Review

Understanding the evolution of restriction-modification systems: Clues from sequence and structure comparisons✪

Janusz M. Bujnicki✉

Bioinformatics Laboratory, International Institute of Molecular and Cell Biology, ks. Trojdena 4, 02-109 Warszawa, Poland; BioInfoBank Institute, Limanowskiego 24A, 60-744 Poznań, Poland

Restriction-modification (RM) systems comprise two opposing enzymatic activities: a restriction endonuclease, that targets specific DNA sequences and performs endonucleolytic cleavage, and a modification methyltransferase that renders these sequences resistant to cleavage. Studies on molecular genetics and biochemistry of RM systems have been carried out over the past four decades, laying foundations for modern molecular biology and providing important models for mechanisms of highly specific protein–DNA interactions. Although the number of known, relevant sequences 3D structures of RM proteins is growing steadily, we do not fully understand their functional diversities from an evolutionary perspective and we are not yet able to engineer new sequence specificities based on rational approaches. Recent findings on the evolution of RM systems and on their structures and mechanisms of action have led to a picture in which conserved modules with defined function are shared between different RM proteins and other enzymes involved in nucleic acid biochemistry. On the other hand, it has been realized that some of the modules have been replaced in the evolution by unrelated domains exerting similar function. The aim of this review is to give a survey on the recent progress in the field of structural phylogeny of RM enzymes with special emphasis on studies of sequence–structure–function relationships and emerging potential applications in biotechnology.

Abbreviations: RM, restriction-modification; MTase, methyltransferase; ENase, endonuclease; M/HsdM, R/HsdR, S/HsdS, protein subunits responsible for: modification, restriction and sequence recognition; TRD, target recognition domain; m^{6}A, N^6-methyladenine; m^{4}C, N4-methylcytosine; m^{5}C, C5-methylcytosine; hm^{5}C, C5-hydroxymethylcytosine; AdoMet, S-adenosyl-L-methionine.

Restriction-modification (RM) systems occur exclusively in unicellular organisms and their viruses. They comprise opposing intracellular enzyme activities: DNA endodeoxyribonuclease (ENase), that recognizes and cleaves its target site, and a DNA methyltransferase (MTase), that transfers methyl group from *S*-adenosyl-L-methionine (AdoMet) onto specific nucleobases within the target, thereby protecting it from the action of the ENase. Methylation occurs either at adenine or cytosine, yielding N6-methyladenine (m^6A), N4-methylcytosine (m^4C) or C5-methylcytosine (m^5C). In symmetrical sequences, the same base is methylated on both strands. The methyl groups lie in the major groove of the DNA helix, in positions that do not interfere with base-pairing, but that change the "epigenetic" information content of DNA. For instance, methylation of only one strand of the target (hemimethylation) is usually sufficient to prevent cleavage by sterically hindering binding of the ENase to the target. This guarantees that after DNA replication hemimethylated daughter duplexes eventually become fully re-methylated rather than being cleaved [1].

RM systems were originally suggested to evolve as a defense mechanism against phage infection and other types of DNA invasion [2], and serve evolutionary purposes by producing gene-size fragments of foreign DNA to be integrated into the host chromosome *via* recombination [3]. There are also cases of seemingly quite typical RM proteins known, which are involved in quite sophisticated physiological processes, such as regulating competence for DNA uptake [4]. Moreover, some DNA repair systems in Eubacteria can be regarded as descendants of RM systems or *vice versa*. Presently there is a large body of evidence that many RM systems are highly mobile elements involved in various genome rearrangements, and that many of them exhibit "selfish" behavior, regardless of potential benefits for the host they may confer (reviewed in ref. [5]). With the abundance of literature it is beyond the scope of this review to fully cover all research articles on the biochemistry and genetics of RM systems; instead I will focus on the recent studies of their sequences and structures and rather recommend several excellent reviews that can provide a complementary viewpoint to that presented in this article [6–10].

CLASSIFICATION OF RM SYSTEMS

RM systems were subdivided into three basic types (I, II, and III) based on the number and organization of subunits, regulation of their expression, cofactor requirements, enzymatic mechanism, and sequence specificity [1]. However, further types and subtypes have been proposed as new, distinct RM systems have been discovered. The biochemical properties of the "novel" systems are intermediate to those of the "old" ones, and their most striking feature is that they seem to combine protein domains originated from the "old" systems in unprecedented structural contexts. Recently, a novel nomenclature for the "subtypes" has been proposed at the "DNA Enzymes: Structures & Mechanisms" conference in Bangalore (December 2000) [11].

Type I RM systems

Type I are the most complex systems — they comprise three subunits: S for sequence recognition, M for modification and R for restriction (reviewed in refs. [7, 8, 12]). The S and M subunits form a DNA:m^6A MTase (with a stoichiometry of M_2S_1), which recognizes and modifies DNA within the specific sequence and exhibiting a strong preference for hemimethylated DNA, which is quite unusual among prokaryotic MTases. The complex of all three subunits ($R_2M_2S_1$) becomes a potent restriction enzyme [13]. A schematic diagram showing the complex architecture of type I RM protein is presented in Fig. 1. If the type I ENase encounters unmodified target, it
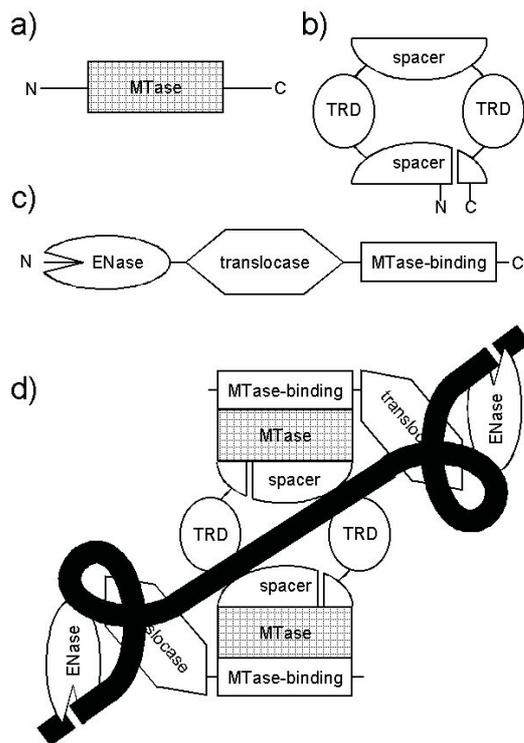
Figure 1. Schematic organization of typical type I RM enzymes, exemplified by *Eco*KI [12].

a) The M (HsdM) subunit comprising a single MTase module with N- and C- terminal extensions, **b)** the S (HsdS) subunit that exhibits circular pseudosymmetry, comprising variable TRDs and conserved spacer domains, **c)** the R (HsdR) subunit comprising modules implicated in DNA cleavage, DNA translocation and binding to the M$_2$S complex, **d)** proposed architecture of the M$_2$R$_2$S complex recognizing its bipartite target using two TRDs, generating DNA loops and cleaving DNA at a distance. For the sake of clarity, only one M$_2$R$_2$S complex is shown, although dimerization is necessary for DNA translocation and cleavage to occur [16], and the aspect of other possible interactions between the domains cleavage is ignored.

dimerizes rapidly [14] and initiates an ATP-dependent translocation of DNA towards itself simultaneously from both directions [15]. This process causes the extrusion or contraction of DNA loops and results in extensive supercoiling of DNA. Cleavage is elicited at variable distance from the recognition sequence once translocation stalls [16]. Since type I systems cleave DNA nonspecifically at

considerable distances from the unmethylated target sequences, they have so far failed to provide useful analytical reagents for modern molecular biology.

Type III RM systems

Type III systems were initially grouped together with type I systems as one family of ATP-dependent restriction enzymes [17]. However, once it was recognized that they comprised only two subunits (termed M or Mod for modification and R or Res for restriction), their recognition sites were only 5–6 bp long and not bipartite, and they cleaved at about 25 bp downstream of the recognition sequence, they were classified as a novel type [18]. Nevertheless, they are mechanistically similar to type I enzymes: the M subunit alone acts as a MTase and in a complex with the R subunit elicits ATP-dependent DNA translocation and cleavage [19]. A schematic diagram showing the domain architecture of type III RM proteins is shown in Fig. 2. AdoMet is required for methylation, but also for the efficient cleavage [20]. Type III ENases do not digest the substrate completely, leaving some fraction of sites always uncut. Another peculiarity of type III systems is that they methylate only one strand of the target, which leads to generation of unmethylated targets after each round of chromosome replication. However, it has been found that cleavage by type III enzymes requires two copies of the target sequence in a head to head orientation. In contrast only one sequence copy is needed for methylation to occur, which promotes re-methylation rather than degradation of the unmethylated strand [21]. It has been recently shown that type III enzymes exhibit R$_2$M$_2$ stoichiometry and that two such complexes cooperate in double stranded (ds) DNA cleavage on the 3′ side of either recognition site [22]. Interestingly, the top strand is cut by the ENase proximal to the cleavage site, while the bottom strand is cut by the distal ENase in the collision complex.
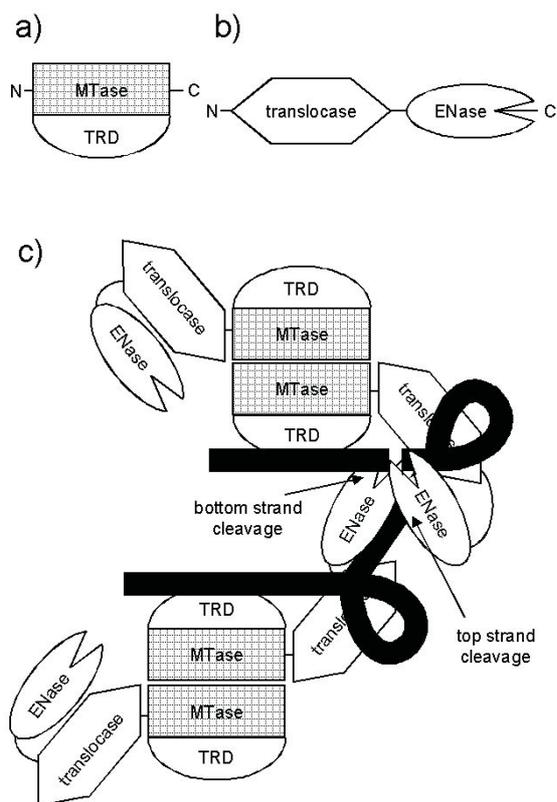
Figure 2. Schematic organization of typical type III RM enzymes, exemplified by *Eco*PI [22].

a) The M (Mod) subunit comprising a MTase module with the TRD localized within an insert, **b)** the R (Res) subunit comprising modules implicated in DNA cleavage and DNA translocation, **c)** proposed architecture of the M_2R_2 complex comprising two enzymes bound to sites in a head to head orientation. For the sake of clarity only one R and one M subunit in each complex interacts with the DNA and possible contacts between elements other than the ENase domains are ignored.

Type II RM systems

Type II systems are the simplest and most abundant of the RM systems, with MTase and ENase activity exerted by two distinct enzymes encoded by gene pairs. The archetypal ("orthodox") type II enzymes recognize short palindromic sequences 4 to 8 bp in length and methylate or cleave within or immediately adjacent to the recognition sequence, however numerous exceptions to that rule have been identified (see below). Type II ENases and

MTases have been intensively studied from the structure–function perspective — they are the only RM proteins for which crystal structures have been solved to date (September 2001: atomic coordinates for 12 ENases and 7 MTases are available; see Table 1).

Table 1. Structurally characterized ENases and MTases

ENase	PDB	MTase	PDB
*Bam*HI	2bam	M.*Dpn*M*	2dpm
*Bgl*I	1dmu	M.*Hae*III	2dct
*Bgl*II	1d2i	M.*Hha*I	5mht
*Bso*BI	1dc1	M.*Mbo*II*	1g60
*Cfr*10I	1cfr	M.*Pvu*II*	1boo
*Eco*RI	1ckq	M.*Rsr*I*	1eg2
*Eco*RV	1az0	M.*Taq*I	1g38
*Fok*I	1fok		
*Mun*I	1d02		
*Nae*I	1iaw		
*Ngo*MIV	1fiu		
*Pvu*II	1pvi		

The most representative entry from the Protein Data Bank (PDB) [23] (http://www.rcsb.org) has been chosen for each enzyme, with the preference for protein-DNA complexes and structures solved at possibly highest resolution. * indicates the enzymes for which protein-DNA cocrystal structures are not available.

Type II MTases (Fig. 3a) are the most diverse — though DNA: m^6A MTases are common to all major types of RM systems, so far all Prokaryotic m^4C and m^5C-generating enzymes were classified as *bona fide* type II, with only few exceptions among "solitary" enzymes believed to be very closely related to type II MTases. Type II m^5C MTases became a paradigm for nucleic acid enzymes that induce "flipping" of the target base into the catalytic pocket [24, 25]. They also served as a model for the studies on mechanism of AdoMet-dependent methylation of nucleic acids [26–28] and helped to understand the mode of action of different types of DNA MTases from Prokaryota [6] and Eukaryota [29]. They usually function as monomers that
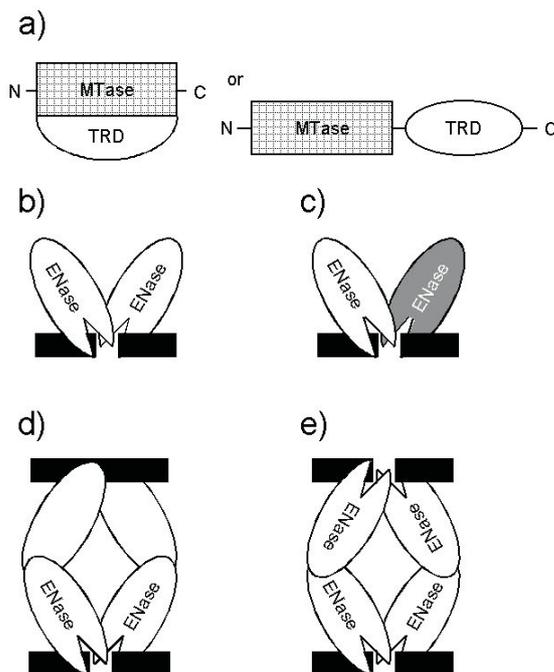
Figure 3. Type II RM enzymes

a) The "standalone" MTase comprising a MTase module with the TRD localized within an insert or fused to its C-terminus, **b)** the orthodox type II ENase homodimer, **c)** the type IIT heterodimer, **d)** the type IIE homodimer that uses two pairs of distinct domains for binding two identical sequences, **e)** the type IIF homotetramer that cleaves two sites in a concerted re-

catalyze methylation of the specific base in both strands of the palindromic target in two separate reactions.

Type II ENases, owing to their outstanding sequence specificity, became an indispensable tool in recombinant DNA technology, with applications in both basic science and molecular medicine. They have been also used as a model system for studying aspects of specific protein–DNA interactions and mechanisms of Mg^{2+}-dependent phosphodiester hydrolysis (which, ironically, have not yet been established for any RM enzyme) [10]. The orthodox type II ENases are homodimers (Fig. 3b) that cleave DNA in two strands producing a $5'$-phosphate and a $3'$-OH end; depending on orientation of the two subunits in respect to each other and to the recognized sequence, they can produce blunt ends like *Eco*RV [30] or *Pvu*II [31] or sticky ends with $5'$-overhangs

(like *Eco*RI [32] or *Bam*HI [33]), or $3'$-over-hangs (like *Bgl*I [34]). Over time, several subtypes of type II RM enzymes with distinct properties have been identified (shown in Figs. 3c–e, 4, 5 and 6).
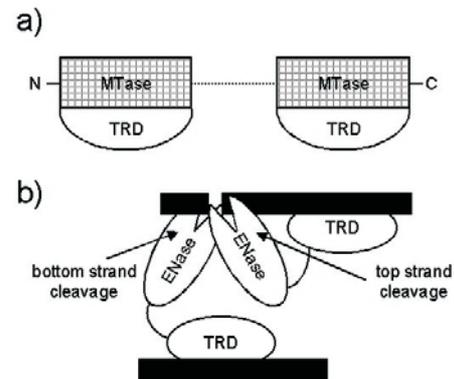

Figure 4. Type IIS RM enzymes.

a) the MTase component comprises two type II-like MTase domains fused within a single polypeptide or two separate enzymes (a dotted line shows the presence of a possible linker sequence) or a single MTase able to methylate different sequences on both strands of the target, **b)** the type IIS ENase homodimer bound to two targets [43] and generating a ds break at a fixed distance in respect to one of the sites (compare with Fig. 2c).

Type IIT restriction endonucleases are composed of two different subunits (Fig. 3c). For instance, *Bpu*10I is a heterodimer that recognizes an asymmetric sequence (it probably evolved from an orthodox homodimeric type II enzyme in which two subunits diverged, or is a hybrid of two related type II systems) [35]. On the other hand, *Bsl*I is a heterotetrameric enzyme ($\alpha_2\beta_2$) that recognizes a palindromic sequence [36].

Type IIE ENases, like *Eco*RII or *Nae*I are allosterically activated by binding of a second recognition sequence and therefore require two recognition sites for cleavage (Fig. 3d). They have two separate binding sites for the identical "target" and "effector" DNA sequences [37, 38].

Type IIF enzymes, like type IIE require binding of two identical sequences for cleavage, however they cleave them both in a con-

certed reaction (Fig. 3e). Those proteins characterized to date are tetrameric, for example *Ngo*MIV [39] and *Sfi*I [40].

Type IIS ENases cut at a fixed distance near their short, asymmetric target site [41]. This makes them similar to type III enzymes, but type IIS ENases do not require ATP or AdoMet or the presence of the MTase subunit for cleavage. They exist as monomers, with the DNA recognition and cleavage functions located on distinct domains (Fig. 4); however a dimerization of cleavage domains from two DNA-bound complexes is obligatory for ds DNA cleavage, as demonstrated for *Fok*I [42, 43]. Since the TRDs of type IIS ENases effectively interact with two sites, of which only one is cut is a single catalytic event, they can be regarded as a subclass of type IIE enzymes. Because of the unusual bipartite structure, type IIS ENases have proven particularly useful in creating chimeric enzymes by attaching the nonspecific cleavage domain to the DNA-binding domain of transcription factors [44–46].

The enzyme N.*Bst*NBI related to type IIS ENases has been characterized as a "nicking" ENase, which cleaves only on the top strand 4 bp away from its recognition sequence [47]. Interestingly, it has been shown that its close homologs, *Mly*I and *Ple*I introduce nicks prior to ds cleavage, which presumably occurs only after the ENases dimerize [48]. Hence, it has been suggested that the peculiar limited bottom strand cleavage activity of N.*Bst*NBI results from the inability of its cleavage domain to dimerize. These results suggest that type IIS enzymes exert ds DNA cleavage in a similar manner to type III enzymes, i.e. the top strand is cut by the ENase bound to the target sequence proximal to the cleavage site, while the bottom strand is cut by the distal ENase (Figs. 3, 4).

Type IIS MTases must methylate an asymmetric target, hence this kind of RM systems comprises two MTases specific for each strand, which may methylate different bases, like adenine (GGTG**A**) and cytosine (TCAC**C**)

in the case of *Ngo*BVIII [49], or one fusion protein with two MTase domains with distinct specificities, like in the case of *Fok*I (GG**A**TG and C**A**TCC), [50]. Another possibility is to employ a MTase, which recognizes a degenerated sequence and is able to methylate both strands, like it has been suggested for G**A**STC-specific (S=G or C) M.*Bst*NBI (unpublished data cited in ref. [48]) or for the hypothetical SS**A**TSS-specific ancestor of the C-terminal MTase domain of M.*Fok*I [50].

Type IIG (formerly type IV) RM systems are composed of two MTases, of which one modifies both strands of the asymmetric substrate, while the other modifies only one strand, but in addition exhibits also the ENase activity (Fig. 5), cutting the target 16/14 bp in
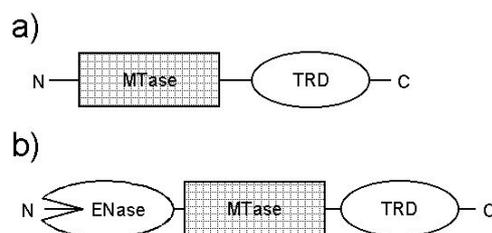

Figure 5. Schematic organization of type IIG RM enzymes.

a) the type II-like MTase, **b)** the ENase/MTase subunit, whose mechanism of interaction with the target or the possible multimerization mode is unknown, but may be related to that of type III and type IIS ENases (Figs. 2c and 4b)

3′ direction from the recognition site [51]. Some type IIG enzymes exhibit peculiar biochemical properties that make them similar to type III enzymes (see below): For instance *Eco*57I cleaves the substrate only partially and is stimulated by AdoMet [51], while for *Bse*MII AdoMet is essential for cleavage [52]. On the other hand, cleavage at a fixed distance from the target resembles both type IIS and type III enzymes. Hence, type IIG enzymes were suggested to be the evolutionary link between type III and type IIS systems, however this hypothesis has never been supported by a genuine phylogenetic study [53].

Type IIB (formerly type V or "*Bcg*I-like") RM systems encode both ENase and MTase activities within one polypeptide chain, similarly to the type IIG bifunctional ENase/MTase, but with the ability to modify both strands of the symmetric, bipartite target sequence [54]. The pattern of cleavage, which makes them distinct from other types, results from unprecedented combination of previously known features: all type IIB enzymes cleave DNA on both sides of their binding site (like type I ENases) at a fixed distance (like type IIs, IIG and III), resulting in excision of a short DNA fragment (Fig. 6). Some
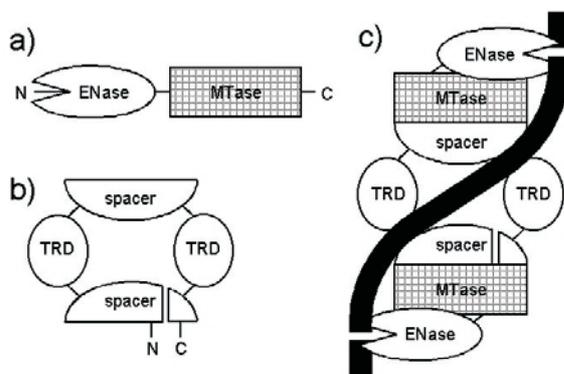

Figure 6. Schematic organization of an archetypal type IIB RM enzyme *Bcg*I [54].

a) The ENase/MTase subunit, **b)** the S subunit, **c)** proposed architecture of the (MR)$_2$S complex of the *Bcg*I RM system that cleaves DNA at a limited distance at both sides of its bipartite type-I like target (compare with Fig. 1d). The aspect of dimerization required for the bilateral cleavage is ignored for clarity and because it is unclear if and how the four ENase domains of the [(MR)$_2$S]$_2$ complex cooperate during the cleavage.

of them, like *Bcg*I, require a separate subunit (S) to bind to DNA and recognize the target, but others, like *Cje*I [55] and *Hae*IV [56] seem to exert all three functions with one chain. The S subunit of the *Bcg*I RM system is related to the type I S subunits, while in *Cje*I the S subunit is fused to the C-terminus of the ENase/MTase subunit. In the *Hae*IV RM system, no region homologous to the typical S subunits has been identified to date, but it is

likely that its TRD maps to the C-terminus [56].

Generally, many type IIB enzymes exhibit various peculiarities, which may be or may be not specific to other proteins of this class. For instance, *Hae*IV was shown to release an asymmetric fragment after cleavage [56] and *Bcg*I requires two bipartite target sites for cleavage [57] similarly to the enzymes of types I, IIE, IIF, and III. It is tempting to speculate that type IIB enzymes are a compact variant of type I enzymes that lack the DNA translocase module, but may show the same mechanism of DNA binding and cleavage on both sides of the target (compare Figs. 1 and 6).

To my best knowledge, interactions between a pair of the ENase domains, each cleaving one strand of the double strand target, has been shown only for the orthodox type II and related "standalone" ENases (types IIT, IIE and IIF) and for the ENase modules of type IIS and type III RM enzymes. It is tempting to speculate that other RM enzymes, including type I, type IIG and type IIB ENases also require a dimer of ENase domains to exert cleavage as opposed to a single domain that would introduce two nicks in both strands of the target, thereby making a ds break. If this hypothesis is corroborated by experiment, it would be interesting to learn if in those complex enzymes that possess two ENase domains, the catalytically competent dimers are formed *in cis* (i.e. by the ENase domains of a multiprotein complex bound to the same target) or *in trans* (i.e. by the ENase domains that belong to different proteins, as in the case of type IIS enzymes). Remarkably, different *in trans* configurations can be envisaged for proteins with more than two ENase domains in the catalytic unit [22].

Some type II RM enzymes recognize lengthy, discontinuous sites, such as *Sfi*I (GGCCNNNNNGGCC), *Bgl*I (GCCNNNNNGC) or *Xcm*I (CCANNNNNNNNNTGG), but most likely they acquired this functional peculiarity independently in the evolution [58] and they have not been classified as a separate

type or subtype. There have been several excellent reviews articles in the last decade focusing on various aspects of type II ENases [9, 10, 59–61], however only recently experimental and computational studies on their sequences and structures provided new data and interpretations, considerably broadening our view on these enzymes and their relationship to other protein families (see the paragraph devoted to the ENase domain within the subsequent section of this paper).

RM systems of other types

There are also some RM systems that do not fit into any of these classes — they likely represent genuine hybrids of "regular" types, which arose by fusions of their separated components, but so far no robust phylogenetic study has been undertaken to infer the pathways of their evolution. For example it has been also suggested that type II ENases may couple with type I MTases with a cognate sequence specificity, giving rise to the chimerical "type I&$^1/_2$" systems (G.G. Wilson, cited as personal communication in ref. [6]). On the other hand, the *Lla*I system consists of four proteins, one of which is a fusion of two type II-like m^{6}A MTases, a typical IIS MTase similar to *Fok*I (see above) [62] and the other three are remotely related to the McrBC nuclease (see below). There are also RM systems comprised of multiple ENases and MTases; in several such cases, like *Dpn*II [63] or *Bcn*I [64], one of the two MTases of the same specificity may also methylate single stranded DNA.

Solitary ENases

Paradoxically, the first restriction enzymes described were McrA (RglA) and McrBC (RglB) from *E. coli,* which do not form a part of a RM system since they do not associate functionally with any particular MTase and their ENase activity is not inhibited by methylation of the target. Conversely, they specifically recognize and cleave sequences containing methylated or hydroxymethylated cytosine (m^{4}C, m^{5}C or hm^{5}C, respectively), unless it is glucosylated as in wild type T-even coliphages [65, 66]. Together with the *E. coli* Mrr enzyme, which targets modified adenine or cytosine in a poorly defined sequence context [67] and *Streptococcus pneumoniae Dpn*I ENases [68] they make up a separate type of modification-directed restriction (MDR) enzymes. Another unusual enzyme of this class is *Pvu*Rts1I, which restricts DNA containing hm^{5}C, even when it is glucosylated. A MTase-like gene has been found near *Pvu*Rts1I, but neither its activity as a modification MTase nor influence on the *Pvu*Rts1I-mediated restriction could be demonstrated [69]. The MDR enzymes can be thought of as free-standing predecessors of RM system components or as nucleases that abandoned RM systems (for instance following the "death" of their cognate MTase) to become "ENases on the loose". Alternatively, the MDR systems may be seen as products of the "arms race" between bacteria developing new defensive weapons against T-even phages and the viruses protecting their DNA using increasingly more complex modifications (reviewed in ref. [70]).

Another class of sequence-specific nucleases, whose relationships with restriction enzymes were not known until very recently, are the so called "homing" ENases (reviewed in refs. [71, 72]). A large number of these enzymes has been identified in Eukaryotic nuclear and organellar genes, but there are also a few, which have been found in Prokaryota and their phages. They function in dissemination of certain mobile introns and inteins by cleavage of long, asymmetric, and degenerate sequences. Creation of recombinogenic ends promotes gene conversion, which leads to duplication of the intron. Homing ENases and some freestanding intergenic ENases, which share functional properties and sequence similarities, can be grouped into three families of presumably independent evolutionary origin (LAGLIDADG, HNH, and GIY-YIG) [73]. In

this review I will refer only to the structural data on members of HNH and GIY-YIG families, which are relevant to the evolutionary studies on genuine restriction enzymes.

Solitary MTases

Another group of enzymes related to RM enzymes are DNA MTases not associated with restriction enzymes. They are generally thought to be involved in gene regulation, chromosome replication, and DNA repair, though only few enzymes of this category are characterized in enough detail to justify unequivocal definition of their physiological function. The best studied examples is the GATC-specific Dam (DNA m^{6}A MTase) of *E. coli* and related γ-Proteobacteria, which has been implicated in numerous regulatory processes including control of expression of virulence determinants, and in methyl-directed mismatch repair (reviewed in ref. [74]). The mismatch-specific MutHSL excision apparatus uses Dam methylation to distinguish between the parental and daughter strands after chromosome replication. Nevertheless, Dam is not essential for viability [75]. The GANTC-specific m^{6}A MTase CcrM is an essential enzyme involved in cell-cycle control of Caulobacter [76]. Another well-studied "solitary" MTase is the CCWGG-specific Dcm (DNA m^{5}C MTase) of *E. coli*, whose function however still remains a mystery [77]. Mismatches resulting from spontaneous deamination of m^{5}C to U are repaired by the so called very short patch (VSP) system, which includes the C(T:G or U:G mismatch)WGG-specific single-strand nicking ENase Vsr [78]. Interestingly, both the Dam-associated nicking ENase MutH and the Dcm-associated Vsr are evolutionarily related to genuine restriction enzymes [79, 80].

Other MTases not associated with *bona fide* restriction enzymes are specified by viral genomes or conjugative plasmids, and serve to self-protect the invasive DNA from restriction endonucleases when it enters a new host.

Some phages carry MTases with Dam-like specificity, but it is unclear whether they have regulatory functions or serve to counteract restriction enzymes with cognate specificities [6]. An intriguing group of "antirestriction" MTases has been identified in several *Bacillus subtilis* phages — these enzymes can each recognize and m^{5}C-methylate several different targets, which are also targets for RM systems of the host. Based on the analysis of the multispecific MTases carried out by Trautner's group a modular model of MTase organization has been proposed, in which specificity of the core enzyme was achieved by a combination with a variety of sequence-specific modules [81, 82].

STRUCTURAL AND FUCTIONAL DOMAINS OF RM SYSTEMS

Dryden [6] suggested that the MTase composed of the target-recognizing domain (TRD; see next section), catalytic subdomain and AdoMet-binding subdomain can be thought of as the structural core of a typical RM system. In this respect, the RM system is made up by association of the MTase with a DNA cleavage (ENase) module and in some cases a DNA translocase module. Thus, all polypeptide subunits either exert their activity in a protein complex containing MTase, which interacts with the target DNA sequence *via* its TRD, or they have functional autonomy owing to a separate TRD analog. For instance the ENase module can exist as a separate protein comprising one or more structural domains (type II systems, Figs. 3b–e, 4), or as a fusion with the DNA translocase module (type I and III, Figs. 1, 2) or with the MTase module (type IIG and IIB, Figs. 5, 6). The orthodox type II ENases developed their own target-recognizing elements, functioning either as a clearly distinguishable TRD or an ensemble of loops protruding from the catalytic interface. On the other hand, the multifunctional R subunits of type I and type III RM systems exert their function of DNA translocase/ENase

only when complexed with the MTase. In type I R subunits a special domain responsible for establishing protein–protein contacts has been identified in the C-terminus [12] (Fig. 1); to my knowledge, such domain has not been delineated to date in primary structures of type III R subunits. The apparent modular architecture of all enzyme types suggested that shuffling of a quite limited repertoire of modules and domains conferring particular functions is the main force driving their functional diversification (Figs. 1–6).

The target recognition domain (TRD)

Target recognition domains have been operationally defined as regions responsible for sequence-specific binding of RM proteins to the target DNA. They have been initially (and most clearly) defined for mono- and multi-specific m^5C MTases [81] and the S subunits of type I RM systems [83], in which they are long, variable sequences, surrounded by well conserved motifs. In the multi-specific m^5C MTases from several bacteriophages of *Bacillus subtilis*, certain mutations in the variable region can abolish one target specificity while leaving the others intact. By mapping the mutations and studying the specificity of chimeric proteins, Trautner and coworkers determined that each target sequence is recognized by its own TRD and defined its minimal size as approximately 40 amino acids. Nevertheless, they failed to generate enzymes with novel specificities by shuffling of gene fragments except for instances where entire TRDs were exchanged [81, 84–87]. TRD swapping has also been successfully applied to alter the DNA sequence specificity of mono-specific m^5C MTases from Bacteria and Eukaryota [88, 89], in agreement with the conclusion of a recent phylogenetic study focused on the m^5C MTase family ([90], J.M. Bujnicki, unpublished).

In type I RM systems, which recognize two short defined regions separated by a non-specific spacer of fixed length, each of these re-

gions is recognized by an independent TRD (reviewed in ref. [91]). Most of the S subunits carry two separable TRDs, each approximately 150 aa in length, within a single polypeptide. It has been proposed that the TRDs and the "conserved" domains in the S subunits have a circular organization (Fig. 1) providing the symmetry for their interaction with the other subunits and with the bipartite, asymmetric DNA target [92]. However, a naturally or artificially truncated S subunit comprising a single TRD and a set of conserved motifs can function as a dimer, specifying the bipartite, symmetric DNA target, suggesting that the present day S subunits are the result of a gene duplication [93]. The conserved regions can be thought of as a scaffold upon which TRDs are mounted, allowing them to be swapped among type I RM systems to generate new specificities. Indeed, natural combinatorial variation of the S subunits and the half-subunits in certain type I RM systems have been reported [91, 94–96].

By analogy, the large variable regions found in most m^4C and m^6A MTases were also predicted to function as TRDs [97]. X-Ray crystallographic studies of the m^5C MTases M.*Hha*I [98] and M.*Hae*III [99], m^6A MTases M.*Taq*I [100], M.*Dpn*M [101] and M.*Rsr*I [102], and m^4C MTase M.*Pvu*II [28] demonstrated that the TRDs of all these proteins (excepting the pair of m^5C MTases) are structurally dissimilar (Fig. 7). It is not clear if these similar TRDs result from independent gene fusion events or evolutionary convergence. Based on structure prediction and random mutagenesis, Dryden and coworkers suggested that the TRDs of type I enzymes may be similar to the TRDs of m^5C MTases [103, 104]. Nevertheless, it is unclear to what degree the "alternative" TRDs are conserved in individual MTase subfamilies and if there are novel types of TRD yet to be discovered. For instance, sequence analysis demonstrated that certain mono-specific MTases possess several variable regions, which may share the function of a spatially-discontinuous TRD [97, 105]. Some
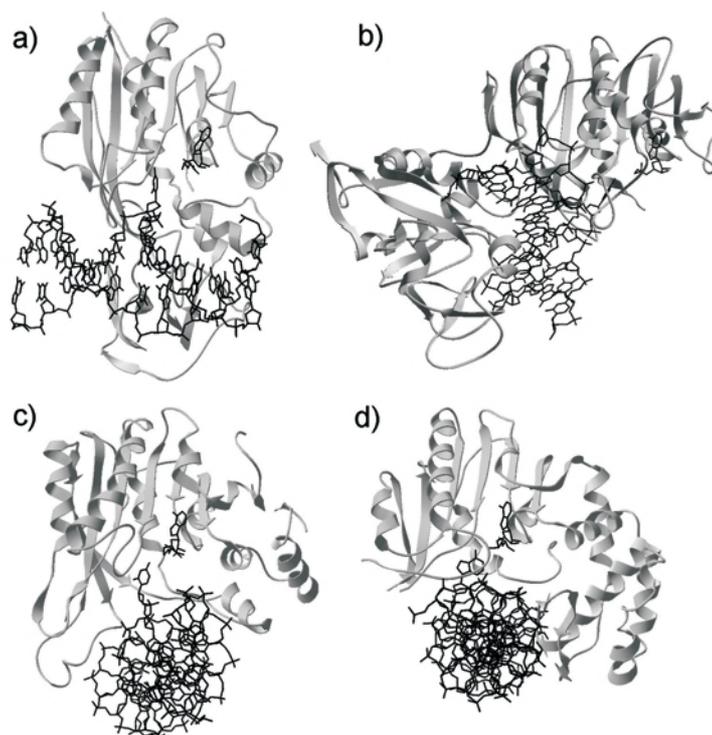
Figure 7. Cartoon diagrams of four structurally characterized DNA MTases depicting similarities between their catalytic domains and differences between their TRDs.

The core of the consensus MTase fold, recognizable by the 7-stranded β-sheet, is in the same relative orientation in all four images. **a)** The m^5C MTase M.*Hha*I co-crystalized with its target DNA (PDB coordinate file 5mht [109]), the TRD is "behind" the DNA, **b)** the γ-m^6A MTase M.*Taq*I (1g38 [110]) co-crystalized with its target DNA, the C-terminal TRD is on the left hand side, **c)** the α-m^6A MTase M.*Dpn*M (2dpm [101]) manually docked to its target, the TRD (localized within an insert in the catalytic domain) is on the right hand side, **d)** the β-m^4C MTase M.*Pvu*II (1boo [28]) manually docked to its target DNA, the proposed TRD (localized within an insert in the catalytic domain that maps to the upper left hand side of the image) is disordered in the crystal of the DNA-free form and therefore not shown.

small MTases seem TRD-less, and it has been suggested that their specificity determinants reside within the short loops protruding from the catalytic face of the catalytic domain [106, 107]. Moreover, even the typical TRD-containing enzyme M.*Eco*RV (and presumably its numerous homologs) has recruited residues from at least two loops in the catalytic domain to make specific protein–DNA contacts [108]. In addition, it is not known how the series of TRDs are arranged in the multispecific m^5C MTases, or how these complex enzymes interact with their multiple targets.

ENases also have to achieve sequence specificity. In the type I systems, the ENase specificity is provided by the same S subunit that is used by the MTase. Type II ENases, which interact with their DNA targets independently from their cognate MTases, may recognize target sequences using either an autonomous TRD fused to the catalytic domain, an ensemble of elongated loops projected from the catalytic domain or combination of both (reviewed in ref. [10]). Generally, the first strategy is characteristic for type IIS enzymes that cleave at a distance and the latter two strategies for most other type II enzymes. For instance, X-ray crystallography demonstrated that type IIS *Fok*I endonuclease comprises a non-specific cleavage domain and a large, compact TRD composed of three subdomains resembling helix-turn-helix domains [111, 112]. Similar bipartite architecture, albeit comprising structurally dissimilar TRDs and catalytic domains, has been predicted from computational sequence analysis for the type IIS en-

zymes *Bfi*I [113] and *Mbo*II [114], and for homing nucleases from the GIY-YIG super-family [115]. It should be stressed that identification of potential TRDs in sequences of restriction enzymes is particularly difficult, since unlike in MTases the catalytic domains of ENases contain no obviously conserved sequence motifs, which renders the simplistic criterion of sequence variability inadequate. Moreover, the key functions of type II restriction enzymes, i.e. multimerization, sequence-specific DNA binding and cleavage are interwoven such that some regions and residues are crucial for more than one aspect of the ENase function [10].

The MTase domain

  The MTase domain, which transfers the methyl group from AdoMet onto the target base, is the only truly conserved domain among RM systems; that is, representatives of only one of several unrelated protein families known to catalyze this kind of reaction have been identified in the context of RM systems (reviewed in ref. [116]). Other enzymes, which generate different modifications to inhibit restriction, are evolutionarily unrelated and structurally dissimilar, including the only enzyme that generates a chemically similar product, the tetrahydrofolate-dependent cytosine-C5 hydroxymetyltransferase of T-even coliphages [117]. The conserved "MTase fold" is characterized by an α/β domain with a central seven-stranded β-sheet sandwiched between two layers of α-helices (Figs. 7, 8a). It strongly resembles the architecture of the duplicated Rossmann-fold, with the only exception of a characteristic β-hairpin, involving strands 6 and 7, which is absent from Rossmann-fold proteins [118]. All DNA MTase structures exhibit very similar fold, with only minor variations of orientation and number of peripheral secondary structural elements. The approximate two-fold pseudo symmetry reflects the structural similarity of the AdoMet binding site to the target nucleo-
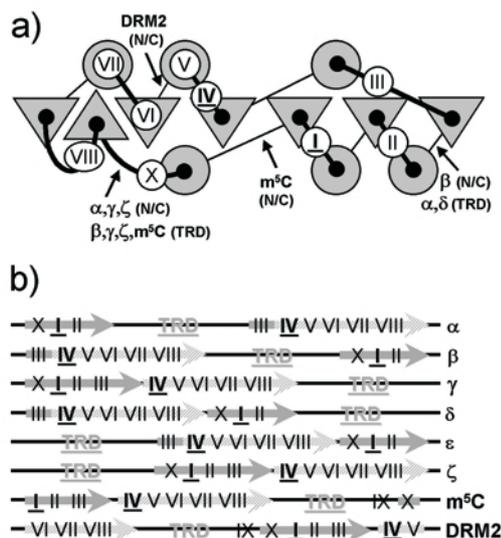

Figure 8. Conserved fold and variable topology of the common MTase domain.

a) The "circularized" topology diagram with triangles representing β-strands, circles representing α- and 3_{10}-helices, and connecting lines representing loops; the thick lines correspond to the loops at the catalytic face of the protein that harbor residues that take part in binding and catalysis. Circled Roman numerals represent nine motifs, the key motifs I and IV shown in bold and underlined. Arrows show the topological breakpoints (N/C for generation of N- and C-termini) and sites of TRD insertion characteristic for the individual classes of MTases. **b)** The linear organization of six classes of amino-MTases (-) postulated in ref. [97] and m^{5}C MTases (the prevailing archetypal topology labeled as m^{5}C, and the two underrepresented classes and DRM2). The AdoMet-binding region is shown as a solid arrow, the catalytic region is shown as a striped arrow. Conserved motifs are labeled accordingly.

tide-binding active site. This observation has led to the suggestion that the ancestral MTase arose after gene duplication converted an AdoMet-binding protein into a protein that bound two molecules of AdoMet and that the two halves then diverged [119]. An alternative hypothesis has been put forward that various MTases could have originated independently from Rossmann-fold proteins [101]. Supporting this view, a subsequent phylogenetic study using both atomic coordinates and corresponding amino-acid sequences suggested that MTases exhibiting the "typical fold" origi-

nated from one common Rossmann-fold ancestor [118].

Based on the methylated nucleotide that is generated, DNA MTases can be divided into three different groups: m^6A, m^4C, and m^5C MTases. m^6A and m^4C MTases methylate the exocyclic amino group of the nucleobase and are collectively termed "amino-MTases", while m^5C MTases methylate the C-5 atom of cytosine. It has been suggested that m^4C and m^6A MTases are more closely related to each other than to m^5C MTases [97]. Remarkably, certain m^6A MTases display cryptic m^4C activity on mismatched cytosines [120] and some m^4C MTases may methylate mismatched adenine [121]. Moreover, experimental and bioinformatics studies suggested that m^4C-specific enzymes may have evolved independently multiple times from m^6A MTases, although no consensus has been reached regarding the evolutionary pathways leading to the present-day distribution of specificities [105, 106, 120, 122]. Recently, it has been shown that a change of the target base specificity from m^6A to m^4C is possible with only a few amino acid substitutions. In an elegant experiment Roth and Jeltsch reduced the size of the target base binding pocket of M.*Eco*RV by site-directed mutagenesis, generating an enzyme variant that no longer methylated adenine and whose activity towards mismatched cytosine was reduced only 17-fold [108, 123]. Nevertheless, such variant was not able to methylate cytosine if it was base-paired with guanine, suggesting that additional mutations are needed to change the base flipping mechanism of amino-MTase.

Amino-acid sequence alignments of MTases revealed 9 relatively weakly conserved motifs and a variable region, localized differently in distinct families [124, 125] (Fig. 8b). Based on the results of X-ray crystallography of m^5C MTase *Hha*I [98] and on structure-based multiple sequence alignment, motifs IV–VIII were assigned to the active-site subdomain, motifs X and I-III to the AdoMet-binding subdomain, and the variable region with the

adjacent motif IX (present only in m^5C MTases) was recognized as the TRD, suggested to be acting as an autonomous structural and functional domain [6, 97, 126]. That alignment has been validated and its details refined by comparison with crystal structures of m^6A MTases *Taq*I [100], *Dpn*M [101], and *Rsr*I [102] and m^4C MTase *Pvu*II [28].

According to the possible linear arrangements of the AdoMet-binding subdomain, the active site subdomain, and the variable region assumed to function as a TRD, the amino-MTases were subdivided into 6 classes: $\alpha, \beta, \gamma, \delta, \varepsilon$ and ζ [97] (Fig. 8). The majority of known DNA amino-MTases fall into the α, β, and γ classes, with no *bona fide* γ-m^4C MTases discovered yet. M.*Ngo*MXV and its homolog M.*Lmo*A118I are the only experimentally characterized m^4C MTases relatively closely similar to γ-m^6A MTases, however they lack a well-defined TRD [106, 127]. Similarly, sequence analysis and structure prediction for a small group of viral γ-like Dam MTases indicated that due to the lack of TRD they cannot be put into any of the proposed classes [107, 128]. Besides, we have identified two families of enzymes closely related to DNA amino-MTases, namely 16S rRNA: guanine-N2 MTases and the HemK family of putative nucleic acid MTases that possess a large variable region at the N-terminus, and therefore should be classified as putative members of the ζ class [129, 130]. It has been also found that the m^4C MTase M.*Mwo*I exhibits the δ architecture [131], rather than previously proposed β [97]. Nearly all m^5C MTases differ from the group γ MTases only in the position of motif X, corresponding to a helix packing against the central beta-sheet next to motif I: in m^5C MTases it is as the C-terminus, while in γ MTases it is in N-terminus. Nevertheless, two exceptions to this rule have been identified: the M.*Bss*HII MTase, which is a typical member of the ζ class with the TRD at the N-terminus followed by the conserved motifs IX, X, I–VIII [132], and a family of putative *de novo* DNA MTases from *Arabidopsis* and

maize (DRM2), that contain a MTase module with a unique arrangement of motifs: VI–VIII–TRD–IX–X, I–V [133] (Fig. 8). Based on careful sequence analysis and molecular modeling it has been proposed that the atypical architecture of M.*Bss*HII is not a result of a simple gene permutation event, but rather a series of recombination events between of fragments of genes coding for up to three different m^5C MTases [134].

Lately, models of circular permutation during evolution of m^4C [105] and m^6A MTases [135] have been proposed. Jeltsch argued that the domain permutation process needs duplication of a MTase gene, producing one enzyme with two catalytic domains. For instance, after formation of new start and stop codons in a hypothetical tandem $\gamma\gamma$-class MTase, a ζ- or β-like permutant would arise. This model corresponds to the widely accepted concept that a permuted protein may arise naturally from tandem repeats by extraction of the C-terminal portion of one repeat together with the N-terminal portion of the subsequent repeat, if the protein's N and C termini are in close spatial proximity [136]. Although the idea itself offers a plausible explanation for the origin of permutants within many protein families, the only duplicated m^6A MTases known to date are the type IIS enzymes of the $\alpha\alpha$-class, whose permutation would eventually produce enzymes of the δ or ε classes that have not been identified to date. M.*Mwo*I, the only plausible candidate for the δ class known to date, is closely related to β MTases, and its putative TRD seems to have "jumped" from the position in the middle of the protein to the C-terminus without convincing evidence for duplication of the entire MTase gene (Ref. [131] and J.M. Bujnicki and M. Radlinska, unpublished data). In my opinion, simple interconversions of topologies from $\gamma\gamma$ to β or from $\beta\beta$ to γ are rather implausible, since the TRDs of known MTases from β and γ classes are unrelated [100, 102]. Moreover, the N- and C-termini of M.*Taq*I, the only γ-m^6A MTase whose 3D structure is known,

are quite distant in space [100]. Still, this scheme may be valid for enzymes, which have not been identified yet, or whose sequences have not been studied in enough detail. Nonetheless, I believe that in most cases permutation of m^4C and m^6A MTases occurred *via* intragenic relocations of gene segments (i.e. "domain shuffling" [137]), which left no evident intermediates or fusions and rearrangements of gene fragments [105], rather than solely according to the "duplicate and get rid of redundant termini" scheme. However, to my knowledge, no systematic study has been published, which would infer the evolutionary history of shuffled fragments of MTase domains in enzymes other than M.*Bss*HII [134].

ENase domain

ENase exerts the second key activity of the RM system and therefore could be predicted to exhibit the degree of conservation at least similar to that of the MTase counterpart. However, among numerous ENase sequences known there are only a few that exhibit statistically significant similarity. The lack of sequence conservation has led to speculation that despite common features, such as a requirement for Mg^{2+} and outstanding sequence specificity, most ENases may be unrelated to one another [138]. Initially, the only similarities were detected between type II izoschizomers, enzymes with identical cleavage specificity, which may be regarded as direct descendants of one ancestor, transferred horizontally to different hosts [59, 139]. Nevertheless, X-ray crystallographic studies of 13 seemingly dissimilar type II ENases demonstrated unequivocally that they share a common structural core and metal-binding/catalytic site, arguing for extreme divergence rather than independent evolution of a similar fair-sized domain (for the most recent reviews see [10, 38, 61, 140]). This domain, termed "PD-(D/E)XK" for a very weakly conserved signature of the active site, turned out to be common to other nucleases, including phage

exonuclease [141], two Archaeal Holliday junction resolvases Hjc [142, 143], phage T7 Endonuclease I [144], transposase TnsA [145] and two enzymes exerting ssDNA nicking in the context of methyl-directed and very short patch DNA repair: MutH [79] and Vsr [80]. It is particularly interesting that MutH and Vsr are genetically linked with DNA MTases Dam and Dcm, respectively. Since the sequences of structurally characterized PD-(D/E)XK cleavage domains seemed too divergent for "regular" phylogenetic analysis, a structure-based treeing has been carried out in a similar manner to that performed for MTase domains [140]. From this and other structure-based comparative studies it can be concluded that the PD-(D/E)XK superfamily can be divided into two lineages, roughly corresponding to "5′ four-base overhang cutters" like *Eco*RI or *Bam*HI that interacts with the target DNA predominantly *via* an α-helix and a loop and the "blunt end cutters" like *Pvu*II and *Eco*RV that use a β-strand for DNA recognition [38]. A hypothetical evolutionary scenario of evolution of the two main ENase lineages based on comparison of publicly available crystal structures is shown in Fig. 9.

Recently, despite limitations resulting from extreme divergence of the PD-(D/E)XK domain, state-of-the-art algorithms for sequence comparisons and structure prediction allowed to identify it in a variety of other genuine and putative nucleases, including the (m^6A or m^5C)-specific restriction enzyme Mrr and its homologs, the McrC subunit of the (m^4C, m^5C or hm^5C)-specific restriction enzyme McrBC, the hm^5C-specific restriction enzyme *Pvu*Rts1I, herpesvirus alkaline exonucleases, Archaeal-type Holliday junction resolvases Hjc, various proteins containing the NTPase module like the RecB and DNA2 nuclease families or other enzymes involved in DNA recombination and repair [146–151]. It has been also found out that the catalytic domain of tRNA splicing endonuclease EndA bears striking resemblance to the minimal core of the PD-(D/E)XK fold [152], although it developed

the RNase A-like active site in a distinct location [153]. It is tempting to speculate that EndA may be related to a "common ancestor" of the PD-(D/E)XK superfamily (Fig. 9), however this hypothesis must await a thorough structure-based phylogenetic study with atomic coordinates of more ancient nucleases available.

Ironically, following the series of crystallographic studies suggesting common origin of all ENase domains in restriction enzymes and related DNA repair and recombination enzymes, bioinformatics studies provided evidence that some *bona fide* type II ENases are in fact diverged members of other well-studied nuclease superfamilies, unrelated to the PD-(D/E)XK enzymes (Fig. 10). It has been found that the N-terminal part of the type IIS restriction enzyme exhibits low sequence similarity to an EDTA-resistant nuclease (Nuc) of *Salmonella typhimurium*, and the relationship of these nuclease domains has been confirmed experimentally [113]. We have also identified the Nuc-like domain in type II restriction enzymes *Ngo*FVII, *Ngo*AVII, and *Cgl*I (J.M. Bujnicki, M. Radlińska, V. Siksnys, unpublished data). Another evolutionarily unrelated nuclease domain, similar to the catalytic domain of nucleases from the HNH superfamily, has been identified in the m^5C-specific restriction enzyme McrA, type II restriction enzymes *Hpy*I, *Nla*III, *Sph*I, *Sap*I, *Nsp*HI, *Nsp*I and *Kpn*I, and in type IIS enzyme *Mbo*II and its homologs from *Helicobacter pylori* by our group [114, 154] and by Eugene Koonin's group [147]. We have also found that type II enzymes *Eco*29kI, *Ngo*MIII, *Ngo*AIII, and *Mra*I are homologous to the GIY-YIG endonuclease domain present in certain homing endonucleases and DNA repair and recombination enzymes [114] and that the *Hgi*DII enzyme is related to the DNA repair enzyme MutL, which also possesses a distinct fold (J.M. Bujnicki, unpublished, and P. Friedhoff, cited as personal communication in ref. [10]). Presently, most of these predictions await experimental confirmation, however even in the
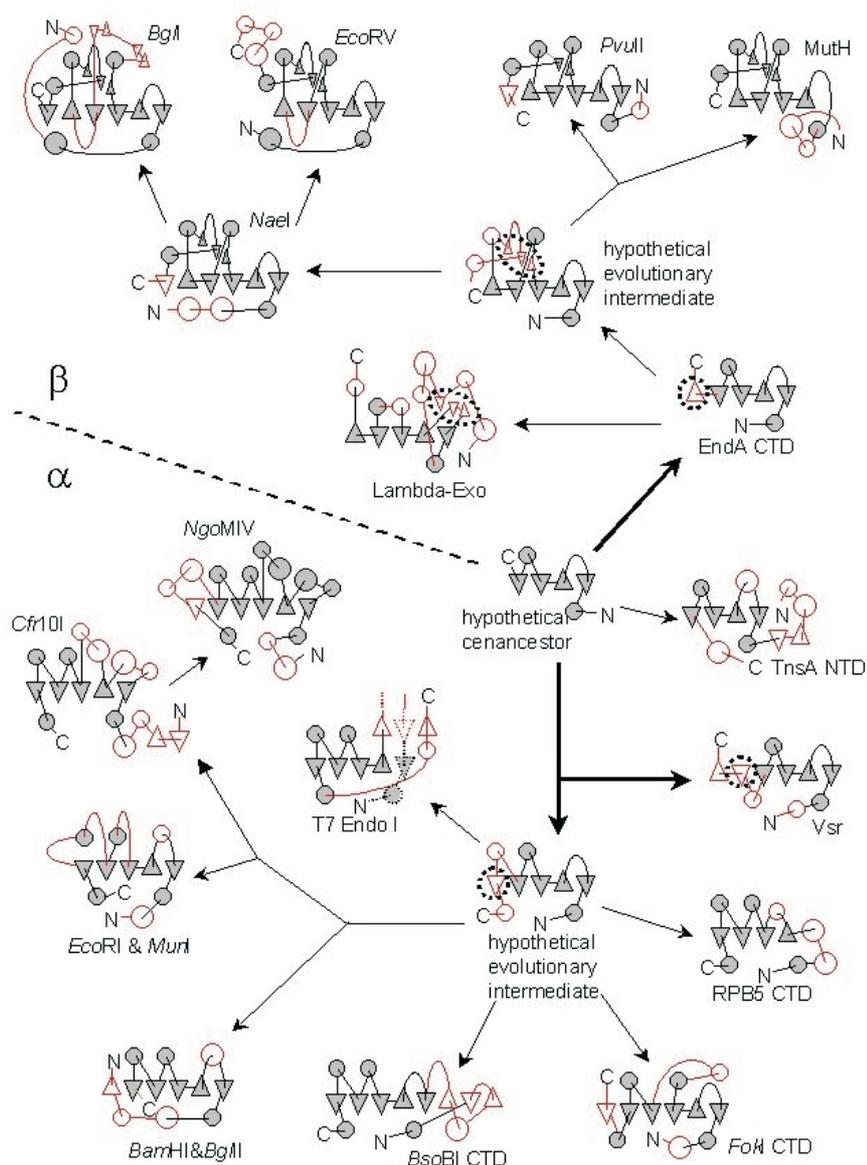
Figure 9. Proposed scheme of evolution of the PD-(D/E)XK family of proteins that depicts radiation and divergence of the α and β subfamilies of restriction enzymes [38, 140].

Secondary structural elements in the topological diagrams are coded as described in Fig. 7a. Evolutionary steps (acquisition and loss of structural elements) are indicated by arrows, elements that are conserved in a given step and in a given sub-lineage are shaded, novel elements are shown in white. The major features that allow distinction between the two lineages are depicted by dotted circles: i) the directionality of the 5th β-strand (parallel in the α-lineage and antiparallel in the β-lineage) and ii) the appearance of an additional small β-sheet that participates in target recognition in the β-lineage. The additional β-sheet of λ-exo and other β-enzymes is a topologically different and hence independently acquired feature. Other peculiarities are the unusual left-handed β-α-β element at the C-teriminal edge of the β-strand in Vsr [80], as opposed to the typical right-handed structure in other proteins, and the fact that the core of T7 Endo I is made of fragments of two polypeptide chains forming a swapped dimer [144].

absence of crystal structures of ENases with any of the three "alternative" folds it became clear that restriction enzymes have evolved on multiple occasions. Moreover, analysis of the various combinations of structural modules present in homing endonucleases and type IIS
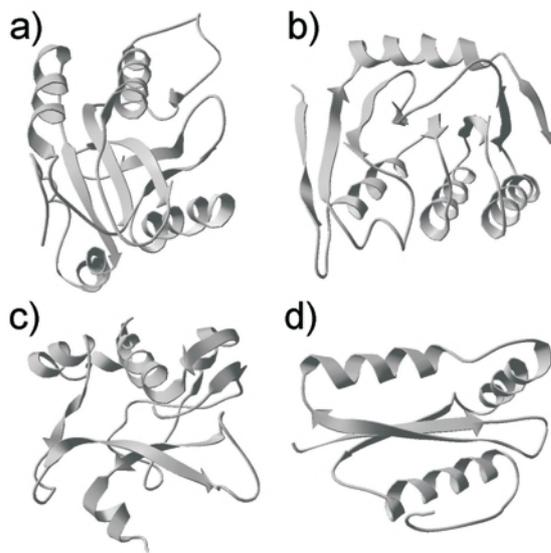
Figure 10. Cartoon diagrams of four structurally and evolutionarily distinct nuclease families, whose members have been identified as alternative ENase domains in the context of RM systems.

a) The canonical PD-(D/E)XK domain exemplified by a non-specific cleavage domain of *Fok*I (PDB code 1fok [111]), which shows relatively few elaborations of the minimal common fold as compared to other, sequence-specific enzymes (Fig. 8), **b)** a Mg^{2+}-independent and hence EDTA-resistant Nuc/phospholipase D domain; a homology model of *Ngo*AVII (J.M. Bujnicki, unpublished data) based on coordinates of the *S. typhimurium* nuclease (1byr), **c)** the DNase domain of *coli*cin E7, a HNH superfamily member (7cei [155]), **d)** a model of the GIY-YIG nuclease catalytic domain obtained from an *ab initio* folding simulation based on published NMR restraints [156].

and certain multimodular type II restriction enzymes suggests that the "remote cutters" arose independently multiple times from various combinations of "cleavage domains" and TRDs with alternative folds and therefore represent an interesting example of convergent evolution.

DNA translocase (helicase-like) domain

All type I and III restriction enzymes, together with the modification-dependent enzyme McrBC, require two recognition sites in linear DNA and nucleotide triphosphate (NTP) hydrolysis before DNA cleavage can occur [70]. Type I and III restriction enzymes require ATP for activity (reviewed in ref. [8]), while McrBC requires GTP [157]. Type I enzymes and McrBC exhibit a similar mechanism: they translocate along DNA from their recognition sites in a reaction powered by NTP hydrolysis until they encounter a block to translocation, which stimulates DNA cleavage [158, 159]. The block is normally another enzyme molecule translocating from another site or a topological barrier resulting from supercoiling of the loop between the two enzymes, explaining the dependence of reaction on two sites. However, other non-specific blocks to translocation, such as a bound repressor or a Holliday junction also stimulate cleavage. One peculiarity of type I enzymes is that they do not turn over in the cleavage reaction, but they hydrolyze ATP long after DNA cleavage has stopped [160]. In contrast, type III enzymes, require a specific contact between the two translocating enzyme molecules and non-specific blocks are inhibitory [19]. Bickle and coworkers demonstrated that cooperation between two enzymes is necessary for ds DNA cleavage, since each translocating enzyme complex cuts only one strand of DNA [22].

The R subunit of all type I RM systems and the Res subunit of all III RM systems comprise two modules: a large DNA translocase module, exhibiting sequence similarity to certain DNA and RNA helicases (Fig. 11a) [161] and a small PD-(D/E)XK cleavage domain (Figs. 1c, 2b). In type I enzymes the PD-(D/E)XK domain is located at the N-terminus of the DNA translocase domain, while in type III enzymes it is located at its C-terminus [12], implying another case of sequence permutation in RM proteins.

Helicases are enzymes that separate duplex DNA or RNA into single strands with the help of ATP; on the basis of sequence comparison, they have been classified into five "superfamilies" (reviewed in refs. [162, 163]). How-
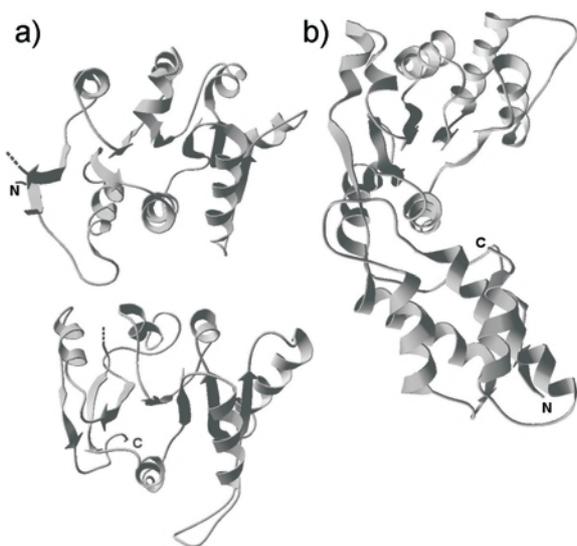
Figure 11. Cartoon diagrams of components of the DNA translocase modules in NTP-dependent restriction enzymes.

a) The two RecA-like domains of the *Eco*AI R subunit homology-modeled (J.M. Bujnicki, unpublished) based on atomic coordinates of the ATP-dependent "DEAD-box" proteins Mj0669 and Eif-4A (1hv8 and 1qva, respectively). The detailed mode of protein–DNA interactions and the mutual position of the two domains in the active enzyme is unknown, **b)** the *E. coli* McrB monomer homology-modeled (J.M. Bujnicki, unpublished) based on atomic coordinates of the AAA+-superfamily members RuvB (1hqc), Cdc6P (1fnn) and the D2 domain of *N*-ethylmaleimide-sensitive fusion protein (1d2n).

ever, many proteins containing motifs common to one or more of the "superfamilies" and described initially as "putative helicases", do not appear to catalyze an unwinding reaction [163]. Remarkably, the strand separation and translocation activity could not be demonstrated for type I and III ENases, however it is believed that they accomplish dsDNA translocation *via* a helicase-like mechanism [12]. The DNA translocase module of type I and III ENases belongs to the large group of evolutionarily related enzymes, which includes helicase superfamilies I and II and various DNA recombination and repair enzymes [12, 146]. This module spans two structurally similar domains, whose fold is related to that

of the RecA protein [164], and several regions, which are not conserved between "superfamilies" and which in type I ENases were suggested to form additional domains required for protein–protein interactions [12] (Fig. 1c).

McrBC is the only known nuclease, which requires GTP [157]. Deletion mutagenesis studies demonstrated that the N-terminal domain of McrB, missing from the naturally truncated form McrB$_S$, is solely responsible for DNA binding and can be regarded as the TRD [165, 166]. On the other hand, GTP-binding motifs were identified in the amino-acid sequence of the central and C-terminal region of McrB [66], which also harbors determinants for binding of McrC [167]. However, site-directed mutagenesis studies suggested that McrB is functionally and presumably structurally distinct from the classic GTP-binding proteins [168]. Recently, based on extensive bioinformatics analysis, it has been suggested that the GTPase module of McrB is related to the so-called AAA-ATPases (ATPases associated with a variety of cellular activities) [169, 170], as well as the DnaA and RuvB helicases, the Clp/Hsp100 family, clamp loading subunits for DNA polymerase, dynein motors and other proteins that appear to function as molecular matchmakers in the assembly, operation, and disassembly of diverse protein machines or DNA–protein complexes [171] (Fig. 11b). In many cases, AAA domains assemble into hexameric rings that are likely to change their shape during the ATPase cycle (reviewed in ref. [172]). However, the results of gel filtration and scanning transmission electron microscopy analysis indicate that McrB and its truncated version McrB$_S$ form forms single heptameric rings as well as tetradecamers, with the latter being more stable when McrC is bound [173]. However, the location and exact stoichiometry of McrC in the McrBC nuclease could not be identified. Moreover, it is still unclear, why McrBC is dependent on GTP and not on ATP, like virtually all of its homologs.

Regulatory proteins

The characterization of type II RM systems has shown that some systems contain other components in addition to the requisite endonuclease and methyltransferase. One of these is the C (controller) protein, which has been proposed to allow establishment of RM systems in new hosts by delaying the appearance of restriction activity; its gene generally precedes and in some cases partially overlaps the ENase gene [174]. C proteins have not yet been structurally characterized, but their amino-acid sequences reveal that they are probably helix-turn-helix proteins similar to numerous known activators and repressors of gene expression (reviewed in ref. [175]). Sequence comparisons have identified a conserved DNA sequence element termed a "C box" immediately upstream of most C genes [176]. It has been shown that C.*Pvu*II and C.*Bam*HI are DNA-binding proteins that bind to the C box and by autogenous activation of the polycistronic *pvuIICR* or *bamHICR* promoter contribute to the temporal activation of the ENase gene expression (ref. [177] and A. Sohail, I. Ghosh, R.M. Fuentes, and J.E. Brooks, unpublished results cited therein). It has been also demonstrated that there is some cross-complementation between the C genes from different RM systems [178].

Kobayashi and coworkers reported that some type II RM systems on plasmids resist displacement by a plasmid bearing RM systems with ENase and MTase of distinct specificity but the C protein of the same specificity. An apparent cell suicide results from chromosome cleavage at unmodified sites by prematurely expressed ENase from an incoming RM system [179]. In general, C genes were found to play important roles in the maintenance, establishment, and mutual exclusion of RM systems. These roles are reminiscent of the strategies of temperate bacteriophages [180] and are in accord with the "selfish gene" hypothesis for the spread and maintenance of RM gene complexes [5, 181] (see also below).

The regulatory protein from the unusual *Lla*I RM system [62] was shown to enhance expression of *Lla*I restriction at a post-transcriptional level rather than to function as a transcriptional activator, despite its sequence similarity to HTH proteins [182]. Similarly, regulation of the ENase activity by inhibiting intracellular subunit association was reported for the *Pvu*II enzyme and a 28-amino-acid peptide, designated W.*Pvu*II [183].

Other elements associated with RM systems

There have been several reports of the close association between enzymes involved in DNA mobility and RM systems. Genes and partial genes encoding phage-like integrases and other proteins from the tyrosine recombinase (Int) superfamily occur next to the *sinIR* [184], *accIM* [185], *ecoHK31IM*, and *eaeIM* genes [186]. Genes for putative proteins similar to DNA invertases and resolvases are found near the *Pae*R7I [187], *Bgl*II [188], and *Apa*LI [189] RM systems. A complete copy of the IS*982* element with a DDE-superfamily transposase-encoding gene was identified between the *llaKR2IR* and *llaKR2IM* genes [190]; a putative transposase was also found in the intergenic area between the *eco47IR* and *eco47IIM* genes [191]. These proteins may facilitate the transfer of RM genes among different bacterial strains.

Genomic context of evolution, structure, and function of RM systems

Currently hundreds of sequences of functionally characterized DNA MTases and ENases are available in public databases [192]. Although this number is still growing, we are also faced with a virtual explosion in the number of sequences of putative RM protein deduced from data produced by numerous Prokaryotic genome-sequencing projects. 75% of completely sequences genomes appear to contain multiple RM systems (up to two dozens in the case of *Helicobacter pylor*i J99),

most of which have never been assayed biochemically. However, as emphasized based on the recent results of genome-wide analyses carried out for putative RM systems of *H. pylori* J99 [193], *H. pylori* 26695 [194] and Cyanobacterium *Anabaena* strain PCC7120 [122], many of the candidate genes are in fact pseudogenes in various states of decomposition. Nevertheless, as demonstrated for the *Hpy*99I system, which has been identified based on sequence analysis and subsequently characterized biochemically, the remaining active RM genes, may be a rich source of novel specificities [193]. Evidently, the genome-based screening method has several important advantages over conventional methods employing testing the crude cell extracts for their restriction activity: it can save the fermentation of large amount of microbes, which may pathogenic or very difficult to grow; and allows cloning and expression of RM systems, whose activity is not detectable in cell extracts.

Genome-wide comparisons carried out for pairs of related strains of: ε-Proteobacteria *H. pylori* [195] and Archaea *Pyrococcus abyssii* and *P. horikoshii* [196] suggested that the presence of RM systems is often associated with various types of genome polymorphisms. It has been noted, that certain chromosomal loci in different strains of related bacteria may be associated with unrelated or very remotely related RM systems that exhibit different specificities [197]. This suggests that the representational difference analysis may be used for isolation of novel RM systems based on genomic sequence analysis even if the sequence of the genome of the strain of interest is not available.

From the data generated by combined theoretical and experimental genomic approaches many more surprises can be expected, not only as a result of enzymes with new specificities or new "types" combining old domains in unprecedented manner, but also because some RM systems may comprise novel domains, not related to those described in this article. For instance, it seems plausible that some restriction enzymes comprise cleavage domains homologous to the LAGLIDADG, AP, RusA, RuvC/RNase H or other nuclease superfamilies [147, 198], rather than the PD-(D/E)XK, HNH, GIY-YIG and Nuc superfamilies described to date. On the other hand, the numerous ongoing structural genomics programs will undoubtedly provide more insight into cases like the yeast RPB5 subunit of RNA polymerase from *Saccharomyces cerevisiae*, which comprises a PD-(D/E)XK-like domain without the nuclease active site [199] or the EndA enzyme, whose PD-(D/E)XK-like domain acquired the RNase A-like active site on an opposite face of the protein [153]. The latter case is especially interesting, since it suggests that additional binding or catalytic sites could be engineered in structures of restriction ENases from the PD-(D/E)XK superfamily [152].

The existence of specific relationships between certain restriction enzymes and other evolutionarily conserved nucleases inferred from structural studies and sequence comparisons on a genome scale suggests that they have arisen on multiple occasions from different nuclease lineages [147]. It is tempting to speculate that most of restriction ENases evolved as self-propagating, "selfish" elements from DNA repair enzymes or other cellular nucleases, however the available data do not allow to draw definite conclusions. Nonetheless, in the course of comparative analysis of sequences and structures of various nucleases carried out by our group and by others it became clear that the major families of sequence-specific restriction enzymes are related to either structure-specific or nonspecific nucleases [114, 140, 146, 147, 149, 150, 154, 198]. It suggests that evolutionary pathways leading from non-specific nucleases to highly sequence-specific restriction enzymes or *vice versa* can be inferred, provided sufficient number of sequences and structures corresponding to "evolutionary intermediates". Even though many putative RM genes are in-

active, their sequences may aid in generation of multiple sequence alignments and phylogenetic trees. The use of "intermediate sequences" is also helpful in molecular modeling, where one attempts to predict the three-dimensional structure of a protein of interest based on sequence alignment to a homologous protein of known structure [200, 201].

Such information could guide mutagenesis experiments aiming at rational engineering of restriction enzymes with new specificities. To date, attempts to change the specificity of type II restriction enzymes using site-directed or random mutagenesis were rather unsuccessful [202, 203]. It has been concluded that even for the very well characterized restriction enzymes, like *Eco*RV, properties that determine specificity and selectivity are difficult to model on the basis of the available structural information [204]. However, with the broad range of enzymes with different specificities in hand one can systematically analyze the structure–function relationships and follow the evolutionary history of selected families of RM proteins. Since MTases show much greater sequence similarity than ENases, several projects have been launched aiming at engineering of their specificity based on phylogenetic analysis and identification of mutations correlated with functional modifications. To date, there has been no spectacular success; it has been concluded that the evolutionary pathway for specificity change leads through a stage of relaxed specificity (ref. [108], S. Klimasauskas, personal communication, J.M. Bujnicki and M. Radlinska, unpublished). It suggests that best targets for specificity engineerning would be not the highly specific enzymes studied presently, but the "sloppy" ones [205], which make only a few key protein–DNA contacts to recognize their target (or rather a broad range of targets). A similar approach seems applicable for engineering of ENases with novel specificities. In my opinion, engineering specificity into polypeptide loops of inherently non-spe-

cific cleavage domains that are able to bind to DNA on their own seems more promising than modifying the highly elaborated DNA-binding surface of enzymes like *Eco*RV. Unfortunately, only a few crystal structures are available for the non-specific nucleases [112, 206] and none for the "sloppy" MTases. Our unpublished results suggest that the three-dimensional structure of certain ENases can be predicted based on results of sequence-structure threading even in the absence of significant sequence similarity, however it remains to be verified experimentally if such models are of sufficient resolution to guide knowledge-based redesign of DNA-binding determinants. Nevertheless, it is obvious that good insight into evolutionary plasticity of functionally important elements in RM proteins can be obtained in the course of comparative analysis carried out using advanced computational methods. In my opinion the elusive goal of creating MTases and ENases with novel specificities will be achieved only if the large-scale bioinformatics and experimental approaches are combined.

CONCLUSIONS

This review covers recent results on the structure and evolution of RM enzymes. One immediately obvious fact is the rapid acceleration in the production of new data in this field. This has allowed the demonstration of phylogenetic and mechanistic links between RM enzymes and other proteins that often possess similar biochemical or enzymatic properties. The wealth of new data becoming available should help to answer many open questions concerning the structure–function relationships of RM proteins. No doubt the approach of functional genomics will play a significant role in identifying genes coding for novel ENases and MTases, and the newly developed computational tools will guide their experimental characterization and engineering, suggesting that a new era in research on the

structure and function of RM systems has just begun.

———————

REFERENCES

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

2. Arber, W. & Dussoix, D. (1962) Host specificity of DNA produced by *Escherichia coli*. Host controlled modification of bacteriophage λ. *J. Mol. Biol.* **5**, 18–36.

3. Arber, W. (1979) Promotion and limitation of genetic exchange. *Science* **205**, 361–365.

4. Lacks, S.A., Ayalew, S., de la Campa, A.G. & Greenberg, B. (2000) Regulation of competence for genetic transformation in *Streptococcus pneumoniae*: Expression of dpnA, a late competence gene encoding a DNA methyltransferase of the *Dpn*II restriction system. *Mol. Microbiol.* **35**, 1089–1098.

5. Kobayashi, I., Nobusato, A., Kobayashi-Takahashi, N. & Uchiyama, I. (1999) Shaping the genome-restriction-modification systems as mobile genetic elements. *Curr. Opin. Genet. Dev.* **9**, 649–656.

6. Dryden, D.T. (1999) Bacterial DNA methyltranferases; in: *S-Adenosylmethionine-dependent Methyltransferases: Structures and Functions* (Cheng, X. *et al.,* eds.) pp. 283–340, World Scientific Inc., Singapore.

7. Murray, N.E. (2000) Type I restriction systems: Sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.* **64**, 412–434.

8. Rao, D.N., Saha, S. & Krishnamurthy, V. (2000) ATP-dependent restriction enzymes. *Prog. Nucleic Acid Res. Mol. Biol.* **64**, 1–63.

9. Pingoud, A. and Jeltsch, A. (1997) Recognition and cleavage of DNA by type-II restriction endonucleases. *Eur. J. Biochem.* **246**, 1–22.

10. Pingoud, A. & Jeltsch, A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* **29**, 3705–3727.

11. Pingoud, A., Jeltsch, A., Maxwell, A. & Sherratt, D. (2001) Enzymes that keep DNA under control: Meeting: DNA enzymes: structures and mechanisms. *EMBO Rep.* **2**, 271–276.

12. Davies, G.P., Martin, I., Sturrock, S.S., Cronshaw, A., Murray, N.E. & Dryden, D.T. (1999) On the structure and operation of type I DNA restriction enzymes. *J. Mol. Biol.* **290**, 565–579.

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

14. Ellis, D.J., Dryden, D.T., Berge, T., Edwardson, J.M. & Henderson, R.M. (1999) Direct observation of DNA translocation and cleavage by the *Eco*KI endonuclease using

atomic force microscopy. *Nat. Struct. Biol*. **6**, 15–17.

15. Yuan, R., Hamilton, D.L. & Burckhardt, J. (1980) DNA translocation by the restriction enzyme from *E. coli* K. *Cell* **20**, 237–244.

16. Berge, T., Ellis, D.J., Dryden, D.T., Edwardson, J.M. & Henderson, R.M. (2000) Translocation-independent dimerization of the *Eco*KI endonuclease visualized by atomic force microscopy. *Biophys. J.* **79**, 479–484.

17. Boyer, H.W. (1971) DNA restriction and modification mechanisms in bacteria. *Annu. Rev. Microbiol.* **25**, 153–176.

18. Kauc, L. & Piekarowicz, A. (1978) Purification and properties of a new restriction endonuclease from *Haemophilus influenzae* Rf. *Eur. J. Biochem.* **92**, 417–426.

19. Meisel, A., Mackeldanz, P., Bickle, T.A., Kruger, D.H. & Schroeder, C. (1995) Type III restriction endonucleases translocate DNA in a reaction driven by recognition site-specific ATP hydrolysis. *EMBO J.* **14**, 2958–2966.

20. Bist, P., Sistla, S., Krishnamurthy, V., Acharya, A., Chandrakala, B. & Rao, D.N. (2001) S-Adenosyl-L-methionine is required for DNA cleavage by type III restriction enzymes. *J Mol. Biol*. **310**, 93–109.

21. Meisel, A., Bickle, T.A., Kruger, D.H. & Schroeder, C. (1992) Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature* **355**, 467–469.

22. Janscak, P., Sandmeier, U., Szczelkun, M.D. & Bickle, T.A. (2001) Subunit assembly and mode of DNA cleavage of the type III restriction endonucleases *Eco*P1I and *Eco*P15I. *J. Mol. Biol*. **306**, 417–431.

23. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242.

24. Roberts, R.J. & Cheng, X. (1998) Base flipping. *Annu. Rev. Biochem.* **67**, 181–198.

25. Hornby, D.P. & Ford, G.C. (1998) Protein-mediated base flipping. *Curr. Opin. Biotechnol.* **9**, 354–358.

26. Ho, D.K., Wu, J.C., Santi, D.V. & Floss, H.G. (1991) Stereochemical studies of the C-methylation of deoxycytidine catalyzed by *Hha*I methylase and the N-methylation of deoxyadenosine catalyzed by *Eco*RI methylase. *Arch. Biochem. Biophys.* **284**, 264–269.

27. Ahmad, I. & Rao, D.N. (1996) Chemistry and biology of DNA methyltransferases. *Crit. Rev. Biochem. Mol. Biol.* **31**, 361–380.

28. Gong, W., O'Gara, M., Blumenthal, R.M. & Cheng, X. (1997) Structure of *Pvu*II DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment. *Nucleic Acids Res.* **25**, 2702–2715.

29. Vertino, P.M. (1999) Eukaryotic DNA methyltransferases; in: *S-Adenosylmethionine-dependent Methyltransferases: Structures and Functions* (Cheng, X. *et al.,* eds.) pp. 341–372, World Scientific Inc., Singapore.

30. Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K. & Wilson, K.S. (1993) The crystal structure of *Eco*RV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.* **12**, 1781–1795.

31. Cheng, X., Balendiran, K., Schildkraut, I. & Anderson, J.E. (1994) Structure of *Pvu*II endonuclease with cognate DNA. *EMBO J.* **13**, 3927–3935.

32. Kim, Y., Grable, J.C., Love, R., Green, P.J. & Rosenberg, J.M. (1990) Refinement of *Eco*RI endonuclease crystal structure: A revised protein chain tracing. *Science* **249**, 1307–1309.

33. Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. & Aggarwal, A.K. (1995) Struc-

ture of *Bam*HI endonuclease bound to DNA: Partial folding and unfolding on DNA binding. *Science* **269**, 656–663.

34. Newman, M., Lunnen, K., Wilson, G., Greci, J., Schildkraut, I. & Phillips, S.E. (1998) Crystal structure of restriction endonuclease *Bgl*I bound to its interrupted DNA recognition sequence. *EMBO J.* **17**, 5466–5476.

35. Stankevicius, K., Lubys, A., Timinskas, A., Vaitkevicius, D. & Janulaitis, A. (1998) Cloning and analysis of the four genes coding for *Bpu*10I restriction-modification enzymes. *Nucleic Acids Res.* **26**, 1084–1091.

36. Hsieh, P.C., Xiao, J.P., O'loane, D. & Xu, S.Y. (2000) Cloning, expression and purification of a thermostable nonhomodimeric restriction enzyme, *Bsl*I. *J Bacteriol.* **182**, 949–955.

37. Kruger, D.H., Barcak, G.J., Reuter, M. & Smith, H.O. (1988) *Eco*RII can be activated to cleave refractory DNA recognition sites. *Nucleic Acids Res.* **16**, 3997–4008.

38. Huai, Q., Colandene, J.D., Chen, Y., Luo, F., Zhao, Y., Topal, M.D. & Ke, H. (2000) Crystal structure of *Nae*I-an evolutionary bridge between DNA endonuclease and topoisomerase. *EMBO J.* **19**, 3110–3118.

39. Deibert, M., Grazulis, S., Sasnauskas, G., Siksnys, V. & Huber, R. (2000) Structure of the tetrameric restriction endonuclease *Ngo*MIV in complex with cleaved DNA. *Nat. Struct. Biol.* **7**, 792–799.

40. Bilcock, D.T. & Halford, S.E. (1999) DNA restriction dependent on two recognition sites: Activities of the *Sfi*I restriction-modification system in *Escherichia coli*. *Mol. Microbiol.* **31**, 1243–1254.

41. Szybalski, W., Kim, S.C., Hasan, N. & Podhajska, A.J. (1991) Class-IIS restriction enzymes — a review. *Gene* **100**, 13–26.

42. Bitinaite, J., Wah, D.A., Aggarwal, A.K. & Schildkraut, I. (1998) *Fok*I dimerization is required for DNA cleavage. *Proc. Natl. Acad. Sci U.S.A.* **95**, 10570–10575.

43. Vanamee, E.S., Santagata, S. & Aggarwal, A.K. (2001) *Fok*I requires two specific DNA sites for cleavage. *J. Mol. Biol.* **309**, 69–78.

44. Kim, Y.G. & Chandrasegaran, S. (1994) Chimeric restriction endonuclease. *Proc. Natl. Acad. Sci U.S.A.* **91**, 883–887.

45. Kim, Y.G., Smith, J., Durgesha, M. & Chandrasegaran, S. (1998) Chimeric restriction enzyme: Gal4 fusion to *Fok*I cleavage domain. *Biol. Chem.* **379**, 489–495.

46. Chandrasegaran, S. & Smith, J. (1999) Chimeric restriction enzymes: what is next? *Biol. Chem.* **380**, 841–848.

47. Morgan, R.D., Calvet, C., Demeter, M., Agra, R. & Kong, H. (2000) Characterization of the specific DNA nicking activity of restriction endonuclease N.*Bst*NBI. *Biol. Chem*. **381**, 1123–1125.

48. Higgins, L.S., Besnier, C. & Kong, H. (2001) The nicking endonuclease N.*Bst*NBI is closely related to Type IIs restriction endonucleases *Mly*I and *Ple*I. *Nucleic Acids Res.* **29**, 2492–2501.

49. Gunn, J.S. & Stein, D.C. (1997) The *Neisseria gonorrhoeae* S.*Ngo*VIII restriction/modification system: A type IIs system homologous to the *Haemophilus parahaemolyticus Hph*I restriction/modification system. *Nucleic Acids Res.* **25**, 4147–4152.

50. Friedrich, T., Fatemi, M., Gowhar, H., Leismann, O. & Jeltsch, A. (2000) Specificity of DNA binding and methylation by the M.*Fok*I DNA methyltransferase. *Biochim. Biophys. Acta* **1480**, 145–159.

51. Janulaitis, A., Petrusyte, M., Maneliene, Z., Klimasauskas, S. & Butkus, V. (1992) Purification and properties of the *Eco*57I restriction endonuclease and methylase-prototypes of a new class (type IV). *Nucleic Acids Res.* **20**, 6043–6049.

52. Jurenaite-Urbanaviciene, S., Kazlauskiene, R., Urbelyte, V., Maneliene, Z., Petrusyte, M., Lubys, A. & Janulaitis, A. (2001) Characterization of *Bse*MII, a new type IV restriction-modification system, which recognizes the penta-nucleotide sequence 5′-CTCAG(N)(10/8). *Nucleic Acids Res.* **29**, 895–903.

53. Janulaitis, A., Vaisvila, R., Timinskas, A., Klimasauskas, S. & Butkus, V. (1992) Cloning and sequence analysis of the genes coding for *Eco*57I type IV restriction-modification enzymes. *Nucleic Acids Res.* **20**, 6051–6056.

54. Kong, H. (1998) Analyzing the functional organization of a novel restriction modification system, the *Bcg*I system. *J. Mol. Biol.* **279**, 823–832.

55. Vitor, J.M. & Morgan, R.D. (1995) Two novel restriction endonucleases from *Campylobacter jejuni*. *Gene* **157**, 109–110.

56. Piekarowicz, A., Golaszewska, M., Sunday, A.O., Siwinska, M. & Stein, D.C. (1999) The *Hae*IV restriction modification system of *Haemophilus aegyptius* is encoded by a single polypeptide. *J Mol. Biol.* **293**, 1055–1065.

57. Kong, H. & Smith, C.L. (1897) Does *Bcg*I, a unique restriction endonuclease, require two recognition sites for cleavage? *Biol. Chem.* **379**, 605–609.

58. Gormley, N.A., Bath, A.J. & Halford, S.E. (2000) Reactions of *Bgl*I and other type II restriction endonucleases with discontinuous recognition sites. *J Biol. Chem.* **275**, 6928–6936.

59. Jeltsch, A. & Pingoud, A. (1996) Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. *J. Mol. Evol.* **42**, 91–96.

60. Jeltsch, A., Wenz, C., Wende, W., Selent, U. & Pingoud, A. (1996) Engineering novel restriction endonucleases: Principles and applications. *Trends Biotechnol.* **14**, 235–238.

61. Kovall, R.A. & Matthews, B.W. (1999) Type II restriction endonucleases: Structural, functional and evolutionary relationships. *Curr. Opin. Chem. Biol.* **3**, 578–583.

62. O'Sullivan, D.J., Zagula, K. & Klaenhammer, T.R. (1995) In vivo restriction by *Lla*I is encoded by three genes, arranged in an operon with *llaIM*, on the conjugative *Lactococcus* plasmid pTR2030. *J. Bacteriol.* **177**, 134–143.

63. de la Campa, A.G., Kale, P., Springhorn, S.S. & Lacks, S.A. (1987) Proteins encoded by the *Dpn*II restriction gene cassette. Two methylases and an endonuclease. *J. Mol. Biol.* **196**, 457–469.

64. Merkiene, E., Vilkaitis, G. & Klimasauskas, S. (1998) A pair of single-strand and double-strand DNA cytosine-N4 methyltransferases from *Bacillus centrosporus*. *Biol. Chem.* **379**, 569–571.

65. Revel, H.R. (1967) Restriction of non-glucosylated T-even bacteriophage: Properties of permissive mutants of *Escherichia coli* B and K12. *Virology* **31**, 688–701.

66. Dila, D., Sutherland, E., Moran, L., Slatko, B. & Raleigh, E.A. (1990) Genetic and sequence organization of the mcrBC locus of *Escherichia coli* K-12. *J. Bacteriol.* **172**, 4888–4900.

67. Waite-Rees, P.A., Keating, C.J., Moran, L.S., Slatko, B.E., Hornstra, L.J. & Benner, J.S. (1991) Characterization and expression of the *Escherichia coli* Mrr restriction system. *J Bacteriol.* **173**, 5207–5219.

68. Lacks, S. & Greenberg, B. (1977) Complementary specificity of restriction endonucleases of *Diplococcus pneumoniae* with respect to DNA methylation. *J. Mol. Biol.* **114**, 153–168.

69. Janosi, L., Yonemitsu, H., Hong, H. & Kaji, A. (1994) Molecular cloning and expression of a novel hydroxymethylcytosine-specific restriction enzyme (*Pvu*Rts1I) modulated by glucosylation of DNA. *J. Mol. Biol.* **242**, 45–61.

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

71. Jurica, M.S. & Stoddard, B.L. (1999) Homing endonucleases: Structure, function and evolution. *Cell Mol. Life Sci.* **55**, 1304–1326.

72. Gimble, F.S. (2000) Invasion of a multitude of genetic niches by mobile endonuclease genes. *FEMS Microbiol. Lett.* **185**, 99–107.

73. Belfort, M. & Perlman, P.S. (1995) Mechanisms of intron mobility. *J. Biol. Chem.* **270**, 30237–30240.

74. Marinus, M.G. (1996) Methylation of DNA; in *Escherichia coli and Salmonella typhimurium* (Neidhardt, F.C., ed.) pp. 782–791, ASM Press, Washington DC.

75. Peterson, K.R., Wertman, K.F., Mount, D.W. & Marinus, M.G. (1985) Viability of *Escherichia coli* K-12 DNA adenine methylase (dam) mutants requires increased expression of specific genes in the SOS regulon. *Mol. Gen. Genet.* **201**, 14–19.

76. Wright, R., Stephens, C. & Shapiro, L. (1997) The CcrM DNA methyltransferase is widespread in the alpha subdivision of proteobacteria and its essential functions are conserved in *Rhizobium meliloti* and *Caulobacter crescentus*. *J. Bacteriol.* **179**, 5869–5877.

77. Gomez-Eichelmann, M.C. & Ramirez- Santos, J. (1993) Methylated cytosine at Dcm (CCATGG) sites in *Escherichia coli*: Possible function and evolutionary implications. *J. Mol. Evol.* **37**, 11–24.

78. Lieb, M. & Bhagwat, A.S. (1996) Very short patch repair: Reducing the cost of cytosine methylation. *Mol. Microbiol.* **20**, 467–473.

79. Ban, C. & Yang, W. (1998) Structural basis for MutH activation in *E.coli* mismatch repair and relationship of MutH to restriction endonucleases. *EMBO J.* **17**, 1526–1534.

80. Tsutakawa, S.E., Jingami, H. & Morikawa, K. (1999) Recognition of a TG mismatch: The crystal structure of very short patch repair endonuclease in complex with a DNA duplex. *Cell* **99**, 615–623.

81. Behrens, B., Noyer-Weidner, M., Pawlek, B., Lauster, R., Balganesh, T.S. & Trautner, T.A. (1987) Organization of multispecific DNA methyltransferases encoded by temperate *Bacillus subtilis* phages. *EMBO J.* **6**, 1137–1142.

82. Tran-Betcke, A., Behrens, B., Noyer-Weidner, M. & Trautner, T.A. (1986) DNA methyltransferase genes of *Bacillus subtilis* phages: Comparison of their nucleotide sequences. *Gene* **42**, 89–96.

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

84. Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B. & Trautner, T.A. (1988) Sequential order of target-recognizing domains in multispecific DNA-methyltransferases. *EMBO J.* **7**, 2601–2609.

85. Lange, C., Jugel, A., Walter, J., Noyer-Weidner, M. & Trautner, T.A. (1991) 'Pseudo' domains in phage-encoded DNA methyltransferases. *Nature* **352**, 645–648.

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

87. Trautner, T.A., Pawlek, B., Behrens, B. & Willert, J. (1996) Exact size and organization of DNA target-recognizing domains of multispecific DNA-(cytosine-C5)-methyltransferases. *EMBO J.* **15**, 1434–1442.

88. Mi, S. & Roberts, R.J. (1992) How M.*Msp*I and M.*Hpa*II decide which base to methylate. *Nucleic Acids Res.* **20**, 4811–4816.

89. Pradhan, S. & Roberts, R.J. (2000) Hybrid mouse-prokaryotic DNA (cytosine-5) methyl-transferases retain the specificity of the parental C-terminal domain. *EMBO J.* **19**, 2103–2114.

90. Bujnicki, J.M. & Radlinska, M. (1999) Molecular phylogenetics of DNA 5mC-methyltransferases. *Acta Microbiol. Pol.* **48**, 19–33.

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

92. Kneale, G.G. (1994) A symmetrical model for the domain structure of type I DNA methyltransferases. *J. Mol. Biol.* **243**, 1–5.

93. MacWilliams, M.P. & Bickle, T.A. (1996) Generation of new DNA binding specificity by truncation of the type IC *Eco*DXXI hsdS gene. *EMBO J.* **15**, 4775–4783.

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

95. Dybvig, K., Sitaraman, R. & French, C.T. (1998) A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13923–13928.

96. Schouler, C., Gautier, M., Ehrlich, S.D. & Chopin, M.C. (1998) Combinational variation of restriction modification specificities in *Lactococcus lactis*. *Mol. Microbiol.* **28**, 169–178.

97. Malone, T., Blumenthal, R.M. & 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.

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

99. Reinisch, K.M., Chen, L., Verdine, G.L. & Lipscomb, W.N. (1995) The crystal structure of *Hae*III methyltransferase convalently complexed to DNA: An extrahelical cytosine and rearranged base pairing. *Cell* **82**, 143–153.

100. Labahn, J., Granzin, J., Schluckebier, G., Robinson, D.P., Jack, W.E., Schildkraut, I. & Saenger, W. (1994) Three-dimensional structure of the adenine-specific DNA methyltransferase M.*Taq*I in complex with the cofactor S- adenosylmethionine. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10957–10961.

101. Tran, P.H., Korszun, Z.R., Cerritelli, S., Springhorn, S.S. & Lacks, S.A. (1998) Crystal structure of the *Dpn*M DNA adenine methyltransferase from the *Dpn*II restriction system of *Streptococcus pneumoniae* bound to *S*-adenosylmethionine. *Structure* **6**, 1563–1575.

102. Scavetta, R.D., Thomas, C.B., Walsh, M.A., Szegedi, S., Joachimiak, A., Gumport, R.I. & Churchill, M.E. (2000) Structure of *Rsr*I methyltransferase, a member of the N6-adenine beta class of DNA methyltransferases. *Nucleic Acids Res.* **28**, 3950–3961.

103. Dryden, D.T., Sturrock, S.S. & Winter, M. (1995) Structural modelling of a type I DNA methyltransferase. *Nat. Struct. Biol.* **2**, 632–635.

104. O'Neill, M., Dryden, D.T. & Murray, N.E. (1998) Localization of a protein-DNA interface by random mutagenesis. *EMBO J.* **17**, 7118–7127.

105. Bujnicki, J.M. & Radlinska, M. (1999) Molecular evolution of DNA-(cytosine-N4) methyltransferases: Evidence for their polyphyletic origin. *Nucleic Acids Res.* **27**, 4501–4509.

106. Radlinska, M., Bujnicki, J.M. & Piekarowicz, A. (1999) Structural characterization of two tandemly arranged DNA methyltransferase genes from *Neisseria gonorrhoeae* MS11: N4-cytosine specific M.*Ngo*MXV and non-

functional 5-cytosine-type M.*Ngo*Morf2P. *Proteins* **37**, 717–728.

107. Radlinska, M. & Bujnicki, J.M. (2001) Cloning of enterohemorrhagic *Escherichia coli* phage VT-2 Dam methyltransferase. *Acta Microbiol. Pol.* **50**, 151–156.

108. Beck, C., Cranz, S., Solmaz, M., Roth, M. & Jeltsch, A. (2001) How does a DNA interacting enzyme change its specificity during molecular evolution? A site directed mutagenesis study at the DNA binding site of the DNA-(adenine-N^6)-methyltransferase *Eco*RV. *Biochemistry* **40**, 10956–10965.

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

110. Goedecke, K., Pignot, M., Goody, R.S., Scheidig, A.J. & Weinhold, E. (2001) Structure of the N6-adenine DNA methyltransferase M.*Taq*I in complex with DNA and a cofactor analog. *Nat. Struct. Biol.* **8**, 121–125.

111. Wah, D.A., Hirsch, J.A., Dorner, L.F., Schildkraut, I. & Aggarwal, A.K. (1997) Structure of the multimodular endonuclease *Fok*I bound to DNA. *Nature* **388**, 97–100.

112. Wah, D.A., Bitinaite, J., Schildkraut, I. & Aggarwal, A.K. (1998) Structure of *Fok*I has implications for DNA cleavage. *Proc. Natl. Acad. Sci U.S.A.* **95**, 10564–10569.

113. Sapranauskas, R., Sasnauskas, G., Lagunavicius, A., Vilkaitis, G., Lubys, A. & Siksnys, V. (2000) Novel subtype of type IIs restriction enzymes. *J Biol. Chem.* **275**, 30878–30885.

114. Bujnicki, J.M., Radlinska, M. & Rychlewski, L. (2001) Polyphyletic evolution of type II restriction enzymes revisited: Two independent sources of second-hand folds revealed. *Trends Biochem. Sci.* **26**, 9–11.

115. Kowalski, J.C., Belfort, M., Stapleton, M.A., Holpert, M., Dansereau, J.T., Pietrokovski, S., Baxter, S.M. & Derbyshire, V. (1999) Configuration of the catalytic GIY-YIG domain of intron endonuclease I-*Tev*I: Coincidence of

116. Dixon, M., Fauman, E.B. & Ludwig, M.L. (1999) The black sheep of the family: AdoMet-dependent methyltransferases that do not fit the consensus structural fold; in: *S-Adenosylmethionine-dependent Methyltransferases: Structures and Functions* (Cheng, X., *et al.*, eds.) pp. 39–54, World Scientific Inc., Singapore.

117. Song, H.K., Sohn, S.H. & Suh, S.W. (1999) Crystal structure of deoxycytidylate hydroxymethylase from bacteriophage T4, a component of the deoxyribonucleoside triphosphate-synthesizing complex. *EMBO J.* **18**, 1104–1113.

118. Bujnicki, J.M. (1999) Comparison of protein structures reveals monophyletic origin of the AdoMet-dependent methyltransferase family and mechanistic convergence rather than recent differentiation of N4-cytosine and N6-adenine DNA methylation. *In Silico Biol.* **1**, 1–8, http://www.bioinfo.de/isb/1999–01/ 0016/.

119. Lauster, R. (1989) Evolution of type II DNA methyltransferases. A gene duplication model. *J. Mol. Biol.* **206**, 313–321.

120. Jeltsch, A., Christ, F., Fatemi, M. & Roth, M. (1999) On the substrate specificity of DNA methyltransferases. Adenine-N6 DNA methyltransferases also modify cytosine residues at position N4. *J. Biol. Chem.* **274**, 19538–19544.

121. Jeltsch, A. (2001) The cytosine N4-methyltransferase M.*Pvu*II also modifies adenine residues. *J. Biol. Chem.* **382**, 707–710.

122. Matveyev, A.V., Young, K.T., Meng, A. & Elhai, J. (2001) DNA methyltransferases of the Cyanobacterium *Anabaena* PCC7120. *Nucleic Acids Res.* **29**, 1491–1506.

123. Roth, M. & Jeltsch, A. (2001) Changing the target base specificity of the *Eco*RV DNA methyltransferase by rational *de novo* protein-design. *Nucleic Acids Res.* **29**, 1–8.

124. Posfai, J., Bhagwat, A.S., Posfai, G. & Roberts, R.J. (1989) Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.* **17**, 2421–2435.

125. Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V. & Janulaitis, A. (1989) Sequence motifs characteristic of DNA[cytosine-N4]methyltransferases: Similarity to adenine and cytosine-C5 DNA-methylases. *Nucleic Acids Res.* **17**, 9823–9832.

126. Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R.J. & Wilson, G.G. (1994) The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* **22**, 1–10.

127. Bujnicki, J.M. & Radlinska, M. (2001) Cloning and characterization of M.*Lmo*A118I, a novel DNA:m4C methyltransferase from the *Listeria monocytogenes* phage A118, a close homolog of M.*Ngo*MXV. *Acta Microbiol. Pol.* **50**, 151–156.

128. Piekarowicz, A. & Bujnicki, J.M. (1999) Cloning of the Dam methyltransferase gene from *Haemophilus influenzae* bacteriophage HP1. *Acta Microbiol. Pol.* **48**, 123–129.

129. Bujnicki, J.M. & Radlinska, M. (1999) Is the HemK family of putative *S*-adenosylmethionine-dependent methyltransferases a "missing" zeta subfamily of adenine methyltransferases? A hypothesis. *IUBMB Life* **48**, 247–250.

130. Bujnicki, J.M. (2000) Phylogenomic analysis of 16S rRNA:(guanine-N2) methyltransferases suggests new family members and reveals highly conserved motifs and a domain structure similar to other nucleic acid amino-methyltransferases. *FASEB J.* **14**, 2365–2368.

131. Bujnicki, J.M. & Rychlewski, L. (2000) Divergence and retroconvergence in the evolution of sequence specificity and reaction mechanism of DNA methyltransferases and their relatives; in: *Proceedings of the IUBMB Symposium "DNA Enzymes: Structures and Mechanisms"*, Anonymous pp. 61, Bangalore, India.

132. Xu, S.Y., Xiao, J.P., Posfai, J., Maunus, R.E. & Benner, J.S. (1997) Cloning of the *Bss*HII restriction-modification system in *Escherichia coli*: *Bss*HII methyltransferase contains circularly permuted cytosine-5 methyltransferase motifs. *Nucleic Acids Res.* **25**, 3991–3994.

133. Cao, X., Springer, N.M., Muszynski, M.G., Phillips, R.L., Kaeppler, S. & Jacobsen, S.E. (2000) Conserved plant genes with similarity to mammalian de novo DNA methyltransferases. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4979–4984.

134. Bujnicki, J.M. (2000) Homology modelling of the DNA 5mC methyltransferase M.*Bss*HII. is permutation of functional subdomains common to all subfamilies of DNA methyltransferases? *Int. J. Biol. Macromol.* **27**, 195–204.

135. Jeltsch, A. (1999) Circular permutations in the molecular evolution of DNA methyltransferases. *J. Mol. Evol.* **49**, 161–164.

136. Heinemann, U. & Hahn, M. (1995) Circular permutation of polypeptide chains: Implications for protein folding and stability. *Prog. Biophys. Mol. Biol.* **64**, 121–143.

137. Heringa, J. & Taylor, W.R. (1997) Three-dimensional domain duplication, swapping and stealing. *Curr. Opin. Struct. Biol.* **7**, 416–421.

138. Heitman, J. (1993) On the origins, structures and functions of restriction-modification enzymes. *Genet. Eng. N.Y.* **15**, 57–108.

139. Jeltsch, A., Kroger, M. & Pingoud, A. (1995) Evidence for an evolutionary relationship among type-II restriction endonucleases. *Gene* **160**, 7–16.

140. Bujnicki, J.M. (2000) Phylogeny of the restriction endonuclease-like superfamily inferred from comparison of protein structures. *J. Mol. Evol.* **50**, 39–44.

141. Kovall, R.A. & Matthews, B.W. (1998) Structural, functional and evolutionary relationships between lambda-exonuclease and the

type II restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7893–7897.

142. Bond, C.S., Kvaratskhelia, M., Richard, D., White, M.F. & Hunter, W.N. (2001) Structure of Hjc, a Holliday junction resolvase, from *Sulfolobus solfataricus. Proc. Natl. Acad. Sci U.S.A.* **98**, 5509–5514.

143. Nishino, T., Komori, K., Tsuchiya, D., Ishino, Y. & Morikawa, K. (2001) Crystal structure of the Archaeal Holliday junction resolvase Hjc and implications for DNA recognition. *Structure* **9**, 197–204.

144. Hadden, J.M., Convery, M.A., Declais, A.C., Lilley, D.M. & Phillips, S.E. (2001) Crystal structure of the Holliday junction resolving enzyme T7 endonuclease I. *Nat. Struct. Biol.* **8**, 62–67.

145. Hickman, A.B., Li, Y., Mathew, S.V., May, E.W., Craig, N.L. & Dyda, F. (2000) Unexpected structural diversity in DNA recombination: The restriction endonuclease connection. *Mol. Cell.* **5**, 1025–1034.

146. Aravind, L., Walker, D.R. & Koonin, E.V. (1999) Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res.* **27**, 1223–1242.

147. Aravind, L., Makarova, K.S. & Koonin, E.V. (2000) Holliday junction resolvases and related nucleases: Identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res.* **28**, 3417–3432.

148. Kvaratskhelia, M., Wardleworth, B.N., Norman, D.G. & White, M.F. (2000) A conserved nuclease domain in the Archaeal Holliday junction resolving enzyme Hjc. *J. Biol. Chem.* **275**, 25540–25546.

149. Bujnicki, J.M. & 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. Biotechnol.* **3**, 69–72.

150. Bujnicki, J.M. & Rychlewski, L. (2001) Identification of a PD-(D/E)XK-like domain with a novel configuration of the endonuclease ac-

tive site in the methyl-directed restriction enzyme Mrr and its homologs. *Gene* **267**, 183–191.

151. Bujnicki, J.M. & Rychlewski, L. (2001) The herpesvirus alkaline exonuclease belongs to the restriction endonuclease PD-(D/E)XK superfamily: Insight from molecular modeling and phylogenetic analysis. *Virus Genes* **22**, 219–230.

152. Bujnicki, J.M. & Rychlewski, L. (2001) Unusual evolutionary history of the tRNA splicing endonuclease EndA: Relationship to the LAGLIDADG and PD-(D/E)XK deoxyribonucleases. *Protein Sci.* **10**, 656–660.

153. Li, H., Trotta, C.R. & Abelson, J. (1998) Crystal structure and evolution of a transfer RNA splicing enzyme. *Science* **280**, 279–284.

154. Bujnicki, J.M., Radlinska, M. & Rychlewski, L. (2000) Atomic model of the 5-methylcytosine-specific restriction enzyme McrA reveals an atypical zinc-finger and structural similarity to $\beta\beta\alpha$ Me endonucleases. *Mol. Microbiol.* **37**, 1280–1281.

155. Ko, T.P., Liao, C.C., Ku, W.Y., Chak, K.F. & Yuan, H.S. (1999) The crystal structure of the DNase domain of *coli*cin E7 in complex with its inhibitor Im7 protein. *Structure Fold. Des.* **7**, 91–102.

156. Bujnicki, J.M., Rotkiewicz, P., Kolinski, A. & Rychlewski, L. (2001) Three-dimensional modeling of the I-*Tev*I homing endonuclease catalytic domain, a GIY-YIG superfamily member, using NMR restraints and Monte Carlo dynamics. *Protein Eng.* **14**, 717–721.

157. Sutherland, E., Coe, L. & Raleigh, E.A. (1992) McrBC: A multisubunit GTP-dependent restriction endonuclease. *J. Mol. Biol.* **225**, 327–348.

158. Janscak, P., MacWilliams, M.P., Sandmeier, U., Nagaraja, V. & Bickle, T.A. (1999) DNA translocation blockage, a general mechanism of cleavage site selection by type I restriction enzymes. *EMBO J.* **18**, 2638–2647.

159. Panne, D., Raleigh, E.A. & Bickle, T.A. (1999) The McrBC endonuclease translocates DNA

in a reaction dependent on GTP hydrolysis. *J Mol. Biol*. **290**, 49–60.

160. Yuan, R. (1981) Structure and mechanism of multifunctional restriction endonucleases. *Annu. Rev. Biochem.* **50**, 285–319.

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

162. Gorbalenya, A.E. & Koonin, E.V. (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.* **3**, 419–429.

163. Hall, M.C. & Matson, S.W. (1999) Helicase motifs: The engine that powers DNA unwinding. *Mol. Microbiol.* **34**, 867–877.

164. Korolev, S., Yao, N., Lohman, T.M., Weber, P.C. & Waksman, G. (1998) Comparisons between the structures of HCV and Rep helicases reveal structural similarities between SF1 and SF2 super-families of helicases. *Protein Sci.* **7**, 605–610.

165. Gast, F.U., Brinkmann, T., Pieper, U., Kruger, T., Noyer-Weidner, M. & 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.

166. Pieper, U., Schweitzer, T., Groll, D.H. & Pingoud, A. (1999) Defining the location and function of domains of McrB by deletion mutagenesis. *Biol Chem.* **380**, 1225–1230.

167. Panne, D., Raleigh, E.A. & Bickle, T.A. (1998) McrBs, a modulator peptide for McrBC activity. *EMBO J.* **17**, 5477–5483.

168. Pieper, U., Schweitzer, T., Groll, D.H., Gast, F.U. & Pingoud, A. (1999) The GTP-binding domain of McrB: more than just a variation on a common theme? *J. Mol. Biol.* **292**, 547–556.

169. Confalonieri, F. & Duguet, M. (1995) A 200-amino acid ATPase module in search of a basic function. *BioEssays* **17**, 639–650.

170. Patel, S. & Latterich, M. (1998) The AAA team: Related ATPases with diverse functions. *Trends Cell Biol.* **8**, 65–71.

171. Neuwald, A.F., Aravind, L., Spouge, J.L. & 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.

172. Vale, R.D. (2000) AAA proteins. Lords of the ring. *J Cell Biol.* **150**, F13–F19

173. Panne, D., Muller, S.A., Wirtz, S., Engel, A. & Bickle, T.A. (2001) The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. *EMBO J.* **20**, 3210–3217.

174. Tao, T., Bourne, J.C. & Blumenthal, R.M. (1991) A family of regulatory genes associated with type II restriction-modification systems. *J. Bacteriol.* **173**, 1367–1375.

175. Wintjens, R. & Rooman, M. (1996) Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *J. Mol. Biol.* **262**, 294–313.

176. Rimseliene, R., Vaisvila, R. & Janulaitis, A. (1995) The *Eco72IC* gene specifies a trans-acting factor which influences expression of both DNA methyltransferase and endonuclease from the *Eco*72I restriction-modification system. *Gene* **157**, 217–219.

177. Vijesurier, R.M., Carlock, L., Blumenthal, R.M. & Dunbar, J.C. (2000) Role and mechanism of action of C.*Pvu*II, a regulatory protein conserved among restriction-modification systems. *J Bacteriol.* **182**, 477–487.

178. Ives, C.L., Sohail, A. & Brooks, J.E. (1995) The regulatory C proteins from different restriction-modification systems can cross-complement. *J. Bacteriol.* **177**, 6313–6315.

179. Nakayama, Y. & Kobayashi, I. (1998) Restriction-modification gene complexes as selfish gene entities: Roles of a regulatory system in their establishment, maintenance and apoptotic mutual exclusion. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6442–6447.

180. Campbell, A. (1994) Comparative molecular biology of lambdoid phages. *Annu. Rev. Microbiol.* **48**, 193–222.

181. Kobayashi, I. (1998) Selfishness and death: Raison d'etre of restriction, recombination and mitochondria. *Trends Genet.* **14**, 368–374.

182. O'Sullivan, D.J. & Klaenhammer, T.R. (1998) Control of expression of *Lla*I restriction in *Lactococcus lactis*. *Mol. Microbiol.* **27**, 1009–1020.

183. Adams, G.M. & Blumenthal, R.M. (1995) Gene *pvuIIW*: A possible modulator of *Pvu* II endonuclease subunit association. *Gene* **157**, 193–199.

184. Karreman, C. & de Waard, A. (1988) Cloning and complete nucleotide sequences of the type II restriction-modification genes of *Salmonella infantis*. *J. Bacteriol.* **170**, 2527–2532.

185. Brassard, S., Paquet, H. & Roy, P.H. (1995) A transposon-like sequence adjacent to the AccI restriction-modification operon. *Gene* **157**, 69–72.

186. Lee, K.F., Shaw, P.C., Picone, S.J., Wilson, G.G. & Lunnen, K.D. (1998) Sequence comparison of *Eco*HK31I and *Eae*I restriction–modification systems suggest an intergenic transfer of genetic material. *Biol. Chem.* **379**, 437–442.

187. Vaisvila, R., Vilkaitis, G. & Janulaitis, A. (1995) Identification of a gene encoding a DNA invertase-like enzyme adjacent to the *Pae*R7I restriction-modification system. *Gene* **157**, 81–84.

188. Anton, B.P., Heiter, D.F., Benner, J.S., Hess, E.J., Greenough, L., Moran, L.S., Slatko, B.E. & Brooks, J.E. (1997) Cloning and characterization of the *Bgl*II restriction-modification system reveals a possible evolutionary footprint. *Gene* **187**, 19–27.

189. Xu, S.Y., Xiao, J.P., Ettwiller, L., Holden, M., Aliotta, J., Poh, C.L., Dalton, M., Robinson, D.P., Petronzio, T.R., Moran, L., Ganatra, M., Ware, J., Slatko, B. & Benner, J. (1998) Cloning and expression of the *Apa*LI, *Nsp*I, *Nsp*HI, *Sac*I, *Sca* I and *Sap*I restriction-modification systems in *Escherichia coli*. *Mol. Gen. Genet.* **260**, 226–231.

190. Twomey, D.P., McKay, L.L. & O'Sullivan, D.J. (2000) Molecular characterization of the *Lactococcus lactis Lla*KR2I restriction-modification system and effect of an IS982 element positioned between the restriction and modification genes. *J. Bacteriol.* **180**, 5844–5854.

191. Stankevicius, K., Povilionis, P., Lubys, A., Menkevicius, S. & Janulaitis, A. (1995) Cloning and characterization of the unusual restriction-modification system comprising two restriction endonucleases and one methyltransferase. *Gene* **157**, 49–53.

192. Roberts, R.J. & Macelis, D. (2001) REBASE-restriction enzymes and methylases. *Nucleic Acids Res.* **29**, 268–269.

193. Kong, H., Lin, L.F., Porter, N., Stickel, S., Byrd, D., Posfai, J. & Roberts, R.J. (2000) Functional analysis of putative restriction-modification system genes in the *Helicobacter pylori* J99 genome. *Nucleic Acids Res.* **28**, 3216–3223.

194. Vitkute, J., Stankevicius, K., Tamulaitiene, G., Maneliene, Z., Timinskas, A., Berg, D.E. & Janulaitis, A. (2001) Specificities of eleven different DNA methyltransferases of *Helicobacter pylori* strain 26695. *J Bacteriol.* **183**, 443–450.

195. Nobusato, A., Uchiyama, I., Ohashi, S. & Kobayashi, I. (2000) Insertion with long target duplication: A mechanism for gene mobility suggested from comparison of two related bacterial genomes. *Gene* **259**, 99–108.

196. Chinen, A., Uchiyama, I. & Kobayashi, I. (2000) Comparison between *Pyrococcus horikoshii* and *Pyrococcus abyssi* genome sequences reveals linkage of restriction-modification genes with large genome polymorphisms. *Gene* **259**, 109–121.

197. Claus, H., Friedrich, A., Frosch, M. & Vogel, U. (2000) Differential distribution of novel restriction-modification systems in clonal lineages of *Neisseria meningitidis*. *J Bacteriol.* **182**, 1296–1303.

198. Lilley, D.M. & White, M.F. (2000) Resolving the relationships of resolving enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9351–9353.

199. Todone, F., Weinzierl, R.O., Brick, P. & Onesti, S. (2000) Crystal structure of RPB5, a universal eukaryotic RNA polymerase subunit and transcription factor interaction target. *Proc. Natl. Acad. Sci U.S.A.* **97**, 6306–6310.

200. Park, J., Teichmann, S.A., Hubbard, T. & Chothia, C. (1997) Intermediate sequences increase the detection of homology between sequences. *J. Mol. Biol.* **273**, 349–354.

201. Sauder, J.M., Arthur, J.W. & Dunbrack, R.L. (2000) Large-scale comparison of protein sequence alignment algorithms with structure alignments. *Proteins* **40**, 6–22.

202. Lanio, T., Jeltsch, A. & Pingoud, A. (1998) Towards the design of rare cutting restriction endonucleases: Using directed evolution to generate variants of *Eco*RV differing in their substrate specificity by two orders of magnitude. *J. Mol. Biol.* **283**, 59–69.

203. Schottler, S., Wenz, C., Lanio, T., Jeltsch, A. & Pingoud, A. (1998) Protein engineering of the restriction endonuclease *Eco*RV-structure-guided design of enzyme variants that recognize the base pairs flanking the recognition site. *Eur. J. Biochem.* **258**, 184–191.

204. Lanio, T., Jeltsch, A. & Pingoud, A. (2000) On the possibilities and limitations of rational protein design to expand the specificity of restriction enzymes: A case study employing *Eco*RV as the target. *Protein Eng.* **13**, 275–281.

205. Radlinska, M. & Piekarowicz, A. (1998) Cloning and characterization of the gene encoding a new DNA methyltransferase from *Neisseria gonorrhoeae*. *Biol. Chem.* **379**, 1391–1395.

206. Miller, M.D., Tanner, J., Alpaugh, M., Benedik, M.J. & Krause, K.L. (1994) 2.1 A structure of *Serratia endonuclease* suggests a mechanism for binding to double-stranded DNA. *Nat. Struct. Biol.* **1**, 461–468.