J. Mol. Biol. (1994) **236**, 1011–1021

The Domains of a Type I DNA Methyltransferase

Interactions and Role in Recognition of DNA Methylation

Laurie P. Cooper and David T. F. Dryden†

Institute of Cell and Molecular Biology
The University of Edinburgh, Darwin Building
Mayfield Road, Edinburgh EH9 3JR, U.K.

The DNA methyltransferases of type I restriction-modification systems are trimeric enzymes composed of one DNA specificity (S) subunit and two modification (M) subunits. The S subunit contains two large regions, each of which recognizes one part of the split, asymmetrical DNA target sequence. Each M subunit contains an amino acid motif for binding the methyl group donor and cofactor, *S*-adenosyl methionine. The *Eco*KI methyltransferase has a strong preference for methylating a hemimethylated DNA target rather than an unmodified target. We have used partial proteolytic digestion of *Eco*KI methyltransferase to generate polypeptide domains that we have identified by amino acid sequencing. The S subunit was cut into two large, folded domains each containing one DNA binding region. Binding of DNA partially protected the S subunit from digestion. The M subunit was also cut into two large domains joined together by a short flexible loop, and a C-terminal tail region. The short loop contained part of the *S*-adenosyl methionine binding motif, and cofactor binding protected the loop and the two large domains from proteolysis. The C-terminal domain of M remained associated with the N-terminal domain of the S subunit even after the rest of the protein had been digested. The conformation of the tail region of the M subunit was sensitive to the methylation state of DNA in ternary complexes also containing *S*-adenosyl methionine, and could differentiate between unmethylated and hemimethylated DNA substrates.

Keywords: methyltransferase; *S*-adenosylmethionine; limited proteolysis; DNA methylation

1. Introduction

The type I restriction and modification (R/M‡) systems combine the functions of site-specific DNA methylation and cutting in one large, oligomeric enzyme (Yuan, 1981; Wilson & Murray, 1991; Bickle & Kruger, 1993; Heitman, 1993). In most other R/M systems these functions are found in separate enzymes. The particular activity manifested by a type I system largely depends on the presence or absence of N6 methylated adenine nucleotide bases in the asymmetrical, bipartite DNA target sequence. Type I enzymes have been divided into three families, IA, IB and IC, by a variety of empirical tests (Murray *et al.*, 1982) but they are all formed from the same three types of subunit, each of which has a specific function. The S, M and R subunits are responsible for DNA target specificity (S), DNA methylation (M) and DNA cutting (R), and are the products of the *hsd S*, *M* and *R* genes.

The S subunits include two large regions of 150 to 180 amino acid residues, each of which binds specifically to one part of the DNA target sequence. Short sequences close to the DNA recognition regions are believed to have a structural role in both subunit assembly and in determining the length of the non-specific DNA spacer between the two parts of the DNA target (Figure 1(a): Gough & Murray, 1983; Argos, 1985; Fuller-Pace & Murray, 1986; Gann *et al.*, 1987; Cowan *et al.*, 1989; Kannan *et al.*, 1989; Price *et al.*, 1989; Gubler & Bickle, 1991; Gubler *et al.*, 1992). The DNA recognition regions of different type I systems have very little sequence similarity unless they recognize a common target, but the smaller structural regions are highly conserved within a type I family.

† Author to whom all correspondence should be addressed.

‡ Abbreviations used: R, restriction; M, modification; S, specificity; AdoHcy, *S*-adenosyl homocysteine; AdoMet, *S*-adenosyl methionine; Mes, 2-[*N*-morpholino] ethanesulphonic acid; mtase, methyltransferase; PVDF, polyvinylidene difluoride.

0022–2836/94/091011–11 $08.00/0

© 1994 Academic Press Limited

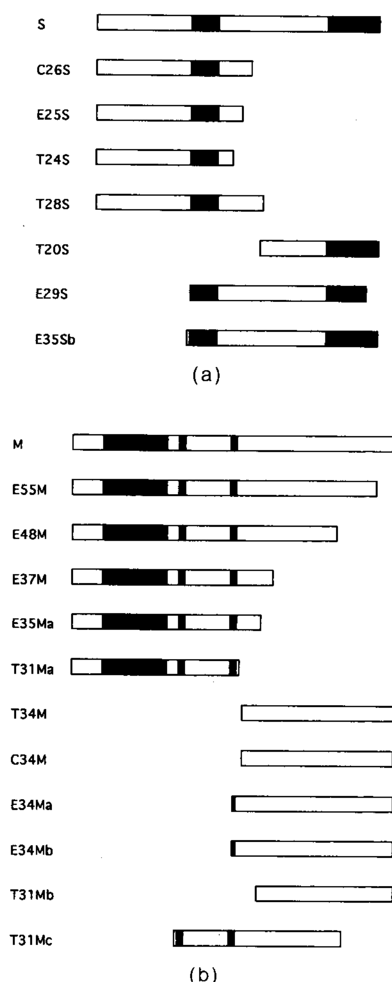

Figure 1. (a) A map of the S subunit (463 amino acid residues) showing the variable regions from amino acid residues 2 to 158 and 202 to 377, which recognize the DNA target, and the central and carboxyl conserved regions (shaded) from residues 159 to 201 and 378 to 464. Conserved regions are similar in all type IA enzymes and are thought to play a structural role. Below the S subunit are aligned all the sequenced proteolytic fragments of S subunit described in this work. (b) A map of the M subunit (529 amino acid residues) showing the m* region from residues 48 to 153, region I from 173 to 179 and region II from 262 to 269 as shaded blocks. Regions I and II play a role in AdoMet binding and catalysis. Region II contains the sequence NPPF from residues 266 to 269 and comprises the D/N PP Y/F consensus sequence found in all N(6) adenine mtases. Below the M subunit are aligned all the sequenced fragments of the M subunit.

The M subunits contain two conserved amino acid elements believed to be involved in binding of the cofactor and methyl group donor, *S*-adenosyl methionine, and the catalysis of DNA methylation (Figure 1(b): Hattman *et al.*, 1985; Vershon *et al.*, 1985; Loenen *et al.*, 1987; Chandrasegaran & Smith, 1988; Klimasauskas *et al.*, 1989; Sharp *et al.*, 1992; Guyot *et al.*, 1993; Kossykh *et al.*, 1993). A series of single amino acid changes in the m* region of the M subunit of the *Escherichia coli* K12 type IA system that abolish the normally strong preference of this system for methylating hemimethylated substrates are found (Kelleher *et al.*, 1991).

The R subunits all contain amino acid sequences characteristic of ATP binding sites and DNA helicases, consistent with the possible role of this subunit in ATP dependent DNA translocation and cutting (Gorbalenya & Koonin, 1991; Murray *et al.*, 1993).

An active DNA methyltransferase can be formed from the M and S subunits that displays the same methylation properties as the complete R/M enzyme. Several of these mtases have been characterized biochemically (Suri & Bickle, 1985; Patel *et al.*, 1992; Taylor *et al.*, 1992, 1993; Dryden *et al.*, 1993; Powell *et al.*, 1993). In particular, the *Eco*KI mtase of the *E. coli* K12 type IA system, has been purified to homogeneity (Dryden *et al.*, 1993). It is a trimeric enzyme of molecular mass 169 kDa containing two M and one S subunits. The M subunits each bind one molecule of AdoMet with a dissociation constant of 3·6(±0·4) μM. Binding causes a conformational change enhancing mtase affinity for the DNA target sequence, A$\underline{A}$C $(N)_6$ G$\underline{T}$GC (Powell *et al.*, 1993). The mtase, as observed for the complete R/M enzyme, methylates a hemimethylated sequence at least 20-fold faster than an unmodified target. There is one methylatable adenine base on each strand of DNA in the target sequence at the underlined positions.

We have used limited proteolysis of the *Eco*KI mtase in the presence of its substrates to delineate the domain structure of the enzyme and to observe the effects of substrates on the protein conformation. In general, limited proteolysis preferentially cuts solvent-exposed loops and domain-linking regions to yield protein domains that have the same structure as in the intact protein and which may retain some of the function. Some examples are *lac* repressor (Geisler & Weber, 1977), λ repressor (Pabo *et al.*, 1979), thermolysin (Fontana *et al.*, 1986; Vita *et al.*, 1989) and *Eco*RI mtase (Reich *et al.*, 1991).

We have located the sites of proteolytic attack on the *Eco*KI mtase by sequencing the fragments. These sites and the resulting domains show a sensitivity to the methylation state of the DNA and are affected differently by AdoMet and inhibitor analogues, strongly suggesting their involvement in substrate recognition and conformational flexibility. Analytical gel filtration chromatography of partially digested mtase shows that some domains remain specifically associated with each other, suggesting that they form part of the intersubunit contacts in the mtase.

2. Materials and Methods

The buffer used in all proteolytic digestions was 20 mM Tris, 20 mM Mes, 200 mM NaCl, 10 mM $MgCl_2$, 7 mM β-mercaptoethanol, 0·1 mM EDTA adjusted to pH 8·0 with NaOH. The mtase was prepared as described previously (Dryden *et al.*, 1993) and was transferred into the proteolysis buffer using a PD10 desalting column (Pharmacia) and used at a concentration of 6 μm. Recent

preparations of the mtase substituted a Superdex 200 (Pharmacia) gel filtration step for the published Sephacryl S300 gel filtration step. This improved the purity of the preparation by removing the small amount of free M subunit present in earlier samples, but had no observable affect on the proteolysis results. AdoMet, AdoHcy and sinefungin were from Sigma and used at a concentration of 100 μM. DNA oligonucleotide duplexes of the sequence: 5′-TGTCTAGATATCGGCCTAACCACGTGGT-GCGTACGAGCTCAGGCG-3′; (OSWEL DNA synthesis, The University of Edinburgh) and its complementary strand were used at a duplex concentration of 9 μM, conditions that should bind all of the mtase in a 1:1 protein to DNA complex (Powell *et al.*, 1993). The hemimethylated substrate was methylated at the underlined adenine base. The concentration of the proteases (Boehringer) used was 0·2 μM, except in some gels where the concentration was 2 μM and digestion proceeded for shorter periods as indicated. The concentrations of stock solutions of the reaction components were obtained by measurements of UV absorption using absorbance coefficients (Powell *et al.*, 1993; Carrey, 1989). The reactions (150 μl total volume) were performed at 37 °C with elastase and 25 °C for trypsin and chymotrypsin. Trypsin reactions were carried out in glass tubes, and elastase and chymotrypsin reactions in plastic tubes. Portions (15 μl) were removed at various times, mixed with 15 μl of SDS sample buffer, boiled immediately and loaded onto SDS/12% polyacrylamide gels for electrophoresis according to the method of Laemmli (1970). The gels were developed using formaldehyde fixing and Coomassie blue staining (Steck *et al.*, 1980) and scanned with a Shimadzu CS-930 densitometer to give relative band intensities. Some gels were electroblotted onto Immobilon P PVDF membranes (Millipore), stained with Amido black and the N termini of selected bands sequenced (Matsudaira, 1989).

The true molecular weights and mean residue weights of the M and S subunits are 59,300, 112·1 and 51,300, 110·75, respectively, but on electrophoresis the M subunit molecular weight appeared too large by 1900, which corresponds to about 17 amino acid residues, and the S subunit appeared too small by 4400, corresponding to 40 amino acid residues. The molecular weights of the proteolysis fragments obtained from the gel measurements were corrected for this discrepancy, taking into account their molecular weight relative to the M or S subunit. In this way a minimum size for M subunit fragments and a maximum size for S subunit fragments could be obtained. Also calculated directly from the gel molecular weights was a maximum and minimum size for the M and S subunits, respectively, to give a conservative estimate of the fragment size. The proteolytic fragments were noted in the form XabY, where X was E, C or T for elastase, chymotrypsin or trypsin, ab was the approximate molecular weight of the fragment and Y was M or S, denoting the subunit origin.

Analytical gel filtration of proteolytic digests was performed by passing the reaction mixture through a Pharmacia FPLC Superose 6 gel filtration column. The column separated the small protease from the larger mtase digestion fragments almost immediately, hence stopping the digestion. Fractions collected were analysed on SDS/12% polyacrylamide gels with silver staining according to the procedure of Ansorge (1985). Glutaraldehyde crosslinking of FPLC gel filtration fractions was performed according to the procedure of Jaenicke & Rudolph (1989). The FPLC gel filtration column was calibrated using 16 proteins of known molecular weight ranging from 12,500 to 670,000; however, the assignment of molecular weight was complicated by the anomalous behaviour of the mtase and its subunits on gel filtration. The mtase (M_2S_1), an inactive M_1S_1 form and the M subunit eluted at volumes corresponding to molecular weights of 210,000, 140,000 and 85,000 instead of at 169,000, 110,000 and 59,000. This behaviour may lead to an overestimate of the molecular weights of proteolytic fragments by this method.

3. Results

The proteases degraded the M and S subunits into many discrete large fragments (Figures 2 to 4). The relative amount of these fragments changed with time. While several fragments accumulated in considerable quantities and represented stable folded protein fragments or domains, the faintest bands on the gels are likely to be relatively unstable fragments that were easily degraded further. In some instances the formation of fragments could be at least partially blocked by the presence of substrates. Some of the proteolysis sites that generated the observed fragments may have become accessible only after certain other sites had been cleaved. Therefore, not all of the observed cleavage sites need to be on the surface of the protein. The fragments most likely to represent protein domains would be those that were stable or influenced by substrate binding.

(a) *Identification of fragments*

The amino acid regions of a type IA S subunit are shown in Figure 1(a). The short conserved regions are highly conserved sequences throughout the family while the two larger regions that recognize DNA are not conserved and are referred to as variable regions. The important amino acid regions of the M subunit are shown in Figure 1(b).

The N-terminal amino acid residues of the four most stable fragments, E55M, E48M, T34M and C34M, and some of the less abundant fragments were identified (Table 1 and Figure 1). The E27 polypeptide visible in Figure 2(d) and (e) gave an N-terminal sequence not present in the mtase but that was identified as the N terminus of elastase (Hartley & Shotton, 1971: and see Table 1).

The S subunit was cut by all three proteases to give essentially two groups of fragments consisting either of the N-terminal variable region plus the central conserved region and about 25% of the carboxyl variable region, fragments T28S, T24S, C26S and E25S, or the central conserved region plus the carboxyl variable and conserved regions, fragments E29S and E35Sb (Figure 1(a)). The T28S fragment and the T20S fragment were probably the result of proteolysis at the same site, arginine 269.

Figure 1(b) and Table 1 show that the M subunit could be cut by trypsin and chymotrypsin just after amino acid region II to give T34M and C34M representing a stable C-terminal domain. Elastase cut the M subunit just at the start of region II to give fragments E34Ma and E34Mb. These C-terminal fragments of M were however not resistant to

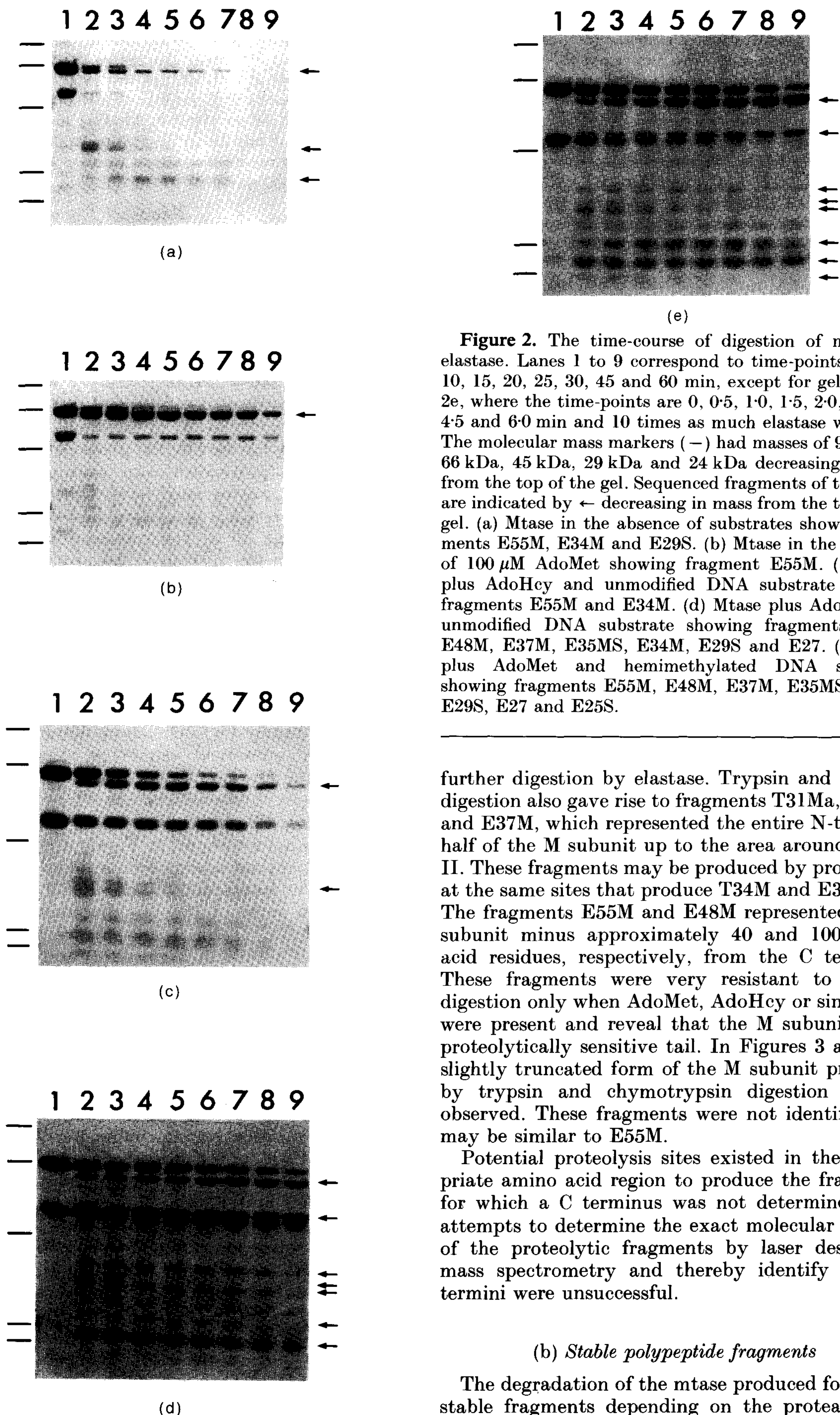

Figure 2. The time-course of digestion of mtase by elastase. Lanes 1 to 9 correspond to time-points of 0, 5, 10, 15, 20, 25, 30, 45 and 60 min, except for gels 2d and 2e, where the time-points are 0, 0·5, 1·0, 1·5, 2·0, 2·5, 3·0, 4·5 and 6·0 min and 10 times as much elastase was used. The molecular mass markers (—) had masses of 97·4 kDa, 66 kDa, 45 kDa, 29 kDa and 24 kDa decreasing in mass from the top of the gel. Sequenced fragments of the mtase are indicated by ← decreasing in mass from the top of the gel. (a) Mtase in the absence of substrates showing fragments E55M, E34M and E29S. (b) Mtase in the presence of 100 μM AdoMet showing fragment E55M. (c) Mtase plus AdoHcy and unmodified DNA substrate showing fragments E55M and E34M. (d) Mtase plus AdoMet and unmodified DNA substrate showing fragments E55M, E48M, E37M, E35MS, E34M, E29S and E27. (e) Mtase plus AdoMet and hemimethylated DNA substrate showing fragments E55M, E48M, E37M, E35MS, E34M, E29S, E27 and E25S.

further digestion by elastase. Trypsin and elastase digestion also gave rise to fragments T31Ma, E35Ma and E37M, which represented the entire N-terminal half of the M subunit up to the area around region II. These fragments may be produced by proteolysis at the same sites that produce T34M and E34Ma, b. The fragments E55M and E48M represented the M subunit minus approximately 40 and 100 amino acid residues, respectively, from the C terminus. These fragments were very resistant to further digestion only when AdoMet, AdoHcy or sinefungin were present and reveal that the M subunit has a proteolytically sensitive tail. In Figures 3 and 4, a slightly truncated form of the M subunit produced by trypsin and chymotrypsin digestion can be observed. These fragments were not identified but may be similar to E55M.

Potential proteolysis sites existed in the appropriate amino acid region to produce the fragments for which a C terminus was not determined. Our attempts to determine the exact molecular weights of the proteolytic fragments by laser desorption mass spectrometry and thereby identify their C termini were unsuccessful.

(b) *Stable polypeptide fragments*

The degradation of the mtase produced four large stable fragments depending on the protease used

Table 1
Fragments of the M and S subunit

Fragment	N-terminal sequence†	Subunit origin	Estimated mol. weight‡ ($\times 10^{-3}$)	Putative polypeptide§
E55M	MNNND	M	53·3–55·8	M1–A476 to S498
E48M	MNNND	M	45·0–48·5	M1–H410 to E433
E37M	MNNND	M	35·8–37·0	M1–G319 to D330
E35Ma	MNNND	M	33·9–35·0	M1–P302 to D312
E35Sb	(A)SFDLI	S	35·0–38·0	S151–S464
E34Ma	(V)ATNPP	M	31·9–33·0	A264–E529
E34Mb	(A)TNPPF	M	31·9–33·0	T265–E529
T34M	(R)TFVHP	M	32·3–34·2	T28S0–E529
C34M	(F)VHPTS	M	33·0–34·6	V282–E529
T31Ma	MNNND	M	30·0–31·0	M1–P267 to N276
T31Mb	(R)AAVVV	M	30·0–31·0	A306–E529
T31Mc	(R)EVVQD	M	30·0–31·0	E169–K436 to S445
E29S	(I)NIPIP	S	31·7–31·9	N156–A442 to L444
T28S	SAGKL	S	28·3–30·7	S2–S256 to E277
E27	VVGGT	—	27·0	Elastase
C26S	SAGKL	S	25·7–27·9	S2–T233 to L253
E25S	SAGKL	S	24·6–26·9	S2–F223 to K242
T24S	SAGKL	S	23·0–25·0	S2–K208 to N227
T20S	(R)FLE_S	S	19·8–21·5	F270–A449 to S464

† The amino acid code letters in parentheses are the amino acid residues preceding the cleavage point.
‡ The molecular weights have been estimated, as described in Materials and Methods, to give the broadest and most conservative estimate of the fragment's size.
§ Some of the molecular weights of fragments are larger than is feasible given the sequences of the M and S subunits.

and the substrates present. Figure 2(a) to (c) show the formation of a truncated form of the M subunit when elastase was used. This fragment, E55M, was rapidly degraded if AdoMet or its analogues were absent but was very stable when they were present. Fragments of M subunit slightly larger than E55M were also produced by trypsin or chymotrypsin digestion though it was not determined if they were similar to E55M. When DNA and AdoMet were both present the E55M fragment was further degraded to a fragment E48M (Figure 2(d) to (e)). E48M was stable when DNA and AdoMet were present but it was formed from E55M more rapidly with unmodified DNA+AdoMet than with hemimethylated DNA+AdoMet. This implied that E55M was less stable with unmodified DNA and AdoMet than with hemimethylated DNA and AdoMet. E48M was not visible and may have been formed only transiently, if at all, when AdoHcy was substituted for AdoMet (Figure 2(c)).

Two fragments, C34M and T34M, were formed under very similar conditions when the mtase alone or mtase plus DNA were degraded by trypsin or chymotrypsin (Figures 3 and 4). They appeared at the same rate as the M subunit was degraded and were stable for over an hour. When AdoMet, AdoHcy or sinefungin were present these fragments were formed in only small amounts as the now protected M subunit was more slowly degraded.

(c) *Assembly of fragments*

The possibility that some of the larger proteolytic fragments of M and S subunits remained assembled in a multimeric structure maintaining subunit–subunit contacts as found in the intact mtase was investigated by passing the products of a proteolysis reaction through an analytical gel filtration column. The digestion of the mtase by trypsin and of the mtase plus AdoMet by elastase were examined to determine if the T34M and E55M fragments were constituents of larger complexes. Figure 5 shows the pure mtase eluting as a single peak after 15·25 ml, corresponding to a molecular weight of 210,000.

The E55M fragment eluted after 17·0 ml as a single peak with an apparent molecular weight of 68,000 (Figure 5). Its monomeric nature was confirmed by glutaraldehyde crosslinking of this fraction, which showed a single species of molecular weight 55,000 on SDS-PAGE (data not shown). Small amounts of other degradation products and of intact mtase were observed in other fractions as determined by silver staining of SDS-containing gels of the fractions (not shown).

The digestion of the mtase by trypsin produced a complex elution pattern showing a broad inhomogeneous band eluting between approximately 15 and 17 ml and a single peak at 17·7 ml (Figure 5). This implied that the broad band was composed of several protein species with molecular weights between 200,000 and 68,000 whose relative proportions changed with the length of digestion. The single peak at 17·7 ml corresponded to species of approximately 30,000 molecular weight. A silver-stained gel of the fractions (Figure 6) showed that the broad, large molecular weight band was a heterogeneous mixture of species including the M subunit, a truncated M subunit, the S subunit and fragments T34M and T24S. The major components of this band were T34M and T24S, which appeared in approximately equal amounts. This strongly implied that T34M and T24S were associated speci-

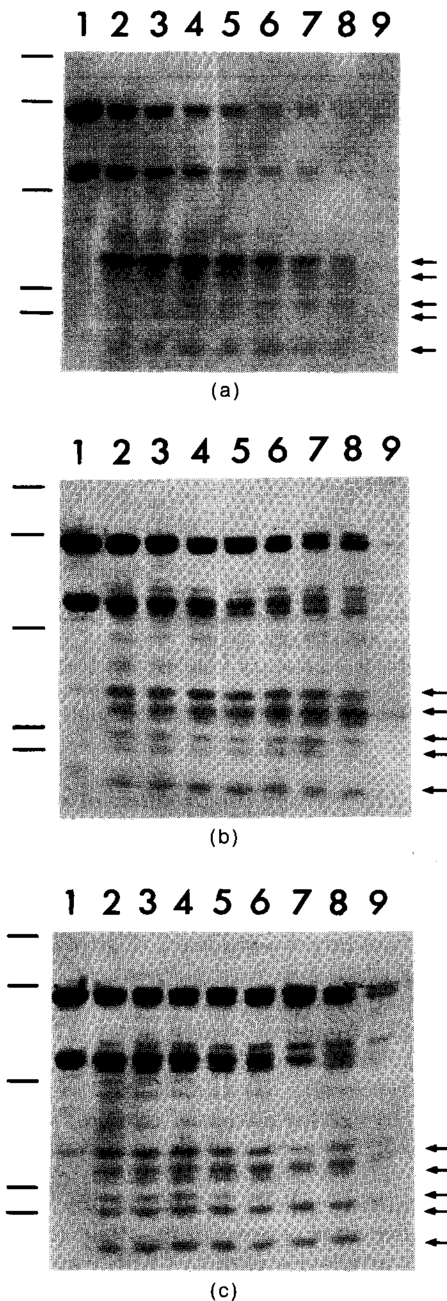

Figure 3. Mtase digestion by trypsin with the same time-points for lanes 1 to 9 and molecular mass markers as in Figure 2(a) to (c). (a) Mtase alone showing fragment T34M, T31M, T28S, T24S and T20S. (b) Mtase plus 100 μM AdoMet showing the same fragments as Figure 2(a). (c) Mtase plus AdoMet and unmodified DNA showing the same fragments as Figure 2(a).

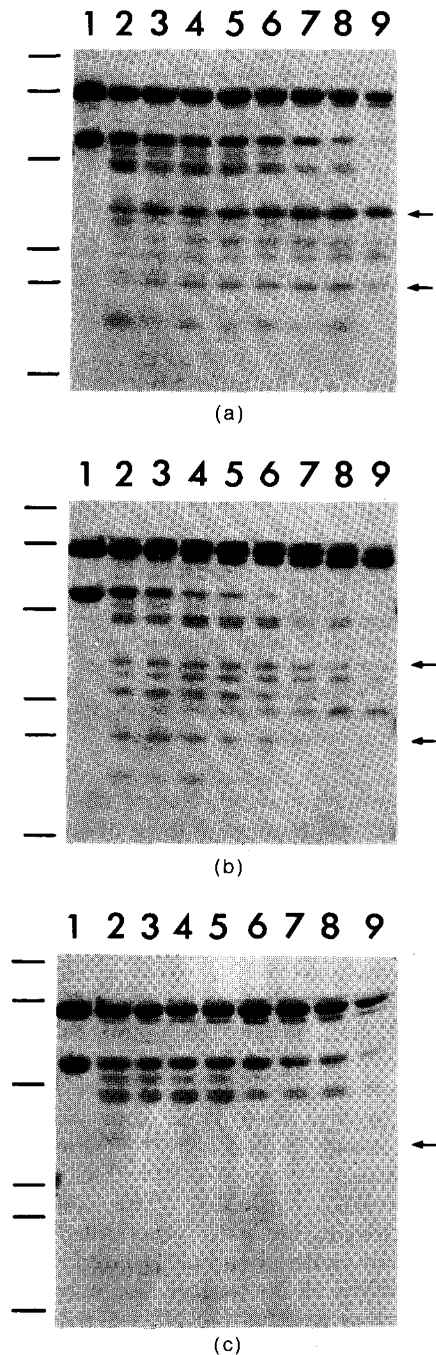

Figure 4. Mtase digestion by chymotrypsin with the same time-points for lanes 1 to 9 as in Figure 2(a) to (c). The molecular mass markers are the same except for the addition of a 18·4 kDa marker at the bottom of the gel. (a) Mtase alone showing fragments C34M and C26S. (b) Mtase plus 100 μM AdoMet showing fragments C34M and C26S. (c) Mtase plus AdoMet and unmodified DNA showing fragment C34M.

fically with each other in a 1:1 ratio. The other species observed in the large molecular weight band presumably formed intact or slightly digested mtase. There was insufficient material in the fractions to show association of the fragments by glutaraldehyde crosslinking. In the absence of T24S, the T34M fragment was still stable and eluted from the column as a monomeric species at 17·7 ml.

We did not examine chymotrypsin-digested products because we expected that the C34M fragments produced by chymotrypsin digestion would behave in the same way as T34M. Complexes of the mtase with DNA were not examined since the complex dissociated on the column.

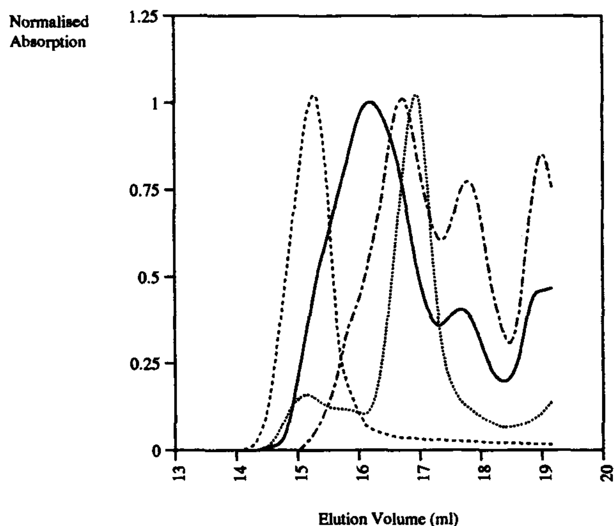

Figure 5. Normalized elution profiles of proteolytic digestions of the mtase from a FPLC Superose 6 gel filtration column. The profiles correspond to pure undigested mtase (— — —); mtase plus 100 μM AdoMet after 30 min digestion by elastase (· · · ·); mtase after 5 min digestion by trypsin (—); and mtase after 15 min digestion by trypsin (— · —). The flow-rate was 25 ml/h, the column volume 25 ml and absorption at 280 nm was monitored.

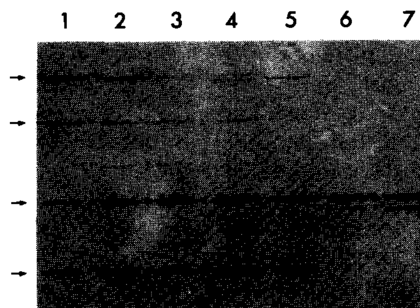

Figure 6. A silver-stained SDS/polyacrylamide gel of fractions eluting from a Superose 6 FPLC gel filtration column: 200 μl of a 5 min tryptic digestion of the mtase in the absence of any substrates was applied directly to the column giving the elution profile shown in Figure 5. Each lane shows consecutive 0·4 ml fractions starting once 15·4 ml of buffer had passed through the column after sample application. The molecular weight ranges of the assembled protein complexes as represented in each fraction are approximately 185,000 to 152,000, 152,000 to 127,000, 127,000 to 97,000, 97,000 to 60,000, 60,000 to 38,000, 38,000 to 30,000 and 30,000 to 27,500 in fractions 1 to 7 respectively, as estimated from the calibration curve for the column. The arrows (→) indicate, from top to bottom, the M subunit, the S subunit, fragment T34M and fragment T24S.

(d) *Rate of degradation of M and S subunits*

The time dependence of the amounts of fragments present was fairly complex as they initially accumulated and then degraded, but the degradation of the M and S subunits followed a simple first-order rate equation with rates that were influenced by substrates (see Table 2). The absolute rates of proteolysis of M or S subunits varied with the protease used but the changes in relative rates due to substrate binding were found to be the same. The complexes with AdoMet and hemimethylated DNA at pH 8 were active methylation complexes and some fully modified DNA–AdoHcy–mtase complexes would have been formed. The formation of these active complexes was reduced by using one-tenth of the incubation time and ten times more protease. In these cases the rate constants for mtase degradation shown in Table 2 were multiplied by ten to match the other proteolysis conditions.

The rate of degradation of the M subunit with

Table 2
Degradation of the M and S subunits

Protease	Elastase		Chymotrypsin		Trypsin	
Subunit	M	S	M	S	M	S
Ligand			Rate (min^{-1})†			
None	0·252 ± 0·022	0·366 ± 0·033	0·059 ± 0·021	0·069 ± 0·010	0·208 ± 0·029	0·211 ± 0·024
AdoMet	0·075 ± 0·016	0·382 ± 0·049	0·016 ± 0·006	0·149 ± 0·036	0·073 ± 0·017	0·114 ± 0·052
AdoHcy	0·097 ± 0·028	0·333 ± 0·047	0·025 ± 0·005	0·182 ± 0·020	0·106 ± 0·011	0·148 ± 0·034
Sinefungin	0·167 ± 0·010	0·654 ± 0·151	0·015 ± 0·012	0·099 ± 0·009	0·093 ± 0·022	0·243 ± 0·025
Unmod DNA	0·287 ± 0·018	0·284 ± 0·034	0·043 ± 0·019	0·076 ± 0·012	ND§	ND
Hemi DNA	0·198 ± 0·018	0·133 ± 0·040	0·010 ± 0·004	0·014 ± 0·009	ND	ND
AdoHcy + Unmod	0·076 ± 0·018	0·088 ± 0·021	0·013 ± 0·008	0·058 ± 0·019	ND	ND
AdoHcy + Hemi	0·068 ± 0·018	0·141 ± 0·022	ND	ND	0·125 ± 0·032	0·089 ± 0·011
AdoMet + Unmod	0·049 ± 0·011	0·088 ± 0·024	0·011 ± 0·019	0·024 ± 0·007	0·052 ± 0·022	0·037 ± 0·006
AdoMet + Unmod‡	0·058 ± 0·014	0·187 ± 0·069	ND	ND	ND	ND
AdoMet + Hemi‡	0·021 ± 0·007	0·122 ± 0·026	ND	ND	ND	ND

† The degradation rates (min^{-1}) were determined by fitting a 1st order rate equation to the protein band areas determined by densitometry of the gel normalized to time 0.

‡ Ten times as much elastase was used in these experiments to allow proteolysis to proceed much faster than DNA methylation. The rates were then multiplied by 10 to match the rest of the Table.

§ ND, Not done.

each of the three proteases was two- to threefold slower when either AdoMet, or it analogues AdoHcy or sinefungin, were present. The addition of unmodified DNA alone had no effect on the rate of degradation but hemimethylated DNA had some effect. AdoHcy and DNA together protected the M subunit between two- and fourfold in a virtually identical manner to AdoHcy alone, but AdoMet + DNA appeared to give a fourfold or higher degree of protection particularly if the DNA was hemimethylated. DNA therefore had a slight protective effect on the M subunit when AdoMet and possibly when AdoHcy were present.

The S subunit was virtually unaffected by the addition of AdoMet, AdoHcy, sinefungin or even unmodified DNA. Hemimethylated DNA by itself protected the S subunit two- to threefold and the combination of either type of DNA and either AdoMet or AdoHcy gave a two- to fivefold protection of the subunit.

4. Discussion

(a) *S subunit structure*

Proteases cut the S subunit into two groups of fragments consisting of either the N-terminal variable region plus the central conserved region and the start of the C-terminal variable region, or the central conserved region plus the C-terminal variable and conserved regions. This suggests that the S subunit forms two large folded domains held together by the two conserved regions or some portion of these conserved regions. The binding of DNA slows the digestion of the S subunit by either directly covering proteolysis sites or changing the subunit conformation. Binding of AdoMet has no effect on digestion of S subunit even though its binding causes a conformational change in the mtase (Powell *et al.*, 1993).

These results are consistent with the S subunit structure, based on extensive nucleotide and amino acid sequence comparisons of S subunits from all three type I families, of two large DNA binding domains formed by the variable regions with the smaller conserved regions separating the DNA binding domains (Gough & Murray, 1983; Fuller-Pace & Murray, 1986; Kannan *et al.*, 1989; Abadjieva *et al.*, 1993).

The N-terminal fragments, C26S, E25S and T24S, are approximately equal in size, which suggests that there may exist a small region sensitive to all three proteases. Although we have not identified the C termini of these fragments and there exist many possible proteolysis sites, we note that the sequences KRFR and VFKK starting at amino acid residues 194 and 222, respectively (Gough & Murray, 1983), both give adjacent sites for all three proteases. The VFKK sequence is predicted to form a secondary structure turn (Argos, 1985) and may be the target most likely to give fragments of these sizes. Turns are usually found on protein surfaces (Branden & Tooze, 1991).

The C-terminal fragments, E35S and E29S, are due to cleavage very close to the junction between the N-terminal variable region and the central conserved region, indicating that this junction is exposed to solvent. A sequence motif at this junction also occurs at the junction between the C-terminal variable and conserved regions and the motif is predicted to form a turn in the secondary structure (Argos, 1985).

(b) *M subunit structure*

Proteolysis of the M subunit shows that it comprises an N-terminal domain containing the m* region, which influences the enzyme specifity for hemimethylated DNA (Kelleher *et al.*, 1991) and region I of the AdoMet binding motif, a short solvent-exposed polypeptide loop containing region II of the AdoMet binding motif, a C-terminal domain involved in subunit–subunit contacts (section (d), below) and lastly a tail region involved in protein–DNA interactions (section (c), below).

The whole M subunit, except the tail region, is substantially stabilized by the binding of AdoMet or its analogues, all of which prevent proteolysis of the loop joining the N and C-terminal domains. Cofactor binding must change the conformation of the loop by making it rigid or less solvent-exposed, possibly by inducing it to close around the cofactor. Once the loop is cut, the N-terminal domain is rapidly digested but the C-terminal domain is digested only slowly by trypsin and chymotrypsin. Region II found within this loop is found in all adenine mtases (Chandrasegaran & Smith, 1988; Klimasauskas *et al.*, 1989) and has recently been shown to be important for binding AdoMet and catalysis. Site-directed mutagenesis of this region in the *Eco*KI mtase results in dramatic changes in activity with no changes in binding affinity or structure (D. Willcock, unpublished results). Changes in binding affinity and catalysis have been found in similar experiments on the Dam mtases (Guyot *et al.*, 1993; Kossykh *et al.*, 1993). In the *Taq*I adenine mtase structure, region II is also found on a surface loop (Anderson, 1993).

(c) *Protein–DNA interactions*

The unexpected increase in the susceptibility to proteolysis of fragment E55M of the M subunit in ternary complexes resulting in fragment E48M, indicates that the tail region of M changes conformation in these complexes. This change is triggered by DNA and AdoMet binding and is sensitive to the pattern of methylation on the DNA, but is not triggered if AdoHcy is substituted for AdoMet. The behaviour of the tail region suggests that it may contact the DNA substrate possibly *via* the DNA binding motif in this region postulated by amino acid sequence comparisons (Loenen *et al.*, 1987). Alternatively the sensitivity of the tail region to DNA may be a secondary effect caused by a conformational change elsewhere in the mtase. We have not been able to detect any affinity for DNA

binding by purified M subunits (unpublished results).

Whatever role the tail region fulfils, it clearly helps the mtase to differentiate between an active, efficient methylation complex, an active, inefficient methylation complex that would restrict DNA in the presence of the R subunit, and an inactive, inhibited complex (Burckhardt *et al.*, 1981). DNA binding studied by gel retardation of ternary complexes containing various methylated DNA substrates and AdoMet or AdoHcy also shows that the mtase can distinguish between different ternary complexes (Powell *et al.*, 1993).

(d) *Subunit–subunit contacts*

The DNA target of a type I enzyme consists of two "half sites" each of which contains a methylatable adenine base. The conformational differences in the enzyme that distinguish hemimethylated DNA from unmethylated DNA imply that the enzyme adopts a different structure around a methylated half site than around an unmethylated half site. They also imply that there is communication between the structures around each half site to ensure that fully unmodified substrates do not get methylated. If the structures around each half site acted independently, then the specificity for hemimethylated DNA could not be achieved. This situation may arise for type IA enzymes in the presence of the phage λ Ral protein (Leonen & Murray, 1986), in the m* mutants of the *Eco*KI mtase (Kelleher *et al.*, 1991) or in the type IB systems exemplified by *Eco*AI (Suri & Bickle, 1985), where in each case methylation of unmodified sequences is efficient. Co-operative interactions between half sites are most likely to involve changes in the interfaces of the domains of the S and M subunits (Yuan, 1981; Taylor *et al.*, 1993).

Part of the interface between the M and S subunits occurs between the C-terminal half of M (fragment T34M) and the N-terminal half of S (fragment T24S) since they associate to form a large molecular weight complex as determined by gel filtration (Figures 5 and 6).

The extent of the interface on the M subunit can be further confined to a region smaller than T34M since the M subunit remains bound to the S subunit in ternary complexes even when elastase has cleaved off the tail region to form fragment E48M. Therefore the region of the M subunit that forms an interface with the S subunit is contained within the C-terminal domain of the M subunit and does not have to include the tail region of the M subunit.

The T24S fragment of the S subunit contacts the M subunit but the proteolysis results do not define the interface more precisely. However, it is possible to conclude that only the central conserved region of S within fragment T24S is involved in contacts with the C-terminal domain of the M subunit. It has been proposed that the central and carboxyl conserved regions of the S subunit are involved in interfaces with the M subunit (Gough & Murray, 1983; Gann *et al.*, 1987; Kannan *et al.*, 1989) and strong evidence for this has recently been found. Gann (1988) constructed a truncated version of the S subunit of the type IA system of *Salmonella enterica serovar potsdam* that lacked the N-terminal variable region. The S subunit in this bacterium is virtually identical in size and definition of the amino acid regions with the S subunit shown in Figure 1(a). He found that the presence of this truncated subunit, equivalent to our E35S fragment, lowered the restriction and modification efficiency of the normal system *in vivo*, presumably by competing with the normal S subunits for binding the M subunits. This result suggests that the interface between fragment T24S and T34M does not require the N-terminal variable region of T24S. This reduces the interface on T24S to the central conserved region and the start of the C-terminal variable region. More recently, it has been found that the N-terminal variable and central conserved region of two type IC S subunits are sufficient to bind M subunits and support DNA methylation (Abadjieva *et al.*, 1993; Meister *et al.*, 1993). Furthermore, Meister *et al.* (1993) found that the last ten amino acid residues of the central conserved region were not required for function. Hence only part of the central conserved region is required to contact the M subunit.

It is probable that the interface between the S subunit and the second M subunit in the mtase involves the carboxyl conserved region of S. All S subunits of type I systems contain a central conserved region and either an amino or carboxyl conserved region (Kannan *et al.*, 1989; Abadjieva *et al.*, 1993). Parts of these conserved regions are amino acid repeats that are present in all three type I families while the rest of these regions are conserved within a family (Argos, 1985; Kannan *et al.*, 1989). The same repeats occur in both conserved regions suggesting that these regions of type I S subunits all have the same functions of binding the two M subunits and defining the length of the non-specific spacer in the DNA target. The truncated S subunits of the type IC S subunits appear to form dimers that recognize a bipartite symmetrical DNA target and which bind two M subunits (Abadjieva *et al.*, 1993; Meister *et al.*, 1993). This indicates that the central conserved region of the type IC S subunits can successfully fulfil the role of the carboxyl conserved region and supports the conclusion reached by sequence comparisons that the two regions have the same function.

The recent determination of the type II *Hha*I cytosine mtase structure (Cheng *et al.*, 1993) shows that it contains two distinct folded domains, one for DNA target sequence recognition, the other for catalysis. The preliminary crystal structure data on the *Taq*I adenine mtase (Anderson, 1993) and the limited proteolysis of the *Eco*RI type II mtase (Reich *et al.*, 1991) suggest that they also have a two-domain structure with perhaps the same separation of DNA recognition and catalysis as in *Hha*I mtase. Our type I mtase, *Eco*KI, contains two

DNA recognition domains and two catalytic domains, hence one might suggest that its structure would be like that of two type II mtases, one stacked upon the other around a 2-fold rotational axis (Abadjieva *et al.*, 1993; Meister *et al.*, 1993). Instead of each DNA binding domain being on the same polypeptide as a catalytic domain, they are joined together to form the S subunit, and the catalytic domains are joined to other domains to form the two M subunits. The extra domains and tail of the M subunit hold the complex together and communicate the methylation state of one DNA target half site to the other.

Further subunit–subunit and subunit–DNA contacts remain to be elucidated as do the mechanism of action of the m* region (Kelleher *et al.*, 1989) and the tail region of the M subunit defined in this paper, both of which appear to have a role in recognizing DNA methylation. Our results complement those obtained by construction of deletion mutants (Gann, 1988; Abadjieva *et al.*, 1993; Meister *et al.*, 1993) and suggest further targets for mutagenesis and deletion experiments. Furthermore, the structure formed by the T34M and T24S fragments, being much smaller than the mtase, may be a suitable subject for X-ray crystallography.

This work was supported by the Medical Research Council and the Wolfson Foundation who awarded a Wolfson Fellowship to D.T.F.D. We are grateful to Professor Noreen Murray and our colleagues in her laboratory for their support and advice. We thank the WELMET Protein Characterisation Facility, University of Edinburgh, for the protein sequencing, Graham Brown and Frank Johnston for photography and Dr K. Firman and Professor T. Bickle for communicating results prior to publication. This is a publication from the Edinburgh Centre for Molecular Recognition.

References

Abadjieva, A., Patel, J., Webb, M., Zinkevich, V. & Firman, K. (1993). A deletion mutant of the type IC restriction endonucleose, *Eco*R124I expressing a novel DNA specificity. *Nucl. Acids Res.* **21**, 4435–4443.

Anderson, J. E. (1993). Restriction endonucleases and modification methylases. *Curr. Opin. Struct. Biol.* **3**, 24–30.

Ansorge, W. (1985). Fast and sensitive detection of protein and DNA bands by treatment with potassium permanganate. *J. Biochem. Biophys. Meth.* **11**, 13–20.

Argos, P. (1985). Evidence for a repeating domain in type I restriction enzymes. *EMBO J.* **4**, 1351–1355.

Bickle, T. A. & Kruger, D. H. (1993). Biology of DNA restriction. *Microbiol. Rev.* **57**, 434–450.

Branden, C. & Tooze, J. (1991). In *Introduction to Protein Structure*, pp. 18–19, Garland Publishing Inc., New York.

Burckhardt, J., Weisemann, J., Hamilton, D. L. & Yuan, R. (1981). Complexes formed between the restriction endonuclease *Eco*K and heteroduplex DNA. *J. Mol. Biol.* **153**, 425–440.

Carrey, E. A. (1989). Peptide mapping. In *Protein Structure: A Practical Approach* (Creighton, T. E., ed.), pp. 117–144, IRL Press, Oxford.

Chandrasegaran, S. & Smith, H. O. (1988). Amino acid sequence homologies among twenty-five restriction endonucleases and methylases. In *Structure and Expression. From Proteins to Ribosomes* (Sarma, R. H. & Sarma, M. H., eds), pp. 149–156, Adenine Press Inc. New York.

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

Cowan, G. M., Gann, A. A. F. & Murray, N. E. (1989). Conservation of complex DNA recognition domains between families of restriction enzymes. *Cell,* **56**, 103–109.

Dryden, D. T. F., Cooper, L. P. & Murray, N. E. (1993). Purification and characterisation of the methyltransferase from the type I restriction and modification system of *Escherichia coli* K12. *J. Biol. Chem.* **268**, 13228–13236.

Fontana, A., Fassina, G., Vita, C., Dalzoppo, D., Zamai, M. & Zambonin, M. (1986). Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry,* **25**, 1847–1851.

Fuller-Pace, F. V. & Murray, N. E. (1986). Two DNA recognition domains of the specificity polypeptides of a family of type I restriction enzymes. *Proc. Nat. Acad. Sci., U.S.A.* **83**, 9368–9372.

Gann, A. A. F. (1988). Recognition domains of type I restriction enzymes. Ph.D. thesis, University of Edinburgh.

Gann, A. A. F., Campbell, A. J. B., Collins, J. F., Coulson, A. F. W. & Murray, N. E. (1987). Reassortment of DNA recognition domains and the evolution of new specificities. *Mol. Microbiol.* **1**, 13–22.

Geisler, N. & Weber, K. (1977). Isolation of the amino-terminal fragment of lactose repressor necessary for DNA binding. *Biochemistry,* **16**, 938–943.

Gorbalenya, A. E. & Koonin, E. V. (1991). Endonuclease (R) subunits of type I and type III restriction-modification enzymes contain a helicase-like domain. *FEBS Letters,* **291**, 277–281.

Gough, J. A. & Murray, N. E. (1983). Sequence diversity among related genes for recognition of specific targets in DNA molecules. *J. Mol. Biol.* **166**, 1–19.

Gubler, M. & Bickle, T. A. (1991). Increased protein flexibility leads to promiscuous protein–DNA interactions in type IC restriction-modification systems. *EMBO J.* **10**, 951–957.

Gubler, M., Braguglia, D., Meyer, J., Piekarowicz, A. & Bickle, T. A. (1992). Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. *EMBO J.* **11**, 233–240.

Guyot, J.-B., Grassi, J., Hahn, U. & Guschlbauer, W. (1993). The role of the preserved sequence of Dam methylase. *Nucl. Acids Res.* **21**, 3183–3190.

Hartley, B. S. & Shotton, D. M. (1971). Pancreatic elastase. In *The Enzymes* (Boyer, P. D., ed.), 3rd edit., vol. 3, pp. 323–373, Academic Press, New York.

Hattman, S., Wilkinson, J., Swinton, D., Schlagman, S., MacDonald, P. M. & Mosig, G. (1985). Common evolutionary origin of the phage T4 *dam* and host *Escherichia coli dam* DNA-adenine methyltransferase genes. *J. Bacteriol.* **164**, 932–937.

Heitman, J. (1993). On the origins, structures and

functions of restriction-modification enzymes. In *Genetic Engineering* (Setlow, J. K., ed.), vol. 15, pp. 57–108, Plenum Press, New York.

Jaenicke, R. & Rudolph, R. (1989). Folding proteins. In *Protein Structure: A Practical Approach* (Creighton, T. E., ed.), pp. 311–330, IRL Press, Oxford.

Kannan, P., Cowan, G. M., Daniel, A. S., Gann, A. A. F. & Murray, N. E. (1989). Conservation of organisation in the specificity polypeptides of two families of type I restriction enzymes. *J. Mol. Biol.* **209**, 335–344.

Kelleher, T. E., Daniel, A. S. & Murray, N. E. (1991). Mutations that confer *de novo* activity upon a maintenance methyltransferase. *J. Mol. Biol.* **221**, 431–440.

Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V. & Janulaitis, A. (1989). Sequence motifs characteristic of DNA [cytosine-N4] methyltransferases: similarity to adenine and cytosine-C5 DN6A-methylases. *Nucl. Acids Res.* **17**, 9823–9831.

Kossykh, V. G., Schlagman, S. L. & Hattman, S. (1993). Conserved sequence motif DPPY in region IV of the phage T4 Dam DNA-[N^6-adenine]-methyltransferase is important for *S*-adenosyl-L-methionine binding. *Nucl. Acids Res.* **21**, 3563–3566.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)*, **227**, 680–685.

Loenen, W. A. M. & Murray, N. E. (1986). Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage λ. *J. Mol. Biol.* **190**, 11–22.

Loenen, W. A. M., Daniel, A. S., Braymer, H. D. & Murray, N. E. (1987). Organisation and sequence of the *hsd* genes of *Escherichia coli* K-12. *J. Mol. Biol.* **198**, 159–170.

Matsudaira, P. T. (1989). *A Practical Guide to Protein and Peptide Purification for Microsequencing.* Academic Press, New York.

Meister, J., MacWilliams, M., Hubner, P., Jutte, H., Skrzypek, E., Piekarowicz, A. & Bickle, T. A. (1993). Macroevolution by transposition: drastic modification of DNA recognition by a type I restriction enzyme following Tn5 transposition. *EMBO J.* **12**, 4585–4591.

Murray, N. E., Gough, J. A., Suri, B. & Bickle, T. A. (1982). Structural homologies among type I restriction-modification systems. *EMBO J.* **1**, 535–539.

Murray, N. E., Daniel, A. S., Cowan, G. M. & Sharp, P. M. (1993). Conservation of motifs within the unusually variable polypeptide sequences of type I restriction and modification enzymes. *Mol. Microbiol.* **9**, 133–143.

Pabo, C. O., Sauer, R. T., Sturtevant, J. M. & Ptashne, M. (1979). The λ repressor contains two domains. *Proc. Nat. Acad. Sci., U.S.A.* **76**, 1608–1612.

Patel, J., Taylor, I., Dutta, C. F., Kneale, G. & Firman, K. (1992). High level expression of the subunits and intact DNA methyltransferase of the type I restriction and modification system *Eco*R124. *Gene,* **112**, 21–27.

Powell, L. M., Dryden, D. T. F., Willcock, D. F., Pain, R. H. & Murray, N. E. (1993). DNA recognition by the *Eco*K methyltransferase: the influence of DNA methylation and the cofactor *S*-adenosyl-L-methionine. *J. Mol. Biol.* **234**, 60–71.

Price, C., Lingner, J., Bickle, T. A., Firman, K. & Glover, S. W. (1989). Basis for changes in DNA recognition by the *Eco*R124 and *Eco*R124/3 type I DNA restriction and modification enzymes. *J. Mol. Biol.* **205**, 115–125.

Reich, N. O., Maegley, K. A., Shoemaker, D. D. & Everett, E. (1991). Structural and functional analysis of *Eco*RI DNA methyltransferase by proteolysis. *Biochemistry,* **30**, 2940–2946.

Sharp, P. M., Kelleher, J. E., Daniel, A. S., Cowan, G. M. & Murray, N. E. (1992). Roles of selection and recombination in the evolution of type I restriction-modification systems in enterobacteria. *Proc. Nat. Acad. Sci., U.S.A.* **89**, 9836–9840.

Steck, G., Leuthard, P. & Burk, R. R. (1980). Detection of basic proteins and low molecular weight peptides in polyacrylamide gels by formaldehyde fixation. *Anal. Biochem.* **107**, 21–24.

Suri, B. & Bickle, T. A. (1985). *Eco*A: The first member of a new family of type I restriction modification systems. Gene organisation and enzymatic activities. *J. Mol. Biol.* **186**, 77–85.

Taylor, I., Patel, J., Firman, K. & Kneale, G. (1992). Purification and biochemical characterisation of the *Eco*R124 Type I modification methylase. *Nucl. Acids Res.* **20**, 179–186.

Taylor, I., Watts, D. & Kneale, G. (1993). Substrate recognition and selectivity in the type IC DNA modification methylase M. *Eco*R124I. *Nucl. Acids Res.* **21**, 4929–4935.

Vershon, A. K., Youderian, P., Weiss, M. A., Susskind, M. M. & Sauer, R. T. (1985). Mnt repressor–operator interactions: altered specificity requires N-6 methylation of operator DNA. In *Sequence Specificity in Transcription and Translation* (Calendar, R. & Gold, L., eds), pp. 209–218, Alan R. Liss Inc., New York.

Vita, C., Fontana, A. & Jaenicke, R. (1989). Folding of thermolysin fragments. Hydrodynamic properties of isolated domains and subdomains. *Eur. J. Biochem.* **183**, 513–518.

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

Yuan, R. (1981). Structure and mechanism of multifunctional endonucleases. *Annu. Rev. Biochem.* **50**, 285–315.

Edited by J. Karn

(Received 12 October 1993; accepted 7 December 1993)