, 1994).

The bound DNA has the characteristic of the β-form DNA with standard Watson-Crick hydrogen bonds between the base pairs except for a completely distorted G-C base pair (containing the target cytosine). The cytosine has "flipped completely out of the DNA helix" (Figure 1.11; Klimasaukas *et al.*, 1994).

Figure 1.10 Ribbon diagrams showing the domain organisation of M·*Hha*I

[The large (amino acids 1-193 and amino acids 304- 327) and small (amino acids 194-275) domains are marked. The AdoMet is omitted from (a). The cleft is large enough to accommodate DNA. Side view showing the DNA binding cleft. Shown in (b) is the view into the cleft. Reproduced from Cheng *et al.* 1993)]

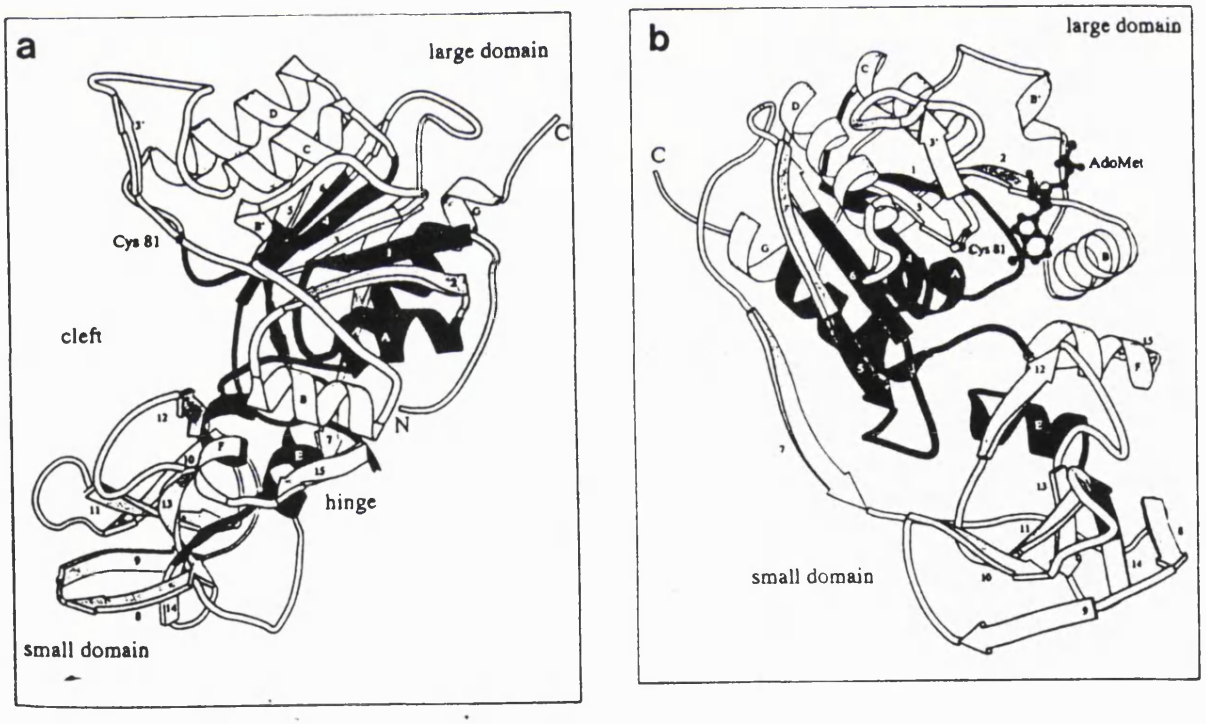
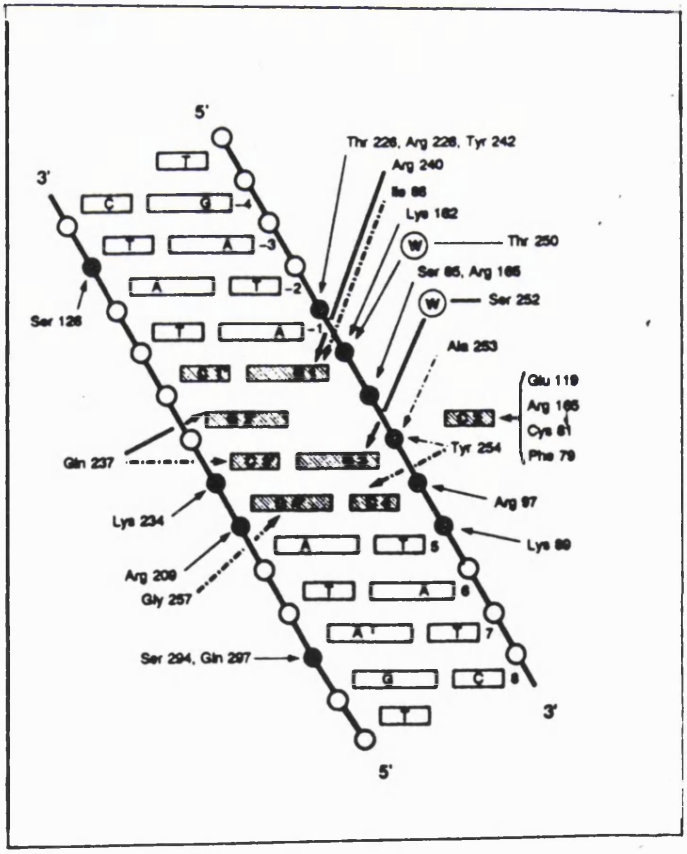


Figure 1.11 Schematic representation showing the specific base and phosphate contacts between M·*Hha*I and DNA

[The DNA is represented as a cylindrical projection. The recognition bases and contacted phosphates are shaded. Base contacts are shown with a thick line, phosphate contacts with a thin line, and contacts through main chain atoms with a dashed line. The symbol (W) indicates the water-mediated contacts. Reproduced from Klimasaukas *et al.* (1994)]



Patel (1994) describes the intermolecular interactions as "a reaching out between extended segments on the protein and the DNA .....a handshake at the molecular level". This is the first structure where it has been shown that a protein distorts the DNA by causing a base to flip out of the DNA helix.

1.9 GENE ORGANISATION OF RESTRICTION MODIFICATION SYSTEMS

The restriction and modifications genes are generally closely linked. In some systems the genes have opposite orientations; (Figure 1.12; see Wilson and Murray 1991, for review). The organisation may be the result of natural selection (Section 1.7.3) and has been exploited in the cloning of restriction modification systems (Section 1.10).

1.9.1 REGULATION OF RESTRICTION MODIFICATION SYSTEMS

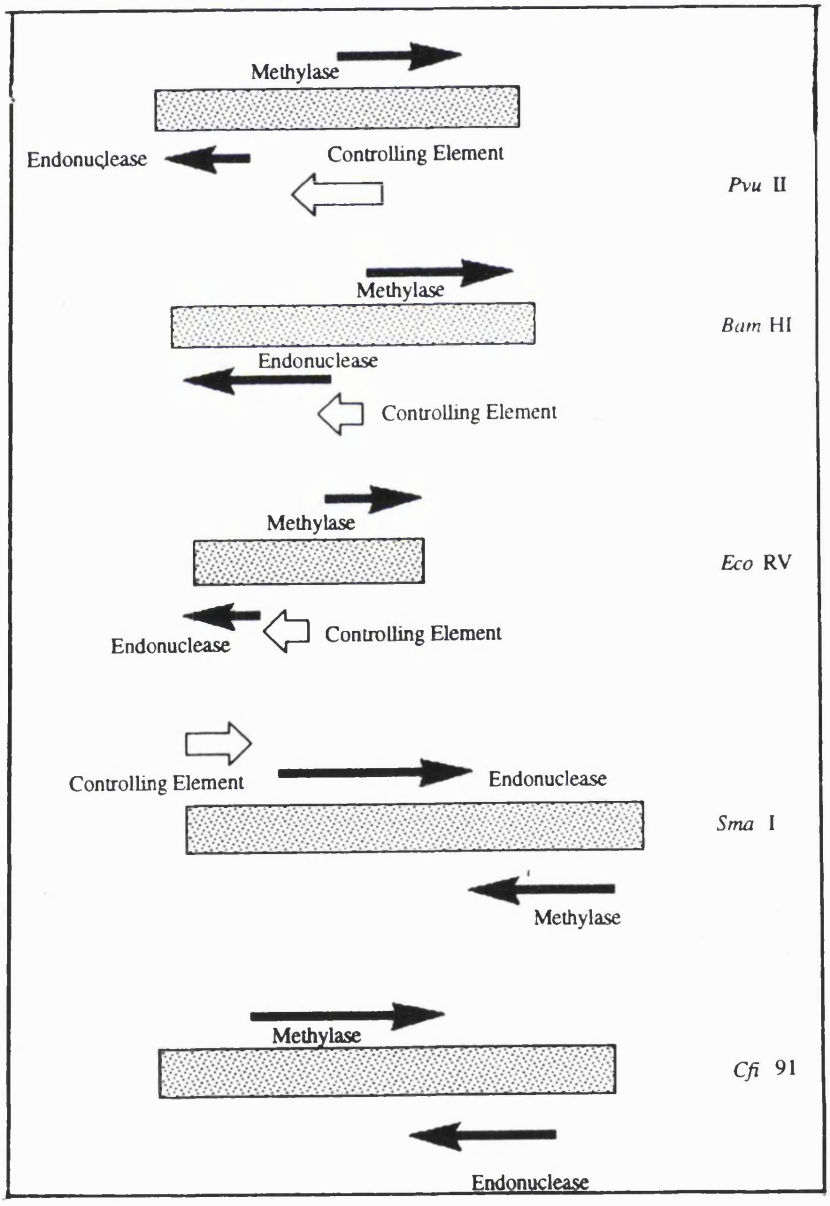
Restriction modification systems are regulated in different ways. Analysis of the *Bam*HI restriction modification system sequences showed two large open reading frames (ORFs). These corresponded to the ENase and MTase respectively. The two genes were divergently orientated. In the intergenic region between the two genes, a third small ORF aligned in the same orientation and overlapping the open reading frame of the endonuclease gene was observed (Figure 1.12). This was 306 nucleotides long and could code for a protein of 13351 r Daltons It had been postulated that this ORF may play a regulatory role (Brooks *et al.*, 1991).

This small ORF of *Bam*HI was designated *bamHIC* for *Bam*HI controlling element. *bamHIC* can act on both *bamHIR* and *bamHIM* via a protein product, C·*Bam*HI. Since C·*Bam*HI contains a potential helix-turn-helix motif which is typical of site specific DNA

binding proteins, the authors hypothesised that C·*Bam*HI acted as a transcriptional activator of *bamHIR* and a repressor of *bamHIM* (Ives *et al.*, 1992).

Figure 1.12 Gene organisation of some restriction modification systems.

[The ORF thought to be involved in regulation is shown for *Pvu*II, *Bam*HI, *Eco*RV, *Sma*I. The figure was adapted from Tao *et al.* (1991). These observations are discussed in greater detail in Section 1.9.1.]



A similar mode of action has been proposed for C·*Pvu*II (Tao *et al.*, 1991; Tao and Blumenthal, 1992). Sequence alignments of the *Bam*HI small ORF with deduced amino acids sequences of the third ORFs identified in other restriction modification systems including *Eco*RV, *Sma*I and *Pvu*II (Tao *et al.*, 1991; Siksnys *et al.*, 1994) revealed a family of related C-proteins (Figure 1.12; Tao *et al.*, 1991).

The nucleotide sequence of the genes encoding M·*Taq*I and R·*Taq*I have been analysed (Barney *et al.*, 1992b). The C terminus of the *taqIM* gene overlapped the N terminus of the *taqIR* gene by thirteen codons. This also occurs with the *Taq*I isoschizomer *Tth*HB81 (Barney *et al.*, 1992a). It is thought that this overlap plays a role in regulating *taqIR* expression. The authors speculated that the overlap region in the *Taq*I and *Tth*HB81 systems permits the formation of a hairpin structure which would promote termination of transcription.

Similarly *Cfr*9I has both endonuclease and methylase gene in the same transcriptional orientation, but differing in their translational phases (Lubys *et al.*, 1994). The ORF of the MTase and the ORF of the ENase overlap by 4 bases (Figure 1.11). In the case of the endonuclease gene, 16 bases upstream from the ATG start codon is the sequence TAAGGA, predicted to be Shine-Dalgarno sequence. A complementary nucleotide sequence is located within the methylase gene. Formation of secondary mRNA structure involving these complementary sequences cannot be excluded. Since initiation of R·*Cfr*9I translation may be hindered by mRNA secondary structure, its synthesis could be initiated by disrupting mRNA structure with an oncoming ribosome followed by a translation restart. The authors suggested that the cell may exploit this feature to permit endonuclease expression only after the methylase had already been synthesised (Lubys *et al.*, 1994). Such a mechanism of regulation could ensure protection of the host DNA against destruction by the endogenous ENase.

1.9.2 BASE COMPOSITION AND FLANKING REGIONS OF RESTRICTION MODIFICATION SYSTEMS

Examination of the ORFs of the restriction modification systems has shown the systems to have a preference for rare codons. The overall frequency of optimal codon selection for each of the proteins in the *Pst*I restriction modification system is less than that predicted even on a random basis (Walder *et al.*, 1984). There are 86 leucine residues in the two proteins. Of these none utilise the optimal codon CUG but instead are coded for by UUA.

The pattern of codon usage in the *Pst*I genes is very similar to that observed for the *Eco*RI genes. The overall frequency of optimal codon selection for the restriction endonuclease is 38% and the methylase 40%. The optimal codon for leucine CUG is rarely used and again UUA is predominant. The coding systems are AT rich for *Pst*I (62% and 66% for the endonuclease and methylase respectively) and similarily for the *Eco*RI genes (65%). However the flanking regions of both systems are also highly AT rich.

For the *Bam*HI restriction-modification system the base composition within the sequenced region is 31% GC which is substantially lower than the 47% GC reported for *B. amyloliquefaciens*. This trait is emphasised when examining codon usage in the genes. The *Bam*HI gene codon usage values are very skewed in their preference for A or U over G or C in the third positions (approximately 75% in each). The codon preference does not resemble that of strongly expressed genes of *B. subtilis* (which is a closely related species) suggesting the restriction modification genes are not strongly expressed. Many of the RM gene sequences reported to date show unusually low GC content and third position A/U preference (Karyagina *et al.*, 1993). The GC content of the genes of the *Mbo*I restriction modification system is approximately 30% which is lower than the

average of *Moraxella bovis* which is 41-44.5% (Ueno *et al.*, 1993). This has also been seen for the *Mbo*II restriction modification system (Bockage *et al.*, 1991). It has been suggested by the authors that this reflects a relatively recent evolutionary origin of restriction modification systems from a progenitor organism with low GC content. However as the restriction modification systems show no homology this argument seems flawed.

Thermus aquaticus is a thermophile and might be expected to be GC rich. For the RM genes the disproportion is slight; 59% GC and 52% GC for the R and M respectively. In non coding sequences, there is substantially more GC content (66%) [Slatko *et al.*, 1987]. However, codon usage reflects a preference for GC in the third position. This system also differs in that it has seven *Taq*I site in the endonuclease gene and none in the methylase gene. This is unusual and not widespread. It is postulated that this asymmetric distribution of sites could be important in the regulation of the endonuclease gene although the actual mechanism is unknown. It is thought that in an unmodified cell, interaction between the endonuclease with the *Taq*I site on the restriction endonuclease gene might interrupt transcription of the gene and lead to a deletion or death of the cell. If the DNA was hemimethylated the expressed endonuclease gene would not be able to prevent further endonuclease synthesis until the cell was fully modified.

Surrounding the ORFs of restriction modification systems, other ORFs have been identified. The *Hpa*II restriction modification system revealed two adjacent genes Kulakauskas *et al.*, 1994). Upstream and partly overlapping *hpaIIM* is the coding sequence for a protein that resembles the very-short-patch -repair endonuclease (*Vsr*) of *E. coli.* Further upstream is the coding sequence for a protein that resembles valyl-tRNA synthetase (*ValS*).

1.10 CLONING RESTRICTION MODIFICATION SYSTEMS

1.10.1 INTRODUCTION

Sixteen years ago, the first paper appeared describing the cloning of a complete restriction modification system in *E. coli* (Mann *et al.*, 1978). The cloned restriction modification systems provide material for investigating the biology of DNA modification and restriction and the biochemistry of DNA-protein interactions. They also constitute new, often more convenient or prolific sources from which restriction and modification enzymes can be purified. An example of over-expression of a restriction endonuclease is the *Kpn*I restriction-modification system, cloned and expressed in *E. coli* (Hammond *et al.*, 1990). The genes for restriction and methylation were isolated sequentially (Section 1.10.3). Using a two step method, the strain was constructed with two compatible plasmids, an inducible plasmid with the *Kpn*I restriction gene which is activated at elevated temperatures and a pBR322 derivative expressing the methylase. The new strain produced 10 million units of *Kpn*I restriction endonuclease activity/g of wet weight cells. This was several 1000 fold higher than the level of *Kpn*I produced by the source bacteria *K. pneumoniae* (Hammond *et al.*, 1990).

There are many different methods of cloning restriction modification systems. The cloning methods use the functional properties of the restriction modification system, i.e., the ability to restrict foreign DNA and the ability to methylate DNA to isolate the clones.

1.10.2 PLASMIDS CODING FOR RESTRICTION MODIFICATION SYSTEMS

Natural plasmids can be subcloned if the restriction modification systems are coded within the plasmid. Purification of the plasmid, i.e., removal of chromosomal DNA prior to cutting and ligation, reduces the complexity of the DNA. The resulting fragments can

then be characterised. This is preferable to preparing genomic libraries and screening for restriction modification systems. *Eco*RI, *Eco*RII and *Eco*RV, *Pae*R7I and *Pvu*II are all found on naturally occurring plasmids and have all been cloned in this manner (Wilson 1988).

The *Sso*II RM system was also cloned in this method (Karyagina *et al.*, 1990). *Shigella sonnei* strain 47 contains a complex array of plasmids. The *S. sonnei* 47 plasmids were transferred to *E. coli* by both transformation and conjugation. The plasmids were selected by susceptibility to phage infection (Section 1.10.3). The genes for the endonuclease and the methylase are located on two small plasmids (designated -P6 and P4 respectively) of *S. sonnei* 47.

1.10.3 SELECTION BASED ON SUSCEPTIBILITY TO PHAGE INFECTION

Some of the earliest restriction modification systems were cloned on the principle that cells that restrict are less susceptible to phage infection. Clones containing the *Hha*II (Mann *et al.*, 1978) and *Pst*I genes (Walder *et al.*, 1984) were isolated by their ability to survive infection. For the *Pst*I cloning experiments, transformants carrying hybrid plasmids of *P. stuartii* 164 chromosomal DNA and pBR322 were screened using bacteriophage λ. Survivors were assayed for restriction activity. However, this method has been found to be of limited value as cloned restriction modification genes do not always manifest sufficient phage resistance to confer selective survival.

1.10.4 SELECTION BASED ON MODIFICATION

A more popular method of cloning restriction modification genes is based on selection using DNA modification. Methylase selection involves screening for methylase clones by exposing DNA from transformed hosts with the corresponding restriction endonuclease.

Survival indicates the presence of the methylase gene, since the DNA of the host is modified and insensitive to attack by the restriction endonuclease. Where restriction and modification are closely linked, both genes can be cloned simultaneously.

The *Ban*I restriction modification system was cloned in a single step method. The chromosomal DNA from *B. aneurinolyticus* was partially digested with *Sau*3AI. The DNA was ligated to *Bam*HI digested pBR322 and used for transformation of *E. coli*. Plasmid DNA was isolated from approximately 10 000 ampicillin resistant transformants. Plasmid DNA was digested with *Ban*I and reintroduced into *E. coli*. The resultant transformants were assayed for restriction activity. The genes for both restriction and modification were isolated on a 3.1kb fragment (Maekawa *et al.*, 1990).

Methylase selection does not necessarily yield a complete restriction system. Even attempts to clone larger regions of DNA failed to produce an active endonuclease gene in the *Bam*HI and *Dde*I systems (Howard *et al.*, 1986; Brooks *et al.*, 1989). In these cases the failure stemmed from the introduction of the endonuclease gene into a host which is not adequately protected by methylation. These restriction modification systems were cloned by a multi-stage procedure;

(i) The methylase gene was cloned by *in vitro* selection.

(ii) The genes were manipulated to achieve full methylation of the host.

(iii) The modified cells were used as recipients of the restriction gene in a separate round of cloning.

The *Dde*I restriction modification system (Howard *et al.*, 1986) was cloned in this two step procedure as were the *Bam*HI (Brooks *et al.*, 1989) and the *Sty*LTI systems (De-Backer and Colson, 1991).

The genes for the *Sfi*I restriction-modification system were isolated and expressed via a different method. The restriction endonuclease was purified to homogeneity, the N terminal sequence determined and used to make an oligonucleotide probe which was used to screen the *Sfi*I modified libraries (VanCott, 1990).

A potential obstacle to cloning RM genes lies in the discovery that some strains of *E. coli* react adversely to cytosine or adenine modification. The methylation of adenine induces the SOS repair response (Heitman and Model, 1987; Raleigh *et al.*, 1988). Adenine methylated DNA in certain sequences is recognised by the product of a locus named *mrr* . It has been suggested that *mrr* encodes an endonuclease that cleaves adenine methylated DNA and that DNA scission induces SOS. Cytosine methylases foreign to *E. coli* also induce restriction. Restriction is due to two genetically distinct systems *mcr*B and *mcr*A (modified cytosine restriction) that differ in their sequence specificities (Table 1.3). When cloning RM systems it is therefore advisable to use mutant strain of *E. coli* ($mcrA^-$, $mcrB^-$, mrr^-), in which these systems are defective.

Table 1.3. Methylation systems in *E. coli*

System	Sequence Recognised
mrr	GAC where A is methylated
	CAG where A is methylated
mcr A	CC*GC where C* is methylated
mcr B	TC where C is methylated
	GC where C is methylated

Piekarowicz *et al.* (1988) constructed a mutant strain which possessed a temperature sensitive *mcr*B gene. The method used the principle that restriction is due to the degradation of the incoming methylated plasmid DNA by the products of the genes

*mcr*B, *mcr*A and *mrr*. The strategy for cloning DNA MTase genes assumes that cloned MTase and a temperature sensitive mutation for one of these new types of restriction systems would be viable and able to form colonies at the permissive temperature (42°C) but unable to form colonies at the non permissive temperature (30°C). The inability of such transformants to grow at 30°C provides a simple method for the identification of DNA methylation clones. Since the *mcr*B nuclease recognises a great variety of sequences, it is possible to use the strain to clone various types of DNA methylase genes.

Recently, the same group have constructed strains which are temperature sensitive for the *mcr* and *mrr* systems but also further modified to include the *lac* Z gene fused to the damaged inducible *din* D locus of *E. coli* (Piekarowicz *et al.*, 1990). The detection of recombinant plasmids encoding restriction and modification enzymes is a simple one step procedure that is based on induction at the restrictive temperature of the *lac* Z gene. Transformants encoding the modification genes are detected as blue colonies on LB agar plates supplemented with X-Gal, hence reducing the need to replica plate thousands of colonies. Using this method the authors have cloned a variety of different methyltransferase genes from many diverse species, and also the thermostable *Taq*I and *Tth*111I restriction enzymes (Fomenkov *et al.* 1994).

1.11 THERMOPHILIC BACTERIA

Microorganisms that grow at high temperatures are referred to as "thermophiles" and belong to either of two groups, the Archea or the Bacteria (Table 1.4). There is great diversity in the metabolism and ecology of thermophilic bacteria, having a range of T_{max} and T_{opt}, pH_{opt} and growth conditions (see Kristjansson and Stetter (1992) for review). Thermophiles can be found in most bacterial metabolic groups. Phylogenically, thermophilic bacteria are usually the oldest representatives of their respective groups in

the bacterial tree, implying that thermophily arose very early in the evolution of bacteria (Hartman *et al.*, 1989; Achenbach-Richer *et al.*, 1987; Huber and Stetter, 1992).

Table 1.4 Comparison of Bacteria and Archaea

[Reproduced from Kristjansson and Stetter (1992)]

Character	Bacteria	Archaea
Cell wall components	Murein	Pseudomurein
Membrane lipids	Glycerol fatty acid esters	Glycerol isopropyl ethers
Square and flat structures	-	+
Endospores	+	-
tRNA "common arm" contains	Ribothymidine	Pseudouridine or 1-methylpseudouridine
Methyonyl initiator tRNA formylated	+	-
Introns in genes	-	+
Eukaryotic RNA polymerase	-	+
Special coenzymes	-	+ (only in Methanogens)
Max growth temperature	90°C	110°C
Complete photosynthesis	+	-
Methanogenesis	-	+
Calvin cycle used in CO_2 fixation	+	-

1.11.1 THE GENUS *THERMUS*

Large collections of strains of *Thermus* have been isolated from Japan, the former USSR, Iceland, Mexico, Yellowstone National Park, USA, New Zealand and Portugal from hot springs (Georgia *et al.*, 1993) where temperatures ranged from 53°C to 86°C and in pH conditions higher than 6.5 (Williams, 1992).

Brock and Freeze (1969) first described the properties of the genus *Thermus* and its type strain *Thermus aquaticus* in 1969. *Thermus* strains, with the exception of *Thermus*

filiformis which always produces very long intertwined filaments, grow as pleomorphic rod shaped cells with short filaments.

All strains stain a Gram negative, are cytochrome oxidase positive and nonmotile in liquid cultures and do not form flagellae. The red pigmented *T. ruber* is usually easily distinguished by the presence of a red pigment and lower growth temperature optimum. The DNA:DNA homology within the *T. ruber* species is high but low compared with yellow and colourless strains (Sharp and Williams, 1988). *T. filiformis* showed great homology with non filamentous *Thermus* strains isolated from New Zealand hot springs but has low DNA:DNA homology with the species *T. aquaticus*, *T. thermophilus* and *T.brockianus*. (Georgia, *et al.* 1993). The authors suggest all the isolates which showed homology with *T. filiformis* should be regarded as strains of the *T. filiformis* despite the inappropriateness of this name to cellular morphology of all but the type strain (Georgia *et al.*, 1993).

There is relatively little known about the genetics of *Thermus*. From the number of sequences available for genes coding for proteins from *Thermus spp.* a consensus promoter has been identified (Figure 1.13, Bergquist and Morgan 1992) This sequence is very different from the *E. coli* consensus promoter sequence and from the 23S/5S ribosomal operon of *Thermus thermophilus* (Hartmann *et al.*, 1987). The consensus ribosomal binding site was within 2-10 base pairs of the translational start site. Only five genes were suitable for the analysis of termination signals: four of the five of these genes contained inverted repeats located from 12-41 downstream of the stop codon (Bergquist and Morgan, 1992).

Figure 1.13 Consensus promoter sequence for *Thermus spp.* cloned genes.

[Bold type, present in all sequences; upper case type, present in 75% or more sequences, lower case type, present in 50% or more sequence. (Bergquist and Morgan, 1992)]

Thermus spp. **consensus promoter sequence**

Gg-CCTC-C--GG-CG-9-14bp-CCtTTa-

E. coli **promoter sequence**

tc-TTGACat-t-9bp-t-tg-TAtAaT-

1.11.2 CLONED THERMOPHILIC ENZYMES

Thermophilic enzymes are of fundamental interest to biochemists because they provide a means for investigating the molecular mechanisms which control the folding, unfolding and flexibility of protein structures. Thermophilic bacteria are a popular source of genes encoding highly active thermostable proteins since the major component of thermostability is genetically encoded.

Cloning genes encoding thermophilic proteins into mesophiles removes the difficulty and cost of growing organisms at high temperatures. There are also lower yields of biomass produced from some thermophilic fermentations compared with biomass yields produced by mesophiles. Ease of purification is another important reason for cloning thermophilic enzymes into mesophilic hosts. The recovery of intracellular enzymes usually follows a

defined sequence of operations. This involves enzyme release and removal of cell debris, often followed by a sequence of chromatographic separations (Bonnerjea *et al.*, 1986). Each step of the process requires a compromise between yield, purification and the cost of the materials and labour involved. A considerable number of such steps are often involved in enzyme purification and the cumulative effect of yield loss at each stage can lead to low final product yields (Fish and Lilly, 1984).

The use of heat treatment as a means of partially purifying thermostable proteins from more labile proteins is inexpensive, rapid and can result in both high purification and yield of thermostable proteins. Under conditions, where the thermophilic enzyme is stable, cellular proteins of the mesophilic host are heat denatured and precipitated. This method has already been used successfully in the purification of many recombinant proteins expressed in mesophilic hosts (e.gRüdiger *et al.*, 1995; Sota *et al.*, 1994; Gabelsberger *et al.*, 1993; Lee *et al.*, 1993; Koyama *et al.*, 1993; Sakamoto *et al.*, 1994; Arnone *et al.*, 1992).

Several genes have been cloned from *Thermus spp.* (reviewed by Bergquist and Morgan, 1992). *T. aquaticus* possesses a type II restriction modification system (*Taq*I) which has been cloned (Slatko *et al.*, 1987). The genes for the *Taq*I restriction modification system were cloned as described in Section 1.10.4. Genetic modification of the cloned DNA allowed a significant increase in the level of expression. Commercially the cloned thermophilic endonucleases serve both as isoschizomers of mesophilic endonucleases and as novel enzymes with unique cleavage sites. Table 1.5 lists the commercial sources that sell thermophilic endonucleases.

However, none of the cloned *Thermus* genes have attracted the scientific attention as the cloned *Taq* polymerase from *T. aquaticus*. *Taq* polymerase is used in the PCR

The PCR invented by Mullis and Faloona (1987) has found a variety of uses in

diagnostic medicine, anthropology, taxonomy and molecular biology. It is an *in vitro* method for primer directed amplification of specific sequences, defined at their 5' and 3' ends by specific synthetic oligonucleotide primers. The advantage of *Taq* polymerase in the PCR reaction is its half life at the DNA denaturing temperature ($t_{1/2}$ at 95°C, 40 minutes). Over the past five years a number of other thermophilic polymerases have been isolated and are used commercially including Pfu™ and Vent™ polymerase (Table 1.5).

The cloning of *Taq* polymerase illustrated the difficulties of manipulating thermophilic genes and their products, which may show minimal enzymatic activity at the usual growth temperature of the cloning host (37°C). Expression of the DNA polymerase in the host organism was low and polymerase activity could not be demonstrated in the recombinant libraries. Isolation of the gene was only achieved by screening using an antibody to the *Taq* polymerase protein (Lawyer *et al.*, 1989). With both the *Taq*I endonuclease and the *Taq* polymerase there does not seem to be a barrier to the expression of the thermophilic protein in *E. coli* with the thermophilic enzymes found to maintain their thermal stability after synthesis in *E. coli.*

Table 1.5 Commercial Sources of Restriction Enzymes

[Companies selling restriction endonucleases are listed. Data obtained from REBASE, Roberts and Macelis (1994).

The list is non exhaustive. The thermostable polymerases sold by some of the companies are also included.]

Company	Thermostable polymerase
American Allied Biochemical, Inc.	
Amersham Life Sciences USB	*Taq* DNA polymerase isolated from *Thermus aquaticus* Thermozyme™ DNA polymerase isolated from *Thermus brockianus* *Tth* DNA polymerase isolated from *Thermus thermophilus* HB8 Δ*Taq* ® Version 2.0 DNA polymerase is derived from *Thermus aquaticus* DNA polymerase but is genetically modified to improve its properties for sequencing
Angewandte GentechnologieSysteme	
Appligene	
Boehringer-Mannheim	*Taq* DNA polymerase isolated from *Thermus aquaticus* BM
CHIMERx	
Fermentas MBI	
Life Technologies Inc. Gibco-BRL	
NBL Gene Sciences Limited	
New England Biolabs	VentR™ DNA polymerase* (VentR™(exo⁻)* and Deep VentR™ DNA polymerase†) *isolated and cloned into *E.coli* from the archaebacterium *Thermococcus litoralis* †isolated and cloned into *E.coli* from *Pyrococcus species* GB-D
Pharmacia Biotech Inc.	*Taq* DNA polymerase and *Tth* DNA polymerase isolated from *Thermus aquaticus* (YT1) and *Thermus thermophilus* respectively
Promega Corporation	*Taq* DNA polymerase and *Taq* DNA polymerase (Sequencing Grade) Both polymerases isolated from *Thermus aquaticus* (YT1)
Sigma	
Stratagene	*Pfu* DNA polymerase*, Recombinant Exo⁻ *Pfu* DNA polymerase* and *Taq* DNA polymerase† *isolated and cloned from *Pyrococcus furiosus* †isolated from *Thermus aquaticus* (YT1)
Takara Shuzo Co.Ltd	

1.12 AIMS

A thermostable restriction endonuclease has been isolated by Cowan *et al.,* (unpublished observations) from a novel thermophilic microorganism thought to be the extreme thermophile *T. filiformis.* The cells are long multicellular, non branching filaments with intermediate "bulbs" and terminal spherical swellings. The strain designated Rot34A1 grows optimally at 70°C and was isolated from a thermal site in Rotorua, New Zealand. The enzyme named *Tfi*I, in accordance to the rules proposed by Smith and Nathans (1973), recognises the sequence $GA^{A}/_{T}TC$ which is a hitherto unknown specificity and has not been reported from either mesophilic or thermophilic sources. The more degenerate pentanucleotide recognition site GANTC (where N=any nucleotide) has been reported from a number of sources (see Kessler and Manta (1990) for primary references). The enzyme is remarkably thermostable, surviving at room temperature for several weeks and having a half life of more than six hours at 70°C. The novel specificity and high thermal stability makes *Tfi*I, a valuable addition to the catalogue of commercial DNA manipulating enzymes.

The aims of the project were to characterise the Rot34AI restriction modification system (*Tfi*I restriction modification system) via a number of biochemical and molecular genetic techniques.

2 Materials and Methods

2.1 CULTURES

2.1.1 GROWTH AND MAINTENANCE OF ROT34A1

The strain designated Rot34A1 was maintained as glycerol stocks [20% glycerol in water (v/v)] and freeze dried vials and stored at -70°C. Cells were firstly grown on Castenholz Medium D agar plates overnight at 70°C and then used as inoculum for liquid culture. Liquid cultures were grown in 2.5L shake flasks containing 1L of Castenholz Medium D media at 70°C shaken at 150 rpm in an orbital shaker for 18 - 24 hours.

2.1.2 GROWTH AND MAINTENANCE OF *E. COLI* STRAINS

The *E. coli* strains used are shown in Table 2.1 with the strain characteristics. Strains were maintained as glycerol stocks and stored at -70°C. Selective media was used where appropriate [ampicillin ($20\mu gmL^{-1}$) and tetracycline ($10\mu gmL^{-1}$)] and IPTG ($40\mu gmL^{-1}$) and XGal ($10\mu gmL^{-1}$) were added when necessary. Cultures were grown at 37°C for 18-24 hours unless otherwise stated.

2.2 PREPARATION OF MEDIA AND BUFFERS

All media components and chemicals were supplied by Aldrich Chemical Company, Dorset, UK; BDH Laboratory Suppliers, Leicestershire, UK; Difco Laboratories, Michigan, USA; Gibco BRL Life Technologies, Paisley, UK; Oxoid, Hampshire, UK and Sigma Chemical Company Ltd., Dorset, UK. (In general all chemicals were of AnalaR quality or equivalent. Radionucleotides was supplied by Amersham International, Buckinghamshire UK.

Table 2.1 Bacterial Strain Genotypes

Strain	Genotype	Reference
GM2163	F-*ara*-14 *leu*B6 *thi*$^{-}$1 *fhuA*31 *lacY1 tsx*-78 *galK2 galT22 supE*44 *hisG4 rpsL*136 (Str^{r}) *xyl-5 mtl-1 dam*13:: Tn9 (Cam^{r}) *dcm*-6 *mcrB*1 *hsd*R2 ($r_k^-m_k^+$) *mcrA*	Woodcock *et al.*, 1989
JM107	F',traD36, *lacI*^{q}Δ (*lacZ*)*M15 proAB/ e*14^{-}(mcrA^{-}) Δ(*lac*$^{-}$*proAB*) *thi gyr*A96 (Nal^{r}) *end*A1 *hsd*R17 ($r_k^-m_k^+$) *rel*A1 *sup*E44, *rel*AI *sup*E44 *mcrA*	Yanisch-Perron *et al.*, 1985
LE392	e14-(*mcrA*),*hsd*R514, *sup*E44, *sup*F58,*lacY*1 or Δ(*lacIZY*)6, *gal*K2, *galT*22, *metB*1, *trpR*55	Silhavy *et al.*, 1984
P2392	LE392 (P2 lysogen)	Silhavy *et al.*, 1984
XL1-Blue MRA	Δ(*mcrA*)*183*,Δ(*mcrCB-hsdSMR-mrr*)173, *end*A1, *sup*E44,*thi*-1,*recA*1,*gyrA*96,*relA*1, *lac*	Jerpseth *et al.*, 1992
XL1-Blue MRA (P2)	XL1-Blue MRA (P2 lysogen)	Jerpseth *et al.*, 1992

2.2.1 MEDIA

All media were autoclaved for 20 minutes, 121°C at 15psi.

2.2.1.1 GROWTH MEDIA FOR ROT34A1

Castenholz Medium D (Hudson *et al.*, 1984)

Nitrilotriacetic acid	$0.1gL^{-1}$
$CaSO_4.2H_2O$	$0.06gL^{-1}$
KNO_3	$0.103gL^{-1}$
$NaNO_3$	$0.689gL^{-1}$
Na_2HPO_4	$0.111gL^{-1}$
NaCl	$0.008gL^{-1}$
Yeast Extract	$1gL^{-1}$
Bacteriological Peptone	$1gL^{-1}$

Liquid media was made up to 1L with distilled water and the pH adjusted to 8.2 with NaOH. Solid media was made up as above except bacteriological agar was added ($15gL^{-1}$). After autoclaving, $MgSO_4.H_2O$ was added at a concentration of $0.1gL^{-1}$.

2.2.1.2 GROWTH MEDIUM FOR *E. COLI* AND PHAGE STRAINS

(i) Nutrient agar	$28\ gL^{-1}$
(ii) Nutrient broth	$25\ gL^{-1}$
(iii) NZY broth	
NaCl	$5gL^{-1}$
$MgSO_4.H_2O$	$2gL^{-1}$
Yeast Extract	$5gL^{-1}$
NZ Amine	$10gL^{-1}$

(iv) NZY top agar

NZY broth	
Agarose	0.7% (w/v)

(v) TB media

NaCl	$5gL^{-1}$
Bactotryptone	$10gL^{-1}$

2.2.2 BUFFERS

(i) 50x phosphate buffer (500mM pH 7.0)

27.22g of KH_2PO_4 and 68.47g of K_2HPO_4 was dissolved in 1L of distilled water and autoclaved.

(ii) 10mM potassium phosphate buffer pH 7.0

5.44g of KH_2PO_4 and 13.7g of K_2HPO_4 was dissolved in 1L of distilled water After autoclaving 5mL of β-mercaptoethanol was added to give a final concentration of 7mM. The buffer was then diluted 1 in 10 to give a 10mM potassium phosphate buffer.

(iii) TE buffer

10mM Tris-HCl pH 7.5
1mM Na_2EDTA

(iv) 10 x Restriction Buffer (10xRB)

500mM	Tris-HCl pH 7.5
50mM	$MgCl_2$

(v) Stop Mixture

40% (w/v) sucrose
0.1M Na_2EDTA
0.015 $mgmL^{-1}$Bromophenol blue

(vi) Tris -Borate Electrophoresis Buffer (TBE)

A 10x stock of TBE was prepared by mixing 108g of Trizma Base, 55g of Boric acid and 9.3g of Na_2EDTA to 1L of distilled water. The stock solution was then autoclaved. TBE was made up by diluting the stock solution 1 in 10 and adding EtBr to a final concentration of $0.5\mu gmL^{-1}$.

(vii) Protein solubilisation solution

125mM Tris-HCl pH 6.8
4% (w/v) SDS
0.002% (w/v) bromophenol blue
20% (v/v) glycerol
10% (v/v) β-mercaptoethanol

(viii) 2xSDS Loading buffer

0.1M Tris-HCl pH 6.8
4% (w/v) SDS
0.2% (w/v) bromophenol blue
20% (v/v) glycerol
200mM DTT (added fresh from 1M stock)

(ix) 10x methylase buffer

500mM NaCl
500mM Tris-HCl (pH 8.0)
100mM Na_2EDTA

(x) 10x Ligase buffer

660mM Tris-HCl (pH 7.5)
100mM $MgCl_2$
100mM dithiothreitol
1mM ATP

(xi) GAPDH 10x Salt Buffer

0.3M Triethanolamine pH 7.5
0.5M KCl pH 7.5
0.05M $MgCl_2$
2mM Na_2EDTA pH 7.5

(xii) GAPDH Assay Mix

100µL GAPDH 10x salt buffer
100µL 100mM 3 PGA (Sigma)
50µL 10mM NADH
3µL 4.5U PGK (yeast)
H_2O 697µL - xµL
xµL sample

(xiii) Phenol Chloroform

50mL of 1M Tris-HCl pH 7.5, 0.2M Na_2EDTA pH 8.0 and 200mL of distilled water was added to 500g of phenol (BDH AnalaR). The mixture was shaken to dissolve the phenol crystals. The mixture separated into two layers and the lower (phenol) layer was aliquoted into a separate container and an equal volume of chloroform added. The phenol chloroform was stored at 4°C and the buffer saturated phenol stored at -20°C.

(xiv) Birnboim Solution

(Birnboim and Doly, 1979)

25mM Tris-HCl pH 7.5
10mM Na_2EDTA
0.9%(w/v) glucose
2mgmL^{-1} lysozyme

(xv) SDS-NaOH solution

(*Solution II*)

1% (w/v) SDS
0.2N NaOH

(xvi) 4M sodium acetate pH 6.0

82.01g of sodium acetate were dissolved in 200mL of water and 10mL of glacial acetic acid added. The solution was made up to 250mL with H_2O.

(xvii) TE-RNase solution

10mM Tris-HCl pH 7.5
1mM Na_2EDTA
20µgmL^{-1} RNaseI

(xviii) SM buffer (L)

5.8g NaCl
2.0g $MgSO_4$
50mL 1M Tris-HCl pH 7.5
5mL 2% (w/v) gelatin

(xix) *Solution III*

5M potassium acetate	60mL
glacial acetic acid	11.5mL
distilled H_2O	28.5mL

(xx) 10x CAP buffer

0.5M Tris-HCl pH 9.3
10mM $MgCl_2$
1mM $ZnCl_2$
10mM spermidine

(xxi) pH buffers

Buffers for different pHs were prepared as described Dawson *et al.* (1986) (pp 426-441).

a. Glycine -NaOH buffer pH 8.6 - 10.6 at 25°C

b. Citric acid-Na_2HPO_4 buffer pH 2.6 - 7.6 at 25°C

c. Na_2HPO_4 - NaH_2PO_4 buffer pH 5.8 - 8.0 at 25°C

d. Carbonate buffer pH 9.7 - 10.9 at 25°C

e. Tris-HCl buffer pH 7.1 - 8.9 at 25°C

2.3 AGAROSE GEL ELECTROPHORESIS

Gel electrophoresis was performed using 1% (w/v) agarose gels unless otherwise stated. 1g of agarose was added to 100mL of TBE (see Section 2.2.2) and dissolved by boiling. Molten agarose was poured onto glass plates surrounded in autoclave tape and a comb inserted 10 mm from the top of the plate. Once set, the autoclave tape and comb were removed and the gel submerged into 1x TBE. The gels were electrophoresed at 200V for 1-2 hours or 20V overnight.

2.4 RADIOACTIVITY

Incorporation of radioactivity was measured by addition the samples into 5mL of Ecoscint A (National Diagnostics, Georgia, USA) and measured using the Packard TriCarb Liquid Scintillation Analyser Model 1900 CA.

The conditions for measurement are shown below.

Cycles	1
Count Time (min)	10.0
# Counts/Vial	1
Radionuclide?	^3H
Data Mode?	cpm

2.5 PREPARATION OF DIALYSIS TUBING

Gloves were used when handling the dialysis tubing. Dialysis tubing was prepared by cutting to the appropriate size (10 -20cm) and boiled in water with a few crystals of Na_2EDTA added. The tubing was washed with distilled water.

2.6 RESTRICTION DIGESTIONS AND LIGATIONS

Restriction digestions were performed by the method of Sambrook *et al.*, 1989. All reactions were carried out at the appropriate temperature for the endonuclease in restriction buffer. Ligations were performed in ligation buffer at 4°C overnight unless otherwise stated. Aliquots was tested for digestion and ligation by electrophoresing samples in parallel on 1% (w/v) agarose gels.

2.7 DNA PREPARATION

2.7.1 LARGE SCALE PREPARATION OF ROT34AI CHROMOSOMAL DNA

Cells (25-100mL) were grown to exponential phase (18-24 hours) and harvested by centrifugation. The cells were resuspended in 1mL of TE. A few crystals of lysozyme were added and the cells incubated for 15-30 minutes at 37°C. 0.1mL of 10% (w/v) SDS and 0.5mL of $10mgmL^{-1}$ pronase (which had been self digested at

37°C for 2 hours) was added and the extract incubated at 37°C for one hour. 4mL of TE was then added and RNase to give a final concentration of 20μgmL^{-1}. The extract was incubated at 37°C for 1 hour. 0.5mL of 5M NaCl was added and 10mL of ethanol was layered on top. The precipitated DNA was removed from the upper ethanol layer using a flame blunted pasteur pipette. The DNA was dissolved into 9mL of TE, 9.9g of caesium chloride and 0.5mL of 10mgmL^{-1} ethidium bromide was added and then centrifuged at 36K for 48 hours in a Beckman 70.1Ti rotor using a Beckman L7 ultracentrifuge. The DNA was viewed under long wave UV light as a fluorescing band and removed by inserting a syringe through the side of the tube. The ethidium bromide was removed from the DNA using isopropanol saturated with caesium chloride. The DNA was then diluted to 5mL and 0.5mL of 5M NaCl and 10mL of ethanol added. This was incubated at -20°C for 30 minutes. The precipitated DNA was then recovered by centrifugation at 10K for 10 minutes in a Sorval type SS34 rotor at 4°C. The pellet was dried at 50°C and resuspended in TE.

2.7.2 LARGE SCALE PREPARATION OF PLASMID DNA

Plasmid DNA was isolated by alkaline lysis from 400mL cultures of cells as described in Sambrook *et al.*, 1989 (pp 1.38-1.39). Cells were harvested by centrifugation and the pellet resuspended in 10mL of *Solution I* (50mM glucose, 25mM Tris-HCl (pH 8.0), 10mM Na_2EDTA). A few crystals of lysozyme were added and incubated at 37°C for 30 minutes. 20mL of *Solution II* (see Section 2.2.2) was added. The mixture was mixed by gently inverting the centrifuge bottle. 15mL of *Solution III* (see Section 2.2.2) was added and the contents mixed by shaking the bottle several times. The bottle was stored on ice for 10 minutes before centrifuging the bacterial lysate to remove the cell debris from the supernatant and

adding 0.6 volume of isopropanol. The bottle was stored at room temperature for 10 minutes before recovering the nucleic acids by centrifugation. The pellet was washed with 70% ethanol and dissolved in TE. The plasmid was purified by the caesium chloride gradient method either using the Beckman 70.1Ti rotor (48K for 24 hours) or the Beckman VTi 80 (70K for twelve hours) rotor. Another method for isolating plasmid DNA was using Promega Wizard™ Maxipreps DNA purification system (A7270) as per the manufacturers instruction. Instead of the suggested vacuum pump, the column was packed, covered with Nescofilm and washed under pressure by using a 50mL syringe (without the needle attached) inserted into a hole in the Nescofilm.

2.7.3 SMALL SCALE PREPARATION OF PLASMID DNA

Plasmid DNA was isolated from 5mL nutrient broth cultures grown overnight in universal bottles. 1.5mL of culture were harvested by centrifugation in a microfuge for 2 minutes at full speed. The supernatant was removed and the pellet resuspended in 100µL of Birnboin solution (see Section 2.2.2) and incubated at room temperature for 5 minutes. 200µL of SDS-NaOH solution (see Section 2.2.2) and inverted to mix. 150µL of 4M sodium acetate pH 6.0 (see Section 2.2.2) was added, the mixture vortexed for 10 seconds and stored on ice for 5 minutes. The mixture was centrifuged at 13 000 rpm and the supernatant transferred to another eppendorf tube. An equal volume of phenol chloroform (see Section 2.2.2) was added and mixed by vortexing. The mixture was centrifuged for 5 minutes and the top layer removed to another tube. 1mL of ethanol was added, vortexed and the tube allowed to stand on ice for 5 minutes. The tube was centrifuged for 10 minutes and the supernatant discarded. 1mL of ethanol was added to the tube taking care not to disturb the pellet and the tube centrifuged again for two minutes keeping the tube

in the same orientation. The supernatant was again discarded and the pellet dried in a 50°C oven. When dried, 100µL of TE-RNase (see Section 2.2.2) was used to resuspend the DNA pellet and the DNA stored at -20°C.

2.7.4 SIZE FRACTIONATION OF ROT34A1 CHROMOSOMAL DNA

Trial digestions of the chromosomal Rot34AI DNA with *Sau*3AI were carried out to find the optimum concentration for partial digestion to produce the correct size fragments (between 23 and 9kb). 2.5µg of chromosomal DNA was diluted with TE and restriction buffer with 2 units of *Sau*3AI to give a final volume of 200µL. This was incubated at 37°C and 20µL samples were removed at various time intervals and stop mixture added. The samples were electrophoresed overnight at 20V in parallel on a 0.7% (w/v) agarose gel, with appropriate λ markers.

The method was scaled up to digest 100µg of chromosomal DNA in 200µL of buffer. The reaction was stopped by heat denaturing the enzyme at 70°C for 10 minutes. The protein was removed by phenol chloroform extraction and the DNA precipitated. In general ethanol precipitation was carried out by addition of 1/10 volume of 4M sodium acetate or 5M NaCl and two volumes of ethanol, mixing and holding at -20°C for thirty minutes before pelletting the DNA by centrifugation. The ethanol was removed by drying the pellet at 55°C and the DNA resuspended in 200µL TE buffer.

A 12mL sucrose gradient of 40-10% was formed by using a gradient mixer. The DNA was layered on top and centrifuged in a Beckman SW41 rotor for 24 hours at 20°C at 22K. A small hole was pierced at the bottom of the tube and the fractions collected drop wise (approximately 400µL) in 1.5mL eppendorf tubes. An aliquot

(10-20μL) of each fraction was removed and 5μL of stop mix added and then electrophoresed in parallel on a 0.7% (w/v) agarose gel with appropriate λ markers overnight at 20V. The appropriate fractions (see Chapter 6) were then pooled, diluted with three volumes with TE and ethanol precipitated. The pellet was resuspended in 5-10μL TE.

2.8 PREPARATION OF COMPETENT CELLS

A universal bottle containing 5mL nutrient broth; 20mM $MgSO_4$ was inoculated with a loop of cells from a freshly grown plate. The cells were grown overnight at 37°C in an orbital shaker and then used to inoculate 200mL of nutrient broth containing 20mM $MgSO_4$. The cells were grown to mid exponential phase (approximately 3-4 hours) at 37°C in an orbital shaker. The cells were harvested in sterile centrifuge pots at 4°C at 10K in a Sorval GSA rotor and washed in 40mL ice cold 75mM $CaCl_2$ containing 15% glycerol. The cells were reharvested at 4°C at 10K in a Sorval GSA rotor and resuspended in 5mL of ice cold 75mM $CaCl_2$ containing 15% glycerol. The cells were aliquoted in 200μL samples and stored at -70°C.

2.9 TRANSFORMATION AND PLATING OF *E. COLI* STRAINS

2.9.1 TRANSFORMATIONS

Transformations were carried out by adding the DNA (1-5μg) to the thawed competent cells and placing the cells on ice for 45 minutes. The cells were then heat shocked for 10 minutes by placing in a 37°C water bath. The cells were added to 5mL of nutrient broth and grown at 37°C for 3-4 hours or overnight before plating on the appropriate media.

2.9.2 REPLICA PLATING

Cells were plated onto the appropriate media and grown at 37°C overnight. Sterile velvet pieces were placed over a wooden block and gently pressed onto the cells. The cells were transferred onto a second plate containing selective agar and finally on to a plate containing the original media.

2.10 PACKAGING DNA

DNA was packaged using the Gigapack II XL packaging extract (Stratagene Ltd. Cambridge, UK) The ligation mixture was added to the "Freeze Thaw" extract and placed on ice. 15µL of "Sonic" extract was then added and the mixture stirred with a glass pasteur pipette avoiding air bubbles. The mixture was centrifuged for 1-2 seconds (using the pulse setting) and incubated at 22°C for two hours. 500µL of SM buffer and 20µL of chloroform were added and mixed gently. The debris was sedimented by centrifugation and the supernatant stored at 4°C.

2.11 TRANSFECTION OF *E. COLI*

A single colony of P2392 or LE392 was added to 10mL of TB broth supplemented with 0.2% (w/v) maltose and 10mM $MgSO_4$ and the cells grown either overnight at 30°C or for six hours at 37°C. The cells were harvested and resuspended in 5mL of sterile 10mM $MgSO_4$. SM buffer was used to dilute the packaged ligation mixture (10^0, 10^{-1}, 10^{-2}).The cells were aliquoted into 200µL samples and 10µL of the packaged material added to the cells. The cells were incubated at 37°C for 20-30 minutes. The top agar was melted and aliquoted into 3mL samples and the cells added. The top agar was then poured onto a dried NZY plate, allowed to set and incubated overnight at 37°C.

2.12 AMPLIFICATION OF LIBRARIES

2.12.1 AMPLIFICATION OF THE PLASMID LIBRARY

Plasmid libraries were amplified by scraping the colonies from the plates using sterile 20% (v/v) glycerol. Half of the stock was used to inoculate 400mL of nutrient broth supplemented with the appropriate antibiotic the other half stored at -20°C. The cells were grown at 37°C overnight in an orbital shaker and the cells harvested by centrifugation. Plasmid DNA was isolated by the large scale method.

2.12.2 AMPLIFICATION OF THE PHAGE LIBRARY

The phage was transfected as previously described but 600µL of cells, 6.5mL of melted top agar and 150mm plates of bottom agar were used. (No more than 300µL of phage volume per 600µL of cells were added). The plates incubated at 37°C for 6-8 hours. The plates were then overlaid with 8-10 mL of SM buffer. The plates were shaken at 4°C overnight on a shaking platform.

2.13 PREPARATION OF ROT 34A1 CELL EXTRACTS

Cells were harvested after overnight growth by centrifugation (9000 rpm for 45 minutes in a Sorval GSA rotor), washed in 10mM potassium phosphate buffer pH 7.0 (see Section 2.2.2) and cell pellet collected by centrifugation (18 500 rpm for 30 minutes in a Sorval SS34 rotor). Cell extracts were prepared by sonication using a MSE Soniprep sonicator. Cells were sonicated by five cycles of one minute sonication at 8 microns with one minute cooling. Cell debris was removed by centrifugation (18 500 rpm at 4°C for 30 minutes in a Sorval SS34 rotor). The supernatant was used immediately.

2.14 ENZYME ASSAYS

2.14.1 *Tfi*I ENDONUCLEASE DETECTION ASSAY

2μL - 10μL of samples were added to λ DNA test mix [0.25μg DNA, 20μL restriction buffer (see Section 2.2.2)]. The mixtures were incubated at 65°C for 1-2 hours. Stop mixture (see Section 2.2.2) was then added to the mixtures and the samples electrophoresed in parallel on a 1% (w/v) agarose gel.

2.14.2 *Tfi*I ENDONUCLEASE ACTIVITY ASSAY

2μL-10μL of the extract was added to an activity assay mixture (2.5μg λ DNA, restriction buffer to 200μL). The assay tube was then incubated at 65°C and 20μL samples removed at various time intervals. 7μL of stop mixture was added to the samples and the samples were electrophoresed in parallel on a 1% (w/v) agarose gel.

1 Unit was defined as the amount of activity capable of total digestion of 1 μg of λ DNA at 65°C in one hour.

2.14.3 *Tfi*I METHYLASE DETECTION ASSAY

(i) Protection Assay

10μL of extracts were incubated with 1μg of pUC18 in methylase buffer (see Section 2.2.2) and 80μM S-adenosyl methionine. Control tubes did not contain S-adenosyl methionine. The tubes were then incubated at 55°C overnight. The DNA was then precipitated with ethanol and 5M NaCl at -20°C and dried at 50°C. The DNA was resuspended in TE containing restriction buffer. *Tfi*I (1-2 units) was then added and the tubes incubated at 65°C for 1 hour. Stop mixture was added to the

samples. The samples were then electrophoresed in parallel on a 1% (w/v) agarose gel.

(ii) The methylation reaction were carried out as above except 2.2μM [^{3}H] S-adenosyl-L-methionine was included in the reaction mixture. Following incubation, samples were removed, adsorbed onto 2.5mm Whatman FR81 paper circles and washed with 3mL volumes of 0.2M $(NH_4)HCO_3$, H_2O and 100% (v/v) ethanol (three washes of each) The circles were dried at room temperature and counted as described earlier.

2.14.4 *Tfi*I METHYLASE ACTIVITY ASSAY

2μL, 4μL and 10μL of sample were incubated with 2.5μg of pUC18 and methylase buffer at 55°C to a final volume of 50μL. At regular time intervals 10μL aliquots were removed and DNA precipitated by ethanol and 5M NaCl at -20°C and dried at 50°C. The DNA was resuspended in TE and restriction buffer. *Tfi*I (1-2 Units) was added and the reaction mixture incubated at 65°C for 1 hour. The DNA was electrophoresed in parallel on 1% (w/v) agarose.

1 Unit is defined as the activity required to protect 1μg of pUC18 DNA in 1 hour at 55°C against *Tfi*I endonuclease activity.

2.14.5 GAPDH ACTIVITY ASSAY

20μL of sample was added to 100μL PGK salt mixture, 100μL 100mM 3PGA, 50μL 10mM NADH, 3μL PGK (yeast 4.5U) and made up to 1000μL with distilled water. 50μL of 40mM ATP pH 7.5 was added to start the reaction. A decrease in absorbance was measured at 340nm.

2.15 DETERMINATION OF PROTEIN CONCENTRATION

Protein concentrations were determined by the Bradford assay (Bradford 1976) using BioRad Protein Assay reagent. Bovine serum albumin was used as a standard to generate standard curves.

2.16 CHROMATOGRAPHY

2.16.1 PHOSPHOCELLULOSE CHROMATOGRAPHY

2.16.1.1 PREPARATION OF PHOSPHOCELLULOSE

10g of Whatman Cellulose Phosphate (P11) was added to 1L of 0.5M NaOH incubated at room temperature for 5 minutes before decanting. The gel was washed with distilled water until the filtrate pH fell below 11. The alkali treated phosphocellulose was then stirred into 0.5M HCl and incubated at room temperature for 5 minutes. The supernatant was decanted and washed in a vacuum funnel until the filtrate pH was above 3.0. The slurry was added to 50x phosphate buffer (see Section 2.2.2) and incubated in this buffer for a minimum of 10 minutes before packing the cellulose phosphate into a Pharmacia XK26 column. The column was pre-equilibrated with 10mM potassium phosphate buffer (containing 0.7mM β-mercaptoethanol)pH 7.0.

2.16.1.2 LOW PRESSURE PHOSPHOCELLULOSE CHROMATOGRAPHY

The cell extract was loaded on the phosphocellulose column using a Pharmacia P1 peristaltic pump at a constant flow rate of 2-3mLmin^{-1}. The column was washed with 10mM potassium phosphate buffer (containing 0.7mM β-mercaptoethanol) pH 7.0 and the bound protein eluted with a linear salt gradient (0-1M NaCl in 10mM potassium phosphate buffer (containing 0.7mM β-mercaptoethanol) pH 7.0). The

column was then washed with 10mM potassium phosphate buffer (containing 0.7mM β-mercaptoethanol and 1M NaCl) pH 7.0. 8mL fractions were collected and assayed for *Tfi*I endonuclease and methylase activity. The active fractions were dialysed overnight at 4°C in a 50% (v/v) glycerol 10mM potassium phosphate buffer (containing 0.7mM β-mercaptoethanol and 200mM NaCl). The samples were then stored at -20°C.

2.16.2 FAST PROTEIN LIQUID GEL FILTRATION CHROMATOGRAPHY

2.16.2.1 HILOAD SUPERDEX 200 PREP GRADE™ CHROMATOGRAPHY

5mL samples of *Tfi*I extracts were applied to a Pharmacia HiLoad 16/60 Superdex 200 prep grade™ column. The column was pre-equilibrated with 10mM potassium phosphate buffer pH 7.0, (containing 0.7mM β-mercaptoethanol and 200mM NaCl) at a flow rate of 0.5mLmin^{-1} and 4mL fractions were collected. Protein was monitored at 280nm and alternate fractions were assayed for *Tfi*I endonuclease activity. The active fractions were pooled and dialysed overnight at 4°C in 50% (v/v) glycerol;10mM potassium phosphate buffer pH 7.0 (containing 0.7mM β-mercaptoethanol).

2.16.2.2 SUPEROSE 6™ CHROMATOGRAPHY

200μL samples of *Tfi*I extracts were applied to a Pharmacia Superose 6 HR 10/30 prepacked column. The column was pre-equilibrated with 20mM potassium phosphate (containing 0.7mM β-mercaptoethanol) buffer pH 7.0 at a flow rate of 0.5 mLmin^{-1} and 1mL fractions were collected. Protein was monitored at 280nm and fractions were assayed for *Tfi*I endonuclease activity.

2.16.2.3 SUPERDEX 75™ CHROMATOGRAPHY

200μL samples of *Tfi*I extracts were applied to a Pharmacia Superdex 75 HR 10/30 prepacked column. The column was pre-equilibrated with 10mM potassium phosphate buffer pH 7.0.(containing 0.7mM β-mercaptoethanol and 200mM NaCl) at a flow rate of 0.5mLmin^{-1} and 1mL fractions were collected. Protein was monitored at 280nm and fractions were assayed for *Tfi*I endonuclease activity.

2.16.3 ADSORPTION CHROMATOGRAPHY

2.16.3.1 LOW PRESSURE ION EXCHANGE CHROMATOGRAPHY

*Tfi*I extracts were loaded onto a Pharmacia HiLoad Q Sepharose High Performance™ or a Pharmacia HiLoad S Sepharose High Performance™ pre-equilibrated with 10mM potassium phosphate (containing 0.7mM β-mercaptoethanol) buffer pH 7.0 at a flow rate of 2.5mLmin^{-1}. The column was washed with the above buffer and *Tfi*I endonuclease eluted with a linear salt gradient (0-1M NaCl with 10mM potassium phosphate (containing 0.7mM β-mercaptoethanol) buffer pH 7.0). 8mL fractions were collected and assayed for enzyme activity. The active fractions were pooled and dialysed overnight at 4°C in 50% (v/v) glycerol;10mM potassium phosphate, (containing 0.7mM β-mercaptoethanol and 200mM NaCl) buffer pH 7.0.

2.16.3.2 FAST PROTEIN LIQUID ION EXCHANGE CHROMATOGRAPHY

*Tfi*I samples were applied to a Pharmacia Mono Q HR 5/5 column pre-equilibrated with 10mM potassium phosphate (containing 0.7mM β mercaptoethanol) buffer pH 7.0 at a flow rate of 1mLmin^{-1}. The column was washed in the above buffer and *Tfi*I endonuclease eluted with a step gradient of 0-150mM NaCl and 150mM NaCl-

225mM NaCl in 10mM potassium phosphate (containing 0.7mM β–mercaptoethanol) buffer pH 7.0. 1mL fractions were collected and assayed for enzyme activity. Protein was monitored at 280nm.

2.16.3.3 AFFINITY CHROMATOGRAPHY (HEPARIN)

A Bio-Rad Econo-Pac Heparin Cartridge™ was pre-equilibrated as follows;

(i) 10mM potassium phosphate (containing 0.7mM β-mercaptoethanol) buffer pH 7.0 was pumped though the cartridge at 2mLmin^{-1} for 10 minutes.

(ii) The cartridge was washed with 10mM potassium phosphate (containing 0.7mM β-mercaptoethanol and 1M NaCl) buffer at 5mLmin^{-1} for 10 minutes.

(iii) The cartridge was inverted and washed with 10mM potassium phosphate (containing 0.7mM β-mercaptoethanol) buffer pH 7.0 for 10 minutes at 5mLmin^{-1}.

(iv) The cartridge was inverted and the flow rate reduced to 2mLmin^{-1} and washed with 10mM potassium phosphate (containing 0.7mM β-mercaptoethanol) buffer pH 7.0 .

The sample was applied to the cartridge and the active fractions eluted with a linear salt gradient (0-1M NaCl with 10mM potassium phosphate (containing 0.7mM β-mercaptoethanol) buffer pH 7.0). The fractions (2.5mL) were assayed for *Tfi*I endonuclease activity and the active fractions pooled and dialysed overnight at 4°C in a 50% (v/v) glycerol 10mM potassium phosphate (containing 0.7mM β-mercaptoethanol) buffer pH 7.0.

2.17 PROTEIN CONCENTRATION

2.17.1 PROTEIN PRECIPITATION

Protein samples were precipitated using a solution of 15% (v/v) trichloroacetic acid (TCA). TCA solution was added to the protein solution to give a final concentration of 3.75% (v/v). The samples were placed on ice for 30 minutes and then pelleted at 10 000 rpm for 10 minutes. The supernatant was removed and the pellet washed with ethanol:ether (1:1). The pellets were dried at 55°C. The pellets were resuspended in protein solubilisation solution (see Section 2.2.2).

2.17.2 PROTEIN CONCENTRATION BY ULTRAFILTRATION

Proteins were concentrated using Centricon-10 concentrators and Microcon-10 micro concentrators (supplied by Amicon) and used as per the manufacturer's instructions unless stated otherwise.

2.18 SDS POLYACRYLAMIDE GEL ELECTROPHORESIS

Analytical gel electrophoresis was carried out in 1.5mm thick 12.5% (v/v) polyacrylamide mini gels as described by Sambrooks *et al.*, 1989 using a Höefer Tall Mighty Small 11cm Vertical Slab Unit SE 280™. Samples were concentrated and suspended in protein solubilisation solution and were heated at 90°C for 3 minutes prior to loading. The gels were electrophoresed at 75mA for 2-3 hours or 10mA overnight.

2.19 SILVER STAINING

Silver staining of SDS PAGE gels was achieved using the BioRad Silver Stain Plus™ (Catalogue No. 161-0449). Gels were fixed and stained as detailed in the manufacturers instructions. Gels were stored in 5% (v/v) acetic acid.

2.20 INTERNAL SEQUENCE DETERMINATION

An aliquot of Mono Q purified *Tfi*I extract was concentrated by ultrafiltration with the addition of 2x SDS loading buffer. This sample was gel purified, blotted and digested by trypsin. The peptide fragments were separated and internal sequence of some fragments were obtained. This data was obtained by Dr. J. Hsuan of the Ludwig Institute, Department of Biochemistry and Molecular Biology, UCL. The sequence alignments shown were also determined by Dr. J. Hsuan.

2.21 THIN LAYER CHROMATOGRAPHY

20μg of lambda DNA was incubated with the 40μL of phosphocellulose purified methylase, methylase buffer, 0.8mM S-adenosyl methionine and 2.2μM [3H] S-adenosyl-L-methionine at 55°C overnight. An aliquot was removed and tested for protection against *Tfi*I. The DNA was then ethanol precipitated and resuspended in 48μL of distilled water. 5μL aliquots were removed in duplicate to test for incorporation.

The DNA was denatured by boiling for 5 minutes. 2μL of calf intestinal alkaline phosphatase, 5μL of phosphodiesterase and CAP buffer (see Section 2.2.2) were added and the mixture incubated at 37°C for 20 hours. The samples were then desalted using a 1mL Sephadex G10 column. The fractions collected were assayed for incorporation of radioactivity. The samples showing the highest counts were concentrated by freeze drying.

The samples were resuspended in 10μL of water and spotted on pre coated TLC sheets (polyethyleneimine cellulose on polyester with 254nm fluorescent indicator supplied by Aldrich Chemical Company), with 20nM (10μL) standards of N^6 methyl 2' deoxyadenosine, 5 methyl 2' deoycytidine, 2' deoxyadenosine, 2'

deoxycytidine, 2' deoxyguanosine and thymidine. Two different solvent systems were used (i) 80:20 ethanol:water (v/v) and (ii) 66:33:1 isobutyric acid:water: ammonium hydroxide (v/v). The plates were visualised using a 254nm UV light box. Spots corresponding to the N^6 methyl 2'-deoxyadenosine and 5-methyl 2'-deoxycytidine were scraped off and analysed for radioactivity.

50µg of Rot34A1 chromosomal DNA was hydrolysed and analysed as above.

2.22 EFFECT OF TEMPERATURE ON ENDONUCLEASE ACTIVITY

100µL of *Tfi*I was incubated at various temperatures and aliquots removed at different time (1, 5, 10, 20, 30, 60 minutes) and stored on ice. Activity assays were then carried out.

2.23 pH PROFILE

Buffers of a range of pH (see Section 2.2.2) were prepared and incubated with pUC18 and 2 units of *Tfi*I and incubated at 70°C for 1 hour. Activity assays at the different pHs were carried out.

2.24 EFFECT OF SALT

A range of buffers with 0-700mM NaCl were prepared and incubated at 70°C with 0.5µg of pUC18 and 2 units of *Tfi*I for 1 hour. The reactions were then terminated with stop mixture and the DNA electrophoresed in parallel lanes on a 1% (w/v) agarose gel. Activity assays at different salt concentrations were carried out.

2.25 EDTA AND METAL ION STUDY

Buffers with a range of Na_2EDTA (0-10mM) concentrations were prepared and incubated with 0.5µg of pUC18 and 2 units of *Tfi*I and incubated at 70°C for 1

hour. The reactions were then terminated with stop mixture and the DNA electrophoresed in parallel lanes on a 1% (w/v) agarose gel.

10x stock solutions of Mn^{2+}, Co^{2+}, Cu^{2+} and Zn^{2+} were prepared in 500mM Tris HCl (pH 7.5). 0.5µg of pUC18 was incubated with TE, each of the above buffers and 2 U of *Tfi*I. Samples were incubated at 65°C for 1 hour. Stop mixture was then added and the samples electrophoresed.

2.26 STAR ACTIVITY STUDIES

Various conditions and chemicals were used to induce star activity:

(i) High glycerol concentration

Different concentrations of glycerol in TE buffer was added to pUC18 and restriction buffer. *Tfi*I was added and the samples incubated at 65°C for 1 hour. Samples were removed, stop mixture added and electrophoresed in parallel on 1% (w/v) agarose gels.

(ii) Low ionic strength (<25mM)

Tris buffers (pH 7.5) of 2mM, 5mM and 10mM were incubated with 2 U of *Tfi*I, 0.5µg of pUC18 and 5mM $MgCl_2$ for 1 hour at 65°C. The samples were removed, stop mixture added and samples electrophoresed.

Chapter 3. Purification of *Tfi*I restriction endonuclease

3.1 INTRODUCTION

This chapter describes the purification of *Tfi*I endonuclease. The purified protein was used to obtain internal sequence data. Such data can be used in the design of oligonucleotide probes with which to screen genomic libraries of the parent organism. Partial purification of the enzyme was achieved by a number of sequential chromatographic methods. The choice of chromatographic steps was influenced by the known properties of Class II endonucleases, i.e. the two fold symmetry exhibited by endonucleases and the dimeric form of active endonucleases exist in aqueous solutions (Moldrich, 1982). The endonuclease was found to co-purify with Rot34A1 glyceraldehyde 3-phosphate dehydrogenase (GAPDH). Purity was determined by SDS PAGE and by internal sequence analysis.

3.2 PRODUCTION OF CELL BIOMASS

Thermus filiformis strain Rot34A1 was grown as described in Section 2.1.1. Cell yields were approximately 0.75-0.9gL^{-1} of culture. Temperature variations of +/- 5°C did not alter the yield of biomass, although the T_{opt} for *T. filiformis* is 65°C (Hudson *et al.*, 1984). As Rot34A1 is a filamentous organism, it was expected that agitation would effect cell growth. It was observed that stirrer speeds greater than 200 rpm decreased cell biomass.

3.3 PREPARATION OF CELL EXTRACT

Cells were harvested and disintegrated by sonication. A colour change was observed (from a bright yellow paste to a translucent yellow solution) during sonication and was indicative of complete cell lysis. Cell debris was removed by centrifugation, the

supernatant collected and stored at 4°C prior to adsorption chromatography. Restriction endonuclease activity and protein concentration was measured as described in Chapter 2 (Section 2.14.1 and 2.15).

3.4 PHOSPHOCELLULOSE CHROMATOGRAPHY

Whatman phosphocellulose (P11) is a cation exchanger based on phosphate derivatised fibrous cellulose. The functionality of phosphocellulose enables it to act like a pseudo-affinity medium for DNA-binding proteins and enzymes. Cellular nucleic acids (DNA and RNA) do not bind to phosphocellulose, thus avoiding the need for prior removal of these species before chromatography (Ward *et al.*, 1991). Steps designed to remove nucleic acids such as streptomycin and ammonium sulphate precipitations are thus eliminated. Phosphocellulose also offers another advantage as fewer dialysis steps are necessary (Greene *et al.*, 1978).

The cell extract was diluted with 10mM potassium phosphate buffer (containing 0.7mM β–mercaptoethanol) pH 7.0 after collection in order to reduce its ionic strength and applied to a phosphocellulose column as described in Section 2.16.1.2. Fractions were assayed for endonuclease activity and protein elution followed by a UV monitor attached to a chart recorder.

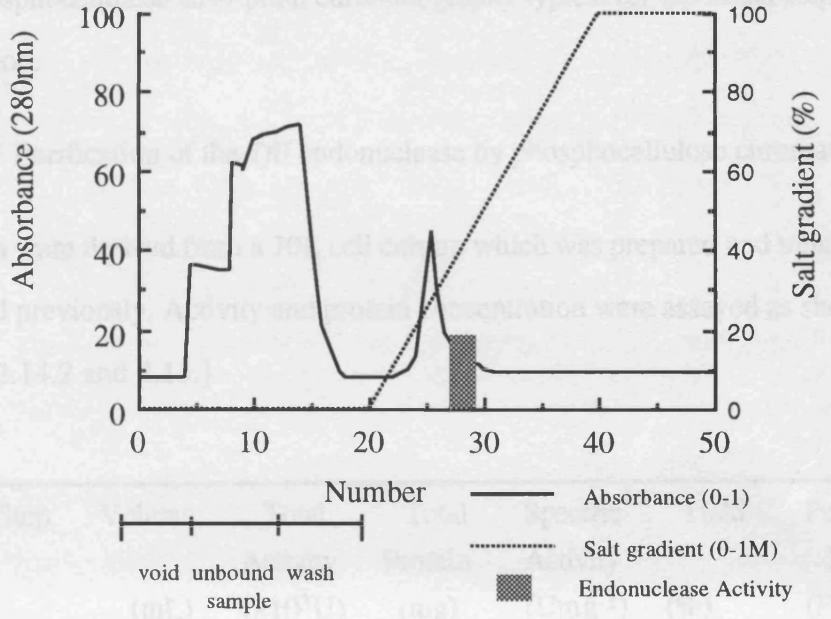
Figure 3.1a shows a typical protein and salt gradient profile obtained for phosphocellulose chromatography. Figure 3.1b shows the endonuclease activity profile for phosphocellulose chromatography. The enzyme eluted between 350mM and 450mM NaCl. Enzyme activity was measured as described (Section 2.14.2) after dialysis of the combined active fraction. The apparent yield for the endonuclease after phosphocellulose chromatography was usually greater than the activity assayed in the crude extract.

Figure 3.1a describes the profile obtained from a 3L culture preparation. Cells were harvested, cell extract prepared and 50mL of cell extract loaded onto a 40mL phosphocellulose column as described previously. The void [Number 0-4] was collected separately as was the unbound sample [Numbers 5-10]. The column was washed with 80mL of 10mM potassium phosphate buffer (containing 0.7mM β-mercaptoethanol) pH 7.0 [Numbers 11-20]. To elute the endonuclease a salt gradient (120mL 0-1M NaCl) was applied [Number 20 -35] and 8mL fractions were collected. The column then washed with 10mM potassium phosphate buffer (containing 0.7mM β -mercaptoethanol and 1M NaCl) pH 7.0 [Number 35-50].

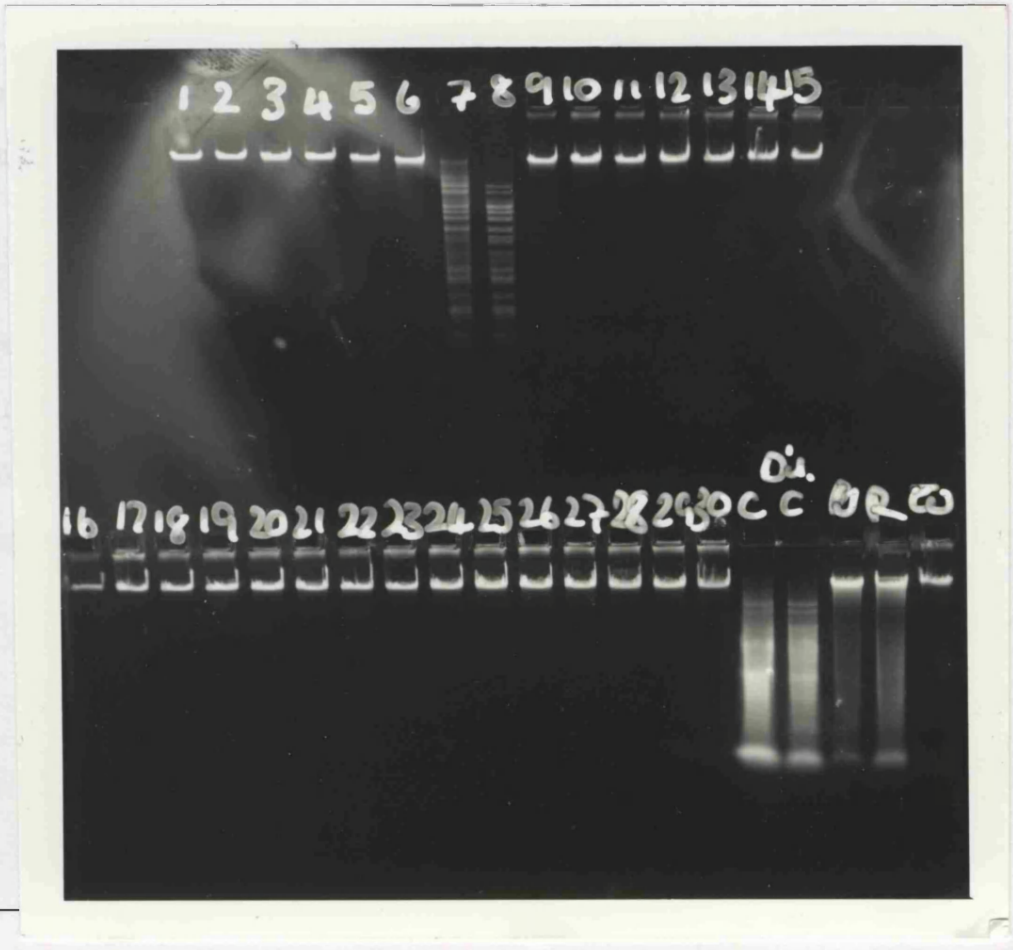
Figure 3.1b shows the photograph of the endonuclease detection assay. Each lane represents a 5μL of each gradient fraction[1-25]| and each high salt wash fraction [25-30] assayed and electrophoresed as described in Section 2.14.1. The crude extract [C] the diluted crude extract [Dil. C], the void sample [V], unbound sample [R] and low salt wash [W] were also treated as above. The figure shows the endonuclease was eluted between fraction 7 and 8 which corresponds to 350mM -450mM NaCl concentration.

Figure 3.1 Elution profile of Rot34AI crude extract from phosphocellulose chromatography

(a)



(b)



The pooled fractions from phosphocellulose elution were separated from the yellow pigment and free from the *Tfi*I methylase enzyme (Chapter 5). Table 3.1 shows data from phosphocellulose adsorption chromatography typical for the initial steps of purification.

Table 3.1 Purification of the *Tfi*I endonuclease by phosphocellulose chromatography

[The data were derived from a 10L cell culture which was prepared and treated as described previously. Activity and protein concentration were assayed as shown in Section 2.14.2 and 2.15.]

Step	Volume (mL)	Total Activity ($x10^3U$)	Total Protein (mg)	Specific Activity (Umg^{-1})	Yield (%)	Purification Factor (Fold)
crude extract	25	10	200	50	100	1
P11	15	15	9.9	1515	150	30

The lower apparent activity of the endonuclease in the crude extract suggested that the activity may be inhibited, possibly by the binding of other proteins to substrate DNA in the assay. This may prevent the enzyme binding to the nucleic acids.

3.5 ION EXCHANGE CHROMATOGRAPHY

Following phosphocellulose chromatography, ion exchange chromatography was performed. Both S-Sepharose and Q-Sepharose chromatography gels were assessed. Activities and recovery were calculated as described previously. In 10mM potassium phosphate buffer (containing 0.7mM β–mercaptoethanol) pH 7.0, the endonuclease bound to both columns. Table 3.2 shows the purification data derived from Q-Sepharose and S-Sepharose chromatography.

Table 3.2 Purification of the *Tfi*I endonuclease by Q-chromatography and S-chromatography

[The data were generated from a comparative test of Q and S-Sepharose chromatography. 5mL of P11 extract diluted and loaded onto pre-equilibrated Q-Sepharose and S-Sepharose columns independently. The endonuclease was eluted as described previously.]

Step (Fraction)	Volume (mL)	Total Activity ($x10^3U$)	Total Protein (mg)	Specific Activity (Umg^{-1})	Yield (%)	Purification Factor (Fold)
P11	5	2.5	1	2 500	100	1
Q Sepharose	4	2.0	0.4	5 000	80	2
S Sepharose	4	1.0	0.5	2 000	30	0.8

The recovery achieved with S-Sepharose was significantly lower than with Q Sepharose (Table 3.2). Q-Sepharose chromatography was performed as described in

Section 2.16.3.1. The dialysed phosphocellulose fraction was diluted four fold before application to the column. Fractions were collected and assayed for activity. The endonuclease eluted between 300mM and 500mM NaCl. Figure 3.2a shows a typical protein profile obtained from Q-Sepharose chromatography. Figure 3.2b shows the *Tfi*I endonuclease detection assay of fractions from Q-Sepharose chromatography. The purification from Q Sepharose is shown in Table 3.3.

Table 3.3 Purification of the *Tfi*I endonuclease by Q-Sepharose chromatography

[The data were derived from a 10L cell culture prepared and applied to a Q-Sepharose column after phosphocellulose chromatography (described in earlier sections). Activity and protein concentration were assayed as shown in 2.14.2 and 2.15.].

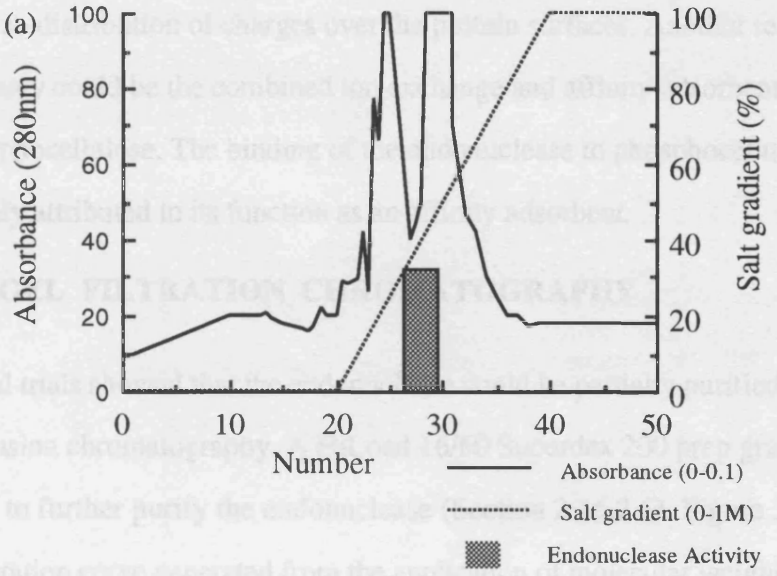
Step	Volume (mL)	Total Activity ($x10^3$U)	Total Protein (mg)	Specific Activity (Umg^{-1})	Yield (%)	Purification Factor (Fold)
P11	15	15	9.9	1515	100	1
Q Sepharose	10	10	6	1667	66	1.1

Approximately 66% of the activity was recovered from Q-Sepharose chromatography although only a small increase in specific activity was obtained from this step. The endonuclease bound more tightly to the S-Sepharose column (data not shown) than the Q-Sepharose column. However the phosphocellulose column (which is also a cationic exchanger) and the Q-Sepharose bound the endonuclease with approximately equal affinity, eluting the active fractions between 300 and 500mM.

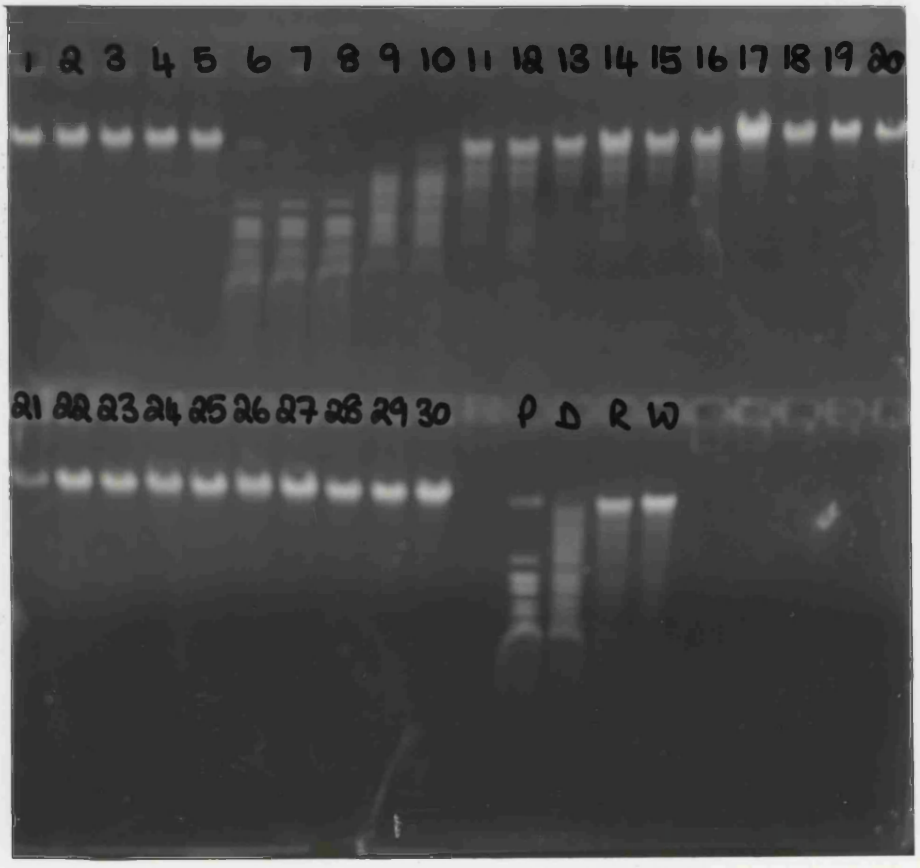
The data describes the profile obtained from a 3L culture preparation. After phosphocellulose chromatography, 10mL of P11 extract was diluted and loaded onto a 50mL Pharmacia HiLoad Q-Sepharose column as described previously (2.15.3.1). The void [50mL, Number 0- 4] was collected separately as was the unbound sample [50mL, Numbers 5-9]. The column was washed with 100mL of 10mM potassium phosphate buffer (containing 0.7mM β -mercaptoethanol) pH 7.0 [Numbers 10 -20]. To elute the endonuclease a salt gradient (160mL, 0-1M NaCl) was applied [Number 20 -40] and 8mL fractions were collected. The column then washed with 80mL of 10mM potassium phosphate buffer (containing 0.7mM β -mercaptoethanol and 1M NaCl) pH 7.0 [Number 41-50].

Figure 3.2b shows a photograph of an *Tfi*I endonuclease detection assay achieved for Q chromatography. Each lane represents a 5μL of each gradient fraction [1- 20] and each high salt wash fraction [21- 30] assayed and electrophoresed as described in Section 2.14.1. The P11 extract [P], the diluted P11 extract [D], the unbound sample [R] and low salt wash [W] were also treated as above (Void sample not shown). The figure shows the endonuclease was eluted between fraction 6 and 10 which corresponds to 300mM -500mM NaCl concentration.

Figure 3.2 Elution profile from Q-chromatography



(b)



The binding of the endonuclease to both cationic and anionic gels in 10mM potassium phosphate buffer (containing 0.7mM β-mercaptoethanol) pH 7.0 could be the result of other non-electrostatic interactions, (e.g., hydrophobic and van der Waal forces) or the uneven distribution of charges over the protein surfaces. Another reason for this anomaly could be the combined ion exchange and affinity adsorbent properties of phosphocellulose. The binding of the endonuclease to phosphocellulose could be largely attributed to its function as an affinity adsorbent.

3.6 GEL FILTRATION CHROMATOGRAPHY

Initial trials showed that the endonuclease could be partially purified using size exclusion chromatography. A HiLoad 16/60 Superdex 200 prep grade column was used to further purify the endonuclease (Section 2.16.2.1). Figure 3.3 shows the calibration curve generated from the application of molecular weight standards to the column.

Aliquots were applied to the column and the elution profile shown below (Figures 3.4a and b) is typical of the separation achieved. The majority of the endonuclease eluted between fractions 21 and 23, though activity was detected in the adjacent fractions. Activity and protein concentration were measured as previously described.

The majority of endonuclease activity eluted between 65 kDa and 77 kDa with the peak activity at approximately 71kDa (Figure 3.3). Table 3.4 shows the purification of the endonuclease by HiLoad Superdex 200 chromatography. This step was not as efficient as Q Sepharose chromatography as less than 50% of the activity was recovered and the purification factor was not significantly greater than the previous step. However the protein elution profile (Figure 3.4a) showed that the method was successful in separating the endonuclease from larger proteins.

Figure 3.3 Calibration curve for Superdex 200 chromatography
[The figure shows the calibration curve generated from the application of 1mg of Dextran, β amylase, bovine serum amylase, carbonic anhydrase and alcohol dehydrogenase to a pre-equilibrated HiLoad 16/60 Superdex 200 column with 10mM potassium phosphate buffer (containing 0.7mM β–mercaptoethanol and 200mM NaCl) pH7.0. The column was washed at a flow rate of 0.5mLmin^{-1} and protein absorbance was monitored at 280nm. The corresponding endonuclease fractions are also shown.]

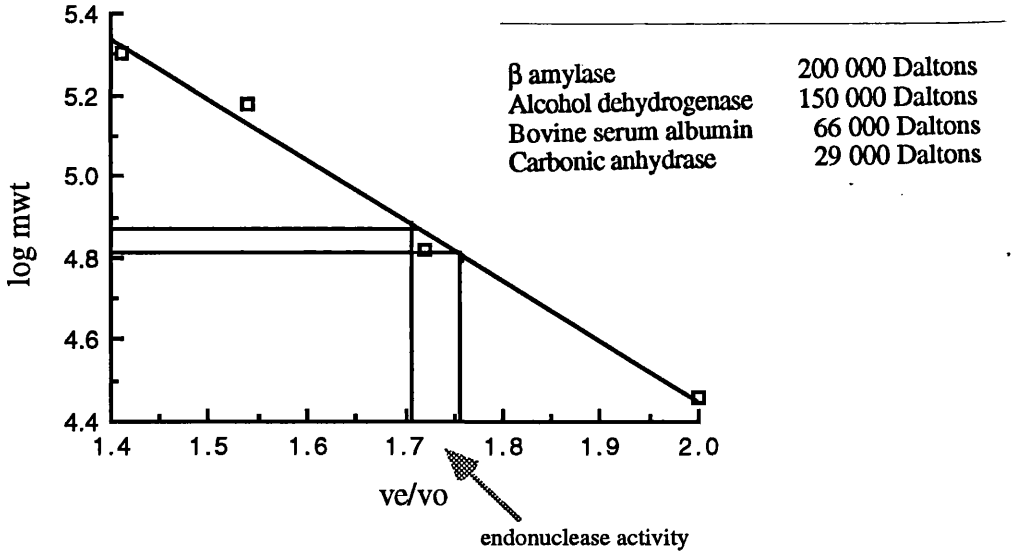


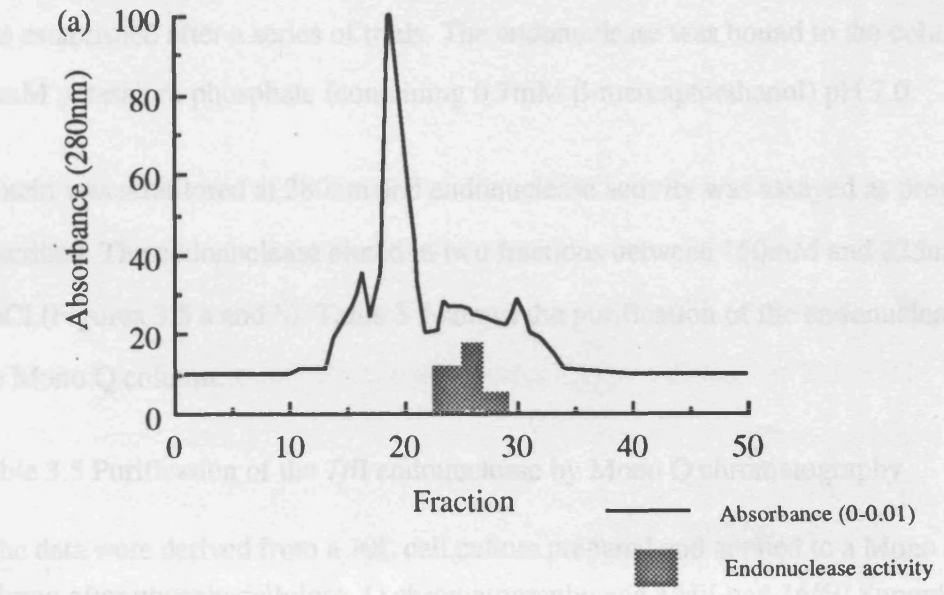
Table 3.4 Purification of the *Tfi*I endonuclease by gel filtration chromatography
[The data were derived from a 10L cell culture prepared and applied to a HiLoad 16/60 Superdex 200 prep grade column after phosphocellulose and Q chromatography (described in earlier sections). Activity and protein concentration were assayed as shown in 2.14.2 and 2.15.]

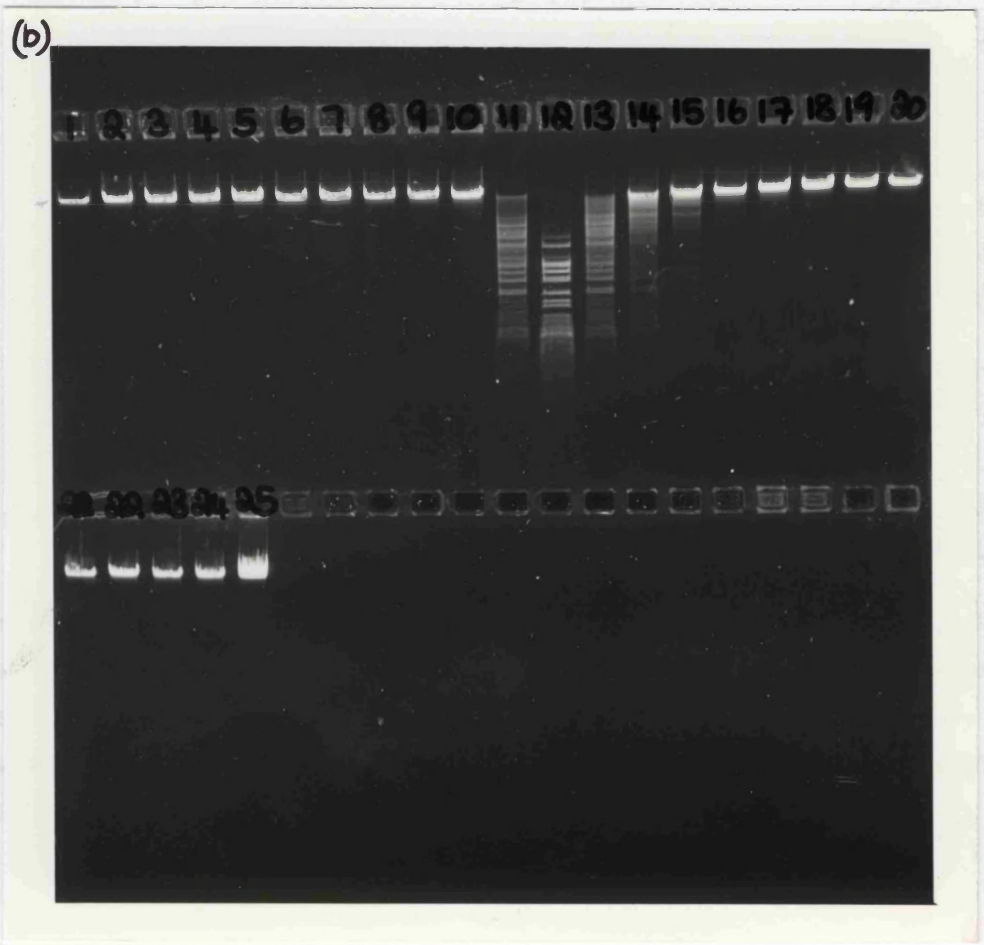
Step	Volume (mL)	Total Activity (x10^{3}U)	Total Protein (mg)	Specific Activity (Umg^{-1})	Yield (%)	Purification Factor (Fold)
Q Sepharose	10	10	6	1667	100	1
Superdex 200	16	4.8	2.4	2000	48	1.1

The figure 3.4a shows the protein profile for HiLoad 16/60 Superdex 200 chromatography. 5mL samples were applied to a pre-equilibrated column (with 10mM potassium phosphate buffer (containing 0.7mM β-mercaptoethanol and 200mM NaCl) pH 7.0 at flow rate of 0.5mLmin^{-1}. 4mL fractions were collected.

Figure 3.4b shows a photograph of a *Tfi*I endonuclease detection assay. The figure shows the endonuclease detection assay for HiLoad 16/60 Superdex 200 chromatography. Each lane represents aliquots (2.5μL) of two fraction (from 3.4a) assayed and electrophoresed as described in Section 2.14.1. The figure shows the endonuclease was eluted between fraction 21 and 28.

Figure 3.4 Elution profile for dialysed active Q chromatography fractions from HiLoad 16/60 Superdex 200 chromatography





3.7 MONO Q CHROMATOGRAPHY

Mono Q chromatography was performed as detailed in Section 2.16.3.2. The protocol was established after a series of trials. The endonuclease was bound to the column in 10mM potassium phosphate (containing 0.7mM β-mercaptoethanol) pH 7.0.

Protein was monitored at 280nm and endonuclease activity was assayed as previously described. The endonuclease eluted in two fractions between 150mM and 225mM NaCl (Figures 3.5 a and b). Table 3.5 shows the purification of the endonuclease from the Mono Q column.

Table 3.5 Purification of the *Tfi*I endonuclease by Mono Q chromatography

[The data were derived from a 10L cell culture prepared and applied to a Mono Q column after phosphocellulose, Q chromatography and a HiLoad 16/60 Superdex 200 prep grade column (described in earlier sections). Activity and protein concentration were assayed as shown in 2.14.2 and 2.15.]

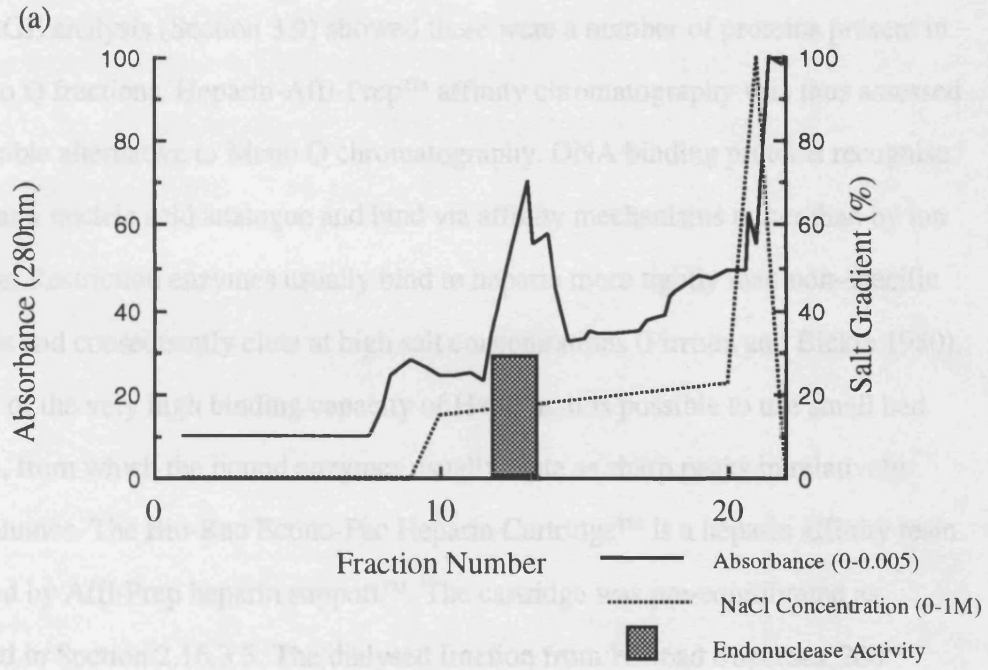
Step	Volume (mL)	Total Activity ($x10^3U$)	Total Protein (mg)	Specific Activity (Umg^{-1})	Yield (%)	Purification Factor (Fold)
Superdex 200	5	3	1.8	1667	100	1
Mono Q	2	1	1.2	5000	33	3

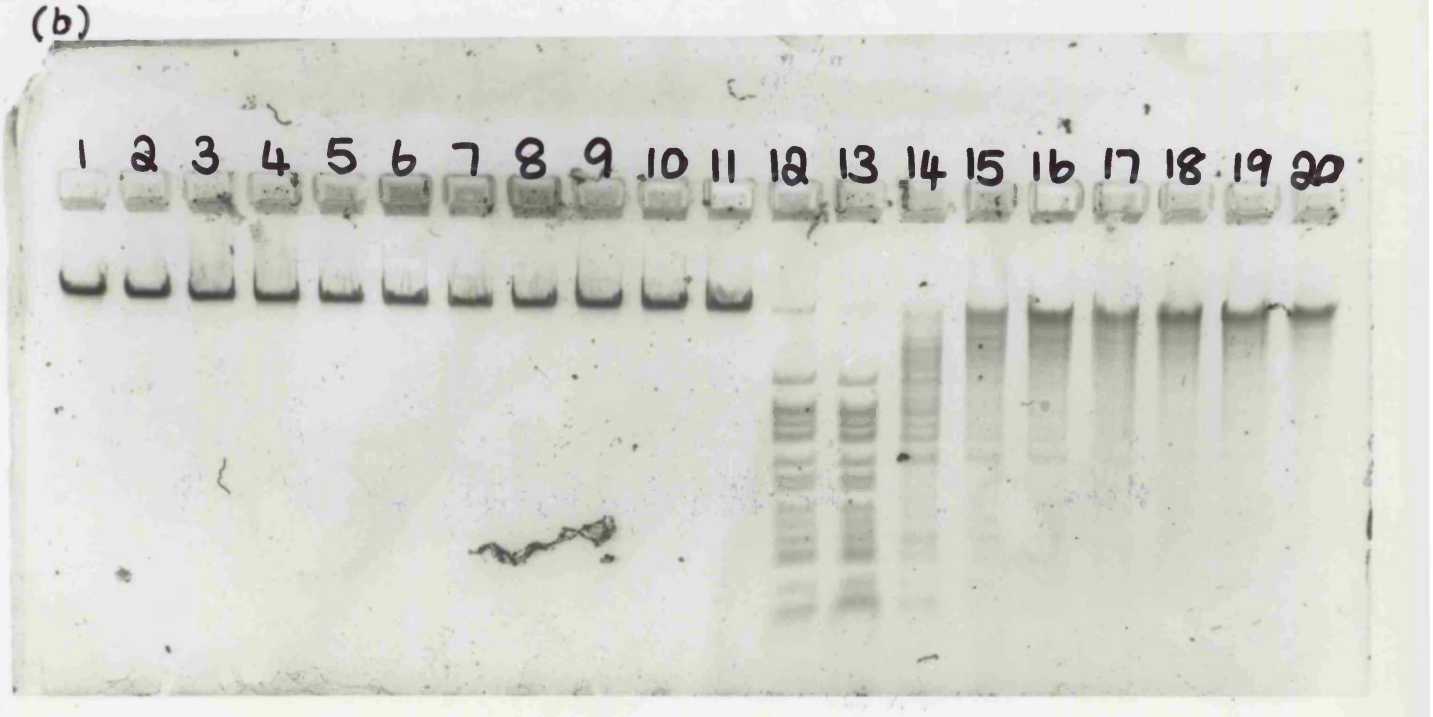
The data in Table 3.5 indicates that only 33% of the activity loaded onto the column was recovered from this step. However the purification factor was greater than that of the earlier steps. SDS-PAGE analysis of the fractions showed that there were six protein bands present in the Mono Q active fraction. These results are discussed in greater detail in Section 3.9.

The figure (3.5a) describes the profile obtain from the application of 1mL of the dialysed Heparin fraction (Section 3.8) to a Mono Q column. The sample was diluted to 5mL, and 1mL fractions were collected. The void [Number 1[was collected separately as was the unbound sample [Numbers 2-6]. The column was washed with 4mL of 10mM potassium phosphate buffer (containing 0.7mM β -mercaptoethanol) pH 7.0 [Numbers 7-9]. To elute the endonuclease a salt gradient (10mL 150-225mM NaCl) was applied [Number 10-20].The salt gradient was then increased to 100% -1M NaCl [21] and decreased to 0% -0M NaCl [22].

Figure 3.5b shows a photograph of an *Tfi*I endonuclease detection assay achieved for Mono Q chromatography. The figures shows the endonuclease detection assay. Each lane represents a 5μL of each fraction [1-20] (Number 1-20, Figure 3.5a) assayed and electrophoresed as described in Section 2.14.1. The figure shows the majority of the endonuclease was eluted between fraction 12 and 13 which corresponds to 145mM and 155mM NaCl concentration.

Figure 3.5 Elution profile from Mono Q chromatography





3.8 HEPARIN-AFFI-PREP™ AFFINITY CHROMATOGRAPHY

SDS-PAGE analysis (Section 3.9) showed there were a number of proteins present in the Mono Q fractions. Heparin-Affi-Prep™ affinity chromatography was thus assessed as a possible alternative to Mono Q chromatography. DNA binding proteins recognise heparin as a nucleic acid analogue and bind via affinity mechanisms rather than by ion exchange. Restriction enzymes usually bind to heparin more tightly than non-specific nucleases and consequently elute at high salt concentrations (Pirrotta and Bickle 1980). Because of the very high binding capacity of Heparin, it is possible to use small bed volumes, from which the bound enzymes usually elute as sharp peaks in relatively small volumes. The Bio-Rad Econo-Pac Heparin Cartridge™ is a heparin affinity resin supported by Affi-Prep heparin support™. The cartridge was pre-equilibrated as described in Section 2.16.3.3. The dialysed fraction from HiLoad Superdex 200 chromatography was diluted 1 in 4 with 10mM potassium phosphate buffer (containing 0.7mM β–mercaptoethanol) pH 7.0. and applied to the cartridge. After various trials the endonuclease was eluted using a 0-500mM NaCl gradient.

Figures 3.6 a and b show the protein and endonuclease activity profiles obtained from Heparin-Affi-Prep™ chromatography. Although only 20% of the activity was recovered with this step there was approximately 7 fold purification (Table 3.6).

Table 3.6 Purification of the *Tfi*I endonuclease by Heparin-Affi-Prep™ affinity chromatography

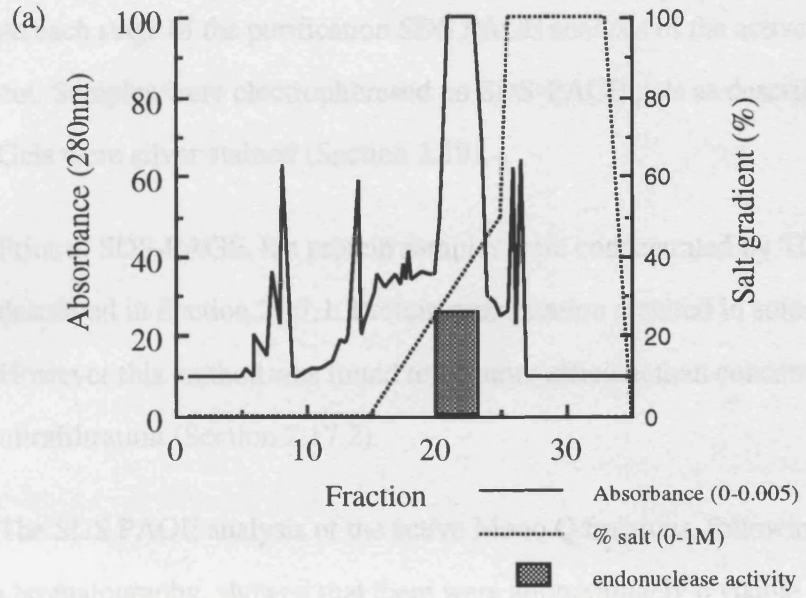
[The data were derived from a 10L cell culture preparation. Samples were treated as described in the text and activity and protein concentration were determined as previously shown (Section 2.14.2 and 2.15).]

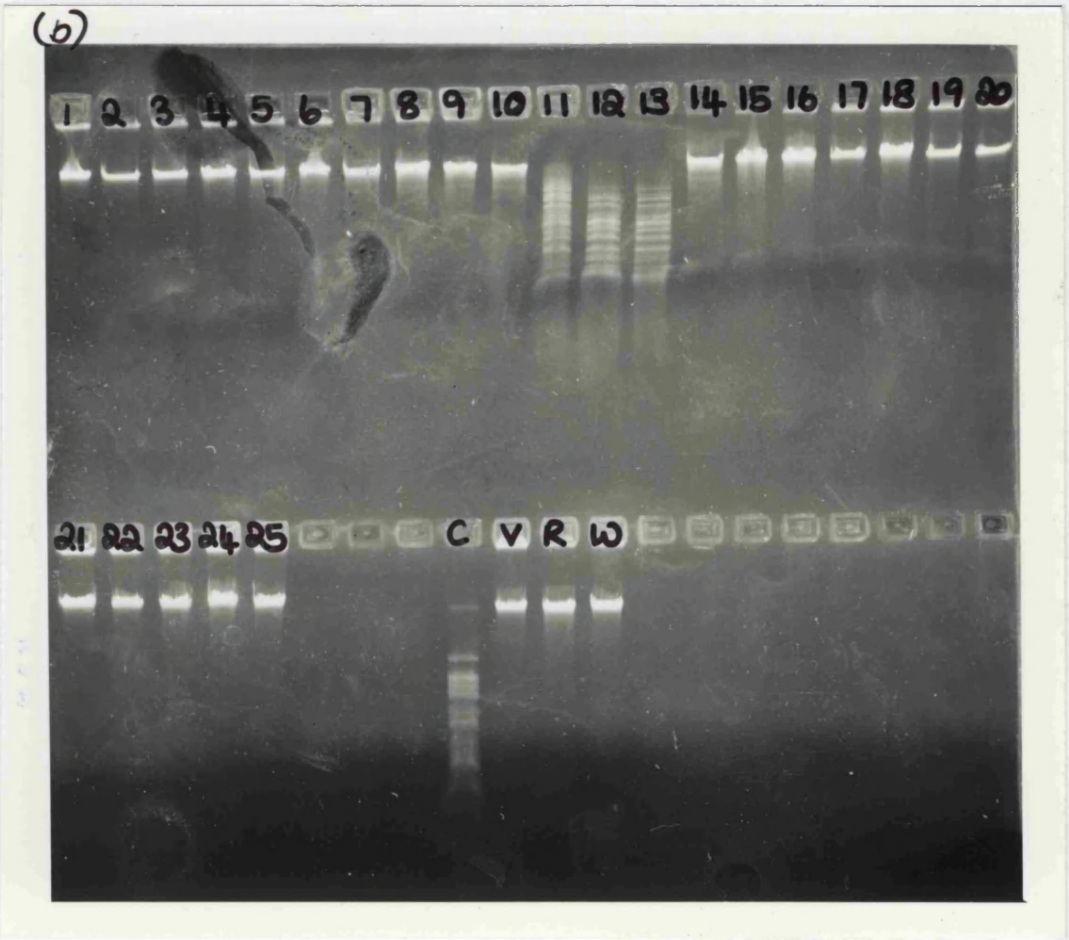
Step	Volume (mL)	Total Activity ($x10^3U$)	Total Protein (mg)	Specific Activity (Umg^{-1})	Yield (%)	Purification Factor (Fold)
Superdex 200	16	4.8	2.4	2000	100	1
Heparin Affi-Prep™	5	1.0	0.07	14286	20	7

The figure describes the profile obtain from the application of 5mL of the dialysed gel filtration fraction to a BioRad Heparin Econo-Pac Cartridge™ column. The sample was diluted to 20mL, and 2.5mL fractions were collected. The void [Number 1- 2] was collected separately as was the unbound sample [Numbers 3-10]. The column was washed with 12.5 mL of 10mM potassium phosphate buffer (containing 0.7mM β -mercaptoethanol) pH 7.0 [Numbers 10-14]. To elute the endonuclease a salt gradient (50mL 0-500mM NaCl) was applied [Number 15-25]. The column then washed with 10mM potassium phosphate buffer (containing 0.7mM β-mercaptoethanol and 1M NaCl) pH 7.0 [Number 26-35].

Figure 3.6 b shows the photograph of the endonuclease detection assay. Each lane represents 5μL of each gradient fraction [1-20] and each high salt wash fraction [21-25] assayed and electrophoresed as described in Section 2.14.1. The crude extract [C], the void sample [V], unbound sample [R] and low salt wash [W] were also treated as above. The figures shows the endonuclease was eluted between fraction 10 and 13 which corresponds to 250mM and 325mM NaCl concentration.

Figure 3.6 Elution profile from Heparin-Affi-Prep™ affinity chromatography





3.9 SDS-PAGE ANALYSIS OF ENDONUCLEASE FRACTIONS

At each stage of the purification SDS PAGE analysis of the active fractions was carried out. Samples were electrophoresed on SDS-PAGE gels as described in Section 2.18. Gels were silver stained (Section 2.19).

Prior to SDS-PAGE, the protein samples were concentrated by TCA precipitation as described in Section 2.17.1. Protein precipitation resulted in some loss of protein. However this method was found to be more efficient than concentration by ultrafiltration (Section 2.17.2).

The SDS PAGE analysis of the active Mono Q fractions, following gel filtration chromatography, showed that there were approximately 6 visible bands (data not shown) upon silver staining. To determine the band corresponding to the endonuclease activity, two further experiments involving elution of active protein samples were performed. Samples from the active Mono Q fractions were concentrated by ultrafiltration (Section 2.17.2) and electrophoresed in duplicate on 10% (v/v) SDS gels. After electrophoresis the gels were cut in half, one half washed in Triton X-100 (two washes of 30 minutes each, followed by one 10 minute wash with distilled water) and the other half of the gel silver-stained. The gels were aligned and the Triton X-100 treated gels were treated by one or other of the procedures described below:

(i) Bands were excised and sliced into tiny fragments. The fragments were incubated with restriction buffer and λ DNA at 65°C overnight. The reactions were stopped by addition of Stop mixture and the samples were electrophoresed on a 1% (w/v) agarose gel to detect digestion.

(ii) Bands were isolated and extracted from the gel using the Höefer GE 200 SixPac Gel Eluter™ at 100V for two hours as described in the manual (Höefer GE 200 SixPac Gel Eluter™ Manual). The resulting sample was incubated with restriction buffer and λ DNA at 65°C overnight. The reactions were stopped by the addition of stop mixture and electrophoresed as described previously.

No activity was detected from either treatment. When the samples recovered using the gel elution system (ii) were re-electrophoresed on SDS gels it was found that very little protein was recovered with this system. The failure of the gel elution protocol was attributed to insufficient protein loaded onto the gel initially and loss of protein during the concentration step and elution steps.

After Heparin-Affi-Prep™ affinity chromatography, the dialysed sample was loaded on the Mono Q column and eluted as described in Section 3.6. Figure 3.7 shows a silver stained SDS-PAGE gel of protein fractions at the different stages of the purification. The SDS PAGE gels showed a 37kDa protein band, free of the contaminating proteins (within the detection levels of the staining system). Figure 3.8 shows a silver stained SDS PAGE gel of fractions from the final stages of purification. Table 3.7 shows the complete purification table for the *Tfi*I endonuclease.

Figure 3.7 Silver stained SDS-PAGE gel showing the initial stages of purification

[from left, lane 1, crude extract (2µL); lane 2, molecular weight markers; lane 3, diluted (2µL, 1 in 10) crude extract; lane 4, diluted (2µL, 1 in 20) crude extract; lane 5, P11 fraction (1mL); lane 6, Q-Sepharose fraction (1mL); lane 7, Superdex 200 fraction (1mL); lane 8, Heparin-Affi-Prep™ fraction (1mL); lane 9, Heparin-Affi-Prep™ fraction (1mL); lane 10, molecular weight markers]

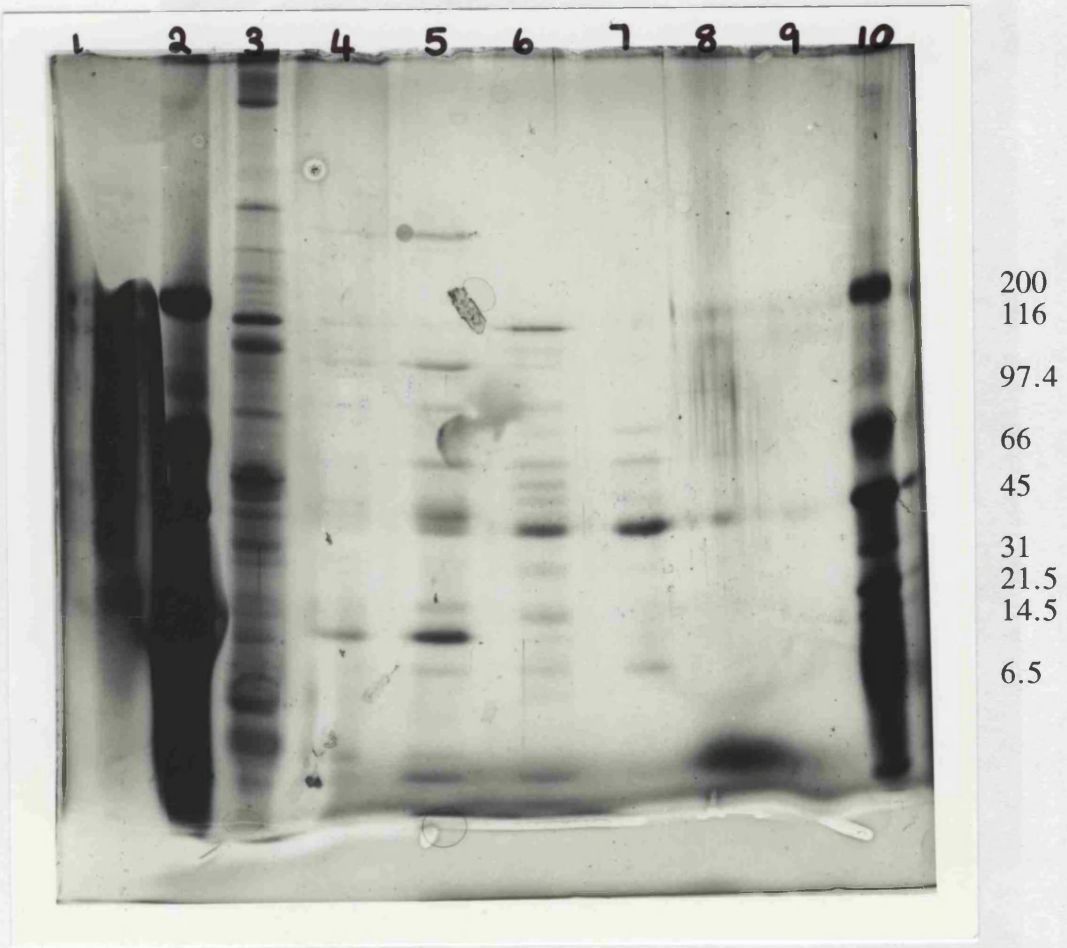


Figure 3.8 Silver stained SDS PAGE gel showing the final stages of purification [from left, lane 1, pre-stained molecular weight markers; lane 2, molecular weight marker; lane 3, P11 fraction (1mL); lane 4, Q-Sepharose fraction (1mL); lane 5, Superdex 200 fraction (1mL); lane 6, Heparin-Affi-Prep™ fraction (2mL); lane 7, Mono Q fraction (2mL)]

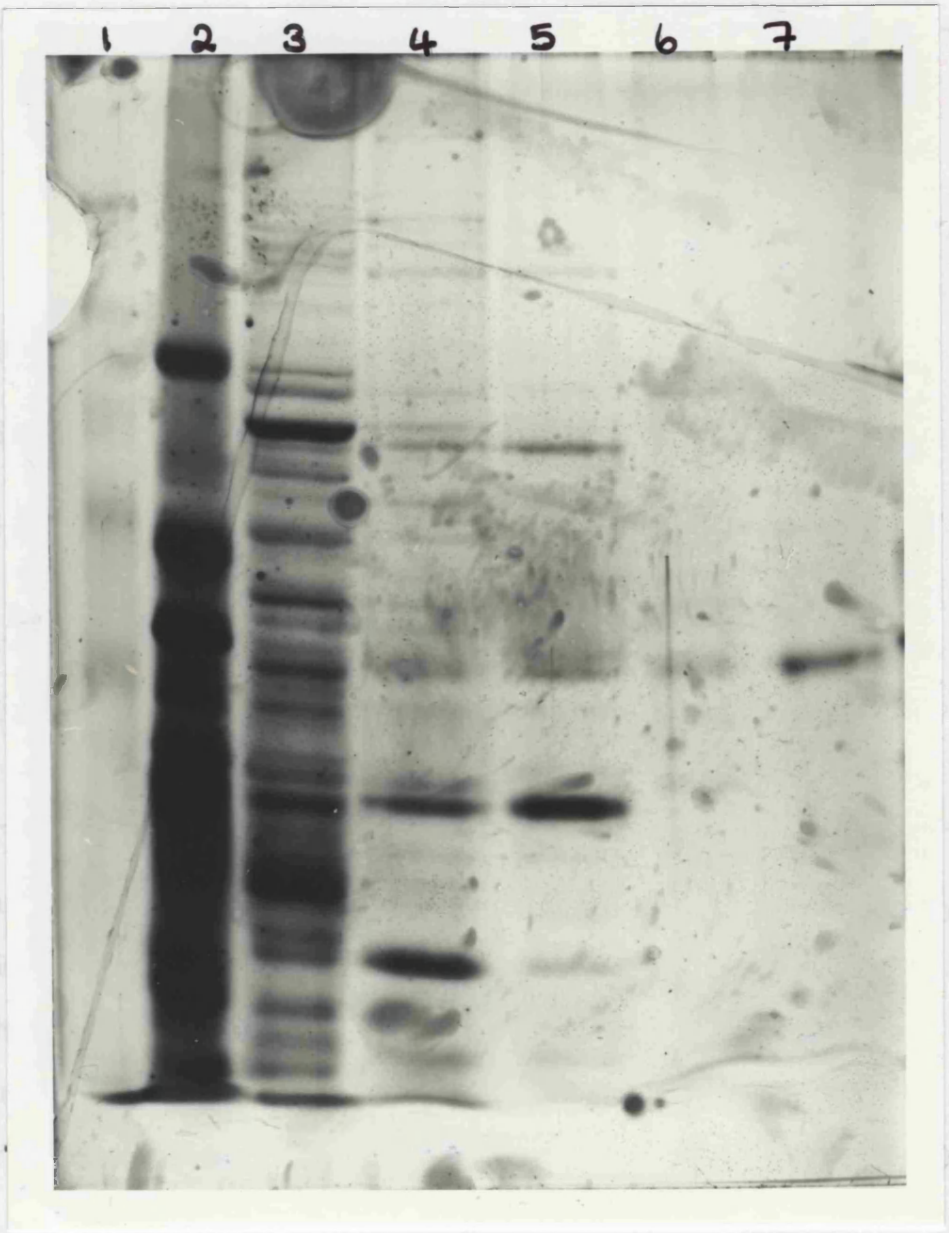


Table 3.7 Purification of *Tfi*I endonuclease

[The data were derived from a 10L cell culture preparation. Samples were treated as described in text and activity and protein concentration were determined as previously shown (Section 2.14.2 and 2.15).]

Step	Volume (mL)	Total Activity ($x10^3U$)	Total Protein (mg)	Specific Activity (Umg^{-1})	Yield (%)	Purification Factor (Fold)
crude extract	25	10	200	50	100	1
P11	15	15	9.9	1515	150	30
Q-Sepharose	10	10	6	1667	100	33.3
Superdex 200	16	4.8	2.4	2000	48	40
Heparin-Affi-Prep™	5	1.0	0.07	14286	10	286
Mono Q	2	0.8	0.018	44444	8	889

The data from Table 3.7 shows that the different chromatographic steps were highly variable in the recovery of activity and purification of the endonuclease. Phosphocellulose was used as an initial step in the purification. This resulted in an apparent 30 fold increase in specific activity, although the activity recovered was greater than present in the crude extract. This was attributed to non-specific inhibition

of the endonuclease, via protein-protein or protein-DNA interactions. Q-Sepharose chromatography resulted in a further 1.3 fold increase, with 33% loss of activity. However gel filtration chromatography showed a 1.2 fold increase but with over 50% loss of activity. Heparin-Sepharose chromatography resulted in a further 7 fold increase in specific activity although the activity recovered was just over 21%. SDS-PAGE analysis (Figure 3.7) showed only one major band of 37kDa with one or two larger proteins. The final step (Mono Q chromatography) showed a 3 fold increase in specific activity with only a modest loss in activity. The active fractions from this column appeared to be homogenous based upon SDS-PAGE analysis (Figure 3.8, Lane 7).

3.10 INTERNAL SEQUENCE DETERMINATION

Internal sequence data was determined by Dr. J. Hsaun as described in Section 2.20. To prepare the protein for sequencing it was necessary to concentrate the protein. Neither TCA precipitation nor ultrafiltration were efficient methods for concentrating the protein. TCA precipitation was found to be too harsh with low protein concentration solutions and ultrafiltration resulted in the protein binding irreversibly to the filter. Fractions from the Mono Q column were concentrated as described in Section 2.17.2 by ultrafiltration with the addition of 2x SDS loading buffer (Section 2.2.2) to prevent loss of the protein. Following gel electrophoresis the protein was blotted onto nitrocellulose and the band digested with trypsin. The peptide fragments were separated and prepared for sequencing.

Part of the internal sequence of the 37kDa protein was determined. The internal amino acid sequence data suggested that the internal sequence showed significant homology with *T. aquaticus* GAPDH (glyceraldehyde 3-phosphate dehydrogenase). Figure 3.9 shows the computer alignment of the two sequences. The Heparin-Sepharose affinity chromatography fraction and the Mono Q chromatography fraction were subsequently

assayed for GAPDH activity (Section 2.14.5) and GAPDH activity was detected. It was concluded that the protein isolated and sequenced was not *Tfi*I endonuclease and that *Tfi*I endonuclease co-migrates with the GAPDH. It is known that GAPDH is present at high levels in many cells. Attempts to separate the endonuclease from the GAPDH protein are discussed in Chapter 4 and Chapter 7.

Figure 3.9 Computer generated alignment of the *T. aquaticus* D-glyceraldehyde 3-phosphate dehydrogenase and the protein isolated from Rot34AI
[Data prepared by Dr J Hsuan. Each letter denotes an amino acid. Each peptide was analysed by HPLC analysis. The top lane represents the known amino acid sequence for *T. aquaticus* D-glyceraldehyde 3-phosphate dehydrogenase and the bottom lane the internal amino acid sequence obtained from the protein isolated from Rot34A1. | indicates where the sequences are identical.]

```
1     MKVGINGFGR IGRQVFRILH SRGVEVALIN DLTDNKTLAH LLKYDSIYHR

51    FPGEVAYDDQ YLYVDGKAIR ATAVKDPKEI PWAEAGVGVV IESTGVFTDA
      |||||:||::  ||||||    |||:||| :: |    |||:| ||||||||||
      FPGEVGYDEE NLYVDGK    ATAIKDPAQL P VXLGVGLV IESTGVFTDA

101   DKAKAHLEGG AKKVIITAPA KGEDITIVMG VNHEAYDPSR HHIISNASCT
      :|                    |||||||:| |||| ||||
      EK                    GEDITIVLG VNHEQYDPS

151   TNSLAPVMKV LEEAFGVEKA LMTTVHSYTN DQRLLDLPHK DLRRARAAAI
               | |:|||||||                              ||||
               V LDEAFGVEK                              AAAI

201   NIIPTTTGAA KATALVLPSL KGRFDGMALR VPTATGSISD ITALLKREVT
      |||||||||                        || ||||||
      NIIPTTTGA                     E XPTPTGSISD

251   AEEVNAALKA AAEGGPLKGIL AYTEDEIVLQ DIVMDPHSSI VDAKLTKALG

301   NMVKVFAWYD NEWGYANRVA DLVELVLRKG V
          ||| || ||      || |||||:
          VFAPYD NE      VA DLVELI
```

Chapter 4 Characterisation of *Tfi*I endonuclease

4.1 INTRODUCTION

Class II restriction modification systems share similar physical and functional characteristics (see Introduction). The systems are composed of two separate proteins, the endonuclease and the methylase. Whereas the methylases show primary amino acid homology, the endonucleases do not show any amino acid sequence homology, except in the case of certain isoschizomers (see section 1.7.2 and 1.8.1). The elucidation of the crystal structures have illustrated tertiary structural similarities between certain endonucleases. They usually exist as homodimers, though some endonucleases are known to form tetrameric structures (Wilson, 1991).

Class II endonucleases usually recognise DNA sites which exhibit rotational symmetry. It has been demonstrated that under non-standard conditions, restriction endonucleases are capable of cleaving sequences which are similar but not identical to their defined recognition sequence. This altered specificity has been termed "star" activity (Polisky *et al.*, 1975).

Class II restriction endonucleases require Mg^{2+} for cleavage of DNA. Other divalent metal ions can be substituted as the co-factor in the reaction mixture. However the substitution can result in a relaxed specificity (Hsu and Berg, 1978).

Mesophilic enzymes are typically relatively labile molecules compared with their thermophilic homologues and thus are adversely affected when exposed to extreme conditions. Thermophilic enzymes, in general, show greater tolerance to such conditions.

This chapter discusses the characteristics of *Tfi*I endonuclease. The native form of *Tfi*I endonuclease was determined by gel filtration chromatography and SDS PAGE analysis. The effects of temperature, pH, salt and detergents on *Tfi*I endonuclease are discussed.

The influence of various divalent metal ions on *Tfi*I endonuclease activity and specificity were also investigated. "Star" activity was induced by physical conditions and the influence of these conditions and the ability of *Tfi*I endonuclease to cleave single stranded DNA are discussed in this chapter.

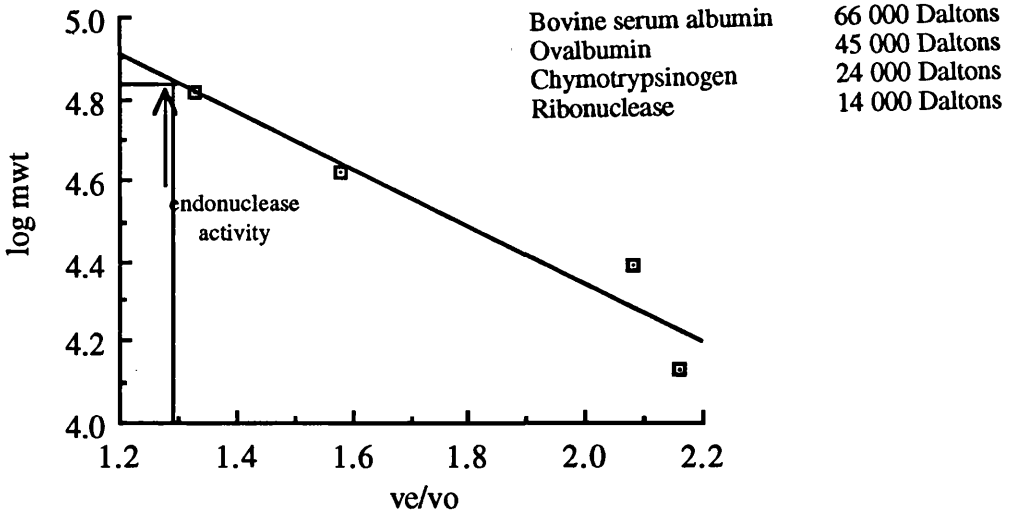
4.2 GEL FILTRATION CHROMATOGRAPHY

4.2.1 ESTIMATION OF MOLECULAR WEIGHT OF *Tfi*I ENDONUCLEASE

200µL of the purified *Tfi*I extract from Mono Q chromatography (Section 3.8) was applied to a Pharmacia Superdex 75 column. The endonuclease was eluted as described in Section 2.16.2.3. The figure below shows the calibration curve for Superdex 75.

Figure 4.1 Calibration curve for Superdex 75 chromatography

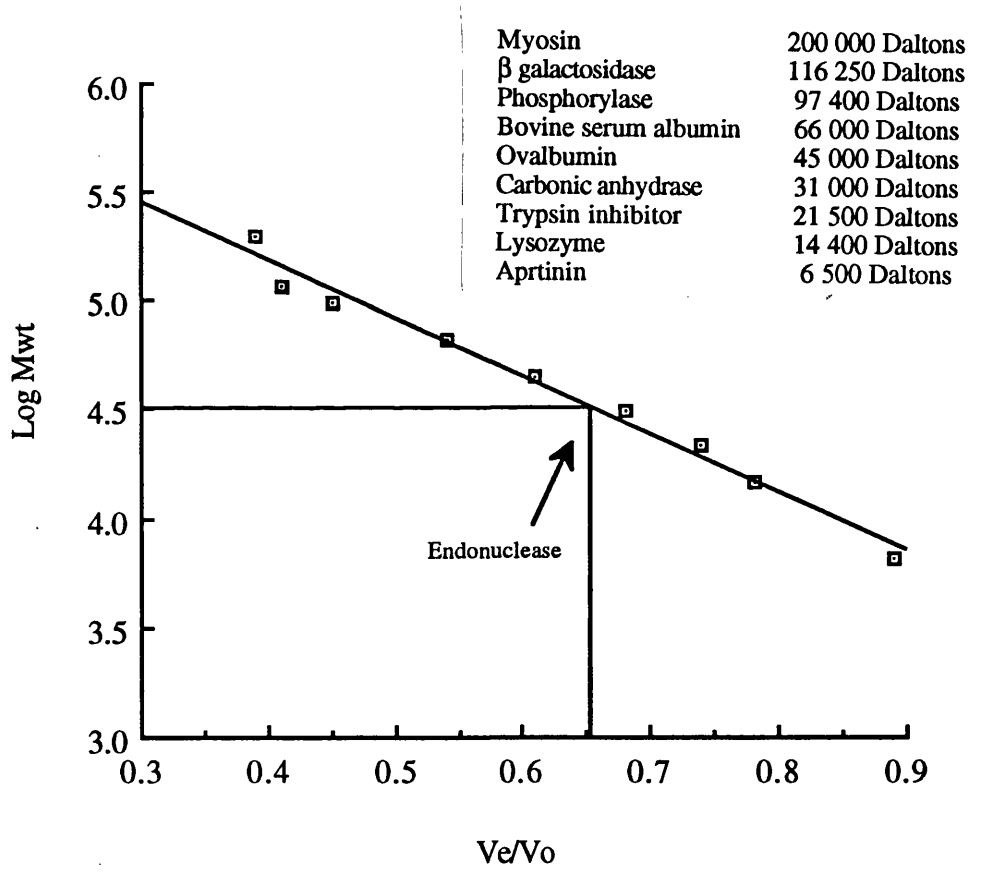
[The calibration curve was generated by the application of 1mg aliquots of Dextran, BSA, ovalbumin, chymotrypsinogen and ribonuclease to a pre-equilibrated Pharmacia Superdex 75 column. The proteins were eluted by washing the column at a flow-rate of 0.5mLmin^{-1} and protein absorbance was monitored at 280nm. The corresponding endonuclease fractions are also shown.]



*Tfi*I endonuclease was found to elute at approximately 72 kDa (Figure 4.1). The SDS PAGE analysis (Figure 4.2) showed that *Tfi*I endonuclease has a subunit molecular weight value of 37kDa. This indicates *Tfi*I endonuclease exists as a dimer.

Figure 4.2 Calibration curve for SDS PAGE analysis of *Tfi*I endonuclease

[The calibration curve was generated from the SDS PAGE analysis of Mono Q purified *Tfi*I endonuclease electrophoresed in parallel with BioRad Mid Range Molecular Weight Markers™.]



4.2.2 SEPARATION OF THE *Tfi*I ENDONUCLEASE AND ROT34A1 GAPDH ENZYME

Data from the purification and internal sequence determination of *Tfi*I endonuclease (see Section 3.10) showed the *Tfi*I endonuclease co-purified with the GAPDH of Rot34A1. The protein profiles from Superdex 75 chromatography showed a second peak in the void volume. It was postulated this protein may be the Rot34AI GAPDH protein. Gel filtration chromatography was used to separate the two enzymes. Superose 6 chromatography separates a broader range of proteins (between 5000 and $5x10^6$ MW) than Superdex 75 chromatography. The purified Mono Q fraction (Section 3.8) was dialysed overnight against 20mM potassium phosphate buffer pH 7.0 containing 50% (v/v) glycerol and 0.7mM β mercaptoethanol at 4°C. The protein profile (Figure 4.3) of the Superose 6 chromatography showed two peaks, corresponding to approximately 75 kDa and 160 kDa (Figure 4.4).

Figure 4.3 Elution profile for dialysed Mono Q fractions from Superose 6 chromatography

[The figure shows the protein profile for Superose 6 chromatography. 200µL of the dialysed Mono Q fraction was applied to a pre-equilibrated column (with 20mM potassium phosphate buffer (containing 0.7mM β-mercaptoethanol) pH 7.0) at a flow-rate of 0.5mLmin^{-1}. 1mL fractions were collected.]

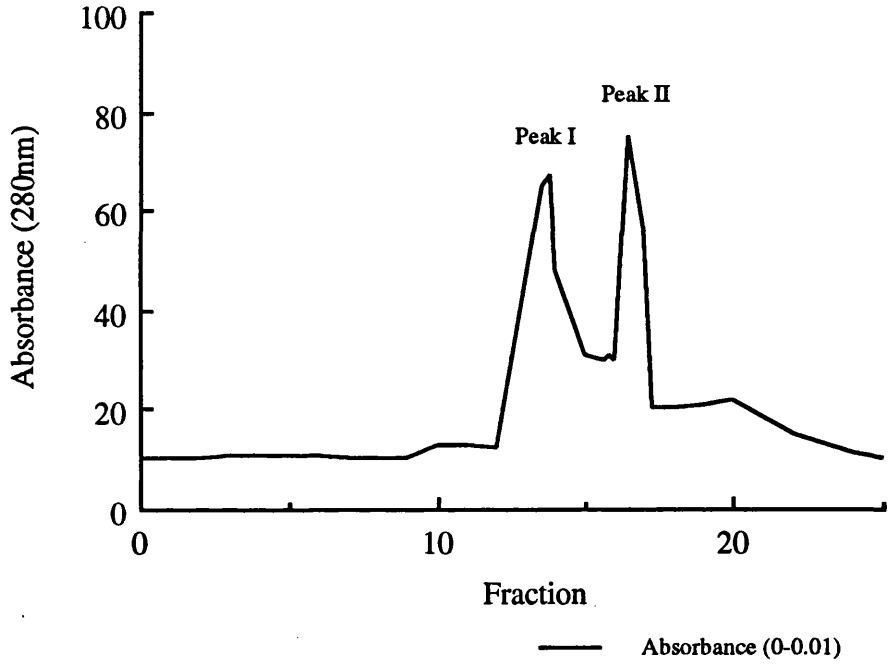
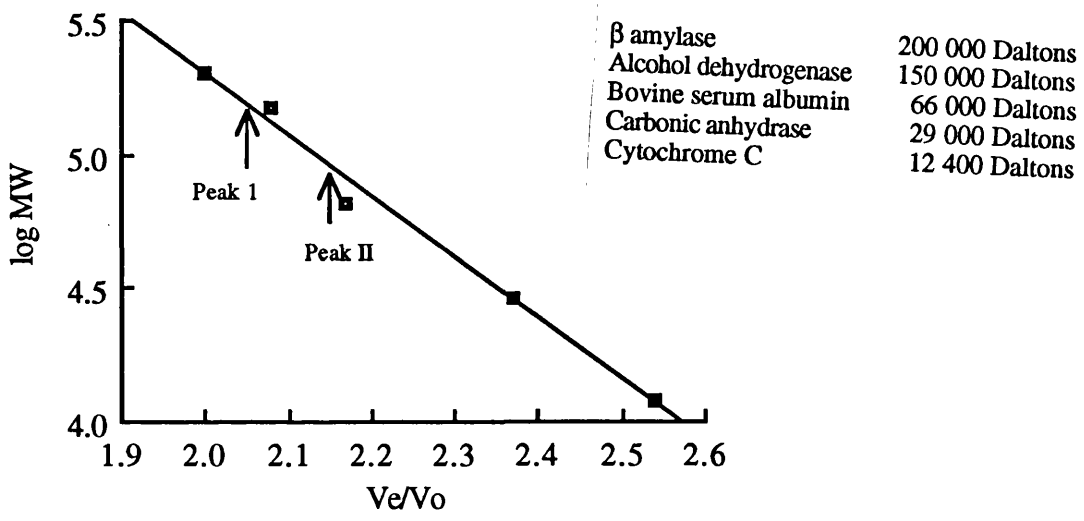


Figure 4.4 Calibration curve for Superose 6 chromatography

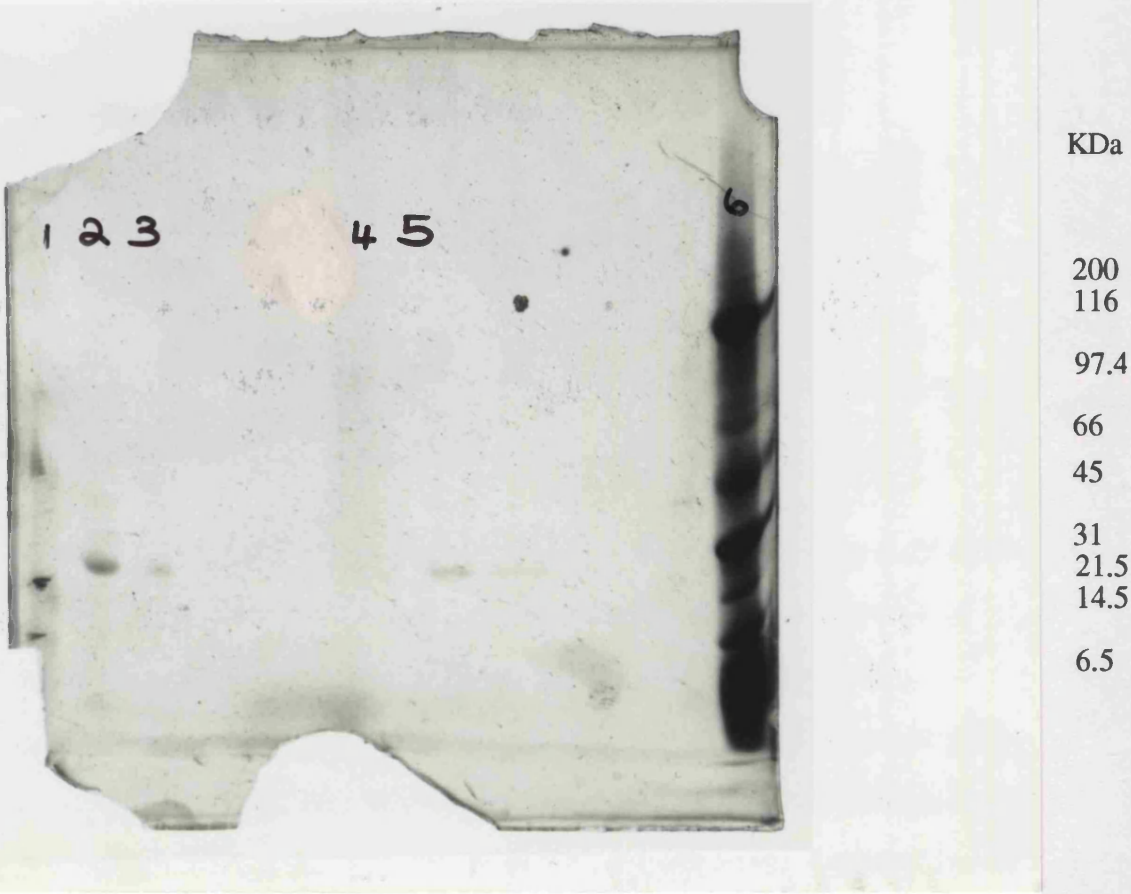
[The calibration curve was generated by the application of 1mg of Dextran, β amylase, alcohol dehydrogenase, albumin, carbonic anhydrase and cytochrome C to a pre-equilibrated Pharmacia Superose 6 column. The proteins were eluted by washing the column at a flow-rate of 0.5mLmin^{-1} and protein absorbance was monitored at 280nm. The protein peaks were also shown. Peak I corresponded to approximately 160 KDa and Peak II to 70 KDa.]



The fractions corresponding to Peaks I and II were assayed for both *Tfi*I endonuclease and GAPDH activity. Endonuclease activity was not detected in fractions corresponding to Peak I but was detected in fractions corresponding to Peak II. This confirmed the earlier data from Superdex 75 chromatography which indicated that *Tfi*I endonuclease has is a dimer of approximately 72 kDa. No GAPDH activity was detected in either of the fractions corresponding to Peak I and II. The fractions corresponding to Peaks I and II were concentrated by ultrafiltration with the addition of 2x SDS loading buffer. The concentrated fractions were electrophoresed on 12% (v/v) SDS acrylamide gels. Figure 4.5 shows the SDS PAGE gel of the fractions.

Figure 4.5 Silver stained SDS PAGE gel showing fractions from Superose 6 chromatography

[lane 1, pre-stained marker; lane 2, Fraction 13; lane 3, Fraction 14 (Peak I); lane 3, Fraction 15; lane 4, Fraction 16; lane 5, Fraction 17 (Peak II); lane 6, molecular weight marker.]



The silver stained SDS PAGE gel (Figure 4.5) showed only one band at 37 kDa for both Peak I and II. This suggested the Mono Q fraction contained two proteins i.e., a dimer of 70 KDa (Peak II) and a tetramer of approximately 160 KDa (Peak I). Peak II showed *Tfi*I endonuclease activity thus confirming the data that *Tfi*I endonuclease exists in solution as a dimer. Peak I could be a non-active tetrameric form of *Tfi*I endonuclease. Another supposition is that the tetrameric protein is in fact the Rot34A1 GAPDH protein. This is further supported by the knowledge that the *T. aquaticus* GAPDH protein is tetrameric. However no GAPDH activity was detected in this fraction and the data alone is not conclusive evidence that the tetrameric protein is Rot34A1 GAPDH.

4.3 pH PROFILE OF *Tfi*I ENDONUCLEASE

The pH profile of *Tfi*I endonuclease was determined by incubating pUC18 and *Tfi*I endonuclease in different buffers of pH ranging from 2.6 to 10.9. Figure 4.6 shows the profile obtained when 1 Unit of *Tfi*I endonuclease was incubated at 65°C with 0.5µg of pUC18 and 10mM $MgCl_2$ for 1 hour. A partially purified *Tfi*I extract containing $1U\mu L^{-1}$ activity was used in these experiments. pUC18 was found to denature below pH 3.0.

A second experiment was carried out to the measure *Tfi*I endonuclease activity at different pHs. 5 units of *Tfi*I endonuclease was added to 2.5µg of pUC18 in buffers containing 10mM $MgCl_2$, between pH 6 and pH 11 and the activity measured as described previously; i.e., aliquots were removed at regular time intervals and the reactions stopped by the addition of stop mixture. The samples were then electrophoresed in parallel on a 1% (w/v) agarose gel. The activity assays were carried out in duplicate. Figure 4.7 shows the residual activities at different pHs.

Figure 4.6 Digestion of pUC18 at different pHs for *Tfi*I endonuclease enzyme

[0.5µg of pUC18 was incubated with 10mM $MgCl_2$ and buffer of different pH at 65°C for 1 hour. The reactions were stopped and the samples electrophoresed as described previously.]

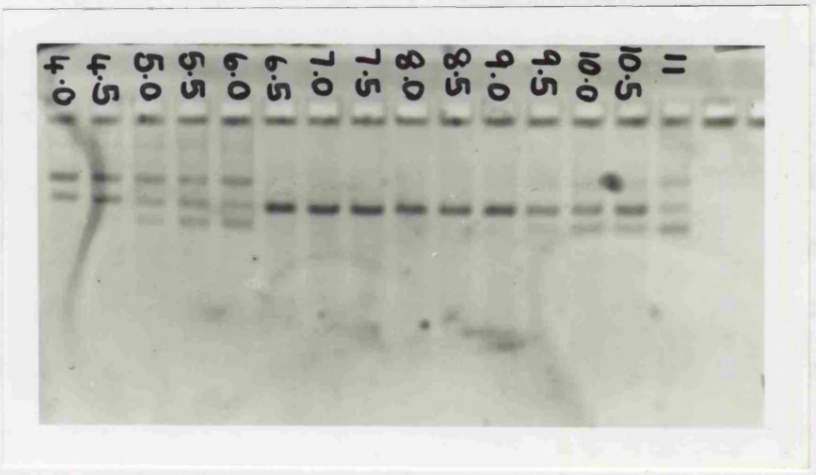
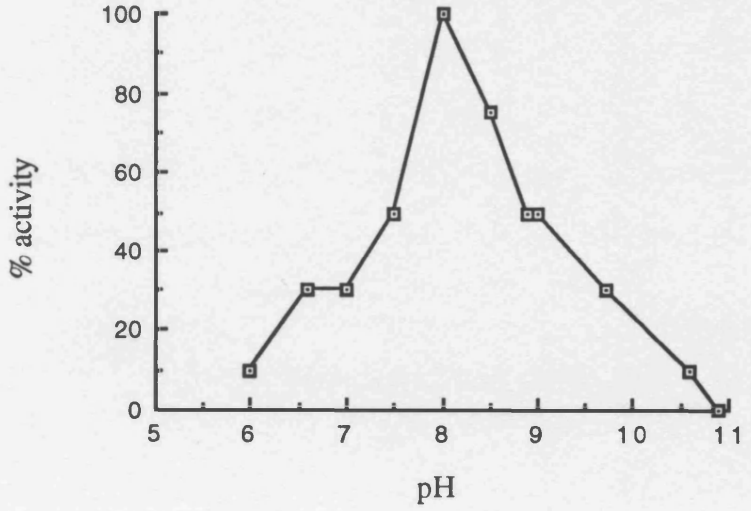


Figure 4.7 pH profile of *Tfi*I endonuclease

[Percentage activity (% activity) was calculated as the amount of activity at the specified pH compared to the activity of 5 units of *Tfi*I endonuclease under normal buffer conditions (Tris-HCl buffer pH 7.5).]



As seen from Figures 4.6 and 4.7, *Tfi*I endonuclease has a pH optimum of 8.0 and is active over a broad range of pH between pH 6.0 and 10.

4.3 THERMOSTABILITY OF *Tfi*I ENDONUCLEASE

Rot34A1 has a growth temperature of 65°C and the endonuclease is active at 70°C. The thermostability of *Tfi*I endonuclease was assessed by incubating 100 units (100µL) of *Tfi*I endonuclease at the different temperatures - 60°C, 65°C, 70°C, 75°C, 80°C, 85°C, 90°C, 95°C, 100°C. 20µL of *Tfi*I endonuclease was removed at specified times (1, 5, 10, 20, 30 and 60 minutes) and stored on ice. An aliquot was removed from each sample and assays carried out as previously described. The half life of *Tfi*I endonuclease at the various temperatures are shown in Table 4.1. This experiment was carried out in duplicate and using 10µL (10 units) of endonuclease for the activity assays.

Figure 4.8 shows the thermostability of the *Tfi*I endonuclease at the different temperatures. The experiments were carried out using a partially purified *Tfi*I endonuclease preparation. The Mono Q purified *Tfi*I endonuclease was found to be unstable at room temperature (data not shown).

Figure 4.8 Thermostability of *Tfi*I endonuclease

[Percentage activity (% activity) was calculated as the amount of activity at the specified temperature and compared with the activity of 10 units of *Tfi*I endonuclease under normal buffer conditions at 65°C.]

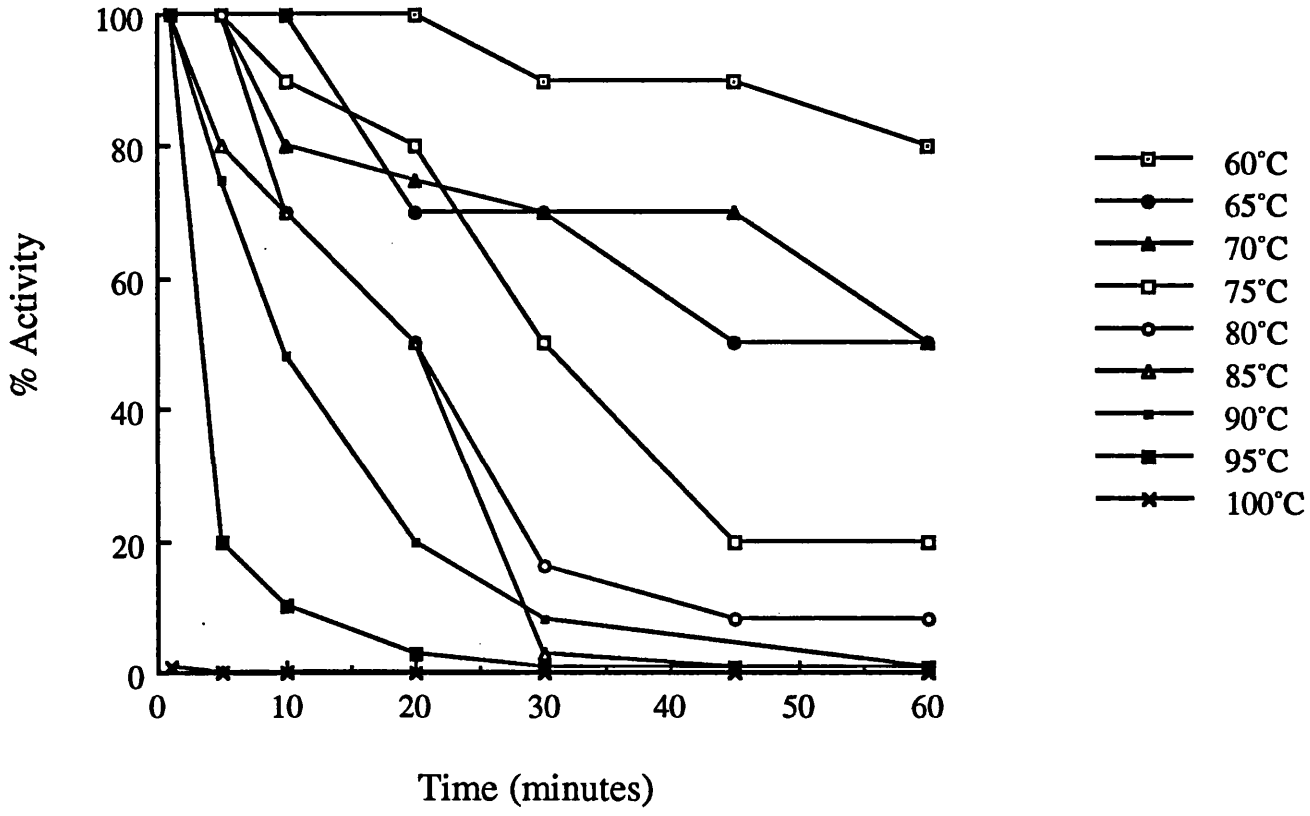


Table 4.1 Thermostability of *Tfi*I endonuclease

[Data were derived from Figure 4.8]

Incubation Temperature (°C)	Half Life (minutes)
60	>60
65	45
70	60
75	30
80	20
85	20
90	10
95	5
100	< 1

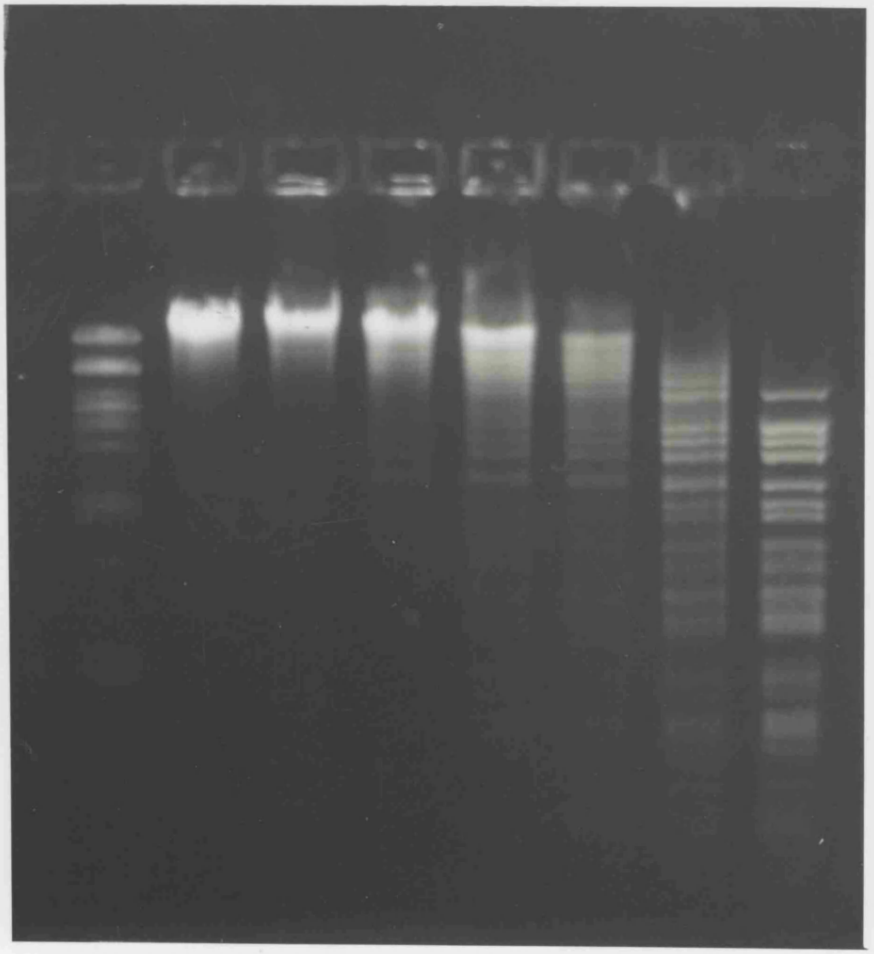
The data showed that at 100°C *Tfi*I endonuclease is very unstable and loses all activity in 1 minute. After 5 minutes there was only 20% of the activity was present at 95°C and 80% of the activity retained at 90°C. At 60°C the half life is greater then 60 minutes. Rot34A1 has a growth temperature of 65°C and at this temperature the endonuclease has a half life of 45 minutes.

4.4 EFFECT OF DETERGENTS

The activity of *Tfi*I endonuclease in the presence of SDS and Triton X100 (1% -10% w/v and v/v respectively) was examined. The endonuclease did not show any activity in the presence of SDS at all concentrations and showed activity only in the presence of 1% (v/v) Triton X100 (Figure 4.9).

Figure 4.9 Activity assay of *Tfi*I endonuclease in the presence of 1% (v/v) Triton X100

[1% (v/v) Triton X100 was prepared by the addition of 1µL of Triton X100 to 100µL of TE buffer containing 10mM $MgCl_2$ and λ DNA. 2 units of *Tfi*I ($1U\mu L^{-1}$) was added to 200 µL of this buffer and aliquots removed at regular time intervals over a two hour incubation at 65°C. The reactions were stopped by the addition of stop mixture added and the samples electrophoresed in parallel on 1% (w/v) agarose gel.]



Total digestion was only achieved at two hours. Percentage activity (%) was calculated as the amount of activity in a buffer containing 1% (v/v) Triton X100 compared to the activity of 2 units of *Tfi*I endonuclease under normal buffer conditions (Tris-HCl buffer pH 7.5). As seen from the photograph only 25% of the activity was present in Triton X100. The photograph also shows that there was no detectable "star activity".

4.5 EFFECT OF SALT ON *Tfi*I ENDONUCLEASE

Endonuclease activity can be affected by the presence of salt. For some endonucleases NaCl is an absolute requirement for cleavage whereas others do not require salt. The effect of salt on *Tfi*I endonuclease was investigated (Table 4.2).

Table 4.2 The effect of salt on *Tfi*I endonuclease

[0.5µg of pUC18 and 2µL of *Tfi*I endonuclease ($1U\mu L^{-1}$) were incubated with a range of buffers (TE buffer containing 0-700mM NaCl and 10mM $MgCl_2$) at 65°C for an hour. The reactions were stopped by the addition of stop mixture and the samples electrophoresed in parallel on 1% (w/v) agarose gel.]

NaCl Concentration (mM)	*Tfi*I endonuclease activity
0	✓
100	✓
200	✓
300	✓
400	not detectable
500	not detectable
600	not detectable
700	not detectable

The table shows that *Tfi*I endonuclease does not require NaCl for activity but is inhibited at salt concentrations above 400mM NaCl. Quantitative assays showed that only 30% of activity remained with 300mM NaCl.

4.6 STAR ACTIVITY STUDIES

Various extreme conditions induce "star" activity in endonucleases (Polisky *et al.*, 1975; Tikchinenko *et al.*, 1978; Barany, 1988). Star activity results from a relaxation of enzyme specificity. The conditions normally associated with star activity were tested with *Tfi*I endonuclease. All reactions were carried out at 65°C for one hour unless otherwise indicated. The reactions were stopped by the addition of stop mixture and electrophoresed on 1% (w/v) agarose gels. Different DNA substrates, λ, pBR322 and pUC18 were used in these experiments and their normal cleavage patterns are shown in Appendix I.

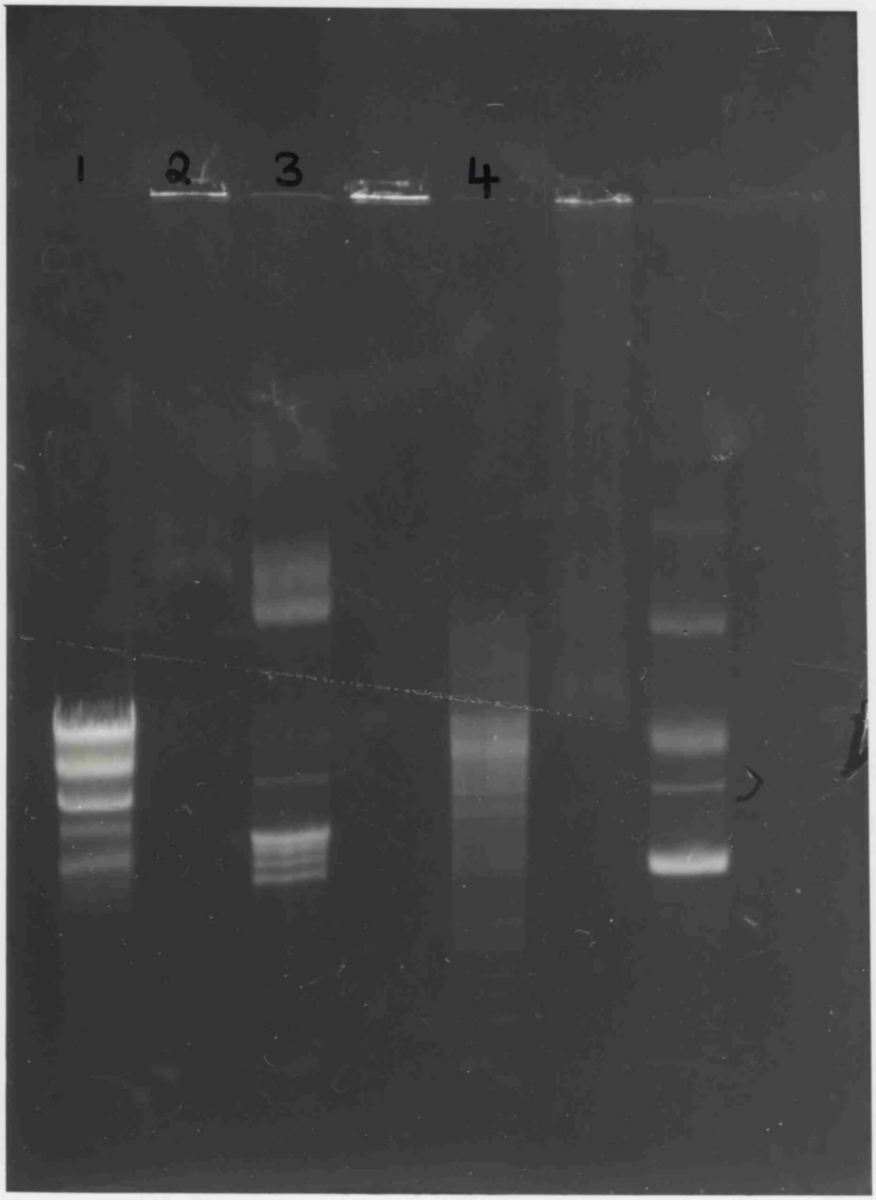
(i) High glycerol concentration

Concentrations of glycerol above 30% (v/v) were found to cause non-specific digestion of λ and pUC18. Figure 4.10 (overleaf) shows the digestion patterns of pUC18 and λ DNA in the presence of *Tfi*I and 30% (v/v) glycerol. These digestion patterns were found to vary between different experiments. Glycerol concentrations up to 20% (v/v) did not cause any star activity even when incubated for four hours at 65°C (data not shown).

(ii) Low ionic strength

*Tfi*I endonuclease exhibited some star activity in buffers below 10mM. Figure 4.11 shows star activity with pUC18 in different low ionic strength buffers. An additional band of approximately 400 base pairs was observed with pUC18. However longer incubations in 5mM buffers *Tfi*I endonuclease did not always produce the 400 base pair band (data not shown). An activity assay of λ DNA in 5mM Tris-HCl buffer did not show any detectable star activity (Figure 4.11).

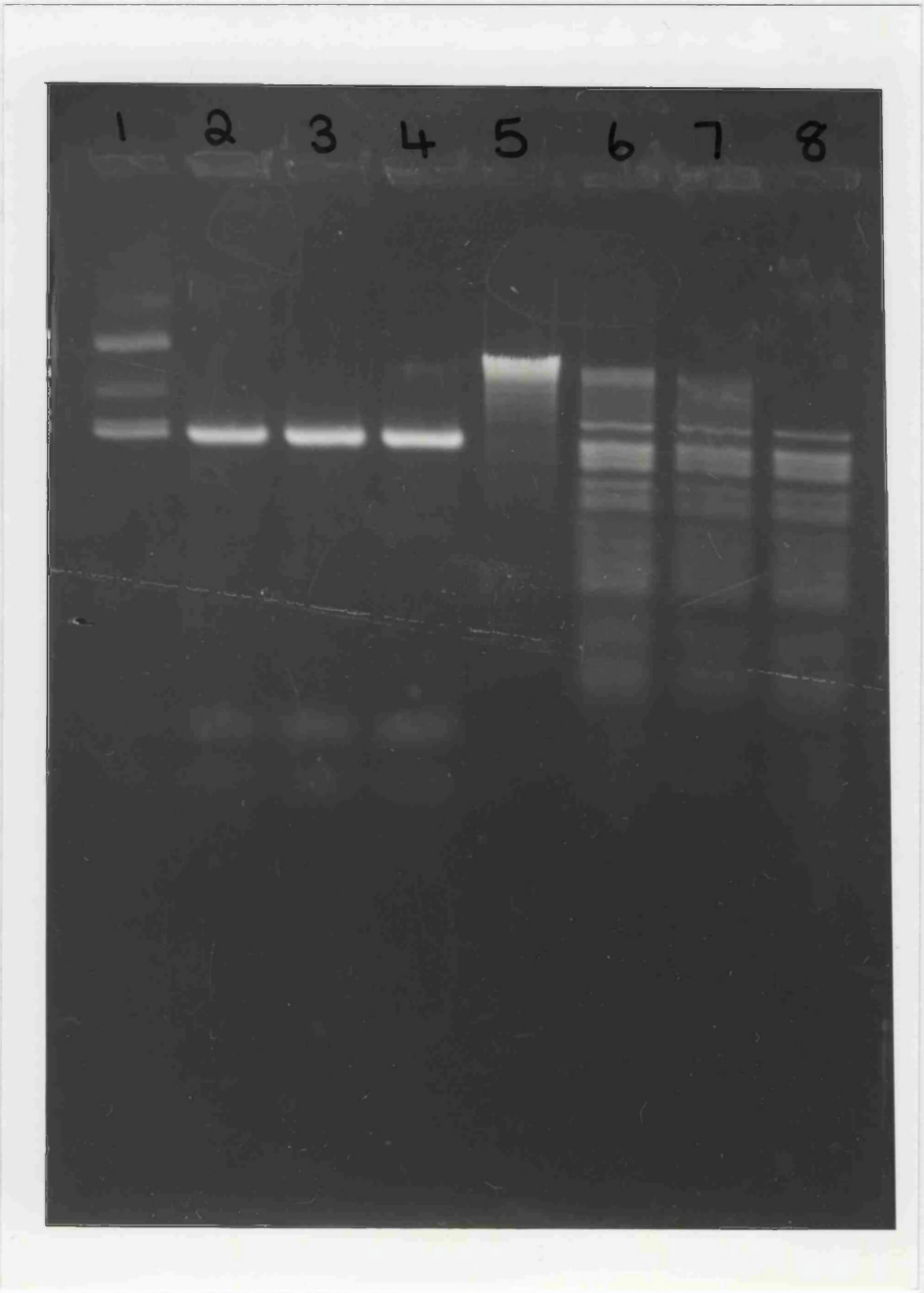
Figure 4.10 Digestion of pUC18 and λ DNA in the presence of 30% (v/v) glycerol [Lane 1, DNA marker; lane 2, pUC18 incubated at 65°C in restriction buffer containing 30% (v/v) glycerol (no enzyme) for 2 hours; lane 3, pUC18 incubated at 65°C in restriction buffer containing 30% (v/v) glycerol and 2μL of *Tfi*I endonuclease ($1U\mu L^{-1}$) for 2 hours; lane 4, λ incubated at 65°C in restriction buffer containing 30% (v/v) glycerol and 2μL of *Tfi*I endonuclease ($1U\mu L^{-1}$) for 2 hours. (λ incubated at 65°C in restriction buffer containing 30% (v/v) glycerol (no enzyme) not shown)]



Additional bands in both substrates

Figure 4.11 *Tfi*I activity on pUC18 and λ DNA in low ionic strength buffers

[The photograph show the activity of *Tfi*I endonuclease in the presence of low ionic strength buffers. Two hour incubation at 65°C of 2μL of *Tfi*I endonuclease ($1U\mu L^{-1}$), 0.5μg of pUC18, 10mM $MgCl_2$ and different Tris-HCl buffers (2mM, 5mM and 10mM). Lane 1, λ DNA-*Hin*dIII marker; lane 2, 2mM Tris-HCl; lane 3, 5mM Tris-HCl; lane 4, 10mM Tris-HCl. Activity assay in 5mM Tris-HCl with 2.5 μg of λ DNA, lane 5, 1 minute; lane 6, 30 minutes; lane 7, 60 minutes; lane 8, 120 minutes]

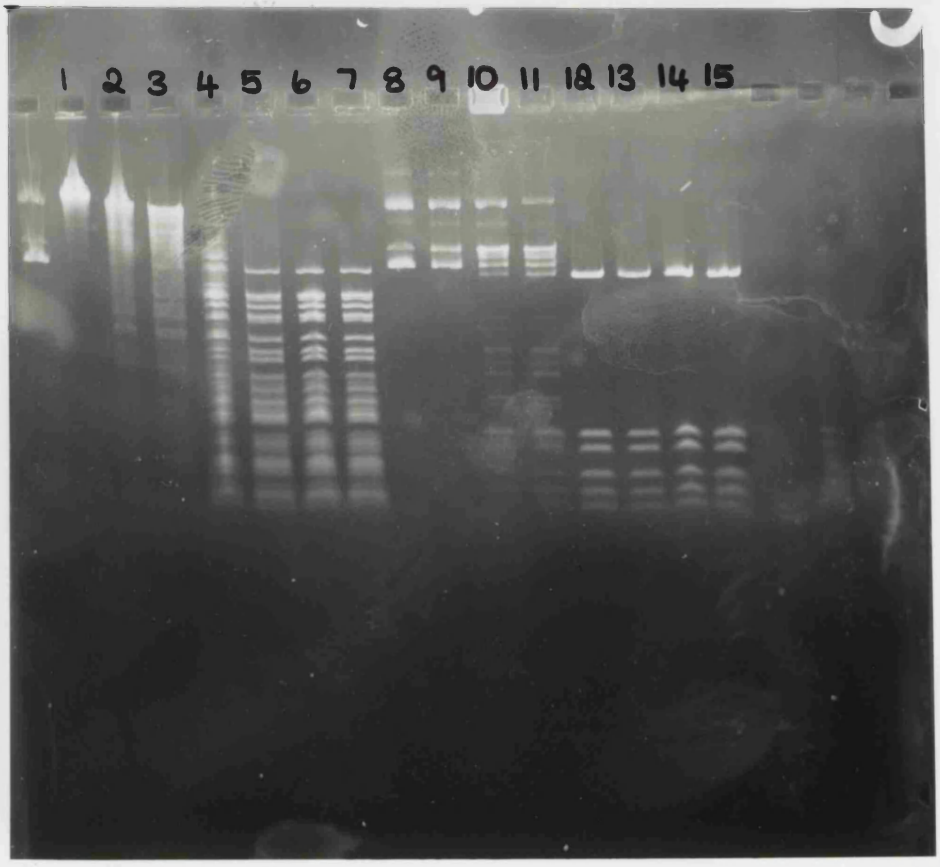


(iii) High pH

High pH (up to pH 11) did not induce star activity. Above pH 11, *Tfi*I endonuclease was not active (Section 4.2). After two hours at 65°C, no star activity was observed with either substrate. Figure 4.12 shows the activity assays of *Tfi*I endonuclease with two different substrates, λ DNA and pBR322, at pH 10.

Figure 4.12 *Tfi*I endonuclease activity in high pH buffer

[Activity assays were carried out as previously described with 2μL of *Tfi*I endonuclease ($1U\mu L^{-1}$) except TE buffer was replaced by carbonate buffer pH 10 and two different DNA substrates (λ DNA and pBR322) were used. Lanes 1-7 shows aliquots of λ DNA digestion at different times (1, 5, 15, 30, 60, 120, 180 minutes) and lanes 8-15 show aliquots of pBR322 digestions at different times (1, 5, 15, 30, 60, 90, 120, 180 minutes).]



4.7 EDTA AND METAL ION STUDY

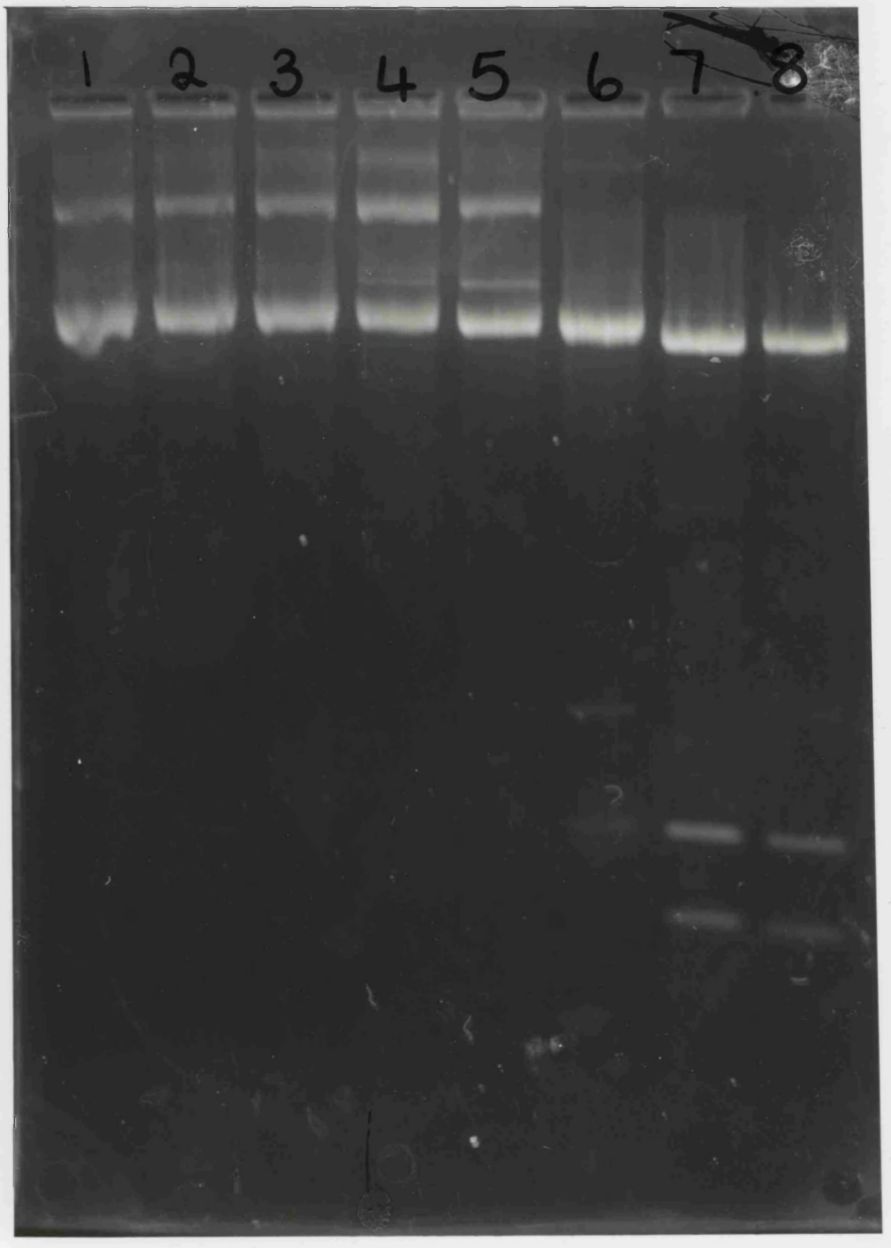
Na_2EDTA concentrations of greater than 10mM inhibited *Tfi*I endonuclease (data not shown). When endonuclease samples inhibited with Na_2EDTA were supplemented with an excess of different metal ions (see Section 2.23), no *Tfi*I endonuclease activity was detected with Co^{2+}, Cu^{2+} and Zn^{2+}. However replacement of Mg^{2+} ion with Mn^{2+} ions did show activity and a relaxation in the specificity (Figure 4.13).

The photograph shows that with 5mM $MnCl_2$, there are three additional bands at approximately 400, 800 and 1700 base pairs. pUC18 has two *Tfi*I sites 150 base pairs apart, thus giving two fragments, one of approximately 2500 base pairs and the other 150 base pairs. The additional fragments are the result of relaxed specification of the endonuclease.

The endonuclease also exhibits reduced activity with Mn^{2+} ions (¼U of *Tfi*I did not digest 2.5 µg of pUC18 to completion in 60 minutes). Incubation with Mn^{2+} ions for longer incubation times time resulted in further digestion of the 800 and 1700 base pair fragment to give three fragments, 2500, 400 and 200 base pairs (data not shown). Examination of the DNA sequence did show two possible sites for cleavage, **AATAC** and **CATAC**. Both differ from the recognition sequence (**GA$^{A}/_{T}$AC**) by only one base pair. However without detailed restriction enzyme analysis or sequencing the cleaved fragments is not possible to map the exact location of the cleavage sites of the Mn^{2+} induced star activity.

Figure 4.13 Activity of *Tfi*I endonuclease with Mn^{2+} ion replacing Mg^{2+} ions

[Activity assay was performed as described in Section 2.13.2 in the presence of 5mM $MnCl_2$ instead of 5mM $MgCl_2$ using 2µL of *Tfi*I endonuclease ($1U\mu L^{-1}$). Sample times were 1, 5, 10, 15, 20, 30, 60, 90 minutes (left to right).]

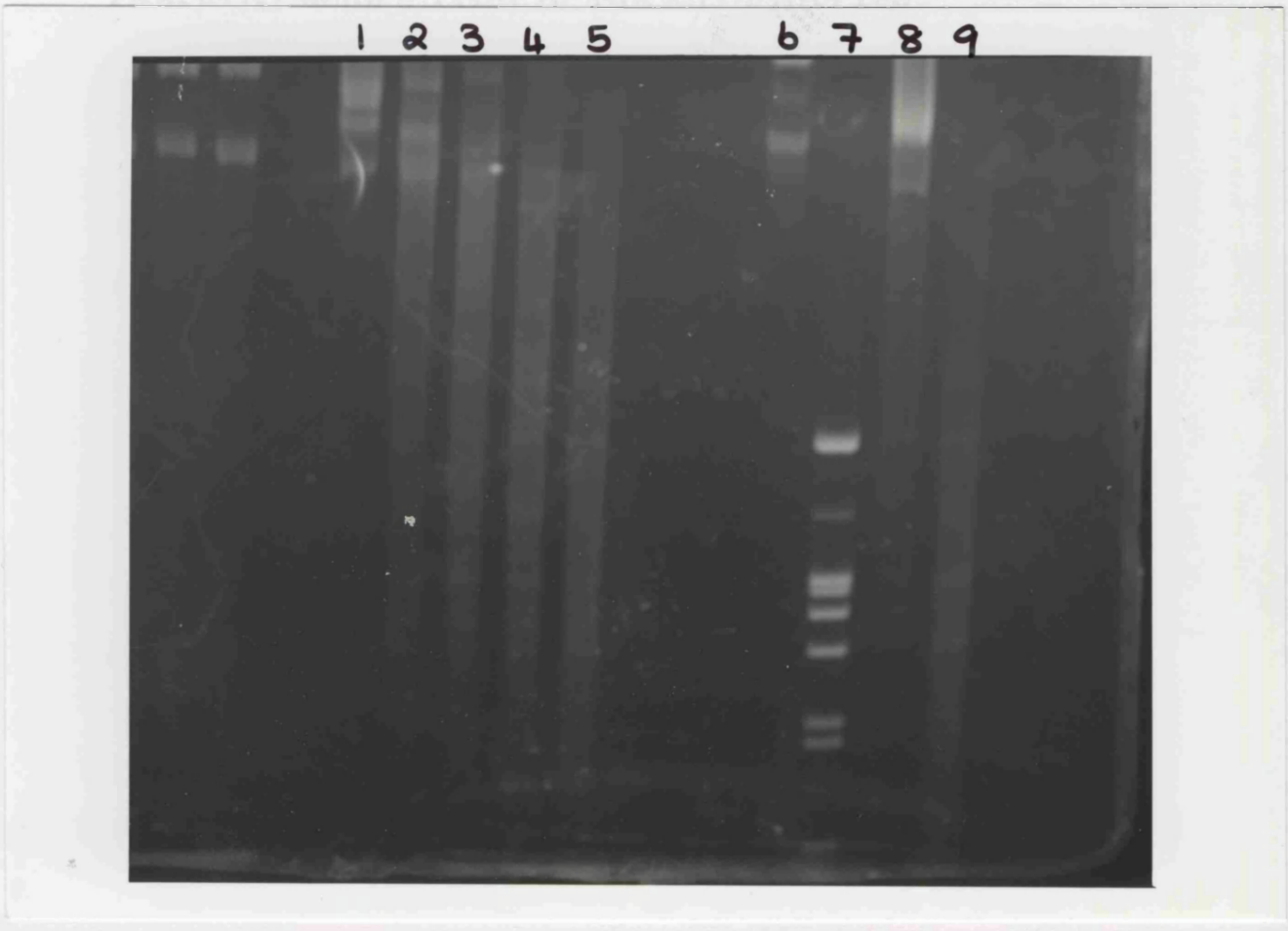


4.8 SINGLE STRANDED DNA

When tested with M13 and pQR700 (this plasmid was constructed and isolated in its single stranded form in this laboratory and is known to have 10 *Tfi*I sites), *Tfi*I endonuclease was found to cut single stranded DNA non specifically. Figure 4.14 shows the digestion of M13 and pQR700 with *Tfi*I.

Figure 4.14 Activity of *Tfi*I on single stranded substrates

[Lane 1-5 the activity assay of 2µL of *Tfi*I endonuclease (1UµL^{-1})with M13 as a substrate at different time intervals (1, 5, 10, 20, 30 minutes) over 30 minutes incubation at 65°C; lane 6, M13 incubated at 65°C with no enzyme for 30 minutes; lane 7, *Tfi*I digested pQR700 (ds form); lane 8, pQR700 (ss form) with no enzyme incubated at 65°C for 30 minutes; lane 9, digested pQR700 with 2µL of *Tfi*I endonuclease (1UµL^{-1}) at 65°C for 30 minutes]



Chapter 5 Partial purification and characterisation of the *Tfi*I methylase

5.1 INTRODUCTION

Modification methylases protect the host DNA from digestion by endogenous restriction endonucleases by incorporating methyl groups into the recognition sequence of their cognate endonucleases at either the adenine or cytosine residues (McClelland, 1983). These enzymes require s-adenosyl-L-methionine for their activity.

The *Tfi*I methylase was partially purified and a single step method was developed to separate the methylase from the endonuclease. The site at which the *Tfi*I methylase incorporates methyl groups into DNA was determined.

5.2 PARTIAL PURIFICATION OF THE *Tfi*I METHYLASE

5.2.1 PHOSPHOCELLULOSE CHROMATOGRAPHY

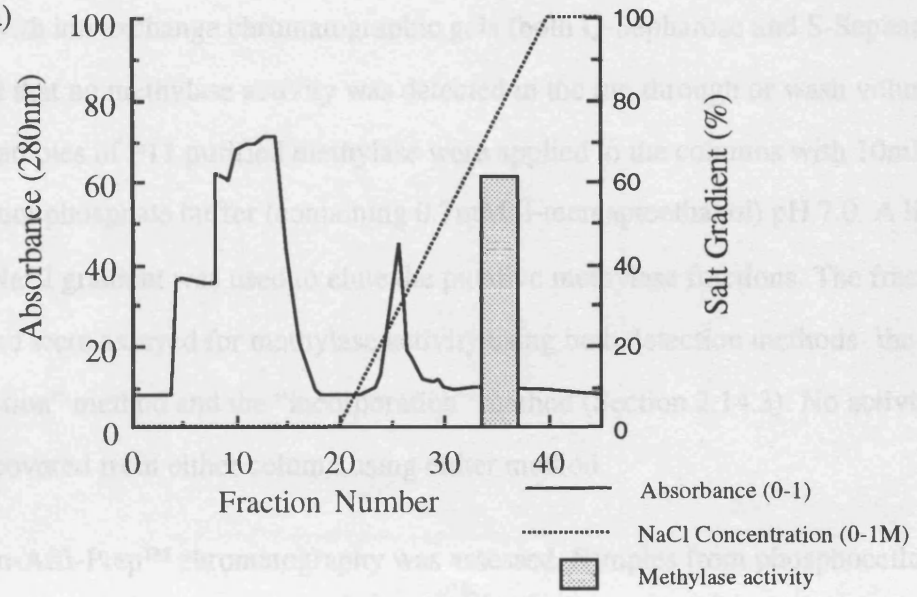
Cell extract was prepared as described previously (Section 2.13 and Section 3.3). The extract was applied to a column of phosphocellulose and the methylase was eluted with a linear salt gradient (0-1M NaCl). Protein was monitored using a UV monitor attached to a chart recorder. Fractions were assayed for methylase activity (Section 2.14.3 (i)). The methylase eluted between 700mM and 900mM NaCl (Figure 5.1). It was found that phosphocellulose purified *Tfi*I methylase was unstable. Even in 10mM potassium phosphate buffer (containing 0.7mM β mercaptoethanol and 50% (v/v) glycerol) pH 7.0 total activity was lost after overnight storage at -20°C. Subsequently, the phosphocellulose fractions were dialysed in 10mM potassium phosphate buffer (containing 0.7mM β mercaptoethanol and 50% (v/v) glycerol) pH 7.0 at room temperature for four hours before application onto the Heparin column.

Figure 5.1a Phosphocellulose chromatography. Cells were harvested from a 5L culture, cell extract prepared and 50mL of cell extract loaded onto a 40mL phosphocellulose column as described previously. The void [Fractions 0- 4] was collected separately as was the unbound sample [Fractions 5-10]. The column was washed with 80mL of 10mM potassium phosphate buffer (containing 0.7mM β mercaptoethanol) pH 7.0 [Fractions 11-20]. To elute the methylase a salt gradient (160mL 0-1M NaCl) was applied [Fractions 20 -40] and 8mL fractions were collected. The column then washed with 10mM potassium phosphate buffer (containing 0.7mM β mercaptoethanol and 1M NaCl) pH 7.0 [Fractions 41-45].

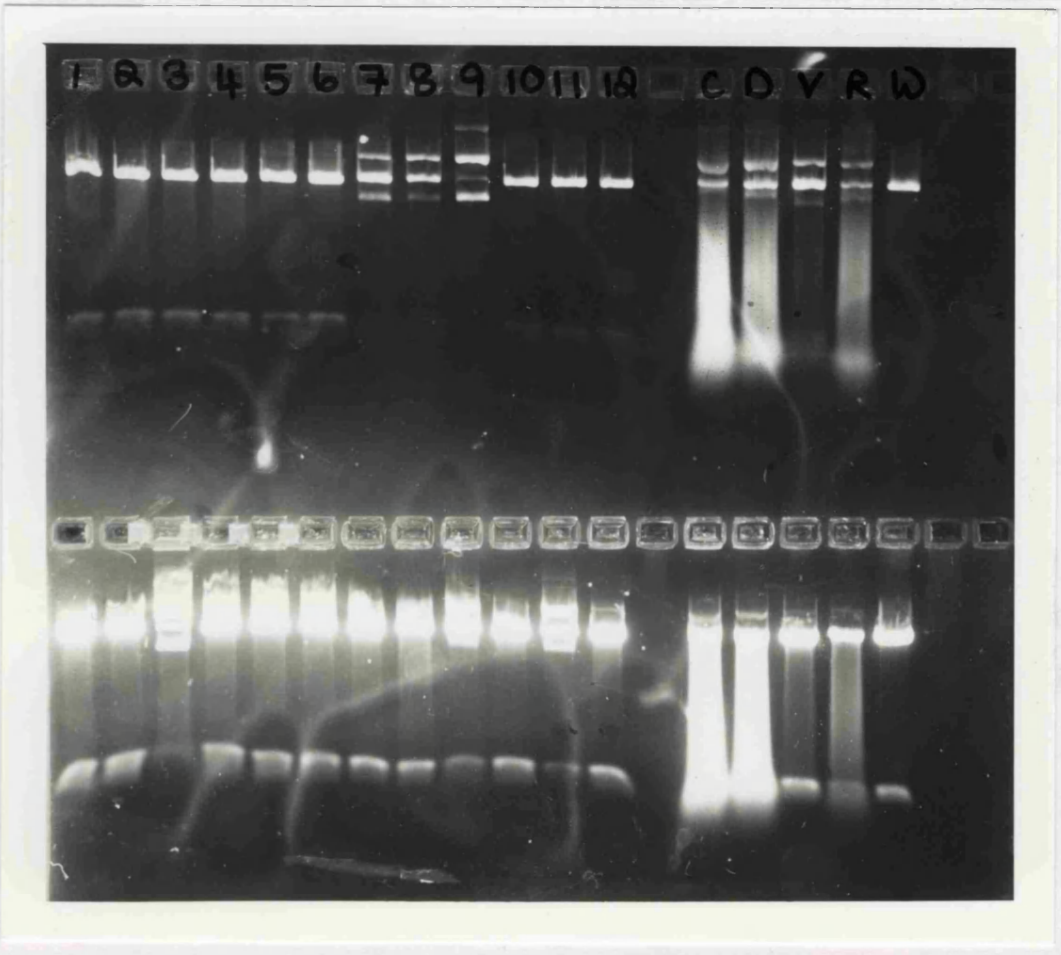
Figure 5.1b The methylase (protection) detection assay. The top lane shows the positive control and the bottom lane shows the negative control. Each track [1-12] represents the incubation of 5μL of every other fraction from the phosphocellulose column (Fractions 20 -45, Figure 5.1a) with the methylase assay mixture (Section 2.14.3). The crude extract [C], diluted crude [D], void volume [V] (Fractions 0-4, Figure 5.1a), unbound sample [L] (Fractions 5-10, Figure 5.1a), and low salt wash [W](Fractions 11-20, Figure 5.1a) were treated in a similar manner.

Figure 5.1 Elution profile of Rot34AI crude extract from phosphocellulose chromatography

(a)



(b)



5.2.2 HEPARIN-AFFI-PREP™ CHROMATOGRAPHY

Trials with ion exchange chromatographic gels (both Q-Sepharose and S-Sepharose) showed that no methylase activity was detected in the run through or wash volumes when samples of P11 purified methylase were applied to the columns with 10mM potassium phosphate buffer (containing 0.7mM β-mercaptoethanol) pH 7.0. A linear 0-1M NaCl gradient was used to elute the putative methylase fractions. The fractions collected were assayed for methylase activity using both detection methods- the "protection" method and the "incorporation" method (Section 2.14.3). No activity was recovered from either column using either method.

Heparin-Affi-Prep™ chromatography was assessed. Samples from phosphocellulose fractions were diluted four fold with 10mM potassium phosphate buffer (containing 0.7mM β mercaptoethanol) pH 7.0. The diluted extract was applied to a Bio-Rad Econo-Pac Heparin Cartridge (2.16.3.3). The methylase was eluted using a linear 0-1M salt gradient.

The protection detection methylase assay [2.14.3 (i)] did not show any methylase activity after a twelve hour incubation at 55°C. However, after a similar incubation the incorporation assay [2.14.3(ii)] showed methylase activity in the eluted fractions between 500mM and 700mM NaCl.

Figure 5.3 shows elution profile for Heparin chromatography. Figure 5.4 shows a SDS PAGE gel separation of proteins in the phosphocellulose and Heparin fractions. Table 5.1 shows the purification data derived from the partial purification of the *Tfi*I methylase.

Figure 5.2 Elution profile for P11 methylase fraction from Heparin-Affi-Prep chromatography

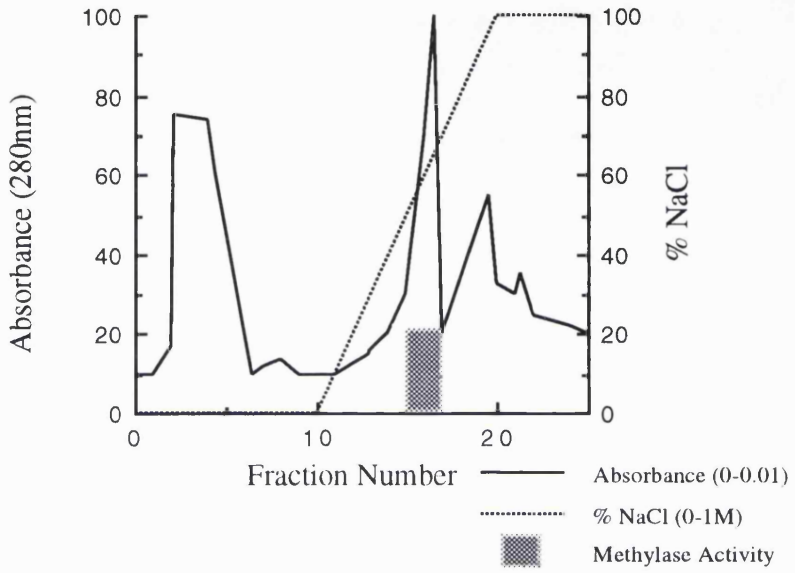


Figure 5.2 shows the protein and salt profile obtained from Heparin-Affi-Prep chromatography. 5mL of P11 purified methylase extract was diluted (4 fold) and loaded onto a BioRad Heparin Econo-Pac Cartridge™ as described in the text. The void [Fractions 0- 2] was collected separately as was the unbound sample [Fractions 3-6]. The column was washed with 10mL of 10mM potassium phosphate buffer (containing 0.7mM β-mercaptoethanol) pH 7.0 [Fractions 7-10]. To elute the methylase a salt gradient (50mL 0-1M NaCl) was applied [Fractions 10 -20] and 2.5mL fractions were collected. The column was then washed with 10mM potassium phosphate buffer (containing 0.7mM β-mercaptoethanol and 1M NaCl) pH 7.0 [Fractions 21-25].

Table 5.1 Partial purification of the *TfiI* methylase

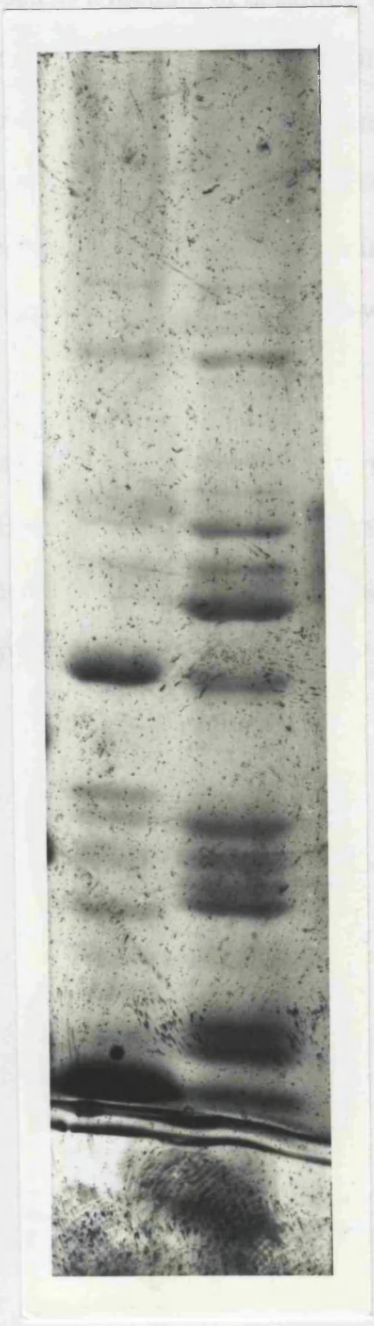
[The table shows a 10L cell culture prepared and applied successively to a phosphocellulose column and a Bio Rad Heparin Econo Pac™ cartridge. Activity and protein concentration were measured as described in 2.14.4 and 2.15.]

STEP	Volume (mL)	Total Activity (U)	Total Protein (mg)	Specific Activity (Umg^{-1})	Yield (%)	Purification Factor (Fold)
Crude extract	25	7500	200	37.5	100	1
Phospho-cellulose	8	2000	0.32	6250	25	158
Heparin Affi-Prep™	10	1500	0.25	6000	19	150

The first chromatographic step, phosphocellulose chromatography resulted in a 150 fold increase in the purification factor, but with a 75% loss in methylase activity. The second step was not very efficient, with a purification factor of 0.96 fold. However 75% of the activity was retained. The low specific activity of the Rot34A1 methylase preparations led to an initial detection assay involving overnight incubations at 55°C and activity measurements made over six hour incubations. The stability of the methylase during these long incubations would influence the activity detected.

Figure 5.3 Photograph of the silver stained SDS PAGE gel showing the initial steps of purification of the methylase

[Lane 1, P11 fraction (1mL); Lane 2, Heparin fraction (2mL)]



The Heparin active fractions lost all activity when dialysed against 10mM potassium phosphate buffer (containing 0.7mM β mercaptoethanol and 50% (v/v) glycerol) pH 7.0. Subsequently the methylase was used immediately.

Gel filtration was assessed as a method of desalting the sample and to calculate the molecular weight of the *Tfi*I methylase. 5mL of purified heparin extract was applied to a pre-equilibrated HiLoad 16/60 Superdex 200 column with 10mM potassium phosphate buffer (containing 0.7mM β-mercaptoethanol and 200mM NaCl) pH 7.0. The column was washed with the buffer at 0.5mLmin^{-1} and 4mL fractions collected. No methylase activity was detected using either of the two detection methods after a twelve hour incubation.

One possibility for the apparent instability of the partially purified methylase was the presence of proteases. However, attempts to stabilise the Heparin purified methylase fractions by the addition of protease inhibitors did not enhance stability. The purification of the methylase did not proceed further.

5.3 DETERMINATION OF METHYLATION SITE OF THE *Tfi*I METHYLASE

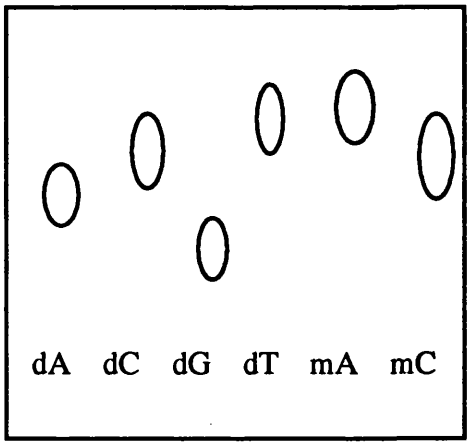
20μg of λ DNA was methylated with Rot34AI methylase. The methylase was purified by phosphocellulose chromatography as described in Section 5.2.1. The DNA was then hydrolysed to nucleosides using phosphodiesterase and calf intestinal alkaline phosphatase (Section 2.21). The samples were then desalted and concentrated. The samples were resuspended in 10μl of water and spotted onto pre-coated TLC sheets.

Two different solvent systems were used for the separation of the nucleosides (i) 80:20 ethanol:water (v/v) and (ii) 66:33:1 isobutyric acid:water: ammonium hydroxide (v/v). The plates were visualised using a 254nm UV light. Table 5.2 shows the migration of the standards. Figures 5.5 and 5.6 show pictorial representations of the migration of standards in the two solvent systems.

After thin layer chromatography of the labelled DNA hydrolyses the area corresponding to the N^6 methyl 2' deoxyadenosine and 5 methyl 2' deoycytidine were "scraped off" and counted for radioactivity. Table 5.3 shows the incorporation of the [^{3}H] S-adenosyl-L-methionine in the these areas. (Section 2.21)

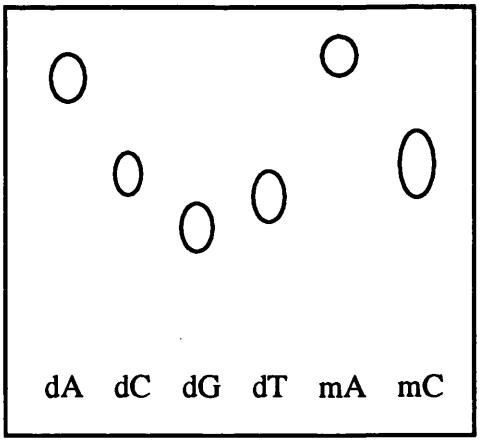
Similar hydrolytic treatment of Rot34AI chromosomal DNA did not result in any nucleosides visible to the eye in the areas corresponding to the N^6 methyl 2' deoxyadenosine and 5 methyl 2' deoycytidine.

Figure 5.4 The separation of nucleosides with the 80:20 ethanol:water (v/v) solvent system



The figure shows the separation of the nucleosides with the 80:20 ethanol:water (v/v) solvent system. The expected migration of the N^4 methylcytosine would be with N^6 methyl 2' deoxyadenosine in this solvent system.

Figure 5.5 The separation of nucleosides with the 66:33:1 isobutyric acid:water: ammonium hydroxide (v/v) solvent system



The figure shows the separation of the nucleosides with 66:33:1 isobutyric acid:water: ammonium hydroxide (v/v).solvent system. The expected migration of the N^4 methylcytosine would be with 5-methylcytosine in this solvent system.

Table 5.2 Cellulose TLC of deoxynucleosides

Compound Name (Symbol)	Rf Values Solvent A[a]	Solvent B[b]
Deoxyadenosine (dA)	0.61	0.78
Deoxycytidine (dC)	0.65	0.73
Deoguanosine (dG)	0.50	0.63
Thymidine (dT)	0.86	0.69
5-methyldeoxycytididine (m^5dC)	0.73	0.74
N^6-methyldeoxyadenosine (m^6dA)	0.88	0.79

[a] Solvent A was 80:20 ethanol:water (v/v).

[b] Solvent B was 66:33:1 isobutyric acid:water: ammonium hydroxide (v/v).

Table 5.3 Incorporation of [^{3}H] S-adenosyl-L-methionine in *TfiI* methylase methylated DNA

	Solvent A		Solvent B	
Identity	m^5dC	$m^4dC + m^6dA$	m^6dA	$m^4dC + m^5dC$
Rf values	0.75	0.90	0.83	0.72
	^{3}H incorporation (cpm)[a]			
	320	53	62	420

[a] ^{3}H incorporation was measured for 10 minute counting periods as described in Section 2.4.

As seen from Table 5.3, the area with significant counts corresponded to 5-methylcytosine migration. As *TfiI* recognises the sequence $GA^A/_TTC^*$, it can be concluded that the external cytosine is methylated by *TfiI* methylase (* indicates the base methylated) at the ^5C.

Chapter 6 Cloning Strategies for the *Tfi*I restriction modification system

6.1 INTRODUCTION

Various methods have been used to clone the restriction and modification genes of different organisms (Section 1.10 Introduction). The method selected to screen for the endonuclease gene of *Thermus filiformis* in this study was based on the modification method (see Section 1.10.4). The selection of positive clones in this method comprises of exposing recombinant plasmids from a library of *T. filiformis* to *Tfi*I endonuclease digestion. The method depends on expression of the cloned methylase gene and the consequent *in vivo* modification of those plasmid molecules that carry the gene.

A library of recombinant plasmids is prepared and propagated in *E. coli* to allow expression and modification to occur. The plasmid population is then purified and digested with the corresponding endonuclease that cleaves unmodified molecules. Undigested molecules, recovered by transformation, are usually found to be modified and to carry the methylase gene.

Since the genes for restriction and modification are often closely linked, the genes can be cloned simultaneously if the association is not disrupted during cloning. If the association is disrupted, the clones carrying just the methylase gene are isolated. Many restriction modification genes have been isolated in this manner including the genes for the restriction modification systems of *Nae*I and *Nco*I and the modification genes of *Fnu*DI and *Xba*I (VanCott and Wilson 1988).

The plasmid vectors must contain a selectable marker and *Tfi*I sites. pUC18, pUC19 and pBR322 were chosen as they contain *Tfi*I sites and selectable markers (see Appendix II

and III for plasmid maps). Libraries were also constructed in lambda vectors (Appendix IV).

This chapter discusses the construction and screening of various gene libraries to locate the restriction and modification genes of Rot34A1. However the genes for the *Tfi*I restriction modification system were not isolated from the plasmid libraries after extensive work with these libraries. The problems associated with creating gene libraries are discussed in this chapter. A library created using a λ vector has been created. The N-terminal data for the endonuclease was not elucidated in this study. Further work with this library was not possible.

6.2 CREATING AND SCREENING LIBRARIES IN PLASMID VECTORS

Libraries were created in pUC18, pUC19 and pBR322. Different strategies were adapted to create libraries and for screening in the plasmid vectors. pUC18 and pUC19 have two *Tfi*I sites and pBR322 has six sites for *Tfi*I.

Chromosomal Rot34A1 was isolated as described in Section 2.7.1. The chromosomal DNA was digested with a number of restriction endonucleases (Table 6.1). It was found that the endonucleases *Bam*HI, *Kpn*I and *Pst*I cut Rot34A1 chromosomal DNA into a wide spectrum of fragment sizes and therefore these enzymes were used for cloning Rot34A1 DNA.

Escherchia coli strain JM107 was initially used as the transforming host but the numbers of recombinant colonies were lower than expected. From this observation it was concluded that JM107 may have a weakly expressed *mcrB* restriction system. The *mcr*B restriction system is known to induce SOS by encoding a restriction endonuclease which

cleaves cytosine methylated DNA (Section 1.10.4). At the time of carrying out the initial cloning experiments, the site of methylation by *Tfi*I methylase (Chapter 5) was not known. To avoid the possibility of a restriction system that methylated DNA GM2163 was selected as the host organism because it has been mutated to have the *mcr*B restriction system abolished. However, GM2163 does not have a modified *lacZ* gene allowing for screening for recombinants on agar containing X-gal plus IPTG. JM107 was also used to calculate the ratio of recombinants to parental vector.

Table 6.1 Restriction endonuclease cleavage sites in Rot34AI chromosomal DNA

Endonuclease	Cut sites in Rot34AI
*Bam*HI	✓
*Bgl*II	–
*Eco*RI	–
*Hin*dIII	–
*Kpn*I	✓
*Pst*I	✓
*Sal*GI	–
*Xho*I	–

– indicates either no digestion or insufficient digestion of chromosomal DNA

✓ indicates digestion of chromosomal DNA to give fragment sizes suitable for cloning i.e. between 23kb and 4kb

6.2.1 CREATING LIBRARIES IN pUC18 AND pUC19

Chromosomal Rot34A1 DNA was digested with *Bam*HI, *Kpn*I and *Pst*I. Vectors pUC18 and pUC19 were also digested with these enzymes. The libraries were created using 1μg of vector and 2-3μg of digested chromosomal DNA. Ligations were performed as described in Section 2.6. An aliquot of each ligation (10%) was transformed into JM107 to calculate the number of recombinants. Cells were grown overnight at 37°C in 5mL nutrient broths. After twelve hours growth, the cells were plated onto nutrient agar plates supplemented with ampicillin. Table 6.2 summarises the creation of the pUC-derived libraries.

Table 6.2 Creation of libraries using pUC-derived vectors

[The number of recombinants refers to the number of recombinants as calculated from the transformation of the 10% of the ligated mixture into JM107. The remainder was transformed into GM2163. [a] calculated for GM2163 libraries]

Treatment of chromosomal Rot34A1 DNA	Treatment of vector DNA	Total number of recombinants[a]
*Bam*HI digested DNA fragments	*Bam*HI digested pUC18	≅ 3000
*Kpn*I digested DNA fragments	*Kpn*I digested pUC19	≅ 3000
*Pst*I digested DNA fragments	*Pst*I digested pUC18	≅ 4000

6.2.2 STRATEGY FOR SCREENING THE pUC18 AND pUC19 LIBRARIES

The strategy adopted for screening the pUC derived libraries was as detailed in Section 6.1. The libraries were amplified in GM2163 as described in Section 2.12.1. Plasmid DNA was isolated as described in 2.7.2. 1µg of the plasmid DNA was challenged with *Tfi*I (Q Sepharose purified). The challenged DNA was transformed into GM2163. The transformed cells were grown for twelve hours at 37°C and plated out on media supplemented with ampicillin.

The number of colonies from the *Tfi*I treated *Bam*HI pUC18 library exceeded 500, the *Kpn*I pUC19 library exceeded 1 000 and the *Pst*I pUC18 library had over 800 colonies after challenge with *Tfi*I. Screening each colony separately was not feasible. Screening a selection of each library by rapid disruption of bacterial colonies (Sambrook et *al.*, 1989, 1.32) revealed that the majority of the colonies contained only the vector.

This was thought to be the result of incomplete digestion of the plasmid library by *Tfi*I. In a mixed population of transformed cells, cells containing only the vector will have an advantage over other recombinants and will grow faster and thus will be the predominant species in the library.

The screening procedure was modified. The challenged libraries were pooled and amplified as above, then a second challenge with *Tfi*I was incorporated. The challenged libraries were plated on to nutrient agar plates supplemented with ampicillin. The numbers in the libraries were reduced by 2 fold, however the frequency of recombinants was not significantly increased. It was found that a third round of screening did not reduce the numbers in the libraries or increase the frequency of recombinants. Table 6.3 shows the number of recombinants in the challenged libraries at each stage of the screening protocol.

A selection of these colonies were grown overnight in 5mL nutrient broths at 37°C. 2mL of cells were then harvested and sonicated to break open the cells. The cell extract was heated to 70°C for 10 minutes to remove endogenous nuclease activity and incubated overnight with λ and restriction buffer. No isolates were found to contain *Tfi*I restriction endonuclease activity. The remainder of the cell culture (3mL) was harvested and plasmid DNA was isolated. The DNA was digested with *Tfi*I. All recombinants were found to be digested by *Tfi*I, and were found to be one of two types, smaller than pUC18 or pUC19 or the vector alone. The smaller plasmid was found to have lost one *Tfi*I site. Restriction digestion analysis of the smaller recombinant revealed that a deletion had occurred. This deletion was mapped to the 150 base pair fragment separating the *Tfi*I sites (data not shown).

The number of recombinants were calculated as described previously. 10% of the library was transformed into JM107 and grown overnight and plated onto nutrient agar supplemented with X-Gal, IPTG and ampicillin. Assuming JM107 was inducing the SOS repair via the *mcrB* cleavage mechanism on recognising Rot34A1 methylated DNA, the frequency and number of recombinants would be reduced. Given the doubling time of *E. coli* (approximately 20 minutes) comparisons between the number of transformants into JM107 and GM2163 cell populations after twelve hours growth would not be possible. The numbers obtained from the JM107 libraries would be lower and not an accurate representation.

The numbers of recombinants in the pUC-derived libraries is thought to be one reason for not isolating the restriction modification genes. The size of a genomic library is dependent on the size of the genome and the size of the inserted DNA fragment.

If x is the insert size and y is the size of the haploid genome then a library of

$$N = \frac{\ln(1-p)}{\ln(1-x/y)}$$

clones will have a probability p of containing any particular DNA sequence (Clarke and Carbon, 1976). Assuming 2kb fragments for the insert fragments and the same genomic size as *E. coli* (4.2×10^6 bp) for Rot34A1, the number of recombinants necessary to give a representative library would have to be 9000. If the number of recombinants in the Rot34A1 libraries were between 3000 and 4000, the probability would be reduced by 20% assuming 2kb fragments. However the size of the insert fragments were known to be between 1 and 20kb so statistically it should have still been possible to isolate the genes.

If *Bam*HI, *Kpn*I or *Pst*I sites were within the *Tfi*I restriction modification ORFs the gene sequence would disrupted and the genes would not be expressed.

Several different post methylation selection steps may have enhanced screening of non methylase clones such as the use of calf intestinal phosphatase and bacterial phosphatase. The use of nucleases such as exonuclease III and Bal-3I nuclease have been reported to render non methylated plasmids less likely to religate and survive (VanCott 1990).

Table 6.3 pUC-derived libraries at different stages of the screening protocol

[1μg of the pooled plasmid DNA was subjected to *Tfi*I digestion. The challenged DNA was transformed into GM2163 (90%) and JM107 (10%). The cultures were grown overnight and the culture plated out on to appropriate plates. The resultant plasmid library from GM2163 was pooled, plasmid DNA isolated and 1μg of the plasmid DNA was challenged with *Tfi*I. The challenged DNA was then transformed into GM2163 (90%) and JM107 (10%) grown overnight in a 5mL nutrient broth at 37°C. The cultures were plated on to media containing ampicillin, IPTG, and X-Gal to calculate number of recombinants.]

	Number of recombinants	1st round of *Tfi*I digestion Number of recombinants[a]	2nd round of *Tfi*I digestion Number of recombinants[b]
*Bam*HI digested library	3000	50	30
*Kpn*I digested library	3000	100	60
*Pst*I digested library	4000	80	40

[a]The number refers to the number of recombinant colonies that were present in 5mL of nutrient broth after an initial round of *Tfi*I digestion.

[b] The number refers to the number of recombinant colonies that were present in 5mL of nutrient broth after a second round of *Tfi*I digestion.

6.2.3 CREATING LIBRARIES IN pBR322

The pUC plasmids have only two *Tfi*I sites (150 bp apart) whereas pBR322 has six sites within the plasmid (see Appendix III). The sites are at positions 852, 1006, 1304, 1525, 2029, 2448. The number of sites ensures that the vector will be more likely to be inactivated on *Tfi*I digestion.

Screening was also hindered by the lack of a visual screen when transforming the pUC derived libraries into GM2163. All survivors had to be screened individually to check for restriction and methylase activity. The number of "vector only" recombinants was high and the lack of a selectable marker made screening very difficult. *Bam*HI and *Pst*I have sites in pBR322, both are found on different antibiotic genes (Appendix III). Inactivation of one or other of these genes could be utilised as a screen.

The number of *Tfi*I sites in pBR322 and the antibiotic selection made pBR322 an attractive cloning vector. Libraries using *Bam*HI and *Pst*I were constructed and screened for the *Tfi*I restriction modification genes. To counter the possibility that *Bam*HI and *Pst*I sites may exist in the restriction modification genes of Rot34A1 libraries were made using partially digested chromosomal DNA.

A *Pst*I pBR322 library was constructed with totally *Pst*I digested chromosomal DNA and *Pst*I digested pBR322 (Table 6.4).

*Pst*I partially digested chromosomal DNA was prepared as described by Polisson (1991). 22.5μg of chromosomal Rot34A1 DNA was partially digested with *Pst*I: DNA was diluted with TE and restriction buffer to give a final volume of 450μL was divided into one 100μL aliquot and seven, 50μL aliquots. To the 100μL tube 40 units of *Pst*I was added to achieve 8.0 units of enzymes per μg of DNA. 50μL was withdrawn from the

first tube and transferred to the second tube to achieve 4 units *Pst*Iμg^{-1} and so on, each succeeding tube receiving half of the previous amount of *Pst*I. The tubes were incubated at 37°C for one hour, then heat treated at 70°C for 10 minutes and an aliquot electrophoresed on 1% (w/v) agarose gel. 2μg of chromosomal DNA exhibiting moderate digestion (fragments between 6 and 4 kb) was precipitated with *Pst*I digested pBR322. The DNA was precipitated following phenol-chloroform extraction and resuspended in 50μL of TE buffer. T_4 ligase was added with ligase buffer and incubated at 8°C overnight. This library was designated as partially digested *Pst*I pBR322 library (Table 6.4).

Aliquots of the ligated DNA was transformed into the *E. coli* strain GM2163 as described in Section 2.8 and grown for twelve hours at 37°C in 5mL of nutrient broth. 100μL of the culture were plated in duplicate on nutrient agar plates supplemented with tetracycline and on nutrient agar plates supplemented with tetracycline and ampicillin. This provided a method for calculating the ratio of recombinants in the culture and a method for calculating the number of recombinants in the library.

*Pst*I cleaves pBR322 in the ampicillin (Amp) resistance gene. Therefore cloning into the *Pst*I site should result in clones that do not survive on plates supplemented with ampicillin. However the clones should grow successfully on tetracycline (Tc) plates. The ratio of recombinants and "vector only" clones can be calculated by plating an aliquot onto both ampicillin and tetracycline nutrient agar supplemented plates and nutrient agar supplemented tetracycline plates (Table 6.4).

In a similar fashion *Bam*HI cleaves pBR322 in the tetracycline resistance gene. Therefore cloning into the *Bam*HI site should result in clones that do not survive on plates supplemented with tetracycline. The clones should grow on ampicillin plates.

The ratio of recombinants and "vector only" clones can be calculated by plating an aliquot onto both ampicillin and tetracycline plates and tetracycline plates.

A *Bam*HI pBR322 library was constructed with totally *Bam*HI digested chromosomal DNA and *Bam*HI digested pBR322 (Table 6.4).

Chromosomal DNA was size fractionated using *Sau*3A1 as described in Section 2.7.4. Figure 6.1 shows the gel showing aliquots of the fractionated DNA. A *Sau*3A1 partial digested library in pBR322 was created (Table 6.4).

Figure 6.1 *Sau*3AI fractionated chromosomal Rot34A1 DNA

[DNA was size fractionated as described in 2.7.4. 20µL of each aliquot was electrophoresed in parallel on a 0.7% (w/v) agarose gel overnight at 20V. Lane 1 shows the λ-*Hin*dIII marker and lane 2-20 show the aliquots of the sucrose gradient. 5µg of size fractionated Rot34A1 chromosomal DNA (lane 18 and 19), corresponding to between 4 and 6kb was precipitated after phenol chloroform extraction and ligated to 1µg of linearised pBR322 (digested with *Bam*HI). The DNA was precipitated following phenol extraction and resuspended in 50µL of TE buffer. 4µL of T_4 ligase ($10UµL^{-1}$) was added with ligase buffer and incubated at 8°C overnight.]

Table 6.4 Quantitation of the recombinants in the pBR322 libraries

[An aliquot of the ligation was transformed into GM2163 and grown overnight at 37°C in 5mL of nutrient broth. The cell culture was plated out onto the appropriate selective nutrient agar plates and also on to nutrient agar plates containing both ampicillin and tetracycline to calculate the number of recombinant colonies.]

Treatment of Rot34A1 chromosomal DNA	Treatment of pBR322	Recombinant colonies
*Pst*I digested chromosomal DNA fragments[a]	*Pst*I digested pBR322	6100
*Pst*I partially digested chromosomal DNA fragments	*Pst*I digested pBR322	7500
*Bam*HI digested chromosomal DNA fragments[a]	*Bam*HI digested pBR322	8000
*Sau*3A1 partial digested (4-6kb) fragments	*Bam*HI digested pBR322	10 000

[a] 1μg of pBR322 and 2.5μg of Rot34AI chromosomal DNA was digested with the appropriate endonuclease. The DNA was precipitated following phenol extraction and resuspended in 50μL of TE buffer. 4μL of T_4 ligase ($10U\mu L^{-1}$) was added with ligase buffer and incubated at 8°C overnight.

6.2.4 STRATEGY FOR SCREENING pBR322 LIBRARIES

Aliquots (10μL) of the pBR322 libraries were amplified and plasmids isolated as described in 2.7.2. and 2.12.2. 1μg of the plasmid DNA was subjected to *Tfi*I digestion. The plasmids were then transformed into GM2163 and grown overnight in 5mL broths. The cells were plated onto the appropriate plates i.e., Amp plates for *Bam*HI generated libraries, and Tc plates for *Pst*I generated libraries.

A selection of recombinants were grown overnight in 5mL nutrient broths. 1.5mL of culture was disintegrated and cell extract incubated with λ DNA and restriction buffer at 65°C for restriction endonuclease activity. None of the isolates exhibited any endonuclease activity.

3mL of the cell culture from the recombinants grown overnight were harvested and plasmid DNA was isolated (Section 2.7.3). The plasmid DNA was digested with *Tfi*I to test for methylation of the plasmid. If the clones carried the active methylase gene, the *Tfi*I sites within the vector would have been methylated and would not be cleaved by *Tfi*I. All isolates were found to be digested by *Tfi*I. Therefore no isolate was shown to contain either endonuclease or methylase activity. Table 6.5 show the number of colonies isolated after *Tfi*I digestion and the results of the replica plate tests.

Figure 6.2 shows a selection of the screening procedure of *Sau*3AI pBR322 library. All the clones from this library were of identical size containing a 4200 base pair insert. The appearance of a fragment when the clones were digested with *Tfi*I which electrophoresed just above the 6.5 kb marker corresponds with an insert of 4200 base pair insert with no *Tfi*I sites (Figure 6.2).

Table 6.5 The effect of *Tfi*I digestion on the pBR322 generated libraries

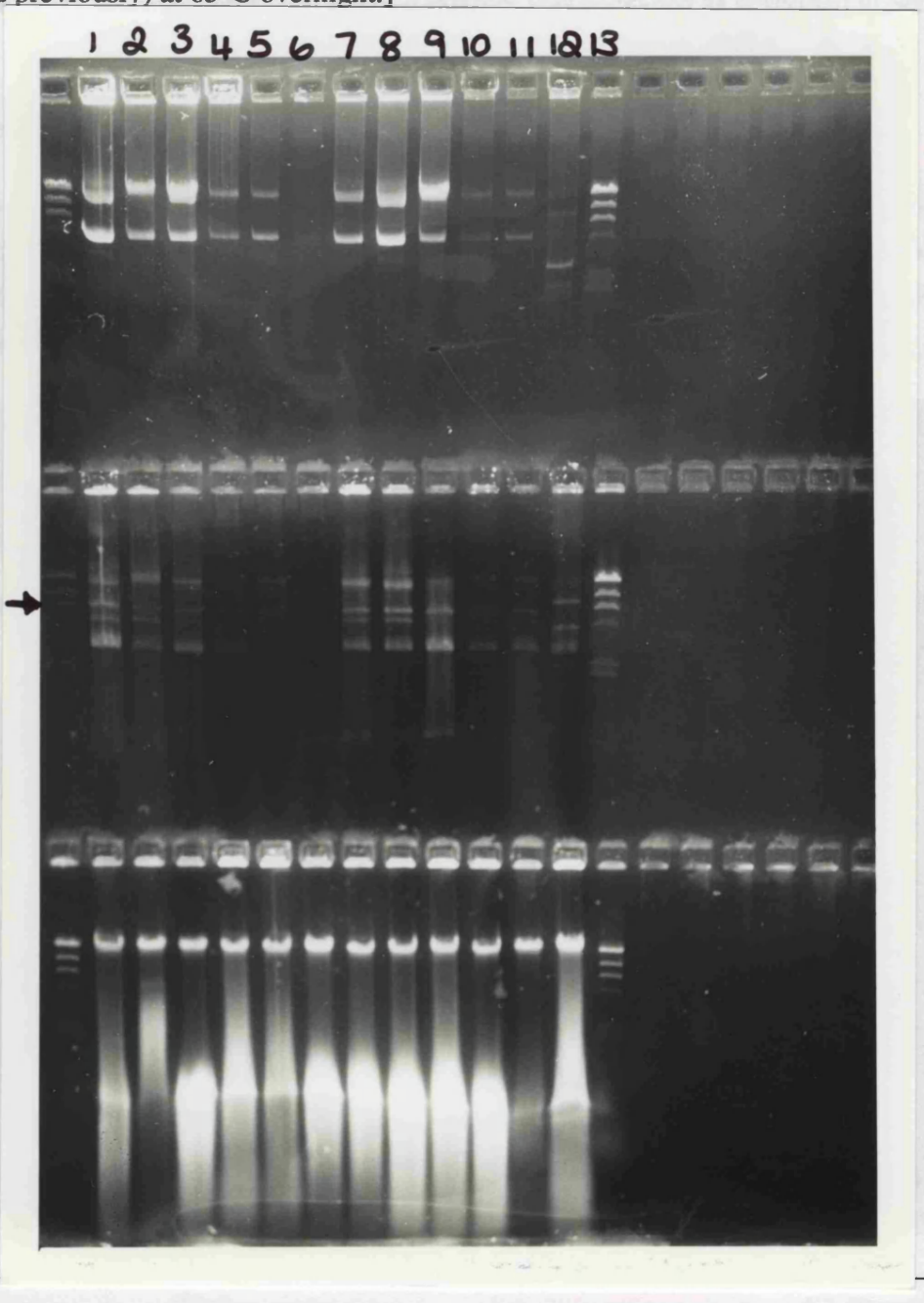
Library	Number of colonies[a]	Number of recombinants[b]
*Pst*I digested chromosomal DNA generated library	74	2
*Pst*I partially digested chromosomal DNA generated library	125	3
*Bam*HI digested chromosomal DNA generated library	100	6
*Sau*3A1 partial digested (4-6kb) generated library	45	12

[a] The number of colonies from 5mL of cell culture, harvested and resuspended in 1mL of nutrient broth and plated onto nutrient agar plates supplemented with the appropriate antibiotic.

[b]The values correspond to the number of recombinants that did not grow on Amp and Tc plates when replica plated. For the *Bam*HI generated libraries recombinants were Amp^R Tc^S and for the *Pst*I libraries recombinants were Amp^S Tc^R.

Figure 6.2. Analysis of pBR322 *Sau*3A1 partial digested generated colonies

[The top row shows 20µL of undigested plasmid DNA isolated from the pBR322 *Sau*3A1 partial digested generated colonies (Lanes 1-12) as described in Section 2.7.3. Lane 13 undigested pBR322 isolated as above. The middle row shows *Tfi*I digestion of 20µL of the plasmid DNA. The 7 kb fragment with the 4.2 kb insert flanked by the pBR322 sequences is shown. The bottom row shows the electrophoresis of samples of 0.5µg of λ DNA incubated with restriction buffer and cell extract (20µL) (prepared from each of the above cell cultures as described previously) at 65°C overnight.]



The construction and the screening of the plasmid libraries did not result in the isolation of the restriction modification genes of Rot34A1.

6.3 CREATING A LIBRARY IN LAMBDA DASH™ II

100µg of chromosomal Rot34A1 DNA was size fractionated as described in Section 2.7.4 (Figure 6.3). The precipitated size fractionated DNA was ligated to Lambda DASH™II (Appendix IV). Figure 6.4 shows aliquots of the ligation before and after incubation at 4°C overnight. The ligation mixture was packaged as described in Section 2.10.

Figure 6.3 Size fractionation of chromosomal Rot34A1 DNA

[DNA was size fractionated as described in 2.7.3. The photograph shows 20µL aliquots of some of the fractions collected from the sucrose gradient electrophoresed in parallel on 0.7% (w/v) agarose gel overnight at 20V. The fraction corresponding to between 9 and 23kb (Fraction 20) was precipitated and resuspended in 3µL of TE buffer.]

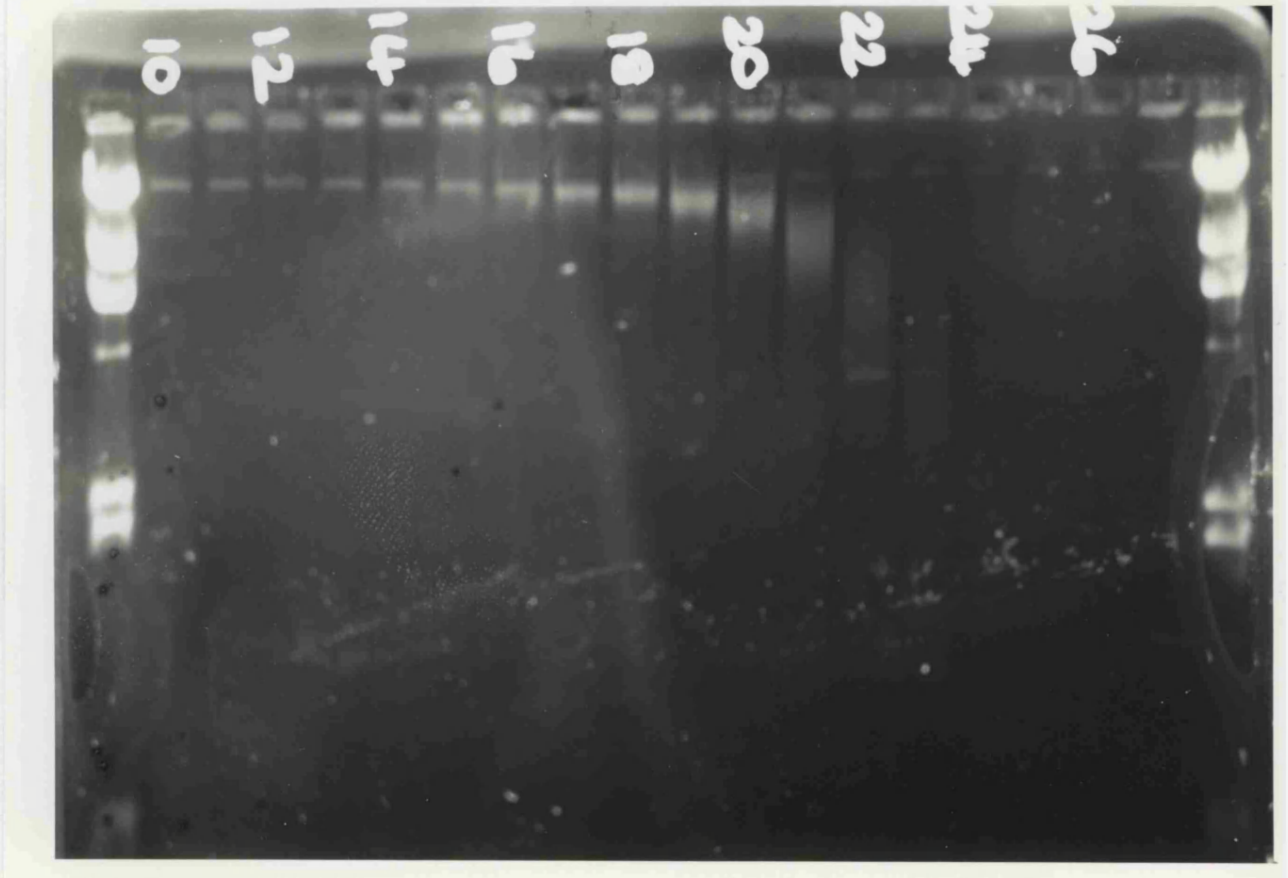
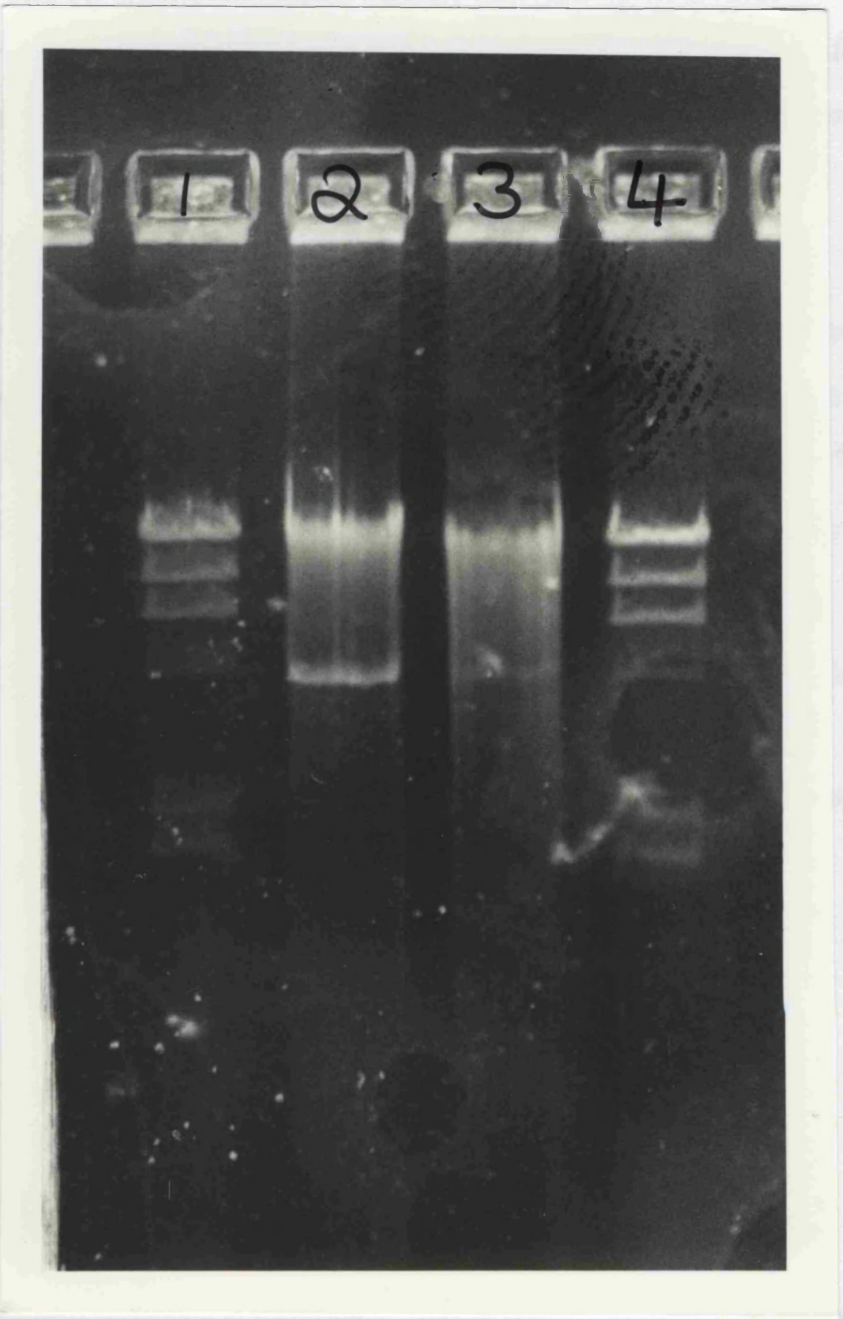


Figure 6.4 Lambda DASH™II ligation with size fractionated chromosomal Rot34A1 DNA

[1µg of Lambda DASH™ which had been pre-digested with *Bam*HI was added to a ligation mixture with the 3µL of size fractionated chromosomal DNA and 0.5µL of ligase (10 unitsμL^{-1}) and ligase buffer. Lane 1, λ-*Hin*dIII; lane 2, pre-ligation aliquot; lane 3, post-ligation aliquot; lane 4, λ-*Hin*dIII.]



Initially, P2392 and LE392 had been used to calculate the frequency of recombinants. However LE392 was found to contain *E. coli* restriction systems (i.e., the *mcrB* restriction system) which have a negative effect on the efficiency of DNA cloning [Stratagene Lambda DASH™ II instruction manual (1993)]. Both recombinant and non recombinant phage will grow on XL1-Blue MRA, but only recombinant phage will grow on XL1-Blue MRA (P2). The XL1-Blue MRA and XL1-Blue MRA (P2) are *mcr*$^-$, *mcrB*$^-$ and *mrr*$^-$; these modifications have been demonstrated to cause up to a 10 fold increase in the yield of recombinant phage containing methylated DNA. In addition, these strains have been further modified to enhance the stability of non-standard DNA structures [Stratagene Lambda DASH™ II instruction manual (1993)].

A reduction in the number of recombinants was observed in the libraries created in both P2392 compared with XL1-Blue MRA (P2). The frequency of recombinants was lower with the P2392 strain and the number of recombinants were also lower. Table 6.6 shows the number of recombinants and their frequency in both strains.

Screening the library was attempted by the method outlined for plasmid libraries was attempted The P2392 library was amplified as described Section 2.12.2. An aliquot (5mL) of the amplified DNA was prepared by the mini prep isolation of λ DNA (Promega Protocols and Applications Guide pp 140-141). The DNA recovered was not of sufficient purity to digest the DNA with *Tfi*I.

Table 6.6 Libraries in Lambda DASH™II in different *E. coli* strains

[An aliquot of the library was serially diluted and transfected into the *E. coli* strains and plated as described in Section 2.11. The frequency of recombinants refers to the percentage of recombinants calculated from transfecting various dilutions of the library into XL1-Blue MRA and LE392.]

Strain	Number of Recombinants	Frequency of recombinants (%)
P2392	2500	60
XL1-Blue MRA (P2)	6900	70

The lambda DASH™II libraries were not screened due to the failure of the endonuclease purification to generate internal sequence data which could be used to design oligonucleotide probes (see Chapter 3). The m^5C-MTases have ten conserved amino acid motifs (see Section 1.7.2) and two of these motifs (DPF-GSGT and TSPPY) have some homology with the m^4C-MTases and the m^6A-MTases. Screening of the library by probes from homology data of m^5C-MTases however was not possible as the back translated DNA sequences of the homology regions were highly degenerate (see Chapter 7).

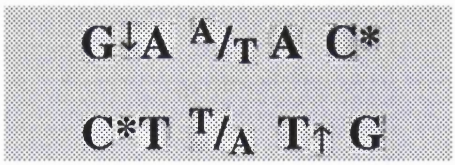
Chapter 7 General Discussion

7.1 INTRODUCTION

Thermus filiformi strain Rot34A1 contains a novel Class II restriction modification system. The endonuclease, available commercially from New England Biolabs, cleaves at a novel specificity, $GA^{A}/_{T}AC$ (Figure 7.1).

Figure 7.1 *Tfi*I Restriction Modification System Cleavage and Methylation Sites

[The arrows indicate the cleavage site of the endonuclease (Cowan *et al.*, unpublished observations) and the asterisks indicate the methylation site by the methylase (this study).]



7.2 PURIFICATION AND CHARACTERISATION OF *Tfi*I ENDONUCLEASE

Chapter Three described the partial purification of *Tfi*I endonuclease. This purification and subsequent characterisation gave an insight in the biochemical properties of *Tfi*I restriction endonuclease.

The purification data showed the endonuclease to have a greater binding capacity for matrices which show some structural similarity to DNA, i.e., phosphocellulose and heparin. Historically phosphocellulose and heparin have been used to purify restriction endonucleases and other DNA binding proteins (Greene *et al.*, 1978; Pirrota and Bickle,

1980; Ward *et al.*, 1991). These pseudo DNA affinity columns have the added advantage of being relatively inexpensive and, in the case of heparin, recyclable.

The steps in the purification were optimised for sequential purification and resulted in the endonuclease having a final activity of 4000 Units $g^{-1}L^{-1}$ of cell culture (Chapter 3, Table 3.7). This value is approximately ten fold lower than those reported for other endonucleases (Lobos and Vásquez, 1993). The final specific activity of *Tfi*I endonuclease was not however a "true" indication of the homogeneous endonuclease specific activity as the final fraction also contained GAPDH activity.

DNA affinity chromatography was not utilised in this study but has greatly facilitated the purification of other sequence specific DNA binding proteins (Kadonaga and Tijan, 1986). The use of DNA as a chromatographic tool for the isolation of restriction endonucleases to homogeneity [Blanks and McLaughlin, 1988; Viatakis and Bouriotis, 1991(a and b)] with high purification factors and recoveries have been reported (Viatakis and Bouriotis, 1991). The affinity binding of the proteins to DNA affinity matrices may have purified the endonuclease from the GAPDH protein. The development of sequence specific DNA matricies (Pozidis *et al.*, 1993) will also increase the ease of purification of restriction endonucleases. The development of sequence specific DNA matricies will also increase the ease of purification of restriction enzymes (Ostrowski and Bomsztyk 1993).

The molecular weight (M_r) of *Tfi*I endonuclease estimated under denaturing conditions (12% SDS-polyacrylamide gel electrophoresis) was 37 000 (Section 4.2). The M_r of native form of *Tfi*I endonuclease as estimated by gel filtration (Superdex 75) was approximately 75 000. Since the native molecular weight is approximately double the determined denatured mass it is concluded that *Tfi*I endonuclease is a dimer in its active form, a composition typical of most Class II endonucleases (Wilson, 1991).

The 37 kDa protein band isolated and internal sequence determined by gel electrophoresis and HPLC analysis was found to be homogenous to the *T. aquaticus* GAPDH enzyme, (Chapter 3 Section 3.10). Subsequent analysis of the final Mono Q fractions revealed GAPDH activity. Based on this information and the initial observations of gel filtration analysis of *Tfi*I endonuclease, two different gel filtration matrices (Superdex 75 and Superose 12) were used to separate the different protein species in the Mono Q fractions. These matrices are primarily analytical and were used successfully to separate the *Tfi*I endonuclease from the contaminating protein. These results came towards the end of the project however and it was not possible to obtain protein sequence data from the purified *Tfi*I endonuclease.

Recovery of the endonuclease was poor with all the gel filtration matrices examined in this study. The endonuclease eluted over a broad size range using Superdex 200 and resulted in over 50% loss of activity. Activity assays of fractions eluted from Superdex 75 chromatography involved incubation times greater than twelve hours. Silver stained SDS PAGE analysis of the active fractions showed very little protein (Chapter 4, Figure 4.4). One reason for poor recovery of the endonuclease was the stability of the endonuclease in buffers containing no glycerol. The addition of 50% (v/v) glycerol to the buffer would have increased the pressure applied to the column above its maximum operating pressure.

Endonucleases can show partial activity in the absence of monovalent ions (Lobos and Vásquez, 1993), while others exhibit star activity in the presence of low salt (Gingeras and Brooks 1983). On assessment of the effect of NaCl, *Tfi*I endonuclease activity was found to decrease with very high NaCl concentrations and it was shown that *Tfi*I endonuclease was active over a broad range of NaCl concentrations; i.e., between 0 and 300mM NaCl. Above 300mM NaCl the endonuclease totally was inhibited (Chapter 4, Section 4.5).

The optimal pH for *Tfi*I endonuclease activity was determined to be 8.0, though *Tfi*I endonuclease exhibits activity over a broad pH range [pH 6.0 and 11 (Chapter 4, Section 4.3)]. This is in common with most restriction endonucleases although some do exhibit star activity at elevated pHs (Polisky *et al.*, 1975). However no star activity was detected at high pH with *Tfi*I endonuclease.

The optimum conditions for catalytic function for an endonuclease can parallel the growth conditions for the organism from which the endonuclease was isolated (although this does not always reproduce the internal environment of the cell). The halophile *Halobacterium halobium* produces an endonuclease that requires 3M NaCl to function (Schinzel and Burger, 1986). *T. filiformis* was isolated from an alkaline hot stream (pH 8.0, 78°C) and *Tfi*I endonuclease reaction conditions mimic certain of the growth conditions of Rot34AI; i.e. exhibiting high tolerance to extreme alkaline pH and temperature. As expected from a thermophilic endonuclease, *Tfi*I endonuclease showed significant thermostability over a range of temperatures (Chapter 4, Section 4.4). Rot34A1 had a growth temperature of 65°C and at this temperature the endonuclease had a half life of greater than 30 minutes. This has been observed for other thermophilic organisms: for example, the *Bst*LVI restriction endonuclease (a thermostable isoschizomer of *Cla*I) has shown wide pH working range and temperature tolerance at higher temperatures (Lobos and Vásquez, 1993). The thermostability may be higher in *vivo* as the cell will contain cofactors that enhance thermostability which are not present *in vitro*.

Not all restriction modification systems show this parallel with activity and growth. The *Dsa* family of Class II endonucleases (*Dsa*I-VI) isolated from *Dactylococcopsis salina* showed differing requirements for optimal activity (Laue *et al.*, 1991). The optimal temperature of the six *Dsa* endonucleases varied from 37°C to 60°C and the preferred

incubation conditions also varied between low and high salt conditions and between different pH values (Laue *et al.*, 1991).

Inactivation of the *Tfi*I endonuclease was observed in the presence of SDS and in Triton X100 (concentrations greater than 1%) under otherwise normal assay conditions (Chapter 4, Section 4.6). Most proteins and enzymes isolated from thermophilic sources have been demonstrated to be stable to protein denaturing agents (Zeikus, 1979), although not necessarily at their optimal activity temperature. In constrast the *Bst*LVI restriction endonuclease did not lose activity when it was assayed in the presence of several denaturing agents such as 30% (v/v) Triton X100; 4% (v/v) formamide and 0.2M urea (Lobos and Vásquez, 1993).

The activity of *Tfi*I endonuclease, as with other class II endonucleases, is influenced by divalent cations [reviewed by Modrich and Roberts (1982)]. The metal ion is required for catalysis rather than DNA binding (Halford and Johnson, 1980; Terry *et al.*, 1983; Taylor *et al.*, 1991; Aiken *et al.*, 1991). In the binding to DNA in the absence of metal ions, different restriction endonucleases show very different levels of specificity for their recognition sites. The *Eco*RI endonuclease binds more tightly to its recognition sequence than to any other sequence (Halford and Johnson, 1980; Terry *et al.*, 1983). In contrast, the absence of divalent ions causes the *Eco*RV endonuclease to bind equally well to specific (cognate) and non specific (non cognate) double stranded sites (Taylor *et al.*, 1991).

The ability of any metal ion to satisfy the binding interactions of the metal ion binding sites depends in part on the radius of the ion. Mg^{2+} ions have been substituted by other divalent ions, though there is usually a reduction in endonucleolytic activity. Typically for endonucleases such as *Eco*RI and *Eco*RV, Mn^{2+} causes about a 20 fold reduction in activity, Co^{2+} causes an even larger reduction, while Ca^{2+} fails to support any activity

(Vermote and Halford, 1992). The failure of Ca^{2+} to support activity could be because Ca^{2+} has a greater effective radius than the other divalent metals [Mg^{2+}, Co^{2+}, Cu^{2+}, and Mn^{2+} (Kretsinger and Nelson 1976; Khoo *et al.*, 1984)] and this might influence binding and catalysis. *Tfi*I endonuclease showed no activity in the presence of Co^{2+}, Cu^{2+} Ca^{2+} and Zn^{2+}. Binding studies of *Tfi*I endonuclease to DNA was not possible as all characterisation involved using partially purified endonuclease preparations (see Chapter Four for experimental detail). Endonuclease activity was detected with only Mg^{2+} and Mn^{2+} ions. Substitution of Mg^{2+} ions with Mn^{2+} ions resulted in approximately 50% decrease in endonucleolytic activity. Mn^{2+} ions also caused reduced specificity. These observations are similar to those of other Class II restriction endonucleases (Vermote and Halford, 1992; Luke *et al.*, 1987; Selent *et al.*, 1992).

Star activity of *Tfi*I endonuclease was also observed in the presence of certain chemicals, such as 30% (v/v) glycerol and in low ionic strength buffers (Chapter 4, Section 4.7). There was a difference in this activity compared with the star activity of the Mn^{2+}-induced endonuclease. The end products of digestion varied for different experiments suggesting that, unlike the situation with Mn^{2+}, the endonuclease did not discriminate for any particular site. This observation is similar to the star activity of other endonucleases (Izsvák and Duda, 1989).

Robinson and Sligar (1993) have proposed a mechanism for star activity. They demonstrated that water activity affected site specific recognition of DNA by *Eco*RI, and that the extent of *Eco*RI star activity depended strongly upon this activity. The loss of specificity accompanied decreased water activity, suggesting one or more water molecules were implicated as part of the molecular basis of specificity. The authors suggested a model for the effect of water activity where "bound" water was involved in specificity; release of this bound water led to the decreased specificity manifested as star activity.

One candidate for this "bound" water in star activity is the resident water that is complexed in the crystal structure of *Eco*RI with the oligo-deoxyribonucleotide containing its target recognition sequence. The bound water molecule was thought to be involved in a bridged hydrogen bond between Arg^{200} and Arg^{203} in *Eco*RI and G_1 in the recognition sequence. Increased osmotic pressure would serve to draw this water out of the protein-DNA interface. The loss of this water-mediated contact between enzyme and substrate may be responsible for the observed loss of specificity (Robinson and Sligar, 1993). Mutations at Arg^{200} have also been shown to relax the specificity of *Eco*RI, by removing protein-DNA interactions (Heitman and Model 1990;). This has been confirmed by homology and crystallography studies (Kim *et al.*, 1990; Siksnys *et al.* 1994).

Mn^{2+} induced star activity shows relaxed specificity but does still discriminate between sites. Vermote and Halford (1992) suggested that Mg^{2+} "locks" the enzyme onto the DNA at its cognate site whereas Mn^{2+} is thought to "trap" the enzyme at either cognate or non cognate sites. However the non cognate sites may only differ from the recognition sequence by one or two bases and cleavage does occur *in vivo* (Taylor and Halford, 1992). The atomic radius of Mn^{2+} is greater than Mg^{2+} (Shannon, 1976) and this probably causes less specific interactions between the protein and DNA.

The inability of endonucleases to discriminate between cognate and non cognate sites in the presence of Mn^{2+} has been observed for other enzymes involved in DNA metabolism. For example, Mn^{2+} enhances the rate at which DNA polymerases utilise either non complementary nucleotides or nucleotide analogues (Kunkel and Loeb, 1979; Tabor and Richardson, 1989) and recently this has been exploited to allow *Taq* DNA polymerase to utilise mRNA as a template, rather than DNA (Jones, 1993).

Some restriction endonucleases can produce site specific cleavage on single stranded substrates (Horiuchi and Zinder, 1975; Blakesley and Wells, 1975; Godson and Roberts, 1976). These reactions are usually at 10 to 50% of the rate for double stranded DNA cleavage. Detailed mechanistic studies, however, demonstrated that "single strand" cleavage probably occurs at transitionally formed double-stranded sites (Blakesly *et al.*, 1977). Most restriction endonucleases cleave single stranded DNA with low specificity. The non specific digestion of single stranded DNA by *Tfi*I (Chapter 4, Section 4.10) is similar to other Class II endonucleases.

7.3 CHARACTERISATION OF THE *Tfi*I METHYLASE

The purification steps for the *Tfi*I methylase (consisting of phosphocellulose chromatography followed by heparin chromatography) did not result in a homogenous preparation. The steps were not optimal for recovery (Chapter 5, Table 5.1). Optimisation was hindered by the low yield of the methylase coupled with the apparent instability of the methylase upon purification. This led to lengthy incubation times to detect and measure activity. The initial phosphocellulose purification step, separating the methylase from the endonuclease has been used successfully in purification strategies for other restriction modification systems (McClelland, 1981). Other groups have also reported low yields of the methylase enzyme from restriction modification systems (Lobos and Vásquez, 1993).

The low yield of the methylase compared with the endonuclease could be the result of the cells regulation of restriction modification enzymes. Cells have different ways of regulating production of restriction modification enzymes. One mechanism is that the recognition sequences for restriction modification systems in the genome of host organisms are often statistically under-represented (Schroeder *et al.*, 1986; Krüger *et al.*, 1989; Ehrlich and Wang, 1981). Hence the cells methylase requirements are relatively

low. This mechanism could exist in Rot34A1 as attempts to identify the methylated base by hydrolysis and subsequent thin layer chromatography analysis of chromosomal Rot34A1 DNA were not sucessful (data not shown).

Determination of the methylation site of the *Tfi*I methylase was achieved by thin layer chromatography after hydrolysis of *Tfi*I methylated DNA. The chromatography (Chapter 5, Table 5.3) showed that the *Tfi*I methylase methylated the cytosine 3' end of the *Tfi*I recognition sequence ($GA^{A}/_{T}AC$) and methylated on the ^{5}C. Based on the definitions defined by Pósfai *et al.* (1989), this would suggest that *Tfi*I methylase is a member of the $m^{5}C$-MTase class.

7.4 CLONING STRATAGIES FOR ISOLATING THE *Tfi*I RESTRICTION MODIFICATION GENES

The perfect genomic library would contain DNA sequences representative of the entire genome, in a stable form, as a manageable number of overlapping clones. The cloned fragments would be large enough to contain whole genes and their flanking sequences. On the other hand they should be small enough to be mapped easily by restriction enzyme analysis. Most important, a library should be both easy to construct from small amounts of starting material and easy to screen for the sequence of interest (Kaiser and Murray, 1986).

The lack of success in the isolation of either the genes for the *Tfi*I restriction modification system was partly due to the size and number of libraries screened. Of the seven plasmid libraries constructed in this study using three different restriction enzymes, three had fewer than 4000 colonies (Chapter 6, Table 6.2 and 6.4). None were found to contain the genes of *Tfi*I restriction modification system. In comparison, the cloning of the *Sfi*I restriction modification system involved using a total of sixteen different restriction endonucleases to construct fifty four libraries (VanCott 1990).

As discussed in Chapter 6, there were problems associated with screening the libraries. The initial choice of host vector, GM2163, meant that there was no blue/white visual screen for the pUC derived libraries in this strain and this in turn led to difficulties in screening the libraries. Other *E. coli* strains which are *mcrC*$^-$, *mrr*$^-$ and *mcrBC*$^-$, but do contain visual screens, would have been more appropriate hosts.

Rot34AI DNA contains methylated cytosines. Methylated DNA affects the transfection frequency into *E. coli* (Wang *et al.*, 1984). In this study it was observed that methylated DNA is transformed into strain JM107 at a lower frequency than unmethylated DNAs, agreeing with findings from Blumenthal *et al.* (1985). JM107 may contain a weakly expressed *mcrB* system which induces restriction of the methylated chromosomal DNA by inducing SOS repair in the cell. Thus the recombinants would lose their methylated DNA by restriction.

The effect of restriction and modification systems on plasmid transformation has been reported: Tanaka (1979) reported a high interference on transformation. Bron *et al.* (1988) found little effect. Studies have shown most strains of *E. coli* appear to be methylase sensitive, in that recombinant plasmids carrying the active methylase gene transform poorly into these strains (Blumenthal *et al.*, 1985). The identification of the *mcrC, mrr* and *mcrBC* systems (Raleigh and Wilson, 1986; Raleigh *et al.*,1988; Dila *et al.*,1990) have given some insight into why the transformation of modified DNA is not efficient in some strains of *E. coli*. The restriction modification system of *Proteus vulgaris* was cloned into *E. coli* strains HB101 and RR1 which are known to have mutated *mcrC* and *mrr* genes (Sambrooks *et al.*, 1989). Many restriction modification systems have been cloned into HB101 after unsuccessful attempts at cloning into other hosts (Walder *et al.*, 1981; Maekawa *et al.*, 1990).

The *Bam*HI restriction modification system is regulated by a small intergenic open reading frame, *bamHIC*, in both *E. coli* and *B. subtilis*. The *bamHIC* was found to affect *McrBC* restriction of methylase containing plasmids. The transformation efficiency of methylase containing plasmids into the *McrBC*$^+$ host was 100 fold lower than that seen for *McrBC*$^-$ host (Ives *et al.*, 1992). The presence of *bamHIC* on a second plasmid however alleviated *McrBC* restriction; transformation efficiency increased 100 fold. This suggests the control of the *McrBC* is similar to that of the *Bam*HI restriction modification system and other restriction modification systems.

VanCott (1990) found transformation of host cells at different temperatures was required for the initial cloning of the *Sfi*I restriction modification system. For example pUC19 derived libraries were grown at 42°C when attempts to grow at 37°C were found to be unsuccessful. It was thought that since the optimal temperature for *Sfi*I restriction endonuclease is 50°C it might be similarly so for the methylase. The first methylase clone was obtained at 42°C. (After successful cloning it was determined that the optimal *in vivo* methylase activity in *E. coli* was 37°C and the optimal *in vitro* temperature was 40°C).

The *Tfi*I restriction modification system is a thermophilic system; growth of the transformed pUC-derived recombinants at slightly higher temperatures may have facilitated adequate expression of the methylase to confer resistance. In some systems the failure to clone a restriction modification system stemmed from the level of expression of the methylase gene (Howard *et al.*, 1986; Brooks *et al.*, 1989; Hammond *et al.*, 1990; Düsterhöft *et al.*, 1991(a); Erdmann *et al.*, 1991).

In this study, the number of recombinants varied between the pUC-derived and the pBR322 libraries (Chapter 6 Tables 6.2 and 6.3). pBR322 derived clones were found to transform *E. coli* with more efficiency than the pUC derived clones. This was also observed by VanCott (1990) and may reflect the greater stability of pBR322.

VanCott (1990) constructed five different vectors in attempts to clone the *Sfi*I restriction system; two pBR322 derivatives, two pUC19 derivatives and a pBR328 derivative. The latter yielded the only active *Sfi*I methylase clone. It was believed that cloning into pBR328 was successful because it had a higher copy number than the stable plasmid pBR322 and greater stability than the very high copy plasmid pUC19 (VanCott 1990). Kapfer *et al.*, (1991) cloned the *Bsu*FI restriction endonuclease gene into a low copy number, but stable plasmid.

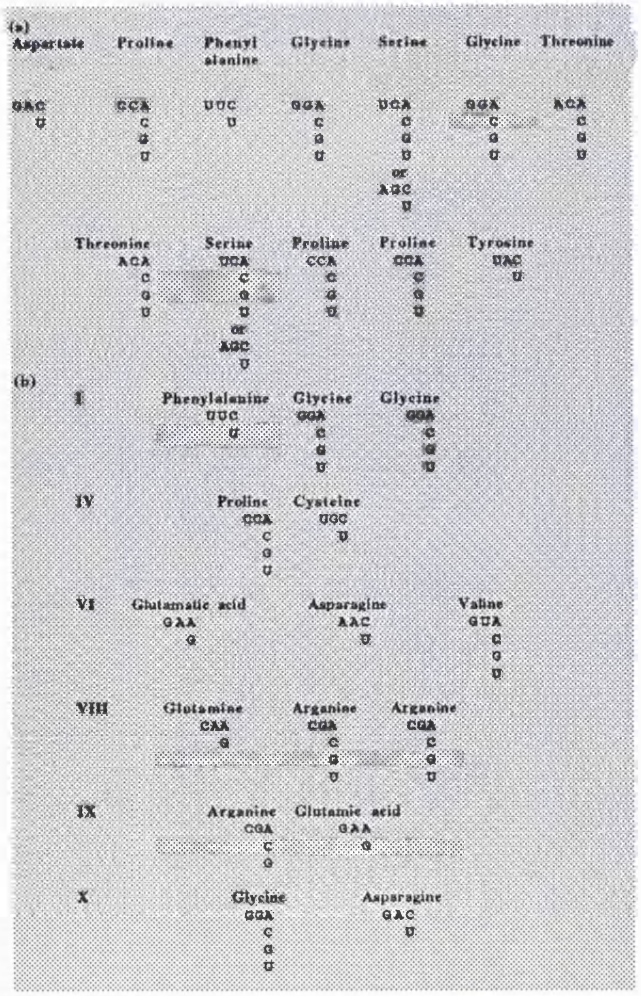
The problems associated in isolating restriction modification genes can be illustrated by the cloning of the *Hgi*CII restriction modification system. The genes, *hgiCIIR* and *hgiCIIM* which encode the *Hgi*CII restriction modification system were seemingly unclonable using a stepwise cloning method. They were finally cloned using a PCR-based technique (Erdmann *et al.*, 1992).

The *hgiCIIM* was cloned using selection based on modification using an isoschizomer of *Hgi*CII, *Ava*II. However all attempts to clone the closely linked *hgiCIIR* and *hgiCIIM* genes in a single step resulted in deletions spanning parts of the coding regions of *hgiCIIR*. Cloning of the missing 3' terminal part of this gene was achieved by applying an inverted PCR technique. All attempts to construct an enzymatically active endonuclease failed (Erdmann *et al.*, 1992). This is in contrast to the restriction modification system of *Hgi*BI from *H. giganteus* strain *Hpg*5 which was cloned in a single step [Düsterhölt *et al.*, 1991(b)]. This suggests that problems in the isolation of restriction modification genes are not just species dependent but also system dependent.

Future work in this study for the isolation of the *Tfi*I restriction modification system genes would involve screening the lambda library (Chapter 6). The internal sequence of the endonuclease was not elucidated (Chapter 3) and hindered plans to screen the lambda library with labelled oligonucleotide probes based on the protein sequence. The final

separation of the endonuclease by gel filtration did not yield sufficient protein to gain any internal or N terminal sequence data and was obtained too late in the project. An alternative screening strategy using the high degree of homology between the MTases was considered. The m^5C-MTases share ten common motifs of which five regions are highly conserved and also share two homologous motifs with the m^4C-MTases and m^6A-MTases (Figure 7.2).

Figure 7.2 Conserved regions within the methylases
[(a) Conserved motifs shared by all the methylases (i.e., m^5C-MTases, m^4C-MTases and m^6A-MTases (Klimasauskas *et al.,* 1989) (b) The six highly conserved motifs (I, IV, VIII, IX and X) shared by all the m^5C-MTases (Pósfai *et al.,* 1989). Below the back translated DNA sequences are shown]



Originally, when the methylation site of the *Tfi*I methylase was not known, probes designed to these motifs (Figure 7.2a) were examined. However the design of primers to these regions showed that the back translated DNA sequences are extremely degenerate (Figure 7.2a). Probes designed to the six conserved motifs shared by the m^5C-methylases (Figure 7.2b) would be another route in the isolation of the restriction modification genes. One consideration is the codon preference in restriction modification systems which do not always adhere to the codon usage of the host organism, thus making this a problematic strategy.

7.5 CONCLUSIONS

In conclusion,

(i) Rot34A1 (*T. filiformis* strain) contains a novel thermostable restriction modification system.

(ii) The methylase recognition site of the *Tfi*I methylase has been elucidated.

(iii) Various characteristics of *Tfi*I endonuclease have been determined, such as molecular weight, thermostability, pH, ability to cleave single stranded DNA and salt requirement (Table 7.1).

(iii) Star activity of the endonuclease has been investigated under different conditions. Two kinds of star activity were found; the indiscriminate endonucleolytic activity exhibited in certain buffers and the relaxed specificity exhibited by Mn^{2+} induced star activity.

Table 7.1 Characteristics of *Tfi*I endonuclease

Native molecular weight (M_r)	75 000
Subunit molecular weight (M_r)	37 000
pH optimum	8.0
Thermostability	half life greater than 40 minutes at 65°C
Salt requirement	None (inhibited at NaCl concentrations greater than 300mM)

(iv) Although this study did not isolate the genes for restriction and modification from the plasmid libraries screened, the creation of a lambda phage library does meet the requirements of a complete genomic library. Further work includes screening this library.

Chapter Eight References

Achenbach-Richer, L., Gugta, R., Stetter, K.O. and Woese, C.R. (1987) Were the original eubacteria thermophiles? *Syst. Appli. Microbiol.* **9**, 34-39.

Aggarwal, A.K. (1995) Structure and function of restriction endonucleases. *Curr. Opin. Struc. Biol.* **5**, 11-19.

Aiken, C., Fisher, E.W. and Gumport, R.I. (1991) The specific binding, bending and unwinding of DNA by *Rsr*I endonuclease, an isoschizomer of *Eco*RI endonuclease *J. Biol. Chem.* **266**, 19063-19069.

Arber, W. and Linn, S. (1969). DNA modification and restriction. *Ann. Rev. Biochem.* **38**, 467-500.

Arnone, M.I., Birolo, L., Cubellis, M.V., Nitti, G., Marino, G. and Sannia, G. (1992) Expression of a hyperthermophilic aspartate aminotransferase in *Escherichia coli. Biochim. Biophys. Acta* **1160**, 206-212.

Athanasiadis, A., Vlassi, M., Kotsifaki, D., Tucker, P.A., Wilson, K.S. and Kokkinidis, M. (1994) Crystal structure of *Pvu*II reveals extensive homologies to *Eco*RV. *Nature Struct. Biol.* **1**, 469-475.

Barany, F. (1988) How *Taq*I endonuclease recognizes its cognate sequence. *Gene* **74**, 63-65.

Barany, F., Danzitz, M., Zebala, J. and Mayer, A. (1992a) Cloning and sequencing of genes encoding the *Tth*HB81 restriction and modification enzymes: comparison with the isochizomeric *Taq*I enzymes. *Gene* **112**, 3-12.

Barany, F., Slatko, B., Danzitz, M., Cowburn, D., Schildkraut, I. and Wilson, G.G. (1992b) The corrected nucleotide sequences of the *Taq*I restriction and modification enzymes reveal a thirteen codon overlap. *Gene* **112**, 91-95.

Bellemare, G. and Potvin, C. (1990) Classification of type II restriction endonucleases and cloning of non identical cohesive end fragments without self polymerisation using non palindromic oligodeoxyribonucleotide adapters. *Gene* **101**, 67-74.

Bergquist, P.I. and Morgan, H.W. (1992) The molecular genetics and biotechnological application of enzymes from extremely thermophilic eubacteria. Ch. 2 in Molecular Biology and Biotechnology of Extremophiles, Eds. Herbert, R.A. and Sharp, R.J. Published by Blackie and Son Limited Ist Ed. pp 44-76.

Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**, 1513-23.

Blakesley, R.W., and Wells, R.D. (1975) "Single stranded" DNA from fX174 and M13 is cleaved by certain restriction endonucleases. *Nature* **257**, 421-422.

Blakesley, R.W., Dodgson, J.B., Nes, I.F., and Wells, R.D. (1977) Duplex regions in "single stranded" øX174 are cleaved by a restriction endonuclease from *Haemophilus aegypticus. J. Biol. Chem.* **252**, 7300-7306.

Blanks, R. and McLaughlin, L.W. (1988) Anoligonucleotide affinity column for the isolation of sequence specific DNA binding proteins. *Nucleic Acids Res.* **16**, 10283-99.

Blumenthal, R.M., Gregery, S.A. and Cooperider, J.S. (1985) Cloning of a restriction modification system from *Proteus vulgaris* and it's uses in analysing a methylase sensitive phenotype in *Escherichia coli* . *J. Bacteriol.* **164**, 501-509.

Bocklage, H., Heeger, K. and Muller-Hill, B. (1991) Cloning and characterization of *Mbo*II restriction modification system. *Nucleic Acids Res.* **19**, 1007-1013.

Bonnerjea, J., Oh, S., Hoare, M. and Dunnill, P. (1986) Protein purification - the right step at the right time. *Bio/technology* **4**, 954-958.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-54.

Brock, T.D.and Freeze, H. (1969) Genus *Thermus*. *J. Bacteriol.* **98**, 289-297.

Bron, S., Jannière, L. and Ehrlich, S.D. (1988) Restriction and modification in *Bacillus subtilis* Marburg 168: Target sites and effect on plasmid transformation. *Mol. Gen. Genet.* **211**, 186-189.

Brooks, J.E., Benner, J.S., Heiter, D.F., Silber, K.R., Sznyter, L.A., Jager-Quinton, T., Moron, L.S., Slatko, B.E., Wilson, G.G. and Nwankwo, D.O. (1989) Cloning the *Bam*HI restriction modification system. *Nucleic Acids Res.* **17**, 979-997.

Brooks, J.E., Nathan, P.D., Landry, D., Sznyter, L.A., Waite-Rees, P., Ives, C.L., Morgan, S.L., Slatko, B.E. and Benner, J.S. (1991) Characterization of the cloned *Bam* HI restriction modification system: its nucleotide sequence, properties of the methylase and expression in hetrologous hosts. *Nucleic Acids Res.* **19**, 841-850.

Chandrasegaren, S. & Smith, H.O. (1988) Amino acid sequence homologies among twenty five restriction endonucleases and methylases. Structure & Expression Volume I: In *Proteins to Ribosomes*, Eds., Sarma, R.H. & Sarma, M.H. Adenine Press pp 149-156.

Cheng, X., Balendiran, K., Schildkraut, I., Anderson, J.E. (1994) Structure of *Pvu*II endonuclease with cognate DNA. *EMBO J.* **13**, 3927-3935.

Cheng, X., Kumar, S., Pósfai, J., Pflugrath, J.W. and Roberts,R.J. (1993) Crystal structure of the *Hha*I DNA methyltransferase complexed with s-adenosyl-L-methionine. *Cell* **74**, 299-307.

Clarke, L. and Carbon, J. (1976) A colony bank containing synthetic *Col*EI hyrid plasmids representative of the *E.coli* genome. Cell **9**, 91-9.

Cowan, D.A. *et al.* (unpublished observations).

Cowlondre, C., Miller, J.H. Farabaugh, P.J. and Gilbert, W. (1978) Molecular basis of base substitution hotspots in *Escherichia coli. Nature* **274**, 775-790.

Dalgaard, J.Z., Garrett, R.A. and Belfort, M. (1993) A site specific endonuclease encoded by a typical archaeal intron. *Proc. Natl. Acad. Sci. USA.* **90**, 5414-5417.

Dawson, R.M.C., Elliot, D.C., Elliot, W.H. and Jones, K.M. (1986) pH, buffers and physiological media. Ch. 18 in *Data for Biochemical Research*, 3rd Ed. Oxford Science Publications pp 426- 441.

De-Backer, O. and Colson, C. (1991) Transfer of the genes for the *Sty*LTI restriction modification system of *Salmonella typhimurium* to strains lacking modification ability results in death of recipient cells and degradation of their DNA. *J.Bacteriol.* **173**, 1328-1330.

Degtyarev, S.K., Rechkunova, N.I., Kolyhalov, A.A., Dedkov, V.S. and Zhilkin, P.A. (1990) II-Q restriction endonucleases - new class of type II enzymes. *Nucleic Acids Res.* **18**, 5807-5810.

Dila, D., Suntherland, E., Moran, L., Slatko, B. and Raleigh, E.A. (1990) Genetic and sequence organization of the *mcr* BC locus of *Escherichia coli* . *J. Bacteriol.* **111**, 4888-4900.

Düsterholt, A., Erdmann, D. and Kröger, M. (1991a) Stepwise cloning and molecular characterization of the *Hgi* DI restriction modification system from *Herpetiosiphon giganteueus* Hpa 2. *Nucleic Acids Res* . **19**, 1049-1056.

Düsterholt, A., Erdmann, D. and Kröger, M. (1991b) Isolation and genetic structure of the *Ava*II isoschizomeric restriction modification system *Hgi*BI from *Herpetiosiphon giganteueus Hpg*5: M·*Hgi*BI reveals high homology to M·*Ban*I. *Nucleic Acids Res.* **19**, 3207-3211.

Ehrlich, M. and Wang, R.Y. (1981) 5-methylcytosine in eukaryotic DNA. *Science* **212**, 1350-7.

Ehrlich, M., Gama-Sosa, M.A., Carreira, L.H., Ljungdahl, L.G., Kuo, K.C. and Gehrke, C.W. (1985) DNA methylation in thermophilic bacteria: N^4 methylcytosine, 5-methylcytosine and N^6-methyladenine. *Nucleic Acids Res.* **13**, 1399-1412.

Erdmann, D., Düsterhöft, A. and Kröger, M. (1991) Cloning and molecular characterisation of the *Hgi*CI restriction modification system from *Herpetosiphon giganteus* strain *Hpg*9 reveals high similarities to the isoschizomeric restriction modification system *Ban*I. *Eur. J. Biochem.* **202**, 1247-1256l.

Erdmann, D., Horst, G., Düsterhöft, A. and Kröger, M. (1992) Stepwise cloning and genetic organisation of the seemingly unclonable *Hgi*CII restriction modification system from *Herpetosiphon giganteus* strain *Hpg*9, using PCR technique. *Gene* **117**, 15-22.

Fish, N.M. and Lilly, M.D. (1984) The interactions between fermentation and protein recovery. *Bio/technology* **2**, 623-627.

Fomenkov, A., Xiao, J., Dila, D., Raleigh, E. and Xu, S. (1994) The "endo-blue method" for direct cloning of restriction endonuclease genes in *E.coli. Nucleic Acids Res.* **22**, 2399-2403.

Fredrick, C.A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobson, L., Wang, B.C., Greene, P.J., Boyer, H.W. and Rosenberg, J.M. (1984) Kinked DNA in the crystalline complex with *Eco*RI endonuclease. *Nature* **309**, 327-331.

Gabelsberger, J., Liebl, W. and Schleifer, K-H. (1993) Purification and properties of recombinant β-glucosidase of the hyperthermophilic bacterium *Thermotoga maritima. Appl. Microbiol. Biotechnol.* **40**, 44-52.

Georgia, G., Smith, K.E. and Williams, R.A.D. (1993) DNA:DNA homology and cellular components of *Thermus filiformis* and other strains of *Thermus* from New Zealand hot springs. *FEMS Micro. Lett.* **107**, 145-150.

Gingeras, T.R. and Brooks, J.E. (1983) Cloned restriction/modification system from *Pseudomonas aeruginosa. Proc. Natl. Acad. Sci. USA* **80**, 402-406.

Godson, G.N. and Roberts, R.J. (1976) A catalogue of cleavages of øX174, S13, G4 and ST-1 DNA by 26 different restriction endonucleases. *Virology* **73**, 561-567.

Greene, P.J., Heyneker, H.L., Bolivar, F., Rodriguez, R.L., Betlach, M.C., Covarrubias, A.A., Backman, K., Russel, D.J., Tait, R. and Boyer, B.W. (1978) A general method for the purification of restriction enzymes *Nucleic Acids Res.* **5**, 2373-2380.

Halford, S.E. and Johnson, N.P. (1980) The *Eco*RI restriction endonuclease with bacteriophage λ DNA. *Biochem. J.* **191**, 593-604.

Hammond, A.W., Gerard, G.F. and Chatterjee, D.K. (1990) Cloning the *Kpn*I restriction modification system in *Escherichia coli*. *Gene* **97**, 97-102.

Hartman, R.K., Wolters, J., Kröger, B., Schultze, S., Specht, T. and Erdmann, V.A. (1989) Does *Thermus* represent another deep eubacterial branching? *Syst. Appli. Microbiol.* **11**, 243-249

Hartmann, R.K., Ulbrich, N. and Erdmann, V.A. (1987) An unusual operon constellation: in Thermus thermophilus HB8 the 23S/5S rRNA operon is a separate entity from the 16S rRNA operon. *Biochemie* **69**, 1097-104.

Heitman, J. and Model, P. (1987) Site specific methylases induce the SOS DNA repair response in *Escherichia coli. J. Bacteriol.* **169**, 3243-3250.

Heitman, J. and Model, P. (1990) Mutants of the *Eco*RI endonuclease with promiscuous substrate specificity implicate residues involved in substrate recognition. *EMBO J.* **9**, 3369-78

Herrero, A., Elhai, J., Holn, B. and Wolk, C.P. (1984), Infrequent cleavage of cloned *Anabaena variabilis* DNA by restriction endonucleases from *A. variabilis. J. Bacteriol.* **160**, 781-784.

Horiuchi, K. and Zinder, N.D. (1975) Site specific cleavage of single stranded DNA by a *Haemophilus* restriction endonuclease. *Proc. Natl. Acad. Sci. USA* **72**, 561-567.

Howard, K.A., Card, C., Benner, J.S., Callahan, H.L., Maunus, R., Silber, K., Wilson, G. and Brooks, J.E. (1986) Cloning the *Dde*I restriction modification system using a two step method. *Nucleic Acids Res.* **14**, 7939-7951.

Hsu, M. and Berg. P. (1978) Altering the specificity of restriction endonuclease: effect of replacing Mg^{2+} with Mn^{2+}. *Biochemistry* **17**, 131-138.

Huber, R. and Stetter, K.O. (1992) The Thermotogales: hyperthermophilic and extremely thermophilic bacteria. Ch. 7 in *Thermophilic Bacteria*, Ed. Kristjansson, J.K. CRC Press, Boca Raton, FL. pp 185-195.

Hudson, J.A., Morgan, H.W. and Daniel, R.M. (1984) Isolation and characterisation of a new caldactive filamentous bacterium. *FEMS Micro. Lett.* **22**, 149-153.

Iszvák, Z. and Duda, E. (1989) "Star" activity and complete loss of specificity of *Ceq*I endonuclease. *Biochem. J.* **258**, 301-303.

Ives, C.L., Nathan, P.D. and Brooks, J.E. (1992) Regulation of the *Bam*HI restriction modification system by a small intergenic open reading frame, *bam*HIC, in both *Escherichia coli* and *Bacillus subtilis*. *J. Bacteriol.* **174**, 7194-7201.

Jack, W.E., Greenough, L., Dorner, L.F., Xu, Shuang-yong, Strzelecka, T., Aggarwal, A.K. and Schildkraut, I. (1991) Overexpression, purification and crystallization of *Bam*HI endonuclease. *Nucleic Acids Res.* **19**, 1825-1829.

Janulaitis, A.A., Klimasauskas, S., Petrusyte, M., and Butkus, V. (1983) Cytosine modification in DNA by *Bcn*I methylase yields N^4 methylcytosine. *FEBS Lett.* **161**, 131-134.

Jerpseth, B., Greener, A., Short, J.M., Viola, J. and Kretz, P.L. (1992) XL1-Blue MRF' *E. coli* cells: *McrA*$^-$, *McrCB*$^-$, *McrF*$^-$, *Mrr*$^-$, *HsdR*$^-$ derivative of XL1-Blue cells. *Strategies* **5** (3), 81-83.

Jones, M.D. (1993) Reverse transcription of mRNA by *Thermus aquaticus* DNA polymerase followed by polyerase chain reaction amplification. *Meth. Enzymol.* **218**, 413-9.

Kadonaga, J.T. and Tijan, R. (1986) Affinity purification of sequence specific DNA binding proteins. *Proc. Natl. Acad. Sci. USA* **83**, 5889-93.

Kaiser, K. and Murray, N.E. (1986) The use of phage lambda replacement vectors in the construction of representative genomic DNA libraries Ch. 1 in *DNA Cloning Volume 1 a practical approach.* Ed. Glover, D.M. IRL Press pp 1-48.

Kapfer, W., Walter, J. and Trautner, T.A. (1991) Cloning, characterization and evolution of the *Bsu*FI restriction endonuclease gene of *Bacillus subtilis* and purification of the enzyme. *Nucleic Acids Res.* **19**, 6457-6463.

Karyagina, A.S., Lunin, V.G. and Nikolskaya, I.I. (1990) Characterization of the genetic determinants of *Sso*II restriction endonuclease and modification methyltransferases. *Gene* **87**, 113-118.

Karyagina, A.S., Lunin, V.G., Degtyyarenko, K.N., Uvarov, V.Y. and Nikolskaya, I.I. (1993) Analysis of the nucleotide and derived amino acid sequences of the *Sso*II restriction endonuclease and methyltransferase. *Gene* **124**,13-19.

Kaszubska, W., Aiken, C., O'Connor, C.D. and Gumport, R.I. (1989) Purification, cloning and sequence analysis of *Rsr*I DNA methyltransferase: lack of homology between two enzymes, *Rsr*I and *Eco*RI, that methylate the same nucleotide in identical recognition sequences. *Nucleic Acid Res.* **17**, 10403-25.

Kessler, C. and Höltke, H.J.(1986) Specificity of restriction endonucleases and methylases - a review (Edition 2). *Gene* **47**, 1-153.

Kessler, C. and Manta, V. (1990) Specificity of restriction endonucleases and DNA modification methyltransferases - a review (Edition 3). *Gene* **92**, 1-248.

Khoo, T.C., Cowan, D.A., Daniel, R.M. and Morgan, H.W. (1984) Interactions of calcium and other metal ions with caldolysin, the thermostable proteinase from *Thermus aquaticus* strain T351. *Biochem. J.* **221**, 407-413.

Kim, Y-C., Grable, J.C., Love, R., Greene, P.J. and Rosenberg, J.M. (1990) Refinement of *Eco*RI endonuclease crystal structure: a revised protein chain tracing. *Science* **249**, 1307-1308.

Kiss, A., Pósfai, G., Keller, C.C., Venetianer, P. and Roberts, R.J. (1985) Nucleotide sequence of the *Bsu*RI restriction modification system. *Nucleic Acids Res.* **13**, 6403-21.

Klimasauskas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994) *Hha*I methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357-369.

Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus,V. and Janulaitis, A. (1989) Sequence motifs characteristic of DNA (cytosine-N4) methyltransferase: similarity to adenine and cytosine C5 DNA methylases. *Nucleic Acids Res.* **17**, 9823-9836.

Kong, H., Morgan, R.D., Maunus, R.E. and Schildkraut, I. (1993) A unique restriction endonuclease, *Bcg*I, from *Bacillus coagulans. Nucleic Acids Res.* **21**, 987-991.

Kossykh, V.G., Repyk, A.V. and Hattman, S. (1993) Sequence motifs common to the *Eco*RII restriction endonuclease and the proposed sequence specificity domain of three DNA-[cytosine-C5] methyltransferase. *Gene* **125**, 65-68.

Koyama, T., Obata, S., Osabe, M., Takeshita, A., Yokoyama, K., Uchida, M., Nishino, T. and Ogura, K. (1993) Thermostable farnesyl diphosphate synthetase of *Bacillus stearothermophilus:* Molecular cloning, sequencing determination, overproduction and purification, *J. Biochem.* **113**, 355-363.

Kretsinger, R.H. and Nelson, D.J. (1976) Calcuim in Biological Systems. *Coord. Chem. Rev.* **18**, 29-124.

Kristjansson, J.K. and Stetter, K.O. (1992) Thermophilic Bacteria Ch. 1 in *Thermophilic Bacteria*, Ed.Kristjansson, J.K. CRC Press, Boca Raton, FL. pp 1-18

Kroger, M. and Hobom, G. (1984) The nucleotide recognised by the *Escherichia coli* A restriction and modification enzyme. *Nucleic Acids Res.* **12**, 887-99.

Krüger, D.H., Schroeder, C., Santibanez-Koref, M. and Reuter, M. (1989) A virus encoded methylase inhibitor and evidence for selection of methylase recognition sites in viral genomes. *Cell. Biophys.* **15**, 87-95.

Kulakauskas, S., Barsomian, J.M., Lubys, A., Roberts, R.J. and Wilson, G.G. (1994) Organisation and sequence of the *Hpa*II restriction modification system and adjacent genes. *Gene* **142**, 9-15.

Kunkel, T.A. and Loeb, L.A. (1979) On the fidelity of DNA replication. Effect of metal ion activation and deoxyrionucleotide pools on *in vitro* mutagenesis. *J. Biol. Chem.* **254**, 5718-25.

Lacks, S. and Greenberg, B. (1977) Complementary specificity of restriction endonucleases of *Diplococcus pneumoniae* with respect to DNA methylation. *J. Mol. Biol.* **114**, 153-168

Laue, F., Evans, L.R., Jarsch, M., Brown, N.L. and Kessler, C. (1991) A complex family of class II restriction endonucleases *Dsa* I-VI in *Dactylococcipsis salia.. Gene* **97**, 87-95.

Lawyer, F.C., Stoffel, S., Saiki, R.K., Myambo, K., Drummond, R. and Gelfand, D.H. (1989) Isolation characterisation and expression in *Escherichia coli* of the DNA polymerase gene from *Thermus aquaticus. J. Biol. Chem.* **264**, 6427-6435

Lee, Y-E., Ramesh, M.V. and Zeikus, J.G. (1993) Cloning, sequencing and biochemical characterisation of xylose isomerase from *Thermoanaerobacterium saccharolyticum* strain B6A-R1. *J. Gen. Microbiol.* **139**, 1227-1234.

Levy, W.P. and Welker, N.E. (1981) Deoxyribonucleic-acid modification methylase from *Bacillus Stearothermophilus. Biochemistry* **20**, 1120-1127.

Liu, Q., Chen, X., Zhao, X., Chen, Y. and Chen, D. (1992) The effect of methylation outside the recognition sequence of restriction endonuclease *Pvu*II on its cleavage efficency. *Gene* **113**, 89-93.

Lobos, C. and Vásquez, C. (1993) Purification and characterisation of *Bst*LVI restriction endonuclease, a thermostable isoschizomer of *Cla*I from *Bacillus steaothermophilus* LV. *Biochim. Biophys. Acta.* **1171**, 295-298.

Lubys, A., Menkevicius, A., Butkus, V. and Janulaitis, A. (1994) Cloning and analysis of translational control for genes encoding the *Cfr*91 restriction modification system. *Gene* **141**, 85-89.

Luke, P.A., McCallum, S.A. and Halford, S.E. (1987) The *Eco*RV restriction endonuclease. *Gene. Amplif. Anal.* **5**, 185-207.

Maekawa, Y., Yasukawa, H. and Kawakami, B. (1990) Cloning and nucleotide sequences of the *Ban*I restriction modification genes in *Bacillus aneurinolyticus. J. Biochem.* **107**, 645-649.

Mann, M.B., Roa, R.N. and Smith, H.O. (1978) Cloning of restriction and modification genes in *E. coli*: the *Hha*II system from *Haemophilus haemolyticus. Gene* **3**, 97-112.

Marinus, M.G. (1987) DNA methylation in *Escherichia coli.,Annu. Rev. Genet.* **21**, 113-31.

McClarin, J.A., Fredrick, C.A., Wang, B.C., Greene, P., Boyer, H.W., Grable, J. and Rosenberg, J.M. (1986) Structure of the DNA- *Eco*RI endonuclease recognition complex at 3Å resolution. *Science* **234**, 1526-1541.

McClelland, M. (1981) Purification and characterization of two new modification methylases; M*Cla*I from *Caryophanon latum* and M*Taq*I from *Thermus aquaticus* YTI. *Nucleic Acids Res.* **9**, 6795-6804.

McClelland, M. (1983) The effect of site specific methylation on restriction endonuclease cleavage (update). *Nucleic Acids Res.* **11**, r169-73.

McClelland, M., Nelson, M. and Raschke, E. (1994) Effect of site specific methylation on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Res.* **22**, 3640-3659.

Meselson, M. and Yaun, R. (1968) DNA restriction enzyme from *E. coli*. *Nature* **217**, 1110-1114.

Modrich, P. (1982)Studies on sequence recognition by type II restriction and modification enzymes. *CRC Crit. Rev. Biochem.* **13**, 287-323.

Modrich, P. and Roberts, R.J. (1982) in *Nucleases* Eds. Linn, S.M. and Roberts, R.J. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York pp 109-154.

Mullis, K.B. and Faloona, F. (1987) Specific synthesis *in vitro* via a polymerase catalyzed chain reaction. *Meth. Enzymol.* **155**, 335-50.

Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1994) The structure of restriction endonuclease *Bam*HI and its relationship to *Eco*RI. *Nature* **368**, 660-664.

Nölling, J., van Eggen, J.M., Eggen, R.I.L. and de Vos, W.M. (1992) Modular organisation of related Archeael plasmids encoding different restriction modification systems in *Methanobacterium thermoformicicum*. *Nucleic Acids Res.* **20**, 6501-6507.

Ostrowski, J. and Bomsztyk, K. (1993) Purification of DNA-binding proteins using tandem DNA-affinity column. *Nucleic Acids Res.* **21**, 1045-1046.

Patel, D.J. (1994) A molecular handshake. *Nature* **367**, 688-690.

Piekarowicz, A., Yuan, R. and Stein, D.S. (1988) Construction of a temperature sensitive mutation for the direct identification of plasmids encoding DNA methylases. *Gene* **74**, 233-235.

Piekarowicz, A., Yuan, R. and Stein, D.S. (1990) A new method for rapid identification of genes encoding restriction and modification enzymes. *Nucleic Acids Res.* **19**, 1831-1835.

Pirrotta, V. and Bickle, T.A. (1980) General purification schemes for restriction endonucleases. *Meth. Enzymol.* **65**, 89-95.

Polisky, B., Greene, P., Garfin, D.E., McCarthy, B.J., Goodman, H.M. and Boyer, H.W. (1975) Specificity of substrate recognition by the *Eco*RI restriction endonuclease. *Proc. Natl. Acad. Sci. USA* **72**, 3310-4.

Polisson, C. (1991) A type II restriction endonuclease *Apo*I, obtainable from *Arthrobacter protphormiae* and a process for producing the same. Europeon Patent Application. 0 539 160 A1.

Pósfai, J., Bhagwat, A.S., Pósfai, G. and Roberts, R.J. (1989) Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.* **17**, 2421-2435.

Pozidis, C., Viatakis, G. and Bouriotis, V. (1993) Sequence specific DNA affinity chromatography: application of a group specific adsorbant for the isolation of restriction endonucleases. *J. Chromatogr.* **630**, 151-7.

Raleigh, E.A. and Wilson, G. (1986) *Escherichia coli* K12 restriction DNA containing 5' methylcytosine. *Proc. Natl. Acad. Sci. USA* **83**, 9070-9074.

Raleigh, E.A., Murray, N.E., Revel, H., Blumenthal, R.M., Westaway, D., Reith, A.D., Rigby, P.W.J., Elhai, J. and Hanahan, D. (1988) *Mcr* A and *Mcr* B restriction phenotypes of some *E. coli* strains and implications for gene cloning. *Nucleic Acids Res.* **16**, 1563-1575.

Roberts, R.J. and Macelis, D. (1992) Restriction enzymes and their isoschizomers. *Nucleic Acids Res.* **20** (suppl.), 2167-2180.

Roberts, R.J. and Macelis, D. (1994) REBASE- restriction enzymes and methylases. *Nucleic Acids Res.* **22**, 3628-3639.

Robinson, C.R. and Slingar, S.G. (1993) Molecular recognition mediated by bound water, a mechanism for star activity of the restriction endonuclease *Eco*RI. *J. Mol. Biol.* **234**, 302-306.

Rüdiger, A., Jorgensen, P.L. and Antranikian, G. (1995) Isolation and characterisation of a heat stable pullulanase from the hyperthermophilic Archaeon *Pyrococcus woesii* after cloning and expression of its gene in *E. coli.* *Appl. Environ. Microbiol.* **61**, 567-575.

Sakamoto, Y., Nakajima, H., Nagata, K., Esaki, N. and Soda, K. (1994) Large scale production of thermostable alanine dehydrogenase from recombinant cells. *J. Ferment. Bioeng.* **78**, 84-87.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning A Laboratory Manual. 2nd Ed. Cold Spring Harbor Laboratory Press.

Schinzel, R. and Burger, K.J. (1986) A site specific endonuclease activity in *Halobacterium Halobium. FEMS Micro. Lett.* **37**, 325-329.

Schroeder, C., Jurkschat, H., Meisel, A., Reich, J-G. and Krüger, D. (1986) Unusual occurrence of *Eco*PI and *Eco*P15 recognition sites and counterselection of type II methylation and restriction sequences in bacteriophage T7 DNA. *Gene* **45**, 77-86.

Selent, U., Ruter, T., Kohler, E., Liedtke, M., Theiking, V., Alves, J., Oelgeschlager, T., Wolfes, H., Peters, F. and Pingoud, A. (1992) A site directed mutagenesis study to identify amino acid residues involved in the catalytic function of the restriction endonuclease *Eco*RV. *Biochemistry* **31**, 4808-15.

Shannon, R.D. (1976) Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr. Sect. A.* **32** 751-767.

Sharp, R.J. and Williams, R.A.D. (1988) Properties of *Thermus ruber* strains isolated from Icelandic hot springs and DNA:DNA homolgy of *Thermus ruber* and *Thermus aquaticus*. *Appl. Environ. Microbiol.* **54**, 2049-2052.

Shaw, P.C., Clark, D.R., Kam, K.M. and Mok, Y.-K. (1991) Isolation and characterisation of thermophilic endonucleses. *ASM Abstracts* **91**, 199.

Shibata, T., Watabe, H., Kaneko, T., Iino, T. and Ando, T. (1984) On the nucleotide sequence recognised by a eukaryotic site specific endonuclease, Endo *Sce*I from yeast. *J. Biol. Chem.* **259**, 10499-10506.

Siksnys, V., Zarecteckaja, N., Vaisvila, R., Timinskas, A., Stakenas, P., Butkus, V. and Janulaitis, A. (1994) CAATTG specific restriction modification *mun*I genes from *Mycoplasma*: sequence similarities between R·*Mun*I and R·*Eco*RI. *Gene* **142**, 1-8.

Silhavy, T.J., Berman, M.L. and Enquist, L.W. (1984) Experiments in Gene Fusions, Cold Spring Harbor Laboratory : New York pp xi-xii.

Sklar, R., Altman, D. and Sager, R. (1986) An endonuclease from *Chlamydomonas reinhardtii* that cleaves the sequence TATA. *J. Biol. Chem.* **261**, 6806-6810.

Slatko, B.E., Benner, J.S., Jager-Quinton, T., Moran, L.S., Simcox, T.G., VanCott, E.M. and Wilson, G.G. (1987) Cloning sequencing and expression of the *Taq* I restriction modification system . *Nucleic Acids Res.* **15**, 9781-9796.

Smith, H.O. and Nathans, D. (1973) A suggested nomenclature for bacterial host modification and restriction systems and their systems. *J. Mol. Biol.* **81**, 419-424.

Sota, H., Arunwanich, P., Kurita, O., Uozumi, N., Honda, H., Iijima, S. and Kobayashi, T. (1994) Molecular cloning of thermostable β-glucosidase gene from a thermophilic anaerobe NA10 and its high expression in *E. coli*. *J. Ferment. Bioeng.* **77**, 199-201.

Steitz, T.A. (1990) Structural studies of protein nucleic acid interaction: the sources of seqence-specific binding. *Q. Rev. Biophys.* **23**, 205-280.

Stephenson, F.H., Ballard, B.T., Boyer, H.W., Rosenberg, J.M. and Greene, P.J. (1989) Comparison of the nucleotide and amino acid sequences of the *Rsr*I and *Eco*RI restriction endonucleases. *Gene* **85**, 1-13.

Szybaski, W., Blumenthal, R.M., Brooks, J.E., Hattman, S. and Raleigh, E.A. (1988) Nomenclature for bacterial genes encoding for class II restriction endonucleases and modification methyltransferases. *Gene* **74**, 279-280.

Tabor, S. and Richardson, C.C. (1989) Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage-T7 DNA-polymerase and *Escherichia Coli* DNA polymerase-I. *Proc. Natl. Acad. Sci. USA* **86**, 4076-4080.

Tanaka, T., Kawano, N. and Oshima, T. (1981) Cloning of 3-isopropylmalate dehydrogenase gene of an extreme thermophile and partial purification of the gene product. *Biochemistry* **89**, 677-682.

Tao, T. and Blumenthal, R.M. (1992) Sequence and characterisation of *pvuIIR*, the *Pvu*II endonuclease gene and of *pvuIIC*, its regulatory gene. *J. Bacteriol.* **174**, 3395-3398.

Tao, T., Bourne, J.C. and Blumenthal, R.M. (1991) A family of regulatory genes associated with type II restriction modification systems. *J. Bacteriol.* **173**, 1367-1375.

Tao, T., Walter, J., Brennan, K.J., Cotterman, M.M. and Blumenthal, R.M. (1989) Sequence internal homology and high level expression of the gene for a DNA (cytosine N4) methyltransferase M·*Pvu* II. *Nucleic Acids Res.* **17**, 4161-4175.

Taylor, J.D. and Halford, S.E. (1992) The activity of the *Eco*RV restriction endonuclease is influenced by flanking DNA sequences both inside and outside the DNA protein complex. *Biochemistry* **31**, 90-97.

Taylor, J.D., Badcoe, I.G., Clarke, A.R. and Halford, S.E. (1991) *Eco*RV restriction endonuclease binds all DNA sequences with equal affinity. *Biochemistry* **30**, 8743-8753.

Taylor, W.R. (1986) The classification of amino acid conservation. *J. Theor. Biol.* **119**, 205-218.

Terry, B., Jack, W.E., Rubin, R.A. and Modrich, P. (1983) Thermodynamic parameters governing interaction of *Eco*RI endonuclease with specific and non specific DNA sequences. *J. Biol. Chem.* **258**, 9820-9825.

Thielking, V., Selent, U., Köhler, E., Wolfes, H., Pieper, U., Geiger, R., Urbanke, C., Winkler, F.K. and Pingoud, A. (1991) Site directed mutagenesis studies with *Eco*RV restriction endonuclease to identify regions involved in recognition and catalysis. *Biochemistry* **30**, 6416-6422.

Tikchinenko, T.I., Karamov, E.V., Zavizion, B.A. and Naroditsky, B.S. (1978) *Eco*RI activity: enzyme modification or activation of accompanying endonuclease? *Gene* **4**, 195 -212.

Topal, M.D. and Conrad, M. (1993) Changing endonuclease *Eco*RII Tyr308 to Phe abolishes cleavage but on recognition: possible homology with the Int-family of recombinases. *Nucleic Acids Res.* **21**, 2599-2603.

Ueno, T., Ito, H., Kimizuka, F., Kotani, H. and Nakajima, K. (1993) Gene structure and expression of the *Mbo*I restriction modification system. *Nucleic Acids Res.* **21**, 2309-2313.

VanCott, E.M. (1990) Method for producing the *Sfi* I restriction endonuclease and methylase. European Patent Application 9031 4311.3.

VanCott, E.M. and Wilson, G.G. (1988) Cloning the *Fnu*DI, *Nae*I, *Nco*I and *Xba*I restriction modification systems. *Gene* **74**, 55-59.

Vermote, C.L.M. and Halford, S.E. (1992) *Eco*RV restriction endonuclease: Communication between catalytic metal ions and DNA recognition. *Biochemistry* **31**, 6082-6089.

Viatakis, G. and Bouriotis, V. (1991a) Sequence specific DNA affinity chromatography: application to the purification of *Eco*RI and *Sph*I. *Anal. Biochem.* **195**, 352-7.

Viatakis, G. and Bouriotis, V. (1991b) Affinity partitioning of restriction endonucleses. Application to the purification of *Eco*RI and *Eco*RV. *J. Chromatogr.* **538**, 311-21.

Vipond, B. and Halford, S.E. (1995) Specific DNA recognition by *Eco*RV restriction endonuclease induced by calcium ions. *Biochemistry* **34**, 1113-1119.

Walder, R.Y., Hartley, J.L., Donelson, J.E. and Walder, J.A. (1981) Cloning and expression of the *Pst*I restriction modification system in *Escherichia coli Proc. Natl. Acad. USA* **78**, 1503-1507.

Walder, R.Y., Walder, J.A. and Donelson, J.E. (1984) The organisation and complete nucleotide sequence of the *Pst*I restriction modification system. *J. Biol. Chem.* **259**, 8015-8026.

Wang, R.Y., Shenoy, S. and Ehrlich, M. (1984) DNA methylation inhibits the transfecting activity of replicative- form *phi* X174 DNA. *J. Virol.* **49** 674-679.

Ward, J.M., Wallace, L.J., Cowan, D. and Shadbolt, P. (1991) Phosphocellulose as a tool for rapid purification of DNA-modifying enzymes. *Analytica Chimica Acta.* **249**, 195-200.

Watabe, H., Shibata, T. and Ando, T. (1981) Site specific endo-deoxyribonucleases in eukaryotes: endonucleases of yeast, *Saccharomyces* and *Pichia. J. Biochem.* **90**, 1623-1632.

Williams, R.A.D. (1992) The Genus *Thermus* (Ch 3) in *Thermophilic Bacteria*, Ed.Kristjansson, J.K. CRC Press, Boca Raton, FL. pp 51-62

Wilson, G.G. (1988) Cloned restriction modification systems - a review *Gene* **74**, 281-289.

Wilson, G.G. (1991) Organisation of restriction modification systems. *Nucleic Acids Res.* **19**, 2539-2566.

Wilson, G.G. and Murray, N.E. (1991) Restriction and modification systems. *Annu. Rev. Genet.* **25**, 585-627.

Winkler, F.K., Banner, D.W., Oefner, D., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K. and Wilson, K.S. (1993) The crystal structure of *Eco*RV endonuclease and of it's complexes with cognate and non-cognate DNA fragments. *EMBO J.* **12**, 1781-1795.

Withers, B.E. and Dunbar, J.C. (1993) The endonuclease isoschizomers, *Sma*I and *Xma*I, bend DNA in opposite orientations. *Nucleic Acids Res.* **21**, 2571-2577.

Woodcock, D.M., Crowther, P.J., Doherty, J., Jefferson, S., DeCruz, E., Noyer-Weidner, M., Smith, S.S., Micheal, M.Z. and Graham, M.W. (1989) Quantitative evaluation of *Escherichia coli* strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic Acids Res.* **17**, 3469-3478.

Xia, Y., Burbank, D.E., Uher, L., Rabussay, D. and Van Etten, J.L. (1986) A virus infection of a *Chlorella*-like green alga induces a DNA restriction endonuclease with a specific sequence specificity. *Nucleic Acids Res.* **15**, 6075-6090.

Xu, G.L., Kapler, W., Walter, J. and Trautner, T.A. (1992) *Bsu*BI- an isospecific restriction and modification system of *Pst*I: characterisation of the *Bsu*BI genes and enzymes. *Nucleic Acids Res.* **20**, 6517-6523.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: Nucleotide sequences of M13mp18 and pUC19 vectors. Gene **33**, 103.

Zebela, J.A., Choi, J., Trainor, G.L. and Barany, F. (1992) DNA recognition of base analogue and chemical modified substrates by the *Taq*I restriction endonuclease. *J. Biol. Chem.* **267**, 8106-8116.

Zeikus, J.G. (1979) Thermophilic bacteria, ecology, physiology and technology. *Enz. Micro. Tech.* **1**, 243-251.

Zhang, B., Tao, T., Wilson, G.G. and Blumenthal, R.M. (1993) The M·*Alu*I DNA-(cytosine C5)-methyltransferase has an unusually large, partially dispensible, variable region. *Nucleic Acids Res.* **21**, 905-911.

Zhang, Y., Nelson, M., Nietfeldt, J.W., Burbank, D.E. and Van Etten, J.L. (1992) Characterization of *Chlorella* virus PBCV-1 *Cvi*AII restriction and modification system. *Nucleic Acids Res* **20**, 5351-5356.

APPENDIX I

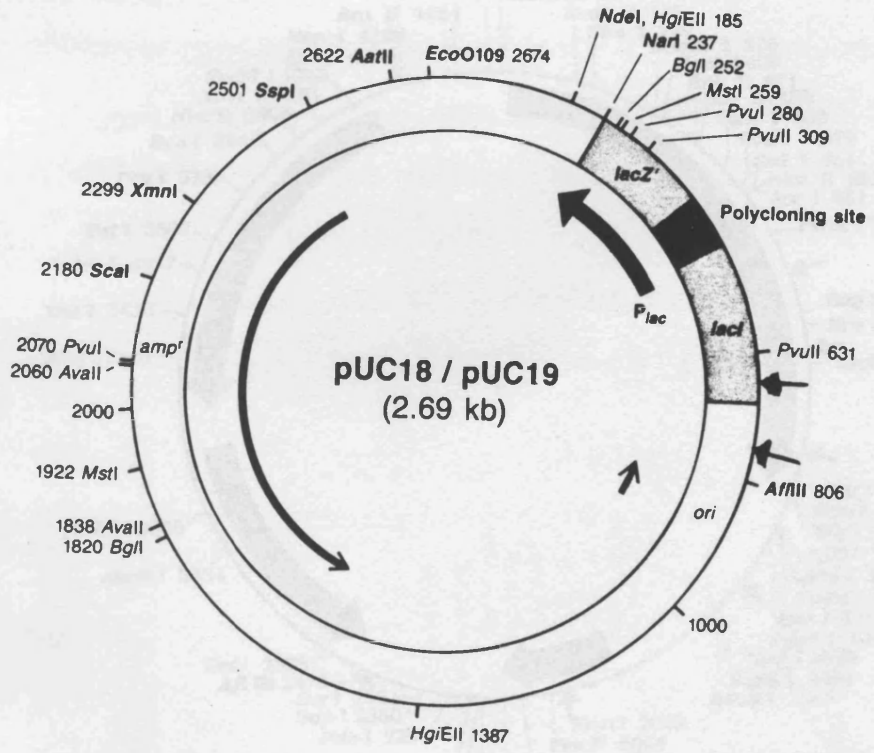
DIGESTION OF pBR322, pUC18 and λ with *Tfi*I ENDONUCLEASE

[Lane 1, λ-*Pst*I; Lane 2, pBR322-*Tfi*I; Lane 3, pUC18-*Tfi*I; Lane 4, λ-*Tfi*I; Lane 5, λ-*Hin*dIII.]

APPENDIX II

MAP OF pUC19/pUC18

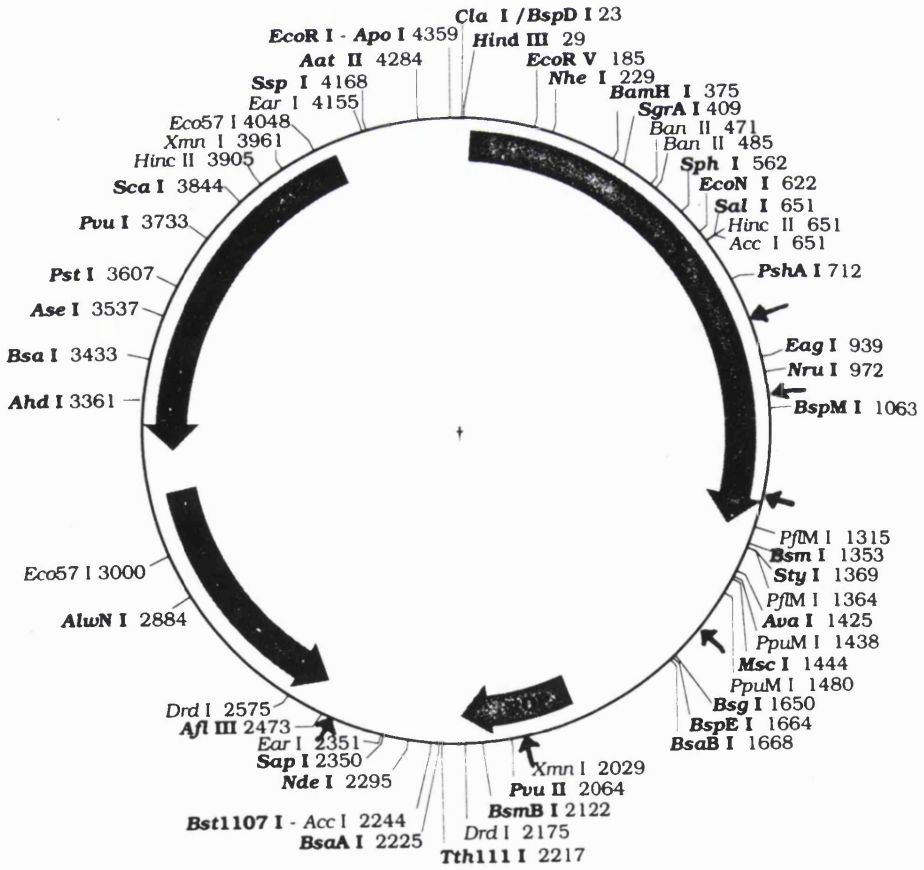
These vectors are high copy number *E. coli* plasmid cloning vectors. They are identical except they contain polycloning sites arranged in opposite orientations. pUC vectors express the amino terminal fragment of the *lacZ* gene product (β-galactosidase) and display α-complementation in appropriate hosts. *Tfi*I sites are marked with a ↑ (positions 641 and 781). The map also shows the relative position of the antibiotics resistance gene and the origin of replication. Reproduced from Sambrooks *et al.*, 1989.



APPENDIX III

MAP OF pBR322

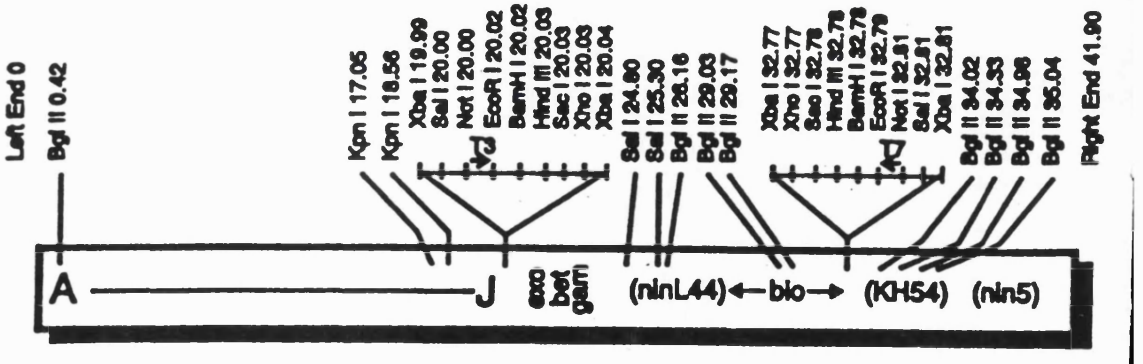
pBR322 is an *E. coli* plasmid vector. The map shows the restriction sites of those enzymes that cut the molecule once or twice; the unique sites are shown in bold type. *Tfi*I sites are marked with a ↑ (positions 852, 1006, 1304, 1525, 2029 and 2448). The map also shows the relative position of the antibiotics resistance genes, rop (mediates the activity of RNase I) and the origin of replication. Reproduced from New England Biolabs 1995 Catalog.



APPENDIX IV

MAP OF LAMBDA DASH™ II REPLACEMENT VECTOR

Lambda DASH™ II is a replacement vector used for the cloning large fragments of genomic DNA. The Lambda DASH™ II system takes advantage of spi (sensitive to P2 inhibition) selection. Lambda phages containing active *red* and *gam* genes are unable to grow on host strains that contain P2 phage lysogens. Lambda phage without these genes are able to grow on strains lysogenic for P2 such as XL1-Blue MRA (P2), a P2 lysogen of XL1-Blue MRA [this study]. The red and gam genes in the Lambda DASH™ II DNA are located on the stuffer fragment; therefore, the wild type Lambda DASH™ II phage can not grow on XL1-Blue MRA (P2). When the stuffer fragment is replaced by an insert, the recombinant Lambda DASH™ II becomes *red*$^-$/*gam*$^-$, and the phage is able to grow on the P2 lysogenic strain. Target DNA cloned into the *Bam*HI sites of the Lambda DASH™ II vector may be removed by digestion with *Not*I. The unique arrangement of the Lambda DASH™ II polylinker permits the isolation of the insert and flanking T3 and T7 bacteriophage promoters as an intact cassette by digestion with *Not*I. Reproduced from Stratagene Lambda DASH™II/*Bam*HI Vector Kit.



APPENDIX V

Strains of the genus *Thermus* have provided a number of thermostable and biotechnologically significant enzymes including restriction modification enzymes, some of which are commercially available. However the importance of these enzymes lies more in their sequence specificity than in their thermal stability. For example,*Tsp*EI isolated by Raven *et al.* (1993) generates cohesive termini compatible with those of *Eco*RI and therefore has an application in producing partial digests of DNA for subsequent ligation into *Eco*RI digested cloning vectors.

Over twenty restriction endonucleases, each with a different DNA recognition site, have been discovered within the species and strains of the genus *Thermus* and several of these enzymes have no known mesophilic isoschizomers (Table AI, see Roberts and Macelis, 1994; Duffield and Cossar, 1995, for primary sources). Welch and Williams (1995) have isolated two different endonucleases from sixteen isolates of *Thermus* from neutral and alkaline hot water springs in the south-west region of Iceland. *Tsp*4CI (ACN/GT) has a novel specificity and *Tsp*8EI is an isoschizomer of the mesophilic enzyme *Bgl*I (GCCNNNN/NGGC).

Several restriction modification systems from *Thermus* have been characterised (see Duffield and Cossar, 1995, for review and primary sources). The most extensively studied of these systems is the *Taq*I restriction modification system isolated from *Thermus aquaticus* YT1 which has been cloned and characterised (Slatko *et al.*, 1987). The cloning of the *Taq*I restriction modification system was accomplished by selecting *in vitro* for self modified plasmids.

Table AI Restriction Endonucleases from *Thermus* species

Enzyme	Sequence
*Taq*I	TCGA
*Taq*II*	GACCGAA
	CACCCA
*Taq*XI (*Aeu*II)	CCW[a]GG
*Tfi*I	GATWC
*Tru*2011 (*Xho*II)	RGATCY[b]
*Tru*I	GGWCC
*Tru*II (*Mbo*I)	GATC
*Tse*AI (*Sdu*I)	GD[c]GCH[d]C
*Tse*I (*Tsp*EI)	GCW[a]GC
*Tsp*45I	GTS[e]AC
*Tsp*4CI	ACN/GT
*Tsp*5041 (*Xma*III)	CGGCG
*Tsp*5071 (*Bsp*MII)	TCCGGA
*Tsp*5601 (*Hae*III)	GGCC
*Tsp*8EI	GCCNNNN/NGGC
*Tsp*RI	CASTG
*Tth*111I	GACN[f]NNGTC
*Tth*111II	CAARCA
*Ttm*I (*Mae*II /*Tsp*49I)	ACGT
*Ttm*II (*Bse*PI)	GCGCGC

[a] W= A or T

[b] D= not C (A or G or T)

[c] Y= C or T

[d] H= not G (A or C or T)

[e] S= G or C

[f] N= A or C or G or T

**Taq*II differs from other restriction enzymes in recognising two distinct sequences

In vivo selection using phage for restricting clones was found to be less reliable as *Taq*I is an example of a cloned restriction modification system that modifies well but restricts poorly. The nucleotide sequence of the genes encoding methyltransferase *Taq*I (M·*Taq*I) and restriction endonuclease *Taq*I (R·*Taq*I) with the recognition sequence, TCGA, were analysed in clones isolated from independent libraries. Barany *et al.* 1992(b) redetermined the nucleotide sequence of *Taq*I restriction modification genes as 421 and 263 codons long, respectively. The C terminus of the *taq*IM gene overlaps the N terminus of the *taq*IR gene by 13 codons. Removal of the overlapping codons did not interfere with *invivo* M·*Taq*I activity.

The roles of metal cofactors Mg^{2+} and Mn^{2+} on the specificity of *Taq*I endonuclease was investigated using steady state and single turnover kinetics (Cao *et al.*, 1995). In the presence of Mg^{2+}, stringent discrimination of *Taq*I against single base-pair changes (star sites) was manifested by the loss of tight, specific binding in the early stage of the enzymatic cycle. In the presence of Mn^{2+}, relaxed specificity for a star site sequence is attributed to formation of three distinct classes of the ternary complexes: the highly activated *Taq*I-cognate-Mn^{2+} complex; the partially activated *Taq*I-star-Mn^{2+} complex; and the ground state, inactive *Taq*I-non-specific-Mn^{2+} complex. In addition to a high affinity for a *Taq*I-DNA complex, Mn^{2+} also binds to *Taq*I in a DNA-independent fashion. This may facilitate enzyme activation, which could account for the observed relaxation in substrate specificity. Thus, the authors speculated that the *Taq*I-DNA-Mn^{2+} complex could be formed by either of two pathways: *Taq*I binding to DNA followed by the binding of Mn^{2+} or *Taq*I first binding to Mn^{2+} followed by the addition of DNA. The inactive, non-specific *Taq*I-star-Mg^{2+} complex virtually

prohibits transition state interactions, but a *Taq*I-star-Mn^{2+} complex attains a measurable single-turnover rate. In the late stages of the enzymatic cycle, high affinity of Mn^{2+} to a *Taq*I-DNA complex and to the *Taq*I enzyme may also account for a slower rate of product release.

The *Thermus aquaticus* DNA methyltransferase M·*Taq*I (EC 2.1.1.72) methylates N^6 of adenine in the specific double-helical DNA sequence TCGA by transfer of $-CH_3$ from the cofactor S-adenosyl-L-methionine. The X-ray crystal structure at 2.4 Å resolution of this enzyme in complex with S-adenosylmethionine has been determined (Labahn *et al.*, 1994). The structures shows alpha/beta folding of the polypeptide into two domains of about equal size. They are arranged in the form of a C with a wide cleft suitable to accommodate the DNA substrate. The N-terminal domain is dominated by a nine-stranded beta-sheet; it contains the two conserved segments typical for N-methyltransferases which form a pocket for cofactor binding. The C-terminal domain is formed by four small beta-sheets and alpha-helices. The three-dimensional folding of M·*Taq*I is similar to that of the cytosine-specific *Hha*I methyltransferase, where the large beta-sheet in the N-terminal domain contains all conserved segments and the enzymatically functional parts, and the smaller C-terminal domain is less structured.

The *Tth*HB8I restriction and modification system from *Thermus thermophilus* HB8 is an isoschizomer of the *Taq*I restriction modification system (i.e., the palindromic recognition sequence T↓CGA). The genes were cloned in *E.coli.* (Barany *et al.*, 1992a). Selection was based on the ability of methylated plasmids to remain intact upon digestion with *Taq*I. The genes have the same transcriptional orientation, with the last 13 codons of the methyltransferase (MTase) overlapping the first 13 codons of the endonuclease (ENase). Nucleotide sequence analysis of the *Tth*HB8I ENase

revealed a single chain of 263 amino acid (aa) residues that share a 77% identity with the corrected isoschizomeric *Taq*I ENase. Likewise, the *Tth* MTase (428 aa) shares a 79% identity with the corrected sequence of the *Taq*I MTase. This high degree of amino acid conservation suggests a common origin between the *Taq* and *Tth* RM systems. However, codon usage and G+C content for the R-M genes differed markedly from that of other cloned *Thermus* genes. The authors suggests that these R-M genes were only recently introduced into the genus *Thermus* as a bacterial virus or provirus- like plasmid analogous to *Eco*PI, *Eco*P15 or the *Chlorella* type viruses.

Additional References

Raven, N.D., Kelly, C.D., Carter, N.D., Eastlake, P., Brown, C and Williams, R.A. (1993) A new restriction endonuclease, *Tsp*EI, from the genus *Thermus* that generates cohesive termini compatible with those of *Eco*RI. Gene **131**(1) 83-6

Cao, W., Mayer, A.N. and Barany, F. (1995) Stringent and relaxed specificities of *Taq*I endonuclease: interactions with metal cofactors and DNA sequences. Biochemistry **34** (7) 2276-83

Duffield, M.L. and Cossar, D. (1995) Enzymes of *Thermus* and their properties. Ch 4 in *Thermus* Species, Eds. Sharp, R. and Williams, R. Published by Plenum Press pp 93-141.

Labahn, J., Granzin, J., Schluckebier, G., Robinson, D.P., Jack, W.E., Schildkraut, I and Saenger, W. (1994) Three-dimensional structure of the adenine-specific DNA methyltransferase M·*Taq* I in complex with the cofactor S-adenosylmethionine. *Proc.Natl.Acad.Sci.U.S.A.* **91**(23): 10957-61

Welch, S.G. and Williams, R.A. (1995) Two thermostable type II restriction endonucleases from Icelandic strains of the genus *Thermus*: *Tsp*4C I (ACN/GT), a novel type II restriction endonuclease, and *Tsp*8E I, an isoschizomer of the mesophilic enzyme *Bgl* I (GCCNNNN/NGGC). *Biochem.J.* **309** (Pt 2): 595-9