

Surface Labelling of the Type I Methyltransferase *M.EcoR124I* Reveals Lysine Residues Critical for DNA Binding

I. A. Taylor, M. Webb and G. G. Kneale*

Biophysics Laboratories
School of Biological Sciences
University of Portsmouth
Portsmouth PO1 2DT, UK

The type IC methyltransferase *M.EcoR124I* consists of a specificity subunit (HsdS) and two methylation subunits (HsdM). Using chemical modification, we have investigated the accessibility of lysine residues in the free enzyme and in the complex with its DNA recognition sequence. A total of 41 of the 109 lysine residues in the enzyme are susceptible to modification, of which 19 are located in the HsdS subunit and 11 in each of the two HsdM subunits. DNA binding results in extensive protection of lysine residues in the HsdS subunit, while those in the HsdM subunit are only protected weakly. The DNA binding activity of the methylase is abolished when a small fraction of the accessible lysine residues are modified. Peptide mapping and N-terminal sequencing has been used to locate the rapidly modified lysine residues in HsdS that are critical for DNA binding. Highly modified residues (K297, K261 and K327) are found in the C-terminal variable domain that is responsible for DNA recognition, but others (K196, K203 and K210) are found in the conserved regions that had not previously been implicated in DNA binding.

© 1996 Academic Press Limited

Keywords: DNA methyltransferases; chemical modification; DNA–protein interactions; restriction–modification systems; subunit interactions

*Corresponding author

Introduction

Enzymes that methylate specific DNA sequences are ubiquitous in living cells, reflecting the importance of DNA methylation in processes as varied as gene expression, cell differentiation, restriction-modification and DNA repair. Structural and mechanistic studies to date have been most successful in the analysis of type II modification methylases, and the recent crystal structure of *M.HhaI* complexed to DNA has shown the remarkable manipulation of DNA structure that such enzymes must perform (Klimasauskas *et al.*, 1994; Kumar *et al.*, 1994).

Type I methyltransferases (MTases) are more complex enzymes (Wilson & Murray, 1991; Bickle & Kruger, 1993), having separate subunits for DNA

recognition (HsdS) and methyltransferase activity (HsdM). In contrast to the compact and symmetrical 4 to 6 bp sequence typically recognised by type II enzymes, type I MTases specifically recognise an asymmetric bipartite DNA sequence, in which the centres of the two halves of the recognition sequence are separated by approximately one helical turn of DNA. This is accomplished by means of two discrete DNA recognition domains (variable domains) in the HsdS subunit (Cowan *et al.*, 1989; Gubler *et al.*, 1992) connected by a central conserved domain, which is homologous to other conserved regions at the N and/or C termini of the protein (Fuller-Pace & Murray, 1986; Kannan *et al.*, 1989; Tyndall *et al.*, 1994). The HsdS subunit interacts with two HsdM subunits to form a large multisubunit complex (Taylor *et al.*, 1992; Powell *et al.*, 1993), probably *via* the conserved domains of the former (Cooper & Dryden, 1994). Deletion mutants of HsdS give rise to enzymes with symmetrical DNA recognition sequences, suggesting that two copies of the truncated HsdS subunits are combined in the mutant enzymes (Abadjieva *et al.*, 1993; Meister *et al.*, 1993). Internal sequence homologies suggest a novel circular organisation of the HsdS domains,

Abbreviations used: MTase, methyltransferase; Adomet, *S*-adenosyl-*L*-methionine; p.p.m., parts per million; HPLC, high pressure liquid chromatography; CNBr, cyanogen bromide; α -MSH, α -melanocyte stimulating hormone; DFP, di-isopropyl fluorophosphate; PTH, phenylthiohydantoin; ATZ, aminoacetylthiozolinone; TFA, trifluoroacetic acid.

which locates the two HsdM subunits symmetrically with respect to the target sites on the DNA (Kneale, 1994).

Two type I methyltransferases, *M. EcoR124I* and *M. EcoKI*, have been over-expressed and purified in sufficient quantities for detailed biochemical and biophysical analysis. Both enzymes have been well characterised in terms of their subunit composition, domain structure, DNA binding characteristics and enzyme activity (Taylor *et al.*, 1992, 1993; Dryden *et al.*, 1993; Powell *et al.*, 1993; Powell & Murray, 1995; Webb *et al.*, 1995).

M. EcoR124I consists of two copies of the HsdM subunit (each 58 kDa) and one HsdS subunit (46 kDa), forming a trimeric enzyme (162 kDa) with a subunit stoichiometry of M_2S_1 . The HsdS subunit alone is insoluble, and the interaction with HsdM is essential to maintain its solubility and structural integrity (Patel *et al.*, 1992; Kneale, 1994). *M. EcoR124I* recognises the sequence GAAN₆RTCG, and binds to a 30 bp DNA duplex containing this sequence with high affinity ($K_d = 10^{-8}$ M) (Taylor *et al.*, 1993). In common with all other type I MTases, the enzyme uses the cofactor *S*-adenosyl-L-methionine (Adomet) to methylate individual adenine residues at the N-6 position on opposite strands of the DNA recognition sequence. Methylation of the target adenine on either strand of the DNA recognition sequence reduces the DNA binding affinity around 30-fold (Taylor *et al.*, 1993).

X-ray solution scattering and circular dichroism have been used to determine the structural parameters of *M. EcoR124I* and its complex with a specific 30 bp DNA duplex (Taylor *et al.*, 1994). There is a dramatic reduction in the overall dimensions of the enzyme when bound to the DNA, accompanied by a large conformational change in the DNA itself. This structural transition appears to involve a large rotation of the HsdM subunits to clamp the DNA, and is driven, at least in part, by non-sequence specific interactions between the protein and DNA outside of the DNA recognition sequence. The resulting DNA-protein complex is very compact and DNase I footprinting shows that the DNA is protected on both strands over 20 bp or more (D. R. Mernagh & G. G. Kneale, unpublished results).

Chemical modification of amino acid residues provides a valuable tool for analysis of enzyme structure and function (Lundblad & Noyes, 1984). Using reagents specific for different amino acids, such techniques have been used extensively to identify key residues involved in protein-nucleic acid interactions (e.g. see Tsugita & Kneale, 1985; Sams & Matthews, 1988; Plyte & Kneale, 1991).

A particularly useful reagent for the modification of lysine residues is formaldehyde, in the presence of the mild reducing agent sodium cyanoborohydride (Jentoft & Dearborn, 1979; Thomas, 1989). Under reducing conditions formaldehyde will convert the ϵ -amino group of lysine to the *N,N*-dimethyl derivative by reductive methylation. Conversion of lysine to the dimethyl derivative is a

relatively small change, and the residue keeps its ionisation properties with only a small change in pK_a . Mild treatment is therefore unlikely to significantly perturb the native enzyme structure.

Reductive methylation has been successfully applied to the investigation of the interactions of the core histones with DNA in the nucleosome (Lambert & Thomas, 1986) and the interaction of the linker histone H5 with nucleosomes in long chromatin (Thomas & Wilson, 1986). We report the effects of reductive methylation of *M. EcoR124I* and show that DNA binding confers preferential protection on residues in the HsdS subunit. We identify a number of lysine residues which are rapidly modified in the MTase, principally those in and around the C-terminal variable domain of the HsdS subunit. Modification of these highly accessible lysine residues is shown to completely abolish DNA binding.

Results

Calibration of the reductive methylation reaction

In order to quantify the extent of modification of *M. EcoR124I*, and thus determine the number of accessible lysine residues in the enzyme and the enzyme-DNA complex, it was necessary to calculate the specific activity of the reagent and to characterise the product of the reductive methylation. The peptide hormone α -MSH (Ac-SYSMEHFRWGKPV-NH₂) is ideal for this purpose. The peptide contains just a single lysine, has a blocked N terminus and the presence of the two aromatic residues allow its concentration to be determined accurately from its UV absorbance at 280 nm.

Modification of α -MSH was carried out as described in Materials and Methods and the modified peptide analysed by fast atom bombardment mass spectroscopy (Figure 1). The spectrum of unmodified α -MSH shows a single species at $M_r = 1666$ corresponding to the monoprotonated species (the $[M + H]^+$ ion) produced from the α -MSH molecule. The spectrum of modified peptide has a major component corresponding to the mass of dimethylated α -MSH ($M_r = 1694$) and a minor component (2 to 3%) corresponding to the monomethylated peptide. To confirm the nature of the modification, larger quantities of the modified α -MSH were prepared and ¹H NMR spectra were recorded of the modified and unmodified peptide (data not shown). The NMR spectrum of modified α -MSH resembles that of the unmodified peptide, except for a large singlet at 2.90 p.p.m. found only in the modified peptide with an integrated area equivalent to six protons, consistent with full modification to dimethyllysine under these conditions. By measuring the radioactivity of the modified peptide it was therefore possible to obtain the true specific activity of the reagent.

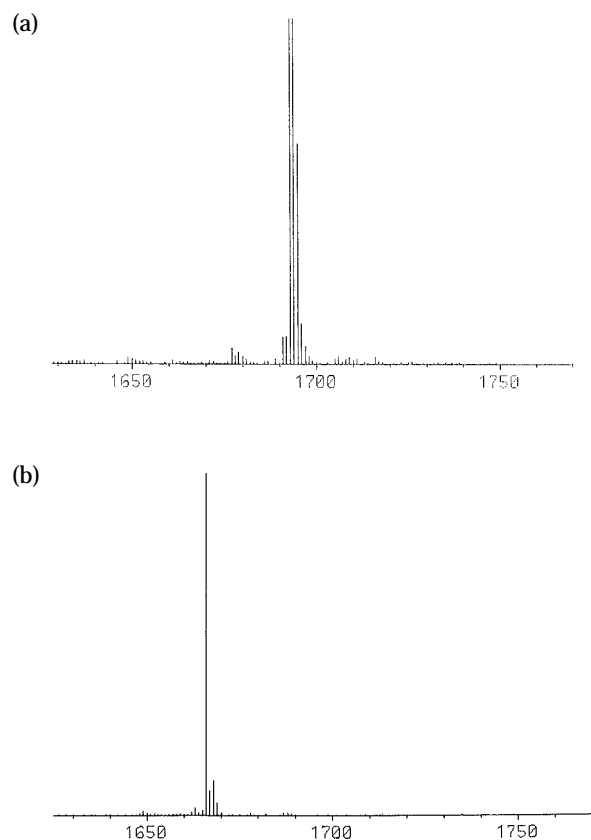


Figure 1. Analysis of unmodified (b) and modified (a) α -MSH by fast atom bombardment mass spectroscopy. Measurements were recorded on a Jeol DX303/DA500 double focusing mass spectrometer (accelerating potential 4 kV). Scanning was performed from $m/z = 50$ to $m/z = 4500$ in 53.6 seconds.

Modification of *M. EcoR124I*

To find the number of surface accessible lysine residues in *M. EcoR124I*, the protein was reductively methylated and the average number of modified lysine residues determined at various time points during the reaction (Figure 2(a)). The kinetic curve can be fitted by a first-order rate equation. Thus, only two classes of lysine residues can be clearly distinguished: those that are readily modified and those that are buried within the protein and are totally unmodified during the time scale of the reaction. The extent of modification at the apparent end-point of the reaction suggests that 40 lysine residues are surface accessible in the free methylase. Since there are a total of 109 lysine residues in the trimeric enzyme, 69 of these are inaccessible. The low percentage (37%) of surface accessible lysine residues may reflect the multisubunit and multi-domain structure of the enzyme.

A parallel experiment on the methylase-DNA complex was also carried out (Figure 2(a)). The rate of modification of the lysine residues is reduced by a factor of 3 in the DNA-protein complex, although the end-point of the reaction is much the same as for the free enzyme (Table 1). It is clear that around

two-thirds (on average) of the previously accessible lysine residues are protected by DNA binding during the early stages of the reaction. Nevertheless, the protection afforded by DNA binding is far from complete, since the lysine residues in the DNA-protein complex are slowly modified. This contrasts with the residues that are buried in the methylase, which show no appreciable modification during the time scale of the reaction.

Differential modification of HsdM and HsdS subunits

The products of the modification reactions were further analysed by densitometry of fluorographed SDS-containing gels at various times during the reaction (Figure 2(b)). The most striking feature is the difference in the relative band intensities when the DNA-protein complex is modified, especially during the early stages of the reaction. In contrast, for the free methylase the relative intensity of the two bands is constant throughout. In the free methylase the ratio of the radiolabel incorporated into the subunits is $M:S = 1.1$ throughout the experiment. Since the lysine content of the HsdM and HsdS subunits in the intact enzyme (M_2S) is in the ratio 70:39, this indicates that the fraction of accessible lysine residues in HsdS is almost twice that in the HsdM.

When the MTase is complexed with a 30 bp DNA duplex containing its recognition sequence, the ratio ($M:S$) of the radiolabel incorporated is about four in the early part of the reaction, and remains less than two after five hours. Clearly, lysine residues in the HsdS subunit are much more extensively protected from modification than those in HsdM when the methylase binds to its DNA recognition sequence.

By combining the data in Figure 2, it is possible to estimate the rate of modification of the HsdS and HsdM subunits separately. This can be done both for the free methylase and for the complex with DNA (Figure 3). In the free methylase, both subunits are modified at a similar rate, but the total number of modified residues differs; ~ 18 of the 39 lysine residues in HsdS are accessible, compared with ~ 11 of the 35 lysine residues in each HsdM subunit (see Table 1).

Differential protection is clearly seen when the accessibility of the lysine residues of individual subunits in the methylase-DNA complex is compared. For lysine residues in the HsdM subunit, very limited protection is conferred by the interaction with DNA; indeed, after five hours none of the surface accessible lysine residues in HsdM is protected from modification in the DNA-protein complex.

In contrast, the rate of modification of HsdS is reduced by an order of magnitude when the enzyme is bound to its cognate DNA recognition sequence. After 30 minutes an average of about nine residues in the HsdS subunit are modified in the free enzyme, but only about one is modified in the

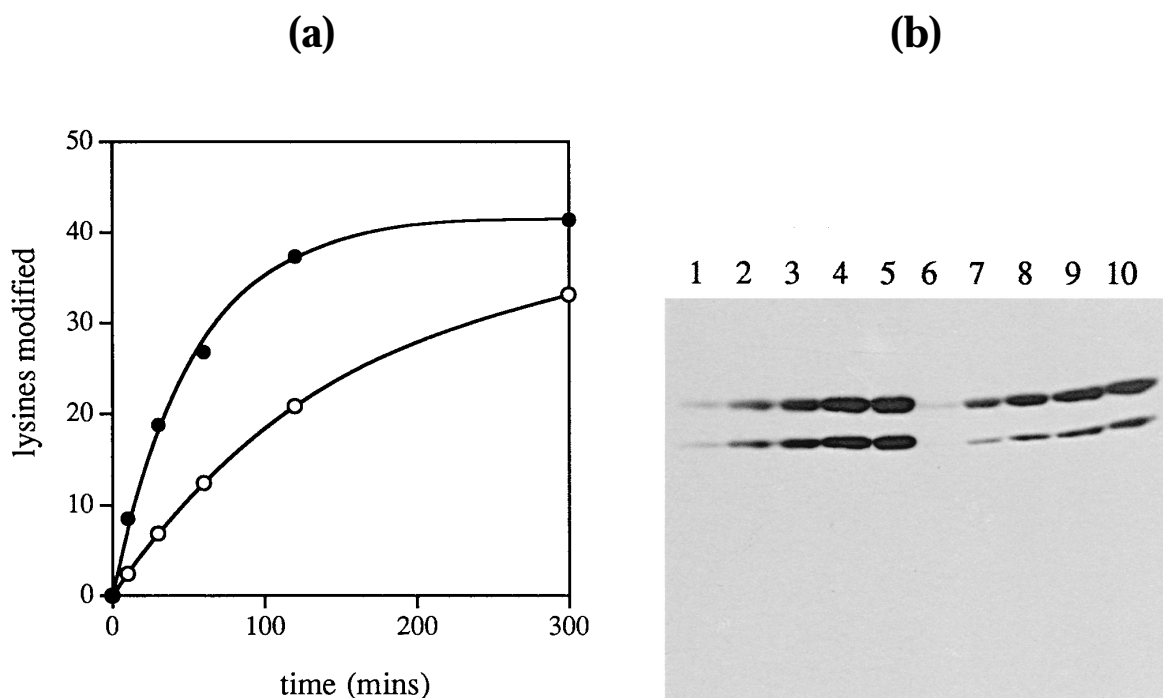


Figure 2. Time course of modification of *M. EcoR124I*. (a) Reductive methylation was carried out on the free MTase (●) and the complex of the MTase with a 30 bp DNA duplex (○) and the amount of label incorporated calculated from the specific activity of the reagent. Curves were fitted as described in Table 1 to obtain the rate and the total number of lysine residues modified. (b) Fluorography of an SDS-containing gel following modification of MTase (tracks 1 to 5) and the MTase–DNA complex (tracks 6 to 10) after 10, 30, 60, 120 and 300 minutes, respectively, to determine the ratio of radiolabel incorporation into the individual subunits.

DNA–protein complex. Even after two to three hours, when modification of the free enzyme is complete, only a third of the lysine residues on the surface of HsdS have reacted in the complex. Nevertheless, kinetic analysis indicates that most, if not all, of the lysine residues that are accessible in the free enzyme are eventually modified in the complex.

Effect of lysine modification on DNA binding

To investigate the effect of lysine modification on the DNA binding properties of the methylase, the

Table 1. Modification of the *EcoR124I* methylase and the methylase–DNA complex

	DNA	Rate (min^{-1})	Lysine residues modified	Total lysine residues
MTase	–	0.019	41	109
MTase	+	0.006	39	109
HsdS	–	0.023	19	39
HsdS	+	0.003	18	39
HsdM	–	0.018	11	35
HsdM	+	0.009	12	35

The rate of modification (k) and the total number of lysine residues modified in the reaction (A) were estimated by fitting the data in Figure 2 and Figure 3 to the first-order rate equation $y = A(1 - e^{-kt})$. The possibility of modification of the N-terminal amino group in each of the subunits has been ignored, since the number of such groups is small in comparison with the number of lysine residues present.

enzyme was modified for various times and incubated with an equimolar amount of a 30 bp duplex containing the *EcoR124I* recognition sequence. The ability of the protein to form a DNA–protein complex was analysed by gel retardation (Figure 4). It is clear that the free methylase is rapidly inactivated by the reaction, with DNA binding activity totally abolished following only brief (10 to 15 minutes) modification. In contrast, when the methylase is modified in the presence of the DNA duplex, the resulting protein retains almost native DNA binding activity, even after prolonged modification. This indicates that surface accessible lysine residues that are not protected in the DNA–protein complex do not inhibit DNA binding by some indirect mechanism, and confirms that the rapidly labelled lysine residues in the free enzyme are protected in the complex.

Comparison of the rate of inactivation of DNA binding by modification of free methylase ($k = 0.1557 \text{ min}^{-1}$) with the rate of inactivation of methylase modified in the complex ($k = 0.0006 \text{ min}^{-1}$) indicates a ~ 250 -fold decrease in the rate of inactivation when the modification is carried out in the presence of the bound DNA duplex. Thus, the decrease in DNA binding ability falls off very much faster than can be accounted for simply by the average levels of modification in the two experiments (which differ by a factor of about

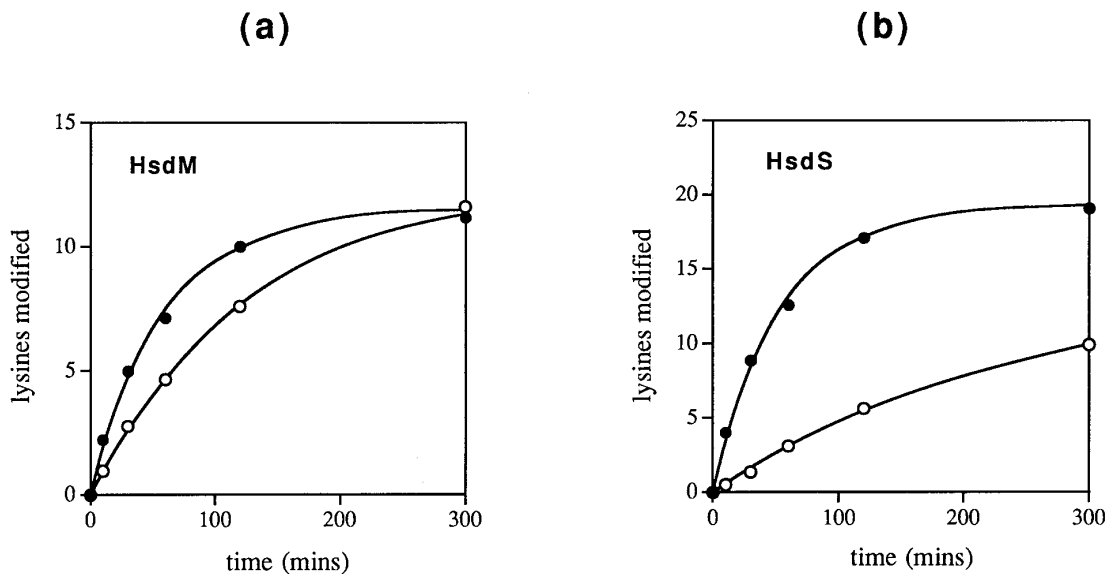


Figure 3. Effect of DNA binding on the modification of (a) HsdM and (b) HsdS. Modification was performed on the MTase (●) and the MTase–DNA complex (○) and the data shown in Figure 2 used to determine the number of lysine residues modified at each time point in the reaction. The curves were fitted by the parameters shown in Table 1.

10). Clearly, those lysine residues that are rapidly modified in the free methylase, and which are protected in the complex with DNA, are absolutely critical for DNA binding activity.

If one compares the kinetics of inactivation (Figure 4) and modification (Figures 2 and 3) it is apparent that when only ~8 of the 40 surface accessible lysine residues are modified (about four in the HsdS subunit and about two in each of the HsdM subunits), DNA binding is almost abolished. Of course, these numbers are statistical averages over the entire population of modified protein molecules. It is not possible from these data to determine whether, for example, there are four lysine residues that are fully modified, or whether a greater number are partially modified. The ability to fit the kinetic curves by a single rate equation would, at first sight, suggest that all lysine residues are modified at broadly similar rates. However, if only a few lysine residues were rapidly modified, this would be difficult to detect from the kinetic data. Peptide mapping, however, can be used to investigate the labelling of individual lysine residues, although for a protein of this size, there are considerable experimental difficulties with this approach.

Pulse chase modification of *M. EcoR124I*

One approach to selective labelling of residues involved in DNA binding is to modify residues in the complex with unlabelled reagent, then to remove the DNA and subsequently modify the newly exposed residues with labelled reagent (Tsugita & Kneale, 1985). For subsequent mapping, it is preferable to also modify the remaining (buried) residues following denaturation of the

protein, i.e. the “cold-hot-cold” technique (Thomas, 1989). Since the most strongly protected lysine residues are in HsdS, subsequent peptide mapping procedures were limited to this subunit. Furthermore, since in our experiments virtually all of the lysine residues in HsdS are protected from modification in the DNA–protein complex, there is little to be gained by the initial “cold-labelling” of the complex, thus simplifying the procedure.

To investigate the location of the lysine residues, pulse labelling with [³H] formaldehyde was followed by an analysis of the labelled peptides obtained by cleavage of the purified HsdS subunit. *M. EcoR124* was modified briefly by reductive methylation for varying times (10 to 20 minutes depending on the particular batch of formaldehyde used in the modification). Times of modification were adjusted to achieve approximately 85% inactivation of DNA binding, as judged by gel retardation assays of the modified protein. After separation of the subunits and subsequent modification of the denatured HsdS subunit with unlabelled reagent, the protein was subjected to peptide mapping.

Identification of labelled lysine residues

Digestion of reductively methylated proteins with trypsin allows selective cleavage at arginine residues as dimethyllysine is resistant to cleavage (Thomas, 1989). Even so, the number of fragments that can be produced by trypsin digestion is very high, due to the large size of the polypeptide. Potential cleavage sites in HsdS are shown in Figure 5(b). The lack of any arginine residues in the N-terminal region of the protein meant that this region could not be digested into sufficiently small

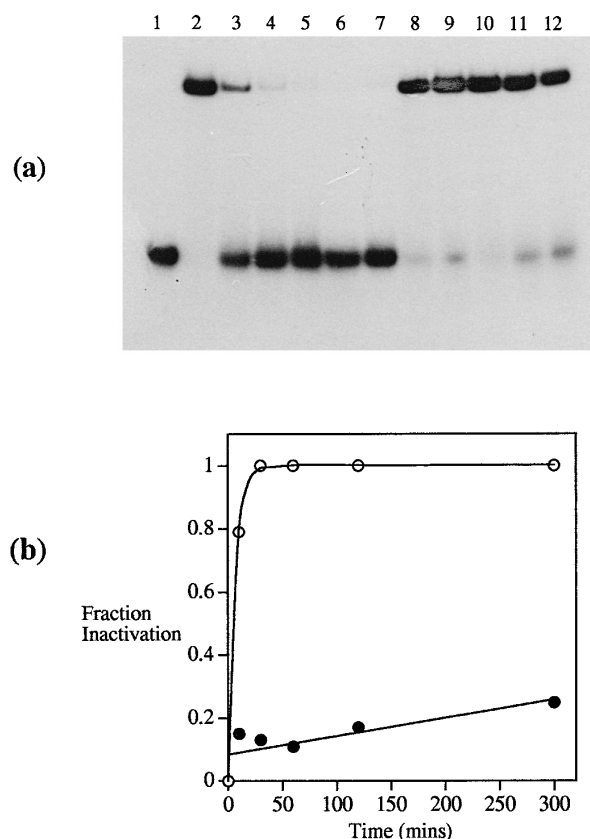


Figure 4. DNA binding analysis of the modified *M. EcoR124I*. (a) Gel retardation analysis following reductive methylation of the MTase and the MTase-DNA complex. Proteins were modified for various times, purified and incubated with an equimolar amount of a ^{32}P -labelled 30 bp DNA duplex (237 nM). Lane 1, no protein; lane 2, unmodified MTase; lanes 3 to 7, modification of free MTase for 10, 30, 60, 120 and 300 minutes; lanes 8 to 12, modification of MTase-DNA complex for 10, 30, 60, 120 and 300 minutes. (b) Inactivation of DNA binding following reductive methylation of the MTase (○) and the MTase-DNA complex (●) as a function of a time of modification. Ratios of free and bound DNA were obtained by densitometry of the autoradiograph shown in (a).

fragments for sequence analysis. Our experiments, therefore, concentrated on the C-terminal domain of HsdS.

HsdS modified by the above procedure was digested with trypsin and the products separated by reverse phase HPLC (C4), giving rise to numerous peaks (Figure 5(a)). Those peaks having apparently high specific activities were further purified on a C8 column. The fractions were counted for radioactivity and those with the highest labelling were analysed by automated Edman degradation. Four tryptic peptides were analysed in this way, together with a peptide (P4) obtained by CNBr cleavage. Of the five peptides sequenced, P4 is the most extensively labelled, followed by P1 and P5. Peptides P2 and P3 are labelled to a much lesser degree but their lysine content is low (Table 2).

The amino acid released at each cycle was determined, together with the amount of radioactivity released at each step (Figure 6). Such a scheme allows comparison to be made of the labelling of individual amino acids in each of the peptides sequenced. Several general observations could be made from analysis of the sequencing data. Firstly, significant quantities of radioactivity were only seen in cycles where modified lysine was expected. The release of radioactivity was invariably accompanied by the presence of a peak not identifiable in the set of standard PTH amino acids. There was also no significant release of any unmodified lysine.

The 3.5 kDa peptide P1 (D190-N220) results from cleavage at R189 on the distal side of the central conserved domain and the peptide extends just into the C-terminal variable region. The amount of labelling at each residue was quantified for the 30 cycles of the Edman reaction. Labelling appears to be fairly uniform at K196, K203 and K210 when account is taken of the decrease in yield per cycle. The region covered by this peptide has been implicated in interactions with the HsdM subunit. However, the lysine residues in this region appear not to be buried in the HsdS-HsdM interface, since all three are modified quite rapidly.

The remaining four peptides span most of the C-terminal variable domain of HsdS. The presence of peptide P2 (K221-R240) in the digest is something of an anomaly as it appears to be the result of tryptic cleavage at N220. Nevertheless, cleavage at this site is helpful for the analysis. This peptide is derived from the proximal side of the C-terminal variable domain and contains two lysine residues, K221 and K236, only the former of which is appreciably labelled.

Tryptic digestion releases two additional 3.5 kDa peptides P3 and P5 (T241-R273 and A324-R356) derived entirely from within the C-terminal variable domain. P3 contains a single lysine (K261) that is quite highly labelled. P5 contains five lysine residues, a number of which must be highly labelled given the specific activity of this peptide (Table 2). It was only possible to sequence the first five residues of this peptide, but this allowed the labelling of the two adjacent residues, K326 and K327, to be determined. From this, it is clear that K327 is at least ten times more highly labelled than its neighbour.

Digestion with CNBr releases the 2 kDa peptide P4 (K297-M313) from the C-terminal variable domain. This peptide contains the most highly labelled lysine of all, K297, together with K308 which is labelled to a significantly lower extent.

Discussion

It is clear that the majority of the lysine residues that are protected by DNA binding are to be found in the HsdS subunit, consistent with the role of this subunit in providing DNA sequence specificity. There are also lysine residues in the HsdM subunit

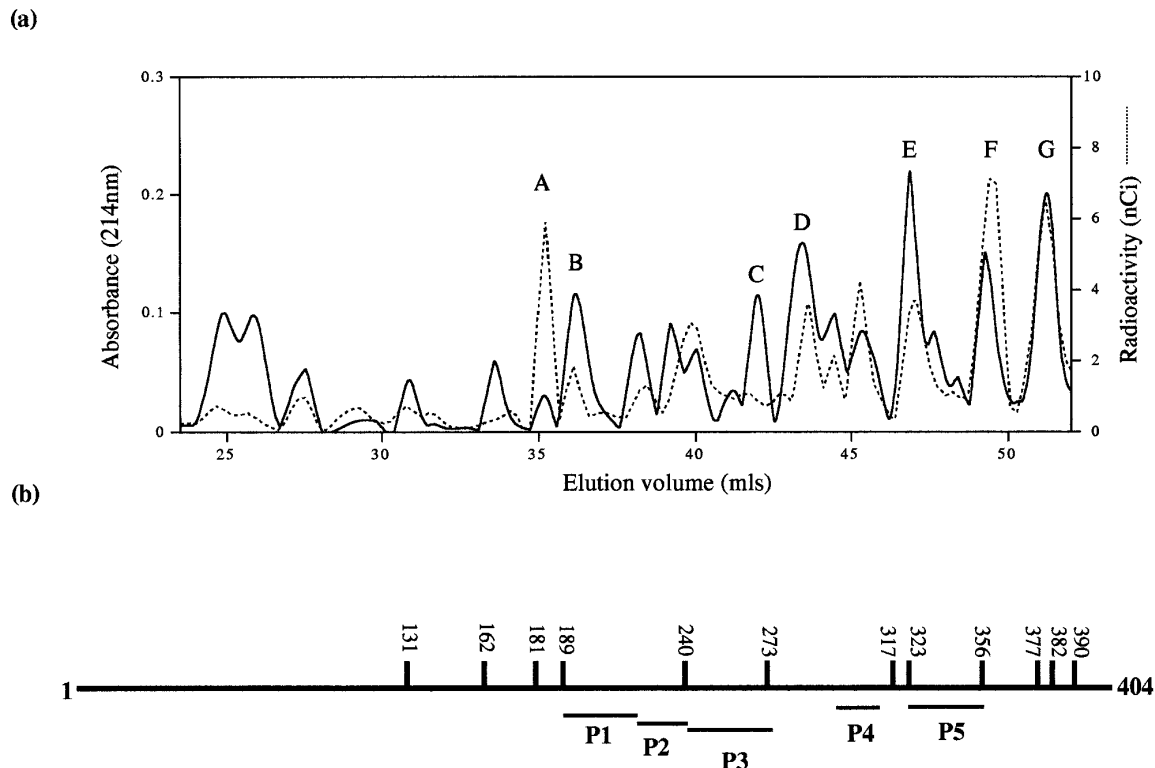


Figure 5. Peptide mapping of the HsdS subunit. (a) Separation of tryptic peptides by reverse phase HPLC (C4). The absorbance at 214 nm and radioactivity were determined for each fraction. Peaks labelled A to G were further purified on a C8 column, and the peaks having the highest radioactivity were sequenced. (b) Potential trypsin digestion sites (arginine residues) in modified HsdS and the location of the sequenced peptides. Peptide P4 was isolated from a CNBr digestion.

that are protected, but to a much lesser extent. These may include residues involved in non-specific interactions with the phosphate groups of DNA, since such interactions are known to extend well beyond the recognition sequence, and may contribute to the stabilisation of the compacted form of the methylase as it wraps around the DNA (Taylor *et al.*, 1994). It would appear that the contacts the HsdM subunit makes with the DNA are not as intimate as those made by the HsdS subunit, which has to make sequence specific contacts with bases in the DNA recognition sequence.

The conserved domains of the HsdS subunit of type I enzymes have been implicated in interactions with the HsdM subunit (Cooper & Dryden, 1994),

and this has been thought to explain the conservation of these sequences within each family. However, analysis of sequences of antirestriction (Ard) proteins reveals sequences which are homologous to only part of the central conserved region (equivalent to the A domain defined by Kneale, 1994), suggesting that this alone may constitute the HsdM binding site (Belogurov & Delver 1995). Some weight is given to this notion from the present work, since we find there is significant labelling within the conserved domains of the HsdS subunit, in particular in and around the region previously referred to as region D (Kneale, 1994; Figure 7). Thus, lysine residues 196, 203 and 210 in this region are quite highly exposed on the surface of the free

Table 2. Location of the labelled lysine residues

Peptide	Sequence	No. cycles ^a	Lysine content	Specific radioactivity ^b (nCi/nmol peptide)	Average modification (per lysine)
P1	D190-N220	30	4	199	0.27
P2	K221-R240	20	2	44	0.11
P3	T241-R273	22	1	39	0.21
P4	K297-M313	17	2	644	1.10
P5 ^c	A324-R356	5	5	308	0.36

^a Number of Edman degradation cycles used to sequence the peptide.

^b Calculated from the radioactivity of the peptide, the specific activity of the reagent and quantitation of the peptide from the sequencing reaction.

^c It was not possible to obtain accurate data for more than five cycles for this peptide.

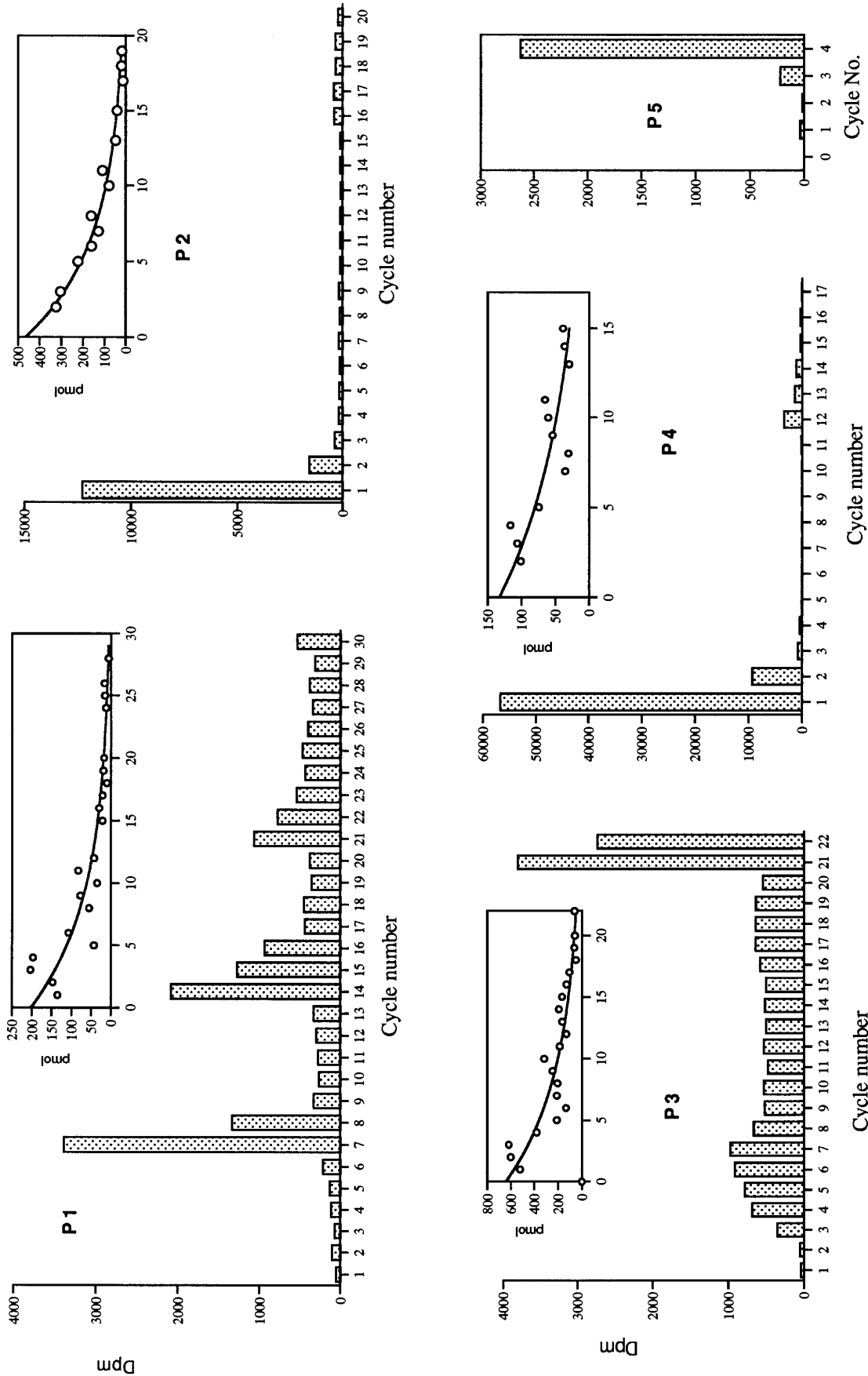


Figure 6. Identification of labelled lysine residues. N-terminal sequencing of peptides following tryptic digestion of pulse-labelled HsdS was performed by automated Edman degradation for each of the peptides P1 to P5. The combination of radioactivity and the pmol released at each cycle allows estimation of the specific radioactivity of individual amino acid residues. Inset: The yield at each cycle where PTH derivatised N,N-dimethyl lysine is released.

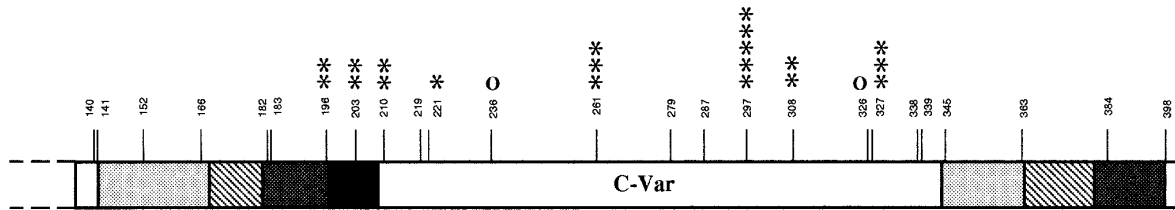


Figure 7. C-terminal sequence of HsdS showing the extent of radiolabelling of lysine residues. The extent of labelling is indicated by the number of asterisks (*). Insignificant levels of radiolabel incorporation are denoted (O). The C-terminal variable domain is indicated (C-Var), together with the central and C-terminal conserved domains (shaded) as described elsewhere (Kneale, 1994).

MTase, and are unlikely to be at the interface between HsdS and HsdM subunits; they are, however, protected in the DNA-protein complex. Moreover, their reductive methylation contributes to the abolition of DNA binding.

But if these lysine residues are involved in DNA binding rather than interactions with HsdM, why should they be conserved between enzymes that have different DNA recognition sequences? We suggest that they are conserved for their contribution to non-specific DNA binding *via* electrostatic interactions with phosphate groups in and around the DNA recognition sequence. This suggests that there are common structural elements in the architecture of the DNA binding domains, despite the absence of any recognisable DNA binding motif. However, lysine residues are not found at equivalent sites in the homologous

(“repeated”) sequences found at the N terminus of the HsdS subunit (Figure 8). Thus, if this region contacts the second HsdM subunit in the methylase, as is believed to be the case, it would not be able to make the non-specific DNA contacts equivalent to those proposed above for the central conserved domain.

The view that the variable domains are responsible for DNA binding, and the conserved domains for protein-protein interactions, could thus be an oversimplification. Clearly, elements of the variable domains confer specificity (presumably *via* hydrogen bonding to the exposed edges of the bases) but there may also be non-specific contacts between the conserved domains and the phosphodiester backbone of the DNA. Much interest has centred on the sequences that are located close to the C-terminal border of the “conserved domains” and the N-terminal border of the “variable domains”, including the LP-GWEW and KEGEVEW “motifs” that are conserved in type IA/IB and type IC families, respectively (Cowan *et al.*, 1987; Gann *et al.*, 1987; Kneale, 1994). It has been suggested that these sequences may play a “structural” role (Chen *et al.*, 1995). On the basis of the arguments outlined above, we suggest that they could also play a role in non-specific DNA binding.

It is less surprising that a number of lysine residues (261, 297 and 327) in the C-terminal variable region are highly accessible to modification, and are implicated in DNA binding; however, other lysine residues in this region are not (residues 236, 326 and to a lesser extent 221). These latter residues are presumably buried (or partially so) within the folded C-terminal variable domain of HsdS, or possibly at the interface with other HsdS domains or with the HsdM subunit. The difference in accessibility (at least tenfold) between lysine 326 and 327 is particularly notable; although the two residues are adjacent in the sequence, they must be located in very different environments

In the absence of a three-dimensional structure for *M. EcoR124I* (or any other type I MTase), site-directed mutagenesis experiments can now be targeted at specific residues of the HsdS subunit which are likely to be involved in interactions with DNA, in order to establish their individual contribution to DNA binding.

CENTRAL	
EcoR124I	YRDQLLSFK KEGEVEW KTLGEIG KWY
EcoDXXI	YRDQLLSFK KEGEVEW KTLGEIG NFT
EcoPrrI	YRDQLLSFK EDGKR KTLGEIG IMKMR
N-TERM	
EcoR124I	YLEKLLD--GVEVEWVTLGSMADIG
EcoDXXI	YLEKLLD--GVEVEWLPLGEITKYE
EcoPrrI	YLEKLM D--GVEVEWLPLSKVFNL R
CONSENSUS	
Central	<u>YRDQLLSFKKEGEVEWKTLGEIG</u> Kxx
N-term	<u>YLEKLLD--GVEVEWLPLG</u> xxxxxx
BOTH	YxxxLLxxxxxxEVEWxxxLGxxxxxx

Figure 8. Conserved sequences found in the D domain of type IC HsdS polypeptides. The rapidly modified lysine residues 196, 103 and 210 are shown in bold. Gaps have been allowed to optimise the alignment between central and N-terminal sequences. The consensus sequence for each region indicates residues that are shared by at least two of the three enzymes (identities, shared between all three are underlined; X denotes that there are no sequence identities at a given site). Residues shared between the two consensus sequences for each repeat are shown below.

Materials and Methods

Protein purification

M.EcoR124I was overexpressed in *E. coli* JM109(DE3) from plasmid pJS4M (Patel *et al.*, 1992). The multisubunit enzyme was purified to homogeneity from crude cell extracts by ion exchange and heparin chromatography (Taylor *et al.*, 1992). The purity and monodispersity of preparations were routinely analysed by reverse phase HPLC, SDS/polyacrylamide gel electrophoresis and by analytical gel filtration on a Superose 12 (10/30) column.

Quantification of the effective specific activity of [³H]formaldehyde

α -Melanocyte stimulating hormone (α -MSH) was purchased from Sigma Chemicals and further purified by reverse phase HPLC on a Waters μ bondpak C18 125 Å 10 μ m column. A sample of 100 nmol of α -MSH was dissolved in 0.5 ml of 10 mM Hepes (pH 7.5). The peptide was modified for three hours by addition of 2 mM [³H]formaldehyde/10 mM sodium cyanoborohydride and then by a second addition of this reagent for a further three hours. Modified peptide was purified again by reverse phase chromatography and the extent of lysine modification was determined by fast atom bombardment mass spectrometry and ¹H NMR spectroscopy. The specific activity of the modified peptide and hence that of a fully modified single lysine residue was calculated from the UV absorbance at 280 nm ($\epsilon_{280} = 7000$) combined with the amount of incorporated radioactivity measured by scintillation counting. Two separate batches of [³H]formaldehyde were used in the various modification experiments and calibrations were carried out routinely on both. The effective specific activity of the two batches were found to be 145 nCi/nmol and 93 nCi/nmol.

Modification of surface accessible lysine residues

A 6.2 μ M solution of MTase was prepared in 50 mM NaCl, 1 mM EDTA, 10 mM Hepes (pH 7.5). The protein modification was initiated by the addition of 20 mM sodium cyanoborohydride, 7 mM [³H]formaldehyde and the mixture incubated at 24°C. At timed intervals 200 μ g samples were withdrawn, quenched with 50 mM glycine and dialysed exhaustively at 4°C with multiple changes into 50 mM NaCl, 1 mM EDTA, 10 mM Tris-HCl (pH 7.5). The concentration of protein was determined from the UV absorbance ($\epsilon_{280} = 160,4000$) and the amount of incorporated radioactivity measured by liquid scintillation counting. The extent of lysine modification at each time point was determined by comparison with the α -MSH standard. Experiments were also carried out where a 30 bp DNA duplex (duplex AB) containing the enzyme recognition site was added at equimolar concentration to the protein to form the DNA-protein complex (Taylor *et al.*, 1993). Reductive methylation was carried out as described above. Complexes from various time points in the reaction were applied to a TSK-GEL DEAE-NPR anion exchange HPLC column equilibrated in 100 mM NaCl, 1 mM EDTA, 10 mM Tris-HCl (pH 7.5). The DNA and protein were separated by application of an increasing gradient of NaCl (100 mM to 500 mM in 15 minutes followed by 500 mM to 1000 mM in five minutes). The extent of lysine modification at each time point was determined as described above. The kinetic data from these experiments were fitted to the

single exponential $y = A(1 - e^{-kt})$ using a non-linear least-squares procedure. This gives the effective rate constant for the modification reaction, and the limit value for the total number of lysine residues accessible.

Fluorography of SDS/polyacrylamide gels

Samples from the time course of each modification reaction were analysed on 13.5% (w/v) polyacrylamide/SDS gels. Gels were electrophoresed for one hour at 125 V, stained with Coomassie brilliant blue, destained and then soaked in Amplify™ (Amersham) for 30 minutes. The gels were then dried at 60°C under vacuum and exposed to preflashed X-ray film at -70°C. The intensity of bands on the fluorograph was analysed by densitometry, a number of exposures being taken to ensure a linear response.

Gel retardation analysis

The DNA binding ability of MTase modified in the presence and absence of DNA was assayed at various time points during the reaction by a gel retardation assay. A 30 bp DNA duplex containing the enzymes recognition sequence (duplex AB) was prepared as described (Taylor *et al.*, 1994). Equimolar amounts of modified protein and ³²P end-labelled duplex were mixed and the free and bound DNA species separated by electrophoresis. The gels were dried under vacuum at 80°C for 30 minutes and exposed to X-ray film. The autoradiographs were analysed by densitometry to quantitate the fraction of DNA bound for each sample.

Pulse labelling

In a typical pulse labelling experiment, 18 nmol of MTase was prepared in 50 mM NaCl, 1 mM EDTA, 10 mM Hepes (pH 7.5) to give a protein concentration of 9 μ M. The protein was modified for ten minutes at 24°C with 20 mM sodium cyanoborohydride/9 mM [³H]formaldehyde. The reaction was quenched with 50 mM glycine and dialysed exhaustively with multiple changes into 50 mM NaCl, 1 mM EDTA, 10 mM Tris-HCl (pH 7.5) at 4°C. The extent of lysine modification was determined from the protein absorbance at 280 nm and the amount of radioactivity incorporated. The ability of modified protein to bind to the DNA recognition sequence was assayed by gel retardation to determine the fraction of protein still active after modification.

Modified protein was adjusted to 8 M urea, 10 mM DTT and briefly incubated at room temperature. The solution was acidified by addition of 100 mM acetic acid and then centrifuged to remove any insoluble material. The protein was applied to semipreparative C4 (300 Å, 5 μ m) reverse phase column (Vydac) equilibrated in 5% (v/v) acetonitrile, 0.05% (v/v) trifluoroacetic acid. The HsdS and HsdM proteins were separated by application of an increasing gradient of acetonitrile (5% to 35% in five minutes then 35% to 75% in 40 minutes). Peaks corresponding to HsdS and HsdM proteins were collected and stored lyophilised at -20°C. Modified HsdS and HsdM protein were redissolved in 8 M urea, 10 mM Hepes (pH 7.5) and the amount of modified lysine determined for each polypeptide from the protein absorbance at 280 nm (HsdS, $\epsilon_{280} = 74,200$; HsdM, $\epsilon_{280} = 43,100$) and the level of incorporated radioactivity.

Prior to peptide mapping, the radiolabelled HsdS was chased with unlabelled reagent mixture to fully modify the remaining sites when exposed in 8 M urea. The

reaction mixture was made 10 mM DTT, incubated briefly at room temperature then acidified with 100 mM acetic acid. The mixture was centrifuged to remove any insoluble material then applied to VYDAC semipreparative C4, 300 Å, 5 µm reverse phase (flow rate was 3 ml/minute, eluent monitored at 254 nm) equilibrated in 5% acetonitrile, 0.05% TFA. The fully modified HsdS protein was eluted by application of an increasing gradient of acetonitrile (5% to 35% in five minutes, then 35% to 75% in 40 minutes), lyophilised and stored at -20°C.

Tryptic cleavage

Typically 2 nmol aliquots of modified pulse-labelled HsdS protein were dissolved in 100 µl of 0.9% (v/v) formic acid. The samples were then diluted fivefold with distilled water and titrated to pH 8.0 by the addition of 35 µl of 2M Tris base. Digestions were then carried out by the addition of 1:20 (w/w) trypsin (Sigma sequencing grade) for 24 hours at 37°C. The protease was added in three additions over the total digestion period. The digests were terminated by the addition of 1 mM DFP and stored frozen until required.

Tryptic digests of HsdS were adjusted to 8 M urea, 10 mM DTT, incubated briefly at room temperature and then acidified by addition of 100 mM acetic acid. Insoluble material was removed by centrifugation and the digests were applied to a BioRad analytical C4 (300 Å, 5 µm) reverse phase HPLC column equilibrated in 2% acetonitrile/0.05% TFA. Peptides were eluted with an increasing gradient of acetonitrile (2% to 35% in 43 minutes followed by 35% to 60% in 20 minutes). A 5% aliquot of each fraction was taken and the level of incorporated radioactivity determined by liquid scintillation counting. All peaks were lyophilised and stored at -20°C.

Peaks from the C4 separation were further purified by HPLC on a ZORBAX R_x-C8 (80 Å, 5 µm) reverse phase column (Dupont). Samples were redissolved in 500 µl of 8 M urea 100 mM acetic acid and applied to the column equilibrated in 2% acetonitrile/0.05% TFA. Bound peptides were eluted with a linear gradient of acetonitrile 2% to 48% in 70 minutes. A 10% proportion of each peak eluted from a single C8 run was taken and the amount of incorporated radioactivity determined. The remainder was lyophilised and stored at -20°C for subsequent N-terminal sequencing.

CNBr cleavage

In a typical digestion 7 nmol of pulse-labelled HsdS protein was dissolved in 200 µl of 70% formic acid. Fresh CNBr was added to a concentration of 25 mM. The mixture was incubated in the dark at room temperature for 18 hours and then lyophilised. The digestion products were redissolved in 100 µl of 70% formic acid and applied to Superose 12 HR (10/30) size exclusion column equilibrated in 70% formic acid. A 5% aliquot was taken from each fraction and analysed by liquid scintillation counting. Fractions were lyophilised and stored at -20°C.

Cleavage at methionine 296 and 313 results in the release of a 2 kDa peptide. Fractions containing this peptide were pooled and further purified by reverse phase HPLC on the ZORBAX C8 column. Elution conditions were the same as for the tryptic peptides. A 10% aliquot was used to determine the amount of incorporated radioactivity of the purified peptide, and the remainder was lyophilised and stored at -20°C.

Protein sequencing

Purified peptides from the C8 column were sequenced by automated Edman degradation on an Applied Biosystems pulsed liquid amino acid sequencer and the PTH derivatised amino acids identified at each cycle by on-line HPLC analysis. In general a fraction of each labelled peptide was sequenced as described above. The remaining proportion was derivatised and cleaved as in the usual sequencing reaction but, after ATZ derivatisation, a proportion of the reaction was diverted to a fraction collector allowing the amount of radioactivity released at each cycle to be determined by liquid scintillation counting.

Acknowledgements

We thank Lesley Canfield for NMR spectroscopy and Nigel Armstrong for mass spectroscopy. Amino acid sequencing was performed at the University of Southampton (courtesy of Lawrence Hunt and Dr M. G. Gore). We are grateful to Dr A. W. Thorne for helpful discussions on aspects of protein chemistry. The work was supported by a project grant and a prize studentship from the Wellcome Trust.

References

- Abadjieva, A., Patel, J., Webb, M., Zinkevich, V. & Firman, K. (1993). A deletion mutant of the type IC restriction endonuclease *EcoR124I* expressing a novel specificity. *Nucl. Acids Res.* **21**, 4435-4443.
- Belogurov, A. A. & Delver, E. P. (1995). A motif conserved among the type I restriction modification enzymes and antirestriction proteins. *Nucl. Acids Res.* **23**, 785-787.
- Bickle, T. A. & Kruger, D. H. (1993). The biology of restriction. *Microbiol. Rev.* **57**, 434-450.
- Chen, A., Powell, L. M., Dryden, D. T. F., Murray, N. E. & Brown, T. (1995). Tyrosine 27 of the specificity polypeptide of *EcoKI* can be UV crosslinked to a bromodeoxyuridine-substituted DNA target sequence. *Nucl. Acids Res.* **23**, 1177-1183.
- Cooper, L. P. & Dryden, D. T. F. (1994). The domains of a type I DNA methyltransferase. *J. Mol. Biol.* **236**, 1011-1021.
- 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 characterization of the methyltransferase from the type I restriction and modification system of *Escherichia coli* K12. *J. Biol. Chem.* **268**, 13228-13236.
- 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. Natl Acad. Sci. USA*, **83**, 9368-9372.
- Gann, A. A. F., Campell, A. J. B., Collins, A. F. W. & Murray, N. E. (1987). Reassortment of recognition domains and the evolution of new specificities. *Mol. Microbiol.* **1**, 13-22.
- 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.
- Jentoft, N. & Dearborn, D. G. (1979). Labelling of proteins by reductive methylation using sodium cyanoborohydride. *J. Biol. Chem.* **254**, 4359–4365.
- Kannan, P., Cowan, C. M., Daniel, A. S., Gann, A. A. F. & Murray, N. E. (1989). Conservation of organization in the specificity polypeptides of the two families of type I restriction enzymes. *J. Mol. Biol.* **209**, 335–344.
- Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. (1994). *HhaI* methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–369.
- Kneale, G. G. (1994). A symmetrical model for the domain structure of type I DNA methyltransferases. *J. Mol. Biol.* **243**, 1–5.
- Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R. J. & Wilson, G. G. (1994). The DNA (cytosine-5) methyltransferases. *Nucl. Acids Res.* **22**, 1–10.
- Lambert, S. F. & Thomas, J. O. (1986). Lysine-containing DNA binding regions on the surface of the histone octamer in the nucleosome core particle. *Eur. J. Biochem.* **160**, 191–201.
- Lundblad, R. L. & Noyes, C. M. (1984). *Chemical Reagents for Protein Modification* vol. I–II, CRC Press, Boca Raton, FA.
- Meister, J., MacWilliams, M., Hubner, P., Jutte, H., Skrzypek, E., Piekarowicz, A. & Bickle, T. (1993). Macroevolution by transposition: drastic modification of DNA recognition by a type I restriction enzyme following Tn5 transposition. *EMBO J.* **12**, 4585–4591.
- Patel, J., Taylor, I. A., Dutta, C., Kneale, G. G. & Firman, K. (1992). High level expression of the cloned genes encoding the subunits and the intact methyltransferase, *M.EcoR124I*. *Gene*, **112**, 21–27.
- Plyte, S. E. & Kneale, G. G. (1991). The role of tyrosines in the DNA binding site of the Pfl gene 5 protein. *Protein Eng.* **4**, 553–560.
- Powell, L. M. & Murray, N. E. (1995). *S*-Adenosyl methionine alters the DNA contacts of the *EcoKI* methyltransferase. *Nucl. Acids Res.* **23**, 967–974.
- Powell, L. M., Dryden, D. T. F., Willcock, D. F., Pain, R. H. & Murray, N. E. (1993). DNA recognition by the *EcoK* methyltransferase. *J. Mol. Biol.* **234**, 60–71.
- Sams, C. F. & Matthews, S. (1988). Diethyl pyrocarbonate reaction with the lactose repressor protein affects both inducer and DNA binding. *Biochemistry*, **27**, 2277–2281.
- Taylor, I. A., Patel, J., Firman, K. & Kneale, G. G. (1992). Purification and biochemical characterisation of the *EcoR124I* type I modification methylase. *Nucl. Acids Res.* **20**, 179–186.
- Taylor, I. A., Watts, D. & Kneale, G. G. (1993). Substrate recognition and selectivity in the type IC DNA modification methylase *M.EcoR124I*. *Nucl. Acids Res.* **21**, 4929–4935.
- Taylor, I. A., Davis, K. G., Watts, D. & Kneale, G. G. (1994). DNA binding induces a major structural transition in a type I methyltransferase. *EMBO J.* **13**, 5772–5778.
- Thomas, J. O. (1989). Chemical radiolabelling of lysines that interact with DNA in chromatin. *Methods Enzymol.* **170**, 369–385.
- Thomas, J. O. & Wilson, C. M. (1986). Selective radiolabelling and identification of a strong nucleosome binding site on the globular domain of H5. *EMBO J.* **5**, 3531–3537.
- Tsugita, A. & Kneale, G. G. (1985). Identification of lysine residues at the binding site of the bacteriophage pfl DNA binding protein. *Biochem. J.* **228**, 193–199.
- Tyndall, C., Meister, J. & Bickle, T. A. (1994). The *Escherichia coli* *pr* region encodes a functional type IC DNA restriction system closely integrated with an anticodon nuclease gene. *J. Mol. Biol.* **237**, 266–274.
- Webb, M., Taylor, I. A. & Kneale, G. G. (1995). Probing the domain structure of the type IC DNA methyltransferase *M.EcoR124I* by limited proteolysis. *J. Mol. Biol.* **280**, 181–190.
- Wilson, G. G. & Murray, N. (1991). Restriction and modification systems. *Annu. Rev. Genet.* **25**, 585–627.

Edited by J. Karn

(Received 18 December 1995; received in revised form 9 February 1996; accepted 14 February 1996)