

CHAPTER 1

General Introduction

CHAPTER 1

General Introduction

1.1 Site-specific recombination

Genetic recombination systems function in a wide range of programmed DNA rearrangements in microbes as well as in other organisms. The biological functions of DNA, including regulation of gene expression, DNA replication, and site-specific recombination often require two or more segments of DNA or other segments in addition to DNA to communicate with each other. For example, proteins act as bridges between distant sites on the same or different molecules of DNA (Parker and Halford, 1991).

Site-specific recombination is an important developmental process in both prokaryotes and eukaryotes. It includes alteration of gene expression patterns, part of the transposition pathway of certain transposable elements, integration and excision of bacteriophages into and out of the bacterial chromosome, control of plasmid cellular copy number and normal segregation of bacterial chromosomes during cell division (reviewed by Sadowski, 1986; Stark *et al.*, 1992). Site-specific recombination requires pairing of two recombination sites (synapsis), four strand scissions, coordinated rearrangement of the eight broken ends and four ligation events (rejoining). Rearrangement of the broken ends at the two recombination sites results in recombination. High-energy cofactors like ATP are not required to resynthesize the phosphodiester bonds that break.

Site-specific recombination systems require recombination enzymes (recombinases) which catalyze the cutting and rejoining of the participating DNA molecules. Some recombination systems also require accessory factors such as integration host factor (IHF), aminopeptidase A (PepA) and arginine repressor (ArgR). The recombinases and accessory proteins form a highly organized protein-DNA complex with the recombination site. All recombinases known show amino acid sequence similarities to either the *Tn3* resolvase/ invertase family or the λ integrase (Int) family of site-specific recombinases (reviewed by Sadowski, 1986; Stark *et al.*, 1992).

Recombinases belonging to the Int family, which show homology with the integrase of phage lambda can catalyze inversion, deletion and formation of cointegrates (an intermolecular reaction) with nearly equal frequencies (Figure 1.1). Recombinases which belong to the *Tn3* resolvase/ invertase family include both DNA resolvases and DNA invertases. Resolvase are highly specialized in catalyzing deletions between two directly oriented repeats harbouring a *res* site, and are very inefficient at making inversions when the repeated sequences are in inverse orientation. DNA invertases have the opposite characteristic: in the presence of an enhancer sequence and a host factor, they are efficient at promoting inversions between two inversely oriented repeats, but deletions between two directly oriented repeats occur only at a very low frequency.

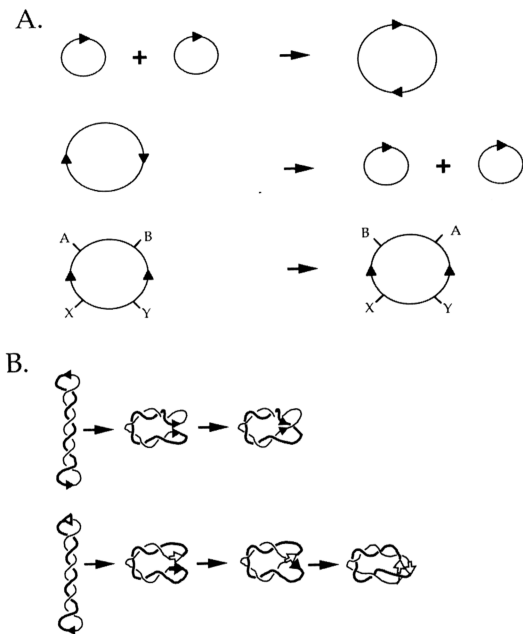


Figure 1.1. Physical consequences of site-specific recombination. Recombination sites are shown as arrows whose direction indicates the orientation of recombination sites. **Panel A:** (Top) Intermolecular and intramolecular recombination. Intermolecular recombination results in the fusion of the substrate DNA circles to form a single product circle. (Middle) Intramolecular recombination of a substrate circle containing directly repeated sites results in the formation of two product circles; this reaction is variously called resolution, deletion, or excision. (Bottom) Intramolecular recombination of a substrate circle containing inverted sites results in the inversion of the DNA segment between the sites. **Panel B:** Intramolecular recombination on supercoiled DNA. A negative supercoiled substrate is shown with the arrows representing the recombination sites. The domains between the sites are represented as thick and thin lines. (Top) A negative supercoiled substrate containing directly repeated sites is shown. The recombination sites are juxtaposed by random collision, which entraps the interwound substrate supercoils so that the product circles are multiply linked catenanes. (Bottom) A negative supercoiled inversion substrate is shown. The sites are synapsed and recombination occurs. Six supercoils are trapped in the synaptic complex. During the process of strand exchange, a positive interdomainal node is introduced. The recombination product is a seven-noded knot. The diagrams shown in A and B (top) were taken from Craig (1988) while that of B (bottom) from Kanaar and van de Putte (1987).

1.2 Xer site-specific recombination system

Xer is an acronym for chromosomal ColE1 recombination functions. Xer site-specific recombination is employed in two related biological processes: the stable inheritance of multicopy plasmids and the normal partitioning of the *E. coli* chromosome to daughter cells at cell division.

The Xer site-specific recombination system utilized by the multicopy plasmid ColE1 and its natural relatives promote rapid recombination between two *cer* sites on dimeric molecules, resolving them into monomeric substrates during cell division thus ensuring stable plasmid inheritance (Summers and Sherratt, 1984; Sherratt, 1993). The *cer*, which is a 250 bp DNA region, acts only in *cis* to resolve plasmid multimers (generated from *recA*-dependent homologous recombination) to monomers, thus maximizing copy number and stability (Figure 1.2). Xer site-specific recombination at *cer* acts at approximately 30 bp core recombination site and also requires approximately 200 bp of accessory DNA sequences adjacent to the core site (Figure 1.3; Summers and Sherratt, 1988; Stirling *et al.*, 1988). Four *E. coli* proteins are required for Xer site-specific recombination at *cer*: the recombinases, XerC and XerD and the accessory proteins, ArgR and PepA. In addition to *cer*, plasmid pSC101 specifies a recombination site, *psi*, that is important in resolving plasmid multimers. Recombination at *psi* does not require ArgR, but requires XerC, XerD, PepA and accessory DNA sequences (Figure 1.3).

Xer site-specific recombination

- stable inheritance of multicopy plasmids
- normal partition of the *E. coli* chromosome to daughter cells at cell division

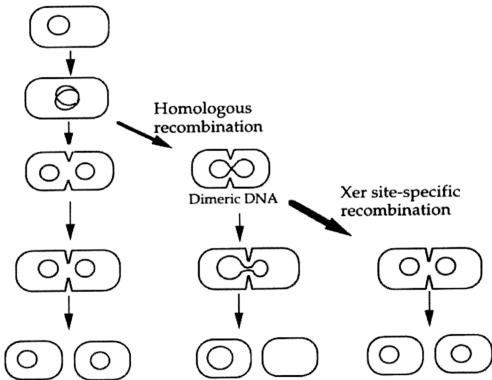
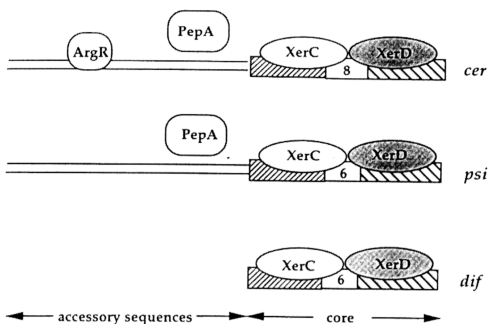


Figure 1.2. Diagrammatic representation of the role of Xer site-specific recombination system in the segregation of multicopy plasmids. During the course of normal replication and partition (left pathway), the plasmids are properly segregated into daughter cells. Host homologous recombination system tend to make dimers (and other multimers) from plasmid copies, thus limits the availability of separate molecules for partition. This results in plasmid free cell (middle pathway). Xer site-specific recombination system that act on *cer* and related sites converts any dimers formed into monomers, thus allowing proper segregation into daughter cells (right pathway). For simplicity, only two plasmid molecules are shown while the host chromosome is not shown.



	XerC	central region	XerD
ColE1 <i>cer</i>	GGTGCGTACAA	TTAAGGGA	TTATGGTAAAT
CloDF13 <i>parB</i>	GGTACCGATAA	--GGGATG	TTATGGTAAAT
<i>cer</i> 8-1	GGTGCGTACAA	TTGGGATG	TTATGGTAAAT
(type I hybrid)			
<i>cer</i> 6-1	GGTGCGTACAA	--GGGATG	TTATGGTAAAT
(type II hybrid)			
pNTP16 <i>nmr</i>	GGTGCGCGTAA	-TGAGACG	TTATGGTAAAT
pSC101 <i>psi</i>	GGTGCGCGCAA	--GATCCA	TTATGTTAAAC
<i>E. coli dif</i>	GGTGCGCATAA	--TGTATA	TTATGTTAAAT
	-8	1	8
consensus	TG GTAT GG C AA GA CGGC		G T TTATG TAA T C

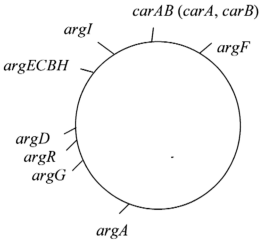
Figure 1.3. Xer site-specific recombination sites. Panel A: An illustration of the *cer*, *psi*, and *dif* recombination sites. Panel B: Xer core recombination sites. The sequence of several core recombination sites in Xer site-specific recombination system is shown. The XerC and XerD binding sites are separated by 6-, 7-, or 8-bp central region of variable sequences. Base pair coordinates for *dif* are indicated. Recombination at *dif* does not require accessory DNA sequences and the accessory proteins ArgR and PepA.

The Xer site-specific recombination system is also utilized by the *E. coli* chromosome to convert chromosomal dimers (arising from homologous recombination during or after chromosome replication) into segregable chromosome monomers. Recombination at *dif*, a 33 bp recombination locus at the replication terminus of the *E. coli* chromosome requires XerC and XerD but is independent of ArgR and PepA (Figure 1.3) (Blakely *et al.*, 1993). Recombination at *dif* occurs intramolecularly and intermolecularly (Figure 1.1) (Blakely *et al.*, 1991; Blakely *et al.*, 1993). It differs from recombination at *cer* and *psi*, in which the recombination at these two sites occur only intramolecularly (Stirling *et al.*, 1988; Colloms *et al.*, 1998).

1.3 Arginine repressor of *Escherichia coli* K-12 (ArgR)

Arginine repressor of *E. coli* K-12 (ArgR) has two known disparate functions. ArgR is the product of the gene *argR*, located on the *E. coli* K-12 chromosome. ArgR, in the presence of its co-repressor ligand, L-arginine, controls the synthesis of the ten enzymes of arginine biosynthesis and also its own synthesis (Lim *et al.*, 1987). The term “regulon” was coined to describe the arginine pathway, in which a single repressor controls the expression of separate genes (Maas and Clark, 1964). The eight structural genes of the arginine biosynthetic pathway are scattered in five regions of the *E. coli* chromosome (Figure 1.4; Maas, 1994; Cunin *et al.*, 1986; Jacoby and Gorini, 1969). The genes are organized into nine transcriptional units, of which the ArgR binding sites (ARG boxes, see below) overlaps the promoters (reviewed by Maas, 1994). It was shown that for each of the arginine transcription units, there is at least one

A.



B.

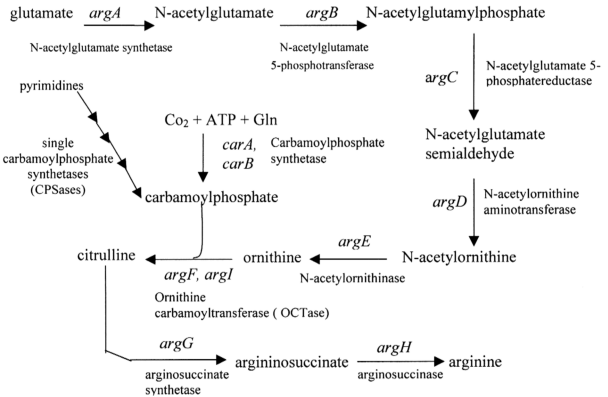


Figure 1.4. Organization of *arg* genes in *E. coli* K-12. Panel A: Chromosomal location of the genes of the arginine regulon. **Panel B:** The arginine biosynthetic pathway. The ten enzymes with the nine transcriptional genes are shown. Note that the synthesis of CPSases is controlled by arginine and pyrimidine (Cunin *et al.*, 1986; Maas, 1994).

control region, consisting of two 18-bp imperfect palindromic sequences, called ARG boxes. Each ARG box is separated by a variable spacer of two or three base pairs and overlapping the promoter region (Figure 1.5; Cunin *et al.*, 1983).

In addition to its role in arginine biosynthesis, ArgR plays an important role as an accessory factor in Xer site-specific recombination at *cer* and related sites present in multicopy plasmids (Stirling *et al.*, 1988a; Burke *et al.*, 1994; Chen *et al.*, 1997). Genetics experiments have indicated that the role of ArgR in site-specific recombination at *cer* is in ensuring that recombination is preferentially intramolecular (Summers, 1989). ArgR (and PepA) is implicated as accessory proteins that are required for resolution selectivity (Summers, 1989), the process that ensures that Xer recombination at *cer* is intramolecular (McCulloch *et al.*, 1994). It is also believed that ArgR (and PepA) have a structural role in recombinational synapse formation and that specific interactions between these proteins and/or the recombinases are essential steps in the Xer recombination reaction (Merican, 1995).

A single ArgR subunit contains 156 amino acids (16.5 kDa) and is composed of a basic N-terminal domain (residues 1-79), which is responsible for DNA binding site and an acidic C-terminal domain (residues 80-156), which is responsible for oligomerization and L-arginine binding (Burke *et al.*, 1994; Van Duyne *et al.*, 1996). It is thought to bind its target sequences as a 98 kDa hexamer (Lim *et al.*, 1987). Mutational analysis have implicated residues in the N-terminal

domain for DNA binding while the C-terminal domain is involved in L-arginine binding and oligomerization (Burke *et al.*, 1994; Tian and Maas, 1994).

The structure of the oligomerization and L-arginine binding domain of ArgR has been determined using X-ray diffraction methods at 2.2 Å resolution with bound L-arginine and at 2.8 Å in the unliganded form (Figure 1.6; Van Duyne *et al.*, 1996). ArgR hexamer is formed from two trimers, each with tightly packed hydrophobic cores. Each subunit has an α/β fold containing a four-stranded antiparallel β -sheet and two antiparallel α -helices. Six molecules of L-arginine bind at the trimer-trimer interface, each making ten hydrogen bonds to the protein including a direct ion pair that crosslink the two protein trimers.

ArgR is a DNA binding protein which binds to DNA targets of 18 bp consensus sequences (ARG box) that consist of related 9 bp dyad symmetrical elements preceding the L-arginine biosynthetic genes. Regulation of L-arginine biosynthesis in *E. coli* is determined by the affinity of ArgR binding to the ARG boxes (Charlier *et al.*, 1992; Tian *et al.*, 1992). The affinities for ARG boxes, ranges from 10^{-10} to 10^{-9} M, with the *argF* operator DNA having the strongest affinity (reviewed by Maas, 1994). Binding of wild-type ArgR to the *argF* operator DNA gives a significant DNA bending about 70° as reported by Tian *et al.* (1992) and about 90° as reported by Burke *et al.* (1994).

Hexameric ArgR binds symmetrically and simultaneously to four helical turns of the DNA on both the major and minor DNA grooves, covering adjacent

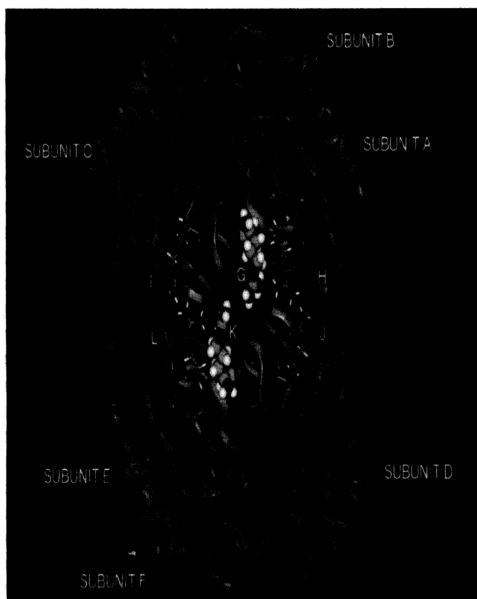


Figure 1.6. ArgRc-L-arginine structure. Six ArgRc subunits, A, B, C, D, E, and F; and six L-arginine molecules G, H, I, J, K and L. The ArgRc-L-arginine binding area is highlighted with the purple ribbon showing the interaction of L-arginine molecules G and K with protein subunits A, C, D, and E (Van Duyne *et al.*, 1996).

pairs of ARG boxes which are separated by three bp, but the binding is only to one face of the DNA (Charlier *et al.*, 1992). The major-groove contacts are largely responsible for the specificity of binding but that the minor-groove contacts contribute also to the specificity of binding (Wang *et al.*, 1998). The repressor can bind just to one ARG boxes but its affinity is about 100-fold lower than to two ARG boxes (Tian *et al.*, 1992). The presence of another ARG box largely stimulates the binding of the properly located second box. Optimal distance separating two ARG boxes is three base pairs but one base pair more or less does not abolish the stimulation effect. This effect will be abolished completely by introduction of two or more additional base pairs unless a full helical turn is introduced (Charlier *et al.*, 1992; Tian *et al.*, 1992). The DNA binding domain of ArgR belongs to the winged helix-turn-helix (WHTH) family of DNA binding proteins (Sunnerhagen *et al.*, 1997).

The specific binding of ArgR to the *cer* site was described by Stirling *et al.* (1988). Using a DNase I footprinting technique, ArgR was shown to bind to a single ARG box about 100 bp upstream of the *cer* core recombination site, to which the two recombinases, XerC and XerD, bind. A pseudo-ARG box with poor sequence homology to that of the consensus sequence can be identified just after the eighteenth bp of the first ARG box but there is no direct evidence for ArgR binding or any functional role of this pseudo-ARG box in Xer site-specific recombination at *cer* (Merican, 1995).

1.4 ArgR homologues

ArgR homologues have been found in other groups of bacteria. This includes *Bacillus subtilis* (AhrC, North *et al.*, 1989), *Salmonella typhimurium* (Lu *et al.*, 1992), *Haemophilus influenzae* (Fleischmann *et al.*, 1995), *Bacillus stearothermophilus* (Dion *et al.*, 1997; Jianping *et al.*, 1999), *Streptomyces clavuligerus* (Rodriguez-Garcia *et al.*, 1997), and *Mycobacterium tuberculosis* (Cole *et al.*, 1998).

The *B. subtilis* AhrC protein is 35% identical in amino acid sequence to ArgR in the C-terminal half of the protein but only 19% identical in the N-terminal half; with 27% identity overall. AhrC can substitute for ArgR in *E. coli* to mediate Xer site-specific recombination (Stirling *et al.*, 1988) as well as to repress the expression of the L-arginine biosynthetic genes (Smith *et al.*, 1989). In addition, the DNA binding sites for AhrC show sequence similarity to the DNA binding sites of ArgR (Czaplewski *et al.*, 1992). The more related *Haemophilus influenzae* ArgR is also interchangeable with *E. coli* K-12 ArgR (Chen *et al.*, 1997).

1.5 ArgR mutants

Mutants of ArgR have been obtained by chemical (hydroxylamine; Tian and Maas, 1994) and site-directed mutagenesis (Burke *et al.*, 1994; Chen *et al.*, 1997). There are two types of *argR* mutants; transdominant and super-repressor. Transdominant mutants may be deficient either in binding to DNA or in binding to L-arginine, or in hexamer formation, whereas the super-repressor mutants

behave as L-arginine auxotrophs as a result of hyper-repression of the arginine biosynthetic enzymes in the absence of L-arginine. The super-repressor mutants are defective in binding L-arginine. These mutants showed enhanced DNA-binding in the absence of L-arginine compared to that of the wild-type repressor (Tian and Maas, 1994; Burke *et al.*, 1994; Chen *et al.*, 1997). One of the super-repressor mutants that had been extensively studied (and used in this study) is ArgR D128N, D129V (abbreviated for simplicity as ArgRNV). ArgRNV is proficient in repression as well as in supporting Xer site-specific recombination and appears to bind DNA as trimers in an L-arginine independent fashion (Burke *et al.*, 1994).

It is interesting to note that wild-type ArgR of *E. coli* B is a naturally occurring super-repressor 'mutant' of ArgR compared to the ArgR wild type of *E. coli* K-12. The ArgR of *E. coli* B differs from ArgR of *E. coli* K-12 in having leucine instead of proline at position 70 of the protein. The level of repression of L-arginine biosynthetic enzymes in *E. coli* B is not correlated with L-arginine concentrations. It is because in the absence of L-arginine, the ArgR of *E. coli* B bound to the *argR* operator sites much stronger than ArgR of *E. coli* K-12 did. This evidence indicates that L-arginine-free of ArgR *E. coli* B, but not L-arginine-free of ArgR *E. coli* K-12, is able to repress enzyme synthesis under physiological conditions, independent of the presence of L-arginine (Tian *et al.*, 1994; Gorini *et al.*, 1961).

1.6 Fusion protein system

Fusion protein involved the construction of gene fusion in which the coding sequences of two or more genes are fused together by genetic approaches. The protein segment fused to the protein of interest is usually a short peptide or small proteins that would give less alteration of the chemical and biological properties of the protein of interest. The fusion protein system widely used comprises fusion of a target protein to short pieces of well-defined peptides (Poly-His, biotinylated peptide, F lag-epitope, c-myc epitope or HA-tag) or small proteins (bacterial Glutathione S-transferase, Maltose binding Protein, jellyfish Green Fluorescent Protein, Thioredoxin, β -galactosidase, Protein A, β -glucuronidase, and VSV-Glycoprotein). Fusion protein systems allow expression and purification of the protein of interest in large quantity. Purification of certain protein in large quantity is very important in order to understand more about the structure and function of the protein in cellular processes. In addition to facilitating the expression and purification of proteins, fusion protein systems have also been utilized in the study of transport mechanism (Silhavy *et al.*, 1977), gene expression (Zabeau and Stanley, 1982), protein purification (Moks *et al.*, 1987; Uhlen *et al.*, 1983; Lofdahl *et al.*, 1983), plasmid replication (Light and Molin, 1982) and stoichiometric determination of the number of protein subunit that binds DNA (Hope and Struhl, 1987).

1.7 Objectives of study

A structure-function relationship study of ArgR is important in order to understand more about its role in Xer site-specific recombination. Thus, the

objectives of this study are to construct elongated ArgR proteins for both ArgRWT and ArgRNV using protein fusion system and to examine the biological properties of these fusion constructs. The biological property of the protein in Xer site-specific recombination will be determined using *in vivo cer*-mediated recombination assay using suitable reporter plasmids.