Restriction systems in *M. methylotrophus*.

by

Alan Christopher Boyd.

A thesis submitted for the degree of Doctor of Philosophy of the University of Leicester.

Department of Biochemistry

University of Leicester

September 1983

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

Column chromatography of *M. methylotrophus* extracts revealed the presence of two restriction endonucleases, *Mme*I and *Mme*II. *Mme*I was active against DNA from normal *E. coli* strains, but *Mme*II could only cleave DNA from *dam*$^-$ strains.

A general computer method was developed for determining restriction enzyme recognition sequences, and was successfully applied to the *M. methylotrophus* enzymes. *Mme*I recognizes the novel sequence 5'-TCCPuAC-3', and *Mme*II the sequence 5'-GATC-3', the same as that recognized by the *dam* methylase of *E. coli*, and several other restriction enzymes, including *Sau*3A. The recognition sequence of *Mme*I was confirmed by comparing the cleavage patterns it produced with sequenced DNAs (ΦX174, SV40, M13) with those predicted and simulated by computer. The partially-purified preparation of *Mme*I used always gave incomplete digestion.

Mutants lacking *Mme*II activity (*mmeB*$^-$) were isolated, and shown to be about one thousand-fold better recipients of unmodified RP4 than wild-type *M. methylotrophus*. A test for *dam* methylation in *E. coli* using this mutant was devised. An unusual Tn*5* mutant, CBM22, was isolated while screening for *mmeB*::Tn*5* mutants. This strain was shown to lack an endonuclease active against *dam*$^+$ DNA which *M. methylotrophus* possesses, probably *Mme*I, but conclusive proof that the missing activity was *Mme*I was not obtained. The mutant allele in CBM22 was designated *endA*::Tn*5*, after the strain was shown to be *mmeB*$^+$. CBM22 appears to be a mutator strain: enhanced mutation rates to trimethoprim, tetracycline and nalidixic acid resistance were observed. The mutator phenotype was exploited to aid in the construction of *endA mmeB* double mutants.

A quantitative comparison of the *in vivo* restriction profiles of endonuclease mutants was carried out using filter mating. Under this regime, *endA* restriction *in vivo* was not detectable.

This thesis is dedicated to

everyone who knows me.

STATEMENT

The accompanying thesis submitted for the degree of Ph. D., entitled "Restriction systems in *M. methylotrophus.*", is based on work conducted by the author in the Department of Biochemistry of the University of Leicester mainly during the period between October 1979 and October 1982.

All the work recorded in this thesis is original unless otherwise acknowledged in the text or by references. None of the work has been submitted for another degree in this or any other University.

Signed: ..... Date: 2/9/83...

ACKNOWLEDGEMENTS

I would like to acknowledge all those people who have helped, in one way or another, to ease me through my three-year PhD project - especially those I have forgotten to mention by name below.

My gratitude, first of all, goes to my supervisor Bill Brammar for his invaluable help and stoic patience.

In my early days in Lab. 124, I was helped considerably by Barry Ely and Nigel Grinter - in return for pool and pinball lessons. My appreciation also goes to Annie, for laughing at my jokes; and to Judy, for being an inexhaustible source of lost scissors, scalpels etc. Thanks also go to Jean Dean, for reminding me when my hair got too long. Jack Stanford made sure that glassware was kept meticulously clean and sterile, and was sorely missed when he left, half way through my project. Thanks to Wil, Dave, Chris, Pam, Mahdu, Alister, Liu, Mark and everyone else who worked in or passed through the lab. I am grateful to members (and ex-members) of the ICI Joint Lab., especially Dave Byrom, Geoff Sharpe, Dave Pioli, Pete Meacock and John Windass for their advice and, occasionally, free beer.

Finally, my thanks go to the Computer Lab personnel, especially Dave Roberts; and to Derek Peacock in the C.S.U. for the use of word-processing facilities. Photographic enlargements were made by Ian Riddell.

I acknowledge the financial support of the Science and Engineering Research Council and Imperial Chemical Industries.

CONTENTS

ABBREVIATIONS. 1

1. MEDIA AND REAGENTS.

1.1. METHYLOTROPH MEDIA.

1.1.1. AS1 salts (4x). 2
1.1.2. AS1 trace elements (x1000). 2
1.1.3. AS1 minimal medium. 2
1.1.4. AS1 minimal agar. 2
1.1.5. -70°C storage medium. 3

1.2. *E. COLI* MEDIA.

1.2.1. L broth (LB). 3
1.2.2. 2x yeast tryptone (2x YT). 3
1.2.3. M9 salts (5x). 3
1.2.4. M9 minimal medium (MM). 4
1.2.5. L agar (LA). 4
1.2.6. BBL agar. 4
1.2.7. P1 agar. 4
1.2.8. M9 minimal agar (MA). 4
1.2.9. Water top agar (WA). 4
1.2.10. BBL top agar. 4
1.2.11. Stab medium. 5
1.2.12. P1 adsorption medium. 5
1.2.13. Xg plates. 5

1.3. BUFFERS.

1.3.1. Phage buffer (Φ B). 5
1.3.2. P1 storage buffer. 6
1.3.3. TE buffer. 6
1.3.4. Tris-acetate buffer (TA). 6
1.3.5. Agarose gel loading buffer. 6
1.3.6. Tris-borate EDTA buffer (TBE). 6
1.3.7. Polyacrylamide gel loading buffer. 6
1.3.8. T4 DNA ligase buffer (10x). 7
1.3.9. Extract buffer (EB). 7
1.3.10. Storage buffer (SB). 7
1.3.11. 666 buffer. 7
1.3.12. Pronase buffer. 7

1.4. MISCELLANEOUS SOLUTIONS.

1.4.1. Lysis solution. 8
1.4.2. Alkaline SDS solution. 8
1.4.3. "5M" acetate solution. 8
1.4.4. CsCl solutions for block gradients. 8
1.4.5. 30% acrylamide solution. 8

2. ANTIBIOTICS. 9

3. METHODS.

3.1. *IN VIVO* METHODS.

3.1.1. Plate matings. 10
3.1.2. Filter matings. 10
3.1.3. Replica plating/mating. 11
3.1.4. Transformation of *E. coli*. 12
3.1.5. P1 lysates. 13
3.1.6. P1 transduction. 14
3.1.7. Detection of plasmid restriction in *M. methylotrophus*. 14
3.1.8. Isolation of a *mmeB* mutant induced by UV light. 16
3.1.9. An *in vivo* test for *dam* methylation. 16

3.2. *IN VITRO* METHODS - DNA.

3.2.1. Ethanol precipitation of DNA. 17
3.2.2. Rapid purification of DNA on G-50 Sephadex columns. 18
3.2.3. Preparation of plasmid DNA. 18
3.2.4. Preparation of λ DNA. 19
3.2.5. Restriction. 20
3.2.6. Ligation. 21
3.2.7. Agarose gel electrophoresis. 22
3.2.8. Polyacrylamide gel electrophoresis. 24
3.2.9. Electroelution of DNA fragments. 25

3.3. *IN VITRO* METHODS - ENZYMES.

3.3.1. Small-scale preparation of *Mme*I and *Mme*II. 25
3.3.2. Purification of *Mme*I. 27
3.3.3. Purification of *Mme*II. 27

BACTERIAL STRAINS. 30

INTRODUCTION.

1. Restriction and modification of coliphage λ *in vivo*. 33
2. Arber's theory. 33
3. Type I enzymes. 34
4. Type III enzymes. 35
5. Type II enzymes. 37
6. *Eco*RI. 38
7. Methylation of GATC sequences. 41
8. Roles of restriction endonucleases in vivo. 43-1
9. Methylophilus methylotrophus: an industrial microorganism. 43-3

Results and Discussion: CHAPTER 1.

1.1. Computer method for determining recognition sequences. 44
1.2. Relationship between molecular weight and gel mobility. 50

Results and Discussion: CHAPTER 2.

2.1. Detection of *Mme*I activity *in vitro*. 53
2.2. Mapping of *Mme*I sites in pBR322. 54
2.3. Determination of *Mme*I recognition sequence. 57
2.4. Generation of partials - theory. 60
2.5. Confirmation of *Mme*I recognition sequence. 62

2.6. *Mme*I sites in M13 and M13mp7. 64

2.7. Comparison of *Mme*I with analogous enzymes. 66

2.8. Cleavage point of *Mme*I. 67

Results and Discussion: CHAPTER 3.

3.1. Detection of *Mme*II activity *in vitro*. 68

3.2. Identification of *Mme*II recognition sequence. 69

3.3. Detection of *Mme*II restriction *in vivo*. 70

3.4. Isolation of UV-induced *Mme*II mutant (*mmeB*). 71

3.5. UV mutant lacks *Mme*II *in vitro*. 72

3.6. Cleavage point of *Mme*II. 73

Results and Discussion: CHAPTER 4.

4.1. Use of *M. methylotrophus* to monitor *dam* genotype. 76

4.2. P1 transduction of *dam*. 76

4.3. Tn*5* mutagenesis of *M. methylotrophus*. 77

4.4. CBM22 - a mutator strain?. 81

4.5. Characterization of mutants *in vitro*. 82

Results and Discussion: CHAPTER 5.

5.1. Simple way of isolating *mmeB* mutants. 84

5.2. Construction of double mutants. 85

5.3. Quantitative comparison of mutant restriction phenotypes. 89

5.4. Modification in endonuclease mutants. 91

CONCLUSION.

1. Prologue. 93

2. Purification and other biochemical aspects. 95

3. The *endA* gene product - a DNA repair enzyme? 96

4. Potential applications of restrictionless mutants. 97-1

REFERENCES. 98

ADDITIONAL REFERENCES. 109

TABLES

Table 1-1 47, 48

Table 1-2 47, 48

Table 2-1 58, 59

Table 5-1 89, 90

FIGURES

Figure 1-1 51, 52

Figure 2-1 55, 56

Figure 2-2 57, 58

Figure 2-3 62, 63

Figure 2-4 63, 64

Figure 2-5 66, 67

Figure 3-1 68, 69

Figure 3-2 72, 73

Figure 4-1 82, 83

APPENDICES.

ABBREVIATIONS.

Nucleotides. (Used when discussing computer output.)

X or Pu	Purine
Y or Py	Pyrimidine
N	A, C, G or T
1	A or C
9	G or T
2	G or C
8	A or T

Antibiotics.

Ap	ampicillin
Km	kanamycin
Nal	nalidixic acid
Rif	rifampicin
Sm	streptomycin
Tc	tetracycline
Tp	trimethoprim

Miscellaneous.

R/M	restriction / modification
w.t.	wild type
RT	room temperature
O/N	overnight
SDS	sodium dodecyl sulphate
bp	base pair(s)
kb	kilobase(s)
EtBr	ethidium bromide
$AmSO_4$	ammonium sulphate
$SmSO_4$	streptomycin sulphate
AdoMet	S-adenosyl methionine

1. MEDIA AND REAGENTS.

Media and reagents are sterilized by autoclaving for 15 min. at 15 p.s.i.

1.1. METHYLOTROPH MEDIA.

1.1.1. AS1 salts (4x).

To make 1l: dissolve $NaH_2PO_4.2H_2O$, 5.6g and K_2HPO_4, 7.6g in 0.8l sterile water, then add $(NH_4)_2SO_4$, 7.2g and stir well to dissolve completely. Add slowly, with stirring, a 100ml solution of $MgSO_4$ (0.8g). Make up to 1l with sterile water. Adjust pH to 6.8, if necessary.

1.1.2. AS1 trace elements (x1000).

To make 1l: $CuSO_4.5H_2O$, 0.02g; $MnSO_4.4H_2O$, 0.01g; $ZnSO_4.7H_2O$; $CaCO_3$, 1.8g; 1M HCl, 36.6ml; make up to 1l with water.

1.1.3. AS1 minimal medium.

To make 1l: 750ml water; 250ml 4x AS1 salts; 9ml CH_3OH; 1ml 1000x trace elements.

1.1.4. AS1 minimal agar.

To make 400ml: 300ml molten Difco Bacto agar (7g/300ml); 100ml 4x AS1

salts 3.6ml CH_3OH; 0.4ml 1000x trace elements.

1.1.5. -70°C storage medium.

N.B. *M. methylotrophus* strains are viable for no more than 14 days on plates at 4°C: it is therefore convenient to maintain stock cultures at -70°C. This is achieved by growing a 0.8ml overnight culture of the strain in a 1.8ml screw-capped polypropylene vial (Nunc). 0.8ml of 20% glycerol in water is added, and the vial frozen and stored at -70°C. Strains kept in this way are viable for at least 3 years.

1.2. *E. COLI* MEDIA.

1.2.1. L broth (LB).

To make 1l: bactotryptone, 10g; yeast extract, 5g; NaCl, 5g; glucose, 1g. Adjust pH to 7.0 with 5N NaOH. For λ growth, add $MgSO_4$ to a final concentration of 10mM.

1.2.2. 2x yeast tryptone (2x YT).

To make 1l: bactotryptone, 16g; yeast extract, 10g; NaCl, 5g.

1.2.3. M9 salts (5x).

To make 1l: Na_2HPO_4, 30g; KH_2PO_4, 15g; NaCl, 2.5g; NH_4Cl, 5g; after autoclaving, add 5ml 1M $MgSO_4$ and 5ml 0.1M $CaCl_2$.

1.2.4. M9 minimal medium (MM).

To make 1l: 800ml water; 200ml 5x M9 salts; 10ml 20% glucose.

1.2.5. L agar (LA).

As L broth but without glucose and solidified by adding 10g/l of Oxoid No. 1 agar prior to autoclaving.

1.2.6. BBL agar.

To make 1l: trypticase (Baltimore Biological Labs.) 10g; NaCl, 5g; Oxoid No. 1 agar, 10g. For λ growth, add $MgSO_4$ to a final concentration of 10mM.

1.2.7. P1 agar.

As BBL agar, except that 10ml of 1M $CaCl_2$ are added after autoclaving.

1.2.8. M9 minimal agar (MA).

As M9 minimal medium, but solidified with 10g/l Oxoid No. 1 agar.

1.2.9. Water top agar (WA).

Difco Bacto agar, 6.5g/l.

1.2.10. BBL top agar.

As BBL agar, but with 6.5g/l Difco Bacto agar instead of the Oxoid agar.

1.2.11. Stab medium.

L broth solidified with 6.5g/l Difco Bacto agar and including 20μg/ml cysteine and 40μg/ml thymine. Stab vials containing 1.2ml of stab medium are inoculated with fresh cells using a sterile wire and sealed by wrapping Nescofilm (Nippon Shoji Kaisha) around the tops to prevent evaporation. Storage is at room temperature.

1.2.12. P1 adsorption medium.

To make 100ml: 10ml LB; 1ml 1M $CaCl_2$; 89ml water.

1.2.13. Xg plates.

25ml MA plates (+ nutritional supplements if required) overlaid with 2.5ml WA containing 10μl 100mM IPTG (in water), 20μl 2% (w/v) Xg (in dimethylformamide) plus transfected cells (see METHODS).
Xg = 5-Bromo-4-chloro-3-indolyl-β-D-galactoside.
IPTG = Isopropyl-β-D-thio-galactopyranoside. (Solutions of the above stored at -20°C.)

1.3. BUFFERS.

1.3.1. Phage buffer (Φ B).

To make 1l: KH_2PO_4, 3g; Na_2HPO_4, 7g; NaCl, 5g; after autoclaving add

1M $MgSO_4$, 1ml; 0.1M $CaCl_2$, 1ml; 1% gelatine, 1ml.

1.3.2. P1 storage buffer.

20mM Tris.HCl (pH 7.2); 10mM $MgSO_4$; 1% (w/v) ammonium acetate. (Walker and Anderson, 1970)

1.3.3. TE buffer.

10mM Tris.HCl (pH 8.0); 1mM EDTA.

1.3.4. Tris-acetate buffer (TA).

40mM Tris; 20mM CH_3COONa; 1mM EDTA; adjust pH to 8.2 with glacial acetic acid.

1.3.5. Agarose gel loading buffer.

10mM Tris.HCl (pH 7.5); 20mM EDTA; 10% glycerol; 0.01% bromophenol blue, 2mg/ml agarose. After melting the agarose mixture at 100°C, allow to set, and convert the slurry to a smooth emulsion by repeatedly passaging it through a hypodermic syringe. Store at 4°C.

1.3.6. Tris-borate EDTA buffer (TBE).

90mM Tris.HCl (pH 8.3); 90mM borate; 2.5mM EDTA.

1.3.7. Polyacrylamide gel loading buffer.

20% Ficoll; 0.025% bromophenol blue in 0.5x TBE.

1.3.8. T4 DNA ligase buffer (10x).

660mM Tris.HCl (pH 7.5); 66mM $MgCl_2$; 100mM DTT. (Recommended by Bethesda Research Laboratories, Inc.)

1.3.9. Extract buffer (EB).

10mM K_2HPO_4-KH_2PO_4 pH 7.0; 7mM β-mercaptoethanol; 1mM EDTA; 10% (v/v) glycerol. For long term storage of columns, 1mM NaN_3 was included. (Greene *et al.*, 1978)

1.3.10. Storage buffer (SB).

10mM K_2HPO_4-KH_2PO_4 pH 7.0; 7mM β-mercaptoethanol; 1mM EDTA; 50% (v/v) glycerol. (Greene *et al.*, 1978)

1.3.11. 666 buffer.

6mM Tris.HCl (pH 7.5); 6mM $MgCl_2$; 6mM β-mercaptoethanol.

1.3.12. Pronase buffer.

20mM Tris.HCl (pH 7.5); 0.1M NaCl; 1mM EDTA; 0.002% Triton X-100.

1.4. MISCELLANEOUS SOLUTIONS.

1.4.1. Lysis solution.

25mM Tris.HCl (pH 8.0); 10mM EDTA (pH 8.0); 50mM glucose. Add lysozyme to 1mg/ml immediately before use. (Adapted from Ish-Horowicz and Burke, 1981)

1.4.2. Alkaline SDS solution.

NaOH, 0.2N; SDS, 1%. Always make up fresh. (Adapted from Ish-Horowicz and Burke, 1981)

1.4.3. "5M" acetate solution.

Mix equal volumes of 3M CH_3COOK and 2M CH_3COOH. pH should be ca. 4.8 (Adapted from Ish-Horowicz and Burke, 1981)

1.4.4. CsCl solutions for block gradients.

Stock solutions are made by combining 65% CsCl (w/w) in water with Φ buffer in the following proportions:

1.7g/ml CsCl = 4.66ml 65% CsCl + 2.00ml Φ buffer,

1.5g/ml CsCl = 2.22ml 65% CsCl + 1.78ml Φ buffer,

1.3g/ml CsCl = 1.33ml 65% CsCl + 2.11ml Φ buffer.

1.4.5. 30% acrylamide solution.

Acrylamide, 8.7g; bis-acrylamide 0.3g; made up to 30ml with deionised water. Store in a foil-covered bottle at 4°C.

2. ANTIBIOTICS.

Levels of antibiotics used for *M. methylotrophus* are generally different from those used for *E. coli*. Stock solutions are made up in sterile water, and at 100x the final concentration required for *E. coli*, except where otherwise indicated. Storage is at 4°C. The final concentrations used for both species are tabulated below in μg/ml.

	E. coli	*M. methylotrophus*	Comments
Ap	100	100	
Km	25	10	
Nal	25	50	Sodium salt
Rif	50	5	Dissolve in 0.2M HCl
Sm	100	100	For chromosomal Sm^R
Sm	10	10	For plasmid Sm^R
Tc	10	2.5	
Tp	15	150	Lactate form

3. METHODS.

3.1. *IN VIVO* METHODS.

3.1.1. Plate matings.

This was the method of choice for the routine inter- or intrageneric transfer of conjugative plasmids such as RP4 and R751 (Datta *et al.*, 1977; Jobanputra and Datta, 1974). For matings involving *M. methylotrophus*, it was found essential to use fresh cells; in matings between *E. coli* strains, however, cells from plates up to two months old could be used successfully. Plates were prepared selective for the transconjugant and contraselective for the donor and recipient strains. One half of the plate was spread with 50μl of a thick suspension (c. 10^9 cells/ml) of either parent. When dry, a wire loop inoculated with cells of the other parent was streaked from the cell-free half of the plate into the lawn of cells on the other half. After O/N incubation, growth of transconjugants should be evident only at the intersection of the two strains; lack of growth anywhere else acts as a control, indicating that neither parent alone can grow on the plate. Transconjugants were purified twice on plates of the same type before further characterization. Sometimes, the method was varied slightly in that aliquots of a parent strain liquid culture were dropped onto a lawn of the other parent - the principle is the same. It was often found convenient to perform several matings on the same plate.

3.1.2. Filter matings.

For chromosome mobilization and quantitative estimation of plasmid transfer frequencies, matings were performed without selection on nitrocellulose filters (Sartorius membranfilters, 0.45μm pore size, 25mm diameter). The filter was placed on a sintered plate attached to a Buchner flask, allowing suction to be applied *via* a water pump. Usually, 0.5ml of each parent (from saturated O/N cultures) were mixed, and applied to the filter. Gentle suction was applied to draw off the fluid. The filter was removed using fine forceps and placed on the surface of a non-selective agar plate. Mating was allowed to proceed at 30°C or 37°C, where appropriate, for at least two hours for plasmid transfer, or at least sixteen hours for chromosome mobilization. After mating, filters were transferred to Universal glass bottles containing 5ml of M9 or AS1 salts, depending on the recipient. Vigorous vortexing was applied to resuspend the cells and break up the mating pairs. Dilutions were plated out on non-selective media to obtain viable counts, and on selective plates to obtain transconjugants. Control filters with donor only and recipient only cells were always included.

3.1.3. Replica plating/mating.

It is often desirable to screen hundreds of individual clones for various properties such as drug resistance, etc. A convenient way of achieving this is by replica plating. The original protocol for replica plating utilized velvet pads as a printing medium; however, I have found it convenient to use Whatman No. 1 filters for this purpose. In either case, a master plate containing approximately 100-300 well-separated colonies or 50-100 patches manually prepared using a suitable grid was utilized. A standard replicating block was set up in the following way: three 9.0cm diameter filters were placed on the block and covered with a single 12.5cm diameter filter. The locking ring was

then put on, care being taken to avoid wrinkling the filters. The surface of the filter assembly was flamed briefly before use to reduce contamination. The master plate was applied to the block, avoiding air bubbles and smudging. A reference mark was made on the rim of the master plate to indicate orientation before making prints on the relevant test plates, which were then similarly marked. It was found that the master plate would be regenerated if incubated along with the other plates, so no direct copy was necessary. Up to five print plates could be made of each master using this technique. Replica mating is a slight variation of the method in which the master plate contains one parent strain, and the print plates are coated with a lawn of the other parent strain prior to replicating. In this work, replica mating was extensively used in the isolation of *M. methylotrophus* mutants (*q. v.*).

3.1.4. Transformation of *E. coli*.

Plasmid DNA molecules and phage genomes can be introduced into *E. coli* cells by a process known as transformation. Unlike, for example, *Pneumococcus* species, *E. coli* is not naturally competent to take up DNA molecules. There is a technique consisting of treating cells with calcium chloride which makes a certain proportion of the population competent, and it is this one which is outlined below. (N.J. Grinter, pers. comm.) Sterile plastic apparatus is used throughout, as calcium treated cells are easily lysed by trace amounts of detergent present in laboratory glassware. A fresh O/N of the strain was diluted 1/20 in 10ml LB, and grown at 37°C until cell density reached 2 x 10^8/ml - usually 100'. The culture was rapidly chilled on ice, and cells pelleted at 4°C. The pellet was resuspended in 10ml cold 100mM $CaCl_2$ and spun again at 4°C. This time, the cells were resuspended in 5ml cold 100mM $CaCl_2$ and left on ice for 20' before centrifugation. Finally, the

pellet was resuspended in 0.5ml cold 100mM $CaCl_2$. The cells were now competent, and could be stored on ice for up to 24 hours. 1 volume of DNA was added to 2 volumes of cells - usually 0.1ml DNA in TE to 0.2ml cells - and heat shocked at 42°C in a water bath for 2'. The cells were left on ice for 1 hour, then 9 volumes of LB added - usually 2.7ml - and the cells incubated with gentle aeration for at least 1 hour to allow the DNA taken up to express its genes. Dilutions were plated out on the relevant selective media. Alternatively, for M13mp7 DNA transformation ("transfection"), after one hour on ice, 0.2ml transfected cells (JM101) were added to the WA top layer mix described above (1.2.13.) at 42°C, before being poured immediately onto MA plates supplemented with thiamine.

3.1.5. P1 lysates.

Before using P1kc to transduce markers, it is necessary to make a lysate of the appropriate donor strain. Unfortunately, P1kc forms very small plaques on *E. coli* and lysates are sometimes difficult to titre. For plate lysates, thick (50ml) L agar plates containing 10mM $CaCl_2$ were prepared. These plates were not dried, and were used where possible on the same day. The recipient strain was grown to late log phase, about 8 x 10^8 cells/ml, in LB + 10mM $CaCl_2$. 10^5 - 10^6 P1kc phage were mixed with 0.2ml aliquots of the culture and allowed to adsorb for 25'. 2.5 - 3.0ml of BBL top agar + 10mM $CaCl_2$ were added to each mixture and the suspension poured onto a pre-warmed plate. After 8 - 10 hours (or O/N) incubation, complete lysis was usually evident. The lysate was transferred from the surface of the agar using a glass spreader into a Universal container. A few drops of chloroform were added, and the lysate vortexed vigorously for at least 30 seconds, to lyse and kill remaining cells. The lysate was placed on ice for 30' to solidify agar

particles. It was found best to remove cell debris and agar fragments by spinning the lysate at 18,000 rpm for 10' at 4°C. The supernatant was removed and dilute tenfold in P1 storage buffer. The lysate was titred on any convenient P1-sensitive strain using P1 agar plates and $CaCl_2$ supplemented BBL top agar. Titres from 10^9 - 10^{11} pfu/ml were usually obtained.

3.1.6. P1 transduction.

Generalized transduction by phage P1kc is an important technique for strain construction in *E. coli*. The procedure described here is derived from Lennox (1955). The recipient strain was grown to log phase (about 2 - 3 x 10^8 cells/ml) in LB. Cells were pelleted and concentrated 10-fold by resuspension in P1 adsorption medium. Enough P1 lysate was added to give an m.o.i. of 0.01 - 0.1, and the volume made up to 1ml with P1AM. Phage were allowed to adsorb by incubating the mixture without shaking for 25' at 37°C. 20 μl of 50% sodium citrate were added and the cells spun down at RT. The pellet was resuspended in 10ml LB with 1% sodium citrate and incubated at 37°C for at least 1 hour to allow expression of the transduced genes. Dilutions were plated out on the relevant selective media. N.B., if the selective gene was *rpsL* (Sm^R), the pellet was suspended in 25ml LB and incubated O/N to allow full expression of the Sm^R phenotype, before plating out.

3.1.7. Detection of plasmid restriction in *M. methylotrophus*.

Wild-type *M. methylotrophus* produces a restriction endonuclease, *Mme*II, which recognizes the sequence 5'-GATC-3'. The same sequence is also recognized by the *dam* methylase of *E. coli*. A replica mating technique was devised to allow discrimination between *M. methylotrophus* strains producing

*Mme*II (*mmeB*$^+$) and mutant strains lacking the enzyme (*mmeB*$^-$). The master plate consisted of isolated colonies or gridded patches of the *M. methylotrophus* clones to be tested. Two antibiotic-supplemented AS1 agar plates appropriate to the plasmid in use (i.e., Km or Tc for RP4; Tp for R751; Sm for S-a) were prepared. One was coated with a 0.1ml lawn of a *dam*$^-$ *E. coli* donor strain and the other with a similar lawn of the equivalent *dam*$^+$ strain. The master plate was replica plated first on to the *dam*$^-$ plate and then on to the *dam*$^+$ plate. After 24 - 36 hours incubation, the plates could be read. Those *M. methylotrophus* clones that were *mmeB*$^+$ restricted incoming plasmid DNA which lacked *dam* methylation, and were thus very poor recipients of DNA from a *dam*$^-$ donor. Conversely, *mmeB*$^-$ clones did not restrict unmethylated DNA, and so were efficient recipients, giving rise to a patch of growth of R$^+$ recipients on the first print plate. All clones on the master plate gave rise to patches of growth on the second print plate, since the plasmid DNA transferred in this case was protected against *Mme*II restriction.

Another endonuclease, *Mme*I, which recognizes the sequence 5'-TCCPuAC-3', is produced by *M. methylotrophus*. The isolation of a probable *Mme*I$^-$ mutant (CBM22) was achieved using the following *in vivo* assay. Patches or colonies of *M. methylotrophus* to be tested were replica plated onto a lawn of *E. coli* donor, e.g., CB12 (*dam*$^+$, RP4) pre-spread on an AS1 plate supplemented with an appropriate antibiotic, e.g., Tc. Areas of growth which appeared after 24 - 36 hours were of two clearly distinguishable types: the putative *Mme*I$^-$ mutant CBM22 produced patches 5 - 10 times as dense as *Mme*$^+$ strains. At first, this phenotype was taken to be evidence of *in vivo* restriction, but later work (see Results 4.4.) suggesting that CBM22 may be a mutator strain cast some doubt on this interpretation.

3.1.8. Isolation of a *mmeB* mutant induced by UV light.

0.1ml aliquots of dilutions of the UV-mutagenized O/N culture were plated out on AS1 agar. After 48 hours growth, the five 10^{-5} dilution plates with a total of ca. 3600 colonies were selected for screening. These were replica mated onto a 50μl lawn of CB10 (*dam*$^-$, contains RP4) on AS1 agar + Km plates. After 36 hours incubation, one positive signal - a patch of growth on one replica plate - was detected. The area on the master plate corresponding to the positive signal contained 25 colonies. These were picked with a sterile wire and patched onto a grid plate for further screening. This was again performed by replica mating similar to that just described. One patch proved positive, and was purified by streaking for single colonies three times. The strain thus derived was designated CBM13. *In vitro* analysis of cell extracts later showed that the strain produced no detectable *Mme*II activity.

3.1.9. An *in vivo* test for *dam* methylation.

The availability of *mmeB*$^+$ and *mmeB*$^-$ strains of *M. methylotrophus* made possible a simple and specific *in vivo* test for *dam* methylation. This was used extensively to aid in the construction of *dam*$^-$ strains by P1 transduction. Clones to be tested for *dam* methylation were patched out in a grid pattern on 50ml plates supplemented with an antibiotic appropriate for the resident plasmid. After O/N growth, each plate was replica mated onto two antibiotic-supplemented AS1 agar plates, the first overlaid with a 0.1ml lawn of wild-type *M. methylotrophus*, the second with a similar lawn of CBM13 (*mmeB*). All plates were incubated at 37°C for 24 - 36 hours. Patches from the master plate which gave rise to patches of *M. methylotrophus*/RP4 growth on the first print plate were *dam*$^+$; those which did not were *dam*$^-$. As a

positive control, all patches from the master gave rise to patches of growth on the second print plate.

3.2. *IN VITRO* METHODS - DNA.

3.2.1. Ethanol precipitation of DNA.

In many experimental procedures, a DNA solution has to be treated serially with different enzymes requiring different buffers. The most convenient way of changing the buffer is to precipitate the DNA in the first buffer and resuspend the dried pellet in the second. The procedure followed in this work was as follows. If necessary, protein in the DNA solution was extracted with phenol, and the phenol extracted with ether. The solution was made 0.3M in sodium acetate by adding one ninth volume of 3M stock solution (pH 5.8). To this was added two volumes of cold (-20°C) ethanol. If small amounts of DNA were involved (less than 100ng), Dextran sulphate T200 (Pharmacia) was added to a final concentration of 1μg/ml. This substance acts as an inert DNA carrier (B.K. Ely, pers. comm.). The DNA was precipitated by incubating the mixture at -70°C for two hours or O/N at -20°C. Alternatively, rapid precipitation was achieved by incubation in a dry ice/ethanol bath for 15'. The DNA was pelleted by spinning in a microfuge at 4°C for 15'. After carefully removing the supernatant, the pellet was washed with cold (-20°C) 70% ethanol to eliminate traces of precipitated salt. Finally, the pellet was dried under vacuum before being resuspended in the desired buffer by gentle rotary shaking at 37°C. DNA of high molecular weight, e.g., chromosomal DNA, took a long time (1 - 2 days) to redissolve; usually, 1 - 5 hours was sufficient.

3.2.2. Rapid purification of DNA on G-50 Sephadex columns.

Impurities in DNA solutions often inhibit restriction enzymes and ligase. A simple and rapid way of removing such impurities is to pass the DNA through a small G-50 column. The G-50 Sephadex was equilibrated in TE buffer before use. 1ml of slurry was put in a blue polypropylene tip (Gilson) plugged with siliconised glass wool and spun at 3000 rpm in a bench centrifuge to remove excess buffer. This was repeated until no more buffer could be spun out. Then, a small Eppendorf vial was fitted to collect the effluent, and the DNA solution loaded onto the gel. The column was spun as before until the volume of effluent voided was near to the volume loaded. Recovery of DNA was usually 90 - 100%.

3.2.3. Preparation of plasmid DNA.

The method most often used was a modification of that of Birnboim and Doly (1979) based on an alkaline-SDS extraction. The protocol is suitable for preparation of plasmid DNA from 200ml O/N cultures of the plasmid bearing strain. The cells were spun down at 4°C for 10', 10000 rpm in an MSE 6 x 250ml rotor. The pellet was resuspended in 8ml of lysis solution and left on ice for 5'. 16ml of alkaline SDS were mixed in, and the lysate left on ice for 4'. 12ml cold "5M" acetate were added, and the lysate mixed well on ice until the viscosity of the mixture decreased noticeably, and large amounts of a white floccular precipitate formed. The precipitate was pelleted by centrifugation at 10000 rpm for 10' at 4°C in an MSE 8 x 50ml rotor (no brake). Some debris remained in the supernatant at this stage, and was removed by pouring the fluid through a funnel plugged with glass wool. 20ml isopropanol at RT was added to the filtrate to precipitate nucleic acids. After allowing to stand at RT

for 15', DNA and RNA were pelleted by centrifugation at 10000 rpm for 10' at 20°C in the 8 x 50ml rotor. The pellet was gently rinsed with cold 70% ethanol. Alcohol was removed by ether extraction, and the pellet dried down under vacuum. When dry, it was dissolved in 4.6ml TE buffer. The solution was transferred to a 10ml polycarbonate tube (MSE). 4.752g CsCl were added and dissolved by gentle inversion. Finally, 0.2ml of ethidium bromide solution (10mg/ml) was added, and the volume made up to 10 ml with paraffin oil before spinning for 48 hours in a fixed-angle MSE 10 x 10ml rotor at 40000 rpm (20°C). After the run, two bands were visible under UV illumination: a faint, diffuse upper band of chromosomal DNA and a lower, sharp band of plasmid DNA. The chromosomal DNA was removed from above using a hypodermic syringe with a 38 x 1.1mm needle, and discarded. The plasmid DNA was similarly removed in a total volume of ca. 1ml. Ethidium bromide was removed by several extractions with CsCl-saturated isopropanol. The DNA solution was then dialysed against TE to remove CsCl. Its concentration was estimated by measuring absorption at 260 and 280nm , or by running a sample on a gel with a standard DNA of known concentration. NOTE: If the first CsCl gradient failed to give a sharp plasmid band, the DNA was rebanded in a second CsCl gradient.

3.2.4. Preparation of λ DNA.

The clarified lysate was centrifuged at 21000 rpm for 3 hours at 4°C in an MSE 10 x 100ml rotor to pellet the phage. To each tube was added 5ml Φ buffer, and the phage resuspended by gentle rotary shaking. After pooling, remaining debris was removed as before. Free nucleic acids in the

supernatant were enzymatically removed by treatment with RNase and DNase at a final concentration of 10μg/ml for 1 hour at room temperature. Once again, the phage were pelleted by cenrifugation at 27000 rpm for 3 hours at 4°C. The pellet was resuspended in 3ml Φ buffer by gentle rotary shaking, and the debris removed as before. Phage were usually purified on a CsCl step gradient (Miller, 1972). 3ml of a 1.3g/ml CsCl solution were pipetted into a 14ml polycarbonate centrifuge tube (MSE). Using Pasteur pipettes, 2ml of a 1.5g/ml CsCl solution followed by 2ml of a 1.7g/ml CsCl solution were successively underlaid. The phage preparation was layered on top of the gradient, and the tube spun for 2 hours at 35000 rpm in a 6 x 14ml swing-out rotor (MSE) at 20°C. The phage band in the 1.5g/ml layer was extracted through the side of the tube with a hypodermic syringe. CsCl was removed by dialysis against Φ buffer. If a very pure phage preparation was required, this stage could be followed by an equilibrium CsCl gradient run. For this, the phage suspension was made 41.5% in CsCl by adding 0.71g CsCl/g phage suspension. The volume was made up with 41.5% CsCl in a 14ml tube, and the phage banded by centrifugation at 33000 rpm for 36 hours at 20°C in the 6 x 14ml swing-out rotor. The band was collected through the side of the tube with a hypodermic syringe. Pre-digested pronase was added to a final concentration of 1mg/ml, and the phage dialysed against 500ml pronase buffer for 2 - 3 hours at 37°C. Protein was removed by two extractions with pre-equilibrated phenol. The DNA was exhaustively dialysed against TE buffer to remove all traces of phenol.

3.2.5. Restriction.

Type II restriction endonucleases have a strict requirement for Mg^{++} ions to perform scission: other factors, particularly NaCl concentration, affect the rate

of reaction. With this in mind, assay buffers which optimize the rates of reaction have been described for each restriction enzyme. Both the enzymes produced by *M. methylotrophus* work best in a buffer containing no NaCl i.e., 666 buffer. Reactions were usually performed in a total volume of 20μl containing 0.5 - 2.0 μg of DNA, 2μl 10x assay buffer and 0.5 - 5 units of enzyme, the volume being made up with sterile water. A unit is defined as the amount of enzyme required to completely cleave 1μg of DNA in 1 hour under optimum conditions. For this work, digests were incubated for 1 - 2 hours at 37°C (except for *Taq*I, 70°C) and the reaction stopped by heat killing at 70°C for 5' or addition of loading buffer containing EDTA. If serial digestions were performed, the enzyme which required the least NaCl was added first; this reaction was stopped by heat killing at 70°C for 5', NaCl solution added to give the appropriate concentration and the next enzyme added for the second digestion. For some combinations of enzymes, this procedure did not work; for these, the DNA was ethanol precipitated from the first reaction and resuspended in the approriate buffer before adding the second enzyme. In some cases, for example, *Hin*dIII and *Eco*RI double digests, the enzymes will work simultaneously in the same buffer. Reaction volumes and constituents were scaled up for preparative gels and production of fragments for cloning. Digests involving λ DNA were always heat killed before gel analysis to dissociate the sticky ends (*cos*).

3.2.6. Ligation.

T4 DNA ligase is an enzyme which catalyzes the formation of covalently-joined hybrid DNA molecules from precursors with complementary single-stranded extensions ("sticky ends"). Also, under certain conditions, it can ligate molecules with no single-stranded extensions ("blunt ended"). It is

therefore possible, using ligase in conjunction with restriction enzymes, to construct recombinant DNA molecules which can be introduced into cells by transformation or transfection. The procedure is quite straightforward. In this work, the vector, usually a plasmid or M13 bacteriophage derivative (Messing *et al.*, 1981), was cleaved with a suitable restriction enzyme and mixed with DNA fragments produced by digestion with a restriction enzyme giving similar sticky ends. The total concentration and relative amounts of each type of DNA in the mixture is critical: conditions which maximize the final yield of recombinants can be calculated from the theoretical analyses of Dugaiczyk *et al.* (1975). For most purposes, however, it is sufficient to use "rules of thumb" to calculate the DNA composition of the ligation mix. For example, if DNA was to be cloned into M13mp7, a three-fold molar excess of fragment over vector was used. Similarly, for general purpose cloning into plasmids such as pBR322, the amount of fragment used ranged from equimolar to a three-fold molar excess. The amount of digested vector included was usually from 10 - 100ng in a reaction volume of 20μl. Blunt-end ligations required higher DNA concentrations and were incubated for a longer period (24 - 36 hours). A typical 20μl reaction mix would contain 50ng DNA (fragment(s) + vector), 2μl 10mM ATP, 4μl 10x ligation buffer and 0.1 - 1.0 units of T4 DNA ligase (Boehringer), the volume being made up with sterile water. The reaction mixture was usually incubated at 12°C for 8 - 20 hours before heat killing (70°C, 5') and transformation or transfection. An aliquot of the ligated DNA was usually run on an agarose gel to estimate the extent of ligation.

3.2.7. Agarose gel electrophoresis.

All agarose gels used in this work were run on a horizontal 200ml slab gel apparatus. 1% agarose gels were found to be satisfactory for most purposes,

although lower concentrations were occasionally used to resolve high molecular weight fragments. The agarose was melted in 0.5x TA buffer and allowed to cool to about 60°C. The perspex gel former was made watertight by sealing the two ends with end-pieces and autoclave tape. The sample wells were generated by locating a comb near one end with plasticine, ensuring that the teeth were about 3mm from the bottom of the bed. For assaying column fractions, up to three 12-teeth combs were used in the same gel. Ethidium bromide to a final concentration of 0.5μg/ml was added to the molten agarose prior to pouring. When set, the comb was carefully removed. If covered with clingfilm, the gel could be stored at 4°C for several days before use. The "Concorde" gel apparatus consists of a tank (capacity 2l) with platinum wire electrodes and a platform for the gel. Before loading, the end-pieces were removed and the gel placed on the platform. 2l of 0.5x TA buffer containing 0.5μg/ml EtBr were added, enough to just submerge the gel. The samples containing loading buffer were loaded into the wells using a 100μl Eppendorf micropipettor. Gels were run at maximum current (200mA) driven by an LKB 2103 power supply. Under these conditions, the run-time necessary was from 0.5 - 5 hours, depending on the resolution desired. Often, the heat generated by electrophoresis was sufficient to evaporate a small amount of running buffer: care was taken, therefore, throughout the run to ensure that the gel was still totally submerged. Failure to do this resulted in drying out and shrivelling of the exposed areas of gel.

After the run, the gel was photographed with a Polaroid MP-4 camera under long-wave UV light provided by a transilluminator (Fotodyne 3-3000). A red filter (Kodak Wratten filter No. 9) and Polaroid 4x5 Land film (No. 57) were used. For fragment preparation, the gel was pre-cooled at 4°C for 30':

this made the agarose stiffer and easier to cut in straight lines. Then, the boundaries of the band were marked with scalpel blade nicks using the shortest possible exposure to UV light. The agarose containing the fragment could then be excised easily under normal light.

3.2.8. Polyacrylamide gel electrophoresis.

Polyacrylamide gels are used to resolve small DNA fragments of up to 1.2kb. In this work, 5% gels were used exclusively, but higher percentage gels (up to 20%) also have applications in DNA electrophoresis. To make a 5% gel, a gel kit consisting of two 380 x 197mm glass plates (one with "rabbits ears") separated by plastic 1mm thick spacers, and sealed by tape along three edges was prepared, and inclined at a slight angle on a flat surface. To make a 5% gel, 25ml of 30% acrylamide was mixed with 5ml fresh 1.6% ammonium persulphate solution, 7.5ml 10 x TBE buffer and 112.5ml deionized water. The solution was degassed before adding 100μl undiluted TEMED. This was then poured carefully into the gel kit, avoiding air bubbles. A 14 place comb was inserted at the top of the gel. After polymerization was complete (30' - 60'), the tape and bottom spacer were removed. The kit was installed in the "Shandon" gel apparatus, and held vertical with clips. The comb was removed, and the wells immediately flushed out with 0.5 x TBE to remove partially polymerized acrylamide. 0.5 x TBE was used to fill top and bottom tanks, and any air trapped was removed using a syringe. 5 - 10μl loading buffer was added to samples prior to layering at the bottom of the wells using drawn-out capillaries. Gels were run O/N at 100V, and then stained in 0.5μg/ml EtBr for 30' before being photographed as for agarose gels.

3.2.9. Electroelution of DNA fragments.

The DNA (in agarose) was placed in a dialysis sac (Visking 2-18/32" size) containing sufficient 0.5x TA buffer to surround the segment. The end was tied, and the sac placed parallel to the electrodes on the platform of a 'Concorde' apparatus filled with 2l 0.5x TA plus 0.5μg/ml EtBr. Current was applied (200mA) as for normal submarine gels until the DNA was eluted from the agarose. This event could be followed by illuminating the sac with a hand-held UV lamp (UVGL-58; UV Products): i.e., when the DNA had eluted, a band of fluorescence was visible on the inside surface of the sac towards the anode. To dissociate the DNA from the dialysis membrane, the polarity was reversed for 30 seconds. The DNA solution was transferred to a 10ml conical bottomed polypropylene tube, and the volume reduced to ca. 400μl by repeated extractions with isobutanol, which also removed the EtBr. At this stage, DNA was ethanol precipitated, dissolved in a suitable buffer (usually TE) and further purified by passaging through a small G-50 Sephadex column. Fragments isolated in this way were good substrates for restriction enzymes and ligase.

3.3. *IN VITRO* METHODS - ENZYMES.

3.3.1. Small-scale preparation of *Mme*I and *Mme*II.

The protocol described here is a scaled-down version of the large-scale purification method, and was extensively used to identify which restriction enzyme was not produced in *M. methylotrophus* mutants lacking restriction activity *in vivo*. If the presence or absence of *Mme*II alone is of interest, the volumes and weights given below may be reduced by a factor of five. A 25ml O/N culture of the strain to be tested was spun at 4°C for 10' at 10000

r.p.m. (Solutions were pre-cooled to 4°C and all further operations carried out at the same temperature). 1ml cold EB was added, and the pellet resuspended. The suspension was transferred to a 10ml plastic tube and the cells disrupted by sonication (six 5 second bursts with 25 second cooling intervals, 16 microns peak-to-peak). The lysate was transferred to a 1.6ml Eppendorf vial and the debris pelleted by spinning in a microfuge for 15'. Nucleic acids in the supernatant were precipitated by adding 100μl of a freshly made 20% streptomycin sulphate solution (in water), and inverting gently several times. The precipitate was removed by centrifugation as for the cell debris. 1.0ml of the supernatant was taken off, and saturated to 40% with ammonium sulphate by adding 0.235g of crystals. After all the crystals had dissolved, the mixture was left on ice for at least 15'. The precipitated protein was removed by centrifugation and discarded. A further 67mg of ammonium sulphate was added to the supernatant to make it 50% saturated, and the mixture left on ice for at least 45'. After pelleting the protein precipitate, the supernatant was discarded and the pellet, containing whichever restriction enzymes were present in the cells, was dissolved in 50μl TE buffer. The protein solution was dialysed against TE for at least 90' with at least one change of buffer. The extract prepared in this way was stable for only 24 hours and it was found best to use it immediately. To detect *Mme*II activity, 0.5μg of DNA prepared from a *dam*$^-$ *E. coli* strain was digested with 4μl of the extract in a total volume of 20μl for 60' at 37°C. To assay *Mme*I activity, it was necessary to use 10μl or more of extract and DNA prepared from a *dam*$^+$strain, all other conditions being the same. For unknown reasons, the amount of *Mme*I activity detected by this method varied considerably from experiment to experiment (see Results 4.5.). Often, the extract contained a small amount of RNA, and this was removed at the end of the 60' by adding 2μl of a 200μg/ml stock solution of RNAse (Sigma - ribonuclease A, bovine

pancreas) and incubating for a further 5' at 37°C. After heat killing the endonuclease reaction, the sample was run on a 1% agarose gel in the usual way.

3.3.2. Purification of *Mme*I.

At the beginning of the project, much time was devoted to devising a procedure for purifying an endonuclease activity from *M. methylotrophus*. Even when an activity was identified (*Mme*I), it was found that subsequent attempts to repeat the procedure often failed to produce detectable enzyme. The method is therefore included in Results 2.1., in order to stress its lack of reproducibility: possible reasons for the difficulties encountered in purifying *Mme*I are also given in the same section. Fortunately, enough material from succesful purifications was obtained to enable the recognition sequence of the enzyme to be established.

3.3.3. Purification of *Mme*II.

The method which follows is based on that described by Greene *et al.* (1978). When a probable *Mme*I$^-$ mutant (CBM22) became available, *Mme*II was prepared from it, rather than from wild type *M. methylotrophus*. A 2l culture was grown to saturation O/N with vigorous aeration. Cells were pelleted by centrifugation (4°C, 10000 r.p.m., 10') in an MSE 6 x 250 rotor. All subsequent operations were carried out near to 4°C. Pellets were pooled and resuspended in 25ml cold EB. The cells were lysed using a French press at 12000 p.s.i., and the extract diluted to 45ml with EB. Cell debris was removed by centrifugation (18000 r.p.m., 60') in an MSE 8 x 50 rotor. Nucleic acids in the supernatant were precipitated by the dropwise addition of

freshly-prepared 20% $SmSO_4$ (in EB) to a final concentration of 1%, or until no further precipitate was formed. The precipitate was spun down (18000 r.p.m., 40', MSE 8 x 50) and discarded. The volume of the supernatant was determined, and 235g/l solid $AmSO_4$ added slowly to give ca. 40% saturation. The extract was gently stirred for at least 1 hour using a magnetic stir bar. Protein precipitated at this stage was spun down (18000 r.p.m., 30', MSE 8 x 50) and discarded. A further 60g/l of solid $AmSO_4$ was added to the supernatant to give a saturation of 50%, and this was stirred as before for at least 2 hours - conveniently O/N. The precipitated protein, which contained most of the restriction activity, was pelleted by centrifugation (18000 r.p.m., 30', MSE 8 x 50), and the supernatant discarded. The pellet was dissolved in 1ml EB, and residual salt removed by dialysis against EB for 4 hours, with at least one change of buffer. The extract was loaded onto a pre-equilibrated DE52 (Whatman) column (15mm x 300mm, Pharmacia K15/30), and the column washed with 2 - 3 column volumes of EB. Bound proteins were eluted with a 300ml gradient of 0.0M - 0.5M NaCl in EB. The flow rate was controlled with a peristaltic pump and not allowed to exceed 40ml/h. 60 5ml fractions were collected using a LKB Ultrorac 7000 fraction collector. The eluate was assayed for endonuclease activity by incubating ca. 0.3μg of DNA (from a *dam*$^-$ strain) with 2μl aliquots from alternate fractions in 20μl of 666 buffer for 1 - 3 hours at 37°C. The reactions were heat-killed and run on a 1% agarose gel to identify active fractions. These were pooled, and either (i) dialysed against SB for 2 hours with one change, and then stored at -20°C or (ii) subjected to a further stage of purification by column chromatography. Prior to (ii), the active fractions were dialysed against EB to eliminate NaCl. The pool was applied to a pre-equilibrated phosphocellulose (Whatman P11) column (25cm x 1cm), and washed, as before, with 2 - 3 column volumes of EB. Elution was carried out with a 100ml gradient of 0.0M - 0.5M NaCl in

EB, and 50 2ml fractions were collected. Alternate fractions were assayed as before, and active fractions pooled and dialysed against SB before storing at -20°C. If a more concentrated enzyme preparation was required, the pooled fractions were dialysed against solid polyethylene glycol 6000 (BDH) before dialysis against SB. Enzyme prepared in this way was stable for many months at -20°C without appreciable loss of activity.

BACTERIAL STRAINS

1. Non-methylotrophs (*E. coli*).

Strain	Genotype	Source/reference
CB1	GM31: *dcm gal ara lac xyl thr leu thi tonA tsx rpsL*	Marinus, 1973
CB2	GM48: *dam-3 dcm gal ara lac thr leu thi tonA tsx*	Marinus, 1973
CB3	CB1 (pBR322) Ap^R Tc^R	This work
CB4	CB2 (pBR322 dimer) Ap^R Tc^R	This work
CB9	W3110: *thyA* (S-a) Cm^R Km^R Sm^R Su^R	N.J. Grinter
CB10	CB2 (RP4) Ap^R Tc^R Km^R	This work
CB11	CB2 (S-a) Cm^R Km^R Sm^R Su^R	This work
CB12	J5-2 (λ^+, RP4) *pro met rpoB*	N.J. Grinter
CB15	CB1 (RP4) Ap^R Tc^R Km^R	This work
CB16	CB1 (pAT153) Ap^R Tc^R	This work
CB17	CB2 (pAT153) Ap^R Tc^R	This work
CB18	CB2 *rpsL* spontaneous mutant	This work
CB19	JM101: *lac pro supE thi* (F' ProAB$^+$ *traD36 lacI^q* ZΔM15)	Messing, 1979
CB20	CB2 (pBR322 monomer) Ap^R Tc^R	This work
CB23	ED8654: *supE supF hsdR met*	W.J. Brammar
CB24	CB23 (R751) Tp^R	This work

CB25	CB23 (R751) *dam-3 rpsL* TpR	This work
CB26	CB12 *dam-3 rpsL*	This work
CB33	DS903 (=AB1157 *recF*) (pLG221) *cib::Tn5*	B. M. Wilkins
CB34	W3110 *recA* (pNJ5073) Tc^R, Tp^R	N. J. Grinter
CB35	CB23 *dam-3 rpsL* (pNJ5073) Tc^R, Tp^R	This work
CB36	CB35: plasmid-free	This work

2. Methylotrophs (*M. methylotrophus* except where noted.)

Strain	Genotype	Source/reference
CBM1	*Methylophilus methylotrophus* - wild type	D. Byrom
CBM2	CBM1 (RP4) Ap^R Tc^R Km^R	This work
CBM3	K81: wild type	D. Byrom
CBM8	CBM1 Rif^R	This work
CBM13	CBM1 *mmeB1* UV induced	This work
CBM14	CBM13 (RP4) Ap^R Tc^R Km^R	This work
CBM15	CBM13 Rif^R	This work
CBM18	CBM1 *mmeB2*::Tn*5*	This work
CBM19	CBM1 *mmeB3*::Tn*5*	This work
CBM20	CBM1 *mmeB4*::Tn*5*	This work
CBM21	CBM1 *mmeB5*::Tn*5*	This work
CBM22	CBM1 *endA1*::Tn*5*	This work
CBM23	CBM1 *mmeB6*::Tn*5*	This work
CBM36	CBM22 (pLP13) Ap^R Tc^R	This work

CBM37	CBM22 (pLP31) Ap^R Tc^R	This work
CBM38	CBM22 (pLP51) Ap^R Tc^R	This work
CBM39	CBM22 (pLP90) Ap^R Tc^R	This work
CBM40	CBM22 (pLP99) Ap^R Tc^R	This work
CBM41	CBM22 (pLP107) Ap^R Tc^R	This work
CBM42	CBM22 (pLP122) Ap^R Tc^R	This work
CBM43	CBM22 (pLP130) Ap^R Tc^R	This work
CBM44	CBM22 (pLP139) Ap^R Tc^R	This work
CBM45	CBM22 (pLP146) Ap^R Tc^R	This work
CBM46	CBM22 (pLP149) Ap^R Tc^R	This work
CBM47	CBM22 (pLP155) Ap^R Tc^R	This work
CBM48	CBM1 Nal^R	This work
CBM49	CBM13 Nal^R	This work
CBM50	CBM22 Nal^R	This work
CBM51	CBM1 Sm^R	This work
CBM52	CBM13 Sm^R	This work
CBM53	CBM22 Sm^R	This work
CBM54	CBM21 (R751) Tp^R	This work
CBM55	CBM22 (R751) Tp^R	This work
CBM58	CBM1 (R751) Tp^R	This work
CBM59	CBM1 *mmeB7* spontaneous mutant	This work
CBM60	CBM1 *mmeB8* spontaneous mutant	This work
CBM61	CBM1 *mmeB9* spontaneous mutant	This work
CBM62	CBM22 *mmeB10* spontaneous mutant	This work
CBM63	CBM22 *mmeB11* spontaneous mutant	This work
CBM64	CBM22 *mmeB12* spontaneous mutant	This work

INTRODUCTION.

1. Restriction and modification of coliphage λ *in vivo*.

The complementary phenomena of restriction and modification have long been recognized as features of *Escherichia coli*. The earliest system in which an attempt was made to study the details of both these processes was the interaction between bacteriophage lambda and *E. coli*. It was observed that phages (designated λ.B) which had been grown on one strain, *E. coli* B, gave a very reduced plating efficiency on a strain with a different host specificity, *E. coli* K. It was said that the new host had "restricted" the growth of the phage. Phages from the few plaques which did form from λ.B/*E. coli* K infections, however, plated normally when allowed to re-infect strain K. It appeared that these phages, designated λ.K, had in some way been "modified" by their first passage through the host such that they were resistant to further rounds of restriction by *E. coli* K. The converse of the above experiment gives a similar result: λ.K phages are restricted by *E. coli* B. Moreover, survivors of this first restriction are able to re-infect strain B with normal virulence - they have become modified to λ.B.

2. Arber's theory.

It was Arber (1965) who first proposed a biochemical model to explain the restriction/modification (R/M) phenotype, based on the knowledge that it was the DNA component of the phage which was affected by both processes (Dussoix and Arber, 1962). He suggested that restriction and modification

were caused by two differing enzymatic activies sharing the ability to recognize the same specific nucleotide sequence. The modification activity, it was proposed, somehow searched its substrate DNA for the target sequence, and there made a specific chemical modification to the DNA. The restriction activity searched the DNA for the same sequence, but could only recognize those target sequences not previously modified. At such a sequence the DNA was cleaved, rendering it biologically inactive. This model neatly and economically explains the phage results, and also explains how the host chromosome avoids auto-digestion.

3. Type I enzymes.

The publication of Arber's model led Meselson and Yuan (1968) to search for R/M enzymes in *E. coli* K12. Without modern assay techniques, this task proved quite time-consuming, but an enzyme with the required characteristics was isolated. The enzyme proved to consist of a complex of at least three distinct polypeptides. The complex, later to be called *Eco*K (Smith and Nathans, 1973), was, under suitable conditions, able to perform either restriction or modification. Both activities were found to require Mg^{++} ions, ATP and S-adenosyl methionine. The conjunction of both activities in one complex later proved to be the exception rather than the rule, however, although enzymes of this nature are still referred to as Type I enzymes.

Type I enzymes have been found in *E. coli* strains K-12, B, 15 and A, and are assumed to be evolutionarily related, since they all map in the same region and can be genetically exchanged by P1 transduction (Arber and Linn, 1969). Nevertheless, they all have different specificities. The K-12 system, for obvious reasons, is the best studied, the genetic locus being divided into

three complementation groups *hsdS*, *hsdR* and *hsdM* which map at 98' on the *E. coli* chromosome (Bachmann and Low, 1980). The genes code for the specificity, restriction and methylase activities respectively.

4. Type III enzymes.

A small group of restriction and modification enzymes, (Type III), represented by only four examples *Eco*P1, *Eco*P15, *Hin*fIII and *Hin*el, share many properties in common with Type I enzymes. These enzymes also require Mg^{++} and ATP for cleavage, but do not extensively hydrolyze ATP. For methylation, only AdoMet and Mg^{++} are required. If all three factors are present, AdoMet appears to stimulate restriction activity as well as acting as a methyl group donor for the modification activity. In this latter case, restriction and modification are competing activities of the same enzyme complex. Another major difference between the groups is that Type III enzymes cleave DNA at a fairly fixed distance (some 25bp) from their recognition sequences. An interesting property of Type III enzymes is their apparent inability to restrict DNA to completion, even under ideal assay conditions. In other words, not every site on every molecule is cleaved, resulting in complex gel patterns when restricted DNA is subjected to electrophoresis. At present, only one possible explanation for this phenomenon has been put forward. Piekarowicz and Brzezinski (1980), in their study of *Hin*fIII, found that the enzyme's activity changed after storage. Fresh enzyme was able to methylate λ DNA without the addition of AdoMet, and cleaved ColE1 DNA randomly at only one of the five possible cleavage sites. After storage for about six weeks at -20°C, however, the enzyme could no longer methylate λ DNA without the addition of AdoMet, and cleaved ColE1 at more than one site per molecule in its absence. Their explanation for this transition of activities was that freshly-purified enzyme

(which they called *Hin*fIII*) had AdoMet tightly bound to it, which dissociated during storage, converting the enzyme to its *Hin*fIII form. The presence of methyl-donor molecules in the *Hin*fIII* form, therefore, means that methylation and restriction are competing, and explains the incomplete cleavage of ColE1. But this can not be the full story, as in the complete absence of AdoMet, the enzyme is still unable to cleave ColE1 at every site. Clearly, more work needs to be done in order to provide a full explanation of this peculiar property of Type III enzymes. The difficulty in obtaining complete digests has meant that the recognition sequences of all the Type III enzymes have been determined by the indirect method of locating the methylation sites. Briefly, DNA is methylated *in vitro* by the purified methylase in the presence of tritiated AdoMet. The position of the methyl-labelled deoxynucleotide within the DNA is then determined by, for example, two dimensional thin-layer chromatography. The sequence common to all methylation sites examined is assumed to be the recognition sequence of the enzyme.

The first Type III enzyme studied was *Eco*P1, coded for by bacteriophage P1 which in its lysogenic state exists as a stably-inherited plasmid in *E. coli*. It was known classically that such lysogens were able to restrict the growth of phages isolated from non-lysogens (Arber, 1965) Hattman *et al.* (1978) were able to identify the sequence (AGACPy) recognized by the P1 modification methylase. A closely-related restriction/modification system is coded for by the P15 plasmid found in *E. coli* $15T^{-}$ (Reiser and Yuan, 1977). The sequence methylated in strains harbouring P15 was found to be CAGCAG (Yuan *et al.*, 1980). The two other known Type III enzymes, *Hin*fIII and *Hin*eI, are also closely related and were isolated from *Haemophilus influenzae* strains Rf and Re respectively (Piekarowicz and Brzezinski, 1980; Piekarowicz, 1982). Both enzymes methylate the same sequence (CGAAT) and cleave about 25bp 3' of

it. The common feature of all three recognition sequences is their asymmetry. Thus, Type III enzymes share two properties with certain Type II enzymes (such as *Mbo*II) which recognize non-palindromic sequences and cleave the DNA a certain distance away (see below).

5. Type II enzymes.

Of far greater significance to the development of molecular biology, however, are the Type II R/M enzymes, which are generally simpler in form and function than the Type I and III enzymes. Type II restriction enzymes, in contrast to the other two classes, cleave the DNA at sites close to or within the recognition sequence, and only require Mg^{++} as a cofactor. Many generate self-complementary, single-stranded DNA extensions in the cleavage process, allowing the possibility of recombining the fragments thus produced *in vitro* and creating novel DNA molecules. Again in contrast to Type I and III enzymes, all known Type II modification enzymes are physically separate from their cognate restriction enzymes and require only Mg^{++} and AdoMet as cofactors: In addition, no Type II restriction or modification enzyme made up of distinct subunits has yet been found.

The first Type II enzyme to be studied in detail was discovered in *Haemophilus influenzae* Rd by Smith and Wilcox (1970). In contrast to *Eco*K, no modification activity was found to be associated with the restriction activity and, moreover, it turned out that the enzyme actually cleaved consistently within its recognition sequence. This enzyme was subsequently completely characterized and its recognition sequence determined. It is now known as *Hin*dII (Kelly and Smith, 1970). The cognate modification system was also investigated (Roy and Smith, 1973) and the enzyme responsible turned out to

be a methylase - i.e., it modifies the DNA by covalently attaching a methyl group to a specific base or bases within the recognition sequence. New techniques, especially agarose slab gel electrophoresis, with DNA staining by ethidium bromide (Sharp *et al.*, 1973), greatly simplified the assaying of column fractions for endonucleolytic activity and a wealth of knowledge about restriction enzymes and their recognition sequences has been accumulated in the last decade (see review by Roberts, 1983).

An interesting subset of Type II restriction enzymes of relevance to this work consists of enzymes recognizing irregular (non-palindromic) sequences. Less than a dozen different specificities of this kind are known (Roberts, 1983), and all but one (*Gdi*II) have the property that they cleave the DNA at some distance outside the recognition sequence. This distance (5 - 18bp) is usually invariable for any particular enzyme. Most such enzymes produce staggered breaks ranging from *Bbv*I, which creates four base 5' projections, to *Tth*111II, which creates two base 3' projections. Interestingly, one bacterial species, *Moraxella bovis*, produces Type II enzymes of both kinds: *Mbo*I, which cleaves GATC sequences, and *Mbo*II which recognizes the sequence GAAGA. The work described below demonstrates that *M. methylotrophus* is a similar species.

Much less, however, is known about Type II modification enzymes, perhaps because they are of little use as tools in the molecular cloning and sequencing of DNA, which, of course, provide the main impetus for research on Type II restriction enzymes. For a survey of known bacterial modification systems, see Brooks and Roberts, 1982.

6. *Eco*RI.

The best characterized Type II enzyme, *Eco*RI, is produced by pMB1, a naturally occurring plasmid ciosely related to CoiE1 (Greene *et al.*, 1981). Because the genes coding for the *Eco*RI restriction and modification system are on a multi-copy plasmid, it is easy to purify large amounts of enzyme to homogeneity: consequently, most of our knowledge about the physical properties of Type Ii enzymes has come from study of this system. The complete nucleotide sequence of the *Eco*Ri locus has been determined independently by two groups (Greene *et al.*, 1981; Newman *et al.*, 1981). Interestingly, although each group sequenced the same segment of two related but different plasmids (pMB1 and pMB4 respectively) which had been propagated separately as laboratory strains for 10 years, the sequences obtained were identical. The genes appear to belong to a single operon, with the endonuclease gene (831bp) transcribed first, followed by the methylase gene (978bp). The coding sequences are separated by an A/T rich, intergenic region 29bp long. Molecular weights calculated from their predicted amino-acid sequences are 31,063 and 38,048 daltons for the endonuclease and methyiase respectively. It is perhaps disappointing to note that computer comparison of the nucleotide sequences of the genes and amino-acid sequences of the proteins revealed no striking similarities which might have indicated a common evolutionary origin. Also, a computer simulation of possible secondary structures adopted by each enzyme revealed marked differences in form. If X-ray crystallographic studies bear this observation out, it suggests that the enzymes have evolved convergently from different ancestors to recognize the same sequence. This would raise an interesting logical problem, because the evolutionary steps involved in such a convergent route would of necessity be much more complex than in the conventionally held theory of divergent evolution from a common ancestor. The latter idea is easy to describe: the ancestral gene coded for the methyiase alone, and gene

duplication gave rise to a copy which evolved to cleave rather than methylate the sequence still recognized by both enzymes. In the convergent evolution model, however, it is difficult to explain how a primitive endonuclease could evolve at all in the absence of a methylase already protecting the host chromosome from degradation. There is also the problem of accounting for the present day proximity of the genes. Unfortunately, neither of these hypotheses can be tested empirically in a reasonable time, so they will probably remain speculative.

A peculiarity of the *Eco*Ri endonuclease is its relaxed specificity under certain assay conditions, when it seems to recognize and cleave certain AATT sequences rather than the canonical GAATTC sequence. Woodhead *et al.* (1980) found that *Eco*RI* activity was most evident in an assay buffer containing 100mM Tris.HCl (pH 7.6), 75mM NaCi, 1mM $MnCl_2$. Indeed, with Mn^{++} instead of Mg^{++} ions in the assay buffer, only *Eco*RI* activity is seen with NaCl concentrations of up to 100mM in a pH range of 7.2 - 8.2. EcoRI* activity is also induced in assay buffers containing relatively high concentrations (1% to 6%) of organic solvents such as dimethylsuphoxide, dimethylacetamide, dimethylformamide and sulphalane. Tikchonenko *et al.* (1978) comprehensively examined *Eco*RI* activity under a variety of assay conditions, and came to the conclusion now generally accepted that it is a modification of *Eco*RI activity rather than a separate, co-purifying endonuclease. High glycerol concentration also induces *Eco*RI* activity: for example, Pribnow *et al.* (1981) used an assay buffer composed of 20mM Tris.HCl (pH 8.5), 2mM $MgCl_2$, 0.1mM EDTA, 20% glycerol with a five-fold excess of enzyme for their *Eco*RI* digests. Interestingly, they found that not all AATT sequences in the DNA were cleaved by the *Eco*RI* enzyme: in particular, NAATTM sequences in which N and M were both different from G and C respectively were resistant to cleavage.

Evidently, *Eco*RI* activity is modulated by the bases flanking the central AATT sequence. Other restriction enzymes, for example *Bam*HI (George *et al.*, 1980), have been shown to exhibit relaxed specificities under various exotic assay conditions.

7. Methylation of GATC sequences.

In *E. coli*, N^6-methyladenine (6-MeA) accounts for 0.38 mole per cent of deoxynucleotide residues (Marinus and Morris, 1975). Thus, approximately one in sixty adenine residues are methylated. While a small proportion of this methylation is due to the activity of M.*Eco*K, most of it is the result of *dam* methylase activity. Marinus and Morris (1973) isolated three *dam*$^-$ mutants, all of which had substantially reduced levels of 6-MeA in their DNA. Mapping experiments (Marinus, 1973) positioned the *dam*-3 allele at 74' on the genetic map of *E. coli*, between the *rpsL* (formerly *strA*) and *trpS* loci. The transduction of the *dam*-3 allele to strains which had not been heavily mutagenized allowed detailed comparative studies of the effects of the *dam* lesion. Perhaps surprisingly, there were several phenotypic differences between *dam*$^+$ and *dam*$^-$ strains: but it was not until McGraw and Marinus (1980) were able to isolate *dam*$^+$ revertants of a *dam*$^-$ strain that the specific effect of the *dam* lesion alone could be assessed. They described six phenotypes other than the primary Dam$^-$ phenotype, loss of adenine methylation:

(i) AP^S, sensitivity to 2-aminopurine;

(ii) UV^S, increased sensitivity to ultraviolet light;

(iii) Vrm^-, inviability of double mutants containing

dam-3 and *recA*, *recB* or *recC* mutations;

(iv) Flh^H, high frequency of homogenotization;

(v) Sli^H, high spontaneous induction of λ prophage and

(vi) Smf^H, high spontaneous mutability.

These secondary effects suggest that the *dam* methylase has a role in DNA repair: some of them can be suppressed by mutations in other genes such as *mutL*, *mutS*, *sbcA* and *sbcB*, all of which are loci implicated in repair or recombination. Glickman *et al.* (1978), in following up the suggestion made by Wagner and Meselson (1976) that methylation may be the signal for post-replicative strand discrimination, obtained evidence that the *dam* methylase was responsible. The Wagner and Meselson model is simple to state: they proposed that the newly-synthesized DNA strand is transiently non-methylated after replication; the mismatch repair system, in scanning the duplex for mismatched bases, can therefore identify and excise the incorrect base (which must be on the unmethylated, "new" strand) and replace it with the base complementary to the base opposite, on the "old" strand. In the absence of methylation, either mismatch repair can not operate at all, or, with no basis for strand discrimination, the system fixes rather than excises the incorrect base half the time. Clearly, in both cases, higher mutation levels will result. The model is attractive because it can account for the secondary phenotypes associated with dam^- strains. The mutator (Smf^H) phenotype is an obvious consequence of loss of strand discrimination. All of the other phenotypes may be consequences of double stranded breaks erroneously produced when the mismatch repair system operates on opposing strands in the same region. Excision repair involves quite extensive exonuclease activity, and clearly the simultaneous removal of opposing strands will result in a cleaved duplex. Double stranded breaks can only be repaired by recombination processes.

accounting for the Vrm^-, Flh^H and Sli^H phenotypes. If the number of such breaks is high, then recombinational repair may not be able to cope, resulting in cell death. This explains the sensitivity of *dam*$^-$ strains to base analogues such as 2-AP (AP^S phenotype); and to UV light (UV^S phenotype); agents which both, in different ways, cause a large number of mismatches to be formed. There is no suggestion in the Wagner and Meselson model that the methylase is responsible for any function other than that of "labelling" DNA strands. Therefore, one prediction of the model is that mutations which suppress some of the secondary phenotypes of *dam*$^-$ strains will be in genes responsible for the mismatch repair system itself. Glickman and Radman (1980) selected mutants of *dam*$^-$ strains which were no longer sensitive to 2-AP. They found that all such mutants had acquired additional lesions, with mutator phenotypes, which mapped at the *mutH*, *mutL* and *mutS* loci: no *dam*$^+$ revertants were obtained. Because such double mutants were viable in a *recA* background, they inferred that these *mut* genes did indeed code for structural or regulatory components of the post-replication mismatch repair system. *dam* is also implicated in the DNA replication system of *E. coli* (Gomez-Eichelmann and Lark, 1977).

8. Roles of restriction endonucleases *in vivo*.

As described above (Introduction 1.), restriction endonucleases were named for their ability to destroy unmodified, invading bacteriophage DNA. This is insufficient reason, however, to suppose that defence is the only (or indeed primary) *in vivo* role of this class of enzymes. What can be said with confidence is that, in all cases so far studied, restriction enzymes are dispensable to the cell: wherever mutants lacking them have been sought, they have been found (delGiudice, 1979; Duncan *et al.*, 1978; this work). Protection against phage attack is not solely a laboratory phenomenon: many types of phage have evolved mechanisms to avoid host-encoded restriction either by harbouring methylation genes (e.g. P1; Hattman *et al.*, 1978); anti-restriction endonuclease genes (e.g. T7; Studier, 1975): or by use of unusual bases in their DNA (e.g. the glucosylated hydroxymethylcytosine-containing DNA of T4; O'Farrell *et al.*, 1980). The powerful effect of a restriction system (MmeII) in greatly reducing (under certain conditions) intergeneric conjugal plasmid transfer to *M.methylotrophus* is described in this work. Unfortunately, no ecological work has been published describing the efficacy or otherwise of restriction systems in protecting bacterial species in a natural environment.

One possible role for restriction enzymes *in vivo* is that of site-specific recombination, a role hinted at by their use *in vitro* to create novel DNA molecules in molecular cloning (see Introduction 5.). In a series of experiments to determine whether a Type II endonuclease could mediate site-specific DNA rearrangements *in vivo*, Chang and Cohen (1977) introduced

plasmids containing EcoRI recognition sites into an E.coli strain which produces the EcoRI endonuclease and found that recombination did indeed take place between the plasmids: subsequent in vitro gel analysis of the recombinants revealed that they had acquired or lost specific EcoRI fragments. These results appear to confirm that the EcoRI endonuclease can perform site-specific recombination in vivo, but whether this is a biologically signficant phenomenon outside the laboratory is another question.

Although there has been considerable speculation on the roles of restriction enzymes other than the two described above, there is very little experimental evidence to support such roles. I will therefore confine myself to instances which are, at least in part, experimentally confirmed.

Lacks and Greenberg (1977) investigated two closely related strains of Diplococcus pneumoniae which produce complementary endonucleases, DpnI and DpnII. Both enzymes recognize the sequence GATC, but DpnI cleaves only when the adenine residues are methylated, and DpnII only when they are not. Hemimethylated DNA is not cleaved by either enzyme. By inference, therefore the corresponding methylases are complementary and, presumably, exchange of genetic material between the strains is subject to severe restriction. In their Discussion, the authors entertain the possibility (supported by strain provenance) that the DpnII strain arose from the DnpI strain by a (reversible) differentiation process. They point out a number of phenotypic differences between the strains, and suggest that the genes responsible for these differences are controlled at the level of

transcription by the complementary restriction/modification systems. That is, each strain contains both systems, but only one is expressed. Each methylase, they postulate, modulates expression (specifically, by methylating promoter or operator sequences) of a different set of genes, so that the differentiated state of the bacterium can be arrived at by switching on one or other of the restriction/modification systems. The presence of a restriction endonuclease simply acts to maintain the current differentiated state. They do not specify the mechanism by which one cell type, albeit rarely, can give rise to the other type. Such a differentiation role for restriction systems cannot be widespread, however, since DpnI and DpnII are the only known pair of complementary restriction enzymes.

A role for restriction enzymes in DNA repair has not been demonstrated: but, as discussed at length above (Introduction 7.), the dam methylase of E.coli clearly plays an important part in the mismatch repair system. It is not inconceivable that cognate restriction/methylation systems in other species act in concert to mediate DNA repair. Evidence is presented in this work which suggests just such a role for the MmeI endonuclease of M.methylotrophus. But much more data must be obtained before this final possibility can be seriously considered.

9. Methylophilus methylotrophus: in industrial microorganism.

The obligate methylotroph Methylophilus methylotrophus (ASI), a gram-negative rod, was isolated from activated sludge by Imperial Chemical

Industries (ICI) workers (D. Byrom, Pers.comm.) for use, on an industrial scale, as a single-cell protein (SCP) organism protein-rich animal feedstuffs. Unfortunately, pure research on the organism has been conducted in widely separated centres, and few results have been published, so much of the available information is anecdotal.

Genetically, M.methylotrophus is a relatively poorly characterized organism. Despite intensive work, no transformation systems or phage have been described, so conventional genetic analysis is impossible. Despite these drawbacks, Windass et al. (1980) were able to improve the carbon conversion efficiency of the organism by replacing its endogenous ammonia assimilation pathway, which uses a two-stage process dependent on glutamine synthesis (GS) and glutamate synthase (GOGAT), and that of E.coli, which uses the energetically more efficient glutamate dehydrogenase (GDH) process. The E.coli gdh gene was first subcloned into a broad host range cloning vector pTB70 (based on R300B). 550 temperature-sensitive mutants of M.methylotrophus were isolated after treatment with the mutagen N-methyl-N'-nitro-N-nitrosoguanidine, and four were found to grow at the non-permissive temperature (37^{o}) after the introduction, by conjungation, of pTB70:gdh. Extracts from these four mutants lack detactable GOGAT activity at $37^{o}C$, and thus the plasmid-borne ammonia assimilation pathway (GDH) must be the one being used by the plasmid-containing mutants. One plasmid-containing mutant was found to give 4-7% higher carbon conversion than the wild-type bacterium. This increase in efficiency is commercially significant, especially since the mutant strain retains the improvement in industrial scale fermentors.

Modern genetic techniques, including cloning, offer many ways to explore the genetics of M.methylotrophus (Brammar, 1981). Most of our present knowledge of this organism's genetics, however, has come from more classical complementation studies in another species, Pseudomonas aeruginosa PAO. Moore et al. (1983), used the broad host range, chromosome-mobilizing plasmid R68.45 to generate R primes containing M.methylotrophus DNA. These were screened for their ability to complement auxotrophies in a large number of well-characterized P.aeruginosa strains to which they had been transferred by conjugation. The most attractive feature of this system was that mutants of M.methylotrophus were not required, because only functional genes of the methylotroph were being assayed. Of course, a potential danger was that M.methylotrophus genes may not have been expressed in the other species, but since complementation of about ten amino-acid auxotrophies and various mutations in purine/pyrimidine metabolism was actually observed, this problem seemed negligible. R primes containing large (>100kb) inserts of M.methylotrophus DNA which complemented a range of P.aeruginosa mutations simultaneously were isolated, thus allowing a redimentary genetic map to be constructed. Clearly, this approach is limited by the availability of suitable P.aeruginosa mutants and their ability to be complemented by M.methylotrophus genes. Until many more M.methylotrophus mutants are isolated and characterized, however, mobilization of R primes into P.aeruginosa remains the most rapid way of genetically analysing the methylotroph.

In order to use M.methylotrophus efficiently as a host organism for cloning (for example) it is essential to have a thorough knowledge of any restriction/modification (R/M) systems it may possess. Such knowledge of

E.coli K12 has led to the universal use of restrictionless (hsdR) mutants in the cloning of heterologous (e.g. eukaryotic) DNA in that species (see Introduction 3.). It was with this goal in mind, therefore, that this work was concentrated eventually into an investigation of the genetics rather than the biochemistry of the two R/M systems discovered in M.methylotrophus. One vindication of this approach came from the observation that P.aeruginosa strains containing conjugative plasmids were unable to transfer them directly to M.methylotrophus (D.Pioli, pers.comm.). Plasmids of interest had first to be transferred to E.coli before being introduced into the methylotroph. In the light of the results presented in this work, this can easily be explained: P.aeruginosa lacks a dam-type methylase which E.coli possesses, and thus is unable to protect its DNA from restriction (after conjugation) in the methylotroph recipient by MmeII. It is extremely fortunate that E.coli does possess a dam methylase, since otherwise routine plasmid transfer between E.coli and M.methylotrophus would have been all but impossible (cf. transfer from dam$^-$ E.coli strains to M.methylotrophus, Results 5.3.).

M.methylotrophus has significant potential as a "safe" host for cloning experiments, since, unlike E.coli, it will not survive in the gut, and is strictly dependent on methanol (or simple derivatives) as carbon source, so containment is much simpler. The lack of a transformation protocol, however, is the main stumbling block to its use, since the initial cloning steps must, at the moment, be carried out in E.coli and suitable vectors containing cloned segments of DNA subsequently transferred to M.methylotrophus by mobilization. Small cloning vectors able to replicate in

both species are being developed for this purpose by G. Sharpe in the ICI Joint Laboratory at Leicester. Clearly, intergeneric transfer of plasmids into M.methylotrophus is going to remain of importance until or unless a transformation protocol can be devised, so knowledge of the methylotroph R/M systems described in this work is of some practical value.

Results and Discussion: CHAPTER 1.

1.1. Computer method for determining recognition sequences.

There are essentially two distinct approaches to determining the recognition sequences of restriction enzymes. The first developed was the obvious one of end-labelling restricted DNA fragments, and subjecting them to sequencing techniques: originally, this was done by two-dimensional thin-layer chromatography of oligonucleotides (Jay *et al.*, 1974); later on, the more powerful sequencing methods of Maxam and Gilbert (1977) and Sanger *et al.* (1977) could be applied. (See also Brown and Smith, 1980 for the use of sequencing ladders in the determination of points of cleavage.) The second method used took advantage of the fact that several DNA replicons (e.g., pBR322, ΦX174 and SV40: Sutcliffe, 1978; Sanger *et al.*, 1978; Fiers *et al.*, 1978) were completely sequenced: if the sites cleaved by a novel endonuclease in such a DNA were accurately mapped, the recognition sequence could be deduced by inspection and comparison of the known sequences around the mapped cut sites. Any unique, common sequence characteristic of a restriction enzyme recognition sequence would almost certainly be the one recognized by the new enzyme. The identification could then be confirmed by comparing the restriction patterns obtained using other sequenced DNAs with those predicted from the proposed recognition sequence. The method is limited by the definition of which sequences are characteristic of restriction enzymes and which are not - a problem which has not been completely solved, and will be considered in greater detail below. A complex correlation process of this nature would be difficult to achieve manually, but is eminently suited to

modern, high-speed computers. Once a suitable program has been written, it can be used to analyse any sequenced DNA without further modification.

Fuchs *et al.* (1978) used this method to correctly deduce the recognition sequence of *Ava*II, GG(A/T)CC. They also published computer-generated tables of locations of palindromic sequences in ΦX174 which could be used to identify recognition sequences of other enzymes cleaving the phage DNA. Unfortunately, the tables were not comprehensive, and would fail to identify palindromic recognition sequences with certain degenerate bases and were not applicable to non-palindromic sequences at all. An earlier attempt by Murray *et al.* (1976), using the end-labelling technique, proposed the incorrect sequence G(A/T)CG(A/T)C for *Ava*II. As more and more restriction enzymes were discovered with novel types of recognition sequence, it became clear that the heuristic method required considerable enhancement to make it generally applicable. A computer program (RECOG) has been developed here which attempts to overcome the drawbacks of the Fuchs program by taking into account many more variations in the formats of recognition sequences.

The problem can be divided into two unequal parts corresponding to enzymes which recognize (a) palindromic and (b) non-palindromic sequences. It is easy to see that part (a) is a far simpler problem to solve than part (b): for example, (ignoring degenerate bases for a moment) there are $4^2 = 16$ possible tetrameric palindromes, but $4^4 - 4^2 = 240$ possible non-palindromic tetramers. The disparity between these classes increases exponentially with the number of bases; so there are, in general, many more cases to consider in part (b) than part (a). The difficulties are far more severe (in both cases) when degenerate bases are allowed, but, unfortunately, of the 89 known (Type II) specificities (Roberts, 1983), 33 include at least one degenerate base, so

this feature cannot be omitted.

The problem of dealing with degenerate base positions, indeed, is the major one in the heuristic approach: it is clear that the diversity of degenerate recognition sequence types is large, and new types are constantly appearing. The program must therefore be provided with sufficient sequence "templates" to rediscover all known specificities, and also be flexible enough to discover a substantial proportion of new specificities, many of which will have degeneracy patterns not previously encountered. A template, in this sense, is defined to be a string of characters which represents a class of possible recognition sequences: for example, pppqqq is the template for all possible (i.e., 64) palindromic hexamers with no degenerate positions (AAATTT - TTTAAA), and ppNqq is the template for all palindromic pentamers where the the central base can be A, C, G or T, such as GCNGC. The program therefore must have the templates 8ppqq8* and p8pq8q, because two known restriction enzymes recognize those sequence types, i.e., *Hae*I (8GGCC8) and *Hgi*AII (G8GC8C). No known restriction enzyme, however, recognizes the type pp88qq, but that template ought to be included in order to anticipate the discovery of an appropriate enzyme. The problem lies in deciding how far to carry the anticipation process. Obviously, the number of different specificities of restriction enzymes in Nature is finite, but there is no way of knowing how many templates are needed to be sure of identifying the recognition sequences of every new enzyme discovered. A practical difficulty also exists in that every new template added to a program based on this method increases the computer time needed to search for appropriate sequences, and experience has shown that this can be a serious drawback. inevitably, a compromise had to be sought, and the one adopted in this work was to define a set of templates which was (a) sufficient to rediscover the known specificities, and (b) general

* See p1 for nucleotide abbreviations 1, 2, 8, 9, X, Y, Pu & Py.

enough, it was hoped, to anticipate the majority of specificities yet to be discovered. The set finally arrived at is shown in Tables 1-1 and 1-2: the first table displaying palindromic templates, and the second one non-palindromic ones. The templates are displayed with the names of type enzymes fitting them, if known, demonstrating the diversity of degeneracy patterns already discovered, and justifying the inclusion of extra templates to anticipate novel specificities. For example, it was felt that the existence of AccI, recognizing the sequence GT19AC, was justification enough to include five other templates with A/C, G/T degeneracies, even though, at the time of writing and using the program, no other known enzyme recognition sequence had such unusual degeneracies. It is gratifying to note that since then (Roberts, 1983) the recognition sequence of *Nsp*BII has been found to be C1GC9G, conforming to the template p1pq9q. The number of templates with no type enzyme is much larger for non-palindromic sequences (see Table 1-2): this is because only a few such enzymes have yet been characterized. It was felt, however, that all the templates devised here could be reasonably justified on a theoretical basis, e.g., the fact that one enzyme (*Tth*111II) recognizes a non-palindromic hexamer with a purine/pyrimidine degeneracy suggested that enzymes recognizing pentamers with a similar degeneracy might also exist. It may well be that many known restriction enzymes whose recognition sequences are not yet determined recognize non-palindromic sequences whose identities, so far, have been obscured by their unforeseeable pattern of degeneracies.

The vast majority of known specificities are six or fewer base pairs long: a small number, however, recognize sequences longer than six base pairs. It is probably significant that none of these sequences have more than six specified bases: no known restriction enzyme recognizes a sequence like AACGCGTT or

1. Non-degenerate tetrameric palindromes, AATT - TTAA:

ppqq
*Alu*I

2. Pentameric palindromes with central degeneracy:

ppNqq	pp8qq	pp2qq
*Hin*fI	*Eco*RII	*Cau*II

3. Non-degenerate hexameric palindromes, AAATTT - TTTAAA:

pppqqq
*Eco*RI

4. Degenerate hexameric palindromes:

ppXYqq	ppYXqq	pp88qq	pp22qq	pp19qq	pp91qq
*Afl*III	*Hin*cII	———	———	*Acc*I	———
pXpqYq	pYpqXq	p8pq8q	p2pq2q	p1pq9q	p9pq1q
*Acy*I	*Ava*I	*Hgi*AI	———	*Nsp*BII	———
XppqqY	YppqqX	8ppqq8	2ppqq2	1ppqq9	9ppqq1
*Hae*II	*Cfr*I	*Hae*I	———	———	———

5. Non-degenerate tetrameric and hexameric palindromes with one to six unspecified central bases: (ppNqq covered in 2. above)

ppNNqq	ppNNNqq	ppNNNNqq	ppNNNNNqq	ppNNNNNNqq
———	———	———	———	———
pppNqqq	pppNNqqq	pppNNNqqq	pppNNNNqqq	pppNNNNNqqq
*Sau*I	———	*Tth*111I	*Xmn*I	*Bgl*I

pppNNNNNNqqq
*Hgl*EII

6. Degenerate hexameric palindromes with one to six unspecified central bases:

(This is an obvious extension of section 4., comprising templates ranging from ppXNYqq to 9ppNNNNNNqq1. Accordingly, there are 6 x 18 = 108 templates altogether. No known enzymes fit any of these templates.)

TABLE 1-1: Palindromic templates used by RECOG.

1. Non-degenerate tetrameric sequences, AAAA - TTTT:

   pppp
   *Mnl*I

2. Non-degenerate pentameric sequences, AAAAA - TTTTT:

   ppppp
   *Mbo*II

3. Degenerate pentameric sequences:

   Xpppp pXppp ppXpp Ypppp pYppp ppYpp

4. Non-degenerate hexameric sequences, AAAAAA - TTTTTT:

   pppppp

5. Degenerate hexameric sequences:

   Xppppp pXpppp ppXppp Yppppp pYpppp ppYppp
                           *Gdi*II   *Tth*111II

TABLE 1-2: Non-palindromic templates used by program RECOG.

CCGATGCATCGG. They are longer than six base pairs only because their sequences are "padded out" with central, unspecified bases; for example *Hgi*EII, recognizing the palindrome ACCNNNNNNGGT. The proposed natural function of restriction enzymes suggests a possible explanation for this upper limit on specificity. if a restriction enzyme is to provide an effective defence against invading phage or plasmid DNA, then there must be at least one recognition sequence for the enzyme on the DNA. The best way to maximize the probability that the invading DNA should be a substrate for the enzyme is for the recognition sequence to be short enough to statistically ensure its occurrence on almost every piece of DNA. Clearly, the shorter the recognition sequence, the more often it will occur by chance in a given length of DNA. A four base pair palindromic sequence will occur about once every 256 ($=4^4$) base pairs, and a similar six base pair sequence will occur about once every 4096 ($=4^6$) base pairs, on a "random" piece of DNA. (Obviously, in reality, these figures vary with the base composition of the recognition sequence and that of the invading DNA, but their order of magnitude is correct.) Since many small DNA phages have genomes of the order of 5kb long, it is obviously disadvantageous for a cell to produce a restriction enzyme recognizing, say, 8bp which would occur only about once every 65536 ($=4^8$) base pairs, and would therefore be absent from most invading DNA molecules. Taking this argument to its logical conclusion, it may be objected that even a 4bp recognition sequence might, by chance, not occur on an invading DNA, and that restriction enzymes with even shorter recognition sequences would be more effective. A bacterium producing a restriction enzyme, however, has to pay a price for the protection thus obtained: it has to methylate all the recognition sequences on its own genome to avoid suicide. The fact that no restriction enzyme recognizing a sequence less than 4bp long has yet been isolated suggests that the energy cost of methylating a much more frequent 2 or 3bp

sequence is so high that the selective advantage afforded by such a hypothetical enzyme is outweighed by the metabolic disadvantage. It is perhaps because of these conflicting evolutionary pressures that, in spite of the manifest diversity of known recognition sequences, none of them possess more than six or less than four (fully-defined) bases. (Some Type I restriction enzymes, e.g. *Eco*B and *Eco*K, recognize sequences containing more than six defined bases, in contradiction of the above argument: however, this class of enzyme is known only in *E. coli*, and caution must therefore be exercised in drawing general conclusions about their function and evolution.)

Palindromes with central tracts of one or more unspecified bases are of special interest. Only six examples of this class have been found, i.e., *Eca*I (GGTNACC), *Sau*I (CCTNAGG), *Tth*111I (GACNNNGTC), *Xmn*I (GAANNNNTTC), *Bgl*I (GCCNNNNNGGC), *Bst*XI (GGANNNNNNTCC) and *Hgl*EII (ACCNNNNNNGGT). Strictly speaking, however, five enzymes which recognize pentameric palindromes containing a single, unspecified base should also be included in this class, i.e., *Asu*I (GGNCC), *Dde*I (CTNAG), *Fnu*4HI (GCNGC), *Hin*fI (GANTC) and *Scr*FI (CCNGG). It is noteworthy that none of the specificities listed above contain degeneracies in the defined bases: but since so few examples are known, it would be rash to conclude that no degenerate specificities are yet to be found. RECOG therefore takes into account this possibility, and generates 18 degenerate templates corresponding to the hexameric series (Table 1-1, 6.) for each number of central bases from one to six, making 108 templates in all. Their inclusion can be justified theoretically by considering these "interrupted hexamers" as being "related" (in some sense) to uninterrupted hexamers. For instance, *Eca*I, recognizing GGTNACC, is related, in this sense, to *Kpn*I which recognizes GGTACC. It is a remarkable fact that all but two of the enzymes cited above can be similarly

matched with an enzyme whose recognition sequence lacks non-specified bases, the remaining pairs being: *Sau*I, *Avr*II; *Tth*111I, *Aat*II; *Xmn*I, *Eco*RI; *Bgl*I, *Nae*I; *Bst*XI, *Bam*HI; *Asu*I, *Hae*III; *Fnu*4HI, *Hha*I; *Hin*fI, *Sau*3A and *Scr*FI, *Hpa*II. The two enzymes lacking a match so far are *Hgi*EII and *Dde*I. (See Roberts, 1983 for sequences.)

The templates described and justified in some detail above form the backbone of RECOG, which can be expanded readily to cope with new templates that may become apparent. Operation of the program is divided into four sections (passes), and various combinations of these are appropriate to searches for palindromic or non-palindromic sequences. A listing of the program is given in Appendix 1-1. Summaries of RECOG operating procedures used to determine the recognition sequences of *Mme*I and *Mme*II are given in Results 2.3. and 3.2.

1.2. Relationship between molecular weight and gel mobility.

It is possible to deduce, in various ways, the size of DNA fragments from their mobilities on agarose or polyacrylamide gels. Still the most commonly used method is to plot mobilities of standard DNA fragments against log(molecular weight) and draw a straight line through the points. Molecular weights of unknowns can then be read from the graph. Unfortunately, on a plot of this kind for agarose gels, linearity of the gel is restricted to fragments less than ca. 6kb; above this size, the graph curves appreciably upwards. Duggleby *et al.* (1981) attempted to improve on this method by fitting a parabola rather than a straight line to standard mobility data, using a simple computer program.

An alternative to this log-linear plot, first described by Southern (1979), uses the idea (based on theoretical considerations) that mobility and size of DNA fragments on agarose gels are related reciprocally. That is, for any particular fragment in a gel track, the product of its size and mobility is roughly constant. A plot of size versus 1/mobility does indeed give a straight line fit, even for fragments greater than 6kb long. Schaffer and Sederoff (1981) developed this idea, and found empirically that a better fit was obtained if the size (S)/ mobility (M) relationship was modified slightly to:

$$(S + a)(M + b) = k,$$

where a, b and k are constant for any particular gel. They, too, wrote a program to fit this equation to mobility data.

Since an important part of the work described in this thesis relied on the ability to accurately estimate DNA fragment sizes from agarose gel mobilities, it was decided at an early stage to compare these two methods. Figure 1-1 summarizes this comparison. The "+" signs indicate size/mobility points from an actual run of a λ/*Hin*dIII/*Eco*RI standard. Molecular weights of λ fragments were deduced from the preliminary (Sanger *et al.*) sequence of the phage DNA, obtained *via* F. Blattner. The narrow line shows the log-linear fit, and the broad line the reciprocal fit. Obviously, the reciprocal fit most nearly intersects with each point, even in the high molecular weight range. At one point, (ca. 17mm mobility), the discrepancy between the estimates of molecular weight approaches 1kb, showing the inadequacy of the log-linear fit in this size range. The reciprocal fit, therefore, and a simple program into which it was incorporated (GELFIT) were used for all further agarose gel

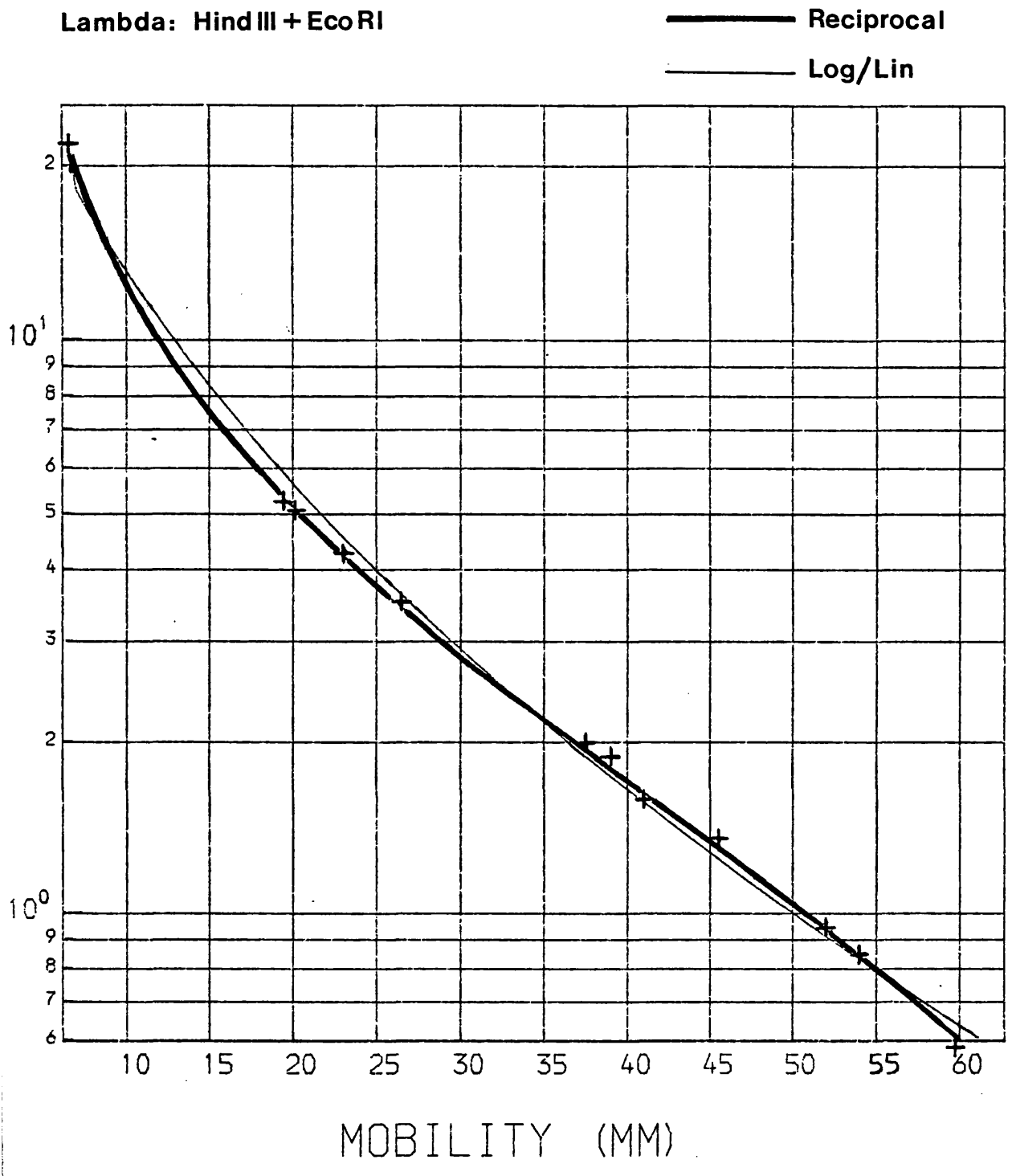

FIGURE 1-1: Comparison of the reciprocal and log-linear fits to agarose gel mobilities of a DNA molecular weight standard. This example is based on a λ HindIII/EcoRI double digest.

analyses. Experience showed that errors of, on average, 1 - 2% in fragment size estiamtes could be expected using this method. GELFIT also works for polyacrylamide gels, if fragments greater than 1kb are not included.

Results and Discussion: CHAPTER 2.

2.1. Detection of *Mme*I activity *in vitro*.

Many variations of the basic purification procedure (Greene *et al.*, 1978) were tried until one was found which resulted in detection of endonuclease activity. One source of the failure of early attempts may have been the use of frozen (-70°C) cell paste as a source of enzyme. Such paste took several hours to thaw at 4°C, with stirring; the resulting sludge was extremely viscous; and it was obvious that most of the cells had lysed in the freeze-thawing process. It may be that under these conditions endonucleases are rapidly degraded. Success was only achieved when fresh cells were used, and lysed rapidly using a French press (see Methods 3.3.2.). Eventually, however, endonuclease activity was detected in the eluate from a phosphocellulose column (at ca. 0.2 - 0.25M NaCl), and a sample was dialysed against storage buffer and kept at -20°C. The rest was applied to a hydroxylapatite column, but no activity was detected in the eluate. The activity eluted from the first column, therefore, was used for all subsequent analyses. Subsequent attempts to purify *Mme*I were rarely successful, even when the mechanics of purification apparently proceeded well. It may well be that amounts of *Mme*I in the cell are so low that purification on the scale attempted here was usually not sufficient to detect activity. It was felt that the large investment of time and materials needed to devise a foolproof purification protocol for *Mme*I was not scientifically justified, especially since the genetic investigation of the restriction system of *M. methylotrophus* was turning out to be so fruitful. Nevertheless, enough material was obtained from the successful columns to characterize the enzyme to the point of deducing its recognition sequence. Simple assays of *Mme*I activity in a variety of buffers showed that it was

inhibited by high (>50mM) NaCl concentration, required Mg^{++} ions, but did not require ATP or S-adenosyl methionine. The latter three observations confirm that the enzyme is Type II and not Type I or III. 666 buffer proved to be the one in which *Mme*I activity was maximal, and was used subsequently in all digests. One property of *Mme*I observed from an early stage was that a large number of partial digest products were always seen. Unlike the partials produced by other Type II enzymes however, these could not be distinguished from complete digest products by inspection of band intensity; (i.e., partials are normally considerably less fluorescent than complete digest bands of greater mobility) since the difference of intensities of the bands was not sufficient to allow such discrimination. (See Results 2.7. for discussion of this point.)

2.2. Mapping of *Mme*I sites in pBR322.

In order to use the computer methods described earlier to identify restriction enzyme recognition sequences, it was necessary to map the cleavage sites formed by *Mme*I in a sequenced DNA, in this case pBR322. Unfortunately, it was clear (as described above) that the enzyme never produced a complete digest - bands corresponding to partial digestion products were always found as well as limit digestion products. In the early stages, this complicated mapping attempts. The first site to be accurately mapped was one at coordinate 2860, later to be called M4. Initially, this was defined by a serial digestion experiment. In a *Hin*dIII/*Mme*I double digest of pBR322, two prominent bands of ca. 1520bp (I) and 1690bp (II) were seen, in relative isolation from several other bands close together (digest not shown). 10μg of plasmid DNA was digested with these two enzymes, and run on a preparative

1% agarose gel. Bands I and II were excised, and ca. 1μg DNA purified from each. The fragments were digested with *Hae*III, and run on a 5% polyacrylamide gel, next to a *Hae*III digest of total plasmid DNA (Figure 2-1). In the digest of intact pBR322, 16 bands (one a doublet) representing fragments > 50bp can be seen, which correspond well to the sizes of *Hae*III fragments predicted from the pBR322 sequence. The accompanying digests of I and II reveal only the *Hae*III fragments A, D, F and N. Reference to the restriction map of pBR322 (Figure 2-2) shows that these fragments are contiguous, in the order D-N-F-A (2953 - 4345). Clearly, the *Hin*dIII sites at one end of each of fragments I and II define the A end of this DNA tract. The simplest interpretation of the gel, then, is that the *Mme*I sites at the other ends of I and II lie within the *Hae*III E fragment.

But here was a paradox: two obviously different-sized DNA fragments (I and II), when purified and re-digested with *Hae*III, gave apparently identical restriction patterns. Moreover, no additional bands, which might be expected by the presence of at least one *Mme*I site within *Hae*III E, seemed to be produced. Because of this ambiguity, the only fact which could be deduced with near certainty from this experiment was that there was one (at least) *Mme*I site at, or anticlockwise of, the end of *Hae*III D (2953), probably within *Hae*III E. The presence of exonuclease in the fragment preparations was blamed for the ambiguities in this experiment. (It turned out, however, that there were indeed two sites in *Hae*III E, and that the fragments thus generated co-migrated, respectively for I and II, with the F and N *Hae*III fragments of pBR322. Close comparison of tracks 1 and 2 in Figure 2-1 reveals that the suspected doublets are brighter than their single counterparts.)

Inspection of the pBR322 sequence (Sutcliffe, 1978) revealed that the

FIGURE 2-1: 5% polyacrylamide gel analysis of a HaeIII restriction digest of purified pBR322 MmeI/HindIII fragments. 1, pBR322/HaeIII; 2, 1690bp MmeI-HindIII fragment/HaeIII; 3, 1520bp MmeI-HindIII fragment/HaeIII. Fragments common to all three tracks are A, D, F and N. (See Fig. 2-2.)

*Mme*I site thus positively identified was within the largest *Taq*I fragment (2576 - 4019) and also within the largest *Sau*3A fragment (1668 - 3041): observations which were confirmed in *Mme*I/*Taq*I and *Mme*I/*Sau*3A double digests. In neither case was the large fragment completely digested by *Mme*I although there was a clear diminution of intensity of the band. The size of the extra band obtained in a *Mme*I/*Taq*I double digest was 1155bp: taken with the previous result, this positions site M4 at about coordinate 2860 (= 4019 - 1155).

In the case of the *Sau*3A double digest, two additional bands of sizes 1208bp and 1029bp are formed by *Mme*I activity, indicating an additional *Mme*I site in the region 1668 - 3041 (M3). (This was the other site hinted at by the *Hae*III experiment described above.) Clearly, the 1208bp band is due to M4 (2860 - 1668), so M3 is either at coordinate 2700 (= 1668 + 1029) or at 2012 (= 3041 - 1029). The position of this site and the two remaining ones were determined by consideration of a *Mme*I/*Pst*I double digest of pBR322. *Pst*I cuts pBR322 uniquely at coordinate 3609. If M3 is located at 2012, a fragment of length 1597 (= 3609 - 2012) would be expected; conversely, the other possible location for M3 would generate a fragment of 909bp (= 3609 - 2700). Inspection of an actual *Pst*I/*Mme*I double digest of pBR322 reveals no fragment in the size range 1100 - 2300bp: there is, however, a fragment of approximately 920bp. This positions M3 at 2700. Of the four fragments less than 1100bp long in the double digest (ca. 1050, 970, 920 and 750bp), two are accounted for by M3 and M4 - i.e., M3 - 3609 = ca. 920bp, M4 - 3609 = ca. 750bp. The other two fragments must be due to the presence of at least two additional *Mme*I sites (M1 and M2) in pBR322. Obviously, each site may be either clockwise or anticlockwise of the *Pst*I site: but only the interpretation that M1 and M2 are both clockwise of the *Pst*I site is consistent

with the fragment length data derived from a *Mme*I digest of pBR322. Given this fact, the double digest data position M1 at ca. 217 (= 3609 + 970 - 4362) and M2 at ca. 297 (= 3609 + 1050 - 4362). Using these coordinates, it is easy to deduce that the four complete *Mme*I digest products of pBR322 would have lengths of ca. 2403, 1719, 160 and 80bp. The two smallest fragments are not seen on 1% agarose gels, either because they are produced in insignificant quantities, or because the bands are too diffuse. Bands closely corresponding to the predicted sizes of the two largest fragments are seen, however, and all other bands can be accurately interpreted as being partials; there is no need to invoke the existence of a fifth *Mme*I site.

The derivation of a more accurate map of *Mme*I sites in pBR322 was achieved by use of the program described by Schroeder and Blattner (1978). The data from *Mme*I, *Mme*I/*Sau*3A, *Mme*I/*Taq*I, *Mme*I/*Pst*I digests were used to obtain refined *Mme*I coordinates of 212, 296, 2697 and 2874. See Appendices 2-1 and 2-2 for the input to and output from the Schroeder/Blattner program which produced these coordinates. Compare with Figure 2-2 for identification of relevant restriction sites. It should be noted that the use of this program, in conjunction with accurate fragment length determination using GELFIT, can give restriction maps with site positions in error by, on average, only 1 - 2%. This combination of programs has been extensively used in Leicester by D. W. Burt (pers. comm.) for the accurate mapping of recombinant DNA carried in the λ vector L47 (Loenen and Brammar, 1980).

2.3. Determination of *Mme*I recognition sequence.

The gel analyses described above accurately define four *Mme*I sites in pBR322. The computer approach was used to identify the recognition sequence

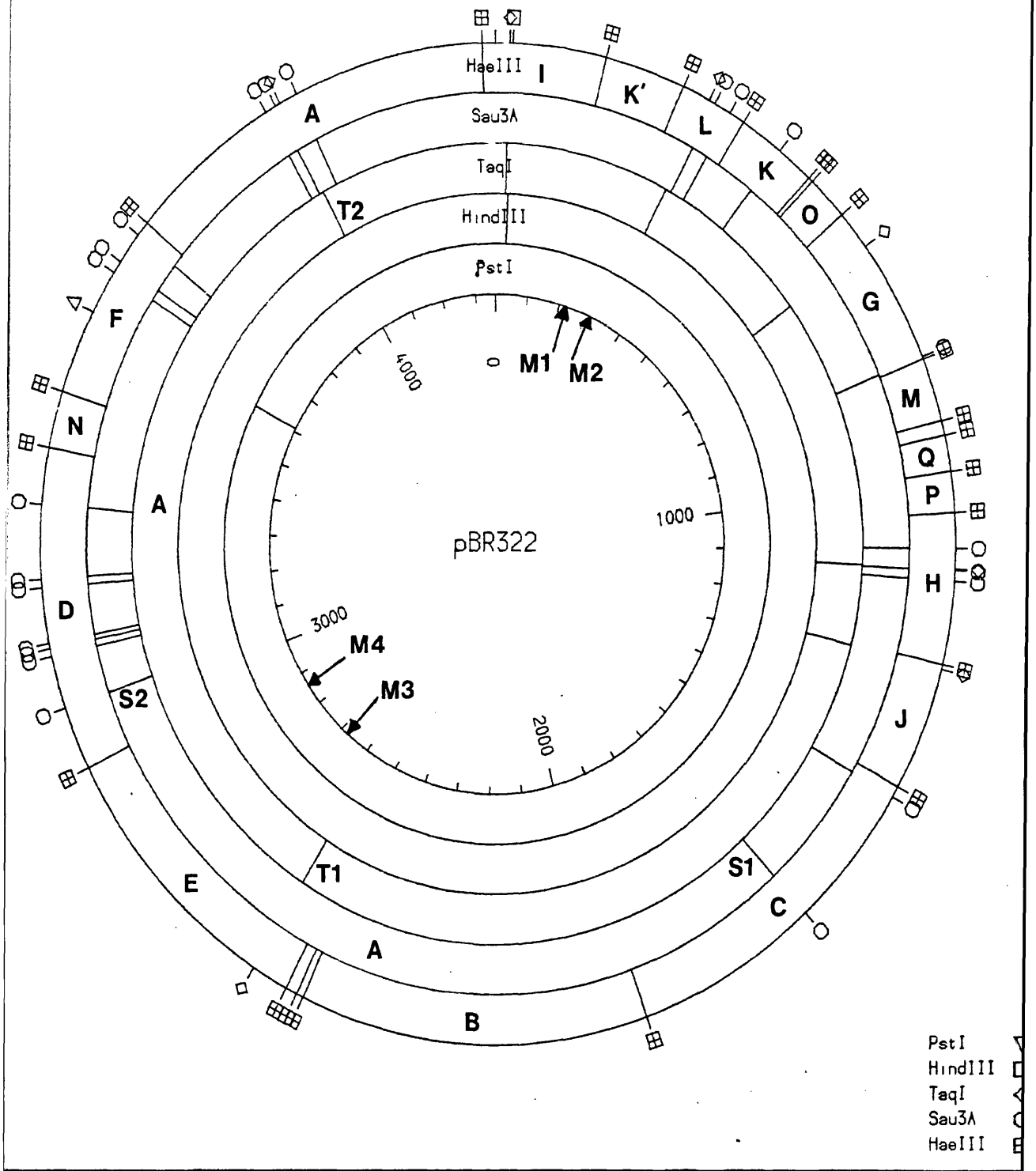

FIGURE 2-2: Restriction map of plasmid pBR322 (4362bp). Enzyme sites and fragments used in mapping the four MmeI sites (M1 - M4) are shown.

of the enzyme. Naturally, the first idea to be tested was that *Mme*I recognized a palindromic sequence, (tetramer, pentamer or hexamer), as do most other known restriction enzymes. The section of the program RECOG which deals with palindromic sequences allows the user to select a size range (window) in which a restriction fragment lies, and predicts, for each sequence, how many fragments would lie within the window if that sequence was cleaved. From the refined gel data, it was deduced that the four complete digestion products of *Mme*I activity on pBR322 had sizes of approximately 2401, 1701, 176 and 84 base pairs. Allowing an error margin of 2% each way, a window of 2350bp - 2450bp was defined. Using RECOG, only five palindromic sequences were identified which could produce a single fragment in that size range. These were 1ATAT9, 1GGCC9, C2AT2G, TC88GA and XCATGY which occur respectively 3, 4, 3, 9 and 4 times in pBR322. Clearly, then, the first and third sequences can be rejected, simply because they occur less than four times. The other three can also be rejected because their locations are very far from the sites mapped for *Mme*I: 1GGCC9 is found at coordinates 594, 828, 938 and 2516; TC88GA is at 489, 2841, 2939, 3037, 3120, 3194, 4202, 4307 and 4356; and XCATGY occurs at coordinates 561, 1815, 2109 and 2474. These data prove that *Mme*I recognizes none of the palindromic tetramers, pentamers or hexamers covered by the set of templates used by RECOG.

Before concluding, however, that *Mme*I did not recognize a palindromic sequence, it was necessary to eliminate hexameric palindromes interrupted by a central tract of one to six unspecified bases, such as the one recognized by *Bgl*I, GCCNNNNNGGC. The program RECOG was run for each of these six additional classes of palindromes, using the 2350 - 2450 window described above. Sequences which could produce a single fragment of this size are shown in Table 2-1. (N.B., only those sequences which occur four or more

4	2440	1GGCC9 (1578)	4	2374	A9TNNNNA1T (1005)
4	2449	XCATGY (1254)	4	2407	9GGNNNNCC1 (1437)
9	2352	TC88GA (1008)	5	2381	TA9NNNN1TA (855)
			6	2352	C2CNNNNG2G (1005)
4	2351	1GTNAC9 (1077)	7	2424	GA9NNNN1TC (561)
4	2356	8CTNAG8 (1115)	9	2363	1TGNNNNCA9 (818)
4	2410	8TGNCA8 (1242)	10	2397	2CGNNNNCG2 (711)
5	2378	9ATNAT1 (803)			
7	2447	AGXNYCT (767)	5	2367	AYCNNNNNGXT (1009)
			5	2367	GXANNNNNTYC (614)
4	2352	2TCNNGA2 (1008)	6	2375	T8ANNNNNT8A (900)
4	2401	AT9NN1AT (1817)	9	2384	TA8NNNNN8TA (463)
5	2359	CG9NN1CG (1404)	10	2381	2CGNNNNNCG2 (657)
5	2385	C1GNNC9G (883)			
5	2442	XAGNNCTY (908)	4	2350	XGGNNNNNNCCY (1339)
7	2390	2ACNNGT2 (696)	4	2352	GA9NNNNNN1TC (1283)
8	2432	2GCNNGC2 (866)	4	2373	1ACNNNNNNGT9 (1437)
			4	2384	AG1NNNNNN9CT (962)
4	2356	CG8NNN8CG (850)	4	2444	AC2NNNNNN2GT (1082)
4	2375	CT9NNN1AG (1044)	5	2368	2CCNNNNNNGG2 (882)
5	2370	TT2NNN2AA (1125)	7	2357	TYANNNNNNTXA (1056)
5	2383	1TTNNNAA9 (1039)	7	2359	9CANNNNNNTG1 (1351)
5	2418	TC9NNN1GA (914)	8	2439	2GCNNNNNNGC2 (682)
5	2435	C2CNNNG2G (1470)	9	2438	GC2NNNNNN2GC (760)
6	2390	XTTNNNAAY (786)			

TABLE 2-1: Palindromic sequences which are candidates for the recognition sequence of *Mme*I. Each sequence is preceded by two numbers: (i) the number of occurrences in pBR322, (ii) the size of the largest gap between adjacent sequences (which would correspond to the largest fragment seen on a gel). The number which follows (in brackets) is the size of the second largest gap between adjacent sequences. Thus, the sequence 1GGCC9 occurs 4 times, with the two largest gaps being 2440 and 1578bp long. (Modified from program RECOG output.)

times are shown, and the hexamers already discussed are included for completeness.) The number in brackets following each sequence is the predicted size of the second largest fragment which it could produce. inspection of these figures reveals that none of the sequences could yield a fragment within 2% of the size of the second largest fragment actually produced by *Mme*I, i.e., 1667 - 1735. Having exhausted all the palindromic sequence types tested by RECOG, it could be concluded either that *Mme*I recognized a palindrome of unforeseen type, or that it recognized a non-palindromic sequence.

Recourse was therefore made to the section of RECOG which dealt with non-palindromic sequences. A print-out of the successful run of the program is given in Appendix 1-2. The operation of the program is in two stages. Pass 1 operates on the following data: (a) DNA sequence (pBR322), (b) length of non-palindrome to be considered (= hexamer, 6), (c) number of subsequences to be generated (= number of sites, 4), (d) length of subsequence to search (100bp) and (e) coordinates of mapped sites (212, 296, 2697 and 2874). RECOG then extracts four 100bp subsequences from the total pBR322 sequence, centred on each cut site coordinate. Every hexamer in the subsequences is then parsed using the non-palindromic templates (Table 1-2) to generate a file of hexamers which are potential recognition sequences. For example, the first hexamer of subsequence 1, CGGTAC, is output as eight hexamers: CGGTAC, YGGTAC, CXGTAC, CGXTAC and complementary sequences GTACCG, XTACCG, GYACCG, GTXCCG. Pass 1 is complete when all subsequences have been treated in this way.

The file of hexamers is then sorted and processed through Pass 4 of RECOG, which simply picks out sequences common to all four subsequences

and prints them out. Appendix 1-2 shows that the refined mapping data of *Mme*I sites in pBR322 were sufficiently accurate for RECOG to discover a unique sequence GTYGGA (= TCCXAC) common to all four short regions around the cut sites. Given the failure of RECOG to find any palindromic sequences with this property, the non-palindromic sequence found was a very likely candidate for the recognition sequence of *Mme*I. In support of this contention, a computer search of the whole of pBR322 for this sequence showed that it only occurred four times, in the same orientation, at coordinates 197 - 202 (M1), 284 - 289 (M2), 2664 - 2669 (M3) and 2848 - 2853 (M4), close to the mapped cut sites. Since the located sequence GTYGGA actually resides on the negative strand of pBR322, as it is normally drawn, it will be referred to henceforth by its complement TCCXAC. Most other restriction enzymes which recognize irregular sequences cut the DNA some distance away (see Introduction 5.): the difference between the mapped cut sites, and the actual location of TCCXAC sequences, suggested strongly that the enzyme *Mme*I cut the DNA a short distance to the 3′ of its putative recognition sequence.

2.4. Generation of partials - theory.

The peculiar property that *Mme*I possesses of generating partial as well as complete digestion products led to the writing of a small set of computer programs (GELSIM) which simulate the gel patterns that should be obtained if all possible digestion products were produced. The theory behind the programs is very simple, but differs slightly between linear and circular molecules. For example, if a linear molecule is partially digested at a single site, three distinct, linear fragments are produced: but if a circular molecule is so digested, only one such fragment will arise. (Undigested circular molecules are discounted because their migration properties on gels are less well

understood than those of linear molecules, and cannot at the moment be predicted.) By simple mathematical induction, two formulae can be written down which give the total number of fragments obtainable from a partial digest of either linear or circular molecules. They are as follows:

$$T = (N+1)(N+2)/2 \qquad \text{(for linear molecules);} \qquad (1)$$

$$T = N^2-N+1 \qquad \text{(for circular molecules)} \qquad (2)$$

In both cases, the number of possible products (T) increases with the square of the number of cleavage sites (N); indeed, as N increases, T tends to N^2 for circular molecules and to half that value for linear molecules: e.g., a hundred sites partially cleaved in a circular molecule generates 9901 products, but only 5151 products in a linear molecule.

It has proved instructive (and largely correct) to suppose that all possible cleavage products are generated in *Mme*I digests, and to use GELSIM to calculate fragment lengths of these products from the coordinates of the *Mme*I sites within the DNA molecule concerned. From these data, given the equation relating molecular weight to gel mobility based on the reciprocal relationship discussed above (Results 1.2.) it is possible to construct a graphical representation of *Mme*I digests of any sequenced DNA molecule. With appropriate linear scaling, the correspondence between the actual gel photographs and simulations is striking and unequivocal. GELSIM is flexible enough to allow the user to specify a site at which a circular molecule (such as a plasmid or phage genome) has been linearized, thus enabling the simulation of double digests. Simulation of digests by enzymes which do not generate partials is also possible, using GELSIM.

2.5. Confirmation of *Mme*I recognition sequence.

The most convincing way of proving that TCCPuAC is the recognition sequence for *Mme*I is by digesting other sequenced DNA molecules with the enzyme, and by showing that the digest pattern obtained is the same as that deduced from the sequence. GELSIM offers an objective way of comparing the theoretical and actual gel patterns. Since the sequence TCCPuAC had been deduced from analysis of digests of pBR322 DNA, the obvious first step was to simulate these gel patterns (Figure 2-3 (a)). By way of illustration, a *Mme*I digest of pBR322, with four TCCPuAC sequences at coordinates 197, 284, 2664 and 2848, should generate (by formula (1) above) 13 cleavage products, of sizes (bps): 4362, 4275, 4178, 2651, 2564, 2467, 2380, 1982, 1895, 1798, 1711, 184 and 87. Using GELSIM, and the appropriate size/mobility equation, these data were transformed into a graphical representation of the gel. (See Appendix 3-1 for example run of GELSIM package to simulate a pBR322/*Mme*I digest.) Similar calculations were performed for *Mme*I digests of pBR322 linearized with *Pst*I, or pre-digested with *Taq*I, and *Sau*3A. Figure 2-3 (a) shows these simulations displayed alongside the actual gel pictures, demonstrating the striking correspondence between them. The size/mobility equations for tracks 1 - 4 were derived, using GELFIT, from *Sau*3A band mobilities: the equation for 5 and 6 was derived from a λ/*Hin*dIII/*Eco*RI standard. The identity of TCCPuAC as the recognition sequence was confirmed beyond doubt by simulation of *Mme*I digests of other sequenced DNAs, namely ΦX174, SV40 and (with a small qualification described below) M13 and M13mp7. The first comparison was with a *Mme*I digest of *Pst*I-digested ΦX174 RF DNA. The DNA sequence of ΦX174 (Sanger *et al.*, 1978) contains five TCCPuAC sequences at coordinates -225, 2691, -3237, 5197 and 5376. (A coordinate preceded by a negative sign implies that the complementary

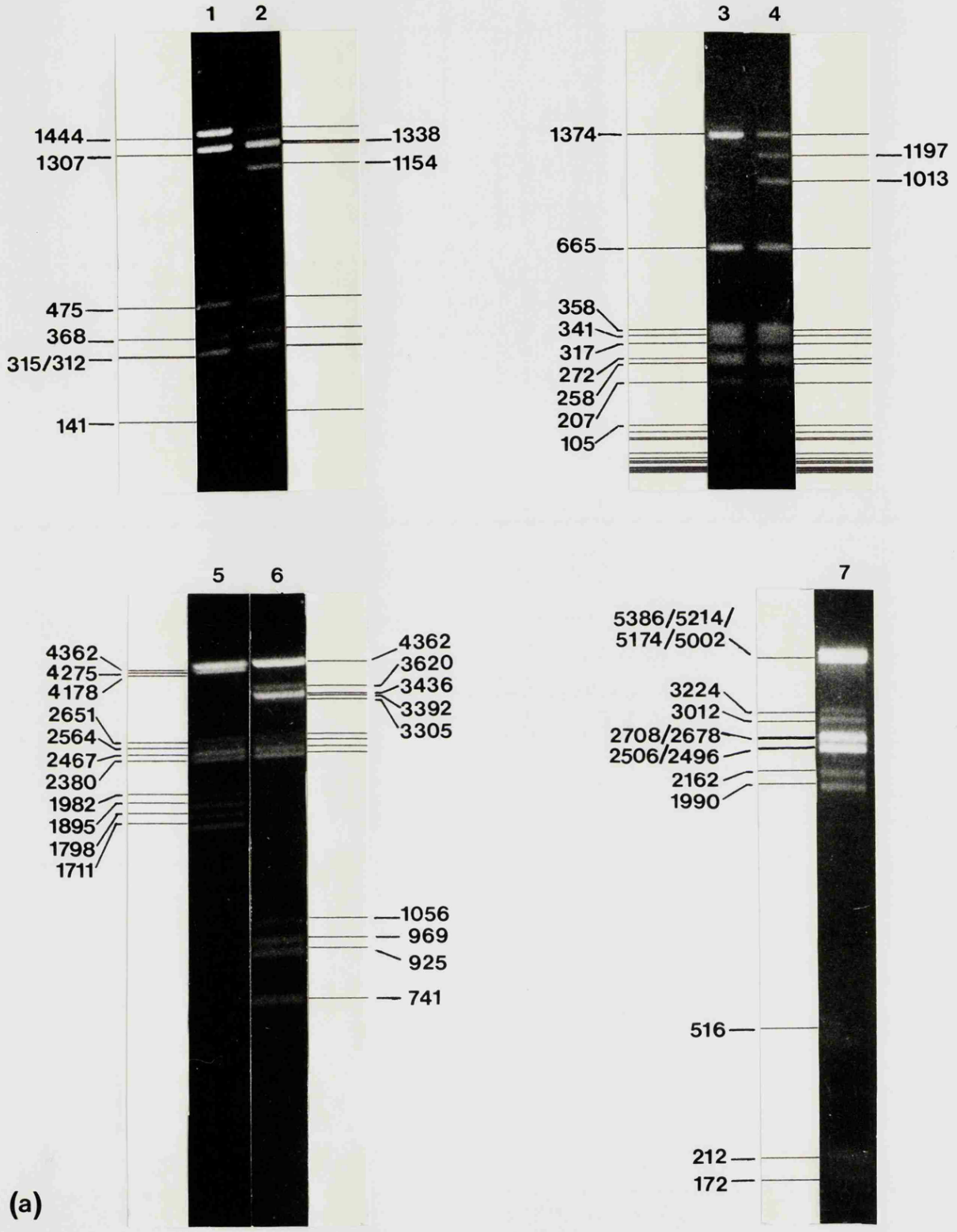

FIGURE 2-3:(a) Agarose gel analysis of pBR322 (1-6) and ØX174 (7) digests used to deduce and confirm the MmeI recognition sequence 5'-TCCPuAC-3'. Gel photographs are shown flanked by appropriately scaled computer simulations. Enzymes used were: 1, TaqI; 2, TaqI/MmeI; 3, Sau3A; 4, Sau3A/MmeI; 5, MmeI; 6, PstI/MmeI; 7, PstI/MmeI. (184bp and 87bp bands predicted in the simulations of 5 and 6 were cut off and were not visible on the gel.)

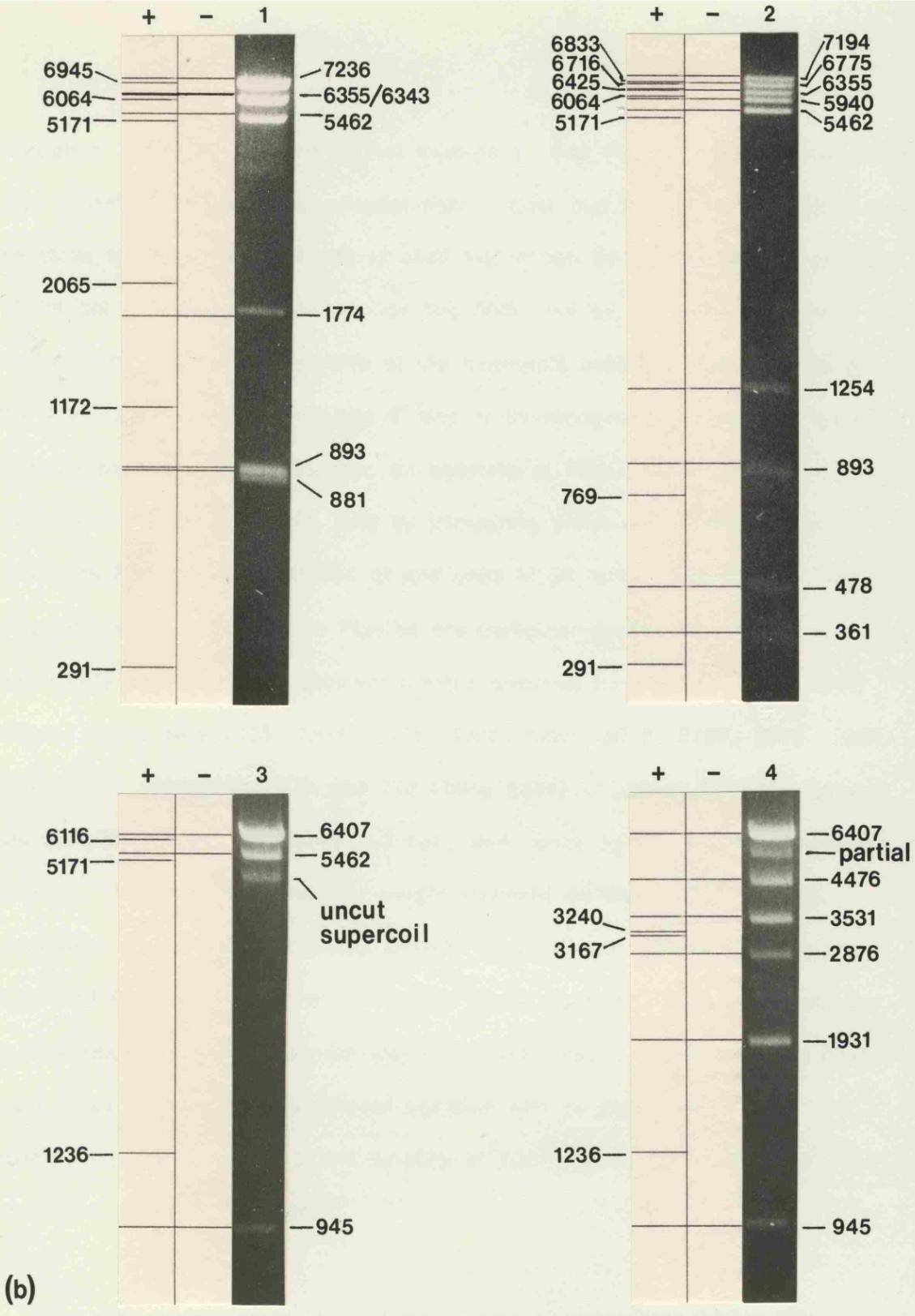

FIGURE 2-3 (continued): (b) Agarose gel analysis of M13mp7 (1, 2) and M13 (3, 4) RF digests. Each track is flanked by two computer simulations: the "+" pattern is that expected if all predicted MmeI sites were cleaved; the "-" pattern is that obtained if the doubtful MmeI site (see Fig. 2-5) is excluded. Enzymes used were: 1, MmeI; 2, MmeI/EcoRI; 3, MmeI; 4, MmeI/BamHI. (291bp bands predicted by the "+" simulations of 3 and 4 were cut off and were invisible on the gel.)

sequence GTPyGGA is found at that location.) See Figure 2-4 for a restriction map of ΦX174 DNA showing relevant sites. Note that the fifth site at 5376 is so close to the unique *Pst*I site at 5386 that it can be ignored for the purpose of the gel simulation. Because, on this DNA, not all the *Mme*I sequences are in the same orientation, the size of the fragments produced depends how far the enzyme's cut site is from the 3' end of its recognition sequence. As this had not been determined exactly, an estimate of 12bp, based on experimentation with GELSIM, and by comparing *Mme*I with similar enzymes (such as *Tth*111II), was arrived at and used in all further simulations. All fragment lengths displayed in Figures are computer-predicted, and not those empirically derived. The fragment lengths deduced from a ΦX174/*Pst*I/*Mme*I digest, then, were 5386, 5214, 5174, 5002, 3224, 3012, 2708, 2678, 2506, 2496 2161, 1990, 516, 212 and 172 (base pairs). Use of GELSIM produced the simulation shown in Figure 2-3 (a), and, once again, the correlation is striking. Because the molecular weight standard on the gel used in this simulation did not work, the size/mobility equation was derived by using the predicted fragment lengths in the double digest track themselves as standards. In this case, however, auto-correlation is legitimate: because so many figures were fitted to the simple reciprocal equation with so little error (< 3% maximum), any doubt as to the veracity of TCCPuAC as the recognition sequence of *Mme*I is removed.

The DNA sequence of the eukaryotic virus genome SV40 (Fiers *et al.*, 1978) contains two TCCPuAC sequences at coordinates 1020 and -4564. Thus, there are two complete digest products of 3514 and 1729bp, as well as the partial full-length linear fragments of 5243bp. These are visible in the actual gel photograph (not shown), but are somewhat obscured by bands produced by the small amount of *Mme*II contamination in the *Mme*I preparation.

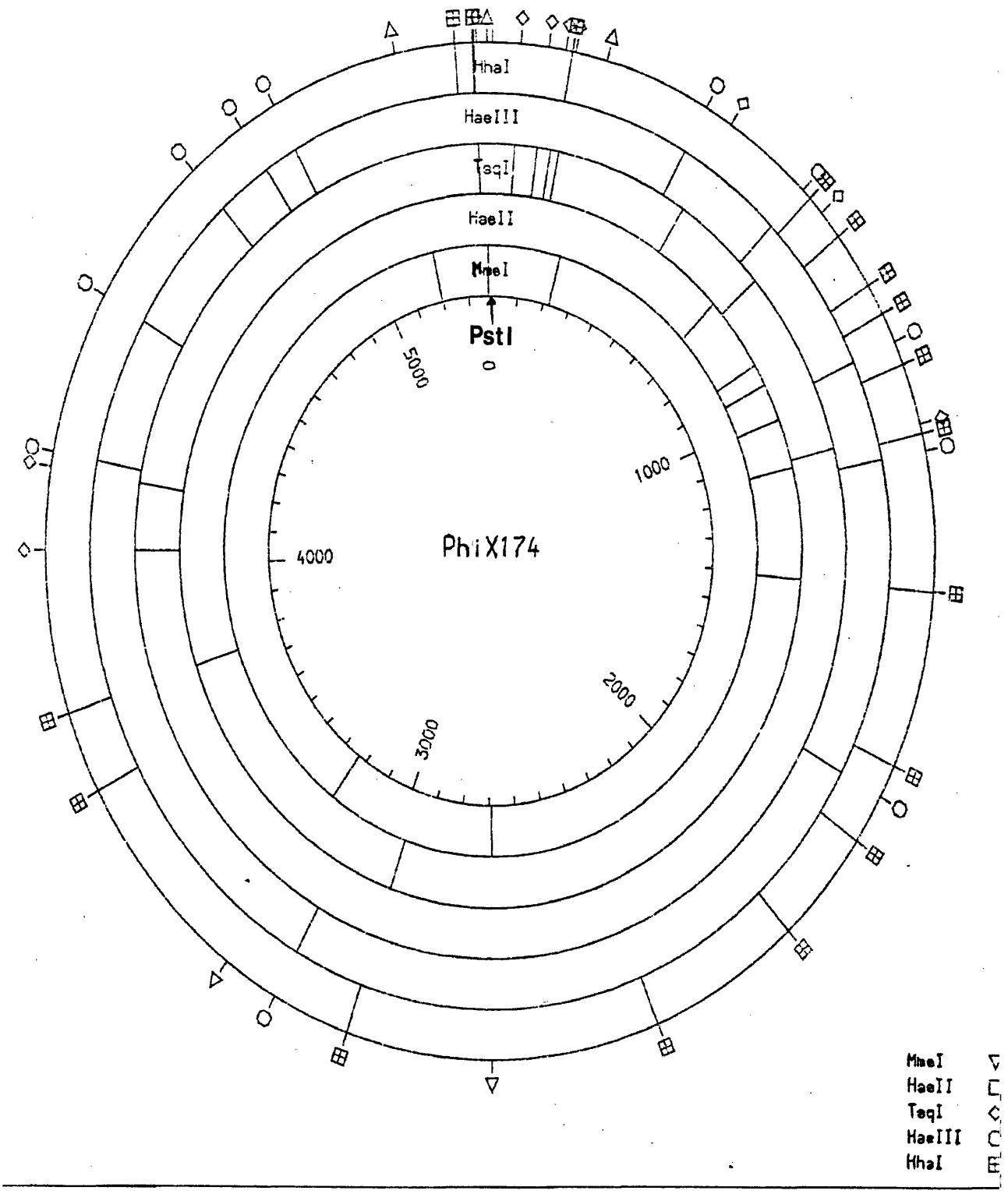

FIGURE 2-4: Restriction map of bacteriophage ØX174 (5386bp). The positions of five MmeI sites (predicted from the sequence) are shown in relation to the unique PstI and other enzyme sites.

(Since SV40 is of eukaryotic origin, its GATC sequences are not methylated, and hence are substrates for *Mme*II).

2.6. *Mme*I sites in M13 and M13mp7.

The identity of the *Mme*I recognition sequence was confirmed by comparing actual DNA digest patterns of sequenced DNAs (SV40, ΦX174) with those predicted from the sequence. As was seen above, the correlation is good enough to prove without reasonable doubt that *Mme*I recognizes TCCPuAC and cuts ca. 12bp 3' of this sequence. Unexpectedly, when another sequenced DNA, M13mp7, was digested with *Mme*I, there were fewer bands visible than the sequence data predicted. M13mp7 is a hybrid DNA molecule, constructed by Messing *et al.* (1981) and consists of the genome of the filamentous, single-stranded DNA bacteriophage M13 fused to a portion of the *lac* operon of *E. coli*, containing a synthetic multi-enzyme cloning site. it is widely used, along with other closely-related M13 derivatives, for the cloning of DNA fragments and preparation of DNA template for the Sanger dideoxy sequencing method (1977). Because it is a hybrid of sequenced DNA molecules, its sequence has been deduced rather than directly determined. A computer search of the sequence (obtained from N.i.M.R, Mill Hill, London *via* B.K. Ely) revealed four TCCPuAC sequences at coordinates -300, 5441, -5762 and 6613. GELSIM predicts 13 fragments from these data, with lengths of 7236, 6945, 6355, 6343, 6064, 5462, 5171, 2065, 1774, 1172, 893, 881 and 291bp. Clearly, (Figure 2-3 (b)), not all of these fragments are present. Careful examination of those fragments which were present suggested that one predicted *Mme*I site (at coordinate 5441, within the M13 portion) was not being cut at all. If this site is removed from the GELSIM analysis, only 7 fragments are predicted, with lengths of 7236, 6355, 6343, 5462, 1774, 893 and 881

bps. As shown in Figure 2-3 (b), the gel simulation based on these lengths is a good fit to the actual photograph. Also shown is a similar analysis of an M13mp7/*Eco*RI/*Mme*I double digest, which supports the idea that the M13mp7 sequence at 5441 is resistant to *Mme*I cleavage. That is, seven extra fragments, of which at least one (769bp) should be visible, are absent from the gel photograph. Several explanations for this observation suggested themselves: (a) sequences flanking TCCPuAC at 5441 prevent *Mme*I activity, (b) the sequence is methylated at this position or (c) there has been a mutation in M13 (since it was sequenced) changing the *Mme*I site there. In order to investigate possibility (c), I wrote to Professor J.G.G. Schoenmakers, in whose laboratory M13 had been sequenced (van Wezenbeek *et al.*, 1980) describing these results. In his reply, he stated that there was a possibility that the sequence in question was not TCCAAC, since in some sequencing runs TCTAAC could be read instead. He enclosed some M13 RF DNA from the strain actually used for sequencing, and this was subjected to further analysis.

From the M13 sequence, with *Mme*I sites at coordinates -300, 5441 and -5762, GELSIM predicts 7 fragments with lengths of 6407, 6116, 5462, 5171, 1236, 945 and 291bp. Only three fragments, plus a small amount of uncut supercoiled M13 RF, are actually visible on the gel, however, corresponding to 6407, 5462 and 945bps, which are the lengths predicted by GELSIM if the sequence at 5441 is discounted. This disproves possibility (c), and strengthens the idea that there may have been a sequencing error at 5441 - 5446. Circumstantial evidence for this view can be gained by considering the homologous regions of the closely-related phages fd and f1 (Sugimoto *et al.*, 1978; Beck and Zink, 1981). (The nucleotide sequence homologies between these phages and M13 are 97.0% and 99.1% respectively.) Significantly, neither of these phages has the *Mme*I site corresponding to the M13 site at

5441, while retaining the other two sites: both have the sequence TCTAAC instead, which is not a substrate for *Mme*I. On balance, then, it seemed likely that a small sequencing error had occurred, giving rise to the results described above. It is the simplest explanation, and therefore takes precedence over possibilities (a) and (b) above. The digests of M13 RF with *Mme*I, and *Mme*I/*Bam*HI are shown in Figure 2-3 (b), together with the simulations. The absence of 3240, 3167 and 1236bp bands in the latter double digest photograph is consistent with the above interpretation. Note also the persistence of a 5462bp *Mme*I fragment in track 4 due to incomplete cleavage by *Bam*HI. (See Figure 2-5 for restriction map of M13mp7.)

2.7. Comparison of *Mme*I with analogous enzymes.

The results described above place *Mme*I in the subset of restriction enzymes which recognize non-palindromic sequences (see Introduction 5.). All but two of these enzymes, *Tth*111II (CAAXCA) and *Gdi*II (YGGCCG), have non-degenerate recognition sequences: *Mme*I brings the total up to three. Since, apparently, *Gdi*II cleaves within its recognition sequence (Roberts, 1983), the closest known analogue to *Mme*I is *Tth*111II. In another respect, however, there is a more similar enzyme, *Hga*I, which recognizes the non-palindrome GACGC (Sugisaki, 1978). Brown and Smith (1977) discovered that *Hga*I, like *Mme*I (Results 2.1.), gave persistent non-stoichiometric digestion products. They suggested that this may be due to the heterogeneity of the sequences actually cleaved, as distinct from the recognition sequences. *Hph*I behaves in a similar way (unpublished results cited in Brown and Smith, 1977 - see also Kleid, 1980). It is possible, therefore, that the generation of partials by *Mme*I is not caused by impurities in the enzyme preparation, but is an inherent property of the enzyme.

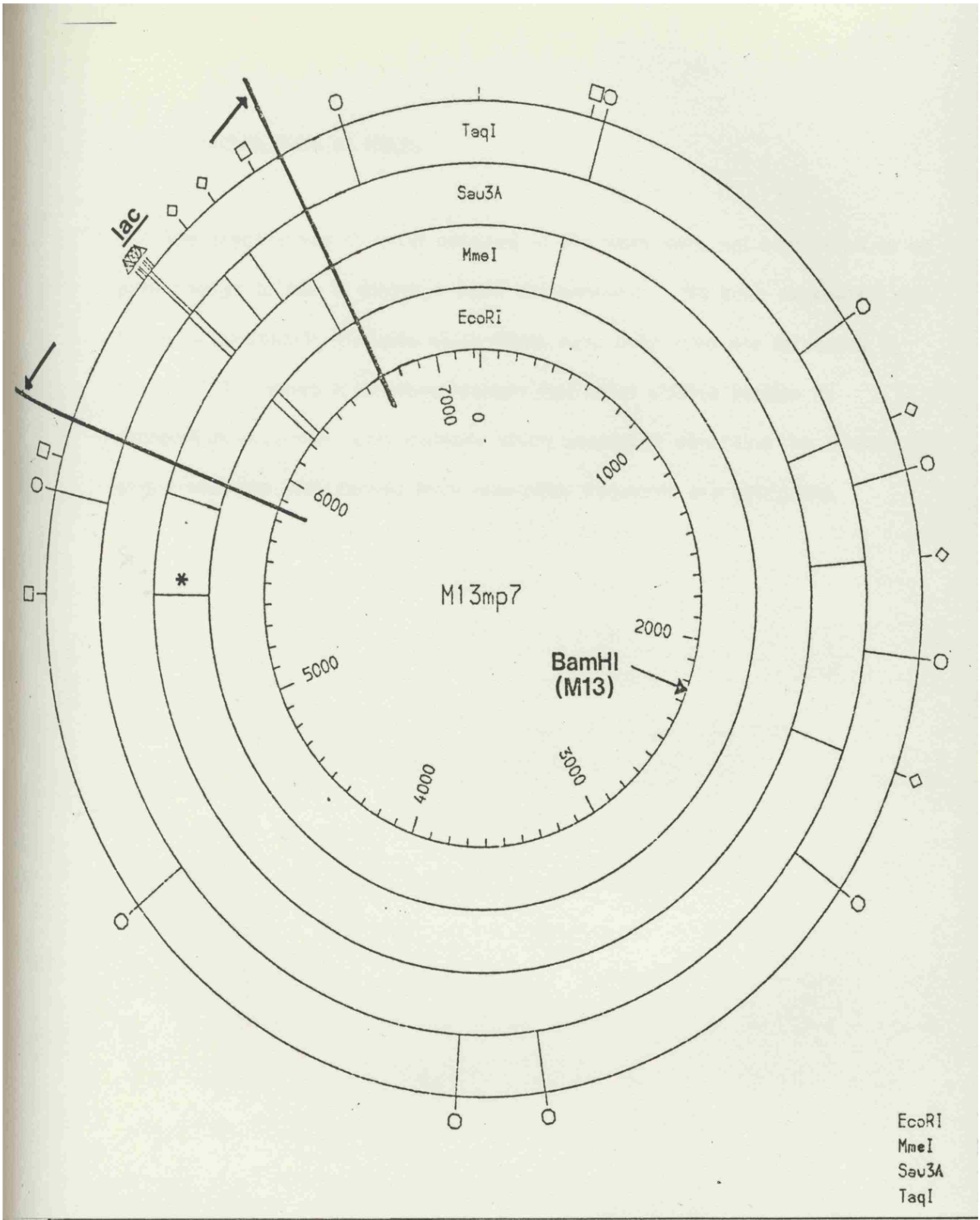

FIGURE 2-5: Restriction map of cloning vector M13mp7 RF (7236bp). The four MmeI sites predicted from the sequence are indicated, including the one (marked with an asterisk) which is not detectably cleaved by MmeI *in vitro*.

2.8. Cleavage point of *Mme*I.

The preparations of *Mme*I obtained in this work were not considered to be pure enough to use in cleavage point determination. No such experiment was therefore attempted: methods which might have been tried are discussed in Results 1.1. Since it is almost certain that *Mme*I cleaves outside its recognition sequence, only methods which accurately determine the lengths of single-stranded DNA derived from restriction fragments are approriate.

Results and Discussion: CHAPTER 3.

3.1. Detection of *Mme*II activity *in vitro*.

A second endonuclease activity was not detected in *M. methylotrophus* until about half way through the project: this was for the simple reason that, until then, *dam*-modified λ DNA had been used as the substrate for assaying column fractions; and it subsequently turned out that the new enzyme, *Mme*II, could not digest such DNA. The first hint of *Mme*II activity was found when developing a modification to the purification protocol for *Mme*I, which involved ammonium sulphate fractionation of nucleic acid free crude extract. pAT153 DNA from a *dam*$^-$ strain was used as the assay substrate for the fractions, and gel analysis of the digests clearly revealed endonuclease activity in the 40 - 50% fraction. (pAT153 is a deletion derivative of pBR322 - see Twigg and Sherratt, 1980). This extract was subsequently loaded onto a phosphocellulose column, and eluted with a NaCl gradient. The column fractions were assayed both with *dam*$^-$ and *dam*$^+$ DNA, and restriction activity was observed only when *dam*$^-$ DNA was used (not shown). it was obvious that the enzyme was completely unable to cleave *dam*-modified DNA, explaining the failure to detect its activity previously.

Subsequent experimentation with purification techniques led to the protocol described in Methods 3.3.3., involving DE52-cellulose chromatography of an ammonium sulphate fraction followed by P11 phosphocellulose chromatography of the peak fractions from the first column. Figure 3-1 is a representative gel assay of DE52 fractions during the purification of *Mme*II from, in this case,

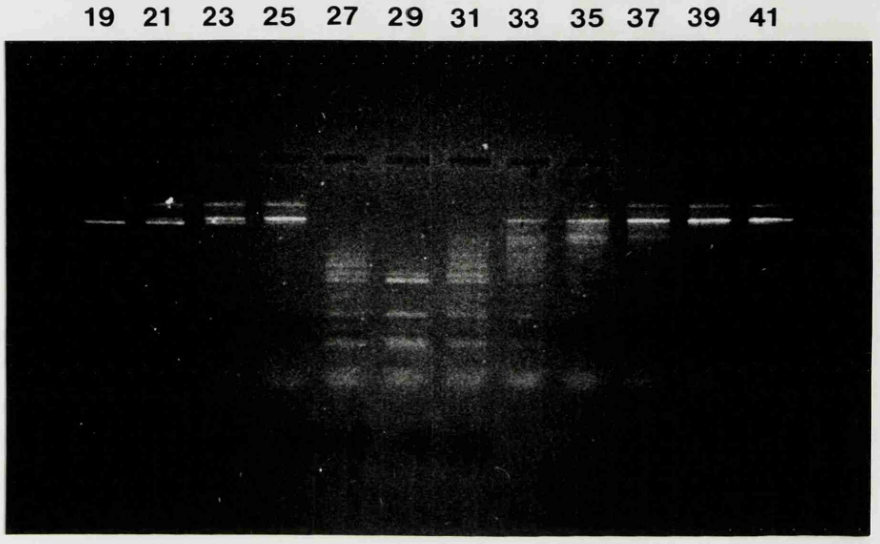

FIGURE 3-1: Agarose gel assay of DE52 column fractions in the purification of MmeII from CBM22. The peak of activity (fraction 29) is clearly visible. Substrate DNA was "dam$^-$" pAT153 from CB17.

CBM22 (see Results 4.3.). fractions from the first column. Purification of *Mme*II proved to be much easier and reproducible than that of *Mme*I, perhaps indicating a higher activity of the former enzyme in the cell. The partially purified enzyme was assayed in a variety of buffers, and it was found to be inhibited by high NaCl concentrations (cf. *Mme*I), and to work best in 666 buffer.

3.2. identification of *Mme*II recognition sequence.

The power of the computer approach to identification of endonuclease recognition sequences was clearly seen when applied to the newly-found enzyme *Mme*II: a single digest of pBR322 with partially purified material yielded several bands, the largest two of which represented fragments of approximately 1360bp and 650bp. The program RECOG was run to attempt to identify palindromes fitting these data. A window of 670 - 1385 was selected, and 128 palindromes were found which could produce a single fragment in this size range. Fortunately, this fragment was (a) the largest and (b) between 1300bp and 1400bp long for only four sequences, namely YCATGX, 2CCGG2, ATNAT and GATC, with predicted lengths of 1390, 1395, 1332 and 1374 base pairs respectively. Only the last, GATC, produces a fragment within 2% of the estimated size. Before using RECOG to explore the possibility that *Mme*II recognized a longer palindrome, or even a non-palindrome, it was decided to test the hypothesis that GATC was the recognition sequence. This is the sequence recognized by a large class of restriction enzymes produced by diverse bacterial species, but most notable among them are *Sau*3A and *Mbo*I (Stobberingh *et al*, 1977; Geiinas *et al.*, 1977) from *Staphylococcus aureus* 3A and *Moraxella bovis* respectively. Significantly, it is also the sequence recognized by the *dam* methylase of *E. coli* (Modrich and Geier, 1979).

*Sau*3A is not affected by *dam* methylation, but *Mbo*I (and *Mme*II) are, and cannot cleave *dam* modified DNA. The existence of other restriction enzymes which recognize GATC made the confirmation of this sequence as the one recognized by *Mme*II a simple matter. pBR322 DNA prepared from a dam^- strain was digested (i) by *Mbo*I, (ii) by *Mme*II and (iii) by *Mbo*I and *Mme*II. The DNA was run on a 1% agarose gel, and no difference was observed between the banding patterns of all three digests, apart from the presence of a small amount of partially digested DNA in the *Mme*II track (not shown, but compare the *Sau*3A digest of pBR322 in Figure 2-3 (a) with the *Mme*II digests of the same DNA in Figure 3-2). This experiment demonstrated unequivocally that *Mme*II recognizes the same sequence as *Mbo*I, namely, GATC.

3.3. Detection of *Mme*II restriction *in vivo*.

Of the many characterized restriction enzymes, only a few have been shown to restrict (degrade) foreign, invasive DNA *in vivo*. The best known example is the *Eco*K (Type I) enzyme of *Escherichia coli* K-12, which efficiently destroys the unmodified DNA of a phage such as λ (Dussoix and Arber, 1962). It was therefore of interest to attempt to demonstrate *in vivo* restriction in *Methylophilus methylotrophus*. Unfortunately, no phage has been found which infects this species, so the only way to investigate *in vivo* restriction was to look at its effect, if any, on the conjugal transfer of broad-host range plasmids. It was already known that plasmid (RP4) transfer from *E. coli* to *M. methylotrophus* was very efficient: however, this at first discouraging fact could be explained by considering the *In vitro* properties of *Mme*II. In particular, as was shown above, DNA from a dam^+ strain of *E. coli* is resistant to cleavage by *Mme*II. Therefore, even if *Mme*II did restrict *in vivo*, DNA transferred from "normal" (dam^+) *E. coli*, being methylated, would be

protected against cleavage. Based on this knowledge, a simple way of probing for *in vivo* restriction immediately suggested itself, i.e., *dam*$^-$ strains of *E. coli* should be poor donors to *M. methylotrophus*, because all the GATC sequences in the plasmid, being unmethylated, would be substrates for *Mme*II. To test this, 0.05ml of an O/N culture of *M. methylotrophus* was spread on half of an AS1 agar plate containing 10μg/ml Km. Into this lawn was streaked (a) CB10 (*dam*$^-$/RP4) and (b) CB15 (*dam*$^+$/RP4). After 36 hours incubation at 37°C, there was a clear difference between the streaks: on the *dam*$^+$ streak, there was confluent growth of *M. methylotrophus*/RP4 transconjugants indicating efficient transfer; on the *dam*$^-$ streak, only a few (<20) transconjugant colonies had appeared. The simplest explanation of this result is that *Mme*II does restrict incoming, unmodified plasmid DNA *in vivo*, but it could be argued that the observed difference was due to a difference between the donor strains unrelated to *dam* methylation: however, when the above experiment was repeated with K81 (CBM3) as recipient instead of *M. methylotrophus*, no difference in mating efficiency was seen between the two donor strains. K81 is an obligate methylotroph of unassigned genus (D. Byrom, pers. comm.) crude extracts of which appear to contain no endonucleases active on λ DNA (data not shown). A more convincing demonstration that *dam* methylation was solely responsible for the difference in mating efficiency was obtained when the first experiment was repeated with otherwise (essentially) isogenic *dam*$^+$/*dam*$^-$ RP4 donors, CB12 and CB26. The result was identical.

3.4. Isolation of UV-induced *Mme*II mutant (*mmeB*).

The experiments described above strongly suggest a link between the *in vitro* activity of a partially purified endonuclease *Mme*II and the existence *in vivo*

of a restriction system effective only on DNA donated by a *dam*$^-$ *E. coli* strain. Genetics offers the best way of confirming that the same enzyme is responsible for both activities. Specifically, if a mutant unable to restrict *dam*$^-$ DNA *in vivo* could be isolated, and if it was shown not to contain *Mme*II activity *in vitro*, then the conclusion would follow automatically. Soon after *in vivo* restriction was demonstrated, therefore, an experiment was devised to isolate such a mutant. Details of the UV mutagenesis and screening procedures are given in Methods 3.1.8. The rationale behind the experiment is based on the prediction that a restrictionless mutant of *M. methylotrophus* should be as good a recipient as wild-type *M. methylotrophus* of both *dam*$^+$ and *dam*$^-$ DNA from *E. coli*. In the screening process, then, a mutant colony would manifest itself by giving rise to a confluent patch of *M. methylotrophus*/RP4 transconjugants on the print plates coated with *dam*$^-$ donor. At the first attempt, of approximately 3600 colonies screened, one clone with the desired phenotype was isolated (CBM13).

3.5. UV mutant lacks *Mme*II *in vitro*.

After purification, crude extracts of an overnight culture of CBM13 were assayed for *Mme*II activity. The same procedure (see Methods 3.3.1.) was carried out simultaneously with five crude extracts of wild-type *M. methylotrophus*. (The extracts had been mixed, and labelled A - F by a colleague before the small-scale enzyme purification was attempted to eliminate unconscious bias.) *Dam*$^-$ pBR322 DNA was digested with each of the six extracts and run on a 1% agarose gel. Five of the extracts gave the distinctive *Mme*II banding pattern, but the one derived from CBM13 did not (Figure 3-2). Hence it was proved conclusively that *Mme*II restricts DNA both *in vitro* and *in vivo*.

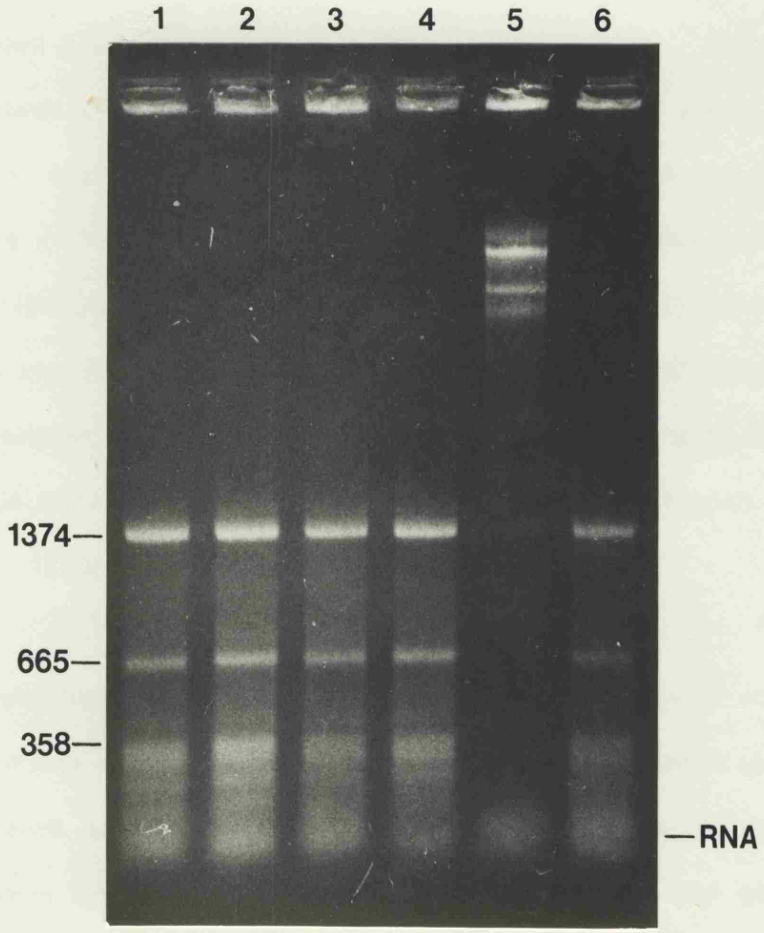

FIGURE 3-2: Agarose gel analysis of pBR322 treated with extracts from CBM1 (w.t.) and CBM13 (mmeB⁻). CBM1 extracts (1-4, 6) all gave characteristic MmeII bands which are absent from the digest produced by the CBM13 extract (5). Traces of MmeI activity are visible in track 5.

3.6. Cleavage point of *Mme*II.

One biochemical property of *Mme*II of interest is the position of the double-stranded break within the recognition sequence GATC. This property has already been determined for three other enzymes with a similar specificity, namely *Sau*3A, *Mbo*I and *Dpn*I (Roberts, 1983). Both *Sau*3A and *Mbo*I cleave at the 5' side of the guanine residue leaving a single-stranded four base extension on the DNA duplex ("sticky" ends). In contrast, *Dpn*I (which can cleave GATC sequences only if both adenine residues are methylated) cleaves the duplex between the adenine and thymine residues, generating fragments with no single stranded extensions at all, or "blunt" ends (Modrich and Geier, 1979; Lacks, 1980)

As a preliminary to determining the cut-point of *Mme*II by *in vitro* techniques, it was decided to test the possibility that the enzyme generated the same sticky ends as *Sau*3A by an *in vivo* assay. If this was the case, then *Mme*II fragments should anneal with *Bam*HI-cleaved M13mp7 DNA generating recombinant molecules. After ligation and transfection, these would then give rise to white plaques on Xgal indicator plates. If no white plaques were detected, then the conclusion would be that *Mme*II was cleaving differently from *Sau*3A. The experiment was performed in the following way. 0.5μg of *dam*$^-$ pBR322 DNA was digested with (a) column purified *Mme*II and (b) *Sau*3A. 50ng aliquots of each were mixed, in separate reactions, to 50ng *Bam*HI-cleaved M13mp7 RF, and 0.1u ligase (Boehringer) added in the presence of ligase buffer and 0.4mM ATP. Both mixtures were incubated O/N at 10°C. Competent cells of JM101 (CB19) were transfected with the ligation mixes, and immediately plated out on Xg indicator plates (see Methods 3.1.4.1). The total blue/white plaque counts for (a) and (b) were 1135/0,

1002/302 respectively. Whereas 23% of the plaques resulting from the *Sau*3A cloning were recombinant (white), no recombinant plaques at all were obtained from the attempted *Mme*II cloning. This would be expected if *Mme*II produced different sticky ends from the 4bp 5' extensions produced by *Bam*HI and *Sau*3A. Alternatively, the *Mme*II enzyme preparation may have contained contaminants (exonucleases or phosphatases) able to damage any sticky ends produced by the restriction enzyme. A further cloning experiment was done to investigate the latter possibility. 0.5μg *dam*$^+$ pAT153 DNA was digested with (a) *Sau*3A and (b) *Sau*3A + *Mme*II. /Since the plasmid DNA itself (being *dam*-modified) was resistant to *Mme*II cleavage, the ends produced in (b) would all have *Sau*3A ends. Any phosphatase or exonuclease in the *Mme*II preparation, therefore, would damage the *Sau*3A ends, reducing the yield of recombinants in (b) compared to the yield from (a). Ligation reactions and transfection of JM101 were carried out exactly as before, and transfected cells plated out on Xg plates. The blue/white plaque counts were 387/85 and 470/38 for (a) and (b) respectively, giving recombinant percentages of 18.0% and 7.5%. Although the figure for (b) is less than half that for (a), it suggests that the *Mme*II preparation was not so grossly contaminated that contamination alone could account for the inability to clone *Mme*II fragments of pBR322 into the *Bam*HI site of M13mp7 RF. The simplest interpretation of the above is that *Mme*II cleaves GATC sequences differently from *Sau*3A.

The position of the *Mme*II cut site still remains to be established. One attempt to determine the cleavage point was made, using thin-layer chromatography (TLC) of mononucleotides produced by λ exonuclease digestion of ^{32}P-labelled *Mme*II fragments, following exactly the method used by Molemans *et al.* (1982) to determine the cleavage points of *Cau*I and *Cau*II. Their method was based on a simple TLC procedure for identifying

monoribonucleotides developed by Volckaert and Fiers (1977). Unfortunately, the attempt did not work, and there was no time left to repeat this experiment. The point of cleavage is of biochemical interest, especially if it differs from that of *Sau*3A, as seems likely.

Results and Discussion: CHAPTER 4.

4.1. Use of *M. methylotrophus* to monitor *dam* genotype.

The striking difference between the mating efficiencies of *dam*$^+$ and *dam*$^-$ strains of *E. coli* with *M. methylotrophus* suggests a simple test for the presence or absence of *dam* methylation in a strain. Firstly, a broad host-range plasmid such as RP4 has to be introduced into the strain; then, the strain must be tested for its mating efficiency with wild-type *M. methylotrophus*. If this is high, then the original strain is *dam*$^+$: if barely detectable, the strain is *dam*$^-$. As a positive control, the strain bearing RP4 can be tested for its mating efficiency with CBM13 (*mmeB1*). This should be high for any donor, regardless of the *dam* genotype. A rapid, replica mating technique which can be used to screen hundreds of clones simultaneously is given in Methods 3.1.3. and 3.1.9.

4.2. P1 transduction of *dam*.

One application of the test is the construction of *dam*$^-$ strains of *E. coli*, which are useful for the preparation of DNA susceptible to cleavage by restriction enzymes inhibited by *dam* methylation, e.g., *Mbo*I (*Mme*II), *Bcl*I and *Xba*I. Strains CB25 and CB26 were constructed by P1 transduction of CB23 and CB12 respectively, using this test to detect *dam*$^-$ clones. In order to transduce a non-selectable marker like *dam*, it is necessary to transduce a nearby selectable marker, and screen a few transductants for the presence of the other gene. A convenient marker in this case is a point mutation in the *rpsL* gene which confers high-level streptomycin resistance on the cell. Accordingly, a spontaneous Sm^R mutant of CB2 was isolated (CB18), and a

P1kc lysate of the strain prepared. A suitable plasmid (e.g., RP4 or R751) was introduced by conjugation into the *dam*$^+$, Sm^S strain, and the *rpsL* allele introduced via the CB18 P1kc lysate. 100 Sm^R transductants were patched out in a grid pattern and subjected to the screening procedure described in Methods 3.1.9. About 5% of the clones were *dam*$^-$, a slightly lower cotransduction frequency than the published 10% value (Marinus, 1973) There is another method for screening for the *dam*$^-$ allele which relies on the fact that *dam*$^-$ strains are highly sensitive to the base analogue 2-aminopurine (2AP). *Dam*$^-$ strains will not grow on minimal plates containing 200μg/ml 2AP (Glickman *et al.*, 1978) In one experiment to construct CB36, 100 Sm^R transductants were screened by the replica mating test, and were also replica plated onto a 2AP plate. Three *dam*$^-$ clones were detected by the first test, and only these three were sensitive to 2AP. Therefore, the two tests give identical results, showing in particular that no *dam*$^-$ transductants were missed by the new test described here.

4.3. Tn*5* mutagenesis of *M. methylotrophus*.

Genetically, *M. methylotrophus* is a poorly characterized organism. Part of the reason for this is the difficulty of obtaining mutants requiring certain amino acids, because the cell apparently lacks the necessary permeases (D. Byrom, pers. comm.), so a lesion in the biosynthetic pathway would be lethal. Lack of a transducing phage (or efficient sex factor) means that the construction of multiply-marked strains is also difficult (Brammar, 1981; Holloway, 1981). Modern cloning techniques, however, allow the possibility of examining genes from *M. methylotrophus* in isolation, and in a well-characterized genetic background such as *E. coli*. The latter point was the rationale behind the next step in the genetic characterization of restriction genes in *M. methylotrophus*.

This was to generate *mmeB* mutants using the transposon Tn*5* as a mutagen, an approach which stably inserts a selectable marker (Km^R) into the gene (see Heilmann *et al.*, 1980). The idea was then to clone out the transposon with flanking DNA sequences and physically characterize the *mmeB* gene. Fortunately, a convenient system for efficiently introducing Tn*5* into the *M. methylotrophus* chromosome was already available in the form of a ColI plasmid derivative produced by B.M. Wilkins and G. Boulnois (pers. comm.) This plasmid, pLG221 (Boulnois, 1981), has a broad-host range for conjugation, but a narrow one for replication. *M. methylotrophus* is one of the organisms for which this is true, so an effective way of making Tn*5* mutants of *M. methylotrophus* is to mate it with an *E. coli* strain harbouring pLG221 and to select for Km^R colonies. Since the plasmid cannot replicate in *M. methylotrophus*, the only Km^R cells that can arise must be due to transposition of Tn*5* onto the chromosome. Tn*5* is one of the least site-specific transposons, and can insert apparently at random along a stretch of DNA (Heilmann *et al.*, 1980; Calos and Miller, 1980). Thus, only 10^3 - 10^4 Km^R colonies need be screened to detect a mutant in any particular gene, assuming a genome of about the same size as *E. coli*.

0.5ml of O/N DS903 (pLG221) were filter mated with an equal volume of O/N CBM1 (see Methods 3.1.2.) at 37°C for 22 hours. The filter was transferred to 5ml AS1 salts and vortexed to remove cells. 5 x 0.1ml were plated out on medium containing 10μg/ml Km, and the remainder spun down. The pellet was resuspended in 1ml AS1 salts, and 5 x 0.1ml plated out as before. (The remaining 0.5ml was diluted to 10ml with AS1 minimal medium plus Km, and shaken O/N at 37°C, but was not used, since the number of Km^R colonies which grew on the plates just described were sufficient to carry out the screening process.) The plates were incubated at 37°C and mutant

colonies allowed to grow for two days. Each of the five plates from the concentrated cell suspension contained about 14,500 Km^R colonies in all: therefore, the filter itself must have had approximately 3 x 10^4 such cells in a total population of approximately 10^9 cells. The frequency of transposition to the chromosome in this experiment, then, was about 3 x 10^{-5}.

The next step was to screen the Km^R colonies for the restrictionless phenotype, in a way similar to that described for the isolation of CBM13. Unfortunately, since the colonies were already Km^R, use of RP4 limited the choice of antibiotics for screening to Ap or Tc. Ap has the drawback that cells containing RP4 secrete β-lactamase into the medium, and can therefore allow Ap^S cells in the vicinity to grow. So, although wild-type *M. methylotrophus* is extremely sensitive to it, Tc was the antibiotic used. The five plates were replica mated onto medium containing 2.5μg/ml Tc and overlaid with 0.1ml lawns of CB10 (*dam*$^-$, RP4). After three days incubation, ten positive signals were identified, although background growth was considerable on some plates. Because the colonies on the regenerated master plates were essentially confluent by then, 1cm diameter circles around the sources of the positive signals were isolated and excised. These were transferred to liquid medium plus Km and grown O/N, before being diluted and plated out on Km plates. The ten plates containing ca. 200 colonies each were subjected to a further round of screening exactly as described before. This time, only eight plates gave positive signals: all colonies from the masters that corresponded to these signals were picked and streaked for singles on Km plates (24 in all). Two single colonies from each of these purification plates were picked and grown up O/N in liquid medium. 10μl of each culture was spotted onto each half of Tc plates overlaid (on one half) with CB10 (*dam*$^-$, RP4) and (on the other) with CB15 (*dam*$^+$, RP4). Patches on which transfer was equally

efficient from each donor were taken to correspond to restrictionless mutants. CBM1 and CBM13 controls were included on each plate. Although this test identified some definitely restrictionless mutants, there was doubt about some clones. The availability of *dam*$^{+}$/*dam*$^{-}$ strains of *E. coli* containing R751 instead of RP4 (CB24 and CB25 respectively) allowed the repetition of the above test on plates containing Tp. This was much more conclusive, and, using this test, 16 clones corresponding to 6 different regions on the original five plates proved to be restrictionless. Six representative clones were selected and purified twice on Km plates before being assigned strain numbers CBM18 - CBM23. On the basis of the simple plate matings described above, all but one (CBM22) of the six Tn*5* mutants appeared phenotypically indistinguishable from the original *mmeB* mutant, CBM13. The exceptional strain appeared to be a better recipient than CBM1 (or CBM13) of *dam*-modified DNA: in contrast, as a recipient of unmodified DNA, it appeared much better than CBM1, but slightly and reproducibly worse than CBM13. One explanation that sprang immediately to mind was that this isolate did not lack *Mme*II, but had lost the other endonucleolytic activity, *Mme*I. Plate matings, however, give essentially qualitative results, and any idea of this kind could only be assessed through more rigorous, quantitative mating experiments - to be described below (Results 5.3.). It should perhaps be noted that the five *mmeB*::Tn*5* mutants are not necessarily independent: because the original cross was performed on a single filter, and the filter incubated for almost a day on minimal medium, there is a possibility that some isolates are siblings. Until the mutants are physically characterized, there is no way of knowing how they are related. All further analyses, therefore, were conducted with the same strain CBM21, taken to be representative of the others. Small scale assays of CBM21 extracts (cf. Results 3.5.) confirmed that this strain does not produce *Mme*II (not shown) but does produce *Mme*I (Figure 4-1 (b)).

The success of this experiment clearly demonstrates the power of the relatively new technique of using transposons to generate mutant genes with selectable markers inserted in them. The completely unexpected bonus of isolating a potentially *Mme*I$^-$ mutant was particularly gratifying, and characterization of this mutant became the dominant objective of the project. So, the original motive behind the Tn*5* mutagenesis, that of physically characterizing the *mmeB* gene, was postponed due to pressure of time. As will be seen below, CBM22 has some interesting properties which justified this change of direction.

4.4. CBM22 - a mutator strain?.

As a preliminary to chromosome mobilization experiments, it was necessary to construct derivatives of CBM1, CBM13 and CBM22 with chromosomal antibiotic-resistance markers for counterselection of donors. Mutations to nalidixic acid resistance (50μg/ml) in *M. methylotrophus* occur at a low frequency of ca. 10^{-9} (D. Byrom, pers. comm.), so this seemed an ideal drug to use. Accordingly, ca. 10^{10} cells of each strain were plated out on media containing 50μg/ml Nal, and incubated at 37°C. After two days, a few (<10) mutant colonies of Nal^R CBM1 and CBM13 appeared, but on the CBM22 plates, there were at least a thousand colonies (albeit of variable size). This phenomenon was at first thought to be an effect of Tn*5*, which CBM22 contains: but when a similar experiment was carried out with equal amounts of CBM22 and CBM21 (both Tn*5* mutants), only CBM22 had an (apparent) increased mutation rate to nalidixic acid resistance. So the effect seemed to be restricted to the specific Tn*5* mutant CBM22. Two possibilities existed: either that the lesion in CBM22 was specifically affecting the mechanism of nalidixic acid resistance; or that CBM22 was, simply, a mutator strain. To

test this, the mutation rate of CBM22 to drug resistance against a variety of antibiotics was measured. At this late stage, only a simple experiment was performed to investigate the possible mutator phenotype. 0.1ml of O/N cultures of CBM1 and CBM22 were each spread on AS1 plates containing Ap, Tc, Rif, Sm, Nal or Tp. After 2 days growth, CBM22 gave significantly more mutant colonies (10^2 - 10^4 x) than CBM1 (10 - 100 colonies) on Tc, Nal and Tp. No elevation of mutation rate in CBM22 to Ap, Rif or Sm resistance was observed. The Tp^R and Tc^R mutants were small, slow growing and thus distinguishable from *M. methylotrophus* strains containing plasmid-borne genes for these antibiotic resistances. These results support the idea that CBM22 is a mutator strain, but more sophisticated experiments, such as the measurement of reversion rates of well characterized nonsense mutants, need to be done in order to strengthen the hypothesis. A further piece of evidence supporting the hypothesis is presented in Results 5.2.

4.5. Characterization of mutants *in vitro*.

When the UV-induced mutant CBM13 was isolated, a rapid, small-scale purification technique was all that was required to show that the strain lacked *Mme*II activity. Naturally, it was desirable to assay extracts of the novel Tn*5*-induced mutant CBM22 for endonuclease activity. Detection of *Mme*I activity using the small-scale method is not usually as unequivocal as detection of *Mme*II, probably due to the small amount of intracellular *Mme*I enzyme. Several attempts were made to try and demonstrate that CBM22 lacked *Mme*I activity, but none of them were completely unequivocal. It can be said with certainty, however, on the basis of such experiments, that CBM22 does produce *Mme*II, but lacks an endonuclease which is produced by CBM1 active against *dam*-modified DNA. The best demonstration of this is seen in Figures

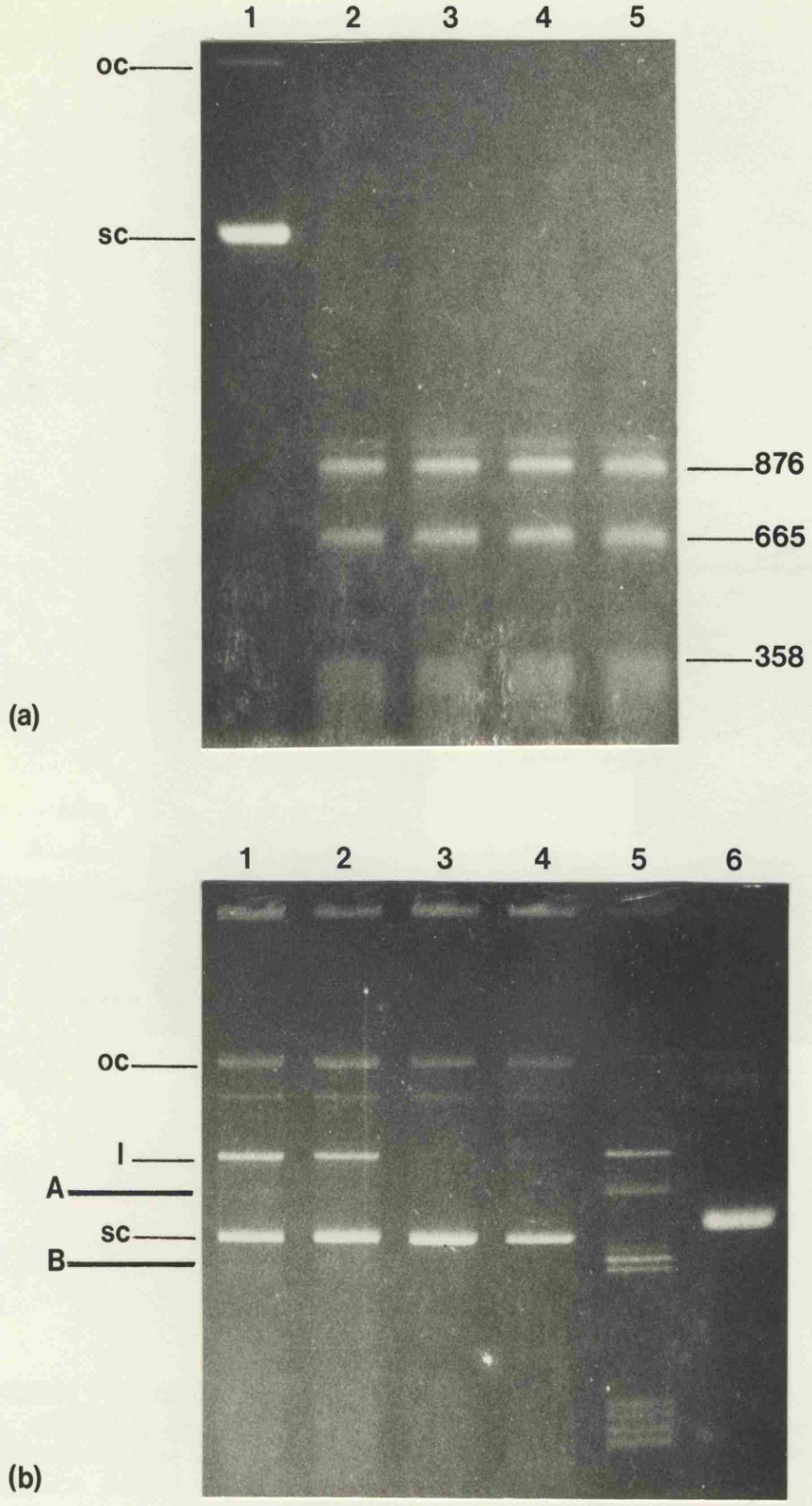

FIGURE 4-1: Agarose gel analysis of pAT153 treated with extracts from CBM1 (w.t.) and CBM22 (endA⁻). (a) "dam⁻" DNA: 1, uncut; 2 & 3, CBM1; 4 & 5, CBM22. (b) "dam⁺" DNA: 1 & 2, CBM1; 3 & 4, CBM22; 5, MmeI; 6, uncut. A and B indicate band patterns characteristic of MmeI activity. Tracks 1-4 of (b) show that a substantial amount of linearization (and traces of A and B bands) are produced by CBM1 extracts but not by CBM22 extracts. Note that both extracts exhibit MmeII activity in (a), tracks 2-5. (sc = supercoil; oc = open circle; l = linear.)

(c)

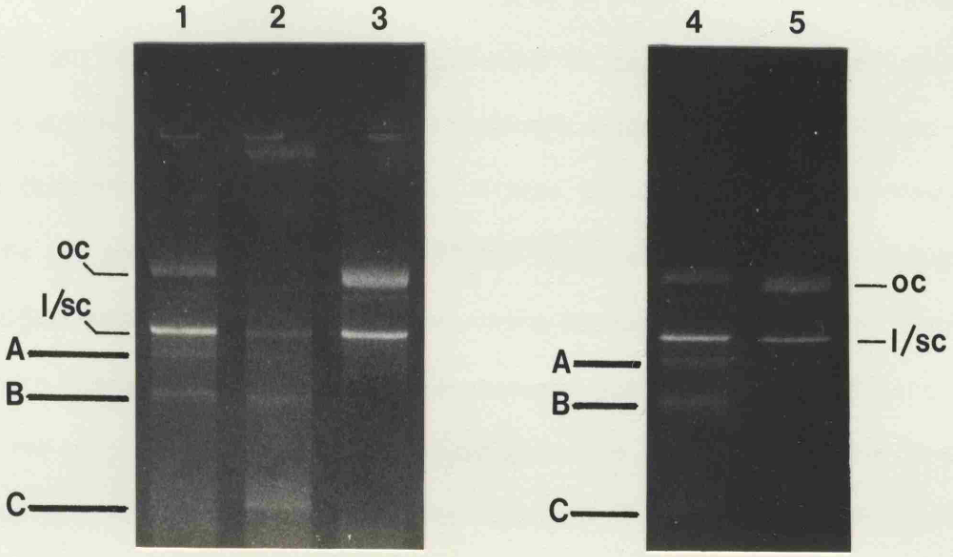

FIGURE 4-1 (continued): (c) "dam$^+$" DNA: 1, CBM1; 2, CBM21 (mmeB$^-$); 3, CBM22; 4, MmeI; 5, uncut. A, B and C indicate band patterns characteristic of MmeI activity. As in (a) and (b), only the CBM22 extract failed to produce these bands. (Note that the gel was run for such a short distance that supercoils apparently co-migrated with linear DNA.)

4-1 (a) and (b), where crude extracts of CBM1 and CBM22 were assayed for endonuclease activity on both *dam*$^+$ and *dam*$^-$ DNA. Very clearly, both strains (Figure 4-1 (a)) produce *Mme*II, as unmodified DNA is digested identically by each extract. A difference is clearly seen, however, in the digests of *dam*-modified DNA: here, the CBM1 extracts have substantially cleaved the supercoiled plasmid DNA into a linear form; no such conversion by the CBM22 extracts is visible (Figure 4-1 (b)). Traces of other *Mme*I bands formed by CBM1, but not by CBM22 are also indicated (A and B). Figure 4-1 (c) shows a separate experiment in which extracts from CBM21 (*mmeB*) and CBM22 were compared with a CBM1 extract. Traces of bands indicative of *Mme*I activity are again missing from the CBM22 track, although they do appear in the CBM1 and CBM21 tracks, perhaps more convincingly than in the previous figures. Thus, CBM22 lacks an endonuclease possessed by CBM1, and though it is tempting to surmise that this endonuclease is *Mme*I, the *in vitro* evidence, is, as yet, insufficient to confirm this suspicion. Accordingly, the mutant gene in CBM22 is designated *endA1*::Tn*5* until its true identity can be ascertained.

Results and Discussion: CHAPTER 5.

5.1. Simple way of isolating *mmeB* mutants.

Two ways of isolating *mmeB* mutants were described above: (a) UV mutagenesis (Results 3.4.) and (b) Tn*5* mutagenesis (Results 4.3.). Towards the end of the project, a third method was devised which is simpler and quicker than the first two. It was based on the idea that the *Mme*II restriction barrier against *dam*-unmodified DNA was so strong that virtually all incoming plasmids would be restricted, and fail to establish themselves. In a cross between *M. methylotrophus* and a *dam*$^-$ *E. coli* strain harbouring a broad-host range plasmid, then, the recipients should be at least enriched for *mmeB* mutant cells which had lacked *Mme*II in the original *M. methylotrophus* population, and were thus good recipients for such unmodified DNA. An experiment to test this idea was conducted as follows. 10ml O/N cultures of CBM1 and CB35 (*dam*$^-$, pNJ5073) were concentrated tenfold by centrifugation, and filter-mated at 37°C for ca. 24 hours. pNJ5073 is a derivative of an RP4::Tn*7* plasmid and is Ap^S, Km^S, Tc^R, Tp^R, Sm^R, Su^R, Tra^+ but unable to transpose. It is highly unstable, and rapidly lost from cultures grown without selection (Grinter, 1983). Cells were removed from the filter in 10ml AS1 salts and spun down before being resuspended in 0.5ml salts. Five Tc containing plates were spread with aliquots of this suspension and incubated for two days at 37°C. After this time, ca. 500 Tc^R colonies in all had arisen: evidence of the strength of *Mme*II restriction *in vivo*. 100 colonies were taken from a single plate and patched onto two AS1 plates without antibiotics. It was hoped, given the instability of pNJ5073, that most of the cells in the fully-grown patch would be plasmid free, and thus not exhibit incompatibility towards incoming plasmids of the same group. The next day, the master

plates were replica mated onto Km supplemented AS1 plates overlaid with (a) CB26 (*dam*$^-$, RP4), then (b) CB12 (*dam*$^+$, RP4). All plates were incubated at 37°C for 48 hours. The result was as expected: 67 patches had given rise to confluent growth on both copy plates, indicating lack of *Mme*II restriction. All the others had retained the restriction system, since they formed no patches of growth on the *dam*$^-$ donor plates. Using the *dam*$^+$ donor plate, it was also established that all clones had retained the *Mme*II modification system, since they transferred RP4 equally well to CBM51 and CBM52. Mutants obtained in this way were purified, checked for loss of pNJ5073 (lack of Tc and Tp resistance) before being stored. CBM59, CBM60 and CBM61 are independent, spontaneous *mmeB* mutants obtained in this way. The method is recommended for the introduction of *mmeB* mutations into *M. methylotrophus* strains with other markers.

5.2. Construction of double mutants.

It was clear from the results described above (4.3. and 4.5.) that CBM22 produces *Mme*II and that *in vivo*, it restricts the entry of *dam*$^-$ plasmid DNA (albeit less efficiently than CBM1, see Results 5.2.). The construction of a derivative of CBM22 which lacked *Mme*II activity, then, was of considerable interest. Initially, the method considered best for the strain construction was to mobilize out the *endA1*::Tn*5* marker into a Sm^R *mmeB1* strain, CBM52, selecting for Km^R Sm^R (Holloway, 1981). The most efficient way of achieving this (in *M. methylotrophus*) is to use RP4 primes containing fragments of *M. methylotrophus* chromosomal DNA. These primes mobilize the chromosome by virtue of recombination between the regions of homology between plasmid and chromosome immediately prior to conjugation, in a manner presumed to be similar to that observed with F′ mobilization in *E. coli*. Fortunately, a series

of RP4 primes containing *M. methylotrophus* DNA had been constructed *in vitro* by D. Pioli (pers. comm.). Random *Hin*dIII fragments of *M. methylotrophus* DNA were cloned into RP4 cleaved at the single *Hin*dIII site within the Km^R gene, and recombinants introduced into *E. coli* by transformation. These were ideal for the mobilization experiment because they had lost the Km^R gene. Mobilization of a chromosomal marker by an RP4 prime is most efficient when the chromosomal part of the plasmid is adjacent to the marker, so the first task was to attempt to identify a prime of this kind. The RP4 primes were introduced into CBM22 by conjugation with an assortment of twelve *E. coli* strains containing the plasmids, selecting for Tc^R.

A great deal of time was spent in trying to create a double mutant by mobilisation, but the difficulty of simply screening transconjugants for both characters meant that not enough clones could be tested simultaneously, resulting in a lack of success. Therefore, the experiments will only be described briefly, as a different technique described afterwards was immediately successful. In one experiment, 0.1ml of each RP4' O/N culture was mixed with an equal volume of CBM52, and dilutions plated out on medium containing streptomycin and kanamycin. After two days growth, the number of Sm^R Km^R colonies on plates spread with the undiluted mixture ranged from 0 to 14 (with only pLP31 and pLP149 giving no colonies at all). The expectation had been that an RP4' with homology near to the chromosomal Tn*5*, and in the correct orientation, would give rise to chromosomal mobilization orders of magnitude more frequently than more "distant" primes. Clearly, this had not happened here. At the time, however, there was little reason to suspect that the doubly-resistant colonies which had grown were not the result of *bona fide* (albeit inefficient) chromosome mobilization. Thus, the strain giving rise to most (14) Sm^R Km^R colonies, CBM42 (containing pLP122), was chosen for

further analysis. At the same time, 25 Sm^R Km^R colonies (resulting from the best donors CBM36 and CBM42) were gridded onto a plate containing Sm and Km. When the patches had grown, they were replica plated onto medium containing tetracycline. All patches gave rise to confluent growth on Tc, indicating that each clone had (probably) received an intact RP4', even if chromosome mobilization had also occurred. This is apparently a common feature of RP4' mobilization in *E. coli* (N.J. Grinter, pers. comm.) and distinguishes it from F' mobilization, in which recipients are almost always F^-.

In an experiment of this nature, Sm^R Km^R transconjugants can arise in two distinct ways: either (as desired) by mobilization of chromosomal Tn*5*, or by transposition of Tn*5* onto the RP4' and subsequent plasmid transfer without concomitant chromosome transfer. Transconjugants which have arisen by the second route can be identified by taking advantage of the fact that they now contain a Tc^R, Ap^R, Km^R sex factor, and should be able to transfer all three markers simultaneously and efficiently to another host, for example, *E. coli*. To test this, the plate with 25 clones was replica mated onto a minimal agar plate containing 25μg/ml Km and coated with a 0.1ml lawn of CB23. In all but one case, efficient transfer of kanamycin resistance to *E. coli* was observed. In this mobilization experiment, then, the most frequent event was the transposition of Tn*5* to the RP4' and subsequent transfer, by simple conjugation, to the Sm^R recipient CBM52. (Although not tested, it is likely that the exceptional clone was also an RP4'::Tn*5* plasmid, but was unable to transfer itself to *E. coli* because the transposon had inactivated a *tra* gene). It could be argued that the conditions of the experiment, i.e., the immediate plating out of the mating mixture on selective media, was likely to bias the results in favour of recovering RP4'::Tn*5* plasmids, rather than mobilizing the chromosome, since transfer of the former would be much quicker than the

latter event. Even when the mating was performed on a filter, however, most of the recipients (as before) were capable of transferring kanamycin resistance efficiently to *E. coli*, suggesting transposition rather than chromosome mobilization as the most frequent event.

The approach which finally yielded a double mutant relied on evidence accumulated earlier that CBM22 was a mutator strain (Results 4.4.). It had been shown (Results 5.1.) that most *M. methylotrophus* cells which succeeded in receiving plasmid DNA from a *dam*$^-$ donor were spontaneous *mmeB* mutants. Combining these two facts with the observation that CBM22, in simple plate matings, was a 10 - 100 fold better recipient of *dam*$^-$ DNA than CBM1, results in the idea that the difference may not be solely due to the alleviation of *Mme*I restriction *in vivo*, but may be due to the higher proportion of *mmeB* mutants in the original culture of CBM22. To test this, therefore, CBM22 was filter mated O/N with CB35, the *dam*$^-$ strain carrying the unstable RP4 derivative pNJ5073 coding for Tc^R and Tp^R. (The donor cells were concentrated tenfold prior to mating, a slight modification to the normal procedure.) Dilutions were plated out on medium containing Tc and Tp, and cells allowed to grow for two days. 96 well-isolated Tc^R Tp^R colonies were picked and patched onto two antibiotic-free minimal plates. Each plate had a control patch of CBM13 and CBM22. The idea behind growing the patches on unsupplemented plates was to allow the unstable plasmid to segregate out, and to maximize the number of plasmid-free cells in the fully-grown patch. When the plates were grown, then, they were replica-mated onto plates supplemented with Tc and Ap, and overlaid with 0.1ml lawns of CB26 (*dam*$^-$, RP4), after being replica plated onto similar plates lacking donor cells. Two days later, none of the patches had grown on the control plates, but all 96 test patches had given rise to confluent growth of RP4-containing recipients at a higher efficiency, it seemed,

than either CBM13 and CBM22. This was an encouraging sign, since double mutants would be expected to have this phenotype. Many of the cells in the test patches, however, would have retained the Tc^R plasmid pNJ5073, and may thus have influenced the result of the plate mating, so it was necessary to purify a few clones from the patches and be certain that pNJ5073 was eliminated before re-testing the restriction phenotype. Cells from four out of the 96 patches were arbitrarily picked, and streaked for single colonies on minimal plates. Two single colonies from each of these plates were repurified on similar plates, and simultaneously tested for tetracycline resistance on Tc-supplemented plates. Four of the six colonies tested were Tc^S, so three belonging to distinct patches on the original plate were retained (CBM62 - 64). These were subsequently found to be Tp^S as well, indicating complete loss of the plasmid pNJ5073. Plate mating tests using *dam*$^-$ donors and comparing these three clones with CBM13 and CBM22 as recipients confirmed the earlier observation that they appeared significantly better recipients than the strain from which they were derived, CBM22. Only a slight improvement over CBM13 was detected, however. As has been stated before, plate matings only give an indication of conjugation efficiencies, and filter matings (see below) are necessary to obtain reliable, quantitative data on the efficiencies of different strains.

5.3. Quantitative comparison of mutant restriction phenotypes.

The availability of *M. methylotrophus* mutants lacking endonucleases e.g., CBM13 (*mmeB*), CBM22 (*endA*) and CBM64 (*mmeB endA*) allowed a quantitative comparison of their restriction phenotypes. The recipient abilities of the above strains (and CBM1) were thus tested in crosses with RP4-containing *dam*$^+$ and *dam*$^-$ *E. coli* donors, CB12 and CB26. 0.5ml of

O/N donor was mixed with 0.5ml of O/N recipient in all eight possible pairwise crosses, and filter-mated on non-selective AS1 plates for 24 hours. After this time, cells were removed into 5ml AS1 salts, and dilutions plated out onto AS1 plates supplemented with Ap and Tc. (The inclusion of both antibiotics was necessary to eliminate any effects of the possible mutator phenotype of *endA* strains on the experiment.) The results of one experiment are shown in Table 5-1. Viable counts of *M. methylotrophus* cells on the filter at the end of mating were also performed, and each was found to contain $5 \times 10^9 - 10^{10}$ cells. Careful study of this table reveals many interesting features. If the section with a dam^+ donor is considered first, it is immediately apparent that all four recipients have similar colony counts. With such a donor, the status of the *Mme*II restriction system is irrelevant, so the similarity between the CBM1 and CBM13 figures is predictable. On plate matings (see Methods 3.1.7. and Chapter 4), however, CBM22 was easily distinguishable from CBM1: so the results of the top half of Table 5-1 with respect to the *endA* allele are surprising. It may be that part of the effect seen on plate matings is due to the mutator phenotype of CBM22: but this cannot be the whole story, because control plate matings in which no donor was present showed that the growth of antibiotic resistant mutants was insufficient to form patches as dense as those obtained when donor was present. Taken alone, the filter mating result suggests that *endA* (= *Mme*I?) restriction does not operate on plasmid DNA *in vivo*: but the problem of accounting for the plate mating effect remains. One possible interpretation is that *endA* does restrict *in vivo*, and that plate matings (in which selection is applied immediately, so only a limited number of matings per recipient can take place) can detect this restriction, but that the *endA* system is unable to cope with the repeated rounds of mating encountered on a filter. This would be easy to test by carrying out much shorter (30 minutes) filter matings. Inspection of the results obtained with a dam^- donor reveal very

DONOR:	RECIP: DILN:	CBM1 w.t.	CBM13 *mmeB*	CBM22 *endA*	CBM64 *mmeB endA*
	0	CONF.	CONF.	CONF.	CONF.
CB12	-2	CONF.	CONF.	CONF.	CONF.
dam$^+$	-4	150	33	125	96
	-6	2	0	1	1
	0	50	CONF.	CONF.	CONF.
CB26	-2	0	1400	30	2000
dam$^-$	-4	0	5	0	28
	-6	0	0	0	0

TABLE 5-1: Results of quantitative mating experiment. DILN. = logarithm of dilution factor: CONF. = confluent growth.

clearly the effect of the *mmeB* restriction system. About a thousand times as many CBM13 cells succesfully received RP4 as CBM1 cells, and work described above has shown that most of the CBM1 R^+ recipients are likely to be spontaneous *mmeB* mutants present in the original mating population (see Results 5.1.). It is interesting to note that CBM22 was a hundredfold better recipient than CBM1. The simplest way to account for this is to repeat the hypothesis that CBM22 is a mutator, and generates *mmeB* mutants at a greater rate than CBM1. This was, of course, the rationale behind the isolation of *endA mmeB* double mutants (Results 5.2.) and Table 1-1 shows that such a mutant (CBM64) is as good a recipient as CBM13 of dam^- DNA (but no better). Again, the similarity between the figures for CBM13 and CBM64 suggest that the effect of *endA* on *in vivo* restriction is not detectable by long filter matings. Although CBM64 has not been shown to lack *Mme*II activity directly, the phenotypic difference between it and CBM22 as recipients of dam^- DNA is clear enough to be able to assign the *mmeB* description with confidence. The other two double mutants, CBM62 and CBM63, have also been shown to be as good recipients of dam^- DNA as CBM13.

5.4. Modification in endonuclease mutants.

R751-containing derivatives of CBM21 and CBM22 were constructed, and were found to be equally good donors to CBM51 ($mmeB^+$) and CBM52 ($mmeB^-$). Thus, both mutants have retained the M.*Mme*II modification system. An RP4 derivative of CBM13 (CBM14) was shown, in a similar way, to contain this system. (See also Results 5.1.)

The cognate modification system of *endA*, if it exists, was not investigated, since results described in the previous section had suggested that true *endA*

restriction might not be detectable *in vivo*.

CONCLUSION.

1. Prologue.

The amount of literature on the subject of restriction enzymes, which now (Roberts, 1983) number nearly 400, including over 90 different specificities, is extensive. Of these, the vast majority are Type II enzymes, reflecting the current trend in molecular biology towards research likely to yield practical applications in biotechnology. A typical paper in this area, therefore, will describe purification procedures, optimum assay conditions, recognition sequence and point of cleavage of a restriction enzyme with little or no investigation of biological aspects, such as *in vivo* activity. Indeed, so finely focused is the general approach that now, many bacterial species are known only by specific name, culture conditions and the restriction enzyme(s) they produce. This parsimony, however, is amply justified by the fact that the ready availability of a large number of distinct well-characterized restriction enzymes has made possible the investigation of molecular biological phenomena not accessible by classical techniques. For example, much more has been learnt about eukaryotic gene organization in the last ten years than in all previous history; a feat which would have been impossible without restriction enzymes.

Because of the general narrowness of research into restriction enzymes, our detailed knowledge of the operation of restriction and modification systems *in vivo* is largely confined to work on *E. coli* enzymes including *Eco*K, RI, RII, P1 and P15. (See Introduction for references.) With a few exceptions, restriction *in vivo* has not even been demonstrated in most other species from which restriction enzymes have been isolated. One exception here is the work

of Duncan *et al.* (1978) on *Bacillus globiggi* who were able to show that *Bgl*I and *Bgl*II (recognizing GCCNNNNNGGC and AGATCT respectively) were effective in restricting phage DNA *in vivo* by isolating mutants lacking either or both enzyme activities. Each single mutant, as expected, was more susceptible to infection by phage grown on a non-modifying host; and the double mutant was indifferent to the methylation state of the phage.

The only other species in which a detailed genetic study of restriction and modification systems has been carried out is the "type" Gram-positive organism *Bacillus subtilis*. A profusion of different R/M systems has been found in various isolates of this one species (Shibata *et al.*, 1976; Roberts, 1983), perhaps hinting at the relative mobility (and, indeed, dispensability) of the genes involved. Although phage-borne R/M systems have been found in *B. subtilis* (Noyer-Weidner *et al.*, 1981; Witmer and Franks, 1981), several chromosomal loci do exist as well. Ikawa *et al.* (1981) were able to construct (by transduction) a *B. subtilis* strain containing four R/M systems (mapping in distinct chromosomal regions) from parental strains possessing one system each. Two of the five genes investigated (*hsrR*, *hsrC*) mapped at the same site; and an $hsrR^{+}hsrC^{+}$ strain could not be constructed, suggesting that this pair of genes is allelic (cf. *Eco*K, *Eco*B). None of the systems mapped in regions containing known prophages. This example suggests that R/M systems (like transposons) are genetically mobile in nature, and might be considered analogous to sexual isolation mechanisms found in some eukaryote species which are closely-related: i.e., two populations of the same bacterial species with differing R/M systems would not be able to efficiently exchange genetic information by chromosome mobilization, for example.

2. Purification and other biochemical aspects.

As described in Results 1. and 2., the development of large-scale purification methods for both *M. methylotrophus* restriction enzymes was curtailed when protocols sufficiently effective to yield active and substantially exonuclease-free preparations were achieved. Obviously, there is scope for considerable improvement here - especially in the purification technique for *Mme*I. It may be that, for *Mme*I, a departure from the conventional ion-exchange chromatography route would be more effective: for example, affinity chromatography with DNA-cellulose, or dye-linked agarose beads, might work better (Baksi *et al.*, 1978).

Turning now to assay conditions, Type II enzymes in general work best in slightly alkaline buffers, at physiological temperatures appropriate to the organism from which they were purified (Pohl *et al.*, 1982; Woodhead *et al.*, 1981). The critical component in the assay buffer, apart from the mandatory presence of Mg^{++} ions, is the ionic strength. Both *M. methylotrophus* enzymes conform to this criterion, being markedly inhibited by NaCl concentrations above 50mM. Why the optimum salt concentration should vary so much between enzymes of similar size and function is not clear. An assay buffer containing acetate rather than chloride ions, however, has been used for a large variety of enzymes with considerable success (O'Farrell *et al.*, 1980; N.J. Grinter, pers. comm.).

Further biochemical analysis of *M. methylotrophus* restriction enzymes would be greatly facilitated by the cloning and consequent increased expression of the genes concerned. Progress to this end has been made by the isolation of *mmeB*::Tn*5* and *endA*::Tn*5* insertion mutants. Km^R recombinants from a

"shotgun" cloning of chromosomal DNA from these mutants could easily be constructed. Given the likely close proximity of restriction and modification genes (cf. *Eco*RI: Newman *et al.*, 1981; Greene *et al.*, 1981), it is possible (in the case of *mmeB*::Tn*5*) that some clones with sufficiently long inserts would express *dam*-type modification, which could be assayed *in vivo*. Such a clone could then be used as a probe to search for the *mmeB*$^+$ gene in a suitable *M. methylotrophus* gene library. Only a few genes for chromosomally-located R/M systems have yet been cloned: one example is the *Hha*II system (Mann *et al.*, 1978). Expression of cloned *M. methylotrophus* genes in *E. coli* has been described by Grinter (1983).

3. The *endA* gene product - a DNA repair enzyme?

It seems appropriate, finally, to comment and speculate on the link, if any, between the apparent mutator phenotype of CBM22 and its lack of an endonuclease. While it is, of course, possible that each trait is caused by a separate mutation (Tn*5* insertion), the likelihood is that a single mutational event creating CBM22 has occurred. Mutator genes in *E. coli* are usually found to be involved in DNA repair pathways (Introduction 7.). Although there are dangers involved in generalising from a single mutant strain, it seems possible that the *endA* product could be an enzyme involved in a *M. methylotrophus* DNA repair system, and that its loss results in enhanced mutation rates. The involvement of an endonuclease in such a pathway would not be surprising: DNA repair requires that the duplex be broken (cleaved) at least on one strand, by a nuclease. What would be novel is that if the link between *Mme*I and *endA* could be firmly established, then, for the first time, it would mean an *in vivo* role for a restriction enzyme other than that of protection from invading DNA had been established. In circumstantial support

of this idea. Results 5.3. suggest that *endA* restriction against foreign DNA *in vivo* is ineffectual, in contrast to that of *mmeB*. It is tempting to speculate that the main role of some restriction enzymes (perhaps confined to those which, like *Mme*I, recognize non-palindromic sequences) is as part of DNA repair systems. A necessary first step in investigating this possibility is the isolation of more *endA* mutants in order to verify that the mutator phenotype is always linked to the loss of an endonuclease. Also, it needs to be firmly established that the endonuclease missing from CBM22 is, indeed, *Mme*I.

4. Potential applications of restrictionless mutants.

The work described above demonstrates that MmeII activity *in vivo* is a formidable barrier to transfer of conjugative plasmids from *dam*⁻ *E.coli* strains to *M.methylotrophus*. Mutations in the *mmeB* gene encoding the MmeII enzyme alleviate the restriction so that transfer frequencies from methylating and non-methylating donors are indistinguishable (Results 5.3.). The effect of mutations in the putative gene (*endA*) encoding MmeI on *in vivo* restriction is less clear and certainly less striking, and so will not be considered further. Restrictionless (*mmeB*) mutants are potentially useful in the future development of this organism as a host for cloning experiments (see Introduction 9.). It has already been demonstrated, in a very simple experiment not described above, that *P.aeruginosa* (RP4) is as proficient a donor of RP4 as *E.coli* (RP4) to CMB13 (*mmeB*); whereas transfer from the pseudomonad to wild-type *M.methylotrophus* is undetectable in plate matings. It is therefore possible to transfer plasmids directly from *P.aeruginosa* to *M.methylotrophus* using a *mmeB* recipient, thus avoiding the use of *E.coli* as an intermediate host. Clearly, this will also be true of any gram-negative species lacking a *dam*-type methylase which may be required to act as a donor of useful, plasmid-borne genetic material to *M.methylotrophus*.

Indeed, one potential use of *mmeB* mutants is as a probe for the detection of *dam*-like methylation systems in other genera. All that is needed is to introduce a broad host range plasmid (such as RP4 or R751) into the species of interest, and to test for transfer efficiency into both wild-type *M.methylotrophus* and a *mmeB* mutant. Those species with a *dam*-lik

methylase will transfer equally well to both methylotroph strains, but those (like P.aeruginosa) lacking one will only transfer well to a mmeB strain. Since no gram-negative species has yet been described in which RP4 is not maintained, this method should have wide application.

In the search for a transformation protocol for M.methylotrophus (D. Pioli, pers.comm.), the plasmid DNA (R300B) being tested is isolated from the methylotroph, and is therefore already protected from endogenous endonucleases. If such a protocol is discovered, then the use of restrictionless mutants as hosts will be of even greater importance for M.methylotrophus than it is for E.coli, since MmeII sites are far commoner than EcoK sites, and heterologous (e.g. eukaryotic) DNA cloned into methylotroph vectors will almost certainly contain some GATC sequences, which are substrates for MmeII. Although the lack of a transducing phage, or well-characterized Hfr-equivalent of M.methylotrophus is a considerable obstacle to the construction of multiply-marked derivatives, the method described in Results 5.1. allows the easy isolation of spontaneous mmeB mutants of existing strains. Restrictionless mutants appear to be as viable as the wild-type, so the routine introduction of mmeB alleles into important M.methylotrophus strains is recommended. G. Sharpe, who is developing vectors for M.methylotrophus (see Introduction 9.), uses CMB13 as a recipient in most experiments, and in many cases observes slightly better transfer into this strain than into wild-type even from dam$^+$ donors (pers.comm. This may be accounted for by suggesting that some dam sites in some plasmid molecules escape methylation before being mobilized into M.methylotrophus.

Finally, restrictionless mutants may well be of use in the search for phages which can multiply in M.methylotrophus. Extensive screens for phage which attack the methylotroph have been carried out (D.Byrom, pers.comm using material from sewage, soil samples etc. as phage sources, so far unsuccessfully. It is possible that DNA phages which grow on M.methylotrophus do exist, but whose DNA is umethylated and therefore subject to (especially) MmeII restriction. The normal hosts for such phages are envisaged as species related to M.methylotrophus which lack the MmeII restriction system. Clearly the isolation of a generalized transducing phage like coliphage PI (Lennox, 1955) whould greatly facilitate the further genetic analysis of the methylotroph. Future phage screenings should therefore be carried out using a mmeB mutant of M.methylotrophus as host, to circumvent the potential restriction problem altogether.

REFERENCES*

ARBER W: "Host controlled modification of bacteriophage." Ann. Rev. Microbiol. (1965) 19, 365-378.

ARBER W and LINN S: "DNA modification and restriction." Ann. Rev. Biochem. (1969) 38, 467-500.

BACHMANN BJ and LOW KB: "Linkage map of *Escherichia coli* K-12, Edition 6." Microb. Revs. (1980) 44, 1-56.

BAKSi K, ROGERSON DL and RUSHiZKY GW: "Rapid, single-step purification of restriction endonucleases on Cibacron Blue F3GA-agarose." Biochemistry (1978) 17, 4136-4139.

BECK E and ZINK B: "Nucleotide sequence and genome organization of filamentous bacteriophages f1 and fd." Gene (1981) 16, 35-38.

BIRNBOIM HC and DOLY J: "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." Nuci. Acid Res. (1979) 7, 1513-1523.

BOULNOIS GJ: "Colicin Ib does not cause plasmid promoted phage infection of *Escherichia coli* K-12." Molec. Gen. Genet. (1981) 182, 508-510.

* See p109 for additional references.

BRAMMAR WJ: "Possible applications of modern genetic techniques to SCP organisms" (1981). in Microbial growth on C1 compounds (Ed. Dalton, H.), London. Heyden and son. pp 312-316.

BROOKS JE and ROBERTS RJ: "Modification profiles of bacterial genomes." Nucl. Acid Res. (1982) 10, 913-934.

BROWN NL and SMITH M: "Cleavage specificity of the restriction endonuclease isolated from *Haemophilus gallinarum* (*Hga*I)." Proc. Natl. Acad. Sci. USA (1977) 74, 3213-3216.

BROWN NL and SMITH M: "A general method for defining restriction enzyme cleavage and recognition sites." Meths. Enzymol. (1980) 65, 391-404.

CALOS MP and MILLER JH: "Transposable elements." Cell (1980) 20, 579-595.

DATTA N, HEDGES RW, SHAW EJ, SYKES RB and RICHMOND MH: "Properties of an R factor from *Pseudomonas aeruginosa*." J. Bacteriol. (1977) 108, 1244-1249.

DUGAICZYK A, BOYER HW and GOODMAN HM: "Ligation of *Eco*RI endonuclease generated DNA fragments into linear and circular structures." J. Mol. Biol. (1975), 96, 171-184.

DUGGLEBY RG, KINNS H and ROOD JI: "A computer program for determining the size of DNA restriction fragments." Anal. Biochem. (1981) 110, 49-55.

DUNCAN CH, WILSON GA and YOUNG FE: "Biochemical and genetic properties of site-specific restriction endonucleases in *Bacillus globigii*." J. Bacteriol. (1978) 134, 338-344.

DUSSOIX D and ARBER W: "Host specificity of DNA produced by *Escherichia coli*. II. Control over acceptance of DNA from infecting phage lambda." J. Mol. Biol. (1962) 5, 37-49.

FIERS W, CONTRERAS R, HAEGEMAN G, ROGIERS R, van de VOORDE A, van HEUVERSWYN H, van HERREWEGHE J, VOLCKAERT G and YSEBAERT M: "Complete nucleotide sequence of SV40 DNA." Nature (1978) 273, 113-120.

FUCHS C, ROSENVOLD EC, HONIGMAN A and SZYBALSKI W: "A simple method for identifying the palindromic sequences recognized by restriction endonucleases: the nucleotide sequence of the *Ava*II site." Gene (1978) 4, 1-23.

GELINAS RE, MYERS PA and ROBERTS RJ: "Two sequence-specific endonucleases from *Moraxella bovis*." J. Mol. Biol. (1977) 114, 169-179.

GEORGE J, BLAKESLEY RW and CHIRIKJIAN JG: "Sequence specific endonuclease *Bam*HI. Effect of hydrophobic reagents on sequence recognition and catalysis." J. Biol. Chem. (1980) 255, 6521-6524.

GLICKMAN BW and RADMAN M: "*Escherichia coli* mutator mutants deficient in methylation-instructed DNA mismatch correction." Proc. Natl. Acad. Sci. USA (1980) 77, 1063-1067.

GLICKMAN B, van den ELSEN P and RADMAN M: "Induced mutagenesis in *dam*$^-$ mutants of *E. coli*: a role for 6-methyladenine residues in mutation avoidance." Molec. Gen. Genet. (1978) 163, 307-312.

GOMEZ-EICHELMANN MC and LARK KG: "Endo R.*Dpn*I restriction of *E. coli* DNA synthesized *in vitro*. Evidence that the ends of Okazaki pieces are determined by template deoxynucleotide sequence." J. Mol. Biol. (1977) 117, 621-635.

GREENE PJ, GUPTA M, BOYER HW, BROWN WE and ROSENBERG JM: "Sequence analysis of the DNA encoding the *Eco*RI endonuclease and methylase." J. Biol. Chem. (1981) 256, 2143-2153.

GREENE PJ, HEYNEKER HL, BOLIVAR F, RODRIGUEZ RL, BETLACH MC, COVARRUBIAS AA, BACKMAN F, RUSSEL DJ, TAIT R and BOYER HW: "A general method for the purification of restriction enzymes." Nucl. Acid Res. (1978) 5, 2373-2380.

GRINTER NJ: "A broad-host-range cloning vector transposable to various replicons." Gene (1983) 21, 133-143.

HATTMAN S, BROOKS JE and MASUREKAR M: "Sequence specificity of the P1 modification methylase (M.*Eco*P1) and the DNA methylase (M.*Eco dam*) controlled by the *E. coli dam* gene." J. Mol. Biol. (1978) 126, 367-380.

HEILMANN H, BURKARDT H-J, PUHLER A and REEVE JN: "Transposon mutagenesis of the gene encoding the bacteriophage P1 restriction endonuclease." J. Mol. Biol. (1980) 144, 387-396.

HOLLOWAY BW: "The application of *Pseudomonad*-based genetics to methylotrophs" (1981). In Microbial growth on C1 compounds (Ed. Dalton, H.), London. Heyden and son. pp 317-324.

IKAWA S, SHIBATA T, MATSUMOTO K, IIJIMA T, SAITO H and ANDO T: "Chromosomal loci of genes controlling site-specific restriction endonucleases of *Bacillus subtilis*." Molec. Gen. Genet. (1981) 183, 1-6.

ISH-HOROWICZ D and BURKE JF: "Rapid and efficient cosmid cloning." Nucl. Acid Res. (1981) 9, 2989-2998.

JAY E, BAMBARA R, PADMANHABAN R and WU R: "DNA sequence analysis: a general, simple and rapid method for sequencing large oligodeoxyribonucleotide fragments by mapping." Nucl. Acid Res. (1974) 1, 331-371.

JEFFREYS AJ and FLAVELL RA: "The rabbit β-globin gene contains a large insert in the coding sequence." Cell (1977) 12, 1097-1108.

JOBANPUTRA RS and DATTA N: "Trimethoprim R factors in enterobacteria from clinical specimens." J. Med. Microbiol. (1974) 7, 169-178.

KELLY TJ and SMITH HO: "A restriction enzyme from *Haemophilus influenzae*. II. Base sequence of the recognition site." J. Mol. Biol. (1970) 51, 393-409.

KLEID DG: "Purification and properties of the *Hph*I endonuclease." Meths. Enzymol. (1980) 65, 163-166.

LACKS SA: "Purification and properties of the complementary endonucleases *Dpn*I and *Dpn*II." Meths. Enzymol. (1980) 65, 138-146.

LENNOX ES: "Transduction of linked genetic characters of the host by bacteriophage P1." Virol. (1955), 1, 190-206.

LOENEN WAM and BRAMMAR WJ: "A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes." Gene (1980) 20, 249-259.

MANN MB, RAO RN and SMITH HO: "Cloning of restriction and modification systems in *E. coli*: the *Hha*II system from *Haemophilus haemolyticus*." Gene (1978) 3, 97-112.

MARINUS MG: "Location of DNA methylation genes on the *E. coli* K-12 genetic map." Molec. Gen. Genet. (1973) 127, 47-55.

MARINUS MG and MORRIS NR: "Isolation of DNA methylase mutants of *E. coli* K-12." J. Bacteriol. (1973) 114, 1143-1150.

MARINUS MG and MORRIS NR: "Pleiotropic effects of a DNA adenine methylation mutation (*dam*-3) in *E. coli* K-12." Mut. Res. (1975) 28, 15-26.

MAXAM AM and GILBERT W: "A new method for sequencing DNA." Proc. Natl. Acad. Sci. USA (1977) 74, 560-564.

McGRAW BR and MARINUS MG: "Isolation and characterization of Dam^+revertant s and suppressor mutations that modify secondary phenotypes of *dam*-3 strains

of *Escherichia coli* K-12." Molec. Gen. Genet. (1980) 178, 309-315.

MESELSON M and YUAN R: "DNA restriction enzyme from *E. coli*." Nature (1968) 217, 1110-1114.

MESSING J: "A multipurpose cloning system based on the single-stranded DNA bacteriophage M13." Recombinant DNA Tech. Bull. (1979) 2, 43-48.

MESSING J, CREA R and SEEBURG PH: "A system for shotgun DNA sequencing." Nucl. Acid Res. (1981) 9, 309-321.

MILLER JH: "Experiments in Molecular Genetics." (1972) (Cold Spring Harbor Laboratory, New York)

MODRICH P and GEIER GE: "Recognition sequence of the *dam* methylase of *E. coli* K-12 and mode of cleavage of *Dpn*I endonuclease." J. Biol. Chem. (1979) 254, 1408-1413.

MOLEMANS F, van EMMELO J and FIERS W: "The sequence specificity of endonucleases *Cau*I and *Cau*II isolated from *Chloroflexus aurantiacus*." Gene (1982) 18, 93-96.

MURRAY NE, BRAMMAR WJ and MURRAY K: "Lambdoid phages that simplify the recovery of *in vitro* recombinants." Molec. Gen. Genet. (1977) 150, 53-61.

NEWMAN AK, RUBIN RA, KIM S-H and MODRICH P: "DNA sequences of structural genes for *Eco*RI DNA restriction and modification enzymes." J. Biol. Chem. (1981) 256, 2131-2139.

O'FARRELL PH, KUTTER E and NAKANISHI M: "A restriction map of the bacteriophage T4 genome." Molec. Gen. Genet. (1980) 179, 421-435.

PIEKAROWICZ A and BRZEZINSKI R: "Cleavage and methylation of DNA by the restriction endonuclease *Hin*fIII isolated from *Haemophilus influenzae* Rf." J. Mol. Biol. (1980) 144, 415-429.

PIEKAROWICZ A: "*Hin*eI is an isoschizomer of *Hin*fIII restriction endonuclease." J. Mol. Biol. (1982) 157, 373-381.

POHL FM, THOMAE R and KARST A: "Temperature dependence of the activity of DNA-modifying enzymes: endonucleases and DNA ligase." Eur. J. Biochem. (1982) 123, 141-152.

PRIBNOW D, SIGURDSON DC, GOLD L, SINGER BS, NAPOLI C, BROSIUS J, DULL TJ and NOLLER HF: "rII cistrons of bacteriophage T4. DNA sequence around the intercistronic divide and positions of genetic landmarks." J. Mol. Biol. (1981) 149, 337-376.

REISER J and YUAN R: "Purification and properties of the P15 specific restriction endonuclease from *E. coli*." J. Biol. Chem. (1977) 252, 451-456.

ROBERTS RJ: "Restriction and modification enzymes and their recognition sequences." Nucl. Acid Res. (1983) 11, r135-r167.

ROY PH and SMITH HO: "DNA methylases of *Haemophilus influenzae* Rd. I. Purification and properties. ii. Partial recognition site base sequences." J. Mol. Biol. (1973) 81, 427-459.

SANGER F, COULSON AR, FRIEDMANN T, AIR GM, BARRELL BG, BROWN NL, FIDDES JC, HUTCHISON CA, SLOCOMBE PM and SMITH M: "The nucleotide sequence of bacteriophage ΦX174." J. Mol. Biol. (1978) 125, 225-246.

SANGER F, NICKLEN S and COULSON AR: "DNA sequencing with chain termination inhibitors." Proc. Natl. Acad. Sci. USA (1977) 74, 5463-5467.

SCHAFFER HE and SEDEROFF RR: "improved estimation of DNA fragment lengths from agarose gels." Anal. Biochem. (1981) 115, 113-122.

SCHROEDER JL and BLATTNER FR: "Least-squares method for restriction mapping." Gene (1978) 4, 167-174.

SHARP PA, SUGDEN B and SAMBROOK J: "Detection of two restriction endonuclease activities in *Haemophilus parainfluenzae* using analytical agarose-ethidium bromide electrophoresis." Biochemistry (1973) 12, 3055-3063.

SHIBATA T, IKAWA S, KIM C and ANDO T: "Site-specific DNAses in *Bacillus subtilis* and other *Bacillus* strains." J. Bacteriol. (1976) 128, 473-476.

SMITH HO and NATHANS D: "A suggested nomenclature for bacterial host modification and restriction systems and their enzymes." J. Mol. Biol. (1973) 81, 419-423.

SMITH HO and WILCOX KW: "A restriction enzyme from *Haemophilus influenzae*. I. Purification and general properties." J. Mol. Biol. (1970) 51, 379-391.

SOUTHERN EM: "Measurement of DNA lengths by gel electrophoresis." Anal. Biochem. (1979) 100, 319-323.

STOBBERINGH EE, SCHIPHOF R and SUSSENBACH JS: "Occurrence of a class II restriction endonuclease in *Staphylococcus aureus*." J. Bacteriol. (1977) 131, 645-649.

SUGIMOTO K, SUGISAKI H, OKAMOTO T and TAKANAMI M: "Nucleotide sequence of bacteriophage fd DNA." Nucl. Acid Res. (1978) 5, 4495-4503.

SUGISAKI H: "Recognition sequence of a restriction endonuclease from *Haemophilus gallinarum*." Gene (1978) 3, 17-28.

SUTCLIFFE JG: "Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322." Cold Spring Harb. Symp. Quant. Biol. (1978) 43, 77-90.

TIKCHONENKO TI, KARAMOV EV, ZAVIZION BA and NARODITSKY BS: "*Eco*RI* activity: enzyme modification or activation of accompanying endonuclease?" Gene (1978) 4, 195-212.

TWIGG AJ and SHERRATT DJ: "Trans-complementable copy-number mutants of plasmid ColE1." Nature (1980) 283, 216-218.

van WEZENBEEK PMGF, HULSEBOS TJM and SCHOENMAKERS JGG: "Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: comparison with phage fd." Gene (1980) 11, 129-148.

VOLCKAERT G and FIERS W: "A micromethod for base analysis of ^{32}P-labelled

oligoribonucleotides." Anal. Biochem. (1977) 83, 222-227.

WAGNER R and MESELSON M: "Repair tracts in mismatched DNA heteroduplexes." Proc. Natl. Acad. Sci. USA (1976) 73, 4135-4139.

WALKER DH and ANDERSON TF: "Morphological variants of coliphage P1" J. Virol. (1970) 5, 765-782.

WITMER H and FRANKS M: "Restriction and modification of bacteriophage SP10 DNA by *Bacillus subtilis* Marburg 168: stabilization of SP10 DNA in restricting hosts preinfected with a heterologous phage SP18." J. Virol. (1981) 37, 148-155.

WOODHEAD JL, BHAVE N and MALCOLM ADB: "Cation dependence of restriction endonuclease *Eco*RI activity." Eur. J. Biochem. (1981) 115, 293-296.

YUAN R, HAMILTON DL, HADI SM and BICKLE TA: "Role of ATP in the cleavage mechanism of the *Eco*P15 restriction endonuclease." J. Mol. Biol. (1980) 144, 501-519.

ADDITIONAL REFERENCES.

CHANG S and COHEN SN: "In vivo site-specific genetic recombination promoted by the EcoRI restriction endonuclease." Proc.Natl.Acad.Sci. USA (1977) 74, 4811-4815.

delGIUDICE L: "Method for isolating restriction and modificationless mutants of E.coli K-12." J.Bacteriol. (1979) 137, 673-676.

LACKS S and GREENBERG B: "Complementary specificity of restriction endonucleases of Diplococcus pneumoniae with respect of DNA methylation". J.Mol.Biol. (1977) 114, 153-168.

MOORE AT, NAYUDU M and HOLLOWAY BW: "Genetic mapping in Methylophilus methylotrophus ASI." J.Gen.Microbiol (1983) 129, 785-799.

MURRAY K, HUGHES SG, BROWN JS and BRUCE SA: "Isolation and characterization of two sequence-specific endonucleases from Anabaena variabilis." Biochem.J. (1976) 159, 317-322.

STUDIER PW: "Gene 0.3 of bacteriophage T7 acts to overcome the DNA restriction system of the host." J.Mol.Biol. (1975) 94, 283-295.

WINDASS JD, WORSEY MJ, PIOLI EM, PIOLI D, BARTH PT, ATHERTON KT, DART EC, BYROM D, POWELL K and SENIOR PJ: "Improved conversion of methanol to single-cell protein by Methylophilus methylotrophus." Nature (1980) 287, 396-401.

Appendix 1-1: Program RECOG listing.

```
00001          REM ****************************************************************
00002          REM *                                                              *
00003          REM * PROGRAM WRITTEN BY CHRIS BOYD, DEPT. OF BIOCHEMISTRY,       *
00004          REM * UNIVERSITY OF LEICESTER, 1979-1982                          *
00005          REM *                                                              *
00006          REM * ALL RIGHTS RESERVED.  NO COPYING (OUTSIDE LEICESTER) WITHOUT *
00007          REM * PERMISSION.  NO LIABILITY ACCEPTED FOR ERRORS ARISING FROM   *
00008          REM * USE OF THIS PROGRAM.                                        *
00009          REM *                                                              *
00010          REM ****************************************************************
00011          REM
00012          REM
00013          REM *** PROGRAM <<< RECOG >>>
00014          REM
00015          REM *** DETECTS 4/5/6 BP PALINDROMIC REGIONS
00016          REM     IN A DEFINED DNA SEQUENCE
00017          REM
00018          REM     INPUT IS ON FILE TAPE1 WHICH HAS FORMAT (E.G.),
00019          REM          AAAAAA<CR>AAAAAC<CR>...ETC
00020          REM
00021          REM     SEQUENCES ARE DEALT WITH IN THE ORDER
00022          REM      TETRA-, PENTA-, AND HEXAMERS
00023          REM
00024              DIM C0(200),L0(200),C1(50),C2(50)
00025          REM
00026          REM *** FILES USED
00027          REM
00028              FILE £2="TAPE1"
00029              RESTORE £2
00030          REM
00031              FILE £20="TAPE20"
00032              RESTORE £20
00033          REM
00034              FILE £3="ASYM"                    'STORES IRREGULAR SEQUENCES
00035              RESTORE £3
00036          REM
00037              FILE £4="RESULTS"
00038              RESTORE £4
00039          REM
00040          REM *** REAL CONSTANTS
00041          REM
00042              N1=1
00043          REM
00044          REM *** STRING CONSTANTS
00045          REM
00046              B8$="ACX19YGT"                    'FOR COMPLEMENTATION ROUTINE
00047              L8=LEN(B8$)                       'ALSO FOR FNC$
00048          REM
00049          REM *** WHICH PASS ?
00050          REM
```

```
00051            PRINT "PASS 1, 2, 3 OR 4?"
00052            INPUT A0
00053            ON A0 GOTO 00100,00600,02100,03400
00054            PRINT "ABORTED"
00055            STOP                                'A0<1 OR A0>4 - RUN ABORTED
00056        REM ------------------------------------------------------------
00057        REM
00058        REM *** PASS 1 ***
00059        REM
00060        REM *** THIS SECTION CREATES THE HEXAMER FILE FROM DNA SEQUENCE
00061        REM     ON FILE "DNA1" - OUTPUT IS "TAPE1".
00062        REM
00063    00100   FILE £1="DNA1"
00064            RESTORE £1
00065        REM
00066        REM *** INPUT SEQUENCE AND FORM FILE OF ALL HEXAMERS
00067        REM
00068            LO$=""                              'CLEAR SEQUENCE STRING
00069            PRINT "HOW MANY BASE PAIRS IN EACH SEQUENCE (4-6)"
00070            INPUT H0                            'AND "N"AMER FLAG
00071    00200   NODATA £1, 00300
00072            INPUT £1,L$
00073            IF L$(1:1)="*" THEN 00200
00074            LO$=LO$+L$
00075            GOTO 00200                          'READ IN NEXT LINE
00076        REM
00077    00300   PRINT "DO YOU WANT TO ANALYSE SUBSEQUENCES (Y/N)"
00078            INPUT A$
00079            IF A$="N" THEN 00400                'DO TOTAL SEQUENCE
00080            IF A$<>"Y" THEN 00300
00081        REM
00082        REM *** SECTION TO OUTPUT ANALYSIS FOR EACH SUBSEQUENCE
00083        REM
00084            PRINT ""
00085            PRINT "HOW MANY SUBSEQUENCES ARE THERE"
00086            INPUT Q0
00087            PRINT £20 USING "£££££££££",Q0      'OUTPUT NUMBER TO TAPE20
00088            PRINT ""
00089        REM
00090            PRINT "HOW LONG SHOULD EACH SUBSEQUENCE BE (BP)"
00091            INPUT W1
00092            W1=INT(W1/2)                        'CONVERT TO HALF-LENGTH
00093            PRINT ""
00094        REM
00095            PRINT "ENTER CENTRAL COORDINATES"
00096            FOR I0=1 TO Q0
00097             INPUT L0
00098             L1$=LO$(L0-W1:L0+W1)
00099             N0=I0
00100             GOSUB 04500                        'ANALYSE SUBSEQUENCE
00101            NEXT I0
00102        REM
00103            GOTO 00500                         'FINISHED
00104        REM
00105        REM *** OUTPUT ANALYSIS OF TOTAL SEQUENCE
00106        REM
00107    00400   N0=1                                'DUMMY ARGUMENT
00108            L1$=LO$                            'WHOLE SEQUENCE
00109            GOSUB 04500                        'ANALYSE THE LOT
00110        REM
00111        REM *** RUN FINISHED - INFORM USER
00112        REM
00113    00500   PRINT "OUTPUT ON 'TAPE20' & 'TAPE1' - RENAME AS 'TAPE1' FOR PASS 2"
00114            PRINT ""
00115            PRINT "PASS 1 COMPLETE"
00116        REM
00117            STOP                               'PASS 1 COMPLETED
00118        REM ------------------------------------------------------------
00119        REM
00120        REM *** PASS 2 ***
00121        REM
00122        REM *** READ IN DATA AND TEST FOR PALINDROMICITY
00123        REM
00124    00600   NODATA £2, 02000
00125            INPUT £2, P6$                      'FRAGS NOT STORED IN ARRAYS PER SE
00126        REM
00127        REM *** IS THE SEQUENCE PALINDROMIC?
00128        REM     DEAL WITH TETRAMERS FIRST
00129        REM
00130            F4$=F6$(1:4)                       'EXTRACT TETRAMER FROM HEXAMER
00131            F0=0                               'IRREGULARITY DETECTION FLAG
00132            IF F4$<>FNC$(F4$) THEN 00700       'DO PENTAS IF NOT PALINDROMIC
00133            PRINT £4,"  "+F4$+","+STR$(N1,"£££££")         'ELSE OUTPUT
00134        REM
00135        REM *** DEAL WITH PENTAS - FOR NOW, ONLY SEQUENCES WITH BASE 3
00136        REM     REDUNDANCY (1,9,8 OR N) WILL BE CONSIDERED
00137        REM
00138            GOTO 00800                         'SKIP OVER IRREGULAR OUTPUT
00139    00700   PRINT £3,"  "+F4$+","+STR$(N1+2,"£££££")       'OUTPUT IRR. SEQ.
00140            PRINT £3,"  "+FNC$(F4$)+","+STR$(N1+2,"£££££") 'AND COMPLEMENT
00141    00800   F5$=F6$(1:5)
00142            F$=F5$(1:2)+F5$(4:200)
00143            IF FNC$(F$)<>F$ THEN 01000         'DO HEXAS IF NO PALINDROMICITY
00144            PRINT £4," "+F5$(1:2)+"N"+F5$(4:200)+","+STR$(N1,"£££££")
00145        REM
00146            IF F5$(3:3) <> "A" AND F5$(3:3) <> "T" THEN 00900
00147            PRINT £4," "+F5$(1:2)+"8"+F5$(4:200)+","+STR$(N1,"£££££")
00148            GOTO 01100                         'NOW DO HEXAS
00149        REM
00150    00900   PRINT £4," "+F5$(1:2)+"2"+F5$(4:200)+","+STR$(N1,"£££££")
```

```
00151            GOTO 01100                          'SKIP OVER IRREGULAR OUTPUT
00152         REM
00153    01000    PRINT £3," "+F5$+","+STR$(N1+2,"£££££")        'OUTPUT IRR. SEQ.
00154             PRINT £3," "+FNC$(F5$)+","+STR$(N1+3,"£££££")  'AND COMPLEMENT
00155         REM
00156         REM *** NOW DO HEXAS
00157         REM
00158    01100    IF F6$<>FNC$(F6$) THEN 01200        'TEST STRICT COMPLEMENTARITY
00159             FO=1                                'PALINDROMICITY DETECTED
00160             PRINT £4,F6$+","+STR$(N1,"£££££")
00161         REM
00162         REM *** ACHIEVE INTERNAL RESOLUTION BY LOOP
00163         REM
00164    01200    FOR IO=1 TO 3
00165              F$=FNZ$(F6$,IO,"","")
00166              IF F$<> FNC$(F$) THEN 01800        'TEST NEXT POSSIBILITY
00167              FO=1                               'PALINDROMICITY DETECTED!
00168         REM
00169         REM *** TEST PURINE/PYRIMIDINE PALINDROMICITY FIRST
00170         REM
00171              L$=F6$(IO:IO)                      'LEFTMOST BASE
00172              R$=F6$(7-IO:7-IO)                  'RIGHTMOST BASE
00173         REM
00174              IF NOT(((L$="A") OR (L$="G")) AND((R$="C") OR (R$="T"))) THEN 01300
00175              PRINT £4,FNZ$(F6$,IO,"X","Y")+","+STR$(N1,"£££££")
00176              GOTO 01400
00177         REM
00178    01300    IF NOT(((L$="C") OR (L$="T")) AND((R$="A") OR (R$="G"))) THEN 01400
00179             PRINT £4,FNZ$(F6$,IO,"Y","X")+","+STR$(N1,"£££££")
00180         REM
00181         REM *** TEST A/C (=1) G/T (=9) PALINDROMICITY
00182         REM
00183    01400    IF NOT(((L$="A") OR (L$="C")) AND((R$="G") OR (R$="T"))) THEN 01500
00184             PRINT £4,FNZ$(F6$,IO,"1","9")+","+STR$(N1,"£££££")
00185             GOTO 01600
00186         REM
00187    01500    IF NOT(((L$="G") OR (L$="T")) AND((R$="A") OR (R$="C"))) THEN 01600
00188             PRINT £4,FNZ$(F6$,IO,"9","1")+","+STR$(N1,"£££££")
00189         REM
00190         REM *** TEST A/T (=8) G/C (=2) PALINDROMICITY
00191         REM
00192    01600    IF NOT(((L$="A") OR (L$="T")) AND((R$="A") OR (R$="T"))) THEN 01700
00193             PRINT £4,FNZ$(F6$,IO,"8","8")+","+STR$(N1,"£££££")
00194             GOTO 01800
00195         REM
00196    01700    IF NOT(((L$="G") OR (L$="C")) AND((R$="G") OR (R$="C"))) THEN 01800
00197             PRINT £4,FNZ$(F6$,IO,"2","2")+","+STR$(N1,"£££££")
00198    01800    NEXT IO
00199         REM
00200            IF FO=1 THEN 01900
```

AXXRECY USER: BCA *** BASIC FORMATTER *** BASIC 3.4 80209 83/08/27.

```
00201            PRINT £3,F6$+","+STR$(N1+3,"£££££")          'OUTPUT IRR. SEQ.
00202            PRINT £3,FNC$(F6$)+","+STR$(N1+3,"£££££")    'AND COMPLEMENT
00203         REM
00204    01900    N1=N1+1
00205             GOTO 00600
00206    02000    RESTORE £4
00207         REM
00208         REM *** PASS 2 COMPLETE - INFORM USER
00209         REM
00210             PRINT "IRREGULAR DATA ON FILE 'ASYM', PALINDROMIC DATA ON 'RESULTS'"
00211             PRINT "- SORT EITHER INTO 'TAPE1' FOR PASS 3 ENTRY"
00212             PRINT ""
00213             PRINT "PASS 2 COMPLETE"
00214             STOP                                 'PASS 2 COMPLETED
00215         REM ------------------------------------------------------------------
00216         REM
00217         REM *** PASS 3 ***
00218         REM
00219         REM *** IN THIS SECTION, STATISTICS ARE
00220         REM     GENERATED FROM THE SORTED SITE DATA FROM PASS 1 - THERE
00221         REM     ARE TWO OUTPUT FILES, £6="SITES" & £7="NUMBERS":
00222         REM     INPUT IS ON £2="TAPE1"
00223         REM
00224         REM *** DECLARE FILES
00225         REM
00226    02100    FILE £6="SITES"                      'STORES SEQUENCES & COORDINATES
00227             RESTORE £6
00228         REM
00229             FILE £7="NUMBERS"                    'FOR NO. CUTS,LEN(LONGEST),4/5/6
00230             RESTORE £7
00231         REM
00232             FILE £8="WINDOWS"                    '*** FOR SELECTED SIZE RANGE
00233             RESTORE £8
00234         REM
00235             PRINT "TYPE IN WINDOW BOUNDS X1,X2 - (LOW VALUE, HIGH VALUE)"
00236             INPUT X1,X2
00237             T1=0                                 'COUNT FOR KNOWN ENZYMES SITES
00238             PRINT "TYPE IN ADDITIONAL CUT SITES (0 TO EXIT)"
00239         REM
00240    02200    T1=T1+1
00241             INPUT C1(T1)
00242             C2(T1)=C1(T1)
00243             IF C1(T1)<>0 THEN 02200
00244             T1=T1-1
00245         REM
00246         REM *** INPUT FILE (FROM TAPE20) = "TAPE1" ***
00247         REM
00248             PRINT "WHAT IS THE LENGTH OF THE DNA SEQUENCE (BPS)"
00249             INPUT LO
00250             NODATA £2, 03300
```

```
00251             INPUT £2,S$(1),C0(1)              'FIRST SEQUENCE & COORDINATE
00252   02300     T0=1                              'RUNNING TOTAL FOR EACH SEQUENCE
00253             W0=0                              'RESET WINDOW COUNT
00254   02400     NODATA £2, 03300
00255             T0=T0+1
00256             INPUT £2,S$,C0(T0)                'SEQUENCE & COORDINATE
00257             IF S$<>S$(1) THEN 02500           'RESET WHEN S$ CHANGES
00258             GOTO 02400                        'REPEAT INPUT OF SEQUENCES
00259         REM
00260   02500     C0=C0(T0)
00261             T0=T0+T1
00262             PRINT £6,S$(1),                   '& OUTPUT SEQUENCE
00263             FOR I1=1 TO T1
00264              C0(T0-1-T1+I1)=C1(I1)
00265             NEXT I1
00266         REM
00267             FOR I1=1 TO T0-1
00268         REM
00269         REM *** INSERT TO COPE WITH EXTRA USER SUPPLIED CUTS ***
00270         REM
00271              FOR I2=1 TO T1
00272               IF C1(I2)>C0(I1) THEN 02600     'NO ACTION BRANCH
00273               FOR I3=T0-2 TO I1 STEP -1
00274                C0(I3+1)=C0(I3)
00275               NEXT I3
00276               C0(I1)=C1(I2)
00277               C1(I2)=1E50
00278   02600       NEXT I2
00279         REM
00280         REM *** END INSERT
00281         REM
00282              PRINT £6 USING "£££££" ,C0(I1);  '& AND EACH COORDINATE
00283              IF I1/12<>INT(I1/12) THEN 02700
00284              PRINT £6," ",
00285   02700      NEXT I1
00286             MAT C1=C2
00287             PRINT £6,""
00288             PRINT £6,"","(  ";
00289         REM
00290         REM *** NOW PRINT TO FILE "NUMBERS"
00291         REM
00292             PRINT £7 USING "£££",T0-1,        'NO.OF FRAGMENTS FOR THIS SEQUENCE
00293         REM
00294         REM *** CALCULATE LONGEST FRAGMENT LENGTH
00295         REM
00296             FOR I1=1 TO T0-1
00297              IF I1<T0-1 THEN 02800
00298              L0(I1)=C0(1)+L0-C0(I1)           '*** IF A PLASMID
00299              GOTO 02900
00300         REM
00301   02800      L0(I1)=C0(I1+1)-C0(I1)           'STORE FORWARD DIFFERENCES
00302   02900      IF I1/12<>INT(I1/12) THEN 03000
00303              PRINT £6,""
00304              PRINT £6,TAB(13);
00305   03000      PRINT £6 USING "£££££",L0(I1),
00306         REM
00307         REM *** WINDOW CALCS
00308         REM
00309              IF L0(I1)>X2 OR L0(I1) <X1 THEN 03100
00310              W0=W0+1                          'INC WIND COUNT FOR THIS SEQUENCE
00311   03100      NEXT I1
00312              PRINT £6,"  )"
00313         REM
00314         REM *** WORK OUT MAXIMUM
00315         REM
00316             M1=L0(1)
00317             IF T0<=2 THEN 03200
00318         REM
00319             FOR I1=2 TO T0-1
00320              M1=MAX(L0(I1),M1)
00321             NEXT I1
00322         REM
00323         REM *** OUTPUT RESULTS
00324         REM
00325   03200     PRINT £7 USING "£££££",M1,
00326             PRINT £7," "+S$(1)
00327             PRINT £8 USING "£££W",W0,
00328             PRINT £8 USING "£££££",T0-1,
00329             PRINT £8 USING "£££££",M1,
00330             PRINT £8," "+S$(1)
00331             S$(1)=S$                          'SWITCH ANCHOR VALUES
00332             C0(1)=C0
00333             GOTO 02300                        'SEQUENCE DEALT WITH: NOW GET NEXT
00334         REM
00335         REM *** PASS 3 OVER - INFORM USER
00336         REM
00337   03300     PRINT "DATA ON FILES 'SITES', 'WINDOWS' AND 'NUMBERS'"
00338             PRINT ""
00339             PRINT "PASS 3 COMPLETE"
00340             STOP                              'PASS 3 COMPLETED
00341         REM ------------------------------------------------------------------
00342         REM
00343         REM *** PASS 4 ***
00344         REM
00345         REM *** SECTION TO PARSE MERGED, SORTED "TAPE20" FILE
00346         REM     AND TO GENERATE FILE OF REDUNDANCIES
00347         REM
00348         REM
00349         REM *** GET NUMBER OF SEQUENCES
00350         REM
```

```
00351    03400      INPUT £20,S0                          'RETRIEVE NUMBER OF SUBSEQUENCES
00352               REM
00353    03500      NODATA £20, 03900
00354               INPUT £20,A$
00355    03600      IF A$(8:8)<>"1" THEN 03500
00356               REM
00357               REM *** HAVE POTENTIAL RUN
00358               REM
00359                  A0$=A$(1:6)                        'GET ROOT SEQUENCE
00360                  FOR I0=2 TO S0
00361    03700          NODATA £20, 03900
00362                   INPUT £20,A$
00363                   A0=VAL(A$(8:8))
00364                   B$=A$(1:6)
00365                   IF B$=A0$ AND A0=I0 THEN 03800    'BRANCH IF RUN MAINTAINED
00366                   IF B$=A0$ AND A0=I0-1 THEN 03700  'LOOP IF DUPLICATE FOUND
00367                   GOTO 03600                        'EXIT LOOP - RUN FINISHED EARLY
00368               REM
00369    03800          NEXT I0
00370                   PRINT ""
00371                   PRINT RPT$("*",67)
00372                   PRINT "*";RPT$(" ",65);"*"
00373                   PRINT "*** RUN FOUND - SEQUENCE IS >>>",A0$,"<<< ***"
00374                   PRINT "*";RPT$(" ",65);"*"
00375                   PRINT RPT$("*",67)
00376                   PRINT ""
00377                   GOTO 03500
00378               REM
00379               REM *** PASS 4 COMPLETED
00380               REM
00381                  PRINT ""
00382    03900         PRINT "PASS 4 COMPLETE"
00383                  STOP                               'PASS 4 COMPLETED
00384               REM ------------------------------------------------------------------------
00385               REM
00386               REM *** FUNCTION TO OUTPUT DEGENERATE STRINGS
00387               REM
00388                  DEF FNM$(X$)
00389               REM
00390                   FOR I9=1 TO 3
00391                    X9$=X$                           'MAKE COPY OF HEXAMER
00392                    Z9$=X9$(I9:I9)
00393                    IF Z9$="A" OR Z9$="G" THEN X9$(I9:I9)="X"
00394                    IF Z9$="C" OR Z9$="T" THEN X9$(I9:I9)="Y"
00395                    PRINT £20 USING "<£££££ £",X9$,N0          'OUTPUT DEGENERATE
00396                   NEXT I9
00397               REM
00398                   FNM$=""
00399                   FNEND
00400               REM
```



```
00401               REM *** FUNCTION TO COMPLEMENT A DNA SEQUENCE
00402               REM
00403               REM     USES GLOBAL VARIABLES AS FOLLOWS -
00404               REM         B8$ = STRING OF BASE CODES
00405               REM         L8 = LEN(B8$)
00406               REM
00407                   DEF FNC$(X$)                       '>>>>> COMPLEMENTATION FUNCTION
00408                    C$=""
00409                    FOR Q7=1 TO LEN(X$)
00410                     X1$=X$(Q7:Q7)
00411                     FOR Q8=1 TO L8                   'USE STRING COMPARISON
00412                      IF X1$ <> B8$(Q8:Q8) THEN 04000
00413                      C$ =  B8$(1+L8-Q8:1+L8-Q8) +C$
00414                      GOTO 04100                      'DO NEXT
00415    04000            NEXT Q8
00416                     C$ = X1$ + C$                    'IF HERE, MUST BE "8" OR "N"
00417    04100           NEXT Q7
00418                    FNC$=C$
00419                   FNEND
00420               REM
00421               REM *** FUNCTION TO FILL ?S IN HEXAMERS
00422               REM
00423                   DEF FNZ$(X9$,I9,L9$,R9$)           '(SOURCE STR,INDEX,L. STR,R. STR)
00424                    IF I9 > 1 THEN 04200
00425                    X8$=L9$+X9$(2:5)+R9$
00426                    GOTO 04400
00427    04200           IF I9 > 2 THEN 04300
00428                    X8$=X9$(1:1)+L9$+X9$(3:3)         'LEFT HALF
00429                    X8$=X8$+X9$(4:4)+R9$+X9$(6:200)   'RIGHT HALF
00430                    GOTO 04400
00431    04300           X8$=X9$(1:2)+L9$+R9$+X9$(5:6)
00432    04400           FNZ$=X8$
00433                   FNEND
00434               REM
00435               REM *** SUBROUTINE TO OUTPUT HEXAMERS
00436               REM
00437    04500          L1$=L1$+L1$(1:H0-1)                'MAKE TERMINALLY REDUNDANT
00438                   L1=LEN(L1$)                        'LENGTH OF SEQUENCE
00439                   FOR K0=1 TO L1-H0+1
00440                    A$=L1$(K0:K0+H0-1)
00441                    A1$=A$(1:3)+A$(H0-2:H0)           'MOD FOR BGLI TYPE ENZYMES
00442               REM ** A1$=A$
00443                    PRINT £2,A1$                      'OUTPUT "N"AMER
00444                    PRINT £20 USING "<£££££ £",A1$,N0 'OUTPUT TO TAPE20
00445                  P0$=FNM$(A1$)                  'AND DEGENERATES REM !!
00446                    PRINT £20 USING "<£££££ £",FNC$(A1$),N0
00447                  P0$=FNM$(FNC$(A1$))              'AND DEGENERATES REM !!
00448                   NEXT K0
00449                   RETURN
00450               REM
```

Appendix 1-2: Example RECOG run.

```
/GET,DNA1=PBR322
/RECOG
PASS 1, 2, 3 OR 4?
? 1
HOW MANY BASE PAIRS IN EACH SEQUENCE (4-6)
? 6
DO YOU WANT TO ANALYSE SUBSEQUENCES (Y/N)
? Y

HOW MANY SUBSEQUENCES ARE THERE
? 4

HOW LONG SHOULD EACH SUBSEQUENCE BE (BP)
? 100

ENTER CENTRAL COORDINATES
? 212
? 296
? 2697
? 2874
OUTPUT ON 'TAPE20' & 'TAPE1' - RENAME AS 'TAPE1' FOR PASS 2

PASS 1 COMPLETE
    45.746 CP SECONDS EXECUTION TIME
/RENAME,TAPE1=TAPE20
RENAME,TAPE1=TAPE20.
/GET,LGODNA
/LGODNA
     5.944 CP SECONDS EXECUTION TIME.
/RECOG
PASS 1, 2, 3 OR 4?
? 4

**************************************************************************
*                                                                        *
*** RUN FOUND - SEQUENCE IS >>>              GTYGGA          <<< ***
*                                                                        *
**************************************************************************

PASS 4 COMPLETE
    11.194 CP SECONDS EXECUTION TIME
```

Appendix 2-1: Input to MAP program.

Appendix 2-2: Output from MAP program.

```
10
PST
T2
M1
M2
S1
T1
M3
M4
S2
PST
5,10,1945
1,5,2417
9,10,572
1,9,3790
6,10,1037
1,6,3325
1,2,406
2,10,3956
6,2,1443
2,6,2919
5,9,1373
9,5,2989
5,6,908
6,5,3454
9,2,978
2,9,3384
5,7,1029
5,8,1208
8,2,1155
8,3,1706
8,4,1771
3,8,2684
4,8,2593
7,3,1879
7,4,1978
3,7,2489
4,7,2391
1,4,1054
1,3,970
7,10,919
8,10,751
0,0,0
1,-752
0
```


SITE	SITE NUMBER	MAP COORDINATE	INTERVAL
PST	1	-752.0	0.0
T2	2	-345.5	406.5
M1	3	212.2	557.8
M2	4	296.4	84.1
S1	5	1669.1	1372.7
T1	6	2578.8	909.7
M3	7	2697.4	118.6
M4	8	2873.5	176.1
S2	9	3045.2	171.7
PST	10	3617.6	572.4

FRAGMENT SIZE ADJUSTMENTS

SITES		FRAGMENT		ORIGINAL	COMPUTED	CHANGE
S1	PST	5	10	1945.0	1948.5	.2%
PST	S1	1	5	2417.0	2421.1	.2%
S2	PST	9	10	572.0	572.4	.1%
PST	S2	1	9	3790.0	3797.2	.2%
T1	PST	6	10	1037.0	1038.8	.2%
PST	T1	1	6	3325.0	3330.8	.2%
PST	T2	1	2	406.0	406.5	.1%
T2	PST	2	10	3956.0	3963.1	.2%
T1	T2	6	2	1443.0	1445.3	.2%
T2	T1	2	6	2919.0	2924.3	.2%
S1	S2	5	9	1373.0	1376.1	.2%
S2	S1	9	5	2989.0	2993.5	.1%
S1	T1	5	6	908.0	909.7	.2%
T1	S1	6	5	3454.0	3459.9	.2%
S2	T2	9	2	978.0	978.9	.1%
T2	S2	2	9	3384.0	3390.7	.2%
S1	M3	5	7	1029.0	1028.4	-.1%
S1	M4	5	8	1208.0	1204.5	-.3%
M4	T2	8	2	1155.0	1150.6	-.4%
M4	M1	8	3	1706.0	1708.3	.1%
M4	M2	8	4	1771.0	1792.5	1.2%
M1	M4	3	8	2684.0	2661.3	-.9%
M2	M4	4	8	2593.0	2577.1	-.6%
M3	M1	7	3	1879.0	1884.4	.3%
M3	M2	7	4	1978.0	1968.6	-.5%
M1	M3	3	7	2489.0	2485.2	-.2%
M2	M3	4	7	2391.0	2401.0	.4%
PST	M2	1	4	1054.0	1048.4	-.5%
PST	M1	1	3	970.0	964.2	-.6%
M3	PST	7	10	919.0	920.2	.1%
M4	PST	8	10	751.0	744.1	-.9%

```
TYPE 0 TO EXIT, 1 FOR ADDITIONAL DATA
GOODBYE
```

Appendix 3-1: Example GELSIM run.

```
/CALL,AXXGRFP(DNA1=PBR322)
PROCEDURE TO PRODUCE PARTIAL DIGESTION SIMULATIONS
FIRST, SELECT ENZYME

 *** Sequence is from file "PBR322" ***


 The DNA sequence is circular and is  4362 bp's long.

  Enter "A" for total enzyme search, "S" for own sequence(s) or "0" to continue
? 0
 Enzymes currently available are :-

  1:AatII     2:AccI      3:AcyI      4:AfiII     5:AflIII    6:AhaIII    7:AluI
  8:ApaI      9:AsuI     10:AvaI     11:AvaII    12:AvaIII   13:AvrII    14:BalI
 15:BamHI    16:BbvI     17:BclI     18:BglI     19:BglII    20:BinI     21:BssHII
 22:BstXI    23:CauII    24:CfrI     25:ClaI     26:DdeI     27:EcaI     28:EcoRI
 29:EcoRI*   30:EcoRII   31:EcoRV    32:Fnu4HI   33:GdiII    34:HaeI     35:HaeII
 36:HaeIII   37:HgaI     38:HgiAI    39:HgiCI    40:HgiEII   41:HgiJII   42:HhaI
 43:HincII   44:HindIII  45:HinfI    46:HinGUII  47:HpaI     48:HpaII    49:HphI
 50:KpnI     51:MboII    52:MlaI     53:MluI     54:MmeI     55:MnlI     56:MstI
 57:NaeI     58:NarI     59:NcoI     60:NdeI     61:NruI     62:NspBII   63:NspCI
 64:PstI     65:PvuII    66:RsaI     67:RshI     68:SalI     69:SauI     70:Sau3A
 71:ScaI     72:ScrFI    73:SduI     74:SfaNI    75:SmaI     76:SnaI     77:SphI
 78:SstI     79:SstII    80:StuI     81:TaqI     82:ThaI     83:Tth111I  84:Tth111II
 85:XbaI     86:XhoI     87:XhoII    88:XmaIII   89:XmnI
 Now list enzymes required in search: use code numbers as specified in the
 above list.  For example, to look for sites cut by the first two enzymes listed
 type "1,2" in response to the question mark. (Enter "0" to exit)
? 54

 You have chosen the following enzymes -

 MmeI


 **** ENZYME: MmeI

 The sequence TCCXAC is non-symmetrical and occurs    4 times

 List of coordinates:-
     1-  TCCGAC     +  215
     2-  TCCGAC     +  302
     3-  TCCGAC     + 2682
     4-  TCCAAC     + 2866
 *************************

 List of fragment lengths (in base pairs) -

  2380  (   303- 2682)
  1711  (  2867-  215)
   184  (  2683- 2866)
    87  (   216-  302)
* Site results on files "DUMP" & "LCFILE" *
* Type "SHPRINT,LCFILE,ULC" for hard copy outpu

NOW TYPE IN SEQUENCE LENGTH -
? 4362
? E

AND SEQUENCE CONFORMATION (L OR C)
? C
? E
```

```
    EXECUTING AXXPART...
     There are  13  partial products                (C)

    SORTED PARTIALS DATA ON FILE TAPE20
    END OF PART1
    PROCEDURE FINISHED
    *
    /LIST,F=TAPE20
       87,!
      184,!
     1711,!
     1798,!
     1895,!
     1982,!
     2380,!
     2467,!
     2564,!
     2651,!
     4178,!
     4275,!
     4362,!
    EOI ENCOUNTERED.
    /
    /
    /
    /
    /GET,PBRMME
    /LIST,F=PBRMME
    48.139712, 22.9704, -0.853018
    1
    pBR322
    MmeI
       87,   184,  1711,  1798,  1895,  1982,  2380,  2467,  2564,  2651,
     4178,  4275,  4362,     0,
    EOI ENCOUNTERED.
    /
    /
    /
    /CALL,AXXGRFP,S=2PART(TRACKS=PBRMME)

    INPUT FILE HAS DATA FOR  1 SIMULATION(S)

    *
    /EXEC,CCPROC

08.00.46.UCLP, AA22,      0.252KLNS.
```