Nucleoside triphosphate-dependent restriction enzymes

Citation for published version:
Dryden, DTF, Murray, NE & Rao, DN 2001, 'Nucleoside triphosphate-dependent restriction enzymes'
Nucleic Acids Research, vol. 29, no. 18, pp. 3728-3741. DOI: 10.1093/nar/29.18.3728

Digital Object Identifier (DOI):
10.1093/nar/29.18.3728

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Nucleic Acids Research

Publisher Rights Statement:
Copyright 2001 Oxford University Press; all rights reserved.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
SURVEY AND SUMMARY

Nucleoside triphosphate-dependent restriction enzymes

D. T. F. Dryden*†, Noreen E. Murray† and D. N. Rao‡

Department of Chemistry, University of Edinburgh, Joseph Black Building, The King’s Buildings, Mayfield Road, Edinburgh EH9 3JJ, UK, †Institute of Cell and Molecular Biology, University of Edinburgh, Darwin Building, The King’s Buildings, Mayfield Road, Edinburgh EH9 3JR, UK and ‡Department of Biochemistry, Indian Institute of Science, Bangalore 560 012, India

Received February 5, 2001; Revised and Accepted March 26, 2001

ABSTRACT

The known nucleoside triphosphate-dependent restriction enzymes are hetero-oligomeric proteins that behave as molecular machines in response to their target sequences. They translocate DNA in a process dependent on the hydrolysis of a nucleoside triphosphate. For the ATP-dependent type I and type III restriction and modification systems, the collision of translocating complexes triggers hydrolysis of phosphodiester bonds in unmodified DNA to generate double-strand breaks. Type I endonucleases break the DNA at unspecified sequences remote from the target sequence, type III endonucleases at a fixed position close to the target sequence. Type I and type III restriction and modification (R-M) systems are notable for effective post-translational control of their endonuclease activity. For some type I enzymes, this control is mediated by proteolytic degradation of that subunit of the complex which is essential for DNA translocation and breakage. This control, lacking in the well-studied type II R-M systems, provides extraordinarily effective protection of resident DNA should it acquire unmodified DNA. The only well-documented GTP-dependent restriction enzyme, McrBC, requires methylated target sequences for the initiation of phosphodiester bond cleavage.

INTRODUCTION

Restriction–modification (R-M) systems are composed of pairs of opposing enzyme activities: an endonuclease and a DNA methyltransferase (mtase). The endonucleases recognise specific sequences and catalyse cleavage of double-stranded DNA. The modification mtases catalyse the addition of a specific sequences and catalyse cleavage of double-stranded DNA methyltransferase (mtase). The endonucleases recognise pairs of opposing enzyme activities: an endonuclease and a restriction–modification (R-M) systems are composed of enzymes that form one functional holoenzyme with both restriction and and modification functions carried out by the same enzyme. These enzymes have been reviewed in detail (1–4).

Type I systems are the most complex with both restriction and modification functions carried out by the same enzyme. They comprise three subunits, named Hsd for host specificity for DNA, which can form one holoenzyme. These enzymes recognise asymmetric bipartite sequences and cleave DNA thousands of base pairs away from the recognition sequence (5,6).

Type III R-M systems consist of two subunits (Mod and Res) that form one functional holoenzyme with both restriction and modification activity. This class of enzymes recognises specific asymmetric sequences and cuts DNA at a fixed distance 25–27 bp to one side of the recognition site (6,7).

The above-mentioned R-M systems restrict unmodified DNA, but there are other systems that specifically recognise and cut modified DNA. These modification-dependent restriction systems (MDRS) have no associated mtase. McrBC, a well-characterised member of MDRS, is a complex nucleotide-requiring enzyme, comprising two types of subunits, which cuts DNA up to 3000 bp away from the target site sequence. Methylated DNA is an absolute requirement for DNA cleavage.

This review deals with nucleoside triphosphate-dependent restriction enzymes, which are being investigated as model systems to understand how proteins communicate with two DNA binding sites separated by large distances. Nucleoside triphosphate-dependent restriction enzymes, type I, type III R-M systems and McrBC, have been shown to possess either ATP or GTP hydrolysis activity. These enzymes usually interact with two DNA binding sites that can be separated by several thousand base pairs of DNA. In all these systems, the proteins have been shown to remain bound to the recognition site while translocating DNA past themselves (Fig. 1). Translocation by type I and type III systems requires the presence of DEAD-box

*To whom correspondence should be addressed. Tel: +44 131 650 4735; Fax: +44 131 650 6453; Email: david.dryden@ed.ac.uk
Translocation occurs from both sides of the target sequence for type I enzymes. DNA translocation, driven by ATP hydrolysis, starts and pulls DNA in towards the enzyme (thin arrow), which remains bound to its target sequence. Translocation occurs from both sides of the target sequence for type I enzymes.

**Figure 1.** A model of DNA translocation for type I and type III DNA restriction enzymes and MDRS (based upon ref. 95). A single enzyme molecule (oval) binds to its DNA target sequence (black rectangle) within a longer stretch of DNA. DNA translocation, driven by ATP hydrolysis, starts and pulls DNA in towards the enzyme (thin arrow), which remains bound to its target sequence. Translocation occurs from both sides of the target sequence for type I enzymes.

Figure 1

amino acid motifs, more commonly associated with DNA and RNA helicases. DNA cleavage is believed to be triggered when two translocating enzyme complexes collide or are stalled.

**TYPE I R-M ENZYMES**

**Overview**

Type I R-M enzymes are complex, oligomeric proteins comprising three types of subunits, HsdS (∼50 kDa), HsdM (50–60 kDa) and HsdR (∼140 kDa) (5,6,8–11). Those type I systems that have been examined genetically and biochemically in enteric bacteria have been grouped into four families, IA, IB, IC and ID, based upon genetic complementation, DNA hybridisation and antibody cross-reactivity (9). The archetypal members of the type IA are EcoBI and EcoKI, EcoAI for type IB, EcoR124I and EcoR124II for type IC, and StyBLI for ID. Many new type I systems are being identified or postulated from genomic sequences of non-enteric bacteria and archaea. These new systems may not fit within the experimental definition of the known families.

**Subunits of type I R-M enzymes**

**HsdS.** HsdS specifies the DNA target sequence for the enzyme and serves as a core subunit to which the others bind. HsdS contains two target recognition domains (TRDs) of ∼150 amino acids each, the N-terminal TRD recognises the 5′ part of the bipartite target sequence and the C-terminal TRD recognises the 3′ part of the target (12–16). All known target sequences for type I R-M systems comprise a 3 or 4 bp 5′ sequence separated by 6–8 non-specific bp from a 4 or 5 bp 3′ sequence. The amino acid sequences of TRDs are poorly conserved between HsdS subunits from different systems, showing <30% identity unless they recognise the same target sequence when levels of identity are much higher. This level of identity, when coupled with predictions of secondary structure, suggested that the TRDs had a common tertiary structure with a DNA binding region similar in structure to that found in the crystallographic structure of the TRD of the type II mtase HhaI (17). The predicted DNA interaction region, comprising a loop–β-strand–loop region, of the TRD has received strong support from random and site-directed mutagenesis (18,19). The TRDs are separated by a short amino acid spacer region which is highly conserved between systems within a type I family and less strongly between families. The similarity between different families in this region is mostly confined to short sequences, which also show similarity to sequences at the N- and C-termini of the subunits (13,20–22). This similarity suggested that the HsdS subunits possess ∼2-fold rotational symmetry (22,23) and this has been confirmed by the construction of various deletion and fusion derivatives of hsdS (24–27), and the finding that the N-terminal TRD of StySKI, with a similar amino acid sequence to the C-terminal TRD of EcoR124I, recognises a DNA sequence complementary to that of the C-terminal TRD of EcoR124I (16). HsdS varies in solubility in vitro depending upon the type I system; the HsdS subunit of EcoKI is insoluble (28), but those of EcoR124I and EcoAI are soluble although that of EcoR124I is only soluble as fusion protein (27,29).

HsdM. The HsdM subunits are responsible for binding the methylation cofactor AdoMet, determining the methylation status of the target sequence and carrying out methylation of adenine bases. Methylation is probably achieved using a base-flipping mechanism in which the target base is rotated 180° out of the DNA helix into the enzyme catalytic site (30,31). For all type I R-M enzymes with known target sequences, one adenine targeted for methylation is located within the top strand of the DNA in the 5′ part of the target and the second adenine is located within the bottom strand in the 3′ part of the target sequence. If both adenines are unmethylated, the restriction reaction is triggered if the HsdR subunits are present. Hemimethylated DNA is the preferred target for rapid methylation by EcoKI (32,33) and EcoR124I (34), but EcoAI rapidly methylates unmethylated and hemimethylated targets (32). The EcoKI and EcoR124I mtases take hours to produce measurable methylation of unmethylated DNA.

A complex including two HsdM subunits appears to be required for methylation activity and the detection of methylation status (27,33,35). Proteins containing only one HsdS and one HsdM can be prepared with the EcoKI system by relying on the unstable nature of the mtase, but this dimer is inactive as an mtase even though it can still bind to its target sequence (33,36). HsdM contains amino acid motifs common to all adenine mtases (37,38). Mutations affecting two particularly well-conserved motifs, I and IV, cause loss of AdoMet binding and loss of catalysis, respectively (39). These defects cause loss of mtase function as found for mtases from type II R-M systems. These motifs are found in the central third of the subunit and form a domain whose structure has been modelled upon the catalytic domain of mtases of type II R-M systems (40). The first third of EcoKI HsdM contains amino acids involved in establishing the preference of this system for methylating hemimethylated DNA (41). Mutations cause this discrimination between hemi and unmethylated DNA to be reduced. The C-terminal domain appears to be at least partly
The HsdR subunit. The addition of two HsdR subunits to the mtase (M2S1) completes the type I enzyme and confers the ability to cleave DNA containing unmethylated sites (44). Cleavage occurs at locations remote from the target sequence in a process involving translocation of DNA (45–47). The enzyme remains attached to the target site and large loops of DNA are produced by the translocation process (48,49).

HsdR comprises several domains defined by limited proteolysis, sequence analysis and mutational analysis for EcoKI (50) and EcoAI (51). Near the N-terminus, there is a proteolytically defined domain of 400 amino acids containing an amino acid motif common to all endonucleases. A C-terminal domain of 300 amino acids is apparently involved in contacting the mtase core. In between the endonuclease domain and the C-terminal domain is a region with sequence similarity to domains IA and 2A of DNA and RNA helicases (52). This region contains the DEAD-box amino acid motifs (53–58) implicated in ATP binding, ATP hydrolysis and DNA translocation. Extensive sequence analysis and mutational experimentation have demonstrated the importance of all seven conserved DEAD-box motifs for ATP hydrolysis and DNA translocation. DNA cleavage is affected indirectly (50,57,58). The presence of these helicase motifs implies that type I enzymes use a helicase activity to move DNA past the enzyme until the cleavage site is reached (54). However, there is no evidence as yet that a helicase mechanism involving DNA strand-separation is actually operational in type I enzymes although the DNA movement (translocation) is clearly revealed by DNA cleavage assays on linear and circular DNA (44,50,59,60) and on DNA containing Holliday junctions (61) as well as the displacement of short oligonucleotides from triplex regions of DNA (62), electron microscopy (48,49) and atomic force microscopy (63,64) of protein–DNA complexes and, most significantly, by in vivo measurements of the entry of phage T7 DNA into cells in which the EcoKI system provides the only means of entry for the phage DNA (58,65).

Assembly and control of biochemical activity. The R subunits bind to the mtase trimer with varying affinities depending upon the type I system; EcoKI forms a stable pentamer (44), EcoR124I forms a stable tetramer R3M2S1 with the second HsdR binding weakly (66) while in EcoAI, neither of the HsdR subunits appears to bind strongly to the mtase trimer (27,32). This assembly process may play a role in the control of activity in vivo where mtase activity appears prior to restriction activity upon establishment of a type I system in a new host (44,66). However, as discussed later, the role of assembly in the control of the EcoKI system is likely to be less important than for those systems, such as EcoR124I and EcoAI, in which one or both R subunits are weakly bound to the mtase. In these cases, the relative level of mtase and restriction activities will change as the concentration of subunits builds up in a new host. The mtase will assemble before the entire enzyme, establishing modification before restriction. The entire enzyme will only form once sufficient subunits have been synthesised for their concentration to exceed the dissociation constant for binding HsdR.

DNA translocation and cleavage experiments. The restriction reaction is a multi-step process requiring ATP binding, ATP hydrolysis for translocating DNA past the enzyme and Mg2+ for DNA cleavage (67). The addition of HsdR to the EcoKI mtase core enhances the DNA binding affinity in the presence of cofactors but curiously in their absence DNA specificity is absent and the enzyme binds well to any DNA sequence (68). This good binding to non-specific DNA may allow the enzyme to diffuse linearly along the DNA looking for its target sequence (69). The footprint of EcoKI changes on cofactor binding, suggesting a change in disposition of the HsdR relative to the mtase core (68,70). This change is not observed for EcoR124I (29).

DNA translocation driven by ATP hydrolysis occurs prior to cleavage even on supercoiled substrates and does not appear to depend on relaxation of topological stress by a nicking or topoisomerase activity. Mutations in the DEAD-box motifs abolish translocation but leave some single-strand nicking activity (57,58). This nicking activity is very weak compared with the cleavage activity of the native enzyme. Mutants in the endonuclease motif which cannot cleave or nick DNA can still translocate DNA as effectively as the native enzyme in vivo (58). This conclusion was drawn from experiments in which EcoKI promotes the transfer of the T7 chromosome from the phage particle to the bacterium (Fig. 2).

The efficiency with which a DNA molecule is cleaved by type I restriction enzymes depends upon the number of target sites present on the DNA. The presence of two or more unmethylated target sites is required for cleavage of linear DNA but circular DNA can be effectively cleaved even if only one site is present (27,43,44,46,50,59–61). If the circular DNA containing one target site is concatenated with a DNA circle lacking a target site, only the DNA containing the target site is cleaved (71). Cleavage occurs roughly half way between two successive target sites on linear or circular DNA if the sites are recognised by the same enzyme. This is particularly evident for the EcoKI system but the type I enzymes, EcoR124I and EcoAI, can also cut near to a target site sequence. Type I enzymes from different families can cooperate to cut linear DNA containing one site for each enzyme but the cleavage position, although not well defined, is not necessarily half way between sites suggesting different translocation rates. Cleavage can also be induced by the presence of complex branched DNA structures on DNA containing one site (61) but not by other non-type I proteins bound to the DNA between two type I target sites (59). The cleavage occurs via two successive nicking reactions. A strong block to translocation appears to be the trigger for DNA cleavage. This block can be the collision of two translocating type I enzymes or a fixed DNA junction (61) (Fig. 3).

Models of DNA translocation and cleavage
The collision model, originally proposed by Studier and Bandyopadhyay (72), has two type I enzymes bound to the same piece of DNA at different unmethylated sites (Fig. 3).
The enzymes remain at these sites and pull in DNA from both sides simultaneously. This generates expanding loops of DNA coming out from the enzyme–DNA complex and these loops have been visualised in both relaxed and highly twisted forms by electron microscopy (48,49). The bi-directional translocation would be expected to produce two expanding loops but usually only one loop was visible in the electron micrographs. This may indicate that the loops are not stable and it has recently...
been found that positive supercoiling induced by the action of EcoAI can only be maintained as long as ATP hydrolysis continues (73). Therefore, electron microscopy sample preparation may not be able to trap all of the looped structures. Observation, using atomic force microscopy, of EcoKI bound to DNA containing two target sequences in the absence of ATP failed to show individual protein molecules bound to the two sites (63,64). Instead, the two proteins had dimerised and collapsed the DNA into a small volume. The addition of ATP then caused further movement in the DNA until cleavage occurred. This suggests that collision of the enzymes occurs prior to ATP hydrolysis and DNA translocation. This could occur through normal diffusional processes of a DNA chain. The addition of ATP to this dimer bound at two sites on the DNA would then cause translocation and production of expanding DNA loops but the DNA between the two sites would now be on a contracting loop. Cleavage would occur when the loop could contract no further. This model is formally equivalent to the collision model (72) but overcomes the problem of requiring the two translocating complexes to move closer and closer together while dragging expanding loops of DNA behind themselves through a cytoplasm crowded with other macromolecules (74). The initial DNA cleavage by EcoBI is followed by the release of nucleotides or short oligonucleotides (75,76). The 5′ ends of DNA produced by EcoKI and EcoBI may be refractory to the polynucleotide kinase reaction (46,77,78).

After cleavage, ATP hydrolysis continues but the enzyme does not dissociate from the DNA and so, in the restriction reaction, type I enzymes do not turn over (32,48,77,78). Hence, stoichiometric amounts, i.e. one type I enzyme per site, are required for cleavage of DNA. The continued ATPase activity may be due to the enzyme cycling repetitively at the end of the DNA. Alternatively, the enzyme may be able to move backwards and forwards on the DNA or fall off the end of the DNA allowing translocation to begin again back at the target sequence.

**TYPE III R-M ENZYMES**

**Enzyme description**

Type III R-M enzymes are complex molecules that exert both modification and restriction activities. They are composed of two different polypeptides, Res (106 kDa) and Mod (75 kDa), products of the res and mod genes, respectively (6–8,79). Type III R-M enzymes described to date are specified by phage P1 and the related plasmid p15B of Escherichia coli (80) and by the bacteria Haemophilus influenzae (81) and Salmonella enterica serovar typhimurium (82,83). Just as phage lambda revealed host specificities in E.coli, the HP1cl phage was used to study R-M systems in H.influenzae, H.inflIII and HinI were discovered from H.influenzae serotypes RI and Re, respectively. Both these enzymes recognise the sequence 5′-CGAT-3′. The enzymes have been purified and used to demonstrate a requirement for more than one site for DNA cleavage (84). The only type III R-M system that has been characterised in Gram-positive bacteria is the LlaFI system in Lactococcus lactis (85). Most of the discussion below pertains to EcoP11 and EcoP15I, the only type III systems to have been described in detail.

Several observations implied that the res and mod genes of EcoP11 and EcoP15I were transcribed as single units (86) leading to the conclusion that translation of res mRNA was due to ribosomal shuffling from the terminator to the initiator codon, an initiation factor-independent event. The Mod subunit functions as a mtase whereas restriction activity requires the cooperation of both Res and Mod subunits.

Both EcoP11 and EcoP15I mtases belong to the class of N6 adenine mtases. The amino acid sequences of these enzymes include the conserved motifs that are responsible for AdoMet binding and catalysis (86). Mutations in these motifs result in loss of activity (87–90). Reddy and Rao (91) further demonstrated that the cysteine at position 344 in EcoP15I DNA mtase was necessary for DNA binding and, therefore, for activity. More recently, they demonstrated that EcoP15I DNA mtase stabilised the target base extrahelically and suggested that the EcoP15I DNA methylase elicits a large structural distortion within the recognition sequence, possibly flipping the target adenine (92).

**DNA translocation and cleavage experiments**

One of the unique characteristics of type III R-M systems is a non-symmetrical recognition sequence, which can be methylated on only one strand (Table 1). These enzymes cleave DNA 25–27 bp downstream of the recognition sequence.

Res associates with Mod to form an active endonuclease of stoichiometry (Res)2(Mod). The stoichiometry has not yet been rigorously determined. By analysing the cleavage of T7 DNA, which contains 36 EcoP15I sites in the same head-to-tail orientation, and T3 DNA, which contains pairs of sites in the reciprocal head-to-head and tail-to-tail orientations, Meisel et al. (93) showed that for type III enzymes to restrict DNA, two unmethylated sites are required in the head-to-head orientation. In other words, cleavage required a palindromic sequence with variable spacer length. These enzymes require ATP hydrolysis and Mg2+ for restriction (94,95) in contrast to the conclusions drawn from early experiments (96,97).

Meisel et al. (95) studied DNA cleavage in a substrate in which a Lac repressor binding site was flanked by two head-to-head recognition sequences. Only in the presence of isopropyl-β-D-thiogalactopyranoside, which caused the Lac repressor to dissociate from the DNA, was cleavage observed suggesting that the DNA is translocated through the enzyme. ATP hydrolysis provides the energy for such a tracking mechanism. Translocation positions the two inversely oriented enzyme–DNA complexes appropriately for cleavage to occur (95). It has also been demonstrated that EcoP11 and EcoP15I are able to interact functionally to restrict a target DNA molecule carrying only one EcoP11 and one EcoP15I recognition site (98). Reich et al. (99) showed that EcoP15I proceeds to cleave DNA efficiently even in the case of two adjacent head-to-head oriented recognition sites and this cleavage was abolished in the presence of non-hydrolysable ATP analogues instead of ATP. These results, therefore, confirm the role of ATP hydrolysis for the phosphodiester bond cleavage. They also found a 36 base footprint symmetrical in both strands in DNase I footprinting experiments and the presence of ATP caused a change in the footprint pattern. These and other results clearly implicate a role for ATP in DNA recognition (100,101), DNA translocation (95) and DNA cleavage (95,99) by type III restriction enzymes.
Based on amino acid sequence comparisons, Gorbalenya and Koonin (53) suggested that the Res subunits of type III enzymes contain the DEAD-box motifs present in helicase superfamily II. Saha and Rao (89), using EcoP1I restriction enzyme and several DNA substrates, were unable to detect any classical strand-separation helicase activity. The type III restriction enzymes remain bound at their recognition sequences while translocating DNA past themselves, whereas the classical helicase proteins move along DNA causing strand separation during the process. The difference in the modes of action of the type III restriction enzymes and the helicases may explain why no helicase activity was evident. Changes were made by site-directed mutagenesis in two of the seven motifs (89,90). Mutations in motif I affected ATP hydrolysis and resulted in loss of DNA cleavage activity, while mutations in motif II decreased ATP hydrolysis but had no effect on DNA cleavage. These results, therefore, suggest that motif I is involved in coupling DNA restriction to ATP hydrolysis.

**Models of DNA translocation and cleavage.** The model of DNA cleavage by type III restriction enzymes is very similar to that for type I restriction enzymes (54,95) and postulates that when an enzyme bound to a recognition site starts tracking along the DNA, it produces a DNA loop of increasing size until it collides with another enzyme bound to another site and tracking in the opposite direction (Fig. 4). The collision triggers DNA cleavage a fixed distance away from the recognition sequence and independent of the length of DNA originally separating the two head-to-head recognition sequences. Which of the two recognition sites of the pair is selected for cleavage is a random event. As both cleavage products contain the original recognition sites, it has been postulated that the enzyme molecules can continue translocating DNA after cleavage. However, type III enzymes turn over in the restriction reaction and, therefore, must eventually dissociate from the recognition sequence after DNA cleavage. Prior to dissociation, the enzyme can methylate the site. Although the details about the release of the enzyme from DNA substrate are not known, it has been suggested the dissociation is facilitated by this methylation.

**METHYLATION-DEPENDENT RESTRICTION SYSTEMS**

The McrBC system

*Escherichia coli* K12 codes for at least three restriction endonucleases that recognise and cleave DNA containing modified bases. Such processes were first recognised phenomenologically with T-even bacteriophages lacking glucosylation of their 5-hydroxymethyl cytosine residues. These enzymes are encoded by the *mcrA*, *mcrBC* and *mrr* genes, respectively (102–104). The McrA (105–107) and Mrr (108) systems have only been partially characterised and do not appear to have a requirement for nucleotide hydrolysis. The *mcrBC* system does require nucleotide hydrolysis and has been more thoroughly characterised.

### Table 1. Salient features of nucleoside triphosphate-dependent restriction enzymes

<table>
<thead>
<tr>
<th>Features</th>
<th>Type I</th>
<th>Type III</th>
<th>MDRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genes</td>
<td>hsdR, hsdM and hsdS</td>
<td>res and mod</td>
<td>mcrB and mcrC</td>
</tr>
<tr>
<td>Polypeptides</td>
<td>HsdR, HsdM and HsdS</td>
<td>Res and Mod</td>
<td>McrB&lt;sub&gt;1&lt;/sub&gt;, McrB&lt;sub&gt;2&lt;/sub&gt; and McrC</td>
</tr>
<tr>
<td>Enzyme activity</td>
<td>ENase, MTase and ATPase</td>
<td>ENase, MTase and ATPase</td>
<td>ENase and GTPase</td>
</tr>
<tr>
<td>Cofactor requirements for nucleic activity</td>
<td>AdoMet, ATP, Mg&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>ATP, Mg&lt;sup&gt;2+&lt;/sup&gt; (AdoMet)</td>
<td>GTP, Mg&lt;sup&gt;2+&lt;/sup&gt;</td>
</tr>
<tr>
<td>Cofactor requirements for mtase activity</td>
<td>AdoMet</td>
<td>AdoMet, Mg&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>–</td>
</tr>
<tr>
<td>DNA recognition sites</td>
<td>Asymmetrical bipartite</td>
<td>Asymmetrical</td>
<td>Bipartite *RmC(N)&lt;sub&gt;40-80&lt;/sub&gt;RmC</td>
</tr>
<tr>
<td>Location of DNA methylation</td>
<td>On both strands within target</td>
<td>On one strand within target</td>
<td>None</td>
</tr>
<tr>
<td>DNA cleavage requirements</td>
<td>Generally two recognition sites in any orientation</td>
<td>Two recognition sites in a head-to-head orientation</td>
<td>Methylated recognition site</td>
</tr>
<tr>
<td>Position of DNA cleavage</td>
<td>Cut DNA sites remote from the recognition site and approximately half-way between two recognition sites</td>
<td>Cut DNA close to one recognition sequence between the two head-to-head sites</td>
<td>Cut DNA close to one or the other parts of the recognition site</td>
</tr>
<tr>
<td>DNA translocation</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Examples</td>
<td>IA EcoKI AAC(N)&lt;sub&gt;3&lt;/sub&gt;GTGC</td>
<td><em>EcoP1I</em> AGACC</td>
<td>McrBC</td>
</tr>
<tr>
<td></td>
<td>IA EcoBI TGA(N)&lt;sub&gt;3&lt;/sub&gt;TGCT</td>
<td><em>EcoP1I</em> CAGCAG</td>
<td>RmC(N)&lt;sub&gt;40-48&lt;/sub&gt;RmC</td>
</tr>
<tr>
<td></td>
<td>IB EcoAI GAG(N)&lt;sub&gt;3&lt;/sub&gt;GTCA</td>
<td><em>HinIII</em> CGAAT</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>IC EcoR124I GAA(N)&lt;sub&gt;3&lt;/sub&gt;RTCG</td>
<td><em>SlyLTI</em> CAGAG</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>ID SlySBLI CGA(N)&lt;sub&gt;3&lt;/sub&gt;TACC</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
McrC neither binds DNA nor affects the protein–DNA interaction but is required for catalysis of the cleavage reaction. Recent site-directed mutagenesis experiments indicate that McrC harbours the catalytic centre for DNA cleavage (U. Pieper and A. Pingoud, personal communication). McrB S alone, or in the presence of McrC, cannot support restriction in vivo (119).

Figure 4. Schematic diagram showing the model proposed for restriction by type III R-M systems (95). The enzymes are shown as dimers of one res and one mod subunit but a further dimerisation to form (res)2(mod)2 may occur (see Fig. 3B). (A) Enzymes bind two inversely oriented recognition sequences. (B) ATP-dependent DNA-translocation commences resulting in (C) collision of two complexes. (D) DNA cleavages are introduced on the recognition side of the loop.

McrB_L is responsible for sequence-specific DNA binding with the DNA binding domain residing in the N-terminal 160 amino acids (120). The sequence of McrB suggests that the GTP-binding site is located in the C-terminal half of the molecule where three sequences characteristic of guanine-nucleotide binding proteins are located (121,122). This has been confirmed by the analysis of a deletion mutant specifying a polypeptide nearly identical to McrB5. This truncated protein binds and hydrolyses GTP in a manner similar to wild-type (120). Early data suggested that DNA binding was stimulated but not dependent on GTP or McrC (123). However, Stewart et al. (124) have found that McrBC requires GTP and Mg^{2+} to form a clearly defined protein–DNA complex in gel retardation experiments. Mutations in McrB that lead to reduction in GTP binding and/or hydrolysis can affect DNA binding, suggesting that the two activities are coupled in McrB5 (125). Pieper et al. (122) demonstrated that the steady state rate of GTP hydrolysis was much faster than the steady state rate of DNA hydrolysis, clearly suggesting that one DNA cleavage event is associated with the hydrolysis of many molecules of GTP.

Experiments on DNA translocation and cleavage

The recognition sequence for McrBC is RmCN(40–80)RmC, where R stands for a purine residue. Cleavage occurs between the two modified cytosine residues at multiple positions on both strands. DNA cleavage by McrBC requires at least two RmC sites separated by 40–3000 bp (121,126). Cleavage is near one recognition element. A peculiar feature of this enzyme is that the relative disposition of the two recognition elements is not critical, i.e. the RmC elements can appear on both strands at each site, or on only one strand at each site, either in the same strand or on opposite strands (121). DNA cleavage in vitro by McrBC requires that McrB_L and McrC are present in a specific ratio (127). McrB_L modulates the activity of the cleavage complex by changing the effective ratio of McrB to McrC.

Panne et al. (128) have shown that DNA cleavage of circular DNA by McrBC required only one methylated recognition site, whereas the linearized form of this substrate was not cleaved. It was also shown that the linearized substrate could be cleaved if a Lac repressor was bound adjacent to the recognition site, clearly implying that communication between two remote sites was achieved by DNA translocation and not by DNA looping. A mutant form of McrBC with defective GTPase activity could cleave DNA substrates with closely spaced recognition sites but not substrates with sites far apart. These results suggest that McrBC translocates DNA in a reaction dependent on GTP hydrolysis.

The results obtained with McrBC are reminiscent of DNA cleavage by type I and type III restriction enzymes. However, McrBC differs from the type I and III restriction enzymes in several aspects. (i) McrBC enzyme does not generate nicked intermediates even transiently, in contrast to type I restriction enzymes and probably type III enzymes. (ii) Both type I and III enzymes exhibit both endonuclease and methylase activities, whereas McrBC exhibits only endonuclease activity. (iii) McrBC requires GTP as a cofactor in the DNA cleavage reaction whereas the type I and III enzymes use ATP as a cofactor. In the case of type I enzymes, AdoMet is also required for restriction. Recent work on type III enzymes shows that AdoMet is also required for DNA cleavage (D.N.Rao, unpublished results). (iv) A single molecule, McrB_L, is responsible for DNA recognition and nucleotide binding whereas DNA binding and nucleotide binding reside on different subunits in type I and type III enzymes. (v) Type I and III restriction enzymes hydrolyse ATP in a DNA-dependent manner but McrBC hydrolyses GTP even in the absence of DNA.
It has been shown that the quaternary structure of the McrBC endonuclease depends on binding of cofactors. In the presence of Mg^{2+} and GTP, GDP or GTP-\gamma-S, McrB_{\gamma} and McrB_{\beta} form high molecular weight oligomers. Oligomer formation is not dependent upon the presence of DNA. These oligomeric forms have been shown to preferentially interact with McrC. Analysis by electron microscopy for both McrB_{\gamma} and McrB_{\beta} reveals ring-shaped oligomers with a central channel (D.Panne and T.A.Bickle, unpublished results).

Model of DNA cleavage by MDRS

McrB_{\beta} binds to DNA only in the presence of GTP and, therefore, the first step in the process would be GTP binding to the protein to trigger DNA binding. The second step in the process is the interaction of McrC, which binds to the DNA–GTP–McrBL complex only in the presence of GTP. This stimulates GTP hydrolysis and ensures translocation. After two translocating complexes meet, DNA cleavage is triggered close to only one of the recognition sites.

Other potential nucleoside triphosphate-dependent R-M systems

It has been observed that DNA cleavage by the type II restriction endonuclease, CviJI, is affected by the addition of ATP and AdoMet. CviJI, isolated from Chlorella-like green algae infected with phycodnavirus IL-3A, recognises and cleaves DNA containing 5''-RGGCY-3' (129). However, in the presence of ATP, a star-like activity has been observed and cleavage was shown to occur at 5''-RGCGY-3', 5''-RGCR-3' and 5''-YGCGY-3'. It has been suggested that ATP causes a conformational change, which alters the enzyme specificity. In the presence of AdoMet, restriction activity by CviJI is increased. These properties are reminiscent of type I and type III restriction enzymes. The open reading frame encoding CviJI has some regions of sequence homology with various DNA-binding proteins, including some of which also bind ATP. Over a short stretch of the gene encoding CviJI, 30–35% identity with the res gene of EcoP11 was observed. (129,130).

Regulation of enzymatic activities

Regulation of the endonuclease activity of a R-M system is expected to be critical because unmodified targets in the bacterial chromosome should make it, and consequently the bacterium, vulnerable to the restriction endonuclease. One extreme example of this problem follows the transfer of genes encoding an R-M system to a bacterium in which the chromosome is unmethylated. Transcriptional regulation of some of the genes encoding type II R-M systems has been demonstrated (131). The potential for transcriptional regulation exists for type I and type III systems where a separate promoter is found for the gene encoding the subunit essential for endonuclease activity, but experiments find no support for control at the level of transcription (37,132–135). Translational, or even post-translational, control has been invoked.

With hindsight, an early clue to the regulation of some type I R-M systems was the demonstration that restriction in E.coli K-12 is alleviated in response to irradiation by UV-light (136). Very much later, direct evidence for a regulatory mechanism came from the demonstration that a mutation (hsdC) in E.coli strain C, a naturally restriction and modification-deficient (r-m-) strain, prevented the acquisition of the hsd genes from E.coli K-12 (133). Ryu and co-workers (137) also showed that there is a long lag before the wild-type E.coli C recipient becomes restriction proficient following the acquisition of the genes encoding EcoKl, thereby allowing time for the methylation of target sequences. When modification proficiency was monitored, by checking the modification acquired by xwir.0 during a single round of infection, it was found that the cells only become fully modification proficient shortly before they become restriction proficient (S.Makovets and N.E.Murray, unpublished results). It appears that restriction activity is modulated (alleviated) during this lag period within which the many unmethylated targets (~600) in the recipient chromosome must become hemimethylated. For EcoKI the efficiency of methylation of unmodified target sequences is very low (32,44), thereby extending the length of the lag period. The cellular function required to modulate the endonuclease activity, the function missing in the hsdC derivative of E.coli C, is a protease specified by the genes clpX and clpP (138). Together the products of these genes comprise the ClpXP protease, while ClpX itself can function as a substrate-specific chaperone (139). For EcoKI and EcoAI, representatives of the type IA and IB families, the temporary alleviation of restriction in response to agents that damage DNA, and the ability to acquire hsd genes, have been shown to require ClpXP. The regulation of restriction activity in both contexts may be viewed as restriction alleviation (RA).

A recent analysis of the role of ClpXP in RA has led to the identification of a molecular pathway that protects the bacterial chromosome from attack by type IA and IB systems. RA was shown to correlate with the ClpXP-dependent loss of HsdR and the consequent acquisition of an r-m' phenotype (140). Early evidence indicated that HsdR was degraded only if it formed part of a functional complex. This finding promoted the concept of a remarkably specific control mechanism, effective only after the relevant pathway had been initiated but able to act before any damage was inflicted on the bacterial chromosome (Fig. 5) (140). Recent experiments provide direct evidence in favour of this model. It has been shown that the proteolytic degradation of HsdR is prevented by missense mutations that identify each of the seven motifs essential for the ATP-dependent translocation that precedes DNA breakage. Proteolysis, however, is not affected by those mutations that permit DNA translocation but block endonuclease activity. It was concluded that the HsdR subunits of EcoKl are recognised by ClpXP only after the enzyme has initiated the restriction pathway but before the signal that stimulates DNA cleavage (141).

The possibility that the proteolytic control of restriction activity correlates with the conformation of HsdR, the polypeptide associated with DNA translocation, raises the question of whether other complex systems, those in which DNA translocation features in their restriction mechanism, are also susceptible to a similar type of control. An easily detectable indicator of the potential for this type of control is the induction of RA in response to agents that damage DNA. The alleviation of restriction in response to treatment with 2-aminopurine (2-AP) has been demonstrated for members of each of the four families of type I R-M systems identified in enteric bacteria (140,142; S.Makovets and N.E.Murray, unpublished results), but the mechanism of RA for type IC and ID systems and its potential relevance to the general control of
ATP-dependent translocation

ClpXP attacks translocating complexes
and degrades HsdR

Figure 5. A model for the mechanism of control of restriction activity of EcoKI. The ATP-dependent translocation of DNA by EcoKI is dependent upon the enzyme binding to unmodified target sequences. When such target sequences elicit translocation of the resident bacterial chromosome, ClpXP recognises HsdR, removes this subunit from the complex and degrades it.

restriction activity remain to be determined. The relative stabilities of intermediates in the assembly pathway of type IC R-M systems may suffice to explain the ease with which their genes are transmitted from one bacterial strain to another (143), but they do not provide a molecular explanation for RA in response to treatment with external agents such as UV-light and 2-AP. It is known that neither transmission of the plasmid-borne IC genes by conjugation (134) nor the induction of RA by 2-AP is dependent on ClpXP (S.Makovets and N.E.Murray, unpublished results), but alternative control systems may modulate restriction activity.

RA in response to external agents has been shown to occur for MDRS (144), although the potential advantage of this alleviation is not obvious.

RA has not been documented for a type III R-M system but early experiments (145) showed that the modification function of EcoP1I was detectable within a few minutes after P1 infection while restriction was evident only after ~1 h. Both protein assembly and proteolysis appear to play an important role in the control of endonuclease activity for EcoP1I and EcoP15I. It was shown that Res is stabilised in the presence of Mod in vivo (135). These authors postulated that the correct folding of Res into an active and stable conformation was promoted by its interaction with Mod; in the absence of Mod, improperly folded Res would be more susceptible to degradation. Thus, Mod protects Res from proteolysis by direct protein–protein interactions. The amount of Mod and Res synthesised is not the only factor known to influence the level of EcoP1I restriction activity. Studies of E.coli strains resistant to streptomycin, as the result of mutations in ribosomal genes that affect the efficiency and accuracy of translation, suggest some means of translational control (135,146). It has been shown that inefficient translation of mRNA can lead to the proteolytic degradation of incomplete polypeptides (147). Incomplete Res subunits may be a substrate for degradation. In contrast to EcoP1I, the chromosomally located genes for StyLTI, the type III R-M system of S.enterica serovar typhimurium LT2, lack the necessary control to permit their transfer to another strain (82).

BIODIVERGENT PERSPECTIVES

More than 3000 R-M systems have been detected (148). Genomic sequences of eubacteria, archaea and actinomycetes usually indicate the presence of at least one R-M system. R-M systems are encoded by algal viruses as well as bacteriophages, but there is no evidence for their presence in eukaryotes. Genomic sequences permit the detection of putative type I and type III R-M systems and surveys of these sequences indicate that complex R-M systems are not predominantly associated with any group of bacteria. The previous bias towards their presence in members of the Enterobacteriaceae reflects the bacteria commonly used for genetic analyses. Present evidence indicates that nucleoside triphosphate-dependent R-M systems are found in the majority of bacterial genera (5,148).

Why are some restriction systems as complex as those surveyed in this review? This question has not been answered but it is apparent that important features for the control of restriction activity are dependent upon the assembly of these complex oligomeric enzymes and, in the case of type I R-M systems, upon the complex pathway that leads to DNA breakage.

The chromosomally encoded type I systems of Enterobacteriaceae are distinguished by their extraordinary allelic diversity, somewhat similar to that found for rfb locus encoding the O-antigen where it is suggested that the diversity is important in influencing pathogenesis (149). In Mycoplasma pulmonis site-specific recombination can invert a segment of DNA within the hsdS gene to create a system with a new specificity (150). This switching of specificities is reminiscent of the phase variation of virulence determinants in pathogenic bacteria. It is postulated (140) that the tight regulation of restriction activity provided by ClpXP could permit the effective activation of dormant R-M genes by genetic switches, in addition to enhancing the acquisition of new specificities by gene transfer.

The mod genes of the type III R-M systems identified in strains of H.influenzae and Pasteurella haemolytica include short repeated nucleotide sequences. In H.influenzae such tetranucleotide repeats are known to be located within genes that appear to encode proteins relevant to the nature of the outer membrane. De Bolle et al. (151) have shown that the number of repeats within the 5′ region of the mod gene of H.influenzae influences the rate of phase variation, suggesting that the activity of the R-M system, like functions relevant to pathogenesis, is subject to phase variation. Similarly, a pentanucleotide repeat in the mod gene of P.haemolytica may modulate expression of a resident type III system (152). R-M genes have also been identified as candidates for phase variation in Neisseria meningitidis (153). Present evidence suggests that both allelic diversity and phase variation may be characteristics of systems relevant to the preferential survival
of bacteria; the most convincing examples being those identified in pathogenic bacteria.

The complex R-M systems are well suited for variable expression and variation in sequence specificity. A specificity subunit common to the enzymes required for modification and restriction, or an endonuclease lacking a cognate modification enzyme, permit change in specificity without the need to co-evolve two enzymes with the same specificity. In addition, the target sequences of type I or III R-M systems offer more scope for the evolution of new sequence specificities than the simple rotationally symmetrical, non-interrupted sequence recognised by most type II R-M systems (1). The bipartite target sequence of a type I R-M system provides particular potential for variation, its general requirement being two adenine residues, the substrates for methylation, situated 8–11 bp apart. For these systems both the spacing and combination of TRDs can contribute to changes in sequence specificity.

It is difficult to determine the driving force for the selection of alternative specificities encoded by allelic hsd genes and, consequently, to comprehend the biological relevance of the diversity. The classical explanation for the diversity of sequence specificity is one dependent upon selection by phages. It has been shown that when bacteria share a laboratory habitat with phages, mutants resistant to phages are rapidly selected. Nevertheless, it seems reasonable that bacteria encoding an R-M system with a different specificity are likely to be at an initial advantage when colonising a new habitat as they can restrict the resident unprotected phages (154–156). This relatively short-lived advantage may be sufficient to impose frequency-dependent selection for diversity. Data are not available to determine whether allelic diversity is found in most bacterial species but preliminary supportive evidence from the sequences of two *Helicobacter* genomes suggests allelic genes for type I R-M systems (5).

Kobayashi and co-workers, in particular, have documented the death of bacteria when R-M genes are lost and have argued that R-M genes are selfish (157). They have shown that when type II R-M genes are lost, residual endonuclease activity attacks unmodified targets in the bacterial chromosome (158). No such problem has been detected for type I R-M systems (134,159). *Escherichia coli*, as already indicated, has an elegant, fail-safe mechanism to guard against chromosomal damage as the result of a resident type IA R-M system, even under the vulnerable conditions of extensive DNA damage.

The mechanism by which RA protects unmodified chromosomal DNA raises new questions about the distinction between DNA defined as self or foreign. In the absence of modification, unmodified chromosomal DNA evokes RA while unmodified phage DNA provokes restriction (140,141). Experiments in the 1960s confirmed the expectation that R-M systems reduce the linkage between genes transferred by conjugation (160,161), and it was appreciated as early as 1973 that DNA breaks induced by restriction endonucleases could be recombinogenic [S.Lederberg cited in Radding (162)]. A reconciliation of the apparent conflict between the established role of RecBCD (ExoV) in the degradation of the DNA fragments produced by the restriction of foreign (phage) DNA (163) and the potential role of RecBCD in salvaging the products of DNA breakage provides the basis for appreciating how restriction can resolve an understanding of the role of special sequences, Chi, in moderating DNA degradation and promoting recombination (164). R-M systems are predicted to modulate the flow of genetic information between bacterial strains, enhancing the opportunity for the acquisition of advantageous sequences in the absence of deleterious ones (165). The cutting of DNA at non-specific sequences into fragments likely to contain Chi sequences may make type I R-M systems particularly advantageous in this context (166).

**CONCLUSION**

At present it seems that complex R-M systems should be considered as enzymes relevant to DNA replication, repair and recombination. Inevitably they will influence the transfer of genetic information. Further advances in our understanding of these complex systems will be aided by structural analysis, single molecule manipulation of translocating enzymes, genomic sequence analysis and classical molecular genetics.

**ACKNOWLEDGEMENTS**

The authors thank Natalie Honeyman for preparing the manuscript. Figures 2 and 3 are from the PhD thesis of Graham P. Davies, Figure 5 from that of Svetlana Makovets. We are grateful to these former colleagues for their help. D.T.F.D. thanks the Royal Society for a University Research Fellowship, N.E.M. thanks the MRC for their support, D.N.R. thanks the INSA-Royal Society Exchange Programme and acknowledges financial support from CSIR and DAE, Government of India for work done in his laboratory.

**REFERENCES**


