

## **Site-specific recombination**

Unlike homologous recombination that utilizes rather long stretches of homology between partner DNA molecules, site-specific recombination targets relatively short DNA sites with well-defined sequences. Whereas a large number of proteins with distinct biochemical activities cooperate to carry out homologous recombination, it is usually a single protein or a pair of proteins that carry out the catalytic steps of site-specific recombination. In some instances, the site-specific recombinase may be aided by one or more accessory proteins in synapsing the DNA partners and assembling the chemically competent recombination complex. Because of their overall simplicity, these systems have served as models for investigating the mechanisms of phosphoryl transfer reactions during recombination and related reactions.

## **Conservative site-specific recombinases**

As the name implies, during conservative site-specific recombination, specific phosphodiester bonds are broken and exchanged between two DNA partners, at the same time conserving the energy of these scissile bonds. To accomplish this, the recombinase mediates the strand cleavage and joining steps through a transesterification mechanism. The reaction is completed without degradation or synthesis of DNA and without the requirement for exogenous input of high-energy cofactors such as ATP.

Only two classes of conservative site-specific recombinases have been discovered so far: the invertase/resolvase family (also called the serine family) and the lambda integrase family (also called the tyrosine family). A few members of the two families have been studied in biochemical and structural detail with respect to the arrangement of the recombination complex, the organization of the recombinase active site, the chemical mechanisms of strand exchange and the conformational changes accompanying the reaction. Interestingly, these recombination systems are present almost exclusively in prokaryotes. Whereas no serine recombinase has been detected in eukaryotes, the eukaryotic tyrosine recombinases have been identified only in a small number of related yeast species belonging to a common genus. Either conservative site-specific recombination emerged after the splitting of the evolutionary lineages leading to prokaryotes and eukaryotes, or it was somehow lost

in the latter lineage. The resurgence of tyrosine recombination within a restricted yeast genus perhaps represents a freak evolutionary accident that imported the system through the horizontal transmission of a genetic element harboring it. Consistent with this notion, the yeast site-specific recombinases are encoded by extra-chromosomal selfish DNA elements that apparently confer no advantage to their hosts nor cause them any harm under normal steady state conditions.

## **The serine and tyrosine families: substrate recognition and reaction pathways**

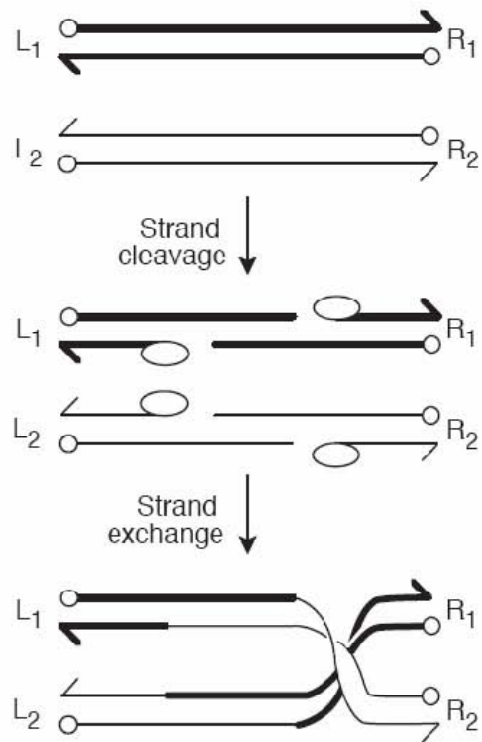
According to the old nomenclature, the two families of site-specific recombinases are distinguished by the type of reactions they carry out. The invertase/resolvase family mediates DNA inversion or DNA resolution (or excision). The integrase family derives its name from the Int protein of phage lambda, the prototype member of the family, which integrates the lambda genome into the bacterial chromosome. Under appropriate conditions, the same protein catalyzes phage excision as well. The new names define the families by the type of chemistry they utilize during recombination: one relies on serine and the other on tyrosine as their active site nucleophiles during DNA breakage.

The general reaction pathways followed by the serine and tyrosine families are illustrated in Fig. 1. For both families, the chemical steps of the reaction take place at core DNA sites that are quite similar in their overall organization. In a target site, two identical (or nearly identical) sequences, each comprising between one and two turns of DNA border a short DNA segment (less than a turn; commonly referred to as the strand exchange region or spacer) in a head to head (inverted) fashion. Each repeated element provides the binding site for a monomer of the recombinase. The serine recombinases in general appear to bind their target sites as dimers. By contrast, the tyrosine recombinases exist in solution and bind DNA as monomers, and establish strong monomer-monomer interactions following binding. In a rare case, exemplified by the XerC/XerD system, the target site contains the binding sequences for two distinct recombinase subunits, namely, the XerC and XerD proteins. The active recombinase is constituted by these two proteins acting in concert. The recombination reaction occurs in the context of two core sites and the four bound recombinase proteins brought together in a synaptic complex with a

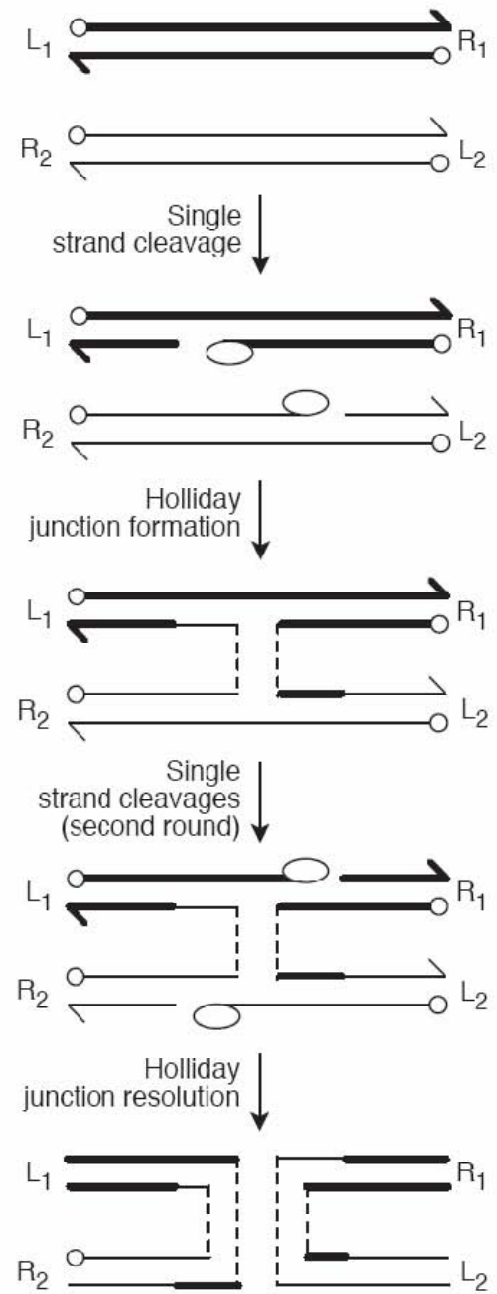
fixed geometry. In those recombination systems that utilize accessory proteins and DNA sites, the multiple DNA-protein interactions result in a complex synapse that is often characterized by a precisely defined DNA topology. The target phosphodiester bonds on the two DNA strands of a substrate (defining the left and right limits of strand exchange) are specified by the specific interaction of a recombinase subunit with its binding element.

Figure 1. Conservative site-specific recombination. A. The serine recombination reaction is initiated by double strand breaks in the two DNA partners arranged in a parallel fashion. Their left and right ends are marked by L and R, respectively. A 180 degree relative rotation of the cleaved recombination complex followed by strand joining completes recombination. Note that a half-twist is introduced in each of the DNA substrates, and a right handed crossing is formed between them. B. The tyrosine recombination reaction is initiated by the cleavage and exchange of one pair of strands to form a Holliday junction intermediate. Cleavage and exchange of the second pair of strands completes recombination. Note that the sites are arranged in an antiparallel manner, and there is no DNA crossing introduced between the recombination products. The knobs and split arrowheads in A and B represent the 5' and 3' ends, respectively.

2A



2B



The typical serine recombinase makes concerted double strand breaks in DNA, forming a 5'-phosphoserine bond and an adjacent 3'-hydroxyl group on each strand (Figure. 1A). In a long-held strand exchange model, now supported by crystallographic structural data, the cleaved recombination complex rotates through 180 degrees about its horizontal axis to bring the DNA strands in the strand joining configuration. This conformational transition may be described as an isomerization step- a term that has been more often used in the context of the Holliday junction dynamics during tyrosine recombination (see below). The 3'-hydroxyl groups now carry out nucleophilic attacks on the 5'-O-serine phosphodiester bonds to restore the DNA strands in the recombined state and release the recombinase from its covalent linkage to DNA. Note that the strand exchange step introduces two positive half-twists in the double helix, one each in each of the exchanged segments (equivalent to the relaxation of a negative supercoil), and one right handed crossing between the double helical partners.

If the recombining sites are in head-to-head (inverted) orientation, as is the requirement for an invertase, the outcome is the inversion of DNA between the recombining sites. If the sites are in head-to-tail (direct) orientation, as is mandatory for a resolvase, the consequence is the deletion of DNA between the recombining sites. However, the two reactive DNA sites are organized in a parallel geometry (left to right fashion for both DNA partners) within the synaptic structure during serine family recombination. The geometry of the synapse is the same for the inversion reaction (invertase) and the deletion reaction (resolvase).

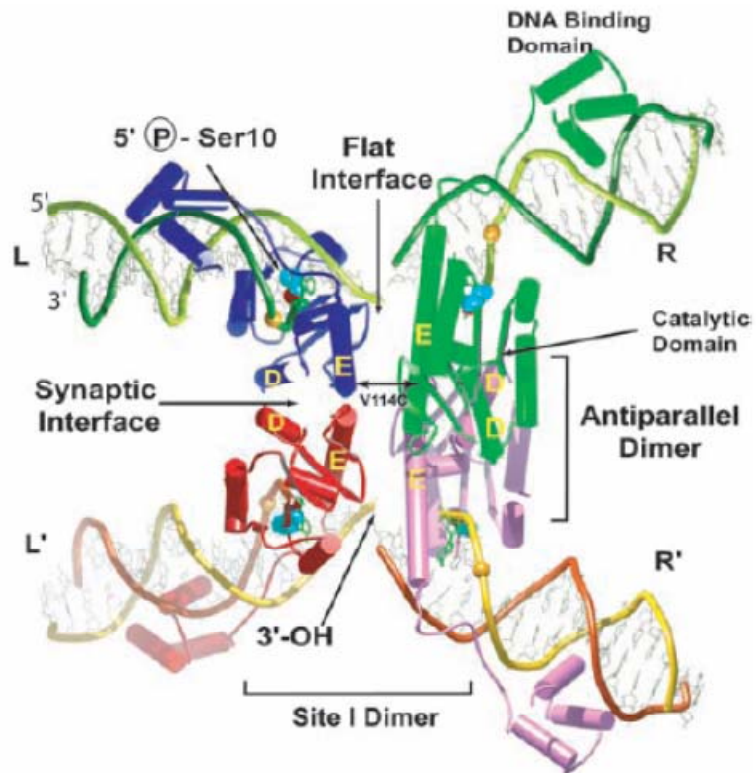
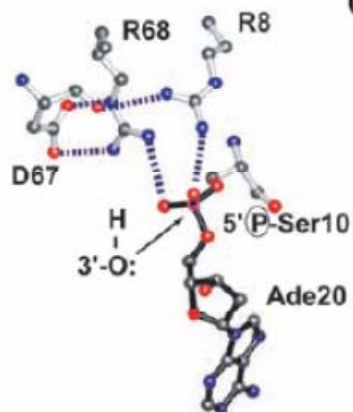
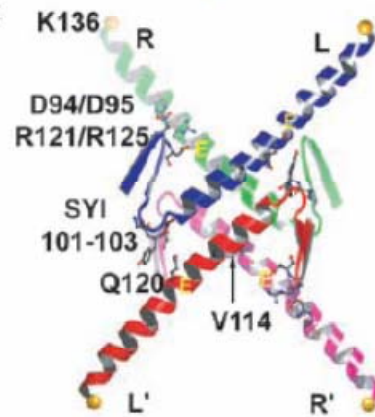
In contrast to the serine recombinase, the tyrosine recombinase makes single stranded DNA breaks on 'like' strands (either top or bottom) in the recombining partners using an active site tyrosine as the cleavage nucleophile (Figure 1B). The cleavage reaction follows the classical type IB topoisomerase mechanism, and the products are a 3'-O-phosphotyrosine bond and a 5'-hydroxyl group. The attack by the hydroxyl groups on the phosphotyrosine bonds in an inter-substrate orientation results in the formation of a Holliday intermediate. Reorganization of the recombination complex (or the isomerization of the Holliday junction) followed by the break/union of the second pair of strands completes recombination. Note that, unlike in the serine recombination case, the two DNA sites are arranged in opposite orientations (antiparallel geometry; left right in one case and right to left in the other), and the recombination products do not cross each other.

**IMPORTANT:** You should not get confused by the global direction of the recombining sites in the DNA substrate and the local arrangement of these sites within the protein-DNA complex. The global direction (head-to-head or head-to-tail) determines whether the result of recombination will be inversion or deletion. The local geometry (parallel or anti-parallel) has no effect on the final outcome of recombination: inversion or deletion. PLEASE PAY CLOSE ATTENTION WHEN WE DISCUSS THESE ASPECTS IN CLASS.

## **Crystal structures of recombinase-DNA complexes provide evidence for the mechanism and conformational dynamics of recombination**

Quite recently, the structure of the cleaved intermediate of resolvase recombination in which the enzyme subunits are covalently linked through Ser-10 to the scissile phosphates of two DNA duplexes has been solved (see Fig.2, below). This structure is an important step forward in understanding the mechanism of the recombination reaction. The structure suggests that the resolvase tetramer must undergo substantial conformational change as the presynaptic complexes (each DNA duplex bound by a resolvase dimer) come together in the recombination complex. The protein tetramer is arranged inside with the cleaved DNA duplexes lying on the outside. This will require large movement of the DNA and protein components in order to complete strand exchange. The two dimers linked to the cleaved ends are held together by a flat protein surface, which can facilitate the 180 degrees of rotation of one pair of dimers (at the left) with respect to the other (at the right).

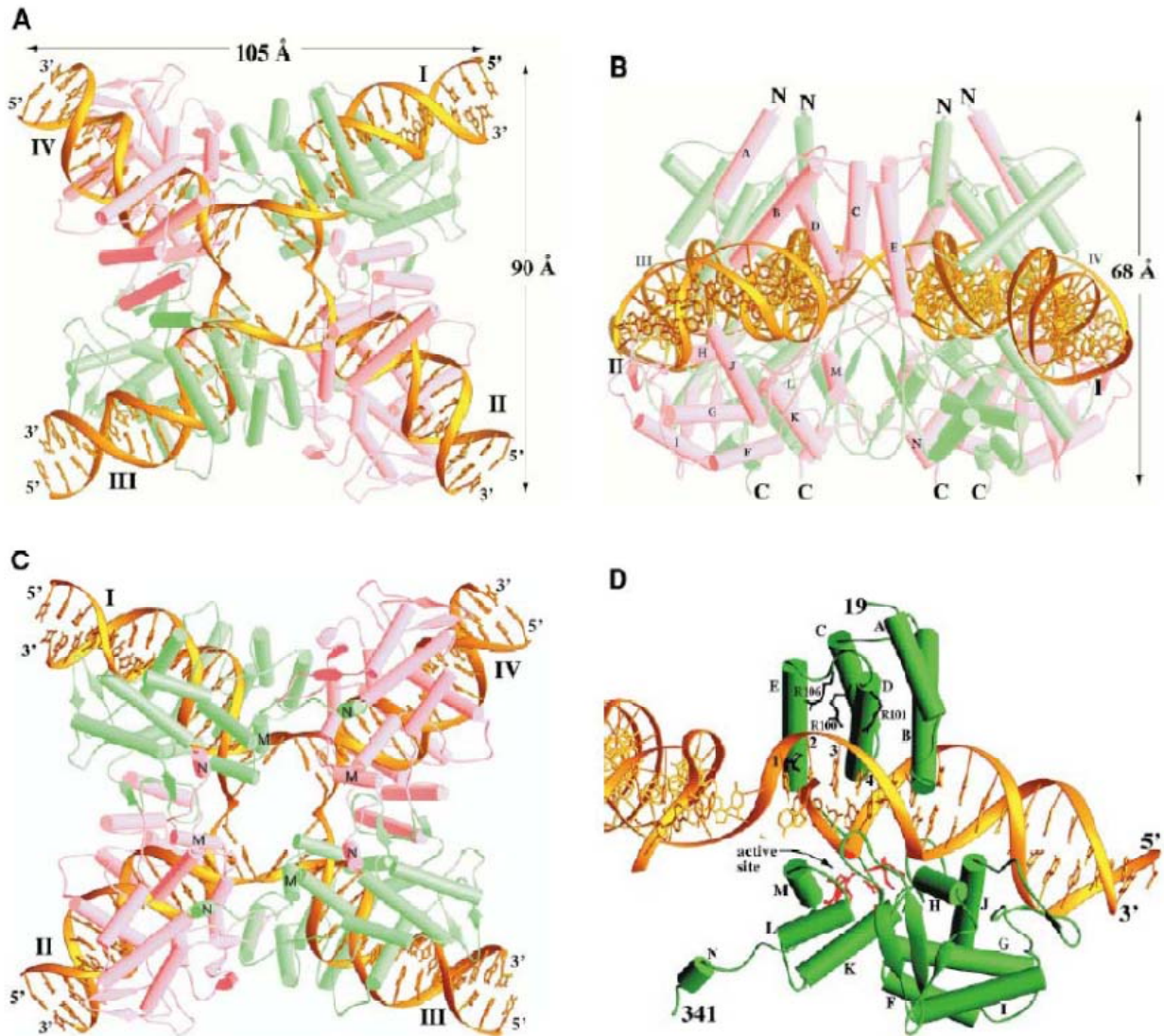
Figure 2. In this structure of the cleaved resolvase recombination intermediate, the two DNA duplexes are shown in green and orange. The resolvase monomers at the left are shaded in blue and red. They form an anti-parallel dimer (the way the E helices are arranged will help you appreciate the anti-parallel orientation). The resolvase monomers at the right are shaded in green and magenta, and form the anti-parallel right dimer. If you look carefully, you can see the Ser-10 in each subunit (blue-red; the red indicates the oxygen) covalently linked to the 5'-phosphate of adenine 20. See how flat the protein dimer interfaces are at the cleaved DNA junction. The left dimer (blue and red) and the right dimer (green and magenta) can rotate easily with respect to each other using this interface and bring the broken DNA ends in correct juxtaposition for strand joining in the recombinant mode.

**A****B****C**

The combined information from solution studies and crystal structures offer an impressively detailed picture of the active sites of tyrosine recombinases. In particular, the structures of the different Cre-DNA complexes and the Flp-DNA complexes provide rich insights into the possible mechanistic features of recombination. The figure below (Figure 3) shows the structure of Cre in association with the Holliday junction intermediate. Two monomers of Cre are shown in green; their partners are shown in pink. Because of the arrangement of the four protein monomers and the two DNA partners, only two of the four active sites are active at one time. This is the reason that strand exchange takes place in two separate steps. The first step generates the Holliday junction intermediate. This junction then undergoes an isomerization step, as depicted by the top (A) and bottom (B) panels of the figure. The result of this conformational change is that the first pair of active sites is inactivated, and the second pair is activated. The new active sites promote the second strand exchange step. As a result, the Holliday junction is resolved, and the recombinant products are formed.

Figure 3. The structure of the Cre recombinase complexed with DNA was solved by the van Duyne group. The four DNA arms of the Holliday junction are shown in gold. The four subunits of Cre are shown in green (two at the north west and south east corners) and in pink (two at the southwest and north east corners). A, B show different perspectives of the complex by rotating it. A and C show the two isomeric forms of the Holliday junction intermediate.





## Biological consequences of site-specific recombination

The types of DNA rearrangements promoted by a large number of site-specific recombination systems and their physiological consequences underscore one of the fundamental attributes of life: the capacity to employ the same or similar chemical mechanisms to bring about vastly different end results (Figure 4). It is almost axiomatic in biology that a solution arrived at in the context of a certain biochemical challenge is certain to be adopted and refined by evolution to be deployed in the context of a variety of related challenges.

The integration and excision reactions mediated by the lambda Int protein act as critical developmental switches in the phage's life style (Figure 4A). Integration leads to the quiescent lysogenic state, and excision triggers the multiplicative lytic pathway.

Related site-specific recombination systems are responsible for the dissemination of integrons or bacterial genes that are usually associated with antibiotic resistance or pathogenesis. Similarly, the conjugative transposons Tn916 and Tn1545 and related genetic elements move from the genome of a donor to that of a recipient bacterium by site-specific integration. Apparently, the 'transposon' is excised as a circle, and a single strand is transferred to the recipient via conjugation, which is then recircularized in double stranded form and finally integrated by site-specific recombination. The transposase of a conjugative transposon is in effect an integrase, and these elements are best described by their new name 'constins': conjugative, self-transmissible, integrating elements.

Programmed DNA rearrangements by site-specific recombination have been described in *B. subtilis* and in cyanobacterium. During sporulation of *B. subtilis*, a relatively large element is excised in the terminally differentiating mother cell to configure the functional form of the *sigK* gene (coding for a specific sigma factor). Similarly, site-specific deletion of intervening DNA elements occur in cyanobacterial heterocysts to establish the functional configurations of operons involved in nitrogen fixation.

The prokaryotic DNA inversion systems can control gene expression by reversing the orientation of a promoter or switching protein coding strands or by both means (Figures 4B and C). The Hin mediated DNA inversion controls phase variation in *Salmonella* by promoter inversion. In the 'on' state, the promoter drives the expression of one flagellin gene and the repressor for a second flagellin gene that is controlled by its own promoter. In the 'off' state, the lack of the repressor turns on the second flagellin gene. The ability to change its flagellar antigen gives an invading *Salmonella* population the opportunity to escape the host immune system. Of course, in order for the inversion reaction to serve its intended purpose, the frequency of recombination has to be tightly controlled so as to express only one type of flagellin protein in any one *Salmonella* bacterium. The Gin recombination helps bacteriophage Mu to fuse one of two variable protein domains to a constant domain in its tail fiber protein. The variable domains specify the phage's infectivity towards two distinct sets of bacterial hosts. Thus, there is a good chance that a phage population emerging from an infectious cycle will not be faced with extinction even if one bacterial type is absent in its ecological niche.

The resolvases encoded by transposons such as Tn3 or  $\gamma\delta$  are important in reducing the cointegrate intermediate formed during replicative transposition to a simple

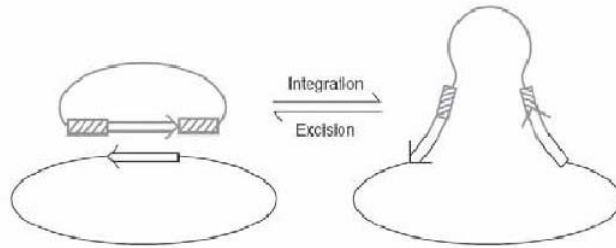
integrant (Figure 4D). Replicative transposition can coalesce the 'donor' plasmid and the 'recipient' chromosome into one circle containing two head to tail copies of the transposon (the cointegrate). Since resolvase mediated recombination releases the donor in its original configuration, it is now free to be potentially transmitted to another bacterium, having already deposited a copy of the transposon in the genome of its current host.

The XerC/XerD and related recombination systems ensure the equal segregation of bacterial chromosomes and certain plasmid genomes during cell division (Figure 4E). Due to homologous recombination, there is a finite probability that the duplicated circular chromosomes are present in a dimeric form. The action by XerC/XerD resolves the dimer into monomers that can then be partitioned into the daughter cells. The Cre recombinase is believed to play a similar partitioning role in the propagation of the unit copy bacteriophage/plasmid P1.

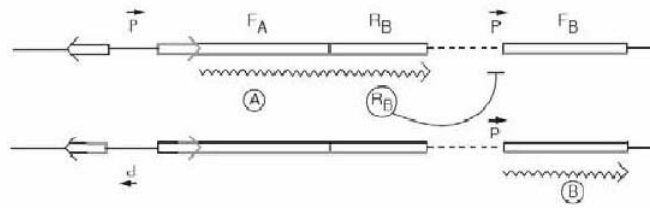
The Flp recombinase is central to the copy number control of the 2 micron yeast plasmid. In the event of a drop in copy number, a replication coupled recombination reaction is thought to trigger the amplification process that restores the steady state plasmid density. Because of the asymmetric location of the replication origin with respect to the *FRT* sites, the proximal site is duplicated first by the bidirectional fork. A recombination between the unreplicated *FRT* and one of the duplicated *FRT*s inverts one fork with respect to the other (Figure 4F). The unidirectional forks can spin out multiple copies of the plasmid by a bifurcated rolling circle mechanism. The tandem copies of the plasmid in the amplification product can be resolved into monomers by Flp recombination or by homologous recombination.

Figure. 4. Biological effects of recombination. The reactions A-F are explained in the text. A. A site-specific recombinase acting in the forward direction integrates an invading circular genome into a destination genome. The reaction may be reversible under appropriate conditions B. FA and FB represent genes coding for two variant forms of a protein; RB is the repressor for FB. P stands for transcription promoters. C. VA and VB are the variable regions and C the constant region of the two alternative forms of a protein. D. Tn denotes the transposon and the hatched bars the duplicated target sequence. E. The replication origin of a bacterial plasmid/chromosome is indicated by the dot; the arrows are the recombination sites for dimer resolution. F. The filled bars represent inverted recombination sites within the yeast plasmid. The arrows show direction of replication.

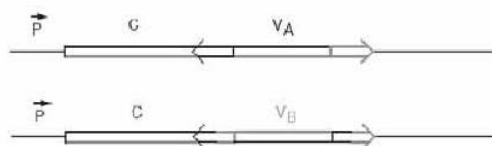
15A



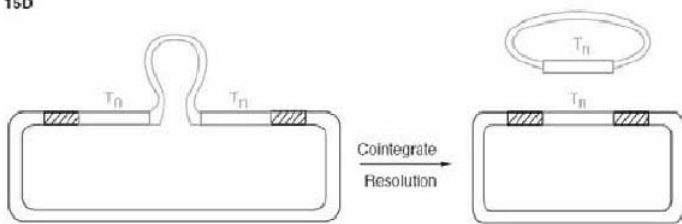
15B



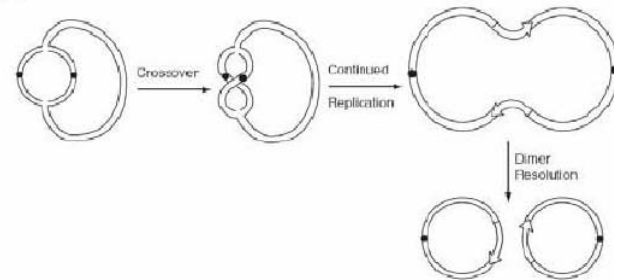
15C



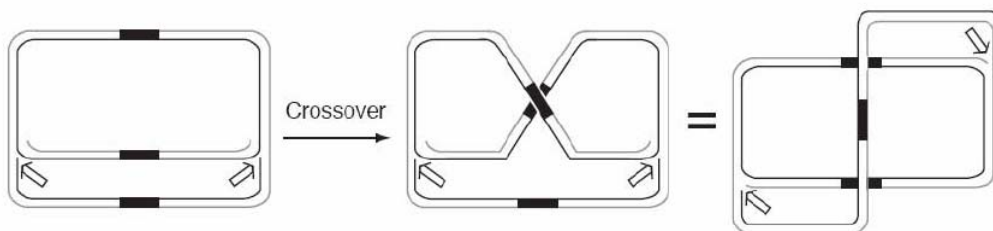
15D



15E

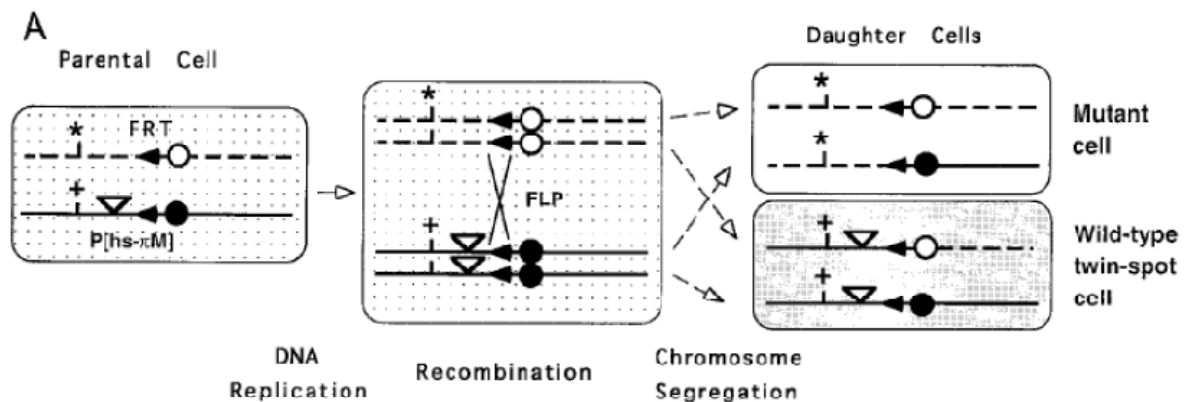


15F



# Applications of site-specific recombination in basic biology and biotechnology

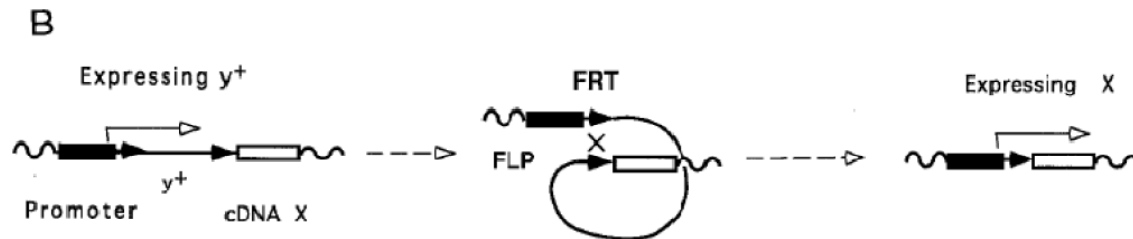
## Basic science applications



1. Site-specific recombination has been used as an important tool to study regulatory events that control development of organisms. In particular, the Flp-FRT (Flp recombination target site) system has been used in *Drosophila* to create mosaic flies by directed recombination between homologous chromosomes. This technique is useful in following cell lineages and cell fates. The Flp protein is expressed at a particular time in development from an inducible promoter such as the heat sock promoter to trigger recombination. In the example shown below, one chromosome carries a mutation (indicated by the asterisk) while its homolog carries the wild type allele (indicated by '+'). Flp recombination between the homologues will generate two sister chromatid pairs, each of which harbors \* on one sister and + on the other. When the chromosomes segregate, both mutant alleles can end up in one cell, and both wild type alleles end up in the other. The descendants of the former will all be mutants, and the descendants of the latter will all be wild type. The wild type and mutant cells are distinguished by the presence of a cell type marker associated with the wild type locus (indicated by the triangle). The preferred markers used now a days contain GFP (Green fluorescent protein) fused to proteins that localize to specific organelles (nucleus, for example) or to the plasma membrane. The marker can be visualized directly by fluorescence microscopy.

2. One can also delete a particular gene at a given stage of development, and follow the consequence of the deletion during further development. Here the locus of interest is

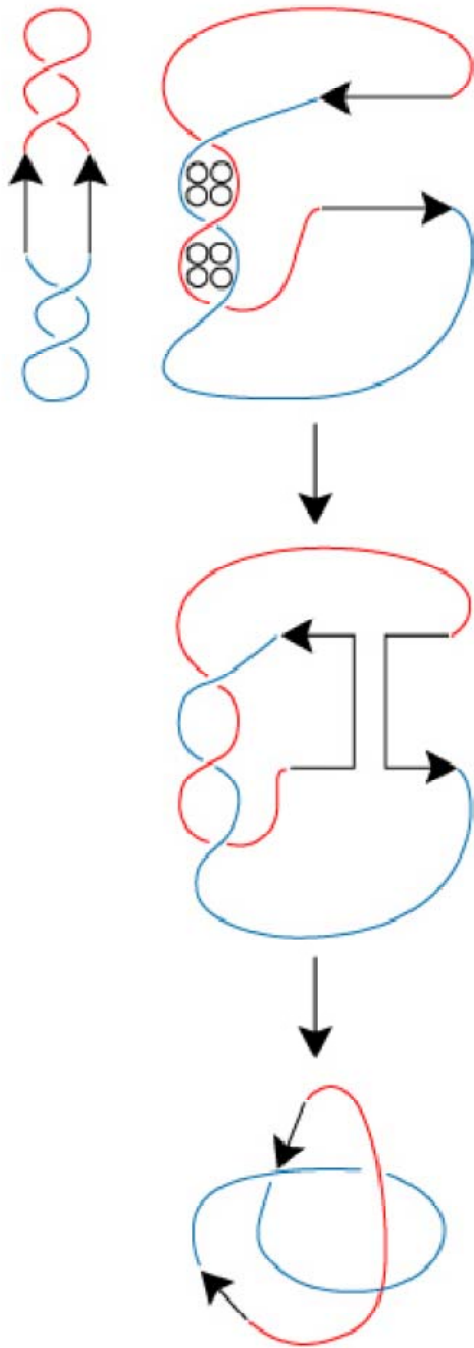
flanked by two copies of FRT in a direct or head-to-tail orientation (see diagram above). The Flp protein is induced from a regulated promoter. If one expresses the Flp gene from a tissue specific promoter, the deletion will occur only in that particular tissue. The effects of removing the gene function in a given tissue or a set of tissues can be followed.



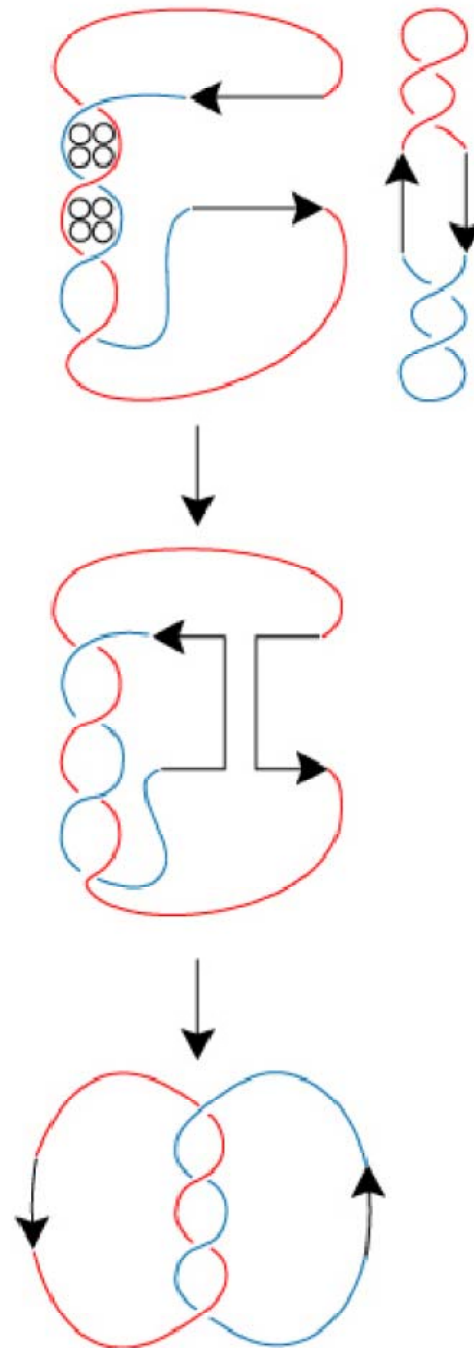
3. One can induce the expression of a gene at a desired point in development via site-specific recombination. For example, imagine a gene X engineered in such a way that its promoter is oriented in the non-functional direction. The promoter is flanked by two FRT sites in the direct or head-to-head orientation. When the Flp protein is induced at the appropriate time, recombination by Flp will invert the DNA segment containing the promoter. It is now turned around, and acquires the functional orientation, thus turning on the gene X of interest.

4. Since recombination exchanges DNA strands at specific target sites, it can seal off DNA domains. [It is roughly analogous to tying your shoe lace. Once tied, the crossings you have made by passing the lace through the eyelets of your shoes are stably preserved]. Imagine that a particular protein machine acting on DNA has assembled, on a circular substrate, a DNA synapse in which a finite number of DNA crossings are trapped. We can tie the shoe lace now by carrying out a site-specific recombination reaction with the target sites placed strategically so as to trap only the specific crossings made by the protein machine and not accidental non-specific crossings. By suitable methods one can count the trapped crossings (these are in the form of DNA knots or DNA links), and thus define the path of the double helix in complex DNA-protein assemblies. The diagram below indicates how the principle of this 'difference DNA topology' method. We will explain the rationale in greater detail in class.

Inversion



Deletion



## **Biotechnological applications**

1. Site specific recombination can be used to introduce a foreign DNA segment into a genomic locale where the recombination target site has been inserted by previous manipulations. The Input DNA is a circular vector harboring the foreign DNA and containing a copy of the recombination site. Recombination between the two target sites mediated by the site-specific recombinase causes the integration of the circle into the genome. [the integration reaction is equivalent to the phage integration reaction shown in Figure 7.]

2. Site-specific recombination can also be used to excise out a previously integrated DNA segment from the genome. It is equivalent to the reversal of the integration reaction. A related reaction is the excision of a native genomic locus surrounded by two copies of the recombination target site in head-to-tail orientation. Practical applications include the excision of pathogenic islands from bacterial strains, and production of large circles encompassing genomic regions for sequencing purposes.

3. In general, site-specific recombination has a variety of practical applications in agriculture, medicine and biotechnology- in the generation of transgenic plants and animals, in the production of bio-factories and in gene therapy.