

Altering the sequence specificity of *HaeIII* methyltransferase by directed evolution using *in vitro* compartmentalization

Helen M.Cohen¹, Dan S.Tawfik² and Andrew D.Griffiths^{3,4}

¹MRC Centre for Protein Engineering and ³MRC Laboratory for Molecular Biology, Cambridge CB2 2QH, UK and ²Weizmann Institute of Science, Rehovot 76 100, Israel

⁴To whom correspondence should be addressed.
E-mail: griff@mrc-lmb.cam.ac.uk

Engineering the specificity of DNA-modifying enzymes has proven extremely challenging, as sequence recognition by these enzymes is poorly understood. Here we used directed evolution to generate a variant of *HaeIII* methyltransferase that efficiently methylates a novel target site. *M.HaeIII* methylates the internal cytosine of the canonical sequence GGCC, but there is promiscuous methylation of a variety of non-canonical sites, notably AGCC, at a reduced rate. Using *in vitro* compartmentalization (IVC), libraries of *M.HaeIII* genes were selected for the ability to efficiently methylate AGCC. A two-step mutagenesis strategy, involving initial randomization of DNA-contacting residues followed by randomization of the loop that lies behind these residues, yielded a mutant with a 670-fold improvement in catalytic efficiency (k_{cat}/K_m^{DNA}) using AGCC and a preference for AGCC over GGCC. The mutant methylates three sites efficiently (AGCC, CGCC and GGCC). Indeed, it methylates CGCC slightly more efficiently than AGCC. However, the mutant discriminates against other non-canonical sites, including TGCC, as effectively as the wild-type enzyme. This study provides a rare example of a laboratory-evolved enzyme whose catalytic efficiency surpasses that of the wild-type enzyme with the principal substrate.

Keywords: directed evolution/*HaeIII* methyltransferase/*in vitro* compartmentalization/sequence specificity/substrate promiscuity

Introduction

Restriction–modification (R–M) systems are widespread in bacteria, with over 240 DNA specificities known (Roberts *et al.*, 2003). Each R–M system consists of a DNA methyltransferase which methylates a specific sequence, usually 4–8 bp long, and a restriction enzyme that digests unmethylated (foreign) DNA, providing the bacteria with a defence against phage infection.

The structure of the cytosine-C⁵ methyltransferase *M.HaeIII* was solved in complex with DNA (Reinisch *et al.*, 1995). The protein comprises two domains separated by a large cleft that holds the DNA. The target base is flipped out of the helix into the catalytic site, as first observed in the structure of the *M.HhaI*:DNA complex (Klimasauskas *et al.*, 1994). One domain contains the catalytic site and the cofactor (*S*-adenosyl-L-methionine; AdoMet) binding site. The structure

of this domain is conserved in all DNA methyltransferases and in the majority of AdoMet-dependent methyltransferases (Schluckebier *et al.*, 1995; Cheng and Blumenthal, 1999; Martin and McMillan, 2002). The second domain is responsible for DNA recognition and varies widely in sequence, size and structure from one enzyme to another.

Despite the wealth of natural R–M systems, there is a continuing effort to engineer systems with novel specificities, as both methyltransferases and restriction endonucleases are extremely valuable tools in molecular biology. Alteration of the sequence specificity of DNA methyltransferases by protein engineering has, however, proved very challenging and has been limited to the exchange of entire domains between methyltransferases, thus conferring the specificity of the new TRD on the recipient enzyme (Gann *et al.*, 1987; Klimasauskas *et al.*, 1991; Gubler *et al.*, 1992; Mi and Roberts, 1992; Walter *et al.*, 1992; Trautner *et al.*, 1996). An attempt to alter the specificity of a bacteriophage methyltransferase by the fusion of two TRDs with different DNA specificities led to an enzyme with a degenerate ‘relaxed’ specificity that was not predicted from the specificities of the two parent TRDs (Lange *et al.*, 1996).

An experiment to alter the specificity for the methylated base from adenine to cytosine demonstrated the separate roles of the catalytic and target recognition domains (Roth and Jeltsch, 2001). The *M.EcoRV* variant, which had mutations in the conserved catalytic domain, displayed a 22-fold preference for cytosine as the target base, but only when located in a CT mismatch and with no increase in catalytic efficiency compared with methylation of the same substrate by the wild-type enzyme. In fact, the designed mutations abolish methylation of adenine without improving methylation of cytosine.

An attractive alternative to rational design is selection (Griffiths and Tawfik, 2000; Brannigan and Wilkinson, 2002). An *in vivo* selection system, based on the principle of methylated DNA being rendered resistant to cleavage by a restriction endonuclease, has been used to clone methyltransferases from genomic DNA and to select active variants from libraries (Szomolanyi *et al.*, 1980; Kiss *et al.*, 2001; Vilkaitis *et al.*, 2002). A library of mutants of *M.SinI*, target site GGWCC (W = A or T), was selected for the ability to methylate the degenerate site, GGNCC (N = A, C, G or T) (Kiss *et al.*, 2001). The best variant isolated by this method showed a more relaxed specificity at the central base, although the original target site was still methylated twice as efficiently as the new site and the methylation activity on both sites was 5- to 10-fold lower than the activity of the wild-type enzyme on the canonical site.

Redesigning restriction endonucleases to create new sequence specificities has also proved difficult. Based on the crystal structure of *EcoRV* in complex with various DNA substrates, attempts have been made to lengthen the 6 bp recognition sequence to 8 bp (Schottler *et al.*, 1998; Lanio *et al.*,

2000). As with DNA methyltransferases, the most promising results have arisen by selection from libraries of variants. A semi-rational approach was used to select a mutant of *EcoRV* with 25-fold higher activity on sites flanked by AT rather than GC, although several other sites were cut only slightly less efficiently (Lanio *et al.*, 1998). More recently, an experiment in directed evolution produced a mutant of *BstYI* (specificity RGATCY, where R = A or G and Y = C or T) with a 12-fold preference for AGATCT over AGATCC or GGATCT and no detectable cleavage of GGATCC (Samuelson and Xu, 2002). Also, selection of random libraries of the bifunctional restriction endonuclease *Eco57I*, in which the restriction endonuclease activity had been inactivated by point mutation, for the ability to methylate and protect the sequence CTGGAG from cleavage by *GsuI* (the wild-type *Eco57I* target sequence is CTGAAG) led to the isolation of a mutant which could methylate both of the above sequences. Restoration of the cleavage activity yielded a mutant restriction endonuclease with the specificity CTGRAG (Rimseliene *et al.*, 2003).

Despite the difficulties that have been encountered when trying to generate R–M systems of novel specificity in the laboratory, the abundance of different specificities that occur in bacteria and archaea indicates that this process must have occurred many times in nature. Non-canonical activity against star sites (sites that differ by one base pair from the target sequence) or other DNA sites has been reported both for restriction endonucleases and methyltransferases. It has been proposed that many enzymes have promiscuous activity against molecules that are not their principal substrate and that new enzymes may evolve by improvements in the enzyme's ability to catalyse the conversion of one of these poor substrates (Jensen, 1976; O'Brien and Herschlag, 1999; Copley, 2003; James and Tawfik, 2003). The enhancement of an existing promiscuous activity has been proposed as a possible route for the evolution of R–M systems with novel specificity (Beck *et al.*, 2001; Cohen *et al.*, 2002). This is also a strategy that has been successfully used for directed evolution of enzymes with altered substrate specificity, for example the evolution of a β -glucuronidase with β -galactosidase activity (Matsumura and Ellington, 2001) and the alteration of the site specificity of Cre recombinase (Buchholz and Stewart, 2001).

M.HaeIII methylates a range of other sites in addition to the target site GGCC, albeit with reduced efficiency, the most frequently methylated star site being AGCC (Cohen *et al.*, 2002). In this study we used *in vitro* compartmentalization (IVC) to evolve a variant of *M.HaeIII* with novel sequence specificity by enhancing this activity towards AGCC.

IVC is a completely *in vitro* system which allows activity-based selection of DNA methyltransferases (Tawfik and Griffiths, 1998; Lee *et al.*, 2002), in addition to other enzymes (Ghadessy *et al.*, 2001; Griffiths and Tawfik, 2003) and ligands (Doi and Yanagawa, 1999; Sepp *et al.*, 2002). In nature, genotype–phenotype linkage results from compartmentalization of the genome by the cell membrane. In IVC, this linkage is achieved by compartmentalizing single genes, not in cells, but in the aqueous compartments of a water-in-oil emulsion (Figure 1).

Using IVC to select *M.HaeIII* libraries for the ability to methylate AGCC, we isolated a variant with a 670-fold improvement in catalytic efficiency (k_{cat}/K_m^{DNA}) on this site.

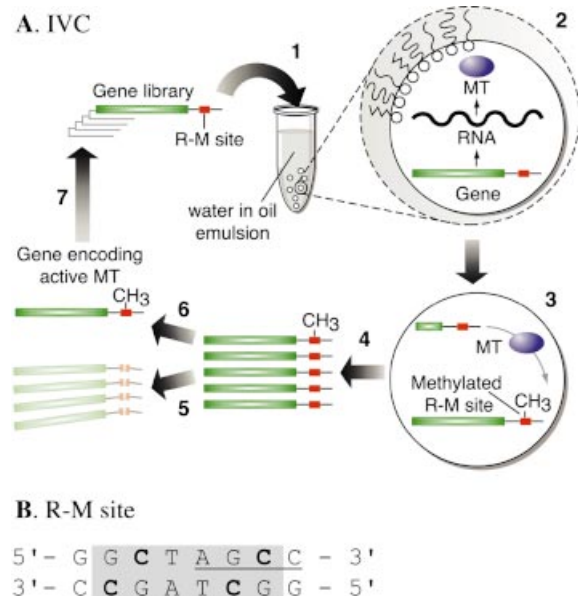


Fig. 1. Selection of DNA methyltransferases by *in vitro* compartmentalization (IVC). (A) Schematic representation of the selection procedure. An *in vitro* transcription/translation reaction mixture containing a library of genes encoding mutant methyltransferases (MT), each with restriction–methylation (R–M) sites appended to the gene is dispersed to form a water-in-oil emulsion with typically one gene per aqueous compartment (1). The genes are transcribed and translated within their compartments (2). Proteins with methyltransferase activity methylate the R–M sites (3). Compartmentalization prevents the methylation of genes in other compartments. The emulsion is broken, all reactions are stopped and the aqueous compartments combined (4). Digestion with the cognate restriction enzyme results in the digestion of unmethylated genes (which do not encode active methyltransferases) (5) and the survival of methylated genes (which encode active methyltransferases) (6). The surviving genes can be amplified using the polymerase chain reaction and compartmentalized for further rounds of selection (7). (B) The sequence of the (R–M) site. Library A has a single *NheI* site (encoded in pIVEX.1s) and library B three *NheI* sites (encoded in pIVEX.3s). The 6 bp recognition site of *NheI* is shaded grey. Digestion by *NheI* is blocked by C⁵ methylation at any of the cytosines in bold type. The two AGCC sites, which overlap each end of the *NheI* site, are underlined.

Materials and methods

Oligonucleotides

See supplementary material (available at *PEDS* online).

Plasmids

Complementary oligonucleotides Mon1 and Mon2 were annealed as described (Cohen *et al.*, 2002), phosphorylated by T4 polynucleotide kinase (New England Biolabs) and ligated into pIVEX2.2bNde (Roche), which had been digested with *Bam*HI and treated with calf-intestinal phosphatase (Roche), to create the vector pIVEX.1s. pIVEX.3s was created in a similar manner by ligation of the annealed oligonucleotides 5SN1 and 5SN2 into pIVEX2.2bNde cut with *Bam*HI and *Sac*I.

The N-terminal FLAG-tagged *M.HaeIII* gene was subcloned from plasmid pET-MHaeIII (Cohen *et al.*, 2002) into pIVEX.1s using *Nco*I and *Sac*I, to create pIVEX.1s.MHaeIII.

The single mutant R225A was synthesized via the strategy of PCR and ligation described previously (Griffiths and Tawfik, 2003). The template for the first PCR was pIVEX.1s.MHaeIII.

Oligonucleotides T1.1R and LMB2.1 were used to amplify the N-terminal fragment and T1.5 and pIV-B1bio for the C-terminal fragment. Digestion with *BsmBI*, ligation and capture via biotin were performed as described. The full-length gene was amplified by PCR using primers LMB2.2 and pIV-B2bio, digested with *NcoI* and *SacI*, and ligated into pIVEX.3s to create pIVEX.3s.R225A.

Substrates for methylation

*HaeIII*sub, a 350 bp PCR product containing a single *HaeIII* site, was prepared by PCR from the template pGEM-4z (Promega) using primers hmc3 (5'-biotin) and hmc4 (5'-digoxigenin). *NheI*sub, a 593 bp PCR product containing a single *NheI* site that is overlapped by the *HaeIII* star site AGCC, was prepared by PCR from pIVEX.1s using primers pIV-B9DIG (5'-digoxigenin) and LMB2.9bio (5'-biotin).

Synthetic 30 bp DNA fragments were made by annealing complementary oligonucleotides. A hemi-methylated GGCC site was formed by annealing Gsub (top strand) and Gsub-meth (bottom strand). Hemi-methylated AGCC was made from Asub (top strand) and Asub-meth (bottom strand), CGCC from Csub (top strand) and Csub-meth (bottom strand), TGCC from Tsub (top strand) and Tsub-meth (bottom strand), and GGCT from Asub-meth (top strand) and Asub (bottom strand).

Library synthesis

Libraries were synthesized using oligonucleotides to encode the diversified residues, as described (Griffiths and Tawfik, 2003). Library A was prepared by PCR amplification from pIVEX.1s.MHaeIII. Primers LMB2.1 and TRD1Fo were used to amplify the N-terminal fragment and pIV-B1bio and TRD1Ba for the C-terminal fragment. Library B was prepared by PCR amplification from pIVEX.3s.R225A using LMB2.1 and TRD1.3R for the N-terminal fragment and the pIV-B1bio and TRD1.3 for the C-terminal fragment. Each library was amplified by PCR with primers pIV-B2bio and LMB2.2.

Library selection by IVC

Selection by IVC was performed essentially as described (Tawfik and Griffiths, 1998) with the following modifications. The EcoPro T7 *in vitro* transcription/translation system (Novagen) was used and the oil mix was as follows: light mineral oil (Sigma) containing 4.5% (w/w) Span 80 (Fluka) and 0.5% w/w Triton X-100 (Fischer). DNA was added to the reaction mix to a final concentration of 0.1 nM. Glycerol increases the activity of *M.HaeIII* against non-canonical sites (Cohen *et al.*, 2002) and was added to the reaction at a final concentration of 16% by volume. Emulsions were incubated at 25°C for 4 h. The restriction endonuclease *NheI* (New England Biolabs) was used to digest unmethylated genes.

After each round of selection the emulsion was broken, the genes captured on streptavidin-coated beads and treated with *NheI*. Undigested genes were amplified by PCR using primers LMB2.11 and pIV-B11. To avoid PCR artefacts and to re-append the *NheI* restriction sites before the next round, the DNA was excised using *NcoI* and *SacI*, ligated into pIVEX.1s (library A) or pIVEX.3s (library B) and re-amplified with primers LMB2.1 and pIV-B1bio which anneal in the vector sequence outside the annealing sites of the primers used in the previous PCR (Griffiths and Tawfik, 2003). Selected libraries were cloned into pIVEX.1s or pIVEX.3s for initial characterization by *in vitro* expression and sequencing.

In vitro translation and assay of selected genes

Libraries, selected libraries or cloned genes were assayed for methylation activity by a digoxigenin-biotin ELISA-based method essentially as described (Tawfik and Griffiths, 1998). PCR-amplified DNA was added to EcoPro T7 transcription/translation system to 2 nM and the reaction mixture was incubated at 25°C for 90 min. A 2 µl volume of *in vitro* transcription/translation reaction mixture was added to the methylation reaction containing 10 nM methylation substrate, 10 mM EDTA, 80 µM AdoMet, 50 nM Tris-HCl, pH 7.4, 50 mM NaCl, 1 mM dithiothreitol in a final volume of 20 µl. Methylation reaction mixtures were incubated at 25°C and aliquots removed after 2 min to 16 h, for quenching as described (Tawfik and Griffiths, 1998). Methylated DNA was bound to streptavidin-coated plates, digested by either *HaeIII* or *NheI* in NEB buffer 2 (50 mM NaCl, 10 mM Tris-HCl, pH 7.9, 10 mM MgCl₂, 1 mM dithiothreitol) (New England Biolabs) and undigested DNA was detected as described.

Expression and purification of methyltransferases

The genes for *M.HaeIII* and the mutant T29 were sub-cloned from pIVEX derivatives into pTrc99 (Amersham Biosciences) using *NcoI* and *SacI* restriction sites and transformed into *E.coli* MC1061 [F⁻ *araD139* Δ(*ara-leu*)7696 *galE15 galK16* Δ(*lac*)X74 *rpsL* (Str^r) *hsdR2* (r_k⁻ m_k⁺) *mcrA mcrB1*]. Cultures were grown at 37°C to OD 0.6, induced by the addition of 1 mM isopropyl-(β)-D-thiogalactoside and transferred to 20°C for 4 h. Methyltransferases were purified by ion exchange using S-Sepharose as described (Chen *et al.*, 1991) and stored at -80°C. The concentration of active sites was determined by an electrophoretic mobility shift assay (EMSA) (Vilkaitis *et al.*, 2001) (see supplementary material).

Kinetic assays of methyltransferase activity

Purified wild-type *M.HaeIII* and the T29 mutant were tested for the ability to protect *NheI*sub and *HaeIII*sub from digestion by *NheI* and *HaeIII*, respectively, using a digoxigenin-biotin ELISA-based method as above (Tawfik and Griffiths, 1998). Each 20 µl methylation reaction contained 0.5 nM active sites of enzyme and 10 nM substrate DNA. Methylation reaction mixtures were incubated at 25°C and aliquots removed and quenched after 2–30 min.

Five double-stranded 30 bp DNA substrates were used to determine the sequence specificity of *M.HaeIII* and the T29 mutant. Each contains a single hemi-methylated *HaeIII* site (GGCC) or star site (AGCC, CGCC, TGCC or GGCT). Methylation reactions were performed in *M.HaeIII* reaction buffer supplemented with 5–6000 nM DNA, 10 mM EDTA and 1 µM *S*-adenosyl-L-[methyl-³H]methionine (20 Ci/mmol) (Amersham). For each enzyme and each DNA substrate the DNA concentration range was within the range 0.2–10.0K_m (see supplementary material). Reactions were pre-heated to 37°C and started by the addition of 1–5 nM active sites of purified enzyme. At least four aliquots of 20 µl were taken within the linear range of the reaction (between 1 and 20 min) and quenched by the addition of 45 µl of 160 µM unlabelled *S*-adenosyl-L-methionine (AdoMet) in 300 mM sodium acetate, pH 5.0. Triplicate 20 µl samples were spotted on to Multiscreen DE plates (Millipore). The plates were washed three times with 200 µl of 0.2 M ammonium hydrogen carbonate and once with 100 µl of ethanol and allowed to air dry. The filters were punched out from the plates and counted in 1 ml of scintillant (National Diagnostics) in a Beckmann LS6000SC scintillation

counter. The initial rate was calculated from measurements taken within the first 5% of the reaction, to avoid product inhibition by *S*-adenosylhomocysteine.

Kinetic constants were determined by fitting plots of the initial reaction velocity (v_0) versus DNA concentration $[S]$ to the Michaelis–Menten model, $v_0 = k_{\text{cat}}[E]_0[S]/(K_m + [S])$, using the Levenberg–Marquardt algorithm, as implemented in Kaleidagraph (Synergy Software, Reading, PA) (see supplementary material). $[E]_0$ is the initial concentration of enzyme active sites (determined by EMSA). k_{cat} is the apparent turnover number and K_m^{DNA} is the apparent Michaelis–Menten constant of the enzyme for DNA, as measured at 1 μM AdoMet. This concentration is above the K_m^{AdoMet} of the wild-type enzyme and T29 mutant. Previously, *M.EcoRI* was tested with a similar variety of DNA substrates with little variation in the K_m^{AdoMet} (Reich *et al.*, 1992), leading to the assumption that the K_m^{AdoMet} of *M.HaeIII* (and mutants) is similar for all the DNA substrates used.

K_m^{AdoMet} was measured by methyltransferase assays containing 1–5 nM enzyme, saturating concentrations of DNA (2 μM of the substrate containing a hemi-methylated GGCC site) and 50–3000 nM *S*-adenosyl-L-[methyl- ^3H]methionine in *M.HaeIII* buffer. Reactions were carried out as above and plots of initial reaction velocity versus AdoMet concentration were fitted to the Michaelis–Menten model, where $[S]$ is the initial concentration of AdoMet and K_m is the apparent Michaelis–Menten constant of the enzyme for AdoMet.

Results

Selection strategy for methylation at star sites

The strategy for selecting active methyltransferases by IVC requires a restriction enzyme to digest the genes encoding inactive enzymes, leaving the methylated genes encoding active enzymes intact (Figure 1A). There are no known restriction enzymes with the sequence specificity AGCC, but restriction sites may be protected from digestion if they overlap methylation sites. *NheI* is blocked by C⁵ methylation of any of the cytosines in its recognition site (GCTAGC) (Roberts *et al.*, 2003). By appending *NheI* sites that overlap with the *M.HaeIII* star site AGCC, genes encoding methyltransferases which methylate this star site can be selected (Figure 1B). Plasmids pIVEX.1s and pIVEX.3s, which contain one or three such *NheI* sites, respectively, were used for library construction and selection. This strategy can also potentially lead to selection for C⁵ methylation of GGCT.

Selection of library A for methylation of AGCC

Crystallographic studies show that the TRD of *M.HaeIII* contains two target recognition loops (loops I and II) which contact the target DNA sequence. These two loops are thought to be responsible for almost all the base-specific interactions with the target site (Reinisch *et al.*, 1995). Loop I residues interact with the first three base pairs of the target DNA whereas loop II residues interact with the rest of the target site.

Library A was synthesized by introducing diversity at five codons in loop I: Arg225, which is the only residue thought to make a contact to the first base pair (to G1; there are apparently no amino acid interactions with the complementary base C1'); Ser224, which is thought to interact with both Arg225 and C2'; Ser217, Asn232 and Glu233, which differ between *M.HaeIII* and its closest homologue *M.FnuDII*, which also methylates GGCC (Figure 2). The template for this library was the

Library A

	217						224 225					232 233					
<i>M.HaeIII</i>	S	Y	S	T	I	F	M	S	R	N	R	V	R	Q	W	N	E
Library A	X	Y	S	T	I	F	M	X	X	N	R	V	R	Q	W	X	X

Library B

	260	261	262
<i>M.HaeIII</i>	N	L	N
Library B	X	X	X

Fig. 2. Library design. The regions of *M.HaeIII* that are mutated in library A and library B are shown. Amino acid numbering is the same as in the untagged protein. Amino acids in bold type were diversified as follows. Library A: codon 217 is replaced with WSW, encoding CRST; codon 224 replaced with RVC, encoding ADGNST; codon 225 replaced with NNS, encoding all 20 amino acids; codon 232 replaced with RAW, encoding DENK; and codon 233 replaced with SAA, encoding E and N. In library B codons 260, 261 and 262 are each replaced by NNS, fully randomizing the amino acids at all three positions. N = A, C, G or T, R = A or G, S = C or G, V = A, C or G, W = A or T.

M.HaeIII gene cloned in pIVEX.1s with an N-terminal FLAG tag (Chiang and Roeder, 1993). All library members therefore contained a single *NheI* site that could be protected from digestion by methylation of the central cytosine of AGCC (Figure 1B).

Library A was selected for methylation activity on these star sites by IVC as described (Tawfik and Griffiths, 1998; Lee *et al.*, 2002), using *NheI* restriction endonuclease to digest unmethylated DNA (Figure 1). The enrichment of genes encoding active methyltransferases was monitored by *in vitro* translation of the DNA amplified after each round of selection and assaying the ability of the *in vitro* translated protein to protect the DNA fragment *NheI*sub from digestion by *NheI*. *NheI*sub is a 593 bp PCR product containing a single *NheI* site overlapped by AGCC sites. Unmethylated *NheI*sub was cut to completion but substrate DNA methylated at these star sites remains uncut. In a 16 h reaction, *in vitro* translated (unselected) library A protected 2% of the DNA from digestion. After the fourth round of IVC, 26% of the DNA was protected (Figure 3).

The selected genes were sub-cloned and individual clones screened by *in vitro* translation and assayed using *NheI*sub. Out of 20 clones with detectable activity, nine methylated *NheI*sub significantly faster than wild-type *M.HaeIII* and were sequenced. These genes all contained the mutation R225A but no other mutations were common to all clones. Ser224 was not mutated in any of the sequenced genes. The single mutant R225A was constructed and assayed as above and displayed the same level of activity as the fastest selected mutant, a 20-fold increase in the rate of methylation of *NheI*sub compared with the wild-type enzyme (Table I).

Selection of library B for methylation of AGCC

A second library, library B, was synthesized using the single mutant R225A in pIVEX.3s. All genes therefore contained three *NheI* sites, all of which must be methylated for any gene to be selected. Three residues were randomized, Asn260, Leu261 and Asn262, which are not thought to make direct DNA contacts but are adjacent to Arg225 in the crystal structure of wild-type *M.HaeIII* (Reinisch *et al.*, 1995) (Figures 2 and 6). Residues 260–262 lie behind loop I and

Table I. Sequence and methylation activity of selected clones

Clone	Residue ^a				Other mutations ^a	AGCC activity ^b	GGCC activity ^b	Activity ratio AGCC:GGCC
	225	260	261	262				
Wild-type	R	N	L	N		<0.05	50	<0.001
^c R225A	A	N	L	N		0.96	1.2	1
T29	A	L	M	W	D-6G	33	5.1	7
T7	A	L	F	W	K200N V265A	3.9	0.12	30
T28	A	L	S	W	F59L M223T	3.0	0.41	7
T26	A	L	T	W	Q230H H270R	2.9	0.32	9
T45	A	L	W	W	K182M	2.3	0.32	7
T1	A	L	R	W	N41S	2.0	0.10	20
T40	A	L	C	W	K-4I I183N N191I N232D	1.8	0.071	30
T9	A	C	L	D	I25M N202D	0.94	0.13	7
T31	A	L	H	F	K-4R N298S	0.55	0.086	6
T46	A	L	Q	H	I34M D84G Q124R I133T C334R	0.46	0.20	2
T44	A	L	E	W	A313T	0.40	0.16	3
T50	A	L	Q	W	K317M	0.29	0.098	3
T42	A	L	A	W	Q250H	0.20	0.015	10

^aAmino acid numbering is the same as in the untagged protein, with the FLAG residues numbered -7 to 0.

^bPCR-amplified genes were translated *in vitro* and tested for the ability to protect *NheI*sub and *HaeIII*sub from digestion by *NheI* and *HaeIII*, respectively.

*NheI*sub contains a single *NheI* site, overlapped by the site AGCC. *HaeIII*sub contains a single GGCC site. Values indicate the rate of DNA protection in fmol/min/ μ l IVT.

^cThe R225A mutation was identified from the selection of library A for AGCC methylation and then served as the starting point for library B from which all the subsequent clones were derived (see text).

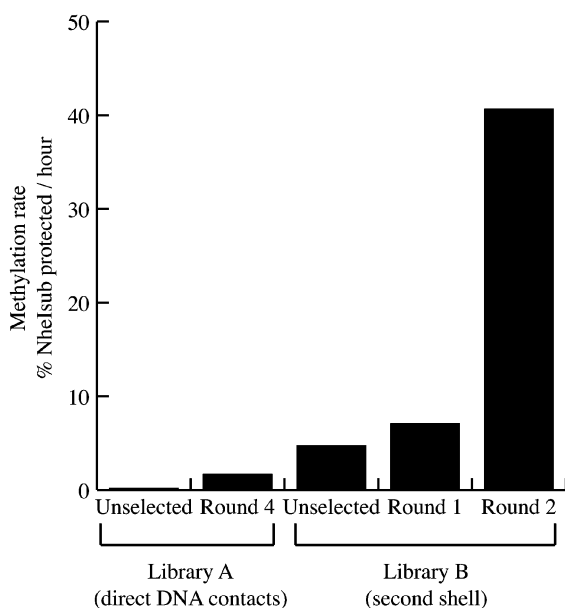


Fig. 3. AGCC methylation activity of the unselected and selected libraries. AGCC methylation was determined by measuring the rate of protection of the 593 bp DNA *NheI*sub against digestion by the restriction endonuclease *NheI*.

the side chain of Asn260 is thought to form a hydrogen bond with Arg225.

This library was put through three rounds of selection by IVC and the enrichment of the genes encoding active methyltransferases monitored as above. The length of methylation assays was reduced from 16 to 1 h, owing to the higher activity of the pool of genes.

Some 4.7% of the DNA remained uncut after methylation by *in vitro* translated, unselected library B, suggesting the presence of a high frequency of active methyltransferase

genes in this library even before selection. After one or two rounds of selection, 7.0 and 41% of the substrate DNA, respectively, remained uncut (Figure 3). A third round of selection, however, did not increase the methyltransferase activity of the library, probably because, after round 2, the library already predominantly comprised clones with sufficient activity to methylate all three AGCC sites during the 4 h incubation.

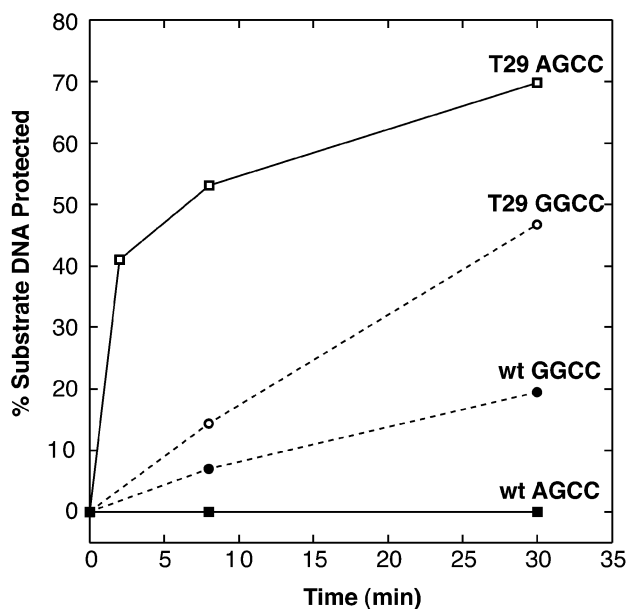
Selected genes were sub-cloned after the second and third rounds of IVC. Individual genes were translated *in vitro* and assayed for the ability to protect *NheI*sub and *HaeIII*sub from digestion by *NheI* or *HaeIII*, respectively (*HaeIII*sub is a 350 bp PCR product containing a single *HaeIII* site). The number of variants showing detectable activity on *NheI*sub after the second round of selection was 10 out of 22 tested and after the third round 37 out of 56 tested. Thirteen of the 78 variants were more efficient at methylating AGCC (and hence protecting *NheI*sub) than methylating GGCC (protecting *HaeIII*sub). These genes were sequenced and the methylation activities on *NheI*sub and *HaeIII*sub are listed in Table I. Seven of these variants methylated *NheI*sub faster than the single mutant R225A.

Activity and sequence specificity of the improved variants

The fastest mutant, T29, which has five amino acid changes (D-6G, R225A, N260L, L261M, N262W), was found to methylate *NheI*sub over 600 times faster than *M.HaeIII* (Table I). This mutant protein and the wild-type *M.HaeIII*, each with an N-terminal FLAG tag, were purified and tested for the ability to protect *NheI*sub (593 bp) and *HaeIII*sub (350 bp) from digestion by *NheI* (detecting methylation of AGCC) and *HaeIII* (detecting methylation of GGCC), respectively (Figure 4). Using 10 nM substrate DNA, the initial rate of methylation of *HaeIII*sub (GGCC) by the wild-type enzyme is 0.17 M/min/M enzyme and there was no detectable methylation of *NheI*sub (AGCC) after 30 min. In contrast, the T29 mutant methylates *NheI*sub (AGCC) very efficiently, with an

Table II. Apparent k_{cat} and K_m^{DNA} values for methylation of 30 bp, hemi-methylated DNA substrates by *M.HaeIII* and the T29 mutant

DNA substrate	<i>M.HaeIII</i>			T29		
	k_{cat} (min^{-1})	K_m^{DNA} (nM)	$k_{\text{cat}}/K_m^{\text{DNA}}$ ($\text{s}^{-1} \text{M}^{-1}$)	k_{cat} (min^{-1})	K_m^{DNA} (nM)	$k_{\text{cat}}/K_m^{\text{DNA}}$ ($\text{s}^{-1} \text{M}^{-1}$)
AGCC	0.17 ± 0.02	2400 ± 700	1180	0.79 ± 0.04	17 ± 3	7.8×10^5
GGCC	0.20 ± 0.01	130 ± 20	2.6×10^4	0.73 ± 0.06	51 ± 10	2.4×10^5
CGCC	0.17 ± 0.02	1700 ± 450	1700	1.7 ± 0.18	30 ± 10	9.4×10^5
TGCC	0.06 ± 0.003	1040 ± 150	960	0.27 ± 0.02	210 ± 30	2.1×10^4
GGCT	0.09 ± 0.003	580 ± 70	2600	0.63 ± 0.04	1300 ± 250	8.1×10^3

**Fig. 4.** Rate of methylation of long DNA substrates. The percentage of 10 nM substrate DNA protected from digestion by either *HaeIII* restriction endonuclease (GGCC) or *NheI* restriction endonuclease (AGCC) is plotted against time. 0.5 nM wild-type (wt) *M.HaeIII* or T29 mutant was used to methylate either the 350 bp substrate DNA *HaeIII*sub (GGCC) or the 593 bp substrate DNA *NheI* sub (AGCC).

initial rate of 4.1 M/min/M enzyme, 24 times faster than methylation of GGCC by the wild-type enzyme. The T29 mutant also methylates *HaeIII*sub (GGCC), but the initial rate of 0.36 M/min/M enzyme is 11 times slower than methylation of *NheI*sub (AGCC).

The sequence specificity of the wild-type enzyme and the T29 mutant was investigated further using several different 30 bp substrates, each containing a single hemi-methylated target site for either wild-type *M.HaeIII* (GGCC), the selected specificity (AGCC) or other star sites of *M.HaeIII* (CGCC, TGCC or GGCT). The star sites are asymmetric (in contrast to the palindromic wild-type site GGCC) and therefore contain a second potential target site for methylation on the complementary strand. For example, an AGCC substrate also has a potential GGCT substrate on the complementary strand. The use of hemi-methylated DNA ensures that methylation of just one target sequence can be assayed (Friedrich *et al.*, 2000). The apparent K_m^{AdoMet} was 620 ± 60 nM for *M.HaeIII* and 440 ± 90 nM for T29. The apparent K_m^{DNA} and k_{cat} of the wild-type and mutant T29 are listed in Table II (also see supplementary material).

Methylation of the star sites by wild-type *M.HaeIII* was 10- to 27-fold less efficient (in terms of $k_{\text{cat}}/K_m^{\text{DNA}}$) than methylation of GGCC, with the preference for GGCC being mainly due to the high K_m^{DNA} for other substrates (Table II). Previously, we had observed that AGCC is methylated more frequently than any other non-canonical site, both *in vitro* and *in vivo* (Cohen *et al.*, 2002), but this was not the case in this study. This difference may be due to a number of factors: the use of hemi-methylated substrates, the sequence context of each site or the use of short oligonucleotides in place of an 832 bp PCR fragment or a 6426 bp plasmid. The sequence context of R-M sites is known to influence the efficiency of digestion by restriction endonucleases (Thomas and Davis, 1975). The length of the DNA substrate may also be significant, as DNA methyltransferases are thought to scan DNA and methylate target sites processively, over distances of several hundred bases (Urig *et al.*, 2002) and significant differences in the activity of DNA methyltransferases on short and long substrates have been reported previously (Reich and Mashhoon, 1991; Cheng and Blumenthal, 1999). However, previous studies have shown that the kinetic parameters of bacterial DNA methyltransferases are almost identical when measured on otherwise identical unmethylated and hemi-methylated substrates (Dryden, 1999; Lindstrom *et al.*, 2000).

Wild-type *M.HaeIII* shows a 22-fold preference (in terms of $k_{\text{cat}}/K_m^{\text{DNA}}$) for the canonical site over AGCC (Table III, Figure 5). T29 shows a dramatic change in both catalytic efficiency and sequence specificity. The k_{cat} for methylation of AGCC is increased 4.6-fold compared with the wild-type and the K_m^{DNA} is 140-fold lower than that of *M.HaeIII* (Table II). Hence mutant T29 has a 670-fold improvement in catalytic efficiency ($k_{\text{cat}}/K_m^{\text{DNA}}$) on AGCC compared with wild-type *M.HaeIII* and shows a 3-fold preference for methylation of AGCC over GGCC (in terms of $k_{\text{cat}}/K_m^{\text{DNA}}$) (Table III, Figure 5). T29 also efficiently methylates CGCC, but the star sites TGCC and GGCT are methylated more slowly.

We cannot rule out the possibility that flanking sequence may affect the observed sequence specificity; however, large increases in rates of AGCC methylation (relative to wild-type) were observed with the T29 mutant using both short oligonucleotide and long DNA substrates in which the DNA was in a completely different sequence context.

Discussion

Eighteen amino acids of *M.HaeIII* are thought to interact with the DNA target bases or phosphate backbone, based on the crystal structure (Reinisch *et al.*, 1995). These interactions include multiple hydrogen bonds to the bases and a network of

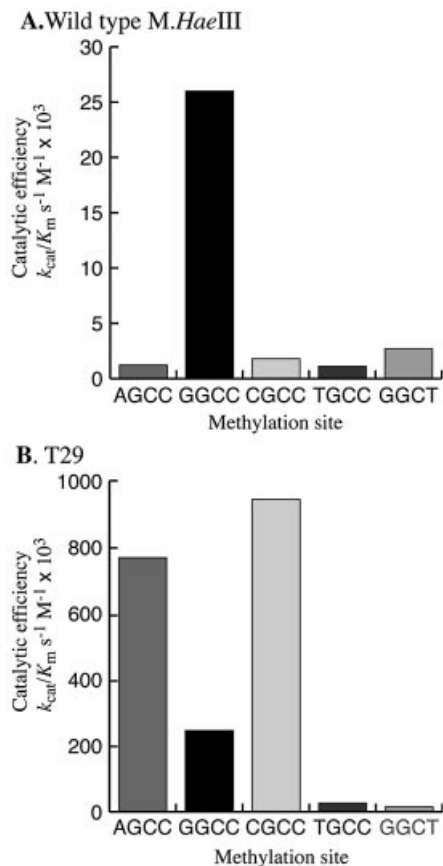


Fig. 5. Catalytic efficiency (k_{cat}/K_m^{DNA}) of wild-type *M.HaeIII* and mutant T29 using hemi-methylated 30 bp DNA substrates.

Table III. Improvements in catalytic efficiency on 30 bp, hemi-methylated DNA substrates

DNA substrate	$k_{cat}/K_m^{\text{DNA}^a}$		Ratio of catalytic efficiency: T29: <i>M.HaeIII</i>
	<i>M.HaeIII</i>	T29	
AGCC	0.045	30.0	670
GGCC	1.000	9.2	9
CGCC	0.065	36.2	560
TGCC	0.037	0.81	22
GGCT	0.100	0.31	3

^a k_{cat}/K_m^{DNA} values of wild-type *M.HaeIII* and T29, normalized with k_{cat}/K_m^{DNA} *M.HaeIII* GGCC = 1.

water-mediated hydrogen bonds, similar to that seen in the high-resolution structure of the *M.HhaI*:DNA complex, is also likely to exist (Klimasauskas *et al.*, 1994). Even with the benefit of the crystal structure of the *M.HaeIII*:DNA complex, any attempt to alter the sequence specificity by rational design would be very challenging as the crystal structure reveals only one of several states in the multistep pathway of DNA binding and recognition (Klimasauskas *et al.*, 1998). Also, the lack of structural and sequence conservation between TRDs of DNA methyltransferases with different specificities and the lack of any known methyltransferase with the desired specificity (AGCC) preclude the use of sequence homology to predict the effects of mutations in the TRD.

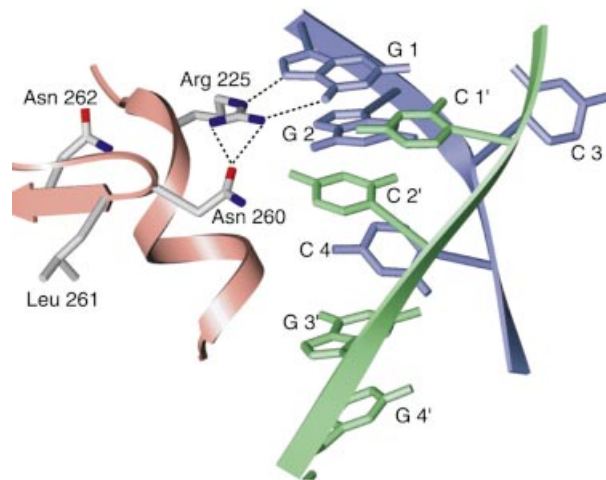


Fig. 6. Model of the interaction of *M.HaeIII* and the first base of the DNA target GGCC. The residues mutated in T29 are shown with atoms coloured: grey, carbon; red, oxygen; blue, nitrogen. DNA strands are coloured purple (methylation target) and green (complementary strand). The figure was prepared with the program SETOR (Evans, 1993) using the published crystal structure (Reinisch *et al.*, 1995).

The normal target for methylation by *M.HaeIII* is the sequence GGCC, but methylation of other DNA substrates also occurs with lower efficiency. We set out to take advantage of this promiscuity to evolve an enzyme which efficiently methylates AGCC as we had previously observed that this site is methylated more frequently than any other non-canonical site, both *in vitro* and *in vivo* (Cohen *et al.*, 2002).

To isolate efficient enzymes which methylate the new target AGCC we used a two-stage strategy of mutation and selection using IVC. The first library (library A) contained diversity at residues in target recognition loop I. The best variants selected all contained the R225A mutation, which was then used as the starting point for a second library (library B) with mutations in the residues which lie behind loop I. Selection of this second library led to the isolation of a number of mutants with further improvements in activity. The fastest variant, T29, has a 670-fold improvement in catalytic efficiency (k_{cat}/K_m^{DNA}) on the 30 bp substrate containing AGCC. This is due to a decrease in K_m^{DNA} of 140-fold and an increase in k_{cat} of 4.6-fold (Table II). In nature and in the laboratory there is selection pressure for the evolution of enzymes with K_m s that match the substrate concentration (Fersht, 1999; Griffiths and Tawfik, 2000). Assuming a mean droplet diameter of 2.6 μm in the emulsion (Tawfik and Griffiths, 1998), the concentration of a single gene inside a droplet is 0.18 nM. This low concentration of DNA may have driven the evolution of enzymes with low K_m^{DNA} values.

Many aspects of enzyme function have been modified by directed evolution, but improvements in catalytic efficiency of the principal reaction and alterations to the catalytic mechanism and the substrate specificity remain a challenge (Brannigan and Wilkinson, 2002). Directed evolution of a phosphotriesterase by IVC led to a 63-fold increase in k_{cat} for the principal reaction, but k_{cat}/K_m of this fast mutant was only 2-fold higher than that of the wild-type (Griffiths and Tawfik, 2003). Although laboratory-evolved enzymes may show dramatic switches in substrate specificity, the catalytic efficiency with the new substrate is usually orders of magnitude lower than that of the wild-type enzyme using the principal substrate; for

example, aspartate aminotransferase was selected for the ability to use the non-cognate substrate valine (Yano *et al.*, 1998). The best variant had an impressive 10^5 -fold increase in k_{cat}/K_m with valine, but the catalytic efficiency of $110 \text{ s}^{-1} \text{ M}^{-1}$ was 1000-fold lower than that of the wild-type with the principal substrate, aspartate ($120\,000 \text{ s}^{-1} \text{ M}^{-1}$). Our results provide a rare example of a laboratory-evolved enzyme where the catalytic efficiency with the new substrate surpasses that of the wild-type enzyme with its normal substrate.

Interestingly, the mutant T29 was more efficient than wild-type *M.HaeIII* with all DNA substrates tested and used CGCC slightly more efficiently than AGCC (Table III, Figure 5). Although three substrates, AGCC, CGCC and GGCC, are methylated with high efficiency, T29 has not simply lost the ability to discriminate the first base pair and shows greater selectivity against thymine in position 1 than does the wild-type enzyme. T29 also shows better discrimination than the wild-type against thymine in position 4. Overall, the sequence specificity of T29 is VGCC (V = A, C or G). Wild-type *M.HaeIII* exhibits a 22-fold preference for GGCC over AGCC whereas T29 has a 3.3-fold preference for AGCC over GGCC, representing a 70-fold switch in specificity.

The improvement in methylation of CGCC and GGCC is surprising, but not completely unexpected, as there was no selective advantage in using these substrates but no disadvantage either. 'Negative selection' against unwanted activities is probably important for the development of a new substrate specificity, as demonstrated in the directed evolution of Cre recombinase and tRNA synthetase variants with altered specificity (Buchholz and Stewart, 2001; Santoro and Schultz, 2002; Santoro *et al.*, 2002). IVC may also be used for simultaneous positive selection for methyltransferases with the desired sequence specificity in parallel with negative selection against enzymes with an alternative or degenerate specificity. After breaking the emulsion and capturing the DNA on beads, the DNA is digested first with *NheI*, so that only those genes methylated at AGCC remain. Digestion with *HaeIII* then releases those genes that are unmethylated at GGCC from the beads. However, a single round of selection of library B in this way did not improve the specificity of the pool of genes (data not shown).

The permutations of mutations that confer improvements in activity towards AGCC are shown in Table I. The consensus that emerged from the randomized loops over the course of selection provides clues to the importance of each residue in sequence recognition. Arg225 is replaced by alanine in every improved variant selected from library A and the R225A mutation alone confers a 20-fold increase in activity using AGCC (Table I). At least some of the additional mutations present in the T29 mutant must significantly contribute to increased activity on AGCC as T29 is 34 times more active than R225A on AGCC (Table I). It seems likely that the N260L mutation is important as this mutation was found in 12 out of 13 mutants selected from library B. Likewise, the N262W mutation is probably significant as it was found in 10 out of 13 mutants selected from library B. In contrast, L261 was mutated to a wide range of different sequences in the selected clones (Table I). Despite this, the L261M mutation found in T29 may still be important for optimizing activity as 10 other clones also contain the R225A, N260L and N262W mutations yet have lower activity than T29 on AGCC (Table I). None of the amino acid substitutions in the TRD of T29 would be likely to occur in

a library generated by error-prone PCR, as each requires more than one base change per codon.

Nine sequences of GGCC-specific C^5 methyltransferases (including *M.HaeIII*) are known (Roberts *et al.*, 2003). In all cases except for *M.HaeIII*, the residue corresponding to Asn262 is aromatic (usually tryptophan). It is interesting that Asn262 is replaced by tryptophan in the mutant T29. It is possible that this tryptophan makes a hydrophobic interaction with the side chain residue 264 (phenylalanine in all GGCC-specific C^5 methyltransferases, including *M.HaeIII*). The side chain of Arg225 is thought to form hydrogen bonds with the first base, G1, and with the side chain of Asn260. In all nine enzymes Arg225 is conserved and the amino acid at position 260 is Asn or Asp. In T29 the substitution of Asn260 with Leu may compensate for the mutation R225A by removing a polar side chain from the core of the TRD.

The four amino acids mutated in the TRD of T29 are highlighted in Figure 6. The crystal structure of the *M.HaeIII*:DNA complex suggests that recognition of the first base pair is achieved by hydrogen bonds between side chain of Arg 225 and the first base, G1, and between Arg229 and a non-bridging phosphoryl oxygen in the backbone immediately 5' to G1. No hydrogen bonds to the complementary cytosine, C1', are predicted (Reinisch *et al.*, 1995). In T29, the putative hydrogen bonding interaction between Arg225 and G1 certainly cannot be replaced by Ala225. Removal of this base-specific interaction might be expected to increase the K_m^{DNA} , decrease the catalytic efficiency and cause the complete loss of specificity towards the first base pair of the target site (leading to the specificity NGCC), but these changes were not observed. This is a surprising result which illustrates the power of directed evolution by IVC to solve the problem of sequence recognition in unexpected ways, creating enzymes with sequences that would never have been predicted by rational design.

It is generally thought that direct contacts to the DNA bases are the primary determinant of sequence specificity in DNA binding proteins (Garvie and Wolberger, 2001). However, the DNA sequence also affects the conformation of the phosphate backbone (el Hassan and Calladine, 1996). Interactions with the phosphate groups of the backbone (indirect readout) can play a role in sequence recognition, as observed in DNA recognition by the restriction enzyme *EcoRV* (Winkler *et al.*, 1993), the trp repressor (Otwinowski *et al.*, 1988) and the glucocorticoid receptor (Luisi *et al.*, 1991). In the case of *M.HaeIII*, 11 amino acids are thought to interact with the DNA phosphate backbone and discrimination of the first base pair by mutant T29 might be entirely due to backbone contacts.

Alternatively, the mutations in T29 may cause conformational changes in the TRD. It is also possible that some of the residues responsible for recognizing the first base pair were not identified in the crystal structure of the *M.HaeIII*:DNA complex, as this structure reveals only one of several states in the multistep pathway of DNA binding and recognition (Klimasauskas *et al.*, 1998).

These results illustrate the power of directed evolution for protein engineering problems that are too complex to solve by rational design and demonstrate the usefulness of IVC for the rapid selection of enzymes with novel properties. IVC offers many advantages over *in vivo* screening or selection methods, including flexibility in the length of reactions, the contents of the compartments and the ability to work with substrates or enzymes that show *in vivo* toxicity. In this case the methylation reaction took place over 4 h in a solution of 16% glycerol. This

increases the likelihood of non-canonical methylation, makes the system very sensitive and allows the selection of enzymes with very low activities, which is highly beneficial when attempting to select a novel enzyme activity. The use of PCR-generated libraries is also extremely advantageous, as many alternative substrates could be incorporated, for example mismatched or unnatural bases, to adapt this procedure for the selection of enzymes with alternative desirable activities.

Acknowledgements

We thank Greg Winter, Philipp Holliger and Paul Dear for critical reading of the manuscript and Leo James for the preparation of Figure 6. H.M.C. was supported by a Medical Research Council studentship. D.S.T. was supported by an Israel Science Foundation grant. Our groups are members of the European Network on Directed Evolution, ENDIRPRO.

References

- Beck, C., Cranz, S., Solmaz, M., Roth, M. and Jeltsch, A. (2001) *Biochemistry*, **40**, 10956–10965.
- Brannigan, J.A. and Wilkinson, A.J. (2002) *Nat. Rev. Mol. Cell Biol.*, **3**, 964–970.
- Buchholz, F. and Stewart, A.F. (2001) *Nat. Biotechnol.*, **19**, 1047–1052.
- Chen, L., MacMillan, A.M., Chang, W., Ezaz-Nikpay, K., Lane, W.S. and Verdine, G.L. (1991) *Biochemistry*, **30**, 11018–11025.
- Cheng, X. and Blumenthal, R.M. (eds) (1999) *S-Adenosylmethionine-dependent Methyltransferases: Structures and Functions*. World Scientific, Singapore.
- Chiang, C.M. and Roeder, R.G. (1993) *Pept. Res.*, **6**, 62–64.
- Cohen, H.M., Tawfik, D.S. and Griffiths, A.D. (2002) *Nucleic Acids Res.*, **30**, 3880–3885.
- Copley, S.D. (2003) *Curr. Opin. Chem. Biol.*, **7**, 265–272.
- Doi, N. and Yanagawa, H. (1999) *FEBS Lett.*, **457**, 227–230.
- Dryden, D.T.F. (1999) In Cheng, X. and Blumenthal, R.M. (eds) (1999) *S-Adenosylmethionine-dependent Methyltransferases: Structures and Functions*. World Scientific, Singapore, pp. 283–340.
- el Hassan, M.A. and Calladine, C.R. (1996) *J. Mol. Biol.*, **259**, 95–103.
- Evans, S.V. (1993) *J. Mol. Graph.*, **11**, 134–138.
- Fersht, A. (1999) *Structure and Mechanism in Protein Science*. Freeman, San Francisco.
- Friedrich, T., Fatemi, M., Gowhar, H., Leismann, O. and Jeltsch, A. (2000) *Biochim. Biophys. Acta*, **1480**, 145–159.
- Gann, A.A., Campbell, A.J., Collins, J.F., Coulson, A.F. and Murray, N.E. (1987) *Mol. Microbiol.*, **1**, 13–22.
- Garvie, C.W. and Wolberger, C. (2001) *Mol. Cell*, **8**, 937–946.
- Ghadessy, F.J., Ong, J.L. and Holliger, P. (2001) *Proc. Natl Acad. Sci. USA*, **98**, 4552–4557.
- Griffiths, A.D. and Tawfik, D.S. (2000) *Curr. Opin. Biotechnol.*, **11**, 338–353.
- Griffiths, A.D. and Tawfik, D.S. (2003) *EMBO J.*, **22**, 24–35.
- Gubler, M., Braguglia, D., Meyer, J., Piekarowicz, A. and Bickle, T.A. (1992) *EMBO J.*, **11**, 233–240.
- James, L.C. and Tawfik, D.S. (2003) *Trends Biochem. Sci.*, **28**, 361–368.
- Jensen, R.A. (1976) *Annu. Rev. Microbiol.*, **30**, 409–425.
- Kiss, A., Posfai, G., Zsurka, G., Rasko, T. and Venetianer, P. (2001) *Nucleic Acids Res.*, **29**, 3188–3194.
- Klimasauskas, S., Nelson, J.L. and Roberts, R.J. (1991) *Nucleic Acids Res.*, **19**, 6183–6190.
- Klimasauskas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994) *Cell*, **76**, 357–369.
- Klimasauskas, S., Szyperski, T., Serva, S. and Wuthrich, K. (1998) *EMBO J.*, **17**, 317–324.
- Lange, C., Wild, C. and Trautner, T.A. (1996) *EMBO J.*, **15**, 1443–1450.
- Lanio, T., Jeltsch, A. and Pingoud, A. (1998) *J. Mol. Biol.*, **283**, 59–69.
- Lanio, T., Jeltsch, A. and Pingoud, A. (2000) *Protein Eng.*, **13**, 275–281.
- Lee, Y.F., Tawfik, D.S. and Griffiths, A.D. (2002) *Nucleic Acids Res.*, **30**, 4937–4944.
- Lindstrom, W.M., Jr, Flynn, J. and Reich, N.O. (2000) *J. Biol. Chem.*, **275**, 4912–4919.
- Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R. and Sigler, P.B. (1991) *Nature*, **352**, 497–505.
- Martin, J.L. and McMillan, F.M. (2002) *Curr. Opin. Struct. Biol.*, **12**, 783–793.
- Matsumura, I. and Ellington, A.D. (2001) *J. Mol. Biol.*, **305**, 331–339.
- Mi, S. and Roberts, R.J. (1992) *Nucleic Acids Res.*, **20**, 4811–4816.
- O'Brien, P.J. and Herschlag, D. (1999) *Chem. Biol.*, **6**, R91–R105.

- Otwinowski, Z., Schevitz, R.W., Zhang, R.G., Lawson, C.L., Joachimiak, A., Marmorstein, R.Q., Luisi, B.F. and Sigler, P.B. (1988) *Nature*, **335**, 321–329.
- Reich, N.O. and Mashhoon, N. (1991) *Biochemistry*, **30**, 2933–2939.
- Reich, N.O., Olsen, C., Osti, F. and Murphy, J. (1992) *J. Biol. Chem.*, **267**, 15802–15807.
- Reinisch, K.M., Chen, L., Verdine, G.L. and Lipscomb, W.N. (1995) *Cell*, **82**, 143–153.
- Rimseliene, R., Maneliene, Z., Lubys, A. and Janulaitis, A. (2003) *J. Mol. Biol.*, **327**, 383–391.
- Roberts, R.J., Vincze, T., Posfai, J. and Macelis, D. (2003) *Nucleic Acids Res.*, **31**, 418–420.
- Roth, M. and Jeltsch, A. (2001) *Nucleic Acids Res.*, **29**, 3137–3144.
- Samuelson, J.C. and Xu, S.Y. (2002) *J. Mol. Biol.*, **319**, 673–683.
- Santoro, S.W. and Schultz, P.G. (2002) *Proc. Natl Acad. Sci. USA*, **99**, 4185–4190.
- Santoro, S.W., Wang, L., Herberich, B., King, D.S. and Schultz, P.G. (2002) *Nat. Biotechnol.*, **20**, 1044–1048.
- Schluckebier, G., O'Gara, M., Saenger, W. and Cheng, X. (1995) *J. Mol. Biol.*, **247**, 16–20.
- Schottler, S., Wenz, C., Lanio, T., Jeltsch, A. and Pingoud, A. (1998) *Eur. J. Biochem.*, **258**, 184–191.
- Sepp, A., Tawfik, D.S. and Griffiths, A.D. (2002) *FEBS Lett.*, **532**, 455–458.
- Szomolanyi, E., Kiss, A. and Venetianer, P. (1980) *Gene*, **10**, 219–225.
- Tawfik, D.S. and Griffiths, A.D. (1998) *Nat. Biotechnol.*, **16**, 652–656.
- Thomas, M. and Davis, R.W. (1975) *J. Mol. Biol.*, **91**, 315–328.
- Trautner, T.A., Pawlek, B., Behrens, B. and Willert, J. (1996) *EMBO J.*, **15**, 1434–1442.
- Urig, S., Gowher, H., Hermann, A., Beck, C., Fatemi, M., Humeny, A. and Jeltsch, A. (2002) *J. Mol. Biol.*, **319**, 1085–1096.
- Vilkaitis, G., Merkiene, E., Serva, S., Weinhold, E. and Klimasauskas, S. (2001) *J. Biol. Chem.*, **276**, 20924–20934.
- Vilkaitis, G., Lubys, A., Merkiene, E., Timinskas, A., Janulaitis, A. and Klimasauskas, S. (2002) *Nucleic Acids Res.*, **30**, 1547–1557.
- Walter, J., Trautner, T.A. and Noyer-Weidner, M. (1992) *EMBO J.*, **11**, 4445–4450.
- Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K. and Wilson, K.S. (1993) *EMBO J.*, **12**, 1781–1795.
- Yano, T., Oue, S. and Kagamiyama, H. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 5511–5515.

Received September 10, 2003; accepted September 23, 2003

Edited by Greg Winter