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 λ 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 λ 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 [(bacte)iro]phages]. Phages were first described during World War I and pioneering work with λ 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 R2M2S1 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
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 λ 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 λ infection 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 λ propagated on *E. coli* C (called λ.C or λ.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.

![Figure 1](image)

**Figure 1.** The Bertani and Weigle experiment (1953). The discovery of a barrier to infection of λ 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 λ propagated on *E. coli* C (called λ. C or λ.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.

**THE ARBER AND DUSSOIX 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 (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 λ 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 λ. 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 λ-susceptible 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 (EcoP11), for which he was awarded the Nobel Prize for Medicine in 1978. With Daisy Dussoix, he published the first formal evidence that EcoP11 restricted or modified λ 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 ENZYMOLGY**

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 unmethylated 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 palindromic 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 physiological conditions (18,24).

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 hsds 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 λ hsds phage was extended in vivo via reintegration and excision from the E.coli chromosome to generate λhsdRMS, a phage also capable of restriction (42).

Finally, RA by bacteriophages deserves special mention (43–48). In contrast to λ, the T3 and T7 phages grew interexchangeably 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 EcoB1) 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 EcoK1, 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 hsdsK, 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. 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, StySBL1 (3,4,5,55,56).

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

<table>
<thead>
<tr>
<th>TRD</th>
<th>Enzyme</th>
<th>Recognition sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>StySPI</td>
<td>AAC (N6) GTRCG</td>
</tr>
<tr>
<td></td>
<td>StyLTII</td>
<td>GAG (N6) RTAVYG</td>
</tr>
<tr>
<td></td>
<td>StySJ</td>
<td>GAG (N6) GTRCG</td>
</tr>
<tr>
<td></td>
<td>StySQ</td>
<td>AAC (N6) RTAVYG</td>
</tr>
<tr>
<td></td>
<td>EcoR124I</td>
<td>GAA (N6) RTCG</td>
</tr>
<tr>
<td></td>
<td>EcoR124II</td>
<td>GAA (N7) RTCG</td>
</tr>
<tr>
<td></td>
<td>EcoDXXI</td>
<td>TCA (N7) RTCG</td>
</tr>
<tr>
<td></td>
<td>EcoDXXI</td>
<td>TCA (N8) TCA</td>
</tr>
</tbody>
</table>

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), StyLTII and StySJ the N-terminal GAG-recognition domain (green) and finally StyLTII and StySQ both recognize RTAYG 3' (orange). (b) EcoR124I differs from EcoR124II in a TALE 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. EcoDXXIXII 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). λ 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 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 M2S1 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 Lr 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 superfamilies (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 γ class of Type II MTases. This suggested a primordial MTase gene for Type I and Type II enzymes with conserved α-helix/β-strand/α-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 $R_5M_2S_1$ 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 α-helices surrounding β-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 λ [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
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

---

**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 Hhal and the methyltransferase domains in the catalytic domains of TaqI. Space filling shows sites of mutations resulting in loss of specificity and activity. (C) A 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].
gene (105), identified as a Clpε 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 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 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 (burning 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 Archaebacteria 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 conception 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