

Biological consequences of site-specific recombination: integration, excision, deletion

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 (discussed below). 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.

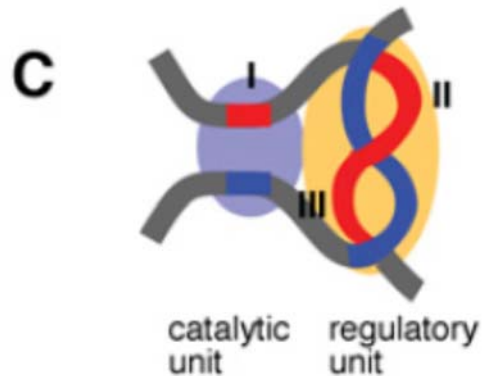
Site-specific recombination may occur between target sites located on two separate DNA molecules (intermolecular recombination) or between sites located within a single DNA molecule (intramolecular recombination). In the latter case the sites may be oriented in the same direction (direct orientation) or in opposite directions (inverted orientation).

Recombination between direct sites will result in the excision (or deletion) of DNA between the two sites as a circle. This reaction is called a deletion or excision (or resolution) reaction. The Tn3 resolvase (a serine recombinase that we discussed) catalyzes such a reaction. The term resolvase implies that this class of recombinases act only on directly oriented sites to promote the resolution reaction. They do not act on inverted sites or sites present on separate DNA molecules.

Recombination between inverted sites will result in the inversion of DNA between the target sites. The reaction is called a DNA inversion reaction. Recombinases that bring about inversion exclusively are called invertases. We will see a couple of examples of these enzymes. Invertases act only on target sites in the inverted orientation. They will not act on target sites in direct orientation or on target sites located on separate DNA molecules.

Strict invertases and resolvases often require negative supercoiling of DNA to organize the functional recombination synapse. For example, we noted earlier that the resolvase dimers bound to sub-sites II and III serve a regulatory function to bring the sub-sites I in parallel geometry within the recombination synapse. During this process three DNA supercoils are specifically trapped. These supercoils are formed by the crossing of the two DNA domains separated by the sub-sites I. Such supercoils are called inter-domainal supercoils. Supercoils formed within a domain (the DNA on one side of sub-sites I or the other side) are called intra-domainal supercoils. The topological architecture of the resolvase is often referred to as a [-3] synapse, referring to the three inter-domainal supercoils sequestered by the synapse. The requirement of supercoiling implies that resolvases (or invertases) act only on circular DNA molecules (or in pseudo-circular

domains formed within linear DNA molecules). To illustrate the [-3] synapse, I have reproduced below an earlier figure that we saw. If we join the ends of the DNA at the bottom and those at the top, you can see the contour of the circular negatively supercoiled DNA.

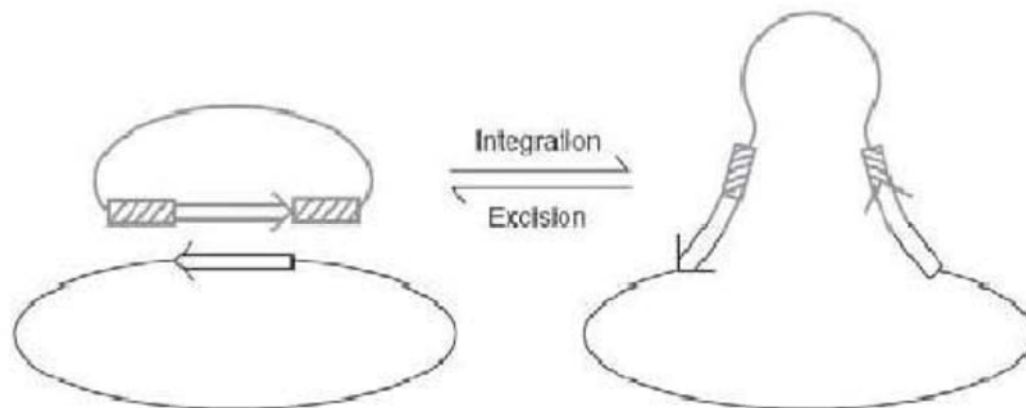


Some recombinases are not fastidious like the resolvase/invertase enzymes. They act on circular or linear molecules, and carry out inter-molecular recombination as well as inversion or deletion reactions. Examples of this type of recombinases include the tyrosine recombinases Cre and Flp. Because of their relaxed behavior towards DNA topology and site orientation, these 'simple' recombinases have been utilized for biotechnological applications in a variety of organisms.

Switching between developmental patterns

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

A



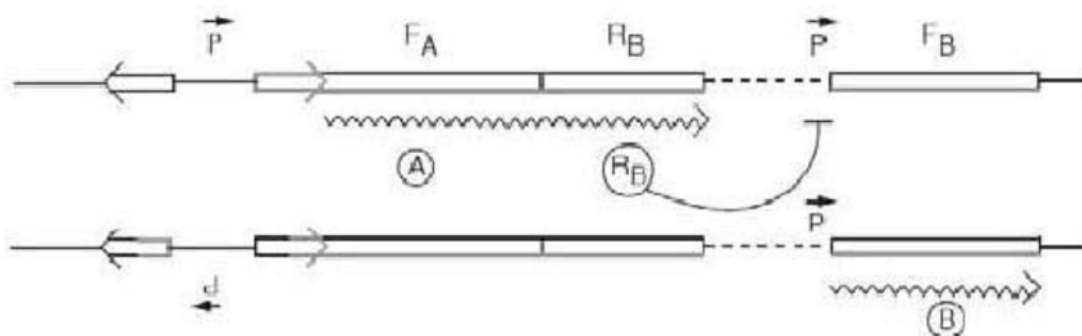
The integrase is a complex member of the tyrosine recombinase family. It requires the *E. coli* host factor IHF (integration host factor) to carry out the phage integration reaction. The phage DNA and the bacterial chromosome share a core sequence between which sites the Int recombinase brings about DNA breakage and joining. The site on the phage DNA (called the phage attachment site) is more complex than the bacterial attachment site. The phage site contains binding sites for the IHF protein and the phage coded Xis protein. In addition, there are also binding sites for the Fis protein (also a host protein). The excision reaction is chemically a perfect reversal of the integration reaction. However, this reaction requires, in addition to int and IHF, the Xis protein as well.

The forward reaction, namely, the integration of the phage into the *E. coli* chromosome, signifies lysogenization. In this state, the phage replicates as part of the chromosome and almost all genes of the phage are kept in a repressed state. Phage excision triggers the lytic phase of phage development. The excised phage genome expresses genes required for phage DNA replication as well as all the structural proteins required for producing mature phage. These phage particles emerge after lysis of the infected cell, and are ready to carry out infection of new host cells.

Thus, tyrosine recombination by lambda int determines whether the phage will become lysogenic (and remain quiescent) or turn lytic (and kill the host).

Altering gene expression patterns: Phase variation in bacteria

The host immune system responds to bacterial infection by mounting antibodies to proteins expressed on the bacterial surface. An abundant surface protein is flagellin, the subunit that is assembled into higher order structure called flagellum. A single flagellum may contain up to 3000 flagellin subunits, providing an attractive target for host immunity. The *Salmonella* bacterium utilizes site-specific recombination as a mechanism to escape the host immune response by displaying a flagellum containing a second type of flagellin subunit. This switching between the



two alternative flagellins is referred to as phase variation.

Phase variation is mediated by the action of a serine recombinase called Hin (stands for H-segment inversion). Hin is an invertase, and acts on target sites in inverted orientation to cause DNA inversion. In the Salmonella genome, the Hin target sites (shown as oppositely directed arrows in the Figure above) are located on either side of a promoter that drives the expression of two genes, one coding for the expression of Flagellin-A and the other coding for a repressor protein, which we may call R-B. Unlinked from this locus, there is a second flagellin gene in the Salmonella genome that codes for Flagellin-B. The expression of this gene is controlled by its own promoter, which, in turn, is regulated by the R-B repressor. When Flagellin A is expressed, the co-expressed R-B repressor will turn off the promoter for the Flagellin-B gene. Recombination mediated by Hin will invert the orientation of the promoter flanked by its target sites. With the promoter in the 'off' position, neither Flagellin-A nor R-B genes will be expressed. In the absence of the R-B repressor, the promoter for Flagellin-B is turned on. Under this condition, the flagellum assembled on the bacterial surface will contain Flagellin-B, and not Flagellin-A.

The frequency of the Hin-mediated inversion reaction is fine-tuned to meet the challenge of the host's immune system. When the immune response is targeted to say Flagellin-A (being the major antigen in an infecting population), cells that switch to Flagellin-B can escape this response. They fly under the radar of immune surveillance (so to speak), and may sustain the infection.

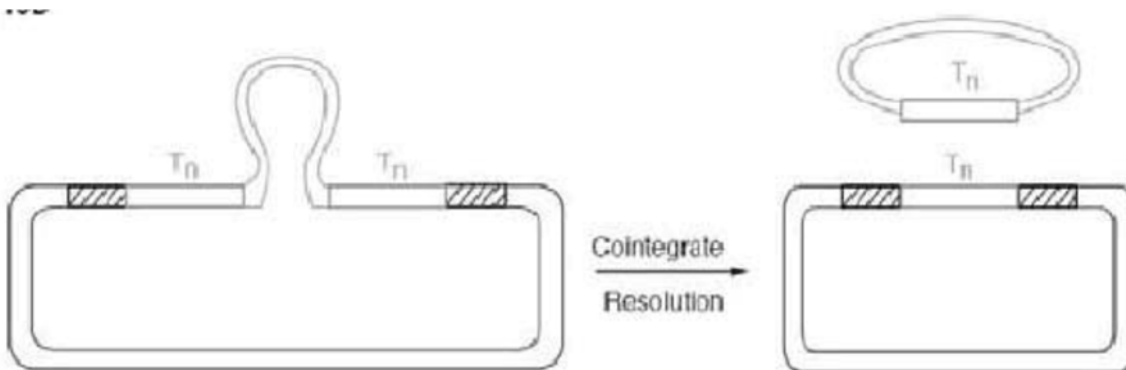
Antigenic variation is a classical strategy used by bacterial pathogens and viruses to evade the innate defense mechanisms of their hosts. The initial infection by a pathogen and the relevant immune response will result in immunologic memory of the foreign antigens encountered. If the same pathogen, displaying the same antigens, re-infects the host, the memory cells are activated, and the infection is quickly cleared. However, if the pathogen has an altered antigenic display, it has a reasonable chance of establishing an active infection.

Resolution of co-integrate intermediates formed during replicative transposition

The biological function of the resolvase site-specific recombinase that we considered in some detail is to promote an important step in the transposition of the transposon Tn3. Transposons are mobile genetic elements that can move from one location within a genome to a second location within the same genome or a different genome. For example, a transposon present within a plasmid may move to the chromosome of its host. Transposons often carry antibiotic resistance

genes, and are responsible for the spread of such genes within bacterial populations. {If time permits, we may discuss the different types of chemical mechanisms utilized by transposons for their mobility.}

Tn3 is an example of a replicative transposon. When it transposes from a circular 'donor' DNA (a plasmid, for example) to a second circular 'recipient' DNA (say, a bacterial chromosome), the transposon is duplicated in the process, and the two circles become conjoined in an intermediate structure called the co-integrate. Within the co-integrate, there are two copies of the transposon and two copies of a short sequence from the target site, which also gets duplicated during transposition. To complete the transposition event, the co-integrate must be resolved to leave one copy of the transposon in the recipient DNA, while retaining one copy of the transposon at its original donor location. The advantage of this mode of transposition is that the donor can now go through a second transposition event.

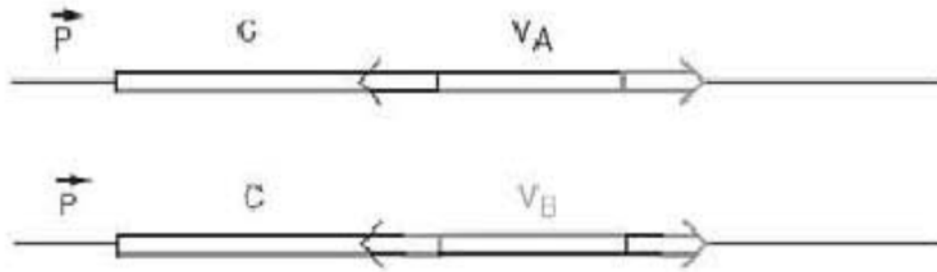


The resolution of the Tn3 co-integrate is mediated by the resolvase enzyme (see the figure above). The target site for resolvase is located within Tn3. Within the co-integrate, the Tn3 copies are in direct orientation, and so are the resolvase target sites located within Tn3. The action of resolvase at these sites results in the excision of the donor DNA containing one copy of the transposon. The other copy of the transposon is integrated into the recipient DNA. This transposition product is called a 'simple integrant'.

Altering host-specificity in bacterial viruses

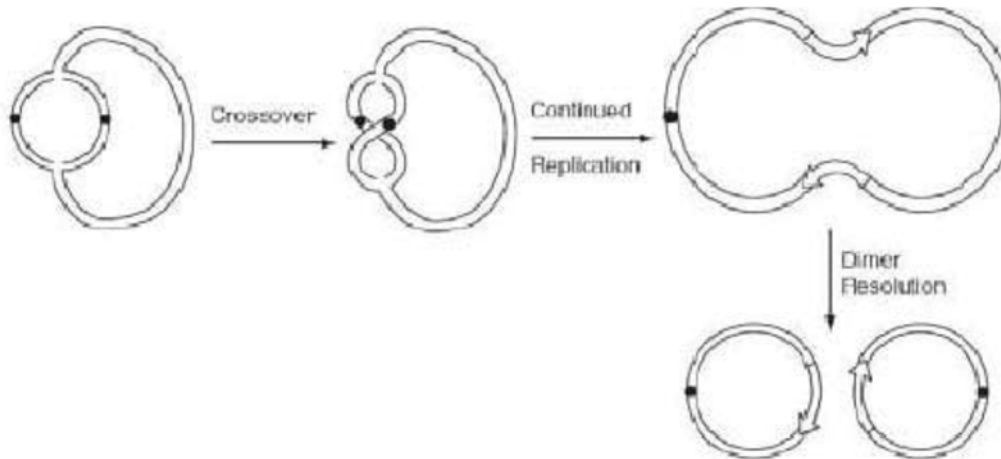
The bacteriophage Mu provides an example where site specific recombination is utilized to change the specificity of the phage in its infectivity towards susceptible bacterial populations. The recombination system harbored by the phage consists of the Gin recombinase (a serine site-specific recombinase, which is an invertase) and the invertible DNA segment G which is bordered

by the target sites for Gin. The G-segment is located within the coding region for the tail fiber protein of the phage (see Figure below).



This protein is organized into two domains, the N-terminal domain (C for constant) that associates with the phage tail and a C-terminal domain (VA or VB; V for variable) that interacts with the receptor of the bacterial host. The target sites for Gin are located on either side of the variable region. Gin recombination helps to fuse one of two variable protein domains the constant domain. 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.

Facilitating chromosome segregation during bacterial cell division



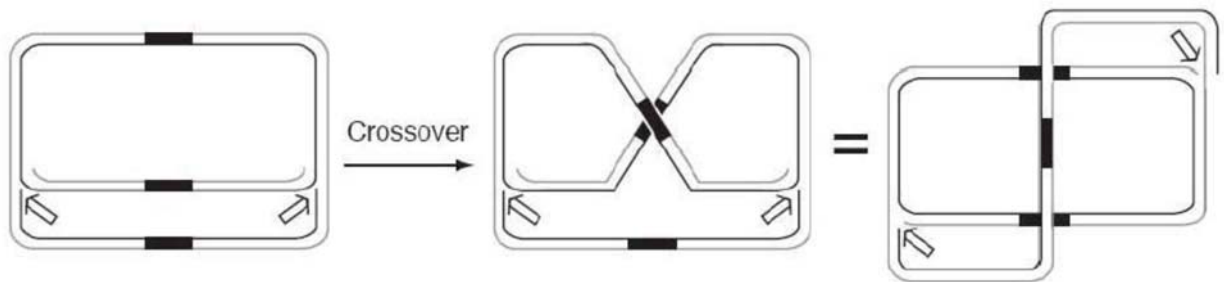
The XerC/XerD (tyrosine recombinase) and related recombination systems ensure the equal segregation of bacterial chromosomes and certain plasmid genomes during cell division (see the figure above). Due to homologous recombination, there is a finite probability that the duplicated circular chromosomes 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 target site for XerC/XerD is located near the terminus of replication. In the chromosome dimer, these sites are present in direct orientation. The action of XerC/XerD will resolve the dimer into two monomer chromosome circles, which can be segregated into daughter cells.

The Cre recombinase (also a tyrosine recombinase) is believed to play a similar partitioning role in the propagation of the unit copy bacteriophage/plasmid P1.

Copy number amplification in yeast plasmids

The Flp recombinase (a tyrosine recombinase) is central to the copy number control of the 2 micron yeast plasmid. This plasmid is a selfish DNA element present in the nucleus, and has a copy number of about 60 molecules per cell. The plasmid has the capacity to replicate efficiently during each cell cycle, and also partition the replicated molecules evenly between daughter cells. A rare missegregation event will result in a daughter cells with fewer than the steady state plasmid copy number. It is under this condition that the Flp protein comes into action during bidirectional replication of the plasmid.



A replication coupled recombination reaction is thought to trigger the amplification process that restores the steady state plasmid density. The Flp recombination target (*FRT*) sites are present in inverted orientation within the plasmid genome. One *FRT* site is proximal to the replication origin, the other distal to it. Because of the asymmetric location of the replication origin with respect to the *FRT* sites, the proximal *FRT* 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 (see Figure above; the *FRT* sites are shown as the filled rectangles). 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.