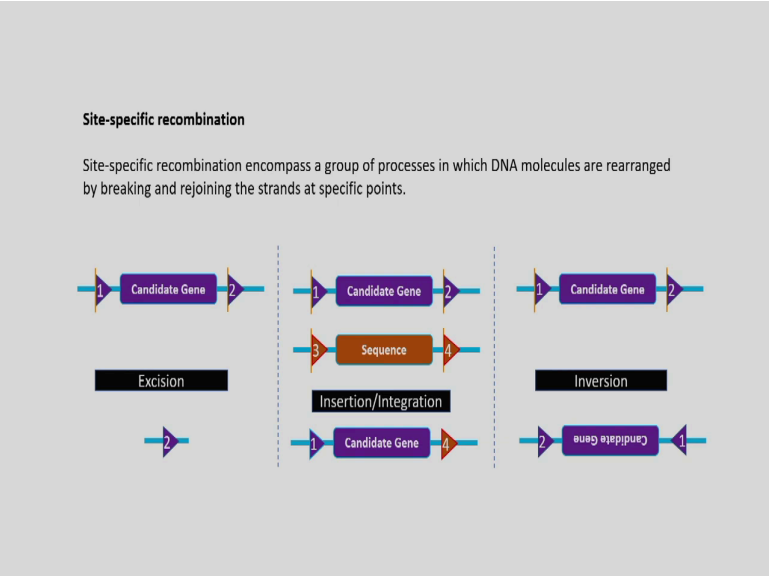
**Genome Editing and Engineering**
**Prof. Utpal Bora**
**Department of Bioscience and Bioengineering**
**Indian Institute of Technology, Guwahati**

**Module - 03**
**Recombination**
**Lecture - 02**
**Site Specific Recombination**

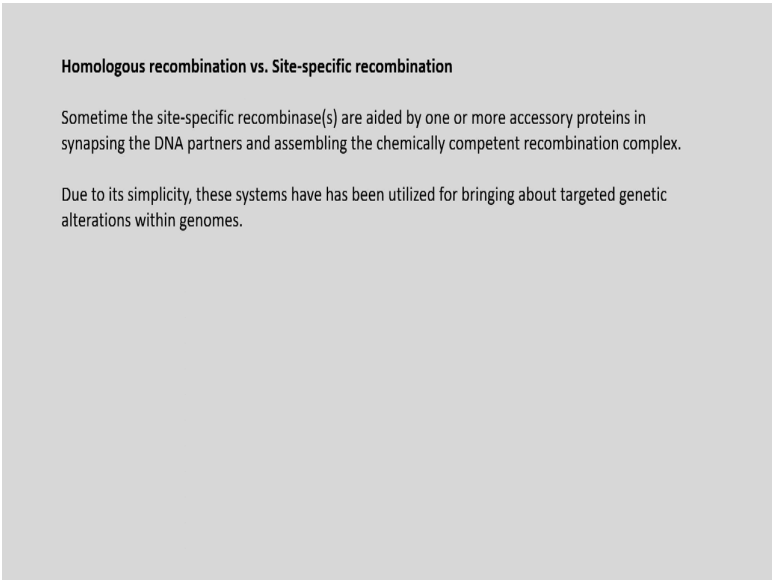
Welcome to module 3 of my course Genome Editing and Engineering. In this module we have been discussing about the Recombination process. Today we are going to discuss about the Site Specific Recombination.

(Refer Slide Time: 00:45)




Site specific recombination encompasses a group of processes in which DNA molecules are rearranged by breaking and re joining the strands at specific points. So, it involves excision insertion or integration and inversion. We will be discussing how these happens in the course of our lecture.

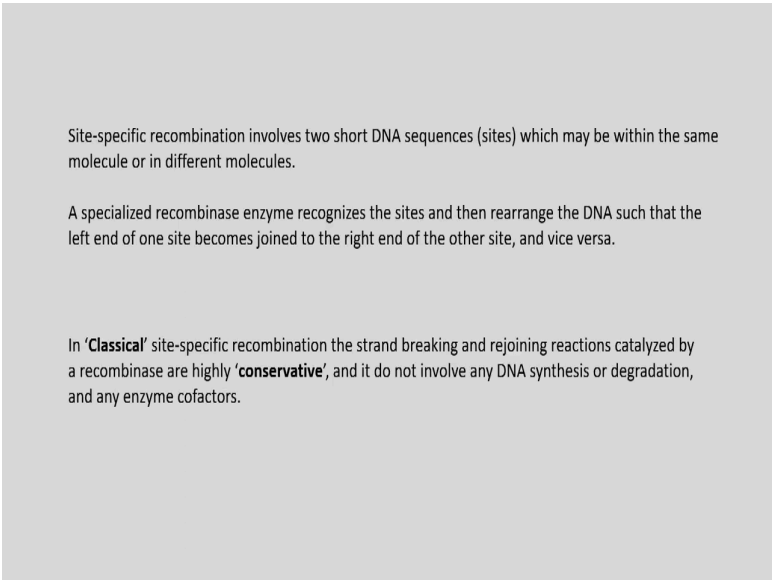
(Refer Slide Time: 01:07)




Let us have a understanding of the homologous recombination versus site specific recombination. In the last lecture we discussed about the hr which utilizes long stretches of homology between partner DNA molecules for double strand break repairs. This process involves a large number of proteins with distinct biochemical activities which cooperate to carry out homologous recombination.

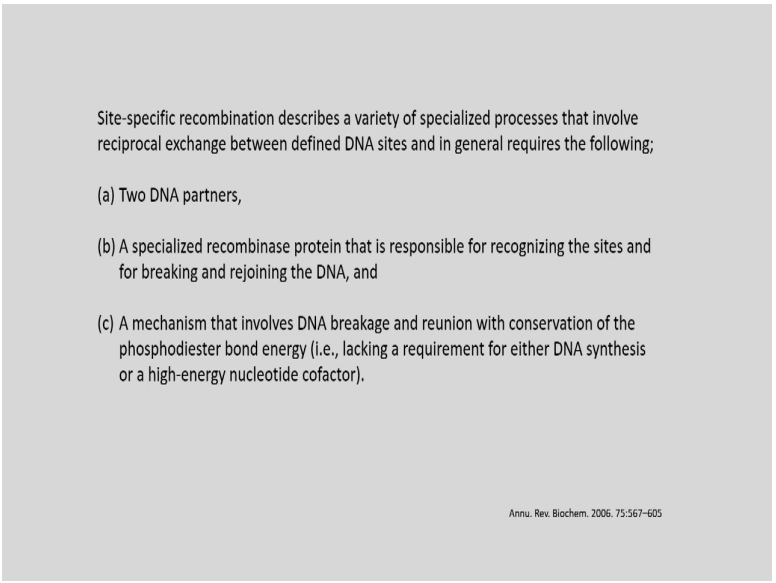
In contrast site specific recombination targets relatively short DNA sites with well defined sequences and usually a single protein or a pair of proteins is or are involved in carrying out the catalytic steps of site specific recombination. Sometimes the site specific recombination are aided by one or more accessory proteins in synopsis the DNA partners and assembling the chemically competent recombination complex. Due to its simplicity these systems has been utilized for bringing about targeted genetic alterations within genomes.

(Refer Slide Time: 02:19)




Site specific recombination involves two short DNA sequences or sites which may be within the same molecule or in different molecules. A specialized recombinase enzyme recognizes the sites and then rearrange the DNA such that the left end of one side becomes joined to the right end of the other side and vice versa. In classical site specific combination the strand breaking and rejoining reactions are catalyzed by recombinase and are highly conservative. It do not involve any DNA synthesis or degradation or any enzyme co factors.

(Refer Slide Time: 03:04)

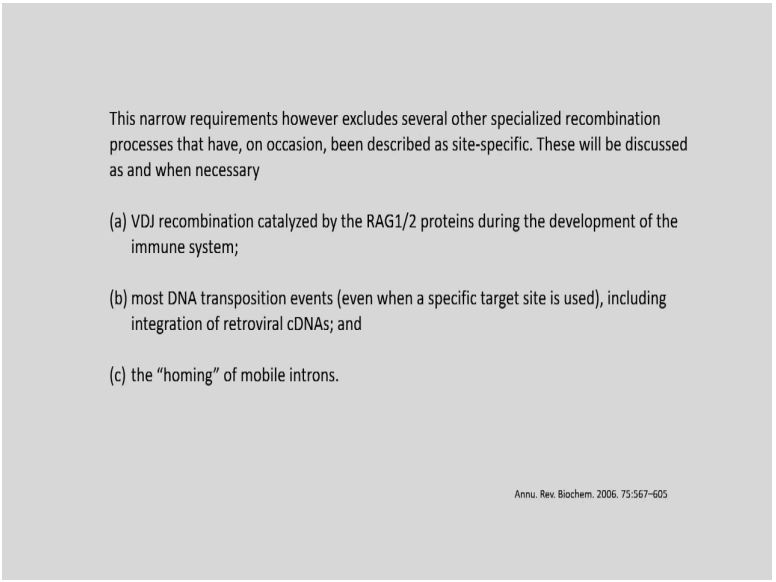


Briefly site specific recombination are a set of specialized processes that involve a reciprocal exchange between defined DNA sites and in general it requires the following. Number 1; it requires two DNA partners. It requires a specialized recombinase protein that is responsible for recognizing the sites and for breaking and rejoining the DNA. And thirdly, a mechanism that involves DNA breakage and reunion with conservation of the phosphodiester bond energy.

(Refer Slide Time: 03:41)

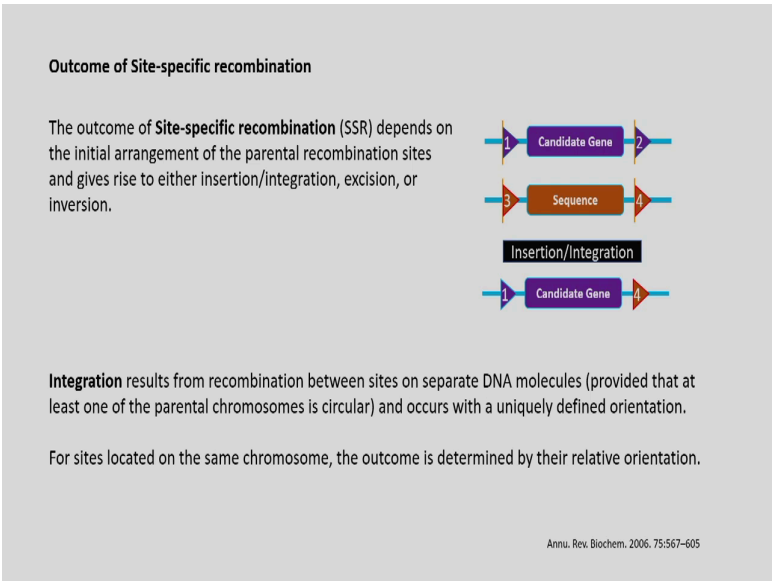

Some examples of SSR or Site Specific Recombination which meet the above criterion are the integration of bacteriophage lambda into the E coli chromosome. The resolution of cointegrates derived from transposition of Tn3 related transpositions and a DNA inversions responsible for flagellar phase variation in salmonella.

(Refer Slide Time: 04:06)




This narrow requirement; however, excludes several other specialized recombination processes that have been on occasion been described as site specific recombination these will be discussed as in when necessary. Some of these are VDJ recombination catalyzed by the RAG 1 2 proteins during the development of the immune system or immune reactions. Most DNA transposition events even when a site specific target is used including integration of retroviral cDNAs and the homing of mobile introns.

(Refer Slide Time: 04:42)

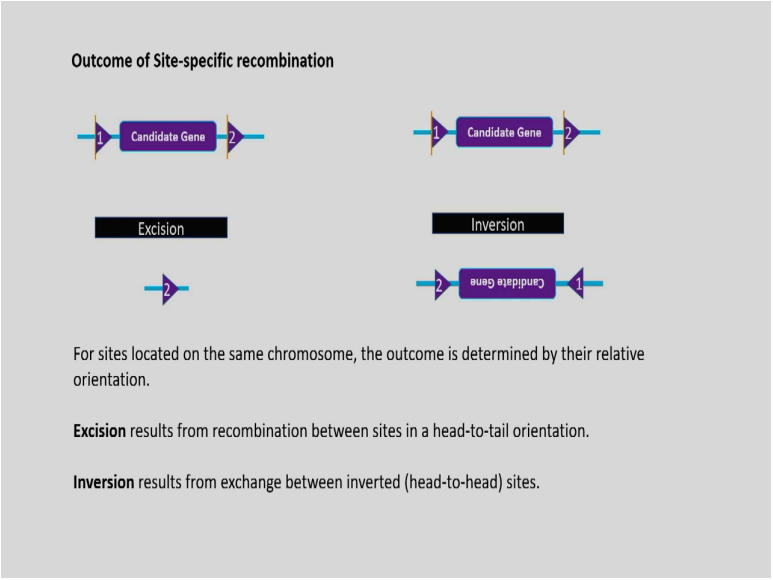


What is the outcome of the process of site specific recombination? You can see in this figure there is a candidate gene and then there is a sequence of interest. And these are flanked by say the candidate gene is flanked by some genetic sequence 1 and 2 the target sequence is flanked by some sequence 3 and 4. And if these candidate gene is to be integrated you can see there is a change in the flanking genes the candidate gene carrying along with it the flanking gene 1 to the sequence site which is now having on the right side the flanking gene 4.

So, this is just the schematics of the outcome of the site specific recombination. So, this; however, depends on the initial arrangement of the parental recombination sites and gives rise to either insertion integration or excision. Sometimes instead of these insertion there may be excision and there may also be inversion.

Now, this integration results from recombination between sites on separate DNA molecules provided that at least one of the parental chromosome is circular and occur with a uniquely defined orientation. For sites located on the same chromosome the outcome is determined by their relative orientation.

(Refer Slide Time: 06:17)



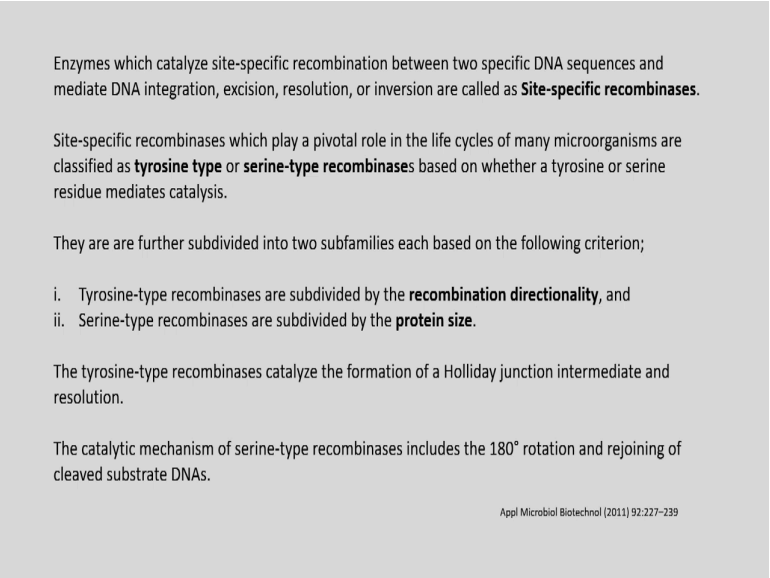

Site specific recombination may also lead to excision of genes. So, there is a candidate gene over here. Now, as excision has happened. You can see the gene is no longer in the scene and it has disappeared and in the case of inversion you can see here the gene has been inverted. For sites located on the same chromosome the outcome is determined by the relative

orientation. Excision results from recombination between sites in a head to tail orientation while inverse and results from exchange between inverted head to head sites.

(Refer Slide Time: 06:55)


So, this is one of the important papers which was published in the PNAs and gets in Cox reported the FLP recombinase and its enzymatic activity.

(Refer Slide Time: 07:15)



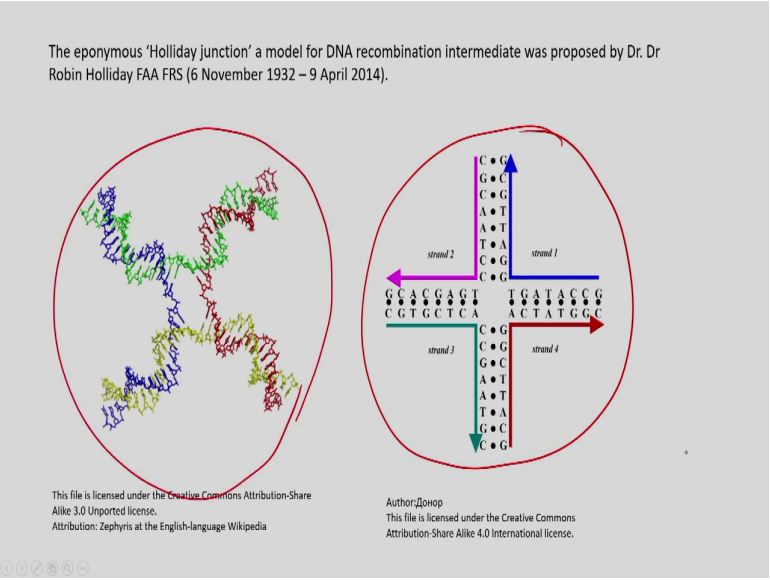

Enzymes which catalyze site specific recombination between two specific DNA sequences and mediate DNA integration, excision, resolution or inversion are called as site specific

recombinases. These site specific recombinases which play a pivotal role in the life cycles of many microorganisms are classified into two types; tyrosine type or serine type recombinases.

And this classification is based on whether a tyrosine or serine residue mediates the catalysis. They further divided or subdivided into two subfamilies each based on the following criterion. The tyrosine type combinations are subdivided by the direction of the recombination or recombination directionality while the serine type recombination are subdivided by the size of the protein.

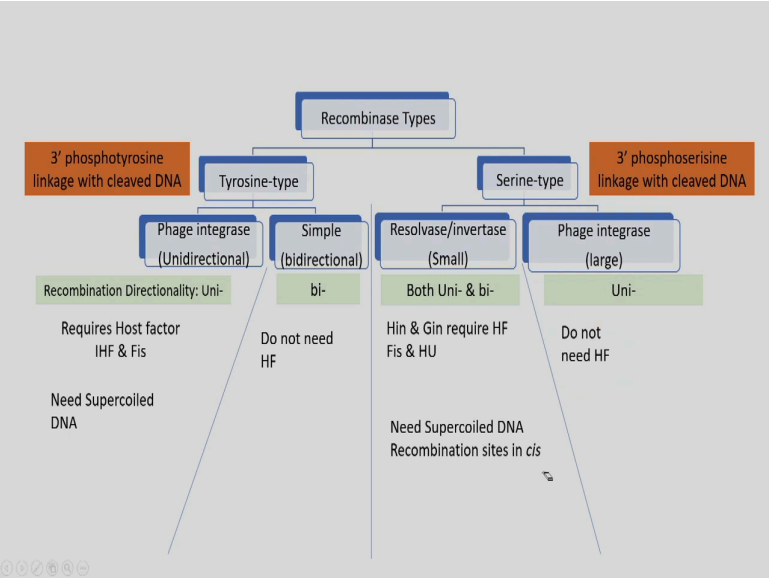
The tyrosine type recombinases catalyzes the formation of a Holliday junction intermediate and resolution. The catalytic mechanism of serine type recombination includes the 180 degree rotation and rejoining of cleaved substrate DNA.

(Refer Slide Time: 08:42)




And you can see here eponymous Holliday junction a model for DNA recombination intermediate proposed by doctor robin Holliday. And this is the simplified diagram where four strands of DNA participant and this is the cartoon diagram of the DNA double strands.

(Refer Slide Time: 09:11)



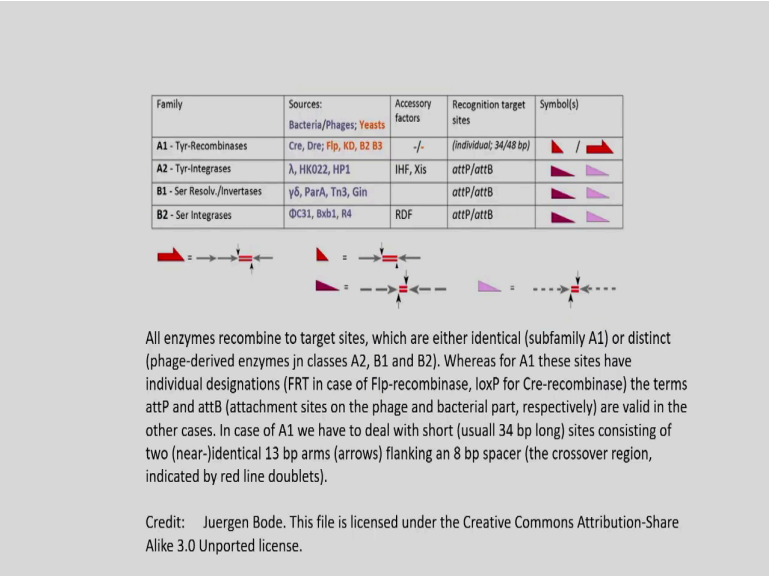

So, briefly we have up till now discussed about the different type of classification of the site specific recombinases. Broadly they fall into two types; tyrosine type and the serine type. And here the 3 prime phosphotyrosine linkage with cleaved DNA and here you have the 3 prime phosphoserisine linkage with cleaved DNA. And these tyrosine type are further subdivided into phase integrase which may be unidirectional.

And simple type which may be bi directional whereas, the serine types are divided depending on the size like the resolvase or invertase they may be small and the phase integrases which are very very large. And there are other features which are important to these recombinases.

For example, the phase integrase requires the host factors IHF and Fis and these needs super coiled DNA these do not need the HF and the serine types particularly the small resolvases and integrases also have some kind of directionality. They may be both uni and bidirectional and the requirements of HF may vary.

Hin and Gin required HF Fis and HU, they also need super coiled DNA and recombination sites in cis. Whereas, the large serine type phase integrases do not need any HF host factors.

(Refer Slide Time: 11:13)



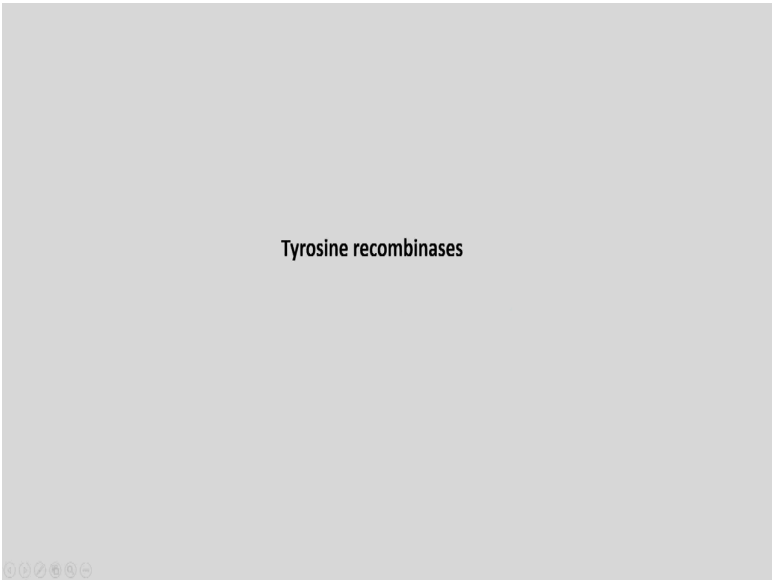

All enzymes recombine to target sites which are either identical or distinct whereas, for A1 these sites have individual designations FRT in case of Flp recombinases loxP for Cre recombinases the terms attP and attB are attachment sites on the phage and bacterial part respectively.

And these are valid in other cases as well in case of A1 we have to deal with short usual thirty-four base pair long sites consisting of two near identical thirteen base pair arms flanking an 8 base pair spacer the crossover region indicated by red line doublets over here.

So, let us look into this table to understand this discussion. We have the types of various recombinases tyrosine and serine you have tyrosine recombinases and tyrosine integrases, serine resolvases invertases or serine integrases. And these are the sources bacteria phage and yeast and you have the various members like Cre, Dre, Flp, KD you have for example, in the case of integrases lambda integrases in serine resolvases you have gamma delta ParA Tn3 Gin and so on.

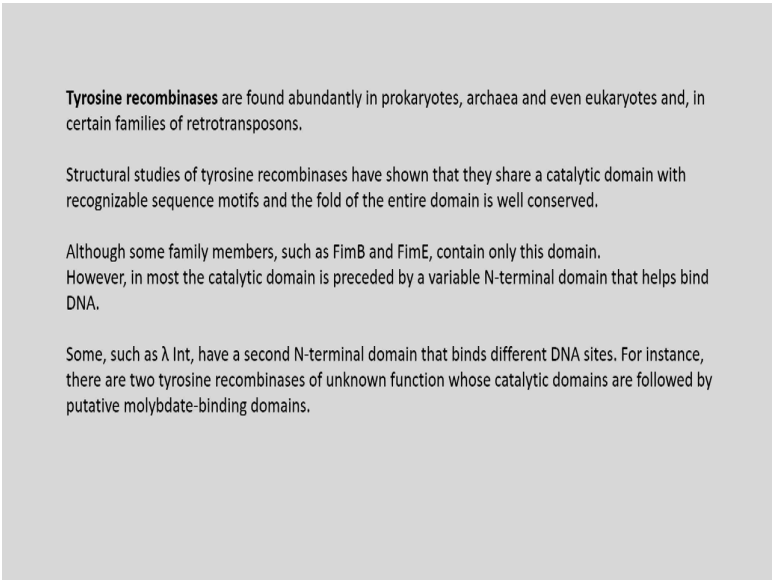
And in certain cases they require accessory factors already discussed earlier and some members do not and these are recognition sites in phage and in the bacteria as discussed earlier.

(Refer Slide Time: 13:22)




Let us first discuss about the features of tyrosine recombinases and what are the type of reactions they carry or the mechanisms involve.

(Refer Slide Time: 13:36)

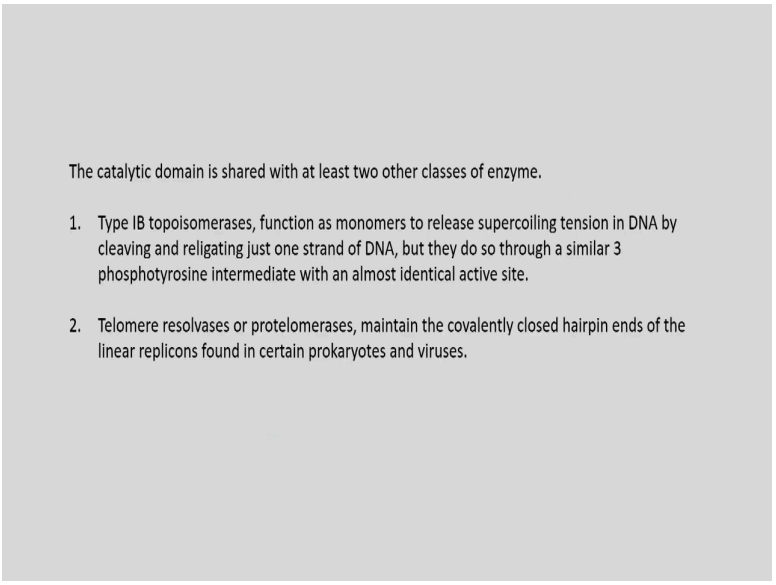



So, these tyrosine recombinases are found abundantly in prokaryotes, archaea and even eukaryotes and in certain families of retrotransposons. Various structural studies of these tyrosine recombinases have shown that they share a catalytic domain with recognizable sequence motives and the full of the entire domain is well conserved. Although some family

members such as FimB and FimE contain only this domain. However, in most the catalytic domain is preceded by a variable N terminal domain that helps binding to DNA.

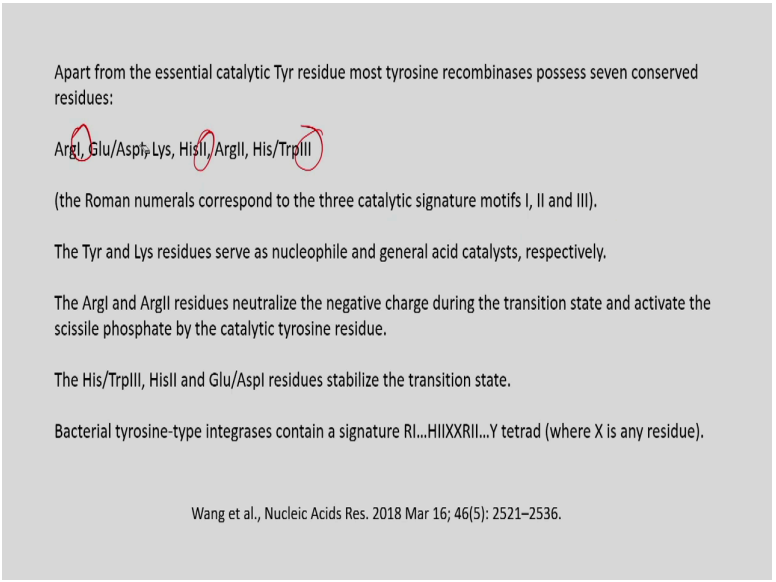
Some such is lambda intregases have a second n terminal domain that binds different DNA sites. For instance there are two tyrosine recombinases of unknown function whose catalytic domains are followed by putative molybdite binding domains.

(Refer Slide Time: 14:47)




The catalytic domain is shared with at least two other classes of enzyme; type 1B isomerases function as monomers to release super coiling tension in DNA by cleaving and religating in just one strand of DNA, but they do so through a similar 3 phosphotyrosine intermediate with an almost identical active site. The second one is the telomere resolvases or protelomerases maintain the covalently closed hair pin ends of the linear replications found in certain prokaryotes and viruses.

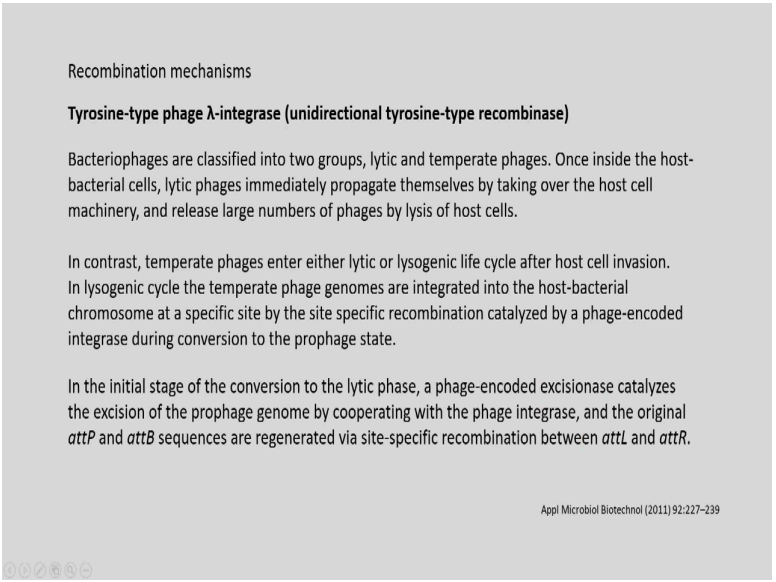
(Refer Slide Time: 15:17)




Apart from the essential catalytic tyrosine residue most tyrosine recombinases possess seven conserved residues as listed here. ArgI, Glu AspI, Lys histidine II, ArgII, His and Trypsin 3 tryptophan 3. These roman numerals here I, II and III correspond to the three catalytic motifs I, II and III. The tyrosine and lysine residue serve as nucleophilic as a nucleophile and general acid catalyst respectively.

The ArgI and ArgII residues neutralize the negative charge during the transition state and activate the scissile phosphate by the catalytic tyrosine residue. The His Trp III and His II and Glu Asp I residues stabilizes the transition state bacterial tyrosine type integrases continuous signature RI HIIXXRII Y tetrad where X can be any residue.

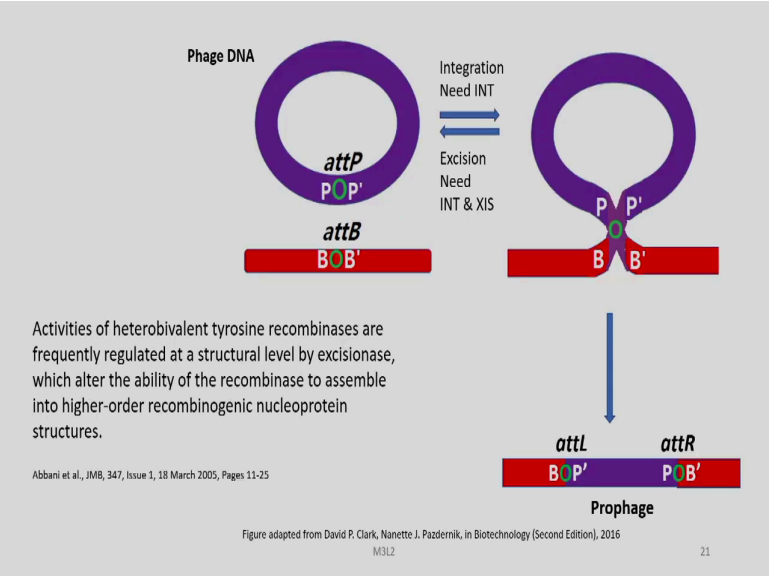
(Refer Slide Time: 16:39)




Now, what is the mechanism by which this recombination happen? Tyrosine type first lambda integrase which is a unidirectional tyrosine type recombinase. Let us study its mechanism first. Bacteriophages are classified into two groups lytic and temperate phages once inside the host bacterial cells lytic phages immediately propagate themselves by taking over the host cell machinery and release large number of phages by lysis of four cells. In contrast to these the temperate phages enter either lytic or lysogenic life cycle after host cell invasion.

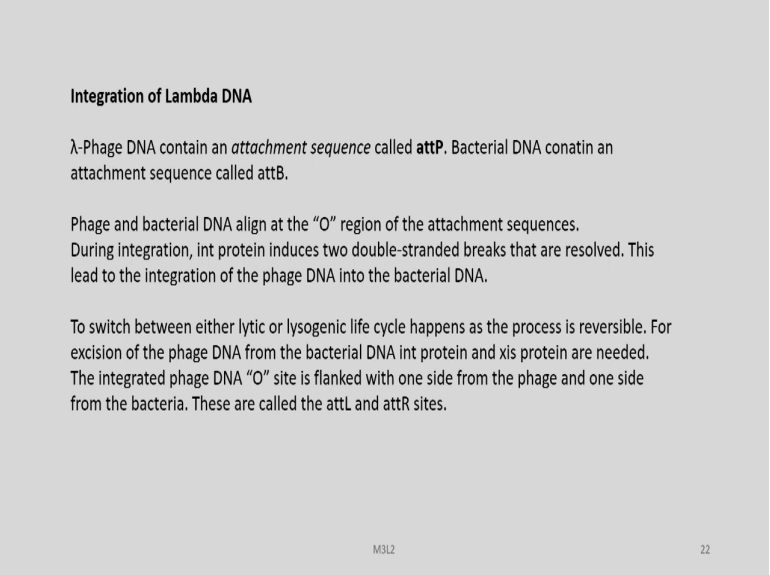
In lysogenic cycle the temperate phages genomes are integrated in the host bacterial chromosome at a specific site by the site specific recombination catalyzed by a phage encoded integrase during conversion to the prophage state. In the initial stage of the conversion to the lytic phage a phage encoded excisionase is catalyzes the excision of the prophage genome by cooperating with the phage integrase n the original attP and attB sequences are regenerated via site specific recombination between attL and attR. So, the integrase and the excisionase carry out opposite functions.

(Refer Slide Time: 18:22)




We have discussed about the attP and attB sequences earlier. So, you can see this location of these sequences over here. Now in the case of integration you need integrate and excision would require integrases as well as the excisionase and these gives rise to the prophase. The activities of hetero bifunctional tyrosine recombinases are frequently regulated at the structural level by excisionase which alter the ability of the recombinase to assemble into higher order recombinogenic nuclear protein structures.

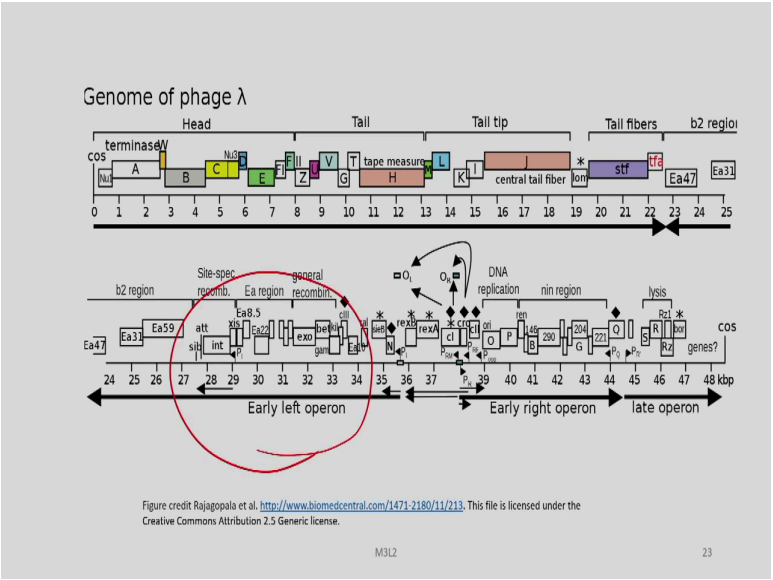
(Refer Slide Time: 19:09)



Integration of the lambda DNA. Lambda phage DNA contain an attachment sequence called attP as shown in this figure and the bacterial DNA contain an attachment sequence called attB. Phage and bacterial DNA align at the origin of the attachment sequences during integration, int protein induces two double strand breaks that are resolved. This lead to the integration of the phage DNA into the bacterial DNA.

To switch between either lytic or lysogenic life cycle the process as it happens is reversible. For excision of prophase phage DNA from the bacterial DNA int protein and xis protein is needed as already discussed earlier. The integrated phage DNA O site is flanked with one side from the phage and one side from the bacteria. These are called the attL and attR sites.

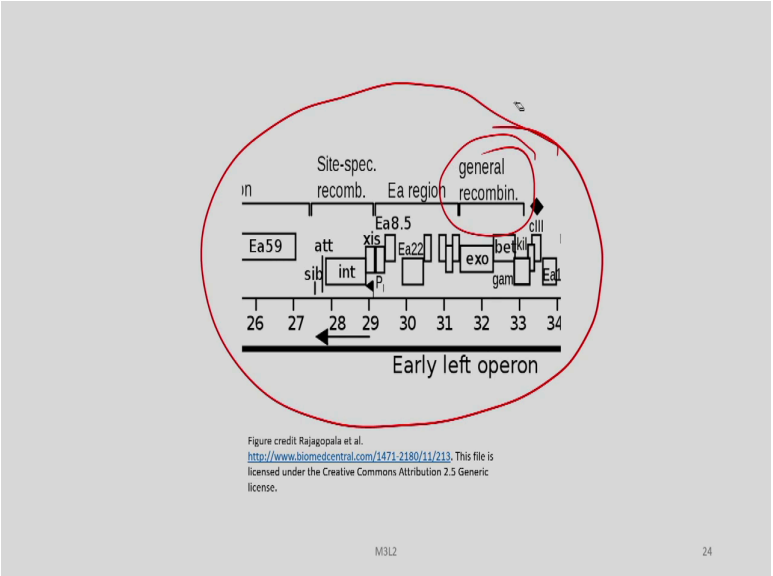
(Refer Slide Time: 20:14)




And we can see here the genome of the phage lambda which has certain sequences corresponding to the head, tail, tail tip and tail fibres. And these are not expressed together these are some early genes and some are late genes accordingly they are known as the early operon or the late operon and those which are located on the left side of the genome are the early left operon.

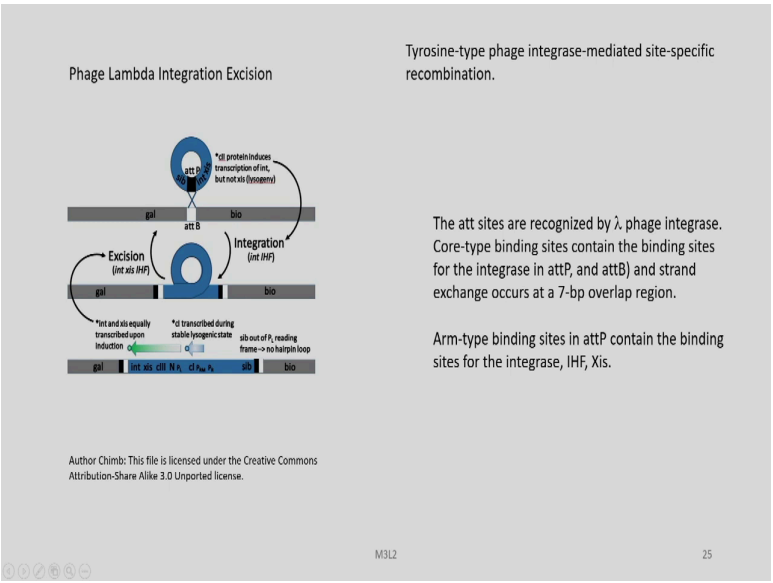
And those which are expressed towards the end are the late operon. And here you can see the location of the site specific recombination and the att sites as well as the integrase and the xis. So, these maps give us an idea that all these elements required for site specific recombination are under the early left operon.

(Refer Slide Time: 21:42)




So, this is a magnified view of this diagram showing the various components here as we have already discussed as well as the elements or sequences which are involved in the general recombination.

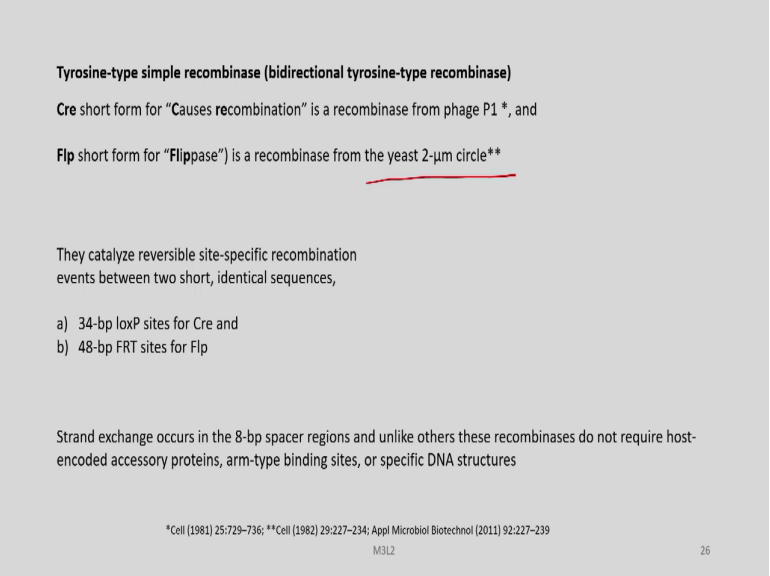
(Refer Slide Time: 22:00)




How does the phage lambda get integrated? So, or how does the excision of the phage lambda takes place? So, in the first case we can see the integration is happening and in the second case the excision is occurring. The tyrosine type phage integrase mediate site specific recombination the att sites are recognized by phage integrase. Core type binding sites

containing the binding sites for the integrase in attP and attB and strand exchange occurs at a 7 base pair overlap region. Arm type binding sites in attP contain the binding sites for the integrase IHF and Xis ionase. Let us now discuss about the bidirectional tyrosine type recombinases.

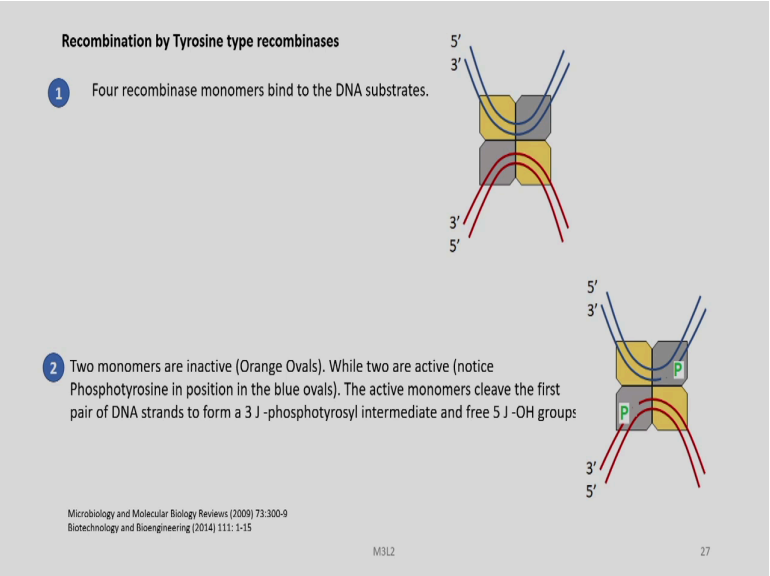
(Refer Slide Time: 23:14)




You can see here the term causes recombination and the C and r e re are combined together to form Cre. So, Cre is short form for causes recombination and this Cre is a recombination enzyme or recombination enzyme from phage P1. Similarly the Flp is a short form from Flippase and this is recombinases from the yeast 2 micron plasmid cycle. Both of these catalyzed reversible site specific recombination events between two short identical sequences.

And these sequences are 34 base pairs loxP sites for Cre and 48 base pairs FRT sites for Flp. Strand exchange occurs in the 8 base pair regions spacer regions and unlike others these recombination do not require host encoded accessory proteins arm type binding sites or specific DNA structures.

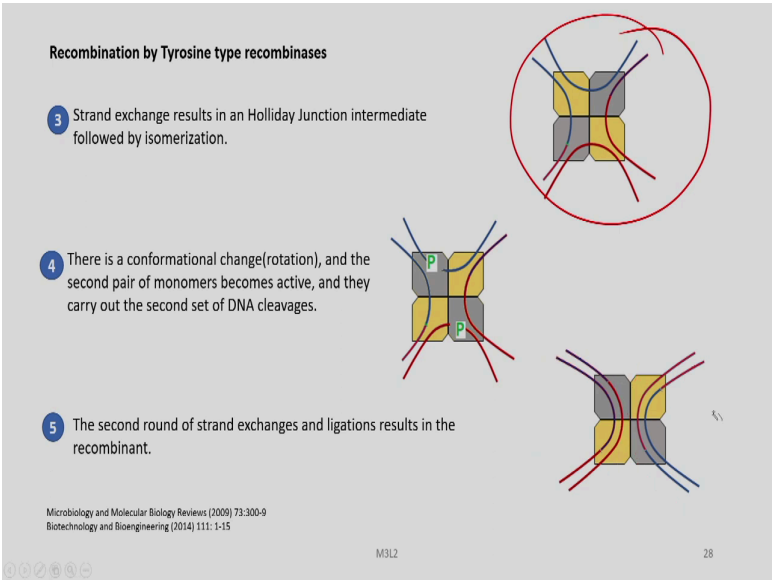
(Refer Slide Time: 24:25)




How does the recombination by tyrosine type recombinases occur? So, in the first step four recombinase monomers bind to the DNA substrates. So, this is a DNA substrate here another DNA substrate here and you can see here the monomers 1, 2, 3, 4 which together bind to the DNA substrates as shown in the picture.

Two of these monomers are inactive these l ones for example, are inactive while two are active ok. You can notice here phosphotyrosine in position in the blue ovals here which is appearing little bit greyish the active monomers cleave the first pair of DNA strands to form a 3 J phosphortyrosil intermediates and free 5 J hydroxyl groups.

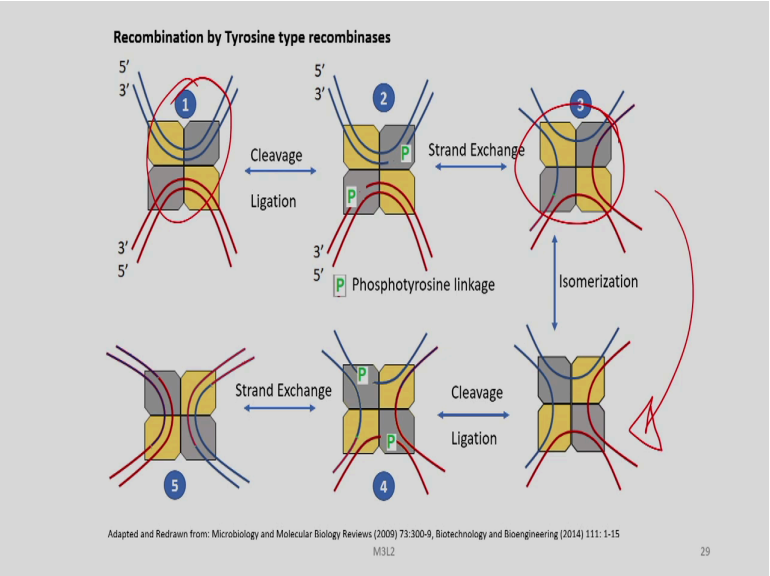
(Refer Slide Time: 25:41)




Now, after these there is a strand exchange which results in Holliday junction intermediate as you can see here and this is followed by isomerisation. Now there is a conformational change in the next stage or a rotation takes place. And the second pair of monomers which were earlier inactive now becomes active and they carry out the second set of DNA cleavage.

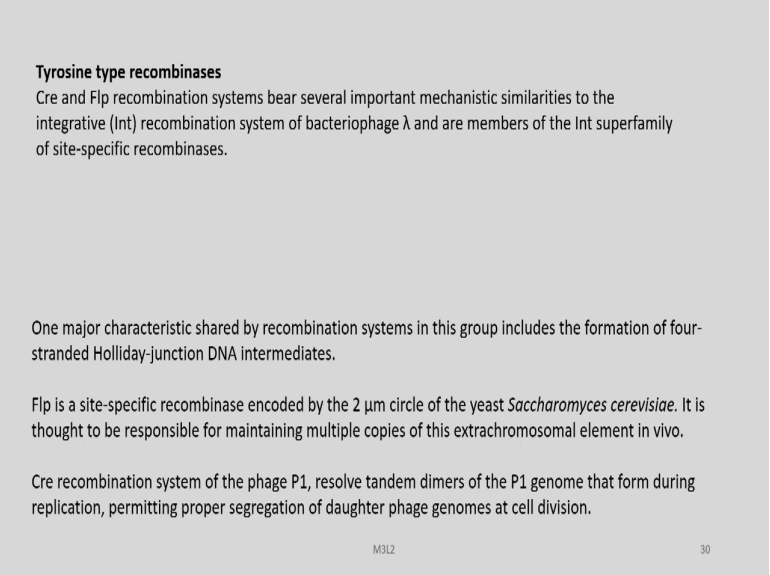
The second round of strand exchanges in ligations result in the recombinants. Now you can see here a molecule two different molecules DNA molecules painted blue and red after recombination process becomes a hybrid molecule as shown by these colours having both blue strand and the red strand. And both the products are now recombinant products.

(Refer Slide Time: 26:48)




Recombination by tyrosine type recombinases. So, here the two strands binds here as already told to you and then there is a strand exchange which forms the Holliday junction. And this is the isomerization step which we could not see in the earlier pictures and finally, the second cleavage occurs and then finally, a recombinant DNA products are produced. So, this is basically as you can see a five step procedure or six step procedure including the isomerization step.

(Refer Slide Time: 27:37)

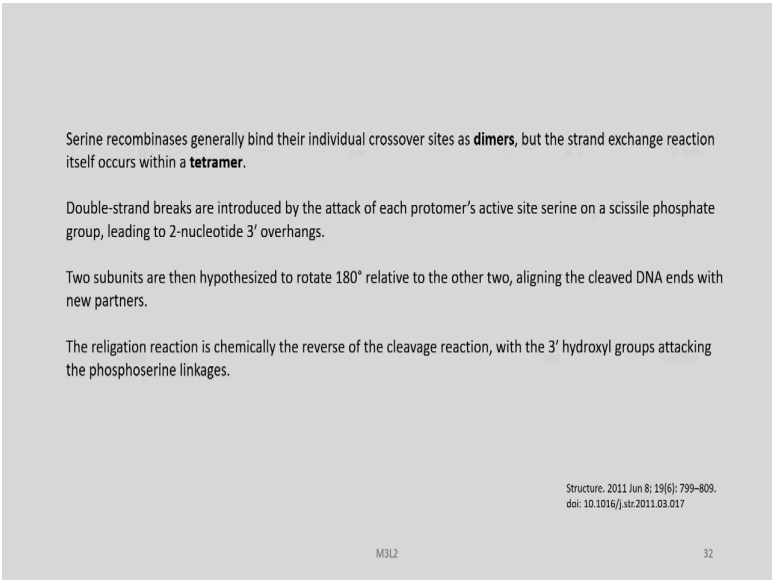


Cre and Flp recombination systems bear several important mechanistic similarities to the integrative recombination system of bacteriophage lambda and are members of the Int super family of site specific recombinases. One major characteristics shared by recombination systems in this group includes the formation of four stranded Holliday junction DNA intermediates.

Flp is a site specific recombinase encoded by the 2 micron circle of the yeast saccharomyces cerevisiae. It is thought to be responsible for maintaining multiple copies of these extra chromosomal elements in vivo. Cre recombination system of the phage 1 phage P1 resolve tandem dimers of the P1 genome that form during replication permitting the proper segregation of daughter phage genomes at a cell division.

(Refer Slide Time: 28:37)

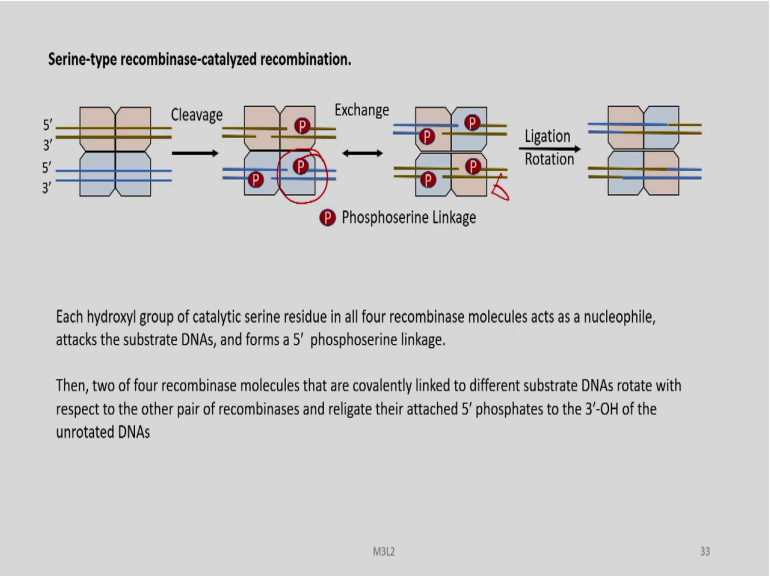
(Refer Slide Time: 28:46)




Let us now discuss about the serine recombinases. The serine recombinases generally bind their individual crossover sides as dimers, but the strand exchange reaction itself occurs within a tetramer. Double strand breaks are introduced by the attack of each promoters active site serine on a scissile phosphate group leading to 2 nucleotide 3 prime overhangs. Two subunits are then hypothesized to rotate 180 degree relative to the each others aligning the cleaved DNA ends with new partners.

The religation reaction is chemically the reverse of the cleavage reaction with the 3 prime hydroxyl groups attacking the phosphoserine linkages.

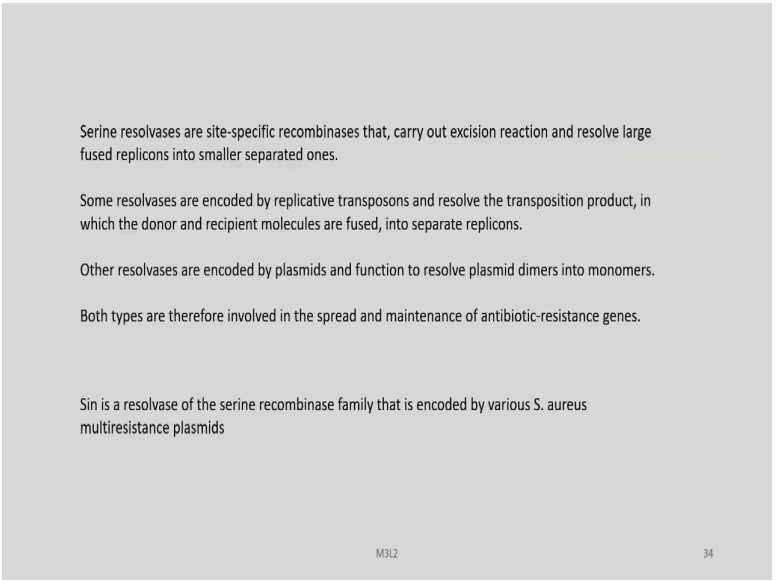
(Refer Slide Time: 29:26)




You can see here the serine type recombination catalyzed reaction. So, you have four units and they are similarly binding two different DNA molecules. And then two of these are on the top and bottom and they bind to separate DNA strands. Each hydroxyl group of catalytic serine residue in all four recombinase molecules acts as a nucleophile. And it attacks the substrate DNAs and forms 5 prime phosphoserine linkages as shown over here. And then ligation and rotation takes place and then the products are formed.

Two of the four recombinase molecules that are covalently linked to different substrate DNAs rotate you can see here there is a rotation. So, these one of these blue has moved up while the pink colour unit has come down. The DNA has rotated with respect to the other pair of the recombinases and religate their attached 5 prime phosphates to the 3 prime of the unrotated DNA.

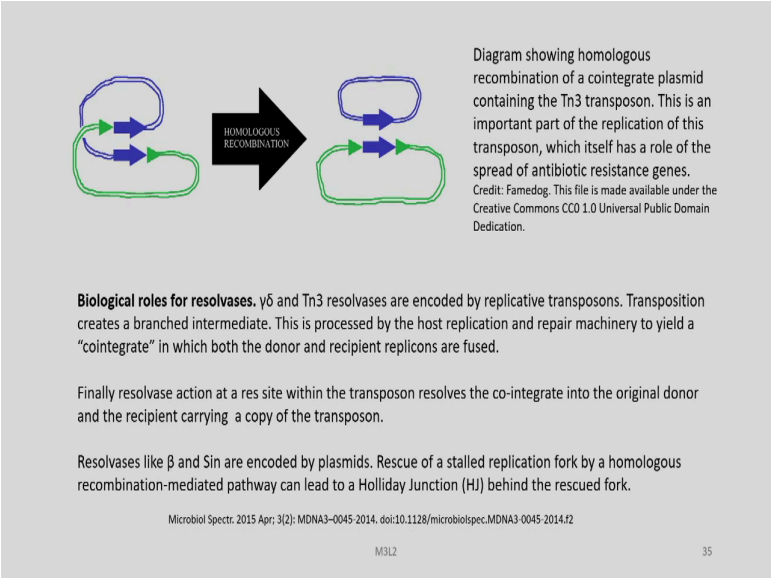
(Refer Slide Time: 31:00)




So, this is a little bit different from the earlier case. Here there is no any formation of the Holliday junction. Serine resolvases are site specific recombinases that carry out excision reaction and resolve large fused replicons into smaller separated ones. Some resolvases are encoded by replicative transposons and resolve the transposition product in which the donor and recipient molecules are fused into separate replicons.

Other resolvases are encoded by plasmids and function to resolve plasmid dimers into monomers. Both types are therefore, involved in the spread and maintenance of antibiotic resistant genes. Sin as a resolves of the serene recombinase family that is encoded by various staphylococcus aureus multi resistance plasmids.

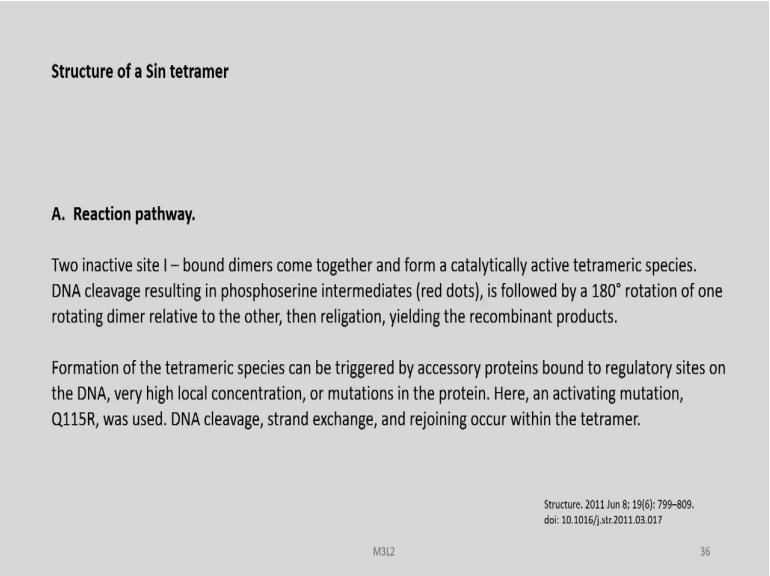
(Refer Slide Time: 31:48)




Now, what are the biological rules of resolvases? Gamma delta and Tn3 resolvases are encoded by replicative transposons. Transposition creates a branched intermediate. This is processed by the host replication and repair machinery to yield a co integrate in which both the donor and recipient replications are replicons are fused. Finally, resolvase action at the res site within the transposon transposon resolves the co integrate into the original donor and the recipient carrying a copy of the transposon.

Resolvases like beta and Sin are encoded by plasmids. Rescue of a stall replication followed by a homologous recombination mediate pathway can lead to a Holliday junction behind the rescued fork. In this diagram you can see homologous recombination of a co integrate plasmid containing the T 3 Tn3 transposon. This is an important part of the replication of this transposon which itself has a role of the spread of antibiotic resistant genes.

(Refer Slide Time: 32:57)



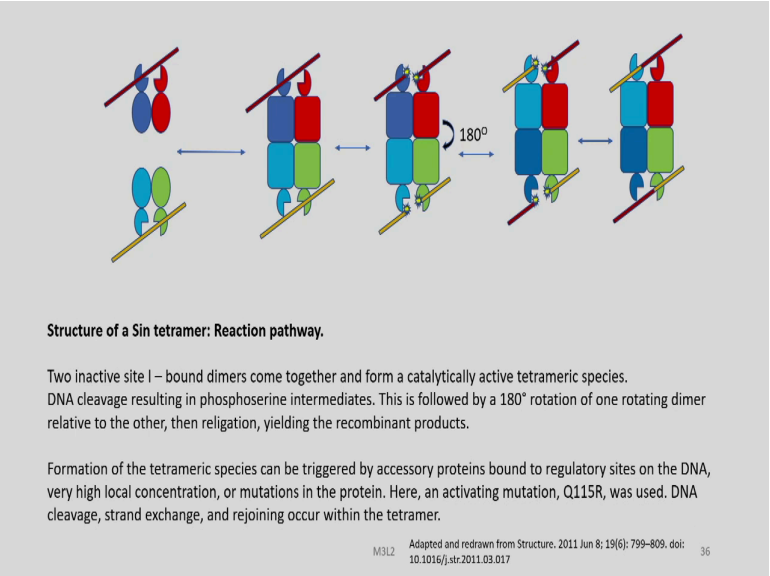

Let us see the structure of a sin tetramer and in the reaction pathway there are two inactive site one which bound and which are bound dimers and come together and form a catalytically active tetrameric species. DNA cleavage resulting in phosphorserine intermediates is followed by 180 degree rotation of 1 rotating dimer relative to the other. Then religation happens yielding the recombinant products. Formation of the tetrameric species can be triggered by a accessory protein bound to regulatory sites on the DNA.

Let us study about the biological roles of resolvases. Gamma, delta and Tn3 resolvases are encoded by replicative transposons. Transposition creates a branched intermediate this is processed by the host replication and repair machinery to yield a co integrate in which both the donor and the recipient replications replicons are fused.

Finally resolvase action take place at the res site with the transposon resolve within the transposon resolves the co integrate into the original donor and recipient carrying a copy of the transposon. So, in this diagram you can see the homologous recombination of a co integrate plasmid containing Tn3 transposon. This is an important part of the replication of this transposon which itself has a role on the spread of antibiotic resistance genes.

Resolvases like beta and Sin are encoded by plasmids rescue of a stalled replication fork by a homologous recombination mediated pathway can lead to a Holliday junction behind the rescued fork.

(Refer Slide Time: 34:55)

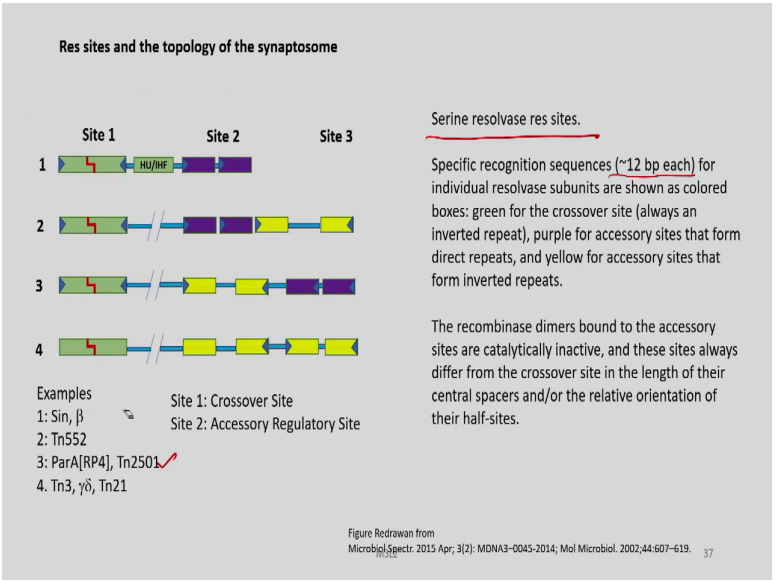



Let us look into the structure of a Sin tetramer and the reaction pathway. Here there are two inactive site 1 bound dimers which come together and form a catalytically active tetrameric species. So, this is inactive and these are also inactive when they come together they form a catalytically active tetrameric species. DNA cleavage resulting in phosphoserine intermediates this is followed by 180 degree rotation of 1 rotating dimer or relative to the other. And then religation happens and it yields the recombinant products.

So, you can see here the cleavage occurring and then there is a 180 degree rotation. So, these goes up while this comes down over here and due to these rotation you can see here the differently coloured DNA molecule which represents a different DNA molecule is flipped and due to ligation it is joined with the other DNA molecule.

So, this is one DNA molecule denoted by one colour, this is another DNA molecule denoted by another colour and here you can get a recombinant product as denoted by the presence of both the coloured fragments of DNA and in both the products.

(Refer Slide Time: 36:55)

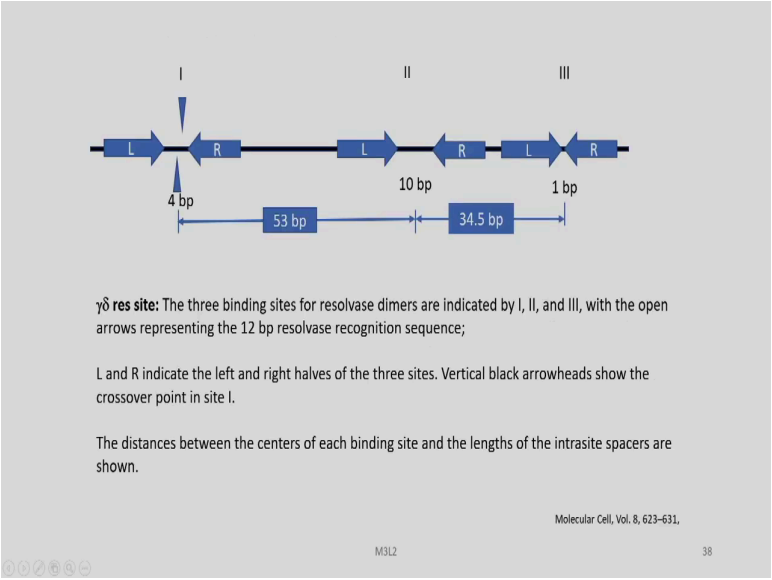



Res sites and topology of the synaptosome. Serine resolvases has res sites. Specific recognition sequences around 12 base pair each for individual resolvase subunits are shown as coloured boxes. This green is for the crossover side always an inverted repeat. Purple for accessory site that form direct repeats and yellow for accessory sites that form inverted repeats.

The recombinase dimers bound to the accessory sites are catalytically inactive and these sites always differ from the crossover site in the length of their central spaces and relative orientation of their half sides. So, site 1 is basically the crossover site and site 2 is the accessory or the regulatory site. And there is a site 3 as well at the end of site 2. So, there are various examples for this particular pattern you have sin and beta.

For the second type you have Tn552 then for the third type you have ParA and then T 2501 and the fourth type you have Tn3, gamma delta and Tn21.

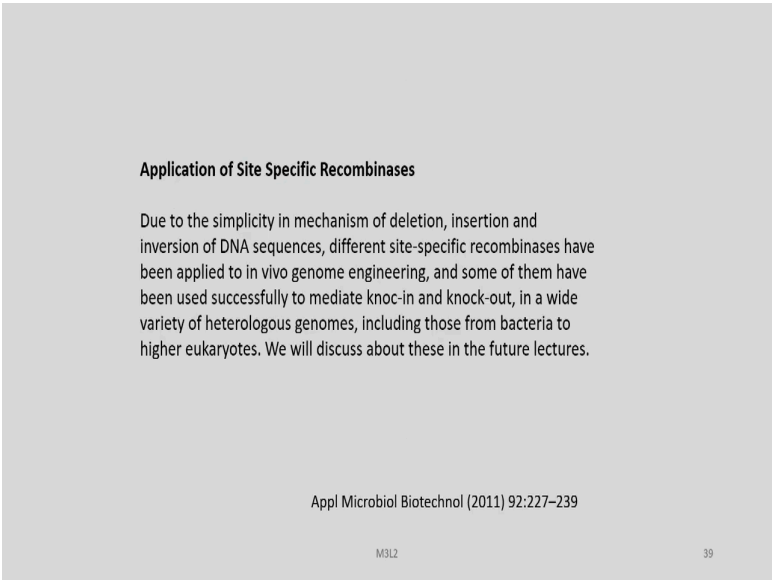
(Refer Slide Time: 38:36)




The gamma delta res site. So, you have site 1, site 2 and site 3 as shown earlier. The three binding sites for resolvase dimers are indicated by 1, 2 and 3 with the open arrows representing the twelve base pair resolvase recognition sequence. This L and R indicate the left and right hubs of the three sides.

So, for first site L and R second site also L and R and third also L and R. The vertical arrows show the cross over point at site 1 here. The distance between the centres of each binding site and the length of the intersect spaces are also shown in this figure around 53 base pairs here and 34.5 base pairs.

(Refer Slide Time: 39:42)




What are the various application of site specific recombinases? Due to the simplicity in mechanism of deletion insertion and inversion of the DNA sequences different site specific recombinases have been applied to in vivo genome engineering. And in some of them it has been used successfully to mediate knock in and knockout.

We will be discussing about the use of some of these site specific recombinases in our later lectures for producing knock in and knockout organisms. And in a wide variety of heterologous genomes including those from bacteria to higher organisms.

Thank you for your patient hearing.