Restriction-Modification Systems as Minimal Forms of Life

I. Kobayashi

1 Introduction

A restriction (R) endonuclease recognizes a specific DNA sequence and introduces a double-strand break (Fig. 1A). A cognate modification (M) enzyme methylates the same sequence and thereby protects it from cleavage. Together, these two enzymes form a restriction-modification system. The genes encoding the restriction endonuclease and the cognate modification enzyme are often tightly linked and can be termed a restriction-modification gene complex. Restriction enzymes will cleave incoming DNA if it has not been modified by a cognate or another appropriate methyltransferase (Fig. 1B). Consequently, it is widely believed that restriction-modification systems have been maintained by bacteria because they serve to defend the cells from infection by viral, plasmid, and other foreign DNAs (*cellular defense hypothesis*).

An alternative hypothesis for the maintenance of restriction-modification systems is based on the observation that several restriction-modification gene complexes in bacteria are not easily replaced by competitor genetic elements because their loss leads to cell death (*post-segregational killing*; Naito et al. 1995; Handa et al. 2001; Sadykov et al. 2003; Figs. 1C, 2B). This finding led to the proposal that these complexes may actually be one of the simplest forms of life, similar to viruses, transposons, and homing endonucleases. This *selfish gene hypothesis* (Naito et al. 1995; Kusano et al. 1995; Kobayashi 1996, 1998, 2001) is now supported by many lines of evidence from genome analysis and experimentation.

A third type of hypothesis that explains why restriction-modification systems are present assumes that they aid the generation of diversity (Arber 1993; Price and Bickle 1986; *variation hypothesis*). Supporting this notion is that these systems are indeed associated with genome variation in a number of different ways (Sect. 2). However, such restriction-modification-associated

I. Kobayashi
Division of Molecular Biology, Institute of Medical Science,
University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

Nucleic Acids and Molecular Biology, Vol. 14
Alfred Pingoud (Ed.)
Restriction Endonucleases

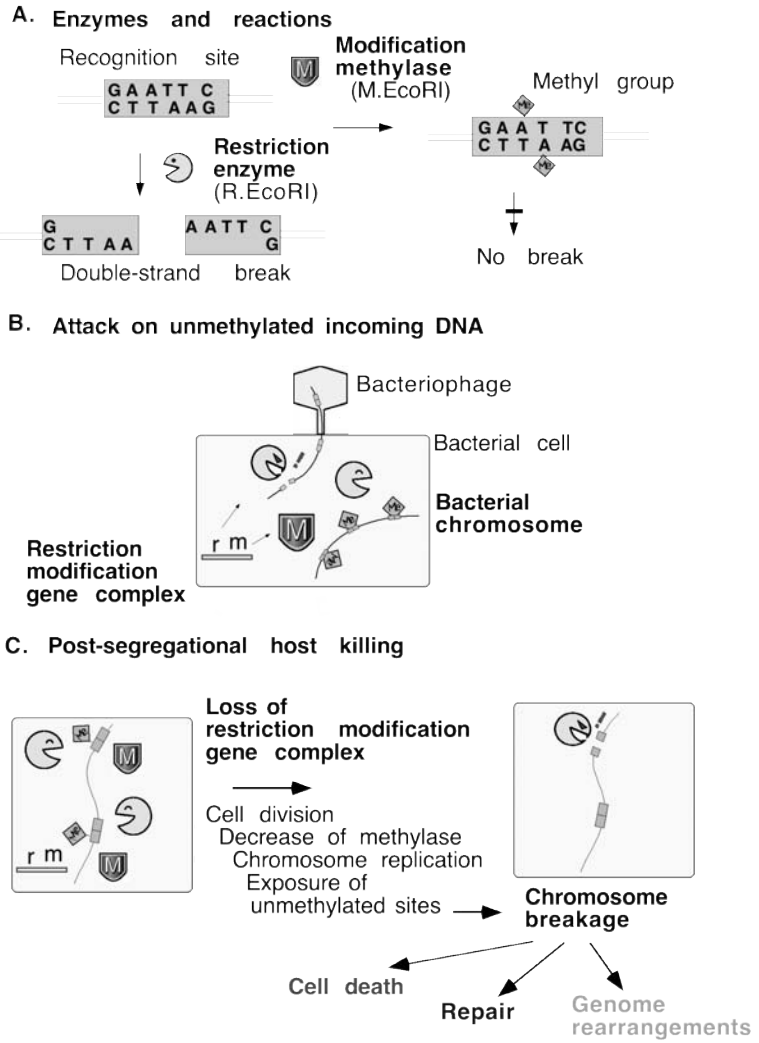

Fig. 1. Action of a restriction-modification gene complex. **A** Restriction enzyme (toxin) and modification methyltransferase (antitoxin). The antitoxin (modification enzyme) protects the targets of the toxin (restriction enzyme) by methylation. **B** Attack on incoming DNA. An attack on invading DNA that is not appropriately methylated is likely to be beneficial to the restriction-modification gene complex and to its host. **C** A simple dilution model for post-segregational killing. After loss of the restriction-modification gene complex, the toxin (restriction enzyme) and antitoxin (modification enzyme) will become increasingly diluted through cell division. Finally, too few modification enzyme molecules remain to defend all the recognition sites present on the newly replicated chromosomes. Any one of the remaining molecules of the restriction enzyme can attack these exposed sites. The chromosome breakage then leads to extensive chromosome degradation, and the cell dies unless the breakage is somehow repaired. The chromosome breakage may stimulate recombination and generate a variety of rearranged genomes, some of which might survive. *rm* Restriction-modification gene complex. Reproduced from *Nucleic Acids Research* (Kobayashi 2001)

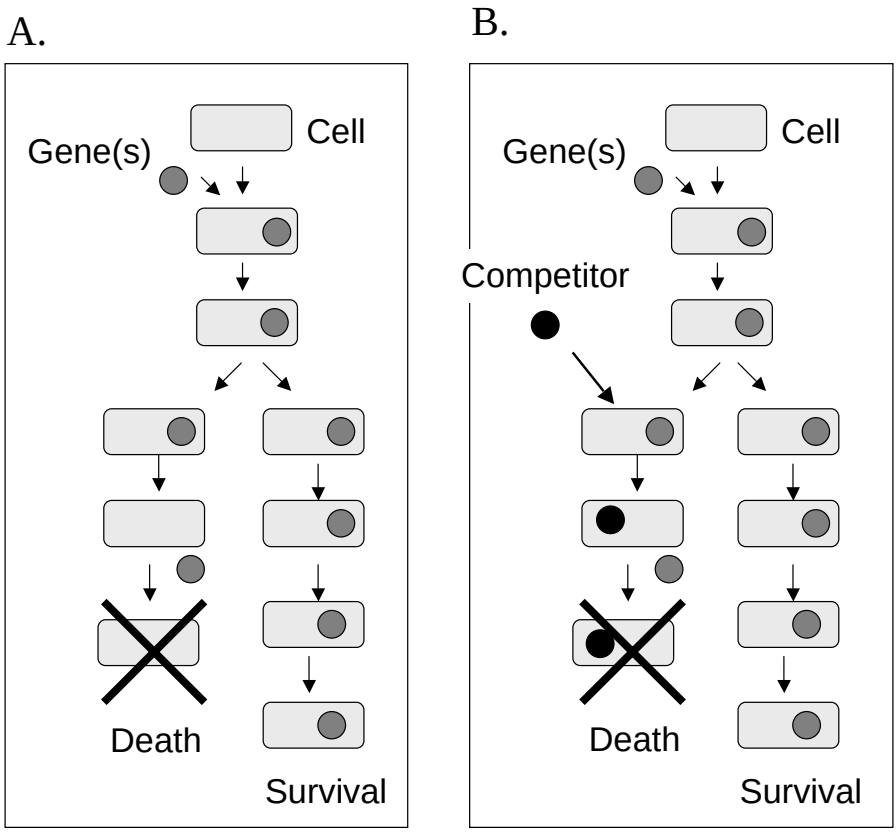

Fig. 2. The principle of post-segregational killing. **A** Once established in a cell, the addiction gene complex is difficult to eliminate because its loss, or some sort of threat to its persistence, leads to cell death. Intact copies of the gene complex survive in the other cells of the clone. **B** Advantage in competitive exclusion. A specific case of post-segregational cell killing showing the fight the gene complex raises against an incoming competing genetic element

genome variation can also be explained by the selfish gene hypothesis, as will be outlined (Sect. 6).

In this chapter, I will first review the evidence supporting the notion that some restriction-modification gene complexes behave as mobile genetic elements that may induce genome variability (Sect. 2). Next, I will describe their attacks on the genome and their consequences and then present the selfish gene hypothesis in detail (Sect. 3). This is followed by a review of the gene organization of these complexes and how they are regulated in relation to their life cycle (Sect. 4). The competition that exists between restriction-modification gene complexes is then described along with the other types of intragenomic interactions involving these complexes (Sect. 5). The effect of the parasitic selfish behavior of restriction-modification complexes on the

genome, in particular, their ability to induce mutagenesis and recombination, will then be discussed. This will illustrate how the host-parasite-type interactions between restriction-modification complexes and the genome contribute to genomic evolution (Sect. 6). How the selfish gene point of view can aid the classification of these complexes is described in Section 7. The next section (Sect. 8) discusses how these systems can be utilized in practical terms. The penultimate section (Sect. 9) proposes that the attack on the host by restriction-modification systems upon their disturbance reflects a general feature of genes that are assembled in a chromosome. The last section (Sect. 10) draws some conclusions.

This work owes much to other publications and databases. Particularly helpful were a brief but insightful review on programmed cell death in bacteria (Yarmolinsky 1995), my own reviews on restriction-modification systems (Kobayashi 2001) and on post-segregational killing systems (Kobayashi 2003b), and an extensive database on restriction enzymes and their genes, namely, REBASE (Roberts et al. 2003b) [http://rebase.neb.com]. To minimize the number of citations, only a few of the possible references have been cited. I welcome feedback on the novel view of restriction enzymes that is detailed in this chapter.

2 Genomics and Mobility of Restriction-Modification Systems

The decoding of several bacterial genomes has provided ample evidence of the variability and potential mobility of restriction-modification systems. Here I will review these lines of evidence. The question of how this variability/mobility is generated will be addressed again in Section 3.4.

2.1 Genomics

The restriction-modification gene homologues that have been identified in completely sequenced bacterial genomes are listed in REBASE. Some of these genomes – for example, those of *Haemophilus influenzae, Methanococcus jannaschii, Helicobacter pylori, Neisseria meningitidis, Neisseria gonorrhoeae,* and *Xylella fastidiosa* – have impressive numbers of restriction-modification gene homologues. Many restriction-modification gene homologues are specific to one strain within a given species, as has been noted for *Escherichia coli* [REBASE], *N. gonorrhoeae* [REBASE] and *H. pylori* (Alm et al. 1999; Nobusato et al. 2000a) [REBASE]. For example, comparison of the modification enzyme homologues in two completely sequenced strains of *H. pylori* revealed that while many pairs were very similar to each other, some homologues occurred in only one strain (Alm et al. 1999; Nobusato et al. 2000a).

2.2 Horizontal Gene Transfer Inferred from Evolutionary Analyses

Various types of evolutionary analyses suggest that restriction-modification genes have undergone extensive horizontal transfer between different groups of microorganisms (Table 1 (4); Kobayashi et al. 1999; Kobayashi 2001). Early studies found that close homologues occur in distantly related organisms such as Eubacteria and Archaea (archaebacteria) (e.g. Nolling and de Vos 1992). Extensive sequence alignment and phylogenetic tree construction now provide strong support for this point (Nobusato et al. 2000a; Bujnicki 2001). The incongruence in the same species of the phylogenetic tree of the methyltransferases with the tree of ribosomal RNA genes is additional evidence of the extensive horizontal transfer that the restriction-modification genes appear to have experienced (Nobusato et al. 2000a). Moreover, the GC content and/or codon usage of restriction-modification genes often differ from those of the majority of the genes in the genome (Jeltsch and Pingoud 1996; Alm et al. 1999; Nobusato et al. 2000a; Chinen et al. 2000b). This indicates that some restriction-modification genes may have joined the genome relatively recently by horizontal transfer from distantly related bacteria.

2.3 Presence on Mobile Genetic Elements

Sometimes there are hints for the molecular basis of the variability and horizontal transfer of restriction-modification gene complexes. One of these hints is that these complexes are often found on a variety of mobile genetic elements [Table 1 (2)]. For example, many restriction-modification gene complexes reside on plasmids (Table 2, B). Many of the cases of strain-specific restriction-modification systems in *E. coli* can be explained by their presence on plasmids [REBASE]. Moreover, some restriction-modification gene homologues have been found in a prophage in the chromosome [Table 1 (2)]. Others are on transposons, conjugative transposons (or integrative conjugative elements), genomic islands, and integrons [Table 1 (2)]. Restriction-modification gene homologues are also sometimes found to be linked with mobility-related genes, although the significance of this linkage is less clear than with the above cases (Xu et al. 1998; Vaisvila et al. 1995).

2.4 Genomic Contexts and Genome Comparison

Close examination of the genomic neighborhood of restriction-modification gene homologues and its comparison with a closely related genome also sometimes provide hints as to how restriction-modification gene complexes can enter a genome [Table 1 (3), Fig. 3].