

## SURVEY AND SUMMARY

# The other face of restriction: modification-dependent enzymes

Wil A. M. Loenen<sup>1,\*</sup> and Elisabeth A. Raleigh<sup>2,\*</sup>

<sup>1</sup>Leiden University Medical Center, P.O. Box 9600 2300RC Leiden, The Netherlands and <sup>2</sup>New England Biolabs Inc., 240 County Road Ipswich, MA 01938-2723, USA

Received April 15, 2013; Revised July 2, 2013; Accepted July 10, 2013

### ABSTRACT

The 1952 observation of host-induced non-hereditary variation in bacteriophages by Salvador Luria and Mary Human led to the discovery in the 1960s of modifying enzymes that glucosylate hydroxymethylcytosine in T-even phages and of genes encoding corresponding host activities that restrict non-glucosylated phage DNA: *rgIA* and *rgIB* (restricts glucoseless phage). In the 1980's, appreciation of the biological scope of these activities was dramatically expanded with the demonstration that plant and animal DNA was also sensitive to restriction in cloning experiments. The *rgI* genes were renamed *mcrA* and *mcrBC* (modified cytosine restriction). The new class of modification-dependent restriction enzymes was named Type IV, as distinct from the familiar modification-blocked Types I–III. A third *Escherichia coli* enzyme, *mrr* (modified DNA rejection and restriction) recognizes both methylcytosine and methyladenine. In recent years, the universe of modification-dependent enzymes has expanded greatly. Technical advances allow use of Type IV enzymes to study epigenetic mechanisms in mammals and plants. Type IV enzymes recognize modified DNA with low sequence selectivity and have emerged many times independently during evolution. Here, we review biochemical and structural data on these proteins, the resurgent interest in Type IV enzymes as tools for epigenetic research and the evolutionary pressures on these systems.

### GENETICS, BIOCHEMISTRY AND STRUCTURES

#### Historical sketch

Like conventional modification-blocked restriction, modification-dependent restriction originally was diagnosed

owing to its biological effects, when interstrain DNA transfer was unexpectedly inhibited. At the start, phages were the investigatory vehicles, moving between *Escherichia coli* K12, *E. coli* B and *E. coli* C or *Shigella dysenteriae* Sh (1,2). Later, reduced plasmid, phage or chromosomal transfer was found when alien modification patterns were present (3–5). Incoming DNA needed the endogenous (for *E. coli* K12) modification of Am6ACN(6)GTGC (M.EcoKI; the A opposite the underlined T is also modified) and Gm6ATC (Dam); Cm5CWGG (Dcm) occasionally had effects (6). ‘Outgoing’ DNA was better accepted in many taxa without any of these (7–10).

Progress in cloning and sequencing of restriction enzyme (REase) genes, other nucleases, methyltransferase (MTase) genes and motor proteins began to feed data into efforts to classify sequences and abstract from them signatures predictive of particular functions, e.g. (11–15). Such signatures often correlate with physical protein domains. These domains can be split off from the original protein and added to another and will then operate (mostly) as they are supposed to. This result is the basis for protein tagging with reporters and epitopes by molecular biologists. As we see from the structural organization of modification-dependent REases, this apparently is also the basis for a mix-and-match evolutionary process in real life—grab a DNA-binding domain here, a nuclease domain there, and you’ve got a site-specific (sort of) nuclease! Sometimes, a dimerization surface or a regulatory domain is needed as well.

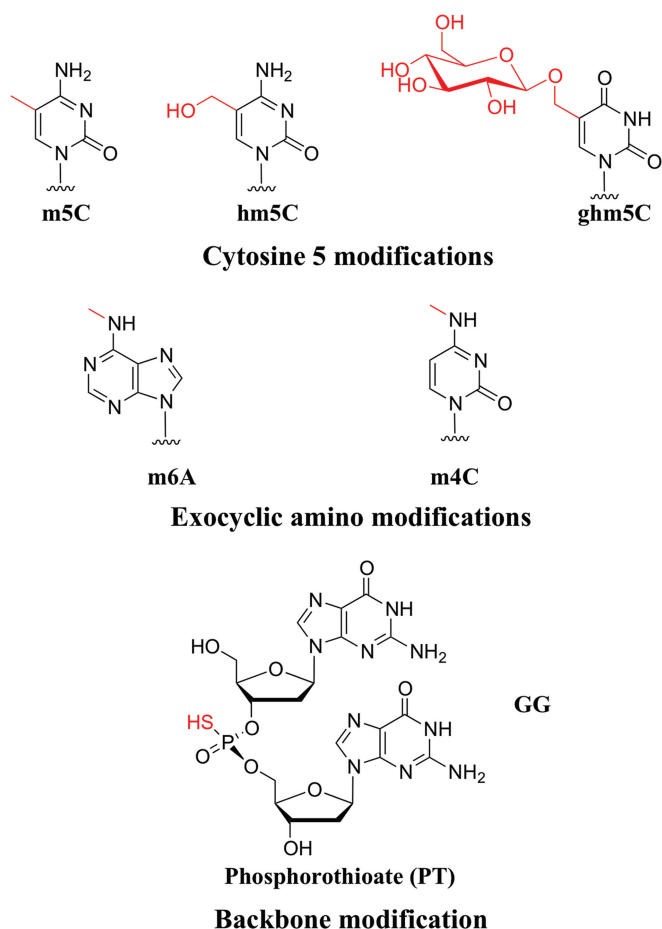
Finally, with the advent of massive genome sequencing, bioinformatic analysis has become a hypothesis generator so that well-chosen biological and enzymatic tests can (hopefully) allow quick creation of strains and enzymes for further research (16).

#### What biological DNA modifications are there?

Biological DNA modifications have been studied for many years, and much is known about their distribution

\*To whom correspondence should be addressed. Tel: +31 85 878 0248; Email: w.a.m.loenen@lumc.nl  
Correspondence may also be address to Elisabeth A. Raleigh. Tel: +1 978 380 7238; Fax: +1 978 921 1350; Email: Raleigh@neb.com

The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.



**Figure 1.** DNA modifications recognized by Type IV enzymes. Enzymatic DNA modifications in the major groove of double-stranded DNA are methylation at cytosine C5 or N4, or at adenine N6; and glucosylation of a pre-existing 5-hydroxymethylcytosine. The beta-glucosyl derivative is shown; other configurations and other sugars are known to be added by some phages. hm5C is incorporated during replication, after conversion of the dCTP pool to hmdCTP. Phosphorothioate modification of the backbone is carried out postsynthetically. Other biological DNA modifications are known. Only those shown to elicit action of characterized Type IV enzymes are shown here.

and the enzymes involved (17–19). Well-known base modifications are C-5-methylcytosine (m5C), N4-methylcytosine (m4C) and N6-methyladenine (m6A) (Figure 1). These are widely distributed in cellular organisms, particularly prokaryotes. Other base modifications have long been known in bacteriophage, prominently 5-hydroxymethylcytosine (hm5C) and derivatives of it with sugar residues attached (ghm5C) (Figure 1), and 5-hydroxymethyluracil (hm5U). Unusual modifications of adenine have also been studied in phage Mu [Mom modification (20)]. Fairly recently, as methods for detection of low frequency modifications have improved, some of these exotic base modifications have also been recognized in higher organisms [hm5C: (21,22)] and lower eukaryotes [hm5U and the sugar-derivatized J base: (23)]. Bioinformatic investigation of coding sequences related to modification enzymes suggests that additional unrecognized base modifications may still be discovered (24,25).

It is not only bases in DNA that may be modified. Enzymatic sulfur modification of the phosphodiester backbone of DNA (PT-DNA, for phosphothioester DNA; Figure 1) has recently been discovered in prokaryotes (26–28). PT-modified DNA is widespread; modification is found in local sequence contexts compatible with sequence-specific addition, and the similarity relationships among the *dnd* genes encoding the modification machinery are consistent with extensive horizontal transfer, as is found for conventional restriction-modification (R-M) systems (29). This opens still further vistas for research on the nature and biological consequences of modification and restriction.

Some of these modifications play important other roles in the life of the host cell, besides restriction wars: in replication timing in prokaryotes and in transcription regulation in prokaryotes and eukaryotes [e.g. (30–33)]. This topic will not be addressed here, except to note that the modifying enzymes that have acquired regulatory effects in bacteria are normally conserved within a clade, unlike cognate-modifying enzymes that accompany R-M systems, which are sporadically distributed (34,35).

### Molecular action: what they do

#### Diversity of modification dependence

Modifications that protect against conventional REases include m5C, hm5C, ghm5C, m6A, m4C and, most recently, PT DNA (with sulfur replacing a non-bridging oxygen). Neither hm5C nor ghm5C are known to be added site-specifically; instead, they are found as universal substitutions in phage DNA. The inverse could also be true: for each protective modification in Figure 1, there are enzymes that attack DNA only when the modification is present (Tables 1 and 2). Many of the enzymes were described only recently and are distinct from the classical examples. Many of the other modifications found in phages (18,19) might be the object of undiscovered Type IV enzymes. hm5U and its glucosylated derivative, J base, and the Mu modification, N6 (1-acetoamido) adenine, would be interesting substrates.

Those modification-dependent enzymes that are classified as Type IV in REBASE (50) have been segregated (Table 1) from those classified as Type IIM (Table 2). The distinction between Type IIM and Type IV appears to reflect production of defined bands on a gel in the reported characterizations. This distinction may be misleading, as bands on a gel can result from substrate choice in some cases (see further later in the text). As no other fundamental property unites the Type IV enzymes, or distinguishes Type IIM from Type IV, these authors advocate adding Type IIM to the Type IV class.

For the most part, those functions acting on hm5C also act on m5C though with varying efficiency. EcoK Mrr, for which only *in vivo* evidence is available, may be an exception—it does not interfere with growth of hm5C-containing T-even phages. However, phage-encoded restriction inhibitors may confound interpretation of negative results obtained *in vivo* (see later in the text, ‘Phage-host arms race’).

**Table 1.** DNA modifications<sup>a</sup> that elicit cleavage by Type IV enzymes

Protein	m5C	hm5C	ghm5C	m4C	m6A	PT	References
EcoKMcrA	(+)	(+)	(-)	NT	(-)	NT	(3,36)
ScoA3McrA	+	NT	NT	NT	(+)	+	(37,38); Sco4631
EcoKMcrBC	+	+	-	(+)	-	NT	(3,39,40)
BanUMcrB	(+)						(41)
BanUMcrB3	(+)						(41)
EcoKMrr	(+)	(-)	(-)	(-?)	(+)	NT	(42,43)
BanUMrr	(+)				(+?)		(41)
ScoA3Mrr	(-)				(+)		(37); Sco4213
ZmoMrr	(+?)				(+?)		(44); ZMO1932; Zmrr
SauUSI	+	+	-	-	-?	NT	(45)
SauNewI	(+)						(46); NWMN_2386
SepRPMcrR	(+)						(46); SERP2052
ScoA3I	(+)						(37); Sco2863
PvuRtsII family	+/-	+	+	NT	-	NT	(47,48)
GmrSD	-	-	+	NT	-	NT	(49)
ScoA3II+III	(+)				(-)		(37); Sco3261-62

<sup>a</sup>Modifications: m5C: 5-methylcytosine; hm5C: 5-hydroxymethylcytosine; ghm5C: glucosylated hydroxymethylcytosine; m4C: *N*4-methylcytosine; m6A: *N*6-methyladenine; PT: phosphorothioation of non-bridging oxygen in DNA linkages, also called S-DNA.

+/-: at least 100-fold less activity on this substrate than on substrates with + entry.

(-), (+), based on *in vivo* restriction of phage infection or plasmid transformation with appropriate host mutant configurations; *in vitro* cleavage results have not been reported.

(+?) either m5C or m6A is recognized; these were not distinguished in the reported experiments.

-?: m6A sites tested were not cleaved, but few modified sequences were tested.

NT: not tested.

Where the name found in REBASE (and listed at the left) is not the same as that used in the cited report, the genomic locus\_ID is given in the References column, or the name used in the publication.

**Table 2.** DNA modifications<sup>a</sup> that elicit cleavage by other modification-dependent enzymes (Type IIM)

Protein	m5C	hm5C	ghm5C	m4C	m6A	PT	References
DpnI	-	-	-	-	+	NT	(50)
MspJI family	+	+	-	-	-	NT	(51,52)
SgeI	+						(53)
AoxI	+						
BisI	+						
BlsI	+						
GlaI	+						(54)
GluI	+						
KroI	+						
Mall	+						
MteI	+						
PcsI	+						

<sup>a</sup>Modifications: m5C: 5-methylcytosine; hm5C: 5-hydroxymethylcytosine; ghm5C: glucosylated hydroxymethylcytosine; m4C: *N*4-methylcytosine; m6A: *N*6-methyladenine; PT: phosphorothioation of non-bridging oxygen in DNA linkages, also called S-DNA. +/-: at least 100-fold less activity on this substrate than on substrates with + entry.

(-), (+), based on *in vivo* restriction of phage infection or plasmid transformation with appropriate host mutant configurations; *in vitro* cleavage results have not been reported.

(+?) either m5C or m6A is recognized; these were not distinguished in the reported experiments.

-?: m6A sites tested were not cleaved, but few modified sequences were tested.

NT: not tested.

Where the name found in REBASE (and listed at the left) is not the same as that used in the cited report, the genomic locus\_ID is given in the References column, or the name used in the publication.

Other Mrr-related enzymes from *Bacillus anthracis*, *Streptomyces coelicolor* and *Zymomonas mobilis* (identified bioinformatically, see later in the text) were also tested for activity *in vivo*. Transformation efficiency is reduced when a plasmid is prepared from a modifying host, compared with the same plasmid from a non-modifying host; this reduction is alleviated when the corresponding Mrr-related gene is disrupted. The specificity of this test depends on how thorough the genetic investigation was; if Dam<sup>-</sup> Dcm<sup>-</sup> EcoKM<sup>+</sup> DNA transforms better than fully modified DNA, modification specificity could be either m6A or m5C or both, hence the question marks in the table.

The four systems listed for *S. coelicolor* 3A constitute a particularly exemplary analysis of this kind (37). In this case, all four candidate R-M systems were deleted individually and together so that the effect of each could be tested, and each system was established in the related non-restricting host *Streptomyces lividans*. For ScoA3Mrr, the effect of removing modifiable sites from the test plasmid was also examined (for M.EcoKI).

### Diversity of functional organization

Unlike the classic Type IIP enzymes such as EcoRI and BamHI, in which catalytic residues are embedded within sequence-recognition structural elements, the modification-dependent enzymes known so far exhibit separation of DNA binding and cleavage into different domains on the same protein, or even into different polypeptide chains (Table 3 and 4). In this they resemble Type I, Type IIS or Type III enzymes, modification-blocked enzymes that also separate recognition and cleavage. For those also,

**Table 3.** Characteristics of Type IV restriction enzymes

Protein	Subunits/ domains	DNA Binding	Endonuclease domain	NTP hydrolysis	Recognition site <sup>a</sup> , comment and references
EcoKMcrA	N-terminal	DBD		–	<i>In vitro</i> cleavage not reported (Y > R)m5CGR bound (55,56); hm5C/m5C discrimination (57)
	C-terminal		H-N-Hc		Bioinformatic ID (58); required for damage to DNA <i>in vivo</i> (57)
ScoA3McrA				–	Some CmCWGG and some S-DNA (PT modified) sites are cleaved (38)
EcoKMcrBC	N-terminal	DBD?			Not related to EcoMcrA (38)
	C-terminal		H-N-Hc		37% identical to EcoMcrA (38)
EcoKMcrBC	McrB-N	DUF3578		GTP	Rm5C(N30-35)-(N30-3000)-Rm5C <sup>b</sup>
	McrB-C McrC		PD-(D/E)XK?	P-loop NTPase	McrB binds DNA (40) via its N-terminal domain (59), by extrahelical modified base (60) (61,62,63)
EcoKMrr	N-terminal	Mrr-N		–	Bioinformatic ID (64); Required for cleavage (39,65)
	C-terminal		Mrr-cat (D/E).. (D/E/Q) × K		m6A or m5C; sequence specificity ambiguous (42,43,66)
SauUS1	N-terminal		PLDc-2	ATP or dATP	Sm5CNGS; two copies required for cleavage (45)
	Middle C-terminal	DUF3427		P-loop NTPase	(45)
PvuRtsII			PD-(D/E)XK?	–	(45); DBD? mC(N11-13/N9-10)G 2-base extensions (47)
EcoCTGmrSD		?	?	UTP >> GTP, CTP	Bioinformatic ID (64)
	GmrS GmrD		Motifs suggested DUF1524	DUF262	Cuts T-even DNA (49) (71); To be confirmed To be determined

<sup>a</sup>Recognition sites (represented 5'→3') are those determined *in vitro* by binding or cleavage experiments.

<sup>b</sup>McrBC cleavage results in a double-strand cut near one Rm5C site (72,73,74) but requires cooperation of two sites (39,40) or a translocation block (73). The sites may be on different daughters across a fork (75). These are separated by 30–3000 (39,72,74) and may be on either strand (39,76); disposition of opposing nicks is not tightly constrained (73), and minor cleavage clusters are found ~40, ~50 and ~60 nt from the m5C (74). Degeneracy abbreviations: **B** = C or G or T; **D** = A or G or T; **H** = A or C or T; **K** = G or T; **M** = A or C; **N** = A or C or G or T; **R** = A or G; **S** = C or G; **V** = A or C or G; **W** = A or T; **Y** = C or T.

Cleavage positions are listed as (N# to top cut/# to bottom cut). If no number is listed, the position of cleavage is not determined. Space between numbers (e.g. PvuRtsII N11-13/N9-10) indicates the range of positions at which cleavage may occur.

multiple evolutionary events apparently have occurred to connect nuclease domains to recognition moieties (81).

### Nuclease domains

Enzymes that recognize modified DNA with minimal sequence selectivity have emerged at least six times, as exemplified by the McrA, McrBC, SauUSI, Mrr, PvuRtsII and GmrSD families. These exemplars are discussed in more detail later in the text. In brief, nuclease domains have been attached covalently or (for McrC) via protein–protein interaction to domains with DNA binding and regulatory functions.

EcoKMcrA carries a C-terminal H-N-Hc nuclease domain identified bioinformatically (58,82) (Figure 2). This nuclease domain is also found in modification-blocked nucleases (81). The purified binding-competent protein did not cleave under a variety of buffer conditions and cofactor additions (55). ScoA3McrA is designated 'McrA' due to its possession of a similar nuclease domain. For this enzyme, cleavage depends on Mn<sup>2+</sup> or Co<sup>2+</sup> (38) and occurs at a variable distance from PT-modified sites. Modification-blocked H-N-H REases also often exhibit unusual metal ion requirements [e.g. (83)].

McrBC: The required McrC component (39,40) is the nuclease moiety (65) (Figure 3). Mutational analysis

confirms that it is a PD-(D/E)XK nuclease (65), in agreement with bioinformatic classification (64). Cleavage results when McrC associates with full-length McrB:GTP complex bound to DNA and GTP is hydrolyzed (72). LlaJI, a modification-blocked restriction activity, exhibits a similar organization (85), although cleavage could not be demonstrated *in vitro*.

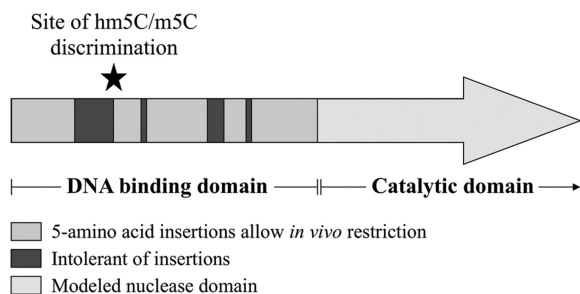
The classic modification-dependent enzyme DpnI also carries a PD-(D/E)XK motif (see further later in the text).

Mrr: EcoKMrr contains a variant of the PD-(D/E)XK motif (68,69) with the Mrr-N (*E. coli* K12) presumed DNA-binding domain. MspJI (see further later in the text) also carries a nuclease domain in this family. As with McrA, McrBC, and SauUSI, nuclease domain similarity does not in itself dictate modification preference properties: the single-chain R-M system LlaGI has conserved motifs characteristic of the *E. coli* Mrr protein, but this enzyme does not target methylated DNA (86).

SauUSI: This is a modification-dependent enzyme with a phosphodiesterase cleavage domain akin to one originally identified in phospholipase-D (45). Mutation of any of the four conserved catalytic residues abolishes *in vitro* activity. This cleavage domain is also found in stand-alone

**Table 4.** Characteristics of other modification-dependent enzymes (Type IIM)

Protein	Subunits/ domains	DNA Binding	Endonuclease Domain	Recognition site	Comment and references
DpnI family	N-terminal C-terminal	wHTH	PD..(D/E)XK	G m6A TC	13 characterized isoschizomers (50,77,78,79)
MspJI family				R m6A TC	(78)
MspJI				mC with preferences	Second copy stimulates cleavage
				5 mCNNR(N9/13)	(51)
	N-terminal C-terminal	SRA-like	Mrr-cat (D/E)..(D/E/Q)XK		(80)
FspEI	Like MspJI			C m5C(N12/16)	(52)
LpnPI	Like MspJI			C m5CDG(N10/14)	(52)
AspBHI	Like MspJI			YS m5CNS(N8/12)	(52)
RlaI	Like MspJI			V m5CW	(52)
SgrTI	Like MspJI			C m5CDS(N10/14)	(52)
SgeI	Like MspJI			m5CNNR(N9/13)	49% identical to MspJI; (53)
No family assigned					Information from <a href="http://www.sibenzyme.com/products/m2_type">http://www.sibenzyme.com/products/m2_type</a>
AoxI				RG m5CY	
BisI				G m5C NGC	
BlsI				RYN R Y	At least two m5C required
GlaI				R m5C GY	(54)
GluI				GmC NG m5C	
KroI				G  m5CCGGC	
MteI				GmCG m5C NGm5CGm5C	
PesI				m5CG(N5 N2)m5CG	
PkrI				Gm5CN G m5C	At least 3 m5C required



**Figure 2.** McrA functional domains. Domain function was inferred indirectly from genetic analysis by Anton & Raleigh (2004) (57). Many mutations in the N-terminal domain spared some activity in one or more of three functional tests (grey segments) while others were deficient in all activities (black segments). One mutation (asterisk) was fully active on m5C-containing substrates, but fully inactive in the hm5C challenge *in vivo*. Most mutations in the C-terminus (pale grey segment) retained function in one test that was interpreted as measuring m5C binding ability. A predicted structural model by Bujnicki, Radlinska and Rychlewski (2000) (58) for this C-terminal region is compatible with these results.

nucleases and modification-blocked REases (87,88). Interestingly, two of the PLDc nuclease activities have been shown to work by a transesterification reaction like that used by topoisomerases and transposases (87,89).

PvuRts1I has an apparently unusual nuclease domain [i.e. not yet identified by sources curated by the NCBI Conserved Domain Database (90)]. However, this enzyme was included in a categorization of PD-(D/E)XK families (64); a tentatively identified divalent metal ion binding site, Block B (47), corresponds to Block D of Bujnicki and Rychlewski (64). Cleavage requires Mg<sup>2+</sup> ions.

EcoCTGmrSD: Functional organization is less clear but several possible nuclease motifs were identified in GmrS (71). Cleavage buffer contained Ca<sup>2+</sup> and Mg<sup>2+</sup> ions, and UTP.

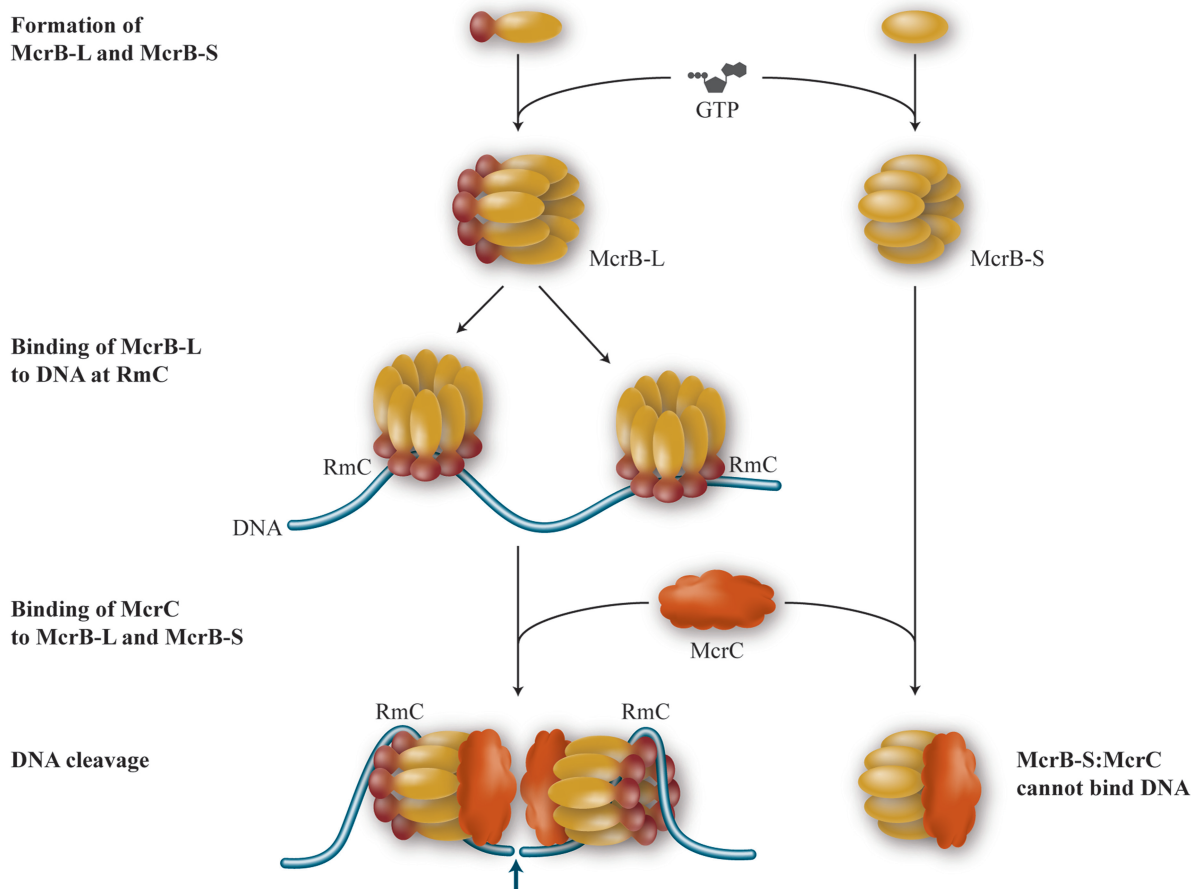
### Sequence context recognition

Many of the modification-dependent enzymes characterized so far have little sequence specificity, in contrast to conventional modification-blocked REases. Relatively complete characterization of sequence preference and cleavage position has been carried out for Type IV enzymes EcoKMcrBC, SauUSI and PvuRts1I (Table 3) and for Type IIM DpnI and the MspJI family (Table 4). Progress has been made with binding recognition for EcoKMcrA. Cleavage conditions have been achieved for Sco3AMcrA (Table 3). For all of these, recognition of surrounding sequence context is degenerate, with preference for a neighboring base and frequently a requirement for two sites with suitable separation. DpnI is in some respects an exception, see later in the text.

The remaining nucleases in Table 4 are less well characterized. The recognition sites might form a related series. It will be interesting to learn more about the relationships among these, and how the requirement for multiple modified positions is specified, e.g. for BlsI and PkrI.

### McrA binding domains

The two 'McrA' enzymes are not similar in their N-termini, with homology limited to the C-terminal nuclease domain. For EcoKMcrA (Figure 2), there is good genetic evidence that base recognition lies in the N-terminus. Extensive mutagenesis using insertion of



**Figure 3.** McrBC Assembly Model. Two proteins are expressed from *mcrB* *in vivo*. Both the complete protein (McrB-L) and a small one missing the N-terminus (McrB-S; top row) bind GTP, forming high-order multimers detected by gel filtration (second row). When visualized by scanning transmission electron microscopy, these appear as heptameric rings with a central channel. Rings of McrB-L in top views show projections that may correspond to the N-terminal DNA-binding domain (red segment). Both forms can then associate with McrC, judged again by gel filtration. McrB-L:GTP can bind to its specific substrate (RmC) in the absence of McrC (third row); in its presence, the substrate is cleaved (fourth row). GTP hydrolysis is required for cleavage (arrow): a supershifted binding complex forms in the presence of GTP-gamma-S, but no cleavage occurs. Translocation accompanies GTP hydrolysis; double-stranded cleavage requires collaboration between two complexes, or a translocation block. The path of the DNA in the figure is arbitrary, as is the conformation of McrC. Modified from Bourniquel, A.A. and Bickle, T.A. Complex restriction enzymes: NTP-driven molecular motors. Bourniquel and Bickle (84) with permission. Copyright © 2002 Elsevier Masson SAS. All rights reserved.

five-amino acid linkers and classification with three functional tests allowed assignment of DNA recognition to the N-terminal portion, with the C-terminal H-N-H domain implicated in cleavage. Of particular note, a mutation discriminated *in vivo* between hm5C and m5C was found in the N-terminal domain (57). The mutant was able to fully restrict bacteriophage lambda modified by M.HpaII, but not at all phage T4 containing hm5C. *In vitro*, modification-dependent binding was achieved with the full-length His-tagged protein (55,56), yielding a putative recognition site (Y > R)mCG. This recognition site is compatible with *in vivo* observations (3,91).

Presumably, the N-terminus of Sco3A McrA also recognizes the DNA. Recognition of both m5C and the phosphorothioate (PT) moiety must be accommodated in the final reaction. As either modification is sufficient to elicit cleavage, more than one domain could be involved. Cleavage occurred near some but not all Dcm-modified sites (Cm5CWGG). Both synthetic

PT-containing oligonucleotides and unmethylated PT-modified plasmid were also cleaved on both sides of a symmetrically modified site. PT modification is thought to be sequence-specific (26,29), but the details are not yet clear.

#### **Novel McrB binding domain**

DNA-binding resides in the McrB N-terminal domain, (40,59,60), whereas translocation and cleavage coordination reside in the C-terminal AAA+ regulatory and translocation domain (61–63,72,73). The complete translation product of *mcrB*, McrB-L, binds DNA specifically (40,72) via its N-terminal 161 amino acids (aa) (59). The crystal structure of this domain (McrB-N) in complex with DNA has recently been published (60).

McrB-N uses a strategy first discovered for DNA-MTase action (92): it flips the C base out of the DNA helix into a binding pocket for inspection. The pocket is large enough to accommodate C, m5C, hm5C or m4C, but too small if a

glucose moiety is attached. Conserved residues Y64 and L68 were noted to make van der Waals contact with the methyl group of the flipped out m5C; these contacts are missing when the pocket contains C.

The flipping action can be compared with, but is distinct from that of, eukaryotic m5C-specific regulatory proteins that use the SET and Ring-finger-Associated (SRA) domain to read DNA modification state (Figure 4A). This domain is found in most eukaryotes, in accessory proteins (e.g. UHRF1/NP95/SUVH5) of the DNMT1 maintenance MTase, (93–95). Despite the similar strategy, the McrB-N domain is not homologous but displays a distinct protein fold (60). Binding is accomplished from the minor groove, and extraction of the C creates a 30° bend toward the major groove, resembling a glycosylase in this respect (Figure 4A). The eukaryotic proteins form a crescent from which loops project to wrap around the DNA, with recognition mediated through both major and minor grooves (94). For McrB-N, the authors suggest that the purine preference in the 5' position might result from flexibility constraints or interaction with a non-conserved aa that occupies space left by the flipped base. Substitutions of this aa (Y41A or Y41Q) compromised binding activity.

#### **Sequence specificity, novel phenotype and structural model of Mrr**

In 1987, Heitman and Model discovered Mrr when they found that transfer of various foreign m6A MTases induced an SOS response due to DNA damage (42). This response to the presence of an incompatible MTase remains the principal evidence that the *E. coli* K12 Mrr protein cleaves DNA. Related proteins discussed later in the text (Type IIM) have been more tractable for *in vitro* work. No concise description of the Mrr recognition sequence has been forthcoming, although several studies have examined the spectrum of incompatible MTases (43,66,96,97). Both adenine and cytosine MTases confer sensitivity.

Mrr is also responsible for DNA damage that does not depend on methylation at all, foreign or otherwise. High hydrostatic pressure (HP) induces the SOS DNA damage response and lethality (98). The response did not depend on the activity of the endogenous MTases of *E. coli* K12 but did depend on both the presence of wild-type *mrr* and the integrity of the SOS signal generation pathway. Possibly, HP elicits a non-enzymatic modification or a structural change in DNA helicity that is acted on by Mrr. This HP phenotype was used to characterize *mrr* mutants, which were fitted into a computer-assisted model of the Mrr protein (67). An N-terminal DNA-binding winged helix was proposed, with a C-terminal nuclease domain previously identified (69). The functional importance of several conserved residues was confirmed. Several of the selected mutants with null phenotypes were isolated in a region far from the active site or binding surface identified bioinformatically. These could affect interaction with a component of the HP response. This intriguing collection of informative mutants will illuminate *in vitro* characterization.

#### **Type IIM binding domains**

Type IIM enzymes of two families are well-characterized with respect to cleavage (Table 4). Crystal structures for both have recently appeared.

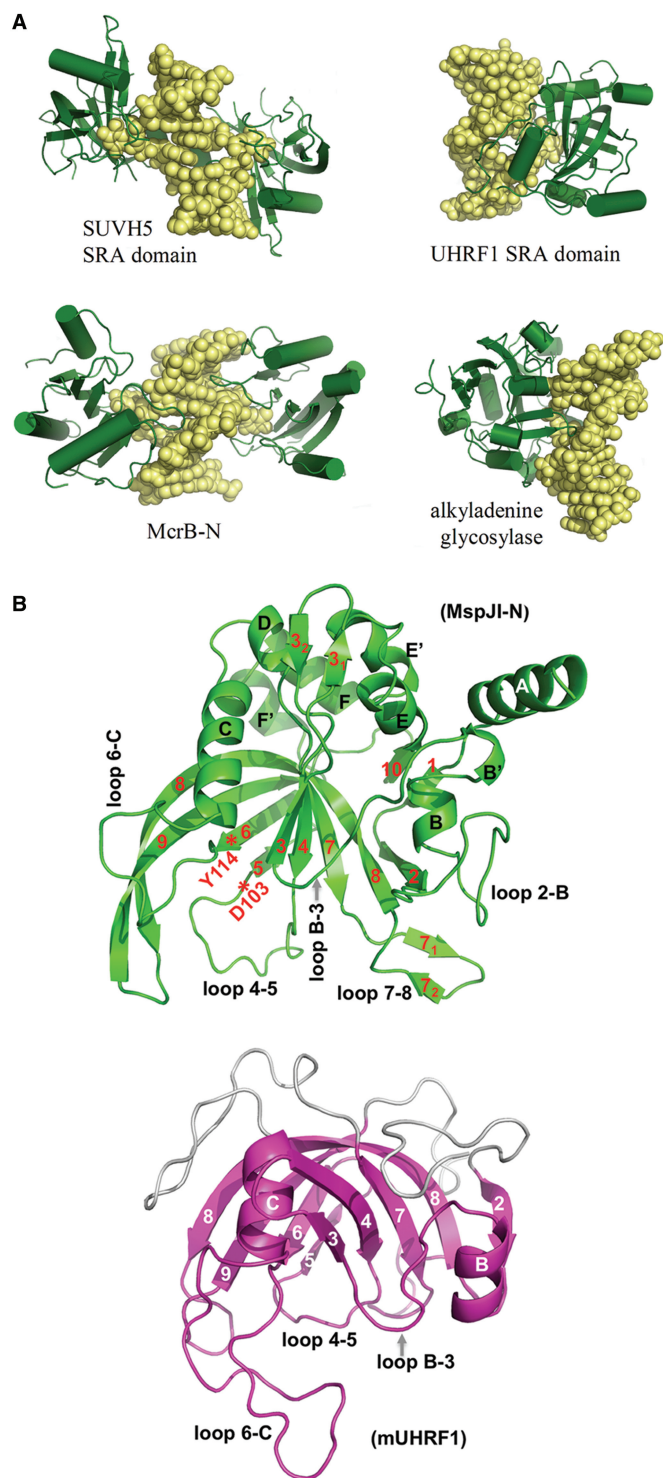
#### **DpnI: winged-helix DNA recognition**

Unusually for modification-dependent enzymes, DpnI cleaves a four-base site (Gm6ATC) with high fidelity (77,99) to leave blunt ends when both strands of the site are methylated. At low concentration, the enzyme nicks the modified strand of a hemimethylated site (100). The behavior of the enzyme with respect to modification patterns within the canonical GATC site—modification of C or A, one strand or both—has been thoroughly explored (50). However, only recently has cleavage of non-canonical adenine-methylated sites been examined. Siwek and co-workers (78) found evidence for considerable relaxation of specificity at the outer base. This experiment used substrates modified by a highly non-specific adenine MTase, extensive DpnI cleavage, cloning of the fragments and sequencing of the borders.

Structure determination in the presence of DNA and validation experiments (78) place this enzyme together with the other modification-dependent enzymes, in that two domains segregate the cleavage function from sequence recognition. It also separates DpnI from the others, in that the cleavage domain also possesses some modification and sequence specificity. The main recognition is accomplished by a monomeric winged-helix domain, which binds in the major groove and recognizes the modifications on both strands in the same event. The structure does not reveal a cleavage-competent complex, however, because the cleavage domain is far from the DNA. Filter-binding experiments validated the ability of the C-terminal domain to bind alone, to do so more tightly to fully methylated than to hemimethylated oligonucleotides, and to compete with the full-length enzyme, reducing cleavage by it. Expression of the N-terminal cleavage domain alone (in low yield) allowed validation of its cleavage activity. Surprisingly, this cleavage was itself dependent on modification state and sequence of the substrate. Modeling based on the structure of the blunt-end-producing Type IIP enzyme PvuII allowed prediction that the cleavage domain approaches from the minor groove. Complete understanding of double-stranded cleavage will depend on understanding the dynamic transformations that allow the cleavage domain to approach and act at the site.

#### **MspJI coupling of cleavage with DNA recognition**

The six members of the MspJI family use the Mrr-cat version of the PD-(D/E)XK nuclease to cut at defined locations to one side relative to the modified base (12 bases on the modified strand, 16–17 on the other; Figure 5A); only one modified base is required for double-strand cleavage to occur (unlike McrBC) (51,52). However, these enzymes are stimulated by the presence of a second site in *cis* or in *trans*. Symmetrically modified sites (such as m5CpG:m5CpG in mammalian DNA) yield prominent bands of defined size (~32 bp) containing a mixed population of sequences each with a m5C in the



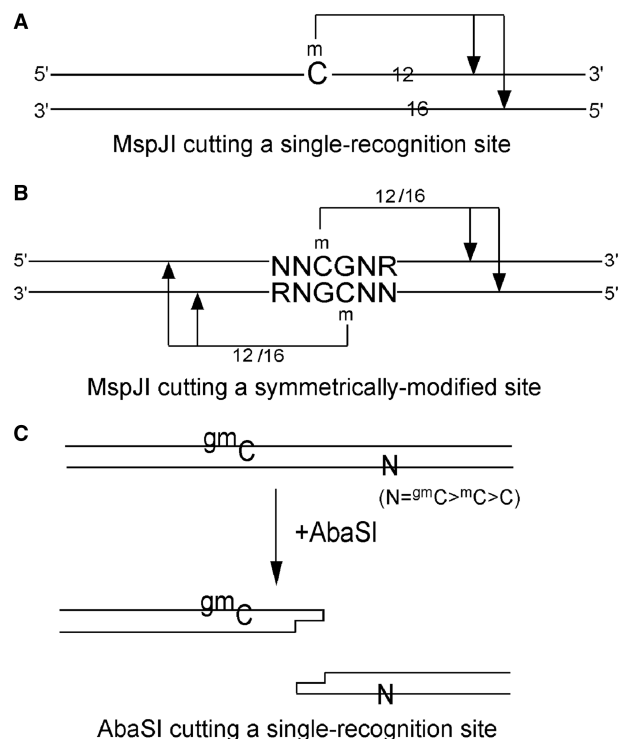
**Figure 4.** McrB-N in comparison to other base-flipping proteins. (A) SRA domains SUVH5 (3Q0C) and UHRF1 (2ZKD) use loops extending from a crescent formed from two beta sheets to flip C or m5C from undeformed B-form DNA into a pocket (top row), whereas McrB-N (3SSC; bottom row) uses loops from one beta-sheet to distort the DNA and flip the base. It resembles the human alkyladenine glycosylase (1BNK) (bottom row) in bending the DNA toward the major groove, while flipping the base via the minor groove. Figure 5 of Sukackaite *et al.* (60). (B) The SRA-like hemi-methylated 5mC recognition domains. A ribbon model of the N-terminal domain of the MspJI structure (4F0Q and 4F0P; left) compared with the SRA domain of UHRF1 (PDB 3FDE; right). The crescent shape formed

middle (Figure 5B). This behavior is recapitulated by the PvuRts1I group of enzymes (exemplified by AbaSDFI in Figure 5C), except that the distances are shorter and recognition of modification state is less well understood.

During the characterization of MspJI, Dcm (Cm5CWGG) sites were the first recognized substrate, yielding a clear banding pattern (51). Cleavage of differently modified plasmids and designed oligonucleotide substrates allowed a good assessment of both modification and sequence specificity. This family shows preference for particular bases nearby, similar to McrBC.

**MspJI DNA recognition is mediated by an SRA-like domain**

Recently, the crystal structure of MspJI without DNA has been resolved at 2.05 Å (80). Search of the Molecular Modeling Database at NCBI (101) using VAST (102)



**Figure 5.** Schematic diagrams of cleavage positions for MspJI and AbaSDFI. Cleavage of both strands is elicited by a singly modified site for both MspJI (A) and AbaSDFI (C). Cleavage position is fixed relative to the modified site, but with a four-base 5' extension for MspJI and a two-base 3' extension for AbaSDFI. When a site is symmetrically-modified (as for CpG sites in mammalian DNA), a 32 base-pair fragment is excised from the DNA (B). (A) Figure 2a and (B) Figure 3a reprinted with permission from Cohen-Karni *et al.* (52). (C) Figure 5a from Wang *et al.* (47).

**Figure 4.** Continued by interacting beta sheets and helices  $\alpha$ B and  $\alpha$ C are the conserved features of the SRA domain highlighted here. Loops on the concave side of UHRF1 participate in flipping the base, and similar loops presumably do so for MspJI. Two of these vary in length among family members and may play roles in sequence context specificity. Figure 2a and b from Horton *et al.* (80).



showed that the N-terminal domain was structurally similar to that of the eukaryotic SRA domain, with a crescent-shaped beta-sheet structure from which loops project (see Figure 4B and discussion earlier in the text, McrB). This structural homology allowed modeling of the DNA-bound structure, with a flipped m5C. The enzyme in the crystal is a tetramer, in which two monomers form a back-to-back dimer via the C-terminal regions that comprise the endonuclease. Two back-to-back dimers generate a tetrameric protein with two cleavage domains positioned (as in the Type IIP enzyme HindIII, used for modeling the C-terminal cleavage domain interaction with DNA) so that a 4-base 5' extension would be created on cleavage of modeled DNA. Cleavage is most efficient at molar ratios that allow all four SRA-like domains to be occupied—too much enzyme prevents cleavage from occurring.

### Tracking and dimerization

#### *McrBC as translocase*

Bourniquel and Bickle (84) have reviewed much of the enzymology of McrBC, which will be briefly summarized here. The Raleigh, Bickle and Pingoud laboratories have contributed to the following consistent picture of the *in vitro* reaction. EcoKMcrBC cleavage results in a double-strand cut near one RmC site (72–74) but requires cooperation of two sites (39,40) or a translocation block (73). The sites may be on different daughters across a fork (75). These are separated by 30–3000 bp (39,72,74) and may be on either strand (39,76); cleavage occurs ~30–35 bases from the modified base, with opposite nicks not tightly constrained (73), and minor cleavage clusters are found ~40, ~50 and ~60 nt from the m5C (74). hm5C DNA elicits cleavage also (39). A ring structure is formed by 5–7 molecules of McrB in the presence of GTP (Figure 3) (103); this complex can bind to a recognition element in DNA. In the presence of McrC, translocation of the complex occurs and cleavage ensues when translocation is blocked. Collision of translocating complexes, a protein barrier or a topological barrier will elicit double-strand cleavage adjacent to one recognition element or the other. The enzyme will cleave when recognition elements are on opposite sides of a forked structure (75). This would allow action *in vivo* to prevent entry of a MTase gene even with rare sites.

Structurally, the McrB protein is proposed to be a member of the AAA+ protein family of NTPases (104), many of which form ring-shaped complexes and participate in molecular machines. ‘Sensor’ segments found in these proteins have been shown in some cases to play roles in coupling NTPase activity to intersubunit communication and movement (105). Two of three elements of the GTP-binding motif proposed by Dila *et al.* (61) were validated by mutational analysis (65,62). The third proposed motif element was identified as amino acids NTAD by Dila *et al.* Alignment of AAA+ NTPases in (104) found this aligned with the motif designated Sensor-1 in (105). An interesting result was that mutations here unexpectedly appeared to abrogate interaction with McrC instead of changing which NTP would be

productive (62). It may instead play a role in coordinating GTP binding and hydrolysis with DNA binding, interaction with McrC and cleavage.

Intracellularly, the story becomes more complex, as the *mcrB* gene encodes two products of 51 and 33 kD, McrB-L and McrB-S, the latter one starting from an in-frame internal translation start site (106). Both *in vivo* and *in vitro*, McrB-S can interfere with the function of McrB-L, at least in part by forming complexes with McrC unable to bind DNA (107). Both species can form multimeric rings in the presence of GTP (103), as is usual for AAA+ NTPases (104).

#### *SauUSI requires two sites and ATP hydrolysis*

SauUSI was originally annotated as a putative helicase from *Staphylococcus aureus* sp. A single polypeptide is sufficient for activity both *in vitro* and also *in vivo* as a clone in *E. coli*, using modified phage  $\lambda$  as a challenge. The amino acid sequence contains a PLDc domain at the N-terminus. This contains a phosphodiesterase motif originally identified in Phospholipase D (108); it was validated by mutagenesis of four catalytic residue candidates. In the middle, ATPase and helicase motifs were proposed to account for ATP dependence of cleavage activity. A Domain of Unknown Function was identified at the C-terminus (Pfam DUF3427) (108) and was proposed to recognize the substrate (45).

The purified enzyme cleaves modified DNA containing m5C and hm5C but not m4C in the presence of ATP or dATP but not other nucleotides. The negative result for m4C is firm: plasmids modified at the same site by an m5C MTase (Dcm) or an m4C MTase (M.PspGI) were tested. The former (Cm5CWGG) was sensitive, whereas the latter was resistant. Thus, the sequence preference is likely to be satisfied. m6A is likely not a substrate, but few m6A-containing sites were examined.

Like McrBC, SauUSI requires the presence of two sites for efficient cleavage. Presumably, the ATPase activity participates in monitoring the presence of two sites, as for other nucleotide-dependent REases, including McrBC. The mechanism of communication is unknown. The enzyme belongs to a family of highly similar orthologues found in other sequenced *Staphylococci* (Tables 1 and 2), and more distant homologues can be found in sequenced bacterial and archaeal genomes.

## EVOLUTIONARY PRESSURES ON R-M SYSTEMS

### Evolution by selfish propagation

One way to understand the massive variety of restriction systems, and their sporadic distribution, is to locate the evolutionary drivers of enzyme diversification in the enzyme genes themselves, as selfish elements. Work from the Kobayashi laboratory has elaborated clear examples of selfish behavior in some Type II enzyme systems (109,110), in which the host becomes ‘addicted’ to the R-M system. Once a cell has acquired an R-M pair, loss of the genes results in death of that cell’s descendants, as the REase is frequently still present and able to act on the genome following loss of methylation activity. In this

perspective, the role played by modification-dependent enzymes is host defense, to exclude systems with 'foreign' MTase patterns, and prevent the cell from loading up with parasites. The exclusion event is accompanied by the death of the cell (111–113). Weak sequence specificity of Type IV enzymes could then result from the need to control entry of a wide variety of invading systems.

The selfish aspect certainly plays a role in R-M population biology, but cannot be the whole story. Type II R-M systems can still be lost, by inactivation of the R gene first. Moreover, Type I systems escape this scenario with complex control of cleavage activity: the restriction assembly includes a methylation assembly to begin with; therefore, the R protein cannot act unless the MTase is present; in addition, failure of the methylation activity in an intact complex leads to abrogation of R activity, sometimes by action of the ClpXP protease specifically on the R protein (114–116).

Furthermore, in population terms, a cell that acquired and became addicted to an R-M system should lose in competition with a sibling that never received the system. Two factors could counter this. First, acquisition could be accompanied by an increase in the total number of copies of the R-M system in the population, as proposed for invading transposable elements. This overreplication results in more copies of the system created than are lost, whether to suicide or to other selective disadvantage [see e.g. (117,118)]. R-M gene amplification within a cell has been reported experimentally (119) but spread in a population has not been demonstrated yet. A second factor that could counter the disability of addiction is localization of competition. In a structured environment (colonies on a plate or biofilm on large or small surfaces), killing of segregants preserves limiting nutrients for lineages that retain the toxin/antitoxin pair (120,121). Much of the real world is structured, so this is an important condition

### Evolution by phage-host arms race

A second perspective supposes that the modification-dependent Type IV enzymes arose from the competitive coevolutionary interaction between phages and their hosts. This was first enunciated by Revel and Luria (2) and most recently by Black and coworkers (122); see also (123). That is, hosts used modification-blocked restriction to defend against phage infection; T-even phages developed methods of substituting modified bases for the ordinary ones; hosts developed Type IV enzymes in defense; phages added sugar or other modifications (19) to thwart Type IV enzymes; hosts extended Type IV enzymes to accommodate these decorations; finally, phages developed protein inhibitors specific for these enzymes as well. T4 phages deliver a protein inhibitor (IPI\*) along with the DNA on infection, which allows growth in the presence of EcoCTGmrSD. The locus responsible for this inhibitor is highly variable among relatives of T4, as *gmrSD* is in enteric bacteria (both in distribution and in aa sequence). When phage with different IPI alleles were tested for protection from cloned EcoCTGmrSD and its homolog EcoUTGmrSD, specificity was evident: one or

the other or both or neither of the two homologs was counteracted in individual cases (122). This variability of the outcome supports the idea that phage-host interaction drives at least some of these developments.

In this perspective, the weak sequence selectivity of the Type IV systems might simply reflect the lack of endogenous targets for the enzyme. As the host does not present any hm5C and the phage is completely substituted, selection for sequence-specificity is weak. Selection would act to spare any co-resident MTases. This differs from Type II enzymes, where the MTase and REase must co-evolve to allow the host to survive. Each Type IV system is compatible with some suite of Type I-III MTases (and thus the R-M systems as a whole). Methylated or hydroxymethylated bases may not be recognized at all (EcoCTGmrSD), or the system may require one specific base in addition to the modified one (McrA, McrBC, MspJI and PvuRtsII). MTases modifying sites not including that base are then compatible, as Dcm (CmCWGG) is compatible with the McrA, McrBC and Mrr systems in *E. coli* K12.

Type IV systems that restrict methylated bases in a weakly specified sequence context confer an additional advantage in competition with phages. Many phages, such as  $\lambda$ , have not evolved the nucleotide-substitution strategy used by the T-even phages. These phages normally carry the modification pattern of the most recent host; if the last host expressed an MTase creating a susceptible site, the Type IV enzyme of the new host will destroy the invader and limit the infection. This may be accompanied by the death of the individual infected; therefore, protection can be conferred on the sibling population (111).

A further implication of this scenario considers the fate of a population invaded by phage. Phage survival of restriction occurs at biologically relevant frequencies ( $10^{-6}$ – $10^{-2}$ ). The survivors of restriction carry the particular methylation pattern of the particular cell and thus are resistant to all restriction systems it might have carried (Types I–IV). A bacterial population as a whole then benefits from mechanisms that diversify the suite of R-M systems so that such surviving phage do not have free access to the entire population. The extreme variability of R-M system content in isolates of the same species is compatible with this idea [see e.g. (124); REBASE Genomes <http://tools.neb.com/~vincze/genomes/>]. Such variability also limits and shapes interstrain gene transfer (115,125,126).

The Raleigh laboratory has built on elegant genetic work with Type I enzymes in the Murray laboratory (127–130) to investigate a locus designated the 'Immigration Control locus', or ICR, that exemplifies variable R-M content. Alternative DNA segments containing R-M systems are located at the same defined location in most *E. coli* chromosomes between the *yjiS* and *yjiA* genes. The ICR in the non-restricting *E. coli* C strain [used in the original definition of the R-M phenomenon (131)] contains a remnant of a Type I enzyme R gene and is 13 kilobase shorter than the same region in *E. coli* K12 (132). The mechanism of segment replacement is still unknown. The ICR would be an example of the 'defense islands' analyzed by the Koonin group (133). 'Defense

islands' contain genes that can defend against phage or other invading DNA; these exhibit bioinformatic properties similar to 'mobilome islands' containing mobilization genes (transposases for example). However, the mechanism of mobilization has not been identified for the ICR.

## FINAL THOUGHTS

The extreme diversity of R-M systems that recognize ordinary DNA seems likely to be approached by the diversity of Type IV restriction systems. Type IV enzymes are hard to find, as most detection methods depend on development of genetic systems for each taxon, or on serendipity. Those characterized so far mostly stem from initial genetic investigation of limits on infection, transformation or transduction. Barriers encountered provide leads to the genes involved. Bioinformatic analysis has helped to identify relatives, which may be more tractable to biochemical investigation than the example originally found. This approach has pitfalls: the gene encoding MspJI was first thought to code for an enzyme recognizing an unmodified site because it is immediately adjacent to an (inactive, it is now thought) cytosine MTase gene. Providentially, the first expression host was devoid of sensitive sites, whereas the first test substrate contained some (51). A combination of biological experiments with bioinformatics and biochemistry will be needed to reveal the full spectrum of Type IV enzymes that may lurk within the vast universe of unidentified ORFs in bacterial systems. One might begin with those strains whose genomes carry few Type II systems: *Bacillus*, or *Corynebacterium*, as opposed to *Helicobacter* or *Neisseria* [see the Genomes section of REBASE (50)].

The role of 'defense islands' and their relation to the 'mobilome' in bacterial population biology remains to be determined. If a defense island is similar to a mobilome island, there should be a mechanism of mobilization nearby, which would boost the contribution of 'overreplication' to the account of selections acting on R-M systems. R-M systems of all types can be found on or adjacent to known mobilizing elements (134,135), but have not been shown to move experimentally.

On another note, it may turn out that evolutionarily there is a continuum between the apparently modification-dependent and modification-blocked paths. One relative of McrBC predicted by bioinformatics analyses is LlaI, a system that recognizes an unmodified sequence and requires two MTases to support it (136). The enzyme BamHI prefers to cleave DNA with m6A in its GGATCC site, and mutants can be isolated that require this modified base (137). Are there native systems similarly protected by modification of one position in the recognition site but dependent on modification at a different one? An interesting evolutionary series can be imagined.

## ACKNOWLEDGEMENTS

The authors thank Rich Roberts and David Dryden for encouragement and critical comment. They are grateful to

Tom Bickle, Virginius Siksny, Yu Zheng and Xiadong Cheng for permission to reproduce figures. They also thank Ivan Correa for help drawing base structures, and Tasha José for illustration assistance. They also thank the anonymous reviewers of a previous manuscript for care taken with the review process, greatly improving the present document.

## FUNDING

New England Biolabs (E.A.R.). Funding for open access charge: New England Biolabs (to E.A.R.).

*Conflict of interest statement.* None declared.

## REFERENCES

- Luria,S.E. and Human,M.L. (1952) A nonhereditary, host-induced variation of bacterial viruses. *J. Bacteriol.*, **64**, 557–559.
- Revel,H.R. and Luria,S.E. (1970) DNA-glucosylation in T-even phage: genetic determination and role in phage-host interaction. *Annu. Rev. Genet.*, **4**, 177–192.
- Raleigh,E.A. and Wilson,G. (1986) Escherichia coli K-12 restricts DNA containing 5-methylcytosine. *Proc. Natl Acad. Sci. USA*, **83**, 9070–9074.
- Noyer-Weidner,M., Diaz,R. and Reiners,L. (1986) Cytosine-specific DNA modification interferes with plasmid establishment in Escherichia coli K12: involvement of rglB. *Mol. Gen. Genet.*, **205**, 469–475.
- Blumenthal,R.M., Gregory,S.A. and Cooperider,J.S. (1985) Cloning of a restriction-modification system from *Proteus vulgaris* and its use in analyzing a methylase-sensitive phenotype in *Escherichia coli*. *J. Bacteriol.*, **164**, 501–509.
- Sohail,A., Lieb,M., Dar,M. and Bhagwat,A.S. (1990) A gene required for very short patch repair in Escherichia coli is adjacent to the DNA cytosine methylase gene. *J. Bacteriol.*, **172**, 4214–4221.
- MacNeil,D.J. (1988) Characterization of a unique methyl-specific restriction system in *Streptomyces avermitilis*. *J. Bacteriol.*, **170**, 5607–5612.
- Vertes,A.A., Inui,M., Kobayashi,M., Kurusu,Y. and Yukawa,H. (1993) Presence of mrr- and mcr-like restriction systems in coryneform bacteria. *Res. Microbiol.*, **144**, 181–185.
- Macaluso,A. and Mettus,A.M. (1991) Efficient transformation of *Bacillus thuringiensis* requires nonmethylated plasmid DNA. *J. Bacteriol.*, **173**, 1353–1356.
- Holmes,M.L., Nuttall,S.D. and Dyal-Smith,M.L. (1991) Construction and use of halobacterial shuttle vectors and further studies on Haloferax DNA gyrase. *J. Bacteriol.*, **173**, 3807–3813.
- Pósfai,J., Bhagwat,A.S., Pósfai,G. and Roberts,R.J. (1989) Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.*, **17**, 2421–2435.
- Malone,T., Blumenthal,R.M. and Cheng,X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.*, **253**, 618–632.
- Timinskas,A., Butkus,V. and Janulaitis,A. (1995) Sequence motif characteristics for DNA [cytosine-N4] and DNA [adenine-N6] methyltransferases. Classification of all DNA methyltransferases. *Gene*, **157**, 3–11.
- Dickman,M.J., Ingleston,S.M., Sedelnikova,S.E., Rafferty,J.B., Lloyd,R.G., Grasby,J.A. and Hornby,D.P. (2002) The RuvABC resolvasome. *Eur. J. Biochem.*, **269**, 5492–5501.
- Kovall,R.A. and Matthews,B.W. (1998) Structural, functional, and evolutionary relationships between lambda-exonuclease and the type II restriction endonucleases. *Proc. Natl Acad. Sci. USA*, **95**, 7893–7897.
- Roberts,R.J., Chang,Y.C., Hu,Z., Rachlin,J.N., Anton,B.P., Pokrzywa,R.M., Choi,H.P., Faller,L.L., Guleria,J., Housman,G. *et al.* (2010) COMBEX: a project to accelerate the functional

- annotation of prokaryotic genomes. *Nucleic Acids Res.*, **39**, D11–D14.
17. Rae, P.M. and Steele, R.E. (1978) Modified bases in the DNAs of unicellular eukaryotes: an examination of distributions and possible roles, with emphasis on hydroxymethyluracil in dinoflagellates. *Biosystems*, **10**, 37–53.
  18. Warren, R.A.J. (1980) Modified bases in bacteriophage DNAs. *Annu. Rev. Microbiol.*, **34**, 137–158.
  19. Gommers-Ampt, J.H. and Borst, P. (1995) Hypermodified bases in DNA. *FASEB J.*, **9**, 1034–1042.
  20. Swinton, D., Hattman, S., Crain, P.F., Cheng, C.S., Smith, D.L. and McCloskey, J.A. (1983) Purification and characterization of the unusual deoxynucleoside, alpha-N-(9-beta-D-2'-deoxyribofuranosylpurin-6-yl)glycinamide, specified by the phage Mu modification function. *Proc. Natl Acad. Sci. USA*, **80**, 7400–7404.
  21. Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
  22. Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
  23. Borst, P. and Sabatini, R. (2008) Base J: discovery, biosynthesis, and possible functions. *Annu. Rev. Microbiol.*, **62**, 235–251.
  24. Kaminska, K.H. and Bujnicki, J.M. (2008) Bacteriophage Mu Mom protein responsible for DNA modification is a new member of the acyltransferase superfamily. *Cell Cycle*, **7**, 120–121.
  25. Iyer, L.M., Tahiliani, M., Rao, A. and Aravind, L. (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle*, **8**, 1698–1710.
  26. Zhou, X., He, X., Liang, J., Li, A., Xu, T., Kieser, T., Helmann, J.D. and Deng, Z. (2005) A novel DNA modification by sulphur. *Mol. Microbiol.*, **57**, 1428–1438.
  27. Ou, H.Y., He, X., Shao, Y., Tai, C., Rajakumar, K. and Deng, Z. (2009) dndDB: A Database Focused on Phosphorothioation of the DNA Backbone. *PLoS One*, **4**, e5132.
  28. Xu, T., Liang, J., Chen, S., Wang, L., He, X., You, D., Wang, Z., Li, A., Xu, Z., Zhou, X. et al. (2009) DNA phosphorothioation in *Streptomyces lividans*: mutational analysis of the dnd locus. *BMC. Microbiol.*, **9**, 41.
  29. Wang, L., Chen, S., Vergin, K.L., Giovannoni, S.J., Chan, S.W., DeMott, M.S., Taghizadeh, K., Cordero, O.X., Cutler, M., Timberlake, S. et al. (2011) DNA phosphorothioation is widespread and quantized in bacterial genomes. *Proc. Natl Acad. Sci. USA*, **108**, 2963–2968.
  30. Løbner-Olesen, A., Skovgaard, O. and Marinus, M.G. (2005) Dam methylation: coordinating cellular processes. *Curr. Opin. Microbiol.*, **8**, 154–160.
  31. Low, D.A. and Casadesu, J. (2008) Clocks and switches: bacterial gene regulation by DNA adenine methylation. *Curr. Opin. Microbiol.*, **11**, 106–112.
  32. Freitag, M. and Selker, E.U. (2005) Controlling DNA methylation: Many roads to one modification. *Curr. Opin. Genet. Dev.*, **15**, 191–199.
  33. Prohaska, S.J., Stadler, P.F. and Krakauer, D.C. (2010) Innovation in gene regulation: The case of chromatin computation. *J. Theor. Biol.*, **265**, 27–44.
  34. Seshasayee, A.S.N., Singh, P. and Krishna, S. (2012) Context-dependent conservation of DNA methyltransferases in bacteria. *Nucleic Acids Res.*, **40**, 7066–7073.
  35. Anton, B.P. (2011) *Dissertation*. Boston University, Boston, Massachusetts, USA.
  36. Raleigh, E.A., Trimarchi, R. and Revel, H. (1989) Genetic and physical mapping of the mcrA (rgIA) and mcrB (rgIB) loci of *Escherichia coli* K-12. *Genetics*, **122**, 279–296.
  37. González-Cerón, G., Miranda-Olivares, O.J. and Servín-González, L. (2009) Characterization of the methyl-specific restriction system of *Streptomyces coelicolor* A3(2) and of the role played by laterally acquired nucleases. *FEMS Microbiol Lett*, **301**, 35–43.
  38. Liu, G., Ou, H.Y., Wang, T., Li, L., Tan, H., Zhou, X., Rajakumar, K., Deng, Z. and He, X. (2010) Cleavage of phosphorothioated DNA and methylated DNA by the type IV restriction endonuclease ScoMcrA. *PLoS Genet.*, **6**, e1001253.
  39. Sutherland, E., Coe, L. and Raleigh, E.A. (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease. *J. Mol. Biol.*, **225**, 327–348.
  40. Kruger, T., Wild, C. and Noyer-Weidner, M. (1995) McrB: a prokaryotic protein specifically recognizing DNA containing modified cytosine residues. *EMBO J.*, **14**, 2661–2669.
  41. Sitaraman, R. and Leppla, S.H. (2012) Methylation-dependent DNA restriction in *Bacillus anthracis*. *Gene*, **494**, 44–50.
  42. Heitman, J. and Model, P. (1987) Site-specific methylases induce the SOS DNA repair response in *Escherichia coli*. *J. Bacteriol.*, **169**, 3243–3250.
  43. Waite-Rees, P.A., Keating, C.J., Moran, L.S., Slatko, B.E., Hornstra, L.J. and Benner, J.S. (1991) Characterisation and expression of the *Escherichia coli* Mrr restriction system. *J. Bacteriol.*, **173**, 5207–5219.
  44. Kerr, A.L., Jeon, Y.J., Svenson, C.J., Rogers, P.L. and Neilan, B.A. (2010) DNA restriction-modification systems in the ethanologen, *Zymomonas mobilis* ZM4. *Appl. Microbiol. Biotech.*, **89**, 761–769.
  45. Xu, S.Y., Corvaglia, A.R., Chan, S.H., Zheng, Y. and Linder, P. (2011) A type IV modification-dependent restriction enzyme SauUSI from *Staphylococcus aureus* subsp. aureus USA300. *Nucleic Acids Res.*, **39**, 5597–5610.
  46. Monk, I.R., Shah, I.M., Xu, M., Tan, M.W. and Foster, T.J. (2012) Transforming the untransformable: application of direct transformation to manipulate genetically *Staphylococcus aureus* and *Staphylococcus epidermidis*. *mBio*, **3**, e00277–00211.
  47. Wang, H., Guan, S., Quimby, A., Cohen-Karni, D., Pradhan, S., Wilson, G., Roberts, R.J., Zhu, Z. and Zheng, Y. (2011) Comparative characterization of the PvuRtsI family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. *Nucleic Acids Res.*, **39**, 9294–9305.
  48. Janosi, L., Yonemitsu, H., Hong, H. and Kaji, A. (1994) Molecular cloning and expression of a novel hydroxymethylcytosine-specific restriction enzyme (PvuRtsI) modulated by glucosylation of DNA. *J. Mol. Biol.*, **242**, 45–61.
  49. Bair, C.L. and Black, L.W. (2007) A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. *J. Mol. Biol.*, **366**, 768–778.
  50. Roberts, R.J., Vincze, T., Posfai, J. and Macelis, D. (2010) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **38**, D234–D236.
  51. Zheng, Y., Cohen-Karni, D., Xu, D., Chin, H.G., Wilson, G., Pradhan, S. and Roberts, R.J. (2010) A unique family of Mrr-like modification-dependent restriction endonucleases. *Nucleic Acids Res.*, **38**, 5527–5534.
  52. Cohen-Karni, D., Xu, D., Apone, L., Fomenkov, A., Sun, Z., Davis, P.J., Kinney, S.R.M., Yamada-Mabuchi, M., Xu, S.Y., Davis, T. et al. (2011) The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. *Proc. Natl Acad. Sci. USA*, **108**, 11040–11045.
  53. Lubys, A., Vitkute, J., Lubiene, J. and Janulaitis, A. (2011) Restriction enzymes and their applications. US Patent Application Publication 20110207139 A1, <http://appft.uspto.gov/netahtml/PTO/search-bool.html> (27 August 2013, date last accessed).
  54. Tarasova, G.V., Nayakshina, T.N. and Degtyarev, S.K. (2008) Substrate specificity of new methyl-directed DNA endonuclease GlAI. *BMC. Mol. Biol.*, **9**, 7.
  55. Mulligan, E.A. and Dunn, J.J. (2008) Cloning, purification and initial characterization of *E. coli* McrA, a putative 5-methylcytosine-specific nuclease. *Protein Expr. Purif.*, **62**, 98–103.
  56. Mulligan, E.A., Hatchwell, E., McCorkle, S.R. and Dunn, J.J. (2010) Differential binding of *Escherichia coli* McrA protein to DNA sequences that contain the dinucleotide m5CpG. *Nucleic Acids Res.*, **38**, 1997–2005.
  57. Anton, B.P. and Raleigh, E.A. (2004) Transposon-mediated linker insertion scanning mutagenesis of the *Escherichia coli* McrA endonuclease. *J. Bacteriol.*, **186**, 5699–5707.
  58. Bujnicki, J.M., Radlinska, M. and Rychlewski, L. (2000) Atomic model of the 5-methylcytosine-specific restriction enzyme McrA reveals an atypical zinc finger and structural similarity to betabetaalphaMe endonucleases. *Mol. Microbiol.*, **37**, 1280–1281.

59. Gast, F.U., Brinkmann, T., Pieper, U., Krüger, T., Noyer Weidner, M. and Pingoud, A. (1997) The recognition of methylated DNA by the GTP-dependent restriction endonuclease McrBC resides in the N-terminal domain of McrB. *Biol. Chem.*, **378**, 975–982.
60. Sukackaite, R., Grazulis, S., Tamulaitis, G. and Siksnys, V. (2012) The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. *Nucleic Acids Res.*, **40**, 7552–7562.
61. Dila, D., Sutherland, E., Moran, L., Slatko, B. and Raleigh, E.A. (1990) Genetic and sequence organization of the mcrBC locus of *Escherichia coli* K-12. *J. Bacteriol.*, **172**, 4888–4900.
62. Pieper, U., Brinkmann, T., Krüger, T., Noyer Weidner, M. and Pingoud, A. (1997) Characterization of the interaction between the restriction endonuclease McrBC from *E. coli* and its cofactor GTP. *J. Mol. Biol.*, **272**, 190–199.
63. Pieper, U., Schweitzer, T., Groll, D.H., Gast, F.U. and Pingoud, A. (1999) The GTP-binding domain of McrB: more than just a variation on a common theme? *J. Mol. Biol.*, **292**, 547–556.
64. Bujnicki, J.M. and Rychlewski, L. (2001) Grouping together highly diverged PD-(D/E)XK nucleases and identification of novel superfamily members using structure-guided alignment of sequence profiles. *J. Mol. Microbiol. Biotech.*, **3**, 69–72.
65. Pieper, U. and Pingoud, A. (2002) A mutational analysis of the PD...D/EXK motif suggests that McrC harbors the catalytic center for DNA cleavage by the GTP-dependent restriction enzyme McrBC from *Escherichia coli*. *Biochemistry*, **41**, 5236–5244.
66. Kelleher, J.E. and Raleigh, E.A. (1991) A novel activity in *Escherichia coli* K-12 that directs restriction of DNA modified at CG dinucleotides. *J. Bacteriol.*, **173**, 5220–5223.
67. Orlowski, J., Mebrhatu, M.T., Michiels, C.W., Bujnicki, J.M. and Aertsen, A. (2008) Mutational analysis and a structural model of methyl-directed restriction enzyme Mrr. *Biochem. Biophys. Res. Commun.*, **377**, 862–866.
68. Aravind, L., Makarova, K.S. and Koonin, E.V. (2000) SURVEY AND SUMMARY: holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res.*, **28**, 3417–3432.
69. Bujnicki, J.M. and Rychlewski, L. (2001) Identification of a PD-(D/E)XK-like domain with a novel configuration of the endonuclease active site in the methyl-directed restriction enzyme Mrr and its homologs. *Gene*, **267**, 183–191.
70. Steczkiewicz, K., Muszewska, A., Knizewski, L., Rychlewski, L. and Ginalski, K. (2012) Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic Acids Res.*, **40**, 7016–7045.
71. Bair, C.L., Rifat, D. and Black, L.W. (2007) Exclusion of glucosyl-hydroxymethylcytosine DNA containing bacteriophages is overcome by the injected protein inhibitor IPI. *J. Mol. Biol.*, **366**, 779–789.
72. Stewart, F.J. and Raleigh, E.A. (1998) Dependence of McrBC cleavage on distance between recognition elements. *Biol. Chem.*, **379**, 611–616.
73. Panne, D., Raleigh, E.A. and Bickle, T.A. (1999) The McrBC endonuclease translocates DNA in a reaction dependent on GTP hydrolysis. *J. Mol. Biol.*, **290**, 49–60.
74. Pieper, U., Groll, D.H., Wünsch, S., Gast, F.U., Speck, C., Mücke, N. and Pingoud, A. (2002) The GTP-dependent restriction enzyme McrBC from *Escherichia coli* forms high-molecular mass complexes with DNA and produces a cleavage pattern with a characteristic 10-base pair repeat. *Biochemistry*, **41**, 5245–5254.
75. Ishikawa, K., Handa, N., Sears, L., Raleigh, E.A. and Kobayashi, I. (2011) Cleavage of a model DNA replication fork by a methyl-specific endonuclease. *Nucleic Acids Res.*, **39**, 5489–5498.
76. Stewart, F.J., Panne, D., Bickle, T.A. and Raleigh, E.A. (2000) Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J. Mol. Biol.*, **298**, 611–622.
77. Lacks, S. and Greenberg, B. (1977) Complementary specificity of restriction endonucleases DpnI and DpnII with respect to DNA methylation. *J. Mol. Biol.*, **114**, 153–168.
78. Siwek, W., Czapinska, H., Bochtler, M., Bujnicki, J.M. and Skowronek, K. (2012) Crystal structure and mechanism of action of the N6-methyladenine-dependent type IIM restriction endonuclease R.DpnI. *Nucleic Acids Res.*, **40**, 7563–7572.
79. Lacks, S. and Greenberg, B. (1975) A deoxyribonuclease of *Diplococcus pneumoniae* specific for methylated DNA. *J. Biol. Chem.*, **250**, 4060–4066.
80. Horton, J.R., Mabuchi, M.Y., Cohen-Karni, D., Zhang, X., Griggs, R.M., Samaranayake, M., Roberts, R.J., Zheng, Y. and Cheng, X. (2012) Structure and cleavage activity of the tetrameric MspJI DNA modification-dependent restriction endonuclease. *Nucleic Acids Res.*, **40**, 9763–9773.
81. Cymerman, I.A., Obarska, A., Skowronek, K.J., Lubys, A. and Bujnicki, J.M. (2006) Identification of a new subfamily of HNH nucleases and experimental characterization of a representative member, HphI restriction endonuclease. *Proteins*, **65**, 867–876.
82. Gorbalenya, A.E. (1994) Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family. *Protein Sci.*, **3**, 1117–1120.
83. Chan, S.H., Opitz, L., Higgins, L., O'Loane, D. and Xu, S.Y. (2010) Cofactor requirement of HpyAV restriction endonuclease. *PLoS One*, **5**, e9071.
84. Bourniquel, A.A. and Bickle, T.A. (2002) Complex restriction enzymes: NTP-driven molecular motors. *Biochimie*, **84**, 1047–1059.
85. O'Driscoll, J., Heiter, D.F., Wilson, G.G., Fitzgerald, G.F., Roberts, R. and van Sinderen, D. (2006) A genetic dissection of the LlaJI restriction cassette reveals insights on a novel bacteriophage resistance system. *BMC Microbiol.*, **6**, 40.
86. Smith, R.M., Diffin, F.M., Savery, N.J., Josephsen, J. and Szczelkun, M.D. (2009) DNA cleavage and methylation specificity of the single polypeptide restriction-modification enzyme LlaGI. *Nucleic Acids Res.*, **37**, 7206–7218.
87. Interthal, H., Pouliot, J.J. and Champoux, J.J. (2001) The tyrosyl-DNA phosphodiesterase Tdp1 is a member of the phospholipase D superfamily. *Proc. Natl Acad. Sci. USA*, **98**, 12009–12014.
88. Sasnauskas, G., Halford, S.E. and Siksnys, V. (2003) How the BfiI restriction enzyme uses one active site to cut two DNA strands. *Proc. Natl Acad. Sci. USA*, **100**, 6410–6415.
89. Sasnauskas, G., Connolly, B.A., Halford, S.E. and Siksnys, V. (2007) Site-specific DNA transesterification catalyzed by a restriction enzyme. *Proc. Natl Acad. Sci. USA*, **104**, 2115–2120.
90. Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M.K., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J. et al. (2012) CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.*, **41**, D348–D352.
91. Kravetz, A.N., Zakharova, M.V., Beletskaya, I.V., Sineva, E.V., Denjuchametov, M.M., Petrov, S.I., Glatman, L.I. and Solonin, A.S. (1993) The cleavage sites and localization of genes encoding the restriction endonucleases Eco1831I and EcoHI. *Gene*, **129**, 153–154.
92. Klimauskas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–369.
93. Hashimoto, H., Vertino, P.M. and Cheng, X. (2010) Molecular coupling of DNA methylation and histone methylation. *Epigenomics*, **2**, 657–669.
94. Rajakumara, E., Law, J.A., Simanshu, D.K., Voigt, P., Johnson, L.M., Reinberg, D., Patel, D.J. and Jacobsen, S.E. (2011) A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation *in vivo*. *Genes Dev.*, **25**, 137–152.
95. Sharif, J. and Koseki, H. (2011) Recruitment of Dnmt1 roles of the SRA protein Np95 (Uhrf1) and other factors. *Prog. Mol. Biol. Transl. Sci.*, **101**, 289–310.
96. Tesfazgi Mebrhatu, M., Wywiał, E., Ghosh, A., Michiels, C.W., Lindner, A.B., Taddei, F., Bujnicki, J.M., Van Melderren, L. and Aertsen, A. (2011) Evidence for an evolutionary antagonism between Mrr and Type III modification systems. *Nucleic Acids Res.*, **39**, 5991–6001.
97. Grant, 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.

98. Aertsen, A. and Michiels, C.W. (2005) Mrr instigates the SOS response after high pressure stress in *Escherichia coli*. *Mol. Microbiol.*, **58**, 1381–1391.
99. Wei, H., Therrien, C., Blanchard, A., Guan, S. and Zhu, Z. (2008) The Fidelity Index provides a systematic quantitation of star activity of DNA restriction endonucleases. *Nucleic Acids Res.*, **36**, e50.
100. Lu, L., Patel, H. and Bissler, J.J. (2002) Optimizing DpnI digestion to detect replicated DNA. *Biotechniques*, **33**, 316–318.
101. Madej, T., Address, K.J., Fong, J.H., Geer, L.Y., Geer, R.C., Lanczycki, C.J., Liu, C., Lu, S., Marchler-Bauer, A., Panchenko, A.R. *et al.* (2011) MMDB: 3D structures and macromolecular interactions. *Nucleic Acids Res.*, **40**, D461–D464.
102. Gibrat, J.F., Madej, T. and Bryant, S.H. (1996) Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol.*, **6**, 377–385.
103. Panne, D., Müller, S.A., Wirtz, S., Engel, A. and Bickle, T.A. (2001) The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. *EMBO J.*, **20**, 3210–3217.
104. Neuwald, A.F., Aravind, L., Spouge, J.L. and Koonin, E.V. (1999) AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.*, **9**, 27–43.
105. Ogura, T. and Wilkinson, A.J. (2001) AAA+ superfamily ATPases: common structure–diverse function. *Genes Cells*, **6**, 575–597.
106. Beary, T.P., Braymer, H.D. and Achberger, E.C. (1997) Evidence of participation of Mcr(B) in McrBC restriction in *Escherichia coli* K-12. *J. Bacteriol.*, **179**, 7768–7775.
107. Panne, D., Raleigh, E.A. and Bickle, T.A. (1998) McrBs, a modulator peptide for McrBC activity. *EMBO J.*, **17**, 5477–5483.
108. Punta, M., Coghill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J. *et al.* (2011) The Pfam protein families database. *Nucleic Acids Res.*, **40**, D290–D301.
109. Naito, T., Kusano, K. and Kobayashi, I. (1995) Selfish behavior of restriction-modification systems. *Science*, **267**, 897–899.
110. Kobayashi, I. (2001) Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.*, **29**, 3742–3756.
111. Fukuda, E., Kaminska, K.H., Bujnicki, J.M. and Kobayashi, I. (2008) Cell death upon epigenetic genome methylation: a novel function of methyl-specific deoxyribonucleases. *Genome Biol.*, **9**, R163.
112. Ishikawa, K., Fukuda, E. and Kobayashi, I. (2010) Conflicts targeting epigenetic systems and their resolution by cell death: novel concepts for methyl-specific and other restriction systems. *DNA Res.*, **17**, 325–342.
113. Fukuyo, M., Sasaki, A. and Kobayashi, I. (2012) Success of a suicidal defense strategy against infection in a structured habitat. *Sci. Rep.*, **2**, 238.
114. Makovets, S., Titheradge, A.J. and Murray, N.E. (1998) ClpX and ClpP are essential for the efficient acquisition of genes specifying type IA and IB restriction systems. *Mol. Microbiol.*, **28**, 25–35.
115. Murray, N.E. (2002) Immigration control of DNA in bacteria: self versus non-self. *Microbiology*, **148**, 3–20.
116. Makovets, S., Powell, L.M., Titheradge, A.J.B., Blakely, G.W. and Murray, N.E. (2004) Is modification sufficient to protect a bacterial chromosome from a resident restriction endonuclease? *Mol. Microbiol.*, **51**, 135–147.
117. Orgel, L.E. and Crick, F.H.C. (1980) Selfish DNA: the ultimate parasite. *Nature*, **284**, 604–607.
118. Kidwell, M.G. and Lisch, D.R. (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution*, **55**, 1–24.
119. Sadykov, M., Asami, Y., Niki, H., Handa, N., Itaya, M., Tanokura, M. and Kobayashi, I. (2003) Multiplication of a restriction-modification gene complex. *Mol. Microbiol.*, **48**, 417–427.
120. Chao, L. and Levin, B.R. (1981) Structured habitats and the evolution of anticompetitor toxins in bacteria. *Proc. Natl Acad. Sci. USA*, **78**, 6324–6328.
121. Magnuson, R.D. (2007) Hypothetical functions of toxin-antitoxin systems. *J. Bacteriol.*, **189**, 6089–6092.
122. Rifat, D., Wright, N.T., Varney, K.M., Weber, D.J. and Black, L.W. (2008) Restriction endonuclease inhibitor IPI\* of bacteriophage T4: a novel structure for a dedicated target. *J. Mol. Biol.*, **375**, 720–734.
123. Labrie, S.J., Samson, J.E. and Moineau, S. (2010) Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.*, **8**, 317–327.
124. Lin, L.F., Posfai, J., Roberts, R.J. and Kong, H. (2001) Comparative genomics of the restriction-modification systems in *Helicobacter pylori*. *Proc. Natl Acad. Sci. USA*, **98**, 2740–2745.
125. Milkman, R., Raleigh, E.A., McKane, M., Cryderman, D., Bilodeau, P. and McWeeny, K. (1999) Molecular evolution of the *Escherichia coli* chromosome. V. Recombination patterns among strains of diverse origin. *Genetics*, **153**, 539–554.
126. Humbert, O., Dorer, M.S. and Salama, N.R. (2011) Characterization of *Helicobacter pylori* factors that control transformation frequency and integration length during inter-strain DNA recombination. *Mol. Microbiol.*, **79**, 387–401.
127. Barcus, V.A., Titheradge, A.J.B. and Murray, N.E. (1995) The diversity of alleles at the hsd locus in natural populations of *Escherichia coli*. *Genetics*, **140**, 1187–1197.
128. Sharp, P.M., Kelleher, J.E., Daniel, A.S., Cowan, G.M. and Murray, N.E. (1992) Roles of selection and recombination in the evolution of type I restriction-modification systems in enterobacteria. *Proc. Natl Acad. Sci. USA*, **89**, 9836–9840.
129. Loenen, W.A.M., Daniel, A.S., Braymer, H.D. and Murray, N.E. (1987) Organization and sequence of the hsd genes of *Escherichia coli* K-12. *J. Mol. Biol.*, **198**, 159–170.
130. Suri, B. and Bickle, T.A. (1985) EcoA: the first member of a new family of type I restriction modification systems. Gene organization and enzymatic activities. *J. Mol. Biol.*, **186**, 77–85.
131. Bertani, G. and Weigle, J.J. (1953) Host-controlled variation in bacterial viruses. *J. Bacteriol.*, **65**, 113–121.
132. Sibley, M.H. and Raleigh, E.A. (2004) Cassette-like variation of restriction enzyme genes in *Escherichia coli* C and relatives. *Nucleic Acids Res.*, **32**, 522–534.
133. Makarova, K.S., Wolf, Y.I., Snir, S. and Koonin, E.V. (2011) Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.*, **193**, 6039–6056.
134. Furuta, Y., Abe, K. and Kobayashi, I. (2010) Genome comparison and context analysis reveals putative mobile forms of restriction-modification systems and related rearrangements. *Nucleic Acids Res.*, **38**, 2428–2443.
135. Furuta, Y. and Kobayashi, I. (2012) Restriction-modification systems as mobile epigenetic elements. In: Roberts, A.P. and Mullany, P. (eds), *Bacterial Integrative Mobile Genetic Elements*. Landes Bioscience, pp. 85–103.
136. O'Sullivan, D.J. and Klaenhammer, T.R. (1998) Control of expression of LlaI restriction in *Lactococcus lactis*. *Mol. Microbiol.*, **27**, 1009–1020.
137. Whitaker, R.D., Dorner, L.F. and Schildkraut, I. (1999) A mutant of BamHI restriction endonuclease which requires N6-methyladenine for cleavage. *J. Mol. Biol.*, **285**, 1525–1536.