The EMBO Journal Vol.17 No.23 pp.7118–7127, 1998

Localization of a protein–DNA interface by random mutagenesis

Mary O'Neill, David T.F.Dryden and Noreen E.Murray[1]

Institute of Cell and Molecular Biology, University of Edinburgh, Darwin Building, King's Buildings, Edinburgh EH9 3JR, UK

[1]Corresponding author
e-mail: noreen.murray@ed.ac.uk

The type I restriction and modification enzymes do not possess obvious DNA-binding motifs within their target recognition domains (TRDs) of 150–180 amino acids. To identify residues involved in DNA recognition, changes were made in the amino-TRD of *Eco*KI by random mutagenesis. Most of the 101 substitutions affecting 79 residues had no effect on the phenotype. Changes at only seven positions caused the loss of restriction and modification activities. The seven residues identified by mutation are not randomly distributed throughout the primary sequence of the TRD: five are within the interval between residues 80 and 110. Sequence analyses have led to the suggestion that the TRDs of type I restriction enzymes include a tertiary structure similar to the TRD of the *Hha*I methyltransferase, and to a model for a DNA–protein interface in *Eco*KI. In this model, the residues within the interval identified by the five mutations are close to the protein–DNA interface. Three additional residues close to the DNA in the model were changed; each substitution impaired both activities. Proteins from twelve mutants were purified: six from mutants with partial or wild-type activity and six from mutants lacking activity. There is a strong correlation between phenotype and DNA-binding affinity, as determined by fluorescence anisotropy.

Keywords: methyltransferase/protein–DNA interaction/ random mutagenesis/restriction–modification

Introduction

Proteins can recognize specific nucleotide sequences within DNA molecules by the combined effects of different types of interactions. These include specific patterns of hydrogen bonding, hydrophobic interactions and electrostatic interactions of amino acid residues with the bases and the phosphate backbone of the DNA. Water molecules and the flexibility of proteins and DNA can also aid the recognition process (Rhodes *et al.*, 1996). Specificity can be achieved using α-helices, β-strands or peptide loops to form a backbone structure for the presentation of specific amino acids to the DNA target sequence. Protein–DNA interfaces have been identified in some complexes, and in the case of a phage repressor, in which a helix is used for recognition, a rational change of amino acids on one side of the helix has led to a repressor with a new, but predicted, target sequence (Wharton and Ptashne, 1985). No general recognition code by which amino acids recognize bases exists, but the deduction of one may be possible in the favourable case of the α-helix used by zinc fingers (reviewed in Choo and Klug, 1997). Such a code may be very difficult or impossible to find for other proteins that use more flexible β strands or polypeptide loops to recognize their DNA targets.

Type I restriction and modification (R–M) systems are complex enzymes that recognize a specific nucleotide sequence in DNA by unknown mechanisms. A type I R–M enzyme is made up of three subunits encoded by the genes *hsdR*, *hsdM* and *hsdS*. HsdM and HsdS comprise a methyltransferase (M_2S_1), which on association with HsdR forms an endonuclease ($R_2M_2S_1$). The methyltransferase modifies hemimethylated target sequences produced after each round of DNA replication, while the endonuclease cleaves unmethylated DNA from foreign sources such as viruses (for review see Wilson and Murray, 1991; Redaschi and Bickle, 1996).

The HsdS subunit of type I systems confers sequence specificity. It recognizes an asymmetric bipartite nucleotide sequence; *Eco*KI, for example, recognizes the sequence AAC-N_6-GTGC. HsdS contains two target recognition domains (TRDs), each of ~160 amino acids; the N-terminal TRD interacts with the trinucleotide (5′) component of its target sequence and the C-terminal TRD with the tetranucleotide (3′) component. Each TRD functions independently of the other and they can be interchanged to generate novel hybrid specificities (Fuller-Pace *et al.*, 1984; Nagaraja *et al.*, 1985; Cowan *et al.*, 1989). However, a single TRD is not sufficient for DNA recognition, as a truncated HsdS subunit with only one TRD dimerizes to produce an enzyme which recognizes a symmetric bipartite nucleotide sequence (Abadjieva *et al.*, 1993; Meister *et al.*, 1993).

No type I R–M enzyme has been crystallized and there is negligible information to imply which amino acids within a TRD interact with DNA. Random mutations were introduced into the 5′ region of *hsdS* coding for the 160 amino acids that include the N-terminal TRD, with the aim of identifying those amino acids that confer specificity. In such experiments, it is important that the changes should be made randomly throughout the coding sequence of the TRD in order to avoid any bias imposed by the experimenters. It was anticipated that those amino acids that could be changed without altering the R–M activity would be unlikely to be involved in target recognition. Amino acids for which substitutions resulted in the loss of both restriction and modification activity (i.e. an r^-m^- phenotype) would include those involved in a specific interaction with DNA.

As the result of random mutations, 79 of the 160

residues within the TRD were altered, and 94 of the 101 substitutions failed to lead to the loss of both restriction and modification activities, although some impaired restriction. Most of those rare mutations that resulted in the abolition of R–M activity were within the interval between residues 80 and 110. On the basis of a model derived from sequence alignments, this same interval has been predicted to form the protein–DNA interface comprising two loops flanking a β-strand (Sturrock and Dryden, 1997). The analysis of the region identified by random mutagenesis was extended by using site-directed mutagenesis to change some residues that the model predicts to be of significance. Enzymes from a selection of mutants were purified, and their properties, particularly DNA binding, examined. A strong correlation was found between phenotype and DNA-binding affinity as determined by fluorescence anisotropy.

Results

Random mutagenesis of a TRD

The substrate for mutagenesis by PCR was a plasmid including *hsdM* and *hsdS*, the genes encoding the methyltransferase component of *Eco*KI. The entire nucleotide sequence of the region of *hsdS* subjected to mutagenesis was determined for each plasmid generated from the mutagenesis experiments. Plasmids with one or more mutations were tested to determine whether the HsdS subunit still conferred specificity to the *Eco*KI complex; the $hsdM^+S^+$ plasmid and many derivatives with mutations in *hsdS* restored both restriction and modification activities to an $hsdR^+M^-S^-$ bacterium.

Only a minority of mutant plasmids had a single mutation; most included multiple changes that resulted in two or more amino acid substitutions (see Table I). Where mutants had more than one amino acid change, but remained proficient in restriction and modification, it was assumed that each mutation was benign. It could have been argued that one mutation suppressed the negative effect of the other, but in all cases this seemed unlikely either because the amino acid changes were conservative or, more importantly, because the same mutation in combination with another mutation also failed to produce a phenotypic effect (Reidhaar-Olson and Sauer, 1988). In those instances where double and triple mutations were associated with a defect in restriction and modification, the component single mutations were remade by site-directed mutagenesis and the phenotype conferred by each single amino acid substitution was determined.

Three libraries of random mutations were made. In the first, the region encoding the entire TRD was the target of mutagenesis. One hundred mutants were obtained: 42 residues within the TRD were altered but only four substitutions within the TRD resulted in the loss of restriction and modification activities. These substitutions (S83P, G85W, G91D and C108R) are loosely clustered in the second half of the TRD.

Each half of the TRD was the target of mutagenesis in the two additional libraries. Twenty-eight residues were altered in the amino half of the TRD (residues 1–75) and 21 in the carboxy half (residues 75–160). Three substitutions (T57P, F107S and G141V) resulted in the loss of restriction and modification activities.

In all, substitutions were obtained for 79 amino acid residues within the TRD, only seven of which resulted in the loss of both restriction and modification activities, the phenotype anticipated for mutations that abolish specificity. The mutants, from which these conclusions are drawn, are listed in Table I. The mutations which have little or no effect on restriction and modification, as assessed using the multicopy plasmid as a test system, are widely distributed throughout the TRD and include non-conservative changes. In contrast, the mutations that abolish restriction and modification activity are not randomly located; five of the seven occur in the interval between residues 80 and 110. On the basis of phenotypes therefore, the mutations imply that this interval includes amino acids at the protein–DNA interface. It is remarkable that in the model structure of the TRD of *Eco*KI (Sturrock and Dryden, 1997), the region between residues 84 and 121 is predicted to form a loop–β-strand–loop that interacts with the major groove of the DNA.

Table I. Mutants obtained from random mutagenesis

Mutant[a]	Substitutions[b]	Mutant	Substitutions
r^+m^+		**r^+m^+**	
M5	P75L, S93P, I122V	M153	F125L
M6	P99L, I122V	M157	I147T
M10,15,39	I122V	M159	F125Y
M12	E170G	M160	G109S
M17	I33V I122V	M162	V90A
M23	A46T	M168	H97N, C108S
M25	V15L	M175	H97Y
M27	K72R, K87R, I122V	M176	P99Q
M36	D153G	M177	F118L
M40	S70G	M181	E114V, Y131H
M43	V62G, I122V	M205	L41S, N47S
M54	T26P, I122V, R132Q	M206	I12F, R23Q, I33V
M56	P99L	M217	E69D
M57	S86N, I122V	M225	S2N I12T, K28R
M61	N65D	M253	Y27H, D39G, F54V
M64	E8G, E76G	M255	K64E E69G
M65	S136P	M261	S2N, I49V
M66	S86R, F118S	M262	A3V
M68	I135V	M264	V15A, L59M, N65S
M69	K166E	M271	G9E K29T
M82	I80M, F121L	**r^-m^-**	
M76	M82I	M42	**S83P**, F121L, Y131H
M84	E30D, A81V	M58	**A177E**
M86	S70R, E76G	M59	**G91D**
M88	Q96R	M60	**C108R**
M93	N51S	M77[c]	V18G, S139P
M96	V18A	M78	**G85W**
M97	I168V	M155	**F107S**
M98	S2N, T20A, S119F	M186	S139P, **G141V**
M104	K148N	M193[c]	Y131N, I135T
M121	S129P	M254	V18A, I44V, **T57P**
M129	P99S A123T	**r^-m^+**	
M133	R132Q	M11	L41S, T57M, E170G
M142	K134R	M212	K5E
M145	L98I	M251	Q50L
M146	F121L	M257	K37N
M149	A140V	M268	Q31R, N47S
		M269	A13T, N34Y, L59M

[a]Omitted from the table are: 18 mutants with four or more base changes, 18 with insertions or deletions of bases causing a shift in the reading frame, seven with silent changes and five with chain termination codons. Also, 15 mutations in HsdM are not included.
[b]Residues that appear more than once are underlined; those that cause loss of restriction and modification activity are outlined in bold.
[c]The individual mutations of M77 and M193 do not cause loss of activity.

Site-directed mutagenesis

(a) Of residues within the putative loops. The TRD of the *Hha*I methyltransferase (M.*Hha*I) interacts with the DNA target through both phosphate and base contacts (Cheng and Blumenthal, 1996). An alignment (Figure 1) of the amino acids of the loop–β-strand–loop structure of M.*Hha*I with the amino-TRD of *Eco*KI has been deduced by Sturrock and Dryden (1997). It includes those amino acids of M.*Hha*I that make contacts with bases and phosphate groups of the DNA backbone, and five of the seven substitutions in *Eco*KI that confer an r⁻m⁻ phenotype.

Three residues within this segment of M.*Hha*I (Y242, T250 and Y254) interact with the phosphate backbone. Fortuitously, the residue in *Eco*KI (F107) that aligns with Y254 had already been changed to serine in M155 (Table I). This change is one of the seven that confers an r⁻m⁻ phenotype. Substitutions were made for the amino acids that align with Y242 and T250. Both substitutions, H95L and S103E, abolished the activity of *Eco*KI.

In M.*Hha*I, residue R240 makes contact with the guanine residue 5′ to the target cytosine (Klimasauskas *et al.*, 1994). *Eco*KI has a lysine residue at position 92 that could play a similar role to R240 in M.*Hha*I. Arginine and isoleucine were substituted for this lysine residue. The very conservative change (K92R) had a small effect on the restriction phenotype, and the less conservative one (K92I) abolished restriction and modification.

(b) Of other residues. Two substitutions at positions remote from the protein–DNA interface abolished the restriction and modification activities. The nature of these two substitutions (T57P and G141V) suggests that they could impose major changes. Both T57P and G141V were substitutions identified in multiple mutants (M254 and M186, respectively) and therefore the component mutations were remade. Alternative changes were made within each of the two codons at the same time: T57P/A and G141V/A. For residue 57, only the change from threonine to proline abolished both restriction and modification activities, whereas either of the substitutions at position 141 abolished both. In the model of Sturrock and Dryden (1997), G141 is located within a region predicted to be α-helical and the loss of glycine is likely to perturb the structure.

All substitutions and their associated phenotypes are summarized in Figure 2.

Additional phenotypic tests

Some mutants (K5E, K37N, Q50L, T57A, I135T and S139P) were shown to be defective in restriction, although they retained the ability to modify DNA (r⁻m⁺). This is an unexpected phenotype for a mutation in *hsdS*, but the phenotypes shown in Table I were derived from tests in which *hsdM* and *hsdS* were present on a multicopy plasmid, and *hsdR* was in the bacterial chromosome. In these tests it is likely that an abnormally high level of the modification enzyme could interfere with restriction. Some *hsdS* mutations were therefore transferred to the bacterial chromosome and the phenotypes reassessed when each of the three genes, *hsdR*⁺, *hsdM*⁺ and *hsdS*⁻, was present as a single copy (Table II). The mutants analysed represent the three different phenotypic classes.

```
                    G P1       P2  P4
(228-282)  EEEEEE        EEE                          EE        HHHHHH
HhaI       RLGIVG--KGGQGERIYSTRGIAITLSAYGGGIFAKT---GGYLVNGKTRKLHPRECARV
            ::: :  |:  |.. .   .:. :::|: | : :..    |::.: ....|.... :.:
EcoKI      VIAMSSGSKSVVGKSAHQHLPFECSFGAFCGVLRPEKLIPSGFIAHFTKSSLYRNKISSL
(79-138)   EEEEE         EEE                          EE        HHHHHH

           LOOP 1  β-STRAND   LOOP2
```

Fig. 1. Part of the alignment of the amino-TRD of *Eco*KI with that of M.*Hha*I, showing the agreement between secondary structure despite the low sequence identity of 14% over the entire TRD (Sturrock and Dryden, 1997). Residues defined in bold typeface are those that contact the phosphate groups in M.*Hha*I and the proposed corresponding residue in *Eco*KI. Y242 contacts the first phosphate group, T250 the second phosphate and Y254 the fourth. Other phosphates within the target sequence are contacted by amino acids outside the TRD. R240 in the β-strand contacts the guanine 5′ to the target cytosine. E denotes β-strand and H denotes α-helix derived from M.*Hha*I (Klimasauskas *et al.*, 1994) or the structural prediction for *Eco*KI (Sturrock and Dryden, 1997).

Those mutations associated with either an r⁻m⁺ or an r⁻m⁻ phenotype in the previous test (Table I) retained their respective phenotypes. Several mutations previously classified as r⁺m⁺ (V.Doronina, personal communication) conferred an r⁻m⁺ phenotype in the tests where the ratio of HsdR to HsdM and HsdS is expected to be normal (Table II). The m⁺ phenotype indicates that these mutations do not abolish specificity but have a more subtle effect on restriction. It is possible, for example, that the enzyme binds to its target sequence but is unable to undergo the ATP-dependent conformational change associated with the formation of the tight complex that is required for DNA translocation (Bickle *et al.*, 1978; Powell *et al.*, 1998b).

Protein molecular weight determination and AdoMet binding of S-adenosylmethionine

Proteins selected to represent the three phenotypes were purified using the method developed for the wild-type enzyme. These included those from five r⁻m⁻ mutants (G91D, K92I, H95L, S103E and F107S) where the substitution was predicted to be at the protein–DNA interface. Low levels of protein and loss from proteolysis prevented purification of the protein containing the H95L (r⁻m⁻) change. No alterations in the purification procedure were required for the other proteins, which suggests that these amino acid changes did not cause any major folding defects within each subunit. The molecular weights of the proteins were examined by analytical gel filtration. Some protein preparations were found to be mixtures of the M_2S_1 trimer and the M_1S_1 dimer (Table III). The presence of large amounts of M_1S_1 may be an artefact due to the small scale of the protein purification; the first preparation of the protein containing the substitution K92R was predominantly M_1S_1, and the second preparation was predominantly M_2S_1. However, it is important to note that the dimeric form of *Eco*KI methyltransferase, although inactive, can still bind DNA. It exists in equilibrium with the active trimeric form, with a dissociation constant estimated to be 15 nM in the absence of DNA (Dryden *et al.*, 1997). Binding of DNA shifts the equilibrium towards the trimeric form, which binds with a higher affinity than the dimeric form (Powell *et al.*, 1998a).

UV crosslinking of tritiated *S*-adenosylmethionine (AdoMet) was used to confirm that each protein was able to bind the cofactor (Figure 3). None of the proteins showed any significant difference in the amount of cofactor

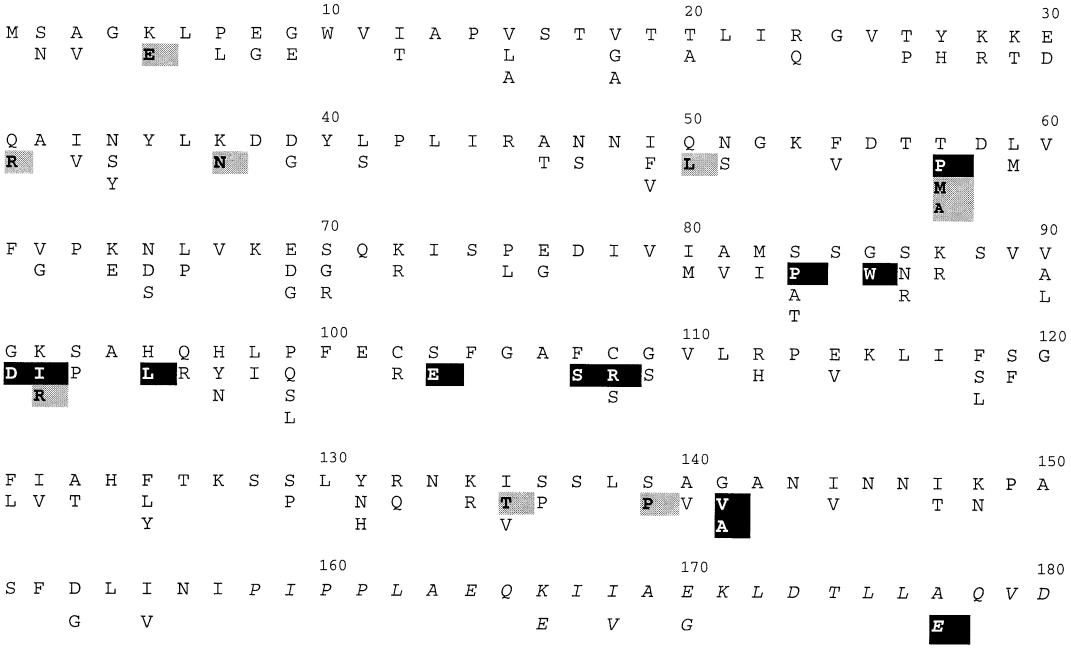

Fig. 2. A summary of the substitutions analysed in the amino-TRD of *Eco*KI. Sequence comparisons suggest that the carboxy-end of the TRD is at residue 157 (Gann *et al.*, 1987). Residues 158–180 were altered when the *Bam*HI–*Sal*I fragment was the target for mutagenesis (see Materials and methods). The substitutions are listed in Table I. Residues shaded in black are those that when altered abolish restriction and modification activity. The residues shaded in grey are those that are associated with a defect in restriction but not modification (r^-m^+). The remaining mutations have no effect on activity.

Table II. Phenotypes associated with amino acid changes

Mutant	Phenotype	
	Mutation on plasmid	Mutation on chromosome
K5E	r^-m^+	r^-m^+
K37N	r^-m^+	r^-m^+
Q50L	r^-m^+	r^-m^+
S86N	r^+m^+	r^-m^+
K92R	r^+m^+	r^-m^+
K92I	r^-m^-	r^-m^-
H97Y	r^+m^+	r^+m^+
P99Q	r^+m^+	r^-m^+
C102R	r^+m^+	r^+m^+
F107S	r^-m^-	r^-m^-
G109S	r^+m^+	r^+m^+
I135T	r^-m^+	r^-m^+
S139P	r^-m^+	r^-m^+
A140V	r^+m^+	$r^{+/-}m^+$

binding. The trimeric and dimeric forms of *Eco*KI can both bind AdoMet efficiently (Powell *et al.*, 1998a).

Fluorescence polarization analysis of DNA binding

Table III shows the values obtained for the K_d and maximum anisotropy change for the binding of DNA by the *Eco*KI methyltransferase. The average anisotropy value in these experiments for the hexachlorofluorescein-labelled DNA in the absence of protein was 0.110. Most protein–DNA complexes had the same maximum anisotropy change, indicating a similar protein:DNA ratio, assumed to be 1:1. The protein containing the change G109S showed a larger anisotropy change, which suggests the presence of larger complexes. Figure 4 shows some typical fluorescence anisotropy titration results.

Amino acid substitutions that confer an r^+m^+ phenotype when the *Eco*KI genes are expressed from a multicopy plasmid, and either an r^-m^+ or r^+m^+ phenotype when expressed from single copies of the genes, do not prevent DNA binding. The K_d values are 0.7–52 times that of the wild-type enzyme (Table III), but the higher K_d values are usually associated with protein preparations that are mixtures of the dimer and trimer. Although these proteins show up to a 50-fold reduction in DNA-binding affinity relative to the wild type, they are clearly distinguishable from the majority of proteins, which show no activity *in vivo.*

Of the proteins which show an r^-m^- phenotype, all except T57P show no binding to DNA, even at the highest protein concentrations. The protein containing the substitution T57P, though devoid of activity *in vivo*, displays DNA-binding affinity *in vitro* (Figure 4). The remainder of the proteins displaying an r^-m^- phenotype *in vivo* are unable to bind to DNA at concentrations of up to 500 nM. This suggests that their affinity for DNA must be at least in the micromolar range.

Gel retardation analysis of DNA binding

The technique of gel retardation can help to stabilize weak protein–DNA interactions by restricting the free diffusion of protein–DNA complexes during gel electrophoresis: the 'cage' effect (Fried, 1989). We observed DNA binding by all of the proteins analysed, including those that did not show any DNA binding in the fluorescence polarization assay. It was found that proteins which retained an r^+m^+ or $r^{+/-}m^+$ phenotype, even when expressed from genes on the chromosome, bound to the DNA as well as the wild-type enzyme. The only exception to this was the protein containing the substitution G109S, which showed 17.5-fold weaker binding despite being active *in vivo*. This reduced affinity should be at least partly due to the

Table III. DNA binding as determined by fluorescence anisotropy and gel retardation

Methyltransferase Mutation	Protein purity[b]	Putative location of amino acid[c]	Fluorescence anisotropy			Gel retardation[a]		Phenotype	
			Maximum anisotropy change	K_d [nM][d]	Free energy difference kJ/mol	K_d [nM]	Ratio of K_d for non-specific to specific DNA	*hsd* genes on chromosome	*hsd* genes on plasmid
Wild type	M_2S_1		0.056	7.1		2	20	r^+m^+	r^+m^+
T57P	M_2S_1	N-terminal	0.055	84.7	6.1	15	1.6	r^-m^-	r^-m^-
S86N	$M_1S_1 > M_2S_1$	Loop 1	0.059	103	6.3	10	4	r^+m^+	r^-m^+
G91D	M_2S_1	Loop 1	0	nd[e]	>>12[f]	100	–	r^-m^-	r^-m^-
K92I	M_2S_1	Loop 1/β-strand	0	nd	>>12	25	6	r^-m^-	r^-m^-
K92R	M_2S_1	Loop 1/β-strand	0.061	372	9.8	6	20	r^+m^+	r^-m^+
H97Y	M_2S_1	Loop 2	0.062	5.1	–0.8	2	12.5	r^+m^+	r^+m^+
P99Q	$M_1S_1 > M_2S_1$	Loop 2	0.059	81	6.0	22	1.8	r^+m^+	r^-m^+
S103E	M_2S_1	Loop 2	0	nd	>>12	30–100	>2–7	r^-m^-	r^-m^-
F107S	$M_1S_1 = M_2S_1$	Loop 2	0	nd	>>12	22	6.3	r^-m^-	r^-m^-
G109S	$M_1S_1 > M_2S_1$	Loop 2	0.095	203	8.3	35	1.1	r^+m^+	r^+m^+
A140V	M_2S_1	α-helix	0.062	16.2	2.0	2	20	r^+m^+	$r^{+/-}m^+$
G141A	$M_1S_1 > M_2S_1$	α-helix	0	nd	>>12	90	1.8	r^-m^-	r^-m^-

[a]All gel retardation experiments were performed at least twice.
[b]The presence of M_1S_1 is indicated in some protein preparations. Its presence is influenced by the protein purification procedure and does not necessarily reflect an intrinsic fault in the protein due to the amino acid substitution.
[c]See Figure 1.
[d]Errors in all fitted anisotropy parameters are ±10%.
[e]nd, no detectable binding at up to 500 nM protein.
[f]Calculated assuming a K_d of at least 1μM.

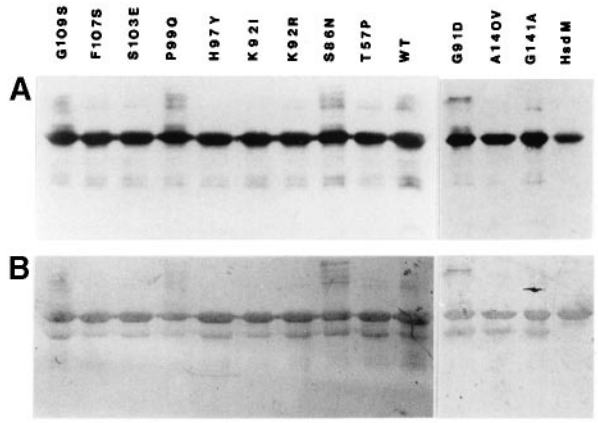

Fig. 3. AdoMet binding assay. Proteins were tested for their ability to bind [^{3}H]AdoMet by UV crosslinking of the co-factor to the protein. The enhanced autoradiogram shows crosslinking predominantly to one subunit (**A**). This was identified as HsdM by Coomassie Blue staining of the total blotted protein (**B**).

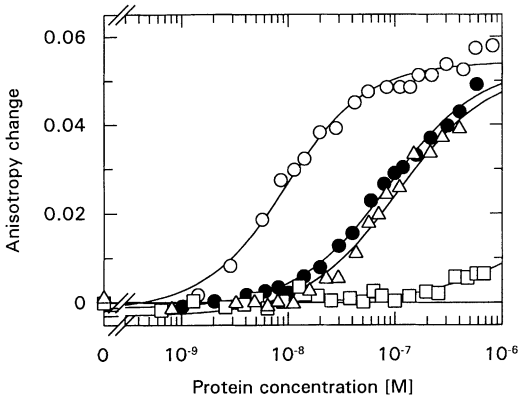

Fig. 4. The increase in fluorescence anisotropy for the labelled DNA duplex containing an *Eco*KI binding site as a function of protein concentration. Data and fitted curves are shown for the wild-type enzyme (○) and the proteins with amino acid changes T57P (●), S86N (Δ) and K92I (□). The data for proteins H97Y and A140V are similar to the wild-type data, the data for P99Q and G109S are similar to those for T57P and S86N, and all of the other proteins show anisotropy changes similar to that for K92I (see Table III for phenotypes).

low amount of trimeric methyltransferase in this protein preparation. Proteins isolated from mutants that showed an r^-m^+ phenotype when the *hsd* genes were located in the chromosome showed 1.5- to 11-fold weaker binding than the wild-type enzyme. Proteins which were inactive *in vivo* showed a 7.5- to 50-fold reduction in binding affinity. These trends are comparable with the results obtained with fluorescence anisotropy. In general, there was a good correlation between activity *in vivo* as assessed from phenotypes and the binding affinities determined *in vitro*.

The degree of discrimination between DNA containing the *Eco*KI target sequence and non-specific DNA was also measured by gel retardation. It was found that the ability to recognize the *Eco*KI target sequence was reduced for proteins which bound more weakly to the target sequence. The binding of non-specific DNA by the wild-type enzyme was too weak to observe by the fluorescence anisotropy technique under conditions matching the gel retardation experiment (Powell *et al.*, 1998a). This suggests that the cage effect may be responsible for the observed binding affinities for non-specific DNA.

Discussion

Restriction enzymes must recognize their target sequences with great precision, otherwise the cutting of modified DNA as the result of recognition errors could have severe consequences. Nevertheless, no structural motif characteristic of target recognition has been identified. For type I R–M systems, the most definitive experiments merely correlate entire TRDs of 150–180 amino acids with the recognition of the tri- or tetra-nucleotide component

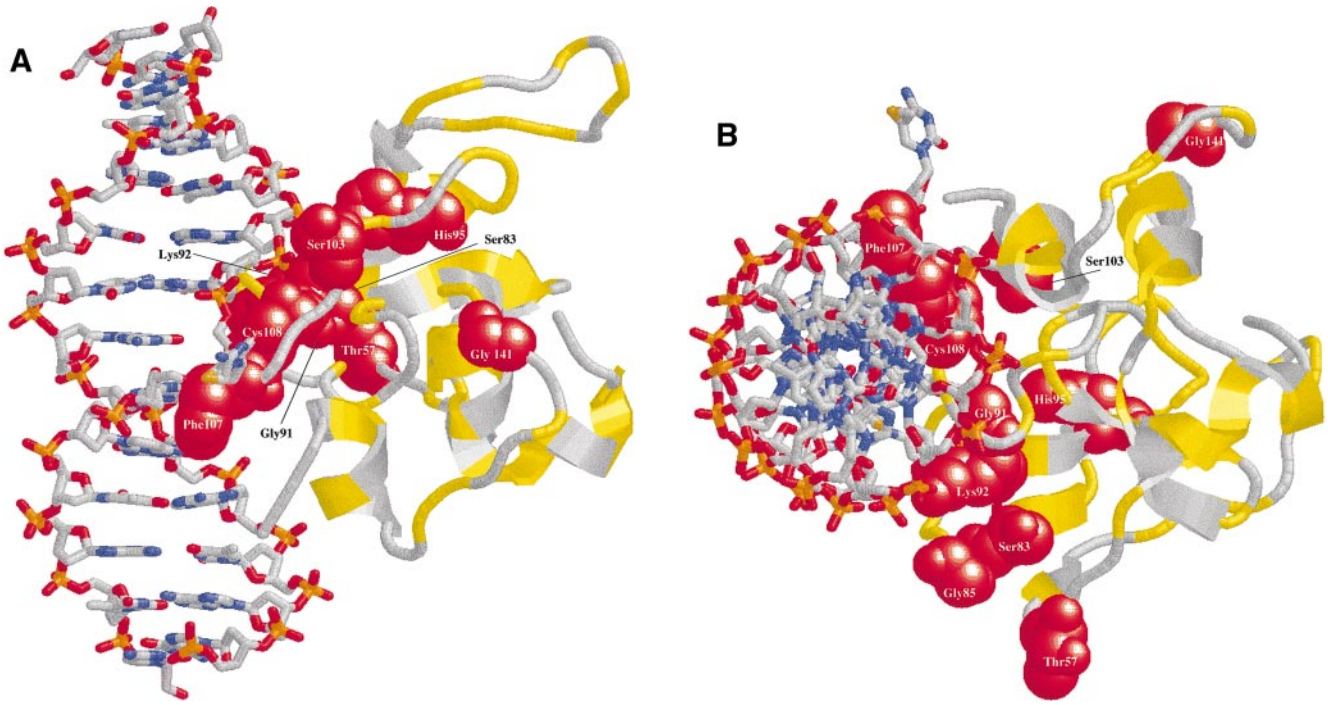

Fig. 5. A model of amino acids 43–157 of the amino-TRD of *Eco*KI interacting with DNA as proposed by Sturrock and Dryden (1997). (**A**) Side-view; (**B**) bird's-eye view. Residues in red are those mutations that cause the loss of restriction and modification activity (Table I and Figure 2). H95 and S103 were the target of site-directed mutagenesis. Residues in yellow are those that do not affect activity. The DNA structure is from the complex of M.*Hha*I with its DNA target sequence (Klimasauskas *et al.*, 1994) and therefore shows an extra-helical cytosine rather than an adenine which would be the methylation target of *Eco*KI. The part of the TRD omitted from the model does not include any changes that cause loss of activity.

sequences (Cowan *et al.*, 1989). We therefore applied random mutagenesis to the segment of the specificity genes of *Eco*KI that is responsible for the recognition of the 5′-trinucleotide component of the target sequence. Random mutagenesis was chosen with the aim of identifying those amino acid substitutions that were innocuous as well as those that impaired specificity. In this context, it is noteworthy that, for the majority of amino acid substitutions, the genetic tests failed to detect any effect on the restriction and modification phenotypes. Unexpectedly common, but not obviously localized, were substitutions in HsdS that only impaired restriction but not modification. Such substitutions could affect the interaction of HsdS with HsdR (Zinkevich *et al.*, 1992; Weiserova and Firman, 1998) or they could influence the interaction of the complex with its DNA target.

The structure of a complex of M.*Hha*I with its target sequence shows that two loops, and the β-strand between them, interact with DNA (Klimasauskas *et al.*, 1994). This region includes several very short conserved amino acid sequences which are found in the TRDs of C5-methyltransferases (Cheng and Blumenthal, 1996; Lange *et al.*, 1996). Comparisons of the sequences of 51 TRDs of type I R–M systems have resulted in a multiple sequence alignment that supports a distant relationship between the structure of the type-I TRDs and the known structure of the TRD of M.*Hha*I, a monomeric DNA methyltransferase (Sturrock and Dryden, 1997). Of the 51 TRD sequences examined, the N-terminal TRD of *Eco*KI shows the closest similarity to the TRD of M.*Hha*I, sufficient to suggest that *Eco*KI might, like M.*Hha*I, interact with DNA via two loops and a β-strand. On the basis of the alignment, these structures are predicted to occur between amino acids 84 and 121 of the N-terminal TRD of *Eco*KI. The initial aim of the random mutagenesis was to identify amino acids close to the DNA, but the data are also relevant as a test of the predictive value of the model.

The majority of the mutations that conferred an r^-m^- phenotype, the phenotype anticipated for mutations leading to loss of specificity, were loosely clustered. The significance of this finding is enhanced when the substitutions are examined in the light of the structural model of Sturrock and Dryden (1997). In the context of this structural model, the loosely clustered mutations that confer an r^-m^- phenotype are preferentially located at the protein–DNA interface (Figure 5); amino acid residues 91, 92, 103 and 107 appear close to the interface and our data have shown them to be important for DNA binding. Chen *et al.* (1995) found that tyrosine 27 in the HsdS subunit of *Eco*KI could be crosslinked to DNA, implying its close proximity to DNA. However, replacing tyrosine with either cysteine or phenylalanine does not confer an r^-m^- phenotype (M.O'Neill, unpublished data). Y27 is not within the part of the TRD included in the model of Sturrock and Dryden (1997), and the phenotype of mutants with the changes Y27F and Y27C indicates that the tyrosine residue is not relevant to DNA specificity. Taylor *et al.* (1996) have reported experiments in which the exposed lysine residues in *Eco*R124I were identified by chemical modification. Lysine residues 261, 297 and 327 in the carboxy-TRD were susceptible to modification especially in the absence of bound DNA. K297 is the most strongly modified residue and it lies within the second of the two proposed loops (Sturrock and Dryden,

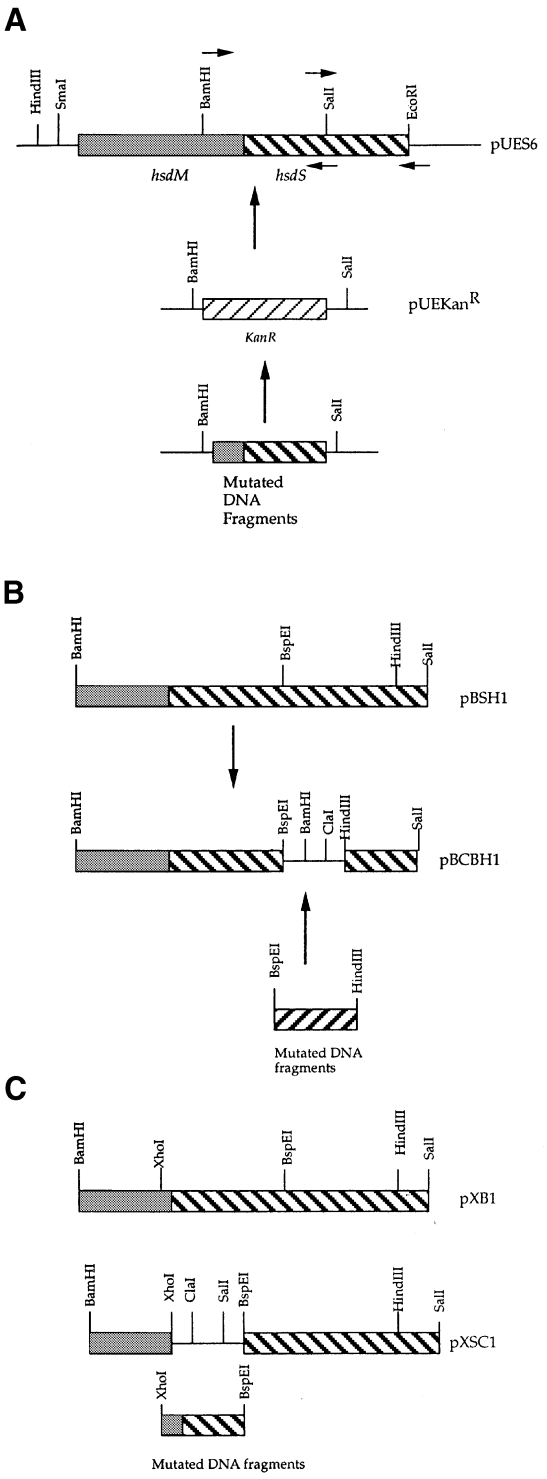

Fig. 6. Plasmids used in random mutagenesis. (**A**) A *Sma*I–*Eco*RI fragment encoding *hsdMS* was cloned in pUC18 to give pUES6. Arrows indicate the PCR primers used to amplify the DNA and make the *Eco*RI and *Sal*I sites. (**B**) A novel *Bsp*EI site was made within the *Bam*HI–*Sal*I fragment of pUES6 to give pBSH1. The 233 bp *Bsp*EI–*Hin*dIII fragment was replaced with a 18 bp stuffer fragment to give pBCBH1. The replacement of the stuffer with mutated DNA fragments was confirmed by screening for loss of the *Cla*I site in pBCBH1. (**C**) A novel *Xho*I site was made within the *Bam*HI–*Sal*I fragment of pBSH1to give pXB1. The 230 bp *Xho*I–*Bsp*EI fragment was replaced with another 18 bp stuffer fragment to give pXSC1 and substituted DNA fragments were identified in the same way as were pBSH1 and pBCBH1.

1997). The three lysine residues identified by chemical modification are conserved in the amino TRD of *Sty*SKI (Thorpe *et al.*, 1997), which recognizes the same target sequence as the carboxy-TRD of *Eco*R124I. The circumstantial evidence that is available therefore is consistent with the present data that implicate interaction of the putative loop–β-strand–loop with the target sequence.

The T57P and the G141A substitutions are outside the confines of the protein–DNA interface predicted by the model; nevertheless, they confer an r^-m^- phenotype. This may indicate that the model has incorrectly identified the interface or that the model is incomplete. However, the protein containing T57P may have a catalytic defect since the protein has the same subunit structure as the wild-type enzyme and can still bind DNA. In addition, it was found that the substitution T57A did not affect modification, implying that DNA specificity was maintained. Therefore, there is no reason to implicate residue T57 in DNA specificity.

The effect of the substitution of glycine 141 with alanine is more difficult to explain. The protein with this substitution was purified largely in the M_1S_1 form, but this does not explain the defect in DNA binding. The inactivity may be attributed to a severe defect in the structure of the dimer, an indirect conformational change affecting the protein–DNA interface or a direct effect at the interface. The last of these three alternatives would imply that the model, based on the relationship with M.*Hha*I, incompletely defines the protein–DNA interface.

Comparing the normal *Eco*KI methyltransferase with proteins that contain amino acid substitutions but can still bind to their target sequence, one can calculate the free energy difference for DNA binding between the two proteins. The free energy difference is rather small, equivalent to the loss of at most one hydrogen bond in the protein–DNA complex. This suggests that any structural perturbation due to the amino acid change is small. The free energy changes, and by implication any structural perturbation at the protein–DNA interface, are much larger for amino acid changes that abolish DNA binding.

We conclude that the loss of DNA binding in some of our mutants is due to the disruption of the protein–DNA interface by a combination of steric hindrance, electrostatic interactions and loss of hydrogen bonding. The results of the random mutagenesis procedure on the amino-TRD of *Eco*KI show that most of the mutations leading to an r^-m^- phenotype were confined to a limited region of the TRD which had been predicted to interact with DNA (Sturrock and Dryden, 1997). Mutations that did not alter the phenotype were spread throughout the TRD. Further site-directed mutagenesis of the putative DNA-binding region produced changes in the phenotype consistent with the structural model. Amino acid changes that had little effect on the modification phenotype, r^+m^+ or r^-m^+, had little effect on DNA binding. Four amino acid changes in the predicted DNA-binding region that abolished modification activity *in vivo* prevented DNA binding *in vitro*. Most of the amino acid substitutions support the structural model and the methods used to derive it. Our study opens the way for further analysis of DNA sequence recognition by type I R–M enzymes and identifies other amino acids potentially involved in the formation of an active protein–DNA complex.

Materials and methods

Bacterial strains

All the bacterial strains used were derivatives of *Escherichia coli* K-12. C600 (Appleyard, 1954) was used as a restriction-proficient (r^+m^+), λ-sensitive strain. Two r^-m^- derivatives of 71–18 (Messing *et al*., 1977) were used for complementation tests. In the first, NM522 (Gough and Murray, 1983), *hsdM* and *S* are deleted; in the second, NM521 *kan* (this work), the *kan*^{r} 'Genblock' of Pharmacia, is substituted for a segment of *hsdM* and *S*. The *kan*^{r} marker facilitated monitoring the transfer of *hsdS* mutations to the *hsd* operon in the bacterial chromosome. A *sup*° Δ (*hsd*) strain, NM679 (King and Murray, 1995), was used to select λ*hsd* phages lacking amber mutations (Webb *et al*., 1996), and as the host for propagating $hsdM^+S^-$ plasmids for the production of mutant methyltransferase for purification. The *hsdR endA recA* strain, DH5α (Grant *et al*., 1990), was used as a competent host for the recovery and amplification of plasmid DNA for sequence analysis following random mutagenesis; XL1-Blue was provided for use with the Quickchange™ mutagenesis kit of Stratagene.

Plasmids

The plasmid used for analysing random mutations in segments of *hsdS* included both *hsdM* and *S*, to enable tests for K-specific modification, but it required new restriction sites to permit substitution of mutated DNA segments for the wild-type sequences. New targets were incorporated, none of which altered an amino acid sequence. To ensure that all members of each random library included the mutagenized DNA, the rationale of displacing a readily identifiable 'stuffer' fragment, rather than the wild-type sequence, was adopted.

The *hsdM* and *S* genes were subcloned from pJFMS (Dryden *et al*., 1993) using the PCR method of overlap extension (Ho *et al*., 1989). A *Bam*HI–*Sal*I fragment and a *Sal*I–*Eco*RI fragment (see Figure 6A) were amplified by the PCR, and the two fragments were joined in a second PCR reaction. The resulting *Bam*HI–*Eco*RI fragment was cloned in pUC18. The *Bam*HI site is a natural sequence within *hsdM*, *Sal*I provides a new target within the centre of *hsdS* and *Eco*RI is a new target that separates the *hsd* genes from irrelevant bacterial DNA present in pJFMS. The *Sma*I–*Bam*I fragment including *hsdM* was isolated from pJFMS and added to reconstruct the *hsdMS* sequence in pUES6 (Figure 6A).

A *Bam*HI–*Sal*I fragment including kan^r replaced the *Bam*HI–*Sal*I fragment of pUES6. The *kan*^{r} gene (Pharmacia) from pGEM3 *kan* (a gift from F.Fuller-Pace) was first subcloned as a *Pst*I fragment in pBluescript to provide the flanking *Bam*HI and *Sal*I targets. Mutagenesis of the sequence coding for the entire amino-TRD commonly resulted in multiple amino acid changes. The target for mutagenesis was therefore split into two segments of equal size in order to maximize single mutations. To permit this, a unique *Bsp*EI site was introduced (pBSH1 in Figure 6B). pBSH1 was made by recloning the *hsdMS* genes in a derivative of pUC18 that lacked the *Hin*dIII target in the polylinker. An 18 bp stuffer fragment including diagnostic targets then replaced the *Bsp*EI–*Hin*dIII fragment of *hsdS*.

In pXB1 (Figure 6C), a unique *Xho*I site was generated within a sequence only 30 bp from the 3′ end of *hsdM* of pBSH1. This permitted mutagenesis of the 5′ end of *hsdS* in the presence of very little *hsdM* sequence. The *Xho*I–*Bsp*EI fragment in pXB1 was replaced with an 18 bp stuffer fragment.

pJES23 has the *Sma*I–*Eco*RI fragment from pUES6 cloned in the expression vector pJF118 (Furste *et al*., 1986).

DNA manipulations

Klenow enzyme, T4 DNA ligase and all the restriction enzymes were supplied by NEB Biologicals, alkaline phosphatase by Epicentre Technologies and the *Taq* polymerase ('Red Hot' polymerase) by Advanced Biotechnologies.

Plasmid DNA was isolated from DH5α or XL1-Blue strains using Qiagen Mini-prep kits or Nucleon Mini-prep kits (Amersham), and DNA sequences were determined using a Perkin-Elmer ABI Prism 377 DNA Sequencer.

Site-specific mutations were made, unless otherwise stated, by the 'Quick Change' site-directed mutagenesis kit (Stratagene).

Random mutagenesis using PCR

The frequency with which mutations are found in a segment of DNA is a function of the polymerase error rate and the number of cycles in the PCR (Eckert and Kunkel, 1991). Therefore, the frequency of mutation can be manipulated experimentally by increasing the number of cycles and/or increasing the error rate of the polymerase. This latter factor can be adjusted by altering parameters such as the length of time at the extension temperature (72°C) (Zhou *et al*., 1991), the addition of transition metal ions, e.g. Mn^{2+} (Leung *et al*., 1989), and the dNTP composition and concentration (Spee *et al*., 1993). Two methods were used to generate random mutations: the first required the addition of $MnCl_2$ to the reaction and was adapted from Leung *et al.* (1989). A PCR comprising 10 ng of template, 200 μM dNTPs and 200 nM primers was supplemented with 0.5 mM $MnCl_2$. Reaction conditions were 35 cycles at 95°C for 1 min, 58°C for 1 min, 72°C for 3.5 min. The second method, using limiting dNTPs and a dITP supplement, was carried out as described in Spee *et al.* (1993), although the PCR conditions were as above.

Generating random mutations within the TRD

All the mutants are listed in Table I. The first target for mutagenesis, the 766 bp *Bam*HI–*Sal*I fragment, generated mutants 1–100. One PCR used a supplement of $MnCl_2$ and the second was prepared using dITP and limiting dNTPs. DNA obtained from the mutagenic PCRs was digested with *Bam*HI and *Sal*I and ligated in pUEKan^R. Ampicillin-resistant transformants of DH5α that had lost resistance to kanamycin were presumed to have the correct fragment. Analysis of the DNA of the transformants showed that in 99% of them the *kan* gene had been replaced with a *Bam*HI–*Sal*I fragment corresponding to the 766 bp fragment of the TRD.

A second library of mutants (those with numbers between 101 and 200) was obtained by random mutagenesis of the 233 bp *Bsp*EI–*Hin*dIII fragment. The mutagenized DNA was digested with *Bsp*EI and *Hin*dIII, and cloned in pBCBH1. Transformants of DH5α were screened for the loss of the stuffer fragments; most had replaced the stuffer fragment with mutagenized *hsdS* DNA. Both methods of mutagenesis were used to generate mutations; the dNTP method gave a lower yield of DNA and therefore the majority of mutants came from the PCR with $MnCl_2$.

A third library of mutants (with numbers between 201 and 275) was obtained from random mutagenesis of the *Xho*I–*Bsp*EI fragment of pXB1 using $MnCl_2$. The mutagenized DNA was cloned in pXSC1 and treated in the same way as pBSH1 and pBCBH1.

Transfer of mutations to the chromosome

Mutations were transferred to the chromosome from a λ *hsd* phage by homologous recombination (Gough and Murray, 1983). The *hsdS* gene tagged with *kan*^{r} was transferred to the chromosome of the *hsd*$^+$ strain 17–81 to generate the r^-m^- derivative NM521 *kan*. *kan* was then replaced by *hsd* genes including point mutations in *hsdS*; any Kan^S strains should have acquired the *hsdS* region that had been subjected to mutagenesis.

Test for restriction and modification

The $hsdR^+M^-S^-$ strain NM522 transformed with $hsdM^+S^-$ plasmids, and $hsdR^+M^+S^-$ strains in which the mutation was on the chromosome, were tested for their ability to restrict and modify unmodified λ phages (λ *vir*.0), as described by Fuller-Pace *et al*. (1985).

Protein and substrate preparation

Methyltransferases were purified from 500 ml cultures of NM679 carrying derivatives of pJES23. The protein concentrations were determined by UV spectroscopy as described previously (Dryden *et al*., 1993, 1997).

Synthetic, 21 bp unmethylated DNA oligonucleotide duplexes containing a hexachlorofluorescein label on the top strand were prepared as described previously (Powell *et al*., 1998a) and used to compare DNA binding by the methyltransferases via observation of the change in fluorescence anisotropy of the label which occurs upon protein binding. The DNA sequence of the top strand was 5′-hexachlorofluorescein-GCCT**AAC**CACGTG**GTGC**GTAC-3′. The 21 bp fluorescent duplex containing the *Eco*KI target has the same sequence as the central portion of the 45 bp duplex used in gel retardation (Powell *et al*., 1993).

Measurement of molecular weight and AdoMet binding

Protein was applied to a FPLC Superose12 gel filtration column (Pharmacia) which had been calibrated with proteins of known molecular weight. The buffer used was 20 mM Tris–HCl, 0.1 M NaCl, 6 mM $MgCl_2$, 7 mM β-mercaptoethanol pH 8; the column flow rate was 1 ml/min and the elution profile was monitored at 280 nm. To determine the extent of AdoMet binding, protein and [^{3}H-methyl]AdoMet (Amersham) were mixed together at a concentration of 3 μM in 20 mM Tris, 20 mM MES, 0.2 M NaCl, 10 mM $MgCl_2$, 7 mM β-mercaptoethanol, 0.1 mM EDTA pH 8. Samples were then exposed to UV radiation for 20 min to crosslink the protein to the AdoMet. The samples were run

on a 10% SDS–polyacrylamide gel and transferred by electroblotting to a PVDF membrane. The dried membrane was then coated with enhanced autoradiography scintillating wax (EABiotech) as per the manufacturer's instructions and exposed to preflashed X-ray film at –70°C. After development of the film, the PVDF blot was stripped of the EA-wax by washing it with toluene and the proteins stained with Coomassie Blue. This allowed a direct matching of the cross-linked proteins with the X-ray film image.

Measurement of DNA binding with fluorescent oligonucleotides

The buffer used for fluorescence anisotropy experiments was 20 mM Tris–HCl, 0.1 M NaCl, 6 mM $MgCl_2$, 7 mM β-mercaptoethanol pH 8. AdoMet (New England Biolabs) was present at 100 μM to ensure saturation of binding sites (Powell *et al.*, 1993). Anisotropy measurements were performed at 25°C on 400 μl samples in an Edinburgh Instruments FS900T fluorimeter for DNA concentrations of 5 nM. The excitation wavelength was 530 nm, and the emission wavelength 570 nm with bandwidths of 3.6 and 10 nm, respectively. The excitation pathlength was 10 mm, the emission pathlength 2 mm. Small amounts of protein were added to the DNA solution in the cuvette using a microlitre syringe, and gently stirred. The cuvette was not removed from the instrument for these additions. The fluorescence intensity was measured with crossed and parallel polarizer orientations and the anisotropy calculated. The broad emission spectrum of the probe allowed the instrumental G factor to be set to 1 ± 0.01 by small adjustments of the emission slitwidths. Each protein titration was repeated at least in duplicate. Each titration took ~45 min to complete. Data were fitted to a single-site binding equation which accounted for significant concentrations of complex when the DNA concentration was similar to the K_d (Heyduk and Lee, 1990). The binding of protein to hexachlorofluorescein-labelled 21 bp duplexes caused changes of $<10\%$ in the fluorescence emission intensity of the fluorescent probe, thereby nearly satisfying the assumption in the binding equation (Heyduk and Lee, 1990) that the quantum yield was invariant (Hill and Royer, 1997). The change in anisotropy reflected the increase in size and change of shape of the DNA duplex upon protein binding (Jameson and Sawyer, 1995; Hill and Royer, 1997).

Acknowledgements

We wish to thank Nicola Preston for running the automated DNA sequencing service, Laurie Cooper for protein purification and Natalie Honeyman for typing. In addition we are particularly indebted to Dr Shane Sturrock for constructing the computer model shown in Figure 5 and advice on its interpretation. D.T.F.D. thanks the Royal Society for a University Research Fellowship. This work was supported from grants from the Medical Research Council and the Royal Society.

References

Abadjieva,A., Webb,M., Patel,J., Zinkevich,V. and Firman,K. (1993) Deletions within the DNA recognition subunit of M.*Eco*R124I that identify a region involved in protein–protein interactions between HsdS and HsdM. *J. Mol. Biol.*, **241**, 35–43.

Appleyard,R.K. (1954) Segregation of lambda lysogenicity during bacterial recombination in *Escherichia coli* K-12. *Genetics*, **39**, 440–452.

Bickle,T.A., Brack,C. and Yuan,R. (1978) ATP-induced conformational changes in the restriction endonuclease from *Escherichia coli* K12. *Proc. Natl Acad. Sci. USA*, **75**, 3099–3103.

Chen,A., Powell,L.M., Dryden,D.T.F., Murray,N.E. and Brown,T. (1995) Tyrosine 27 of the specificity polypeptide of *Eco*KI can be UV crosslinked to a bromodeoxyuridine-substituted DNA target sequence. *Nucleic Acids Res.*, **23**, 1177–1183.

Cheng,X. and Blumenthal,R.M. (1996) Finding a basis for flipping bases. *Structure*, **4**, 639–645.

Choo,Y. and Klug,A. (1997) Physical basis of a protein–DNA recognition code. *Curr. Opin. Struct. Biol.*, **7**, 117–125.

Cowan,G.M., Gann,A.A.F. and 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. and 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.

Dryden,D.T.F., Cooper,L.P., Thorpe,P.H. and Byron,O. (1997) The *in vitro* assembly of the *Eco*KI Type I DNA restriction/modification enzyme and its *in vivo* implications. *Biochemistry*, **36**, 1065–1076.

Eckert,K.A. and Kunkel,T.A. (1991) The fidelity of DNA polymerases used in the polymerase chain reaction. In Quirke,T. and Taylor,G. (eds) *Polymerase Chain Reaction I: a Practical Approach.* IRL Press, Oxford, UK, pp. 225–244.

Fried,M.G. (1989) Measurement of protein–DNA interaction parameters by electrophoresis mobility shift assay. *Electrophoresis*, **10**, 366–376.

Fuller-Pace,F.V., Bullas,L.R., Delius,H. and Murray,N.E. (1984) Genetic recombination can generate altered restriction specificity. *Proc. Natl Acad. Sci. USA*, **81**, 6095–6099.

Fuller-Pace,F.V., Cowan,G.M. and Murray,N.E. (1985) *Eco*A and *Eco*E: Alternatives to the *Eco*K family of Type I restriction and modification systems of *Escherichia coli*. *J. Mol. Biol.*, **186**, 65–75.

Fürste,J.P., Pansegrau,W., Frank,R., Blöcker,H., Scholz,P., Bagdasarian, M. and Lanka,E. (1986) Molecular cloning of the plasmid RP4 primase region in a multi-host-range *tacP* expression vector. *Gene*, **48**, 119–131.

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

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

Grant,S.G.N., Jessee,J., Bloom,F.R. and Hanahan,D. (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc. Natl Acad. Sci. USA*, **87**, 4645–4649.

Heyduk,T. and Lee,J.C. (1990) Application of fluorescence energy transfer and polarisation to monitor *Escherichia coli* cAMP receptor protein and *lac* promoter interaction. *Proc. Natl Acad. Sci. USA*, **87**, 1744–1748.

Hill,J.J. and Royer,C.A. (1997) Fluorescence approaches to study of protein–nucleic acid complexation. *Methods Enzymol.*, **278**, 390–416.

Ho,S.N., Hunt,H.D., Horton,R.M., Pullen,J.K. and Pease,L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51–59.

Jameson,D.M. and Sawyer,W.H. (1995) Fluorescence anisotropy applied to biomolecular interactions. *Methods Enzymol.*, **246**, 283–300.

King,G. and Murray,N.E. (1995) Restriction alleviation and modification enhancement by the Rac prophage of *Escherichia coli* K-12. *Mol. Microbiol.*, **16**, 769–777.

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

Lange,C., Wild,S. and Trautner,T.A. (1996) Identification of a subdomain within DNA- (cytosine-C5)-methyltransferases responsible for the recognition of the 5′ part of their DNA target. *EMBO J.*, **15**, 1443–1450.

Leung,D.W., Chen,E. and Goeddel,D.V. (1989) A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction technique. *J. Methods Cell Mol. Biol.*, **1**, 11–15.

Meister,J., MacWilliams,M., Hubner,P., Jutte,H., Skrzypek,E., Piekarowicz,A. and 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.

Messing,J., Gronenborn,B., Müller-Hill,B. and Hofschneider,P.H. (1977) Filamentous coliphage M13 as a cloning vehicle: Insertion of a *Hin*dIII fragment of the *lac* regulatory region in M13 replicative form *in vitro*. *Proc. Natl Acad. Sci. USA*, **9**, 3642–3646.

Nagaraja,V., Shepherd,J.C.W. and Bickle,T.A. (1985) A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA-sequence specificity. *Nature*, **316**, 371–372.

Powell,L.M., Dryden,D.T.F., Willcock,D.F., Pain,R.H. and 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.

Powell,L.M., Connolly,B.A. and Dryden,D.T.F. (1998a) The DNA binding characteristics of the trimeric *Eco*KI methyltransferase and its partially assembled dimeric form determined by fluorescence polarisation and DNA footprinting. *J. Mol. Biol.*, **283**, 947–961.

Powell,L.M., Dryden,D.T.F. and Murray,N.E. (1998b) Sequence-specific DNA binding by *Eco*KI, a type IA DNA restriction enzyme. *J. Mol. Biol.*, **283**, 963–976.

Redaschi,N. and Bickle,T.A. (1996) DNA restriction and modification systems. In Neidhardt,F.C. (ed.), *Escherichia coli and Salmonella.* 2nd edn. ASM Press Washington, DC, pp. 773–781.

Reidhaar-Olson,J.F. and Sauer,R.T. (1988) Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science*, **241**, 53–57.

Rhodes,D., Schwabe,J.W.R., Chapman,L. and Fairall,L. (1996) Towards an understanding of protein–DNA recognition. *Phil. Trans. R. Soc. Lond. B.*, **351**, 501–509.

Spee,J.H., de Vos,W.M. and Kuipers,O. (1993) Efficient random mutagenesis method with adjustable mutation frequency by use of PCR and dITP. *Nucleic Acids Res.*, **21**, 777–778.

Sturrock,S.S. and Dryden,D.T.F. (1997) A prediction of the amino acids and structures involved in DNA recognition by type I DNA restriction and modification enzymes. *Nucleic Acids Res.*, **25**, 3408–3414.

Taylor,I.A., Webb,M. and Kneale,G.G. (1996) Surface labelling of the type I methyltransferase M.*Eco*R124I reveals lysine residues critical for DNA binding. *J. Mol. Biol.*, **258**, 62–73.

Thorpe,P.H., Ternent,D. and Murray,N.E. (1997) The specificity of *Sty*SKI, a type I restriction enzyme, implies a structure with rotational symmetry. *Nucleic Acids Res.*, **25**, 1694–1700.

Webb,J.L., King,G., Ternent,D., Titheradge,A.J.B. and Murray,N.E. (1996) Restriction by *Eco*KI is enhanced by co-operative interactions between target sequences and is dependent on DEAD box motifs. *EMBO J.*, **15**, 2003–2009.

Weiserova,M. and Firman,K. (1998) Isolation of a non-classical mutant of the DNA recognition subunit of the Type I restriction endonuclease R.*Eco*R124I. *Biol. Chem.*, **379**, 585–589.

Wharton,R.P. and Ptashne,M. (1985) Changing the binding-specificity of a repressor by redesigning an alpha-helix. *Nature*, **316**, 601–605.

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

Zhou,Y., Zhang,X. and Ebright,R.H. (1991) Random mutagenesis of gene-sized DNA molecules by use of PCR with *Taq* DNA polymerase. *Nucleic Acids Res.*, **19**, 6052.

Zinkevich,V., Heslop,P., Glover,S.W., Weiserova,M., Hubácek,J. and Firman,K. (1992) Mutation in the specificity polypeptide of the Type I restriction endonuclease R · *Eco*K that affects subunit assembly. *J. Mol. Biol.*, **227**, 597–601.

Received September 11, 1998; revised October 12, 1998; accepted October 14, 1998