

## SURVEY AND SUMMARY

# Tracking EcoKI and DNA fifty years on: a golden story full of surprises

Wil A. M. Loenen\*

Department of Medical Microbiology, University Maastricht, Maastricht, The Netherlands

Received October 3, 2003; Revised and Accepted October 30, 2003

### ABSTRACT

1953 was a historical year for biology, as it marked the birth of the DNA helix, but also a report by Bertani and Weigle on ‘a barrier to infection’ of bacteriophage  $\lambda$  in its natural host, *Escherichia coli* K-12, that could be lifted by ‘host-controlled variation’ of the virus. This paper lay dormant till Nobel laureate Arber and PhD student Dussoix showed that the  $\lambda$  DNA was rejected and degraded upon infection of different bacterial hosts, unless it carried host-specific modification of that DNA, thus laying the foundations for the phenomenon of restriction and modification (R-M). The restriction enzyme of *E. coli* K-12, EcoKI, was purified in 1968 and required S-adenosylmethionine (AdoMet) and ATP as cofactors. By the end of the decade there was substantial evidence for a chromosomal locus *hsdK* with three genes encoding restriction (R), modification (M) and specificity (S) subunits that assembled into a large complex of >400 kDa. The 1970s brought the message that EcoKI cut away from its DNA recognition target, to which site the enzyme remained bound while translocating the DNA past itself, with concomitant ATP hydrolysis and subsequent double-strand nicks. This translocation event created clearly visible DNA loops in the electron microscope. EcoKI became the archetypal Type I R-M enzyme with curious DNA translocating properties reminiscent of helicases, recognizing the bipartite asymmetric site AAC(N6)GTGC. Cloning of the *hsdK* locus in 1976 facilitated molecular understanding of this sophisticated R-M complex and in an elegant ‘pas de deux’ Murray and Dryden constructed the present model based on a large body of experimental data plus bioinformatics. This review celebrates the golden anniversary of EcoKI and ends with the exciting progress on the vital issue of restriction alleviation after DNA damage, also first reported in 1953, which involves intricate control of R subunit activity by the

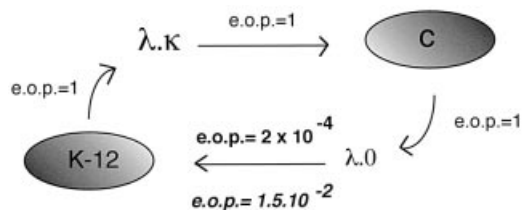
bacterial proteasome ClpXP, important results that will keep scientists on the EcoKI track for another 50 years to come.

### INTRODUCTION

Bacteria employ numerous survival strategies to protect themselves against foreign invaders, including viruses [(bacterio)phages]. Phages were first described during World War I and pioneering work with  $\lambda$  and P1 (1,2) proved indispensable for molecular and biochemical studies on the first ‘classical’ R-M enzyme, EcoKI (originally known as EcoK). Recent reviews by main contributors to the field, constantly ‘refuelled’ by the Swiss National Science Foundation and the Medical Research Council UK, describe EcoKI and related Type I enzymes as ‘sophisticated molecular machines’ or ‘NTP-driven motors’, ample proof of major progress as well as the complexity of this enzyme (3–6).

This review article tracks the EcoKI enzyme along the DNA over the past 50 years, from Bertani and Weigle who first described a barrier to infection of bacteria by phage, via Arber and Dussoix who proved the involvement of DNA, to the genetic, biochemical and EM data of several laboratories in the 1960s and 1970s that provided evidence for a large  $R_2M_2S_1$  protein complex with peculiar translocating properties encoded by three genes *hsdR*, *hsdM* and *hsdS*. Cloning, sequencing and further genetic analysis followed and today much is known about the assembly and conformational changes of EcoKI in functional complexes, putative coevolution of R and M subunits, and gene duplication and shuffling of target recognition domains (TRDs) of the S subunit. Improved sequence alignments and comparisons of EcoKI with 3-D structures of other DNA binding proteins, including Type II REases, MTases and helicases, provide substantial support for similar structural domains in all Type I enzymes. These theoretical assumptions are backed up by extensive genetic, biochemical and biophysical experimental evidence. Taken together results support the notion of primordial genes for R-M and translocase/helicase activity, despite low primary sequence identity. In addition to its role in the distinction between ‘self’ and ‘foreign’ DNA that favors survival of the individual cell, EcoKI may be important to populations and play a role in bacterial evolutionary processes by creating DNA ends of high recombinogenic potential.

\*To whom correspondence should be addressed at present address: PO Box 35082, 3005 DB Rotterdam, The Netherlands. Tel: +31 10 418 6266; Fax: +31 10 422 5607; Email: wil.loenen@medmic.unimaas.nl



**Figure 1.** The Bertani and Weigle experiment (1953). The discovery of a barrier to infection of  $\lambda$  in its natural host, *E. coli* K-12, if the phage had been previously propagated on *E. coli* C. This led to the phenomenon of R-M. *Escherichia coli* K-12 possesses a R-M system, EcoKI, absent in *E. coli* C. Phage  $\lambda$  propagated on *E. coli* C (called  $\lambda.C$  or  $\lambda.0$ ) is not protected from restriction by EcoKI and forms plaques with low efficiency (e.o.p.) on *E. coli* K-12 as compared to *E. coli* C (bold). Pretreatment of cells with UV leads to RA (italic). The mechanism behind this phenomenon was recently elucidated by Svetlana Makovets and Noreen Murray.

This survey starts and ends with phage, as *λhsd* aided Makovets and Murray in the elucidation of the intricate control of EcoKI restriction via ClpXP to avoid damage to the resident bacterial chromosome. Due to limitations of space, regrettably only landmark papers can be individually listed and the reader is referred to reviews on Type II and III, as well as other Type I systems for further references (3–10).

## DISCOVERY OF A BARRIER TO INFECTION OF $\lambda$ IN *E. COLI* K-12

Perhaps symbolically in the year Hillary and Tenzing ‘crowned’ Mount Everest, Joe Bertani and Jean Weigle reported a barrier to  $\lambda$  infection in its natural host, *E. coli* K-12, if the phage had been previously propagated on *E. coli* C ( $\lambda.C$ ) (11). However, with a frequency of about  $2 \times 10^{-4}$ , phage would overcome this barrier and these few progenitors were now fully capable again of growth on *E. coli* K-12 ( $\lambda.K$ ). Figure 1 demonstrates this experiment, but similar results were obtained with phage P2 going from *Shigella* to the restricting *E. coli* B, thus providing evidence for a more general nature of this phenomenon (11,12). These findings provided the first evidence that (i) bacteria applied an identification (modification) tag to their DNA and (ii) DNA lacking this tag (upon transfer from one strain to another) was recognized as ‘non-self’ and restricted. This ‘host-controlled variation’ proved phenotypic and not genetic, and determined by the bacterium in which the virus had been produced. Modifications by successive hosts were not cumulative, but mutually exclusive. Apparently, *E. coli* C lacked a R-M system, permitting propagation of  $\lambda.K$  and  $\lambda.C$  equally well. The authors also reported that the barrier was lifted by pretreatment of host cells with UV (restriction alleviation, RA), a phenomenon recently readdressed (see below).

## THE ARBER AND DUSOIX EXPERIMENTS

Booming bacterial and phage genetics, protocols for isotope labeling of phage and cell components, high-speed centrifugation and EM spreading, all led to major mid-century progress with respect to transduction of bacterial genes, polarity of bacterial conjugation, the role of the viral DNA and the structure and semi-conservative replication of the DNA itself



**Figure 2.** Tracking EcoKI, Nobel Laureate Werner Arber and Noreen Murray.

(1,2). All these proved crucial to EcoKI research. Chance brought Bertani’s former postdoc Werner Arber (Fig. 2) to EcoKI in 1960, when he went to work on ‘the nature of radiation damage to genetic material and its repair mechanisms’ (13). Arber had been introduced to phage  $\lambda$  by Eduard and Grete Kellenberger in Geneva and his first job was to render *E. coli* B and the radiation-resistant strain B/r sensitive to  $\lambda$ . This serendipitous choice of *E. coli* B (that harbors the EcoBI R-M system closely related to EcoKI) led to the notion that B and B/r could easily be made  $\lambda$ -sensitive by transduction, but these transductants did not allow efficient propagation of phage. Arber quickly realized that this was due to the ‘barrier’ described 7 years earlier and this inspired his pioneering work on the R-M systems of EcoKI, EcoBI and that of phage P1 (EcoPII), for which he was awarded the Nobel Prize for Medicine in 1978. With Daisy Dussoix, he published the first formal evidence that EcoPII restricted or modified  $\lambda$  at the DNA level, events occurring within minutes after infection of the host (14,15). Further evidence that R-M targeted the DNA came from observations that hybrid DNA molecules consisting of one strand with host-specific modification and the other lacking it were not degraded. Upon transformation such DNA gave rise to new phage progeny. Strain-specific R-M directly affected both phage and cellular DNA without causing mutations, facts we now take for granted (16–20).

## GENETICS AND ENZYMOLOGY

The Arber and Dussoix papers generated widespread interest. Elegant complementation tests using partial diploids between the EcoKI and EcoBI systems and mutational analyses soon identified three *hsd* (host specificity for DNA) genes, *hsdR*,

*hsdM* and *hsdS*. The *hsdK* and *hsdB* loci were allelic and located near the *thr* locus on the *E. coli* chromosome, currently called the 'immigration control region' (18,21). Marker rescue experiments provided the first evidence for two-step degradation of restricted DNA: a fast endonucleolytic cut by a putative highly specific REase, followed by a slower breakdown of cleavage products by one or more less specific nucleases. The definite assignment of the three genes to R-M functions was uncertain, but at least one of the functions involved in restriction was also involved in modification (16–20).

The isolation of the EcoKI enzyme from cell extracts proved problematic to Matt Meselson and Bob Yuan, but success followed in 1968 after the educated guess that EcoKI (like EcoP1I) might require ATP for activity plus the discovery that S-adenosylmethionine (AdoMet) was an essential cofactor *in vivo* (21–23). The complex contained all three gene products, restricted unmodified DNA in the presence of ATP and AdoMet, but methylated such DNA in the absence of ATP, leading to protection of the DNA against subsequent endonuclease attack. Purification of M.EcoBI (the modification component of the EcoBI R-M system), the first modification enzyme to be isolated, showed transfer of two methyl groups from radiolabeled AdoMet to a target site on double-strand phage fd DNA, yielding 6-methylaminopurine (m6A), confirming Arber's hypothesis that modification might be due to methylation at specific DNA sites (16,18,21). In this period of consolidation, many other facts became clear, e.g. R-M acted within minutes of viral DNA entry and could take place on non-replicating DNA, restriction efficiency (reduced plating of phage) varied from enzyme to enzyme and depended on the number of specificity sites, gene dosage and physiological conditions (18,24).

## RESTRICTION AND MODIFICATION COMING OF AGE

By 1972, seven REases had been purified: EcoKI, EcoBI, EcoP1I, HindII, EcoRI, HindIII (21) and EcoRII (K.Murray, personal communication). These data and the palendromic nature of the DNA recognition sequences of the latter four enzymes were all reported at the special restriction enzyme meeting organized by Werner Arber in Switzerland that year (N.Murray, personal communication). The discovery and potential use of 'sticky ends' caused a lot of excitement and made this one of the most memorable meetings in the history of restriction enzymes. The *in vitro* AdoMet requirement of EcoKI (and EcoBI) paralleled that *in vivo*: *E. coli* deprived of AdoMet should yield unmodified DNA, but such DNA was apparently not degraded by the resident R-M, leading to the inevitable conclusion that both enzymes needed AdoMet *in vivo* for restriction as well as modification. The authors apparently gazed into a crystal ball by mentioning that AdoMet, perhaps together with ATP, might be part of a complex control system 'in order to protect cells against suicide when AdoMet was too scarce to support proper modification' (21). This issue was rigorously investigated in the 1990s by Svetlana Makovets and Noreen Murray and shown to involve proteolytic processing via the ClpXP pathway (see below).

In 1972 denaturation and renaturation experiments brought the message that EcoBI cut the DNA of f1 phage away from the recognition site (25). The rest is history: EcoKI and EcoP1I proved likewise, the other enzymes cleaved at their target site and this contributed to the current classification of R-M systems into Type I (EcoKI, EcoBI), II (EcoRI, HindIII) and III (EcoP1I), and more recently Type IV (methyl-dependent enzymes), based on subunit composition, cofactor requirement, structure and location of recognition and cut site (though some enzymes appear to be intermediary) (8,26).

In the last part of their 1972 review, Meselson *et al.* (21) provide interesting insight in their discussion on the interaction of REases/MTases with their target site. The authors comment that 'more on general structural grounds than because of experimental evidence, it has often been stated that proteins could not easily recognize specific nucleotide sequences on the DNA in the double-helical configuration ... active sites of enzymes are generally located within grooves or cavities within the enzyme molecule'. Without accepting these generalizations as universally valid, they speculate that perhaps a structural rearrangement facilitated sequence recognition. Though their model involving crossover of the DNA strands at the site of 2-fold rotational symmetry to allow 'insertion' of the restriction site into the enzyme gained no experimental support, the concept for rearrangement of the DNA helix was essentially correct, as witnessed later by the base flipping mechanism by MTases (27) and the kinking of the helix by EcoKI (28).

## THE SEVENTIES: NEW PLAYERS IN THE FIELD

EcoRI heralded the era of DNA cloning, excluding Type I and III enzymes like EcoKI, EcoBI and EcoP1I that cut away from the target site. Within 10 years >200 Type II REases were identified, and the classification into Type I, II and III firmly established (29–32). Type II systems consisted of separate REases and MTases recognizing the same DNA sequence, while prototype III EcoP1I REase was a heterodimer cleaving the DNA 25–28 bases 3' to a non-palindromic target site, 5'AGACC (29,30). This sequence was puzzling as it allowed methylation of one strand only, leaving the DNA vulnerable to restriction after replication, a riddle now solved as EcoP1I requires two recognition sites inverted with respect to each other for cutting (6). By 1980 there was common agreement that Type I restriction involved the *hsdR*, *hsdM* and *hsdS* genes (29). Mutations in *hsdS* or *hsdM* yielded an  $r^{-}m^{-}$  phenotype, while those in *hsdR* only affected restriction ( $r^{-}m^{+}$ ). The REase consisted of all three gene products, while the MTase required two, but a definitive assignment of subunits to genes awaited sequencing of the chromosomal locus (33,34).

Tom Bickle, a 'diploma' student of Arber, continued research into EcoP1I and related Type III enzymes as well as several Type I systems (6,35,36). With Bob Yuan and colleagues, he provided details on the complicated reaction mechanism of EcoKI (30,37). Interaction of EcoKI with AdoMet caused allosteric changes allowing specific DNA binding, an interaction stable in the absence of ATP. After the addition of ATP, three different events could occur depending on the methylation state of the target site: EcoKI released the DNA if bound to a modified site, but modified the second strand if the site was hemimethylated. When the enzyme was

bound to unmodified DNA, a profound structural change occurred and the DNA was cleaved away from the recognition site at a highly variable distance up to 7000 bases away. The latter process was accompanied by ATP hydrolysis, and ATP turnover continued long after the cutting event. Two years later, Yuan and colleagues provided the first evidence for translocation, when their EM pictures showed large DNA loops protruding from the enzyme (38).

Noreen Murray chose EcoKI to study site-specific endonucleases after moving to Edinburgh in 1968, with the aim of manipulating its target sequences. After the realization that EcoKI cut away from its target site, EcoRI was successfully used for vector construction and cloning of many important genes, e.g. DNA polymerase, T4 DNA ligase and also the *hsdMS* genes (39), the latter by selecting for phages that were protected against host restriction via self-modification, a principle widely employed later to clone R-M genes in plasmids. In addition to outstanding papers in the field of genetic engineering, Noreen Murray would become the main contributor to the story of EcoKI and related enzymes (3,4,40,41). The  $\lambda$  *hsdMS* phage was extended *in vivo* via reintegration and excision from the *E.coli* chromosome to generate  $\lambda$ *hsdRMS*, a phage also capable of restriction (42).

Finally, RA by bacteriophages deserves special mention (43–48). In contrast to  $\lambda$ , the T3 and T7 phages grew interchangeably on *E.coli* K, B and C, indicating that their DNA was resistant to restriction. This antirestriction function mapped to the first phage gene expressed after infection, *0.3* or *ocr* (overcoming classical restriction). Ocr blocked EcoKI (and EcoBI) R-M completely and was made in large amounts [in the case of T3 Ocr about 10 000 molecules per cell (45)], perhaps suggesting an additional, as yet unidentified role, e.g. an interaction with *E.coli* RNA polymerase (46). T3 Ocr also hydrolyzed AdoMet, a function separable from the blocking activity and absent in the T7 protein. The recently elucidated crystal structure indicates that a T7 Ocr dimer resembles a piece of bent DNA, thus mimicking the kinked DNA substrate of EcoKI (28). Ard (alleviation of restriction of DNA) proteins of various conjugative plasmids share with Ocr a conserved sequence motif and a highly acidic protein surface (49,50), and predicted elongated shape (51), suggesting similar physical inhibition of EcoKI. However, recent evidence with respect to ArdA of plasmid R16 points towards considerable modification activity by EcoKI under conditions in which ArdA completely blocks restriction (in contrast to T7 Ocr). This makes it unlikely that Ard interferes in an identical way with EcoKI as Ocr (51). A further discussion of these interesting proteins, including the ‘artful’ way in which their gene products are transiently expressed from single-stranded promoters thus avoiding destruction of their incoming DNA, is outside the scope of this review (see 52).

## THE EIGHTIES: EVOLUTIONARY ASPECTS OF TYPE I ENZYMES





During the 1980s, the total number of REases rose to 700 with 120 different specificities, and a large survey of 864 *E.coli* strains showed restrictive properties in about 25% of these, confirming widespread occurrence (53). When Noreen Murray joined the EMBL laboratory in Heidelberg in 1980, there was no information about the DNA recognition domains of

proteins and comparisons of Type I polypeptides seemed an excellent choice. Sequencing of the *hsdS* genes of EcoKI, EcoBI and the novel EcoDI definitively allocated DNA target specificity to the S subunit, which interacted interchangeably with all R and M subunits (33). But just as helix–turn–helix motifs in other DNA binding proteins made headlines, sequence comparisons of these three *hsdS* genes identified only small conserved regions flanking large variable regions. With diligent assistance of excellent PhD students, this disappointing result turned into a success story, as comparisons with other Type I enzymes yielded particularly relevant data about a bipartite DNA recognition domain and led to the elucidation of the evolution of the *hsdS* gene structure. *hsdK* probes showed strong homology between the *hsdK*, *hsdB* and *hsdD* genes, weak hybridization to three R-M systems from *Salmonella*, but surprisingly enough none at all to the *EcoA* genes of *E.coli* 15T<sup>-</sup>. The implication that EcoAI constituted a novel Type I subgroup was confirmed using antibodies against EcoKI that failed to cross-react with EcoAI (54). This lack of sequence conservation gave the current subdivision into Type IA enzymes (prototype EcoKI and EcoBI), Type IB (prototype EcoAI) followed by the plasmid-encoded EcoR124I (prototype IC) and finally, prototype ID, StySBLI (3,4,8,55,56).

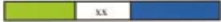



The *Salmonella* systems referred to above were identified by Len Bullas and colleagues, who not only reported the first Type I systems in *Salmonella typhimurium* (encoding StySB) and *Salmonella potsdam* (encoding StySP), but, by chance, in their P1 transduction experiments isolated a third specificity called StySQ (57,58). Following up Bullas’ inspired guess that a recombination event could generate a novel specificity such as StySQ (57), Noreen Murray and coworkers provided the first solid evidence for this idea (3,4,59–62). Heteroduplex analysis of the *Salmonella* genes proved the major breakthrough with respect to a novel mechanism of DNA recognition by the S subunit, as it suggested the presence of a hybrid gene (59). The N-terminal non-conserved domain (currently called TRD) of the S polypeptide of the StySQ strain (the transduction-derived recombinant between *S.typhimurium* and *S.potsdam*) was the same as that of StySP, and both proteins recognized the 5′-AAC trinucleotide within the target site: 5′-AAC(N6) RTAYG-3′. Similarly, StySQ and StySB bound the tetranucleotide TAYG-3′ and shared the second protein domain. Moreover the crossover between StySB and StySP occurred in the central conserved region, splitting the two S domains that recognized the 5′ and 3′ DNA target sites (Fig. 3) (59–62). The organization of variable protein domains involved in DNA recognition on both sides of a central conserved region obviously allows evolutionary diversification of Type I specificities, by the simple expedient of crossovers in that conserved region (62). Further support for the model linking the N-terminal TRD with recognition of the 5′ nucleotide sequence (with one exception a trinucleotide), and the C-terminal TRD with the downstream tetranucleotide (sometimes pentanucleotide) came from studies in the 1990s that also implicated the conserved regions in subunit interactions (3,4). Based on large-scale analysis of TRDs from different Type I classes, the above data and the model for the Type IC EcoR124I S subunit, *hsdS* is currently believed to be a duplication of a primordial gene, with divergent evolution leading to the present arrangement (4,6,40,55,63). Probably a large pool of TRDs already existed before evolution led to

## Evolution of new Type I R-M specificities

a) rearrangements of variable regions (TRD) within the *hsdS* gene results in R-M systems with hybrid specificities

<i>hsdS</i> gene structure	Enzyme	Recognition sequence
	StySPI	AAC (N6) GTRC
	StyLTIII	GAG (N6) RTAYG
	StySJ	GAG (N6) GTRC
	StySQ	AAC (N6) RTAYG

b) alterations in the length of the central region of the *hsdS* gene or the combination of two truncated HsdS polypeptides with one TRD generates new specificities

	EcoR124I	GAA (N6) RTCG
	EcoR124II	GAA (N7) RTCG
	EcoDXXI	TCA (N7) RTTC
	EcoDXXIsI	TCA (N8) TCA

**Figure 3.** Evolution of new Type I specificities via domain shuffling of the N-terminal and C-terminal TRD (a), or alterations in the length of the spacer between the two TRDs (b). Black regions indicate conserved regions involved in subunit interactions and determine the length of the spacer between the two recognition sequences. (a) StySQ is an *in vivo* recombinant of StySPI and the enzymes share the 5' AAC-recognition domain (red), StySPI shares the C-terminal domain with StySJ (blue), StyLTIII and StySJ the N-terminal GAG-recognition domain (green) and finally StyLTIII and StySQ both recognize RTAYG 3' (orange). (b) EcoR124I differs from EcoR124II in a TAEI tetrapeptide in the central conserved region, which is repeated two and three times, respectively, resulting in an extra base in the spacer between the tri- and tetranucleotides of the recognition sequence. EcoDXXIsI results from a transposon insertion in *hsdS* of EcoDXXI, leading to a truncated polypeptide. Such a polypeptide functions as a dimer on a symmetrical trinucleotide sequence with a specificity dictated by the single TRD.

different bacterial species (64). The central region between the TRDs would serve as a spacer between the two DNA recognition domains (Fig. 3), as supported by the transposon insertion mutant of EcoDXXI, and the differences in spacing of the adenines methylated by the closely related EcoR124I and EcoR124II proteins (3,4). In this way, differences in spacer length plus the divergent sequences of R and M between the Type I subfamilies would strongly enhance the potential for diversification of target sequences (4,6,40,55,63). Despite this shuffling of whole domains, it is difficult to relax specificity of these domains, a feature reminiscent of the efforts made with respect to adapting the specificity of Type II enzymes (41).

Finally, the cloning of the *hsdK* locus also allowed the determination of the first DNA sequences for the R and M subunits of a Type I enzyme and identification of two promoters, one for expression of *hsdR* and one for *hsdM* plus *hsdS* (33,34,42). Though disappointing at the time, the lack of any evidence for transcriptional control of expression of *hsdR* had yet another happy sequel 10 years later with evidence for post-translational control (see below).

## VIRAL RESTRICTION ALLEVIATION

Alleviation of restriction (RA) of Type I enzymes by foreign invaders is accomplished in many different ways, including virus-encoded modifications to the DNA, physical interference with REases and counter selection of target sites (44,65).  $\lambda$  Ral protein produces an additional variant of RA as it cannot protect its own unmodified DNA, but rescues unmodified superinfecting phages from EcoKI restriction (66,67). Experiments with the cloned *ral* gene showed that the alleged 'RA' could be accounted for by 'modification enhancement' of unmodified DNA (67). EcoKI clearly distinguished unmodified, hemimethylated and fully modified DNA (9,30), but showed an unusual preference for methylation of hemimethylated targets. Ral appeared to switch EcoKI from maintenance to a *de novo* MTase by enhancing methylation of unmodified sites. Interestingly, Ral-independent EcoKI m\* mutants with similar activity could be isolated after selection for  $\lambda$  *hsdMS* phages with enhanced modification properties (68). The mutations mapped in *hsdM* in line with later data that assigned the discriminatory capacity with respect to the methylation state of the DNA to the M<sub>2</sub>S<sub>1</sub> complex (see below). These valuable mutants should open the way to further *in vitro* approaches to understand this feature of EcoKI at the molecular level, and precisely how the distinction between the methylated and hemimethylated state takes place. Experiments with Ral wait overproduction of this small acidic protein, which have thus far proved elusive both *in vivo* and *in vitro* (D.Dryden, personal communication). The analogous Lar protein, present in some *E.coli* chromosomes, whose deduced primary sequence is quite distinct from that of Ral (69), also proved difficult, but is perhaps more amenable to handling with improved technology.

## THE NINETIES: MODELING ECOKI

Forty years after the first EcoKI paper, reviews mention >2400 Type II REases with >200 specificities, but only about 20 Type I enzymes and five Type III (35,70). Despite an initially rather astounding lack of similarity at the amino acid level even when sharing the same DNA recognition site (with few exceptions), as well as lack of conservation between these enzymes and their cognate MTases, Type II REases currently appear to share one of two common building plans with distinct conserved folding, the EcoRI and EcoRV super-families (7). The 3-D structures of Type II REases and MTases contributed crucial information to the remarkable progress in the elucidation of structural domains of EcoKI over the past 10 years in the absence of crystal data.

Type I enzymes were initially identified in enterobacteriaceae only, as detection was limited to *in vivo* restriction of phage, available *hsd* DNA probes and antibodies, or complementation tests (35,36,40,41,56,58). However, by the end of the century improved sequencing and computer prediction programs showed putative Type I systems in many non-enteric species (8). Type IA, IB and ID constituted chromosomal alleles, but Type IC was plasmid-borne, though one chromosomal member mapped elsewhere (41,71). Rather than biological tests, high level sequence identity of R and M polypeptides within subgroups (80–99%, even from different species) was used to classify these proteins. Between

subclasses, alleles were rather dissimilar (20–35% identity irrespective of the host) showing little evidence of common ancestry. However, early coevolution was suggested, as it was unlikely that Type I R-M systems evolved more than once (41,72), a view supported by the recent avalanche of putative Type I genes in newly sequenced bacterial genomes that include distantly related species with intermediate levels of identity. Such uncommon allelic diversity, reminiscent of other genes that discriminate ‘self’ from ‘non-self’, e.g. the mammalian major histocompatibility complex, led to the concept of ‘primitive bacterial immune system’.

Purification of the 170 kDa EcoKI MTase proved that the capacity of the EcoKI R-M system to distinguish hemimethylated DNA from unmodified DNA strictly required AdoMet and resided within the trimeric  $M_2S_1$  complex (73,74). Mutational analysis of two conserved motifs common to m6A MTases (in M) confirmed that the N-terminal one interacted with AdoMet, while mutations in the second motif, predicted to be part of a surface-exposed loop, prevented enhanced DNA binding or catalysis (75,76). This first evidence for structural similarity between EcoKI and Type II MTases inspired a major tour de force by David Dryden and colleagues (77,78). Long hampered by the absence of EcoKI crystals for structural studies, they used amino acid sequence comparisons and tertiary structure modeling, backed up by extensive mutational analyses from Noreen Murray’s lab, to identify six sequence motifs common to the  $\gamma$  class of Type II MTases. This suggested a primordial MTase gene for Type I and Type II enzymes with conserved  $\alpha$ -helix/ $\beta$ -strand/ $\alpha$ -helix repeats. In the model, the M subunits of EcoKI (linked by the S subunit) would clamp the DNA and resemble two Type II MTases stacked together (Fig. 4). Alignment of the primary sequence of 51 Type I TRDs combined with secondary structure predictions to strengthen true distant similarities led to a tentative tertiary structure for the TRD, resembling the known TRD of the Type I Mtase M.HhaI, and was backed up by extensive mutational analysis (79). Combining the data for M and S, each TRD would fit into the major groove and recognize the DNA, with the M subunits arranged on either side of S allowing them to encircle the DNA and gain access to the methylation targets via base flipping. By extension, all MTases may derive from a common ancestor with one monomeric TRD associated with a separate catalytic subunit [some Type II MTases, e.g. AquI, have this form (8,80)]. The above model was supported by a wide variety of experimental biochemical data, including DNA footprinting, fluorescence anisotropy, gel retardation, protein–DNA cross-linking and measurements of the hydrodynamic shape of wild-type protein and mutants (73–79,81–86).

In addition to the above studies on the domains and structure of the M and S subunits of the MTase moiety of EcoKI, generating support for common ancestry of Type I and II TRDs and methyltransferase functions, much progress was made with respect to the curious translocation function of the pentameric  $R_2M_2S_1$  Type I complex. Mutational analysis of seven conserved ‘DEAD box’ motifs in the R subunit showed impaired ATPase and endonuclease activity in all mutants, without having an effect on modification (87–89). ‘DEAD boxes’ are putative helicase-like motifs and are potentially involved in translocation (90–93). Mutations in motif I and II affected ATP binding and ATP hydrolysis, respectively, while

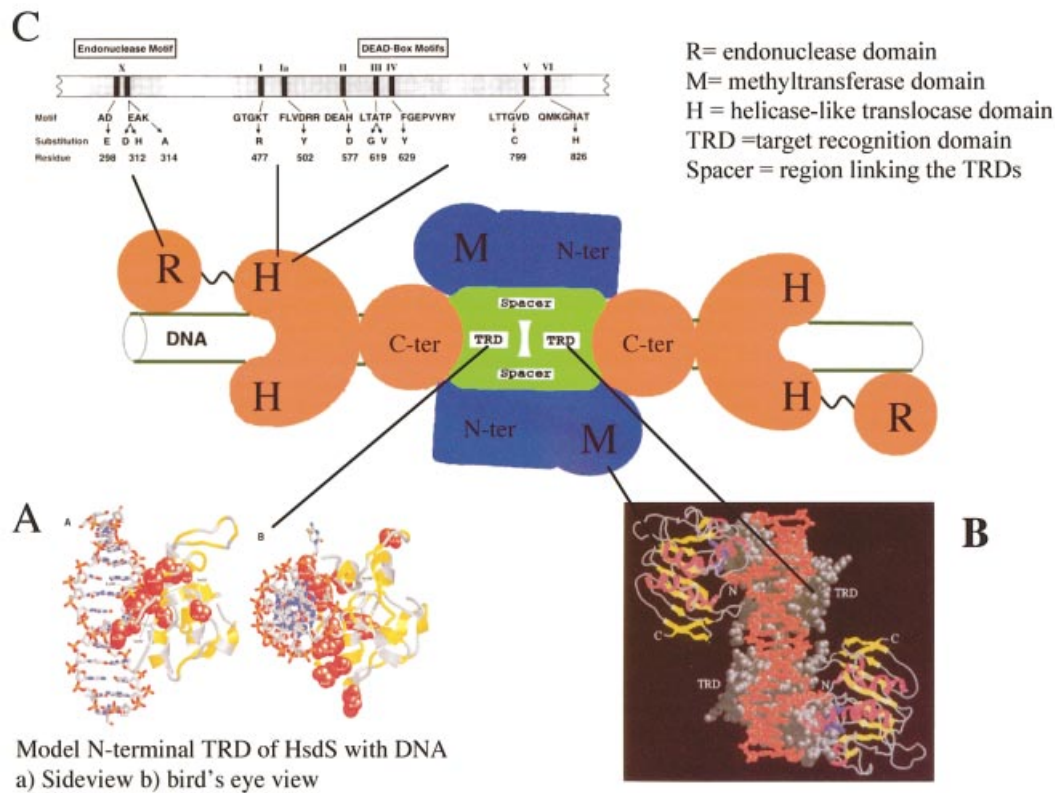
alterations in the other motifs inactivated the translocation properties. Purified mutant proteins did not linearize supercoiled DNA, had negligible ATPase activity *in vitro* and failed to translocate DNA *in vivo*, thus providing evidence for the coupling of ATP hydrolysis to translocation. The *in vivo* translocation assay used deserves particular mention, as it uses EcoKI-mediated transfer of T7 DNA from the phage head into the cell (94).

Sequence alignments and secondary structure predictions supported folding of the EcoKI DEAD box motifs into a structure resembling domains IA and 2A of helicase superfamily 2 with the motifs clustered around a cleft between two domains, each with  $\alpha$ -helices surrounding  $\beta$ -sheet cores (89,95). Recent evidence suggests that translocases and helicases share the DEAD box motifs that fold into a RecA-like domain with a large cleft through which the DNA is either pushed or pulled (92,93). While true helicases split the base pairs, translocases would use ATP hydrolysis to change the relative positions of additional domains to guide the DNA towards and into this cleft via DNA–backbone interactions. Such evidence would fit in with the notion that there is no obvious need for strand separation activity of Type I REases, nor for the recognition of specific base sequences during translocation.

In the N-terminal region of the R subunit, a conserved motif X, also present in other Type I and Type III enzymes, resembles the known active site in Type II REases and the RecB family of endonucleases (96). Mutations in this region abolish nicking and cutting activity, but have no effect on the ATPase activity or DNA translocation *in vivo* (89,95). Thus, translocation could be uncoupled from restriction and cutting was firmly relegated to a region distinct from the translocation domain. A common structure for the endonucleolytic site in all REases appears likely.

## MODEL FOR THE LOCATION OF THE CUT SITE

The debate on the choice of the cut site with respect to the recognition site started in the 1980s and led to the collision model (97). Cutting would be stimulated when two translocating complexes met between adjacent sites, consistent with *in vivo* evidence for cutting between targets (98). As a single site in a phage genome elicited restriction, it was suggested that at high enzyme-to-DNA ratios the excess enzyme would be involved in cutting. Evidence for the implied cooperative interactions between sites was confirmed by *in vivo* experiments in the absence of the *ral* gene of  $\lambda$  [or its chromosomal homologue *lar*, which alters the behavior of EcoKI on unmodified DNA (87)] and by atomic force microscopy (AFM) *in vitro*, which showed that linear DNA with two sites was more efficiently cleaved than that with only one target (99). AFM, in contrast to EM, allows gentle sample preparation via non-covalent attachment of protein–DNA complexes to a mica surface in aqueous solution. This further suggested that two EcoKI complexes bound their respective recognition sites on opposite sides of the target plasmid, dimerized, looped the DNA and cut it about 7 min after addition of ATP, due to stalling of the complex upon excessive DNA supercoiling or maximal contraction of the DNA loop between the two bound EcoKI molecules. This supported the model in which two target sites are preferentially used and cutting takes place in



**Figure 4.** Model of the structure of the EcoKI restriction complex [adapted from Davies *et al.* (95), reproduced with permission from Elsevier. Copyright 1999]. The two TRDs of the specificity subunit HsdS (green) recognize the two halves of the DNA recognition site AAC(N6)GTGC. The TRDs are linked by conserved sequence regions which function as subunit interfaces and also define the length of the non-specific DNA sequence in the middle of the recognition site. Two HsdM modification subunits (blue) bind to the conserved regions of HsdS via their N- and C-terminal domains. They wrap around the DNA helix on the opposite side of S, allowing access of the methyltransferase domain of M to DNA, presumably using base flipping as described for Type II MTases. Two HsdR subunits (orange) associate with M and S via the C-terminus. The central part of the protein is involved in translocation and contains 'DEAD box' motifs, characteristic of helicases (H). These motifs probably fold into two domains (IA and 2A) to form a cleft through which the DNA would pass (resembling a 'RecA-like' structure that may be common to all helicases/translocases). EcoKI belongs to helicase superfamily 2, whose members are believed to guide the DNA via regions outside the IA and IIA domains towards the cleft involving interactions with the DNA backbone (and not the bases), in line with the function of EcoKI as a DNA translocase rather than helicase. In the N-terminus of HsdR is a motif, characteristic of endonucleases (R). The enzyme binds the target site via HsdM and HsdS using AdoMet as cofactor for binding and distinguishing between hemimethylated and unmodified DNA. If unmodified, the enzyme undergoes a large conformational change and translocates the DNA past itself, while remaining bound to the recognition site, creating large loops visible by EM and AFM, concomitant with ATP hydrolysis. The model rests on extensive genetic, biochemical and biophysical evidence (see text for further details and references). (A) A model of amino acids 43–157 from the N-terminal TRD of EcoKI interaction with DNA [reproduced with permission from Sturrock, S.S. and Dryden, D.T.F. (1997) *Nucleic Acids Res.*, **25**, 3408–3414; 79]. (B) A front view from a partial model of a Type I MTase bound to DNA constructed using two copies of the structure of Type II MTases bound to DNA [from Dryden *et al.* (77) with permission from Nature Publishing Group (<http://www.nature.com/nsb/>)]. The TRD regions are based on the structure of the TRD from HhaI and the methyltransferase domains in the catalytic domains of TaqI. Space filling shows sites of mutations resulting in loss of specificity and activity. (C) Section of the HsdR subunit showing mutational analysis of conserved endonuclease and 'DEAD box' (helicase-like) motifs [from Davies *et al.* (89), reproduced with permission from Elsevier. Copyright 1999].

between (97). Translocase mutants were also capable of dimerization that might occur between any two occupied (not only adjacent) EcoKI sites (99,100), although interaction between adjacent sites is most probable. In all cases cutting would be due to stalling of the complex, as a result of topological constraints or a block by, e.g. a Holliday junction (3,4,97,101,102).

## RESTRICTION ALLEVIATION REVISITED

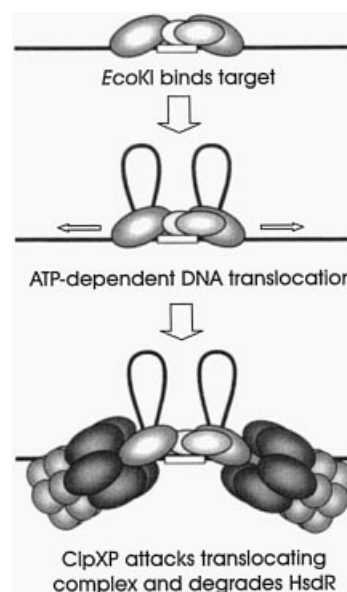
In the 1960s gene transfer experiments via transduction, conjugation and transformation laid the foundations of our current knowledge about the *hsd* locus (16,18). These and more recent data support the notion that (in contrast to 'selfish'

Type II genes) EcoKI genes easily replace alleles with different target recognition sites and thus modification protection pattern and vice versa, or alternatively mutant genes encoding non-modifying proteins (103–105). This led to a reinvestigation of the issue of RA in situations that lead to unmodified restriction sites in the bacterial chromosome, e.g. UV-induced (or other) DNA damage, repair, transfer of *hsd* genes and recombination (3,4,11,104–110). After conjugational transfer of *hsd* genes restriction activity is suppressed during many generations, whereas functional expression of the *hsdM* and *hsdS* genes is immediate [though complete modification requires several generations due to the large number of chromosomal targets (>600)]. Establishment of the *hsd* locus in the recipient was affected by a mutation in the *hsdC*

gene (105), identified as a Clp<sup>-</sup> mutant. From the study of the *hsdC* mutation, J. Ryu hinted the possibility of ClpXP regulation on restriction activity of EcoKI to N. Murray (J.Ryu, personal communication).

Sophisticated *in vivo* tests for restriction activity by EcoKI using complementation between  $\lambda$ *hsdK* and host mutants, as well as western blots, led Svetlana Makovets and Noreen Murray to the discovery of intricate post-translational control of the R subunit of EcoKI by the ATP-dependent ClpXP protease (107–110). A wild-type ClpXP complex in the recipient cell proved essential for efficient transfer of the *hsd* genes into a cell with unmodified chromosomal DNA. Rather elegantly the authors used the R and M mutants of EcoKI described above and established that the R subunit was degraded by ClpXP only (a) when modification was impaired and (b) after assembly of a specific DNA/EcoKI-translocation-proficient complex (Fig. 5). Protein complexes with mutations in the R polypeptide affecting ATP hydrolysis or DNA translocation (the above ‘DEAD box’ mutants), or wild-type R proteins in an MTase-deficient background (defects in motif II in the M subunit) did not lose their R subunit. Hence a modification-deficient EcoKI complex that could still bind AdoMet (wild-type motif I in the M subunit) would retain endonuclease activity. Finally, R proteins with a defect in the endonucleolytic site (the above motif X), but with intact translocase (‘DEAD box’ motifs) would be degraded like wild-type enzyme, showing that EcoKI became sensitive to ClpXP-mediated degradation before the final step in the restriction pathway. Western blots showed selective degradation of the R, but not the M polypeptide after DNA damage generated unmethylated DNA. This ClpXP control of restriction pointed towards an extraordinary control of R activity to prevent chromosomal destruction, if modification became even temporarily insufficient. Similar protection of chromosomal DNA was noted for the Type IB enzyme EcoAI, but not for plasmid-borne Type IC EcoR124I (107–109). In addition these important data showed that foreign (phage) DNA is destroyed by a reconstituted restriction-proficient EcoKI complex with a mutation in methyltransferase (motif II) while all (600-odd) chromosomal target sequences remain unharmed, leading to the concept of ‘self’ and ‘non-self’, and protection of self via control of autorestriction (4). Note that this mechanism does not interfere with additional alternative mechanisms for control of REase activity of EcoKI, e.g. via subunit assembly, as perhaps suggested for EcoKI by *in vitro* data and postulated for other enzymes (110–112).

An interesting hypothesis associates RA with R-M systems in which DNA translocation is an integral part of the restriction pathway. RA thus may play a role during incorporation of unmodified DNA into the recipient chromosome upon conjugational transfer. The random cleavage process of Type I enzymes, together with its great potential for diversification of sequence specificity via TRD domain shuffling, may be of crucial importance in facilitating general recombination, as they will place the recombinogenic ends at variable distances from Chi-sites that promote RecBC-mediated recombination. Large chunks of DNA entering the cell via transduction or conjugation would be processed to smaller recombinogenic fragments by Type I enzymes and thus may be of importance to populations as well as individual cells (4,41,113). This would fit in with the notion that, though



**Figure 5.** Model for the mechanism of ClpXP-dependent control of restriction. When EcoKI binds to an unmodified chromosomal target sequence, translocation starts. However, ClpXP recognizes the R subunit in the translocation-proficient state and destroys it, thereby preventing further translocation and cutting of the chromosome [from Murray (3) with permission from the American Society for Microbiology. Copyright 2000].

widespread, R-M systems are not essential for survival. Such a biological role for R-M systems like EcoKI would make sense, as chromosomes should be stable, not static. A certain low amount of influx of foreign material would be desirable, whether in this manner or e.g. via insertion elements. Introduction of small regions could enhance survival upon changing conditions, leading to mosaic sequences and slow evolution towards new species (4,41,113), a topic of interest also to Arber till today (114).

## CONCLUSIONS

Starting with a humble barrier to phage in 1953, EcoKI often blocked and frustrated scientific dreams over the last 50 years. Undeterred, dedicated scientists employed different techniques and explored many novel ideas along untrodden paths, with the present golden model to show for it. The unpleasant finding that EcoKI was not a site-specific endonuclease (barring it from lucrative cloning schemes) led to the interesting issues of DNA translocation and intricate Clp-mediated control of restriction. The unexpected lack of helix–turn–helix motifs in the DNA binding domains turned into an evolutionary triumph with the revolutionary idea that recombination generates new specificities. For the next 50 years, the role of translocation by Type I enzymes in chromosome integrity and evolution, the importance of recognition of the methylated state of the DNA with respect to self/non-self (but also to gene expression and cell cycle control in cancer biology, a topic outside the scope of this review), and finally dissection of the RA pathways by bacterial hosts and their enemies await further exploration. The resurgence of viral and plasmid antirestriction proteins that appear to target Type I restriction enzymes in a DNA sequence-independent way opens the way to gene transfer

experiments in Archaeobacteria and other non-enteric species, and may prove to be of both fundamental and commercial value in this respect. Finally, application of the avalanche of new computer programs in this post-genomic era may help to make sense of so many ATGCs in the database that may be novel Type I enzymes. This review starts and ends with Joe Bertani, as without his concoction called LB [an abbreviation for 'Lysogeny Broth' and not 'Luria-Bertani' broth! (115)], our bacteria (and his favorite phage P2) would not have grown so happily all these years, nor our experiments materialized.

## ACKNOWLEDGEMENTS

I wish to thank the following scientists for published and unpublished information used for this review: Werner Arber, Joe Bertani, Tom Bickle, Bill Brammar, Brian Wilkins, David Dryden and Noreen Murray. I am grateful to David Dryden and Noreen Murray for permission to use their figures and for critical reading of the manuscript. The idea to write up the EcoKI story was met with enthusiasm by colleagues and friends. Unfortunately, two of these contributors to the EcoKI Type I story, Gill Cowan and Brian Wilkins, have died since, and I therefore would like to dedicate this review to them.

## REFERENCES

- Stent,G.S. (1960) *Papers on Bacterial Viruses*. Little, Brown and Company, Boston, Toronto.
- Holmes,F.L. (2001) Meselson, Stahl and the Replication of DNA. *A History of the Most Beautiful Experiment in Biology*. Yale University Press, New Haven and London.
- Murray,N.E. (2000) Type I restriction systems. Sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.*, **64**, 412–434.
- Murray,N.E. (2002) Immigration control of DNA in bacteria: self versus non-self. *Microbiology*, **148**, 3–20.
- Dryden,D.T.F., Murray,N.E. and Rao,D.N. (2001) Nucleoside triphosphate-dependent restriction enzymes. *Nucleic Acids Res.*, **29**, 3728–3741.
- Bourniquel,A. and Bickle,T.A. (2002) Complex restriction enzymes: NTP-driven molecular motors. *Biochimie*, **84**, 1047–1059.
- Pingoud,A. and Jeltsch,A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res.*, **29**, 3705–3727.
- Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2003) REBASE: restriction enzymes and methylases. *Nucleic Acids Res.*, **31**, 418–420.
- Suri,B., Nagaraja,V. and Bickle,T.A. (1994) Bacterial DNA modification. *Curr. Top. Microbiol. Immunol.*, **180**, 1–9.
- Dryden,D.T.F. (1999) Bacterial DNA methyltransferases. In Cheng,X. and Blumenthal,R.M. (eds), *S-Adenosylmethionine-Dependent Methyltransferases: Structures and Functions*. World Scientific Publishing, Singapore, pp. 283–340.
- Bertani,G. and Weigle,J.J. (1953) Host controlled variation in bacterial viruses. *J. Bacteriol.*, **65**, 113–121.
- Luria,S.E. (1953) Host-induced modification of viruses. *Cold Spring Harbor Symp. Quant. Biol.*, **18**, 237–244.
- Arber,W. (1978) Promotion and limitation of genetic exchange. In *Les Prix Nobel 1978. The Nobel Foundation 1979*. Almquist and Wiksell International, Stockholm, Sweden.
- Arber,W. and Dussoix,D. (1962) Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage lambda. *J. Mol. Biol.*, **5**, 18–36.
- Dussoix,D. and Arber,W. (1962) Host specificity of DNA produced by *Escherichia coli*. II. Control over acceptance of DNA from infecting phage lambda. *J. Mol. Biol.*, **5**, 37–49.
- Arber,W. (1965) Host-controlled modification of bacteriophage. *Annu. Rev. Microbiol.*, **19**, 365–378.
- Boyer,H.W. (1964) Genetic control of restriction and modification in *Escherichia coli*. *J. Bacteriol.*, **88**, 1652–1660.
- Arber,W. and Linn,S. (1969) DNA modification and restriction. *Annu. Rev. Biochem.*, **38**, 467–500.
- Boyer,H.W. and Roulland-Dussoix,D. (1969) A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J. Mol. Biol.*, **41**, 459–472.
- Glover,S.W. and Colson,C. (1969) Genetics of host-controlled restriction and modification in *Escherichia coli*. *Genet. Res.*, **13**, 227–240.
- Meselson,M., Yuan,R. and Heywood,J. (1972) Restriction and modification of DNA. *Annu. Rev. Biochem.*, **41**, 447–466.
- Meselson,M. and Yuan,R. (1968) DNA restriction enzyme from *E.coli*. *Nature*, **217**, 1110–1114.
- Wood,W.B. (1966) Host specificity of DNA produced by *Escherichia coli*: bacterial mutations affecting the restriction and modification of DNA. *J. Mol. Biol.*, **16**, 118–133.
- Franklin,N.C. and Dove,W.F. (1969) Genetic evidence for restriction targets in the DNA of phages lambda and phi 80. *Genet. Res.*, **14**, 151–157.
- Horiuchi,K. and Zinder,N.D. (1972) Cleavage of bacteriophage phi DNA by the restriction enzyme of *Escherichia coli* B. *Proc. Natl Acad. Sci. USA*, **69**, 3220–3224.
- Roberts,R.J., Belfort,M., Bestor,T., Bhagwat,A., Bickle,T.A., Bitinaite,J., Blumenthal,R.M., Degtyarev,S.Kh., Dryden,D.T.F., Dybvig,K. et al. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–1812.
- Roberts,R.J. and Cheng,X. (1998) Base flipping. *Annu. Rev. Biochem.*, **67**, 181–198.
- Walkinshaw,M.D., Taylor,P., Sturrock,S.S., Atanasiu,C., Berge,T., Henderson,R.M., Edwardson,J.M. and Dryden,D.T.F. (2002) Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. *Mol. Cell*, **9**, 187–194.
- Glover,S.W. (1980) Genetics of restriction/modification systems. *Biochem. Soc. Trans.*, **8**, 395–396.
- Bickle,T.A. (1980) Adenosine triphosphate-requiring restriction enzymes. *Biochem. Soc. Trans.*, **8**, 396–397.
- Gingeras,T.R. (1980) Restriction endonucleases and their recognition sequences. *Biochem. Soc. Trans.*, **8**, 397–398.
- Yuan,R. (1981) Structure and mechanism of multifunctional restriction endonucleases. *Annu. Rev. Biochem.*, **50**, 285–315.
- Gough,J.A. and Murray,N.E. (1983) Sequence diversity among related genes for recognition of specific targets in DNA molecules. *J. Mol. Biol.*, **166**, 1–19.
- Loenen,W.A.M., Daniel,A.S., Braymer,H.D. and Murray,N.E. (1987) Organisation and sequence of the *hsd* genes of *Escherichia coli* K-12. *J. Mol. Biol.*, **198**, 159–170.
- Bickle,T.A. and Kruger,D.H. (1993) Biology of DNA restriction. *Microbiol. Rev.*, **57**, 434–450.
- Bickle,T.A. (1993) The ATP-dependent restriction enzymes. In Linn,S.M., Lloyd,R.S. and Roberts,R.J. (eds), *Nucleases*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 89–109.
- Bickle,T.A., Brack,C. and Yuan,Y. (1978) ATP-induced conformational changes in the restriction endonuclease from *Escherichia coli* K-12. *Proc. Natl Acad. Sci. USA*, **75**, 3099–3103.
- Yuan,R., Hamilton,D.L. and Burckhardt,J. (1980) DNA translocation by the restriction enzyme from *Escherichia coli* K. *Cell*, **20**, 237–244.
- Borck,K., Beggs,J.D., Brammar,W.J., Hopkins,A.S. and Murray,N.E. (1976) The construction of transducing derivatives of phage lambda. *Mol. Gen. Genet.*, **146**, 199–207.
- King,G. and Murray,N.E. (1994) Restriction enzymes in cells, not Eppendorfs. *Trends Microbiol.*, **2**, 465–469.
- Barcus,V.A. and Murray,N.E. (1995) Barriers to recombination: restriction. In Baumberg,S., Young,J.P.W., Saunders,S.R. and Wellington,E.M.H. (eds), *Population Genetics of Bacteria*. Society for General Microbiology, Cambridge University Press, Cambridge, UK, pp. 31–58.
- Sain,B. and Murray,N.E. (1980) The *hsd* (host specificity) genes of *E.coli* K-12. *Mol. Gen. Genet.*, **180**, 35–46.
- Studier,F.W. and Movva,N.R. (1976) SAMase gene of bacteriophage T3 is responsible for overcoming host restriction. *J. Virol.*, **19**, 136–145.
- Kruger,D.H. and Bickle,T.A. (1983) Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol. Rev.*, **47**, 345–360.

45. Spoerel, N., Herrlich, P. and Bickle, T.A. (1979) A novel bacteriophage defence mechanism: the anti-restriction protein. *Nature*, **278**, 30–34.
46. Ratner, D. (1974) The interaction of bacterial and phage proteins with immobilised *Escherichia coli* RNA polymerase. *J. Mol. Biol.*, **88**, 373–383.
47. Bandyopadhyay, P.K., Studier, F.W., Hamilton, D.L. and Yuan, R. (1985) Inhibition of the type I restriction-modification enzymes EcoB and EcoK by the gene 0.3 protein of bacteriophage T7. *J. Mol. Biol.*, **182**, 567–578.
48. Atanasiu, C., Su, T.-J., Sturrock, S. and Dryden, D.T.F. (2002) Interaction of the *ocr* gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. *Nucleic Acids Res.*, **30**, 3936–3944.
49. Zavilgelski, G.B. (2000) Antirestriction. *Mol. Biol.*, **34**, 724–732.
50. Wilkins, B.M. (2000) Plasmid promiscuity: meeting the challenge of DNA immigration control. *Environ. Microbiol.*, **4**, 495–500.
51. Thomas, A.T., Brammar, W.J. and Wilkins, B.M. (2003) Plasmid R16 *ArdA* protein preferentially targets restriction activity of the Type I restriction-modification system EcoKI. *J. Bacteriol.*, **185**, 2022–2025.
52. Bates, S., Roscoe, R.A., Althorpe, N.J., Brammar, W.J. and Wilkins, B.M. (1999) Expression of leading region genes on IncII plasmid ColIb-P9: genetic evidence for single-stranded DNA transcription. *Microbiology*, **145**, 2655–2662.
53. Janulaitis, A., Kaslauskienė, R., Lazarevičiūtė, L., Gilvonauskaitė, R., Steponavičienė, D., Jagelavicius, M., Petrusyte, M., Bitinaite, J., Vezeviciute, Z., Kiuduliene, E. and Butkus, V. (1988) Taxonomic specificity of restriction modification enzymes. *Gene*, **74**, 229–232.
54. Murray, N.E., Gough, J.A., Suri, B. and Bickle, T.A. (1982) Structural homologies among type I restriction–modification systems. *EMBO J.*, **1**, 535–539.
55. Wilson, G.G. and Murray, N.E. (1991) Restriction and modification systems. *Annu. Rev. Genet.*, **25**, 585–627.
56. Titheradge, A.J.B., King, J., Ryu, J. and Murray, N.E. (2001) Families of restriction enzymes: an analysis prompted by molecular and genetic data for type I restriction and modification systems. *Nucleic Acids Res.*, **29**, 4195–4205.
57. Bullas, L.R., Colson, C. and 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**, 166–172.
58. Bullas, L.R., Colson, C. and Neufeld, B. (1980) Deoxyribonucleic acid restriction and modification systems in *Salmonella*: chromosomally located systems of different serotypes. *J. Bacteriol.*, **141**, 275–292.
59. Fuller-Pace, F.V., Bullas, L.R., Delius, H. and Murray, N.E. (1984) Genetic recombination can generate altered restriction specificity. *Proc. Natl Acad. Sci. USA*, **81**, 6095–6099.
60. Fuller-Pace, F.V. and 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**, 9368–9372.
61. Nagaraja, V., Shepherd, J.C. and Bickle, T.A. (1985) A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity. *Nature*, **316**, 371–372.
62. Cowan, G.M., Gann, A.A. and Murray, N.E. (1989) Conservation of complex DNA recognition domains between families of restriction enzymes. *Cell*, **56**, 103–109.
63. Kneale, G.G. (1994) A symmetrical model for the domain structure of type I DNA methyltransferases. *J. Mol. Biol.*, **243**, 1–5.
64. Thorpe, P.H., Ternent, D. and Murray, N.E. (1997) The specificity of StySKI, a type I restriction enzyme, implies a structure with rotational symmetry. *Nucleic Acids Res.*, **25**, 1694–1700.
65. 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.
66. Zabeau, M.S., Friedman, M., Van Montagu, M. and Schell, J. (1980) The *ral* gene of phage lambda. I. Identification of a non-essential gene that modulates restriction and modification in *E.coli*. *Mol. Gen. Genet.*, **179**, 63–73.
67. Loenen, W.A.M. and Murray, N.E. (1986) Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage lambda. *J. Mol. Biol.*, **190**, 11–22.
68. Kelleher, J.E., Daniel, A.S. and Murray, N.E. (1991) Mutations that confer *de novo* activity upon a maintenance methyltransferase. *J. Mol. Biol.*, **221**, 431–440.
69. King, G. and Murray, N.E. (1995) Restriction alleviation and modification enhancement by the *rac* prophage of *E.coli* K-12. *Mol. Microbiol.*, **16**, 769–777.
70. Roberts, R.J. and Halford, S.E. (1993) Type II restriction enzymes. In Linn, S.M., Lloyd, R.S. and Roberts, R.J. (eds), *Nucleases*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 35–88.
71. Barcus, V.A., Titheradge, A.J. and Murray, N.E. (1995) The diversity of alleles at the *hsd* locus in natural populations of *Escherichia coli*. *Genetics*, **140**, 1187–1197.
72. Sharp, P.M., Kelleher, J.E., Daniel, A.S., Cowan, G.M. and Murray, N.E. (1992) Roles of selection and recombination in the evolution of type I restriction-modification systems in enterobacteria. *Proc. Natl Acad. Sci. USA*, **89**, 9836–9840.
73. Dryden, D.T.F., Cooper, L.P. and Murray, N.E. (1993) Purification and characterization of the methyltransferase from the type I restriction and modification system of *Escherichia coli* K12. *J. Biol. Chem.*, **268**, 13228–13236.
74. Powell, L.M. and Murray, N.E. (1995) S-Adenosyl methionine alters the DNA contacts of the EcoKI methyltransferase. *Nucleic Acids Res.*, **23**, 967–974.
75. Willcock, D.F., Dryden, D.T.F. and Murray, N.E. (1994) A mutational analysis of the two motifs common to adenine methyltransferases. *EMBO J.*, **13**, 3902–3908.
76. Cooper, L.P. and Dryden, D.T.F. (1994) The domains of a type I DNA methyltransferase: interactions and role in recognition of DNA methylation. *J. Mol. Biol.*, **236**, 1011–1021.
77. Dryden, D.T.F., Sturrock, S.S. and Winter, M. (1995) Structural modelling of a type I DNA methyltransferase. *Nat. Struct. Biol.*, **2**, 632–635.
78. Sturrock, S.S. and Dryden, D.T.F. (1997) A prediction of the amino acids and structures involved in DNA recognition by type I DNA restriction and modification enzymes. *Nucleic Acids Res.*, **25**, 3408–3414.
79. O'Neill, M., Dryden, D.T.F. and Murray, N.E. (1998) Localization of a protein–DNA interface by random mutagenesis. *EMBO J.*, **17**, 7118–7127.
80. Pinarbasi, H., Pinarbasi, E. and Hornby, D.P. (2003) The small subunit of M. AquI is responsible for sequence-specific DNA recognition and binding in the absence of the catalytic domain. *J. Bacteriol.*, **185**, 1284–1288.
81. O'Neill, M., Powell, L.M. and Murray, N.E. (2001) Target recognition by EcoKI: the recognition domain is robust and restriction-deficiency commonly results from the proteolytic control of enzyme activity. *J. Mol. Biol.*, **307**, 951–963.
82. Chen, A., Powell, L.M., Dryden, D.T.F., Murray, N.E. and Brown, T. (1995) Tyrosine 27 of the specificity polypeptide of EcoKI can be UV crosslinked to a bromodeoxyuridine-substituted DNA target sequence. *Nucleic Acids Res.*, **23**, 1177–1183.
83. Powell, L.M., Dryden, D.T.F., Willcock, D.F., Pain, R.H. and Murray, N.E. (1993) DNA recognition by the EcoK methyltransferase: the influence of DNA methylation and the cofactor S-adenosyl-L-methionine. *J. Mol. Biol.*, **234**, 60–71.
84. Powell, L.M., Connolly, B.A. and Dryden, D.T.F. (1998) The DNA binding characteristics of the trimeric EcoKI methyltransferase and its partially assembled dimeric form determined by fluorescence polarisation and DNA footprinting. *J. Mol. Biol.*, **283**, 947–961.
85. Powell, L.M., Lejeune, E., Hussain, F.S., Cronshaw, A.D., Kelly, S.M., Price, N.C. and Dryden, D.T.F. (2003) Assembly of EcoKI methyltransferase requires the C-terminal region of the HsdM modification subunit. *Biophys. Chem.*, **103**, 129–137.
86. Powell, L.M., Dryden, D.T.F. and Murray, N.E. (1998) Sequence-specific DNA binding by EcoKI, a type IA DNA restriction enzyme. *J. Mol. Biol.*, **283**, 963–976.
87. Webb, J.L., King, G., Ternent, D., Titheradge, A.J.B. and Murray, N.E. (1996) Restriction by EcoKI is enhanced by cooperative interactions between target sequences and is dependent on DEAD box motifs. *EMBO J.*, **15**, 2003–2009.
88. Davies, G.P., Powell, L.M., Webb, J.L., Cooper, L.P. and Murray, N.E. (1998) EcoKI with an amino acid substitution in any one of seven DEAD-box motifs has impaired ATPase and endonuclease activities. *Nucleic Acids Res.*, **26**, 4828–4836.
89. Davies, G.P., Kemp, P., Molineux, I.J. and Murray, N.E. (1999) The DNA translocation and ATPase activities of restriction-deficient mutants of EcoKI. *J. Mol. Biol.*, **292**, 787–796.

90. Gorbalenya, A.E. and Koonin, E.V. (1991) Endonuclease (R) subunits of type-I and type-III restriction-modification enzymes contain a helicase-like domain. *FEBS Lett.*, **291**, 277–281.
91. Hall, M.C. and Matson, S.W. (1999) Helicase motifs: the engine that powers DNA unwinding. *Mol. Microbiol.*, **34**, 867–877.
92. Singleton, M.R. and Wigley, D.B. (2003) Multiple roles for ATP hydrolysis in nucleic acid modifying enzymes. *EMBO J.*, **22**, 4579–4583.
93. Singleton, M.R. and Wigley, D.B. (2002) Modularity and specialization in Superfamily 1 and 2 helicases. *J. Bacteriol.*, **184**, 1819–1826.
94. Garcia, L.R. and Molineux, I.J. (1999) Translocation and specific cleavage of bacteriophage T7 DNA *in vivo* by EcoKI. *Proc. Natl Acad. Sci. USA*, **96**, 12430–12435.
95. Davies, G.P., Martin, I.N., Sturrock, S.S., Cronshaw, A., Murray, N.E. and Dryden, D.T.F. (1999) On the structure and operation of type I DNA restriction enzymes. *J. Mol. Biol.*, **290**, 565–579.
96. Titheradge, A.J.B., Ternent, D. and Murray, N.E. (1996) A third family of allelic *hsd* genes in *Salmonella enterica*: sequence comparisons with related proteins identify conserved regions implicated in restriction of DNA. *Mol. Microbiol.*, **22**, 437–447.
97. Studier, F.W. and Bandyopadhyay, P.K. (1988) Model for how type I restriction enzymes select cleavage sites in DNA. *Proc. Natl Acad. Sci. USA*, **85**, 4677–4681.
98. Brammar, W.J., Murray, N.E. and Winton, S. (1974) Restriction of  $\lambda$ trp bacteriophages by *Escherichia coli* K. *J. Mol. Biol.*, **90**, 633–647.
99. Ellis, D., Dryden, D.T.F., Berge, T., Edwardson, J.M. and Henderson, R.M. (1998) Direct observation of DNA translocation and cleavage by atomic force microscopy. *Nat. Struct. Biol.*, **5**, 15–17.
100. Berge, T., Ellis, D.J., Dryden, D.T.F., Edwardson, J.M. and Henderson, R.M. (2000) Translocation-independent dimerization of the EcoKI endonuclease visualized by atomic force microscopy. *Biophys. J.*, **79**, 479–484.
101. Taylor, A.F. and Smith, G.R. (1990) Action of RecBCD enzyme on cruciform DNA. *J. Mol. Biol.*, **211**, 117–134.
102. Janscak, P., MacWilliams, M.P., Sandmeier, U., Nagaraja, V. and Bickle, T.A. (1999) DNA translocation blockage, a general mechanism of cleavage site selection by type I restriction enzymes. *EMBO J.*, **18**, 2638–2647.
103. O'Neill, M., Chen, A. and Murray, N.E. (1997) The restriction-modification genes of *Escherichia coli* K-12 may not be selfish: they do not resist loss and are readily replaced by alleles conferring different specificities. *Proc. Natl Acad. Sci. USA*, **94**, 14596–14601.
104. Prakash-Cheng, A., Chung, S.S. and Ryu, J. (1993) The expression and regulation of *hsdK* genes after conjugative transfer. *Mol. Gen. Genet.*, **241**, 491–496.
105. Prakash-Cheng, A. and Ryu, J. (1993) Delayed expression of *in vivo* restriction activity following conjugal transfer of *Escherichia coli hsdK* (restriction-modification) genes. *J. Bacteriol.*, **175**, 4905–4906.
106. Prakash, A., Valinluck, B. and Ryu, J. (1991) Genomic *hsd-Mu(lac)* operon fusion mutants of *Escherichia coli* K-12. *Gene*, **99**, 9–14.
107. Makovets, S., Titheradge, A.J.B. and Murray, N.E. (1998) ClpX and ClpP are essential for the efficient acquisition of genes specifying type IA and IB restriction systems. *Mol. Microbiol.*, **28**, 25–35.
108. Makovets, S., Doronina, V.A. and Murray, N.E. (1999) Regulation of endonuclease activity by proteolysis prevents breakage of unmodified bacterial chromosomes by type I restriction enzymes. *Proc. Natl Acad. Sci. USA*, **96**, 9757–9762.
109. Doronina, V.A. and Murray, N.E. (2001) The proteolytic control of restriction activity in *Escherichia coli* K-12. *Mol. Microbiol.*, **39**, 416–428.
110. Makovets, S., Powell, L.M., Titheradge, A.J.B., Blakely, G.W. and Murray, N.E. (2003) Is modification sufficient to protect a bacterial chromosome from a resident restriction endonuclease? *Mol. Microbiol.*, DOI: 10.1046/j.1365-2958.2003.03801.x.
111. Dryden, D.T.F., Cooper, L.P., Thorpe, P.H. and Byron, O. (1997) The *in vitro* assembly of the EcoKI type I DNA restriction/modification enzyme and its *in vivo* implications. *Biochemistry*, **36**, 1065–1076.
112. Janscak, P., Dryden, D.T.F. and Firman, K. (1998) Analysis of the subunit assembly of the type IC restriction-modification enzyme EcoR124I. *Nucleic Acids Res.*, **26**, 4439–4445.
113. Price, C. and Bickle, T.A. (1986) A possible role for DNA restriction in bacterial evolution. *Microbiol. Sci.*, **3**, 296–299.
114. Arber, W. (2003) Elements for a theory of molecular evolution. *Gene*, **317**, 3–11.
115. Bertani, G. (2004) Lysogeny at mid-twentieth century: P1, P2 and other experimental systems. *J. Bacteriol.*, in press.