RECOMBINATIONAL AND TRANSCRIPTIONAL REGULATION

WITHIN THE CLASS 1 INTEGRON

by

CARAN ASTOR CAGLE

(Under the Direction of Anne O. Summers)

ABSTRACT

The integron is a site-specific recombination system that uses IntI integrase to mediate the insertion and excision of gene cassettes among bacterial genomes, notably the dissemination of antibiotic resistance genes. Integrons carry three promoters, a well conserved integrase promoter (P_{int}), and seven variants of a pair of promoters (P_c and P_2) responsible for gene cassette transcription. I investigated integron regulation at the levels of recombination and transcription. Qualitative PCR and protein pull-down assays done here did not identify growth-phase- dependent, or accessory protein(s) involvement in recombination regulation. However, others work revealed growth phase involvement as recombination products increased in the transition from log to stationary phase, and crystallography revealed site-directed recognition through specific substrate conformations. Competitive electro-mobility binding assays done in this body of work suggested additional levels of recombination regulation through differing IntI1 folding conformations resulting in the preference of one recognition site over another. Bi-directional transcriptional fusion data of P_{int}, and P_c and P_2 established the weak expression of P_{int} and confirmed P_c and P_2 strength; in contrast to prior speculation, deletion of 3-bp within the P_2 spacer region as it exists in four other versions, resulted in dramatically reduced but not

completely inhibited expression of P_2. The transcription regulator predictor program PRODORIC 8.9 predicted FIS, LexA, and IHF sites. *In vitro* binding assays confirmed the direct interaction of all three proteins with the promoter region. Transcriptional fusion data showed FIS repressed P_{int} and the cassette promoters; however, LexA did not repress P_{int} or P_2 in the native IntI state (i.e. without the 3-bp deletion resulting in formation of a LexA site). IHF activated P_{int} and the cassette promoters, and did not affect P_2 with 3-bp deletion. Finally, H-NS, a transcription regulator with no sequence specific recognition sites, directly repressed P_{int} and P_2 promoters with 3-bp deletion, but activated the cassette promoters (P_c and P_2) *in vivo*. Integrase mediated recombination is an intricate multi-faceted network of regulatory components entailing growth phase and conformations of the substrate DNA and IntI. Furthermore, the integrase and cassettes are regulated transcriptionally through equally complicated nucleo-protein complexes involving FIS, LexA, IHF, and H-NS.

INDEX WORDS: integron, integrase, intI1, regulation, recombination, transcriptional regulators, convergent promoters, FIS, IHF, H-NS, antibiotic resistance

RECOMBINATIONAL AND TRANSCRIPTIONAL REGULATION WITHIN THE CLASS 1 INTEGRON

by

CARAN ASTOR CAGLE

B.S., Virginia Polytechnic Institute and State University, 2001

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2009

RECOMBINATIONAL AND TRANSCRIPTIONAL REGULATION

WITHIN THE CLASS 1 INTEGRON

by

CARAN ASTOR CAGLE

Major Professor: Anne O. Summers

Committee: Anna Karls
Sidney Kushner
John Maurer

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2009

DEDICATION

This work is dedicated to my friend and fellow lab member, Bijal Patel, who was always in the lab with me on weekends and helped keep me sane and laughing throughout all the frustration. I wish you were here.

ACKNOWLEDGEMENTS

I would like to thank my major professor, Anne O. Summers, for all of her guidance and support, in addition to all of my committee members for all of their helpful insight and suggestions. I would like to thank the other members of my lab, both past and present, for all their assistance and for creating such a wonderful work environment. I'd like to thank my friends and family for all their encouragement. Finally, I would like to thank my husband for his endless support and insightful perspective over the years that kept me going with this classic line, "It's supposed to be hard, if it were easy everybody would do it."

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

LITERATURE REVIEW

INTRODUCTION

Mobile genetic elements (MGE) are segments of DNA that can mediate their own transfer, spreading from cell to cell or from genome to genome, and consist of bacteriophages, transposons, insertion sequences, plasmids, and integrons (Shapiro, 1995). They are believed to play a large role in evolution of bacterial genomes generating genetic diversity through the acquisition of advantageous genes or operons. They have proven to be invaluable tools for genetic engineering in both prokaryotes and eukaryotes. However, in nature they can contribute to numerous human health related concerns by promoting immune system evasion through antigenic variation among pathogenic bacteria (Norris, 2006), facilitating the movement of pathogenicity islands (Hacker and Kaper, 2000), and spreading antibiotic resistance genes (Mazel, 2006). This review focuses on mechanisms and regulation of those elements that use conservative site-specific recombination.

Conservative site-specific recombination is a DNA exchange mechanism that begins with the alignment of two short (largely homologous) central regions, sometimes referred to as the crossover site. Recombination occurs via a two-step transesterfication mechanism that requires no ATP or DNA synthesis/repair for completion, resulting in a reciprocal exchange of DNA that contains no breaks. In nature, this process underlies many biological phenomena such as the

integration/excision of phage genomes, control of gene expression, and the resolution of chromosome and plasmid dimers to monomers thereby ensuring stable inheritance of both to daughter cells (Grindley *et al*., 1982; Smith and Thorpe, 2002; Grindley *et a*l., 2006).

Site-specific recombinases are typically divided into two groups referred to as either serine or tyrosine recombinases. Each group is named according to the conserved amino acid responsible for covalently binding DNA during the recombination reaction. Both groups are ancient and well represented throughout eubacteria and archea, in addition to more recent discoveries made within eukaryotes (Grindley *et al.,* 2006). Although, the two recombinase types differ in amino acid sequence and the mechanism by which the site-specific reaction occurs, they use a common core DNA recognition site structure. A core site consists of a short central crossover region whose sequence usually determines the directionality of each site. Flanking each core site is a pair of symmetrically positioned perfect or imperfect inverted repeats that typically bind one recombinase each (Fig. 1.1). Once bound, recombinases initiate distinct recombination reactions determined by the location and orientation of core sites leading to the inversion, excision, or integration of DNA adjacent to the sites (Grindley *et al*., 1982; Smith and Thorpe, 2002; Grindley *et a*l., 2006). For example, two core sites on the same DNA molecule arranged in inverse orientation will result in inversion of the DNA segment as is the case for the Hin invertase involved in switching flagella antigen in *Salmonella*. Alternatively, resolution, a form of excision, is the recombination of two sites on the same DNA molecule arranged in the same orientation as direct repeats, resulting in the separation of two DNA circles as in the XerC/D (Blakely *et al*., 1993) and Tn*3* resolvase (Boocock *et al*., 1986; Grindley *et al*., 2006) systems for chromosome/plasmid dimer resolution and plasmid co-integrate resolution, respectively. Lastly, integration is the reverse reaction of excision and also entails

recombination of two sites on physically separate DNA molecules as seen for the insertion of lambda phage genome into its *E. coli* host chromosome.

Not surprisingly, since recombinases can initiate multiple reactions, investigation of their regulatory mechanisms has uncovered strategies as diverse as the members of each group. Regulation can come in the form of recombinase gene expression, interaction with host encoded accessory protein(s), or at physical checkpoints within the recombination pathway itself. Regulation via gene expression often involves reorientation of a promoter(s) via recombination or transcriptional interference among multiple promoters arranged in tandem, convergent or divergent orientations. An example of regulation through the interaction of a mobile element and host encoded accessory protein(s) is seen in the lambda recombinase as it uses excisionase XIS, and the global regulator IHF, to control lambda integration/excision and dictate the direction and frequency of each reaction. Finally, regulation via checkpoints in the recombination pathway refers to the start order of cleavage and the steric constraints imposed by parallel versus anti-parallel core site alignments that prevent misaligned sites from successful completion of recombination.

Here I focus on what is known regarding the mechanism and regulation of the integron, a genetic element that uses a unique type of tyrosine site-specific recombinase. To fully portray the distinctiveness of its mechanism and speculate on its regulation, I first describe the forms of regulation employed by other recombinases, particularly of the three better defined structurally distinct serine site-specific recombinases, followed by similar characterization of three distinct tyrosine site-specific recombinases (Cre, Xer, λ), that represent the simplicity, variation and complexity, respectively. Finally, I consider what is known of the mechanism and potential

forms of regulation of the distinct integron integrase in light of its possible evolution and importance in dissemination of antibiotic resistant gene cassettes among bacteria.

FORMS OF REGULATION

Transcription

Transcription can be regulated during initiation, elongation, termination, or message turnover. Regulation during initiation and elongation involves the recruitment and allosteric change of RNA polymerase from closed to open complex, these effects are directed through specific transcription factors acting as either activators or repressors and will be discussed later in this review. Regulation during termination entails either “Rho-independent” or “Rho-dependent” termination. During “Rho-indepentent” termination, transcription stops after the newly synthesized RNA molecule forms a G-C hairpin, followed by a run of U's (Richardson and Greenblatt, 1996). "Rho-dependent" termination releases the newly synthesized mRNA from the elongation complex after protein factor "Rho" destabilizes the interaction between the template and the mRNA. Lastly, message turnover refers to the stability and half-life of mRNA transcripts (Richardson and Greenblatt, 1996).

Promoter Arrangement

In addition to these classical types of transcriptional regulation, mobile elements also control transcription in two other ways: (1) the initiation or cessation of transcription through recombination of a segment containing a promoter involved in expression of host or element encoded genes as seen in both the Hin and Fim invertase recombination systems (Zieg *et al*., 1977; Olsen and Klemm, 1994), (2) transcription interference (TI), through the direct negative impact of one transcriptional activity on another through the interactions of multiple promoters

in either tandem, convergent or divergent orientations. The former is not commonly done by site-specific recombinases, while the later is found in various MGEs and naturally within the bacterial chromosomes as shown in *E. coli* Table 1.1 (Shearwin *et al*., 2005).

There are five currently defined TI mechanisms; promoter competition, sitting duck, occlusion, collision, and roadblock (Shearwin *et al*., 2005; Fig.1.2). Not all mechanisms inhibit transcription at the same point. Promoter competition affects transcription initiation and can occur in promoters arranged tandemly, convergently, or in overlapping divergent pairs. Promoter competition occurrence is defined by RNA polymerase (RNAP) occupation of only one promoter, preventing RNAP occupation of the nearby second promoter leading to enhanced the activity of the first (Fig. 1.2a). Sitting duck interference affects transcription initiation and can be found in either tandem or convergent promoter pairs. It occurs when the RNAP complex bound to one promoter is slow to transition from open to closed elongation complex. This complex is dislodged by the elongating RNAP complex traveling from the neighboring more efficient "complex transitioning" promoter (Fig.1.2b) as seen in promoters P_R and P_L of bacteriophage 186 (Callen *et al*., 2004; Shearwin *et al*., 2005). However, mathematical modeling indicates this form of interference is highest when the ratio of open complex formation is equal to closed complex. If the ratio is less than one, the complex is not stationary for long and therefore less likely to be dislodged, and if it is greater than one, the complex is dislodged and rapidly loads another (Shearwin *et al*., 2005; Sneppen *et al*., 2005). Occlusion interference affects transcription initiation and can be found in tandem or convergent promoter pairs. It occurs when rapid firing from a stronger promoter leads to multiple elongating complexes that transcribe over a nearby promoter, preventing RNAP from binding (Fig.1.2c). However, since

the transit time of an RNAP across a standard promoter is only 2-3 seconds, this form of interference is considered slight and requires a very strong promoter to occur.

Unlike the first three interferences described, collision interference affects transcription elongation and can be found only in convergent promoter pairs. Natural examples are rare, but one speculated to occur with the *pR* and *pRE* promoters of bacteriophage lambda. Collision interference happens when two elongating complexes run into each other resulting in the displacement of one or both complexes and premature transcription termination (Fig. 1.2d). It is not yet understood whether both complexes "fall off" after collision, if one complex is rescued by a host factor while the other "falls off," or if a third complex stimulates one of the stalled complexes (Epshtein and Nudler, 2003; Roberts and Parks, 2004; Shearwin *et al*., 2005).

Finally, roadblock interference also affects transcription elongation but would be found in tandem or convergent promoter pairs. There are few known examples, but the possibility of its occurrence is based on the knowledge that DNA-bound Lac repressor can block the progress of RNAP complex. Thus theoretically an extremely tightly bound open RNAP complex if not dislodged could act as a "roadblock" versus the previously described "sitting duck." (Fig. 1.2e) (Shearwin *et al*., 2005)

Transcription Factors

Accessory proteins referred to as transcription factors (TFs) consist of a DNA binding domain and an "allosteric-metabolite interaction" which allows TFs to express or repress genes as circumstance dictates (Pollack and Lyer, 2002; Martinez-Antonio and Collado-Vides, 2003). They are typically classified as either local or global and are defined by the ability to regulate operons located within one metabolic pathway or within different metabolic pathways, respectively (Gottesman, 1984; Martinez-Antonio and Collado-Vides, 2003). For both local and

global TFs, the decision to activate or repress an operon within a metabolic pathway is based on environmental cues. Global TFs are capable of recognizing and responding to a large number of changes in environment.

TFs directly interact with target sites located upstream, downstream, or overlapping promoters, acting as activators that recruit RNA polymerase (RNAP) or repressors that interfere with RNAP binding (Barnard *et al*., 2004; Bintu *et al*., 2005). In addition to actively recruiting or blocking RNAP, bound accessory proteins can also alter promoter conformation leading to inhibited RNAP recognition (Barnard *et al*., 2004). Activation can occur through simple, cooperative, or synergistic means. Repression can occur through simple, cooperative, or "DNA looping" (Bintu *et al.,* 2005). Simple regulation involves one TF bound to a single operator site; cooperative regulation involves the sequential interaction of two or more TFs bound at adjacent operator sites. Synergistic activation also involves the interaction of two or more TFs, but their interaction simultaneously activates transcription. Lastly, repression via DNA looping refers to bound TF provoking conformational changes in the promoter region inhibiting efficient RNAP binding or successful transcription (Bintu *et al.,* 2005).

Work with systems having simple activation and/or repression has provided the basis of understanding for TF promoter regulation, but most naturally occurring promoters are modulated by cooperative regulation between repressors, activators or both (Barnard *et al.,* 2004). It is estimated that only 8% of the genes in *E. coli* are known or predicted TFs, therefore it is necessary for many of them to serve multiple functions and work in combination to achieve regulation (Blattner *et al*., 1997; Perez-Rueda and Collado-Vides, 2000; Martinez-Antonio and Collado-Vides, 2003). In fact, analysis of the *E. coli* genome done through the RegulonDB database indicates only 20% of the known TFs regulate just one or two genes, while 49% of

regulated genes are controlled by multiple TFs (Martinez-Antonio and Collado-Vides, 2003). Interestingly, in *E. coli* 51% of all genes, including other TFs, are directly modulated by one or a combination of seven global regulatory proteins: CRP (Cyclic AMP Receptor Protein), FNR (Fumarate and Nitrate Reduction), IHF (Integration Host Factor), FIS (Factor for Inversion Stimulation), ArcA (type of anaerobic response regulator), H-NS (Histone-like Nucleoid Structuring protein), and Lrp (Leucine-Responsive regulatory Protein) (Martinez-Antonio and Collado-Vides, 2003). The interaction of these seven global TFs on the transcriptional regulatory network is overlapping and complex. Each member can potentially play multiple regulatory roles dictated by environmental prompting of activator and/or repressor action. Regulatory actions can be applied to itself, on another global or local TF, and finally in collaboration with another global or a local TF as co-regulators of another gene (Fig. 1.3 and Table 1.2). This elaborate network of TF groupings and subsets acts almost like a combination lock, but instead of only one possible correct combination, multiple combinations are used to create a cooperative decision-making process that provides the cell the flexibility to change as its environment alters and subsequent cell needs change.

Accessory Proteins for Recombinases

In addition to their actions as TFs, many of the proteins discussed above act as recombination accessory proteins, the most common of which are nucleoid structuring DNA binding proteins IHF, H-NS, and FIS. The anaerobic response regulator, ArcA is not used as widely but is also an accessory protein involved in Xer resolvase recombination activity discussed in more detail later. However, while the action of ArcA is not known nor required for recombination, it is thought to act as an enhancer of recombination activity. IHF, H-NS, and FIS often play critical roles in a number of differing recombination systems. Each protein acts in a

similar fashion through the promotion of supercoiling. Once bound IHF bends DNA 150º (Rice *et al.,* 1996) while FIS bends DNA 45-95º (Funnel and Phillips, 2004).

The widespread presence of these proteins or homologues, the ambiguous nature of the sequence recognition sites for FIS and IHF, and the complete absence of a conserved recognition site in the case of H-NS, creates ideal accessory factors. These features create a set of proteins readily available in most organisms, and capable of binding at multiple locations (Table 1.3). All three are subject to growth phase regulation. FIS is abundant in exponential growth phase ranging from 30,000 dimers per cell then sharply decreasing to less than 1,000 dimers per cell in stationary phase. IHF does not fluctuate as dramatically, having approximately 6,000 dimers per cell in exponential growth and increasing to 27,500 dimers per cell in stationary phase while H-NS has 10,000 to 20,000 dimers per cell in exponential growth and drops to 7,500 during stationary phase (Azam *et al.,* 1999).

Regulation of Recombination via Checkpoints

Host and/or element encoded accessory proteins are not always needed for successful recombination. Often, simple systems with only one type of recombinase and two identical core sites use recombination regulation checkpoints. Checkpoints are physical limitations which impose steric constraints within the crossover region in order to ensure correct strand breaks, DNA exchange, and product stability. Checkpoints work on several levels, but in most cases the asymmetry of the central crossover region which provides directionality and determines polarity of the site drives checkpoint formation (Grindley *et al*., 2006). In general, the asymmetry of the crossover region permits it to assume two possible bends that accommodate the conformation of the DNA in the initial synaptic complex, determining the first checkpoint, location and strand cleavage. The second checkpoint occurs after cleavage and is defined by parallel versus anti-

parallel core site alignments where only central regions in anti-parallel orientation can productively recombine. Finally, the third checkpoint occurs after recombination whereby the central region association with the initial substrate is broken through adoption of the alternate bend, preventing recombination from reverting back to original components. Without these checkpoints, simple systems would not be able to distinguish between intermolecular and intramolecular reactions leading to inversion, excision, or insertion (Grindley *et al*., 2006). However, the use of checkpoints is not exclusive to simple systems. The complex integron integrase recombination system discussed below relies on a checkpoint.

The method and degree of the regulation of recombination is diverse. However, the physical mechanism of site-specific recombination has the same common steps carried out by one of two distinct enzyme families, the serine or tyrosine recombinases, named in reference to the lytic nucleophile used by the recombinase. Their mechanisms and specific examples are discussed below.

SERINE RECOMBINASES

Structure of Serine Recombinases

Serine recombinases are also sometimes referred to as the resolvase/invertase family. This family is the less understood of these two recombinase families even though much information has been gathered from the study of four systems in particular: γδ and Tn*3* resolvases, both involved in co-integrate resolution; Hin invertase, involved in switching the flagella antigen in *Salmonella;* and Gin invertase, involved in switching the tail fiber in bacteriophage Mu. Despite its name this family also has insertion/excision function and is composed of three different structural groups (Smith and Thorpe, 2002). The first structural

group is the best studied and it consists of small recombinases like those found in Tn*3* and Hin which are involved in resolution and inversion. Members of this group typically have less than 200 amino acids forming two distinct domains. The 140 amino acid N-terminal domain is involved in protein-protein interaction and catalysis. The remaining 60 amino acids in the C-terminus bind DNA with a conserved helix-turn-helix (HTH) motif (Smith and Thorpe, 2002). Within this structural group there is smaller subgroup that contains only resolvases that have the same domain and sizes as the parent group, but contain an additional 100 amino acids at the C-terminus with unknown function. Preliminary work with two members in this subgroup, IS*Xc5* and Tn*5044*, found the additional region is essential to Tn*5044* but not IS*Xc5* recombination activity (Smith and Thorpe, 2002).

The second structural group is not as well known, with only 30 identified members and it encompasses integrases like *Streptomyces* temperate phage ΦC31 (Rowley *et al*., 2008) and transposases like TnpX from *Clostridium perfringens* transposon Tn*4451* (Crellin and Rood 1997). Members in this group encode large recombinases that vary in size from 441-772 amino acids, and like the resolvase/invertase group, the catalytic domain lies in the N-terminal region. However, immediately following this catalytic domain is a 220 amino acid region including 10 highly conserved residues. The N-proximal portion of this long region, is believed to be the DNA binding domain despite the fact it shares little homology with the DNA binding domain of the rest of the resolvase/invertase group. The third structural group is the least understood. One example is the transposase from IS*607*, a two gene mobile element found in *Helicobacter pylori* (Kersulyte *et al*., 2000). Its small recombinase is only 200 residues, and is used exclusively for transposition. Also, unlike the other groups the DNA binding domain is located N-terminal to the catalytic domain.

Mechanism of Serine Recombinases

Despite differences in structure, all members of this family contain a catalytic domain with two clusters of conserved residues one of which includes the serine nucleophile for which the group was named (Smith and Thorpe, 2002; Grindley *et al*., 2006). All members of this family use the same basic mechanism for recombination. The two-step transesterfication reaction requires no ATP or DNA synthesis or repair for completion. It needs only the assembly of a complex consisting of two DNA core sites, each bound by one pair of recombinases at the core half sites. After synaptic complex formation, the recombinases are activated and the conserved serine nucleophiles directly attack the sugar phosphate backbone of the donor and recipient DNA molecules within the crossover region of each core site creating double-stranded breaks with 2-bp overhangs (Fig. 1.4). These breaks conserve the energy from the broken phosphodiester bonds via the direct transfer of phosphoryl groups to the serine side chains forming a covalent DNA-recombinase complex at the 5' ends, leaving a free hydroxyl at the 3' ends of the DNA in each core site. Once cleavage of both sites is complete, the complex undergoes a single right-handed 180º rotation of one half of the complex relative to the other in order to relax negative supercoiling of the DNA substrate. This rotation rearranges broken ends in the correct configuration in order to allow the free hydroxyl groups at the 3' ends of each core site to attack of the phosphoseryl bonds on the opposite core site followed by the ligation of the crossover regions and recombination (Smith and Thorpe, 2002; Grindley *et al*., 2006).

Group 1 Serine Resolvases/Invertases

This first group contains the well characterized Tn*3* resolvase and Hin invertase. Tn*3* resolvase recognizes a 120-bp core site referred to as the *res* site, which contains three distinct binding sites, each binding one resolvase dimer (Grindley *et al*., 1982; Grindley *et al*., 2006).

All three binding sites consist of a pair of 12-bp recognition sequences flanking a spacer region of varied length consisting of 4-bp, 10-bp, and 1-bp for sites I, II, and III, respectively. Occupancy of all three sites is necessary for recombination, but only site I contains the region of cleavage and crossover. Sites II and III serve a regulatory role, acting as accessory sites preventing inversion and integration reactions through usage of a "topological filter" (Boocock *et al.*,1986; Grindley *et al*., 2006). The filter is enabled when sites II and III of two separate *res* sites are occupied by resolvases resulting in trapped negative supercoils that can either promote recombination by resolvases bound at site I of *res* sites in the correct orientation, or inhibit synapsis between *res* site in an incorrect orientation or on separate DNA molecules.

Hin invertase is involved in switching flagella antigen in *Salmonella* and recognizes two 26-bp sites *hixL* and *hixR* flanking a 996-bp H DNA segment (Zieg and Simon, 1980). The segment contains the coding region for the invertase (*hin* gene) and a promoter (depending on fragment orientation) drives transcription of the H2 flagellin gene (*fliB*) located 5-bp downstream from the *hixR* site. Within the *hin* gene, there is an enhancer region containing two binding sites separated by 48-bp for FIS dimers (Huber *et al*., 1985; Merickel *et al*., 1998; Smith and Thorpe, 2002). Supercoiled DNA permits a direct interaction between bound FIS and recombinase, activating catalysis and increasing inversion rates 1000-fold (Johnson *et al*., 1986). FIS involvement is common to several invertases, and HU is another accessory protein unique to Hin that also increases inversion, whose role is not yet determined (Johnson *et al*., 1986). Surprisingly, unlike the directionality regulation seen in resolvases, invertases use regulatory elements to moderate the frequency of recombination. Correct orientation of sites is not determined until the 2-bp crossover, whereby if the crossover region cannot correctly base pair, multiple DNA exchanges are made to restore initial sequence.

In addition to the components that regulate recombination within these examples, transcriptional regulation is also observed. Tn*3* encodes two divergent promoters within the *res* site that transcribe *tnpA* (transposase) and *tnpR* (resolvase) genes. The *tnpR* promoter is autoregulated by binding the resolvase to the *res* site (Wisehart *et al*., 1983). It has been speculated that Hin invertase is also subject to transcriptional control, due to transcriptional fusion data indicating a *hin* promoter in the short DNA segment between *hix* site and start of *hin* gene. (Craig *et al*., 2002).

Group 2 Serine Integrases and Transposases

The second group is not as well understood, but contains integrases like that of ΦC31 and transposases like TnpX of Tn*4451*. The mechanism regulating ΦC31 recombination is unclear, but like other phages it recognizes a 39-bp phage attachment site *attP*, 34-bp bacterial attachment site *attB*, and insertion creates hybrid sites called *attL* and *attR* (Groth *et al*., 2000; Smith and Thorpe, 2002). Little conserved sequence exists between sites, and crossover regions vary in length from 2 to 12-bp (Smith and Thorpe, 2002). Its integrase binds *attP, attB, attL,* and *attR* with equal affinities, and catalyses integration without host or phage encoded accessory proteins *in vitro* (Thorpe and Smith, 1998; Smith and Thorpe, 2002). It also has no topological requirement, recombining linear or supercoiled DNA. However, some form regulation/directionality is evident as indicated by the lack of synaptic complex formation and excision between any two sites *in vivo* or *in vitro,* despite equal binding site affinities (Thorpe and Smith 1998; Smith and Thorpe, 2002). Recent work has identified a motif within the C-terminal domain of the ΦC31 integrase itself, involved in directing bound integrase to form a synaptic complex probably through protein–protein interactions (Rowley *et. al*, 2008).

Tn*4451* is capable of both integration and excision using its recombinase TnpX (Lyras *et al*., 2004). During excision, the transposon circularizes itself to form a stable intermediate with an arrangement that provides a strong promoter which transcribes the *tnpX* gene (Bannam *et al*., 1995). TnpX has no known specific insertion recognition sequence, but has been shown to have a much higher binding affinity for either the left or right transposon ends than to known insertion recognition targets (Adams *et al*., 2004). Tn*4451* regulation is unclear. However, its differing binding affinities and ability to function without any host or transposon-encoded accessory factors (Lyras *et al*., 2004) has led to speculation that directionality is regulated through differences in the TnpX-DNA synaptic complex formation and subsequent conformational changes (Adams *et al*., 2004).

Group 3 Other Serine Transposases

The third structural group contains other transposases like IS*607*, and is an interesting mix of structural groups one and two. IS*607* carries two genes, *orfA* and *orfB*, with the recombinase encoded by *orfA* acting as a transposase (Kersulyte *et al*., 2000; Smith and Thorpe, 2002). Surprisingly, *orfB* has protein-level homology to one of two putative transposase genes found in IS*605* and IS*1535*, but was not required for transposition. Group three is similar to the resolvases and invertases of group one in encoding a small recombinase, but the recombinase acts only to integrate and excise IS*607* and similar to the group two TnpX recombinase of Tn*4451*, forms a circular intermediate and has no sequence specific core recognition site. It can also integrate and excise without any host or element encoded accessory proteins. (Kersulyte *et al*., 2000; Smith and Thrope 2002). Currently there is no model regarding regulation of its recombination or expression.

TYROSINE RECOMBINASES

Structure of Tyrosine Recombinases

Tyrosine recombinases, also sometimes referred to as the λ integrase family in reference to its most well studied member, comprise a large group of site-specific recombinases containing over 1000 members by sequence similarity. They are structurally and functionally diverse group consisting of resolvases, invertases, integrases, and transposases. Most tyrosine recombinases have an N-terminal domain that binds DNA, and all share a well conserved C-terminal catalytic region containing the critical tyrosine residue for which the family is named. This latter domain also has a highly conserved nonadjacent five residue active site motif, RKHRH, which participates in acid-base catalysis. The significance and function of all residues is not known, but it is thought that the first arginine assists the lysine in protonation of the 5' hydroxyl after cleavage. The first histidine is not as well conserved and mutagenesis does not lead to inactivity, but it is postulated to act as a base in accepting a proton from the attacking tyrosine (Grindley *et al*., 2006). The second histidine serves an important role forming a hydrogen bond to the cut phosphate (Stivers *et al.*, 2000; Grindley *et al.*, 2006).

Mechanism of Tyrosine Recombinases

Like serine recombinases, all members of this family have the same basic mechanism for recombination involving a two-step transesterfication reaction requiring no ATP or DNA synthesis or repair for completion, only the assembly of a complex consisting of two core sites, each bound by a recombinase dimer at the half-sites (Grindley *et al*., 1982; Grindley *et al*., 2006; Fig. 1.5). However, unlike serine recombinases, only one pair is catalytically active at each step. Consequently, after complex assembly the conserved tyrosine nucleophiles in one recombinase pair directly attack the sugar phosphate backbone of the DNA within the crossover region of

each core site. During cleavage a direct transfer of the phosphoryl group to the tyrosine side chain conserves the energy from the broken phosphodiester bond, and results in a covalent DNA-recombinase intermediate at the 3' end and a free hydroxyl at the 5' end of the DNA in each core site. The free 5' hydroxyl groups in each core site attack either the phosphotyrosyl bond at the 3' end of the original DNA strand or the phosphotyrosyl bond of the opposite core site resulting in formation of original substrate or strand exchange, respectively. If strand exchange occurs, the recombination intermediate known as a Holliday junction (HJ) is formed. Resolution of the Holliday junction is done through isomerization of the entire complex in order to activate the second pair of bound recombinases and inactivate the first pair enabling cleavage and stand exchange of the other untouched DNA strands within the two core sites using the same mechanism (Grindley *et al*., 2006; Fig. 1.5).

Examples of Tyrosine Recombinases

Cre Recombination

Cre recombinase (Cyclization REcombination) is the product of the P1 bacteriophage *cre* gene (Nunes-Duby *et al*., 1998) and is one of two site-specific recombinases encoded on the bacteriophage P1 (Sternberg *et al.*, 1986). It recognizes the 34-bp *loxP* site consisting of two 13-bp inverted repeats flanking an asymmetric 8-bp central region with sequence ATGTATGC (Hoess and Abremski, 1984). Biologically, Cre serves to ensure maintenance and inheritance of P1. Bacteriophage P1 is a large linear 100-kb virion that circularizes after infection, and does not integrate into the host genome, but instead maintains itself as a plasmid within the host cell. Typically P1 relies on host cell homologous recombination at its terminal redundancies to circularize itself. However, in the event that this cannot occur due to infection of a recombination deficient host, Cre will circularize the virion through site-specific

recombination of *loxP* sites located within the terminal redundancies (Hochman *et al*.,1983). In addition, P1 maintains itself as a plasmid, Cre acts to resolve any plasmid dimer formation that may result after DNA replication. This ensures the stable inheritance of P1 to each daughter cell (Austin *et al*., 1981).

Cre is considered one of the simplest site-specific recombinase systems because it requires only the *loxP* site and the Cre recombinase itself for activity *in vitro* (Abremksi and Hoess, 1985). It is through the asymmetry of the site that directionality of recombination is determined. For example, if two directly repeated *loxP* sites are on the same DNA strand, then the DNA segment between those two sites is excised as a covalently closed circle. However, if the two *loxP* sites are inverted, then recombination will result in inversion of the segment. Consequently, the simplicity and easy manipulation of its system has proven to be extremely useful tool in genomics in both prokaryotes and eukaryotes. However, little is understood about its regulation in its native P1 environment. The *cre* gene within P1 has three promoters P_{R1}, P_{R2}, and P_{R3} which are located 304-bp, 124-bp, and 11-bp respectively from the Cre start codon (Sternberg *et al*., 1985). Transcriptional fusions indicate all three promoters are weak, but equal in strength to each other. P_{R3} is thought to be biologically irrelevant, as it does not contain a discernable Shine Delgarno sequence and is not translated. Both P_{R1} and P_{R2} are predicted to contain *dam* methylation sites within their -35 hexamers, however only P_{R1} is sensitive with a 3-4 fold transcriptional increase in *dam*- host (Sternberg *et al*., 1985). It is postulated that immediately following P1 infection, P_{R1} expression would be highest since newly replicated unmethylated P1 DNA would be present for a short period of time. Its high expression levels combined with P_{R2} would subsequently increase the amount of Cre available to circularize P1. After cyclization and methylation, the combined expression of the P_{R2} and less active P_{R1} would

be adequate to maintain Cre levels in the host cell needed to resolve P1 dimers (Sternberg *et al*., 1985).

Not only has the regulation of *cre* expression been investigated, but so has the regulation of its mechanism. Although Cre has shown itself to be a promiscuous recombinase capable of functioning with only the minimal 34-bp *loxP* site *in vitro*, recombination appears to be regulated *in vivo* showing functional specificity and playing a more predominant role as a resolvase than an invertase (Adams *et al*., 1992). Construction of a hybrid *loxP* site next to the regulatory region of the related Xer recombinase, and sequence comparisons of regulatory regions led to the identification of a putative ArgR binding site 97-bp upstream of the *loxP* site (Paul and Summers, 2004). While ArgR is a known regulator in the Xer recombinase system, when tested it was shown to have no direct role in Cre recombination in P1 (MacDonald *et al*., 2008). Currently, there is no model to explain Cre regulation observed *in vivo*, but there is some speculation that is may be FtsK dependent. (MacDonald *et al*., 2008)

Xer Recombination

The Xer system found in *E. coli*, like Cre, resolves, multimers created during cell chromosome replication and like other site-specific recombinases it contains a core site with characteristics described above. There is a pair of 11-bp inverted repeats recognized for recombinase binding that flank the short central region where the crossover occurs. However, Xer is a more complex system than Cre recombination because of the number of recombinases involved, the number of core sites recognized, and the requirement of additional accessory proteins. Xer recombination uses two different recombinases XerC and XerD, chromosomally encoded from the *xerC* and *xerD* genes, respectively. It functions to ensure only single copy chromosome inheritance, by resolving chromosome dimers. Binding of each recombinase is site

specific and dictated by the sequence of the flanking inverted repeats (Blakely *et al*., 1993; Hayes and Sherratt, 1997). XerD has a higher binding affinity for its site, but XerC is always the first to cleave (Blakely *et al*., 1993). Regardless of their distinctions, they interact as a pair because a 5-10 residue deletion of XerC in the C-terminus leads to a 20-fold reduction in XerD binding (Spiers and Sherratt, 1999), while the two recombinases are located in different regions of the chromosome and only share 37% identity with each other, they are still the most closely related of all known tyrosine recombinases (Blakely *et al*., 1993). However, genome database comparisons using the conserved C-terminus of each recombinase found that *Bacillus, Mycobacterium, Haemophilus, Helicobacter, Chlamydia, Rickettsia, Treponema, Pseudomonas, Vibrio, Bordatella, Neisseria, Staphylococcus* and *Enterococcus* species all had two Xer homologues with 23-69% identity, suggesting the mechanism of chromosome dimer resolution found in *E. coli* is highly conserved (Recchia and Sherratt, 1999).

In addition to use of two different recombinases, the Xer system also differs from Cre by its recognition of multiple core sites (*cer, psi, dif, dib, cer3, cer6,* etc.) found on either the chromosome or various plasmids. The three most studied are the *cer* site found on plasmid ColE1 containing an 8-bp central region (TTAAGGGA), the *psi* site found on pSC101 containing a 6-bp central region (GATCCA), and the *dif* site found on the chromosome containing a 6-bp central region (TGTATA). Each of these core sites dictates a different recombinational specificity. The basis of specificity is better understood at the *cer* and *psi* core sites, which depend on host encoded accessory proteins. The binding sites for these host TF proteins are 178-bp and 158-bp from the XerC binding site for *cer* and *psi*, respectively. The 178-bp region of the *cer* site binds PepA, a leucine aminopeptidase, and ArgR the arginine biosynthesis repressor, and the 158-bp region of the *psi* site binds PepA and ArcA, an anaerobic

repressor. These proteins serve structural, not catalytic roles. Binding a single PepA hexamer (Reijns *et al*., 2005) is critical at both sites to act as an architectural element by winding two core sites around each other in a nucleoprotein complex that brings sites together in the topography needed to support recombination (Alen *et al*, 1997; Colloms *et al*., 1998). It is unclear what roles ArgR and ArcA play in *cer* and *psi*, respectively, since they both bind, but neither is essential for recombination to occur. Hybrid sites constructed from Cre *loxP* and *cer* or *psi* accessory sequences indicate neither ArgR nor ArcA can initiate recombination if bound alone, but each increases the efficiency of recombination at their site (Gourlay and Colloms, 2004).

Recombination regulation at the chromosomal *dif* site is quite different and does not require adjacent accessory protein binding, although other cellular factors may be involved since *E. coli* XerCD cannot successfully complete recombination at *dif* sites on supercoiled plasmids *in vitro* (Tecklenburg *et al*., 1995). The current model supports the involvement of the C-terminal domain of the FtsK protein which is involved in chromosome segregation (Yu *et al*., 1998). Cells lacking FtsK, are defective at *dif* site recombination (Steiner, *et al.* 1999). However, since the formation of the recombination intermediate Holiday Junction (HJ) is not hindered in the absence of FtsK, it is thought that FtsK performs a role similar to PepA and provides a conformational change necessary to allow for initiation of catalysis by XerD leading to complete recombination (Barre *et al.*, 2000).

The regulation of XerCD expression is subject to some speculation. It is odd that the two *xer* gene products work in tandem, but are not near each other on the chromosome and are divergently transcribed, which is very different from the other site-specific recombination system that use two recombinases, FimB and FimE, located adjacent to each other (Klemm, 1986). Investigation of neighboring *xer* genes in *E. coli* found the *recJ* gene co-expressed with *xerD* and

the *uvrD* and *uvrQ* genes proximal to the *xerC* gene. The possible although unlikely coincidence that all three neighboring genes have either a direct or indirect involvement in recombination, repair, or stress response has led to speculation regarding their involvement in *xerC* and *xerD* regulation. Little research has been done to address this possibility, although support for this idea was demonstrated when mutations within *xerC* were shown to increase the basal level of SOS expression (Hendricks *et al*., 2000).

Lambda Integrase Recombination

Lambda integrase is the most complex and well-studied of all the tyrosine site-specific recombinases. It is found on the chromosome of the temperate lysogenic bacteriophage λ and is responsible for the integration and excision of the lambda chromosome into and out of the *Escherichia coli* host chromosome. Its complexity derives from multiple interacting core sites (*attP, attB, attL* and *attR*) and the regulation and direction of recombination driven by a network of host and phage encoded accessory proteins (Landy and Ross, 1977; Nash and Robertson, 1981). Like other tyrosine site-specific recombination systems, the host chromosome core site *attB* consists of two recombinase binding sites called B and B' that flank a 7-bp central crossover region; while the bacteriophage core site *attP* consists of two recombinase binding sites C and C', that flank a homologous 7-bp central region. Once the lambda bacteriophage has integrated in host chromosome, the resulting *attL* and *attR* core sites are created consisting of the same 7-bp homologous core crossover region each with flanking recombinase binding sites consisting of B and C' and C and B' respectively. (Landy and Ross, 1977)

In addition to its multiple sites, lambda's complexity as a recombination system can also be attributed to its two distinct integrase binding sites. The first type is the core-type consisting of the two binding sites within the core sites that flank the central crossover region (B, B', C, C')

and are characteristic of all members of this family. The second type of site is a group referred to as arm-type. Distinction between types is determined through Int domain recognition, with the large carboxy-terminal domain of Int recognizing core-type sites, while arm-type sites are recognized by the amino-terminal domain (Landy, 1989). Lambda has four core-type sites discussed above and five arm-type binding sites: P_1, P_2, P'_1, P'_2, and P'_3. All five arm-type sites contain the characteristic consensus 7-bp central region. Despite this similarity, Int has a higher binding affinity to arm-type sites than core. λ also uses occupancy of different sites (in addition to other factors described below) to drive recombination directionality and it is currently believed that P_1, P'_2, and P'_3 occupancy is required for insertion, while only occupancy of P'_1 is needed for excision.

Lambda recombination is subject to regulation through host accessory proteins IHF, FIS, and phage accessory protein XIS. IHF has no catalytic activity, but plays an architectural role with three binding sites referred to as H_1, H_2, and H' adjacent to *attP*. When all sites are occupied, a nucleoprotein complex is formed between IHF and integrase bound to *attP* arm-type sites P_1, P'_2, and P'_3. Within this complex, bound IHF bends DNA so that integrase bound with high affinity to arm-type sites is brought near the core site to encourage integrase binding to low affinity core-type sites. Once bound this permits catalytically active integrase pairs to bind with host *attB* leading to integration of phage DNA (Rice *et al.*, 1996). Additionally the host encoded DNA bending protein FIS increases insertion recombination 2-fold *in vivo* (Ball *et al.*, 1991). Currently the details of FIS involvement are still elusive, and initial in vitro work did not corroborate this finding. However more recent in vitro work with conditions matching the natural physiology of the cell using suboptimal levels of integrase support the initial in vivo work

(Esposito *et al*., 2003). These results led to the speculation that FIS directly binds the F site bending DNA and inducing a topological change that allows integrase interactions.

Excision also requires IHF, but is driven by the phage encoded protein XIS, whose presence is inhibitory to integration. Like IHF, XIS does not have catalytic activity, but acts as another essential architectural protein whose three binding sites referred to as X_1, $X_{1.5,}$ and X_2 are located adjacent to *attR*. XIS bound at $X_{1.5}$ forms intermolecular contacts with the Xis monomers bound at X_1 and $X_{2,}$ forming of a micronucleoprotein filament over the entire Xis binding region which physically prevents IHF and integrase from occupying the P_1 H_1 P'_3 sites that are critical to integrative recombination (Abbani *et al*., 2007). Oddly despite its essential role in determining the directionality of recombination, XIS has little to no binding affinity for any of its sites *in vitro* (Papagiannis *et al*., 2007). This low affinity is countered by FIS, whose binding site called the F site partially overlaps the X_2 site. FIS binds the F site with 100-fold greater affinity than XIS to any of its three sites (Papagiannis *et al*., 2007). This might lead to the erroneous conclusion that FIS completes with XIS for the F/X_2 binding sequence consequently reducing the number of bound XIS leading to decreasing excisive recombination efficiency. However, FIS binding strongly stimulates excisive recombination by 20-fold (Thompson *et al*, 1987). The two proteins work cooperatively, with FIS enhancing XIS binding by recruiting XIS to the X_2 site. It is postulated that once bound, XIS can then recruit other XIS monomers to the X_1 and $X_{1.5}$ sites (Papagiannis *et al*., 2007). This hypothesis is supported by the finding that three XIS monomers and one FIS dimer are all bound in a nucleoprotein complex. (Sun *et al*., 2006) However, it should be noted that XIS binding is not wholly dependent on FIS. Integrase bound at the arm P_2 site located adjacent to XIS binding sites, recruit XIS and lead to a 16-fold increase in XIS binding (Thompson *et al*., 1987).

Lambda integrase transcription regulation is just as convoluted as the regulation of its recombination mechanism with multiple variables that allow integrase to be made from different transcripts in different conditions. Controlling integrase expression ensures prophage insertion and excision during conditions suited to lysogeny. This regulatory system involves five major components: prophage proteins CII and N, *sib*, a site of endonucleolytic RNase III cleavage, and two tandem promoters P_{int} and P_L. During infection when sufficient prophage protein CII is made, P_{int} transcription of integrase is initiated (Reichardt, 1975). These transcripts end at the T_{int} terminator and contain an incomplete *sib* site protecting them from RNase degradation (Plunkett and Echols, 1989), but transcripts initiating from the stronger P_L promoter encode the phage protein N, which suppresses the T_{int} terminator producing transcripts with the entire *sib* sequence subsequently leading to exonucleolytic degradation of the *int* message. However, after insertion, the P_L transcript no longer contains the *sib* site due to a permutation of prophage genes that occurs during insertion and Int is then translated from this transcript (Court *et al*., 2007).

INTEGRONS

Distinctions Among Integrons

Integrons are unique members of the tyrosine site-specific recombinase family that are found in chromosomes, transposable elements and/or conjugative plasmids. They move gene cassettes that encode a variety of gene products, most notably antibiotic resistance. However, they differ from other members in the family in three distinct ways: site variability, protein structure/mechanism, and their further division into separate distinct classes. Unlike most other tyrosine recombinase family members whose successful recombination relies on sequential strand exchange within a prototypical core site composed of a pair of highly conserved 9-13bp inverted binding sites separated by a 6-8bp spacer region of identical sequence, integrons use

only a single strand exchange and have an highly flexible core recognition site sequence with no requirements for central region sequence homology. The significance of this versatility is great, as the only other site-specific systems able to tolerate non-homology within the central region are found in Tn*916* and Tn*1545* (Craig *et al*., 2002). Still, like other tyrosine recombinase family members, integron recombinases have five conserved RKHRH residues and mediate recombination using a conserved tyrosine nucleophile. However, they differ in that they contain an additional domain of 20 to 22 amino acids with a conserved motif of 15 predominantly non-polar amino acids in the catalytic C terminus forming an additional α-helix (Messier and Roy, 2001; Craig *et al*., 2002; MacDonald, 2006). Deletion or substitution of some of some of these conserved amino acids can eliminate DNA binding and/or recombination activity due to their involvement in orchestrating the unique stepwise assembly of the synaptic complex between recognition site and the recombinase (Messier and Roy, 2001; MacDonald *et al*., 2006) . Finally, integron recombinases differ from other tyrosine recombinases as they are further divided into four separate classes based on amino acid sequence homology.

Structure of Integrons

Integrons are composed of a variable region which encodes mobile, non-self-replicating elements referred to as cassettes which contain an ORF and integrase-specific recombination site called *attC* (Fig. 1.6; Hall and Stokes, 1993). This variable region is flanked by conserved 5' and 3' regions. The flanking 5' region is comprised of the integrase gene (*intI*) itself, an adjacent recombination site called *attI* where new cassettes are typically added, and at least two convergent promoters (P_{int} and P_c). The P_{int} drives transcription of *intI*, P_c directs transcription of acquired gene cassettes within the variable region, and a rare secondary cassette promoter P_2 is sometimes present due to a 3-bp addition creating ideal 17-bp spacing for adventitiously spaced

-35 and -10 hexamers. The conserved 3' region consists of a quaternary ammonium compound resistance gene (*qacE*Δ1), a sulfonamide resistance gene (*sulI)*, and an open reading frame of unknown function (ORF 5) (Fig. 1.6; Hall and Stokes, 1993).

The *attI* site encoded in the 5' conserved region is approximately 65bp in length, with little sequence identity among the several integron classes. However *attI1, attI2, and attI3* sites all carry the prototypical core site composed of two imperfect inverted repeats (L and R) with the sequence RYYYAAC and GTTRRRY (where R is a purine and Y is a pyrimidine), separated by a 6 to 8-bp spacer region (depending on integron class) of varied sequence. In addition to the core site, the *attI* site contains two imperfect direct repeats (DR1 and DR2) located upstream that are thought to serve in recombination and will be discussed later. The *attC* site at the end of each cassette was previously referred to as the "59-base element," as it was originally thought to be only 59-bases long (Hall *et al.*, 1991). The *attC* sites of gene cassettes can move between the integron classes and have great variation in sequence and length (57 to 141 bases) (Recchia and Hall, 1995). Despite length and sequence differences, *attC* sites do have some common features as each contains two potential core sites (L' and R') and (L" and R"), arranged as imperfect inverted repeats and the RYYYAAC and GTTRRRY core site recognition sequences, separated by a 6 to 8-bp spacer region of varied sequence.

Classes of Integrons

There are at least 90 distinct integron classes, most of which are located on chromosomes (Mazel, 2006; Boucher *et al*., 2007). However, only four classes have been well studied with class 1 being the best characterized, but all sharing 40-59% identity with each other (Collis *et al.*, 2002). Classes 1, 2, and 3 are mobile integrons involved in the dissemination of antibiotic resistance due to their linkage to mobile elements like transposons (e.g. Tn*7*), and conjugative

plasmids (e.g. NR1). Class 1 in particular is found extensively in clinical and environmental isolates in both Gram-positive and Gram-negative bacteria (Nandi *et al*., 2004), with over 80 different resistance gene cassettes (Mazel, 2006). In comparison, the original (Tn*7*) class 2 integron has only six different resistance cassettes, a difference attributed to a mutation in codon 179 of the class 2 integrase that creates a stop codon resulting in a truncated, non-functioning protein (Hansson *et al*., 2002). A functional class 2 integron lacking this mutation was recently discovered in environmental (Barlow and Gobius, 2006) and clinical samples (Marquez et al., 2008). Class 3 integrons are not widely found, but are functional (Collis *et al*., 2002) and have been found with two different resistance cassettes (Correia *et al.*, 2003)

Class 4 which has been renamed VchIntIA, found on chromosome 2 of *Vibrio cholerae* is referred to as a superintegron (Mazel *et al*., 1998). Since its discovery in the 1990's, other superintegrons were found throughout the proteobacteria, pseudomonads, xanthomonads, and several other bacterial groups, all sharing the same general characteristics (Mazel, 2006). First, they were all located on the chromosome and not associated with mobile elements as is the case with classes 1, 2, and 3 (Mazel, 2006). Second, the superintegrons typically carry a large number of cassettes, over 200 in some cases (Chen *et al*., 2003). In contrast, class 1 integrons have never been found with more than eight cassettes (Naas *et al*., 2001). Third, the *attC* sites of these cassettes had greater than 80% identity (Mazel, 2006). Thus, it is not surprising that IntI from each class preferentially recognizes its own *attI* site, but it is surprising to find that IntI from each class is able to recombine *attC* sites of different sequences and sources with comparable efficiency to the cognate protein (Collis *et al*., 2002; Hansson *et al*., 2002; Biskri *et al*., 2005). These findings suggest each integrase uses the same recombination mechanism for *attC*, but a class specific mechanism when recombining at *attI.*

Mechanisms of Integron Recombination

The recombination mechanisms used by integrons is very distinct from other site-specific systems due to their ability to recombine dissimilar core sites with the variable recognition and spacer sequences described above. In addition, integrons use of only a single strand exchange by cleavage and crossover occurring between the G and the first T in the R region of *attC* and the corresponding R' region of *attI* (Hannson, *et al.*,1997; Stokes *et al.*, 1997). This result led to speculation that the subsequent Holiday Junction intermediate formed is later resolved through DNA replication or an unidentified resolvase. Determining the binding preferences of IntI1 within both *attI* and *attC* sites has led to model which supports resolution via DNA replication.

IntI1 preferentially binds double-stranded *attI*, but favors single stranded *attC*, specifically, only the bottom strand of single stranded *attC* (Francia *et al*., 1999). Despite the extreme variability in *attC* sites, this result was observed in both class one (*aadA1* and *aadA7*) and VchIntIA *VCR2/1* cassettes, suggesting it may be a general characteristic found in other classes (Francia *et al*., 1999; Bouveir *et al*., 2005). A closer inspection *in vitro* of a single-stranded *attC* oligo found it could form a double-stranded-like configuration by binding its imperfect inverted repeats (L" to L' and R" to R') at either end of the site (Hall *et al*., 1991; Stokes *et al*., 1997; Rowe-Magnus *et al*., 2003; Bouvier *et al*., 2005; Fig. 1.7). This finding was later confirmed when mutations altering the potential cruciform formation interfered with IntI binding (Johansson *et al*., 2004). These findings led to the current model in which single stranded *attC* made available by replication or transcription is bound by IntI thereby stabilizing the single stranded form. Recombination then takes place via a strand exchange between single-stranded *attC* and double-stranded *attI* leading to a HJ intermediate that is resolved through an additional replication step (Bouvier *et al*., 2005; Fig. 1.7). Other recombinases use a second

strand exchange in order to resolve the HJ intermediate. However if done in this system it would result in either formation of the original substrates via reverse recombination, or the formation of an abortive covalent linear molecule (Bouvier *et al*., 2005).

Integron-type Integrase Recombination

IntI1 integrase can insert or excise gene cassettes in the form of covalently closed circular ds-DNA intermediates (Collis and Hall, 1992) or through co-integrate formation or resolution (Shearer and Summers, 2009). Recombination needs only IntI1, and can occur with any combination of sites, i.e. between *attI* x *attC*, *attC* x *attC*, or *attI* x *attI*. Both *attI* and *attC* sites have occasionally been found to recombine with non-specific secondary sites (2^{o}rs) (Francia *et al*., 1993; Recchia, *et al*., 1994; Hansson *et al*., 1997).

Despite its versatility and lack of prototypical site-specific recombination sites, distinct preferences are apparent among sites as new cassettes are typically added at *attI* (Recchia *et al*., 1994). This finding was supported by a study of all potential recombination site combinations catalyzed by IntI1, which found *attI* x *attC* to be the most efficient although highly variable with recombination frequencies ranging from 1.1×10^{-2} to 8.5×10^{-5} among seven different *attC* sites (Collis *et al*., 2001). Next best was *attC* x *attC*, and *attI* x *attI* was least efficient by 10-fold or more. However, while *attI* x *attC* was preferred in most cases, the frequency of this combination was highly variable (Collis *et al*., 2001).

The question of how the integrase recognizes and cleaves the correct location despite the great sequence diversity among *attI* and *attC* sites was answered for *attC* with a crystal structure of V. *cholerae* integrase (VchIntIA) bound to a single *attC* bottom strand of a *V. cholerae* superintegron repeat (VCR) sequence. The site recognition occurs through DNA secondary structure by positioning two outward flipped bases (T12" and G20") that occur when single-

stranded *attC* folds into a cruciform via its imperfect inverted repeats (R', L', R", L"). These two extrahelical bases interact with the integrase in *cis* and in *trans* and are thought to correctly position the protein on the DNA and mediate the stepwise assembly of the synaptic complex. The remaining integrase-DNA interactions are non-specific phosphate contacts (MacDonald *et al*., 2006). VchIntIA binds as a dimer to the two antiparallel *Vibrio cholera* repeat bottom strand (VCR_{bs}) duplexes. The N- and C-terminal domains of the attacking subunit fold at Lys 160 and wrap around the DNA forming a clamp. The non-attacking subunit interacts with extrahelical nucleotide T12" in *cis* and is inserted between His240 and Pro232 dictating the integrase dimer position on the DNA. The G20" extrahelical nucleotide eight bases from T12" interacts in *trans* with the attacking subunit across the synaptic interface, binding in a hydrophobic pocket formed by Trp157 and Trp219. This latter interaction ensures proper substrate placement and orientation for recombination and accounts for the IntI integrase's ability to recombine *attC* sites with little sequence similarity. However, the answers to the integron's dual site specificity is not apparent as *attI* does not recombine single-stranded (MacDonald *et al*., 2006).

Regulation of Integron Recombination and IntI Integrase Expression

Regulation of integron integrase recombination is not completely understood. IntI is the only mobile element protein known to be necessary for site-specific recombination, and there are no recognition sites for DNA binding proteins commonly used in other site-specific recombination systems adjacent to either *attI* or *attC* sites (Martinez and de la Cruz, 1988; Hallet and Sherratt, 1997). Recent discoveries regarding the recombination mechanism indicate possible dependence on the availability and abundance of single-stranded *attC*. However, single-stranded DNA concentration cannot be the only factor. Previous work found that the two directly repeated integrase binding sites (DR1 and DR2) in *attI* are critical for *attI* x *attC*

recombination, but are not required for *attI* x *attI* recombination. This led to speculation that additional integrase molecules bound at DR1 and DR2 serve a yet unexplained accessory function possibly similar to what is seen in lambda integrase (Recchia *et al*., 1994; Hansson *et al*., 1997; Partridge *et al*., 2000).

Little is known about the transcription of *intI1* and or of the captured gene cassettes. P_{int} lies within the 5' conserved region and drives transcription of *intI1*. It has little sequence variation in class 1, but differs among other integron classes. Much attention has been directed toward understanding cassette expression. With few exceptions most gene cassettes lack their own promoter, transcripts beginning at the P_c promoter in the 5' conserved region, are read as one transcript, although only the most *attI* proximal cassettes are detected phenotypically (Stokes and Hall, 1991; Collis and Hall, 1995; Recchia and Hall, 1995). Indeed, cassette expression is reduced when situated downstream of one or more other cassettes (Collis and Hall, 1992). Interestingly, the reduction in cassette expression levels is not only dependent on the number of cassettes before it, but also on the identity of each preceding gene cassette. This observation led to speculation that premature transcriptional termination at the 3' end of some gene cassettes may arise from formation of cruciform structures in their *attC*'s (Collis and Hall, 1995). This explanation seems unlikely as all cassettes are predicted to form this cruciform structure as it necessary for IntI integrase recognition of *attC*.

Cassette expression also varies due to polymorphisms in the promoter. Initial sequence analysis revealed four distinct naturally occurring versions of the cassette promoter(s) (Table 1.4). The relative strength of each version was compared to *tac* promoter and showed a wide range of relative transcription proficiency varying from 0.2 to 6.5 of the activity of de-repressed *tac*, and this resulted in their being named 'strong', 'weak', hybrid' and 'weak + 2^{nd}',

respectively (Levesque *et al*., 1994). The secondary promoter in 'weak + 2nd' refers to P_2 which becomes active due to a 3-bp insertion that creates a consensus 17-bp spacer region between -10 and -35 hexamers. However, as more integrons were characterized three additional versions were identified, 'hybrid 2' (Lagatolla *et al*., 2006; Colinon *et al*., 2007), 'strong + 2nd' (Pournaras *et al*., 2005), and 'hybrid 1 + 2nd' (Gonzalez-Zorn *et al*., 2005). The relative strengths of each were determined and compared to the initial four and found 'hybrid 2' equal to 'weak + 2nd', 'hybrid 1 + 2nd' greater than 'hybrid 1' alone, and 'strong + 2nd' showed a slight increase over 'strong' alone (Papagiannitis *et al*., 2009; Table 1.4). The variability among cassette promoters may have evolved in order to accommodate varied expression observed among *attC* sites.

Interestingly, cassettes have a very long leader region as P_c is located approximately 200-bp upstream from cassettes in the *attI* site, but there is no published work regarding any role for the long leader in regulation of cassette expression. There is, however, some evidence of promoter involvement in recombination frequency. Substrates with the strong cassette promoter (P_c) recombined cassettes with lower frequency compared to the hybrid or weak versions (Collis *et al*., 2001). Moreover, a transcriptome analysis of the gene cassettes in the superintegron in *V. cholerae* strain N16961 comparing *hapR* (virulence factor TcpP regulator), *rpoS* (stationary/starvation phase sigma factor), or *rpoN* (nitrogen-limitation sigma factor) mutant strains, revealed that the majority of cassettes increased transcription at high cell densities and in response to stress or in stationary phase (Yildiz *et al*., 2004). The authors speculated that cassettes are activated in stressed or non-replicating populations to provide a survival advantage to cells bearing beneficial cassettes conferring stress resistance. However, it is unclear whether integrase expression also increases (Yildiz *et al*., 2004). Recently a chromosomal and a mobile

integron comparison revealed that the SOS response controlled transcriptional regulation of integrase promoter (P_{int}) and will be addressed in Chapter 3 (Guerin *et al*., 2009).

Translational control has received little attention as most gene cassettes are thought to carry their own Shine-Dalgarno sequence. However 25% of antibiotic resistance genes do not have a plausible translation initiation region (TIR) consisting of an initiation codon, spacer region, or a Shine-Dalgarno sequence (Hanau-Bercot *et al*., 2002). Speculation regarding their translation revolves around a small ORF found in the class 1 integron only, which is located in the *attI* site (Fig. 1.8). This so called ORF-11 encodes 11 amino acids, uses a consensus Shine-Delgarno sequence, and is located 2 to 62-bp upstream of gene cassettes in the "X region" (depending on cassette). It has no known function, but separate experiments show its involvement in at least two cassettes providing an initiation codon to support *aacC1* expression in Tn*1696* (Wohlleben *et al*., 1989); *aacA-luc* translational fusions revealed it provides a ribosome binding site for the cassette provided initiation codon (Hanau-Bercot *et al*., 2002). Interestingly, cassette insertion often results in placement of a stop codon in frame with ORF-11. The significance or impact of the stop codon is not known, but it is possible that terminated ribosomes are not released, but instead go on to restart translation at the cassette initiation codon. It is also possible that ORF-11 and its neighboring cassette are subject to translational coupling, seen when start and stop codons are close to each other or overlap (Hanau-Bercot *et al*., 2002). Regardless, apart from cases of ORF-11 fusion, the mechanism of translational initiation of those cassettes without a TIR is unclear.

Integron and Cassette Origins

Class 1 integrons are homologous to transposons derived from Tn*402* embedded in larger transposons (Brown *et al.*, 1996); class 2 are homologous to Tn*7* derivatives (Sundstrom *et al*.,

1991), and class 3 more distantly homologous to class 1, are also located in a Tn*402* related transposon, but in an opposite orientation from class 1 (Collis *et al*., 2002). How, where, and when did they come to exist? A hypothesis regarding the evolution of the class 1 was recently introduced after the discovery of two class 1 integrons in environmental bacteria isolated from sediment samples, which appear to be chromosomally located and predate their Tn*402* association (Stokes *et al*., 2006). Further analysis revealed the integrons could be isolated from several non-pathogenic members of *β-proteobacteria,* none of which carried known antibiotic resistance cassettes and despite their chromosomal location, were mobile via an unknown mechanism as Tn*402* features were absent (Gillings *et al.*, 2008). These findings led the authors to conclude that the class 1 integron originated with this group (or one similar) and was subsequently incorporated into a plasmid carrying the Tn*402* transposon. The source of this *β-proteobacteria* was an agricultural pond which might have experienced selective pressures of antibiotic use (Gillings *et al*., 2008). Integrons of classes 2 and 3 may have similar environmental ancestors as a functional class 2 was isolated from *Providencia stuartii* carrying no antibiotic resistance cassettes (Barlow and Gobius, 2006) and a chromosomal class 3 integron from *Delftia* (Xu *et al*., 2007).

Cassettes are thought to be recruited from superintegrons through an unknown mechanism. Comparisons of superintegron gene cassettes from different *Vibrio* hosts show the majority of them are host-specific, in contrast to gene cassettes in mobile elements that have different codon-use even in the same mobile element implying different origins (Rowe-Magnus *et al.*, 2003; Mazel, 2006). Consequently, it would appear mobile integron classes 1, 2, and 3 have been moving cassettes recruited from different superintegrons as well as from other related sources yet to be discovered. Interestingly, no antibiotic resistance cassette associated with

classes 1, 2, and 3 from clinical isolates has been identified in a superintegron. However, several superintegron cassettes encode genes highly homologous to antibiotic resistance genes, which if exposed to drug selection pressure would have the potential to result in a resistance phenotype (Rowe-Magnus, 1999).

Significance of Integrons

Integron involvement in the dissemination of antibiotic resistance genes has facilitated the rapid formation of multi-drug resistance gene arrays. For example, only six years after the introduction and large scale production of streptomycin, tetracycline, and chloramphenical, *Shigella dysenteriae* isolates resistant to all of three antibiotics were identified (Mazel, 2006). Currently, over 80 different gene cassettes from the class 1 integron have been described conferring resistance to all known β–lactams, aminoglycosides, chloramphenical, trimethoprim, streptothricin, rifampin, erythromycin, fosfomycin, lincomycin, and antiseptics of the quaternary-ammonium-compound family (Mazel, 2006). In addition to antibiotic resistance, some superintegrons encode virulence factors and pathogenicity determinants in *Vibrio* species (Ogawa and Takeda, 1993; Smith and Siebeling, 2003). However, interest in them should not be driven exclusively by the threat they pose toward human health. Their origins and involvement as a general gene capture system promoting bacterial adaptation means they are potentially holding a cornucopia of undiscovered proteins which could result in tremendous value to both, the applied science and the biotechnology, as well as offer further insight into bacterial diversity and evolution.

SUMMARY

The serine and tyrosine site-specific recombinases have evolved as two distinct families with each family using different mechanisms for DNA synapsis, cleavage, strand exchange, and forms of regulation. Despite the distinctions between the two families, members within each do share some basic characteristics. However, there can be no doubt that the integron integrase recombinase system is a member of the tyrosine family that posses several unique characteristics: (i) it recombines using a novel mechanism of site recognition and cleavage resulting in a single strand exchange, then resolving HJ intermediates via cell replication; (ii) it possesses dual site specificity requiring both single (*attC*) and double stranded (*attI*) forms of its recognition sites, both of which have little to no sequence homology; (iii) it contains an additional conserved domain not found in other tyrosine recombinase family members; and (iv) it is further divided into separate related classes on the basis of sequence homology.

The major role integrons play in the dissemination of antibiotic resistance among bacterial populations and their unquantified impact in bacterial evolution assigns them tremendous importance. Although recent discoveries have identified the mechanism of cleavage and resolution at its *attC* site, there is still much to learn as the reactions at the *attI* site remain unknown. There has also been little work on how expression of the integrase or the gene cassettes is regulated, although both transcriptional and translational control appear to be involved for the superintegron and class 1 integron.

REFERENCES

Abbani, M., Papagiannis, C.V., Sam, M.D., Cascio, D., Johnson, R.C., and R.T. Clubb. (2007) Structure of the cooperative Xis-DNA complex reveals a micronucleoprotein filament that regulates phage lambda intasome assembly. *Proc. Natl. Acad. Sci*. **104:**2109-2114.

Abremski, K., and R. Hoess. (1985) Phage P1 Cre-*loxP* site-specific recombination. Effects of DNA supercoiling on catenation and knotting of recombinant products. *J. Mol. Biol*. **184:** 211-220.

Adams, V., Lucet, I.S., Lyras, D., and J.I. Rood. (2004) DNA binding properties of TnpX indicate that different synapses are transformed in the excision and integration of the Tn*4451* family. *Mol. Micro*. **53:**1195-1207.

Adams, D.E., Bliska, J.B., and N.R. Cozzarelli. (1992) Cre-lox recombination in *Escherichia coli* cells. Mechanistic differences from the *in vitro* reaction. *J. Mol. Biol.* **226:**661-673.

Adrian, P.V.,Thomson, C.J., Klugman, K.P., and S.G.B. Amyes. (2002) New gene cassettes for trimethoprim resistance, *dfr13*, and streptomycin-spectinomycin resistance, *aadA4*, inserted on a class 1 integron. *Antimicrob Agents Chemother.* **44:**355-61.

Alen, C., Sherratt, D.J., and S.D. Collums. (1997) Direct interaction of aminopeptidase A with recombinase site DNA in Xer site-specific recombination. *J. EMBO.* **16:**5188-5197.

Austin, S., M. Ziese, and N. Sternberg. (1981) A novel role for site-specific recombination in maintenance of bacterial replicons. *Cell*. **25:**729-736.

Azam, T.A., Iwata, A., Nishimura, A., Ueda, S., and A. Ishihama. (1999) Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol*. **181:**6361-6370.

Ball, C.A. and R.C. Johnson. (1991) Multiple effects of Fis on integration and the control of lysogeny in phage λ. *J. Bacteriol*. **173:**4032-4038.

Bannam, T.L., Crellin, P.K., and J.I. Rood. (1995) Molecular genetics of the chloramphenical-resistance transposon Tn*4451* from *Clostridium perfringens*: the TnpX site-specific recombinase excises a circular transposon molecule. *Mol. Microbiol*. **16:** 535-551.

Barlow, R. and K. Gobius. (2006) Diverse class 2 integrons from beef cattle sources. *J. Antimcirob. Chemother.* **58:**1133-1138.

Barnard, A., Wolfe, A., and S. Busby. (2004) Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. *Current Opinions in Microbiology*. **7:**102-108.

Barre, F., Aroyo, M., Colloms, S.D., Helfrich, A., Cornet, F., and D.J. Sherratt. (2000) FtsK functions in the processing of a Holliday junction intermediate during bacterial chromosome segregation. *Genes Dev.* **14:**2976-2988.

Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., and R. Phillips. (2005) Transcriptional regulation by the numbers: applications. *Current Opinion In Genetics & Development.* **15:**125-135.

Biskri, L., Bouvier, M., Guerout, A.M., Boisnard, S., and D. Mazel. (2005) Comparitive study of class 1 integron and *Vibrio cholerae* superintegron integrase activities. *J. Baceriol*. **187:**1740-1750.

Blakely, G., May, G., McCulloch, R., Arciszewska, L.K., Burke, M., Lovett, L.T., and D. J. Sherratt. (1993) Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E. coli* K-12. *Cell* **75:** 351-361.

Blattner, F.R., Plunkett, G., Bloch, C.A, Perna, N.T, Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., and G. Mayhew. (1997) The complete genome sequence of *Escherichia coli*. *Science*. **277:** 1453-1462.

Boocock, M.R., Brown, J.L., and D.J. Sherratt. (1986) Structural and catalytic properties of specific complexes between Tn*3* resolvase and the recombination site res. *Biochem. Soc. Trans*. **14:**214-16.

Boucher, Y. Labbate, M., Koenig, J. and H.W. Stokes. (2007) Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends Microbiol.* **15:**301-309.

Bouvier, M., Demarre, G., and D. Mazel. (2005) Integron cassette insertion: a recombination process involving a folded single strand substrate. *J. EMBO* **24:**4356-4367.

Brown, H.J., Stokes, H.W., and R.M. Hall. (1996) The integrons In0, In2, and In5 are defective transposon derivatives. *J. Bacteriol.* **178:**4429-4437.

Callen, B.P., Shearwin, K.E., and J.B. Egan. (2004) Transcriptional interference between convergent promoters caused by elongation over the promoter. *Mol. Cell*. **14:**647-656.

Chen, C.Y.,Wu, K.M. Chang, Y.C., Chang, C.H., Tsai, H.C., Liao, T.L., Liu, Y.M., Chen, H.J., Shen, B.T., Li, J.C., Su, T.L., Shao, C.P., Lee, C.T., Hor, L.I., and S.F. Tsai. (2003) Comparative genome analysis of *Vibrio vulnificus*, a marine pathogen. *Genome Res.* **13:**2577-2587.

Colinon, C., Miriagou, V., Carattoli, A., Luzzaro, F., and G.M. Rossolini. (2007) Characterization of the IncA/C plasmid pCC416 encoding VIM-4 and CMY-4 β-lactamases. *J. Antimicrob. Chemother*. **60:**258-262.

Collis, C.M., Kim, M. Stokes, H.W., and R.M. Hall. (2002) Integron-encoded IntI integrases preferentially recognize the adjacent cognate attI site in recombination with a 59-be site. *Mol. Microbiol.*. **46:**1415-1427.

Collis, C.M., Kim, M., Partridge, S., Stokes, H.W., and R.M. Hall. (2002) Characterization of the class 3 integron and the site-specific recombination system it determines. *J. Bacteriol.* **184:**3017-3026.

Collis, C.M., Recchia, G.D., Kim, M., Stokes, H.W., and R.M. Hall. (2001) Efficiency of recombination reactions catalyzed by class 1 integron integrases IntI1. *J. Bacteriol*. **183:**2535-2542.

Collis, C.M. and R.M. Hall. (1995) Expression of Antibiotic Resistance genes in the integrated cassettes of integrons. *Antimicrobial Agents and Chemotherapy*. **39:**155-162.

Collis, C.M. and R.M. Hall. (1992) Gene cassettes from the insert region of integrons are excised as covalently closed circles. *Mol . Microbiol*. **6:**2875-2885.

Collis, C.M. and R.M. Hall. (1992) Site-specific deletion and rearrangement of integron insert genes catalyzed by the integron DNA integrase. *J. Bacteriol*. **174:**1574-1585.

Colloms, S.D., Alén, C., and D.J. Sherratt. (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.

Correia, M. Boavida, F., Grosso, F., Salgado, M.J., Lito, L.M., Cristino, J.M., Mendo, S. and A. Duarte. (2003) Molecular Characterization of a New Class 3 Integron in *Klebsiella pneumonia.* **47:** 2838-2843.

Court, D.L, Oppenheim, A.B., and S. Adhya. (2007) A new look at bacteriophage λ genetic networks. *J. Bacteriol*. **189:** 298-304.

Craig, N.L., Craigie, R., Gellert, M., and A.M. Lambowitz (ed). (2002) Mobile DNA II. American Society for Microbiology, Washington, D.C.

Crellin, P.K., and J.I. Rood. (1997) The resolvase/invertase domain of the site-specific recombinase TnpX is functional and recognizes a target sequence that resembles the junction of the circular form of the *Clostridium perfringens* transposon Tn*4451*. *J. Bacteriol.* **179**:5148–5156.

Epshtein, V. and E. Nudler. (2003) Cooperation between RNA polymerase molecules in transcription elongation. *Science.* **300:**801-805.

Esposito, D., and G.F. Gerard. (2003) The *Escherichia coli* Fis protein stimulates bacteriophage lambda integrative recombination in vitro. *J. Bacteriol*. **185:**3076-3080.

Francia, M.V., Zabala, J.C., de la Cruz, F., and J.M. Garcia-Lobo. (1999) The IntI1 integron integrase preferentially binds single-stranded DNA of the *attC* site. *J Bacteriol*. **181:**6844-6849.

Francia, M.V., de la Cruz, F., and M. Garcia Lobo. (1993) Secondary sites for integration mediated by the Tn*21* integrase. *Mol. Microbiol*. **10:**823-828.

Funnel, E.F., and G.J. Phillips (ed.). (2004) Plasmid Biology. American Society for Microbiology, Washington, D.C.

Gama-Castro, S., Jiménez-Jacinto, V., Peralta-Gil, M., Santos-Zavaleta, A., Peñaloza-Spinola, M.I., Contreras-Moreira, B., Segura-Salazar, J., Muñiz-Rascado, L., Martínez-Flores, I., Salgado, H., Bonavides-Martínez, C., Abreu-Goodger, C., Rodríguez-Penagos, C., Miranda-Ríos, J., Morett, E., Merino, E., Huerta, A.M., Treviño-Quintanilla, L., and J. Collado-Vides. (2008) RegulonDB (version 6.0): gene regulation model of *Escherichia coli K-12* beyond

transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Research.* **36:**D120-4.

Gillings, M., Boucher, Y., Labbate, M., Holmes, A., Krishnan, S., Holley, M., and H.W. Stokes. (2008) The Evolution of class 1 integrons and the rise of antibiotic resistance.

Glinkowska, M., Majka, J., Messer, W., and G. Wegrzyn. (2003) The mechanism of regulation of bacteriophage λ p_R promoter activity by *Escherichia coli* DnaA protein. *J. Biol. Chem.* **278:**22250-22256.

Gonzalez-Zorn, B. Catalan, A., Escudero, J.A., Dominguez, L., Teshager, T., Porrero, C., and M.A. Moreno. (2005) Genetic basis for dissemination of *armA. J. Antimicrob. Chemother*. **56:**583-585.

Gorth, A.C., and M.P. Calos. (2004) Phage Integrases: Biology and Applications. *J. Mol. Biol.* **335:**667-678.

Gottesman, S. (1984) Bacterial regulation: global regulatory networks. *Annu. Rev. Genet*. **18:**415-441.

Gourlay, S. and S. Colloms. (2004) Control of Cre recombination by regulatory elements from the Xer recombination systems. *Mol. Microbiol.*. **52:**53-65.

Grindley, N.D.F., K.L. Whiteson, and P.A. Rice. (2006) Mechanisms of site-specific recombination. *Annu. Rev. Biochem*. **75:** 567-605.

Grindley, N.D.F., Lauth, M.R., Wells, R.G., Wityk, R.J., Salvo, J.J., and R.R. Reed. (1982) Transposon-mediated site-specific recombination: identification of three binding sites for resolvase at the res sites of gamma delta and Tn*3*. *Cell.* **30**:19-27.

Groth, A.C., Olivares, E.C., Thyagarajan, B., and M.P. Calos. (2000) A phage integrase directs efficient site-specific integration in human cells. *Proc. Natl. Acad. Sci*. **97:**5995-6000.

Guerin, E., Cambray, G., Sanchez-Alberola, N., Campoy, S., Erill, I., Da Re, S., Gonzalez-Zorn, B., Barbé, J., Ploy, M., and D. Mazel. (2009) The SOS response controls integron recombination. *Science*. **324:**1034.

Hallet, B., and D.J. Sherratt. (1997) Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol. Rev.* **21:**157-178.

Hanau-Bercot, B., Podglajen, I., Casin, I., and E. Collatz. (2002) An intrinsic control element for translational initiation in class 1 integrons. *Mol. Microbiol*. **44:**119-130.

Hacker, J., and J.B. Kaper. (2000) Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol.* **54:**641-79

Hall, R.M., and H.W. Stokes (1993) Novel DNA elements which capture genes by site-specific recombination. *Genetica* **90**: 115–132.

Hall, R.M., Brookes, D.E., and H.W. Stokes. (1991) Site-specific insertion of genes into integrons: role of the 59-base element and determination of the recombination crossover point. *Mol. Microbiol*. **5:**1941-1959.

Hanau-Bercot, B., Podglajen, I., Casin, I., and E. Collatz. (2002) An intrinsic control element for translational initiation in class 1 integrons. *Mol. Microbiol*. **44:**119-130.

Hansson, K., Sundstrom, L., Pelletier, A., and P.H. Roy. (2002) IntI2 integron integrase in Tn*7. J. Bac*. **184:**1712-1721.

Hansson, K., Skold, O., and L. Sundstrom. (1997) Non-palindromic *attI* sites of integrons are capable of site-specific recombination with one another and with secondary targets. *Mol. Microbiol.* **26:**441-453.

Hayes, F. and D.J. Sherratt. (1997) Recombinase binding specificity at the chromosome dimer resolution site dif of Escherichia coli. *J. Mol. Biol.* **266:**525-537.

Hendricks, E.C., H., Szerlong, T., Hill, and P. Kuempel. (2000) Cell division, guillotining of dimer chromosomes and SOS induction in resolution mutants (*dif*, *xerC* and *xerD*) of *Escherichia coli*. *Mol Microbiol* **36**: 973–981.

Hochman, L., Seger, N., Sternberg, N., and G. Cohen. (1983) Site-specific recombinational circularization of bacteriophage P1 DNA *Virology*. **131:**11-17.

Hoess, R.H. and K. Abremski. (1984) Interaction the bacteriophage P1 recombinase Cre with the recombining site *loxP*. *Proc. Natl. Acad. Sci*. **81:**1026-1029.

Huber H.E., Iida, S., Arber, W., and T.A. Bickle. (1985) Site-specific DNA inversion is enhanced by a DNA sequence element in cis. *Proc. Natl. Acad. Sci.* **82:**3776-3780.

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

Johnson, R.C., Bruist, M.F., and M.I. Simon. (1986) Host protein requirements for in vitro site-specific DNA inversion. *Cell.* **46:**531-539.

Kersulyte, D., Mukhopadhyay, A.K., Shirai, M. Nakazawa, T., and D.E. Berg. (2000) Functional organization and insertion specificity of IS607, a chimeric element of *Helicobacter pylori*. *J. Bacteriol*. **182:**5300-5308.

Klemm, P. (1969) Two regulatory *fim* genes, *fimB* and *fimE*, control the phase variation of type 1 fimbriae in *Escherichia coli*. *EMBO. J.* **5:**1389-1393.

Lagatolla, C., Edalucci, E., Dolzani, L, Riccio, M.L., De Luca, F., Medessi, E., Rossolini, G.M., and E.A. Tonin. (2006) Molecular evolution of metallo-β-lactamase-producing

Pseudomonas aeruginosa in a nosocomial setting of high-level endemicity. *J. Clin. Microbiol.* **44:**2348-2353

Landy, A. and W. Ross. (1977) Viral integration and excision: structure of the lambda at sites. *Science* **197:**1147-1160.

Landy, A. (1989) Dynamic, structural, and regulatory complexity in the site-specific recombination. *Annu. Rev. Biochem*. **58:**913-949.

Lévesque, C., Brassard, S., Lapointe, J., and P.H. Roy. (1994) Diversity and relative strengths of tandem promoters for the antibiotic-resistance genes of several integrons. *Gene.* **142:**59-54.

Liebert, C.A., Hall, R.M., and A.O. Summers. (1999) Transposon Tn*21*, flagship of the floating genome. *Microbiol Mol Biol Rev*. **63:**507-22.

Lyras, D., Adams, V., Lucet, I.S., and J.I. Rood. (2004) The large resolvase TnpX is the only transposon-encoded protein required for transposition of the Tn*4451/3*-family of integrative mobilizable elements. *Mol. Microbiol*. **51:**1787-1800.

MacDonald, A.I., Y. Lu, E.A. Kilbride, A. Akopian, and S.D. Colloms. (2008) PepA and ArgR do not regulate Cre recombination at the bacteriophage P1 *loxP* site. *Plasmid* **59:**119-126.

MacDonald, D., Demarre, G., Bouvier, M., Mazel, D., and D.N. Gopaul. (2006) Structural basis for broad DNA-specificity in integron recombination. *Nature*. **440:**1157-1162.

Marquez, C., Labbate, M., Ingold, A.J., Chowdhury, P.R., Ramirez, M.S., Centron, D., Borthagaray, G., and H.W. Stokes. (2008) Recovery of a functional class 2 integron from an *Escherichia coli* strain mediating a urinary tract infection. *Antimicrobial Agents and Chemotherapy.* **52:**4153-4154

Martinez-Antonio, A. and J. Collado-Vides. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Current Opinions in Microbiology*. **6:**482-489.

Martinez, E. and F. de la Cruz. (1988) Transposon Tn*21* encodes a RecA-indendent site-specific integration system. *Mol. Gen. Genet*. **211:**320-325.

Mazel, D. (2006) Integrons: agents of bacterial evolution. *Nature Rev. Microbiol*. **4:**608-620.

Mazel, D., Dychino, B., Webb, V.A., and J. Davies. (1998) A distinctive class of integron in the *Vibrio cholerae* genome. *Science.* **280:**605-608.

Merickel, S.K., Haykinson, M.J., and R.C. Johnson. (1998) Communication between Hin recombinases and Fis regulatory subunits during coordinate activation of Hin-catalyzed site-specific DNA inversion. *Genes Dev.* **12:**2803-2816.

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

Münch, R., Hiller, K., Barg, H., Heldt ,D., Linz, S., Wingender, E. and D. Jahn. (2003) PRODORIC: prokaryotic database of gene regulation. *Nucleic Acids Res.* **31**, 266-269.

Naas, T., Mikami, Y., Imai, T., Poirel, L. and Nordmann, P. (2001) Characterization of In53, a class 1 plasmid- and composite transposon-located integron of Escherichia coli which carries an unusual array of gene cassettes. *J. Bacteriol*. **183:**235-249.

Nandi, S., Maurer, J.J., Hofacre, C., and A.O. Summers. (2004) Gram-positive bacteria are a major reservoir of class 1 antibiotic resistance integrons in poultry litter. *Proc. Natl. Acad. Sci.* **101:**7118-7122.

Nash, H.A. and C.A. Robertson. (1981) Purification and properties of the Escherichia coli protein factor required for λ integrative recombination. *J. Biol. Chem*. **256:**9246-9253.

Norris, S.J. (2006) Antigenic variation with a twist – the *Borrelia* story. *Mol. Microbiol.* **60:**1319-1322.

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

Ogawa, A. and T. Takeda. (1993) The gene encoding the heat stable enterotoxin of *Vibrio cholerae* is flanked by 123-base pair direct repeats. *Microbiol. Immunol*. **37:**607-616.

Olsen, P.B. and P. Klemm. (1994) Localization of promoters in the fim gene cluster and the effect of H-NS on the transcription of fimB and fimE. *FEMS Microbiol Lett.* **116:**95-100.

Papagiannis, C.V., Sam, M.D., Abbani, M.A., Yoo, D., Casico, D., Clubb, R.T., and R.C. Johnson. (2007) Fis targets assembly of the Xis nucleoprotein filament to promote excisive recombination by phage lambda. *J. Mol. Biol*. **367:**328-343.

Papagiannitsis, C.C, Tzouvelekis, L.S., and V. Miriagou. (2009) Relative strengths of the Class 1 Integron promoter hybrid 2 and the combinations of strong and hybrid 1 with an active P2 promoter. *Antimicrob. Agents Chemother*. **53:**277-280.

Partridge, S.R., Recchia, G.D., Scaramuzzi, C., Collis, C.M., Stokes, H.W., and R.M. Hall. (2000) Definition of the attI1 site of class 1 integrons. *Microbiology*. **146:**2855-2864.

Paul, S., and D. Summers. (2004) ArgR and PepA, accessory proteins for XerCD-mediated resolution of ColE1 dimers, are also required for stable maintenance of the P1 prophage. *Plasmid*. **52:** 63-68.

Perez-Rueda, E., and J. Collado-Vides. (2000) The repertoire of DNA-binding transcriptional regulators in *Escherichia coli K-12*. *Nucleic Acids Research.* **28:** 1838-1847.

Plunkett, G. 3rd and H. Echols. (1989). Retroregulation of the bacteriophage lambda int gene: limited secondary degradation of the RNase III-processed transcript. *J. Bacteriol*. **171:**588-92.

Pollack, J.R., and V.R. Lyer. (2002) Characterizing the physical genome. *Nat. Genet*. **32:515-**521.

Pournaras, S., Ikonomidis, A., Tzouvelekis, L.S., Tokatlidou, D., Spanakis, N., Maniatis, A.N., Legakis, N.J., and A. Tsakris. (2005) VIM-12, a novel plasmid-mediated metallo-β-lactamase from *Klebsiella pneumoniae* that resembles a VIM-1/VIM-2 hybrid. *Antimicrob. Agents Chemother*. **49:**5153-5156.

Recchia, G.D., and D.J. Sherratt. (1999) Conservation of Xer site-specific recombination genes in bacteria. *Mol. Microbiol*. **34:** 1146-1148.

Recchia, G.D., and R.M. Hall. (1995) Gene cassettes: a new class of mobile elements. *Microbiology.* **141**:3015-3027.

Recchia, G.D., Stokes, H.W., and R.M. Hall. (1994) Characterization of specific and secondary recombination sites recognized by the integron DNA integrase. *Nucleic Acids Res.* **22:**2071-2078.

Reichardt, L.F. (1975) Control of bacteriophage lambda repressor synthesis after phage infection: the role of the N, cII, cIII and cro products. *J. Mol. Biol*. **93:**267-288.

Reijns, M., Lu, Y., Leach, S., and S.D. Colloms. (2005) Mutagenesis of PepA suggests a new model for the Xer/cer synaptic complex. *Mol. Microbiol.* **57:**927-41.

Rice, P.A., Yang, S.W., Mizuuchi, K., and H.A. Nash. (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell.* **87:**1295-1306.

Richardson, J. P. & J.L. Greenblatt. (1996). Control of RNA chain elongation and termination. *Escherichia coli and Salmonella: Cellular and Molecular Biology*. F. Neidhardt (ed.). pp 822–848. American Society for Microbiology. Washington, DC.

Roberts, J. and J.S. Parks. (2004) Mfd, the bacterial transcription repair coupling factor: translocation, repair, and termination. *Curr. Opin. Microbiol*. **7:**120-125.

Rowe-Magnus, D.A., Guerout, A.M., Biskri, L., Bouige, P., and D. Mazel. (2003) Comparative analysis of superintegrons: engineering extensive genetic diversity in the vibrionaceae. *Genome Res.* **13:**428-442.

Rowe-Magnus, D.A., Guerout, A.M., and D. Mazel. (1999) Super-integrons. *Res. Microbiol.* **150:**641-651.

Rowley, P.A., Smith, M.C., Younger, E., and M.C. Smith. (2008) A motif in the C-terminal domain of phiC31 integrase controls the directionality of recombination. *Nucleic Acids Research*. **36:**3879-3891.

Shapiro, J. A. (1995) The discovery and significance of mobile genetic elements. In Mobile Genetic Elements (Sherrat, D. J., ed.). IRL Press, Oxford, 1–17.

Shearer, J.E.S and A. O. Summers. (2009) Intracellular steady-state concentration of integron recombination products varies with integrase level and growth phase. *J Mol Biol.* **386:**316-331.

Shearer, J.E.S. (2007) *In vivo* integron integrase (IntI1) abundance affects the type and amount of recombination products differentially during the growth cycle. Ph.D. Thesis. University of Georgia.

Shearwin, K.E., Callen, B.P., and J.B. Egan. (2005) Transcriptional interference- a crash course. *Trends in Genetics*. **21:**340-345.

Smith, A., and R.J. Siebeling. (2003) Identification of genetic loci required for capsular expression in *Vibrio vulnificus*. *Infect. Immun*. **71:**1091-1097.

Smith, M.C.M., and H.M. Thorpe. (2002) Diversity in the serine recombinases. *Mol. Microbiol.* **44:**299-307.

Sneppen, K., Dodd, I.B., Shearwin, K.E., Palmer, A.C., Schubert, R.A., Callen, B.P., and J.B. Egan. (2005) A Mathematical Model for Transcriptional Interference by RNA Polymerase Traffic in *Escherichia coli*. *J. Mol. Biol*. **346:**399-409.

Spiers A.J., and D.J. Sherratt. (1999) C-terminal interactions between the XerC and XerD site-specific recombinases. *Mol Microbiol.* **32:**1031-42.

Steiner, W., Liu, G., Donahie, W.D., and P. Kuempel. (1999) The cytoplasmic domain of the FtsK protein is required for resolution of chromosome dimers. *Mol. Microbiol.* **27:**257-268.

Sternberg, N., B. Sauer, R. Hoess, and K. Abremski. (1986) Bacteriophage P1 cre gene and its regulatory region: evidence for multiple promoters and for regulation by DNA methylation. *J. Mol. Biol*. **187:** 197-212.

Stivers, J.T., G.J. Jagadeesh, B. Nawrot, W.J. Stec, and S. Shuman. (2000) *Biochemistry* **39:**5561-72.

Stokes, H.W., Nesbo, C., Holley, M. Bahl, M., Gillings, M., and Y. Boucher. (2006) Class 1 integrons predating the association with Tn402-like transposition genes are present in a sediment microbial community. *J. Bacteriol*. **188:**5722-5730.

Stokes, H.W., O'Gorman, D.B., Recchia, G.D., Parsekhian, M., and R.M. Hall. (1997) Structure and function of 59-base element recombination sites associated with mobile gene cassettes. *Mol. Microbiol*. **26:**731-745.

Stokes, H.W. and R.M. Hall. (1991) Sequence analysis of the inducible chloramphicol resistance determinant in the Tn1696 integron suggests regulation by translational attenuation. *Plasmid.* **26:**10-19.

Sun, X., Mierke, D.F., Biswas, T., Lee, S.Y., Landy, A., and M. Radman-Livaja. (2006) Architecture of the 99 bp DNA-six-protein regulatory complex of the lambda att site. *Mol. Cell.* **24:**569-580.

Sundstrom. L., Roy, P.H., and O. Skold. (1991) Site-specific insertion of three structural gene cassettes in transposon Tn*7. J. Bacteriol*. **173:**3025-3028.

Tecklenburg, M., Naumer, A., Nagappan, O., and P. Kuempel. (1995) The *dif* resolvase locus of the *Escherichia coli* chromosome can be replaced by a 33-bp sequence, but function depends on location. *Proc. Natl. Acad. Sci*. **92:**1352-1356.

Thompson, J.F., Moitoso de Vargas, L., Koch, C., Kahmann, R., and A. Landy. (1987). Cellular factors couple recombination with growth phase: characterization of a new component in the λ site-specific recombination pathway. *Cell.* **50:**901-908.

Thorpe, H.M. and M.C.M. Smith. (1998) *In vitro* site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc. Natl. Acad. Sci.* **95:**5505-5510.

Wisehart, W. L., Machida, C., Ohtsubo, H., and E. Ohtsubo. (1983). *E. coli* RNA polymerase binding sites and transcription initiation sites in the transposon Tn*3*. *Gene* **24:**99-113.

Wohlleben, W., Arnold, W., Bissonnette, L., Pelletier, A., Tanguay, A., and P.H. Roy. (1989) On the evolution of Tn*21*-like multiresistance transposons; sequence analysis of the gene

(*aacC1*) for gentamicin acetyltransferase 3-1 (ACC(3)-1), another member of the Tn*21*-based expression cassette. *Mol. Gen. Genet.* **217:**202-208.

Xu, H., Davies, J., and V. Miano. (2007) Molecular characterization of class 3 integrons from *Delftia* spp. *J. Bacteriol*. **189:**6276-6283.

Yildiz, F.H., Liu, X.S., Heydorn, A., and G.K. Schoolnik. (2004) Molecular analysis of rugosity in a *Vibrio cholerae* O1 El Tor phase variant. *Mol. Microbiol*. **53:**497-515.

Yu, X., Weihe, C.E., and W. Margolin. (1998) Role of the C terminus of FtsK in *Escherichia coli chromosome* segregation. *J. Bacteriol*. **180:**6424-6428.

Zieg, J., and M. Simon. (1980) Analysis of the nucleotides sequence of an invertible controlling element. *Proc. Natl. Acad. Sci.* **77:**4196-4200.

Zieg, J., Silverman, M., Hilmen, M., and M. Simon. (1977) Recombinational switch for gene expression. *Science*. **196:**170-172.

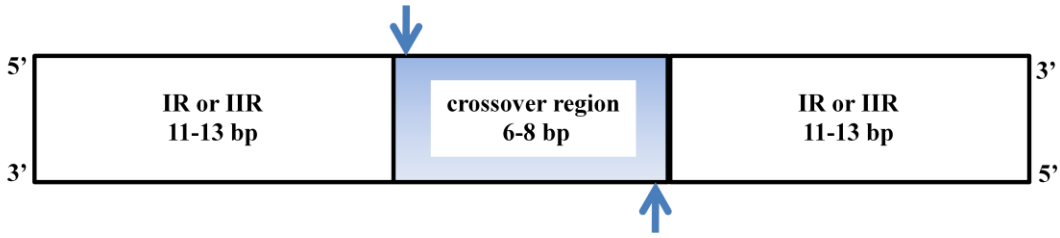


Figure 1.1 Typical core site structure

White rectangles represent recombinase binding sites defined through their symmetrical inverted repeats (IR) or imperfect inverted repeats (IIR). The blue rectangle is the shorter central crossover region where recombinase mediated cleavage occurs. Cleavage points are indicated with blue arrows. Figure modified from reference (Craig *et al*., 2002).

Table 1.1 *Escherichia coli* promoter pair arrangements

4,462 Known or Predicted Promoter Pairs[a]		
Arrangement	**Non-Overlapping Promoter Pairs[b]**	**Overlapping Promoter Pairs[c]**
Tandem	166	292
Convergent	54	54
Divergent	ND	89

[a] Data compiled in RegulonDB database (Gama-Castro et al., 2008); analyzed by (Shearwin *et al*., 2005).
[b] Non-overlapping pairs defined by start sites greater than 40-bp and less than 200-bp apart, ND means not determined.
[c] Overlapping pairs defined by start sites less than 40-bp apart.

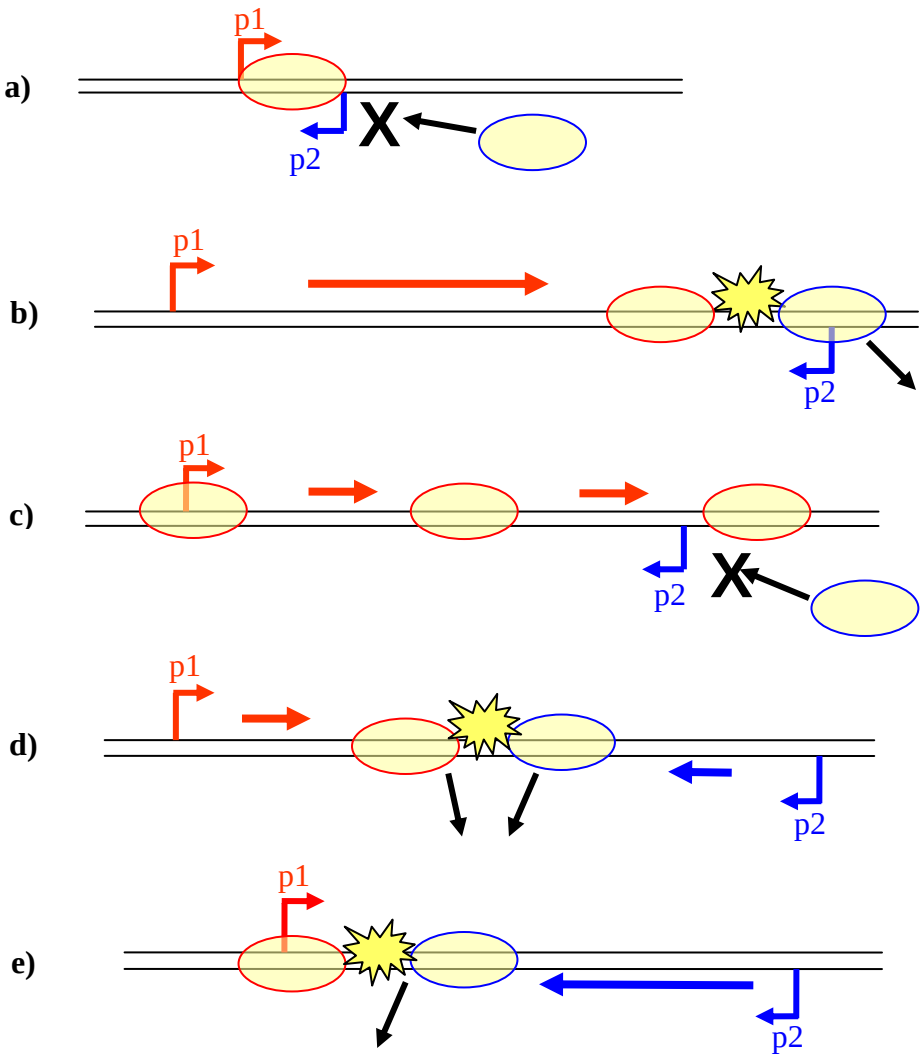

a)
p1
p2
X
b)
p1
p2
c)
p1
p2
X
d)
p1
p2
e)
p1
p2

Figure 1.2 Forms of transcriptional interference

The five currently defined forms of transcriptional interference are shown. Yellow ovals represent RNA polymerase, and the red or blue lines correspond to cognate promoters (angle arrows) and direction of transcription (straight arrows). Black arrows represent removal of RNA polymerase. Black "X" indicates RNA polymerase inability to load. Yellow stars indicate RNA polymerase impact. Promoter interference can occur with **(a)** promoters arranged tandemly, convergently, or in overlapping divergent pairs **(b)** sitting duck and **(c)** occlusion occurs with tandem or convergent promoter pairs. **(d)** Collision occurs only in convergent pairs. **(e)** Roadblock could theoretically occur in tandem or convergent promoter pairs. Figure modified from reference (Shearwin *et al*., 2005).

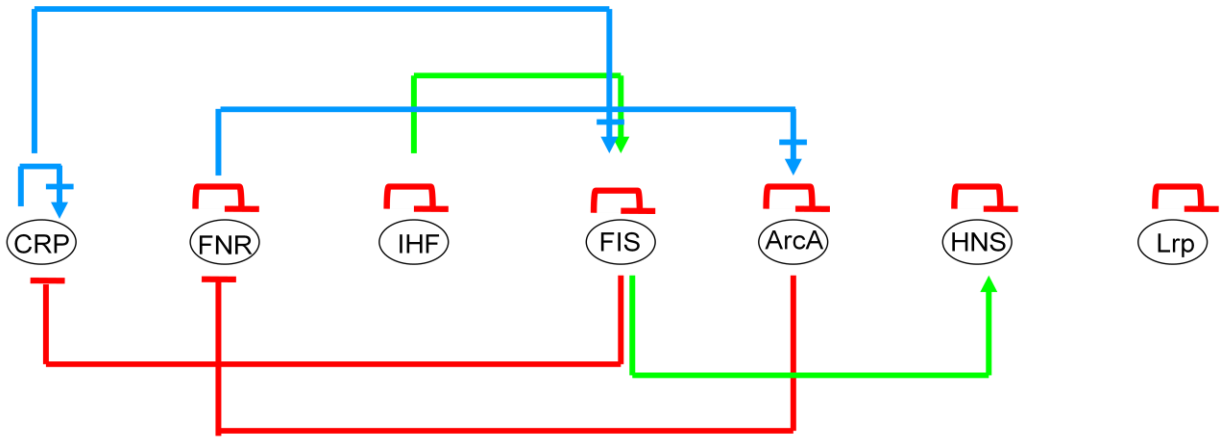


Figure 1.3 Global regulatory network interactions between the seven global regulators in *Escherichia coli.*

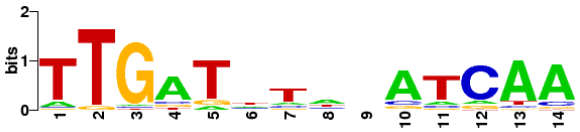
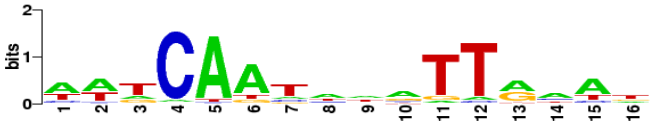
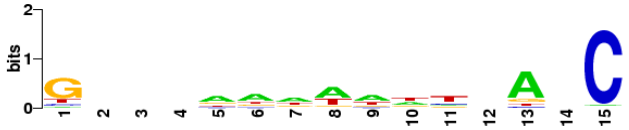
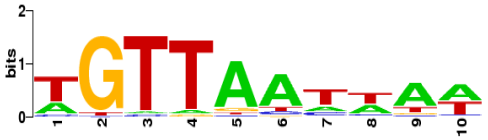
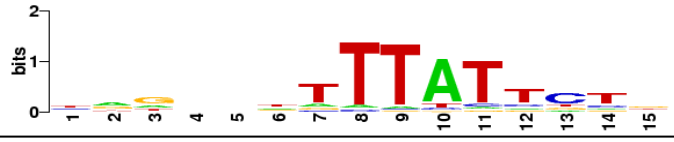
Regulatory actions are indicated by red lines (repressor), green arrows (activator), or blue arrows with line (dual function acting as both, repressor or activator on the same system as the situation dictates). All global regulators negatively self-regulate except ArcA, which is activated and repressed by FNR. All data compiled from http://regulondb.ccg.unam.mx/ (Gama-Castro *et al*., 2008).

Table 1.2 Network interactions among global regulators

Global regulator [a]	Regulated by[b]	Number of genes regulated[c]	Functional distribution[d]	Number of co-regulators [e]
CRP	CRP CRP, FIS	414	79 11 10	101
FNR	ArcA, FNR	282	63 33 4	54
IHF	IHF	217	60 39 1	60
FIS	CRP, IHF CRP, FIS	211	66 34 <1	49
ArcA	FNR, FNR	155	28 65 7	43
Hns	CspA, FIS, GadX, HNS	144	29 71 <1	47
Lrp	GadE, Lrp	93	35 65 0	32

[a] Global regulators according to Martinez-Antonio and Collado-Vides, 2003.
[b] Activators (green) and repressors (red) are listed for each global regulator.
[d] Fraction of each global regulator that acts as an activator (green), repressor (red), or dual function (blue) acting as both, activator or repressor within the same system.
[b-e] All data compiled from http://regulondb.ccg.unam.mx/, (Gama-Castro *et al*., 2008).

Table. 1.3 Global regulator recognition sequence weight matrices[a]

CRP	
FNR	
IHF	
FIS	
ArcA	
HNS	**No conserved sequence**
Lrp	

[a] Weight matrices from http://www.prodoric.de/, (Münch *et al*., 2003).

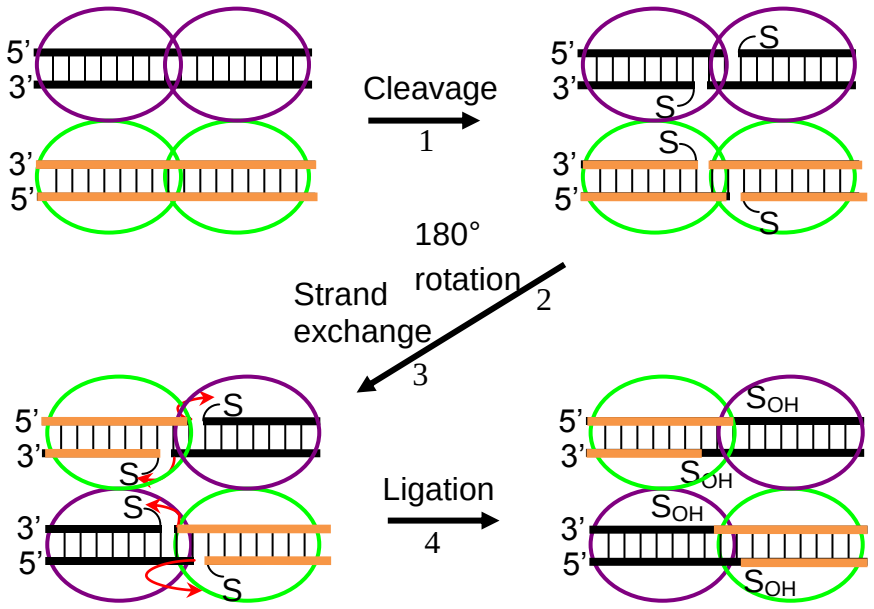


Figure 1.4 Serine recombination mechanism

Circles represent catalytic recombinase subunits. **(1)** The serine nucleophile of the recombinase subunits cleave all four DNA strands creating double stranded breaks with 2-bp overhangs and then forming protein-DNA covalent linkages with the 5' ends, while leaving 3' hydroxyls unreacted (free). **(2)** The complex rotates 180º prior to strand exchange, breaking and remaking four H-bonds in the process. **(3 and 4)** The 3' hydroxyls then attack the phosphoserine linkages to ligate the DNA. Figure modified from reference (Shearer, 2007)

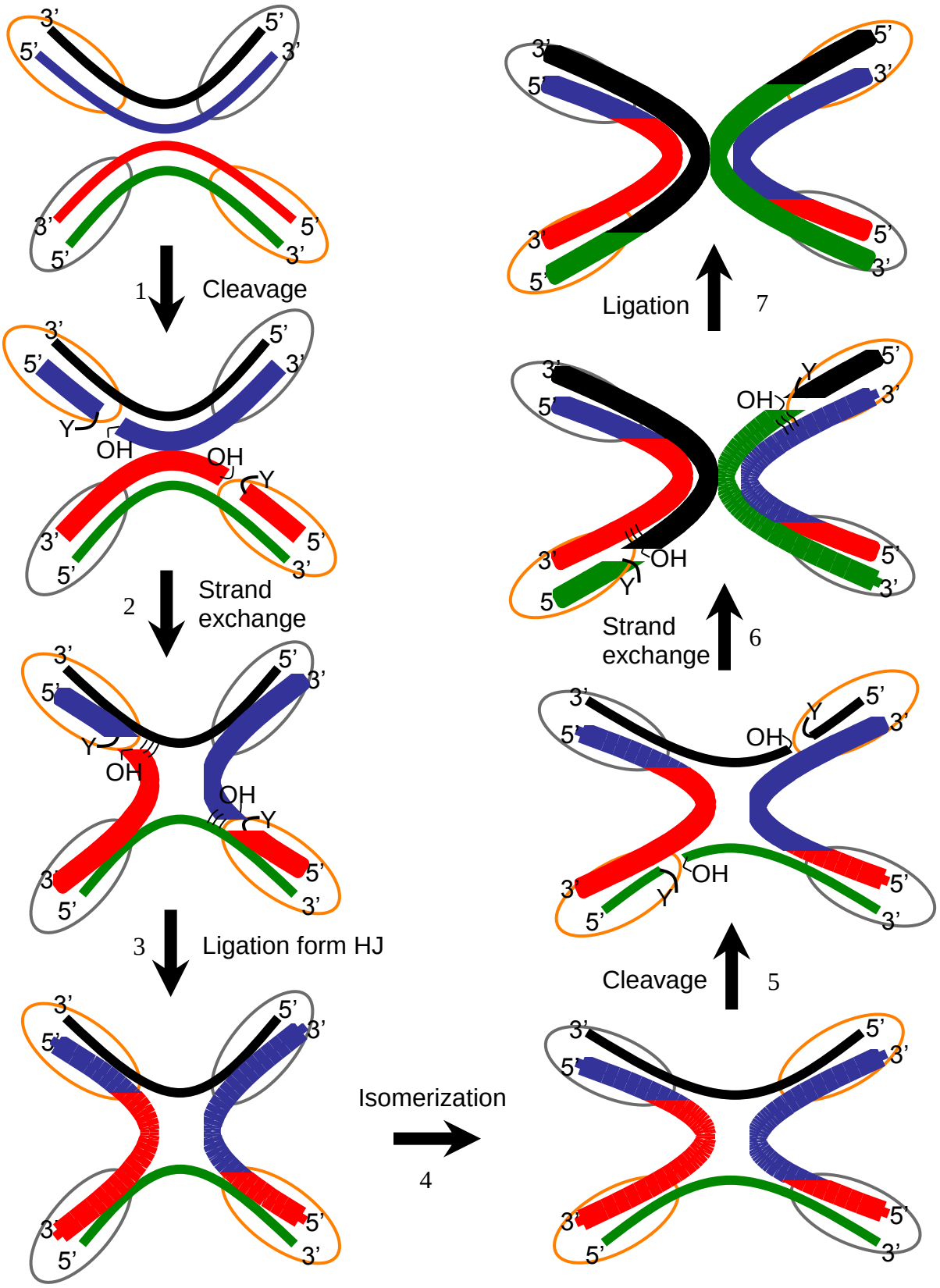

3’
5’
5’
3’
3’
5’
5’
3’
1
Cleavage
Y
OH
OH
Y
2
Strand
exchange
3
Ligation form HJ
Isomerization
4
5
Cleavage
6
Strand
exchange
7
Ligation

Figure 1.5 Tyrosine recombination mechanism

Ovals represent recombinase subunits bound to each half site; the active subunits are shown in orange and inactive subunits are gray. Thick lines indicate DNA strands being modified or already modified by recombinase activity. **(2)** The conserved tyrosine nucelophiles in the active pair cleave one DNA strand forming covalent phosphotyrosine links with the 3' end and free hydroxyls at the 5' end of the DNA in each core site. **(3)** The free 5' hydroxyls in each DNA core site attack the phosphotyrosyl bond of the opposite core site ligating to form the recombination intermediate known as a Holliday junction (HJ). **(4)** Holliday junction resolution occurs after complex isomerization activates the second pair of bound recombinases and inactivates the first pair enabling **(5)** cleavage, **(6)** stand exchange, and **(7)** ligation of the other DNA strands in order to complete the reaction. Figure from reference (Shearer, 2007) with permission.

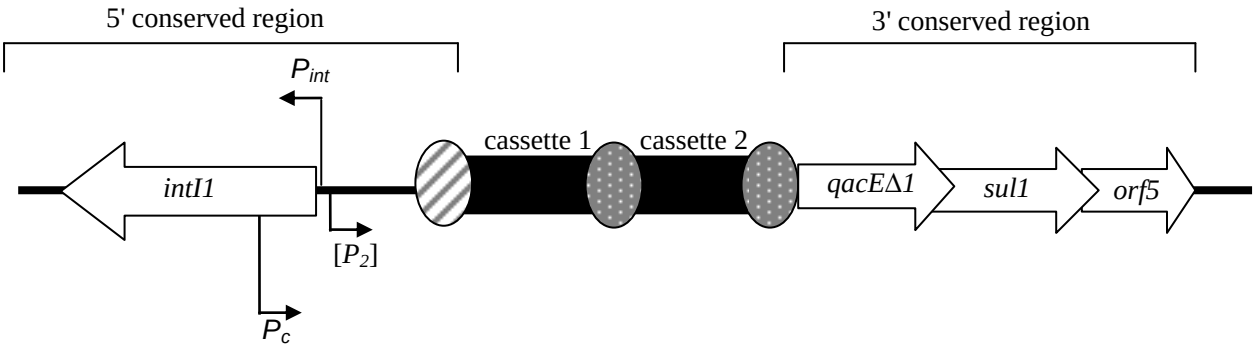


Figure 1.6 Class 1 integron structure

The general structure of a class one integron. Conserved genes *intI*, *qacEΔ1*, *sul1*, and *orf5* are shown in white block arrows indicating direction of transcription, and the variable region containing inserted gene cassettes are shown in black. Integrase recognition sites *attI* , is where new cassettes are added using the *attC* , carried with each gene cassette. Integron promoters P_c and P_2 transcribe inserted cassettes, but P_2 is bracketed to indicate it is not present in all instances of the class 1 integron. P_2 is a polymorphism of class 1 integrons that arises from a 3-bp insertion between potential RNAP recognition hexamers changing their non-ideal 14-bp spacing to the ideal 17-bp spacing. P_{int} transcribes *intI*. Figure modified from reference (Liebert *et al*., 1999)

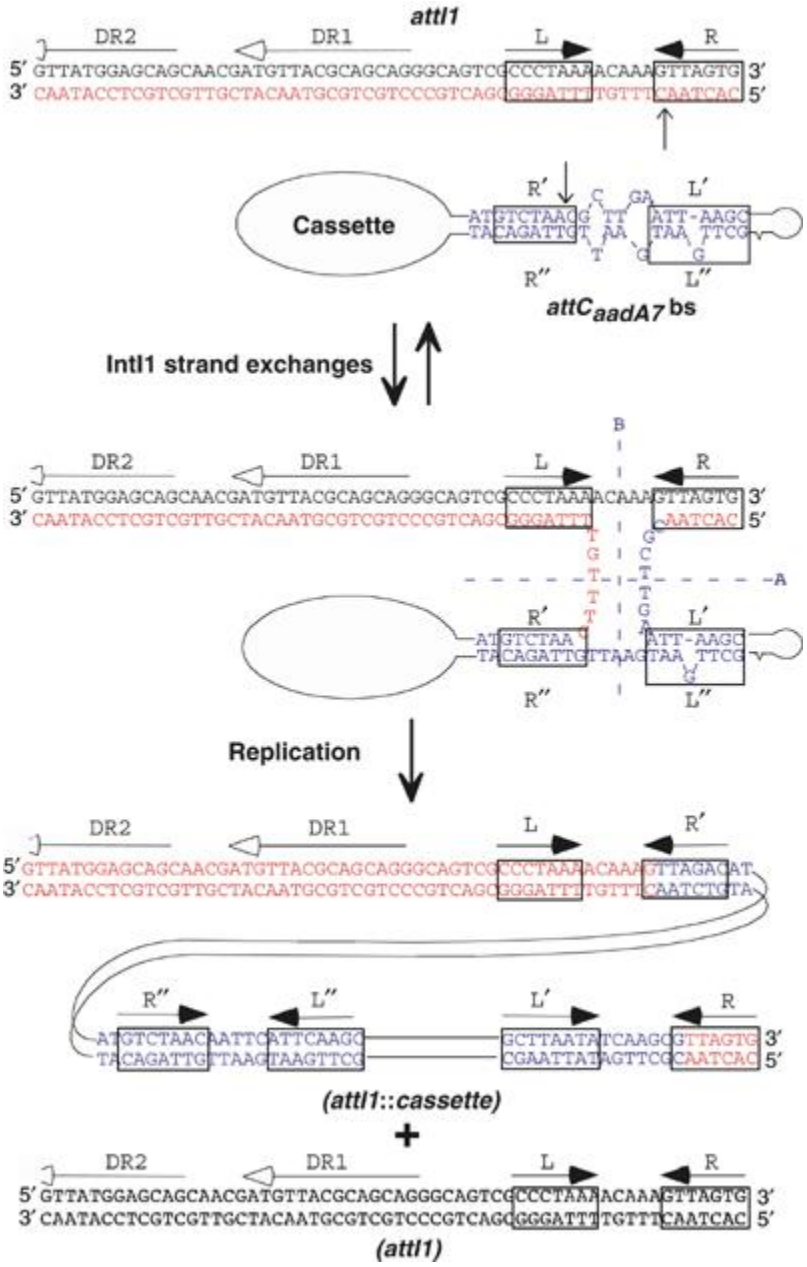

attI1
DR2
DR1
L
R
5′ GTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAACAAAGTTAGTG 3′
3′ CAATACCTCGTCGTTGCTACAATGCGTCGTCCCGTCAGCGGGATTTTGTTTCAATCAC 5′
Cassette
R′
L′
R″
L″
attC aadA7 bs
IntI1 strand exchanges
B
A
Replication
R″
L″
L′
R
(attI1::cassette)
+
(attI1)

Figure 1.7 Integron recombination mechanism model

Recombination between the aadA7 bottom strand ss-*attC* (blue) folded upon itself pairing its imperfect inverted repeats of the core sites (R' and L') and (R'' and L'') with ds-*attI* (red). Putative IntI1 binding domains and crossover positions are indicated by boxes and vertical arrows, respectively. Cleavage and strand exchange use prototypical tyrosine site-specific recombination, but differs in HJ resolution. Typical resolution through the *A* axis reverses the recombination resulting in the original substrates. Resolution through the *B* axis is unsuccessful, resulting in covalently linear molecules. The model predicts successful recombination uses a replication step. Figure from reference (Bouvier *et al*., 2005) with permission.

Table 1.4 Integron cassette promoter strengths

Promoter Version	-35 hexamer	-10 hexamer	Strength[a]
Strong	TTGACA	TAAACT	6.5
Strong + P_2			7.8
Weak	TGGACA	TAAGCT	0.2
Weak + P_2			3.2
Hybrid 1	TGGACA	TAAACT	0.7
Hybrid 1 + P_2			2.5
Hybrid 2	TTGACA	TAAGCT	3.2
P_2	TTGTTA	TACAGT	

[a] The strength of each promoter version has been normalized to derepressed P_{tac} (Lévesque *et al*., 1994; Papagiannitis *et al*., 2009).

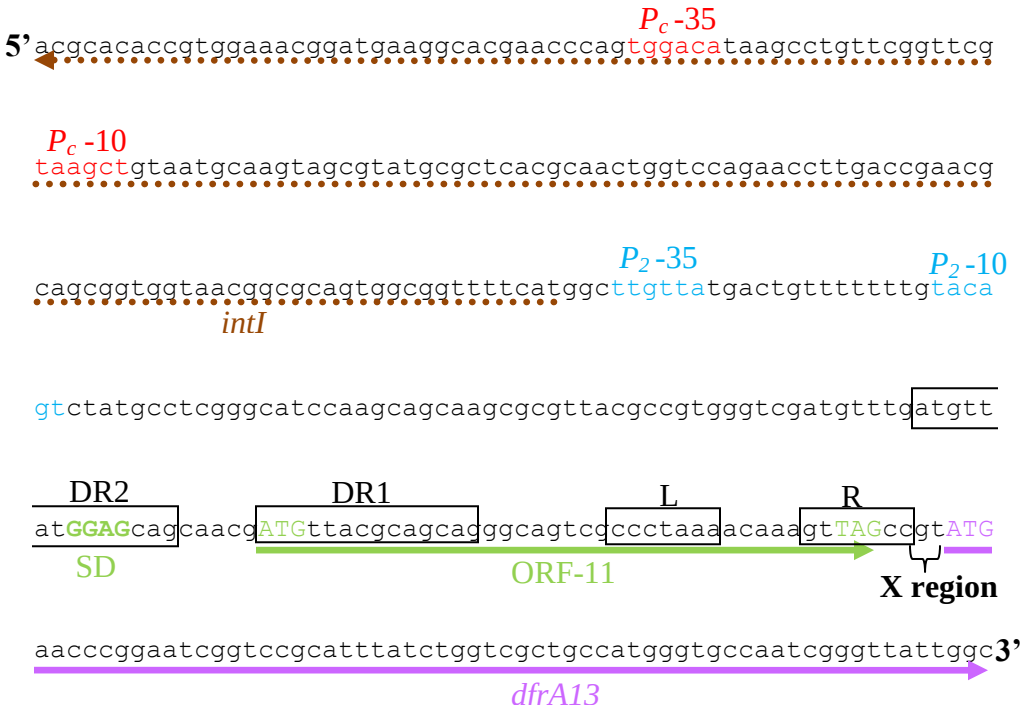


Figure 1.8 ORF-11 position in the class 1 integron

Sequence from accession number Z50802 (Adrian *et al*., 2000). Integron cassette promoters P_c and P_2 are shown in red and blue, respectively. The *attI* site boundaries and IntI binding sites are boxed and labeled direct repeat 1 (DR1), direct repeat 2 (DR2), left (L), and right (R) core site. Dotted brown arrow indicates *intI* translation of the opposite strand. Solid green and purple arrows show ORF-11 and gene cassette *dfrA13* translation, respectively. Start codons of each and the stop codon of ORF-11 are indicated in capitals with corresponding colors. ORF-11's putative Shine-Dalgarno (SD) indicated in green capitals. The variable X region (depending on cassette), is bracketed (Hanau-Bercot *et al*., 2002)

CHAPTER 2

INVESTIGATING CLASS 1 INTEGRON RECOMBINATION CONTROL VIA GROWTH PHASE, ACCESSORY PROTEINS, AND SITE COMPETITION

OVERVIEW

The integron integrase IntI1 mediates the insertion and excision of gene cassettes. Although *in vitro* this activity requires only the integrase for recombination (Martinez and de la Cruz, 1988), its dual site specificity involving two distinctly different recognition sites, *attI* and *attC*, implys that an additional unidentified factor(s) may dictate recombination *in vivo*. I investigated three possible levels of recombination control: 1) accessory protein involvement with a pull down assay consisting of whole cell lysate containing overexpressed integrase complexed with biotinylated *attI* site; 2) site competition and/or interaction directed by specific IntI conformations with competitive binding assays between *attC* and *attI* with purified integrase; and 3) growth phase dependency through qualitative PCR of recombinant junctions between donor and recipient plasmids. Neither accessory protein involvement nor growth phase dependency could be determined within the limits of the techniques used here. However, site competitions/interactions revealed *attI* presence and order of addition does affect IntI-*attC* complex formation, suggesting site specific IntI allostery may play a role in recombination frequency or orientation.

INTRODUCTION

Integrons are mobile genetic elements in the tyrosine site-specific recombinase family most recognized for their role in the dissemination of antibiotic resistance. There are several integron classes defined by sequence comparison ranging from 40-59% identity (Collis *et al.*, 2002). The class 1 integron is the best studied and the current paradigm for integron function, containing a variable region flanked by conserved 5' and 3' regions. Unlike the conjugative plasmids and transposons in which they are typically found, integrons themselves are not mobile. Instead, the variable region contains mobile, non-self-replicating elements referred to as cassettes which are open reading frames (ORF) immediately followed by an integrase-specific recombination site called *attC*. The integron encodes the site-specific integrase (IntI) which inserts or excises gene cassettes in the form of covalently closed circular intermediates or through co-integrate formation and resolution (Stokes and Hall, 1989; Martinez and de la Cruz, 1990; Collis and Hall, 1992; Shearer and Summers, 2009).

The 5' conserved region flanking the cassettes encodes the integrase gene (*intI*), and an adjacent recombination site called *attI* where new cassettes are typically inserted (Recchia *et al*., 1994). This 5' conserved region also has at least two convergent promoters; P_{int} driving transcription of *intI*, and P_c which transcribes inserted cassettes. There also can be a rare secondary cassette promoter P_2 due to a 3-bp addition creating a consensus 17-bp spacing between adventitiously placed -35 and -10 RNAP recognition hexamers. The 3' conserved region on the opposite side of the inserted cassettes encodes a quaternary ammonium compound resistance gene (*qacE*Δ1), a sulfonamide resistance gene (*sulI)*, and an open reading frame of unknown function (ORF 5) (Hall and Stokes, 1993).

The *attI* site in the 5' conserved region is approximately 65-bp long, and has a core site composed of two imperfect inverted repeats (L and R), RYYYAAC and GTTRRRY, (where R is a purine and Y is a pyrimidine), separated by a 6 to 8-bp spacer of variable sequence in different class 1 integrons. In addition to the core site, the *attI* site contains two imperfect direct repeats (DR1 and DR2) located upstream of the core site, not found in the *attC* site. The *attC* site at the 3' end of each cassette varies more in sequence and in length from 57 to 141-bp (Recchia and Hall, 1995). However, *attC* sites do have some common features including two possible core sites (L' and R') and (L" and R"), arranged as imperfect inverted repeats. Each has also RYYYAAC and GTTRRRY core recognition sequences, and they are separated by 6 to 8-bp spacers of varied sequence. Single strand cleavage occurs between the G and the first T in the R region of *attC* and the corresponding R' region of *attI* (Hannson, *et al.*,1997; Stokes *et al.*, 1997).

Recombination can occur between any combination of sites; *attI* x *attC*, *attC* x *attC*, or *attI* x *attI*, but *attI* x *attC* is the most efficient, followed by *attC* x *attC*, then *attI* x *attI* (Francia *et al*., 1993; Recchia, *et al*., 1994; Hansson *et al*., 1997). Interestingly, although *attI* x *attC* is preferred, the absolute recombination frequency depends on the cassette, suggesting unidentified factors that modulate complex formation (Collis *et al*., 2001). In addition, *in vitro* experiments revealed IntI1 preferentially binds double stranded *attI*, but only the bottom single strand of *attC* (Francia *et al*., 1999; Johansson *et al*., 2004). This has led to speculation that observed growth phase dependent abundance of recombination products reflects the availability of single stranded substrate arising during DNA replication and transcription (Bouvier *et al*., 2005; Shearer and Summers, 2009). However, the exact mechanism controlling integron recombination at any level remains unknown. In this study, I investigated possible host-encoded accessory protein

involvement, integrase direction of site recognition or complex recruitment, and growth phase dependency as possible modulators of recombination.

MATERIALS AND METHODS

Bacterial strains, plasmids, and culture conditions

Bacterial strains, plasmids, and primers used in this study are listed in Table 2.1.

Integrase Purification

The Int-His_6 protein was expressed as described previously (Johansson *et al*., 2004), but with the following modification. Induced cells were harvested by centrifugation at 4°C, and resuspended in 20 ml of lysis buffer (20 mM [Na_2HPO_4-NaH_2PO_4] pH 7.2, 200 mM NaCl, 1 mM imidazole) supplemented with 0.1mM protease inhibitor phenylmethyl sulfonylflouride (PMSF) and 10% glycerol. Cells were lysed with two passes through the French Press at 16,000 psi. Insoluble material was separated from soluble by centrifugation at 11,000 r.p.m for 40 min at 4ºC and the latter was passed through a Whatman 25 mm GD/X 0.2 um glass fiber filter. The suspension was immediately loaded on a HiTrap-chelating column (Amersham Pharmacia Biotech) charged with Ni^{2+}. Integrase was eluted with a linear gradient of imidazole and NaCl ranging from 1 mM imidazole and 200 mM NaCl to 800 mM imidazole and 40 mM NaCl. Eluted fractions were pooled and loaded on a HiPrep 26/10 desalting column (Amersham Pharmacia Biotech) and eluted with 50 mM Tris pH 7.5, 0.2 mM EDTA, 1 mM DTT buffer. Protein concentrations were determined by using a Bradford protein assay (Bio-Rad). 500 ul aliquots of the purified integrase were stored at –70°C in 10% glycerol.

Accessory Protein Pull Down Assay

Double-stranded DNA amplicand containing the *attI* recognition site of Tn*21* in R100 was prepared using PCR with synthesized primer pair biotinylated attI414U and attI414L (Sigma) to amplify a 414 bp region. Primer pair attI414U and attI137L (Sigma) were used to amplify a shorter 137 bp region containing the same *attI* recognition site. Biotinylated double-stranded attI414 was bound to 25 ul of streptavidin magnetic particles as recommended by the manufacturer (Roche). Densitometry comparison of biotinylated DNA before addition to streptavidin magnetic particles was compared to wash buffer and ran on a 3% metaphor gel for 1 h at 130 V to determine DNA binding efficiency. Accessory protein binding was done with SK1592 (pSU2056) IPTG induced lysate or purified Int-His_6 and based on protocol described previously (Gabrielsen *et al*., 1989), but with modified protein binding buffer to (50 mM Tris, 100 mM KCl, 1 mM EDTA, 5% glycerol, 1 mM DTT, pH 7.5) and wash buffer(s) for non-specific or weakly bound proteins (50 mM Tris, 100-300 mM KCl, 1 mM EDTA, 5% glycerol, 0.5 mg/ml BSA, pH 7.5). Accessory protein elution was done with either wash buffer containing1M KCl elution or heating bead-*attI*-protein mix 95ºC for 5 min. 20ul of all wash and elution steps were run in a pre-run 12% Tris SDS polyacrylamide gel (Bio-Rad) for 1 h at 100 V and stained with Coomassie Blue.

Electrophoretic Mobility Shift Assay

Single-stranded dephosphorylated 62-base oligodeoxynucleotide *aadA7* bottom strand (Sigma) named *attCaadA7bs* (Bouvier *et al*., 2005) was 5' end-labeled with radioactive phosphate using [γ-^{32}P]ATP and polynucleotide kinase. All subsequent binding reactions contained 8 nM labeled *attCaadA7b* and 160 nM purified integrase in binding buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM CHAPS, 0.2 mM EDTA, 5% glycerol 1 mM DTT, 0.7 ug BSA) in 20 ul reaction

(Johannson *et al*., 2004). *Competitive EMSA* was done with increasing concentrations 4-400 nM of unlabeled 137-bp *attI* amplicand. Reactions were incubated 30 min at 25°C.

Order-of-Addition Competitive EMSA

Reagent combinations and incubations periods were as follows: *attI* + *attC* 3 min, plus integrase 25 min; *attC* + integrase 25 min, plus *attI* 3 min; *attI* + integrase 25 min, plus *attC* 3 min. All binding reactions were loaded on a 5% non-denaturing polyacrylamide gel (Bio-Rad) and run at 80 V for 1 h at 25°C in 1× TBE buffer.

Growth Phase Experiments

Integron-carrying test strains CB454(pICV8)(pRMH14) and CB454(pICV8)(pRMH14) (pSU2056) and control strains CB454(pICV8)(no integron), CB454(pRMH14)(pSU2056) (no target), and CB454(pICV8::aadA2) were streaked on LB 25 ug/ml streptomycin plate and grown at 37ºC overnight. Ten colonies from each plate were used to inoculate 25 ml of fresh LB containing no selective antibiotics. Cells were grown at 37ºC shaking 250 r.p.m with 1.5 ml samples taken after 1, 5, 10, 15, 20, 25, 30, 35, 40 h growth. Cells were harvested by centrifugation 10,000 r.p.m at room temperature and resuspended in 1 ml of cell freezing medium then stored in -70ºC. The 5' junction primer pair (aadA2-U1 and 4E-L2) were used in PCR to detect cassette insertion and cointegrate events (Fig. 2.4), and cointegrate junction primer pair (8E-U1and intL1) were used to detect cointegrate formation only (Fig. 2.4). PCR reactions (50 ul) were programmed for 1 min at 95ºC followed by, 40 cycles of 1 min at 95ºC, 1 min at 60ºC, 2 min at 72ºC, and a final elongation step of 5 min at 72ºC.

RESULTS

Accessory protein interaction and attI site recognition.

The *attC* and *attI* sites have no sequence similarity apart from the inverted repeat RYYYAAC and GTTRRY core site, and yet prior work had found (Recchia *et al*., 1994) incoming cassettes are preferentially inserted at the *attI* site. It has been shown in certain assays that excess integrase could influence the insertion site because the two direct repeats (DR1 and DR2) in *attI* bind integrase and are critical for *attI* x *attC* recombination, but not for *attI* x *attI* recombination (Recchia *et al*., 1994; Partridge *et al*., 2000; Hansson *et al*., 1997). This phenomenon is similar to lambda where its integrase recognizes two distinct binding sites, core and arm, whose relative occupancy is thought to influence excision or insertion in conjunction with host encoded accessory proteins (Thompson *et al*., 1987; Landy, 1989; Rice *et al*., 1996). It was hypothesized that IntI behaves in a similar way using additional IntI and/or host encoded accessory proteins to direct recognition or complex interaction.

To examine this possibility I used magnetic streptavidin coated beads carrying a biotinylated 414-bp *attI* amplicand to capture IntI and any host proteins that might bind with it. Although *attI* is only 65-bp in length, surrounding sequence was included to cover any potential accessory protein binding sites as seen in other site-specific recombinase systems like lambda (Thompson *et al*., 1987; Ball *et al.*, 1991). By densitometry of electrophoresed biotinylated DNA samples before and after addition to beads, I found approximately 79% of DNA bound to the magnetic beads. However, integrase in a lysate of cells overexpressing His6-integrase did not bind this DNA, i.e. neither the 38.3 kDa integrase nor any host proteins as seen by Coomassie staining (Fig. 2.1, lanes 6-9). The only band eluting from the beads was the BSA in the binding and elution buffers, which runs at 80.6 kDa (Fig. 2.1, lanes 2-4, lanes 6-8).

Considering the possibility that the salt elution was not sufficient to remove tightly bound proteins, I also heated the beads at 95ºC for 5 min before removing the elution buffer from them. This resulted in the appearance of 5 bands, but the control sample without cell extract or integrase showed the same 5 bands (Fig. 2.1, lanes 5, 9). In case the 414-bp amplicand blocked IntI binding sites by hybridizing to itself due to secondary structure formation, a smaller 137-bp amplicand was used, but this yielded the same results (data not shown). Finally, I attempted to 'bait' the 137-bp amplicand with purified integrase and discovered the affinity of integrase for ds- DNA *attI* was too low to be seen in this assay (data not shown). Consequently it is also likely that any accessory proteins bound to the amplicand or in conjunction with integrase are also not abundant enough to be detected with this method. However, during the course of my work, others showed recombination between the single-stranded $attC_{aadA7}$ and double-stranded *attI* and discovered that double- versus single-stranded core site DNA directs recombination, not accessory proteins involvement (Bouvier *et al*., 2005).

Integrase binding affinity among mixed site populations

Recent *in vitro* work on single sites determined that pure IntI has a 5-fold higher affinity for ss-*attC* versus ds-*attI* (Bouvier *et al*., 2005). However, in nature these sites are all present at once, so IntI binding to *attC* may be affected by *attI*. In fact, folding differences observed in IntI bound to each half-site in the VchIntIA-VCR_{bs} co-crystal suggest that binding preferences could be attributed to different IntI conformations specific to either *attI* or *attC* site (MacDonald *et al*., 2006). A β-4,5 hairpin in different positions in the attacking and non-attacking subunit of the IntI dimer may be a molecular switch allowing alternate binding modes and providing specificity based on sequence of ds-*attI* and structure in ss-*attC* (MacDonald *et al*., 2006). This

would imply at least three IntI conformations: one for ds-*attI*; one for the non-attacking subunit on ss-*attC*; and finally one for the attacking subunit on ss-*attC*.

I examined this possibility through simultaneous and sequential competition protocols for mobility shift assays on IntI and ss-*attC* complex formation, which is believed to be one integrase molecule per half-site (Johannson, 2004). Simultaneous competition reactions with labeled ss-*attC*, unlabeled ds-*attI*, and IntI showed that excess *attI* did not compete with *attC* binding when they are added together (Fig. 2.2, Fig 2.3a). However, at very high excess *attI* does change the distribution between complexes I and II (Fig. 2.2, lane 3-9), compared to *attC* only (Fig. 2.2, lane 2). Interestingly, at low relative concentrations *attI* seemed to enhance IntI binding (i.e. increasing the intensity of complex II).

Conversely, when equal or half molar concentrations of *attI* (when compared to *attC*) were pre-incubated with IntI, then followed by *attC* incubation, complex II formation was not detected (Fig 2.3b). This suggests that when added alone, *attI* provokes IntI to adopt a conformation that prevents its binding to *attC* in the same way as when they are both present. I ruled out that the *attI* pre-incubation interference with IntI-*attC* complex formation was due to the short interaction period between IntI and *attC* by doing separate brief interactions without *attI* which resulted in formation of both complexes (although mostly complex I) similar to Fig. 2.2 lane 2 (data not shown). It is not possible that 8 nM *attI* has sequestered all of the 160 nM IntI so the conversion of IntI to an *attC* unreactive form by *attI* alone must be a catalytic rather than stoichiometric process. Unfortunately, pre-incubations of *attC* and integrase, followed by *attI* addition were not reproducible. Nonetheless, my preliminary findings were consistent with earlier speculation that different integrase conformations affect site preference and stable complex formation. This idea was further supported by another group who found that mutation

of the conserved aspartic acid at position 161 of IntI1, thought to have a central role in multimer assembly, increased *attI* x *attC* recombination and decreased *attC* x *attC* recombination (Demarre *et al*., 2007).

Growth phase dependency in strains expressing wildtype or high integrase levels.

Sequence variation and favored interactions are not the only differences between recognition sites. *In vitro* experiments revealed IntI1 preferentially binds double-stranded *attI*, but only the bottom strand of single-stranded *attC* (Francia *et al*., 1999; Johansson *et al*., 2004). This result led to speculation that growth phase might restrict the availability of single-stranded substrate to periods of active synthesis and transcription. To investigate growth phase variation in integrase-mediated recombination, I used qualitative PCR to detect recombination products every 5 h over 45 hours in strains carrying a cassette donor plasmid (pRMH14) and a recipient plasmid (pICV8), in the presence or absence of additional integrase (pSU2056). The donor plasmid (pRMH14) carried a wildtype *intI*, *attI* site, and two gene cassettes, *aacA4* and *aadA2* (Fig 2.5a). The recipient plasmid (pICV8) carried a truncated *intI*, complete *attI* site, and a truncated immobile *aadA1* cassette (Fig 2.5a). Excision of one or both cassettes from pRMH14 was detected with 5' junction primers resulting in a 1276-bp and 1915-bp PCR amplicand, respectively (Fig. 2.4b; Fig. 2.4c). However, integrons not only mediate complete cassette insertion and excision, but also cointegrate formation and resolution between plasmids (Shearer and Summers, 2009). Note that *attC* x *attI* and *attI* x *attI* recombination would result in the same 1276-bp and 1915-bp respectively, and consequently would be indistinguishable from complete cassette insertion. However, the cointegrate junction primers detected only cointegrate formation indicated by 2019-bp or 1380-bp amplicands (Fig 2.4b, Fig 2.4c).

With additional integrase, the 1276-bp (*attC* x *attI*) amplicand was present throughout the entire 45 h growth period, but surprisingly the 1915-bp (*attI* x *attI*) and 1380-bp (co-integrate) amplicands were either diminished or not detected at the 5, 10, and 15 h time points and appeared in stationary phase or in log phase (Fig 2.5 a,b). This cannot be attributed to high integrase concentrations, as strains grown without additional integrase had the same result (data not shown). These results seem to agree with other findings suggesting a growth phase dependency, but do not coincide with the expected time of occurrence as ss-DNA would be most abundant when cells were in early stationary phase (Fig 2.5 c). However, when PCR was subsequently done on 5, 10, and 15 h time points at 1/20, 1/100, 1/500 dilutions of DNA the missing1915-bp and 1380-bp amplicands appeared, indicating initial results arose from PCR interference due to excess DNA and cell debris. Therefore, qualitative PCR is not sensitive enough to discern whether integrase recombination is influenced by the growth cycle.

DISCUSSION

At the inception of this work, little was known regarding regulation of recombination mediated by the IntI and its two recognition sites *attI* and *attC*. Sequence variation among recognition sites, and the potential for *attC* to form cruciform structure clearly demonstrates a versatile recombinase with dual site specificity. However, despite the ambiguous nature of the recognition sites, the recombination site preference observed may be based in part on DNA substrates and differing IntI binding affinities, indicating some other element also controls site interactions. My pull-down assay explored the possibility of host encoded accessory proteins participation, but proved insufficiently sensitive to detect association with any protein, including IntI.

However, others work determined *attC* recognition is modulated via double versus single-stranded substrate. Specifically, the bottom strand of ss-*attC* was the preferred catalytically cleavable substrate *in vivo,* but was not observed for either strand of ss-*attI* (Bouvier *et al*., 2005). These authors established DNA site preference using suicide conjugation experiments that prevented the transferred plasmid from replicating in the recipient cell by withholding a replication protein (II), provided by a *pir* gene in the donor genome. The transfer of top or bottom strand of the recognition site was determined by its orientation relative to the *oriT* sequence. Suicide conjugation showed recombination was 1000-fold higher with the $attC_{aadA7}$ bottom strand than with top strand consistent with a distinct secondary structure resembling a double strand site (Bouvier *et al*., 2005). Such a structure is not restricted to this cassette only, as all *attC* sites exhibit potential cruciform structures (Hall *et al.,* 1991, Stokes *et al.*, 1997; Rowe-Magnus *et al.* 2003).

In addition, an ss-*attC*-integrase co-crystal showed how IntI could bind ss-*attC* without accessory proteins. The *attC*-integrase complex co-crystal was made with class 4 *Vibrio* integrase (VchIntIA) and its corresponding bottom single-stranded *attC* site (VCR_{bs}) confirmed that DNA secondary structure is important in IntI recognition and orientation (MacDonald *et al*., 2006). However, whether host factors initiate *attI* recognition, mediate differences between sites, or preferred substrates forms, is still an open question.

I also investigated competition and interaction between *attC* and *attI* and found the order of addition to be very important. *attI* does not compete with *attC* for IntI binding if the two DNA's are mixed before adding the integrase. Rather, in this case having *attI* present actually enhances IntI binding to *attC* enhancing a second supershifted complex. Conversely, when IntI is pre-incubated with *attI*, its binding of *attC* is limited to formation of complex I only. These

results could be explained by differing IntI folding conformations provoked by the single- or double- stranded substrate mixes. The pre-incubation of IntI and *attC* (Fig. 2.2., lane 2) results in the formation of predominantly complex I, but in order for recombination to occur, the initial *attC*-IntI complex must bind a second IntI subunit. It is possible that when *attI* is present initially with IntI (Fig. 2.3b), the *attI*-IntI complex interacts with the *attC*-IntI forming a recombination synapse which promotes binding of the second IntI subunit to *attC*. The *attI*-IntI complex could induce this *attC*-IntI conformation to occur more effectively than a second IntI-*attC* complex since the *attI*-IntI complex is more stable and less mobile (Demarre *et al*. 2007). Also, the recombination synapses are probably fleeting interactions, consequently they would not easily be detected on a gel. However, when the *attI* is pre-incubated with integrase, IntI recognition of *attC* was inhibited and only complex I was seen. This could be due to: (1) the flexibility of the ss-*attC* allowing IntI to bind I any conformation, but if *attC* has not yet bound an IntI in the correct conformation, the second IntI cannot bind and make complex II; or (2) the pre-incubated *attI* reactions catalyzed the formation of some IntI to the ds-*attI* specific conformation. This idea could be tested with a longer incubation period following the addition of *attC*, to see if the second complex would eventually appear as an equilibrium is reached.

Finally, I examined the possibility of growth phase dependent regulation of recombination. Unfortunately, qualitative PCR reactions of recombinant junctions between donor and recipient plasmids were not sensitive enough to answer this question. However, later quantitative dilution PCR experiments determined that both intracellular integrase concentration and growth phase affected integron recombination because recombination products increased through late log phase in both high and natural integrase expression (Shearer and Summers, 2009). Products continued to rise during stationary phase in strains with high integrase

concentrations, but decreased in the natural expression strain. In addition, natural integrase expression strains preferred *attI* x *attI* over *attI* x *attC* recombination throughout the growth cycle, while high integrase expression strains showed a similar preference until stationary phase when *attI* x *attC* recombination predominated (Shearer and Summers, 2009).

In summary, my work investigated three possible routes to modulate recombination based on either work with IntI1 or related recombinases. Two approaches proved to be limited by the sensitivity of the available techniques. So, accessory protein involvement could not be confirmed and the recombinant site preferences could not be quantified. However, it is probable that dual site specificity and interaction are determined through distinct IntI folding differences as suggested by the differing IntI conformations visible in the crystal structure and supported by my site competition results reported here. In the future it would be interesting to find the basis for the increase in junctions during late log phase. Does it result from an increase in integrase and if so is it controlled by transcription or translation? Or are integrase concentrations constant throughout growth, but environmental signals alter the IntI conformation in order to direct site recognition and/or interaction between sites?

ACKNOWLEDGEMENTS

I thank Dr. Lars Sundström University of Uppsala for his generous gift of the His-tagged IntI overexpression vector p2352 (Johansson *et al*., 2004).

REFERENCES

Ball, C.A. and R.C. Johnson. (1991) Multiple effects of Fis on integration and the control of lysogeny in phage λ. *J. Bacteriol*. **173:**4032-4038.

Bouvier, M., Demarre, G., and D. Mazel. (2005) Integron cassette insertion: a recombination process involving a folded single strand substrate. *EMBO J.* **24:**4356-4367.

Collis, C.M., Kim, M. Stokes, H.W., and R.M. Hall. (2002) Integron-encoded IntI integrases preferentially recognize the adjacent cognate attI site in recombination with a 59-be site. *Mol. Microbiol.*. **46:**1415-1427.

Collis, C.M., Recchia, G.D., Kim, M., Stokes, H.W., and R.M. Hall. (2001) Efficiency of recombination reactions catalyzed by class 1 integron integrases IntI1. *J. Bacterial*. **183:**2535-2542.

Collis, C.M. and R.M. Hall. (1992) Gene cassettes from the insert region of integrons are excised as covalently closed circles. *Mol . Microbiol*. **6:**2875-2885.

Demarre, G., Frumerie, C., Gopaul, D., and D. Mazel. (2007) Identification of key structural determinants of the IntI1 integron integrase that influence attC x attI1 recombination efficiency. *Nucleic Acids Research*. **35:**6475-6489.

Francia, M.V., Zabala, J.C., de la Cruz, F., and J.M. Garcia-Lobo. (1999) The IntI1 integron integrase preferentially binds single-stranded DNA of the *attC* site. *J Bacteriol*. **181:**6844-6849.

Francia, M.V., de la Cruz, F., and M. Garcia Lobo. (1993) Secondary sites for integration mediated by the Tn*21* integrase. *Mol. Microbiol*. **10:**823-828.

Gabrielsen, O.S., Hornes, E., Korsnes, L., Ruet, A., and T.B. Oyen. (1989) Magnetic DNA affinity purification of yeast transcription factor tau--a new purification principle for the ultrarapid isolation of near homogeneous factor. *Nucleic Acids Res.* **11:**6253-6267.

Hall, R.M., and H.W. Stokes (1993) Novel DNA elements which capture genes by site-specific recombination. *Genetica* **90**: 115–132.

Hall, R.M., Brookes, D.E., and H.W. Stokes. (1991) Site-specific insertion of genes into integrons: role of the 59-base element and determination of the recombination crossover point. *Mol. Mirobiol*. **5:**1941-1959.

Hansson, K., Skold, O., and L. Sundstrom. (1997) Non-palindromic *attI* sites of integrons are capable of site-specific recombination with one another and with secondary targets. *Mol. Microbiol.* **26:**441-453.

Kushner, S.R. (1978) An improved method for transformation of *Escherichia coli* with CotEl-derived plasmids. In Genetic engineering. (Boyer, H.W. and S. Nicosia, ed.), Elsevier/North Holland Publishing Co., Amsterdam, 17-23.

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

Landy, A. (1989) Dynamic, structural, and regulatory complexity in the site-specific recombination. *Annu. Rev. Biochem*. **58:**913-949.

MacDonald, D., Demarre, G., Bouvier, M., Mazel, D., and D.N. Gopaul. (2006) Structural basis for broad DNA-specificity in integron recombination. *Nature*. **440:**1157-1162.

Martinez, E., and F. de la Cruz. (1990) Genetic elements involved in Tn*21* site-specific integration, a novel mechanism of the dissemination of antibiotic resistance genes. *EMBO. J.* **9:**1275-1281.

Martinez, E., and F. de la Cruz. (1988) Transposon Tn*21* encodes a RecA-independent site-specific integration system. *Mol. Gen. Genet.* **211:**320-325.

Partridge, S.R., Recchia, G.D., Scaramuzzi, C., Collis, C.M., Stokes, H.W., and R.M. Hall. (2000) Definition of the attI1 site of class 1 integrons. *Microbiology*. **146:**2855-2864.

Recchia, G.D., and R.M. Hall. (1995) Gene cassettes: a new class of mobile elements. *Microbiology.* **141**:3015-3027.

Recchia, G.D., Stokes, H.W., and R.M. Hall. (1994) Characterization of specific and secondary recombination sites recognized by the integron DNA integrase. *Nucleic Acids Res.* **22:**2071-2078.

Rice, P.A., Yang, S.W., Mizuuchi, K., and H.A. Nash. (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell.* **87:**1295-1306.

Rowe-Magnus, D.A., Guerout, A.M., Biskri, L., Bouige, P., and D. Mazel. (2003) Comparative analysis of superintegrons: engineering extensive genetic diversity in the vibrionaceae. *Genome Res.* **13:**428-442.

Schneider, K. and C.F. Beck (1986) Promoter-probe vectors for the analysis of divergently arranged promoters. *Gene*. **42:**37-48.

Shearer, J.E.S and A. O. Summers. (2009) Intracellular steady-state concentration of integron recombination products varies with integrase level and growth phase. *J Mol Biol.* **386:**316-331.

Stokes, H.W., O'Gorman, D.B., Recchia, G.D., Parsekhian, M., and R.M. Hall. (1997) Structure and function of 59-base element recombination sites associated with mobile gene cassettes. *Mol. Microbiol*. **26:**731-745.

Stokes, H.W., Tomaras, C., Parsons, Y., and R. M. Hall. (1993) The partial 3'-conserved segment duplications in the integrons In6 from pSa and In7 from pDGO100 have a common origin. *Plasmid*. **30:**39-50.

Stokes, H.W. and Hall, R.M. (1989) A Novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol. Microbiol*. **3:**1669-1683.

Thompson, J.F., Moitoso de Vargas, L., Koch, C., Kahmann, R., and A. Landy. (1987). Cellular factors couple recombination with growth phase: characterization of a new component in the λ site-specific recombination pathway. *Cell.* **50:**901-908.

Table 2.1 Strains, plasmids, and primers

Strains	**Phenotype/Description**	**Reference**
CB454	*lacZ lacY galK thi rpsL recA*	Schneider and Beck, 1986
SK1592	*galK thi* T1 *rec*$^+$ *hsm*$^+$ *hsrR4 endA sbcB15*	Kushner, 1978
Plasmids	**Phenotype/Description**	**Reference**
pRMH14	Km^R, Sm^R, Su^R, Tra$^+$, IntI1$^+$, donor plasmid	Stokes *et al*., 1993
pICV8	Zeo^R, recipient plasmid	Shearer and Summers, 2009
pSU2056	Ap^R, IntI1 P_{lac} overexpression vector	Martinez and de la Cruz, 1990
pICV8::aadA2	Zeo^R, Sm^R, primer pair control plasmid	Shearer and Summers, 2009
p2352	Ap^R, His-tagged IntI1 overexpression vector	Johansson *et al*., 2004
Primers	**Sequence**	**Reference**
aadA2-U1 5' junction	5'GCCGGTTATTGCGCTGTACCAAATG3'	Shearer and Summers, 2009
4E-L2 5' junction	5'GCCTATGCCTACAGCATCCAGGGTGAC3'	Shearer and Summers, 2009
8E-U1 coint junction	5'CCTCGTTAAAGGACAAGGACCTGAG3'	Shearer and Summers, 2009
int-L1 coint junction	5'CGCGCTGAAAGGTCTGGTCATAC3'	Shearer and Summers, 2009
attI414U	5'CTCGATGACGCCAACTACCT3'	This study
attI414L	5'AGCCAGGACAGAAATGCCTC3'	This study
attI137L	5'GTCGATGTTTGATGTTATGGA3'	This study
PCR amplicand & oligo	**Description**	**Reference**
aadA7bs (oligo)	ss-*attC* bottom strand from *aadA7*, used in EMSA	Bouvier *et al*., 2005
*attI*137 (amplicand)	ds-*attI* containing 4 IntI binding sites used in EMSA	This study

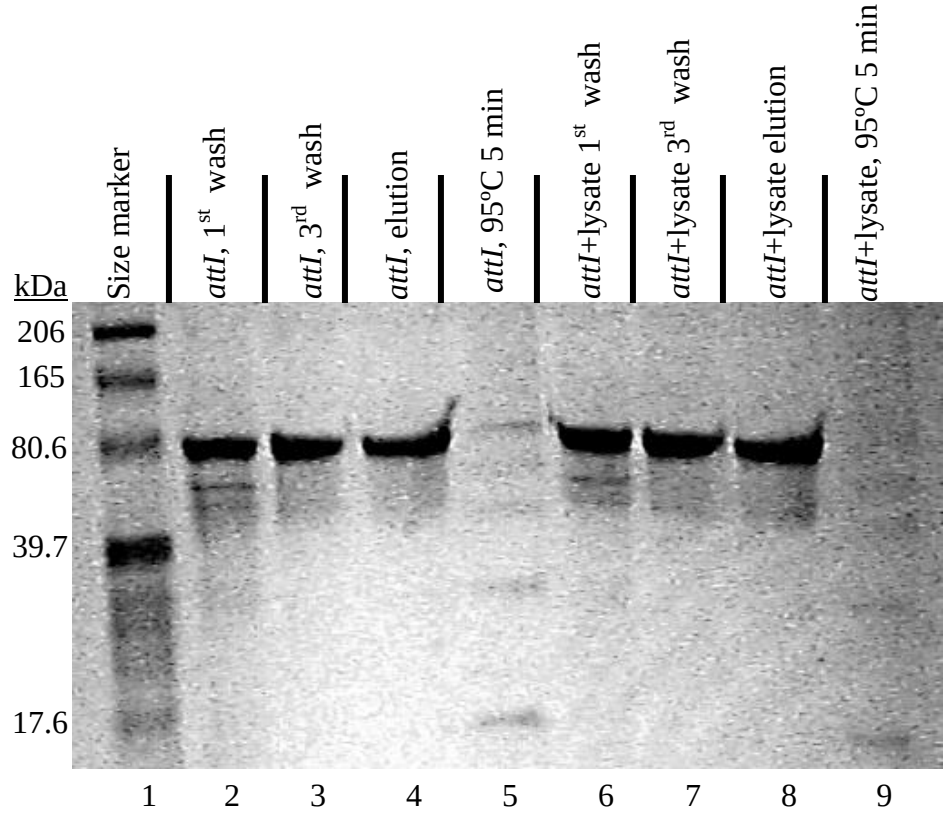


Figure 2.1 Pull-down assay to identify possible accessory factors

Lanes 2, 414-bp ds-*attI* amplicand washed with 50 uL wash buffer 1 (50 mM Tris, 100 mM KCl, 1 mM EDTA, 5% glycerol, 0.5 mg/mL BSA, pH 7.5); Lane 3, ds-*attI* amplicand washed with 50 uL wash buffer 3 (50 mM Tris, 300 mM KCl, 1 mM EDTA, 5% glycerol, 0.5 mg/mL BSA, pH 7.5); Lane 6, ds-*attI* amplicand + induced protein lysate, washed with wash buffer 1; Lane 7, ds-*attI* amplicand + induced protein lysate, washed with wash buffer 3; Lanes 4 and 8, 50 uL elution buffer (50 mM Tris, 1 M KCl, 1 mM EDTA, 5% glycerol, 0.5 mg/mL BSA, pH 7.5); Lanes 5 and 9, heat elution. Lanes 2-4, 6-8, the 80 kDa band is BSA present in both wash and elution buffers.

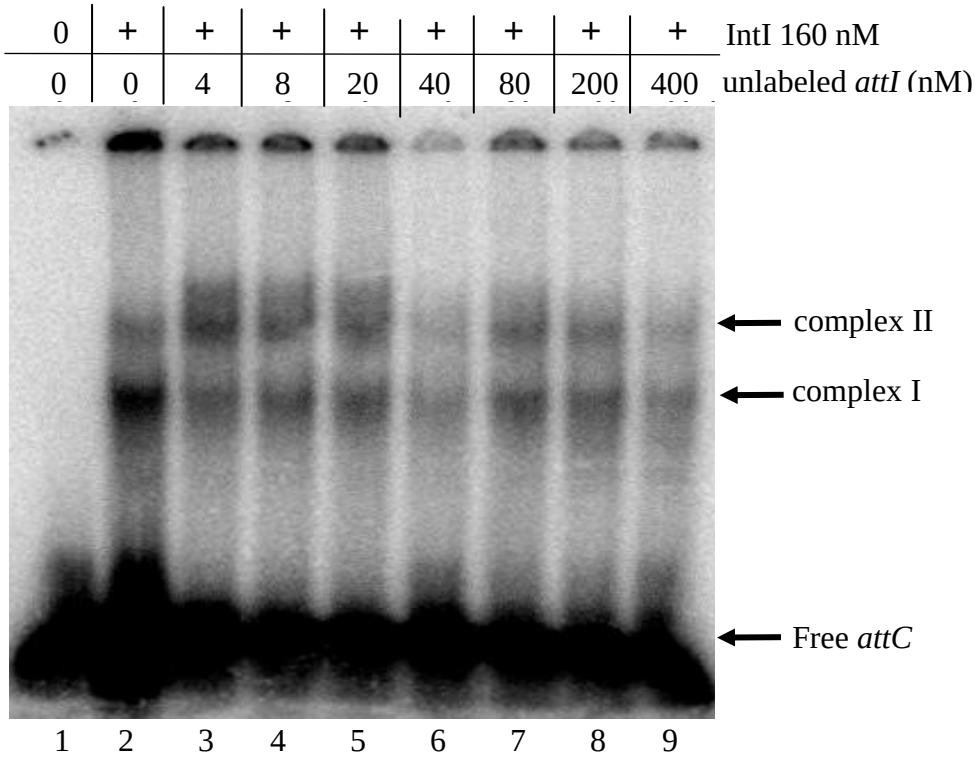


Figure 2.2 ds-*attI* alters ss-*attC*-integrase binding complex

Lane 1, 8 nM P^{32} 62-bp *aadA7* bottom strand *attC*; Lane 2, 8 nM P^{32} *attC*, 160 nM integrase; Lanes 3-9, 8 nM P^{32} *attC*, 160 nM integrase, and 4-400 nM of 137-bp unlabeled *attI*.

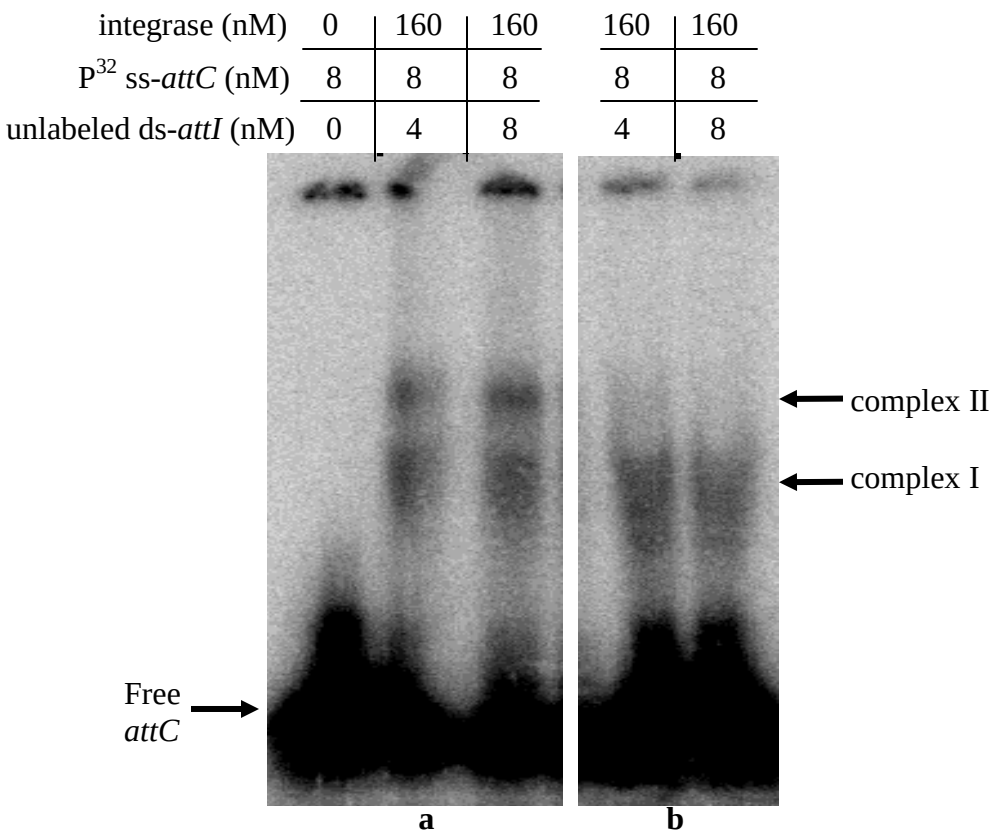
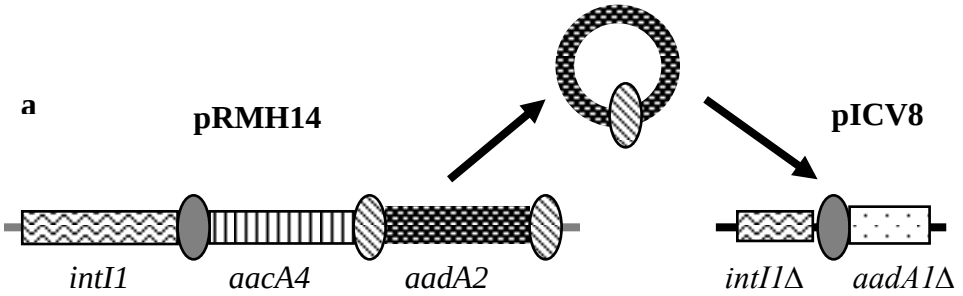
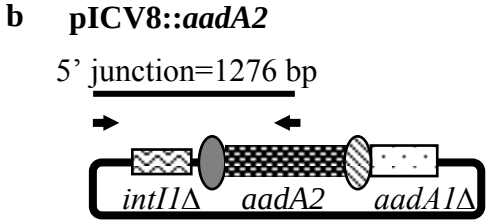
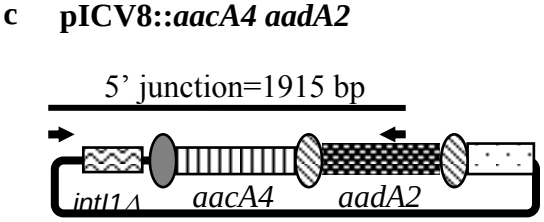


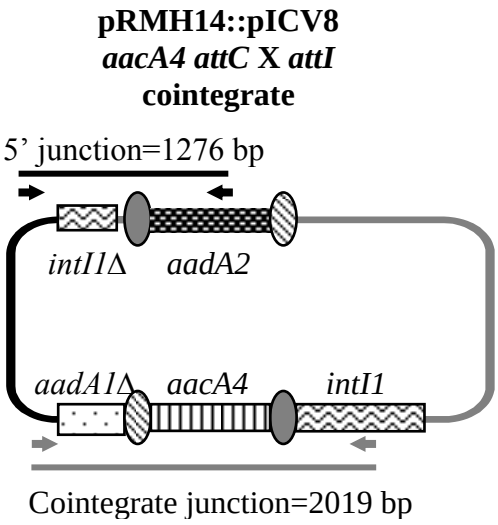
Figure 2.3 ds-*attI* site competes with ss-*attC* site, limiting complex formation

Sequential competitive mobility shifts consisting of 8 nM *attC*, 160 nM integrase, and 4 or 8 nM unlabeled *attI* incubated as follows: **(a)** *attI* + *attC* incubated 3 min, integrase added and incubation continued 25 min; **(b)** *attI* + integrase incubated 25 min, *attC* added and incubation continued 3 min**.**


a
pRMH14
pICV8
intI1
aacA4
aadA2
intI1Δ
aadA1Δ


b
pICV8::aadA2
5' junction=1276 bp
intI1Δ
aadA2
aadA1Δ


c
pICV8::aacA4 aadA2
5' junction=1915 bp
intI1Δ
aacA4
aadA2


pRMH14::pICV8
aacA4 attC X attI
cointegrate
5' junction=1276 bp
intI1Δ
aadA2
aadA1Δ
aacA4
intI1
Cointegrate junction=2019 bp

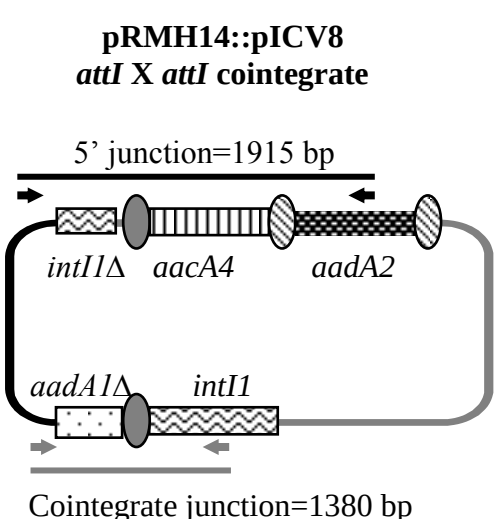
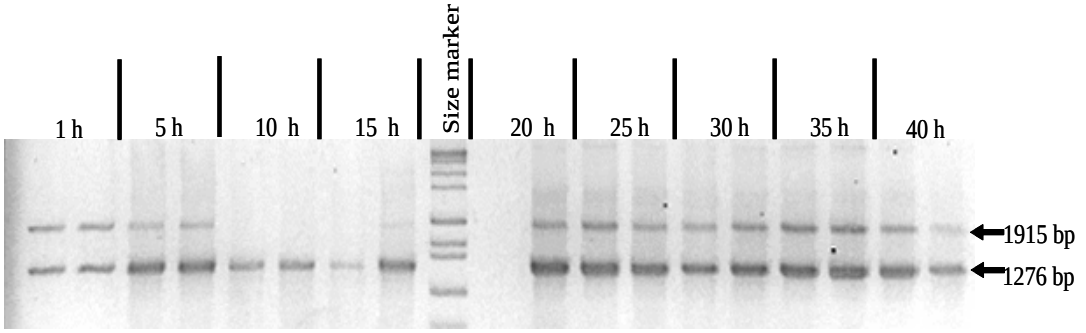

pRMH14::pICV8
attI X attI cointegrate
5' junction=1915 bp
intI1Δ
aacA4
aadA2
aadA1Δ
intI1
Cointegrate junction=1380 bp

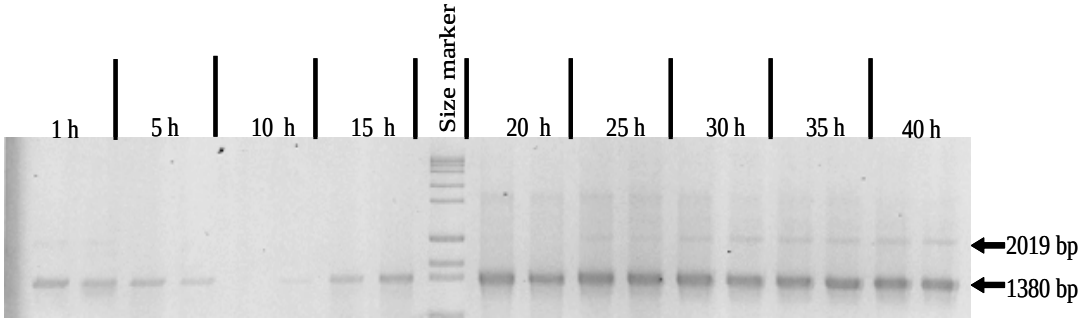
Figure 2.4 Integron cassette capture and cointegrate formation between donor and recipient plasmids.

(a) Prototypical transfer mechanism of a gene cassette in the form of covalently closed circle from donor plasmid pRMH14 to recipient plasmid pICV8. (**b)** Measure of *aadA2* single cassette transfer (top) or cointegrate formation (bottom) indicated by 1276-bp amplicand using 5' junction primer pair. In addition, cointegrate formation only (bottom) between pRMH14 *aacA4 attC* x pICV8 *attI* indicated by 2019-bp amplicand using cointegrate junction primer pair. (**c)** Measure of *aadA2* and *aacA4* double cassette transfer (top) or cointegrated formation (bottom) indicated by 1915-bp amplicand using 5' junction primer pair. In addition, cointegrate formation only (bottom) between pRMH14 *attI* x pICV8 *attI* indicated by 1380-bp amplicand using cointegrate junction primer pair. Figure from (Shearer and Summers, 2009) with permission.

a


1 h
5 h
10 h
15 h
Size marker
20 h
25 h
30 h
35 h
40 h
1915 bp
1276 bp

b


1 h
5 h
10 h
15 h
Size marker
20 h
25 h
30 h
35 h
40 h
2019 bp
1380 bp

c

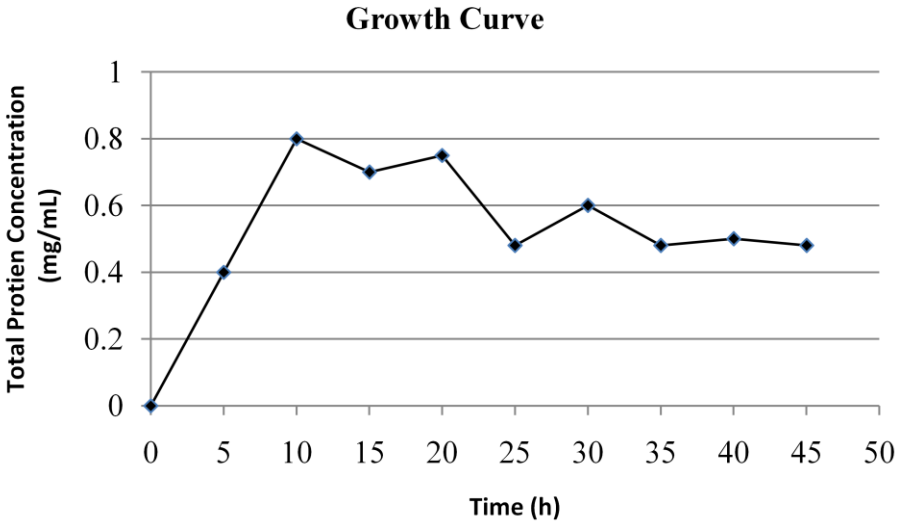

Growth Curve
Total Protien Concentration (mg/mL)
1
0.8
0.6
0.4
0.2
0
0
5
10
15
20
25
30
35
40
45
50
Time (h)

Figure 2.5 PCR amplification of the 5' and cointegrate junctions demonstrated reduced recombination during early stationary phase.

(a) 5' junction PCR of CB454(pRMH14)(pICV8)(pSU2056). Insertion of *aadA2* only is indicated by 1276-bp amplicand and insertion of both *aacA4* and *aadA2* is indicated by 1915-bp amplicand. Duplicate PCRs shown for each timepoint, except 20 h, due to a loading error. (**b)** Conintegrate junction PCR of CB454(pRMH14)(pSU2056). pRMH14:pICV8 (*attI x attI*) cointegrate recombination is indicated by 1380-bp amplicand and *aacA4 attC* x *attI* cointegrate is indicated by 2019-bp amplicand. Duplicate PCRs shown for each timepoint. **(c)** CB454(pRMH14)(pICV8)(pSU2056) growth curve indicated via total protein concentration.

CHAPTER 3

REGULATION OF THE INTEGRASE AND CASSETTE PROMOTERS OF THE CLASS 1 INTEGRON BY NUCLEOID-ASSOCIATED PROTEINS

OVERVIEW

Integron integrase IntI1 mediates the exchange of antibiotic resistance gene cassettes in the integron. I examined the transcriptional strength and the roles of global regulators in the Tn*21* integron (In2) integrase promoter (P_{int}) and cassette promoters (P_c and P_2). The -10 and -35 hexamers of P_{int} are close to consensus for σ^{70} RNAP, but there are seven versions of P_c and P_2 cassette promoters among class 1 integrons. In In2, the P_2 -10 hexamer overlaps the P_{int} -10 hexamer by 3-bp, but in four versions the P_2 promoter has only a 14-bp spacer and is assumed to be inactive. Using bi-directional transcriptional fusions, I found P_{int} expression to be relatively weak compared to that of P_c and P_2. The 14-bp spacer of the P_2 promoter is not inactive, but very weak compared to the 17-bp P_2 spacer. The regulator predictor program, (PRODORIC 8.9), identified conserved regulators FIS, LexA, and IHF sites near the promoter region. I found with transcriptional fusions that FIS repressed integrase and cassette expression, and LexA repressed P_{int} and the P_2 with 14-bp spacer, which is also a LexA site. IHF activated P_{int} and the P_2 cassette promoter with the 17-bp spacer, but not P_2 with 14-bp spacer. Lastly, H-NS a regulator with no sequence specific recognition site gave varied results, but generally repressed P_{int} and P_2 with a 14-bp spacer. The involvement of growth phase-regulated nucleoid-associated

proteins FIS and IHF suggests transcriptional control based on nucleo-protein structures that may subject integron promoters to growth phase regulation.

INTRODUCTION

The class 1 integron integrase is a unique member of the tyrosine site-specific recombinase family. Class 1 integrase is found in transposable elements and/or on broad-host-range conjugative plasmids (Stokes and Hall, 1989; Hall and Collis, 1995) and is important in the dissemination of multi-drug resistant bacteria. The integron consists of two conserved regions flanking a variable region in which are located multiple gene cassettes. These cassettes are mobile, non-self-replicating DNA elements encoding an open reading frame (ORF) and an integrase-specific recombination site called *attC* (Hall and Collis, 1995; Hansson *et al*., 1997). The conserved region 5' to the cassettes encodes the integrase gene (*intI1*) and an adjacent recombination site, *attI* where cassettes are typically inserted (Hall and Stokes, 1993).

Most inserted gene cassettes lack their own promoter (Stokes and Hall, 1991; Tolmasky and Crosa, 1993), and are transcribed as one transcript by a promoter (P_c), which is located in the divergently transcribed *intI1* gene (Stokes and Hall, 1991; Collis and Hall, 1995; Recchia and Hall, 1995). Cassettes closest to the *attI* insertion point have higher expression than those situated further from *attI* (Collis and Hall, 1992). A rare secondary cassette promoter (P_2) occurs in In2 and other integrons due to a 3-bp insertion that generates a 17-bp spacer adventitiously placed between -35 and -10 hexamers. Without this insertion P_2 with only a 14-bp spacer is thought to be nonfunctional. The integrase promoter P_{int} reads convergently towards P_c, and overlaps with the P_2 -10 hexamer.

Little research on the transcription of gene cassette arrays has been done, but a recent comparison of *intI* transcription of chromosomal *V. cholera* integron and mobile class 1 integron of *E. coli* showed LexA regulation of both (Guerin *et al.,* 2009). P_{int} is highly conserved in all variants (Zhang *et al*., 2000), but cassette promoters are much more variable. Currently, there are seven known cassette promoter variants (Table 3.2). The expression of each variant was tested and named relative to strength as follows: 'strong', 'strong + 2nd', 'weak', 'weak + 2nd', 'hybrid 1', 'hybrid 1 + 2nd', and 'hybrid 2' (Lévesque *et al*., 1994; Papagiannitis *et al*., 2009). Papagiannitis *et al*., 2009 speculate the cassette promoter variability may help the establishment of other factors shown to influence expression in multicassette arrays like cassette position and differing recombination frequency observed among *attC* sequences.

Interestingly, P_c transcription results in a very long leader as it is located approximately 200-bp upstream of the *attI* cassette insertion site. The transcript has extensive potential for strong secondary structures but has not been investigated for a role in cassette expression. Recent transcriptome analysis of gene cassettes in the superintegron of *V. cholerae* strain N16961 using mutants in *hapR* (regulator of virulence factor TcpP), *rpoS* (stationary phase sigma factor), or *rpoN* (nitrogen stress sigma factor) revealed that cassettes are positively controlled by HapR and RpoS and negatively controlled by RpoN. HapR increases at high cell densities and RpoS increases in response to nutrient stress and during stationary phase. The authors speculate that cassette expression increases in crowded, stressed, or non-replicating bacteria to provide a survival advantage by exchanging putative beneficial cassettes (Yildiz *et al*., 2004).

Here I used bidirectional transcriptional fusions of the convergent promoters P_{int} *(lacZ* reporter), and $P_c + P_2$ or P_2 (*phoA* reporter) to the Tn*21* integron (In2) and mutants of them to

ask: (i) what is the activity of P_{int} alone and when competing with one or both cassette promoters; (ii) what is the activity of P_2 with 14-bp or 17-bp spacer. I also examined the roles of global regulators: FIS (factor inversion stimulation), IHF (integration host factor), and SOS regulon repressor LexA, predicted to have binding sites in this promoter region. Transcriptional fusion constructs were tested in *fis*, *ihf*, and *lex* hosts and purified proteins were used in electrophoretic mobility shift assays (EMSA) with various promoters. I also examined the global regulator H-NS (histone-like nucleoid structuring) which recognizes structure not sequence.

MATERIALS AND METHODS

Bacterial strains, plasmids, and culture conditions

Bacterial strains and promoter constructs are described in Table 3.1. *E. coli* strains were grown in Luria-Bertani (LB) broth at 37°C and 250 rpm and supplemented with 100 μg/ml ampicillin (Ap) or 50ug/ml kanamycin (Km). pWT, pI2-17, and pI templates were constructed by PCR amplification of the promoter region of In2 in Tn*21* of plasmid R100 incorporating *BamHI* and *HindIII* restriction sites using primer pairs IP212U & IP212L, IPU & IP134L, and IPU & IP97L (Table 3.1). PCR reactions (50 ul) were programmed for 5 min at 95ºC followed by, 30 cycles of 30s at 95ºC, 30s at 57.4ºC, 30s at 72ºC, and a final elongation step of 5 min at 72ºC. Amplicands were digested with *Bam*HI and *Hind*III and ligated to *Bam*HI- and *Hind*III- digested pCB267 (Klaus and Beck, 1985). pI2-14 was generated from pI2-17 template by PCR amplification with primers IP131U & IP131L using QuikChange Site-Directed Mutagenesis Kit (Stratagene).

β-Galactosidase Assay

Overnight cultures in LB broth were diluted 1/20 with fresh LB containing 100 ug/ml Ap. The cells were grown an additional 6 h and a 1 ml aliquot taken cells were pelleted and washed and

resuspended in duplicate (Miller, 1972). Cells were lysed according to (Zhang and Bremer, 1995), except lysis was done at 25ºC for 30 min in a microtiter plate. 100 ul of *o*-nitrophenyl-β-d-galactopyranoside (ONPG) 5 mg/ml was added to each reaction which was then incubated at 25ºC for 60 min. During incubation, A_{405} (ONPG) and A_{550} (cell debris) readings of sample wells were done 0, 15, 30, 45, 60 min after addition of ONPG to determine ONPG hydrolysis saturation and assay for β-Galactosidase activity. Each reported activity is the average of three independent experiments done in duplicate. Host strain background activity done at the same time was subtracted in each experiment.

Alkaline Phosphatase Assay

Cells were prepared as for the β-galactosidase assay except, that the cells were washed with 1 mM Tris-HCl, pH 8.0. The washed cells were resuspended in 1 ml of 1 mM Tris-HCl, pH 8.0. Cells were lysed with 1 mM Tris-HCl, pH 8.0, 0.8 mg/ml CTAB, and 0.4 mg/ml sodium deoxycholate at 25ºC for 30 min in a microtiter plate. 100 ul of 104 phosphatase substrate 5mg/ml (Sigma) was added to each reaction which was then incubated at 25ºC for 60 min. During incubation, A_{405} and A_{550} readings of sample wells were done 0, 15, 30, 45, 60 min after addition of phosphatase to determine saturation and assay for alkaline phosphatase activity. Each reported activity is the average of three independent experiments done in duplicate and host strain background activity done at the same time was subtracted in each experiment.

Electrophoretic Mobility Shift Assay (EMSA)

PCR amplicands from pWT, pI2-17, pI2-14, and pI were cleaned with QIAquick PCR purification kit (Qiagen). Cleaned pI2-14 PCR was also gel extracted with QIAquick gel extraction kit (Qiagen) to remove PCR template (pI2-14). FIS, IHF, and H-NS at 0, 5, 50, 500, 2500, and 5000 nM were mixed with 5 nM DNA in 10 μl binding buffer 1 [20 mM Hepes (pH

7.5), 100 mM NaCl, 5% (v/v) glycerol, 100 μg/ml bovine serum albumin (BSA), 2 mM DTT, and 1 mM EDTA] for 30 min at 25ºC. Reactions with LexA were done at 0, 20, 100, 200, 1000, 2000 nM with 5 nM DNA in 10 ul binding buffer 2 [10 mM HEPES (pH 7.9), 10 mM Tris (7.9), 5% (v/v) glycerol, 50 mM KCl, 1 mM EDTA, 1 mM DTT, 50 ug/ml BSA] (Mazón *et al*., 2004). Mixed protein EMSAs were done in buffer 1 with predetermined minimum protein binding concentration (5 or 500 nM) for one protein and incubated with 5, 50, or 500 nM concentrations of the other protein (IHF, FIS, LexA, or H-NS) present at start of reaction. All 10 ul of each binding reaction was immediately loaded onto a 5% nondenaturing polyacrylamide ready gel (Bio-Rad) and run at 25ºC in 1x TBE buffer for 5 min at 120 V, then 40 min at 80 V and stained with Syber Green (Invitrogen). Purified FIS, IHF, H-NS, and LexA were generous gifts from Dr. Anna Karls (University of Georgia, GA), Dr. Ishita Mukerji (Wesleyan University, CT), Dr. Sylvie Rimsky (ENS Cachan, France), and Dr. John Little (University of Arizona, AZ), respectively.

RESULTS

Promoter strength of P_{int}, P_c and/or P_2 when isolated or competing

Relative promoter strength has been determined for the cassette promoter variants (Lévesque *et al*., 1994; Papagiannitis *et al.*, 2009). However, these represent only part of the integron transcription mechanics as P_c and P_2 converge and the latter partially overlaps the integrase promoter (P_{int}). P_{int} expressed best alone (P_i), increasing 6.5-fold in wild type host without the competing cassette promoters P_c and P_2 (Fig. 3.2). Surprisingly, deletion of P_c only (P_{i2}) weakens P_{int} expression making it almost 12-fold less active than when alone (Fig. 3.2 b). Moreover, in the absence of P_c by the removal of three CCC's ($P_{i2\text{-}14}$, resulting in the 14-bp

spacer) there is very little P_{int} expression (Fig. 3.2 b), due to LexA binding to its cognate site formed by the 3-base deletion (Guerin *et al.*, 2009).

Cassette expression previously tested in Tn*2603*, found P_2 accounted for approximately 90% of total cassette expression (Lévesque *et al*., 1994). My *phoA* transcriptional fusions to P_2 alone (P_{i2}) and to P_c and to P_2 (P_{i2c}) indicate that P_c contributed comparatively little to cassette expression and agreed with previous findings (Fig. 3.3b). It has been assumed that P_2 with the 14-bp spacer ($P_{i2\text{-}14}$) is non-functional. Indeed, although not completely 'off' ($P_{i2\text{-}14}$) has less than 1% of P_2 function (Fig. 3.3a).

Predicted FIS, IHF, and LexA transcriptional regulator binding sites

I initially looked at commonly used accessory proteins in other site-specific recombinases in regards to recombinational control. Later, I returned to them as possible transcription regulators since investigation of transcriptional control of the integrase and cassette promoter(s) was in preliminary stages. To examine this point I used the http://www.prodoric.de (Münch *et al*., 2003) virtual footprint tool to search the entire integron promoter sequence of Tn*21*. PRODORIC predicted several potential binding sites for FIS and IHF, multifunctional proteins that can serve as recombination accessory elements, nucleoid compaction proteins, and global transcriptional regulators; which influence multiple phenotypes and regulate genes from different metabolic pathways (Gottesman, 1984; Luijsterburg *et al*., 2006; Dorman, 2009). All of these proteins have highly degenerate 15 to 18-bp sites, therefore, binding locations for each protein are based on a program generated probability score (Münch *et al*., 2003; Table 3.3; Fig. 3.1).

Prior research indicates LexA repression of P_{int} (Guerin *et al*., 2009), but note that PRODORIC only predicted LexA sites in P_2 with the 14-bp spacer region ($P_{i2\text{-}14}$), found in four promoter variants. LexA sites were predicted on both the top and bottom strands of $P_{i2\text{-}14}$ having

high scores of 11.19 and 12.50, respectively (Table 3.3; Figure 3.1). LexA represses the 43 genes of the SOS regulon in *E. coli* (Gama-Castro *et al*., 2008), including itself (Little *et al*., 1981) by binding a 16-bp recognition sequence.

FIS and LexA repress integron promoters

Mutants lacking FIS showed a 25% increase in expression of P_{int} and P_2 (compared to wildtype), regardless of whether P_c was present (P_{i2C} and P_{i2}; Fig. 3.2b, Fig. 3.3b). When both the cassette promoters were deleted (P_i; Fig. 3.2a) P_{int} increased 2.5-fold without FIS. Interestingly, FIS also impacts P_{int} and P_2 with the 14-bp spacer ($P_{i2\text{-}14}$), whose expressions increase 16- and 3.5-fold, respectively, when FIS is absent (Fig. 3.2b, Fig.3.3a). This was unexpected as the formation of a LexA site in $P_{i2\text{-}14}$ was expected to result in complete repression of both by LexA(Guerin *et al*., 2009).

P_{int} expression was repressed by LexA increasing 23-fold in the absence of LexA ($P_{i2\text{-}14}$, Fig. 3.2b), P_2 with the short spacer is repressed only 3-fold by LexA ($P_{i2\text{-}14}$, Fig. 3.3a), and LexA has no effect on either P_{int} or P_c and/or P_2 expression if the 17-bp spacer is present. Therefore, it is likely there is no LexA site in the common P2 17-bp spacer (P_{i2c}, P_{i2}; Fig. 3.2b, Fig. 3.3b), but P_{int} expression in P_i increases 2-fold without LexA even though it lacks a LexA site (Fig. 3.2a).

There were three predicted sites for IHF involvement (Fig. 3.1). In the absence of IHF, P_{int} expression decreased 30% in P_{i2c} and P_c and/or P_2 decreased only 15-25% in P_{i2} (Fig. 3.2., Fig. 3.3), but neither decrease was statistically significant. From these data alone, IHF may be a weak activator of P_{int} and its role with the cassette promoters is not clear.

In vitro binding of LexA, FIS, and IHF to the integron promoter region

Effects on integron promoters fluctuations in *lex*, *fis*, and *ihf* could arise from direct interaction of those proteins with the integron promoters, or indirectly by regulation of some

other transcription factor. Electrophoretic mobility shift assays (EMSAs) with purified LexA, FIS, and IHF were used to examine this point. As expected, LexA had a high affinity for P_{i214}, partially retarded when FIS was present at equal molarites to P_{i214} and resulted in one complex formation, but barely retarded DNA with P_{int} and P_2 with 17-bp spacer (P_{i2}) (Fig. 3.4 a,b). FIS bound with equally high affinity to wild type promoter region (P_{i2c}) and resulted in 5 complexes (Fig. 3.4c). The same high binding affinities were observed between FIS and the other three promoter amplicands, but the number of complexes at low FIS concentrations differed among P_{i2c}, P_{i2}, P_{i2-14}, and Pi. Equal molar concentrations of P_{i2c}, P_{i2}, P_{i2-14}, or Pi with FIS resulted in the formation of five, four, four, and two distinct complexes, respectively (Fig. 3.4 d). It is unclear whether these complexes represent up to five separate FIS binding sites.

IHF formed stable complexes with all promoter amplicands, but only at a 100-fold excess (Fig. 3.5), which could be considered non-specific. However the significance of IHF involvement was not discounted here since the majority of IHF binding is non-specific (Yang and Nash, 1995; Arfin *et al*., 2000; Ali *et al*., 2001). Two IHF complexes were observed with DNA lacking the P_c promoter (P_{i2-14} and P_{i2}; Fig 3.5 b, c), but not with wild type (P_{i2c}) or P_{int} alone (P_i) (Fig. 3.5 a, d). Since both complexes (iI and iII) form with amplicands (P_{i2-14} and P_{i2}), this indicates P_2 length does not affect IHF binding (Fig 3.5 b,d).

FIS and IHF are most abundant during distinct periods of the growth cycle. During early exponential phase FIS is present at 30,000 dimers per cell, and IHF is present at 6,000 dimers per cell (Azam *et al*., 1999). However, in the transition from exponential to stationary phase, IHF peaks at 27,500 dimers per cell, and FIS decreases to less than 1000 dimers per cell (Azam *et al*., 1999). LexA's cellular concentration does not depend on growth phase, but increases when DNA is damaged (Kelley, 2006). So I asked whether these proteins affect each other's binding

to the promoter region. FIS and IHF or LexA and IHF were incubated with the wildtype promoter (P_{i2c}; Fig 3.6a) or with P_2 with 14-bp containing the LexA site ($P_{i2\text{-}14}$) (Fig.3.6 b). IHF did not compete with FIS or LexA binding, except at very high concentrations (Fig.3.6, lanes 7 and 8).

The effect of H-NS on integron promoter regions

H-NS is a nucleoid packing protein that does not bind via a sequence specific site, but instead through DNA secondary structure recognition, more specifically curved DNA with A/T rich regions (Dame *et al*., 2001; Navarre *et al*., 2006). H-NS is a global regulator that controls approximately 150 genes in *E. coli*, typically as a repressor (Gama-Castro *et al*., 2008). P_{int} expression increases without H-NS in all promoter constructs, except P_{i2} when expression is not detectable (Fig. 3.7 a, b). Repression varies in each construct increasing 5-fold in $P_{i2\text{-}14}$, 40% in P_i, and wildtype P_{i2c} does not change in the absence of H-NS (Fig. 3.7a, b). However, the cassette promoters are activated by H-NS, decreasing 35-40% in both P_{i2c} and P_{i2} (Fig. 3.7 c). Surprisingly, P_2 with 14-bp ($P_{i2\text{-}14}$) expression increases 60% without H-NS (Fig.3.7 d), suggesting H-NS has dual function, not only regarding integrase and cassette promoters, but possibly among different cassette promoter variants. H-NS directly interacts with the promoter region of P_{i2c} forming one or two retarded complexes (Fig. 3.8 a) as well as to P_{i2} and $P_{i2\text{-}14}$ (data not shown). In all promoter amplicands complex formation occurred at a range of 1 to 100-fold excess, except with P_i when binding was not detected (Fig. 3.8b). I predict H-NS does not bind to P_i which contains only three of seven A's in the A-tract of the P_2 spacer region, as this A-tract is the most likely H-NS binding site based on sites in other systems. I can not determine affinity, as I was unable to establish consistent repeatable binding ratios with H-NS.

DISCUSSION

The class 1 integron has seven different cassette promoter variants, but a highly conserved integrase promoter (P_{int}). There has been little study of P_{int} expression and no information on regulatory factors that may affect P_{int} and the cassette promoters. Here I establish that alone P_{int} is a relatively strong promoter and its weak expression in In2 can be ascribed to direct competition with P_c and P_2. I also determined the minor assistance P_c might contribute to P_{int} expression and speculate that although this P_c variant is the weakest of all cassette promoters, its expression may assist P_{int} slightly by occluding RNAP binding to the stronger P_2. However, differences could be attributed to an artifact of the constructs stemming from secondary structure formation at the start of *lacZ* reporter in P_{i2}. I also confirmed that the 14-bp spacer does render P_2 essentially inactive. However, although P_2 expression is reduced, P_{int} expression is barely detectable due to LexA binding at the acquired site in the P_2 spacer, as also noted in other recent work (Guerin *et al*., 2009).

My work shows for the first time the direct repressive action of FIS on both integrase and cassette promoters. Although, LexA represses the variants with the P_2 14-bp promoter ($P_{i2\text{-}14}$), FIS repression affects both P_{i2} and $P_{i2\text{-}14}$. I also show for the first time the direct interaction of IHF, activating P_{int} expression and possibly cassette promoters. Finally, my work suggests the direct involvement of H-NS, but I am unable to speculate on the role it may play in regulation due to result variability and do not include in it in my model.

On the basis of these observations, I propose a model (Fig. 3.9) in which LexA and/or FIS binding result in moderately repressed cassette expression. Here I show IHF is unable to displace FIS or LexA, or bind the unoccupied promoter region unless present at high concentrations. Therefore, depending on promoter variant, complete or partial repression of P_{int}

and partial cassette promoter repression occurs during early exponential growth when FIS concentrations are highest and IHF are lowest. However, as cells transition from exponential to stationary growth, FIS concentrations dramatically decrease, while IHF concentrations increase. Consequently, at this point IHF would displace FIS, resulting in the up-regulation of integrase and cassette promoters. It is possible upregulation may be due to a more desirable structure in the DNA as FIS and IHF bend DNA upon binding 40-90º and 140º, respectively. Still for those promoter variants that have P_2 14-bp, it is unclear whether IHF would displace LexA as its concentration, which is not growth phase dependent, may be too high.

I propose promoter variants provide varied depths of transcriptional regulation. Each variant is likely subject to the dual involvement, but separate interactions of FIS repression and IHF activation. The varied growth phase dependent concentrations of each suggest a form of transcriptional regulation supported by recent evidence which demonstrated increased recombinational products during the transition from log to stationary phase (Shearer and Summers, 2009). However, the four variants lacking a fully functional P_2 have an additional LexA dependent SOS response controlled mode of regulation recently demonstrated in *V. cholera* chromosome and *E. coli* plasmid integrons (Guerin *et al*., 2009). The benefit of this immediate response is apparent as the SOS response is induced by sub-lethal doses of antibiotics like trimethoprim, ciprofloxacin, and β-lactams (Lewin *et al*., 1991; Drlica *et al*., 1997; Miller *et al*., 2004; Butala *et al*., 2009). Together these proteins provide the integron an immediate situational response dependent on environmental stimuli, and/or a more gradual form of regulation dependent on growth phase.

ACKNOWLEDGEMENTS

We thank Anna Karls for providing purified FIS, IHF, and helpful advice regarding this work, Didier Mazel for providing both the parent and *lex* strains, Sylvie Rimsky and Colin Corcoran for providing purified H-NS and the H-NS overexpression plasmid. We also thank Natalie Strynadka for providing the LexA overexpression plasmid, and John Little and Ishita Mukerji for purified wild type LexA and IHF, respectively.

REFERENCES

Ali, B.M.J. Amit, R., Braslavasky, I., Oppenheim, A.B., Gileadi, O., and J. Stavans. (2001). Compaction of single DNA molecules induced by binding of integration host factor (IHF). *Proc. Natl. Acad. Sci*. USA. **98:** 10658-10663.

Arfin, S.M., Long, A.D., Ito, E.T., Tolleri, L., Riehle, M.M., Paegle, E.S., and G.W. Hatfield. (2000). Global gene expression profiling in *Escherichia coli* K12: The effects of integration host factor. *J. Biol. Chem*. **275:**29672-29684.

Azam, T. A., Iwata, A., Nishimura, A., Ueda, S., and A. Ishihama. (1999) Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol*. **181:**6361-6370.

Brown, H.J., Stokes, H.W. and R.M. Hall. (1996) The integrons In0, In2, and In5 are defective transposon derivatives. *J. Bacteriol*. **178:**4429-4437.

Butala, M, Zgur-Bertok, D., and S.J.W. Busby. (2009) The bacterial LexA transcriptional repressor. *Cell Mol. Life Sci.* **66:**82-83.

Colinon, C., Miriagou, V., Carattoli, A., Luzzaro, F., and G.M. Rossolini. (2007) Characterization of the IncA/C plasmid pCC416 encoding VIM-4 and CMY-4 β-lactamases. *J. Antimicrob. Chemother*. **60:**258-262.

Dame, R.T., Wyman, C., and N. Goosen. (2001) Structural basis for preferential binding of H-NS to curved DNA. *Biochimie*. **83:**231-234.

Dorman, C.J. (2009) Nucleoid-associated proteins and bacterial physiology. *Adv Appl Microbiol*. **67:**47-64

Drlica, K., and X. Zhao. (1997) DNA gyrase, topoisomerase IV, and the 4-quinolones. *Microbiol. Mol. Biol. Rev*. **61:**377-392.

Gama-Castro, S., Jimenez Jacinto, V., Peralta-Gil, M., Santos-Zavaleta, A., Peñaloza-Spindola, M., Contreras-Moreira, B., Segura-Salazar, J., Muñiz-Rascado, L., Martinez-Flores, I., Salgado, H., Bonavides-Martinez, C., Abreu-Goodger, C., Rodriguez-Penagos, C., Miranda-Rios, J., Morett, E., Merino, E., Huerta, A.M. and J. Collado-Vides. (2008) RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation**.** *Nucleic Acids Research*. **36:** Database issue:D120-4

Gonzalez-Zorn, B. Catalan, A., Escudero, J.A., Dominguez, L., Teshager, T., Porrero, C., and M.A. Moreno. (2005) Genetic basis for dissemination of armA. *J. Antimicrob. Chemother*. **56:**583-585.

Gottesman, S. (1984) Bacterial regulation: global regulatory networks. *Annu Rev Genet*. **18:**415-441.

Guerin, E., Cambray, G., Sanchez-Alberola, N., Campoy, S., Erill, I., Da Re, S., Gonzalez-Zorn, B., Barbé, J., Ploy, M., and D. Mazel. (2009) The SOS response controls integron recombination. *Science*. **324:**1034.

Hall, R.M., and C.M. Collis. (1995) Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Mol. Microbiol*. **15:**593-600.

Hall, R.M., and H.W. Stokes (1993) Novel DNA elements which capture genes by site-specific recombination. *Genetica* **90**: 115–132.

Hall, R.M. and C. Vockler. (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.

Hansson, K., Skold, O., and L. Sundstrom. (1997) Nonpalindromic attI sites of integrons are capable of site-specific recombination with one another and with secondary targets. *Mol. Microbiol*. **26:**441-453.

Kelley, W.L. (2006) Lex marks the spot: the virulent side of SOS and a closer look at the LexA regulon. *Mol. Microbiol*. **62:** 1228-1238.

Lagatolla, C., Edalucci, E., Dolzani, L, Riccio, M.L., De Luca, F., Medessi, E., Rossolini, G.M., and E.A. Tonin. (2006) Molecular evolution of metallo-β-lactamase-producing *Pseudomonas aeruginosa* in a nosocomial setting of high-level endemicity. *J. Clin. Microbiol.* **44:**2348-2353.

Lewin, C.S. and S.G. Amyes. (1991) The role of the SOS response in bacteria exposed to zidovudine or trimethoprim. *J. Med. Microbiol*. **34:** 329-332.

Liebert, C.A., Hall, R.M., and A.O. Summers. (1999) Transposon Tn*21*, flagship of the floating genome. *Microbiol Mol Biol Rev*. **63:**507-22.

Little, J.W., Mount, D.W., C.R., Yanisch-Perron. (1981) Purified lexA protein is a repressor of the recA and lexA genes. *Proc Natl Acad Sci*. **78:**4199-4203.

Luijsterburg, M.S., Noom, M.C., Wuite, G.J., and R.T. Dame.(2006) The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective *J Struct Biol*. **156:**262-72.

Mazón, G., Erill, J., Campoy, S., Cortés, P., Forano, E., and J. Barbé. (2004) Reconstruction of the evolutionary history of the LexA-binding sequence. *Microbiology.* **150:**3783-3795.

Miller, J. H. (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Miller, C., Thomsen, L.E., Gaggero, C. Mosseri, R., Ingmer, H. and S.N. Cohen. (2004) SOS response induction by beta-lactams and bacterial defense against antibiotic lethality. *Science.* **305:** 1629-1631.

Münch, R., Hiller, K., Barg, H., Heldt ,D., Linz, S., Wingender, E. and D. Jahn. (2003) PRODORIC: prokaryotic database of gene regulation. *Nucleic Acids Res.* **31**, 266-269.

Navarre, W.W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S.J., F.C. Fang. (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science*. **313:**236-238.

Papagiannitsis, C.C, Tzouvelekis, L.S., and V. Miriagou. (2009) Relative strengths of the Class 1 Integron promoter hybrid 2 and the combinations of strong and hybrid 1 with an active P2 promoter. *Antimicrob. Agents Chemother*. **53:**277-280.

Partridge, S.R, Brown, H.J., and R.M. Hall. (2002) Characterization and movement of the class 1 integron known as Tn*2521* and Tn*1405*. *Antimicrob Agents Chemother*. **46:**1288-1294.

Pournaras, S., Ikonomidis, A., Tzouvelekis, L.S., Tokatlidou, D., Spanakis, N., Maniatis, A.N., Legakis, N.J., and A. Tsakris. (2005) VIM-12, a novel plasmid-mediated metallo-β-lactamase from *Klebsiella pneumoniae* that resembles a VIM-1/VIM-2 hybrid. *Antimicrob. Agents Chemother*. **49:**5153-5156.

Schneider, K. and C.F. Beck (1986) Promoter-probe vectors for the analysis of divergently arranged promoters. *Gene*. **42:**37-48.

Shearer, J.E.S and A. O. Summers. (2009) Intracellular steady-state concentration of integron recombination products varies with integrase level and growth phase. *J Mol Biol.* **386:**316-331.

Stokes, H.W. and R.M. Hall. (1989) A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol. Microbiol*. **3:**1669-1683.

Stokes, H.W. and R.M. Hall. (1991) Sequence analysis of the inducible chloramphenical resistance determinant in the Tn*1696* integron suggests regulation by translational attenuation. *Plasmid.* **26:** 10-19.

Tolmasky, M.E. and J.H. Crosa. (1993) Genetic organization of antibiotic resistance genes (*aac(6')-lb*), *aadA*, and *oxa9*) in the multiresistance transposon Tn*1331*. *Plasmid.* **29:**31-40.

Yang, S.W. and H.A. Nash. (1995) Comparison of protein binding to DNA *in vivo* and *in vitro*: Defining an effective intracellular target. *EMBO J.* **14:**6292-6300.

Yildiz, F.H., Liu, X.S., Heydorn, A., and G.K. Schoolnik. (2004) Molecular analysis of rugosity in a *Vibrio cholerae* O1 El Tor phase variant. *Mol. Microbiol*. **53:**497-515.

Zhang, Z., Schwartz, S., Wagner, L., and W. Miller. (2000) A greedy algorithm for aligning DNA sequences. *J. Comput Biol.* **7:**203-14.

Zhang, X. and H. Bremer. (1995) Control of the *Escherichia coli rrnB* P1 promoter strength by ppGpp. *J Biol Chem.* **270:** 11181-11189.

Table 3.1 Strains, plasmids, and primers

Strains	**Genotype**	**Source[a]**
MG1656-197	*lacZ-* derivative of MG1655	Guerin *et al.*, 2009
MG1656-430	*lacZ-*, *ΔsulA*, *ΔlexA*	Guerin *et al.*, 2009
BW25113	*Δ(araD-araB)567*, *ΔlacZ4787*(::rrnB-3), *rph1*, *Δ(rhaD-rhaB)568*, *hsdR514*	CGSC
JW3229-1	*Δ(araD-araB)567*, *ΔlacZ4787*(::rrnB-3), *Δfis779::kan*, *rph1*, *Δ(rhaD-rhaB)568*, *hsdR514*	CGSC
JW1702-1	*Δ(araD-araB)567*, *ΔlacZ4787*(::rrnB-3), *ΔihfA786::kan*, *rph1*, *Δ(rhaD-rhaB)568*, *hsdR514*	CGSC
JW1225-2	*Δ(araD-araB)567*, *ΔlacZ4787*(::rrnB-3), *Δhns746::kan*, *rph-1*, *Δ(rhaD-rhaB)568*, *hsdR514*	CGSC
Plasmids	**Genotype**	
pCB267	Bi-directional (*phoA* and *lacZ*) transcription vector, Ap^R	Schneider and Beck, 1986
pWT	P_{int}, P_2, P_c reporter insertion; referred to as (P_{i2c}), derivative of pCB267	This study
pI2-17	P_{int} reporter insertion, $P_2$17bp spacer; referred to as (P_{i2}), derivative of pCB267	This study
pI2-14	P_{int} reporter insertion, $P_2$14bp spacer; referred to as ($P_{i2\text{-}14}$), derivative of pCB267	This study
pI	P_{int} reporter insertion; referred to as (P_i), derivative of pCB267	This study
Primers		
IP97L	5'-ATGACTAAGCTTTTGGGGTACAGTCTAT-3'	
IP134L	5'-GGTGGTAAGCTTGCAGTGGCGGTTTT-3'	
IPU	5'-TCGTTGGATCCCCATAACATCAAACAT-3'	
IP212U	5'-GGATCCCCACGGCGTAACG-3'	
IP212L	5'-AAGCTTACGAACCCAGTGGACAT-3'	
IP131U	5'-GGCATAGACTGTACAAAAAAACAG-3'	
IP131L	5'-CTGTTTTTTTGTACAGTCTATGCC-3'	

[a]CGSC, *E. coli* Genetic Stock Center; http://cgsc.biology.yale.edu/index.php

Table 3.2 Naturally occurring integron promoter combinations

Promoters	-35[a]	-10[b]	Location	Reference
Strong	TTGACA	TAAACT	R388 (In3)	Partridge *et al*., 2002
Strong + P_2	TTGACA TTGTTA	TAAACT TACAGT	P2873 (In-h12)	Pournaras *et al*., 2005
Weak	TGGACA	TAAGCT	pSCH884 (In5)	Brown *et al*., 1996
Weak + P_2	TGGACA TTGTTA	TAAGCT TACAGT	Tn*21* (In2)	Liebert *et al*., 1999
Hybrid 1	TGGACA	TAAACT	R46 (In1)	Hall and Vockler, 1987
Hybrid 1 + P_2	TGGACA TTGTTA	TAAACT TACAGT	pMUR050	Gonzalez-Zorn *et al*., 2005
Hybrid 2	TTGACA	TAAGCT	pCC416 (In-t4) chromosome (In7)	Colinon *et al*., 2007 Lagatolla *et al*., 2006

[a,b] -35 and -10 hexamers of cassette promoter variants as compared to consensus -35 (TTGACA) and -10 (TATAAT).

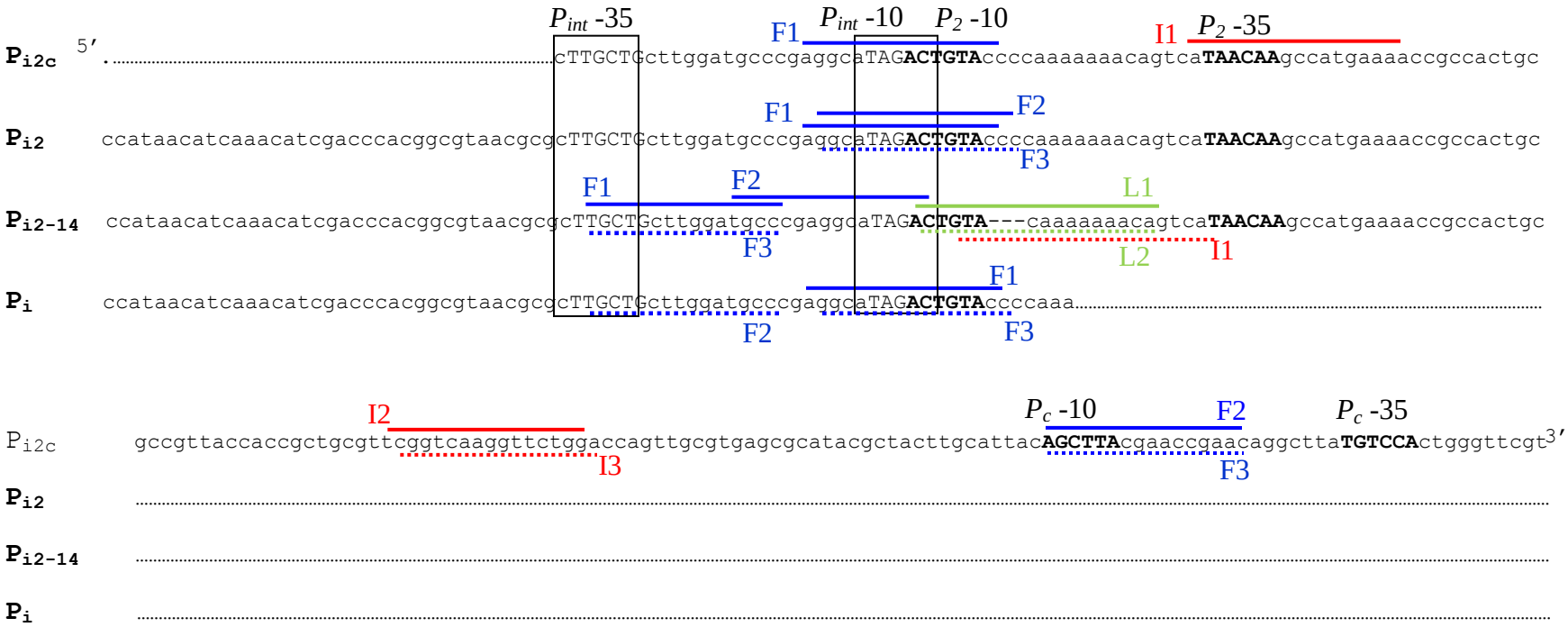

P_{int} -35
P_{int} -10
P_2 -10
P_2 -35
P_c -10
P_c -35
P_{i2c} 5′ ..... cTTGCTGcttggatgcccgaggcaTAGACTGTAccccaaaaaaacagtcaTAACAAgccatgaaaaccgccactgc
P_{i2} ccataacatcaaacatcgacccacggcgtaacgcgcTTGCTGcttggatgcccgaggcaTAGACTGTAccccaaaaaaacagtcaTAACAAgccatgaaaaccgccactgc
P_{i2-14} ccataacatcaaacatcgacccacggcgtaacgcgcTTGCTGcttggatgcccgaggcaTAGACTGTA---caaaaaaacagtcaTAACAAgccatgaaaaccgccactgc
P_i ccataacatcaaacatcgacccacggcgtaacgcgcTTGCTGcttggatgcccgaggcaTAGACTGTAccccaaa.....
P_{i2c} gccgttaccaccgctgcgttcggtcaaggttctggaccagttgcgtgagcgcatacgctacttgcattacAGCTTAcgaaccgaacaggcttaTGTCCActgggttcgt 3′
P_{i2} .....
P_{i2-14} .....
P_i .....
F1
F2
F3
L1
L2
I1
I2
I3

Figure 3.1 Predicted promoters and transcription regulator binding sites

Alignment of promoters P_{i2c}, P_{i2}, $P_{i2\text{-}14}$, P_i. P_{int}, -35 and -10 hexamers are boxed. Cassette promoters (P_2 and P_c) hexamers overlap and converge with P_{int} and are in boldface. Predicted FIS (blue), IHF (red), and LexA (green) binding sites identified by http://www.prodoric.de/ are numbered; those on the top strand are solid lines; those on the bottom strand are dotted lines. The 3C polymorphism in $P_{i2\text{-}14}$ is indicated by dashes. (Münch *et al.*, 2003)

Table 3.3 Probability scores[a] of predicted transcription regulator sites[b]

Transcription factor	Probability score	P_{i2C}	P_{i2}	P_{i2-14}	P_i
FIS	Max=3.56 Avg=2.87 N=126	F1=3.15 F2=3.15 (F3)=3.09	F1=2.93 F2=2.83 (F3)=2.88	F1=2.74 F2=2.74 (F3)=2.74	F1=3.01 (F2)=2.75 (F3)=2.84
IHF	Max=7.67 Avg=6.23 N=91	I1=5.84 I2=5.74 (I3)=5.89	NP	(I1)=5.63	NP
LexA	Max=12.92 Avg=11.45 N=55	NP	NP	L1=11.19 (L2)=12.50	NP

[a] Maximum and average score are calculated for each, N=total examples currently in database, predicted by http://www.prodoric.de/(Münch *et al*., 2003); parentheses, predicted on bottom strand; NP, not predicted.
[b] Sites are numbered as in Fig. 3.1.

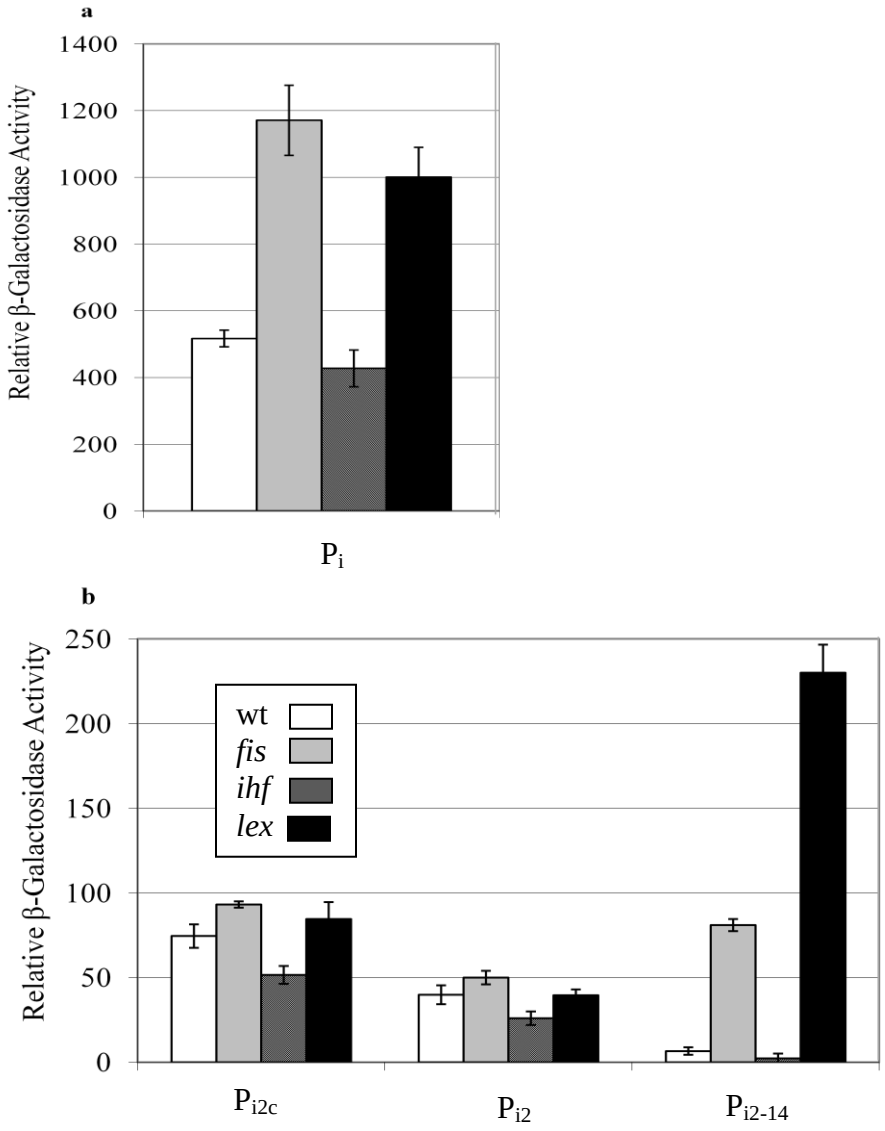


Figure 3.2 Expression from P_{int} in global regulator mutant strains

β-galactosidase expression relative to other integron promoters from **(a)** P_{int} alone (P_i); **(b)** P_{int} converging with cassette promoters P_c and P_2 (P_{i2c}), P_2 only (P_{i2}) or disabled P_2 with 14-bp spacer (P_{i2-14}) in wildtype (white), *fis* (light grey), *ihf* (dark grey), or *lex* (black) strains.

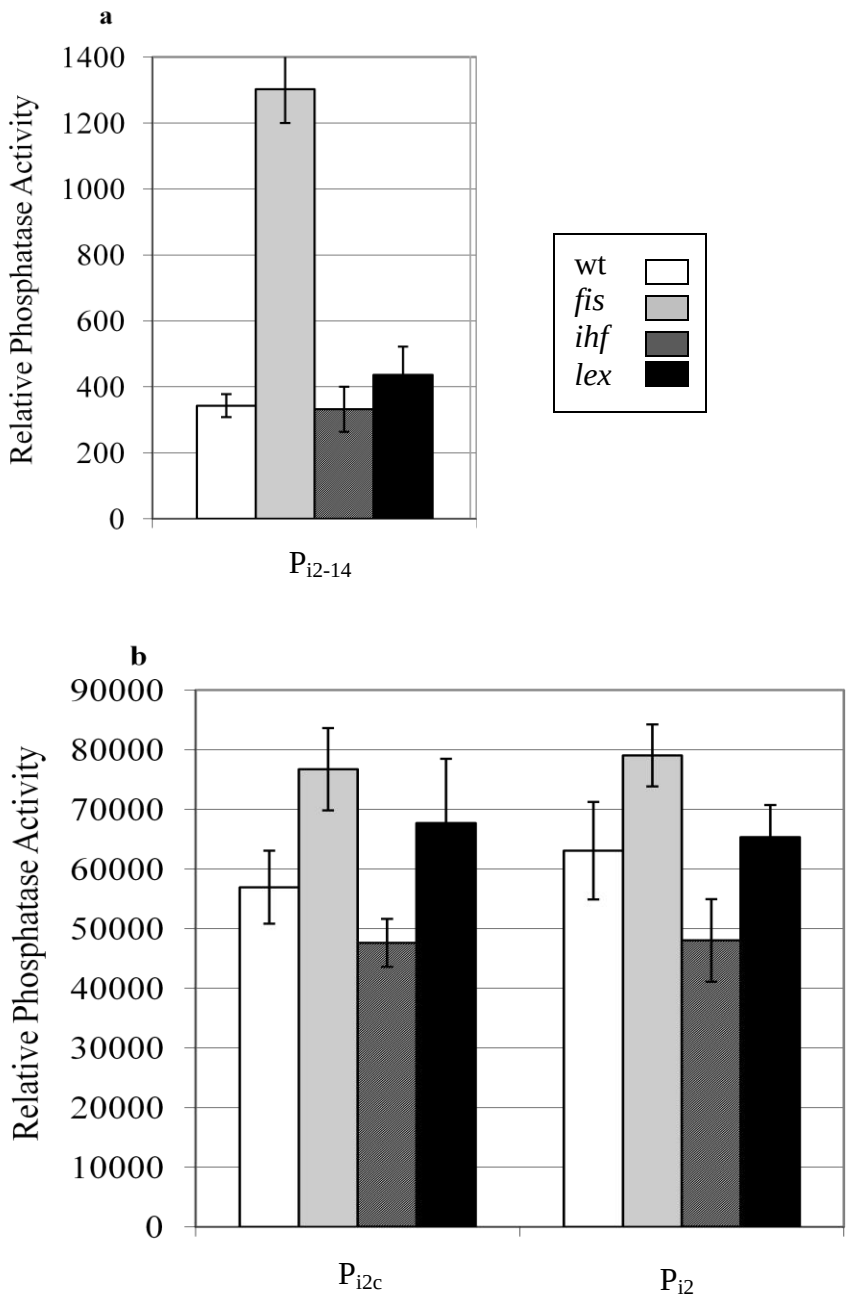


Figure 3.3 Expression from cassette promoters in global regulator mutant strains

Phosphatase expression relative to other integron promoters from **(a)** P_2 with 14-bp spacer ($P_{i2\text{-}14}$); **(b)** P_c and P_2 (P_{i2c}) and P_2 only (P_{i2}) in wildtype (white), *fis* (light grey), *ihf* (dark grey), or *lex* (black) strains.

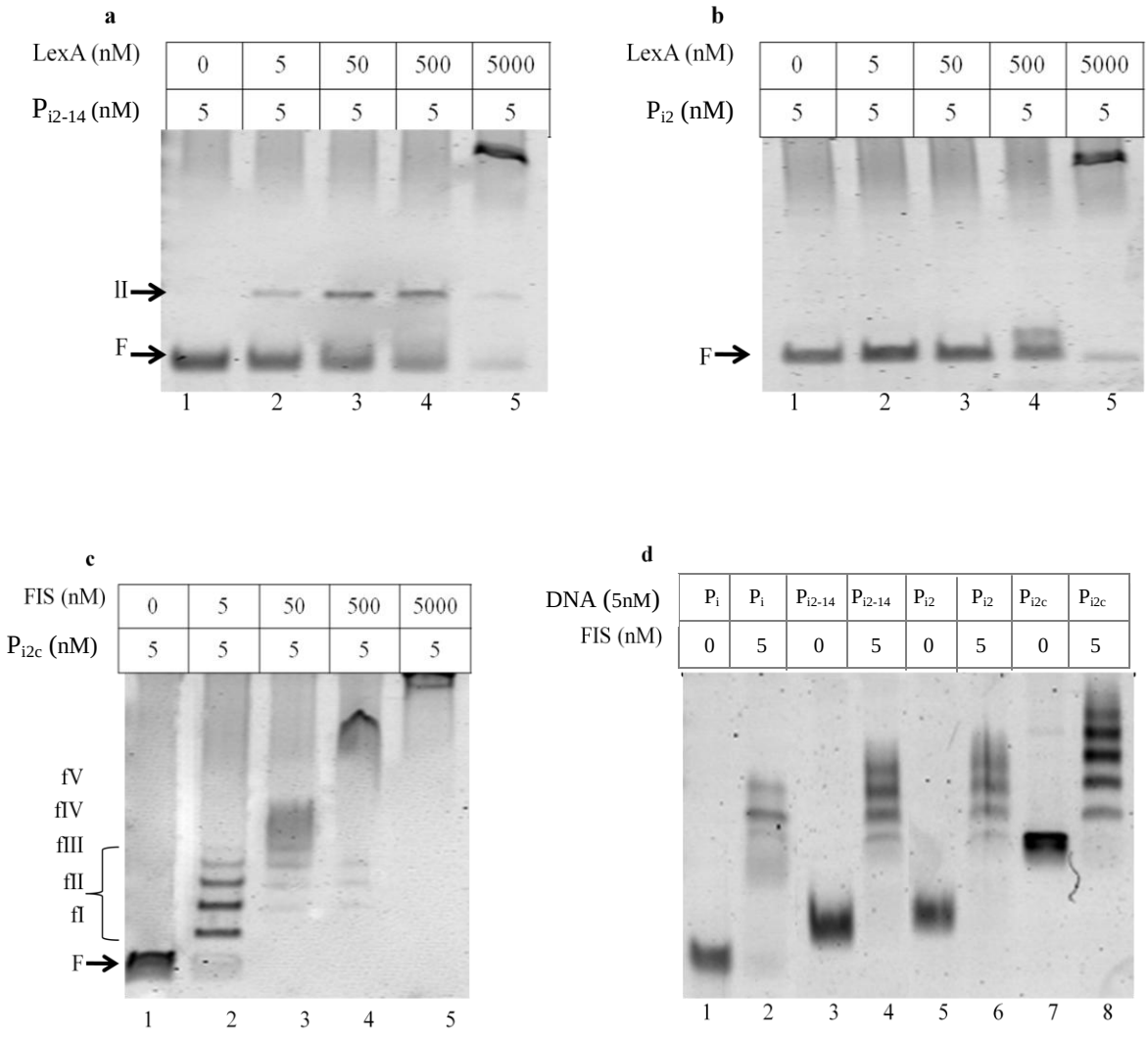


Figure 3.4 Binding of LexA and FIS to various integron promoter regions

LexA bound to **(a)** P_{i2-14}, lanes 2-5 or **(b)** P_{i2}, lanes 2-5. FIS bound to **(c)** P_{i2c}, lanes 2-5 and **(d)** DNA is 5 nM in each reaction and bound to P_i, lane 2; P_{i2-14}, lane 5; P*i2*, lane 6; and P_{i2c}, lane 8. Free (F) DNA is indicated by arrows; LexA complex lI indicated by arrow; FIS complexes fI, fII, fIII, fIV, fV are indicated by bracket.

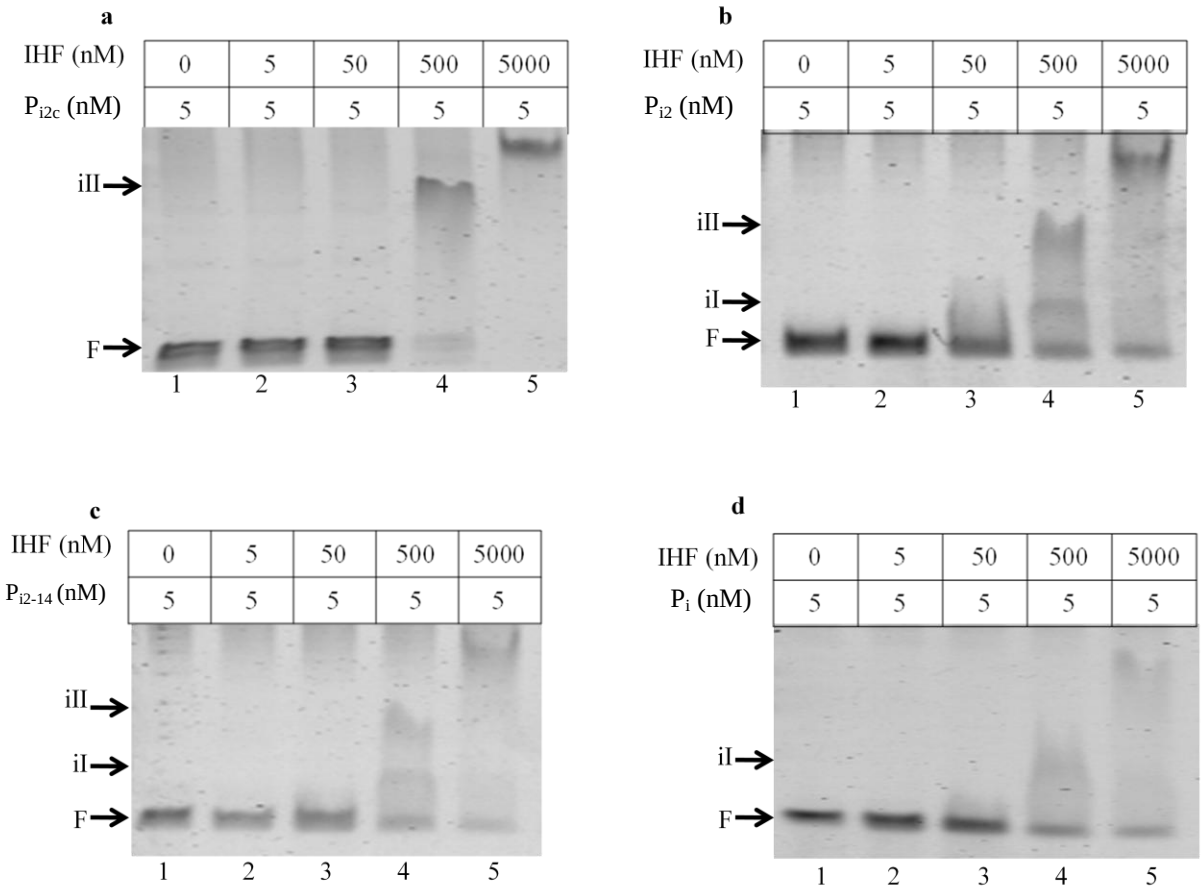


Figure 3.5 Binding of IHF to various integron promoter regions

IHF bound to **(a)** P_{i2c}, **(b)** P_{i2}, **(c)** $P_{i2\text{-}14}$, and **(d)** P_i. Free (F) DNA is indicated by arrows. IHF complexes iI and iII are indicated by arrows.

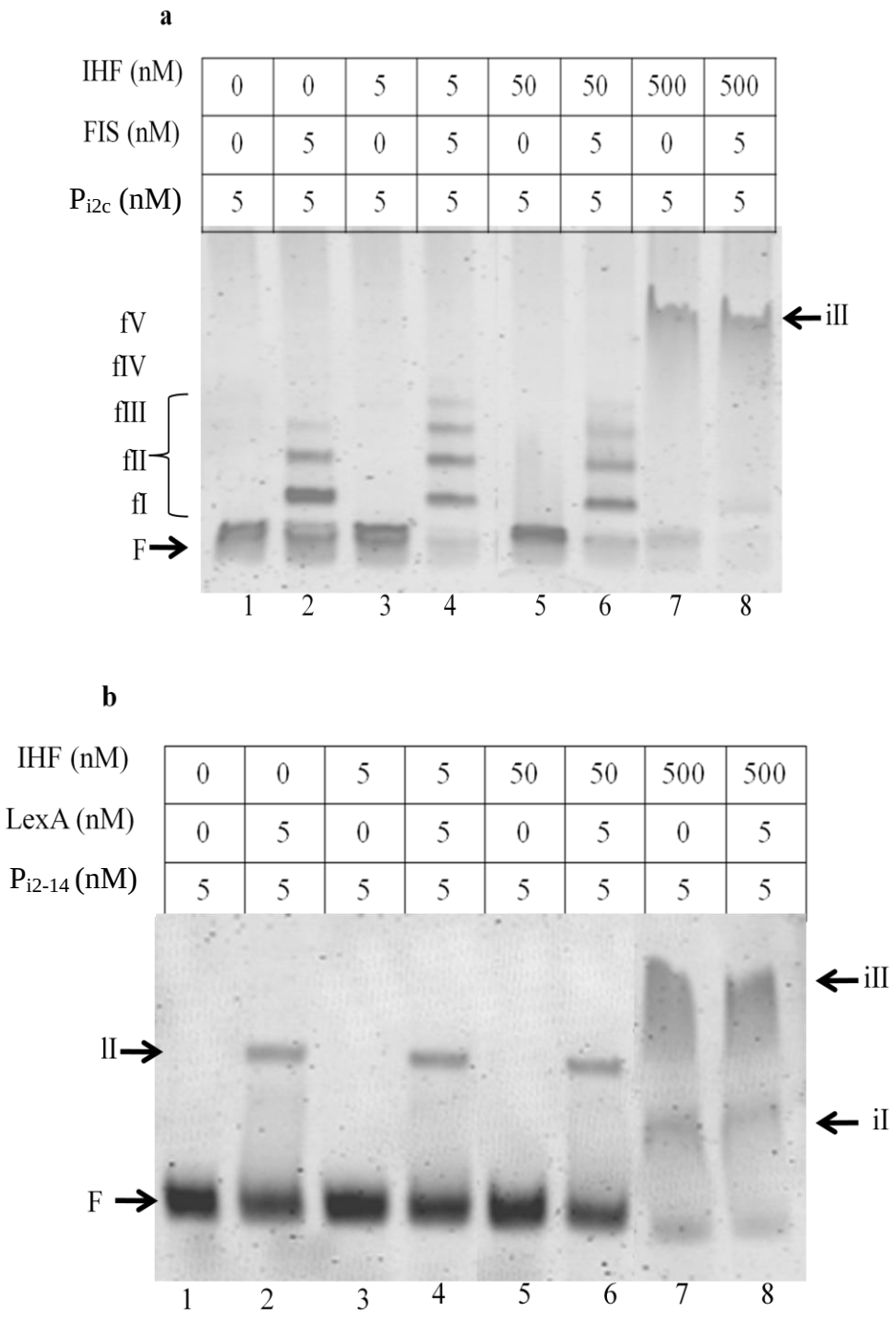


Figure 3.6 FIS and LexA compete with IHF for binding to promoter region

(a) FIS incubated with P_{i2c} and increasing IHF; **(b)** LexA incubated with $P_{i2\text{-}14}$ and increasing IHF. Free (F) DNA is indicated by arrows. IHF complexes iI and iII, and FIS complexes fI, fII, fIII, fIV are indicated by arrows and a bracket, respectively.

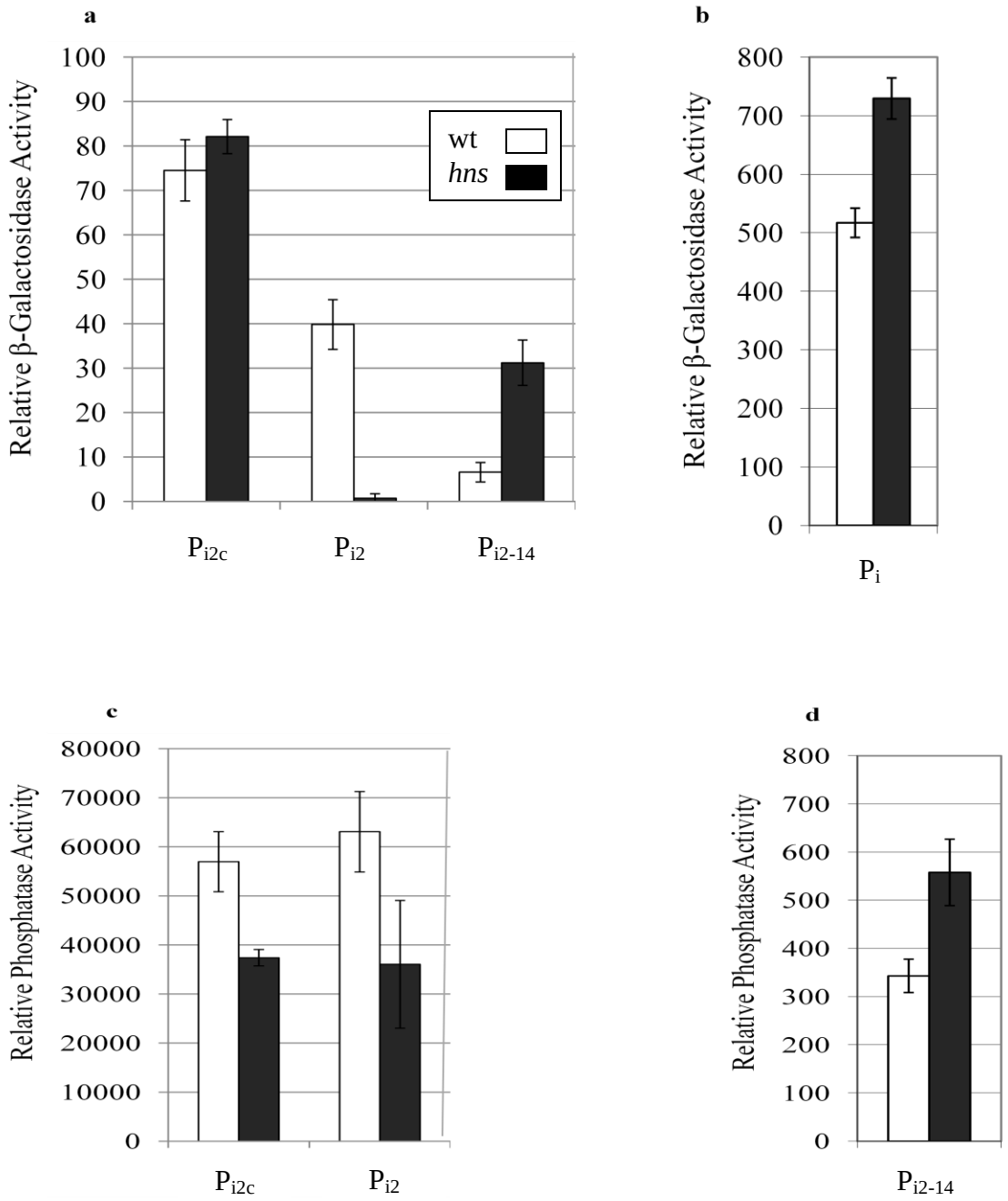


Figure 3.7 Effect of H-NS on P_{int}, and cassette promoters

β-galactosidase expression from P_{int} relative to converging cassette promoters **(a)** P_c and P_2 (P_{i2c}), P_2 only (P_{i2}) or P_2 with the 14-bp spacer (P_{i2-14}); **(b)** P_{int} alone. **(c)** Relative phosphatase expression from cassette promoters P_{i2c} or Pi2; **(d)** P_{i2-14}. Wildtype (white) and *hns* (black) hosts.

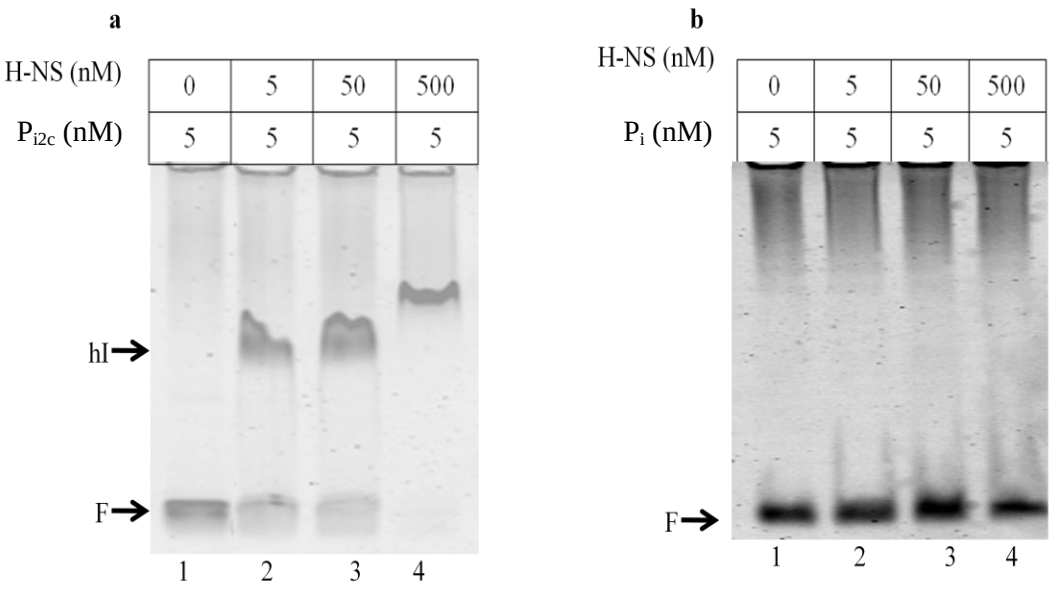


Figure 3.8 Selective binding of H-NS

H-NS interaction with **(a)** P_{i2c}, but not **(b)** P_i. Free (F) DNA and H-NS complex (hI) are indicated by arrows.

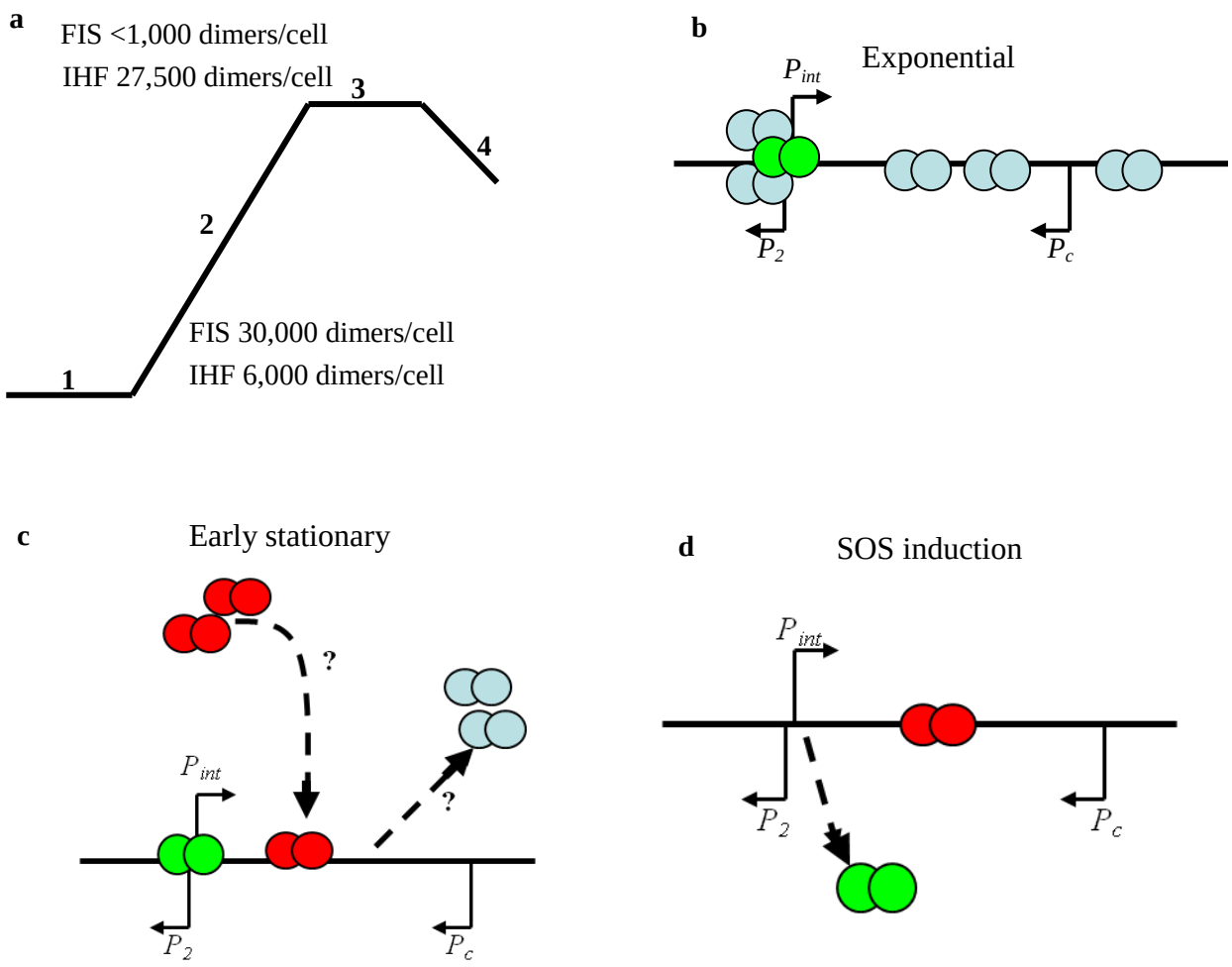


Figure 3.9 Model of integron convergent promoter regulation influenced by global regulators

(a) FIS and IHF concentrations during lag (1), exponential (2), stationary (3) and death (4) phase. **(b)** Exponential phase loading of transcription repressor FIS (blue) and LexA (green). **(c)** Early stationary phase IHF (red) increases, and displaces FIS. **(d)** SOS induction autoproteolysis of LexA and derepression of P_{int}.

Chapter 4

DISSERTATION SUMMARY

Integrons carry a genetic system that mediates the insertion and/or excision of varied gene cassettes (Stokes and Hall, 1989; Hall and Stokes, 1993). They occur among natural populations of Gram positive and Gram negative bacteria in clinical and environmental settings, and has facilitated the spread of antibiotic resistance (Goldstein *et al*., 2001; Nandi *et al*., 2004; Rowe-Magnus and Mazel, 2002). Reactions are carried out by a site-specific recombinase (IntI) integrase, and both this recombinase and the inserted gene cassettes are transcribed by the integron promoters P_{int} and P_c respectively (Lèvesque *et al*., 1994). Integrases belong to a larger family of tyrosine site-specific recombinases and are unique because they: (1) carry an additional catalytic domain not found in other members (Messier and Roy, 2001), (2) require little sequence specificity for site recognition (Recchia and Hall, 1995), and (3) differ among themselves in having seven variants of the cassette promoter (Papagiannitis *et al.*, 2009).

Chapter 2 describes my investigation of three forms of control that might contribute to the regulation of recombination itself, focusing on possible accessory protein involvement, on IntI protein allostery, and on culture growth phase. Several tyrosine recombinases use host-encoded accessory proteins like FIS and/or IHF to assist in recombination, but there are no known cofactors involved in the integron integrase mediated recombination (Hallet and Sherratt, 1997; Grindley *et al*., 2006). To address this possibility, I used protein pull-down assay based on

biotinylated PCR-amplified *attI* site attached to streptavidin-coated magnetic beads, but could not isolate accessory proteins or integrase. The detection limitations of this assay prevented it from answering the question of accessory protein involvement. Subsequently, the Mazel group reported an *attC*-VchIntIA co-crystal showing *attC* binding without accessory proteins. Their involvement in *attI* recognition remains an open question (Bouvier *et al*., 2005).

However, the crystal structure showed differing IntI conformations between the active *attC* bound and inactive integrases indicating site binding involves IntI allosteric change. My initial findings indicated that *attI* stabilized previously formed *attC*-IntI complexes, and pre-incubation of IntI with *attI* inhibited formation of *attC*-IntI complex supported IntI allosteric changes. My findings suggest *attI* stabilizes a preferred *attC*-IntI complex perhaps due to formation of the more stable *attI*-IntI complex (Demarre *et al*. 2007). However, if *attI* is pre-incubated with integrase, IntI binding of *attC* is partly inhibited perhaps due to formation of an IntI conformation specific to binding *attI* only. I also visualized recombination products between donor and recipient plasmids during the growth cycle, but qualitative PCR was not sensitive enough to detect changes in product abundance during the growth cycle. Later, quantitative dilution PCR done by another lab group member showed that recombination products increased through late log phase in both high and natural integrase expression, continued to rise during stationary phase in strains with high integrase, but decreased in the natural expression strain (Shearer and Summers, 2009).

Chapter 3 covers my work on regulation of integrase and cassette transcription. I used bi-directional transcriptional fusions measuring P_{int} expression of *lacZ* and cassette promoters expression of *phoA* determined the promoter strength of integrase promoter (P_{int}) alone and in competition with the natural cassette promoters (P_c and P_2) of In2, and also compared the

cassette promoters' strengths. P_{int} is a strong promoter, but interference from convergent cassette promoters decreases its expression 6-fold. Surprisingly, P_{int} is assisted by the more distant P_c whose removal decreases P_{int} expression by 45-47% of its maximum strength in wildtype. I found previously determined cassette promoters' relative strengths to be correct (Lèvesque *et al*., 1994; Papagiannitis *et al.*, 2009), but discovered that P_2 with the 14-bp spacer (predicted to be inactive) actually does function although at only 1% of the 17-bp version of this promoter.

I used the bi-directional transcription fusions to investigate the impact, if any, of predicted global transcription regulators in *fis*, *lex*, *ihf*, and *hns* strains and confirmed the direct interaction of each through electrophoretic mobility shift assays with purified FIS, LexA, IHF, and H-NS on promoter amplicands. Results indicated FIS and LexA repress all promoters, but the effect of LexA only occurs when the P_2 has the 14-bp spacer which is a LexA site no present in P_2 with 17-bp spacer. IHF is a weak activator of P_{int} transcription, but has no detectable role for the cassette promoters. H-NS activates cassette promoters with the P_2 17-bp spacer, and represses P_{int} and P_2 with the 14-bp spacer. The effects of these transcription factors on the different P_2 spacers, indicates distinct but related regulatory mechanisms among the seven cassette promoter versions. In addition, the growth phase dependence of FIS and IHF and the DNA damage dependence response action of LexA indicate two distinct global systems controlling integrase gene expression. It is premature to speculate how H-NS is involved. LexA provides integrons in the cell a rapid response to DNA damage. FIS, H-NS, and possibly IHF afford a more gradual control dependent on growth phase. However, either the DNA damage response evolved first in the integron, or it is strongly selected for because the promoter variant not controlled by LexA (i.e. P_2 with 17-bp) is less common in the Genbank database (Table 4.1; Zhang *et al*., 2001).

The work presented here opens a whole new dimension for understanding integron function. To obtain a more complete picture, message stability and translational regulation of both integrase and cassette transcripts must be investigated. Initial work investigating cassette translational control reported the potential for separate mechanisms dependent on the specific cassette, because 25% of gene cassettes do not have a plausible translation initiation region (TIR) consisting of the initiation codon, Shine-Dalgarno sequence, or spacer region (Hanau-Bercot *et al.*, 2002). These authors reported a short open reading frame (ORF-11) overlapping the *attI* site enhanced translation of aminoglycoside 6'-N-acetlytransferase by providing an appropriate TIR that the cassette did not have itself (Hanau-Bercot *et al.*, 2002). This work was done with the weak cassette promoter version; it would be interesting to use translational fusions to investigate other versions on cassettes with and without an appropriate TIR, as well as the translation of P_{int} transcripts.

In future work, the effect of DNA damage and nucleo-protein complexes on recombination should be examined. In recent work quantifying recombination products though the growth cycle (Shearer and Summers, 2009), the natural integrase was expressed by P_{int} facing the strong P_c promoter and the 14-bp P_2. So, integrase expression would have been repressed by LexA. Indeed, since these experiments were done in a *recA* background even DNA damage would not overcome LexA repression so P_{int} and P_2 with 14-bp would have been continually repressed.

Finally, investigation of the involvement of additional transcription factors should be pursued. During construction of transcription fusions used in chapter 3 experiments, two random point mutations at positions -36A and -38T (relative to the P_2 -35 hexamer, indicated in Fig. 4.1) occurred in separate otherwise wildtype constructs. The former resulted in a 66% decrease and

the latter in complete inhibition of cassette promoters' expression. -36A and -38T are in a predicted IHF site, but their changes are not predicted to be significant (Münch *et al.,* 2003). Results could be due to another unidentified factor.

BLAST searches also revealed a recurring mutation -12C (relative to the P_c -10 hexamer, indicated in Fig.4.1). Interestingly, -12C occurs naturally only in weak and hybrid 2 cassette promoter versions (Table 4.1). It forms a binding site for AlgU (Münch *et al.,* 2003), a sigma factor of *Pseudomonas aeroginosa* involved in stress response and alginate production (Martin *et al*., 1994). Further research in this area could offer insight into the possibility of multiple transcriptional regulatory mechanisms that have evolved over time using fairly non-specific ubiquitous host proteins in varied bacterial populations.

REFERENCES

Bouvier, M., Demarre, G., and D. Mazel. (2005) Integron cassette insertion: a recombination process involving a folded single strand substrate. *EMBO J.* **24:**4356-4367.

Demarre, G., Frumerie, C., Gopaul, D., and D. Mazel. (2007) Identification of key structural determinants of the IntI1 integron integrase that influence attC x attI1 recombination efficiency. *Nucleic Acids Research*. **35:**6475-6489.

Goldstein, C., Lee, M.D., Sanchez, S., Hudson, C., Phillips, B., Register, B., Grady, M., Liebert, C., Summers, A.O., White, D.G., and J. J. Maurer. (2001) Incidence of class 1 and 2 integrases in clinical and commensal bacteria from livestock, companion animals, and exotics. *Antimicrob. Agents Chemother*. **45:**723-726.

Grindley, N.D. F., Whiteson, K.L., and P.A. Rice. (2006) Mechanisms of Site-Specific Recombination. *Annual Review of Biochemistry*. **75:**567-605.

Hall, R.M., and H.W. Stokes (1993) Novel DNA elements which capture genes by site-specific recombination. *Genetica* **90**: 115–132.

Hallet, B., and D.J. Sherratt. (1997) Transposition and site-specific recombination: adapting DNA cut-an-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiology Reviews.* **21:**157-178.

Hanau-Bercot, B., Podglajen, I., Casin, I., and E. Collatz. (2002) An intrinsic control element for translational initiation in class 1 integrons. *Mol. Microbiol*. **44:**119-130.

Lévesque, C., Brassard, S., Lapointe, J., and P.H. Roy. (1994) Diversity and relative strengths of tandem promoters for the antibiotic-resistance genes of several integrons. *Gene.* **142:**59-54.

Martin, D.W., Schurr, M.J, Yu, H., and V. Deretic. (1994) Analysis of promoters controlled by the putative sigma factor AlgU regulating conversion to mucoidy in *Pseudomonas aeruginosa*: relationship to sigma E and stress response. *J. Bacteriol*. **176:**6688-6896.

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

Münch, R., Hiller, K., Barg, H., Heldt ,D., Linz, S., Wingender, E. and D. Jahn. (2003) PRODORIC: prokaryotic database of gene regulation. *Nucleic Acids Res.* **31**, 266-269.

Nandi, S., Maurer, J.J., Hofacre, C., and A.O. Summers. (2004) Gram-positive bacteria are a major reservoir of Class 1 antibiotic resistance integrons in poultry litter. *Proc. Natl. Acad. Sci.* **101:**7118-7122.

Papagiannitsis, C.C, Tzouvelekis, L.S., and V. Miriagou. (2009) Relative strengths of the Class 1 Integron promoter hybrid 2 and the combinations of strong and hybrid 1 with an active P2 promoter. *Antimicrob. Agents Chemother*. **53:**277-280.

Recchia, G.D., and R.M. Hall. (1995) Gene cassettes: a new class of mobile elements. *Microbiology.* **141**:3015-3027.

Rowe-Magnus, D.A., and D. Mazel. (2002) The role of integrons in antibiotic resistance gene capture. *Int. J. Med. Microbiol*. **292:**115-125.

Shearer, J.E.S and A. O. Summers. (2009) Intracellular steady-state concentration of integron recombination products varies with integrase level and growth phase. *J Mol Biol.* **386:**316-331.

Stokes, H.W. and R.M. Hall. (1989) A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol. Microbiol*. **3:**1669-1683.

Zhang, Z., Schwartz, S., Wagner, L., and W. Miller. (2000) A greedy algorithm for aligning DNA sequences. *J. Comput Biol.* **7:**203-14.

Table 4.1 BLAST search comparison of the percentage occurrence and sample source of integron cassette promoter versions

Cassette Promoter	Strength (%)[a]	Occurrence (%)[b]	-12C allele present (%)[c]	Environmental[d]
Strong	650	20	0	No
Weak	20	28	20	Yes
Hybrid 1	70	42	0	No
Weak+2nd	320	4	0	No
Hybrid 2	320	4	20	Yes
Hybrid 1+2nd	250	1	0	No
Strong+2nd	780	<1	0	No

[a] The strength of each promoter version has been normalized to derepressed P_{tac} and indicated as a percent of derepressed P_{tac} expression (Lévesque *et al*., 1994; Papagiannitis *et al*., 2009).
[b,c] BLAST search using nucleotide collection nr/nt (all GenBank + RefSeq Nucleotides + EMBL + DDBJ + PDB sequences excluding HTGS0,1,2, EST, GSS, STS, PAT, WGS. No longer "non-redundant"), http://www.ncbi.nlm.nih.gov/, (Zhang *et al*., 2000).
[d] BLAST search using environmental samples (env_nt) **(**nucleotide sequences from environmental samples, including those from Sargasso Sea and Mine Drainage projects), http://www.ncbi.nlm.nih.gov/, (Zhang *et al*., 2000)

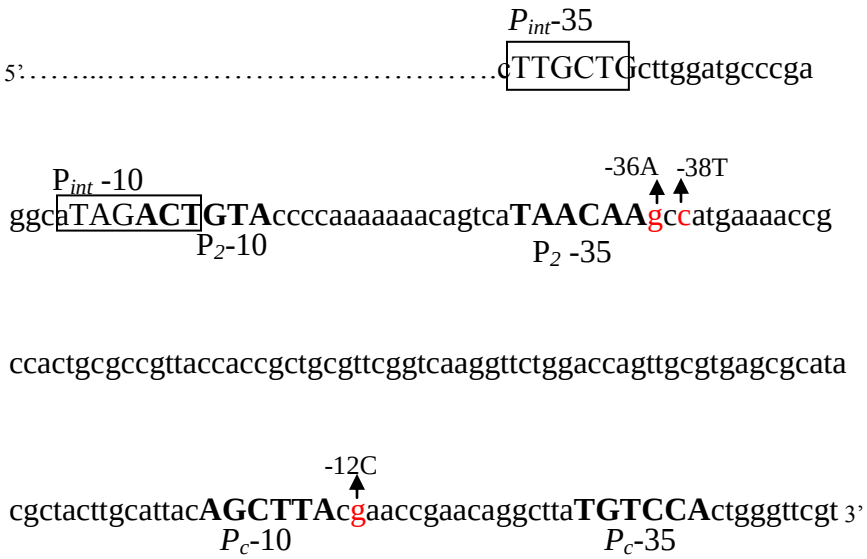


Figure 4.1 Point mutations in the integron promoter region

P*int* -35 and -10 hexamers are boxed. Cassette promoters (P_2 and P*c*) hexamers reading to the left (5') are bolded. Mutations are shown in red with arrow; -36A and -38T refer to allele change and position relative to the P_2 -35 hexamer in boldface. -12C refers to allele change and position relative to the P_c -10 hexamer in boldface.