The biology of restriction and anti-restriction
Mark R Tock and David TF Dryden

The phenomena of prokaryotic restriction and modification, as well as anti-restriction, were first discovered five decades ago but have yielded only gradually to rigorous analysis. Work presented at the 5th New England Biolabs Meeting on Restriction-Modification (available on REBASE, http://www.rebase.com) and several recently published genetic, biochemical and biophysical analyses indicate that these fields continue to contribute significantly to basic science. Recently, there have been several studies that have shed light on the still developing field of restriction-modification and on the newly re-emerging field of anti-restriction.

Addresses
School of Chemistry, The King’s Buildings, The University of Edinburgh, Edinburgh, EH9 3JJ, UK

Corresponding author: Dryden, David TF (david.dryden@ed.ac.uk)

Introduction
Bacterial restriction-modification (R-M) systems function as prokaryotic immune systems that attack foreign DNA entering the cell [1]. Typically, R-M systems have enzymes responsible for two opposing activities: a restriction endonuclease (REase) that recognizes a specific DNA sequence for cleavage and a cognate methyltransferase (MTase) that confers protection from cleavage by methylation of adenine or cytosine bases within the same recognition sequence. REases recognize ‘non-self’ DNA (Figure 1), such as that of phage and plasmids, by its lack of characteristic modification within specific recognition sites [2]. Foreign DNA is then inactivated by endonucleolytic cleavage. Generally, methylation of a specific cytosine or adenine within the recognition sequence confers protection from restriction. Host DNA is normally methylated by the MTase following replication, whereas invading non-self DNA is not. However, the ability of phage and some plasmids to acquire host modification and to escape restriction suggests that the R-M barrier is imperfect, and highlights an evolutionary arms race between bacterial genomes (R-M systems) and parasitic DNA molecules (anti-restriction systems) [3].

The prevalence of R-M enzymes among eubacteria and archaea [4**] indicates that they might have more than one role. Possibilities include defence against newly encountered phage [1], the acquisition of beneficial genetic code at low energetic cost [5], or the propagation of selfish genetic elements [6].

In this review, we will describe how R-M systems are classified into four major groups according to their subunit composition, recognition site, cofactor requirement and cleavage position [7**]. Table 1, and the various anti-restriction systems deployed by bacterial parasites, Table 2.

Restriction-modification systems
Type I restriction-modification enzymes
Type I R-M enzymes are hetero-oligomeric complexes that typically contain two REase subunits (R) that are required for DNA cleavage, one specificity subunit (S) that specifies the DNA sequence recognized, and two MTase subunits (M) that catalyse the methylation reaction [8,9]. Depending upon the methylation status of DNA, this complex can function as either an REase or an MTase. Unmethylated DNA is targeted for restriction, hemi-methylated molecules are targeted for further methylation, and fully methylated DNA is immune to restriction [10] (Figure 1). The MTases use S-adenosyl methionine (SAM) to methylate the N\(^6\) position of adenine within bipartite, asymmetrical, target sequences [11]. The restriction reaction requires SAM, ATP and Mg\(^{2+}\). Cleavage occurs at a site that is distant from the recognition sequence and it is preceded by ATP-dependent DNA translocation during which the REases remain attached to their recognition site [12–15,16**,17–19]. Cleavage is prompted by stalled translocation; for example, when two translocating enzymes collide, each cuts one strand of the duplex [18].

Type I R-M systems are widely spread in prokaryotes, with the discovery of approximately 600 putative enzymes to date. An example is EcoKI, which was discovered because of its ability to limit phage propagation by factors of between \(10^3\) and \(10^8\) [19]. On the basis of genetic complementation and molecular evidence, four families of Type I R-M enzymes have been determined (IA–ID) [20]; a fifth family, Type IE R-M, was recently proposed [21].

Type II restriction-modification enzymes
Type II R-M enzymes are an ever-expanding collection of over 3650 different R-M systems. The MTases share several conserved amino acid motifs, but the REase proteins contain such dissimilar amino acid sequences and
different behaviours that they are classified into 11 overlapping subclasses [2,7**]. Most, but not all, type II R-M systems contain separate REase and MTase enzymes. Restriction is usually Mg$^{2+}$-dependent and the MTase requires SAM as a methyl donor. The REase and MTase recognize the same DNA sequence, which is typically a 4–

---

**Figure 1**

The function of R-M systems, as illustrated by Type I R-M enzymes. These enzymes recognize the methylation state of their specific target sequence. Fully methylated DNA (shown as two green circles on the target sequence on the host DNA) is recognized to be part of the bacterial genome. Hemimethylated DNA (a single green circle on host DNA target sequence) is recognized as newly replicated bacterial DNA, and the MTase (M) modifies the other strand by methylation using the cofactor SAM. However, invading DNA, for example a phage genome, generally lacks specific modification (red circles on the target sequence of phage DNA) and is recognized to be foreign by the REase (R) and cleaved into harmless fragments.

---

**Table 1**

<table>
<thead>
<tr>
<th>Characteristics and organization of the genetic determinants of different classes of R-M systems.</th>
<th>Type I</th>
<th>Type II</th>
<th>Type III</th>
<th>Type IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example R-M system</td>
<td>EcoKI</td>
<td>EcoRI</td>
<td>EcoP1I</td>
<td>EcoMcrBC</td>
</tr>
<tr>
<td>Genes</td>
<td>hasdR, hasdM, hasdS</td>
<td>hasdR, hasdM</td>
<td>hasdR, hasdM</td>
<td>mcrB, mcrC</td>
</tr>
<tr>
<td>Subunits</td>
<td>Three different subunits (R, M and S) combine to form R$_3$M$_2$S$_1$ and M$_1$S$_1$</td>
<td>Two different subunits (R and M) combine to form R$_3$ or M$_1$</td>
<td>Two different subunits (mod and res) combine to form mod$_2$res$_2$</td>
<td>Two different subunits are present, McrB and McrC</td>
</tr>
<tr>
<td>Enzyme activities</td>
<td>REase, MTase and ATPase</td>
<td>REase or MTase</td>
<td>REase, MTase and ATPase</td>
<td>REase and GTPase</td>
</tr>
<tr>
<td>Co-factors required for DNA cleavage</td>
<td>ATP, SAM, Mg$^{2+}$</td>
<td>Mg$^{2+}$</td>
<td>ATP, Mg$^{2+}$ (SAM)</td>
<td>GTP, Mg$^{2+}$</td>
</tr>
<tr>
<td>Co-factors required for methylation</td>
<td>SAM</td>
<td>SAM</td>
<td>SAM</td>
<td>No methylation</td>
</tr>
<tr>
<td>Recognition sequence</td>
<td>Asymmetric and bipartite, e.g. EcoKI, 5’AAC(N$_5$–4000)AGACC</td>
<td>Mostly symmetric, e.g. EcoRI, 5’GAATTC</td>
<td>Asymmetric, e.g. EcoP11, 5’AGACC</td>
<td>Bipartite and methylated, e.g. EcoMcrBC, 5’RmC(N$_{50}$–4000)RmC</td>
</tr>
<tr>
<td>Cleavage site</td>
<td>Variable locations 1000 bp from recognition site</td>
<td>Fixed location at or near the recognition site</td>
<td>Fixed location 25–27 bp from recognition site</td>
<td>Between methylated bases at multiple sites</td>
</tr>
<tr>
<td>DNA translocation</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*a A full description of R-M classifications is given in [7**].
8 base pair (bp) palindrome. All Type II REases cleave within or adjacent to this specific DNA sequence to generate a defined restriction pattern of products that have both a 5'-PO4 and a 3'-OH termini [2]. Some Type II REases are active as homodimers, with each monomer cutting one strand in a co-ordinated fashion to generate double-strand breaks. Other Type II REases are able to act as monomers or tetramers, and there is evidence that many Type II REases must bind to two or more copies of their recognition site before the DNA is cleaved [22,23**]. Type II REases generally act as monomers to modify a specific base of their recognition sequence on each strand of the duplex. Methylation of a cytosine occurs at either the N4 or the C6 position, and methylation of adenine occurs at the N6 position [24].

It is known that the expression of some Type II R-M systems is controlled by repressor-like proteins known as C-proteins. For example, a C-protein has been demonstrated to activate expression of the PvuII restriction gene [25], and the structure of the C-protein from the AhdI R-M system was recently solved [26*].

**Type III restriction-modification enzymes**

Type III R-M enzymes are less complex but share many similarities with Type I R-M enzymes [8]. They are hetero-oligomers that consist of a ‘mod’ subunit, which is required for substrate recognition and modification, and a restriction subunit (res), which is only active when associated in a res3mod2 complex [27]. Modification requires SAM and restriction is dependent on the presence of both Mg2+ and ATP. For cleavage to occur, a Type III R-M enzyme must interact with two inversely oriented copies of its 5–6 bp asymmetric recognition sequence [28]. As with Type I R-M enzymes, cleavage is preceded by DNA translocation during which two res3mod2 complexes maintain contact with their recognition sequence [8,29]. Stalled DNA translocation and/or collision of res3mod2 complexes initiates cleavage by each monomer at a point that is 25–27 bases from the recognition site; one DNA strand is cut by each complex. In all known cases, Type III mod subunits are able to act independently of their cognate res subunits and can methylate the N6 position of adenine on only one strand to produce a hemi-methylated DNA molecule. Since only completely unmethylated DNA is cut by Type III REases, this methylation confers protection.

To date, Type III R-M systems have been identified almost exclusively in phage and in Gram-negative bacteria, with the best studied examples being EcoP11 and EcoP15I [30]. Type III R-M enzymes are encoded by contiguous genes that have a high degree of sequence similarity between different systems. Approximately 130 different putative Type III R-M systems have been identified throughout the prokaryotic genomes that have been sequenced, which suggests that these systems might be widely spread in these organisms.

**Type IV restriction-modification enzymes**

Type IV R-M enzymes are REases that will only cleave DNA substrates that have been modified, for example bases that have been methylated, hydroxymethylated and glucosyl-hydroxymethylated. This group contains 227 putative enzymes. The best studied example within this group, MerBC from *E. coli* K12 [31], shows homology to the AAA+ motor proteins such as the DnaA and RuvB proteins that are involved in DNA replication. MerBC requires Mg2+ for activity and is the only known nuclease to use GTP for cleavage and for translocation of DNA [32]. This enzyme detects two copies of a dinucleotide sequence, consisting of a purine followed by a cytosine.
methylated at either the N4 or the C5 position, which are separated by between 40 and 3000 nucleotides, and preferentially cuts 30 bp away from one of the sites [33]. As in Type I R-M systems, McrBC remains bound to its recognition sites during translocation, the stalling of which initiates cleavage.

**Anti-restriction strategies**

The detection of phage that have acquired modification for protection from the host and the apparent promiscuity of conjugal plasmids begs the question of how they evade destruction by bacterial R-M systems. Although evidence suggests that R-M systems are an imperfect barrier to invasion by foreign DNA, it is clear that phage and plasmids employ additional strategies to avoid restriction [3,30,34] (Table 2). Mechanisms that involve modification of the phage genome, transient occlusion of restriction sites, subversion of host R-M activities, and direct inhibition of restriction enzymes are reviewed below.

**DNA sequence alteration**

Changes to the DNA sequence that remove recognition sites from phage and plasmid genomes enable them to evade restriction [3,30]. In some cases, a reduction in the number of recognition sites is sufficient to enable phage to avoid cleavage. For example, EcoRII must bind to two copies of its target sequence before cleavage can occur, but the distance between EcoRII sites in the phage T3 and phage T7 genomes is so large that the DNA of these phage is resistant to cleavage [22,30]. It seems highly probable that phage have lost restriction sites as a result of counter selection imposed by host R-M systems.

The phage T7 genome contains another feature that enables it to avoid restriction by Type III R-M enzymes: all of the EcoP1I sites in phage T7 DNA are in the same orientation rather than in the head to head formation that is required for cleavage [28].

Some phage also incorporate unusual bases within their DNA as protection against restriction in hosts that carry appropriate R-M systems. For example, many *B. subtilis* phage replace thymine with 5-hydroxymethyluracil [35], and T-even phage genomes contain the unusual base hydroxymethylcytosine [3]. The product of the mom gene protects phage Mu by converting adenine within Type I and Type II recognition sites into N6-(1-acetamido) adenine [4**]. To avoid restriction, several phage genomes encode MTases that modify and protect phage DNA within the bacterial host. For example, SPβ phage encodes an MTase that modifies bases within *B. subtilis* BsuRI recognition sites [35].

**Transient occlusion of restriction sites**

Transient occlusion of restriction sites by phage- and plasmid-encoded DNA-binding proteins enables these DNA molecules to gain modification to protect them from the host. Phage P1 accessory proteins encoded by its *darA* and *darB* genes are present within the phage head. These accessory proteins, DarA and DarB, are co-injected with the DNA into the bacteria and occlude Type IA and Type IB restriction sites by binding to the phage DNA. In this way, P1 DNA avoids restriction by Type I R-M enzymes [36]. An interesting parallel can be drawn between these accessory proteins and the ArsC protein encoded by the IncW plasmid pSa that protects the incoming T-strand during bacterial conjugation [34]. This type of co-transport of anti-restriction protein and DNA is known as Type IV secretion [37], an important pathway in the movement of bacterial virulence effectors into human cells.

**Subversion of restriction-modification activities**

Bacterial R-M activities can be subverted by way of two routes: by stimulation of host MTases to modify phage DNA or by destruction of REase cofactors. An example of the first method is in phage λ Ral protein, which alleviates restriction by stimulating the activity of Type IA MTases [38]. The hybrid phage, λ-ρVS, also encodes an analogous Ral function known as Lar, which both alleviates restriction and enhances modification by the EcoKI Type IA system [20].

As Type I and Type III R-M enzymes require SAM for activity, reduction of the intracellular concentration of this cofactor was found to provide a second method by which to alleviate restriction. This second route has been adopted by phage T3, which encodes a SAM hydrolase that destroys intracellular SAM soon after infection [39]. This does not inhibit Type I and Type III R-M enzymes already bound to SAM, but does prevent newly synthesized enzymes from acquiring its cofactor. SAM hydrolase has also been shown to improve the chances of phage survival in cells that contain the Type III EcoP1 system [40].

**Inhibition of Type I restriction-modification enzymes**

Inhibition of Type I R-M enzymes by direct interaction is the mechanism employed by the most extensively studied anti-restriction protein — the gene 0.3 protein of phage T7 (also known as the overcome classical restriction [Ocr] protein). The first product to be expressed by phage T7 as it enters a bacterium is Ocr [40,41]. It blocks the DNA binding site of resident Type I R-M enzymes, which enables the phage to propagate [42]. However, Ocr is inactive against Type II R-M enzymes. Bioinformatics searches reveal the presence of one Ocr homologue in the Yersinia phage ΦA1122A [43]. As shown in Figure 2, the active Ocr dimer has similar dimensions to those of DNA. By mimicking the size, shape and electrical charge of 24 bp of bent B-form DNA, Ocr prevents DNA from binding to the Type I EcoK1 RjMs21 complex. The binding of Ocr inhibits both the restriction and the modification ability of this enzyme as the affinity of
EcoKI is at least 50-fold greater for Ocr than for DNA
[44]. The only inhibitor anti-restriction protein of which
the structure is known is Ocr. Clearly, this protein has
potential as a tool for mapping the structure of Type I
R-M enzymes for which there is currently no published
atomic structure (see Update).

Conjugative plasmids and transposons, such as IncN plas-
mids and the Tn916 transposon, encode anti-restriction
proteins to allow them to escape restriction. Two such
families of proteins often carried on the same plasmid are
the ArdA (alleviation of restriction of DNA A) and ArdB
systems, which are among the first proteins to be expressed
by an immigrant plasmid as it enters a recipient cell during
conjugation [34]. These proteins transiently block REases,
allowing the plasmid to acquire protective modification.
Although, the exact mechanism by which this is done is not
clear, there is some evidence to suggest that some ArdA
proteins target both the REase and MTase of Type I R-M
systems [45]. The mechanisms by which some ArdA and all
ArdB proteins inhibit Type II R-M enzymes remain
unclear. ArdA and ArdB proteins, like Ocr, are highly acidic
with a net negative charge. They have a putative anti-
restriction motif [46,47]. Combined with secondary struc-
ture predictions that identify broad similarities between
Ard proteins and Ocr, this suggests that Ard proteins might
also mimic aspects of the DNA structure.

Intriguingly, the IncP plasmid-encoded ArdB homologue
KleA lacks a putative anti-restriction motif [48], which
suggests that this might not always be required for anti-
restriction activity. The plasmid-encoded MerR-like pro-
teins involved in bacterial tolerance of heavy metals do
contain the putative anti-restriction motif, but it is dis-
rupted upon binding of heavy metals leading to a loss of
anti-restriction activity [49]. The significance of this
observation is not entirely clear given that the primary
function of these proteins is probably not anti-restriction.

Figure 2

DNA mimicry by the phage T7 anti-restriction protein Ocr (shown in a blue ribbon cartoon form; protein database code: 1S7Z). (a) For each
monomer of Ocr a superposition of the phosphate groups of a 12 bp B-DNA complex (protein database code: 1BNA) was made onto 12 carboxyl
groups of Ocr, with a deviation of only 1.9 Å between phosphate and carboxyl groups (the position of the pairs of amino acid side-chains and
phosphates used in the fit are labelled). The phosphorus atoms are coloured yellow and oxygen atoms of the phosphate groups are coloured purple.
The 12 carboxyl groups are coloured red (oxygen) and black (carbon). The sugar backbones of the DNA dimer chains are coloured in two
shades of green, and the base pairs are omitted for clarity. In (b) the view is rotated by 90° with respect to panel (a) and is coloured as in (a).
The twofold axis lies in the plane of the paper. The vectors describing the direction of the fitted DNA on both halves of the dimer are drawn
as black lines. Their intersection gives a bend angle of 33.6°. Reprinted from [44] with permission.
Unknown methods of restriction evasion

The mechanisms used by several phage to evade restriction remain unresolved. The availability of genome sequences, such as that recently completed for phage T5 [50], are likely to reveal the presence of potential anti-restriction mechanisms. The DNA molecule of phage T5 is insensitive to cleavage by Type I, II and III R-M enzymes [51,52]. In the cases of the Type II EcoRV and Type II EcoRII R-M systems this is not owing to a lack of restriction sites, but is because of the inability of EcoRII and EcoRV MTases to modify wild-type T5 DNA in vivo [52]. This suggests that phage T5 encodes an anti-restriction mechanism that prevents Type II REases and MTases from connecting with the DNA. The method of inhibition of the Type I and III R-M systems is not known, but the phage T5 genome sequence encodes a putative clp protease — an enzyme shown to be involved in alleviation of restriction by some Type I R-M enzymes [53*]. Intriguingly, Mu phage also encodes a clp protein that is implicated in the degradation of the Mu repressor [54]. This might suggest that phage-encoded anti-restriction proteases have a dual role. Phage T5, as well as other instances reviewed by Kruger and Bickle [3], still require analysis.

Conclusions

Bacteria have clearly evolved R-M systems as a defence against invading selfish DNA molecules. The importance of these systems is confirmed by their wide distribution and huge diversity. These systems are a leaky barrier that allows some foreign DNA to enter the cell; however, this might represent a mechanism by which bacteria acquire new genes. Knowledge of this mechanism is important as it should enhance our understanding of horizontal gene transfer associated with the spread of antibiotic resistance genes within the bacterial fauna.

In response to R-M systems, phage and conjugational plasmids have evolved several anti-restriction measures to ensure their survival. Among these, anti-restriction proteins, especially the DNA-mimic Ocr, that directly interact with R-M enzymes to cause inhibition have huge potential as models of protein–nucleic acid and protein–protein interactions and in the production of therapeutic or diagnostic agents. The study of additional anti-restriction proteins might even reveal inhibitors suitable for use in treating bacterial pathogens that are currently resistant to genetic manipulation. Plasmid-encoded proteins that have broad similarities to Ocr might be ideal candidates for such work.

Update

Kim et al. [55] have recently published the structure of a sequence specificity subunit from a putative Type I R-M system. The structure confirms all of the predictions made about the structures of these subunits and their relationship with the other subunits and with the DNA substrate (reviewed in [9]).

Acknowledgements

We would like to acknowledge the following funding bodies for supporting our work on R-M and anti-restriction systems: the Biotechnology and Biological Sciences Research Council (BBSRC), the Engineering and Physical Sciences Research Council (EPSRC), the Medical Research Council (MRC), the Leverhulme Trust and the Darwin Trust. We also thank Noreen Murray and Malcolm Walkinshaw for many fruitful discussions about R-M and anti-restriction.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest


This is a database of information concerning all aspects of R-M systems.


A detailed article that defines R-M systems.


The first application of single molecule force and kinetic measurements (as opposed to many previous single-molecule imaging experiments) to the study of DNA translocation by R-M enzymes.


54. The control of restriction by Type I R-M systems in vivo is shown to rely upon proteolysis. The presence of possible proteases in phage genomes might provide an anti-restriction mechanism for these phage.