CHAPTER 1

Mechanisms of homologous recombination in bacteria

Marie-Agnès Petit

Homologous recombination promotes the pairing between identical – or nearly identical – DNA sequences and the subsequent exchange of genetic material between them. It is an important and widely conserved function in living organisms, from bacteria to humans, that serves to repair double-stranded breaks or single-stranded gaps in the DNA, arising as a consequence of ionizing radiations, ultraviolet (UV) light, or chemical treatments creating replication-blocking adducts (Kuzminov, 1999). More recently, homologous recombination functions were also found in bacteria to rescue replication forks that have stalled for various reasons, such as a missing factor (e.g., the helicase), or a particular difficulty upstream of the fork, such as supercoiling or intense traffic of proteins (Michel et al., 2001).

Besides its molecular role, homologous recombination has played a major role in genome dynamics, by changing gene copy numbers through deletions, duplications, and amplifications: Intrachromosomal recombination between ribosomal operons or between mobile elements scattered into the genome leads to deletion or tandem duplications of large regions within the genome, up to several hundred kilobases (Roth et al., 1996). The duplications are unstable. Mostly they recombine back to the parental organization, and, therefore, remain undetected, except when appropriate selection, by gene dosage mostly, is exerted (Petes and Hill, 1988). In contrast, such duplications are ideal substrate for the diversification of genes: One gene is kept intact whereas the other is mutagenized, which leads to the birth of gene families. Once the duplicated segment has sufficiently diverged, it becomes more stable because of the lack of perfect homology to recombine the duplicated segment. Tandem duplications are also the starting point for further gene amplification, the repetition up to 20- to 100-fold of the tandem array (Kodama et al., 2002).

Finally, homologous recombination is also critical in terms of evolution, by allowing the generation of new allele combinations, and, as a consequence, the possibility to evolve and adapt to new environments, a hallmark of living organisms. In eukaryotes this happens mainly through meiosis, whereas in bacteria and archea the so-called "horizontal transfer" of genes is taking place on a larger scale (Ochman, Lawrence, and Groisman, 2000). Homologous recombination is one of the mechanisms through which such gene transfers occur, in particular during generalized transduction, conjugation, and natural transformation.

Much of what is known at the molecular level about homologous recombination in bacteria is from the in-depth work realized over the last 50 years on *Escherichia coli* (*E. coli*). For more recent reviews on this topic, the reader is referred to Kuzminov (1999) and Cox (2001). This chapter begins with a brief description of the knowledge based on the *E. coli* paradigm, but its main focus is on how other eubacteria resemble or differ from the paradigm. Because this book is on genomic rearrangements, plasmid recombination, which is a field in itself, is excluded.

HOMOLOGOUS RECOMBINATION: THE DNA ACTORS

Toward a definition of homologous recombination

The more processes of homologous recombination are known at the molecular level, the more difficult they are to be defined precisely. Concerning the DNA partners, in the original definition, homologous recombination concerned only events between pairs of chromosome homologs, and, therefore, was restricted to diploid cells. It then appeared that recombination could also concern two sequences at different loci (either in the same or in different chromosomes), the so-called "ectopic recombination." Finally, and especially in bacteria, homologous recombination was found to be a major way to integrate incoming DNA into a genome.

At the molecular level, homologous pairing and strand exchange may occur between two DNA molecules without any consequence at the genetic level, the so-called "non–cross-over" products (see the section "The DNA intermediates" in this chapter). The process is silent phenotypically, but essential molecularly, as it leads to DNA repair. During such a process, however, some point mutations of one molecule engaged into the homologous pairing may be transferred to the other molecule, and lead to gene conversion. This is not only frequent in fungi but also takes place in bacteria (Abdulkarim and Hughes, 1996). Finally, the enzymes that process homologous recombination

intermediates are able to pair sequences that are not identical, but partially diverged or homeologous (see the following section "The DNA products"). To summarize, homologous recombination does not necessarily recombine DNA and does not necessarily involve chromosome homologs, or identical sequences. In bacteria, one may define homologous recombination as all *recA*-dependent events, but even this simple assessment is not always true (see "RecA-independent homologous recombination" in this chapter).

The DNA products

Homologous recombination has its primary consequence at the DNA level; therefore, the main products of the process are first described. Most bacterial genomes are circular, and recombination products leading to linear chromosomes are lethal, so the focus is on circular products.

When the bacterial chromosome recombines with incoming DNA, as is the case during horizontal transfer, two main products are expected: If the entering DNA molecule contains two different stretches of homology with the bacterial chromosome (Fig. 1.1A), recombination can proceed by "double cross-over" (DCO). In this case, the intervening part of the chromosome is exchanged with the incoming DNA, and, therefore, lost. The DCO product is stable because no repeated sequences flank the incoming DNA. If the entering molecule contains a single stretch of homologous DNA and is circular – for instance, a nonreplicative plasmid – recombination occurs by single crossing over (sometimes called "Campbell-type" recombination, Fig. 1.1B). It produces a recombined chromosome in which the incoming

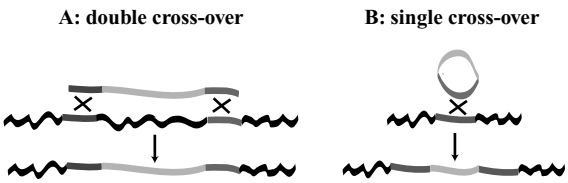

Figure 1.1. Two examples of integration of incoming DNA (flat lines, grey color) into the bacterial chromosome (black wavy lines). **(A)** The incoming DNA is linear and contains two regions of homology (dark grey) with the chromosome, each recombines (between the regions shown as a cross), and the resulting recombinant has integrated the foreign DNA by double cross-over. **(B)** The incoming DNA is circular and contains one region of homology (dark grey) with the chromosome. Upon recombination by single cross-over (shown as a cross), the foreign DNA is flanked by two copies of the homologous region, oriented in parallel.

DNA is flanked by directly repeated sequences, sometimes called "pop-in" recombinant. No DNA has been lost, but the resulting recombinant is unstable because it can "pop out" by homologous recombination. In cases where the introduced DNA confers a selective advantage, such recombinants are maintained. This process is widely used by geneticists to interrupt genes in bacteria such as *Bacillus subtilis*, *Lactococcus lactis*, *Deinococcus radiodurans*, and so on.

Chromosomal DNA can also generate intramolecular recombinants. This happens due to recombination between members of a gene family dispersed in the genome, typically between *rrn* operons or between mobile elements. If the two identical copies are inverted with respect to one another, the product is an inversion of the intervening sequence. If the two copies are oriented the same way, the process is called unequal crossing over because recombination probably takes place in an "unequal way" between sister chromatids behind a replication fork (Fig. 1.2). It leads to one chromosome containing a duplicated stretch flanked by the sequences that served to initiate the cross-over, and the other chromosome deleted for this same stretch of DNA, most likely unable to give a progeny. The chromosome containing the duplicated region is called a merodiploid or partial diploid; it is highly unstable and tends to recombine back to its original configuration. However, the frequency of production of these merodiploid is quite high: At any given time for one particular duplication, around 10^{-4} of an *E. coli* K12 population

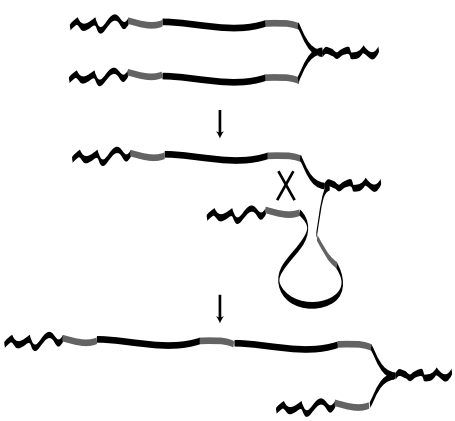

Figure 1.2. Intrachromosomal recombination. If two identical copies of a gene or a mobile element (grey and flat lines) are present on a chromosome (black wavy lines), they can recombine together (shown as cross) behind a replication fork to produce one chromosome with a deletion of the intervening region, and the other with a duplicated region.

is in the merodiploid state (Anderson and Roth, 1977; Petes and Hill, 1988). Such rearrangements offer a chance for the selection of new chromosomal variants and evolution.

The DNA intermediates

It is now generally admitted that a free DNA end, either a double-stranded extremity or a single-stranded gap, is the prerequisite to initiate homologous recombination. Two models based on this assumption and adapted from the review of Kuzminov (1999) are presented in Fig. 1.3. One starts with a double-stranded extremity (Fig. 1.3A), and the other starts with the single-stranded gap (Fig. 1.3B). In both cases, the initial event and the final event are the

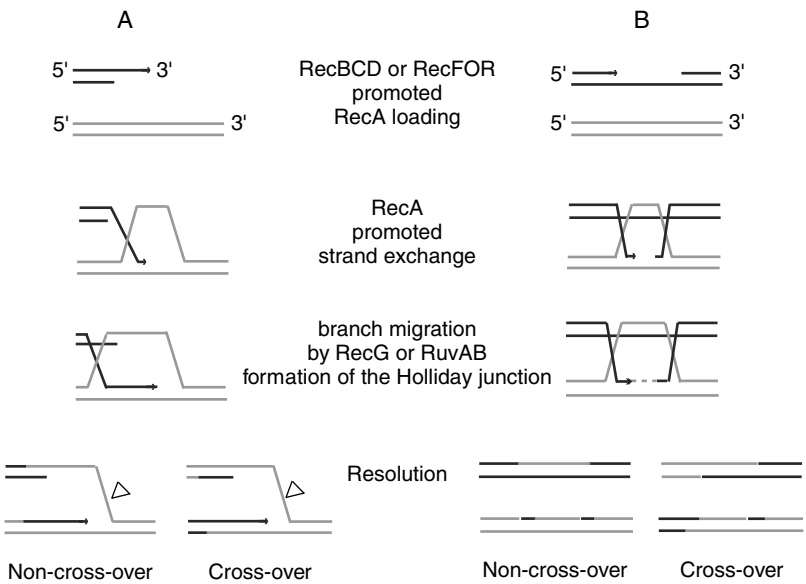

Figure 1.3. Two models showing possible intermediates of the recombination process, as adapted from Kuzminov (1999). Recombination starts from a double-stranded break (column A) or a single-stranded gap (column B) blue molecule, which is processed by RecBCD (left) or RecFOR (right) to load RecA and promote strand pairing with an intact grey molecule (step 2). Each strand of the DNA duplex is drawn, and 5′ and 3′ extremities are indicated on the first lanes. The black arrow shows the 3′ extremity of an invading molecule. Due to the action of the RecG and/or RuvAB helicase, Holliday junctions are created (step 3) and resolved by RuvC or another nuclease into crossed-over or non–crossed-over products (step 4). Dotted lane indicates DNA synthesis, and the white triangle stands for a putative nuclease that would cleave the D loop.

most documented ones, whereas the steps in between are speculative – in particular, the short patch of DNA synthesis by unknown polymerase and the cleavages by unknown endonuclease. Initial to all homologous recombination is the invasion of a single-stranded DNA into a duplex molecule so it pairs and forms the so-called "heteroduplex." Next – and less well characterized – this intermediate is converted into a Holliday junction (HJ), in which the second strand of the invading molecule has also paired with its recipient. This four-stranded structure is able to move around (branch migration) and then to be processed by a specific nuclease into either of two products, depending on the orientation of the cleavage. One cleavage will result in product molecules having received a small patch of the donor DNA, the heteroduplex region; it is called *gene conversion*. The other cleavage orientation will result in the exchange of the flanking sequences, or "crossing over," between the two recombining molecules.

The third model (Fig. 1.4) accounts for the repair of blocked replication fork (Seigneur et al., 1998). Recombination genes are not essential in bacteria (bacteria deleted for *rec* genes are viable), but they can become essential for the viability of replication mutants. In such mutants, replication fork progression is hindered, and a process called *replication fork reversal* is supposed to take place, in which the DNA intermediate is structurally identical to the HJ (Fig. 1.4B). This intermediate is processed by recombination enzymes and rescues the replication fork by removing the replication block either by recombination (Fig. 1.4E) or by trimming the extremity of the new strands (Fig. 1.4D) to allow restart. If the recombination enzyme RecBCD (see Chapter 2) does not process this intermediate, it is subjected to a cleavage by RuvABC (see Chapter 2), which may be lethal (Fig. 1.4C). The little revolution brought about by this model is that recombination enzymes reveal themselves as being closely interconnected with the replication process and more generally involved in the normal life cycle of a bacterium, rather than being specialized in some aspects of lesion repairs, or even more specialized processes, such as conjugation or transformation.

GENETICS AND BIOCHEMISTRY OF HOMOLOGOUS RECOMBINATION IN *E. COLI*

A wealth of genes and proteins play a role in homologous recombination processes in *E. coli*. They are briefly listed, and refer to the recombination models described in Fig. 1.3. Three distinct and successive steps are involved in homologous recombination: presynapsis, during which enzymes prepare the DNA substrate for RecA; synapsis, where RecA bound to DNA promotes

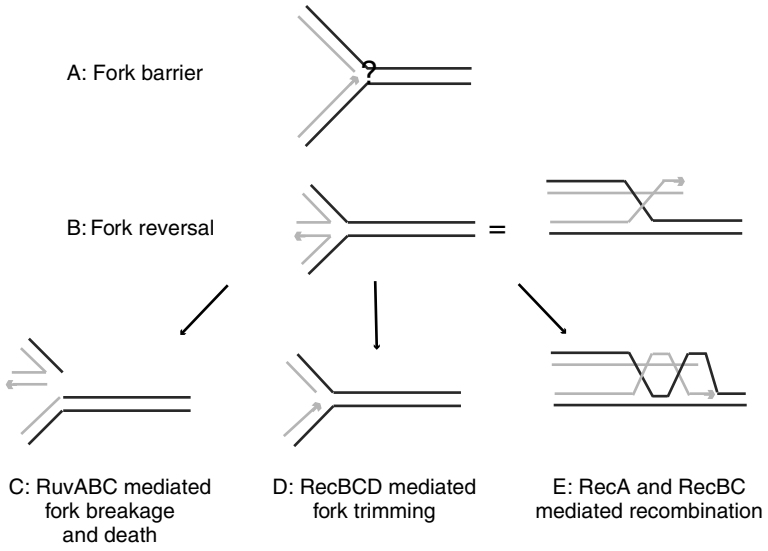

Figure 1.4. A model for the repair of a blocked replication fork by recombination enzymes, adapted from Seigneur et al. (1998). A replication fork is drawn with newly replicated strands in a light grey color, the arrow showing its 3′ extremity. The question mark stands for the block, which leads to a fork reversal, where the two replicated strands have paired (step B). This intermediate can be drawn such that the newly paired region faces its unreplicated homolog, ahead of the reversed fork (step B, right part). Three possible fates for this intermediate have been proposed: breakage if RuvABC arrives on the DNA before RecBCD (step C), fork trimming if RecBCD acts on the exposed double-strand end (step D), and recombination if RecBCD and RecA act in concert (step E). For simplicity the reversed fork is shown with blunt extremities, but a normal fork would tend to produce a 3′ protruding end because of the advance of the leading strand relative to the lagging strand. Such an extremity may be processed by exonucleases to produce a blunt end or be used directly by RecA in step E.

the active search for a homologous molecule; and postsynapsis, where intermediates are processed into products.

Presynapsis

Two sets of proteins are needed in *E. coli* for the processing of DNA ends and the efficient loading of RecA:

- The RecBCD heterotrimer is a dual enzyme that is very well characterized biochemically (Kowalczykowski et al., 1994; Myers and Stahl, 1994) and structurally (Singleton et al., 2004). It is a potent

exonuclease (ExoV) strictly dependent on blunt or nearly blunt ends for its activity, as well as a highly processive helicase. The RecB subunit encodes the nuclease activity, which degrades preferentially in the 3′ to 5′ direction, and the RecD subunit modulates RecB: Upon interaction with a specific DNA sequence, the Chi site, RecD converts the polarity of DNA degradation by RecB, such that a 3′ single-stranded extremity is created, and this promotes the loading of RecA (Anderson and Kowalczykowski, 1997). In addition, both RecB and RecD act as DNA helicases, of opposite polarity and different speed (Dillingham et al., 2003; Taylor and Smith, 2003). RecC is inert enzymatically, it serves as a structural component allowing the physical separation of the two DNA strands, and it is proposed to contain the Chi recognition site (Singleton et al., 2004). How exactly the RecBCD complex recognizes the Chi site is not yet understood. *recB* and *recC* mutants are sensitive to UV and gamma irradiation, and deficient for homologous recombination when the DNA substrate has a double-stranded blunt end, such as during generalized transduction and conjugation (Fig. 1.3A). Interestingly, *recB* or *recC* mutations profoundly affect cell growth, with up to 80% of dead cells in a liquid culture (Capaldo, Ramsey, and Barbour, 1974), and tend to yield suppressor mutations. Such is not the case for *recA* mutants and points to an additional role of the RecBCD complex in the cell, besides recombination, probably its exonucleolytic role, removing useless – or potentially dangerous – linear DNA. The strict dependence of the RecBCD complex for double-stranded ends makes this complex the first actor to act on linear DNA (Fig. 1.3, left side), and one of the key enzymes for the rescue of arrested replication forks (Figs. 1.4D and 1.4E) and for preventing RuvABC mediated fork breakage (Fig. 1.4C).

- The second group of proteins comprises RecF, RecO, and RecR. Although less well characterized at the molecular level, RecF, RecO, and RecR play a key role in preparing substrates for RecA on gapped DNA (Fig. 1.3, right side), to which RecBCD has no access (Morimatsu and Kowalczykowski, 2003). Their role would consist of competing away the single-stranded binding protein SSB to favor RecA loading. In addition to this role, RecFOR may also process double-strand breaks in cells in which the RecBCD complex is absent (Amundsen and Smith, 2003). For this purpose, at least two additional functions are recruited, the RecJ 5′ to 3′ nuclease, and the RecQ helicase (or, in its absence, UvrD or Helicase IV). These accessory functions would serve to trim the blunt extremity into a 3′ single strand for RecA. More recently, it was shown that RecFOR

and RecJ could collaborate with RecBCD when the RecB component has lost its nuclease activity (Ivancic-Bace et al., 2003). *recF*, *recO*, and *recR* mutants are less sensitive than *recB* or *recC* mutants to radiation, but double mutants *recB recF* are nearly as sensitive as the *recA* mutant strain. *recF*, *recO*, or *recR* mutants do not exhibit any particular growth defects. This may reflect the low frequency of gap formation, the exclusive RecFOR substrate, in *E. coli*.

Synapsis

The RecA protein forms a stable filament on single-stranded DNA (ss DNA), which extends in the 5′ to 3′ direction (Roca and Cox, 1997). It promotes ss DNA pairing with a homologous double-stranded DNA, which is the key step of homologous recombination. Once the two DNA molecules have been placed in register, strand exchange can start. Homologous recombination is completely abolished in *recA* mutants, except in two particular cases mentioned in the section "RecA-independent homologous recombination." *recA* mutants are highly sensitive to UV and gamma radiation, and also affected for their growth, with a reduced doubling time and 50% of dead cells in liquid culture (Capaldo, Ramsey, and Barbour, 1974).

The minimal length of DNA on which RecA binds *in vitro* is 8nt, and the smaller stable duplex made by RecA is 15 base pair (bp) long (Hsieh et al., 1992). *In vivo*, the minimal length of homology on which RecA can act is probably as small as 23 to 27 bp, but data vary according to the system used (Lloyd and Low, 1996). RecA also promotes strand exchange between sequences that are not identical *in vitro*, and *in vivo*, an editing process mediated by MutS and MutL aborts such intermediates by a mechanism that remains to be elucidated. As a consequence, genetic exchanges between closely related species with nearly identical DNA sequence (called homeologous sequence, diverged up to 15% or even more) are highly increased when the recipient strain is mutated for the *mutS* or *mutL* function (Rayssiguier, Thaler, and Radman, 1989; Vulic et al., 1997). MutS and MutL were also found to edit intrachromosomal rearrangements between two slightly diverged *rhs* sequences (Petit et al., 1991). A more recent study on a wide spectrum of natural *E. coli* isolates has revealed that such mutant strains are present in 3% to 5% of isolates (Denamur et al., 2002), and they should favor horizontal gene transfer between related species.

In addition to its role at the heart of recombination, RecA of *E. coli* is also responsible for inducing the SOS response (see "The SOS response"),

and for promoting the first step of replication fork reversal in some cases (McGlynn and Lloyd, 2000; Robu, Inman, and Cox, 2001; Seigneur, Ehrlich, and Michel, 2000).

Postsynapsis

Downstream of the RecA-promoted strand exchange, two alternative enzymes process the intermediates, RuvABC and RecG (Sharples, Ingleston, and Lloyd, 1999). RuvA is a DNA-binding protein specific for HJs, and RuvB is a DNA helicase that catalyzes branch migration when bound to RuvA. This helicase promotes the branch migration of HJ and delivers them to the specific nuclease RuvC, which resolves the recombination intermediate by cleaving symmetrically across the junction. Depending on which strands are cleaved, different products are expected, as drawn on Figs. 1.3A and 1.3B. RecG is also a DNA helicase, which favors branch migration of HJ and three-stranded branched structures (Lloyd and Sharples, 1993). Whether a nuclease is also involved to cleave the junctions processed by RecG is not known at present. What prompted the conclusion that *ruvABC* and *recG* encode redundant functions was the genetic observation that single mutants were only partially affected for recombination and partially sensitive to UV radiation, whereas the double mutant was as deficient and as sensitive as a *recA* strain, and affected for viability (Lloyd, 1991).

The outcome of recombination events appears to differ markedly with each situation, the RuvABC complex favoring the cross-over products when recombination is initiated from a double-stranded break, and the non–cross-over products when recombination is initiated from a gap (Cromie and Leach, 2000). Concerning RecG, one study suggests that it favors the cross-overs when recombination is initiated from a gap (Michel et al., 2000).

During the replication fork repair process (Fig. 1.4), a toxic role of RuvABC has been revealed: It recognizes the putative regressed fork intermediate, which has an HJ structure, and cleaves it, which leads to a linear chromosome. This aberrant role is countered by RecBCD, which either degrades the tail or initiates recombination with RecA (Seigneur et al., 1998).

Involvement of DNA replication

Some recombination intermediates (e.g., the one drawn in Fig. 1.3, last step, left side) are converted into replication forks due to the action of a group of seven proteins called collectively the PriA-dependant primosome, composed of PriA, PriB, PriC, DnaT, DnaC, DnaB, and DnaG (Marians, 1999,