UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 214

Mechanisms and DNA Specificity in Site-specific Recombination of Integron Cassettes

CAROLINA JOHANSSON

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2007

ISSN 1651-6206
ISBN 978-91-554-6767-8
urn:nbn:se:uu:diva-7429

Dissertation presented at Uppsala University to be publicly examined in C10:301, BMC, Husargatan 3, Uppsala, Wednesday, January 24, 2007 at 09:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English.

Abstract
Johansson, C. 2007. Mechanisms and DNA Specificity in Site-specific Recombination of Integron Cassettes. Acta Universitatis Upsaliensis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 214. 65 pp. Uppsala.
ISBN 978-91-554-6767-8.

Bacterial resistance to antibiotics has become a serious problem. This is due to the remarkable ability of bacteria to respond and rapidly adapt to environmental changes. Integrons are elements with the capacity for gene capture by an integron-encoded site-specific recombinase called IntI. IntI binds and acts at the recombination sites, *attI* and *attC* resulting in excision and integration of short DNA elements called gene cassettes carrying an *attC* site in the 3' end. Several families of antibiotic resistance genes are borne on gene cassettes in integrons connected to mobile elements. Other cassettes reside in the larger and ancestral superintegrons located on chromosomes in both pathogenic and environmental bacteria. Due to their close connection with lateral gene transfer systems, it is possible that integrons are functionally dependent on those networks. This work presents arguments for such connections. The *attC* of the *aadA1-qacE* cassette junction in Tn*21* was characterized in detail. Like other *attC* sites, it contains two pairs of inverted repeats and is almost palindromic. By using electrophoretic mobility shift assays, this study showed that IntI1 binds only to the bottom strand of *attC*. Upon folding the strand into a hairpin, a few chiral hairpin distortions define both the strand choice and also the appropriate orientation of the highly symmetrical site. Structural recognition also explains the wide sequence variation among *attC* sites. We have documented the initial cleavage step in recombination in IntI extracts and integrase levels in extracts were evaluated by a new method. Mutagenesis and homology modelling were performed to find amino acid residues in IntI1 that are important for recognition of *attC* hairpin-DNA. Comparisons were made with other tyrosine family members to explain how integron integrases differ in site-recognition and also in their mechanism of strand exchange.

Keywords: lateral gene transfer, site-specific recombination, tyrosine recombinase, integron, single-stranded, DNA hairpin

Carolina Johansson, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden

ISSN 1651-6206
ISBN 978-91-554-6767-8
urn:nbn:se:uu:diva-7429 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-7429)

To my loving family

List of papers

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals.

I **Johansson C**, Kamali-Moghaddam M, Sundström L. 2004. Integron integrase binds to bulged hairpin DNA. *Nucleic Acids Res 32*: 4033-4043.

II Hansson K, **Johansson C**, Sundström L. 2006. Long-range effects of mutations in *attC* correlate with the formation of an integron hairpin recombination substrate. Manuscript.

III **Johansson C**, Boukharta L, Eriksson J, Åqvist J, Sundström L. 2006. Mutagenesis and homology modelling of the Tn*21* integron integrase IntI1. Manuscript.

IV **Johansson C**, Samskog J, Sundström L, Wadensten H, Björkesten L, Flensburg J. 2006. Differential expression analysis of *Escherichia coli* proteins using a novel software for relative quantitation of LC-MS/MS data. *Proteomics 6*: 4475-4485.

Contents

Abbreviations

aadA1	gene for aminoglycoside adenylyl transferase of type A1
attI	integron attachment site
attC	cassette attachment site
be	base element
bp	base pair or base pairs
CR	Common region
DR	Direct repeat
ds	double-stranded
FIS	Factor for inversion stimulation
GTTRRRY	R is purine (adenine or guanine); Y is pyrimidine (thymine or cytosine)
GWTMW (or GNT)	W, adenine or thymine; M, adenine or cytosine; N, guanine, adenine, thymine or cytosine
HTH	Helix-turn-helix
HU	Histone-like heat unstable nucleoid protein
ICE	Integrative conjugative element
IHF	Integration host factor
intI	gene for integron integrase
IntI	Integron integrase
IS	Insertion sequence
ISCR	IS common region
LINE	Long interspersed repetitive elements
MGE	Mobile genetic element
MRI	Multiresistance integron
ORF	Open reading frame
RC	Rolling-circle
SINE	Short interspersed repetitive elements
ss	single-stranded
2rs	secondary site

Introduction

It is an astounding thought to outline the dimensions of the dark tragedy that never happened thanks to the invention of antibiotics. Antibiotics came as a great relief from the burden of so many painful and often deadly diseases due to bacterial infections. Since the invention of antibiotic therapy in the 1930's, resistance to almost all useful drugs has appeared in pathogenic organisms (Davies, 1994). This dramatic episode of evolution and gene spread among bacteria is now a major threat to humankind and a concrete manifestation of the ability of microorganisms to respond and rapidly adapt to environmental changes. The rapid and increasing spread of antibiotic resistance is forcing medical care and research to find new solutions to suppress the spread of resistance. New concepts for antibiotic treatment, for example, should include optimization of the therapeutic use and reduction of the the total consumption of antibiotics. If we do not address this serious issue, there may be a worrying future scenario where there will be no antibiotics available to cure bacterial infections. Bacteria gain resistance by spontaneous mutations in either their own housekeeping genes or by acquiring new genetic information due to horizontal gene transfer. Research on the dissemination of antibiotic resistance has led to the discovery of many naturally occuring mobile elements, vehicles on which resistance genes are usually found. These vehicles of spread include transposons, conjugative plasmids and pathogenicity islands. In the middle of the 1980s it was revealed by analysis of sequence data that many unrelated antibiotic resistance genes in gram-negative pathogens reside in tandem clusters flanked by partially related sequences. It was gradually realized that many plasmids might have a specific mechanism for capture of various genes (Cameron *et al*., 1986; Hall and Vockler, 1987; Ouellette *et al.,* 1987; Sundström *et al.*, 1987, 1988). When also an adjacent gene of a site-specific recombinase was found (Ouellette and Roy, 1987; Sundström *et al*., 1988) the contours of a complete gene-packaging system appeared. The term integron was invented (Stokes and Hall, 1989). It was realized early on that integrons probably have a more general role rather than specific dissemination of antibiotic resistance genes (Martinez and de la Cruz, 1988; Stokes *et al*., 1997; Sundström *et al*., 1988).

The rapid spread of antibiotic resistance genes and the fact that multiresistance in clinical isolates correlates with the presence of integrons (Leverstein-van Hall *et al*., 2003), strongly encourage research aiming for an improved understanding of the recombination mechanisms used by integron

systems. The work presented here is focused on analyzing the recognition strategy of integron-borne integrases and the unusual organization of the recombination sites needed for integron-mediated site-specific recombination to occur.

Antibiotics and resistance mechanisms

Antibiotics are drugs capable of inhibiting bacterial growth and some of them have a lethal activity. Most antibiotics are natural compounds produced by bacteria and fungi but modified chemically to be used as drugs. Two very important events in medical history are the discoveries of the antibacterial effects of penicillin and sulfonamides. Alexander Fleming discovered penicillin in 1928 and further studies by the Oxford investigators, Florey and Chain, lead to the introduction of penicillin into clinical practice during World War II. All three shared the Nobel prize in 1945 for their contributions to the discovery (Ligon, 2004). Sulfonamides were discovered by Gerhard Domagk in 1932, and he was awarded the Nobel prize in 1939. The first sulfa drug, prontosil, was in 1935 the first antibacterial agent introduced into clinical practice (Otten, 1986).

Antibiotics are ideally harmful to bacterial cells but are not harmful to human cells by selectively acting on functions that are different in the two types of cells. Most antibiotics used today were developed within the period 1950 to 1970. Antibiotics are classified into groups depending on their modes of activity (Tenover, 2006). Four of those modes are: 1) drugs interfering with the cell wall biosynthesis (ex β-lactams (of which penicillin is one example), cefalosporins, glycopeptide agents), 2) drugs interfering with protein biosynthesis (for example tetracyclines, macrolides, aminoglycosides), 3) drugs interfering with nucleic acid synthesis (for example fluoroquinolones, rifampicin) and 4) drugs interfering with a metabolic pathway (trimethoprim, sulfamethoxazole). The development of new drugs is both time consuming and costly but urgently needed as we face the increasing consequences of the resistance problem. Since the launching of trimethoprim drugs in the 1960's only two new classes of antibiotics which both inhibit protein synthesis, have been introduced into clinical practice, the oxazolidinones and the glycylcyclines. Bacteria developing resistance acquire resistance mechanisms including; efflux pumps preventing the accumulation of the drug, modified protein structure of the bacterium or replacement of the antibiotic target, inactivation of the drug and blocking entrance of the drug into the bacterial cell. Most of the resistance genes known today are thought to originate from antibiotic-producing strains that need these properties to avoid self-destruction and drug resistance was in fact already discovered in 1940. Prior to penicillin use in clinical practice, investigators found enzymes, β-lactamases, that could catalyze the hydrolysis of the β –lactam ring

and potentially interfere with penicillin therapy (Davies, 1994). However, multidrug antibiotic resistance was at that time not thought of as a problem during therapy since the mutation frequency was regarded low. When multiresistant bacteria were eventually discovered in the 1950's it was stated that the pattern of resistance genes must have been gained by horizontal transfer between bacteria rather than by mutations alone, since the resistance developed so quickly (Mazel 2006; Ochman, 2000). Horizontal gene transfer systems and the importance of integrons for the spread of antibiotic resistance genes will be the focus of the next chapters.

Horizontal gene transfer

Inheritance of genes is largely vertical but in most bacteria a significant proportion of the genome is introduced by lateral gene transfer (Beiko *et al*., 2005). This can be studied by comparing the history of genes within an organism where some differ owing to horizontal gene transfer (Gogarten and Townsend, 2005). Mobile genetic elements (MGEs) such as plasmids, bacteriophages and transposons, combined with gene loss, gene duplications and chromosomal rearrangements have greatly facilitated the evolution of prokaryotes by generating changes in their genomes and thereby creating genomic diversity (Ochman *et al*., 2000, Frost *et al*., 2005).

Bacteria with smaller genome size and a smaller endogenous sequence supplement than that of eukaryotes are able to share genetic information via horizontal gene transfer mediated by conjugation, transduction and transformation. The principal process is conjugation that occurs through direct cell to cell contacts (Mazel, Davies, 1999). It allows for genetic exchange between many different bacteria in nature, sometimes also with eukaryotic cells, and is often mediated by genes on plasmids (conjugative plasmids) (Heinemann and Ankenbauer 1993). However, conjugation is by no means restricted to plasmids. There are also integrative and conjugative elements named ICEs with plasmid and phage-like features that may mobilize gene clusters, commonly pathogenicity islands (Burrus and Waldor, 2004). Transduction, another form of horizontal transfer process, is mediated by viruses /bacteriophages. Some bacteriophages are temperate phages meaning that they can integrate as a prophage into the bacterial host chromosome and establish lysogeny. Prophages constitute a substantial part of horizontally acquired DNA in many bacteria. Sometimes bacteriophages mediate transfer of other mobile DNA or bacterial DNA when excision of the prophage is imprecise and these processes are called specialized or generalized transduction (Canchaya *et al*., 2003). The third transfer process is transformation in which free DNA is taken up from the environment (Averhoff and Friedrich, 2003).

To establish the incoming DNA in the genome all these events need support from recombination processes like homologous and illegitimate recombination. Some routes for gene assembly are programmed and in these instances site-specific recombination that join different genomic components together occur (see section below). According to Beiko *et al*., (2005) DNA exchanges are more likely to succeed between close relatives than more distantly related species due to the increased chance for conjugation and homologous recombination events in the former.

It is generally agreed that acquisition of new sequences and genomic expansion is vital for evolution. The main reason for gene duplication or importation of sequences that have an active polypeptide protein fold is the small proportion among the stock of all theoretical sequences endowed with a potential for biological function (Britten, 2005; Ochman, 2000). However, the excess genome size due to duplication events among resident sequences in the genome or horizontal gene transfer necessitates trimming of the genome size for stabilization of the host. Recombination leading to deletions occurs intermittently under reduction episodes. A continuously larger genome will affect the replication rate and cell division and bacteria will lose fitness (Kurland, 2005).

If and how long the acquired gene is kept by bacteria depends on its function and the strong positive environmental selection for it (Ochman *et al*., 2000). Mobile elements often carry antibiotic resistance genes that will increase their likelihood of persistence when antibiotics are present. Other functions that may give a selective advantage are virulence factors such as the cholera toxin encoded by a prophage in *Vibrio cholerae*. Integrons are themselves immobile but these presumed gene-packaging systems are often located within transposons and/or conjugative plasmids and in that way contribute to the gene traffic leading to acquisition of new genes in bacteria.

Recombination

Recombination plays an essential role in many processes that involves rearrangement of genetic information in and among DNA molecules and is important for genome maintenance and evolution. **Homologous recombination** occurs between DNA segments that share extensive sequence homology. It is tightly linked to cell division in eukaryotes, and in bacteria it often occurs during conjugation and replication where it is important for repair of several types of DNA damage. Many proteins have been found to be involved in homologous recombination in *E.coli* of which the RecA, RecBCD and RuvABC proteins are well characterized (Kowalczykowski, 2000). The RecBCD enzyme complex is both a helicase and nuclease that unwinds DNA and processes DNA breaks to generate single-strands for invasion. RecA plays a central role and catalyzes pairing of the ssDNA with homolo-

gous dsDNA and the exchange of DNA strands via formation of a Holliday intermediate that is followed by branch migration facilitated by RuvA and RuvB proteins. The Holliday intermediate consists of a heteroduplex region in which strands from different DNA molecules are joined. The enzyme that resolves the intermediate is the resolvase RuvC. Homologous recombination is divided into RecA-dependent and RecA-independent recombination. The mechanism for the latter has remained cryptic but has been observed to operate with short homologous sequences (Lovett *et al*., 2002; Bi and Liu, 1994). **Illegitimate recombination** occurs at low frequency and involves end-joining between sequences of little or no homology. Double-stranded breaks induced by UV irradiation or other DNA damaging agents are thought to trigger initiation of this type of recombination. It can be grouped into two classes: short homology independent, involving topoisomerase I, DNA gyrase and the DNA binding protein HU, and short homology dependent illegitimate recombination that needs short regions of homology (Ikeda *et al*., 2004). **Transpositional recombination** or transposition is the process when DNA transposons and IS (insertion sequence) elements "jump" from one place to another on the same or a different chromosome or plasmid. Transposition does not need matching between the mobile DNA element and the target and the target selectivity is often low. **V(D)J recombination** shares similarities with transposition and is a mechanism for generating antibody and T-cell-receptor diversity. There is no requirement for homologous sequences and break-ends are directly ligated. **Conservative site-specific recombination** requires short sequences of DNA homology at the actual sites of cutting and resealing for precise DNA rearrangements and can also involve a Holliday intermediate. This class of recombination is used by cells and viruses in a variety of biological processes and will be discussed in more detail below.

A representative selection of proteins

A protein is built up of one or several chains of amino acids that determine its shape and biological function. The number of combinations the 20 different amino acids can be linked are enormous and specified by a gene's DNA sequence, while the number of unique ways a chain can be folded into an active and functioning structure is limited. Proteins perform a wide variety of functions in living cells by acting as hormones, enzymes, antibodies and structural proteins. Proteins often interact with each other to form complexes with synergistic effects.

Enzymes may be catalysts in biochemical reactions and without them almost all chemical processes in the cell would stall. A subset of enzymes are involved in processes connected with nucleic acids and in some of these reactions the enzymes themselves participate by reacting with its substrate

(Mizuuchi, 1997). This is exemplified by enzymes that utilize an OH group on the side chain of a tyrosine or a serine residue. After deprotonation of these proteic OH groups they have the capacity for conducting a nucleophilic attack on a phosphate diester group in DNA to transiently form a covalent protein-DNA linkage. Due to the in-line substitutive nature of the reaction with the phosphate group, the attacking oxygen nucleophile is bound to a phosphate at the expense of losing one of the earlier bound oxygen ligands. An unstable intermediate coordinating five oxygens could possibly appear (Cassano *et al*., 2004). The covalently linked protein-DNA adduct contains phosphotyrosyl (or phosphoseryl) linkages that are highly reactive and passed on further into consecutive nucleophilic reactions yielding nucleic acid products and regained enzyme. The catalysis among this group of proteins is independent of magnesium ions but could rely on magnesium-equivalent effects due to amino acids in the protein structure. In the following sections a selection of proteins will be presented that are important for the understanding of the integron system, the main topic of this thesis.

Site-specific recombinases

Site-specific recombinases are well documented in bacteria and yeast where they promote remarkably accurate DNA rearrangements in processes called conservative site-specific recombination. The recognition of the nucleic acid and progression along the reaction trajectory are directed by highly specific DNA-protein and protein-protein interactions while DNA-DNA sequence homology has only an indirect role (Grindley *et al*., 2006). The outcome of the reaction is dictated by the relative disposition of two recombining sites and importantly, the same reaction can be employed to produce integration, excision or inversion of DNA. Site-specific recombination is used for many biological reactions satisfying cellular as well as viral functions such as phage lysogenization, control of expression and resolution of replicon dimers. In the evolutionary rather than physiological perspective, site-specific recombination introduces a possibility for programmed variation of genomes (Nash, 1996; Sadowski, 1986). For recombination to occur, the DNA of two recombination sites must first be recognized by a site-specific recombinase, individually or pairwise and brought into intimate and often cooperative physical contacts with the protein. The resulting complex can mature into higher ordered assemblies that are activated to form a catalytically competent synaptic complex. Sometimes additional DNA-binding proteins assist in the assembly of the recombinases, in particular with the purpose to regulate the recombination reaction (Echols, 1990; Hallet and Sherratt, 1997; Landy, 1989; Stark *et al*., 1989b). In its earliest step a DNA strand is broken and in following reactions the ends are transferred to phosphate diester groups in the polynucleotide of new partners where rejoining takes place. Site-specific recombinases resemble topoisomerases in that they

conserve the energy of the broken phosphodiester bond in a protein-phosphate linkage and use that energy for subsequent re-ligation of the phosphodiester backbone (Sadowski, 1986). The first well-studied example of site-specific recombination was the integrative process in lysogenization of bacteriophage lambda (λ) (Campbell, 1962; Weisberg and Landy, 1983; Landy, 1989).

On the basis of amino acid sequence homology and biochemical details, site-specific recombinases can be divided into two families: the serine family and the tyrosine family (Grindley *et al*., 2006). The serine recombinases are further divided into small serine recombinases and large serine recombinases (Smith and Thorpe, 2002; Grindley *et al*., 2006). A minor and very conserved subgroup of small serine recombinases, called invertases, build molecular complexes combined with the small DNA-bending protein FIS to act only on oppositely repeated sites disposed to generate inversion. In contrast to invertases the majority of small serine recombinases recognize their sites placed in direct orientation and produce DNA resolution (or excision). These resolvases bind to the DNA without involvement of FIS, and a synaptic complex of different topology is built on multiple repeats of the site core dyad. A relatively large number of serine recombinase protomers act as recombisome structural units when they assembled on the repeated sites (Grindley, 1993). These recombinases use a conserved serine hydroxyl to attack the phosphodiester bond and are relatively highly conserved in primary sequence. The serine in the small type of enzyme is found roughly at amino acid position 10 in the N-terminal domain where the majority of the conserved amino acids also reside (Johnsson, 1995). The best-characterized members of the serine family are the invertases, Gin from bacteriophage Mu and Hin from *Salmonella* sp. and the resolvases of γδ and Tn*3* transposons (Grindley *et al*., 2006 and references therein; Stark *et al*., 1989a). Members of this family carry out the entire recombination reaction in one step, where the four DNA strands are simultaneously cleaved and swapped. It is therefore not surprising that small serine recombinases act in intramolecular reactions but act very inefficiently in intermolecular reactions (Johnsson, 1995). However, the large serine recombinases (Smith and Thorpe, 2002) are similar to tyrosine recombinases and catalyze integration. They have been found in phages and transposons mainly in gram-positive bacteria.

The tyrosine recombinases use a tyrosine hydroxyl as the attacking nucleophile. Tyrosine recombinases mediate recombination in two steps through two consecutive exchanges between pairs of DNA strands via an intermediary Holliday junction. Since the integron system contains an integron integrase that is a member of the tyrosine family of site-specific recombinases the following section will focus on some of the tyrosine recombinase systems.

Tyrosine recombinases

The tyrosine family of recombinases shows more amino acid sequence variation than the small serine recombinases. Argos and coworkers (1986) identified the proteins as a structural family by primary and secondary structure comparisons. Very little homology was observed in the N-terminal halves of the proteins, however analysis of the C-terminus of seven different bacteriophage systems showed conservation, in particular with respect of two regions (called boxes I and II).

The tyrosine recombinases are represented in bacteria, bacterial viruses, archaea and lower eukaryotes and a recent database search resulted in about 1000 related members (Grindley *et al.*, 2006). The best studied members of the family are the phage lambda integrase protein, which promotes integration and excision of the phage genome from the host chromosome in *E.coli* (Landy, 1989); the Cre integrase of the *E.coli* phage P1, the role of which is to maintain the phage genome as a monomeric, unit-copy plasmid in the lysogenic state (Abremski and Hoess, 1984); Flp, which mediates amplification of the yeast 2-micron plasmid copy number by inverting a DNA segment during replication (Vetter *et al*., 1983); and the chromosome dimer resolution proteins XerC/D of most prokaryotes (Blakely *et al.*, 1993). In addition, we notice that the *Haemophilus* phage 1 integrase HP1, was the first structurally described tyrosine recombinase (Hakimi and Scocca, 1994). It has strong resemblance with the P2 phage integrase, another example of tyrosine recombinases studied in depth (Argos *et al.,* 1986; Frumerie *et al.,* 2005).

The tyrosine family members comprise of 300-400 amino acids which can be divided into a structurally varied N-terminal domain and a larger and mostly helical catalytic C-terminal domain. Although the amino acid sequences within the tyrosine family are diverse and interrupted by frequent deletions and insertions, it has been possible to align and compare sequences from their C-terminal domains (Yang and Mizuuchi, 1997). The most conserved feature in the C-terminal domain is a set of six highly conserved residues that are part of the catalytic site (Abremski and Hoess, 1992; Argos *et al*., 1986; Esposito and Scocca, 1997; Grainge and Jayaram, 1999). Almost all six residues reside in two boxes (box I and box II) of marked sequence similarity throughout the family: two arginines, two histidines, one lysine and one tyrosine. The tyrosine provides the nucleophile to break the labile phosphodiester bond adjacent to the bound recombinase monomer. The other mentioned active site residues are either involved in stabilization of the leaving group or in phosphate activation. In addition to the highly conserved box I and box II motifs, three regions of conserved sequence have been identified and named patch I, II and III. These are located around box I and seem to be important in the secondary structure of the proteins (Nunes-Düby *et al*., 1998). Tyrosine recombinases form subfamilies based on significant levels

of sequence similarities beyond the active site residues (Esposito and Scocca, 1997). The focus of this thesis is the integron integrases that form their own phylogenetic branch clearly within this family.

Structures of complete synaptic complexes, including monomeric DNA complexes and/or protein in the absence of DNA (e.g HP1, λ Int, Cre, XerD, Flp, human topoisomerase I and vaccinia virus topoisomerase) are now available (Van Duyne, 2001; Guo *et al*., 1997; Chen *et al*., 2000; Biswas *et al.,* 2005; Redinbo *et al.,* 1998; Kwon *et al*., 1997; Hickman *et al*., 1997; Subramanya *et al*., 1997; Sharma *et al*., 1994). More recently, the first integron integrase, VchIntI1 from *Vibrio cholerae* was crystallized bound to its substrate (MacDonald *et al*., 2006). These structures reveal that integrases have almost identical three-dimensional peptide folds despite their relatively high sequence diversity. The ability to align these three-dimensional structures has been essential for understanding their mechanical similarity.

Mechanism of tyrosine recombinases

The tyrosine recombinases, according to a classical description, typically recognize a core part of their recombination sites consisting of about 30 base pairs (bp) with two inverted integrase-binding repeats of about 11-13 bp in length that are separated by a central region of 6-8 bp. The central regions are asymmetric giving the sites a directionality that determines which one of the two strands in each site that are cleaved first in the recombination reaction. Functional asymmetry has also been observed in sites of the small serine recombinase TnpR from Tn*3* (Blake *et al.,* 1995). The recombination reactions usually occur between identical sites that are unique (as in the Cre/*loxP* system; Van Duyne, 2001) or constant pairs of two partner sites (for example the λ-*attB/attP* system; Dorgai *et al*., 1998). The reaction is performed in two steps by a tetramer of the recombinase protein bound to two sites with antiparallel spacers. For the well-known systems of Cre, Flp and lambda Int the tetramers have twofold and pseudo fourfold symmetry. In the synaptic complex two protomers are in a cleaving (active) and two in a non-cleaving (inactive) conformation at any given time, regulated by the positioning of the active site residues in the two different interfaces. The first strand exchange starts with cleavage of the two top strands at one end of the central regions by nucleophilic attack of two active site tyrosines on the scissile phosphodiester linkages of the paired DNA backbones. Two free 5'-OH are generated after the covalent 3'-phosphotyrosine linkages have been formed. The two strands are then joined across the DNA partners to form a Holliday junction by attack of the 5'-OH groups that will break the 3'-phosphotyrosine bonds. This intermediate is finally resolved into recombinant products by the previously inactive pair of recombinases repeating the cleavage-religation process of the two bottom strands at the other end of the synapsed central regions. Alternatively, the Holliday junction can be resolved back to the substrates by reversal of the first strand exchange step.

During a complete recombination reaction, a conformational change of the Holliday junction intermediate is expected to be necessary to allow for the pairwise activity/inactivity switch of the four recombinases. It is through this isomerization step of the Holliday junction (facilitated if nearly square planar) the tight control of the recombination reaction is achieved (Arciszewska *et al*., 2000; Gopaul and Duyne, 1999; Hallet *et al*., 1999). After recombination, the central region has each strand derived from different parents. Therefore, identity in the central region is expected to be highly important to ensure that a left half-site is rejoined to a right half-site during recombination (Nunes-Düby *et al.,* 1997; Sherratt, 2001).

Evolution of the tyrosine family of recombinases and the type IB topoisomerases

Topoisomerases are able to change DNA topology and play important roles in processes like replication, recombination and transcription. The type 1B topoisomerases form a subclass of topoisomerases that are structurally distinct from all other known topoisomerases but similar to tyrosine recombinases (Champoux, 2001). The type 1B topoisomerases cleave and religate only one strand of DNA and like the tyrosine family of recombinases they use a tyrosine as a nucleophile to attack the DNA yielding a covalent DNA-3'-phosphotyrosyl enzyme intermediate and a free 5'OH-DNA end. Tyrosine recombinases are mostly known in bacteria and yeast while topoisomerases IB have been found almost exclusively in the eukaryotic domain until very recently when bacterial type IB enzymes were discovered (Krogh and Shuman, 2002). The two protein families most likely result from divergent evolution and a common ancestor with strand transferase activity via a tyrosyl-3'phosphodiester intermediate has been proposed (Krogh and Shuman, 2002). Their active site conformation is very similar and most tyrosine recombinases display type I topoisomerase activity *in vitro* (Cheng *et al*., 1998).

Topoisomerase proteins are built up of two evolutionary distinct domains. It is suggested that there has been a common ancestral C-terminal domain containing the tyrosine and the four catalytic amino acids RKRH that then underwent differentiation into a recombinase-type and a topoisomerase-type catalytic domain. Tyrosine recombinases are found to contain a RKHRH pentad of catalytic amino acids whereas all members of the topoisomerase IB family contain a conserved RKKRH pentad in the C-terminal domain. The catalytic amino acids occupy homologous positions in the tertiary structures of the two families of proteins.

During evolution the C-terminal domains were connected to various N-terminal domains forming the two separate families of tyrosine recombinases on one side, and bacterial, poxviral and eukaryotic nuclear type IB topoisomerase on the other. There is no structural similarity at all between the N-terminal domains. Differences among the tyrosine recombinases suggest that

they have evolved independently in parallel by gene fusions events that captured different amino-terminal modules mediating varied biological specificities. Different types of 1B topoisomerases probably evolved serially by fusion of the bacterial/poxvirus-like amino-terminal domain to the ancestral topoisomerase-type catalytic domain whereby horizontal gene transfer events probably played a key role in the appearance of topoisomerase IB in the eukaryotic domain. The nuclear type IB topoisomerases are likely to have evolved from a bacterial/poxviruslike precursor by addition of multiple discrete structural modules (Krogh and Shuman, 2002).

Another class of enzymes that share the catalytic domain with the tyrosine recombinases is the telomere resolvases of certain prokaryotes and viruses (Chaconas *et al*., 2001; Deneke *et al*., 2000). One example, ResT, is discussed below.

A mix of proteins directly or indirectly involved in nucleophilic attacks on DNA/RNA

Replication and transcription are common events in the cell where nucleic acids are processed through nucleophilic reactions. Here the nucleophilic attacks are made by OH groups of nucleotides on phosphodiesters. The operating catalysts are **DNA polymerase** and **RNA polymerase**. The specialized DNA polymerase involved in replication of the ends of eukaryotic linear chromosomes is called **telomerase**. It ensures that the chromosome end (the telomere) is maintained at sufficient length. The telomerase elongates the 3' end of ssDNA by using an RNA template and is therefore both formally and by homology to other proteins a **reverse transcriptase**. This enzyme is related to an operational enzyme for transposition of certain transposons called non-viral retrotransposons. They use the enzyme to synthesize DNA from an RNA copy of the transposon as a template, and a nick at the target site of insertion as a primer. In fact telomere extension and mobility of these elements exemplified by LINEs and SINEs have strong resemblance to one another. Retroviruses like HIV for example, are similar to LINEs and SINEs only from the aspect of using a reverse transcriptase. The recombinational behavour instead depends on a **retroviral integrase** with a similar function to the transposases used by DNA transposons such as cut-and-paste and replicative transposons.

Transposases and retroviral integrases catalyze one-step transesterification reactions mediating transposition events and do not form covalent protein-DNA linkages. Transposition differs from site-specific recombination in being more or less random with respect to the target, by creating duplication at the site of insertion, and by requirement for DNA synthesis and energy cofactors. The transposases use a DDE motif or equivalent to catalyze a direct transesterification reaction via Mg^{2+} cations and water molecules as

nucleophiles to release the reactive transposon ends in the donor sequence. The integration is a reaction when the terminal 3'-OH groups of the transposon attack an uncut target sequence under transposase catalysis. These transposons are exemplified by the cut-and-paste transposons Tn*5*, Tn*10*, Tn7, and the replicative transposons Tn*3*, bacteriophage Mu and Tn*5090/402*.

The donor cleavage is partial or varies in other aspects among many IS elements and transposons. One in particular unusual mechanism called rolling-circle (RC) transposition is used by the IS*91* family of transposable elements. **IS*91*-like transposases** lack the typical DDE motif of most other transposases and instead utilize tyrosine residues to form different covalent 5' DNA-protein linkages during the process (Garcillan-Barcia *et al*., 2002). The transposase TnpA from IS*91* shares similarities with A proteins of RC replicating phages (such as φX174) including a pair of invariant tyrosines that catalyze two successive transesterification reactions during replication initiation and termination (del Pilar Garcillán-Barcia *et al*., 2001).

ResT is a telomere resolvase of *Borrelia burgdorferi* that has a linear chromosome which is rare among prokaryotes. The enzyme opens up and closes DNA ends at replicative intermediates formed during replication thereby generating covalently closed hairpin ends. ResT has a hairpin-binding module (discussed below) and it is reactive via a tyrosine nucleophile that attacks the phosphate backbone of DNA forming a 3'phosphotyrosine intermediate linkage (Chaconas, 2005). The enzyme is related to both cut-and-paste transposases, and to tyrosine recombinases and type 1B topoisomerases (Bankhead and Chaconas, 2004; Deneke *et al*., 2004). An interesting observation is that ResT and tyrosine recombinases XerC/D have some degree of mechanistic and structural similarity. Both proteins act on chromosomal structures, linear and circular, respectively, to ensure stable replication products by solving topological problems.

Conjugative relaxases and the origin of transfer (*oriT*) are the central components in the conjugative mobilization of most plasmids (Parker *et al*., 2005). The relaxases initiate bacterial conjugation by site-specific cleavage of the transferred DNA strand at the *oriT*. Nicking occurs by a transesterification reaction where a catalytic tyrosine residue acts as the nucleophile and mediates a 5'tyrosylphosphodiester linkage. A conjugative relaxase has a strand-transfer activity and enters the recipient cell covalently bound to the 5' end and catalyzes recircularization of the transferred DNA (Parker *et al*., 2005; Draper *et al*., 2005). The TrwC family of relaxases has been suggested to involve a double tyrosine mechanism for DNA processing like the phage A protein of φX174 (César *et al*., 2006).

Topoisomerases, have been discussed previously. These enzymes cover somewhat different biological roles and some of them are important targets in anti-cancer and antimicrobial chemotherapy. Topoisomerases use a tyrosine as the nucleophile in a DNA transesterification reaction. Besides altering supercoiling they can promote catenation and decatenation of circular

DNA. They are divided into type I and type II depending on whether they introduce temporary single- or double-stranded breaks in the DNA prior to strand passage and resealing. The type I topoisomerases are also subclassified into type IA and type IB based on whether they form a 5'- or a 3'phosphotyrosyl adduct, respectively. Topoisomerases play important roles in replication, recombination and transcription (Wang, 1996; Champoux, 2001).

DNA-binding proteins

DNA-binding properties are vital for proteins controlling gene expression, replication, recombination and maintenance of genomes (Pabo and Sauer, 1992). DNA-binding proteins can be grouped into classes that use related structural motifs for recognition. The helix-turn-helix (HTH) proteins form one class, and they were the first class recognized because of their structural similarities. The HTH motif consists of an α-helix, a turn and a second α-helix. Well-known HTH binding proteins include the λ Cro proteins, the λ repressor, the repressors involved in tryptophan and lac operon regulation (Trp repressor and the Lac repressor), the *E.coli* CAP protein and the FIS protein. In the HTH motif the two helices lie with a certain fixed angle relative to each other. The second helix is called the recognition helix and fits into the major groove in DNA. The HTH motif is not stable by itself and always occurs as part of a larger DNA binding domain. Residues outside the HTH motif can have significant roles in recognition by mediating important protein-protein contacts. The HTH-DNA interaction becomes more stable and specific if the HTH motif-containing proteins bind as dimers to two adjacent half-sites that are inverted relative to each other.

Site-specific recognition is mediated by direct hydrogen bonds between the protein side chains and the bases, occasional hydrogen bonds between the polypeptide backbone and the bases, or hydrogen bonds mediated by water molecules and hydrophobic contacts. The larger major groove is more important for site-specific recognition than the minor groove in the DNA. Some amino acid side chains make contact with more than one base and some bases are in contact with several side chains. Hydrogen bonds with purines, especially guanines, appear to be particularly important for recognition. In addition, contact with the DNA backbone, mainly hydrogen bonds and/or salt-bridges to the phosphodiester oxygens are important for site-specific recognition. The role of backbone contact could be to enhance the orientation of specific base-side chain contacts and lock the surrounding peptide chain in a certain position. The structure of DNA is to a limited extent determined by its sequence but in those cases, contacts with the DNA backbone may allow indirect recognition of the sequence such as unpaired nucleotides generating bulges.

Tyrosine recombinases contain HTH motifs that overlap with catalytic residues and participate in DNA binding (Grishin, 2000). In the DNA-stuctures of Flp, Cre and lambda, each domain flanking the crossover site inserts a helix into a major groove, but direct side-chain-base contacts are rather few (Grindley *et al*., 2006). Among integron integrases (discussed in the next chapters) recognition appears to be more structurally dependent where bulges play an important role in hairpin DNA, rather than sequence dependent.

Integrons

Function

Integrons are DNA elements with a capacity for gene capture by a RecA-independent site-specific recombination system (Martinez and de la Cruz, 1988; Stokes and Hall, 1989). The elements were originally noticed due to a modular variation within a narrow segment on plasmids in human pathogens where they carry a broad variety of different antibiotic resistance genes (Cameron *et al*., 1986; Hall and Collis, 1995; Leverstein-van Hall *et al*., 2003; Martinez and de la Cruz, 1990; Ouellette *et al*., 1987; Sundström *et al*., 1988). An integron functions as a natural cloning and expression system with a capacity for generating arrays of genes of enormous diversity. The integration and excision of genes that are part of gene cassettes (Recchia and Hall, 1995b) are mediated by an integron-encoding site-specific recombinase, IntI, that performs recombination between two sites, one in the integron (*attI*) and one in the gene cassette (*attC*) (Collis *et al*., 1993; Collis and Hall, 1992a, 1992b). In contrast to integrative conjugative elements (ICEs; Burrus and Waldor, 2004) integrons are constructed for mobilizing the cassettes but are themselves immobile. However, a few classes of integrons have been included into transposons that are typically found on large conjugative plasmids such as R388 (IncW group) and R46 (IncN group). This means that the integrons, if borne on conjugative plasmids, are directly connected to conjugative transfer mechanisms (Liebert *et al*., 1999; Sundström, 1998; Sundström *et al.*, 1991). The integrons have also been reported on chromosomes in proteobacteria more or less related with *Vibrio* species. This is suggested to be a location of the integron ancestry.

Definition

The essential components of an integron are located in a 5'-conserved segment (CS). The components comprise a gene encoding integron integrase (IntI) which is a site-specific recombinase belonging to the tyrosine family,

an adjacent recombination site *attI* into which gene cassettes are integrated, and P_c for the expression of integrated DNA (Bunny *et al*., 1995; Collis and Hall, 1995; Levesque *et al*., 1994; Stokes and Hall, 1989; Sundström *et al*., 1988; Ouellette and Roy 1987) (Figure 1). The segment in the 3' end of many integrons may also be highly conserved but only a few base pairs in the end are functional (Bissonnette and Roy, 1992; Brown *et al*., 1996; Paulsen *et al*., 1993). Thus, integrons do not necessarily include a gene cassette (Bissonnette and Roy, 1992; Hansson *et al*., 1997), but when integrated the cassette is a functional part of the integron. There are different types of *intI* genes and each type is associated with a cognate *attI* site. These two components together with promoters make up different 5'-conserved segments that form different integron classes.

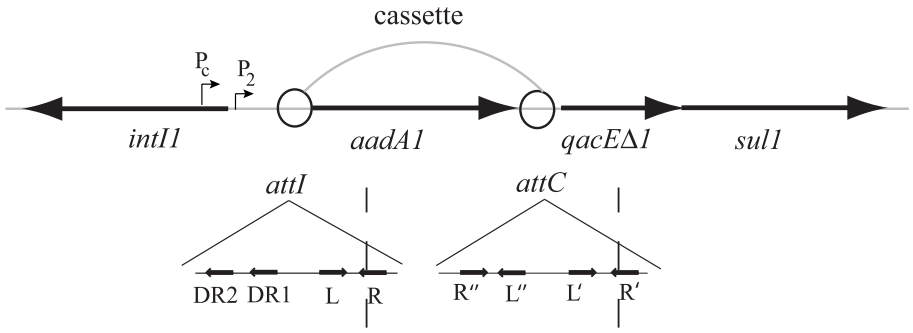

Figure 1: A schematic picture of the organization of the integron in Tn*21* (X12870), carrying the *aadA1*cassette encoding streptomycin/spectinomycin resistance and the following *qacEΔ1-sul1* combination downstream of *attC* (Sundström *et al*., 1988; Paulsen *et al*., 1993). The integron integrase is encoded in the 5'-conserved sequence, to the left. Genes are shown in bold face horizontal arrows and recombination sites as open circles. The curve shows the extent of the *aadA1* cassette. P_c and P_2 are promoters used for expression of integrated gene cassettes. In the lower panel the two primary recombination sites *attI* and *attC* are enlarged to visualize their different repeat organization. In the *attI* site, two direct repeats (DR2 and DR1) are found upstream of two inversely oriented core repeats (L and R) that constitute a simple site and include the recombination cross-over point. The *attC* site on the other hand is almost a perfect palindrome organized in two subsites of dyad symmetry (repeats R" and L" and repeats L' and R'). Vertical dashed lines mark cassette borders.

Since the discovery of chromosomal integrons (superintegrons) (Mazel *et al*., 1998) it is rational to divide the integrons in two major groups: floating integrons (also called resistance integrons, RI; multiresistance integrons, MRI or mobile integrons due to their connection to mobile elements (Mazel, 2006)) and superintegrons. Most integrons in these two groups share an

identical organization with a 5'-conserved region encoding an integrase and a variable region of gene cassettes in the 3'-end (Rowe-Magnus and Mazel, 1999). Most cassettes that have been observed on floating integrons encode resistance to antibiotics and disinfectants. They can be located on conjugative plasmids but also in islands on the chromosome where they are commonly associated with other mobile DNA elements (such as insertion sequences (IS), transposons, and ICEs). Floating integrons typically contain less than six gene cassettes (Rowe-Magnus *et al*., 2002). The floating integrons are divided into classes based on the homology of their integrase genes (Recchia and Hall, 1995b). All members of a given integron class also include the same *attI* sequence and differ only in location and in the cassettes they encode. Five classes of floating integrons are known today (Sørum *et al.*, GenBank accession no. AJ277063; Arakawa *et al.*, 1995; Hochhut *et al*., 2001; Sundström *et al*., 1988; Hall and Vockler 1987; Hansson *et al*., 2002). All integrons independently of their class seem to use the same type of cassettes and this is best illustrated by the *dfr1* cassette found in all five examples (Arakawa *et al.,* 1995; Sundström *et al.*, 1988; Hochhut *et al*., 2001; Sundström and Sköld, 1990; Sørum *et al*., unpublished). Class I integrons are the most prevalent integrons on plasmids and were the first described. These integrons are commonly found in multiresistant gram-negative bacteria. A frequently detected cassette contains *aadA1*, a gene encoding an adenylyl transferase mediating streptomycin-spectinomycin resistance. Trimethoprim resistance determinants, encoding dihydrofolate reductases, as well as gene cassettes encoding β-lactamases are also common. The conserved 3'-segment of most class I integrons contains the *qacEΔ1* gene (encoding an efflux mechanism mediating resistance to quaternary ammonium compounds; Paulsen *et al.*, 1993), the *sulI* gene (encoding dihydropteroate synthase causing resistance to sulfonamides; Sundström *et al.*, 1988) and an open-reading frame (*orf5*) of unknown function (Stokes and Hall, 1989). Class 1 integrons are generally borne on elements similar to Tn*5090/402* or truncates thereof (Brown *et al.*, 1996; Rådström *et al.,* 1994). Tn*21* is a large mercuric ion resistance element that contains one of these truncated forms (Rådström *et al.*, 1994). Class 2 integrons are found on Tn*7,* a non-replicative member of the Mu family of transposons (Hansson *et al*., 2002; Sundström *et al.,* 1988; Sundström *et al*., 1991). Class 3 integrons were originally found on a transferable large plasmid in *Serratia marcescens* (Arakawa *et al*., 1995) and probably in a transposon related to Tn*402* (Collis *et al*., 2002a). Class 4 (also known as IntI9) was discovered on a novel self-transmissible genetic element designed the SXT^{ET} element that is integrated on the large *Vibrio cholerae* chromosome (Hochhut *et al*., 2001). Class 5 was found in *Vibrio salmonicida* on plasmid pRVS1 (Sørum *et al.*, unpublished data; accession AJ277063). Although floating integrons are widely spread among gram-negative species, there are a few cases where class 1

integrons have also been found in gram-positive bacteria (Mazel, 2006 and references therein).

The chromosomally located superintegrons are on average much larger than the floating integrons and may contain hundreds of gene cassettes often with unknown functions (Mazel *et al*., 1998; Rowe-Magnus *et al*., 2002; Rowe-Magnus *et al*., 2001). The coevolution of superintegron integrases with their host genome suggests that superintegrons are sedentary. The first discovered superintegron from the *Vibrio cholerae* genome (Mazel *et al.,* 1998) was originally named class 4 and encodes the integron integrase VchIntIA. Today more than 40 species of proteobacteria have been found to contain a superintegron (Rowe-Magnus *et al*., 2002; Mazel, 2006).

Perhaps the most distinctive difference between floating integrons and superintegrons is the appearance of the *attC* sites – the recombination sites separating the contiguosly integrated gene cassettes. In floating integrons the *attC* sites are highly variable both with respect to their length and sequence while the *attC* sites in each superintegron are closely related. However, *attC* sites connected to superintegrons in different species vary both in length and sequence but share the same organization of internal repeats (Mazel, 2006) (Figure 2).

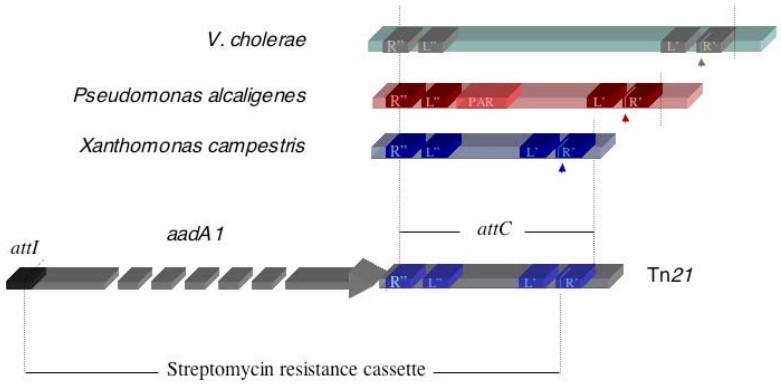

Figure 2. Schematic representation of *attC* sites from different superintegrons and the *attC* site found in Tn*21*. The diagram shows that cassette sites (*attC*) are species-specific and vary in length. The relation between superintegrons and floating integrons is visualized by the cassettes in blue that have been found to be identical in the superintegron in *Xanthomonas campestris* and in the floating integron of Tn*21* (Barker *et al*., 1994; da Silva *et al*., 2002; Johansson *et al*., 2004; Mazel *et al*., 1998; Rowe-Magnus *et al*., 2001; Vaisvila *et al.,* 2001).

Integron integrases, IntIs

By homology and conservation profile the integron integrases may be easily added to the long and growing list of tyrosine family site-specific recombinases (Sundström *et al.,* 1988; Nunes-Düby *et al*., 1998; Ouellette and Roy, 1987). Sequence alignment shows that the six highly conserved residues that are part of the catalytic site in tyrosine recombinases are also present in integron integrases. These residues are an arginine in box I, two histidines, an arginine and a tyrosine in box II and a lysine in patch II (for alignments see Collis *et al*., 2002b). Apart from these conserved residues, integron integrases have a conserved stretch of an additional 16-36 conserved amino acids (different length assumptions: Messier and Roy, 2001; Nield *et al*., 2001; Recchia and Sherratt, 2002) that distinguish them from other tyrosine recombinases. This region is located between patch II and patch III in the C-terminal catalytic domain and is implicated in the recognition of the *attI* and *attC* sites (Messier and Roy, 2001; Paper III). This year the first three-dimensional structure of an integron integrase was presented (at 2.8 angstrom resolution). The analyzed complex contained integron integrase VchIntIA from *Vibrio cholerae* bound to its *attC* substrate $VCR_{bs.}$ The structure shows protein-DNA contacts that highlight the role of the inserted amino acid sequence in the protein for synapse formation (MacDonald *et al*., 2006). Within this region of the polypeptide lies α-helix I_2, an additional secondary structure to the otherwise similar C-terminal organization of Cre with 9 helices, a three-stranded beta sheet and a β-4,5 hairpin (MacDonald *et al.,* 2006; Yang och Mizuuchi, 1997).

More than 40 different types of integron integrases have been identified (Mazel, 2006 and references therein). For floating integrons, the number assigned to each integron class corresponds to the number assigned to the *intI* gene and IntI integrase. For example, class 1 and class 3 integrons encode IntI1 and IntI3 integrases, respectively. When two distinct integron integrases are more than 98 % identical they belong to the same integron class (Collis *et al*., 2002a and references therein). However, for superintegrons the general rule is to name the integrase with its host initials, for example VchIntIA from the *Vibrio cholerae* genome, although numbering of these genetically fixed *intI* genes also occurs in the literature. Integron integrases comprise about 300-350 amino acids. The integrase genes among the five classes of floating integrons share between 39 % and 58 % amino acid sequence identity (Rowe-Magnus *et al*., 2002. However, all the integron integrases described to date (including all known integron integrases of superintegrons) are about 34-94 % identical or 57-96 % similar based on pairwise comparisons (Collis *et al*., 2002a).

Integron integrases are the only known proteins to date described as being required for the movement of gene cassettes. They catalyze both integrative and excisive recombination as shown experimentally *in vivo* in the case of

class 1, 2, 3, 5 and the chromosomal integrons of the *Shewanella oneidensis, Nitrosomonas europaea, Pseudomonas stutzeri* and *Vibrio cholerae* genomes (Abremski and Hoess, 1984; Collis *et al*., 1993; Collis and Hall, 1992a, 1992b; Collis *et al*., 2002a, 2002b; Drouin *et al*., 2002; Hall *et al*., 1991; Hall and Collis, 1995; Hansson *et al.,* 2002; Holmes *et al*., 2003; Leon and Roy, 2003; Martinez and de la Cruz, 1988, 1990; Recchia *et al.,* 1994; Rowe-Magnus *et al*., 2001), Sørum *et al.*, unpublished).

Gene cassettes

The gene cassette is the mobile component of the integron system (Recchia and Hall, 1995b) but contrary to most entities that are comparable from a mechanistic perspective, it does not encode the protein that catalyzes its own movement. Instead the mobility of cassettes is dependent on the integron integrase IntI that is encoded outside of the cassette. IntI interacts with two primary recombination sites, *attI* and *attC*, (Collis *et al.,* 1998 and 2002b; Gravel *et al.,* 1998; Bouvier *et al*. 2005, Paper 1) and the outcomes are excision and integration of cassettes (Hall *et al*., 1991; Collis *et al*., 1993; Collis and Hall, 1992a, 1992b). Usually each gene cassette contains only a single gene and an *attC* site, located downstream of the gene. The gene cassettes are generally named by the gene that they contain. Many of the antibiotic resistance genes found in gram-negative bacteria are part of gene cassettes and the first ones discovered were from antibiotic-resistant clinical isolates. More than 80 different antibiotic resistance cassettes have been identified in class 1 integrons to date, encoding resistance to a broad spectrum of drugs such as all known aminoglycosides, all β-lactams, chloramphenicol, trimethoprim, erythromycin, streptothricin, rifampin, antiseptics and disinfectants (Hall and Collis, 1995, 1998; Mazel and Davies, 1999; Mazel *et al.,* 1998; Rowe-Magnus *et al.,* 1999; Fluit and Schmitz, 2004; Mazel, 2006). In contrast, class 2 integrons have only been associated with six different resistance cassettes (Mazel, 2006). We know today that the antibiotic resistance cassettes are a minor proportion of all cassettes represented in bacterial integrons (Mazel, 2006). Cassettes encoding a diverse range of other functions and several open reading frames (ORFs) of unknown function have been found by the discovery of superintegrons and the estimated total number of cassettes have shifted from the tens to thousands (Hall *et al*., 1991; Rowe-Magnus *et al.,* 2002, Mazel, 2006). Furthermore, it seems likely that in principle any gene can be packaged and mobilized as a cassette.

Gene cassettes vary considerably in length from 262-1549 bp, largely due to differences in the size of the coding part of the gene (Recchia and Hall, 1995b). They are most commonly found in an integrated form but free covalently closed circular molecules have been observed as products of excision (Collis and Hall, 1992b). The cassette circles have been postulated to be important intermediates in the spread of genes (for example resistance

genes) from one integron to another as well as occasionally to non-specific sites. The data presented recently and in this work show that the picture concerning the state of the free intermediate is not yet complete.

Most gene cassettes appear to contain a translational initiation region but although a systematic search is not reported it seems as most cassettes lack a functional promoter (Recchia and Hall, 1995b). Almost all cassettes integrated into an integron are dependent on the promoter, P_c in the 5'-CS of the integron (Levesque, 1994), for their expression and are located in the same orientation. The gene encoding the integron integrase and integrated gene cassettes are oriented in opposite directions with perhaps negligible exceptions (Stokes and Hall, 1989; Coleman *et al*., 2004) (Figure 1). When cassettes are integrated at secondary sites outside an integron context, expression is dependent on the presence of a suitably oriented external promoter (Recchia and Hall 1995b; Segal and Elisha, 1997; Francia *et al*, 1993). In class 1 integrons five variants of P_c are known (Bunny *et al*., 1995; Stokes and Hall, 1989; Sundström *et al*., 1988) and these have been observed to differ in strength over at least a 20-fold range (Collis and Hall, 1995; Levesque *et al*, 1994). In Tn*21* and a few other integron variants, a second promoter P_2 has been created (Sundström *et al.*, 1988) (Figure 1). P_2 is a strong promoter and supports the level of expression when the weakest version of P_c is present. Several cassettes in tandem may be co-transcribed from P_c and long transcripts covering several cassettes have been observed. However, the majority of cassette transcripts are shorter (Collis and Hall, 1995, 2004). Transcripts generally start at P_c but end at points that appear to correspond to the ends of cassettes. However, in general the first cassette closest to the P_c and the 5'-CS of the integron is expressed at the highest level due to premature termination of transcripts (Collis and Hall, 1995). Rearrangement of gene cassettes through excision and reintegration at the favoured *attI* site in the 5'-CS can lead to higher expression of previously distal and weakly expressed genes (Collis and Hall, 1992a; Rowe-Magnus and Mazel, 2001).

Recombination sites

attI

The *attI* site is the preferred target site for integration of gene cassettes (Collis *et al*., 1993; Hansson unpublished; Hall and Collis, 1995). It is adjacent to the *intI* gene and includes the end of the class-specific 5'-CS and the core repeat in the beginning of the first cassette in an integron (Hansson *et al*., 1997). Thus, *attI* occurs only once per integron. The *attI* sites are characteristic for integrons of a particular class but share only limited sequence identity among different classes. The *attI1* site of class 1 integrons is the most extensively studied site of the target site class (Collis *et al*., 1998; Gravel *et al*., 1998; Hall *et al.,* 1999; Hansson *et al.,* 1997; Partridge *et al.*, 2000; Rec-

chia *et al*., 1994). Recently detailed studies of *attI3* were reported (Collis and Hall, 2004). The studies have revealed that *attI1* is composed of four integrase binding sites, two of which are inversely oriented and constitute a simple core site that includes the recombination cross-over point (Figure 1). The additional two IntI1-binding sites are found upstream of the simple site and they are direct repeats known as DR1 and DR2, respectively (also known as strong and weak). Each of the four single binding sites contains GTTRRRY or variations thereof. A full (65-70 bp) *attI1* site is required for efficient recombination with *attC in vivo* (see Paper II; Hansson *et al.,* 1997 och Recchia *et al.,* 1994). DR1 (closest to the simple site) and the simple site are essential for recombination, while DR2 further enhances the recombination efficiency. However, in recombination with a second complete *attI1* site, the simple site of *attI1* is sufficient (Hansson *et al.,* 1997). Recombination occurs in the right-hand repeat of *attI1* (Martinez and de la Cruz, 1990; Recchia and Hall, 1997) and the cross-over point has been localized between the G and TT (Hansson *et al.,* 1997) (Figure 1).

An alignment of *attI1* and *attI3* reveals that DR1 and DR2 are in equivalent positions with respect to the simple site. The finding that the organization and properties of *attI3* are very similar to those of *attI1* indicates that these features are likely to be common to all *attI* sites (Collis and Hall, 2004). The approximate spacing of 20 bp between the integron integrase contact sequences is maintained in the integron in Tn*1696* that introduces an additional DR on a 20 bp insertion. Furthermore, the region 5' of the integrase gene forms a very long but weak palindrome (Sundström personal communication).

The differences in sequence among *attI* sites suggest that IntI may recognize only the *attI* site found adjacent to the *intI* gene (the cognate *attI*). It has been shown by Collis *et al.* (2002b) that IntI proteins exhibit strong specificity for the cognate site *in vivo* and that the requirements are more stringent in integration reactions than in excision reactions. In studies by Collis *et al*., (2002b) and Hansson *et al.* (2002), IntI1 appeared to have a greater ability to recognize non-cognate *attI* sites in excision reactions than either IntI3 or IntI2*179E. The ability of IntI1 and IntI3 to distinguish between *attI1* and *attI3* seems to be determined by differences in the sequences of one or more of the four IntI-binding sites. These differences are likely to reside in the left-hand repeat of the simple site and in DR1 since the strongest binding of IntI3 to *attI3* has been observed in these two repeats (Collis and Hall, 2004) (Figure 1).

In order to understand the mechanisms in the recombination reactions involving *attI* it is important to mention the work by Francia *et al.* (1999). They observed binding of IntI1 to a both double-stranded (ds) and single-stranded (ss) form of *attI1*. The interaction between IntI1 and ss *attI1* was specific for the bottom strand. In addition, the N-terminal domain of IntI1 was required for binding to both forms of *attI1*. However, it remains un-

known how many molecules of IntI1 protein were bound to the ds or ss form of *attI1*. In addition, because a PCR method was used to generate the single-stranded fragments in Francia *et al.,* (1999) small amounts of both singlex and duplex fragments could have been present. Importantly, recent work by Bouvier *et al*., (2005) shows that IntI1 interacts with ds *attI1* but that IntI1 could not bind to ss *attI1*. Previous studies by Gravel *et al.*, (1998), observed four distinct complexes of IntI1 bound to ds forms of *attI1*. Thus, conflicting data supports various models of IntI1-*attI1* complex formation.

Vanhooff *et al.,* (2006) have shown that the tyrosine recombinase TnpI of Tn*4430*, binds to a res site (to resolve transposition intermediates) similar to *attI* to control pairing of the recombination sites and strand exchange. It acts both as catalytic component, and as a regulatory element in the recombination complex to ensure selective interaction between separate recombination sites (see also Mahillon and Lereclus, 1988).

attC

The cassette-associated recombination site *attC* (the terms *attI* and *attC* were introduced by Hansson *et al.,* (1997) is located downstream of the gene in each cassette and was first identified as imperfect- inverted repeat sequences that were related to a consensus known as a 54-base element (be) or 59be (Wiedemann *et al*., 1986; Fling *et al*., 1985; Cameron *et al*., 1986). The *attC* sites are known to vary considerably both in length (from 57-141) and in sequence (Hall *et al*., 1991; Sundström *et al*., 1993; Recchia and Hall, 1995b). It is not well understood how different *attC* sites are accurately recognized by one particular class of integron integrases (Arakawa *et al*., 1995; Recchia and Hall, 1995b; Sundström *et al*., 1988, 1991; Sundström and Sköld, 1990). Notably, IntI proteins that share less than 40% amino acid identity can also efficiently recognize the same *attC* site (Mazel *et al.,* 1998).

Alignments of different *attC* sites have revealed a conserved organization of four short symmetrical sequences that form dyad pairs on either end (Francia *et al*., 1997; Stokes *et al*., 1997) (Figure 1). For simplicity reasons we call these pairs subsites. The inner of the four repeats (L" and L') form imperfect inverted repeats of the flank repeats (R" and R'). The central part between L' and L" varies both in length and sequence among *attC* sites but shows a constant inverted repeat character. With its defined central symmetry axis *attC* forms a large imperfect palindrome. Previously reported *in vivo* data (Hall *et al.,* 1991) indicated the main cross-over point was located between the G and TT in the consensus sequence GTTRRRY in the 3' end of the site. Thus both *attI* and *attC* are composite sites with the last 6 bases of the consensus sequence supplied by the adjacent cassette.

The two subsites, R"L" and L'R', in *attC* resemble simple sites recognized by other tyrosine recombinases that are made up of two inverse repeats separated by a spacer of 6-8 bp. However, neither of the two subsites in *attC* is an efficient recombination site alone and it has been shown that a full

sized *attC* is required to make a proficient site function (Martinez and de la Cruz, 1990; Paper II). In their combined context IntI recombinases distinguish between the two subsites, since recombination exclusively almost occurs within the right-hand subsite. The left-hand subsite has been suggested to form a second alternative simple site or accessory binding site (Francia *et al*., 1997; Recchia and Sherratt, 2002; Stokes *et al*., 1997). Stokes *et al.* (1997) have postulated four putative IntI1 binding domains of *attC*, two at each end.

Recombination positions have been observed in all other repeats except for the L" repeat (Stokes *et al.*, 1997; Paper II). It has been suggested that imperfections in the symmetry of the left-hand end of the *attC* site and an additional base in one of the two repeats make it possible for the integrases to distinguish between the two simple sites (Hall *et al*., 1991). It is likely that the differences in both halves of the *attC* site play a role in ensuring correct orientation of inserted gene cassettes which allows expression of the associated gene from P_c. Use of the left-hand simple site would integrate the gene cassette in an opposite orientation and the cassette must in that case be expressed independently of the normal integron promoter P_c. The mechanism for orientation discrimination appears to be determined by *attC* but has until recently remained to be established (Hall *et al*., 1991). Serine recombinase sites have also been reported to benefit from minor "imperfections" in the matching between the site repeats to define the site orientation (Blake *et al*., 1995).

The binding of integrase to *attC* has been difficult to study by standard techniques and early attempts to obtain complexes of IntI1 bound to ds *attC* in order to map the IntI1 binding sites have failed (Mats Gullberg personal communication; Collis *et al*., 1998; Francia *et al.,* 1999; Gravel *et al.,* 1998; Stokes *et al*., 1997). However, a breakthrough was made in 1999 by Francia *et al.* By using PCR-derived DNA fragments produced by using an excess of either of the primers, it was demonstrated that IntI1 binds strongly and specifically to the bottom strand of single-stranded *attC* DNA but not to the top strand. The integrase was observed to bind to single-stranded DNA with amino acids residing in the C-terminal domain. The same study also reports footprinting data on two of the four repeats. In total a region of at least 40 bp including the cross-over point, was most efficiently protected by IntI1 against DNase1. The authors suggested that single-stranded DNA binding might be a significant activity of IntI1 integrase. Formation of stem-loops structures has been proposed to be of importance for the function of the *attC* site in recombination reactions (Francia *et al.,* 1997; Francia *et al*., 1999; Hall *et al.*, 1991; Stokes *et al*., 1997; Sundström *et al*., 1988).

Recent studies (Johansson *et al*., 2004 (paper I); Bouvier *et al*., 2005; Macdonald *et al.,* 2006) report in further detail single-stranded recognition of *attC* and highlight the importance of the asymmetrical nucleotides within the site. A proposed recombination integration model is presented in which

the bottom strand of *attC* is folded into a hairpin with bulges generating a pseudo double-stranded recombination site. In strong support of the suggested model, the recently completed crystal structure of the integrase complexed on a representation of an *attC* hairpin from the *Vibrio cholerae* integron, was presented in 2006 (MacDonald *et al*.,). Details are further explained and discussed in the section on the mechanisms of integron integrases.

Secondary sites

Besides acting on the primary sites *attI* and *attC*, IntI1 has been found to mediate site-specific recombination between a primary site and a secondary site at low frequency (Francia *et al*., 1993; Francia and Lobo, 1996; Hansson *et al.,* 1997; Recchia *et al*., 1994). A few cassettes situated outside the integron context have been discovered (Recchia and Hall, 1995a; Segal and Elisha, 1997). These secondary sites contain patches of sequence homology with either of the primary sites, *attI* or *attC*. The defined consensus GWTMW (or GNT; Recchia *et al*., 1994) has some similarity to the core repeat (R') in the primary sites where the cross-over occurs (Francia *et al*., 1993). The inserted gene cassette needs to be correctly oriented with respect to a pre-existing promoter for gene expression since gene cassettes generally lack promoters. Integration of gene cassettes at secondary sites is likely to be stable since the hybrid sites that flank the cassette after insertion in a secondary site are probably insufficient for supporting excision. Excision of a gene cassette from a secondary site has not been observed (Recchia and Hall, 1995b and references therein). Thus, through integration and consequent fixation of gene cassettes at secondary sites, genes can be stably moved to new locations. Recombination at secondary sites may play an important role in the evolution of bacterial and plasmid genomes.

Recombination events

Integron integrases mediate site-specific recombination between two sites where the combinations *attI* x *attC*, *attC* x *attC*, *attI* x *attI* and *attI* / *attC* x secondary sites (2rs) are all possible but yield products at widely different frequencies. Integrative recombination events have been observed with all the above combinations whereas excisive events have been clearly documented when involving *attI* x *attC* and *attC* x *attC* (Collis *et al.*, 1993; Collis and Hall, 1992a, 1992b; Collis and Hall, 2004; Collis *et al.,* 2002a, 2002b; Collis *et al*., 2001; Drouin *et al*., 2002; Francia *et al.,* 1997; Francia *et al*., 1993; Francia and Lobo, 1996; Hall *et al*., 1991; Hansson *et al*., 1997; Hansson *et al.,* 2002; Leon and Roy, 2003; Martinez and de la Cruz, 1990; Recchia *et al*., 1994; Biskri *et al*., 2005). The most common combinations of sites used for cassette integration and excision are *attI* x *attC* and *attC* x *attC*. A new cassette is preferably integrated into the *attI* site, thereby ensur-

ing its maximal expression from the integron promoter, P_c. Reactions involving *attI* x *attI* and 2rs x primary sites occur much less frequently but could have evolutionary significance. The movement of cassettes between integrons could follow an alternative pathway such as via formation and subsequent resolution of plasmid replicon fusions using different recombination sites (Hall and Collis, 1995). In all recombination reactions observed crossover has been detected in the right-hand end of the sites (Figure 1) (Hansson *et al.,* 1997; Stokes *et al*., 1997).

As mentioned earlier, different integron integrases can recognize the same *attC* sites and thereby move cassettes from one integron class to another. However cassette integration at *attI in vivo*, occurs only when the integron integrase and *attI* site belong to the same integron class. The requirement for a cognate *attI* site in integrative reactions involving *attI* and *attC* is strict. However, this strict requirement does not hold true for IntI1 in excisive reactions (Collis *et al*., 2002b; Hansson *et al*., 2002). Although at low frequency, the integron integrase IntINeu from *Nitrosomonas europaea* has also been observed to perform deletions in which a non-cognate *attI* is one partner site (Leon and Roy, 2003).

Hansson *et al.* (2002), have shown results suggesting that the orientation of a cassette relative to the surrounding sequence influences the rate of excision. The effect is not fully elucidated but could either be due to the direction and strength of transcription or to the passage of replication forks. However, since we now know that only one strand of the *attC* site is bound by the IntI recombinase (Francia *et al*., 1999; Johansson *et al*., 2004), a third explanation is possible. It is conceivable that a DNA segment harbouring an integron cassette is subject to replication in such a way that one DNA strand is more abundant than the other strand. This occurs at nick-primed replication for instance at conjugative transfer of DNA. Because only the bottom strand forms a hairpin that is eligible for recombination inverting the site may turn recombination on or off.

Origin

Gene cassettes as well as the *intI* genes and *attI* sites found in floating integrons are believed to originate from superintegrons and there are numerous findings supporting this logical theory (Rowe-Magnus *et al.,* 2001; Mazel, 2006). It is well documented that *attC* sites associated with cassettes found in floating integrons are highly diverse while the sequences of *attC* sites from superintegrons are closely related and species-specific. Each gene cassette in a floating integron probably represents a single member harvested from one of the many different cassette pools of superintegrons and selected based on the environmental conditions under evolution. Moreover, it has been observed that integrases from class 1 integrons can recruit cassettes from the superintegron found in *Vibrio cholerae* by recognizing the *attC*s

(named XXRs after the species in which the superintegron resides, here VCRs) of the superintegron cassettes (Mazel *et al*, 1998; Rowe-Magnus *et al*., 2002; Biskri *et al*, 2005). Two gene cassettes *CARB4* and *dfr6* were also found to contain similar *attC* sites to VCRs (Mazel *et al*., 1998). Rowe-Magnus *et al*., (2001) found that 20% of the *attC* sites of the antimicrobial cassettes were virtually identical in length and sequence to the highly related *attC* sites/XXRs in some of the superintegrons. Gene cassettes within the same integron sometimes have marked differences in codon usage while their *attC*s are highly homologous. This observation suggests that the genes have different origins and that the XXRs / *attCs* were added to the genes in a cryptic cassette genesis process inside the species harbouring the superintegron (Rowe-Magnus and Mazel, 2001).

Dendrograms based on the 16S rRNA gene and the *rplT* gene (coding for the ribosomal L20 protein) made by Rowe-Magnus *et al.,* (2001) show that the extent of divergence between the *intI* genes adhere to the line of descent among the bacterial species and correspond with environmental niches of respective host microorganisms (Mazel, 2006). The authors also observed coevolution of the superintegron integrase genes and their cognate sites (XXRs). The XXRs of closely related *intI* genes were found to share extensive similarity. They propose that there is an integron in the ancestral organism of each genus. In other words the integrons are believed to be ancient structures that may have been involved in the evolution of bacterial genomes for hundreds of millions of years (Rowe-Magnus *et al*., 2001; Rowe-Magnus and Mazel, 2001; Vaisvila *et al*., 2001). The fact that VCR cassettes have been found in a strain of *Vibrio metschnikovii* isolated in 1888 (Mazel *et al*., 1998), provides evidence that the integron structures at least pre-date the short antibiotic era that began with the introduction of sulfonamides in the 1930's (Sköld, 2000).

Hall *et al.* (1991) have suggested that the gene cassettes originate from retrotranscribed RNA transcripts based on the observation that the gene cassettes are flanked by very little non-coding sequences. Another argument supporting this theory highlights the fact that the *attC* sites resemble transcription terminators and also possess such a function. It is possible, though, that these sites were added from some other structure (Hall and Collis, 1995). Francia *et al.*(1999) point out that integron integrases should be able to bind ssDNA as a preliminary step for integration if gene cassettes originate from RNA transcripts. In support for this theory, the authors were able to show binding to single-stranded DNA as either *attI* or *attC*.

Finally, the mechanism of cassette genesis remains under speculation (Recchia and Hall, 1997; Hansson *et al.,* 1997; Centron and Roy, 2002), and the original role of integrons in the pre-antibiotic gene transport systems still awaits elucidation. Integrons may, for instance, have been gene factories, gene deposits or hubs in metagenomic transport roads. However, presently the complete picture remains unclear. It is interesting to postulate what hap-

pened when a number of integrons from the vast integron pool detached from their original fixed position among chromosomal housekeeping genes and came adrift. It is not understood whether these laterally transferred entities were unique or whether they were fortuitously selected due to their repertoire of antibiotic resistance genes and thus became established. We know that certain transposable elements have a more intimate relation to integrons than others. Two widely diverged Mu-like elements, Tn*402* (Tn*5090*) and Tn*7* appear to have captured their class 1 and 2 integrons early on. For what reasons are these two elements established to take up combined movable and gene capturing roles on the metagenomic web? Why are both elements Mu-related and why are they not instead Tn*3* elements that appear to be more widespread? Why are integrons so rarely borne on compound transposons flanked by IS elements? Many questions remain unanswered.

Mechanisms of strand exchange of integron integrases in comparison with other tyrosine recombinases

Our understanding of IntI-mediated site-specific recombination has recently made fast progress but pieces are still missing on a detailed, mechanistic level, particularly concerning the terminating steps of the reaction. The new recombination model now supported by the crystal structure of an integron integrase (presented by Bouvier *et al.,* 2005; MacDonald *et al.*, 2006) shows the involvement of a single-stranded *attC* site. However, since it is folded into a hairpin it still fits most classical characteristics of tyrosine recombinase sites because the hairpin may be regarded as a pseudo double-stranded recombination site. In addition, the integron integrases contain the hallmark of six conserved residues in the C-terminal domain. Therefore it seems inevitable that IntI recombinases also share many aspects of the catalytic mechanism with Cre and other well-characterized members of the family (see earlier chapters about tyrosine recombinases and their mechanism). Nevertheless, it is now clear that the integron system differs in a number of ways. First of all, differences arise in the unusual site organization. A typical recombination reaction performed by tyrosine recombinases involves two identical or almost identical simple recombination sites, each consisting of two recombinase-binding core elements of about 11-13 bp, inverted relative to each other and separated by 6-8 bp long central region/spacer (Sadowski, 1986; Sadowski, 1993; Stark, 1992). Each recombinase system contains its own specialized recombinases that mediate the rearrangement of DNA. Many tyrosine recombinase systems use additional accessory proteins for regulation of the recombination reaction. Recognition sites for these proteins are included in the recombination site organization and complement the simple (core) site where the strand cross-overs occur (Colloms *et al*., 1998; Dorgai *et al*., 1998; Nash, 1996). The primary sites recognized by integron

integrases, *attI* and *attC,* are complex and too varied to be recognized by a single protein. Furthermore, the site organization differs from other tyrosine recombination sites. Only the conserved 7 bp sequence, GTTRRRY is found in all *attI* and *attC* sites (Stokes *et al*., 1997). The *attI* sites consist of a simple site downstream of two directed repeats. The *attC* sites are unusual in that they are composed of two simple sites flanking a palindromic central region, while the less frequently used secondary sites do not seem to constitute any type of common site structure at all (see earlier chapters about the different sites). Moreover, the involvement of non-integrase accessory proteins in the recombination reaction appears to be limited (Biskri *et al*., 2005) and the integron system is suggested to utilize additional molecules of their own integrases for accessory functions.

It is remarkable how a single integron integrase can recognize and perform recombination between two sites that are so varied in organization and/or sequence. For comparison, other tyrosine recombinases have been tested for their ability to recognize sites of related enzymes without success. Only substitution and replacement of residues implicated in site discrimination in the two different recombinases tested, or substitution of specific bases in their recombination sites relieved discrimination against the cognate site (Dorgai *et al*., 1998; Dorgai *et al.*, 1995; Gottfried *et al*., 2000; Yagil *et al*., 1995). One could postulate whether integron integrases possess special features allowing for its wider and more flexible site-recognition qualities. As described previously in this thesis, integron integrases share an additional conserved region that is not found among other tyrosine recombinases (Messier and Roy, 2001). This conserved region is known to be required for recombination activity (Messier and Roy, 2001), and the crystal structure of VchIntIA-VCR_{bs} indicates that this region plays a clear role in synapse formation (MacDonald *et al*., 2006).

Another difference between integron integrases and other tyrosine recombinases is the unique ability of integron integrases for mediating site-specific recombination at sites that contain heterologous central regions between the conserved inverted core repeats (Stokes *et al.,* 1997). Sequence identity of the spacer bases in the two recombining sites is critical for progression the recombination in most tyrosine recombinase systems. This has been proposed to relate to the requirement for branch migration of a Holliday junction throughout the spacer or, alternatively the more favoured explanation, strand-swapping (Nunes-Düby *et al*., 1995). Introduction of heterology dramatically reduces the recombination efficiency either by impairing first strand cleavage and rejoining, or by causing the reactions to stall at the Holliday junction intermediate stage (Abremski and Hoess, 1984; Dorgai *et al*., 1998; Dorgai *et al*., 1995; Recchia and Sherratt, 2002).

According to the classical paradigm the integron integrases were supposed to initiate first strand cleavage and form a Holliday junction intermediate similar to typical tyrosine recombinases. The lack of requirement for

central region homology, however, lead to suspicions that IntI1 catalyzes only one DNA strand exchange (Hansson *et al.,*1997; Stokes *et al*., 1997). Strand exchange in the integron system has been confined between two adjacent nucleotide pairs at one end of the spacer in the consensus repeat furthest to the right in both *attI* and *attC* (Hansson *et al.,*1997; Stokes *et al*., 1997). The question has been adressed whether a second strand exchange occurs at precisely the same position as the first strand breakage or, if the reaction halts and is followed by resolution of the resulting Holliday junction intermediate by host-encoded resolvases present in the cell (West, 1994; White *et al.*, 1997). The reaction may otherwise possibly be completed by replication through the Holliday junction intermediate and thereby generate both substrate and recombinant product (Stokes *et al*., 1997). By analogy with other tyrosine recombinases, it was suggested that the integron integrase-mediated cleavage occurs between the A and C on the bottom strand (Recchia and Sherratt, 2002). It was also proposed that a staggered strand exchange is possible since the nucleotides on either side of the central regions are conserved, but that the heterology and strong purine bias in the central regions will limit the reaction to only one strand exchange (Recchia and Sherratt, 2002).

Integron integrases differ from other tyrosine recombinases not only in site-recognition but potentially also in their mechanism of strand exchange. Recent work by Johansson *et al*., (2004) (I), Bouvier *et al.,* (2005) and the crystal structure of VchIntIA-VCR_{bs} by MacDonald *et al.,* (2006) suggest that integron integrases have an incomplete mechanism of strand exchange to other well-known tyrosine recombinases. Despite considerable sequence divergence the *attC* sites are almost perfect palindromic. It has been previously implicated that this palindromic feature may be important in the recombination reaction, such as the formation of *attC* cruciforms (Francia *et al*., 1997; Francia *et al*., 1999; Hall *et al*., 1991; Stokes *et al*., 1997; Sundstrom *et al*., 1988). A structural similarity between extruded cruciforms and the Holliday junctions formed by most tyrosine recombinases fostered speculations that the former could be site-recognition determinants for this group of proteins (Recchia and Sherratt, 2002; Johansson *et al*., 2004 (I).) The recent studies confirm binding of integron integrase to the bottom strand of *attC*, and take into special account that only one strand is exposed under conjugative transfer of DNA. The importance of extrahelical bases in a folded hairpin structure of the bottom strand of the site is highlighted. Using a suicide conjugation assay in which the transferred plasmid is unable to replicate in the recipient cell, Bouvier *et al.,* (2005) measured recombination frequency based on cointegrate formation between ss circular DNA and a plasmid also containing a recombination site. From these data the authors presented a cassette integration model where a single-stranded region of the cassette including the *attC* site (bottom strand) is folded into a hairpin mimicry of a double-stranded recombination site. This *attC* substrate recombines

with a double-stranded *attI* generating a Holliday junction that must be subsequently processed to recombination products by yet unknown cellular factors. The suggested recombination reaction is similar to site-specific recombination catalyzed by other tyrosine recombinases up to the Holliday-junction intermediate. Recombination between two folded *attC* sites and subsequent replication could form a single-stranded cassette circle and two daughter molecules: one with the cassette deleted and the other retaining a full copy (Bouvier *et al*., 2005; see also Chandler, 2006).

Recombination events catalysed by λ integrase, Cre, FLP, XerC/D and others all involve two sequential strand exchanges and rejoining reactions that occur at either end of the central region. A pseudo four-fold symmetrical intermediary Holliday junction must change conformation to allow for the next strand exchange event to occur. All four recombinases in the formed tetrameric complex are pairwise active and cleave one strand each. The crystal structure of the integron integrase from *Vibrio cholerae* shows that two folded *attC* substrates (VCR_{bs}) are bound by four VchIntIA protomers. The two protomers bound to the same substrate do not share equivalent protein-DNA contacts resulting in an overall synapsis with only two-fold symmetry. The secondary structures α-helix I_2 and β-4,5 hairpin encoded by the additional region plus surroundings in the integrase, and the bulges formed in the folded bottom strands are shown to play key roles for this synapse formation by determining which protein pair that is active and discrimination of the two strands for cleavage. Only one pair of VchIntIA molecules is thought to be active. The recombination reaction is assumed to stop at the Holliday junction stage due to an energetically unfavoured isomerization step of the two-fold symmetrical Holliday junction intermediate (MacDonald *et al*., 2006).

In integron integrase-mediated site-specific recombination, depending on the architectures of the recombining sites, there may be more than the usual of four recombinase molecules involved in the reaction. Four protein-DNA complexes have been observed in gelshift assays involving ds *attI1* and IntI1 (Collis *et al*., 1998; Gravel *et al.,* 1998) indicating that *attI* itself can recruit four integron integrase molecules.

Examles of other hairpin binding proteins

Hairpins are considered to be transiently present in DNAs, and are important in cellular processes such as gene regulation, recombination and mutagenesis. DNA domains that contain such structures generally consist of inverted repeats that are preferred target sites for many proteins. **Topoisomerase II**, for example, has been reported to recognize an inverted repeat sequence as part of a strong cleavage site in phage pBR322 DNA (Mauffret *et al*., 1998). The site-specific **XerC/D** recombination system is conserved in essentially all bacteria with circular chromosomes, and ensures proper chromosome segregation at cell division by acting at a site called *dif* near the replication

terminus (Barre and Sherratt, 2002). Curiously, this host-encoded system is utilized by the filamentous phage **CTXφ** (Waldor and Mekalanos, 1996) to integrate its (+) single-stranded genome (encoding cholerae toxin) into the *dif* site of the large chromosome of *Vibrio cholerae*. A region of the single-stranded phage genome folds into a pseudo double-stranded recombination site by internal base pairing. The generated site contains a short hairpin on one strand of the overlap and several short bulges in the XerC/D binding sites. Formation of the short hairpin in the overlap region results in a complementary overlap region with sufficient similarity with the overlap region of *dif* to allow recombination. Only one strand exchange occurs (mediated by XerC) so the formed HJ must be resolved by replication passing through the structure (Val *et al*., 2005).

Relaxases as mentioned earlier, bind to *oriT* and perform a site-specific nick in a conserved core region. The nick provides the priming hydroxyl for a rolling-circle replication process that generates single-stranded DNA. To the left of the core sequence lies an inverted repeat region that varies in sequence and size among several compared *oriT* sequences. This region is recognized by the relaxases. When DNA is single-stranded this region is able to form a hairpin loop in the 3' end of transferred DNA. This is thought to enable recruitment of relaxase and recircularization of the transferred DNA (Parker *et al*., 2005; Williams and Schildbach, 2006).

The prokaryotic transposases of Tn*5* and Tn*10* contain a hairpin binding module consisting of a hydrophobic binding pocket and a conserved YREK motif. At transposition both Tn*5* and Tn*10* are excised from their donor sites through concomitant formation of a hairpin intermediate with covalently closed hairpin ends (Kennedy *et al.,* 1998; Davies *et al*., 2000; Allingham *et al*., 2001; Naumann and Reznicoff, 2002; Rice and Baker, 2001). Another enzyme containing a hairpin-binding module with a known YKEK motif similar to the transposase of Tn*5* and Tn*10* is **ResT** (Bankhead and Chaconas, 2004). ResT is the telomere resolvase of *Borrelia burgdorferi,* the causative agent of Lyme disease. As mentioned previously, ResT has similar active site residues as the tyrosine recombinases and type 1b topoisomerases (Deneke *et al*., 2004). In the telomere resolution reaction, the hairpin-binding module promotes pre-hairpinning of the inverted repeats found in the replicated telomere junctions. This step activates the replication intermediate for cleavage through a two-step transesterification reaction mediated by ResT. ResT has also been found to perform the reverse reaction that fuses telomeres on unrelated DNA molecules (Kobryn and Chaconas, 2005). This discovery is suggested to have played an important role in genome plasticity and evolution of the *Borrelia* genomes (Chaconas, 2005).

Rag1/2 proteins mediate V(D)J recombination in human and other vertebrates (a process for developing B and T lymphocytes) through a process that in part is mechanistically similar to cut-and-paste transposition (e.g. of Tn*5* and Tn*10*). During gene rearrangements of different gene segments,

hairpins are formed in the ends of the coding regions. Rag1 uses a catalytic DDE motif to form DNA hairpins through direct transesterification. Rag1 is proposed to contain a sequence motif that may share similarities with the YREK motif (Lu *et al.,* 2006; Jones and Gellert, 2004).

Relations to other systems

One important characteristic of the integron recombination system is the recognition of hairpins. The need for a strict regulation of integron activity may be one reason for the unusual shape of the recombination sites and perhaps this single-stranded recombination system is restricted to be activated only at occations when DNA is single-stranded.

The importance of single-stranded DNA

DNA is thought to be able to exist in many more different conformations than RNA. However, it is very rarely single-stranded. Single-stranded DNA naturally has a tendency to coil back up. Potential occasions when a major portion of DNA is single-stranded are:

Homologous recombination
Replication
Transcription
Topoisomerization
Conjugative DNA transfer
Natural transformation
RC replication and transposition
Growth of bacteriphage containing single-stranded DNA

Many of these situations have been previously mentioned in this thesis, such as conjugation and the integration of the single-stranded genome of the temperate and filamentous bacteriophage CTXφ phage encoding cholera toxin (Moyer *et al*., 2001).

Replication via a rolling-circle (RC) mechanism occurs in ss DNA viruses such as CTXφ and the icosahedral bacteriophage φX174. Certain transposons like IS*91*, IS*1294* and IS*801* use the (RC) mechanism for transposition, in which the DNA intermediate exists in a single-stranded form (del Pilar Garcillán-Barcia *et al*., 2001; Tavakoli *et al*., 2000; Richter *et al*., 1998). IS common region (ISCR) elements are widely distributed IS*91*-like elements found to be closely linked with many different antibiotic resistance genes. They have been found in Salmonella genomic islands (SGI) and in SXT elements from *Vibrio cholerae*. The ISCR1 element contains *orf513* encoding an IS*91*-like transposase and has been observed closely associated with class 1 integrons (forming the so-called complex integrons) (Toleman

et al., 2006; Partridge and Hall, 2003; Rodriguez-Martinez *et al*., 2006). The ISCR1-linked antibiotic resistance genes are not gene cassettes and must therefore been taken up by another mechanism. The appearance of another system for genetic assembly close to the integron is very interesting. One could perhaps refer to the two linked gene capture mechanisms as primary and secondary integrons. Several clinically highlighted resistance genes have been observed in the secondary integron. Some of them have spread in a short time to world-wide extent (Rodriguez-Martinez *et al.,* 2006; Toleman *et al.,* 2006; Parsons *et al.,* 1991). The proposed RC transposition mechanism of ISCRs is mediated by the encoded putative transposase that shares similarities with, for example, the replication protein (protein A) of φX174 and TnpA of IS*91*. It has been shown for IS*91* and IS*1294* that the RC mechanisms sometimes misidentifies the termination end and replicates more sequence adjacent to the IS element. This could mean that ISCRs are likely to mobilize any gene from any location the element has transposed into. The presence of ISCR1 elements near class 1 integrons, suggests a more efficient system for spread of antibiotic resistance genes has been created, partly by co-mobilizing integron sequence at the RC transposition event, and partly by generating single-stranded DNA that has been shown to be important for integron activity.

Present investigation

Aims of this thesis

Several families of antibiotic resistance genes are part of gene cassettes that can be integrated at or excised from expression sites in integrons. Either of these two reactions is performed by an integron-encoded integrase that belongs to the tyrosine family of site-specific recombinases.

In the papers presented, the general aims were I and II) to find evidence for the involvement of single-stranded DNA in site-specific recombination and I) to characterize strand-and subsite specificity of IntI1 to the *attC* site of the *aadA1-qacE* cassette junction in Tn*21*, II) to probe for internal cooperation effects in *attC* using long-range mutations and a mating-out assay, III) to find and characterize the hairpin binding region in integron integrases and to perform homology modelling of IntI1 with the crystal structure of VchIntIA-VCR_{bs} as template, and IV) to increase our knowledge about the integron integrase IntI1 and optimize conditions for production of large amounts of the protein with maintained stability and high purity.

Results

Paper I:

In this paper we have investigated the binding of the integron integrase, IntI1, to the *attC* site of the *aadA1-qacE* cassette junction in Tn*21*. One aim was to uncover what properties of *attC* direct recombination to only one end of the site that is highly symmetrical and palindromic from end to end. Interruptions in the palindromic symmetry are condensed in a central trinucleotide (called the loop), two different spacers between the two subsites (R"pentamerL" and L'hexamerR') and two single-nucleotide insertions. After preliminary recombination and PCR experiments (not shown) we decided to test the binding of the purified integrase to oligonucleotides with different sequence modifications in either strand. We believed that the asymmetries could include the determinant in the site that mediate recogni-

tion and also ensure that the gene cassettes are inserted in the proper orientation.

It was earlier demonstrated by Francia *et al.* (1999) that IntI1 binds to the bottom strand of *attC* and not to the top strand. This observation was supported in paper I where our constructed oligonucleotides were incubated in the presence of IntI1 and the strand specificity was investigated using an electrophoretic mobility shift assay (EMSA), a well-known method for studying DNA-protein interactions *in vitro*. Our data show that the recognition is dependent on the presence of the few asymmetrical nucleotides. The interpretation of our data suggests that the integron integrases bind to a hairpin stucture with bulges. The asymmetrical nucleotides may play a role in forming specific bulges when the DNA strand folds into a hairpin. Both the nature and positioning of the hairpin distortions would be strand-specific and could hence explain the remarkably strong bias for the bottom strand. Consequently, this bias for one DNA strand directs recombination to the right-hand subsite. The recognition differs from that of other tyrosine recombinases in being specific for a structure rather than for a sequence. This could explain why two *attC* sites that vary in size and sequence still can be recognized by the same integron integrase.

We also observed the initial biochemical reaction of the recombination cycle when we incubated integron integrase with the bottom strand *attC* oligos. We found mobility-shifted complexes in denaturing gels as well as in non-denaturing gels. Due to the proteic nature of the nucleophile in the reaction, the shifted complexes contained a covalent linkage indicating that the cleavage step had taken place. We produced a non-nucleophilic control protein, IntI1Y312F, by site-directed mutagenesis. This protein maintains the ability of the wild type for binding DNA but cannot generate DNA cleavage. When using the non-nucleophilic protein, shifts persisted only under non-denaturing conditions. We concluded from this data that we had observed binding and cleavage activity of IntI1 to the bottom strand of *attC in vitro* and that tyrosine Y312 is the residue responsible for DNA-attack and formation of a covalent phosphotyrosine linkage. This was the first biochemical step of the recombination demonstrated to occur with high efficiency in a cell-free system.

Paper II:

The paper studies recombinational effects of mutations in *attC* using conjugative cotransfer indicating replicon fusion due to *attI/attC* crossover. The conjugative plasmid R388Δ1 containing an *attI* site, the non-mobilizable pSU18/19 derived plasmid containing variants of the *attC* site and the integron integrase from Tn*21* supplied in *trans* by the plasmid 2101 were used for the co-mating assay (Hansson *et al*., 1997).

A set of progressive deletions in the *aadA1-qacE* junction-area in Tn*21* defined a fully recombination-proficient *attC* of 65 bases. This is similar to the deletion subcloning analysis made previously by Martinez and de la Cruz (1988 and 1990). Our data in addition, identified cross-over points in *attC*. The positions of these points support a previously suggested pattern of L and R repeats in *attC* by Stokes *et al*. (1997) and Francia *et al*., (1997). Thus, the *attC* site contains two closely related subsites each built-up of two inverted repeats R"L" and L'R', respectively. The preferred recombination cross-over point is found in R' and this work showed that recombination events also occur (at a lower frequency) in two of the remaining repeats, after a deletion of R' had been made in the *attC* site. Furthermore, analysis of a functional left half-*attC* site from a cassette in Tn*7* showed cross-over points in both repeats (resembling L" and R") but the half-site in Tn*7* and similar constructed derivatives of *attC* revealed imprecise strand transfer suggesting that a full *attC* site is needed for efficient and accurate recombination.

Mutations in the R' repeat had earlier been found to be very responsive. Here we evaluate the current stem-loop model (presented in paper I and updated more recently by Bouvier *et al*., (2005) and MacDonald *et al*., (2006) for the recognition of the *attC* site in Tn*21* by an extensive array of directed long-range mutations in the recombination silent end in and near R". The effect of the mutations on the rate of integrative recombination was assayed in the *in vivo* conjugation-cointegration assay. Mutations in R" also had a strong negative influence on the recombination at R', resulting in an approximate 1000-fold reduction of recombination. The strongest effect was seen in point mutations in the four most conserved bases of R". These observations support our data in paper I and new publications within the field showing that the integron integrase binds to a hairpin structure with bulges generated by folding of the bottom strand of the *attC* site. In such a structure, complementarity is important between R"-R' and L"-L'.

In an integrated cassette the *attC* site is a composite site, meaning that the upstream region (including R", L" and L') and the R' repeat belong to different cassettes. Here we analyzed the differences in recombination efficiencies depending on the origin of the two parts in the composite *attC* site. Our data indicate that mismatches between R" and R' may reduce site activation.

Paper III

In this paper we aimed to perform homology modelling of IntI1 with the crystal structure of VchIntIA-VCR_{bs} as template and to find and characterize a hairpin binding region in integron integrases. When comparing integron integrases with other tyrosine recombinases, they all possess typical conserved motifs and are therefore also thought to share the same catalytic mechanism as that mediated by Cre and the other well-characterized members of the family. However, integron integrases contain a unique extra

stretch of amino acids necessary for activity, and the organization of their recombination sites differs from those recognized by other proteins within the same family. The integron integrases recognize and bind to double-stranded *attI* (Gravel *et al*., 1998) but only to the bottom strand of *attC* (Francia *et al*., 1999; paper I). The crystal structure of the related integron integrase, VchIntIA, from the superintegron in *Vibrio cholerae* bound to a bulged duplex reconstruction of the folded bottom strand of its *attC* substrate, VCR_{bs}, (*V. cholerae* repeat, bottom strand) was reported earlier this year (MacDonald *et al*, 2006). Four VchIntIA monomers are bound to two VCR_{bs} molecules forming a two-fold symmetric complex in which two subunits are in active and attacking conformation. The solved structure clearly highlights the importance of two flipped-out bases, G20" and T12", in a folded hairpin structure of the bottom strand of the *attC* substrate, for correct assembly of the synaptic complex. Two secondary structures in VchIntIA, αI_2 and β-4,5 hairpin found in and near to the unique additional region, have been suggested to play key roles for the synapse formation. In the crystal structure, α-helix I_2 in the attacking subunit is involved in several important DNA contacts in *trans* including contacts with G20" while β-4,5 hairpin in the non-attacking subunit mediates contacts with T12" in *cis*.

We have shown in this work using homology modelling of IntI1 from Tn*21* against the present synaptic complex of VchIntIA that the overall protein folds of the two proteins (sharing 45% identity) are identical. We looked specifically at the area around the six catalytic amino acids (RKHRHY) (involved in *cis* contacts), and the amino acids shown to either constitute αI_2 or β-4,5 hairpin or to be involved in protein-DNA contacts involving the bulges and/or other nucleotides in the VchIntIA-VCR_{bs} structure. The corresponding amino acids in IntI1 fit equally well in the homology model and those that differ from VchIntIA in the structure were found to not disturb the overall fold. When searching for differences in our homology model versus the available structure, we observed more base-specific contacts with VCR_{bs} for IntI1 than for VchIntIA. This was interesting since IntI1 has previously been shown to have a broader recognition profile compared to VchIntIA.

Furthermore, we have made alignments of integron integrases with the hairpin binding motifs found in the transposases of Tn*5* and Tn*10* and the telomere processing protein ResT of *Borrelia burgdorferi* and found similarities with the unique insertion of integron integrases. Based on these alignments, twenty IntI1 mutants were constructed and tested for binding to an $attC_{bs}$ substrate by using EMSA aiming to find and characterize a similar motif in integron integrases. Four amino acids were highly responsive and by using our homology model of IntI1 we could explain the EMSA results for the different IntI1 mutants. Interestingly, one of the mutants K219W that was previously shown to bind *attI* did not bind *attC*. This suggested that K219, situated in α-helix I_2, is involved in recognition and hairpin binding since only *attC* is proposed to form hairpins. Although we did not find IntI1

to utilize the same set of amino acids for hairpin binding as the documented motifs, we concluded that there is a strong connection between the unique additional region among integron integrases and hairpin binding properties.

Paper IV:

The work completed for this paper aimed to increase our knowledge about the protein characteristics of the integron integrase IntI1. It was necessary for our work to produce large amounts of the enzyme, purify and store it with maintained activity. The gene encoding the integron integrase from Tn*21*, *intI1*, is 5'-fused to a sequence coding for a His-tag decamer. The tagged protein conjugate was formed by cloning the IntI1 gene into a pET vector of which expression is based on transcription from a strong promoter of phage T*7*. When gene expression is induced, large quantities of the integrase is produced. The main fraction of the expressed protein is entrapped as inclusion bodies, however expression at a lower temperature gives a substantial fraction of the native and soluble form. Here we used nano-LC/MS-MS techniques and new software, DeCyder™ MS for relative quantitation of LC-MS data, for optimizing growth conditions to obtain maximal yields of pure soluble integrase, followed by chromatographic purification, for binding and mechanistic studies. Crude extracts from cultures of *E.coli* harbouring the expression construct induced at 37ºC for 0, 15, 30, 60 or 240 minutes were digested with trypsin and analyzed. Data from different LC-MS analysis were compared using the software resulting in label-free relative quantitation of several identified peptides. The results indicate that the optimal induction time at 37ºC for obtaining soluble integrase is 60 minutes. This method for evaluation reveals differences sometimes hard to detect using SDS-PAGE. To demonstrate the ability of DeCyder™ to accurately quantify proteins in complex mixtures, three different strains of *E.coli* expressing known and varying levels of dihydrofolate reductase (DHFR) were also used. The results obtained using the presented techniques in this work were consistent with data from traditional spectrophotometric and gel based assays of DHFR content in the different strains.

Discussion

This work presents new data that belong to a recent breakthrough in the study of integrons and metagenomic traffic in general culminating with the determination of the three-dimensional structure of the *Vibrio cholerae* IntI-VCR_{bs} complex by MacDonald *et al.,* (2006). Integron integrases belong to the tyrosine family of recombinases but act on sites that vary both in length and sequence. To these two aspects should be added now if not already earlier, DNA conformation. The *attC* sites are generally organized into two

subsites intervened by a palindromic sequence. Why the terminal repeat pairs influenze each other has for us remained enigmatic until a new model appeared. It has been wondered how the site organization can be so different from the usual recombination site recognized by other tyrosine recombinases. Some have suggested that the sequence between the two R repeats recruits accessory proteins like HU, FIS or IHF that could bend the site, resulting in a closer distance between the two R repeats (the bending model; 1). Unpublished data from our lab have shown that supplements of purified HU and IHF stimulated the excision frequency in an *in vitro* recombination assay (see below) (Gullberg *et al.,* unpublished). However, no such recognition site has yet been identified (Biskri *et al*., 2005 and references therein). It is also possible that the bent appearance might be promoted by the binding of the integron integrase itself.

A second suggestion (the hairpin model; 2) is presented and investigated in paper I. With its palindromic appearance, the *attC* site might extrude hairpin arms and therefore adopt a cruciform shape (2a). In such structure the outer repeats (R" and R') are brought close enough together to be grasped by a postulated dimer of integrase. A special point with the cruciform is that it so obviously resembles a Holliday junction intermediate formed after the first cleavage and rejoining step in the recombination reaction of known tyrosine recombinases. In paper I we showed using EMSA that mutations in L repeats were more responsive than mutations in R repeats and concluded that at least two (L", L') of the four repeats are involved in intramolecular base pairing. Results from the *in vivo* assay in paper II showed that both ends of the more than 60 base pair long site are essential for recombination. This strongly suggests that the ends must come together.

The second version of the hairpin model, 2b, is folding of a single strand of the *attC* site with all four repeats, resulting in a pseudo double-stranded recombination site with normal appearance, consisting of two inverted repeats. A problem with this option is that DNA is seldom single-stranded for more than very short periods during processes like replication and transcription. However, in lateral gene transfer processes the DNA sometimes exists in single-stranded form. Conjugation is one such process and the fact that integrons often are found on conjugative plasmids signifies that this process might be utilized for integron activity.

Bouvier *et al,*. (2005) elaborated the hairpin model (2) further in the now favoured direction (2b) by using a suicide conjugation assay. Cointegrate formation occurred in the recipient due to recombination between the bottom strand of the *attC* site and a double-stranded *attI* site. The recombination rate was similar to the classical assay, where double-stranded *attC* is present on a plasmid that was able to replicate. However, *attC* recombination could not be observed when cells were transformed with double-stranded *attC*-containing plasmids not able to replicate. The top strand did not appear to be involved in recombination, which is consistent with results in paper I where

cleavage occurred without the necessity of top strand presence. The recombination reaction is assumed to stop at the Holliday junction intermediate which necessitates a resolution step by replication or other cellular factors. In the case of replication, both substrate and recombination products would be formed. Bouvier *et al.,* (2005) also suggest that a similar model could be applied for *attC* x *attC* recombination.

In paper II, we have used a method that utilizes the fact that single-stranded DNA is produced and exposed without the normally annealing counter strand at conjugation. Interestingly, *attI* is situated on the conjugative plasmid R388Δ1 whereas the tested *attC* sites are positioned on replicable but non-mobilizable pSU18/19 derivatives. It is crucial for the assay that the p15a-derived replicon (pSU18/19) unlike a wide range of other plasmids, is not mobilized by R388 but leaves the co-transfer of the two plasmids entirely dependent on plasmid fusion (Chang and Cohen, 1978; Martinez and de la Cruz, 1988). According to the current integration model presented by Bouvier *et al*., (2005) cointegration of pSU18/19 derivatives into R388Δ1 should occur when the former plasmid contains single-stranded *attC* and the latter is still double-stranded. We aim to perform a reverse comparable experiment where we put the *attI* site on the smaller plasmid and *attC*-variants on the larger conjugative plasmid R388.

We also reflected on the fact that matters are not always black or white but sometimes gray. Unpublished data from our lab shows recombination using an *in vitro* excision assay based on recombination between double-stranded DNA substrates (Gullberg *et al.,* unpublished). Some of these results are discussed here, as the data suggests that a parallel Campbell-like pathway may exist as well as the proposed single-stranded mechanism. This was somewhat surprising since we at present favour a recombination model which first requires the presence of single-stranded *attC* and secondly, requires replication. The *in vitro* assay does not involve conjugation nor replication. However, DNA may appear in different conformations due to its own flexibility in solution and thereby make at least one step in the reaction possible according to the current model. Reaction mixtures containing purified IntI1 or IntI3 and either plasmid DNA or linear DNA containing an *aadAI* cassette flanked by *attI* and *attC* respectively, were incubated and examined by inverse or conventional PCR for recombination products. The analyzed PCR products had sizes consistent with site-specific recombination between the sites at the *aadA1* cassette ends. The results were identical for IntI1 and IntI3 and thus the restricted specificity of IntI3 for its own class of *attI* site *in vivo* was not observed *in vitro*. It is possible that the specificity is more relaxed in cell-free reactions. Furthermore, the linear DNA recombination substrates gave similar results as was obtained with plasmid substrates, indicating that negative supercoiling is not an absolute requirement for DNA strand-transfer *in vitro*. In addition, mutations in the recombination sites reduced the formation of PCR products according to previous quantitative

measurements *in vivo* (Hansson *et al*., 1997). We believe that we have observed site-specific recombination *in vitro* that can be interpreted as IntI1 and IntI3 cleave both top and bottom strands in the recombination reaction. There is however one drawback with this assay. In our PCR approach we were not able to distinguish whether the generated amplification products show the result of partial (first strand) DNA transfer or complete transfer of both strands. The template for the PCR reaction could in fact either be a stable Holliday junction intermediate or a recombinant product caused by deletion of the *aadA1* cassette. Both templates would give the same PCR product and that was inevitable in the design of the assay. Consequently, the results from our *in vitro* assay could not tell us whether resolution of Holliday structures formed by integron integrases requires accessory replication or repair. Although supplements of both purified HU and IHF were observed to stimulate the reactions, our *in vitro* results support the idea that no other cellular proteins except from the integron integrase itself are absolutely essential. Thus, the question remains to be answered how the first step or second step recombination products are generated in the first place. One could postulate that apart from the single-stranded mechanism, there may exist a parallel Campbell-like pathway as well.

Integron integrases seem to prefer double-stranded *attI* and single-stranded *attC*. It is needed to conduct futher studies in order to understand how both principles act together in the same synaptic complex. Prior to solving the IntI1 crystal structure, homology modelling in paper III has proven how important such a model is when lacking a crystal structure to better understand the reaction mechanism. Although unlikely, one has to consider that reactions catalyzed by VchIntIA may differ in substrate recognition compared to IntI1. In being an integrase that is commonly borne on mobile vehicles, IntI1 may have adapted to function in different host types and may also have obtained extra features valuable for different mechanisms.

The role of IS*91*-like elements in the vicinity of integrons is very exciting because they have a potential for supporting the integron with conversion of its sites into the needed singlex confomation. This could mean that integrons are adapted to the two niches of conjugation events and the rolling-circle replication caused by IS*91*-like elements (del Pilar Garcillan-Barcia *et al.,* 2001, Toleman *et al.,* 2006).

With the sensitive techniques used and published in paper IV very small amounts of IntI1 can be detected and quantitated. This leaves us an expectation that the method could facilitate future studies on integron regulation. This is entirely unknown but the system appears to be strictly controlled. The availability of ss DNA could perhaps be a second principle of recombinational control.

Concluding remarks

Important for the progress in understanding the integron system has been the upgrading of the previously underestimated impact of single-stranded DNA in molecular genetics. A changed focus from duplex to singlex DNA now promises to clarify important functions of integrons and several related processes with impact on the spread of a plague in our time, antibiotic resistance. The diversity of plasmid-borne antibiotic resistance was early connected with modular genetic variation in many plasmids and transposons (Wiedemann *et al*., 1985; Tanaka *et al*., 1985). Scrutinized comparisons among sequences for instance among Tn*21*-like elements revealed single or multiple insertions of genes in a very narrow sector of a highly conserved sequence. Insertions were flanked by the sulfonamide resistance gene, *sul1* on one side and a 337 amino acid *orf* on the opposite side (Swedberg and Sköld, 1983; Cameron *et al*, 1986; Quellette *et al*, 1987; Sundström *et al.,* 1987 and 1988). The sharp distinction and high integrity of the inserted modules promised prompt explanations especially when it was soon realized that the 337 amino acid *orf* was related to the lambda type integrases described as a protein family just earlier by Argos *et al.,* (1986); (Ouellette and Roy, 1987; Sundström *et al.,* 1988). It was early realized that the integrons come in multiple types because two of the inserted cassettes, *aadA1* and *dfr1* were found in a second homing site on Tn*7* encoding a closely related recombinase IntI2 (Sundström *et al.,* 1988; Sundström and Sköld, 1990, Hansson *et al*., 2002). The *dfr1* cassette has later been found to be distributed in all five hitherto decribed classes of resistance integrons (Arakawa *et al*., 1995; Hochhut *et al.,* 2001; Sørum *et al*., GenBank: AJ277063). A rather large number of tyrosine recombinases have been extensively studied and in a few cases in very high detail such as P1 Cre, λ Int, 2-micron yeast plasmid FLP and *E.coli* XerCD. Molecular structure data are available for a few recombinases, with or without their DNA substrates, and to the former group now also for one of the integron integrases, the protein VchIntIA from the small chromosome of *V. cholerae*. These recent data made possible the modelling approach presented in paper III.

Data from 1998 and later have widened our views concerning both the origin and function of integrons. In 1998 the first large chromosomal integron, called superintegron, was described. The superintegrons seem to be distributed to most *Vibrio* genera and less thoroughly to other aquatic or environmental species. Already a few years ago sequence alignments and mutagenesis data uncovered a unique amino acid sequence among integron integrases (Messier and Roy, 2001; Nunes-Duby *et al*., 1998; Esposito and Scocca, 1997). The reported data since 1999 (Francia *et al.,* 1999) have now matured into the insight that singlex DNA has a role in integron recombination. We have also noticed that single-stranded DNA is reported to play the role as a molecular recognition signal in other processes. For instance the

CTXφ phage encoding the cholera toxin in *V. cholerae* integrates in the form of ssDNA (Val *et al.,* 2005). The presence of both a superintegron and the CTXφ prophage, two singlex-dependent mechanisms, in the same *Vibrio* species could tell us something about the genetic transfer across aquatic populations of pathogenic and environmental bacteria.

Finally, new questions are now raised: First, to what degree is the Campbell model as proposed by Collis and Hall (1992b) still relevant for integron cassette recombination? Are parallel mechanisms running side by side or is the hairpin model describing the unique and also essential integron function? Do we have to create new models combining elements from both models? Other questions include what type of biological or functional motivation there was for elaborating a hairpin recognition mechanism? Is it used solely to widen the register of sites that the integrases have to operate on in order to recruit new genes? Is it instead more relevant to regard it as a regulatory function that boosts cassette migration in connection with rare horizontal gene transfer events that are known for exposing single-stranded DNA in the cell? After those questions it could perhaps even be reasonable to challenge the assumption that the integron integrases indeed belong to the same class as other tyrosine family of recombinases? In total, though, we approach the moment when we can provide answers to the long unanswered question why the site organization is so different from the usual recombination site recognized by other tyrosine recombinases.

Acknowledgements

After all these years of enjoyful work as a PhD student I have many people to thank:

Docent **Lars Sundström** my excellent supervisor for giving me the opportunity to learn more about the fascinating world of science by accepting me as a PhD student. Thank you for generously have guided me through this period with your great knowledge and expertise and for your constant belief in me. I have enjoyed our yearly trips to conferences and meetings where you have introduced me to many interesting and talented scientists.

Professor **Dan Andersson**, my co-supervisor for all support.

My opponent, Dr. **Didier Mazel** and members of the committee for taking time going through this thesis and coming to Sweden.

Professor **Ulf Lindahl**, my examiner for all support.

The members of the Microbiology group, both past and present: **Ola Sköld, Göte Swedberg, Kristina Lundberg, Ewa Boström, Claës Linder**, **Agneta Hortlund and Hans Darberg** for all encouragement, interesting discussions and always being very friendly. **Maria Jönsson** and **Yvonne Qvarnström** for being the best roomies one could have! Maria, also for your invaluable computer expertise. **Masood Kamali-Moghaddam** for taking care of me and teaching me all the basics during my first months as a PhD student and even after you had left the lab. Thank you for all support! **Karin Hansson** for your fantastic and interesting work on integrons! **Mats Gullberg** for great teamwork and teaching me real-time PCR. **John Flensburg** presently at GE Heathcare Uppsala -Thank you so much for our fruitful collaboration. **Jens Eriksson** for your excellent work when staying in our lab. **Cecilia Norman-Setterblad** for teaching me how to do alignments and for letting me join you in your car from Stockholm every Monday morning. **Peter Rådström**, thank you for letting me come to Lund and learn more about PCR.

Thanks also to:

Jenny Samskog, **Henrik Wadensten** and **Lennart Björkesten** at GE Healthcare, Uppsala for collaboration with Paper IV.
Lars Boukharta and **Johan Åqvist** at ICM for a fun modelling work and fruitful collaboration.
Robert Robinsson formly at IMBIM for modelling our PCR mutant and **Stefan Knight** (SLU) for fruitful IntI1 discussions.

Malin Grape and **Göran Kronvall** from Karolinska institut, Stockholm.

Professor **David Sherratt** for valuable advice, visits in Oxford and for organizing fantastic conferences. Also thanks to his wonderful wife Dr **Lidia Arciszewska.**

Professor **Peter Geiduschek** for planning a collaboration**,** Dr. **Garry Blakely** for footprinting advice and Professor **Anca Segall** for sending IHF.

All students: **Alexander Jernberg, Björn Eklund, Emma af Bjerkén, Henrik Boije, Jenny Karlsson, Jenny Sandberg, Liina Annuk, Linn Skärberg, Linus Nystrand and Oskar Ström**. THANK YOU!

The 337 group: **Elisabeth**, **Clara, Lina** and **Carlos** for interesting and stimulating meetings in a very friendly atmosphere.

IMBIOL group for interesting Tuesday seminars.

Allison Jones for kindly taking your time and go through this thesis.

Present and former colleagues in the B9:3 corridor **Ali, Aline, Amal, Aneta, Anh-Tri, Camilla, Chi, Dorothe, Elina, Elisabeth, Emanuel, Erik, Eva G, Eva H, Helena, Jin-Ping, Juan, Krysztof, Lars, Lena, Maarten, Maria W, Marion, Marta B, Martha E, Nadja, Per, Rebecka, Sabrina, Sindhu, Sofia, Ulf, and Wei** for delicious Friday cakes, friendly smiles, laughs, pep-talks, advice and interesting discussions.

Anh-Tri and **Maarten** for being such good friends!

Jonny Wernersson for being so friendly and for all help and support and for understanding my experimental problems.

Maria, Yvonne, Anne, Ulrika, Mia, Josef, **Kristina** and **Ted** for fantastic teamwork when teaching.

Åke Engström for protein analysis. **Inger Eriksson** for help with sequencing, **Santanu Dasgupta** for strains, **Kerstin Lidholt** and **Olav Nordli** for all

help with computer related problems, questions and purchase and other technical matters, **Barbro Lowisin** and **Erika Enström** for PhD and thesis related issues and **Rehné Eklund, Marianne Wigenius, Maud Pettersson and Lisa Holmsäter** and of course **Tony Grundin** and **Piotr Wlad.**

Other IMBIM colleagues for making IMBIM such a great place to be!

EUDRESP2 for funding and **Marco Oggioni** for being the coordinator, **Anna Maria Lundins Stipendiefond** at Smålands nation and **Apotekarsocieteten** enabling attendance at interesting conferences and courses.

Friends outside BMC-you know who you are! Thank you so much for your friendship, belief in me, pep-talks and chocolate. Also, a big thank you for convincing me that there is a world outside BMC.

Friskis&Svettis

Anders' family: **Inger** and **John**, **Lisa** and **Jan-Erik**, and sister and brothers **Christian**, **Marina**, **Cecilia**, **Amanda** and their families for all encouragement.

Finally, **My loving family**. Without your support, phone calls, visits, home made cakes and raggsockor, I would never have made it. **Mamma Anita**, **Pappa Stig**, **Sister Hanna** and **Kenneth**, and my dear **Mormor Maita**. LOVE YOU!

Last but most important My Fiancé **Anders** for your endless belief in me, taking care of me and our home. You mean everything! LOVE YOU, PUSS!

References

Abremski K, Hoess R. 1984. Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J Biol Chem* **259:** 1509-1514.

Abremski KE, Hoess RH. 1992. Evidence for a second conserved arginine residue in the integrase family of recombination proteins. *Protein Eng* **5:** 87-91.

Allingham J S, Wardle S J, Haniford D B. 2001. Determinants for hairpin formation in Tn10 transposition. 2001. *The EMBO Journal* **20:** 2931-2942.

Arakawa Y, Murakami M, Suzuki K, Ito H, Wacharotayankun R, Ohsuka S, Kato N, Ohta M. 1995. A novel integron-like element carrying the metallo-beta-lactamase gene blaIMP. *Antimicrob Agents Chemother* **39:** 1612-5.

Arciszewska LK, Baker RA, Hallet B, Sherratt DJ. 2000. Coordinated control of XerC and XerD catalytic activities during Holliday junction resolution. *J Mol Biol* **299:** 391-403.

Argos P, Landy A, Abremski K, Egan JB, Haggard-Ljungquist E, Hoess RH, Kahn ML, Kalionis B, Narayana SV, Pierson LS, 3rd, *et al.*, 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. *Embo J* **5:** 433-440.

Averhoff B, Friedrich A. 2003. Type IV pili-related natural transformation systems: DNA transport in mesophilic and thermophilic bacteria. *Arch Microbiol* **180:** 385-93. Review.

Bankhead T, Chaconas G. 2004. Mixing active-site components: a recipe for the unique enzymatic activity of a telomere resolvase. *Proc Natl Acad Sci U S A* **101:** 13768-73.

Barker A, Clark CA, Manning PA. 1994. Identification of VCR, a repeated sequence associated with a locus encoding a hemagglutinin in Vibrio cholerae O1. *J Bacteriol* **176:** 5450-5458.

Barre FX, Sherratt DJ. 2002. Xer Site-Specific Recombination: Promoting Chromosome Segregation. *Mobile DNA II.* Washington DC: ASM Press.

Beiko RG, Harlow TJ, Ragan MA. 2005. Highways of gene sharing in prokaryotes. *Proc Natl Acad Sci U S A* **102:** 14332-7.

Bi X, Liu LF.1994. recA-independent and recA-dependent intramolecular plasmid recombination. Differential homology requirement and distance effect. *J Mol Biol* **235:** 414-23.

Biskri L, Bouvier M, Guerout AM, Boisnard S, Mazel D. 2005. Comparative study of class 1 integron and Vibrio cholerae superintegron integrase activities. *J Bacteriol*. **187:** 1740-50.

Bissonnette L, Roy PH. 1992. Characterization of In0 of Pseudomonas aeruginosa plasmid pVS1, an ancestor of integrons of multiresistance

plasmids and transposons of gram-negative bacteria. *J Bacteriol* **174:** 1248-1257.

Biswas T, Aihara H, Radman-Livaja M, Filman D, Landy A, Ellenberger T. 2005. A structural basis for allosteric control of DNA recombination by lambda integrase. *Nature* **435**: 1059-66.

Blake DG, Boocock MR, Sherratt DJ, Stark WM. 1995. Cooperative binding of Tn3 resolvase monomers to a functionally asymmetric binding site. *Curr Biol* **5:** 1036-46.

Blakely G, May G, McCulloch R, Arciszewska LK, Burke M, Lovett ST, Sherratt DJ. 1993. Two related recombinases are required for site-specific recombination at dif and cer in E. coli K12. *Cell* **75:** 351-361.

Bouvier M, Demarre G, Mazel D. 2005. Integron cassette insertion: a recombination process involving a folded single strand substrate. *EMBO J* **24:** 4356-67.

Britten RJ. 2005. The majority of human genes have regions repeated in other human genes. *Prot Natl Acad Sci U S A* **102:** 5466-70.

Brown HJ, Stokes HW, Hall RM. 1996. The integrons In0, In2, and In5 are defective transposon derivatives. *J Bacteriol* **178:** 4429-4437.

Bunny KL, Hall RM, Stokes HW. 1995. New mobile gene cassettes containing an aminoglycoside resistance gene, aacA7, and a chloramphenicol resistance gene, catB3, in an integron in pBWH301. *Antimicrob Agents Chemother* **39:** 686-693.

Burrus V, Waldor MK. 2004. Shaping bacterial genomes with integrative and conjugative elements. *Res Microbiol* **155:** 376-386.

Campbell AM, 1962. Episomes. *Adv. Genet* **11:** 101-45.

Cameron FH, Groot Obbink DJ, Ackerman VP, Hall RM. 1986. Nucleotide sequence of the AAD(2") aminoglycoside adenylyltransferase determinant aadB. Evolutionary relationship of this region with those surrounding aadA in R538-1 and dhfrII in R388. *Nucleic Acids Res* **14:** 8625-8635.

Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brussow H. 2003. Phage as agents of lateral gene transfer. *Curr Opin Microbiol* **6:** 417-24. Review.

Cassano AG, Anderson VE, Harris ME. 2004. Understanding the transition states of phosphodiester bond cleavage: insights from heavy atom isotope effects. *Biopolymers* **73:** 110-29. Review.

Centron D, Roy PH. 2002. Presence of a group II intron in a multiresistant Serratia marcescens strain that harbors three integrons and a novel gene fusion. *Antimicrob Agents Chemother* **46:** 1402-9.

Cesar CE, Machon C, de la Cruz F, Llosa M. 2006. A new domain of conjugative relaxase TrwC responsible for efficient oriT-specific recombination on minimal target sequences. *Mol Microbiol* **62:** 984-96.

Chaconas G. 2005. Hairpin telomeres and genome plasticity in Borrelia: all mixed up in the end. *Mol Microbiol* **58:** 625-35. Review.

Chaconas G, Stewart PE, Tilly K, Bono JL, Rosa P. 2001. Telomere resolution in the Lyme disease spirochete. *EMBO J* **20:** 3229-37.

Champoux JJ. 2001. DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem* **70:** 369-413. Review.

Chandler M. 2006. Molecular biology: Singled out for integration. *Nature* **440:** 1121-1122. News and Views.

Chang AC, Cohen SN. 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol* **134:** 1141-56.

Chen Y, Narendra U, Iype LE, Cox MM, Rice PA. 2000. Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol Cell* **6:** 885-897.

Cheng, C, Kussie P, Pavletich N, Shuman S. 1998. Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases. *Cell* **92:** 841-50.

Coleman N, Tetu S, Wilson N, Holmes A. 2004. An unusual integron in Treponema denticola. *Microbiology* **150** (Pt 11): 3524-6.

Collis CM, Grammaticopoulos G, Briton J, Stokes HW, Hall RM. 1993. Site-specific insertion of gene cassettes into integrons. *Mol Microbiol* **9:** 41-52.

Collis CM, Hall RM. 1992a. Site-specific deletion and rearrangement of integron insert genes catalyzed by the integron DNA integrase. *J Bacteriol* **174:** 1574-1585.

Collis CM, Hall RM. 1992b. Gene cassettes from the insert region of integrons are excised as covalently closed circles. *Mol Microbiol* **6:** 2875-2885.

Collis CM, Hall RM. 1995. Expression of antibiotic resistance genes in the integrated cassettes of integrons. *Antimicrob Agents Chemother* **39:** 155-162.

Collis CM, Hall RM. 2004. Comparison of the structure-activity relationships of the integron-associated recombination sites attI3 and attI1 reveals common features. *Microbiology* **150:** 1591-1601.

Collis CM, Kim MJ, Partridge SR, Stokes HW, Hall RM. 2002a. Characterization of the class 3 integron and the site-specific recombination system it determines. *J Bacteriol* **184:** 3017-26.

Collis CM, Kim MJ, Stokes HW, Hall RM. 1998. Binding of the purified integron DNA integrase Intl1 to integron- and cassette-associated recombination sites. *Mol Microbiol* **29:** 477-490.

Collis CM, Kim MJ, Stokes HW, Hall RM. 2002b. Integron-encoded IntI integrases preferentially recognize the adjacent cognate attI site in recombination with a 59-be site. *Mol Microbiol* **46:** 1415-1427.

Collis CM, Recchia GD, Kim MJ, Stokes HW, Hall RM. 2001. Efficiency of recombination reactions catalyzed by class 1 integron integrase IntI1. *J Bacteriol* **183:** 2535-2542.

Colloms SD, Alen C, Sherratt DJ. 1998. The ArcA/ArcB two-component regulatory system of Escherichia coli is essential for Xer site-specific recombination at psi. *Mol Microbiol* **28:** 521-530.

da Silva AC, Ferro JA, Reinach FC, Farah CS, Furlan LR, Quaggio RB, Monteiro-Vitorello CB, Van Sluys MA, Almeida NF, Alves LM, *et al.,* 2002. Comparison of the genomes of two Xanthomonas pathogens with differing host specificities. *Nature* **417:** 459-463.

Davies J. 1994. Inactivation of antibiotics and the dissemination of resistance genes. *Science **264:*** 375-382.

Davies D R, Goryshin I Y, Reznikoff W S, Rayment I. 2000. Three-Dimensional Structure of the Tn5 Synaptic Complex Transposition Intermediate. *Science* **289:** 77-85

del Pilar Garcillan-Barcia M, Bernales I, Mendiola MV, de la Cruz F. 2001. Single-stranded DNA intermediates in IS91 rolling-circle transposition. *Mol Microbiol* **39:** 494-501.

Deneke J, Burgin AB, Wilson SL, Chaconas G. 2004. Catalytic residues of the telomere resolvase ResT: a pattern similar to, but distinct from, tyrosine recombinases and type IB topoisomerases. *J Biol Chem* **279:** 53699-706.

Deneke J, Ziegelin G, Lurz R, Lanka E. 2000. The protelomerase of temperate Escherichia coli phage N15 has cleaving-joining activity. *Proc Natl Acad Sci U S A* **97:** 7721-6.

Dorgai L, Sloan S, Weisberg RA. 1998. Recognition of core binding sites by bacteriophage integrases. *J Mol Biol* **277:** 1059-1070.

Dorgai L, Yagil E, Weisberg RA. 1995. Identifying determinants of recombination specificity: construction and characterization of mutant bacteriophage integrases. *J Mol Biol* **252:** 178-188.

Draper O, Cesar CE, Machon C, de la Cruz F, Llosa M. 2005. Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells. *Proc Natl Acad Sci U S A* **102:** 16385-90.

Drouin F, Melancon J, Roy PH. 2002. The IntI-like tyrosine recombinase of Shewanella oneidensis is active as an integron integrase. *J Bacteriol* **184:** 1811-1815.

Echols H. 1990. Nucleoprotein structures initiating DNA replication, transcription, and site-specific recombination. *J Biol Chem* **265:** 14697-14700.

Esposito D, Scocca JJ. 1997. The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res* **25:** 3605-3614.

Fling ME, Kopf J, Richards C. 1985. Nucleotide sequence of the transposon Tn7 gene encoding an aminoglycoside-modifying enzyme, 3"(9)-O-nucleotidyltransferase. *Nucleic Acids Res* **13:** 7095-106.

Fluit AC, Schmitz FJ. 2004. Resistance integrons and super-integrons. *Clin Microbiol Infect* **10:** 272-288.

Francia MV, Avila P, de la Cruz F, Garcia Lobo JM. 1997. A hot spot in plasmid F for site-specific recombination mediated by Tn21 integron integrase. *J Bacteriol* **179:** 4419-4425.

Francia MV, de la Cruz F, Garcia Lobo JM. 1993. Secondary-sites for integration mediated by the Tn21 integrase. *Mol Microbiol* **10:** 823-828.

Francia MV, Garcia Lobo JM. 1996. Gene integration in the Escherichia coli chromosome mediated by Tn21 integrase (Int21). *J Bacteriol* **178:** 894-898.

Francia MV, Zabala JC, de la Cruz F, Garcia Lobo JM. 1999. The IntI1 integron integrase preferentially binds single-stranded DNA of the attC site. *J Bacteriol* **181:** 6844-6849.

Frost LS, Leplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol.* **3:** 722-32. Review.

Frumerie C, Sylwan L, Ahlgren-Berg A, Haggard-Ljungquist E. 2005. Cooperative interactions between bacteriophage P2 integrase and its accessory factors IHF and Cox. *Virology* **332:** 284-94.

Garcillan-Bracia MP, Bernales I, Mendiola MV et al. 2002. IS91 rolling-circle transposition. In: *Mobile DNA II*. Washington: ASM Press.

Gogarten JP, Townsend JP. 2005.Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol.* **3:** 679-87. Review.
Gopaul DN, Duyne GD. 1999. Structure and mechanism in site-specific recombination. *Curr Opin Struct Biol* **9:**14-20.
Gottfried P, Yagil E, Kolot M. 2000. Core-binding specificity of bacteriophage integrases. *Mol Gen Genet* **263:** 619-624.
Grainge I, Jayaram M. 1999. The integrase family of recombinase: organization and function of the active site. *Mol Microbiol* **33:** 449-456.
Gravel A, Fournier B, Roy PH. 1998. DNA complexes obtained with the integron integrase IntI1 at the attI1 site. *Nucleic Acids Res* **26:** 4347-4355.
Grindley ND. 1993. Analysis of a nucleoprotein complex: the synaptosome of gamma delta resolvase. *Science,* **262:** 738-40.
Grindley ND, Whiteson KL, Rice PA. 2006. Mechanisms of site-specific recombination. *Annu Rev Biochem* **75:** 567-605.
Grishin NV. 2000. Two tricks in one bundle: helix-turn-helix gains enzymatic activity. *Nucleic Acids Res* **28:** 2229-33.
Guo F, Gopaul DN, van Duyne GD. 1997. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389:** 40-46.
Hakimi JM, Scocca JJ. 1994. Binding sites for bacteriophage HP1 integrase on its DNA substrates. *J Biol Chem* **269:** 21340-21345.
Hall RM, Brookes DE, Stokes HW. 1991. Site-specific insertion of genes into integrons: role of the 59-base element and determination of the recombination cross-over point. *Mol Microbiol* **5:** 1941-1959.
Hall RM, Collis CM. 1995. Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Mol Microbiol* **15:** 593-600.
Hall RM, Collis CM. 1998. Antibiotic resistance in gram-negative bacteria: the role of gene cassettes and integrons. *Drug Resist Updates* **I:** 109-119.
Hall RM, Collis CM, Kim MJ, Partridge SR, Recchia GD, Stokes HW. 1999. Mobile gene cassettes and integrons in evolution. *Ann N Y Acad Sci* **870:** 68-80.
Hall RM, Vockler C. 1987. The region of the IncN plasmid R46 coding for resistance to beta-lactam antibiotics, streptomycin/spectinomycin and sulphonamides is closely related to antibiotic resistance segments found in IncW plasmids and in Tn21-like transposons. *Nucleic Acids Res* **15:** 7491-7501.
Hallet B, Arciszewska LK, Sherratt DJ. 1999. Reciprocal control of catalysis by the tyrosine recombinases XerC and XerD: an enzymatic switch in site-specific recombination. *Mol Cell* **4:** 949-959.
Hallet B, Sherratt DJ. 1997. Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol Rev* **21:** 157-178.
Hansson K, Skold O, Sundstrom L. 1997. Non-palindromic attl sites of integrons are capable of site-specific recombination with one another and with secondary targets. *Mol Microbiol* **26:** 441-453.
Hansson K, Sundstrom L, Pelletier A, Roy PH. 2002. IntI2 integron integrase in Tn7. *J Bacteriol* **184:** 1712-21.
Heinemann JA, Ankenbauer RG. 1993. Retrotransfer of IncP plasmid R751 from Escherichia coli maxicells: evidence for the genetic suffi-

ciency of self-transferable plasmids for bacterial conjugation. *Mol Microbiol* **10**: 57-62.
Hickman AB, Waninger S, Scocca JJ, Dyda F. 1997. Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 A resolution. *Cell* **89:** 227-237.
Holmes AJ, Holley MP, Mahon A, Nield B, Gillings M, Stokes HW. 2003. Recombination activity of a distinctive integron-gene cassette system associated with Pseudomonas stutzeri populations in soil. *J Bacteriol* **185:** 918-28.
Hochhut B, Lotfi Y, Mazel D, Faruque SM, Woodgate R, Waldor MK. 2001. Molecular analysis of antibiotic resistance gene clusters in vibrio cholerae O139 and O1 SXT constins. *Antimicrob Agents Chemother*. **45:** 2991-3000.
Ikeda H, Shiraishi K, Ogata Y. 2004. Illegitimate recombination mediated by double-strand break and end-joining in Escherichia coli.*Adv Biophys* **38:** 3-20. Review.
Johansson C, Kamali-Moghaddam M, Sundstrom L. 2004. Integron integrase binds to bulged hairpin DNA. *Nucleic Acids Res* **32:** 4033-4043.
Johnsson RC. 1995. Site-specific recombinases and their interactions with DNA. In: DNA-Protein: Structural Interactions: IRL Press.
Jones JM, Gellert M. 2004. The taming of a transposon: V(D)J recombination and the immune system. *Immunol Rev* **200:** 233-48. Review.
Kennedy AK, Guhathakurta A, Kleckner N, Haniford DB. 1998; Tn10 transposition via a DNA hairpin intermediate. *Cell* **95:**125-134.
Kobryn K, Chaconas G. 2005. Fusion of hairpin telomeres by the B. burgdorferi telomere resolvase ResT implications for shaping a genome in flux. *Mol Cell* **17:** 783-91
Kowalczykowski SC. 2000. Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem Sci* **25:** 156-65. Review.
Krogh BO, Shuman S. 2002. A poxvirus-like type IB topoisomerase family in bacteria. *Proc Natl Acad Sci U S A* **99:** 1853-8.
Kurland CG. 2005. What tangled web: barriers to rampant horizontal gene transfer. *Bioessays.* **27:** 741-7.
Kwon HJ, Tirumalai R, Landy A, Ellenberger T. 1997. Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science* **276:**126-131.
Landy A. 1989. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu Rev Biochem* **58:** 913-949.
Leon G, Roy PH. 2003. Excision and integration of cassettes by an integron integrase of Nitrosomonas europaea. *J Bacteriol* **185:**2036-2041.
Leverstein-van Hall MA, HE MB, AR TD, Paauw A, Fluit AC, Verhoef J. 2003. Multidrug resistance among Enterobacteriaceae is strongly associated with the presence of integrons and is independent of species or isolate origin. *J Infect Dis* **187:** 251-259.
Levesque C, Brassard S, Lapointe J, Roy PH. 1994. Diversity and relative strength of tandem promoters for the antibiotic-resistance genes of several integrons. *Gene* **142:** 49-54.
Liebert CA, Hall RM, Summers AO. 1999. Transposon Tn21, flagship of the floating genome. *Microbiol Mol Biol Rev* **63:** 507-522.
Ligon BL. 2004. Penicillin: its discovery and early development. *Semin Pediatr Infect Dis* **15:** 52-7.

Lovett ST, Hurley RL, Sutera VA Jr, Aubuchon RH, Lebedeva MA. 2002. Crossing over between regions of limited homology in Escherichia coli. RecA-dependent and RecA-independent pathways. *Genetics* **160:** 851-9.

Lu CP, Sandoval H, Brandt VL, Rice PA, Roth DB. 2006. Amino acid residues in Rag1 crucial for DNA hairpin formation. *Nat Struct Mol Biol* **13:** 1010-5.

MacDonald D, Demarre G, Bouvier M, Mazel D, Gopaul DN. 2006. Structural basis for broad DNA-specificity in integron recombination. *Nature* **440:** 1157-62.

Mahillon J, Lereclus D. 1988. Structural and functional analysis of Tn4430: identification of an integrase-like protein involved in the co-integrate-resolution process. *EMBO J* **7:** 1515-26.

Martinez E, de la Cruz F. 1988. Transposon Tn21 encodes a RecA-independent site-specific integration system. *Mol Gen Genet* **211:** 320-325.

Martinez E, de la Cruz F. 1990. Genetic elements involved in Tn21 site-specific integration, a novel mechanism for the dissemination of antibiotic resistance genes. *Embo J* **9:** 1275-1281.

Mauffret O, Amir-Aslani A, Maroun RG, Monnot M, Lescot E, Fermandjian S. 1998. Comparative structural analysis by [1H,31P]-NMR and restrained molecular dynamics of two DNA hairpins from a strong DNA topoisomerase II cleavage site. *J Mol Biol* **283:** 643-55.

Mazel D. 2006. Integrons: agents of bacterial evolution. *Nat Rev Microbiol* **4:** 608-20. Review.

Mazel D, Dychinco B, Webb VA, Davies J. 1998. A distinctive class of integron in the Vibrio cholerae genome. *Science* **280:** 605-608.

Mazel D, Davies J. 1999. Antibiotic resistance in microbes. *Cell Mol Life Sci*. **56 :** 742-54. Review.

Messier N, Roy PH. 2001. Integron integrases possess a unique additional domain necessary for activity. *J Bacteriol* **183:** 6699-6706.

Mizuuchi K. 1997. Polynucleotidyl transfer reactions in site-specific DNA recombination. *Genes Cells* **2:** 1-12. Review.

Moyer KE, Kimsey HH, Waldor MK. 2001. Evidence for a rolling circle mechanism of phage DNA synthesis from both replicative and integrated forms of CTXphi. *Mol Microbiol* **41:** 311-23.

Nash HA. 1996. *Site-specific recombination: integration, excision, resolution, and inversion of defined DNA segments.* Washington, DC: ASM press.

Naumann T A, Reznikoff W S. 2002. Tn5 Transposase Active Site Mutants. *the Journal of Biological Chemistry* **277:** 17623-17629.

Nield BS, Holmes AJ, Gillings MR, Recchia GD, Mabbutt BC, Nevalainen KM, Stokes HW. 2001. Recovery of new integron classes from environmental DNA. *FEMS Microbiol Lett* **195:** 59-65.

Nunes-Düby SE, Azaro MA, Landy A. 1995. Swapping DNA strands and sensing homology without branch migration in lambda site-specific recombination. *Curr Biol* **5:** 139-48.

Nunes-Duby SE, Kwon HJ, Tirumalai RS, Ellenberger T, Landy A. 1998. Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res* **26:** 391-406.

Nunes-Duby SE, Yu D, Landy A. 1997. Sensing homology at the strand-swapping step in lambda excisive recombination. *J Mol Biol* **272:** 493-508.

Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405:** 299-304.

Otten H.1986. Domagk and the development of the sulphonamides. *J Antimicrob Chemother* **17:** 689-96.

Ouellette M, Bissonnette L, Roy PH. 1987. Precise insertion of antibiotic resistance determinants into Tn21-like transposons: nucleotide sequence of the OXA-1 beta-lactamase gene. *Proc Natl Acad Sci U S A* **84:** 7378-7382.

Ouellette M, Roy PH. 1987. Homology of ORFs from Tn2603 and from R46 to site-specific recombinases. *Nucleic Acids Res* **15:** 10055.

Pabo CO, Sauer RT. 1992.Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem* **61:** 1053-95. Review.

Parker C, Becker E, Zhang X, Jandle S, Meyer R. 2005. Elements in the co-evolution of relaxases and their origins of transfer. *Plasmid* **53:** 113-8. Review.

Parsons Y, Hall RM, Stokes HW. 1991. A new trimethoprim resistance gene, dhfrX, in the In7 integron of plasmid pDGO100. *Antimicrob Agents Chemother* **35:** 2436-9.

Partridge SR, Hall RM. 2003. In34, a complex In5 family class 1 integron containing orf513 and dfrA10.*Antimicrob Agents Chemother* **47:** 342-9.

Partridge SR, Recchia GD, Scaramuzzi C, Collis CM, Stokes HW, Hall RM. 2000. Definition of the attI1 site of class 1 integrons. *Microbiology* **146** *(Pt 11)*:2855-2864.

Paulsen IT, Littlejohn TG, Radstrom P, Sundstrom L, Skold O, Swedberg G, Skurray RA. 1993. The 3' conserved segment of integrons contains a gene associated with multidrug resistance to antiseptics and disinfectants. *Antimicrob Agents Chemother* **37:** 761-768.

Radstrom P, Skold O, Swedberg G, Flensburg J, Roy PH, Sundstrom L. 1994. Transposon Tn5090 of plasmid R751, which carries an integron, is related to Tn7, Mu, and the retroelements. *J Bacteriol* **176:** 3257-68.

Recchia GD, Hall RM. 1995a. Plasmid evolution by acquisition of mobile gene cassettes: plasmid pIE723 contains the aadB gene cassette precisely inserted at a secondary site in the incQ plasmid RSF1010. *Mol Microbiol* **15:** 179-187.

Recchia GD, Hall RM. 1995b. Gene cassettes: a new class of mobile element. *Microbiology* **141** *(Pt 12)*: 3015-3027.

Recchia GD, Hall RM. 1997. Origins of the mobile gene cassettes found in integrons. *Trends Microbiol* **5:** 389-394.

Recchia GD, Sherratt DJ. 2002. Gene acquisition in bacteria by integron-mediated site-specific integration. *Mobile DNA II.* Washington DC: ASM Press.

Recchia GD, Stokes HW, Hall RM. 1994. Characterisation of specific and secondary recombination sites recognised by the integron DNA integrase. *Nucleic Acids Res* **22:** 2071-2078.

Redinbo MR, Stewart L, Kuhn P, Champoux JJ, Hol WG. 1998. Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science* **279:**1504-13.

Rice PA, Baker TA. 2001. Comparative architecture of transposase and integrase complexes. *Nat Struct Biol* **8:** 302-7. Review.

Richter GY, Bjorklof K, Romantschuk M, Mills D.1998. Insertion specificity and trans-activation of IS801. *Mol Gen Genet* **260:** 381-7.

Rodriguez-Martinez JM, Poirel L, Canton R, Nordmann P. 2006. Common region CR1 for expression of antibiotic resistance genes. *Antimicrob Agents Chemother* **50:** 2544-6.

Rowe-Magnus DA, Guerout AM, Mazel D. 1999. Super-integrons. *Res Microbiol* **150:** 641-651.

Rowe-Magnus DA, Guerout AM, Mazel D. 2002. Bacterial resistance evolution by recruitment of super-integron gene cassettes. *Mol Microbiol* **43:** 1657-69.

Rowe-Magnus DA, Guerout AM, Ploncard P, Dychinco B, Davies J, Mazel D. 2001. The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. *Proc Natl Acad Sci U S A* **98:** 652-7.

Rowe-Magnus DA, Mazel D. 1999. Resistance gene capture. *Curr Opin Microbiol* **2:** 483-488.

Rowe-Magnus DA, Mazel D. 2001. Integrons: natural tools for bacterial genome evolution. *Curr Opin Microbiol* **4:** 565-569.

Sadowski P. 1986. Site-specific recombinases: changing partners and doing the twist. *J Bacteriol* **165:** 341-347.

Sadowski PD. 1993. Site-specific genetic recombination: hops, flips, and flops. *Faseb J* **7:** 760-767.

Segal H, Elisha BG. 1997. Identification and characterization of an aadB gene cassette at a secondary site in a plasmid from Acinetobacter. *FEMS Microbiol Lett* **153:** 321-326.

Sharma A, Hanai R, Mondragon A. 1994. Crystal structure of the amino-terminal fragment of vaccinia virus DNA topoisomerase I at 1.6 A resolution.*Structure **2:*** 767-77.

Sherratt DJ. 2001. *Site-specific Recombination in Chromosome Function*: Macmillan Publishers Ltd, Nature publishing group.

Skold O. 2000. Sulfonamide resistance: mechanisms and trends. *Drug Resist Updat* **3:** 155-160.

Smith MC, Thorpe HM, 2002. Diversity in the serine recombinases. *Mol. Microbiol* **44:** 299-307. Review.

Stark WM, Boocock MR, Sherratt DJ. 1989b. Site-specific recombination by Tn3 resolvase. *Trends Genet* **5:** 304-309.

Stark WM, Boocock MR, Sherratt DJ. 1992. Catalysis by site-specific recombinases. *Trends Genet 8*: 432-439.

Stark WM, Sherratt DJ, Boocock MR. 1989a. Site-specific recombination by Tn3 resolvase: topological changes in the forward and reverse reactions. *Cell* **58:** 779-90.

Stokes HW, Hall RM. 1989. A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol Microbiol* **3:** 1669-1683.

Stokes HW, O'Gorman DB, Recchia GD, Parsekhian M, Hall RM. 1997. Structure and function of 59-base element recombination sites associated with mobile gene cassettes. *Mol Microbiol* **26:** 731-745.

Subramanya HS, Arciszewska LK, Baker RA, Bird LE, Sherratt DJ, Wigley DB. 1997. Crystal structure of the site-specific recombinase, XerD. *Embo J* **16:** 5178-5187.

Sundstrom L. 1998. The potential of integrons and connected programmed rearrangements for mediating horizontal gene transfer. *APMIS Suppl* **84:** 37-42.

Sundstrom L, Skold O. 1990. The dhfrI trimethoprim resistance gene of Tn7 can be found at specific sites in other genetic surroundings. *Antimicrob Agents Chemother* **34:** 642-50.

Sundstrom L, Swedberg G, Skold O. 1993. Characterization of transposon Tn5086, carrying the site-specifically inserted gene dhfrVII mediating trimethoprim resistance. *J Bacteriol* **175:** 1796-805.

Sundstrom L, Radstrom P, Swedberg G, Skold O. 1988. Site-specific recombination promotes linkage between trimethoprim- and sulfonamide resistance genes. Sequence characterization of dhfrV and sulI and a recombination active locus of Tn21. *Mol Gen Genet* **213:** 191-201.

Sundstrom L, Roy PH, Skold O. 1991. Site-specific insertion of three structural gene cassettes in transposon Tn7. *J Bacteriol* **173:** 3025-3028.

Sundstrom L, Vinayagamoorthy T, Skold O. 1987. Novel type of plasmid-borne resistance to trimethoprim. *Antimicrob Agents Chemother* **31:** 60-6.

Swedberg G, Skold O. 1983. Plasmid-borne sulfonamide resistance determinants studied by restriction enzyme analysis. *J Bacteriol* **153:** 1228-37.

Tanaka M, Matsushita K, Yamamoto T. 1985. Genesis of a complex transposon encoding the OXA-1 (type II) beta-lactamase gene. *Antimicrob Agents Chemother* **28:** 227-34.

Tavakoli N, Comanducci A, Dodd HM, Lett MC, Albiger B, Bennett P. 2000. IS 1294, a DNA element that transposes by RC transposition. *Plasmid* **44:** 66-84.

Tenover FC. 2006. Mechanisms of antimicrobial resistance in bacteria. *Am J Infect Control* **34**(5 Suppl 1): S3-10; discussion S64-73. Review.

Toleman MA, Bennett PM, Walsh TR. 2006. Common regions e.g. orf513 and antibiotic resistance: IS91-like elements evolving complex class 1 integrons. *J Antimicrob Chemother* **58:** 1-6.

Vaisvila R, Morgan RD, Posfai J, Raleigh EA. 2001. Discovery and distribution of super-integrons among pseudomonads. *Mol Microbiol* **42:** 587-601.

Val ME, Bouvier M, Campos J, Sherratt D, Cornet F, Mazel D, Barre FX. 2005. The single-stranded genome of phage CTX is the form used for integration into the genome of Vibrio Cholerae. *Mol Cell* **19:** 559-66.

Vanhooff V, Galloy C, Agaisse H, Lereclus D, Revet B, Hallet B. 2006. Self-control in DNA site-specific recombination mediated by the tyrosine recombinase TnpI.*Mol Microbiol* **60:** 617-29.

Van Duyne GD. 2001. A structural view of cre-loxp site-specific recombination. *Annu Rev Biophys Biomol Struct* **30:** 87-104.

Vetter D, Andrews BJ, Roberts-Beatty L, Sadowski PD. 1983. Site-specific recombination of yeast 2-micron DNA in vitro. *Proc Natl Acad Sci U S A* **80:** 7284-7288.

Yagil E, Dorgai L, Weisberg RA. 1995. Identifying determinants of recombination specificity: construction and characterization of chimeric bacteriophage integrases. *J Mol Biol* **252:** 163-177

Yang W, Mizuuchi K. 1997. Site-specific recombination in plane view. *Structure.* **5:** 1401-6. Review.
Waldor MK, Mekalanos JJ. 1996. Lysogenic conversion by filamentous phage encoding cholera toxin. *Science* **272:** 1910-4.
Wang JC. 1996. DNA topoisomerases. *Annu Rev Biochem.* **65:** 635-92. Review.
Weisberg RA, Landy A. 1983. Site-specific recombination in phage lambda, p. 211-250. In R.W. Hendrix, J.W. Roberts, F.W.Stahl, and R.A. Weisberg, ed. Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
West SC. 1994. The processing of recombination intermediates: mechanistic insights from studies of bacterial proteins. *Cell* **76:** 9-15.
White MF, Giraud-Panis MJ, Pohler JR, Lilley DM. 1997. Recognition and manipulation of branched DNA structure by junction-resolving enzymes. *J Mol Biol* **269:** 647-664.
Wiedemann B, Meyer JF, Nies BA, Kratz J.1985. Transposable multiresistance. *J Antimicrob Chemother* **16:** 416-7.
Wiedemann B, Meyer JF, Zuhlsdorf MT. 1986. Insertions of resistance genes into Tn21-like transposons. *J Antimicrob Chemother* **18** Suppl C: 85-92.
Williams SL. Schildbach JF. 2006. Examination of an inverted repeat within the F factor origin of transfer: context dependence of F TraI relaxase DNA specificity. *Nucleic Acids Res* **34:** 426-35.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 214

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-7429

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2007