# Homologous Genetic Recombination during Bacterial Conjugation

**Susan M Rosenberg,** *Baylor College of Medicine, Houston, Texas, USA*
**Mohammad R Motamedi,** *University of Alberta, Edmonton, Canada*

Secondary article

Article Contents



Conjugation is the sexual transfer of deoxyribonucleic acid (DNA) from one bacterium directly into another. The transferred DNA can replace DNA of similar sequence in the recipient cell's chromosome by homologous genetic recombination, a process that both exchanges lengths of similar DNA sequences and repairs broken chromosomes, using similar chromosomes as templates.

## Introduction

Homologous genetic recombination is the exchange of stretches of deoxyribonucleic acid (DNA) sequence between DNA molecules with similar sequences, creating new linked combinations of alleles of genes. The exchange is precise, occurring in register, neither adding nor deleting nucleotides at the borders. Exchange is also *general*, in that it can occur anywhere along the lengths of two similar or *homologous* DNA molecules. Recombination is an important force in evolution because it generates new cellular and organismal combinations of alleles on which natural selection acts, and may even be a greater force in the generation of genetic diversity in evolution than mutation, which creates the new alleles. Recombination is also required for proper segregation of chromosomes in meiosis in eukaryotes, for repair of damaged DNA, for promoting some mutations in bacteria, yeast and vertebrates (Rosenberg, 1997; Harris *et al*., 1999) and is intimately connected with DNA replication (discussed below). Because recombination is a major controller of genetic and genomic stability, it is also an important force in the generation of human cancers. Humans who carry mutant alleles of genes encoding *homologues* of recombination proteins found in the bacterium *Escherichia coli* (i.e. proteins of similar function and sequence presumed to be derived from common ancestors) display altered DNA recombination and are genetically predisposed to cancer (Ellis, 1997). Recombination is thus also crucial to the daily lives of cells and organisms such that mutants that fail to recombine DNA, or do too much recombination, usually have grossly abnormal genetic stability, growth and viability. (See Ellis (1997), Rosenberg (1997) and Harris *et al*. (1999) for discussions of recombination, genetic instability, mutation and cancer.)

Bacteria, particularly *E. coli*, have long been used as highly tractable model systems for the study of recombination. At present, much is understood about how recombination occurs, although the precise details of the mechanism(s) have not yet been solved. The mounting discoveries of eukaryotic homologues of the bacterial recombination proteins have made it obvious that the mechanisms learned from bacteria provide important insights into this fundamental process in all organisms.

### Bacterial sex

Recombination is also important in the lives of bacteria. This entry focuses on recombination during *conjugation*, a form of bacterial sex (see Firth *et al*., 1996) understood most thoroughly from studies of *E. coli* and its relatives. Many other modes of gene transfer occur in bacteria. The ability of bacterial DNAs to recombine depends on how similar the DNAs are, and can be used as one definition of species. More distantly related bacteria are unable to recombine due to their more divergent DNA sequences, and thus may be thought of as separate species. However gene transfer between apparent species (both *horizontal* (nonsexual) and *vertical* (sexual) *gene transfer*) is so prevalent as to make the species concept difficult. Control of recombination at the level of sequence similarity and interspecies gene transfer are discussed later in this article.

Conjugation is possible only in bacteria that harbour a special DNA element called a *sex plasmid*. These are small, autonomously replicating, usually circular DNA molecules carrying genes that encode an array of functions required for the plasmids' maintenance and transfer into other cells. It has been speculated that such plasmids are parasitic DNA elements that use bacteria to transport themselves to other bacteria and even to eukaryotic cells. The functions they encode include *sex pili* (singular: *pilus*), long, tube-like structures on the cell surface that let the plasmid-bearing (*male* or *donor*) cell attach to cells without the plasmid (*female* or *recipient* cells). A special sexual or

*transfer origin* of replication is present on these plasmids, as well as genes encoding the *transfer proteins* required to move the sexually replicated plasmid DNA into a recipient cell. The DNA enters the recipient as a linear single strand and is copied to become a linear duplex molecule, which can then recircularize to produce another sex plasmid, turning the recipient into a male (**Figure 1a**).

The roughly 100-kb *F* plasmid is the sex plasmid used for most studies of conjugational recombination. F exists in three forms in *E. coli*: $F^+$ is the free-living plasmid separate from the chromosome. *F′* (F-prime) plasmids are also extrachromosomal, but carry on them a region of chromosomal DNA that they have acquired from bacteria (such that cells with an F′ can be *merodiploid*: having a chromosomal segment or gene represented more than once in the genome (here, once in the normal chromosomal location and once on the F′)). F can integrate into the bacterial chromosome creating *Hfr* (high frequency recombination) cells. Any of these forms can be transferred into recipient cells. Only $F^+$ and F′ can recircularize independently of homologous recombination proteins in the recipient (**Figure 1a**). Hfr DNA will be lost after transfer unless the linear double-stranded molecule recombines with the recipient's circular bacterial chromosome, replacing a length of similar DNA sequence in the recipient (**Figure 1b**). For the transfer of bacterial (i.e. not plasmid)

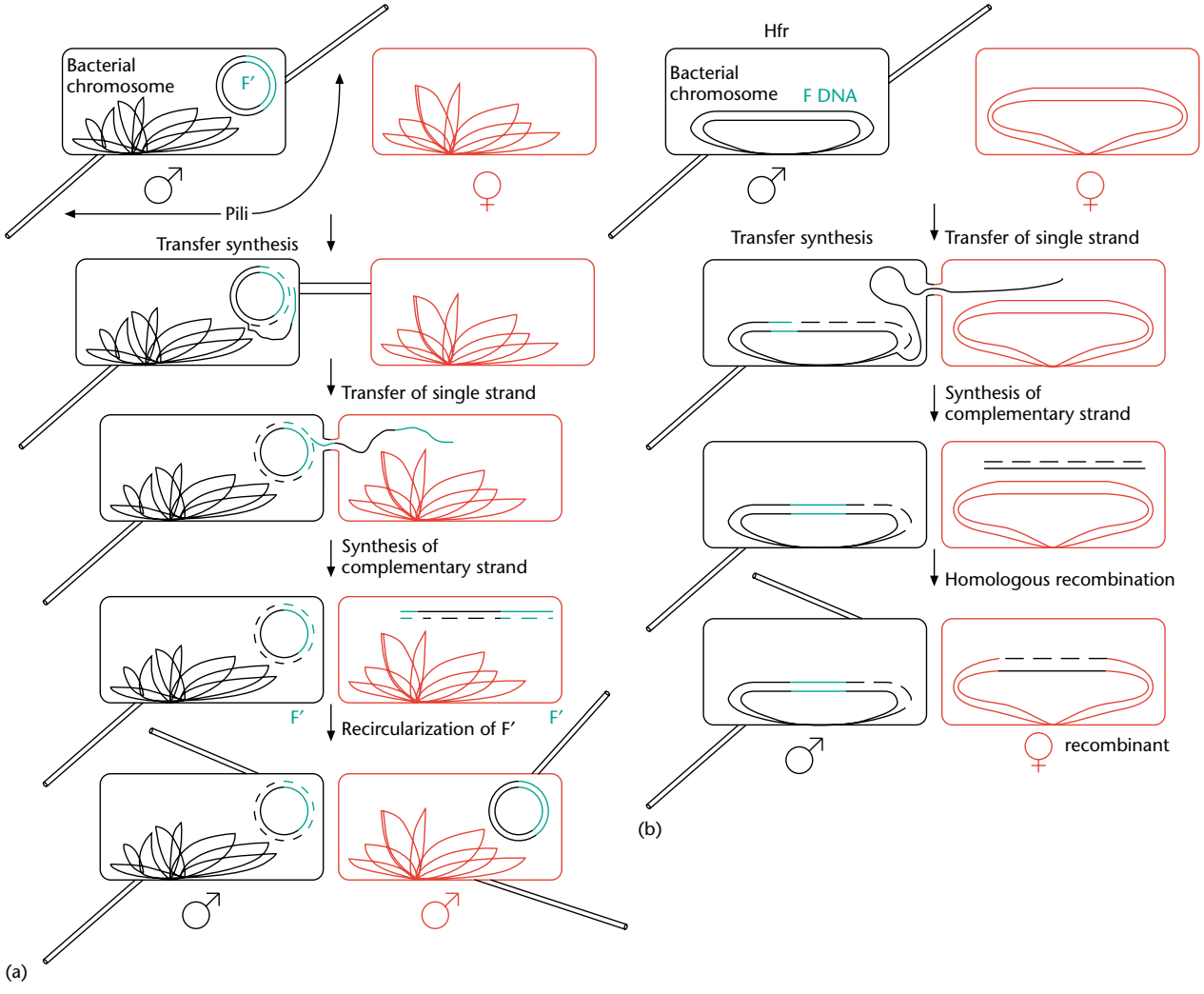



**Figure 1** Conjugal transfer of DNA mediated by bacterial sex plasmids. (a) Transfer of extrachromosomal sex plasmid DNA such as $F^+$ or F′ DNA. An F′ is illustrated with green F DNA and black DNA from the bacterial chromosome. Dashed lines represent newly synthesized DNA. Each line of the plasmid DNA represents a single DNA strand (polynucleotide chain). (b) Transfer of bacterial DNA of an Hfr strain, followed by homologous recombination of the transferred DNA into the recipient cell's chromosome. The usual folded and packaged (*nucleoid*) structure of the bacterial chromosome shown in (a) is not drawn in (b) for clarity in illustrating the recombination of Hfr DNA with the recipient chromosome (b). No real difference in chromosome structures is implied.

DNA sequences into recipient bacteria, the F′ and Hfr forms are most important. This text focuses on the mechanism(s) by which linear DNA, including DNA transferred by Hfrs, is recombined with homologous DNA in bacteria. Because some other examples of recombination of linear DNA have been central to illuminating recombination mechanism(s), these are also reviewed here as model systems for understanding both conjugal recombination and double-strand break-repair (discussed below).

## Experimental Approaches

### Hfr and F′ crosses in *Escherichia coli*

A. J. Clark (Berkeley, USA) and his colleagues isolated many different mutant derivatives of *E. coli* that were unable to perform homologous recombination during conjugation. These were identified as mutants that could not incorporate transferred Hfr DNA stably, a process that requires recombination with the recipient chromosome (**Figure 1b**), but could receive transferred DNA that did not require recombination to function. These mutants are recombination-deficient, and were named *rec*. *rec* mutants define a large number of genes whose protein products participate in the reactions of homologous recombination. **Table 1** is a list of recombination genes and proteins in *E. coli,* including several that were found to participate in recombination after their identification in other processes (and so are not named '*rec*'). The first discovered and most famous is RecA without which there is essentially no homologous recombination in *E. coli*. RecA is conserved throughout all organisms examined so far, having homologues that are important for recombination and DNA repair in humans, other eukaryotes, and the Archaea.

#### RecBCD enzyme and RecA protein

An important feature of recombination that occurs during bacterial conjugation is that one of the recombining DNA molecules is linear, in contrast with the circular nonconjugating (vegetative) bacterial chromosome. In *E. coli*, the proteins used for recombination of linear DNA are different from those used to recombine circular DNA. The most important protein for recombination of linear DNA is the RecBCD enzyme.

RecBCD is a heteromultimeric enzyme with subunits encoded by the *recB*, *recC* and *recD* genes. The RecB and RecC subunits are essential for all activities of this enzyme. RecD appears to be a regulatory subunit that allows the enzyme to switch back and forth between two states. With the RecD subunit plugged in, RecBCD is an *exonuclease*, an enzyme that breaks the phosphodiester backbone of DNA, destroying the DNA from a DNA end. RecBCD is a potent double-strand-dependent exonuclease. When RecD is not making effective contact with the RecBC part of the enzyme, RecBC promotes DNA recombination, probably by virtue of its DNA unwinding or *helicase* activity.

RecBCD loads on to DNA only at double-strand ends. Such ends bracket the linear DNA after conjugal transfer and synthesis of the complementary strand (**Figure 1b**), and also result from DNA damage (double-strand breaks in DNA). RecBCD uses the double-strand ends of linear transferred DNA as entry points, and, by way of reactions discussed below, exposes stretches of single-stranded DNA at those ends, as a first step in the recombination process. The single-strand DNA is a substrate for the RecA protein, which catalyses *synapsis* of the single-strand ends to an intact (unbroken) DNA molecule as part of the recombination process (discussed below).

RecBCD and RecA are thus critical proteins for recombination of the transferred linear duplex (double-stranded) DNA with the recipient cell's intact circular chromosome. These proteins, and the homologous recombination process that they catalyse, are also the cell's primary means of repairing chromosomal double-strand DNA breaks that occur during normal growth in vegetative cells. For double-strand break-repair, the molecule with which the broken chromosome recombines is a *sister chromosome*, a product of DNA replication. Growing bacterial cells often have several sister chromosomes present per cell. These allow broken chromosomes to be reconstructed by homologous recombination. Sexual transfer of DNA probably benefits bacteria in the long term, allowing them to acquire useful gene combinations and new genes. However, the ongoing need to repair broken DNA is probably the job for which recombination has evolved. This need is illustrated by the fact that strains with defective *recA* or *recBC* genes are exceedingly slow-growing and apparently sickly. Homologous recombination also serves both the purposes of facilitating sexual DNA information exchange and repair of DNA damage in eukaryotes.

#### Other recombination proteins identified from studies of *Escherichia coli* mutants defective in conjugal recombination

A long list of *rec* genes and their corresponding proteins were identified in studies of conjugal recombination (**Table 1**; Clark and Sandler, 1994). Interestingly, loss-of-function mutations of many of these do not cause dramatic loss of recombination proficiency, or sickly phenotype in otherwise nonmutant cells. This is true of *rec* genes *F*, *G*, *J*, *N*, *O*, *Q*, *R* and other important recombination genes not named *rec*, such as *ruvA*, *ruvB* and *ruvC*. These *rec* mutants were not found as described above; rather, many of these were isolated as mutants that conveyed conjugal recombination deficiency in cells already lacking other recombination enzymes (Clark and Sandler, 1994). These encode very interesting proteins, not all of whose functions are under-

**Table 1** *Escherichia coli* recombination genes and proteins[a]

| Gene name[b] | Protein encoded | Complex that protein is part of | Relevant biochemical functions | Recombination phenotype of cells lacking this protein[c] | Eukaryotic homologues | Stage of recombination believed to act during | Role in recombination in *E. coli* |
|---|---|---|---|---|---|---|---|
| *recB* | RecB | RecBCD enzyme (essential subunit) | RecBCD is a double-strand exonuclease and helicase | Rec$^-$ | | Presynapsis | 1. Creates single-strand DNA used by RecA for strand invasion and exchange<br>2. Degrades linear duplex DNA |
| *recC* | RecC | RecBCD enzyme (essential subunit) | RecBCD is a double-strand exonuclease and helicase | Rec$^-$ | Some similarity to Nucl[d] endo/exonuclease | Presynapsis | 1. Creates single-strand DNA used by RecA for strand invasion and exchange<br>2. Degrades linear duplex DNA |
| *recD* | RecD | RecBCD enzyme (regulatory subunit) | Subunit required for nuclease activity of RecBCD, not required for helicase or recombination | Hyper-Rec | | Presynapsis | Required for modulation of the RecBCD enzyme at Chi sites at which RecBCD switches out of exonuclease mode, and retains helicase activity creating single-strand DNA ends used for strand-invasion after coating with RecA |
| *recJ* | RecJ | | 5′-single-strand-dependent exonuclease | HypoRec when ExoI, ExoVII absent | | Presynapsis?/ unknown | Unknown; suggested to help preserve single-strand DNA created by RecBCD so that it can be used by RecA in strand invasion |
| *xonA* or *sbcB* | ExoI | | 3′-single-strand-dependent exonuclease | HypoRec when RecJ, ExoVII absent | | Presynapsis/ unknown | Unknown; suggested to help preserve single-strand DNA created by RecBCD so that it can be used by RecA in strand invasion |
| *xseA* | ExoVII | ExoVII subunits encoded by *xseA* and *xseB* | Single-strand-dependent exonuclease on 5′- and 3′-ends | HypoRec when Exol, RecJ absent | | Presynapsis/ unknown | Unknown; suggested to help preserve single-strand DNA created by RecBCD so that it can be used by RecA in strand invasion |
| *recQ* | RecQ | | Helicase | None | *BLM*[d], *WRN*[d], *SGS1*, *Rqh1*, *Q1/RecQL*[d] *UTAS*[a] | Presynapsis?/ unknown | Unknown; suggested to help create single-strand DNA used by RecA in strand invasion |
| *ssb* | Single-strand binding protein (SSB) | | Coats single-strand DNA preventing formation of intra-molecular base pairs (secondary structure) | Unknown | (Many eukaryotic SSBs exist and may have similar functions but have not evolved from a common ancestor) | Presynapsis/ synapsis | May aid RecA-mediated strand-exchange reactions.<br>SSB is thought to remove secondary structure from single-strand DNA prior to coating of that DNA with RecA |

*continued*

**Table 1** – *continued*

| Gene name[b] | Protein encoded | Complex that protein is part of | Relevant biochemical functions | Recombination phenotype of cells lacking this protein[c] | Eukaryotic homologues | Stage of recombination believed to act during | Role in recombination in *E. coli* |
|---|---|---|---|---|---|---|---|
| *recF* | RecF | RecFOR may sometimes act as a complex | | Slight HypoRec | | Presynapsis/ synapsis | Unknown; hypothesized may help RecA displace SSB from single-strand DNA; hypothesized may allow 5′-DNA end invasions |
| *recO* | RecO | RecFOR may sometimes act as a complex | | None | | Presynapsis/ synapsis | Unknown; hypothesized may help RecA displace SSB from single-strand DNA; hypothesized may allow 5′-DNA end invasions |
| *recR* | RecR | RecFOR may sometimes act as a complex | | None | | Presynapsis/ synapsis | Unknown; hypothesized may help RecA displace SSB from single-strand DNA; hypothesized may allow 5′-DNA end invasions |
| *recA* | RecA | | 1. Strand exchange;<br>2. Coprotease | $Rec^-$ | *RAD51*[d], *DMC1*[d], *RAD55*, *RAD57*, *RAD51B/hREC2*[d], *RAD51C*[d], *XRCC2*[d], *XRCC3*[d], *rhp51*, *DLH11*, *uvsC* | Synapsis | 1. Coats single-stranded DNA and catalyses its invasion of duplex DNA, searches for sequence complementarity, after which base pairing occurs to form a bimolecular, heteroduplex strand-exchange intermediate<br>2. Sensor molecule for activation of the SOS DNA damage repair response in which RecA acts as a coprotease, posttranslationally regulating the expression of many other genes |
| *priA* | PriA | Replisome | Primasome assembly in DNA replication | Mild HypoRec | | Postsynapsis?/ unknown | Unknown; may help establish DNA replication at sites of strand exchange |
| *ruvA* | RuvA | RuvAB | Strand-exchange junction-specific binding helps load RuvB DNA helicase | Slight HypoRec; $Rec^-$ when cells $RecG^-$ | | Postsynapsis | Unknown; hypothesized branch migration that extends heteroduplex joints of one polarity but disrupts heteroduplex joints of the opposite polarity; may be partially redundant with RecG |
| *ruvB* | RuvB | RuvAB | Strand-exchange junction-specific DNA helicase | Slight HypoRec; $Rec^-$ when cells $RecG^-$ | | Postsynapsis | Unknown; hypothesized branch migration that extends heteroduplex joints of one polarity but disrupts heteroduplex joints of the opposite polarity; may be partially redundant with RecG |

*continued*

**Table 1** – *continued*

| Gene name[b] | Protein encoded | Complex that protein is part of | Relevant biochemical functions | Recombination phenotype of cells lacking this protein[c] | Eukaryotic homologues | Stage of recombination believed to act during | Role in recombination in *E. coli* |
|---|---|---|---|---|---|---|---|
| *ruvC* | RuvC | RuvABC | Holliday junction-specific endonuclease | Slight HypoRec; Rec$^-$ when cells RecG$^-$ | | Postsynapsis/ resolution | Unknown; hypothesized cleavage of strand-exchange junctions; may be partially redundant with RecG |
| *recG* | RecG | | Strand-exchange junction-specific helicase | Slight HypoRec; Rec$^-$ when cells RuvA$^-$, RuvB$^-$ or RuvC$^-$ | | Postsynapsis | Unknown; hypothesized branch migration that extends heteroduplex joints of one polarity but disrupts heteroduplex joints of the opposite polarity; hypothesized partially redundant with Ruv system |
| *recN* | RecN | | | None | | Unknown | Unknown |

[a] Other proteins that perform housekeeping roles in DNA metabolism probably also participate in recombination. These include DNA ligase, some to all of the four known topoisomerases of *E. coli*, and others.

[b] The genes are listed in the order that their protein products are thought to act in recombination (see column 7).

[c] Some of the recombination or *rec* genes that appear not to affect recombination phenotype much (e.g. *recF*, *recO*, *recR*, *recQ*, *recN*) may either be redundant with other functions that substitute when that gene is defective or may play more important roles in recombination of DNA substrates other than the double-strand linear DNA that is processed during conjugational recombination. For example, their gene products may be more important for recombination of the circular bacterial chromosome and/or for DNA repair at single-strand rather than double-strand breaks. *See* Figure 3 for a possible model for how these proteins might help promote recombination (also Clark and Sandler, 1994).

[d] These homologues known from mammals (including humans).

stood. The implication that their losses of function cause recombination deficiency only when other recombination proteins are also absent is that these proteins provide functions that are in some way redundant. This need not mean that the apparently substituting proteins do exactly the same jobs, although that is one possibility. It could be that some parts of the recombination process occur *via* one or another set of biochemical steps or *pathway*, and that such alternative pathways use different proteins. In nonmutant cells, more than one pathway to recombination may exist, all of them using RecA and RecBCD (when the DNA recombining is linear), but using different sets of proteins for other stages of the reactions.

Studies of conjugal recombination identified many recombination proteins whose subsequent biochemical characterizations have been central to understanding how linear DNA recombines in *E. coli*. Because linear DNA is the most recombinationally active form in all organisms tested, these reactions define important paradigms for all organisms.

## Phage-mediated transductional crosses

The viruses of bacteria, or *phages*, have been invaluable tools for the study of nearly all aspects of molecular biology, including the mechanisms of recombination. Phages also provide bacteria with a second avenue for exchange of their DNA. A few phages, e.g. *E. coli phage P1* and *Salmonella typhimurium phage P22*, are capable of *generalized transduction*, the transfer of small, random pieces of bacterial chromosomal DNA to another bacterium inside of a phage *capsid* (protein coat). Infrequently during a phage infection, bacterial DNA is packaged, instead of a phage chromosome, into a few of the capsids (about $10^{-2}$ of all capsids). When the infection culminates with lysis of the host cell and release of the phage particles, the *transducing particle* (a phage capsid containing bacterial DNA) can attach to a new bacterium, injecting the piece of bacterial DNA. Usually, this DNA is linear, not autonomously replicating, and will be lost unless it recombines with similar sequences in the *recipient* bacterium. The recombination of these linear DNA fragments with the circular chromosome of the recipient appears in nearly all respects to be similar to conjugal recombination. For example, transductional recombination requires RecA, RecBCD and nearly every protein used in conjugal recombination (Masters, 1996).

Generalized transduction allows bacterial geneticists to construct new strains of bacteria in which only a small region is replaced with DNA from another strain. This allows reliable creation of sets of strains that are genetically identical (*isogenic*) except for a single variable: the replaced DNA segment carrying a different allele of some gene of interest. The transduction is *generalized* because almost any region of bacterial DNA can be transduced. P1 packages and transduces about 100 kb of the roughly 4500-kb *E. coli* host genome per capsid. Phage P22 transduces segments of about 45 kb.

## Contributions of the study of phage DNA recombination to understanding bacterial recombination mechanisms

Because phage provide smaller, physically manipulable model DNA substrates, phage work and phage workers have been central to the definition of DNA intermediate structures, important regulatory sequences and molecular mechanisms of recombination in bacteria.

### Heteroduplex DNA underlies the precision of homologous recombination

Phage worker Matthew Meselson (Cambridge, USA, then in Pasadena, USA) realized that a powerful means to ensure the remarkable precision of homologous recombination would be by basepairing a strand of one DNA molecule with a complementary strand of another molecule with which it recombines. Such regions of intermolecular basepairing, called *heteroduplex* DNA, were suggested by early studies (1950s) of phage. After recombination between phages of two different genotypes (carrying different alleles of one or more genes), some single phage, which carry single DNA molecules in their capsids, were observed to give rise to progeny carrying both parental alleles of a single gene. This suggested that the sequences specifying each allele might be present on complementary basepaired strands, i.e. heteroduplex DNA.

Heteroduplex DNA is now known to be the central DNA intermediate in homologous recombination in all organisms. The existence of heteroduplex DNA as the key joint in recombination was also postulated by fungal recombination students H. L. K. Whitehouse and P. J. Hastings (1963, Cambridge, UK) and independently by Robin Holliday (1962, 1964, Mill Hill, UK). In *E. coli*, the synapsis events begun by RecBCD and RecA lead to the formation of *heteroduplex DNA junctions* (also called Holliday junctions) between the two recombining molecules, ensuring precise, high-fidelity exchange between sequences that are identical, or very nearly so, within the joint region.

### Do parental DNA molecules contribute their own material, or only their sequence information to their recombinant progeny?

Meselson and phage worker Jean Weigle (Pasadena, USA) articulated three general mechanisms that could, in principle, underlie recombination (**Figure 2**). The DNA molecules recombining could be broken and rejoined one to another, creating a hybrid molecule composed of material from each parental DNA molecule (*break-join*

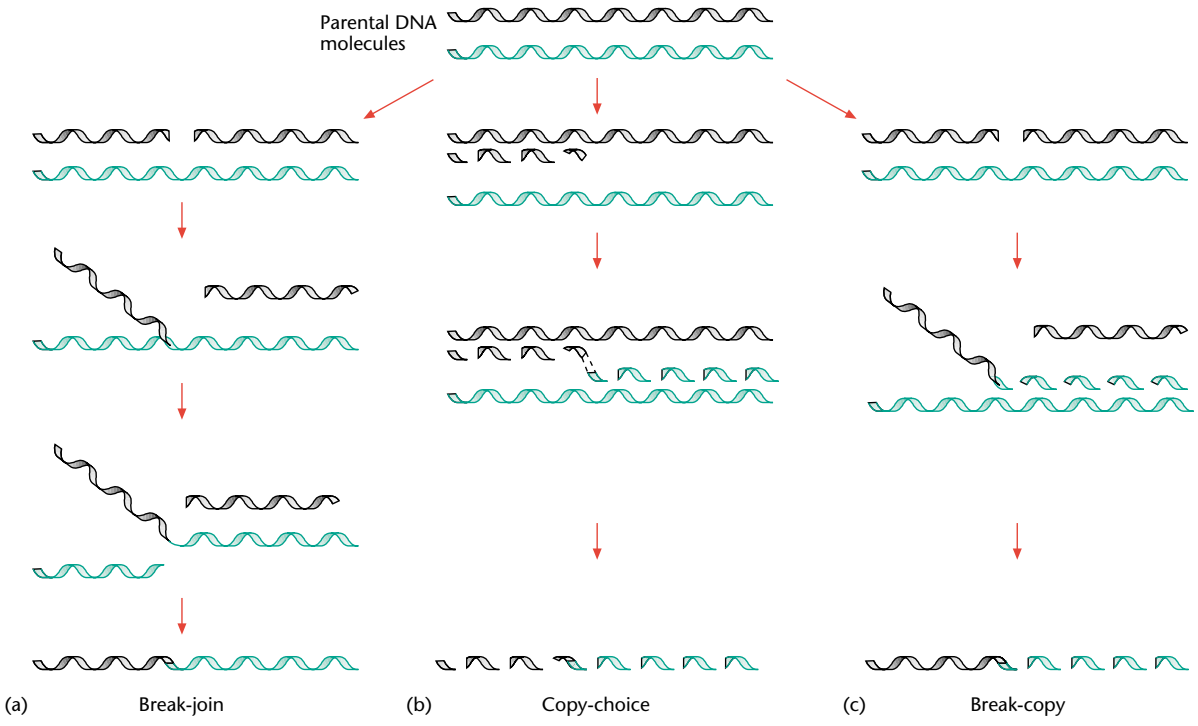



**Figure 2** General hypotheses for possible mechanisms of DNA recombination, as formulated by Meselson and Weigle. In this figure only, double-stranded DNA is represented by helices. Newly synthesized DNA is symbolized by broken helices. (a) Break-join models create recombined DNA from the parental DNA material. (b) Copy-choice models form recombinant DNA without any material contribution of parental DNA, by switching the template for DNA replication from one molecule to another. To have no material contribution of parental DNA, the replication would have to be *conservative* (not the standard *semiconservative* segregation of DNA strands). (c) Break-copy models use part of one parental DNA molecule to initiate DNA replication using another molecule as a template, joining the two by heteroduplex DNA.

recombination, **Figure 2a**). Alternatively, recombinants could carry the information from each parental molecule but carry no material contribution from the parental DNA. This could occur if recombination were a replication process in which part of one parental molecule was copied, then the template switched to copy part of the other parental molecule (*copy-choice* recombination, **Figure 2b**). A third possibility combining these two models is that one broken molecule (for example, the linear DNA transferred in conjugation) might synapse with the other parental molecule and then initiate DNA replication to copy the information from the unbroken parent (*break-copy* recombination, later demonstrated by Gisela Mosig (Nashville, USA) and colleagues to occur in *E. coli phage T4*).

Studies of phage λ (lambda) DNA recombining *via* the RecBCD system of *E. coli* demonstrated the existence of recombinants containing most, if not all, parental DNA material, indicating a break-join mechanism for RecBCD-mediated (including conjugal) recombination. The break-join paradigm for *E. coli* recombination has reigned for more than 30 years. Recently, attention is shifting back towards break-copy models. In a few unusual genetic systems, connections between recombination proteins and DNA replication (Rosenberg, 1997; Harris *et al.*, 1999) and between replication proteins and recombination (PriA in **Table 1**) have emerged. A snag in the renewed interest in break-copy models is that all of the evidence so far has been indirect or circumstantial, whereas the demonstration of break-join recombinants was direct. Here, we will review RecBCD-mediated break-join recombination, which does demonstrably occur. Our recent work (M.R. Motamedi, S.K. Szigety and S.M. Rosenberg, unpublished data) shows directly that a major fraction of RecBCD-mediated recombination actually proceeds *via* a break-copy route.

## Phage λ, Chi sites and double-strand DNA ends

The 48-kb DNA chromosome of phage λ provides one of the most important models for elucidating DNA elements in, and molecular mechanisms of, recombination *in vivo* for several reasons. First, it can be analysed both physically and genetically, and its replication can be controlled experimentally. Such analyses, done in the Meselson, Maurice Fox (Cambridge, USA), and Franklin Stahl (Oregon, USA) laboratories, demonstrated the roles of heteroduplex DNA, provided details of heteroduplex strand polarities in recombinant molecules and allowed examination of the involvement of DNA replication in recombination.

Second, λ is naturally devoid of *Chi sites*. Chi is a small sequence (5′GCTGGTGG3′) that controls conjugal recombination and double-strand break-repair in *E. coli* by modulating the activity of RecBCD and of RecA. Chi (crossover hotspot instigator) promotes recombination near itself and up to 20 kb away. The bacterial chromosome is saturated with Chi sites, which are present about every 4 kb, making Chi hard to study there. Its absence in λ let workers see its effect on *E. coli* recombination (studies done with λ lacking the λ recombination genes) by comparing recombination in λ strains carrying $Chi^+$ mutations, with naturally $Chi^\circ$ λ. Extensive genetic and physical characterization of Chi sites in the Stahl lab and then biochemical characterization (Gerald Smith, Seattle, USA, and Steven Kowalczykowski, California, USA, laboratories, and others) have led to elaboration of the detailed possible molecular mechanisms of recombination discussed below.

Third, phage λ has a single, defined, double-strand endonucleolytic cut site, *cos*, at which the circular replicating λ chromosome is linearized during packaging. This was used by Stahl and colleagues to demonstrate that Chi works only on linear molecules, and to provide evidence that RecBCD loads on to double-strand ends *in vivo*, interacts with Chi sites that it encounters downstream, and is switched to recombination mode at Chi, *via* RecD modulation of the enzyme. These features and many aspects of data from protein biochemistry are combined into a model for a possible molecular mechanism of recombination of linear DNA in the next section.

## Mechanisms Promoting Recombination

A model for a molecular mechanism of recombination of linear DNA in *E. coli* is shown in **Figure 3**. That it is a *model* indicates that some steps in the picture are hypothetical. We will attempt to convey the degree of certainty (or lack of it) with which each step is understood or hypothesized. The following steps follow the diagram in **Figure 3**.

### Presynapsis

(a) RecBCD enzyme loads onto double-strand ends of DNA.

(b) RecBCD degrades the DNA until it reaches its recognition sequence, Chi. The exonuclease activity of RecBCD is a combination of helicase (unwinding) and single-strand endonuclease (cutting) activity.

(c) The interaction of RecBCD with Chi is believed to induce a change in the enzyme's structure and activity, leaving the enzyme in a state equivalent to a RecBC enzyme without the RecD subunit. Though not shown in this picture, this interaction is facilitated by RecA.

(d) RecBC($D^-$) has no nuclease activity. The Chi-activated enzyme in this state is proposed to continue to travel along DNA, unwinding the DNA and producing single-strands with 5′- and 3′-ends. Other current models differ as to the exact means by which single-strand DNA is generated at Chi sites, and whether the DNA ends generated are of both or one polarity; however, all current models envision creation of single-strand DNA at Chi sites. Single-strand binding protein (SSB) is presumed to bind to the single-stranded DNA, removing regions of intrastrand base pairing.

(e) RecA protein then coats the single-stranded DNA. RecF, RecO and RecR proteins are postulated to aid RecA in coating of single-strand DNA. The role(s) of RecQ DNA helicase in recombination is not yet understood, but may be in creation of single-strand DNA. Single-strand DNA exonucleases ExoI, ExoVII and RecJ promote recombination, but at which stage is not yet known. Here we illustrate the possibility that they help perpetuate single-strandedness by degrading one single strand so that its complement cannot reanneal, thus driving the equilibrium in this reaction downwards (in **Figure 3**) towards recombination.

**Figure 3** is not meant to imply that each act of recombination has a 3′-heteroduplex joint at one end and a 5′-joint at the other. We merely suggest that the joints can be of one or the other polarity.

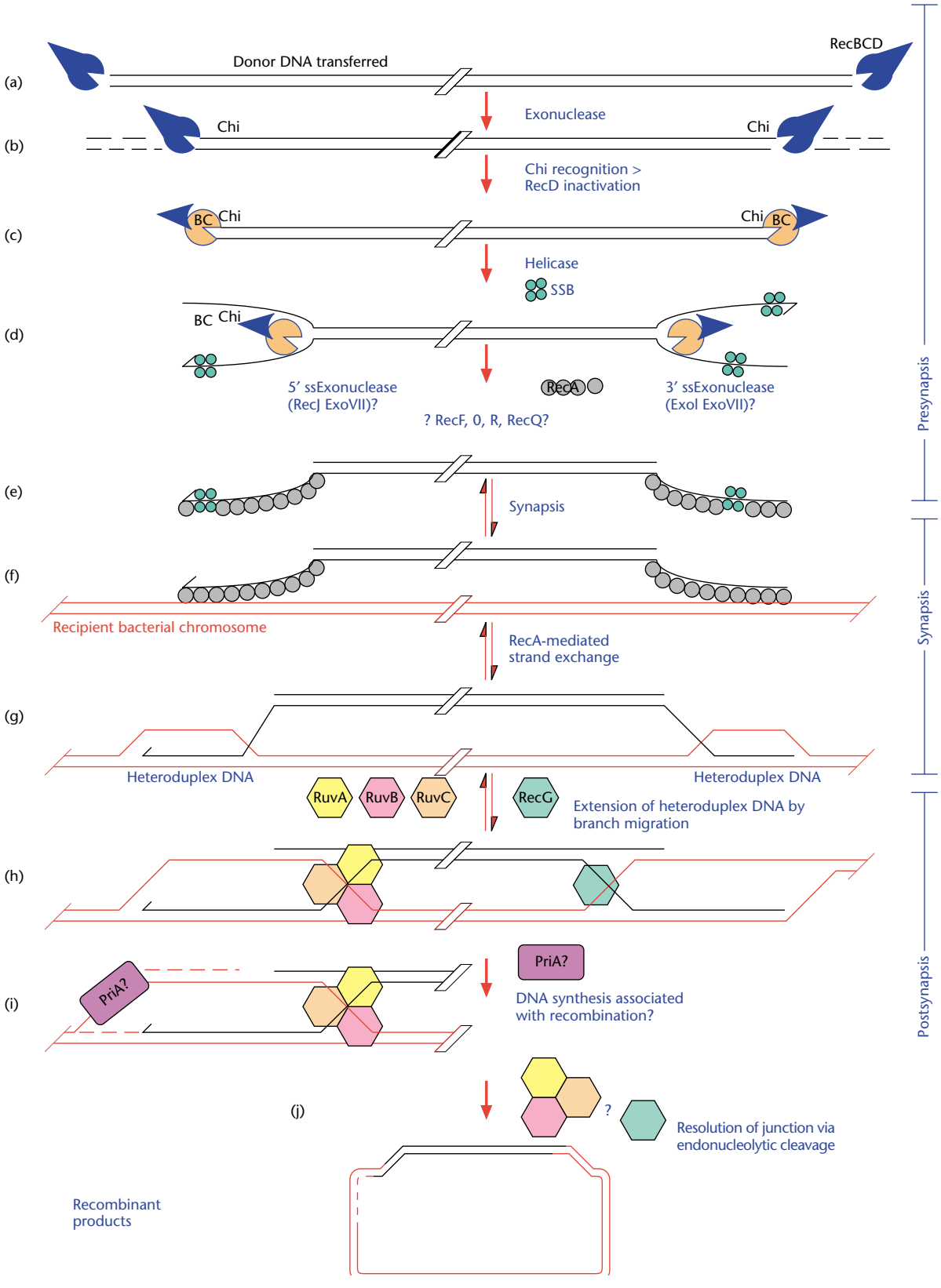
### Synapsis

(f) RecA initiates interaction of the single-strands with duplex DNA and conducts a search for similar DNA sequences in the duplex. RecA recognizes and uses Chi-containing, GC-rich single strands more efficiently than other DNA, thus coupling RecBCD and Chi recognition to synapsis.

(g) When a region of complementarity with the single strand is found, RecA promotes strand exchange, creating heteroduplex DNA joints that hold the recombining molecules together, precisely in register.

### Postsynapsis

(h) The RuvA, B and C proteins appear to act together and bind DNA specifically at strand-exchange junctions. RecG protein also has this ability. RecG and RuvB (with RuvA) are both DNA helicases that are specific for crossed-strand junctions (including four-stranded *Holliday junctions*). The junction-specific helicase activities of these complexes probably extend the regions of heteroduplex DNA by *branch migration* of the junctions.


(a)
Donor DNA transferred
RecBCD
(b)
Chi
Exonuclease
Chi
Chi recognition >
RecD inactivation
(c)
BC
Chi
Chi
BC
Helicase
SSB
(d)
BC
Chi
5′ ssExonuclease
(RecJ ExoVII)?
RecA
3′ ssExonuclease
(ExoI ExoVII)?
? RecF, 0, R, RecQ?
(e)
Synapsis
(f)
Recipient bacterial chromosome
RecA-mediated
strand exchange
(g)
Heteroduplex DNA
Heteroduplex DNA
RuvA
RuvB
RuvC
RecG
Extension of heteroduplex DNA by
branch migration
(h)
(i)
PriA?
PriA?
DNA synthesis associated
with recombination?
(j)
?
Resolution of junction via
endonucleolytic cleavage
Recombinant
products
Presynapsis
Synapsis
Postsynapsis

(i) A few lines of indirect evidence have suggested that DNA synthesis may sometimes be associated with recombination (e.g. Rosenberg, 1997; Harris *et al.*, 1999). In this diagram, synthesis is indicated at one of the two heteroduplex joints, the 3′-joint, by dashed lines. Whether synthesis is normally associated with recombination, the extent of such synthesis, whether it is located at joint regions and the fate of newly synthesized DNA in recombinant molecules are topics of active current investigation (M.R. Motamedi, S.K. Szigety and S.M. Rosenberg, unpublished data).

### Resolution

(j) RuvC is a DNA *endonuclease*, an enzyme that breaks phosphodiester bonds in the DNA backbone internally, that is, not at a DNA end. RuvC is a junction-specific endonuclease, that probably works with RuvA and RuvB in this activity. RuvC is hypothesized to cut at strand exchange (Holliday) junctions to produce break-join recombinants. The apparent redundancy of the RuvABC system with RecG (**Table 1**) suggested that these two might provide alternative resolution systems; however, the details of which components catalyse what kind(s) of resolution events *in vivo* are not yet clear.

## Interspecies Crosses

The genomes of eubacteria, and probably also eukaryotes, appear to be somewhat of a patchwork of segments of DNA that have come from other species *via* interspecies and horizontal genetic transmission. Such transmission from bacteria to eukaryotes is responsible for the symbiotic association that allows nitrogen fixation in plant roots, for formation of some tumours in plants and animals, and for the spread through bacterial populations of advantageous genes (for the bacteria) such as those that allow pathogenicity and resistance to antibiotics. Between closely related species, the DNAs, once transferred, can become incorporated into the foreign genome by recombination with similar DNA sequences. However, mechanisms exist to inhibit recombination of nonidentical sequences.

## Mechanisms aborting recombination

### Mismatch repair corrects DNA polymerase errors and prevents recombination of homeologous DNAs

Phage and bacterial molecular geneticist Miroslav Radman (Paris) and colleagues have pioneered a profound understanding of recombination between the imperfectly identical DNAs (called *homeologous* DNAs) of different species. Studying Hfr-mediated and transductional recombination between *Salmonella typhimurium* and the 85% identical DNA of *E. coli*, the Radman group has developed the concept of abortion of recombination (Matic *et al.*, 1995). The idea is that certain steps in the recombination reactions are reversible (e.g. steps(e)–(h) in **Figure 3**), or can be blocked, and if the reverse reactions are favoured, recombination is decreased dramatically. Their work showed that this occurs when the DNAs recombining are not perfectly identical, such that base mispairs occur in the heteroduplex joints of recombination. This is an important factor that normally impedes interspecies genetic transfer.

Abortion of homeologous recombination is an active process mediated by proteins that recognize mispaired bases in a heteroduplex DNA joint. These proteins, the *mismatch repair* proteins constitute a DNA repair system conserved among eubacteria and eukaryotes (but apparently absent in the few archaebacteria whose genomes have been sequenced). Mismatch repair provides the major route to mutation avoidance from DNA polymerase errors (**Figure 4a**). In *E. coli*, when polymerase errors occur, leaving an incorrect base paired with the template strand (**Figure 4a–i**), MutS protein binds the mismatched base and is then bound by MutL protein, which coordinates between the bound mismatch and two other proteins, MutH, a DNA endonuclease (**Figure 4a–ii**), and MutU, a DNA helicase. Together these proteins nick the newly synthesized strand of DNA (**Figure 4a–iii**), unwind it (**Figure 4a–iv**) and remove a segment of the new DNA strand that contains the mismatched base. The single-strand gap created is then filled by resynthesis, inserting the correct base (**Figure 4a–v**). This process helps avoid mutation from polymerase errors.

In the distantly related bacterium *Streptococcus pneumoniae*, and in eukaryotes, MutS and MutL homologues exist, as do helicases similar in function to MutU, but MutH homologues have not been found, nor has the enzyme (not discussed here) that instructs MutH as to the identity of the new strand. In these other species, it is hypothesized that single-strand nicks in the newly

**Figure 3** A model for a molecular mechanism of recombination of linear DNA in *Escherichia coli*, illustrated for conjugation. The model is synthesized (with modifications) from three sources (Rosenberg and Hastings, 1991; Harris *et al.*, 1996; Razavy *et al.*, 1996). Step-by-step description of this model appears in the text (Mechanisms Promoting Recombination). Each strand of duplex DNA is represented as a line. In single strands, arrowed ends represent 3′-ends; nonarrowed ends represent 5′-ends. Dashed lines represent newly synthesized DNA. The black DNA molecule represents linear DNA that will recombine with the (red) circular bacterial chromosome. SSB, single-strand binding protein.

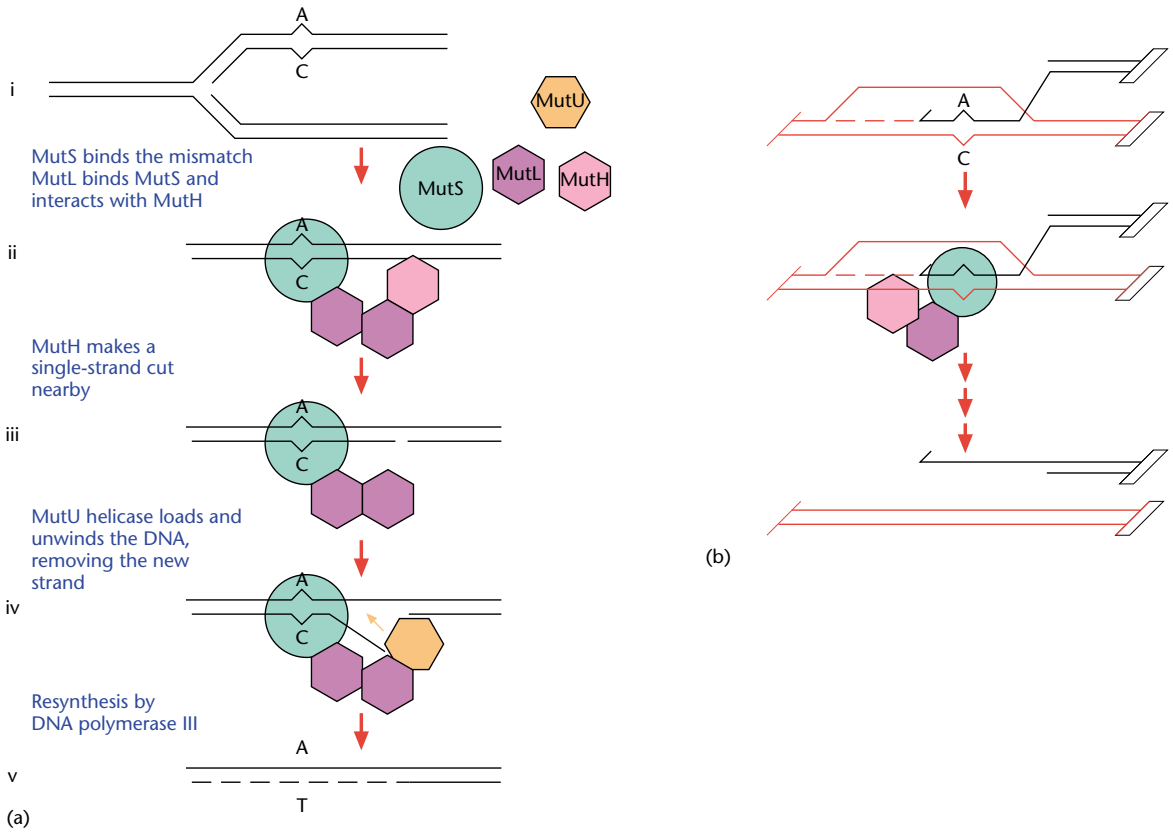



**Figure 4** Mismatch repair promotes the fidelity of DNA replication and recombination. (a) Mismatch repair proteins recognize incorrectly paired bases (shown) and 1–4 base insertion or deletion single-strand loops (not shown) that result from DNA polymerase errors. Step by step discussion of this process appears in the text (Mechanisms Aborting Recombination). (b) Abortion of recombination between imperfectly identical DNA sequences by mismatch repair proteins. Mismatch repair proteins bind to mispairs (shown) and 1–4 base insertion/deletion loops (not shown) that result from heteroduplex DNA formation between homeologous DNAs during recombination, and impede recombination. The mechanism(s) by which mismatch repair proteins inhibit recombination after they have bound the mispairs in heteroduplex DNA is not yet understood (three downward arrows).

synthesized strand direct repair to that strand without the need of a MutH-like endonuclease.

During recombination, mismatch repair proteins attack heteroduplex DNA joints that contain mispaired bases and block continuation of recombination by mechanisms not fully understood (Matic *et al*., 1995). In **Figure 4b**, MutS is shown binding a mismatch in homeologous heteroduplex DNA with the help of MutL. Work from the Radman and Paul Modrich (Durham, North Carolina, USA) laboratories indicates that MutS and MutL are the most important players for impeding homeologous recombination. They appear to do battle with RecA during the strand-exchange process, thereby blocking completion of homeologous strand exchange. How this occurs is not understood. They impede strand exchange only when the strands are similar but not identical. MutH and MutU are less important players in blocking homeologous recombination. How they do so and at what stage in recombination they act are unknown. MutH appears to act in the subset of recombination events that include DNA synthesis. This is presumed to be dictated by the requirement of MutH for acting on newly synthesised DNA strands. Understanding the mechanism of abortion of homeologous recombination by mismatch repair proteins will require further study.

## Other mechanisms aborting recombination

Some of the proteins that normally promote recombination may also act to inhibit recombination. This is suggested by a few lines of evidence from work in *E. coli*, and also by the phenotype of humans with mutant alleles of *E. coli recQ* homologues. RecQ is a helicase, and humans with defective alleles of either of two human homologues of this gene display hyperrecombination (and have genetic diseases, Bloom and Werner syndromes, that predispose them to cancer; Ellis, 1997). This suggests that a role for some helicases may be to decrease recombination. For example, if strand-exchange intermediates were unwound by a helicase, recombination could be prevented. In *E. coli*, our studies of a mutation process that requires DNA recombination and DNA synthesis (perhaps using inter-

mediates such as in **Figure 3i**; Rosenberg, 1997; Harris *et al*., 1999), suggest that RecG junction-specific DNA helicase may sometimes inhibit one class of recombination events, those with DNA synthesis associated. This could occur by branch migration of a strand-exchange junction in the wrong direction, so that heteroduplex DNA is unwound rather than extended (the reverse of the process in **Figure 3h**; Harris *et al*., 1996).

A potentially important role of DNA replication in recombination may be to combat these abortion mechanisms, both abortion by mismatch repair and by helicases not specific to mismatched DNAs. For example, if replication extends the joint in homeologous recombination (**Figure 4b**), then mispairs in the heteroduplex may have a smaller disruptive effect because replication can produce a long, stable intermolecular joint without any base mispair (suggested by Michael Resnick, North Carolina, USA; and by Harris *et al*., 1996). Similarly, replication could combat unwinding in nonmismatched substrates.

### Environmental influence on interspecies genetic transmission

Evidence from our laboratory, Radman's and Malcolm Winkler's (Houston, USA) has indicated that under stressful conditions, such as starvation, the mismatch repair system may be transiently downregulated (reviewed by Rosenberg, 1997). One profound consequence of this possibility is that organisms may experience heightened ability to recombine homeologous DNAs, and thus to acquire new genes from other species, when they are stressed. An active area of future work will be to understand the mechanisms and extent of generality of such environmental influence on genetic (in)stability.

## Summary

Conjugation is a form of bacterial sex in which DNA is delivered unidirectionally from a donor to a recipient, mediated by sex plasmids. When Hfr DNA is transferred, it can be incorporated into the recipient genome *via* homologous recombination. The linear donor DNA recombines using the bacterial double-strand break-repair system, in which RecBCD enzyme loads at duplex ends, creates single-strand DNA at Chi sites, and the single-strands invade the recipient DNA to form heteroduplex joints mediated by RecA. The joints can be resolved to produce break-join recombinants.

Homologous recombination is exquisitely sensitive to DNA sequence identity in the regions where heteroduplex DNA joints form, such that imperfectly identical DNAs are prevented from recombining by mismatch repair proteins. However, homeologous recombination does sometimes occur, and can be an important component of horizontal and sexual passing of genes between different species. Such radical genetic change appears to occur more often during stress. This may be an important factor in the evolution of genomes and organisms in the wild.

The bacterial genetic systems discussed here have analogous and homologous counterparts in other organisms, including simple and complex eukaryotes. Because molecular mechanisms of recombination and other aspects of genetic change are so much more accessible in the experimentally tractable bacterial and phage systems, these provide important models from which the paradigms of genetic stability, instability, change and evolution are understood in all organisms.

## Acknowledgements

Supported in part by grants R01-GM53158 and R01-AI43917 from the National Institutes of Health, USA, and by the Medical Research Council of Canada.

## References

Clark AJ and Sandler SJ (1994) Homologous genetic recombination: the pieces begin to fall into place. *Critical Reviews in Microbiology* **20**: 125–142.

Ellis NA (1997) DNA helicases in inherited human disorders. *Current Opinion in Genetics and Development* **7**: 354–364.

Firth N, Ippen-Ihler K and Skurray RA (1996) Structure and function of the F factor and mechanism of conjugation. In: Neidhardt FC *et al.* (eds) *Escherichia coli and Salmonella Cellular and Molecular Biology*, 2nd edn, pp. 2377–2401. Washington, DC: ASM Press.

Harris RS, Ross KJ and Rosenberg SM (1996) Opposing roles of the Holliday junction processing proteins of *Escherichia coli* in recombination-dependent adaptive mutation. *Genetics* **142**: 681–691.

Harris RS, Kong Q and Maizels N (1999) Somatic hypermutation and the three R's: repair, replication and recombination. *Reviews in Mutation Research* **436**: 157–178.

Masters M (1996) Generalized transduction. In: Neidhardt FC *et al.* (eds) *Escherichia coli and Salmonella Cellular and Molecular Biology*, 2nd edn, pp. 2421–2441. Washington, DC: ASM Press.

Matic I, Rayssiguier C and Radman M (1995) Interspecies gene exchange in bacteria: the role of SOS and mismatch repair systems in evolution of species. *Cell* **80**: 507–515.

Razavy H, Szigety SK and Rosenberg SM (1996) Evidence for both 3′ and 5′ single-strand DNA ends in intermediates in Chi-stimulated recombination *in vivo*. *Genetics* **142**: 333–339.

Rosenberg SM (1997) Mutation for survival. *Current Opinion in Genetics and Development* **7**: 829–834.

Rosenberg SM and Hastings PJ (1991) The split-end model for homologous recombination at double-strand breaks and at Chi. *Biochimie* **73**: 385–397.

## Further Reading

Hendrix RW, Roberts JW, Stahl FW and Weisberg RA (eds) (1983) *Lambda II*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Kowalczykowski SC, Dixon DA, Eggleston AK, Lauder SD and Rehrauer WM (1994) Biochemistry of homologous recombination in *Escherichia coli*. *Microbiological Reviews* **58**: 401–465.

Lloyd RG and Low KB (1996) Homologous recombination. In: Neidhardt EC, Curtis R III, Ingraham JL *et al*. (eds) *Escherichia coli and Salmonella Cellular and Molecular Biology*, 2nd edn: pp 2236–2255. Washington, DC: ASM Press.

Modrich P and Lahue R (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer. *Annual Review of Biochemistry* **65**: 101–133.

Myers RS and Stahl FW (1994) χ and RecBCD enzyme of *Escherichia coli*. *Annual Review of Genetics* **28**: 49–70.

Neidhardt FC, Curtiss R III, Ingraham JL *et al*. (eds) (1996) *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd edn. Washington, DC: ASM Press.

Roca AI and Cox MM (1997) RecA protein: structure, function, and role in recombinational DNA repair. *Progress in Nucleic Acid Research and Molecular Biology* **56**: 129–223.

Smith GR (1991) Conjugational recombination in *E. coli*: myths and mechanisms. *Cell* **64**: 19–27.

Snyder L and Champness W (1997) *Molecular Genetics of Bacteria*. Washington, DC: ASM Press.

West SC (1994) The processing of recombination intermediates: mechanistic insights from studies of bacterial proteins. *Cell* **76**: 9–15.