Annu. Rev. Genet. 1991. 25:585–627
Copyright © 1991 by Annual Reviews Inc. All rights reserved

RESTRICTION AND MODIFICATION SYSTEMS

Geoffrey G. Wilson

New England Biolabs, Beverly, Massachusetts 01915

Noreen E. Murray

Institute of Cell and Molecular Biology, University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh EH9 3JR, Britain

KEY WORDS: endonuclease, methyltransferase, isoschizomer, molecular evolution, protein-DNA sequence recognition

---

CONTENTS

INTRODUCTION 586
BIOLOGY OF RESTRICTION AND MODIFICATION 586
Restriction and modification enzymes 586
Occurrence of R-M systems 588
Restriction and modification of viruses 590
Cloning restriction and modification genes 591
CHARACTERISTICS OF R-M SYSTEMS 592
Type I systems 592
Type II systems 594
Type IIs systems 601
Type III systems 603
Other system types 604
Modification-requiring systems 604
Regulation of expression 605
CONTRASTS AND COMPARISONS AMONG R-M SYSTEMS 606
Type I systems 606
Type II systems 609
Type II endonucleases 612
Methyltransferases 613
DISCUSSION 616

0066-4197/91/1215-0585$02.00

INTRODUCTION

The phenomenon of restriction and modification was first encountered in the early 1950s. Certain strains of bacteria were found to inhibit ("restrict") the propagation of viruses grown previously on different strains (see ref. 2 for an early review). The effect was traced to sequence-specific endonucleases (143), of which some produced discrete DNA fragments upon cleavage (17, 24, 79, 101, 157). The endonucleases proved useful for analyzing and rearranging DNA, and this encouraged the search for more of them (152, 179). They are common among bacteria (175, 200); over two hundred different specificities have now been discovered (102, 178).

The genes for over one hundred restriction-modification (R-M) systems have been cloned, and the genes for approximately sixty have been sequenced. While the systems appear to play equivalent roles in vivo and comprise enzymes that often behave alike in vitro, they are exceedingly diverse at the genetic level (37). R-M systems illustrate extreme examples of both convergent and divergent evolution. In this review we summarize our general understanding of R-M systems and discuss their evolutionary relationships. (For more specialized reviews see (10, 14, 15, 46, 58, 113, 114, 202, 212, 216, 242, 244.))

BIOLOGY OF RESTRICTION AND MODIFICATION

Restriction and Modification Enzymes

Restriction-modification systems comprise pairs of opposing intracellular enzyme activities: an endodeoxyribonuclease (ENase) and a DNA-methyltransferase (MTase). The enzymes interact with specific sequences of nucleotides (nt) in DNA. The sequences usually comprise four to eight defined nt; they can be continuous or interrupted, symmetric or asymmetric, unique or degenerate. The enzymes recognize double-stranded (ds) DNA; a few also recognize single-stranded (ss) DNA (35, 109, 234; S. Chandrasegaran, personal communication). ENases and MTases from the same system ("cognate" enzymes) recognize the same sequences. In some R-M systems, the two activities are combined in a single, multi-subunit enzyme, but in most systems they are separate.

RESTRICTION ENDONUCLEASES Restriction endonucleases catalyze double-strand cleavage of DNA. Cleavage occurs once for each occurrence of the recognition sequence, and is accomplished by hydrolysis of one phosphate-deoxyribose bond in the backbone of each DNA strand. Usually, the reaction occurs without energy input. In many systems, cleavage occurs at a fixed position with respect to the recognition sequence, either within the sequence

or a few bases to one side of it. In others, hydrolysis takes place at an indefinite distance from the recognition sequence. Restriction endonucleases commonly require Mg^{2+} or a similar divalent cation; some also require, or are stimulated by, ATP or *S*-adenosylmethionine (AdoMet). Cleavage generally occurs on the 5'-side of the phosphate, leaving DNA fragments with 5'-phosphoryl and 3'-hydroxyl termini. This is probably an adaptation to facilitate the repair of inadvertently nicked DNA molecules, since fragments with these ends can be rejoined by DNA ligase, unlike fragments with 3'-phosphoryl and 5'-hydroxyl ends.

MODIFICATION METHYLTRANSFERASES Modification methyltransferases catalyze the addition of a methyl group to one nucleotide in each strand of the recognition sequence (reviewed in 202, 212, 244). AdoMet always serves as the methyl donor and is thus an essential cofactor for methylation. The methylated nucleotides lie within the recognition sequence; their identities and positions vary from MTase to MTase. Usually, the same base is methylated on both strands. Adenine and cytosine are the only bases known to be methylated. The products of methylation appear to be confined to *N*6-methyladenine (m^6A) (3, 25, 51, 75, 115, 183), 5-methylcytosine (m^5C) (24, 77), and *N*4-methylcytosine (m^4C) (31, 55, 90). The methyl groups lie in the major groove of the DNA helix, in positions that do not interfere with base-pairing.

METHYLATION The endonucleases of R-M systems are crucially sensitive to methylation of the recognition sequence. Usually, methylation prevents cleavage. "Cognate" methylation, the kind conferred by the natural partner of a restriction endonuclease, always prevents cleavage. This is an adaptation to protect the cell's own DNA from digestion. Noncognate methylation, that occurring elsewhere in the recognition sequence, sometimes prevents cleavage, but not always (141). For most endonucleases, neither the nature of the cognate methylation nor the effects of noncognate methylation can be predicted in advance. Methylation probably prevents cleavage by sterically hindering enzyme: substrate binding. Depending on the exact topology of the protein:DNA interface, methyl groups in some positions will interfere with binding, but will not do so in others.

HEMIMETHYLATION Although both strands of the recognition sequence usually become methylated, cognate methylation of just one strand ("hemimethylation") is sufficient to prevent cleavage. This is important physiologically because hemimethylated sequences arise continuously as the DNA replicates: passage of a replication fork through fully methylated sequences leaves daughter duplexes that are methylated on only one strand.

Eventually the other strands become methylated, but until they do, the integrity of the chromosome depends on the DNA remaining resistant to cleavage.

Since the natural substrates of methyltransferases are hemimethylated, the enzymes might be expected to prefer this form of the recognition sequence. This preference is true for eukaryotic CpG methyltransferases (12, 70) but generally not for bacterial methyltransferases.

Occurrence of R-M Systems

Restriction-modification systems occur exclusively in unicellular organisms, mainly bacteria. They have not been found in eukaryotes, although certain *Chlorella* viruses encode them (223), as do some bacterial viruses (85). They are common among bacteria from all ecological niches and taxonomic groups (87, 178, 180). R-M systems were originally detected by their effects on viral propagation (2); today they are usually detected by assaying cell extracts for the ability to fragment DNA. Over 10,000 bacterial strains have been examined to date, and restriction enzymes—or more precisely, sequence-specific endonucleases—have been found in roughly one-quarter (62, 87, 102, 178).

These assays have limitations, and some of the bacteria scored as negative might in fact be positive. Thus, endonucleases are easily overlooked if they occur at low level; if they are accompanied by excessive exonuclease activity; if they cleave infrequently; or if they have unusual requirements, or unusual substrates. Perhaps most bacteria have restriction systems of one form or another. Approximately one half of the bacteria that possess restriction enzymes possess several enzymes; usually two or three, but occasionally six (123), or even seven (46, 110).

NOMENCLATURE By convention, R-M systems are designated by a three letter acronym derived from the name of the organism in which they occur (203). The first letter comes from the genus, and the second and third letters from the species. The strain designation, if any, follows the acronym. Different systems in the same organism are distinguished by Roman numerals. Thus, *Fnu*DI, *Fnu*DII, and *Fnu*DIII, for example, are three systems from *Fusobacterium nucleatum* strain D. Restriction and modification enzymes are sometimes distinguished by the prefixes R· and M·, as in R·*Hae*III (from *Haemophilus aegyptius*), and M·*Hin*PII (from *H. influenzae* P1). These prefixes are generally omitted if the context is unambiguous. The genes for R-M systems are at times designated *hsd* (for "host-specificity of DNA"), but most often their designations are based on the name for the system (169). The genes for the *Bam*HI system, for example, are designated *bamHIR* and *bamHIM*.

GENE LOCI Far more is known about cloned restriction and modification genes than about the genes in their natural states. *R* and *M* genes are usually closely linked; often they overlap, or are separated by, a few nt. In organisms with several R-M systems, the genes for the systems themselves are sometimes linked, e.g. *Hae*II and *Hae*III (199), and *Nla*III and *Nla*X (116). Some R-M systems are chromosomally encoded (e.g. *Eco*A, *Eco*B and *Eco*K, *Sty*LTI). Others are encoded by plasmids (e.g., *Eco*RI, *Eco*RII, *Eco*RV, *Eco*DXXI, *Eco*R124). A few systems are encoded by temperate bacterial viruses (*Eco*P1), or provirus-like plasmids (*Eco*P15); others are encoded by virulent viruses, such as those of the unicellular algae, *Chlorella* (194, 206, 223). R-M systems have not been found on transposable elements.

SOLITARY ENDONUCLEASES AND METHYLTRANSFERASES Endonucleases and methyltransferases most often accompany one another, as components of R-M systems. Less frequently they occur alone, in contexts that may or may not be relevant to restriction and modification. *Escherichia coli* possesses two such unaccompanied methyltransferases, DNA-adenine methyltransferase (Dam) and DNA-cytosine methyltransferase (Dcm) (140). Dam functions in methyl-directed mismatch repair (7), whereas Dcm is associated with very short patch repair (91, 128, 167). Many viruses from *E. coli* also code for Dam (42, 78, 84, 145, 188, 189).

Some bacterial viruses encode unaccompanied methyltransferases that help them to evade restriction. *Bacillus* phage H2, for example, is immune to the *Bam*HI system of its host because it codes for a methyltransferase of the same specificity (41). *Bacillus* phages ϕ3T, ρ11 and SPR encode unusual methyltransferases with multiple specificities (5, 8, 222, 235). *Dpn*I, the best characterized unaccompanied restriction endonuclease, cleaves the sequence Gm^{6}ATC, provided the sequence is methylated; cleavage is prevented by the absence of methylation (117). Few such endonucleases have been discovered, perhaps because they are scarce, but more likely because they are difficult to detect.

INTRON ENDONUCLEASES Certain mobile group I introns encode unaccompanied endonucleases that function in the dissemination ("homing") of the introns. The recognition sequences of these endonucleases are asymmetric, degenerate, and extended, spanning approximately 20 bp. Intron insertion disrupts the recognition sequences and so prevents further cleavage (9, 52, 120).

Two intron endonucleases from *Saccharomyces cerevisiae* have been characterized: I-*Sce*I (previously, *omega* transposase), from an intron in the large subunit of the mitochondrial rRNA gene (40, 98, 147), and I-*Sce*II (previously, paI 4 endonuclease) from an intron in the cytochrome oxidase gene

(186). Another rRNA intron, in *Physarum polycephalum,* codes for I-*Ppo* (150). All three endonucleases cleave their targets in a staggered fashion, leaving fragments with 3' ss termini of 4 nt. A third yeast endonuclease, HO, functions not in intron homing, but in mating-type switching (154).

Intron endonucleases are not confined to eukaryotes. An intron in the DNA polymerase gene of the bacterium *Thermococcus litoralis* encodes an endonuclease that cleaves in the same fashion as those above (R. Morgan, F. Perler & D. Comb, personal communication). An intron in the thymidylate synthase gene of phage T4 from *E. coli* (166, 195) codes for an endonuclease, I-TevI, that cleaves to leave fragments with 3' ss termini of 2 nt (39).

Restriction and Modification of Viruses

Bacteria with R-M systems are less susceptible to new viral infections than are cells that lack R-M systems (2, 15, 23, 174). This susceptibility can be assessed by measuring the probability of plaque formation ("efficiency of plating": eop). The eop is around 1 for viruses such as λ propagating on nonrestricting *E. coli,* but much lower—typically 10^{-3} to 10^{-5}—for the same phage propagating on restricting cells. The severity of restriction depends on the system and the virus; and up to a point, the eop decreases logarithmically as the number of sites in the viral DNA molecule increases. This approximates single-hit kinetics, and signifies that restriction at a single site is enough to prevent viral reproduction. For cloned R-M systems, the eop can be considerably lower, 10^{-8} or less (71; G. Wilson, unpublished observations).

Viruses that engender successful infections become immune to restriction and multiply thereafter with an eop of 1. The immunity is host-specific and nonheritable: it protects the virus from only the current restriction system, and only as long as the virus propagates continuously on the current host. Modification of viruses is the result of mistaken identity. The cell avoids restricting its own DNA by modifying it, and occasionally the viral DNA is also modified. R-M systems function independently in bacteria that possess multiple systems and, as a result, the overall eop is the product of the individual eops. For cells with more than two systems, the overall eop is exceedingly small. *Neisseria gonorrhoeae,* for example, has multiple systems, which might explain why no phage able to plate on it has yet been found (46, 50).

FUNCTION OF R-M SYSTEMS The prevailing view is that R-M systems function to protect cells from foreign DNA molecules, particularly viruses (163). Of the few known systems actually tested for this ability, most have been found to protect. Bacteriophages, for their part, sometimes encode antirestriction systems (113, 114) and often are deficient in restriction sites (191, 193), indicating that restriction has influenced bacteriophage evolution, but not that it is the sole function of R-M systems. Possible alternative functions include DNA recombination or repair (163); regulation of gene

expression; and acquisition of foreign genes by a process akin to molecular cloning (176, 177). There is little evidence to support such alternative functions.

We know that: (*a*) systems from various bacteria restrict and modify, including those of *E. coli; Sal*I *(Streptomyces); Dpn*I and *Dpn*II *(Streptococcus); Sty*SB, *Sty*SP, and *Sty*LTI *(Salmonella); Bam*HI and *Bsu*RI *(Bacillus);* and several lactococcal systems (92). (*b*) Other systems, untested in situ, restrict and modify when transferred to *E. coli,* including: *Ban*I, *Dde*I, *Hae*II, *Hha*I, *Hha*II, *Hin*P1I, *Fnu*DI, *Nla*IV, *Pae*R7I, *Pst*I, and *Pvu*II (19, 83, 134, 219, 230, 237; G. Wilson, unpublished observations). (*c*) R-M systems are dispensable, and restriction-negative mutants display no apparent deficiencies other than a susceptibility to viral infection. (*d*) No special features differentiate the enzymes known to restrict and modify from those whose properties have not been tested.

The simplest explanation for these observations is that R-M systems function only to protect against infection. The multiplicity and variety of R-M systems is also consistent with this role. Protection does not necessarily occur at the level of the individual, however, since cells might be killed by viral infection whether or not the viral DNA were eventually restricted. When cells are killed, restriction benefits not the individual, but its siblings in the immediate vicinity, since it shields them from infection by what would otherwise be the progeny of the restricted virus. Thus, R-M systems appear to function to check the spread of infectious DNA molecules within populations of clonal bacteria.

Cloning Restriction and Modification Genes

Attempts to clone restriction and modification genes began in the mid-1970s. Selections based on both the restriction phenotype and the modification phenotype have been used to isolate clones. The genes for most of the *Eco*K system were initially cloned by selecting for recombinants that could modify (21). The genes for the *Hha*II, *Eco*K, *Eco*RII, *Eco*RI, *Pst*I, and *Eco*RV systems were cloned by selecting for, or screening for, recombinants that could restrict (22, 69, 112, 138, 153, 184, 230). The restriction phenotype is expressed less consistently, among clones, than is the modification phenotype, and so a variation of the first method is used predominantly. The variation relies on restriction enzymes rather than restricting hosts to select for modified recombinants (138); since its introduction (88, 104, 105, 215), the procedure has enabled over 80 complete and 40 incomplete R-M systems to be cloned (236).

Selection is performed as follows: bacterial DNA fragments are ligated to a vector (usually, a plasmid), and introduced into competent cells (usually, *E. coli*) to prepare a library. The vector must contain sites for the system to be cloned. The transformants are grown for a short period to allow recombinants

carrying the methyltransferase gene to become modified, then DNA is prepared from the entire plasmid population. The DNA is digested with a restriction enzyme (usually, the cognate endonuclease) that cleaves only the unmodified DNA, and molecules that survive digestion are recovered by transformation. Alternatively, selection of the plasmid population is accomplished in vivo, by transformation of a host that restricts the unmodified DNA. Either way selects for *M* gene clones, and because *M* and *R* genes are seldom far apart, recombinants carrying both genes can usually be recovered (for further discussion, see 134).

MULTISTEP CLONING Cloning the genes for an R-M system in one step demands that the host can tolerate the presence of the endonuclease before its DNA becomes completely modified. *E. coli* can cope with this trauma, to a degree. The SOS system participates in the repair of endonuclease-induced damage (82) to the extent that even R^+ M^- derivatives of some systems are viable (67, 237; G. Wilson, unpublished observations). On the other hand, many systems, including *Dde*I, *Bam*HI, *Hgi*DI, *Kpn*I, and *Sty*LTI, can only be cloned in *E. coli* if the cells are modified beforehand. The genes for these systems must be cloned sequentially, *M* first, and then *R* (26, 47, 54, 73, 83). Perhaps the methyltransferases of these systems prefer hemimethylated substrates, and thus are slow to methylate the DNA of naive hosts.

CHARACTERISTICS OF R-M SYSTEMS

R-M systems are classified according to enzyme composition and cofactor requirements, recognition sequence symmetry, and cleavage position (242).

Type I Systems

Type I systems have been found in *E. coli* and its relatives, *Citrobacter* and *Salmonella*. Cleavage occurs at variable distances from the recognition sequence, and so does not produce discrete fragments easily visualized by gel electrophoresis. As a consequence, type I systems are often detected in vivo, by the restriction and modification of phages—a method that fails if the bacteria are phage-resistant. Fewer than a dozen natural type I systems have been discovered. They may well be more common than this figure implies, although no general screen exists. Currently, type I systems are grouped into three families: K (also referred to as IA (15)), which includes *Eco*B, *Eco*D, *Eco*K, *Sty*SB, and *Sty*SP; A (or IB), which includes *Eco*A, *Eco*E and *Cfr*A; and R124 (or IC), which includes *Eco*R124 and *Eco*DXXI. The methyltransferases of the K family are the only prokaryotic methyltransferases known to display a strong preference for hemimethylated DNA (99, 212).

Type I systems are the most complex known (14, 15, 242). They comprise three polypeptides, R (restriction), M (modification), and S (specificity), and

the resulting complex is both an endonuclease and a methyltransferase. The following details derive mainly from *Eco*B and *Eco*K; the other systems have been less well studied. In the presence of AdoMet, the endonuclease binds to the recognition sequence irrespective of its state of methylation (241). ATP may not be essential for discrimination between methylated, hemimethylated, and unmethylated target sequences (30), but nevertheless it affects each response. If the recognition sequence is methylated, ATP stimulates dissociation of the enzyme from the DNA; if the sequence is hemimethylated, ATP stimulates methylation of the other strand; if the sequence is unmethylated, a tightly bound complex forms in the presence of ATP (or a nonhydrolyzable analogue), and following DNA translocation, cleavage occurs at a position remote from the recognition sequence (16, 241). During DNA translocation, the endonuclease remains at the recognition sequence and the DNA loops past it in a process dependent upon ATP hydrolysis (16, 57, 58, 243). The bound complex does not turn over (60, 143).

DNA translocation is thought to occur in both directions from the recognition complex, cleavage occurring when neighboring complexes collide. When cleavage reactions are synchronized, the sizes of the resulting fragments are consistent with cutting occurring midway between adjacent recognition sequences (209). Cleavage apparently results from two nicks (60, 143) and may require two enzyme molecules. The nature of the ends of the fragments is unclear; they are refractory to labeling by the transfer of phosphates to 5' ends (60, 148). It is not known what selective advantage is conferred by cutting at indeterminate distances from the recognition sequence, but it has been proposed that random cleavage serves a critical role in facilitating general recombination (163).

In the absence of the R subunit, the M and S subunits form an ATP-independent methyltransferase (212). Among the known systems, the methylated base is always m^6A. The constraints on the endonuclease—that the M subunit is a required part of the complex, and that AdoMet is necessary for cleavage—have safety aspects that ensure that restriction occurs only if the cell has the wherewithal to modify. This is a feature of other R-M systems too, possibly an adaptation to prevent restriction of the cell's own DNA if AdoMet becomes limiting or if the methyltransferase becomes dysfunctional.

Type I recognition sequences are asymmetric and bipartite; they comprise two components, one of 3 nt, the other of 4 or 5 nt, separated by a nonspecific spacer of 6 to 8 nt (e.g. *Eco*K: $AACN_6GTGC$). Methylation affects one adenine in the top strand of the 5' (3-nt) component, and one adenine in the bottom strand of the 3' (4- or 5-nt) component. The S subunit has two specificity domains (target recognition domains: TRDs), one for recognizing each component (43, 65, 68, 93). There are two M subunits per S subunit (D. T. Dryden, personal communication), so each TRD might associate with a separate M subunit. The TRDs function independently, and their reassortment

by genetic recombination results in enzymes with new specificities (63, 66, 151, 164). Type I systems are the only R-M systems observed to undergo major changes in specificity.

The genes for type I systems are transcribed in the same direction, *M* and *S* as an operon, and *R* separately. The *M* gene precedes *S*, and *R* may be either upstream or downstream (Table 1). The R subunits are large, ~1000 amino acids (aa), and the M and S subunits between 450 and 600 aa each. The *Eco*K endonuclease is believed to be a complex of two R, two M, and one S subunits, and to have a molecular mass of ~440 kd (144). The evolutionary success of such enormous, complicated enzymes may lie in their modularity. The specificity of the system can be changed by changing just one module, the S subunit, whose specificity itself derives from interchangeable TRD modules.

Type II Systems

Type II systems are the simplest and the most numerous. Their specificities are exceedingly varied: about 150 have been identified so far (102, 178). Although most type II enzymes discovered today are duplicates of previously identified specificities ("isoschizomers"), around 20 new specificities are identified each year (62). Some specificities are common: isoschizomers of *Hae*III, *Eco*RII, and *Pst*I, for example, have each been found over 50 times. Other specificities are rare: that of *Xba*I, for example, has been found only once (245). Type II endonucleases and methyltransferases act independently and have simple requirements: the endonucleases require Mg^{2+}, the methyltransferases require AdoMet. The endonucleases range in length from 157 to 576 amino acids (aa), the average being 300 aa; the methyltransferases from 228 to 587 aa, the average being 400 aa (Table 1).

Type II recognition sequences are essentially symmetric. Some sequences are continuous (e.g. GAATTC); others are interrupted. The interruptions can be short (GANTC) or relatively long ($CCAN_9TGG$). The sequences comprise four to eight specific nt, and they vary in base composition (GGCC; TTAATTAA). Some are unique (GTCGAC); others are degenerate (GTYRAC; nt abbreviations are in the footnotes to Table 1).

Symmetric sequences are economical sequences: one protein can react with both strands of the duplex since it appears the same regardless of orientation. Type II endonucleases generally act as homodimers, an association that facilitates the coordinated cleavage of both strands. Type II methyltransferases appear to act as monomers, and methylate the duplex one strand at a time. This is consistent with the hemimethylated nature of their customary substrates.

Cleavage by type II endonucleases occurs symmetrically within the recognition sequence. Some endonucleases cleave on the 5' side of the dyad axis, producing fragments with 5' ss termini of various lengths (e.g.

Table 1 Properties and gene organizations of R-M systems[a]

System / Organism	Type	Specificity[b]	Gene organization[c]: Endonuclease	Gene organization[c]: Methyltransferase	Refs[d]
1. Aligned genes, *R* first					
AccI *Acinetobacter calcoaceticus*	II	GT'MKAC	366	541 m^6A_γ	e
BanI *Bacillus aneurinolyticus*	II	G'GYRCC	354	428 m^5C	137
Bsp6I *Bacillus sphaericus*	II	GC'N_2GC	220	315 m^5C	f
BsuRI *Bacillus subtilis R*	II	GG'CC	576	436 m^5C	105
DdeI *Desulfovibrio desulfuricans*	II	C'TNAG	240	415 m^5C	214
EcoA *Escherichia coli 15T⁻*	I	GAGN$_7$GTCA		489 m^6A_γ; 589	43, 64, g
EcoE *Escherichia coli A58*	I	GAGN$_7$ATGC		490 m^6A_γ; 594	43, 64, g
EcoK *Escherichia coli K-12*	I	AACN$_6$GTGC	1090	529 m^6A_γ; 464	68, 132
EcoRI *Escherichia coli RY13*	II	G'AATTC	277	326 m^6A	69, 153
FnuDI *Fusobacterium nucleatum D*	II	GG'CC	284	344 m^5C	e
HgiCI *Herpetosiphon giganteus Hpg9*	II	G'GYRCC	345	420 m^5C	59
HgiDI *Herpetosiphon giganteus Hpa2*	II	GR'CGYC	359	309 m^5C	54

Table 1 (*Continued*)

System / Organism	Specificity[b] Type	Gene organization[c] Endonuclease	Methyltransferase	Refs[d]
2. Aligned genes, *M* first				
BsuBI *Bacillus subtilis*	II CTGCAG	311	501 $m^6A\gamma$	240
Cfr9I *Citrobacter freundii RFL9*	II C'CCGGG	330	300 $m^4C\beta$	108, f
DpnII *Streptococcus pneumoniae*	II 'GATC	288	256 $m^6A\beta$; 284 $m^6A\alpha$	119, 139
EagI *Enterobacter agglomerans*	II C'GGCCG	301	401 m^5C	h
EcoP1 *Escherichia coli phage P1*	III AGACC	970	646 $m^6A\beta$	85
EcoP15 *Escherichia coli 15T⁻ [p15B]*	III CAGCAG		645 $m^6A\beta$	85
EcoR124 *Escherichia coli [R124]*	I $GAAN_6RTCG$		406; 520 $m^6A\gamma$	164
EcoR124/3 *Escherichia coli [R124/3]*	I $GAAN_7RTCG$	1033	410; 520 $m^6A\gamma$	164
FokI *Flavobacterium okeanokoites*	IIs GGATG 9/13	578	647 $m^6A\alpha$ + $m^6A\alpha$	107, 133
HaeIII *Haemophilus aegyptius*	II GG'CC	317	330 m^5C	199, e
HgaI *Haemophilus gallinarum*	IIs GACGC 5/10	488	358 m^5C; 357 m^5C	e
HgiBI *Herpetosiphon giganteus Hpg5*	II G'GWCC	274	437 m^5C	53

Table 1 *(Continued)*

System Organism	Type	Specificity[b]	Gene organization[c] Endonuclease	Methyltransferase	Refs[d]
HgiGI *Herpetosiphon giganteus Hpa1*	II	GR'CGYC		m^5C	53
HindIII *Haemophilus influenzae Rd*	II	A'AGCTT	300	309 $m^6A\beta$	e
HpaI *Haemophilus parainflu-enzae*	II	GTT'AAC	254	314 $m^6A\beta$	227
MwoI *Methanobact-erium wolfei*	II	$GCN_5'N_2GC$		256 $m^4C\beta$	e
NgoPII *Neisseria gonorrhoea P9*	II	GG'CC	278	341 m^5C	211
NlaIII *Neisseria lactamica*	II	CATG'	230	334 $m^6A\alpha$	116 e
RsrI *Rhodopseudo-monas sphaeroides*	II	G'AATTC	276	319 $m^6A\beta$	97 207 208
SalI *Streptomyces albus*	II	G'TCGAC	315	587 $m^6A\gamma$	1
Sau3AI *Staphylococcus aureus 3A*	II	'GATC	489	412 m^5C	192
StySB *Salmonella typhimurium LT2*	I	$GAGN_6RTAYG$		529 469 $m^6A\gamma$	63 66 g
StySP *Salmonella potsdam*	I	$AACN_6GTRC$		529 463 $m^6A\gamma$	63 65 g
XbaI *Xanthomonas badrii*	II	T'CTAGA	209	423 $m^6A\beta$	e

Table 1 (*Continued*)

System Organism	Specificity[b] Type		Gene organization[c] Endonuclease	Methyltransferase	Refs[d]
XmaI *Xanthomonas malvacaerum*	II	C'CCGGG	333	300 $m^4C\beta$	e
3. Divergent genes					
BamHI *Bacillus amyloliquefaciens*	II	G'GATCC	213	423 $m^4C\beta$	27 224
BglI *Bacillus globigii*	II	GCCN4'NGGC		348 $m^4C\beta$	e
EcoRV *Escherichia coli J62[pLG74]*	II	GAT'ATC	245	298 $m^6A\alpha$	22
Eco72I *Escherichia coli RFL72*	II	CAC'GTG	211	372 m^5C	f
MspI *Moraxella sp.*	II	C'CGG	262	418 m^5C	131
MunI *Micoplasma sp.*	II	C'AATTG	202	229 $m^6A\beta$	f
MvaI *Micrococcus varians RFL19*	II	CC'WGG	259	454 $m^4C\alpha$	108 f
PstI *Providencia stuarti*	II	CTGCA'G	326	507 $m^6A\gamma$	231
PvuII *Proteus vulgaris*	II	CAG'CTG	157	323 $m^4C\beta$	4 217 218
4. Convergent genes					
BcnI *Bacillus centrosporus*	II	CC'SGG	211	$m^4C\alpha$	162 f
BsuFI *Bacillus subtilis*	II	CCGG	395	409 m^5C	94 232

Table 1 (*Continued*)

System Organism	Specificity[b] Type		Gene organization[c] Endonuclease	Methyltransferase	Refs[d]
HgiCII *Herpetosiphon giganteus Hpg9*	II	G'GWCC	273	437 m^5C	59
HgiEI *Herpetosiphon giganteus Hpg24*	II	G'GWCC	274	437 m^5C	59
HhaII *Haemophilus haemolyticus*	II	G'ANTC	227	228 $m^6A\beta$	190
HincII *Haemophilus influenzae Rc*	II	GTY'RAC	257	518 $m^6A\gamma$	86 228
HinfI *Haemophilus influenzae Rf*	II	G'ANTC	262	358 $m^6A\beta$	36
MboII *Moraxella bovis*	IIs	GAAGA 8/7	416	260 $m^6A\beta$	20
NcoI *Nocardia corallina*	II	C'CATGG	287	422	e
NdeI *Neisseria denitrificans*	II	CA'TATG	368	478 $m^6A\alpha$	229
NgoMI *Neisseria gonorrhoea MS11*	II	G'CCGGC	286	313 m^5C	i
PaeR7I *Pseudomonas aeruginosa pMG7*	II	C'TCGAG	246	531 $m^6A\gamma$	219
Sau96I *Staphylococcus aureus PS96*	II	G'GNCC	261	430 m^5C	213 e
TaqI *Thermus aquaticus YT1*	II	T'CGA	263	419 $m^6A\gamma$	198 j
TthHB8I *Thermus thermophilus HB8*	II	T'CGA	263	427 $m^6A\gamma$	246 j

CC'SGG; AT'CGAT; G'ANTC; A'AGCTT; 'CCWGG); others cleave in the center, producing flush termini (AG'CT); yet others cleave on the 3' side, producing 3' ss termini (CGAT'CG; CTGCA'G). Methylation also occurs symmetrically within the recognition sequence. Some methyltransferases produce m^4C; others produce m^5C; and yet others produce m^6A. No invariant relationship exists between the positions of cleavage and methylation, but the methylated base often flanks the cleavage site. m^5C is a common modification among mesophiles, but not among thermophiles (55, 56). One explanation for this difference is that at elevated temperatures, increased deamination of m^5C is detrimental (55).

Type II restriction and modification enzymes possess separate TRDs. The specificities of cognate enzymes are the same, or nearly so. This coincidence is economical, but not essential: methylation of sequences that include the endonuclease's sequence as a subset also achieves protection. Perhaps during the evolution of type II systems, the methyltransferases with which endonucleases first associate are of broader specificity, and once the partnership has formed, the specificities of the enzymes converge, the methyltransferases becoming more specific, or the endonucleases less specific. The *C*viJI system has yet to achieve specificity coincidence: the endonuclease recognizes RGCY, but the methyltransferase recognizes RGCB (194), and perhaps even VGCB (M. Nelson & J. Van Etten, personal communication). This is also true for the type IIs system *Hga*I: the endonuclease recognizes GACGC, but both methyltransferases recognize GACGY, albeit less efficiently than they recognize just GACGC (J. Barsomian and G. Wilson, unpublished observations).

Given the matching specificities of cognate endonucleases and methyltransferases, it might be supposed that both had a common ancestor. This appears not to be the case: 44 type II systems have now been analyzed and no similarities indicative of a common ancestry have been seen. Nonetheless, cognate enzymes might be expected to have at least similar TRDs, since they recognize identical sequences. Again, comparisons have failed to reveal any such similarities (37). One explanation is that the endonucleases and methyltransferases employ different strategies to recognize their targets because they act in different multimeric forms. Thus, endonuclease subunits need recognize only one half of the sequence, whereas methyltransferases must recognize the entire sequence. Possession of separate TRDs probably restrains type II systems from adopting new specificities during evolution, because a change in the specificity of either component would be detrimental without a corresponding change in the specificity of the other component.

Type IIs Systems

Type IIs R-M systems (shifted cleavage) resemble type II systems in that the endonucleases and methyltransferase act independently and have similar re-

Table 1 Properties and gene organizations of R-M systems[a]

System / Organism	Type	Specificity[b]	Gene organization[c] Endonuclease	Gene organization[c] Methyltransferase	Refs[d]
CfrlOI *Citrobacter freundii RFL 10*	II	R'$\underline{C}$CGGY		343; m^5C	f
Eco57I *Escherichia coli RFL57*	IIs	C$\underline{T}$GA$\underline{A}$G 16/14	993; R + m^6A_γ	544; m^6A_γ	f
EcoRII *Escherichia coli (N3)*	II	'C$\underline{C}$WGG	404	477; m^5C	13, 111, 205
SfiI *Streptomyces fimbriatus*	II	GGCCN$_4$'NGGCC	269	421; m^4C_β	e
SinI *Salmonella infantis*	II	G'GW$\underline{C}$C	230	461; m^5C	96
SmaI *Serratia marcescens*	II	C$\underline{C}$C'GGG	247	292; m^4C_α	80

[a]Only systems that have been substantially sequenced are listed.

[b]One strand only of the recognition sequence is shown, printed 5' to 3'. Strings of unspecified nt are depicted N_i, where i specifies the number of nts. Standard abbreviations are:

R: A or G	M: A or C	S: C or G	H: A or Y	V: C or R
Y: C or T	K: G or T	W: A or T	B: G or Y	D: T or R

Cleavage positions are indicated by apostrophes (type II systems), and by numerals (other systems); nts that probably become methylated are indicated by *underlining*. For type II systems, the cleavage and methylation positions on only one strand are shown, the complementary strand being identical. For the other systems, the cleavage and methylation positions on both strands are shown; $\underline{A}$ and $\underline{C}$ signify that that base is methylated, $\underline{T}$ or $\underline{G}$ signify that the complementary nt is methylated.

[c]Genes are depicted as arrows; the directions indicate transcriptional orientations. Broad arrows identify genes that have been substantially sequenced, thin arrows, genes that have not been sequenced. Filled arrows signify *R* genes; open arrows, *M* genes; cross-hatched arrows, *S* genes; stippled arrows, potential *C* genes. The diagrams are oriented so that the *R* genes are always on the left. The genes are drawn to scale, and their lengths (in codons) are printed above the arrows. The intervals between the genes are drawn to scale; they are depicted by lines, or by the absence of a gap. If the length of the interval is not known, the genes are shown separated by a gap.

The probable methylation products are indicated below the methyltransferase genes. The organizational classes to which the methyltransferases belong are indicated by Greek subscripts. The classes (and their conserved motif orders) are: m^6A_α (F-G-G then DPPY); m^6A_β (DPPY then F-G-G); m^6A_γ (LEP-G-G then NPPY); m^4C_α (F-G-G then SPPY); and m^4C_β (SPPY then F-G-G). Greek subscripts are used for convenience; a convention for distinguishing the methyltransferase classes has yet to be agreed upon.

[d]References that relate to the nt sequences of the genes are given. References that deal with the discovery and characterization of the enzymes (102, 141, 178), or with the cloning of the genes (236) can be found elsewhere.

[e]J. Barsomian, J. Benner, S. Chen, E. Van Cott, K. Lunnen, M. Meda, L. Moran, R. Morgan, D. Nwankwo, F. Perler, B. Slatko, B-H. Zhang and G. Wilson, unpublished observations

[f]D. Butkiene, V. Butkus, A. Lubys, S. Menkevicius, P. Povilonis, R. Vaisvila, N. Zareckaja, and A. Janulaitis, personal communication.

[g]G. Cowan, A. Daniel, J. Kelleher, N. Murray, unpublished observations.

[h]L. Sznyter, L. Moran, B. Slatko, and J. Brooks, personal communication.

[i]R. Chien and D. Stein, personal communication.

[j]F. Barany, personal communication.

functions in the hemimethylated mode, or perhaps a second methyltransferase specific for the other strand remains to be identified.

Several type IIs systems are irregular and might better be classified as new types. The *Eco*57I (CTGAAG) system, for example, comprises a joint endonuclease-methyltransferase and a separate methyltransferase. The former behaves as if it were a type-IIs endonuclease fused to one of a pair of type IIs methyltransferases: it cleaves to one side of the sequence and methylates the sequence on one strand. Cleavage is stimulated by AdoMet, but not by ATP. The separate methyltransferase behaves like a type-II enzyme and methylates both strands (89, 159). The *Bcg*I (CGAN$_6$TGC) endonuclease behaves as though its recognition sequence were symmetrical. It cleaves outside of the sequence, on both sides, excising a 34-bp fragment that contains the recognition sequence (H-M. Kong & I. Schildkraut, personal communication). Although AdoMet is required for cleavage, *Bcg*I has no intrinsic methyltransferase activity.

Type IIs systems are an order of magnitude less common than type II systems. Given their complexity—twice that of type II systems—it is not obvious why they exist at all. One explanation may be that asymmetric sequences are more abundant than symmetric sequences: of all the sequences four to six nt in length, over 95% are asymmetric.

Type III Systems

Four type III specificities are known (178). The M subunit confers sequence specificity for both restriction and modification. On its own, the M subunit acts as a methyltransferase; as a heterodimer with the R subunit, it acts as a methyltransferase and endonuclease. Methylation requires AdoMet. Cleavage requires ATP and is stimulated by AdoMet. In the presence of both cofactors, the heterodimer competes with itself by modifying and restricting in the same reaction (15). As a consequence, digestions do not go to completion. Type III enzymes are large, roughly the size of their type I counterparts (85).

Type III recognition sequences are asymmetric, uninterrupted, and five to six nt in length. Cleavage occurs on one side of the recognition sequence, approximately 25 nt away. Methylation takes place on only one strand, in apparent violation of the rule that both strands must be modified to protect daughter duplexes from cleavage. However, type III endonucleases require the presence of two recognition sites in the DNA, in opposite orientations. Cleavage occurs only if neither site is modified, and since one site or the other remains modified following replication, protection is preserved (114). It is not clear how close two sites must be to participate in cleavage.

Extreme strand-bias is seen in the DNA of phage T7 with respect to *Eco*P15 sites (CAGCAG). T7 is not restricted by *Eco*P15, yet its DNA contains 36 CAGCAG sequences. The sequences all occur in the same orientation,

quirements. In other respects, however, they represent a different kind of system altogether. Approximately 30 type IIs specificities have been discovered, but few have been well characterized (216).

Type IIs recognition sequences are asymmetric and uninterrupted; they range in length between four and seven nts. The endonucleases are roughly twice as large as type II endonucleases and appear to act as monomers. Cleavage occurs on only one side of the recognition sequence, at a defined distance—under 20 bp—from the sequence. Cleavage of the two strands is usually staggered by one to four nt (178, 216). Type IIs endonucleases probably possess separate domains for recognition and cleavage; perhaps one recognition domain and two cleavage domains—one for each strand.

In the *Fok*I (GGATG) and *Hga*I (GACGC) systems, modification is accomplished by a pair of methyltransferases, one methyltransferase for each strand. The *Hga*I methyltransferases are independent proteins, encoded by adjacent genes (J. Barsomian, G. Wilson, unpublished observations). In contrast, the *Fok*I methyltransferases are fused together in a single, bifunctional enzyme (107, 121, 133, 210). Whereas the *Hga*I methyltransferases closely resemble one other (45% identity)[1], and probably diverged from a common ancestor, the *Fok*I methyltransferases are dissimilar (17%) and might have arisen independently (116, 133).

In the *Bbv*I system (GCAGC), the modification enzyme is a single methyltransferase that modifies both strands symmetrically (J. Barsomian, G. Wilson, unpublished observations). Evidently, M·*Bbv*I recognizes the sequence as symmetric (i.e. GCWGC), whereas R·*Bbv*I recognizes it as asymmetric.

In aa sequence, type IIs methyltransferases are much like type II enzymes. Probably they differ only in that they bind to asymmetric, as opposed to symmetric, nt sequences. Type II methyltransferases modify both strands of the sequence because they can bind it in both orientations, methylating first one strand, and then the other. Type IIs methyltransferases, on the other hand, modify only one strand because they can bind to the sequence in only one orientation.

A corollary of the dual nature of type IIs modification is that in some systems different nucleotides are methylated on each strand: an A on one strand, and a C on the other. This is true for *Alw*26I (GTCTC), *Eco*31I (GGTCTC), and *Esp*3I (CGTCTC) (18). It might also be true for systems like *Mnl*I (CCTC) that have only pyrimidines on one strand and purines on the other. The *Mbo*II system (GAAGA) appears to code for only a single, m^{6}A-type, methyltransferase (20, 142). Perhaps *Mbo*II is unusual and only

[1]Throughout this article, % identity refers to the value computed by the UWGCG program "GAP", scoring perfect amino acid matches only, and run with a gap penalty = 3, and a gap length penalty = 0.1.

product of a single gene immediately upstream of *Eco*K. The purified Mrr protein degrades the m^{5}C-containing DNA of phage XP12 from *Xanthomonas oryzae* with little apparent specificity (226).

EXCISION REPAIR SYSTEMS Restriction can be mimicked by DNA repair enzymes. The uracil-DNA glycosylase (Ung) of *E. coli* (225), for example, reduces the plating efficiency of phages that contain uracil in partial replacement for thymine (233). Because of base-pairing promiscuity, uracil is mutagenic in DNA, and Ung functions in concert with dUTPase (Dut) to exclude it from the hereditary material. Phage grown in Dut$^-$, Ung$^-$ cells incorporate uracil into their DNA and appear to be restricted when plated on Ung$^+$ hosts (233). "Restriction" probably results from double-stranded gaps introduced into the DNA by the AP endonuclease, during repair of the spaces left following removal of the uracils (reviews: 185, 220).

Mcr and Mrr might be more like repair enzymes than restriction enzymes. Deamination of m^{5}C creates thymine, which, if ignored, is mutagenic because it causes C · G to T · A transitions. Preventing this requires the excision of either the mismatched thymines (91) or the methylated cytosines. Mcr and Mrr are active on m^{5}C-containing DNA and may possibly fulfill one of these roles. An excision repair function in yeast has been implicated in the removal of m^{5}C from DNA (61) and an hm^{5}C-specific glycosylase has been identified in calf thymus (33).

OTHER ORGANISMS Functions comparable to Mcr and Mrr have been identified in microorganisms other than *E. coli*. The mycoplasma *Acholeplasma laidawii* JA1 restricts DNA if it contains m^{5}C, regardless of the sequence context (196, 197). *Streptomyces avermitilis* restricts DNA if it contains m^{6}A or m^{5}C in various sequences (136). *Bacillus thuringiensis* behaves similarly (135). Together with Mcr and Mrr, these systems might represent the first examples of a ubiquitous family of protective mechanisms that operate alongside the more familiar restriction-modification systems.

Regulation of Expression

It is not clear to what extent the expression of *R* and *M* genes is regulated. Our ignorance stems partly from neglect and partly from a lack of familiarity with the diverse bacteria in which R-M systems occur. Cloning the genes in *E. coli* makes the systems more accessible, but it also disturbs their expression, rendering conclusions about their regulation uncertain. In the type I systems, the *R* genes have different promoters from the *M* and *S* genes, providing the opportunity for independent regulation. This applies also to those type II systems in which the genes have opposite orientations. The genes for most type II systems have the same orientations, however, and usually abut so closely that they are unlikely to have separate promoters. This arrangement

suggesting a strong evolutionary aversion to inverted sites, presumably due to past restriction by *Eco*P15, or an isoschizomer (114, 191).

Other System Types

More types of R-M systems will likely be discovered, and novel variants of existing systems recognized. A potential variant is the Dcm-VSP repair system of *E. coli*. "Very short patch" repair converts mismatched T·G base pairs to C·G base pairs, specifically in the sequence C*T*WGG/CCWGG (91, 129). One way this mismatch arises is by deamination of m^5C in the sequence Cm^5CWGG. This sequence first becomes methylated by the action of Dcm, and hence VSP functions to maintain Dcm sites. The gene for VSP, *vsr*, is adjacent to the *dcm* gene (204); the matching specificities of Dcm and VSP and the close proximity of their genes are features reminiscent of R-M systems. Dcm-VSP may have diverged from an R-M system akin to *Eco*RII (CCWGG). VSP and R·*Eco*RII are dissimilar, but Dcm and M·*Eco*RII are closely similar (65% identity) (13, 74, 111, 204, 205).

Modification-Requiring Systems

MCR AND MRR The systems discussed so far are all blocked by methylation. Certain systems behave in the opposite way. RglA and RglB were the original names for two such systems in *E. coli* that *r*estrict *gl*ucoseless variants of the T-even phages, T2, T4, and T6 (76, 174). Wild-type T-even phage have glucosylated 5-hydroxymethylcytosine (glu-hm^5C) in their DNA instead of cytosine. They are immune to restriction by the Rgl systems, and to restriction by many R-M systems as well (173). Phage mutants that contain hydroxymethylcytosine (hm^5C) instead of glu-hm^5C are restricted by RglA and RglB; but since hm^5C is not a common modification, it was not clear why *E. coli* possessed the Rgl systems nor why the phages used glu-hm^5C for protection, instead of the traditional m^5C.

An explanation emerged when it was discovered that the Rgl systems restrict not only DNA that contains hm^5C, but also DNA that contains m^5C (155, 171), and in the case of RglB, DNA that contains m^4C too (19). The systems are now referred to as McrA and McrBC (for *m*odified *c*ytosine *r*estriction) to more accurately reflect their functions (168). A third system in *E. coli*, Mrr (*m*odified adenine *r*ecognition and *r*estriction), restricts DNA-bearing foreign modifications, including some involving m^6A (81, 226) and at least one involving m^5C (100). The sequence-specificities of the Mcr and Mrr systems are not known.

McrA is encoded by a prophage-like element, e14 (172). McrBC is encoded by adjacent genes (*mcrB* and *mcrC*) immediately downstream of the *Eco*K genes (170, 182). A third *mcr* gene, *mcrD*, is adjacent to *mcrC* and may play a regulatory role (156). The McrBC complex cleaves suitably m^5C-modified DNA in a reaction that requires GTP (48). Mrr appears to be the

identical effect on restriction and modification. Consequently, type I systems have greater opportunities for evolutionary diversification than do type II or type IIs systems, because a change in the specificity of one activity is automatically accompanied by the same change in the specificity of the other activity. Type I systems can be divided into families whose members have nearly identical M and R subunits but different S subunits. All type I systems have similar enzymatic organization and co-factor requirements, and the sequences recognized are always bipartite.

Three families of type I systems are currently known. The archetypal member of the first family is *Eco*K, from *E. coli* K-12. Evidence for its relatedness to other systems came first from complementation tests that demonstrated that subunits could be exchanged between systems of different specificity (28), and subsequently from molecular comparisons involving nucleic acid hybridization and antibody cross-reactivity (44, 149). The K family includes *Eco*B, *Eco*D, and *Eco*K, as well as *Sty*SB and *Sty*SP from *Salmonella typhimurium*. The genes for these systems have the same chromosomal locations, probably a consequence of the common ancestry between *E. coli* and *S. typhimurium*. However, the genes for members of the A family of type I systems (*Eco*A, *Eco*E, and *Cfr*A) also have the same location, yet their genes differ markedly from those of the K family. The genes for both members of the R124 family (*Eco*R124 and *Eco*DXXI) are plasmid-borne, although very similar gene sequences have been identified in the chromosome of *E. coli* K-12 (130).

COMPARISONS BETWEEN TYPE I FAMILIES The nt sequences of the *R*, *M*, and *S* genes for at least one member of each family are now known. The gene order in the chromosomal families (A and K) is *R*, *M*, and *S* and in the plasmid-borne family, *M*, *S*, and *R* (Table 1). The R polypeptides in all three families contain a motif resembling an ATP-binding site. The M polypeptides contain motifs common to DNA adenine-methyltransferases, including the F-G-G motif that might form the AdoMet-binding site and the NPPY motif that might form part of the active site (108, 125, 201). The S polypeptides contain repeats suggestive of gene duplication, a feature that may correlate with the bipartite structure of the recognition sequence (93).

The M polypeptides from the K family share about 30% identity with those from the A family; both share about 22% identity with the M polypeptide from *Eco*R124. The corresponding values for the R polypeptides of *Eco*K, *Eco*E, and *Eco*R124 are 22% and 15–18%. Some similarity is evident between M genes at the nt level, particularly between members of the K and A families. These families are, therefore, more similar to each other than to the R124-family. Within two short repeats, the S polypeptides from all three families share some similarity (93). Outside of these repeats, however, there is little or no similarity; overall, the S polypeptides share only 18% identity.

could enable coordinated expression of the genes; the gene orders seem unimportant: in some systems the *R* gene comes first and in others, the *M* gene (Table 1).

A priori, we expect the expression of *R* genes to be regulated so that the endonuclease level would parallel the extent of modification: if the DNA were modified, the *R* gene would be expressed, and if the DNA were unmodified, the *R* gene would be silent. This would reduce the danger to the cell if events temporarily interfered with modification. It would also facilitate the establishment of R-M systems in naive cells, a factor of importance to temperate phage systems and to plasmid-borne systems that might regularly encounter new hosts.

In several R-M systems, recognition sites for the cognate enzymes occur in positions that could influence transcription. Thus, two *Pae*R7I sites occur before the *paeR7IM* gene (219); two *Fok*I sites occur before the *fokIR* gene (133); two MboII sites occur within the *MboIIR* gene (20); and numerous *Taq*I sites occur within the taq*IR* gene (198). These strategic locations suggest that the sites form binding sites for regulatory factors whose affinities would vary according to whether the sites were methylated or unmethylated (133, 198). The factors might be repressors or inducers, or the methyltransferase or endonuclease proteins themselves. As yet, no evidence supports this model. Moreover, since such sites are lacking in the majority of systems, even if the mechanism were to operate in the examples above, it would not be universal.

Open reading frames (ORFs) of ~100 codons precede the *R* genes in several systems in which the gene orientations differ, including *Bam*HI, *Eco*RV, *Pvu*II and *Sma*I (217); and *Eco*72I, *Mun*I, and *Mva*I (A. Janulaitis, personal communication). The *Bam*HI and *Pvu*II ORFs, termed *C* genes (control), stimulate endonuclease expression in *cis* and in *trans* (217). The *Bam*HI ORF is system-specific; it functions both in *E. coli* and *Bacillus* and reduces methyltransferase expression in addition to increasing endonuclease expression (27; P. Nathan, C. Ives, J. Brooks, personal communication). The *Pvu*II ORF appears to affect only endonuclease expression (217). The derived amino acid sequences for the *Bam*HI, *Eco*RV, *Pvu*II, and *Sma*I C proteins contain a highly conserved sequence that resembles the helix-turn-helix motif of DNA-binding proteins (27, 217). This suggests that the C proteins act as transcriptional repressors and inducers, comparable to the lambdoid phage repressors.

CONTRASTS AND COMPARISONS AMONG R-M SYSTEMS

Type I Systems

FAMILIES OF TYPE I SYSTEMS The specificity of type I R-M systems is conferred by the specificity subunit, S. Changes in the specificity of S have an

*Sty*SP (K-family) recognize the same 5' sequence component (AAC). Their amino variable regions show 90% identity (65); their carboxy variable regions are dissimilar, even though they recognize quite similar 3' sequences (GTGC and GTRC) (66). Likewise, *Eco*A and *Eco*E (A family) recognize the same 5' sequence component (GAG), and their amino variable regions show 80% identity. *Sty*SB recognizes the same 5' sequence as *Eco*A and *Eco*E, and although it belongs to a different family, its amino variable region is nevertheless closely similar (44% identity).

While the variable regions equate with TRDs, no information implies which amino acids interact with which nucleotides. There is no evidence for helix-turn-helix motifs in S polypeptides. The 5' component of the *Eco*R124 recognition sequence, GAA, is the same as the half-sequence of *Eco*RI. Some aa sequence similarity has been noted between the *Eco*R124 S polypeptide and the section of the *Eco*RI endonuclease responsible for sequence recognition (164).

CHANGES IN SEQUENCE SPECIFICITY Changes in type I specificities have been observed in the laboratory. Recombination between the *S* genes of *Sty*SB and *Sty*SP led to the reassortment of TRDs (43, 63, 66, 151). This first occurred serendipitously—and was correctly diagnosed—before the organization of the S polypeptide was known (29). The similarities between the 5'-TRDs of *Eco*K and *Sty*SP and of *Eco*A and *Eco*E suggest that reassortment also occurs in the wild.

A second kind of specificity change may also have arisen by recombination. Unequal crossing-over within a natural 12-bp duplication in the *S* gene of *Eco*R124 generated a triplication that resulted in the insertion of four extra aa (71, 164). The insertion caused the interval between the two components of the recognition sequence to be lengthened by one nt: the specificity of *Eco*R124 is GAAN$_6$RTCG; that of the mutant (*Eco*R124/3) is GAAN$_7$RTCG (165).

The properties of *Eco*R124/3 suggest a degree of flexibility in the interdomain linker (71) and reveal a further dimension to the evolutionary versatility of the S polypeptide. Specificity can change not only by reassortment of TRDs, but also by altering the spacing between them. Mutation so far has failed to change the specificity of individual TRDs. However, mutants from *E. coli* K-12 have recently been isolated that behave as if the methyltransferase has lost its specificity for hemimethylated DNA and is now active on unmethylated recognition sequences (99).

Type II Systems

Type II R-M and type I systems are organized quite differently and their proteins share few similarities. Over 60 type II systems have been characterized; based on aa sequence, the methyltransferases can be grouped into

The similarities between type I families, though modest, are sufficient to consider the families to have a common origin. The differences between them imply a strong selective pressure for diversification. This pressure is presumed to be for new specificities, and hence for changes in *S*, but it is not known whether diversification of *M* is necessary to achieve maximal variation of specificity.

COMPARISONS WITHIN TYPE I FAMILIES The *hsdM* genes for *Eco*K, *Eco*B, *Sty*SB, *Sty*SP, *Eco*A, and *Eco*E have been sequenced (see Table 1 for citations), allowing comparisons within both the K and the A families. The M subunits of *Eco*K, *Eco*B, *Sty*SB, *Sty*SP are almost indistinguishable (95% identity), as are those for *Eco*A and *Eco*E (90% identity). The evolutionary separation between organisms can be estimated by analyzing the silent (synonymous) nt changes that have accumulated in their genes. The values obtained when *E. coli* and *Salmonella* genes of the K family are compared are similar to those obtained when other pairs of homologous genes are compared between the two species (194b). In contrast, the values obtained when the comparisons are confined to *E. coli*, however, are abnormally high; silent changes between the *M* genes of *Eco*A and *Eco*E, for example, are as high as those that occur between species, the possible result of genetic exchange between species or of a selective pressure in favor of divergence.

Early interest in type I R-M systems focused on the specificity of their interactions with DNA. Consequently, comparisons within families were confined to S polypeptides. It was argued that related S polypeptides might differ in only minor ways that could identify the amino acids involved in DNA recognition. The S polypeptides of the K-family were found to share about 40% identity. The similarities were due to the presence of two conserved sequences, each following a longer variable region of around 150 aa. The variable regions are no more alike than sequences chosen at random (66). It was suggested that the variable regions were TRDs, and that each recognized one component of the bipartite recognition sequence (68).

Support for this proposal came from hybrid S polypeptides generated by crossing-over between the central conserved regions of different S genes. Reciprocal hybrids between *Sty*SB (GAGN$_6$RTAYG) and *Sty*SP (AACN$_6$GTRC) displayed the predicted novel specificities: GAGN$_6$GTRC (SJ) and AACN$_6$RTAYG (SQ) (63, 66, 151). Minor differences within the central conserved sequence of one recombinant S polypeptide were removed so that the amino variable region derived from *Sty*SP, and the rest of the protein derived from *Sty*SB (43). Its specificity confirmed that the amino variable region determines the 5' component of the recognition sequence and, by inference, that the carboxy variable region determines the 3' component.

Amino acid sequence comparisons among S polypeptides of K- and A-family members also correlate the variable regions with TRDs. *Eco*K and

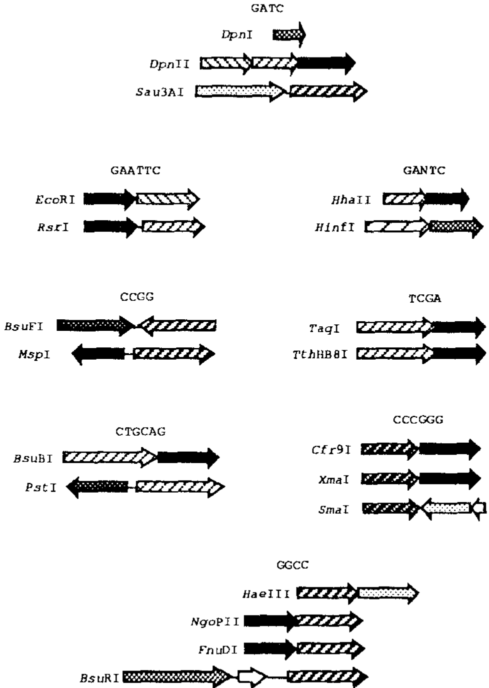

Figure 1 Comparisons between isoschizomeric type II systems. The genes for systems of similar specificity are grouped and shaded to indicate degrees of similarity. *R* genes are stippled or solid; *M* genes are cross-hatched. Within each group, similar shading or cross-hatching signifies aa sequence similarity. (See text for descriptions.)

PARTIALLY RELATED SYSTEMS Several isoschizomeric systems are similar in one respect but dissimilar in others. *Eco*RI and *Rsr*I (GAATTC) have similar gene organizations and their enzymes catalyze identical reactions. The endonucleases are very similar (49% identity) but the methyltransferases are not (16% identity) (97, 207). The systems must have arisen independently; the endonucleases probably had a common origin but they associated with unrelated methyltransferases (Figure 1, center left).

*Bsu*FI and *Msp*I (CCGG) have similar components, but their gene organizations differ (94, 131). The methyltransferases are closely similar (46% identity) and the endonucleases somewhat similar (94, 232). A comparable situation holds for *Bsu*BI and *Pst*I (CTGCAG) (231, 240). In these examples, the systems must have arisen independently, but their component enzymes may have had common origins (Figure 1, center left).

six discrete classes, but the endonucleases are heterogeneous and defy grouping. Endonucleases that catalyze identical reactions sometimes have similar aa sequences; those that catalyze different reactions always have dissimilar aa sequences. Among the methyltransferases, similarities also depend to an extent upon specificity, but there are additional, skeletonlike similarities that depend upon enzymatic reaction mechanisms: enzymes that methylate in the same way have similar skeletons; enzymes that methylate in different ways have different skeletons.

Three aspects of type II systems can be evaluated when assessing their relatedness: gene organization; endonuclease sequence; and methyltransferase sequence. Among nonisoschizomeric systems, those comprising enzymes of different sequence-specificity, the gene organizations are sometimes similar (Table 1) but the endonucleases always differ, and the methyltransferases usually differ too. This suggests that nonisoschizomeric systems arose independently and converged towards similar gene organizations. This seems to be true for some isoschizomeric systems too, but other isoschizomeric systems evidently are related—either indirectly, in that their components are similar but their organizations differ—or directly, in that all three aspects of the systems match.

UNRELATED SYSTEMS The isoschizomeric *Dpn*I, *Dpn*II, and *Sau*3AI systems (recognition sequence GATC) appear to be unrelated. The gene organizations of these systems differ, as do the enzymes (Figure 1, top). *Dpn*I is an unaccompanied endonuclease; it cleaves in the center of the recognition sequence if the adenines are methylated (118). *Dpn*II comprises an endonuclease and two methyltransferases. The endonuclease cleaves on the 5' side of the sequence if the adenines are unmethylated (117); the methyltransferases methylate the adenines but differ from each other (35, 119, 139). *Sau*3AI comprises an endonuclease and one methyltransferase (192). The endonuclease cleaves at the same position as R·*Dpn*I, whether the adenines are methylated or not; the methyltransferase methylates the cytosines. The aa sequences of the three endonucleases are dissimilar, as are those of the methyltransferases (119, 192).

DISTANTLY RELATED SYSTEMS Some isoschizomeric systems are marginally similar and it is unclear whether they diverged from a very distant ancestor, or whether they evolved independently. *Hha*II and *Hin*fI (recognition sequence: GANTC), for example, have similar gene organizations (36, 190); the endonucleases catalyze identical reactions and the methyltransferases probably do too. The endonucleases share 19% identity and the methyltransferases 18% (37). The disparity between these two systems is surprising, considering both come from *Haemophilus* (Figure 1, center right).

*Mwo*I (GCNNNNN'NNGC), *Bgl*I GCCNNNN'NGGC) and *Sfi*I (GGCCNNNN'NGGCC) (K. Lunnen, E. Van Cott, G. Wilson, unpublished observations). The lack of similarity among endonucleases suggests that they might have evolved independently. If this is indeed the case, then a minimum estimate for the number of independent endonucleases that have evolved is ~1000. Thus, 150 different type II specificities are currently known, and probably an equal number remain undiscovered. It is reasonable to assume that each specificity evolved more than once; hence the figure of 1000. Alternatively the lack of similarity among endonucleases might reflect a strong selective advantage in favor of divergence that has obscured the evidence of common origins. In spite of their diversity, endonucleases might be fairly homogeneous at the three-dimensional level. Several have been crystallized, including *Eco*RI, *Eco*RV, *Fok*I, *Hha*II and *Taq*I (6, 38, 45, 106, 181, 238), and *Bam*HI, *Hin*dIII, *Nde*I, *Pae*R7I, *Pvu*II, and *Sfi*I (49), but only for *Eco*RI has a structure been published (103).

Methyltransferases

The genes for approximately 90 methyltransferases have been sequenced. The proteins display similarities that depend mainly upon the nature of their methylated products. They can be grouped into three major classes, corresponding to the three modifications, m^5C, m^4C, and m^6A; the latter two classes can be further subdivided by aa sequence organization. Methyltransferases share only one aa sequence motif in common (F-G-G); the motif varies somewhat from class to class, but it can be identified in nearly every enzyme. The ubiquity of the motif makes it a candidate for the AdoMet-binding domain (108, 125, 201).

5-CYTOSINE METHYLTRANSFERASES The m^5C-MTases form a relatively homogeneous class of 40 members. The amino-termini of the m^5C-MTases are variable. The F-G-G motif is the first common sequence; it may be only a few aa from the amino-terminus, or as many as 100 aa. The m^5C-MTases share around ten common motifs that occur in invariant order (127, 161). The least variable motif is GFPCQ, the cysteine of which forms a transient covalent bond with carbon-6 of cytosine during an intermediate stage in the transmethylation reaction (158, 239). The ten common motifs also occur in eukaryotic methyltransferases: in the human CpG methyltransferase (D. Hornby, personal communication), and in the final 500 aa of the mouse CpG methyltransferase (11; T. Bestor, personal communication).

Towards the carboxy-termini of the m^5C-MTases between conserved regions VIII and IX is a variable region of 80–120 aa. This region is thought to be the TRD. Evidence for this comes from elegant studies of *Bacillus* phage methyltransferases (5, 8, 222, 235; see below) and from aa sequence com-

CLOSELY RELATED SYSTEMS A few isoschizomeric systems are so similar that they are clearly homologous i.e. descendants of the same ancestral system. *Taq*I and *Tth*HB8I (TCGA), both from *Thermus*, are alike; the gene organizations are the same and the enzymes catalyze identical reactions. The methyltransferases are very similar (81% identity), as are the endonucleases (75% identity) (198, 246; F. Barany, personal communication).

The same holds for *Cfr*9I and *Xma*I (CCCGGG). The genes are similarly organized; the endonucleases catalyze the same reactions, and the methyltransferases probably do too. The methyltransferases are almost indistinguishable (81% identity), as are the endonucleases (75% identity) (108; A. Lubys, S. Menkevicius, V. Butkus & A. Janulaitis, personal communication, K. Lunnen, G. Wilson, unpublished observations). In contrast, the *Sma*I system must have arisen independently: the genes for *Sma*I converge, while those for *Cfr*9I and *Xma*I are aligned (80; Figure 1, center right). And, whereas the *Sma*I and *Cfr*9I methyltransferases catalyze identical reactions and are very similar (55% identity), the endonucleases cleave at different positions and are dissimilar (17% identity; Figure 1, center right).

*Bsu*RI, *Fnu*DII, *Hae*III, and *Ngo*PII (GGCC) exemplify isoschizomeric systems that appear to have evolved three times. The methyltransferases of all four systems are similar, but not the endonucleases. The *Fnu*DI and *Ngo*PII systems are homologous; their genes are similarly organized, and the methyltransferases (64% identity) and endonucleases (59% identity) are very similar (211; B.-H. Zhang, G. Wilson, unpublished observations). The *Hae*III system must have arisen independently, since its gene organization is different and its endonuclease is dissimilar (B.-H. Zhang, G. Wilson, unpublished observations; Figure 1, bottom). The same is true for the *Bsu*RI system, the *R* gene of which is unusually large and is separated from the *M* gene by an unusually large interval (105). The methyltransferases of all four systems probably had a common origin, but the endonucleases appear to have had three different origins.

Type II Endonucleases

The notion that DNA specificity might be understood by comparing the aa sequences of type II endonucleases is simplistic. Endonucleases of different specificity differ so thoroughly that comparisons among them are uninformative. No similarities have been reported between endonucleases that recognize different sequences, nor between endonucleases that recognize the same sequence but cleave in different positions. Similarities are seen occasionally among endonucleases that catalyze identical reactions, but these are usually so close as to signify common ancestry. No obvious similarities are evident between endonucleases that recognize related but nonidentical sequences, for example *Bcn*I (C'CSGG) and *Mva*I (C'CWGG) (32); or

It might be possible to interconvert m^4C- and m^6A-MTases by mutation or domain-swapping. Motif I in the m^4C-MTases is usually SPPY (108).

HETERODIMERIC METHYLTRANSFERASES Type II methyltransferases are usually monomers, except for *Aqu*I (CYCGRG), which comprises two polypeptides of 248 and 139 aa (95). The polypeptides are encoded by adjacent, similarly oriented genes, the larger gene preceding the smaller. The large polypeptide contains sequence motifs I through VIII common to m^5C-MTases, and the small subunit contains sequence motifs IX and X. The junction between the genes interrupts the section of the enzyme corresponding to the variable region of monomeric m^5C-MTases.

The genes for M·*Aqu*I resemble an m^5C-MTase gene that contains a translational stop and restart in the section coding for the variable region. The structure of M·*Aqu*I reinforces the notion that enzymes evolve in steps, first by the association of functional polypeptide units, then by the joining of those units through gene fusion. M·*Aqu*I may represent an early stage in the evolution of methyltransferases, in which the final fusion to link the units into a continuous chain has yet to occur.

MULTISPECIFIC METHYLTRANSFERASES Most methyltransferases recognize a single DNA sequence, albeit one that can be degenerate. Two exceptions have been found: type IIs MTases such as M·*Fok*I, which combine two methyltransferases of complementary specificity into one protein (116, 133, 210), and the multispecific *Bacillus* phage m^5C-MTases that methylate several unrelated sequences (221). The latter enzymes exemplify extreme biological economy, and their analysis has enabled the TRDs of m^5C-MTases to be mapped with precision.

Bacillus phages ϕ3T, ρ11s and SPR encode closely related methyltransferases that methylate GGCC and one or two additional sequences besides. The major difference between these MTases and the monospecific m^5C-MTases occurs in the variable region. The region is 80–120 aa in the monospecific MTases and up to 250 aa in the multispecific MTases. The increased length is due to the presence of up to four motifs, each of which is a TRD of approximately 40 aa (8, 235). Comparable motifs can be discerned, in single copy, in the variable region of many of the monospecific m^5C-MTases (127).

Point mutations that eliminate methylation altogether occur throughout the multispecific m^5C-MTases; presumably they affect catalytic functions needed regardless of the sequence recognized (72, 235). Mutations that abolish the methylation of only one sequence occur in just the variable region (235). Mutations that affect different specificities define different TRDs; the TRDs occur sequentially (235), with no space or overlap between them (T. Traut-

parisons: enzymes that catalyze identical reactions—*Bsu*RI, *Hae*III, and *Ngo*-PII, for example; or *Bsu*FI and *Msp*I—have similar variable regions unlike enzymes that recognize different sequences (131, 199, 211, 232). The variable regions of enzymes that recognize the same sequence but methylate different cytosines—*Hpa*II and *Msp*I, for example—differ (34, 131). No resemblance between variable region sequences and the helix-turn-helix motifs of DNA binding proteins has been reported.

N-6 ADENINE METHYLTRANSFERASES The m^{6}A-MTases are more diverse than the m^{5}C-MTases (78, 124, 126). Approximately 40 have been characterized that can be divided into three subclasses according to aa sequence organization (108). The two most conserved sequence motifs among the m^{6}A-MTases are F-G-G (motif I) and DPPY (motif II). The motifs occur in numerical order in one subclass (α)[2], and in the reverse order in another subclass (β) (108, 125). In the third subclass (γ), the motifs occur closer together in numerical order; motif II is usually NPPY, and motif I is LEP—G-G. (Motif I in the α class is often LEPF-G-G.).

The position of the TRD in the M^{6}A-MTases has not been mapped. A variable region of 100–200 aa separates motifs I and II in the α- and β-type methyltransferases, and the TRD probably occupies part of this region. The *Eco*P1 (AGACC) and *Eco*P15 (CAGCAG) methyltransferases are similar in all but this region (85). Mutants of the phage T2 and T4 DNA-adenine methyltransferases have been isolated that decrease specificity (*dam*^{h} mutants) and increase substrate discrimination (*dam*^{c} mutants) (146). The mutations affect neighboring amino acids in the variable region, implicating this region in specificity determination. The specificities of the wild-type phage enzymes differ with respect to noncanonical methylation ("star" activity) (187); the enzymes have identical variable regions, however, indicating that sections of the proteins outside the variable region affect specificity (146).

N-4 CYTOSINE METHYLTRANSFERASES Only recently have the sequences for m^{4}C-MTases become available (27, 80, 108, 218; G. Wilson, unpublished observations). They are organized along the same lines as m^{6}A-MTases and occur with the same two permuted motif organizations, α and β. The methylation reaction mechanism is incompletely understood (160), but the similarity between m^{4}C-MTases and m^{6}A-MTases suggests a common mechanism. Indeed, viewed from the major groove of DNA, the local environments of the exocyclic nitrogens of adenine and cytosine appear similar.

[2]The classifications α, β, and γ are personal preferences, used here for convenience; a uniform classification has not been agreed upon.

Approximately two hundred specificities have now been discovered. This wide range reflects the circumstance that numerous DNA sequences can be targets for restriction. Independent origins aside, however, why do enzymes of different specificity differ so markedly from one another? Does it signify that radically different aa sequences are needed to recognize slightly different nt sequences? Or that comparable recognition tasks can be accomplished by entirely different aa sequences? Probably both answers are true to some extent. The three-dimensional structures of various endonucleases will likely be solved in the next few years, and will help to elucidate these questions.

ACKNOWLEDGMENTS

We have omitted certain aspects of restriction and modification that are better reviewed by others. In the interest of brevity, we have done scant justice to the work of many scientists, and thus ask indulgence. We are indebted to numerous colleagues for unpublished information and for discussions, especially Francis Barany, Timothy Bestor, David Dryden, Stanley Hattman, Arvydas Janulaitis, Manfred Kröger, Richard Roberts, Paul Sharp, Barton Slatko, Daniel Stein, and Thomas Trautner. We thank the members of our laboratories for help and discussions, including Vicki Barcus, Janet Barsomian, Keith Lunnen, Marta Meda, Lynn Powell, Elizabeth Van Cott, Damien Willcock, and Bo-Hong Zhang.

Literature Cited

1. Alvarez, M. A., Chater, K. F., Rodicio, M. R. 1990. Transcriptional analysis of the genes for the SalI restriction modification system. *NEB Workshop Biol. DNA Modif., Berlin, 2nd*, p. 91
2. Arber, W. 1971. Host controlled variation. In *The Bacteriophage Lambda*, ed. A. D. Hershey, pp. 83–96. Cold Spring Harbor, NY: Cold Spring Harbor Lab. Press
3. Arber, W. 1965. Host specificity of DNA produced by *Escherichia coli* V. The role of methionine in the production of host specificity. *J. Mol. Biol.* 11:247–56
4. Athanasiadis, A., Gregoriu, M., Thanos, D., Kokkinidis, M., Papamatheakis, J. 1990. Complete nucleotide sequence of the PvuII restriction enzyme gene from *Proteus vulgaris*. *Nucleic Acids Res.* 18:6434
5. Balganesh, T. S., Reiners, L., Lauster, R., Noyer-Weidner, M., Wilke, K., et al. 1987. Construction and use of chimeric SPR/ϕ3T DNA methyltransferases in the definition of sequence recognizing enzyme regions. *EMBO J.* 6:3543–49
6. Barany, F. 1988. Overproduction, purification and crystallization of TaqI restriction endonuclease. *Gene* 65:167–77
7. Barras, F., Marinus, M. G. 1989. The great GATC: DNA methylation in *E. coli*. *Trends Genet.* 5:139–43
8. Behrens, B., Noyer-Weidner, M., Pawlek, B., Lauster, R., Balganesh, T. S., et al. 1987. Organization of multispecific DNA methyltransferases encoded by temperate *Bacillus subtilis* phages. *EMBO J.* 6:1137–42
9. Belfort, M. 1989. Bacteriophage introns: parasites within parasites? *Trends Genet.* 5:209–13
10. Bennett, S. P., Halford, S. E. 1989. Recognition of DNA by Type II restriction enzymes. In *Current Topics in Cellular Regulation*, ed. B. L. Horecker, pp. 57–104. New York: Academic
11. Bestor, T., Laudano, A., Mattaliano, R., Ingram, V. 1988. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells: the carboxy-terminal domain of the mammalian enzyme is related to bacterial restriction methyltransferases. *J. Mol. Biol.* 203:971–83
12. Bestor, T. H., Ingram, V. M. 1983.

ner, personal communication). The TRDs of ϕ3T I and SPR have each been assigned specificities, but two of the TRDs in ρ11s appear to be nonfunctional (8). These "silent" TRDs differ from functional TRDs by only a few amino acids (122); they might represent TRDs in the making, or TRDs that have lost function by mutation. The final TRD in the variable regions of the multi-specific MTases recognizes GGCC. The aa sequence of the TRD is similar to part of the variable region of the monospecific methylases that also recognize this sequence.

HYBRID METHYLTRANSFERASES Chimeric methyltransferase have been constructed by interchanging TRDs among the multispecific MTases (5, 222). The specificities of the chimeras reflect the specificities of their TRDs. New specificities, per se, have not been created, but novel combinations of the parental specificities have. Thus, one chimera, between ρ11s (GDGCHC and GGCC) and SPR (CCWGG, CCGG and GGCC), methylates GDGCHC, CCGG, and GGCC; another, between ρ11s and ϕ3T I (GCNGC and GGCC), methylates GDGCHC, GCNGC, and GGCC (222); and a third, between ϕ3T I and SPR, methylates GCNGC, CCGG and GGCC (5).

DISCUSSION

Close gene linkage is a feature of all restriction-modification systems. The genes are usually adjacent and often overlap. Despite this consistent linkage, R-M systems are genetically diverse. Type I systems are the least diverse; they occur as families whose members have similar enzymatic subunits, but different specificity subunits. Type II systems appear not to occur as families; while the systems may often have similar organizations, the enzymes usually have dissimilar sequences. Type II endonucleases of different specificity are entirely dissimilar; type II methyltransferases are somewhat similar, but the similarities are skeletal and may reflect mechanistic constraints rather than common ancestry. Many type II systems appear to have formed independently as partnerships between miscellaneous genes that were initially separate, but became linked due to a persistent selective advantage.

The nature of that advantage might be similar to that which compelled the clustering of biosynthesis genes. Interdependent functions are more efficiently coregulated, and more likely to be inherited as a unit if their genes are closely linked than separate. Since both activities of R-M systems are needed to form a restriction barrier, only systems inherited in whole confer any benefit; indeed, systems inherited in part could be detrimental. Remarkably, in R-M systems, particularly the type II systems, faced with the problem of viral infection, bacteria seem to have chanced upon the same solution so many times, evolving a multitude of equivalent systems, albeit differing in the particulars of specificity.

ing and characterization of the *Hpa*II methylase gene. *Nucleic Acids Res.* 18:1377–83
35. Cerritelli, S., Springhorn, S. S., Lacks, S. A. 1989. *Dpn*A, a methylase for single-stranded DNA in the *Dpn*II restriction system, and its biological function. *Proc. Natl. Acad. Sci. USA* 86:9223–27
36. Chandrasegaran, S., Lunnen, K. D., Smith, H. O., Wilson, G. G. 1988. Cloning and sequencing of the *Hin*fI restriction and modification genes. *Gene* 70:387–92
37. Chandrasegaran, S., Smith, H. O. 1988. Amino acid sequence homologies among twenty-five restriction endonucleases and methylases. In *Structure and Expression*, ed. M. H. Sarma, R. H. Sarma, pp. 149–56. Guilderland, NY: Adenine
38. Chandrasegaran, S., Smith, H. O., Amzel, M. L., Ysern, X. 1986. Preliminary x-ray diffraction analysis of HhaII endonuclease-DNA cocrystals. *Proteins* 1:263–66
39. Chu, F. K., Maley, G., Pedersen-Lane, J., Wang, A. M., Maley, F. 1990. Characterization of the restriction site of a prokaryotic intron-encoded endonuclease. *Proc. Natl. Acad. Sci. USA* 87:3574–78
40. Colleaux, L., D'Auriol, L., Galibert, F., Dujon, B. 1988. Recognition and cleavage site of the intron-encoded *omega* transposase. *Proc. Natl. Acad. Sci. USA* 85:6022–26
41. Connaughton, J. F., Kaloss, W. D., Vanek, P. G., Nardone, G. A., Chirikjian, J. G. 1990. The complete sequence of the *Bacillus amyloliquefaciens* proviral H2, *Bam*HI methylase gene. *Nucleic Acids Res.* 18:4002
42. Coulby, J. M., Sternberg, N. L. 1988. Characterization of the phage P1 *dam* gene. *Gene* 74:191
43. Cowan, G. M., Gann, A. A. F., Murray, N. E. 1989. Conservation of complex DNA recognition domains between families of restriction enzymes. *Cell* 56:103–9
44. Daniel, A. S., Fuller-Pace, F. V., Legge, D. M., Murray, N. E. 1988. Distribution and diversity of *hsd* genes in *Escherichia coli* and other enteric bacteria. *J. Bacteriol.* 170:1775–82
45. D'Arcy, A., Brown, R. S., Zabeau, M., van Resandt, R. W., Winkler, F. K. 1985. Purification and crystallization of the EcoRV restriction endonuclease. *J. Biol. Chem.* 260:1987–90
46. Davies, J. K. 1989. DNA restriction and modification systems in *Neisseria gonorrhoeae*. *Clin. Microbiol. Rev.* 2:S78-S82
47. De Backer, O., Colson, C. 1991. Two-step cloning and expression in *Escherichia coli* of the DNA restriction-modification system *Sty*LTI of *Salmonella typhimurium*. *J. Bacteriol* 173:1321–27
48. Dila, D., Sutherland, E., Moran, L., Slatko, B., Raleigh, E. A. 1990. Genetic and sequence organization of the *mcr*BC locus of *Escherichia coli* K-12. *J. Bacteriol.* 172:4888–900
49. Dorner, L., Robinson, D., Schildkraut, I. 1990. Crystallization of restriction endonucleases and methylases. See Ref. 1, p. 142
50. Duff, M. K., Davies, J. K. 1988. Multiple restriction-modification systems in *Neisseria gonorrhoeae*. *Gene* 74:227–28
51. Dugaiczyk, A., Hedgpeth, J., Boyer, H. W., Goodman, H. M. 1974. Physical identity of the SV40 deoxyribonucleic acid sequence recognized by the EcoRI restriction endonuclease and modification methylase. *Biochemistry* 13:503–12
52. Dujon, B. 1989. Group I introns as mobile genetic elements: facts and mechanistic speculations—a review. *Gene* 82:91–114
53. Düsterhöft, A. 1990. *Characterization of restriction modification systems* Hgi*BI (GGWCC) and* Hgi*DI*/Hgi*GI (GRCGYC) and* Hgi*DII (GTCGAC)*. PhD thesis. Giessen, Germany: Justus-Liebig-Universität
54. Düsterhöft, A., Erdmann, D., Kröger, M. 1991. Stepwise cloning and molecular characterization of the *Hgi*DI restriction-modification system from *Herpetosiphon giganteus Hpa*2. *Nucleic Acids Res.* 19:1049–56
55. Ehrlich, M., Gama-Sosa, M. A., Carreira, L. H., Ljungdahl, L. G., Kuo, K. C., et al. 1985. DNA methylation in thermophilic bacteria: N^4-methylcytosine, 5-methylcytosine, and N^6-methyladenine. *Nucleic Acids Res.* 13:1399–412
56. Ehrlich, M., Wilson, G. G., Kenneth, C. K., Gehrke, C. W. 1987. N^4-Methylcytosine as a minor base in bacterial DNA. *J. Bacteriol.* 169:939–43
57. Endlich, B., Linn, S. 1985. The DNA restriction endonuclease of *Escherichia coli* B. II. Further studies of the structure of DNA intermediates and products. *J. Biol. Chem.* 260:5729–38
58. Endlich, B., Linn, S. 1981. Type I restriction enzymes. *Enzymes* 14:137–56
59. Erdmann, D. 1991. *Characterization of restriction modification systems* Hgi*CI (GGYRCC) and* Hgi*EI*/Hgi*CII*

Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc. Natl. Acad. Sci. USA* 80:5559–63

13. Bhagwat, A. S., Johnson, B., Weule, K., Roberts, R. J. 1990. Primary sequence of the EcoRII endonuclease and properties of its fusions with beta-galactosidase. *J. Biol. Chem.* 265:767–73
14. Bickle, T. A. 1982. The ATP-dependent restriction endonucleases. In *Nucleases*, ed. S. M. Linn, R. J. Roberts, pp. 85–108. Cold Spring Harbor, NY: Cold Spring Harbor Lab. Press
15. Bickle, T. A. 1987. DNA restriction and modification systems. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, ed. F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, H. E. Umbarger, pp. 692–96. Washington, DC: Am. Soc. Microbiol.
16. Bickle, T. A., Brack, C., Yuan, R. 1978. ATP-induced conformational changes in the restriction endonuclease from *Escherichia coli* K-12. *Proc. Natl. Acad. Sci. USA* 75:3099–3103
17. Bigger, C. H., Murray, K., Murray, N. E. 1973. Recognition sequence of a restriction enzyme. *Nature New Biol.* 244:7–10
18. Bitinaite, J., Maneliene, Z., Grigaite, R., Menkevicius, S., Butkus, V., et al. 1990. *Alw*26I, *Eco*31I and *Esp*3I—type IIs methyltransferases capable of modifying different bases in complementary DNA strands. See Ref. 1, p. 128
19. Blumenthal, R. M., Gregory, S. A., Cooperider, J. S. 1985. Cloning of a restriction-modification system from *Proteus vulgaris* and its use in analyzing a methylase-sensitive phenotype in *Escherichia coli*. *J. Bacteriol.* 164:501–9
20. Bocklage, H., Heeger, K., Müller-Hill, B. 1991. Cloning and characterization of the *Mbo*II restriction-modification system. *Nucleic Acids Res.* 19:1007–13
21. Borck, K., Beggs, J. D., Brammar, W. J., Hopkins, A. S., Murray, N. E. 1976. The construction in vitro of transducing derivatives of phage lambda. *Mol. Gen. Genet.* 146:199–207
22. Bougueleret, L., Schwarzstein, M., Tsugita, A., Zabeau, M. 1984. Characterization of the genes coding for the *Eco*RV restriction and modification system of *Escherichia coli*. *Nucleic Acids Res.* 12:3659–76
23. Boyer, H. W. 1971. DNA restriction and modification mechanisms in bacteria. *Annu. Rev. Microbiol.* 25:153–76
24. Boyer, H. W., Chow, L. T., Dugaiczyk, A., Hedgpeth, J., Goodman, H. M. 1973. DNA substrate site for the EcoRII restriction endonuclease and modification methylase. *Nature New Biol.* 244:40–43
25. Brockes, J. P., Brown, P. R., Murray, K. 1972. The deoxyribonucleic acid modification enzyme of bacteriophage P1: Purification and properties. *Biochem. J.* 127:1–10
26. Brooks, J. E., Benner, J. S., Heiter, D. F., Silber, K. R., Sznyter, L. A., et al. 1989. Cloning the *Bam*HI restriction modification system. *Nucleic Acids Res.* 17:979–97
27. Brooks, J. E., Nathan, P. D., Landry, D., Sznyter, L. A., Waite-Rees, P. A., et al. 1991. Characterization of the cloned *Bam*HI restriction modification system: its nucleotide sequence, properties of the methylase, and expression in heterologous hosts. *Nucleic Acids Res.* 19:841–50
28. Bullas, L. R., Colson, C. 1975. DNA restriction and modification systems in Salmonella. III. SP, a *Salmonella potsdam* system allelic to the SB system in *Salmonella typhimurium*. *Mol. Gen. Genet.* 139:177–88
29. Bullas, L. R., Colson, C., van Pel, A. 1976. DNA restriction and modification systems in Salmonella. SQ, a new system derived by recombination between the SB system of *Salmonella typhimurium* and the SP system of *Salmonella potsdam*. *J. Gen. Microbiol.* 95:166–72
30. Burckhardt, J., Weisemann, J., Hamilton, D. L., Yuan, R. 1981. Complexes formed between the restriction endonuclease EcoK and heteroduplex DNA. *J. Mol. Biol.* 153:425–40
31. Butkus, V., Klimasauskas, S., Kersulyte, D., Vaitkevicius, D., Lebionka, A., et al. 1985. Investigation of restriction-modification enzymes from M. varians RFL19 with a new type of specificity toward modification of substrate. *Nucleic Acids Res.* 13:5727–46
32. Butkus, V., Padegimiene, A., Menkevicius, S., Butkiene, D., Timinskas, A., et al. 1990. Sequencing and primary structure comparison of MvaI and BcnI RM enzymes recognizing similar sequences—CC(A,T)GG and CC(C,G)GG. See Ref. 1, p. 84
33. Cannon, S. V., Cummings, A., Teebor, G. W. 1988. 5-Hydroxymethylcytosine DNA glycosylase activity in mammalian tissue. *Biochem. Biophys. Res. Commun.* 151:1173–79
34. Card, C. O., Wilson, G. G., Weule, K., Hasapes, J., Kiss, A., et al. 1990. Clon-

84. Hsu, M. Y., Inouye, M., Inouye, S. 1990. Retron for the 67-base multicopy single-stranded DNA from *Escherichia coli*: A potential transposable element encoding both reverse transcriptase and Dam methylase functions. *Proc. Natl. Acad. Sci. USA* 87:9454–58
85. Hümbelin, M., Suri, B., Rao, D. N., Hornby, D. P., Eberle, H., et al. 1988. Type III DNA *res*triction and *mod*ification systems *Eco*P1 and *Eco*P15: Nucleotide sequence of the *Eco*P1 operon, the *Eco*P15 *mod* gene and some *Eco*P1 *mod* mutants. *J. Mol. Biol.* 200:23–29
86. Ito, H., Sadaoka, A., Kotani, H., Hiraoka, N., Nakamura, T. 1990. Cloning, nucleotide sequence, and expression of the *Hinc*II restriction-modification system. *Nucleic Acids Res.* 18:3903–11
87. Janulaitis, A., Kazlauskiene, R., Lazareviciute, L., Gilvonauskaite, R., Steponaviciene, D., et al. 1988. Taxonomic specificity of restriction-modification enzymes. *Gene* 74:229–32
88. Janulaitis, A., Povilionis, P., Sasnauskas, K. 1982. Cloning of the modification methylase gene of *Bacillus centrosporus* in *Escherichia coli*. *Gene* 20:197–204
89. Janulaitis, A., Vaisvila, R., Klimasauskas, S., Petrusyte, M., Maneliene, Z., et al. 1990. Cloning and sequencing of the genes coding *Eco*57I restriction-modification system—prototype of type IV enzymes. See Ref. 1, p. 12
90. Janulaitis, A., Klimasauskas, S., Petrusyte, M., Butkus, V. 1983. Cytosine modification in DNA by BcnI methylase yields N^4-methylcytosine. *FEBS Lett.* 161:131–34
91. Jones, M., Wagner, R., Radman, M. 1987. Mismatch repair of deaminated 5-methylcytosine. *J. Mol. Biol.* 194:155–59
92. Josephensen, J., Klaenhammer, T. 1990. Stacking of three different restriction and modification systems in *Lactococcus lactis* by cotransformation. *Plasmid* 23:71–75
93. Kannan, P., Cowan, G. M., Daniel, A. S., Gann, A. A. F., Murray, N. E. 1989. Conservation of organization in the specificity polypeptides of two families of type I restriction enzymes. *J. Mol. Biol.* 209:335–44
94. Kapfer, W., Walter, J., Trautner, T. A. 1990. Characterization of the *Bsu*FI restriction endonuclease—comparison to the isoschizomer R. *Msp*I. See Ref. 1, p. 93
95. Karreman, C., de Waard, A. 1990. *Agmenellum quadruplicatum* M. *Aqu*I, a novel modification methylase. *J. Bacteriol.* 172:266–72
96. Karreman, C., de Waard, A. 1988. Cloning and complete nucleotide sequences of the type II restriction-modification genes of *Salmonella infantis*. *J. Bacteriol.* 170:2527–32
97. Kaszubska, W., Aiken, C., O'Connor, C. D., Gumport, R. I. 1989. Purification, cloning and sequence analysis of *Rsr*I DNA methyltransferase: lack of homology between two enzymes, *Rsr*I and *Eco*RI, that methylate the same nucleotide in identical recognition sequences. *Nucleic Acids Res.* 17:10403–25
98. Kawasaki, K., Takahashi, M., Natori, M., Shibata, T. 1991. DNA sequence recognition by a eukaryotic sequence-specific endonuclease, Endo.SceI, from *Saccharomyces cerevisiae*. *J. Biol. Chem.* 266:5342–47
99. Kelleher, J. E., Daniel, A. S., Murray, N. E. 1991. Mutations that confer de novo methyltransferase activity. *J. Mol. Biol.* Submitted
100. Kelleher, J. E., Raleigh, E. A. 1991. A novel activity in *E. coli* K-12 that directs restriction of DNA modified at CG dinucleotides. *J. Bacteriol.* In press
101. Kelly, T. J. J., Smith, H. O. 1970. A restriction enzyme from *Hemophilus influenzae* II. Base sequence of the recognition site. *J. Mol. Biol.* 51:393–409
102. Kessler, C., Manta, V. 1990. Specificity of restriction endonucleases and DNA modification methyltransferases—a review. *Gene* 92:1–248
103. Kim, Y., Grable, J. C., Love, R., Greene, P. J., Rosenberg, J. M. 1990. Refinement of EcoRI endonuclease crystal structure: A revised protein chain tracing. *Science* 249:1307–9
104. Kiss, A., Baldauf, F. 1983. Molecular cloning and expression in *Escherichia coli* of two modification methylase genes of *Bacillus subtilis*. *Gene* 21:111–19
105. Kiss, A., Pósfai, G., Keller, C. C., Venetianer, P., Roberts, R. J. 1985. Nucleotide sequence of the *Bsu*RI restriction-modification system. *Nucleic Acids Res.* 13:6403–21
106. Kita, K., Kotani, H., Hiraoka, N., Nakamura, T., Yonaha, K. 1989. Overproduction and crystallization of FokI restriction endonuclease. *Nucleic Acids Res.* 17:8741
107. Kita, K., Kotani, H., Sugisaki, H., Takanami, M. 1989. The FokI restric-

(GGWCC). PhD thesis. Giessen, Germany: Justus-Liebig-Universität
60. Eskin, B., Linn, S. 1972. The deoxyribonucleic acid modification and restriction enzymes of *Escherichia coli* B. II. Purification, subunit structure, and catalytic properties of the restriction endonuclease. *J. Biol. Chem.* 247:6183–91
61. Feher, Z., Schlagman, S. L., Miner, Z., Hattman, S. 1989. The UV excision-repair system of *Saccharomyces cerevisiae* is involved in the removal of methylcytosines formed in vivo by a cloned prokaryotic DNA methyltransferase. *Curr. Genet.* 16:461–64
62. Frey, B., Wedlich, A., Kaluza, K., Herz, G., Lechner, M., et al. 1990. Screening for novel class II restriction endonucleases. See Ref. 1, p. 92
63. Fuller-Pace, F. V., Bullas, L. R., Delius, H., Murray, N. E. 1984. Genetic recombination can generate altered restriction specificity. *Proc. Natl. Acad. Sci. USA* 81:6095–99
64. Fuller-Pace, F. V., Cowan, G. M., Murray, N. E. 1985. *Eco*A and *Eco*E: alternatives to the *Eco*K family of type I restriction and modification systems of *Escherichia coli*. *J. Mol. Biol.* 186:65–75
65. Fuller-Pace, F. V., Murray, N. E. 1986. Two DNA recognition domains of the specificity polypeptides of a family of type I restriction enzymes. *Proc. Natl. Acad. Sci. USA* 83:9368–72
66. Gann, A. A. F., Campbell, A. J. B., Collins, J. F., Coulson, A. F. W., Murray, N. E. 1987. Reassortment of DNA recognition domains and the evolution of new specificities. *Mol. Microbiol.* 1:13–22
67. Gingeras, T. R., Brooks, J. E. 1983. Cloned restriction/modification system from *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. USA* 80:402–6
68. Gough, J. A., Murray, N. E. 1983. Sequence diversity among related genes for recognition of specific targets in DNA molecules. *J. Mol. Biol.* 166:1–19
69. Greene, P. J., Gupta, M., Boyer, H. W., Brown, W. E., Rosenberg, J. M. 1981. Sequence analysis of the DNA encoding the *Eco*RI endonuclease and methylase. *J. Biol. Chem.* 256:2143–53
70. Gruenbaum, Y., Cedar, H., Razin, A. 1982. Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature* 295:620–22
71. Gubler, M., Bickle, T. A. 1991. Increased protein flexibility leads to promiscuous protein-DNA interactions in type IC restriction-modification systems. *EMBO J.* 10:951–57
72. Gunthert, U., Lauster, R., Reiners, L. 1986. Multispecific DNA methyltransferases from *Bacillus subtilis* phages. Properties of wild-type and various mutant enzymes with altered DNA affinity. *Eur. J. Biochem.* 159:485–92
73. Hammond, A. W., Gerard, G. F., Chatterjee, D. K. 1990. Cloning the *Kpn*I restriction-modification system in *Escherichia coli*. *Gene* 97:97–102
74. Hanck, T., Gerwin, N., Fritz, H. -J. 1989. Nucleotide sequence of the *dcm* locus of *Escherichia coli* K12. *Nucleic Acids Res.* 17:5844
75. Hattman, S. 1971. Variation of 6-methylaminopurine content in bacteriophage P22 deoxyribonucleic acid as a function of host specificity. *J. Virol.* 7:690–91
76. Hattman, S., Fukasawa, T. 1963. Host-induced modification of T-even phages due to defective glucosylation of their DNA. *Proc. Natl. Acad. Sci. USA* 50:297–300
77. Hattman, S., Gold, E., Plotnik, A. 1972. Methylation of cytosine residues in DNA controlled by a drug resistance factor. *Proc. Natl. Acad. Sci. USA* 69:187–90
78. Hattman, S., Wilkinson, J., Swinton, D., Schlagman, S., Macdonald, P. M., et al. 1985. Common evolutionary origin of the phage T4 dam and host *Escherichia coli* dam DNA-adenine methyltransferase genes. *J. Bacteriol.* 164:932–37
79. Hedgpeth, J., Goodman, H. M., Boyer, H. W. 1972. DNA nucleotide sequence restricted by the RI endonuclease. *Proc. Natl. Acad. Sci. USA* 69:3448–52
80. Heidmann, S., Seifert, W., Kessler, C., Domdey, H. 1989. Cloning, characterization and heterologous expression of the *Sma*I restriction-modification system. *Nucleic Acids Res.* 17:9783–96
81. Heitman, J., Model, P. 1987. Site-specific methylases induce *SOS* DNA repair response in *Escherichia coli*. *J. Bacteriol.* 169:3243–50
82. Heitman, J., Zinder, N. D., Model, P. 1989. Repair of the *Escherichia coli* chromosome after in vivo scission by the EcoRI endonuclease. *Proc. Natl. Acad. Sci. USA* 86:2281–85
83. Howard, K. A., Card, C., Benner, J. S., Callahan, H. L., Maunus, R., et al. 1986. Cloning the *Dde*I restriction-modification system using a two-step method. *Nucleic Acids Res.* 14:7939–51

Braymer, H. D., Murray, N. E. 1987. Organization and sequence of the *hsd* genes of *Escherichia coli* K-12. *J. Mol. Biol.* 198:159–70

133. Looney, M. C., Moran, L. S., Jack, W. E., Feehery, G. R., Benner, J. S., et al. 1989. Nucleotide sequence of the *Fok*I restriction-modification system: separate strand specificity domains in the methyltransferase. *Gene* 80:193–208
134. Lunnen, K. D., Barsomian, J. M., Camp, R. R., Card, C. O., Chen, S. Z., et al. 1988. Cloning type-II restriction and modification genes. *Gene* 74:25–32
135. Macaluso, A., Mettus, A. M. 1991. Efficient transformation of *Bacillus thuringiensis* requires nonmethylated plasmid DNA. *J. Bacteriol.* 173:1353–56
136. MacNeil, D. J. 1988. Characterization of a unique methyl-specific restriction system in *Streptomyces avermitilis*. *J. Bacteriol.* 170:5607–12
137. Maekawa, Y., Yasukawa, H., Kawakami, B. 1990. Cloning and nucleotide sequence of the *Ban*I restriction-modification genes in *Bacillus aneurinolyticus*. *J. Biochem.* 107:645–49
138. Mann, M. B., Rao, R. N., Smith, H. O. 1978. Cloning of restriction and modification genes in *E. coli*: The *Hha*II system from *Haemophilus haemolyticus*. *Gene* 3:97–112
139. Mannarelli, B. M., Balganesh, T. S., Greenberg, B., Springhorn, S. S., Lacks, S. A. 1985. Nucleotide sequence of the *Dpn*II DNA methylase gene of *Streptococcus pneumoniae* and its relationship to the *dam* gene of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 82:4468–72
140. Marinus, M. G. 1987. DNA methylation in *Escherichia coli*. *Annu. Rev. Genet.* 21:113–31
141. McClelland, M., Nelson, M. 1988. The effect of site-specific methylation on restriction endonucleases and DNA modification methyltransferases—a review. *Gene* 74:291–304
142. McClelland, M., Nelson, M., Cantor, C. R. 1985. Purification of MboII methylase (GAAGmA) from *Moraxella bovis*: site specific cleavage of DNA at nine and ten base pair sequences. *Nucleic Acids Res.* 13:7171–82
143. Meselson, M., Yuan, R. 1968. DNA restriction enzyme from *E. coli*. *Nature* 217:1110–14
144. Meselson, M., Yuan, R., Heywood, J. 1972. Restriction and modification of DNA. *Annu. Rev. Biochem.* 41:447–66
145. Miner, Z., Hattman, S. 1988. Molecular cloning, sequencing and mapping of the bacteriophage T2 *dam* gene. *J. Bacteriol.* 170:5177–84
146. Miner, Z., Schlagman, S. L., Hattman, S. 1989. Single amino acid changes that alter the DNA sequence specificity of the DNA- methyltransferase (Dam) of bacteriophage T4. *Nucleic Acids Res.* 17:8149–57
147. Monteilhet, C., Perrin, A., Thierry, A., Colleaux, L., Dujon, B. 1990. Purification and characterization of the in vitro activity of I-Sce I, a novel and highly specific endonuclease encoded by a group I intron. *Nucleic Acids Res.* 18:1407–13
148. Murray, N. E., Batten, P. L., Murray, K. 1973. Restriction of bacteriophage lambda by *Escherichia coli* K. *J. Mol. Biol.* 81:395–407
149. Murray, N. E., Gough, J. A., Suri, B., Bickle, T. A. 1982. Structural homologies among type I restriction-modification systems. *EMBO J.* 1:535–39
150. Muscarella, D. E., Ellison, E. L., Ruoff, B. M., Vogt, V. M. 1990. Characterization of I-*Ppo*, an intron-encoded endonuclease that mediates homing of a group I intron in the ribosomal DNA of *Physarum polycephalum*. *Mol. Cell. Biol.* 10:3386–96
151. Nagaraja, V., Shepherd, J. C. W., Bickle, T. A. 1985. A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity. *Nature* 316:371–72
152. Nathans, D., Smith, H. O. 1975. Restriction endonucleases in the analysis and restructuring of DNA molecules. *Annu. Rev. Biochem.* 44:273–93
153. Newman, A. K., Rubin, R. A., Kim, S-H., Modrich, P. 1981. DNA sequences of structural genes for *Eco*RI DNA restriction and modification enzymes. *J. Biol. Chem.* 256:2131–39
154. Nickoloff, J. A., Singer, J. D., Heffron, F. 1990. In vivo analysis of the *Saccharomyces cerevisiae* HO nuclease recognition site by site-directed mutagenesis. *Mol. Cell. Biol.* 10:1174–79
155. Noyer-Weidner, M., Diaz, R., Reiners, L. 1986. Cytosine-specific DNA modification interferes with plasmid establishment in *Escherichia coli* K12: involvement of rglB. *Mol. Gen. Genet.* 205:469–75
156. Noyer-Weidner, M., Krüger, T., Grund, C., Wild, C. 1990. Characterization of the *mcr*B loci of *E. coli* K12 wild type and mutant strains. See Ref. 1, p. 45
157. Old, R., Murray, K., Roizes, G. 1975.

tion-modification system. I: Organization and nucleotide sequences of the restriction and modification genes. *J. Biol. Chem.* 264:5751–56

108. Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkienė, D., Butkus, V., et al. 1989. Sequence motifs characteristic of DNA [cytosine-N4] methyltransferases: similarity to adenine and cytosine-C5 DNA-methylases. *Nucleic Acids Res.* 17:9823–32
109. Koncz, C., Kiss, A., Venetianer, P. 1978. Biochemical characterization of the restriction-modification system of *Bacillus sphaericus*. *Eur. J. Biochem.* 89:523–29
110. Korch, C., Hagblom, P., Normark, S. 1983. Sequence-specific DNA modification in *Neisseria gonorrhoeae*. *J. Bacteriol.* 155:1324–32
111. Kossykh, V., Repyk, A., Kaliman, A., Buryanov, Y. 1989. Nucleotide sequence of the *Eco*RII restriction endonuclease gene. *Biochim. Biophys. Acta* 1009:290–92
112. Kossykh, V. G., Buryanov, Y. I., Bayev, A. A. 1980. Molecular cloning of *Eco*RII endonuclease and methylase genes. *Mol. Gen. Genet.* 178:717–18
113. Krüger, D. H., Bickle, T. A. 1983. Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol. Rev.* 47:345–60
114. Krüger, D. H., Bickle, T. A., Reuter, M., Pein, C-D., Schroeder, C. 1990. DNA methylation and restriction processes in *Escherichia coli*: insights by use of bacterial viruses T3 and T7. In *Nucleic Acid Methylation*, ed. G. A. Clawson, pp.113–24. New York: Liss
115. Kuhnlein, U., Linn, S., Arber, W. 1969. Host specificity of DNA produced by *Escherichia coli*, XI. In vitro modification of phage fd replicative form. *Proc. Natl. Acad. Sci. USA* 63:556–62
116. Labbé, D., Höltke, H. J., Lau, P. C. K. 1990. Cloning and characterization of two tandemly arranged DNA methyltransferase genes of *Neisseria* lactamica: An adenine-specific M. *Nla*III and a cytosine-type methylase. *Mol. Gen. Genet.* 224:101–10
117. Lacks, S., Greenberg, B. 1977. Complementary specificity of restriction endonucleases of *Diplococcus pneumoniae* with respect to DNA methylation. *J. Mol. Biol.* 114:153–68
118. Lacks, S., Greenberg, B. 1975. A deoxyribonuclease of *Diplococcus pneumoniae* specific for methylated DNA. *J. Biol. Chem.* 250:4060–66
119. Lacks, S. A., Mannarelli, B. M., Springhorn, S. S., Greenberg, B. 1986. Genetic basis of the complementary *Dpn*I and *Dpn*II restriction systems of *S. pneumoniae:* an intercellular cassette mechanism. *Cell* 46:993–1000
120. Lambowitz, A. M. 1989. Infectious introns. *Cell* 56:323–26
121. Landry, D., Looney, M. C., Feehery, G. R., Slatko, B. E., Jack, W. E., et al. 1989. M.*Fok*I methylates adenine in both strands of its asymmetric recognition sequence. *Gene* 77:1–10
122. Lange, C., Jugel, A., Noyer-Weidner, M., Trautner, T. A. 1990. Silent target recognizing domains (TRDs) in multispecific DNA-methyltransferases and their activation. See Ref. 1, p. 25
123. Laue, F., Evans, L. R., Jarsch, M., Brown, N. L., Kessler, C. 1991. A complex family of class-II restriction endonucleases, *Dsa*I-VI, in *Dactylococcopsis salina*. *Gene* 97:87–95
124. Lauster, R. 1989. Close relationship between the HinfI and *Dpn*A DNA-methyltransferase. *Nucleic Acids Res.* 17:4402
125. Lauster, R. 1989. Evolution of type II DNA methyltransferases. *J. Mol. Biol.* 206:313–21
126. Lauster, R., Kriebardis, A., Guschlbauer, W. 1987. The GATATC-modification enzyme *Eco*RV is closely related to the GATC-recognizing methyltransferases *Dpn*II and dam from E. coli and phage T4. *FEBS Lett.* 220:167–76
127. Lauster, R., Trautner, T. A., Noyer-Weidner, M. 1989. Cytosine-specific type II DNA methyltransferases: a conserved enzyme core with variable target-recognizing domains. *J. Mol. Biol.* 206:305–12
128. Lieb, M. 1987. Bacterial genes mutL, mutS, and dcm participate in repair of mismatches at 5-methylcytosine sites. *J. Bacteriol.* 169:5241–46
129. Lieb, M. 1985. Recombination in the λ repressor gene: evidence that very short patch (VSP) mismatch correction restores a specific sequence. *Mol. Gen. Genet.* 191:118–25
130. Linder, P., Doelz, R., Gubler, M., Bickle, T. A. 1990. An anticodon nuclease gene inserted into a hsd region encoding a type I DNA restriction system. *Nucleic Acids Res.* 18:7170
131. Linn, P. M., Lee, C. H., Roberts, R. J. 1989. Cloning and characterization of the genes encoding the *Msp*I restriction modification system. *Nucleic Acids Res.* 17:3001–11
132. Loenen, W. A. M., Daniel, A. S.,

tures and functions of DNA glycosylases. *Mutat. Res.* 236:161–72

186. Sargueil, B., Hatat, D., Delahodde, A., Jacq, C. 1990. In vivo and in vitro analyses of an intron-encoded DNA endonuclease from yeast mitochondria. Recognition site by site-directed mutagenesis. *Nucleic Acids Res.* 18:5659–65
187. Schlagman, S. L., Hattman, S. 1989. The bacteriophage T2 and T4 DNA-methyltransferase (Dam) sequence specificities are not identical. *Nucleic Acids Res.* 17:9101–12
188. Schlagman, S. L., Miner, Z., Feher, Z., Hattman, S. 1988. The DNA [adenine-N6] methyltransferase (Dam) of bacteriophage T4. *Gene* 73:517–30
189. Schneider-Scherzer, E., Auer, B., de Groot, E. J., Schweiger, M. 1990. Primary structure of a DNA (N^6-Adenine)-methyltransferase from *Escherichia coli* virus T1. *J. Biol. Chem.* 265:6086–91
190. Schoner, B., Kelley, S., Smith, H. O. 1983. The nucleotide sequence of the *Hha*II restriction and modification genes from *Haemophilus haemolyticus*. *Gene* 24:227–36
191. Schroeder, C., Jurkschat, H., Meisel, A., Reich, J-G., Krüger, D. 1986. Unusual occurrence of *Eco*P1 and *Eco*P15 recognition sites and counterselection of type II methylation and restriction sequences in bacteriophage T7 DNA. *Gene* 45:77–86
192. Seeber, S., Kessler, C., Götz, F. 1990. Cloning, expression and characterization of the *Sau*3AI restriction and modification genes in *Staphylococcus carnosus* TM300. *Gene* 94:37–43
193. Sharp, P. M. 1986. Molecular evolution of bacteriophages: evidence of selection against the recognition sites of host restriction enzymes. *Mol. Biol. Evol.* 3:75–83
194. Shields, S. L., Burbank, D. E., Grabherr, R., Van Etten, J. L. 1990. Cloning and sequencing the cytosine methyltransferase gene M. *Cvi*JI from *Chlorella* virus IL-3A. *Virology* 176:16–24

194a. Sharp, P. M. 1991. *J. Molec. Evol.* 33:3736-54

195. Shub, D. A., Gott, J. M., Xu, M. Q., Lang, B. F., Michel, F., et al. 1988. Structural conservation among three homologous introns of bacteriophage T4 and the group I introns of eukaryotes. *Proc. Natl. Acad. Sci. USA* 85:1151–55
196. Sladek, T. L., Maniloff, J. 1987. Endonuclease from Acholeplasma laidlawii strain JA1 associated with in vivo restriction of DNA containing 5-methylcytosine. *Isr. J. Med. Sci.* 23: 423–26
197. Sladek, T. L., Nowak, J. A., Maniloff, J. 1986. Mycoplasma restriction: identification of a new type of restriction specificity for DNA containing 5-methylcytosine. *J. Bacteriol.* 165:219–25
198. Slatko, B. E., Benner, J. S., Jager-Quinton, T., Moran, L. S., Simcox, T. G., et al. 1987. Cloning, sequencing and expression of the *Taq*I restriction-modification system. *Nucleic Acids Res.* 15:9781–96
199. Slatko, B. E., Croft, R., Moran, L. S., Wilson, G. G. 1988. Cloning and analysis of the *Hae*III and *Hae*II methyltransferase genes. *Gene* 74:45–50
200. Smith, H. O. 1979. Nucleotide sequence specificity of restriction endonucleases. *Science* 205:455–62
201. Smith, H. O., Annau, T. M., Chandrasegaran, S. 1990. Finding sequence motifs in groups of functionally related proteins. *Proc. Natl. Acad. Sci. USA* 87:826–30
202. Smith, H. O., Kelly, S. V. 1984. Methylases of the type II restriction-modification systems. In *DNA Methylation: Biochemistry and Biological Significance*, ed. A. Razin, H. Cedar, A. D. Riggs, pp. 39–71. New York: Springer-Verlag
203. Smith, H. O., Nathans, D. 1973. A suggested nomenclature for bacterial host modification and restriction systems and their enzymes. *J. Mol. Biol.* 81:419–23
204. Sohail, A., Lieb, M., Dar, M., Bhagwat, A. S. 1990. A gene required for very short patch repair in *Escherichia coli* is adjacent to the DNA cytosine methylase gene. *J. Bacteriol.* 172:4214–21
205. Som, S., Bhagwat, A. S., Friedman, S. 1987. Nucleotide sequence and expression of the gene encoding the *Eco*RII modification enzyme. *Nucleic Acids Res.* 15:313–32
206. Stefan, C., Xia, Y., Van Etten, J. L. 1991. Molecular cloning and characterization of the gene encoding the adenine methyltransferase M. *Cvi*RI from *Chlorella* virus XZ-6E. *Nucleic Acids Res.* 19:307–11
207. Stephenson, F. H., Ballard, B. T., Boyer, H. W., Rosenberg, J. M., Greene, P. J. 1989. Comparison of the nucleotide and amino acid sequences of the *Rsr*I and *Eco*RI restriction endonucleases. *Gene* 85:1–13
208. Stephenson, F. H., Greene, P. J. 1989.

Recognition sequence of restriction endonuclease III from *Hemophilus influenzae*. *J. Mol. Biol.* 92:331–39

158. Osterman, D. G., DePillis, G. D., Wu, J. C., Matsuda, A., Santi, D. V. 1988. 5-Fluorocytosine in DNA is a mechanism-based inhibitor of HhaI methylase. *Biochemistry* 27:5204–10
159. Petrusyte, M., Bitiniaite, J., Menkevicius, S., Klimasauskas, S., Butkus, V., et al. 1988. Restriction endonucleases of a new type. *Gene* 74:89–91
160. Pogolotti, A. L., Ono, A., Subramaniam, R., Santi, D. V. 1988. On the mechanism of DNA-adenine methylase. *J. Biol. Chem.* 263:7461–64
161. Pósfai, J., Bhagwat, A. S., Pósfai, G., Roberts, R. J. 1989. Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.* 17:2421–35
162. Povilionis, P. I., Lubys, A. A., Janulaitis, A. A. 1988. Cloning of the restriction-modification genes of *Bacillus centrosporus* in *Escherichia coli*. *Genetika* 24:210–15
163. Price, C., Bickle, T. A. 1986. A possible role for DNA restriction in bacterial evolution. *Microbiol. Sci.* 3:296–99
164. Price, C., Lingner, J., Bickle, T. A., Firman, K., Glover, S. W. 1989. Basis for changes in DNA recognition by *Eco*R124 and *Eco*R124/3 type I DNA restriction and modification enzymes. *J. Mol. Biol.* 205:115–25
165. Price, C., Shepherd, J. C. W., Bickle, T. A. 1987. DNA recognition by a new family of type I restriction enzymes: a unique relationship between two different DNA specificities. *EMBO J.* 6: 1493–97
166. Quirk, S. M., Bell-Pedersen, D., Belfort, M. 1989. Intron mobility in the T-even phages: high frequency inheritance of group I introns promoted by intron open reading frames. *Cell* 56: 455–65
167. Radman, M., Wagner, R. 1986. Mismatch repair in *Escherichia coli*. *Annu. Rev. Genet.* 20:523–38
168. Raleigh, E. A., Benner, J., Bloom, F., Braymer, H. D., DeCruz, E., et al. 1991. Note on nomenclature to restriction of modified DNA in *E. coli*. *J. Bacteriol.* In press
169. Raleigh, E. A., Murray, N. E., Revel, H., Blumenthal, R. M., Westaway, D., et al. 1988. McrA and McrB restriction phenotypes of some *E. coli* strains and implications for gene cloning. *Nucleic Acids Res.* 16:1563–75
170. Raleigh, E. A., Trimarchi, R., Revel, H. 1989. Genetic and physical mapping of the *mcrA (rglA)* and *mcrB (rglB)* loci of *Escherichia coli* K-12. *Genetics* 122:279–96
171. Raleigh, E. A., Wilson, G. 1986. *Escherichia coli* K-12 restricts DNA containing 5-methylcytosine. *Proc. Natl. Acad. Sci. USA* 83:9070–74
172. Ramalingam, R., Prasad, R., Raja, M. C., Dharmalingam, K. 1990. Cloning and expression of genes encoding RglA (McrA) and RglB (McrB) proteins. See Ref. 1, p. 43
173. Revel, H. R. 1983. DNA modification: glucosylation. In *Bacteriophage T4*, ed. C. K. Mathews, E. M. Kutter, G. Mosig, P. Berget, pp. 156–65. Washington, DC: Am. Soc. Microbiol.
174. Revel, H. R., Luria, S. E. 1970. DNA-glucosylation in T-even phage: genetic determination and role in phage-host interaction. *Annu. Rev. Genet.* 4:177–92
175. Roberts, R. J. 1978. Restriction and modification enzymes and their recognition sequences. *Gene* 4:183–93
176. Roberts, R. J. 1978. Restriction endonucleases. In *Microbiology 1978*, pp. 5–9. Washington, DC: Am. Soc. Microbiol.
177. Roberts, R. J. 1978. Restriction endonucleases: a new role in vivo? *Nature* 271:502
178. Roberts, R. J. 1990. Restriction enzymes and their isoschizomers. *Nucleic Acids Res.* 18: 2331–65 (Suppl.)
179. Roberts, R. J. 1977. The role of restriction endonucleases in genetic engineering. In *Recombinant Molecules: Impact on Science Society*, ed. R. F. J. Beers, E. G. Barrett, pp. 21–32. New York: Raven
180. Roberts, R. J., 1976. Restriction endonucleases. *CRC Crit. Rev. Biochem.* 4:123–64
181. Rosenberg, J. M., McClarin, J. A., Frederick, C. A., Wang, B-C., Grable, J., et al. 1987. Structure and recognition mechanism of EcoRI endonuclease. *Trends Biol. Sci.* 12:395–98
182. Ross, T. K., Achberger, E. C., Braymer, H. D. 1989. Nucleotide sequence of the McrB region of *Escherichia coli* K-12 and evidence for two independent translational initiation sites at the *mcrB* locus. *J. Bacteriol.* 171:1974–81
183. Roy, P. H., Smith, H. O. 1973. DNA methylases of *Hemophilus influenzae* Rd I. Purification and properties. *J. Mol. Biol.* 81:427–44
184. Sain, B., Murray, N. E. 1980. The *hsd* (host specificity) genes of *E. coli* K12. *Mol. Gen. Genet.* 180:35–46
185. Sakumi, K., Sekiguchi, M. 1990. Struc-

complete nucleotide sequence of the *Pst*I restriction-modification system. *J. Biol. Chem.* 259:8015–26

232. Walter, J., Noyer-Weidner, M., Trautner, T. A. 1990. The amino acid sequence of the CCGG recognizing DNA methyltransferase M. *Bsu*FI: implications for the analysis of sequence recognition by cytosine DNA methyltransferases. *EMBO J.* 9:1007–13
233. Warner, H. R., Duncan, B. K. 1978. In vivo synthesis and properties of uracil-containing DNA. *Nature* 272:32–34
234. Wells, R. D., Neuendorf, S. K. 1981. Cleavage of single-stranded viral DNAs by certain restriction endonucleases. In *Gene Amplification and Analysis,* ed. J. G. Chirikjian, pp. 101–11. Holland: Elsevier
235. Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., et al. 1988. Sequential order of target-recognizing domains in multispecific DNA-methyltransferases. *EMBO J.* 7:2601–9
236. Wilson, G. G. 1991. Organization of restriction-modification systems. *Nucleic Acids Res.* 19:2539
237. Wilson, G. G. 1988. Type II restriction-modification systems. *Trends Genet.* 4:314–18
238. Winkler, F. K., Darcy, A., Blocker, H., Frank, R., Vanboom, J. H. 1991. Crystallization of complexes of EcoRV endonuclease with cognate and non-cognate DNA fragments. *J. Mol. Biol.* 217:235–38
239. Wu, J. C., Santi, D. V. 1987. Kinetic and catalytic mechanism of HhaI methyltransferase. *J. Biol. Chem.* 262:4778–86
240. Xu, G., Kapfer, W., Walter, J., Trautner, T. A. 1990. Cloning and characterization of the *Bsu*BI restriction and modification genes—comparison to the *Pst*I restriction and modification system. See Ref. 1, p. 85
241. Yuan, R., Bickle, T. A., Ebbers, W., Brack, C. 1975. Multiple steps in DNA recognition by restriction endonuclease from *E. coli* K. *Nature* 256:556–60
242. Yuan, R., Hamilton, D. L. 1984. Type I and type III restriction-modification enzymes. See Ref. 202, pp. 11–37
243. Yuan, R., Hamilton, D. L., Burckhardt, J. 1980. DNA translocation by the restriction enzyme from *E. coli* K. *Cell* 20:237–44
244. Yuan, R., Smith, H. O. 1984. The restriction and modification DNA methylases: an overview. See Ref. 202, pp. 73–80
245. Zain, B. S., Roberts, R. J. 1977. A new specific endonuclease from *Xanthomonas badrii*. *J. Mol. Biol.* 115:249–55
246. Zebala, J. A., Mayer, A. A., Barany, F. 1990. *Taq*I endonuclease insertion mutants and a comparison of its sequence with the *Tth*HB8I isoschizomer. See Ref. 1, p. 16

Nucleotide sequence of the gene encoding the *Rsr*I methyltransferase. *Nucleic Acids Res.* 17:503
209. Studier, F. W., Bandyopadhyay, P. K. 1988. Model for how type I restriction enzymes select cleavage sites in DNA. *Proc. Natl. Acad. Sci. USA* 85:4677–81
210. Sugisaki, H., Kita, K., Takanami, M. 1989. The FokI restriction-modification system. II: Presence of two domains in *Fok*I methylase responsible for modification of different DNA strands. *J. Biol. Chem.* 264:5757–61
211. Sullivan, K. M., Saunders, J. R. 1989. Nucleotide sequence and genetic organization of the *Ngo*PII restriction-modification system of *Neisseria gonorrhoeae*. *Mol. Gen. Genet.* 216:380–87
212. Suri, B., Nagaraja, V., Bickle, T. A. 1984. Bacterial DNA modification. *Curr. Top. Microbiol. Immunol.* 108:1–9
213. SziláK, L., Venetianer, P., Kiss, A. 1990. Cloning and nucleotide sequence of the genes coding for the *Sau*96I restriction and modification enzymes. *Nucleic Acids Res.* 18:4659–64
214. Sznyter, L. A., Slatko, B., Moran, L., O'Donnell, K. H., Brooks, J. E. 1987. Nucleotide sequence of the *Dde*I restriction-modification system and characterization of the methylase protein. *Nucleic Acids Res.* 15:8249–66
215. Szomolányi, E., Kiss, A., Venetianer, P. 1980. Cloning the modification methylase gene of *Bacillus sphaericus* R in *Escherichia coli*. *Gene* 10:219–25
216. Szybalski, W., Kim, S. C., Hasan, N., Podhajska, A. J. 1991. Class-IIs restriction enzymes—a review. *Gene*. In press
217. Tao, T., Bourne, J. C., Blumenthal, R. M. 1991. A family of regulatory genes associated with type II restriction-modification systems. *J. Bacteriol.* 173:1367–75
218. Tao, T., Walter, J., Brennan, K. J., Cotterman, M. M., Blumenthal, R. M. 1989. Sequence, internal homology and high-level expression of the gene for a DNA-(cytosine N^4)-methyltransferase, M·*Pvu*II. *Nucleic Acids Res.* 17:4161–75
219. Theriault, G., Roy, P. H., Howard, K. A., Benner, J. S., Brooks, J. E., et al. 1985. Nucleotide sequence of the *Pae*R7 restriction/modification system and partial characterization of its protein products. *Nucleic Acids Res.* 13:8441–61
220. Tomilin, N. V., Aprelikova, O. N. 1989. Uracil-DNA glycosylases and DNA uracil repair. *Int. Rev. Cytol.* 114:125–79
221. 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
222. Trautner, T. A., Balganesh, T. S., Pawlek, B. 1988. Chimeric multispecific DNA methyltransferases with novel combinations of target recognition. *Nucleic Acids Res.* 16:6649–58
223. Van Etten, J. L., Xia, Y., Burbank, D. E. 1989. Viruses infecting a eukaryotic green alga are a new source of DNA methyltransferases and DNA site-specific endonucleases. *J. Cell Biochem. Suppl.* 0:199
224. Vanek, P. G., Connaughton, J. F., Kaloss, W. D., Chirikjian, J. G. 1990. The complete sequence of the *Bacillus amyloliquefaciens* strain H, cellular *Bam*HI methylase gene. *Nucleic Acids Res.* 18:6145
225. Varshney, U., Hutcheon, T., van de Sande, J. H. 1988. Sequence analysis, expression, and conservation of *Escherichia coli* uracil DNA glycosylase and its gene (ung). *J. Biol. Chem.* 263:7776–84
226. Waite-Rees, P. A., Keating, C. J., Moran, L. S., Slatko, B. E., Hornstra, L. J., et al. 1991. Characterization and expression of the *Escherichia coli* Mrr restriction system. *J. Bacteriol.* In press
227. Waite-Rees, P. A., Moran, L. S., Slatko, B. E., Hornstra-Coe, L., Benner, J. S. 1990. Cloning, purification and characterization of the *Hinc*II and *Hpa*I restriction and modification systems. See Ref. 1, p. 86
228. Waite-Rees, P. A., Nwankwo, D. O., Moran, L. S., Slatko, B. E., Wilson, G. G., et al. 1991. Cloning, characterization and sequence of the *Hinc*II restriction-modification system and its similarity to the *Hind*II restriction-modification system. *Nucleic Acids Res.* In press
229. Waite-Rees, P. A., Polisson, C., Silber, K. R., Moran, L. S., Slatko, B. E., et al. 1991. Cloning, characterization and sequence of the *Nde*I restriction and modification system. *Nucleic Acids Res.* In press
230. Walder, R. Y., Hartley, J. L., Donelson, J. E., Walder, J. A. 1981. Cloning and expression of the *Pst*I restriction-modification system in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 78:1503–7
231. Walder, R. Y., Walder, J. A., Donelson, J. E. 1984. The organization and